**Project Title: Chinook Data Analytics - Unlocking Business Insights from a Digital Media Store**

**Project Overview:**

This project involves an in-depth exploration and analysis of the **Chinook database**, a comprehensive sample dataset representing a digital media store. Leveraging SQL (Structured Query Language), the goal is to extract, transform, and analyze various facets of the store's operations, including sales performance, customer behaviour, employee efficiency, and media catalogue trends. This project will demonstrate proficiency in a wide range of SQL concepts, from basic data retrieval to complex analytical queries, ultimately providing actionable business insights.

**Context:**

As of July 2025, understanding customer purchasing habits, optimizing inventory, and evaluating sales strategies are more critical than ever for digital media businesses. This project simulates a real-world scenario where a database professional is tasked with providing data-driven recommendations to improve the Chinook store's profitability and operational efficiency.

**Project Objectives:**

The project will address key business questions and involve the following objectives:

**Database Understanding & Schema Navigation:**

* + Thoroughly understand the Chinook database schema, including all tables (Artists, Albums, Tracks, Customers, Employees, Invoices, Invoice Lines, Genres, Media Types, Playlists, Playlist Track) and their relationships (Primary Key/Foreign Key constraints).
  + Create an Entity-Relationship Diagram (ERD) or describe the key relationships within the database.

**Core Data Retrieval & Manipulation:**

* + Practice basic SELECT statements with WHERE, ORDER BY, and LIMIT/TOP clauses.
  + Demonstrate proficiency in INSERT, UPDATE, and DELETE operations for data management (e.g., adding a new track, updating customer details).

**Advanced Querying & Data Aggregation:**

* + **Joining Tables:** Utilize various types of JOIN clauses (INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN) to combine data from multiple tables to answer complex questions (e.g., "List all tracks by a specific artist and their genre").
  + **Aggregation and Grouping:** Employ GROUP BY and aggregate functions (COUNT, SUM, AVG, MIN, MAX) with HAVING clauses to perform statistical analysis (e.g., "Calculate total sales per country," "Find the average invoice amount for each customer").
  + **Subqueries & Common Table Expressions (CTEs):** Apply subqueries and/or CTEs for multi-step data processing and improved query readability (e.g., "Find customers who have spent more than the average customer spending").
  + **Window Functions:** Explore the use of window functions for more sophisticated analytical tasks (e.g., "Rank employees by sales performance within their respective departments").

**Business Insights & Reporting:**

* + **Sales Analysis:**
    - Identify top-performing artists and genres by revenue.
    - Analyze monthly/quarterly sales trends over the four-year period.
    - Determine the average value of an invoice and the average number of items per invoice.
  + **Customer Behaviour Analysis:**
    - Segment customers based on their total spending.
    - Identify the top 10 most valuable customers.
    - Determine the most popular genres or artists among specific customer groups.
  + **Employee Performance:**
    - Evaluate sales performance of individual employees and their contribution to overall revenue.
    - Analyze employee hierarchy and reporting structures.
  + **Media Catalogue Analysis:**
    - Identify the distribution of tracks by genre and media type.
    - Analyze the average duration of tracks per album or genre.

**Technology Used:**

* **Database:** Chinook Database
* **Language:** SQL (MySQL)
* **Tools:** MySQL Workbench

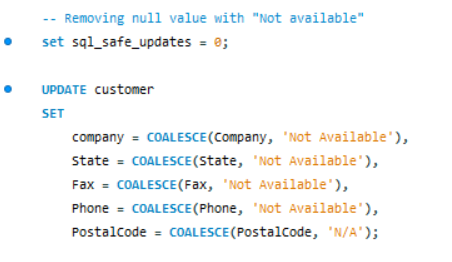
**Deliverables:**

* A documented SQL script file (.sql) containing all queries used to achieve the project objectives, clearly commented and organized by objective.
* A brief project report summarizing the key findings, insights derived from the SQL queries, and potential business recommendations.

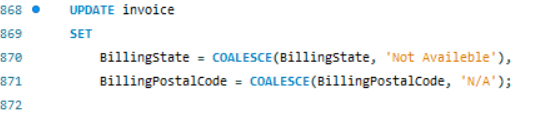
**Data Cleaning**

**Replace Null value in each table** 🡪

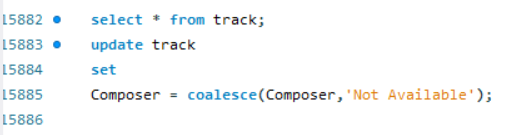
**Customer Table --**



**Invoice Table –**



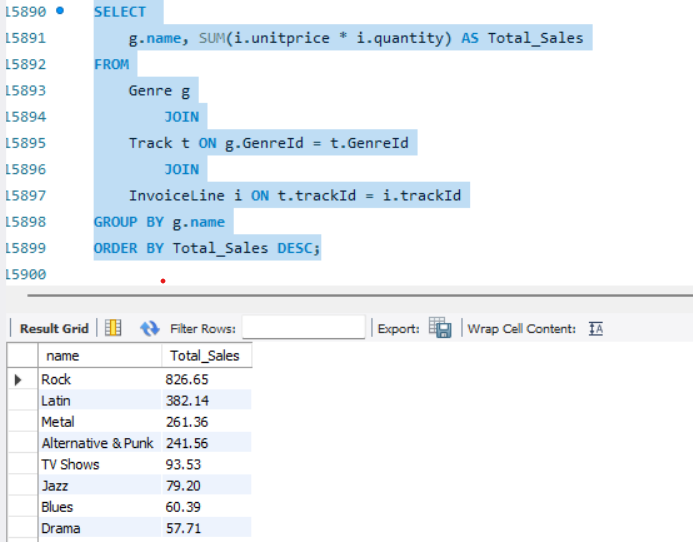
**Table Track –**



**Data Analysis**

**Q.1 Total Sales by Genre:**

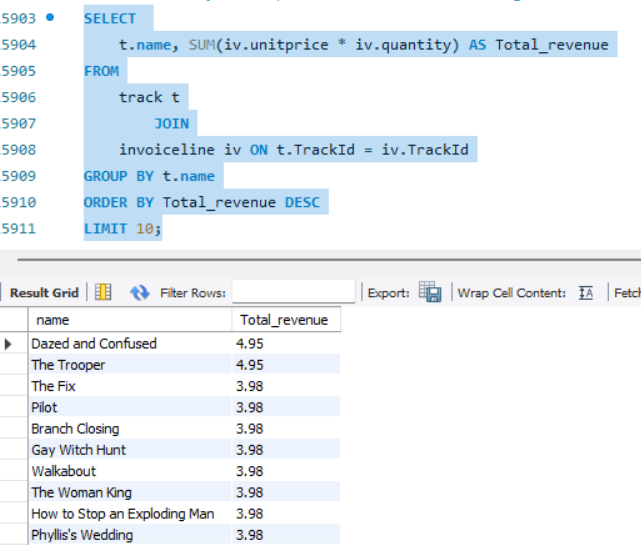
🡪



**Q.2 Top 10 Selling Tracks:**

-- Identify the top 10 Track Names that have generated the highest total revenue. Include the track name and total revenue.

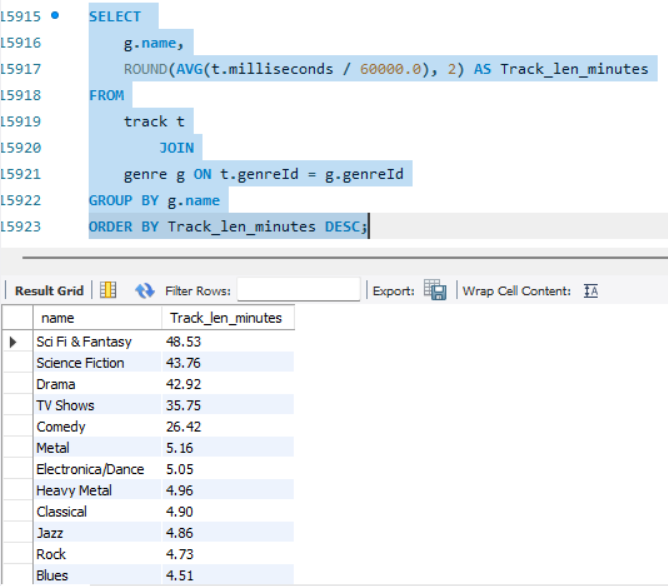
🡪



**Q.3 Average Track Length by Genre:**

-- For each Genre Name, calculate the average Milliseconds of its tracks. Convert milliseconds to minutes.

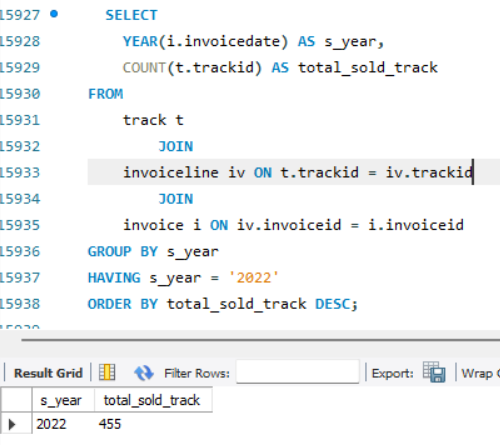
🡪



**Q.4 Tracks Sold in a Specific Year:**

-- Count the total number of Invoice Line items that occurred in the year 2022.

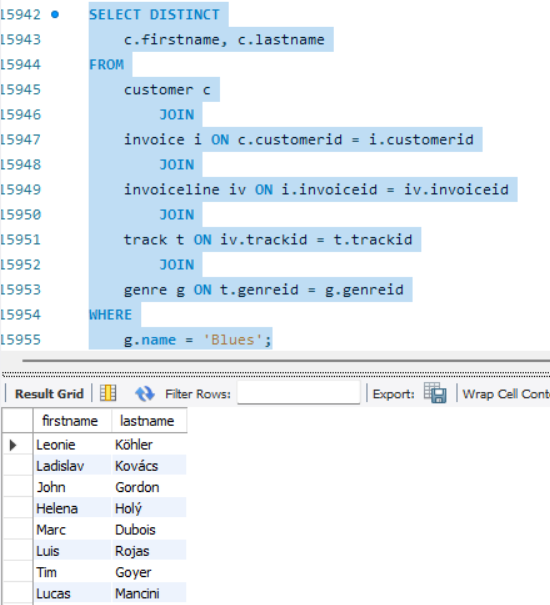
🡪



**Q.5 Customers Who Purchased 'Blues' Genre:**

-- List FirstName, LastName of Customers who have purchased at least one Track from the 'Blues' Genre. Do not list duplicates.

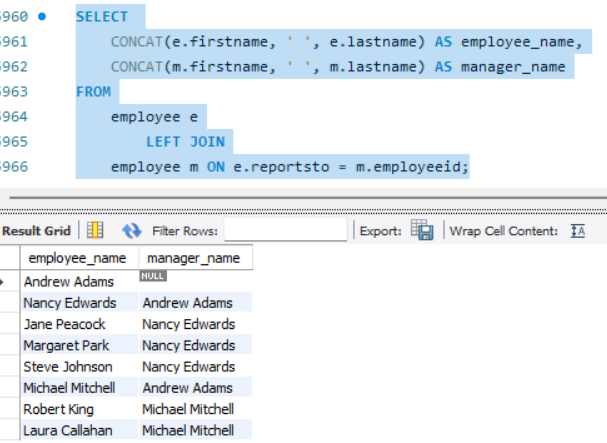
🡪



**Q.6 Employees and Their Direct Manager:**

-- List Employee FirstName, LastName (as 'Employee Name') and their manager's FirstName, LastName (as 'Manager Name').

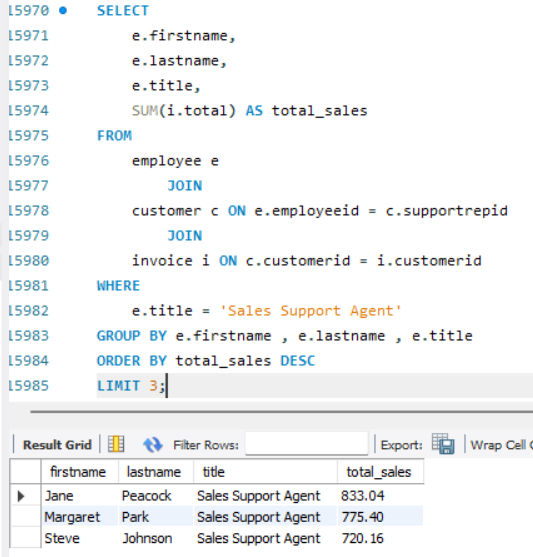
🡪



**Q.7 Top 3 Support Representatives by Sales:**

-- Identify the top 3 Employees (FirstName, LastName, Title) who are 'Sales Support Agent's and have generated the most total sales (Invoice, Total) from their supported customers.

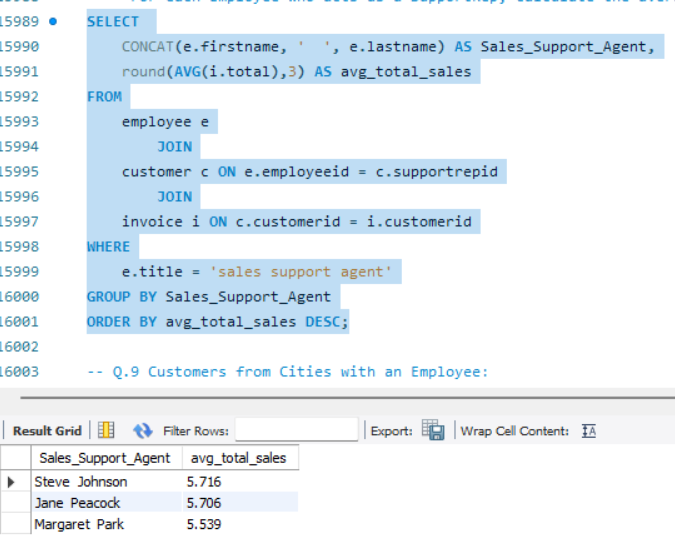
🡪



**Q.8 Average Sales per Employee (as Support Rep):**

-- For each Employee who acts as a SupportRep, calculate the average Total of the Invoices associated with their customers

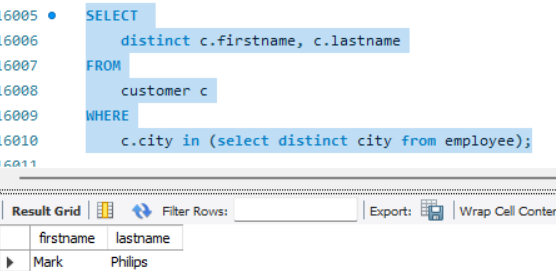
🡪



**Q.9 Customers from Cities with an Employee:**

-- Find the FirstName, LastName of Customers who live in the same City as at least one Employee. Do not list duplicates.

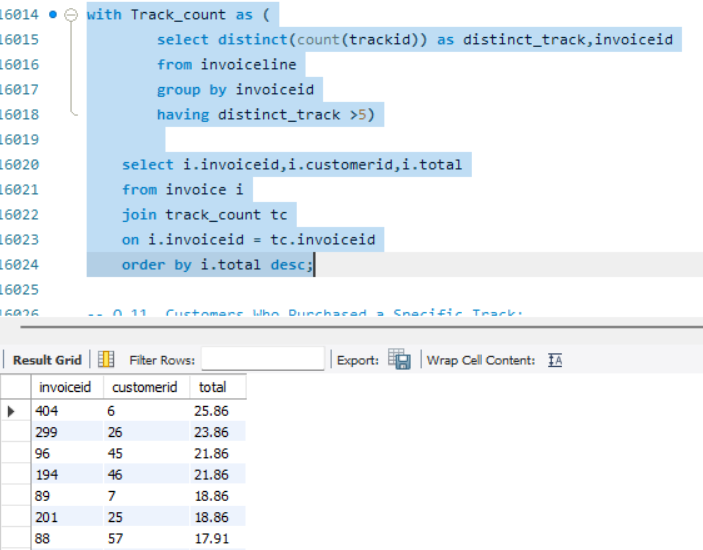
🡪



**Q.10 Invoices with More Than 5 Tracks:**

-- List InvoiceId, CustomerId, and Total for invoices that contain more than 5 distinct Tracks.

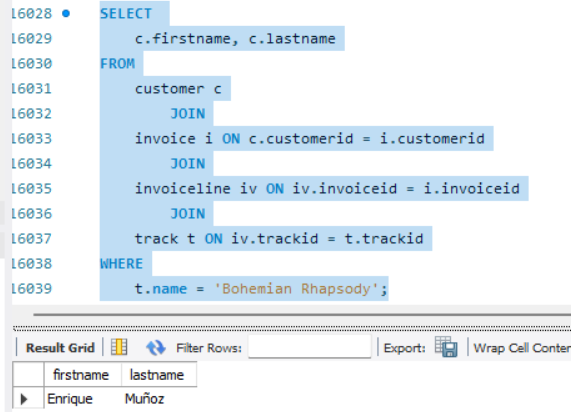
🡪



**Q.11 Customers Who Purchased a Specific Track:**

-- List FirstName, LastName of Customers who purchased the track 'Bohemian Rhapsody’.

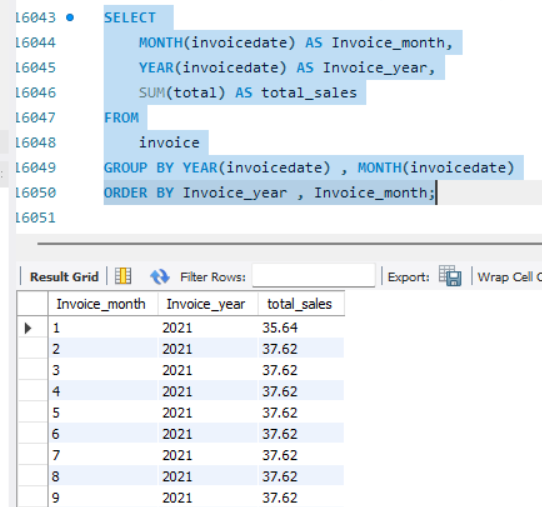
🡪



**Q.12 Monthly Sales Trend:**

-- Calculate the total sales (Invoice.Total) for each month of each year. Display InvoiceYear, InvoiceMonth, and TotalSales, ordered chronologically.

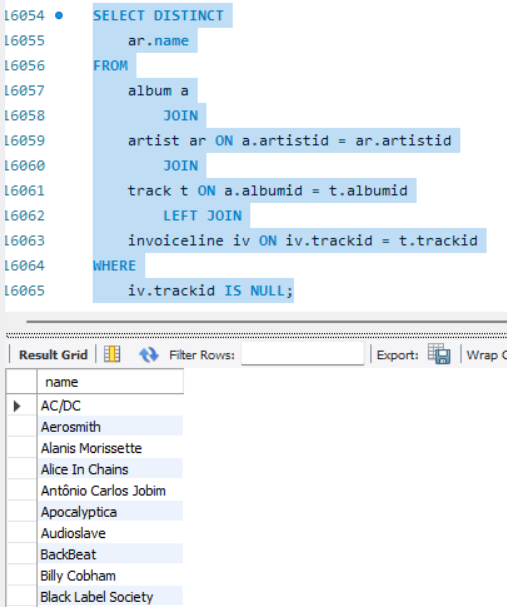
🡪



**Q.13 Artists Whose Tracks Have Never Been Sold:**

-- List Artist Names whose Albums have Tracks that have never appeared in any Invoice Line.

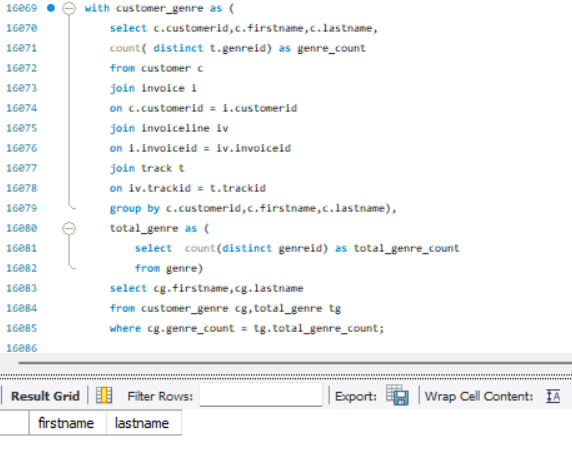
🡪



**Q.14. Customers Who Purchased Every Genre:**

-- Find the FirstName and LastName of Customers who have purchased at least one Track from every single Genre available in the store.

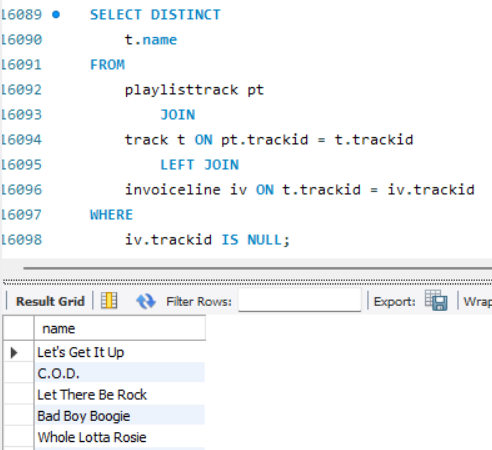
🡪



**Q.15 Tracks Available in Playlists but Not Sold:**

-- Find Track Names that exist in at least one Playlist but have never been sold in an Invoice Line.

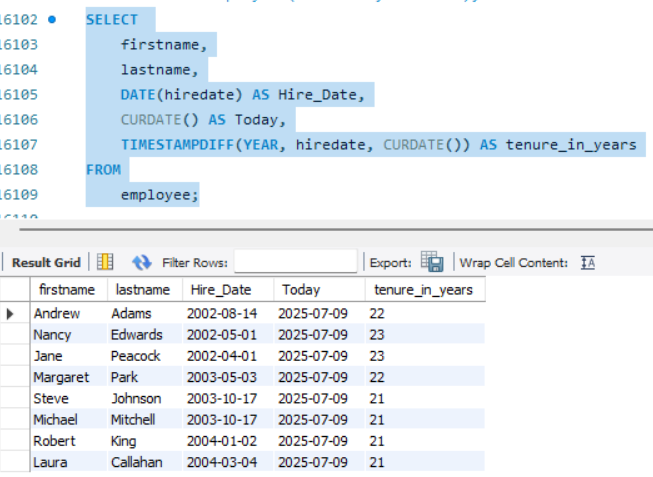
🡪



**Q.16 Employee Tenure (in years):**

-- For each Employee (FirstName, LastName), calculate their tenure in years from their Hire Date to the current date.

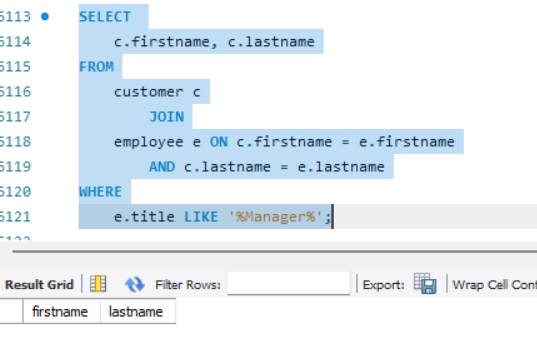
🡪



**Q.17 Customers Who Are Managers:**

-- Identify if any Customer (FirstName, LastName) is also an Employee and holds a Title like 'Manager' or 'General Manager'.

🡪



**Q.18 Customer Acquisition by Support Rep:**

-- For each SupportRep (FirstName, LastName), count the number of new Customers they acquired in each year based on Customer.SupportRepId and Customer.

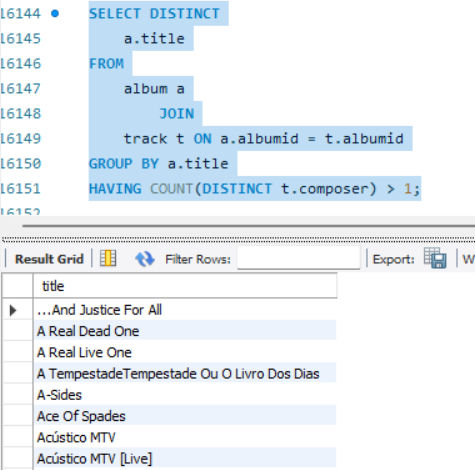
🡪



**Q.19 Albums with Tracks from Multiple Composers:**

-- List Album Titles that contain tracks from more than one unique Composer.

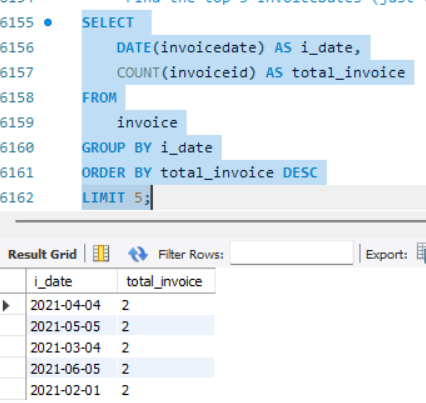
🡪



**Q.20 Top 5 Busiest Invoice Dates:**

-- Find the top 5 Invoice Dates (just the date part) that had the highest number of Invoices.

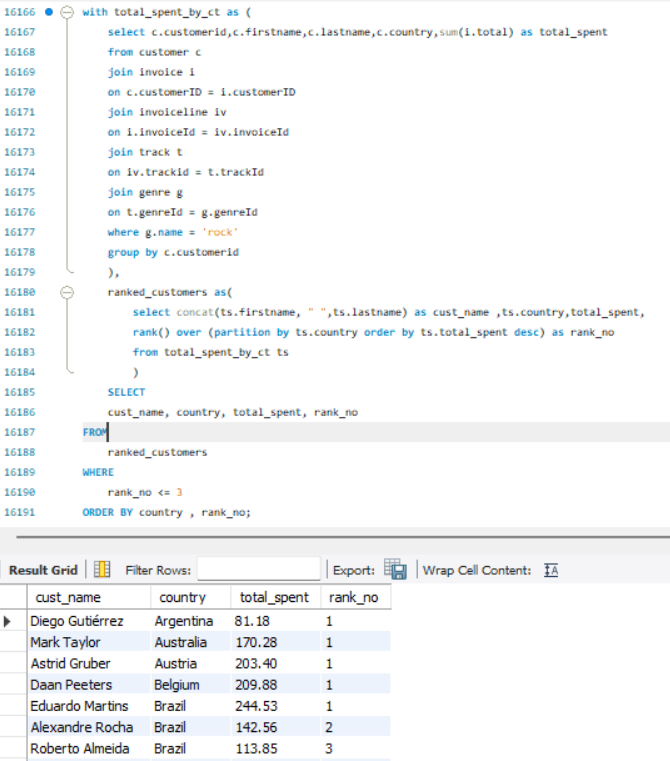
🡪



**Q.21 Top Spenders by Genre:**

-- Find the top 3 customers in each country who have spent the most money on "Rock" music. For each customer, list their full name, country, and total amount spent on Rock music.

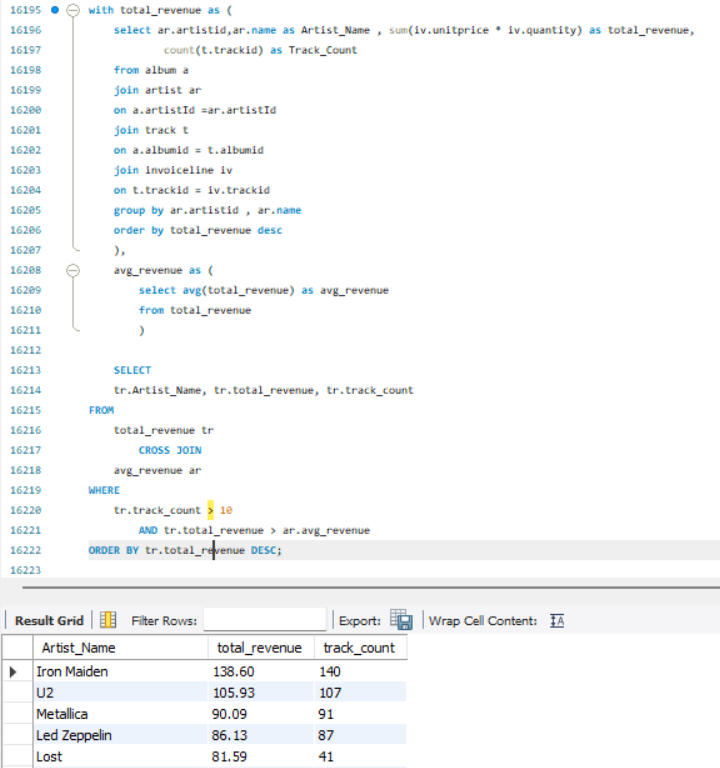
🡪



**Q.22 Artist Popularity by Track Count and Revenue:**

-- Identify artists who have more than 10 tracks and whose total revenue generated from sales is above the average revenue generated by all artists. Display the artist’s name, the number of tracks, and their total revenue, ordered by revenue in descending order.

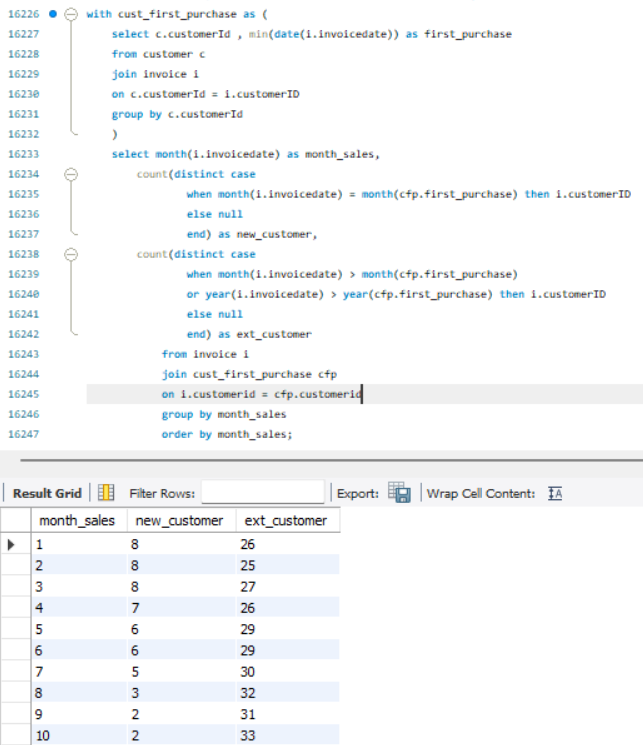
🡪



**Q.23 Customer Retention Analysis (Monthly):**

-- For each month, calculate the number of new customers acquired and the number of existing customers who made a purchase in that month.

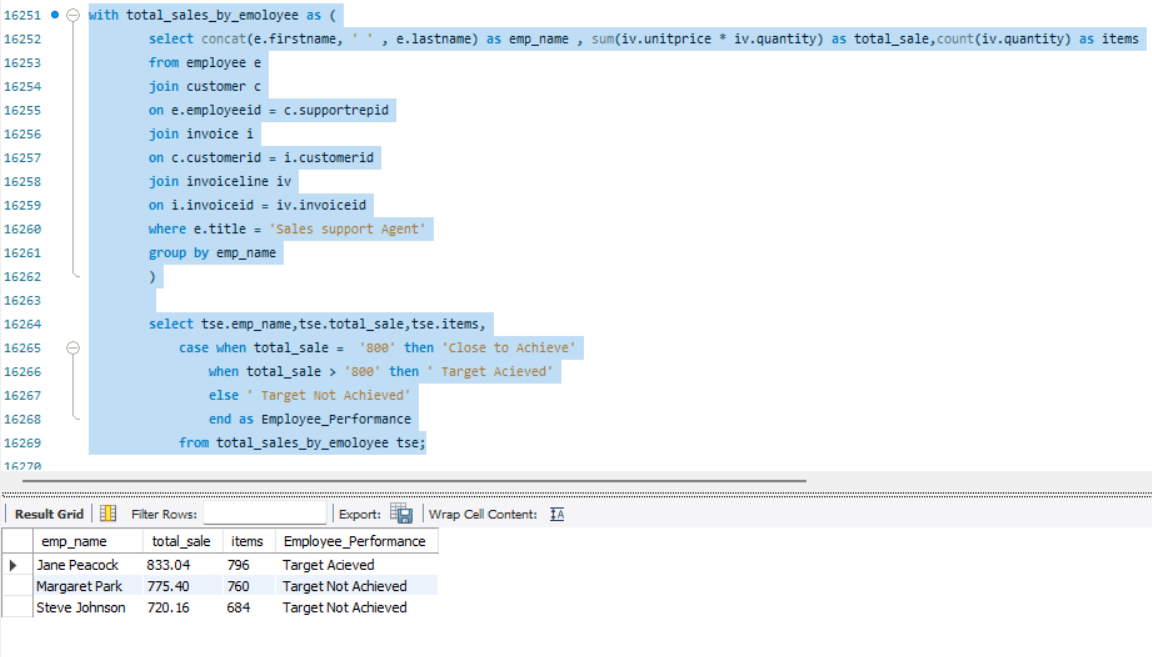
🡪



**Q.24 Employee Performance vs. Sales Targets:**

-- Assume a sales target of $800 for each invoice line item. For each sales support agent, calculate their total sales, the number of invoice line items they processed, and the percentage by which they exceeded or fell short of their hypothetical sales target.

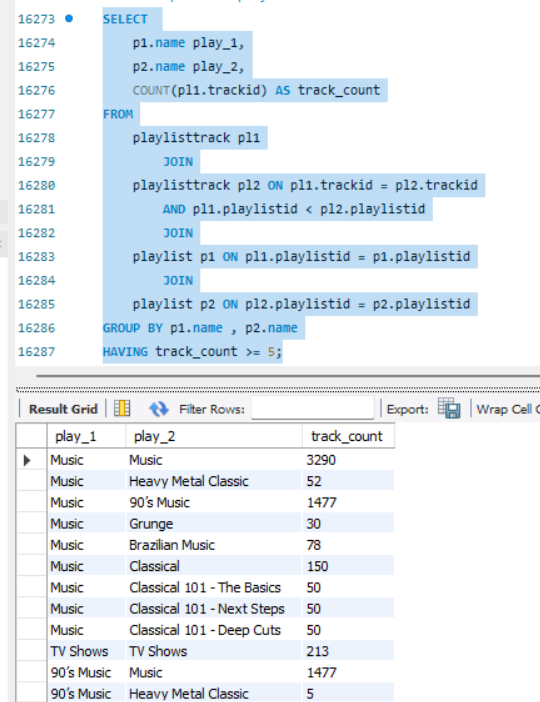
🡪



**Q.25 Playlist Content Overlap:**

-- Find pairs of playlists that share at least 5 common tracks. For each pair, display the names of both playlists and the count of shared tracks.

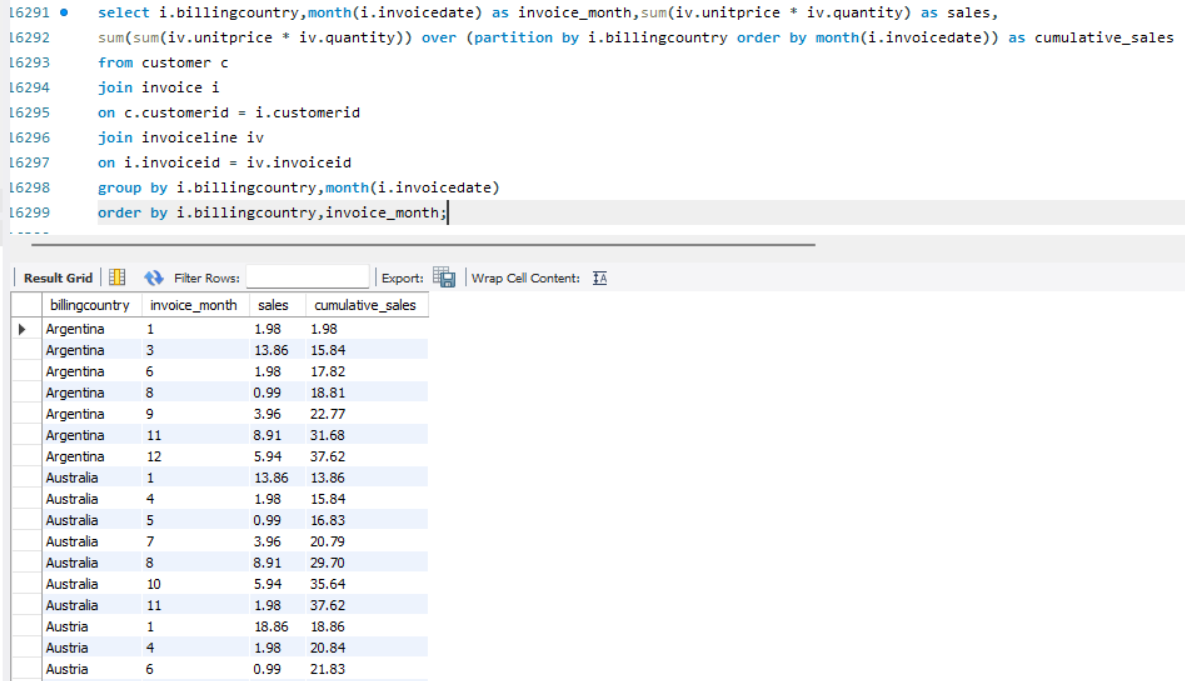
🡪



**Q.26 Cumulative Sales by Country and Month:**

-- Calculate the cumulative total sales for each country, month by month.

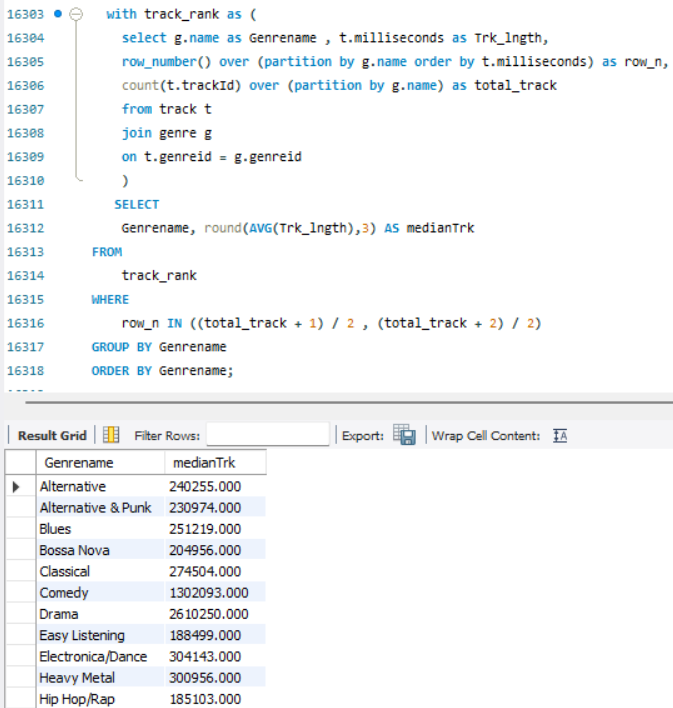
🡪



**Q.27 Median Track Length by Genre:**

-- For each genre, find the median track length (in milliseconds).

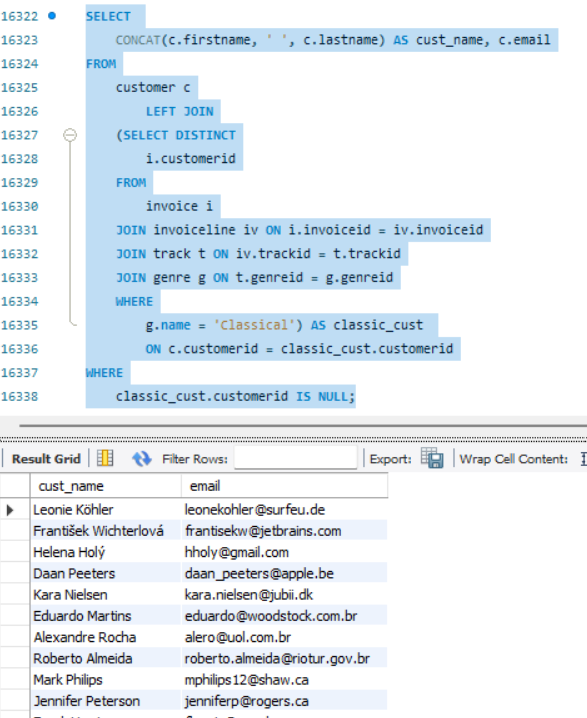
🡪



**Q.28 Customers Who Haven't Purchased a Specific Genre:**

-- Find all customers (full name and email) who have never purchased a "Classical" music track.

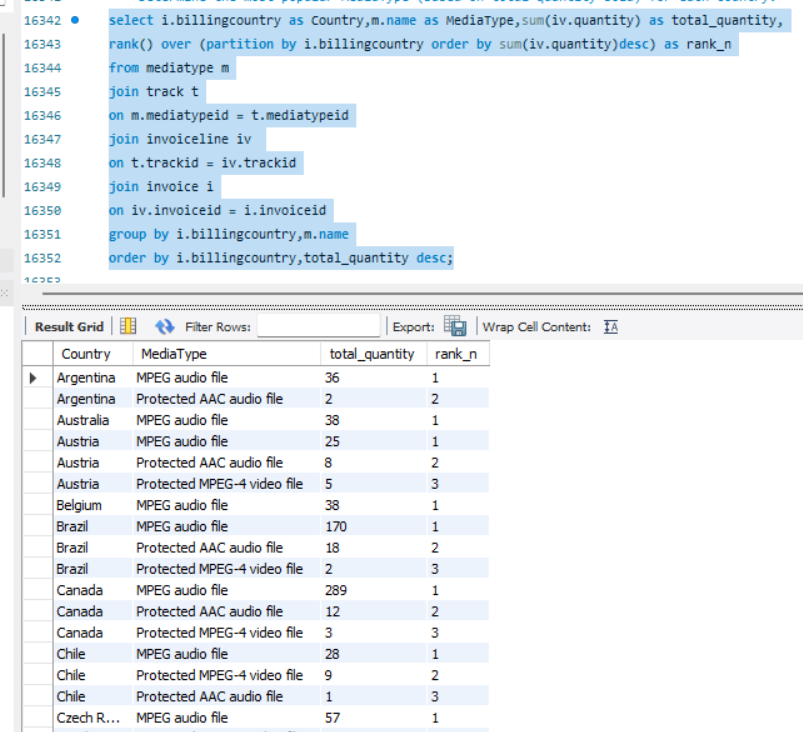
🡪



**Q.29 Most Popular Media Type per Country:**

-- Determine the most popular MediaType (based on total quantity sold) for each country.

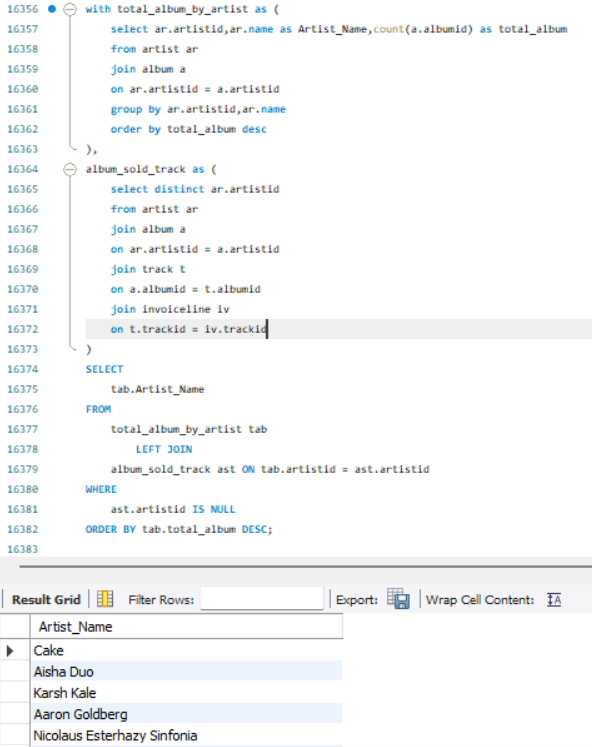
🡪



**Q.30 Artist with Most Albums Without Any Tracks Sold:**

-- Find the artist(s) who have the most albums in the database, but none of their tracks have ever been sold.

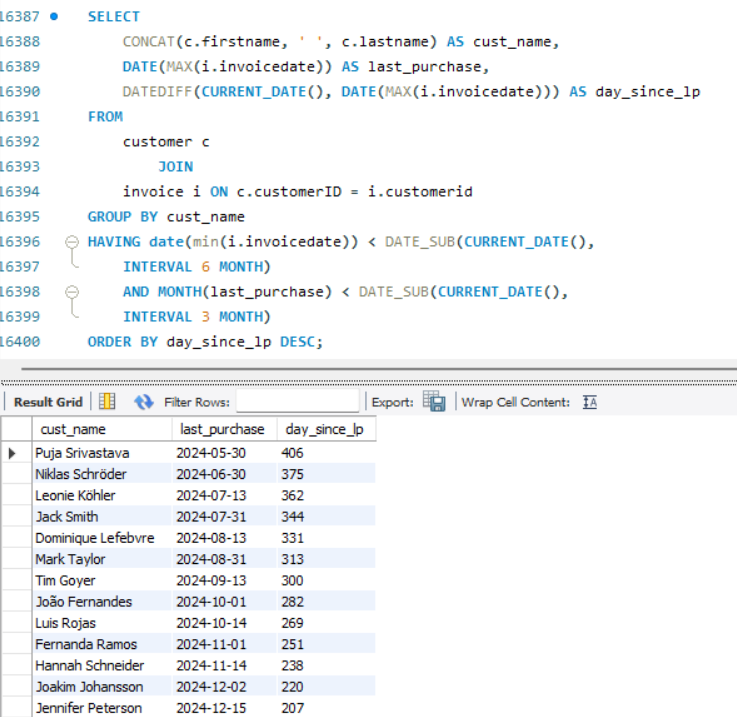
🡪



**Q.31 Churn Prediction - Customers at Risk:**

-- Identify customers who made their first purchase more than 6 months ago, but haven't made any purchases in the last 3 months. For these customers, list their full name, their last purchase date, and the total number of days since their last purchase.

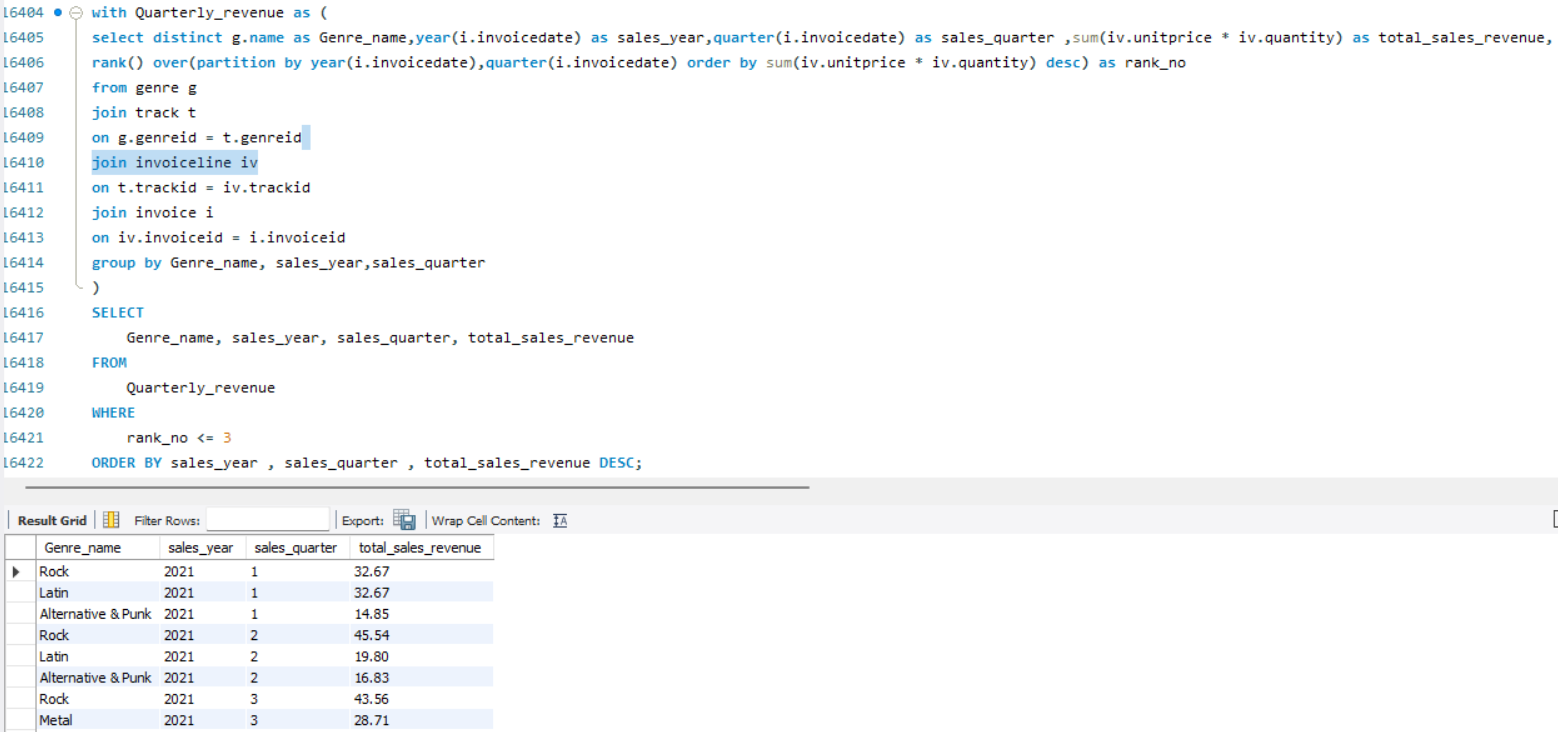
🡪



**Q.32 Genre Popularity Shift Over Time (Quarterly Analysis):**

-- For each quarter of each year present in the data, determine the top 3 most popular genres by total sales revenue. Display the year, quarter, genre name, and the total revenue for that genre in that quarter.

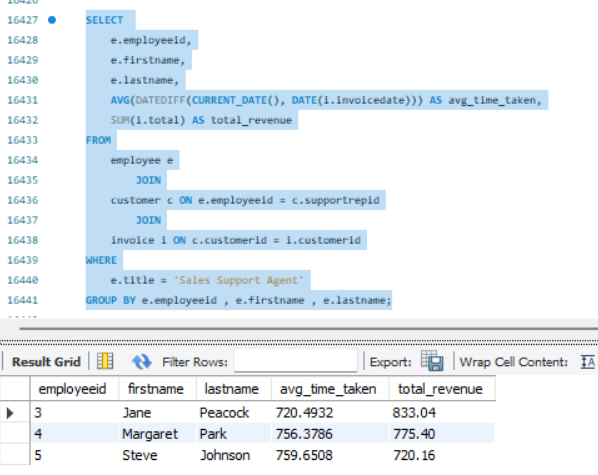
🡪



**Q.33 Employee Efficiency - Average Time to Close an Invoice:**

-- For each sales support agent (employee), calculate the average time (in days) between the Invoice Date and the date the invoice was paid (assume Billing Country is a proxy for "closed" if Invoice Date is the only date field). If no 'paid' date exists, assume its CURRENT\_DATE. Also, show the total revenue generated by each agent.

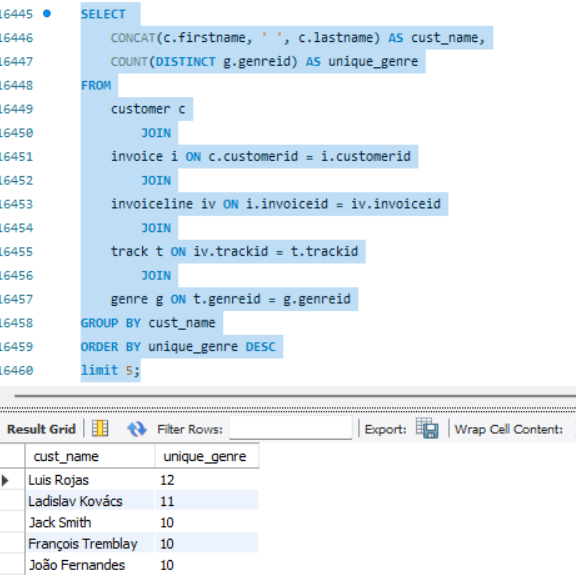
🡪



**Q.34 Most Diverse Customer Portfolios:**

-- Identify the top 5 customers who have purchased tracks from the most unique genres. For each of these customers, list their full name, and the count of unique genres they've purchased from.

🡪

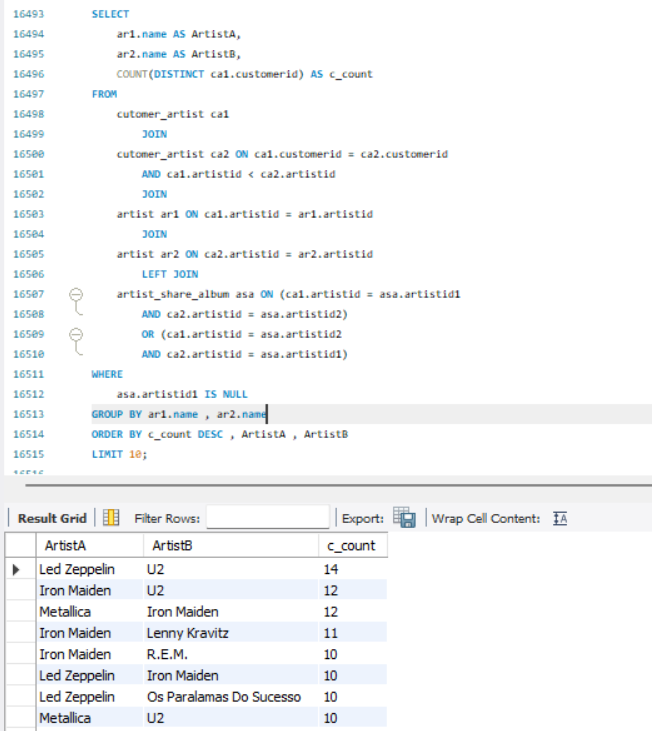


**Q.35 Artist-to-Artist Collaboration Network (Indirect Sales Influence):**

-- Identify pairs of artists where a customer who purchased a track from Artist A also purchased a track from Artist B, and these two artists are not directly related by having tracks on the same album. Quantify the "collaboration" by the number of unique customers who bought from both. Order the results by the collaboration count in descending order.

🡪





**Q.36 Impact of Track Length on Sales:**

-- Divide tracks into 5 equal length 'quintiles'. For each quintile, calculate the total sales revenue and the average number of times tracks from that quintile were purchased. `

🡪



**Report:**

"To better understand the operational dynamics of our simulated digital media store, this report utilizes SQL to query the Chinook database. Our objective is to identify top-performing areas and potential opportunities for growth."

**Insights:**

* **Most Selling Genre**: -

We can identify Popularity of Genre and most selling genre is “Rock”

* **Top Selling Tracks**: -

Top selling Tracks are “Dazed and Confused”, “The Trooper”, “The Fix”, etc.

* **Top 3 Employee as Sales representative**: -

Top 3 Employees are “Jane Peacock”, “Margret Park”, “Steve Johnson”

* **Monthly Trend Analysis**: -

We can see that Sales are increasing after 1st month.

* **Employee Tenure**: -

Most senior employee is “Nancy Edwards” who is working for 23 years.

* **Artist with no Track sold**: -

Artists whose track have not sold are “AC/DC”, “Aerosmith”, “Alice in chains” etc.

* **Most Popular Artists**: -

Most Popular Artists are “Iron Maiden”, “U2”, “Lost” etc.

* **Most Popular MediaType by Country**: -

In Argentina “MPEG audio file”, “Protected AAC audio file”,

In Australia “MPEG audio file”,

These are the Popular media types in each country.

* **Customer at Risk**: -

“Pooja Srivastav”, “Niclas Schröder”, “Leonie Köhler”, these are one of the customers who made their last purchase 350 days ago and losing interest in buying.

* **Most valuable Customer**: -

“Loise Rojas”, “Ladislav Kovacs”, “Jack Smith”

These are one of the most valuable customers as they have most diverse portfolio.