# **Sprint 2 Data Analytics**

## Nivell 1

## - Ejercicio 1

A partir de los documentos adjuntos (estructura\_datos y datos\_introducir), importa las dos tablas. Muestra las características principales del esquema creado y explica las diferentes tablas y variables que existen. Asegúrate de incluir un diagrama que ilustre la relación entre las diferentes tablas y variables.

Tras importar los datos en mysql he procedido a realizar el diagrama que, como se puede visualizar consta de 2 tablas, la tabla "Company" y la tabla "Transaction", teniendo una relación de 1 a N. En este caso la empresa puede tener un identificador único pero múltiples transacciones.

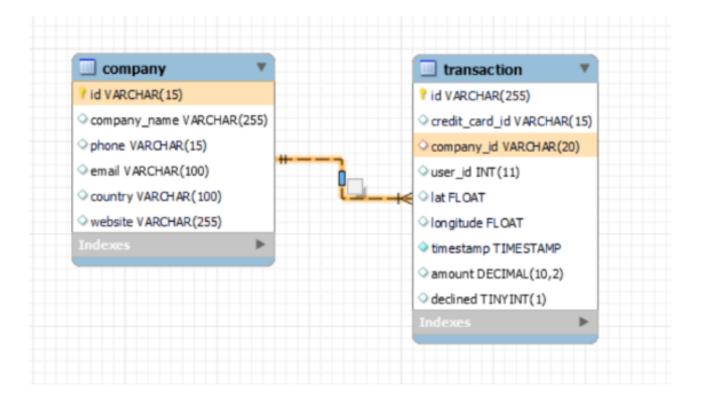


Tabla 1. Company: tabla de Empresa Variables

id: Identificador y Primary key (PK). Se trata del identificador único para cada empresa.
Cuando queremos asegurarnos al 100% que una columna es la PK podemos hacer la siguiente query:





En primer lugar, seleccionamos el id de la tabla "company" y contamos los registros agrupados por id, filtrando aquellos con más de una aparición usando HAVING COUNT(\*) > 1 para incluir aquellos resultados que tienen más de una fila. Como no nos devuelve ninguna fila, podemos deducir que no hay duplicados y que por tanto es nuestra PK.

company\_name: Nombre de la empresa.

phone: Teléfono de la empresa.

email: Correo electrónico de la empresa.

country: País en el cual está ubicada la empresa.

website: Web de la empresa.

#### Tabla 2. Transaction: Tabla de transacción

#### **Variables**

id: Identificador único para cada transacción. Es la PK de esta tabla.

credit\_card\_id: Identificador de las tarjetas de crédito.

**company\_id**: Identificador de la compañía. Indica a qué empresa pertenece la transacción y es la **FK**.

La FK o Foreign Key es un campo o campos que nos ayudan a relacionarlos con una PK de otra tabla. En este caso la PK de la tabla 1 Company.id está relacionada con Company\_id.

**user\_id:** Identificador para el usuario que realizó la transacción.

lat: Latitud de la ubicación donde se realizó la transacción.

**longitude:** Longitud de la ubicación donde se realizó la transacción.

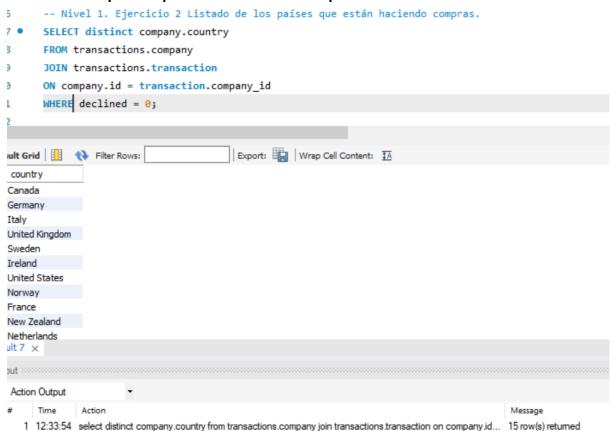
timestamp: Marca de tiempo cuando se realizó la transacción.

amount: Monto de la transacción.

declined: Indicador de si la transacción fue rechazada.

## - Ejercicio 2

### Utilizando JOIN realizarás las siguientes consultas: Listado de los países que están haciendo compras.



En esta consulta seleccionamos todos los países de la tabla company, eliminando aquellos duplicados con distinct. Después, con una JOIN unimos las tablas "company" y "transaction" a través de la PK (id) de la tabla "Company" y la Fk (company\_id) de la tabla "Transaction". Por último, para asegurarnos de que solo se incluyen las transacciones que no fueron rechazadas usamos el (declined = 0). Nos da un resultado de 15 países.

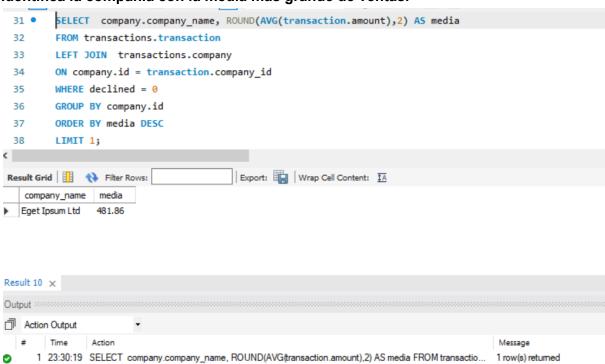
### Desde cuántos países se realizan las compras.





En esta consulta seleccionamos el contaje de los países, siempre con distinct para que al igual que en la consulta anterior no se nos repitan dichos países y en esta ocasión he puesto el alias "num\_países". Después, con una JOIN unimos las tablas "company" y "transaction" a través de la PK (id) de la tabla "Company" y la Fk (company\_id) de la tabla "Transaction". Y volvemos a incluir (declined = 0) para incluir sólo las transacciones que no fueron rechazadas. El resultado nos da 15.

#### Identifica la compañía con la media más grande de ventas.



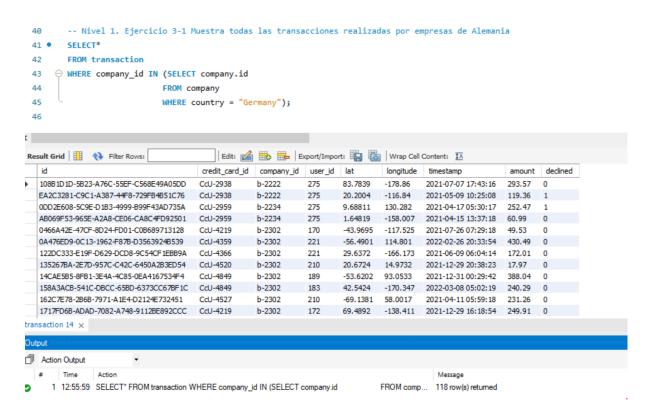
En primer lugar, seleccionamos el nombre de la empresa (company\_name) y la media de las cantidades de transacción (ROUND(AVG(transaction.amount), 2)) redondeada a dos decimales. Se realiza una unión externa izquierda (LEFT JOIN) entre las tablas "Company" y "Transaction" a través de la PK (id) de la tabla "Company" y la Fk (company\_id). Solo se consideran las transacciones que no fueron rechazadas (declined = 0). Los resultados se agrupan por el id de la empresa (GROUP BY company.id), se ordenan por la media en orden descendente (ORDER BY media DESC), y se limita el resultado a una sola fila (LIMIT 1), es decir, la empresa con la media de transacciones más alta.

-ROUND redondea a las decimales que le pongas.

# - Ejercicio 3

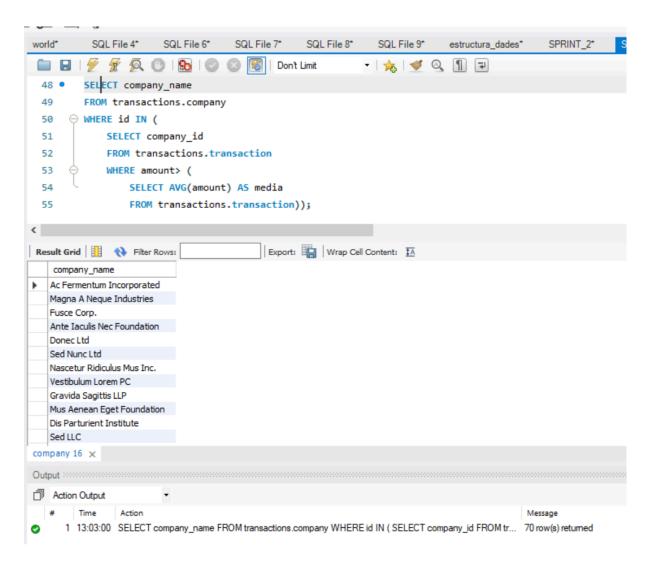
#### Utilizando solo subconsultas (sin utilizar JOIN):

Muestra todas las transacciones realizadas por empresas de Alemania.



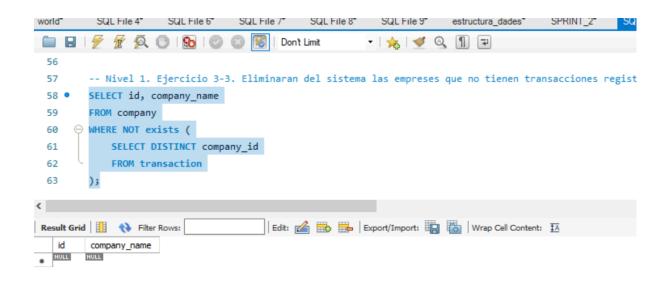
En primer lugar, seleccionamos todas las columnas de la tabla "Transaction" con un SELECT \*. Después, filtramos con un WHERE, seleccionando las transacciones cuyos company\_id están en la lista de id de empresas (company.id) que a su vez están en el país "Germany" (WHERE country = "Germany"). Es decir, devuelve todas las transacciones que pertenecen a empresas ubicadas en Alemania.

Lista las empresas que han realizado transacciones por un amount superior a la media de todas las transacciones.



En esta query, seleccionamos los nombres de las empresas (company\_name) de la tabla "Company". Después empezamos las subconsultas, filtrando las empresas (WHERE id IN) cuyo "id" está en la lista de "company\_id" de transacciones (transactions.transaction) donde el monto (amount) es mayor que el promedio (AVG(amount)) de todos los montos de transacción. Es decir, devuelve los nombres de las empresas que tienen transacciones con montos superiores al promedio global de todas las transacciones.

Se eliminarán del sistema las empresas que no tienen transacciones registradas, entrega la lista de estas empresas.



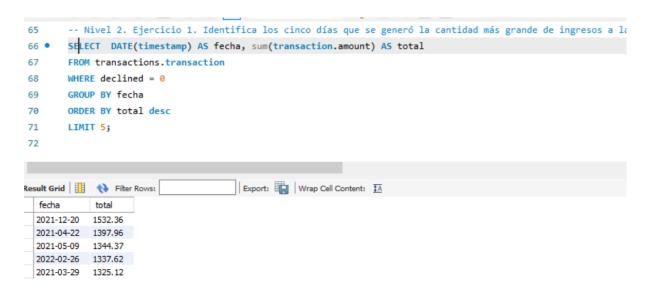


En esta query, seleccionamos el "id" y el "company\_name" de la tabla "company". Después filtramos las empresas donde no existe ninguna transacción asociada, para ello usamos la cláusula NOT EXISTS y así verificamos que no hay ningún "company\_id" en la tabla "Transaction" que coincida con el "id" de la empresa en cuestión. En resumen, devuelve las empresas que no tienen ninguna transacción registrada en la tabla "Transaction".

## Nivel 2

# Ejercicio 1

Identifica los cinco días que se generó la cantidad más grande de ingresos a la empresa por ventas. Muestra la fecha de cada transacción junto con el total de las ventas.



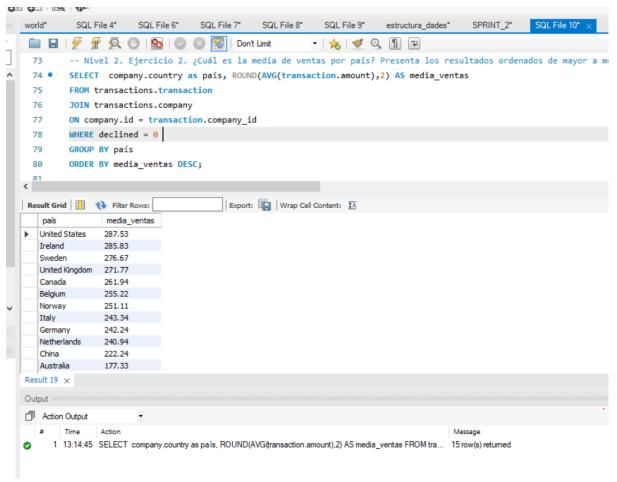


En primer lugar, seleccionamos la fecha de las transacciones (DATE(timestamp) AS fecha) y la suma de los montos de transacción (sum(transaction.amount) AS total) poniéndoles en el porceso alias. Después, filtramos las transacciones donde no fueron rechazadas (declined = 0). Luego, agrupamos las transacciones por fecha (GROUP BY fecha) y las ordenamos por el total de montos en orden descendente (ORDER BY total DESC). Finalmente, limitamos el resultado a las 5 fechas con el total de montos más altos (LIMIT 5). Es decir, con todo ello mostramos las 5 fechas con la mayor cantidad total de transacciones no rechazadas.

-Date hace que aparezca la fecha pero sin la hora.

## Ejercicio 2

¿Cuál es la media de ventas por país? Presenta los resultados ordenados de mayor a menor media.

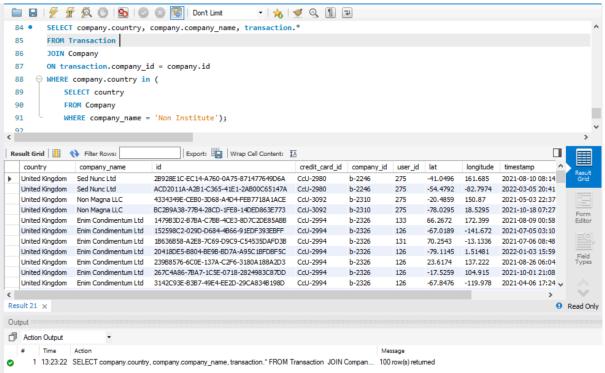


Seleccionamos el país de la empresa (company.country AS país) y la media de ventas (ROUND(AVG(transaction.amount),2) AS media\_ventas), a ambos les ponemos alias para usar más tarde. Después unimos (JOIN) la tabla "transaction" con la tabla "company" usando el id de la empresa y el company\_id de la transacción. Luego filtramos las transacciones donde no fueron rechazadas (declined = 0). Más tarde, agrupamos las transacciones por país (GROUP BY país) y se ordenan en orden descendente según la media de ventas (ORDER BY media\_ventas DESC). En resumen, muestra la media de ventas de las empresas por país, ordenadas de mayor a menor.

## - Ejercicio 3

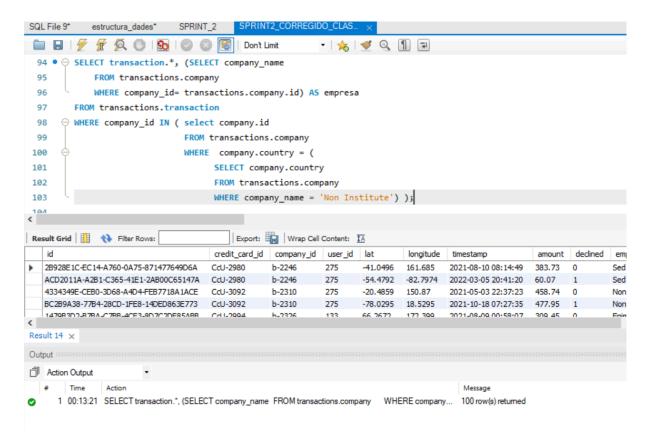
En tu empresa, se plantea un nuevo proyecto para lanzar algunas campañas publicitarias para hacer competencia a la compañía "Non Institute". Para lo cual, te piden la lista de todas las transacciones realizadas por empresas que están situadas en el mismo país que esta compañía.





En esta consulta seleccionamos el país de la empresa (company.country), el nombre de la empresa (company.company\_name), y todas las columnas de la tabla "Transaction" (transaction.\*). Se realiza una unión (JOIN) entre las tablas "Transaction" y "Company" basándose en el company\_id de la transacción y el id de la empresa. Filtramos los resultados donde el país de la empresa está incluido (IN) en la lista de países donde el nombre de la empresa es "Non Institute". En resumen, la consulta devuelve todas las transacciones de las empresas cuyo nombre coincide con "Non Institute", junto con el país y el nombre de la empresa asociada a cada transacción.

#### Muestra el listado aplicando solo subconsultas.

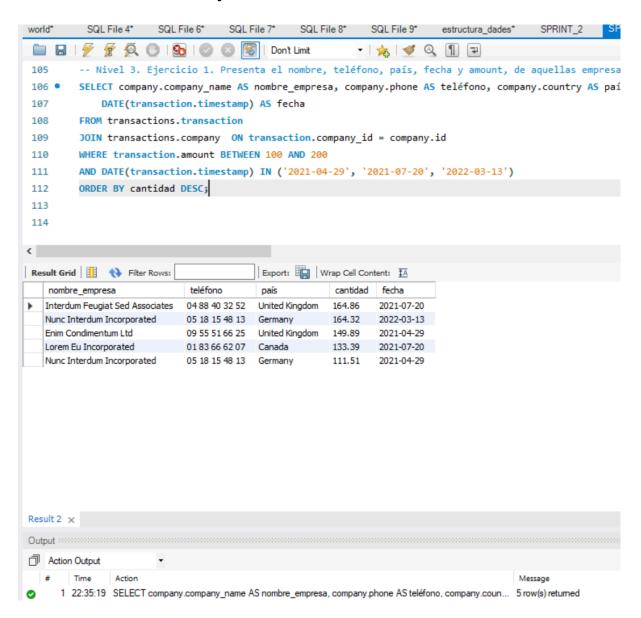


En esta query seleccionamos todas las columnas de la tabla "transaction" (transaction.\*) y agregamos una columna adicional bajo el alias "empresa", que contiene el nombre de la empresa asociada a cada transacción. Lo hace mediante una subconsulta que busca el company\_name en la tabla "company" donde el "company\_id" coincide con el "id" de la empresa. Luego, filtramos las transacciones para incluir solo aquellas cuyo "company\_id" pertenece a empresas que están en el mismo país que la empresa llamada 'Non Institute'. La subconsulta interna busca el country de la empresa 'Non Institute' y luego selecciona las transacciones de todas las empresas en ese país.

## Nivel 3

# Ejercicio 1

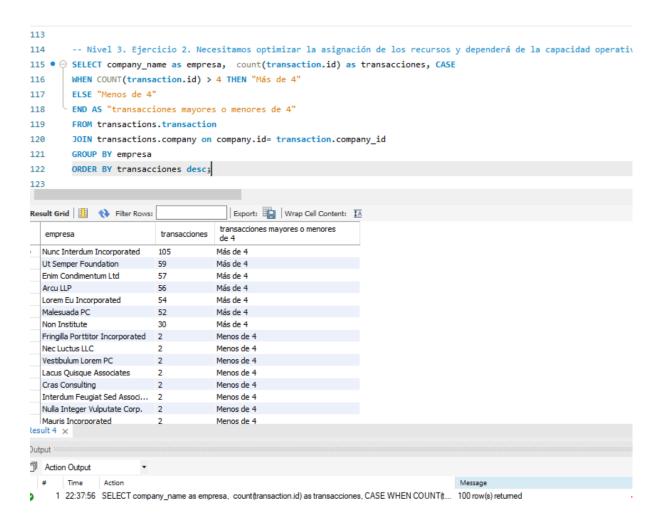
Presenta el nombre, teléfono, país, fecha y amount, de aquellas empresas que realizaron transacciones con un valor comprendido entre 100 y 200 euros y en alguna de estas fechas: 29 de abril del 2021, 20 de julio del 2021 y 13 de marzo del 2022. Ordena los resultados de mayor a menor cantidad.



En primer lugar seleccionamos el nombre de la empresa (company.company\_name), el teléfono (company.phone), el país (company.country), el monto de la transacción (transaction.amount), y la fecha de la transacción (DATE(transaction.timestamp)). Después unimos las tablas transaction y company usando transaction.company\_id y company.id. Filtramos las transacciones con un monto entre 100 y 200 (BETWEEN 100 AND 200) y que ocurrieron en las fechas específicas ('2021-04-29', '2021-07-20', '2022-03-13'). Los resultados se ordenan por el monto de la transacción en orden descendente (ORDER BY cantidad DESC.

# Ejercicio 2

Necesitamos optimizar la asignación de los recursos y dependerá de la capacidad operativa que se requiera, por lo cual te piden la información sobre la cantidad de transacciones que realicen las empresas, pero el departamento de recursos humanos es exigente y quiere un listado de las empresas donde especifiques si tienen más de 4 transacciones o menos.



En esta query seleccionamos el nombre de la empresa (company\_name AS empresa), el número de transacciones (COUNT(transaction.id) AS transacciones) a ambos les ponemos alias y usamos una etiqueta que indica si la empresa tiene más de 4 transacciones o menos (CASE ... END AS "transacciones mayores o menores de 4").

CASE funciona como un condicional, es decir, si se da que el contaje del "id" de la tabla "Transacción" es mayor a 4 entonces se printará "más de 4", sino se printará "menos de 4. Después, unimos como siempre las tablas "transaction" y "company" usando "company.id" y "transaction.company\_id". Luego, agrupamos los resultados por nombre de empresa (GROUP BY empresa) y los ordenamos por el número de transacciones en orden descendente (ORDER BY transacciones DESC).