Objective Questions

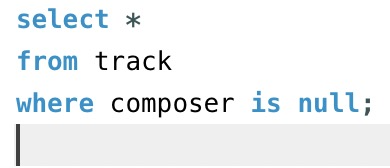
1. **Does any table have missing values or duplicates? If yes, how would you handle it?**

Yes, some columns in the tables have missing values. They are found by running queries.

Example:

SELECT \* FROM artist

WHERE name is null.



Missing values are found in the following columns in the dataset:

Customer

Employee

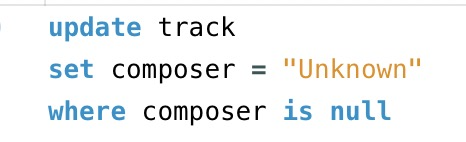
Example MYSQL response after running the query:

978 row(s) affected Rows matched: 978 Changed: 978 Warnings: 0

We need to set the database to allow updates. It can be done by using the following query:

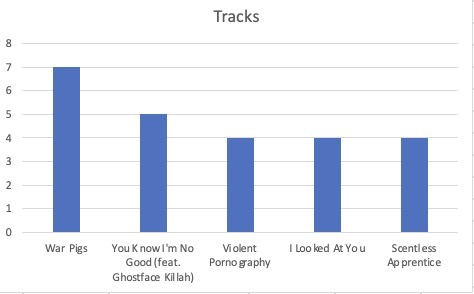
SET SQL\_SAFE\_UPDATES =1;

The missing values are further updated by using ALTER TABLE.



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

Top selling Tracks in USA:



SELECT t.name, SUM(il.quantity) AS quantity

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

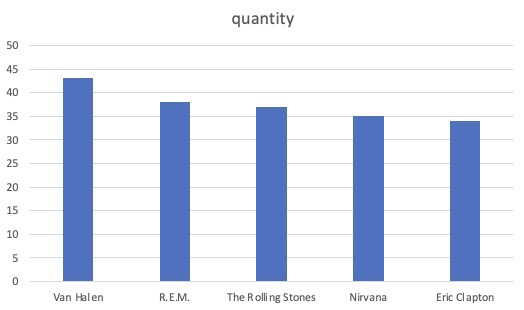
invoice\_line il ON i.invoice\_id = il.invoice\_id INNER JOIN track t ON il.track\_id = t.track\_id

WHERE c.country = 'USA'

GROUP BY t.name

ORDER BY quantity DESC;

Top artists in USA:



Van Halen is the best artist in USA.

Approach to find the best artist:

SELECT art.name, SUM(il.quantity) AS quantity

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

invoice\_line il ON i.invoice\_id = il.invoice\_id INNER JOIN track t ON il.track\_id = t.track\_id

INNER JOIN album a on t.album\_id = a.album\_id INNER JOIN artist art ON a.artist\_id = art.artist\_id

WHERE c.country = 'USA'

GROUP BY art.name

ORDER BY quantity DESC;

**Top Genre of the Top artist in USA:**

SELECT g.genre\_id, g.name AS genre\_name,SUM(il.quantity) AS total\_sold

FROM

invoice\_line il

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

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

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

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

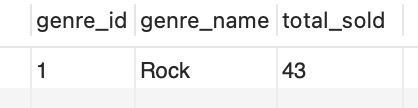
INNER JOIN artist a ON al.artist\_id = a.artist\_id

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

WHERE c.country = 'USA' AND a.artist\_id = 152

GROUP BY g.genre\_id, g.name

ORDER BY total\_sold DESC;



Note: The top artist in USA only has data on ‘Rock’ genre so it is provided.

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

The demographics of Chinook’s customer base can be divided into three major types. On the basis of Country, State, City.

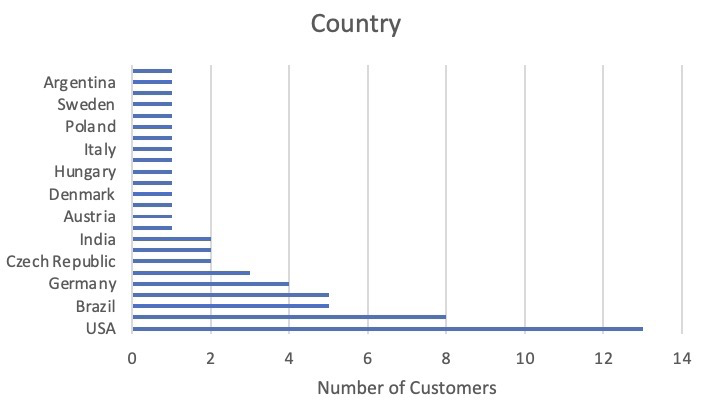
Basis of country:

SELECT country, COUNT(\*) AS NumberOfCustomers

FROM customer

GROUP BY country

ORDER BY NumberOfCustomers DESC;



Basis of State:

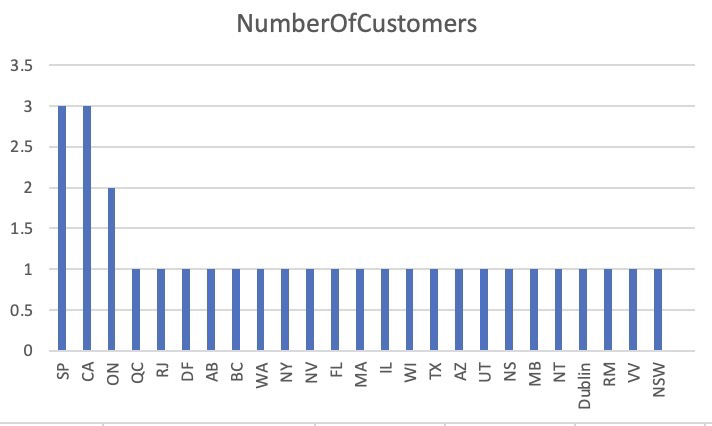
SELECT state as state, COUNT(customer\_id) AS NumberOfCustomers

FROM customer

WHERE state != 'Unknown State'

GROUP BY state

ORDER BY NumberOfCustomers DESC;



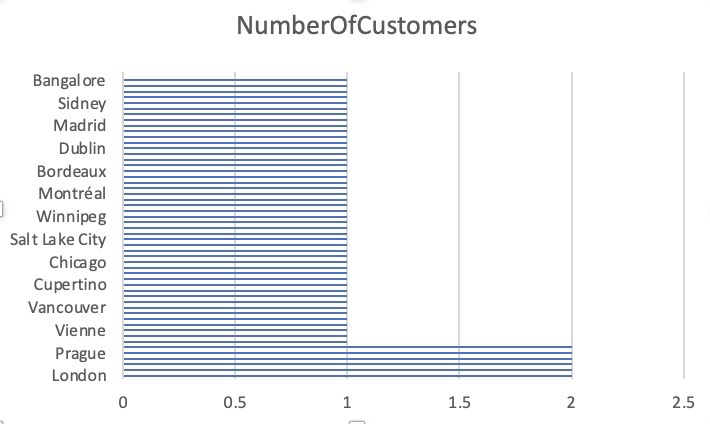
Basis of City:

SELECT city as city, COUNT(customer\_id) AS NumberOfCustomers

FROM customer

GROUP BY city

ORDER BY NumberOfCustomers DESC;



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

The total revenue, invoices for each country. It can be achieved by using the below query.

Revenue VS Country:

SELECT c.country,

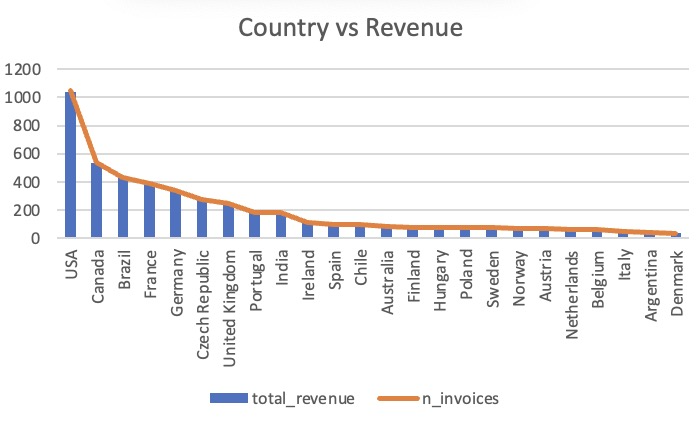
SUM(il.unit\_price \* il.quantity) AS total\_revenue,

COUNT(il.invoice\_id) AS n\_invoices

FROM invoice I JOIN invoice\_line il ON il.invoice\_id = i.invoice\_id JOIN customer c ON c.customer\_id = i.customer\_id

GROUP BY c.country

ORDER BY total\_revenue DESC, n\_invoices DESC;



The total revenue, invoices for each country. It can be achieved by using the below query.

Revnue vs State:

SELECT COALESCE(c.state, 'N/A') AS state,

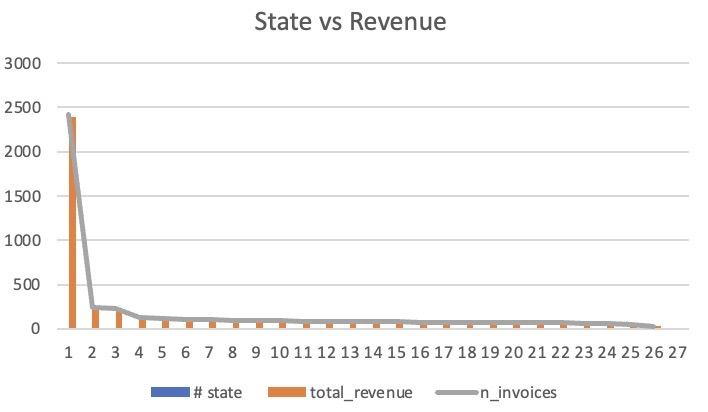
SUM(il.unit\_price \* il.quantity) AS total\_revenue,

COUNT(il.invoice\_id) AS n\_invoices

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

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

GROUP BY state

ORDER BY total\_revenue DESC, n\_invoices DESC;

The total revenue, invoices for each country. It can be achieved by using the below query.

Revnue vs City:

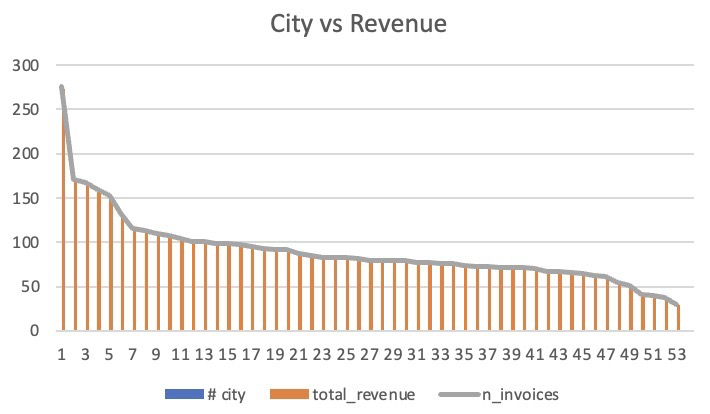
SELECT

COALESCE(c.city, 'N/A') AS city,SUM(il.unit\_price \* il.quantity) AS total\_revenue, COUNT(il.invoice\_id) AS n\_invoices

FROM invoice I JOIN invoice\_line il ON il.invoice\_id = i.invoice\_id JOIN customer c ON c.customer\_id = i.customer\_id

GROUP BY city

ORDER BY total\_revenue DESC, n\_invoices DESC;



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

To calculate the total revenue for each country and rank them in order

and sort by selecting the top 5 ranked customers.

MYSQL QUERY:

with cte1 as

(select customer\_id,billing\_country,sum(total) as customer\_total,rank() over(partition by billing\_country order by sum(total) desc) as rank

from invoice

group by customer\_id,billing\_country)

select customer\_id,billing\_country,customer\_total,rank

from cte1

where rank <=5

order by billing\_country asc



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



MYSQL QUERY:

with ranked\_tracks as(

select c.customer\_id, c.first\_name, c.last\_name, t.track\_id, t.name as track\_name, sum(il.quantity)as total\_sales from customer c

join invoice i on i.customer\_id=c.customer\_id

join invoice\_line il on il.invoice\_id=i.invoice\_id

join track t on t.track\_id=il.track\_id

group by c.customer\_id,c.first\_name,c.last\_name,t.track\_id,t.name

)

select customer\_id,first\_name,last\_name,track\_id,track\_name,total\_sales

from ranked\_tracks

order by total\_sales desc;

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

After considering many parameters we can observe that there is a significant decrease in the purchasing behaviour of some customers

and increase in purchasing behaviour of some customers.



MYSQL QUERY:

SELECT

c.customer\_id,

CONCAT(c.first\_name, ' ', c.last\_name) as customers,

YEAR(i.invoice\_date) AS year,

COUNT(i.invoice\_id) AS purchase\_count

FROM

customer c

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

GROUP BY

c.customer\_id, customers, YEAR(i.invoice\_date)

ORDER BY

c.customer\_id, customers, YEAR(i.invoice\_date);

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

The customer churn rate is **1.6949.**

WITH MostRecentInvoice AS (

SELECT

MAX(invoice\_date) AS most\_recent\_invoice\_date

FROM

invoice

),

CutoffDate AS (

SELECT

DATE\_SUB((SELECT most\_recent\_invoice\_date FROM MostRecentInvoice), INTERVAL 1 YEAR) AS cutoff\_date

),

ChurnedCustomers AS (

SELECT

c.customer\_id,

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

HAVING

last\_purchase\_date IS NULL OR last\_purchase\_date < (SELECT cutoff\_date FROM CutoffDate)

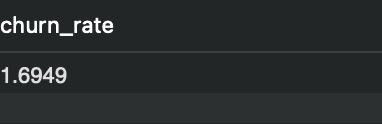
)

SELECT

(SELECT COUNT(\*) FROM ChurnedCustomers) / COUNT(\*) \* 100 AS churn\_rate

FROM

customer;



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

Observation:

The most popular genre is ROCK.

MYSQL QUERY:

WITH genre\_counts AS (

SELECT

g.name AS genre\_name,

COUNT(g.genre\_id) AS genre\_count

FROM

customer c

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

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

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

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

WHERE

c.country = 'USA'

GROUP BY

g.name

),

total\_count AS (

SELECT

COUNT(g.genre\_id) AS total\_count

FROM

customer c

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

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

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

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

WHERE

c.country = 'USA'

)

SELECT

gc.genre\_name,

gc.genre\_count,

(gc.genre\_count \* 100 / tc.total\_count) AS percentage

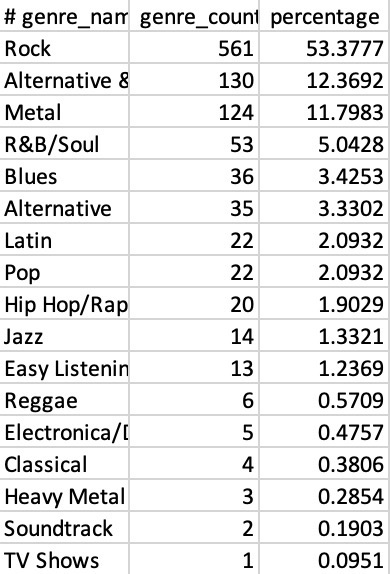
FROM

genre\_counts gc

CROSS JOIN total\_count tc

ORDER BY

gc.genre\_count DESC;



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

We can find the customers who have purchased different tracks by counting the distinct genres.

MYSQL QUERY:

SELECT

c.first\_name, count(distinct g.name) AS genre\_count

FROM

customer c

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

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

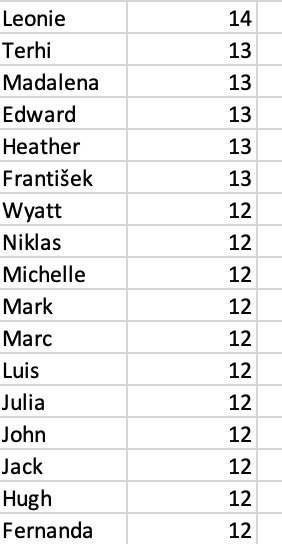
INNER JOIN track t ON il.track\_id = t.track\_id

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

GROUP BY c.first\_name

HAVING count(distinct g.name) >=3

ORDER BY genre\_count DESC



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

TOP 5 Genres in USA are:



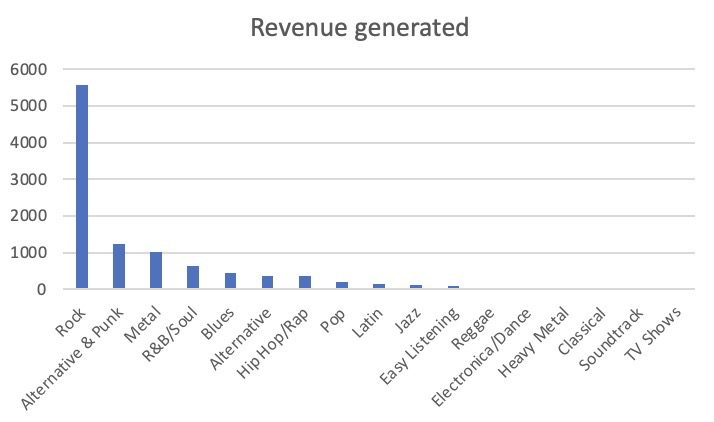
select g.name,sum(i.total) as genre\_sum, rank() over(order by sum(i.total) desc) as rankk

from customer c inner join invoice i on c.customer\_id = i.customer\_id inner join invoice\_line il on i.invoice\_id = il.invoice\_id inner join track t on il.track\_id = t.track\_id

inner join genre g on t.genre\_id = g.genre\_id

where c.country = "USA"

group by g.name



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

To identify the customers who have not made a purchase in the last 3 months we need to filter the customer ids based on the invoice date.

The below query can be used:

WITH cte1 as

(SELECT customer\_id,invoice\_date,rank() over(partition by customer\_id order by invoice\_date desc) as rankk

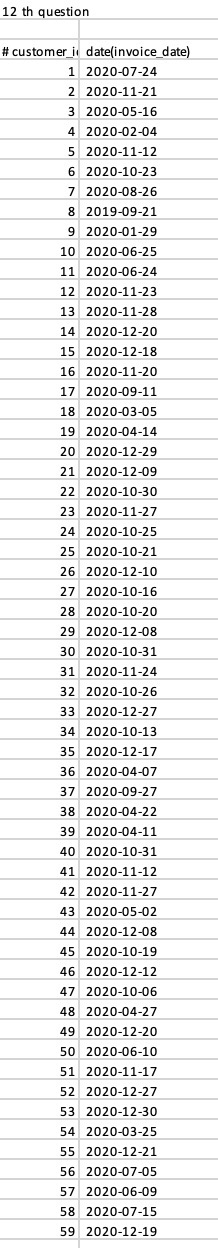
FROM invoice

WHERE invoice\_date<'2024-05-09')

SELECT customer\_id,date(invoice\_date)

FROM cte1

WHERE rankk=1



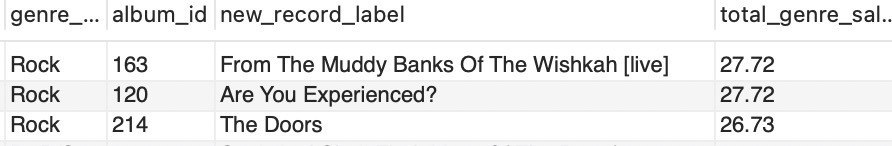
Subjective Questions

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.**

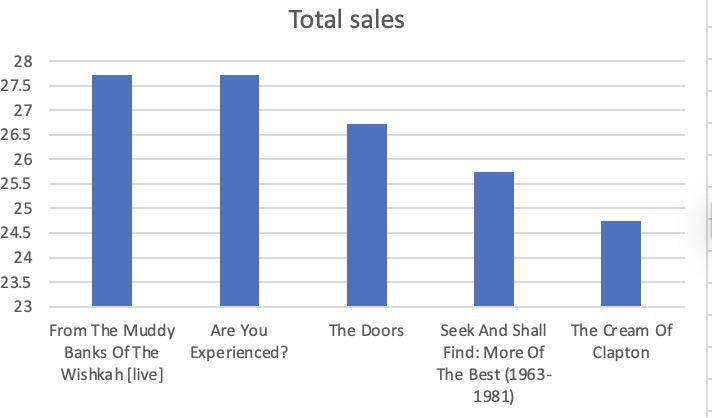
Upon analysis of the genre sales, we found that the top 3 genres based on sales are:



We cannot sort if the record labels are new or old because no date was specified in the dataset provided. So, let’s take the albums from the top 3 genres in USA.



These three albums are generating the most buzz which is directly proportional to the revenue generated. These are the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.



Code:

SELECT

g.genre\_id,

g.name AS genre\_name,

al.album\_id,

al.title AS new\_record\_label,

SUM(il.unit\_price \* il.quantity) AS total\_genre\_sales,

DENSE\_RANK() OVER (ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS Ranking

FROM

genre g

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

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

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

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

INNER JOIN album al on t.album\_id = al.album\_id

WHERE

c.country = 'USA'

GROUP BY

g.genre\_id, g.name, al.album\_id,

al.title

ORDER BY

total\_genre\_sales DESC;

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

Approach:

First, we find the top genres in USA based on the below provided code snippets and then we also find the most popular genres in the rest of the countries. India and France are some countries where the unique customer base is located so special attention has been provided.

Top selling genres in USA:



Code snippet:

SELECT

g.name,

SUM(i.total) AS genre\_sum,

RANK() OVER (ORDER BY SUM(i.total) DESC) AS rankk

FROM

customer c

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

WHERE

c.country != 'USA'

GROUP BY

g.name;

Rock is the Top-selling genre in many countries they are:



Top selling genres in India:



Code snippet:

SELECT

g.name,

SUM(i.total) AS genre\_sum,

RANK() OVER (ORDER BY SUM(i.total) DESC) AS rankk

FROM

customer c

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

WHERE

c.country = ‘India’

GROUP BY

g.name;

Top selling genres in France:



Code snippet:

SELECT

g.name,

SUM(i.total) AS genre\_sum,

RANK() OVER (ORDER BY SUM(i.total) DESC) AS rankk

FROM

customer c

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

INNER JOIN

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

WHERE

c.country = 'France'

GROUP BY

g.name;

**Commonalities:**

* Rock dominates 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 India, France, Brazil, showing a more diverse taste in music genres in that country.

**Differences:**

* France has a different taste in music they prefer genres like Easy listening and Metal.
* Metal has a notable presence in Canada and Brazil but does not appear as a top genre in other countries listed.

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?**

Insights:

* Long Term customers tend to purchase more frequently
* Long Term customers spend more and have a larger basket size compare to short term customers
* Long Term customers have large basket size and more order value compared to short term customers.

Customer Loyalty and Retention Strategies:

* Use customer data to personalize marketing messages, product recommendations, and services.
* Create a loyalty program that rewards customers for repeat purchases or referrals.
* Create valuable content (blogs, videos, newsletters) that educates and engages customers.
* Use email marketing campaigns to re-engage inactive customers.
* Use automated tools for personalized email campaigns, follow-ups, and loyalty programs.

WITH Customerinsights AS (

SELECT

c.customer\_id,

COUNT(i.invoice\_id) AS purchase\_frequency,

SUM(il.quantity) AS total\_items\_purchased,

SUM(i.total) AS total\_spent,

AVG(i.total) AS avg\_order\_value,

DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) AS customer\_tenure\_days

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.customer\_id

),

Customersegment AS (

SELECT

customer\_id,

purchase\_frequency,

total\_items\_purchased,

total\_spent,

avg\_order\_value,

customer\_tenure\_days,

CASE

WHEN customer\_tenure\_days >= 365 THEN 'Long-Term'

ELSE 'New'

END AS customer\_segment

FROM

Customerinsights

)

SELECT

customer\_segment,

ROUND(AVG(purchase\_frequency),2) AS avg\_purchase\_frequency,

ROUND(AVG(total\_items\_purchased),2) AS avg\_basket\_size,

ROUND(AVG(total\_spent),2) AS avg\_spending\_amount,

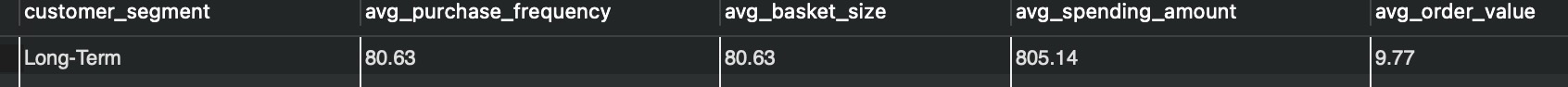
ROUND(AVG(avg\_order\_value),2) AS avg\_order\_value

FROM

Customersegment

GROUP BY

customer\_segment;



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 factors?**

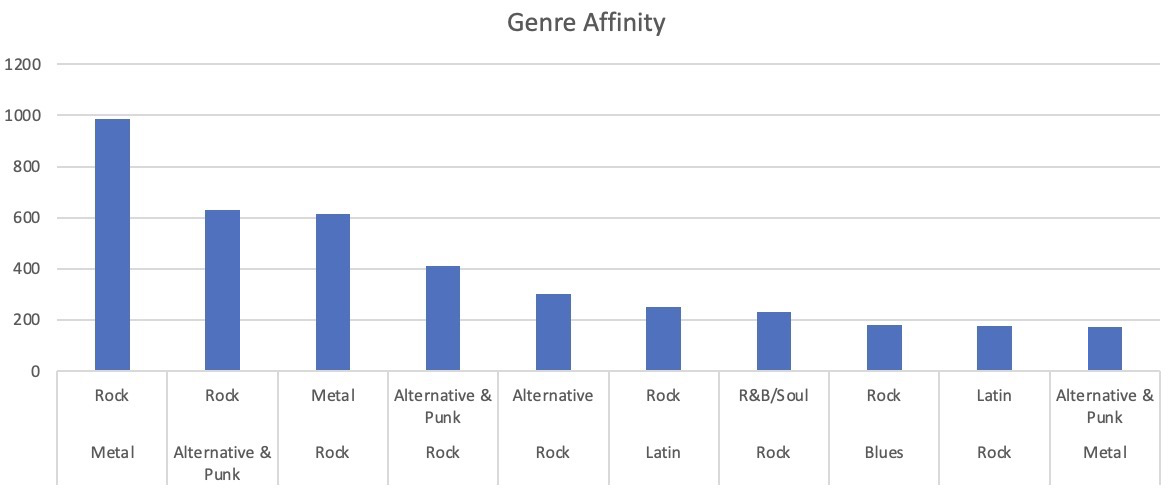
**Recommendations and cross-selling initiatives:**

* Bundle offers and Combo offers should be provided on the genres such as rock and metal, rock and alternative & punk which would increase overall sales.
* Money should be spent on advertising these genres together to boost sales.
* Provide sponsorships for concerts of the artists to perform together to get more reach to the brand and increase brand value.
* Make custom playlists by mixing all the albums which have the most affinity.

**Product Affinity Analysis:**

Now we divide the entire product affinity into three sub categories they are genre affinity, artist affinity and album affinity.

**Genre Affinity:**

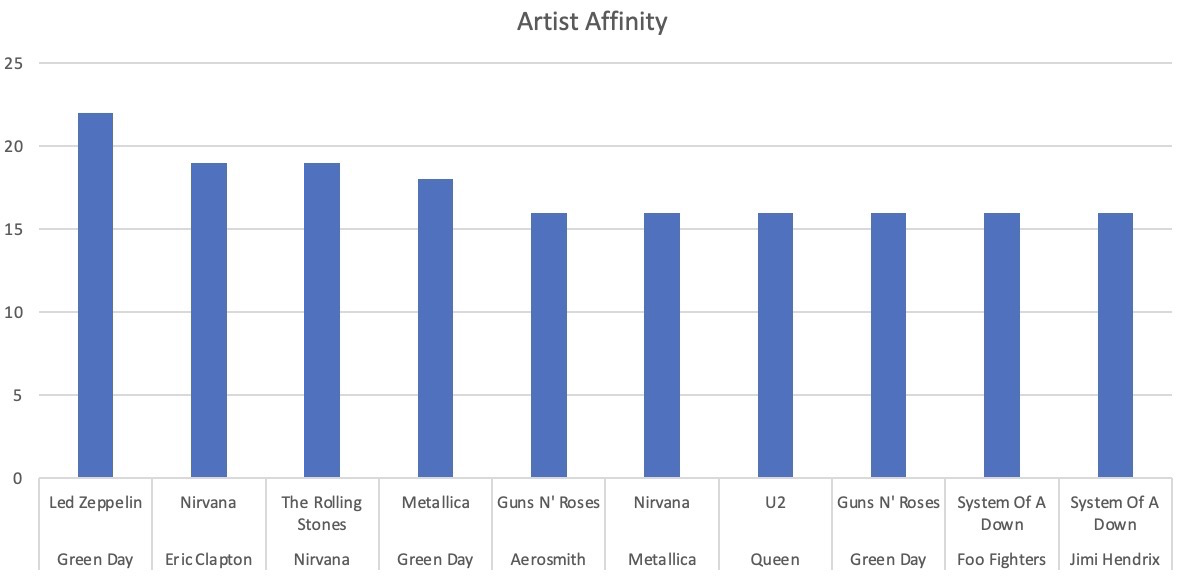


* Rock and Metal are the genres which are purchased together the most with a purchase count of 933.
* Alternative & Punk and Rock: Purchased together 629 times
* Rock and Metal: Purchased together 613 times
* Rock and Alternative & Punk: Purchased together 412 times

Recommendations:

* Bundle offers and Combo offers should be provided on the genres such as rock and metal, rock and alternative & punk which would increase overall sales.
* Money should be spent on advertising these genres together to boost sales.

**Artist Affinity:**

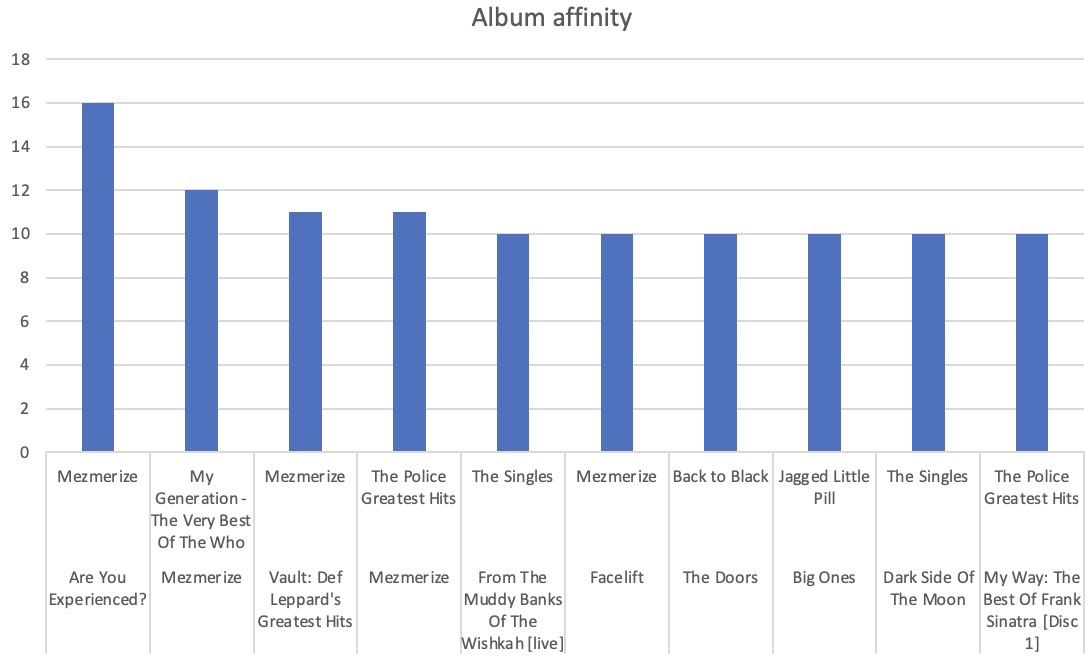


* Led Zeppelin and Green Day are the artists which are the most purchased together 22 times.
* Eric Clapton and Nirvana: Purchased together 19 times
* Nirvana and The Rolling Stones: Purchased together 19 times
* Green Day and Metallica: Purchased together 18 times

Recommendations:

* Pair up the artists and provide Bundle offers and Combo offers such as led zeppelin and green day, Eric Clapton and nirvana, Nirvana and Rolling stones.
* Provide sponsorships for concerts of the artists to perform together to get more reach to the brand and increase brand value.

**Album Affinity:**



The albums which are purchased together the most are Mezmerize and are you experienced? They are purchased together 22 times.

Recommendations:

* Promote the albums such as Mezmerize and Are you experienced? together focus more on bill boards to advertise these albums.
* Make custom playlists by mixing all the albums which have the most affinity.

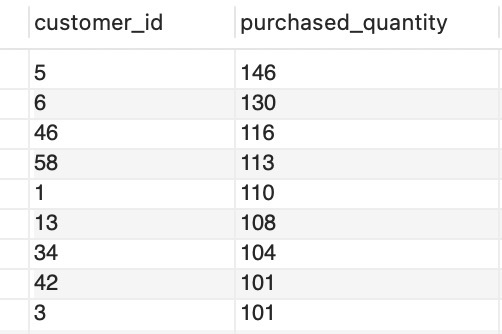
1. **Regional Market Analysis: Do customer purchasing behaviours and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?**

Regional Market Analysis:

Upon analysing the market and dividing the market into three segments like low, high and mid valued customer segments.

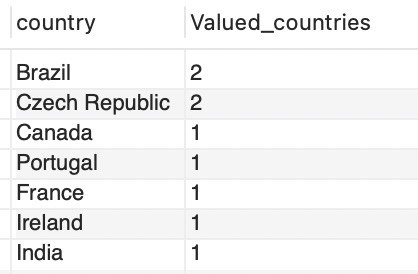
The high valued customers are from Brazil, Czech Republic, Canada, Portugal, France, Ireland and India.

First, we try to find who are our high value customers by filtering the customers purchasing quantity. We are considering the customers whose basket size is more than 100.

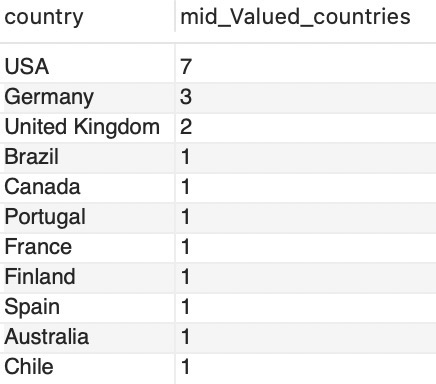


Now considering the customers who purchased between 80- 100 as medium valued customers and the customers below 80 as low value customers.

Now let’s try to find the countries in which the most valued customers are from and sort them based on categories.



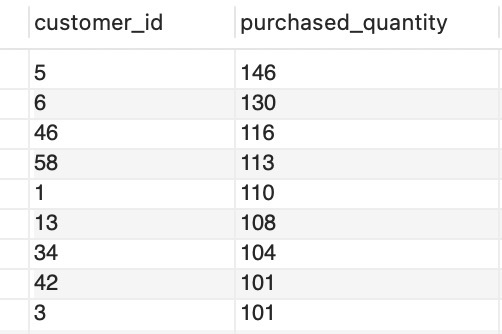
Medium valued customers countries:



Low valued customers countries:



So, from the above analysis we can say that the high valued customers are from Brazil, Czech Republic, Canada, Portugal, France, Ireland and India.



**Correlation with local demographic or economic factors:**

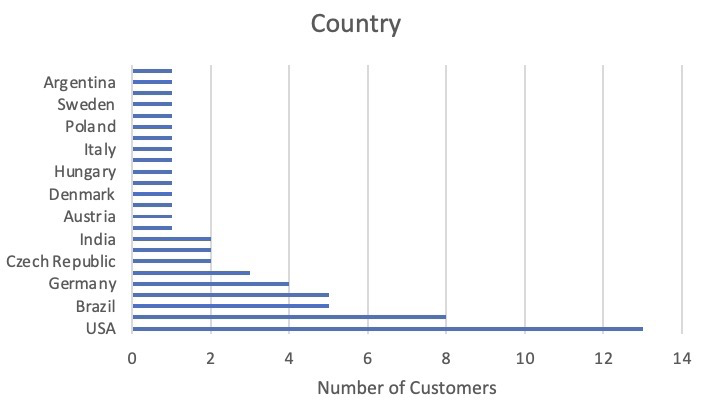
* Countries with higher GDP per capita and disposable income (e.g., Canada, Ireland, and France) tend to have customers who spend more on entertainment and digital products like music.
* Countries with a rich music culture and a strong preference for both local and international music (e.g., Brazil, France, and India), customers are more engaged and likely to purchase music.
* Countries with a more mature digital economy and e-commerce infrastructure (e.g., Ireland, Czech Republic, and Portugal) provide an environment conducive to higher online spending on music.
* Countries with higher GDP per capita and disposable income (e.g., Canada, Ireland, and France) tend to have customers who spend more on entertainment

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?**

As no data is available on the customers location and age the only parameter, we can consider to segment them is again the demographics of the customer and then further classify them based on the state and its location.

Customer can be divided into three areas:

* **High-Value Areas:** Prioritize exclusive products, adopt premium pricing strategies, and implement personalized marketing campaigns to maximize customer value.
* **Moderate Areas:** Enhance customer engagement with loyalty programs, regular communication, and attractive value-for-money offers.
* **High Churn Areas:** Identify the causes of customer churn and develop retention strategies, such as offering discounts, improving customer service, and using satisfaction surveys to gather feedback.





The above charts show the demographics of the low valued customers location. The customers from these locations have purchased the least and are most likely to churn. So, we need to run Ad campaigns in order to increase customers in such areas.

1. **Customer Lifetime Value Modelling: 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?**

To leverage the customer data and predict customer lifetime value and identifying patterns among customers who have stopped purchasing. The following are the steps that needs to be looked into:

* Tenure: It helps to calculate the duration of the customer relationship to identify common characteristics and patterns among customers.
* Purchase History: Purchase history help in calculating the total sales of the particular customer which in turn helps us to determine if the customer is a high valued customer or a low valued customer which helps a lot in customer segmentation.
* Engagement: Customer engagement is a key factor in purchasing behaviour and it would also effect the trend of sales.
* Churn rate: Calculate the time since the last purchase for each customer.

Customer data such as tenure, purchase history and engagement are very helpful to predict the lifetime value of different customer segments. Churn analysis could also be effectively done by using the customer data.

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?**

We currently do not have direct access to data on promotional campaigns, such as discounts, events, or email marketing. However, if this data were available, we could evaluate the impact of these campaigns on customer acquisition, retention, and overall sales using the following method:

**Method for Evaluating Impact:**

**1. Determine Promotion Timeframes:**

* Use a data source or field that logs the dates or periods when promotions were conducted.

**2. Separate Data Sets:**

* Divide the data into two categories: periods with promotions and periods without promotions.

**3. Measure Customer Acquisition:**

* Compare the number of new customers gained during promotional periods against those gained during non-promotional periods.

**4. Assess Customer Retention:**

* Analyze repeat purchase rates to evaluate customer retention during and following promotional periods.

**5. Compare Sales Metrics:**

* Examine metrics such as total sales, average order value, and basket size for promotional periods compared to non-promotional periods.

This method would help us understand the effectiveness of promotional efforts in achieving desired business objectives.

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

If no predefined questions are provided, the following approach involves a systematic, exploratory analysis of the available data:

**Steps for Exploratory Data Analysis**

**1. Review Data Structure:**

* Identify key entities and relationships (e.g., Employee, Customer, Invoice).
* Determine critical metrics such as sales performance and customer demographics.

**2. Perform Data Exploration and Quality Check:**

* Check for missing values, duplicates, and inconsistencies.
* Clean the data by imputing missing values or removing incomplete records.

**3. Analyze Sales Performance:**

* Examine total sales, average order value, and top-selling products across different time periods, regions, and customer segments.

**4. Conduct Customer Analysis:**

* Analyze customer demographics and segment customers based on purchase history.
* Calculate churn rates and estimate Customer Lifetime Value (CLV).

**5. Evaluate Market and Regional Trends:**

* Assess sales performance across different regions and correlate with external factors, such as local demographics.

**6. Identify Product Affinity and Cross-Selling Opportunities:**

* Detect frequently purchased product combinations to identify cross-selling opportunities.

**7. Analyze Promotional Effectiveness:**

* Measure the impact of promotional activities on sales and customer behavior.

**8. Develop Strategic Insights:**

* Identify growth opportunities, refine pricing strategies, and design targeted marketing plans.

**9. Document and Report Findings:**

* Summarize insights with visualizations and provide actionable recommendations.

**Conclusion**

This approach involves a thorough exploration of data to uncover insights that can guide strategic decisions, improve customer retention, optimize sales, and identify growth opportunities. The process is iterative and continually evolves as new insights emerge.

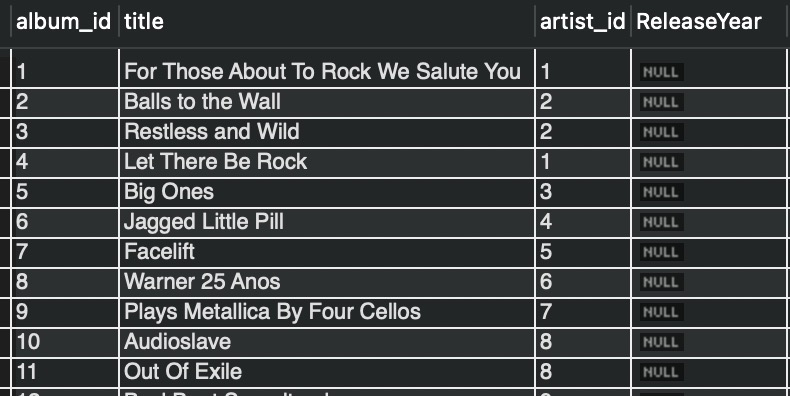
1. **How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?**

We can alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album by using the following SQL query:

ALTER TABLE Album

ADD COLUMN ReleaseYear INTEGER;

select \* from album;



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.**

**SQL QUERY:**

WITH customer\_tracks AS (

SELECT

i.customer\_id,

SUM(il.quantity) AS total\_tracks

FROM

invoice i

JOIN

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

GROUP BY

i.customer\_id

),

total\_customer\_spending AS (

SELECT

c.country,

c.customer\_id,

SUM(i.total) AS total\_spent,

ct.total\_tracks

FROM

customer c

JOIN

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

JOIN

customer\_tracks ct ON c.customer\_id = ct.customer\_id

GROUP BY

c.country, c.customer\_id, ct.total\_tracks

)

SELECT

cs.country,

COUNT(DISTINCT cs.customer\_id) AS number\_of\_customers,

ROUND(AVG(cs.total\_spent),2) AS average\_amount\_spent\_per\_customer,

ROUND(AVG(cs.total\_tracks),2) AS average\_tracks\_purchased\_per\_customer

FROM

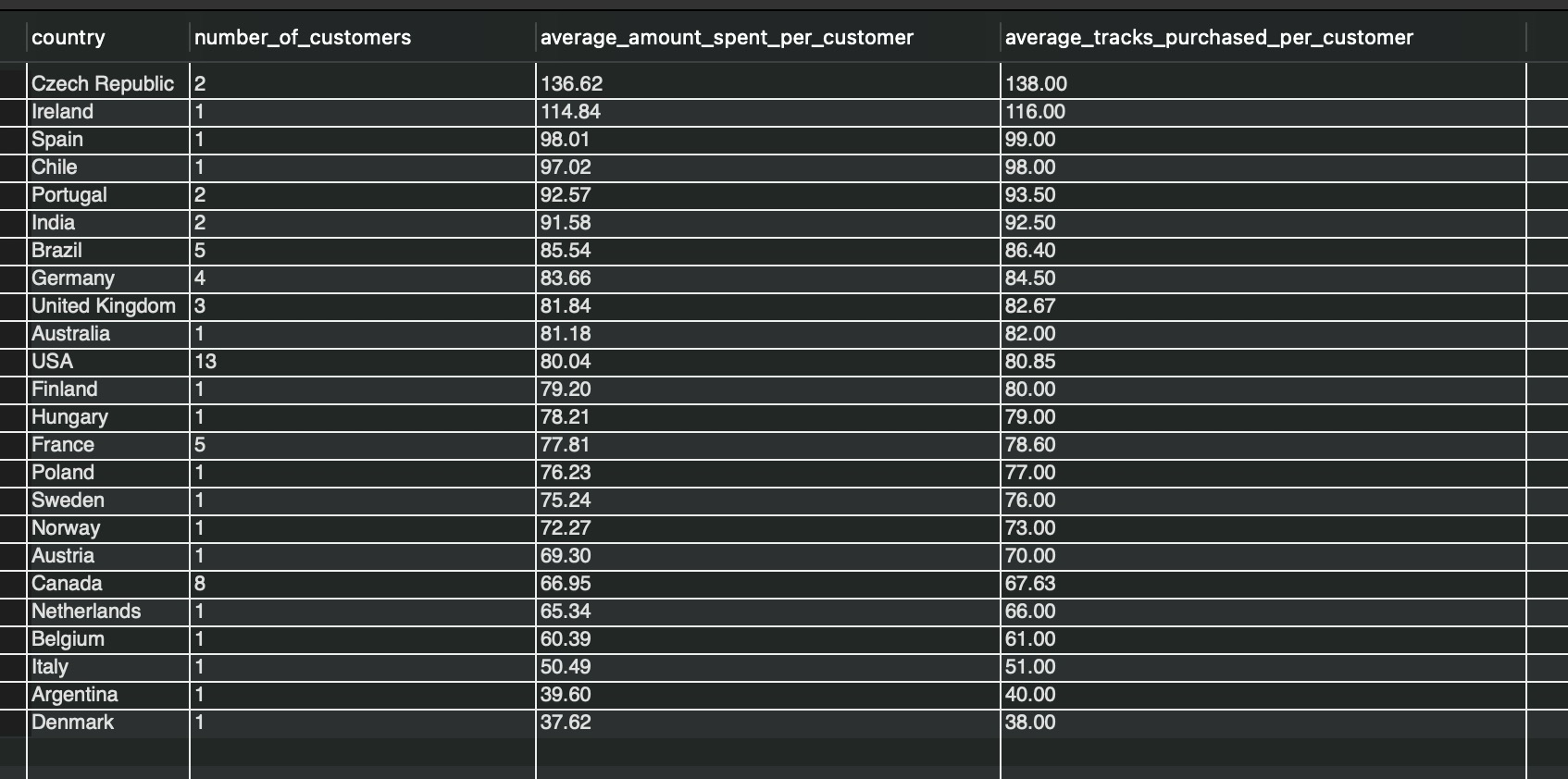
total\_customer\_spending cs

GROUP BY

cs.country

ORDER BY

average\_amount\_spent\_per\_customer DESC;

****

**Conclusion:**

* European Countries like Denmark, Italy, Belgium, Netherlands are the countries where the lowest number of customers are located.
* Chinook Music store has the greatest number of customers are from USA followed by Canada.
* Czech Republic has the highest amount spent per customer among all the countries.