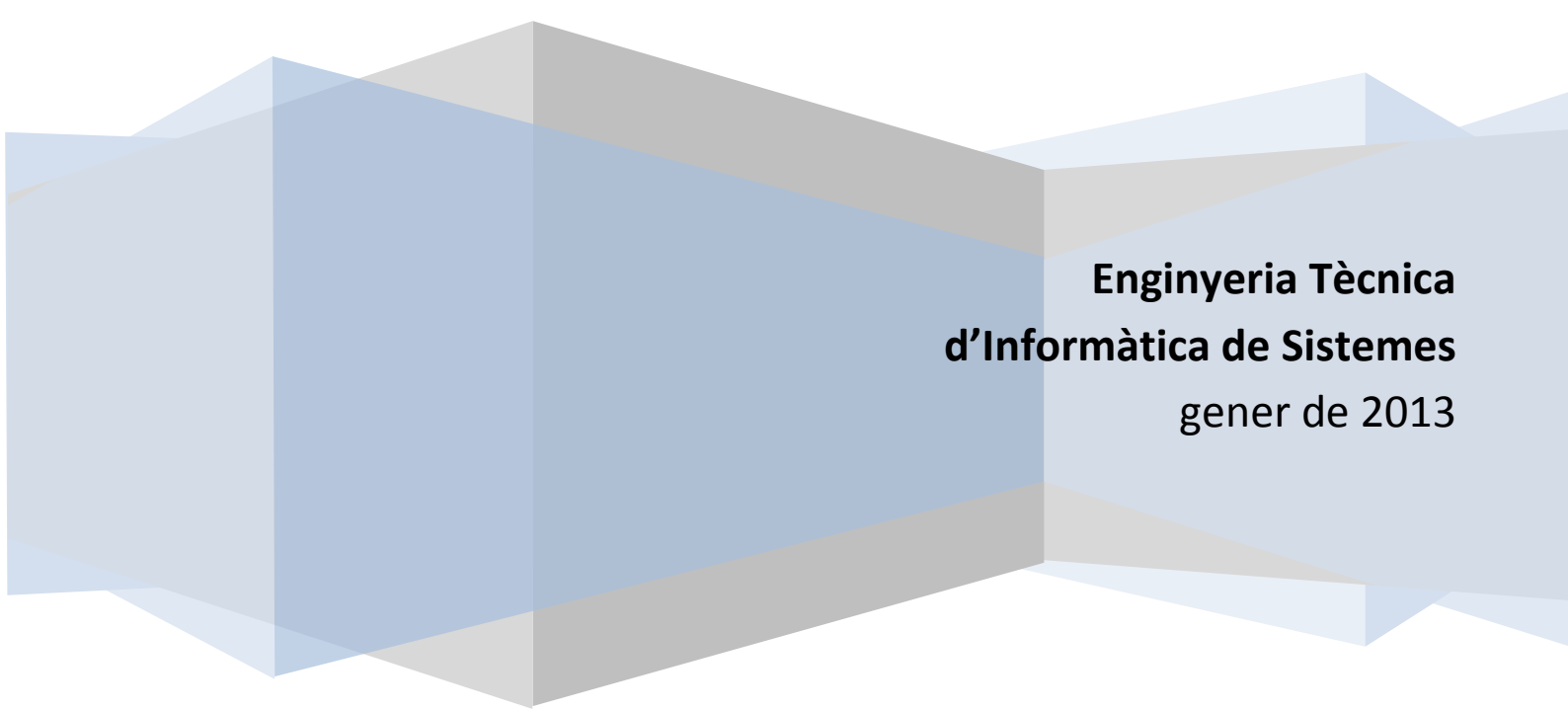


Construcció i explotació d'un magatzem de dades per a l'anàlisi de vendes d'una cadena de supermercats

Memòria del Treball Final de Carrera

Jesús Segura del Olmo

Tutor: Carles Llorach Rius



**Enginyeria Tècnica
d'Informàtica de Sistemes**
gener de 2013

Dedicatòria i agraïments

Abans de res, voldria agrair i dedicar aquest treball a la meva família i amics, als *stöuts* i a la meva parella, Júlia. Per la seva ajuda, l'eterna paciència i els ànims rebuts .

Resum

El Grup Líder en Distribució de Proximitat (GLDP) va decidir fer un estudi a Catalunya per determinar la continuïtat de la seva xarxa d'establiments i conèixer millor el comportament de les vendes per tal de capgirar-ne l'evolució negativa dels darrers anys.

És per això que GLDP va decidir encarregar-nos, com a consultoria externa independent, la creació d'un magatzem de dades, mitjançant el qual es pogués obtenir una sèrie d'informes on es pogués veure l'evolució trimestral de les vendes mitjançant una eina de Business Intelligence: Pentaho Report Designer.

Els passos que vam seguir per aconseguir el resultat final van ser:

- Estudiar les dades inicials. Aquestes es dividien en arxius de tres tipus: uns .csv amb la informació sobre els productes, un full d'excel amb dades sobre els establiments i una sèrie d'arxius d'accés amb les dades sobre les vendes i les vendes al detall.
- Creació d'un pla de treball que ens assegurés aconseguir les nostres fites dins el temps estipulat.
- Realitzar el model conceptual i el multidimensional sobre els quals crear els dissenys físics i lògics.
- Procés ETL - Extracció, transformació i càrrega de dades.
- Creació de les taules finals amb l'eina Mysql Workbench.
- Exportar el cub a Pentaho Report Designer per a la realització dels informes.

Paraules clau:

TFC, Ad-hoc, Base de Dades, ETL, Magatzem de dades, Data Warehouse, Olap, Mysql, Pentaho, GLDP, Establiment.

Nom de l'àrea del Treball Final de Carrera:

Magatzems de dades.

Índex de continguts

1. Introducció	7
1.1. El projecte.....	7
1.2 Justificació del TFC.....	7
1.3 Objectius.....	8
1.3.1 Objectius concrets del projecte	8
1.4 Enfocament i metodologia	9
1.5 Planificació temporal.....	10
1.5.1 Descripció de tasques.....	10
1.5.2 Calendari previst	10
1.5.3 Diagrama de Gantt	12
1.6 Productes obtinguts	13
1.7. Breu descripció dels altres capítols de la memòria.....	13
2. Anàlisi	14
2.1 Descripció	14
2.2 Informes a realitzar	14
2.3 Dades d'origen	15
2.4 Diagrama de casos d'ús.....	18
2.5. Elements per a l'anàlisi multidimensional	19
2.6. Especificacions tècniques.....	20
2.7 Anàlisi conceptual	20
2.7.1 Identificar els fets.....	20
2.7.2 Granularitat	21
2.7.3 Escollir les dimensions.....	21
2.7.4 Atributs de les dimensions	21
2.7.5 Distingir descriptors de jerarquies	21
2.7.6 Mesures.....	22
2.7.7 Cel·les	22
2.7.8 Restriccions d'integritat	22
2.7.9 Viabilitat	22
3. Disseny.....	24
3.1. Descripció	24
3.2 Disseny lògic	24
3.3 Disseny físic	25
3.4 Procediments ETL	27
3.4.1 Extracció	28
3.4.2 Transformació	28
3.4.3 Càrrega	29

4. Implementació	31
4.1 Descripció	31
4.2 Accés a la Base de Dades.....	31
4.3 Accés a l'eina Business Intelligence	34
4.4 Informes:	36
4.4.1 Total de vendes i marge net del grup:	36
4.4.2 Import mitjà de venda per soci i establiment	38
4.4.3 Percentatge de vendes respecte el total del grup	40
4.4.4 Preus màxims i mínims per tipologia d'establiment i característiques del producte.	42
4.4.5 Rànquing d'establiments per número de vendes i volum total.....	44
4.4.6 Top ten productes	46
4.4.7 Percentatge de marques blanques per habitant	48
4.4.8 Categorització clients ABC.....	50
4.4.9 Anàlisi de compra per impuls.....	52
4.4.10 Distribució estacional vendes	54
5. Conclusions.....	56
6. Línies de futur.....	57
7. Glossari	58
8. Bibliografia.....	60
8. 1 Documents	60
8.2 Llibres de text	60
8.3 Enllaços d'interès a Internet	60

Índex de figures

Figura 1. Taula fites semestre	11
Figura 2. Taula resum calendari	12
Figura 3. Diagrama de Gantt	12
Figura 4. Fitxer proporcionat per GLDP.....	15
Figura 5. Arxiu d'establiments.....	15
Figura 6. Arxius de venda i venda al detall.....	16
Figura 7. Arxius de producte per any	17
Figura 8. Diagrama de casos d'ús	19
Figura 9. Diagrama model conceptual	23
Figura 10. Esquema disseny lògic.....	25
Figura 11. Disseny físic: taula establiment.....	25
Figura 12. Disseny físic: taula producte	26
Figura 13. Disseny físic: taula venda	26
Figura 14. Disseny físic: taula client	27
Figura 15. Disseny físic: taula demarcació	27
Figura 16. Disseny físic: taula data	27
Figura 17. Càrrega arxius productes	30
Figura 18. Càrrega arxius vendes i vendes al detall	30
Figura 19. Taula usuaris i contrasenyes	31
Figura 20. Login base de dades	32
Figura 21. Esquemes MySQL Workbench	32
Figura 22. Establiment connexió amb Pentaho Report Designer	34
Figura 23. Disseny de la consulta	35
Figura 24. Informe 1: total de vendes i marge net del grup	36
Figura 25. Informe 2: import mitjà de venda per soci i establiment	38
Figura 26. Informe 3: Percentatge de vendes respecte el total del grup	40
Figura 27. Informe 4: preus màxims i mínims per tipologia d'establiment i característiques del producte	42
Figura 28. Informe 5: rànkning d'establiments per número de vendes i volum total.....	44
Figura 29. Informe 6: top ten productes.....	46
Figura 30. Informe 7: percentatge de marques blanques per habitant	48
Figura 31. Informe 8: categorització clients ABC	50
Figura 32. Informe 9: Anàlisi de compra per impuls	52
Figura 33. Informe 10: Distribució estacional vendes.....	54

1. Introducció

1.1. El projecte

Els magatzems de dades són una col·lecció de dades organitzades per a donar suport a processos de presa de decisions. Les seves característiques són les següents: estan orientats al tema (l'estructura es divideix per temes d'interès enlloc de fer-ho segons una funcionalitat), les dades han de passar per un procés d'integració i, a més, la informació és històrica i no volàtil (per garantir que podrem veure quan ocorre un succés i que no perdrem cap dada, ni que sigui errònia).

En la construcció dels DW (magatzems de dades) el punt més important que s'ha de tenir en compte és l'obtenció d'informació necessària, ja que amb aquesta podrem desenvolupar noves estratègies que permetin millorar els resultats de l'empresa.

1.2 Justificació del TFC

En el Treball de Final de Carrera (TFC) hem tractat el cas particular del Grup Líder en Distribució de Proximitat (GLDP) que ens demanava crear un DW capaç d'analitzar informació compresa en un conjunt d'arxius en formats diferents (CSV, XLS i MDB). Aquestes dades, a més, s'havien de complementar amb altres que hem hagut de buscar externament, com ara el nombre d'habitants per zona, en pàgines com la de l'IDESCAT.

El principal problema al qual s'enfrontava el GLDP era que explotar la informació dels diferents arxius resultava costós i encara més comparar les dades de diferents punts de venda. Per això, construir el DW havia de permetre emmagatzemar la informació dels diferents establiments, a més de poder-la creuar entre si.

Un cop finalitzat el TFC, el client ha obtingut un DW, segons definició de Bill Inmon:

- **Orientat a temes.** - Les dades a la base de dades estan organitzades de manera que tots els elements de dades relatives al mateix esdeveniment o objecte del món real quedin unides entre si.

- **Variants en el temps.** - Els canvis produïts en les dades al llarg del temps queden registrats perquè els informes que es puguin generar reflecteixin aquestes variacions.
- **No volàtil.** - La informació no es modifica ni s'elimina.
- **Integrat.** - La base de dades conté les dades de tots els sistemes operacionals de l'organització, i aquestes dades han de ser consistents.

1.3 Objectius

En aquest projecte hem mirat d'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, així com la utilització d'eines que faciliten la generació d'informes.

Bàsicament, durant el TFC els punts clau a assolir han estat els següents:

- Bon coneixement sobre el món dels DW i el seu funcionament i construcció.
- Aprendre a agafar i analitzar dades des del DW i comprovar el seu funcionament en la presa de decisions.
- Conèixer el funcionament dels programes i entorns utilitzats.
- Desenvolupar una presentació multimèdia amb els resultats del projecte.

1.3.1 Objectius concrets del projecte

Des d'un punt de vista més pràctic, els objectius genèrics del projecte s'han concretat en els següents:

Importació de les dades

Hem creat una sèrie de taules temporals dins el sistema de gestió de bases de dades MySQL. Aquí és on hem afegit tota la informació inicial proporcionada per GLDP. En aquest procés hem hagut de corregir totes les errades i inconsistències; a més, també hem donat formats similars a les dades que posteriorment havien d'anar a les taules definitives del magatzem de dades.

Creació del magatzem de dades

A partir de la informació donada per GLDP, hem dissenyat un magatzem de dades que contingui totes les taules necessàries per a poder desenvolupar posteriorment els informes.

Generació d'informes

Un cop tota l'aplicació va estar desenvolupada, aquesta havia de ser capaç de generar uns informes automatitzats que mostressin la informació sol·licitada en cada cas.

1.4 Enfocament i metodologia

Per a desenvolupar el nostre magatzem de dades hem descomposat el treball en diferents fases. Aquestes es corresponen a un model en cascada genèric.

1. Anàlisi de requeriments.

En aquesta fase hem recollit tota la informació de GLDP per a fer-nos una idea de la magnitud del projecte i poder començar a desenvolupar les idees prèvies.

2. Disseny del sistema.

Hem creat els models de dades físics, lògics i conceptuals per a deixar definides totes les estructures que ens facilitaran la creació de la base de dades.

3. Creació de la base de dades.

Hem treballat amb Mysql Workbench per a crear una sèrie de taules on poder afegir posteriorment totes les dades

4. Extracció, Transformació i Càrrega.

En aquesta fase hem dut a terme la fase ETL amb la qual hem carregat totes les dades, ja tractades, a la nostra base de dades.

5. Creació del cub.

Un cop les dades estan es disposició de ser treballades hem pogut crear el cub del qual partiran els informes que realitzarem amb posterioritat.

6. Realització dels informes.

En aquesta fase hem creat els informes demanats amb l'eina Pentaho Report Designer.

7. Manteniment.

Amb tots els passos enllestits només ens manca fer feines de manteniment.

1.5 Planificació temporal

Ara podem veure un llistat de les tasques que s'han realitzat durant el semestre i com es relacionen amb la memòria del projecte.

1.5.1 Descripció de tasques

- **TASCA 1.** Estudi d'abast: Hem definit els objectius a desenvolupar i, posteriorment, hem planificat totes les activitats del semestre.
- **TASCA 2.** Estudi del DW: Hem efectuat la introducció als DW i la redacció del "Capítol 1. Introducció als Magatzems de Dades"
- **TASCA 3.** Estudi del Programari: Ens hem endinsat en el programari del TFC, centrant-nos en MySQL WorkBench. També hem redactat el "Capítol 2. Software i entorn de treball".
- **TASCA 4.** Construcció del DW: Ens hem centrat en els punts vistos anteriorment dins el "Capítol 3. Construcció del DW", treballant per la integritat de les dades, l'emmagatzematge i l'explotació de dades, etc.
- **TASCA 5.** Anàlisi de dades: Hem discutit sobre la interpretació que ens proporciona la informació extreta del DW, i aquesta tasca s'ha vist reflectida en el "Capítol 4. Informes".
- **TASCA 6.** Acabament del TFC: Hem realitzat els últims retocs al TFC, hem escrit els agraïments, conclusions, resums, etc. Participació al debat del fòrum.

1.5.2 Calendari previst

A més de les tasques a realitzar, el següent calendari ens mostra un seguit de fites a superar durant el semestre:

Dia	Mes	Fita
10	Setembre	Inici TFC – DW
19	Setembre	Inici PAC 1
2	Octubre	Lliurament PAC 1
3	Octubre	Inici PAC 2
6	Novembre	Lliurament PAC 2
7	Novembre	Inici PAC 3
19	Desembre	Lliurament PAC 3
20	Desembre	Inici LLIURAMENT FINAL
7	Gener	Lliurament del LLIURAMENT FINAL

Figura 1. Taula fites semestre

Vam preveure que hauríem de dedicar entre 15 i 20 hores setmanals al TFC per arribar a unes 300 hores estimades de feina.

Setmana	Fita	Tasca
0		Revisió material, apunts, etc.
1		Inici TASCA 1
	1	Inici TFC
	2	Inici PAC 1
2		Acabar TASCA 1 i PAC 1
3	3	Lliurament PAC 1
	4	Inici PAC 2
		Inici TASCA 2
4		Acabar TASCA 2
5		Inici TASCA 3 / Acabar TASCA 3
6		Inici TASCA 4
7		Continuar TASCA 4
8	5	Lliurament PAC 2
	6	Inici PAC 3
		Continuar TASCA 4

9		Continuar TASCA 4
10		Continuar TASCA 4
11		Continuar TASCA 4
12		Acabament TASCA 4
13		Inici TASCA 5
14	7	Lliurament PAC 3
	8	Inici LLIURAMENT FINAL
		Acabar TASCA 5
15		Inici TASCA 6
16		Acabar TASCA 6
17	9	Lliurament LLIURAMENT FINAL

Figura 2. Taula resum calendari

1.5.3 Diagrama de Gantt

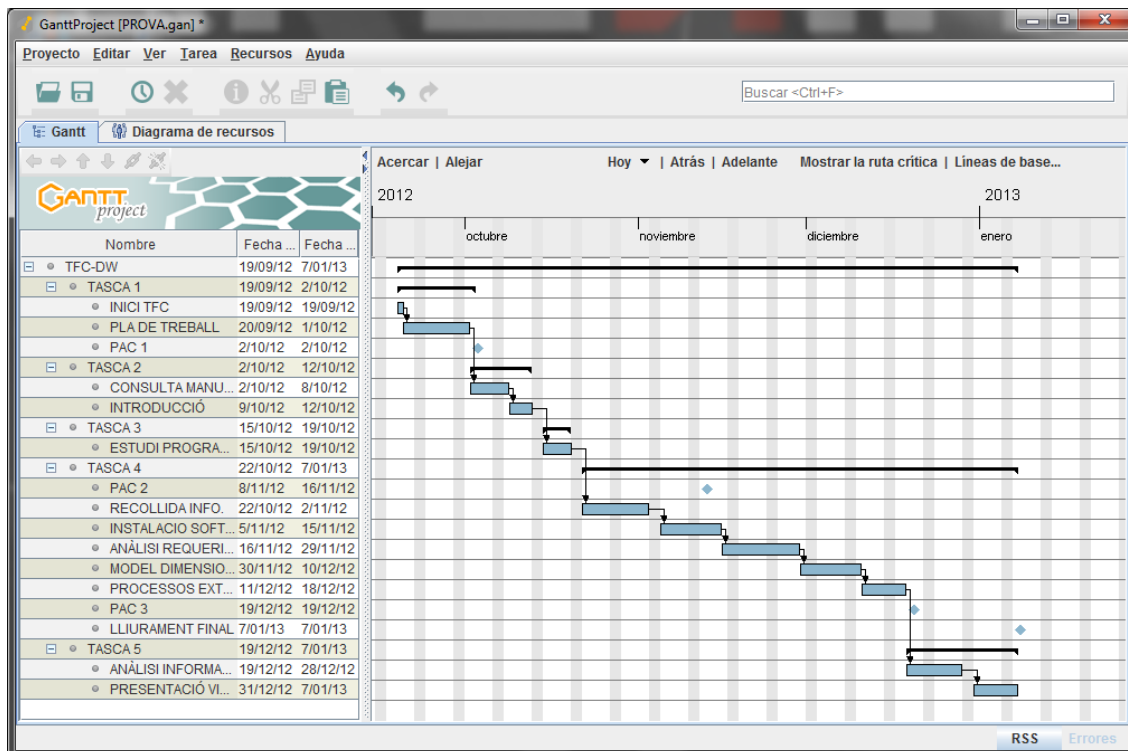


Figura 3. Diagrama de Gantt

1.6 Productes obtinguts

Un cop acabada tota la feina del magatzem de dades, hem obtingut els següents productes:

- Pla de treball
- Anàlisi de requeriments
- Disseny del magatzem de dades
- Magatzem de dades (Màquina virtual)
- Informes (Màquina virtual)
- Presentació Virtual

1.7. Breu descripció dels altres capítols de la memòria

Anàlisi i disseny

En aquest apartat tractarem l'anàlisi de requeriments de l'usuari GLDP i veurem el disseny conceptual, lògic i físic per a la creació del cub.

Implementació.

Tractarem el disseny final de la base de dades i la implementació d'aquesta, juntament amb la creació de la base de dades i el procés ETL. A més, podrem veure les captures sobre els informes realitzats.

2. Anàlisi

2.1 Descripció

En els apartats anteriors hem definit com organitzarem i enfocarem el volum de treball de tot el projecte. Hem tractat els objectius i l'estructura, hem organitzat les tasques i fites a complimentar i, a més, hem parlat sobre els materials necessaris per a dur a terme el projecte.

En aquest capítol, podrem veure els requeriments funcionals i no funcionals on parlarem dels diagrames de casos d'ús i del maquinari i programari necessaris. Aquest punt ha condicionat molt la BBDD, ja que parteix directament de la informació facilitada pel client.

2.2 Informes a realitzar

Des de el GLDP ens van encarregar els següents informes per a realitzar:

- Total vendes i marge net del grup
- Import mitjà de compra per soci i establiment
- Percentatge de vendes respecte el total del grup (per establiment i per demarcació)
- Preus màxims i mínims per tipologia d'establiment i característiques del producte.
- Rànquing d'establiments per nombre de vendes i volum total
- "Top ten" de productes
- Percentatge de vendes de "marques blanques" per habitants
- Categorització de Clients A/B/C
- Anàlisi de compra per impuls
- Distribució setmanal i estacionalitat de vendes

2.3 Dades d'origen

Inicialment se'ns va proporcionar un fitxer comprimit amb els següents arxius :

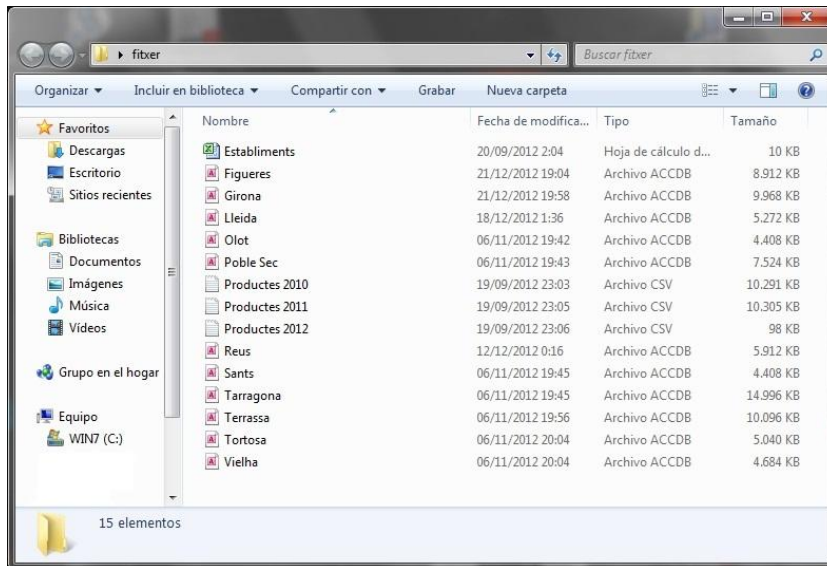


Figura 4. Fitxer proporcionat per GLDP

Vam poder catalogar aquests arxius en tres tipus:

Arxiu d'establiments:

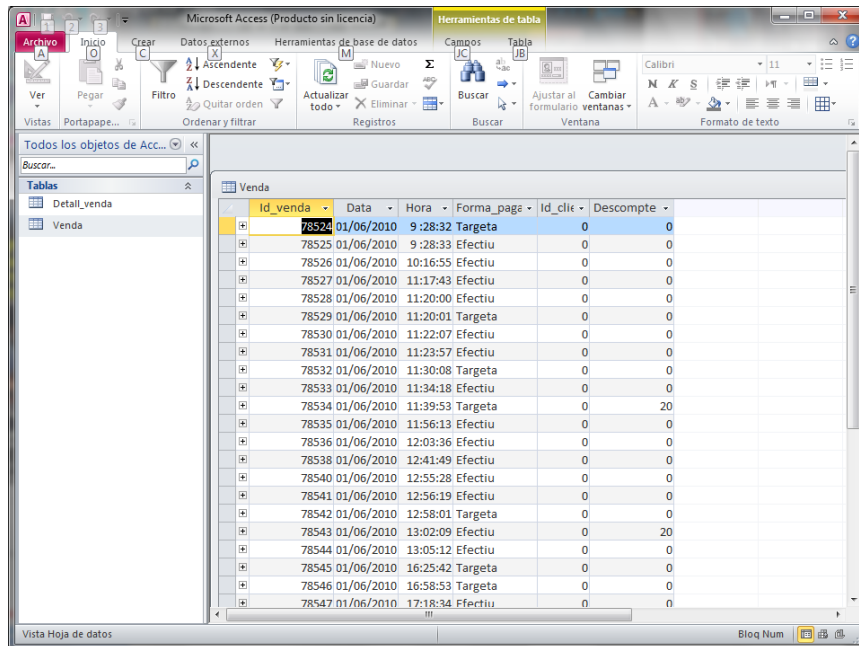
Un petit arxiu d'excel on hi ha tota la informació referent als establiments: nom, superfície, dia del soci...

	A	B	C	D	E	F	G	H	I	J	K	L
1	Establiment		GLDP FIGUERES									
2		Superfície	380m2									
3		Dia del soci	Dimarts									
4		Cost fix mensual	86.000 €									
5		Tipologia	Petit supermercat									
6												
7	Establiment		GLDP GIRONA									
8		Superfície	648m2									
9		Dia del soci	Dijous									
10		Cost fix mensual	145.000 €									
11		Tipologia	Supermercat									
12												
13	Establiment		GLDP LLEIDA									
14		Superfície	490m2									
15		Dia del soci	Dimecres									
16		Cost fix mensual	123.000 €									
17		Tipologia	Supermercat									
18												
19	Establiment		GLDP OLOT									
20		Superfície	286m2									
21		Dia del soci	Dilluns									
22		Cost fix mensual	75.400 €									
23		Tipologia	Petit supermercat									
24												

Figura 5. Arxiu d'establiments

Arxius de venda i venda al detall

Són una sèrie d'arxius d'access amb les dades referents a les vendes i vendes detallades d'un establiment en un període de temps. Es establiments a tractar són els de: Girona, Lleida, Olot, Poble Sec, Reus, Sants, Tarragona, Terrassa, Tortosa i Vielha. Tot i que també s'ha inclòs el de Figueres degut a una inconsistència entre l'arxiu de vendes i vendes al detall ha quedat fora de l'anàlisi.



Id_venta	Data	Hora	Forma_pago	Id_cliente	Descompte
78524	01/06/2010	9:28:32	Targeta	0	0
78525	01/06/2010	9:28:33	Efectiu	0	0
78526	01/06/2010	10:16:55	Efectiu	0	0
78527	01/06/2010	11:17:43	Efectiu	0	0
78528	01/06/2010	11:20:00	Efectiu	0	0
78529	01/06/2010	11:20:01	Targeta	0	0
78530	01/06/2010	11:22:07	Efectiu	0	0
78531	01/06/2010	11:23:57	Efectiu	0	0
78532	01/06/2010	11:30:08	Targeta	0	0
78533	01/06/2010	11:34:18	Efectiu	0	0
78534	01/06/2010	11:39:53	Targeta	0	20
78535	01/06/2010	11:56:13	Efectiu	0	0
78536	01/06/2010	12:03:36	Efectiu	0	0
78538	01/06/2010	12:41:49	Efectiu	0	0
78540	01/06/2010	12:55:28	Efectiu	0	0
78541	01/06/2010	12:56:19	Efectiu	0	0
78542	01/06/2010	12:58:01	Targeta	0	0
78543	01/06/2010	13:02:09	Efectiu	0	20
78544	01/06/2010	13:05:12	Efectiu	0	0
78545	01/06/2010	16:25:42	Targeta	0	0
78546	01/06/2010	16:58:53	Targeta	0	0
78547	01/06/2010	17:18:34	Efectiu	0	0

Figura 6. Arxius de venda i venda al detall

Dins l'arxiu s'observa:

- Venda:
 - Id_venta: Nombre que identifica una venda determinada.
 - Data: Dia, mes i any en el que s'ha efectuat la compra.
 - Hora: Hora en la que s'ha efectuat la compra.
 - Forma_pagament: Si s'ha efectuat la compra amb targeta o en efectiu.
 - Id_cliente: Identificació del client (en cas que ho sigui) que ha efectuat la compra
 - Descompte: Valor del descompte en la venda.
- Venda detall:
 - Id_venta: Nombre que identifica una venda determinada.

- Id_línia: Ordre de venda dels productes
- Unitats: Nombre d'unitats del mateix producte comprat.
- Id_producte: Identificador del producte que participa en la transacció.

Arxius de producte per any.

Tres arxius .csv repartits entre els anys 2010, 2011 i 2012 amb un llistat de productes que poden ser venuts.

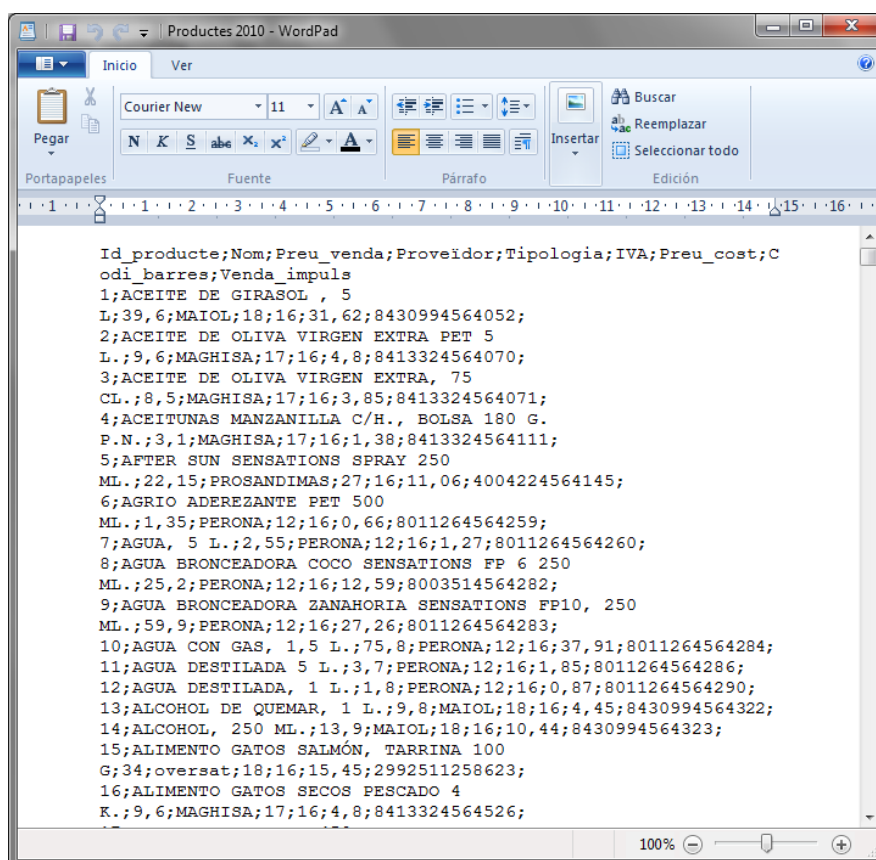


Figura 7. Arxius de producte per any

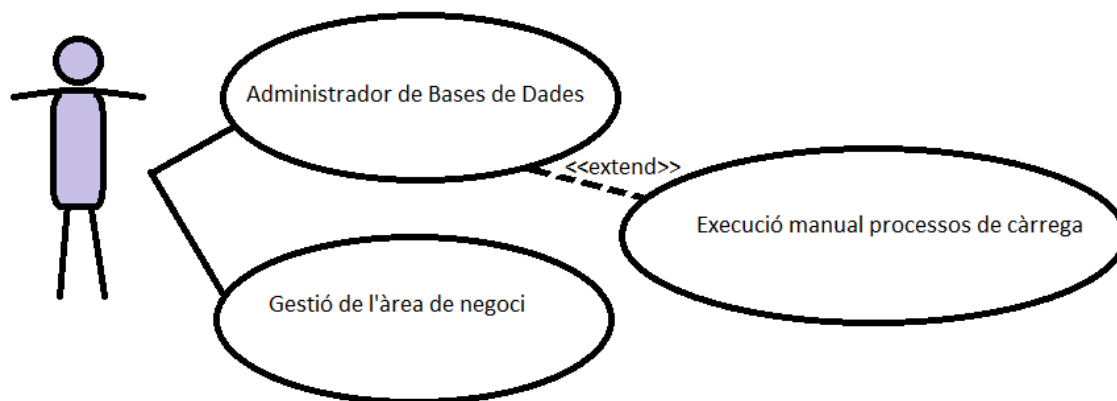
Dins d'aquests arxius podem observar els següents camps:

- Id_producto: Identificador del producte en un any determinat
- Nom: Nom del producte.
- Preu_venta: Preu de venda al públic
- Proveedor: Proveïdor que distribueix el producte

- Tipologia: Tipus de producte.
- IVA: Impost relacionat amb el producte
- Preu_cost: Preu de cost per el qual ha de pagar l'establiment.
- Codi_barres: Identificador del producte en forma de codi de barres.
- Venda_impuls: Valor que indica si un producte està considerat de compra per impuls, el qual vol dir, normalment que es petit i de poc preu: caramels, bolígrafs...

2.4 Diagrama de casos d'ús

Administrador



Tècnic GLDP



Analista GLDP

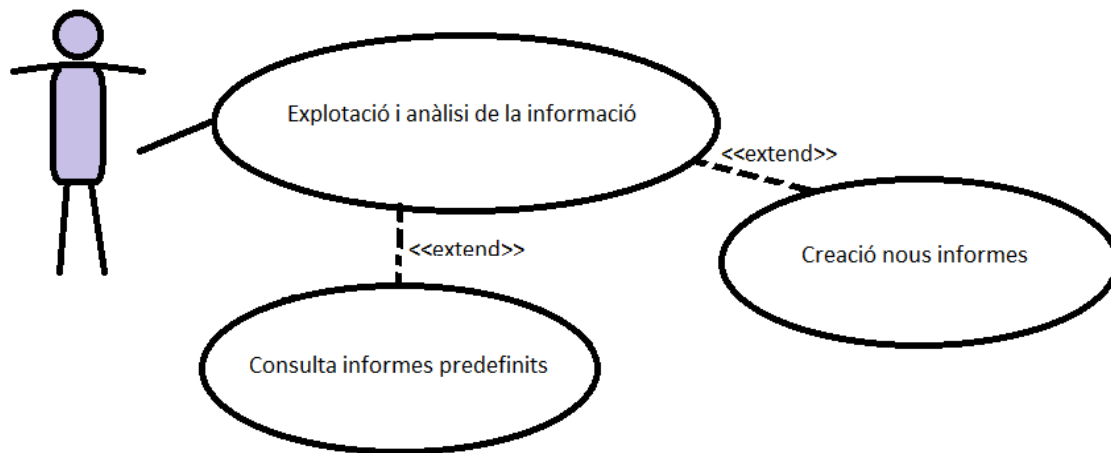


Figura 8. Diagrama de casos d'ús

2.5. Elements per a l'anàlisi multidimensional

Gràcies a les dades proporcionades pel GLDP podem fer l'anàlisi multidimensional:

- **Fets:** Podem considerar com a fet les dades d'una venda en una data determinada.
- **Mesures:** La mesura del fet serà el preu de venda i de cost, que estarà en els camps:
 - Preu_venda.
 - Preu_cost.
- **Dimensions:** Podem considerar 5 dimensions:
 - Data: Quan s'efectua la venda.
 - Demarcació: Ubicació de l'establiment on s'efectua la venda.
 - Producte: Objecte posat en venda.
 - Establiments: Emplaçaments on s'efectua la venda.
 - Client: Persona que efectua una compra, pot tenir un identificador propi de la botiga que el converteix en soci.
 - Descriptors: Per a les nostres dimensions tenim els següents descriptors.
 - Data: Any.
 - Demarcació: Zona, província.

2.6. Especificacions tècniques

El nostre projecte està realitzat sobre un sistema operatiu Windows XP amb 512MB de memòria RAM. Tot seguit descrivim el programari necessari per a realitzar-lo:

- Microsoft Access: Programa que ens ha permès d'obrir els arxius .ACCDB i fer la migració a Mysql mitjançant una connexió ODBC.
- Mysql-connector-ODBC-5.1.11: Driver utilitzat per a la migració dels arxius d'access.
- Microsoft Excel: Eina utilitzada per a obrir l'arxiu Establiments.
- Sistema de gestió de base de dades Mysql: És el sistema de gestió de bases de dades escollit per a treballar el nostre projecte.
- Mysql Workbench: Programa on hem creat les taules temporals i definitives.
- Pentaho Report Designer: Eina empleada per a realitzar els informes

2.7 Anàlisi conceptual

En aquest apartat veurem els passos a seguir per a obtenir l'estrella que modelitza el disseny en el qual s'han estructurat les dades de partida que s'han recollit dels fitxers proporcionats pel GLDP.

S'han hagut de seguir diverses fases per elaborar el disseny conceptual. Així doncs, per a crear el model en estrella en el qual estructurarem les dades, primerament s'han identificat aquestes en fets, dimensions i atributs. Per això mateix, hem necessitat definir un cub OLAP per a poder crear el magatzem de dades multidimensional. Dins aquest cub tenim tota la informació necessària organitzada en dimensions, i les dades, en fets.

2.7.1 Identificar els fets

La taula de fets és la taula principal dins el magatzem de dades. Cada fet el podríem veure com les vendes i vendes detallades per cada establiment en una localitat determinada en una data concreta. Aquest seria el centre de l'estrella.

2.7.2 Granularitat

Entenem com a granularitat l'especificitat en la qual es defineix un nivell de detall dins una taula. Calia vigilar-la, ja que amb una granularitat baixa podríem retardar les consultes o perdre operativitat en els reports. Per altra banda, amb una molt alta podríem perdre massa detall en les dades. En el nostre cas, la granularitat donada ens és correcta.

2.7.3 Escollir les dimensions

Les dimensions que ens han definit l'hipercub són les següents:

- Data: L'any en el que treballem.
- Demarcació: La localització de l'establiment en la que es desenvolupa la consulta o informe.
- Client: Persona que efectua una venda.
- Establiment: Ubicació on s'efectua la venda.
- Producte: Part que s'ofereix en la venda.

2.7.4 Atributs de les dimensions

Un cop definides les dimensions calia trobar els atributs d'aquestes dimensions.

- La dimensió Data està compresa entre l'any 2005 i 2011.
- La dimensió Demarcació està composta per les comunitats autònomes espanyoles.

2.7.5 Distingir descriptors de jerarquies

Dins els atributs que hi ha en una dimensió podem distingir dos classes: d'una banda, tenim els que s'utilitzen per a agrupar i, de l'altra, els que serveixen únicament per a seleccionar. D'aquí hem extret quins s'estableixen com a descriptors, permetent fer seleccions basades ens els seus valors, i quins possibiliten fer agrupacions formant jerarquies.

- En la dimensió Data hem pogut donar una jerarquia d'agregació que vagi en: setmanes, mesos, estacions i anys.
- En la dimensió Demarcació hem determinat una jerarquia donada per 2 nivells: província i zona.

2.7.6 Mesures

En el nostre cas particular hem pogut contemplar com a mesura:

- Preu.
- Data.
- Descompte.

2.7.7 Cel·les

En aquest pas hem definit quines de les possibles cel·les que s'extreuen de l'anàlisi poden ser considerades interessants per al nostre model i, per tant, han de ser emmagatzemades i quines no. En el nostre cas, hem emmagatzemat les vendes i les vendes al detall.

2.7.8 Restriccions d'integritat

En aquest pas s'han definit les bases i s'han establert les restriccions d'integritat. El conjunt inicial de dimensions, que hem trobat després de definir el grànul, havia de donar lloc a una Base. Substituint dimensions en aquesta, segons les dependències funcionals que hi ha amb la resta de dimensions, s'han obtingut les altres bases de l'espai.

2.7.9 Viabilitat

Un cop finalitzat l'estudi conceptual havíem d'estudiar la viabilitat del disseny. En el nostre cas, es tractava de 11 demarcacions amb uns 30.000 registres es vendes i 150.000

aproximadament en vendes detallades. Dins un registre de 256 Bytes estaríem parlant per una banda de 0,26 MB i pel segon cas de 1,3 MB. Unes xifres assolibles

Un cop finalitzats tots els passos anteriors, el resultat ha estat el següent:

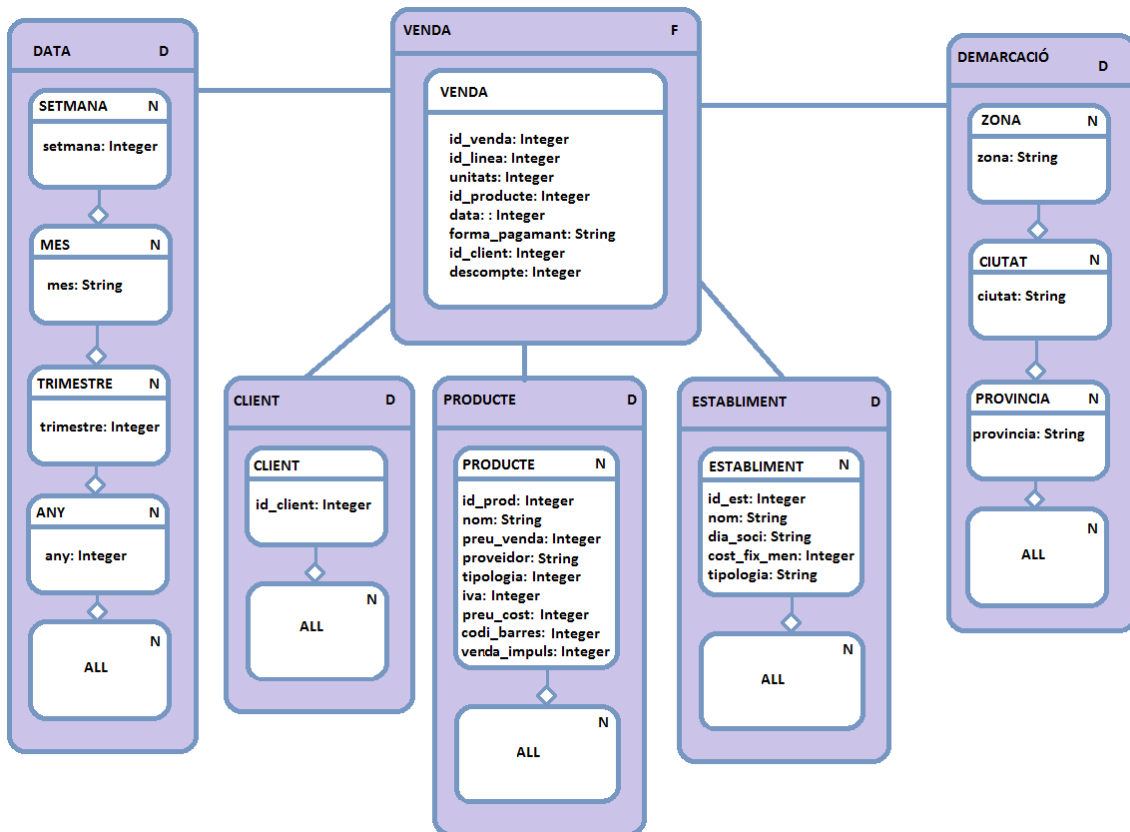


Figura 9. Diagrama model conceptual

3. Disseny

3.1. Descripció

En aquest apartat, definirem dos nivells de disseny: lògic i físic. En el primer, depurarem el model conceptual anteriorment tractat. Per últim, al model físic, donarem forma a les taules que finalment seran poblades a la base de dades.

Per últim veurem el procés ETL (extreure, transformar i carregar): aquest procés ens permetrà moure les dades des de els fitxers, reformatejar-los i netejar-los per finalment carregar-los a la BBDD. Dins aquest pas també tractarem els possibles errors puntuals que ens trobarem en el nostre cas particular.

3.2 Disseny lògic

Un cop acabat el model conceptual, tocava passar al disseny lògic. Aquesta vegada, hem continuat des de l'anàlisi de l'estrella definida en els passos anteriors i, a partir d'ella, el nostre disseny lògic havia de crear una taula per al Fet i una altra per cadascuna de les dimensions que composaven el nostre model. Posteriorment, quan ja estaven definides les taules, s'havien de lligar aquestes entre si afegint claus primàries de les dimensions a les taules de Fet en forma de clau forana.

- Establiment(Id establiment, nom, superfície, dia_soci, cost_fix_men, tipologia)
- Producte (Id producte, nom, preu_venda, proveïdor, tipologia, IVA, preu_cost, codi_barres, venda_impuls)
- Venda (Id establiment, Id producte, id demarcació, Id data, id venda, id_client, id_linea, unitats, forma_pagament, descompte)
- Client (id_client)
- Demarcació (id demarcació, nom_zona, nom_ciutat, nom_provincia)
- Data (Id_data, any, trimestre, mes, setmana)

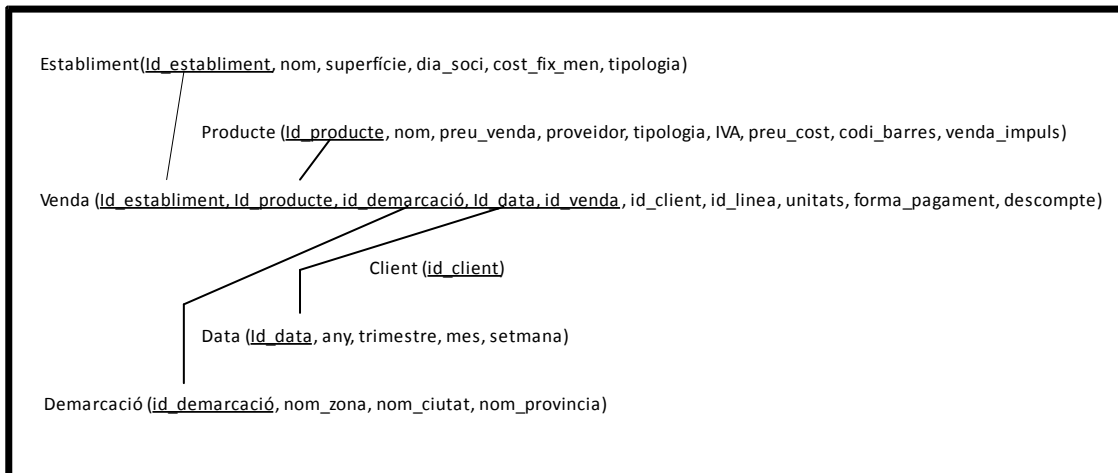


Figura 10. Esquema disseny lògic

3.3 Disseny físic

En aquest punt del disseny ja podíem centrar-nos en el disseny físic que havia de contenir les relacions esmentades, juntament amb totes les dades.

Establiment

Nom	Tipus de dada	Nul	Clau primària
Id_establiment	Number (10,0)	-	1
nom	String	-	-
superfície	String	-	-
dia_soci	String	-	-
cost_fix_men	String	-	-
tipologia	String	-	-

Figura 11. Disseny físic: taula establiment

Producte

Nom	Tipus de dada	Nul	Clau primària
Id_producte	Number (10,0)	-	1
nom	String	-	-
preu_venda	Number (10,3)	-	-
proveïdor	String	-	-
tipologia	String	-	-
IVA	Number (6,0)	-	-
preu_cost	Number (10,3)	-	-
codi_barres	String	-	-
venda_impuls	Number (6,0)	-	-

Figura 12. Disseny físic: taula producte

Venda

Nom	Tipus de dada	Nul	Clau primària
id_establiment	Number (10,0)	-	1
id_producte	Number(10,0)	-	1
id_demarcacio	Number(10,0)	-	1
id_data	Number(10,0)	-	1
id_venda	Number(10,0)	-	1
id_client	Number(10,0)	-	-
id_linea	Number(10,0)	-	-
unitats	Number(10,0)	-	-
forma_pagament	String	-	-
descompte	Number(3,0)	-	-

Figura 13. Disseny físic: taula venda

Client

Nom	Tipus de dada	Nul	Clau primària
Id_client	Number (10,0)	-	1
Id_demarcacio	Number (10,0)	-	1

Figura 14. Disseny físic: taula client

Demarcació

Nom	Tipus de dada	Nul	Clau primària
Id_demarcacio	Number (10,0)	-	1
nom_zona	String	-	-
nom_ciutat	String	-	-
nom_provincia	String	-	-

Figura 15. Disseny físic: taula demarcació

Data

Nom	Tipus de dada	Nul	Clau primària
any	Number(4,0)	-	1

Figura 16. Disseny físic: taula data

3.4 Procediments ETL

Aquest procediment és el que permet agafar les dades de les diverses fonts i manipular-les de forma convenient per a poder, posteriorment, carregar-les en les diferents taules que componen el magatzem de dades.

3.4.1 Extracció

En el nostre cas concret, aquesta primera part ja estava feta degut a que el GLDP ens havia passat una sèrie de arxius amb les dades d'interès extretes. Aquests fitxers corresponen a:

- Catàleg de productes: Un arxiu per cadascun dels anys, amb tots els productes oferts en algun dels seus punts de venda.
- Relació d'establiments: Un arxiu amb la fitxa descriptiva de cada establiment.
- Detall de vendes: Un arxiu per cada establiment amb la relació de vendes.

3.4.2 Transformació

En aquest punt, en primer lloc, hem hagut de revisar l'estat de les dades. Ens interessava que totes estiguessin donant la mateixa informació i que aquesta fos coherent amb el que s'esperava trobar.

En el nostre cas particular ja se'ns informava que, tot i tenir estructures semblants, no tots els establiments ens donarien la informació de forma similar. Hi havia establiments que proporcionaven dades en valors absoluts i d'altres que ho feien amb valors mitjos, la qual cosa feia necessari unificar-les. Podíem observar que moltes vendes s'havien efectuat donant números negatius: tal vegada havíem d'entendre que eren transaccions mal efectuades des de caixa que havien hagut de ser corregides... tot i això sorprenia que fossin sempre amb els mateixos productes.

A l'hora de fer la transformació, primerament, vam començar per les taules de tipus "demarcació"_venda_detall que les vam ajuntar amb les de "demarcació"_venda mitjançant un natural join, donant així les taules demarcació_venda_total.

Posteriorment, a les taules de productes que havien sortit de carregar els arxius cvs se'ls hi va afegir a cadascuna una columna amb l'any. Això és degut a que els identificadors depenen de l'any i per tant aquestos dos camps, id_producte i any, són les claus primàries en la taula definitiva productes_total.

A més, vam crear manualment varies taules degut a la senzillesa d'aquestes. Primerament la taula Data, que conté els anys compresos de l'estudi. La taula Demarcació, que conté dades corresponents a les ciutats o barris on hi ha un establiment. Finalment, també vam crear a mà la taula Establiment amb les dades corresponents a aquests.

La taula `usuaris` la vam extreure agafant els `id_usuari` de cada taula de tipus "demarcació" `_venda_total`. Creiem així que no hi ha cap criteri que reguli les id's donades als clients de GLDP i que depenen solament de cada establiment, per tant, la demarcació també serà clau principal de cada client.

Per últim, vam crear la taula `f_venda_total`, la nostra taula de fets. Aquesta es va omplir afegint les taules "demarcació" `_venda_total` i les fk de les taules relacionades. A més vam incorporar la clau primària `id_transacció` per a tenir l'identificador de cada venda.

3.4.3 Càrrega

Quan les dades ja havien estat depurades i transformades correctament, s'havia de passar a la càrrega de les mateixes. En el cas d'un DW totes les dades s'han de mantenir de forma històrica, així doncs, no es poden ni sobreesciure ni esborrar-se posteriorment. Bàsicament el procés seguit ha estat el següent:

- Llegir les dades dels diferents arxius.
- Inserir les dades en taules temporals.
- Transformar les dades segons era convenient.
- Fer la inserció en les taules definitives.

Arxiu Establiments (Excel)

Hem creat amb el Mysql Workbench unes taules que hem omplert de forma manual. El procés ha estat realitzat així degut a que estàvem davant de només onze camps a introduir, per tant la feina no es complicava gaire.

Arxius Productes 2010, Productes 2011 i Productes 2012 (csv)

En primer lloc vam crear taules amb els camps necessaris per a posar cadascun dels fitxers. Posteriorment, les vam omplir des del Mysql Workbench mitjançant la comanda:

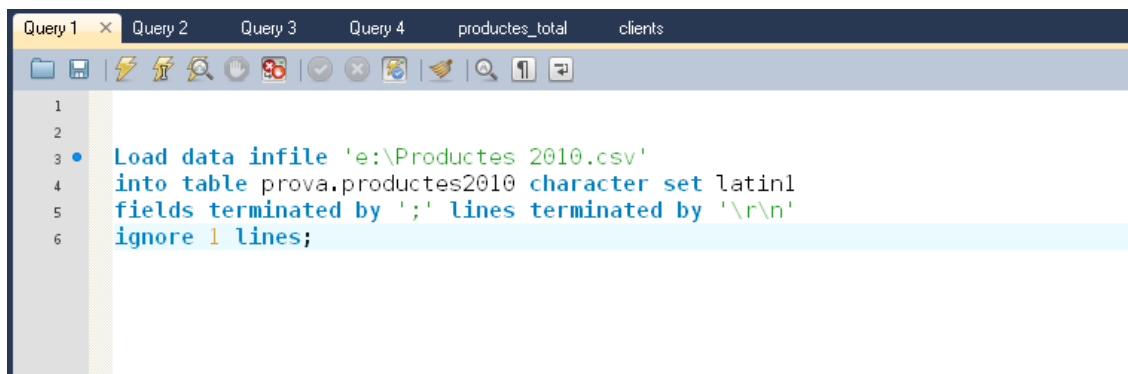


Figura 17. Càrrega arxius productes

Arxius de vendes i vendes al detall (Access)

Per a realitzar aquest pas, el primer que vam fer va ser descarregar-nos el Mysql Connector ODBC 5.1.11. Un cop instal·lat, vam anar a "Panel de control" i des de "Herramientas administrativas" vam accedir a "Administrador de orígenes de datos" ODBC, des d'aquí vam crear una nova connexió ODBC.

Un cop creada la connexió, des de l'access ja vam poder accedir a cada taula de cada arxiu i les vam poder anar exportant a ODBC, seleccionant com a origen de dades la nostra connexió. Així, al tornar al Mysql Workbench ja teníem la taula creada i exportada.

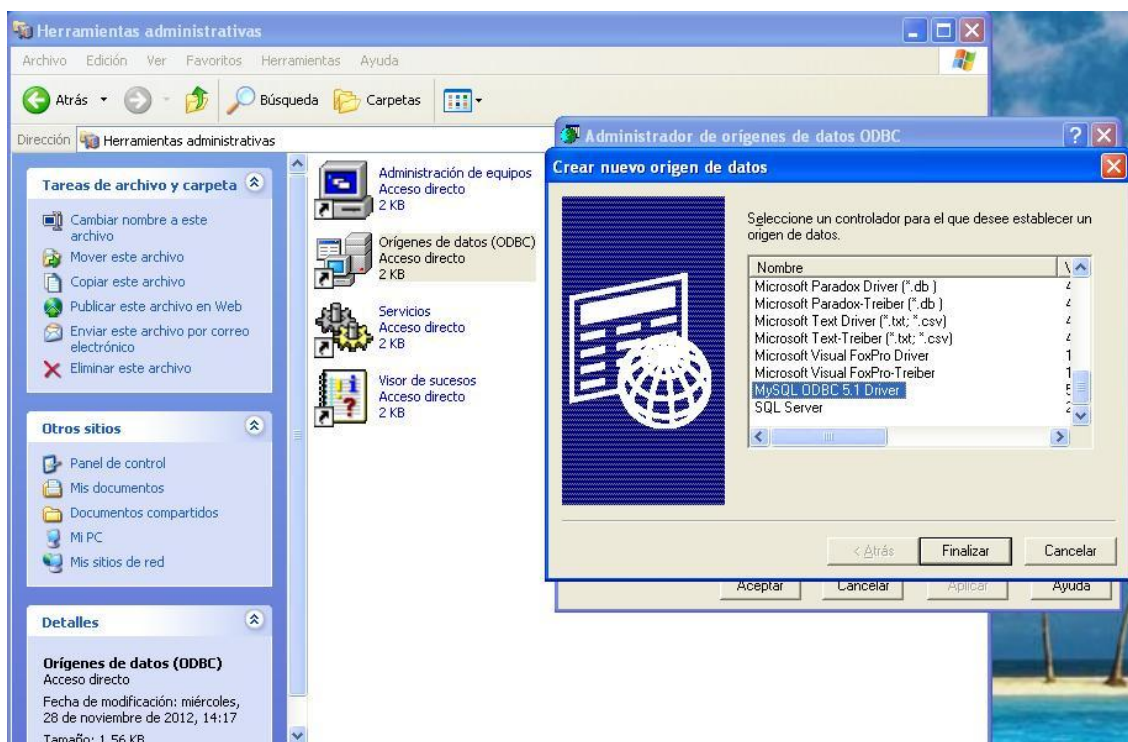


Figura 18. Càrrega arxius vendes i vendes al detall

4. Implementació

4.1 Descripció

Un cop acabades les parts anteriors del treball de final de carrera, ja ens vam poder posar a implementar la base de dades. Per a fer-ho hem utilitzat l'eina Mysql Workbench, una eina visual de disseny de bases de dades que integra desenvolupament de programari, administració de bases de dades, disseny de bases de dades, creació i manteniment per al sistema de base de dades Mysql.

4.2 Accés a la Base de Dades

Per accedir al sistema de gestió de bases de dades hem utilitzat el logins i les contrasenyes que ens proporcionava el GLDP.

Usuari	Contrasenya
root	tfcmmmd
tfcmmmd	tfcmmmd

Figura 19. Taula usuaris i contrasenyes

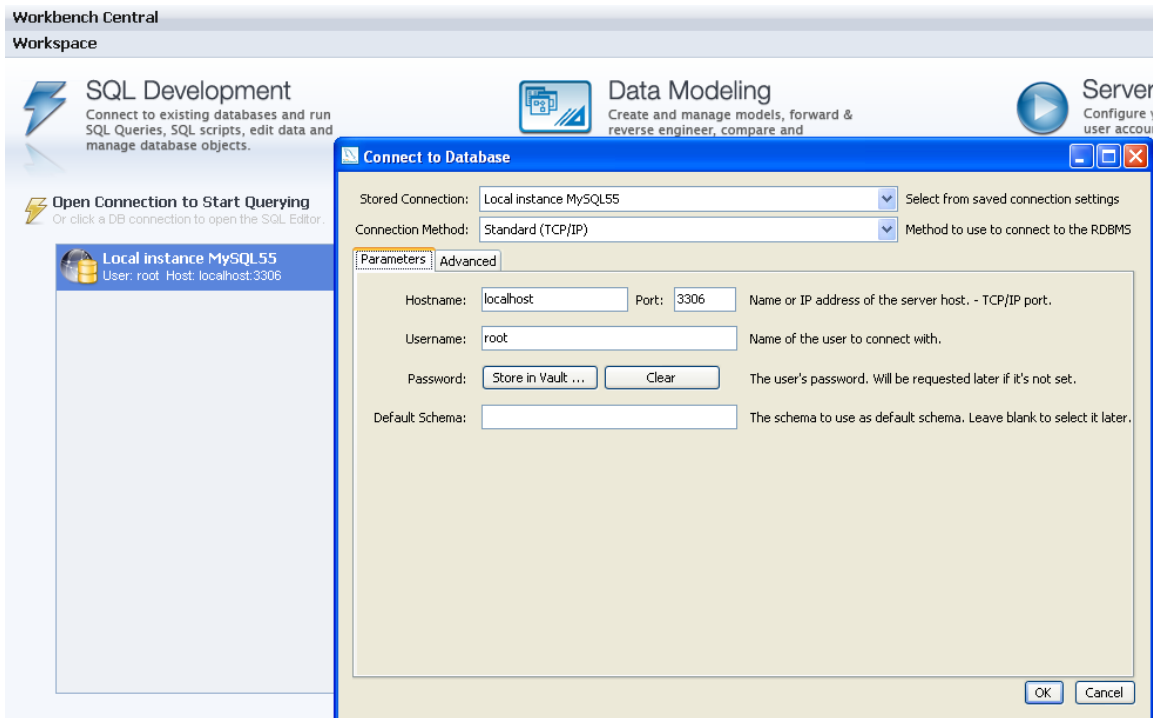


Figura 20. Login base de dades

Un cop dins del sistema de gestió de dades vam poder definir, en primer lloc, l'Schema prova, on vam crear les taules temporals, i després l'Schema GDLP, on hem creat les taules finals.

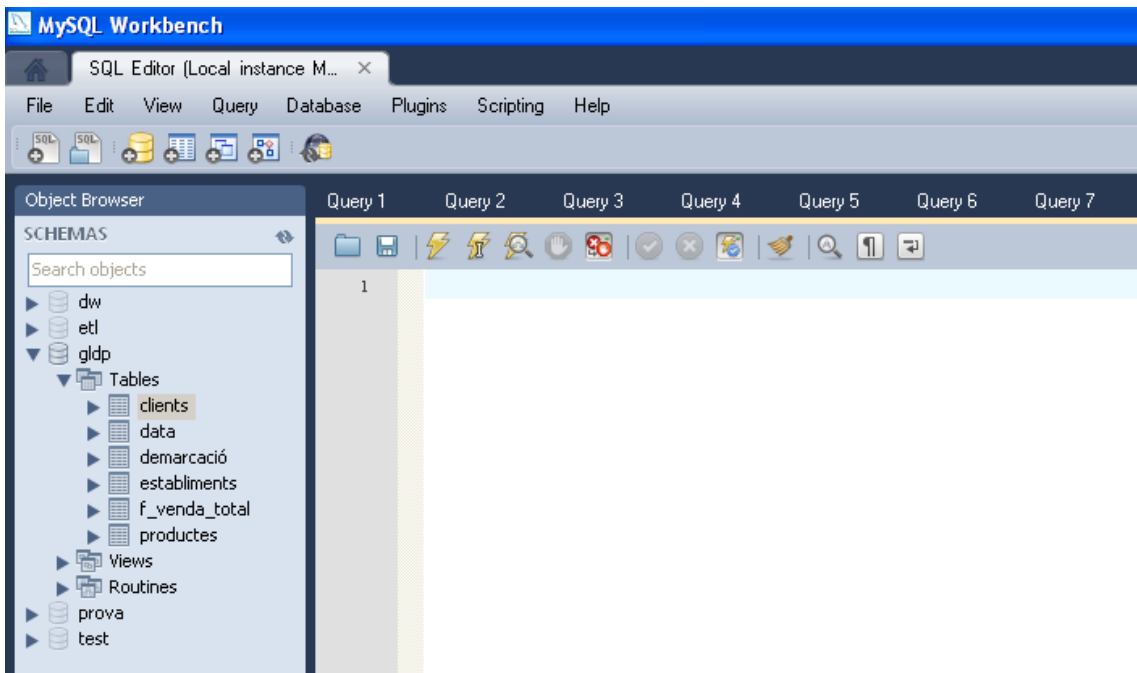


Figura 21. Esquemes MySQL Workbench

A més de crear les taules, és des d'aquí que hem pogut fer les operacions de LOAD dels arxius .csv i la seva corresponent transformació, comentades amb anterioritat en l'apartat que tractava sobre el procés ETL. En el cas concret dels productes de l'any 2010 hem realitzat els següents passos:

- Carregar les dades.

```
Load data infile 'e:\Productes 2010.csv'  
into table prova.productes2010 character set latin1  
fields terminated by ';' lines terminated by '\r\n'  
ignore 1 lines;
```

- Eliminar les files buides o sense id.

```
Delete From prova.productes2010 WHERE ID_PRODUCTE = 0;
```

- Convertir els valors numèrics per a un millor tractament posterior.

```
Update prova.productes2010  
Set Preu_venda = replace(Preu_venda, ',', '.');  
Update prova.productes2010  
Set Preu_cost = replace(Preu_cost, ',', '.');  
Posem a 0 el preu de cost allà on no hi és.  
UPDATE prova.productes2010  
SET preu_cost='0'  
WHERE preu_cost='';
```

4.3 Accés a l'eina Business Intelligence

Amb tot el magatzem de dades creat, només falta obrir la connexió amb el programa Pentaho Report Designer. Per a fer-ho, primerament, establirem la connexió.

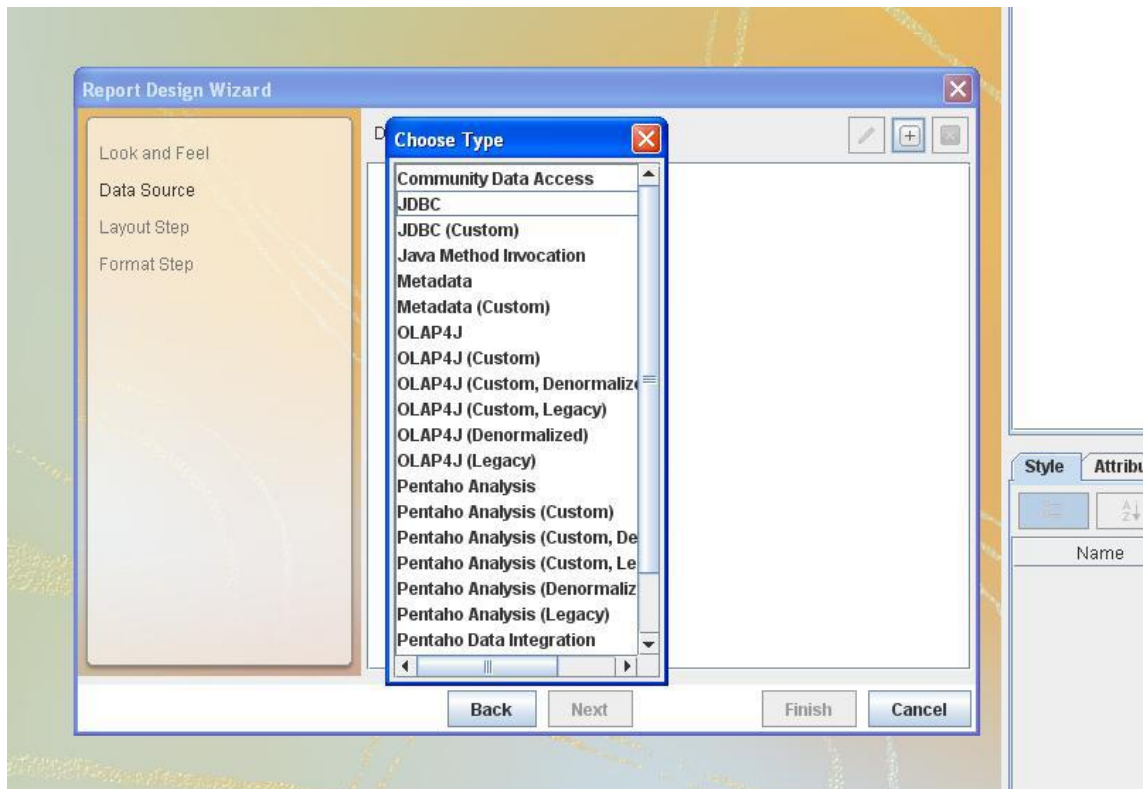


Figura 22. Establiment connexió amb Pentaho Report Designer

Un cop definida la connexió, editarem la consulta afegint al llenç aquelles taules necessàries. Posteriorment, seleccionarem els camps desitjats i les condicions.

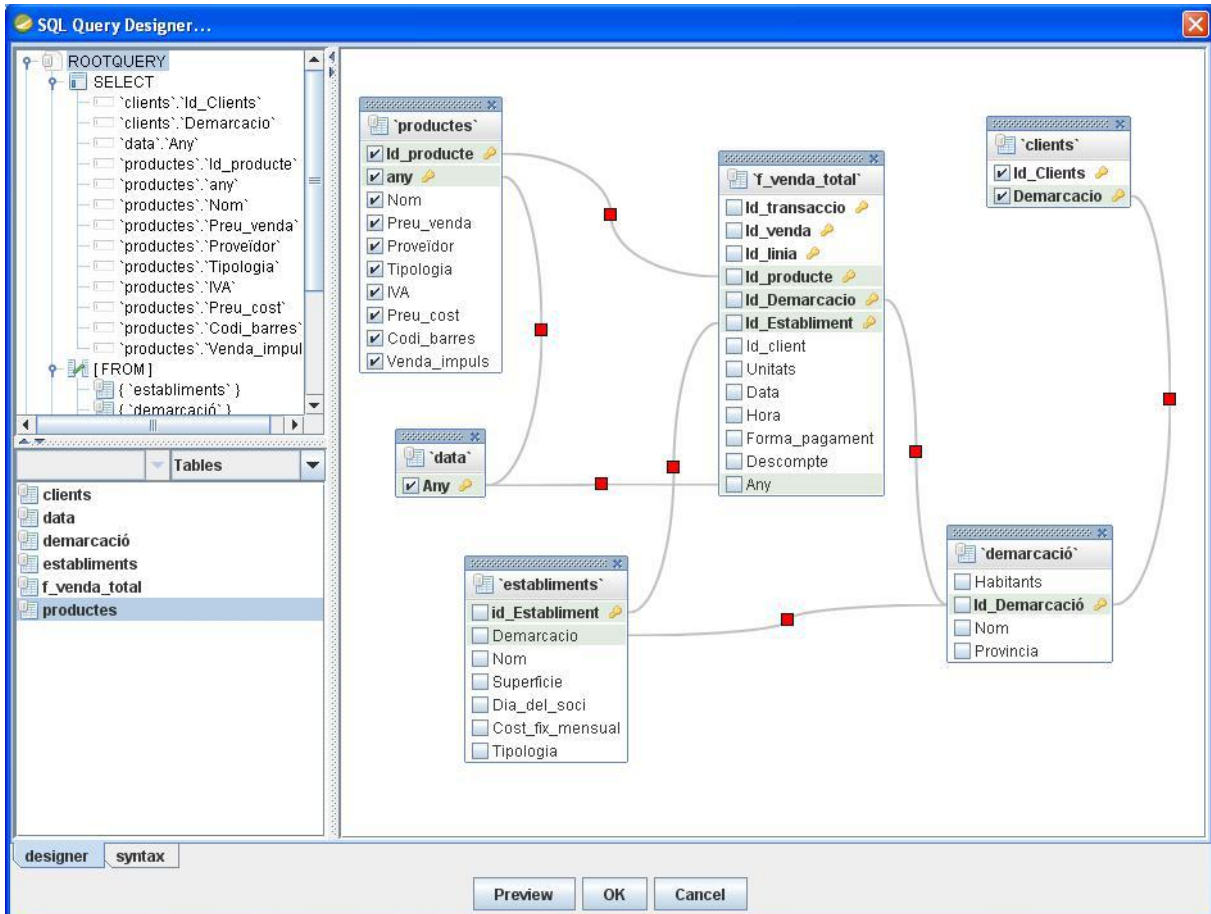


Figura 23. Disseny de la consulta

4.4 Informes:

4.4.1 Total de vendes i marge net del grup:

Per aquesta consulta hem calculat el total de vendes en un any i li hem restat les despeses generades pel lloguer de les botigues i el cost dels productes venuts. El resultat seria el total de beneficis nets que entren en un any al grup comercial sencer.

Print Preview

Report Export View Help

Any 2010

Auto-Update on selection Update

enero 06, 2013 @ 11:00

Total de vendes i marge net del grup

Vendes	3.320.791.431,72
Cost Productes	2.214.152.802,13
Cost establiments	388.356.247,200
Beneficis	1.104.898.629,59

1/1

Figura 24. Informe 1: total de vendes i marge net del grup

SELECT

```
SUM(productes.preu_venda*f_venda_total.unitats)      AS Vendes,  
SUM(productes.preu_cost*f_venda_total.unitats) AS Cost_Productes,  
SUM(establiments.cost_fix_mensual*12)  AS Cost_establiments,  
SUM(productes.preu_venda*f_venda_total.unitats) -  
SUM(productes.preu_cost*f_venda_total.unitats) -(establiments.cost_fix_mensual*12) AS  
Beneficis
```

FROM

```
`establiments` INNER JOIN `f_venda_total` ON `establiments`.`id_Establiment` =  
`f_venda_total`.`Id_Establiment`  
  
INNER JOIN `productes` ON `f_venda_total`.`Id_producte` = `productes`.`Id_producte`  
  
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`  
  
AND `data`.`Any` = `productes`.`any`  
  
WHERE  
  
`data`.`Any` = ${Any}
```

4.4.2 Import mitjà de venda per soci i establiment

GLDP ens va demanar en aquesta consulta que filtréssim un informe amb l'import mitjà de vendes per soci i establiment. Per fer-ho, hem agafat totes aquelles vendes produïdes per socis i les hem mostrat juntament el nombre de clients socis, el total de vendes i els mitjà per compra i per any.

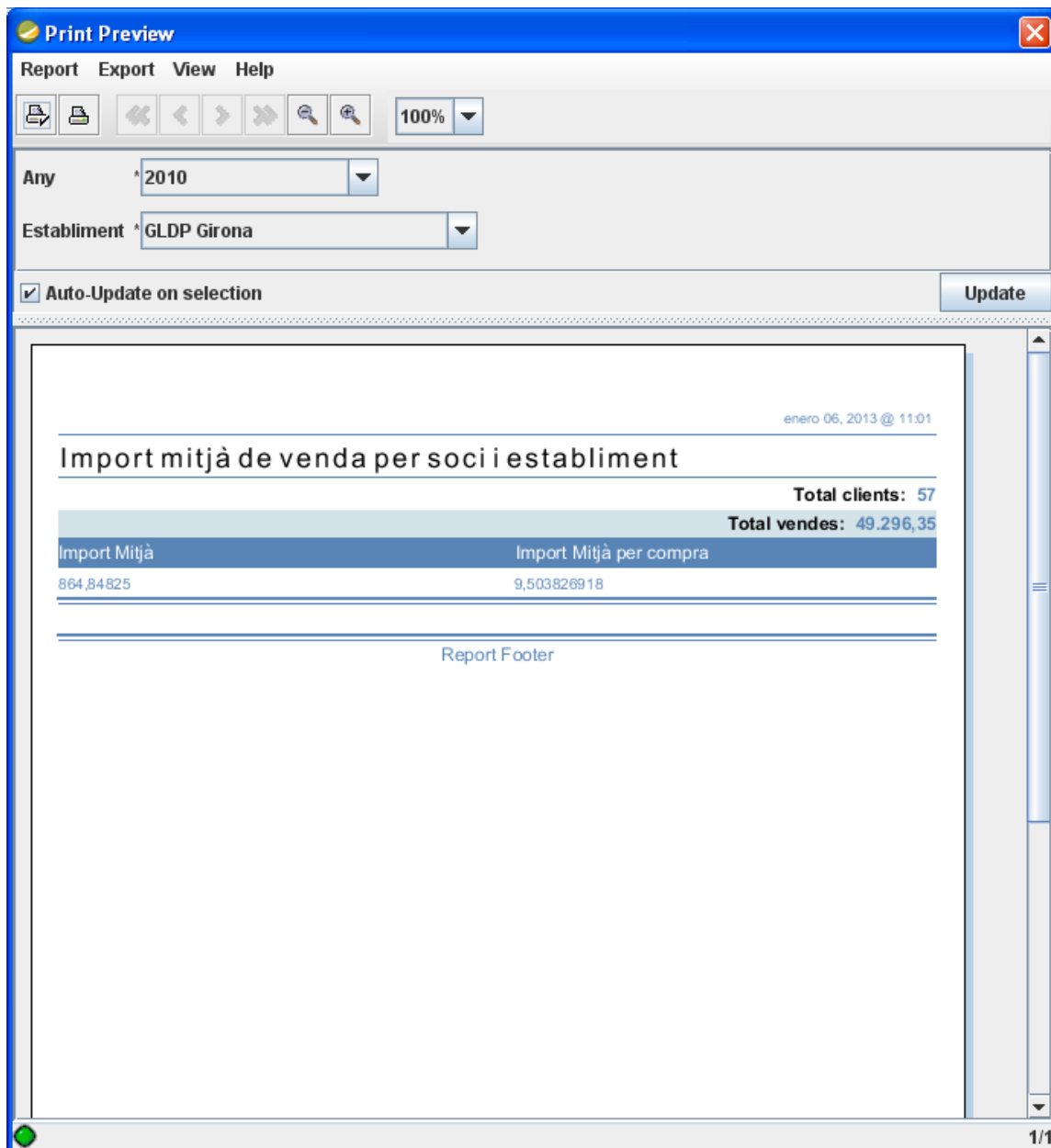


Figura 25. Informe 2: import mitjà de venda per soci i establiment

```
SELECT
    sum(`productes`.`Preu_venda`) / count( DISTINCT`clients`.`Id_Clients`) AS
Import_Mitjà_per_any,
    count( DISTINCT`clients`.`Id_Clients`) AS Total_clients,
    sum(`productes`.`Preu_venda`) AS Total_vendes,
    sum(`productes`.`Preu_venda`) / count( DISTINCT`clients`.`Id_Clients`) / count(DISTINCT
f_venda_total.id_venda) AS Import_Mitjà_per_compra
FROM
    `f_venda_total` INNER JOIN `clients` ON `f_venda_total`.`id_client` = `clients`.`Id_Clients`
    INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
    INNER JOIN `productes` ON `f_venda_total`.`Id_producte` = `productes`.`Id_producte`
    INNER JOIN `establiments` ON `f_venda_total`.`Id_Establiment` =
`establiments`.`id_Establiment`
    AND `data`.`Any` = `productes`.`any`
WHERE
    f_venda_total.any = ${Any}
AND establiments.nom = ${Establiment}
AND f_venda_total.id_client > 0
```

4.4.3 Percentatge de vendes respecte el total del grup

En aquest informe el GLDP ens va fer estudiar el tant per cent de vendes que genera cada establiment (segons la demarcació) sobre el total de vendes generades pel grup en una determinada data.

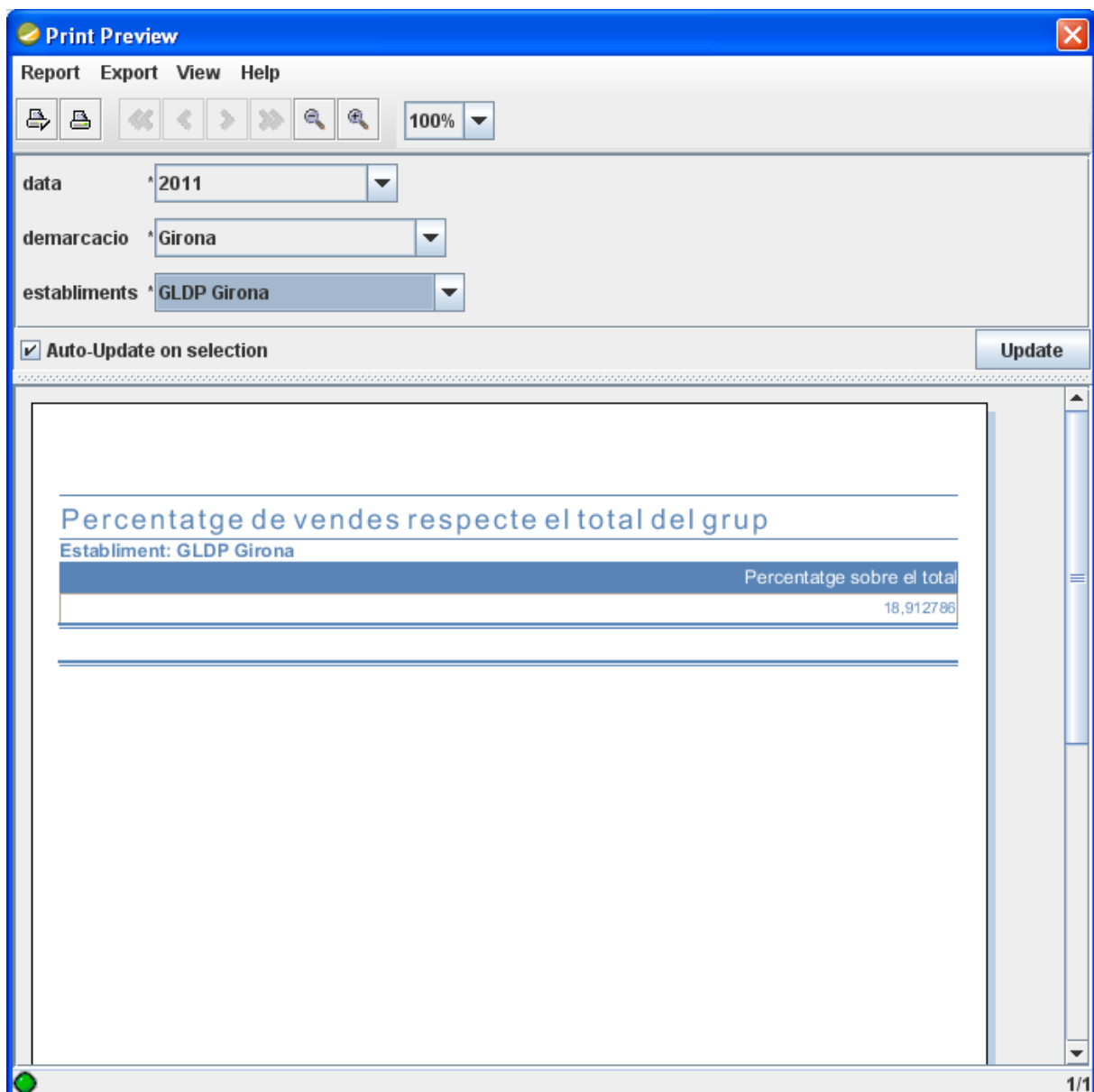


Figura 26. Informe 3: Percentatge de vendes respecte el total del grup

SELECT

```
SUM(productes.preu_venda*f_venda_total.unitats)/(SELECT  
SUM(productes.preu_venda*f_venda_total.unitats)
```

FROM

```
`data` INNER JOIN `f_venda_total` ON `data`.`Any` = `f_venda_total`.`Any`  
INNER JOIN `productes` ON `data`.`Any` = `productes`.`any`  
AND `productes`.`Id_producte` = `f_venda_total`.`Id_producte`
```

WHERE

```
`data`.`Any` = ${data} * 100 AS Percentatge_per_botiga
```

FROM

```
`data` INNER JOIN `f_venda_total` ON `data`.`Any` = `f_venda_total`.`Any`  
INNER JOIN `productes` ON `data`.`Any` = `productes`.`any`  
AND `productes`.`Id_producte` = `f_venda_total`.`Id_producte`  
INNER JOIN `demarcació` ON `f_venda_total`.`Id_Demarcacio` =  
`demarcació`.`Id_Demarcació`  
INNER JOIN `establiments` ON `f_venda_total`.`Id_Establiment` =  
`establiments`.`id_Establiment`  
AND `demarcació`.`Id_Demarcació` = `establiments`.`Demarcacio`
```

WHERE

```
`data`.`Any` = ${data}  
AND `demarcació`.`nom` = ${demarcacio}  
AND `establiments`.`nom` = ${establiments};
```

4.4.4 Preus màxims i mínims per tipologia d'establiment i característiques del producte.

En aquest informe vam extreure els preus màxims i mínims dels productes. Aquests van ser filtrats segons la tipologia d'establiment i la tipologia del mateix producte. Per a realitzar-ho hem fet la unió de dos consultes on buscàvem el màxim i mínim dels preus dels productes.

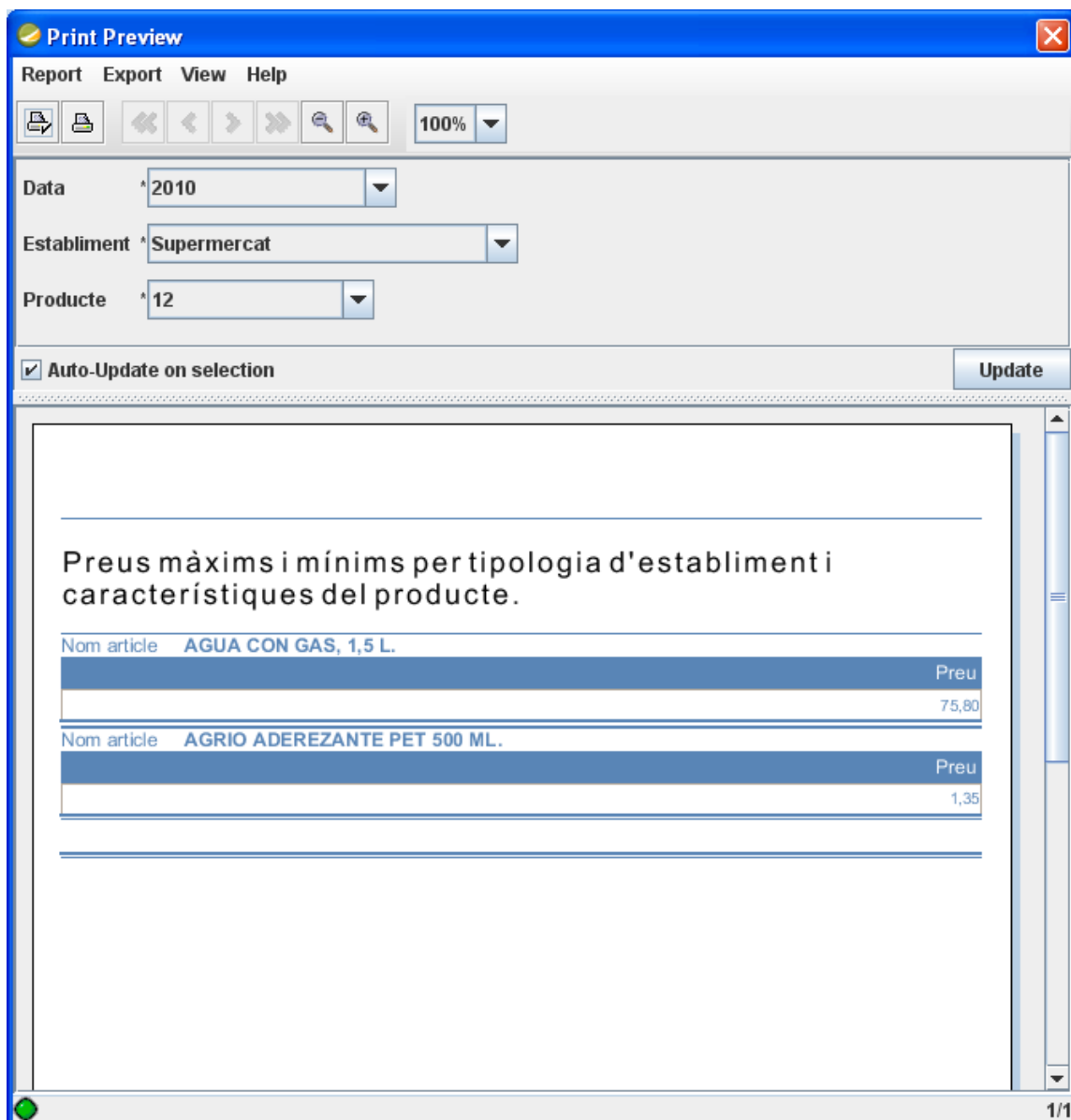
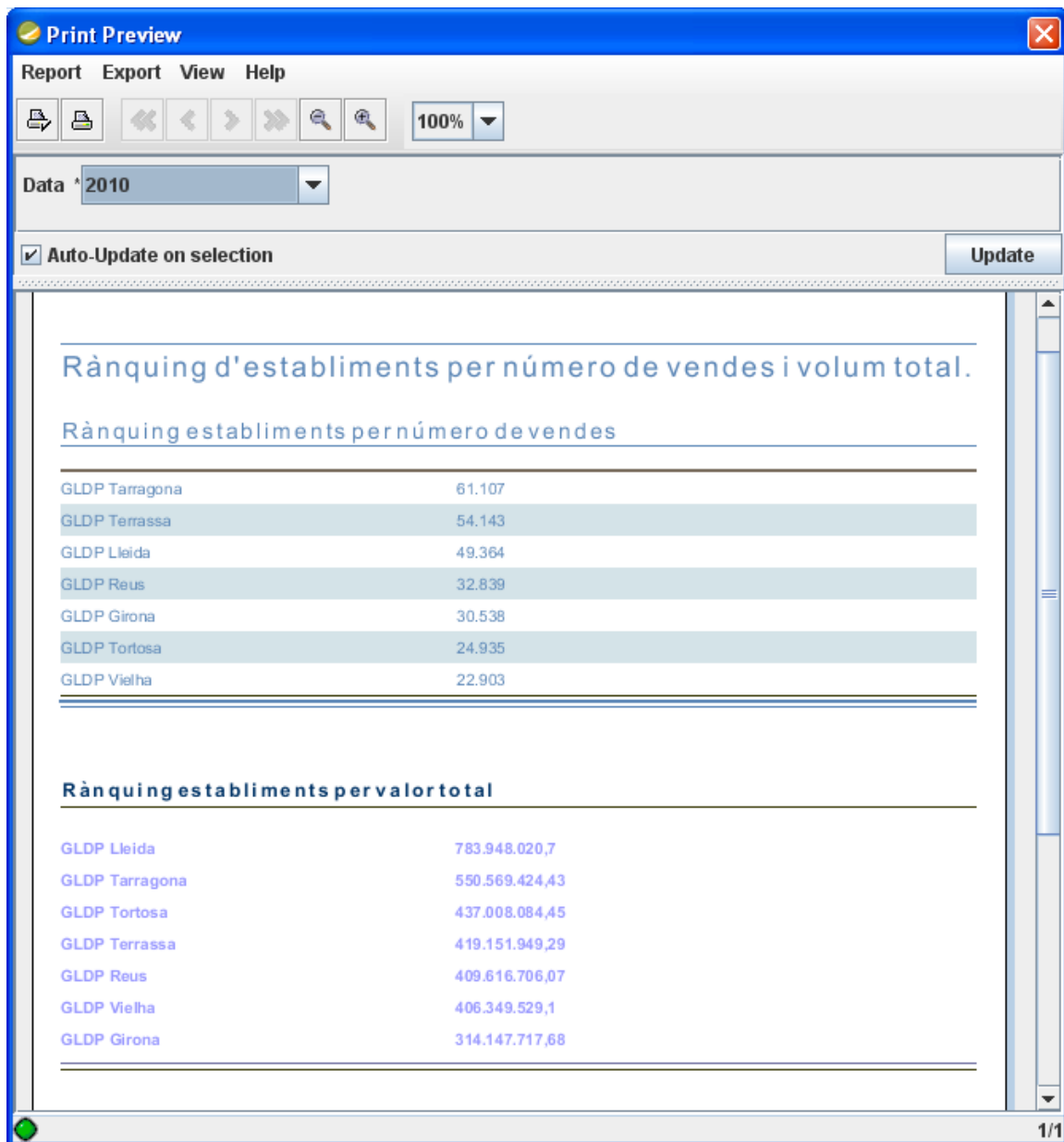


Figura 27. Informe 4: preus màxims i mínims per tipologia d'establiment i característiques del producte

```
SELECT DISTINCT
  `productes`.`nom`,
  `productes`.`preu_venda`
FROM
  `f_venda_total` INNER JOIN `productes` ON `f_venda_total`.`id_producte` =
  `productes`.`id_producte`
  INNER JOIN `establiments` ON `f_venda_total`.`id_Establiment` =
  `establiments`.`id_Establiment`
  INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`
WHERE
  preu_venda = ( SELECT MAX(preu_venda) FROM `productes` WHERE productes.tipologia =
  ${Producte})
  AND establiments.tipologia = ${Establiment}
  AND productes.tipologia = ${Producte}
  AND DATA.any = ${Data}
UNION
SELECT DISTINCT
  `productes`.`nom`,
  `productes`.`preu_venda`
FROM
  `f_venda_total` INNER JOIN `productes` ON `f_venda_total`.`id_producte` =
  `productes`.`id_producte`
  INNER JOIN `establiments` ON `f_venda_total`.`id_Establiment` =
  `establiments`.`id_Establiment`
  INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`
WHERE
  preu_venda = ( SELECT MIN(preu_venda) FROM `productes` WHERE preu_venda > 0 AND
  tipologia = ${Producte} )
  AND establiments.tipologia = ${Establiment}
  AND productes.tipologia = ${Producte}
  AND DATA.any = ${Data};
```

4.4.5 Rànquing d'establiments per número de vendes i volum total.

Per a realitzar aquest informe hem realitzat dos consultes diferents. La primera mostra el rànquing d'establiments per número de vendes. La segona l'hem creada amb una subconsulta que agafa el nombre d'establiments per valor total. Totes dues estan filtrades pels anys.



Print Preview

Report Export View Help

Data: 2010

Auto-Update on selection Update

Rànquing d'establiments per número de vendes i volum total.

Rànquing establiments per número de vendes

GLDP Tarragona	61.107
GLDP Terrassa	54.143
GLDP Lleida	49.364
GLDP Reus	32.839
GLDP Girona	30.538
GLDP Tortosa	24.935
GLDP Vielha	22.903

Rànquing establiments per valor total

GLDP Lleida	783.948.020,7
GLDP Tarragona	550.569.424,43
GLDP Tortosa	437.008.084,45
GLDP Terrassa	419.151.949,29
GLDP Reus	409.616.706,07
GLDP Vielha	406.349.529,1
GLDP Girona	314.147.717,68

1/1

Figura 28. Informe 5: rànquing d'establiments per número de vendes i volum total

```
SELECT
    count(`f_venda_total`.`id_venda`) AS nombre_vendes,
    `establiments`.`nom`
FROM
    `establiments` INNER JOIN `f_venda_total` ON `establiments`.`id_Establiment` =
`f_venda_total`.`id_Establiment`
    INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
WHERE
    `data`.`Any` = ${Data}
GROUP BY
    nom
ORDER BY
    nombre_vendes DESC
```

```
SELECT
    SUM(f_venda_total.unitats*productes.preu_venda) AS preu,
    `establiments`.`Nom`
FROM
    `f_venda_total` INNER JOIN `productes` ON `f_venda_total`.`id_producte`
= `productes`.`id_producte`
    INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
    AND `data`.`Any` = `productes`.`any`
    INNER JOIN `establiments` ON `f_venda_total`.`id_establiment` =
`establiments`.`id_Establiment`
WHERE
    `data`.`Any` = ${Data}
GROUP BY
    f_venda_total.id_establiment
ORDER BY
    preu DESC
```

4.4.6 Top ten productes

En aquesta consulta mostrem el rànquing dels 10 productes estrella de cada any que més han sigut venuts al grup GLDP. De la consulta s'han eliminat aquells que no estan dins els id's donats inicialment per l'usuari, per a tenir-los separats els hi vam donar el valor identificador 999999999.

Top Ten productes	
Nom producte: ENELDO,	Total vendes 39.215
Nom producte: CUBO AGUA C/PICO, 13 L	Total vendes 10.367
Nom producte: DESODORANTE FRESH SENSATIONS, 150 ML.	Total vendes 9.289
Nom producte: CREMA MANOS Nº 3 ANTIMANCHAS SENSATIONS 125 ML	Total vendes 6.930
Nom producte: GALLETAS MARIA TOSTADA, 4X200 G.	Total vendes 6.921
Nom producte: GALLETAS MARIA, 4X200 G.	Total vendes 6.898
Nom producte: CREMA ANTIARRUGAS CAVIAR SENSATIONS 50 ML.	Total vendes 6.663
Nom producte: CUAJADA POLVO ESTUCHE 48 G	Total vendes 5.462
Nom producte: GALLETAS PERROS MIX, 400 G.	Total vendes 5.087
Nom producte: CREMA CORPORAL SOFT SENSATIONS 200 ML.	Total vendes 4.114

Figura 29. Informe 6: top ten productes

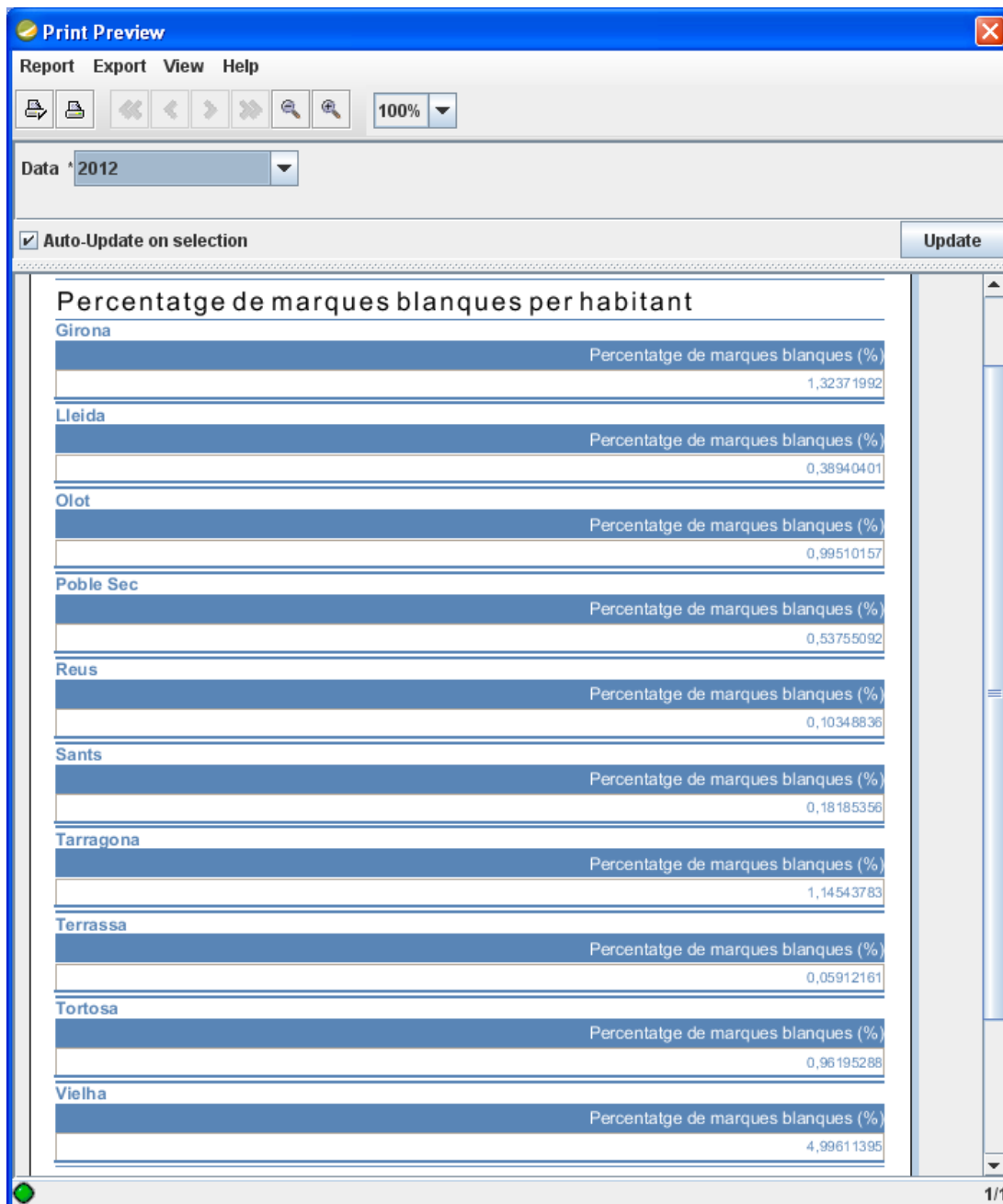
```
SELECT DISTINCT
  `productes`.`Nom`,
  COUNT(`f_venda_total`.`id_venda`) AS total_vendes
FROM
  `f_venda_total` INNER JOIN `productes`
ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`

WHERE `f_venda_total`.`id_producte` < 999999999
AND `data`.`Any` = ${Data}

GROUP BY `productes`.`Nom`
ORDER BY `total_vendes` DESC
LIMIT 10;
```

4.4.7 Percentatge de marques blanques per habitant

GLDP, en aquest cas, ens va demanar el percentatge de marques blanques pel total d'habitants que té cada demarcació en un any. Nosaltres vam considerar que els barris de Barcelona, per la seva mida, contenen com a zones úniques (tot i considerar que, per exemple, el Poble Sec podria tenir dos establiments en un futur).



The screenshot shows a 'Print Preview' window with a menu bar (Report, Export, View, Help) and a toolbar with navigation and zoom controls. The data is for the year 2012. The report title is 'Percentatge de marques blanques per habitant'. The data is presented in a table with columns for the district name and the percentage value.

District	Percentatge de marques blanques (%)
Girona	1,32371992
Lleida	0,38940401
Olot	0,99510157
Poble Sec	0,53755092
Reus	0,10348836
Sants	0,18185356
Tarragona	1,14543783
Terrassa	0,05912161
Tortosa	0,96195288
Vielha	4,99611395

Figura 30. Informe 7: percentatge de marques blanques per habitant


```
SELECT
    Demarcació.nom,
    SUM(f_venda_total.id_venda)/Demarcació.habitants/(SELECT
SUM(f_venda_total.id_venda)/Demarcació.habitants
FROM
    `demarcació` INNER JOIN `f_venda_total`
ON `demarcació`.`Id_Demarcació` = `f_venda_total`.`Id_Demarcacio`
    INNER JOIN `productes`
ON `f_venda_total`.`Id_producte` = `productes`.`Id_producte`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
    AND `data`.`Any` = `productes`.`any`

WHERE
`data`.`Any` = ${Data})*100 AS Marques_Blanques_Habitant

FROM
    `demarcació` INNER JOIN `f_venda_total`
ON `demarcació`.`Id_Demarcació` = `f_venda_total`.`Id_Demarcacio`
    INNER JOIN `productes`
ON `f_venda_total`.`Id_producte` = `productes`.`Id_producte`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
    AND `data`.`Any` = `productes`.`any`

WHERE `productes`.`proveïdor`='GLDP'
AND `data`.`Any` = ${Data}
GROUP BY `demarcació`.`nom`
```

4.4.8 Categorització clients ABC

Per a realitzar aquest informe, primer de tot, vam seleccionar els clients socis i la seva demarcació, els dos elements necessaris per a identificar-los. Posteriorment, vam separar els socis segons el volum de compres anuals que realitzen sobre la venda en els que suposen menys de 40%, el 40%-80% i més del 80%.

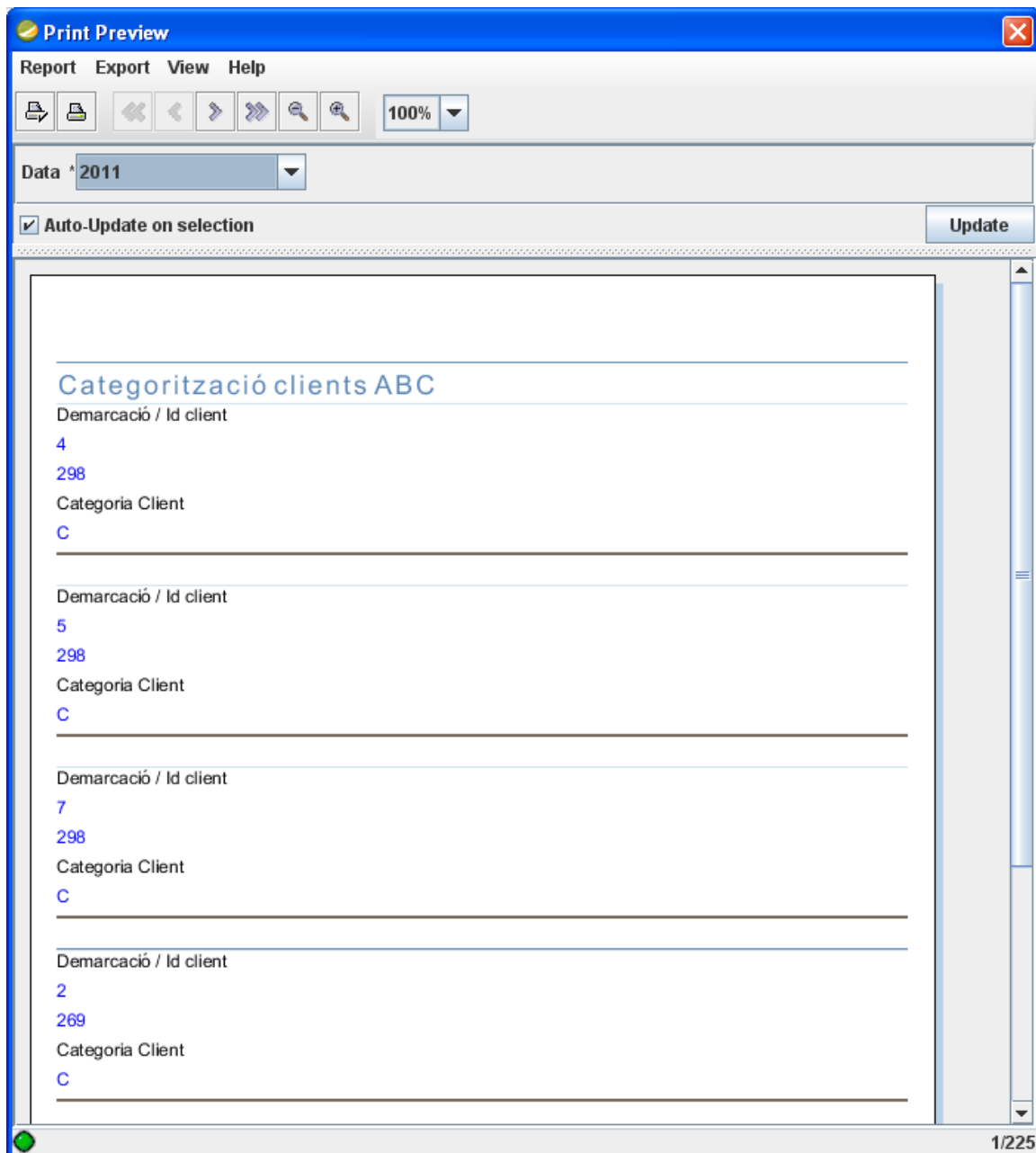


Figura 31. Informe 8: categorització clients ABC

```
SELECT `clients`.`Id_Clients`,`clients`.`demarcacio`,

CASE WHEN sum(`productes`.`Preu_venda`)<=(SELECT AVG(`productes`.`Preu_venda`)*0.40
      FROM   `f_venda_total` INNER JOIN `productes` ON
`f_venda_total`.`Id_producte` = `productes`.`Id_producte`
      INNER JOIN `clients` ON `f_venda_total`.`Id_client` = `clients`.`Id_Clients`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
      AND `data`.`Any` = `productes`.`any` WHERE `data`.`Any` = ${Data})
THEN 'A'

ELSE (CASE WHEN sum(`productes`.`Preu_venda`)<=(SELECT
AVG(`productes`.`Preu_venda`)*0.80

FROM   `f_venda_total` INNER JOIN `productes` ON `f_venda_total`.`Id_producte` =
`productes`.`Id_producte`
      INNER JOIN `clients` ON `f_venda_total`.`Id_client` = `clients`.`Id_Clients`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
      AND `data`.`Any` = `productes`.`any` WHERE `data`.`Any` = ${Data})
THEN 'B'
ELSE 'C' END)
END AS Clients_ABC

FROM
  `f_venda_total` INNER JOIN `productes` ON `f_venda_total`.`Id_producte` =
`productes`.`Id_producte`
  INNER JOIN `clients` ON `f_venda_total`.`Id_client` = `clients`.`Id_Clients`
INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`

WHERE `data`.`Any` = ${Data}

GROUP BY `clients`.`Id_Clients`,`clients`.`Demarcacio`
ORDER BY Clients_ABC;
```

4.4.9 Anàlisi de compra per impuls

En aquest informe vam decidir mostrar la quantitat i tant per cent total de la venda per impuls que es genera en un establiment durant un any. Entenem per venda d'impuls tots aquells productes venuts que estan marcats com a tal en la base de dades i que, en principi, responen a unes característiques comunes: objecte petit, de poca necessitat, pensat per a ser comprat per suggeriment.

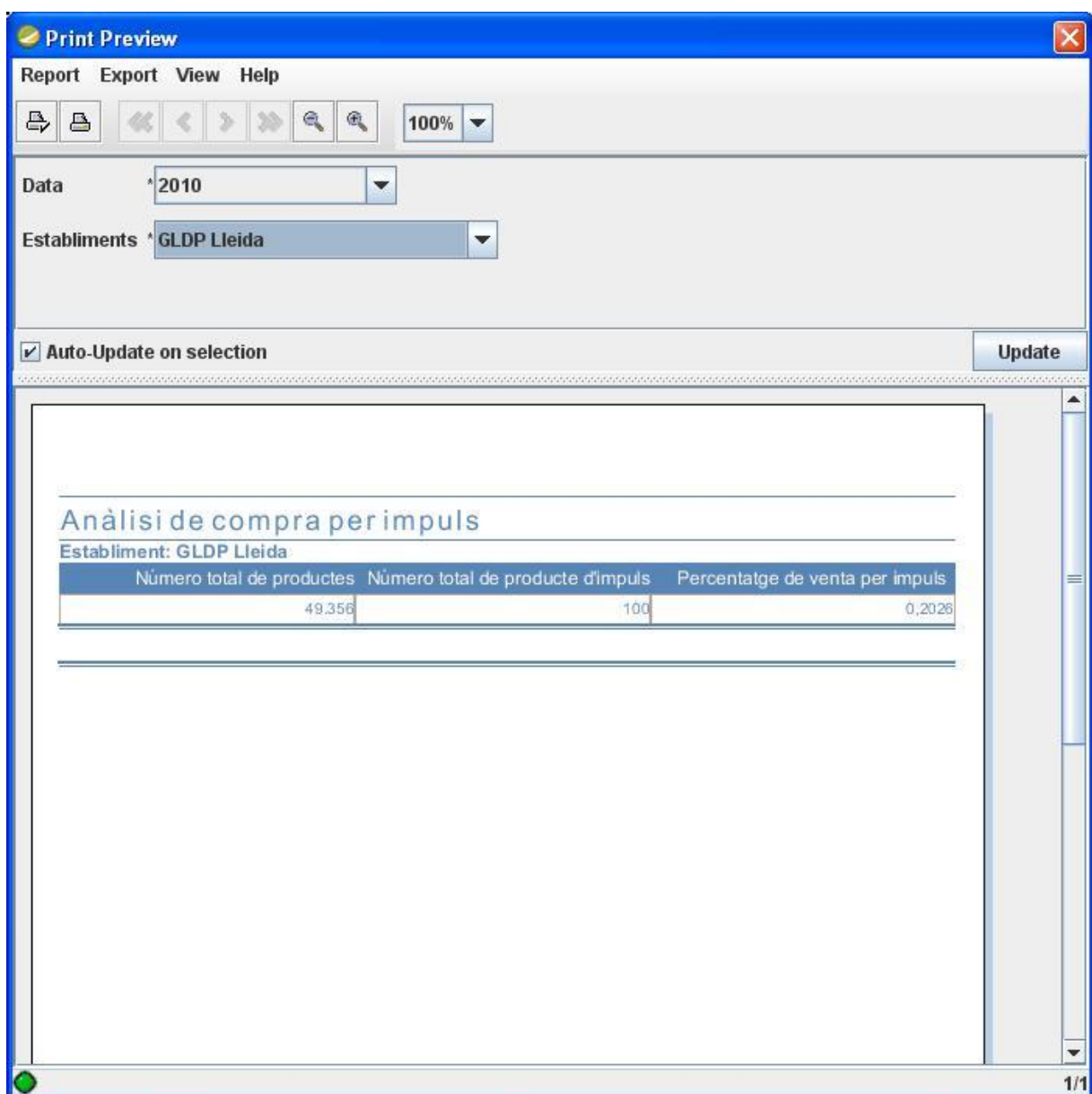


Figura 32. Informe 9: Anàlisi de compra per impuls

```
SELECT
  `establiments`.`nom`,
  COUNT(productes.preu_venda*f_venda_total.unitats) AS Total_productes_impuls,
(SELECT
  COUNT(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
  INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`
WHERE
  `data`.`any` = ${Data}
AND `establiments`.`nom` = ${Establiments}) AS Total_producte,

  COUNT(productes.preu_venda*f_venda_total.unitats)/(SELECT
  COUNT(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
  INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`
WHERE
  `data`.`any` = ${Data}
AND `establiments`.`nom` = ${Establiments})*100 AS Percentatge_impuls

FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
  INNER JOIN `data` ON `f_venda_total`.`Any` = `data`.`Any`
  AND `data`.`Any` = `productes`.`any`
WHERE
  `data`.`any` = ${Data}
AND `establiments`.`nom` = ${Establiments}
AND productes.venda_impuls= 'Sí'
```

4.4.10 Distribució estacional vendes

Per a realitzar el aquest informe vam dividir la distribució anual de vendes en les diferents estacions de l'any. En l'exemple veiem el cas de GLDP Olot, establiment obert que te la seva primera venda el dia 20 de gener de 2011.

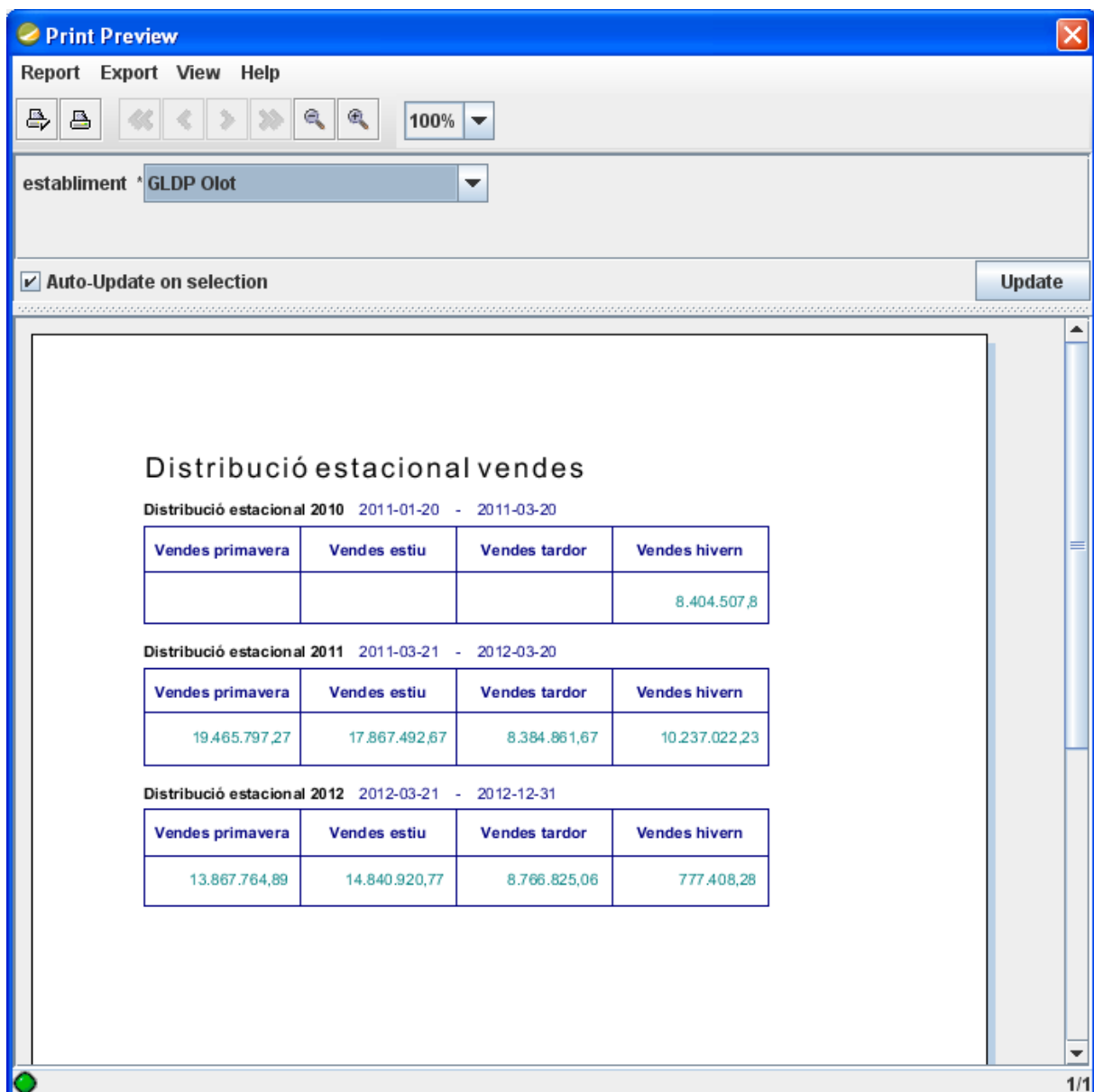


Figura 33. Informe 10: Distribució estacional vendes

```
--Any 2010
SELECT
  MIN((SELECT DATE(`f_venda_total`.`data`)))AS Data_inici,
  MAX((SELECT DATE(`f_venda_total`.`data`)))AS Data_fi,
(SELECT
  SUM(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
WHERE
  `f_venda_total`.`data` BETWEEN '2010-03-21' AND '2010-06-20'
AND `establiments`.`nom` = ${establiment}) AS Vendes_primavera,

(SELECT
  SUM(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
WHERE
  `f_venda_total`.`data` BETWEEN '2010-06-21' AND '2010-09-20'
AND `establiments`.`nom` = ${establiment}) AS Vendes_estiu,

(SELECT
  SUM(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
WHERE
  `f_venda_total`.`data` BETWEEN '2010-09-21' AND '2010-12-20'
AND `establiments`.`nom` = ${establiment}) AS Vendes_tardor,

(SELECT
  SUM(productes.preu_venda*f_venda_total.unitats)
FROM
  `establiments` INNER JOIN `f_venda_total`
  ON `establiments`.`id_Establiment` = `f_venda_total`.`id_Establiment`
  INNER JOIN `productes` ON `f_venda_total`.`id_producte` = `productes`.`id_producte`
WHERE
  `f_venda_total`.`data` BETWEEN '2010-12-21' AND '2011-03-20'
AND `establiments`.`nom` = ${establiment}) AS Vendes_hivern
```

5. Conclusions

Un cop arribats a aquest punt, podem fer un breu recordatori a nivell personal de tots aquells passos que s'han tractat durant el TFC i de la feina realitzada durant el semestre.

He après a fer un estudi preliminar de les dades del qual he partit per a fer el model conceptual, el disseny lògic i el disseny físic. Posteriorment, he estudiat els processos ETL amb els quals he carregat les dades pertinents a taules temporals que m'ha servit per a crear el magatzem de dades. Per últim, he treballat amb eines Business Intelligence que m'ha servit per a treure els informes finals, demanats en l'enunciat del TFC. Ha sigut una feina dura però molt motivadora, en la que he pogut aprendre sobre el sistema de gestió de base de dades Mysql, l'entorn Pentaho i, especialment, el llenguatge SQL.

Un dels grans problemes que he tingut fent aquest treball ha estat precisament l'entorn Pentaho: degut a molts maldecaps amb el programa Spoon, vaig decidir fer les càrregues dels fitxers d'Access de forma manual. Jo ho vaig trobar més ràpid i fàcil tot i que, és clar, no és un procés automatitzat i possiblement sigui per la falta d'experiència en l'entorn Pentaho. D'altra banda, la màquina virtual també va suposar un contratemps elevat a l'hora de treballar amb ella, ja que pràcticament no tenia espai i les consultes dels informes podien arribar a tardar quaranta-cinc minuts. A més, he de reconèixer que per circumstàncies personals i laborals no sempre he disposat del temps necessari que un projecte de tal magnitud demana. Així doncs, voldria disculpar-me per les mancances o errors que es puguin trobar en el treball.

Tot i això, la feina realitzada ha sigut gratificant i les sensacions finals són bones. M'ha agradat posar en pràctica els coneixements obtinguts durant el semestre i veure com, poc a poc, el magatzem de dades s'anava creant. Ha sigut una bona experiència, crec que molt similar a la que es pot trobar en un entorn laboral real.

6. Línies de futur

Ara, un cop entregat el treball realitzat, seria moment de parlar amb el client, GLDP, per a validar el producte i posar en comú possibles tasques futures. Entre les quals es podria preveure les següents:

- Posada en comú per part dels establiments a l'hora de proporcionar les dades: el mateix sistema per als productes i un únic sistema per a identificar als socis. També corregir errors com el de l'establiment de Figueres, on no hi ha correlació entre les vendes i les vendes al detall.
- Crear una sola manera d'identificar els productes. Que no depengui de l'any o que almenys guardi el mateix identificador per al mateix producte independentment de l'any.
- Revisar el sistema de lloguers dels establiments, ja que pot fluctuar amb el temps i no està considerat.
- Acabar d'arreglar errors i mancances en el magatzem de dades.
- Automatitzar el procés ETL i tots aquells que facilitin el treball a l'usuari.
- Proporcionar accés a les consultes des de la web.

7. Glossari

Ad-hoc: Solució que permet a l'usuari de personalitzar una consulta en temps real, en comptes d'estar lligat a les consultes predissenyades per a informes.

Atribut: Valor específic d'una instància determinada.

Base de Dades: Conjunt de dades que pertanyen a un mateix context i es troben emmagatzemades sistemàticament per al seu ús a posteriori.

Cel·la: Mesura dins d'una mateixa estructura.

Descriptor: Atribut en un nivell que no conté informació jeràrquica.

Dimensió: Element d'anàlisi d'un fet.

ETL: Procés de extracció, transformació i càrrega d'unes dades determinades.

Fet: Part assignada per a ser analitzada

Granularitat: Nivell de detall en una taula.

Magatzem de dades: Conjunt de dades orientat a una temàtica concreta, integrat, variable en el temps i no volàtil que dóna suport a la presa de decisions.

Mesura: Atribut numèric d'una cel·la.

Mysql: Sistema de gestió de bases de dades relacional, multifil i multiusuari.

Mysql Workbench: Eina visual de disseny de bases de dades que integra desenvolupament de programari, administració de bases de dades, disseny de bases de dades, creació i manteniment per al sistema de base de dades Mysql.

OLAP: Acrònim d'*On-Line Analytical Processing* (processament analític en línia). Solució utilitzada en el camp de Business Intelligence i que té com a objectiu agilitzar la consulta de grans quantitats de dades.

Pentaho: Plataforma Business Intelligence.

Pentaho Report Designer: Programa de la plataforma Pentaho encarat a la creació d'informes.

Sistema de gestió de bases de dades: Conjunt de programes informàtics dissenyats per a facilitar la gestió d'un conjunt de dades en una base de dades

SQL: Acrònim de *Structured Query Language*. Llenguatge declaratiu d'accés a bases de dades relacionals que permet especificar diversos tipus d'operacions en aquestes.

8. Bibliografia

8.1 Documents

- “Treball final de carrera”, UOC
- Apunts de l'assignatura “Models multidimensionals i magatzems de dades (MMMD)”, UOC

8.2 Llibres de text

- “The Data Warehouse Toolkit. (2nd Edition)”, Ralph Kimball and Margy Ross.
- “Building the Data Warehouse (3rd Edition)”, W. H. Inmon.
- “The Data Warehouse ETL Toolkit”, Joe Caserta and Ralph Kimball.

8.3 Enllaços d'interès a Internet

- <http://www.inmoncif.com>
- <http://www.rkimball.com>
- <http://www.dwreview.com>
- <http://www.1keydata.com/datawarehousing/datawarehouse.html>
- <http://www.intranetjournal.com/features/datawarehousing.html>
- <http://community.pentaho.com>
- <http://dev.mysql.com>