**SQL MODULE PROJECT**

**OBJECTIVE Q’S**

1. *Does any table have missing values or duplicates? If yes how would you handle it ?*

For identifying missing values in the tables, we use the general code lines:

SELECT \* FROM <tablename>

WHERE <first column of table> IS NULL

OR <second column> IS NULL

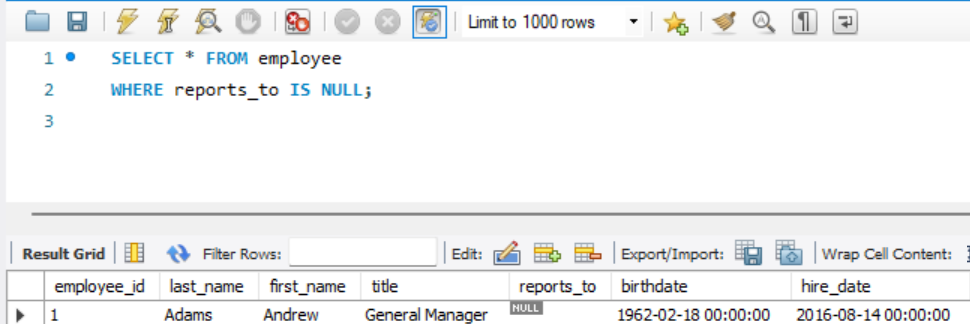
**.**

**.**

**.**

OR <final row> IS NULL;

A concrete application of the above:



For identifying duplicates:

SELECT column1, column2, ..., COUNT(\*)

FROM <tablename>

GROUP BY <uniquecolumn(s)>

HAVING COUNT(\*) > 1;

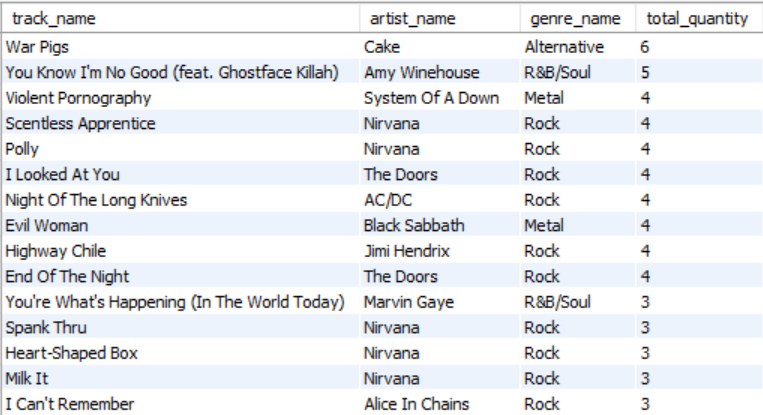
A concrete application of the above:

Hence, we find several missing values in:

* The ‘reports\_to’ column of ‘employee’.
* Multiple rows of the ‘composer’ column of ‘track’.
* Multiple rows of ‘company’, ‘state’, ‘fax’ and ‘postal\_code’ columns of ‘customer’ etc.

And no duplicates. The missing values may be handled via the COALESCE function, and we can specify either a carefully considered default value, or an aggregated mean.

1. *Find the top-selling tracks and top artist in the USA and identify their most famous genres.*



Approach used: SUM aggregation function, Joining relevant tables and filtering for USA-only.

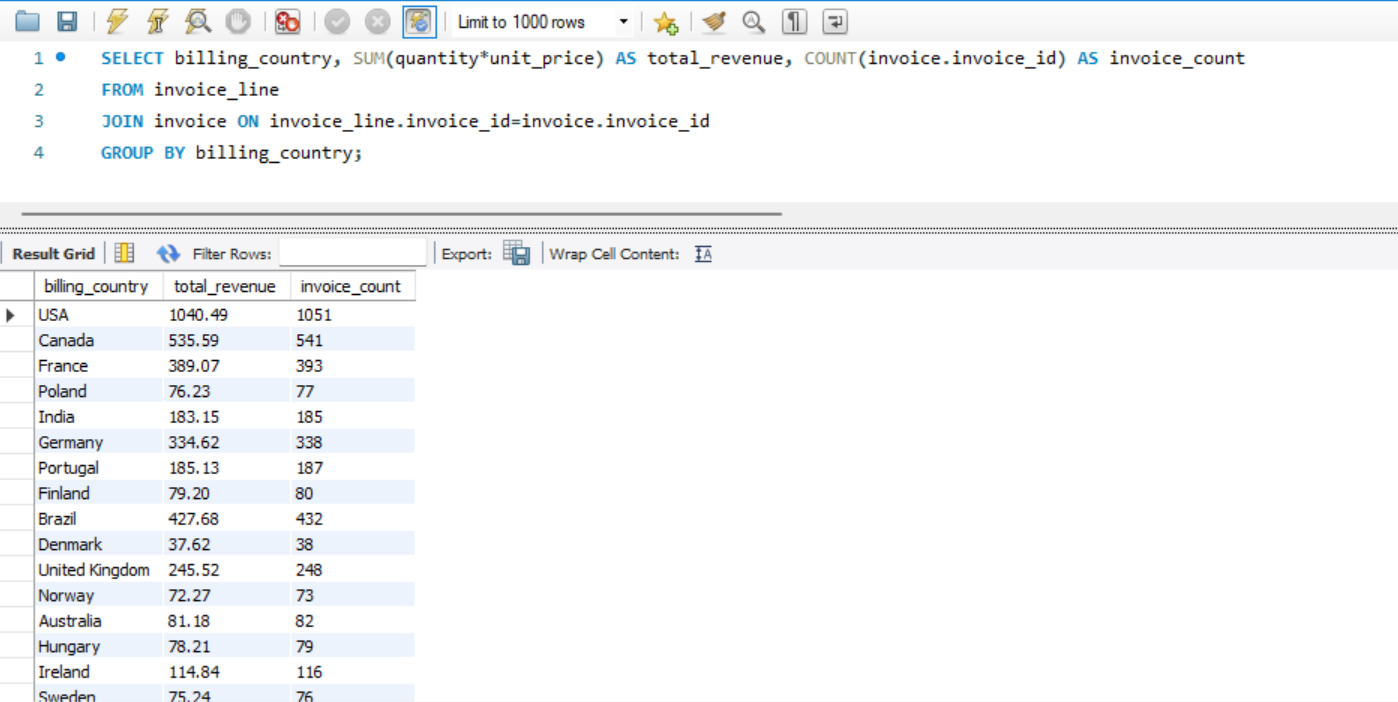
1. *What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?*



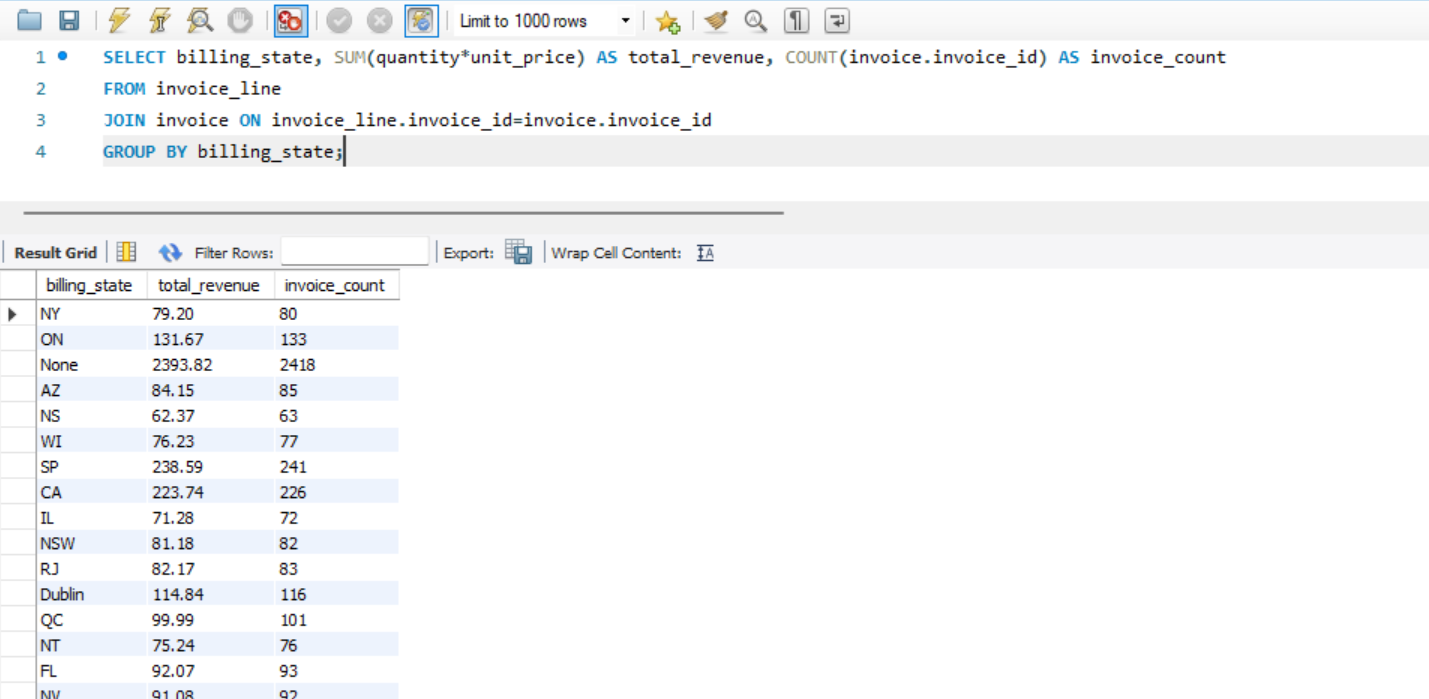
Approach used: COUNT aggregation function.

1. *Calculate the total revenue and number of invoices for each country, state, and city.*

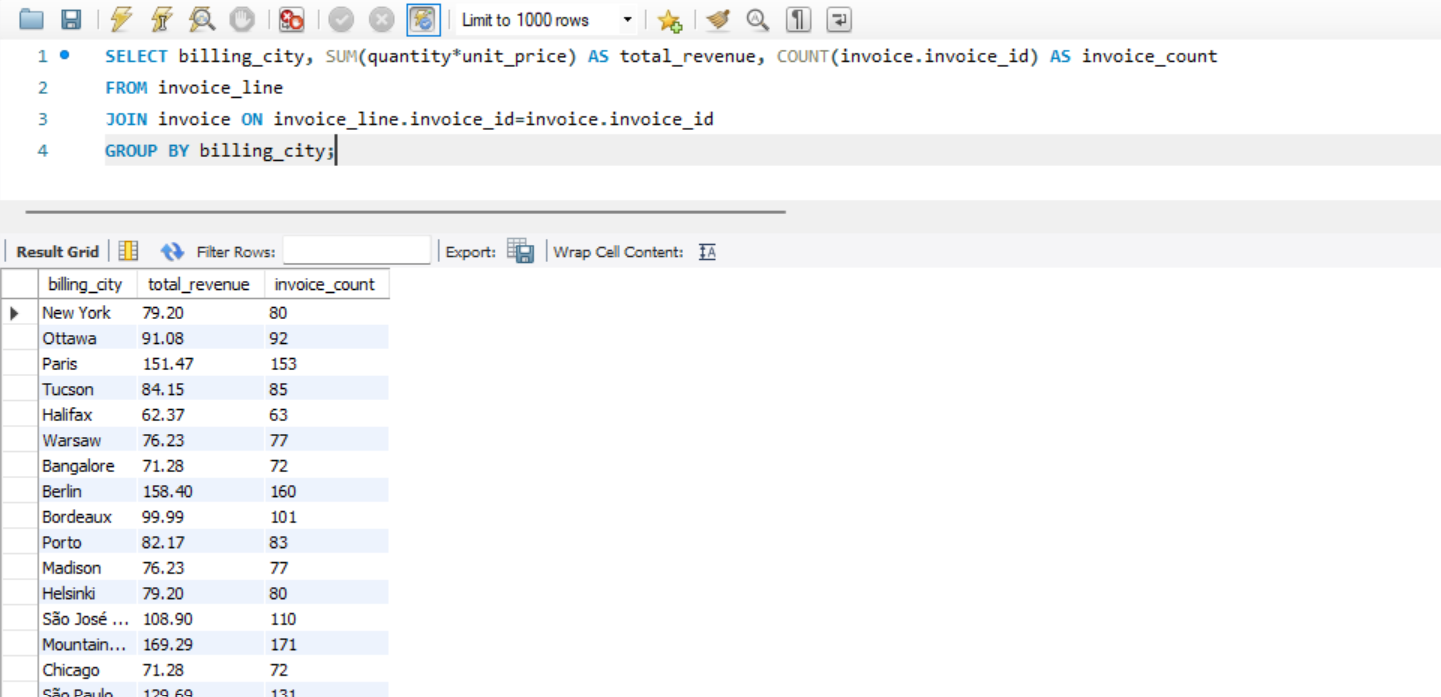
By country:



By state:

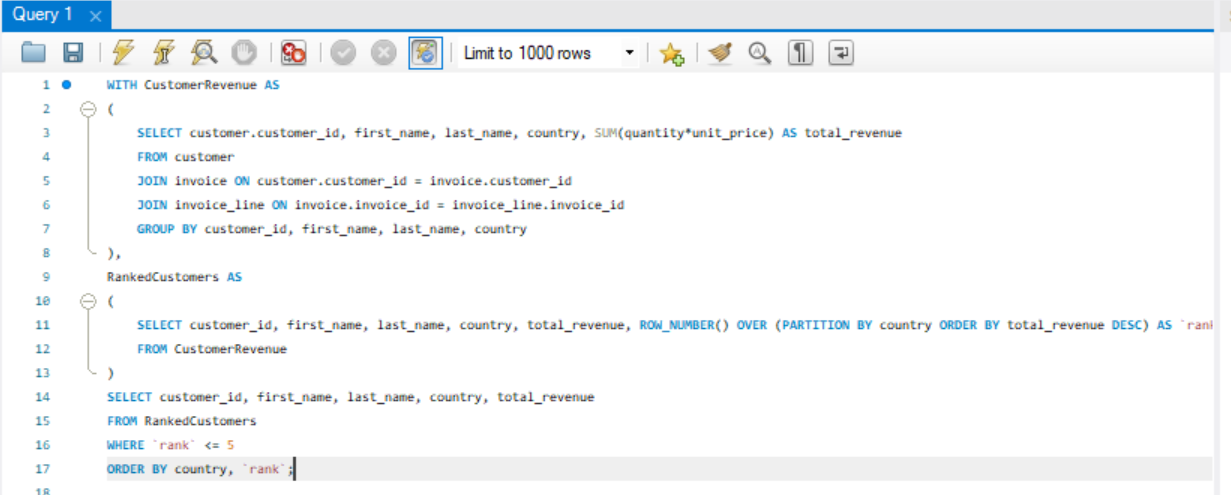


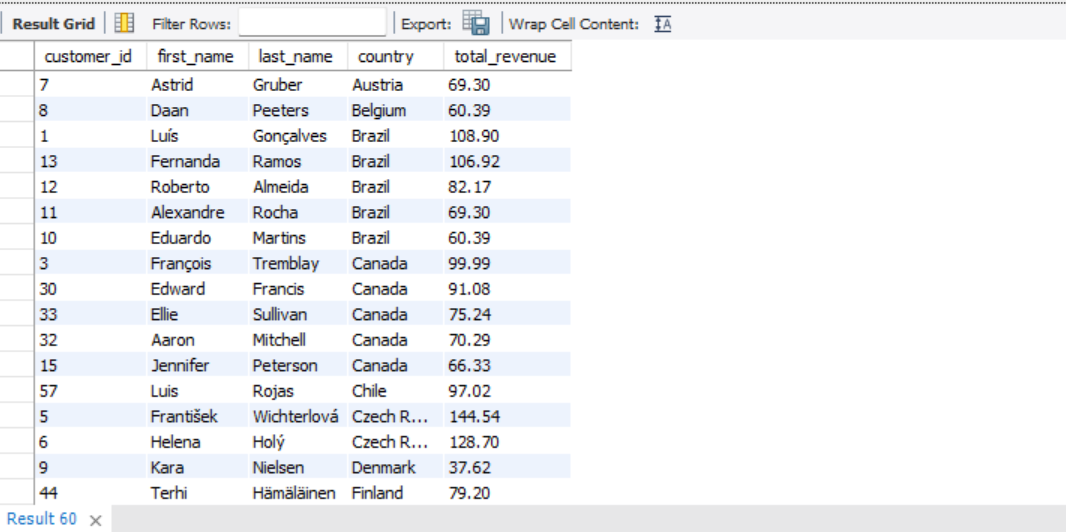
By city:



Approach used: Three queries apiece for country, state and city, each using SUM and COUNT aggregation functions for total revenue and invoice count, respectively.

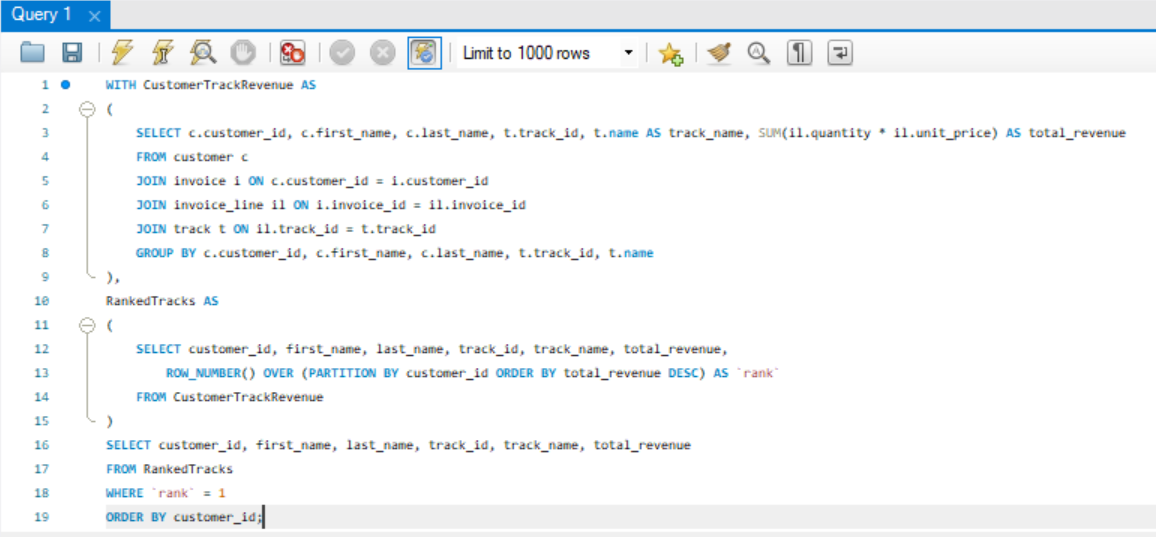
1. *Find the top 5 customers by total revenue in each country.*





Approach used: CTEs (one for all customer revenues, another for deriving the top 5 from these), Window function and SUM aggregation.

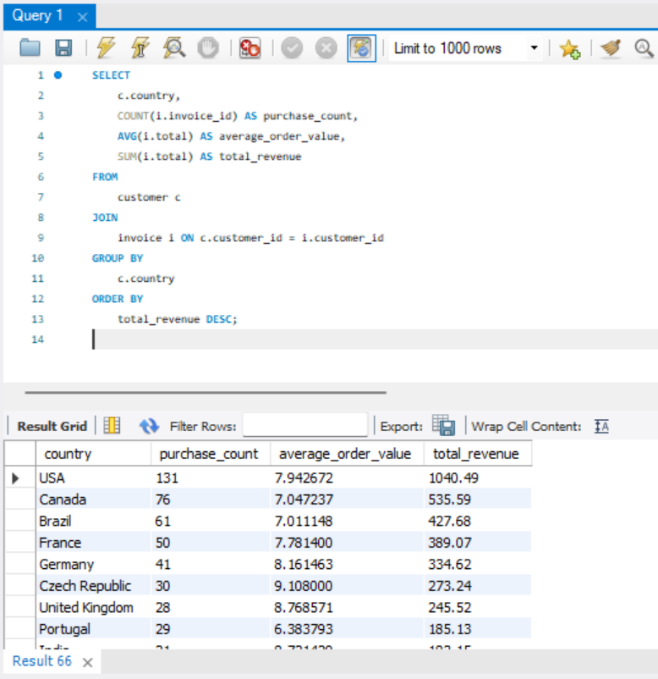
1. *Identify the top-selling track for each customer.*





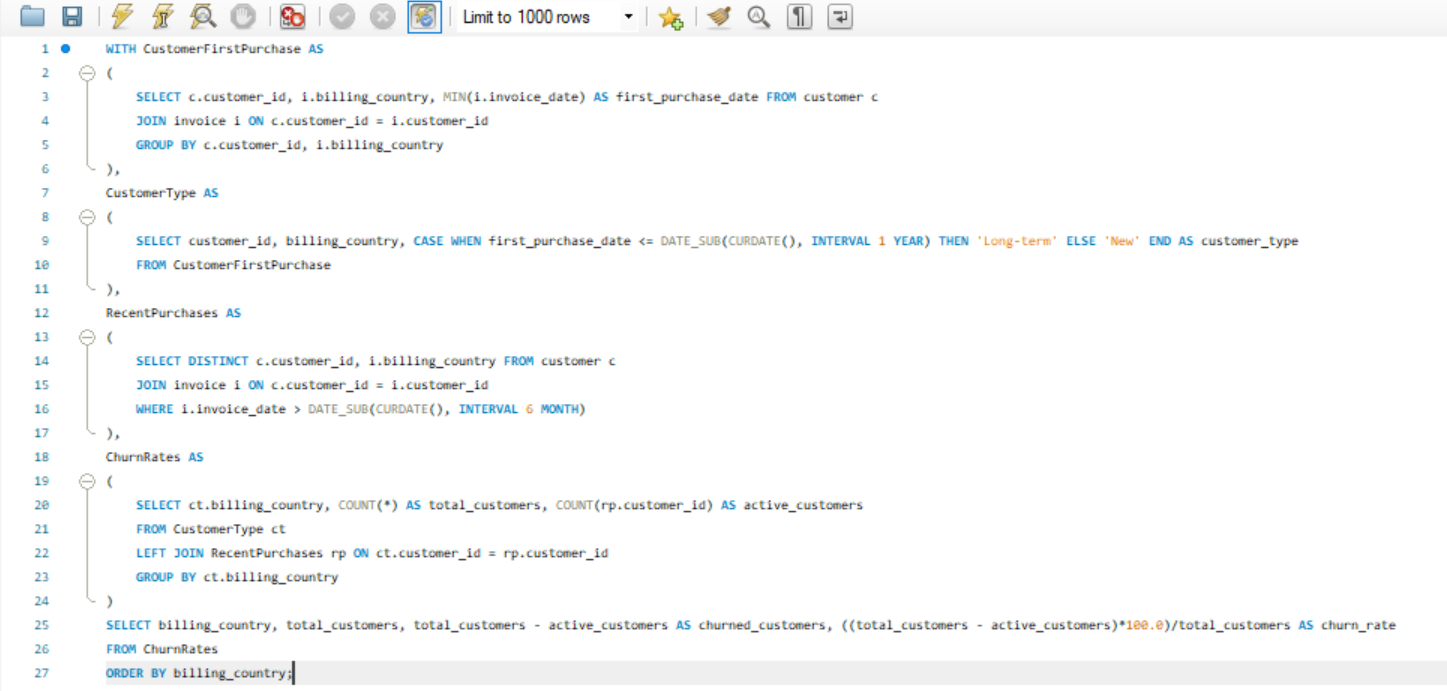
Approach used: CTEs (one for all customer tracks purchased and another for deriving the top-selling track from these), Window function and SUM aggregation.

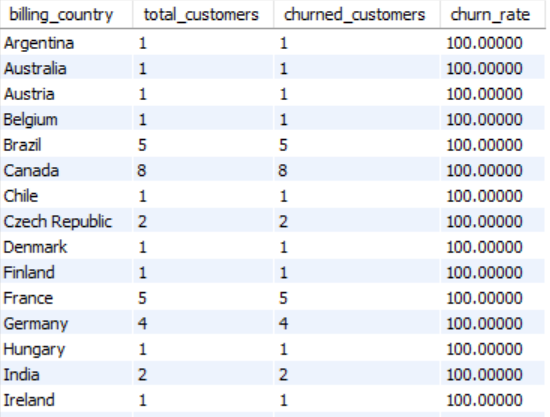
1. *Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?*

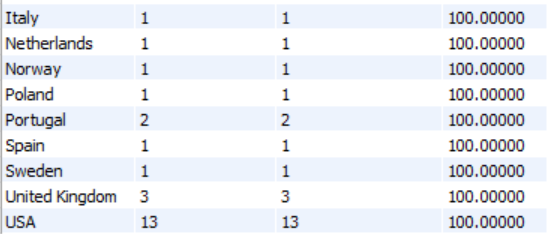


Approach used: COUNT, SUM and AVERAGE aggregation for separate metrics to gauge customer purchasing behavior.

1. *What is the customer churn rate?*

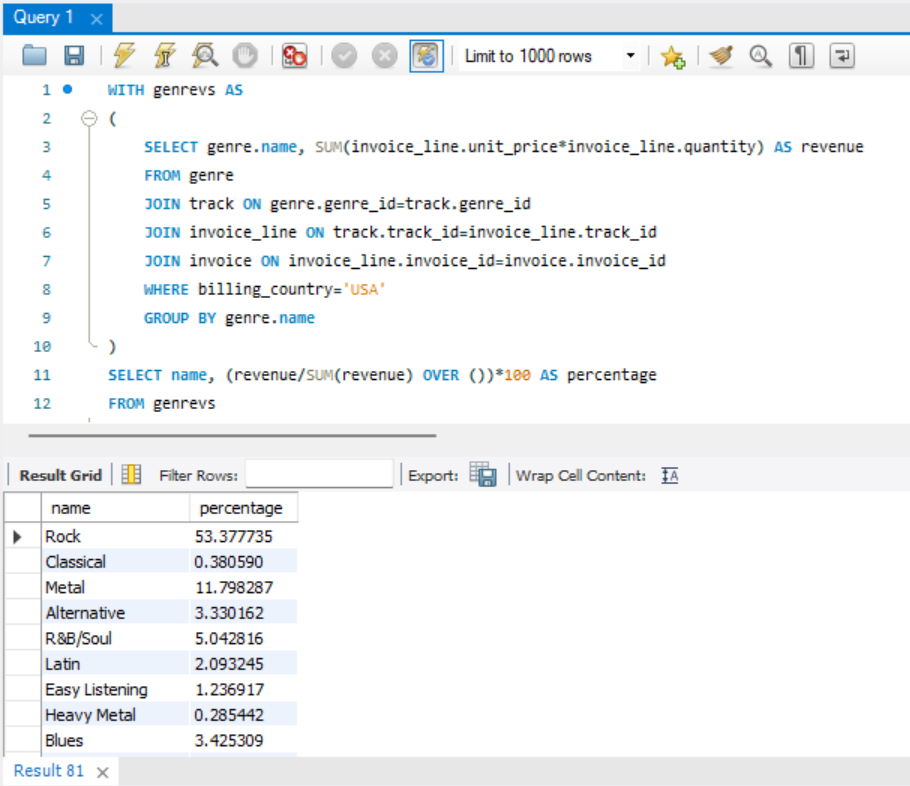


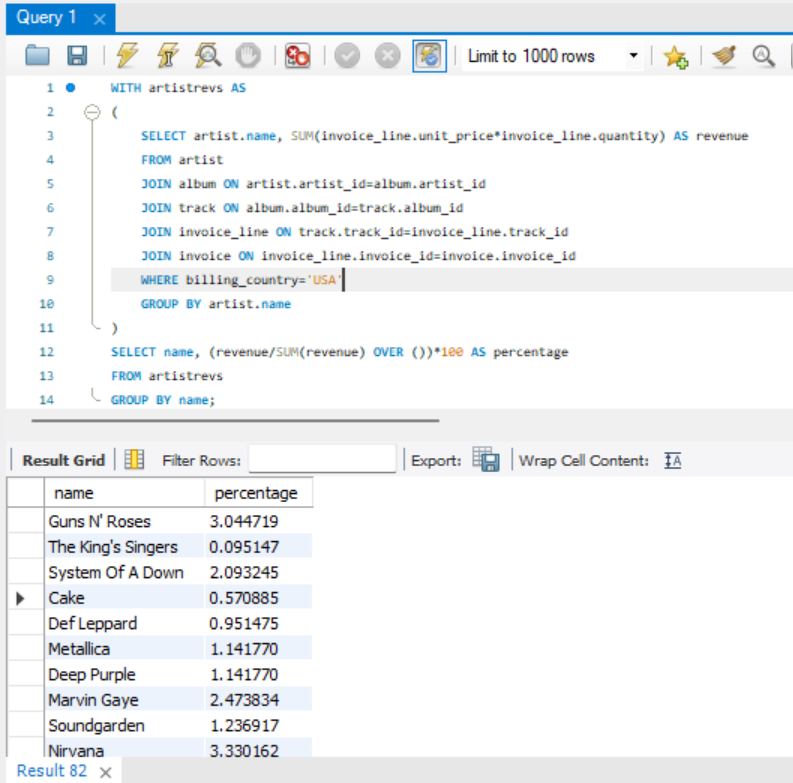




Approach used: CTEs (four: one to calculate customers’ first transaction, another to classify them as long or short-term based on this, a third to check for purchases within the last month and the last to calculate final churn rates), MIN and COUNT aggregation functions.

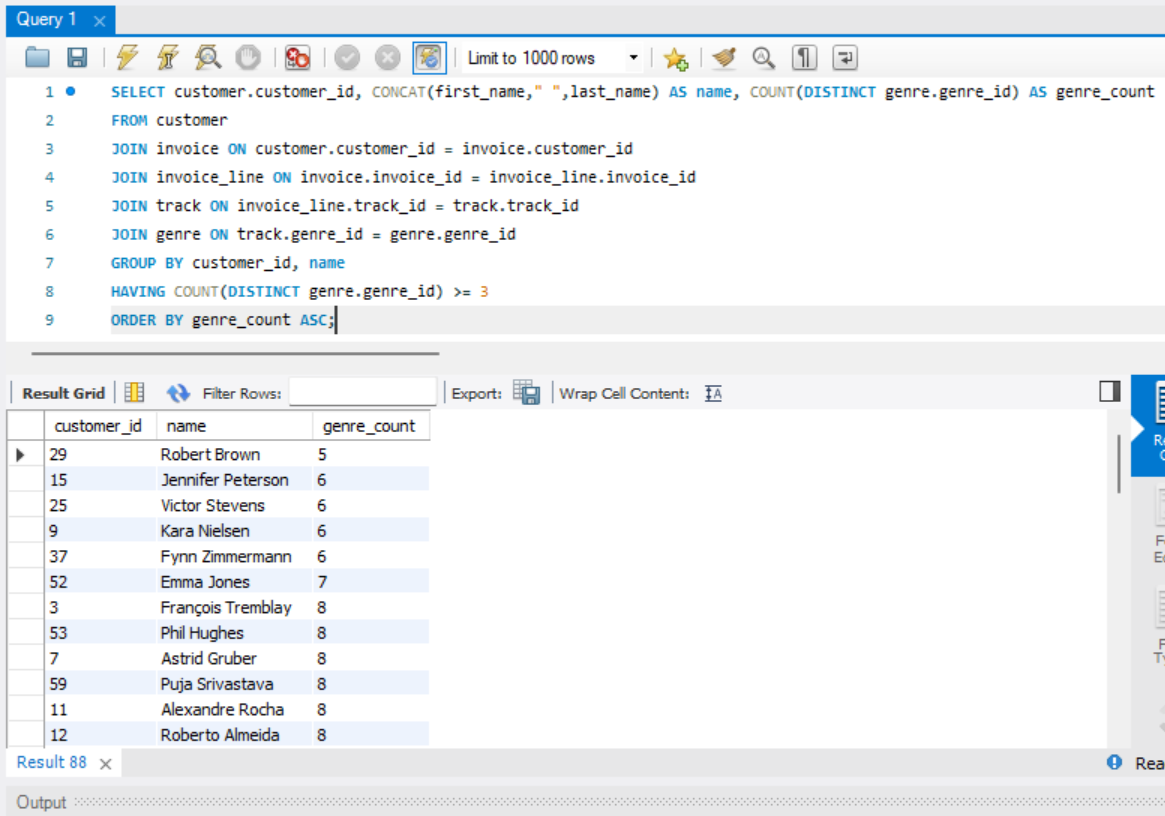
1. *Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.*





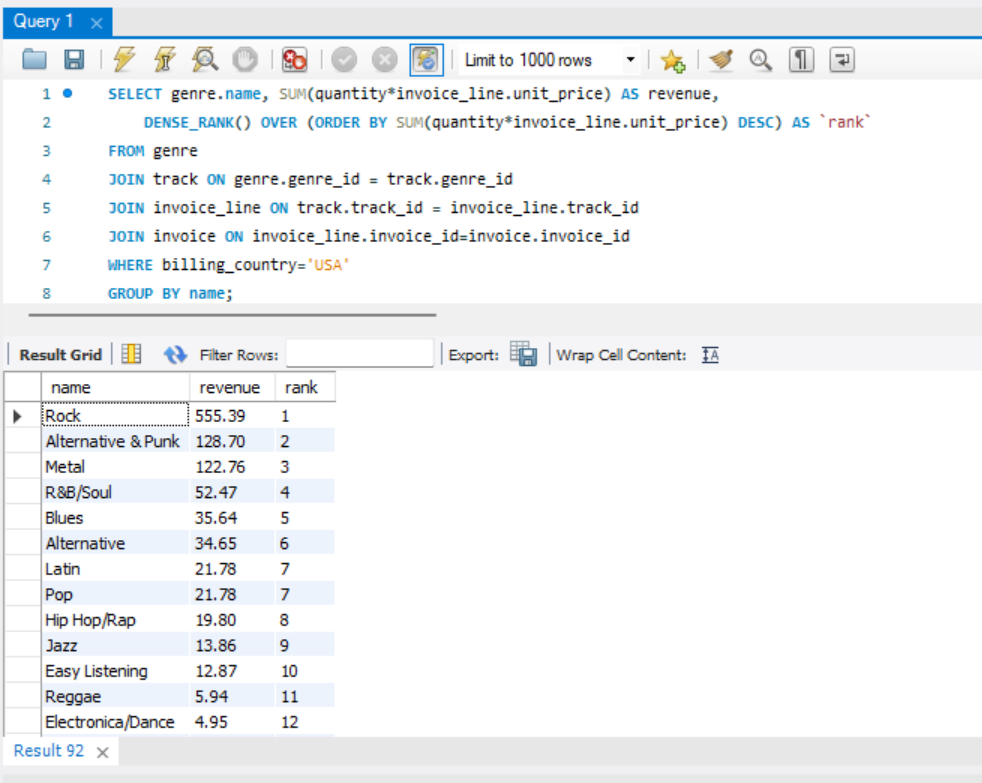
Approach used: 2 queries for genres and artists each, CTEs (to calculate revenue per genre/artist), followed by a customized function for percentage, SUM+Window function aggregation

1. *Find customers who have purchased tracks from at least 3 different+ genres.*

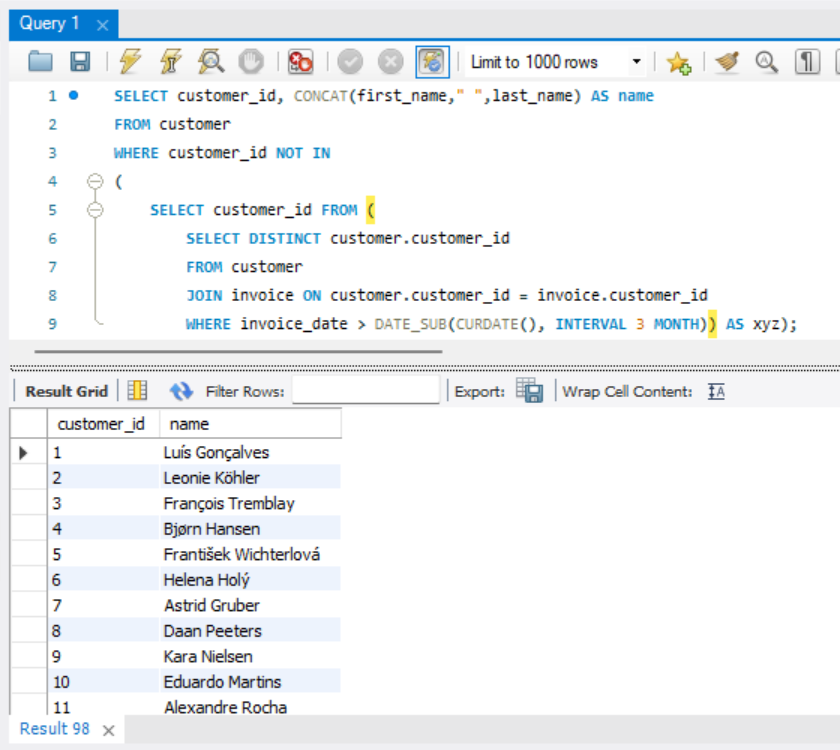


Approach used: COUNT aggregation function and use of the ‘DISTINCT’ keyword to ensure selection of 3+ different genres only.

1. *Rank genres based on their sales performance in the USA*

**

Approach used: SUM aggregation and Dense Rank Window Function with Joining of the relevant tables.

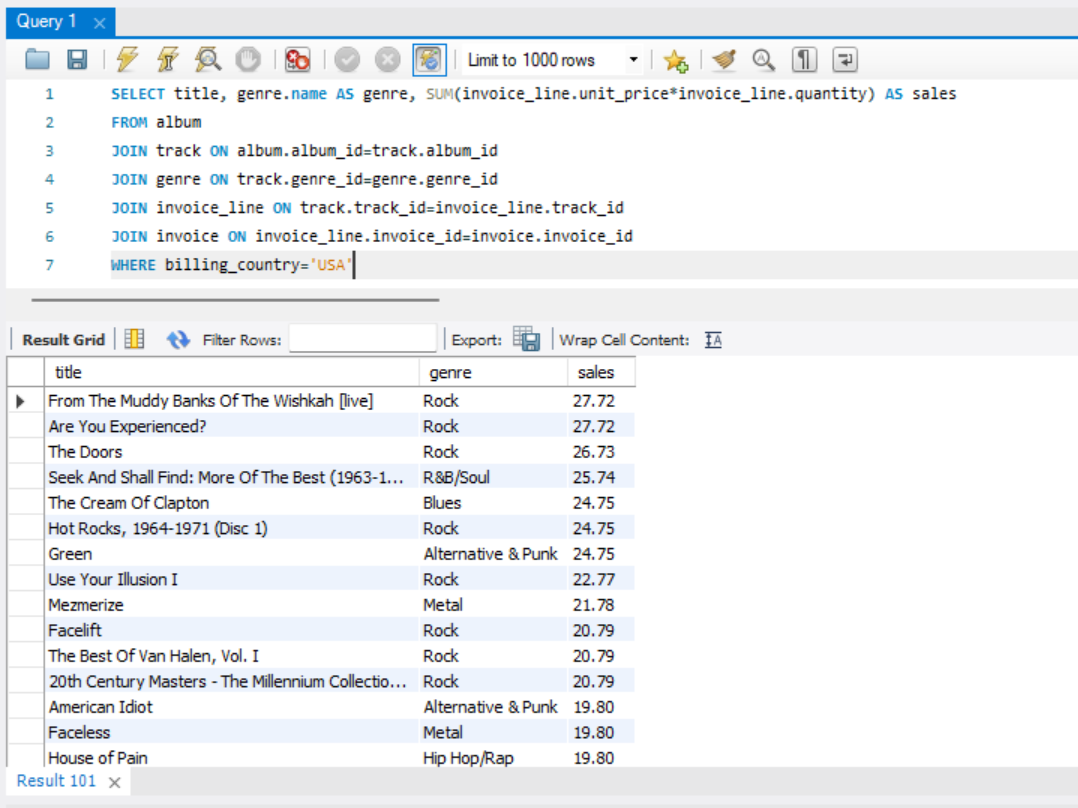
1. *Identify customers who have not made a purchase in the last 3 months*

Approach used: Subquery (to compile a list of those who have made purchases in the last 3 months), in conjunction with ‘NOT IN’ clause to isolate the remainder.

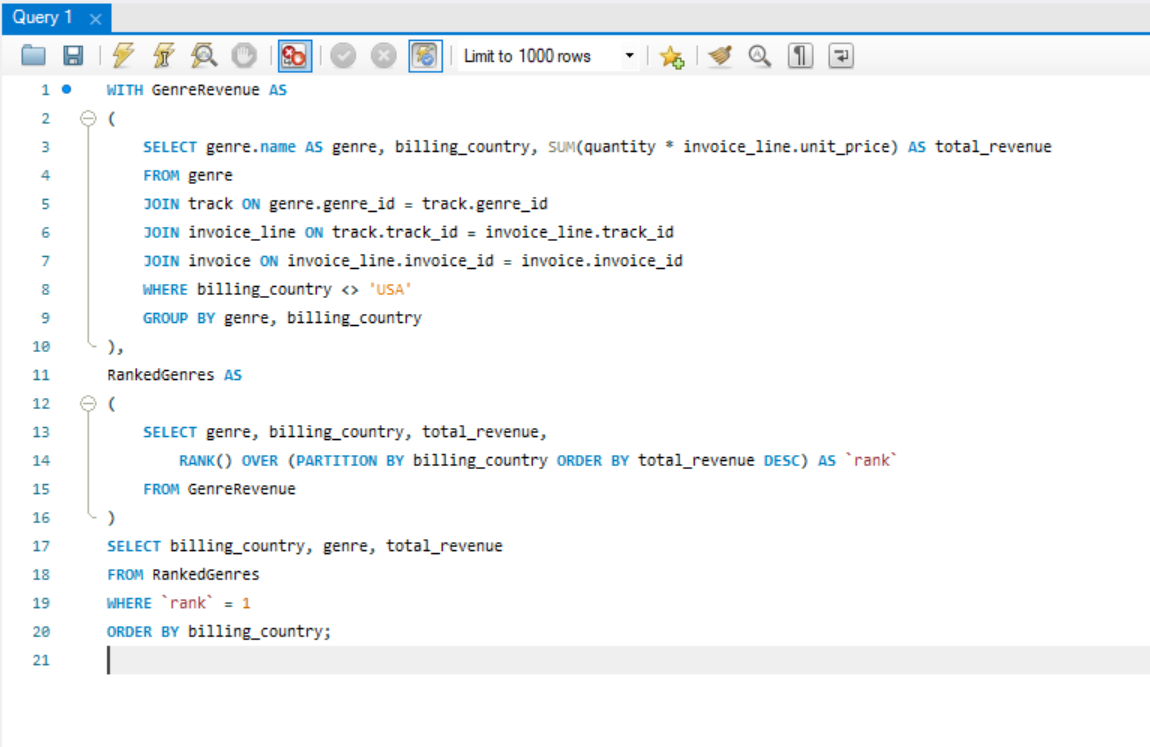
**SUBJECTIVE Q’S**

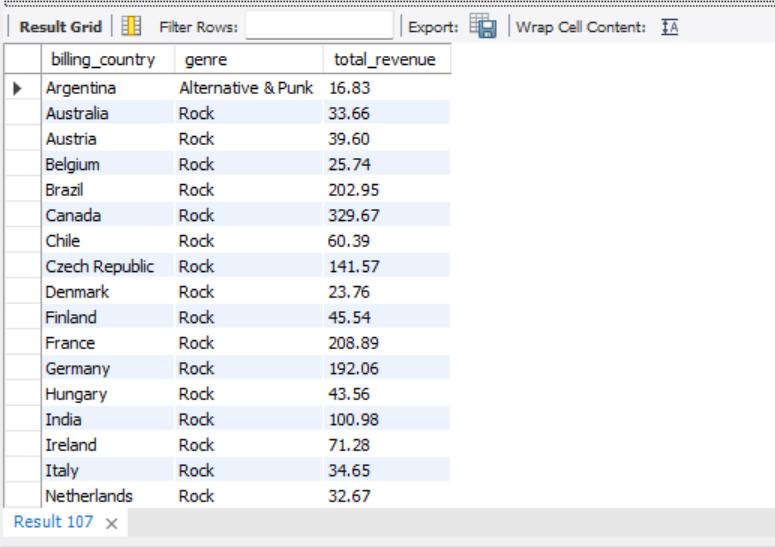
1. *Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.*

Below, we find that albums from the Rock genre dominate sales, while R&B/Soul, Blues and Alternative/Punk offer stiff competition. Albums from these latter three genres (Seek And Shall Find, The Cream Of Clapton, Green) thus seem fit for promotion, each showing potential to dethrone Rock:



1. *Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.*

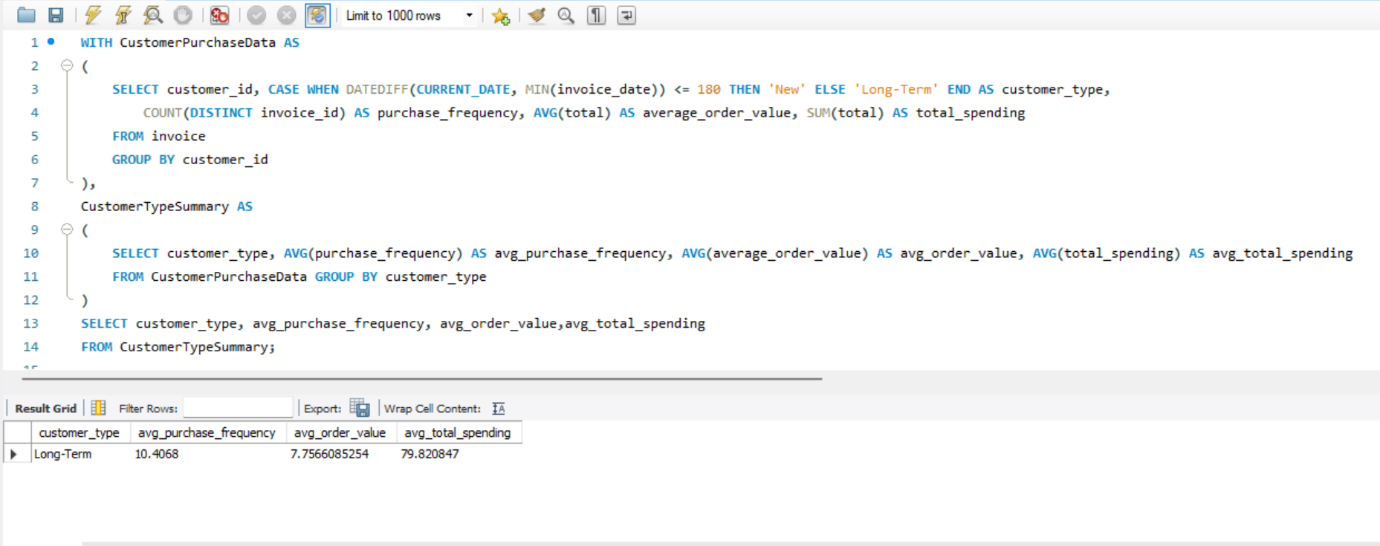




We find that Rock overwhelmingly outsells other genres in all non-USA countries except Argentina, where Alternative & Punk keeps its lead.

*3. Customer Purchasing Behavior Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?*

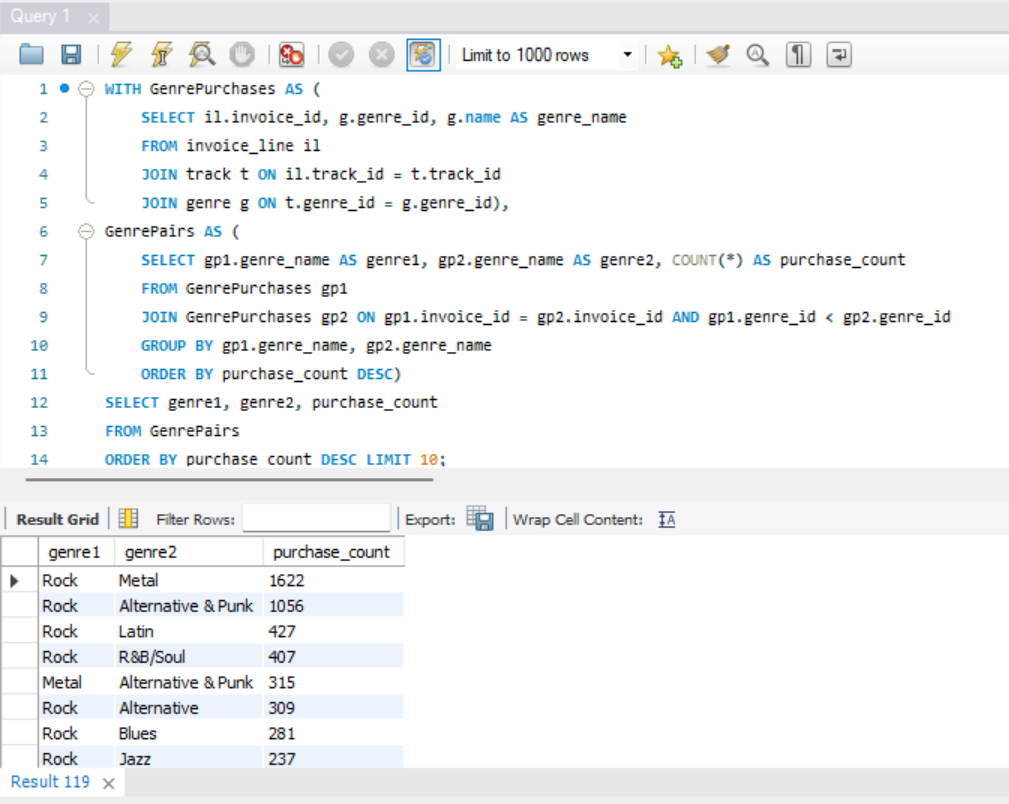
1. **Purchase Frequency:** Typically, long-term customers make purchases more frequently. New customers might have a lower purchase frequency as they are still exploring the value and offerings of the service.
2. **Basket Size (Average Order Value):** Long-Term Customers often have larger basket sizes as they tend to buy more products per transaction. New customers might start with smaller basket sizes and get acquainted with the product offerings.
3. **Total Spending Amount:** Cumulatively, long-term customers generally spend more over time. The total spending amount for new customers is likely lower initially but can increase as they become more engaged.



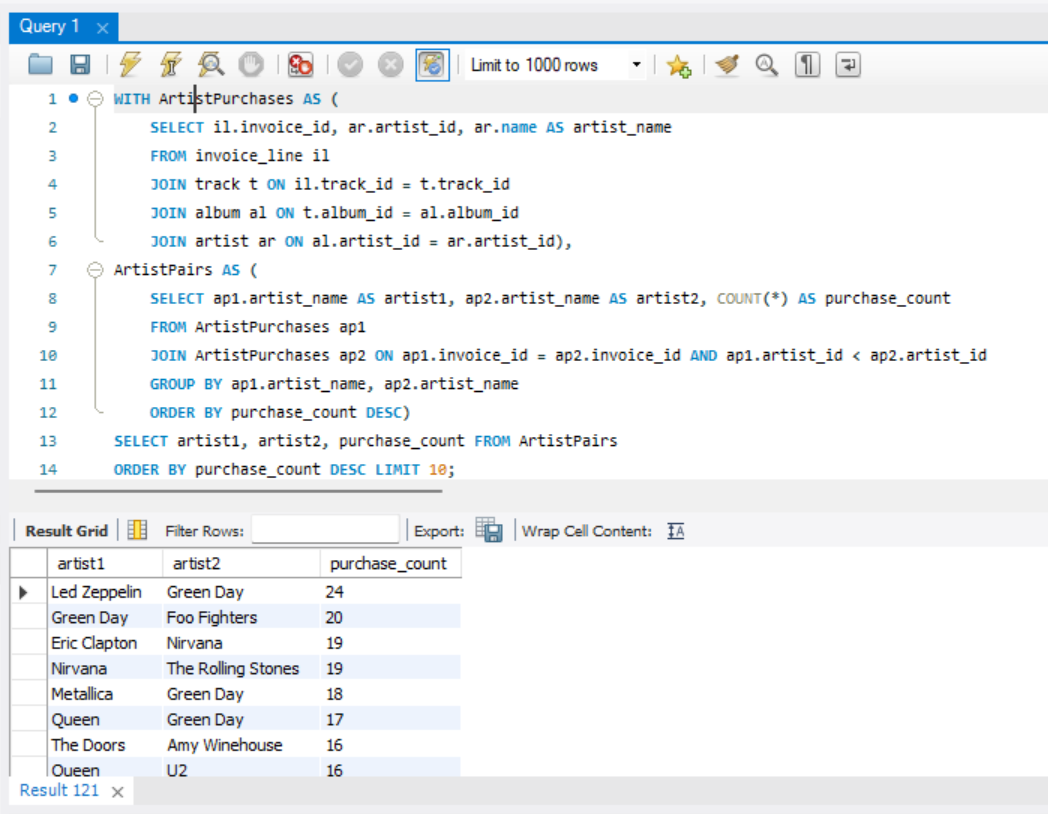
Insights/Strategies:

1. Long-term customers have higher purchase frequency and spending. **Strategy**: Implement loyalty programs that reward frequent purchases and higher spending.
2. New customers have lower engagement initially. **Strategy**: Use personalized marketing to engage new customers. This can include welcome emails, tailored product recommendations, and targeted promotions to encourage repeat purchases and higher basket sizes.
3. New customers need time to become familiar with the products. **Strategy**: Enhance onboarding process to educate new customers about the benefits of the products.
4. Long-term customers are valuable due to their consistent spending and loyalty. **Strategy**: Develop retention campaigns focused on long-term customers.
5. Different behaviors between new and long-term customers. **Strategy**: Segment customers based on their tenure and purchasing habits to tailor marketing strategies effectively. Use data analytics to predict which new customers are likely to become long-term.
6. *Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?*

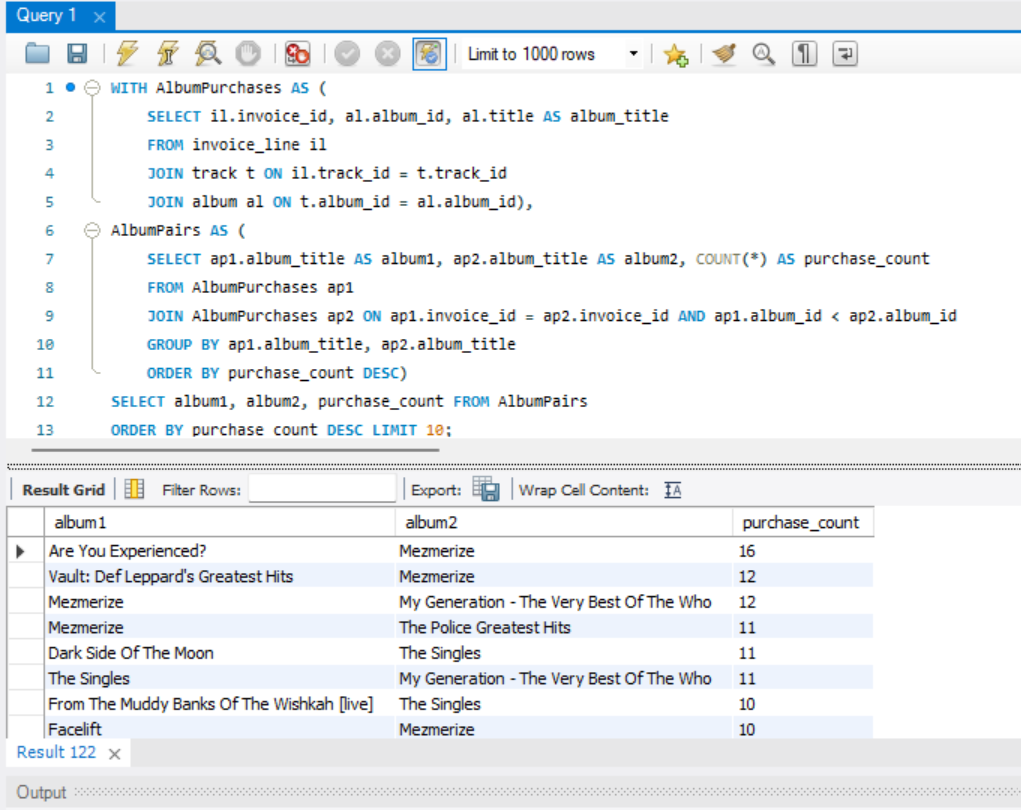
Genres:



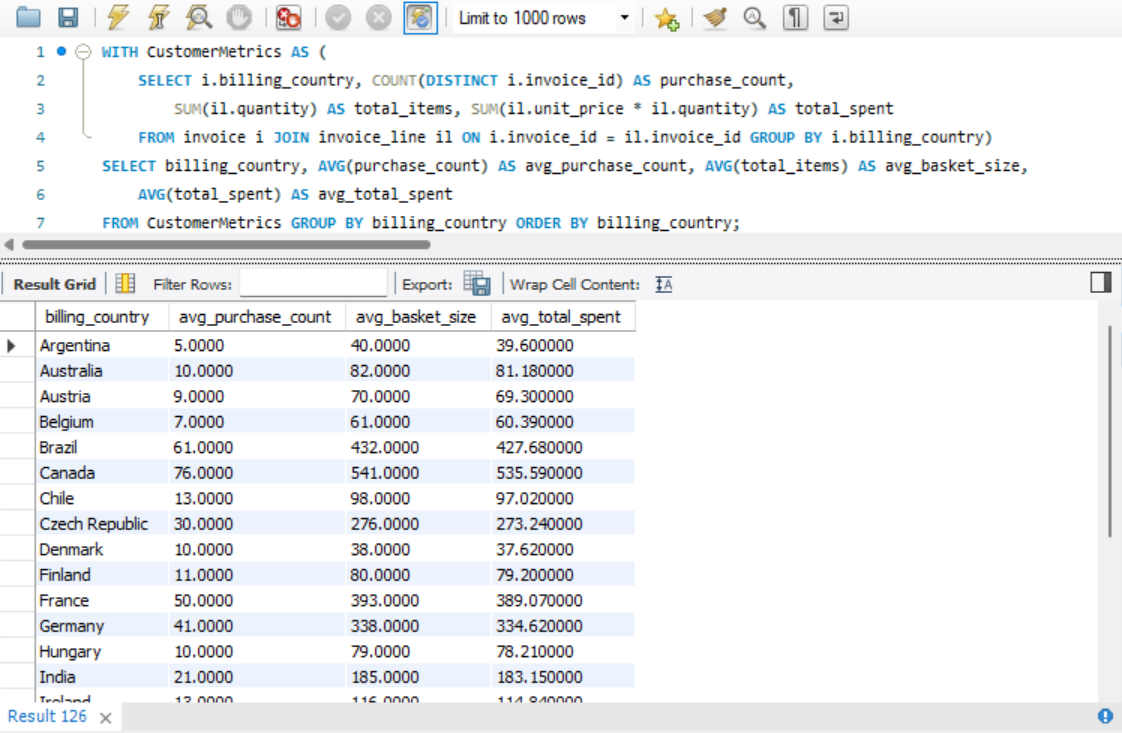
Artists:

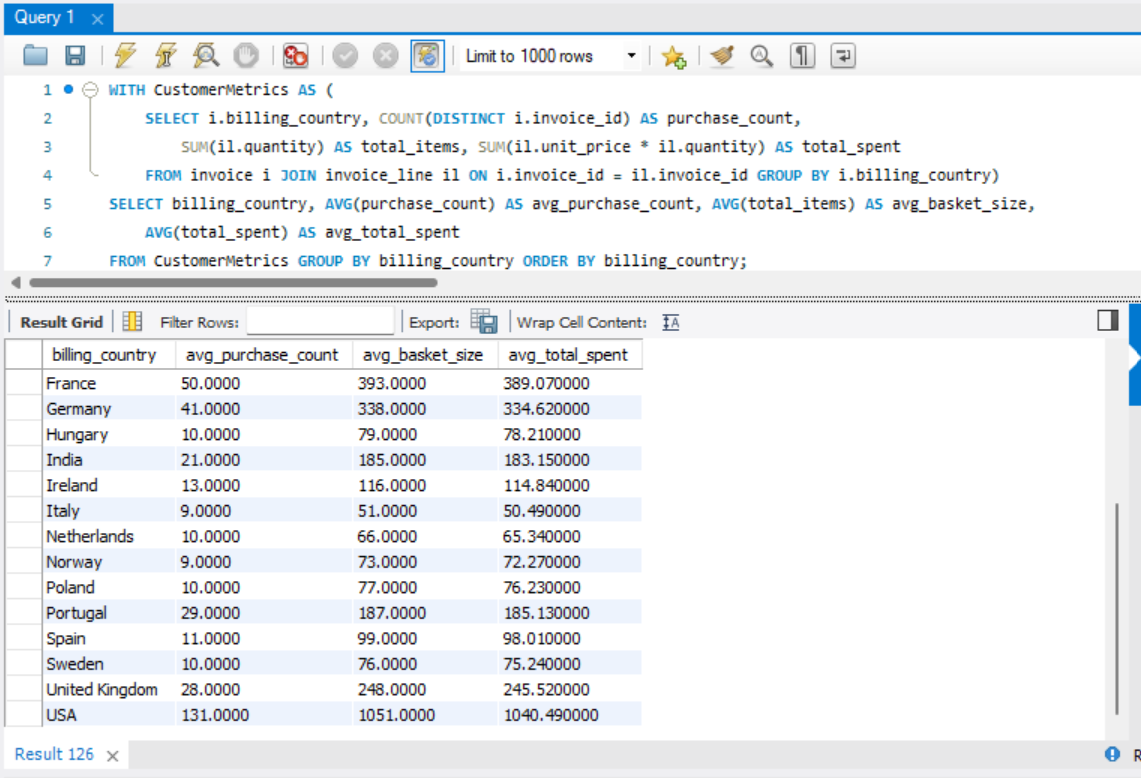


Albums:



Each of the above data could guide recommendation initiatives in sequential ways. Lucrative genre pairings could inform the choice of artists to populate a user’s recommendation section, and this in turn may suggest which of their albums the user should listen to first.

1. *Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?*



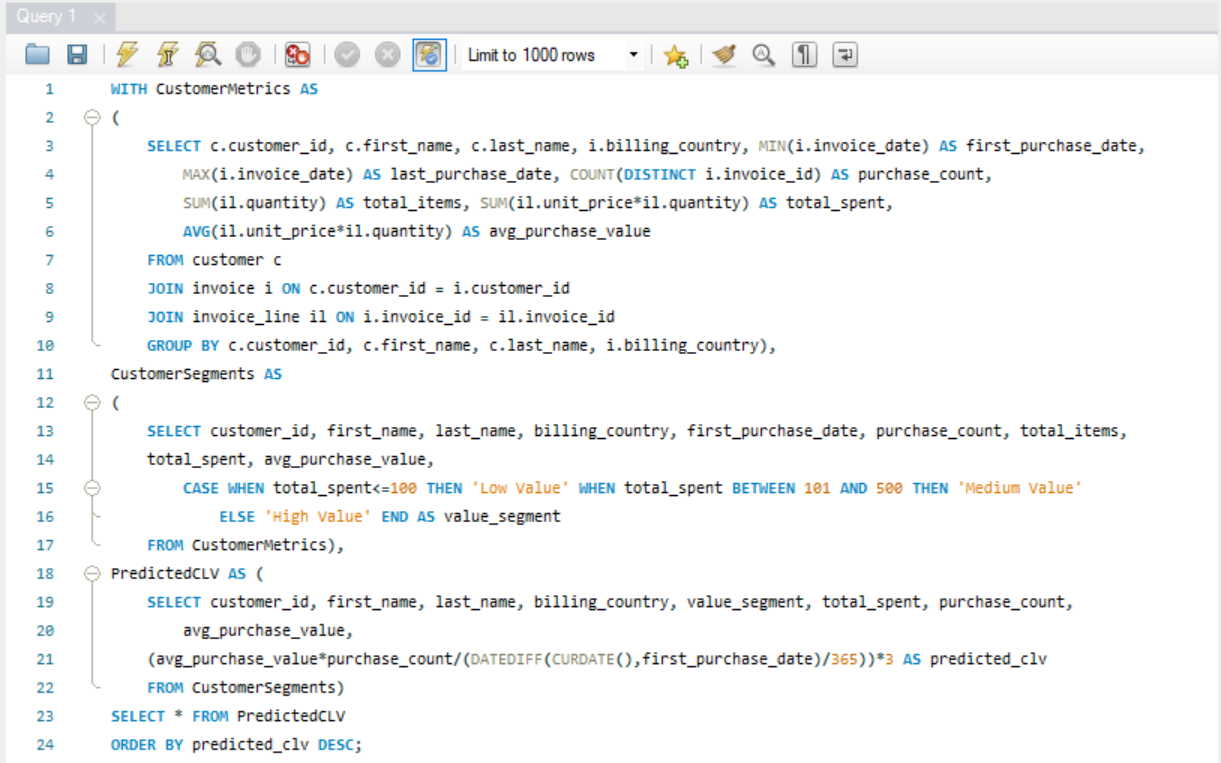
Thus, we find greatest activity among the major European and North American nations. Italy is an exception to this pattern, while India also stands out for its relatively significant purchasing and being the only Asian region represented here.

1. *Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?*
2. Customers with infrequent purchases are more likely to churn.
3. Customers with lower total spending might not see enough value in the service to remain loyal. Conversely, high spenders might churn if they find a better value elsewhere.
4. Some regions might have higher churn rates due to local competition, economic factors, or cultural preferences. Analyzing regional churn patterns can provide insights into these factors.
5. New customers are more likely to churn compared to long-term customers.
6. Customers who primarily purchase during promotions may churn when there are no discounts available.
7. Customers who have niche preferences may churn if their preferred products or genres are no longer available or updated.





1. *Customer Lifetime Value Modelling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns among customers who have stopped purchasing?*





1. *If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?*

* Gather data on the promotional activities, including the type of promotion, the duration, and the target audience. It is integrated with customer data, sales transactions, and engagement metrics.
* We next compare the number of new customers during the promotional period against a control period without promotions.
* For customer retention, analyze repeat purchase behavior by comparing the frequency and recency of purchases before, during, and after the promotional campaigns. This method can reveal whether customers acquired during promotions tend to remain active longer/return more frequently compared to those acquired outside of promotional periods.
* Overall sales impact can be measured by evaluating total revenue and average order value during the campaign periods relative to non-campaign periods.
* Assessing the effectiveness of different types of promotions can provide deeper insights.

1. *How would you approach this problem, if the objective and subjective questions weren't given?*

Without specific questions provided, we would:

* Determine the goals of the analysis, e.g., is the company looking to increase sales/reduce churn/improve customer satisfaction/target marketing efforts etc.
* Understand what metrics are most important to the business, such as total revenue, average order value, purchase frequency, customer lifetime value, etc.
* Analyze customers by their location, purchasing patterns and trends over time.
* Identify which countries generate the most revenue.
* Determine which countries have the highest/lowest average order values.
* Find out if there are specific countries with higher customer acquisition or retention rates.
* Look for correlations between customer demographics (if available) and purchasing behavior.

1. *How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?*

We use the following query in order to accomplish this operation:

ALTER TABLE Albums ADD ReleaseYear INTEGER;

1. *Chinook is interested in understanding the purchasing behavior of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write a SQL query to provide this information.*



