

Paritat de gènere a la premsa local digital de Mallorca:

el cas de *Diariodemallorca.es* i *Ultimahora.es*

UOC

Treball final de màster

Aitor Francisco Vallespir Rivas

Màster universitari en periodisme i comunicació digital: dades i noves narratives

Curs acadèmic 2023-2024

Universitat Oberta
de Catalunya



Fitxa resum del treball final de màster

Títol: Paritat de gènere a la premsa local digital de Mallorca: el cas de *Diariodemallorca.es* i *Ultimahora.es*

Modalitat: Professionalitzadora

Autor: Aitor Francisco Vallespir Rivas

Directora del treball: Victòria Oliveres Torrecassana

Professora responsable: Candela Ollé Castellà

Titulació: Màster universitari en periodisme i comunicació digital: dades i noves narratives

Universitat: Universitat Oberta de Catalunya (UOC)

Curs acadèmic: 2023-2024

Data d'entrega: 8 de gener de 2024

Idioma: Català

Resum:

L'estudi examina la paritat de gènere a la premsa local digital de Mallorca, centrant-se en *Diariodemallorca.es* i *Ultimahora.es*. Mitjançant programació avançada i eines d'intel·ligència artificial, la investigació automatitza l'anàlisi del contingut de la premsa digital, examinant la representació de les dones en aquests mitjans de comunicació. La metodologia de l'estudi integra de manera innovadora la intel·ligència artificial per a l'anàlisi textual, contribuint en eficiència en temps i recursos per obtenir informació valuosa sobre la representació de gènere en el periodisme. Els resultats posen de manifest l'estat actual de la paritat de gènere als principals mitjans de premsa digital de Mallorca i ofereixen un enfocament nou per a l'anàlisi de mitjans, que combina la recerca acadèmica i les aplicacions pràctiques en el periodisme. Les troballes de l'estudi es publiquen a una pàgina web de format periodístic en la qual, mitjançant nou notícies i setze infografies interactives, es fa difusió de les dades més rellevants. El web es complementa amb una secció de dades obertes per posar a l'abast de tothom les microdades, les estadístiques i el codi font creat per al projecte.

Paraules clau:

paritat de gènere, igualtat de gènere, dones als mitjans, representació de les dones, anàlisi dels mitjans, anàlisi textual, disparitats de gènere en el periodisme, IA al periodisme, transformació dels mitjans digitals, tecnologia periodística, lideratge periodístic i gènere

Título en español: Paridad de género en la prensa local digital de Mallorca: el caso de *Diariodemallorca.es* y *Ultimahora.es*

Resumen: El estudio examina la paridad de género en la prensa local digital de Mallorca, centrándose en *Diariodemallorca.es* y *Ultimahora.es*. Mediante programación avanzada y herramientas de inteligencia artificial, la investigación automatiza el análisis del contenido de la prensa digital, examinando la representación de las mujeres en estos medios de comunicación. La metodología del estudio integra de forma innovadora la inteligencia artificial para el análisis textual, contribuyendo en la eficiencia del tiempo y los recursos para obtener información valiosa sobre la representación de género en el periodismo. Los resultados ponen de manifiesto el estado actual de la paridad de género en los principales medios de prensa digital de Mallorca y ofrecen un enfoque novedoso para el análisis de medios, que combina la investigación académica y las aplicaciones prácticas en el periodismo. Los hallazgos del estudio se publican en una página web de formato periodístico en la que, mediante nueve noticias y dieciséis infografías interactivas, se hace difusión de los datos más relevantes. La web se complementa con una sección de datos abiertos para poner al alcance de todo el mundo los microdatos, las estadísticas y el código fuente creado para el proyecto.

Palabras clave: paridad de género, igualdad de género, mujeres en los medios, representación de las mujeres, análisis de medios, análisis textual, disparidades de género en el periodismo, IA en el periodismo, transformación de los medios digitales, tecnología periodística, liderazgo periodístico y género

English Title: Gender Parity in Mallorca's Local Digital Press: The Case of *Diariodemallorca.es* and *Ultimahora.es*

Abstract: This study examines gender parity in Mallorca's local digital press, focusing on *Diariodemallorca.es* and *Ultimahora.es*. Using advanced programming and artificial intelligence tools, the research automates the analysis of digital press content, scrutinizing the representation of women in these media outlets. The study's methodology innovatively integrates AI for textual analysis, contributing in time and resources efficiency to obtain valuable insights into gender representation in journalism. The findings highlight the current state of gender parity in the main Mallorca's digital press outlets and offer a novel approach to media analysis, blending academic research and practical journalism applications. The results of the study are published on a journalistic-style website where, through nine news stories and sixteen interactive infographics, the most relevant data is disseminated. The website is complemented by an open data section to make the microdata, statistics and source code created for the project available to everyone.

Keywords: gender parity, gender equality, women in media, women's representation, media analysis, textual analysis, gender disparities in journalism, AI in journalism, digital media transformation, journalism technology, journalism leadership and gender

Índex

1. Introducció	5
2. Objectius i justificació	6
3. Marc teòric.....	7
3.1. De l'ús d'eines digitals i intel·ligència artificial a l'àmbit de la comunicació	7
3.2. Dels estudis previs sobre representació de gènere als mitjans	8
3.3. Dels referents per a l'anàlisi i visualització de dades	9
4. Metodologia	11
4.1. Selecció de la mostra.....	11
4.2. Recopilació i tractament de les dades	12
4.3. Definició de les variables d'anàlisi.....	12
4.4. Visualització de les dades	13
5. Procés d'elaboració.....	15
5.1. Consideracions prèvies.....	15
5.2. Creació dels marcs de dades amb les notícies	15
5.3. Proves amb els models d'OpenAI.....	17
5.4. Creació d'altres marcs de dades complementaris	19
5.5. Errades amb el codi de l'anàlisi i errades en les respostes de la IA	20
5.6. Anàlisi automatitzada de les fotografies.....	21
5.7. Depuració de les dades recuperades via IA	22
5.8. Comprovació de la fiabilitat de l'anàlisi automatitzada amb IA	24
5.9. Creació d'indicadors estadístics	27
5.10. Elaboració dels reportatges, el web i les visualitzacions de dades.....	30
5.11. Costos del projecte i temps de processament	33
6. Conclusions	35
6.1. Respecte dels objectius i la justificació del projecte.....	35
6.2. Respecte de la metodologia i el procés de treball	35
6.3. Respecte del projecte professionalitzador	36
7. Línies de desenvolupament futur	37
8. Bibliografia	40
9. Annexos.....	43

1. Introducció

Molts acadèmics han establert la relació entre els mitjans de comunicació de masses i la construcció de la realitat a través de la paraula publicada (Gamson et al., 1992, p. 387; McLuhan, 1966; Stamenković, 2015; Vogt, 2020, per citar-ne alguns exemples). En aquest sentit, la perspectiva de gènere és un altre dels factors en els quals els mitjans són determinants en els seus vessants informatiu i formatiu (Bach Arús et al., 2000; Gallego Ayala, 2014).

El treball té la intenció d'analitzar i difondre si es compleix amb la paritat de gènere en els dos principals mitjans digitals locals de Mallorca. La idea sorgeix a partir d'estudis anteriors que tractat la qüestió. Tot i això, aquest exercici no vol ser una repetició d'informes passats ni una simple actualització de les dades.

El plantejament d'aquest treball final de màster és, d'una banda, fer una revisió de com es pot facilitar l'anàlisi de gènere sobre allò publicat i, de l'altra, aportar noves eines per a aquesta tasca. Amb aquesta finalitat, es pretén aprofundir en els avantatges que les tecnologies de la informació (i més concretament, els llenguatges de programació i la intel·ligència artificial) proporcionen als periodistes a l'hora de recopilar dades per a històries periodístiques.

Principalment, la recerca d'informació se centrarà en l'anàlisi quantitativa i automatitzada dels textos periodístics mitjançant eines de programació que facilitin l'extracció de les dades, la transformació i l'estudi de les variables d'interès.

Un cop obtingudes les dades, s'explotaran en un format híbrid: un conjunt d'articles periodístics dins d'una pàgina web que inclourà infografies interactives per tal d'explorar els conjunts de dades. La finalitat és que aquestes eines permetin abastar diversos públics i millorar així la difusió de les conclusions.

2. Objectius i justificació

Els principals objectius d'aquest treball final de màster són:

- Desenvolupar un codi de programació reutilitzable que permeti aconseguir dades de premsa escrita digital.
- Automatitzar tant com sigui possible l'anàlisi de les dades recopilades d'acord amb les variables d'estudi que es determinin.
- Esbrinar el grau de fiabilitat de les eines d'intel·ligència artificial per tal de contribuir a l'anàlisi de textos periodístics.
- Estudiar si ha augmentat la representació de les dones als mitjans de Mallorca.
- Publicar les dades de la recerca de forma comprensible a través d'una plana web.

A més, també es plantegen com a objectius secundaris controlar les despeses del projecte per reduir els costos derivats de l'ús d'eines d'intel·ligència artificial i determinar el temps de processament automatitzat necessari per a un conjunt de notícies de premsa digital.

El producte final per lliurar serà una pàgina web amb un reportatge periodístic i un conjunt d'infografies interactives, a més del codi de programació comentat per tal de compartir-lo en obert i posar-lo a l'abast de tothom.

La justificació d'aquest treball final de màster es basa en les següents motivacions que en suporten l'oportunitat:

1. *Necessitat d'eines automatitzades i reutilitzables:* El periodisme digital està en evolució constant i necessita poder recopilar i analitzar dades de manera eficient. La realització d'un codi programable i reutilitzable no només beneficia aquest projecte, sinó que també pot servir com una eina valuosa per a futurs estudis sobre mitjans o per a usos periodístics.
2. *Avanços en la intel·ligència artificial:* L'ús de la intel·ligència artificial per a l'anàlisi de textos pot facilitar l'estudi sobre articles periodístics. Ben utilitzada, aquesta tecnologia pot servir per a una anàlisi més profunda i pot contribuir a observar patrons i tendències que no sempre són evidents.
3. *Conscienciar socialment sobre la representació de les dones als mitjans:* Una altra de les motivacions del projecte és estudiar si la representació de les dones als mitjans de comunicació de Mallorca ha millorat amb els anys.
4. *Contribució professional al periodisme:* Aquest treball vol contribuir al periodisme amb l'exploració de nous mètodes per recopilar, analitzar i mostrar dades de manera atractiva. A més, s'atenen les premisses d'eficiència en els recursos i reducció de costos, factors claus per a la sostenibilitat de projectes periodístics digitals.

En resum, aquest treball final de màster busca contribuir a l'estudi de les qüestions de gènere i, al mateix temps, proporcionar eines pràctiques i innovadores que poden ser aplicades en el món laboral. Atès el vessant professionalitzador i sense oblidar l'àmbit acadèmic propi de la universitat, el fet de combinar recerca i aplicació pràctica volen garantir que el projecte tingui una rellevància i utilitat significatives en el periodisme digital actual.

3. Marc teòric

3.1. De l'ús d'eines digitals i intel·ligència artificial a l'àmbit de la comunicació

La digitalització de les redaccions dels mitjans ha hagut d'anar acompanyada en molts casos de formació per als professionals de la comunicació. Tot sovint, en programari nou o en competències digitals bàsiques. Altres vegades, respecte dels processos d'elaboració i tractament de la informació (Bunce, 2019).

Com expliquen Luciana i Denis Renó, “aquests nous models i eines per a la producció de notícies, com ara el *big data*, han provocat canvis en la configuració de les redaccions de diversos països, malgrat el conservadorisme característic del periodisme en general” (Renó i Renó, 2015). Tot i que els mitjans cada vegada demanden professionals més polivalents (Mayo-Cubero, 2022), són comptades les experiències capdavanteres en el si de les empreses mediàtiques.

Algunes redaccions ja han incorporat l'ús de la intel·ligència artificial (IA) i, més en particular la IA generativa. Un cas comú és la redacció automatitzada de notícies, una experiència que de vegades s'han fet famosa pels errors que han inclòs aquests textos (Drummond, 2023) i per a la qual les empreses tecnològiques n'ultimen nous models (Mullin i Grant, 2023).

Però les experiències no queden aquí. Segons el recent informe, les tasques encomanades a aquestes tecnologies han anat més enllà per a certs mitjans:

“En general, l'ús exitós de les tecnologies d'IA va variar molt entre els enquestats, tot i que un gran nombre van considerar avantatjoses les eines de transcripció i edició d'àudio. El raspap (scraping) de pàgines web, el monitoratge de xarxes socials, la generació d'imatges, els sistemes de recomanació i altres eines de distribució també es van esmentar com a aplicacions d'IA d'èxit” (Beckett i Yaseen, 2023, p. 22).

En els darrers anys, han tingut lloc avenços significatius en la recerca sobre la capacitat de les IA quant a la comprensió lectora automàtica i el processament del llenguatge natural. Més concretament, i en relació amb l'àmbit del present treball, en la lectura de manera comprensiva de notícies i altres documents per després respondre preguntes sobre ells. En són exemples:

- *Deepmind i la Universitat de Stanford*: El projecte Deepmind de Google va fer entrenar un algoritme per llegir i entendre notícies. En el seu cas, varen emprar com a referència textos del *Daily Mail* i la CNN. Posteriorment, un grup de recerca de la Universitat de Stanford va millorar-ne l'algoritme per poder respondre preguntes amb major precisió (Gearin, 2016).
- *Els models que han desafiat el Conjunt de Dades per Respondre Preguntes de Stanford o SQuAD*: El 2016 es va crear un conjunt de més de 100.000 preguntes que podien ser respostes amb els coneixements de la Viquipèdia, anomenat Stanford Question Answering Dataset o SQuAD (Rajpurkar et al., 2016). El 2018 Microsoft i Alibaba varen ser les dues primeres companyies que varen dissenyar models que varen superar

lleugerament la puntuació humana per al mateix conjunt de preguntes (Linn, 2018). Des d'ençà, el qüestionari s'ha actualitzat a la segona versió. Actualment, hi ha una trentena de models de gran llenguatge d'IA capaços de superar la puntuació humana (Rajpurkar, 2023).

- *Semantic Scholar*: La iniciativa Semantic Scholar, de l'Allen Institute for Artificial Intelligence, resumeix i classifica articles acadèmics per tal d'identificar els més adients per a una consulta (Woolston i Perkel, 2020).

Aquests avenços mostren les creixents capacitats de la IA per entendre i processar informació textual complexa. Això no obstant, els desenvolupaments evolucionen envers la millora de la precisió, fiabilitat i aplicabilitat d'aquests models.

De fet, més recentment, les eines d'IA han assolit un abast generalista des de l'obertura al públic de ChatGPT a finals de novembre de 2022 (OpenAI, 2022). En tan sols, un any el ressò d'aquest model d'IA generativa ha modificat les rutines de treball de molts sectors, inclòs el de la comunicació. En l'àmbit acadèmic, se n'han explorat les possibilitats per a recomanar notícies (Li et al., 2023) o per a la verificació d'articles periodístics (Caramancion, 2023).

En aquest sentit, un dels objectius d'aquest treball serà veure el nivell de fiabilitat de les respostes obtingudes pels models generatius preentrenats (GPT, per les sigles en anglès) de la companyia OpenAI.

3.2. Dels estudis previs sobre representació de gènere als mitjans

Els mitjans de comunicació es configuren com a actors que intervenen en la creació de la realitat i de les identitats socials (Giró, 1999). El fet de tenir ressò als mitjans publicats permet establir una primacia de certes opinions i discursos, així com construir referents socials per als lectors, oients o espectadors. Tot sovint, per les presses de les rutines d'elaboració de la informació, els professionals de la comunicació s'obliden "del poder que això representa i de l'impacte que poden tenir els seus enfocaments en la construcció de la realitat" (Buades Àvila, 2018, p. 14).

Així mateix, les persones que formen part d'una redacció trien els marcs mentals amb els quals expliquen el relat al seu públic. Aquest procés de *framing* s'ha explicat pels estudiosos com "el procés de selecció, interpretació i avaluació d'aspectes particulars de la realitat percebuda (Comas d'Argemir, 2015, p. 122) o bé com "l'angle que adopten els i les periodistes per abordar una història" (Kitzinger, 2004, p. 15).

No cal oblidar que a l'escenari mediàtic es creen uns rols socials (Goffman, 2006), al mateix temps que els mitjans tenen el poder de marcar les regles d'aquests rols socials. "Tenen el poder d'exposar nous problemes, de redefinir la naturalesa d'un tema, de popularitzar noves formes discursives, de generar espais de creació de noves identitats, de mantenir falses associacions d'idees, de construir estereotips... Són un element fonamental per construir la idea que tenim del món" (Kitzinger, 2004, citat per Buades Àvila, 2018, p. 15-16).

En aquest sentit, alguns estudis sobre la presència de les dones als mitjans de comunicació han conclòs que la representació de les dones als mitjans és "escassa i força estereotipada" i

que “els discursos dels mitjans de comunicació tendeixen a negar la veu de les dones com a subjectes de la informació, la qual cosa significa que la imatge del món [...] és predominantment masculina” (Altés Rufias, 2014, p. 12).

Per aquest motiu, des de l'any 1995, es realitza el Global Media Monitoring Project (GMMP). Es defineix com “l'estudi longitudinal més extens i durador sobre gènere als mitjans de comunicació del món” (WACC Global, 2021). Compta amb el suport d'ONU Dones i cada cinc anys monitoren els mitjans a més d'un centenar de països a través de grups locals de voluntaris, generalment vinculats al sector acadèmic.

El darrer informe, que data de 2020, mostra que, de totes les persones esmentades a les notícies dels mitjans i països analitzats, només un 25% són dones. L'Orient Mitjà és el territori on menys s'esmenten les dones (17% del total de subjectes). A més, les notícies relacionades amb política són les que tenen una menor presència de persones del sexe femení, amb un 20% sobre el total (Macharia et al., 2020, p. 20-22).

El GMMP ha servit de base per a dos exercicis posterior d'àmbit regional, que al mateix temps constitueixen els precedents més semblants al que es pretén assolir amb aquest treball. Són, d'una banda, el document *Representació i tractament de les dones i el gènere en la informació sobre economia* (Altés Rufias i Majó Vázquez, 2012) i, de l'altra, l'informe *Un dia a les notícies en clau de gènere: protagonistes i periodistes als mitjans de comunicació de les Illes Balears* (Altés Rufias, 2014).

Tot i que abasten diferents àmbits temporals i geogràfics, els tres treballs volen analitzar la presència mediàtica de les dones a partir de fitxes de buidatge omplertes per col·laboradors externs. Tots tres tenen en compte variables d'anàlisi quantitativa i qualitativa i comparteixen certs elements metodològics.

L'apartat metodològic d'aquesta dissertació ressegueix la senda d'aquests treballs. La finalitat serà tractar d'automatitzar algunes de les variables dels estudis esmentats per tal de substituir les fitxes de buidatge o monitoratge amb bases de dades generades mitjançant codi de programació, intel·ligència artificial generativa i tècniques d'anàlisi de dades.

3.3. Dels referents per a l'anàlisi i visualització de dades

Tant als mitjans de comunicació com a la societat civil existeixen casos de reportatges de periodisme de dades creats a partir de raspats de fonts d'informació. Aquests exercicis encara són excepcionals i el seu esforç tot sovint justifica la creació d'un web específic per consultar totes les històries periodístiques associades al projecte. També és habitual que incloguin infografies, cercadors per a la base de dades, apartats de metodologia o enllaços a les dades en brut. El periodisme nord-americà va ser un dels primers a tractar aquestes tècniques i formats amb projectes com ara:

- *Dollars for Docs* (Tigas et al., 2010): Aquest projecte de *Propublica* va utilitzar el raspat per recopilar dades sobre els pagaments efectuats als professionals sanitaris pels fabricants de medicaments. La visualització de dades permet que els lectors explorin les dades recopilades i afegeix transparència a la indústria farmacèutica.

- *Fatal Force* (Jenkins et al., 2015): El diari *The Washington Post* va desenvolupar una base de dades sobre els trets mortals per part de la policia als Estats Units. En molts de casos, les fonts d'informació eren els mateixos reporters. Tot i això, el gran factor innovador va ser l'exercici de crear una base de dades i mostrar-la a través d'infografies interactives per oferir una anàlisi detallada de cada incident.
- *Gun deaths in America* (Casselman et al., 2016): El blog *Fivethirtyeight*, adquirit més tard per ABC News, va escometre una anàlisi profunda sobre les morts per armes de foc als EUA. Varen utilitzar les dades de diverses fonts públiques, les varen processar amb codi de programació i en varen publicar un web amb visualitzacions impactants. Tant el codi com les dades són disponibles en obert a Github.

Fora dels mitjans de comunicació, existeix l'exemple d'*Inside Airbnb*. Aquesta iniciativa analitza les dades de pisos llogats a Airbnb de diversos llocs del món i fa servir el raspap per recopilar la informació. Les seves visualitzacions permeten copsar com Airbnb afecta els mercats de l'habitatge i contribueix a la gentrificació de certs territoris (Cox et al., 2016).

A Espanya, cada vegada més mitjans disposen d'equips especialitzats en dades, tot i que encara són una minoria. Alguns bons exemples de raspap de dades que serveixen per generar reportatges i visualitzacions de dades els ha fet la Fundació Civio, encara que amb un abast molt superior al d'aquest treball. En són exemples:

- *¿Dónde van mis impuestos?* (Fundación Civio, 2010): Aquest projecte vol mostrar a què es destinen els impostos recaptats a Espanya, d'acord amb el principi de transparència de la despesa pública. Utilitzen dades dels pressupostos generals de l'Estat i fan servir tècniques de *scraping* i programació per processar i presentar les dades en formats interactius.
- *El indultómetro* (Fundación Civio, 2013): El portal web permet veure les dades sobre els indults concedits pel Govern espanyol des de 1996. El projecte combina raspap de dades del BOE, reportatges dels casos més cridaners i infografies per ressaltar les tendències de cada Executiu a l'hora d'atorgar aquesta mesura de gràcia.
- *¿Quién cobra la obra?* (Fundación Civio, 2015): Es tracta d'una investigació sobre la contractació d'obres públiques a Espanya. El projecte també fa ús del *scraping* per recopilar dades de contractes i les visualitza per mostrar qui són els majors beneficiaris dels contractes públics.

No s'han trobat referents locals per al present treball que combinin el raspap amb la intel·ligència artificial. Sí que hi ha alguns exemples internacionals, a una altra escala i amb unes altres finalitats, que combinen raspap, anàlisi estructural i IA generativa. Un cas és l'eina Cyborg de Bloomberg. Entre altres funcions, permet analitzar informes financers d'empreses en qüestió de minuts per facilitar la feina dels periodistes econòmics (Willis, 2019). Una iniciativa semblant la fa servir el diari *The Washington Post* amb l'eina Heliograf (Moses, 2017), que el 2020 varen fer servir per a la cobertura electoral amb missatges d'àudio automatitzats (The Washington Post PR, 2020).

4. Metodologia

La metodologia d'aquest treball final de màster divergeix quant a l'emprada en els informes de gènere i mitjans esmentats a l'apartat 3.2. Com ja s'ha explicat, la recerca dels informes citats es basa en voluntaris que analitzen els mitjans, la qual cosa introdueix un possible biaix subjectiu en la recopilació de les dades. A més, limita l'abast de recopilació i buidatge de notícies i implica uns temps molts més llargs tant per a l'anàlisi quantitativa com per a la qualitativa.

En els darrers anys, s'han posat a l'abast del públic eines d'intel·ligència artificial (IA) basades en models de llenguatge gran (*large language models* o LLM) que poden resoldre la necessitat passada d'un buidatge manual. Aquest treball vol indagar en el potencial d'aquestes eines de cara a l'automatització de l'anàlisi de les dades, tot combinant-les amb les possibilitats de la programació per la recopilació i tractament de la mostra.

4.1. Selecció de la mostra

L'estudi tracta els dos principals mitjans digitals de Mallorca per volum mensual de visitants: *Ultimahora.es*, amb gairebé 5 milions d'usuaris únics l'octubre de 2023, i *Diariodemallorca.es*, amb 2,8 milions el mateix mes (OJD interactiva, 2023). De cara a la tria de les dates per a la mostra, s'han tingut en compte les següents consideracions:

1. Que siguin del mateix dia de la setmana per tal que ambdues capçaleres tinguin:
 - a. les mateixes seccions, atès que poden variar segons el dia de la setmana;
 - b. una xifra semblant de personal, ja que cap de les dues capçaleres té una plantilla de cap de setmana, i
 - c. una quantitat similar d'articles, per les mateixes raons del punt anterior.
2. Que siguin de la mateixa setmana del mateix mes per neutralitzar el factor de l'estacionalitat de la informació (p. ex., a principis de setembre s'inclouran més notícies sobre l'inici del curs escolar) i que no siguin de períodes de vacances (p. ex., durant l'agost o Nadal), per evitar que el volum de notícies sigui més baix de l'habitual.
3. Que a les dates triades no hagi tingut lloc cap notícia de gran impacte o abast que pugui desvirtuar la comparativa interanual (p. ex., un partit de futbol de primera divisió masculina o femenina suposaria modificar la representació usual de la secció d'esports).

D'acord amb aquests criteris, les dates seleccionades són tres dimarts, tots ells de la tercera setmana de setembre: el 21 de setembre de 2021, el 20 de setembre de 2022 i el 19 de setembre de 2023.

Amb aquesta tria, d'entrada, es complementa la mostra de mitjans de comunicació realitzada per a l'informe *Un dia a les notícies en clau de gènere: protagonistes i periodistes als mitjans de comunicació de les Illes Balears*, ja que no incloïa mitjans digitals. Al mateix temps, s'hi amplia l'abast temporal (un dia d'un any en aquell treball enfront d'un dia de tres anys diferents en aquest).

4.2. Recopilació i tractament de les dades

Una vegada seleccionada la mostra, es durà a terme un raspap d'informació (*scraping*) per aconseguir els articles per a l'anàlisi. La informació recuperada se sotmetrà a un procés d'anàlisi estructural (*parsing*) per poder convertir-la en un marc de dades estructurat en forma de taula.

Una vegada recopilats els articles, caldrà eliminar tots aquells que no siguin valuosos per a l'anàlisi, ja sigui perquè no són textos, perquè no són articles periodístics o amb intervenció editorial, perquè estan duplicats o per qualsevol altre motiu que pugui desvirtuar els resultats.

Els marcs de dades seguiran una estructura relacional per tal de mantenir-ne la senzillesa i optimitzar el volum de dades, al mateix temps que es podran ampliar per allotjar la informació que s'obtingui de les variables de l'anàlisi.

4.3. Definició de les variables d'anàlisi

Les variables d'anàlisi estan basades en la darrera versió de la guia ràpida sobre la codificació per a un monitoratge complet del Global Media Monitoring Project (WACC Global, 2020, p. 10-12). En la tria, s'ha donat prioritat a aquelles que siguin quantitatives i automatitzables per aprofitar al màxim el desenvolupament tecnològic.

A més, s'han intentat reflectir moltes de les variables incloses en l'informe elaborat el 2014 a les Balears. D'aquesta forma, salvant les diferències metodològiques i de tipologia dels mitjans analitzats, s'intentaran crear sèries temporals per observar la tendència en la representació del col·lectiu femení.

Atenent aquestes premisses, les variables quantitatives i qualitatives de les quals es vol automatitzar l'anàlisi són:

- Sobre els mitjans:
 - Xifra de notícies per mitjà.
 - Volum de notícies per mitjà i secció.
- Sobre els professionals de la comunicació:
 - Nombre de notícies escrites segons el sexe de les persones autores.
- Sobre els continguts de les notícies:
 - Nombre de persones que apareixen al títol per sexe.
 - Nombre de persones que apareixen al subtítol per sexe.
 - Nombre de persones que apareixen al text per sexe.
 - Nombre de persones que apareixen a la fotografia principal que acompanya la notícia per sexe.
 - De les persones esmentades als articles, si s'esmenta al text:
 - Nom.
 - Sexe.
 - Edat, si s'esmenta a la notícia.

- Ocupació, si s'esmenta a la notícia.
 - Rol familiar, si s'esmenta a la notícia.
 - Si la persona és presentada com a:
 - Protagonista.
 - Portaveu.
 - Veu experta.
 - Testimoni.
 - Víctima o supervivent.
 - Si la persona apareix a:
 - Títol.
 - Subtítol.
 - Text.
 - Nombre de vegades que s'esmenta la persona en el conjunt de l'article.
- Sobre el procés de tractament automatitzat de les dades:
- Costos del tractament mitjançant l'API d'OpenAI.
 - Temps d'execució dels processos automatitzats a través de l'API d'OpenAI.
 - Respecte d'una submostra del conjunt de notícies:
 - Nombre de dades recopilades mitjançant IA que han hagut de ser corregides mitjançant una revisió manual.

4.4. Visualització de les dades

La visualització de les dades és l'etapa final en la presentació i interpretació dels resultats obtinguts en aquest treball. L'objectiu principal d'aquesta fase és transformar les dades complexes en representacions visuals clares, comprensibles i atractives per a qualsevol públic, que facilitin la comprensió de les tendències i patrons identificats en l'estudi. Amb aquesta finalitat, el procés metodològic que es planteja consisteix en quatre etapes:

1. *Elaboració de taules i marc de dades*: Inicialment, les dades recollides i processades seran organitzades en taules i marcs de dades. Aquests formats estructurats són essencials per a una anàlisi detallada i per facilitar la transició cap a una visualització efectiva. Els marcs de dades inclouran sèries estadístiques diverses, en xifres absolutes i relatives, que permetin ressaltar les tendències més importants en les dades.
2. *Selecció d'eines de visualització interactives*: Aquestes eines permetran crear representacions visuals dinàmiques, que poden ser explorades i modificades pels usuaris per a una millor comprensió de les dades. L'aproximació interactiva permetrà els usuaris no només veure les dades de manera estàtica, sinó també explorar diferents capes i dimensions de les dades.
3. *Focus en l'accessibilitat i la comprensió*: La finalitat de la visualització de dades és fer que la informació sigui accessible i fàcil d'entendre per a qualsevol persona, independentment del seu bagatge en anàlisi de dades. Per tant, es prestarà atenció especial a la claredat, la simplicitat i la interpretabilitat de les visualitzacions. Això inclourà l'ús de colors i formes que siguin fàcilment distingibles i la inclusió de textos explicatius i títols clars.

4. *Publicació en el portal web*: Finalment, les visualitzacions seran incorporades a la pàgina web del projecte. Aquesta plataforma actuarà com a mitjà per a la difusió dels resultats de la recerca.

La finalitat que es pretén aconseguir és no només presentar les dades de manera efectiva, sinó també promoure una major comprensió i conscienciació sobre la representació de gènere a la premsa digital de Mallorca.

5. Procés d'elaboració

5.1. Consideracions prèvies

Atesa la natura d'aquest treball, el plantejament inicial és que el codi que es desenvolupi el pugui reutilitzar qualsevol persona amb uns mínims coneixements de programació. Per aquest motiu, el format en què es guarda la informació tècnica és un quadern de Jupyter (*.ipynb*). S'ha utilitzat VisualStudio Code com a eina d'ajuda al disseny del codi. Per la seva versatilitat, el llenguatge de programació triat és Python en la versió 3.12.0.

5.2. Creació dels marcs de dades amb les notícies

La primera passa ha estat recuperar els enllaços de les notícies dels dies de la mostra. Totes dues webs disposen d'una secció d'hemeroteca que permet filtrar per dia. Ara bé, cap de les dues mostra totes les notícies a una sola pàgina, la qual cosa dificulta l'execució del *scraping*. La solució per la qual s'ha optat ha estat mostrar tots els resultats manualment i descarregar-ne el codi font en format HTML.

Cal assenyalar que l'extracció del codi font es podria haver automatitzat amb Selenium (un complement de Python) que permet ser configurat per clicar a botons com "Mostrar más" o canviar de pàgina de resultats. Això no obstant, el desenvolupament d'aquesta opció hauria comportat prou més temps que la solució *ad hoc*.

S'han extret tots els enllaços a notícies, filtrant-los a través del tipus d'objecte HTML. A partir de la llista d'enllaços, s'ha descarregat tota la informació de cada article amb la llibreria Newspaper3k. Aquesta eina permet recuperar els textos complets, encara que les notícies siguin de pagament i dividir cada notícia en camps. Gràcies a Newspaper3k, s'ha pogut obtenir la informació de la data de publicació, el títol de l'article, el text, l'URL de la notícia i l'URL de la imatge mitjançant un procés d'anàlisi estructural (*parsing*).

La secció s'ha recuperat a partir de cada URL mitjançant expressions regulars (p. ex., per a "https://www.diariodemallorca.es/economia/2023/..." es copia el text entre la tercera i la quarta barra inclinada, que és "economia"). El subtítol i el nom de la persona autora de l'article s'han aconseguit des del codi font amb un procediment semblant a l'utilitzat per elaborar el llistat d'adreces web.

Aquest procés s'ha fet de forma separada per a cadascun dels mitjans analitzats, ja que els webs estan dissenyats de manera diferent i ha estat necessari personalitzar la forma de descarregar la informació.

El marc de dades inicial (*df*, per les inicials de *data frame*) conté 1.380 resultats. D'acord amb els criteris establerts a l'apartat 4.2 de la metodologia, s'han filtrat els següents continguts:

- Publicitat. En canvi, s'han mantingut els publireportatges, ja que tenen intervenció editorial d'algun membre de la redacció.
- La secció "Llama la atención" de *Diariodemallorca.es*.
- Notícies d'agenda.

- Sortejos de loteria.
- Tires còmiques, per la dificultat d'automatitzar-ne l'anàlisi.
- Les fotogaleries que no contenen text.
- Articles repetits o que ja estan inclosos dins d'altres notícies.

El marc de dades *df_canvis* és el que conté les notícies ja filtrades, que sumen 1.311 articles. Inclou una columna amb la longitud del camp text. També té columnes separades per a la firma de les notícies amb la finalitat de facilitar l'anàlisi automatitzada (són sols dues columnes, perquè no hi ha cap cas a la mostra on un article l'hagin firmat més de dues persones).

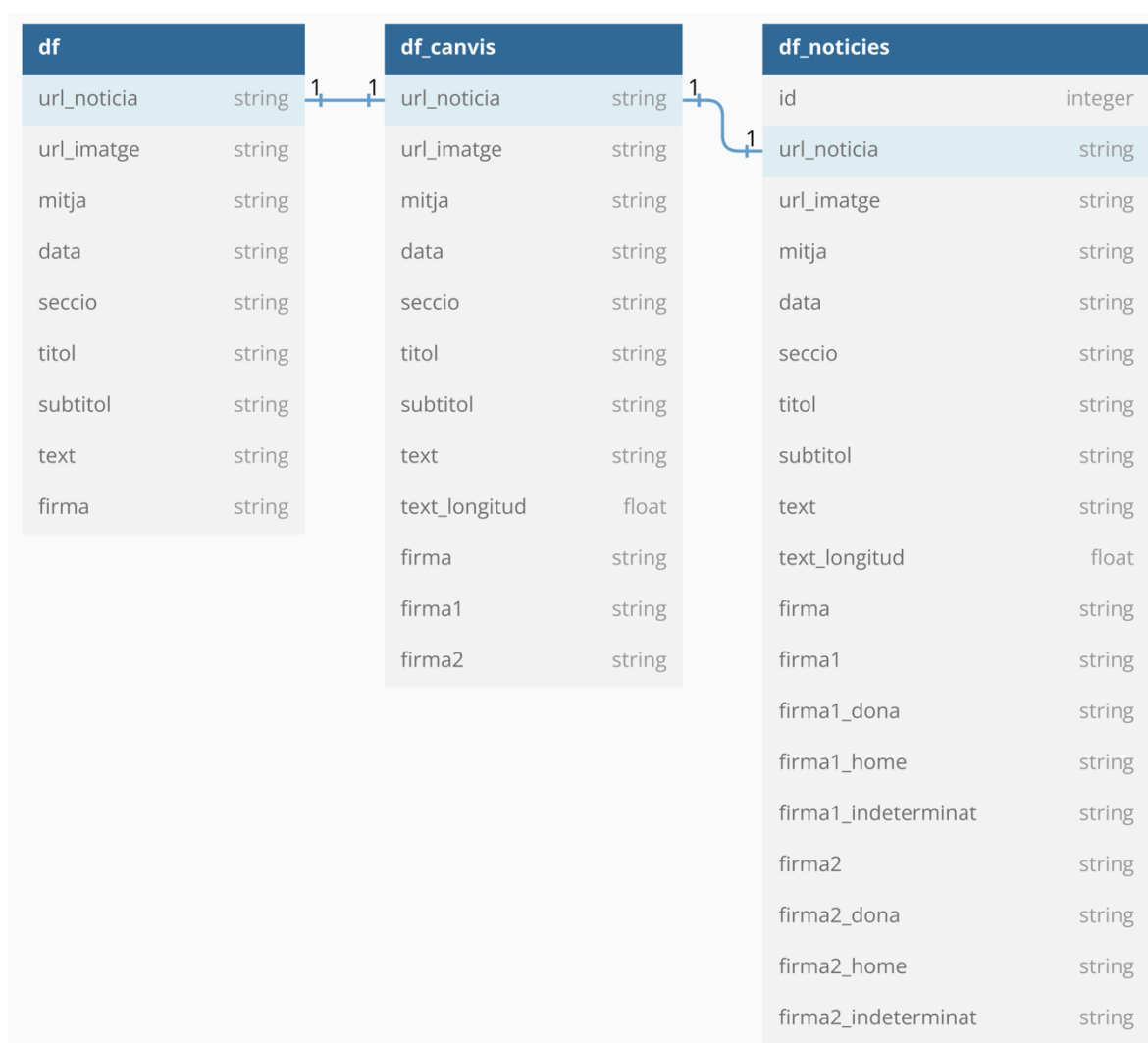


Figura 1. Marcs de dades que contenen les notícies obtingudes a partir de scraping i parsing, i les seves relacions.

En aquesta taula, s'han aplicat tota una sèrie de transformacions per homogeneïtzar les dades. S'han eliminat les hores de les dates. S'han estandarditzat les seccions i s'han suprimit els URL de les imatges d'aquells articles que no contenen fotografies. Mitjançant un mapatge (*mapping*), s'han modificat algunes firmes a manera de pretractament per a l'anàlisi

automatitzada, esclarint el nom propi de diverses persones que firmen amb el nom ocult per inicials, només amb els llinatges o amb pseudònims.

A partir de tota la informació de *df_canvis*, s'ha creat la taula *df_notícies*, que es converteix en el marc de dades central de l'estructura relacional (figura 1). Per aquest motiu, aquesta taula incorpora un camp *ID* que facilita la identificació de cada registre. També s'hi afegeixen tres columnes per cada camp de firma per tal de recollir la informació de la primera variable analitzada (el sexe de la persona autora de cada article).

5.3. Proves amb els models d'OpenAI

Per a l'anàlisi de les firmes, es requeria extreure només una dada mitjançant IA: quin era el sexe més probable per a cada nom. Aquest fet va suposar que fos una variable ideal per provar el funcionament de l'API d'OpenAI per a l'anàlisi automatitzada de notícies. Per estalviar costos i temps de processament, s'ha creat un marc de dades nou amb tots els possibles valors dels camps de firma que conté 275 valors de periodistes o articulistes diferents.

Durant les proves per a aquesta tasca, s'han testejat dos models d'OpenAI: text-davinci-003 i gpt-3.5-turbo-0613. Tots dos models són dels més barats que ofereix OpenAI l'octubre de 2023. Les proves varen mostrar que text-davinci-003 necessita més informació de context i retorna resultats més inconsistents.

En canvi, el model gpt-3.5-turbo-0613 ha estat capaç d'acomplir una tasca pautada amb major precisió. L'objectiu de l'*input* enviat a la IA era que retornés una resposta sempre idèntica perquè fos fàcil de categoritzar i afegir al marc de dades existent.

Després de diverses proves, s'ha descartat enviar els noms dels quals ja se sap que no són atribuïbles a una persona. En aquest cas, eren tots aquells que contenen les paraules "Agència", "Agències" o "Redacció". En els primers intents, s'havia donat tres opcions de resposta a la IA: dona, home o desconegut.

S'ha observat que quan s'atorgava al model una tercera opció per autoritzar-li que assignés un gènere desconegut als noms, la IA donava preferència a aquesta possibilitat en comptes de triar un sexe. L'assignació del valor *desconegut* ha tingut lloc de manera especialment repetida amb noms poc comuns en espanyol o català (p. ex., un cas especialment problemàtic ha estat el del nom Idoya).

Finalment, la indicació (*prompt*) enviada al model ha reduït les opcions a dues per forçar la IA a prendre una decisió. El text amb les ordres s'ha enviat en anglès i ha estat el següent:

"Determine the gender of the following Spanish or Catalan names. Answer only with 'female' or 'male'. If you think there is a doubt between 'female' and 'male', answer with the most likely option. What gender does the name [...] suggest?"

En la primera iteració, va deixar en blanc el nom Kike i va respondre que Mia no tenia un gènere associat. En repetir la petició només per a aquells dos noms, va assignar el sexe

correctament. Cal assenyalar que va encertar a la primera els noms Eli i Lluc, que podrien ser tant de dona com d'home (en aquest cas, tots dos corresponien a homes).

Per a l'anàlisi dels textos, s'han provat tres models: GPT 3.5 Turbo (tècnicament, gpt-3.5-turbo-0613), GPT-4 (també anomenat gpt-4-0613) i GPT-4 Turbo (una versió més ràpida de GPT-4, que permet donar al model un major context). Aquest darrer model, GPT-4 Turbo, és encara una versió beta llançada al públic per OpenAI el 6 de novembre de 2023 (durant l'elaboració del treball).

S'ha detectat que, si es demanaven els mateixos nou ítems d'anàlisi, però sense explicar a la IA el format de la resposta, l'output era un llistat o un paràgraf explicatiu en els quals es bolcaven tots els elements de la pregunta. Per tal de facilitar el processament de la resposta, la millor opció detectada ha estat sol·licitar al model que contesti amb una expressió regular. Després de diversos intents, la pregunta formulada als models ha estat:

"You're an AI trained to extract human data from texts. Exclude places, dates, organizations, initials, or non-person names. Extract all person names mentioned in the following news article. The text is in Spanish or Catalan. Then, identify the following information only if stated in the text in this order: [1] name, [2] age, [3] job or profession (in only one or two words, in English, excluding family roles), [4] family role (e.g., wife, husband, mother, parent...), [5] if the person is the protagonist of the article (1 for yes, 0 for no), [6] if they are identified as a spokesperson (1=yes, 0=no), [7] if they are identified as an expert (1=yes, 0=no), [8] if they are identified as a witness (1=yes, 0=no), and [9] if they are identified as a victim or survivor (1=yes, 0=no). Every reply has to include always those nine elements per person separated with commas (even though they are blank) and every person will be separated with semicolons, so the information follows a regular expression (e.g., Person1, 25, actress, , 1, 0, 0, 0, 0; Person2, , politician, , 0, 1, 0, 0, 0)."

Les proves amb GPT-3.5 Turbo han mostrat problemes a l'hora d'executar ordres complexes. Mentre que era capaç de retornar una llista de noms únics trobats als articles, de vegades es confonia amb sobrenoms o alguns cognoms que s'esmentaven sense repetir el nom propi. També hi ha inclòs noms de llocs o d'empreses. A més, no va ser capaç de retornar una resposta tot formant una expressió regular.

GPT-4 ha resultat molt millor les indicacions facilitades. Concretament, quan s'ha demanat al model informació complementària de cada nom de persona per completar les variables d'estudi. D'aquesta manera, es pot simplificar el procés d'anàlisi sense haver de fer una petició a la IA per extreure els noms, una altra per obtenir les edats, una tercera per a les professions, i així fins a obtenir tots els paràmetres. El principal inconvenient del model és el preu per ús, 30 vegades més car que GPT-3.5 Turbo.

Finalment, per a l'anàlisi automatitzada dels textos, s'ha optat per GPT-4 Turbo. Tot i no ser una versió estable segons els seus desenvolupadors (OpenAI, 2023), aquest model no ha

mostrat diferències en la informació que recupera del subconjunt de notícies respecte de GPT-4 i té el gran avantatge de ser tres vegades més barat que GPT-4.

5.4. Creació d'altres marcs de dades complementaris

Com s'ha explicat a l'apartat anterior, amb l'ajuda de la IA s'han creat dues taules noves que complementen *df_notícies* (figura 2). El marc de dades *df_firmes* conté la llista única de valors d'autors d'articles (és a dir, de *df_notícies:firma1* i *df_notícies:firma2*) amb el sexe assignat per la IA. L'altre conjunt és *df_persones* i inclou totes les dades necessàries per respondre a les variables d'estudi determinades a l'apartat 4.3 de la metodologia.

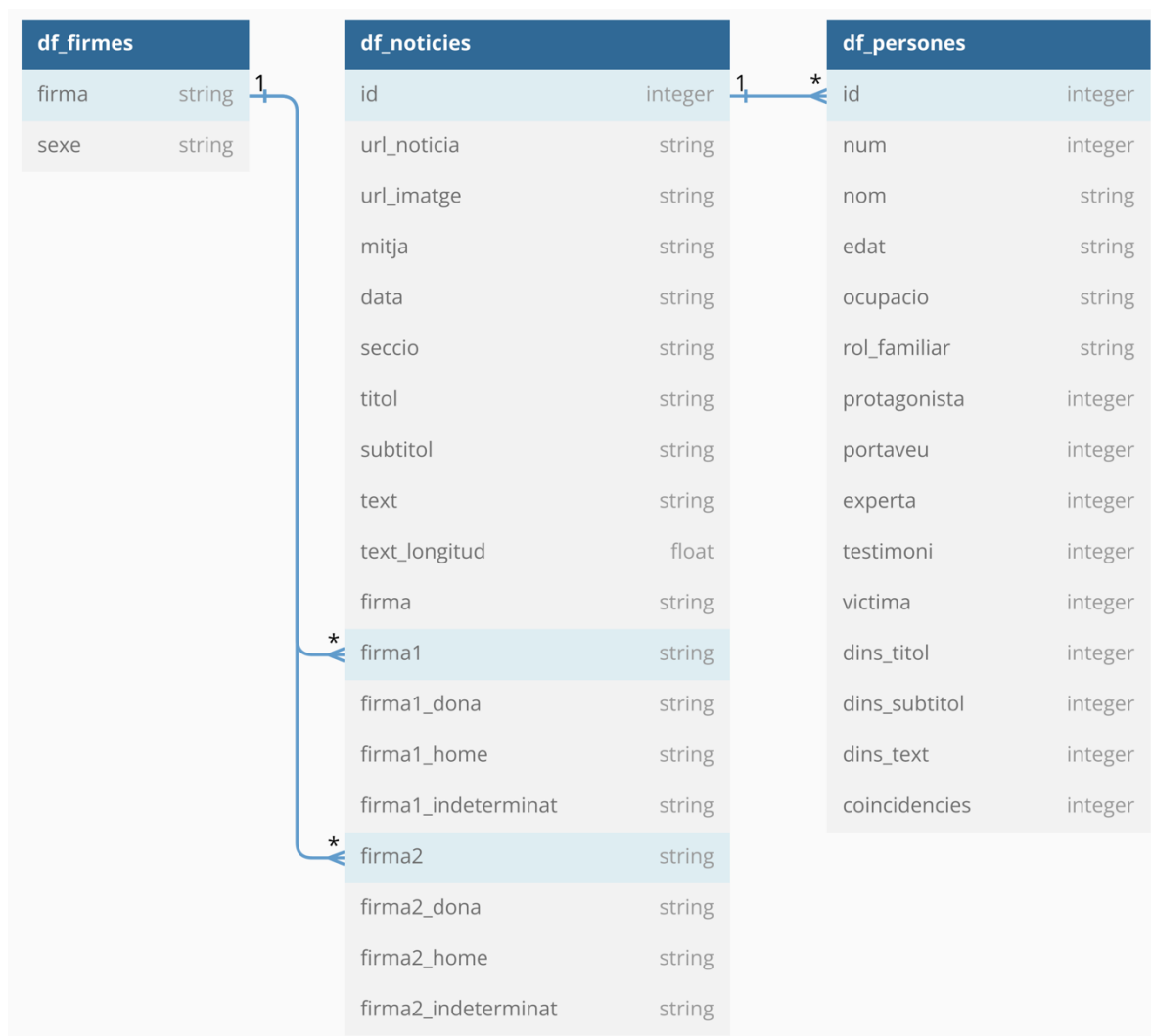


Figura 2. Relacions entre les taules que complementen el marc de dades principal (*df_notícies*).

El camp *id* es correspon amb la notícia de *df_notícies*. La columna *num* és un comptador correlatiu de les persones trobades a cada notícia. Els camps *dins_títol*, *dins_subtítol* i

dins_text es creen fent una cerca dels noms dins dels camps de títol, subtítol i text. La resta de camps de *df_persones* són els generats mitjançant IA.

5.5. Errades amb el codi de l'anàlisi i errades en les respostes de la IA

Després d'haver executat el codi per recuperar la informació dels textos, s'han observat tot un seguit de missatges d'error que no es varen detectar amb el subconjunt emprat per a les proves. En concret, en el *prompt* no s'havia previst una resposta estàndard per al cas que no hi hagués cap nom propi a la notícia. El codi tampoc havia contemplat l'opció que la IA inclogués un signe de puntuació final a les respostes, la qual cosa ha marcat com a error aquelles rèpliques que acabaven amb un punt i coma o amb un punt final. En altres casos, la IA ha retornat missatges inconsistents amb les indicacions fetes, bé perquè no seguien una expressió regular, bé perquè incloïen notes o missatges addicionals.

Totes les respostes obtingudes a través de l'API s'han emmagatzemat a uns fitxers de registre (*logs*) i s'han reprocessat posteriorment per evitar repetir les peticions a la IA amb els costos subsegüents que això suposa.

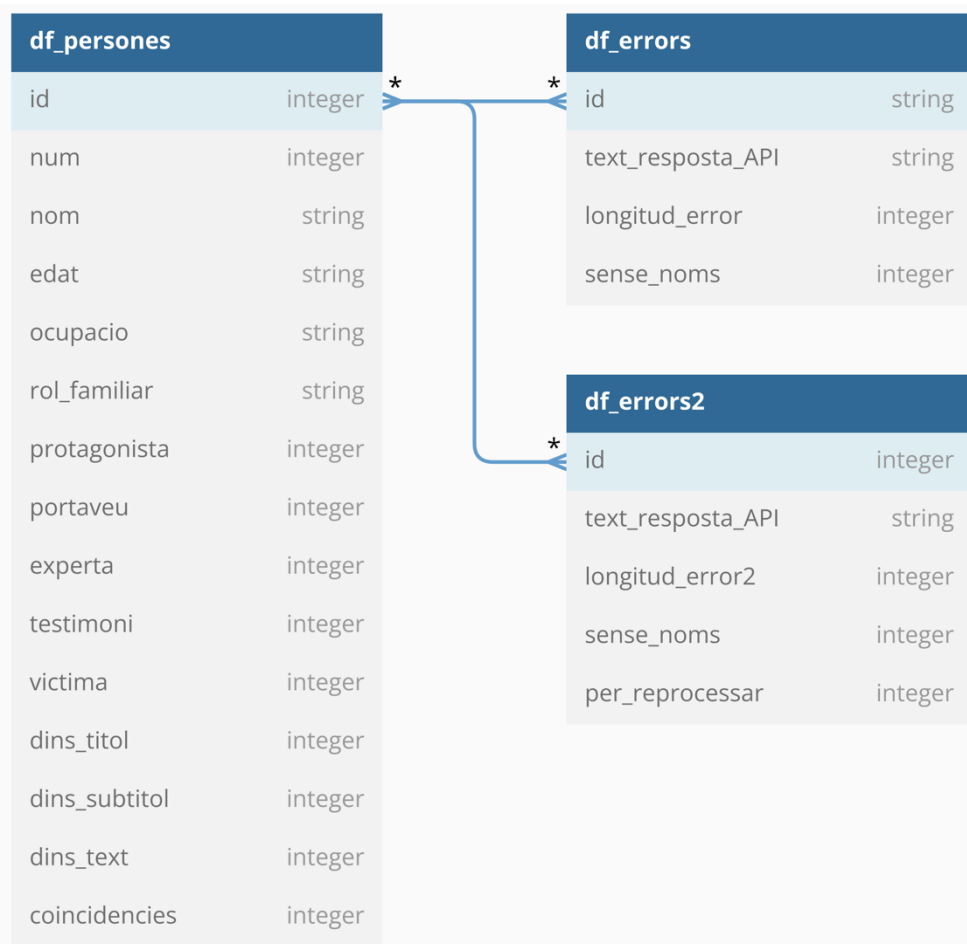


Figura 3. Relacions entre el marc de dades *df_persones* i les taules que emmagatzemen els errors en el processament automatitzat de les notícies.

Els missatges d'aquests *logs* s'han transformat en dues taules (figura 3). El marc de dades *df_errors* conté els missatges del primer intent d'anàlisi amb les 1.311 notícies. S'han marcat totes aquelles en les quals la IA ha especificat que no hi havia cap nom a la notícia.

La resta de textos s'han enviat de nou a través de l'API després de depurar el codi. Els errors d'aquest segon intent figuren a *df_errors2*, que en aquesta ocasió s'ha procedit a revisar manualment pel poc volum de casos no processats correctament. En aquesta taula, s'ha tornat a incloure una columna per marcar notícies sense noms i també s'ha establert un sistema per marcar les notícies que s'havien de reprocessar amb codi de Python, tota vegada s'havien corregit les respostes del model GPT-4 Turbo per conformar les expressions regulars que se li havien sol·licitat.

5.6. Anàlisi automatitzada de les fotografies

Amb el gruix del marc de dades *df_persones* analitzat, s'ha procedit a provar l'anàlisi automatitzada de les fotografies. Per a aquesta tasca, OpenAI només ofereix un model, anomenat GPT-4 Vision o bé *gpt-4-vision-preview*. Es tracta d'un model en proves, que permet estudiar fotografies que s'envien bé a través d'un URL bé carregant-les des d'un arxiu.

El sistema de tarifes es basa fonamentalment en la resolució de la imatge que s'ha d'analitzar. Per al processament del gruix de les imatges, s'ha determinat que el paràmetre *detail* sigui *low*, de forma que només envia imatges de 512 × 512 píxels. Per al reprocessament de les imatges que han donat error, el mateix paràmetre s'ha assignat a *high* per tal d'evitar la compressió de les fotografies.

Es va crear un subconjunt de proves per veure el funcionament de l'anàlisi d'imatges. En dur a terme els tests, s'ha observat que alguns URL d'imatges de *Diariodemallorca.es* recuperats amb la llibreria *Newspaper3k* són erronis, principalment amb notícies basades amb vídeo. L'error és del tipus 404. Per aquest motiu, a l'inici del processament s'ha inclòs una funció que comprova la validesa de l'enllaç. En cas que l'adreça no sigui accessible, l'afegeix a un llistat d'imatges processades per prevenir els errors.

A la fase de proves, la taxa d'errades de l'API fregava el 90 % amb missatges com ara *"I'm sorry, but I can't assist with that request"* o *"I'm sorry, but I can't assist with images involving people"*. En alguns casos, la resposta de la IA argumentava motius ètics. La taxa d'errors es va invertir després d'incloure a l'inici del *prompt* la frase *"I'll show you a photograph published in a newspaper"*. D'aquesta manera, el model va acceptar fer l'anàlisi de les imatges. El *prompt* complet ha estat el següent:

"I'll show you a photograph published in a newspaper. Determine how many people are in this image. Also, determine their gender. Return an answer with 4 numbers separated by commas following a regular expression: total of people, females, males, unknown gender. Example: 5,2,3,0. If there is no people, return 0,0,0,0"

L'oració final es va incloure atès que, del 10 % aproximat de respostes fallides, gairebé totes eren degudes al fet que el retorn de la IA no s'ajustava al patró amb aquelles fotografies de

paisatges, ciutats o edificis que no contenien persones. La resta es devien a fotografies de grans grups (p. ex., manifestacions o plans generals de carrers) o a imatges borroses, com ara un *frame* extret d'un vídeo enregistrat amb un mòbil. El temps de processament de les 1.302 imatges, incloent-hi els URL trencats, ha estat d'una hora i deu minuts.

5.7. Depuració de les dades recuperades via IA

En aquest punt, s'ha estudiat si hi havia files repetides mitjançant un sistema de clau composta. Per a un mateix codi de notícia, no hi podia haver un nom repetit. No s'han observat files repetides, per la qual cosa no ha estat necessari eliminar resultats.

A continuació, s'han analitzat els valors únics de *df_persones* a partir de patrons (p. ex., el nombre de caràcters, si contenen punts o guionets o si no tenen espais). També es repassen les seves professions per detectar irregularitats (p. ex., amb la professió *brand* s'han eliminat els valors Carolina Herrera o Sandro Paris que no feien referència als dissenyadors de moda, sinó a les seves empreses). En aquest cas, sí que s'han eliminat algunes files incloses en el marc de dades perquè no eren persones.

La següent passa ha consistit a analitzar el sexe de les persones. Aquest paràmetre no s'havia especificat intencionadament al *prompt* de l'API per fer el bolcat de les notícies, per tal de processar-ho a partir de llistats de valors únics. Els valors únics s'han relacionat al marc de dades *df_sexepersones*. Amb l'experiència del procés anterior d'anàlisi de les firmes, primer s'ha procedit a fer mapatges manuals de valors fàcils de confondre, com ara noms artístics o sigles.

Com a novetat, en aquesta ocasió els noms s'envien en tongades de 12 valors per reduir costos. Cada vegada que s'envia informació a l'API s'ha de repetir el que s'anomena missatge de sistema, que són les instruccions bàsiques al model per establir quin és el camp d'expertesa de la IA i com s'ha de comportar. Amb les tongades, es poden analitzar més noms fent servir menys *tokens* (les unitats bàsiques de significat dels models de llenguatge que serveixen a OpenAI per establir el sistema de tarifació).

El missatge de sistema enviat ha estat el següent:

"Determine the gender of the following 12 names. Answer with 1 for female, 2 for male, 3 for unknown, separating the answers with commas. If you think there is a doubt, between 'female' and 'male' answer with the most likely option. Return an answer with 12 numbers separated by commas following a regular expression. Example: 2,1,2,1,1,2,2,1,2,2,1,2."

El missatge d'usuari, és a dir, la instrucció concreta de la tasca, ha estat:

"The 12 names, separated by semicolons, are: [...]"

Aquest mètode no només ha suposat una reducció de costos, sinó també un escurçament notable del temps de resposta de l'API. A tall d'exemple, amb *df_firmes* es varen analitzar 275

noms en 16 minuts i 30 segons, mentre que amb *df_sexepersones* s'han estudiat 3.084 noms en 8 minuts i 11 segons.

La metodologia del Global Media Monitoring Project (GMMP) estableix que alguns valors s'han de codificar a partir dels valors d'unes taules predefinides per tal de facilitar l'anàlisi estadística posterior. Per això, les següents fases han estat parametritzar les respostes de l'API a partir dels codis del GMMP.

Quant al rol familiar, el marc teòric demana explicitar únicament si s'inclou un rol familiar (però no quin) de la persona esmentada a la notícia. Amb aquesta finalitat, s'han dividit els valors trobats per l'API en dos grups, segons si la resposta era un rol familiar (p. ex., espòs, germana...) o no (p. ex. amiga). Finalment, aquests valors s'han convertit en 0 o 1.

El procediment ha estat més complex per codificar les professions. Per a aquest camp, el GMMP estableix 28 possibles valors. La taula amb els codis i descripcions del GMMP s'han extret de la guia de monitoratge (WACC Global, 2020) i s'han transcrit a la taula *df_categoriaocupació*. S'ha creat una llista de valors únics de professions i s'han dissenyat les instruccions per al model de llenguatge d'IA, que en aquest cas han estat més llargues, ja que havien d'incloure la taula de codificació. El missatge de sistema ha estat:

“Match each of the provided occupations to one of the categories by answering only the category ID number. These are the categories:

0 Not stated

1 Royalty, monarch, deposed monarch...

2 Politician, member of parliament...

3 Government employee, public servant, spokesperson...

4 Police, military, paramilitary, militia, fire officer

5 Academic expert, lecturer, teacher

6 Doctor, dentist, health specialist

7 Health worker, social worker, childcare worker

8 Science/technology professional, engineer...

9 Media professional, journalist, film-maker...

10 Lawyer, judge, magistrate, legal advocate...

11 Business person, exec, manager, stockbroker...

12 Office or service worker, non-management worker

13 Tradesperson, artisan, labourer, truck driver...

14 Agriculture, mining, fishing, forestry

15 Religious figure, priest, monk, rabbi, mullah, nun...

16 Activist or worker in civil society organization, NGO, trade union

17 Sex worker

18 Celebrity, artist, actor, writer, singer, TV personality

19 Sportsperson, athlete, player, coach, referee

20 Student, pupil, schoolchild

21 Home-maker, parent

22 Child, young person

23 Villager or resident

24 Retired person, pensioner

25 Criminal, suspect

26 Unemployed

27 Other

If you think there is a doubt between several options, answer with the lowest ID number.

Return an answer with 12 numbers separated by commas following a regular expression. Example: 2,0,27,5,1,2,2,4,5,8,9,12."

Per completar les instruccions, el missatge d'usuari amb la tasca ha estat:

"The 12 occupations, separated by semicolons, are: [...]"

El codi s'ha executat sense errors en 69 segons. Finalment, els valors de la taula paramètrica *df_sexepersones* s'han inclòs en una nova columna de *df_persones* que emmagatzema el codi d'acord amb la numeració del GMMP.

5.8. Comprovació de la fiabilitat de l'anàlisi automatitzada amb IA

En primer lloc, s'han revisat una per una totes les assignacions de valors realitzades a les ocupacions d'acord amb la taula de codificació del GMMP. La conclusió és que la fiabilitat de la IA per acomplir aquesta tasca ha estat del 86,5 %. Hi ha 76 valors canviats manualment (un 13,5 %) d'un total de 563 ocupacions. Les dades han estat calculades a partir del llistat de valors únics d'ocupacions que figura al marc de dades *df_ocupacions*.

Tot seguit, s'ha fet una anàlisi de les persones detectades. Amb aquesta finalitat, s'ha creat una submostra aleatòria d'anàlisi (*df_revisió*) amb el 10 % de les notícies (131 de les 1.311) i s'han extret les persones detectades en aquestes notícies, que s'han revisat manualment. Per avaluar la fiabilitat de la IA a l'hora d'identificar persones a les notícies, s'ha realitzat un enfocament estadístic. Les mètriques emprades són:

- els veritables positius (identificacions correctes per IA),
- els falsos negatius (persones no detectades), i
- els falsos positius (entitats que no són individus o persones duplicades).

A partir de l'anàlisi automatitzada, el model d'IA va trobar 4.380 persones. El nombre d'individus dins el subconjunt de notícies revisades era de 566 registres (*df_revisió_persones*). Durant la comprovació manual, s'han suprimit 36 falsos positius i s'han afegit 21 falsos negatius que la IA no havia detectat. Per al càlcul de la fiabilitat, s'han fet el càlcul de tres estadístiques considerades rellevants (Kundu, 2022):

- Precisió: Mesura el valor predictiu positiu; és a dir, la precisió de la IA per identificar persones, és a dir, quantes de les persones identificades eren realment persones.

$$\text{Precisió} = \frac{\text{Veritables positius}}{\text{Veritables positius} + \text{Falsos positius}}$$

- Reclam o sensibilitat: Expressa la capacitat de la IA per trobar totes les persones reals.

$$\text{Reclam} = \frac{\text{Veritables positius}}{\text{Veritables positius} + \text{Falsos negatius}}$$

- Puntuació F1: És la mitjana de precisió i reclam, generalment emprada per mesurar la fiabilitat del model.

$$\text{Puntuació F1} = \frac{2 \times \text{Precisió} \times \text{Reclam}}{\text{Precisió} + \text{Reclam}}$$

A partir dels càlculs, s'observa que:

- La precisió és del 93,64 %; és a dir, que de tots els individus que la IA va identificar com a persones, el 93,64 % es va identificar correctament.
- El reclam o sensibilitat és del 96,19 %, cosa que mostra que la IA va identificar amb èxit el 96,19 % de les persones reals que hi havia als articles.
- La puntuació F1, a tall d'indicador global de la fiabilitat del model, és del 94,90 %.

Aquestes mètriques suggereixen que l'actuació de la IA ha estat força fiable a l'hora d'identificar persones a les notícies. Tanmateix, és important tenir en compte que aquests càlculs es basen en el subconjunt comprovat manualment i no en tota la mostra.

La següent passa ha estat analitzar la fiabilitat dels paràmetres de cada persona a partir de les modificacions fetes a mà a cada columna del marc de dades. En aquest cas, s'han comparat els valors a *df_revisió_persones* (la submostra aleatòria) amb els de la revisió manual (*df_revisió_persones2*) amb els següents resultats:

- Fiabilitat en analitzar nom: 99,82 %
- Fiabilitat en analitzar sexe: 97,28 %
- Fiabilitat en analitzar edat: 100,00 %
- Fiabilitat en detectar ocupació¹: 46,10 %

¹ Respecte de les professions, només s'ha tingut en compte la columna *ocupació*, ja que la columna *categoria_ocupació* s'ha analitzat a l'inici d'aquest epígraf.

- Fiabilitat en analitzar rol familiar: 99,46 %
- Fiabilitat en analitzar protagonista: 96,01 %
- Fiabilitat en analitzar portaveu: 97,28 %
- Fiabilitat en analitzar experta: 99,27 %
- Fiabilitat en analitzar testimoni: 99,64 %
- Fiabilitat en analitzar víctima: 99,64 %

Cal assenyalar que el valor de fiabilitat en la detecció de les ocupacions és anormalment baix respecte de la resta de valors. El problema no rau en l'anàlisi errònia d'aquesta dada que fa la IA (ja comprovada a l'inici d'aquest epígraf), sinó en la omissió del camp.

L'observació de les dades ha permès detectar un comportament arbitrari de la IA a l'hora d'obtenir les dades de la professió. El *prompt* enviat a l'API sol·licitava que la IA expliqués la feina, si s'esmentava a la notícia en una o dues paraules. En notícies amb molts subjectes, tot sovint la resposta del model ha estat deixar en blanc aquest apartat. S'han trobat tres exemples on aquesta conducta ha estat freqüent:

- *Notícies sobre sèries o pel·lícules o premis de cinema*: En aquests casos, s'esmenten moltes directores o actrius que de vegades s'intercalen amb personatges. Cal destacar que aquest apartat té la notícia amb la xifra més alta de persones esmentades (54 subjectes en una informació sobre la prèvia del Festival de Sitges 2023). La fiabilitat a l'hora de descartar personatges de ficció ha estat bastant bona, però la IA no ha identificat el rol (director, actriu, guionista...) de cada individu.
- *Cròniques polítiques o parlamentàries*: En aquestes notícies, el model de vegades ha mostrat problemes per distingir qui era diputada, membre del govern o representant institucional. En canvi, en notícies més curtes o amb menys subjectes, ha estat més precisa a l'hora d'atribuir els rols laborals.
- *Informacions sobre futbol, siguin prèvies o cròniques*: Les notícies esportives, especialment les de futbol inclouen molts subjectes, entre jugadors titulars, suplents, entrenadors i altres membres del cos tècnic i persones de la directiva de l'equip. Tot sovint, aquestes informacions no explicitaven l'ocupació de les persones, tot i que era deduïble pel context. La IA ha mostrat un comportament irregular. En alguns pocs casos, la IA ha entès que la notícia parlava d'esportistes i els ha assignat la professió; en la majoria, ha deixat el camp en blanc.

Una de les característiques de la IA és que, a diferència d'una persona real, la màquina pot proporcionar respostes diferents per a una mateixa pregunta cada vegada que se li demani. L'aleatorietat en l'execució de les ordres per part del model de llenguatge en aquests supòsits ha minvat dràsticament l'indicador de la fiabilitat en la detecció d'ocupacions. Les indicacions facilitades per al camp *ocupació* implicaven un ventall més ampli de respostes (retornar la professió en una o dues paraules) a diferència de la resta de columnes, on només havia de respondre amb l'edat o bé amb un codi binari (0 o 1).

L'assignació de les feines s'incrementa de manera significativa quan el buidatge de les dades el fa un humà. Per aquest motiu, en cas de fer servir el codi en el futur, cal millorar les

instruccions que s'envien a l'API per augmentar el nombre de respostes i, en conseqüència, la precisió del treball.

Finalment, totes aquestes estadístiques s'han desat a un nou marc de dades amb el nom *st_fiabilitat* (annex 2) per facilitar la seva explotació, si escau, en una visualització de dades.

5.9. Creació d'indicadors estadístics

A més de les dades sobre la fiabilitat de les tasques executades per la IA, s'han generat onze marcs de dades amb els següents indicadors estadístics:

- Estadístiques generals (*st_generals*, annex 3):
 - Total de notícies analitzades
 - Notícies sense persones
 - Notícies amb persones
 - Dones esmentades a les notícies
 - Homes esmentats a les notícies
 - Persones de sexe indeterminat esmentades a les notícies
 - Persones per notícia (màxim, mínim i mitjana)
 - Notícies amb una sola persona esmentada
 - Mitjana de persones incloent només les notícies que esmenten persones
- Estadístiques de presència de dones a les notícies (*st_notícies_dones*, annex 4):
 - Notícies sense dones
 - Notícies amb dones
 - Notícies sense dones per mitjà
 - Notícies amb dones per mitjà
 - Notícies sense dones per secció
 - Notícies amb dones per secció
 - Notícies sense dones per mitjà i secció
 - Notícies amb dones per mitjà i secció
- Estadístiques per ocupació (*st_ocupacions*, annex 5):
 - Dones esmentades per ocupació
 - Homes esmentats per ocupació
 - Persones de sexe indeterminat esmentades per ocupació
 - Dones esmentades per ocupació i mitjà
 - Homes esmentats per ocupació i mitjà
 - Persones de sexe indeterminat esmentades per ocupació i mitjà
 - Dones esmentades per ocupació i secció
 - Homes esmentats per ocupació i secció
 - Persones de sexe indeterminat esmentades per ocupació i secció
- Estadístiques per rol familiar (*st_rolfamiliar*, annex 6):
 - Dones amb rol familiar esmentat
 - Dones sense rol familiar esmentat
 - Homes amb rol familiar esmentat

- Homes sense rol familiar esmentat
- Persones de sexe indeterminat amb rol familiar esmentat
- Persones de sexe indeterminat sense rol familiar esmentat
- Estadístiques sobre la funció de les persones a la notícia (*st_funció*, annex 7):
 - Dones amb funció esmentada
 - Homes amb funció esmentada
 - Persones de sexe indeterminat amb funció esmentada
 - Dones sense funció esmentada
 - Homes sense funció esmentada
 - Persones de sexe indeterminat sense funció esmentada
 - Protagonistes total
 - Protagonistes dones
 - Protagonistes homes
 - Protagonistes de sexe indeterminat
 - Portaveus total
 - Portaveus dones
 - Portaveus homes
 - Portaveus de sexe indeterminat
 - Expertes dones
 - Experts homes
 - Expertes de sexe indeterminat
 - Testimonis dones
 - Testimonis homes
 - Testimonis de sexe indeterminat
 - Víctimes dones
 - Víctimes homes
 - Víctimes de sexe indeterminat
- Estadístiques dels elements de titulació (*st_titulació*, annex 8):
 - Dones en títols i subtítols
 - Homes en títols i subtítols
 - Persones de sexe indeterminat en títols i subtítols
 - Dones en títols i subtítols per mitjà
 - Homes en títols i subtítols per mitjà
 - Persones de sexe indeterminat en títols i subtítols per mitjà
- Estadístiques d'autoria de les notícies (*st_firmes*, annex 9):
 - Notícies escrites per periodista o periodistes dones
 - Notícies escrites per periodista o periodistes homes
 - Notícies escrites per periodista o periodistes indeterminats (agències, redacció, sigles...)
 - Notícies escrites per periodistes d'ambdós sexes
 - Notícies escrites per periodista o periodistes dones per mitjà
 - Notícies escrites per periodista o periodistes homes per mitjà
 - Notícies escrites per periodista o periodistes indeterminats per mitjà
 - Notícies escrites per periodistes d'ambdós sexes per mitjà

- Notícies escrites per periodista o periodistes dones per secció
- Notícies escrites per periodista o periodistes homes per secció
- Notícies escrites per periodista o periodistes indeterminats per secció
- Notícies escrites per periodistes d'ambdós sexes per secció
- Notícies escrites per periodista o periodistes dones per mitjà i secció
- Notícies escrites per periodista o periodistes homes per mitjà i secció
- Notícies escrites per periodista o periodistes indeterminats per mitjà i secció
- Notícies escrites per periodistes d'ambdós sexes per mitjà i secció
- Estadístiques de biaix de gènere segons l'autoria (*st_biaixgènere*, annex 10):
 - Dones en notícies segons el sexe de l'autoria
 - Homes en notícies segons el sexe de l'autoria
 - Persones de sexe indeterminat en notícies segons el sexe de l'autoria
 - Expressió del rol familiar en notícies segons el sexe de l'autoria
 - Funció en notícies segons el sexe de l'autoria
- Estadístiques relatives a les fotografies (*st_imatges*, annex 11):
 - Fotografies amb persones
 - Fotografies sense persones
 - Fotografies amb dones
 - Fotografies sense dones
 - Fotografies amb homes
 - Fotografies sense homes
 - Fotografies amb persones de sexe indeterminat
 - Fotografies sense persones de sexe indeterminat
- Estadístiques d'edat (*st_edats*, annex 12):
 - Persones per rang d'edat
 - Persones per rang d'edat i sexe
 - Persones per rang d'edat i ocupació
- Sèries temporals amb dades de l'estudi *Un dia a les notícies en clau de gènere* (Altés Rufias, 2014) (*st_sèrietemporal*, annex 13):
 - Dones per any
 - Homes per any
 - Dones amb rol familiar esmentat per any
 - Homes amb rol familiar esmentat per any
 - Dones protagonistes per any
 - Homes protagonistes per any
 - Dones expertes per any
 - Homes experts per any
 - Dones en fotografies per any
 - Homes en fotografies per any
 - Dones redactores per any
 - Homes redactors per any

Les taules s'han creat per facilitar la redacció dels reportatges. També inclouen descriptors pensats per ser explicatius dins de les visualitzacions de dades i haver de fer el mínim de

modificacions. A més, els percentatges s'han expressat en un format que sigui acceptat per les eines de visualització.

5.10. Elaboració dels reportatges, el web i les visualitzacions de dades

L'etapa final del procés de treball ha consistit en la redacció dels reportatges i el disseny de les visualitzacions de dades, que s'han publicat a una plana web. Tant els articles com les infografies s'han basat en les estadístiques elaborades en l'apartat anterior.

Per a la fase de prototipatge de la pàgina web, s'han establert els requisits següents:

- Fer servir un sistema de gestió de continguts (CMS) extern, que serveixi per configurar i actualitzar el web a través d'un portal d'administració.
- Disposar de plantilles que facilitin el disseny i permetin donar un aspecte més professional al web sense haver de programar les pàgines des de zero.
- Incloure emmagatzematge gratuït, ja que, atès el volum del projecte, la capacitat que es preveu necessitar és inferior a 100 megabytes.
- Permetre la publicació sense necessitat de contractar un domini de primer nivell.
- Utilitzar plantilles CSS per, en cas que sigui necessari, fer canvis massius d'estils en comptes de canvis manuals.
- Permetre la incrustació d'elements externs, preferiblement *iframes*, per a la publicació de les infografies interactives.

Una vegada establerts els mínims, es van considerar tres opcions gratuïtes per gestionar el web: Wordpress, Wix i Squarespace. Es va optar per dissenyar el web amb Wordpress per les següents raons:

- *Permet una major flexibilitat en la personalització sense haver d'escriure codi web:* El sistema de blocs de Wordpress facilita la realització del prototip. El nivell de personalització supera el que és possible fer amb Wix i Squarespace sense haver d'escriure codi. A més, Wordpress disposa de més complements que permeten afegir funcionalitats, tot i que aquests són de pagament.
- *Inclou més plantilles predissenyades:* Després de comprovar les opcions de Wix i Squarespace per a un web de notícies, les possibilitats eren molt més limitades que les que oferia Wordpress gratuïtament.
- *Té més opcions de suport:* Wordpress és una plataforma molt estesa, la qual cosa fa que sigui més fàcil solucionar problemes de disseny gràcies a la seva comunitat d'usuaris, fòrums, tutorials, guies i desenvolupadors.
- *Disposa d'un model de subscripcions amb més graons:* Inicialment, el projecte es basa en un web gratuït. En cas que sigui necessari augmentar l'emmagatzematge, que el tràfic es dispari o que es requereixi alguna funció de pagament, Wordpress permetria una major escalabilitat del portal amb un cost menor de subscripció.

El web és responsiu i s'han arranjat tres versions distintes: per a ordinadors, per a tauletes i per a dispositius mòbils. S'han configurat els textos predefinits per tal que es mostrin en català. També s'han tingut en compte principis d'experiència d'usuari (UX) en el disseny, per

tal que la càrrega de la plana sigui ràpida i la informació sigui fàcilment accessible. El prototip s'ha basat en la plantilla Tenku i s'ha personalitzat per tal que el disseny s'adigui al d'un portal de notícies. La barra de navegació conté cinc opcions:

- Articles: Filtra totes aquelles entrades amb l'etiqueta "Article".
- Infografies: Filtra totes les entrades amb l'etiqueta "Infografia".
- Metodologia: Inclou un resum de l'apartat metodològic d'aquesta memòria.
- Dades obertes: Permet la descàrrega de les microdades, les estadístiques i el codi font, així com posar-se en contacte amb l'autor del projecte.
- Sobre aquest projecte: Referencia l'origen i l'autoria del treball.

El disseny de la portada s'ha individualitzat i s'ha personalitzat a partir de diversos blocs visuals. Els articles inclouen una columna addicional a la dreta per referenciar notícies relacionades, cosa que no passa amb la resta de pàgines. Les pàgines d'infografia premien el gran format per tal de veure'n els detalls amb major resolució, però sense renunciar a l'estètica general del web.

La incrustació de les visualitzacions de dades va ser l'escull més difícil. Tot i que Wordpress permet afegir les infografies a partir d'un enllaç amb un bloc predefinit d'incrustació, aquestes no es mostraven en tot l'ample de la columna de l'article. Per tal de resoldre el problema, va ser necessari afegir codi CSS addicional a l'editor del web.

Per a les visualitzacions, s'han considerat dues eines que permeten fer-les interactives, per tal de complementar les dades amb *tooltips*, filtres o altres elements que faciliten la comprensió de les dades: Flourish i Datawrapper. Totes dues tenen opcions gratuïtes i estan pensades per aconseguir mostrar grans volums d'informació de manera simple i sense tenir coneixements de codi. En aquest apartat, es va considerar l'ús de PowerBI o Tableau. Això no obstant, Flourish i Datawrapper varen ser les opcions preferides, atès que presenten els avantatges següents:

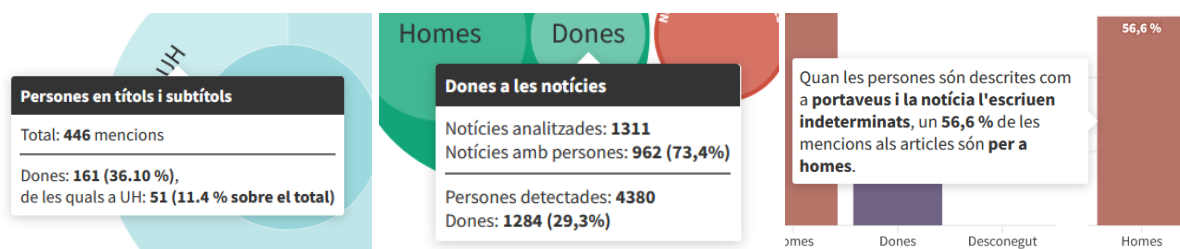
- *Facilitat d'ús i corba d'aprenentatge*: Datawrapper i Flourish tenen interfícies fàcils d'utilitzar, que els fan accessibles per a periodistes. PowerBI i Tableau són més complexos i requereixen més experiència tècnica.
- *Compatibilitat d'integració i publicació*: Tant Datawrapper com Flourish ofereixen una integració perfecta amb plataformes web, especialment amb sistemes de gestió de continguts com Wordpress. Aquesta integració garanteix que les visualitzacions es puguin incrustar fàcilment, a diferència de PowerBI i Tableau, que poden requerir més passes per a la integració.
- *Atractiu estètic*: Datawrapper i Flourish ofereixen plantilles i opcions de disseny especialment adaptades per a finalitats periodístiques, mentre que PowerBI i Tableau tenen dissenys més orientats a negocis.
- *Compatibilitat mòbil*: Atès que el web és responsiu, era adient (tot i que no imprescindible) que l'eina de visualització també adaptés els gràfics. Datawrapper i Flourish ho ofereixen per defecte per a la majoria de tipus de gràfics, mentre que PowerBI i Tableau sovint estan més orientats a la visualització d'escriptori.

- **Cost:** Datawrapper i Flourish ofereixen versions gratuïtes amb moltes opcions. Els seus plans de subscripció són més rendibles en comparació amb els costos més elevats de PowerBI i Tableau.

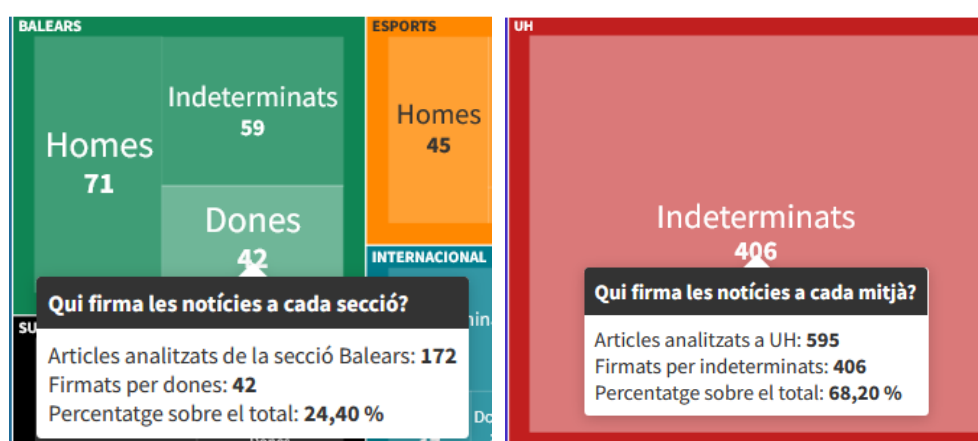
La pàgina web s'ha publicat sota l'URL <https://premsaigenera.wordpress.com/>. Inclou nou articles periodístics i setze infografies interactives. Les visualitzacions de dades s'han dissenyat amb els criteris de representar les dades de manera fidedigna. El tipus de gràfic ha depès de les dades que es volien ressaltar. També s'ha tingut en compte un criteri estètic i s'han plantejat infografies variades que no sempre tinguessin el mateix estil.

Per a moltes de les infografies, s'han configurat finestres emergents personalitzades que permeten detallar, ampliar o complementar les dades (figures 4A, 4B i 4C). Algunes visualitzacions mostren textos diferents segons la pestanya que es mostra o el filtre que s'aplica (figures 5A i 5B). En certs casos, especialment quan un gràfic contenia diverses variables d'anàlisi, s'ha optat per missatges amb text que servissin per esclarir o descriure millor idees complexes, amb la dificultat afegida de fer que aquests missatges predefinitos fossin gramaticalment correctes (figures 6A, 6B i 6C).

Els reportatges i les visualitzacions de dades elaborats estan inclosos en l'annex 14 d'aquest document. Alguns dels continguts, així com el disseny de la pàgina web, es poden observar a l'annex 15.



Figures 4A, 4B i 4C. Exemples de finestres emergents a les infografies.



Figures 5A i 5B. Variacions de les finestres emergents a la infografia titulada "Qui firma les notícies?". Els missatges canvien segons el filtre que s'aplica.



Figures 6A, 6B i 6C. Finestres emergents a la infografia “L’ús de l’edat en la premsa digital de Mallorca”. Els textos mostren variacions per a singular, plural i “Ocupació no esmentada”.

5.11. Costos del projecte i temps de processament

Quant a la part d’anàlisi amb intel·ligència artificial, el projecte ha suposat el processament de:

- 276 noms de periodistes de les firmes d’articles;
- 1.311 títols, subtítols i cossos de notícies, que sumen 656.549 paraules (o bé 3.952.419 caràcters);
- 1.302 imatges;
- 4.380 persones trobades a les notícies amb les variables d’anàlisi associades;
- 3.084 noms distintes de les persones esmentades a les notícies, i
- 564 professions d’aquests subjectes.

Amb aquest volum de dades, els costos en eines d’intel·ligència artificial per a la realització del projecte han estat els següents:

Objecte	Model	Preu en dòlars	Preu en euros
Sexe de les firmes dels articles	GPT 3.5-Turbo	0,09 \$	0,08 €
Títols, subtítols i textos (proves)	GPT-3.5 Turbo i GPT-4 Turbo	2,69 \$	2,52 €
Títols, subtítols i textos (processament)	GPT-4 Turbo	18,49 \$	17,29 €
Títols, subtítols i textos (reprocessament)	GPT-4 Turbo	3,19 \$	2,98 €
Fotografies	GPT-4 Vision	3,20 \$	2,94 €
Sexe de les persones	GPT-4 Turbo	0,69 \$	0,63 €
Codificació de les ocupacions	GPT-4 Turbo	0,22 \$	0,20 €
Total		28,57 \$	26,64 €

Pel que fa als continguts de la pàgina web i amb la finalitat de no utilitzar elements amb drets de propietat intel·lectual, les il·lustracions que acompanyen els articles també s’han generat amb intel·ligència artificial. Aquesta tasca s’ha realitzat amb el model Dall-e 3 d’OpenAI. El processament també s’ha dut a terme amb l’API amb un codi independent al de la resta del projecte. En aquest cas, el cost de generació de les imatges (proves incloses) ha estat de 2,04 dòlars (1,85 euros).

Per altra part, els temps de processament de les dades han estat els següents:

Objecte	Model	Temps		
Sexe de les firmes dels articles	GPT 3.5-Turbo	16 min	30 s	
Títols, subtítols i textos (processament)	GPT-4 Turbo	1 h	15 min	4 s
Títols, subtítols i textos (reprocessament)	GPT-4 Turbo	17 min	38 s	
Fotografies	GPT-4 Vision	1 h	10 min	0 s
Sexe de les persones	GPT-4 Turbo	8 min	11 s	
Codificació de les ocupacions	GPT-4 Turbo	1 min	9 s	
Total		3 h	8 min	32 s

El període de desenvolupament de l’estudi *Un dia a les notícies en clau de gènere* (Altés Rufias, 2014), utilitzat com a referent d’aquest treball, va ser més ampli. Va començar a gestar-se a finals de 2013. D’acord amb l’informe, en varen formar part tres investigadores i onze voluntàries. El febrer de 2014 varen tenir lloc les sessions de formació a les voluntàries, que varen codificar 469 notícies de 12 mitjans de comunicació en aproximadament un mes. Els resultats del treball es varen fer públics a principis de desembre de 2014 (Mèdia.cat, 2014).

Tot i que ambdós treballs no són comparables ni per la metodologia ni per l’abast (l’exercici de 2014 incloïa més variables qualitatives i també una anàlisi qualitativa), val la pena assenyalar que les possibilitats de la intel·ligència artificial i el processament automatitzat han permès que, amb el codi de programació actual, es puguin codificar 1.311 notícies en 3 hores, 8 minuts i 32 segons. A aquest temps de processament, cal afegir-hi un període addicional per a la revisió manual de resultats i la possible introducció de canvis en les indicacions a la IA per refinar els resultats.

6. Conclusions

6.1. Respecte dels objectius i la justificació del projecte

El projecte ha assolit els objectius inicials que es varen plantejar. Primerament, s'ha desenvolupat un codi de programació reutilitzable, aportant un recurs valuós per a futures investigacions en el camp periodístic. Aquest codi ha permès automatitzar l'anàlisi de les dades, demostrant la utilitat i fiabilitat de les eines d'intel·ligència artificial en el context de l'anàlisi periodística.

En aquest procés, s'ha aprofundit en l'estudi de la representació de les dones en els mitjans de comunicació de Mallorca, tot proporcionant una visió detallada i actualitzada en la matèria. Finalment, la publicació de les dades en una plataforma web amb infografies interactives ha facilitat la difusió dels resultats. A més, també s'han complert els objectius secundaris de control de les despeses i del temps de processament per tal de reafirmar l'eficiència de recursos del projecte.

Si s'observa la justificació original del treball, l'anàlisi dels resultats obtinguts demostra que la raó de ser d'aquesta recerca manté la seva validesa. La contribució professional al periodisme s'ha materialitzat en l'exploració de noves metodologies per a la recopilació, anàlisi i presentació de dades, reafirmant la importància de combinar recerca acadèmica amb aplicació pràctica en el periodisme digital.

La necessitat d'automatitzar i reutilitzar eines tecnològiques, un dels arguments fonamentals per desenvolupar aquesta feina, s'ha confirmat amb l'èxit del desenvolupament i aplicació del codi de programació i la seva compartició en obert. A més, l'ús de la intel·ligència artificial en l'anàlisi de textos ha servit per provar la utilitat d'aquest recurs per a la professió periodística.

Queda per demostrar si la difusió dels resultats de la recerca a través d'una pàgina web contribuirà significativament a la conscienciació social sobre la representació de les dones en els mitjans digitals de Mallorca.

6.2. Respecte de la metodologia i el procés de treball

Quant a la metodologia establerta, on es va plantejar l'ús de models de llenguatge d'intel·ligència artificial per a l'anàlisi automatitzada de textos, es pot observar que l'aplicació de la IA ha estat fonamental. A més, ha complert majoritàriament amb les expectatives, tot i alguns desafiaments que han aparegut en la interpretació de les ordres i la classificació de les dades.

Especialment laboriosa ha estat la creació de les indicacions (*prompts*) per al processament de dades. En alguns casos, sobretot en el de l'anàlisi de les ocupacions dels individus, es fa necessari ajustar les instruccions en el futur per tal d'augmentar la fiabilitat i aconseguir la precisió desitjada.

Les dificultats sorgides durant el procés de treball han subratllat la necessitat d'una comprensió més profunda del funcionament de la IA en l'anàlisi textual. Aquestes incidències també suggereixen que, encara que és una eina tecnològica potent, requereix una supervisió i uns ajustaments detallats per a ser emprada de manera òptima.

6.3. Respecte del projecte professionalitzador

Quant al vessant professionalitzador del present treball, s'han aplicat de manera pràctica els coneixements teòrics adquirits en assignatures del màster com ara Periodisme de dades, Anàlisi i visualització de dades, Tècniques avançades d'anàlisi i visualització de dades, Eines i recursos digitals, i Comunicació digital i societat.

Adicionalment, s'ha complert amb el repte personal de promoure noves habilitats tècniques i d'explorar eines que no són d'ús freqüent a les redaccions com ara VisualStudio Code, Python, els models de llenguatge d'OpenAI, Wordpress, Flourish o Datawrapper. A més, s'han analitzat els avantatges i inconvenients que algunes d'aquestes plataformes tenen respecte dels seus competidors.

El resultat final (el web amb les notícies i els infografies interactives) ha suposat el desenvolupament per complet d'un producte periodístic digital, la viabilitat del qual ha quedat palesa amb la publicació del portal.

En conclusió, aquest treball ha tractat de fer una contribució significativa en l'àmbit acadèmic i, fonamentalment, en el periodístic. Des d'una perspectiva acadèmica, aprofundeix en les possibilitats de la intel·ligència artificial per a l'anàlisi de textos periodístics i ofereix una metodologia de treball que pot servir de base per a estudis futurs.

Pel que fa al sector periodístic, a més de facilitar noves eines tecnològiques, proporciona una visió crítica i actualitzada de la representació de gènere en els mitjans locals digitals de Mallorca, un fet essencial en un present en el qual la igualtat de gènere es troba en el punt de mira social.

Les dades obtingudes permeten una comprensió més profunda de com els dos principals digitals de l'illa tracten les qüestions de gènere i poden servir com a referència per a futures investigacions i per a què les redaccions desenvolupin pràctiques periodístiques més igualitàries.

7. Línies de desenvolupament futur

De cara al futur, es poden tenir en compte diverses millores en l'eficiència del codi, de l'anàlisi de dades i, en conseqüència, dels resultats del projecte. Aquestes són algunes de les possibles accions que es poden emprendre per al desenvolupament futur:

1. Crear nous mecanismes per augmentar la fiabilitat de les dades: Un projecte de dades que no només les mostra, sinó que també les genera, ha de poder certificar la fiabilitat dels resultats. Tot i que la intel·ligència artificial (igual que la humana) és fal·lible, es fa imprescindible executar més comprovacions de les respostes generades pel model de llenguatge per tal de reduir el marge d'error. L'augment de la precisió i de la sensibilitat de la IA (i, per extensió, dels resultats) són especialment necessaris en un projecte de dades obertes, on tots els detalls estan disponibles per a l'escrutini públic.
2. Refinar el codi de programació:
 - a. *Millorar els bucles d'anàlisi:* Tot optimitzant els bucles utilitzats per a l'anàlisi de dades, es pot augmentar significativament la velocitat d'execució del codi. Això implica perfeccionar la programació amb un major coneixement de Python.
 - b. *Espaiar les cridades a l'API per evitar errors:* És necessari gestionar més acuradament la freqüència de les sol·licituds a l'API per evitar respostes amb errades pels límits a les sol·licituds o als *tokens*. Una opció és programar temps d'espera variables en funció dels retorns de l'API. La implementació d'un enfocament més estratègic per sincronitzar aquestes sol·licituds minimitzarà les errades i les iteracions.
 - c. *Recomptar internament els tokens:* Una de les evolucions més substancials des de l'inici al final del projecte ha estat la retallada de costos a l'hora de tractar amb l'API, gràcies a l'enviament de dades per tongades. A més d'implementar aquesta metodologia a totes les passes, una altra opció és afegir alguna eina que ajudi a recomptar els *tokens* abans d'enviar-los a l'API per controlar millor els costos dels missatges d'entrada.
 - d. *Incloure més registres d'errades amb més informació:* Al llarg del projecte, es va determinar la necessitat d'incloure fitxers de registre (*logs*) que emmagatzemassin els missatges d'error. En alguns casos, han servit per reprocessar les dades i minimitzar costos; en d'altres, s'han utilitzat per millorar els *prompts* que s'enviaven a l'API. Afegir mecanismes de registre més complets ajudarà a un millor seguiment de les tasques i a una major depuració del codi.
3. Automatitzar el raspat d'informació web (*web scraping*): La implementació d'un raspat web automatitzat eliminaria la necessitat d'introduir manualment el codi font a una carpeta i, eventualment, permetria que el codi s'executàs cada dia per analitzar les notícies sense intervenció manual. Un exemple és l'eina Selenium, un programari

de codi obert que permet automatitzar la navegació a través de pàgines web (p. ex., per clicar a un botó de “Veure més”) i aleshores executar el codi del raspap. Tot i això, tindria l’inconvenient que s’hauria de personalitzar per a cada web de cada diari de l’anàlisi. Això no obstant, aquest enfocament no només estalviaria temps, sinó que també augmentaria la precisió i la fiabilitat de la recollida de dades.

4. Incloure l’anàlisi de les persones de les quals no se n’esmenta el nom: En el present treball, s’ha utilitzat el nom com a referència per determinar el sexe de les persones que sortien a les notícies, però s’ha obviat l’anàlisi d’aquelles mencions genèriques a “una dona” o “un home”, atès que dificultava que la IA determinés a quantes dones o homes diferents es feia referència. Per tal de millorar el model, caldria fer que la IA reculli aquestes referències genèriques i les agrupi per persones en funció del context textual. Aquesta millora repercutiria significativament en l’anàlisi de les notícies de successos, en les quals tot sovint es fa referència a una persona pel seu sexe o per les seves sigles, però s’obvia el nom per raons ètiques o legals.
5. Millorar la gestió i l’anàlisi dels elements audiovisuals:
 - a. *Descarregar els textos i les imatges de les notícies per fer-ne una còpia de seguretat*: Per tal de protegir-se de la pèrdua de dades a causa de canvis a les pàgines web, es fa essencial implementar una funció de descàrrega d’imatges i textos amb finalitats de còpia de seguretat. En el present projecte, s’han trobat errors 404 en algunes imatges de *Diariodemallorca.es*. Després d’haver contactat amb l’equip web del mitjà, han explicat que pot haver passat per la retirada d’una imatge a petició d’una persona afectada o bé, en el cas d’imatges antigues, pels canvis de disseny del web. Tenir còpies de seguretat garantiria tenir un dipòsit fiable del conjunt de l’anàlisi, mantenint la integritat de les dades al llarg del temps per poder consultar-les.
 - b. *Analitzar totes les imatges de la notícia i no només la imatge principal*: La llibreria Newspaper3k que s’ha emprat per recuperar les notícies té la limitació que només ha permès obtenir la imatge primera de cada article. Per aquest motiu, cal ampliar l’abast de l’anàlisi d’imatges per incloure totes les fotografies. Aquest fet es fa especialment important als grans reportatges i les fotogalleries, on l’anàlisi d’una sola imatge pot resultar esbiaixat.
 - c. *Estudiar la viabilitat d’analitzar els vídeos*: Python disposa de llibreries que permeten analitzar vídeos i extreure’n *frames*. A partir d’aquí, es podrien analitzar aquestes captures per realitzar-ne l’anàlisi de gènere. Atesa l’especificitat d’aquesta tasca, és un dels desenvolupaments que poden requerir un major esforç per tal de garantir resultats precisos.
6. Crear una base de dades per tal de connectar-la a les visualitzacions interactives: Una altra possible línia de desenvolupament és la sistematització de les dades recollides amb la creació d’una arquitectura de base de dades. Aquest supòsit permetria connectar-les amb les visualitzacions interactives, de forma que les infografies s’actualitzessin diàriament sense intervenció manual.

7. Crear unes línies mestres d'identitat visual per a les infografies: Ja que moltes visualitzacions comparen els mateixos elements o sèries (p. ex., dones, homes o diaris), és viable estandarditzar els colors i altres elements gràfics de les infografies. Això facilitaria la comprensió dels elements d'anàlisi i contribuiria a crear una identitat visual corporativa per a la pàgina web del projecte.

Aquestes línies de desenvolupament pretenen no només millorar els processos actuals, sinó que també tenen per objectiu ampliar les capacitats del projecte per tal que sigui útil tant per a la professió periodística com per al sector de la recerca en comunicació. Emprendre els avanços proposats suposaria incrementar les capacitats actuals, tot assegurant que les anàlisis continuïn sent fiables, eficaces i rellevants gràcies a l'evolució dels paradigmes tecnològics.

8. Bibliografia

- Altés Rufias, E. (2014). *Un dia a les notícies en clau de gènere: protagonistes i periodistes als mitjans de comunicació de les Illes Balears*. Ajuda en Acció. <https://archivo.amecopress.net/-2020/pdf/UnDiaALesNoticiesFinal.pdf>
- Altés Rufias, E., i Majó Vázquez, S. (2012). *Representació i tractament de les dones i el gènere en la informació sobre economia*. <https://es.scribd.com/doc/112940056/Informe-MMCC-i-Ec>
- Bach Arús, M., Altés Rufias, E., Gallego Ayala, J., Plujà Calderón, M., i Puig Mollet, M. (2000). *El sexo de la noticia: reflexiones sobre el género en la información y recomendaciones de estilo* (M. Bach i Arús, Ed.) [Llibre]. Icaria.
- Beckett, C., i Yaseen, M. (2023). *Generating Change: a global survey of what news organisations are doing with AI*.
- Buades Àvila, C. (2018). *La Manada: un abans i un després? L'impacte del cas dels Sanfermines en el relat comunicatiu de la violència cap a les dones* [Treball final de màster no publicat]. Institut Interuniversitari d'Estudis de Dones i Gènere.
- Bunce, M. (2019). Management and resistance in the digital newsroom [Article]. *Journalism (Londres, Anglaterra)*, 20(7), 890-905. <https://doi.org/10.1177/1464884916688963>
- Caramancion, K. M. (2023). News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking [Article]. *ArXiv.Org*. <https://doi.org/10.48550/arxiv.2306.17176>
- Casselman, B., Conlen, M., Fischer-Baum, R., Aranke, K., i Nield, D. (2016, 13 de juliol). *Gun deaths in America*. FiveThirtyEight. <https://fivethirtyeight.com/features/gun-deaths/>
- Comas d'Argemir, D. (2015). News of partner femicides: The shift from private issue to public problem [Article]. *European Journal of Communication (Londres)*, 30(2), 121-136. <https://doi.org/10.1177/0267323114544865>
- Cox, M., Morris, J., Higgins, T., Corona, A., Lamonaca, L., Mintz, M., i Behn, A. S. (2016). *Inside Airbnb*. <http://insideairbnb.com/>
- Drummond, J. (2023, 28 de setembre). Multiple newsrooms are using AI to optimize work, despite concerns about bias and accuracy. *The Verge*. <https://www.theverge.com/2023/9/28/23894651/ai-newsroom-journalism-study-automation-bias>
- Fundación Civio (2010). *¿Dónde van mis impuestos?* <https://dondevanmisimpuestos.es/>
- Fundación Civio (2013). *El indultómetro*. <https://civio.es/el-indultometro/>
- Fundación Civio (2015). *¿Quién cobra la obra?* <https://civio.es/quien-cobra-la-obra/>
- Gallego Ayala, J. (2014). *De reinas a ciudadanas* (1a ed.) [Llibre]. Editorial UOC.

- Gamson, W. A., Croteau, D., Hoynes, W., i Sasson, T. (1992). Media images and the social construction of reality [Article]. *Annual Review of Sociology*, 18(1), 373-393. <https://doi.org/10.1146/annurev.so.18.080192.002105>
- Gearin, C. (2016, 20 de juny). AI just got a big boost in its ability to understand the news. *New Scientist*. <https://www.newscientist.com/article/2094385-ai-just-got-a-big-boost-in-its-ability-to-understand-the-news/>
- Giró, X. (1999). *Anàlisi crítica del discurs sobre nacionalisme i identitat als editorials de la premsa diària publicada a Catalunya des de la transició fins al govern del PP (1977-1996)*.
- Goffman, Erving. (2006). *Frame analysis: los marcos de la experiencia* [Llibre]. Centro de Investigaciones Sociológicas.
- Jenkins, J., Mathur, M., Muyskens, J., Nakhlawi, R., Ric, S., i Tran, A. B. (2015, 30 de maig). Fatal force: Police shootings database 2015-2023. *The Washington Post*. <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>
- Kitzinger, Jenny. (2004). *Framing abuse media influence and public understanding of sexual violence against children* [Llibre]. Pluto Press.
- Kundu, R. (2022, 16 de desembre). *F1 Score in Machine Learning: Intro i Calculation*. V7labs. <https://www.v7labs.com/blog/f1-score-guide>
- Li, X., Zhang, Y., i Malthouse, E. C. (2023). A Preliminary Study of ChatGPT on News Recommendation: Personalization, Provider Fairness, Fake News [Article]. *ArXiv.Org*. <https://doi.org/10.48550/arxiv.2306.10702>
- Linn, A. (2018, 15 de gener). *Microsoft creates AI that can read a document and answer questions about it as well as a person*. The AI Blog. <https://blogs.microsoft.com/ai/microsoft-creates-ai-can-read-document-answer-questions-well-person/>
- Macharia, S., Djerf-Pierre, M., Ross, K., Edström, M., i López, S. (2020). *Who makes the news?: 6th Global Media Monitoring Project*. https://whomakesthenews.org/wp-content/uploads/2021/11/GMMP2020.ENG_FINAL.pdf
- Mayo-Cubero, M. (2022). A multivariable analysis on news production in Spain: digital newsroom profile, polyvalent journalists and gender perspective [Article]. *Communication & Society*, 35(3), 1-14. <https://doi.org/10.15581/003.35.3.1-14>
- McLuhan, M. (1966). *Understanding media: the extensions of man* [Llibre]. McGraw-Hill Books.
- Mèdia.cat (2014, 4 de desembre). «Les dones protagonitzen el 23% de les notícies dels mitjans a les Balears». *Mèdia.cat*. <https://www.media.cat/2014/12/04/elvira-altes-dones-protagonistes-noticies-mitjans-balears/>
- Moses, L. (2017, 14 de setembre). The Washington Post's robot reporter has published 850 articles in the past year. *Digiday*. <https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/>

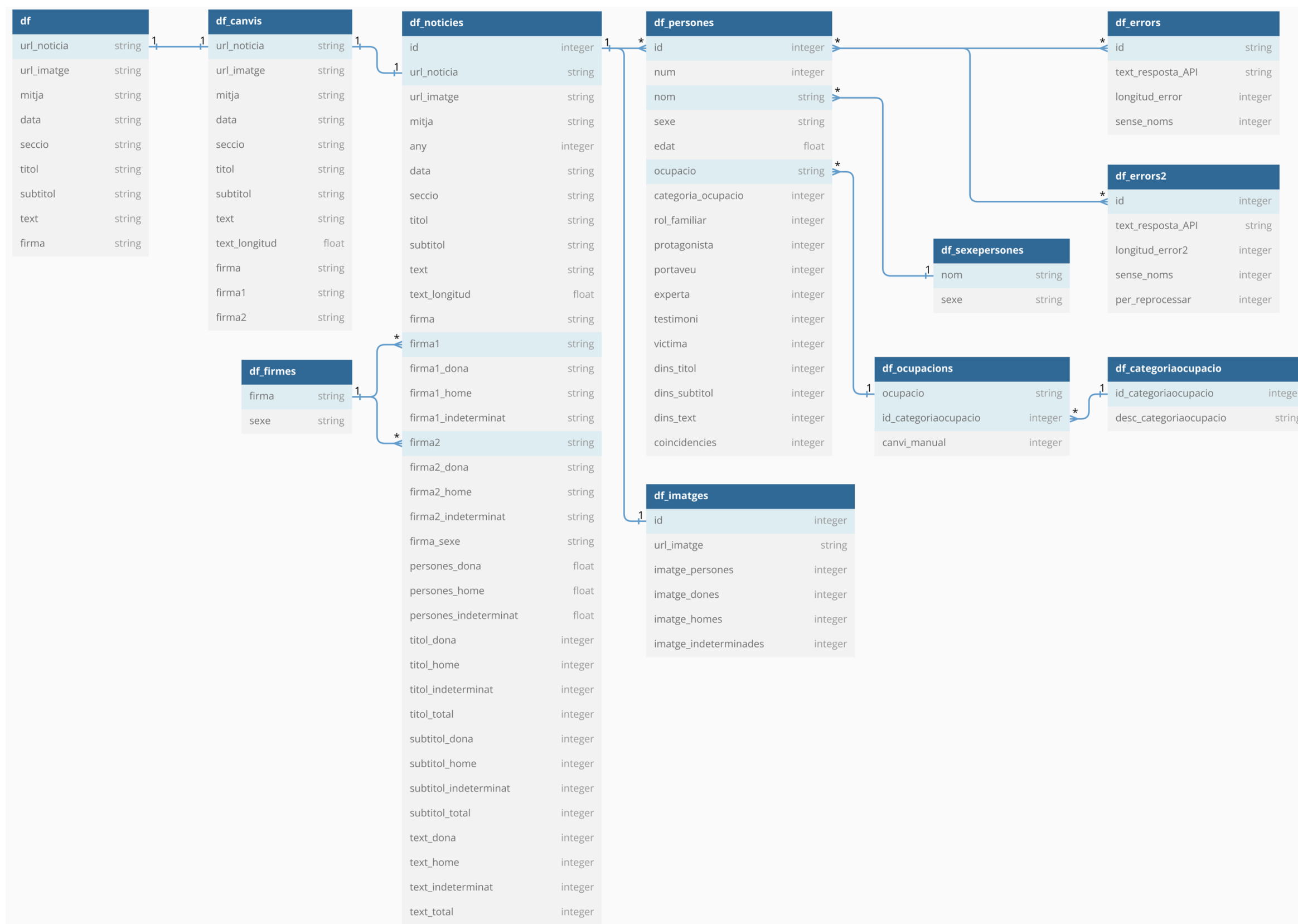
- Mullin, B., i Grant, N. (2023, 19 de juliol). Google Tests A.I. Tool That Is Able to Write News Articles. *The New York Times*. <https://www.nytimes.com/2023/07/19/business/google-artificial-intelligence-news-articles.html>
- OJD interactiva (2023). *Auditoría de medios digitales - Octubre 2023*. <https://www.ojdinteractiva.es/medios-digitales>
- OpenAI (2022, 30 de novembre). *Introducing ChatGPT*. <https://openai.com/blog/chatgpt>
- OpenAI (2023, 6 de novembre). *New models and developer products announced at DevDay*. <https://openai.com/blog/new-models-and-developer-products-announced-at-devday>
- Rajpurkar, P. (2023). *The Stanford Question Answering Dataset*. <https://rajpurkar.github.io/SQuAD-explorer/>
- Rajpurkar, P., Zhang, J., Lopyrev, K., i Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2383-2392. <https://doi.org/10.18653/v1/d16-1264>
- Renó, D., i Renó, L. (2015). Las nuevas redacciones, el «big data» y los medios sociales como fuentes de noticias [Article]. *Estudios sobre el mensaje periodístico*, 21 SE, 131. https://doi.org/10.5209/rev_ESMP.2015.v21.51135
- Stamenković, S. (2015). Journalism and media future: creating identity and reality [Article]. *In Medias Res (Zagreb, Online)*, 4(6), 838-858.
- The Washington Post PR (2020, 13 d'octubre). The Washington Post to debut AI-powered audio updates for 2020 election results. *The Washington Post*. <https://www.washingtonpost.com/pr/2020/10/13/washington-post-debut-ai-powered-audio-updates-2020-election-results/>
- Tigas, M., Grochowski Jones, R., Ornstein, C., i Groeger, L. (2010, 18 d' octubre). Dollars for Docs. *ProPublica*. <https://projects.propublica.org/docdollars/>
- Vogt, D. J. (2020). *Distorted realities: the discursive construction of realities in digital news media*. Stockholm University, Department of Media Studies.
- WACC Global (2020). *GMMP 2020 - Monitoreo completo: guía rápida sobre la codificación*. <https://whomakesthenews.org/wp-content/uploads/2021/03/GMMP-2020-Monitoreo-completo-Guia-rapida-sobre-la-codificacion.pdf>
- WACC Global (2021). *The Global Media Monitoring Project*. <https://whomakesthenews.org/the-gmmp/>
- Willis, H. (2019). Journalism on Autopilot [Article]. *The Quill*, 107(4), 18-21.
- Woolston, C., i Perkel, J. M. (2020). Sliced, diced and digested: science ready in minutes. *Nature (Londres)*, 588(7837), S136-S137. <https://doi.org/10.1038/d41586-020-03415-w>

9. Annexos

Annex 1. Esquema logicofísic de les taules creades a partir del codi.....	44
Annex 2. Estadístiques de fiabilitat de les tasques executades per la IA	45
Annex 3. Estadístiques generals	46
Annex 4. Estadístiques de presència de dones a les notícies	47
Annex 5. Estadístiques per ocupació	50
Annex 6. Estadístiques per rol familiar	66
Annex 7. Estadístiques sobre la funció de les persones a la notícia.....	67
Annex 8. Estadístiques dels elements de titulació	68
Annex 9. Estadístiques d'autoria de les notícies	69
Annex 10. Estadístiques de biaix de gènere segons l'autoria.....	77
Annex 11. Estadístiques relatives a les fotografies.....	81
Annex 12. Estadístiques d'edat.....	82
Annex 13. Sèries temporals	85
Annex 14. Reportatges i infografies elaborats a partir de les dades.....	87
Annex 15. Captures de pantalla del web	112
Annex 16. Codi programat per a l'automatització de l'extracció i anàlisi de les notícies	120

Annex 1. Esquema logicofísic de les taules creades a partir del codi

Les taules es poden descarregar a la secció *Dades obertes* del web del projecte o a través de l'enllaç https://premsaigenere.files.wordpress.com/2023/12/premsaigenere_microdades-1.zip.



Annex 2. Estadístiques de fiabilitat de les tasques executades per la IA

descripció	valor	percentatge
Total de notícies a l'anàlisi	1311	
Notícies per comprovar manualment	131	9,99 %
Notícies comprovades manualment	131	100,00 %
Total de persones a l'anàlisi	4380	
Persones per comprovar manualment	566	12,92 %
Persones detectades correctament per la IA (veritables positius)	530	
Persones eliminades manualment (falsos positius)	36	
Persones afegides manualment (falsos negatius)	21	
Precisió		93,64 %
Reclam		96,19 %
Puntuació F1		94,90 %
Fiabilitat en analitzar nom	550	99,82 %
Fiabilitat en analitzar sexe	536	97,28 %
Fiabilitat en analitzar edat	551	100,00 %
Fiabilitat en analitzar ocupació	254	46,10 %
Fiabilitat en analitzar rol_familiar	548	99,46 %
Fiabilitat en analitzar protagonista	529	96,01 %
Fiabilitat en analitzar portaveu	536	97,28 %
Fiabilitat en analitzar experta	547	99,27 %
Fiabilitat en analitzar testimoni	549	99,64 %
Fiabilitat en analitzar víctima	549	99,64 %

Annex 3. Estadístiques generals

descripció	valor	percentatge
Total de notícies analitzades	1311	100.0
Notícies sense persones	349	26.62090007627765
Notícies amb persones	962	73.37909992372235
Dones totals	1284	29.315068493150687
Homes totals	2932	66.94063926940639
Indeterminades totals	164	3.744292237442922
Màxim de persones a una notícia	54	
Mínim de persones a una notícia	1	
Notícies amb una sola persona	245	18.688024408848207
Mitjana de persones per notícia	3.3409610983981692	
Mitjana de persones per notícia que esmenta persones	4.553014553014553	

Annex 4. Estadístiques de presència de dones a les notícies

descripció	mitjà	secció	valor	percentatge
Notícies sense dones	Total	Total	713	54.385964912280706
Notícies amb dones	Total	Total	598	45.614035087719294
Notícies sense dones per mitjà	DM	Total	378	52.79329608938548
Notícies sense dones per mitjà	UH	Total	335	56.30252100840336
Notícies amb dones per mitjà	DM	Total	338	47.20670391061452
Notícies amb dones per mitjà	UH	Total	260	43.69747899159664
Notícies sense dones per secció	Total	Balears	90	52.32558139534884
Notícies sense dones per secció	Total	Cultura	12	26.08695652173913
Notícies sense dones per secció	Total	Economia	32	60.37735849056604
Notícies sense dones per secció	Total	Esports	65	66.3265306122449
Notícies sense dones per secció	Total	Fotogalleries / Vídeos	44	77.19298245614034
Notícies sense dones per secció	Total	Internacional	36	50.0
Notícies sense dones per secció	Total	Miscel·lània	31	73.80952380952381
Notícies sense dones per secció	Total	Motor	12	85.71428571428571
Notícies sense dones per secció	Total	Nacional	160	46.3768115942029
Notícies sense dones per secció	Total	Opinió	27	49.09090909090909
Notícies sense dones per secció	Total	Palma	16	61.53846153846154
Notícies sense dones per secció	Total	Part forana	25	67.56756756756756
Notícies sense dones per secció	Total	Prensa rosa	4	23.52941176470588
Notícies sense dones per secció	Total	Publireportatges	14	70.0
Notícies sense dones per secció	Total	Reportatges	12	75.0
Notícies sense dones per secció	Total	Societat	29	39.189189189189186
Notícies sense dones per secció	Total	Successos	78	77.22772277227723
Notícies sense dones per secció	Total	Tecnologia	18	90.0
Notícies sense dones per secció	Total	Televisió / Cinema	8	17.391304347826086
Notícies amb dones per secció	Total	Balears	82	47.674418604651166
Notícies amb dones per secció	Total	Cultura	34	73.91304347826086
Notícies amb dones per secció	Total	Economia	21	39.62264150943396
Notícies amb dones per secció	Total	Esports	33	33.6734693877551
Notícies amb dones per secció	Total	Fotogalleries / Vídeos	13	22.807017543859647
Notícies amb dones per secció	Total	Internacional	36	50.0
Notícies amb dones per secció	Total	Miscel·lània	11	26.190476190476193
Notícies amb dones per secció	Total	Motor	2	14.285714285714285
Notícies amb dones per secció	Total	Nacional	185	53.62318840579711
Notícies amb dones per secció	Total	Opinió	28	50.90909090909091
Notícies amb dones per secció	Total	Palma	10	38.46153846153847
Notícies amb dones per secció	Total	Part forana	12	32.432432432432435
Notícies amb dones per secció	Total	Prensa rosa	13	76.47058823529412
Notícies amb dones per secció	Total	Publireportatges	6	30.0
Notícies amb dones per secció	Total	Reportatges	4	25.0
Notícies amb dones per secció	Total	Societat	45	60.810810810810814
Notícies amb dones per secció	Total	Successos	23	22.772277227722775
Notícies amb dones per secció	Total	Tecnologia	2	10.0

Notícies amb dones per secció	Total	Televisió / Cinema	38	82.6086956521739
Notícies sense dones per mitjà i secció	DM	Balears	44	53.65853658536586
Notícies sense dones per mitjà i secció	DM	Cultura	7	22.58064516129032
Notícies sense dones per mitjà i secció	DM	Economia	32	60.37735849056604
Notícies sense dones per mitjà i secció	DM	Esports	36	62.06896551724138
Notícies sense dones per mitjà i secció	DM	Fotogalleries / Vídeos	44	77.19298245614034
Notícies sense dones per mitjà i secció	DM	Internacional	22	50.0
Notícies sense dones per mitjà i secció	DM	Miscel·lània	18	69.23076923076923
Notícies sense dones per mitjà i secció	DM	Motor	12	85.71428571428571
Notícies sense dones per mitjà i secció	DM	Nacional	26	28.888888888888886
Notícies sense dones per mitjà i secció	DM	Opinió	13	46.42857142857143
Notícies sense dones per mitjà i secció	DM	Palma	10	66.66666666666666
Notícies sense dones per mitjà i secció	DM	Part forana	14	77.77777777777779
Notícies sense dones per mitjà i secció	DM	Prensa rosa	0	0.0
Notícies sense dones per mitjà i secció	DM	Publireportatges	11	68.75
Notícies sense dones per mitjà i secció	DM	Reportatges	12	75.0
Notícies sense dones per mitjà i secció	DM	Societat	19	37.254901960784316
Notícies sense dones per mitjà i secció	DM	Successos	44	73.33333333333333
Notícies sense dones per mitjà i secció	DM	Tecnologia	7	87.5
Notícies sense dones per mitjà i secció	DM	Televisió / Cinema	7	16.666666666666664
Notícies sense dones per mitjà i secció	UH	Balears	46	51.11111111111111
Notícies sense dones per mitjà i secció	UH	Cultura	5	33.33333333333333
Notícies sense dones per mitjà i secció	UH	Esports	29	72.5
Notícies sense dones per mitjà i secció	UH	Internacional	14	50.0
Notícies sense dones per mitjà i secció	UH	Miscel·lània	13	81.25
Notícies sense dones per mitjà i secció	UH	Nacional	134	52.54901960784314
Notícies sense dones per mitjà i secció	UH	Opinió	14	51.85185185185185
Notícies sense dones per mitjà i secció	UH	Palma	6	54.54545454545454
Notícies sense dones per mitjà i secció	UH	Part forana	11	57.89473684210527
Notícies sense dones per mitjà i secció	UH	Prensa rosa	4	40.0
Notícies sense dones per mitjà i secció	UH	Publireportatges	3	75.0
Notícies sense dones per mitjà i secció	UH	Societat	10	43.47826086956522
Notícies sense dones per mitjà i secció	UH	Successos	34	82.92682926829268
Notícies sense dones per mitjà i secció	UH	Tecnologia	11	91.66666666666666
Notícies sense dones per mitjà i secció	UH	Televisió / Cinema	1	25.0
Notícies amb dones per mitjà i secció	DM	Balears	38	46.34146341463415
Notícies amb dones per mitjà i secció	DM	Cultura	24	77.41935483870968
Notícies amb dones per mitjà i secció	DM	Economia	21	39.62264150943396
Notícies amb dones per mitjà i secció	DM	Esports	22	37.93103448275862
Notícies amb dones per mitjà i secció	DM	Fotogalleries / Vídeos	13	22.807017543859647
Notícies amb dones per mitjà i secció	DM	Internacional	22	50.0
Notícies amb dones per mitjà i secció	DM	Miscel·lània	8	30.76923076923077
Notícies amb dones per mitjà i secció	DM	Motor	2	14.285714285714285
Notícies amb dones per mitjà i secció	DM	Nacional	64	71.11111111111111
Notícies amb dones per mitjà i secció	DM	Opinió	15	53.57142857142857
Notícies amb dones per mitjà i secció	DM	Palma	5	33.33333333333333
Notícies amb dones per mitjà i secció	DM	Part forana	4	22.22222222222222
Notícies amb dones per mitjà i secció	DM	Prensa rosa	7	100.0

Notícies amb dones per mitjà i secció	DM	Publireportatges	5	31.25
Notícies amb dones per mitjà i secció	DM	Reportatges	4	25.0
Notícies amb dones per mitjà i secció	DM	Societat	32	62.745098039215684
Notícies amb dones per mitjà i secció	DM	Successos	16	26.666666666666668
Notícies amb dones per mitjà i secció	DM	Tecnologia	1	12.5
Notícies amb dones per mitjà i secció	DM	Televisió / Cinema	35	83.33333333333334
Notícies amb dones per mitjà i secció	UH	Balears	44	48.888888888888886
Notícies amb dones per mitjà i secció	UH	Cultura	10	66.666666666666666
Notícies amb dones per mitjà i secció	UH	Esports	11	27.500000000000004
Notícies amb dones per mitjà i secció	UH	Internacional	14	50.0
Notícies amb dones per mitjà i secció	UH	Miscel·lània	3	18.75
Notícies amb dones per mitjà i secció	UH	Nacional	121	47.45098039215686
Notícies amb dones per mitjà i secció	UH	Opinió	13	48.148148148148145
Notícies amb dones per mitjà i secció	UH	Palma	5	45.45454545454545
Notícies amb dones per mitjà i secció	UH	Part forana	8	42.10526315789473
Notícies amb dones per mitjà i secció	UH	Premsa rosa	6	60.0
Notícies amb dones per mitjà i secció	UH	Publireportatges	1	25.0
Notícies amb dones per mitjà i secció	UH	Societat	13	56.52173913043478
Notícies amb dones per mitjà i secció	UH	Successos	7	17.073170731707318
Notícies amb dones per mitjà i secció	UH	Tecnologia	1	8.333333333333332
Notícies amb dones per mitjà i secció	UH	Televisió / Cinema	3	75.0

Annex 5. Estadístiques per ocupació

descripció	mitjà	secció	categoria_ocupació	desc_categoria_ocupació	valor	percentatge
Dona per ocupació	Total	Total	0	No esmentada	681	53.0373831775701
Dona per ocupació	Total	Total	1	Família reial	4	0.3115264797507788
Dona per ocupació	Total	Total	2	Classe política	170	13.239875389408098
Dona per ocupació	Total	Total	3	Funcionariat	45	3.5046728971962615
Dona per ocupació	Total	Total	4	Polícia, bombers i forces armades	2	0.1557632398753894
Dona per ocupació	Total	Total	5	Professorat	17	1.32398753894081
Dona per ocupació	Total	Total	6	Medicina o odontologia	14	1.0903426791277258
Dona per ocupació	Total	Total	7	Altres serveis socio-sanitaris	2	0.1557632398753894
Dona per ocupació	Total	Total	8	Ciència o enginyeria	13	1.0124610591900312
Dona per ocupació	Total	Total	9	Mitjans de comunicació	53	4.127725856697819
Dona per ocupació	Total	Total	10	Sector jurídic	20	1.557632398753894
Dona per ocupació	Total	Total	11	Sector empresarial	12	0.9345794392523363
Dona per ocupació	Total	Total	12	Treballadors de serveis	4	0.3115264797507788
Dona per ocupació	Total	Total	13	Comerç i artesanía	10	0.778816199376947
Dona per ocupació	Total	Total	16	ONG, sindicats i associacions	7	0.5451713395638629
Dona per ocupació	Total	Total	18	Celebritats i artistes	165	12.850467289719624
Dona per ocupació	Total	Total	19	Esports	54	4.205607476635514
Dona per ocupació	Total	Total	20	Estudiantat	4	0.3115264797507788
Dona per ocupació	Total	Total	23	Resident o habitant	2	0.1557632398753894
Dona per ocupació	Total	Total	24	Jubilades	1	0.0778816199376947
Dona per ocupació	Total	Total	25	Criminals o sospitoses	3	0.23364485981308408
Dona per ocupació	Total	Total	27	Altres	1	0.0778816199376947
Dona per mitjà i ocupació	DM	Total	0	No esmentada	413	51.56054931335831
Dona per mitjà i ocupació	DM	Total	1	Família reial	4	0.4993757802746567
Dona per mitjà i ocupació	DM	Total	2	Classe política	97	12.109862671660425
Dona per mitjà i ocupació	DM	Total	3	Funcionariat	21	2.6217228464419478
Dona per mitjà i ocupació	DM	Total	4	Polícia, bombers i forces armades	1	0.12484394506866417

Dona per mitjà i ocupació	DM	Total	5	Professorat	12	1.4981273408239701
Dona per mitjà i ocupació	DM	Total	6	Medicina o odontologia	11	1.373283395755306
Dona per mitjà i ocupació	DM	Total	7	Altres serveis socio-sanitaris	1	0.12484394506866417
Dona per mitjà i ocupació	DM	Total	8	Ciència o enginyeria	6	0.7490636704119851
Dona per mitjà i ocupació	DM	Total	9	Mitjans de comunicació	35	4.369538077403246
Dona per mitjà i ocupació	DM	Total	10	Sector jurídic	15	1.8726591760299627
Dona per mitjà i ocupació	DM	Total	11	Sector empresarial	3	0.37453183520599254
Dona per mitjà i ocupació	DM	Total	12	Treballadors de serveis	2	0.24968789013732834
Dona per mitjà i ocupació	DM	Total	13	Comerç i artesanía	6	0.7490636704119851
Dona per mitjà i ocupació	DM	Total	16	ONG, sindicats i associacions	5	0.6242197253433208
Dona per mitjà i ocupació	DM	Total	18	Celebritats i artistes	117	14.606741573033707
Dona per mitjà i ocupació	DM	Total	19	Esports	45	5.617977528089887
Dona per mitjà i ocupació	DM	Total	20	Estudiantat	4	0.4993757802746567
Dona per mitjà i ocupació	DM	Total	23	Resident o habitant	1	0.12484394506866417
Dona per mitjà i ocupació	DM	Total	25	Criminals o sospitosos	2	0.24968789013732834
Dona per mitjà i ocupació	UH	Total	0	No esmentada	268	55.48654244306418
Dona per mitjà i ocupació	UH	Total	2	Classe política	73	15.113871635610765
Dona per mitjà i ocupació	UH	Total	3	Funcionariat	24	4.968944099378882
Dona per mitjà i ocupació	UH	Total	4	Polícia, bombers i forces armades	1	0.2070393374741201
Dona per mitjà i ocupació	UH	Total	5	Professorat	5	1.0351966873706004
Dona per mitjà i ocupació	UH	Total	6	Medicina o odontologia	3	0.6211180124223602
Dona per mitjà i ocupació	UH	Total	7	Altres serveis socio-sanitaris	1	0.2070393374741201
Dona per mitjà i ocupació	UH	Total	8	Ciència o enginyeria	7	1.4492753623188406
Dona per mitjà i ocupació	UH	Total	9	Mitjans de comunicació	18	3.7267080745341614
Dona per mitjà i ocupació	UH	Total	10	Sector jurídic	5	1.0351966873706004
Dona per mitjà i ocupació	UH	Total	11	Sector empresarial	9	1.8633540372670807
Dona per mitjà i ocupació	UH	Total	12	Treballadors de serveis	2	0.4140786749482402
Dona per mitjà i ocupació	UH	Total	13	Comerç i artesanía	4	0.8281573498964804
Dona per mitjà i ocupació	UH	Total	16	ONG, sindicats i associacions	2	0.4140786749482402
Dona per mitjà i ocupació	UH	Total	18	Celebritats i artistes	48	9.937888198757763
Dona per mitjà i ocupació	UH	Total	19	Esports	9	1.8633540372670807

Dona per mitjà i ocupació	UH	Total	23	Resident o habitant	1	0.2070393374741201
Dona per mitjà i ocupació	UH	Total	24	Jubilades	1	0.2070393374741201
Dona per mitjà i ocupació	UH	Total	25	Criminals o sospitoses	1	0.2070393374741201
Dona per mitjà i ocupació	UH	Total	27	Altres	1	0.2070393374741201
Dona per secció i ocupació	Total	Balears	0	No esmentada	91	63.63636363636363
Dona per secció i ocupació	Total	Balears	2	Classe política	25	17.482517482517483
Dona per secció i ocupació	Total	Balears	3	Funcionariat	6	4.195804195804196
Dona per secció i ocupació	Total	Balears	5	Professorat	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	6	Medicina o odontologia	4	2.797202797202797
Dona per secció i ocupació	Total	Balears	7	Altres serveis socio-sanitaris	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	8	Ciència o enginyeria	3	2.097902097902098
Dona per secció i ocupació	Total	Balears	9	Mitjans de comunicació	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	11	Sector empresarial	2	1.3986013986013985
Dona per secció i ocupació	Total	Balears	13	Comerç i artesanía	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	16	ONG, sindicats i associacions	3	2.097902097902098
Dona per secció i ocupació	Total	Balears	18	Celebritats i artistes	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	19	Esports	2	1.3986013986013985
Dona per secció i ocupació	Total	Balears	24	Jubilades	1	0.6993006993006993
Dona per secció i ocupació	Total	Balears	27	Altres	1	0.6993006993006993
Dona per secció i ocupació	Total	Cultura	0	No esmentada	35	24.137931034482758
Dona per secció i ocupació	Total	Cultura	1	Família reial	1	0.6896551724137931
Dona per secció i ocupació	Total	Cultura	2	Classe política	2	1.3793103448275863
Dona per secció i ocupació	Total	Cultura	4	Polícia, bombers i forces armades	1	0.6896551724137931
Dona per secció i ocupació	Total	Cultura	5	Professorat	4	2.7586206896551726
Dona per secció i ocupació	Total	Cultura	9	Mitjans de comunicació	21	14.482758620689657
Dona per secció i ocupació	Total	Cultura	11	Sector empresarial	1	0.6896551724137931
Dona per secció i ocupació	Total	Cultura	12	Treballadors de serveis	1	0.6896551724137931
Dona per secció i ocupació	Total	Cultura	18	Celebritats i artistes	79	54.48275862068965
Dona per secció i ocupació	Total	Economia	0	No esmentada	23	74.19354838709677
Dona per secció i ocupació	Total	Economia	2	Classe política	5	16.129032258064516
Dona per secció i ocupació	Total	Economia	3	Funcionariat	2	6.451612903225806

Dona per secció i ocupació	Total	Economia	5	Professorat	1	3.225806451612903
Dona per secció i ocupació	Total	Esports	0	No esmentada	52	56.52173913043478
Dona per secció i ocupació	Total	Esports	2	Classe política	5	5.434782608695652
Dona per secció i ocupació	Total	Esports	3	Funcionariat	1	1.0869565217391304
Dona per secció i ocupació	Total	Esports	5	Professorat	2	2.1739130434782608
Dona per secció i ocupació	Total	Esports	19	Esports	32	34.78260869565217
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	0	No esmentada	6	24.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	1	Família reial	1	4.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	2	Classe política	6	24.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	3	Funcionariat	1	4.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	5	Professorat	2	8.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	6	Medicina o odontologia	1	4.0
Dona per secció i ocupació	Total	Fotogalleries / Vídeos	19	Esports	8	32.0
Dona per secció i ocupació	Total	Internacional	0	No esmentada	25	45.45454545454545
Dona per secció i ocupació	Total	Internacional	2	Classe política	18	32.72727272727273
Dona per secció i ocupació	Total	Internacional	3	Funcionariat	2	3.6363636363636362
Dona per secció i ocupació	Total	Internacional	5	Professorat	1	1.8181818181818181
Dona per secció i ocupació	Total	Internacional	7	Altres serveis socio-sanitaris	1	1.8181818181818181
Dona per secció i ocupació	Total	Internacional	8	Ciència o enginyeria	2	3.6363636363636362
Dona per secció i ocupació	Total	Internacional	10	Sector jurídic	2	3.6363636363636362
Dona per secció i ocupació	Total	Internacional	11	Sector empresarial	1	1.8181818181818181
Dona per secció i ocupació	Total	Internacional	16	ONG, sindicats i associacions	1	1.8181818181818181
Dona per secció i ocupació	Total	Internacional	18	Celebritats i artistes	1	1.8181818181818181
Dona per secció i ocupació	Total	Internacional	25	Criminals o sospitoses	1	1.8181818181818181
Dona per secció i ocupació	Total	Miscel·lània	0	No esmentada	6	33.33333333333333
Dona per secció i ocupació	Total	Miscel·lània	2	Classe política	2	11.111111111111111
Dona per secció i ocupació	Total	Miscel·lània	5	Professorat	1	5.555555555555555
Dona per secció i ocupació	Total	Miscel·lània	6	Medicina o odontologia	2	11.111111111111111
Dona per secció i ocupació	Total	Miscel·lània	8	Ciència o enginyeria	2	11.111111111111111
Dona per secció i ocupació	Total	Miscel·lània	13	Comerç i artesanía	5	27.77777777777778
Dona per secció i ocupació	Total	Motor	0	No esmentada	3	75.0

Dona per secció i ocupació	Total	Motor	5	Professorat	1	25.0
Dona per secció i ocupació	Total	Nacional	0	No esmentada	202	58.04597701149425
Dona per secció i ocupació	Total	Nacional	1	Família reial	2	0.5747126436781609
Dona per secció i ocupació	Total	Nacional	2	Classe política	81	23.275862068965516
Dona per secció i ocupació	Total	Nacional	3	Funcionariat	27	7.758620689655173
Dona per secció i ocupació	Total	Nacional	4	Polícia, bombers i forces armades	1	0.28735632183908044
Dona per secció i ocupació	Total	Nacional	5	Professorat	1	0.28735632183908044
Dona per secció i ocupació	Total	Nacional	6	Medicina o odontologia	3	0.8620689655172413
Dona per secció i ocupació	Total	Nacional	8	Ciència o enginyeria	3	0.8620689655172413
Dona per secció i ocupació	Total	Nacional	9	Mitjans de comunicació	6	1.7241379310344827
Dona per secció i ocupació	Total	Nacional	10	Sector jurídic	5	1.4367816091954022
Dona per secció i ocupació	Total	Nacional	11	Sector empresarial	4	1.1494252873563218
Dona per secció i ocupació	Total	Nacional	12	Treballadors de serveis	2	0.5747126436781609
Dona per secció i ocupació	Total	Nacional	18	Celebritats i artistes	10	2.8735632183908044
Dona per secció i ocupació	Total	Nacional	19	Esports	1	0.28735632183908044
Dona per secció i ocupació	Total	Opinió	0	No esmentada	50	84.7457627118644
Dona per secció i ocupació	Total	Opinió	2	Classe política	5	8.47457627118644
Dona per secció i ocupació	Total	Opinió	9	Mitjans de comunicació	2	3.389830508474576
Dona per secció i ocupació	Total	Opinió	18	Celebritats i artistes	1	1.694915254237288
Dona per secció i ocupació	Total	Opinió	19	Esports	1	1.694915254237288
Dona per secció i ocupació	Total	Palma	0	No esmentada	9	56.25
Dona per secció i ocupació	Total	Palma	2	Classe política	3	18.75
Dona per secció i ocupació	Total	Palma	3	Funcionariat	1	6.25
Dona per secció i ocupació	Total	Palma	5	Professorat	1	6.25
Dona per secció i ocupació	Total	Palma	11	Sector empresarial	1	6.25
Dona per secció i ocupació	Total	Palma	23	Resident o habitant	1	6.25
Dona per secció i ocupació	Total	Part forana	0	No esmentada	9	60.0
Dona per secció i ocupació	Total	Part forana	2	Classe política	2	13.333333333333334
Dona per secció i ocupació	Total	Part forana	8	Ciència o enginyeria	2	13.333333333333334
Dona per secció i ocupació	Total	Part forana	16	ONG, sindicats i associacions	1	6.666666666666667
Dona per secció i ocupació	Total	Part forana	19	Esports	1	6.666666666666667

Dona per secció i ocupació	Total	Prensa rosa	0	No esmentada	22	56.41025641025641
Dona per secció i ocupació	Total	Prensa rosa	9	Mitjans de comunicació	2	5.128205128205128
Dona per secció i ocupació	Total	Prensa rosa	18	Celebritats i artistes	14	35.8974358974359
Dona per secció i ocupació	Total	Prensa rosa	19	Esports	1	2.564102564102564
Dona per secció i ocupació	Total	Publireportatges	0	No esmentada	8	80.0
Dona per secció i ocupació	Total	Publireportatges	11	Sector empresarial	1	10.0
Dona per secció i ocupació	Total	Publireportatges	18	Celebritats i artistes	1	10.0
Dona per secció i ocupació	Total	Reportatges	0	No esmentada	11	100.0
Dona per secció i ocupació	Total	Societat	0	No esmentada	47	50.0
Dona per secció i ocupació	Total	Societat	2	Classe política	11	11.702127659574469
Dona per secció i ocupació	Total	Societat	3	Funcionariat	3	3.1914893617021276
Dona per secció i ocupació	Total	Societat	5	Professorat	1	1.0638297872340425
Dona per secció i ocupació	Total	Societat	6	Medicina o odontologia	3	3.1914893617021276
Dona per secció i ocupació	Total	Societat	9	Mitjans de comunicació	2	2.127659574468085
Dona per secció i ocupació	Total	Societat	10	Sector jurídic	4	4.25531914893617
Dona per secció i ocupació	Total	Societat	11	Sector empresarial	1	1.0638297872340425
Dona per secció i ocupació	Total	Societat	13	Comerç i artesanía	3	3.1914893617021276
Dona per secció i ocupació	Total	Societat	16	ONG, sindicats i associacions	1	1.0638297872340425
Dona per secció i ocupació	Total	Societat	18	Celebritats i artistes	10	10.638297872340425
Dona per secció i ocupació	Total	Societat	19	Esports	5	5.319148936170213
Dona per secció i ocupació	Total	Societat	20	Estudiantat	2	2.127659574468085
Dona per secció i ocupació	Total	Societat	23	Resident o habitant	1	1.0638297872340425
Dona per secció i ocupació	Total	Successos	0	No esmentada	24	55.81395348837209
Dona per secció i ocupació	Total	Successos	2	Classe política	4	9.30232558139535
Dona per secció i ocupació	Total	Successos	3	Funcionariat	2	4.651162790697675
Dona per secció i ocupació	Total	Successos	5	Professorat	1	2.3255813953488373
Dona per secció i ocupació	Total	Successos	6	Medicina o odontologia	1	2.3255813953488373
Dona per secció i ocupació	Total	Successos	10	Sector jurídic	6	13.953488372093023
Dona per secció i ocupació	Total	Successos	12	Treballadors de serveis	1	2.3255813953488373
Dona per secció i ocupació	Total	Successos	16	ONG, sindicats i associacions	1	2.3255813953488373
Dona per secció i ocupació	Total	Successos	20	Estudiantat	1	2.3255813953488373

Dona per secció i ocupació	Total	Successos	25	Criminals o sospitoses	2	4.651162790697675
Dona per secció i ocupació	Total	Tecnologia	0	No esmentada	2	66.66666666666666
Dona per secció i ocupació	Total	Tecnologia	8	Ciència o enginyeria	1	33.33333333333333
Dona per secció i ocupació	Total	Televisió / Cinema	0	No esmentada	56	42.10526315789473
Dona per secció i ocupació	Total	Televisió / Cinema	2	Classe política	1	0.7518796992481203
Dona per secció i ocupació	Total	Televisió / Cinema	9	Mitjans de comunicació	19	14.285714285714285
Dona per secció i ocupació	Total	Televisió / Cinema	10	Sector jurídic	3	2.2556390977443606
Dona per secció i ocupació	Total	Televisió / Cinema	11	Sector empresarial	1	0.7518796992481203
Dona per secció i ocupació	Total	Televisió / Cinema	13	Comerç i artesanía	1	0.7518796992481203
Dona per secció i ocupació	Total	Televisió / Cinema	18	Celebritats i artistes	48	36.09022556390977
Dona per secció i ocupació	Total	Televisió / Cinema	19	Esports	3	2.2556390977443606
Dona per secció i ocupació	Total	Televisió / Cinema	20	Estudiantat	1	0.7518796992481203
Home per ocupació	Total	Total	0	No esmentada	1417	48.32878581173261
Home per ocupació	Total	Total	1	Família reial	14	0.47748976807639837
Home per ocupació	Total	Total	2	Classe política	396	13.506139154160982
Home per ocupació	Total	Total	3	Funcionariat	66	2.251023192360164
Home per ocupació	Total	Total	4	Polícia, bombers i forces armades	34	1.1596180081855387
Home per ocupació	Total	Total	5	Professorat	33	1.125511596180082
Home per ocupació	Total	Total	6	Medicina o odontologia	25	0.8526603001364257
Home per ocupació	Total	Total	7	Altres serveis socio-sanitaris	2	0.06821282401091405
Home per ocupació	Total	Total	8	Ciència o enginyeria	42	1.4324693042291952
Home per ocupació	Total	Total	9	Mitjans de comunicació	103	3.5129604365620737
Home per ocupació	Total	Total	10	Sector jurídic	57	1.9440654843110503
Home per ocupació	Total	Total	11	Sector empresarial	68	2.3192360163710775
Home per ocupació	Total	Total	12	Treballadors de serveis	9	0.30695770804911326
Home per ocupació	Total	Total	13	Comerç i artesanía	41	1.3983628922237379
Home per ocupació	Total	Total	14	Sector primari	3	0.10231923601637108
Home per ocupació	Total	Total	15	Clergat	3	0.10231923601637108
Home per ocupació	Total	Total	16	ONG, sindicats i associacions	16	0.5457025920873124
Home per ocupació	Total	Total	18	Celebritats i artistes	258	8.799454297407912
Home per ocupació	Total	Total	19	Esports	340	11.596180081855389

Home per ocupació	Total	Total	25	Criminals o sospitoses	4	0.1364256480218281
Home per ocupació	Total	Total	27	Altres	1	0.034106412005457026
Home per mitjà i ocupació	DM	Total	0	No esmentada	843	48.309455587392556
Home per mitjà i ocupació	DM	Total	1	Família reial	14	0.8022922636103151
Home per mitjà i ocupació	DM	Total	2	Classe política	224	12.836676217765042
Home per mitjà i ocupació	DM	Total	3	Funcionariat	35	2.005730659025788
Home per mitjà i ocupació	DM	Total	4	Policia, bombers i forces armades	18	1.0315186246418337
Home per mitjà i ocupació	DM	Total	5	Professorat	20	1.146131805157593
Home per mitjà i ocupació	DM	Total	6	Medicina o odontologia	13	0.7449856733524355
Home per mitjà i ocupació	DM	Total	8	Ciència o enginyeria	27	1.5472779369627507
Home per mitjà i ocupació	DM	Total	9	Mitjans de comunicació	77	4.412607449856734
Home per mitjà i ocupació	DM	Total	10	Sector jurídic	39	2.2349570200573066
Home per mitjà i ocupació	DM	Total	11	Sector empresarial	37	2.1203438395415475
Home per mitjà i ocupació	DM	Total	12	Treballadors de serveis	3	0.17191977077363896
Home per mitjà i ocupació	DM	Total	13	Comerç i artesanian	27	1.5472779369627507
Home per mitjà i ocupació	DM	Total	14	Sector primari	2	0.1146131805157593
Home per mitjà i ocupació	DM	Total	15	Clergat	1	0.05730659025787965
Home per mitjà i ocupació	DM	Total	16	ONG, sindicats i associacions	11	0.6303724928366762
Home per mitjà i ocupació	DM	Total	18	Celebritats i artistes	193	11.060171919770774
Home per mitjà i ocupació	DM	Total	19	Esports	159	9.111747851002866
Home per mitjà i ocupació	DM	Total	25	Criminals o sospitoses	2	0.1146131805157593
Home per mitjà i ocupació	UH	Total	0	No esmentada	574	48.35720303285594
Home per mitjà i ocupació	UH	Total	2	Classe política	172	14.490311710193765
Home per mitjà i ocupació	UH	Total	3	Funcionariat	31	2.611625947767481
Home per mitjà i ocupació	UH	Total	4	Policia, bombers i forces armades	16	1.3479359730412805
Home per mitjà i ocupació	UH	Total	5	Professorat	13	1.0951979780960404
Home per mitjà i ocupació	UH	Total	6	Medicina o odontologia	12	1.0109519797809603
Home per mitjà i ocupació	UH	Total	7	Altres serveis socio-sanitaris	2	0.16849199663016007
Home per mitjà i ocupació	UH	Total	8	Ciència o enginyeria	15	1.2636899747262005
Home per mitjà i ocupació	UH	Total	9	Mitjans de comunicació	26	2.1903959561920807
Home per mitjà i ocupació	UH	Total	10	Sector jurídic	18	1.5164279696714407

Home per mitjà i ocupació	UH	Total	11	Sector empresarial	31	2.611625947767481
Home per mitjà i ocupació	UH	Total	12	Treballadors de serveis	6	0.5054759898904801
Home per mitjà i ocupació	UH	Total	13	Comerç i artesanía	14	1.1794439764111204
Home per mitjà i ocupació	UH	Total	14	Sector primari	1	0.08424599831508003
Home per mitjà i ocupació	UH	Total	15	Clergat	2	0.16849199663016007
Home per mitjà i ocupació	UH	Total	16	ONG, sindicats i associacions	5	0.42122999157540014
Home per mitjà i ocupació	UH	Total	18	Celebritats i artistes	65	5.475989890480202
Home per mitjà i ocupació	UH	Total	19	Esports	181	15.248525695029485
Home per mitjà i ocupació	UH	Total	25	Criminals o sospitoses	2	0.16849199663016007
Home per mitjà i ocupació	UH	Total	27	Altres	1	0.08424599831508003
Home per secció i ocupació	Total	Balears	0	No esmentada	124	59.90338164251207
Home per secció i ocupació	Total	Balears	2	Classe política	32	15.458937198067632
Home per secció i ocupació	Total	Balears	3	Funcionariat	9	4.3478260869565215
Home per secció i ocupació	Total	Balears	4	Polícia, bombers i forces armades	1	0.4830917874396135
Home per secció i ocupació	Total	Balears	5	Professorat	4	1.932367149758454
Home per secció i ocupació	Total	Balears	6	Medicina o odontologia	3	1.4492753623188406
Home per secció i ocupació	Total	Balears	7	Altres serveis socio-sanitaris	1	0.4830917874396135
Home per secció i ocupació	Total	Balears	8	Ciència o enginyeria	9	4.3478260869565215
Home per secció i ocupació	Total	Balears	9	Mitjans de comunicació	3	1.4492753623188406
Home per secció i ocupació	Total	Balears	10	Sector jurídic	5	2.4154589371980677
Home per secció i ocupació	Total	Balears	11	Sector empresarial	10	4.830917874396135
Home per secció i ocupació	Total	Balears	15	Clergat	1	0.4830917874396135
Home per secció i ocupació	Total	Balears	18	Celebritats i artistes	4	1.932367149758454
Home per secció i ocupació	Total	Balears	19	Esports	1	0.4830917874396135
Home per secció i ocupació	Total	Cultura	0	No esmentada	66	24.087591240875913
Home per secció i ocupació	Total	Cultura	1	Família reial	2	0.7299270072992701
Home per secció i ocupació	Total	Cultura	2	Classe política	2	0.7299270072992701
Home per secció i ocupació	Total	Cultura	3	Funcionariat	3	1.094890510948905
Home per secció i ocupació	Total	Cultura	5	Professorat	5	1.824817518248175
Home per secció i ocupació	Total	Cultura	8	Ciència o enginyeria	2	0.7299270072992701
Home per secció i ocupació	Total	Cultura	9	Mitjans de comunicació	50	18.248175182481752

Home per secció i ocupació	Total	Cultura	11	Sector empresarial	4	1.4598540145985401
Home per secció i ocupació	Total	Cultura	18	Celebritats i artistes	140	51.09489051094891
Home per secció i ocupació	Total	Economia	0	No esmentada	29	51.78571428571429
Home per secció i ocupació	Total	Economia	1	Família reial	1	1.7857142857142856
Home per secció i ocupació	Total	Economia	2	Classe política	10	17.857142857142858
Home per secció i ocupació	Total	Economia	3	Funcionariat	5	8.928571428571429
Home per secció i ocupació	Total	Economia	5	Professorat	1	1.7857142857142856
Home per secció i ocupació	Total	Economia	10	Sector jurídic	1	1.7857142857142856
Home per secció i ocupació	Total	Economia	11	Sector empresarial	2	3.571428571428571
Home per secció i ocupació	Total	Economia	16	ONG, sindicats i associacions	5	8.928571428571429
Home per secció i ocupació	Total	Economia	18	Celebritats i artistes	2	3.571428571428571
Home per secció i ocupació	Total	Esports	0	No esmentada	293	43.92803598200899
Home per secció i ocupació	Total	Esports	1	Família reial	3	0.4497751124437781
Home per secció i ocupació	Total	Esports	2	Classe política	16	2.39880059970015
Home per secció i ocupació	Total	Esports	3	Funcionariat	4	0.5997001499250375
Home per secció i ocupació	Total	Esports	4	Polícia, bombers i forces armades	4	0.5997001499250375
Home per secció i ocupació	Total	Esports	5	Professorat	1	0.14992503748125938
Home per secció i ocupació	Total	Esports	6	Medicina o odontologia	4	0.5997001499250375
Home per secció i ocupació	Total	Esports	9	Mitjans de comunicació	2	0.29985007496251875
Home per secció i ocupació	Total	Esports	11	Sector empresarial	14	2.0989505247376314
Home per secció i ocupació	Total	Esports	12	Treballadors de serveis	1	0.14992503748125938
Home per secció i ocupació	Total	Esports	16	ONG, sindicats i associacions	1	0.14992503748125938
Home per secció i ocupació	Total	Esports	18	Celebritats i artistes	11	1.6491754122938531
Home per secció i ocupació	Total	Esports	19	Esports	313	46.92653673163419
Home per secció i ocupació	Total	Fotogalleries / Vídeos	0	No esmentada	21	53.84615384615385
Home per secció i ocupació	Total	Fotogalleries / Vídeos	2	Classe política	7	17.94871794871795
Home per secció i ocupació	Total	Fotogalleries / Vídeos	4	Polícia, bombers i forces armades	1	2.564102564102564
Home per secció i ocupació	Total	Fotogalleries / Vídeos	9	Mitjans de comunicació	1	2.564102564102564
Home per secció i ocupació	Total	Fotogalleries / Vídeos	13	Comerç i artesanía	1	2.564102564102564
Home per secció i ocupació	Total	Fotogalleries / Vídeos	18	Celebritats i artistes	1	2.564102564102564
Home per secció i ocupació	Total	Fotogalleries / Vídeos	19	Esports	7	17.94871794871795

Home per secció i ocupació	Total	Internacional	0	No esmentada	79	40.932642487046635
Home per secció i ocupació	Total	Internacional	1	Família reial	1	0.5181347150259068
Home per secció i ocupació	Total	Internacional	2	Classe política	72	37.30569948186528
Home per secció i ocupació	Total	Internacional	3	Funcionariat	9	4.66321243523316
Home per secció i ocupació	Total	Internacional	4	Polícia, bombers i forces armades	8	4.145077720207254
Home per secció i ocupació	Total	Internacional	5	Professorat	2	1.0362694300518136
Home per secció i ocupació	Total	Internacional	8	Ciència o enginyeria	3	1.5544041450777202
Home per secció i ocupació	Total	Internacional	9	Mitjans de comunicació	3	1.5544041450777202
Home per secció i ocupació	Total	Internacional	10	Sector jurídic	7	3.6269430051813467
Home per secció i ocupació	Total	Internacional	11	Sector empresarial	1	0.5181347150259068
Home per secció i ocupació	Total	Internacional	12	Treballadors de serveis	1	0.5181347150259068
Home per secció i ocupació	Total	Internacional	16	ONG, sindicats i associacions	1	0.5181347150259068
Home per secció i ocupació	Total	Internacional	18	Celebritats i artistes	5	2.5906735751295336
Home per secció i ocupació	Total	Internacional	25	Criminals o sospitoses	1	0.5181347150259068
Home per secció i ocupació	Total	Miscel·lània	0	No esmentada	12	21.052631578947366
Home per secció i ocupació	Total	Miscel·lània	2	Classe política	2	3.508771929824561
Home per secció i ocupació	Total	Miscel·lània	5	Professorat	2	3.508771929824561
Home per secció i ocupació	Total	Miscel·lània	6	Medicina o odontologia	3	5.263157894736842
Home per secció i ocupació	Total	Miscel·lània	8	Ciència o enginyeria	5	8.771929824561402
Home per secció i ocupació	Total	Miscel·lània	13	Comerç i artesanía	29	50.877192982456144
Home per secció i ocupació	Total	Miscel·lània	15	Clergat	1	1.7543859649122806
Home per secció i ocupació	Total	Miscel·lània	18	Celebritats i artistes	3	5.263157894736842
Home per secció i ocupació	Total	Motor	0	No esmentada	9	42.857142857142854
Home per secció i ocupació	Total	Motor	1	Família reial	2	9.523809523809524
Home per secció i ocupació	Total	Motor	8	Ciència o enginyeria	1	4.761904761904762
Home per secció i ocupació	Total	Motor	9	Mitjans de comunicació	1	4.761904761904762
Home per secció i ocupació	Total	Motor	11	Sector empresarial	5	23.809523809523807
Home per secció i ocupació	Total	Motor	18	Celebritats i artistes	3	14.285714285714285
Home per secció i ocupació	Total	Nacional	0	No esmentada	398	53.066666666666666
Home per secció i ocupació	Total	Nacional	1	Família reial	5	0.6666666666666667
Home per secció i ocupació	Total	Nacional	2	Classe política	216	28.799999999999997

Home per secció i ocupació	Total	Nacional	3	Funcionariat	30	4.0
Home per secció i ocupació	Total	Nacional	4	Polícia, bombers i forces armades	9	1.2
Home per secció i ocupació	Total	Nacional	5	Professorat	6	0.8
Home per secció i ocupació	Total	Nacional	6	Medicina o odontologia	8	1.0666666666666667
Home per secció i ocupació	Total	Nacional	7	Altres serveis socio-sanitaris	1	0.1333333333333333
Home per secció i ocupació	Total	Nacional	8	Ciència o enginyeria	5	0.6666666666666667
Home per secció i ocupació	Total	Nacional	9	Mitjans de comunicació	6	0.8
Home per secció i ocupació	Total	Nacional	10	Sector jurídic	25	3.3333333333333335
Home per secció i ocupació	Total	Nacional	11	Sector empresarial	10	1.3333333333333335
Home per secció i ocupació	Total	Nacional	12	Treballadors de serveis	4	0.5333333333333333
Home per secció i ocupació	Total	Nacional	13	Comerç i artesanía	1	0.1333333333333333
Home per secció i ocupació	Total	Nacional	14	Sector primari	1	0.1333333333333333
Home per secció i ocupació	Total	Nacional	16	ONG, sindicats i associacions	7	0.9333333333333335
Home per secció i ocupació	Total	Nacional	18	Celebritats i artistes	12	1.6
Home per secció i ocupació	Total	Nacional	19	Esports	4	0.5333333333333333
Home per secció i ocupació	Total	Nacional	25	Criminals o sospitoses	2	0.2666666666666666
Home per secció i ocupació	Total	Opinió	0	No esmentada	103	87.28813559322035
Home per secció i ocupació	Total	Opinió	2	Classe política	1	0.847457627118644
Home per secció i ocupació	Total	Opinió	5	Professorat	1	0.847457627118644
Home per secció i ocupació	Total	Opinió	8	Ciència o enginyeria	1	0.847457627118644
Home per secció i ocupació	Total	Opinió	9	Mitjans de comunicació	2	1.694915254237288
Home per secció i ocupació	Total	Opinió	13	Comerç i artesanía	2	1.694915254237288
Home per secció i ocupació	Total	Opinió	18	Celebritats i artistes	8	6.779661016949152
Home per secció i ocupació	Total	Palma	0	No esmentada	16	50.0
Home per secció i ocupació	Total	Palma	2	Classe política	9	28125
Home per secció i ocupació	Total	Palma	11	Sector empresarial	4	12.5
Home per secció i ocupació	Total	Palma	12	Treballadors de serveis	1	3125
Home per secció i ocupació	Total	Palma	18	Celebritats i artistes	2	6.25
Home per secció i ocupació	Total	Part forana	0	No esmentada	13	46.42857142857143
Home per secció i ocupació	Total	Part forana	2	Classe política	7	25.0
Home per secció i ocupació	Total	Part forana	3	Funcionariat	3	10.714285714285714

Home per secció i ocupació	Total	Part forana	5	Professorat	1	3.571428571428571
Home per secció i ocupació	Total	Part forana	8	Ciència o enginyeria	1	3.571428571428571
Home per secció i ocupació	Total	Part forana	11	Sector empresarial	1	3.571428571428571
Home per secció i ocupació	Total	Part forana	16	ONG, sindicats i associacions	1	3.571428571428571
Home per secció i ocupació	Total	Part forana	27	Altres	1	3.571428571428571
Home per secció i ocupació	Total	Prensa rosa	0	No esmentada	14	53.84615384615385
Home per secció i ocupació	Total	Prensa rosa	2	Classe política	1	3.8461538461538463
Home per secció i ocupació	Total	Prensa rosa	9	Mitjans de comunicació	4	15.384615384615385
Home per secció i ocupació	Total	Prensa rosa	11	Sector empresarial	1	3.8461538461538463
Home per secció i ocupació	Total	Prensa rosa	18	Celebritats i artistes	4	15.384615384615385
Home per secció i ocupació	Total	Prensa rosa	19	Esports	2	7.6923076923076925
Home per secció i ocupació	Total	Publireportatges	0	No esmentada	6	54.54545454545454
Home per secció i ocupació	Total	Publireportatges	11	Sector empresarial	3	27.27272727272727
Home per secció i ocupació	Total	Publireportatges	13	Comerç i artesanía	2	18.181818181818183
Home per secció i ocupació	Total	Reportatges	0	No esmentada	15	65.21739130434783
Home per secció i ocupació	Total	Reportatges	5	Professorat	2	8.695652173913043
Home per secció i ocupació	Total	Reportatges	6	Medicina o odontologia	2	8.695652173913043
Home per secció i ocupació	Total	Reportatges	8	Ciència o enginyeria	4	17.391304347826086
Home per secció i ocupació	Total	Societat	0	No esmentada	87	54.037267080745345
Home per secció i ocupació	Total	Societat	2	Classe política	15	9.316770186335404
Home per secció i ocupació	Total	Societat	4	Polícia, bombers i forces armades	1	0.6211180124223602
Home per secció i ocupació	Total	Societat	5	Professorat	8	4.968944099378882
Home per secció i ocupació	Total	Societat	6	Medicina o odontologia	3	1.8633540372670807
Home per secció i ocupació	Total	Societat	8	Ciència o enginyeria	8	4.968944099378882
Home per secció i ocupació	Total	Societat	9	Mitjans de comunicació	5	3.1055900621118013
Home per secció i ocupació	Total	Societat	10	Sector jurídic	2	1.2422360248447204
Home per secció i ocupació	Total	Societat	11	Sector empresarial	6	3.7267080745341614
Home per secció i ocupació	Total	Societat	13	Comerç i artesanía	5	3.1055900621118013
Home per secció i ocupació	Total	Societat	14	Sector primari	1	0.6211180124223602
Home per secció i ocupació	Total	Societat	16	ONG, sindicats i associacions	1	0.6211180124223602
Home per secció i ocupació	Total	Societat	18	Celebritats i artistes	11	6.832298136645963

Home per secció i ocupació	Total	Societat	19	Esports	8	4.968944099378882
Home per secció i ocupació	Total	Successos	0	No esmentada	57	53.271028037383175
Home per secció i ocupació	Total	Successos	2	Classe política	3	2.803738317757009
Home per secció i ocupació	Total	Successos	3	Funcionariat	2	1.8691588785046727
Home per secció i ocupació	Total	Successos	4	Polícia, bombers i forces armades	10	9.345794392523365
Home per secció i ocupació	Total	Successos	6	Medicina o odontologia	2	1.8691588785046727
Home per secció i ocupació	Total	Successos	9	Mitjans de comunicació	2	1.8691588785046727
Home per secció i ocupació	Total	Successos	10	Sector jurídic	17	15.887850467289718
Home per secció i ocupació	Total	Successos	11	Sector empresarial	4	3.7383177570093453
Home per secció i ocupació	Total	Successos	12	Treballadors de serveis	2	1.8691588785046727
Home per secció i ocupació	Total	Successos	14	Sector primari	1	0.9345794392523363
Home per secció i ocupació	Total	Successos	15	Clergat	1	0.9345794392523363
Home per secció i ocupació	Total	Successos	18	Celebritats i artistes	2	1.8691588785046727
Home per secció i ocupació	Total	Successos	19	Esports	3	2.803738317757009
Home per secció i ocupació	Total	Successos	25	Criminals o sospitoses	1	0.9345794392523363
Home per secció i ocupació	Total	Tecnologia	0	No esmentada	8	66.66666666666666
Home per secció i ocupació	Total	Tecnologia	2	Classe política	1	8.333333333333332
Home per secció i ocupació	Total	Tecnologia	8	Ciència o enginyeria	1	8.333333333333332
Home per secció i ocupació	Total	Tecnologia	11	Sector empresarial	2	16.666666666666664
Home per secció i ocupació	Total	Televisió / Cinema	0	No esmentada	67	44.666666666666664
Home per secció i ocupació	Total	Televisió / Cinema	2	Classe política	2	1.3333333333333335
Home per secció i ocupació	Total	Televisió / Cinema	3	Funcionariat	1	0.6666666666666667
Home per secció i ocupació	Total	Televisió / Cinema	8	Ciència o enginyeria	2	1.3333333333333335
Home per secció i ocupació	Total	Televisió / Cinema	9	Mitjans de comunicació	24	16.0
Home per secció i ocupació	Total	Televisió / Cinema	11	Sector empresarial	1	0.6666666666666667
Home per secció i ocupació	Total	Televisió / Cinema	13	Comerç i artesanía	1	0.6666666666666667
Home per secció i ocupació	Total	Televisió / Cinema	18	Celebritats i artistes	50	33.33333333333333
Home per secció i ocupació	Total	Televisió / Cinema	19	Esports	2	1.3333333333333335
Indeterminat per ocupació	Total	Total	0	No esmentada	122	74.39024390243902
Indeterminat per ocupació	Total	Total	2	Classe política	1	0.6097560975609756
Indeterminat per ocupació	Total	Total	6	Medicina o odontologia	1	0.6097560975609756

Indeterminat per ocupació	Total	Total	8	Ciència o enginyeria	1	0.6097560975609756
Indeterminat per ocupació	Total	Total	10	Sector jurídic	1	0.6097560975609756
Indeterminat per ocupació	Total	Total	11	Sector empresarial	1	0.6097560975609756
Indeterminat per ocupació	Total	Total	13	Comerç i artesanía	1	0.6097560975609756
Indeterminat per ocupació	Total	Total	18	Celebritats i artistes	6	3.6585365853658534
Indeterminat per ocupació	Total	Total	19	Esports	30	18.29268292682927
Indeterminat per mitjà i ocupació	DM	Total	0	No esmentada	87	82.0754716981132
Indeterminat per mitjà i ocupació	DM	Total	6	Medicina o odontologia	1	0.9433962264150944
Indeterminat per mitjà i ocupació	DM	Total	8	Ciència o enginyeria	1	0.9433962264150944
Indeterminat per mitjà i ocupació	DM	Total	11	Sector empresarial	1	0.9433962264150944
Indeterminat per mitjà i ocupació	DM	Total	18	Celebritats i artistes	2	1.8867924528301887
Indeterminat per mitjà i ocupació	DM	Total	19	Esports	14	13.20754716981132
Indeterminat per mitjà i ocupació	UH	Total	0	No esmentada	35	60.3448275862069
Indeterminat per mitjà i ocupació	UH	Total	2	Classe política	1	1.7241379310344827
Indeterminat per mitjà i ocupació	UH	Total	10	Sector jurídic	1	1.7241379310344827
Indeterminat per mitjà i ocupació	UH	Total	13	Comerç i artesanía	1	1.7241379310344827
Indeterminat per mitjà i ocupació	UH	Total	18	Celebritats i artistes	4	6.896551724137931
Indeterminat per mitjà i ocupació	UH	Total	19	Esports	16	27.586206896551722
Indeterminat per secció i ocupació	Total	Balears	0	No esmentada	3	75.0
Indeterminat per secció i ocupació	Total	Balears	13	Comerç i artesanía	1	25.0
Indeterminat per secció i ocupació	Total	Cultura	0	No esmentada	5	45.45454545454545
Indeterminat per secció i ocupació	Total	Cultura	18	Celebritats i artistes	6	54.54545454545454
Indeterminat per secció i ocupació	Total	Esports	0	No esmentada	48	60.75949367088608
Indeterminat per secció i ocupació	Total	Esports	2	Classe política	1	1.2658227848101267
Indeterminat per secció i ocupació	Total	Esports	19	Esports	30	37.9746835443038
Indeterminat per secció i ocupació	Total	Internacional	0	No esmentada	3	100.0
Indeterminat per secció i ocupació	Total	Miscel·lània	0	No esmentada	2	100.0
Indeterminat per secció i ocupació	Total	Nacional	0	No esmentada	9	90.0
Indeterminat per secció i ocupació	Total	Nacional	10	Sector jurídic	1	10.0
Indeterminat per secció i ocupació	Total	Opinió	0	No esmentada	20	100.0
Indeterminat per secció i ocupació	Total	Palma	11	Sector empresarial	1	100.0

Indeterminat per secció i ocupació	Total	Part forana	0	No esmentada	1	100.0
Indeterminat per secció i ocupació	Total	Reportatges	0	No esmentada	2	50.0
Indeterminat per secció i ocupació	Total	Reportatges	6	Medicina o odontologia	1	25.0
Indeterminat per secció i ocupació	Total	Reportatges	8	Ciència o enginyeria	1	25.0
Indeterminat per secció i ocupació	Total	Societat	0	No esmentada	3	100.0
Indeterminat per secció i ocupació	Total	Successos	0	No esmentada	13	100.0
Indeterminat per secció i ocupació	Total	Tecnologia	0	No esmentada	4	100.0
Indeterminat per secció i ocupació	Total	Televisió / Cinema	0	No esmentada	9	100.0

Annex 6. Estadístiques per rol familiar

descripció	valor	percentatge
Dones amb rol familiar esmentat	44	3.4267912772585665
Dones sense rol familiar esmentat	1240	96.57320872274143
Homes amb rol familiar esmentat	22	0.7503410641200546
Homes sense rol familiar esmentat	2910	99.24965893587995
Indeterminades amb rol familiar esmentat	1	0.6097560975609756
Indeterminades sense rol familiar esmentat	163	99.39024390243902

Annex 7. Estadístiques sobre la funció de les persones a la notícia

descripció	sexe	valor	percentatge
Amb funció esmentada a la notícia	dona	496	35.78643578643579
Amb funció esmentada a la notícia	home	864	62.33766233766234
Amb funció esmentada a la notícia	indeterminat	26	1.875901875901876
Amb funció esmentada a la notícia	total	1386	100.0
Sense funció esmentada a la notícia	dona	788	26.319305277221112
Sense funció esmentada a la notícia	home	2068	69.07147628590515
Sense funció esmentada a la notícia	indeterminat	138	4.609218436873747
Sense funció esmentada a la notícia	total	2994	100.0
Protagonistes	dona	320	35.437430786267996
Protagonistes	home	568	62.901439645625686
Protagonistes	indeterminat	15	1.6611295681063125
Protagonistes	total	903	100.0
Portaveus	dona	119	44.569288389513105
Portaveus	home	147	55.0561797752809
Portaveus	indeterminat	1	0.37453183520599254
Portaveus	total	267	100.0
Expertes	dona	48	26.373626373626376
Expertes	home	131	71.97802197802197
Expertes	indeterminat	3	1.6483516483516485
Expertes	total	182	100.0
Testimonis	dona	6	31.57894736842105
Testimonis	home	13	68.42105263157895
Testimonis	indeterminat	0	0.0
Testimonis	total	19	100.0
Víctimes	dona	53	38.1294964028777
Víctimes	home	74	53.23741007194245
Víctimes	indeterminat	12	8.633093525179856
Víctimes	total	139	100.0

Annex 8. Estadístiques dels elements de titulació

descripció	mitjà	valor	percentatge
Persones en títols i subtítols	Total	446	100.0
Dones en títols i subtítols	Total	161	36.09865470852018
Homes en títols i subtítols	Total	277	62.10762331838565
Indeterminades en títols i subtítols	Total	8	1.7937219730941705
Persones en títols i subtítols	DM	300	67.2645739910314
Persones en títols i subtítols	UH	146	32.73542600896861
Dones en títols i subtítols	DM	110	36.666666666666664
Dones en títols i subtítols	UH	51	34.93150684931507
Homes en títols i subtítols	DM	186	62.0
Homes en títols i subtítols	UH	91	62.328767123287676
Indeterminades en títols i subtítols	DM	4	1.3333333333333335
Indeterminades en títols i subtítols	UH	4	2.73972602739726

Annex 9. Estadístiques d'autoria de les notícies

descripció	mitjà	secció	valor	percentatge
Notícies escrites per periodista o periodistes dones	Total	Total	220	16.781083142639204
Notícies escrites per periodista o periodistes homes	Total	Total	385	29.36689549961861
Notícies escrites per periodista o periodistes indeterminats	Total	Total	696	53.08924485125858
Notícies escrites per periodistes d'ambdós sexes	Total	Total	10	0.7627765064836003
Notícies escrites per periodista o periodistes dones per mitjà	DM	Total	159	22.206703910614525
Notícies escrites per periodista o periodistes homes per mitjà	DM	Total	257	35.89385474860335
Notícies escrites per periodista o periodistes indeterminats per mitjà	DM	Total	290	40.502793296089386
Notícies escrites per periodistes d'ambdós sexes per mitjà	DM	Total	10	1.3966480446927374
Notícies escrites per periodista o periodistes dones per mitjà	UH	Total	61	10.252100840336134
Notícies escrites per periodista o periodistes homes per mitjà	UH	Total	128	21.51260504201681
Notícies escrites per periodista o periodistes indeterminats per mitjà	UH	Total	406	68.23529411764706
Notícies escrites per periodistes d'ambdós sexes per mitjà	UH	Total	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Successos	14	13.861386138613863
Notícies escrites per periodista o periodistes homes per secció	Total	Successos	53	52.475247524752476
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Successos	34	33.663366336633665
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Successos	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Nacional	34	9.855072463768117
Notícies escrites per periodista o periodistes homes per secció	Total	Nacional	28	8.115942028985506
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Nacional	280	81.15942028985508
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Nacional	3	0.8695652173913043
Notícies escrites per periodista o periodistes dones per secció	Total	Fotogalleries / Vídeos	2	3.508771929824561
Notícies escrites per periodista o periodistes homes per secció	Total	Fotogalleries / Vídeos	5	8.771929824561402
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Fotogalleries / Vídeos	50	87.71929824561403
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Fotogalleries / Vídeos	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Balears	42	24.418604651162788
Notícies escrites per periodista o periodistes homes per secció	Total	Balears	71	41.27906976744186
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Balears	59	34.30232558139535

Notícies escrites per periodistes d'ambdós sexes per secció	Total	Balears	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Cultura	12	26.08695652173913
Notícies escrites per periodista o periodistes homes per secció	Total	Cultura	14	30.434782608695656
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Cultura	20	43.47826086956522
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Cultura	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Opinió	7	12.727272727272727
Notícies escrites per periodista o periodistes homes per secció	Total	Opinió	45	81.81818181818183
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Opinió	3	5.454545454545454
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Opinió	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Televisió / Cinema	9	19.565217391304348
Notícies escrites per periodista o periodistes homes per secció	Total	Televisió / Cinema	9	19.565217391304348
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Televisió / Cinema	28	60.86956521739131
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Televisió / Cinema	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Esports	9	9.183673469387756
Notícies escrites per periodista o periodistes homes per secció	Total	Esports	45	45.91836734693878
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Esports	43	43.87755102040816
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Esports	1	1.0204081632653061
Notícies escrites per periodista o periodistes dones per secció	Total	Internacional	10	13.88888888888889
Notícies escrites per periodista o periodistes homes per secció	Total	Internacional	17	23.611111111111111
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Internacional	44	61.111111111111114
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Internacional	1	1.3888888888888888
Notícies escrites per periodista o periodistes dones per secció	Total	Societat	29	39.189189189189186
Notícies escrites per periodista o periodistes homes per secció	Total	Societat	18	24.324324324324326
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Societat	26	35.13513513513514
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Societat	1	1.3513513513513513
Notícies escrites per periodista o periodistes dones per secció	Total	Reportatges	1	6.25
Notícies escrites per periodista o periodistes homes per secció	Total	Reportatges	11	68.75
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Reportatges	4	25.0
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Reportatges	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Miscel·lània	8	19.047619047619047
Notícies escrites per periodista o periodistes homes per secció	Total	Miscel·lània	13	30.952380952380953

Notícies escrites per periodista o periodistes indeterminats per secció	Total	Miscel·lània	19	45.23809523809524
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Miscel·lània	2	4.761904761904762
Notícies escrites per periodista o periodistes dones per secció	Total	Part forana	13	35.13513513513514
Notícies escrites per periodista o periodistes homes per secció	Total	Part forana	17	45.94594594594595
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Part forana	7	18.91891891891892
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Part forana	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes homes per secció	Total	Tecnologia	1	5.0
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Tecnologia	19	95.0
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Economia	12	22.641509433962266
Notícies escrites per periodista o periodistes homes per secció	Total	Economia	20	37.735849056603776
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Economia	19	35.84905660377358
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Economia	2	3.7735849056603774
Notícies escrites per periodista o periodistes dones per secció	Total	Prensa rosa	3	17.647058823529413
Notícies escrites per periodista o periodistes homes per secció	Total	Prensa rosa	2	11.76470588235294
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Prensa rosa	12	70.58823529411765
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Prensa rosa	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Palma	9	34.61538461538461
Notícies escrites per periodista o periodistes homes per secció	Total	Palma	10	38.46153846153847
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Palma	7	26.923076923076923
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Palma	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes homes per secció	Total	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Publireportatges	20	100.0
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes dones per secció	Total	Motor	6	42.857142857142854
Notícies escrites per periodista o periodistes homes per secció	Total	Motor	6	42.857142857142854
Notícies escrites per periodista o periodistes indeterminats per secció	Total	Motor	2	14.285714285714285
Notícies escrites per periodistes d'ambdós sexes per secció	Total	Motor	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Successos	10	16.666666666666664

Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Successos	32	53.33333333333336
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Successos	18	30.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Successos	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Nacional	32	35.55555555555556
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Nacional	27	30.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Nacional	28	31.11111111111111
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Nacional	3	3.333333333333335
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Fotogaleries / Vídeos	2	3.508771929824561
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Fotogaleries / Vídeos	5	8.771929824561402
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Fotogaleries / Vídeos	50	87.71929824561403
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Fotogaleries / Vídeos	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Balears	18	21.951219512195124
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Balears	37	45.1219512195122
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Balears	27	32.926829268292686
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Balears	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Cultura	9	29.03225806451613
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Cultura	7	22.58064516129032
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Cultura	15	48.38709677419355
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Cultura	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Opinió	6	21.428571428571427
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Opinió	22	78.57142857142857
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Opinió	0	0.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Opinió	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Televisió / Cinema	8	19.047619047619047
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Televisió / Cinema	9	21.428571428571427
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Televisió / Cinema	25	59.523809523809526
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Televisió / Cinema	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Esports	9	15.517241379310345
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Esports	24	41.37931034482759
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Esports	24	41.37931034482759
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Esports	1	1.7241379310344827

Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Internacional	10	22.727272727272727
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Internacional	15	34.090909090909086
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Internacional	18	40.909090909090914
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Internacional	1	2.272727272727273
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Societat	20	39.21568627450981
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Societat	9	17.647058823529413
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Societat	21	41.17647058823529
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Societat	1	1.9607843137254901
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Reportatges	1	6.25
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Reportatges	11	68.75
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Reportatges	4	25.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Reportatges	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Miscel·lània	6	23.076923076923077
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Miscel·lània	10	38.46153846153847
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Miscel·lània	8	30.76923076923077
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Miscel·lània	2	7.6923076923076925
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Part forana	4	22.222222222222222
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Part forana	12	66.66666666666666
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Part forana	2	11.111111111111111
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Part forana	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Tecnologia	1	12.5
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Tecnologia	7	87.5
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Economia	12	22.641509433962266
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Economia	20	37.735849056603776
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Economia	19	35.84905660377358
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Economia	2	3.7735849056603774
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Prensa rosa	3	42.857142857142854
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Prensa rosa	2	28.57142857142857
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Prensa rosa	2	28.57142857142857

Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Prensa rosa	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Palma	3	20.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Palma	8	53.333333333333336
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Palma	4	26.666666666666668
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Palma	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Publireportatges	16	100.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	DM	Motor	6	42.857142857142854
Notícies escrites per periodista o periodistes homes per mitjà i secció	DM	Motor	6	42.857142857142854
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	DM	Motor	2	14.285714285714285
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	DM	Motor	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Balears	24	26.666666666666668
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Balears	34	37.777777777777778
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Balears	32	35.555555555555556
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Balears	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Part forana	9	47.368421052631575
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Part forana	5	26.31578947368421
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Part forana	5	26.31578947368421
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Part forana	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Nacional	2	0.7843137254901961
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Nacional	1	0.39215686274509803
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Nacional	252	98.82352941176471
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Nacional	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Internacional	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Internacional	2	7.142857142857142
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Internacional	26	92.85714285714286
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Internacional	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Cultura	3	20.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Cultura	7	46.666666666666664

Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Cultura	5	33.33333333333333
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Cultura	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Societat	9	39.130434782608695
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Societat	9	39.130434782608695
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Societat	5	21.73913043478261
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Societat	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Palma	6	54.54545454545454
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Palma	2	18.181818181818183
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Palma	3	27.27272727272727
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Palma	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Tecnologia	12	100.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Tecnologia	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Miscel·lània	2	12.5
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Miscel·lània	3	18.75
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Miscel·lània	11	68.75
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Miscel·lània	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Publireportatges	4	100.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Publireportatges	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Esports	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Esports	21	52.5
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Esports	19	47.5
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Esports	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Successos	4	9.75609756097561
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Successos	21	51.21951219512195
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Successos	16	39.02439024390244
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Successos	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Opinió	1	3.7037037037037033

Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Opinió	23	85.18518518518519
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Opinió	3	11.111111111111111
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Opinió	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Prensa rosa	0	0.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Prensa rosa	0	0.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Prensa rosa	10	100.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Prensa rosa	0	0.0
Notícies escrites per periodista o periodistes dones per mitjà i secció	UH	Televisió / Cinema	1	25.0
Notícies escrites per periodista o periodistes homes per mitjà i secció	UH	Televisió / Cinema	0	0.0
Notícies escrites per periodista o periodistes indeterminats per mitjà i secció	UH	Televisió / Cinema	3	75.0
Notícies escrites per periodistes d'ambdós sexes per mitjà i secció	UH	Televisió / Cinema	0	0.0

Annex 10. Estadístiques de biaix de gènere segons l'autoria

descripció	firma	paràmetre	sexe	valor	percentatge
Dones en notícies escrites per periodista o periodistes dones	dones		dona	304	38.97435897435898
Homes en notícies escrites per periodista o periodistes dones	dones		home	465	59.61538461538461
Indeterminades en notícies escrites per periodista o periodistes dones	dones		indeterminat	11	1.4102564102564104
Dones en notícies escrites per periodista o periodistes homes	homes		dona	381	25.115359261700725
Homes en notícies escrites per periodista o periodistes homes	homes		home	1051	69.28147659854977
Indeterminades en notícies escrites per periodista o periodistes homes	homes		indeterminat	85	5.603164139749506
Dones en notícies escrites per periodista o periodistes indeterminats	indeterminats		dona	591	28.815212091662602
Homes en notícies escrites per periodista o periodistes indeterminats	indeterminats		home	1392	67.86933203315456
Indeterminades en notícies escrites per periodista o periodistes indeterminats	indeterminats		indeterminat	68	3.315455875182838
Dones en notícies escrites per periodista o periodistes mixtes	mixtes		dona	8	25.0
Homes en notícies escrites per periodista o periodistes mixtes	mixtes		home	24	75.0
Indeterminades en notícies escrites per periodista o periodistes mixtes	mixtes		indeterminat	0	0.0
Rol familiar en notícies escrites per periodista o periodistes dones	dones	rol_familiar	Total	25	100.0
Rol familiar en notícies escrites per periodista o periodistes homes	homes	rol_familiar	Total	18	100.0
Rol familiar en notícies escrites per periodista o periodistes indeterminats	indeterminats	rol_familiar	Total	24	100.0
Rol familiar en notícies escrites per periodista o periodistes mixtes	mixtes	rol_familiar	Total	0	0.0
Protagonista en notícies escrites per periodista o periodistes dones	dones	protagonista	Total	155	100.0
Protagonista en notícies escrites per periodista o periodistes homes	homes	protagonista	Total	265	100.0
Protagonista en notícies escrites per periodista o periodistes indeterminats	indeterminats	protagonista	Total	478	100.0
Protagonista en notícies escrites per periodista o periodistes mixtes	mixtes	protagonista	Total	5	100.0
Portaveu en notícies escrites per periodista o periodistes dones	dones	portaveu	Total	54	100.0
Portaveu en notícies escrites per periodista o periodistes homes	homes	portaveu	Total	67	100.0
Portaveu en notícies escrites per periodista o periodistes indeterminats	indeterminats	portaveu	Total	143	100.0
Portaveu en notícies escrites per periodista o periodistes mixtes	mixtes	portaveu	Total	3	100.0
Experta en notícies escrites per periodista o periodistes dones	dones	experta	Total	53	100.0
Experta en notícies escrites per periodista o periodistes homes	homes	experta	Total	69	100.0
Experta en notícies escrites per periodista o periodistes indeterminats	indeterminats	experta	Total	56	100.0
Experta en notícies escrites per periodista o periodistes mixtes	mixtes	experta	Total	4	100.0

Testimoni en notícies escrites per periodista o periodistes dones	dones	testimoni	Total	10	100.0
Testimoni en notícies escrites per periodista o periodistes homes	homes	testimoni	Total	5	100.0
Testimoni en notícies escrites per periodista o periodistes indeterminats	indeterminats	testimoni	Total	4	100.0
Testimoni en notícies escrites per periodista o periodistes mixtes	mixtes	testimoni	Total	0	0.0
Víctima en notícies escrites per periodista o periodistes dones	dones	víctima	Total	36	100.0
Víctima en notícies escrites per periodista o periodistes homes	homes	víctima	Total	41	100.0
Víctima en notícies escrites per periodista o periodistes indeterminats	indeterminats	víctima	Total	62	100.0
Víctima en notícies escrites per periodista o periodistes mixtes	mixtes	víctima	Total	0	0.0
Rol familiar atribuït a una dona en notícies escrites per periodista o periodistes dones	dones	rol_familiar	dona	17	68.0
Rol familiar atribuït a un home en notícies escrites per periodista o periodistes dones	dones	rol_familiar	home	8	32.0
Rol familiar atribuït a indeterminades en notícies escrites per periodista o periodistes dones	dones	rol_familiar	indeterminat	0	0.0
Rol familiar atribuït a una dona en notícies escrites per periodista o periodistes homes	homes	rol_familiar	dona	12	66.66666666666666
Rol familiar atribuït a un home en notícies escrites per periodista o periodistes homes	homes	rol_familiar	home	6	33.33333333333333
Rol familiar atribuït a indeterminades en notícies escrites per periodista o periodistes homes	homes	rol_familiar	indeterminat	0	0.0
Rol familiar atribuït a una dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	rol_familiar	dona	15	62.5
Rol familiar atribuït a un home en notícies escrites per periodista o periodistes indeterminats	indeterminats	rol_familiar	home	8	33.33333333333333
Rol familiar atribuït a indeterminades en notícies escrites per periodista o periodistes indeterminats	indeterminats	rol_familiar	indeterminat	1	4.166666666666666
Rol familiar atribuït a una dona en notícies escrites per periodista o periodistes mixtes	mixtes	rol_familiar	dona	0	0.0
Rol familiar atribuït a un home en notícies escrites per periodista o periodistes mixtes	mixtes	rol_familiar	home	0	0.0
Rol familiar atribuït a indeterminades en notícies escrites per periodista o periodistes mixtes	mixtes	rol_familiar	indeterminat	0	0.0
Protagonista dona en notícies escrites per periodista o periodistes dones	dones	protagonista	dona	64	41.29032258064516
Protagonista home en notícies escrites per periodista o periodistes dones	dones	protagonista	home	90	58.06451612903226
Protagonista indeterminat en notícies escrites per periodista o periodistes dones	dones	protagonista	indeterminat	1	0.6451612903225806
Protagonista dona en notícies escrites per periodista o periodistes homes	homes	protagonista	dona	84	31.69811320754717
Protagonista home en notícies escrites per periodista o periodistes homes	homes	protagonista	home	173	65.28301886792453
Protagonista indeterminat en notícies escrites per periodista o periodistes homes	homes	protagonista	indeterminat	8	3.018867924528302

Protagonista dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	protagonista	dona	172	35.98326359832636
Protagonista home en notícies escrites per periodista o periodistes indeterminats	indeterminats	protagonista	home	300	62.76150627615063
Protagonista indeterminat en notícies escrites per periodista o periodistes indeterminats	indeterminats	protagonista	indeterminat	6	1.2552301255230125
Protagonista dona en notícies escrites per periodista o periodistes mixtes	mixtes	protagonista	dona	0	0.0
Protagonista home en notícies escrites per periodista o periodistes mixtes	mixtes	protagonista	home	5	100.0
Protagonista indeterminat en notícies escrites per periodista o periodistes mixtes	mixtes	protagonista	indeterminat	0	0.0
Portaveu dona en notícies escrites per periodista o periodistes dones	dones	portaveu	dona	31	57.407407407407405
Portaveu home en notícies escrites per periodista o periodistes dones	dones	portaveu	home	23	42.592592592592595
Portaveu indeterminat en notícies escrites per periodista o periodistes dones	dones	portaveu	indeterminat	0	0.0
Portaveu dona en notícies escrites per periodista o periodistes homes	homes	portaveu	dona	25	37.3134328358209
Portaveu home en notícies escrites per periodista o periodistes homes	homes	portaveu	home	42	62.68656716417911
Portaveu indeterminat en notícies escrites per periodista o periodistes homes	homes	portaveu	indeterminat	0	0.0
Portaveu dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	portaveu	dona	61	42.65734265734265
Portaveu home en notícies escrites per periodista o periodistes indeterminats	indeterminats	portaveu	home	81	56.64335664335665
Portaveu indeterminat en notícies escrites per periodista o periodistes indeterminats	indeterminats	portaveu	indeterminat	1	0.6993006993006993
Portaveu dona en notícies escrites per periodista o periodistes mixtes	mixtes	portaveu	dona	2	66.66666666666666
Portaveu home en notícies escrites per periodista o periodistes mixtes	mixtes	portaveu	home	1	33.33333333333333
Portaveu indeterminat en notícies escrites per periodista o periodistes mixtes	mixtes	portaveu	indeterminat	0	0.0
Experta dona en notícies escrites per periodista o periodistes dones	dones	experta	dona	18	33.9622641509434
Experta home en notícies escrites per periodista o periodistes dones	dones	experta	home	35	66.0377358490566
Experta indeterminat en notícies escrites per periodista o periodistes dones	dones	experta	indeterminat	0	0.0
Experta dona en notícies escrites per periodista o periodistes homes	homes	experta	dona	16	23.18840579710145
Experta home en notícies escrites per periodista o periodistes homes	homes	experta	home	50	72.46376811594203
Experta indeterminat en notícies escrites per periodista o periodistes homes	homes	experta	indeterminat	3	4.3478260869565215
Experta dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	experta	dona	14	25.0
Experta home en notícies escrites per periodista o periodistes indeterminats	indeterminats	experta	home	42	75.0
Experta indeterminat en notícies escrites per periodista o periodistes indeterminats	indeterminats	experta	indeterminat	0	0.0
Experta dona en notícies escrites per periodista o periodistes mixtes	mixtes	experta	dona	0	0.0
Experta home en notícies escrites per periodista o periodistes mixtes	mixtes	experta	home	4	100.0
Experta indeterminat en notícies escrites per periodista o periodistes mixtes	mixtes	experta	indeterminat	0	0.0
Testimoni dona en notícies escrites per periodista o periodistes dones	dones	testimoni	dona	5	50.0

Testimoni home en notícies escrites per periodista o periodistes dones	dones	testimoni	home	5	50.0
Testimoni indeterminat en notícies escrites per periodista o periodistes dones	dones	testimoni	indeterminat	0	0.0
Testimoni dona en notícies escrites per periodista o periodistes homes	homes	testimoni	dona	0	0.0
Testimoni home en notícies escrites per periodista o periodistes homes	homes	testimoni	home	5	100.0
Testimoni indeterminat en notícies escrites per periodista o periodistes homes	homes	testimoni	indeterminat	0	0.0
Testimoni dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	testimoni	dona	1	25.0
Testimoni home en notícies escrites per periodista o periodistes indeterminats	indeterminats	testimoni	home	3	75.0
Testimoni indeterminat en notícies escrites per periodista o periodistes indeterminats	indeterminats	testimoni	indeterminat	0	0.0
Testimoni dona en notícies escrites per periodista o periodistes mixtes	mixtes	testimoni	dona	0	0.0
Testimoni home en notícies escrites per periodista o periodistes mixtes	mixtes	testimoni	home	0	0.0
Testimoni indeterminat en notícies escrites per periodista o periodistes mixtes	mixtes	testimoni	indeterminat	0	0.0
Víctima dona en notícies escrites per periodista o periodistes dones	dones	víctima	dona	17	47.22222222222222
Víctima home en notícies escrites per periodista o periodistes dones	dones	víctima	home	17	47.22222222222222
Víctima indeterminat en notícies escrites per periodista o periodistes dones	dones	víctima	indeterminat	2	5.555555555555555
Víctima dona en notícies escrites per periodista o periodistes homes	homes	víctima	dona	12	29.268292682926827
Víctima home en notícies escrites per periodista o periodistes homes	homes	víctima	home	23	56.09756097560976
Víctima indeterminat en notícies escrites per periodista o periodistes homes	homes	víctima	indeterminat	6	14.634146341463413
Víctima dona en notícies escrites per periodista o periodistes indeterminats	indeterminats	víctima	dona	24	38.70967741935484
Víctima home en notícies escrites per periodista o periodistes indeterminats	indeterminats	víctima	home	34	54.83870967741935
Víctima indeterminat en notícies escrites per periodista o periodistes indeterminats	indeterminats	víctima	indeterminat	4	6.451612903225806
Víctima dona en notícies escrites per periodista o periodistes mixtes	mixtes	víctima	dona	0	0.0
Víctima home en notícies escrites per periodista o periodistes mixtes	mixtes	víctima	home	0	0.0
Víctima indeterminat en notícies escrites per periodista o periodistes mixtes	mixtes	víctima	indeterminat	0	0.0

Annex 11. Estadístiques relatives a les fotografies

descripció	valor	percentatge
Fotografies amb persones	911	73.40854149879131
Fotografies sense persones	330	26.591458501208702
Fotografies amb dones	459	50.38419319429198
Fotografies sense dones	452	49.61580680570801
Fotografies amb homes	743	81.55872667398462
Fotografies sense homes	168	18.441273326015367
Fotografies amb indeterminades	85	9.330406147091107
Fotografies sense indeterminades	826	90.66959385290889

Annex 12. Estadístiques d'edat

descripció	rang_edat	sexe	categoria_ocupació	desc_ocupació	valor	percentatge
Persones per rang d'edat	Total	Total	Total		4380	100.0
Persones per rang d'edat	Menors d'edat (0-17)	Total	Total		7	0.15981735159817353
Persones per rang d'edat	Adultes (18-64)	Total	Total		62	1.4155251141552512
Persones per rang d'edat	Jubilades (65+)	Total	Total		21	0.4794520547945206
Persones per rang d'edat	Desconeguda	Total	Total		4290	97.94520547945206
Persones per rang d'edat	0-9 anys	Total	Total		2	0.045662100456621
Persones per rang d'edat	10-19 anys	Total	Total		7	0.15981735159817353
Persones per rang d'edat	20-29 anys	Total	Total		18	0.410958904109589
Persones per rang d'edat	30-39 anys	Total	Total		8	0.182648401826484
Persones per rang d'edat	40-49 anys	Total	Total		18	0.410958904109589
Persones per rang d'edat	50-59 anys	Total	Total		12	0.273972602739726
Persones per rang d'edat	60-69 anys	Total	Total		6	0.136986301369863
Persones per rang d'edat	70-79 anys	Total	Total		12	0.273972602739726
Persones per rang d'edat	80-89 anys	Total	Total		5	0.1141552511415525
Persones per rang d'edat	+90 anys	Total	Total		2	0.045662100456621
Persones per rang d'edat i sexe	Menors d'edat (0-17)	dona	Total		6	6.666666666666667
Persones per rang d'edat i sexe	Menors d'edat (0-17)	home	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	Menors d'edat (0-17)	indeterminat	Total		0	0.0
Persones per rang d'edat i sexe	Adultes (18-64)	dona	Total		25	27.77777777777778
Persones per rang d'edat i sexe	Adultes (18-64)	home	Total		31	34.44444444444444
Persones per rang d'edat i sexe	Adultes (18-64)	indeterminat	Total		6	6.666666666666667
Persones per rang d'edat i sexe	Jubilades (65+)	dona	Total		7	7.777777777777778
Persones per rang d'edat i sexe	Jubilades (65+)	home	Total		13	14.444444444444443
Persones per rang d'edat i sexe	Jubilades (65+)	indeterminat	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	0-9 anys	dona	Total		2	2.222222222222223
Persones per rang d'edat i sexe	0-9 anys	home	Total		0	0.0
Persones per rang d'edat i sexe	0-9 anys	indeterminat	Total		0	0.0

Persones per rang d'edat i sexe	10-19 anys	dona	Total		4	4.444444444444445
Persones per rang d'edat i sexe	10-19 anys	home	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	10-19 anys	indeterminat	Total		2	2.222222222222223
Persones per rang d'edat i sexe	20-29 anys	dona	Total		7	7.777777777777778
Persones per rang d'edat i sexe	20-29 anys	home	Total		9	10.0
Persones per rang d'edat i sexe	20-29 anys	indeterminat	Total		2	2.222222222222223
Persones per rang d'edat i sexe	30-39 anys	dona	Total		2	2.222222222222223
Persones per rang d'edat i sexe	30-39 anys	home	Total		6	6.666666666666667
Persones per rang d'edat i sexe	30-39 anys	indeterminat	Total		0	0.0
Persones per rang d'edat i sexe	40-49 anys	dona	Total		9	10.0
Persones per rang d'edat i sexe	40-49 anys	home	Total		8	8.88888888888889
Persones per rang d'edat i sexe	40-49 anys	indeterminat	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	50-59 anys	dona	Total		6	6.666666666666667
Persones per rang d'edat i sexe	50-59 anys	home	Total		5	5.555555555555555
Persones per rang d'edat i sexe	50-59 anys	indeterminat	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	60-69 anys	dona	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	60-69 anys	home	Total		4	4.444444444444445
Persones per rang d'edat i sexe	60-69 anys	indeterminat	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	70-79 anys	dona	Total		5	5.555555555555555
Persones per rang d'edat i sexe	70-79 anys	home	Total		7	7.777777777777778
Persones per rang d'edat i sexe	70-79 anys	indeterminat	Total		0	0.0
Persones per rang d'edat i sexe	80-89 anys	dona	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	80-89 anys	home	Total		4	4.444444444444445
Persones per rang d'edat i sexe	80-89 anys	indeterminat	Total		0	0.0
Persones per rang d'edat i sexe	+90 anys	dona	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	+90 anys	home	Total		1	1.1111111111111112
Persones per rang d'edat i sexe	+90 anys	indeterminat	Total		0	0.0
Persones per rang d'edat i ocupació	Menors d'edat (0-17)	Total		0 No esmentada	3	3.333333333333335
Persones per rang d'edat i ocupació	Menors d'edat (0-17)	Total		19 Esports	1	1.1111111111111112
Persones per rang d'edat i ocupació	Menors d'edat (0-17)	Total		20 Estudiantat	3	3.333333333333335
Persones per rang d'edat i ocupació	Adultes (18-64)	Total		0 No esmentada	28	31.111111111111111

Persones per rang d'edat i ocupació	Adultes (18-64)	Total	2	Classe política	4	4.4444444444444445
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	4	Polícia, bombers i forces armades	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	5	Professorat	2	2.2222222222222223
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	6	Medicina o odontologia	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	9	Mitjans de comunicació	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	10	Sector jurídic	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	11	Sector empresarial	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	12	Treballadors de serveis	2	2.2222222222222223
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	13	Comerç i artesania	3	3.3333333333333335
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	18	Celebritats i artistes	9	10.0
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	19	Esports	6	6.666666666666667
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	20	Estudiantat	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	25	Criminals o sospitoses	1	1.1111111111111112
Persones per rang d'edat i ocupació	Adultes (18-64)	Total	27	Altres	1	1.1111111111111112
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	0	No esmentada	12	13.333333333333334
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	2	Classe política	3	3.3333333333333335
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	8	Ciència o enginyeria	1	1.1111111111111112
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	11	Sector empresarial	1	1.1111111111111112
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	18	Celebritats i artistes	3	3.3333333333333335
Persones per rang d'edat i ocupació	Jubilades (65+)	Total	19	Esports	1	1.1111111111111112

Annex 13. Sèries temporals

descripció	any	sexe	valor	percentatge
Dones per any	2014	dona	253.0	22.792792792792792
Homes per any	2014	home	857.0	77.2072072072072
Dones per any	2021	dona	288.0	27.7992277992278
Homes per any	2021	home	748.0	72.2007722007722
Dones per any	2022	dona	415.0	29.000698812019564
Homes per any	2022	home	1016.0	70.99930118798044
Dones per any	2023	dona	581.0	33.21898227558605
Homes per any	2023	home	1168.0	66.78101772441394
Dones amb rol familiar esmentat per any	2014	dona		7.0
Homes amb rol familiar esmentat per any	2014	home		5.0
Dones amb rol familiar esmentat per any	2021	dona	10.0	0.9652509652509652
Homes amb rol familiar esmentat per any	2021	home	9.0	0.8687258687258688
Dones amb rol familiar esmentat per any	2022	dona	12.0	0.8385744234800839
Homes amb rol familiar esmentat per any	2022	home	7.0	0.4891684136967156
Dones amb rol familiar esmentat per any	2023	dona	22.0	1.257861635220126
Homes amb rol familiar esmentat per any	2023	home	6.0	0.34305317324185247
Dones protagonistes per any	2014	dona		23.0
Homes protagonistes per any	2014	home		77.0
Dones protagonistes per any	2021	dona	84.0	41.17647058823529
Homes protagonistes per any	2021	home	118.0	57.84313725490197
Dones protagonistes per any	2022	dona	106.0	32.21884498480243
Homes protagonistes per any	2022	home	219.0	66.56534954407294
Dones protagonistes per any	2023	dona	130.0	35.13513513513514
Homes protagonistes per any	2023	home	231.0	62.43243243243243
Dones expertes per any	2014	dona		18.0
Homes experts per any	2014	home		82.0
Dones expertes per any	2021	dona	15.0	25.862068965517242
Homes experts per any	2021	home	42.0	72.41379310344827
Dones expertes per any	2022	dona	12.0	16.901408450704224
Homes experts per any	2022	home	58.0	81.69014084507043
Dones expertes per any	2023	dona	21.0	39.62264150943396
Homes experts per any	2023	home	31.0	58.490566037735846
Dones en fotografies per any	2014	dona		24.5
Homes en fotografies per any	2014	home		74.5
Dones en fotografies per any	2021	dona	248.0	29.176470588235293
Homes en fotografies per any	2021	home	602.0	70.82352941176471
Dones en fotografies per any	2022	dona	409.0	31.245225362872418
Homes en fotografies per any	2022	home	900.0	68.75477463712758
Dones en fotografies per any	2023	dona	402.0	31.284046692607003
Homes en fotografies per any	2023	home	883.0	68.715953307393
Dones redactores per any	2014	dona		33.5
Homes redactors per any	2014	home		66.5
Dones redactors per any	2021	dona	61.0	37.19512195121951
Homes redactors per any	2021	home	103.0	62.80487804878049
Dones redactors per any	2022	dona	82.0	34.45378151260504

Homes redactors per any	2022	home	156.0	65.54621848739495
Dones redactors per any	2023	dona	88.0	38.93805309734513
Homes redactors per any	2023	home	138.0	61.06194690265486

Annex 14. Reportatges i infografies elaborats a partir de les dades

Els textos i les visualitzacions de dades han estat extrets de la pàgina web publicada, que es pot consultar a través de l'adreça <https://premsaigenera.wordpress.com/>. Les captures de pantalla de les infografies interactives no inclouen totes les possibles variants de cada gràfic.

Notícia número 1:

Les dades generals

Només 3 de cada 10 persones que surten als diaris digitals de Mallorca són dones

L'anàlisi de les notícies dels dos principals mitjans en línia de l'illa mostra que les dones estan infrarepresentades als mitjans
· *El 54 % dels articles no inclou ni esmenta cap dona*

Els mitjans de comunicació no només són una font de notícies, sinó també un reflex de la dinàmica de la societat. I en aquest ecosistema de representació social, la bretxa de gènere encara és present. L'estudi *Paritat de gènere a la premsa local digital de Mallorca* indica que **només tres de cada deu persones que apareixen en les notícies són dones**.

L'estudi ha detectat 4.380 persones entre el total de textos analitzats. D'aquests individus, 1.284 eren dones (**el 29,3 % del total**) i 2.932 eren homes (el 66,9 %). No s'ha pogut determinar el sexe de les 164 persones restants, que representen el 3,7 % de la mostra.

El treball s'ha dut a terme a partir del buidatge i l'**anàlisi automatitzada mitjançant intel·ligència artificial de 1.311 articles de premsa** dels dos principals periòdics en línia de Mallorca, Diaridemallorca.es (DM) i Ultimahora.es (UH). La mostra han estat totes les publicacions de tres dies complets de cadascun dels mitjans **entre 2021 i 2023**.

De totes les notícies, el 54,4 % no inclou cap dona. La xifra és lleugerament superior a UH (56,3 %) que al DM (52,8 %). Per seccions, Tecnologia, Motor i Successos són els tres àmbits on hi ha una major proporció d'articles sense representació femenina.

A més de posar de manifest la disparitat de gènere, les dades també mostren altres pràctiques del periodisme digital actual. Un exemple és que **un 26,6 % dels articles publicats no esmenten cap persona** i que un 18,7 % de les informacions inclouen únicament un individu,

sigui com a font o com un esment al text. **De mitjana, les notícies analitzades referencien 3,3 persones.**

Les estadístiques calculades a partir de l'estudi reflecteixen un desequilibri sistèmic més profund en com els mitjans representen els gèneres. A més, **les dades permeten plantejar preguntes crítiques sobre la visibilitat i el paper de les dones a les notícies i les narratives socials** més àmplies de les quals formen part. La infrarepresentació als mitjans digitals no només limita el retrat de les dones en l'actualitat: també afecta potencialment la diversitat de perspectives i veus en el discurs públic.

La sobrerepresentació dels homes, un patró que es repeteix

Encara que les veus femenines a les notícies continuen sent minoria, la presència de les dones a les notícies s'ha **incrementat 10 punts des de 2014, passant d'un 23 % aleshores al 33 % actual**. La tendència, tot i ser ascendent, subratlla el canvi lent en les narracions dels mitjans.

L'examen de les funcions discursives que dones i homes tenen als articles periodístics permet observar un biaix de gènere. Els homes tenen el doble de probabilitats de ser els protagonistes dels articles, eclipsant les seves homòlogues femenines. La diferència també es veu en aquells casos en què s'entrevista algú per la seva experiència professional. **Només el 26,4 % de les veus expertes són de dones.**

El fotoperiodisme mostra la mateixa tendència. Gairebé la meitat de les fotografies de les notícies no inclouen dones, la qual cosa disminueix encara més la seva presència a l'ull públic. En canvi, **vuit de cada deu imatges de premsa ensenyen homes**. Els elements de titulació són un altre aspecte crític de la visibilitat de les dones. D'acord amb les dades, sols el 36% dels títols i subtítols inclouen dones.

Les anàlisis també assenyalen un biaix en les pràctiques periodístiques de les redaccions. Tant els redactors homes com les periodistes dones consulten més fonts masculines, però ells ho fan en major proporció que elles. **Els professionals de la comunicació referencien un 69,3 % d'homes**, mentre que aquest percentatge baixa al 59,6 % quan qui escriu la notícia són dones.

Els rols familiars encara s'associen més a les dones

La narrativa dels rols familiars és més present en dones que en homes. Descriure una persona a partir de les seves relacions de parentiu és una pràctica minoritària. Ara bé, quan això passa, **és cinc vegades més probable que una dona sigui definida com a «mare», «filla» o «esposa»** que un home com a «pare», «fill» o «marit».

Pel que fa a la representació per ocupacions, les dones en política, les celebritats i les artistes són les més esmentades a les notícies. Les dades revelen una novetat: **les dones esportistes ocupen la tercera posició** en nombre de mencions.

Cadascun dels aspectes que mostren les dades contribueix a una imatge polièdrica de l'estat de la representació de gènere als mitjans digitals de Mallorca. Les dades obren el camí per a debats sobre com **aconseguir un entorn mediàtic més equilibrat i equitatiu, i proposen que**

la indústria dels mitjans reflexioni per tal d'assolir una representació més equitativa de les dones en les seves informacions.

Notícia número 2:

L'evolució en el temps

La presència de les dones a les notícies s'incrementa 10 punts des de 2014

L'informe 'Un dia a les notícies en clau de gènere', que data de 2014, xifrava en un 23% la presència de dones a les informacions de les Balears

La representació de les dones en els mitjans de comunicació de les Balears està en ascens. Segons l'informe *Un dia a les notícies en clau de gènere*, publicat el 2014, la presència de dones en les notícies era del 23 %. Aquesta xifra va marcar el començament d'un canvi significatiu. Ara, gairebé una dècada després, **la presència del sexe femení en les informacions dels digitals de Mallorca s'ha incrementat fins al 33 %** l'any passat, és a dir, un augment de 10 punts percentuals. Aquest creixement no només destaca la visibilitat més gran de les dones, sinó també una evolució lenta cap a una major paritat de gènere en la cobertura informativa.

L'anàlisi dels diversos rols que les dones han ocupat en les notícies revela canvis importants. **La representació de les dones com a protagonistes en les històries ha experimentat un augment notable.** Passen de representar el 23 % el 2014 al 35 % el 2023. Un altre aspecte destacable és el canvi en la representació de les dones com a expertes. Mentre que en 2014, les dones expertes eren escasses en les notícies (menys de 2 de cada 10), aquesta xifra ha augmentat significativament en els anys recents. **En 2023, gairebé 4 de cada 10 experts citats en les notícies eren dones.**

La **presència de les dones en les fotografies** que acompanyen les notícies també ha experimentat un augment progressiu (del 24,5 % al 31,3 % en nou anys).

Un altre cas el nombre de dones sobre les quals s'esmenta el seu rol familiar ha baixat de forma pronunciada, tot i veure's un repunt l'any passat: del 7 % de 2014 a l'11,26 % del 2023. Tot i això, **cada any es manté la tendència de representar més dones que homes mitjançant relacions de parentiu.**

Un altre aspecte que mostra un lent avenç és la **feminització de les redaccions**. El 2014 el 33,5 % de les peces les firmaven dones, xifra que el 2023 ha passat a ser del 38,9 %. En aquest

període, **Diariodemallorca.es** ha passat a ser dirigit per dones: Maria Ferrer Oliver en va ser la pionera i va deixar el relleu a l'actual directora, Marisa Goñi Sanzberro.

La tendència positiva cap a una major inclusió de les dones en els diferents aspectes de la narrativa mediàtica evidencia un esforç conscient dels mitjans per reflectir una societat més equitativa i diversa. Malgrat aquests avenços, encara queda camí per recórrer per assolir una representació completament equilibrada i justa de totes les veus en els mitjans de comunicació.

Notícia número 3:

Els rols

Els homes són protagonistes dels articles el doble de vegades que les dones

La representació de les dones com a expertes encara és minoritària: només una de cada quatre veus expertes és de sexe femení

Les funcions que dones i homes desenvolupen a les notícies assenyalen desequilibris notables de gènere als mitjans digitals de Mallorca. **Els homes apareixen com a protagonistes en els articles de premsa gairebé el doble de vegades que les dones**, una estadística que destaca la persistència de patrons de gènere tradicionals en el periodisme.

En les informacions en les quals un individu té una funció específica, **els homes són el personatge principal de la notícia en un 62,9 % dels casos**, mentre que les dones ho són només en un 35,4 % d'ocasions. Els rols protagonistes, en particular, mostren una de les divergències més notables de l'estudi.

Tot i això, és encara més colpidora la representació de les dones en rols d'expertes. **Només el 26,4 % de les figures expertes dins de les notícies són dones**, un fet que posa de manifest la infrarepresentació femenina en posicions d'autoritat i coneixement en diversos camps. Aquesta disparitat podria tenir implicacions profundes, no només en la manera com s'informa sobre temes específics, sinó també en la formació de l'opinió pública i en la promoció de la igualtat de gènere en l'àmbit acadèmic i professional.

En contrast, **la representació com a portaveus és més equilibrada**, amb les dones sent presentades en aquesta funció en un **44,6 % dels casos**, en comparació a un 55,1 % dels homes que són esmentats amb aquest rol.

La diferència de gènere a l'hora de presentar una persona com a víctima és menys marcada. **De totes les persones introduïdes com a víctima o supervivent, el 38,1 % són dones**, el 53,2 % són homes i el 8,7 % són persones de les quals la intel·ligència artificial no ha pogut determinar el sexe.

Notícia número 4:

Les imatges

Gairebé la meitat de les fotografies publicades no mostra cap dona

Els homes són presents a més del 80 % de les imatges dels dos principals diaris digitals mallorquins

Gairebé la meitat de les fotografies publicades en notícies no inclouen cap dona. Segons les dades recollides a partir de l'anàlisi automatitzada de representació de gènere en les fotografies, de les 911 imatges que contenien persones, 459 (un 50,4%) mostraven dones. Això implica que **un 49,6 % de les fotografies amb persones no tenien cap dona representada**.

Per contra, els homes tenen una presència dominant en l'àmbit semiòtic del fotoperiodisme. Els registres assenyalen que **els homes surten a un 81,6 % de les fotografies amb persones**. Només un 18,4 % de les imatges no inclouen representació masculina. Aquesta diferència notable posa de manifest una disparitat en la visibilitat de gènere en els mitjans de comunicació digital de Mallorca.

La intel·ligència artificial amb la qual s'ha fet l'estudi no ha pogut determinar el sexe de 85 persones representades en les imatges. Aquesta anàlisi s'ha **basat únicament en un estudi visual i no en les identitats de gènere que puguin manifestar els individus retratats**. En qualsevol cas, les dades apunten a la complexitat en l'anàlisi de gènere i la necessitat d'enfocaments més inclusius en la representació mediàtica.

Per arribar a aquestes conclusions, l'estudi s'ha centrat únicament en la **fotografia principal de cada un dels articles analitzats** i no en totes les imatges que acompanyaven cada notícia. Aquest enfocament subratlla la importància de la imatge destacada en la percepció i interpretació de les notícies i dona una visió més clara de com es representen els gèneres en el primer punt de contacte visual amb l'audiència.

Els resultats de l'anàlisi apunten la **necessitat d'una major inclusió i representació equitativa de totes les identitats de gènere** en els mitjans de comunicació. Així mateix, permet donar pas a una reflexió més profunda sobre les pràctiques editorials en el fotoperiodisme actual de l'illa.

Els redactors homes consulten més fonts masculines que les periodistes

Les notícies d'agència o firmades per la redacció representen més de la meitat dels articles publicats en els digitals mallorquins

Existeix un **biaix en el sexe de les fonts o protagonistes segons el sexe del periodista**. Aquesta afirmació es desprèn de l'anàlisi de les dades de l'autoria de les notícies. Les xifres constaten que els periodistes homes tendeixen a consultar més fonts masculines.

En les notícies escrites per periodistes homes, un 69,3 % de les fonts eren homes, mentre que només un 25,1 % eren dones. Aquest contrast és notable en comparació amb **les notícies escrites per periodistes dones, en què les fonts masculines representen el 59,6 %** i les femenines el 39 %. Els percentatges restants fins a arribar al 100 % es corresponen a persones de les quals no s'ha pogut determinar el sexe mitjançant l'anàlisi textual amb intel·ligència artificial.

Més concretament, **els redactors homes recorren amb més freqüència a veus expertes masculines, escriuen sobre més protagonistes masculins, parlen amb més portaveus homes**, fan servir més homes com a testimonis de fets noticiosos i també presenten més homes com a víctimes. I a més, en tots aquests casos, els redactors homes destaquen les fonts masculines en major proporció que les seves companyes de redacció.

Per exemple, **de cada deu persones expertes amb qui parla un redactor home, set són homes i tres són dones**. La proporció varia lleugerament si es tracta d'una redactora: les expertes s'incrementen a tres de cada deu. L'única funció per la qual les redactores fan servir més fonts femenines és amb les portaveus: **quan la notícia la firma una dona, el 57,4 % de les fonts identificades com a portaveus són dones**. Curiosament, les redactores solen esmentar més el rol familiar en dones que en homes, i també ho fan en major proporció que els redactors homes.

En els teletips d'agència i les notícies firmades per redacció, la tendència continua sent similar, amb un 67,9 % de fonts masculines enfront d'un 28,8 % de veus femenines. En el cas que un article estigui signat per un equip mixt de redactora i redactor, els individus homes predominen amb un 75 % de mencions, respecte del 25 % de subjectes dones.

Aquestes dades posen en relleu un biaix en la representació de gènere en els mitjans de comunicació. La tendència no només afecta la diversitat de veus en les notícies, sinó que també podria tenir implicacions en les perspectives amb les quals s'aborden els diferents temes.

Les notícies d'agència o de redacció són la firma principal

Si s'analitzen les dades d'autoria, **la premsa digital mallorquina es nodreix majoritàriament de firmes anònimes**. Notícies d'agències, articles firmats per redacció, textos sense firma o amb inicials constitueixen més de la meitat de les informacions analitzades.

Les dades assenyalen que **el 53,1 % de les notícies estan signades amb un autoria indeterminada**. Aquesta xifra supera significativament la combinació de notícies escrites per periodistes homes (29,4 %), dones (16,8 %) i firmes mixtes de dona-home (0,7 %).

La prevalença de les agències en el contingut publicat pot tenir diverses implicacions. Per una banda, suggereix la dependència dels mitjans en fonts externes per a la generació de contingut, el que pot afectar l'angle i la profunditat amb què s'aborden els temes. Per altra banda, mostra la possible precarietat de recursos per diferenciar-se de la resta de mitjans de comunicació.

Per mitjans, en el cas de *Diariodemallorca.es* (DM), tan sols el 22,2 % de les notícies són redactades per periodistes dones, mentre que els homes signen el 35,9 % dels articles. **Encara més pronunciada és la diferència a *Ultimahora.es* (UH), on les dones redacten només el 10,3 % dels articles, enfront del 21,5 % que escriuen els homes**. Les notícies d'autoria indeterminada al DM representen el 40,5 %, mentre que a UH es disparen al 68,2 %.

Per seccions, **8 de cada 10 firmes d'opinió són masculines**. Els grans reportatges els firmen homes en 6 de cada 10 casos. I les plantilles de Successos també estan més masculinitzades que altres seccions, amb 5 de cada 10 articles firmats per redactors homes. **Els apartats amb més articles escrits per dones periodistes són Motor i Societat (4 de cada 10, en ambdós casos)** i Palma i Part forana (3 de cada 10, respectivament). Les firmes indeterminades predominen als publireportatges (100 % dels textos), Tecnologia (95 %), fotogaleries i vídeos (87,7 %) i Nacional (81,2 %).

Si es fa una observació més detallada de les seccions dins cada mitjà, UH mostra una absència total d'articles de periodistes dones en seccions com Tecnologia i Premsa rosa. En contrast, **DM presenta una distribució més equilibrada entre periodistes homes i dones en diverses seccions**. Per exemple, en la secció de Cultura, DM té un 29 % signats per redactores, mentre que la xifra és del 20 % a la capçalera del Grup Serra.

Similarment, en seccions com Economia i Motor, el diari d'Editorial Premsa Ibérica mostra una representació gairebé igualitària entre les firmes, una realitat que no es reflecteix a UH. Una altra clau és la **presència dominant d'homes en seccions com Esports, Successos i Opinió en ambdós mitjans**, que suggereix que certes àrees del periodisme continuen fortament masculinitzades.

Els titulars

Els títols i subtítols en clau femenina són minoria a la premsa digital mallorquina

El 36 % de les persones esmentades als títols i subtítols són dones · La proporció gairebé no varia entre mitjans

Les dones estan molt menys representades que els homes en els elements de titulació de les notícies. Es tracta d'una tendència clara que emergeix de l'anàlisi de l'escenari actual del periodisme digital a Mallorca. Segons les dades recollides de més de 1.300 notícies dels dos principals diaris digitals de l'illa, *Diariodemallorca.es* (DM) i *Ultimahora.es* (UH), només **el 36% de les persones incloses en els títols i subtítols són dones**.

Aquesta xifra, que reflecteix una disparitat de gènere, és **similar entre ambdós mitjans estudiats**. De tots els individus esmentats en títols i subtítols del DM, el 36,7 % són dones. A UH, la xifra és lleugerament inferior, amb un 34,9 %. En contraposició, els homes representen el 62,0 % de les mencions del DM i el 62,3 % a UH. El percentatge restant (un 1,3 % al DM i un 2,8 % a UH) corresponen a persones de les quals la intel·ligència artificial que ha analitzat les notícies no ha pogut determinar el sexe.

Aquestes dades posen en evidència una **tendència generalitzada cap a una major representació masculina en els elements més destacats de les notícies**. Aquesta propensió suscita una reflexió sobre com es construeixen i perceben les notícies. Els títols i subtítols són sovint el primer contacte que els lectors tenen amb les notícies, i la manera en què es representen els gèneres en aquests espais pot influir significativament en la percepció dels temes i les persones tractades.

La similitud en les proporcions entre els dos diaris digitals més grans de l'illa també indica que aquesta no és una qüestió aïllada o específica d'un mitjà, sinó una **característica comuna dins del panorama mediàtic de Mallorca**. Aquest fet suggereix la necessitat d'una revisió més profunda de les pràctiques periodístiques i la consideració de com les decisions editorials poden afectar la representació equitativa en la premsa.

Polítiques, celebritats i artistes: les professions femenines més esmentades

Les esportistes i les cineastes i professionals de la comunicació ocupen les posicions 3 i 4 del rànquing de les professions més destacades en la representació de les dones als mitjans

Les polítiques, les celebritats i les artistes han estat les professions femenines més esmentades a *Diariodemallorca.es* (DM) i *Ultimahora.es* (UH). L'anàlisi, realitzat a partir de les notícies publicades per ambdós mitjans en sengles dies de 2021 a 2023, ha permès descobrir les tendències de la premsa digital mallorquina en la representació de les dones en els mitjans de comunicació en una època en què la igualtat de gènere està en el punt de mira.

L'estudi reflecteix que les ocupacions de les dones incloses en les notícies són aquelles que tenen **rellevància (i, tot sovint, protagonisme) en l'esfera pública**. Alhora, les dades també palesen una tendència en els mitjans de comunicació per destacar les dones en rols de lideratge i influència.

Les categories política i cultural ocupen les dues primeres posicions del rànquing, seguida **en tercer lloc per les esportistes**, una àrea tradicionalment dominada per homes, però que ara mostren un canvi cap a una major inclusió. **En quarta posició, se situa el grup de cineastes i professionals dels mitjans**. Les professions menys representades per les dones han estat les de les jubilades, les policies i forces de seguretat i la d'altres serveis socio-sanitaris (que inclou feines com ara infermeria, serveis socials o treball social, i exclou les metgesses, que tenen una categoria pròpia).

Per seccions

A més, l'anàlisi de les seccions dels diaris mostra que les notícies de **Nacional, Cultura, Televisió/Cinema i Local han estat aquelles que més han esmentat dones**. Algunes de les tendències s'apunten com a més lògiques, com ara que a la secció de Cultura i a la de Televisió i Cinema, les més noticiables hagin estat les celebritats, les artistes, les cineastes i les professionals de la comunicació. O que a les seccions d'informació local de les Balears, d'Economia, d'Internacional i de Nacional, sobretot s'esmenten dones en política i funcionàries. A les planes de Successos, són les professionals del sector jurídic (jutgesses, advocades o fiscals) les qui lideren la presència.

Les variables de l'estudi també permeten fer una **comparativa entre mitjans**, que assenyalava les diferents visions de cada mitjà té per explicar la realitat. El **DM ha mostrat una tendència a destacar més les celebritats i artistes**, amb 3,51 punts percentuals més de representació que UH. Per contra, **UH ha posat més èmfasi en les dones polítiques** (2,98 punts percentuals més que el DM) i **el funcionariat femení** (2,12 punts més). Pel que fa a les esportistes, Diariodemallorca.es lidera amb una representació femenina 1,88 punts major que UH. En canvi, la capçalera del Grup Serra representa més les dones empresàries que la seva competidora d'Editorial Prensa Ibérica per 1,22 punts de diferència.

Aquestes dades tenen en compte aquells casos on els articles periodístics esmenten la feina de les persones incloses en les notícies, la qual cosa succeeix en un 47 % dels casos. **En l'altre 53 %, el càrrec de les dones mencionades no figura en el text**. Les xifres de l'estudi permeten que els professionals de la comunicació dels diferents mitjans puguin veure possibles tendències o biaixos en la seva forma de retratar la realitat de les dones en el sector laboral.

Notícia número 8:

Els parentius i les edats

El rol familiar s'esmenta quasi cinc vegades més en les dones que en els homes

El rol familiar s'inclou en l'1,5 % de les persones analitzades i amb molta més freqüència en les de sexe femení · L'edat s'assenyala en el 2 % dels individus, sobretot en celebritats i artistes

Les dones són descrites per les seves relacions de parentiu gairebé cinc vegades més que en el cas dels homes. L'estudi, que ha examinat les notícies dels dos principals diaris digitals de l'illa, *Diariodemallorca.es* i *Ultimahora.es*, revela que **en un 3,42% de les ocasions es menciona el rol familiar de les dones**, mentre que aquesta xifra es redueix a tan **sols un 0,75% en el cas dels homes**.

Això suposa que **és 4,6 vegades més freqüent que el parentiu serveixi per descriure una dona**. Aquesta dada mostra que existeix una disparitat significativa en la representació de rols familiars en els mitjans de comunicació de Mallorca. També posa en relleu que encara existeix una tendència a representar les dones des d'una perspectiva més domèstica o personal, en contrast amb la representació més professional o neutra dels homes.

Tot i això, **és una pràctica minoritària**. Dels 4.380 subjectes analitzats, només en 67 casos (l'1,5 %) s'ha esmentat la relació familiar. «**Filla**», «**esposa**» o «**mare**» són les paraules més comunes que fan servir els i les periodistes per descriure el parentiu de les dones. En el cas dels homes, els termes més utilitzats són «marit», «parella» o «fill».

Les edats, més freqüents en les dones i en les famoses

L'anàlisi de les edats també evidencia que incloure aquesta dada és una pràctica ben reduïda. **El 97,95 % de les persones mencionades en les notícies no tenen una edat assignada**, el que indica una tendència general a ometre aquesta informació. No obstant això, quan l'edat és esmentada, s'esmenta tres vegades més sovint en les persones adultes (1,42 % dels casos) que en les que estan en edat de jubilació (0,48 %).

També **és més habitual que les dones vegin publicada la seva edat (6 % de les esmentades) que els homes (3 %)**. A més, quan s'esmenta l'edat en les notícies, sovint es concentra en figures públiques, com ara celebritats i artistes. Els segueixen esportistes i membres de la classe política.

Les dades de menció de les edats segons l'ocupació suggereixen un interès particular en els aspectes més personals i menys rellevants professionalment de les vides d'aquestes persones i plantegen el dubte sobre la possible frivolitat en el tractament de les persones famoses, especialment en el cas de les dones.

[Notícia número 9:](#)

La fiabilitat de les dades

La tecnologia al servei del periodisme: l'anàlisi de textos amb IA encerta el 94,9 % de les persones

L'ús d'intel·ligència artificial per a la detecció de persones en articles de premsa té una precisió elevada · L'estudi de les professions és la variable que obté les puntuacions més baixes del model

En l'era de la informació digital, l'ús de la intel·ligència artificial (IA) per analitzar grans volums de dades ha esdevingut una **eina imprescindible, també per al periodisme**. L'anàlisi duta a terme en el marc del projecte *Paritat de gènere a la premsa local digital de Mallorca* ha mesurat que la fiabilitat de la IA en identificar i categoritzar les variables de gènere dins de les notícies és força elevada.

Segons l'estudi, el **model GPT-4-Turbo** de l'empresa OpenAI ha aconseguit una puntuació global del 94,9 % en la detecció de persones als articles de premsa. Aquesta fiabilitat s'ha mesurat a través de diversos paràmetres. Destaquen la **precisió (93,64 %)**, que mesura quantes de les persones identificades eren realment persones, i el **reclam o sensibilitat (96,19 %)**, que expressa la capacitat de la IA per trobar totes les persones reals. La puntuació mitjana ponderada d'aquests dos valors és l'anomenada **puntuació F1, que és del 94,9 %**.

Les dades de fiabilitat s'han calculat a partir d'una **revisió manual de 131 de les 1.311 notícies** incloses en l'estudi. És a dir, la fiabilitat global del 94,9 % s'ha d'entendre **en comparació a com una persona humana executaria la mateixa tasca** amb el mateix conjunt d'articles periodístics. Aquests números indiquen que la IA és molt eficaç en identificar correctament les persones a les notícies i en capturar la majoria de les persones reals presents en els articles.

No obstant això, l'anàlisi també ha descobert una àrea significativa de millora en observar altres variables i, més concretament, en l'estudi de les **ocupacions**. En aquest apartat, la IA només ha assolit una **fiabilitat del 46,1 %**. La xifra es deu sobretot a la tendència del model d'ometre la informació sobre l'ocupació, especialment en notícies amb un gran nombre d'individus.

En contrast, la IA ha demostrat una **fiabilitat excepcionalment alta en analitzar altres paràmetres** com el nom (99,82 %), l'edat (100,00 %), el rol familiar (99,46 %) o si les persones esmentades eren descrites com a expertes (99,27 %), testimonis (99,64 %) o víctimes (també 99,64 %). Aquestes puntuacions evidencien l'habilitat de la IA per tractar amb precisió ordres més concretes que la identificació general de la professió a partir del context de la notícia.

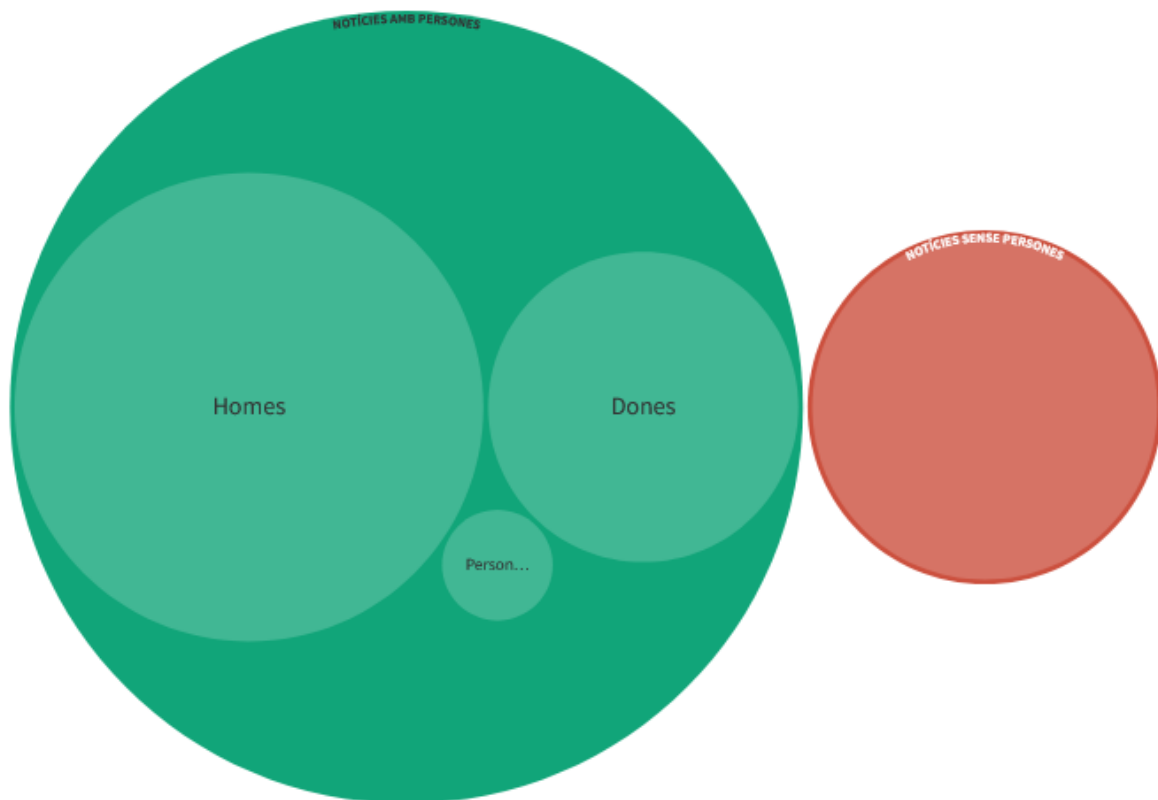
Les dades mostren que l'ús de la IA en el periodisme ofereix un **gran potencial per a l'extracció ràpida, eficient i precisa de dades**. No obstant això, els resultats obtinguts en l'estudi suggereixen que, tot i l'eficàcia general, els models de llenguatge d'intel·ligència artificial encara necessiten ser optimitzats. En aquest sentit, encara és necessària la supervisió humana en l'ús de tecnologies automatitzades, especialment en tasques que requereixen un alt grau de comprensió contextual i matisos.

Infografia número 1:

Les dones a les notícies no arriben al 30 %

Les dones a les notícies no arriben al 30 %

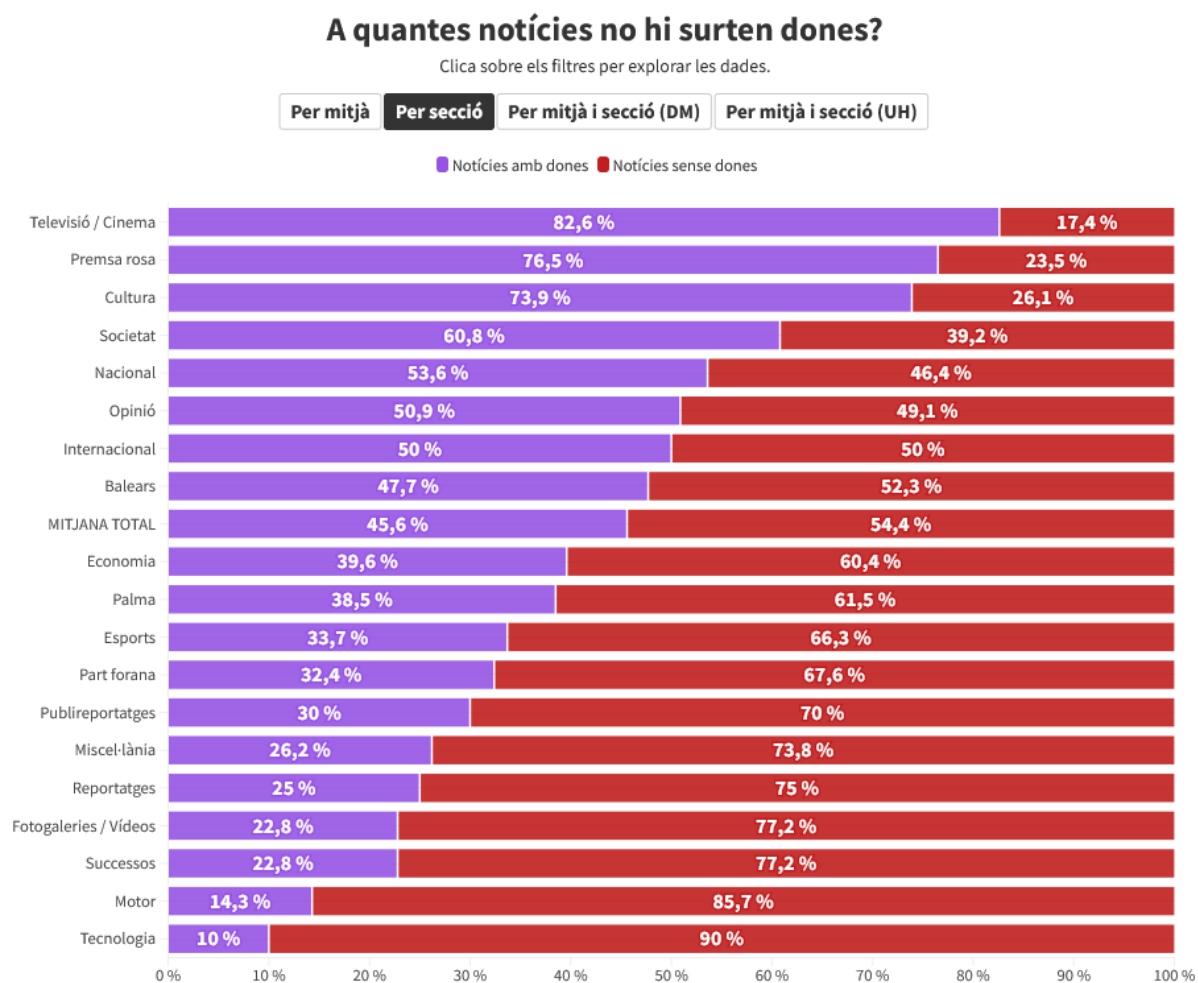
Canvia la mida dels cercles amb el filtre. Clica a un cercle per veure'n els detalls.



Font: [Paritat de gènere a la premsa local digital de Mallorca](#) • Infografia: [Aitor F. Vallespir](#)

Infografia número 2:

A quantes notícies no hi surten dones?



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Infografia número 3:

La disparitat de gènere a la premsa, en cinc indicadors

La disparitat de gènere a la premsa, en cinc indicadors

Per tal que les dones estiguin representades als mitjans de manera paritària i proporcional a la població, tots aquests indicadors haurien d'assolir el 50%.
Clica sobre les dades per veure'n els detalls.

Sexe **Ambdós** Homes Dones

Homes Dones

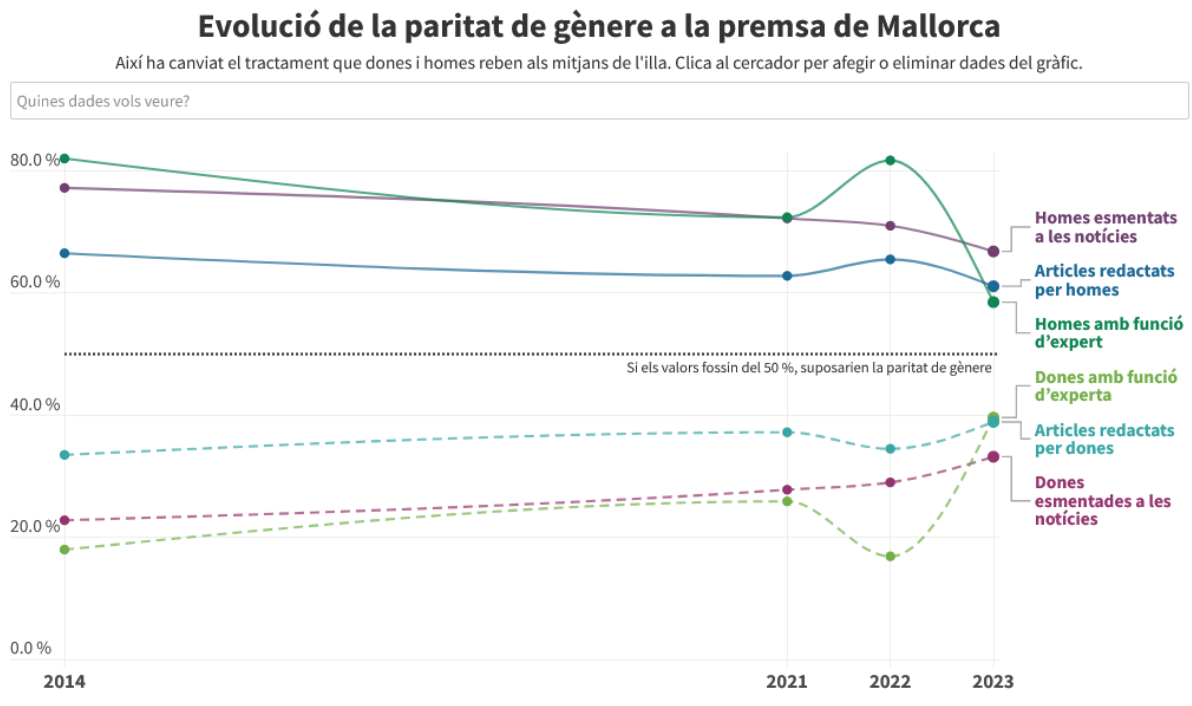


Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Nota: En els anys 2021 a 2023, la suma de dones i homes no és 100%. Els percentatges restants fins al 100% es corresponen a persones de les quals la IA no ha pogut determinar el sexe.

Infografia número 4:

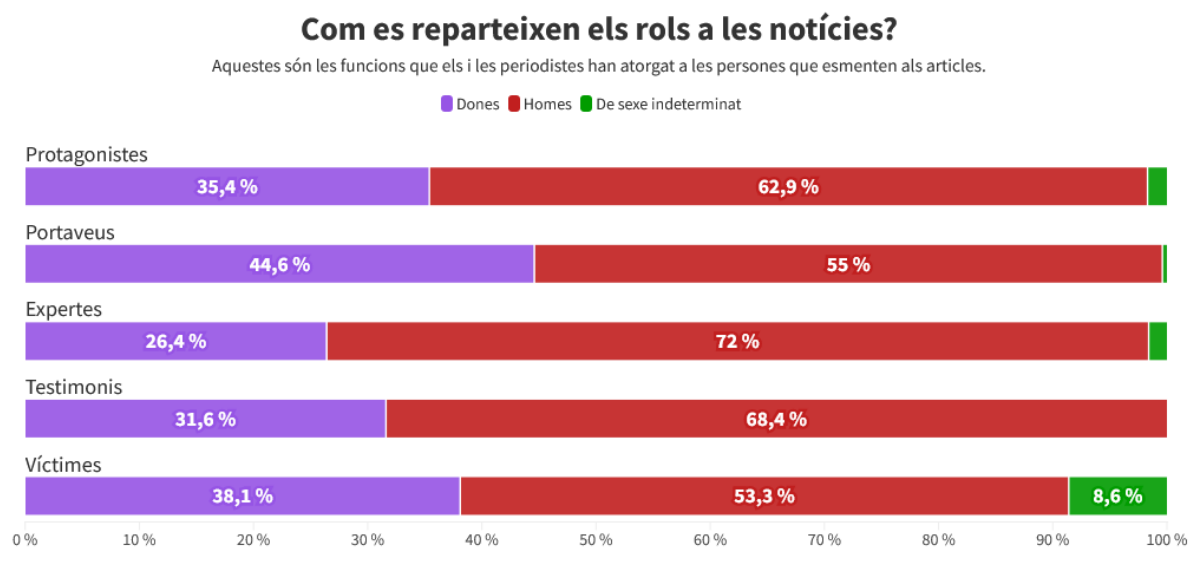
Evolució de la paritat de gènere a la premsa de Mallorca



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Infografia número 5:

Com es reparteixen els rols a les notícies?



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

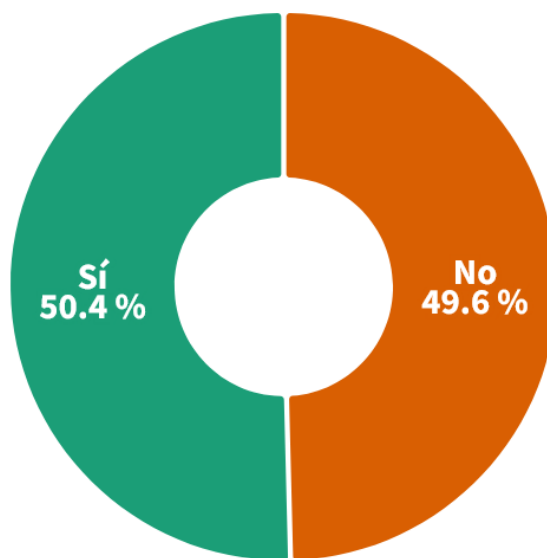
Infografia número 6:

Qui surt a les fotografies?

Qui surt a les fotografies?

La IA ha analitzat 1.241 imatges publicades als dos principals diaris digitals de Mallorca. Utilitza el filtre per veure qui hi apareix.

Les fotografies inclouen persones? dones? homes?

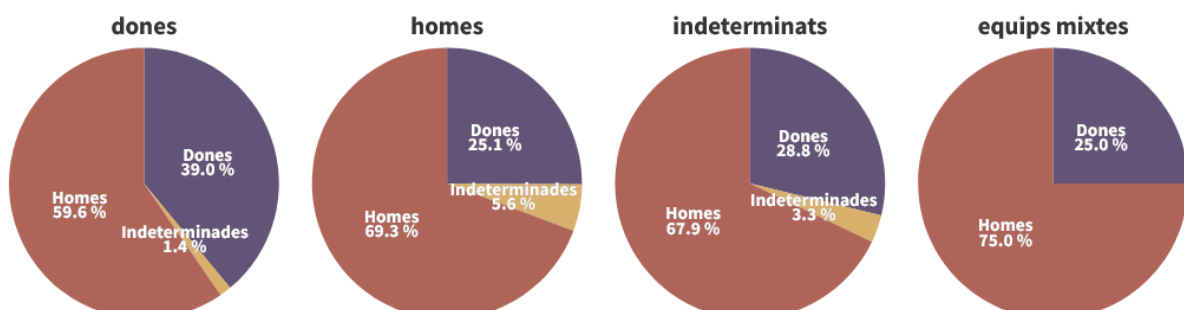


Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Infografia número 7:

Quan la notícia la firmen...

Quan la notícia la firmen...

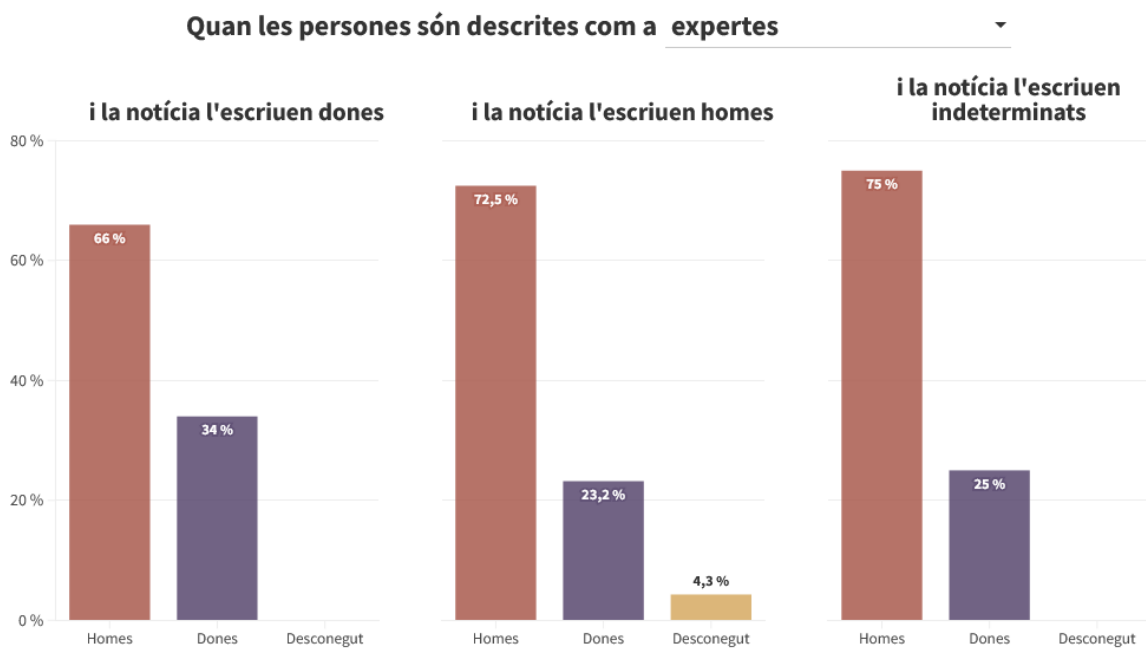


Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Nota: Els redactors indeterminats inclouen agències de notícies; articles firmats per redacció, per inicials de les quals es desconeix l'autoria o per diaris d'un mateix grup editorial, i altres casos semblants.

Infografia número 8:

Quan les persones són descrites com a...



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Nota: Els redactors indeterminats inclouen agències de notícies; articles firmats per redacció, per inicials de les quals es desconeix l'autoria o per diaris d'un mateix grup editorial, i altres casos semblants.

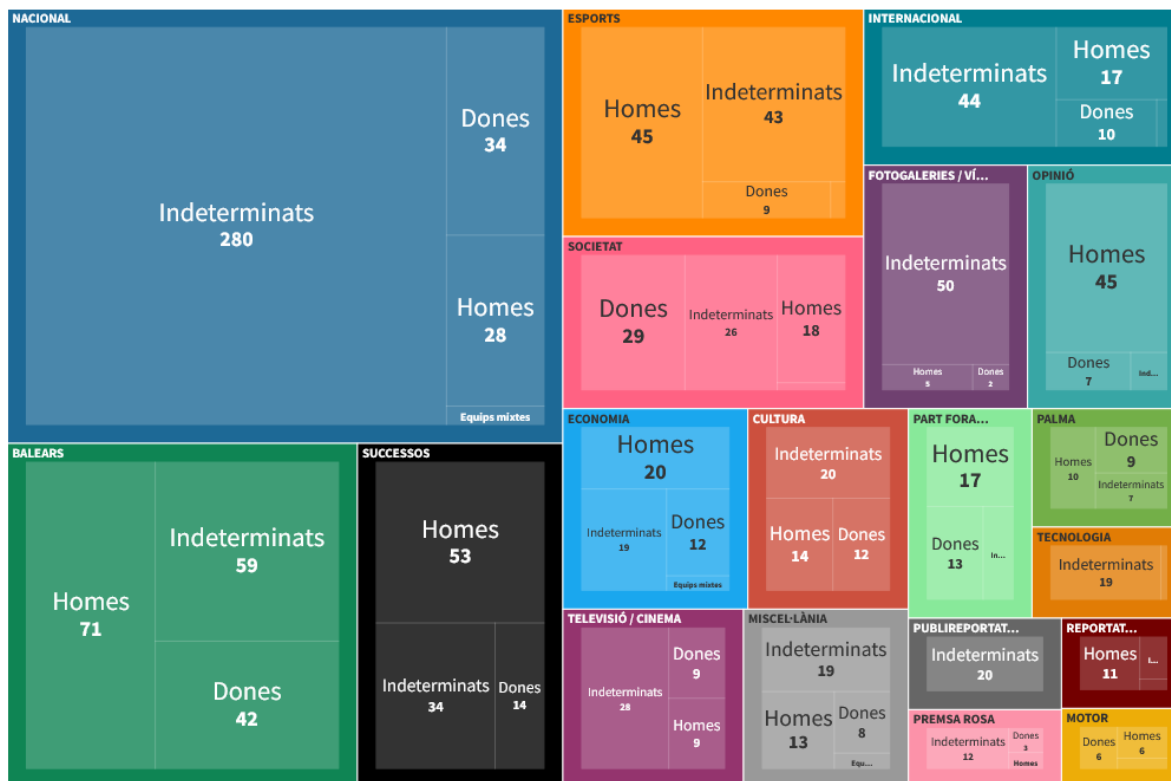
Infografia número 9:

Qui firma les notícies?

Qui firma les notícies?

Canvia de filtre per conèixer més resultats. Clica a un grup per ampliar-lo.

Per sexe Per mitjà i sexe **Per secció i sexe**

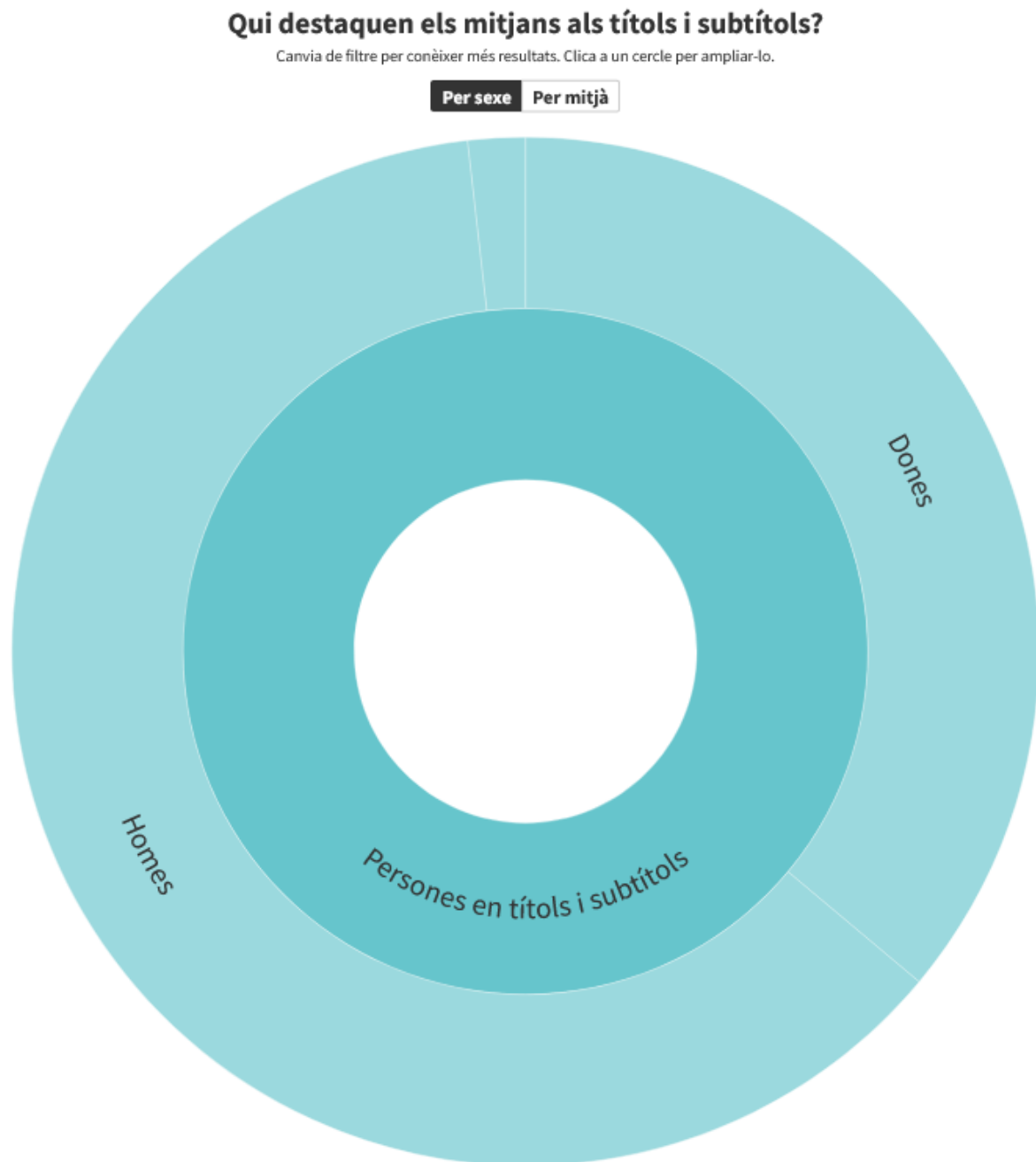


Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Àlgor F. Vallespir

Nota: Els redactors indeterminats inclouen agències de notícies; articles firmats per redacció, per inicials de les quals es desconeix l'autoria o per diaris d'un mateix grup editorial, i altres casos semblants.

Infografia número 10:

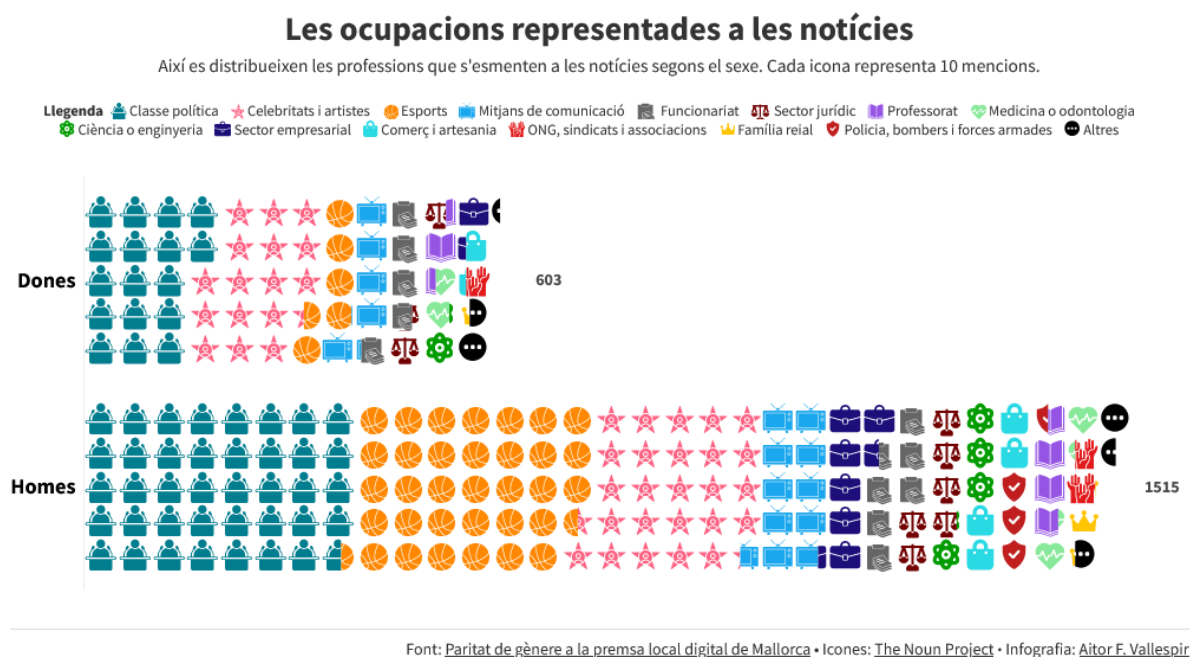
Qui destaquen els mitjans als títols i subtítols?



Font: [Paritat de gènere a la premsa local digital de Mallorca](#) • Infografia: [Aitor F. Vallespir](#)

Infografia número 11:

Les ocupacions representades a les notícies



Infografia número 12:

Les probabilitats de ser presentada com a «filla», «esposa» o «mare»

Les probabilitats de ser presentada com a «mare», «filla» o «parella de»

Clica damunt de les icones per veure més informació.



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir

Infografia número 13:

Els parentius més utilitzats

Els parentius més utilitzats

Aquests són els termes que han fet servir els i les periodistes dels diaris digitals de Mallorca per descriure els rols familiars de les persones de les quals parlaven. Clica sobre una paraula per veure'n més detalls.



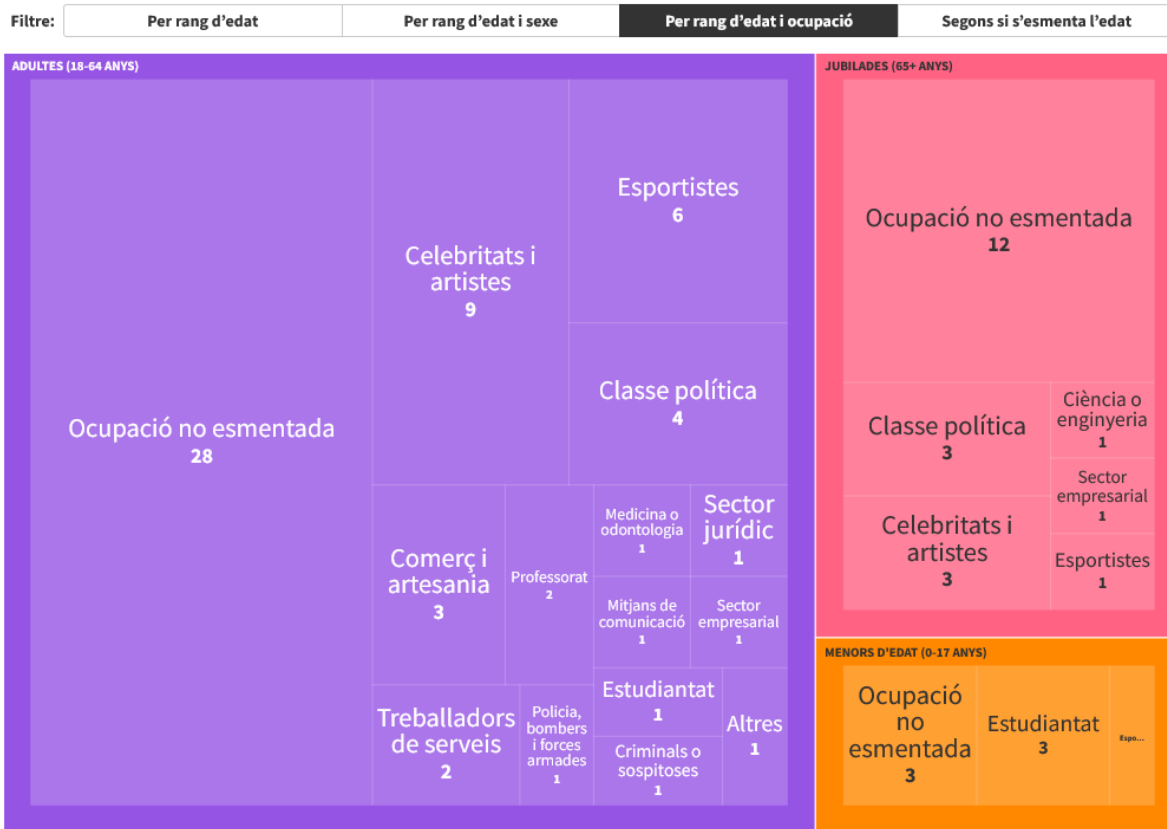
Font: [Paritat de gènere a la premsa local digital de Mallorca](#) • Infografia: [Aitor F. Vallespir](#)

Infografia número 14:

L'ús de l'edat en la premsa digital de Mallorca

L'ús de l'edat en la premsa digital de Mallorca

Canvia de filtre per conèixer més resultats. Clica a un grup per ampliar-lo i torna enrere amb la fletxa.



Font: [Paritat de gènere a la premsa local digital de Mallorca](#) • Infografia: [Àitor F. Vallespir](#)

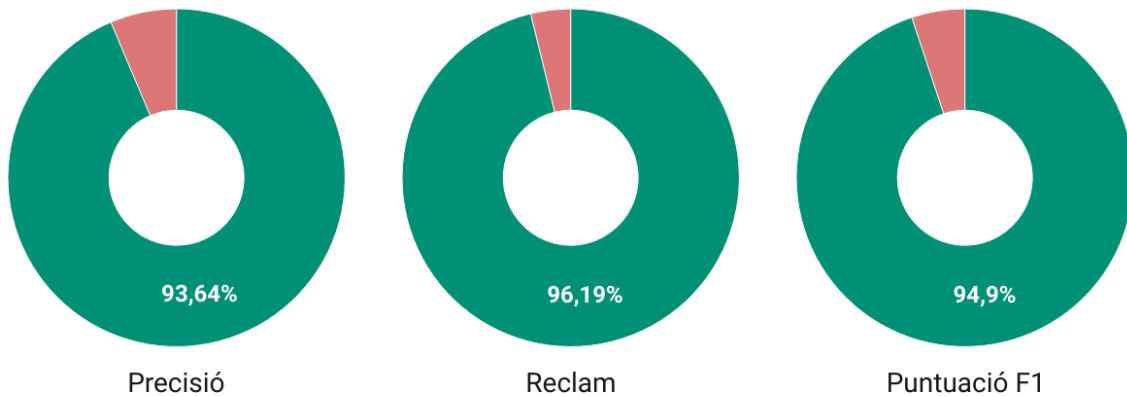
Infografia número 15:

La fiabilitat de la IA per identificar persones a les notícies

La fiabilitat de la IA per identificar persones a les notícies

La precisió i el reclam mesuren si el model s'ha confós en identificar individus i si se n'ha deixat algun per comptabilitzar. La puntuació F1 és la mitjana ponderada de les altres dues mètriques.

■ **Fiabilitat** ■ Error



Gràfic: Aitor F. Vallespir • Font: Paritat de gènere a la premsa local digital de Mallorca • Creat amb Datawrapper

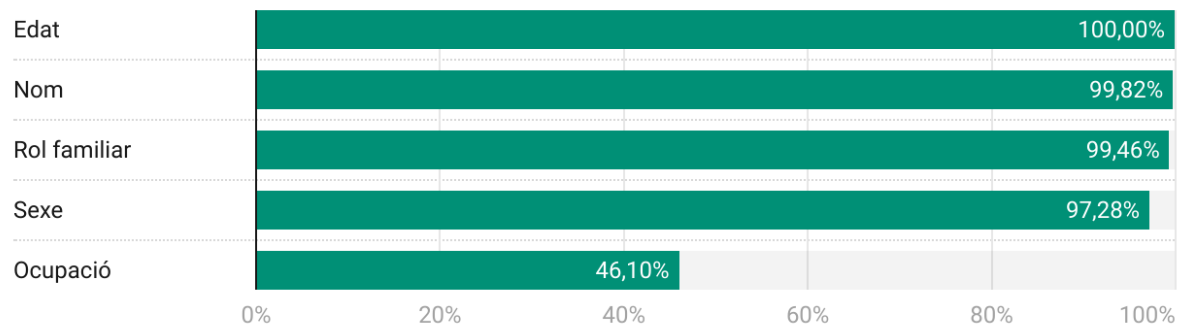
Infografia número 16:

La fiabilitat de la IA en analitzar les persones esmentades a les notícies

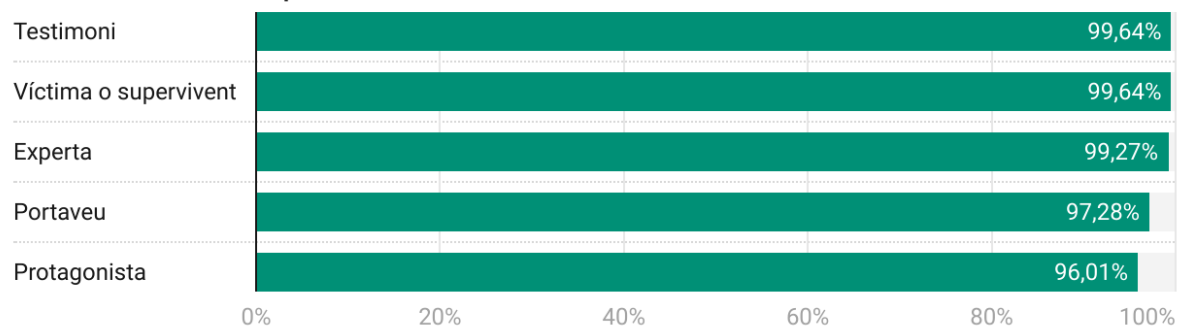
La fiabilitat de la IA en analitzar les persones esmentades a les notícies

L'estudi *Paritat de gènere a la premsa local digital de Mallorca* s'ha centrat en deu paràmetres per analitzar els textos. La detecció de **les ocupacions** ha estat l'únic camp en el qual la IA ha tingut una **fiabilitat baixa** en comparació a la resta de variables.

Fiabilitat en detectar quan, sobre una persona, s'esmenta...



Fiabilitat en detectar si la persona és descrita com a...



Gràfic: Aitor F. Vallespir • Font: Paritat de gènere a la premsa local digital de Mallorca • Creat amb Datawrapper

Annex 15. Captures de pantalla del web

Les captures de pantalla han estat extretes del web que es pot consultar a través de l'adreça <https://premsaigenere.wordpress.com/>.

Captura número 1:

Portada del web

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

Articles Infografies Metodologia Dades obertes Sobre aquest projecte

ARTICLES

Les dades generals
Només 3 de cada 10 persones que surten als diaris digitals de Mallorca són dones
L'anàlisi de les notícies dels dos principals mitjans en línia de l'illa mostra que les dones estan infrarepresentades als mitjans - El 54 % dels articles no inclou ni esmenta cap dona
Aitor F. Vallespí

La evolució en el temps
La presència de les dones a les notícies s'incrementa 10 punts des de 2014
Aitor F. Vallespí

El rol
Els homes són protagonistes dels articles el doble de vegades que les dones
Aitor F. Vallespí

Les imatges
Gairebé la meitat de les fotografies publicades no mostra cap dona
Aitor F. Vallespí

El biaix de gènere
Els redactors homes consulten més fonts masculines que les periodistes
Aitor F. Vallespí

El titular
Els títols i subtítols en clau femenina són minoria a la premsa digital mallorquina
Aitor F. Vallespí

Les ocupacions
Polítics, celebritats i artistes: les professions femenines més esmentades
Aitor F. Vallespí

El parentiu i les edats
El rol familiar s'esmenta quasi cinc vegades més en les dones que en els homes
Aitor F. Vallespí

La fiabilitat de les dades
La tecnologia al servei del periodisme: l'anàlisi de textos amb IA encerta el 94,9 % de les persones
Aitor F. Vallespí

EN XIFRES

El projecte ha revisat tots els continguts publicats a *Diariodemallorca.es* i *Ultimahora.es* dels dies 21 de setembre de 2021, 20 de setembre de 2022 i 19 de setembre de 2023. [Descobrimos un més.](#)

1.311 articles de premsa analitzats dels dos principals diaris digitals de Mallorca

4.380 persones trobades als articles mitjançant intel·ligència artificial, de les quals 1.284 eren dones

713 notícies no esmentaven cap dona i 452 de 911 fotografies tampoc no mostraven cap

INFOGRAFIES

Infografies
A quantes notícies no hi surten dones?
Aitor F. Vallespí

Infografies
Les dones a les notícies no arriben al 30 %
Aitor F. Vallespí

Infografies
Qui firma les notícies?
Aitor F. Vallespí

Infografies
Quan les persones són descrites com a...
Aitor F. Vallespí

1 2 3 4 Pàgina següent →

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'Aitor F. Vallespí

Captura número 2:

Exemple d'article

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

Articles Infografies Metodologia Dades obertes Sobre aquest projecte



El rol

Els homes són protagonistes dels articles el doble de vegades que les dones

La representació de les dones com a expertes encara és minoritària: només una de cada quatre veus expertes és de sexe femení

Aitor F. Vallejo
07/01/2024 · 2 minuts

Las funciones que dones i homes desenvolupen a les notícies assenyalen desequilibris notables de gènere als mitjans digitals de Mallorca. **Els homes apareixen com a protagonistes en els articles de premsa gairebé el doble de vegades que les dones**, una estadística que destaca la persistència de patrons de gènere tradicionals en el periodisme.

En les informacions en les quals un individu té una funció específica, **els homes són el personatge principal de la notícia en un 62,9 % dels casos**, mentre que les dones ho són només en un 35,4 % d'ocasions. Els rols protagonistes, en particular, mostren una de les divergències més notables de l'estudi.

Tot i això, és encara més colpidora la representació de les dones en rols d'expertesa. **Només el 26,4 % de les figures expertes dins de les notícies són dones**, un fet que posa de manifest la infrarepresentació femenina en posicions d'autoritat i coneixement en diversos camps. Aquesta disparitat podria tenir implicacions profundes, no només en la manera com s'informa sobre temes específics, sinó també en la formació de l'opinió pública i en la promoció de la igualtat de gènere en l'àmbit acadèmic i professional.

Com es reparteixen els rols a les notícies?

Aquestes són les funcions que els i les periodistes han atorgat a les persones que esmenten als articles.

Funció	Dones	Homes	No s'ha identificat
Protagonistes	35,4 %	62,9 %	
Portaveus	44,6 %	55 %	
Expertes	26,4 %	72 %	
Testimonis	31,6 %	65,4 %	
Víctimes	38,1 %	53,3 %	8,6 %

Font: Paritat de gènere a la premsa local digital de Mallorca + Infografia Aitor F. Vallejo

En contrast, **la representació com a portaveus és més equilibrada**, amb les dones sent presentades en aquesta funció en un **44,6 % dels casos**, en comparació a un 55 % dels homes que són esmentats amb aquest rol.

La diferència de gènere a l'hora de presentar una persona com a víctima és menys marcada. **De totes les persones introduïdes com a víctima o supervivent, el 38,1 % són dones**, el 53,3 % són homes i el 8,6 % són persones de les quals la intel·ligència artificial no ha pogut determinar el sexe.

LLERGEIX-NE MÉS

- Només 3 de cada 10 persones que surten als diaris digitals de Mallorca són dones
23/12/2023
- Els redactors homes consulten més dones masculines que les periodistes
27/12/2023
- La presència de les dones a les notícies s'incrementa 10 punts des de 2014
24/12/2023
- El rol familiar s'incrementa quasi cinc vegades més en les dones que en els homes
30/12/2023

SOBRE L'AUTOR

Aitor F. Vallejo
Periodista especialitzat en dades.

El projecte Paritat de gènere a la premsa local digital de Mallorca és una adaptació de l'informe *El rol de les dones en les notícies. Fet per les investigadores Elvira Albés Rufas i Sílvia Majó Vázquez el 2014.*

Un web d'Aitor F. Vallejo

Captura número 3:

Exemple d'infografia

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

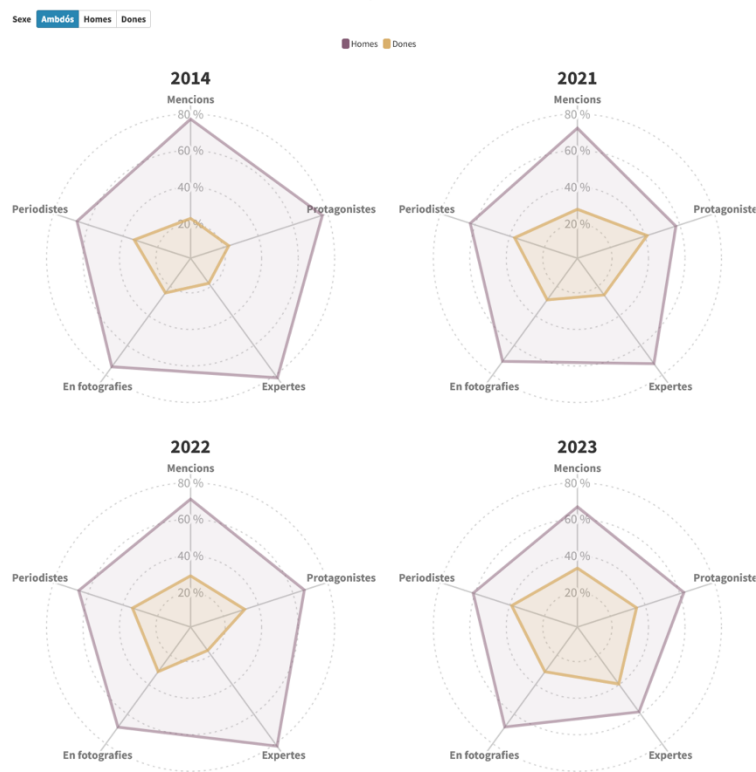
[Articles](#) [Infografies](#) [Metodologia](#) [Dades obertes](#) [Sobre aquest projecte](#)

La disparitat de gènere a la premsa, en cinc indicadors

Aitor F. Vallespir
07/01/2024

La disparitat de gènere a la premsa, en cinc indicadors

Per tal que les dones estiguin representades als mitjans de manera paritària i proporcional a la població, tots aquests indicadors haurien d'assolir el 50%. Clica sobre les dades per veure'n els detalls.



Font: Paritat de gènere a la premsa local digital de Mallorca • Infografia: Aitor F. Vallespir
Nota: En els anys 2021 a 2023, la suma de dones i homes no és 100%. Els percentatges restants fins al 100% es corresponen a persones de les quals la IA no ha pogut determinar el sexe.

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'Aitor F. Vallespir

Captura número 4:

Secció "Articles"

The screenshot shows the 'Articles' section of a website. At the top, there is a dark header with the title 'PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA' in white. Below the header, there is a navigation menu with links: 'Articles', 'Infografies', 'Metodologia', 'Dades obertes', and 'Sobre aquest projecte'. The main content area is titled 'Articles' and lists five articles, each with a date, a title, a subtitle, and a brief description. The articles are:

- 05/01/2024**
La tecnologia al servei del periodisme: l'anàlisi de textos amb IA encerta el 94,9 % de les persones
L'ús d'intel·ligència artificial per a la detecció de persones en articles de premsa té una precisió elevada. L'estudi de les professions és la variable que obté les puntuacions més baixes del model [Llegeix l'article](#)
- 05/01/2024**
El rol familiar s'esmenta quasi cinc vegades més en les dones que en els homes
El rol familiar s'inclou en l'1,5 % de les persones analitzades i amb molta més freqüència en les de sexe femení. L'edat s'assenyala en el 2 % dels individus, sobretot en celebritats i artistes [Llegeix l'article](#)
- 06/01/2024**
Polítics, celebritats i artistes: les professions femenines més esmentades
Les esportistes i les cineastes i professionals de la comunicació ocupen les posicions 3 i 4 del rànquing de les professions més destacades en la representació de les dones als mitjans [Llegeix l'article](#)
- 05/01/2024**
Els títols i subtítols en clau femenina són minoria a la premsa digital mallorquina
El 36 % de les persones esmentades als títols i subtítols són dones. La proporció gairebé no varia entre mitjans [Llegeix l'article](#)
- 07/01/2024**
Els redactors homes consulten més fonts masculines que les periodistes
Les notícies d'agència o firmades per la redacció representen més de la meitat dels articles publicats en els digitals mallorquins [Llegeix l'article](#)

At the bottom of the article list, there is a page number '1 2' and a link 'Pàgina següent »'. Below the article list, there is a dark footer with the following text:

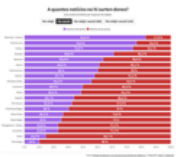
El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'Àlex P. Valdesp

Captura número 5:

Secció "Infografies"






07/01/2024

A quantes notícies no hi surten dones?


Clica aquí per veure la infografia a pantalla completa.
[Descobreix-ne més](#)



07/01/2024

Les dones a les notícies no arriben al 30 %

[Descobreix-ne més](#)



06/01/2024

Qui firma les notícies?

[Descobreix-ne més](#)

1 2 3 4 5 6 Pàgina següent →

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'[Aitor F. Villegas](#)

Captura número 6:

Secció “Metodologia”

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

[Articles](#) [Infografies](#) [Metodologia](#) [Dades obertes](#) [Sol·licitar projecte](#)

Metodologia

Com hem elaborat les bases de dades per als reportatges i les infografies? A continuació, trobaràs un resum del procés de treball que hem fet servir.

Recerca prèvia

Altres estudis han analitzat la representació de les dones als mitjans de comunicació. Generalment, aquests treballs es fan amb persones voluntàries que compleixen feines de buidatge. El nostre ha consistit a agilitzar el procés d'obtenció i anàlisi dels articles dels diaris digitals, tot conservant el màxim de precisió possible.

Selecció de la mostra

L'estudi tracta sobre els dos principals mitjans digitals de Mallorca per volum mensual de visites, segons les dades d'CID Interactiva: *Ultimahora.es*, amb gairebé 5 milions d'usuaris únics l'octubre de 2023, i *Diariodemallorca.es*, amb 2,8 milions el mateix mes.

Les dades triades per a la mostra han estat del mateix dia de la setmana (dimarts), ja que les capçaleres poden tenir seccions diferents segons el dia. També s'ha tingut en compte que sigui de la mateixa setmana del mateix mes, perquè no influeixi l'estacionalitat de la informació. S'han descartat les setmanes de vacances i també les dates en les quals hi va haver alguna notícia de gran impacte o d'alt que pugui desvirtuar la comparativa entre anys.

D'acord amb aquests criteris, les dades seleccionades han estat el 21 de setembre de 2021, el 20 de setembre de 2022 i el 19 de setembre de 2023.

Obtenció de les dades

Les dades dels diaris digitals s'han obtingut a partir d'un raspig d'informació (*scraping*). La informació recuperada s'ha sotmès a un procés d'anàlisi estructural (*paring*) per poder convertir cada notícia en una fila d'una taula.

Tractament de la informació

El treball previ ha consistit a eliminar resultats duplicats, que no tinguessin text o que no fossin valuosos per a l'anàlisi. S'han normalitzat els valors d'ambdós diaris digitals perquè siguin comparables (per exemple, els noms de les seccions). Algunes dades, com ara els noms dels periodistes que signen amb inicials, s'han depurat a mà per aconseguir millor resultat.

Una vegada enllestida la base de dades de les notícies, s'han analitzat els textos de forma automatitzada per extreure les dades de les variables d'anàlisi. La informació s'ha processat amb Python 3.12.0. Les visualitzacions de dades s'han dissenyat amb [Flourish](#) i [Datavizapper](#).

Definició de les variables d'anàlisi

Les variables d'anàlisi es basen en les feines de monitoratge [Global Media Monitoring Project](#), que pots consultar [aquí](#). Algunes variables s'han descartat per controlar els costos i temps del projecte i per simplificar l'anàlisi mitjançant eines d'intel·ligència artificial.

Ús de la intel·ligència artificial

El projecte ha fet servir diversos models GPT d'OpenAI per analitzar les dades. Hem comprovat manualment la veracitat d'una submostra dels resultats. El resultat a l'hora de detectar persones ha estat un índex de [falsedat](#) del 98,9% en comparació a si la tasca l'hagués realitzada una persona humana.

Publicació dels resultats

Tots els resultats de l'estudi els trobaràs en aquest web. Les dades més rellevants són als reportatges i les infografies però, si vols més detall, pots descarregar les microdades a la secció [Dades obertes](#). Amb aquesta feina, volem contribuir que més professionals de la comunicació puguin repetir aquest projecte. Per això, també oferim el [codi font](#) que hem fet servir en obert.

Trobes a faltar alguna informació sobre la metodologia?

[ENVIJA UN CORREU](#)

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Enira Altés Rafas i Sílvia Mayó Vázquez el 2014.

Un web d' [enira i silvia](#)

Captura número 7:

Secció "Dades obertes"

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

[Articles](#) [Infografies](#) [Metodologia](#) [Dades obertes](#) [Sobre aquest projecte](#)

Dades obertes



Microdades

En aquest [enllaç](#), pots descarregar les dades del *scraping* dels diaris digitals i les dades obtingudes amb IA. El fitxer inclou un esquema amb les relacions entre les taules.



Estadístiques

Els conjunts de dades que pots descarregar [aquí](#) inclouen les estadístiques tractades per elaborar els articles i les infografies.

[Su](#)



Codi font

Descarrega [aquí](#) el codi font del projecte en un quadern de Jupyter per a Python 3.12.0. Recorda que per executar el codi necessitaràs una clau d'API d'[OpenAI](#).



Necessites més dades?

Envia un [correu electrònic](#).

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'Antor F. Vallespí



Captura número 8:

Secció "Sobre aquest projecte"

PARITAT DE GÈNERE A LA PREMSA LOCAL DIGITAL DE MALLORCA

[Articles](#) [Infografies](#) [Metodologia](#) [Dades obertes](#) [Sobre aquest projecte](#)

Sobre aquest projecte

El projecte *Paritat de gènere a la premsa local digital de Mallorca: el cas de Diaridemallorca.es i Ultimahora.es* s'ha desenvolupat com a treball final del Màster universitari en periodisme i comunicació digital: dades i noves narratives de la Universitat Oberta de Catalunya.

La tutora del treball ha estat la periodista Victòria Oliveres Torrecassana i la responsable de l'assignatura ha estat la professora Candela Ollé Castellà. Aquest treball final de màster ha estat entregat i defensat el gener de 2024.

El treball ha tingut per objectiu estudiar la presència de dones en els dos principals mitjans digitals locals de Mallorca, alhora que ha volgut aprofundir en els avantatges que les tecnologies de la informació proporcionen als professionals de la comunicació a l'hora de recopilar, analitzar i difondre dades per a històries periodístiques.

El projecte *Paritat de gènere a la premsa local digital de Mallorca* és una adaptació de l'informe *Un dia a les notícies en clau de gènere*, fet per les investigadores Elvira Altés Rufias i Sílvia Majó Vázquez el 2014.

Un web d'[Aïnor F. Vallespir](#)



Annex 16. Codi programat per a l'automatització de l'extracció i anàlisi de les notícies

El codi font que es reproduïx a continuació està programat per a la versió de Python 3.12.0. Es pot descarregar en format quadern de Jupyter (un fitxer amb extensió *.ipynb*) a través de l'enllaç https://premsaigenere.files.wordpress.com/2023/12/premsaigenere_codifont.zip.

1 Càrrega inicial

1.1 Llibreries

```
[ ]: from bs4 import BeautifulSoup # Per llegir el codi HTML de forma més amigable
import re # Per extreure els URL de les notícies i dades dels logs
    ↳ mitjançant expressions regulars
from newspaper import Article # Per llegir cada notícia i dividir-la en camps
import os # Per comprovar si ja s'ha fet l'scraping i no repetir-lo cada
    ↳ vegada
import pandas as pd # Per al tractament i anàlisi de dades
import openai # Per a l'anàlisi automatitzada dels textos
import json # Per obrir l'arxiu amb la clau de l'API
import logging # Per crear logs d'execució de certes tasques de l'API
import datetime # Per anomenar seqüencialment els arxius de log
import requests # Per comprovar si els URL de les imatges encara funcionen
import numpy as np # Per a transformacions a les estadístiques
```

1.2 Arxius ja processats

```
[ ]: df = pd.read_csv('df.csv')
df = df.fillna('')

df_canvis = pd.read_csv('df_canvis.csv')
df_canvis = df_canvis.fillna('')

df_notícies = pd.read_csv('df_notícies.csv')
df_notícies = df_notícies.fillna('')

df_firmes = pd.read_csv('df_firmes.csv')
df_firmes = df_firmes.fillna('')

df_persones = pd.read_csv('df_persones.csv')
df_persones = df_persones.fillna('')

with open('notícies_processades.txt', 'r') as txt:
    lines = txt.readlines()
notícies_processades = set(line.strip() for line in lines)
del lines, txt

df_errors = pd.read_csv('df_errors.csv')
df_errors = df_errors.fillna('')

df_errors2 = pd.read_excel('df_errors2.xlsx')
df_errors2 = df_errors2.fillna('')
```

2 Importació de les dades de *Diariodemallorca.es*

2.1 Obre el codi font HTML

Es carreguen els fitxers HTML dels dies analitzats d'un dels diaris. Els objectes es creen amb les inicials del mitjà i les dues darreres xifres de l'any analitzat.

```
[ ]: # Obre els HTML en mode de només lectura
with open(r'RAW\DM20230919.htm', 'r', encoding='utf-8') as DM23:
    DM23html = DM23.read()

with open(r'RAW\DM20220920.htm', 'r', encoding='utf-8') as DM22:
    DM22html = DM22.read()

with open(r'RAW\DM20210921.htm', 'r', encoding='utf-8') as DM21:
    DM21html = DM21.read()
```

Després d'inspeccionar el codi font a un navegador, s'observa que en el cas de *Diariodemalorca.es* cal filtrar pels objectes *article* amb la propietat *class:lst*.

```
[ ]: # Interpreta el codi font amb BeautifulSoup i filtra els objectes 'article'
      ↳ amb la propietat 'class:lst'
bs = BeautifulSoup(DM23html)
DM23url = bs.find_all('article', {'class': 'lst'})

bs = BeautifulSoup(DM22html)
DM22url = bs.find_all('article', {'class': 'lst'})

bs = BeautifulSoup(DM21html)
DM21url = bs.find_all('article', {'class': 'lst'})
```

2.2 Crea una llista dels URL

S'extreuen els enllaços de cadascun dels fitxers de codi font i s'ajunten a un sol fitxer *txt*.

```
[ ]: DM23_llista = []
DM22_llista = []
DM21_llista = []

# Troba els URL de DM23url
DM23url_str = [str(a) for a in DM23url] # Converteix a string
for url in DM23url_str:
    DM23_llista.append(str(url)) # Afegeix cada URL
DM23url_llista = re.findall(r'itemid="([\^"]+)"', ' '.join(DM23url_str) ) #
↳ Amb una expressió regular, cerca cada URL dins 'itemid' i filtra el text
↳ entre les cometes del codi font.

# Troba els URL de DM22url
DM22url_str = [str(a) for a in DM22url]
for url in DM22url_str:
    DM22_llista.append(str(url))
DM22url_llista = re.findall(r'itemid="([\^"]+)"', ' '.join(DM22url_str) )

# Troba els URL de DM21url
DM21url_str = [str(a) for a in DM21url]
for url in DM21url_str:
    DM21_llista.append(str(url))
DM21url_llista = re.findall(r'itemid="([\^"]+)"', ' '.join(DM21url_str) )
```

```
# Uneix les tres llistes
DMurl_sensefiltrar = DM23url_llista + DM22url_llista + DM21url_llista
```

En una primera iteració, s'observa que la llista resultant conté URL que no són de notícies. Hi ha resultats de publicitat, esqueles, PDF annexos a notícies o altres.

```
[ ]: # Filtra els URL per incloure només els que comencin per "https://www.
↳diariodemallorca.es/" i elimina les esqueles
DMurl = [url for url in DMurl_sensefiltrar if url.startswith('https://www.
↳diariodemallorca.es/') and not url.startswith('https://www.
↳diariodemallorca.es/esquelas/')]

# Desa la llista de l'objecte DMurl en un txt
with open('DMurl.txt', 'w') as DMexport:
    DMexport.write('\n'.join(DMurl))
```

2.3 Descarrega el contingut dels URL

Primer, es creen tres funcions personalitzades per suplir les limitacions de la llibreria Newspaper3k.

```
[ ]: # Crea una funció per extreure la secció de l'URL
def extreu_secció_DM(url):
    patro = r'https://www.diariodemallorca.es/([\w-]+)/' # Defineix el patró
↳de l'expressió regular (el text entre barres després de "https://www.
↳diariodemallorca.es/")
    expressio = re.search(patro, url)
    if expressio:
        secció = expressio.group(1).replace('-', ' ').title() # Converteix
↳els guionets en espais en blanc i posa el nom de la secció en majúscules
↳inicials.
        return secció
    return '' # secció en blanc si no s'acomplís el patró

# Crea una funció per extreure el subtítol del codi HTML
def extreu_subtítol_DM(html):
    bs = BeautifulSoup(html, 'html.parser')
    marcador_subtítol = bs.find('h2',
↳class_='headline-article__extended-subtitle')
    if marcador_subtítol:
        return marcador_subtítol.get_text(strip=True)
    return '' # Subtítol en blanc si no troba el marcador

# Crea una funció per extreure l'autoria de la notícia del codi HTML
def extreu_firma_DM(html):
    bs = BeautifulSoup(html, 'html.parser')
    # Opció 1: Retorna el contingut de 'news-author' dins l'etiqueta <a>
    marcador_firma1 = bs.find('a', class_='news-author')
    if marcador_firma1:
        return marcador_firma1.get_text(strip=True)
```

```

# Opció 2: Retorna el contingut de 'news-author' dins l'etiqueta <span>
marcador_firma2 = bs.find('span', class_='news-author')
if marcador_firma2:
    return marcador_firma2.get_text(strip=True)
# Opció 3: Retorna el contingut de masthead__author (per a alguns
↳articles d'opinió)
marcador_firma3 = bs.find('p', class_='masthead__author')
if marcador_firma3:
    firma_a = marcador_firma3.find('a', class_='txt')
    if firma_a:
        return firma_a.get_text(strip=True)
# Deixa la dada en blanc si no troba cap dels marcadors
return ''

```

A continuació, es crea una funció per crear un marc de dades amb la informació de cada notícia. Els camps es recuperen o bé amb Newspaper3k o bé amb les funcions personalitzades.

```

[ ]: # Funció personalitzada per recuperar les dades de cada notícia per a cada
↳URL
def extreu_info_article_DM(url):
    article = Article(url)
    article.download()
    article.parse()
    html = article.html # Per a les funcions que llegeixen del codi HTML,
↳s'obté el contingut HTML de cada article

    dades = {
        'url_notícia': url, # Copia la URL
        'data': article.publish_date, # Funció data de publicació de
↳Newspaper3k
        'secció': extreu_secció_DM(url), # Funció personalitzada
↳extreu_secció
        'títol': article.title, # Funció title de Newspaper3k
        'subtítol': extreu_subtítol_DM(html), # Funció personalitzada
↳extreu_subtítol
        'text': article.text, # Funció text de Newspaper3k
        'firma': extreu_firma_DM(html), # Funció personalitzada extreu_firma
        'url_imatge': article.top_image if article.top_image else None #
↳Recupera la URL de la imatge principal i, si no n'hi ha, deixa el camp en
↳blanc
    }
    return dades

```

Finalment, s'executa el procés per buidar la informació de tots els articles i es converteix el conjunt en un marc de dades.

```

[ ]: # Comprova si ja s'ha fet l'scraping complet
if not os.path.exists('df.csv'):

    # Si el fitxer 'df.csv' no existeix, s'executa el buidatge
    DMbuidatge = []

```

```

for url in DMurl:
    DMbuidatge_dades = extreu_info_article_DM(url)
    DMbuidatge.append(DMbuidatge_dades)

# Crea una taula amb Pandas
DMdf = pd.DataFrame(DMbuidatge)

# Exporta la taula
DMdf.to_csv('DMdf.csv', index=False)
else:
    pass

```

3 Importació de les dades d'*Ultimahora.es*

3.1 Obre el codi font HTML

Es carreguen els altres tres documents HTML i es reanomenen amb el mateix patró.

```

[ ]: with open(r'RAW\UH20230919.htm', 'r', encoding='utf-8') as UH23: # Obre els
      ↪HTML en mode de només lectura
      UH23html = UH23.read()

with open(r'RAW\UH20220920.htm', 'r', encoding='utf-8') as UH22:
    UH22html = UH22.read()

with open(r'RAW\UH20210921.htm', 'r', encoding='utf-8') as UH21:
    UH21html = UH21.read()

```

En aquest cas, els enllaços apareixen als atributs *href* dels objectes *news-heading*.

```

[ ]: bs = BeautifulSoup(UH23html) # Interpreta el codi font amb BeautifulSoup
UH23url = bs.find_all(class_="news-heading") # Filtra els objectes amb un
      ↪objecte de tipus "news-heading"

bs = BeautifulSoup(UH22html)
UH22url = bs.find_all(class_="news-heading")

bs = BeautifulSoup(UH21html)
UH21url = bs.find_all(class_="news-heading")

```

3.2 Crea una llista dels URL

```

[ ]: UH23_llista = []
UH22_llista = []
UH21_llista = []

# Troba els URL de UH23url
for url in UH23url:
    href = url.find('a').get('href')
    UH23_llista.append(href)

```

```

# Troba els URL de UH2url
for url in UH2url:
    href = url.find('a').get('href')
    UH2_llista.append(href)

# Troba els URL de UH1url
for url in UH1url:
    href = url.find('a').get('href')
    UH1_llista.append(href)

# Uneix les tres llistes
UHurl = UH23_llista + UH22_llista + UH21_llista

# Desa la llista de l'objecte UHurl en un txt
with open('UHurl.txt', 'w') as UHexport:
    UHexport.write('\n'.join(UHurl))

```

3.3 Descarrega el contingut dels URL

Es creen les funcions personalitzades per a la secció, el subtítol i l'autoria.

```

[ ]: # Crea una funció per extreure la secció de l'URL
def extreure_secció_UH(url):
    patro = r'https://www\.ultimahora\.es/([~/]+)/'
    expressio = re.search(patro, url)
    if expressio:
        secció = expressio.group(1).replace('-', ' ').title()
        # Comprova si l'URL comença per "https://www.ultimahora.es/noticias/"
        if secció == 'Noticias':
            # Extreure el text entre les següents barres
            if 'https://www.ultimahora.es/noticias/' in url:
                secció2 = re.search(r'\s/noticias\/([~/]+)\/', url)
                if secció2:
                    secció3 = secció2.group(1).replace('-', ' ').title()
                    return secció3
        return secció
    return ''

# Crea una funció per extreure el subtítol del codi HTML
def extreure_subtítol_UH(html):
    bs = BeautifulSoup(html, 'html.parser')
    marcador_subtítol = bs.find('h2', class_='post-subtitle')
    if marcador_subtítol:
        return marcador_subtítol.get_text(strip=True)
    return '' # Subtítol en blanc si no troba el marcador

# Crea una funció per extreure l'autoria de la notícia del codi HTML
def extreure_firma_UH(html):
    bs = BeautifulSoup(html, 'html.parser')
    # Opció 1: Retorna el contingut de 'news-author' dins l'etiqueta <a>
    marcador_firma1 = bs.find('a', class_='post-author')

```

```

if marcador_firma1:
    return marcador_firma1.get_text(strip=True)
# Opció 2: Retorna el contingut de 'news-author' dins l'etiqueta <span>
marcador_firma2 = bs.find('span', class_='post-author')
if marcador_firma2:
    return marcador_firma2.get_text(strip=True)
# Deixa la dada en blanc si no troba cap dels marcadors
return ''

```

Es crea la funció per descarregar els articles i fer-ne el *parsing*.

```

[ ]: # Funció personalitzada per recuperar les dades de cada notícia per a cada
↳URL
def extreu_info_article_UH(url):
    article = Article(url)
    article.download()
    article.parse()
    html = article.html # Per a les funcions que llegeixen del codi HTML,
↳s'obté el contingut HTML de cada article

    dades = {
        'url_notícia': url, # Copia la URL
        'data': article.publish_date, # Funció data de publicació de
↳Newspaper3k
        'secció': extreu_secció_UH(url), # Funció personalitzada
↳extreu_secció
        'títol': article.title, # Funció títol de Newspaper3k
        'subtítol': extreu_subtítol_UH(html), # Funció personalitzada
↳extreu_subtítol
        'text': article.text, # Funció text de Newspaper3k
        'firma': extreu_firma_UH(html), # Funció personalitzada extreu_firma
        'url_imatge': article.top_image if article.top_image else None #
↳Recupera la URL de la imatge principal i, si no n'hi ha, deixa el camp en
↳blanc
    }
    return dades

```

S'executa la recuperació de les dades i es crea el marc de dades 'UHdf'.

```

[ ]: # Comprova si ja s'ha fet l'scraping complet
if not os.path.exists('df.csv'):

    # Si el fitxer 'df.csv' no existeix, s'executa el buidatge
UHbuidatge = []
for url in UHurl:
    UHbuidatge_dades = extreu_info_article_UH(url)
    UHbuidatge.append(UHbuidatge_dades)

    # Crea una taula amb Pandas
UHdf = pd.DataFrame(UHbuidatge)

```

```

# Exporta la taula
UHdf.to_csv('UHdf.csv', index=False)
else:
    pass

```

4 Fusió dels marcs de dades

```

[ ]: # Comprova si ja s'ha fet l'scraping complet
if not os.path.exists('df.csv'):
    # Afegeix una columna amb el mitjà abans de juntar els marcs de dades
    DMdf['mitjà'] = 'DM'
    UHdf['mitjà'] = 'UH'

    # Si no existeix, afegeix les files a un únic 'dataframe'
    df = pd.concat([DMdf, UHdf]).reset_index(drop=True)

    # Exporta el marc de dades resultant
    df.to_csv('df.csv', index=False)

    # Elimina les variables prescindibles que només en creen en aquest 'loop'
    del DMbuidatge, DMbuidatge_dades, DMdf, UHbuidatge, UHbuidatge_dades, U
    ↪UHdf

#Si l'scraping ja s'ha fet, carrega l'arxiu 'df.csv' i elimina les variables
↪prescindibles
else:
    df = pd.read_csv('df.csv')

```

S'eliminen les variables prescindibles.

```

[ ]: del DM21, DM21_llista, DM21html, DM21url, DM21url_llista, DM21url_str
del DM22, DM22_llista, DM22html, DM22url, DM22url_llista, DM22url_str
del DM23, DM23_llista, DM23html, DM23url, DM23url_llista, DM23url_str
del UH21, UH21_llista, UH21html, UH21url
del UH22, UH22_llista, UH22html, UH22url
del UH23, UH23_llista, UH23html, UH23url
del bs, DMexport, DMurl, DMurl_sensefiltrar, href, UHexport, UHurl, url

```

5 Filtratge dels resultats

Es duplica l'objecte *df* per no perdre l'arxiu original. Conté 1.380 files.

```

[ ]: df_canvis = df.copy()

```

Se suprimeixen les files que tenen la data en blanc (són 4 casos i tots ells són publicitat). Queden 1.376 files.

```

[ ]: df_canvis = df_canvis[df_canvis['data'].notnull()]

```

Es descarten els resultats d'agenda (23 casos), ja que majoritàriament són cartells de l'esdeveniment. El marc de dades ara té 1.353 files.


```
[ ]: df_canvis = df_canvis[~df_canvis['url_notícia'].str.startswith('https://www.
↳diariodemallorca.es/ocio/agenda/')]

```

S'ometen les seccions “Fnd” (també es agenda), “Humor”, “Llama la atención” i “Sorteos Loterías”, perquè tampoc no inclouen contingut per a l'anàlisi. Sumen 13 casos. Queden 1.340 resultats.

```
[ ]: seccions_descartades = ['Fnd', 'Humor', 'Llama La Atencion', 'Sorteos_
↳Loterias']
df_canvis = df_canvis[~df_canvis['secció'].isin(seccions_descartades)]
del seccions_descartades

```

Es detecten 12 files que son enquestes per a un partit de futbol. S'eliminen les 11 files dependents i es manté la principal. El marc de dades passa a tenir 1.329 articles.

```
[ ]: # Descarta les files on subtítol i firma són buits i la secció és 'Rcd_
↳Mallorca'
df_canvis = df_canvis[~(df_canvis['subtítol'].isnull() &_
↳(df_canvis['secció'] == 'Rcd Mallorca') & df_canvis['firma'].isnull())]

```

A *Diariodemallorca.es*, hi ha 6 articles que només inclouen resultats de la loteria. S'eliminen. Queden 1.323 elements.

```
[ ]: # Descarta les files on el subtítol sigui una de les dues possibles variants_
↳sobre Bonoloto o Euromillones
df_canvis = df_canvis[df_canvis['subtítol'] != 'Consulta la combinación_
↳ganadora y las estrellas del sorteo de hoy']
df_canvis = df_canvis[df_canvis['subtítol'] != 'Consulta los resultados de_
↳la combinación ganadora de hoy']

```

Es revisen valors anormals a la columna *text* a partir del nombre de caràcters d'aquest camp.

```
[ ]: # Afegeix una columna amb el càlcul de la longitud de cada text
df_canvis['text_longitud'] = df_canvis['text'].str.len().fillna(0)

# Reordena les columnes
df_canvis = df_canvis[['mitjà', 'url_notícia', 'data', 'secció', 'títol',_
↳'subtítol', 'text', 'text_longitud', 'firma', 'url_imatge']]

# Mostra els descriptors de la columna 'text_longitud'
text_longitud_dades = df_canvis['text_longitud'].describe()
print(text_longitud_dades)

```

```
count      1323.000000
mean       2813.966742
std        1910.448665
min         0.000000
25%        1546.000000
50%        2395.000000
75%        3672.500000
max        21692.000000
Name: text_longitud, dtype: float64

```

```
[ ]: # Mostra els 15 valors més baixos i més alts
text_longitud_bottom15 = df_canvis.nsmallest(15,
↳ 'text_longitud')['text_longitud']
print(text_longitud_bottom15)
```

```
28      0.0
159     0.0
674     0.0
388    76.0
614    90.0
645    90.0
15     98.0
144    98.0
153   126.0
62    152.0
68    152.0
414   152.0
78    229.0
83    259.0
511   267.0
Name: text_longitud, dtype: float64
```

```
[ ]: # Mostra els 15 valors més baixos i més alts
text_longitud_top15 = df_canvis.nlargest(15,
↳ 'text_longitud')['text_longitud']
print(text_longitud_top15)
```

```
994   21692.0
630   16504.0
445   12603.0
218   12125.0
56    11226.0
149   10717.0
208   10450.0
213   10191.0
724    9885.0
725    9701.0
723    9663.0
348    9409.0
205    9117.0
352    8858.0
593    8792.0
Name: text_longitud, dtype: float64
```

Després d'una revisió manual, s'observa que l'*scraping* dels valors més alts és correcte. Això no obstant, s'han produït errades en aquells elements que no disposen d'un text (principalment, fotogaleries). Se suprimeixen del conjunt 12 elements el text dels quals és inferior a 200 caràcters, atès que tots ells són problemes amb l'*scraping*. Es mantenen 1.311 resultats.

```
[ ]: # Filtra el marc de dades segons la longitud de 'text_longitud'
df_canvis = df_canvis[df_canvis['text_longitud'] >= 200]
```

```
# Calcula la longitud total dels textos de la mostra
text_longitud_total = df_canvis['text_longitud'].sum()
print(text_longitud_total)

# Suprimeix les variables creades
del text_longitud_dades, text_longitud_bottom15, text_longitud_top15, \
    ↪text_longitud_total
```

3721844.0

En aquest punt, no es detecten més resultats impossibles d'analitzar i s'inicia la fase de preparació de les dades per al seu posterior processament.

6 Transformació de les dades

Els valors NaN a les cel·les buides se substitueixen per *strings* buits. Només succeeix a les columnes *Subtítol* i *Firma*.

```
[ ]: df_canvis['subtítol'] = df_canvis['subtítol'].fillna('')
df_canvis['firma'] = df_canvis['firma'].fillna('')
```

S'observa que hi ha 32 articles de *DiariodeMallorca.es* i 38 d'*Ultimahora.es* sense fotografia. Tot i això, al camp *URL imatge* mostren el logotip del mitjà o la foto de la persona autora. Se substitueixen els valors perquè es mostrin en blanc.

```
[ ]: # Reemplaça un valor a 'url_imatge' amb un 'string' buit
df_canvis.loc[df_canvis['url_imatge'] == 'https://www.diariodemallorca.es/
    ↪images/og-facebook/diariodemallorca.jpg?id=0-195-0', 'url_imatge'] = ''
df_canvis.loc[df_canvis['url_imatge'] == 'https://uh.gsstatic.es/images/
    ↪apple-icon-180x180.png', 'url_imatge'] = ''
df_canvis.loc[df_canvis['url_imatge'].str.startswith('https://uh.gsstatic.es/
    ↪autor/'), 'url_imatge'] = ''
```

Atès que l'hora de publicació no es farà servir per a l'anàlisi, s'elimina aquesta informació de la columna *Data*.

```
[ ]: # Converteix la data a un string
df_canvis['data'] = df_canvis['data'].astype(str)

# Extreu els 10 primers caràcters
df_canvis['data'] = df_canvis['data'].str.slice(0, 10)

# Converteix la data a format de data sense el 'timestamp'
df_canvis['data'] = pd.to_datetime(df_canvis['data']).dt.strftime('%Y-%m-%d')
```

Es modifiquen els valors de la columna *Secció* per traduir-los al català i homogeneïtzar-los entre ambdós mitjans. Cal assenyalar que la transformació de “Violència de género” per “Sucessos” atèn a la categorització que en fa *Diariodemallorca.es*. Aquest mitjà inclou la subsecció “Violència de género” dins la secció “Caso abierto”, que és el seu apartat de crònica negra.

```
[ ]: # Defineix els valors origen i els valors nous
transformacions_secció = {
    'Local': 'Balears',
```

```

'Mallorca': 'Balears',
'Part Forana': 'Part forana',
'Sucesos': 'Successos',
'Violencia De Genero': 'Successos',
'Comunidades': 'Nacional',
'Elecciones 23J': 'Nacional',
'Vivienda En Mallorca': 'Nacional',
'Sociedad': 'Societat',
'Deportes': 'Esports',
'Rcd Mallorca': 'Esports',
'Atletico Balear': 'Esports',
'Buzzeando': 'Virals',
'Virales': 'Virals',
'Salud': 'Salut',
'Opinion': 'Opinió',
'Fotos': 'Fotogalleries / Vídeos',
'Videos': 'Fotogalleries / Vídeos',
'Tecnologia Videojuegos': 'Tecnologia',
'Medio Ambiente': 'Medi ambient',
'Comunicacion Empresarial': 'Publireportatges',
'Comunicados Empresariales': 'Publireportatges',
'Empresas En Mallorca': 'Publireportatges',
'Especiales': 'Publireportatges',
'Buscando Respuestas': 'Reportatges',
'Tendencias21': 'Reportatges'
}

# Executa la substitució dels valors al camp 'secció'
df_canvis['secció'] = df_canvis['secció'].replace(transformacions_secció)

del transformacions_secció

```

Es modifiquen alguns valors de les seccions que agrupen diversos temes per precisar-ne millor el contingut. En aquest cas, els canvis es fan d'acord amb els seus URL.

```

[ ]: # Marca els criteris de substitució d'acord amb l'URL
c1 = df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.
↳es/vida-y-estilo/tecnologia/')
c2 = df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.
↳es/vida-y-estilo/gente/') | df_canvis['url_notícia'].str.
↳startswith('https://www.ultimahora.es/vips/quiosco-rosa/') |
↳df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.es/
↳vida-y-estilo/moda-belleza/') | df_canvis['url_notícia'].str.
↳startswith('https://www.ultimahora.es/xaloc/tendencias/')
c3 = df_canvis['url_notícia'].str.startswith('https://www.ultimahora.es/
↳xaloc/salud-belleza/')
c4 = df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.
↳es/ocio/gastronomia/') | df_canvis['url_notícia'].str.startswith('https://
↳www.ultimahora.es/xaloc/gastronomia/')

```

```

c5 = df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.
↳es/ocio/tv/') | df_canvis['url_notícia'].str.startswith('https://www.
↳diariodemallorca.es/ocio/cine/') | df_canvis['url_notícia'].str.
↳startswith('https://www.ultimahora.es/vips/television/')
c6 = df_canvis['url_notícia'].str.startswith('https://www.ultimahora.es/
↳xaloc/dudas-respuestas/')
c7 = df_canvis['url_notícia'].str.startswith('https://www.ultimahora.es/
↳xaloc/cajon-sastre/') | df_canvis['url_notícia'].str.startswith('https://
↳www.ultimahora.es/xaloc/historia/') | df_canvis['url_notícia'].str.
↳startswith('https://www.diariodemallorca.es/vida-y-estilo/mascotas/') |
↳df_canvis['url_notícia'].str.startswith('https://www.diariodemallorca.es/
↳ocio/2023/09/19/
↳descubriendo-prehistoria-mallorca-poblados-talayoticos-3803475.html')

# Actualitza els valors de 'secció' segons els criteris de substitució
df_canvis.loc[c1, 'secció'] = 'Tecnologia'
df_canvis.loc[c2, 'secció'] = 'Premsa rosa'
df_canvis.loc[c3, 'secció'] = 'Salut'
df_canvis.loc[c4, 'secció'] = 'Gastronomia'
df_canvis.loc[c5, 'secció'] = 'Televisió / Cinema'
df_canvis.loc[c6, 'secció'] = 'Virals'
df_canvis.loc[c7, 'secció'] = 'Miscel·lània'

del c1, c2, c3, c4, c5, c6, c7

```

Es fa un recompte del repartiment de les notícies per secció.

```

[ ]: secció_recompte = df_canvis['secció'].value_counts().reset_index()
secció_recompte.columns = ['Secció', 'Freqüència']
print(secció_recompte)

```

	Secció	Freqüència
0	Nacional	345
1	Balears	174
2	Successos	101
3	Esports	99
4	Societat	74
5	Internacional	72
6	Fotogalleries / Vídeos	56
7	Economia	53
8	Televisió / Cinema	47
9	Cultura	46
10	Opinió	43
11	Part forana	41
12	Palma	29
13	Tecnologia	20
14	Publireportatges	20
15	Premsa rosa	17
16	Reportatges	16
17	Motor	15
18	Medi ambient	12
19	Gastronomia	10

20	Salut	10
21	Virals	7
22	Miscel·lània	4

Per simplificar les agrupacions, les seccions amb una freqüència inferior a 15 s'agrupen a *Miscel·lània*.

```
[ ]: # Elimina les seccions amb menys articles
secció_recompte = df_canvis['secció'].value_counts()
secció_menysde15 = secció_recompte[secció_recompte < 15].index
df_canvis.loc[df_canvis['secció'].isin(secció_menysde15), 'secció'] =
↳ 'Miscel·lània'

# Comprova els canvis
secció_recompte = df_canvis['secció'].value_counts().reset_index()
secció_recompte.columns = ['Secció', 'Freqüència']
print(secció_recompte)

del secció_menysde15, secció_recompte
```

	Secció	Freqüència
0	Nacional	345
1	Balears	174
2	Successos	101
3	Esports	99
4	Societat	74
5	Internacional	72
6	Fotogalleries / Vídeos	56
7	Economia	53
8	Televisió / Cinema	47
9	Cultura	46
10	Opinió	43
11	Miscel·lània	43
12	Part forana	41
13	Palma	29
14	Tecnologia	20
15	Publireportatges	20
16	Premsa rosa	17
17	Reportatges	16
18	Motor	15

Es corregeixen alguns camps de la columna *Firma* amb els següents criteris: - Si hi ha dues persones autores o més, se separen amb una barra inclinada ('/'). - S'elimina el nom de la persona autora a xarxes socials. - S'afegeixen tots els noms propis coneguts dels redactors perquè sigui més fàcil identificar-ne el sexe. - Totes les agències passen a incloure la paraula "Agència" a l'inici del camp. - Si la informació està feta per més d'una agència, es canvia a "Agències". - S'assigna el valor 'Redacció' a totes les firmes que inclouen la paraula "Redacción", sigles mitjançant les quals no es pot identificar l'autoria (p. ex., "R. P. F.") o són d'autoria dubtosa (p. ex., "Guardia Civil").

```
[ ]: # Marca els criteris de substitució dels noms
transformacions_firma = {
    'A. P. F.': 'Redacció',
```

'A. T. B.': 'Redacció',
 'Adrián González | @adri_gonzalez96': 'Adrián González',
 'Agencia ATLAS / EFE': 'Agències',
 'Agencia ATLAS / EP': 'Agències',
 'Agencia ATLAS / Foto: David Castro': 'Agència Atlas',
 'Agencia ATLAS / Foto: EFE': 'Agència Atlas',
 'Agencia ATLAS / Foto: EP': 'Agència Atlas',
 'Agencia ATLAS / Reuters': 'Agències',
 'Agencia ATLAS | EFE': 'Agències',
 'Agencia ATLAS | Foto: David Castro': 'Agència Atlas',
 'Agencia ATLAS | Foto: EFE': 'Agència Atlas',
 'Agencia ATLAS | Foto: Reuters': 'Agència Atlas',
 'Agencia ATLAS': 'Agència Atlas',
 'Agencias/EP': 'Agències',
 'Agencias': 'Agències',
 'Alba Marqués | @albamgalvez': 'Alba Marqués',
 'Àngels Feroselle Paterna': 'Àngels Feroselle Paterna',
 'Antonio \nPapell': 'Antonio Papell',
 'B. Arzayús': 'Bernardo Arzayús',
 'B. Ramon': 'Bartolomé Ramon',
 'C. Capó': 'Cristòfol Capó',
 'Click': 'Pedro Prieto',
 'Cristina Gallardo, Juan José Fernández': 'Cristina Gallardo / Juan José_↵
 ↵Fernández',
 'D.B.': 'Redacció',
 'Doctor Rodríguez-Mañas': 'Leocadio Rodríguez-Mañas',
 'E.P.': 'Agència Europa Press',
 'Editorial': 'Redacció',
 'Eduardo Costas Catedrático de Genética y Académico': 'Eduardo Costas',
 'efe. palma': 'Agència Efe',
 'EFE': 'Agència Efe',
 'Efe': 'Agència Efe',
 'Elsotanoperdido.com': 'Redacció',
 'Elsotanoperdido': 'Redacció',
 'Elsótanoperdido': 'Redacció',
 'EP': 'Agència Europa Press',
 'Europa Press': 'Agència Europa Press',
 'G. Bosch': 'Guillem Bosch',
 'Guardia Civil': 'Agències',
 'Guillem Sánchez, Jordi Ribalaygue': 'Guillem Sánchez / Jordi_↵
 ↵Ribalaygue',
 'I. Durán': 'Isabel Durán',
 'I. G. / T. S.': 'Redacció',
 'I. G./ T. S.': 'Redacció',
 'I. Olaizola': 'Iñaki Olaizola',
 'Información.es': 'Redacció',
 'INFORMACIÓN': 'Redacció',
 'Iván Carsí | Agustín Perales': 'Iván Carsí / Agustín Perales',
 'J. G. Albalat': 'Jesús G. Albalat',
 'J. L. F./Redacción': 'Joan Lluís Ferrer / Redacció',

'J. L. Ferrer/Redacci3n': 'Joan Llu3s Ferrer / Redacci3n',
'J. L. Ferrer': 'Joan Llu3s Ferrer',
'J. L. Ruiz Collado': 'Juan Luis Ruiz Collado',
'J. M. Arbucias': 'Josep Miquel Arbucias',
'J.E.': 'Redacci3n',
'J.F. Mestre': 'Juan Francisco Mestre',
'J.S.': 'Redacci3n',
'Jaume A.': 'Redacci3n',
'JM Vigarar': 'Jos3 Miguel Vigarar',
'Jorge Napal | Grupo Noticias': 'Jorge Napal',
'Jose Rico | Silvia Martinez': 'Jose Rico / Silvia Mart3nez',
'Juan P. Mart3nez': 'Juan Pedro Mart3nez',
'La Provincia': 'Redacci3n',
'Laura Ballester | @LaBallester': 'Laura Ballester',
'Levante-EMV': 'Redacci3n',
'Luis Vaello (Binance)': 'Luis Vaello',
'Ll. Garcia': 'Lluc Garcia',
'M. Elena Vall3s': 'Mar3a Elena Vall3s',
'M. Mielniezuk': 'Manuel Mielniezuk',
'M. R.': 'Redacci3n',
'M. V.': 'Mat3as Vall3s',
'M.D. Cuenca': 'Maria Dolors Cuenca',
'Mario Saavedra, David Page': 'Mario Saavedra / David Page',
'Meritxell M. Paun3 | Jordi Ribalaygue': 'Meritxell M. Paun3 / Jordi
Ribalaygue',
'Miserol': 'Joan Ferrer Miserol',
'N.C.B.': 'Redacci3n',
'N+1/T21': 'Redacci3n',
'Neomotor': 'Redacci3n',
'P. Bover': 'Ponç Bover',
'P. Cerrada': 'Pedro Cerrada',
'P3rez Gil': 'Juan Carlos P3rez Gil',
'PI Studio': 'Redacci3n',
'Pi Studio': 'Redacci3n',
'Polic3a Local Alc3dia / DM': 'Redacci3n',
'R. D.': 'Redacci3n',
'R. P. F.': 'Redacci3n',
'R. Prieto': 'Rosa Prieto',
'R. S.': 'Redacci3n',
'R.C.': 'Redacci3n',
'Rafa L3pez | @Garcio72': 'Rafa L3pez',
'Rafa Sardiña': 'Rafael Sardiña',
'Ra3l Rodr3guez | Utrech, Holanda': 'Ra3l Rodr3guez',
'Redacci3n Deportes': 'Redacci3n',
'Redacci3n Digital': 'Redacci3n',
'Redacci3n Especiales': 'Redacci3n',
'Redacci3n Local': 'Redacci3n',
'Redacci3n Menorca': 'Redacci3n',
'Redacci3n Monogr3ficos': 'Redacci3n',
'Redacci3n Part Forana': 'Redacci3n',
'Redacci3n T21': 'Redacci3n',


```

'Redacció YoTele': 'Redacció',
'Redacció Yotele': 'Redacció',
'Redacció': 'Redacció',
'RedaccióT21': 'Redacció',
'Reuters': 'Agència Reuters',
'Sara Ledo/EFE': 'Sara Ledo / Agència Efe',
'Servimedia': 'Agència Servimedia',
'Silvia Martinez': 'Silvia Martínez',
'Torres Blasco': 'Juan Torres Blasco',
'VERÓNICA SEDEÑO / AGENCIAS': 'Verónica Sedeño / Agències',
'Vídeo: Agencia ATLAS | Foto: EFE': 'Agència Atlas',
'Vídeo: Agencia Atlas | Foto: EP': 'Agència Atlas',
'Vídeo: Agencia ATLAS | Foto: EP': 'Agència Atlas',
'Vídeo: AGENCIA ATLAS Foto: Agencias': 'Agència Atlas',
'Vídeo: EFE | Foto: EP': 'Agència Efe',
'Vídeo: La Provincia | Foto: EP': 'Redacció'
}

df_canvis['firma'] = df_canvis['firma'].replace(transformacions_firma)

del transformacions_firma

```

Després de l'anàlisi de les autories, es detecten alguns casos d'articles d'opinió que estan agrupats a la secció del tema que tracten en comptes de figurar amb el valor *Opinió*. Mitjançant el nom, es reassignen a la secció *Opinió* aquells casos en els quals la persona autora sempre és columnista. També es reassignen com a *Opinió* els articles el títol dels quals comença per “Opinió” o “Opinión”.

```

[ ]: # Marca els criteris de substitució segons l'autoria
c8 = (df_canvis['firma'] == 'Juan Mestre') | (df_canvis['firma'] == 'Ferran_
↳ Monegal') | (df_canvis['firma'] == 'Augusto Soyer') | (df_canvis['firma']_
↳ == 'Àngels Fermoselle Paterna') | (df_canvis['firma'] == 'Josep Miquel_
↳ Arbucias') | (df_canvis['firma'] == 'Llorenç Riera')
c9 = df_canvis['títol'].str.startswith('Opinió')

# Actualitza els valors de 'secció' segons els criteris de substitució
df_canvis.loc[c8, 'secció'] = 'Opinió'
df_canvis.loc[c9, 'secció'] = 'Opinió'

del c8, c9

```

Hi ha un cas excepcional d'una fotogaleria a la secció *Balears*.

```

[ ]: c10 = df_canvis['url_notícia'] == 'https://www.diariodemallorca.es/mallorca/
↳ 2023/09/19/menorca-talayotica-patrimonio-mundial-92283626.html'
df_canvis.loc[c10, 'secció'] = 'Fotogaleries / Vídeos'

del c10

```

Es modifica el valor d'una notícia sense firma. Se li assigna el valor “Redacció”.

```
[ ]: c11 = df_canvis['url_notícia'] == 'https://www.diariodemallorca.es/
↳empresas-en-mallorca/2021/09/21/
↳fundacion-asima-asprom-colaboran-integracion-57509566.html'
df_canvis.loc[c11, 'firma'] = 'Redacció'

del c11
```

Una vegada fets els darrers canvis de secció, es tornen a calcular les freqüències de cada secció.

```
[ ]: secció_recompte = df_canvis['secció'].value_counts().reset_index()
secció_recompte.columns = ['Secció', 'Freqüència']
print(secció_recompte)

del secció_recompte
```

	Secció	Freqüència
0	Nacional	345
1	Balears	172
2	Successos	101
3	Esports	98
4	Societat	74
5	Internacional	72
6	Fotogalleries / Videos	57
7	Opinió	55
8	Economia	53
9	Televisió / Cinema	46
10	Cultura	46
11	Miscel·lània	42
12	Part forana	37
13	Palma	26
14	Tecnologia	20
15	Publireportatges	20
16	Premsa rosa	17
17	Reportatges	16
18	Motor	14

Com a preparació de l'anàlisi automatitzada, els valors del camp *Firma* se separen en dues columnes (atès que dos és el nombre màxim de firmes per article). Es creen les columnes *Firma1* i *Firma2*.

```
[ ]: # Divideix 'firma' en dues columnes
df_canvis[['firma1', 'firma2']] = df_canvis['firma'].str.split(' / ', 1,
↳expand=True)

# Umpl els valors NaN amb un 'string' buit
df_canvis['firma1'] = df_canvis['firma1'].fillna('')
df_canvis['firma2'] = df_canvis['firma2'].fillna('')
```

Finalment, es reordenen les columnes del marc de dades i es calculen dues estadístiques sobre la longitud del conjunt de dades que influeixen en l'anàlisi automatitzada de textos.

```
[ ]: # Reordena les columnes
df_canvis = df_canvis[['url_notícia', 'url_imatge', 'mitjà', 'data',
↳ 'secció', 'títol', 'subtítol', 'text', 'text_longitud', 'firma',
↳ 'firma1', 'firma2']].reset_index(drop=True)

# Calcula el subtotal de caràcters analitzables
caràcters_títol = df_canvis['títol'].str.len().sum()
caràcters_subtítol = df_canvis['subtítol'].str.len().sum()
caràcters_text = df_canvis['text'].str.len().sum()
caràcters_firma = df_canvis['firma'].str.len().sum()

# Calcula el subtotal de paraules analitzables
paraules_títol = df_canvis['títol'].str.split().apply(len).sum()
paraules_subtítol = df_canvis['subtítol'].str.split().apply(len).sum()
paraules_text = df_canvis['text'].str.split().apply(len).sum()
paraules_firma = df_canvis['firma'].str.split().apply(len).sum()

# Mostra la suma total de caràcters i paraules
print(f"Total de caràcters: {caràcters_títol + caràcters_subtítol +
↳ caràcters_text + caràcters_firma}")
print(f"Total de paraules: {paraules_títol + paraules_subtítol +
↳ paraules_text + paraules_firma}")

# Elimina les variables ja utilitzades
del caràcters_firma, caràcters_subtítol, caràcters_text, caràcters_títol
del paraules_firma, paraules_subtítol, paraules_text, paraules_títol

# Exporta el marc de dades a CSV
df_canvis.to_csv('df_canvis.csv', index=False)
```

Total de caràcters: 3952419

Total de paraules: 656549

7 Anàlisi automatitzada de les dades

7.1 Determinació de les variables d'estudi

Les variables quantitatives que s'analitzaran en aquest punt són: - Dins del marc de dades de la notícia: - Sexe de les persones autors dels articles. - Dins d'un nou marc de dades vinculat al primer, relatiu a cada persona esmentada a cada notícia: - Nom de la persona. - Sexe. - Edat, si s'esmenta a la notícia. - Ocupació, si s'esmenta a la notícia. - Rol familiar, si s'esmenta a la notícia. - Si la persona és presentada com a: - Víctima o supervivent. - Protagonista. - Portaveu. - Veu experta. - Testimoni. - Si la persona apareix a: - Títol. - Subtítol. - Text. - Quantes vegades s'esmenta la persona en el conjunt de l'article.

- A partir de la informació d'aquest nou marc de dades:
 - Número de persones que apareixen al títol per sexe.
 - Número de persones que apareixen al subtítol per sexe.
 - Número de persones que apareixen al text per sexe.

Els camps d'anàlisi estan basats en la guia ràpida de monitoratge complet del *Global Media Monitoring Project*, consultable a <https://whomakesthenews.org/wp->

content/uploads/2021/03/GMMP-2020-Monitoreo-completo-Guia-rapida-sobre-la-codificacion.pdf.

7.2 Tractament preliminar dels marcs de dades

Per tal de mantenir intacte l'objecte `df_canvis` i tenir una traçabilitat dels canvis, es crea un nou duplicat del conjunt. En aquest apartat, el marc de dades s'anomenarà `df_notícies`.

```
[ ]: df_notícies = df_canvis.copy()
```

En el següent fragment de codi, es creen noves columnes al marc de dades ja existent per a les dades de les quals s'automatitzarà la recollida.

```
[ ]: #Crea les noves columnes
df_notícies['firma1_dona'] = ''
df_notícies['firma1_home'] = ''
df_notícies['firma1_indeterminat'] = ''

df_notícies['firma2_dona'] = ''
df_notícies['firma2_home'] = ''
df_notícies['firma2_indeterminat'] = ''

# Crea una columna amb un ID correlatiu per a cada notícia
df_notícies['id'] = df_notícies.index

# Reordena les columnes
df_notícies = df_notícies[['id', 'url_notícia', 'url_imatge', 'mitjà',
↳'data', 'secció', 'títol', 'subtítol', 'text', 'text_longitud', 'firma',
↳'firma1', 'firma1_dona', 'firma1_home',
↳'firma1_indeterminat',
↳'firma2', 'firma2_dona', 'firma2_home',
↳'firma2_indeterminat']].reset_index(drop=True)
```

7.3 Anàlisi de les firmes

Per tal de provar la viabilitat d'automatitzar l'anàlisi quantitativa, es fa una prova amb els camps de firma. Per estalviar temps de processament i costos, es crea una llista única amb els diferents valors de `firma1` i `firma2`.

```
[ ]: # Primer, comprova si 'df_firmes' no existeix
if 'df_firmes' not in locals():

    # Crea la llista única
    df_firmes = pd.concat([df_notícies['firma1'], df_notícies['firma2']]).
↳unique()
    df_firmes = pd.DataFrame(df_firmes, columns=['firma']).
↳sort_values(by='firma').reset_index(drop=True)

    # Estableix el sexe *unknown* a certs valors
    sexe_desconegut = ['Agència', 'Agències', 'Redacció']
```

```

df_firmes['sexe'] = df_firmes['firma'].apply(lambda name: 'indeterminat'
↳if not name or any(keyword.lower() in str(name).lower() for keyword in
↳sexe_desconegut) else '')

del sexe_desconegut

```

Per raons de seguretat, la contrasenya per utilitzar l'API s'emmagatzema a un arxiu diferent que es carrega en aquesta part del codi.

```

[ ]: # Obre l'arxiu json que conté la clau de l'API a un element anomenat 'openai'
def extreu_clau_api(filepath='api-key.json'):
    with open(filepath, 'r') as contrasenya:
        codi = json.load(contrasenya)
        return codi.get('openai')

# Carrega la clau personalitzada de l'API
openai.api_key = extreu_clau_api()

```

Es crea la funció que processa cada firma mitjançant l'API. S'envia un *prompt* en anglès al model en el qual se li demana determinar el sexe del nom i respondre amb una sola paraula, ja que després de diverses proves ha estat la fórmula que ha obtingut una taxa d'èxit més alta. Aquesta paraula s'associa amb un conjunt tancat de resultats, determinats amb una funció *if*. La pregunta (*prompt*) es fa en anglès, atès que a les proves preliminars s'han obtingut millors resultats.

```

[ ]: # Crea una funció per analitzar els valors únics de firma
for index, row in df_firmes.iterrows():
    nom = row['firma']

    # Evita les cel·les que no estan en blanc
    if row['sexe'] != '':
        continue

    # Determina la petició i el context que s'envia al model de GPT
    try:
        resposta = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=[
                {"role": "system",
                 "content": "Determine the gender of the following Spanish
↳or Catalan names."
                 "Answer only with 'female' or 'male'. If you think there is
↳a doubt, between 'female' and 'male' answer with the most likely option.
↳"},
                {"role": "user", "content": f"What gender does the name
↳'{nom}' suggest?"}
            ],
            max_tokens=60
        )

    # Crea un objecte de resultat per ser analitzat

```

```

resultat = resposta.choices[0].message["content"].lower()

# Mostra el llistat d'assignacions de sexe
print(f"{nom}': {resultat}")

# Assigna un valor a cada resposta del model
sexe_firma = 'dona' if "female" in resultat else 'home' if "male" in
↳resultat else ''
df_firmes.at[index, 'sexe'] = sexe_firma

# Guarda cada iteració a un CSV per no perdre el progrés
df_firmes.to_csv("df_firmes.csv", index=False)

# Mostra un error en cas d'haver-n'hi i deixa la cel·la en blanc
except Exception as e:
    print(f"Error amb el nom '{nom}': {e}")
    df_firmes.at[index, 'sexe'] = ''

# Elimina les variables prescindibles si existeixen
for variable in ['index', 'nom', 'resposta', 'resultat', 'row',
↳'sexe_firma']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

El cost total del processament automatitzat de les firmes, incloent-ne les proves, ha estat de 0,09 dòlars (aproximadament, 0,08 euros). Finalitzada l'assignació de sexe a cada nom, es passen els resultats al marc de dades `df_notícies` amb una funció condicional.

```

[ ]: # Itera per cada fila de 'df_notícies'
for index, row in df_notícies.iterrows():

    # Regles d'assignació de valors per a 'firma1'
    firma1_nom = row['firma1']
    if firma1_nom in df_firmes['firma'].values:
        sexe = df_firmes.loc[df_firmes['firma'] == firma1_nom, 'sexe'].
↳values[0]
        if sexe == 'dona':
            df_notícies.at[index, 'firma1_dona'] = 1
            df_notícies.at[index, 'firma1_home'] = 0
            df_notícies.at[index, 'firma1_indeterminat'] = 0
        elif sexe == 'home':
            df_notícies.at[index, 'firma1_dona'] = 0
            df_notícies.at[index, 'firma1_home'] = 1
            df_notícies.at[index, 'firma1_indeterminat'] = 0
        elif sexe == 'indeterminat':
            df_notícies.at[index, 'firma1_dona'] = 0
            df_notícies.at[index, 'firma1_home'] = 0
            df_notícies.at[index, 'firma1_indeterminat'] = 1

```

```

# Regles d'assignació de valors per a 'firma1'
firma2_nom = row['firma2']
if pd.isna(firma2_nom) or firma2_nom == '': # S'introdueix una excepció
↳si la cel·la de 'firma2' és buida per no recomptar firmes que no
↳existeixen
    continue
elif firma2_nom in df_firmes['firma'].values:
    sexe = df_firmes.loc[df_firmes['firma'] == firma2_nom, 'sexe'].
↳values[0]
    if sexe == 'dona':
        df_notícies.at[index, 'firma2_dona'] = 1
        df_notícies.at[index, 'firma2_home'] = 0
        df_notícies.at[index, 'firma2_indeterminat'] = 0
    elif sexe == 'home':
        df_notícies.at[index, 'firma2_dona'] = 0
        df_notícies.at[index, 'firma2_home'] = 1
        df_notícies.at[index, 'firma2_indeterminat'] = 0
    elif sexe == 'indeterminat':
        df_notícies.at[index, 'firma2_dona'] = 0
        df_notícies.at[index, 'firma2_home'] = 0
        df_notícies.at[index, 'firma2_indeterminat'] = 1

del index, row, firma1_nom, firma2_nom, sexe

```

```

[ ]: # Exporta el marc de dades a CSV
df_notícies.to_csv('df_notícies.csv', index=False)

```

7.4 Anàlisi de títols, subtítols i textos

Es crea un log per emmagatzemar el resultat de l'execució del processament automatitzat de dades.

```

[ ]: data_hora = datetime.datetime.now().strftime("%Y%m%d_%H%M%S")
nom_log = f'log_df_notícies_{data_hora}.txt'

logging.basicConfig(filename=nom_log, level=logging.INFO,
                    format='%(message)s')

```

Es crea una funció per identificar les persones i processar les dades que s'extreuen de `df_notícies`.

```

[ ]: def identifica_persones(text_complet, id_fila):
    try:
        resposta = openai.ChatCompletion.create(
            model="gpt-4-1106-preview",
            messages=[
                {"role": "system",
                 "content": "You're an AI trained to extract human data from
↳texts. Exclude places, dates, organizations, initials, or non-person
↳names."

```

```

        "Extract all person names mentioned in the following news_
↳article. The text is in Spanish or Catalan."
        "Then, identify the following information only if stated in_
↳the text in this order: [1] name, [2] age, [3] job or profession (in only_
↳one or two words, in English, excluding family roles), [4] family role (e.
↳g., wife, husband, mother, parent...), [5] if the person is the protagonist_
↳of the article (1 for yes, 0 for no), [6] if they are identified as a_
↳spokesperson (1=yes, 0=no), [7] if they are identified as an expert_
↳(1=yes, 0=no), [8] if they are identified as a witness (1=yes, 0=no), and_
↳[9] if they are identified as a victim or survivor (1=yes, 0=no)."
```

Every reply has to include always those nine elements per_
↳person separated with commas (even though they are blank) and every_
↳person will be separated with semicolons, so the information follows a_
↳regular expression (e.g., Person1, 25, actress, , 1, 0, 0, 0, 0; Person2,_
↳, politician, , 0, 1, 0, 0, 0)."},

```

        {"role": "user",
         "content": text_complet}
    ],
    max_tokens=1000
)

# Extreu i mostra el text de la resposta
resposta_gpt = resposta.choices[0].message['content']
print(f"{id_fila}: ", resposta_gpt)
logging.info(f"{id_fila}: {resposta_gpt}")

# Elimina signes de puntuació finals a la resposta, en cas que n'hi_
↳hagi
if resposta_gpt.endswith(';'):
    resposta_gpt = resposta_gpt[:-1]
if resposta_gpt.endswith('.'):
    resposta_gpt = resposta_gpt[:-1]

# Extreu els noms separats per punts i coma
noms = resposta_gpt.split('; ')
noms = [nom.strip() for nom in noms if nom.strip()]

return noms

# Missatge d'error si la crida a l'API no funciona
except Exception as e:
    print(f"Error: {e}")
    logging.error(f"Error: {e}")
    return []

```

Es crea un nou marc de dades on es desarà la informació extreta per l'API. La nova taula s'anomena *df_persones* i conté les columnes necessàries per explotar la informació estadística-ment i emmagatzemar per separat les respostes del model d'intel·ligència artificial. La informació es processa d'acord amb el format establert per a la resposta de l'API. Si la resposta no té el format establert, es marca com a error. Els ID dels articles vàidament processats

s'emmagatzemen a un fitxer `.txt` per tal d'evitar-ne el reprocessament (i els costos associats) en cas que hi hagi un problema amb l'API.

```
[ ]: # Comprova si existeix l'arxiu 'df_persones.csv' i el carrega si ja existeix
if os.path.exists('df_persones.csv'):
    df_persones = pd.read_csv('df_persones.csv')
    df_persones = df_persones.fillna('')
else:
    # Si no existeix, crea el marc de dades 'df_persones'
    df_persones = pd.DataFrame(columns=['id', 'núm', 'nom', 'edat', 'ocupació', 'rol_familiar',
                                       'protagonista', 'portaveu', 'experta', 'testimoni', 'víctima',
                                       'dins_títol', 'dins_subtítol', 'dins_text', 'coincidències'])

# Es crea una variable que recompta les notícies processades
notícies_processades = set()

# Carrega els ID de les notícies processades
if os.path.exists('notícies_processades.txt'):
    with open('notícies_processades.txt', 'r') as file:
        for line in file:
            notícies_processades.add(line.strip())

# Itera sobre el marc de dades amb les notícies
for index, row in df_notícies.iterrows():
    id_fila = str(row['id'])
    text_complet = f>Title: {row['títol']} \n\nSubtitle: {row['subtítol']} \n\nText: {row['text']}"

    # Comprova si l'article ja s'ha processat
    if id_fila in notícies_processades:
        continue

    persones_info = identifica_persones(text_complet, id_fila)
    article_vàlid = True # Marca cada processament com a vàlid per defecte

    # Processa cada persona trobada a la notícia
    for núm, info_persona in enumerate(persones_info, start=1):

        # Afegeix un espai en blanc per si l'API retorna dues comes seguides
        info_persona = info_persona.replace(',,', ',, ')

        # Divideix la informació de la persona en components individuals
        components = info_persona.split(' ')

        # Comprova que hi ha 9 components a cada resposta i que els cinc
        # darrers valors són binaris
        if len(components) != 9 or not all(c in ['0', '1'] for c in components[4:]):
```

```

print(f"Error: {id_fila}")
logging.error(f"Error: {id_fila}")
article_vàlid = False
break # Si no s'acompleix la condició, atura el processament de
↳l'article

# Crea un diccionari amb els components de la resposta
nom = components[0]
edat = components[1]
ocupació = components[2]
rol_familiar = components[3]
protagonista = components[4]
portaveu = components[5]
experta = components[6]
testimoni = components[7]
víctima = components[8]

# Determina a quina part del text està present cada nom
dins_títol = int(nom.lower() in row['títol'].lower())
dins_subtítol = int(nom.lower() in row['subtítol'].lower())
dins_text = int(nom.lower() in row['text'].lower())

# Recompta la freqüència d'aparició d'un nom al conjunt de la notícia
coincidències = text_complet.lower().count(nom.lower())

# Formata la informació per ser tabulada
persones_trobades = pd.DataFrame([
    'id': row['id'],
    'núm': núm,
    'nom': nom,
    'edat': edat,
    'ocupació': ocupació,
    'rol_familiar': rol_familiar,
    'víctima': víctima,
    'protagonista': protagonista,
    'portaveu': portaveu,
    'experta': experta,
    'testimoni': testimoni,
    'dins_títol': dins_títol,
    'dins_subtítol': dins_subtítol,
    'dins_text': dins_text,
    'coincidències': coincidències
])

# Afegeix la informació al marc de dades 'df_persones'
df_persones = pd.concat([df_persones, persones_trobades],
↳ignore_index=True)

# Comprova si el processament de l'article està marcat com a vàlid
if article_vàlid:

```

```

        # Afegeix l'ID de la notícia a una llista de notícies processades
        ↳ només si l'article és vàlid
        notícies_processades.add(id_fila)
        # Desa els ID de les notícies processades després de cada
        ↳ processament d'article vàlid
        with open('notícies_processades.txt', 'w') as file:
            for pid in notícies_processades:
                file.write(f"{pid}\n")

# Exporta el marc de dades a CSV
df_persones.to_csv('df_persones.csv', index=False)

```

Finalment, s'exporta el nou marc de dades i es fa una neteja de totes les variables creades durant els processos.

```

[ ]: # Exporta el marc de dades a CSV
df_persones.to_csv('df_persones.csv', index=False)

# Elimina variables prescindibles
for variable in ['article_vàlid', 'coincidències', 'components',
↳ 'dins_títol', 'dins_subtítol', 'dins_text', 'edat', 'experta', 'file',
↳ 'id_fila', 'index', 'info_persona', 'line', 'nom',
↳ 'núm', 'ocupació', 'persones_info', 'persones_trobades', 'pid',
↳ 'portaveu', 'protagonista', 'rol_familiar', 'row',
↳ 'testimoni', 'text_complet', 'víctima']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

L'API ha trigat 75 minuts en processar les 1.311 notícies. El cost total de les proves ha estat de 2,69 dòlars (uns 2,52 euros), mentre que el processament final dels textos amb el model GPT-4-Turbo ha suposat un cost de 18,49 dòlars (17,29 euros).

7.4.1 Processament dels errors en l'anàlisi

Es processa el log per conèixer quants errors s'han produït.

```

[ ]: # Recompta els errors
recompte_errors = 0

# Carrega el log
with open('log_df_notícies_20231110_211043.txt', 'r', encoding='utf-8') as
↳ log:
    for line in log:
        if "Error" in line:
            recompte_errors += 1

# Mostra el recompte
print(f"Total d'errors: {recompte_errors}")

```

Total d'errors: 421

S'han processat vàlidament 890 notícies i han tingut lloc 421 errors. El marc de dades `df_errors` llista les respostes amb errors de l'API.

```
[ ]: # Crea llistes per emmagatzemar les dades del nou marc de dades
errors = []
id_filas = None
text_filas = ""

# Carrega el log
with open('log_df_notícies_20231110_211043.txt', 'r', encoding='utf-8') as log:
    for line in log:
        id_error = re.search(r'(\d+):', line)

        if id_error:
            # Cerca l'ID de cada error
            id_filas = id_error.group(1)
            text_filas = line.strip()

        elif "Error" in line and id_filas:
            # Calcula la longitud del missatge previ a l'error
            longitud_error = len(text_filas)

            # Afegeix les dades a la llista d'errors
            errors.append([id_filas, text_filas, longitud_error])

            # Reinicia per al següent error
            id_filas = None
            text_filas = ""

# Copia les dades al marc de dades 'df_errors' i exporta'l a csv
df_errors = pd.DataFrame(errors, columns=['id', 'text_resposta_API',
    'longitud_error'])
df_errors.to_csv('df_errors.csv', index=False)

del errors, id_error, id_filas, line, log, longitud_error, text_filas
```

El 60 % d'aquests errors (254 dels 421) han estat perquè no s'havia previst una excepció per a aquells textos on hi ha cap nom propi, la qual cosa es detecta i corregeix amb el fragment de codi següent.

```
[ ]: # Crea una funció per comprovar els articles que no inclouen noms propis
def comprova_sensenoms(row):
    frases_sensenom = ["no person names", "no names", "no personal names",
    "no se mencionan nombres",
        "ningún nombre", "specific personal names", "specific
    person names", "no hay nombres",
        "ninguna persona", "no person's names", "no explicit
    personal names", "no human names"]
```

```

    # Si hi ha cap de les frases citades o si la resposta té menys de 10
    ↪ caràcters, marca com a article sense noms
    if any(phrase in row['text_resposta_API'].lower() for phrase in
    ↪ frases_sensenom) or row['longitud_error'] < 10:
        return 1
    else:
        return 0

# Crea una columna a 'df_errors' per marcar els articles sense noms
df_errors['sense_noms'] = df_errors.apply(comprova_sensenoms, axis=1)

# Calcula el total d'articles sense noms
total_sense_noms = df_errors['sense_noms'].sum()
print(f"Articles sense noms: {total_sense_noms}")

del total_sense_noms

```

Articles sense noms: 254

Els ID de les 254 notícies que no contenen noms propis s'afegeixen a la llista de notícies processades.

```

[ ]: # Cerca totes les files de 'df_errors' marcades com que no inclouen noms
for index, row in df_errors.iterrows():
    if row['sense_noms'] == 1:
        # Actualitza el conjunt 'notícies_processades'
        if row['id'] not in notícies_processades:
            notícies_processades.add(row['id'])

# Actualitza 'notícies_processades.txt'
with open('notícies_processades.txt', 'w', encoding='utf-8') as file:
    for pid in notícies_processades:
        file.write(f"{pid}\n")

# Elimina les files de 'df_errors' afegides a 'notícies_processades'
#df_errors = df_errors[~df_errors['id'].isin(notícies_processades)]

del file, index, pid, row

```

7.4.2 Reprocessament dels articles

Queden 167 articles per reprocessar d'acord amb el nombre de files de `df_errors`. Es repeteix la mateixa instrucció a l'API amb el propòsit de reduir la xifra d'errors.

```

[ ]: # Itera sobre el marc de dades amb les notícies
for index, row in df_notícies.iterrows():
    id_fil_a = str(row['id'])
    text_complet = f"Title: {row['títol']} \n\nSubtitle: {row['subtítol']}\n
    ↪ \n\nText: {row['text']}"

    # Comprova si l'article ja s'ha processat
    if id_fil_a in notícies_processades:

```

```

        continue

persones_info = identifica_persones(text_complet, id_fila)
article_vàlid = True # Marca cada processament com a vàlid per defecte

# Processa cada persona trobada a la notícia
for núm, info_persona in enumerate(persones_info, start=1):

    # Afegeix un espai en blanc per si l'API retorna dues comes seguides
    info_persona = info_persona.replace(',,', ', ')

    # Divideix la informació de la persona en components individuals
    components = info_persona.split(' ')

    # Comprova que hi ha 9 components a cada resposta i que els cinc
↳darrers valors són binaris
    if len(components) != 9 or not all(c in ['0', '1'] for c in
↳components[4:]):
        print(f"Error: {id_fila}")
        logging.error(f"Error: {id_fila}")
        article_vàlid = False
        break # Si no s'acompleix la condició, atura el processament de
↳l'article

    # Crea un diccionari amb els components de la resposta
    nom = components[0]
    edat = components[1]
    ocupació = components[2]
    rol_familiar = components[3]
    protagonista = components[4]
    portaveu = components[5]
    experta = components[6]
    testimoni = components[7]
    víctima = components[8]

    # Determina a quina part del text està present cada nom
    dins_títol = int(nom.lower() in row['títol'].lower())
    dins_subtítol = int(nom.lower() in row['subtítol'].lower())
    dins_text = int(nom.lower() in row['text'].lower())

    # Recompta la freqüència d'aparició d'un nom al conjunt de la notícia
    coincidències = text_complet.lower().count(nom.lower())

    # Formata la informació per ser tabulada
    persones_trobades = pd.DataFrame([
        'id': row['id'],
        'núm': núm,
        'nom': nom,
        'edat': edat,
        'ocupació': ocupació,

```

```

        'rol_familiar': rol_familiar,
        'víctima': víctima,
        'protagonista': protagonista,
        'portaveu': portaveu,
        'experta': experta,
        'testimoni': testimoni,
        'dins_títol': dins_títol,
        'dins_subtítol': dins_subtítol,
        'dins_text': dins_text,
        'coincidències': coincidències
    })

    # Afegeix la informació al marc de dades 'df_persones'
    df_persones = pd.concat([df_persones, persones_trobades],
↳ ignore_index=True)

    # Comprova si el processament de l'article està marcat com a vàlid
    if article_vàlid:
        # Afegeix l'ID de la notícia a una llista de notícies processades
↳ només si l'article és vàlid
        notícies_processades.add(id_fila)
        # Desa els ID de les notícies processades després de cada
↳ processament d'article vàlid
        with open('notícies_processades.txt', 'w') as file:
            for pid in notícies_processades:
                file.write(f"{pid}\n")

    # Exporta el marc de dades a CSV
    df_persones.to_csv('df_persones.csv', index=False)

```

El reprocessament ha trigat 17 minuts i mig amb un cost de 3,19 dòlars (aproximadament, 2,98 euros). Es processa un altre cop el log per conèixer quants errors s'han produït.

```

[ ]: # Recompta els errors
recompte_errors = 0

# Carrega el log
with open('log_df_notícies_20231111_130409.txt', 'r', encoding='iso-8859-1')
↳ as log:
    for line in log:
        if "Error" in line:
            recompte_errors += 1

# Mostra el recompte
print(f"Total d'errors: {recompte_errors}")

```

Total d'errors: 62

El total de de notícies vàlides puja a 1.249 i queden 62 errors per tractar. El marc de dades `df_errors2` llista les respostes amb la segona tongada d'errors de l'API.

```
[ ]: # Crea llistes per emmagatzemar les dades del nou marc de dades
errors2 = []
id_fila = None
text_fila = ""

# Carrega el log
with open('log_df_notícies_20231111_130409.txt', 'r', encoding='iso-8859-1') as log:
    for line in log:
        id_error2 = re.search(r'(\d+):', line)

        if id_error2:
            # Cerca l'ID de cada error
            id_fila = id_error2.group(1)
            text_fila = line.strip()

        elif "Error" in line and id_fila:
            # Calcula la longitud del missatge previ a l'error
            longitud_error2 = len(text_fila)

            # Afegeix les dades a la llista d'errors
            errors2.append([id_fila, text_fila, longitud_error2])

            # Reinicia per al següent error
            id_fila = None
            text_fila = ""

# Copia les dades al marc de dades 'df_errors2' i exporta'l a csv
df_errors2 = pd.DataFrame(errors2, columns=['id', 'text_resposta_API', 'longitud_error2'])

# Exporta 'df_errors2.csv' només si no existeix
if not os.path.exists('df_errors2.csv'):
    df_errors2.to_csv('df_errors2.csv', index=False)
```

S'afegeixen dues columnes a `df_errors2`: `sense_noms` i `per_reprocessar`. S'exporta el marc de dades i els 62 errors es revisen manualment.

```
[ ]: df_errors2['sense_noms'] = 0
df_errors2['per_reprocessar'] = 0

df_errors2.to_excel('df_errors2.xlsx', index=False)
```

Una vegada revisats els resultats manualment, es troben 14 casos més d'articles que no contenen noms, 46 casos que es poden processar mitjançant expressions regulars on l'API havia comès alguna errada formal i 2 casos on només ha retornat els noms sense les característiques de les persones. La notícia amb ID 1309 es completa manualment a `df_errors2` amb les dades de `df_errors`. La notícia amb ID 180 es reprocessa manualment amb l'API per tercera vegada i s'afegeix també a `df_errors2`.

```
[ ]: df_errors2 = pd.read_excel('df_errors2.xlsx')
```


Primerament, els 14 casos sense noms s'afegeixen a la llista de notícies processades.

```
[ ]: # Cerca totes les files de 'df_errors2' marcades com que no inclouen noms
for index, row in df_errors2.iterrows():
    if row['sense_noms'] == 1:
        # Actualitza el conjunt 'notícies_processades'
        if row['id'] not in notícies_processades:
            notícies_processades.add(row['id'])

# Actualitza 'notícies_processades.txt'
with open('notícies_processades.txt', 'w', encoding='utf-8') as file:
    for pid in notícies_processades:
        file.write(f"{pid}\n")
```

A continuació, la informació de les 48 notícies restants s'incorpora a *df_persones*.

```
[ ]: # Itera cada línia de 'df_errors2'
for index, row in df_errors2.iterrows():

    # Comprova si ja s'ha processat
    if row['id'] in notícies_processades:
        continue

    # Si no s'ha processat i està marcada per reprocessar
    if row['per_reprocessar'] == 1:

        # Troba la notícia origen a 'df_notícies'
        id_fila = df_notícies[df_notícies['id'] == row['id']].iloc[0]

        # Divideix cada línia del log en els components d'informació de cada
        ↳ persona
        persones_info = row['text_resposta_API'].split('; ')
        for núm, info_persona in enumerate(persones_info, start=1):
            info_persona = info_persona.replace(',', ' ', ', ')
            if info_persona.endswith(';'):
                info_persona = info_persona[:-1]
            components = info_persona.split(' ', ')

            # En cas d'error en el reprocessament
            if len(components) != 9 or not all(c in ['0', '1'] for c in
            ↳ components[4:]):
                print(f"Error en el reprocessament: {row['id']}")
                continue

            # Elimina el prefix amb l'ID de la notícia que surt a les
            ↳ entrades del log
            nom = components[0].replace(str(row['id']) + ': ', '')

            # Crea un diccionari amb els components de la resposta
            person_data = {
                'id': row['id'],
```

```

        'núm': núm,
        'nom': nom,
        'edat': components[1],
        'ocupació': components[2],
        'rol_familiar': components[3],
        'protagonista': components[4],
        'portaveu': components[5],
        'experta': components[6],
        'testimoni': components[7],
        'víctima': components[8],
        'dins_títol': int(nom.lower() in id_fila['títol'].lower()),
        'dins_subtítol': int(nom.lower() in id_fila['subtítol']
↳lower()),
        'dins_text': int(nom.lower() in id_fila['text'].lower()),
        'coincidències': id_fila['títol'].lower().count(nom.lower())
↳+ id_fila['subtítol'].lower().count(nom.lower()) + id_fila['text']
↳lower().count(nom.lower())
    }

    # Afegeix la informació de les persones que surten a les
↳notícies marcades per reprocessar
    df_persones = pd.concat([df_persones, pd.
↳DataFrame([person_data]), ignore_index=True)

    # Actualitza la llista de notícies processades
    if row['id'] not in notícies_processades:
        notícies_processades.add(row['id'])
        with open('notícies_processades.txt', 'a', encoding='utf-8') as
↳file:
            file.write(f"{row['id']}\n")

# Exporta el marc de dades a CSV
if len(pd.read_csv('df_persones.csv')) != 6246:
    df_persones.to_csv('df_persones.csv', index=False)

```

```

[ ]: # Elimina variables prescindibles
for variable in ['components', 'data_hora', 'errors2', 'file', 'id_error2',
↳'id_fila', 'index', 'line', 'log',
        'longitud_error2', 'nom_log', 'pid', 'recompte_errors',
↳'row', 'text_complet', 'text_fila']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

7.5 Anàlisi de les fotografies

Primer, es crea una taula on es desen els ID de cada notícia que té imatge (és a dir, es descarten els valors buits a *url_imatge*) i s'afegeixen noves columnes per als valors que s'extreuran amb intel·ligència artificial. També es crea un arxiu per recomptar les imatges ja processades en

cas que l'API doni error.

```
[ ]: # Es comprova si df_imatges ja existeix prèviament per no eliminar les
↳imatges ja processades
if os.path.exists("df_imatges.csv"):
    df_imatges = pd.read_csv("df_imatges.csv").fillna('')
else:
    # Si no existeix, es crea df_imatges amb els ID i els URL
    df_imatges = df_notícies[df_notícies['url_imatge'] != ''][['id',
↳'url_imatge']]

    # S'hi afegeixen quatre columnes per desar les respostes de l'API
    for columna in ['imatge_persones', 'imatge_dones', 'imatge_homes',
↳'imatge_indeterminades']:
        if columna not in df_imatges.columns:
            df_imatges[columna] = ''

# Es carrega l'arxiu amb les imatges ja processades
imatges_processades = set()
if os.path.exists("imatges_processades.txt"):
    with open("imatges_processades.txt", "r") as arxiu:
        imatges_processades = {int(fila.strip()) for fila in arxiu}
```

Es defineixen dues funcions personalitzades. La primera comprova si l'URL de la imatge és vàlida, atès que alguns casos mostren un error 404. Si no és vàlida, servirà per marcar la imatge com a processada. La segona funció és la que envia els URL a l'API, fa el *parsing* de la resposta mitjançant una expressió regular i n'emmagatzema els resultats a les columnes noves de *df_imatges*.

```
[ ]: # Comprova si l'URL és vàlida
def comprova_url(url):
    try:
        response = requests.head(url, allow_redirects=True, timeout=5)
        return response.status_code == 200
    except requests.RequestException:
        return False

# Defineix una funció per analitzar cada imatge
def analitza_imatge(url, row_id):

    # Primer, executa la funció per comprovar si la URL funciona
    if not comprova_url(url):
        print(f"Fila {row_id}: URL no accessible.")
        # Si no funciona, retorna quatre strings buits
        return '', '', '', ''

    # Si l'URL funciona, crida el model GPT-4 Vision amb l'API d'OpenAI
    try:
        # Defineix els paràmetres de la crida de l'API i el prompt
        response = openai.ChatCompletion.create(
            model="gpt-4-vision-preview",
```

```

    messages=[
        {
            "role": "user",
            "content": [
                {"type": "text", "text": "I'll show you a photograph
↳ published in a newspaper. Determine how many people are in this image.
↳ Also, determine their gender. Return an answer with 4 numbers separated
↳ by commas following a regular expression: total of people, females,
↳ males, unknown gender. Example: 5,2,3,0. If there is no people, return
↳ 0,0,0,0"},
                {
                    "type": "image_url",
                    "image_url": {
                        "url": url,
                        "detail": "high"
                    },
                },
            ],
        }
    ],
    max_tokens=300,
)

# Extreu i processa la resposta
resposta = response.choices[0].message.content

# Imprimeix la resposta per facilitar trobar errades
print(f"Fila {row_id}: Resposta: {resposta}")

# Utilitza una expressió regular per trobar un patró de quatre
↳ xifres separades per comes dins de la resposta
check = re.search(r'(\d+),\s*(\d+),\s*(\d+),\s*(\d+)[.,;]?$',
↳ resposta)
if check:
    persones, dones, homes, indeterminades = map(int, check.groups())
    return persones, dones, homes, indeterminades

# Si no es troba l'expressió regular, es mostra un missatge d'error
else:
    print(f"Fila {row_id}: Format erroni.")
    return None, None, None, None

# En cas d'error en la crida a l'API, es mostra el següent missatge
except Exception as e:
    print(f"Error en cridar l'API a la fila {row_id}: {e}")
    return None, None, None, None

# Crea el loop per iterar sobre el marc de dades
for index, row in df_imatges.iterrows():

```

```

# Comprova si la fila encara no s'ha processat
per_processar = row['id'] not in imatges_processades
per_processar = per_processar and (pd.isna(row['imatge_persones']) or
↳row['imatge_persones'] == '')

# Si la fila no s'ha processat, executa la crida a l'API
if per_processar:
    persones, dones, homes, indeterminades =
↳analitza_imatge(row['url_imatge'], row['id'])

    # Si l'URL era invàlida, s'afegeix l'ID a la llista sense
↳processar-la
    if persones == '' and dones == '' and homes == '' and indeterminades
↳== '':
        imatges_processades.add(row['id'])
        with open("imatges_processades.txt", "a") as file:
            file.write(f"{row['id']}\n")

    # En cas contrari, es fa el 'parsing' de la resposta amb l'expressió
↳regular i s'assignen els resultats a la columna corresponent
    else:
        df_imatges.at[index, 'imatge_persones'] = persones
        df_imatges.at[index, 'imatge_dones'] = dones
        df_imatges.at[index, 'imatge_homes'] = homes
        df_imatges.at[index, 'imatge_indeterminades'] = indeterminades

    # Comprova si l'anàlisi de la imatge ha estat satisfactòria i,
↳si sí, desa l'ID de la fila a 'imatges_processades'
    if persones is not None and dones is not None and homes is not
↳None and indeterminades is not None:
        imatges_processades.add(row['id'])
        with open("imatges_processades.txt", "a") as file:
            file.write(f"{row['id']}\n")

    # En acabar cada iteració, guarda 'df_imatges' com un CSV per evitar
↳perdre informació ja analitzada
    df_imatges.to_csv("df_imatges.csv", index=False)

# Finalment, s'eliminen les variables prescindibles
for variable in ['arxiu', 'columna', 'dones', 'file', 'homes',
↳'indeterminades', 'index', 'per_processar', 'persones', 'processed_ids',
↳'row']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

Queden 16 files en blanc, que la IA no analitza i que integren tres supòsits: - L'URL està trencat. - Les fotografies mostren plans generals amb molta gent (p. ex, manifestacions). - Les

imatges són borroses (p. ex., un frame extret d'un vídeo fet amb mòbil).

El temps de processament ha estat d'1 hora i 10 minuts. El cost total de les proves, el processament i el reprocessament d'algunes imatges que varen donar error en primera instància ha estat de 3,20 dòlars (uns 2,94 euros).

7.6 Neteja dels resultats de l'anàlisi de textos

Es detecta que al marc de dades `df_persones` hi ha 42 noms que no es corresponen a persones. S'eliminen les files corresponents.

```
[ ]: # Llista de noms per eliminar
mapeig_noms_errors = [
    "", "A-abogados", "Africa Archives ", "Animalario", "Antònia Font",
    ↪ "Arde Bogotá", "Artola Couture", "Autor del atropello", "Barrancas",
    ↪ "Bildu",
    "Bomba Estéreo", "CEP", "CUP", "Ciudadanos", "Compromís", "Cuenca", "EH",
    ↪ "Bildu", "ERC", "FIBA", "Festival Grec", "Fito y Fitipaldis", "Gemeliers",
    "Gerry & the Pacemakers", "Gilbert & George", "Grip Face", "Junts", "Las",
    ↪ "Brujas de Salem", "Manacor", "Manuela va de fiesta",
    "Memphis", "Más País", "Moby Dick", "Nemo", "PDeCAT", "Pep Merda",
    ↪ "Policía Local", "Principiantes", "Prostitución", "Rayo McQueen", "Santa",
    ↪ "Eugènia", "SKFK", "SPP", "SUP",
    "Sastrería Aldabaldetrek", "Shock 1", "Shock 2", "Sister Sister Brand",
    ↪ "Talisman Jewellery", "Teatre Principal de Palma", "Teatre Principal",
    ↪ "d'Inca",
    "Testigos presenciales", "The Weekend", "UFP", "Vetusta Morla", "familia",
    ↪ "March"
]

# Elimina les files amb aquests noms de df_persones
df_persones = df_persones[~df_persones['nom'].isin(mapeig_noms_errors)]
```

També es detecten sis professions que fan referència a col·lectius o personatges de ficció, per la qual cosa se n'esborren els resultats.

```
[ ]: # Llista d'ocupacions per eliminar
mapeig_ocupacions_errors = ["Brand", "nickname", "character", "band",
    ↪ "collective", "fictional character"]

# Elimina les files amb aquestes ocupacions de df_persones
df_persones = df_persones[~df_persones['ocupació'].
    ↪ isin(mapeig_ocupacions_errors)]
```

Una vegada depurats els resultats, es restableixen els comptadors de la columna `núm` i s'exporten els canvis.

```
[ ]: # Ordena df_persones per 'id', després per 'núm' i finalment per 'nom'
df_persones = df_persones.sort_values(by=['id', 'núm', 'nom'])

# Crea una funció per renumerar la columna 'núm' de df_persones
def reinicia_df_persones_núm(group):
    group['núm'] = range(1, len(group) + 1)
```

```

return group

# S'executa la funció i es restableix l'índex del marc de dades
df_persones = df_persones.groupby('id').apply(reinicia_df_persones_núm).
↳reset_index(drop=True)

# S'exporta a CSV
df_persones.to_csv('df_persones.csv', index=False)

```

7.7 Anàlisi del sexe dels subjectes de les notícies

Seguint el mateix mètode emprat per a l'anàlisi de les firmes, es crea una taula nova (*df_sexepersones*), amb tots els noms diferents que s'han recuperat amb l'anàlisi de textos.

```

[ ]: df_sexepersones = pd.DataFrame({'nom': df_persones['nom'].unique()})
df_sexepersones['sexe'] = ''
df_sexepersones = df_sexepersones.sort_values(by='nom').
↳reset_index(drop=True)

```

A continuació, es mapegen manualment una sèrie de noms per assignar-los el valor indeterminat i evitar processaments innecessaris. S'estableixen els valors de sexe per a un total de 65 noms, fonamentalment sigles i artistes amb pseudònims.

```

[ ]: # Es defineixen els mapejos segons el sexe que s'assignarà
mapeig_noms_indeterminat = ['C', 'E', 'E.M.N.R.', 'F.C.M.F.', 'Fiscal', 'G.O.
↳', 'J.C.', 'J.R.', 'R. R. L.', 'V']

mapeig_noms_dona = [
    'Abogada defensora', 'Alaska', 'Becky G', 'Gata Cattana', 'Karol G', 'La
↳Nena', 'La Prohibida', 'La Terremoto de Alcorcón', 'La serrana', 'Lady
↳Di', 'Le Pen',
    'Letrada de la acusación particular', 'Mala Rodríguez', 'Mi madre', 'Mi
↳mujer', 'Mi tía María', 'Mujer', 'Niña Pastori', 'Papi Robles', 'Sor
↳Tomasseta', 'Sra. V.',
    'Suegra', 'chica', 'QTCinderella'
]

mapeig_noms_home = [
    'A. Sepúlveda', 'Acusado', 'Agresor1', 'Agresor2', 'Arkano',
↳'AuronPlay', 'B. Ramon', 'Bad Bunny', 'Bizarrap', 'C.Tangana', 'Cardenal
↳Despuig', 'Daddy Yankee',
    'De Bruyne', 'De Jong', 'Frank T', 'Kake Portas', 'Kang In Lee', 'Le
↳Normand', 'Maluma', 'Mi padre', 'Mi suegro', 'Mi tío Alfonso', 'Núñez
↳Feijóo', 'Quevedo',
    'Sr. C.', 'Sr. M.', 'Sr. S.', 'Sr. Á.', 'Yerno', 'chico', 'hombre'
]

# Es carreguen els valors a la columna 'sexe' de df_sexepersones
df_sexepersones.loc[df_sexepersones['nom'].isin(mapeig_noms_indeterminat),
↳'sexe'] = 'indeterminat'

```

```

df_sexepersones.loc[df_sexepersones['nom'].isin(mapeig_noms_dona), 'sexe'] =
↳ 'dona'
df_sexepersones.loc[df_sexepersones['nom'].isin(mapeig_noms_home), 'sexe'] =
↳ 'home'

# Finalment, s'eliminen les variables prescindibles
for variable in ['mapeig_noms_dona', 'mapeig_noms_errors',
↳ 'mapeig_noms_home', 'mapeig_noms_indeterminat',
↳ 'mapeig_ocupacions_errors']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

La següent passa és fer els preparatius per enviar a l'API el llistat d'uns 3.000 noms amb una fórmula semblant a la utilitzada per a l'anàlisi de firmes. Es crea un conjunt anomenat *persones_processades* que conté els noms als quals ja se li ha assignat un sexe.

```

[ ]: persones_processades = set()

if os.path.exists("persones_processades.txt"):
    with open("persones_processades.txt", "r") as arxiu:
        for fila in arxiu:
            noms = fila.strip().split('; ')
            persones_processades.update(noms)
else:
    persones_processades = set(df_sexepersones[df_sexepersones['sexe'] !=
↳ '']['nom'])

```

Es creen tongades de 12 noms per reduir les cridades a l'API.

```

[ ]: # Carrega les dades ja processades
if os.path.exists("df_sexepersones.csv"):
    df_sexepersones = pd.read_csv("df_sexepersones.csv").fillna('')
else:
    None

# Afegeix els noms amb sexe ja assignat
noms_sense_sexe = df_sexepersones[df_sexepersones['sexe'] == '']['nom']

# Crea la llista amb les tongades de 12 noms
noms_per_processar = ["; ".join(noms_sense_sexe[i:i+12]) for i in range(0,
↳ len(noms_sense_sexe), 12)]

```

Finalment, es crea la funció que crida a l'API i sol·licita una resposta amb uns valors preassignats. Els valors numèrics s'associen a un sexe per tal de processar-ho localment.

```

[ ]: # Es defineix una funció per enviar una tongada de noms i processar la
↳ resposta
def analitza_sexe_persones(tongada_noms):
    try:

```



```

response = openai.ChatCompletion.create(
    model="gpt-4-1106-preview",
    messages=[
        {
            "role": "system",
            "content": "Determine the gender of the following 12
↳names. Answer with 1 for female, 2 for male, 3 for unknown, separating
↳the answers with commas."
            "If you think there is a doubt, between 'female' and
↳'male' answer with the most likely option."
            "Return an answer with 12 numbers separated by commas
↳following a regular expression. Example: 2,1,2,1,1,2,2,1,2,2,1,2."
        },
        {
            "role": "user",
            "content": f"The 12 names, separated by semicolons, are:
↳{tongada_noms}"
        }
    ],
    max_tokens=100
)

# Processa la resposta
resposta = response.choices[0].message["content"]
print(f"Noms: {tongada_noms}, Resposta: {resposta}")
patró =
↳r'(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+)'
codis_sexe = re.findall(patró, resposta)
codis_sexe = [item for sublist in codis_sexe for item in sublist] if
↳codis_sexe else None
return codis_sexe

# Assenyala els errors de l'API
except Exception as e:
    print(f"Error amb la tongada '{tongada_noms}': {e}")
    return None

# Assigna les etiquetes per a cada sexe
etiqueta_sexe = {'1': 'dona', '2': 'home', '3': 'indeterminat'}

# Itera sobre cada tongada i executa la funció
for tongada in noms_per_processar:
    codis_sexe = analitza_sexe_persones(tongada)

    if codis_sexe:

        # Divideix la llista
        noms_list = tongada.split('; ')

        # Itera sobre cada nom i assigna un sexe

```

```

for nom, codi_sexe in zip(noms_list, codis_sexe):

    # Actualitza df_sexepersones
    df_sexepersones.loc[df_sexepersones['nom'] == nom.strip(),
↳'sexe'] = etiqueta_sexe.get(codi_sexe, '')

    # Actualitza persones_processades
    persones_processades.add(nom.strip())

# Afegeix els noms processats al llistat
with open("persones_processades.txt", "a") as arxiu:
    arxiu.write(f"{tongada}\n")

# En acabar cada iteració, guarda 'df_sexepersones' com un CSV per
↳evitar perdre informació ja analitzada
df_sexepersones.to_csv("df_sexepersones.csv", index=False)

# Finalment, s'eliminen les variables prescindibles
for variable in ['arxiu', 'codi_sexe', 'codis_sexe', 'etiqueta_sexe',
↳'fila', 'nom', 'noms', 'noms_list', 'noms_per_processar',
↳'noms_sense_sexe', 'tongada']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

La durada de l'execució del codi ha estat de 8 minuts i 11 segons. El processament en tongades no només ha permès escurçar el temps de processament, sinó també utilitzar un model de llenguatge més complet (GPT-4 Turbo), en comptes del model GPT-3.5 Turbo que es va fer servir per analitzar les firmes. A més, també ha permès rebaixar dràsticament el cost de processament a 0,69 dòlars (uns 0,63 euros).

Per completar el marc de dades *df_persones*, s'associa cada nom de *df_persones* al sexe que té a *df_sexepersones*.

```

[ ]: # Associa el sexe a cada nom
sexe_nom = df_sexepersones.set_index('nom')['sexe']
df_persones['sexe'] = df_persones['nom'].map(sexe_nom)

# Reordena les columnes
df_persones = df_persones[['id', 'núm', 'nom', 'sexe', 'edat', 'ocupació',
↳'rol_familiar',
↳'protagonista', 'portaveu', 'experta',
↳'testimoni', 'víctima',
↳'dins_títol', 'dins_subtítol', 'dins_text',
↳'coincidències']]

del sexe_nom

```

7.8 Homogeneïtzació dels valors de rol familiar

Es crea una llista única de valors de la columna `rol_familiar` de `df_persones`.

```
[ ]: rols_únics = sorted(df_persones['rol_familiar'].unique().tolist())
```

Atès el baix volum de resultats, els 21 valors únics es transformen a mà. D'acord amb les indicacions del GMMP, s'assigna un 1 si la persona té un rol familiar assignat a la notícia i un 0 en cas contrari. Els valors s'estableixen d'acord amb els mapejos.

```
[ ]: # Defineix una funció per assignar els valors a partir d'un mapeig
def transforma_rol_familiar(rol):
    valor_0 = [0, '', '0', 'amiga', 'amigo', 'friend']

    valor_1 = [1, '1', 'aunt', 'brother', 'daughter', 'Girlfriend', 'hija', '
↳husband', 'lover',
               'madre', 'mother', 'nieta', 'nieto', 'parent', 'partner', '
↳sister', 'son',
               'uncle', 'wife']

    if rol in valor_0:
        return 0
    elif rol in valor_1:
        return 1

# Aplica els canvis al marc de dades, transforma la columna a valors enters,
↳i exporta'ls
df_persones['rol_familiar'] = df_persones['rol_familiar'].
↳apply(transforma_rol_familiar)
df_persones['rol_familiar'].fillna(0, inplace=True)
df_persones['rol_familiar'] = df_persones['rol_familiar'].astype(int)
df_persones.to_csv('df_persones.csv', encoding='utf-8', index=False)

del rols_únics
```

7.9 Homogeneïtzació dels valors d'ocupació

D'una banda, es crea un marc de dades amb les categories d'ocupació que estableix la guia de codificació del Global Media Monitoring Project.

```
[ ]: categories_ocupació_GMMP = {
    'id_categoriaocupació': list(range(28)),
    'desc_categoriaocupació': [
        "Not stated",
        "Royalty, monarch, deposed monarch...",
        "Politician, member of parliament...",
        "Government employee, public servant, spokesperson...",
        "Police, military, paramilitary, militia, fire officer",
        "Academic expert, lecturer, teacher",
        "Doctor, dentist, health specialist",
        "Health worker, social worker, childcare worker",
        "Science/technology professional, engineer...",
```

```

    "Media professional, journalist, film-maker...",
    "Lawyer, judge, magistrate, legal advocate...",
    "Business person, exec, manager, stockbroker...",
    "Office or service worker, non-management worker",
    "Tradesperson, artisan, labourer, truck driver...",
    "Agriculture, mining, fishing, forestry",
    "Religious figure, priest, monk, rabbi, mullah, nun...",
    "Activist or worker in civil society organization, NGO, trade union",
    "Sex worker",
    "Celebrity, artist, actor, writer, singer, TV personality",
    "Sportsperson, athlete, player, coach, referee",
    "Student, pupil, schoolchild",
    "Home-maker, parent",
    "Child, young person",
    "Villager or resident",
    "Retired person, pensioner",
    "Criminal, suspect",
    "Unemployed",
    "Other"
]
}

df_categoriaocupació = pd.DataFrame(categories_ocupació_GMMP)

```

D'altra banda, es relacionen les diferents ocupacions i càrrecs que la IA ha trobat a les notícies.

```

[ ]: # Crea una llista de valors únics d'ocupació a un nou marc de dades
ocupacions_úniques = sorted(df_persones['ocupació'].unique())
df_ocupacions = pd.DataFrame(ocupacions_úniques, columns=['ocupació'])

del categories_ocupació_GMMP, ocupacions_úniques

```

A continuació, es preparen les dades per enviar-les en tongades. Es crea un conjunt anomenat *ocupacions_processades* per allotjar els valors amb un resultat vàlid, en cas que falli la connexió a l'API.

```

[ ]: ocupacions_processades = set()

if os.path.exists("ocupacions_processades.txt"):
    with open("ocupacions_processades.txt", "r") as arxiu:
        for fila in arxiu:
            ocupacions = fila.strip().split('; ')
            ocupacions_processades.update(ocupacions)

```

Es creen tongades de 12 ocupacions per reduir les cridades a l'API.

```

[ ]: # Carrega les dades ja processades
if os.path.exists("df_ocupacions.csv"):
    df_ocupacions = pd.read_csv("df_ocupacions.csv").fillna('')
else:
    None

```

```
# Exclou les ocupacions que ja han estat codificades
ocupacions_sense_codi = df_ocupacions[~df_ocupacions['ocupació'].
↳isin(ocupacions_processades)]['ocupació']

# Crea la llista amb les tongades de 12 ocupacions
ocupacions_per_processar = ["; ".join(ocupacions_sense_codi[i:i+12]) for i_
↳in range(0, len(ocupacions_sense_codi), 12)]
```

Es crea el prompt del marc de dades `df_categoriaocupació` per explicar a la IA quines són les possibles categories i els seus codis.

```
[ ]: prompt_categories = "\n".join([f"{row['id_categoriaocupació']}_
↳{row['desc_categoriaocupació']}" for index, row in df_categoriaocupació.
↳iterrows()])
```

Per simplificar la cridada a l'API, es defineix primer el missatge de sistema i es mostra per comprovar-ne el resultat.

```
[ ]: missatge_sistema = f"Match each of the provided occupations to one of the_
↳categories by answering only the category ID number. These are the_
↳categories:\n{prompt_categories}\n\nIf you think there is a doubt between_
↳several options, answer with the lowest ID number.\nReturn an answer with_
↳12 numbers separated by commas following a regular expression. Example:_
↳2,0,27,5,1,2,2,4,5,8,9,12."
print(missatge_sistema)
```

Match each of the provided occupations to one of the categories by answering only the category ID number. These are the categories:

```
0 Not stated
1 Royalty, monarch, deposed monarch...
2 Politician, member of parliament...
3 Government employee, public servant, spokesperson...
4 Police, military, paramilitary, militia, fire officer
5 Academic expert, lecturer, teacher
6 Doctor, dentist, health specialist
7 Health worker, social worker, childcare worker
8 Science/technology professional, engineer...
9 Media professional, journalist, film-maker...
10 Lawyer, judge, magistrate, legal advocate...
11 Business person, exec, manager, stockbroker...
12 Office or service worker, non-management worker
13 Tradesperson, artisan, labourer, truck driver...
14 Agriculture, mining, fishing, forestry
15 Religious figure, priest, monk, rabbi, mullah, nun...
16 Activist or worker in civil society organization, NGO, trade union
17 Sex worker
18 Celebrity, artist, actor, writer, singer, TV personality
19 Sportsperson, athlete, player, coach, referee
20 Student, pupil, schoolchild
21 Home-maker, parent
22 Child, young person
23 Villager or resident
```

24 Retired person, pensioner
 25 Criminal, suspect
 26 Unemployed
 27 Other

If you think there is a doubt between several options, answer with the lowest ID number.

Return an answer with 12 numbers separated by commas following a regular expression. Example: 2,0,27,5,1,2,2,4,5,8,9,12.

Es crea una funció per, a través d'IA, codificar les ocupacions a un dels grups establerts pel GMMP.

```
[ ]: # Es defineix una funció per enviar una tongada d'ocupacions i processar la
↳ resposta
def codifica_ocupacions(tongada_ocupacions):
    try:
        # Crida l'API amb el prompt complet
        response = openai.ChatCompletion.create(
            model="gpt-4-1106-preview",
            messages=[
                {
                    "role": "system",
                    "content": f"{missatge_sistema}"
                },
                {
                    "role": "user",
                    "content": f"The 12 occupations, separated by
↳ semicolons, are: {tongada_ocupacions}"
                }
            ],
            max_tokens=100
        )

        # Processa la resposta amb una expressió regular
        resposta = response.choices[0].message["content"]
        print(f"Ocupacions: {tongada_ocupacions}, Resposta: {resposta}")
        pattern =
↳ r'(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+),\s*(\d+)'
        categoria_codis = re.findall(pattern, resposta)
        categoria_codis = [item for sublist in categoria_codis for item in
↳ sublist] if categoria_codis else None
        return categoria_codis

    # Mostra els errors de l'API
    except Exception as e:
        print(f"Error amb la tongada '{tongada_ocupacions}': {e}")
        return None
```

Tot seguit, s'executa la funció i es processen els resultats per emmagatzemar els codis a `df_ocupacions`.

```
[ ]: # Itera sobre cada tongada i executa la funció per codificar les ocupacions
for tongada_ocupacions in ocupacions_per_processar:
    codis_categoria = codifica_ocupacions(tongada_ocupacions)

    if codis_categoria:

        # Divideix la llista d'ocupacions de la tongada
        llista_ocupacions = tongada_ocupacions.split('; ')

        # Itera sobre les 12 ocupacions de la tongada i assigna el codi de
        ↳ la categoria atorgat per l'API
        for ocupació, codi_categoria in zip(llista_ocupacions,
        ↳ codis_categoria):

            # Actualitza df_ocupacions a una nova columna on s'indica el
            ↳ codi de la categoria
            df_ocupacions.loc[df_ocupacions['ocupació'] == ocupació.strip(),
            ↳ 'id_categoriaocupació'] = codi_categoria

            # Actualitza el conjunt d'ocupacionsprocessades
            ocupacions_processades.add(ocupació.strip())

        # Afegeix les ocupacions processades a l'arxiu de registre
        with open("ocupacions_processades.txt", "a") as arxiu:
            arxiu.write(f"{tongada_ocupacions}\n")

        # Després de cada iteració, exporta df_ocupacions a un CSV per no perdre
        ↳ dades analitzades
        df_ocupacions.to_csv("df_ocupacions.csv", index=False)
```

Per primera vegada, l'API ha respost sense errors en el primer intent. El temps de processament ha estat d'1 minut i 9 segons. El cost del processament ha estat de 0,22 dòlars (0,20 euros). Finalment, s'eliminen les variables que no es tornaran a fer servir.

```
[ ]: # Elimina les variables innecessàries
for variable in ['arxiu', 'codi_categoria', 'codis_categoria', 'fila',
↳ 'llista_ocupacions',
↳ 'mapeig_codi_categoria_ocupació', 'missatge_sistema',
↳ 'ocupació', 'ocupacions', 'ocupacions_per_processar',
↳ 'ocupacions_sense_codi', 'prompt_categories',
↳ 'prompt_message', 'tongada_ocupacions',
↳ 'imatges_processades', 'notícies_processades',
↳ 'ocupacions_processades', 'persones_processades']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable
```

8 Revisió manual dels marcs de dades

8.1 De les ocupacions (*df_ocupacions*)

El primer marc de dades del qual se'n revisen els resultats és el d'ocupacions. Es comprova l'assignació de tots els 563 valors d'acord amb les taules facilitades del GMMP i es fa el següent mapatge de canvis.

```
[ ]: mapatge_canvis_manuais_ocupacions = {
  'murder victim': 0,
  'jury president': 1,
  'president': 2, 'presidenta': 2, 'first lady': 2, 'leader': 2, 'líder': 2,
  'Ex-leader': 2, 'Ex-president': 2, 'Chief of Staff': 2,
  'chief of diplomacy': 2, 'asesor': 2, 'commissioner': 2, 'Commissioner
of Budgets': 2, 'conseller de Movilidad y Vivienda': 2,
  'consellera': 2, 'conseller': 2, 'conseller insular de Turisme i
Esports': 2, 'consellera insular': 2, 'conseller insular': 2,
  'delegado del Gobierno': 2, 'delegada del Gobierno': 2, 'EU official': 2,
  'EMT director': 2, 'Education Councilor': 2,
  'Education Counselor': 2, 'director general de Asistencia Sanitaria': 2,
  'Director of Economy': 2, 'Deputy Director': 2,
  'Director General': 2, 'Commissioner': 2, 'vice-chancellor': 2, 'Health
Counselor': 2, 'consejero de Sanidad': 2,
  'health counselor': 2, 'CSD President': 2, 'Industria Minister': 2,
  'eurodiputada': 2, 'directora': 2, 'responsable de Urbanismo': 2,
  'director adjunto operativo': 4,
  'experto': 5,
  'neuropsychologist': 6, 'Col·legi Oficial de Psicologia': 6, 'sports
psychologist': 6, 'divulgadora científica y ginecóloga': 6,
  'political analyst': 9, 'financier': 9, 'photographer': 9, 'light
designer': 9,
  'drone pilot': 12, 'hostelería': 12,
  'cocinera': 13, 'cocinero': 13, 'sommelier': 13, 'interior designer': 13,
  'payés': 14, 'presidente de Aspa': 14,
  'Arciprest de Hita': 15,
  'WWF spokesperson': 16, 'president Association of Health Economics': 16,
  'coauthor': 18, 'videocreator': 18, 'streamer': 18, 'comedian': 18,
  'stylist': 18, 'creator': 18,
  'UEFA President': 19, 'former RFEF President': 19, 'football manager': 19,
  'former president of RFEF': 19,
  'ex-football player and president': 19, 'football federation president': 19,
  'escaladora': 19, 'técnico': 19,
  'law graduate': 20,
  'au pair': 27
}
```

A continuació, es fa un registre dels canvis i s'apliquen al marc de dades *df_ocupacions*.

```
[ ]: # Es crea una columna per registrar les modificacions fetes manualment
df_ocupacions['canvi_manual'] = 0
```



```

# Es crea una funció per aplicar el mapatge i marcar la columna
↳ 'canvi_manual' = 1 en cas que s'hagi modificat el valor
↳ d'id_categoriaocupació
def funció_mapatge_ocupacions(row):
    valor_antic = row['id_categoriaocupació']
    valor_nou = mapatge_canvis_manuais_ocupacions.get(row['ocupació'],
↳ valor_antic)
    if row['ocupació'] in mapatge_canvis_manuais_ocupacions:
        row['canvi_manual'] = 1

    row['id_categoriaocupació'] = valor_nou
    return row

# S'executen els canvis del mapatge
df_ocupacions = df_ocupacions.apply(funció_mapatge_ocupacions, axis=1)

```

La conclusió és que hi ha 76 valors canviats (un 13,5 %) d'un total de 563. Una vegada revisats els valors, la parametrització de codis a *df_ocupacions* s'utilitza per crear una nova columna a *df_persones* amb el codi de la ocupació.

```

[ ]: # Associa cada ocupació a df_persones amb un codi de categoria
mapatge_codi_categoria_ocupació = df_ocupacions.
↳ set_index('ocupació')['id_categoriaocupació']
df_persones['categoria_ocupació'] = df_persones['ocupació'].
↳ map(mapatge_codi_categoria_ocupació).fillna(0) # Els valors no trobats
↳ s'assignen a 0 (0 = professió no esmentada)

# Converteix les dades a enters
df_ocupacions['id_categoriaocupació'] =
↳ df_ocupacions['id_categoriaocupació'].astype(int)
df_persones['categoria_ocupació'] = df_persones['categoria_ocupació'].
↳ astype(int)

# Reordena les columnes
df_persones = df_persones[['id', 'núm', 'nom', 'sexe', 'edat', 'ocupació',
↳ 'categoria_ocupació',
                                'rol_familiar', 'protagonista', 'portaveu',
↳ 'experta', 'testimoni',
                                'víctima', 'dins_títol', 'dins_subtítol',
↳ 'dins_text', 'coincidències']]

# Exporta df_persones i df_ocupacions
df_persones.to_csv("df_persones.csv", index=False)
df_ocupacions.to_csv("df_ocupacions.csv", index=False)

del mapatge_codi_categoria_ocupació, mapatge_canvis_manuais_ocupacions

```

8.2 De les persones (*df_persones*)

En aquest apartat, es realitza una comprovació manual aleatòria d'un conjunt de notícies per veure la fiabilitat de la IA a l'hora d'analitzar les dades de les persones. Primer, se selecciona

aleatòriament un conjunt de 131 notícies (un 10 % del total de *df_notícies*) i es trespassen a un nou marc de dades anomenat *df_revisió*.

```
[ ]: # Comprova si l'arxiu ja s'ha exportat abans
if not os.path.exists('df_revisió.csv'):
    # Crea la mostra de 131 files com a df_revisió
    df_revisió = df_notícies.sample(n=131)
    # Exporta a CSV i Excel
    df_revisió.to_csv('df_revisió.csv', index=False)
    df_revisió.to_excel('df_revisió.xlsx', index=False)
# Si l'arxiu ja s'havia creat, es carrega el CSV
else:
    df_revisió = pd.read_csv('df_revisió.csv').fillna('')
```

A continuació, es crea un marc de dades (*df_revisió_persones*) amb les dades de *df_persones* relatives a les 131 notícies de la mostra.

```
[ ]: # Comprova si l'arxiu ja s'ha exportat abans
if not os.path.exists('df_revisió_persones.csv'):
    # Crea la submostra de df_persones per a les files amb el mateix id que
    ↪ les notícies seleccionades a df_revisió
    df_revisió_persones = pd.merge(df_revisió[['id']], df_persones, on='id',
    ↪ how='inner')
    # Exporta a CSV i Excel
    df_revisió_persones.to_csv('df_revisió_persones.csv', index=False)
    df_revisió_persones.to_excel('df_revisió_persones.xlsx', index=False)
# Si l'arxiu ja s'havia creat, es carrega el CSV
else:
    df_revisió_persones = pd.read_csv('df_revisió_persones.csv').fillna('')
```

Es revisen les notícies manualment i es fan canvis a *df_revisió_persones*. Es reanomenen els arxius com a *df_revisió2.xlsx* i *df_revisió_persones2.xlsx*. Es carreguen tots dos arxius i es crea un codi que cerca si hi ha canvis a cada línia respecte de *df_revisió_persones.csv*. Els resultats de la comprovació s'afegeixen a noves columnes al marc de dades *df_revisió_persones2*, una per cadascuna de les columnes que es comparen, que són: - nom - sexe - edat - ocupació - rol_familiar - protagonista - portaveu - experta - testimoni - víctima

Les noves columnes dins *df_revisió_persones2* s'anomenen *CHK_[nom-columna]*.

```
[ ]: # Es carreguen els arxius de la revisió manual
df_revisió2 = pd.read_excel("df_revisió2.xlsx").fillna('')
df_revisió_persones2 = pd.read_excel("df_revisió_persones2.xlsx").fillna('')

# Es crea un marc de dades temporal amb les columnes de df_revisió_persones2
↪ i df_revisió_persones
df_revisió_persones3 = pd.merge(df_revisió_persones2, df_revisió_persones,
↪ on=['id', 'núm'], suffixes=('_2', ''), how='left', indicator=True)

# Es determinen les columnes de la comparativa
columnes_comparativa = ['nom', 'sexe', 'edat', 'ocupació', 'rol_familiar',
↪ 'protagonista', 'portaveu', 'experta', 'testimoni', 'víctima']
```

```

# S'itera per les columnes i es creen les columnes de comprovació
for col in columnes_comparativa:
    check_col = f'CHK_{col}'
    # Es comparen els valors d'ambdós marcs de dades
    df_revisió_persones3[check_col] = (df_revisió_persones3[f'{col}_2'] !=
    ↪df_revisió_persones3[col]).astype(int)

# Extreu les columnes de comprovació
chk_columnes = ['id', 'núm'] + [f'CHK_{col}' for col in columnes_comparativa]
df_check = df_revisió_persones3[chk_columnes]

# Actualitza df_revisió_persones2 amb les columnes de comprovació
df_revisió_persones2 = pd.merge(df_revisió_persones2, df_check, on=['id',
    ↪'núm'], how='left')

# Es defineixen les columnes de comprovació
chk_columns = [f'CHK_{col}' for col in columnes_comparativa]

del check_col, col, df_check, df_revisió_persones3

```

Per a les persones que s'han afegit durant la comprovació manual, les columnes de comprovació mostren sempre 1 tot i que no són discrepàncies de la IA. Aquests valors es canvien a 0.

```

[ ]: # Crea una funció per comprovar si totes les columnes de comprovació són uns
def canvia_1a0(row):
    if row[chk_columns].sum() == len(chk_columns):
        row[chk_columns] = 0
        return row, True
    return row, False

# Executa la funció i recompta els canvis
recompte_tot1 = 0
for index, row in df_revisió_persones2.iterrows():
    fila, tot1 = canvia_1a0(row)
    df_revisió_persones2.loc[index, chk_columns] = fila[chk_columns]
    if tot1:
        recompte_tot1 += 1

```

La informació corregida manualment es copia a *df_persones* i es restableix la numeració de la columna *núm*.

```

[ ]: # Elimina les files l'id de les quals està inclosa a df_revisió
df_persones = df_persones[~df_persones['id'].isin(df_revisió['id'])]

# Exclou les columnes de comprovació
columnes_per_copiar = [col for col in df_revisió_persones2.columns if not
    ↪col.startswith('CHK_')]
df_revisió_persones2_per_copiar = df_revisió_persones2[columnes_per_copiar]

# Afegeix les files corregides manualment

```

```

df_persones = pd.concat([df_persones, df_revisió_persones2_per_copiar],
↳ ignore_index=True)

# Ordena df_persones per 'id', després per 'número' i finalment per 'nom'
df_persones = df_persones.sort_values(by=['id', 'número', 'nom'])

# Crea una funció per renumerar la columna 'número' de df_persones
def reinicia_df_persones_número(group):
    group['número'] = range(1, len(group) + 1)
    return group

# S'executa la funció i es restableix l'índex del marc de dades
df_persones = df_persones.groupby('id').apply(reinicia_df_persones_número).
↳ reset_index(drop=True)

# S'exporta a CSV
df_persones.to_csv('df_persones_corregit.csv', index=False)

```

8.3 Estadístiques de fiabilitat (*st_fiabilitat*)

Completades les revisions manuals, es generen les estadístiques de fiabilitat per a cada columna i es desen al marc de dades *st_fiabilitat*.

```

[ ]: # Extreu les columnes de comprovació
columnes_check = [col for col in df_revisió_persones2.columns if col.
↳ startswith('CHK_')]

# Crea una funció per desar les estadístiques
def desa_estadística(estadístiques_fiabilitat, descripció, valor,
↳ percentatge):
    valor = '' if valor is None or np.isnan(valor) else int(valor)
    percentatge = '' if percentatge is None else percentatge
    estadístiques_fiabilitat.append({
        'descripció': descripció,
        'valor': valor,
        'percentatge': percentatge
    })

# Crea el conjunt
estadístiques_fiabilitat = []

# Calcula les estadístiques de l'anàlisi
percentatge_revisió = (len(df_revisió) / len(df_notícies)) * 100
suma_checked = df_revisió2['checked'].sum()
percentatge_checked = (suma_checked / len(df_revisió2)) * 100
percentatge_persones = (len(df_revisió_persones) / len(df_persones)) * 100
falsos_negatius = recompte_tot1
falsos_positius = recompte_tot1 + (len(df_revisió_persones) -
↳ len(df_revisió_persones2))
veritables_positius = len(df_revisió_persones) - falsos_positius

```

```

precisió = (veritables_positius / (veritables_positius + falsos_positius)) *
↳100
reclam = (veritables_positius / (veritables_positius + falsos_negatius)) *
↳100
puntuació_F1 = (2 * (precisió * reclam) / (precisió + reclam))

# Afegeix les estadístiques al conjunt
desa_estadística(estadístiques_fiabilitat, "Total de notícies a l'anàlisi",
↳len(df_notícies), None)
desa_estadística(estadístiques_fiabilitat, "Notícies per comprovar
↳manualment", len(df_revisió), f"{percentatge_revisió:.2f} %".replace('.',
↳','))
desa_estadística(estadístiques_fiabilitat, "Notícies comprovades
↳manualment", suma_checked, f"{percentatge_checked:.2f} %".replace('.',
↳','))
desa_estadística(estadístiques_fiabilitat, "Total de persones a l'anàlisi",
↳len(df_persones), None)
desa_estadística(estadístiques_fiabilitat, "Persones per comprovar
↳manualment", len(df_revisió_persones), f"{percentatge_persones:.2f} %".
↳replace('.', ','))
desa_estadística(estadístiques_fiabilitat, "Persones detectades correctament
↳per la IA (veritables positius)", veritables_positius, None)
desa_estadística(estadístiques_fiabilitat, "Persones eliminades manualment
↳(falsos positius)", falsos_positius, None)
desa_estadística(estadístiques_fiabilitat, "Persones afegides manualment
↳(falsos negatius)", falsos_negatius, None)
desa_estadística(estadístiques_fiabilitat, "Precisió", None, f"{precisió:.
↳2f} %".replace('.', ','))
desa_estadística(estadístiques_fiabilitat, "Reclam", None, f"{reclam:.2f} %".
↳replace('.', ','))
desa_estadística(estadístiques_fiabilitat, "Puntuació F1", None,
↳f"{puntuació_F1:.2f} %".replace('.', ','))

# Calcula la suma i el percentatge de cada columna
for col in columnes_check:
    zeros = len(df_revisió_persones2) - df_revisió_persones2[col].sum()
    percentatges = (zeros / len(df_revisió_persones2)) * 100
    nom_columna = col[4:]
    descripció = f"Fiabilitat en analitzar {nom_columna}"
    valor = zeros
    percentatge = f"{percentatges:.2f} %".replace('.', ',')
    desa_estadística(estadístiques_fiabilitat, descripció, valor,
↳percentatge)

# Converteix la llista en un marc de dades
st_fiabilitat = pd.DataFrame(estadístiques_fiabilitat)
if not os.path.exists('st_fiabilitat.csv'):
    st_fiabilitat.to_csv('st_fiabilitat.csv', index=False)

# Mostra el marc de dades

```

```
print(st_fiabilitat)
```

		descripció	valor	percentatge
0		Total de notícies a l'anàlisi	1311	
1		Notícies per comprovar manualment	131	9,99 %
2		Notícies comprovades manualment	131	100,00 %
3		Total de persones a l'anàlisi	4380	
4		Persones per comprovar manualment	566	12,92 %
5		Persones detectades correctament per la IA (ve...	530	
6		Persones eliminades manualment (falsos positius)	36	
7		Persones afegides manualment (falsos negatius)	21	
8		Precisió		93,64 %
9		Reclam		96,19 %
10		Puntuació F1		94,90 %
11		Fiabilitat en analitzar nom	550	99,82 %
12		Fiabilitat en analitzar sexe	536	97,28 %
13		Fiabilitat en analitzar edat	551	100,00 %
14		Fiabilitat en analitzar ocupació	254	46,10 %
15		Fiabilitat en analitzar rol_familiar	548	99,46 %
16		Fiabilitat en analitzar protagonista	529	96,01 %
17		Fiabilitat en analitzar portaveu	536	97,28 %
18		Fiabilitat en analitzar experta	547	99,27 %
19		Fiabilitat en analitzar testimoni	549	99,64 %
20		Fiabilitat en analitzar víctima	549	99,64 %

```
[ ]: # Elimina les variables prescindibles
for variable in ['chk_columnes', 'chk_columns', 'col', 'columnes_check',
↳ 'columnes_comparativa', 'columnes_per_copiar',
↳ 'descripció', 'df_revisió_persones2_per_copiar',
↳ 'estadística', 'estadístiques_fiabilitat', 'falsos_negatius',
↳ 'falsos_positius', 'fila', 'index', 'nom_columna',
↳ 'percentatge', 'percentatge_checked', 'percentatge_persones',
↳ 'percentatge_revisió', 'percentatges', 'precisió',
↳ 'puntuació_F1', 'reclam', 'recompte_tot1', 'row',
↳ 'suma_checked', 'sumes', 'tot1', 'valor',
↳ 'veritables_positius', 'zeros']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable
```

9 Indicadors estadístics sobre la recerca

9.1 Estadístiques generals (*st_generals*)

S'afegeixen noves columnes a *df_notícies* a partir de la informació agrupada de *df_persones*.

```
[ ]: # Agrupa per notícia i sexe i recompta
persones_sexe = df_persones.groupby(['id', 'sexe']).size()
↳ reset_index(name='count')
```

```

# Pivot the grouped dataframe
persones_pivot = persones_sexe.pivot(index='id', columns='sexe',
↳values='count').reset_index()

# Reanomena les columnes
persones_pivot.columns = ['id', 'persones_dona', 'persones_home',
↳'persones_indeterminat']

# Fusiona amb df_notícies
df_notícies = df_notícies.merge(persones_pivot, on='id', how='left')

# Substitueix els NaN per 0
df_notícies.fillna({'persones_dona': 0, 'persones_home': 0,
↳'persones_indeterminat': 0}, inplace=True)

```

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_generals*: - Total de notícies analitzades - Notícies sense persones - Notícies amb persones - Dones totals - Homes totals - Indeterminades totals

S'exporten les següents dades (en valor absolut) al marc de dades *st_generals*: - Persones per notícia (màxim, mínim i mitjana) - Persones per notícia amb persones (mitjana)

```

[ ]: # Calcula els valors absoluts
notícies_total = len(df_notícies)
notícies_sense_persones = len(df_notícies[(df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']) ==
↳0])
notícies_amb_persones = len(df_notícies[(df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']) > 0])
dones_totals = df_notícies['persones_dona'].sum()
homes_totals = df_notícies['persones_home'].sum()
indeterminades_totals = df_notícies['persones_indeterminat'].sum()
max_persones_notícia = (df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']).max()
min_persones_notícia = (df_notícies[(df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']) >
↳0]['persones_dona'] + df_notícies['persones_home'] +
↳df_notícies['persones_indeterminat']).min()
notícies_amb_una_persona = len(df_notícies[(df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']) ==
↳1])
mitjana_persones_per_notícia = (df_notícies['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']).
↳mean()
mitjana_persones_per_notícia_amb_persones =
↳(df_notícies[(df_notícies['persones_dona'] + df_notícies['persones_home']
↳+ df_notícies['persones_indeterminat']) > 0]['persones_dona'] +
↳df_notícies['persones_home'] + df_notícies['persones_indeterminat']).
↳mean()

```



```

# Crea 'st_generals'
st_generals = pd.DataFrame({
    'descripció': ['Total de notícies analitzades', 'Notícies sense_
↳persones', 'Notícies amb persones', 'Dones totals',
                'Homes totals', 'Indeterminades totals', 'Màxim de_
↳persones a una notícia', 'Mínim de persones a una notícia',
                'Notícies amb una sola persona', 'Mitjana de persones per_
↳notícia', 'Mitjana de persones per notícia que esmenta persones'],
    'valor': [notícies_total, notícies_sense_persones,
↳notícies_amb_persones, dones_totals, homes_totals, indeterminades_totals,
                max_persones_notícia, min_persones_notícia,
↳notícies_amb_una_persona, mitjana_persones_per_notícia,
↳mitjana_persones_per_notícia_amb_persones],
    'percentatge': [None] * 11,
})

# Calcula els percentatges
st_generals['percentatge'] = [
    (len(df_notícies) / len(df_notícies)) * 100, # Total de notícies_
↳analitzades
    (notícies_sense_persones / len(df_notícies)) * 100, # Notícies sense_
↳persones
    (notícies_amb_persones / len(df_notícies)) * 100, # Notícies amb_
↳persones
    (dones_totals / len(df_persones)) * 100, # Dones totals
    (homes_totals / len(df_persones)) * 100, # Homes totals
    (indeterminades_totals / len(df_persones)) * 100, # Indeterminades_
↳totals
    None, # Màxim de persones a una notícia
    None, # Mínim de persones a una notícia
    (notícies_amb_una_persona / len(df_notícies)) * 100, # Notícies amb una_
↳sola persona
    None, # Mitjana de persones per notícia
    None # Mitjana de persones per notícia que esmenta persones
]

# Exporta a CSV
if not os.path.exists('st_generals.csv'):
    st_generals.to_csv('st_generals.csv', index=False)

```

9.2 Estadístiques de presència de dones a les notícies (*st_notícies_dones*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_notícies_dones*: - Notícies sense dones - Notícies amb dones - Notícies sense dones per mitjà - Notícies amb dones per mitjà - Notícies sense dones per secció - Notícies amb dones per secció - Notícies sense dones per mitjà i secció - Notícies amb dones per mitjà i secció

```

[ ]: # Calcula variables
notícies_sense_dones = len(df_notícies[df_notícies['persones_dona'] == 0])
notícies_amb_dones = notícies_total - notícies_sense_dones

```



```

# Agrupa per 'mitjà' i 'secció'
agrupa_per_mitjà = df_notícies.groupby('mitjà')
agrupa_per_secció = df_notícies.groupby('secció')
agrupa_per_mitjà_secció = df_notícies.groupby(['mitjà', 'secció'])

# Crea una funció per calcular valors i percentatges
def calcula_dades(df_agrupat, nom_columna, condition=lambda x: x == 0):
    recompte = df_agrupat.apply(lambda x: condition(x[nom_columna]).sum())
    recompte_total = df_agrupat.size()
    percentatge = (recompte / recompte_total) * 100
    return recompte, percentatge

# Calcula valors i percentatges de cada estadística
recompte_sense_dones_mitjà, percentatge_sense_dones_mitjà =
    ↪ calcula_dades(agrupa_per_mitjà, 'persones_dona')
recompte_amb_dones_mitjà, percentatge_amb_dones_mitjà =
    ↪ calcula_dades(agrupa_per_mitjà, 'persones_dona', lambda x: x > 0)
recompte_sense_dones_secció, percentatge_sense_dones_secció =
    ↪ calcula_dades(agrupa_per_secció, 'persones_dona')
recompte_amb_dones_secció, percentatge_amb_dones_secció =
    ↪ calcula_dades(agrupa_per_secció, 'persones_dona', lambda x: x > 0)
recompte_sense_dones_mitjà_secció, percentatge_sense_dones_mitjà_secció =
    ↪ calcula_dades(agrupa_per_mitjà_secció, 'persones_dona')
recompte_amb_dones_mitjà_secció, percentatge_amb_dones_mitjà_secció =
    ↪ calcula_dades(agrupa_per_mitjà_secció, 'persones_dona', lambda x: x > 0)

# Combina les dades
data = {
    'descripció': ['Notícies sense dones', 'Notícies amb dones'],
    'mitjà': ['Total', 'Total'],
    'secció': ['Total', 'Total'],
    'valor': [notícies_sense_dones, notícies_amb_dones],
    'percentatge': [(notícies_sense_dones / notícies_total) * 100,
    ↪ (notícies_amb_dones / notícies_total) * 100]
}

# Converteix a un marc de dades
st_notícies_dones = pd.DataFrame(data[['descripció', 'mitjà', 'secció',
    ↪ 'valor', 'percentatge']])
st_notícies_dones = pd.concat([st_notícies_dones,
    pd.DataFrame({'descripció': 'Notícies sense dones per mitjà', 'mitjà':
    ↪ recompte_sense_dones_mitjà.index, 'secció': 'Total', 'valor':
    ↪ recompte_sense_dones_mitjà, 'percentatge':
    ↪ percentatge_sense_dones_mitjà}),
    pd.DataFrame({'descripció': 'Notícies amb dones per mitjà', 'mitjà':
    ↪ recompte_amb_dones_mitjà.index, 'secció': 'Total', 'valor':
    ↪ recompte_amb_dones_mitjà, 'percentatge': percentatge_amb_dones_mitjà}),

```

```

    pd.DataFrame({'descripció': 'Notícies sense dones per secció', 'mitjà':
↳ 'Total', 'secció': recompte_sense_dones_secció.index, 'valor':
↳ recompte_sense_dones_secció, 'percentatge':
↳ percentatge_sense_dones_secció}),
    pd.DataFrame({'descripció': 'Notícies amb dones per secció', 'mitjà':
↳ 'Total', 'secció': recompte_amb_dones_secció.index, 'valor':
↳ recompte_amb_dones_secció, 'percentatge': percentatge_amb_dones_secció}),
    pd.DataFrame({'descripció': 'Notícies sense dones per mitjà i secció',
↳ 'mitjà': recompte_sense_dones_mitjà_secció.index.get_level_values(0),
↳ 'secció': percentatge_sense_dones_mitjà_secció.index.get_level_values(1),
↳ 'valor': recompte_sense_dones_mitjà_secció, 'percentatge':
↳ percentatge_sense_dones_mitjà_secció}),
    pd.DataFrame({'descripció': 'Notícies amb dones per mitjà i secció',
↳ 'mitjà': recompte_amb_dones_mitjà_secció.index.get_level_values(0),
↳ 'secció': percentatge_amb_dones_mitjà_secció.index.get_level_values(1),
↳ 'valor': recompte_amb_dones_mitjà_secció, 'percentatge':
↳ percentatge_amb_dones_mitjà_secció})
], ignore_index=True)

# Exporta a CSV
if not os.path.exists('st_notícies_dones.csv'):
    st_notícies_dones.to_csv('st_notícies_dones.csv', index=False)

```

9.3 Estadístiques per ocupació (*st_ocupacions*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_ocupacions*:

- Dones esmentades per ocupació - Homes esmentats per ocupació - Indeterminades esmentades per ocupació
- Dones esmentades per ocupació i mitjà - Homes esmentats per ocupació i mitjà
- Indeterminades esmentades per ocupació i mitjà - Dones esmentades per ocupació i secció - Homes esmentats per ocupació i secció - Indeterminades esmentades per ocupació i secció

```

[ ]: # Fusiona df_persones i df_notícies per afegir les columnes 'mitjà' i
↳ 'secció'
join_ocupacions = df_persones.merge(df_notícies[['id', 'mitjà', 'secció']],
↳ on='id', how='left')

# Crea una funció per als càlculs
def calcula_dades_ocupacions(df_agrupat, base=None):
    recompte = df_agrupat.size()
    if base is not None:
        percentatge = (recompte / base.reindex(recompte.index, level=0) *
↳ 100).fillna(0)
    else:
        percentatge = (recompte / recompte.sum() * 100).fillna(0)
    return recompte, percentatge

# Assigna els valors de la categoria d'ocupació
# Es tradueixen i resumeixen els textos de df_categoriaocupació
mapatge_ocupació = {
    0: "No esmentada",
    1: "Família reial",

```

```

2: "Classe política",
3: "Funcionariat",
4: "Policia, bombers i forces armades",
5: "Professorat",
6: "Medicina o odontologia",
7: "Altres serveis socio-sanitaris",
8: "Ciència o enginyeria",
9: "Mitjans de comunicació",
10: "Sector jurídic",
11: "Sector empresarial",
12: "Treballadors de serveis",
13: "Comerç i artesania",
14: "Sector primari",
15: "Clergat",
16: "ONG, sindicats i associacions",
17: "Treball sexual",
18: "Celebritats i artistes",
19: "Esports",
20: "Estudiantat",
21: "Treball domèstic",
22: "Infància",
23: "Resident o habitant",
24: "Jubilades",
25: "Criminals o sospitoses",
26: "Aturades",
27: "Altres"
}

# Executa els càlculs
dades = []
for sexe in ['dona', 'home', 'indeterminat']:
    join_sexe = join_ocupacions[join_ocupacions['sexe'] == sexe]

    # Per ocupació
    grup = join_sexe.groupby('categoria_ocupació')
    recompte, percentatge = calcula_dades_ocupacions(grup)
    dades.append(pd.DataFrame({'descripció': f'{sexe.capitalize()} per',
    ↳ocupació', 'mitjà': 'Total', 'secció': 'Total', 'categoria_ocupació':,
    ↳recompte.index, 'valor': recompte, 'percentatge': percentatge}))

    # Per mitjà i ocupació
    grup_mitjà = join_sexe.groupby(['mitjà', 'categoria_ocupació'])
    total_mitjà = join_sexe.groupby('mitjà').size()
    recompte_mitjà, percentatge_mitjà = calcula_dades_ocupacions(grup_mitjà,
    ↳total_mitjà)
    dades.append(pd.DataFrame({'descripció': f'{sexe.capitalize()} per mitjà',
    ↳i ocupació', 'mitjà': recompte_mitjà.index.get_level_values(0), 'secció':,
    ↳'Total', 'categoria_ocupació': recompte_mitjà.index.get_level_values(1),
    ↳'valor': recompte_mitjà, 'percentatge': percentatge_mitjà}))

```

```

# Per secció i ocupació
grup_secció = join_sexe.groupby(['secció', 'categoria_ocupació'])
total_secció = join_sexe.groupby('secció').size()
recompte_secció, percentatge_secció =
↳calcula_dades_ocupacions(grup_secció, total_secció)
dades.append(pd.DataFrame({'descripció': f'{sexe.capitalize()} per
↳secció i ocupació', 'mitjà': 'Total', 'secció': recompte_secció.index.
↳get_level_values(0), 'categoria_ocupació': recompte_secció.index.
↳get_level_values(1), 'valor': recompte_secció, 'percentatge':
↳percentatge_secció}))

# Afegeix la descripció de l'ocupació
for codi_ocupació in dades:
    codi_ocupació['desc_categoria_ocupació'] =
↳codi_ocupació['categoria_ocupació'].map(mapatge_ocupació)

# Desa les estadístiques al marc de dades
st_ocupacions = pd.concat(dades, ignore_index=True)[['descripció', 'mitjà',
↳'secció', 'categoria_ocupació', 'desc_categoria_ocupació', 'valor',
↳'percentatge']]

# Exporta a CSV
if not os.path.exists('st_ocupacions.csv'):
    st_ocupacions.to_csv('st_ocupacions.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['agrupa_per_mitjà', 'agrupa_per_mitjà_secció',
↳'agrupa_per_secció', 'agrupació', 'codi_ocupació', 'dades', 'data',
↳'dones_totals', 'grup', 'grup_mitjà', 'grup_secció',
↳'homes_totals', 'indeterminades_totals', 'join_ocupacions', 'join_sexe',
↳'mapatge_ocupació', 'max_persones_notícia',
↳'min_persones_notícia', 'mitjana_persones_per_notícia',
↳'mitjana_persones_per_notícia_amb_persones',
↳'nom_agrupació', 'notícies_amb_dones', 'notícies_amb_persones',
↳'notícies_amb_una_persona',
↳'notícies_sense_dones', 'notícies_sense_persones',
↳'notícies_total', 'ocupacions_temporal',
↳'percentatge', 'percentatge_amb_dones_mitjà',
↳'percentatge_amb_dones_mitjà_secció', 'percentatge_amb_dones_secció',
↳'percentatge_mitjà', 'percentatge_secció',
↳'percentatge_sense_dones_mitjà',
↳'percentatge_sense_dones_mitjà_secció',
↳'percentatge_sense_dones_secció',
↳'persones_pivot', 'persones_sexe', 'recompte',
↳'recompte_amb_dones_mitjà', 'recompte_amb_dones_mitjà_secció',
↳'recompte_amb_dones_secció',
↳'recompte_mitjà', 'recompte_secció',
↳'recompte_sense_dones_mitjà', 'recompte_sense_dones_mitjà_secció',
↳'recompte_sense_dones_secció',
↳'sexe', 'total_mitjà', 'total_secció', 'total_sexe']:

```

```

try:
    del globals()[variable]
except KeyError:
    pass
del variable

```

9.4 Estadístiques per rol familiar (*st_rolfamiliar*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_ocupacions*: - Dones amb rol familiar esmentat. - Dones sense rol familiar esmentat. - Homes amb rol familiar esmentat. - Homes sense rol familiar esmentat. - Indeterminades amb rol familiar esmentat. - Indeterminades sense rol familiar esmentat.

```

[ ]: # Crea una funció per als càlculs
def calcula_dades_rolfamiliar(sexe, rol_familiar):
    total_persones = len(df_persones[df_persones['sexe'] == sexe])
    recompte = len(df_persones[(df_persones['sexe'] == sexe) &
    ↪(df_persones['rol_familiar'] == rol_familiar)])
    percentatge = (recompte / total_persones) * 100 if total_persones != 0
    ↪else 0
    return recompte, percentatge
df_rolfamiliar = pd.read_csv('rolfamiliar.csv')
df_temporal = df_persones.merge(df_rolfamiliar[['id', 'nom']], on=['id',
    ↪'nom'], how='left', indicator=True)
màscara = df_temporal['_merge'] == 'both'
df_persones.loc[màscara, 'rol_familiar'] = 1

# Executa els càlculs
dades = [
    {"descripció": "Dones amb rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('dona', 1)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('dona', 1)[1]},
    {"descripció": "Dones sense rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('dona', 0)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('dona', 0)[1]},
    {"descripció": "Homes amb rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('home', 1)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('home', 1)[1]},
    {"descripció": "Homes sense rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('home', 0)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('home', 0)[1]},
    {"descripció": "Indeterminades amb rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('indeterminat', 1)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('indeterminat', 1)[1]},
    {"descripció": "Indeterminades sense rol familiar esmentat", "valor":
    ↪calcula_dades_rolfamiliar('indeterminat', 0)[0], "percentatge":
    ↪calcula_dades_rolfamiliar('indeterminat', 0)[1]},
]

# Crea el marc de dades

```

```

st_rolfamiliar = pd.DataFrame(dades)

# Exporta a CSV
if not os.path.exists('st_rolfamiliar.csv'):
    st_rolfamiliar.to_csv('st_rolfamiliar.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['dades', 'df_rolfamiliar', 'df_temporal', 'màscara']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

9.5 Estadístiques sobre la funció a la notícia (*st_funció*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_funció*:

- Dones amb funció esmentada - Homes amb funció esmentada - Indeterminades amb funció esmentada - Dones sense funció esmentada - Homes sense funció esmentada - Indeterminades sense funció esmentada - Protagonistes total - Protagonistes dones - Protagonistes homes - Protagonistes indeterminades - Portaveus total - Portaveus dones - Portaveus homes - Portaveus indeterminades - Expertes dones - Experts homes - Expertes indeterminades - Testimonis dones - Testimonis homes - Testimonis indeterminades - Víctimes dones - Víctimes homes - Víctimes indeterminades

```

[ ]: # Crea dues funcions per recomptar les persones amb o sense funció
def calcula_dades_funció(funció, sexe, descripció):
    if sexe == 'total':
        subconjunt = df_persones[df_persones[funció] == 1]
        total_sexe = len(subconjunt)
    else:
        subconjunt = df_persones[(df_persones['sexe'] == sexe) &
↳(df_persones[funció] == 1)]
        total_sexe = len(df_persones[(df_persones[funció] == 1)])
        recompte = len(subconjunt)
        percentatge = (recompte / total_sexe) * 100 if total_sexe != 0 else 0
        return {"descripció": descripció, "sexe": sexe, "valor": recompte,
↳"percentatge": percentatge}

def calcula_dades_sense_funció(sexe, descripció):
    if sexe == 'total':
        subconjunt = df_persones
    else:
        subconjunt = df_persones[df_persones['sexe'] == sexe]

    if descripció == "Amb funció esmentada a la notícia":
        recompte = len(subconjunt[subconjunt[['protagonista', 'portaveu',
↳'experta', 'testimoni', 'víctima']].sum(axis=1) > 0])
        total_sexe = len(df_persones[df_persones[['protagonista',
↳'portaveu', 'experta', 'testimoni', 'víctima']].sum(axis=1) > 0])
    else: # "Sense funció esmentada a la notícia"

```

```

    recompte = len(subconjunt[subconjunt[['protagonista', 'portaveu',
↳'experta', 'testimoni', 'víctima']].sum(axis=1) == 0])
    total_sexe = len(df_persones[df_persones[['protagonista',
↳'portaveu', 'experta', 'testimoni', 'víctima']].sum(axis=1) == 0])

    percentatge = (recompte / total_sexe) * 100 if total_sexe != 0 else 0
    return {"descripció": descripció, "sexe": sexe, "valor": recompte,
↳"percentatge": percentatge}

# Mapatge dels textos per a la columna 'descripció'
textos_descripció = {
    'protagonista': 'Protagonistes',
    'portaveu': 'Portaveus',
    'experta': 'Expertes',
    'testimoni': 'Testimonis',
    'víctima': 'Víctimes'
}
sexes = df_persones['sexe'].unique().tolist() + ['total']
altres_descripció = ["Amb funció esmentada a la notícia", "Sense funció,
↳esmentada a la notícia"]

# Executa els càlculs
dades = []
for descripció in altres_descripció:
    for sexe in sexes:
        dades.append(calcula_dades_sense_funció(sexe, descripció))

for funció, desc in textos_descripció.items():
    for sexe in sexes:
        dades.append(calcula_dades_funció(funció, sexe, desc))

# Desa el marc de dades
st_funció = pd.DataFrame(dades)

# Exporta a CSV
if not os.path.exists('st_funció.csv'):
    st_funció.to_csv('st_funció.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['dades', 'dades_expertes', 'dades_portaveus',
↳'dades_protagonistes', 'dades_testimonis', 'dades_víctimes',
↳'altres_descripció', 'desc', 'descripció', 'funció',
↳'percentatge_qualsevol_funció', 'percentatge_sense_funció',
↳'recompta_qualsevol_funció', 'recompta_sense_funció',
↳'sexe', 'sexes', 'textos_descripció', 'total_percentatge',
↳'total_persones', 'total_recompte']:
    try:
        del globals()[variable]
    except KeyError:

```

```
pass
del variable
```

9.6 Estadístiques dels elements de titulació (*st_titulació*)

Primer, es creen noves columnes a *df_notícies*.

```
[ ]: # Transforma a dades numèriques
conversió = ['dins_títol', 'dins_subtítol', 'dins_text']
df_persones[conversió] = df_persones[conversió].apply(pd.to_numeric,
↳errors='coerce')

# Agrupa per 'id' i 'sexe' i suma les columnes
columnes_per_agrupar = ['id', 'sexe'] + conversió
agrupació = df_persones[columnes_per_agrupar].groupby(['id', 'sexe']).sum().
↳reset_index()

# Crea una llista per aplicar transformacions a les columnes noves
columnes_noves = []

# Executa la suma per a cada element de la notícia
for element in ['títol', 'subtítol', 'text']:
    suma = agrupació.pivot(index='id', columns='sexe',
↳values=f'dins_{element}')
    suma.columns = [f'{element}_{col}' for col in suma.columns]
    suma[f'{element}_total'] = suma.sum(axis=1)

    columnes_noves.extend(suma.columns)

df_notícies = df_notícies.merge(suma, on='id', how='left')

# Umpl els buits amb zeros
df_notícies[columnes_noves] = df_notícies[columnes_noves].fillna(0)

# Converteix les dades a enters
df_notícies = df_notícies.astype({col: int for col in columnes_noves})
```

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_titulació*: - Dones en títols i subtítols - Homes en títols i subtítols - Indeterminades en títols i subtítols - Dones en títols i subtítols per mitjà - Homes en títols i subtítols per mitjà - Indeterminades en títols i subtítols per mitjà

```
[ ]: # Calcula els totals
total_notícies = len(df_notícies)
suma_títols_subtítols = df_notícies[['títol_dona', 'subtítol_dona',
↳'títol_home', 'subtítol_home', 'títol_indeterminat',
↳'subtítol_indeterminat']].sum().sum()

# Crea un conjunt amb les primeres dades
dades = [
```



```

    {'descripció': 'Persones en títols i subtítols', 'mitjà': 'Total',
    ↪'valor': suma_títols_subtítols, 'percentatge': 100},
    {'descripció': 'Dones en títols i subtítols', 'mitjà': 'Total', 'valor':
    ↪df_notícies[['títol_dona', 'subtítol_dona']].sum().sum(), 'percentatge':
    ↪(df_notícies[['títol_dona', 'subtítol_dona']].sum().sum() /
    ↪suma_títols_subtítols * 100) if suma_títols_subtítols > 0 else 0},
    {'descripció': 'Homes en títols i subtítols', 'mitjà': 'Total', 'valor':
    ↪df_notícies[['títol_home', 'subtítol_home']].sum().sum(), 'percentatge':
    ↪(df_notícies[['títol_home', 'subtítol_home']].sum().sum() /
    ↪suma_títols_subtítols * 100) if suma_títols_subtítols > 0 else 0},
    {'descripció': 'Indeterminades en títols i subtítols', 'mitjà': 'Total',
    ↪'valor': df_notícies[['títol_indeterminat', 'subtítol_indeterminat']].
    ↪sum().sum(), 'percentatge': (df_notícies[['títol_indeterminat',
    ↪'subtítol_indeterminat']].sum().sum() / suma_títols_subtítols * 100) if
    ↪suma_títols_subtítols > 0 else 0}
]

# Calcula estadístiques per mitjà
for mitjà in df_notícies['mitjà'].unique():
    subconjunt = df_notícies[df_notícies['mitjà'] == mitjà]
    suma_per_mitjà = subconjunt[['títol_dona', 'subtítol_dona',
    ↪'títol_home', 'subtítol_home', 'títol_indeterminat',
    ↪'subtítol_indeterminat']].sum().sum()
    dades.append({
        'descripció': 'Persones en títols i subtítols',
        'mitjà': mitjà,
        'valor': suma_per_mitjà,
        'percentatge': (suma_per_mitjà / suma_títols_subtítols * 100)
    })

# Calcula estadístiques per sexe i mitjà
for sexe, nom_columna in [('Dones', 'dona'), ('Homes', 'home'),
    ↪('Indeterminades', 'indeterminat')]:
    for mitjà in df_notícies['mitjà'].unique():
        subconjunt = df_notícies[df_notícies['mitjà'] == mitjà]
        suma_per_mitjà = subconjunt[[f'títol_{nom_columna}',
    ↪f'subtítol_{nom_columna}']].sum().sum()
        total_mitjà = subconjunt[['títol_dona', 'subtítol_dona',
    ↪'títol_home', 'subtítol_home', 'títol_indeterminat',
    ↪'subtítol_indeterminat']].sum().sum()

        dades.append({
            'descripció': f'{sexe} en títols i subtítols',
            'mitjà': mitjà,
            'valor': suma_per_mitjà,
            'percentatge': (suma_per_mitjà / total_mitjà * 100) if
    ↪total_mitjà > 0 else 0
        })

# Crea el marc de dades

```

```

st_titulació = pd.DataFrame(dades)

# Exporta a CSV
if not os.path.exists('st_titulació.csv'):
    st_titulació.to_csv('st_titulació.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['agrupació', 'base_percentatges', 'categoria',
↳ 'columnes_noves', 'columnes_per_agrupar', 'conversió', 'dades', 'element',
↳ 'estadístiques_titulació', 'mitjà', 'nom_columna',
↳ 'percentatge_per_mitjà', 'sexe', 'subconjunt', 'suma', 'suma_per_mitjà',
↳ 'suma_títols_subtítols', 'total_mitjà', 'total_notícies']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

9.7 Estadístiques d'autoria de les notícies (*st_firmes*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_firmes*:

- Notícies escrites per periodista o periodistes dones
- Notícies escrites per periodista o periodistes homes
- Notícies escrites per periodista o periodistes indeterminats (agències, redacció, sigles...)
- Notícies escrites per periodistes d'ambdós sexes
- Notícies escrites per periodista o periodistes dones per mitjà
- Notícies escrites per periodista o periodistes homes per mitjà
- Notícies escrites per periodista o periodistes indeterminats per mitjà
- Notícies escrites per periodistes d'ambdós sexes per mitjà
- Notícies escrites per periodista o periodistes dones per secció
- Notícies escrites per periodista o periodistes homes per secció
- Notícies escrites per periodista o periodistes indeterminats per secció
- Notícies escrites per periodistes d'ambdós sexes per secció
- Notícies escrites per periodista o periodistes dones per mitjà i secció
- Notícies escrites per periodista o periodistes homes per mitjà i secció
- Notícies escrites per periodista o periodistes indeterminats per mitjà i secció
- Notícies escrites per periodistes d'ambdós sexes per mitjà i secció

```

[ ]: # Crea una funció
def calcula_dades_firmes(df):
    total_notícies = len(df)

    # Notícies escrites per periodista o periodistes dones
    cond_dona = (df['firma1_dona'] == 1) & (df['firma2_dona'].isin([1, '']))
    valor_dona = cond_dona.sum()
    percentatge_dona = (valor_dona / total_notícies) * 100

    # Notícies escrites per periodista o periodistes homes
    cond_home = (df['firma1_home'] == 1) & (df['firma2_home'].isin([1, '']))
    valor_home = cond_home.sum()
    percentatge_home = (valor_home / total_notícies) * 100

    # Notícies escrites per periodista o periodistes indeterminats
    cond_indeterminat = df['firma1_indeterminat'] == 1
    valor_indeterminat = cond_indeterminat.sum()
    percentatge_indeterminat = (valor_indeterminat / total_notícies) * 100

```

```

# Notícies escrites per periodista o periodistes d'ambdós sexes
cond_mixt = ((df['firma1_dona'] == 1) & (df['firma2_home'].isin([1]) |
↳df['firma2_indeterminat'].isin([1]))) | \
            ((df['firma1_home'] == 1) & (df['firma2_dona'].isin([1]) |
↳df['firma2_indeterminat'].isin([1])))
valor_mixt = cond_mixt.sum()
percentatge_mixt = (valor_mixt / total_notícies) * 100

# Crea el marc de dades
st_firmes = pd.DataFrame({
    'descripció': [
        'Notícies escrites per periodista o periodistes dones',
        'Notícies escrites per periodista o periodistes homes',
        'Notícies escrites per periodista o periodistes indeterminats',
        'Notícies escrites per periodistes d\'ambdós sexes'
    ],
    'mitjà': ['Total'] * 4,
    'secció': ['Total'] * 4,
    'valor': [valor_dona, valor_home, valor_indeterminat, valor_mixt],
    'percentatge': [percentatge_dona, percentatge_home,
↳percentatge_indeterminat, percentatge_mixt]
})

return st_firmes

# Executa la funció
st_firmes = calcula_dades_firmes(df_notícies)

```

```

[ ]: # Crea una segona funció
def calcula_dades_firmes_per_mitjà(df):
    resultats = []
    for mitjà in df['mitjà'].unique():
        df_mitjà = df[df['mitjà'] == mitjà]
        total_notícies_mitjà = len(df_mitjà)

        valor_dona = ((df_mitjà['firma1_dona'] == 1) &
↳(df_mitjà['firma2_dona'].isin([1, '']))).sum()
        percentatge_dona = (valor_dona / total_notícies_mitjà) * 100 if
↳total_notícies_mitjà > 0 else 0

        valor_home = ((df_mitjà['firma1_home'] == 1) &
↳(df_mitjà['firma2_home'].isin([1, '']))).sum()
        percentatge_home = (valor_home / total_notícies_mitjà) * 100 if
↳total_notícies_mitjà > 0 else 0

        valor_indeterminat = (df_mitjà['firma1_indeterminat'] == 1).sum()
        percentatge_indeterminat = (valor_indeterminat /
↳total_notícies_mitjà) * 100 if total_notícies_mitjà > 0 else 0

```

```

        valor_mixt = (((df_mitjà['firma1_dona'] == 1) &
↳(df_mitjà['firma2_home'].isin([1]) | df_mitjà['firma2_indeterminat'].
↳isin([1]))) | \
                ((df_mitjà['firma1_home'] == 1) &
↳(df_mitjà['firma2_dona'].isin([1]) | df_mitjà['firma2_indeterminat'].
↳isin([1])))
        ).sum()
        percentatge_mixt = (valor_mixt / total_notícies_mitjà) * 100 if
↳total_notícies_mitjà > 0 else 0

    resultats.extend([
        {'descripció': 'Notícies escrites per periodista o periodistes
↳dones per mitjà', 'mitjà': mitjà, 'secció': 'Total', 'valor': valor_dona,
↳'percentatge': percentatge_dona},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳homes per mitjà', 'mitjà': mitjà, 'secció': 'Total', 'valor': valor_home,
↳'percentatge': percentatge_home},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳indeterminats per mitjà', 'mitjà': mitjà, 'secció': 'Total', 'valor':
↳valor_indeterminat, 'percentatge': percentatge_indeterminat},
        {'descripció': 'Notícies escrites per periodistes d\'ambdós
↳sexes per mitjà', 'mitjà': mitjà, 'secció': 'Total', 'valor': valor_mixt,
↳'percentatge': percentatge_mixt}
    ])

    return pd.DataFrame(resultats)

# Executa la funció
st_firmes_mitjà = calcula_dades_firmes_per_mitjà(df_notícies)

# Afegeix les files
st_firmes = pd.concat([st_firmes, st_firmes_mitjà], ignore_index=True)

```

```

[ ]: # Crea una tercera funció
def calcula_dades_firmes_per_secció(df):
    resultats = []

    for secció in df['secció'].unique():
        df_secció = df[df['secció'] == secció]
        total_notícies_secció = len(df_secció)

        valor_dona = ((df_secció['firma1_dona'] == 1) &
↳(df_secció['firma2_dona'].isin([1, ''])))
        .sum()
        percentatge_dona = (valor_dona / total_notícies_secció) * 100 if
↳total_notícies_secció > 0 else 0

        valor_home = ((df_secció['firma1_home'] == 1) &
↳(df_secció['firma2_home'].isin([1, ''])))
        .sum()
        percentatge_home = (valor_home / total_notícies_secció) * 100 if
↳total_notícies_secció > 0 else 0

```

```

        valor_indeterminat = (df_secció['firma1_indeterminat'] == 1).sum()
        percentatge_indeterminat = (valor_indeterminat /
↳total_notícies_secció) * 100 if total_notícies_secció > 0 else 0

        valor_mixt = (((df_secció['firma1_dona'] == 1) &
↳(df_secció['firma2_home'].isin([1]) | df_secció['firma2_indeterminat'].
↳isin([1]))) | \
                ((df_secció['firma1_home'] == 1) &
↳(df_secció['firma2_dona'].isin([1]) | df_secció['firma2_indeterminat'].
↳isin([1]))))).sum()
        percentatge_mixt = (valor_mixt / total_notícies_secció) * 100 if
↳total_notícies_secció > 0 else 0

    resultats.extend([
        {'descripció': 'Notícies escrites per periodista o periodistes
↳dones per secció', 'mitjà': 'Total', 'secció': secció, 'valor':
↳valor_dona, 'percentatge': percentatge_dona},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳homes per secció', 'mitjà': 'Total', 'secció': secció, 'valor':
↳valor_home, 'percentatge': percentatge_home},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳indeterminats per secció', 'mitjà': 'Total', 'secció': secció, 'valor':
↳valor_indeterminat, 'percentatge': percentatge_indeterminat},
        {'descripció': 'Notícies escrites per periodistes d\'ambdós
↳sexes per secció', 'mitjà': 'Total', 'secció': secció, 'valor':
↳valor_mixt, 'percentatge': percentatge_mixt}
    ])

    return pd.DataFrame(resultats)

# Executa la funció
st_firmes_secció = calcula_dades_firmes_per_secció(df_notícies)

# Afegeix les files
st_firmes = pd.concat([st_firmes, st_firmes_secció], ignore_index=True)

```

```

[ ]: # Crea una quarta funció
def calcula_dades_firmes_per_mitjà_i_secció(df):
    resultats = []

    for mitjà, secció in df[['mitjà', 'secció']].drop_duplicates().
↳itertuples(index=False):
        df_filtrat = df[(df['mitjà'] == mitjà) & (df['secció'] == secció)]
        total_notícies = len(df_filtrat)

        valor_dona = ((df_filtrat['firma1_dona'] == 1) &
↳(df_filtrat['firma2_dona'].isin([1, '']))).sum()
        percentatge_dona = (valor_dona / total_notícies) * 100 if
↳total_notícies > 0 else 0

```

```

    valor_home = ((df_filtrat['firma1_home'] == 1) &
↳(df_filtrat['firma2_home'].isin([1, ''])).sum()
    percentatge_home = (valor_home / total_notícies) * 100 if
↳total_notícies > 0 else 0

    valor_indeterminat = (df_filtrat['firma1_indeterminat'] == 1).sum()
    percentatge_indeterminat = (valor_indeterminat / total_notícies) *
↳100 if total_notícies > 0 else 0

    valor_mixt = (((df_filtrat['firma1_dona'] == 1) &
↳(df_filtrat['firma2_home'].isin([1]) | df_filtrat['firma2_indeterminat'].
↳isin([1]))) | \
                    ((df_filtrat['firma1_home'] == 1) &
↳(df_filtrat['firma2_dona'].isin([1]) | df_filtrat['firma2_indeterminat'].
↳isin([1]))))).sum()
    percentatge_mixt = (valor_mixt / total_notícies) * 100 if
↳total_notícies > 0 else 0

    resultats.extend([
        {'descripció': 'Notícies escrites per periodista o periodistes
↳dones per mitjà i secció', 'mitjà': mitjà, 'secció': secció, 'valor':
↳valor_dona, 'percentatge': percentatge_dona},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳homes per mitjà i secció', 'mitjà': mitjà, 'secció': secció, 'valor':
↳valor_home, 'percentatge': percentatge_home},
        {'descripció': 'Notícies escrites per periodista o periodistes
↳indeterminats per mitjà i secció', 'mitjà': mitjà, 'secció': secció,
↳'valor': valor_indeterminat, 'percentatge': percentatge_indeterminat},
        {'descripció': 'Notícies escrites per periodistes d\'ambdós
↳sexes per mitjà i secció', 'mitjà': mitjà, 'secció': secció, 'valor':
↳valor_mixt, 'percentatge': percentatge_mixt}
    ])

    return pd.DataFrame(resultats)

# Executa la funció
st_firmes_mitjà_secció = calcula_dades_firmes_per_mitjà_i_secció(df_notícies)

# Afegeix les files
st_firmes = pd.concat([st_firmes, st_firmes_mitjà_secció], ignore_index=True)

```

```

[ ]: # Exporta a CSV
if not os.path.exists('st_firmes.csv'):
    st_firmes.to_csv('st_firmes.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['st_firmes_mitjà', 'st_firmes_mitjà_secció',
↳'st_firmes_secció', 'total_notícies']:
    try:
        del globals()[variable]

```

```

except KeyError:
    pass
del variable

```

9.8 Estadístiques de biaix de gènere segons l'autoria (*st_biaixgènere*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_biaixgènere*:

- Dones en notícies segons el sexe de l'autoria
- Homes en notícies segons el sexe de l'autoria
- Indeterminades en notícies segons el sexe de l'autoria
- Expressió del rol familiar en notícies segons el sexe de l'autoria
- Funció en notícies segons el sexe de l'autoria

```

[ ]: # Assigna valors de gènere a les firmes de cada article
df_notícies['firma_sexe'] = ''
df_notícies.loc[(df_notícies['firma1_dona'] == 1) &
↳(df_notícies['firma2_dona'].isin([1, ''])), 'firma_sexe'] = 'dones'
df_notícies.loc[(df_notícies['firma1_home'] == 1) &
↳(df_notícies['firma2_home'].isin([1, ''])), 'firma_sexe'] = 'homes'
df_notícies.loc[df_notícies['firma1_indeterminat'] == 1, 'firma_sexe'] =
↳'indeterminats'
df_notícies.loc[((df_notícies['firma1_dona'] == 1) &
↳(df_notícies['firma2_home'].isin([1]) |
↳df_notícies['firma2_indeterminat'].isin([1]))) |
↳((df_notícies['firma1_home'] == 1) &
↳(df_notícies['firma2_dona'].isin([1]) |
↳df_notícies['firma2_indeterminat'].isin([1]))), 'firma_sexe'] = 'mixtes'

# Fusiona temporalment df_persones i df_notícies
df_biaix = pd.merge(df_persones, df_notícies[['id', 'firma_sexe']], on='id')

# Calcula les dades
resultats = []
for tipus in ['dones', 'homes', 'indeterminats', 'mixtes']:
    for sexe in df_persones['sexe'].unique():
        total_sexe = df_biaix[df_biaix['firma_sexe'] == tipus]['sexe'].
↳count()
        valor = df_biaix[(df_biaix['firma_sexe'] == tipus) &
↳(df_biaix['sexe'] == sexe)]['sexe'].count()
        percentatge = (valor / total_sexe) * 100 if total_sexe > 0 else 0
        descripcio = f'{sexe.capitalize()} en notícies escrites per
↳periodista o periodistes {tipus}'
        resultats.append({
            'descripció': descripcio,
            'firma': tipus,
            'paràmetre': '',
            'sexe': sexe,
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix les dades al marc de dades

```

```
st_biaixgènere = pd.DataFrame(resultats)
```

```
[ ]: # Columnes d'anàlisi
categories = ['rol_familiar', 'protagonista', 'portaveu', 'experta',
↳ 'testimoni', 'víctima']

# Calcula les dades per a cada categoria
nous_resultats = []
for categoria in categories:
    for tipus in ['dones', 'homes', 'indeterminats', 'mixtes']:
        # Total count for each 'firma_sexe' and category
        total_firma_categoria = df_biaix[(df_biaix['firma_sexe'] == tipus) &
↳ (df_biaix[categoria] == 1)].shape[0]

        for sexe in df_persones['sexe'].unique():
            # Filter based on 'firma_sexe', 'sexe', and category == 1
            valor = df_biaix[(df_biaix['firma_sexe'] == tipus) &
↳ (df_biaix['sexe'] == sexe) & (df_biaix[categoria] == 1)].shape[0]
            percentatge = (valor / total_firma_categoria) * 100 if
↳ total_firma_categoria > 0 else 0
            descripcio = f'{categoria.capitalize()} {sexe} en notícies
↳ escrites per periodista o periodistes {tipus}'

            nous_resultats.append({
                'descripció': descripcio,
                'firma': tipus,
                'paràmetre': categoria,
                'sexe': sexe,
                'valor': valor,
                'percentatge': percentatge
            })

# Afegeix les dades al marc de dades
nous_st_biaixgènere = pd.DataFrame(nous_resultats)

# Combina els nous resultats amb els que ja hi ha a st_biaixgènere
st_biaixgènere = pd.concat([st_biaixgènere, nous_st_biaixgènere],
↳ ignore_index=True)
```

```
[ ]: # Exporta a CSV
if not os.path.exists('st_biaixgènere.csv'):
    st_biaixgènere.to_csv('st_biaixgènere.csv', index=False)
```

```
[ ]: # Elimina les variables prescindibles
for variable in ['categoria', 'categories', 'descripcio', 'df_biaix',
↳ 'df_filtrat', 'df_filtrat_sexe',
                'nous_resultats', 'nous_st_biaixgènere', 'percentatge',
↳ 'resultats', 'sexe', 'tipus',
                'total_categoria', 'total_firma', 'total_firma_categoria',
↳ 'total_sexe',
```



```

        'total_sexe_categoria', 'valor']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

9.9 Estadístiques de fotografies (*st_imatges*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_imatges*: - Fotografies amb persones - Fotografies sense persones - Fotografies amb dones - Fotografies sense dones - Fotografies amb homes - Fotografies sense homes - Fotografies amb indeterminades - Fotografies sense indeterminades

```

[ ]: # Converteix les dades a enters
df_imatges['imatge_persones'] = pd.to_numeric(df_imatges['imatge_persones'],
↳errors='coerce').fillna(0).astype(int)
df_imatges['imatge_dones'] = pd.to_numeric(df_imatges['imatge_dones'],
↳errors='coerce').fillna(0).astype(int)
df_imatges['imatge_homes'] = pd.to_numeric(df_imatges['imatge_homes'],
↳errors='coerce').fillna(0).astype(int)
df_imatges['imatge_indeterminades'] = pd.
↳to_numeric(df_imatges['imatge_indeterminades'], errors='coerce').
↳fillna(0).astype(int)

# Calcula les estadístiques
total_imatges = len(df_imatges)
total_amb_persones = df_imatges[df_imatges['imatge_persones'] > 0].shape[0]
categories = {
    'Fotografies amb persones': df_imatges['imatge_persones'] > 0,
    'Fotografies sense persones': df_imatges['imatge_persones'] == 0,
    'Fotografies amb dones': (df_imatges['imatge_dones'] > 0) &
↳(df_imatges['imatge_persones'] > 0),
    'Fotografies sense dones': (df_imatges['imatge_dones'] == 0) &
↳(df_imatges['imatge_persones'] > 0),
    'Fotografies amb homes': (df_imatges['imatge_homes'] > 0) &
↳(df_imatges['imatge_persones'] > 0),
    'Fotografies sense homes': (df_imatges['imatge_homes'] == 0) &
↳(df_imatges['imatge_persones'] > 0),
    'Fotografies amb indeterminades': (df_imatges['imatge_indeterminades'] >
↳0) & (df_imatges['imatge_persones'] > 0),
    'Fotografies sense indeterminades': (df_imatges['imatge_indeterminades']
↳== 0) & (df_imatges['imatge_persones'] > 0)
}
resultats = []
for descripcio, condicio in categories.items():
    valor = df_imatges[condicio].shape[0]
    if 'persones' in descripcio:
        percentatge = (valor / total_imatges) * 100
    else:

```

```

        percentatge = (valor / total_amb_persones) * 100 if
↳total_amb_persones > 0 else 0
        resultats.append({'descripció': descripcio, 'valor': valor,
↳'percentatge': percentatge})

# Crea el marc de dades
st_imatges = pd.DataFrame(resultats)

# Exporta a CSV
if not os.path.exists('st_imatges.csv'):
    st_imatges.to_csv('st_imatges.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['categories', 'condicio', 'descripcio', 'percentatge',
↳'resultats', 'total_amb_persones', 'total_imatges', 'valor']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

9.10 Estadístiques d'edat (*st_edats*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_edats*: - Persones per rang d'edat - Persones per rang d'edat i sexe - Persones per rang d'edat i ocupació

Els rangs d'edat considerats són: - Menors d'edat (0-17) - Adultes (18-64) - Jubilades (65-99) - Edat desconeguda - 0-9 anys - 10-19 anys - 20-29 anys - 30-39 anys - 40-49 anys - 50-59 anys - 60-69 anys - 70-79 anys - 80-89 anys - +90 anys

```

[ ]: # Converteix edats a enters i anys en edats
from datetime import datetime
any_actual = datetime.now().year
df_persones['edat'] = pd.to_numeric(df_persones['edat'], errors='coerce')
df_persones['edat'] = df_persones['edat'].apply(lambda x: any_actual - x if
↳x > 1900 else x)

# Defineix els rangs d'edat
rangs_edat = {
    'Total': 'tot',
    'Menors d\'edat (0-17)': (0, 17),
    'Adultes (18-64)': (18, 64),
    'Jubilades (65+)': (65, 99),
    'Desconeguda': 'NaN',
    '0-9 anys': (0, 9),
    '10-19 anys': (10, 19),
    '20-29 anys': (20, 29),
    '30-39 anys': (30, 39),
    '40-49 anys': (40, 49),
    '50-59 anys': (50, 59),
    '60-69 anys': (60, 69),

```

```

    '70-79 anys': (70, 79),
    '80-89 anys': (80, 89),
    '+90 anys': (90, 99)
}

# Crea una llista per desar els resultats
resultats = []

# Calcula totals de files
total_persones = len(df_persones)

# Calcula les dades per a cada rang
for rang, edats in rangs_edat.items():
    if edats == 'tot':
        valor = total_persones
    elif edats == 'NaN':
        valor = df_persones['edat'].isna().sum()
    else:
        valor = df_persones[(df_persones['edat'] >= edats[0]) &
↳(df_persones['edat'] <= edats[1])].shape[0]

    percentatge = (valor / total_persones) * 100 if total_persones > 0 else 0

    resultats.append({
        'descripció': 'Persones per rang d\'edat',
        'rang_edat': rang,
        'sexe': 'Total',
        'categoria_ocupació': 'Total',
        'desc_ocupació': '',
        'valor': valor,
        'percentatge': percentatge
    })

# Desa els resultats al marc de dades
st_edats = pd.DataFrame(resultats)

```

```

[ ]: # Defineix els sexes
sexes = ['dona', 'home', 'indeterminat']

# Calcula el total de persones amb edat
total_persones_amb_edat = df_persones['edat'].dropna().shape[0]

# Inicia una nova llista
nous_resultats = []

# Calcula les dades per rang d'edat i sexe
for rang, edats in rangs_edat.items():
    if edats not in ['tot', 'NaN']:
        for sexe in sexes:

```

```

        valor = df_persones[(df_persones['edat'] >= edats[0]) &
↳(df_persones['edat'] <= edats[1]) & (df_persones['sexe'] == sexe)].
↳shape[0]
        percentatge = (valor / total_persones_amb_edat) * 100 if
↳total_persones_amb_edat > 0 else 0
        nous_resultats.append({
            'descripció': 'Persones per rang d\'edat i sexe',
            'rang_edat': rang,
            'sexe': sexe,
            'categoria_ocupació': 'Total',
            'desc_ocupació': '',
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix els nous resultats a st_edats
noves_edats = pd.DataFrame(nous_resultats)
st_edats = pd.concat([st_edats, noves_edats], ignore_index=True)

```

```

[ ]: # Defineix el rang d'edat i els textos de desc_ocupació
rangs_edat_limitat = {
    'Menors d\'edat (0-17)': (0, 17),
    'Adultes (18-64)': (18, 64),
    'Jubilades (65+)': (65, 99)
}

mapatge_ocupació = {
    0: "No esmentada", 1: "Família reial", 2: "Classe política", 3:
↳"Funcionariat",
    4: "Policia, bombers i forces armades", 5: "Professorat", 6: "Medicina o
↳odontologia",
    7: "Altres serveis socio-sanitaris", 8: "Ciència o enginyeria", 9:
↳"Mitjans de comunicació",
    10: "Sector jurídic", 11: "Sector empresarial", 12: "Treballadors de
↳serveis",
    13: "Comerç i artesanía", 14: "Sector primari", 15: "Clergat", 16: "ONG,
↳sindicats i associacions",
    17: "Treball sexual", 18: "Celebritats i artistes", 19: "Esports", 20:
↳"Estudiantat",
    21: "Treball domèstic", 22: "Infància", 23: "Resident o habitant", 24:
↳"Jubilades",
    25: "Criminals o sospitoses", 26: "Aturades", 27: "Altres"
}

# Crea la llista per als nous resultats
nous_resultats2 = []

# Calcula dades per rang d'edat i ocupació
for rang, edats in rangs_edat_limitat.items():
    for cat_ocupació, desc_ocupació in mapatge_ocupació.items():

```

```

        if edats == 'all':
            valor = df_persones[df_persones['categoria_ocupació'] ==
↳cat_ocupació].shape[0]
        elif edats == 'NaN':
            valor = 0
        else:
            valor = df_persones[(df_persones['edat'] >= edats[0]) &
↳(df_persones['edat'] <= edats[1]) & (df_persones['categoria_ocupació'] ==
↳cat_ocupació)].shape[0]

        if valor > 0:
            percentatge = (valor / total_persones_amb_edat) * 100 if
↳total_persones_amb_edat > 0 else 0
            nous_resultats2.append({
                'descripció': 'Persones per rang d\'edat i ocupació',
                'rang_edat': rang,
                'sexe': 'Total',
                'categoria_ocupació': cat_ocupació,
                'desc_ocupació': desc_ocupació,
                'valor': valor,
                'percentatge': percentatge
            })

# Afegeix els nous resultats a st_edats
noves_edats2 = pd.DataFrame(nous_resultats2)
st_edats = pd.concat([st_edats, noves_edats2], ignore_index=True)

```

```

[ ]: # Exporta a CSV
if not os.path.exists('st_edats.csv'):
    st_edats.to_csv('st_edats.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['any_actual', 'cat_ocupació', 'desc_ocupació', 'edats',
↳'mapatge_ocupació', 'nous_resultats', 'nous_resultats2',
↳'noves_edats', 'noves_edats2', 'percentatge', 'rang',
↳'rangs_edat', 'rangs_edat_limitat', 'resultats', 'sexe',
↳'sexes', 'total_persones', 'total_persones_amb_edat',
↳'valor']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```

9.11 Estadístiques per anys (*st_sèrietemporal*)

S'exporten les següents dades (en valor absolut i percentatge) al marc de dades *st_sèrietemporal*:

- Dones per any - Homes per any - Dones amb rol familiar esmentat per any - Homes amb rol familiar esmentat per any - Dones protagonistes per any - Homes protagonistes per any - Dones expertes per any - Homes experts per any - Dones en fotografies per any - Homes en fotografies per any - Dones redactores per any - Homes redactors per any

```
[ ]: # Es converteix el camp 'data' a any
df_notícies['any'] = pd.to_datetime(df_notícies['data']).dt.year

# Es crea el nou marc de dades
st_sèrietemporal = pd.DataFrame()
```

```
[ ]: # Es calculen els totals de dones i homes per any
persones_per_any = df_notícies.groupby('any')[['persones_dona',
↪ 'persones_home']].sum()

# S'afegeixen manualment les dades de 2014
if 2014 not in persones_per_any.index:
    persones_per_any.loc[2014] = {'persones_dona': 253, 'persones_home': 857}
else:
    persones_per_any.loc[2014, 'persones_dona'] += 253
    persones_per_any.loc[2014, 'persones_home'] += 857
persones_per_any.sort_index(inplace=True)

# Calcula les dades per any i sexe
dades_any_sexe = []
for any, fila in persones_per_any.iterrows():
    total_persones = fila['persones_dona'] + fila['persones_home']

    for sexe, persones in zip(['dones', 'homes'], ['persones_dona',
↪ 'persones_home']):
        valor = fila[persones]
        percentatge = (valor / total_persones) * 100 if total_persones > 0
↪ else 0
        descripcio = f'{sexe.capitalize()} per any'
        dades_any_sexe.append({
            'descripció': descripcio,
            'any': any,
            'sexe': sexe,
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix les dades a st_sèrietemporal
any_sexe = pd.DataFrame(dades_any_sexe)
st_sèrietemporal = pd.concat([st_sèrietemporal, any_sexe], ignore_index=True)
```

```
[ ]: # Fusiona df_persones i df_notícies per id
join_persones_notícies = pd.merge(df_persones, df_notícies[['id', 'any']],
↪ on='id')
```

```
[ ]: # Calcula el total de dones i d'homes per any, independentment del
↪ rol_familiar
total_dones_homes_per_any =
↪ join_persones_notícies[join_persones_notícies['sexe'].isin(['dona',
↪ 'home'])].groupby('any')['sexe'].size()
```

```

# Calcula la xifra de dones i d'homes amb rol familiar esmentat
dades_rolfamiliar =
↳join_persones_notícies[join_persones_notícies['rol_familiar'] == 1].
↳groupby(['any', 'sexe']).size().unstack(fill_value=0)

# Crea una llista per a les noves dades
dades_rolfamiliar_any = []

# S'afegeixen manualment les dades de 2014
dades_rolfamiliar_any.extend([
    {'descripció': 'Dones amb rol familiar esmentat per any', 'any': 2014,
↳'sexe': 'dona', 'valor': None, 'percentatge': 7.00},
    {'descripció': 'Homes amb rol familiar esmentat per any', 'any': 2014,
↳'sexe': 'home', 'valor': None, 'percentatge': 5.00}
])

# Mapatge per al camp 'descripció'
sexe_plural = {'dona': 'Dones', 'home': 'Homes'}

# Calcula les dades per any
for any, fila in dades_rolfamiliar.iterrows():
    total_persones_any = total_dones_homes_per_any.get(any, 0)

    for sexe in ['dona', 'home']:
        valor = fila.get(sexe, 0)
        percentatge = (valor / total_persones_any) * 100 if
↳total_persones_any > 0 else 0

        descripcio = f'{sexe_plural[sexe]} amb rol familiar esmentat per any'

        dades_rolfamiliar_any.append({
            'descripció': descripcio,
            'any': any,
            'sexe': sexe,
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix les dades a st_sèrietemporal
dades_rolfamiliar_any_df = pd.DataFrame(dades_rolfamiliar_any)
st_sèrietemporal = pd.concat([st_sèrietemporal, dades_rolfamiliar_any_df],
↳ignore_index=True)

```

```

[ ]: # Calcula la xifra de dones i d'homes com a protagonistes
dades_protagonista =
↳join_persones_notícies[join_persones_notícies['protagonista'] == 1].
↳groupby(['any', 'sexe']).size().unstack(fill_value=0)

# Crea una llista per a les noves dades

```

```

dades_protagonista_any = []

# S'afegeixen manualment les dades de 2014
dades_protagonista_any.extend([
    {'descripció': 'Dones protagonistes per any', 'any': 2014, 'sexe': 'dona',
    ↪ 'valor': None, 'percentatge': 23.00},
    {'descripció': 'Homes protagonistes per any', 'any': 2014, 'sexe': 'home',
    ↪ 'valor': None, 'percentatge': 77.00}
])

# Calcula les dades per any
for any, fila in dades_protagonista.iterrows():
    total_protagonistes = fila.sum()

    for sexe in ['dona', 'home']:
        valor = fila.get(sexe, 0)
        percentatge = (valor / total_protagonistes) * 100 if
    ↪ total_protagonistes > 0 else 0
        descripcio = f'{sexe_plural[sexe]} protagonistes per any'
        dades_protagonista_any.append({
            'descripció': descripcio,
            'any': any,
            'sexe': sexe,
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix les dades a st_sèrietemporal
dades_protagonista_any_df = pd.DataFrame(dades_protagonista_any)
st_sèrietemporal = pd.concat([st_sèrietemporal, dades_protagonista_any_df],
    ↪ ignore_index=True)

```

```

[ ]: # Calcula la xifra de dones i d'homes com a experts
dades_experta = join_persones_notícies[join_persones_notícies['experta'] ==
    ↪ 1].groupby(['any', 'sexe']).size().unstack(fill_value=0)

# Crea una llista per a les noves dades
dades_experta_any = []

# S'afegeixen manualment les dades de 2014
dades_experta_any.extend([
    {'descripció': 'Dones expertes per any', 'any': 2014, 'sexe': 'dona',
    ↪ 'valor': None, 'percentatge': 18.00},
    {'descripció': 'Homes experts per any', 'any': 2014, 'sexe': 'home',
    ↪ 'valor': None, 'percentatge': 82.00}
])

# Calcula les dades per any
for any, fila in dades_experta.iterrows():
    total_experts = fila.sum()

```



```

    for sexe, descripcio_sexe in zip(['dona', 'home'], ['Dones expertes',
↳ 'Homes experts']):
        valor = fila.get(sexe, 0)
        percentatge = (valor / total_experts) * 100 if total_experts > 0
↳ else 0
        descripcio = f'{descripcio_sexe} per any'

        dades_experta_any.append({
            'descripció': descripcio,
            'any': any,
            'sexe': sexe,
            'valor': valor,
            'percentatge': percentatge
        })

# Afegeix les dades a st_sèrietemporal
dades_experta_any_df = pd.DataFrame(dades_experta_any)
st_sèrietemporal = pd.concat([st_sèrietemporal, dades_experta_any_df],
↳ ignore_index=True)

```

```

[ ]: # Es fusionen df_imatges i df_notícies per id
join_imatges_notícies = pd.merge(df_imatges, df_notícies[['id', 'any']],
↳ on='id')

# S'eliminen les files amb fotografies sense persones
join_imatges_notícies =
↳ join_imatges_notícies[join_imatges_notícies['imatge_persones'] > 0]

# Calcula la suma d'imatges de dones i d'homes per any
imatges_per_any = join_imatges_notícies.groupby('any')[['imatge_dones',
↳ 'imatge_homes']].sum()

# Crea una llista per a les noves dades
dades_imatges_any = []

# S'afegeixen manualment les dades de 2014
dades_imatges_any.extend([
    {'descripció': 'Dones en fotografies per any', 'any': 2014, 'sexe':
↳ 'dones', 'valor': None, 'percentatge': 24.5},
    {'descripció': 'Homes en fotografies per any', 'any': 2014, 'sexe':
↳ 'homes', 'valor': None, 'percentatge': 74.5}
])

# Calcula les dades per any
for any, fila in imatges_per_any.iterrows():
    total_imatges = fila.sum()

    for sexe, imatge in zip(['dones', 'homes'], ['imatge_dones',
↳ 'imatge_homes']):

```

```

    valor = fila[imatge]
    percentatge = (valor / total_imatges) * 100 if total_imatges > 0,
↳else 0
    descripcio = f'{sexe.capitalize()} en fotografies per any'

    dades_imatges_any.append({
        'descripció': descripcio,
        'any': any,
        'sexe': sexe,
        'valor': valor,
        'percentatge': percentatge
    })

# Afegeix les dades a st_sèrietemporal
dades_imatges_any_df = pd.DataFrame(dades_imatges_any)
st_sèrietemporal = pd.concat([st_sèrietemporal, dades_imatges_any_df],
↳ignore_index=True)

```

```

[ ]: # Suprimeix les firmes indeterminades i extreu els anys
dades_firmes = df_notícies[df_notícies['firma1_indeterminat'] != 1].copy()
dades_firmes['any'] = pd.to_datetime(dades_firmes['data']).dt.year

# Converteix strings buits a 0 dins firma2_dona i firma2_home
dades_firmes['firma2_dona'] = pd.to_numeric(dades_firmes['firma2_dona'],
↳errors='coerce').fillna(0)
dades_firmes['firma2_home'] = pd.to_numeric(dades_firmes['firma2_home'],
↳errors='coerce').fillna(0)

# Calcula la suma de firmes de dones i d'homes per any
firmes_per_any = dades_firmes.groupby('any').agg({
    'firma1_dona': 'sum',
    'firma2_dona': 'sum',
    'firma1_home': 'sum',
    'firma2_home': 'sum'
})

# Crea una llista per a les noves dades
dades_firmes_any = []

# S'afegeixen manualment les dades de 2014
dades_firmes_any.extend([
    {'descripció': 'Dones redactores per any', 'any': 2014, 'sexe': 'dones',
↳'valor': None, 'percentatge': 33.5},
    {'descripció': 'Homes redactors per any', 'any': 2014, 'sexe': 'homes',
↳'valor': None, 'percentatge': 66.5}
])

# Calcula les dades per any
for any, fila in firmes_per_any.iterrows():

```

```

total_firmes = fila['firma1_dona'] + fila['firma2_dona'] +
↳fila['firma1_home'] + fila['firma2_home']

for sexe, firmes in zip(['dones', 'homes'], [('firma1_dona',
↳'firma2_dona'), ('firma1_home', 'firma2_home')]):
    valor = fila[firmes[0]] + fila[firmes[1]]
    percentatge = (valor / total_firmes) * 100 if total_firmes > 0 else 0

descripcio = f'{sexe.capitalize()} redactors per any'

dades_firmes_any.append({
    'descripció': descripcio,
    'any': any,
    'sexe': sexe,
    'valor': valor,
    'percentatge': percentatge
})

# Afegeix les dades a st_sèrietemporal
dades_firmes_any_df = pd.DataFrame(dades_firmes_any)
st_sèrietemporal = pd.concat([st_sèrietemporal, dades_firmes_any_df],
↳ignore_index=True)

```

```

[ ]: # Exporta a CSV
if not os.path.exists('st_sèrietemporal.csv'):
    st_sèrietemporal.to_csv('st_sèrietemporal.csv', index=False)

```

```

[ ]: # Elimina les variables prescindibles
for variable in ['any', 'any_sexe', 'dades_any_sexe', 'dades_experta',
↳'dades_experta_any', 'dades_experta_any_df',
↳'dades_firmes', 'dades_firmes_any', 'dades_firmes_any_df',
↳'dades_imatges_any', 'dades_imatges_any_df',
↳'dades_protagonista', 'dades_protagonista_any',
↳'dades_protagonista_any_df', 'dades_rolfamiliar',
↳'dades_rolfamiliar_any', 'dades_rolfamiliar_any_df',
↳'descripcio', 'descripcio_sexe', 'fila',
↳'firmes', 'firmes_per_any', 'imatge', 'imatges_per_any',
↳'join_imatges_notícies',
↳'join_persones_notícies', 'percentatge', 'persones',
↳'persones_per_any', 'sexe', 'sexe_plural',
↳'total_dones_homes_per_any', 'total_experts',
↳'total_firmes', 'total_imatges', 'total_persones',
↳'total_persones_any', 'total_protagonistes', 'valor']:
    try:
        del globals()[variable]
    except KeyError:
        pass
del variable

```