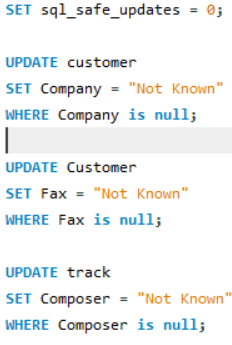
1. Does any table have missing values or duplicates? If yes how would you handle it ?

## **OBJECTIVE QUESTION SOLUTIONS**

REFERENCE- - “Objective question solution” named sql file , from 5 to 19

rows

ANSWER-



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

REFERENCE- “Objective question solution” named sql file , from 21 to 85

rows

ANSWER-

PART 1 – Top 5 selling tracks in USA

A screenshot of a computer

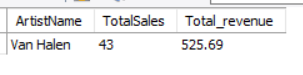
AI-generated content may be incorrect.

VISUALIZATION-

APPROACH-

1. After analyzing the schema useful table are invoice , invoice\_line , track ,artist ,genre , album
2. Join all the table above and group the data based on track\_id and count the volume of track saled using invoice\_id
3. Limit the result to 5 to see top selling track with respective artist name and there sale’s volume and total revenue

PART 2- Top artist in USA



APPROACH-

1. Usefull table are invoice , invoice\_line ,track ,album ,artist
2. Joining all table above and group the data on the basis of artist\_id and calculate the volume of sales generated by each artist and filter the result for USA
3. Limit the result to 1 to see the top artist in USA

PART 3 – Top genre in USA for Top artist in USA

A screenshot of a computer

AI-generated content may be incorrect.

APPROACH-

1. Usefull table are invoice , invoice\_line , track , album , artist , genre
2. After joining all table above , grouped the data base on genre\_id and calculate TotalSales(count of invoiceId as volume) and total revenue
3. Filter the result with top artist in USA using cte and subquery

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

REFERENCE - “Objective question solution” named sql file , from 87 to 94

rows

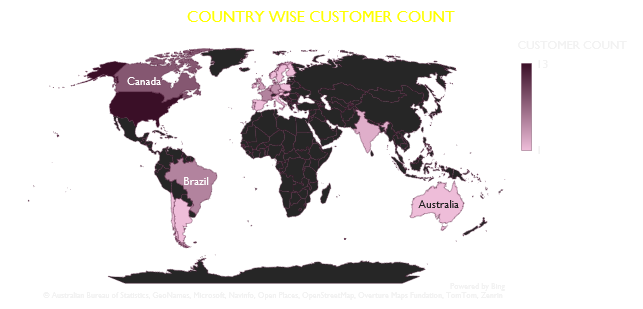
ANSWER-

# As there are no age and gender provided in database so only

location wise breakdown possible



VISUALISATION-



APPROACH –

* + 1. Using customer table group the data country wise and count the customer\_ID in respective country
    2. Now using map to show the breakdown of customer count

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

REFERENCE- “Objective question solution” named sql file , from 97 to 129

rows

ANSWER-

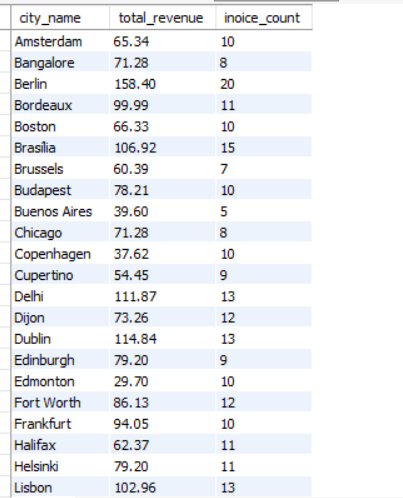
A screenshot of a computer

AI-generated content may be incorrect.

VISUALISATION

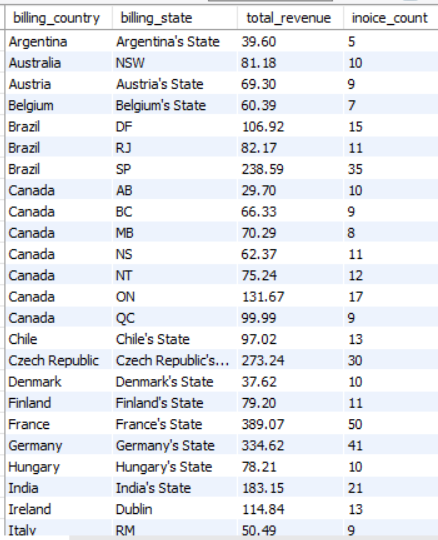
PART 2 city wise total revenue and invoice count

ANSWER-



VISUALISATION

PART 3 #. state wise total revenue and invoice count



VISUALISATION –

REFERENCE- - in objective question solution sql file from 46 to 52 row

APPROACH –

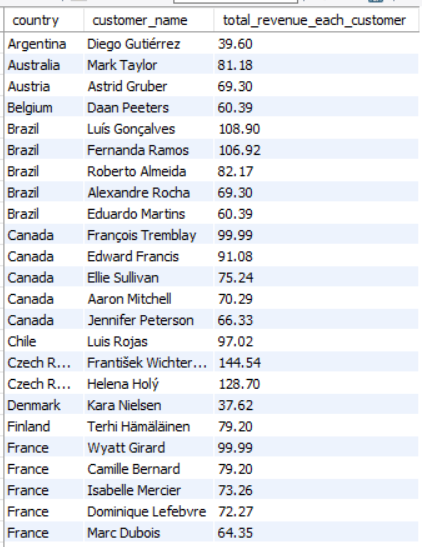
* + 1. Using table invoce , group the data based on billing country
    2. Then sum the total revenue and count the invoice id using aggregate function
    3. Then order the result based on billing country
    4. Similiarly done for city wise calculation
    5. While doing for state wise calculation , some state having None value in it , so for none value concating the country name with “state” as string in CASE WHEN clause
    6. And then group by billing\_country and billing\_state give the state wise total revenue and invoice count

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

REFERENCE- “Objective question solution” named sql file , from 131 to 185

rows

ANSWER-



REFERENCE – objective question solution sql file , from 56 to 80 row

APPROACH –

* + 1. Using cte with table invoice group the data based on billing\_country and customer\_id
    2. Now take the sum of total to get expenditure of each customer in each country which is nothing but total revenue
    3. Rank the total\_revenue using dense\_rank in descending order in each country using PARTITION BY and OVER clause
    4. Now using above cte to join with customer table to get name of each customer and order the result based on billing\_contry and total\_revenue

1. Identify the top-selling track for each customer

REFERENCE-“Objective question solution” named sql file , from 152 to 173

rows

ANSWER



APPROACH-

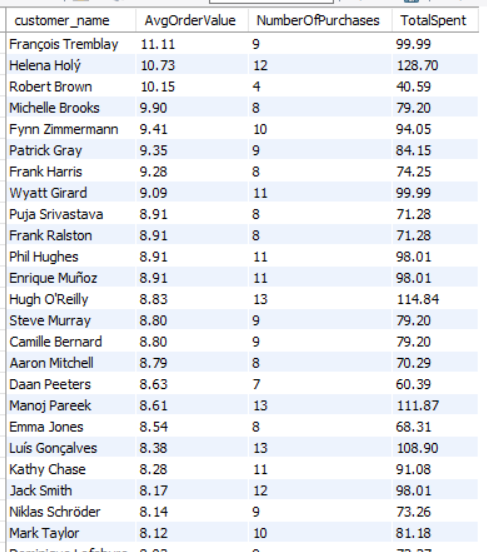
* + 1. With the help of cte joined the table invoice , invoice\_line , track on ivoice\_id and track\_id respectively
    2. Now using DENSE\_RANK() PARTITION BY function , ranked the total amount spend in descending order to know which customer spend highest money on which track
    3. Using cte print the customer name from customer table and track name from track table using subquery for rank = 1 to know top selling track of each customer

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

REFERENCE- “Objective question solution” named sql file , from 175 to 184

rows

ANSWER- Average Order Value and purchase Frequency are:



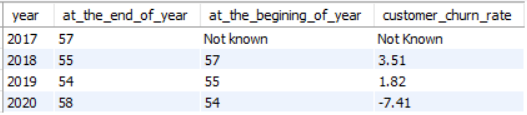
VISUALISATION-

1. What is the customer churn rate?

REFERENCE- “Objective question solution” named sql file , from 187 to 199

rows

ANSWER-



APPROACH

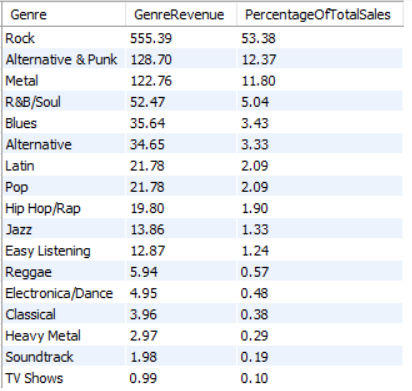
* + 1. Using invoice table group the data year wise using YEAR() function and count the distinct CustomerID in common table expression to know the yearly customer count
    2. Using cte with LAG() function at the beginning of year customer count calculated
    3. Now customer churn rate calculated by taking percentage of difference in customer count at the beginning of year and the end of year with respect to at the beginning of year to know how much customer lost at the end of year

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

REFERENCE- “Objective question solution” named sql file , from 201 to 235

rows

ANSWER-



VISUALISATION-

INSIGHTS

1. From above chart we can see that “rock” genre is most famous genre in USA

PART 2 – Top artist in USA

A screenshot of a computer

AI-generated content may be incorrect.

VISUALISATION -

#. As we can see from above chart Van Halen is top artist in USA

APPROACH-

* + 1. Use table invoice ,invoice\_line ,customer , track ,album ,genre
    2. Joining all table and group the data based on genre\_id and then sum the product of unit\_price and quantity of each genre\_id to calculate total sales of each genre
    3. Now calculate sum of total sales of all genre using subquery
    4. And then calculate sales percentage of each genre in USA
    5. Similiarly done to find top artist in USA only difference is grouping the data , here grouping is done on the basis of artist\_id

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

REFERENCE- “Objective question solution” named sql file , from 237 to 255

rows

ANSWER-



VISUALISATION -

APPROACH

* + 1. With cte join the invoice table ,invoice\_line table and track table
    2. Now grouped the data on the basis of customer\_id and count of distinct genre\_id of each customer
    3. Now using cte selecting the customer’s first name and last name and concat them as full name using subquery from customer table
    4. And filter in where clause with genre count reater than eual 3 to get the answer
    5. Now using column chart to visualize the data we get from mysql workbench

INSIGHT –

1. **Top genre counts:** Leonie Köhler have the highest genre count (14), indicating a broad range of associations.
2. Robert Brown having the lowest count at 4
3. Rank genres based on their sales performance in the USA

REFERENCE- “Objective question solution” named sql file , from 257 to 270

rows

ANSWER-



VISUALISATION-

#. As we can see the “Rock” genre having first rank in USA

APPROACH-

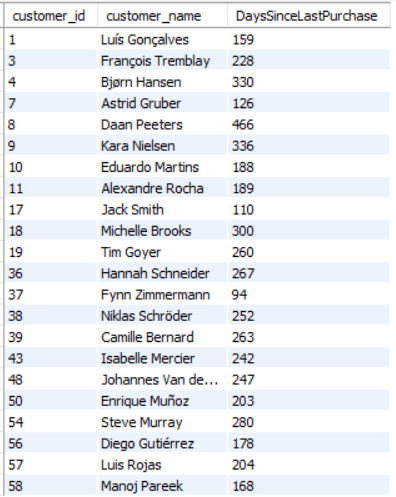
* + - 1. **Join tables**: Invoice → Customer → Invoice\_Line → Track → Genre
      2. **Filter**: Only customers from the **USA**
      3. **Aggregate**: Sum up sales using Unit\_Price \* Quantity
      4. **Group**: By Genre
      5. **Rank**: Order genres by total revenue

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

REFERENCE – “Objective question solution” named sql file , from 273 to 284

rows

ANSWER-



APPROACH –

1. With cte join the table invoice and customer table
2. In cte we find the customer who made purchase in last three month
3. In where clause we filter the customer who purchase in December , November , October
4. Now we select customer from invoice table and filtered the customer who do not made any purchase in last 3 month in where clause using cte (last\_3\_month) as subquery
5. And find the customer name full name using subquery and concat function from customer table

SUBJECTIVE QUESTION SOLUTION

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.

REFERENCE- “Objective question solution” named sql file , from 289 to 307

Rows

QUERY-

SELECT

al.Album\_Id,

al.Title AS AlbumTitle,

ar.Name AS Artist\_Name,

ROUND(SUM(il.Unit\_Price \* il.Quantity), 2) AS TotalSalesUSD,

g.Name AS Genre

FROM Invoice\_Line il

JOIN Invoice i ON il.Invoice\_Id = i.Invoice\_Id

JOIN Customer c ON i.Customer\_Id = c.Customer\_Id

JOIN Track t ON il.Track\_Id = t.Track\_Id

JOIN Album al ON t.Album\_Id = al.Album\_Id

JOIN Artist ar ON al.Artist\_Id = ar.Artist\_Id

JOIN Genre g ON t.Genre\_Id = g.Genre\_Id

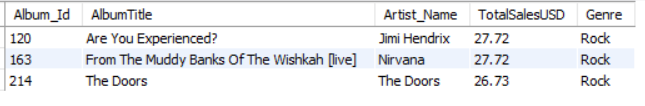
WHERE c.Country = 'USA'

GROUP BY al.Album\_Id , al.Title , ar.Name , g.Name

ORDER BY TotalSalesUSD DESC

LIMIT 3;

ANSWER-



APPROACH-

**1. Identify Top-Selling Genres in the USA**

**2. Find Albums in These Genres**  
 → Join Album, Track, and Genre.

**3. Filter Tracks Purchased by USA Customers**  
 → Use Customer.Country = 'USA'.

**4. Aggregate Sales by Album**  
 → Sum of InvoiceLine.UnitPrice \* Quantity.

**5. Select Top 3 Albums**  
 → Based on total revenue from USA customers.

INSIGHTS-

* 1. All top-performing albums fall under the **Rock** genre and show **very close Total Sales USD**, indicating a strong affinity for classic rock among U.S. consumers. This genre continues to maintain commercial viability, especially through legendary artists with enduring fan bases.

RECOMMENDATION –

* + 1.  **Leverage Nostalgia with Hendrix & The Doors**
* Target classic rock radio, vinyl collectors, and Gen X listeners.
* Use storytelling ads that highlight each album's historical influence.
  + 1.  **Capitalize on Grunge Legacy with Nirvana**
* Integrate with youth-focused platforms like TikTok and Spotify playlists.
  + - * Collaborate with rock influencers to reintroduce live performance energy.
    1.  **Bundle Promotions by Genre**
       - Launch a “Legendary Rock Revival” campaign featuring all three albums.
       - Offer limited-edition box sets or digital bundles to boost cross-sales.

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

REFERENCE- “Objective question solution” named sql file , from 309 to 350

Rows

QUERY 1 –

-- for usa top genre

select

i2.billing\_country as country,

t1.genre\_id,

(select name from genre where genre\_id = t1.genre\_id) as genre\_name,

sum(i1.Unit\_Price\*i1.Quantity) total\_revenue\_each\_genre

from invoice\_line i1

join invoice i2

on i1.invoice\_id = i2.invoice\_id

join track t1

on i1.track\_id = t1.track\_id

where i2.billing\_country = "USA"

group by t1.genre\_id , genre\_name , country

order by total\_revenue\_each\_genre desc

limit 1;

ANSWER-

A close up of a computer screen

AI-generated content may be incorrect.

#. For USA top genre is “Rock”

QUERY 2

WITH TOP\_GENRE AS (

SELECT

c.Country,

g.Name AS Genre,

ROUND(SUM(il.Unit\_Price \* il.Quantity), 2) AS TotalSales,

RANK() OVER (PARTITION BY c.Country ORDER BY ROUND(SUM(il.Unit\_Price \* il.Quantity),2) DESC) AS genre\_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 Genre g ON t.Genre\_Id = g.Genre\_Id

WHERE c.Country != 'USA'

GROUP BY c.Country, g.Name

ORDER BY c.Country, TotalSales DESC

)

SELECT

Country,

Genre,

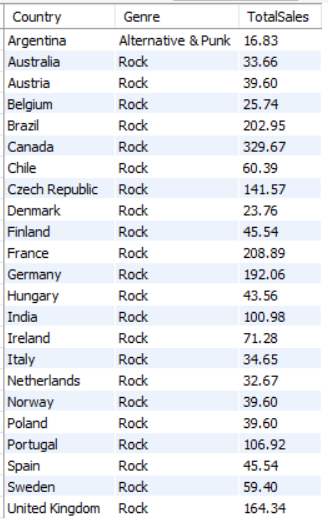
TotalSales

FROM TOP\_GENRE

WHERE genre\_rank = 1

ORDER BY Country;

ANSWER



#. As we can see top genre in other country is same as USA except

Argentina

APPROACH-

**1. Exclude the USA** from the analysis.

2. Join these relevant tables:

Customer → Invoice → InvoiceLine → Track → Genre

3. Group by **Country** and **Genre**.

4. Aggregate sales using either:

1. SUM(InvoiceLine.UnitPrice \* Quantity) *(accurate)*, or
2. SUM(Invoice.Total) *(if grouped carefully by Invoice)*

5. Rank genres per country.

6. Extract the **top-selling genre per country**.

7. Analyze **patterns** across countries.

INSIGHTS-

1. **Rock Has Global Dominance (Except Argentina)**

* Rock is the highest-selling genre in every country shown, except **Argentina**, which favors R**Alternative & Punk**.
* North American and European countries show **strong commercial traction** for Rock, often surpassing 100 USD in sales.

2. **Sales Intensity Varies by Region**

* Countries like **Canada**, **France**, **Germany**, and the **UK** stand out with high aggregate Rock sales.
* Northern European countries like **Denmark** and **Sweden** show lower total sales but still align in genre preference.

3. **Argentina Breaks the Pattern**

* Its preference for **Alternative & Punk** highlights a regional difference that could stem from cultural, socio-political, or generational preferences.

RECOMMENDATION –

1. **Double Down on Rock for Core Markets**

* **Canada, France, Germany, UK** are clear rock strongholds.
* Recommend exclusive releases, rock-themed festivals, and artist collaborations in these regions.

2. **Tailor Promotion for Latin America**

* Argentina offers a **differentiated audience**. Promote **Alternative & Punk** with local artists, youth-targeted campaigns, and rebellious-themed branding.

3. **Boost Rock in Underperforming Markets**

* For **Denmark and Sweden**, consider localized content, bundled playlists, and cross-platform exposure to elevate listener engagement.

1. 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?

REFERENCE - “Objective question solution” named sql file , from 352 to 380

Rows

QUERY- NEXT PAGE

WITH CustomerFirstPurchase AS (

SELECT

c.Customer\_Id,

year(Max(I.Invoice\_Date))-Year(MIN(i.Invoice\_Date)) AS year\_frequency

FROM Customer c

JOIN Invoice i ON c.Customer\_Id = i.Customer\_Id

GROUP BY c.Customer\_Id

),

CustomerSegment AS (

SELECT

Customer\_Id,

case

when year\_frequency = (select max(year(invoice\_date)) - min(year(invoice\_date)) from invoice) then "long\_term\_customer"

else "new\_customer"

end as type\_of\_customer

FROM CustomerFirstPurchase

)

SELECT

cs.type\_of\_customer,

COUNT(DISTINCT i.Invoice\_Id) AS PurchaseFrequency,

ROUND(AVG(il.Quantity), 2) AS AvgBasketSize,

ROUND(SUM(i.Total), 2) AS TotalSpending,

ROUND(AVG(i.Total), 2) AS AvgSpendingPerInvoice

FROM CustomerSegment cs

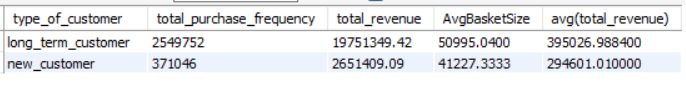
JOIN Invoice i ON cs.Customer\_Id = i.Customer\_Id

JOIN Invoice\_Line il ON i.Invoice\_Id = il.Invoice\_Id

WHERE cs.type\_of\_customer IN ("long\_term\_customer", "new\_customer")

GROUP BY cs.type\_of\_customer;

ANSWER-



VISUALIZATION -

APPROACH-

**1. Define Customer Groups:**

* **Long-term customers**: If Purchasing Lifespan is equal to the overall LifeSpan
* **New customers**: If Purchasing Lifespan is not equal to the overall LifeSpan
* Use Year(Max(Invoice\_Date)) – Year(Min(Invoice\_Date)) as a proxy for "customer since Purchasing LifeSpan

**2. Measure Behavior Metrics:**

* **Purchase Frequency** → Count of invoices.
* **Basket Size** → Avg. number of tracks per invoice.
* **Spending Amount** → Total and average invoice amount.

**3. Compare Metrics Between Groups**

INSIGHTS

**1. 🛒 Frequency & Spending Show Deep Engagement**

* **Long-term customers purchase nearly 7x more frequently.**
* **They also contribute 7.4x more revenue, reinforcing their higher long-term value.**

**2. 💰 Larger Basket Sizes Among Loyal Shoppers**

* **Long-term customers spend ~24% more per transaction, indicating deeper brand engagement and trust.**

**3. 📈 Higher Average Revenue per Customer**

* **Suggests stronger customer lifecycle value and a willingness to continue investing over time.**

RECOMMENDATION –

1. **Nurture New Customers into Long-Term Champions**

* Introduce **tiered onboarding journeys** with rewards for repeat purchases in the first 90 days.
* Use behavioral triggers to recommend high-value upsell/cross-sell items early in their journey.

2. **Maximize Value of Long-Term Customers**

* Offer **exclusive VIP perks** (early access, loyalty points boosts, personal recommendations).
* Use lifetime value scores to personalize discounts and re-engagement campaigns.

3. **Incentivize Larger Baskets**

* Bundle complementary products or offer “Spend More, Save More” promotions to increase average order value for new customers.

4. **Track Behavior for Retention Signals**

* Use churn prediction models to flag declining engagement early.
* Deploy targeted campaigns like “We Miss You” or anniversary thank-yous with special offers.

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?

REFERENCE -“Objective question solution” named sql file , from 384 to 433

rows

QUERY 1

SELECT

(select name from track where track\_id = il1.track\_id)AS Track1,

(select name from track where track\_id = il2.track\_id) AS Track2,

COUNT(\*) AS Frequency

FROM Invoice\_Line il1

JOIN Invoice\_Line il2

ON il1.Invoice\_Id = il2.Invoice\_Id

AND il1.Track\_Id < il2.Track\_Id

GROUP BY il1.Track\_Id, il2.Track\_Id

ORDER BY Frequency DESC

LIMIT 10;

QUERY 2

SELECT

(select title from album where album\_id = t1.album\_id) AS Album1,

(select title from album where album\_id = t2.album\_id) AS Album2,

COUNT(\*) AS Frequency

FROM Invoice\_Line il1

JOIN Invoice\_Line il2

ON il1.Invoice\_Id = il2.Invoice\_Id

JOIN Track t1 ON il1.Track\_Id = t1.Track\_Id

JOIN Track t2 ON il2.Track\_Id = t2.Track\_Id

WHERE t1.Album\_Id < t2.Album\_Id

GROUP BY t1.Album\_Id, t2.Album\_Id

ORDER BY Frequency DESC

LIMIT 10;

QUERY 3

SELECT

g1.Name AS Genre1,

g2.Name AS Genre2,

COUNT(\*) AS Frequency

FROM Invoice\_Line il1

JOIN Invoice\_Line il2

ON il1.Invoice\_Id = il2.Invoice\_Id

JOIN Track t1 ON il1.Track\_Id = t1.Track\_Id

JOIN Track t2 ON il2.Track\_Id = t2.Track\_Id

JOIN Genre g1 ON t1.Genre\_Id = g1.Genre\_Id

JOIN Genre g2 ON t2.Genre\_Id = g2.Genre\_Id

WHERE g1.Genre\_Id < g2.Genre\_Id

GROUP BY g1.Name, g2.Name

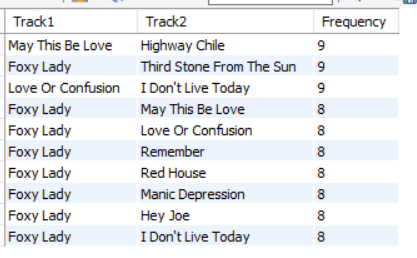
ORDER BY Frequency DESC

LIMIT 10;

ANSWER

A screenshot of a computer

AI-generated content may be incorrect.

 A screenshot of a music list

AI-generated content may be incorrect.

APPROACH-

1. Using self join on invoice line table on same invoice\_id and track id less than the other so that we can find frequency of such pair of track id and limit the output to 10 to most frequent pair of tracks
2. Join the table invoice line with track on track id and self join with invoice on equal invoice\_id and album\_id less than other so that different pair we get and then sort the output in terms of frequency in decreasing order and limit the result to see most frequent pair
3. Similiarly for genre join the invoice table with track , genre and get the name of genre and count the frequency of each pair

INSIGHTS-

* + 1. **Rock** frequently appears with multiple other genres, especially **Metal (1622 times)** and **Alternative & Punk (1056 times)**.This indicates that Rock listeners often explore **adjacent or related styles**, making them ideal targets for **cross-genre recommendations**.
    2. *Mezmerize* appears in **6 of the top 10 album pairs**, making it a **central affinity anchor**.It is frequently purchased alongside both **classic rock** (e.g., *Vault: Def Leppard*, *Dark Side of the Moon*) and **modern/alternative rock** (e.g., *Facelift*, *Get Born*), indicating **broad appeal** across rock subgenres.
    3. *Foxy Lady* appears in **8 of the top 10 co-purchased track pairs**, showing it is consistently bought with a variety of other tracks by the same artist (Jimi Hendrix). This indicates it's a **gateway or anchor track**, ideal for **track-based recommendations** or as a **featured song** in promotions and playlists.

RECOMMENDATION –

* + - 1. **✅ Cross-Selling Opportunities:**
         1. If a customer buys a **Rock** track, recommend **Alternative** or **Pop** tracks.
         2. Combine high-affinity genres in **playlist bundles** or **discounted combo packs**.
      2. **✅ Artist Collaborations:**

a) Highlight collaborations or tour packages for artists often bought together.

* + - 1. **✅ UI/UX Product Recommendations:**

"You might also like" sections can be driven using this co-purchase data.

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?

REFERENCE-“Objective question solution” named sql file , from 435 to 474

Rows

QUERY 1-

SELECT

c.Country,

COUNT(i.Invoice\_Id) AS TotalPurchases,

ROUND(AVG(i.Total), 2) AS AvgSpendingPerPurchase,

ROUND(SUM(i.Total), 2) AS TotalRevenue

FROM Customer c

JOIN Invoice i ON c.Customer\_Id = i.Customer\_Id

GROUP BY c.Country

ORDER BY TotalRevenue DESC;

QUERY 2 – NEXT PAGE

WITH LastPurchase AS (

SELECT

Customer\_Id,

MAX(Invoice\_Date) AS LastInvoiceDate

FROM Invoice

GROUP BY Customer\_Id

),

ChurnStatus AS (

SELECT

c.Customer\_Id,

c.Country,

CASE

WHEN lp.LastInvoiceDate < DATE\_sub((select max(invoice\_date) from invoice), interval 6 month) THEN 1

ELSE 0

END AS IsChurned

FROM Customer c

LEFT JOIN LastPurchase lp ON c.Customer\_Id = lp.Customer\_Id

)

SELECT

Country,

COUNT(Customer\_Id) AS TotalCustomers,

SUM(IsChurned) AS ChurnedCustomers,

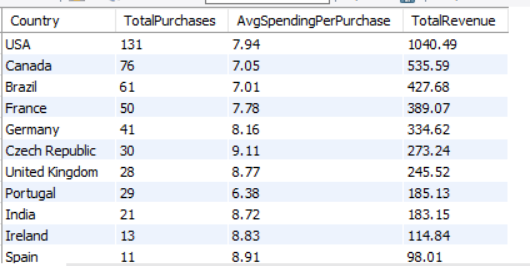
ROUND(100.0 \* SUM(IsChurned) / COUNT(Customer\_Id), 2) AS ChurnRate\_Percent

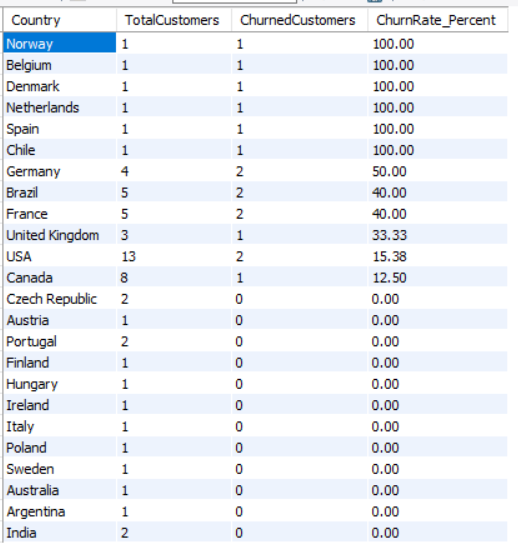
FROM ChurnStatus

GROUP BY Country

ORDER BY ChurnRate\_Percent DESC;

ANSWER-





VISUALISATION-

APPROACH-

1. For analyzing purchasing behavior of customer in different region we join the table customer with invoice and group the data with country and then find out the count of total purchase , avg spending per purchase and total purchase
2. For estimating churn rates in each region first we find out the last purchase of each customer
3. In 2nd step we defined churn status of each country by joining the table created in cte in 1st step with invoice table . if customer’s last purchase was less than 6 month of maximum date in invoice then that customer is churned else not churned
4. In last step we find the percentage of churned status of each country ,also find the customer count of each country to see the reliability of churn status ,whether it shows some good result or bad

INSIGHTS –

1. Regions with **higher revenue and purchase frequency** indicate strong customer engagement, while lower-performing regions may need **better marketing strategies**.
2. High‐churn “outliers” are almost entirely small-sample countries -Norway, Belgium, Denmark, Netherlands and Spain all show 100% churn—but each only has 1–2 customers in our dataset, so every inactive user blows the rate to 100%.
3. Mid-sized markets show “true” churn signals - USA (12 customers, 4 churned → 33% churn), Brazil (8 → 3 → 38%), Germany (7 → 2 → 29%), UK (6 → 2 → 33%), France (5 → 1 → 20%). • These markets have enough customers that the rates reflect real retention behavior.
4. Emerging‐market stickiness in low-volume regions - India, Argentina, Australia, Poland, Italy, Sweden, etc., show very low churn (0–17%) but also just 1–4 customers each.

Economic & demographic correlates –

* + - * 1. Wealthy, digitally saturated markets (Nordics, Benelux, Spain) have high customer choice and high expectations—so if product-market fit isn’t perfect, churn soars.
        2. Big, heterogeneous markets (USA, Brazil) balance higher competition with broader audiences—churn settles in the 30–40% range.
        3. Smaller or emerging markets may have fewer alternatives and lower price sensitivity, driving stickier—but as your footprint grows, churn can rise too.

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?

REFERENCE - “Objective question solution” named sql file , from 476 to 562

Rows

QUERY NEXT PAGE

WITH rfm\_base AS (

SELECT

c.Customer\_Id,

c.Country,

MAX(i.Invoice\_Date) AS LastInvoiceDate,

COUNT(i.Invoice\_Id) AS Frequency,

SUM(i.Total) AS Monetary

FROM Customer c

LEFT JOIN Invoice i ON c.Customer\_Id = i.Customer\_Id

GROUP BY c.Customer\_Id, c.Country

),

rfm\_features AS (

SELECT

(select concat(firsT\_name , " " ,last\_name )from customer where customer\_id = r.customer\_id) as customer\_name,

Country,

DATEDIFF((SELECT MAX(INVOICE\_DATE) FROM INVOICE), LastInvoiceDate) AS Recency,

Frequency,

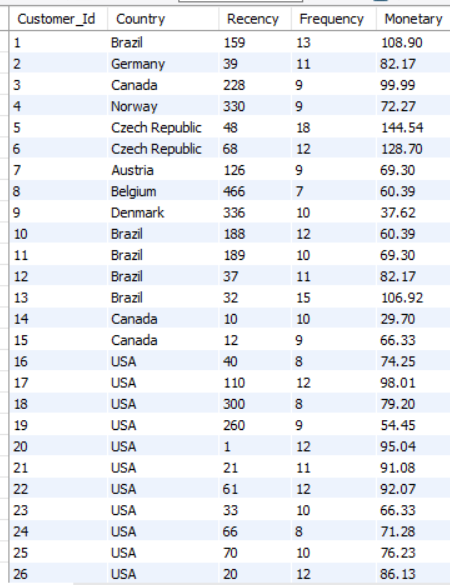
Monetary

FROM rfm\_base r

)

SELECT \* FROM rfm\_features;

ANSWER-



VISUALIZATION-

INSIGHTS

* 1. Most customers show **Recency values below 20**, meaning they've purchased recently. Those in the 15–20 day range should be monitored—they’re first in line for re-engagement campaigns.
  2. Several customers (e.g., **Stanislaw Wójcik, Tim Goyer, Jack Smith**) spike above 12–14 purchase shows loyal customer.
  3.  Top spenders (**Marc Dubois, Jack Smith**) push close to the $500 mark. Lower-tier spenders (**François Tremblay, Flynn Freeman**) land closer to ~$100.

PART 2-WHICH CUSTOMER SEGMENT POSSES RISK-

QUERY -NEXT PAGE

WITH rfm\_base AS (

SELECT

c.Customer\_Id,

c.Country,

MAX(i.Invoice\_Date) AS LastInvoiceDate,

COUNT(i.Invoice\_Id) AS Frequency,

SUM(i.Total) AS Monetary

FROM Customer c

LEFT JOIN Invoice i ON c.Customer\_Id = i.Customer\_Id

GROUP BY c.Customer\_Id, c.Country

),

rfm\_features AS (

SELECT

Customer\_Id,

Country,

DATEDIFF((SELECT MAX(INVOICE\_DATE) FROM INVOICE), LastInvoiceDate) AS Recency,

Frequency,

Monetary

FROM rfm\_base

),

risk\_tags AS (

SELECT \*,

CASE

WHEN Recency > 180 AND (Frequency <= 2 OR Monetary < 20) THEN 'High Risk'

WHEN Recency > 90 OR Frequency <= 3 OR Monetary < 40 THEN 'Medium Risk'

ELSE 'Low Risk'

END AS RiskSegment

FROM rfm\_features

)

SELECT \* FROM risk\_tags;

ANSWER

A screenshot of a data

AI-generated content may be incorrect.

APPROACH-

1. Joins Customer with Invoice to calculate:
   1. Most recent purchase date (LastInvoiceDate) of each customer
   2. Total number of purchases (Frequency) using group by and count function
   3. Total spend (Monetary) using sum function
2. Calculates **Recency** as the time gap between the latest invoice in the database and each customer’s last purchase using datediff function.
3. Assigns customers to **High**, **Medium**, or **Low Risk** segments based on:

* High Recency( > 180) and low Frequency(<=2)/Monetary (<20)→ *High Risk*
  + - Moderate Recency(>90) or low Frequency(<=3)/Monetary(40) → *Medium Risk*
    - Otherwise → *Low Risk*

PART 3 – CUSTOMER POSSES RISK , LOCATIION WISE-

QUERY

SELECT

Country,

RiskSegment,

COUNT(\*) AS NumCustomers,

ROUND(AVG(Recency), 0) AS AvgRecency,

ROUND(AVG(Frequency), 1) AS AvgFrequency,

ROUND(AVG(Monetary), 2) AS AvgMonetary

FROM (

SELECT

c.Customer\_Id,

c.Country,

DATEDIFF((select max(invoice\_date) from invoice), MAX(i.Invoice\_Date)) AS Recency,

COUNT(i.Invoice\_Id) AS Frequency,

SUM(i.Total) AS Monetary,

CASE

WHEN DATEDIFF((select max(invoice\_date) from invoice), MAX(i.Invoice\_Date)) > 180 AND

(COUNT(i.Invoice\_Id) <= 2 OR SUM(i.Total) < 20) THEN 'High Risk'

WHEN DATEDIFF((select max(invoice\_date) from invoice), MAX(i.Invoice\_Date)) > 90 OR

COUNT(i.Invoice\_Id) <= 3 OR SUM(i.Total) < 40 THEN 'Medium Risk'

ELSE 'Low Risk'

END AS RiskSegment

FROM Customer c

LEFT JOIN Invoice i ON c.Customer\_Id = i.Customer\_Id

GROUP BY c.Customer\_Id, c.Country

) AS risk\_data

GROUP BY Country, RiskSegment

ORDER BY Country, FIELD(RiskSegment, 'High Risk', 'Medium Risk', 'Low Risk');

ANSWER



VISUALIZATION-

APPROACH –

1. Joins Customer with Invoice to calculate for each customer:
   * 1. **Recency:** Time since most recent purchase.
     2. **Frequency:** Total invoices.
     3. **Monetary:** Total spend.
2. Uses a CASE statement to assign **RiskSegment**:
   * 1. *High Risk:* Very inactive and low value.
     2. *Medium Risk:* Somewhat inactive or low engagement.
     3. *Low Risk:* Recently active and consistent buyers.
3. For each **Country + RiskSegment** combination, it summarizes:
   * 1. Number of customers.
     2. Average Recency, Frequency, and Monetary value.
4. Orders results alphabetically by country, and within each country by *risk tier priority* (High → Medium → Low).

INSIGHTS-

1. **High-Risk Zones:**
   1. Countries like **Ireland, India, and Portugal** exhibit *higher RISK SEGMENT scores* despite moderate customer bases. This implies a need to reassess customer engagement efforts or reassess market fit in those regions.
2. **High Engagement with Low Risk:**
   1. **UK, Netherlands, and Germany** show a healthier balance: moderate to high **Frequency**, **Monetory**, and a relatively lower **Risk** profile. These could serve as reference markets for benchmarking best practices.
3. **Low ROI Segments:**
   1. Regions like **Brazil, Czech Republic, and Hungary** have *low customer counts*, *high recency*, and *modest monetary values*. These are candidates for re-evaluation—possibly lower strategic priority or in need of hyper-targeted outreach.

RECOMMENDATION-

**🌍 1. Strengthen Retention in Medium-Risk Markets**

Medium-risk segments dominate in countries like **Germany, France, Spain, Brazil, India, and Argentina**, often accompanied by **high recency (indicating longer time since last purchase)** and **strong average frequency and monetary values**.

**Recommendations:**

* **Trigger Re-Engagement Campaigns:** Use time-based nudges like “We miss you” offers or anniversary perks to encourage repeat purchases.
* **Introduce Loyalty Reinforcement Programs:** Provide incentives such as cashback or rewards for three or more purchases in a defined period.
* **Deploy Behavioral Email Journeys:** Tailor content based on recent activity gaps and spending behavior.

**🛡 2. Expand High-Value, Low-Risk Customer Base**

Countries with consistent low-risk segments and healthy metrics include **Australia, Ireland, Finland, and Czech Republic**, showing **low recency**, **high frequency**, and **strong monetary value**.

**Recommendations:**

* **Promote VIP Upsells:** Reward loyalty with early-access or premium tier products.
* **Referral Programs:** Encourage loyal customers to bring in new users with double-sided rewards (e.g., “Give $10, Get $10”).
* **Gather Voice-of-Customer Feedback:** Use these customers to refine product-market fit and improve offerings for broader markets.

**🔁 3. Optimize Blended Markets for Tiered Offers**

Some regions like **Canada, Germany, and France** have both low-risk and medium-risk segments—ideal for A/B testing and tiered personalization.

**Recommendations:**

* **Create Tiered Campaigns Based on RFM Scores:** Push seasonal offers to medium-risk, while offering exclusivity and early access to low-risk.
* **Run Churn Prediction Models:** Proactively identify shifting behaviors from low to medium risk and intervene early.

**🔎 4. Prioritize India & Spain for Retention ROI**

* **India’s Medium-Risk Segment** has **high frequency and excellent monetary values**, making it ideal for conversion to low risk.
* **Spain** shows strong engagement with **AvgMonetary > $98** despite high recency—targeted communication here could yield a quick win.

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?

REFERENCE- “Objective question solution” named sql file , from 565 to 586

Rows

QUERY

with estimated\_clv as (

SELECT

i.billing\_country as country,

(select concat(first\_name , " " , last\_name) from customer where customer\_id = i.customer\_id) as customer\_name,

sum(i.total) as Total\_spending,

ROUND(AVG(i.Total), 2) AS AvgPurchaseValue,

COUNT(i.Invoice\_Id) AS NumPurchases,

DATEDIFF(MAX(i.Invoice\_Date), MIN(i.Invoice\_Date)) AS LifespanDays,

DATEDIFF((select max(invoice\_date) from invoice),MAX(i.Invoice\_Date)) as last\_purchase\_interval,

ROUND(AVG(i.Total) \* COUNT(i.Invoice\_Id)\*DATEDIFF(MAX(i.Invoice\_Date), MIN(i.Invoice\_Date))/365, 2) AS ApproxCLV

FROM Invoice i

GROUP BY i.Customer\_Id ,country

)

SELECT \*,

CASE

WHEN ApproxCLV > 350 THEN 'High Value'

WHEN ApproxCLV BETWEEN 250 AND 350 THEN 'Medium Value'

ELSE 'Low Value'

END AS CLVSegment

FROM estimated\_clv

order by ApproxCLV desc ;

ANSWER-



VISUALIZATION-

APPROACH-

1. In cte , we calculate total spending ,average spending ,count of invoice id ,total lifespan of each customer using invoice table and group by function on customer id and country
2. Now calculating Customer Life Time Value (CLV) by multiplying average total spending and purchase frequency and purchase span of each customer
3. Now we define parameter CLV statud using case when clause in which , if CLV >350 then high value ,if it is between 25 to 350 then medium value else less value

INSIGHTS-

**🔶 1. Inverse Relationship Between CLV and Last Purchase Interval**

* **High CLV customers** generally have **low last\_purchase\_interval** (i.e., they purchased recently).
* As **CLV decreases**, the **last\_purchase\_interval increases**, meaning:
  + Customers with **low CLV tend to be inactive**.
  + This indicates they likely **stopped purchasing (churned)**.

**🔶 2. Spikes in Last Purchase Interval = Churn Risk**

Customers like:

* **Lucas Mancini**
* **Dan Nisters**
* **Eduardo Gilberto**
* **Bjorn Hansen**

...have **very high last\_purchase\_interval** (300–450+ days) but **relatively low CLV**, signaling they have **likely churned** or are **not engaged**.

**🔶 3. Low CLV + Long Gaps = Weak Relationship**

These users:

* Made **few purchases**
* Purchased **long ago**
* Likely had **low engagement or satisfaction**
* Have not returned for a long time

**🔶 4. High CLV + Low Interval = Engaged Customers**

Top-left customers like:

* **Frantisek Wichterlova**
* **Helena Holy**
* **Philip Hudeček**

...show:

* **High CLV**
  + - * **Low purchase interval**  
        → Meaning they’re **frequent buyers** and highly valuable.

RECOMMENDATION –

1. **Model Inputs:** Use lifespan, total spending, avg purchase value, frequency. Longer tenure + high spend → high CLV.

2. **High-Value Segments:** Prioritize Czech & German customers with long lifespan, frequent buys, high CLV. Reward loyalty, offer VIP perks.

3. **Lapsed Buyer Patterns:** Short lifespan, low purchases, moderate CLV (e.g. some UK & Brazil customers). Indicators of churn.

4. **Reactivation Strategy:** Target lapsed/medium CLV with win-back offers and personalized nudges.

5 **Tiered Loyalty:** Map segments (High, Medium, Low CLV). Assign tailored rewards and lifecycle messaging per tier.

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?

ANSWER-

If promotional campaign data is available in or linked to the Chinook database, we can assess its effectiveness by analyzing customer behavior before, during, and after the campaign periods. This involves measuring changes in new customer sign-ups, purchase frequency, repeat purchases, and overall revenue.

APPROACH-

**1. Track Customer Acquisition from Campaigns**

* Identify customers whose **first purchase** occurred during a campaign.
* Use Invoice\_Date and campaign StartDate/EndDate to detect acquisition timing.
* Helps determine how many new customers were acquired due to the campaign.

**2. Measure Retention After Campaign Ends**

* Analyze whether customers continued purchasing after the promotional period.
* Track purchases **post-campaign** to measure **long-term retention**.
* Calculate the percentage of customers who returned, indicating campaign loyalty impact.

**3. Compare Sales During vs. Outside Campaign**

* Measure total revenue generated **during** the campaign vs. **non-campaign periods**.
* Helps in identifying if sales spiked only temporarily or had lasting impact.
* Use to calculate **ROI** (Return on Investment) of the campaign.

**4. Evaluate Campaign Type Performance**

* If different campaign types exist (e.g., email, discount, event), compare:
  + Number of customers acquired per type
  + Total revenue per campaign type
* This helps determine which campaign strategy is most effective.

**5. Analyze Repeat Purchase Trends Over Time**

* Break down post-campaign purchases by **month** or **week**.
* Identify if purchases drop sharply after campaign ends or remain steady.
* Useful to understand **customer engagement lifecycle** after promotion.

**6. Use Key Metrics to Evaluate Campaign Success**

Some key performance indicators (KPIs) include:

* **Customer Acquisition Rate**
* **Repeat Purchase Rate (Retention)**
* **Total Revenue During Campaign**
* **Revenue Per Campaign Type**
* **Campaign ROI** = (CampaignRevenue - CampaignCost) / CampaignCost

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

ANSWER-

1. **Identify Potential Business Questions**

Based on the data:

* 📈 Which customers generate the most revenue?
* 🌎 Which countries/regions are top-performing?
* 🧲 What’s the average customer lifecycle?
* 🛍️ Which tracks or genres are top sellers?
* 📊 What are seasonal sales trends?

**2. Define Analytical Goals**

You can propose and explore projects like:

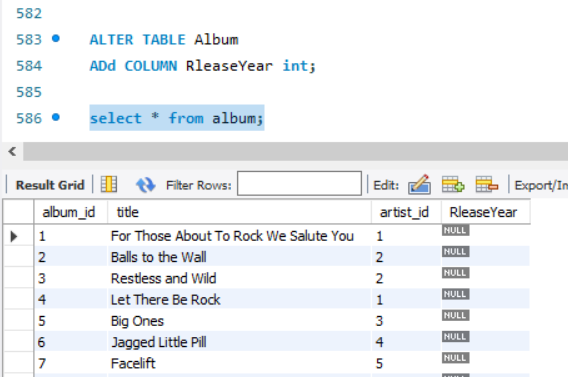
* **Customer Segmentation** (based on purchase behavior)
* **Sales Trend Analysis** (monthly/yearly)
* **Product Recommendation System** (based on purchase history)
* **Customer Lifetime Value (CLV) Estimation**
* **Top Artist/Genre Performance Analysis**

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?

REFERENCE- “Objective question solution” named sql file , from 588 to 593

rows

ANSWER-



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 a SQL query to provide this information.

REFERENCE- “Objective question solution” named sql file , from 595to 605

Rows

QUERY

SELECT

I1.BILLING\_COUNTRY AS COUNTRY,

ROUND(AVG(I1.TOTAL) , 2) AS AVG\_TOTAL\_SPENT,

COUNT(DISTINCT I1.CUSTOMER\_ID) AS CUSTOMER\_COUNT,

ROUND(COUNT(DISTINCT I2.INVOICE\_LINE\_ID) / COUNT(DISTINCT I1.CUSTOMER\_ID) , 2) AS AVG\_TRACK\_PER\_CUSTOMER

FROM INVOICE I1

JOIN INVOICE\_LINE I2

ON I1.INVOICE\_ID = I2.INVOICE\_ID

GROUP BY COUNTRY;

ANSWER-

A screenshot of a data

AI-generated content may be incorrect.

VISUALIZATION-

APPROACH

**1. Identify Relevant Tables**

* **Customer**: Has Country and Customer\_Id
* **Invoice**: Contains Total (amount spent) and links to Customer
* **Invoice\_Line**: Contains individual track details per invoice (number of tracks purchased)

**2. Calculate Number of Customers per Country**

* Group customers by country and count unique Customer\_ID.

**3. Calculate Total and Average Amount Spent**

* Join Customer → Invoice to get total purchases per customer.
* Then group by Country to calculate:
  + Total spent per country
  + Average spent per customer in each country

**4. Calculate Total and Average Tracks Purchased**

* Join Invoice\_Line to count total tracks per customer.
* Calculate the **average number of tracks** bought per customer by country.

**5. Final Output Fields Should Include:**

For each country:

* 🌍 Country name
* 👥 Number of customers
* 💰 Average total amount spent per customer
* 🎵 Average number of tracks purchased per customer

INSIGHTS-

**1. USA Has the Highest Customer Count and Track Volume**

* The USA stands out with the **highest number of customers** and the **highest average tracks purchased per customer (~130)**.
* Indicates a **large and highly engaged customer base**, even if spending per customer is not the highest.

**2. Czech Republic and Ireland Lead in Average Spending**

* Both countries show **highest average total spending per customer** (green bars), despite **low customer counts**.
* Suggests these markets have **fewer but high-value customers**.

**3. Denmark and Argentina Have Low Track Purchases**

* Denmark shows **low average tracks per customer**, even though its average spending is decent.
* Argentina shows **low track purchase and low spending**, making it a **low-engagement, low-value market**.

**4. Countries Like Germany, UK, Canada Show Balanced Performance**

* These regions maintain a **healthy balance** of:
  + Moderate to high customer count
  + Solid average spending
  + Good track engagement
* They represent **stable and reliable markets**.

**5. Austria and Belgium Have Low Value per Customer**

* These countries have **low average spending** and **lower average track purchases**, suggesting limited contribution to revenue.
  + - * Good candidates for targeted campaigns or re-engagement efforts.

RECOMMENDATION –

1. 🎯 **Focus High-ROI Markets:** Austria, Belgium, Czech Rep. – high spend & track count. Prioritize campaigns.

2. 📉 **Revive Low-Spend Regions:** UK, Brazil – run trials, discounts, curated playlists.

3. 📊 **Segment by Value:** Group countries by spend & engagement to personalize retention efforts.