**Objective Questions**

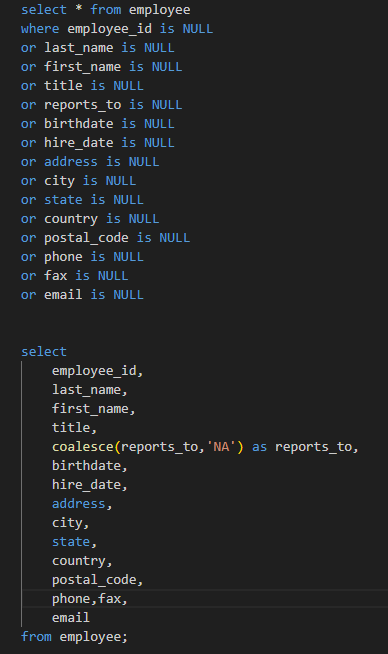
1. Does any table have missing values or duplicates? If yes how would you handle it ?

**Approach -** To check if table contain missing values or duplicates we use the IS NULL function.

**Steps -**

**EMPLOYEE\_TABLE**

1. Check for null values in Employees table.
2. Output shows that column reports\_to contains NULL values
3. Handle the NULL or missing values using coalesce values and replace it with NA

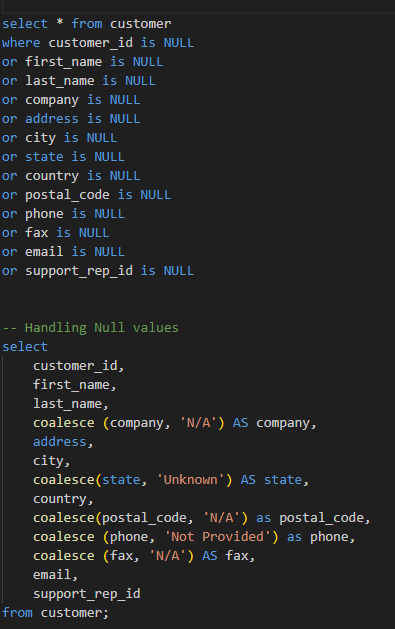
****

**OUTPUT**

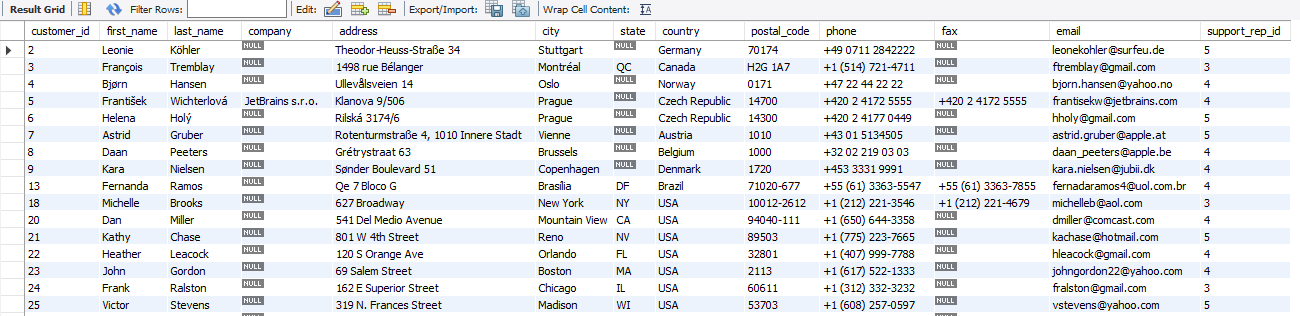
****

**CUSTOMER\_TABLE**

1. Check for NULL values in customer table using is NULL .
2. Output shows company, state, postal\_code, phone and fax contain NULL values
3. Handle the NULL or missing values using coalesce and replace it with ‘NA’

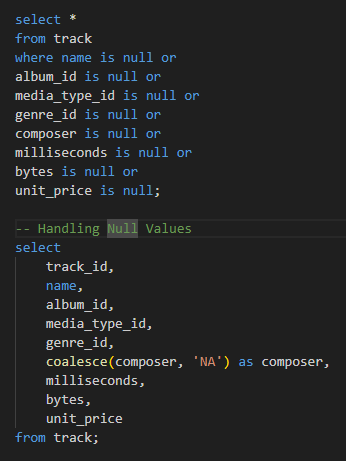


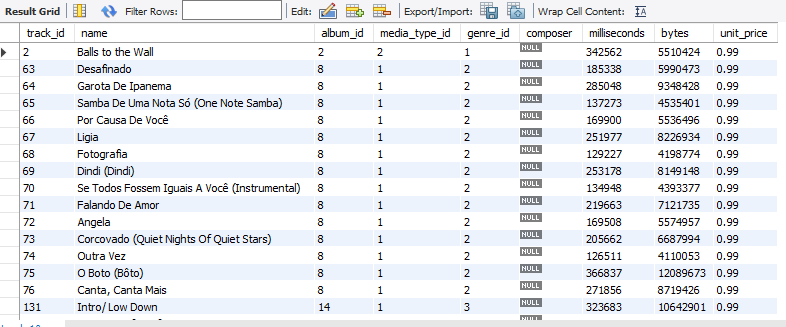
**OUTPUT**

****

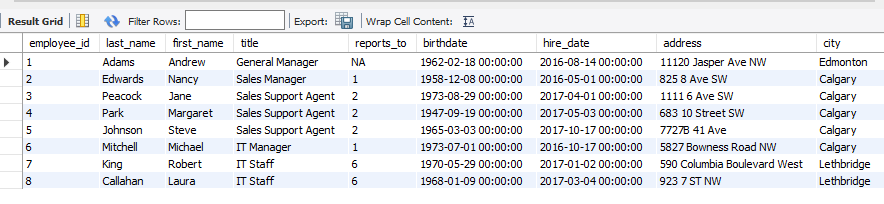
**TRACK\_TABLE**

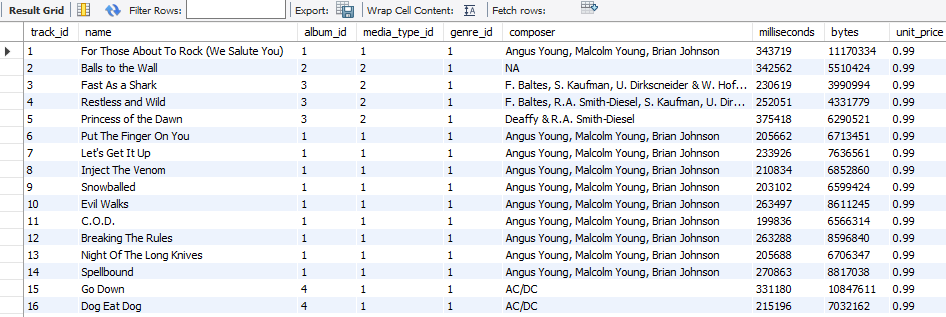
1. Check for missing values in track table using is NULL .
2. Output shows composer contain NULL values
3. Handle the NULL or missing values using coalesce and replace it with ‘NA’

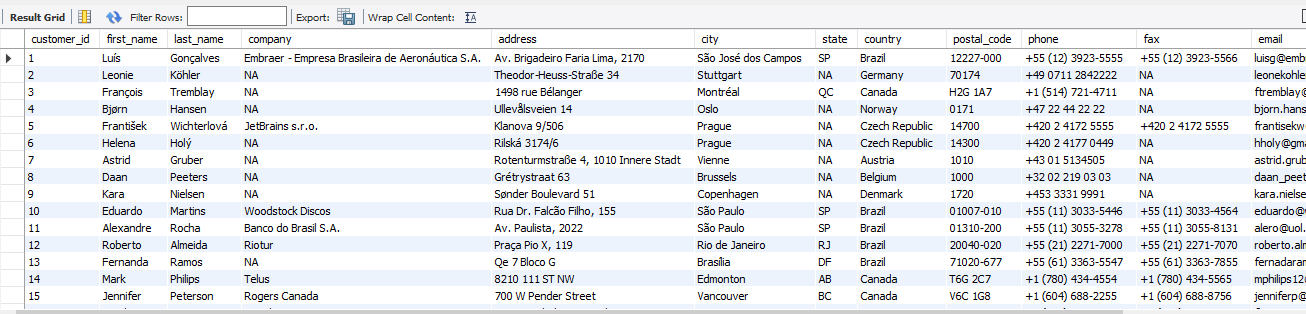




**RESULT-**

1. All other tables were reviewed, and no missing values or duplicates were found.
2. The COALESCE function successfully replaced null values in each table with appropriate default values, preserving data usability without modifying the original records.
3. No duplicate rows were found across primary key columns in any table.



****

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

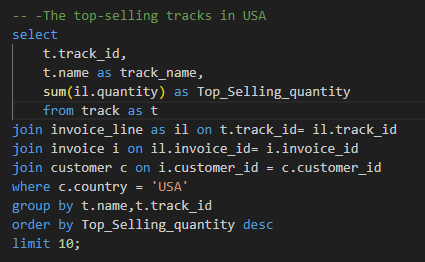
**Approach** - Applying joins, filtering, and aggregation techniques to identify the top-selling tracks, most popular artist, and their most successful genres specifically in the USA.

**Steps -**

* Top-Selling Tracks: Joined track, invoice\_line, invoice, and customer filtered for USA; grouped by track; listed top 10 by quantity sold.
* Top Artist: Joined track, album, artist, invoice\_line, invoice, and customer; filtered for USA; summed sales per artist to find the top seller.
* Popular Genres: Identified top artist via subquery; joined genre, track, album, and invoice\_line; grouped by genre;

**RESULT -**

* Top-Selling Tracks: Found the 15 most popular tracks in the USA based on sales, reflecting customer preferences.
* Top Artist: Identified Van Halen as the top-selling artist in the USA with 43 total sales, indicating high market demand.
* Popular Genres: Listed **Rock** as the most popular genre associated with Van Halen, with 43 total sales, revealing the favored music style among their fans.

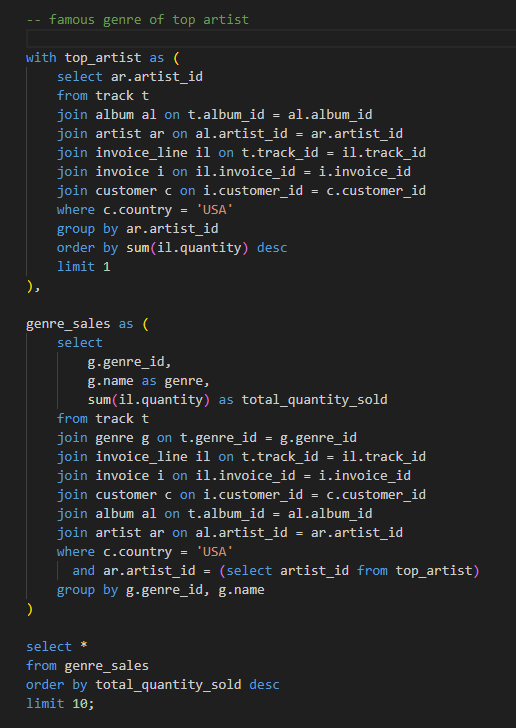


**OUTPUT**

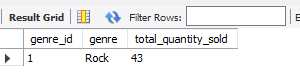
****

**A computer screen shot of a program

Description automatically generated**



**OUTPUT**

****

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

**APPROACH** - To find customer demographic breakdown we use customer table and aggregate on country, city and state and handle the NULL values .

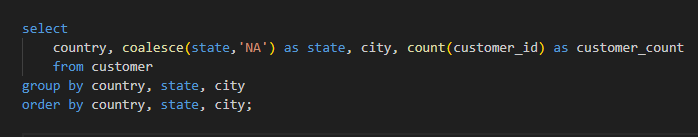
STEPS-

1. Selected country, state, and city from the customer table and counted customers (customer\_id) in each segment.
2. Applied COALESCE to replace nulls in the state column with "Not Available" for cleaner output.
3. Grouped and ordered the results by country, state, and city to structure the demographic distribution clearly.

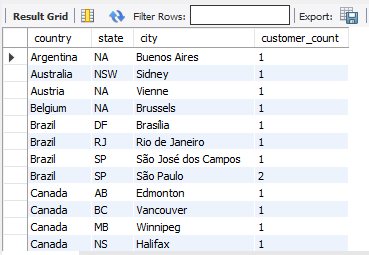
**RESULT** -

Provided a detailed view of customer distribution by location, supporting targeted marketing and engagement strategies.

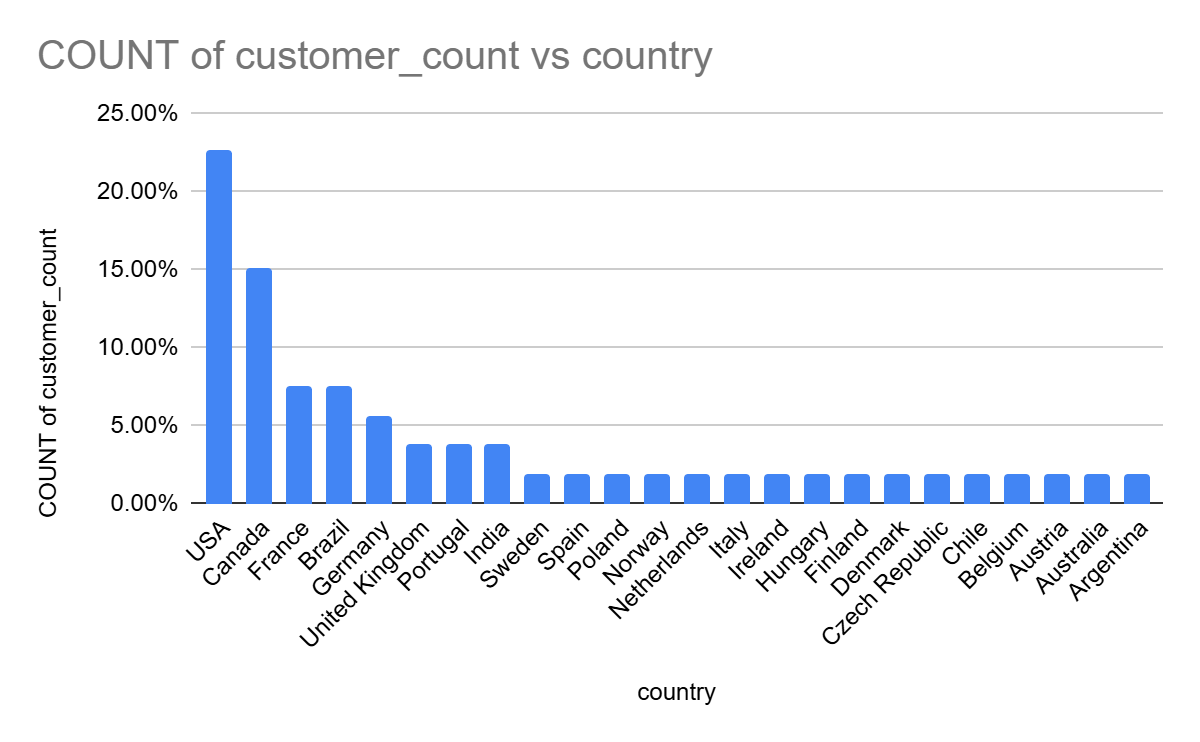
**CODE** -



**OUTPUT** -



**VISUALIZATION -**

****

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

**APPROACH** -

Used the invoice table to aggregate sales data by geographic location (country, state, city).

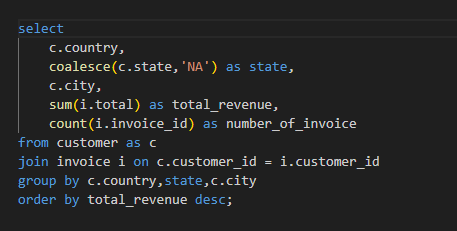
**STEPS** -

1. Selected country, state, and city from the customer table joined on customer table.
2. Replaced null state values with NA using coalesce for clarity.
3. Grouped by country, state, and city, then calculated count of invoices and sum of revenue

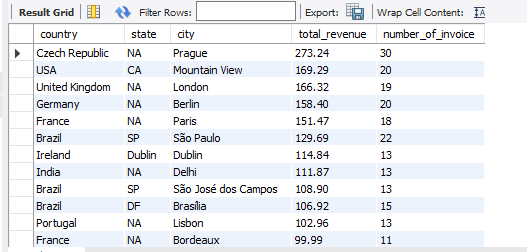
**RESULT** -

1. Output provides total revenue and number of invoices for each location
2. Useful for identifying high-performing markets and areas needing strategic attention

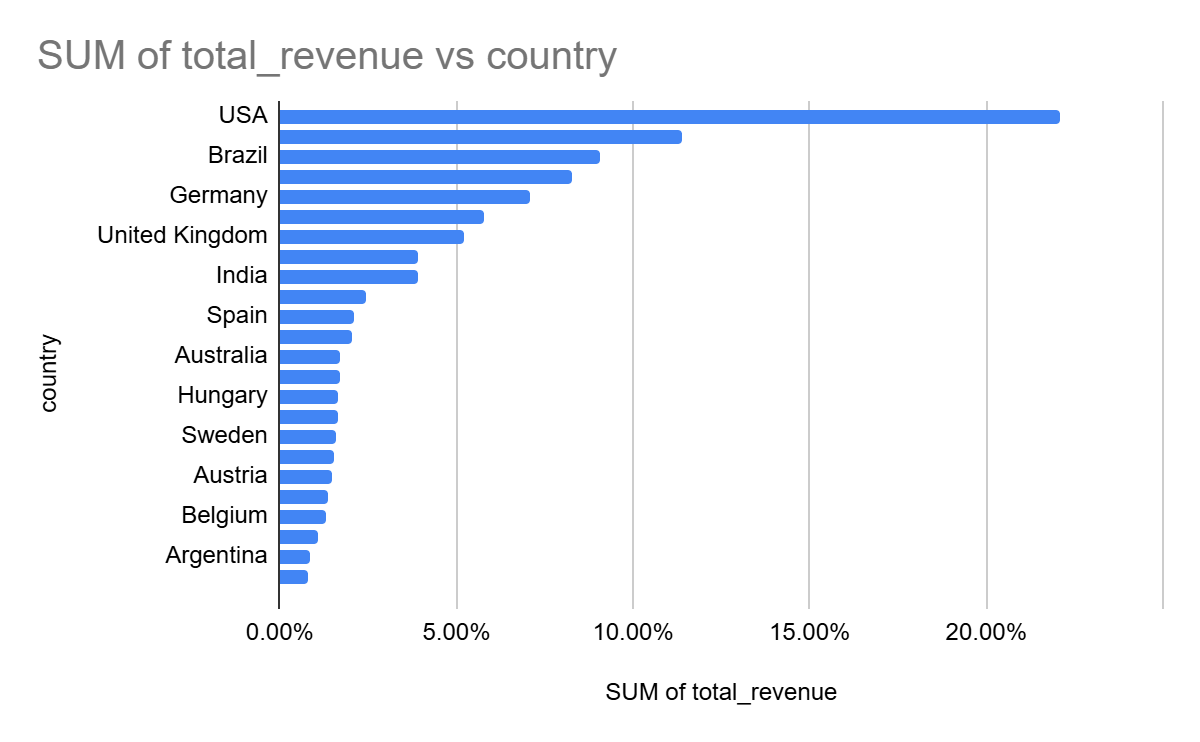
**CODE** -



**OUTPUT** -



**VISUALIZATION**

****

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

**APPROACH** -

1. This query uses the customer and invoice tables.
2. It calculates total revenue per customer and identifies the top 5 customers in each country based on that revenue.

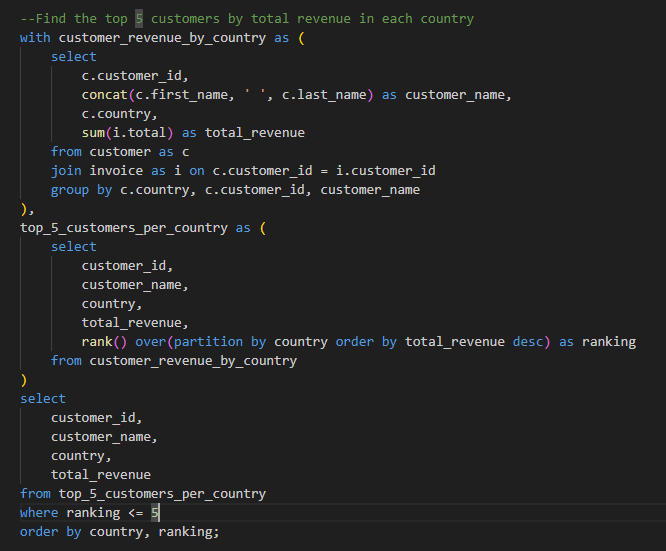
**STEPS** -

1. In the first CTE (customer\_revenue\_by\_country), we calculate total revenue generated by each customer using a join between customer and invoice.
2. We group this data by customer and country.
3. In the second CTE (top\_5\_customers\_per\_country), we rank customers within each country based on their revenue using the rank() window function.
4. Finally, we filter and select only the top 5 customers per country.

**RESULT** -

1. The result shows the top 5 revenue-generating customers in each country, providing a view of regional sales performance.
2. It helps businesses identify key customers per market for targeted engagement.

**CODE** -



**OUTPUT** -



1. Identify the top-selling track for each customer

**APPROACH** -

1. This query uses the customer, invoice, invoice\_line, and track tables.
2. It calculates the total quantity of tracks purchased by each customer and identifies their most frequently purchased track.

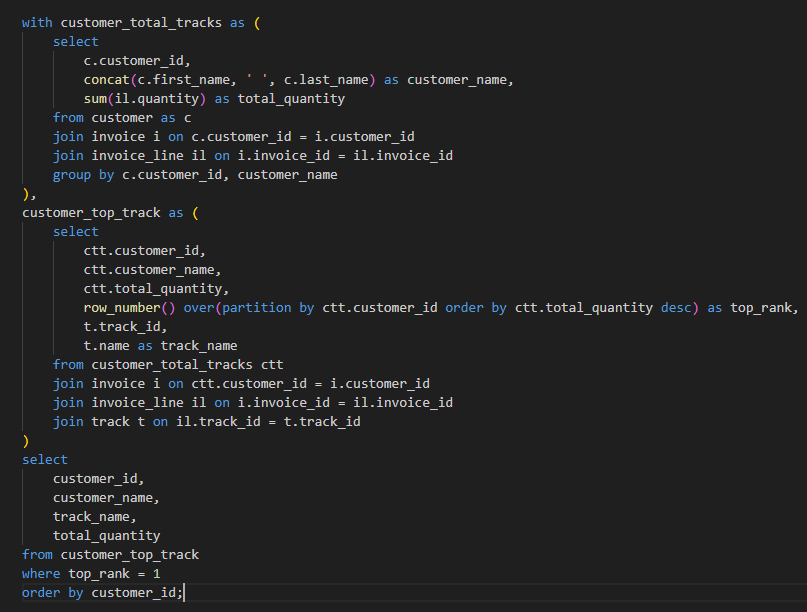
**STEPS** -

1. In the first CTE (customer\_total\_tracks), we calculate the total number of tracks purchased by each customer using joins and aggregation.
2. In the second CTE (customer\_top\_track), we link each customer's invoices to tracks and rank the most frequently purchased track.
3. We use row\_number to determine the top track per customer.
4. The final selection filters only the top-ranked track per customer.

**RESULT** -

1. This identifies each customer's favorite or most purchased track, which can help personalize recommendations.
2. Businesses can use this to analyze track popularity among top customers.

**CODE** -



**OUTPUT** -



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

**APPROACH** -

1. The analysis uses the customer and invoice tables.
2. It evaluates purchase frequency, average order value, and total revenue per customer.

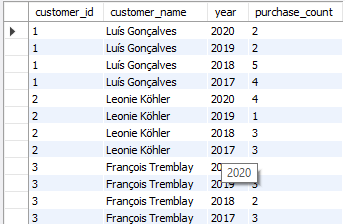
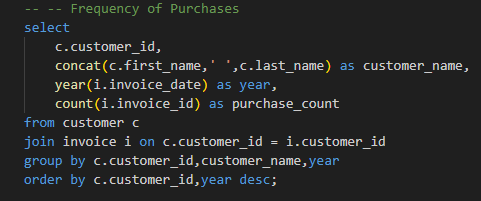
**STEPS** -

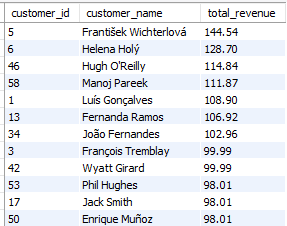
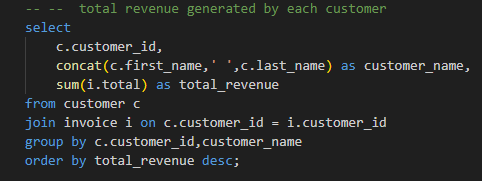
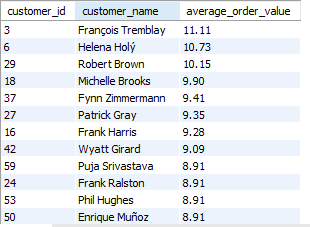
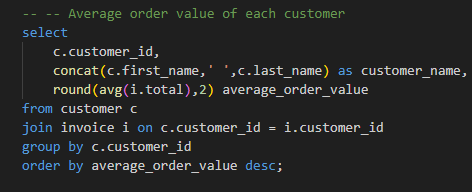
1. Frequency of Purchases: Count the number of invoices per customer, grouped by year.
2. Average Order Value: Calculate the average invoice total per customer.
3. Total Revenue: Sum invoice totals per customer to assess total contribution.

**RESULT** -

1. Some customers make regular purchases across years, indicating consistent engagement.
2. A few customers generate a large share of revenue through frequent purchases or high order values.

**CODE** -





**OUTPUT** -

1. What is the customer churn rate?

**APPROACH** -

1. Calculate the number of customers who have not made a purchase within the last year.
2. Divide the number of inactive customers by the total number of customers to calculate the churn rate.

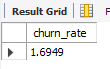
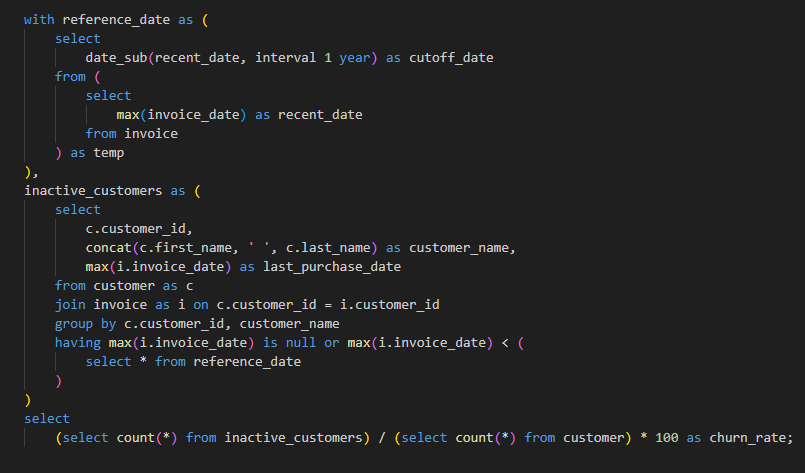
**STEPS** -

1. Determine the **cutoff date** (1 year before the most recent invoice date).
2. Identify **inactive customers**, who have either never made a purchase or have made their last purchase before the cutoff date.
3. Calculate the total number of **inactive customers**.
4. Compute the **churn rate** by dividing inactive customers by the total customer base.

**RESULT** -

1. Churn rate is an important metric for understanding customer retention and business sustainability.
2. A high churn rate indicates a potential issue with customer satisfaction or service quality.
3. The calculated churn rate is **1.69%**, indicating that a small proportion of customers have ceased making purchases over the past year

**CODE** -



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

**APPROACH** -

1. Calculate the total sales for each genre within a specific country (USA) and determine the percentage contribution of each genre to the overall sales.
2. Identify the best-selling genres and their corresponding artists based on sales, ranking the artists within each genre.

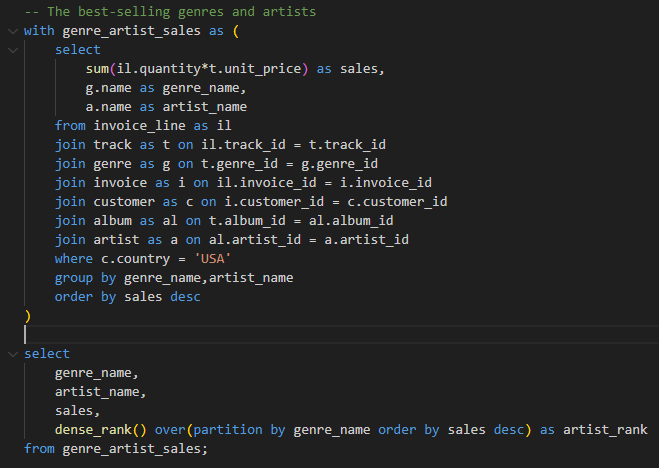
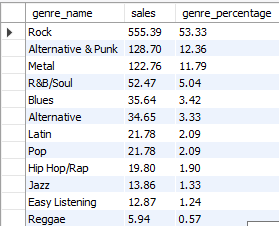
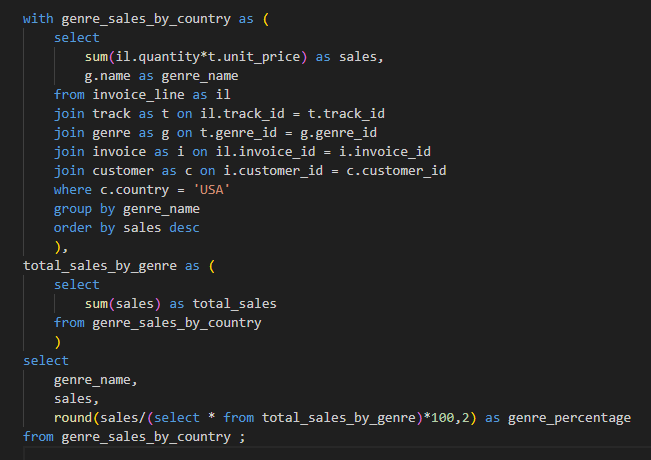
**STEPS** -

1. **Calculate Genre Sales**: Sum the sales for each genre based on the quantity sold and the unit price of tracks in the USA.
2. **Calculate Total Sales**: Calculate the total sales for all genres in the USA.
3. **Compute Genre Contribution Percentage**: Calculate the percentage contribution of each genre to the total sales.
4. **Rank Artists**: Calculate the total sales for each artist in each genre and rank them based on their sales within each genre.

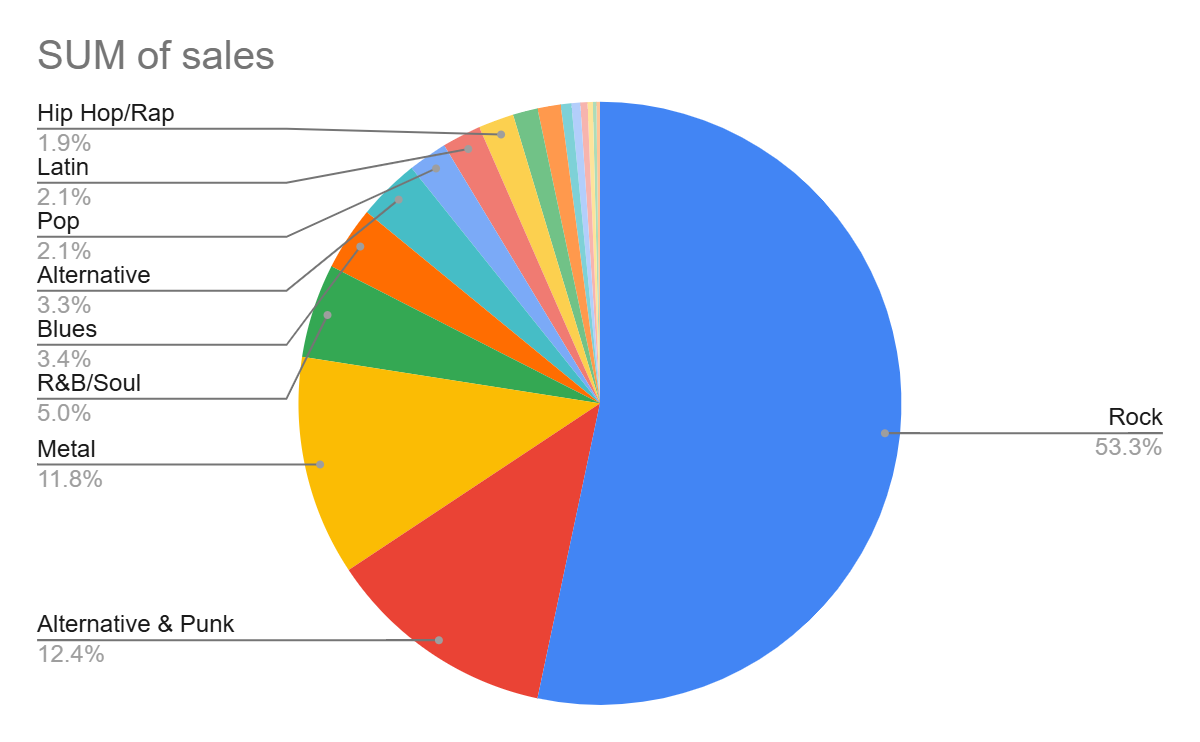
**RESULT** -

1. The genre-wise contribution analysis helps understand which genres are more popular in specific markets.
2. Ranking the artists within genres provides insights into which artists are driving the most sales in each genre.

**CODE** -



**VISUALIZATION** -



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

**APPROACH -** To find customers who have purchased tracks from at least 3 different genres, we need to fetch invoice , track, and customer table and group invoices by customer\_id, first\_name, last\_name and count how many distinct genres their purchases cover.

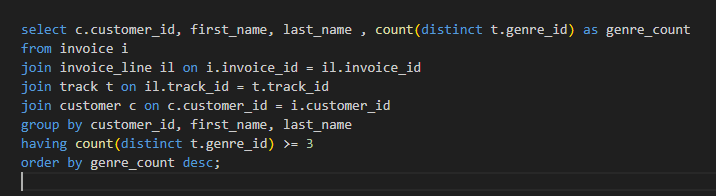
**STEPS**

1. Join invoice, invoice\_line, and track to access customer and genre info.
2. Group results by customer\_id.
3. Count distinct genre\_id values per customer.
4. Filter for those with 3 or more genres.

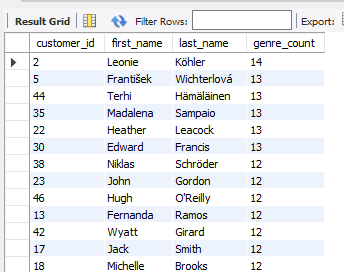
**RESULT**

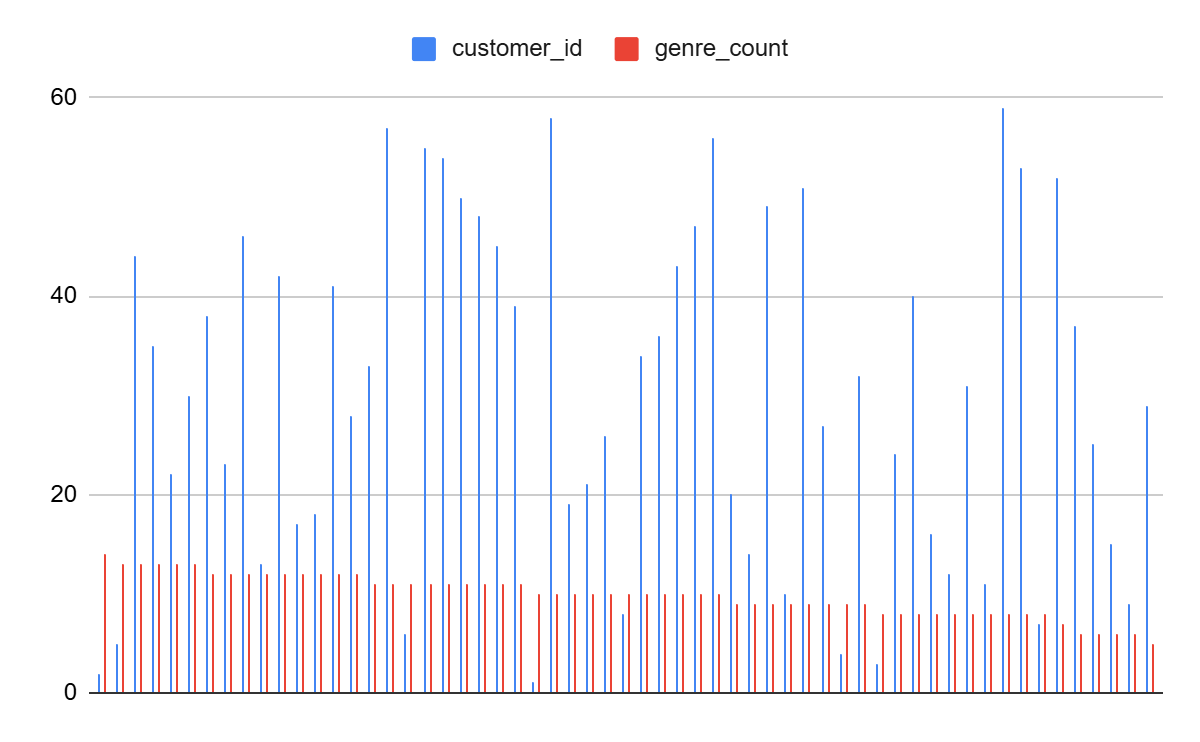
The output shows customer\_id, first\_name, last\_name and genre\_count who bought tracks from 3 or more different genres.

**CODE-**

****

**OUTPUT**

****

****

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

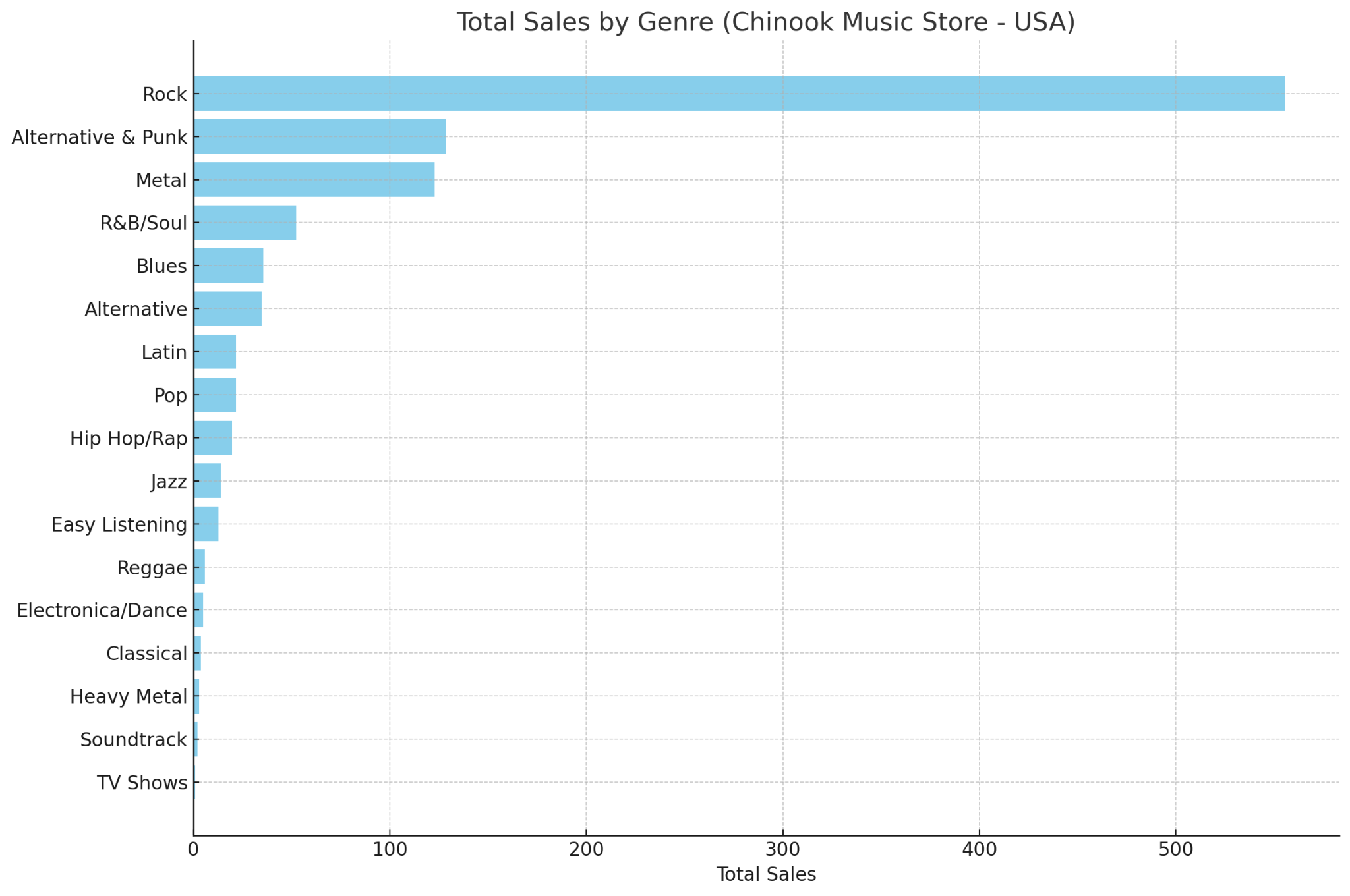
**APPROACH -** To Rank genres based on their sales performance in the USA we need to fetch genre table , track table, invoice\_line table, and invoice table

**STEPS**

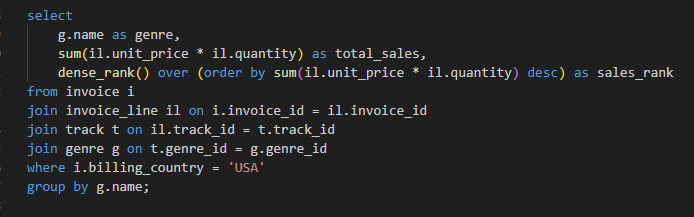
1. Fetch genre\_name from genre\_table
2. Calculate sum(il.unit\_price \* il.quantity) from invoice\_line table
3. Join the tracks table on track\_id, invoice\_table on invoice\_id , invoice\_line table on invoice\_id and tracks table on tracks id
4. Add condition where billing country = “USA”
5. Group them by genre\_name
6. dense\_rank() function assign rank to each of them based on their performance

**RESULT**

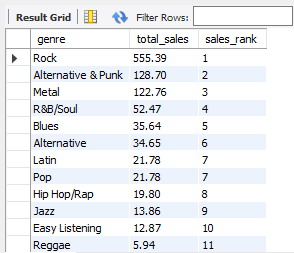
Output contains genre names, their sales performance figures, and their rankings, providing insights into which genres are the most successful in terms of sales in the USA.



**CODE**

****

**OUTPUT**

****

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

**Approach** - To Identify customers who have not made a purchase in the last 3 month,we need to fetch customer table and invoice table

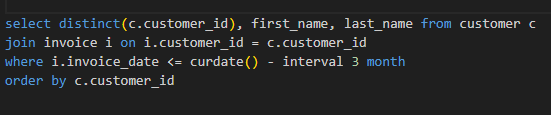
**STEPS**

1. Select customer\_id, first\_name, last\_name form customer
2. Join Cutomer table and invoice table on customer\_id
3. Check invoice date is less than 3 month using curdate() function

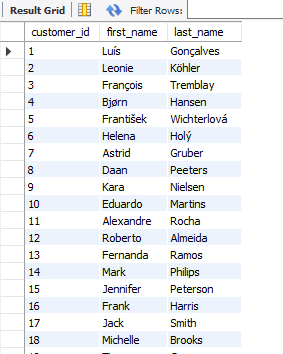
**RESULTS**

The output list contains customer\_id, first\_name, last\_name who have not made a purchase in 3 months

**CODE**



**OUTPUT**



Subjective Questions

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.

**APPROACH**

To Identify the top 3 best selling albums in the USA, we categorized them by genre and title and order by sales rank

**EXPLANATION**

1. The query calculates total sales by multiplying unit price and quantity sold for each track.
2. It groups the results by genre and album title.
3. It ranks albums using dense rank based on their total sales.
4. It filters for customers only in the USA to focus on regional sales performance.
5. It limits the output to the top 3 highest earning albums.

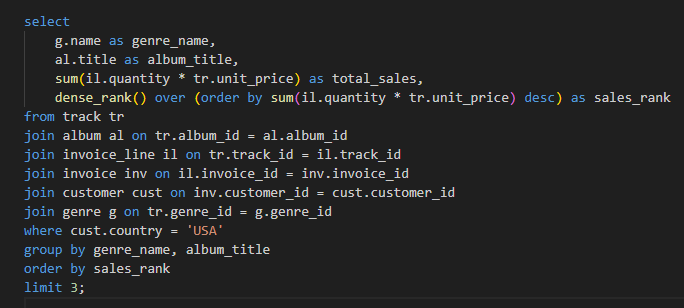
**STEPS**

1. Join track, album, invoice line, invoice, customer, and genre tables to gather all relevant data.
2. Filter customers from the USA.
3. Group data by genre and album title.
4. Calculate total sales using sum of unit price multiplied by quantity.
5. Rank albums using dense rank function.
6. Limit the output to the top 3 results.

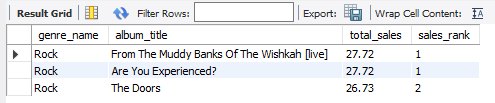
**INSIGHTS**

1. The top 3 albums in the USA market contribute the highest revenue, making them key products for the company.
2. Identifying strong genres and albums helps guide decisions on artist promotions and future inventory.

**CODE**



**OUTPUT**



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

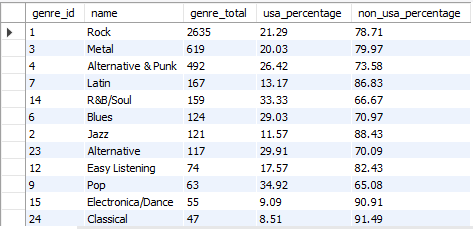
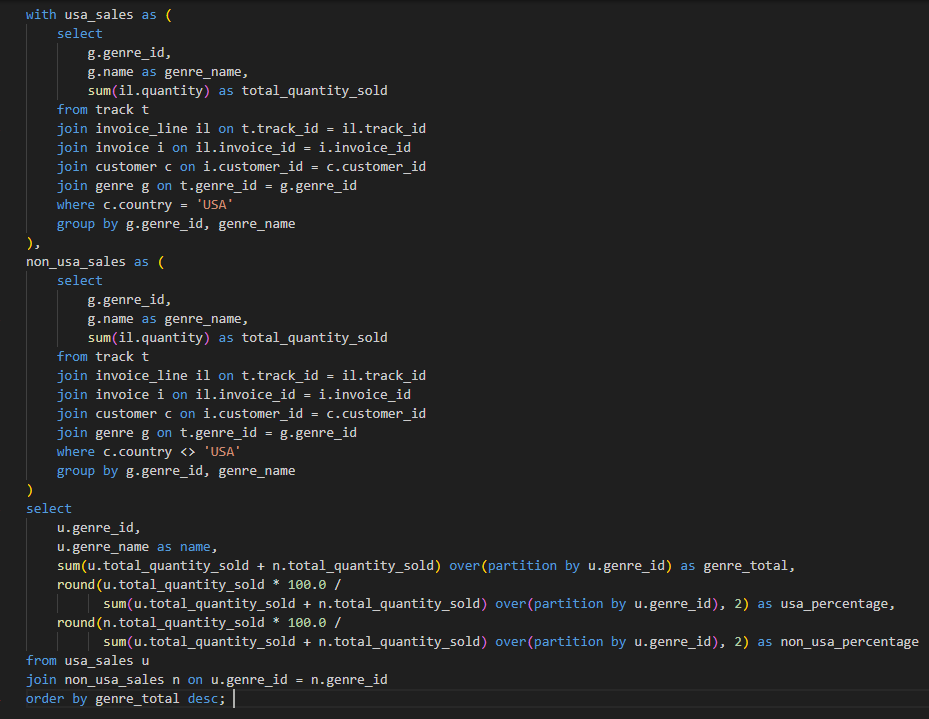
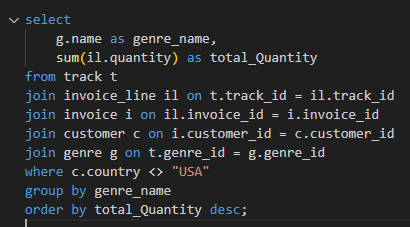
**APPROACH**

1. Identify top-selling music genres outside the USA.
2. Compare genre performance between USA and non - USA regions by sales volume.

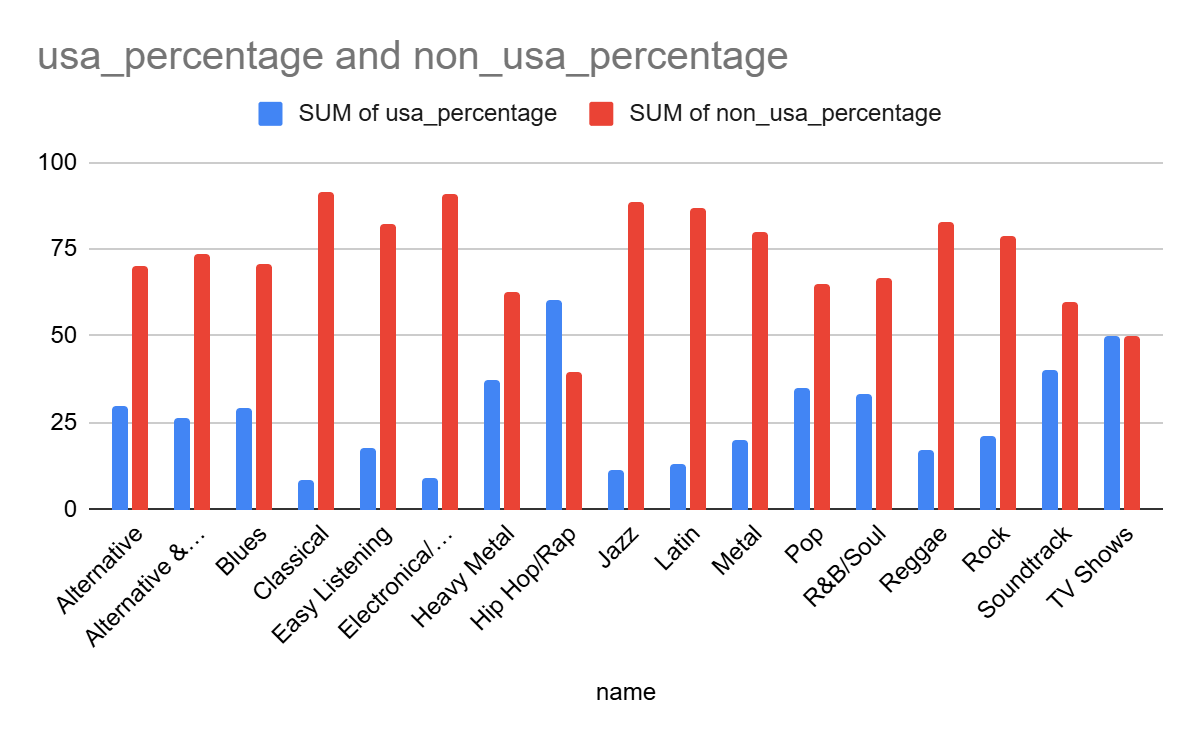
**STEPS**

1. Join relevant tables (track, invoice\_line, invoice, customer, genre) to link genres with customer purchases.
2. Use WHERE clause to filter USA and non-USA sales separately in two CTEs.
3. Group by genre to compute total quantity sold per genre for both regions.
4. Join both datasets on genre\_id.
5. Calculate total sales per genre and compute regional percentage contributions.
6. Order the result by total sales to get a clear genre ranking.

**CODE**

****

**VISUALIZATION**

**INSIGHTS**

1. Strong Presence of Non-USA Sales: Genres like "Classical" (91.49%), "Electronica/Dance" (90.91%), and "Jazz" (88.43%) show a much higher percentage of sales outside the USA. This suggests these genres might have a more global fanbase and are particularly popular in countries other than the USA.
2. Genres with Balanced Sales: Some genres, such as "Soundtrack" (50%) and "TV Shows" (50%), have an equal split between USA and non-USA sales. This indicates that these genres may appeal equally to both domestic and international markets.
3. USA-Dominated Genres: On the other hand, genres like "Hip Hop/Rap" (60.61%) and "Heavy Metal" (37.5%) have a significant share of sales from the USA, highlighting that these genres are particularly popular among American listeners compared to other regions.
4. 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?

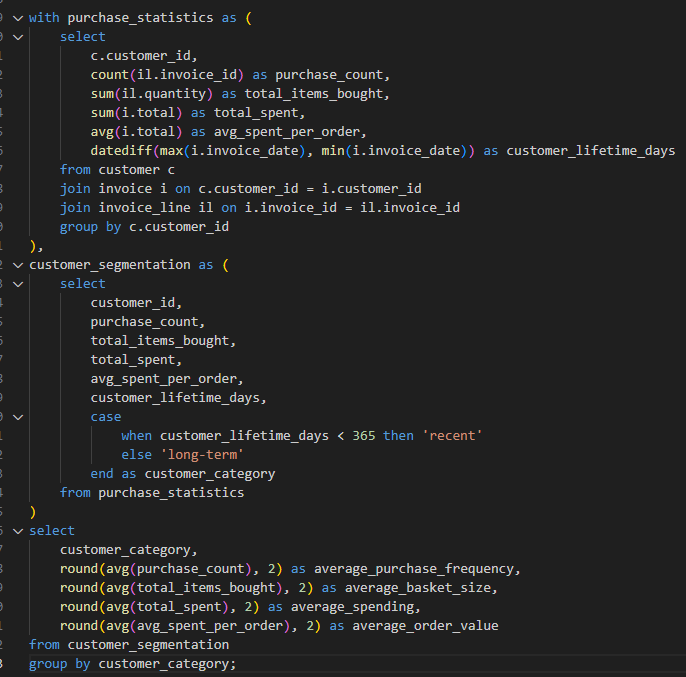
**APPROACH**

1. **Segmentation of Customers**: The first part of the query calculates the basic statistics for each customer, such as the number of purchases, total items bought, total money spent, average amount spent per order, and customer tenure in days.
2. **Classification of Customer Types**: The second part of the query assigns each customer to one of two segments based on their tenure: "New" if the customer has been with the company for less than 365 days and "Long\_Term" if the customer has been with the company for longer.
3. **Average Calculation by Segment**: The final part of the query calculates average statistics for each customer segment, such as the average purchase frequency, basket size, total spending, and average order value.

**STEPS**

1. **CTE: purchase\_statistics**: Collects and aggregates data for each customer, including the number of purchases, total items bought, total money spent, and customer lifetime.
2. **CTE: customer\_segmentation**: Classifies customers as "New" or "Long\_Term" based on their tenure.
3. **Final Select**: Aggregates the statistics for each customer segment and provides average values for the metrics (purchase frequency, basket size, spending amount, and order value).

**CODE**

****

**OUTPUT**

****

**INSIGHTS**

1. **New vs. Long-Term Customers**: This approach helps to understand the behavior of new customers compared to long-term ones, which can be useful for personalized marketing strategies.
2. **Spending Behavior**: The query allows for identifying spending patterns, such as whether new customers are spending less or if long-term customers contribute more to the overall sales.
3. **Customer Retention**: By segmenting customers based on their tenure, the data provides insights into customer retention and can highlight the importance of retaining long-term customers for sustained business growth.
4. 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?

**APPROACH**

To find which music genres, artists, or albums are frequently purchased together by customers we need to find genre affinity, artistaffinity and album affinity.The queries provided analyze affinities between music products (genres, artists, albums) by looking at their co-occurrence in customer purchase transactions. The analysis focuses on identifying which products (genres, artists, albums) are often purchased together by customers.

**EXPLANATION**

**Track Combination (Common to All Analyses)**:

* First, we identify pairs of tracks that are frequently purchased together by customers.
* The track\_combination CTE (Common Table Expression) pairs tracks within the same invoice (invoice\_id), counting how many times each pair of tracks is purchased together. This gives a foundation for exploring affinities between music products at different levels (genre, artist, album).

**Genre Affinity**:

* After identifying track-level co-purchases, the genre\_combination CTE groups the tracks by their genres, counting how often different genres are purchased together based on track co-purchases.
* This helps to find pairs of genres that share customer interest. For example, if customers tend to buy Rock and Blues tracks together, these genres are considered to have a strong affinity.
* The final query then returns the most commonly purchased genre pairs, sorted by the frequency of co-purchase.

**Artist Affinity**:

* In the artist\_combination CTE, we map track-level co-purchases to artists. This is done by identifying the artists of the tracks and counting how often two different artists' tracks are purchased together.
* This helps us understand which artists are frequently bought together. The final output presents pairs of artists that are commonly bought together, which could guide cross-promotion or collaborative marketing between artists.

**Album Affinity**:

* The album\_combination CTE focuses on the relationship between albums. It counts how often tracks from two different albums are purchased together.
* By understanding which albums are frequently purchased together, we can recommend albums to customers based on their current purchase patterns. This could be especially useful for suggesting albums that complement others.

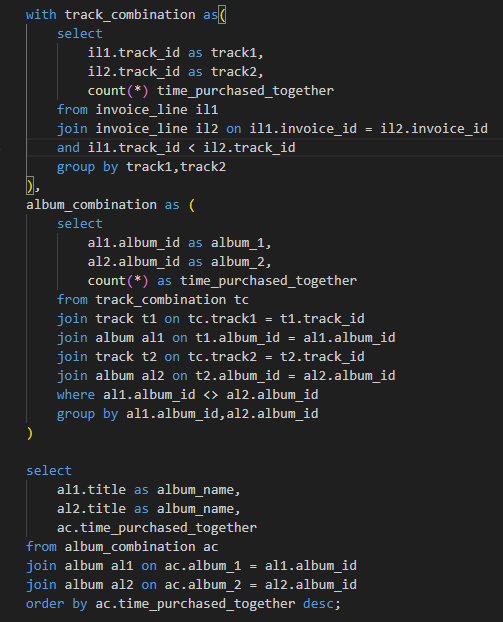
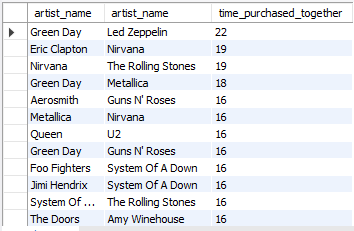
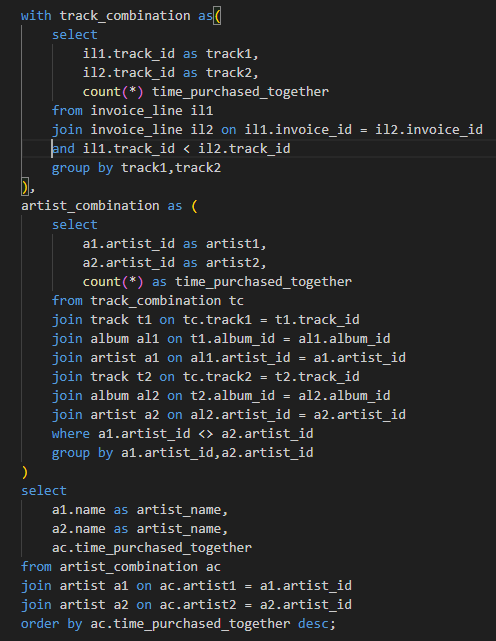
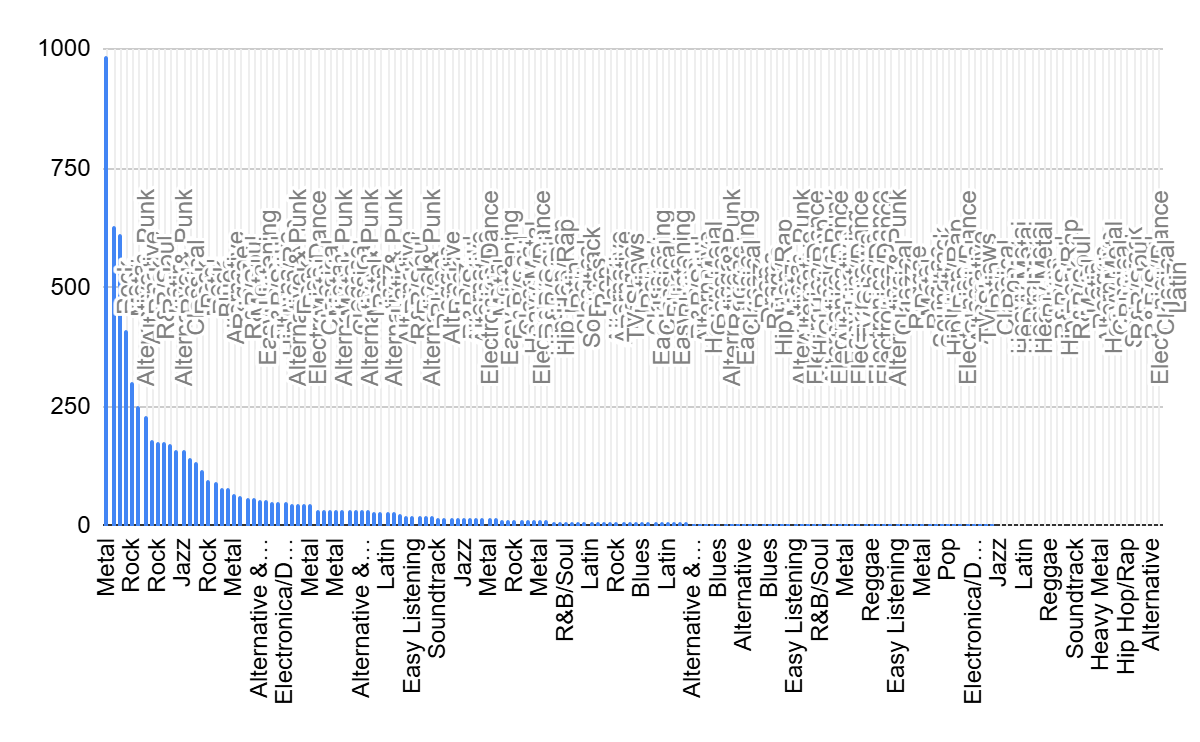
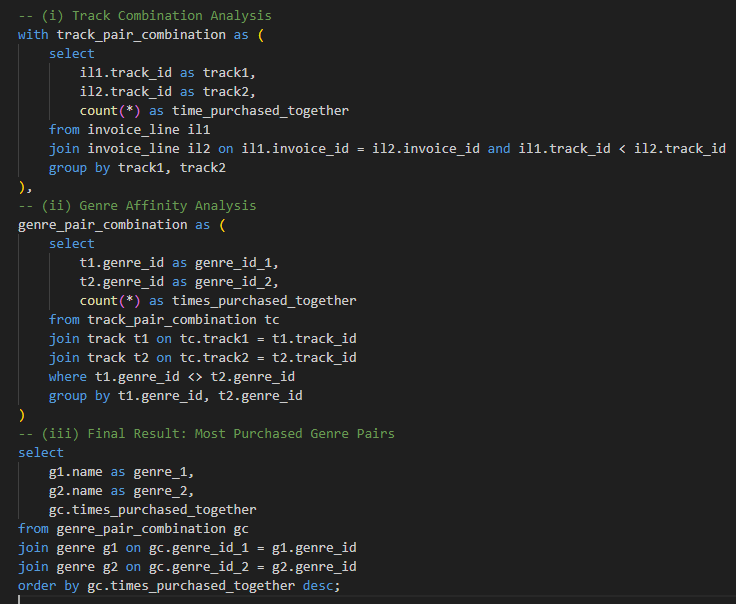
**Insight Generation**:

* For each of the three affinity analyses (genre, artist, album), the results show pairs of related products (genres, artists, albums) that are most commonly purchased together.
* These results can be used for personalized product recommendations or for creating bundles or playlists based on customer preferences.

**Use Cases for Cross-Selling and Recommendations**:

* **Product Bundles**: Use the affinities to create product bundles. For instance, if a customer buys a "Rock" album, recommend a "Blues" album if those genres are frequently bought together.
* **Personalized Recommendations**: Suggest albums or tracks from artists and genres that the customer has already purchased in the past, increasing customer satisfaction and sales.
* **Collaborative Marketing**: Artists or albums with strong affinity relationships can be promoted together in campaigns or through social media collaborations.

**CODE**

****

**INSIGHTS**

1. **Strong Genre Affinity**: If genres like Rock and Metal are often purchased together, it may indicate that customers who enjoy hard rock are also likely to enjoy heavy metal. Marketing these genres together could increase cross-selling.
2. **Artist Partnerships**: If certain artists are frequently purchased together, consider partnerships between those artists. This could lead to joint promotions or concert collaborations.
3. **Album Pairing**: When albums from different artists or genres are often purchased together, it suggests customers are interested in albums with similar themes or styles. Suggesting albums with related tracks could enhance user experience and increase sales.
4. **Cross-Promotion Opportunities**: Use genre, artist, or album combinations for cross-promotion. If certain genres or albums are highly correlated, bundle them together for promotions or create genre-based playlists.
5. **Personalized Marketing Campaigns**: Use the affinity data to send personalized emails or recommendations to customers based on their past purchases and the affinity analysis results.
6. **Enhanced Customer Retention**: By recommending products that a customer is likely to enjoy based on affinity analysis, customer satisfaction and retention could increase, leading to long-term growth in sales.
7. 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?

**APPROACH**

To find customer purchasing behaviors and churn rates vary across different geographic regions or store locations we need following tables

1. **invoice**: Contains data related to customer purchases, including invoice details.
2. **customer**: Stores customer information like location (country, state, city).
3. **customer\_purchase\_summary**: A derived table that aggregates purchase details such as frequency, spending, and average value per customer.
4. **latest\_purchase\_by\_region**: Calculates the most recent purchase date for each customer, categorized by location.
5. **churned\_customers\_by\_region**: Identifies customers who haven't made a purchase in the last year.

**STEPS**

**Customer Purchase Data**:

1. Aggregate data from the invoice table to calculate total purchase frequency, spending, and average value for each customer.
2. Combine this data with customer location information from the customer table.

**Regional Purchase Summary**:

1. Group data by region (country, state, city) and calculate total customers, total purchases, total spending, average order value, and average spending per customer for each region.

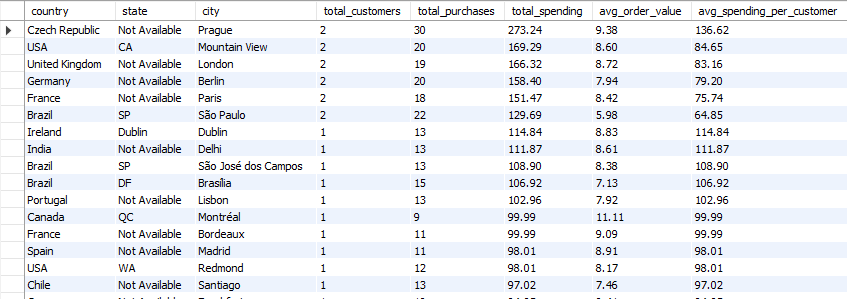
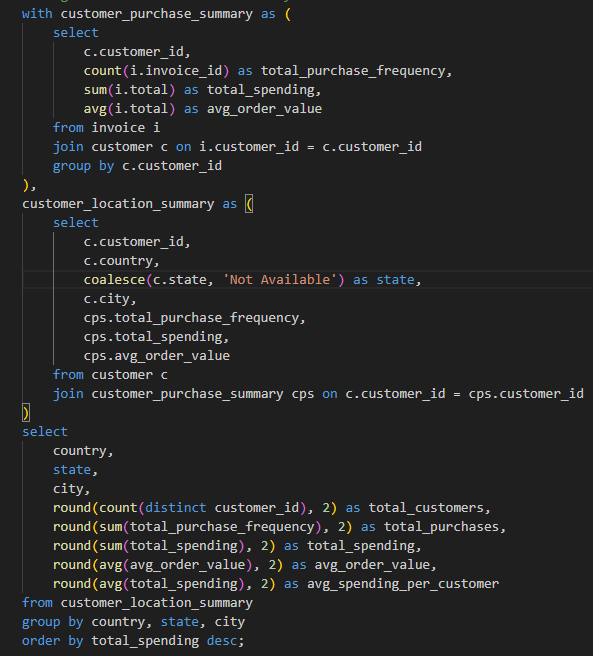
**Churn Calculation**:

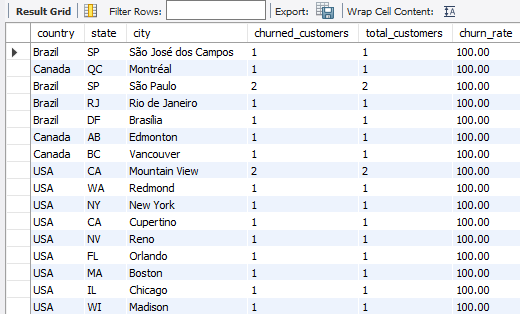
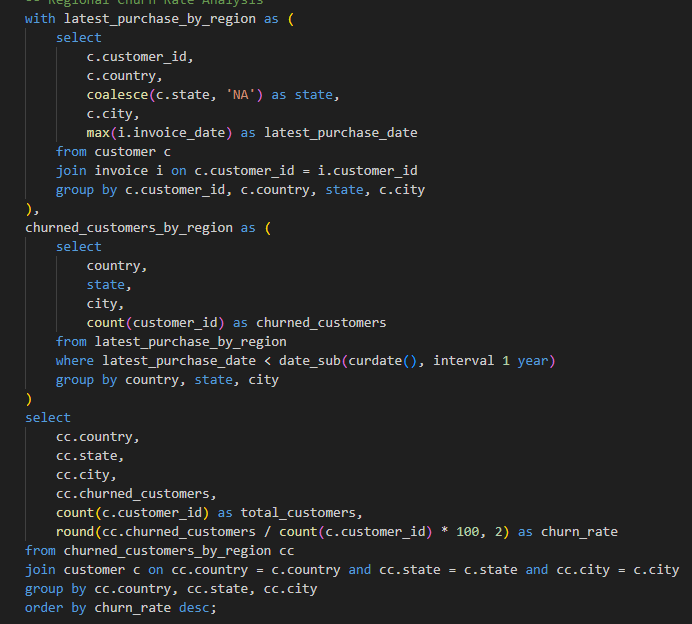
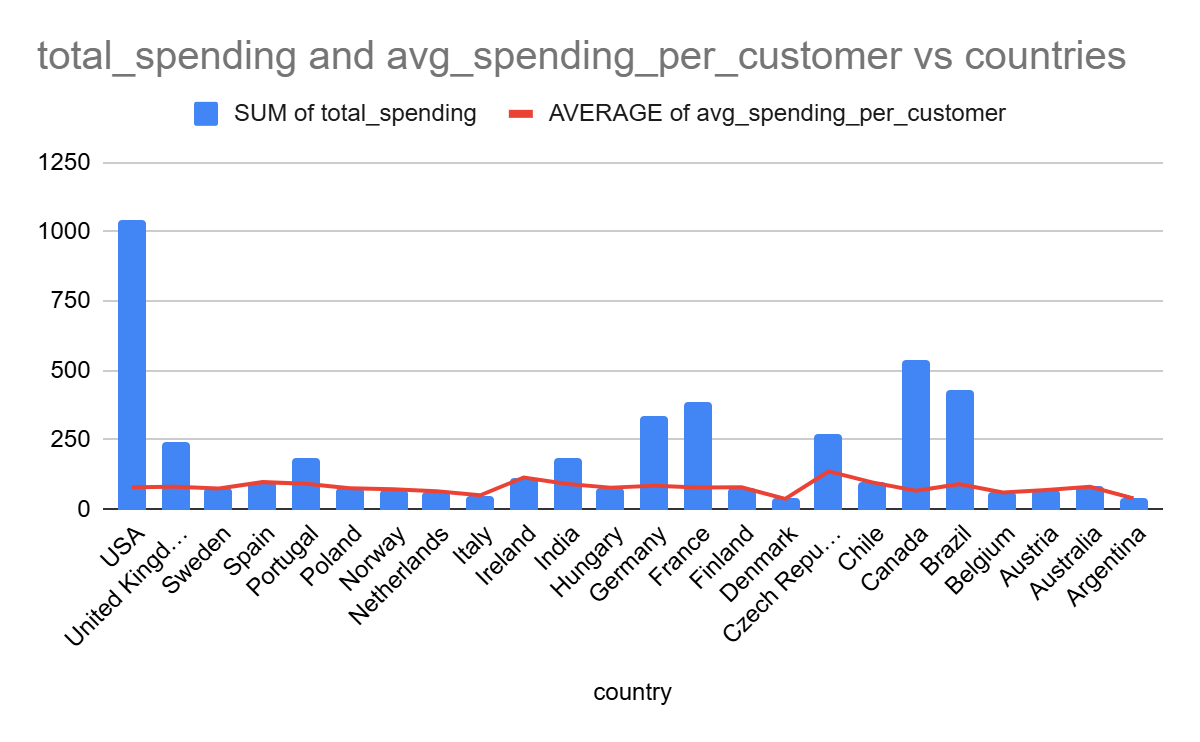
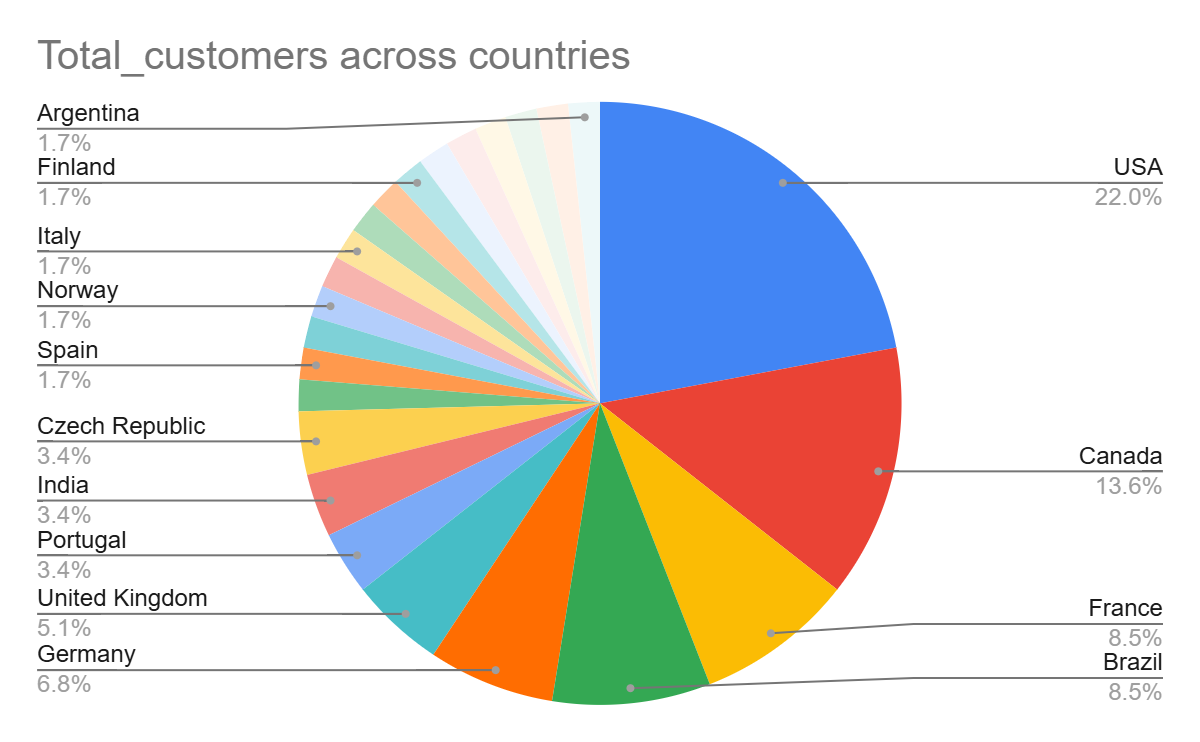
1. Find the most recent purchase date for each customer by joining the customer and invoice tables.
2. Identify churned customers (those with no purchase in the last year).

**Churn by Region**:

1. Calculate the churn rate by dividing churned customers by the total customers for each region.

**CODE**

****

****

**INSIGHTS**

1. **Income Levels and Spending Power**: Regions with higher average spending per customer may correlate with higher income levels or wealthier demographics. Customers in these areas might have more disposable income, leading to higher purchase frequency and total spending.
2. **Population Density and Market Size**: Areas with larger populations or more urbanized regions tend to have a greater number of customers, which can impact the total number of purchases and spending. In more densely populated areas, there could also be more competition, leading to different purchasing behaviors.
3. **Economic Conditions**: Economic factors, such as local economic growth or downturns, can affect consumer confidence and spending habits. A prosperous economic environment typically results in higher customer spending and lower churn, while areas in economic decline may experience higher churn rates and reduced purchasing frequency.
4. **Cultural Preferences**: Local tastes and preferences, influenced by culture or regional trends, can affect the types of products purchased and the average order value. Regions with specific cultural trends may have distinct spending behaviors, leading to variations in the kinds of purchases made.
5. 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?

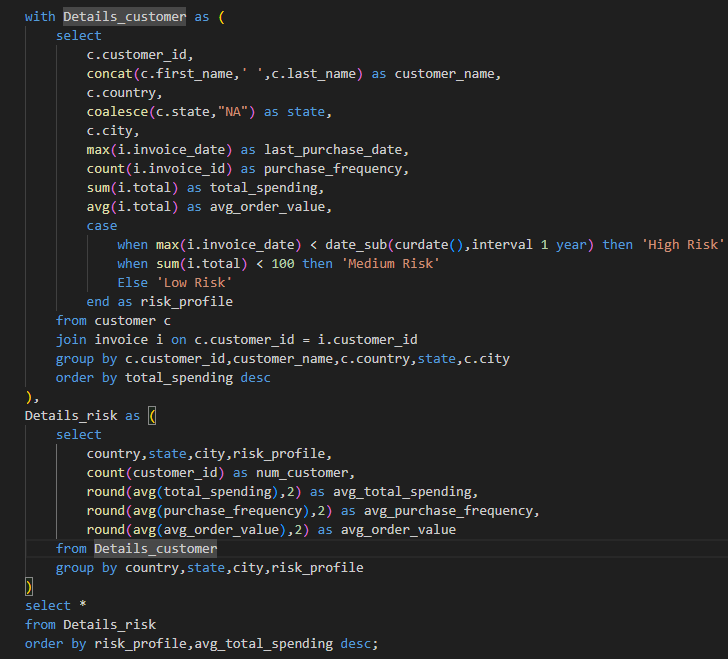
**APPROACH**

1. Used customer and invoice data to analyze behavior based on purchase history and activity.
2. Segmented customers by location and defined categories based on inactivity and spending.

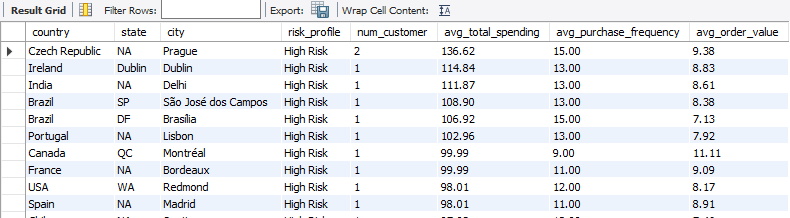
**STEPS**

1. Joined customer and invoice data to compute each customer's total spending, order frequency, average order value, and last purchase date.
2. Classified customers into three categories:
   1. No purchases in the last year.
   2. Total spending less than 100.
   3. Active and spending above threshold.
3. Aggregated data by location to summarize counts, average spending, frequency, and order values.

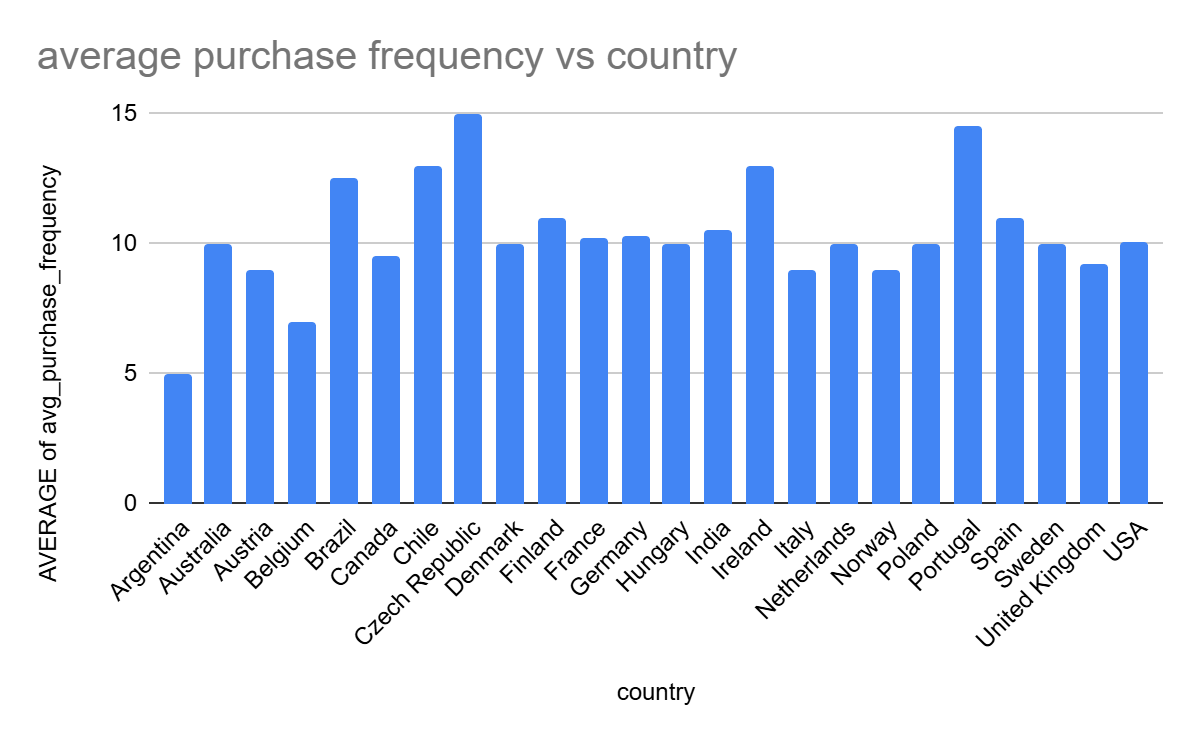
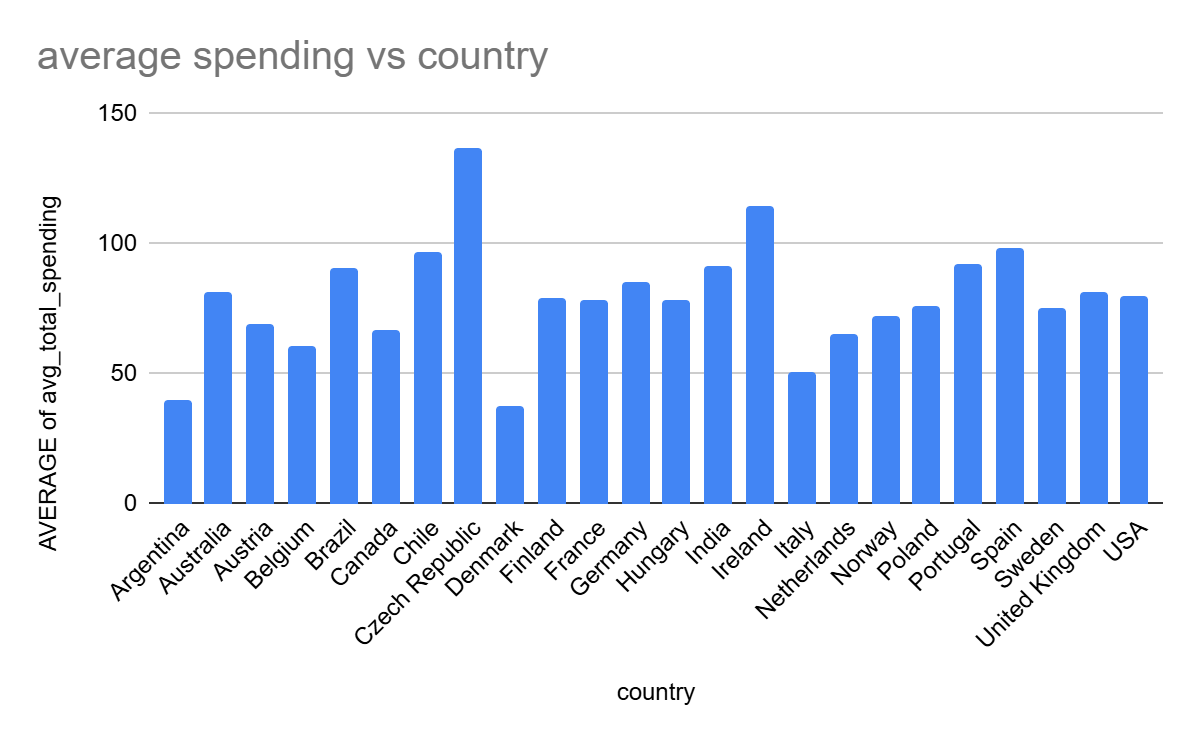
**CODE**

****

**OUTPUT**

****

**VISUALIZATION**

****

**INSIGHTS**

1. Locations like Prague, Dublin, and Delhi show inactive customers with high average spending and frequency, indicating potential value.
2. Inactive customers show reasonable order values, suggesting retention efforts may be effective.
3. Regions like Canada and France have consistent behavior among inactive customers, possibly due to external conditions.
4. Customer Lifetime Value Modeling: 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?

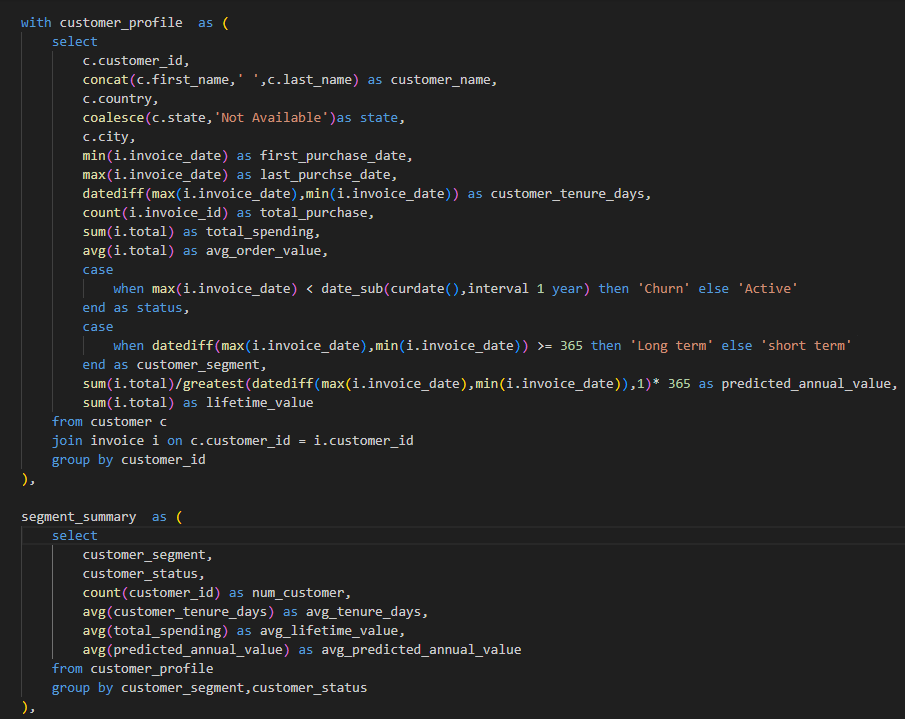
**APPROACH**

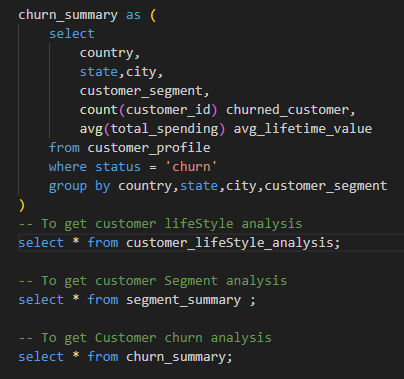
1. Used customer and invoice records to calculate tenure, total spending, order behavior, and determine customer status.
2. Segmented customers by engagement duration and churn status to assess lifetime value trends.
3. Analyzed churned customer patterns across locations and tenure groups.

**STEPS**

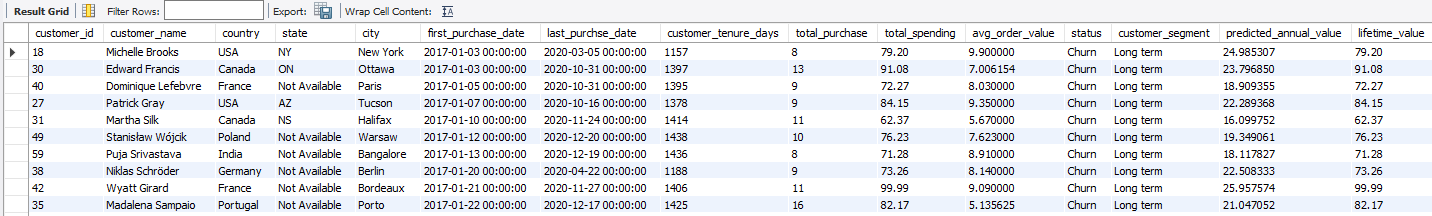
1. Created a profile for each customer with purchase dates, tenure, number of purchases, total spend, and order average.
2. Labeled customers as churned or active based on their last purchase.
3. Grouped customers into long-term and short-term segments using tenure duration.
4. Estimated annual value and total lifetime value for each customer.
5. Summarized customer segments and status to understand trends in spending and retention.
6. Analyzed churned customers across locations and segments to detect patterns.

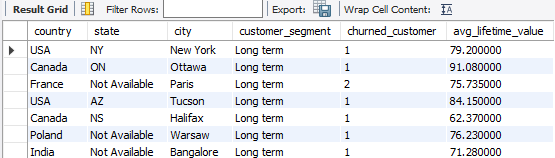
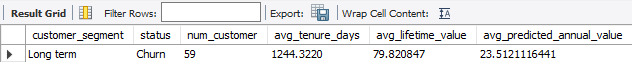
**CODE**

****

****

**OUTPUT**

****

****

**INSIGHTS**

1. Long-term customers tend to have higher predicted and actual lifetime value compared to short-term ones.
2. Most churned users fall under short-term segment, indicating lack of long-term engagement.
3. Cities like Prague, Delhi, and Dublin show higher churn in short-term users despite notable spending.
4. Customers with frequent purchases but short tenures may require better onboarding or loyalty strategies to boost retention.
5. If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?

**APPROACH**

To evaluate the impact of promotional campaigns (such as discounts, events, or email marketing) on customer acquisition, retention, and overall sales, a structured analysis approach can be used

**STEPS**

**1. Customer Acquisition Impact**

* Track the number of new customers acquired during promotional periods compared to similar non-promotional timeframes.
* Use metrics such as sign-up rates, first-time purchases, and acquisition cost per customer to gauge effectiveness.

### **2. Customer Retention and Engagement**

* Measure retention by analyzing repeat purchase behavior post-campaign.
* Compare customer return rates within 30, 60, or 90 days after campaigns against baseline periods.

### **3. Sales Performance and Uplift**

* Calculate the increase in total revenue and average order value during promotional periods.
* Use a pre-post comparison or control-test group approach to isolate campaign effects.
* Adjust for seasonality or external trends to better understand true sales uplift.

### **4. Segment-Level Analysis**

* Evaluate campaign performance across different customer segments (e.g., age, location, spending tier).
* Understand which groups are more responsive to specific types of promotions (e.g., high-value customers reacting to loyalty events vs. new users reacting to first-time discounts).
* Optimize future campaigns based on segment responsiveness.

### **5. Channel Effectiveness**

* Break down results by marketing channel (email, social media, app notifications) to assess which source led to the most conversions or repeat visits.
* Identify the best-performing communication strategies and timing.

### **6. Customer Lifetime Value (CLV) Shift**

* Track whether customers acquired through campaigns show higher or lower lifetime value over time.
* Evaluate if promotions attract long-term loyal customers or short-term deal seekers.

### **7. Campaign Profitability**

* Compare campaign cost against revenue generated to determine return on investment (ROI).
* Monitor metrics like cost per acquisition (CPA), revenue per campaign, and margin dilution due to discounts.

**INSIGHTS**

1. Long-term customers tend to have higher predicted and actual lifetime value compared to short-term ones.
2. Most churned users fall under short-term segment, indicating lack of long-term engagement.
3. Cities like Prague, Delhi, and Dublin show higher churn in short-term users despite notable spending.
4. Customers with frequent purchases but short tenures may require better onboarding or loyalty strategies to boost retention.
5. How would you approach this problem, if the objective and subjective questions weren't given?

**APPROACH**

1. **Explore the Dataset First**: Begin with an open-ended exploration, understanding the structure, columns, data types, and distributions without being guided by specific questions.
2. **Let Data Guide the Questions**: Derive questions organically based on what patterns or anomalies are observed in the dataset.
3. **Use Domain Understanding**: Apply business context (e.g., in a music sales dataset: albums, genres, customer locations) to form meaningful hypotheses.
4. **Prioritize Business Value**: Focus on aspects of the data that could influence strategy, such as sales growth, retention, churn, or user engagement.

**STEPS**

**Data Quality Checks**

* Detect and handle missing values, incorrect formats, duplicates, and inconsistencies.
* Standardize categorical values (e.g., country names, genres).

**Descriptive Statistics**

* Understand basic distributions (e.g., total purchases, average spend).
* Plot data trends to uncover seasonality, anomalies, or spikes.

**Generate Analytical Questions**

* Which genres, albums, or artists have the highest sales?
* How do spending patterns vary by region?
* Who are the most loyal or high-value customers?

**Customer Segmentation**

* Group customers by demographics, purchase frequency, lifetime value, or product preferences.
* Identify segment-specific behaviors (e.g., rock lovers in Europe vs. classical fans in Asia).

**Churn Identification**

* Flag customers who haven't made a purchase in a long time.
* Analyze patterns leading to inactivity (e.g., genre boredom, price sensitivity).

**Comparative and Trend Analysis**

* Compare sales over time, between countries or platforms (mobile vs. desktop).
* Examine product success before and after price changes or promotions.

**Sales & Revenue Deep Dive**

* Identify top and bottom performers (albums, artists, genres).
* Calculate average order value, repeat purchase rate, and most active cities or countries.

**Visual Data Storytelling**

* Use dashboards or visuals (charts, heatmaps) to summarize findings for decision-makers.
* Create segment profiles or customer journeys.

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?

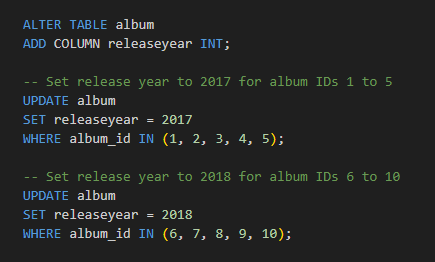
**APPROACH**

1. Extend the existing album table schema by adding a new column to store the release year of each album.
2. Populate the releaseyear field using conditional updates based on album identifiers.

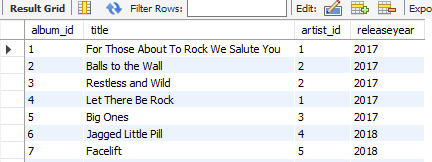
**STEPS**

1. **Schema Modification** Add a new column releaseyear of type INT to the album table to store the year each album was released.
2. **Data Update for Album Batches** Use two UPDATE statements with IN clauses to assign the release year:
   * Albums with album\_id from 1 to 5 are assigned 2017.
   * Albums with album\_id from 6 to 10 are assigned 2018.

**CODE**

****

**OUTPUT**

****

**INSIGHTS**

Using IN improves efficiency and readability by reducing repetition.

The new column allows for:

1. Time-based analysis of album trends.
2. Understanding patterns in releases by year.
3. Enabling reports to group or filter albums by release year.
4. 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 an SQL query to provide this information.

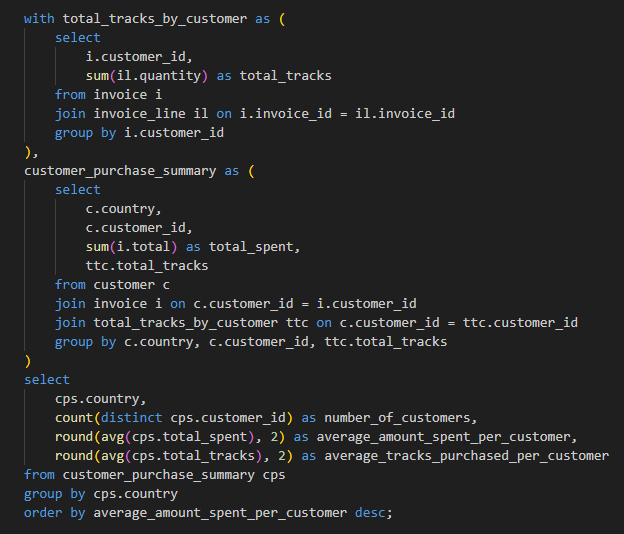
**APPROACH**

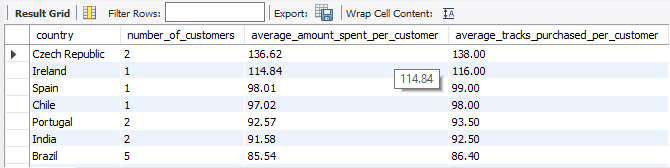
1. Break down the analysis into smaller calculations using subqueries to make the logic easier to follow and more efficient.
2. First, calculate total tracks bought by each customer.
3. Then, summarize total spending and track quantity along with customer location details.
4. Finally, aggregate this by country to get average values.

**STEPS**

1. **Calculate Tracks per Customer**Aggregate the total quantity of tracks each customer purchased using the invoice\_line table.
2. **Calculate Spending per Customer**Join customer data with invoices and the tracks-per-customer CTE to get each customer's total amount spent and total tracks.
3. **Aggregate by Country**Group the data by country to:
   1. Count how many customers are from each location.
   2. Calculate the average spending per customer.
   3. Determine the average number of tracks bought.
4. **Sort Output** Sort the results in descending order of average spending to easily identify high-value regions.

**CODE**

****

**OUTPUT**

**INSIGHTS**

1. Countries with high average spending and track purchases can be prime markets for targeted campaigns.
2. Understanding where customers purchase more tracks or spend more money helps tailor regional marketing and promotions.
3. These metrics can support pricing strategies or promotional planning by country.