

SPRINT 4

Tasca S4.01. Creació de Base de Dades

Descripció

Partint d'alguns arxius CSV dissenyaràs i crearàs la teva base de dades.

Nivell 1

Descàrrega els arxius CSV, estudia'ils i dissenya una base de dades amb un esquema d'estrella que contingui, almenys 4 taules de les quals puguis realitzar les següents consultes:

Primer farem uns comentaris sobre els fitxers csv proporcionats per treballar amb les dades que incorporen.

Els tres fitxers csv, amb els noms “**users_usa**”, “**users_uk**” i “**users_ca**” tenen les mateixes columnes. Per tant, els podrem unir en una sola taula per tenir tots els “users” junts.

COLUMNES	OBSERVACIONS
id	Els “id” no estan repetits, per tant es podrà fer servir de PK.
name	El nom del usuari
surname	El cognom del usuari. Al ser països anglosaxons només tenen un cognom.
phone	Les dades d'aquesta columna no les modificaré, encara que no són igual el format a cada país. S'hauria de conèixer els prefixos per zones, cosa que es podria validar amb el país, la ciutat i en determinats casos amb l'adreça pels telèfons fixes, diferenciar si són mòbils. Desagregar la part que és prefix en una nova columna. Aquí no s'estaria complint allò del valor atòmic
email	L'e-mail del usuari.
birth_date	El format de la data de naixement, ho haurem de modificar en el moment de carregar les dades. Ho convertirem a format “yyyy-mm-dd” per a tenir consistència i facilitat d'ús a les consultes.
country	El país del usuari. En aquest Sprint no crearem taula de països lligada amb les ciutats, telèfons i els codis postal. Però es podria fer. Així es podrien fer validacions de que les dades introduïdes no són errònies.
city	La ciutat del usuari
postal_code	El codi postal del usuari. Cal dir que els països de la Commonwealth tenen lletres i números. En canvi, USA només números.
address	L'adreça del usuari.

El fitxer “**products.csv**” conté les següents columnes amb les seves respectives dades:

COLUMNES	OBSERVACIONS
id	Els “id” no estan repetits, per tant es podrà fer servir de PK.
product_name	El nom del producte.
price	El preu del producte. Inclou el \$ en el seu format. Estableixo que són USD i no CAD. Això s'ha de tenir en compte quan es faci la carrega de les dades a MySQL. El símbol “\$” ho netejaré. Estableixo la hipòtesi que la nostra divisa base són USD.
color	El codi de la referència del color del producte.
weight	El pes del producte. No tinc informació si només es fa servir o bé el sistema mètric o l'imperial. Com que els països son anglosaxons estableixo la hipòtesi de que és imperial, per si hem de fer càlculs més endavant tenir-ho en compte.
warehouse_id	L'identificador del magatzem on es troba el producte.

La informació continguda en aquest fitxer, podria ser molt útil si es decidís ampliar la BDD per a gestió dels inventaris, gestionar els temps d'entrega del producte als clients en funció de les ubicacions. Optimitzar l'espai utilitzat i la proximitat del producte que es ven al consumidor, etc.

El fitxer “**crèdit_cards.csv**” conté les següents columnes amb les seves respectives dades:

COLUMNES	OBSERVACIONS
id	Els “id” no estan repetits, per tant es podrà fer servir de PK.
user_id	Els “id” dels usuaris. Servirà per poder relacionar les dades.
iban	IBAN, el número del compte bancari. Es podria fer una validació, però seria un projecte més gran
pan	El número de la targeta. A l'hora d'importar el fitxer, haurem de netejar els espais que hi ha en alguns registres. No desagregarem em funció del emissor. Es podria fer i obtenir més informació sobre els consumidors segons el tipus de targeta que tenen. S'haurà de treure els espais en blanc d'alguns registres.
pin	El codi pin de la targeta. Farem servir VARCHAR per evitar perdre números per l'esquerra si són ceros.
cvv	El codi de seguretat de la targeta. Idem que l'anterior per si hi ha ceros a la part de l'esquerra.
track1	El track1. Desconeix si aquests codis són per poder fer l'autentificació o la validació de les dades de la targeta. O per un altre ús. Em falta informació.
track2	El track2. Desconeix si aquests codis són per poder fer l'autentificació o la validació de les dades de la targeta. O per un altre ús. Em falta informació.
expiring_date	Data de caducitat de la targeta. El format original és “mm/dd/yy”, s'haurà de transformar en el procés de carrega per homogeneïtzar el format de les dates. Ho convertirem a format “yyyy-mm-dd” per a tenir consistència i facilitat d'ús a les consultes.

El fitxer “**companies.csv**” conté les següents columnes amb les seves respectives dades:

COLUMNES	OBSERVACIONS
company_id	El “id” de la companyia. Ho farem servir de PK.
company_name	El nom de la companyia.
phone	El telèfon de la companyia. Ja he comentat al parlar dels fitxers dels “users” que es podria atomitzar una mica, separant els prefixes. A la vegada que fer controls per validar que no hi ha errors en funció de les ubicacions. I per descomptat, separar entre mòbil i fixe. En aquest Sprint, no faré això, però deixo constància que es podria fer.
email	L'e-mail de la companyia.
country	El país de la companyia
website	La pàgina web de la companyia

En la majoria dels fitxers, el separador que delimita els camps és la “,”. Excepte en el fitxer “**transactions.csv**”, on el separador que delimita és el “;”. Això ho haurem de tenir en compte a l'hora de carregar els fitxers.

El fitxer “**transactions.csv**” conté les següents columnes amb les seves respectives dades:

COLUMNES	OBSERVACIONS
id	El “id” de la transacció. Ho farem servir de PK.
card_id	El card_id serà FK en el model.
business_id	El business_id serà FK en el model.
timestamp	Dia i hora de la transacció.
amount	Import de la transacció.
declined	Indicador de si s'ha rebutjat la transacció. “0” vol dir correcte, “1” vol dir rebutjada.

product_ids	El "id" del producte. S'ha de tenir en compte que hi ha més d'un per registre i els haurem de separar més endavant, es a dir, fer un Split.
user_id	El "id" del usuari serà FK en el model.
lat	La latitud geogràfica
longitude	La longitud geogràfica

El model estrella que farem servir estarà format per les següents taules de dimensions:

- Taula Dim_users
 - El camp birth_date era VARCHAR abans de fer la carrega i després ho he modificat. La captura de sota s'ha fet a posteriori.

Field	Type	Null	Key	Default
id	int	NO	PRI	NULL
name	varchar(50)	YES		NULL
surname	varchar(50)	YES		NULL
phone	varchar(15)	YES		NULL
email	varchar(100)	YES		NULL
birth_date	date	YES		NULL
country	varchar(50)	YES		NULL
city	varchar(50)	YES		NULL
postal_code	varchar(10)	YES		NULL
address	varchar(255)	YES		NULL

- Taula Dim_companies

Field	Type	Null	Key	Default
company_id	varchar(50)	NO	PRI	NULL
company_name	varchar(100)	YES		NULL
phone	varchar(15)	YES		NULL
email	varchar(100)	YES		NULL
country	varchar(50)	YES		NULL
website	varchar(100)	YES		NULL

- Taula Dim_credit_cards

Field	Type	Null	Key	Default
id	varchar(20)	NO	PRI	NULL
user_id	int	YES		NULL
iban	varchar(50)	YES		NULL
pan	varchar(20)	YES		NULL
pin	varchar(4)	YES		NULL
cvv	varchar(3)	YES		NULL
track1	varchar(150)	YES		NULL
track2	varchar(150)	YES		NULL
expiring_date	date	YES		NULL

- Taula Dim_products

Field	Type	Null	Key	Default
id	int	NO	PRI	NULL
product_name	varchar(100)	YES		NULL
price	decimal(10,2)	YES		NULL
colour	varchar(50)	YES		NULL
weight	decimal(10,2)	YES		NULL
warehouse_id	varchar(50)	YES		NULL

- Taula Dim_fecha:

- Aquesta taula la creo per si hem de fer anàlisis del cicle estacional de les vendes més endavant. Conèixer quin és el dia de la setmana amb més volum, el mes, ... Això, per exemple, pot ajudar a gestionar els recursos dels que disposa l'empresa i fer previsions dels que li caldran.

Field	Type	Null	Key	Default
fecha	date	NO	PRI	NULL
dia	int	YES		NULL
dia_setmana	varchar(10)	YES		NULL
mes	int	YES		NULL
mes_nom	varchar(15)	YES		NULL
trimestre	int	YES		NULL
semestre	int	YES		NULL
año	int	YES		NULL

- Taula Fact_transactions:

- Relacionen les taules de dimensions amb aquesta taula de fets que és la de Fact_transactions. Les relacions són de 1 a N

Field	Type	Null	Key	Default
id	varchar(50)	NO	PRI	NULL
card_id	varchar(20)	YES	MUL	NULL
business_id	varchar(20)	YES	MUL	NULL
timestamp	datetime	YES		NULL
amount	decimal(10,2)	YES		NULL
declined	tinyint(1)	YES		NULL
product_ids	varchar(20)	YES		NULL
user_id	int	YES	MUL	NULL
lat	float	YES		NULL
longitude	float	YES		NULL
fecha_id	date	YES	MUL	NULL

- Creem la base de dades

```

1  /* SPRINT_4
2   Tasca S4.01. Creació de Base de Dades
3   Descripció
4   Partint d'alguns arxius CSV dissenyaràs i crearàs la teva base de dades.
5   Nivell 1
6   Descàrrega els arxius CSV, estudia's i dissenya una base de dades amb un esquema d'estrella que contingui, almenys 4 taules de les quals puguis realitzar les següents consultes:
7   */
8
9   -- Cream la BDD
10 • CREATE DATABASE IF NOT EXISTS sprint_4;
11   -- aseguro la connexió
12 • SET NAMES 'utf8mb4';
13 • SET CHARACTER SET 'utf8mb4';
14 • SET sql_mode = '';
15
16   -- Fem servir la BDD
17 • USE Sprint_4;
18
19   -- Cream les taules
20
21   -- Cream les taules de dimensions
22   -- Taula dimensió usuaris. Aquí afegirem els 3 csv d'users.
23 • CREATE TABLE IF NOT EXISTS Dim_users (

```

Output:

#	Time	Action	Message	Duration / Fetch
1	01:17:00	CREATE DATABASE IF NOT EXISTS sprint_4	1 row(s) affected	0.016 sec
2	01:17:00	SET NAMES utf8mb4	0 row(s) affected	0.000 sec
3	01:17:00	SET CHARACTER SET utf8mb4	0 row(s) affected	0.000 sec
4	01:17:00	SET sql_mode = ''	0 row(s) affected	0.000 sec

- La seleccionem per treballar sobre ella

```

16   -- Fem servir la BDD
17 • USE Sprint_4;
18
19   -- Cream les taules
20
21   -- Cream les taules de dimensions
22   -- Taula dimensió usuaris. Aquí afegirem els 3 csv d'users.
23 • CREATE TABLE IF NOT EXISTS Dim_users (

```

Output:

#	Time	Action	Message	Duration / Fetch
1	01:17:00	CREATE DATABASE IF NOT EXISTS sprint_4	1 row(s) affected	0.016 sec
2	01:17:00	SET NAMES utf8mb4	0 row(s) affected	0.000 sec
3	01:17:00	SET CHARACTER SET utf8mb4	0 row(s) affected	0.000 sec
4	01:17:00	SET sql_mode = ''	0 row(s) affected	0.000 sec
5	01:19:10	USE Sprint_4	0 row(s) affected	0.000 sec

- Comencem a crear les taules de dimensions i carregar-les:
 - Taula dimensió usuaris: creació i el load dels 3 csv

```

22   -- Taula dimensió usuaris. Aquí afegirem els 3 csv d'users.
23 • CREATE TABLE IF NOT EXISTS Dim_users (
24     id INT PRIMARY KEY,
25     name VARCHAR(50),
26     surname VARCHAR(50),
27     phone VARCHAR(15),
28     email VARCHAR(100),
29     birth_date VARCHAR(13),
30     country VARCHAR(50),
31     city VARCHAR(50),
32     postal_code VARCHAR(10),
33     address VARCHAR(255)
34 );
35
36   -- Importar Dim_Usuari
37   -- Primer importem "users_usa.csv"
38
39 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv"
40 INTO TABLE Dim_users
41 FIELDS TERMINATED BY ','
42 ENCLOSED BY ''
43 LINES TERMINATED BY '\r\n'
44 IGNORE 1 ROWS
45 (id, name, surname, phone, email, birth_date, country, city, postal_code, address)
46 SET birth_date = STR_TO_DATE(birth_date, "%b %d, %Y");
47

```

Output:

#	Time	Action	Message
1	01:17:00	CREATE DATABASE IF NOT EXISTS sprint_4	1 row(s) affected
2	01:17:00	SET NAMES utf8mb4	0 row(s) affected
3	01:17:00	SET CHARACTER SET utf8mb4	0 row(s) affected
4	01:17:00	SET sql_mode = ''	0 row(s) affected
5	01:19:10	USE Sprint_4	0 row(s) affected
6	01:20:39	CREATE TABLE IF NOT EXISTS Dim_users (id INT PRIMARY KEY, name VARCHAR(50), surname VARCHAR(50), phone VARCHAR(15), email VARCHAR(100), birth_date VARCHAR(13), country VARCHAR(50), city VARCHAR(50), postal_code VARCHAR(10), address VARCHAR(255))	0 row(s) affected
7	01:20:49	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv" INTO TABLE Dim_users FIELDS TERMINATED BY ',' ENCLOSED BY '' LINES TERMINATED BY '\r\n' IGNORE 1 ROWS (id, name, surname, phone, email, birth_date, country, city, postal_code, address)	150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0

```

48 -- Repetir-ho per "users_uk.csv"
49 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv"
50 INTO TABLE Dim_users
51 FIELDS TERMINATED BY ','
52 ENCLOSED BY ""
53 LINES TERMINATED BY '\r\n'
54 IGNORE 1 ROWS
55 (id, name, surname, phone, email, birth_date, country, city, postal_code, address)
56 SET birth_date = STR_TO_DATE(birth_date, "%b %d, %Y");
57
58 -- Repetir-ho per "users_ca.csv"
59 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv"
60 INTO TABLE Dim_users
61 FIELDS TERMINATED BY ','
62 ENCLOSED BY ""
63 LINES TERMINATED BY '\r\n'
64 IGNORE 1 ROWS
65 (id, name, surname, phone, email, birth_date, country, city, postal_code, address)
66 SET birth_date = STR_TO_DATE(birth_date, "%b %d, %Y");
67
68 -- Modifiquem la columna de birth_date de tipus VARCHAR a DATE

```

Action Output

#	Time	Action	Message
3	01:17:00	SET CHARACTER SET utf8mb4'	0 row(s) affected
4	01:17:00	SET sql_mode = ''	0 row(s) affected
5	01:19:10	USE Sprint_4	0 row(s) affected
6	01:20:39	CREATE TABLE IF NOT EXISTS Dim_users (id INT PRIMARY KEY, name VARCHAR(50), surname VARCHAR(50), phone VARCHAR... 0 row(s) affected	0 row(s) affected
7	01:20:49	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0	150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0
8	01:21:56	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0	50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0
9	01:22:08	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0	75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0

■ Modifiquem el tipus de dada de "birth_date": de VARCHAR a DATE

```

68 -- Modifiquem la columna de birth_date de tipus VARCHAR a DATE
69 • ALTER TABLE dim_users
70 MODIFY birth_date DATE;
71

```

Action Output

#	Time	Action	Message
4	01:17:00	SET sql_mode = ''	0 row(s) affected
5	01:19:10	USE Sprint_4	0 row(s) affected
6	01:20:39	CREATE TABLE IF NOT EXISTS Dim_users (id INT PRIMARY KEY, name VARCHAR(50), surname VARCHAR(50), phone VARCHAR... 0 row(s) affected	0 row(s) affected
7	01:20:49	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0	150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0
8	01:21:56	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0	50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0
9	01:22:08	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0	75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0
10	01:23:13	ALTER TABLE dim_users MODIFY birth_date DATE	275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0

○ Taula dimensió empreses: creació i el load

```

72 -- Taula dimensió empreses
73 • CREATE TABLE IF NOT EXISTS Dim_companies (
74     company_id VARCHAR(50) PRIMARY KEY,
75     company_name VARCHAR(100),
76     phone VARCHAR(15),
77     email VARCHAR(100),
78     country VARCHAR(50),
79     website VARCHAR(100)
80 );
81
82 -- Dim_companies
83 -- Carreguem "companies.csv"
84 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/companies.csv"
85 INTO TABLE Dim_companies
86 FIELDS TERMINATED BY ','
87 ENCLOSED BY ""
88 LINES TERMINATED BY '\r\n'
89 IGNORE 1 ROWS
90 (company_id, company_name, phone, email, country, website);
91

```

Action Output

#	Time	Action	Message
1	01:17:00	CREATE DATABASE IF NOT EXISTS sprint_4	1 row(s) affected
2	01:17:00	SET NAMES utf8mb4'	0 row(s) affected
3	01:17:00	SET CHARACTER SET utf8mb4'	0 row(s) affected
4	01:17:00	SET sql_mode = ''	0 row(s) affected
5	01:19:10	USE Sprint_4	0 row(s) affected
6	01:20:39	CREATE TABLE IF NOT EXISTS Dim_users (id INT PRIMARY KEY, name VARCHAR(50), surname VARCHAR(50), phone VARCHAR... 0 row(s) affected	0 row(s) affected
7	01:20:49	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0	150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0
8	01:21:56	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0	50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0
9	01:22:08	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv" INTO TABLE Dim_users FIELDS TERMINATED BY... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0	75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0
10	01:23:13	ALTER TABLE dim_users MODIFY birth_date DATE	275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
11	01:24:36	CREATE TABLE IF NOT EXISTS Dim_companies (company_id VARCHAR(50) PRIMARY KEY, company_name VARCHAR(100), phone... 0 row(s) affected	0 row(s) affected
12	01:24:46	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/companies.csv" INTO TABLE Dim_companies FIELDS TERMINATE... 100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0	100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0

- Taula dimensió targetes de crèdit: creació i el load
 - Aquesta és una de les taules que al fer el load no dona error però no carrega cap dada.

```

92  -- Taula dimensió targetes de crèdit
93 • CREATE TABLE IF NOT EXISTS Dim_credit_cards (
94   id VARCHAR(20) PRIMARY KEY,
95   user_id INT,
96   iban VARCHAR(50),
97   pan VARCHAR(20),
98   pin VARCHAR(4), -- Per si comença per 0 que no l'elimini
99   cvv VARCHAR(3), -- Per si comença per 0 que no l'elimini
100  track1 VARCHAR(150),
101  track2 VARCHAR(150),
102  expiring_date VARCHAR(15)
103 );
104
105 -- Dim_credit_cards
106 -- Carreguem "credit_cards.csv"
107 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/credit_cards.csv"
108 INTO TABLE Dim_credit_cards
109 FIELDS TERMINATED BY ','
110 ENCLOSED BY ""
111 LINES TERMINATED BY '\r\n'
112 IGNORE 1 ROWS
113 (id, user_id, iban, @pan_temp, pin, cvv, track1, track2, @expiring_date_temp)
114 SET
115   pan = REPLACE(@pan_temp, ' ', ''),
116   expiring_date = STR_TO_DATE (@expiring_date_temp, '%m/%d/%y');
117
<

```

Output:

#	Action	Time	Message
8	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv" INTO TABLE Dim_users	01:21:56	FIELDS TERMINATED BY ... 50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0
9	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv" INTO TABLE Dim_users	01:22:08	FIELDS TERMINATED BY ... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0
10	ALTER TABLE dim_users MODIFY birth_date DATE	01:23:13	275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
11	CREATE TABLE IF NOT EXISTS Dim_companies (company_id VARCHAR(50) PRIMARY KEY, company_name VARCHAR(100), phon...	01:24:36	0 row(s) affected
12	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/companies.csv" INTO TABLE Dim_companies	01:24:46	FIELDS TERMINATED BY ... 100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0
13	CREATE TABLE IF NOT EXISTS Dim_credit_cards (id VARCHAR(20) PRIMARY KEY, user_id INT, iban VARCHAR(50), pan V...	01:26:06	0 row(s) affected
14	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/credit_cards.csv" INTO TABLE Dim_credit_cards	01:26:18	FIELDS TERMINATED BY ... 0 row(s) affected Records: 0 Deleted: 0 Skipped: 0 Warnings: 0

No ha donat error i no ha carregat les dades. Passo a fer-ho amb Wizzard.

- Carreguem les dades amb wizzard

The screenshot shows the MySQL Workbench Table Data Import window. At the top, it says "Table Data Import". Below that is a section titled "Import Results" which contains the following text:

```

File C:\Users\Albert\Documents\Personal\MOOCs\ITACADEMY\Data Analytics\Sprint 4\credit_cards.csv was imported in 3.333 s
Table sprint_4.dim_credit_cards has been used
275 records imported

```

At the bottom, there is a log of SQL statements and their execution times:

Time	Statement	OK
18 01:29:08	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK
19 01:29:08	SHOW DATABASES	OK
20 01:29:28	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK
21 01:29:28	SHOW COLUMNS FROM 'sprint_4.dim_credit_cards'	OK
22 01:29:45	PREPARE stmt FROM 'INSERT INTO `sprint_4.dim_credit_cards` (`id`, `user_id`, `iban`, `pan`, `pin`, `cvv`, `track1`, `track2`, `expiring_date`) VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?)'	OK
23 01:29:49	DEALLOCATE PREPARE stmt	OK

- Realitzem les modificacions del tipus de dada i l'actualització de les dades que ens ha quedat pendent al haver-ho fet amb Wizzard.

```

126
127  /*
128  Com que no funciona amb Workbench ni amb Command Line Client o faig amb Wizard.
129  Però hem de treballar el format data de la columna expiring_date i eliminar espais de la columna pan
130 */
131 -- Actualitzar les dades existents
132 • UPDATE Dim_credit_cards
133 SET expiring_date = STR_TO_DATE(expiring_date, '%m/%d/%y');
134
135 -- Modificar la columna per a que sigui de tipus DATE
136 • ALTER TABLE Dim_credit_cards
137 MODIFY expiring_date DATE;
138
139 -- Eliminar espai columna pan
140 • UPDATE Dim_credit_cards
141 SET pan = REPLACE(pan, ' ', '');
142
143 -- Taula dimensió productes
<

```

Output

#	Time	Action	Message
26	01:32:34	SHOW CREATE TABLE Dim_credit_cards	1 row(s) returned
27	01:33:35	UPDATE Dim_credit_cards SET expiring_date = STR_TO_DATE(expiring_date, '%m/%d/%y')	275 row(s) affected Rows matched: 275 Changed: 275 Warnings: 0
28	01:33:43	ALTER TABLE Dim_credit_cards MODIFY expiring_date DATE	275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
29	01:33:53	UPDATE Dim_credit_cards SET pan = REPLACE(pan, ' ', '')	187 row(s) affected Rows matched: 275 Changed: 187 Warnings: 0

- Taula dimensió productes: creació i el load

- Aquesta és l'altre taula que no ha carregat dades i no ha donat error.

```

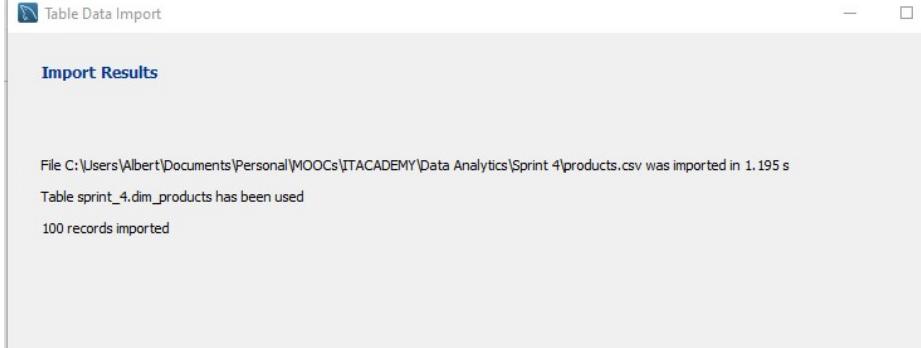
143 -- Taula dimensió productes
144 • CREATE TABLE IF NOT EXISTS Dim_products (
145     id INT PRIMARY KEY,
146     product_name VARCHAR(100),
147     price VARCHAR(10),
148     colour VARCHAR(50),
149     weight DECIMAL(10, 2),
150     warehouse_id VARCHAR(50)
151 );
152
153 -- Dim_products
154 -- Carreguem "products.csv"
155 • LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/products.csv"
156 INTO TABLE Dim_products
157 FIELDS TERMINATED BY ','
158 ENCLOSED BY ''
159 LINES TERMINATED BY '\r\n'
160 IGNORE 1 ROWS
161 (id, product_name, @price_temp, colour, weight, warehouse_id)
162 SET price = REPLACE(@price_temp, '$', '');
163
164 /* 
<

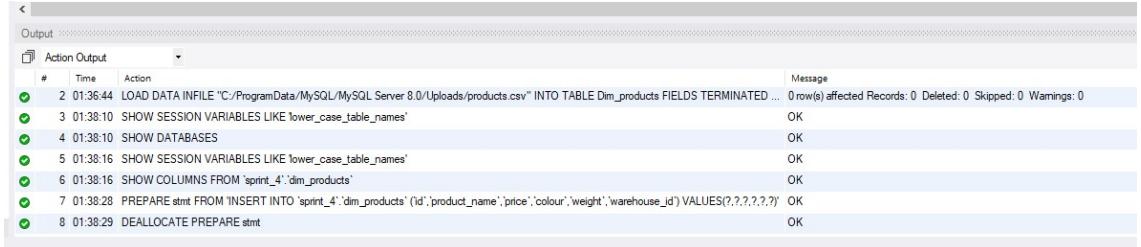
```

Output

#	Time	Action	Message
1	01:36:36	CREATE TABLE IF NOT EXISTS Dim_products (id INT PRIMARY KEY, product_name VARCHAR(100), price VARCHAR(10), colour V... 0 row(s) affected	
2	01:36:44	LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/products.csv" INTO TABLE Dim_products FIELDS TERMINATED B... 0 row(s) affected Records: 0 Deleted: 0 Skipped: 0 Warnings: 0	

- Carreguem les dades amb Wizard





- Eliminarem el símbol del “\$” i modificarem el tipus de dada de la columna “price”.

```

164  /* 
165   Aquí em passa el mateix que amb la taula dim_credit_cards. No puc carregar les dades amb codi i ho faig amb Wizzard.
166   Com que no he pogut fer el replace del $ en la pujada ho faig a continuació.
167 */
168
169  -- Eliminar espai columna pan
170 • UPDATE Dim_products
171   SET price = REPLACE(price, '$', '');
172
173  -- Modifiquem el tipus de dada de la columna
174 • ALTER TABLE Dim_products
175   MODIFY price DECIMAL(10,2);
176
177  -- Taula dimensió fecha

```



- Taula dimensió data : creació

```

177      -- Taula dimensió fecha
178 • CREATE TABLE IF NOT EXISTS Dim_fecha (
179     fecha DATE PRIMARY KEY,
180     dia INT,
181     dia_setmana VARCHAR(10),
182     mes INT,
183     mes_nom VARCHAR(15),
184     trimestre INT,
185     semestre INT,
186     año INT
187 );
188
189 -- Creem un procediment per carregar les dades a la taula
190
191 -- Em dona talls de connexió i ho incremento a 8 hores
<
Output ::::::::::::::::::::
Action Output
# Time Action
10 01:42:23 ALTER TABLE Dim_products MODIFY price DECIMAL(10,2) Message
100 row(s) affected Records: 100 Duplicates: 0 Warnings: 0
11 01:43:35 CREATE TABLE IF NOT EXISTS Dim_fecha ( fecha DATE PRIMARY KEY, dia INT, dia_setmana VARCHAR(10), mes INT, mes_nom VARCHAR(15), trimestre INT, semestre INT, año INT ) Message
0 row(s) affected

```

- Crear procediment per carregar les dades
 - Abans per evitar els problemes de timeout, ho incremento

```

189 -- Creem un procediment per carregar les dades a la taula
190
191 -- Em dona talls de connexió i ho incremento a 8 hores
192 • SET GLOBAL wait_timeout = 28800;
193 • SET GLOBAL interactive_timeout = 28800;
194
<
Output ::::::::::::::::::::
Action Output
# Time Action
12 01:45:09 SET GLOBAL wait_timeout = 28800 Message
0 row(s) affected
13 01:45:14 SET GLOBAL interactive_timeout = 28800 Message
0 row(s) affected

```

- Procediment: LoadDimFecha

```

199 • CREATE PROCEDURE LoadDimFecha()
200 ⚒ BEGIN
201   DECLARE current_fecha DATE;
202   DECLARE limit_date DATE;
203   SET current_fecha = '2020-01-01';
204
205   WHILE current_fecha <= '2030-12-31' DO
206     SET limit_date = DATE_ADD(current_fecha, INTERVAL 999 DAY);
207
208     WHILE current_fecha <= limit_date AND current_fecha <= '2030-12-31' DO
209       INSERT INTO Dim_fecha (fecha, dia, dia_semana, mes, mes_nom, trimestre, semestre, año)
210       VALUES (
211         current_fecha,
212         DAY(current_fecha),
213         DAYNAME(current_fecha),
214         MONTH(current_fecha),
215         MONTHNAME(current_fecha),
216         QUARTER(current_fecha),
217         CASE
218           WHEN QUARTER(current_fecha) IN (1, 2) THEN 1
219           ELSE 2
220         END,
221         YEAR(current_fecha)
222       );
223
224       SET current_fecha = DATE_ADD(current_fecha, INTERVAL 1 DAY);
225     END WHILE;
226   END WHILE;
227 END //;
228
229 DELIMITER ;

```

Output

#	Time	Action	Message
✓ 12	01:45:09	SET GLOBAL wait_timeout = 28800	0 row(s) affected

- Executar el procediment. El haver incrementat el timeout evita que falli.

```

231 • -- Executem el procediment per carregar les dades.
232 CALL LoadDimFecha();

```

Output

#	Time	Action	Message
✓ 1	01:49:08	CALL LoadDimFecha()	1 row(s) affected

- Ara crearem la taula de fets:
 - Taula de transaccions: creació i el load.

```

234    -- Creem la taula de fets
235 • CREATE TABLE IF NOT EXISTS Fact_transactions (
236     id VARCHAR(50) PRIMARY KEY,
237     card_id VARCHAR(20),
238     business_id VARCHAR(20),
239     timestamp DATETIME,
240     amount DECIMAL(10,2),
241     declined BOOLEAN,
242     product_ids VARCHAR(20),
243     user_id INT,
244     lat FLOAT,
245     longitude FLOAT,
246     fecha_id DATE,
247     FOREIGN KEY (user_id) REFERENCES Dim_users(id),
248     FOREIGN KEY (business_id) REFERENCES Dim_companies(company_id),
249     FOREIGN KEY (card_id) REFERENCES Dim_credit_cards(id),
250     FOREIGN KEY (fecha_id) REFERENCES Dim_fecha(fecha)
251 );
252
<

```

Output

#	Time	Action	Message
1	01:52:10	CREATE TABLE IF NOT EXISTS Fact_transactions (id VARCHAR(50) PRIMARY KEY, card_id VARCHAR(20), business_id VARCHAR(20), ...)	0 row(s) affected

Durant el procés de carrega li diem que ignori l'última columna "Fecha_id", ja que ho farem en el següent pas.

```

253    -- Carreguem la taula transactions.csv
254
255 • LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv'
256   INTO TABLE fact_transactions
257   FIELDS TERMINATED BY ';'
258   ENCLOSED BY ''
259   LINES TERMINATED BY '\r\n'
260   IGNORE 1 ROWS
261   (id, card_id, business_id, timestamp, amount, declined, product_ids, user_id, lat, longitude, @ignore);
262
263   -- afegim les dades a la columna fecha_id de fact_transactions amb la data de Timestamp i que estigui a la taula Dim_fecha
264
<

```

Output

#	Time	Action	Message
1	01:52:10	CREATE TABLE IF NOT EXISTS Fact_transactions (id VARCHAR(50) PRIMARY KEY, card_id VARCHAR(20), business_id VARCHAR(20), ...)	0 row(s) affected
2	01:54:37	LOAD DATA INFILE C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv' INTO TABLE fact_transactions FIELDS TERMINAT...	587 row(s) affected Records: 587 Deleted: 0 Skipped: 0 Warnings: 0

■ Afegim les dades a la columna "fecha_id".

```

262   -- afegim les dades a la columna fecha_id de fact_transactions amb la data de Timestamp i que estigui a la taula Dim_fecha
263
264
265 • UPDATE Fact_transactions
266   SET fecha_id = DATE(TIMESTAMP)
267   WHERE DATE(TIMESTAMP) IN (SELECT fecha FROM Dim_fecha);
268
269 /* SPRINT 4
270 Nivell 1 - Exercici 1
<

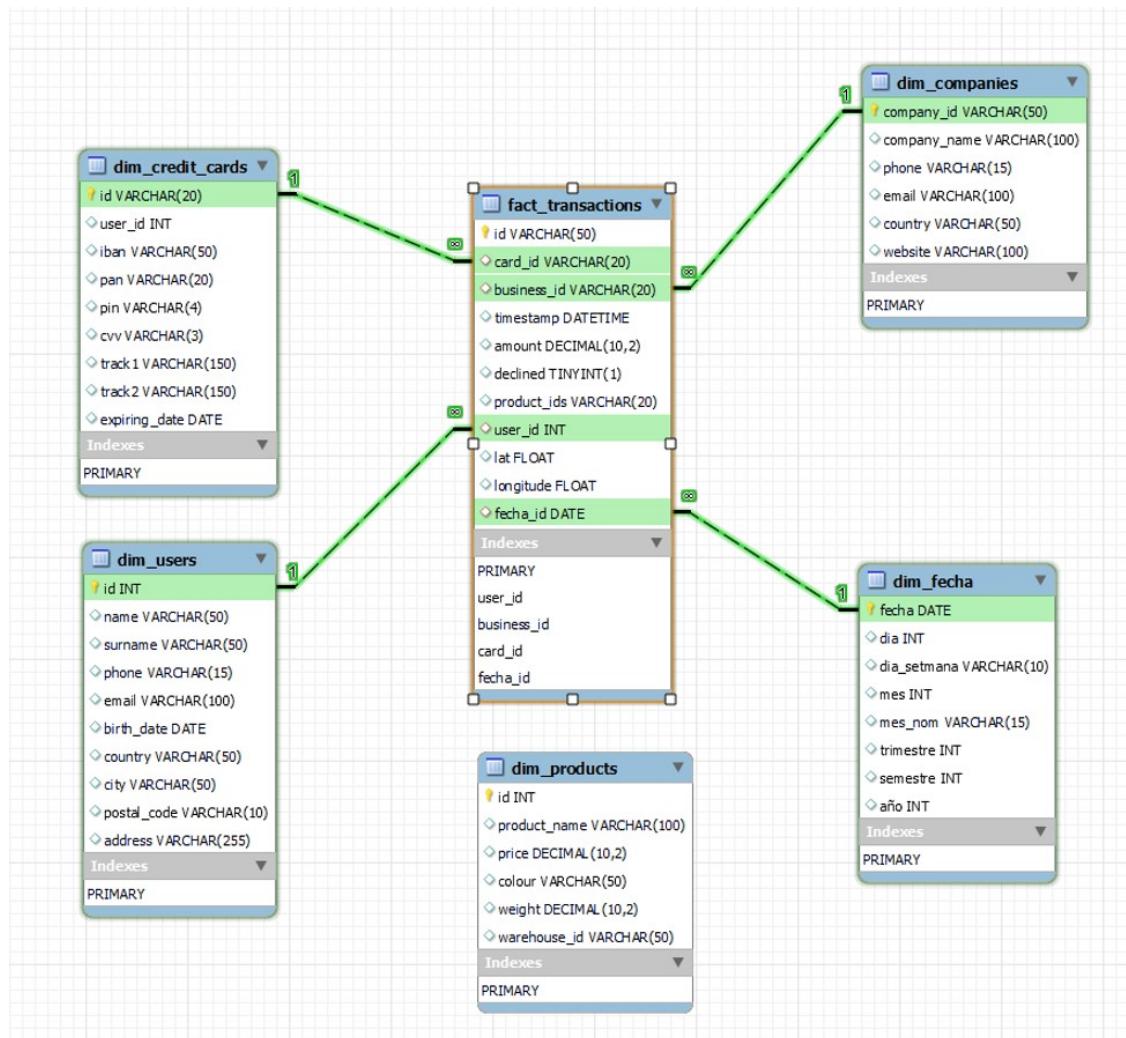
```

Output

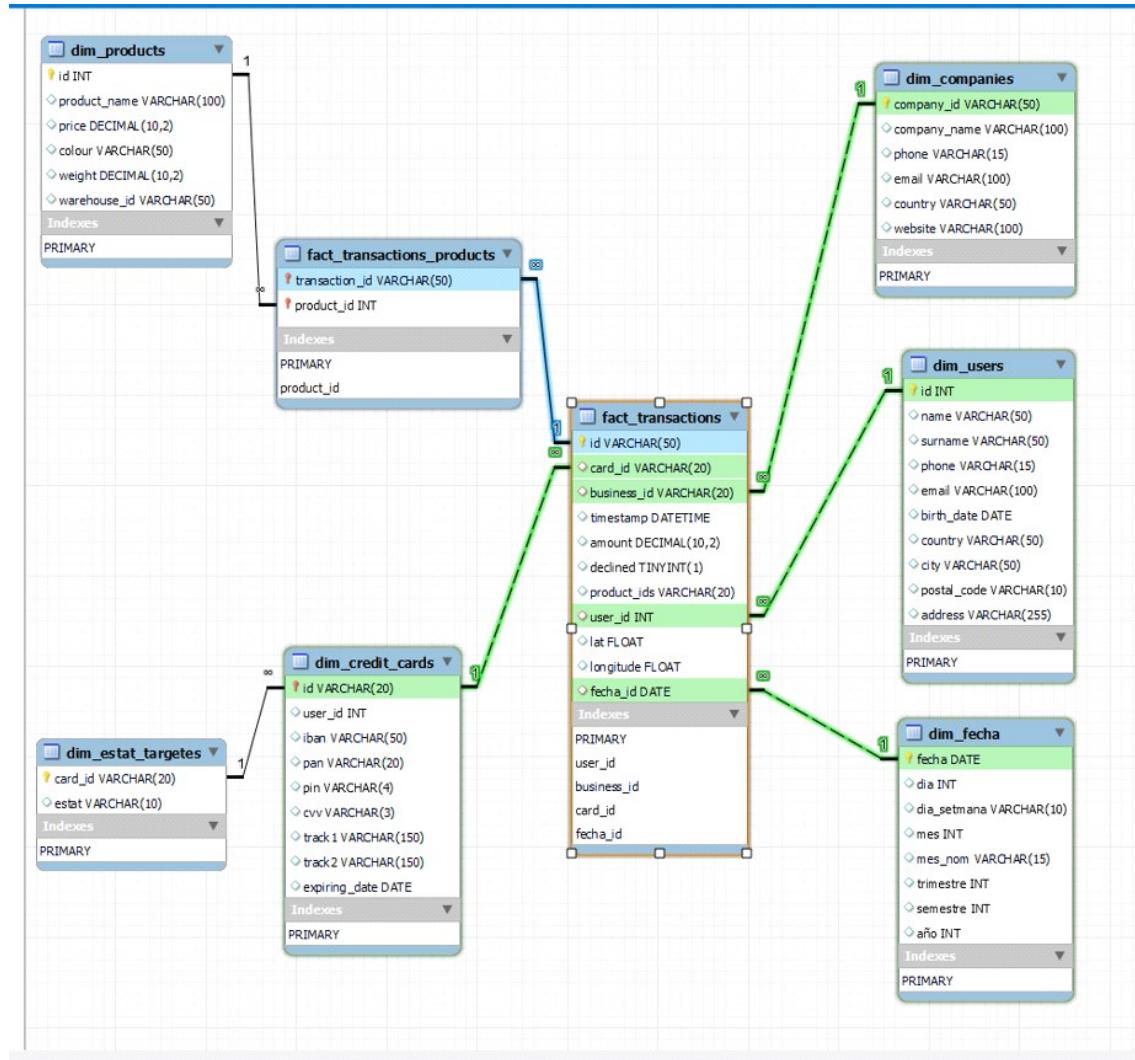
#	Time	Action	Message
1	01:55:52	UPDATE Fact_transactions SET fecha_id = DATE(TIMESTAMP) WHERE DATE(TIMESTAMP) IN (SELECT fecha FROM Dim_fecha)	587 row(s) affected Rows matched: 587 Changed: 587 Warnings: 0

- Adjuntem el diagrama del esquema estrella amb les taules creades i les seves relacions.

En aquest primer diagrama encara no s'ha connectat la taula dim_products, al haver-se de realitzar més endavant.



Aquest és el diagrama que ha quedat al final.



Comprovació de taules creades

- Dim_users.

	id	name	surname	phone	email	birth_date	country	city	postal_code	address
▶	1	Zeus	Gamble	1-282-581-0551	interdum.enim@protonmail.edu	1985-11-17	United States	Lowell	73544	348-7818 Sagittis St.
2	Garrett	Mcconnell	(718) 257-2412	integer.vitae.nibh@protonmail.org	1992-08-23	United States	Des Moines	59464	903 Sit Ave	
3	Ciaran	Harrison	(522) 598-1365	interdum.feugiat@aol.org	1998-04-29	United States	Columbus	56518	736-2063 Tellus St.	
4	Howard	Stafford	1-411-740-3269	ornare.egestas@icloud.edu	1989-02-18	United States	Kailua	77417	Ap #545-2244 Erat. Rd.	
5	Hayfa	Pierce	1-554-541-2077	et.malesuada.fames@hotmail.org	1998-09-26	United States	Sandy	31564	341-2821 Ultrices Av.	
6	Joel	Tyson	(718) 288-8020	gravida.nunc.sed@yahoo.ca	1989-10-15	United States	Nashville	96838	888-2799 Amet Street	
7	Rafael	Jimenez	(817) 689-0478	eget@outlook.ca	1981-12-04	United States	Hillsboro	29874	8627 Malesuada Rd.	
8	Nissim	Franks	(692) 157-3469	egestas.aliquam.fringilla@google.ca	1993-08-01	United States	Jackson	61750	Ap #251-7144 Integer St.	
9	Mannix	Mcclain	(590) 883-2184	aliquam.nisi@outlook.com	1987-01-24	United States	Richmond	35987	647-3080 Lacus. St.	
10	Robert	Mccarthy	(324) 746-6771	fermentum@protonmail.com	1984-04-30	United States	Eugene	85526	P.O. Box 773, 3594 Ornare St.	
11	Joan	Baird	(981) 429-8106	et@outlook.net	1990-02-25	United States	Lincoln	35211	P.O. Box 687, 8917 Ligula St.	
12	Benedict	Wheeler	1-515-824-2855	tincidunt.donec.vitae@hotmail.co.uk	1999-08-06	United States	Lewiston	92393	748-8694 Porttitor Avenue	
13	Allena	Stanton	1-927-753-6488	nroin.enet@protonmail.ca	1990-05-19	United States	Kearney	14447	4457 Ante. Av.	
dim_users 120 ×										

Output: Action Output
Time Action
1 03:12:59 SELECT * FROM dim_users

Message: 275 row(s) returned

- Dim_companies.

88

```
89 •  SELECT * FROM dim_companies;
```

90

Result Grid | Filter Rows: [] | Edit: [] | Export/Import: [] | Wrap Cell Content: []

	company_id	company_name	phone	email	country	website
▶	b-2222	Ac Fermentum Incorporated	06 85 56 52 33	donec.porttitor.tellus@yahoo.net	Germany	https://instagram.com/site
	b-2226	Magna A Neque Industries	04 14 44 64 62	risus.donec.nibh@cloud.org	Australia	https://whatsapp.com/group/9
	b-2230	Fusce Corp.	08 14 97 58 85	risus@protonmail.edu	United States	https://pinterest.com/sub/cars
	b-2234	Convallis In Incorporated	06 66 57 29 50	mauris.ut@aol.co.uk	Germany	https://cnn.com/user/110
	b-2238	Ante Iaculis Nec Foundation	08 23 04 99 53	sed.dictum.pronin@outlook.ca	New Zealand	https://netfix.com/settings
	b-2242	Donec Ltd	01 25 51 37 37	at.iaculis@hotmail.co.uk	Norway	https://nytimes.com/user/110
	b-2246	Sed Nunc Ltd	02 62 64 73 48	nibh@yahoo.org	United Kingdom	https://cnn.com/one
	b-2250	Amet Nulla Donec Corporation	07 15 25 14 74	mattis.integer.eu@protonmail.net	Italy	https://netflix.com/sub/cars
	b-2254	Nascetur Ridiculus Mus Inc.	06 26 87 61 84	suspendisse.dui@icloud.net	United States	https://ebay.com/sub
	b-2258	Vestibulum Lorem PC	02 02 87 33 40	aenean.massा.integer@aol.net	Belgium	https://pinterest.com/sub/cars
	b-2262	Gravida Sagittis LLP	03 81 28 33 97	turpis.vitae@google.ca	Sweden	https://naver.com/site
	b-2266	Mus Aenean Eget Foundation	06 25 15 52 43	mi.duis@hotmail.net	Sweden	https://instagram.com/group/9
	b-2270	Dis Parturient Institute	05 36 29 78 74	nurus@protonmail.orca	Ireland	https://noodle.com/one

dim_companies 121 ×

Output:

Action Output

#	Time	Action
✓	1 03:16:45	SELECT * FROM dim_companies

Message: 100 row(s) returned

- Dim_credit_cards.

90

```
91 •  SELECT * FROM dim_credit_cards;
```

92

Result Grid | Filter Rows: [] | Edit: [] | Export/Import: [] | Wrap Cell Content: []

	id	user_id	iban	pan	pin	cvv	track1	track2	expiring_date
▶	CCU-2938	275	TR301950312213576817638661	5424465566813633	3257	984	%88383712448554646~WovsxeDpvieW~8604...	%87653863056044187~800716333673	2022-10-30
	CdU-2945	274	DO2685476374853747516568689	5142423821948828	9080	887	%8462131609958661~UftufyrsSelmxn~06106...	%84149568437843501=5107140330?1	2023-08-24
	CCU-2952	273	BE45IVQL52710525608255	455645355287	4598	438	%82183285104307501~CddyyttJxwfdg~5907...	%86778580257827162=690685974007	2021-06-29
	CCU-2959	272	CR7242477244335841535	372461377349375	3583	667	%87281111956795320~XocddjBkced~09016...	%84246154489281853=280522391678	2023-02-24
	CCU-2966	271	BG72LKTQ15627628377363	4485668867472765	4900	130	%84728932322756223~JlqvsFBwmqj~7202...	%8231871115999881~90982157845	2024-10-29
	CCU-2973	270	PT8780652813592429456346	545865454343384	8760	887	%84761405253275637~HjnipoBlejr~7108515...	%87816169831446746~1310277229	2025-01-30
	CCU-2980	269	DE3924188183086277136	4024007145845969	5075	596	%87320483593870549~Ookzqxrhpsased~4901...	%82474313962214151=041221913175	2022-07-24
	CCU-2987	268	GE89681434837746781813	376374768776666	2298	797	%84750646345146674~PjmlyrGwrbf~83051...	%85441935173418615=41037453677	2023-10-31
	CCU-2994	267	BH62714428365066765294	34283273252993	7545	595	%8185379794015674~GmqohytJotqrn~2507...	%8411467473024349=6506800955074	2022-02-28
	CCU-3001	266	CY49087426654774581266832110	5117229248332244	9562	867	%8622728875672864~AwiifPmgvy~2808...	%83429355750963453=530526830573	2022-09-16
	CCU-3008	265	LU507216693616119230	448574464433884	1856	740	%87182449430529226~MkoutyhTfdvo~1708...	%86235123731781366=40563816678	2025-04-05
	CCU-3015	264	PS119398216295715968342456821	378466223317389	3246	822	%85776250106724742~QvvzrcCwrov~530...	%83561372148267521=20547316677	2022-01-31
	CCU-3022	263	GT9169516285055977473121857	516413794873951	5610	347	%88456672214337875~XmwohenXwnxht~590...	%87761154174174063=201423342974	2025-04-25

dim_credit_cards 123 ×

Output:

Action Output

#	Time	Action
✓	1 03:18:58	SELECT * FROM dim_credit_cards

Message: 275 row(s) returned

- Dim_products.

92

```
93 •  SELECT * FROM dim_products;
```

94

Result Grid | Filter Rows: [] | Edit: [] | Export/Import: [] | Wrap Cell Content: []

	id	product_name	price	colour	weight	warehouse_id
▶	1	Direwolf Stannis	161.11	#7c7c7c	1.00	WH-4
	2	Tarly Stark	9.24	#919191	2.00	WH-3
	3	duel tourney Lannister	171.13	#d8d8d8	1.50	WH-2
	4	warden south duel	71.89	#111111	3.00	WH-1
	5	skywalker ewok	171.22	#dbdbdb	3.20	WH-0
	6	dooku solo	136.60	#c4c4c4	0.80	WH-1
	7	north of Casterly	63.33	#b7b7b7	0.60	WH-2
	8	Winterfell	32.37	#383838	1.40	WH-3
	9	Winterfell	76.40	#b5b5b5	1.20	WH-4
	10	Karstark Dorne	119.52	#f4f4f4	2.40	WH-5
	11	Karstark Dorne	49.70	#141414	2.70	WH-6
	12	duel Direwolf	181.60	#a8a8a8	2.10	WH-7
	13	nalmatine chewbacca	139.59	#7b7b7b	1.00	WH-8

dim_products 124 ×

Output:

Action Output

#	Time	Action
✓	1 03:19:38	SELECT * FROM dim_products

Message: 100 row(s) returned

- Dim_Fecha.

94

95 • `SELECT * FROM dim_fecha;`

96

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows: |

fecha	dia	dia_semana	mes	mes_nom	trimestre	semestre	año
2020-01-01	1	Wednesday	1	January	1	1	2020
2020-01-02	2	Thursday	1	January	1	1	2020
2020-01-03	3	Friday	1	January	1	1	2020
2020-01-04	4	Saturday	1	January	1	1	2020
2020-01-05	5	Sunday	1	January	1	1	2020
2020-01-06	6	Monday	1	January	1	1	2020
2020-01-07	7	Tuesday	1	January	1	1	2020
2020-01-08	8	Wednesday	1	January	1	1	2020
2020-01-09	9	Thursday	1	January	1	1	2020
2020-01-10	10	Friday	1	January	1	1	2020
2020-01-11	11	Saturday	1	January	1	1	2020
2020-01-12	12	Sunday	1	January	1	1	2020
2020-01-13	13	Monday	1	January	1	1	2020

dim_fecha 125 ×

Output:

Action Output:

Time Action

1 03:21:27 SELECT * FROM dim_fecha

Message: 4018 row(s) returned

- Fact_transactions.

97 • `SELECT * FROM fact_transactions;`

98

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: |

id	card_id	business_id	timestamp	amount	declined	product_ids	user_id	lat	longitude	fecha_id
02C6201E-D90A-1859-B4EE-88D2986D3802	CcU-2938	b-2362	2021-08-28 23:42:24	466.92	0	71, 1, 19	92	81.9185	-12.5276	2021-08-28
0466A42E-47CF-8D24-FD01-C0B689713128	CcU-4219	b-2302	2021-07-26 07:29:18	49.53	0	47, 97, 43	170	-43.9695	-117.525	2021-07-26
063FBA79-99EC-66FB-29F7-25726D1764A5	CcU-2987	b-2250	2022-01-06 21:25:27	92.61	0	47, 67, 31, 5	275	-81.2227	-129.05	2022-01-06
0668296C-CDB9-A883-76BC-2E4C44F8C8AE	CcU-3743	b-2618	2022-01-26 02:07:14	394.18	0	89, 83, 79	265	-34.3593	-100.556	2022-01-26
06CD9A45-9842-6864-DDD-A5E394FEBAA9	CcU-2959	b-2346	2021-10-26 23:00:01	279.93	0	43, 31	92	33.7381	158.298	2021-10-26
07A46D48-31A3-7E87-65B9-0DA902AD109F	CcU-3225	b-2386	2021-06-28 21:11:42	340.87	1	47, 23	272	38.8342	92.1905	2021-06-28
09DE92CE-6F27-2B87-13B5-938582B38E82	CcU-3071	b-2298	2021-05-11 20:40:06	303.05	1	67, 7	275	71.1706	10.5757	2021-05-11
0A476ED9-0C13-1962-F87B-03563924B539	CcU-4359	b-2302	2022-02-26 20:33:54	430.49	0	29, 41, 11	221	-56.4901	114.801	2022-02-26
0BE88087-9066-1707-C64B-9DCE71914B5	CcU-3141	b-2338	2022-03-04 14:54:35	288.81	1	19, 41, 29, 3	272	23.3264	-13.6037	2022-03-04
0C7C3A33-9947-3BC1-846D-7BE300D17598	CcU-3309	b-2434	2021-04-10 20:58:41	103.44	1	89, 31	272	63.3615	-68.6667	2021-04-10
0CE957A6-CCAA-2B7A-6839-8A4B1B324853	CcU-3435	b-2506	2022-02-02 07:29:36	428.69	1	83, 43, 73, 61	269	-69.3537	-10.26	2022-02-02
0DD2E608-5C9E-D1B3-4999-B99F43AD735A	CcU-2959	b-2234	2021-04-17 05:30:17	252.47	1	7, 47, 17	275	9.68811	130.282	2021-04-17
1017AA59-3D5F-7A4C-1997-D151ARD1FA0A	CrlI-3701	h-261R	2021-11-01 01:07:11	447.11	0	37, 13	267	4.77645	-101.554	2021-11-01

fact_transactions 127 ×

Output:

Action Output:

Time Action

1 03:24:12 SELECT * FROM fact_transactions

Message: 587 row(s) returned

Nivell 1 - Exercici 1

Realitza una subconsulta que mostri tots els usuaris amb més de 30 transaccions utilitzant almenys 2 taules.

En aquest exercici he realitzat la consulta:

- Sense rebutjar les declined.

```

269  /* SPRINT 4
270   Nivell 1 - Exercici 1
271   Realitza una subconsulta que mostri tots els usuaris amb més de 30 transaccions utilitzant almenys 2 taules.
272   */
273   -- Sense rebutjar les declined
274 •  SELECT f.user_id, COUNT(f.id) AS Num_transaccions, (SELECT name FROM dim_users WHERE id = f.user_id) AS Nom_usuari
275   FROM fact_transactions AS f
276   GROUP BY user_id
277   HAVING Num_transaccions >30
278   ORDER BY Num_transaccions DESC;
279

```

Result 19 x

Output

Action Output

user_id	Num_transaccions	Nom_usuari
272	76	Hedwig
267	52	Ocean
275	48	Kenyon
92	39	Lynn

Message

507 row(s) affected Rows matched: 507 Changed: 507 Warnings: 0

2 02:02:47 SELECT f.user_id, COUNT(f.id) AS Num_transaccions, (SELECT name FROM dim_users WHERE id = f.user_id) AS Nom_usuari FROM fact_transactions AS f GROUP BY user_id HAVING Num_transaccions >30 ORDER BY Num_transaccions DESC;

2. Rebutjar les declined.

```

280   -- Rebutjan les declined
281 •  SELECT f.user_id, COUNT(f.id) AS Num_transaccions, (SELECT name FROM dim_users WHERE id = f.user_id) AS Nom_usuari
282   FROM fact_transactions AS f
283   WHERE f.declined = 0
284   GROUP BY user_id
285   HAVING Num_transaccions >30
286   ORDER BY Num_transaccions DESC;
287
288  /* SPRINT 4

```

Result 20 x

Output

Action Output

user_id	Num_transaccions	Nom_usuari
92	39	Lynn
267	39	Ocean
272	38	Hedwig

Message

3 row(s) returned

Comparant el resultat es pot començar a veure la solvència de cada un dels clients. Caldrien els motius dels declined per filar més prim i veure que succeeix amb cada client, aquesta dada no la tenim a les que ens han proporcionat.

Nivell 1 - Exercici 2

Mostra la mitjana d'amount per IBAN de les targetes de crèdit a la companyia Donec Ltd, utilitza almenys 2 taules.

Per ser coherents, continuarem separant entre rebutjar o no les declined.

1. Sense rebutjar les declined.

```

288  /* SPRINT 4
289   Nivell 1 - Exercici 2
290   Mostra la mitjana d'amount per IBAN de les targetes de crèdit a la companyia Donec Ltd, utilitza almenys 2 taules.
291 */
292 -- Sense rebutjar les declined
293 • SELECT c.iban, c.pan, AVG(t.amount) AS mitjana
294   FROM dim_credit_cards AS c
295   JOIN fact_transactions AS t
296   ON c.id = t.card_id
297   WHERE t.business_id IN (
298     SELECT company_id
299       FROM dim_companies
300      WHERE company_name = 'Donec Ltd')
301   GROUP BY c.iban;
302

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Action Output | # Time Action | Message | 1 row(s) returned

iban	pan	mitjana
PT87806228135092429456346	5445865454343384	203.715000

Result 21 x

Output

Action Output | # Time Action | Message | 1 row(s) returned

1 02:06:20 SELECT c.iban, c.pan, AVG(t.amount) AS mitjana FROM dim_credit_cards AS c JOIN fact_transactions AS t ON c.id = t.card_id WHERE t.business_id IN (SELECT company_id FROM dim_companies WHERE company_name = 'Donec Ltd') AND t.declined = 0 GROUP BY c.iban;

2. Rebutjar les declined.

```

303 -- Rebutjan les declined
304 • SELECT c.iban, c.pan, AVG(t.amount) AS mitjana
305   FROM dim_credit_cards AS c
306   JOIN fact_transactions AS t
307   ON c.id = t.card_id
308   WHERE t.business_id IN (
309     (SELECT company_id
310      FROM dim_companies
311     WHERE company_name = 'Donec Ltd')
312     AND t.declined = 0
313   GROUP BY c.iban;
314

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Action Output | # Time Action | Message | 1 row(s) returned

iban	pan	mitjana
PT87806228135092429456346	5445865454343384	42.820000

Result 22 x

Output

Action Output | # Time Action | Message | 1 row(s) returned

1 02:10:02 SELECT c.iban, c.pan, AVG(t.amount) AS mitjana FROM dim_credit_cards AS c JOIN fact_transactions AS t ON c.id = t.card_id WHERE t.business_id IN (SELECT company_id FROM dim_companies WHERE company_name = 'Donec Ltd') AND t.declined = 0 GROUP BY c.iban;

Al comparar es pot veure com pot distorsionar el resultat si es tenen en compte les declined. A priori, podríem dir que potser la targeta té algun límit d'import per transacció. Informació que li podria ser útil al departament de pricing, entre d'altres.

Nivell 2 - Exercici 1

Crea una nova taula que reflecteixi l'estat de les targetes de crèdit basat en si les últimes tres transaccions van ser declinades i genera la següent consulta:

Quantes targetes estan actives?

1. Creem la taula Dim_estat_targetes.

```

315  /* SPRINT 4
316  Nivell 2 - Exercici 1
317  Crea una nova taula que reflecteixi l'estat de les targetes de crèdit basat en si les últimes tres transaccions van ser declinades i genera la següent consulta:
318
319 */
320  -- Primer cream la taula
321 • CREATE TABLE IF NOT EXISTS Dim_estat_targetes (
322     card_id VARCHAR(20) PRIMARY KEY,
323     estat VARCHAR(10)); -- aquí serà activa o inactiva, farem un CASE
324

```

Output:

#	Time	Action	Message	Duration
1	02:11:30	CREATE TABLE IF NOT EXISTS Dim_estat_targetes (card_id VARCHAR(20) PRIMARY KEY, estat VARCHAR(10))	0 row(s) affected	0.047s

2. Afegir dades a la nova taula.

```

325  -- Afegim les dades a la nova taula
326  -- Farem servir el case i una subquery
327 • INSERT INTO Dim_estat_targetes (card_id, estat)
328     SELECT card_id,
329         CASE
330             WHEN COUNT(*) = 3 AND SUM(declined) = 3 THEN 'inactiva'
331             ELSE 'activa'
332         END AS estat
333     FROM (
334         SELECT
335             card_id,
336             declined,
337             -- això dona un número de línia per cada transacció en ordre descent que ens fa falta per buscar les 3
338             ROW_NUMBER() OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS vegades
339         FROM Fact_transactions) AS taula_derivada
340     WHERE vegades <= 3
341     GROUP BY card_id;
342
343  -- Quantes targetes estan actives?

```

Output:

#	Time	Action	Message	
1	02:15:11	INSERT INTO Dim_estat_targetes (card_id, estat) SELECT card_id, CASE WHEN COUNT(*) = 3 AND SUM(declined) = 3 THEN inactiva' ELS... 275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0		

3. Connectar amb Dim_credit_cards: afegir FK

```

324
325  -- S'ha de connectar amb Dim_credit_cards: afegir FK
326 • ALTER TABLE Dim_credit_cards
327     ADD CONSTRAINT FK_dim_estat_targetes
328     FOREIGN KEY (id) REFERENCES dim_estat_targetes (card_id);
329

```

Output:

#	Time	Action	Message	
1	11:52:27	ALTER TABLE Dim_credit_cards ADD CONSTRAINT FK_dim_estat_targetes FOREIGN KEY (id) REFERENCES dim_estat_targetes (card_id)	275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0	

4. Executem la consulta.

```

343  -- Quantes targetes estan actives?
344 • SELECT COUNT(*) AS numero_targetes_actives
345     FROM Dim_estat_targetes
346     WHERE estat = 'activa';
347

```

Result Grid:

numero_targetes_actives
275

Result 23:

Output:

#	Time	Action	Message	
1	02:16:16	SELECT COUNT(*) AS numero_targetes_actives FROM Dim_estat_targetes WHERE estat = 'activa' LIMIT 0, 1000	1 row(s) returned	

Nivell 3 - Exercici 1

Crea una taula amb la qual puguem unir les dades del nou arxiu products.csv amb la base de dades creada, tenint en compte que des de transaction tens product_ids. Generar la següent consulta:

Necessitem conèixer el nombre de vegades que s'ha venut cada producte.

1. Crear taula intermitja per poder treballar amb els products_ids i els transaction_id.
Crear la taula temporal.

```

347  /* SPRINT 4
348  Nivell 3 - Exercici 1
349  Crea una taula amb la qual puguem unir les dades del nou arxiu products.csv amb la base de dades creada, tenint en compte que des de transaction tens product_ids.
350  Generar la següent consulta:
351
352  Necesitem conèixer el nombre de vegades que s'ha venut cada producte.
353  */
354  -- Creem una taula intermitja per poder treballar amb els product_ids i els transaction_id
355  CREATE TABLE IF NOT EXISTS fact_transactions_products (
356      transaction_id VARCHAR(50),
357      product_id INT,
358      FOREIGN KEY (transaction_id) REFERENCES fact_transactions(id),
359      FOREIGN KEY (product_id) REFERENCES dim_products(id),
360      PRIMARY KEY (transaction_id, product_id)
361 );
362
363  -- Creació de taula temporal per poder afegir la columna amb els product_ids i separar-los
364  CREATE TEMPORARY TABLE temp_transactions_products (
365      transaction_id VARCHAR(50),
366      product_ids VARCHAR(50)
367 );
368

Output
Action Output
# Time Action Message
# 1 02:49:16 CREATE TABLE IF NOT EXISTS fact_transactions_products ( transaction_id VARCHAR(50), product_id INT, FOREIGN KEY transaction_id... 0 row(s) affected
# 2 02:49:23 CREATE TEMPORARY TABLE temp_transactions_products ( transaction_id VARCHAR(50), product_ids VARCHAR(50)) 0 row(s) affected
0.01
0.01

```

2. Load de les columnes que ens interessen des de transactions.csv.

```

369  -- Carreguem la taula temporal només amb columnes que interessen de transactions.csv
370
371  LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv'
372  INTO TABLE temp_transactions_products
373  FIELDS TERMINATED BY ';'
374  ENCLOSED BY '"'
375  LINES TERMINATED BY '\r\n'
376  IGNORE 1 ROWS
377  (transaction_id, @ignore, @ignore, @ignore, @ignore, @ignore, @product_ids_temp, @ignore, @ignore, @ignore)
378  SET product_ids = REPLACE(@product_ids_temp, ',', ',');
379
380  -- Creem un altre procediment per dividir els ids i alimentar la taula

Output
Action Output
# Time Action Message
# 1 02:52:19 LOAD DATA INFILE C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv INTO TABLE temp_transactions_products FIELDS TE... 587 row(s) affected Records: 587 Deleted: 0 Skipped: 0 Warnings: 0
0.01
0.01

```

3. Creació procediment per separar els ids i alimentar la taula.

```

380  -- Creem un altre procediment per dividir els ids i alimentar la taula
381  DELIMITER //
382
383 • CREATE PROCEDURE split_and_insert_product_ids()
384  BEGIN
385      DECLARE finished INT DEFAULT 0;
386      DECLARE current_transaction_id VARCHAR(50);
387      DECLARE current_product_ids VARCHAR(50);
388      DECLARE single_product_id VARCHAR(50);
389      DECLARE product_pos INT;
390
391      DECLARE temp_cursor CURSOR FOR SELECT transaction_id, product_ids FROM temp_transactions_products;
392      DECLARE CONTINUE HANDLER FOR NOT FOUND SET finished = 1;
393
394      OPEN temp_cursor;
395
396      read_loop: LOOP
397          FETCH temp_cursor INTO current_transaction_id, current_product_ids;
398          IF finished THEN
399              LEAVE read_loop;
400          END IF;
401
402          WHILE LENGTH(current_product_ids) > 0 DO
403              SET product_pos = LOCATE(',', current_product_ids);
404
405              IF product_pos = 0 THEN
406                  SET single_product_id = current_product_ids;
407                  SET current_product_ids = '';
408              ELSE
409                  SET single_product_id = LEFT(current_product_ids, product_pos - 1);
410                  SET current_product_ids = SUBSTRING(current_product_ids, product_pos + 1);
411              END IF;
412
413              INSERT INTO fact_transactions_products (transaction_id, product_id)
414              VALUES (current_transaction_id, single_product_id);
415          END WHILE;
416      END LOOP;
417
418      CLOSE temp_cursor;
419  END //
420
421  DELIMITER ;
422
423 • -- Executar el procediment per dividir-ho i carregar-ho
424  CALL split_and_insert_product_ids();

Output:
Action Output
# Time Action
1 02:52:19 LOAD DATA INFILE C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv INTO TABLE temp_transactions_products FIELDS TE... 587 row(s) affected Records: 587 Deleted: 0 Skipped: 0 Warnings: 0
2 02:54:16 CREATE PROCEDURE split_and_insert_product_ids() BEGIN  DECLARE finished INT DEFAULT 0;  DECLARE current_transaction_id VARCHA... 0 row(s) affected
Message

```

4. Executar procediment.

```

423 • -- Executar el procediment per dividir-ho i carregar-ho
424  CALL split_and_insert_product_ids();

Output:
Action Output
# Time Action
1 02:57:03 CALL split_and_insert_product_ids()
Message
0 row(s) affected

```

5. Consulta per conèixer el nombre de vegades que s'ha venut cada producte.

```
426    -- Consulta per conèixer el nombre de vegades que s'ha venut cada producte
427 • SELECT p.id AS product_id,
428        p.product_name,
429        COUNT(tp.transaction_id) AS nombre_vegades_venut
430   FROM Dim_products AS p
431  LEFT JOIN Fact_transactions_products AS tp
432    ON p.id = tp.product_id
433   GROUP BY p.id, p.product_name
434   ORDER BY nombre_vegades_venut DESC;
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: []

product_id	product_name	nombre_vegades_venut
23	riverlands north	68
67	Winterfell	68
79	Direwolf riverlands the	66
2	Tarly Stark	65
43	duel	65
47	Tully	62
1	Direwolf Stannis	61

Result 24 ×

Output

Action Output

#	Time	Action	Message
1	02:57:03	CALL split_and_insert_product_ids()	0 row(s) affected
2	02:58:00	SELECT p.id AS product_id, p.product_name, COUNT(tp.transaction_id) AS nombre_vegades_venut FROM Dim_products AS p LEFT JOI...	100 row(s) returned