

Bases de Datos con SQL: TAREA FINAL

Máster Data Science, Big Data & Business Analytics 2023-2024

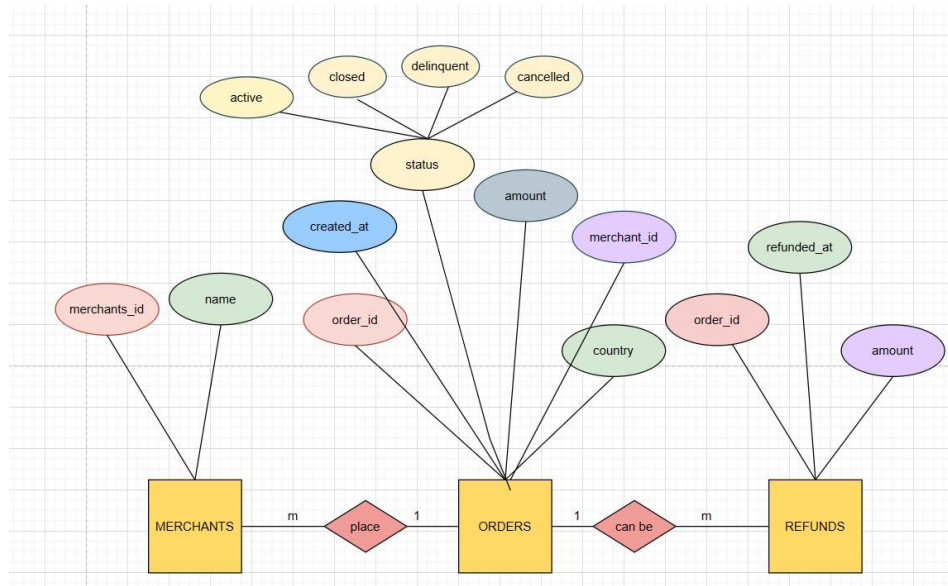
TAREA SQL

Máster Data Science, Big Data & Business Analytics 2023-2024	3
Ejercicio 1	5
Modelo Entidad Relación	5
Modelo Lógico	5
Modelo Físico	6
Ejercicio 2	6
2.1 Total de operaciones y su importe promedio, en Francia, Portugal y España	7
2.2 Los tres países con mayor número de operaciones	8
Ejercicio 3	8
3.1 Total de operaciones por País y comercio	9
3.2 Devoluciones por operación y suma valor de las operaciones.	9
Ejercicio 4	11
4.1 Estudio Riesgo crediticio	11
4.1.1 Número de pedidos por establecimiento y país.	11
4.1.2 Ratio Impagos.....	12
4.1.3 Ratio Devoluciones.....	13

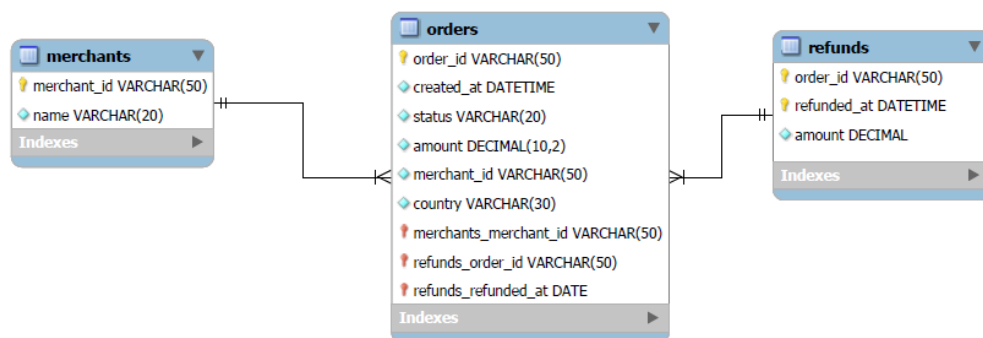
Ejercicio 1

Desarrollar un **Modelo de Entidad-Relación** indicando las entidades, atributos, relaciones y cardinalidad entre las tablas, el **Modelo Lógico** y el **Modelo Físico** mediante DLL.

Modelo Entidad Relación



Modelo Lógico



Modelo Físico

```
CREATE SCHEMA IF NOT EXISTS `prestamos_2015`;  
USE `prestamos_2015`;  
  
CREATE TABLE IF NOT EXISTS `prestamos_2015`.`merchants` (  
  `merchant_id` VARCHAR(50) NOT NULL,  
  `name` VARCHAR(20) NOT NULL,  
  PRIMARY KEY (`merchant_id`));  
  
CREATE TABLE IF NOT EXISTS `prestamos_2015`.`refunds` (  
  `order_id` VARCHAR(50) NOT NULL,  
  `refunded_at` DATETIME NOT NULL,  
  `amount` DECIMAL NOT NULL,  
  PRIMARY KEY (`order_id`, `refunded_at`));  
  
CREATE TABLE IF NOT EXISTS `prestamos_2015`.`orders` (  
  `order_id` VARCHAR(50) NOT NULL,  
  `created_at` DATETIME NOT NULL,  
  `status` VARCHAR(20) NOT NULL,  
  `amount` DECIMAL(10,2) NOT NULL,  
  `merchant_id` VARCHAR(50) NOT NULL,  
  `country` VARCHAR(30) NOT NULL,  
  PRIMARY KEY (`order_id`));
```

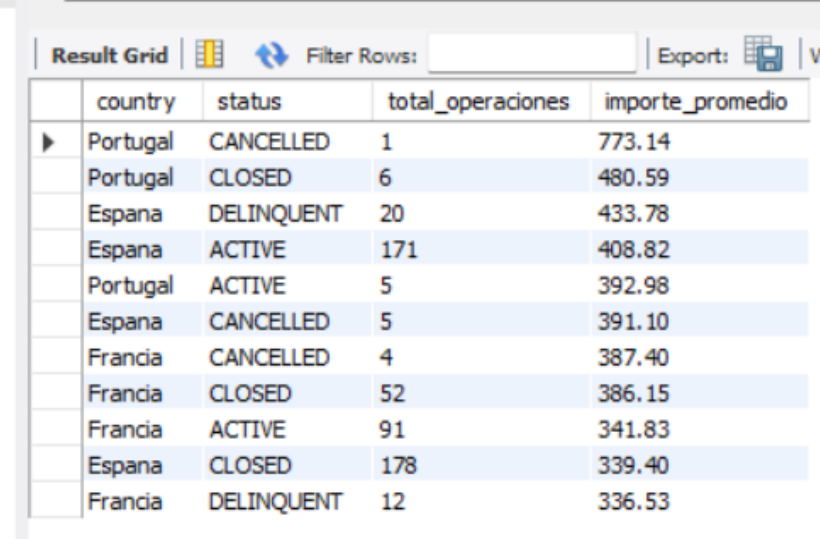
Ejercicio 2

A partir de las tablas incluidas en la base de datos préstamos, vamos a realizar las siguientes consultas:

- 1- Realizamos una consulta donde obtengamos por país y estado de operación, el total de operaciones y su importe promedio. La consulta debe cumplir las siguientes condiciones:
 - a. Operaciones posteriores al 01-07-2015
 - b. Operaciones realizadas en Francia, Portugal y España.
 - c. Operaciones con un valor mayor de 100 € y menor de 1500€
 - d. Ordenamos los resultados por el promedio del importe de manera descendente.

2.1 Total de operaciones y su importe promedio, en Francia, Portugal y España

```
SELECT country, status, COUNT(*) AS total_operaciones, ROUND(AVG(amount), 2) AS
importe_promedio
FROM orders
WHERE created_at > '2015-07-01' AND country IN ('Francia', 'Portugal', 'España') AND
amount > 100 AND amount < 1500
GROUP BY country,
status
ORDER BY
importe_promedio DESC;
```



	country	status	total_operaciones	importe_promedio
▶	Portugal	CANCELLED	1	773.14
	Portugal	CLOSED	6	480.59
	Espana	DELINQUENT	20	433.78
	Espana	ACTIVE	171	408.82
	Portugal	ACTIVE	5	392.98
	Espana	CANCELLED	5	391.10
	Francia	CANCELLED	4	387.40
	Francia	CLOSED	52	386.15
	Francia	ACTIVE	91	341.83
	Espana	CLOSED	178	339.40
	Francia	DELINQUENT	12	336.53

- 2- Realizamos una consulta donde obtengamos los 3 países con el mayor número de operaciones, el total de operaciones, la operación con un valor máximo y la operación con el valor mínimo para cada país. La consulta debe cumplir las siguientes condiciones:
- Excluimos aquellas operaciones con el estado “Delinquent” y “Cancelled”
 - Operaciones con un valor mayor de 100 €

2.2 Los tres países con mayor número de operaciones

```
SELECT country,
COUNT(*) AS total_operaciones,
SUM(amount) AS total_valor_operaciones,
MAX(amount) AS operacion_valor_maximo,
MIN(amount) AS operacion_valor_minimo
FROM orders
WHERE status NOT IN ('Delinquent', 'Cancelled') AND amount > 100
GROUP BY country
ORDER BY total_operaciones DESC
LIMIT 3;
```

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	country	total_operaciones	total_valor_operaciones	operacion_valor_maximo	operacion_valor_minimo
▶	Espana	359	150399.07	2960.87	101.00
	Francia	147	57896.54	1863.98	100.88
	Italia	77	23695.32	1299.00	107.99

Ejercicio 3

A partir de las tablas incluidas en la base de datos prestamos 2015 vamos a realizar las siguientes consultas:

- 1- . Realizamos una consulta donde obtengamos, por país y comercio, el total de operaciones, su valor promedio y el total de devoluciones. La consulta debe cumplir las siguientes condiciones:
 - a. Se debe mostrar el nombre y el id del comercio.
 - b. Comercios con más de 10 ventas.
 - c. Comercios de Marruecos, Italia, España y Portugal.
 - d. Creamos un campo que identifique si el comercio acepta o no devoluciones. Si no acepta (total de devoluciones es igual a cero) el campo debe contener el valor “No” y si sí lo acepta (total de devoluciones es mayor que cero) el campo
 - e. debe contener el valor “Sí”. Llamaremos al campo “acepta_devoluciones”.

Ordenamos los resultados por el total de operaciones de manera ascendente.

3.1 Total de operaciones por País y comercio

```

SELECT orders.country, merchants.merchant_id, merchants.name,
COUNT(orders.order_id) AS total_operaciones, ROUND(AVG(orders.amount), 2) AS
valor_promedio,
CASE
WHEN SUM(CASE WHEN refunds.amount IS NOT NULL THEN 1 ELSE 0 END) >
0 THEN 'Si'
ELSE 'No'
END AS acepta_devoluciones,
COUNT(DISTINCT refunds.order_id) AS total_devoluciones
FROM orders
JOIN merchants ON orders.merchant_id = merchants.merchant_id
LEFT JOIN refunds ON orders.order_id = refunds.order_id
WHERE orders.amount > 0 AND orders.country IN ('Marruecos', 'Italia', 'España',
'Portugal')
GROUP BY orders.country, merchants.merchant_id, merchants.name
HAVING COUNT(orders.order_id) > 10
ORDER BY total_operaciones ASC;

```

country	merchant_id	name	total_operaciones	valor_promedio	acepta_devoluciones	total_devoluciones
Espana	pk_743f2fdec876b75e975c005	Pepe Jeans	11	171.99	No	0
Espana	pk_736c7094ea96eda38b098f56	Massimo Dutti	13	169.88	No	0
Marruecos	pk_317b4fc6fd80a5f8fb2ff216	Calcedonia	13	365.36	Si	2
Espana	pk_c15afcbd3a31b732f097ba7b	Havainas	16	323.02	No	0
Espana	pk_07225590b8fea17e739aa451	Netflix	21	363.57	No	0
Espana	pk_a3aa2fa07c5436f4c8ca1e03	fnac	22	531.84	No	0
Espana	pk_c447a91e755425d163df6837	YouTube music	25	669.28	Si	1
Italia	pk_317b4fc6fd80a5f8fb2ff216	Calcedonia	26	229.30	Si	2
Espana	pk_b9ee4936f19ba28d96f6001e	K-tuin	46	373.36	No	0
Espana	pk_19d9ed34a670cbd04543ec35	Spotify	64	644.26	No	0
Espana	pk_317b4fc6fd80a5f8fb2ff216	Calcedonia	137	314.76	Si	4

- Realizamos una consulta donde vamos a traer todos los campos de las tablas operaciones y comercios. De la tabla devoluciones vamos a traer el conteo de devoluciones por operación y la suma del valor de las devoluciones. Una vez tengamos la consulta anterior, creamos una vista con el nombre orders_view dentro del esquema Prestamos_2015 con esta consulta.

3.2 Devoluciones por operación y suma valor de las operaciones.

```

SELECT *
FROM merchants
JOIN orders ON merchants.merchant_id = orders.merchant_id;
SELECT order_id, COUNT(*) AS conteo_devoluciones, SUM(amount) AS suma_devoluciones
FROM refunds
GROUP BY order_id;

```

Result Grid			
Filter Rows: <input type="text"/>			
Export: Wrap Cell Con			
	order_id	conteo_devoluciones	suma_devoluciones
▶	5c370ef03e78fa6aac5f623	1	727
	5c3ef3f70aee697c1ba7e93a	1	123
	5c3ef3f70aee697c1ba7e93b	1	137
	5c3ef6460aee697c1ba82bcd	1	124
	5c3ef7880aee697c1ba83db8	1	769
	5c3ef8170aee697c1ba8432a	2	163
	5c3ef8170aee697c1ba8432b	1	773
	5c3ef8170aee697c1ba8432c	2	191
	5c3ef8170aee697c1ba8432d	1	236

```

CREATE VIEW orders_view AS
SELECT r.order_id,
COUNT(*) AS conteo_devoluciones,
SUM(r.amount) AS suma_devoluciones
FROM refunds r
GROUP BY r.order_id;

```

Filter objects

- Foreign Keys
- Triggers
- refunds
- Views
 - orders_view
- Stored Procedures
- Functions

ministration
Schemas
ormation

1 • SELECT * FROM prestamos_2015.orders_view;

View: orders_view
Columns:
order_id varchar
conteo_devoluciones bigint
suma_devoluciones decimal

Result Grid			
Filter Rows: <input type="text"/>			
Export: Wrap Cell			
	order_id	conteo_devoluciones	suma_devoluciones
▶	5c370ef03e78fa6aac5f623	1	727
	5c3ef3f70aee697c1ba7e93a	1	123
	5c3ef3f70aee697c1ba7e93b	1	137
	5c3ef6460aee697c1ba82bcd	1	124
	5c3ef7880aee697c1ba83db8	1	769
	5c3ef8170aee697c1ba8432a	2	163
	5c3ef8170aee697c1ba8432b	1	773
	5c3ef8170aee697c1ba8432c	2	191
	5c3ef8170aee697c1ba8432d	1	236

Nota: La tabla refunds contiene más de una devolución por operación por lo que, para hacer el cruce, es muy importante que agrupemos las devoluciones.

Ejercicio 4

A partir de los datos disponibles diseñar una **funcionalidad** a tu elección que permita obtener un insight de interés sobre el caso de uso estudiado.

Para ello debes plantear primeramente en un breve texto el objetivo de tu funcionalidad, la queries desarrollada y una reflexión sobre el insight obtenido. Para ello puedes usar cualquier recurso estudiado en clase.

Algunas funcionalidades podrían ser: segmentación de clientes en función del valor de las operaciones, sistema de alertas para operaciones delictivas, identificación de estacionalidad, etc.. Tienes libertad total para desarrollar tu funcionalidad, lo importante es que tenga tu sello personal.

4.1 Estudio Riesgo crediticio

Con el objeto de estudiar el riesgo crediticio de las operaciones con los diferente Comercios y países, se propone una funcionalidad que nos permita estudiar a los clientes de los comercios, por el número de pedidos en situación de impago, por un lado, y por el otro lado también vamos a estudiar el porcentaje de reembolsos en relación con el número total de pedidos.

Para estudiar el número de pedidos en situación de impago primero vamos a llevar a cabo un análisis exploratorio sobre las el número de pedidos realizados por establecimiento y país.

4.1.1 Número de pedidos por establecimiento y país.

```
SELECT m.name,  
       o.country,  
       COUNT(o.order_id) AS num_orders  
FROM merchants m  
JOIN orders o  
ON m.merchant_id = o.merchant_id  
GROUP BY m.name, o.country;
```

```
1 • SELECT m.name,  
2      o.country,  
3      COUNT(o.order_id) AS num_orders  
4 FROM merchants m  
5 JOIN orders o  
6 ON m.merchant_id = o.merchant_id  
7 GROUP BY m.name, o.country;
```

Result Grid			
Filter Rows:			
Export:			
	name	country	num_orders
	fnac	Italia	8
	Speedo	Espana	7
	Speedo	Francia	7
	Massimo ...	Espana	13
	Massimo ...	Francia	9
	Spotify	Espana	64
	Spotify	Francia	6
	Spotify	Italia	8
	YouTube ...	Italia	5
	YouTube ...	Espana	25
	YouTube ...	Belgica	2
	YouTube ...	Portugal	2

4.1.2 Ratio Impagos

En base a estos datos vamos a desarrollar una funcionalidad que nos indique el número total de pedidos que ha recibido y el porcentaje de pedidos que se encuentran en situación de impago, filtrando los diez primeros por Comercio y País.

```

SELECT name,
       country,
       num_orders,
       CONCAT(ROUND(delinquent_rate), '%') AS delinquent_rate
FROM (
  SELECT m.name,
         o.country,
         COUNT(o.order_id) AS num_orders,
         SUM(CASE WHEN o.status = 'Delinquent' THEN 1 ELSE 0 END) / COUNT(o.order_id) * 100
         AS delinquent_rate,
         ROW_NUMBER() OVER (ORDER BY SUM(CASE WHEN o.status = 'Delinquent' THEN 1
         ELSE 0 END) / COUNT(o.order_id) DESC) AS row_num
  FROM merchants m
  JOIN orders o
  ON m.merchant_id = o.merchant_id
  GROUP BY m.name, o.country
) AS t
WHERE row_num <= 10;

```

```

8 • SELECT name,
9       country,
10      num_orders,
11      CONCAT(ROUND(delinquent_rate), '%') AS delinquent_rate
12 FROM (
13     SELECT m.name,
14           o.country,
15           COUNT(o.order_id) AS num_orders,
16           SUM(CASE WHEN o.status = 'Delinquent' THEN 1 ELSE 0 END) / COUNT(o.order_id) * 100 AS delinquent_rate,
17           ROW_NUMBER() OVER (ORDER BY SUM(CASE WHEN o.status = 'Delinquent' THEN 1 ELSE 0 END) / COUNT(o.order_id) DESC) AS row_num
18     FROM merchants m
19     JOIN orders o
20     ON m.merchant_id = o.merchant_id
21     GROUP BY m.name, o.country
22 ) AS t
23 WHERE row_num <= 10;

```

	name	country	num_orders	delinquent_rate
▶	Lefties	Espana	1	100%
	Loreal	Francia	2	50%
	Netflix	Alemania	2	50%
	Speedo	Francia	7	43%
	Carhart	Francia	3	33%
	K-tuin	Marruecos	10	30%
	Havainas	Francia	8	25%
	Havainas	Marruecos	4	25%
	Leroy Merlin	Espana	4	25%
	Netflix	Belgica	14	21%

4.1.3 Ratio Devoluciones

Ahora estudiaremos en función comerciante y el país, el número total de pedidos que ha recibido y el porcentaje de reembolsos en relación con el número total de pedidos, mostrando los diez comercios con más devoluciones.

```

SELECT name,
       country,
       num_orders,
       CONCAT(ROUND(refund_rate * 100, 2), '%') AS refund_rate
FROM (
  SELECT m.name,
         o.country,
         COUNT(o.order_id) AS num_orders,
         SUM(r.amount) / SUM(o.amount) AS refund_rate,
         ROW_NUMBER() OVER (ORDER BY SUM(r.amount) / SUM(o.amount) DESC) AS row_num
  FROM merchants m
       JOIN orders o
         ON m.merchant_id = o.merchant_id
       LEFT JOIN refunds r
         ON o.order_id = r.order_id
  WHERE o.status IS NOT NULL AND r.amount IS NOT NULL
  GROUP BY m.name, o.country
  HAVING refund_rate IS NOT NULL
) AS t
WHERE row_num <= 10;

```

```

8 • SELECT name,
9     country,
10    num_orders,
11    CONCAT(ROUND(refund_rate * 100), '%') AS refund_rate
12 FROM (
13     SELECT m.name,
14            o.country,
15            COUNT(o.order_id) AS num_orders,
16            SUM(r.amount) / SUM(o.amount) AS refund_rate,
17            ROW_NUMBER() OVER (ORDER BY SUM(r.amount) / SUM(o.amount) DESC) AS row_num
18     FROM merchants m
19          JOIN orders o
20            ON m.merchant_id = o.merchant_id
21          LEFT JOIN refunds r
22            ON o.order_id = r.order_id
23     WHERE o.status IS NOT NULL AND r.amount IS NOT NULL
24     GROUP BY m.name, o.country
25     HAVING refund_rate IS NOT NULL
26 ) AS t
27 WHERE row_num <= 10;

```

Result Grid		Filter Rows:		Export:	Wrap
	name	country	num_orders	refund_rate	
▶	YouTube music	Espana	1	100%	
	Apple music	Francia	1	100%	
	Calcedonia	Portugal	1	100%	
	Calcedonia	Italia	2	100%	
	Havainas	Marruecos	1	100%	
	Kindle	Italia	1	100%	
	Calcedonia	Espana	5	85%	
	Calcedonia	Alemania	5	79%	
	Calcedonia	Francia	3	69%	
	Calcedonia	Marruecos	3	64%	

A la vista de los resultados con el “ratio de impagos” podemos observar las diez empresas con los porcentajes más altos de clientes morosos en relación con los pedidos realizados , resaltar que Netflix tiene impagos tanto en Bélgica como en Alemania y Havainas también tiene impagos en Marruecos y Francia, por otro lado en el resultado de “ratio de devoluciones” observamos el ranking de los diez comercios, que más devoluciones de préstamos vinculados a un compra determinada realizan, en este caso tendríamos que resaltar los porcentajes de Calcedonia que se repiten en diferentes países.

La conclusión final es que habría que revisar las condiciones financieras de crédito-vinculado tanto de Netflix, Havainas como de Calcedonia.

