



Anàlisi de dades sobre la correlació entre delictes de violència de gènere i la climatologia

Autor: Jin Lung Chan
Grau d'Enginyeria Informàtica

Consultor: *Humberto Andrés Sanz*

14 de juny de 2017



Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial-SenseObraDerivada
3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	Anàlisi de dades sobre la correlació entre delictes de violència de gènere i la climatologia.
Nom de l'autor:	Jin Lung Chan
Nom del consultor:	Humberto Andrés Sanz
Data de lliurament (mm/aaaa):	06/2017
Àrea del Treball Final:	Business Intelligence
Titulació:	Grau en Enginyeria informàtica
Resum del Treball (màxim 250 paraules):	
<p>Al segle XXI, vivim en una societat regida per un estat de dret i justícia on es dona per suposat que els membres que la componen coneixen i respecten els drets fonamentals. No obstant això, aquesta afirmació no és del tot certa, ja que si ens aturem a pensar en els maltractaments que pateixen les dones en el marc de la violència de gènere, ens adonarem que els drets i llibertats, concretament el dret a la vida i la integritat física i moral, no és per tots igual.</p> <p>Els organismes públics inverteixen recursos en la prevenció, detecció i ajuda a les víctimes de la violència de gènere ja que és un problema latent a la nostra societat, i una de les eines que disposen són les forces i cossos de seguretat, que treballen tant en l'àmbit preventiu com reactiu.</p> <p>Ben segur que les policies estatals i autonòmiques ja disposen d'analistes que estudiïn les tendències de diferents fenòmens delictius, però des de la desconeixença de l'autor d'aquest treball de quins tipus de recursos disposen, es vol aportar una eina de <i>Business Intelligence</i> per estudiar possibles factors o indicadors que ajudin a prevenir els delictes de violència de gènere des d'una vessant estratègica.</p> <p>Així doncs, a partir del recull de fets delictius cedits pels Mossos d'Esquadra, es vol establir una correlació entre la violència exercida en l'àmbit de la violència de gènere i la climatologia, que permeti a la policia conèixer les tendències i així poder destinar els recursos de manera més eficient.</p>	

Abstract (in English, 250 words or less):

In XXI century, we live in a society lead by the rules of the law and justice where all members should know and respect the fundamental rights. However, this affirmation it is not true at all because if we take a minute and think about the mistreatment suffered by women within the scope of gender violence, then we will realize that the right to life, more specifically, the right to have physical and moral integrity, is not equal for everyone.

In order to prevent, detect and help the victims of gender violence, public organizations invest lots of resources to eradicate this problem which is present in our society. One of those resources is the police forces, entity that help in proactive and reactive way.

Sure it is, that police forces already has analysts that study the trends from different criminal phenomena's, but with the lack of awareness from the author of this project of what resources they have, the author would like to contribute with a business intelligence tool that let to organization know possible factors or indicators to prevent gender violence from a strategic view.

Therefore, through the data hand out by Mossos d'Esquadra, this project will try to establish a correlation between the violence from the gender violence and the climatology, so as to police may know the trends and then invest the resources more effectively.

Paraules clau (entre 4 i 8):

Business Intelligence, MongoDB, Tableau, Violència de gènere.

Índex

1. INTRODUCCIÓ	3
1.1 Context i justificació del treball	4
1.2 Objectius.....	5
1.3 Enfocament i mètode de treball	6
1.4 Planificació de treball.....	7
1.5 Breu sumari de productes obtinguts	10
1.6 Breu descripció dels altres capítols de la memòria	10
2 Cerca de documentació.....	12
2.1 Recerca de dades estadístiques.....	12
2.2.1 MongoDB.....	13
2.2.2 Tableau.....	13
2.2.3 Simba	14
2.2.4 Robomongo	14
3 Entorn i desenvolupament.....	15
3.1 Instal·lació i configuració de programari	15
3.1.1 MongoDB.....	16
3.1.2 Robomongo	17
3.1.3 Simba MongoDB ODBC Driver	18
3.1.4 Tableau.....	20
3.2 ETL sobre les dades recollides.....	21
3.3 Càrrega de dades	21
3.4 Proves	23
4 Explotació, anàlisi i representació de dades.....	27
4.1 Representació gràfica.....	28
4.1.1 Any 2010.....	29
4.1.2 Any 2011.....	30
4.1.3 Any 2012.....	32
4.1.4 Any 2013.....	33
4.1.6 Any 2015.....	35
4.1.7 Any 2016.....	36
4.1.8 Anys 2010-2016.....	37
4.2 Gràfics a partir de camps calculats.....	38
4.3 El mapa com a representació	40
4.4 Dashboard.....	42
5 Conclusions.....	45
6 Bibliografia.....	47

Taula d'il·lustracions

Il·lustració 1 - Diagrama de Gant	9
Il·lustració 2 - Esquema de l'estructura del programari que s'ha implementat	15
Il·lustració 3 - Configuració de Robomongo contra la BBDD de MongoDB	17
Il·lustració 4 - Configuració de l'ODBC de SIMBA	18
Il·lustració 5 - Configuració de la connexió a Tableau	20
Il·lustració 6 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2010	29
Il·lustració 7 - Gràfic de barres sobre els homicidis i la temperatura l'any 2010	30
Il·lustració 8 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2011	31
Il·lustració 9 - Gràfic de barres sobre els homicidis i la temperatura l'any 2011	31
Il·lustració 10 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2012	32
Il·lustració 11 - Gràfic de barres sobre els homicidis i la temperatura l'any 2012	32
Il·lustració 12 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2013	33
Il·lustració 13 - Gràfic de barres sobre els homicidis i la temperatura l'any 2013	33
Il·lustració 14 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2014	34
Il·lustració 15 - Gràfic de barres sobre els homicidis i la temperatura l'any 2014	34
Il·lustració 16 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2015	35
Il·lustració 17 - Gràfic de barres sobre els homicidis i la temperatura l'any 2015	35
Il·lustració 18 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2015	36
Il·lustració 19 - Gràfic de barres sobre els homicidis i la temperatura l'any 2016	36
Il·lustració 20 - Gràfic de barres sobre els delictes de lesions i les temperatures entre els anys 2010 i 2016.....	37
Il·lustració 21 - Creació de camps calculats.....	38
Il·lustració 22 - Gràfic de barres sobre fets denunciats en l'àmbit de la violència de gènere a les diferents comissaries de la regió metropolitana de Barcelona.	39
Il·lustració 23 - Mapa generat a partir de les coordenades i dels fets denunciats a Barcelona ...	41
Il·lustració 24 - <i>Dashboard</i> de Tableau combinat amb un gràfic de barres i un mapa	42
Il·lustració 25 - Dashboard que recull els delictes de lesions i homicidis entre els anys 2010 i 2016 i la seva correlació amb les temperatures.	44

1. INTRODUCCIÓ

1.1 Context i justificació del treball

Actualment existeix una alarma social envers els delictes de violència de gènere i aquest fenomen s'agreuja quan aquesta tipologia delictiva té resultat de mort. Això ha conduït als legisladors a elaborar reformes en el codi penal per tal de garantir i protegir la integritat física i moral de la víctima. No obstant això, i els recursos invertits per la conscienciació de la ciutadania, cada any són moltes vides les que es perden a les mans de les seves parelles.

Per afrontar aquest problema, s'aborda des d'una estratègia reactiva pels organismes relacionats amb la seguretat, és a dir, que la policia actua a requeriment de les víctimes i és llavors, quan ja es disposa d'una denúncia prèvia, que la justícia aplica mesures restrictives i coercitives de llibertat envers l'autor i de protecció envers la persona perjudicada.

Així, aquest projecte pretén aportar una eina a les entitats relacionades amb la seguretat pública, que a partir de un o diversos indicadors es pugui des d'un punt de vista estratègic conèixer quines són les tendències sobre els delictes relacionats amb la violència de gènere i des d'un punt de vista tàctic i operatiu, poder reduir el número de víctimes focalitzant els recursos d'una manera més efectiva i eficient.

Aquesta solució tecnològica, versarà sobre l'estudi i anàlisi de la relació entre la climatologia i el número de víctimes mortals per violència de gènere a diferents èpoques de l'any. És a dir, que si existeix una relació estadística, és podria conèixer quina ha estat la tendència fins a dia d'avui.

El producte final és convertiria en una eina per les forces i cossos de seguretat que permetria realitzar un anàlisi predictiu sobre quines franges de l'any són més propenses a aquesta tipologia delictiva i així es podria abocar els recursos d'una manera més òptima i eficient.

1.2 Objectius

L'abast del projecte consistirà, per una banda recollir les dades de diverses fonts oficials sobre lesions, homicidis i assassinats en l'àmbit de la violència de gènere i també de les temperatures diàries dels últims anys. Per altra banda, el volum de dades que serà tractat com a *Big Data*, serà depurat per posteriorment carregar-ho contra una base de dades i mitjançant un quadre de comandament, s'analitzarà i estudiarà de manera gràfica els fets delictius i la seva correlació amb la temperatura.

L'objectiu final del treball és generar un sistema d'anàlisi predictiu que permeti conèixer quines èpoques de l'any són més tendencioses a la violència de gènere i integrar aquesta eina a organitzacions policials d'àmbit autonòmic o estatal, com ara els Mossos d'Esquadra.

Així, el treball es desgrana en els següents objectius que estan ordenats per ordre descendent de prioritat :

- Recopilar dades de diferents fonts oficials relacionades amb les temperatures durant un període 10 anys
- Recopilar dades de diferents fonts oficials relacionades amb fets delictius englobats en l'àmbit de la violència de gènere durant un període 10 anys.
- ETL sobre les dades recopilades, és a dir, tractar i depurar les dades per poder-les importar i treballar.
- Instal·lar i configurar el programari orientat al *Big Data* per emmagatzemar les dades recopilades i poder-les analitzar.
- Instal·lar i configurar el programari orientat al *Business Intelligence* per representar les dades gràficament.
- Estudiar i analitzar els resultats i extreure conclusions.

Tots aquests objectius formen part essencial del treball i no seria possible dur a terme el projecte sense l'acompliment exitós de cada objectiu. Malgrat això, hi ha tasques que encara que aparentment semblin més trivials, tenen una prioritat superior perquè hi ha fites que depenen d'altres. És a dir, com a exemple, no podem realitzar l'ETL de les dades sense haver-les recopilat abans.

1.3 Enfocament i mètode de treball

Actualment, per eradicar la violència de gènere, s'inverteixen recursos des de dues estratègies oposades:

Des del punt de vista estratègic reactiu, la víctima requereix auxili a la policia i aquest porta davant del jutge a l'autor per a què decreti mesures preventives. És a dir, que si la víctima no posa en coneixement el maltractament no és pot iniciar tot el procés, que engloba les accions penals envers l'autor. L'òrgan encarregat de gestionar la seguretat és el Ministeri de l'Interior.

Per altra banda, des del punt de vista proactiu o preventiu, el Ministeri de Sanitat, Serveis Socials i Igualtat inverteix molts recursos publicitaris per conscienciar a la societat sobre la violència de gènere i la necessitat de la denúncia a través de la policia o telèfons d'ajuda com el 016. Aquest tipus d'estratègia intenta anticipar-se a escenaris que l'organització ha previst.

La combinació d'aquestes dues estratègies frena l'augment de les xifres de les víctimes mortals. No obstant això, i que ambdós ministeris tenen objectius comuns, a l'organigrama general són dos òrgans separats que tenen una metodologia de treball diferent i cap de les dues utilitza les tecnologies com a eina d'anàlisi per predir futurs escenaris.

L'estratègia escollida per aquest treball pretén adaptar polítiques proactives i prospectives al món policial a partir de l'anàlisi de possibles indicadors tècnics, com la que s'ha proposat: Estudiar la relació entre el clima i el número de víctimes mortals en la violència de gènere.

Per últim, l'estratègia proposada serà des d'una perspectiva analítica del Big Data combinada amb una solució de *Business Intelligence*, que serà la més apropiada perquè s'emprà la tecnologia del segle XXI per abordar un problema que ha estat latent des de temps remots.

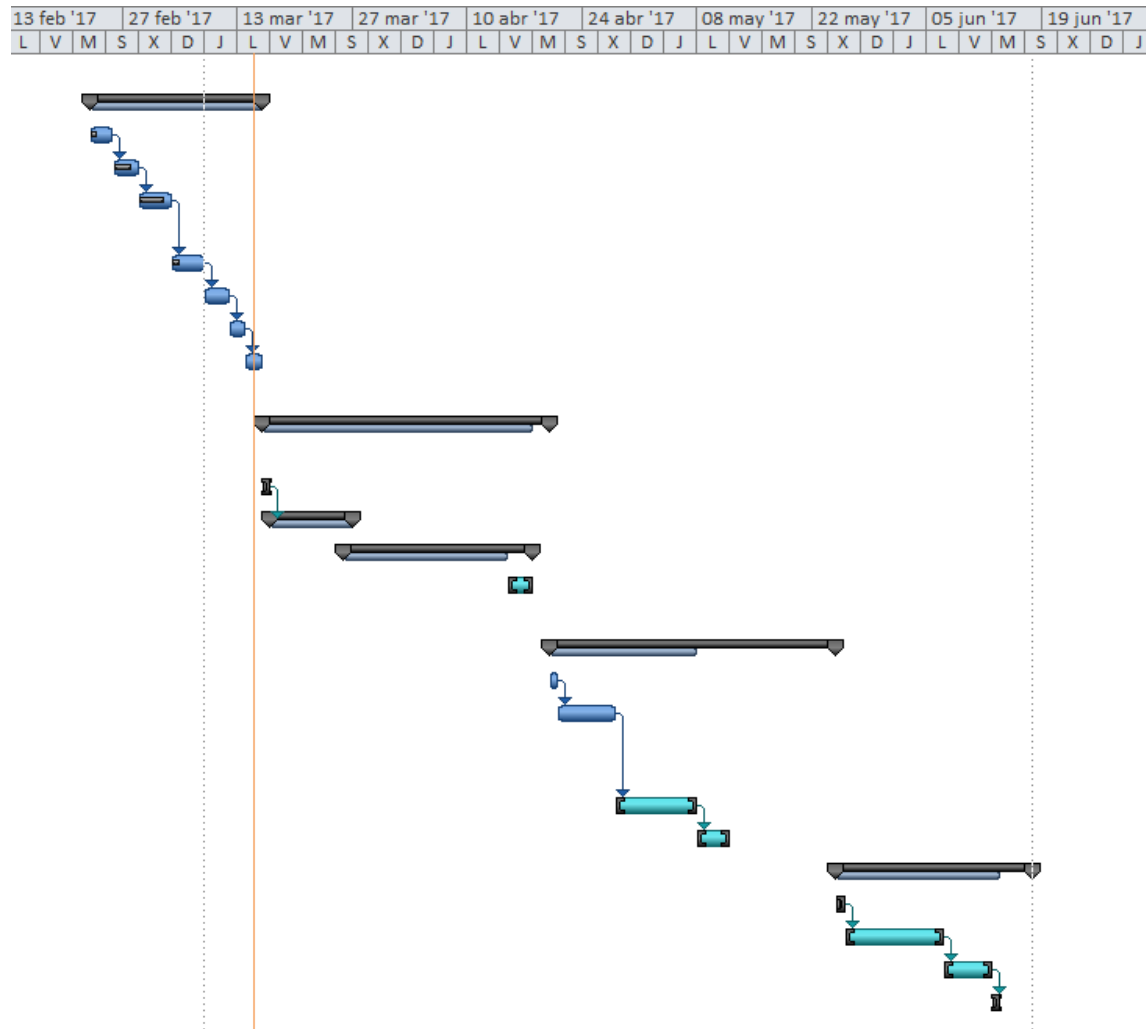
1.4 Planificació de treball

El calendari de la temporalització del treball contempla una planificació temporal de 7 dies laborals per setmana i sense contemplar les dates festives doncs s'entén que el responsable del projecte té una dedicació exclusiva fins el final del treball.

Nom de la tasca	Durada	Inici	Fi	Predecessores	Responsable i Nombres de los recursos
FASE1: PAC1 "Pla de treball"	21 días	jue 23/02/17	mié 15/03/17		Jin Lung Chan
Definir proposta del TFG	3 días	jue 23/02/17	sáb 25/02/17		
Recull d'informació	3 días	dom 26/02/17	mar 28/02/17	3	
Planificació de tasques i abast	4 días	mié 01/03/17	sáb 04/03/17	4	
Redacció del context	4 días	dom 05/03/17	mié 08/03/17	5	
Redacció dels objectius	3 días	jue 09/03/17	sáb 11/03/17	6	
Redacció de l'estratègia	2 días	dom 12/03/17	lun 13/03/17	7	
Maquetació i entrega de la PAC1	2 días	mar 14/03/17	mié 15/03/17	8	
FASE2: PAC2 "Desenvolupament"	35 días	jue 16/03/17	mié 19/04/17		Jin Lung Chan
Aplicar correccions a la PAC2	1 día	jue 16/03/17	jue 16/03/17		
Cerca de documentació	10 días	vie 17/03/17	dom 26/03/17	11	
Recopilació de dades estadístiques per l'estudi	5 días	vie 17/03/17	mar 21/03/17		
Recerca i selecció programari DataWareHouse	5 días	vie 17/03/17	mar 21/03/17		
Recerca i selecció programari Business Intelligence	5 días	mié 22/03/17	dom 26/03/17	14	
Entorn i desenvolupament	23 días	dom 26/03/17	lun 17/04/17	15	
Instal·lació i configuració de programari	7 días	dom 26/03/17	sáb 01/04/17		
ETL sobre les dades recollides al punt 11	5 días	dom 02/04/17	jue 06/04/17	17	
Càrrega de dades	4 días	vie 07/04/17	lun 10/04/17	18	

Proves	4 días	mar 11/04/17	vie 14/04/17	19	
Maquetació i entrega de la PAC2	3 días	sáb 15/04/17	lun 17/04/17	20	
FASE3: PAC3 "Anàlisi"	35 días	jue 20/04/17	mié 24/05/17		Jin Lung Chan
Aplicar correccions a la PAC2	1 día	jue 20/04/17	jue 20/04/17		
Representar resultats a través de quadre de comandament	7 días	vie 21/04/17	jue 27/04/17	23	
Anàlisi dels resultats	10 días	vie 28/04/17	dom 07/05/17	24	
Redactar conclusions	4 días	lun 08/05/17	jue 11/05/17	25	
FASE4: Memòria final TFG	24 días	jue 25/05/17	sáb 17/06/17		Jin Lung Chan
Aplicar correccions a la PAC3	1 día	jue 25/05/17	jue 25/05/17		
Redactar memòria final	12 días	vie 26/05/17	mar 06/06/17	28	
Disseny de la presentació	6 días	mié 07/06/17	lun 12/06/17	29	
Maquetació i entrega del TFG	1 día	mar 13/06/17	mar 13/06/17	30	

El següent diagrama de *Gant* és una planificació i representació de les tasques necessàries per realitzar el projecte associat a la seva relació temporal i els recursos necessaris per realitzar el projecte.



Il·lustració 1 - Diagrama de Gant

1.5 Breu sumari de productes obtinguts

El producte final que s'obtindrà serà una recopilació de dades amb valor estadístic que seran emmagatzemades a un *datawarehouse*, on es tractaran com a Big Data i posteriorment s'explotaran gràficament a partir d'un quadre de comandament amb l'objectiu de realitzar un anàlisi qualitatiu de la informació que proporciona i amb el suport de les eines de *Business Intelligence*.

1.6 Breu descripció dels altres capítols de la memòria

Els capítols versaran sobre les diferents tasques esmentades en el diagrama de *Gant*. Així, la memòria final del projecte contindrà els següents apartats:

Capítol 1- Introducció i posada en context

És el punt de partida del treball que conté una breu introducció sobre quin és la necessitat a cobrir, com es tracta actualment el problema i per què és rellevant el tema. A partir d'aquesta base es desenvolupa la proposta per aconseguir una eina tecnològica que ajudi a una organització.

S'enumeren en un llistat els objectius que es pretenen assolir amb el projecte. A més, s'indica l'estratègia actual i quina és la estratègia proposada. Per últim, s'adjunta una planificació de les tasques amb dates determinades per executar i aconseguir les fites.

Capítol 2- Recollida d'informació

En aquesta fase, es recullen les dades de diferents fonts oficials per poder realitzar la importació en el capítol 4 i l'anàlisi en el capítol 5. Les fonts amb les que es treballaran seran de l'Institut Nacional d'Estadística, Institut Català d'Estadística i d'altres fonts que es determinin necessàries per realitzar el treball.

Capítol 3- Disseny del *datawarehouse*

El disseny del *datawarehouse* consistirà a cerca informació sobre un tipus de base de dades o arquitectura per emmagatzemar dades tipus *Big Data*. Una vegada ja s'ha escollit, s'instal·larà i configurarà per deixar enllestit el *datawarehouse* per importar les dades del capítol 2.

Per altra banda, també és cercarà informació sobre quadres de comandament per la representació de la informació i una vegada seleccionat és deixarà instal·lat i configurat en el *datawarehouse*.

Capítol 4 – ETL

ETL o conegut popularment com “*Export, Transform & Load*”, és a dir, extracció, transformació i càrrega de les dades recollides en el capítol 2. A grans trets, les dades s’hauran de depurar i tractar per a què puguin ser carregades a la base de dades.

Capítol 5- Explotació de dades i anàlisi

Aquest capítol conté dues fases. A la primera fase es realitzaran proves per verificar el correcte funcionament del disseny elaborat en el capítol 3 i verificar que les dades tenen sentit.

La segona fase, consistirà a estudiar les dades a partir d’un quadre de comandament, per localitzar patrons i/o tendències que puguin proporcionar informació de valor a l’organització.

Capítol 6 - Conclusions

L’apartat de conclusions contindrà una síntesi del treball realitzat al llarg del projecte, quines han estat els principals problemes per dur-ho a terme i els resultats obtinguts a partir de l’estudi i anàlisi de les dades.

2 Cerca de documentació

2.1 Recerca de dades estadístiques

La informació necessària per realitzar aquest projecte esta composta de dues parts. La primera part, són les dades relacionades amb la climatologia que en un primer s'havien d'extreure de la pàgina web de l'Institut Nacional d'Estadística i alternativament de l'Institut Estadístic de Catalunya, però en els seus registres només constaven temperatures mitjanes en períodes d'un mes.

Per tal d'obtenir més precisió durant la fase d'anàlisi, s'han aconseguit recollir les dades de l'Agència Estatal de Meteorologia, entitat de la qual disposa d'un registre de temperatures diàries.

La segona part, són dades policials sobre els delictes de violència de gènere que s'ha sol·licitat a la Policia de la Generalitat Mossos d'Esquadra, especificant a la petició que es sol·licita fets coneguts emmarcats en aquest àmbit que incloguin faltes i delictes de lesions, homicidis i assassinats. La resposta facilitada conté fets coneguts a Catalunya des de l'any 2010 fins l'any 2016.

Així doncs, la següent taula descriu el contingut i l'estructura de les dades que s'han recollit a les dues entitats.

Nom del fitxer	Fets_VIGE_BCN.csv	Barcelonaclima.csv
Descripció	Full de càlcul en format Excel que conté els fets delictius coneguts a Catalunya emmarcats en l'àmbit de la violència de gènere des de l'any 2010 fins l'any 2016	Full de càlcul en format Excel que conté les temperatures diàries registrades a Barcelona capital des de l'any 2010 fins l'any 2016.
Camps	<ul style="list-style-type: none">• Data inici fet• Hora inici fet• Tipus fet• Grau del fet• Municipi fet• Districte fet• Fets Localitzats	<ul style="list-style-type: none">• Fecha• TMax• HoraTmax• TMin• HoraTmin• TMed

2.2 Recerca i selecció de programari

2.2.1 MongoDB

Com a *datawarehouse* per aquest tipus de projecte, s'ha optat per **MongoDB**, un tipus de base de dades no relacional, popularment conegut com a *NoSQL* ja que no implementa totes les propietats ACID de les bases de dades tradicionals, sinó que segueix el teorema de CAP. *MongoDB* és un programari de codi obert que no necessita l'existència d'un esquema relacional. i es caracteritza a grans trets per:

- ✓ Escalabilitat
- ✓ Indexació
- ✓ Replicació
- ✓ Balanç de càrrega
- ✓ Emmagatzemant d'arxius

MongoDB emmagatzema les dades en format BSON que són estructures semblants a les taules anomenades documents. Així, l'estructura de les dades en una base de dades de *MongoDB* conté els següents elements:

- Una **base de dades** que conté un conjunt de col·leccions
- Una **col·lecció** que conté un conjunt de documents
- Un **document** que conté un conjunt de camps
- Un **camp** que conté un identificador generat per una clau i un valor
- Una **clau** que acostuma a ser un nom tipus String i un **valor** que identifica el tipus de valor que conté, que podria ser tipus string, integer, float, entre d'altres.

2.2.2 Tableau

Com a eina de *Business Intelligence* per representar, estudiar i analitzar les dades s'ha escollit **TABLEAU**, un programari analític que focalitza el seu desenvolupament en la intel·ligència empresarial amb l'objectiu de millorar els processos de negoci. A més, qualsevol usuari que estigui familiaritzat amb eines ofimàtiques com Excel pot generar amb aquesta aplicació un impacte de qualitat per l'organització.

Així mateix, Tableau és capaç de sincronitzar-se amb diferents tipus de base de dades ja siguin emmagatzemades en un servidor remot o en un núvol, o importar dades des de diferents tipus d'arxius i representar-les gràficament en un *dashboard* personalitzat.

No obstant això, Tableau no és programari de codi obert i la versió completa té un cost d'entre 35\$ i 70\$ al mes en funció de la versió.

2.2.3 Simba

Per establir un canal comunicació entre MongoDB i Tableau, és necessari un programari per connectar les dues parts. Aquest programari s'anomena ODBC (Open Database Connectivity), que és una capa intermitja entre el dataWarehouse i el programari BI que tradueix les consultes de dades en les dues direccions.

A fi d'establir aquest canal de comunicació intermig, s'ha optat per l'ODBC de *Simba Technologies*, que és una empresa especialitzada en desenvolupar productes tipus ODBC, JDBC, OLE DB per a diferents sistemes operatius.

La raó per de selecció d'aquesta empresa és perquè tenen un ODBC desenvolupat específicament per *MongoDB*. Malgrat aquesta avantatge, aquest programari no és de codi obert i té un preu de 199\$.

2.2.4 Robomongo

Robomongo és un programari gratuït que està dissenyat com a eina per administrar gràficament el contingut d'una base de dades amb les avantatges que confereix la capacitat de connexió múltiple a la mateixa o diferents bases de dades i a més el ressaltat de les sintaxis amb colors. Així, a grans trets, és un client per connectar a un servidor i manipular les dades.

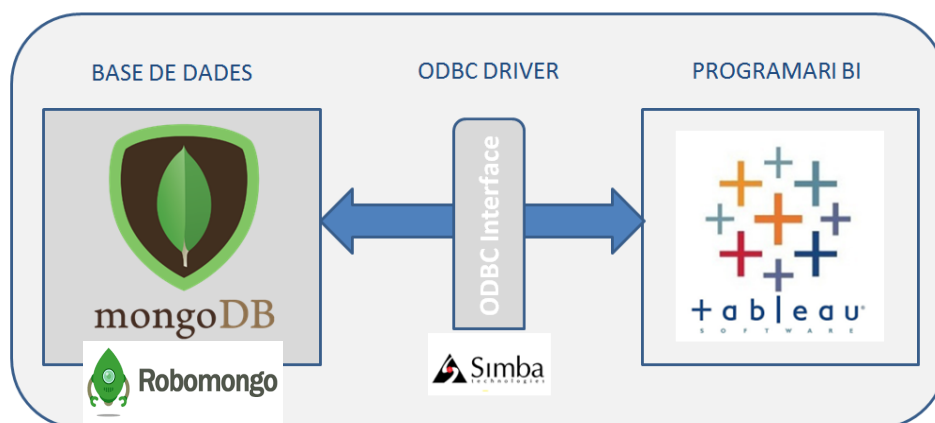
3 Entorn i desenvolupament

La següent taula descriu les característiques del maquinari i programari que s'han utilitzat en aquest projecte.

	Servidor	Client
Hardware	Intel Core i5 2450M @ 2.5Ghz 8 Gb RAM 120Gb	Intel Core i5 2450M @ 2.5Ghz 8 Gb RAM 320Gb
Sistema operatiu	Kali (Debian Distribution 4.6.4-1 x86_64)	Windows 7 Ultimate SP1 64 bits
Adreça IP	192.168.1.135	192.168.1.130
Programari instal·lat	MongoDB 2.6.12	Robomongo 1.0 RC1
		Simba MongoDB ODBC Driver
		Tableau 10.2

3.1 Instal·lació i configuració de programari

El següent esquema és una representació gràfica de l'estructura final del programari del projecte, que és compost per **MongoDB** com a dataWareHouse, **Tableau** com a eina de Business Intelligence, amb **Simba** com a connector ODBC entre els dos programaris. També és representa en el bloc de base de dades **Robomongo**, que és un programari client per testejar i manipular la base de dades des de la màquina client.



Il·lustració 2 - Esquema de l'estructura del programari que s'ha implementat

A continuació es desgranaran les instal·lacions i configuracions d'aquests programaris.

3.1.1 MongoDB

Comencem la instal·lació de MongoDB en el servidor, que conté el sistema operatiu Linux. La instal·lació des d'una distribució *Debian* es realitza executant des de la consola la següent comanda:

```
apt-get install mongodb
```

Un cop instal·lat, s'ha d'arrancar el dimoni per poder connectar-nos al client de la base de dades a través de la comanda `mongod`. Cal destacar que el port que escolta el dimoni per defecte és el **27017**.

```
root@kali:~# mongod
mongod --help for help and startup options
2017-04-16T20:09:58.564-0400 [initandlisten] MongoDB starting : pid=9228 port=27017 dbpath=/data/db 64-bit host=kali
2017-04-16T20:09:58.565-0400 [initandlisten] db version v2.6.12
2017-04-16T20:09:58.565-0400 [initandlisten] git version:
2017-04-16T20:09:58.565-0400 [initandlisten] OpenSSL version: OpenSSL 1.0.2h  3 May 2016
2017-04-16T20:09:58.565-0400 [initandlisten] build info: Linux julia 4.6.0-1-amd64 #1 SMP Debian 4.6.4-1 (2016-07-18) x86_64 BOOST_LIB_VERSION=1_61
2017-04-16T20:09:58.565-0400 [initandlisten] allocator: tcmalloc
```

A continuació, s'ha d'afegir al fitxer de configuració ubicat a `/etc/mongodb.conf` l'adreça IP de la màquina client amb la que ens connectarem per realitzar les consultes, en el nostre cas l'adreça és `192.168.1.130`.

```
root@kali:~# cat /etc/mongodb.conf
# mongod.conf

# Where to store the data.
dbpath=/var/lib/mongodb

#where to log
logpath=/var/log/mongodb/mongod.log

logappend=true

bind ip = 127.0.0.1; 192.168.1.130
#port = 27017
```

Ara ja ens podem connectar a la base de dades a través de la comanda `mongo`. Per defecte es connecta a la base de dades "test".

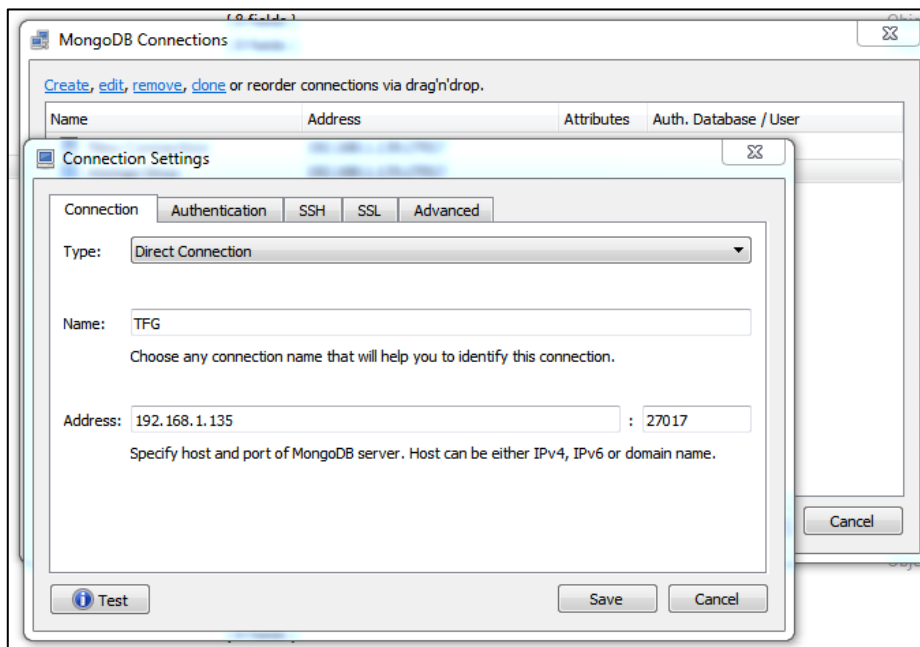
```
root@kali:~# mongo
MongoDB shell version: 2.6.12
connecting to: test
Server has startup warnings:
2017-04-14T19:40:59.727-0400 ** WARNING: --rest is specified without --httpinterface,
2017-04-14T19:40:59.727-0400 ** enabling http interface
```

A partir d'aquest pas, MongoDB ja està preparat per importar dades.

3.1.2 **Robomongo**

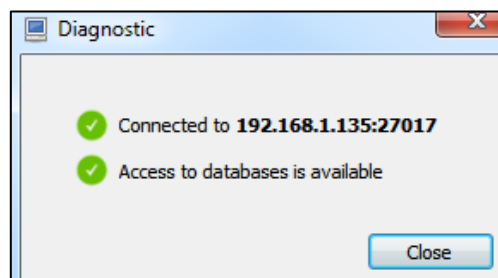
Complementàriament a *MongoDB*, a l'equip client s'instal·larà **Robomongo**, un programari gràfic que es connectarà com a client a la base de dades de *MongoDB* i facilitarà la manipulació de les col·leccions i documents des de l'entorn Windows. Es descàrrega a la màquina client des de la pàgina oficial <https://robomongo.org/>.

L'únic pas a destacar és, a l'hora de crear una connexió amb la nostra base de dades, el qual s'ha d'indicar l'adreça IP del servidor que és 192.168.1.135 i el port per defecte 27017.



Il·lustració 3 - Configuració de Robomongo contra la BBDD de MongoDB

Aquest pas és una prova per corroborar que MongoDB s'ha instal·lat i configurat correctament i està en funcionament. Si s'ha connectat correctament rebrem el següent missatge, que ens confirma que el servidor està disponible i ha aconseguit connectar-se a la base de dades.



La connexió s'ha realitzat a través d'un usuari anònim, això és perquè es tracta d'un entorn de proves. Quan entrés a producció caldria crear els usuaris des de MongoDB i restringir els permisos.

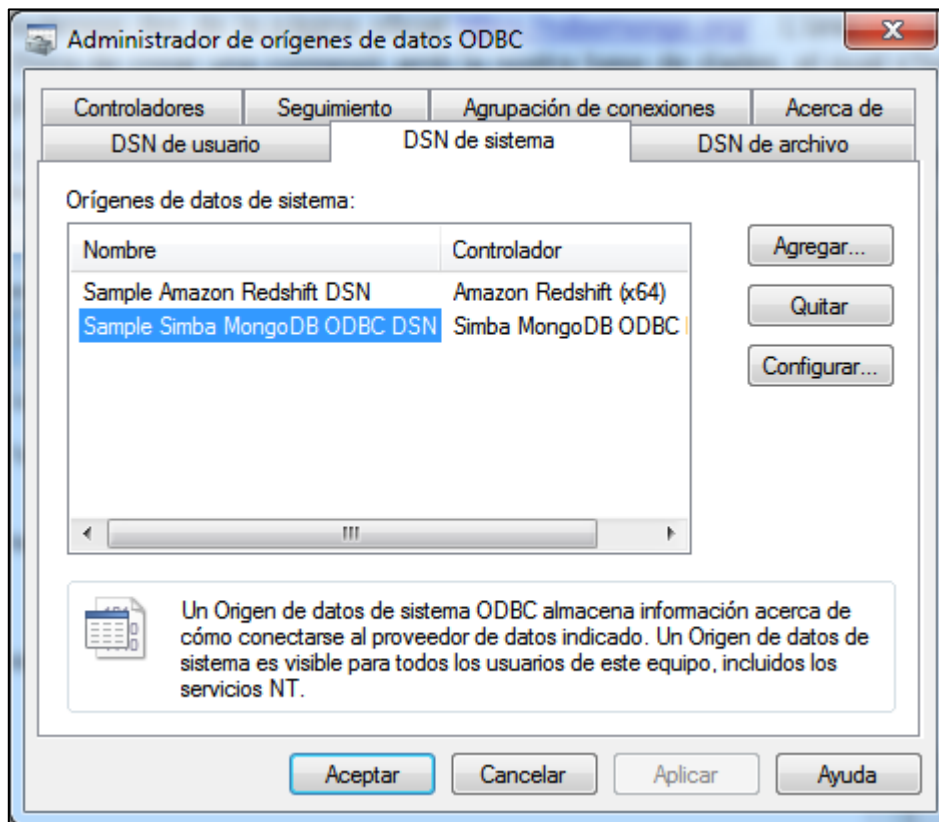
3.1.3 Simba MongoDB ODBC Driver

Per connectar la base de dades de *MongoDB* amb *TABLEAU*, és necessari un pas intermig que consisteix en d'instal·lació d'un *driver* a la màquina client per connectar ambdós programaris. Es descàrrega a la màquina client des de la pàgina oficial <http://www.simba.com/>

La llicència d'avaluació de 30 dies es rep per correu electrònic a l'adreça que es facilitar durant el registre per descarregar el programa. Aquesta llicència s'ha de copiar a la carpeta de repositoris de Tableau per a què conegui totes les funcionalitats que té la base de dades. La llicència s'ha de copiar a la següent ruta:

"C:\Users\XX\Documents\Mi Repositorio de Tableau\Fuentes de datos"

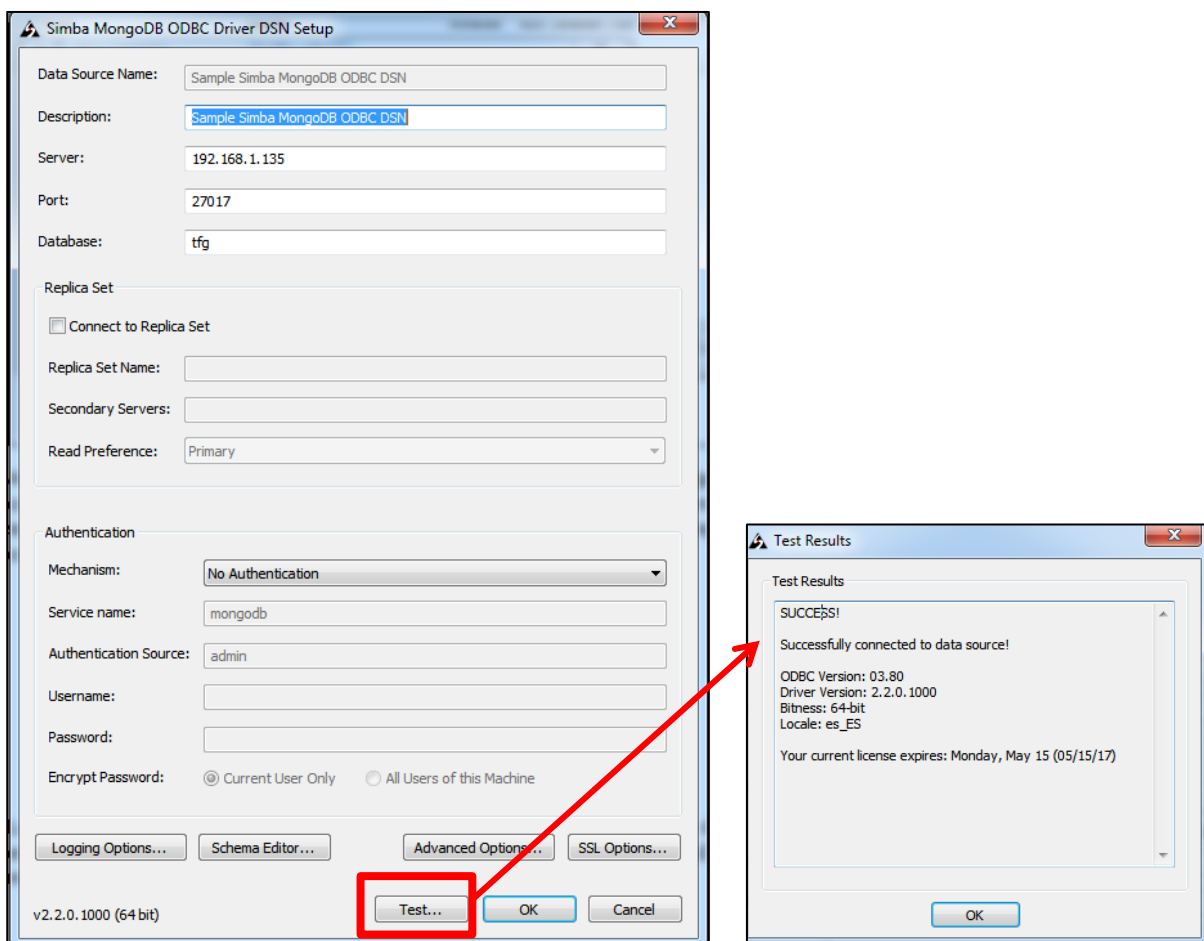
Una vegada instal·lat el programa i la llicència, s'arrancarà el programa i a la pestanya "DSN de sistema" es podrà observar que s'ha instal·lat el *driver* correctament.



Il·lustració 4 - Configuració de l'ODBC de SIMBA

A continuació, s'ha de seleccionar "Sample Simba MongoDB ODBC DSN" i prémer sobre el botó "configurar", tot seguit s'obrirà una finestra s'haurà d'omplir les dades de connexió del servidor, és a dir, l'adreça IP **192.168.1.135**, el port **27017** i la base de dades a la que es connectarà, que serà "**TFG**". Addicionalment, si s'hagués configurat algun tipus de compte d'usuari a MongoDB, seria necessari omplir les caselles d'autenticació.

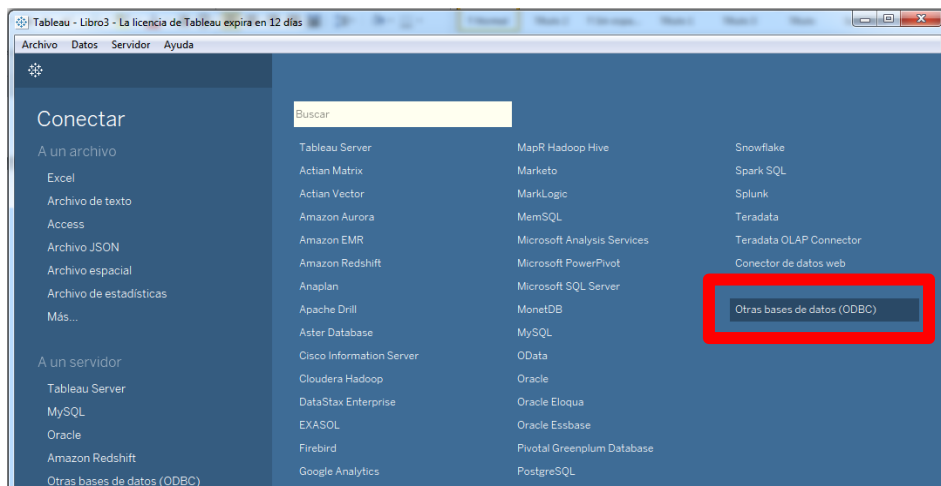
Finalment, s'ha de clicar sobre el botó "Test" i si s'ha configurat correctament les dades del servidor, s'obtindrà un missatge informant què és capaç de connectar-se amb la base de dades.



3.1.4 Tableau

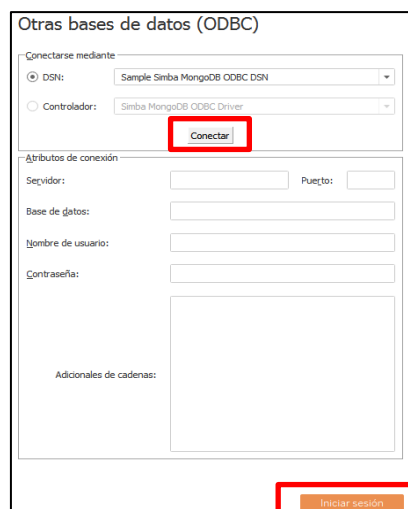
Des de la màquina client, es descarregarà i instal·larà una versió d'avaluació de *Tableau* a través de la seva pàgina web <https://www.tableau.com/products/desktop>.

Una vegada instal·lat, el programa permet extreure dades des de diferents fonts com ara Excel o arxius JSON, i també a diferents tipus de bases de dades tipus SQL, entre d'altres. Com que *MongoDB* no apareix en aquest llistat, hi ha un comodí per altres bases de dades configurades a través d'ODBC, que s'ha configurat en el pas anterior. Per tant, s'ha d'escollir l'opció "Otra base de datos (ODBC)".



Il·lustració 5 - Configuració de la connexió a Tableau

A continuació s'obrirà la següent finestra, on s'haurà d'escollir de la llista de desplegable l'API que s'ha configurat en el pas anterior, és a dir, el "*Simba MongoDB ODBC Driver*". *El seleccionem i cliquem sobre el botó "Conectar" i tot seguit s'activarà el botó inferior "Iniciar sesión"*.



3.2 ETL sobre les dades recollides

L'ETL o “*Export, Transform & Load*” és el procés d'extracció, transformació i càrrega de gran volum de dades des de diferents fonts amb l'objectiu de modificar el seu format d'origen i adequar-ho al format de l'estructura del destí.

A les organitzacions on la depuració constant de les dades és necessària, s'automatitzen parcial o totalment els processos de l'ETL amb un programa específic o mitjançant macros d'Excel per tal de reduir el temps invertit en recursos i minimitzar els errors humans que es produïrien si la depuració és realitzés manualment per una persona física

Així, per aquest projecte, en el primer esglaió de l'ETL s'han extret les dades rebudes per les diferents entitats en format Excel. En el següent pas, la transformació ha consistir en eliminar les files que no tenen utilitat.

Per exemple, en els fitxers descarregats a l'Agència Estatal de Meteorologia a més de disposar de les dades sobre les temperatures, també adjunten valors sobre velocitat del vent, que per l'estudi que es vol realitzar no són necessaris. A més, s'ha adequat el format de les cel·les al tipus de dada que representa, com ara les dates que s'ha hagut d'especificar el format DD/MM/AAAA.

Per altra banda, el fitxer rebut per la Policia de la Generalitat Mossos d'Esquadra, només ha requerit eliminar els accents, dièresis i el caràcter “ç”, ja que la base de dades de *MongoDB* només accepta caràcters en format UTF8.

Per últim, una vegada hem guardat ambdós fitxers en format CSV i abans de carregar les dades en el servidor, podem utilitzar la utilitat ***iconv*** de Linux, que serveix per convertir la codificació dels caràcters al format que li especifiquem, en aquest cas UTF8. Per fer-ho, hem d'executar la següent comanda:

```
iconv -f iso-8859-1 -t utf-8 /Fets_VIGE_BCN.csv > /Fets_VIGE_BCN_ok.csv
```

3.3 Càrrega de dades

El primer pas és crear la base de dades a ***MongoDB*** on es carregaran les dades que s'han depurat en el pas anterior i que després es voldran analitzar. Per fer-ho, s'ha d'executar la comanda `use tfg` en una consola del servidor i es crearà la base de dades.

Per poblar la base de dades, s'ha d'obrir una altre consola en el servidor i executar la comanda `mongoimport` indicant el nom de la base de dades, el nom de la col·lecció on es vol que es buidi el contingut del fitxer, el tipus de fitxer d'origen i la ruta:

```
mongoimport -d tfg -c clima --drop --stopOnError --type csv --file /Barcelonaclima.csv --ignoreBlanks --headerline
```

```
mongoimport -d tfg -c fets --drop --stopOnError --type csv --file /Fets_VIGE_BCN.csv --ignoreBlanks --headerline
```

Per `mongoimport`, els paràmetres que es poden afegir són els següents:

- d** : el nom de la base de dades
- c**: nom de la col·lecció
- drop**: eliminem la col·lecció en el cas que existís alguna amb el mateix nom
- stopOnError**: s'atura la importació en cas d'error
- type**: s'ha d'indicar l'extensió del tipus de fitxer que volem carregar.
- file**: ruta del fitxer
- headerline**: per indica que el fitxer conté una capçalera amb els noms de cada columna

Si la comanda s'ha executat correctament el sistema donarà una resposta semblant a la següent:

```
root@kali:~# mongoimport -d tfg -c clima --drop --stopOnError --type csv --file /Barcelonaclima.csv --ignoreBlanks --headerline
connected to: 127.0.0.1
2017-04-17T05:53:59.065-0400 dropping: tfg.clima
2017-04-17T05:53:59.173-0400 check 9 6201
2017-04-17T05:53:59.323-0400 imported 6200 objects
root@kali:~# mongoimport -d tfg -c fets --drop --stopOnError --type csv --file /Fets_VIGE_BCN.csv --ignoreBlanks --headerline
connected to: 127.0.0.1
2017-04-17T05:54:02.751-0400 dropping: tfg.fets
2017-04-17T05:54:02.767-0400 check 9 537
2017-04-17T05:54:02.777-0400 imported 536 objects
```

Es comprova que s'han creat les col·leccions de cada fitxer amb la comanda

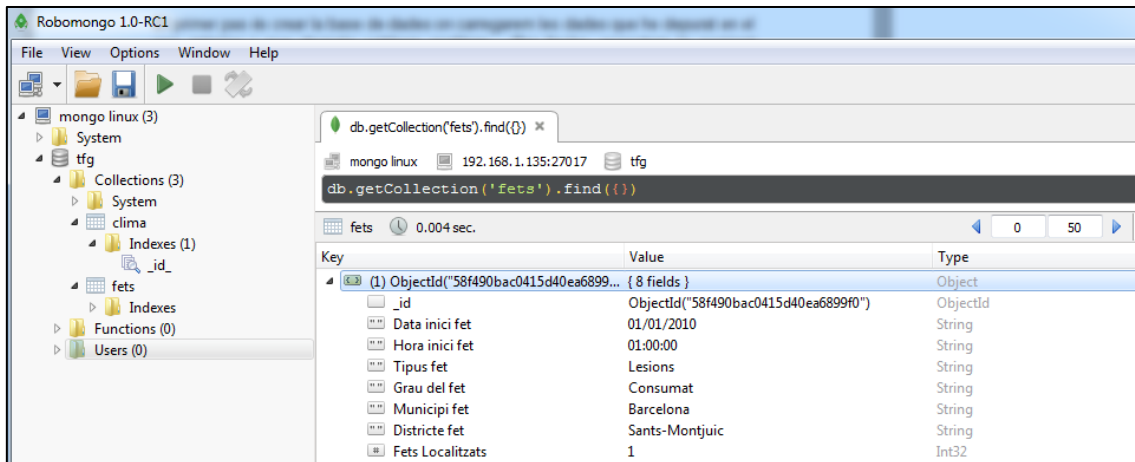
```
show collections
```

Amb la comanda `db.fets.find()` es pot comprovar el contingut de les col·leccions, és a dir, els documents. S'ha de recordar que els documents equivaldria a les files en les taules relacionals i les col·leccions a les taules senceres.

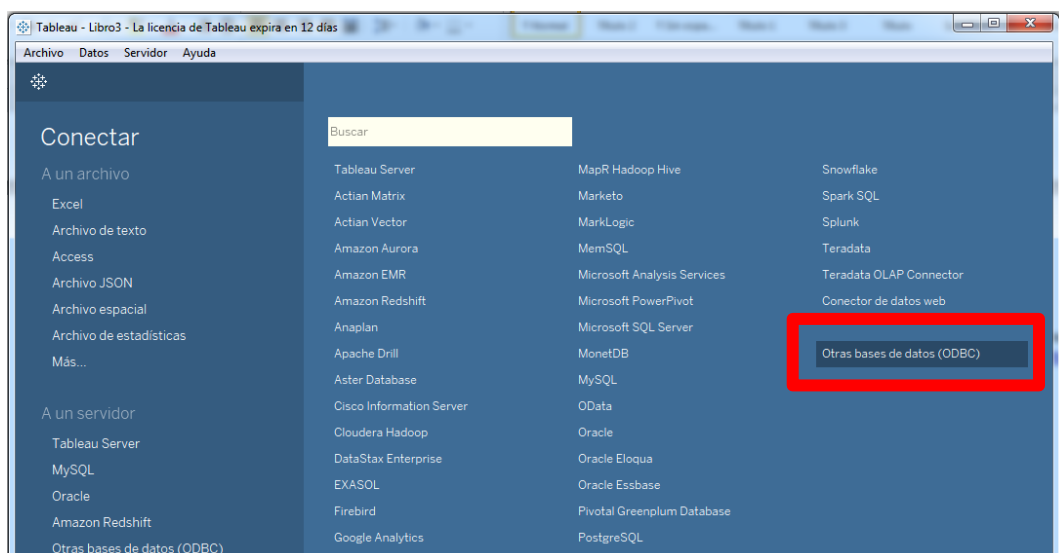
```
> db.fets.find()
{ "_id" : ObjectId("58f490bac0415d40ea6899f0"), "Data inici fet" : "01/01/2010", "Hora inici fet" : "01:00:00", "Tipus fet" : "Lesions",
"Grau del fet" : "Consumat", "Municipi fet" : "Barcelona", "Districte fet" : "Sants-Montjuïc", "Fets Localitzats" : 1 }
{ "_id" : ObjectId("58f490bac0415d40ea6899f1"), "Data inici fet" : "07/01/2010", "Hora inici fet" : "14:35:00", "Tipus fet" : "Lesions",
"Grau del fet" : "Consumat", "Municipi fet" : "Barcelona", "Districte fet" : "Sant Andreu", "Fets Localitzats" : 1 }
{ "_id" : ObjectId("58f490bac0415d40ea6899f2"), "Data inici fet" : "12/01/2010", "Hora inici fet" : "00:05:00", "Tipus fet" : "Lesions",
"Grau del fet" : "Consumat", "Municipi fet" : "Barcelona", "Districte fet" : "Sant Andreu", "Fets Localitzats" : 1 }
{ "_id" : ObjectId("58f490bac0415d40ea6899f3"), "Data inici fet" : "12/01/2010", "Hora inici fet" : "14:00:00", "Tipus fet" : "Falta de
esions", "Grau del fet" : "Consumat", "Municipi fet" : "Barcelona", "Districte fet" : "Eixample", "Fets Localitzats" : 1 }
```

3.4 Entorn de proves

Una vegada s'han carregat les dades, en la màquina client i des de l'aplicació *Robomongo* podem manipular i representar les dades.

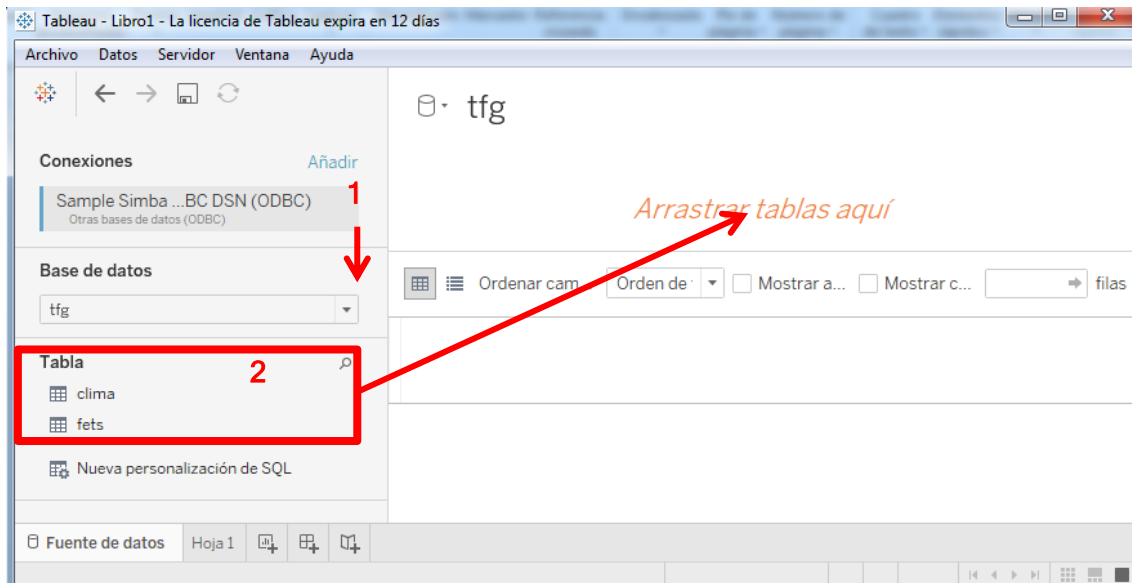


Arranquem l'aplicació Tableau i seleccionem l'opció "Otra base de datos (ODBC)".

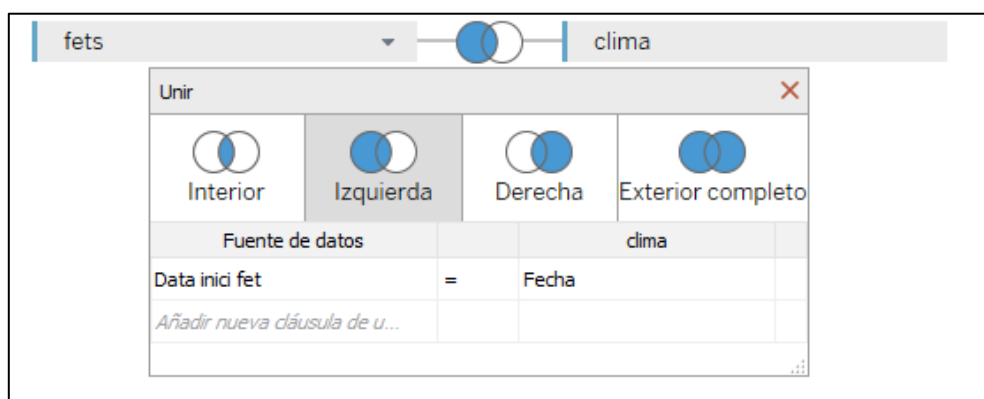


Ens connectem al driver "*Simba MongoDB ODBC Driver*" que ja hem preparat a l'apartat de configuració.

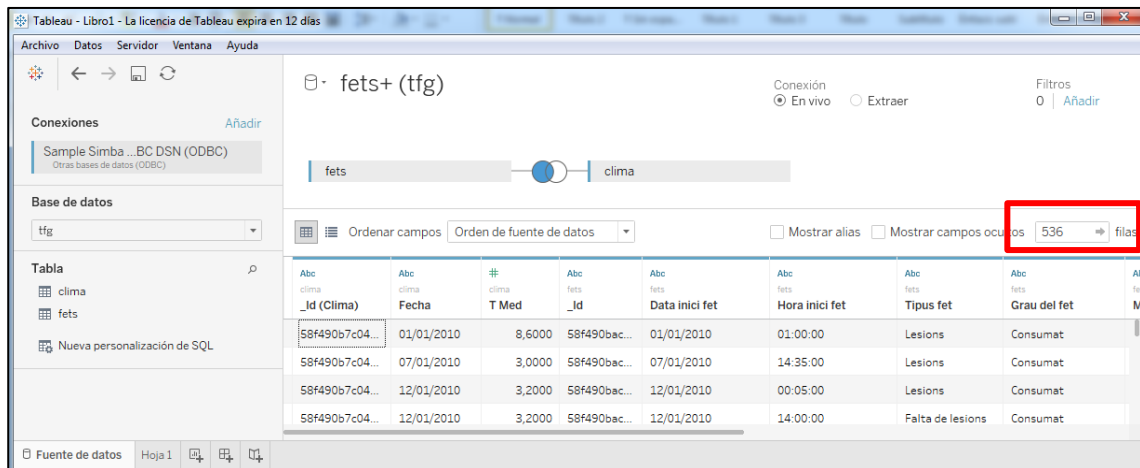
Seleccionem del desplegable el nom de la base de dades, concretament “tfg” que és la que s’ha creat per aquest projecte i en la part esquerra apareixerà les col·leccions que conté: “clima” i “fets”



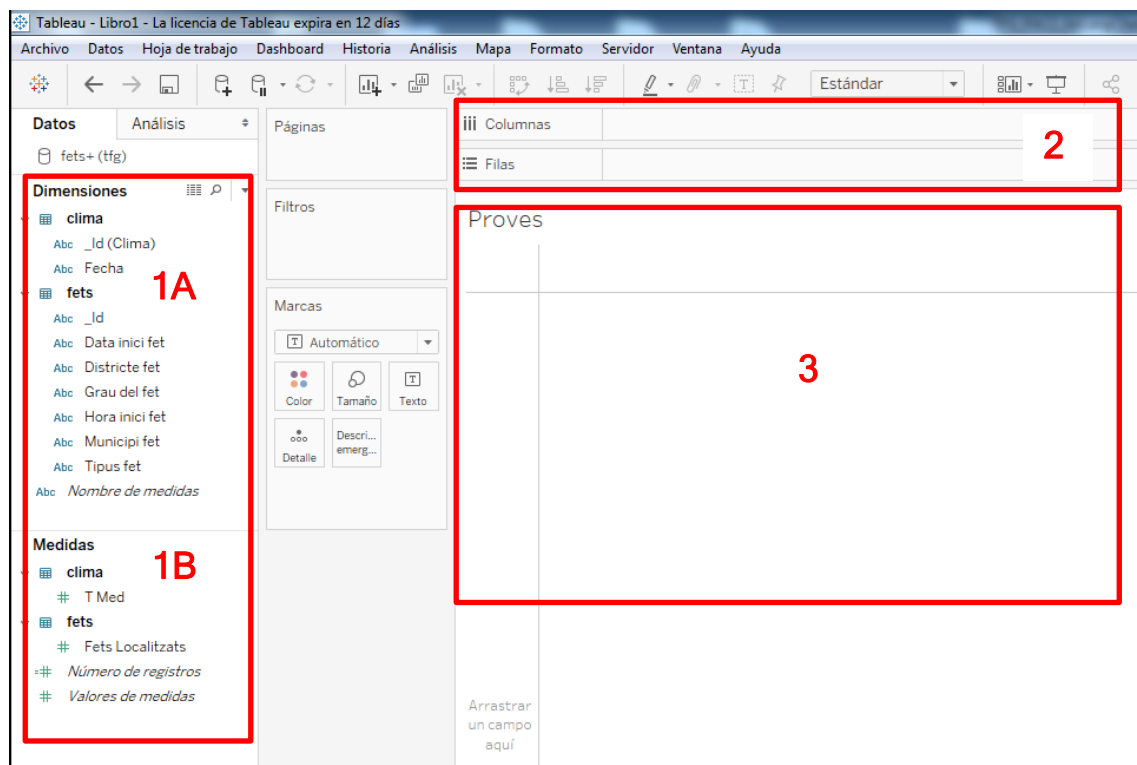
A continuació s’arrastra les col·leccions “clima” i “fets” cap al centre de la pantalla i es mostraran les dades de les col·leccions. Es pot especificar a Tableau quin tipus de conjunt o subconjunt de dades es vol generar. En el cas que ens ocupa, s’ha escollit com a prova que la intersecció dels dos grups de dades amb els camps de nexa “Data inici fet” i “data” del clima.



A continuació es pot observar les dades que s'legeix directament de la base de dades de MongoDB i que amb el conjunt de dades que hem escollit, apareixen 536 registres, que coincideix amb el número de fets que hi ha registrats a Barcelona.



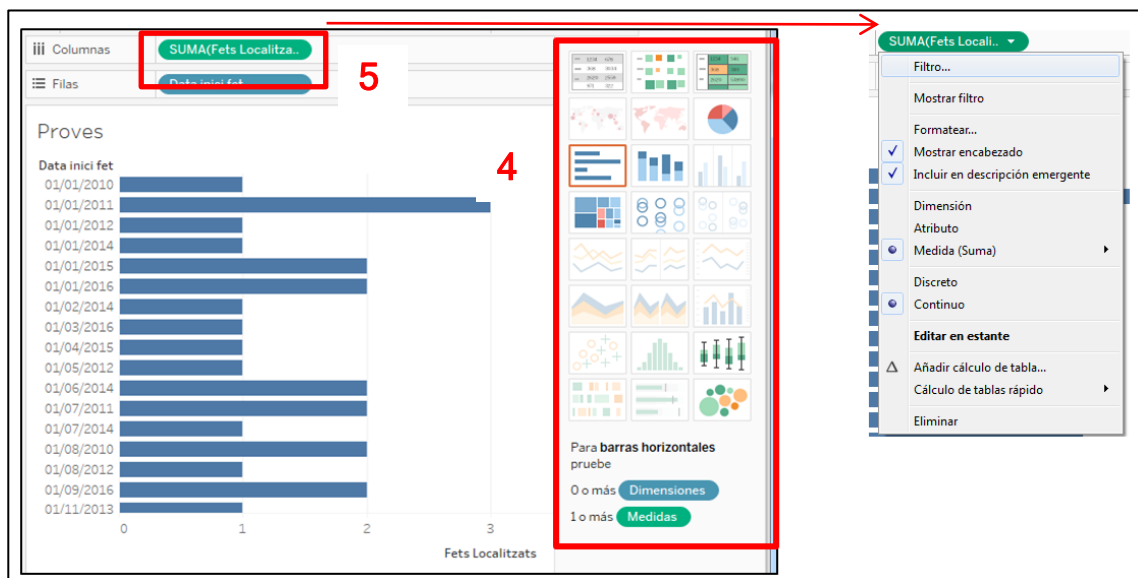
Una vegada s'ha assegurat que Tableau té accés a les col·leccions de la base de dades TFG de MongoDB, podem començar a fer proves a l'etiqueta que hem anomenat com a proves. En aquesta secció podem identificar a grans trets 3 parts:



- **Secció 1:** En aquest apartat podem visualitzar els camps de cada col·lecció. En l'apartat 1A normalment es mostren els camps amb atributs que contenen una cadena o enters que no siguin valors calculats, com ara una suma, producte o mitjana, aquest altres valors es mostrarien a l'apartat 1B.
- **Secció 2:** Per representar un gràfic, podem arrastrar els camps de les col·leccions de la secció 1 cap a la secció 2, en funció de si volem col·locar el valor en el eix de les X (files) o de les Y (columnes).
- **Secció 3:** La representació gràfica dels valors que haguem escollit es representaran en aquest apartat.

Com a exemple, si arrastrem el camp “**Data inici fet**” a l'eix de les X i “**Fets localitzats**”, que és el número de fets coneguts obtindrem el següent gràfic (per defecte gràfic de barres). Observem que a l'apartat de la dreta (secció 4) podem escollir quin tipus de gràfics es vol representar. Cal destacar que només s'habilitaran quan hi hagi prou dades per representar un tipus de gràfic, sinó l'opció restarà en gris.

A més, podem aplicar filtres sobre les dades representades (secció 5).



Amb aquestes proves, s'ha comprovat el funcionament del programari que s'ha instal·lat i que les dades carregades a la base de dades són llegides correctament pel programari BI.

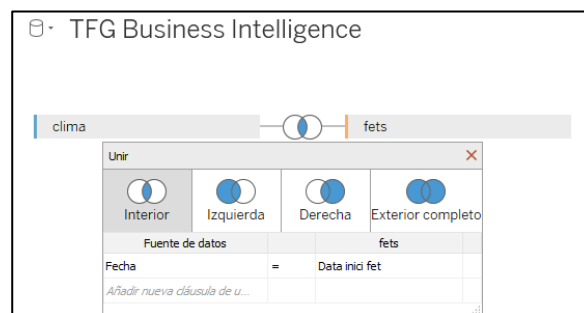
4 Explotació, anàlisi i representació de dades

S'han tingut en compte les següents consideracions prèvies a l'hora de treballar les dades tractades:

- El número de fets registrats/any per homicidi i assassinat (àmbit de violència de gènere) són molt baixos (de l'ordre de 1), per tant, l'anàlisi s'ha centrat en els delictes de lesions que tot i que són mostres de valor baix (<5), són dades representatives superiors a la resta de fets.
- Per problemes d'obtenció de temperatures diàries arreu de Catalunya, només s'han aconseguit aquestes dades relacionades amb la ciutat de Barcelona. Per això, l'anàlisi dels fets es centra en aquesta ciutat.
- S'han utilitzat les dades de temperatura mitjanes, tot i que el més òptim hauria estat conèixer la temperatura a l'hora de cada fet.

Pel que fa a l'àmbit tecnològic, per tal de realitzar l'anàlisi de dades que prèviament s'han sincronitzat de *MongoDB* a *Tableau*, s'ha d'especificar quin tipus de conjunt o subconjunt de dades es volen combinar indicant quin és l'element en comú, en aquest cas només interessa els dies dels fets delictius i les temperatures corresponents a aquell dia.

Per tant, s'ha de combinar de la col·lecció **clima**, el camp "*Fecha*" amb el camp "*Data inici fet*", de la col·lecció de **fets** i escollir la intersecció dels dos conjunts, que només inclourà els valors amb coincidències a les dues col·leccions.



A continuació es pot observar les dades que s'obtenen directament de la base de dades de MongoDB i que amb el conjunt de dades que s'ha escollit, apareixen 532 registres, que coincideix amb el número de fets que hi ha registrats a la ciutat de Barcelona.

Fecha	T Max	Hora Tmax	T Min	Hora Tmin	T Med	Data inici fet	Hora inici fet	Tipus fet	Grau del fet	Municipi fet	Districte fet
01/01/2010	11,40	14:33	5,8	Varias	8,6000	01/01/2010	1:00:00	Lesions	Consumat	Barcelona	Sants-Montjuic
07/01/2010	4,40	7:54	1,5	17:05	3,0000	07/01/2010	14:35:00	Lesions	Consumat	Barcelona	Sant Andreu

4.1 Representació gràfica

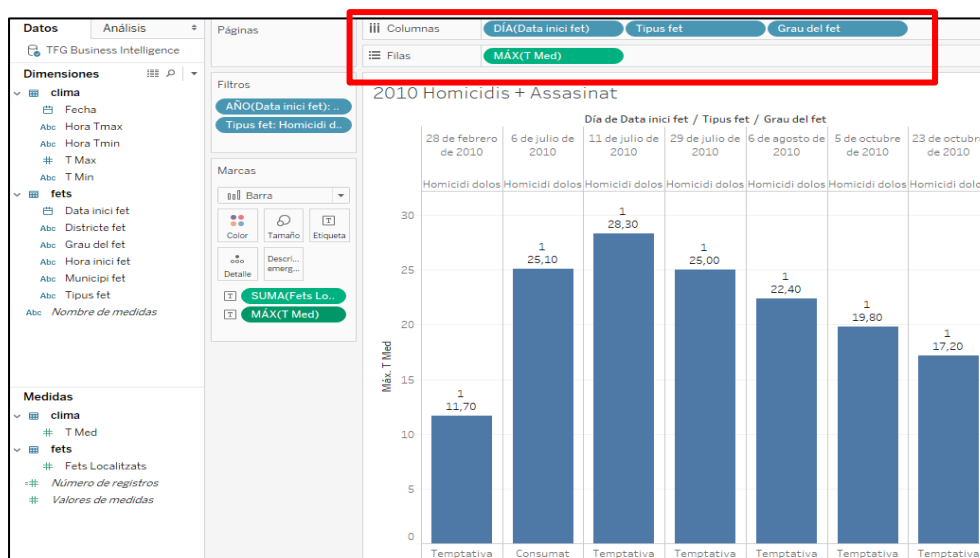
Per realitzar l'anàlisi s'han desgranat els fets per anys amb l'objectiu de poder visualitzar les dades representades i les tendències amb major detall.

Pel que respecta a l'elaboració dels gràfics, s'han utilitzat a l'eix de les Y (files) la temperatura mitjana i a l'eix de les X (columnes) els següents camps:

Data inici fet: Data en què van succeir els fets.

Tipus de fet: El recull de dades conté els diferents delictes emmarcats en l'àmbit de la violència de gènere, és a dir, lesions (faltes i delictes), homicidis dolosos i assassinats.

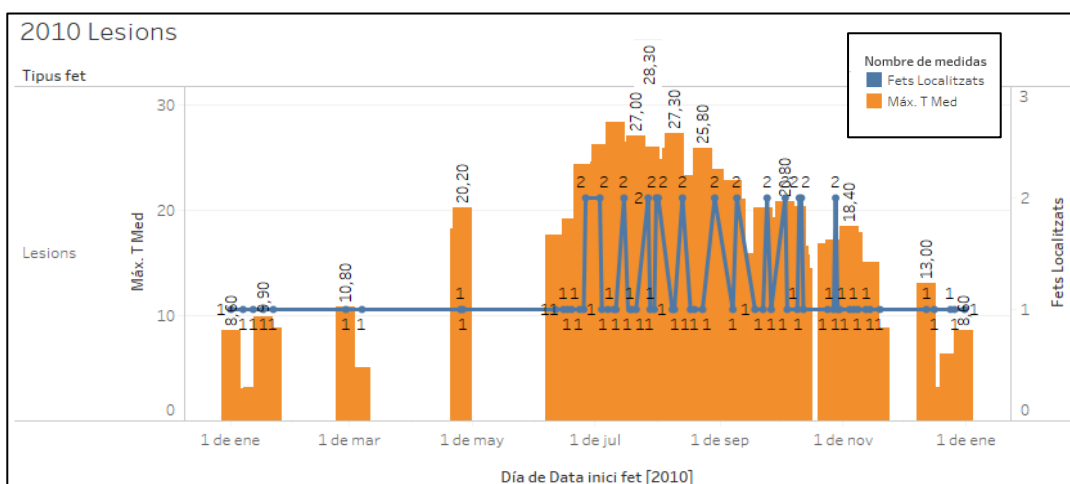
Grau del fet: Diferència el resultat final del tipus de fet, és a dir, si és en grau de temptativa o consumat.



4.1.1 Any 2010

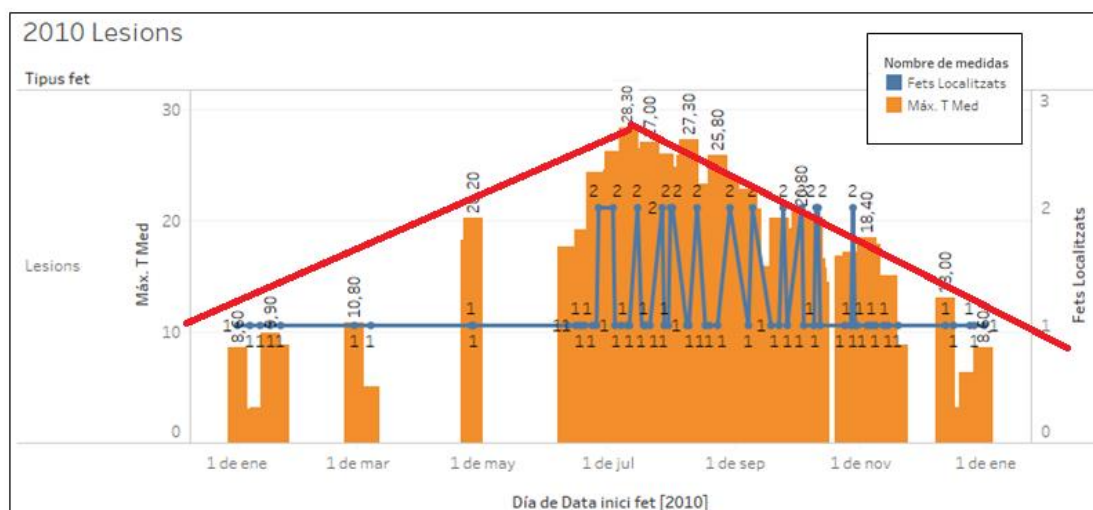
Sobre els delictes de lesions, es van recollir 83 fets a l'any 2010, el següent gràfic representa en color taronja les **temperatures mitjanes** i de color blau en forma de línia de tendència **el número de fets localitzats**.

Tot i que la mostra és petita ja que el número de fets per dia varia entre 1 i 3, podem observar que les lesions es concentren entre els mesos de juliol i setembre, on la temperatura oscil·la entre els 15,30° i 28,30°. Si entrem en més detall i només ens fixem en els dies que hi ha hagut com a mínim 2 fets, el rang de temperatura s'estableix entre els 24,30° i els 28,30°.



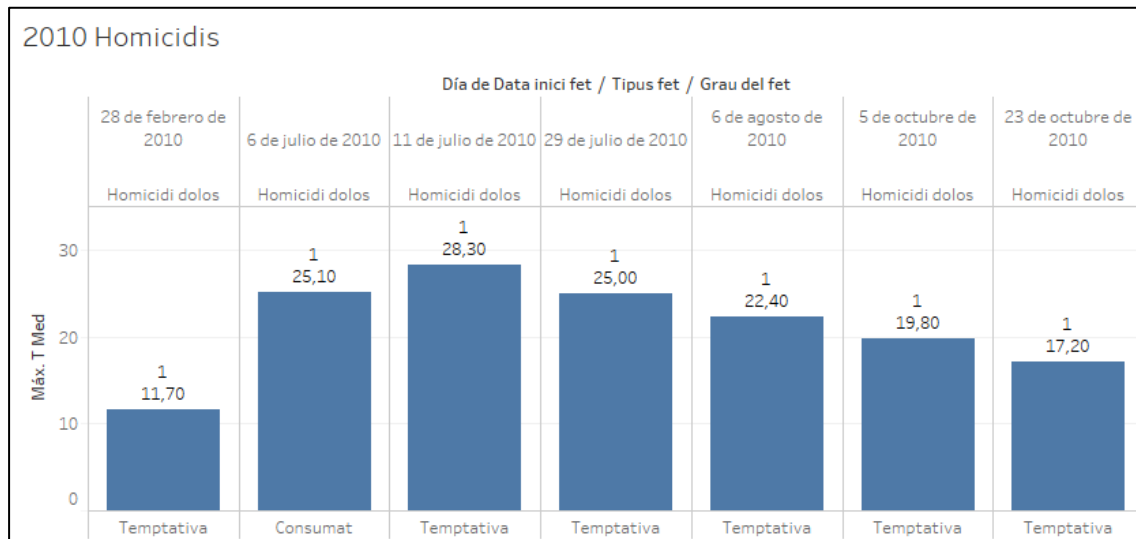
Il·lustració 6 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2010

El següent gràfic és el mateix que l'anterior, però s'ha afegit manualment una línia vermella on es pot observa la forma d'una piràmide on la cresta és situa al 11 de juliol de 2010, sent la temperatura d'aquest dia 28,30°, la 4ª temperatura més alta de l'any 2010.



Respecte els **homicidis dolosos es van produir un total de 7 fets**, dels quals 6 van ser en grau de temptativa i 1 consumat.

5 dels homicidis, és a dir, un 71,42%, és van produir entre els mesos de juliol i octubre on la temperatura oscil·lava entre els 19,80° i 28,30°.



Il·lustració 7 - Gràfic de barres sobre els homicidis i la temperatura l'any 2010

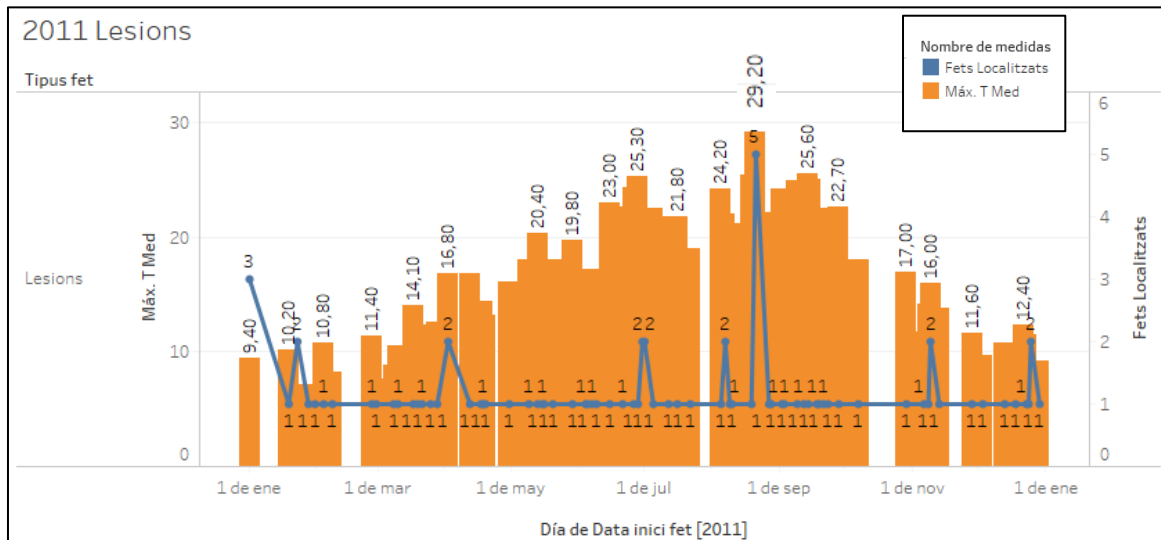
4.1.2 Any 2011

L'any 2011 es van produir 91 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 5.

Cal destacar que el **21 d'Agost** va ser el dia en què es van produir **5 fets** (el nombre de fets més alt de l'any) i coincidint amb el dia que va registrar una temperatura mitjana més alta de **29,20°**, que va ser la més alta durant tot l'any 2011.

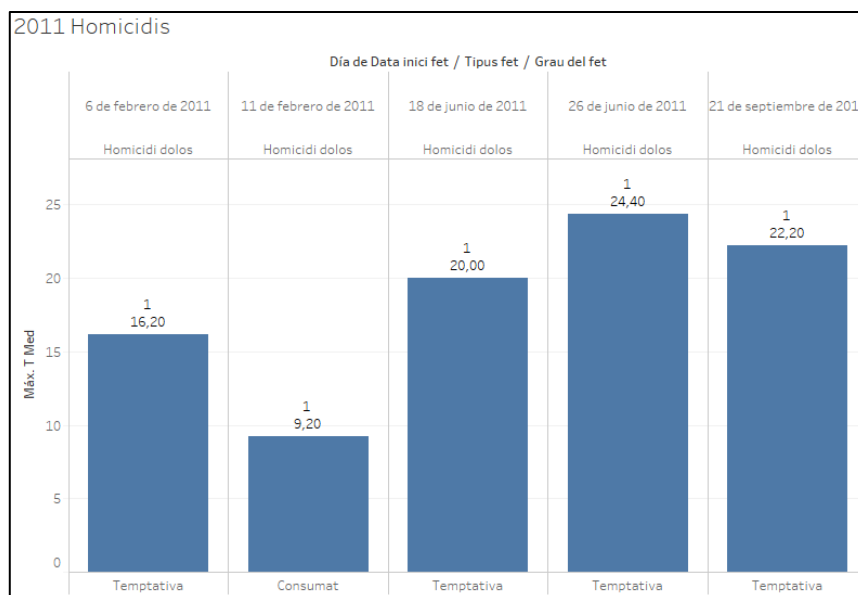
Data inici fet	Districte	Fecha	Grau del fet	Hora inici fet	Tipus fet	T Med	Hora Tmax
21/08/2011	Les Corts	21/08/2011	Consumat	16:20:00	Lesions	29,2000	Varias
21/08/2011	Sant Marti	21/08/2011	Consumat	17:00:00	Lesions	29,2000	Varias
21/08/2011	Ciutat Vella	21/08/2011	Consumat	20:00:00	Lesions	29,2000	Varias
21/08/2011	Eixample	21/08/2011	Consumat	21:45:00	Lesions	29,2000	Varias
21/08/2011	Sant Andreu	21/08/2011	Consumat	5:00:00	Lesions	29,2000	Varias

Per altra banda, no passa despercebut que el segon dia amb més fets produïts durant l'any 2011 va ser l'1 de gener amb 3 fets. A la hipòtesis inicial d'aquest treball es proposava estudiar la relació entre la temperatura i la violència. Malgrat això, també existeixen altres variables i factors com és en aquest cas, que es tractava d'un dia festiu.



Il·lustració 8 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2011

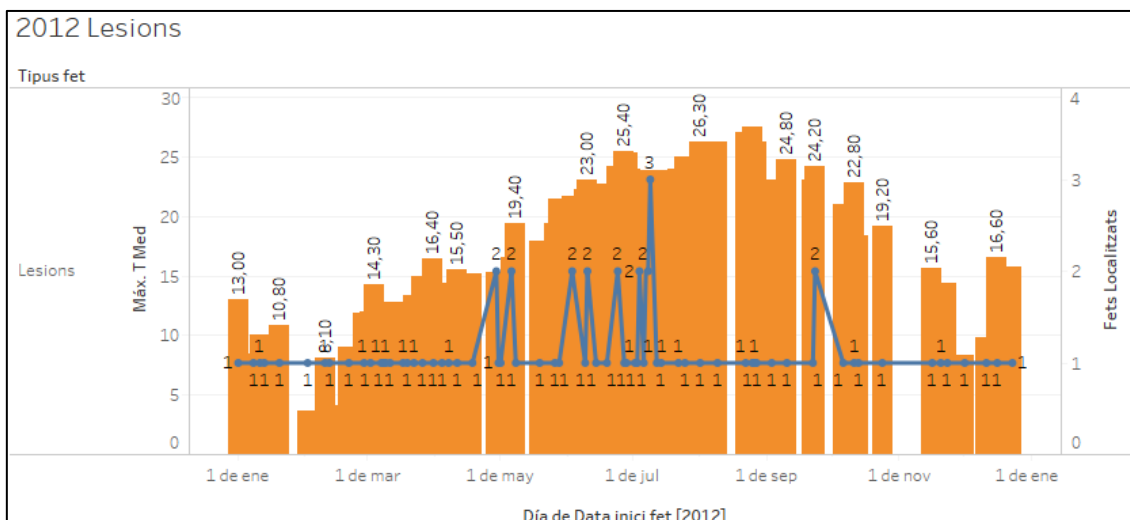
Respecte als **homicidis dolosos es van produir un total de 5 fets**, dels quals només 1 va ser consumat. Els fets es troben més dispersats durant tot l'any i la temperatura dels dies dels fets oscil·lava entre els 9,20° i 24,40°.



Il·lustració 9 - Gràfic de barres sobre els homicidis i la temperatura l'any 2011

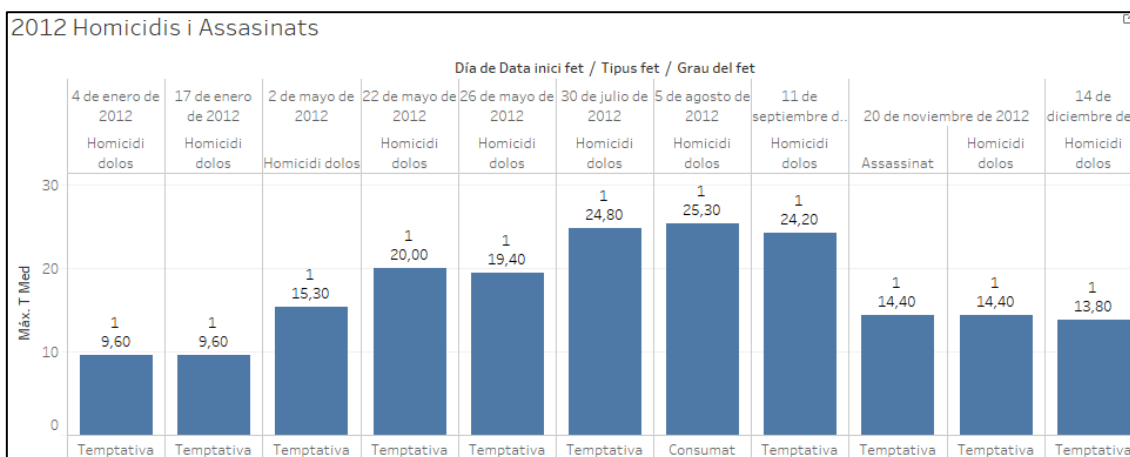
4.1.3 Any 2012

L'any 2012 es van produir 90 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 3. El dia 9 de juliol va ser el dia que es van produir el màxim de fets de l'any, 3 fets i la temperatura mitjana d'aquell dia va ser 23,80°.



Il·lustració 10 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2012

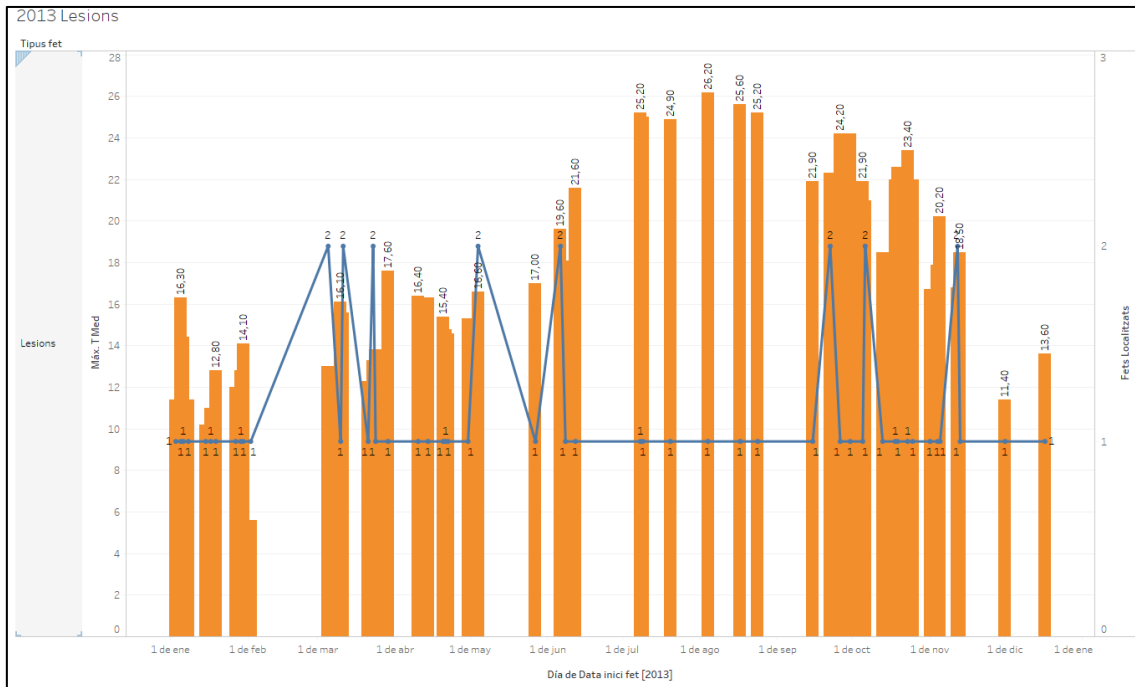
Respecte als **homicidis dolosos es van produir un total de 10 fets**, dels quals només 1 va ser consumat. També es va produir un assassinat el 20 de novembre on la temperatura mitjana era de 14,40°. Els fets es troben més dispersats durant tot l'any i la temperatura dels dies dels fets oscil·lava entre els 9,60° i 25,30°.



Il·lustració 11 - Gràfic de barres sobre els homicidis i la temperatura l'any 2012

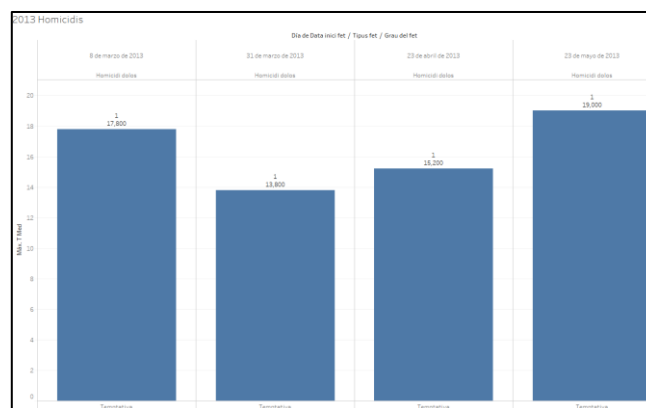
4.1.4 Any 2013

L'any 2013 es van produir 61 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 2. Els dies que es van produir 2 fets, la temperatura oscil·lava entre els 13,0° i els 21,90°.



Il·lustració 12 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2013

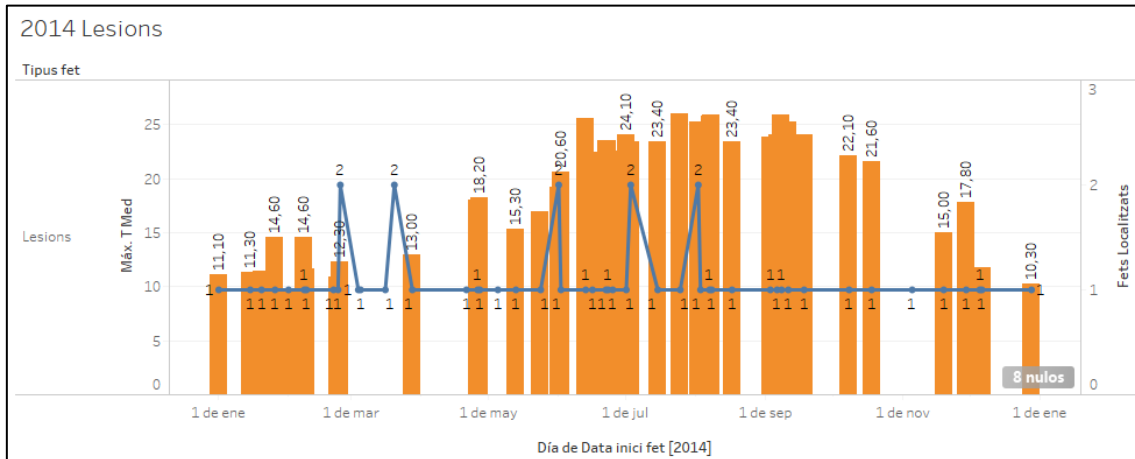
Respecte als **homicidis dolosos** es van produir un total de 4 fets, tots en grau de temptativa i que es van produir agrupats entre el mes de març i maig, amb temperatura entre els 13,80° i 19,20°. Com ja s'ha comentat, poden existir altres factors a part de la temperatura, en aquest cas podria influir que coincidís amb setmana santa.



Il·lustració 13 - Gràfic de barres sobre els homicidis i la temperatura l'any 2013

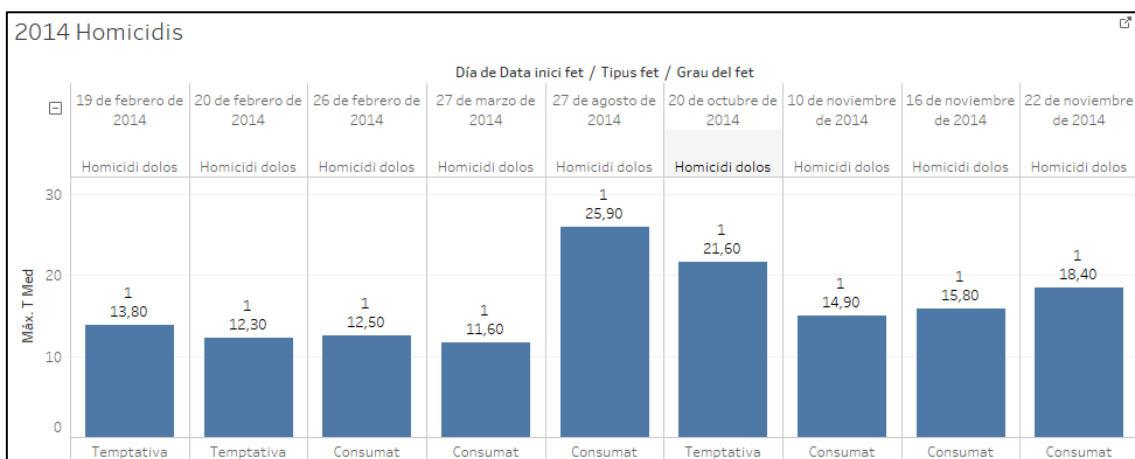
4.1.5 Any 2014

L'any 2014 es van produir 54 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 2. Els dies que es van produir 2 fets, entre el mes de març i agost, la temperatura va oscil·lar entre els 12,30° i els 25,30°.



Il·lustració 14 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2014

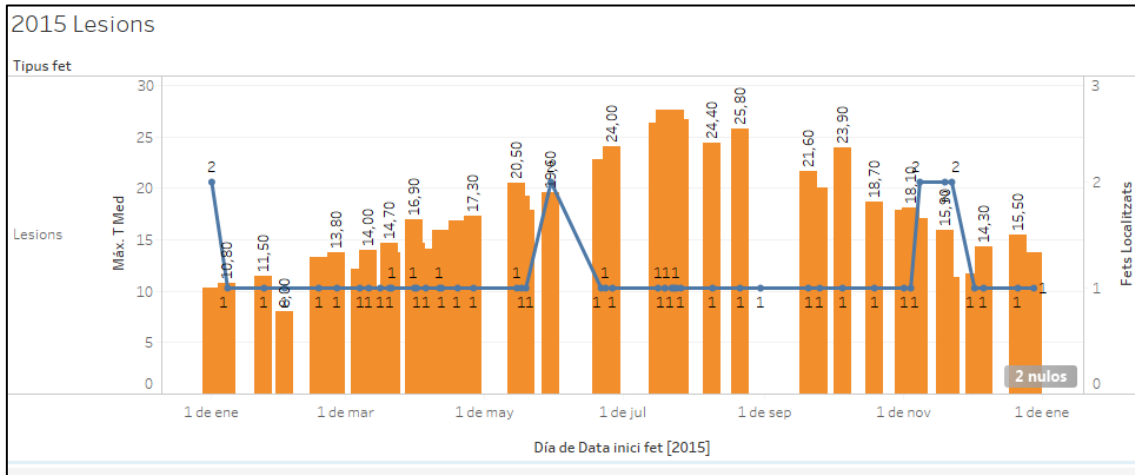
Respecte als **homicidis dolosos es van produir un total de 9 fets**, dels quals 6 van ser consumats. Els fets es troben més dispersats durant tot l'any i la temperatura dels dies dels fets oscil·lava entre els 11,60° i 25,90°.



4.1.6 Any 2015

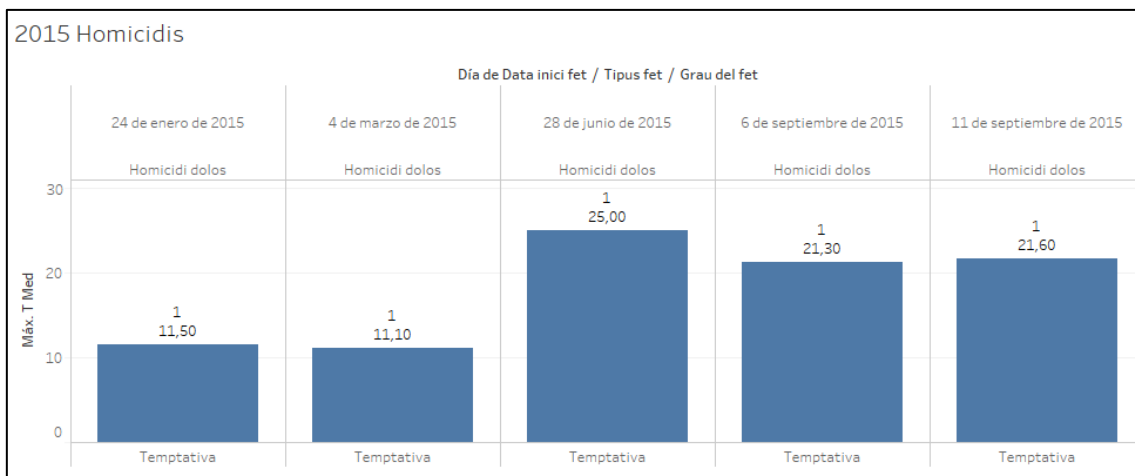
L'any 2015 es van produir 52 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 2. Els dies que es van produir 2 fets, la temperatura va oscil·lar entre els 10,30° i els 19,60°.

Una vegada més, l'1 de gener, és un dels dies que destaca, com ja s'ha dit possiblement perquè es tracta d'un dia festiu.



Il·lustració 16 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2015

Respecte als **homicidis dolosos es van produir un total de 5 fets**, tots en grau de temptativa. Els fets es troben més dispersats entre gener i setembre i la temperatura dels dies dels fets oscil·lava entre els 11,50° i 25°.

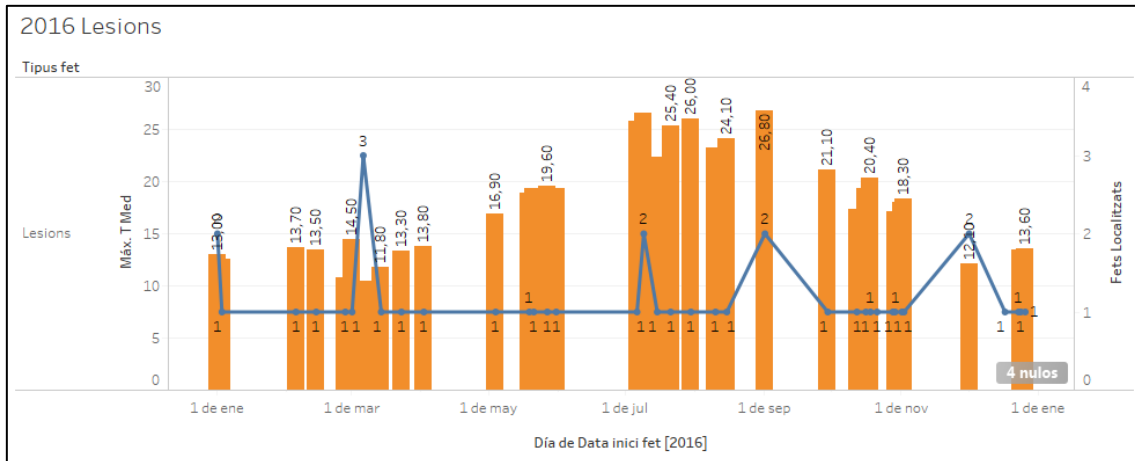


Il·lustració 17 - Gràfic de barres sobre els homicidis i la temperatura l'any 2015

4.1.7 Any 2016

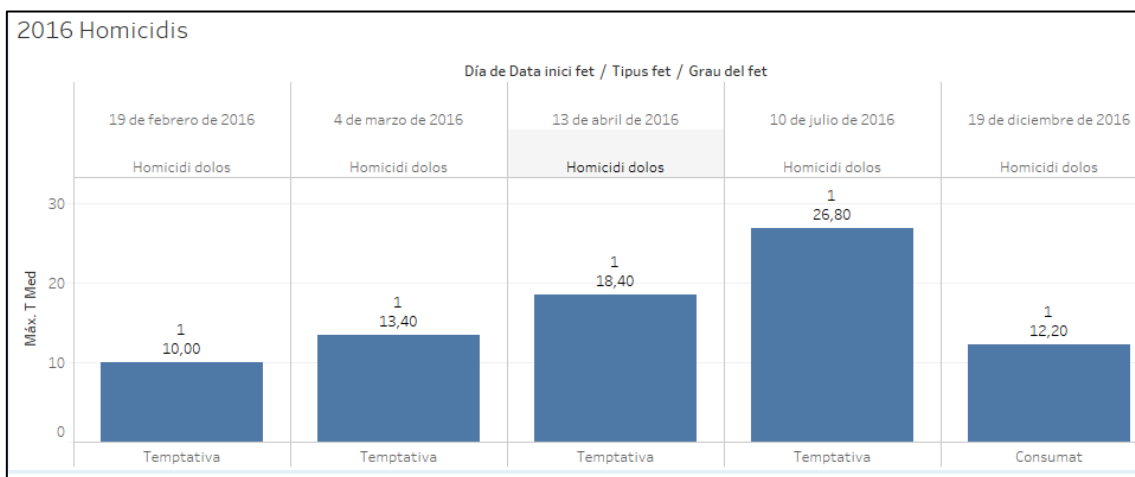
L'any 2015 es van produir 43 fets relacionats amb delictes de lesions i el número de fets per dia varia entre 1 i 3. Els dies que es van produir 2 fets, la temperatura va oscil·lar entre els 10,50° i els 26,80°.

El 6 de març, que va ser el dia en què es van registrar 3 fets era un diumenge, el que equivaldria a un dia festiu.



Il·lustració 18 - Gràfic de barres sobre els delictes de lesions i la temperatura l'any 2015

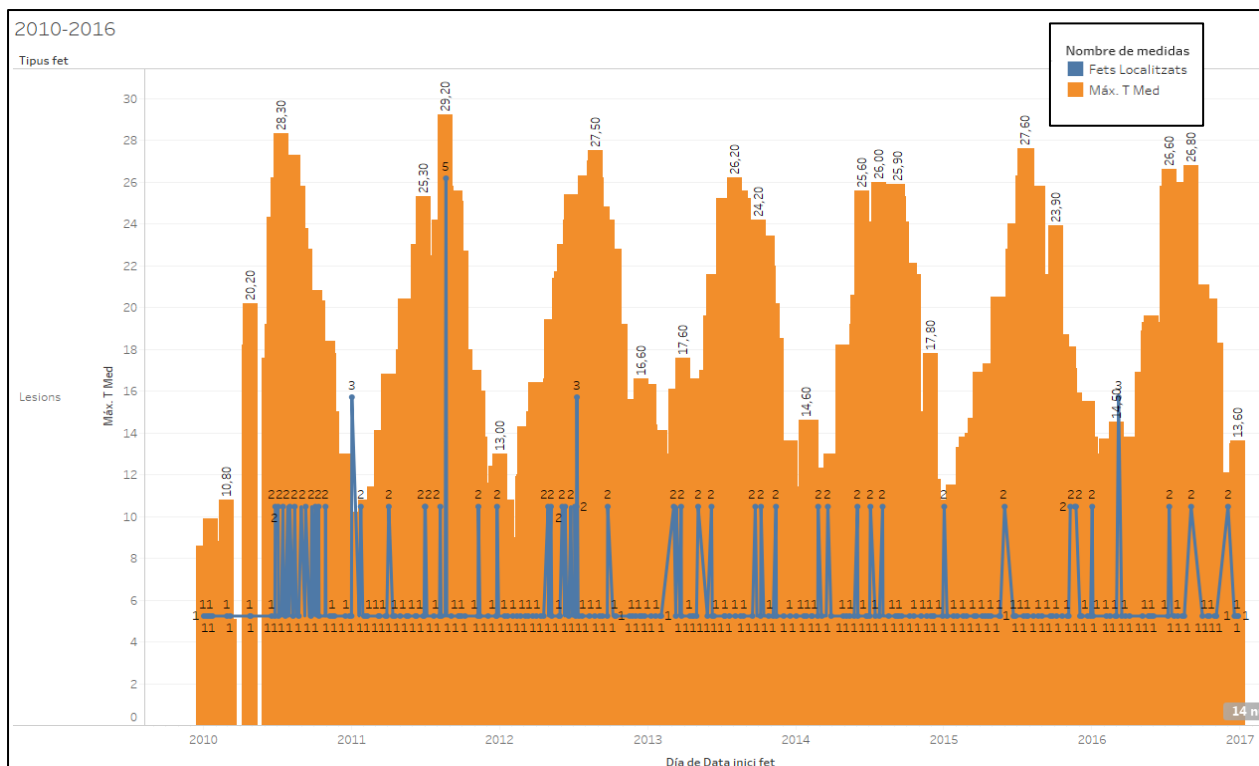
Respecte als **homicidis dolosos es van produir un total de 5 fets**, només 1 de consumat. Els fets es troben més dispersats entre gener i setembre i la temperatura dels dies dels fets oscil·lava entre els 10° i 26,80°.



Il·lustració 19 - Gràfic de barres sobre els homicidis i la temperatura l'any 2016

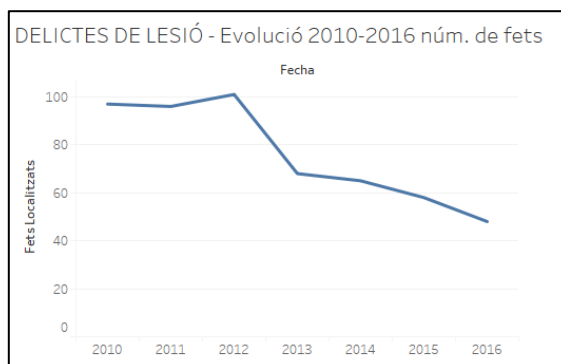
4.1.8 Anys 2010-2016

El següent gràfic representa les temperatures que s'han registrat en el moment de produir-se un o més delictes de lesions. Es pot observar que, aproximadament per sobre dels 20º, es quan succeeixen les baralles, en canvi, quan les temperatures són baixes (<10º) no és registren tants fets en comparació.



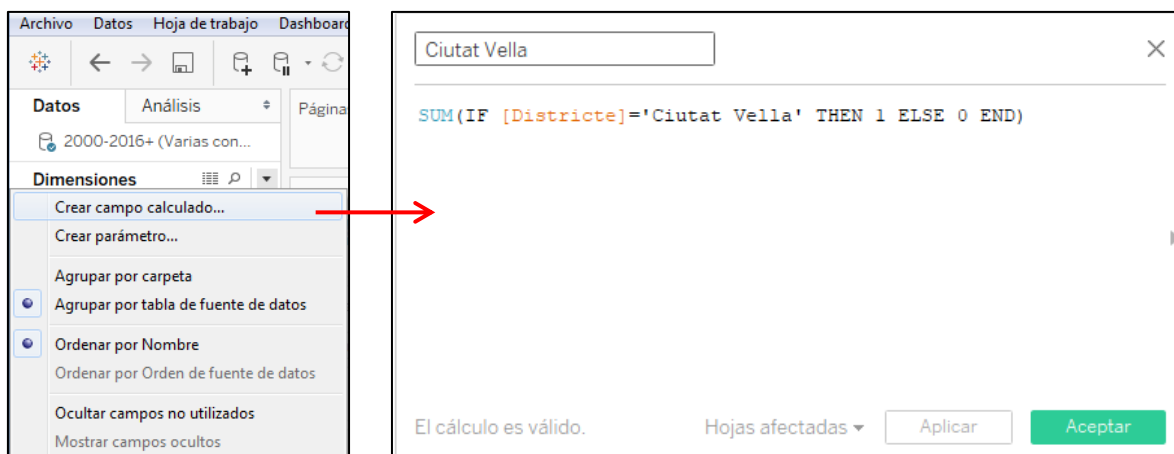
Il·lustració 20 - Gràfic de barres sobre els delictes de lesions i les temperatures entre els anys 2010 i 2016

Tot i que no és un motiu d'estudi en aquest projecte, un fet a destacar que s'ha detectat durant l'anàlisi, és que els **delictes de lesions des de l'any 2010 fins l'any 2016 han anat decreixent**. Un dels motius podria ser que, les campanyes de conscienciació del Ministeri de Sanitat i de Serveis Socials, entre d'altres estigui contribuint a les reduccions d'aquesta tipologia delictiva.



4.2 Gràfics a partir de camps calculats

Tableau permet crear nous camps a partir de càlculs d'altres camps. Com a part de l'estudi, s'han analitzat quines comissaries han rebut més denúncies entre els anys 2010 i 2016. Per fer-ho, s'ha de crear un camp calculat per cada comissaria que consistirà en la suma d'elements que compleixi una condició, concretament, que sigui igual al nom d'un districte.



Il·lustració 21 - Creació de camps calculats

SUM(IF [Districte]='Ciutat Vella' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Eixample' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Sant Marti' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Nou Barris' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Sants-Montjuic' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Horta-Guinardo' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Sant Andreu' THEN 1 ELSE 0 END)

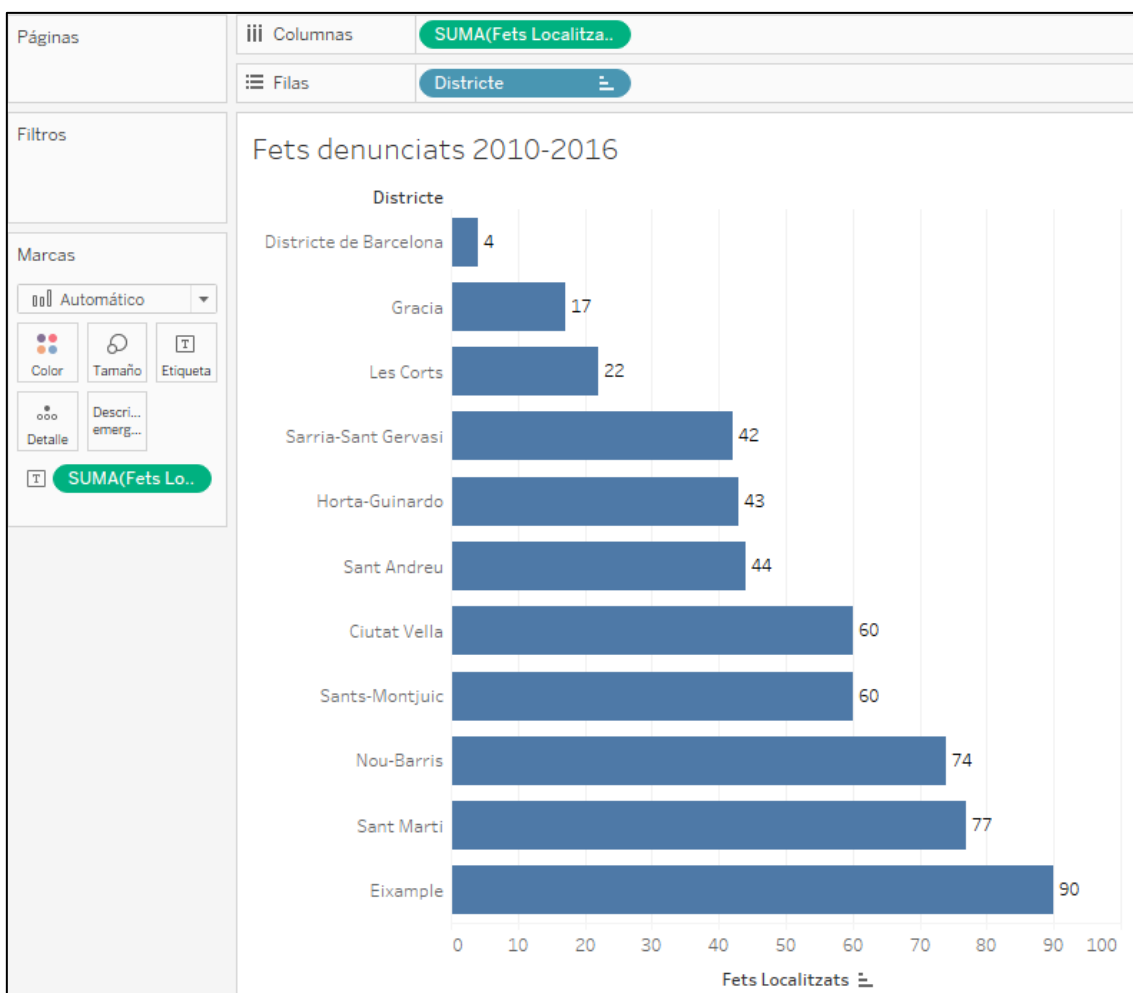
SUM(IF [Districte]='Sarria-Sant Gervasi' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Les Corts' THEN 1 ELSE 0 END)

SUM(IF [Districte]='Gracia' THEN 1 ELSE 0 END)

Així doncs, si es representa en un gràfic de barres horitzontals, quines comissaries de districte han rebut més denúncies per fets emmarcats en l'àmbit de la violència de gènere entre els anys 2010 i 2016, en el següent gràfic podem observar que la llista és encapçalada per la comissaria d'**Eixample amb 90 fets** i la **comissaria de Gràcia amb 17 fets** és la que menys denúncies rep.

Per altra banda, s'ha de tenir present que hi ha 4 fets recollits com a "Districte de Barcelona" que no especifica en quina comissaria van ser recollides les denúncies.



Il·lustració 22 - Gràfic de barres sobre fets denunciats en l'àmbit de la violència de gènere a les diferents comissaries de la regió metropolitana de Barcelona.

4.3 El mapa com a representació

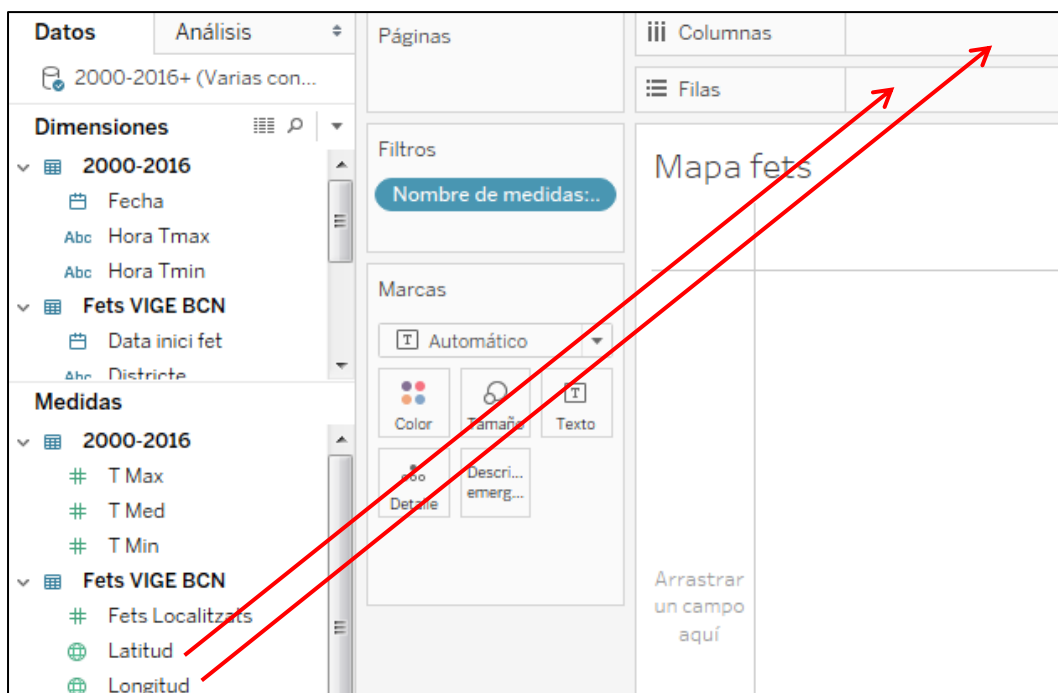
Tableau permet crear diferents tipus de mapes per representar gràficament les dades, per aquest projecte s'ha escollit el tipus de mapa de símbols proporcionals, que són idonis per representar dades quantitatives en ubicacions individuals, podent assignar una mida en funció de la seva magnitud.

Per aquest projecte, únicament és disposa dels districtes on s'han denunciat els fets. Així, s'ha elaborat un llistat de les 10 comissaries de la Regió Policial Metropolitana de Barcelona i s'han afegit dos camps més a la base de dades, sent aquests les coordenades en format UTM (latitud i longitud) de cada comissaria de districte.

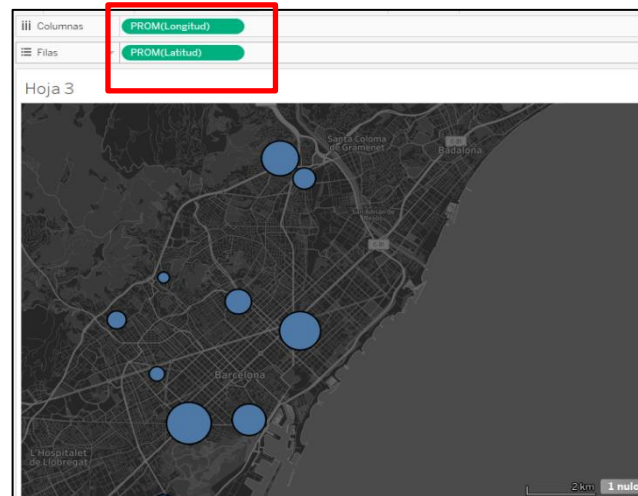
De manera que el JOIN queda representat a la següent il·lustració amb les taules de les **temperatures** juntament amb els **fets** i les dues noves columnes de les **coordenades** de cada comissaria.

#	2000-2016	#	2000-2016	#	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016	Fets VIGE BCN	2000-2016
T Max	Hora Tmax	T Min	Hora Tmin	T Med	Data inici fet	Hora inici fet	Tipus fet	Fets VIGE BCN	Municipi fet	Districte	Fets VIGE B.	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN	Fets VIGE BCN
19,0000	0:00	14,2000	4:23	16,6000	0/10/2010	9:50:00	Lesions	Consumat	Barcelona	Sant Marti	1	41,4012398	2,1892591						
25,1000	13:50	18,7000	5:40	21,9000	5/10/2013	9:50:00	Lesions	Consumat	Barcelona	Nou-Barris	1	41,4493512	2,1823247						
26,7000	2:10	23,6000	22:40	25,2000	8/07/2013	9:49:00	Lesions	Consumat	Barcelona	Gracia	1	41,4160519	2,1416407						
15,3000	14:00	8,4000	8:00	11,8000	6/12/2014	9:35:00	Lesions	Consumat	Barcelona	Sant Marti	1	41,4012398	2,1892591						
14,2000	14:00	6,4000	1:20	10,3000	1/01/2015	9:30:00	Lesions	Consumat	Barcelona	Sant Andreu	1	41,4487657	2,1905183						

En conseqüència, Tableau comptabilitzarà els fets per cada districte i en funció del seu volum, representarà al mapa el cercle amb una mida més o menys gran que la resta. Per fer-ho, s'han d'arrastrar els valors de les longituds i latituds als camps de columnes i files, respectivament.

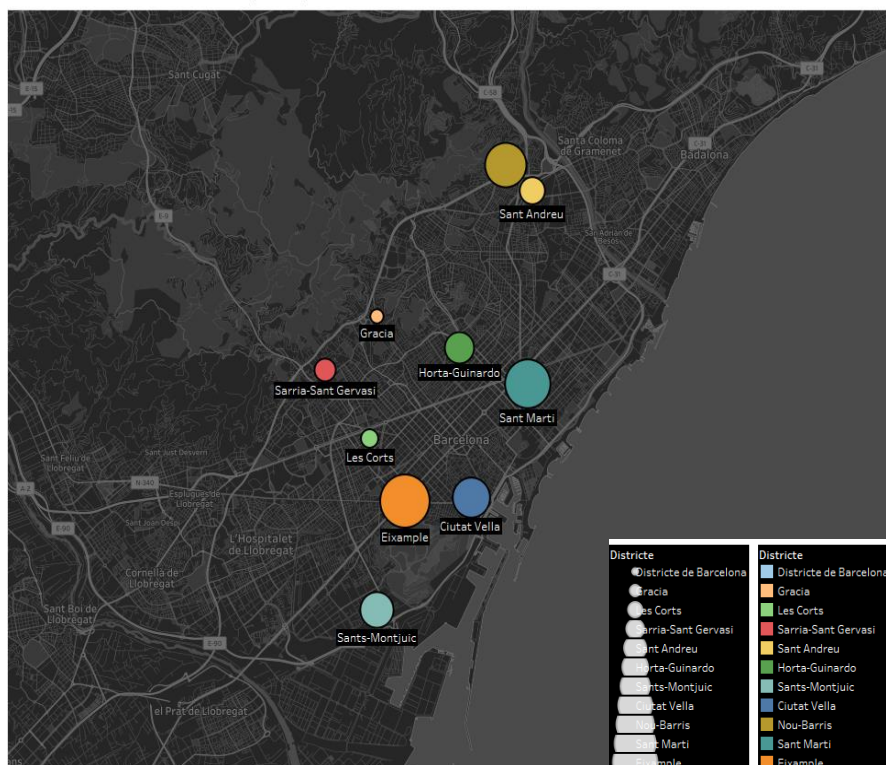


Una vegada arrastrat els valors, es generarà el mapa automàticament. Els punts blaus representen on es troba ubicat cada comissaria on s'han denunciat fets per violència de gènere i la mida del cercle representar el volum de fets recollits, és a dir, quan més gran més fets s'han denunciat.



A continuació, *Tableau* permet donar un format al mapa per a poder distingir amb més claredat els diversos elements i afegir l'etiqueta a cada districte, de manera que el mapa final queda del següent manera.

Fets denunciats 2010-2016 (mapa)



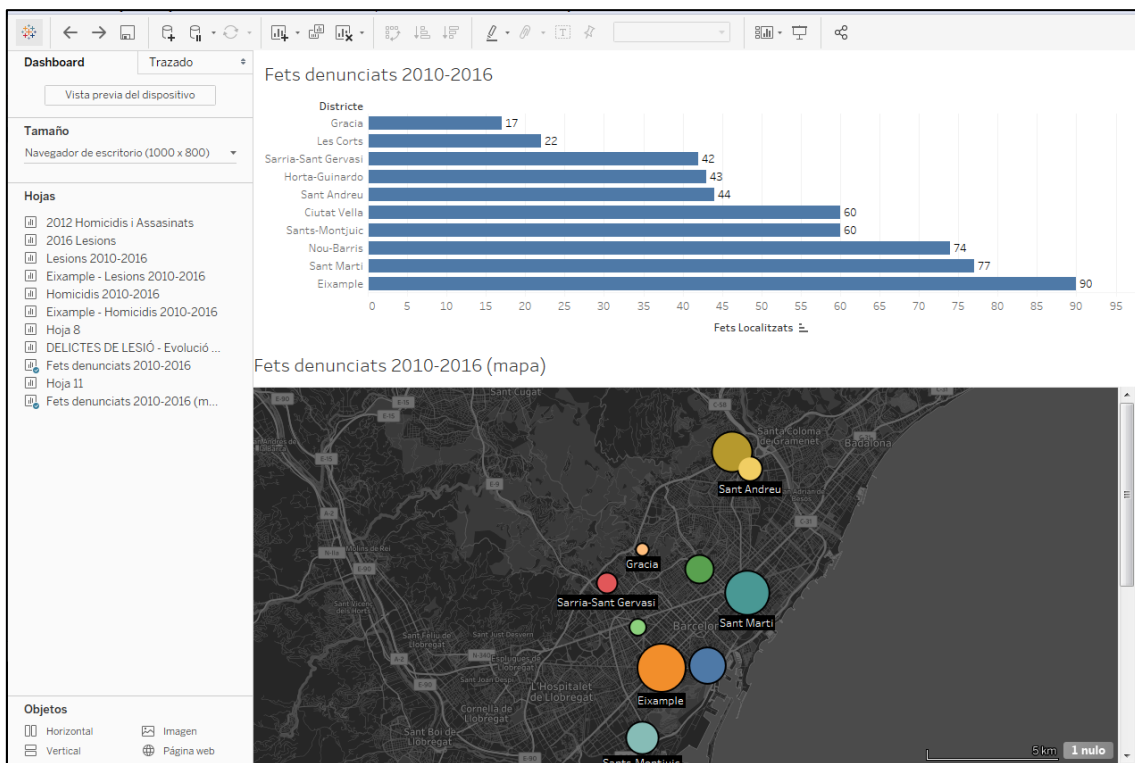
II-Il·lustració 23 - Mapa generat a partir de les coordenades i dels fets denunciats a Barcelona

4.4 Dashboard

Tableau permet crear *dashboards*, que són una representació gràfica dels principals indicadors que afectarien a una organització i que esta orientat a la presa de decisions i que es podria utilitzar com a eina per generar un quadre de comandament.

Un dels seus avantatges, és que en cas de modificació d'altres gràfics, els elements representats al *dashboard* i afectats pels canvis s'actualitzen automàticament.

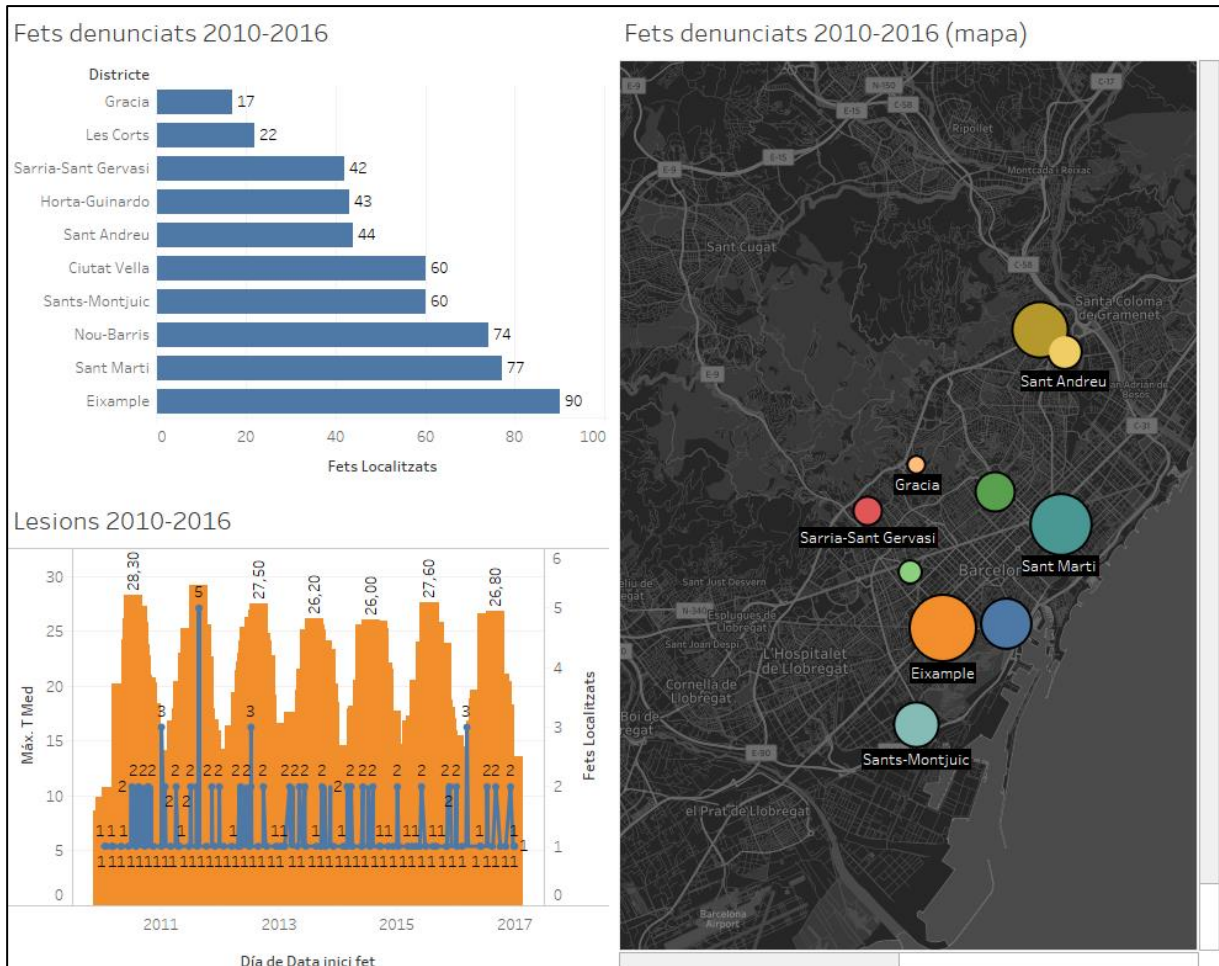
En el cas que ens ocupa, hipotèticament l'organització desitja conèixer quines comissaries són les que han rebut més denúncies per violència de gènere entre els anys 2010 i 2016. A la següent il·lustració extreta del *dashboard*, a simple vista es pot apreciar que Eixample és la comissaria que més denúncies rep (90 fets) i on es troba ubicada geogràficament.



Il·lustració 24 - *Dashboard* de Tableau combinat amb un gràfic de barres i un mapa

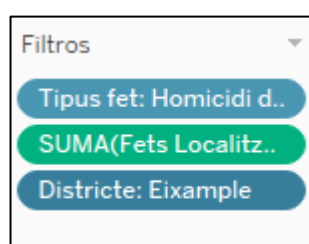
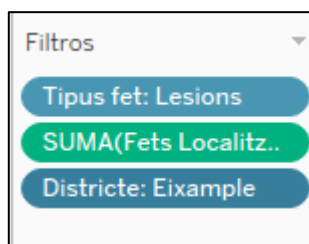
Així, en funció de les necessitats i preguntes que necessiti contestar l'organització, es pot personalitzar el *dashboard* i dedicant poc temps, es poden interpretar les dades dels gràfics.

En el següent gràfic, s'ha afegit el full que conté les lesions entre els anys 2010 i 2016, podent observar que a l'estiu del 2011, concretament el 21 d'agost es van recollir 5 fets i on la temperatura va ser de 29°

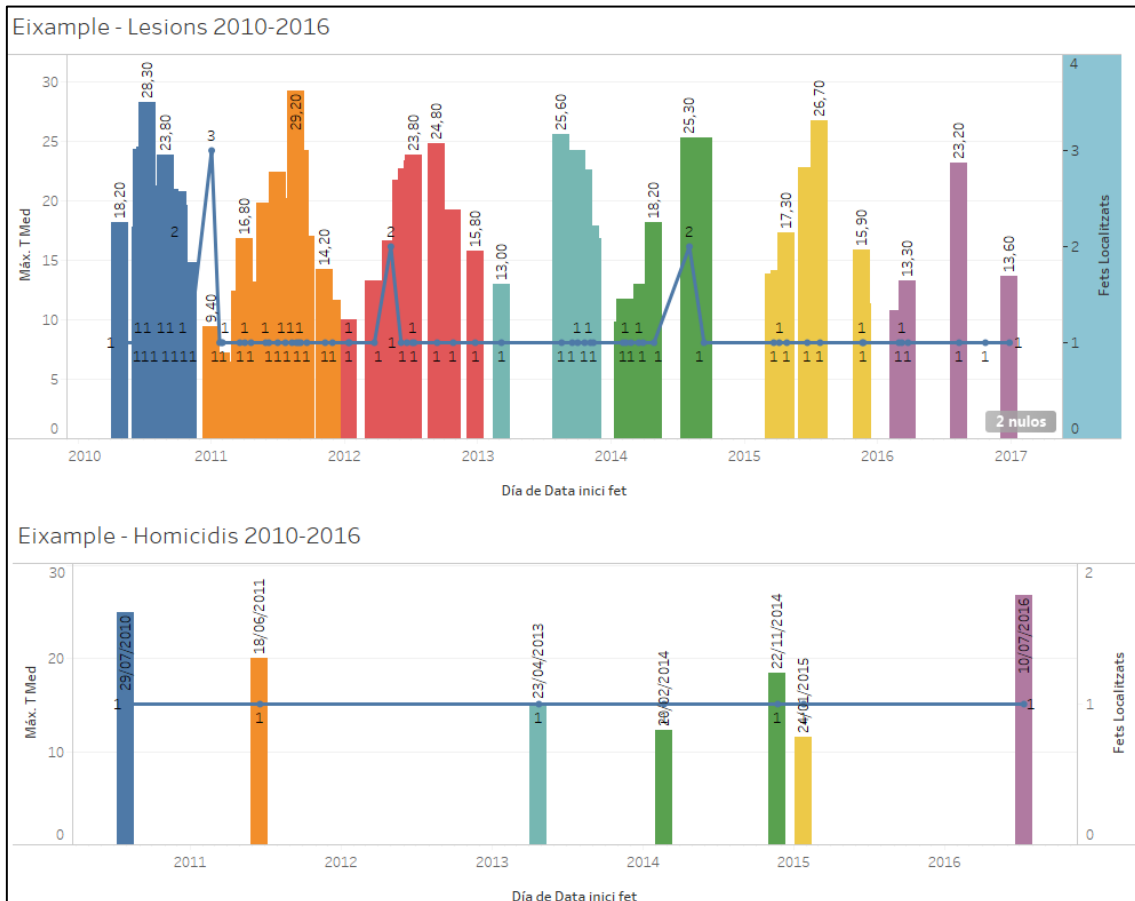


La següent pregunta que possiblement es planteji l'organització podria ser, **dins de l'àmbit de la violència de gènere, quins tipus de delictes es denuncien a la comissaria d'Eixample, a quina època de l'any i quina temperatura feia aquell dia.**

Al mateix *dashboard*, s'escull els gràfics que s'ha generat en el punt 4.1.8 i s'aplica el filtre de "Districte" i es selecciona "Eixample" i també el filtre de "Tipus de fet" un de "Lesions" i altre de "Homicidis"



El resultat és el següent:



Il·lustració 25 - Dashboard que recull els delictes de lesions i homicidis entre els anys 2010 i 2016 i la seva correlació amb les temperatures.

Al gràfic superior, que representa les lesions entre els anys 2010 i 2016 al districte d'Eixample, es dedueix que les barres que tenen forma de piràmides són les èpoques més caloroses de l'any, coincidint amb la temporada d'estiu i alguns dels fets estan agrupats al voltant de les piràmides, és a dir, que quan fa més calor, es podria concloure que hi ha més fets.

Per altra banda, s'observa que l'any 2010, 2011 i 2016 és van produir homicidis al districte d'Eixample on la temperatura era entre els 20° i 27°, no es reportar cap homicidi al districte d'Eixample durant l'any 2012 i entre el 2013 i 2015 els homicidis es van produir per sota de la franja de 20°.

5 Conclusions

Les primeres dificultats amb les què es va topar a l'hora de realitzar el projecte, va ser la recollida de dades oficials, ja que no acostumen a ser públiques, i en que cas ho siguin solen mostrar-se dades sintetitzades que no permeten un estudi en detall.

Per altra banda, l'objectiu inicial d'aquest projecte va ser estudiar l'existència d'una relació entre la climatologia i la violència exercida de persones, i així poder verificar la llei tèrmica de **Adolphe Quételet** (matemàtic, sociòleg, entre d'altres) que enunciacava que l'estiu és produïen més delictes contra les persones i a l'hivern és produïen més delictes contra el patrimoni.

La dificultat ha radicat en el número de fets, ja que en el nostre país, i fins tot en el nostre continent, no és produeixen tants delictes d'homicidis i assassinats com podria succeir en altres comunitats on els civils poden portar armes de foc o tenen un nivell cultural diferent, i possiblement l'índex de mortalitat per aquesta tipologia delictiva sigui més alta. És a dir, que el número de fets analitzats no és prou significativa com per poder confirmar de manera més evident la hipòtesi inicial.

No obstant això, s'han pogut extreure les següents conclusions:

- Els anys que mostren un patró més clar, són els anys 2010 i 2011, els quals confirmen la teoria què a més temperatura més delictes de lesions es cometem, coincidint amb l'estació de temps d'estiu.
- El **21 d'Agost de 2011** es van produir **5 fets** (el nombre de fets més alt d'aquest any), dia que es va registrar una temperatura mitjana de **29,20°** i **que va ser la més alta de tot l'any**.
- Existeixen altres factors, a més de la calor, que podrien induir a la violència, com ara que es tracti de dates assenyalades, per exemple l'1 de gener o setmana santa, tal i com ha quedat plasmat en els períodes 2011, 2013 i 2015.
- L'any 2013 entre el mes de març i maig, **es van cometre un total de 4 homicidis dolosos**, tots en grau de temptativa. La temperatura va oscil·lar entre els 13,80° i 19,20° però probablement, el detonador no va ser la calor sinó que coincidís amb setmana santa.
- La comissaria d'Eixample és la que té més volum de denúncies per delictes de violència de gènere i per tant, es podria abocar més efectius especialitzats en l'atenció de la víctima en aquesta comissaria, desplaçant efectius d'altres comissaries que tinguin menys requeriments com per exemple la comissaria del districte de Gràcia.

- Un fet a destacar que s'ha detectat durant l'anàlisi, encara que no sigui motiu d'estudi en aquest projecte, és que els **delictes de lesions des de l'any 2010 fins l'any 2016 han anat decreixent.**

6 **Bibliografia**

- Mariné Jové, Pere. "Components de la gestió de projectes: les àrees de coneixement" (2012). Universitat Oberta de Catalunya.
- Ley Orgánica 1/2015 (01/03/2017). "Reformas de la Ley Orgánica 1/2015. Código Penal". Capítulo XII. Disponible a http://noticias.juridicas.com/base_datos/Penal/549720-lo-1-2015-de-30-mar-modifica-la-lo-10-1995-de-23-nov-del-codigo-penal.html
- Db-Engines (2017). Data Bases Ranking. Recuperat de <https://db-engines.com/en/ranking>
- MongoDB (2007-2016). Mongo Data Base. Recuperat de <https://www.mongodb.com/es>
- Robomongo (). Robomongo 3T Client. Recuperat de <https://robomongo.org/>
- Simba Technologies (2016). Simba ODBC Driver. Recuperat de <http://www.simba.com/>
- Tableau Software (2016). Tableau. Recuperat de <https://www.tableau.com>
- Kevin Donnelly. "Adolphe Quetelet. Social Physics and the Average Men of Science, 1796-1874" (2015). Sci & Culture in the Nineteenth Century.