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

**CHINOOK\_MUSIC\_PROJECT**

**OBJECTIVE** **QUESTION WITH THEIR ANSWRES**

handle it?

Solution-There are some missing values in the following tables

* 49 Company,29 state,47 fax values are null in the customer table
* 1 reports\_to value is null in the employee table
* 978 composer values are null in the track table

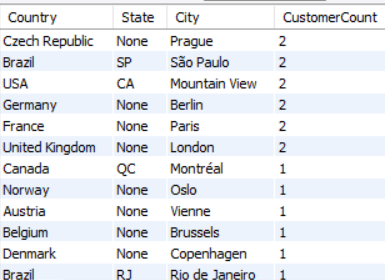
:There are no duplicates values present in data set

: For missing or null values we can use COALESCE function to handle this.

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

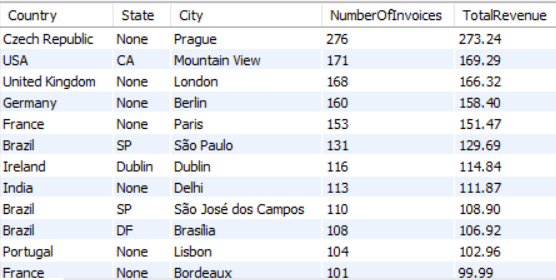
Solution— Below are the top selling tracks and top artist in **USA** their most famous genres

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

Solution- 

In customer table we do not have age and gender so we break customer information according to location and above table shows the demographic breakdown .

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

Solution-

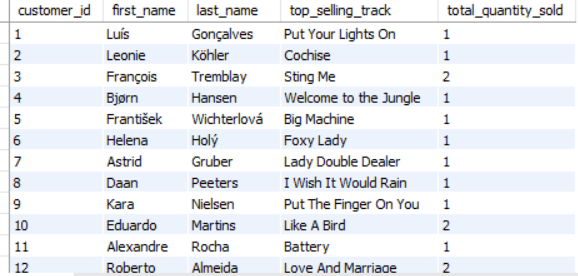
Here we can see all the output and also conclude highest revenue generated by ‘Czech Republic’ which is ‘273.24’

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

Solution- 

Here we can see top 5 customers by total revenue in each country ,if we see argentina have only one record that means only one people belong there but if we see in brazil we got 5 customer with their respective revenue.

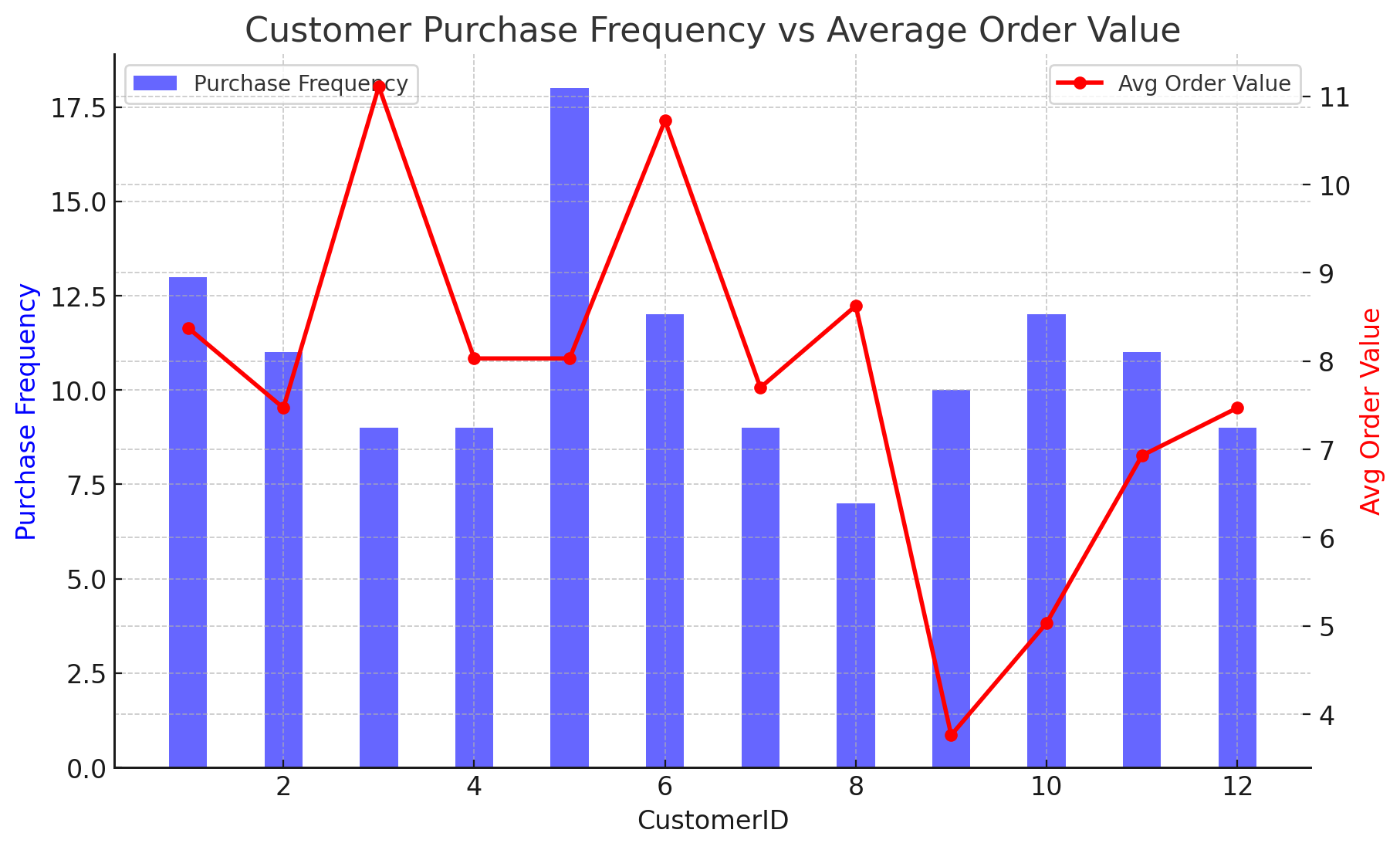
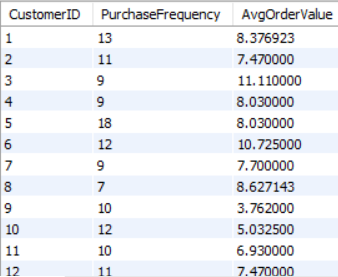
**6**. Identify the top-selling track for each customer?

Solution- 

Here we can see top-selling track for each customer

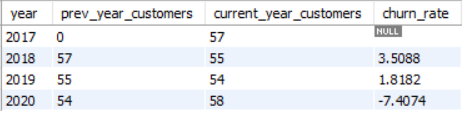
7. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?

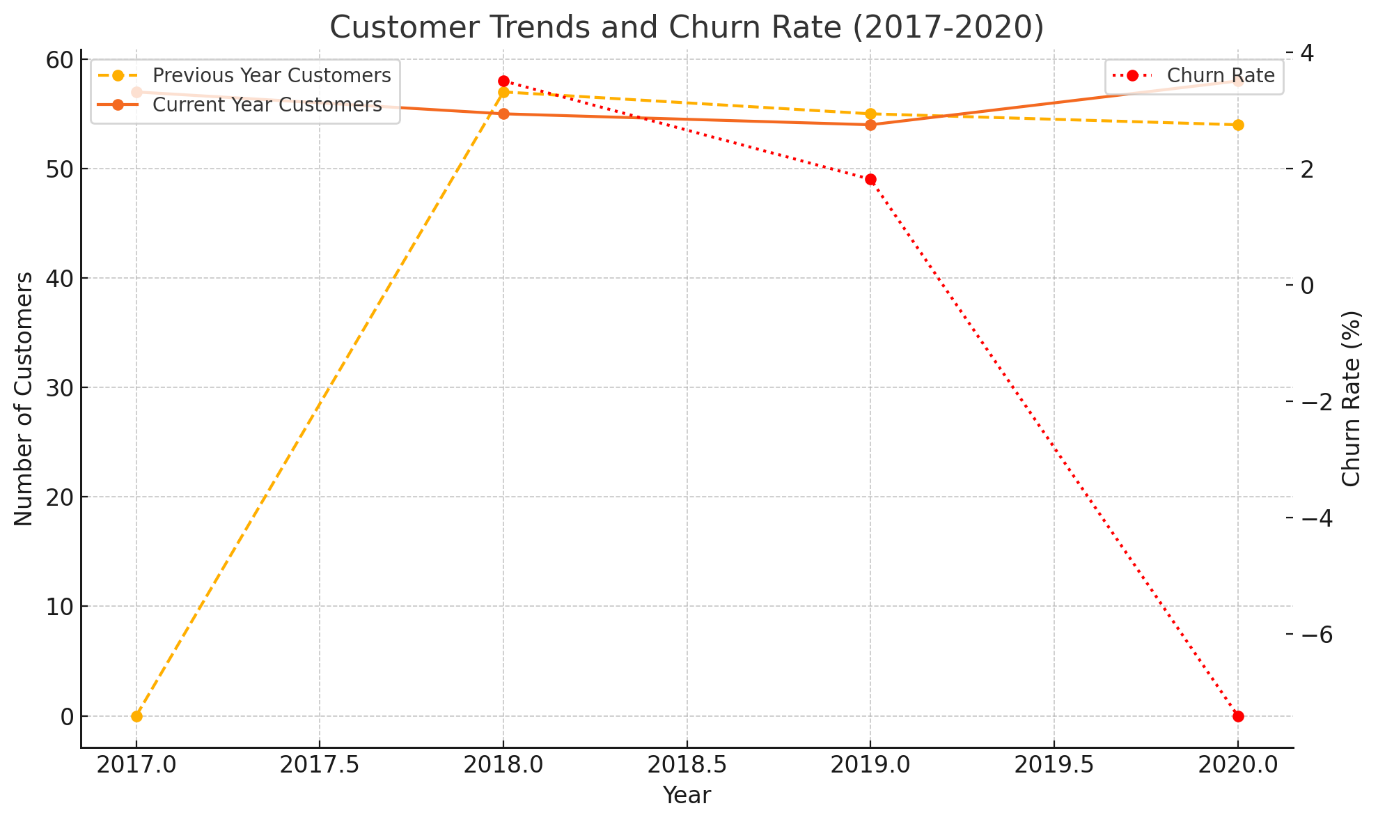
Solution- To identify **patterns or trends in customer purchasing behavior** such as:

**OUTPUT-** 

There is no patterns or trends between frequency of purchases by customers and average order value generated by these customers.

8. What is the customer churn rate?

Solution- 



**Observations and Conclusions:**

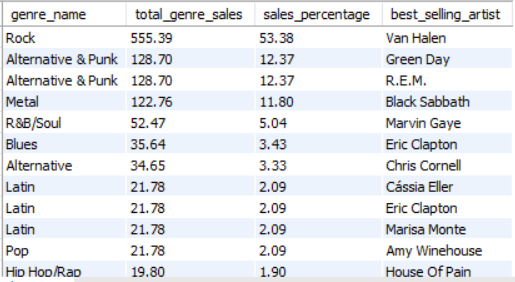
1. **Overall Trend**:
   * The churn rate initially increased but decreased significantly in 2020. This suggests that the business may have made improvements in its customer retention efforts.
2. **Retention Focus**:
   * Between 2017 and 2020, the overall number of customers has decreased , indicating that while churn is improving, attracting new customers might also need attention.

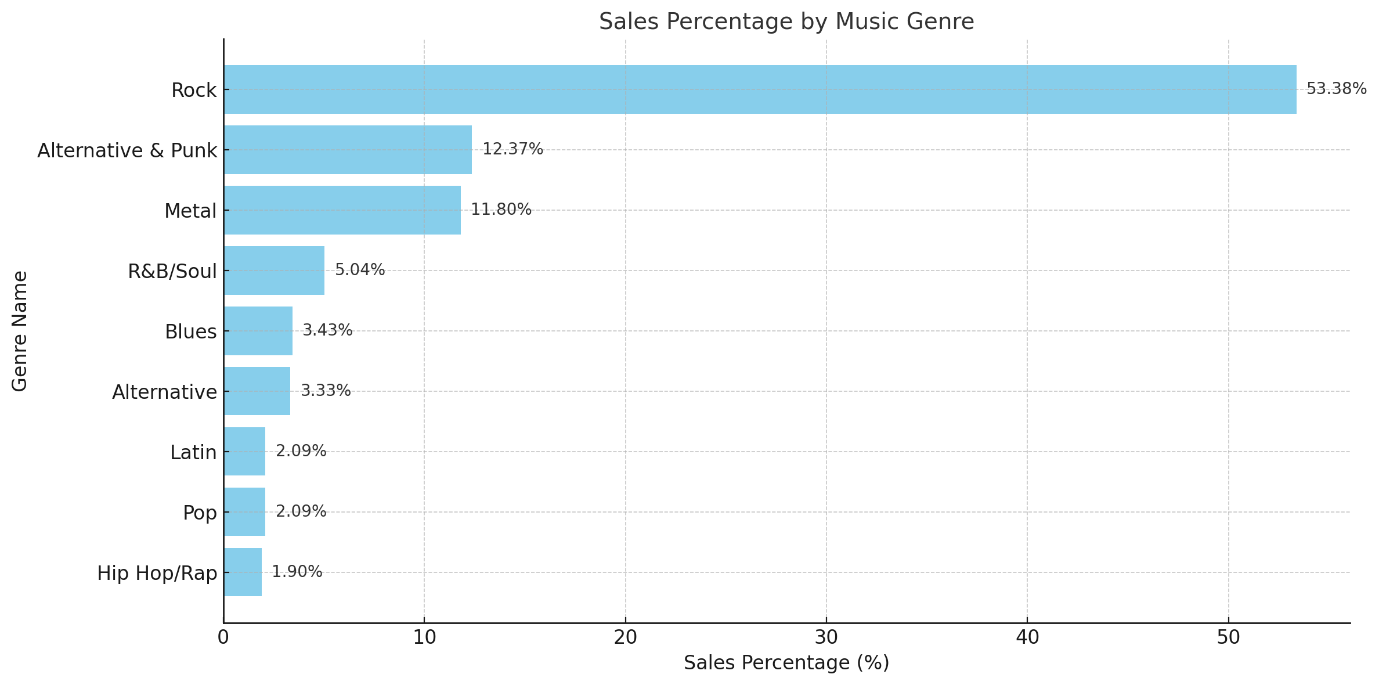
**Conclusion**

. From above output it is clear that with time, company building a strong relation with customer as we can see in 2020 a negative churn rate suggests customer **growth**, meaning more customers were gained than lost. This could be due to effective marketing, increased demand for music services or loyalty programs.

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

Solution- Percentage of Total Sales by Genre in the USA





Here we can see highest genre percentage by ‘Rock’ is ’53.38%’ also

we can see best selling artist ‘**Van Halen’**

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

**Solution**- 

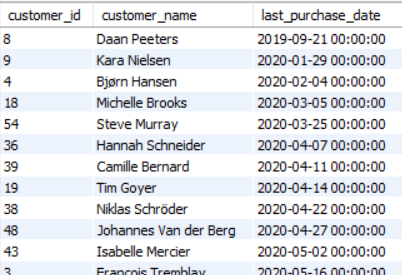
Here are the customers who have purchased tracks from at least 3 different genres in which **‘Leonie Kohler’** have maximum genre count and that is **14.**

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

Solution - 

Here we can see ‘**Rock**’ being on top with total sales of ‘**555.39**’ and **‘Tv Shows’** at the bottom with total sale of ‘**0.99**’

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

Solution- 

These are the customers who have not made a purchase in the last 3 months.

There are total 22 customers who haven’t made any purchase in last three months.

Subjective Questions with their answers

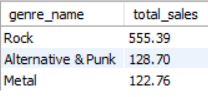
1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis?

Solution- To recommend the three albums from the new record label for advertising and promotion in the USA, we can base the decision on the following steps:

**Steps to Analyze and Recommend:**

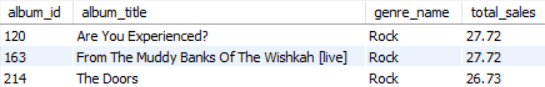
**Identify Top-Selling Genres in the USA**:

* + Use sales data to rank genres by their total revenue contribution in the USA.



Top 3 genres based on sales revenue in the USA.

**Match Genres to Albums**:

* + Find which albums primarily feature tracks from these top-selling genres.
  + 

**Select the Top 3 Albums**:

* + Prioritize albums with the highest sales contribution from top-selling genres.

**Explanation:**

**Top Genres**:

* + The first query identifies the three most profitable genres in the USA.

**Top Albums for Promotion**:

* + The second query matches albums to tracks in the top genres.
  + Filters for purchases in the USA and calculates total sales revenue for each album.
  + Sorts by sales and limits the result to the top 3 albums.

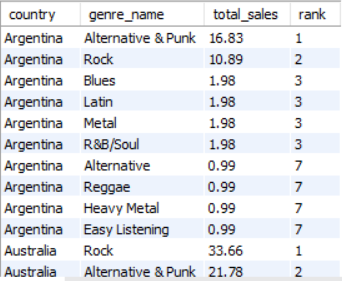
**Recommendations:**

1. Focus marketing efforts on **albums** that align with the top-performing genres identified in the analysis.
2. Highlight tracks and artists from these albums in advertisements.
3. Use the sales trends to tailor promotional messages to target audiences in the USA.

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

**Solution-** To determine the top-selling genres in countries other than the USA and identify commonalities or differences, we can analyze genre sales data while excluding the USA**.**

**Rank Genres by Sales in Countries Other than the USA**

****

.

**Insights:**

**Commonalities:**

* + The commonality between the data regarding USA and rest of the countries is that the ROCK genre is the top genre in all countries.

**Differences:**

* + Certain genres, like Classical, might be more popular in specific regions (e.g., Germany)

**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?

**Solution- Customer Purchasing Behavior Analysis: Comparing Long-Term vs. New Customers**

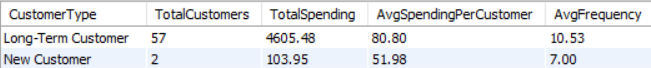
This analysis involves segmenting customers into long-term and new categories based on their first purchase date, then examining purchasing habits such as frequency, basket size, and spending amount.

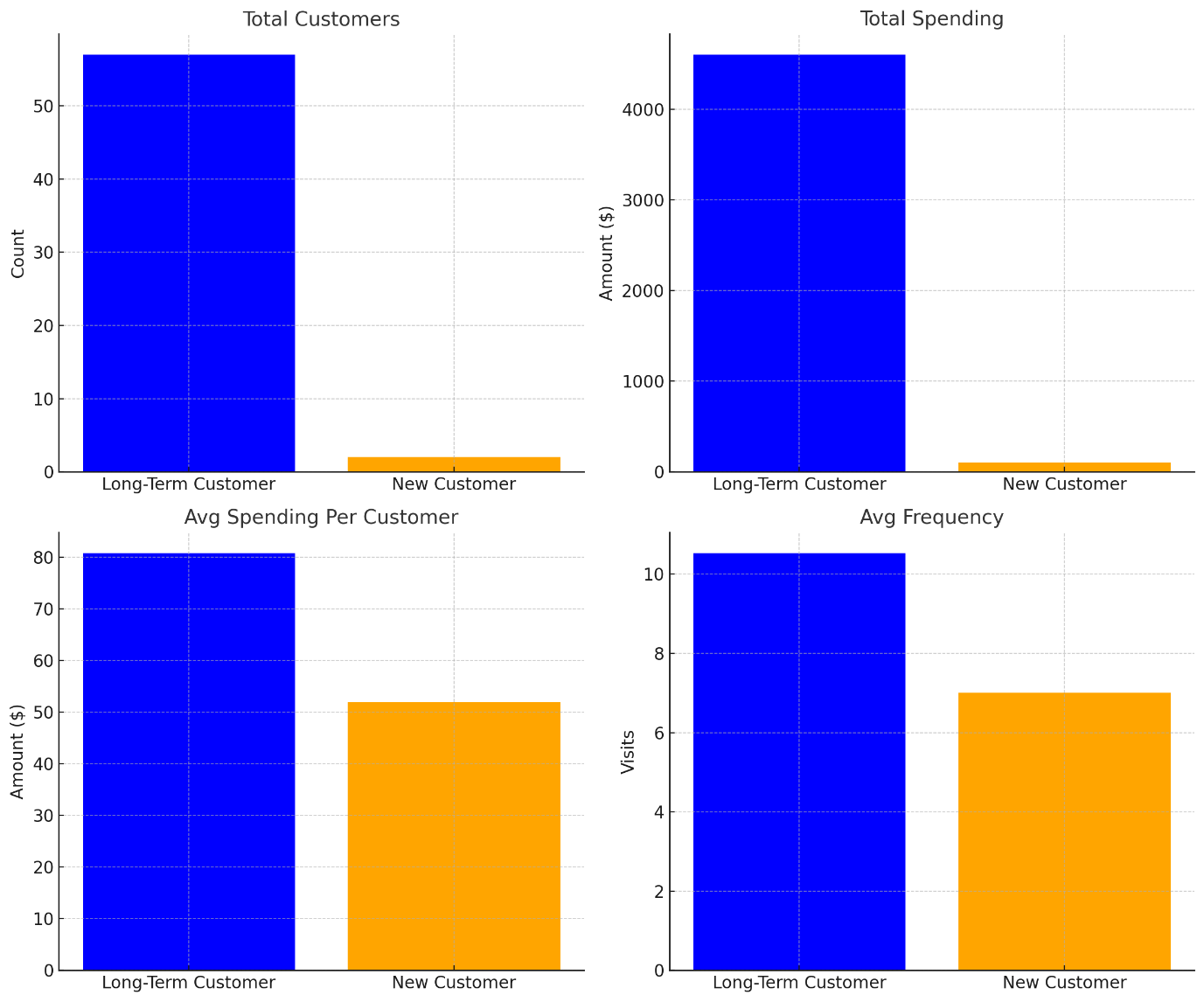
**Customer Purchasing Behavior Analysis: Comparing Long-Term vs. New Customers**

This analysis involves segmenting customers into **long-term** and **new** categories based on their first purchase date, then examining purchasing habits such as frequency, basket size, and spending amount.

**Segment** **Customers into Long-Term and New**

I consider first purchase after 2017 are classified as New Customers else long term





**Insights:**

1. **Purchase Frequency**:
   * Long-term customers make purchases more frequently, indicating stronger engagement and loyalty.
2. **Basket Size**:
   * Long-term customers tend to purchase slightly more items per transaction, possibly due to trust in the brand.
3. **Spending Amount**:
   * Long-term customers spend significantly more on average, highlighting their higher lifetime value.

**Retention Strategies:**

1. **For New Customers**:
   * Introduce welcome offers or discounts to encourage repeat purchases.
   * Provide personalized recommendations based on their initial purchase.
2. **For Long-Term Customers**:
   * Implement a loyalty rewards program to sustain engagement.
   * Offer exclusive deals or early access to new products.
3. **General Improvements**:
   * Use basket size insights to cross-sell or upsell complementary products.
   * Encourage frequency with subscription models or regular engagement campaigns.
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?

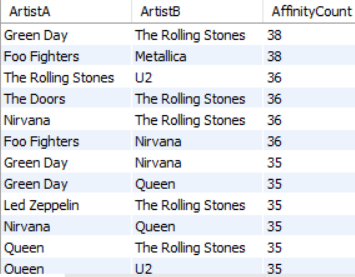
Solution- **Product Affinity Analysis: Music Genres, Artists, or Albums Frequently Purchased Together**

The goal is to understand which items (products) are often bought together by customers, and then use this information to guide product recommendations and cross-selling initiatives.

To start, let's identify which **genres**, **artists**, or **albums** are purchased together by the same customer.



**Identify Other Product Pairings** (Albums, Artists, etc.): You can similarly extend the analysis to **albums**



**Conclusion:**

By analyzing product affinities (frequent co-purchases), you can:

* **Recommend related products** (albums, artists, or genres).
* **Create bundles** of items that customers are likely to purchase together.
* **Target customers with personalized marketing** based on their purchase behavior.

These insights help optimize product recommendations, increase sales, and enhance cross-selling initiatives.

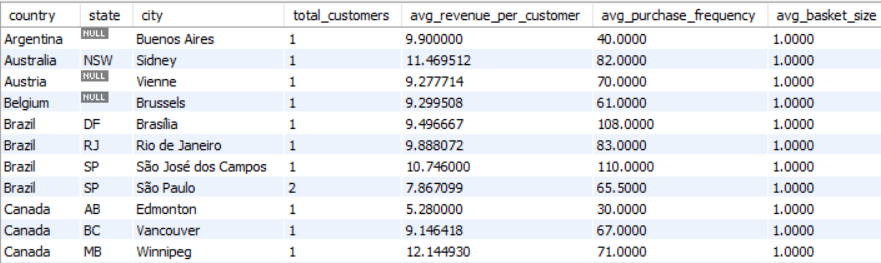
**5**.Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors**?**

**Solution**-**Regional Market Analysis: Customer Purchasing Behaviors and Churn Rates Across Geographic Regions**

Analyzing customer purchasing behavior and churn rates across geographic regions can reveal patterns and correlations with local demographics or economic factors. Here's how to structure the analysis:

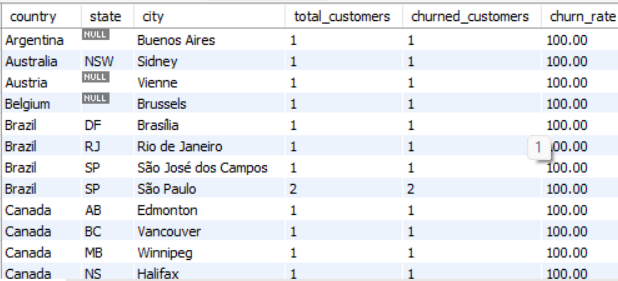
**Purchasing Behaviors by Region**

Purchasing Behaviors by Region

****

Purchasing Behaviors by Region

**Churn Rates by Region**

****

**Correlation with Demographic/Economic Factors**

1. **Demographics:**
   * Regions with younger populations might have higher purchase frequency but lower average basket size due to spending habits.
   * Older populations might show loyalty but lower frequency**.**
2. **Economic Factors:**
   * Wealthier regions likely correlate with higher spending and lower churn.
   * Economically strained areas may show lower revenue per customer and higher churn.

**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?

Solution-**Customer Risk Profiling: Identifying At-Risk Customer Segments**

### Steps for Customer Risk Profiling:

### ***Segmentation by Key Attributes***:

* Segment customers based on **age**, **location**, and **purchase behavior** (e.g., frequency of purchases, average spend, recent purchases).

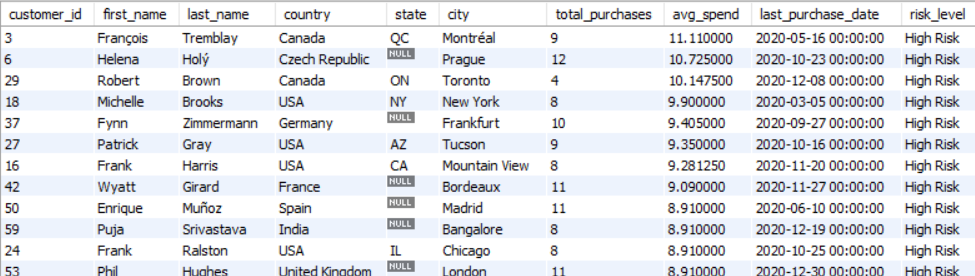
#### **Define Risk Indicators**:

* **High-risk indicators** include:
  + Low engagement.
  + Reduced spending over time.
  + No purchase in recent months.
  + Age and demographic factors suggesting lower disposable income.
  + Location factors suggesting economic decline.

#### **Churn Probability Model**:

* Combine these indicators to predict which segments have the highest **churn probability** or **reduced spending risk**.

**Profile High-Risk Customers**: Since we don't have the birth\_date, we will focus on **purchase history** (e.g., total purchases, average spend, and last purchase date) to identify at-risk customers.



**Insights and Recommendations:**

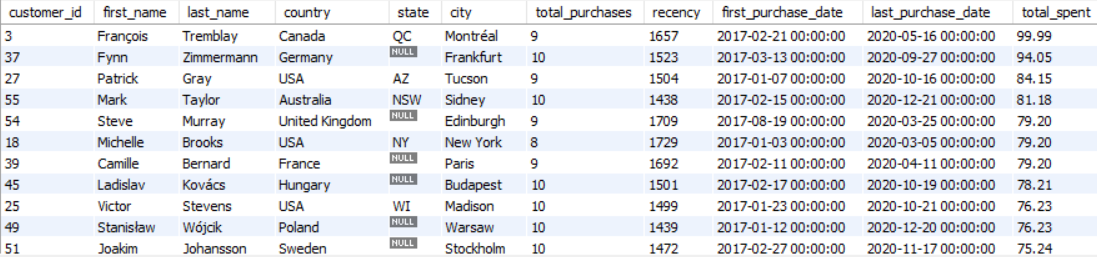
1. **High-Risk Customers**:
   * **Customers with no purchases** or low spending are more likely to churn. These customers should be targeted with **personalized offers**, **special promotions**, or **re-engagement campaigns** to bring them back.
2. **Low Average Spend**:
   * Customers with **low average spend** (e.g., below 5 units) may require incentives or **product bundles** to encourage larger purchases.
3. **No Purchases in 6 Months**:
   * Customers who have not made a purchase in **the last 6 months** are prime candidates for **reactivation campaigns** (e.g., offering them exclusive discounts to re-engage).
4. **Active Customers**:

* Customers who are actively purchasing (with reasonable spend) are less likely to churn. **Loyalty programs** or **cross-selling opportunities** should be considered to increase their spend.

**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?

Solution-**Customer Lifetime Value (CLV) Modeling**

To estimate **Customer Lifetime Value (CLV)** and analyze purchase patterns among customers, we can use available data like **tenure**, **purchase history**, and **spending behavior**. CLV helps businesses identify valuable customer segments, optimize marketing budgets, and design loyalty programs.

CLV Calculation: The following table computes **RFM metrics** and predicts **lifetime value** for each customer segment.

**Explanation:**

1. **RFM Metrics**:
   * **Recency**: To find the number of days since the last purchase.
   * **Frequency**: Total number of purchases
   * **Monetary Value**: Sum of all purchases
2. **Average Order Value**:
   * Gives the average spending per order.
3. **Average Monthly Spend**:Divide total spend by the number of months since the first purchase

**Customer Segmentation**:

**High Value**: Frequent purchases (10+) and high total spending (>500 units).

**Medium Value**: Moderate purchases (5-10).

**Low Value**: Fewer than 5 purchases or low spending.

**Common Patterns of Customers Who Stopped Purchasing:**

To identify customers who have stopped purchasing, analyze their recency and frequency metrics.



**Insights from CLV and Churn Analysis:**

**Patterns in High-Value Customers:**

1. **Consistent Spending**: High-value customers purchase frequently and maintain a steady spending rate.
2. **Short Recency**: They tend to make recent purchases (e.g., within the last 1-3 months).
3. **Preferred Genres**: They often stick to specific genres or artists.

**Patterns in Churned Customers:**

1. **Low Frequency**: Customers who stopped purchasing often had fewer transactions historically.
2. **Low Average Spend**: Churned customers typically have lower average order values.
3. **Long Recency**: They haven't made a purchase in the last 6+ months.

**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?

Solution-To measure the impact of **promotional campaigns** on **customer acquisition, retention, and overall sales**, we can adopt a data-driven approach leveraging available customer and sales data.

**Here's a detailed plan:**

**1. Metrics for Measuring Campaign Impact**

**Customer Acquisition**

* New Customers Acquired: Count of new customers during or after the campaign period.
* Growth Rate in Customers: Percentage increase in new customers compared to the pre-campaign period

**Customer Retention**

* **Retention Rate**: Percentage of existing customers who made purchases during/after the campaign.
* **Reactivation of Churned Customers**: Count of customers who had not purchased in 3+ months but made a purchase during/after the campaign.

**Sales Impact**

* **Sales Growth**: Percentage increase in total sales during/after the campaign.
* **Average Order Value (AOV)**: Changes in average transaction value.
* **Purchase Frequency**: Number of purchases per customer during the campaign compared to before.

**Engagement Metrics (if applicable):**

* **Email Open/Click Rates**: Percentage of customers engaging with email campaigns.
* **Conversion Rate**: Percentage of email recipients who made a purchase.

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

**Solution**-**Understand the Problem Context:**

Identify Chinook's primary business goals such as :

* Increasing revenue.
* Improving customer retention.
* Enhancing customer acquisition.

Define the Core Business Questions:

* "Which geographical regions generate the most revenue?"
* "What is the average purchase behaviour of customers across countries?"
* "What drives customer retention, and where are we losing customers?"
* "Which products or services contribute the most to revenue?"

**Explore and Understand the Data**

* Review the available tables in the Chinook database (e.g., Customer, Invoice, Invoice Line, Track, Genre, etc.).
* Understand the relationships between these tables (e.g., how customers link to invoices, how tracks contribute to purchases).

**Identify Key Metrics**

Customer Insights:

* Total customers by country.
* Average spending per customer.
* Churn rate and retention rate.
* Average number of tracks purchased.

Sales Insights:

* Total revenue by country.
* Most purchased genres or tracks.

Geographical Trends:

* Which countries contribute most to revenue?
* Identify high-value and low-value regions.

**Conclusion**

Without predefined questions, the approach revolves around exploring the data, getting insights and aligning analysis with business goals.

**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?

Solution- To add a new column named ReleaseYear of type INTEGER to the Albums table, you can use the SQL ALTER TABLE statement. Here's the query:

**ALTER TABLE album**

**ADD COLUMN ReleaseYear INTEGER**;

**Explanation:**

1. **ALTER TABLE album**:
   * Specifies the table (album) that we want to modify.
2. **ADD COLUMN ReleaseYear INTEGER**:
   * Adds a new column named ReleaseYear with the data type 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?

Solution- To analyze the purchasing behavior of customers based on their geographical location, we can use the following SQL query:

**SELECT**

**c.country AS Country,**

**COUNT(c.customer\_id) AS NumberOfCustomers,**

**AVG(i.total) AS AvgTotalSpent,**

**AVG(t.TrackCount) AS AvgTracksPurchased**

**FROM customer c**

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

**LEFT JOIN (**

**SELECT**

**il.invoice\_id,**

**COUNT(il.track\_id) AS TrackCount**

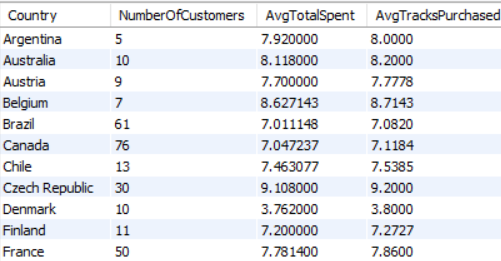
**FROM invoice\_line il**

**GROUP BY il.invoice\_id**

**) t ON i.invoice\_id = t.invoice\_id**

**GROUP BY c.country**

**ORDER BY c.country;**

****

**Explanation:**

1. **customer Table:**
   * Contains customer details, including their geographical location (country).
2. **invoice Table:**
   * Linked to customers via customer\_id.
   * Includes the total amount spent on each invoice.
3. **invoice\_line Table:**
   * Used to count the number of tracks purchased per invoice.
   * A subquery (t) calculates the number of tracks (TrackCount) for each invoice.
4. **Query Steps:**

* Join customer with invoice using customer\_id.
* Use a subquery to count tracks (TrackCount) for each invoice in invoice\_line.
* Join this subquery result (t) with invoice using invoice\_id.
* **Group by country to calculate:**
  + NumberOfCustomers: Count of customers per country.
  + AvgTotalSpent: Average total amount spent by customers per country.
  + AvgTracksPurchased: Average number of tracks purchased per customer (using the subquery result).

**Ordering:**

* The results are ordered by country for readability**.**