# **DATA ANALYTICS**

IT ACADEMY | BCN ACTIVA

# SPRINT 02

Bases de datos relacionales e introducción a SQL







# **PRINCIPALES COMANDOS**

**SELECT**: extraer datos específicos de una o más tablas.

WHERE: filtrar los resultados según condiciones específicas (valores de rango, fechas o otros).

JOIN: combinar datos de dos tablas relacionadas utilizando claves comunes.

**GROUP BY**: agrupar los resultados en base a una columna específica para realizar cálculos como conteo o promedio.

**ORDER BY**: ordena los resultados de manera ascendente (ASC) o descendente (DESC) según una columna.

**CASE**: crear columnas personalizadas basadas en condiciones específicas.

**BETWEEN**: filtra valores que estén dentro de un rango determinado.

IN: seleccionar valores específicos.

**COUNT**: contar el número de registros en una columna específica.

AVG: promedio de valores en una columna.

WHERE NOT IN: filtra registros que no coinciden con una lista de valores específicos.

# **NIVEL 01**

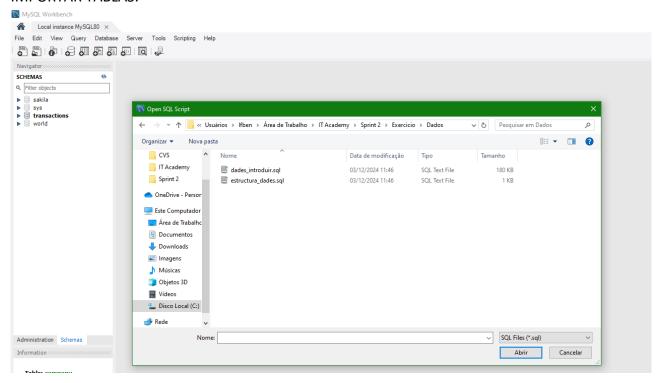
# **EJERCICIO 01**

A partir de los documentos adjuntos (*estructura\_dades* y *dades\_introduir*), 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.

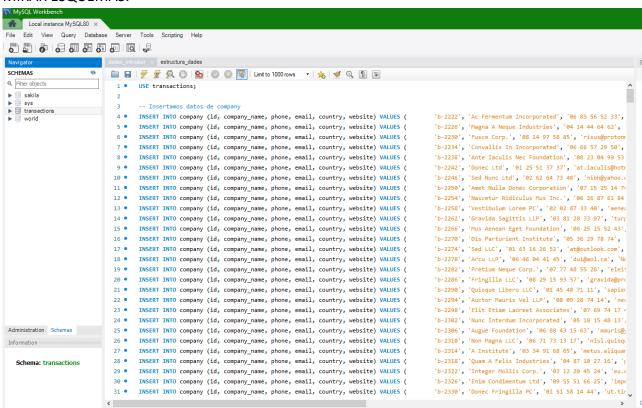
#### Datos a introducir:

Dades\_introduir.sql Estructura\_dades.sql

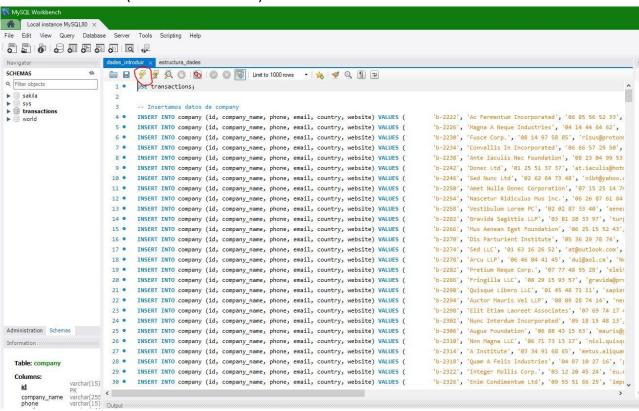
#### **IMPORTAR TABLAS:**



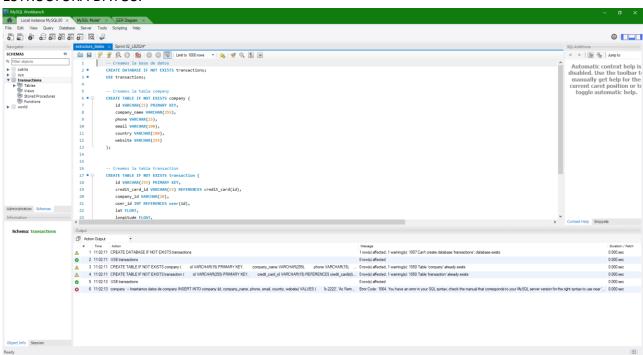
#### **MIRAR ESQUEMAS:**



#### EXECUTAR TABLAS (o ctrl + shift + enter):

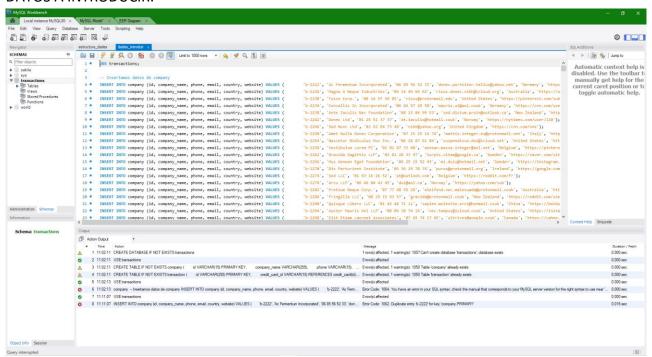


#### **ESTRUCTURA DATOS:**



El archivo describe las informaciones iniciales para la tabla de estructuración de los datos, como *id* (llave primaria), nombre de la empresa, teléfono, email, país e website. Desde esta forma estructurada, podremos linkar los datos con la información de los datos a seguir:

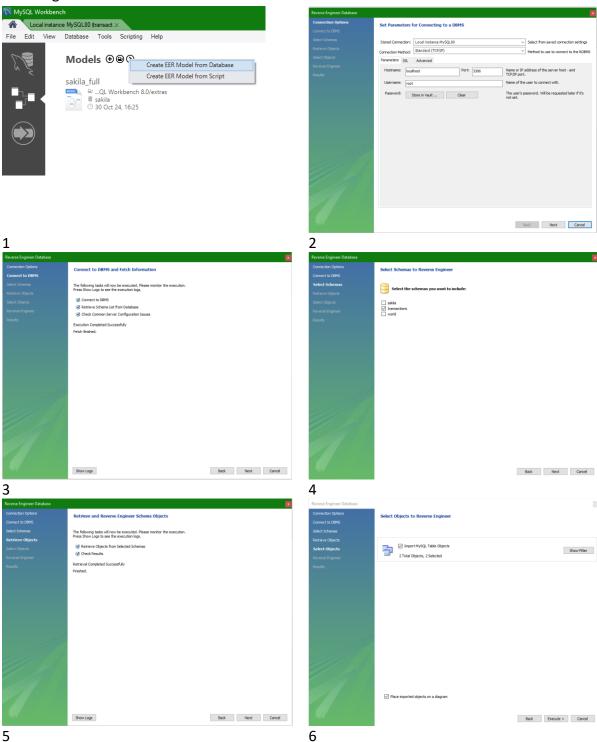
#### **DATOS A INTRODUCIR:**



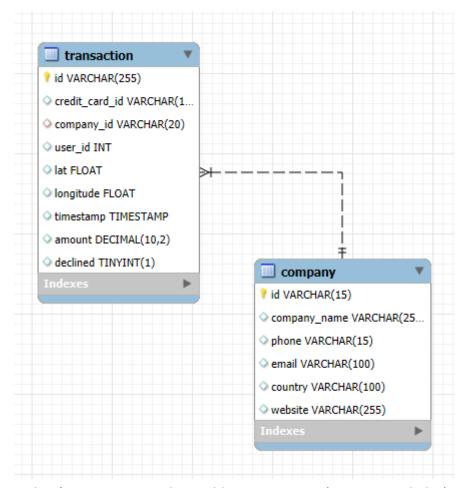
En este, están los datos de los clientes, estucturados de acuerdo con la estructura inicial. Com *id* (llave primaria) y los datos rellenados de los clientes.

### DIAGRAMA RELACIONAL (ER):

#### Como llegar hasta él:

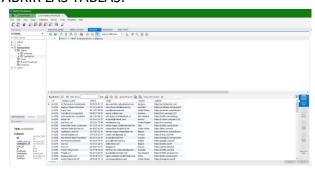


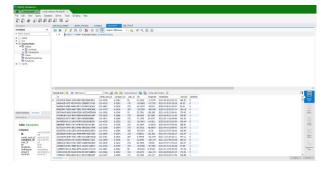




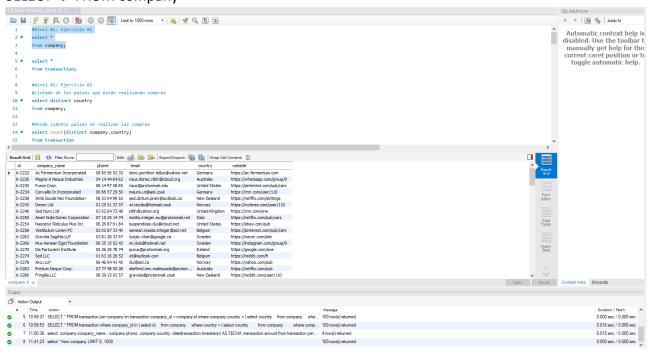
Muestra una relación 1:N entre las tablas company (estructura\_dades) y transaction (dades\_introduir). Una compañía puede estar vinculada a múltiples transacciones. La conexión se establece mediante por el *company\_id* en la tabla transaction, que referencia la columna *id* de la tabla company.

#### **ABRIR LAS TABLAS:**



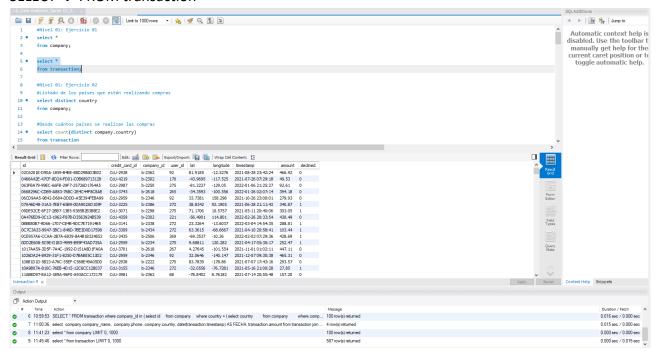


#### SELECT → FROM company



Retorna 587 lineas

#### SELECT → FROM transaction

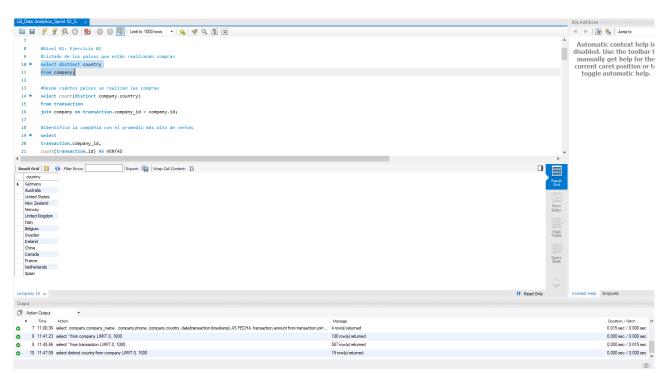


Retorna 587 lineas. Son las dos tablas con los valores existentes. Para poder "verlas" hay que escribir SELECT \* (asterisco = para traer todos los valores introducidos). Despues FROM + el nombre de la tabla que se quiere sacar los datos.

Utilizando JOIN realizarás las siguientes consultas:

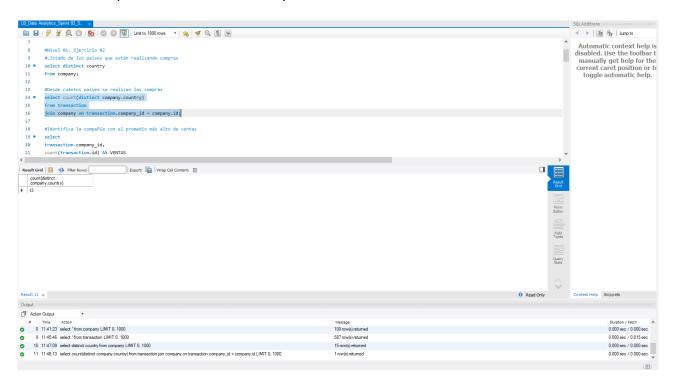
- Listado de los países que están realizando compras.
- Desde cuántos países se realizan las compras.
- Identifica la compañía con el promedio más alto de ventas.

Listado de los países que están realizando compras.



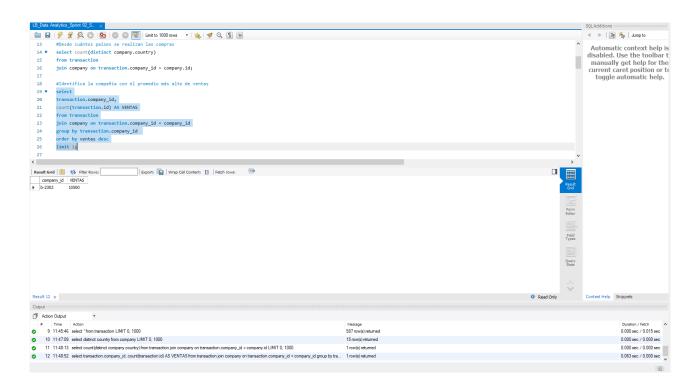
Usando el comando seleccionar (select) con distinct para evitar repeticiones, nos salen 15 países que están realizando compras.

Desde cuántos países se realizan las compras:



Con count contamos el numero de países y el comando distinct para evitar repetirlos. Los campos company\_id de la tabla **transaction** y el campo country de la tabla **company**. Con JOIN para combinar las tablas (conectándolas a través de la relación entre transaction.company\_id y company.id). Nos resultan 15 países.

Se identifica la compañía con el promedio más alto de ventas:

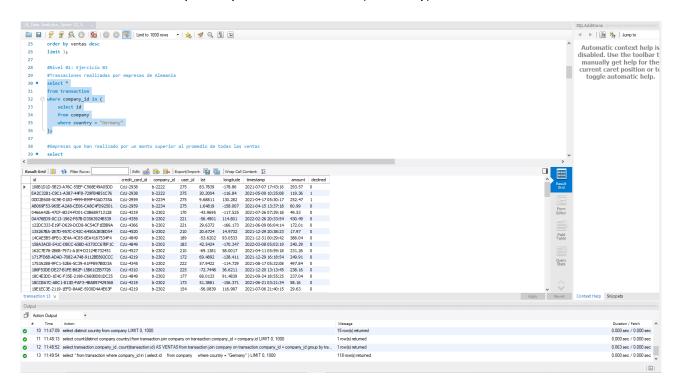


Como resultado, la compañía b\_2302, fue la que realizó más compras con un total de 10500. SELECT para seleccionar el identificador de la empresa (transaction.company\_id) y contar el número de transacciones asociadas con COUNT(transaction.id), asignando el nombre ventas al resultado. Luego, se conectaron las tablas transaction y company mediante JOIN. Con GROUP BY transaction.company\_id, se agrupan las transacciones por empresa, y con ORDER BY ventas DESC para ordenar los resultados de mayor a menor número de transacciones. Por fin, usado el comando limit 1 para delimitar solamente una respuesta.

Utilizando solo subconsultas (sin utilizar JOIN):

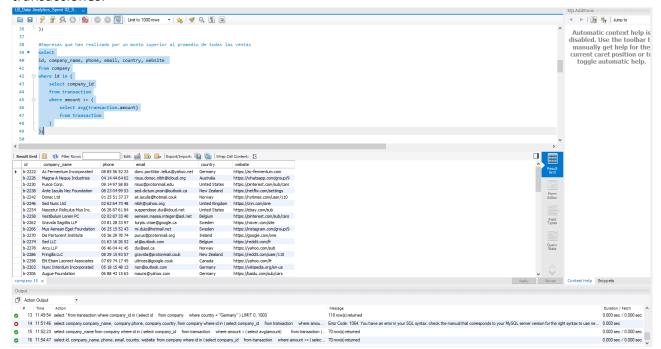
- Muestra todas las transacciones realizadas por empresas de Alemania.
- Lista las empresas que han realizado transacciones por un monto superior al promedio de todas las transacciones.
- Se eliminarán del sistema las empresas que no tienen transacciones registradas; entrega el listado de estas empresas.

Transacciones realizadas por empresas de Alemania (Germany):



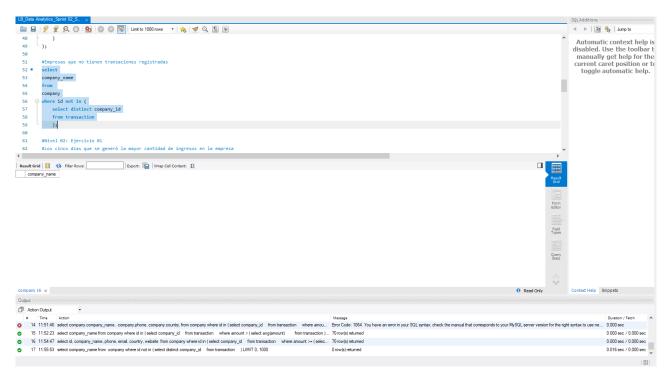
Se utilizó una subconsulta con *WHERE*, primero, seleccioné el *company\_id* de la tabla **transaction** y lo comparé con los *id* de empresas cuya columna *country* sea **Germany** en la tabla **company**.

Empresas que han realizado transacciones por un monto superior al promedio de todas las transacciones:



La subconsulta calculó el promedio general de los montos (SELECT AVG(amount) FROM transaction). Luego, se seleccionaron los company\_id de las empresas cuyas transacciones tienen un monto mayor a este promedio.

Empresas que no tienen transacciones registradas:

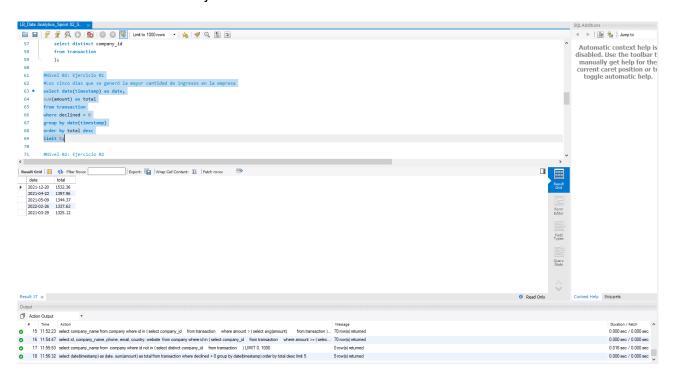


No ha retornado ninguna empresa. La subconsulta dentro de *WHERE NOT IN*, selecciona los *company\_id* desde la tabla **transaction**. Con *NOT IN*, filtrado de las empresas de la tabla **company** cuyo *id* no aparece en los resultados de la subconsulta.

# **NIVEL 02**

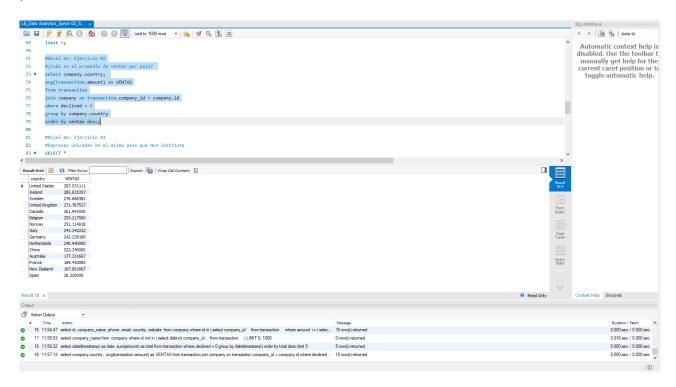
# **EJERCICIO 01**

Identifique los cinco días en los que se generó el mayor ingreso para la empresa por ventas. Muestre la fecha de cada transacción junto con el total de las ventas.



Aquí se presentan los 5 días con mayor ingreso, utilizando el comando de *LIMIT 5*. En la lateral de la fecha si descriminan el valor total de las ventas *SUM(AMOUNT)*.

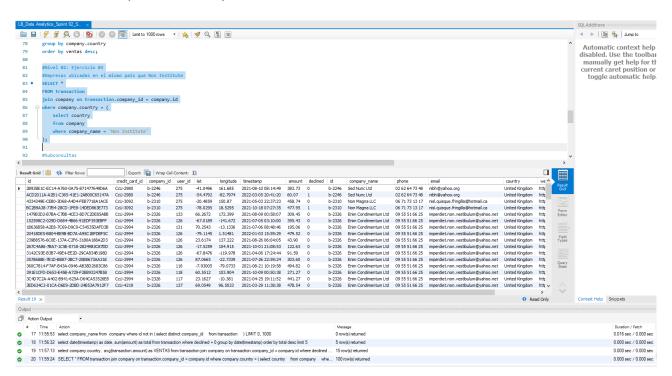
¿Cuál es el promedio de ventas por país? Presente los resultados ordenados de mayor a menor promedio.



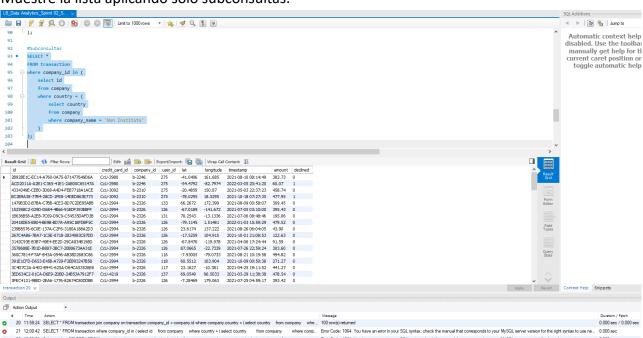
Calculamos el promedio de ventas por país y mostramos los resultados ordenados del mayor a menor. Lo que hice fue tomar las tablas **transaction** (con las transacciones), y **company** (con la información de las empresas). Cálculo del promedio de ventas con *AVG*(*transaction.amount*) (tabla transactions +. columna amount) renomeando como VENTAS con los datos de la tabla **transaction**. Usé *JOIN* para conectar las tablas, relacionando el *company\_id* de las tabla transacciones con el *id* de las empresas de la tabla company. Luego, filtré las transacciones que no fueron rechazadas con un *WHERE declined = 0*. Agrupé los datos por país usando *GROUP BY company.country* (tabla company +. columna country). Para ordenarlo utilicé *ORDER BY* para que los países con mayores promedios de ventas aparezcan primero, con el *DESC* (descendiente).

Su empresa está considerando lanzar campañas publicitarias para competir con la compañía "Non Institute". Para ello, solicitan la lista de todas las transacciones realizadas por empresas ubicadas en el mismo país que dicha compañía.

Muestre la lista aplicando JOIN y subconsultas.



Aquí debemos encontrar todas las transacciones realizadas por empresas que están en el mismo país que **Non Institute**. Para hacerlo, primero conecté dos tablas: **transaction** (que contiene los datos de las transacciones) y **company** (que tiene información sobre las empresas), usando *JOIN*, relacionando la columna *company\_id* de la tabla **transaction** con la columna *id* de la tabla **company**. Usando *WHERE* para buscar solo las empresas ubicadas en el mismo país que **Non Institute**. Para esto, utilicé una subconsulta que selecciona el país (*country* – columan de la tabla **company**) de la empresa **Non Institute**.

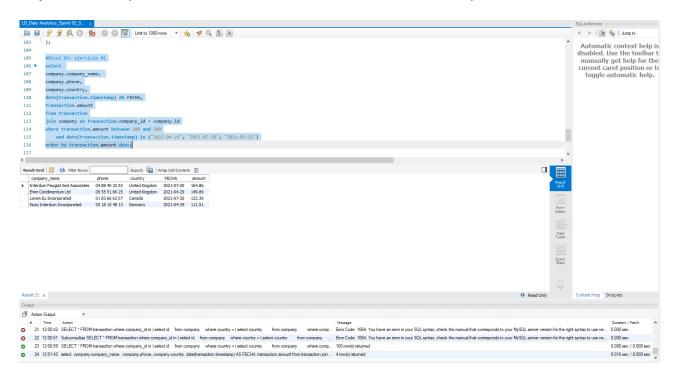


Muestre la lista aplicando solo subconsultas.

Para llegar a este resultado utilizando solo subconsultas, el objetivo fue listar todas las transacciones realizadas por empresas ubicadas en el mismo país que **Non Institute**. Primero, seleccioné toda la información de la tabla **transaction**, si aplicó *WHERE company\_id IN* para limitar las transacciones a aquellas realizadas por empresas específicas. Para obtener esas empresas, utilicé una subconsulta que buscó los *id* de las empresas que están en el mismo país que **Non Institute.** 

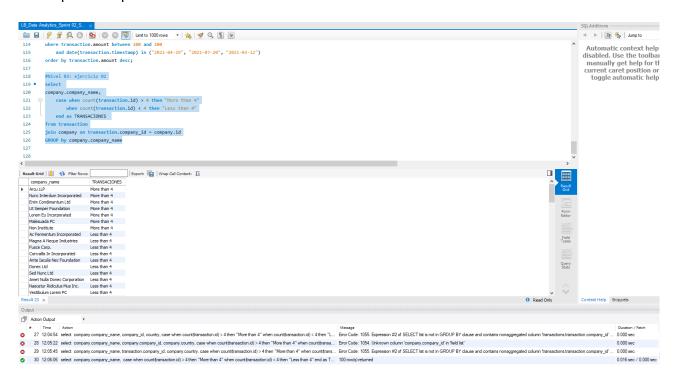
Para determinar el país de **Non Institute**, usé una subconsulta adicional dentro de la anterior, que selecciona el país desde la tabla **company** donde el nombre de la empresa coincide con la empresa e cuestión. Con este dato, la subconsulta principal seleccionó todas las empresas en ese país, y con *WHERE* relacionó las transacciones con esas empresas. El resultado final muestra únicamente las transacciones de empresas que comparten país con **Non Institute**.

Presente el nombre, teléfono, país, fecha y valor (amount) de las empresas que realizaron transacciones con valores entre 100 y 200 euros en alguna de estas fechas: 29 de abril de 2021, 20 de julio de 2021 y 13 de marzo de 2022. Ordene los resultados de mayor a menor valor.



Otra vez, conecté las tablas **transaction** y **company** usando un *JOIN*, relacionando la columna *company\_id* en la tabla **transaction** con la columna *id* en la tabla **company**. Si aplica *WHERE* con *t.amount BETWEEN 100 AND 200* (las transacciones cuyo valor estaba dentro del rango indicado). Si añade la condición con *AND DATE(t.timestamp) IN* para incluir solo las transacciones que ocurrieron en las fechas indicadas para el ejercicio. Por último, para ordenar los resultados de mayor a menor valor, usase *ORDER BY* (ordenar por) *transactio.amount* (tabla transaction +. columna amount) *DESC* (los valores se mostrarán comenzando por el más alto).

Para optimizar la asignación de recursos operativos, es necesario conocer la cantidad de transacciones realizadas por empresa. El departamento de recursos humanos requiere una lista de las empresas especificando si tienen más o menos de 4 transacciones.



Objetivo era listar las empresas y clasificarlas según la cantidad de transacciones que han realizado, indicando si tienen "Más de 4" o "Menos de 4" transacciones. Con la función *CASE WHEN* para crear una columna personalizada llamada TRANSACIONES. Aquí se configura una condición: si el número de transacciones (COUNT(transaction.id)) es mayor a 4, el resultado será "More than 4". Si es menor a 4, mostrará "Less than 4". Aplicando el CASE WHEN dentro del SELECT, lo que permite clasificar los datos directamente en la consulta y al final añadir END para finalizar el comando. Conectar las tablas transaction y company utilizando JOIN, relacionando el company\_id de la tabla transaction con el id de la tabla company. Por ultimo, GROUP BY company.company\_name para agrupar los resultados por el nombre de la empresa y contar sus transacciones. El resultado muestra el nombre de cada empresa y una clasificación indicando si tienen más o menos de 4 transacciones.