**Objective Questions**

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

**Ans :**

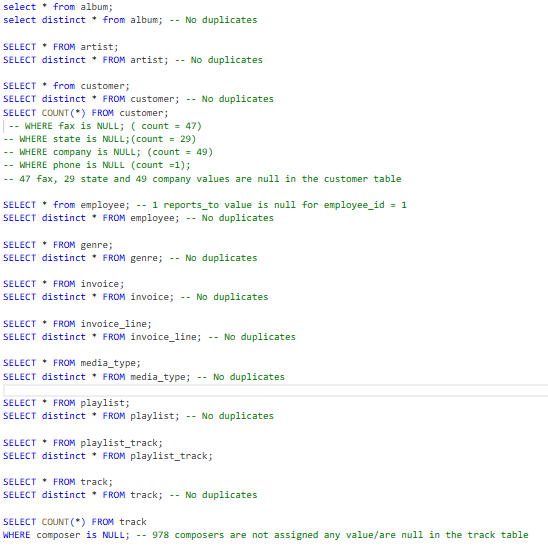
The data provided possess 0 duplicates although there are missing values in 3 tables.

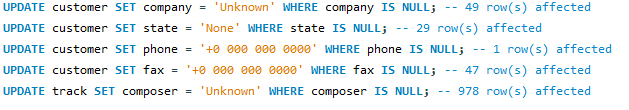
**Customer table**: 47 fax, 29 state and 49 company values are null in the customer table

**Employee table**: 1 reports to value is null for employee\_id = 1

**Track table**: 978 composers are not assigned any value/are null in the track table

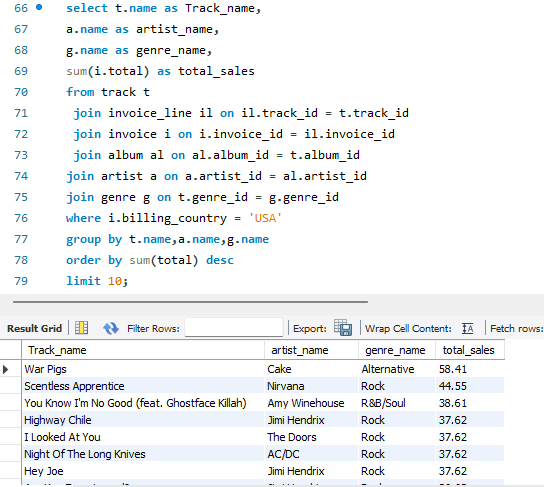
These Null values could be handled easily by using the Coalesce function or UPDATE command in SQL.

****

****

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

**Ans :**

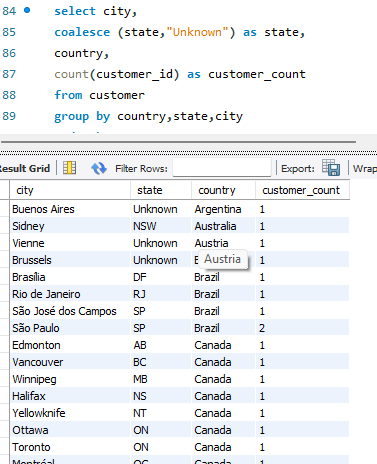
****

Hence, The Top Genre in USA is Rock.

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

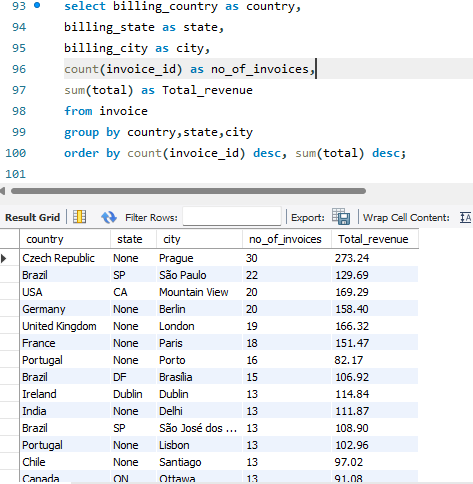
**Ans :** The customer demographic breakdown based on location is highly diversified.

* Chinook's customer base spans 24 countries, with the highest number of customers from the USA.
* The customer table lacks age and gender columns, which makes it difficult to analyze the customer demographic breakdown in detail.

****

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

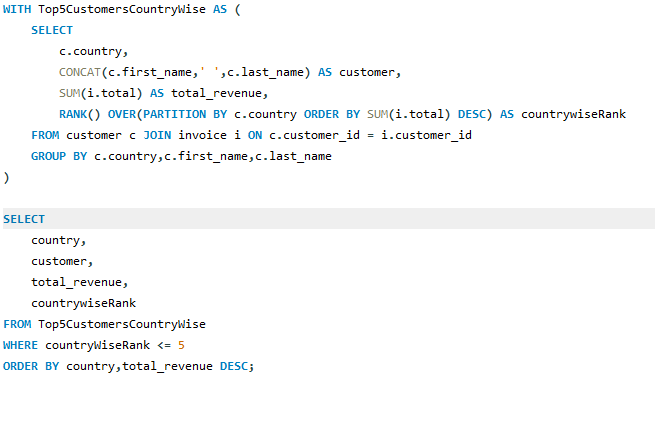
**Ans :**

****

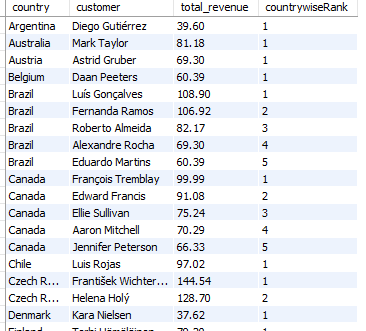
1. **Find the top 5 customers by total revenue in each country**

**Ans :**

**SQL query :**

****

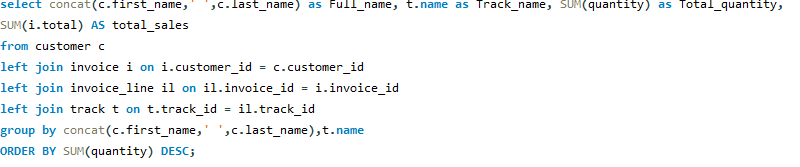
**Output :**



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

**Ans :**

**SQL Query**

****

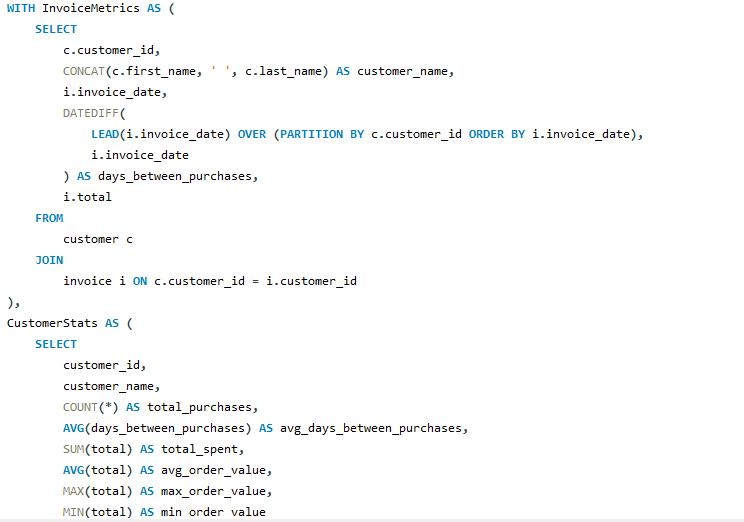
Output :



\*The customers purchased a similar quantity of different tracks, customer.

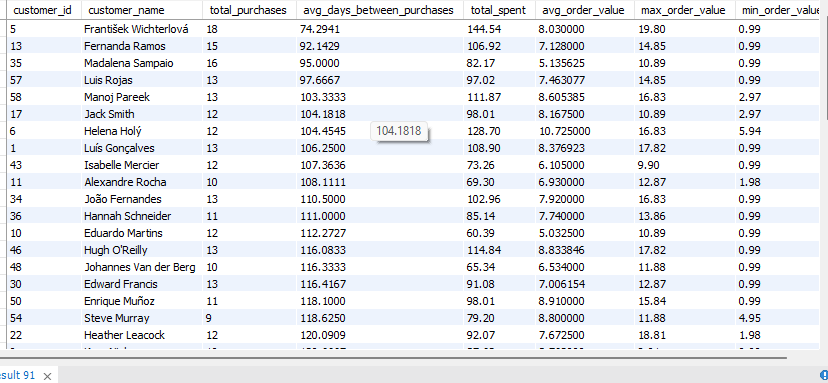
1. **Are there any patterns or trends in customer purchasing behaviour (e.g., frequency of purchases, preferred payment methods, average order value)?**

**Ans :**  **SQL Query :**





**Output :**



The table provides insights into customer behaviour, showing purchase frequency, total spending, and average order values. "Fernanda Ramos" and "František Wichterlová" are frequent buyers, with "František Wichterlová" also spending the most. Customers like "Luis Gonçalves" spend more per transaction, indicated by higher average order values. The avg\_days\_between\_purchases shows some customers have longer gaps between purchases, suggesting lower engagement. Customers like "Mano" have shorter gaps and higher purchase frequency. Focus on high-spending, frequent buyers, and use upselling or rewards to enhance loyalty.

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

**Ans:**

Churn Rate = (Number of customers lost during a period / Number of customers at the start of

the period) x 100

In this case, I have considered a customer to be churned if they have not made any purchase for >180 days between the last purchase date and the second last purchase date.

**SQL Query:**

WITH PreviousCustomerPurchases AS (

SELECT

c.customer\_id,c.first\_name,c.last\_name,DATE(i.invoice\_date) AS invoice\_date,

LEAD(DATE(i.invoice\_date)) OVER(PARTITION BY c.customer\_id ORDER BY invoice\_date DESC) AS prev\_purchase

FROM customer c

JOIN invoice i ON c.customer\_id = i.customer\_id

),

PrevPurchaseRank AS (

SELECT

\*,ROW\_NUMBER() OVER(PARTITION BY customer\_id ORDER BY prev\_purchase DESC) AS prev\_purchase\_rn

FROM PreviousCustomerPurchases

),

PreviousPurchaseDate AS (

SELECT

\*,DATEDIFF(invoice\_date,prev\_purchase) AS days\_since\_last\_purchase

FROM PrevPurchaseRank

WHERE prev\_purchase\_rn = 1

AND DATEDIFF(invoice\_date,prev\_purchase) > 180

ORDER BY days\_since\_last\_purchase DESC

)

SELECT

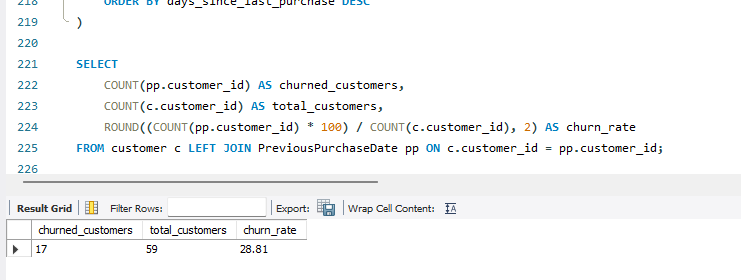
COUNT(pp.customer\_id) AS churned\_customers,

COUNT(c.customer\_id) AS total\_customers,

ROUND((COUNT(pp.customer\_id) \* 100) / COUNT(c.customer\_id), 2) AS churn\_rate

FROM customer c LEFT JOIN PreviousPurchaseDate pp ON c.customer\_id = pp.customer\_id;

**OUTPUT :**



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

**Ans:**

**1. Percentage of total sales contributed by each genre in the USA**

**SQL Query:**

WITH SalesGenreRankUSA AS (

SELECT

g.name AS genre, ar.name AS artist, SUM(i.total) AS genre\_sales,

DENSE\_RANK() OVER( PARTITION BY g.name ORDER BY SUM(i.total) DESC) AS genre\_rank

FROM genre g

LEFT JOIN track t ON g.genre\_id = t.genre\_id

LEFT JOIN invoice\_line il ON t.track\_id = il.track\_id

LEFT JOIN invoice i ON il.invoice\_id = i.invoice\_id

LEFT JOIN album a ON t.album\_id = a.album\_id

LEFT JOIN artist ar ON a.artist\_id = ar.artist\_id

WHERE i.billing\_country = 'USA'

GROUP BY 1,2

),

TotalSalesUSA AS (

SELECT

SUM(i.total) AS total\_sales

FROM invoice\_line il

LEFT JOIN invoice i ON il.invoice\_id = i.invoice\_id

WHERE i.billing\_country = 'USA'

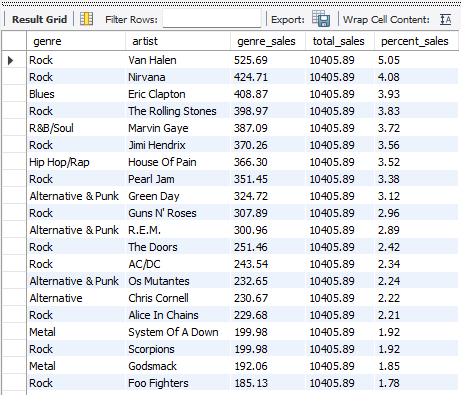
)

SELECT s.genre,s.artist,s.genre\_sales,t.total\_sales, ROUND((s.genre\_sales / t.total\_sales)\* 100,2) AS percent\_sales

FROM SalesGenreRankUSA s JOIN TotalSalesUSA t

ORDER BY s.genre\_sales DESC, s.genre ASC;

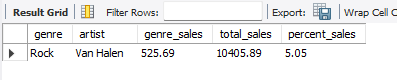
**OUTPUT** :



**2. Best Selling Genre and Artist**

To identify the best-selling genre and artist, we have to include LIMIT 1 at the end of order by which is ORDER BY s.genre\_sales DESC, s.genre ASC LIMIT 1; .We will get the following result:

**OUTPUT :**



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

**Ans :**

**SQL Query :**

SELECT

c.customer\_id,

CONCAT(c.first\_name,' ',c.last\_name) AS customer,

COUNT(DISTINCT t.genre\_id) AS genre\_count,

COUNT(DISTINCT t.track\_id) AS track\_count

FROM customer c

JOIN invoice i ON c.customer\_id = i.customer\_id

JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id

JOIN track t ON il.track\_id = t.track\_id

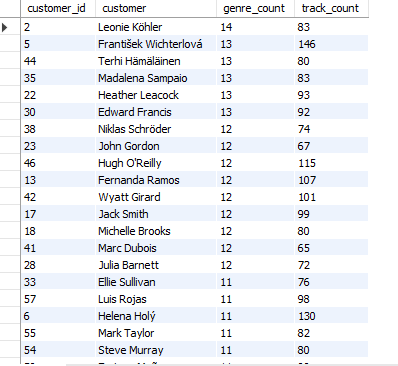
JOIN genre g ON t.genre\_id = g.genre\_id

GROUP BY c.customer\_id,c.first\_name,c.last\_name

HAVING COUNT(DISTINCT g.genre\_id) >=3

ORDER BY genre\_count DESC;

**Output :**



Leonie Köhler is a unique customer who has purchased tracks spanning 14 different genres, showcasing a diverse taste in music.

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

**Ans:**

**SQL Query :**

WITH SalesWiseGenreRank AS (

SELECT

g.name AS genre,

SUM(i.total) AS total\_sales,

DENSE\_RANK() OVER(ORDER BY SUM(i.total) DESC) AS genre\_rank

FROM genre g

LEFT JOIN track t ON g.genre\_id = t.genre\_id

LEFT JOIN invoice\_line il ON t.track\_id = il.track\_id

LEFT JOIN invoice i ON il.invoice\_id = i.invoice\_id

WHERE i.billing\_country = 'USA'

GROUP BY g.name

)

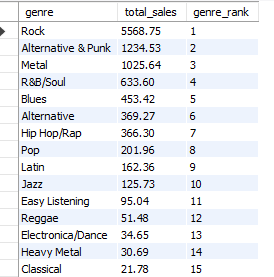
SELECT

genre,total\_sales,genre\_rank

FROM SalesWiseGenreRank

ORDER BY genre\_rank;

**OUTPUT :**



The sales performance of various music genres in the USA reveals a clear dominance of **Rock**, followed by **Alternative & Punk** and **Metal**. Genres like **R&B/Soul**, **Blues**, and **Alternative** also perform moderately well, but their sales are significantly lower than Rock. Overall, the rankings demonstrate a clear divide in genre popularity, with Rock far outperforming others.

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

**Ans :**

**SQL Query :**

WITH CustomerLastPurchase AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

MIN(DATE(i.invoice\_date)) AS first\_purchase\_date,

MAX(DATE(i.invoice\_date)) AS last\_purchase\_date

FROM customer c

JOIN invoice i ON c.customer\_id = i.customer\_id

GROUP BY c.customer\_id, c.first\_name, c.last\_name

),

CustomerPurchases AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

DATE(i.invoice\_date) AS invoice\_date

FROM customer c

JOIN invoice i ON c.customer\_id = i.customer\_id

)

SELECT

clp.customer\_id,

clp.first\_name,

clp.last\_name,

clp.first\_purchase\_date,

clp.last\_purchase\_date

FROM CustomerLastPurchase clp

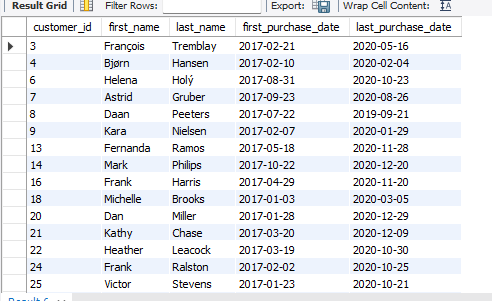
LEFT JOIN CustomerPurchases cp ON clp.customer\_id = cp.customer\_id

AND cp.invoice\_date BETWEEN clp.last\_purchase\_date - INTERVAL 3 MONTH AND clp.last\_purchase\_date - INTERVAL 1 DAY

WHERE cp.invoice\_date IS NULL

ORDER BY clp.customer\_id;

**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.**

**Ans :**

**SQL Query:**

WITH RecommendedAlbums AS (

SELECT

al.title AS album\_name,

a.name AS artist\_name,

g.name AS genre\_name,

SUM(i.total) AS total\_sales,

SUM(il.quantity) AS total\_quantity,

ROW\_NUMBER() OVER(ORDER BY SUM(i.total) DESC) AS sales\_rank

FROM customer c

JOIN invoice i ON c.customer\_id = i.customer\_id

JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id

JOIN track t ON il.track\_id = t.track\_id

JOIN album al ON t.album\_id = al.album\_id

JOIN artist a ON al.artist\_id = a.artist\_id

JOIN genre g ON t.genre\_id = g.genre\_id

WHERE c.country = 'USA'

GROUP BY al.title,a.name,g.name

)

SELECT \* FROM RecommendedAlbums

ORDER BY total\_sales DESC

limit 3;

**OUTPUT :**



**Recommendation:**

The top 3 albums that should be prioritised for advertisements and promotions in the USA based on genre analysis are:-

1. **Seek And Shall Find: More of the Best** (By Marvin Gayne)
2. **From The Muddy Banks Of The Wishkah** (By Nirvana)
3. **Are You Experienced?** (By Jimi Hendrix)

These albums are from the genre R&B/Soul and ROCK respectively as it is the most popular genre in the USA.

**Visualization :**

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

**Ans :**

**SQL Query:**

SELECT

g.genre\_id,

g.name,

SUM(t.unit\_price \* il.quantity) AS total\_revenue\_for\_genre,

SUM(il.quantity) AS total\_tracks\_sold,

COUNT(DISTINCT i.invoice\_id) AS total\_invoices

FROM

track t

LEFT JOIN

genre g ON g.genre\_id = t.genre\_id

LEFT JOIN

invoice\_line il ON il.track\_id = t.track\_id

LEFT JOIN

invoice i ON i.invoice\_id = il.invoice\_id

WHERE

billing\_country != 'USA'

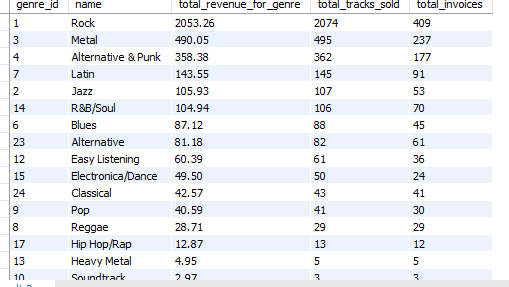
GROUP BY

g.genre\_id, g.name

ORDER BY

total\_revenue\_for\_genre DESC;

**OUTPUT :**



We can observe some similarities in terms of famous genres in countries other than USA vs USA.

* In both the cases, we can observe that Rock is the most famous Genre.
* The 2nd and 3rd places are interchanged in case of countries other than USA. Metal is the second popular genre in countries other than USA, followed by Alternative & Punk.

**Visualization :**

1. **Customer Purchasing Behaviour 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?**

**Ans :**

**SQL Query :**

WITH cte as

(

SELECT i.customer\_id,

MAX(invoice\_date), MIN(invoice\_date),

abs(TIMESTAMPDIFF(MONTH, MAX(invoice\_date),

MIN(invoice\_date))) time\_for\_each\_customer,

SUM(total) sales, SUM(quantity) items,

COUNT(invoice\_date) frequency FROM invoice i

LEFT JOIN customer c on c.customer\_id = i.customer\_id

LEFT JOIN invoice\_line il on il.invoice\_id = i.invoice\_id

GROUP BY 1

ORDER BY time\_for\_each\_customer DESC),

average\_time as (

SELECT AVG(time\_for\_each\_customer) average FROM cte

),

categorization as

(

SELECT \*,

CASE

WHEN time\_for\_each\_customer > (SELECT average from average\_time) THEN "Long-term Customer" ELSE "Short-term Customer"

END category

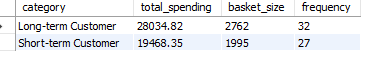
FROM cte

)

SELECT category, SUM(sales) total\_spending, SUM(items) basket\_size, COUNT(frequency) frequency FROM categorization

GROUP BY 1

**OUTPUT :**

****

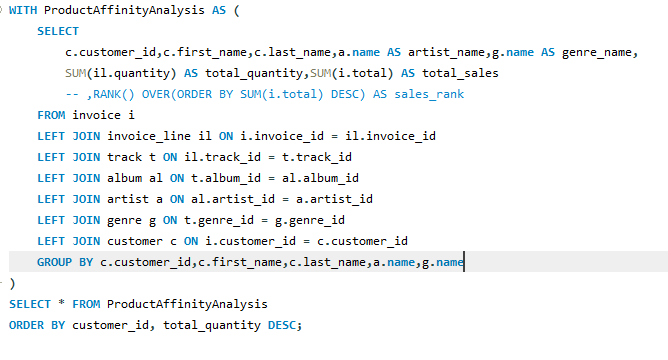
**Visualization :**

**Insights**: - It can be observed that long-term customers tend to have higher spending amounts and larger basket sizes compared to short-term customers.

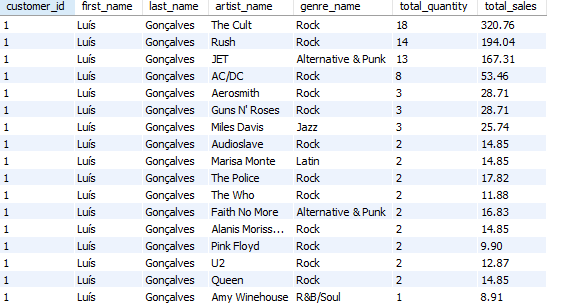
**Recommendations:** - This indicates that customer loyalty is crucial for boosting company revenue, as long-term customers tend to make more purchases than short-term customers. Therefore, the company should prioritize improving customer retention to drive higher sales over time.

1. **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?**

**Ans: SQL Query :**

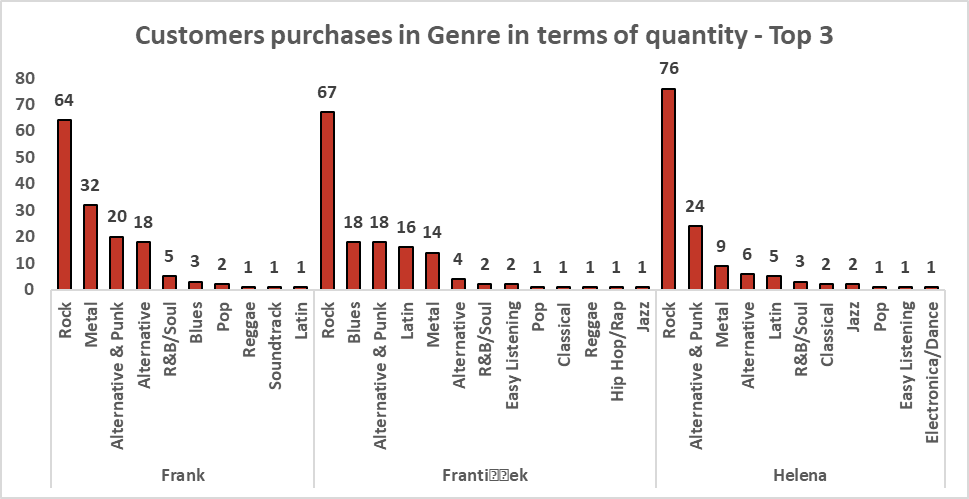
****

**OUTPUT :**

****

* Common Genres:
* All three considered customers favor Rock and Metal genres, indicating a shared preference for these types of music.
* Product Recommendations:
* Suggesting related genres (such as Blues or Alternative) or introducing musicians inside Rock and Metal may work well for clients like Frank who buy a lot of this genre.
* Cross-Selling Initiatives:
* By using their preexisting interests to promote wider musical discovery, expose clients who exhibit a high level of commitment to particular genres to related genres through carefully chosen recommendations.
* Since rock, metal, and alternative and punk music are always in style, make customized playlists for every client that feature their best songs.

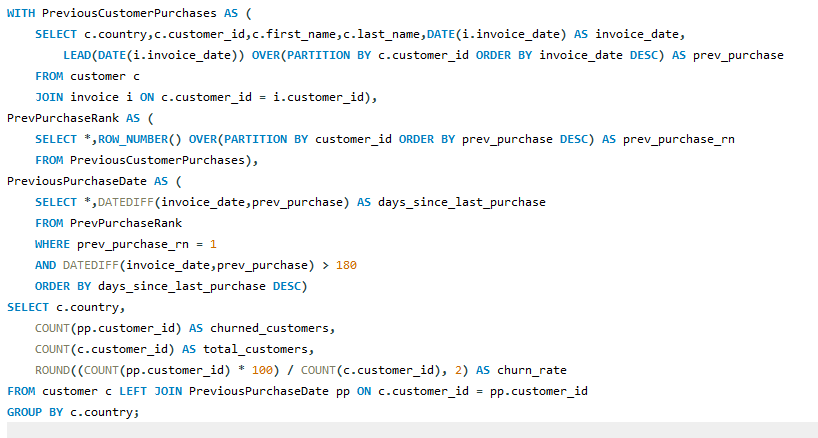
**Visualization :**

****

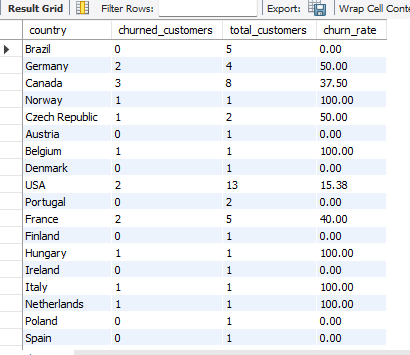
1. **Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?**

**Ans:**

**SQL Query :**



**OUTPUT:**

****

Based on the customer churn, the regional market analysis is performed which gives the

following insights:

• A customer is considered to be churned if they have not made any purchases in the last

6 months.

• We can see that some countries like Finland, Australia, India, Spain etc. have 0 churn

rate which indicates that the customers in these regions are active and make frequent

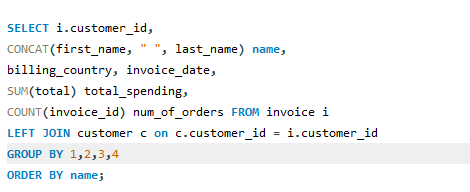
purchases.

**Visualization :**

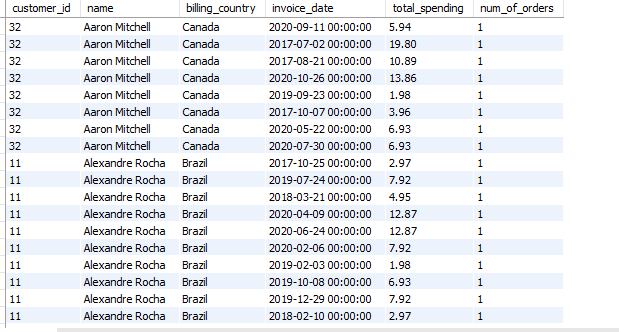
1. **Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?**

**Ans:**

**SQL Query :**

****

**OUTPUT :**

****

After analyzing the data through charts and tables, it is evident that countries with already high spending and order frequency are showing growth, while sales and frequency remain stagnant in other regions. This highlights the need for new promotional campaigns in underperforming countries to reduce the churn rate and sustain or increase spending levels.

Factors contributing to customer churn risk include:

* **Age**: Are younger customers more likely to churn? (If age data were available, it could provide insights.)
* **Gender**: Does gender influence churn rates? (Gender information could reveal trends if provided.)
* **Location**: How does geographical location impact churn? Regional differences may indicate varying engagement levels.
* **Spending Behavior**: High spenders versus infrequent buyers—does spending behavior correlate with churn risk?

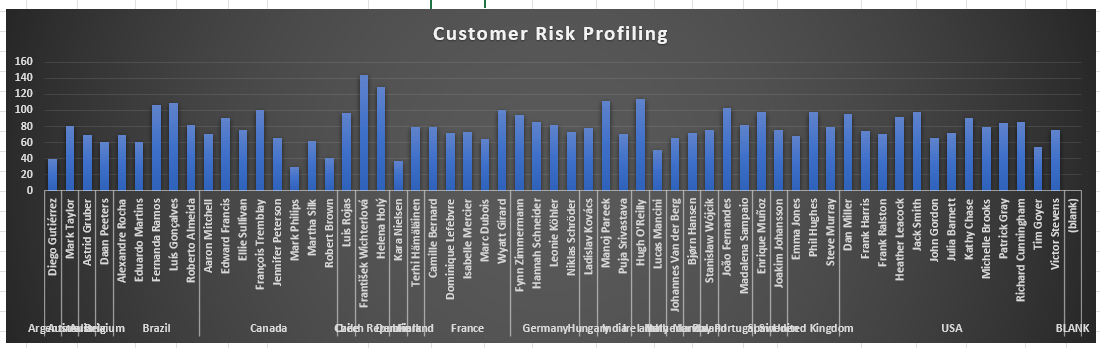
Addressing these factors through targeted strategies could help mitigate churn effectively.

If age and gender data were available, customer segmentation could be more detailed, such as:

* **Young-Male-High-Spenders**
* **Young-Female-High-Spenders**
* **Old-Male-Low-Spenders**
* **Old-Female-Low-Spenders**
* And additional combinations reflecting varying spending patterns and demographics.

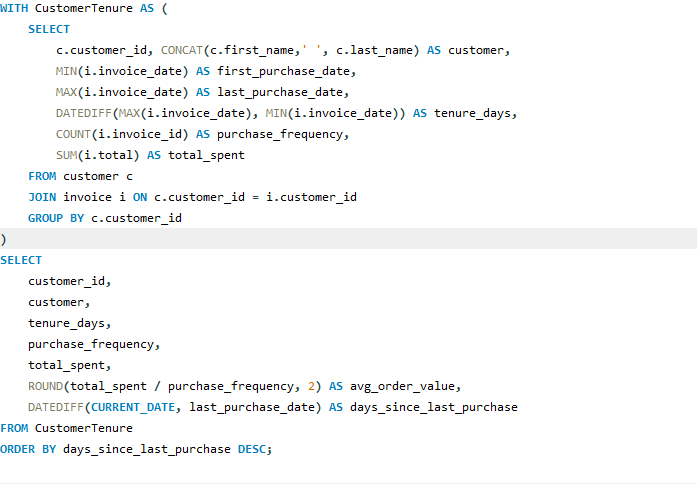
This segmentation would enable more targeted marketing and personalized strategies to optimize customer engagement and retention.

**Visualization :**

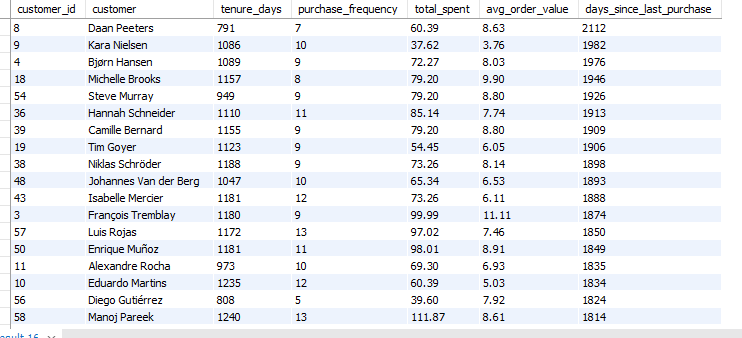
****

1. **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?**

**Ans : SQL Query:**

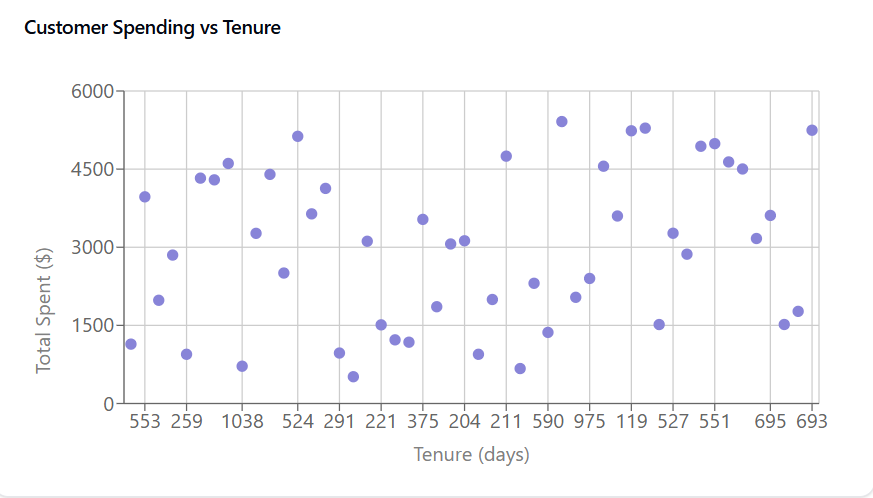
****

**OUTPUT :**

****

1. To predict the lifetime value of a customer segment, analyzing purchase history and customer tenure is essential. These parameters help identify High-Value Customers who significantly contribute to revenue.
2. Customers with short tenure but high purchase activity should be targeted with promotional campaigns to convert them into loyal, long-term customers.
3. Customers who have stopped purchasing often belong to underdeveloped or developing countries, highlighting the influence of economic factors on sales. To address this, companies should leverage promotional channels like social media, articles, and advertisements, while offering discounts to boost sales in these regions.

**Visualization :**



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?**

**Ans:** If data on promotional campaigns were available, I would analyze their impact on:

1. **Customer Acquisition**:
   * Tracking the increase in customer count over time.
   * Monitoring participation in events across different locations.
   * Evaluating the rise in click-through rates from email marketing campaigns.
2. **Customer Retention**:
   * Identifying the number of returning customers who attended events and purchased tracks after a long gap.
   * Assessing the impact of discounts on retaining existing customers.
3. **Sales**:
   * Measuring the sales growth attributed to promotional campaigns.
   * Analyzing trends to identify the most effective campaigns.
   * Tracking new customer acquisition driven by discounts.

This would provide valuable insights into the effectiveness of promotional strategies.

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

**Ans:**  If no specific questions were given, I would start by exploring the dataset broadly to uncover patterns and key insights relevant to customer behavior, sales performance, and promotional effectiveness. Here’s how I would approach it:

* Understand Business Objectives: First, I’d clarify business goals like increasing customer retention, boosting sales, or identifying high-value customer segments. This would help me

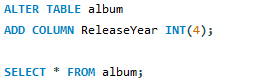
target meaningful insights.

* Data Exploration and Cleaning: I would perform data cleaning to handle any missing or inconsistent entries, then conduct exploratory analysis to understand the data’s structure, distribution, and trends.
* Identify Key Metrics and Segments: I’d establish KPIs such as customer acquisition rate, churn rate, lifetime value (LTV), average order value, and campaign ROI. Grouping by customer demographics, geographical location, and purchase behavior would reveal patterns within each segment.
* Perform Analysis and Modeling:
* Churn Analysis: I’d analyze factors that increase the likelihood of churn, such as recent purchase frequency or discount engagement.
* Campaign Effectiveness: I would examine promotional data by comparing pre- and post-campaign metrics to assess changes in acquisition, retention, and sales.
* Interpret Results for Strategy Recommendations: Based on findings, I’d make data driven recommendations to improve customer retention, target high-value customers, and refine marketing strategies for optimal ROI

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?**

**Ans:** We can use the ALTER TABLE command to add a new column to the Album table**.**

**SQL Query :**

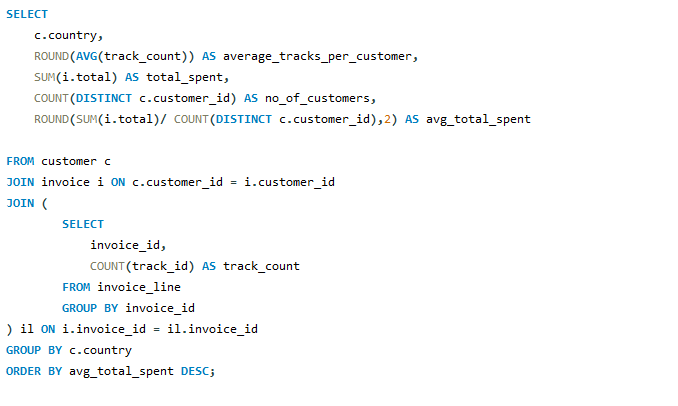
****

**OUTPUT :**

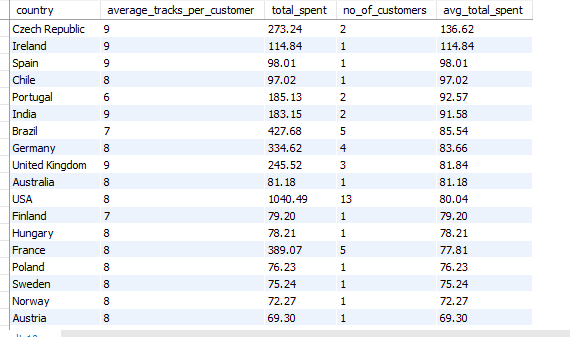
****

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 an SQL query to provide this information.**

**Ans: SQL Query :**

****

**OUTPUT :**

****

The table reveals customer data across countries, showing metrics like the average number of tracks purchased, total spending, and the number of customers. The Czech Republic, Ireland, and Spain have high track averages, while the USA has the largest customer count (13) but lower average spending. Total spending varies, with Brazil and Canada showing higher values, while Chile and Finland have smaller totals. The USA leads in total spending at $10,040.49. Countries like Belgium and Italy have consistent average spending per customer at around $60.39. This data highlights regional differences in purchase frequency and spending behavior.