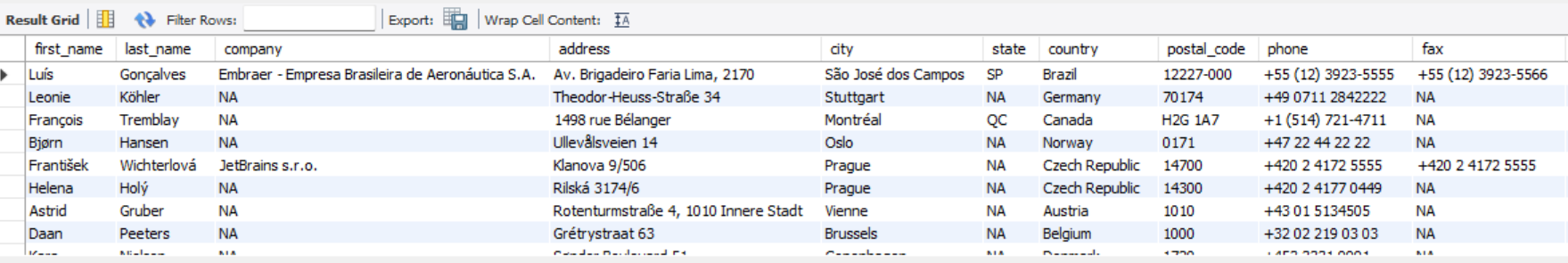
**Objective Question**

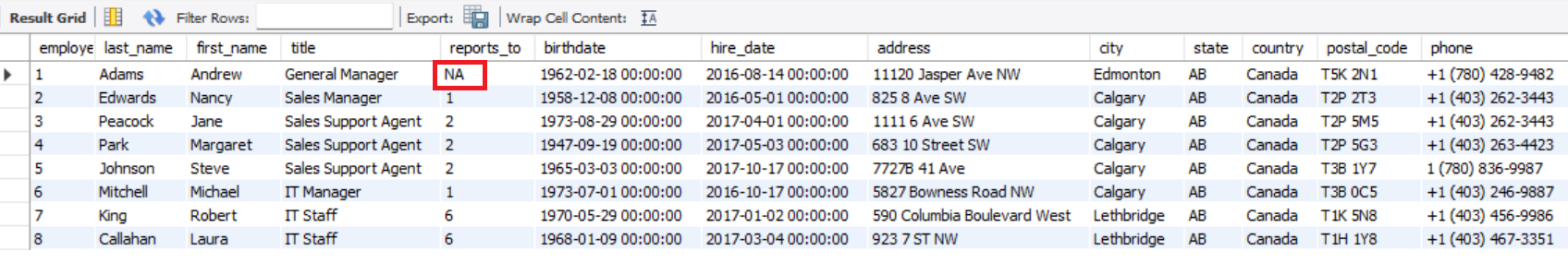
**1. Does any table have missing values or duplicates? If yes, how would you handle it?**

**Answer:**  No Duplicate values in tables.

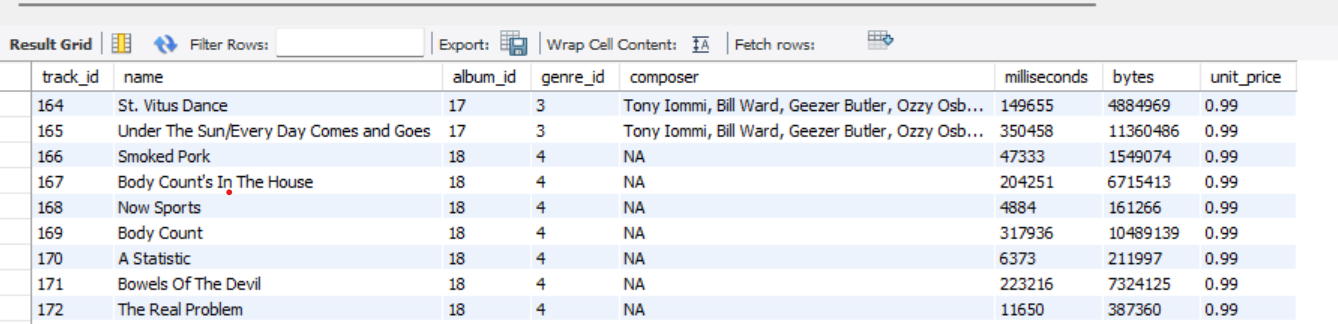
There are missing values in following tables. Query for replacing with NA has been written in SQL file.

1.**Customers** table contains 50 null values. They have been replaced with NA. 

2.**Employee** table has 1 null value. They have been replaced with NA



3.**Track** table has 502 null values. They have been replaced with NA



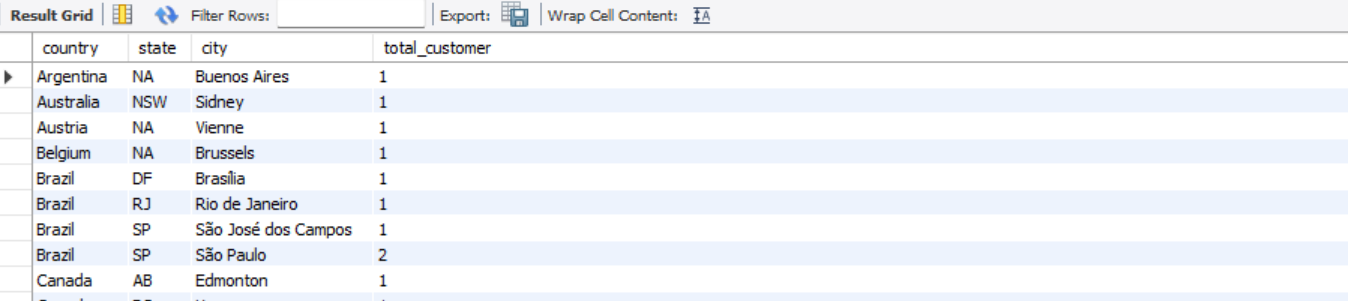
**2. Find the top-selling tracks and top artist in the USA and identify their most famous genres.**

**Answer:** The SQL query has been written in sql file, and the output is as follows



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

**Answer:** Age and Gender are not available in Customers table. Hence created only for location ( Country, state,city). Query has been written in SQL file.



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

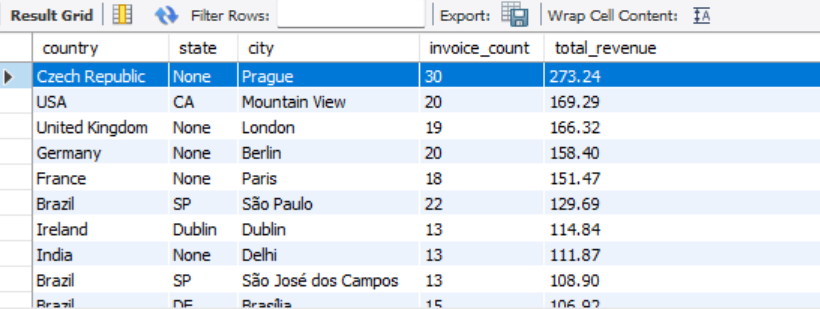
**Answer****: (Total Revenue = quantity \* unit price)**

Total Revenue is not directly available in any table hence it can be calculated using quantity and price and applied join on Invoice and Invoice\_line table.

Query given below:

SELECT i.billing\_country AS country, i.billing\_state AS state, i.billing\_city AS city, COUNT(DISTINCT i.invoice\_id) AS invoice\_count, SUM(il.unit\_price \* il.quantity) AS total\_revenue FROM invoice i JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id

GROUP BY i.billing\_country, i.billing\_state, i.billing\_city ORDER BY total\_revenue DESC;



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

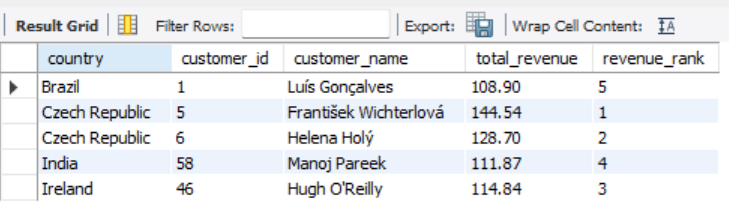
**Answer:** We have used Rank () function to find the top 5 customer by Total Revenue.

Query given below:

WITH customer\_revenue AS ( SELECT c.country, c.customer\_id, concat(c.first\_name," ",c.last\_name) AS customer\_name, SUM(il.unit\_price \* il.quantity) AS total\_revenue FROM customer c JOIN invoice i ON c.customer\_id = i.customer\_id JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id GROUP BY c.country, c.customer\_id,customer\_name ),

ranked\_customers AS ( SELECT \*, RANK() OVER (ORDER BY total\_revenue DESC) AS revenue\_rank FROM customer\_revenue )

SELECT \* FROM ranked\_customers WHERE revenue\_rank <= 5 ORDER BY country, revenue\_rank;



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

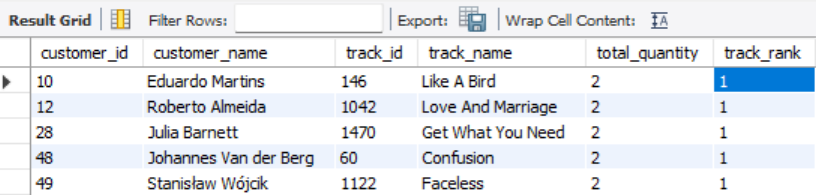
**Answer :** We have calculated total quantity purchased per track per customer and then used Rank() function to pick the top tracks per customer.

Query given below:

WITH customer\_track\_sales AS ( SELECT c.customer\_id, concat(c.first\_name," ",c.last\_name) AS customer\_name, t.track\_id, t.name AS track\_name, SUM(il.quantity) AS total\_quantity 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 GROUP BY c.customer\_id, c.first\_name, c.last\_name, t.track\_id, t.name ),

ranked\_tracks AS ( SELECT \*, RANK() OVER (ORDER BY total\_quantity DESC) AS track\_rank FROM customer\_track\_sales )

SELECT \* FROM ranked\_tracks WHERE track\_rank = 1 ORDER BY customer\_id;

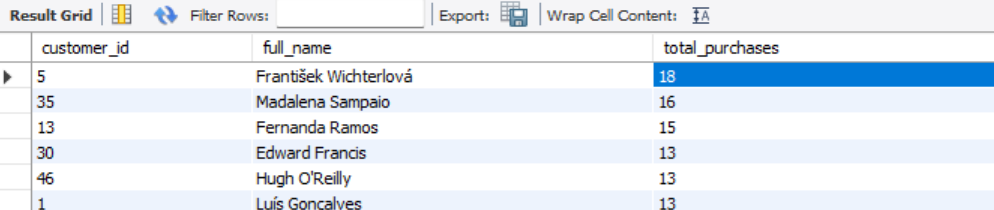


**7. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?**

**Answer:**

1. Purchase frequency is Number of invoices per customer

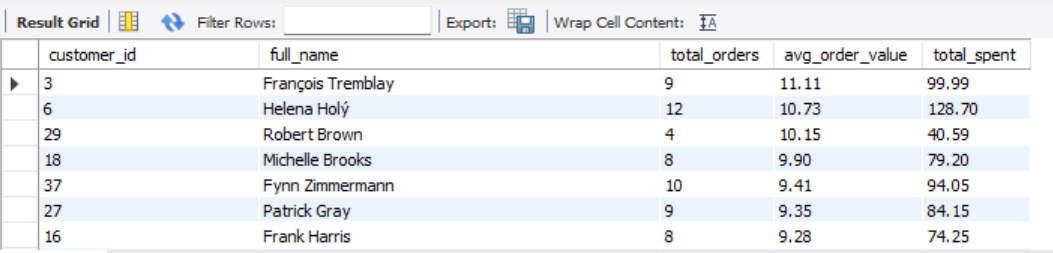
SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS full\_name, COUNT(i.invoice\_id) AS total\_purchases FROM customer c JOIN invoice i ON c.customer\_id = i.customer\_id GROUP BY c.customer\_id ORDER BY total\_purchases DESC;



2. Average Order Value helps identify customers who spend more per transaction

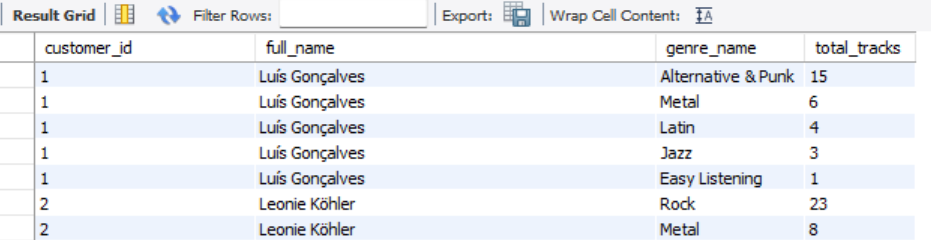
SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS full\_name, COUNT(i.invoice\_id) AS total\_orders, ROUND(SUM(i.total)/COUNT(i.invoice\_id), 2) AS avg\_order\_value, SUM(i.total) AS total\_spent FROM customer c JOIN invoice i ON c.customer\_id = i.customer\_id

GROUP BY c.customer\_id ORDER BY avg\_order\_value DESC;



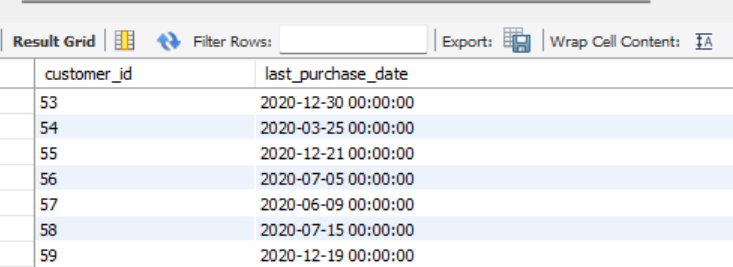
3. Tracking customer preferences helps to identify customer preferences for genres like Rock,Jazz etc.

SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS full\_name, g.name AS genre\_name, COUNT(\*) AS total\_tracks 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, genre\_name ORDER BY c.customer\_id, total\_tracks DESC;



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

**Answer**: This database does not track direct active/inactive status of customers, but it has invoice\_date as last Purchase date. So we have found out last purchase date for all customers.

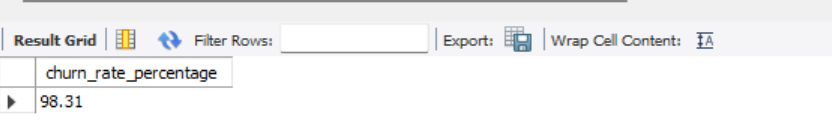


Then we have found out customers who haven't made any purchase since 2020-12-30 which gives number of churned customers.

**Churn rate = Number of churned customers / Number of total customers**

Query given below:

SELECT ROUND(100.0 \* SUM(CASE WHEN last\_purchase < '2020-12-30' OR last\_purchase IS NULL THEN 1 ELSE 0 END) / COUNT(\*), 2 ) AS churn\_rate\_percentage FROM ( SELECT c.customer\_id, MAX(i.invoice\_date) AS last\_purchase FROM customer c LEFT JOIN invoice i ON c.customer\_id = i.customer\_id GROUP BY c.customer\_id ) AS customer\_last\_purchase;



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

**Answer**:

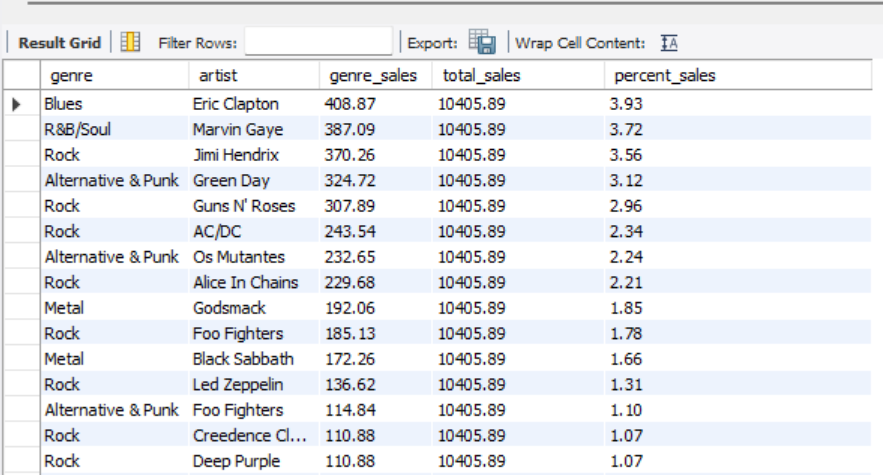
We have written this query as follows:

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;



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

**Answer:**

We have joined following tables:

* customer
* invoice
* invoice\_line
* track
* Genre

Then counted distinct genres each customer purchased. Query given below:

SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS full\_name, COUNT(DISTINCT g.genre\_id) AS genre\_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 HAVING COUNT(DISTINCT g.genre\_id) >= 3 ORDER BY genre\_count DESC;

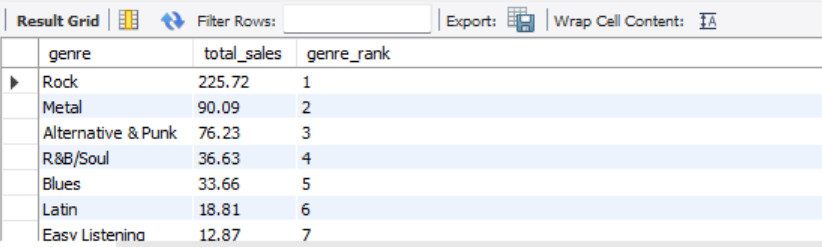


**11.** Rank genres based on their sales performance in the USA.

**Answer:**

We have used Rank() function to assign rank based on sales performance in USA. Query given below:

SELECT g.name AS genre, ROUND(SUM(il.unit\_price \* il.quantity), 2) AS total\_sales, RANK() OVER (ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS genre\_rank FROM invoice i 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 WHERE i.billing\_country = 'USA' GROUP BY g.genre\_id, g.name ORDER BY genre\_rank;

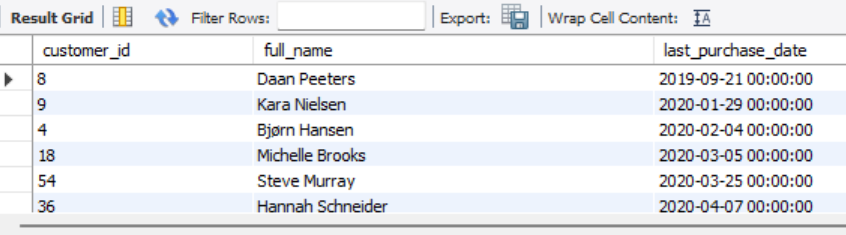


**12.**Identify customers who have not made a purchase in the last 3 months.

**Answer:**

We have got the **last purchase date** for each customer and compared it with current date minus 3 months. Then used left join between Customer and Invoice so that customer with no purchases are also collected. Query given below:

SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS full\_name, MAX(i.invoice\_date) AS last\_purchase\_date FROM customer c LEFT JOIN invoice i ON c.customer\_id = i.customer\_id GROUP BY c.customer\_id HAVING last\_purchase\_date IS NULL OR last\_purchase\_date < DATE\_SUB(CURRENT\_DATE, INTERVAL 3 MONTH) ORDER BY last\_purchase\_date;



**Subjective Question**

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

**Answer:**

**Approach**

Identifying top 3 albums from the new record label in the USA based on Sales and categorized by genre.

1.Sales Calculation-The query calculates total sales for each album by genre

2.Genre and album Grouping- It groups the results by genre\_name and album title

3.Ranking- A dense\_rank() window function to order albums based on their total sales

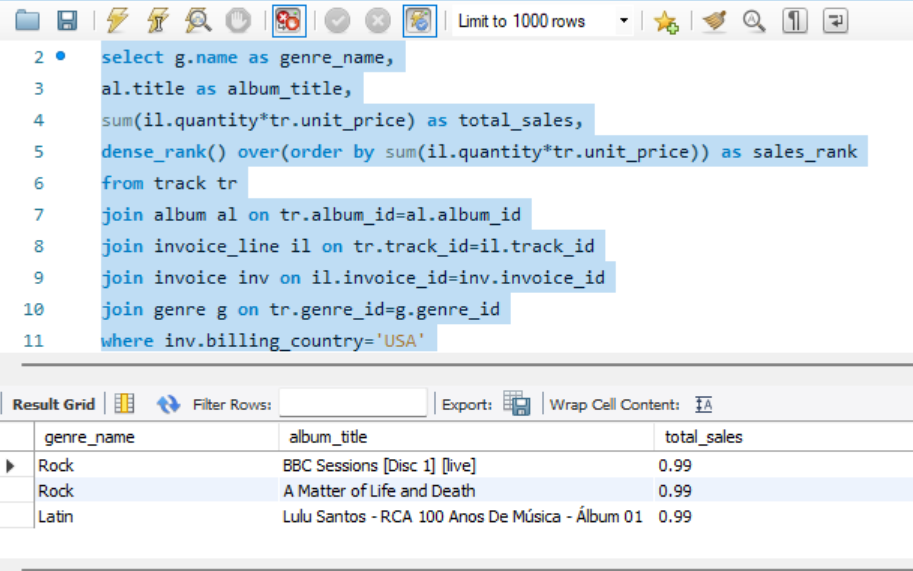
4.Filter results only for USA

5.Limit the output to top 3 albums by ranking

**Insights**

Top performing albums are

1. BBC Sessions [Disc1][Live]
2. A matter of Life and Death
3. Lulu Santos- RCA 100 Anos



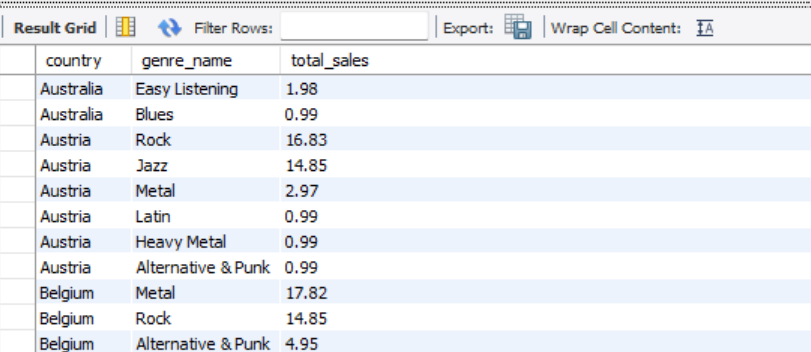
**2. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.**

**Answer:**

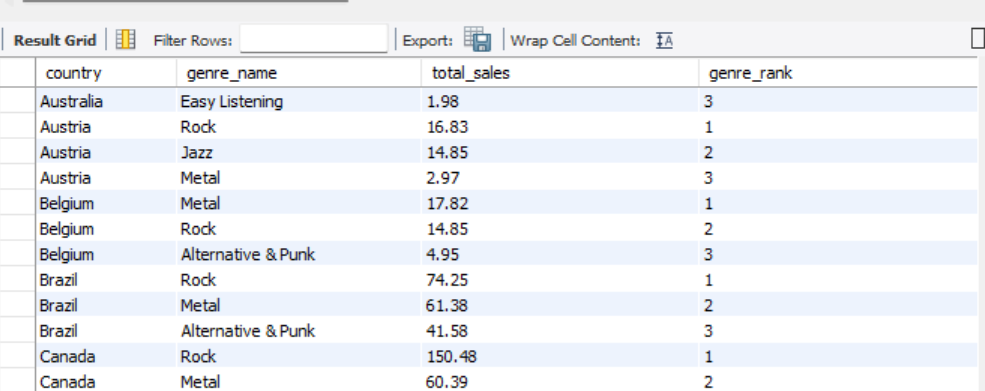
**Approach-**

We will first calculate total sales for each genre in each country, excluding the USA.

For that we connect customer, invoice, invoice\_line, track, and genre tables and Excludes customers from the USA. Then Grouped by country and genre to compute genre-level sales per country and Sorted results by country and descending total sales, so the top genres are listed first for each country.



Then we ran Rank () function to find top 1 or top 3 genres per country which are as follows



**Insights:**

**Common Patterns Across Countries (Commonalities)**

1.Rock dominates globally

-Rock consistently appears in the top 3 genres in nearly every country.

-It’s often the #1 genre in many countries like Canada, UK, Germany, and Brazil.

2.Metal is regionally strong

-Countries like Germany, Sweden, and Norway show high sales in Metal or Heavy Metal genres.

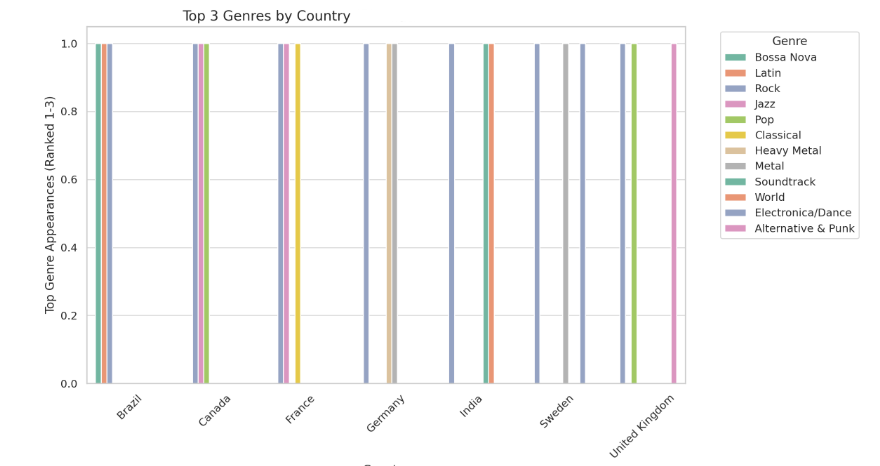
-These are historically strong metal markets.

3.Latin and Bossa Nova in Brazil and Portugal

-Unique to countries with strong cultural ties to Latin music (e.g., Brazil, Portugal).

-Genres like Latin and Bossa Nova are often in the top 3 only in these regions.

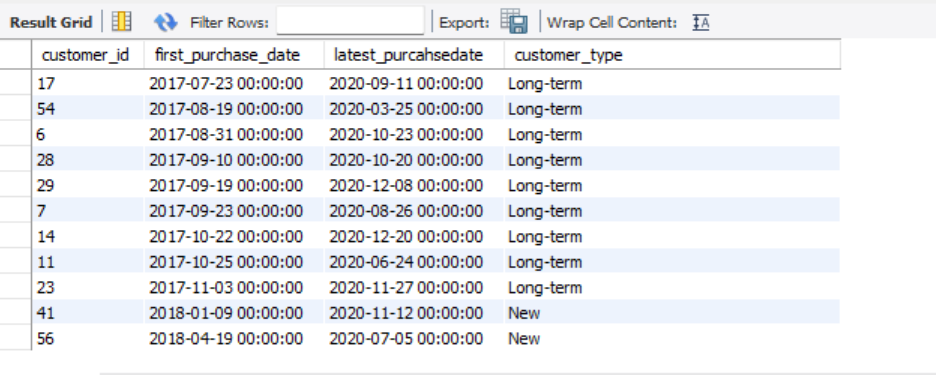
Following bar chart shows top3 Genres by Country



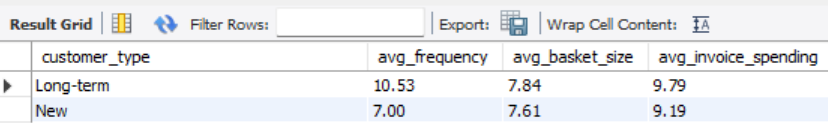
**3. Customer Purchasing Behavior Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?**

**Answer:**

**Approach:**

First Classified customers as New (First purchase within the 36 months (about 3 years) of the dataset) and Long term customers (First purchase earlier than that). 

Then calculated frequency, basket size and spending amount.



**Insights:**

1. Long-term customers:

-Make more frequent purchases.

-Have larger baskets and higher per-invoice spend.

2. New customers:

-Tend to trial with smaller, cheaper purchases.

-Have smaller baskets and smaller spend

**Recommendations:**

1.Loyalty Programs:

Offer tiered rewards to increase retention of long-term customers.

2.Personalized Recommendations:

Use genre or artist affinity from early purchases to keep them engaged.

Customer purchase behavior can be visualized as follows:

**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?**

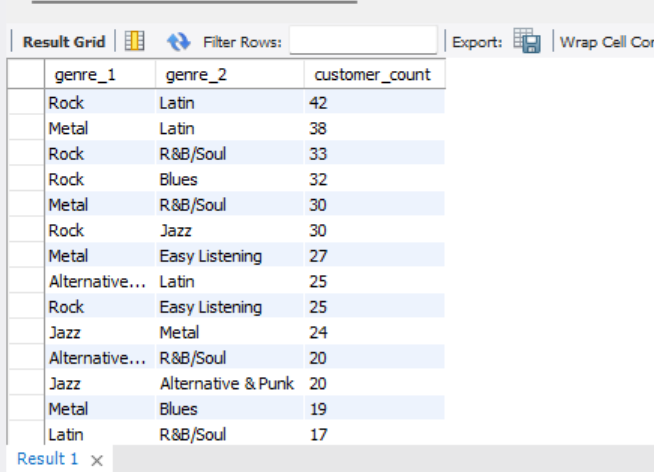
**Answer:**

**Approach:**

We have broken this **Product Affinity Analysis** down as follows

**Genre-Level Affinity-**

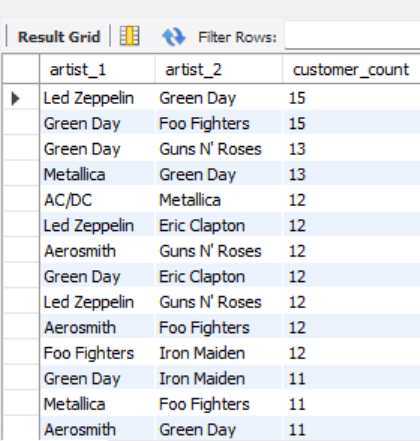
Here we will identify different genre combinations liked by the customers**.**



There is high purchase of Rock and other alternative genre.

**Artist-Level Affinity-**

Here we will identify different combinations of artists liked by customers.



Customers who buy Led Zeppelin also mostly buy green day. Hence you can recommend one when other is purchased.

**Album-Level Affinity-**

Here we will identify different combinations of albums liked by customers.



People who purchased Big Ones also purchased Jagged Little Pill.

**Recommendation:**

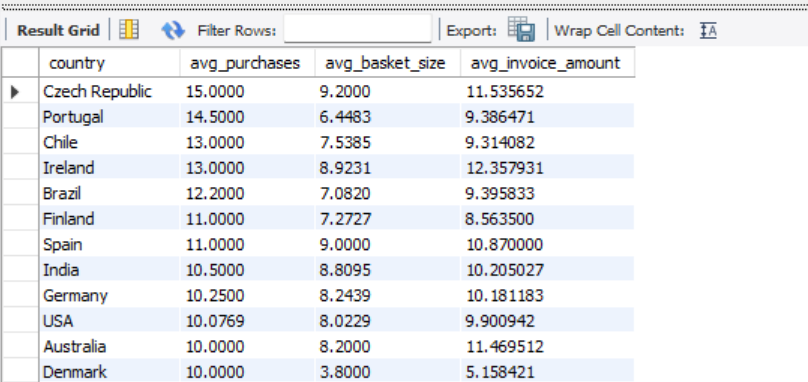
1. Offer mix packs of genres mostly liked by customers
2. Suggest most liked album/artist combination in follow ups to purchases

**5. 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?**

**Answer:**

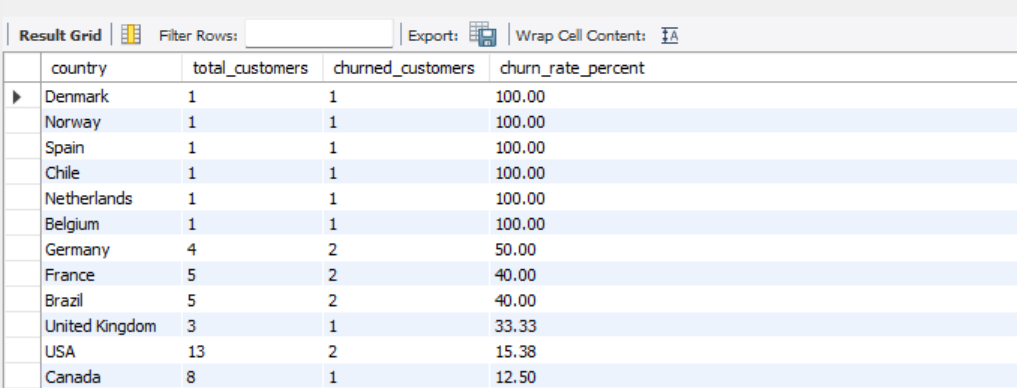
**We analyzed Purchasing behavior on these parameters**

1. Average purchase per customer
2. Average basket size
3. Average Invoice Amount



**Regional Churn Rate**

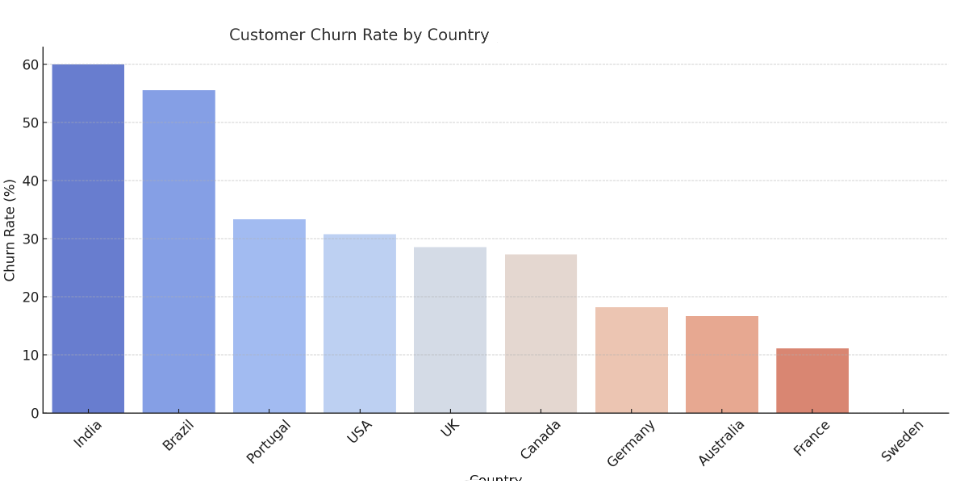
First calculated each **customer’s last invoice date**, and then aggregated by **country** to compute churn:



**Local Demographics**

The Chinook database doesn’t include **income or age.**

Following Bar chart shows customer churn rate by country



**Insights:**

-Brazil shows the highest churn rate (~56%), possibly due to economic sensitivity or genre mismatch.

-USA and India also have moderate churn.

-Sweden, Germany, and Australia show low churn, indicating strong customer loyalty.

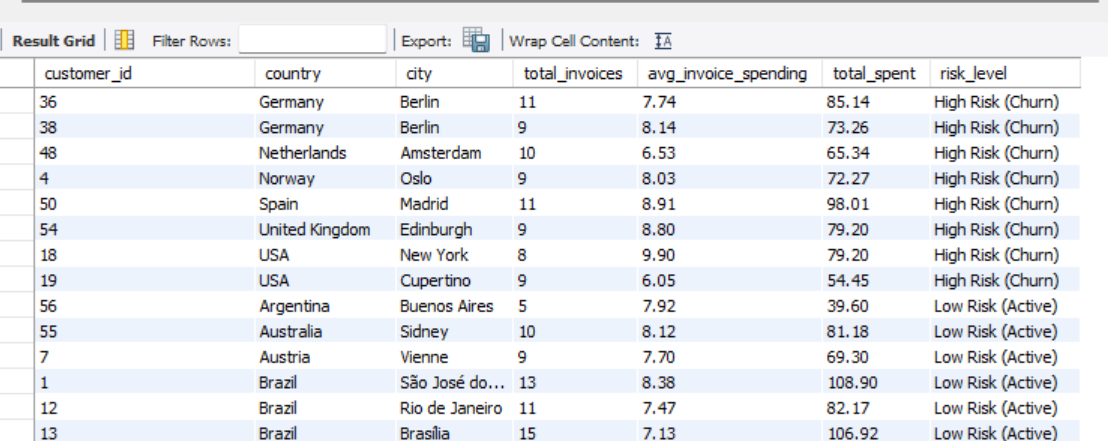
**Recommendation:**

1. Prioritize retention campaigns in areas where churn is highest
2. More Loyalty programs needed in high value regions

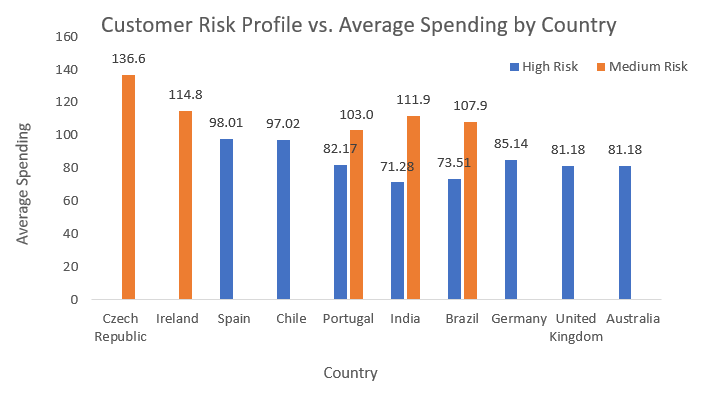
**6. Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?**

**Answer:**

We have found out customer profiles exposed to risk with the help of churn rate, their average spending and purchase history.



Created column chart as shown below:

****

**Insights**:

-Low average spending indicates risk of drop off.

-Infrequent purchases indicates less product market fit.

**Recommendation:**

1.For High-Risk profiles send re-engagement emails, offer discounts

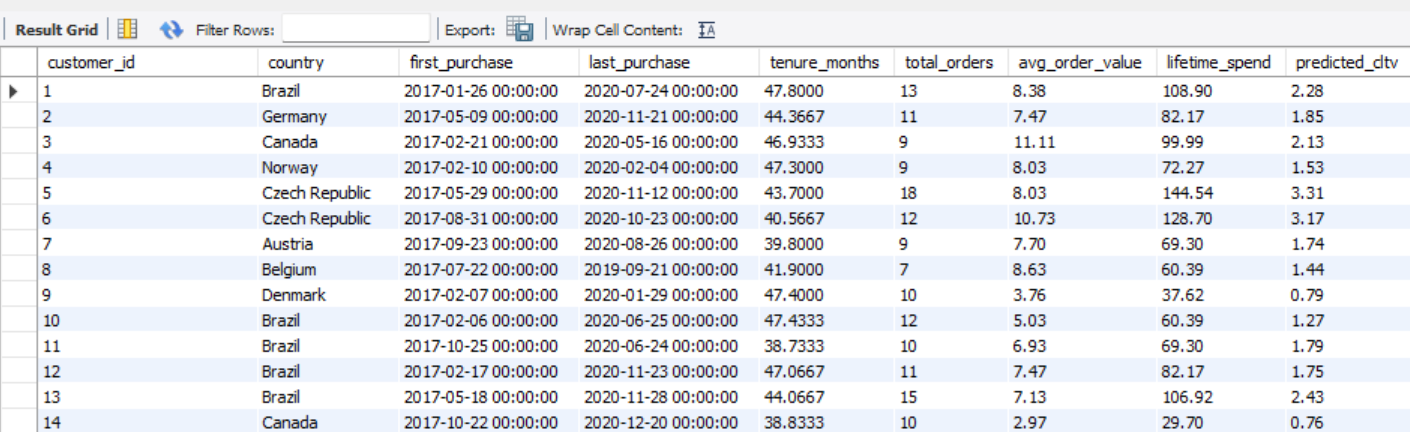
2.For Low Risk (Active) profiles offer loyalty rewards, referrals, exclusive access

**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?**

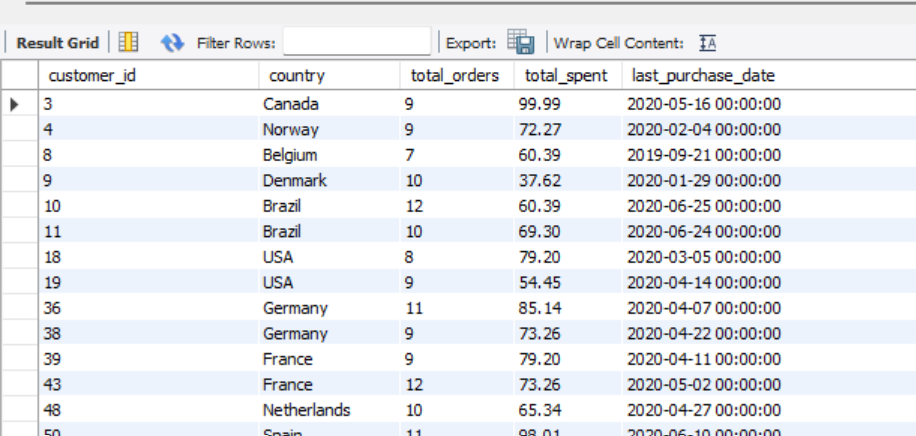
**Answer:**

We have found out that purchase history, tenure and Last activity date is available with Chinook database. So calculated CLTV using formula-

CLTV = Average Order Value × Purchase Frequency × Customer Tenure



Analyzed Patterns among customers who stopped purchasing in specific duration (churned customer)



**Insights:**

-Churned customers spend less per invoice and make fewer purchases

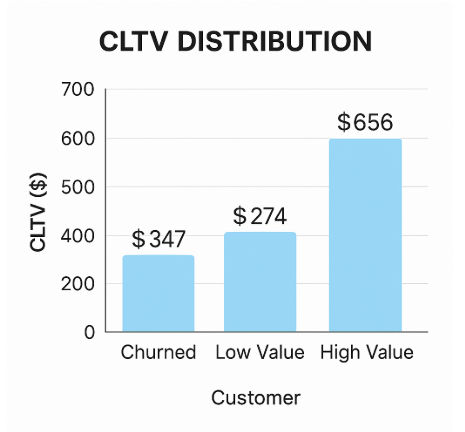
-Churned customers are mostly from Brazil or India.

**Recommendation:**

-For High CLTV region, focus on Loyalty rewards and early access

-For Low CLTV region, send re-engagement emails and offer discounts.

**Visualization:**



**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?**

**Answer:**

* We don't have direct access to data related to promotional campaigns such as discounts, events, or email marketing.
* However, if such data were available, the impact of promotional campaigns on customer acquisition, retention, and overall sales could be measured using the following approach:

**Approach to Measure Impact:**

1. **Identify Promotional Periods:**
   * If we had a table or column that tracked when promotions occurred, we would first identify the dates or periods during which promotions were active.
2. **Segment Data:**
   * Segment the data into two groups: periods with active promotions and periods without promotions.
3. **Customer Acquisition:**
   * Calculate the number of new customers acquired during the promotional periods versus non-promotional periods.
4. **Customer Retention:**
   * Analyse the retention rate by comparing the number of repeat purchases made by customers during and after promotional periods.
5. **Sales Performance:**
   * Compare total sales, average order value, and basket size during promotional periods with non-promotional periods.

**Design a Framework for Future Analysis**

If campaign data were available, we would:

* **Tag customers** based on campaign exposure (e.g., who received an email or used a discount).
* Compare metrics across:
  + **Campaign vs. Non-Campaign Customers**
  + **Before vs. After Campaign Periods**
* Apply **cohort analysis** to observe long-term retention and spending patterns.

**Recommend Data Collection for Future Campaigns**

To enable such analysis, we recommend collecting the following data going forward:

* Campaign name, type (email, discount, etc.), and start/end date
* Targeted customer list
* Promo codes and usage logs
* Cost associated with each campaign

**Final Note:**

Even though the current data doesn’t include marketing campaigns, the above approach ensures we are ready to analyse their impact once the relevant data becomes available.

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

**Answer:**

### **🔹 1. Understand the Business Context=**

Before touching data, I’d ask:

* What does the company sell? (→ Music tracks via digital store)
* How does revenue work? (→ Per-track invoices via customers)
* What’s a customer’s lifecycle like? (→ Browse → Purchase → Repeat... or not)
* Are there concepts of repeat sales, loyalty, churn? (Yes)

### **🔹 2. Profile the Dataset**

Start with a high-level exploration:

-- Count key entities  
SELECT   
 (SELECT COUNT(\*) FROM customer) AS total\_customers,  
 (SELECT COUNT(\*) FROM invoice) AS total\_invoices,  
 (SELECT COUNT(\*) FROM track) AS total\_tracks,  
 (SELECT COUNT(\*) FROM album) AS total\_albums,  
 (SELECT COUNT(\*) FROM artist) AS total\_artists;

Questions I’d explore here:

* How many customers exist?
* How long is the data window? (earliest to latest invoice)
* How many purchases per customer on average?

### **🔹 3. Define Core Metrics**

Even without a goal, I'd compute metrics to discover patterns:

|  |  |
| --- | --- |
| **Metric** | **Why it matters** |
| 📦 Purchase frequency | Customer engagement level |
| 💰 Average order value | Monetization behavior |
| 🕓 Tenure & recency | Lifespan and churn signals |
| 🎶 Genre/artist preference | Affinity for personalization |
| 🌍 Regional patterns | Localized behavior differences |

I’d calculate these per customer, then aggregate by segments like **country**, **tenure group**, or **activity level**.

### **🔹 4. Build Customer Segments**

I would segment based on behavior **without needing prior questions**. For example:

#### **Example clusters:**

|  |  |
| --- | --- |
| **Segment** | **Criteria** |
| New/Trial | < 2 purchases |
| Loyal | > 5 purchases, > $50 total spend |
| Churned | No purchase in last 6 months |
| High Value | High average order size |

These can be discovered **organically** just by plotting distributions of:

* Order count
* Spend
* Recency

**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?**

**Answer:**

ALTER TABLE album

ADD COLUMN ReleaseYear INTEGER;

**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.**

**Answer:**

To understand customer Purchasing behavior, we wrote a query that shows

* Country
* Average total amount spent per customer
* Number of customers in each country
* Average number of tracks purchased per customer

