

# **Construcció i explotació d'un magatzem de dades per a l'anàlisi d'informació immobiliària**

---

## **Treball Fi de Carrera**

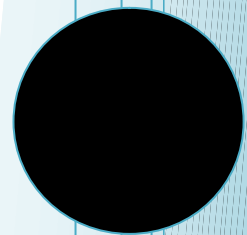
---

Silvia Hildebrandt (ETIG)

Consultor: José Ángel Martín Carballo

8 de juny de 2008

Universitat Oberta de Catalunya



## Resum

Per encàrrec d'una empresa immobiliària s'ha construït un magatzem de dades a partir d'un fitxer en format text amb els registres de 92.555 immobles amb la finalitat de, mitjançant consultes sobre aquestes dades utilitzant eines OLAP, obtenir informació sobre el mercat immobiliari.

Pel disseny conceptual de la base de dades s'han triat el Fet i les Dimensions, la granularitat amb la que es volen fer les anàlisis i s'han determinat els atributs, descriptors, jerarquies i mesures. Pel disseny lògic s'ha triat un model multidimensional en forma d'estrella amb l'oferta d'immobles com a taula de Fet i el temps, la zona i l'immoble com a taules de Dimensions. En el disseny físic s'ha establert l'estructura de les taules que conformen el magatzem de dades i els índex adequats per tal d'optimitzar els tems de resposta de les consultes.

La base de dades s'ha implementat en el sistema de gestió de base de dades Oracle 10g Express Edition i s'han portat a terme els processos de creació de taules, extracció, transformació i càrrega de dades utilitzant els llenguatges SQL i PL/SQL.

Per a l'obtenció de la informació s'ha fet servir l'eina d'explotació de dades Oracle Discoverer amb la que s'han construït l'àrea de negoci i el llibre de treball amb la finalitat de generar les consultes necessàries per cobrir les necessitats del client.

Amb les consultes generades s'ha pogut obtenir informació sobre els immobles que han baixat de preu, l'evolució dels preus de venda i d'oferta al llarg del temps, els preus i superfícies màxims, mínims i mitjans, els temps mitjans de venda i els immobles disponibles i venuts classificats per zona, tipologia i característiques. A més s'ha pogut determinar quines són les característiques tipus d'un immoble estàndard espanyol i andorrà.

## Paraules clau

Magatzem de dades, OLAP, anàlisi multidimensional, ETL, immobiliària

## Àrea del Treball Fi de Carrera

Magatzems de dades

## Índex de continguts

Resum .....	2
Paraules clau.....	2
Àrea del Treball Fi de Carrera.....	2
1  Capítol 1 - Introducció.....	7
1.1  Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC .....	7
1.2  Objectius del TFC.....	8
1.3  Enfocament i mètode seguit.....	8
1.4  Planificació del projecte .....	9
1.5  Productes obtinguts .....	15
1.6  Breu descripció dels altres capítols de la memòria.....	16
1.6.1  Capítol 2. Anàlisi:.....	16
1.6.2  Capítol 3. Disseny:.....	16
2  Capítol 2 - Anàlisi .....	17
2.1  Anàlisi prèvia de les dades proporcionades per l'empresa.....	17
2.2  Diagrama del model conceptual .....	17
2.2.1  Triar el Fet.....	17
2.2.2  Trobar el grànul escaient.....	18
2.2.3  Escollir les Dimensions que s'utilitzaran en l'anàlisi .....	18
2.2.4  Trobar els atributs de cada Dimensió.....	18
2.2.5  Distingir entre descriptors i jerarquies d'agregació.....	19
2.2.6  Decidir quines són les Mesures que interessin .....	19
2.2.7  Definir Cel·les.....	20
2.2.8  Explicitar les restriccions d'integritat .....	20
2.2.9  Estudiar la viabilitat.....	21
3  Capítol 3 - Disseny.....	23
3.1  Diagrama de l'arquitectura programari .....	23
3.2  Diagrama de l'arquitectura maquinari .....	23
3.3  Disseny de la base de dades i diagrama del model físic.....	24
3.3.1  Disseny lògic.....	24
3.3.2  El disseny físic.....	25

3.4	El procés d'extracció, transformació i càrrega de dades .....	28
3.4.1	Extracció.....	28
3.4.2	Transformació.....	28
3.4.3	Càrrega.....	30
3.5	Disseny dels informes creats.....	33
3.5.1	La creació de l'àrea de negoci .....	33
3.5.2	La creació del llibre de treball .....	35
3.6	Descripció de les consultes i informes.....	37
3.6.1	Full 1: Llistat baixada preus inicial .....	37
3.6.2	Full 2: Llistat baixada preus anterior.....	38
3.6.3	Full 3: Preus.....	39
3.6.4	Full 4: Evolució preus d'oferta.....	40
3.6.5	Full 5: Evolució preus de venda.....	41
3.6.6	Full 6: Metres quadrats.....	42
3.6.7	Full 7: Temps de venda.....	43
3.6.8	Full 8: Immobles disponibles.....	44
3.6.9	Full 9: Immobles venuts.....	45
3.6.10	L'immoble estàndard espanyol i andorrà.....	46
4	Conclusions.....	54
5	Línies d'evolució futures.....	55
6	Glossari .....	56
7	Bibliografia consultada.....	58
8	Annexos.....	59
8.1	Annex I- Instruccions per la restauració del sistema.....	59

## Índex de figures

Figura 1. Diagrama de Gantt amb la temporització del projecte proposat.....	14
Figura 2. Esquema UML del disseny conceptual del model multidimensional per al magatzem de dades de l'empresa Sostre per a tothom.....	22
Figura 3. Diagrama de l'arquitectura programari.....	23
Figura 4. Diagrama de l'arquitectura maquinari.....	24
Figura 5. Esquema relacional en forma d'estrella del model multidimensional per al magatzem de dades de l'empresa Sostre per a tothom.....	25
Figura 6. Disseny físic de la base de dades. Es mostra a més el disseny de la taula temporal CATALEG.....	27
Figura 7. Esquema de càrrega de dades a les taules COMARCA, AREA, SUBAREA, IMMOBLE i OFERTA a partir de les dades de la taula CATALEG.....	33
Figura 8. Dades de l'àrea de negoci TFC-Sostre per a tothom.....	34
Figura 9. Jerarquies de l'àrea de negoci TFC-Sostre per a tothom.....	35
Figura 10. Classes d'elements de l'àrea de negoci TFC-Sostre per a tothom.....	35
Figura 11. Elements i disseny de taula del full Llistat baixada preus inicial.....	37
Figura 12. Full 1: Llistat baixada preus inicial.....	38
Figura 13. Elements i disseny de taula del full Llistat baixada preus anterior.....	38
Figura 14. Full 2: Llistat baixada preus anterior.....	39
Figura 15. Elements i disseny de taula del full Preus.....	39
Figura 16. Full 3: Preus.....	40
Figura 17. Elements i disseny de taula del full Evolució preus d'oferta.....	40
Figura 18. Full 4: Evolució preus d'oferta.....	41
Figura 19. Elements i disseny de taula del full Evolució preus de venda.....	41
Figura 20. Full 5: Evolució preus de venda.....	42
Figura 21. Elements i disseny de taula del full Metres quadrats.....	42
Figura 22. Full 6: Metres quadrats.....	43
Figura 23. Elements i disseny de taula del full Temps de venda.....	43
Figura 24. Full 6: Temps de venda.....	44
Figura 25. Elements i disseny de taula del full Immobles disponibles.....	44
Figura 26. Full 8: Immobles disponibles.....	45
Figura 27. Elements i disseny de taula del full Immobles venuts.....	45
Figura 28. Full 9: Immobles venuts.....	46
Figura 29. Immobles espanyols classificats segons la tipologia.....	48
Figura 30. Immobles espanyols classificats segons la planta.....	48
Figura 31. Immobles espanyols classificats segons el nombre de dormitoris.....	49
Figura 32. Immobles espanyols classificats segons els metres quadrats.....	49
Figura 33. Immobles espanyols classificats segons els metres quadrats (visió ampliada).....	50
Figura 34. Immobles espanyols classificats segons el preu.....	50
Figura 35. Immobles andorrans classificats segons la tipologia.....	51
Figura 36. Immobles andorrans classificats segons la planta.....	51
Figura 37. Immobles andorrans classificats segons el nombre de dormitoris.....	52

Figura 38. Immobles andorrans classificats segons els metres quadrats.....	52
Figura 39. Immobles andorrans classificats segons el preu.....	53

## 1 Capítol 1 - Introducció

### 1.1 Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC

Actualment, gràcies a l'abaratiment i a l'increment de prestacions dels equips informàtics, les empreses tenen la capacitat de generar i emmagatzemar grans volums de dades que generalment es troben en sistemes d'informació transaccionals que s'actualitzen contínuament. Però disposar d'aquestes dades no significa necessàriament disposar d'informació. Perquè les dades aportin informació i per tant coneixement per tal de millorar la productivitat, conèixer el mercat i els clients, prendre decisions amb més rapidesa, etc., en definitiva, augmentar la competitivitat, és imprescindible que tinguin una estructura organitzada, integrada, lògica, dinàmica i de fàcil explotació. Això s'aconsegueix amb els magatzems de dades.

A un magatzem de dades es diposita de forma consistent informació procedent de diferents sectors o departaments d'una organització i de forma que les dades relatives al mateix esdeveniment del món real es trobin unides entre si. La informació és no volàtil, no es modifica ni s'elimina i a més es registren els canvis que es produeixen en les dades al llarg del temps (Inmon, 2002). D'altra banda aquestes dades es troben estructurades per poder ser consultades i analitzades (Kimball, 2002). D'una manera més amplia, un magatzem de dades inclou, a més, els mitjans per obtenir, analitzar, extreure, transformar i carregar les dades i la gestió de les mateixes.

Una tècnica específica utilitzada en Intel·ligència Empresarial (*Business Intelligence - BI*) per a l'extracció de coneixement a partir de grans volums de dades és el processament analític en línia (*On-Line Analytical Processing - OLAP*). Mitjançant l'OLAP es poden executar consultes multidimensionals i complexes sobre les dades d'una empresa que permeten a l'usuari una visió més ràpida i interactiva d'aquestes.

En aquest TFC es tracta de construir i explotar un magatzem de dades per a l'empresa immobiliària "Sostre per a tothom" que porta uns anys enregistrant en fulls de càlcul la informació referent als immobles que ofereixen arreu de Catalunya, Comunitat Valenciana, Illes Balears i Andorra. Ha arribat el moment en que ha decidit fer una anàlisi més exhaustiva d'aquesta informació per tal d'ajustar l'oferta i la demanda i poder obtenir un avantatge competitiu respecte altres empreses del sector. Per aquesta raó ha encarregat la construcció

d'un magatzem de dades i l'elaboració d'una sèrie d'informes que recullin les dades d'interès com ara el nombre d'immobles disponibles i venuts, els preus i superfícies màxims, mínims i mitjans, el temps mitjà de venda i un llistat d'immobles que hagin baixat de preu. Totes aquestes dades han de ser presentades per zona, tipologia i característiques de l'immoble. A part d'això, l'empresa desitja conèixer quin és l'immoble estàndard espanyol i andorrà.

## 1.2 Objectius del TFC

L'objectiu principal del projecte és adquirir experiència en el disseny, construcció i explotació d'un magatzem de dades a partir de la informació disponible en una base de dades transaccional.

Paral·lelament, els objectius secundaris que es pretenen assolir amb el TFC són:

1. Familiaritzar-se amb l'arquitectura general d'un magatzem de dades.
2. Introduir-se en els aspectes teòrics del disseny d'un magatzem de dades
3. Adquirir experiència en llenguatges de consultes de bases de dades.
4. Conèixer les tècniques de tractament de dades i la seva integració en el model de dades físic del magatzem de dades
5. Analitzar i comparar les eines de mercat que treballen amb aquests models
6. Realitzar les anàlisis i informes dirigits a usuaris finals del magatzem de dades

## 1.3 Enfocament i mètode seguit

Per desenvolupar el TFC s'han seguit, encara que de manera simplificada donat l'abast més reduït d'aquest treball, les fases típiques de les quals consta un projecte de magatzems de dades<sup>1</sup>:

1. **Recopilació de requeriments:** Per conèixer els requeriments del client generalment es celebren reunions en les que estan presents tots els departaments involucrats. En aquest cas, els requeriments vénen donats a l'enunciat del TFC.
2. **Instal·lació de l'entorn físic:** S'han d'instal·lar els servidors i les bases de dades. Generalment és convenient establir un entorn de desenvolupament i un altre de producció. Inclús és recomanable un tercer entorn per fer les proves. En aquest TFC aquesta fase ha consistit en posar a punt el maquinari i el programari utilitzats i s'ha fet tot en un mateix entorn.
3. **Modelització de dades:** Aquesta es realitza a tres nivells, conceptual, lògic i físic. Els dos primers poden ser unificats en un sol procés en el que s'identifiquen les entitats amb les

---

<sup>1</sup> Data Warehousing Processes. <http://www.1keydata.com/datawarehousing/processes.html>. Data de consulta: 15/03/08



seves claus primàries i atributs, les relacions entre elles i es fa la normalització. Es fa una descripció exhaustiva de les dades sense implementar-les físicament a la base de dades. En la modelització física es converteixen les entitats en taules, els atributs en columnes i les relacions en claus foranes.

4. **ETL:** És la fase d'extracció, transformació i càrrega de dades.
5. **Disseny del cub OLAP:** A partir del coneixement de requeriments s'identifiquen els elements necessaris per realitzar les consultes que satisfacin els requeriments del client.
6. **Desenvolupament de les eines de consulta per l'usuari:** Aquestes eines poden ser comercials o dissenyades a mesura però el més important és que siguin amigables i adaptables a l'empresa client. Generalment s'utilitzen eines que generin informes a través de la xarxa de manera que no sigui necessari instal·lar programari específic a l'ordinador del client. En aquest cas s'ha utilitzat el programari Oracle BI Discoverer.
7. **Optimització del rendiment:** És important que tant el temps d'ETL com el processament de consultes i la generació d'informes estigui optimitzat. Això s'ha aconseguit amb un disseny adequat de l'algoritme de càrrega, la creació d'índex per accelerar les consultes i la tria dels elements que conformen aquestes consultes.
8. **Garantia de qualitat:** Les comprovacions de qualitat generalment les fa el client. En aquest projecte s'han fet les proves necessàries per assegurar el bon funcionament del magatzem de dades.
9. **Entrada en funcionament del producte:** Aquesta fase generalment va associada a un curs d'entrenament del client i està fora de l'abast del TFC.
10. **Manteniment del producte:** Aquesta fase està fora de l'abast del TFC.
11. **Actualitzacions del producte:** Aquesta fase està fora de l'abast del TFC.

## 1.4 Planificació del projecte

El projecte s'ha estructurat en 6 grans blocs la finalització dels quals coincideix, excepte en el blocs 1 i 6, amb el lliurament dels documents i aplicacions generats:

- 1) **INICIACIÓ:** En aquesta fase d'iniciació s'han preparat els materials necessaris pel desenvolupament del TFC
  - a) **Descarregar i organitzar materials de l'assignatura:** Una vegada obertes les aules s'han descarregat tots els materials (pla docent, materials complementaris, etc.)

corresponents a l'assignatura i es s'han creat carpetes al disc dur per tal d'organitzar-los.

- b) **Llegir materials:** S'ha un estudi exhaustiu dels materials disponibles.
- c) **Descarregar i llegir enunciat TFC:** Una vegada publicat l'enunciat del TFC i altres materials s'han descarregat i estudiat.
- d) **Cercar bibliografia:** S'ha fet una cerca de la bibliografia recomanada i d'altres fonts.
- e) **Instal·lar i provar programari:** S'ha instal·lat el programari necessari pel desenvolupament del projecte i es s'han fet proves de funcionament per tal de familiaritzar-se amb el programari nou.

**2) PAC 1 – PLA DE TREBALL I ANÀLISI PRELIMINAR DE REQUERIMENTS:** En aquesta fase s'ha redactat aquest pla de treball on s'indica la planificació estimada de les diferents tasques a realitzar per dur a terme el projecte i un document d'anàlisi preliminar de requeriments que inclou l'enumeració i breu descripció dels elements d'anàlisi identificats (dimensions, atributs, indicadors, etc.) que estan disponibles per als usuaris i el nombre i contingut dels informes que s'han d'implementar. També s'han analitzat les fonts de dades operacionals proporcionades.

- a) **Redacció esborrany Pla de treball:** S'ha redactat una primera versió del pla de treball per tal de lliurar-la abans de la data de lliurament definitiva. D'aquesta manera s'ha pogut incloure en el document final les correccions i recomanacions del consultor.
- b) **Lliurament esborrany Pla de treball**
- c) **Redacció esborrany Anàlisi preliminar:** S'ha redactat una primera versió de l'anàlisi preliminar de requeriments per tal de lliurar-la abans de la data de lliurament definitiva. D'aquesta manera s'ha pogut incloure en el document final les correccions i recomanacions del consultor.
- d) **Lliurament esborrany Anàlisi preliminar**
- e) **Redacció documents definitius:** S'han completat i corregit els esborranys per elaborar els documents definitius.
- f) **Lliurament PAC 1**

**3) PAC 2 – ANÀLISI DE REQUERIMENTS I DISSENY CONCEPTUAL I TÈCNIC:** En aquesta fase s'ha redactat un document amb l'anàlisi detallat de requeriments basat en l'anàlisi preliminar realitzat. A més s'ha elaborat un document de disseny amb la descripció del model dimensional que dona suport a les necessitats dels usuaris, segons l'anàlisi realitzat i el disseny dels procediments d'extracció de dades a alt nivell (processos, pseudocodi, etc.)

- a) **Cerca i estudi de bibliografia:** S'ha fet una cerca i estudi més exhaustius de recursos útils per l'elaboració de les fases 3 i 4.
  - b) **Redacció esborrany Anàlisi requeriments:** Hi havia programada la redacció d'un esborrany de l'anàlisi de requeriments però donat que l'anàlisi preliminar va ser correcte només es van fer algunes puntualitzacions respecte al primer document i no va ser necessari fer un document nou.
  - c) **Lliurament esborrany Anàlisi requeriments:** No es va fer
  - d) **Redacció esborrany Disseny conceptual i tècnic:** Es va redactar una primera versió del disseny conceptual i tècnic per tal de lliurar-la abans de la data de lliurament definitiva. D'aquesta manera s'ha pogut incloure en el document final les correccions i recomanacions del consultor.
  - e) **Lliurament esborrany Disseny conceptual i tècnic**
  - f) **Redacció documents definitius:** S'ha completat i corregit l'esborrany per elaborar el document definitiu.
  - g) **Lliurament PAC 2**
- 4) PAC 3 – IMPLEMENTACIÓ:** En aquesta fase s'ha construït el magatzem de dades, s'han creat les taules, s'ha fet el procés d'extracció, transformació i càrrega de dades i s'han generat els informes que aporten la informació demanada pel client.
- a) **Construcció del magatzem de dades:** S'ha construït la base de dades Oracle i s'han carregat les dades.
  - b) **Familiarització amb l'eina d'explotació de dades:** S'han fet proves de generació de consultes i informes amb Oracle BI Discoverer per tal de familiaritzar-se amb el seu funcionament.
  - c) **Construcció dels informes i anàlisi de la informació:** S'ha construït l'àrea de negoci i el llibre de treball amb les consultes i informes mitjançant l'eina d'anàlisi i consulta d'informes Oracle BI Discoverer.
  - d) **Lliurament PAC 3 provisional.** Hi havia planificat un lliurament previ de la base de dades i dels informes però per falta de temps no es va fer.
  - e) **Elaboració definitiva PAC 3:** S'ha exportat la base de dades i s'ha redactat el document que acompanyava als fitxers generats.
  - f) **Lliurament PAC 3**
- 5) LLIURAMENT FINAL:** En aquesta fase s'ha elaborat la memòria final del TFC i la presentació virtual que es lliuren al tribunal del TFC.

- a) **Elaboració esborrany memòria:** S'ha redactat la memòria que recull tota la informació rellevant sobre el problema plantejat en el TFC, la metodologia emprada per la seva resolució i els resultats obtinguts.
  - b) **Elaboració versió preliminar presentació:** S'ha realitzat una presentació virtual de 30 diapositives que sintetitza el treball realitzat durant el desenvolupament del projecte i els resultats obtinguts.
  - c) **Lliurament documents provisionals:** Hi havia programat el lliurament provisional dels dos documents però només s'ha lliurat una primera versió de la presentació. S'ha considerat que amb l'experiència adquirida durant el semestre en la redacció d'informes i amb les indicacions proporcionades pel consultor sobre el contingut de la memòria no era necessari lliurar una versió preliminar.
  - d) **Elaboració documents definitius:** S'ha completat i corregit l'esborrany de la presentació amb les recomanacions del consultor.
  - e) **Lliurament documents definitius**
- 6) DEBAT VIRTUAL:** En aquesta fase es participarà al debat virtual que es produirà al final del semestre.
- a) **Inici debat**
  - b) **Lectura d'intervencions i participació al debat:** A mesura que es produeixin intervencions al debat virtual es farà una anàlisi de les mateixes i es respondran o faran aportacions noves sempre que es consideri oportú.
  - c) **Fi debat**

Com a dates clau s'han pres les publicades al pla docent i són les que s'especifiquen a continuació:

Inici semestre / inici projecte	28/02/08
Lliurament PAC 1 (Pla de treball i anàlisi preliminar de requeriments)	16/03/08
Lliurament PAC 2 (Anàlisi de requeriments i disseny conceptual i tècnic)	20/04/08
Lliurament PAC 3 (Implementació)	18/05/08
Lliurament final (Memòria i presentació TFC)	08/06/08
Inici debat virtual	18/06/08
Fi debat virtual / fi projecte	26/06/08

En la següent taula es mostren les dates d'inici i finalització estimades per a cadascuna de les tasques i fites descrites anteriorment. Aquest és el calendari que es va dissenyar a l'inici del semestre. En general les dates s'han complert i només s'han produït retards d'alguns dies en les dates de lliurament d'alguns esborrany. On potser s'han produït més desviaments respecte a les dates estimades és a la PAC 3. Les tasques 4.2 i 4.3 van requerir més temps de

l'esperat el que va impossibilitar presentar una versió provisional de la PAC per a la seva revisió.

Tasca / Fita	Data inici	Data fi
TFC - Magatzems de dades	<b>dij 28/02/08</b>	dij 26/06/08
<b>1. Iniciació</b>	<b>dij 28/02/08</b>	<b>dij 06/03/08</b>
1.1 Descarregar i organitzar materials de l'assignatura	dij 28/02/08	dij 28/02/08
1.2 Llegir materials	div 29/02/08	div 29/02/08
1.3 Descarregar i llegir enunciat TFC	dis 01/03/08	dis 01/03/08
1.4 Cercar bibliografia	dill 03/03/08	dima 04/03/08
1.5 Instal·lar i provar programari	dime 05/03/08	dij 06/03/08
<b>2. PAC 1</b>	<b>div 07/03/08</b>	<b>dium 16/03/08</b>
2.1 Redacció esborrany Pla de treball	div 07/03/08	dima 11/03/08
2.2 Lliurament esborrany Pla de treball	dime 12/03/08	dime 12/03/08
2.3 Redacció esborrany Anàlisi preliminar	div 07/03/08	dill 10/03/08
2.4 Lliurament esborrany Anàlisi preliminar	dima 11/03/08	dima 11/03/08
2.5 Redacció documents definitius	dime 12/03/08	dium 16/03/08
2.6 Lliurament PAC 1	dium 16/03/08	dium 16/03/08
<b>3. PAC 2</b>	<b>dill 17/03/08</b>	<b>dium 20/04/08</b>
3.1 Cerca i estudi de bibliografia	dill 17/03/08	dill 24/03/08
3.2 Redacció esborrany Anàlisi requeriments	dima 25/03/08	dill 07/04/08
3.3 Lliurament esborrany Anàlisi requeriments	dima 08/04/08	dima 08/04/08
3.4 Redacció esborrany Disseny conceptual i tècnic	div 28/03/08	dij 10/04/08
3.5 Lliurament esborrany Disseny conceptual i tècnic	div 11/04/08	div 11/04/08
3.6 Redacció documents definitius	dill 14/04/08	div 18/04/08
3.7 Lliurament PAC 2	dium 20/04/08	dium 20/04/08
<b>4. PAC 3</b>	<b>dill 21/04/08</b>	<b>dium 18/05/08</b>
4.1 Construcció del magatzem de dades	dill 21/04/08	dill 28/04/08
4.2 Familiarització amb l'eina d'explotació de dades	dima 29/04/08	dima 29/04/08
4.3 Construcció dels informes i anàlisi de la informació	dime 30/04/08	dij 08/05/08
4.4 Lliurament PAC 3 provisional	div 09/05/08	div 09/05/08
4.5 Elaboració definitiva PAC 3	dill 12/05/08	div 16/05/08
4.6 Lliurament PAC 3	dium 18/05/08	dium 18/05/08
<b>5. Lliurament final</b>	<b>dill 19/05/08</b>	<b>dium 08/06/08</b>
5.1 Elaboració esborrany memòria	dill 19/05/08	dij 29/05/08
5.2 Elaboració versió preliminar presentació	dill 26/05/08	dime 28/05/08
5.3 Lliurament documents provisionals	dij 29/05/08	dij 29/05/08
5.4 Elaboració documents definitius	dill 02/06/08	div 06/06/08
5.5 Lliurament documents definitius	dium 08/06/08	dium 08/06/08
<b>6. Debat virtual</b>	<b>dime 18/06/08</b>	<b>dij 26/06/08</b>
6.1 Inici debat	dime 18/06/08	dime 18/06/08
6.2 Lectura d'intervencions i participació al debat	dime 18/06/08	dij 26/06/08
6.3 Fi debat	dij 26/06/08	dij 26/06/08

A continuació es representa el diagrama de Gantt on es mostra el temps de dedicació previst per a les tasques o activitats que conformen el projecte. Es tracta del diagrama presentat a l'inici del semestre. Com ja s'ha esmentat anteriorment, en general la temporització s'ha complert i només s'han produït lleugers desviaments de les dates però que no han impedit el lliurament puntual de les PAC.

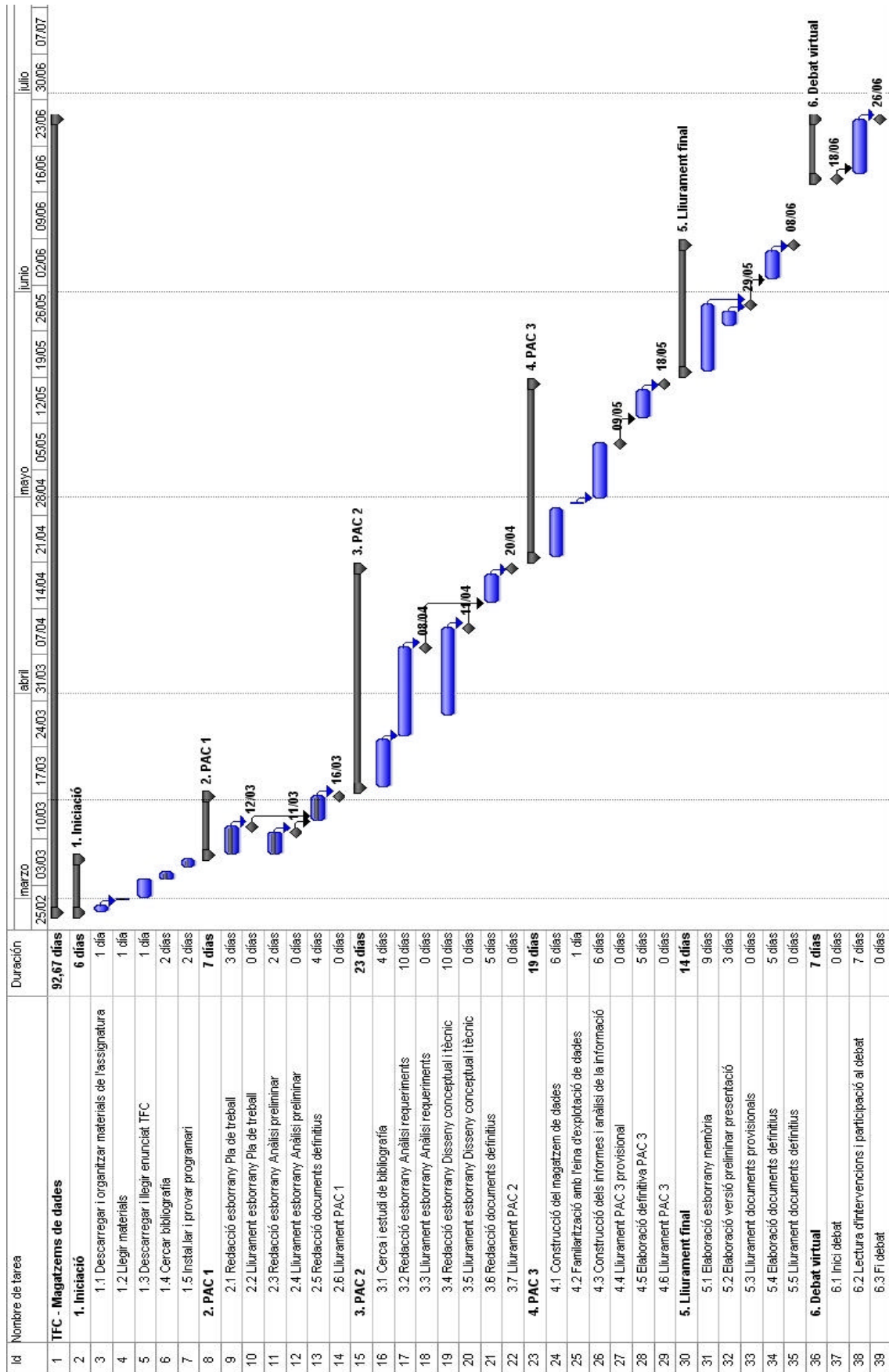


Figura 1. Diagrama de Gantt amb la temporització del projecte proposat

## 1.5 Productes obtinguts

Al llarg del projecte s'han elaborat els següents documents i productes:

- ★ **Pla de treball:** Descripció general del projecte i els seus objectius, metodologia a seguir per desenvolupar el projecte i la seva planificació temporal, recursos de maquinari, programari i bibliogràfics disponibles, possibles incidències que es puguin produir al llarg del projecte i les solucions i precaucions preses.
- ★ **Anàlisi preliminar de requeriments:** Descripció i anàlisi prèvia de les dades proporcionades, primera identificació dels elements d'anàlisi i descripció preliminar dels informes que es lliuraran.
- ★ **Anàlisi de requeriments:** Anàlisi de les dades, identificació dels elements d'anàlisi definitius i descripció dels informes que es lliuraran al client.
- ★ **Disseny:** Disseny conceptual, lògic i físic del model multidimensional emprat per desenvolupar el magatzem de dades, descripció del procés d'extracció, transformació i càrrega de dades (ETL) i primera aproximació de la metodologia de construcció de consultes.
- ★ **Implementació:** Construcció de la base de dades i elaboració dels informes.
- ★ **Memòria final:** L'actual document. Descripció detallada del projecte, els seus objectius, la seva planificació, la metodologia de desenvolupament emprada, els resultats obtinguts i les conclusions.
- ★ **Presentació:** Presentació gràfica del projecte que dona una visió global i resumida de la feina desenvolupada.
- ★ **Producte:** Carpeta que conté tots els elements necessaris per a l'execució del projecte.
  - ★ **crear\_tables.sql:** *script* SQL per a la creació de les taules del magatzem de dades. Per a més informació vegeu l'apartat 3.3.2.
  - ★ **crear\_index.sql:** *script* SQL per a la creació dels índex utilitzats per fer les consultes al magatzem de dades. Per a més informació vegeu l'apartat 3.3.2.
  - ★ **cataleg.ctf:** fitxer amb les instruccions per a la càrrega de la taula CATALEG. S'utilitza juntament amb el fitxer de text proporcionat per la immobiliària que ha de portar el nom `cataleg.dat`. Per més informació vegeu l'apartat 3.3.2.
  - ★ **cataleg.dat:** catàleg d'immobles proporcionat per l'empresa. S'utilitza, juntament amb el fitxer `cataleg.ctf`, per carregar la taula CATALEG.

- ★ **correccio\_errors.sql**: *script* SQL de correcció d'errors detectats a les dades del fitxer cataleg.dat. S'executa sobre les dades carregades a la taula CATALEG. Per més informació vegeu l'apartat 3.4.2.
- ★ **carrega\_dades.pls**: *script* PL/SQL de carrega de dades al magatzem de dades. Per més informació vegeu l'apartat 3.4.3.
- ★ **bd\_silvia.dmp**: base de dades exportada. Per més informació vegeu els apartats 3.4.3 i 8.1.
- ★ **TFC-Sostre per a tothom.eex**: espai de negoci. Per més informació vegeu els apartats 3.5.1 i 8.1.
- ★ **Sostre per a tothom - workbook.DIS**: llibre de treball. Per més informació vegeu els apartats 3.5.2 i 8.1.

## 1.6 Breu descripció dels altres capítols de la memòria

### 1.6.1 Capítol 2. Anàlisi:

En aquest capítol es fa una primera descripció del fitxer de dades proporcionat per l'empresa. A partir de la informació continguda en aquest fitxer es realitza el disseny conceptual del model multidimensional on es tria el Fet que es vol analitzar, les Dimensions o punts de vista des d'els que s'analitzen el Fet, la granularitat i es defineixen els atributs, descriptors, jerarquies i mesures. Finalment es fa un estudi de viabilitat del projecte.

### 1.6.2 Capítol 3. Disseny:

A partir del model conceptual es fa un disseny lògic on els Fets i les Dimensions s'estructuren en taules i es defineixen les seves claus primàries i foranes, que relaciones unes taules amb altres. Finalment es fa el disseny físic, on es decideix com estructurar la informació en les taules per tal de obtenir un bon temps de resposta a les consultes, tenint en compte el volum de dades, l'índex de selectivitat dels atributs o el tipus de consultes més freqüents.

A part del disseny de la base de dades, també es descriu la seva implementació, que consta de la creació de taules i l'extracció, transformació i càrrega de dades.

Finalment es fa una descripció del disseny de les consultes i informes que donen resposta als requeriments plantejats per l'empresa client i es mostren els resultats obtinguts.



## 2 Capítol 2 - Anàlisi

### 2.1 Anàlisi prèvia de les dades proporcionades per l'empresa

El catàleg d'immobles proporcionat per l'empresa es troba en format text i una vegada carregat en una aplicació de gestió de bases de dades com és Microsoft Access s'observa que conté la següent informació:

El catàleg consta de 93.360 registres. No tots els registres corresponen a immobles sinó que alguns, en concret 805, són capçaleres que separen els immobles de les diferents zones. Així doncs, l'empresa disposa d'un total de 92.555 immobles en 3 comunitats autònomes espanyoles (Catalunya, Illes Balears i Comunitat Valenciana) i Andorra.

Per a cada immoble es disposa d'una sèrie de dades com ara:

- ★ Un codi alfanumèric que identifica l'immoble i que és únic
- ★ Un codi alfanumèric que indica la zona on es troba l'immoble (província, comarca, àrea i subàrea).
- ★ La tipologia de l'immoble: si es tracta d'un àtic, un xalet, un dúplex, un estudi o un pis.
- ★ La planta en la que es troba l'immoble.
- ★ El nombre de dormitoris dels que disposa l'immoble.
- ★ Els metres quadrats de l'immoble.
- ★ El preu de venda des del mes en el que es posa a la venda l'immoble fins el mes en el que s'ha venut. El període abastat va des de gener de 2006 fins gener de 2008.

### 2.2 Diagrama del model conceptual

El disseny conceptual d'un model multidimensional consisteix en l'execució iterativa de nou passos. A continuació es presenta una anàlisi exhaustiva d'aquests nou passos aplicats al projecte de la immobiliària *Sostre per a tothom*.

#### 2.2.1 Triar el Fet

Un Fet és un conjunt d'esdeveniments amb dades numèriques associades. En aquest cas a l'empresa l'interessa obtenir informació de dos fets: l'oferta i la venda d'immobles.

En realitat es poden considerar les vendes com un cas concret de les ofertes i ambdues es poden distingir mitjançant un atribut booleà que indiqui si l'immoble corresponent a una oferta ha estat venut o no. Per tant el model contindrà un sol Fet: l'oferta d'immobles.

### **2.2.2 Trobar el grànul escaient**

Un bon disseny demana triar sempre el grànul més petit possible però sense arribar a malbaratar espai o fer inviable el projecte per excés de dades. Si es tria un grànul petit sempre hi ha la possibilitat de calcular dades derivades.

En aquest cas concret la mida de la base de dades no és excessivament gran per tant es pot triar un grànul petit.

Per a la Dimensió temps es tria la granularitat Mes ja que una de les informacions que es demana és el temps mitjà de venda i expressar-lo en l'escala temporal Any seria poc precís.

L'altre Dimensió que presenta Nivells és la zona. També en aquest cas es tria la granularitat més petita, és a dir a nivell de subàrea i es deixarà a l'usuari la decisió de si vol consultar les dades amb una granularitat més gran. D'aquesta manera es podran consultar els immobles oferts en el nivell de zona geogràfica més petit i agrupar-los per conèixer l'oferta a nivell de Comunitat Autònoma o País, per exemple.

Respecte al producte ofert, la granularitat més petita és cada immoble individual. Aquests poden ser agrupats per tipus d'immoble: xalets, àtics, estudis, etc. El més adequat és mantenir la granularitat d'immoble individual per poder fer consultes concretes, com per exemple el llistat de tots el immobles oferts o venuts en un determinat mes, i deixar la possibilitat d'agrupar-los per obtenir dades per tipus d'immoble si l'usuari ho requereix.

### **2.2.3 Escollir les Dimensions que s'utilitzaran en l'anàlisi**

Una vegada definida la granularitat ja es disposa el conjunt de Dimensions que, com s'ha esmentat anteriorment, són: el temps, la zona i el producte que és l'immoble.

Si una combinació d'instàncies d'aquestes Dimensions inicials determinessin la instància d'una altra Dimensió també s'hauria d'afegir al conjunt però no és el cas.

### **2.2.4 Trobar els atributs de cada Dimensió**

Els atributs serveixen per fer agrupacions i seleccions de les dades. Un atribut ha de tenir un domini discret, no ha d'estar abreujat ni codificat i ha de ser fàcilment intel·ligible per a l'usuari.

Per a la Dimensió Zona la única informació de la que es disposa de cada àrea geogràfica és el seu nom per tant aquest serà l'únic atribut a cadascun dels nivells. Si hi haguessin més dades, com per exemple el nombre d'habitants, s'afegirien com atributs.

Per a la Dimensió Temps, la situació és similar. Només es disposa del nom dels mesos i del número de l'any, sense cap més informació.

Per a la Dimensió Immoble sí que hi ha més dades. Cada immoble té un nombre concret de dormitoris i metres quadrats i es troba en una planta determinada per tant aquests seran els seus atributs. També el codi d'identificació de cada immoble és un atribut. Per simplificar les consultes es poden definir rangs i reduir el nombre de valors dels atributs que defineixen les característiques del immobles. Per exemple, per l'atribut dormitoris es poden reduir els valors a: d'un a quatre, de quatre a deu, més de deu dormitoris. Aquests nous atributs serien derivats dels anteriors. No obstant, aquesta divisió per rangs només seria viable si l'empresa estableix prèviament quin són els rangs que li poden resultar més profitosos per les seves consultes. Pel cas del Nivell Tipus només es disposa del nom del Tipus com a atribut.

### **2.2.5 Distingir entre descriptors i jerarquies d'agregació**

Els atributs d'una Dimensió poden ser classificats en:

- ★ Descriptors: serveixen per seleccionar
- ★ Els que defineixen jerarquies d'agregació: serveixen per agrupar

Segons els atributs definits en l'apartat anterior i la granularitat determinada en l'apartat 2.2.2 els atributs que definiran un Nivell dins una jerarquia d'agregació seran: subàrea, àrea, comarca, província, comunitat autònoma i país per a la Dimensió Zona; mes i any per a la Dimensió Temps i Immoble i Tipus per a la Dimensió Immoble. La resta d'atributs, com ara els noms, el codi d'immoble, el nombre de dormitoris i metres quadrats o la planta en la que es troba l'immoble són els descriptors i es col·loquen al Nivell que els pertoca.

### **2.2.6 Decidir quines són les Mesures que interessen**

Utilitzant com a base la informació que demana l'empresa es poden definir les Mesures que interessa analitzar. Per una banda interessen les Mesures "quantitat d'immobles venuts" i "quantitat d'immobles oferts". Ambdues quantitats es poden distingir mitjançant un atribut booleà que indica si l'immoble ofert ja s'ha venut o no. D'altra banda interessa conèixer el temps (mesurat en mesos) que un immoble ha estat en oferta per poder calcular els temps mitjans de venda. A més s'han d'enregistrar els preus amb els que s'ofereixen (i venen) els immobles per calcular els preus màxims, mínims i mitjans. Finalment, l'empresa també demana saber l'evolució dels preus dels immobles i vol fer un llistat dels immobles que han baixat de preu. Amb aquest fi, es creen dues noves mesures, una que indica la variació del preu en un determinat mes respecte al seu preu d'oferta inicial i una altra que indica la

variació del preu respecte al mes anterior. Aquestes mesures tindran valor null el primer mes d'oferta, 0 si el preu no ha variat, un valor positiu si el preu s'ha incrementat i un valor negatiu si pel contrari el preu ha baixat.

### 2.2.7 Definir Cel·les

L'única cel·la que conté el Fet és el de l'Oferta Atòmica. Es tracta de l'oferta d'un immoble amb unes determinades característiques, ofert en un determinat mes i en una subàrea determinada.

### 2.2.8 Explicitar les restriccions d'integritat

Les restriccions d'integritat inherents al model són:

- 1) **Unicitat i entitat de la Base:** La Base indica quines Dimensions identifiquen les Cel·les. Les restriccions que han de complir una Base són:
  - a) Els Nivells que formen una Base han de ser funcionalment independents.
  - b) Les associacions dels Nivells amb la Cel·la han de tenir una multiplicitat mínima de 1 del costat del Nivell
  - c) Dues Cel·les no poden estar associades amb les mateixes instàncies per tots els Nivells de la Base.

El conjunt inicial de Dimensions trobades després de definir el gràdul donen lloc a la Base. Tal i com s'ha esmentat en l'apartat 2.2.3 aquestes dimensions són el Temps, l'Immoble i la Zona per tant una certa oferta d'un immoble s'identifica per un mes, un immoble i una subàrea. S'ha de considerar, però, que per la naturalesa del producte que s'ofereix, un immoble, aquest és únic, es a dir, un immoble determinat només existeix una vegada, en un lloc determinat. No hi pot haver un mateix immoble en dues subàrees diferents així doncs les Dimensions Immoble i Temps són suficients per definir una oferta.

- 2) **Restriccions d'agregació:** Per poder fer una agregació s'han de complir tres condicions:
  - a) **Compatibilitat:** ha d'existir compatibilitat entre la mesura que es vol agregar, la Dimensió i l'operació d'agregació. Per exemple es pot obtenir el total d'immobles venuts en un any, sumant els immobles venuts cadascun dels mesos de l'any. En canvi el nombre d'ofertes anuals no serà la suma d'ofertes mensuals donat que un mateix immoble es pot oferir en mesos diferents i per tant la suma és incompatible amb les ofertes i la Dimensió Temps. En el cas dels immobles venuts es tracta d'una Mesura additiva, que es pot sumar en qualsevol Dimensió però l'oferta d'immobles és una

Mesura semiadditiva ja que no es pot sumar en la Dimensió Temps. Altres operacions que s'utilitzaran per obtenir la informació necessària són la mitjana, el màxim i el mínim, que s'aplicaran per esbrinar els preus i metres quadrats màxims, mínims i mitjans o els temps mitjans de venda.

- b) **Disjuntivitat:** Per no córrer el risc de considerar més d'una vegada la mateixa dada, els Nivells han de contenir instàncies amb conjunts de parts disjunts. En el cas dels immobles, on s'ha definit el Nivell Tipus, cada immoble només pot formar part d'un sol tipus per tant no hi ha perill de considerar un immoble més d'una vegada en el moment d'aplicar una operació d'agregació.
  - c) **Completesa:** Totes les instàncies han de participar com a mínim en una instància dels Nivells immediatament superiors de la jerarquia d'agregació. Un Nivell ha de cobrir completament els Nivells que té per sota. En el aquest cas, si es considera la Dimensió Àrea, succeeix, per exemple, que els immobles que es troben a Andorra només tenen 2 Nivells (país i subàrea) per tant s'han de completar els nivells entremitjos amb les dades que es considerin convenients.
- 3) **Restriccions de transitivitat:** És imprescindible que les operacions d'agregació siguin transitives, és a dir el resultat d'operar amb resultats parcials i el resultat d'operar amb dades atòmiques han de coincidir. Si no ho són s'ha d'operar sempre amb les dades atòmiques.

### 2.2.9 Estudiar la viabilitat

Per saber si l'estrella que forma el model multidimensional és implementable, s'ha d'estimar l'espai que ocuparan totes les dades. Normalment el volum de dades que ocupen les Dimensions és molt inferior al que ocupen els Fets, per tant és suficient considerar l'espai que ocupen les instàncies del Fet. Això s'aconsegueix multiplicant el nombre d'instàncies de cada Nivell que forma la Base. En aquest cas hi ha 92.555 immobles i dades emmagatzemades durant 25 mesos. El nombre màxim d'instàncies serà  $92.555 \times 25 = 2.313.875$ .

S'ha de considerar, però, que aquesta seria la mida de la Cel·la si no es vengués cap immoble i cada mes s'haguessin d'oferir tots els immobles. Com hi haurà mesos en que s'oferiran menys immobles, bé perquè ja s'ha venut algun o perquè un determinat immoble encara no s'ha incorporat al catàleg d'ofertes, el volum de dades serà més petit que la xifra estimada.

Un cop conegut el nombre d'instàncies, s'ha de multiplicar la xifra pel nombre de bytes que ocupa cadascuna. El Fet conté els tres identificadors de les Dimensions que són del tipus

int i les Mesures preu, mesosEnOferta, difPreuInicial i difPreuAnterior (també de tipus int) i venut (el booleà que indica si l'immoble està venut o no). Per tant cada cel·la ocuparà 29 bytes. En total, la mida màxima serà de  $2.313.875 \times 29 \text{ bytes} = 67.102.375 \text{ bytes} = 64 \text{ Mbytes}$ .

Donat que la mida que ocuparà el volum de dades actual no és elevat no hauria d'haver cap problema ni amb l'emmagatzemament ni amb la velocitat de resposta a les consultes o amb l'actualització de les dades per tant es considera que el disseny del model multidimensional és adequat.

En el cas en que el model no fos viable s'hauria de tornar a començar el procés de disseny pel primer punt.

A la Figura 2 es mostra una representació en UML del model escollit.

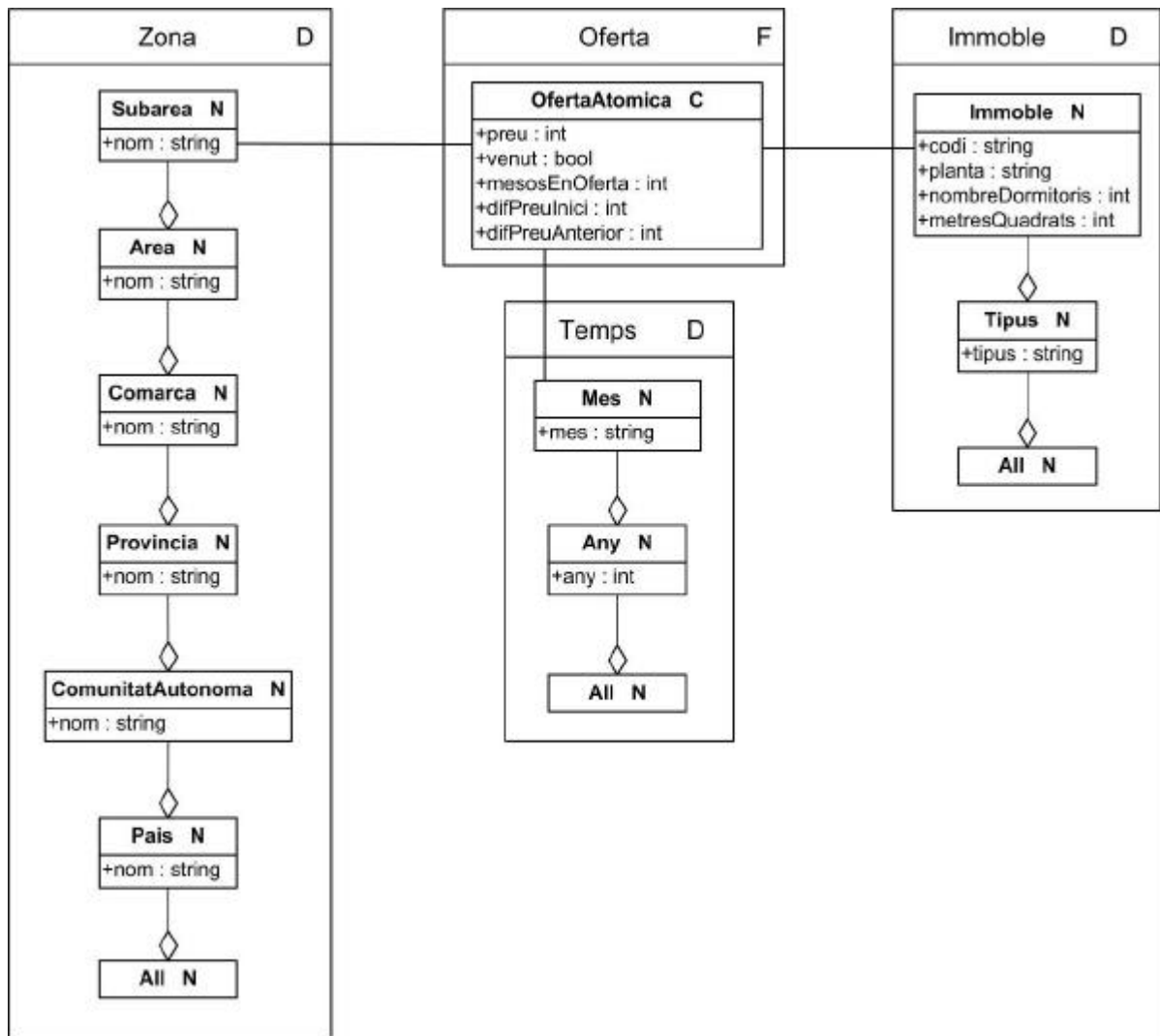


Figura 2. Esquema UML del disseny conceptual del model multidimensional per al magatzem de dades de l'empresa Sostre per a tothom.

## 3 Capítol 3 - Disseny

### 3.1 Diagrama de l'arquitectura programari

Els programaris utilitzats pel desenvolupament del projecte han estat:

- ★ Base de dades Oracle 10g Express Edition: Per a la gestió de la base de dades.
- ★ SQL\*Loader: Per a la càrrega de dades a la taula CATALEG a partir del fitxer catalog.dat proporcionat per l'empresa.
- ★ SQL Developer: Per a la construcció de taules i l'execució del procés ETL sobre la base de dades.
- ★ Oracle BI Discoverer Administrator: Per a la construcció de l'àrea de negoci.
- ★ Oracle BI Discoverer Desktop: Per a la construcció del llibre de treball.

A la Figura 3 es mostra com d'interrelaciona aquest programari en el context del projecte.

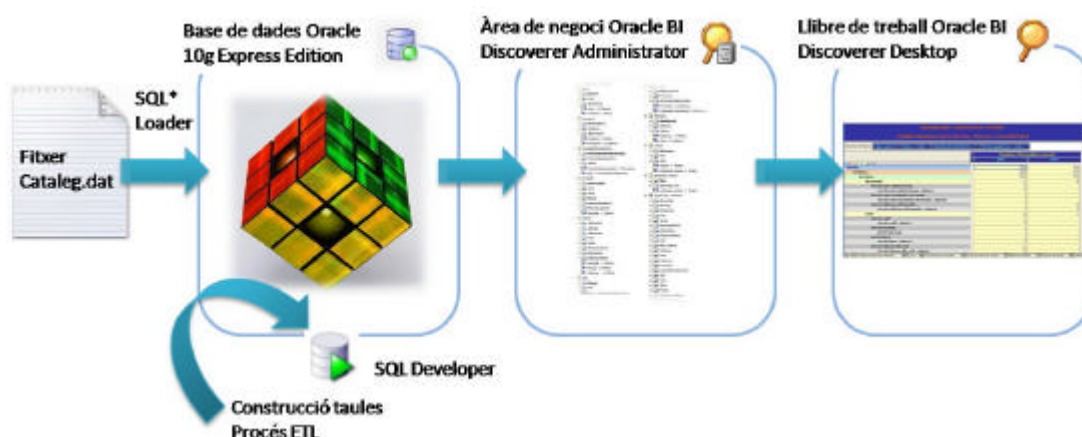


Figura 3. Diagrama de l'arquitectura programari

### 3.2 Diagrama de l'arquitectura maquinari

Per la implantació del magatzem de dades a l'empresa "Sostre per a tothom", la solució més adequada seria centralitzar la base de dades a un servidor de manera que els diferents treballadors de l'empresa s'hi poguessin connectar des dels seus punts de treball per realitzar les consultes pertinents.

En el següent diagrama (Figura 4) es mostra l'arquitectura del programari necessari per portar a terme la implantació.

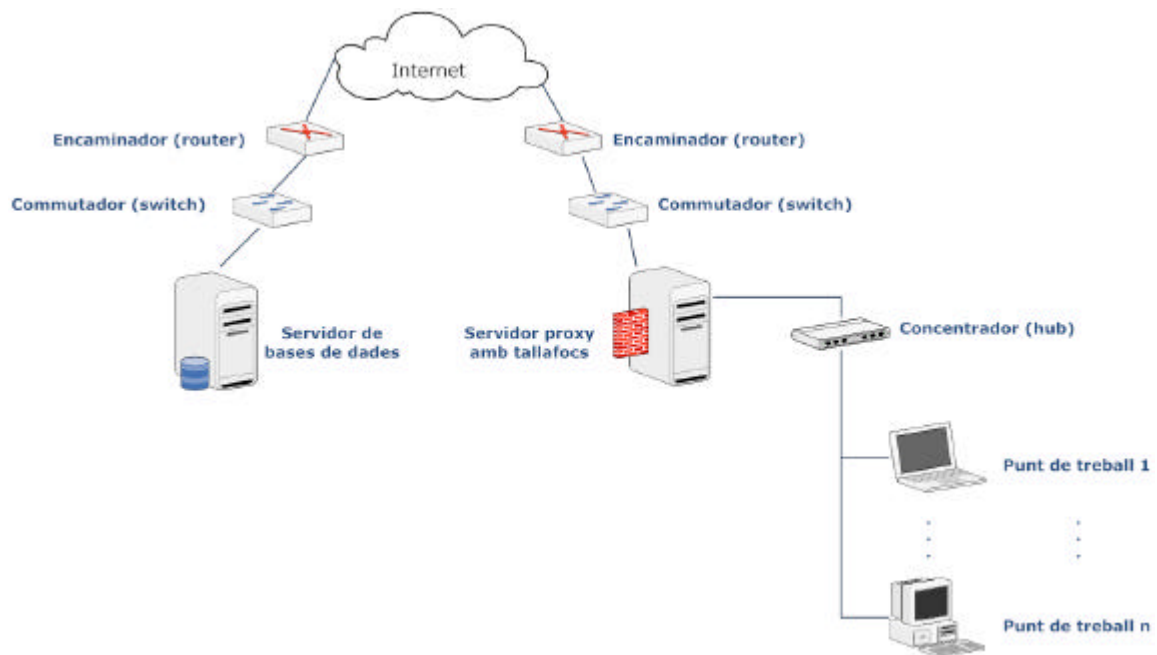


Figura 4. Diagrama de l'arquitectura maquinari

### 3.3 Disseny de la base de dades i diagrama del model físic

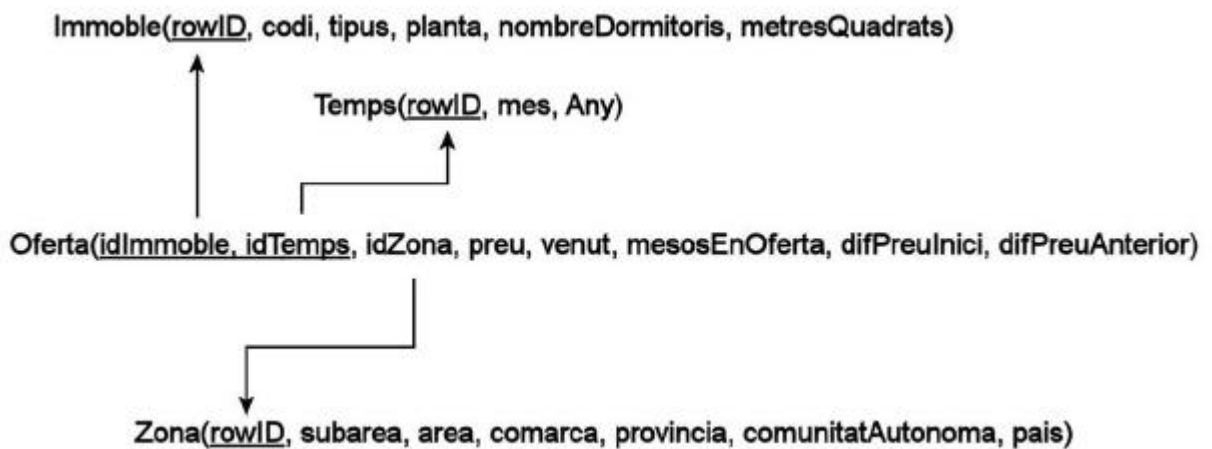
#### 3.3.1 Disseny lògic

Per a la implementació d'un model dimensional en forma d'Estrella, es necessita una taula per al Fet, on cada fila representa una Cel·la, i una taula per a cadascuna de les Dimensions. La taula del Fet està lligada a les taules de Dimensions mitjançant les claus foranes. La taula del Fet conté, a més, les Mesures i les taules de Dimensions contenen els Descriptors. Les jerarquies d'agregació queden implícites en els valors dels atributs de les taules de Dimensions. Els atributs que formen la Base de la Cel·la atòmica constitueixen la clau primària.

Com ja s'ha comentat a l'apartat 2.2.9, l'espai ocupat per la taula Fet és molt superior al de les taules Dimensions per tant és recomanable reduir aquest espai. Una manera d'aconseguir-lo és utilitzant substituïts de la clau primària que ocupin menys espai en les taules Dimensions. D'aquesta manera també es redueix la grandària de la clau primària de la taula Fet. Per exemple, en lloc d'utilitzar com a clau primària de la taula Immoble el codi de l'immoble, es fa servir un identificador de fila i en les taules Temps i Zona es substitueix el nom del mes o de la subàrea per els indicadors de fila corresponents. Aquesta substitució té un altre avantatge i és que si en algun moment es fan canvis en els identificadors, com per exemple una modificació en el codi de l'immoble o en el nom d'una zona, aquests no afectaran a les dades inserides a la taula d'ofertes.



Basat en el disseny conceptual desenvolupat a l'apartat 2, l'esquema relacional de l'Estrella queda com es mostra a continuació (Figura 5).



**Figura 5.** Esquema relacional en forma d'estrella del model multidimensional per al magatzem de dades de l'empresa Sostre per a tothom

### 3.3.2 El disseny físic

A la Figura 6 es mostra el disseny físic de la base de dades. S'han convertit les entitats en taules i les relacions en claus foranes (indicades en cursiva). Els atributs s'han convertit en columnes. Les claus primàries s'indiquen en negreta.

Per a la Dimensió Zona s'ha optat per fer taules independents per a cada subnivell inserint en cada taula una clau forana que fa referència al nivell immediatament superior. D'aquesta manera s'evita una única taula amb un gran volum de dades.

A continuació es fa una descripció detallada de cada taula:

- ★ PAIS: Conté els identificadors i els noms dels països
- ★ COMUNITAT\_AUTONOMA: Conté els identificadors i els noms de les comunitats autònomes i l'identificador del país al que pertanyen.
- ★ PROVINCIA: Conté els identificadors i els noms de les províncies i l'identificador de la comunitat autònoma a la que pertanyen.
- ★ COMARCA: Conté els identificadors i els noms de les comarques i l'identificador de la província a la que pertanyen.
- ★ AREA: Conté els identificadors i els noms de les àrees i l'identificador de la comarca a la que pertanyen.
- ★ SUBAREA: Conté els identificadors i els noms de les subàrees i l'identificador de l'àrea a la que pertanyen.

- ★ TEMPS: Conté cadascuna de les combinacions mes-any des de gener de 2006 fins gener de 2008 i un identificador de cada combinació. S'inclou a més un identificador del mes per tal de poder ordenar els mesos de forma lògica en les consultes i evitar així l'ordenació alfabètica per defecte.
- ★ IMMOBLE: Conté tota la informació que caracteritza un immoble: el seu codi, la seva tipologia, la planta en la que es troba i el nombre de dormitoris i els metres quadrats que té. Cada immoble s'identifica per un valor numèric.
- ★ OFERTA: Conté un històric de totes les ofertes que ha tingut la immobiliària. Cada oferta individual es caracteritza per: l'identificador de l'immoble que està en oferta, l'identificador del mes i l'identificador de la subàrea on es troba l'immoble. A més, el preu d'oferta de l'immoble en aquell mes en concret, si l'immoble s'ha venut en aquell mes o estava en oferta, el temps en mesos que porta en oferta i la diferència de preu respecte al preu inicial i respecte al preu del mes anterior.

En el cas de les taules PAIS, COMUNITAT\_AUTONOMA, PROVINCIA i TEMPS en el moment de la creació es carregen també les dades (referiu-vos al fitxer `crear_taules.sql`).

En l'esquema es mostra també el disseny de la taula CATALEG. Aquesta taula no forma part del magatzem de dades final. És una taula temporal creada a partir de les dades proporcionades per l'empresa i que s'utilitzarà en el procés de càrrega (referiu-vos a l'apartat 3.4.3). Respecte a les dades originals, només s'ha eliminat el camp que indicava el codi de zona donat que s'havien detectat moltes incongruències en aquest camp i no és necessari per adjudicar un immoble a una determinada zona ja que entre els registres dels diferents immobles es troben registres que indiquen en quina zona es troben els immobles.

Les dades de la taula CATALEG es carreguen amb Oracle SQL\*Loader executant la sentència `sqlldr control = catalegctl` des del mateix directori on es troben els fitxers `catalegctl` (que conté les instruccions de càrrega) i `cataleg.dat` (el fitxer proporcionat per l'empresa que conté la informació que es vol carregar).

Per crear les taules s'executa el *script* `crear_taules.sql` des d'Oracle SQL Developer. Les taules s'han construït de manera que els identificadors es generen automàticament començant pel valor 1.

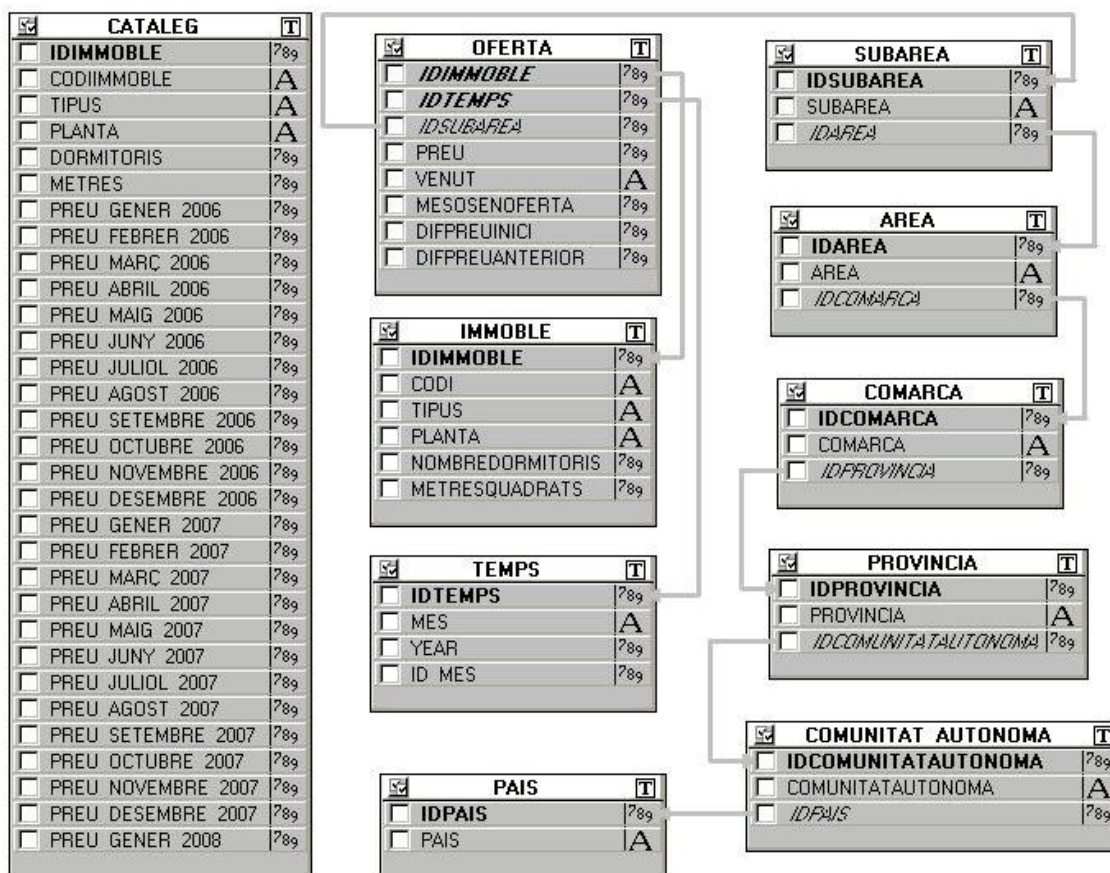


Figura 6. Disseny físic de la base de dades. Es mostra a més el disseny de la taula temporal CATALEG

Per tal d'optimitzar el temps de resposta de les consultes s'ha optat per la utilització d'índex de combinació (*join index*). Aquests es troben definits sobre una clau forana, de manera que contenen els valors d'una taula i apunten a una altra.

Degut a que sempre s'entra en l'índex pel valor corresponent al primer atribut i després es comproven els valors de la resta d'atributs seqüencialment és important triar l'ordre correcte dels atributs. El més adequat es posar en primera posició l'atribut amb la selectivitat més gran. En aquest cas és diferent posar el temps com a primer atribut del Fet i després l'immoble que a l'inrevés. En el primer cas, si es fixa primer el temps en el moment de fer una consulta, després hi haurà una gran quantitat d'immobles per triar els que compleixin les condicions demanades. En canvi, si es fixa primer l'immoble, el nombre de mesos en els que estan en oferta és molt més reduït.

Els *scripts* de creació dels índex utilitzats poden ser consultats al fitxer `crear_index.sql` adjunt.

### **3.4 El procés d'extracció, transformació i càrrega de dades**

El procés ETL (Extract, Transform, Load) permet fusionar dades des de múltiples orígens, donar-les el format adequat, netejar i corregir les dades errònies i finalment carregar-les en un magatzem de dades pel seu posterior anàlisi amb la finalitat d'obtenir una determinada informació.

#### **3.4.1 Extracció**

L'extracció consisteix en ajuntar dades que procedeixen de diferents bases de dades o fitxers, per exemple de diferents departaments d'una mateixa empresa, en un únic fitxer. Generalment les dades fusionades de diferents orígens tenen organitzacions o formats diferents per tant és important convertir-lo a un únic format que permeti els processos posteriors. A més l'extracció de dades ha de tenir el mínim impacte possible en el sistema d'origen donat que si el volum de dades que s'extreu és molt gran podria fer més lent o fins i tot col·lapsar el sistema d'origen impedit el seu funcionament normal. Per tant, aquest procés ha de ser programat en horaris o dies en els que el sistema no estigui gaire carregat de feina.

En aquest cas, l'empresa ja ha proporcionat el fitxer amb les dades que es desitgen analitzar per tant no és necessari fer el procés d'extracció.

#### **3.4.2 Transformació**

En el procés de transformació es tracta d'uniformitzar les dades, d'aplicar certes funcions per donar-les el format adequat per fer la càrrega o realitzar les accions de tractament d'excepcions que es considerin oportunes. Alguns exemples poden ser:

- ★ Substituir caràcters que defineixen grups o classes per xifres.
- ★ Dividir columnes que contenen més d'una dada en varies.
- ★ Eliminar els registres amb dades errònies o substituir aquestes dades pel valor nul o un determinat codi que les identifiqui.
- ★ Donar un únic valor a dades diferents que representen el mateix.

El catàleg "Sostre per a tothom" proporcionat per l'empresa immobiliària no necessita gaires tractaments donat que les dades són força correctes. On si s'han de fer algunes correccions és als camps corresponents al nombre de dormitoris, la planta i els metres quadrats. S'observen casos en els que les dades no tenen sentit, com immobles de menys de 10 metres quadrats o que es troben en la planta 99. També hi ha casos en que un determinat tipus d'immoble ha de complir una determinada condició, com per exemple el camp planta ha

de ser nul pels xalets o els estudis han de tenir 0 dormitoris, així que s'han de fer les transformacions necessàries.

A continuació es mostra una relació dels errors identificats i de les transformacions aplicades a les dades per tal de preparar-les pel procés de càrrega:

- ★ Àtics que es troben en una planta baixa (bj): el valor de la planta s'ha substituït pel valor indefinit (--).
- ★ Estudis que no tenen 0 dormitoris: Per definició un estudi no té dormitoris per tant s'ha posat el valor del dormitori a 0 en tots els estudis.
- ★ Immobles que no són estudis i que tenen 0 dormitoris: S'ha considerat que si un immoble tenia el valor del dormitori a 0 és que no es disposa d'aquesta informació i per tant s'ha posat el valor a NULL.
- ★ Xalets que no tenen la planta a NULL: El camp planta indica la planta en la que es troba l'immoble. Els xalets no es troben a cap planta per tant aquesta informació no és aplicable als xalets i per aquesta raó s'ha posat el valor de la planta a NULL en els xalets.
- ★ Immobles que no són xalets i que tenen la planta a NULL: S'ha considerat que si un immoble no tenia el valor de la planta és que no es disposa d'aquesta informació i per tant s'ha posat el valor a indefinit (--).
- ★ Immobles amb més de 24 dormitoris: S'han trobat alguns casos amb valors exageradament alts al camp dormitori. S'ha considerat 24 com el valor màxim de dormitoris i per tant tots el valor l'excedien s'han posat a NULL.
- ★ Immobles on el valor de la planta era més de 60: Considerant que l'edifici més alt a Espanya té 57 plantes s'han considerat erronis tots els valors de planta que excedien 60 i per tant s'han posat a NULL.
- ★ Immobles amb superfícies de menys de 10 m<sup>2</sup> o més de 3000 m<sup>2</sup>: Els valors de superfície exageradament baixos o exageradament alts s'han considerat erronis i per tant s'han posat a NULL.
- ★ El caràcter Ž: S'ha comprovat que al fitxer original l'apòstrof s'havia substituït pel caràcter Ž. Per no tenir problemes amb el símbol de cometes senzilles que s'utilitza a les seqüències SQL s'ha decidit substituir aquest caràcter per un espai.

Executant el fitxer adjunt `correccio_errors.sql` des d'Oracle SQL Developer aquest errors es poden corregir de manera automàtica.

Abans d'executar cada *script* de correcció d'errors es fa un recompte dels registres erroris. Aquest mateix recompte executat després de la correcció de l'error ha de retornar el valor 0 si la correcció s'ha realitzat amb èxit.

### 3.4.3 Càrrega

Una vegada transformades les dades contingudes a la taula CATALEG s'omplen les taules COMARCA, AREA, SUBAREA, IMMOBLE i OFERTA executant l'*script* `carrega_dades.pls` des d'Oracle SQL Developer.

Amb aquest *script* es fa un recorregut pels 93.360 registres de la taula CATALEG. Segons si el registre indica una zona o un immoble es carregen les dades en unes taules o altres.

Si el registre indica una zona, el primer pas és dividir la cadena en els diferents subnivells, prenent la seqüència `:::` com a indicador d'on s'ha de fer la divisió. Si la cadena té menys de 5 subnivells, aquests s'omplen amb les paraules comarca, area i subarea respectivament. Si en canvi la cadena té més de 5 subnivells, els últims es fusionen en un de sol per denominar la subàrea.

**Exemple 1:** cadena de 4 subnivells: `# :::` girona `:::` alt empordà `:::` àrea de figueres:

- ★ #: aquesta cadena es descarta
- ★ girona: aquesta cadena s'identifica com província. No cal inserir-la a la taula PROVINCIA donat que aquesta taula ja s'ha omplert prèviament.
- ★ alt empordà: aquesta cadena s'identifica com comarca. S'introdueix a la taula COMARCA.
- ★ àrea de figueres: aquesta cadena s'identifica com àrea. S'introdueix a la taula AREA.
- ★ Com no hi ha més subnivells, a la taula SUBAREA s'introdueix la cadena àrea de figueres – subàrea

**Exemple 2:** cadena de 6 subnivells `# :::` barcelona `:::` barcelonès `:::` barcelona `:::` gràcia `:::` vallcarca:

- ★ #: aquesta cadena es descarta

- ★ barcelona: aquesta cadena s'identifica com província. No cal inserir-la a la taula PROVINCIA donat que aquesta taula ja s'ha omplert prèviament.
- ★ barcelonès: aquesta cadena s'identifica com comarca. S'introdueix a la taula COMARCA.
- ★ barcelona: aquesta cadena s'identifica com àrea. S'introdueix a la taula AREA
- ★ gràcia: aquesta cadena s'identifica com subàrea. S'introdueix a la taula SUBAREA
- ★ vallcarca: aquest sisè subnivell es fusiona amb l'anterior i la subàrea passa a denominar-se gràcia· vallcarca.

Si el registre indica un immoble, primer s'omple la taula IMMOBLE amb les dades de codi, tipus, planta, nombre de dormitoris i metres quadrats. Després es fa un recorregut pels camps que indiquen el preu de l'immoble per a cada mes i s'introdueix un registre d'oferta corresponent a aquest immoble en la taula OFERTA. Hi haurà tants registres d'oferta per un mateix immoble com mesos estigui en oferta aquest immoble. Quan es llegeix un primer preu, s'emmagatzema aquest valor com a preu inicial per poder fer els càlculs de diferència de preu respecte al preu inicial. Quan es llegeix l'últim preu s'indica que l'immoble s'ha venut.

Una vegada finalitzat el procés de càrrega, que triga uns 25 minuts aproximadament, tindrem 92.555 registres a la taula IMMOBLE i 1.091.388 registres a la taula OFERTA.

La base de dades s'ha exportat amb el nom de `bd_silvia.dmp`

En el següent pseudocodi es fa una descripció exhaustiva del funcionament de l'algoritme de càrrega:

```
// Mitjançant un cursor es llegeixen els registres de la taula CATALEG
Llegir registre_i
// Cas en que es llegeix un registre que indica una zona
Si registre_i comença per #
    Guardar registre
    llegint_poblacions = true
fi del si
// Cas en que es llegeix un registre que indica un immoble
Si registre_i no comença per #
    // Cas en que el registre anterior indicava una zona
    Si llegint_poblacions == true
        Dividir registre guardat per marcador ::
        Inserir elements a taules Comarca, Area i Subarea
        Guardar idSubarea
        llegint_poblacions = false
    fi del si
// Una vegada omplertes les taules COMARCA, AREA I SUBAREA o si el registre anterior no indicava una
zona, es llegeixen els primers camps del registre i s'omple la taula immoble
```

```

Extreure dades de les columnes 1 a 5 (codi de l'immoble, tipus, planta, nombreDormitoris,
metresQuadrats) i inserir als camps corresponents de la taula immoble
// Llegim els camps corresponents als preus
Inserir els preus d'un immoble en un array
seguir = true
preu_inicial = nul
mesosEnOferta = 1
venut = false
// Es llegeixen els preus amb un bucle
llegir preu_k
// Si hi ha un preu significa que l'immoble està en oferta aquell mes
si preu_k ? nul
    // Si és el primer preu que es llegeix, es considera aquest preu com l'inicial i es posen a nul les
dues    diferències de preu (respecte al preu inicial i respecte al preu del mes anterior)
    si preu_inicial == nul
        preu_inicial = preu_k
        difPreuInici = nul
        difPreuAnterior = nul
    fi del si
    // Si no és el primer preu que es llegeix, es calculen les dues diferències de preu
    si preu_inicial ? nul
        difPreuInici = preu_k - preu_inicial
        difPreuAnterior = preu_k - preu_(k-1)
    fi del si
    // Una vegada calculats els preus s'omple un registre de la taula
    Inserir registre a taula Oferta amb idTemps = k
    mesosEnOferta = mesosEnOferta + 1
fi del si
// Si no hi ha un preu pot significar que l'immoble encara no està en oferta o que s'ha venut el mes
anterior
si preu_k == nul
    // Es pot comprovar si l'immoble s'ha venut el mes anterior si hi ha un preu inicial. Llavors en el
registre del mes anterior es marca que l'immoble s'ha venut i no es continua llegint més preus
    si preu_inicial ? nul
        actualitzar venut = true en registre amb idTemps = k-1
        seguir = false
    fi del si
    // Si no hi ha un preu inicial és que l'immoble no s'ofereix encara i per tant s'ha de consultar el
mes següent
    si preu_inicial == nul
        continuar amb el bucle que llegeix els preus
    fi del si
fi del si
fi del bucle
// Una vegada llegits tots els camps corresponents a un registre es continua llegint els registres de la
taula CATALEG fins finalitzar
fi del cursor

```

A l'esquema de la Figura 7 es mostra aquest funcionament de manera gràfica:



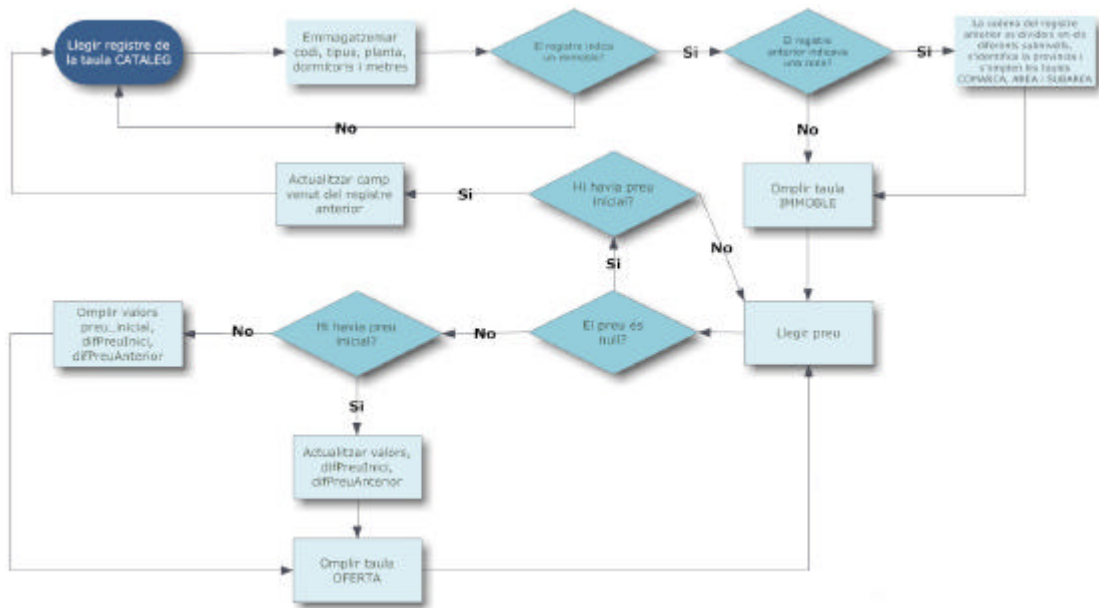


Figura 7. Esquema de càrrega de dades a les taules COMARCA, AREA, SUBAREA, IMMOBLE i OFERTA a partir de les dades de la taula CATALOG.

### 3.5 Disseny dels informes creats

#### 3.5.1 La creació de l'àrea de negoci

Amb el programari Oracle BI Discoverer Administrator s'ha creat l'àrea de negoci anomenada "TFC-Sostre per a tothom". En aquesta àrea de negoci es troben carpetes que contenen els elements de les taules creades amb anterioritat i una carpeta personalitzada "Sostre per a tothom" on s'han inserit tots els elements necessaris per fer les consultes (Figura 8). Com totes les consultes utilitzen pràcticament els mateixos elements s'ha optat per crear una sola carpeta.

A més s'ha creat una jerarquia que indica l'ordenació dels diferents subnivells de les zones i una altra que indica la jerarquia de temps (Figura 9) i les classes d'elements necessàries (Figura 10).

Aquesta àrea de negoci s'ha exportat amb el nom TFC-Sostre per a tothom.eex

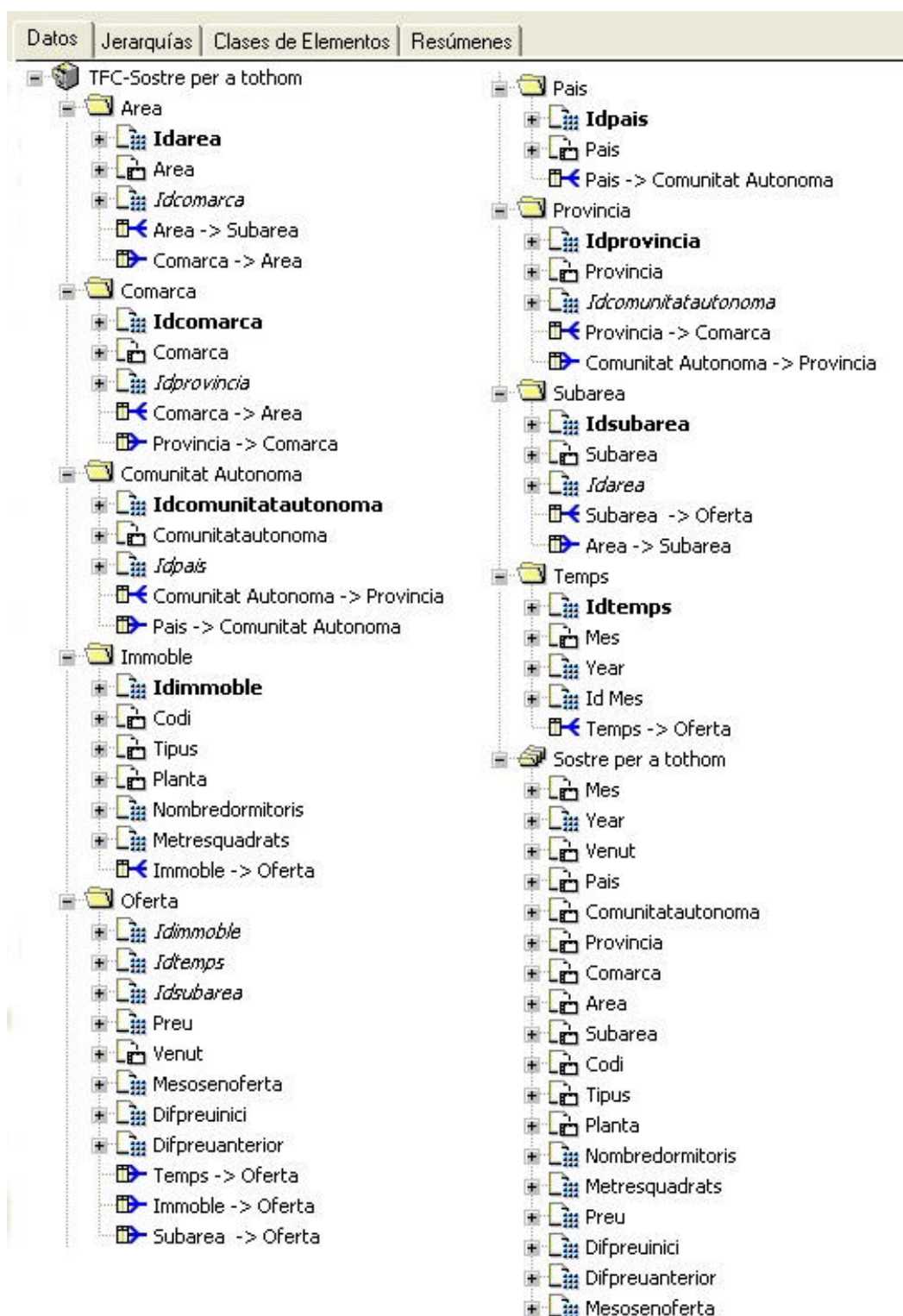


Figura 8. Dades de l'àrea de negoci TFC-Sostre per a tothom

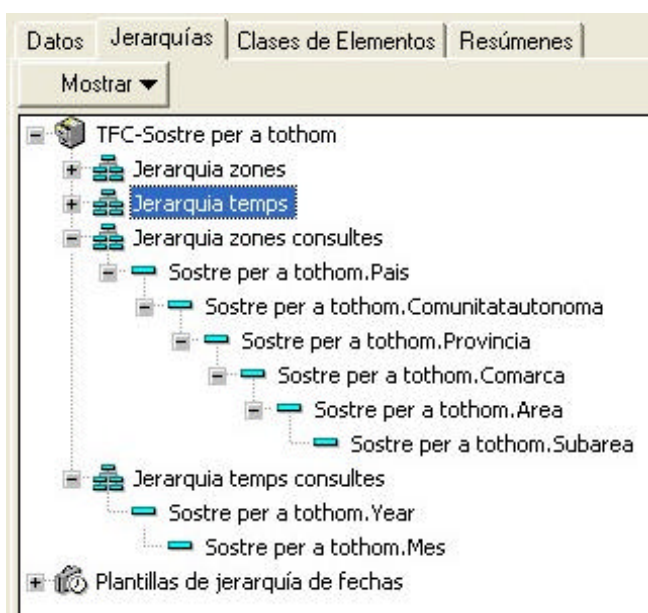


Figura 9. Jerarquies de l'àrea de negoci TFC-Sostre per a tothom

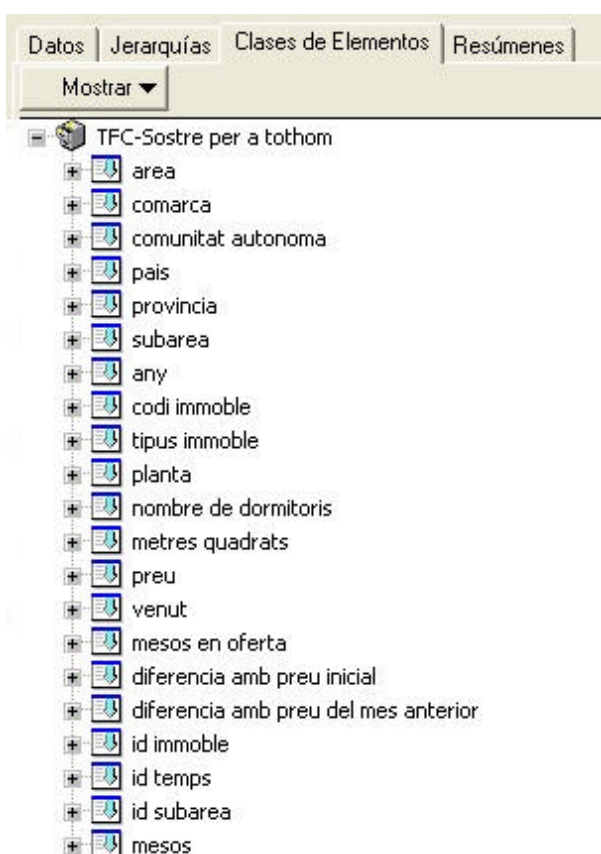


Figura 10. Clases d'elements de l'àrea de negoci TFC-Sostre per a tothom

### 3.5.2 La creació del llibre de treball

Amb el programari Oracle BI Discoverer Desktop s'ha creat el llibre de treball anomenat "Sostre per a tothom – workbook". Aquest llibre consta de 9 fulls cadascun dels quals conté una taula o una matriu que dona resposta a les consultes plantejades per la

immobiliària. Presenten l'avantatge que els resultats poden ser mostrats a nivell general per un gran volum de dades o restringint les condicions, com ara escollint només uns determinats tipus o característiques d'immoble o uns mesos o anys concrets, s'obtenen consultes més concretes.

Per a cada consulta s'han escollit els elements necessaris de la carpeta "Sostre per a tothom" creada a l'àrea de negoci. A la carpeta es troben els següents elements:

- ★ els atributs d'un immoble (codi, tipus, planta, nombre de dormitoris i metres quadrats)
- ★ els diferents subnivells de la Dimensió Zona (país, comunitat autònoma, província, comarca, àrea i subàrea)
- ★ els subnivells de la Dimensió Temps (mes i any)
- ★ les mesures del Fet, preu, venut, mesos en oferta, diferència de preu respecte al preu inicial i diferència de preu respecte al preu anterior.

La major part d'aquest elements són comuns a totes les consultes i només hi ha certs elements que s'utilitzen en casos concrets com per exemple les mesures Difpreuinici i Difpreuanterior que es fan servir a les consultes sobre els immobles que han baixat de preu, la mesura Mesosenoferta per a la consulta referent als temps mitjans de venda o la mesura Venut que s'utilitza en aquells casos en els que es desitja fer una diferenciació entre immobles oferts i immobles venuts.

El llibre de treball s'ha exportat amb el nom `Sostre per a tothom - workbook.DIS`

Segons requeriments de l'empresa, els informes que s'han generat i que es troben al llibre de treball són:

- ★ Llistat d'immobles que han baixat de preu respecte al preu inicial
- ★ Llistat d'immobles que han baixat de preu respecte al preu del mes anterior
- ★ Preus màxims, mínims i mitjans per zona, tipologia i característiques
- ★ Evolució de preus d'oferta per zona, tipologia i característiques
- ★ Evolució de preus de venda per zona, tipologia i característiques
- ★ Metres quadrats màxims, mínims i mitjans per zona, tipologia i característiques
- ★ Temps mitjans de venda per zona, tipologia i característiques
- ★ Immobles disponibles per zona, tipologia i característiques
- ★ Immobles venuts per zona, tipologia i característiques

A continuació es mostren imatges dels fulls del llibre de treball, una breu descripció de com s'han creat i de la informació que proporcionen.

## 3.6 Descripció de les consultes i informes

### 3.6.1 Full 1: Llistat baixada preus inicial

Aquest full s'ha creat amb els següents elements i disseny de taula (Figura 11) i amb la condició  $Difpreuinici < 0$ :

	Pais	Comunitatautonomia	Provincia	Codi	Planta	Tipus	Nombredormitoris	Metresquadrats	Preu	Difpreuinici
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Figura 11. Elements i disseny de taula del full Llistat baixada preus inicial

Aquesta taula mostra tots aquells immobles que han baixat de preu respecte al seu preu inicial. Es mostren les característiques de l'immoble, la seva localització, el seu preu en el mes de la consulta i la rebaixa respecte al preu inicial. Es pot escollir per a què mes i any es vol fer la consulta i el detall en la jerarquia per zones.

En aquest cas concret es mostren immobles de la província de Barcelona que el mes de desembre de 2006 havien baixat de preu respecte a l'inici, ordenats per la seva tipologia (Figura 12).

IMMOBILIÀRIA "SOSTRE PER A TOTHOM"										
LLISTAT D'IMMOBLES QUE HAN BAIXAT DE PREU RESPECTE AL PREU INICIAL										
Elementos de Página: Year: 2006 Mes: desembre										
	Pais	Comunitat autonoma	Provincia	Codi	Planta	Tipus	Nombre de dormitoris	Metres quadrats	Preu	Rebaixa
> 1	Espanya	Catalunya	barcelona	V0000000339604	5	àtics	3	87	396.500	-23.500
> 2	Espanya	Catalunya	barcelona	V0000000486003	4	àtics	4	162	321.000	-6.000
> 3	Espanya	Catalunya	barcelona	V0000000578041	7	àtics	2	142	205.000	-35.000
> 4	Espanya	Catalunya	barcelona	V0000000608149	5	àtics	2	72	228.385	-11.419
> 5	Espanya	Catalunya	barcelona	V0000000694464	4	àtics	4	130	430.000	-12.729
> 6	Espanya	Catalunya	barcelona	V0000000699661	4	àtics	2	122	265.000	-6.000
> 7	Espanya	Catalunya	barcelona	V0000000676352	5	àtics	4	90	229.000	-6.000
> 8	Espanya	Catalunya	barcelona	V0000000679254	4	àtics	2	75	262.000	-3.481
> 9	Espanya	Catalunya	barcelona	V0000000395980	4	àtics	4	120	462.779	-12.020
> 10	Espanya	Catalunya	barcelona	V0000000510028	4	àtics	3	260	618.000	-41.000
> 11	Espanya	Catalunya	barcelona	V0000000526163	5	àtics	2	50	230.000	-7.000
> 12	Espanya	Catalunya	barcelona	V0000000536943	3	àtics	3	130	480.809	-39.427
> 13	Espanya	Catalunya	barcelona	V0000000542993	-	àtics	5	173	1.052.000	-90.000
> 14	Espanya	Catalunya	barcelona	V0000000595723	8	àtics	2	73	478.000	-22.000
> 15	Espanya	Catalunya	barcelona	V0000000599467	5	àtics	2	85	558.041	-36.051
> 16	Espanya	Catalunya	barcelona	V0000000592338	4	àtics	3	70	390.405	-18.631
> 17	Espanya	Catalunya	barcelona	V0000000596471	5	àtics	3	68	255.435	-6.005
> 18	Espanya	Catalunya	barcelona	V0000000619358	8	àtics	3	100	665.000	-34.000
> 19	Espanya	Catalunya	barcelona	V0000000629990	3	àtics	3	168	462.780	-41.759
> 20	Espanya	Catalunya	barcelona	V0000000691399	7	àtics	3	190	1.051.771	-90.229

Figura 12. Full 1: Llistat baixada preus inicial

### 3.6.2 Full 2: Llistat baixada preus anterior

Aquest full s'ha creat amb els següents elements i disseny de taula (Figura 13) i amb la condició  $Difpreuanterior < 0$ :

Sostre per a tothom										
Elementos de Página: Year: Mes										
	Pais	Comunitat autonoma	Provincia	Codi	Planta	Tipus	Nombredormitoris	Metresquadrats	Preu	Difpreuanterior
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Figura 13. Elements i disseny de taula del full Llistat baixada preus anterior

Aquesta taula és similar a l'anterior però en aquest cas mostren tots aquells immobles que han baixat de preu respecte al mes anterior. Es mostren les característiques de l'immoble, la seva localització, el seu preu en el mes de la consulta i la rebaixa respecte al mes anterior. Es pot escollir per a què mes i any es vol fer la consulta i el detall en la jerarquia per zones.

En aquest cas concret es mostren immobles de la província de Barcelona que el mes de febrer de 2007 havien baixat de preu respecte al gener de 2007, ordenats per la seva tipologia (Figura 14).



IMMOBILIÀRIA "SOSTRE PER A TOTHOM"											
LLISTAT D'IMMOBLES QUE HAN BAIXAT DE PREU RESPECTE AL MES ANTERIOR											
Elementos de Pàgina: Year: 2007 Mes: febrer											
	Pais	Comunitat autonoma	Provincia	Codi	Planta	Tipus	Nombre de dormitoris	Metres quadrats	Preu	Rebaxa	
> 1	Espanya	Catalunya	barcelona	VP0000000544237	7	àtics	2	70	238.385	-12020	
> 2	Espanya	Catalunya	barcelona	VP0000000760324	4	àtics	2	80	336.000	-14000	
> 3	Espanya	Catalunya	barcelona	VP000000097376	8	àtics	3	141	900.000	-41416	
> 4	Espanya	Catalunya	barcelona	VP000000097376	8	àtics	3	144	817.000	-42447	
> 5	Espanya	Catalunya	barcelona	VP0000000526163	5	àtics	2	50	227.000	-3000	
> 6	Espanya	Catalunya	barcelona	VP0000000714384	9	àtics	4	212	839.901	-32500	
> 7	Espanya	Catalunya	barcelona	VP0000000762611	4	àtics	2	112	336.567	-6009	
> 8	Espanya	Catalunya	barcelona	VP0000000772365	4	àtics	5	198	907.700	-24020	
> 9	Espanya	Catalunya	barcelona	VP0000000825400	7	àtics	3	65	356.000	-25500	
> 10	Espanya	Catalunya	barcelona	VP0000000844144	4	àtics	2	105	350.000	-9440	
> 11	Espanya	Catalunya	barcelona	VP0000000878533	9	àtics	3	120	596.848	-30042	
> 12	Espanya	Catalunya	barcelona	VP0000000893501	4	àtics	2	40	375.000	-21000	
> 13	Espanya	Catalunya	barcelona	VP0000000903548	5	àtics	1	67	578.000	-12000	
> 14	Espanya	Catalunya	barcelona	VP0000000915855	6	àtics	4	180	961.000	-87000	
> 15	Espanya	Catalunya	barcelona	VP0000000928743	8	àtics	4	125	988.090	-153510	
> 16	Espanya	Catalunya	barcelona	VP0000000943554	8	àtics	3	100	811.367	-24040	
> 17	Espanya	Catalunya	barcelona	VP0000000955162	7	àtics	4	100	336.500	-24100	
> 18	Espanya	Catalunya	barcelona	VY0000000364660	3	àtics	4	103	299.000	-10000	
> 19	Espanya	Catalunya	barcelona	VY0000000407572	10	àtics	3	91	324.500	-12100	
> 20	Espanya	Catalunya	barcelona	VY0000000408882	3	àtics	3	108	325.000	-15000	

Figura 14. Full 2: Llistat baixada preus anterior

### 3.6.3 Full 3: Preus

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 15) i amb la condició Venut='N':

Figura 15. Elements i disseny de taula del full Preus

Aquesta matriu mostra els preus màxims, mínims i mitjans per zona, tipologia i característiques i per a un temps determinat. Es pot escollir el detall en la jerarquia per zones i en la jerarquia de temps. Es pot escollir a més quina tipologia i quines característiques es volen consultar.

En aquest cas concret es mostren els preus màxims, mínims i mitjans a una escala temporal d'any per Espanya en general i per Andorra fins a un subnivell de subàrea. A més s'ha

restringit la consulta als pisos que es troben a la primera planta i que tenen 3 dormitoris (Figura 16).

IMMOBILIÀRIA *SOSTRE PER A TOTHOM*									
PREUS MÀXIMS, MITJANS I MÍNIMS PER ZONA, TIPOLOGIA I CARACTERÍSTIQUES									
Elementos de Página: Tipo: piso Planta: 1 Número de dormitorios: 3 Metres cuadrados: <Todo>									
	PREU MÀXIM			PREU MITJÀ			PREU MÍNIM		
	2006	2007	2008	2006	2007	2008	2006	2007	2008
Espanya	3.000.000	1.250.000	7.250.000	309.065	302.075	295.335	86.000	55.000	66.111
Principat d'Andorra	632.800	632.800	632.800	455.466	456.884	431.690	380.000	299.000	299.000
Andorra	632.800	632.800	632.800	455.466	456.884	431.690	380.000	299.000	299.000
andorra	632.800	632.800	632.800	455.466	456.884	431.690	380.000	299.000	299.000
andorra - comarca	412.915	380.000	380.000	390.573	380.000	380.000	380.000	380.000	380.000
andorra - comarca - area	412.915	380.000	380.000	390.573	380.000	380.000	380.000	380.000	380.000
andorra - comarca - area - subarea	412.915	380.000	380.000	390.573	380.000	380.000	380.000	380.000	380.000
camp - canillo - ordino	632.800	632.800	632.800	449.686	454.164	407.810	418.000	299.000	299.000
massana - andorra - escaldes	595.000	595.000	595.000	595.000	595.194	526.397	595.000	500.000	500.000

Figura 16. Full 3: Preus

### 3.6.4 Full 4: Evolució preus d'oferta

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 17) i amb la condició Venut='N':

Figura 17. Elements i disseny de taula del full Evolució preus d'oferta

Aquesta matriu mostra com evoluciona al llarg del temps els preus d'oferta mitjans per una determinada zona i per una tipologia i característiques determinades. Es pot escollir el detall en la jerarquia per zones i en la jerarquia de temps. Es pot escollir a més quina tipologia i quines característiques es volen consultar.

En aquest cas concret es mostra l'evolució dels preus de tots els immobles d'Espanya i Andorra a una escala temporal de mes pel any 2006 i per una escala temporal d'any pels anys 2007 i 2008. Aquestes dades no es mostren només de forma numèrica sinó també de forma gràfica (Figura 18).



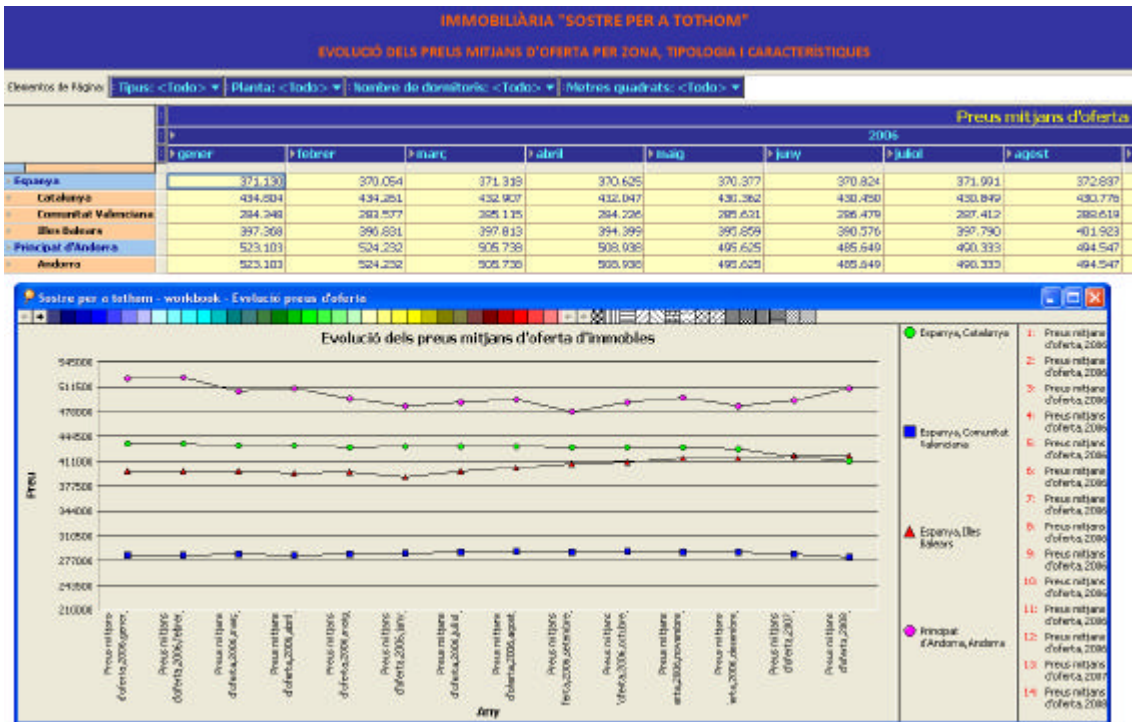


Figura 18. Full 4: Evolució preus d'oferta

### 3.6.5 Full 5: Evolució preus de venda

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 19) i amb la condició Venut='S':

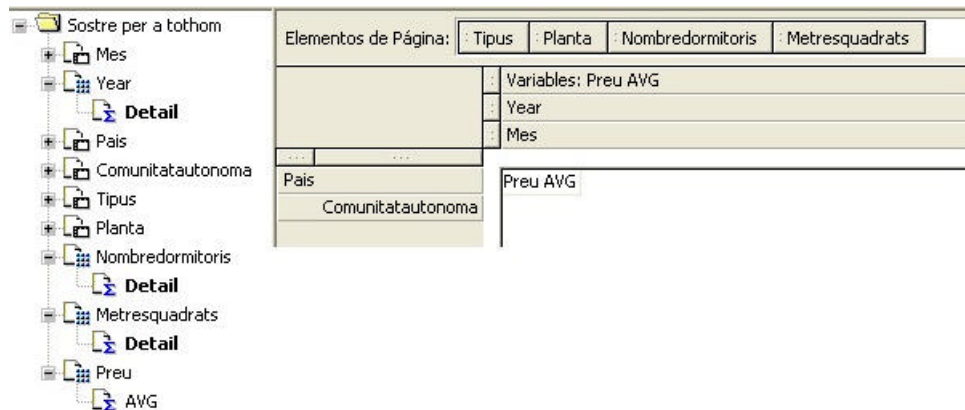


Figura 19. Elements i disseny de taula del full Evolució preus de venda

Aquesta matriu mostra com evoluciona al llarg del temps els preus de venda mitjans per una determinada zona i per una tipologia i característiques determinades. Es pot escollir el detall en la jerarquia per zones i en la jerarquia de temps. Es pot escollir a més quina tipologia i quines característiques es volen consultar.

En aquest cas concret es mostra l'evolució dels preus de tots els immobles d'Espanya i Andorra a una escala temporal de mes pel any 2007 i per una escala temporal d'any pels anys

2006 i 2008. Aquestes dades no es mostren només de forma numèrica sinó també de forma gràfica (Figura 20).

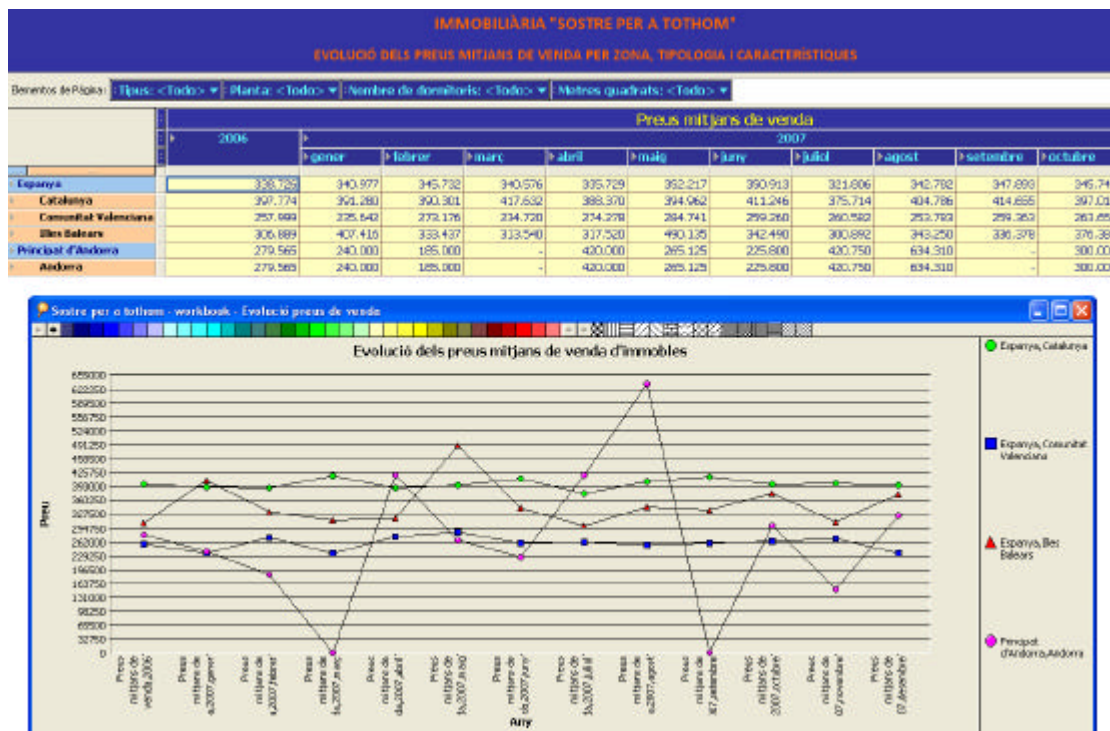


Figura 20. Full 5: Evolució preus de venda

### 3.6.6 Full 6: Metres quadrats

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 21):

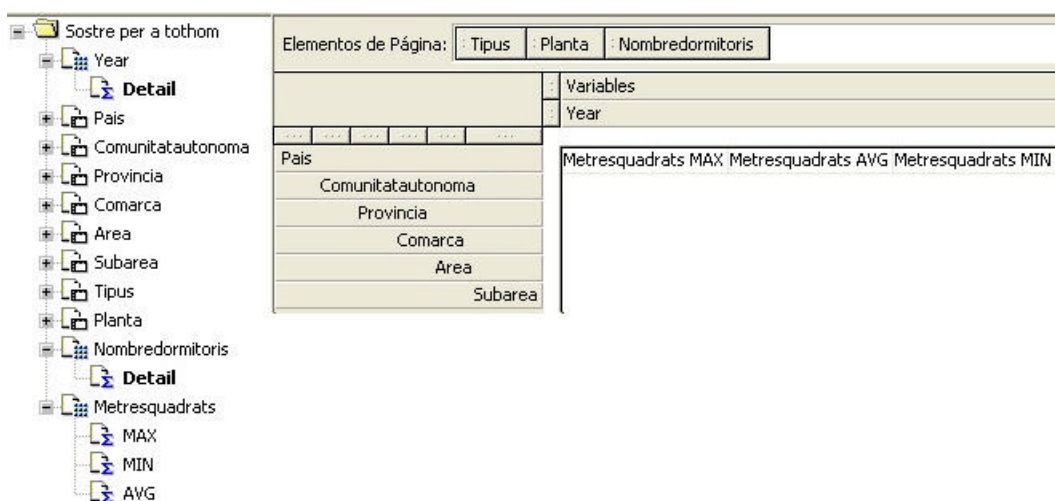


Figura 21. Elements i disseny de taula del full Metres quadrats

Aquesta matriu mostra els metres quadrats màxims, mínims i mitjans per zona, tipologia i característiques i per a un temps determinat. Es pot escollir el detall en la jerarquia

per zones i en la jerarquia de temps. Es pot escollir a més quina tipologia i quines característiques es volen consultar.

En aquest cas concret es mostren els metres quadrats màxims, mínims i mitjans a una escala temporal d'any per Espanya fins a un subnivell de subàrea. A més s'ha restringit la consulta als xalets de 3 dormitoris (Figura 22).

IMMOBILIÀRIA "SOSTRE PER A TOTHOM"									
METRES QUADRATS MÀXIMS, MITJANS I MÍNIMS PER ZONA, TIPOLOGIA I CARACTERÍSTIQUES									
Elementos de Página: Tipus: chalets   Planta: <Todo>   Nombredormitoris: 3									
	METRES QUADRATS MÀXIMS			METRES QUADRATS MITJANS			METRES QUADRATS MÍNIMS		
	2006	2007	2008	2006	2007	2008	2006	2007	2008
Espanya	2427	2427	2427	169	169	169	35	35	35
Catalunya	2427	2427	2427	174	175	173	35	35	35
Barcelona	1000	1750	1750	181	182	180	48	48	48
alt penedès	350	350	350	172	172	170	76	75	75
àrea de sant cugat i osona	300	300	300	149	148	148	78	75	75
àrea de sant cugat i osona - subàrea	350	300	300	140	148	148	78	75	75
àrea de santa margarida i els monjos	329	320	320	240	236	230	132	140	140
àrea de santa margarida i els monjos - subàrea	329	320	320	243	236	230	132	140	140
àrea de vilanova del penedès	200	200	156	152	163	156	136	136	156
àrea de vilanova del penedès - subàrea	200	200	156	152	163	156	136	136	156
arona	347	347	320	160	174	174	71	90	90
àrea de calal	300	300	300	181	193	199	138	140	150
àrea de calal - subàrea	200	300	300	181	193	199	138	140	150
àrea de gualba	360	360	160	160	160	160	160	160	160
pla de sant mart	360	360	160	160	160	160	160	160	160
àrea de piera	347	347	320	151	164	164	71	90	90
àrea de piera - subàrea	347	347	320	151	164	164	71	90	90
àrea de vilanova del carí	300	300	299	262	262	260	230	230	230
àrea de vilanova del carí - subàrea	300	300	300	262	262	260	230	230	230

Figura 22. Full 6: Metres quadrats

### 3.6.7 Full 7: Temps de venda

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 23) i amb la condició  $Venut = 'S'$ :

<ul style="list-style-type: none"> <li>[-] Sostre per a tothom                             <ul style="list-style-type: none"> <li>[+] Year                                     <ul style="list-style-type: none"> <li>[+] Detail</li> </ul> </li> <li>[+] Pais                                     <ul style="list-style-type: none"> <li>[+] Comunitatautonomia   <ul style="list-style-type: none"> <li>[+] Provincia   <ul style="list-style-type: none"> <li>[+] Comarca   <ul style="list-style-type: none"> <li>[+] Tipus   <ul style="list-style-type: none"> <li>[+] Planta   <ul style="list-style-type: none"> <li>[+] Nombredormitoris   <ul style="list-style-type: none"> <li>[+] Detail</li> </ul> </li> <li>[+] Metresquadrats   <ul style="list-style-type: none"> <li>[+] Detail</li> </ul> </li> <li>[+] Mesosenoferta   <ul style="list-style-type: none"> <li>[+] AVG</li> </ul> </li> </ul> </li> </ul> </li> </ul> </li> </ul> </li> </ul> </li> </ul> </li> </ul> </li> </ul>	Elementos de Página: <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>Tipus</td> <td>Planta</td> <td>Nombredormitoris</td> <td>Metresquadrats</td> </tr> </table>	Tipus	Planta	Nombredormitoris	Metresquadrats
	Tipus	Planta	Nombredormitoris	Metresquadrats	
		Variables: Mesosenoferta AVG Year			
		Mesosenoferta AVG			
Pais Comunitatautonomia Provincia Comarca					

Figura 23. Elements i disseny de taula del full Temps de venda

Aquesta matriu mostra els temps mitjans de venda per zona, tipologia i característiques. Es pot escollir el detall en la jerarquia per zones i en la jerarquia de temps. Es pot escollir a més quina tipologia i quines característiques es volen consultar.

En aquest cas concret es mostren els temps mitjans de venda per diverses zones de Catalunya i la Comunitat Valenciana fins a un subnivell de comarca i per Espanya en general. Aquestes dades es mostren pels anys 2006 i 2007. Per l'any 2008 només es disposen dades de gener i en aquest mes no s'ha produït cap venda. S'ha limitat la consulta a pisos de 3 dormitoris (Figura 24).

IMMOBILIÀRIA "SOSTRE PER A TOTHOM"			
TEMPS MITJANS DE VENDA (EN MESOS) PER ZONA, TIPOLOGIA I CARACTERÍSTIQUES			
Elementos de Página: Tipus: pisos   Planta: <Todo>   Nombre de dormitorios: 3   Metres quadrats: <Todo>			
	Temps mitjans de venda en mesos		
	2006	2007	
Espanya	6	12	
Catalunya	6	12	
barcelona	6	12	
girona	6	12	
alt empordà	6	13	
baix empordà	6	13	
gironès	5	11	
la cerdanya	8	11	
la garrotxa	-	10	
la selva	6	12	
pla de l'estany	4	9	
ripollès	4	7	
l'elida	5	10	
alt Urgell - cerdanya	6	5	
noquera	3	10	
pallars jussà	2	13	
segarra	4	11	
segrià	6	9	
solsonès	9	-	
urgell - pla d'urgell	-	7	
val d'aran	-	14	
tarragona	6	12	
Comunitat Valenciana	5	11	
alicante	6	12	
castellón	5	11	
valencia	5	11	
Illes Balears	5	11	

Figura 24. Full 6: Temps de venda

### 3.6.8 Full 8: Immobles disponibles

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 25) i amb la condició Venut='N':

Sostre per a tothom		Elementos de Página: Tipus   Planta   Nombredormitoris   Metresquadrats   Year	
Mes		Variables: Idimmoble	COUNT
Year		Mes	
Detail			
Pais		Pais	Idimmoble
Subarea		Subarea	COUNT
Codi			
COUNT			
Tipus			
Planta			
Nombredormitoris			
Detail			
Metresquadrats			
Detail			

Figura 25. Elements i disseny de taula del full Immobles disponibles



Aquesta matriu mostra la quantitat d'immobles disponibles per zona, tipologia i característiques. Es pot escollir el detall en la jerarquia per zones i la tipologia i característiques que es volen consultar.

En aquest cas concret es mostren tots els immobles que van estar disponibles els diferents mesos de l'any 2006 per a Espanya i Andorra, en aquest últim cas fins el nivell de subàrea (Figura 26).

IMMOBILIÀRIA "SOSTRE PER A TOTHOM"												
NOMBRE D'IMMOBLES DISPONIBLES PER ZONA, TIPOLOGIA I CARACTERÍSTIQUES												
Elementos de Página: Tipus: <Tudo>   Planta: <Tudo>   Nombredormitoris: <Tudo>   Metresquadrats: <Tudo>   Year: 2006												
	Nombre d'immobles disponibles											
	gener	febrer	març	abril	maig	juny	juliol	agost	setembre	octubre	novembre	desembre
Espanya	34695	35317	35294	35970	36827	37642	37973	39043	40292	40845	40822	42081
Principat d'Andorra	60	61	68	70	75	70	79	79	84	87	100	106
andorra - comarca - area - subarea	45	46	50	51	55	56	58	55	54	52	52	52
encamp - castell - ardiso - area - subarea	9	9	10	10	10	11	11	13	17	19	26	31
massana - andorra - escaldes - area - subarea	6	6	8	9	10	10	10	10	12	14	22	23

Figura 26. Full 8: Immobles disponibles

### 3.6.9 Full 9: Immobles venuts

Aquest full s'ha creat amb els següents elements i disseny de matriu (Figura 27) i amb la condició Venut='S':

The screenshot shows a BI tool interface. On the left is a tree view of data elements under the folder 'Sostre per a tothom'. The elements include 'Year', 'Detail', 'Pais', 'Comunitatautonomia', 'Provincia', 'Comarca', 'Area', 'Subarea', 'Codi', 'COUNT', 'Tipus', 'Planta', 'Nombredormitoris', and 'Metresquadrats', each with a 'Detail' icon. On the right, the 'Elements de Página' are 'Tipus', 'Planta', 'Nombredormitoris', and 'Metresquadrats'. Below this, the 'Variables' section lists 'Idimmoble COUNT' and 'Year'. The table design shows a grid with columns for 'Pais', 'Comunitatautonomia', 'Provincia', 'Comarca', 'Area', and 'Subarea', and a single row for 'Idimmoble COUNT'.

Figura 27. Elements i disseny de taula del full Immobles venuts

Aquesta matriu mostra la quantitat d'immobles venuts per zona, tipologia i característiques. Es pot escollir el detall en la jerarquia per zones i en la jerarquia per temps i també la tipologia i característiques que es volen consultar.

En aquest cas concret es mostren els pisos de 4 dormitoris que s'han venut a Espanya i Andorra els anys 2006 i 2007. Pel cas d'Espanya es mostra fins a un subnivell de província i pel cas d'Andorra fins a un subnivell de subàrea (Figura 28).

IMMOBILIÀRIA "SOSTRE PER A TOTHOM"			
NOMBRE D'IMMOBLES VENUTS PER ZONA, TIPOLOGIA I CARACTERÍSTIQUES			
Elementos de Página: Tipus: pisos ▶ Planta: <Todo> ▶ Nombre de dormitorios: 4 ▶ Metres quadrats: <Todo> ▶			
	Nombre d'immobles venuts		
	2006	2007	
Espanya	2262	2428	
Catalunya	1382	1403	
barcelona	1132	1216	
girona	29	34	
lleida	0	12	
tarragona	212	231	
Comunitat Valenciana	818	878	
alicante	158	164	
castellón	43	80	
valencia	617	634	
Illes Balears	02	07	
ibiza-formentera	5	2	
mallorca	55	53	
menorca	2	2	
Principat d'Andorra	1	1	
Andorra	1	1	
andorra	1	1	
encamp - canillo - ordino	1	1	
encamp - canillo - ordino - area	1	1	
encamp - canillo - ordino - area - subarea	1	1	

Figura 28. Full 9: Immobles venuts

### 3.6.10 L'immoble estàndard espanyol i andorrà

Per determinar quines són les característiques, preu i tipologia que caracteritza un immoble estàndard d'Espanya i d'Andorra s'ha optat per classificar els immobles en diferents grups segons aquest paràmetres i mitjançant una representació gràfica d'aquesta classificació, esbrinar quins són els grups més nombrosos. D'aquesta manera s'obté informació visual i clara no només del grup més nombrós sinó també de la proporció que representa aquest grup respecte als altres.

A continuació es mostra l'*script* utilitzat per determinar el nombre d'immobles andorrans per a cada preu. La consulta equivalent per Espanya seria substituint `pais.pais` like 'Principat%' per `pais.pais = 'Espanya'`

```
SELECT
  count(immoble.idimmoble) as nombre_immobles,
  oferta.preu
FROM
  immoble,
  oferta,
  pais,
  provincia,
  comarca,
  comunitat_autonoma,
  area,
  subarea
WHERE
  pais.pais like 'Principat%' and
  oferta.mesosenoferta = 1 and
  oferta.idimmoble = immoble.idimmoble and
  oferta.idsubarea = subarea.idsubarea and
```

```
subarea.idarea = area.idarea and
area.idcomarca = comarca.idcomarca and
comarca.idprovincia = provincia.idprovincia and
provincia.idcomunitatautonoma =
comunitat_autonoma.idcomunitatautonoma and
comunitat_autonoma.idpais = pais.idpais
GROUP BY
oferta.preu
ORDER BY
oferta.preu
```

Per fer la classificació dels immobles espanyols segons els metres quadrats s'ha emprat l'*script* següent. Substituint el paràmetre metresquadrats per tipus, planta i nombredormitoris es poden fer la resta de consultes.

```
SELECT
  count(immoble.idimmoble) as nombre_immobles,
  immoble.metresquadrats
FROM
  immoble,
  oferta,
  pais,
  provincia,
  comarca,
  comunitat_autonoma,
  area,
  subarea
WHERE
  pais.pais = 'Espanya' and
  oferta.mesosenoferta = 1 and
  oferta.idimmoble = immoble.idimmoble and
  oferta.idsubarea = subarea.idsubarea and
  subarea.idarea = area.idarea and
  area.idcomarca = comarca.idcomarca and
  comarca.idprovincia = provincia.idprovincia and
  provincia.idcomunitatautonoma =
  comunitat_autonoma.idcomunitatautonoma and
  comunitat_autonoma.idpais = pais.idpais
GROUP BY
  immoble.metresquadrats
ORDER BY
  immoble.metresquadrats
```

### 3.6.10.1 L'immoble estàndard espanyol

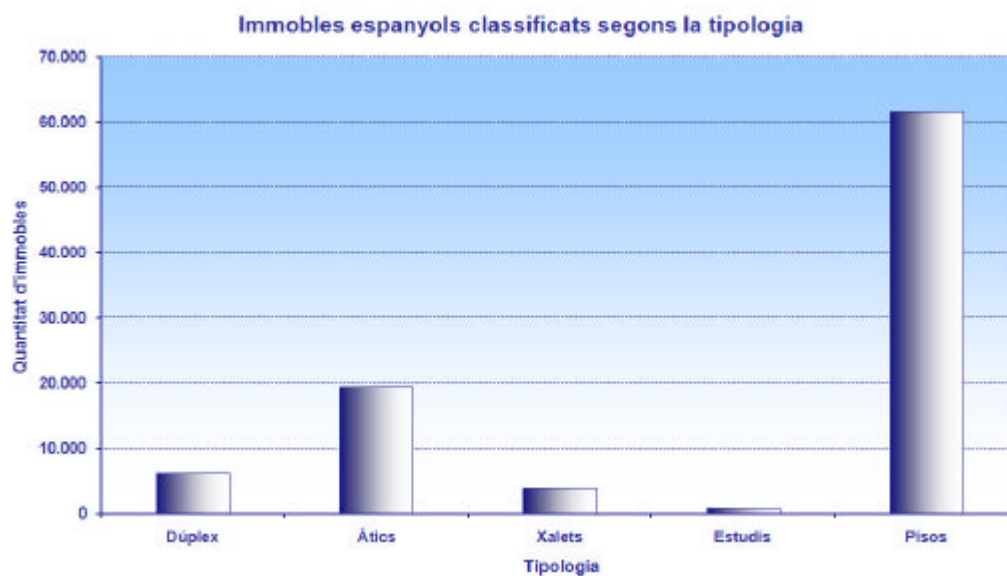


Figura 29. Immables espanyols classificats segons la tipologia

Els pisos són clarament la tipologia més abundant entre els immables espanyols (Figura 29).

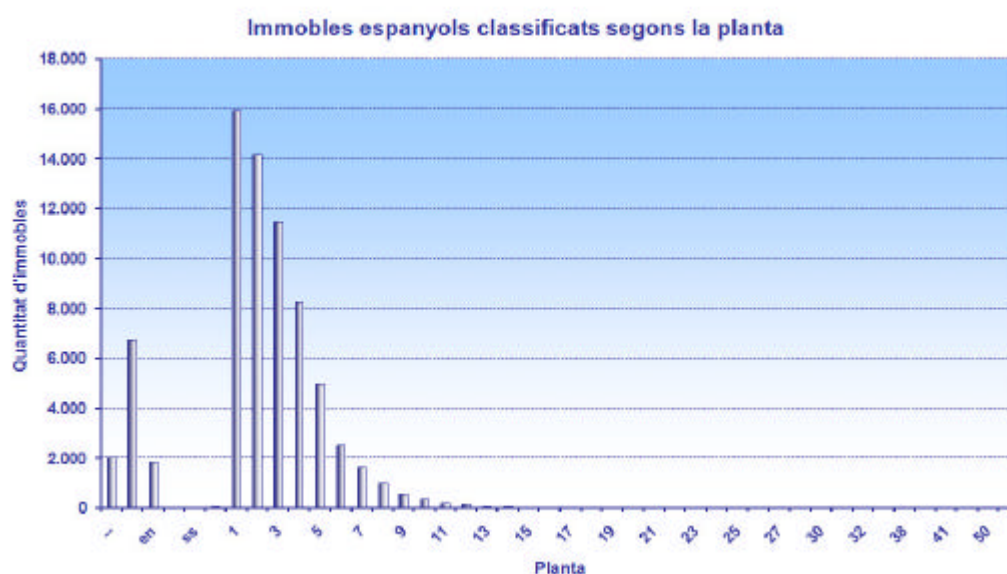


Figura 30. Immables espanyols classificats segons la planta

La majoria dels immables espanyols es troben a la primera planta. La freqüència va disminuint a mesura que augmenta la planta. A partir de la planta 18, el nombre d'immables es pràcticament menyspreable (Figura 30).



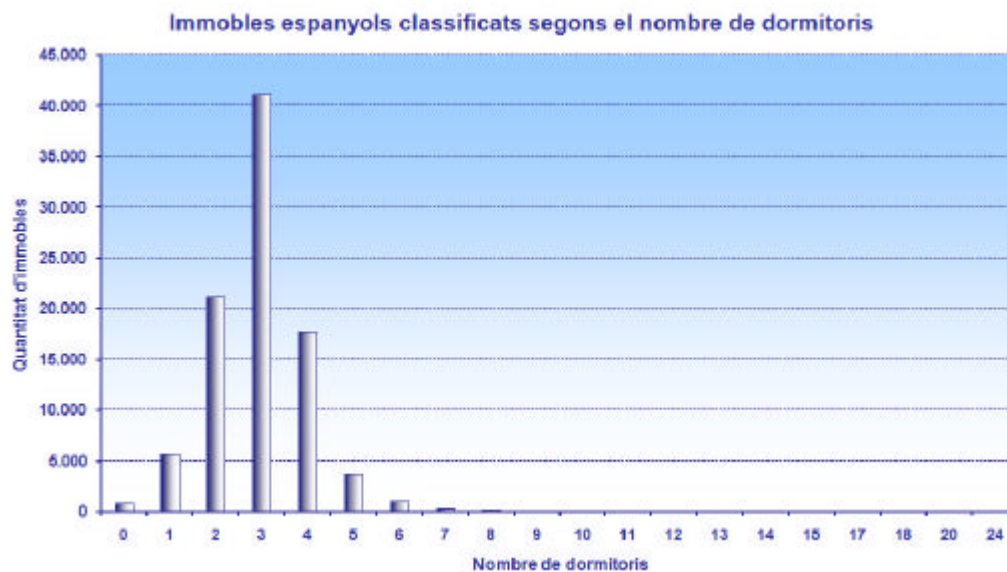


Figura 31. Immobilis espanyols classificats segons el nombre de dormitoris

La majoria dels immobilis espanyols tenen 3 dormitoris. Es pot apreciar una distribució normal entre 0 i 7 dormitoris. Per sobre de 7 dormitoris el nombre d'immobilis és pràcticament menyspreable (Figura 31).

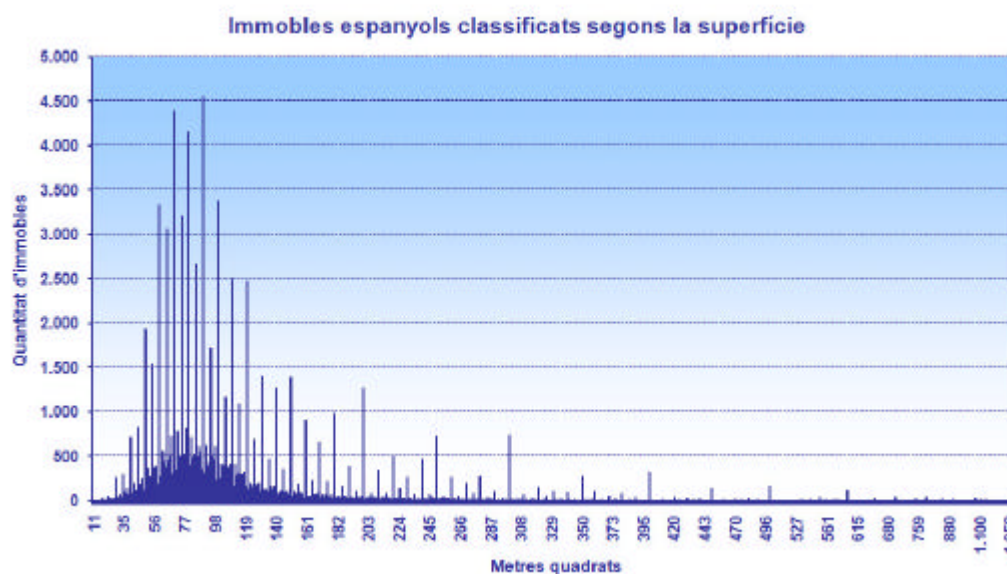


Figura 32. Immobilis espanyols classificats segons els metres quadrats

Degut a la gran diversitat en els metres quadrats és difícil apreciar quina és la superfície més habitual (Figura 32) però limitant el gràfic a un rang de entre 30 i 150 m², que és on es troben la majoria d'immobilis, es pot veure que la superfície més abundant és 90 m² (Figura 33).

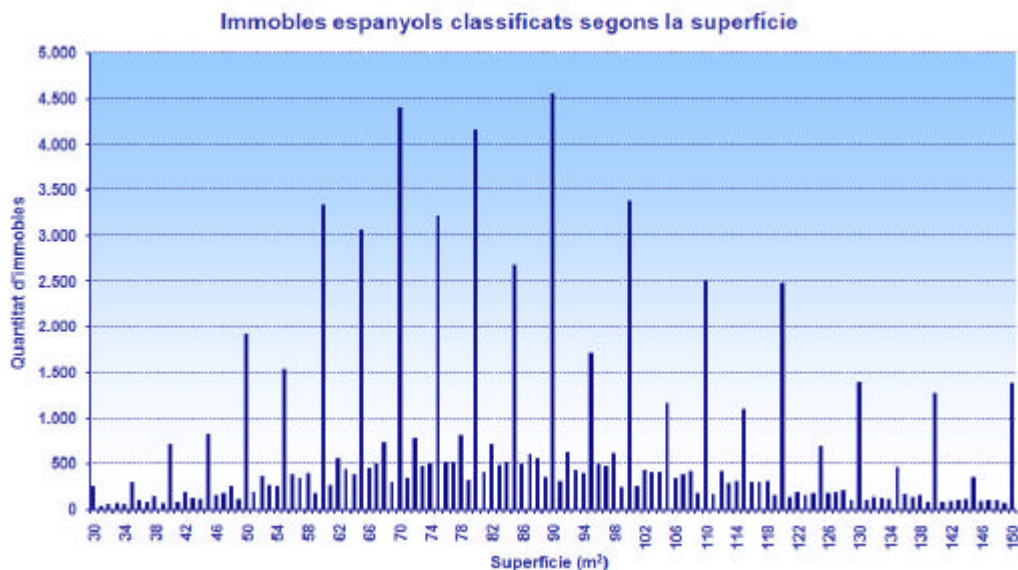


Figura 33. Immobles espanyols classificats segons els metres quadrats (visió ampliada)

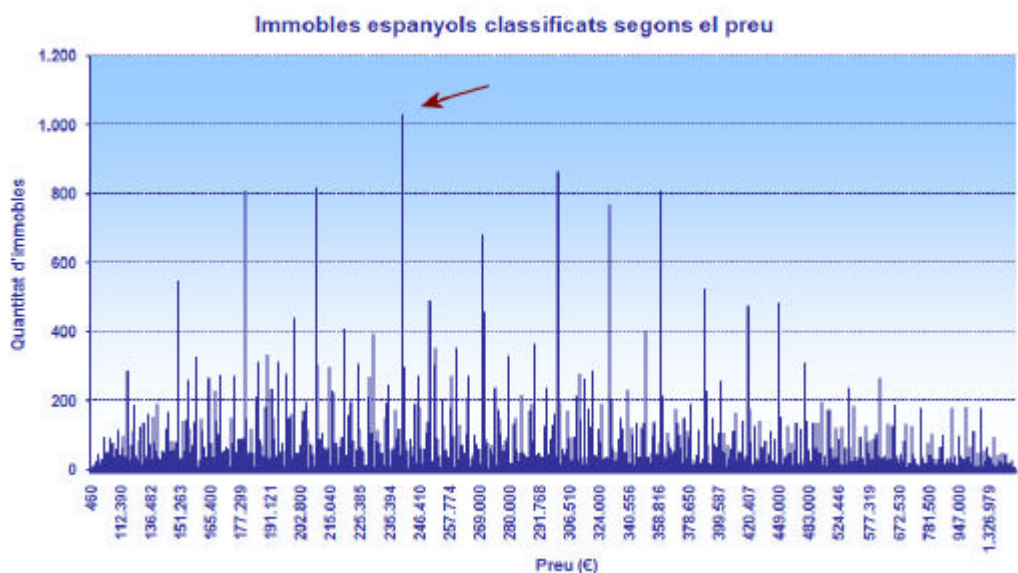


Figura 34. Immobles espanyols classificats segons el preu

Respecte als preus també existeix una gran diversitat però es pot apreciar que el preu que més vegades apareix és 240.000 euros (Figura 34).

Com a conclusió es pot dir que l'immoble estàndard espanyol és un pis de 90 metres quadrats, amb 3 dormitoris, que es troba a la primera planta i que té un preu de 240.000 euros.

### 3.6.10.2 L'immoble estàndard andorrà

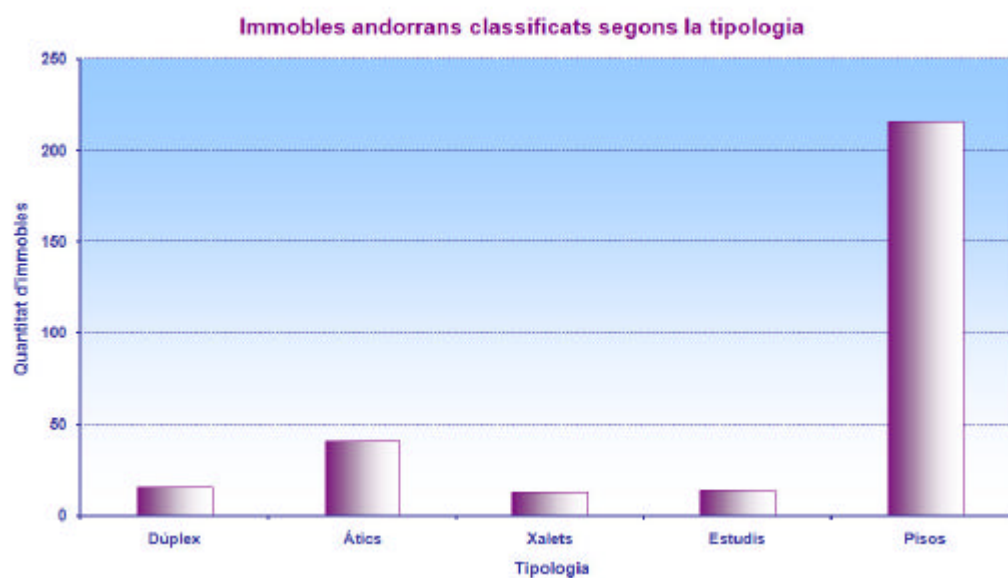


Figura 35. Immables andorrans classificats segons la tipologia

Els pisos són clarament la tipologia més abundant entre els immables andorrans (Figura 35).

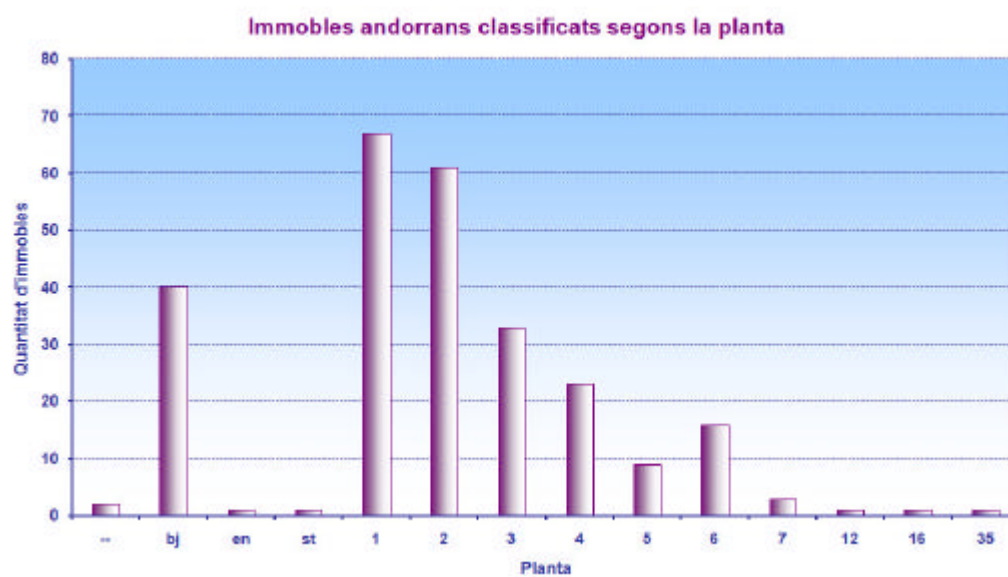


Figura 36. Immables andorrans classificats segons la planta

La majoria dels immables andorrans es troben a la primera planta encara que no hi ha gaire diferència amb els que es troben a la planta 2 (Figura 36).

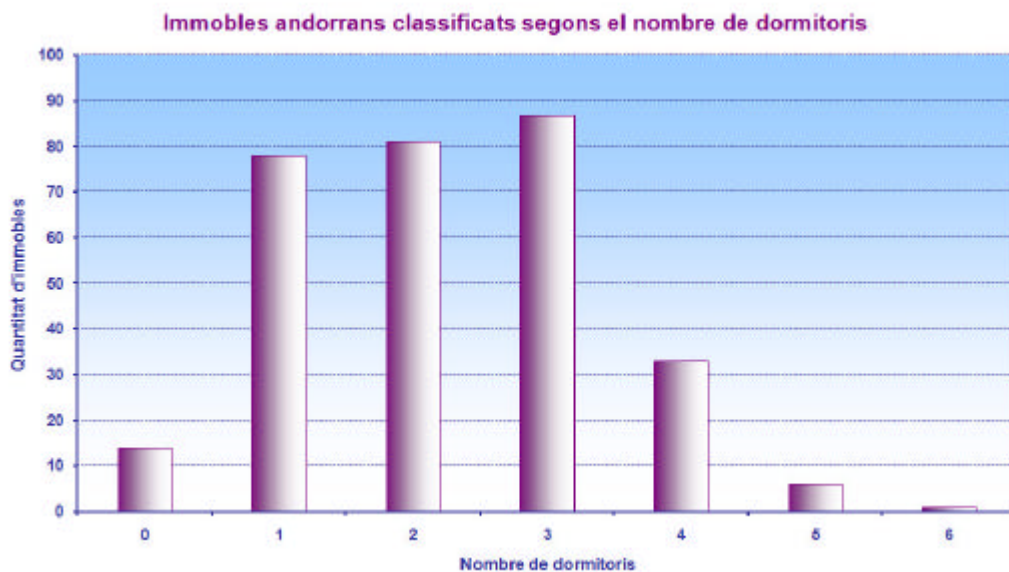


Figura 37. Immables andorrans classificats segons el nombre de dormitoris

Respecte als dormitoris tampoc s'aprecien grans diferències entre el nombre d'immables que tenen 1, 2 i 3 dormitoris però aquest últims són els més nombrosos (Figura 37).

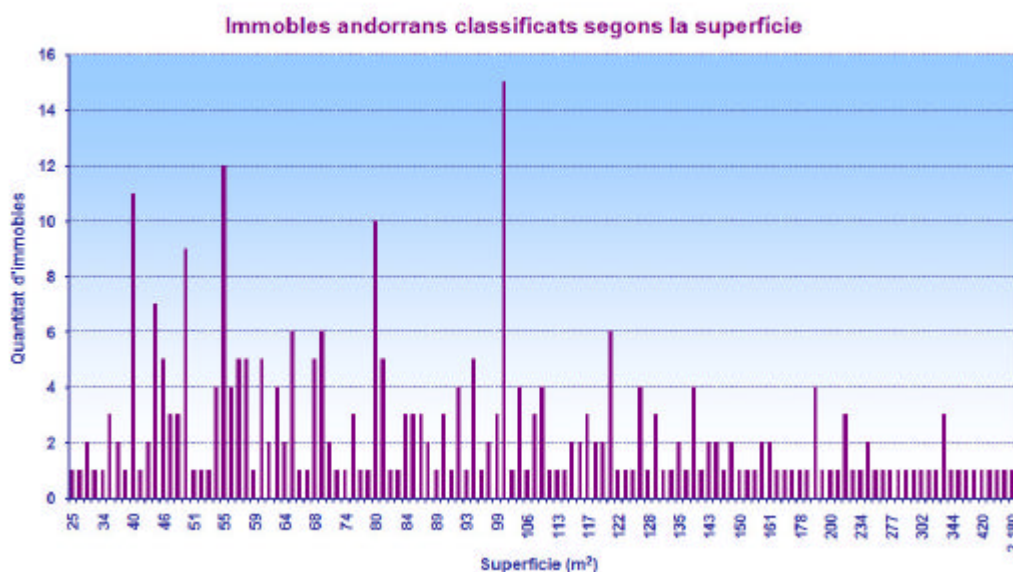


Figura 38. Immables andorrans classificats segons els metres quadrats

La superfície més habitual als immables andorrans és 100 m<sup>2</sup> (Figura 38).

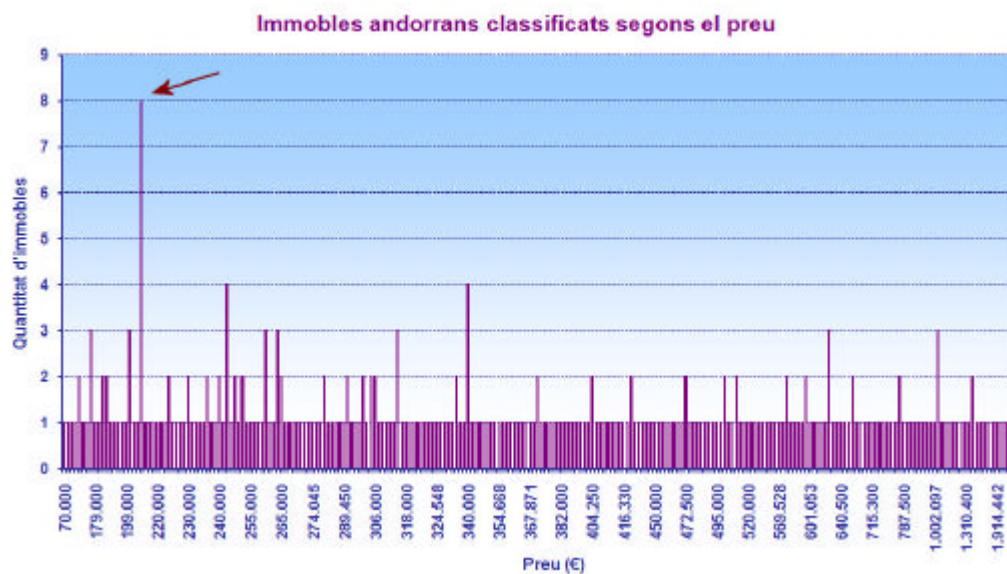


Figura 39. Immobilis andorrans classificats segons el preu

També en el cas dels preus andorrans existeix una gran diversitat però es pot apreciar que el preu que més vegades apareix és 210.000 euros.

Com a conclusió es pot dir que l'immoble estàndard andorrà és un pis de 100 metres quadrats, amb 3 dormitoris, que es troba a la primera planta i que té un preu de 210.000 euros (Figura 39).

## 4 Conclusions

1. S'ha realitzat el procés complet de creació d'un magatzem de dades, partint de l'anàlisi de requeriments, el disseny del model multidimensional i finalment la implementació del magatzem pròpiament dit.
2. Amb el magatzem de dades creat s'han pogut realitzar totes les consultes necessàries per generar els informes que han donat resposta als requeriments del client.
3. Amb unes senzilles consultes SQL s'ha pogut extreure fàcilment la informació necessària per representar de manera visual les característiques que defineixen l'immoble estàndard espanyol i andorrà. Aquests són, pel cas d'Espanya, els pisos de 90 m<sup>2</sup> amb 3 dormitoris, que es troben a la primera planta i amb un preu de 240.000 euros, i pel cas d'Andorra, els pisos de 100 m<sup>2</sup> amb 3 dormitoris, que es troben a la primera planta i amb un preu de 210.000 euros
4. S'han assolit els objectius proposats.



## 5 Línies d'evolució futures

En aquest projecte el que s'ha fet és incloure en el magatzem de dades la informació proporcionada per l'empresa immobiliària i que abastava de gener de 2006 a gener de 2008. No obstant això, per tal que l'explotació del magatzem de dades proporcioni realment informació útil i aprofitable de cara a la presa de decisions, és imprescindible actualitzar aquestes dades i proporcionar al client la formació necessària perquè pugui realitzar aquesta actualització de forma periòdica. Les actualitzacions són possible amb unes mínimes modificacions de l'*script* de càrrega de dades i d'algunes taules de la base de dades com ara la taula TEMPS on s'hauria d'afegir nous mesos i les taules de zones on, en cas de ser necessari, s'haurien d'afegir noves províncies, comunitats autònomes i fins i tot països, preveient que l'empresa expandeixi la seva oferta d'immobles a un àmbit geogràfic més ampli.

Respecte als informes, els que s'han generat en aquest projecte cobreixen totalment els requeriments plantejats per l'empresa però en previsió de que les seves necessitats puguin canviar seria adient que el client pogués crear les seves pròpies consultes. En aquest sentit es considera més usable la versió d'Oracle Discoverer que s'executa en un entorn web, Oracle Discoverer Plus i Oracle Discoverer Viewer, donat a que és un entorn més usable i familiar per a l'usuari.

## 6 Glossari

**Atribut:** Qualsevol detall que serveix per qualificar, identificar, classificar, quantificar o expressar l'estat d'una entitat.

**BI (*Business Intelligence*) Intel·ligència Empresarial:** Conjunt d'estratègies i eines orientades a l'administració i creació de coneixement mitjançant l'anàlisi de dades existents a una organització o empresa.

**cel·la:** Mesurament que fa referència al mateix esdeveniment dins una mateixa estructura

**Cel·la:** Conjunt de cel·les del mateix Fet que estan associades a instàncies del mateix Nivell per a cadascuna de les Dimensions.

**Clau forana:** Una o més columnes de la taula de una base de dades relacional que implementa una relació molt-a-molt que aquesta taula té amb altres taules o amb si mateixa.

**Clau primària:** Una o més columnes de la taula de una base de dades els valors de les quals, combinats entre si, han de ser únics dins la taula.

**Descriptor:** Atribut que es troba en un Nivell. Conté informació no jeràrquica i està definit sobre un domini discret.

**Dimensió:** Punt de vista des del que s'analitza un determinat fet.

**Estrella:** Organització de dades en una base de dades relacional. Al mig es troba la taula del Fet amb les columnes que són les mesures multidimensionals. Les branques de l'estrella són les Dimensions.

**ETL (*Extract, Transform and Load*):** Procés d'extracció, transformació i càrrega de dades que permet obtenir dades de múltiples fonts, formatjar-les i netejar-les i carregar-les en una base o magatzem de dades per a la seva posterior anàlisi.

**Fet:** Activitat de l'empresa objecte d'anàlisi.

**Granularitat:** Grandària d'un objecte respecte a un altre.

**Índex de combinació (*join index*):** Índex definit sobre una clau forana, de manera que té els valors d'una taula i apunta a una altra.



**Jerarquia d'agregació:** Conjunt de relacions entre les instàncies d'una dimensió que indica com s'agrupen unes per formar d'altres.

**Magatzem de dades (*Data warehouse*):** Sistema d'informació que reuneix la informació històrica generada pels diferents departaments d'una organització, unificada i depurada d'errors que permet la consulta d'informació de forma àgil i flexible.

**Mesura:** Atribut numèric d'una Cel·la, normalment additiu.

**Nivell:** Conjunt d'instàncies d'una Dimensió que tenen la mateixa granularitat.

**OLAP (*On-Line Analytical Processing*) Processament analític en línia:** Eines de consulta de grans quantitats de dades que utilitzen estructures multidimensionals o cubs OLAP que contenen dades procedents de bases de dades o sistemes transaccionals. Utilitzat per les empreses per millorar el coneixement sobre l'empresa i per fer informes de vendes, direcció, marketing i similars.

**PL/SQL:** Llenguatge de programació que suporta totes les consultes i manipulació de dades que es fan servir en SQL, però inclou característiques com el maneig de variables, les estructures modulars, les estructures de control de flux i presa de decisions i el control d'excepcions.

**SQL (*Structured Query Language*):** Llenguatge de consulta estructurat utilitzat per crear, modificar i recuperar dades d'un sistema de base de dades relacional.

## 7 Bibliografia consultada

Abelló Gamazo, A. (2003). Disseny multidimensional. A: Magatzems de dades i models multidimensionals (Rius Gavídia, A. i Serra Vizern, M. (coord.)). Fundació per a la Universitat Oberta de Catalunya, Barcelona. ISBN: 84-8429-437-4.

Brownbridge P.R. & Fry, N. (2002). Oracle9i Discoverer Desktop User's Guide Version 9.0.2 for Windows. Oracle Corporation

Inmon, W. H., 2002. Building the Data Warehouse (3rd Edition). John Wiley & Sons, Inc. ISBN: 0-471-08130-2

Kimball, R. & Caserta, J., 2004. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing, Inc. ISBN 0-7645-7923 -1

Kimball, R. & Ross, M., 2002. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition). Wiley Computer Publishing. ISBN 0-471-20024-7

Oracle Corporation (2003). Oracle Discoverer Administrator Tutorial, 10g (9.0.4).

Pribyl B. & Feuerstein S. (2001). Learning Oracle PL/SQL (First Edition). O'Reilly, 424 pages. ISBN: 0-596-00180-0.

Data Warehousing Processes. Data de consulta: 15/3/08.

<http://www.1keydata.com/datawarehousing/processes.html>.

Welcome to the Oracle Database 10g Express Edition Tutorial!. Data de consulta: 22/4/08.

<http://st-curriculum.oracle.com/tutorial/DBXETutorial/index.htm>.

Welcome to the Oracle SQL Developer Tutorial!. Data de consulta: 23/4/08.

<http://st-curriculum.oracle.com/tutorial/SQLDeveloper/index.htm>.

## 8 Annexos

### 8.1 Annex I - Instruccions per la restauració del sistema

Els fitxers necessaris per la restauració del sistema es troben a la carpeta producte.

- ★ Per a la restauració de la base de dades és necessari importar el fitxer bd\_silvia.dmp executant: "oracle\app\oracle\product\10.2.0\server\BIN\imp.exe".
- ★ L'àrea de negoci es pot importar des d'Oracle BI Discoverer Administrator mitjançant: "Archivo : Importar : Agregar archivo" i obrint l'arxiu TFC-Sostre per a tothom.eex
- ★ El llibre de treball es pot obrir des d'Oracle BI Discoverer Desktop mitjançant: "Archivo : Abrir : Mi Computadora : Abrir" i obrint l'arxiu Sostre per a tothom - workbook.DIS