**Learners have to come up with a Report to support the answers to the following questions and suggestions**

**Objective Questions**

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

**Observation:**

Firstly, need to check the null values in the columns of the provided tables e.g. album table. Below is the query to check if there are any null values in any of the columns:

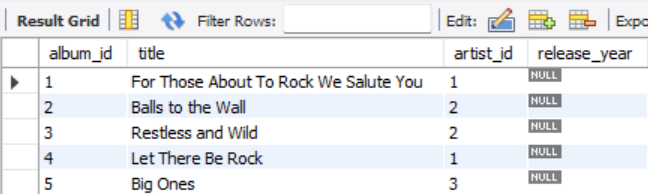
select \* from album

where album\_id is null

or title is null

or artist\_id is null

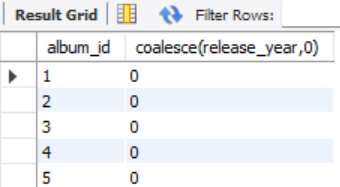
or release\_year is null;



Approach:

Here in the output we can see that the column release\_year has null values. Therefore to handle null values, we can replace the null values of text datatype to N/A and integer datatype to 0 by using COALESCE. Below is the query and output:

select album\_id,coalesce(release\_year,0) from album;



Similarly checked the null values in the other tables and replaced the null values to N/A or 0 in the customer , employee and track table. Following are the queries:

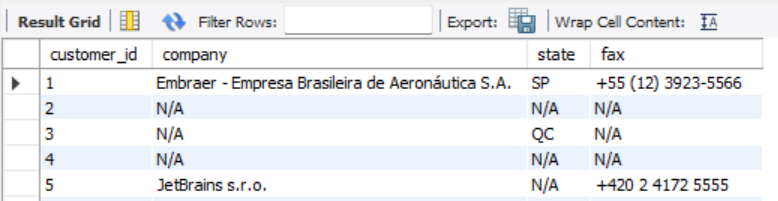
SELECT customer\_id,

COALESCE(company, 'N/A') AS company,

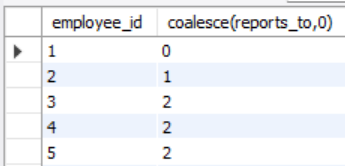
COALESCE(state, 'N/A') AS state,

COALESCE(fax, 'N/A') AS fax

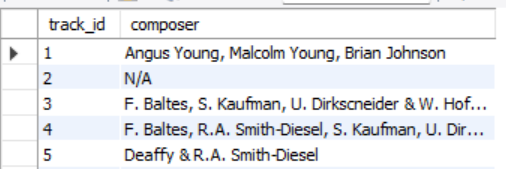
FROM customer;



select employee\_id,coalesce(reports\_to,0) from employee;

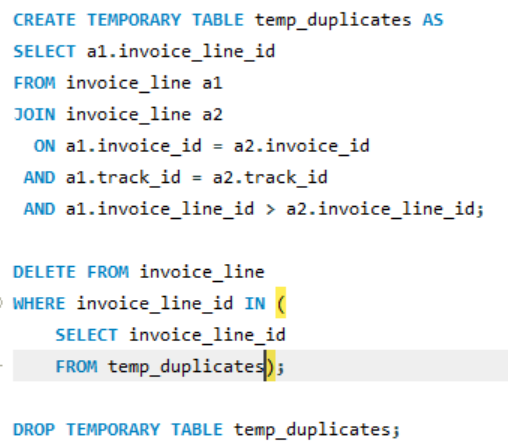


select track\_id, coalesce(composer,'N/A') as composer from track;

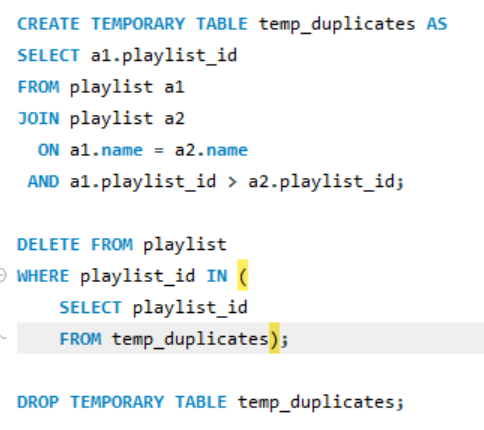


There are also duplicates in the invoice\_line and playlist tables, below are the queries:

**invoice\_line Table**

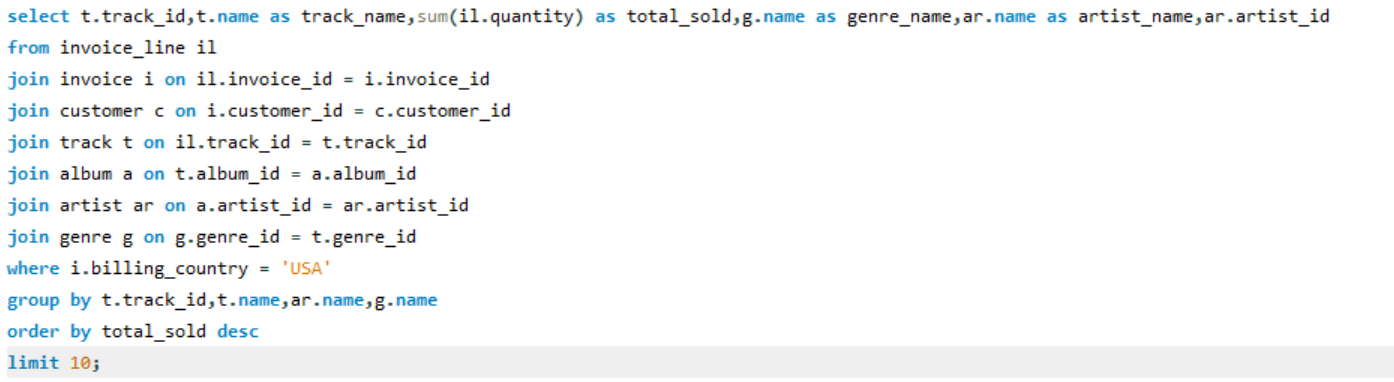


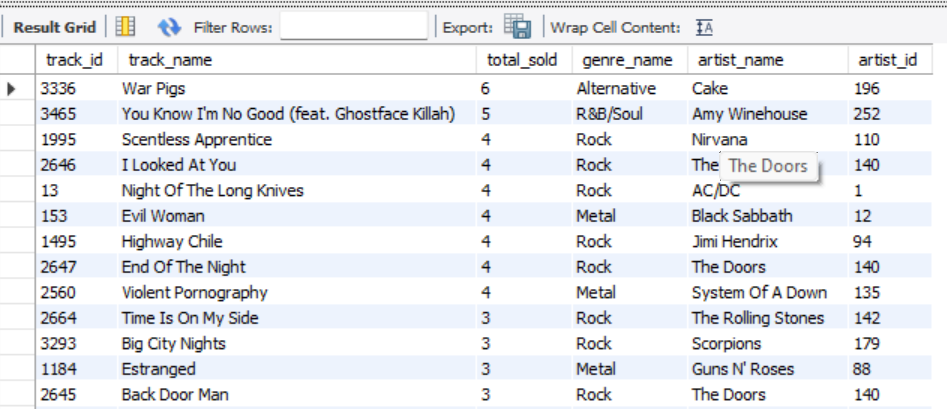
**Playlist Table**



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

**Top selling tracks in USA with their most famous genre:**

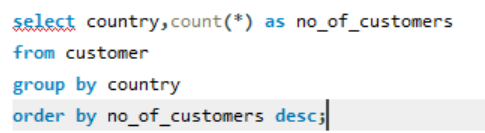


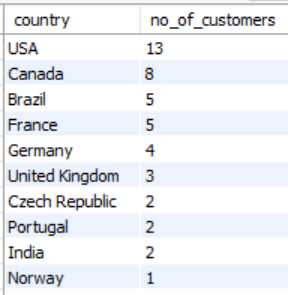


**3.**What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?

Based on dataset we can see the customer demographic breakdown for location i.e. country, state, city. Below are the query and output for the same:

**Country Breakdown:**

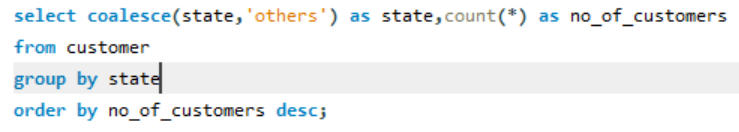


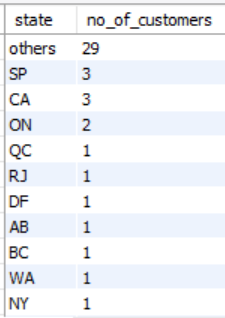


**Insights:**

USA and Canada have the highest no. of customers with 13 and 8 customers, whereas Chile has the lowest no. of customers with 1.

**State Breakdown:**

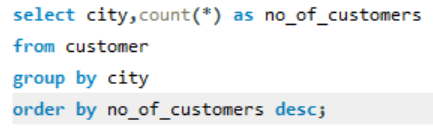


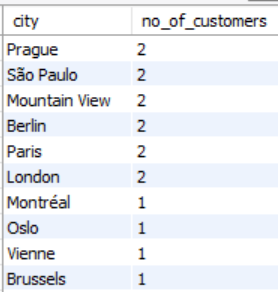


**Insights:**

Missing state information replaced with others as it has significantly more customers than individual states like SP and CA, which only have 3 each. Hence to keep the data accurate, missing states were grouped under "others" instead of most common state, with NSW having the fewest customers at 1.

**City Breakdown:**



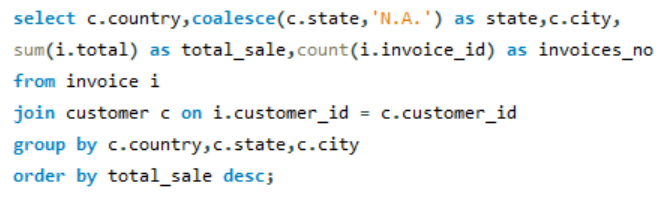


**Insights:**

Prague and Sao Paula are the leading city with the 2 customers each and other cities such as Sidney, Buenos Aires, Santiago, Delhi and Bangalore each with 1 customer.

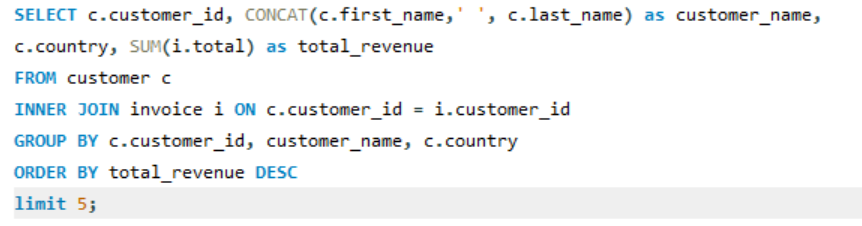
There are no age and gender columns are available in the customer table.

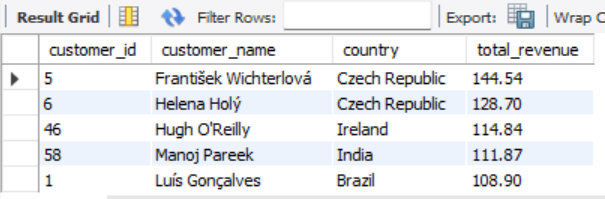
**4.**Calculate the total revenue and number of invoices for each country, state, and city:



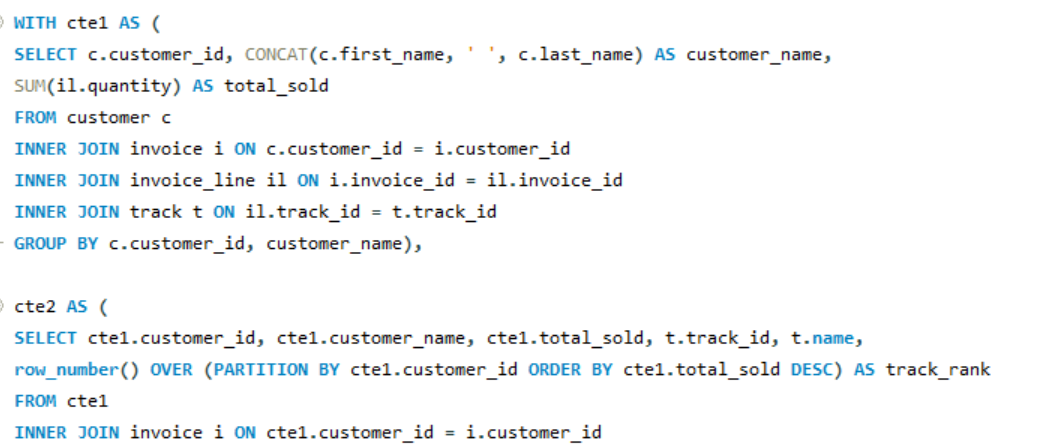
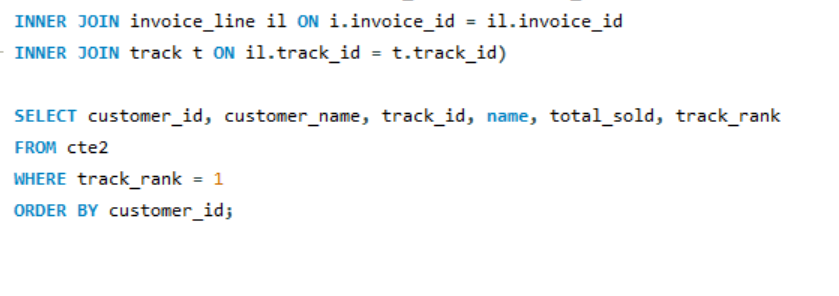


**5.**Find the top 5 customers by total revenue in each country





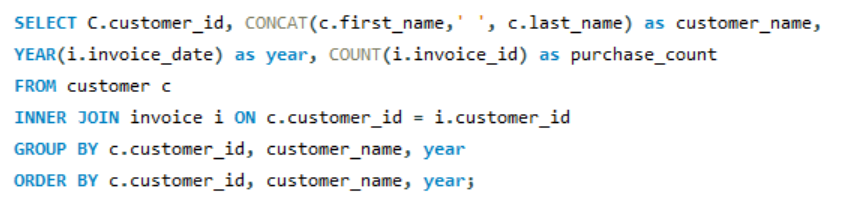
**6.**Identify the top-selling track for each customer



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

**Frequency of purchases:**



****

**Preferred payment mode:**

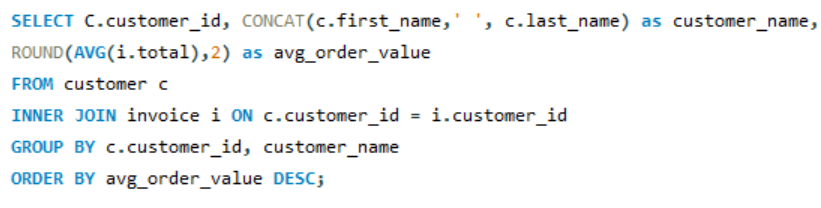
Could not find any payment mode related columns in the given database schema.

**Average order value:**

**Preferred payment mode:**

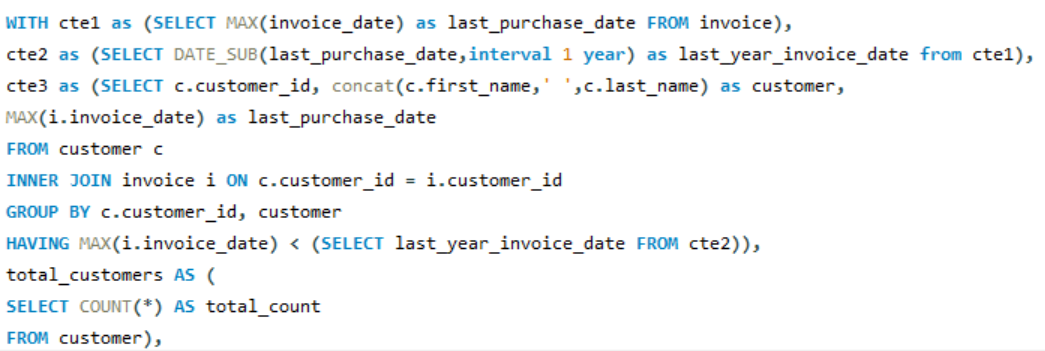
Could not find any payment mode related columns in the given database schema.

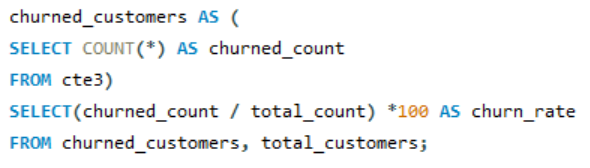
**Average order value:**

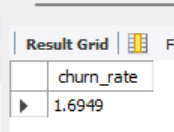




**8.**What is the customer churn rate?

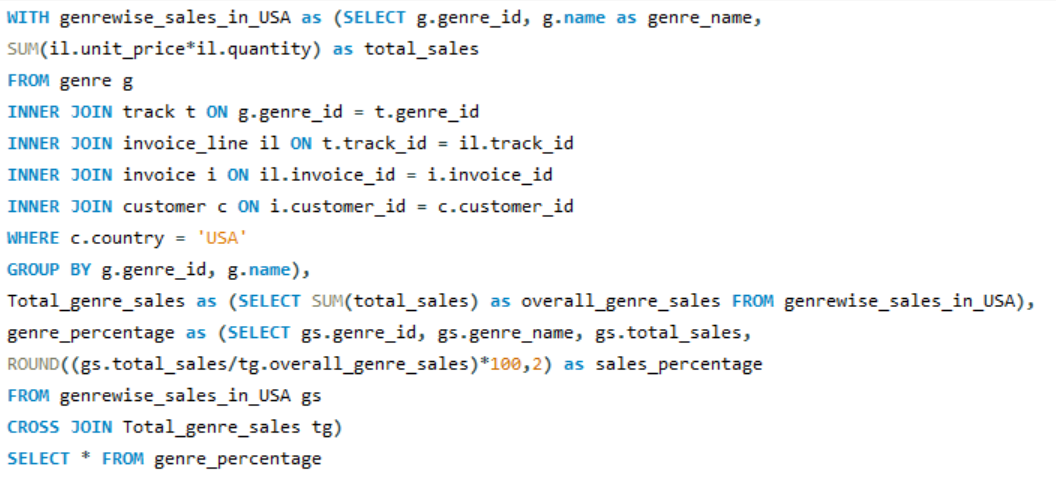
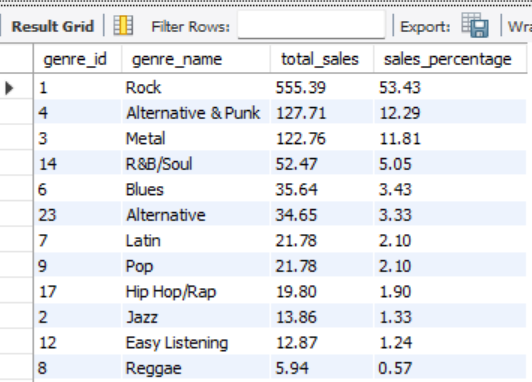




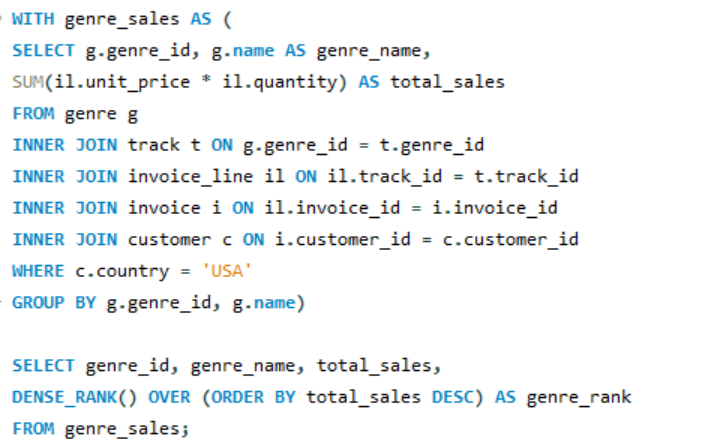


**9.**Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.

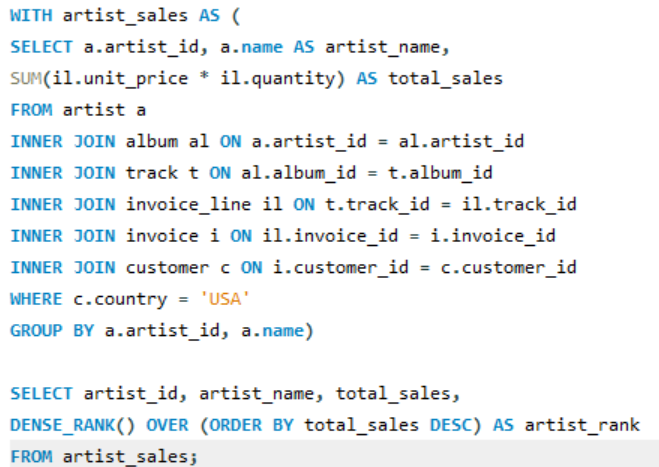
**the percentage of total sales contributed by each genre in USA:**

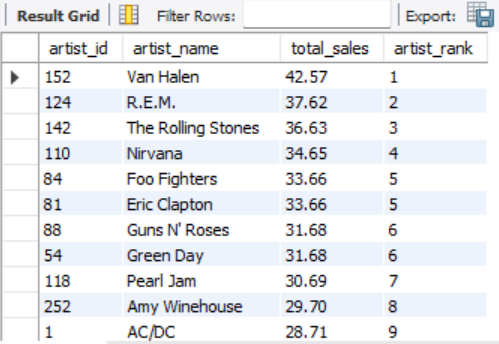
  

**Best-selling genres and artists:**

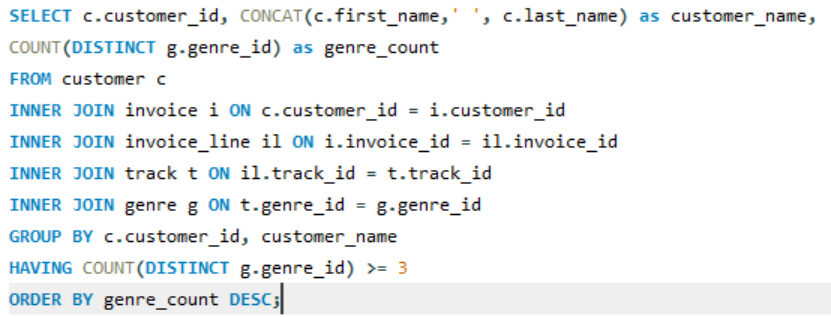


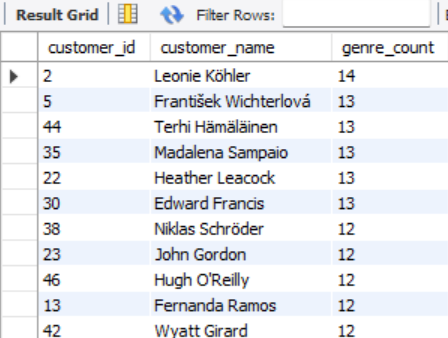




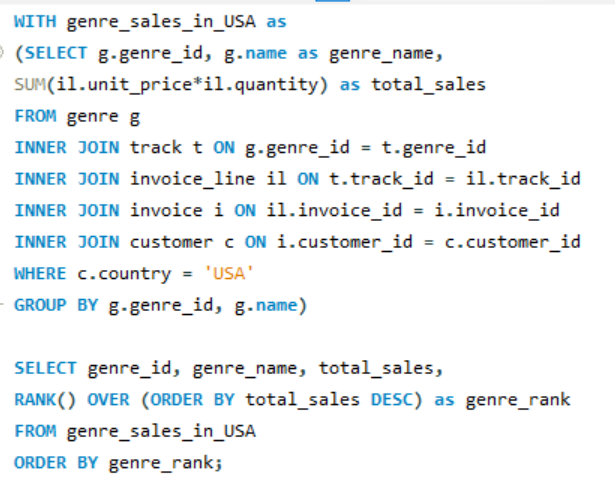


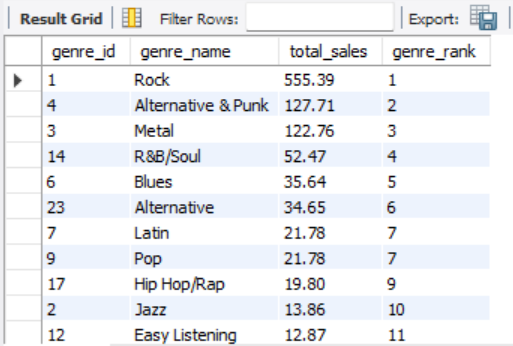
**10.**Find customers who have purchased tracks from at least 3 different genres



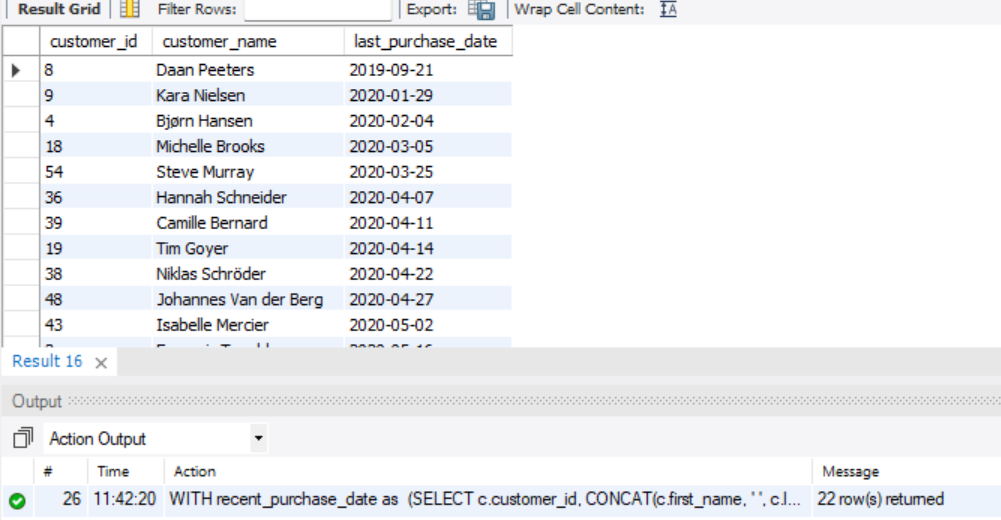
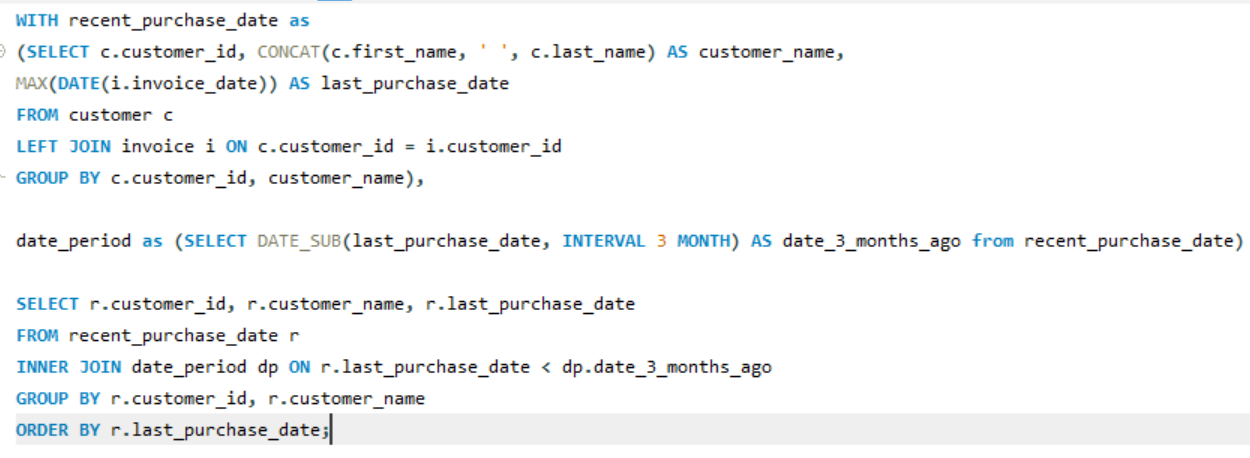


**11.**Rank genres based on their sales performance in the USA



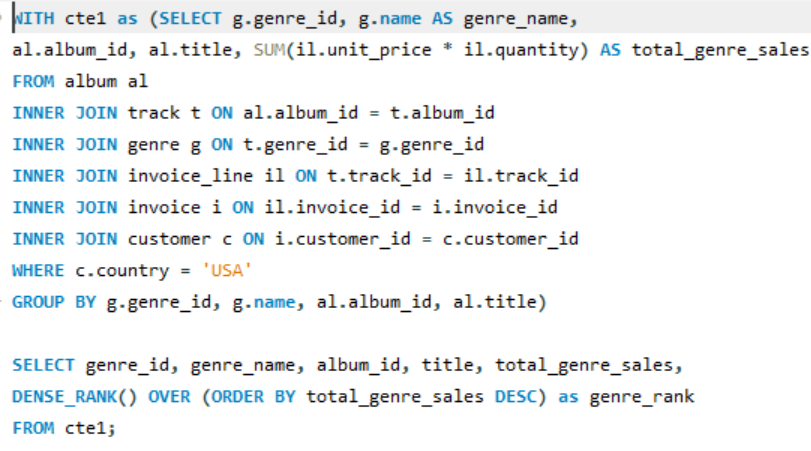


**12.**Identify customers who have not made a purchase in the last 3 months



**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.





The top 3 albums from the above analysis which should be prioritised for advertising and promotion are:

**1. "Are You Experienced?" (Rock)**

* Album ID: 120
* Total Genre Sales: 27.72
* Ranking: 1

**2. "From The Muddy Banks Of The Wishkah [live]" (Rock)**

* Album ID: 163
* Total Genre Sales: 27.72
* Ranking: 1

**3. "The Doors" (Rock)**

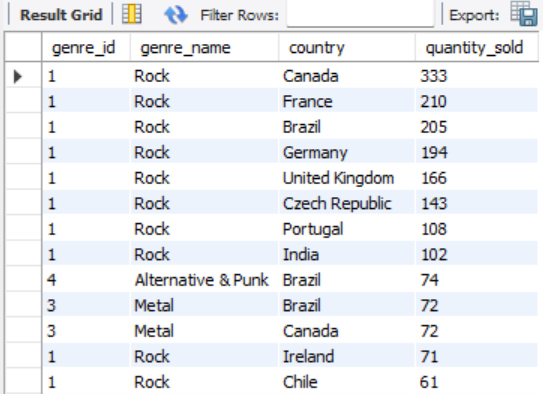
* Album ID: 214
* Total Genre Sales: 26.73
* Ranking: 2

**Insights**

Rock is the popular genre with top 3 albums i.e. the highest record of total sales which should promote more advertising in USA

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





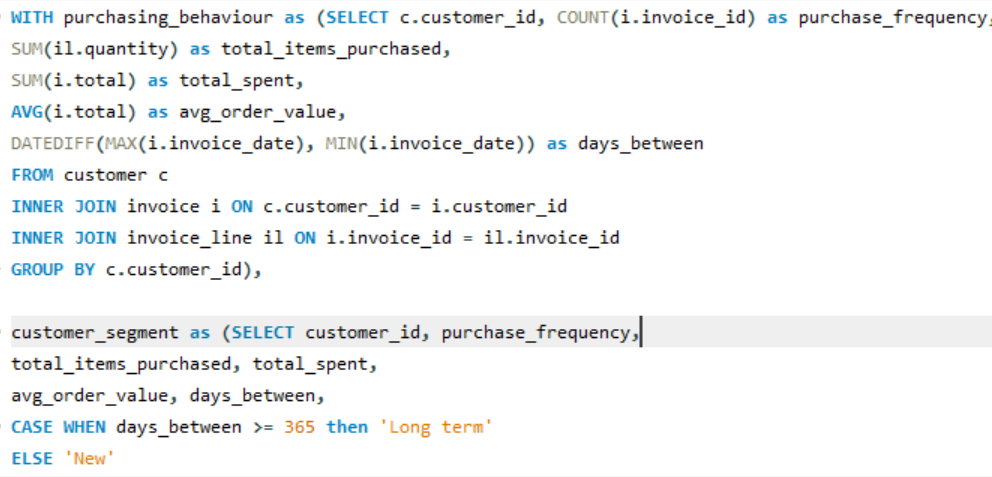
**Commonalities:**

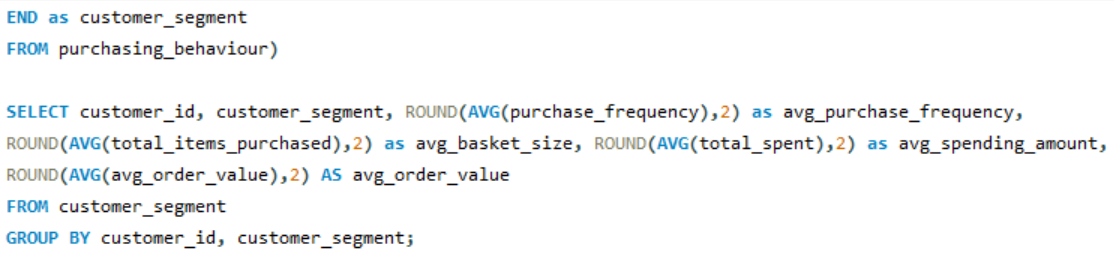
* Rock is the top genre in almost all the countries listed, indicating a strong global preference for Rock music outside the USA.
* Metal and Alternative & Punk are also popular in Brazil, showing a more diverse taste in music genres in that country.
* Rock is the leading genre in every country mentioned except where Metal or Alternative & Punk stand out.

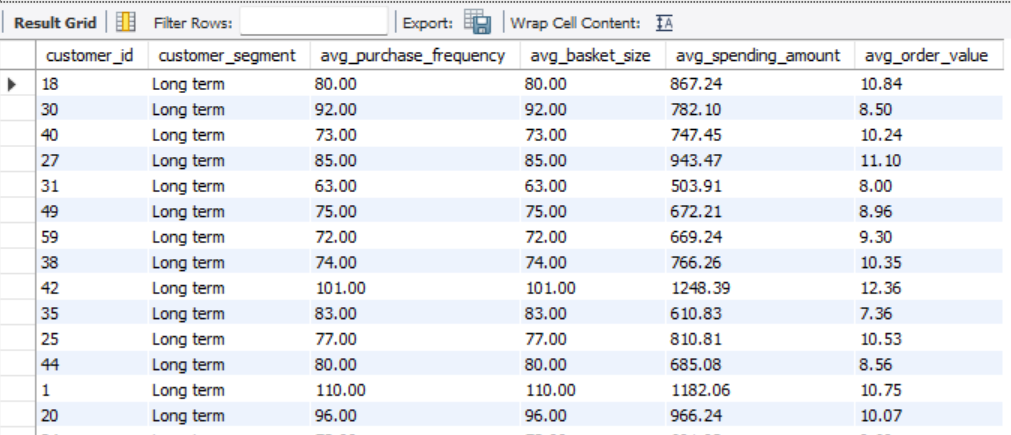
**Differences:**

* Brazil stands out with significant sales in both Alternative & Punk and Metal genres, in addition to Rock, showing a more varied musical taste compared to other countries.
* Metal is prominent in Canada and Brazil but not in other countries listed.
* The sales volume for genres other than Rock is generally lower in all non-USA countries, it clearly leads in popularity across all non-USA countries

**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?







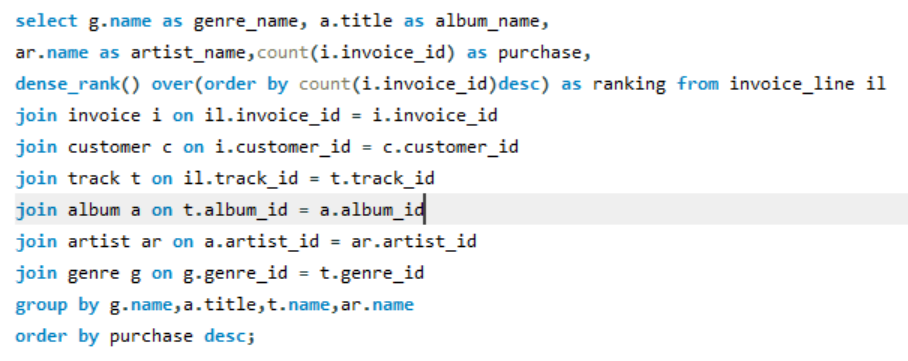
**Insights:**

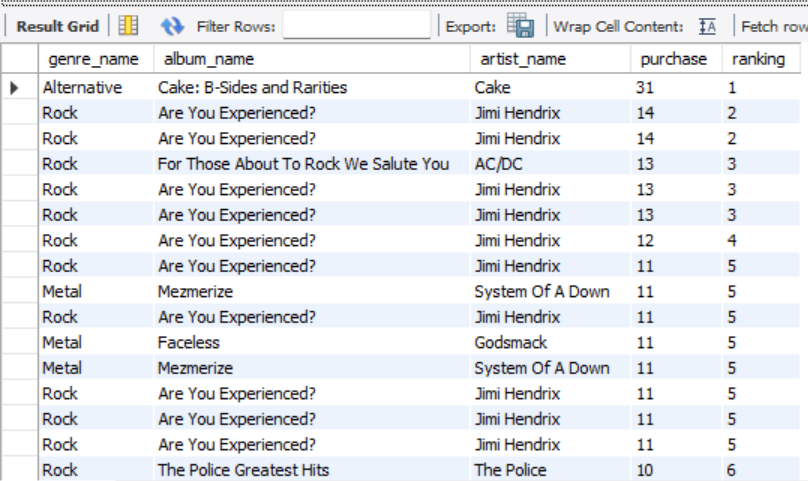
* **Purchase Frequency**: We can check if long-term customers buy more often than new customers.
* **Basket Size**: We can see if long-term customers generally buy more items in each order compared to new customers.
* **Spending Amount**: We can find out if long-term customers spend more money overall, which might show they are more loyal and satisfied.
* **Order Value**: We can determine if long-term customers make higher-value purchases, indicating they trust and prefer our products.

**Recommendations:**

These insights will help with customer loyalty and retention by rewarding long-term customers special offers or targeted marketing campaigns to encourage repeat purchases from new customers.

**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?





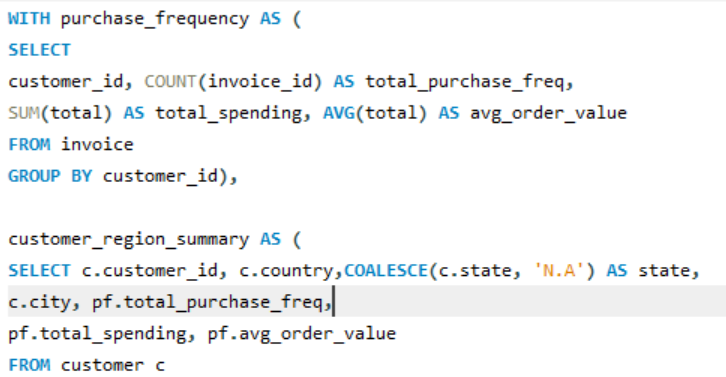
**Insights:**

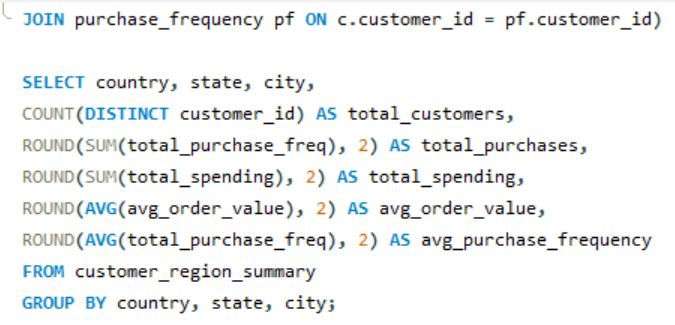
* Genre: Rock is the dominant genre, Metal is the second most frequent and Alternative appears once but has the highest number of purchases for a single album.
* Album: Cake's "B-Sides and Rarities" has the highest number of purchase and "Are You Experienced?" shows multiple purchase, indicating sustained popularity.
* Artist: Jimi Hendrix is the popular, with other artists like Cake, AC/DC, System of a Down, Godsmack, The Police.

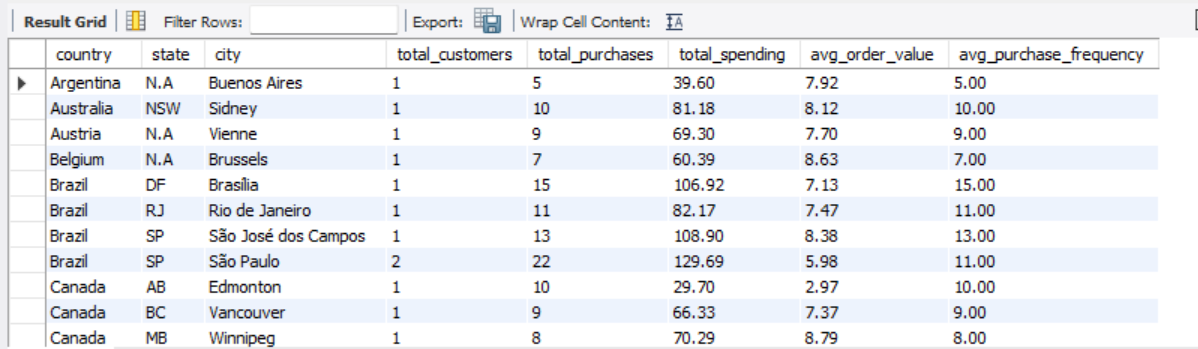
We can recommend these music suggestions to the customers in order to enhance the cross-selling initiatives

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

-- -- Customer Purchasing Behaviors by Region -- --

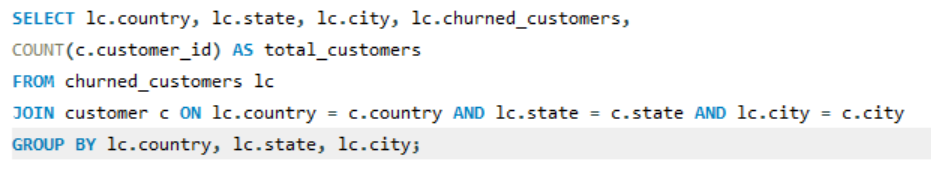


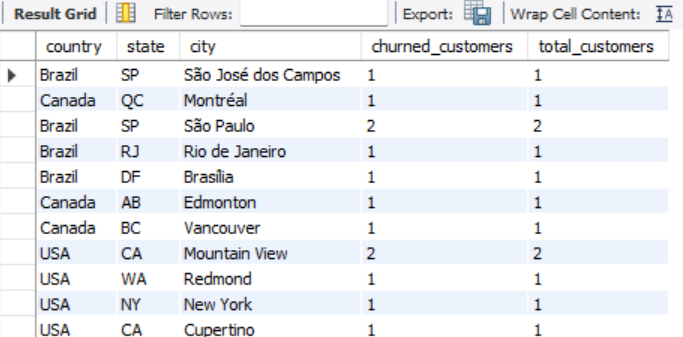




-- -- Churn Rate by Region -- --







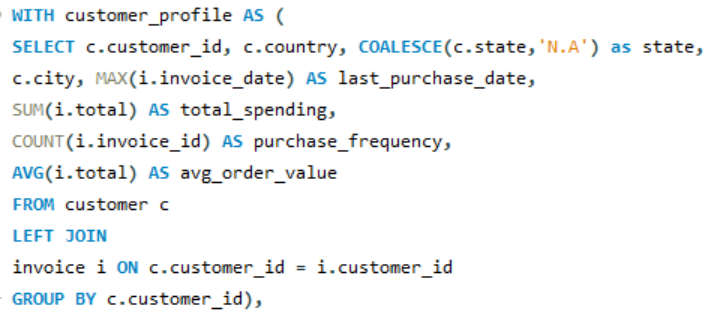
**Correlation with Local Factors:**

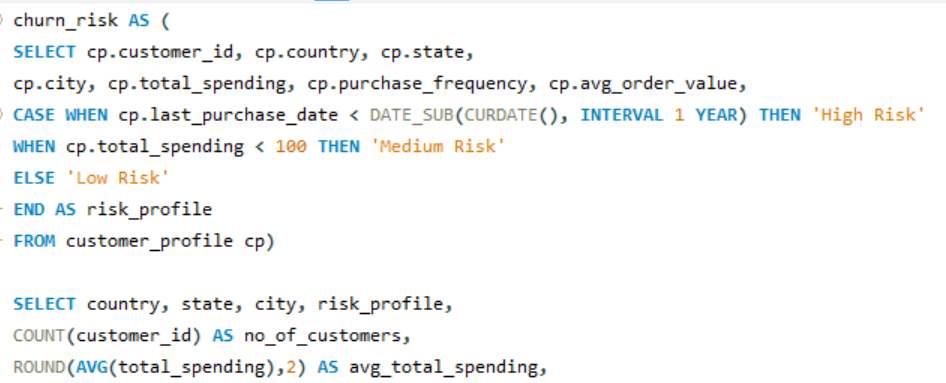
* **Affluent Areas:** High spending and frequency suggest higher income and better economic conditions.
* **Emerging Markets:** Moderate spending with high frequency points to budget-conscious middle-class consumers.
* **High Churn in Competitive Areas:** Competition and high living costs may affect loyalty.

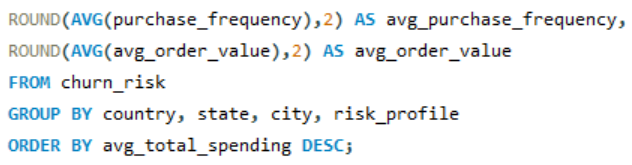
**Recommendations:**

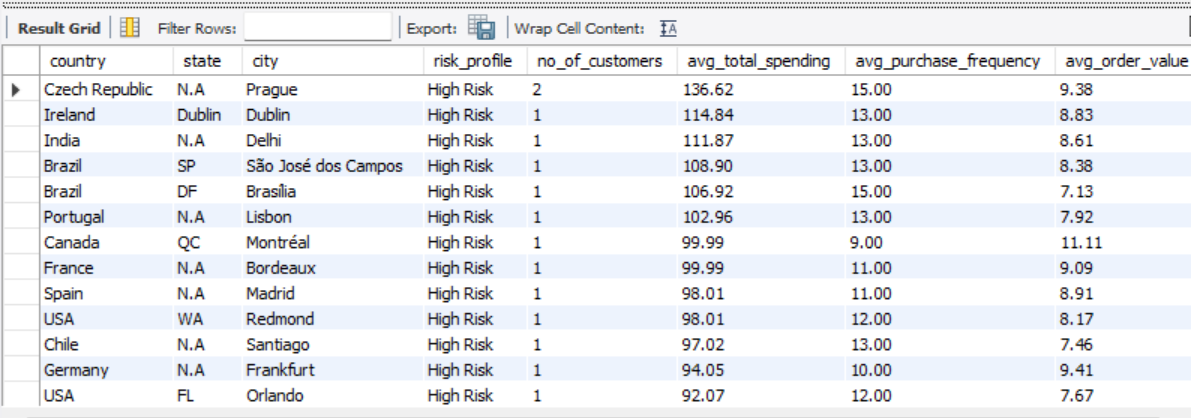
* **High-Value Areas:** Offer exclusive products and premium pricing.
* **Moderate Areas:** Enhance engagement with loyalty programs and value offers.
* **High Churn Areas:** Address churn with discounts, improved service, and feedback collection.

**6.**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?









**Insights**

1. **High Spending with High Frequency (High Churn Risk):**

* Regions: Prague, Czech Republic; Delhi, India; Brasília, Brazil.
* Risk Factors: High expectations and increased competition may lead to churn if service or deals decline.

1. **Moderate Spending with Moderate Frequency:**

* Regions: Mountain View, USA; Stuttgart, Germany.
* Risk Factors: Changes in the economy or satisfaction levels can reduce spending or increase churn.

1. **Low Spending with High Frequency:**

* Regions: Yellowknife, Canada; Stockholm, Sweden.
* Risk Factors: Customers looking for good deals are sensitive to price increases or reduced promotions.

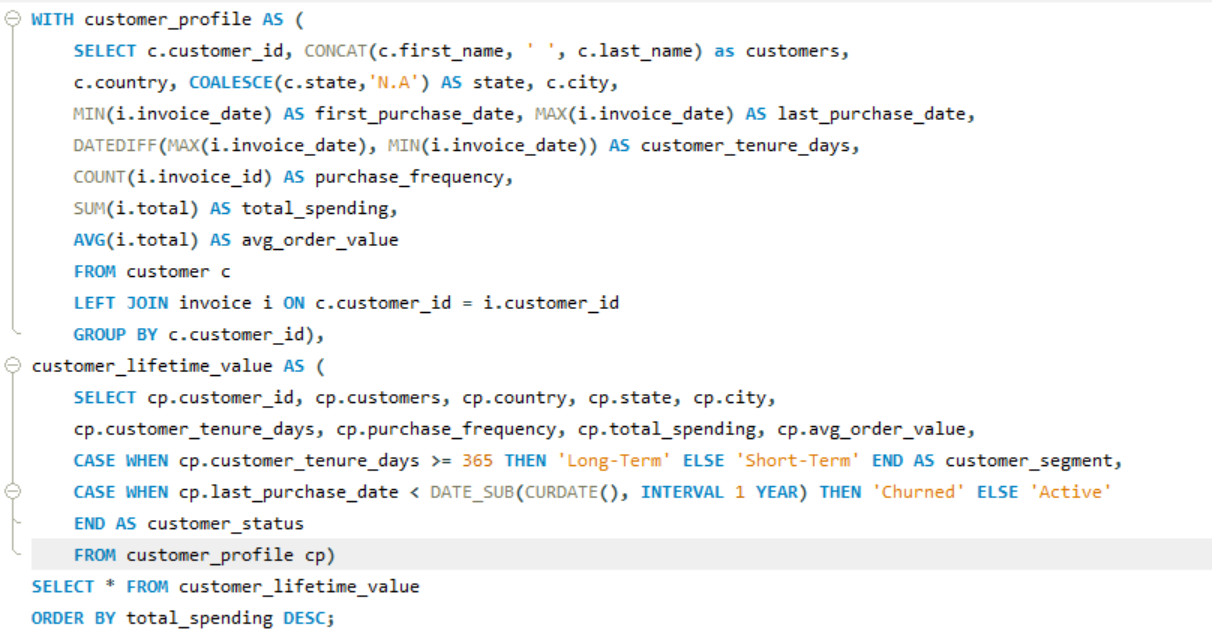
1. **Low Spending with Low Frequency (High Churn Risk):**

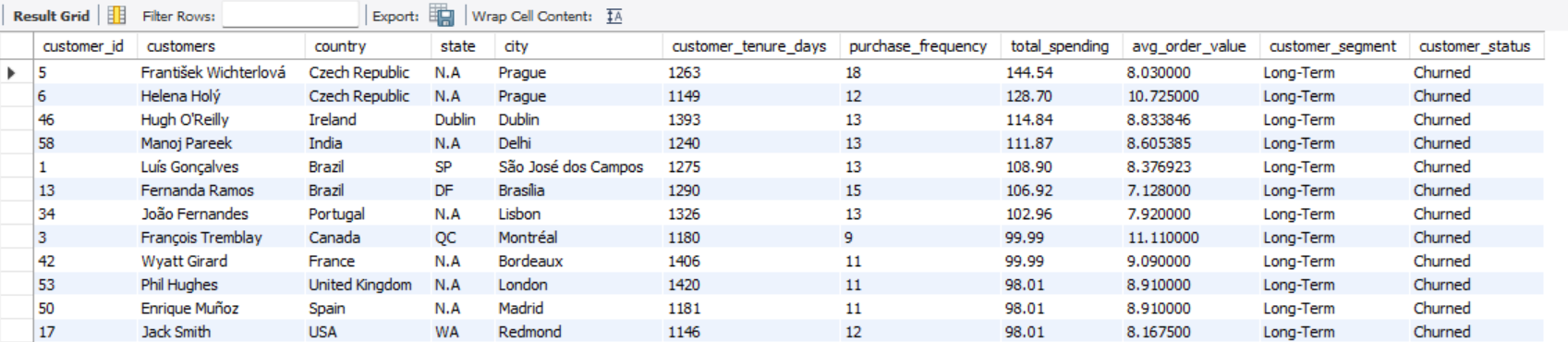
* Regions: Edmonton, Canada; Copenhagen, Denmark.
* Risk Factors: Customers with low engagement and high price sensitivity are at risk of churning if offerings aren't attractive or prices rise.

**Factors Contributing to High Risk:**

Economic instability, competitive marketplaces, declining service/product quality, and unmet demographic needs.

**7.**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:**

* Analyse the different types of customer behaviour by their tenure days, purchase frequency, total spend and average spend.

* Customer tenure helps in identify whether customers are new or have been around for a while.
* By customer status we came to know which customers have stopped purchasing and which are still engaged.
* Purchase Frequency shows how often customers make purchases.
* Total Spending is the total revenue generated by each customer. This metric is crucial for understanding customer value.
* Identify the spending behavior of customer by average order value

**Observed Purchase Patterns:**

* Higher Engagement: Frequent purchases reduce churn risk.
* Lower Order Value: Customers with low average order values but long tenure may be at risk if don’t meet customer expectations and needs.

**Insights:**

* High Lifetime value customers: Example: František Wichterlová from Prague with high spending and average order value. Need targeted retention strategies.
* Moderate Lifetime value customers: Example: Jack Smith from Redmond. Reward to increase engagement and loyalty programs.
* Low Lifetime value customers: Example: Mark Philips from Edmonton with lower spending and order values. More price-sensitive or less engaged.
* Churn can occur even with long tenure. Example: Helena Holý from Prague.

8.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:**

**Email Marketing:** Send an email campaign on a specific date e.g. on festive season and track how many people open to know the email opening rate and compare it with the purchase.

**Events:** Organize music events in cities where the artists live and compare sales data from before and during the event to see if there's an increase in purchases in those cities.

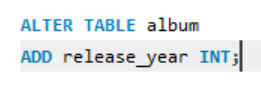
**Customized Discounts:** Offer discounts during specific events and then track sales before, during, and after the discount period to see how effective the discounts were.

**Client Retention:** Promotions and engagement strategies in order to retain clients and monitor retention rates and customer feedback to measure success.

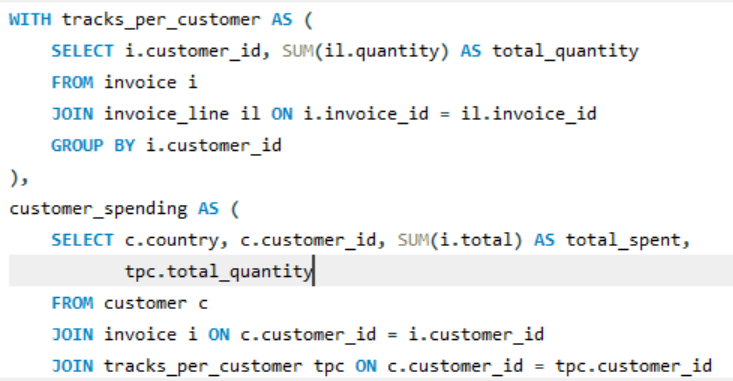
**9.**How would you approach this problem, if the objective and subjective questions weren't given?

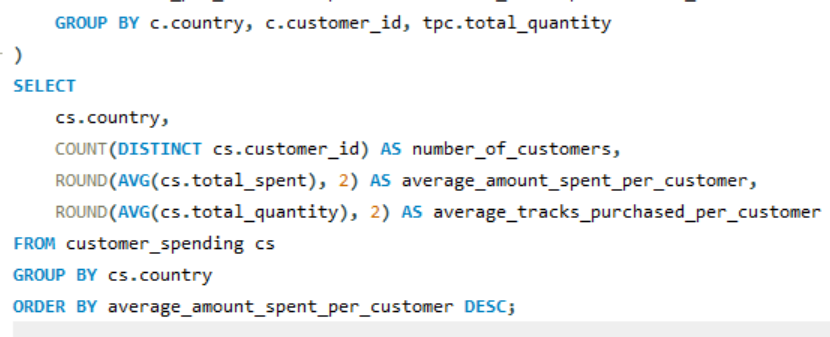
* Identify entities and relationships (e.g., Employee, Customer, Invoice).
* Determine key metrics like sales performance and customer demographics.
* Look for missing values or duplicates and clean the data as needed.
* Analyse total sales, average order value, and top-selling products by time, location, and customer group.
* Examine customers purchase history and churn rates
* Analyze sales across different location and link these sales to demographics and also find popular product combinations to use for cross-selling.
* Review promotions effect on sales and customer behaviour.
* Identify growth opportunities, optimize pricing, and create targeted marketing strategies.

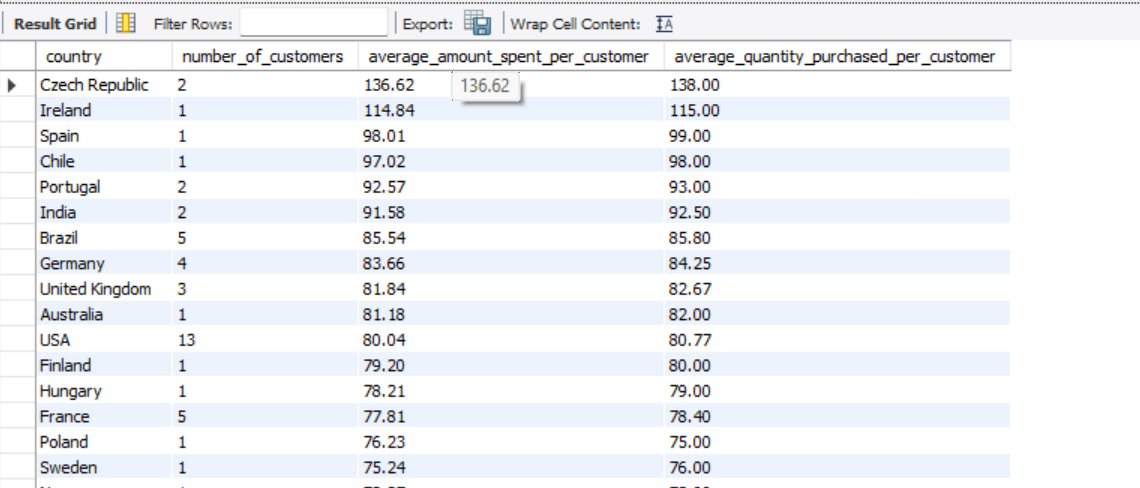
**10.**How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?



**11.**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.







**Insights:**

The maximum number of customers are from the countries like the USA, Germany, France, Canada and Brazil with the average amount spent per customer and average quantity purchased per customer.