Objective Questions

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

NO duplicates value but NULL values are present   
 **customer Table:**

* fax: 47 null values
* state: 29 null values
* company: 49 null values

**employee Table:**  reports\_to: 1 null value for employee\_id = 1

**track Table:** composer: 978 null values

**QUERY** Select \* from album;

Select \* from artist;

Select Count(\*) from customer

where company is NULL ;

Select Count(\*) from customer

where state is NULL;

Select Count(\*) from customer

where fax is NULL ; -- 49 company, 29 state, 47 fax values are null in the customer table

Select \* from employee ;-- 1 reports\_to is null in this table

Select \* from genre;

Select \* from invoice;

Select \* from invoice\_line;

Select \* from media\_type;

Select \* from playlist;

Select \* from playlist\_track;

Select count(\*) from track

where composer is NULL;

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

**QUERY**WITH TopTracks AS (

SELECT

t.track\_id,

t.name AS track\_name,

SUM(il.quantity) AS total\_sales

FROM invoice\_line il

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

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

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

WHERE c.country = 'USA'

GROUP BY t.track\_id

),

-- Find top artists in the USA

TopArtists AS (

SELECT

a.artist\_id,

a.name AS artist\_name,

SUM(il.quantity) AS total\_sales

FROM invoice\_line il

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

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

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

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

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

WHERE c.country = 'USA'

GROUP BY a.artist\_id

)

-- Combine top-selling tracks and artists

SELECT

t.track\_name,

a.artist\_name,

g.name AS genre\_name,

SUM(il.quantity) AS total\_sales

FROM TopTracks t

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

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

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

JOIN TopArtists a ON tr.album\_id = a.artist\_id

GROUP BY t.track\_name, a.artist\_name, g.name

ORDER BY total\_sales DESC

LIMIT 10;

**RESULT  
**

Therefore, top genre is **Rock**.

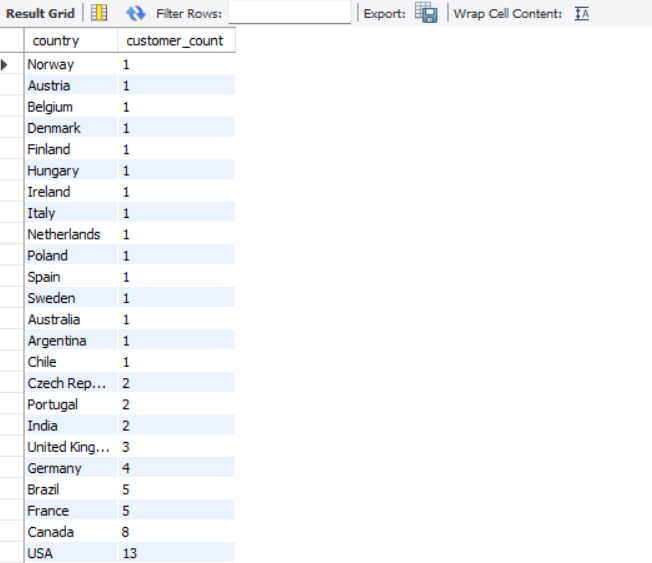
**3.What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?**

**Query**

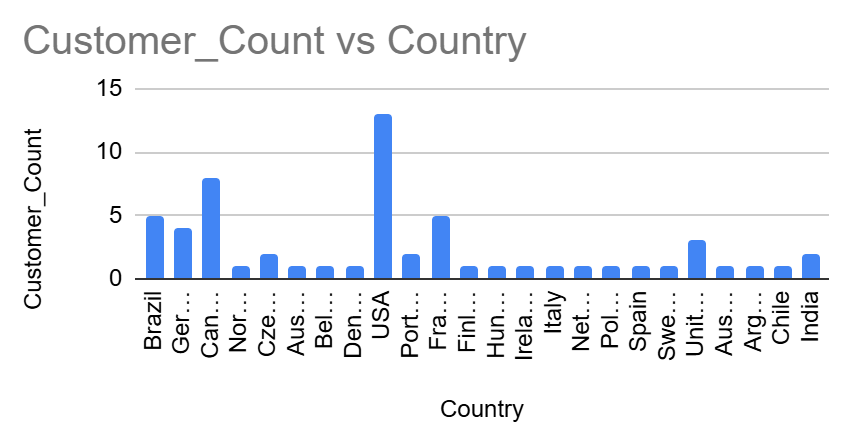
select country,count(\*) as customer\_count

from customer

group by country;

**Result  
 **

* Based on location **USA** is having maximum number of customers.
* The customer table doesn’t have age and gender columns to understand the customer breakdown.



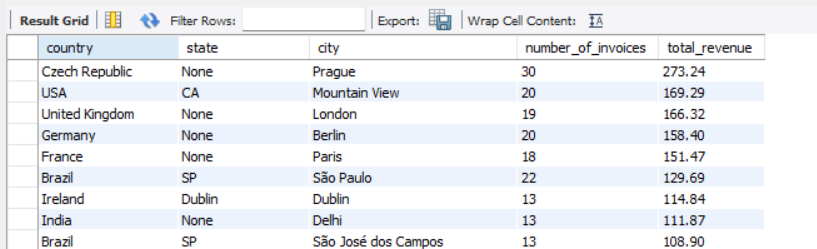
**4.Calculate the total revenue and number of invoices for each country, state, and city**:

**Query**Select billing\_country as country, billing\_state as state, billing\_city as city, count(invoice\_id) as number\_of\_invoices, sum(total) as total\_revenue

from invoice

group by billing\_country,billing\_state,billing\_city

order by total\_revenue desc;

**Result  
**

* Prague (Czech Republic) leads in revenue generation.
* São Paulo (Brazil) records a high invoice count but comparatively lower revenue**.**

A graph with red and blue lines

AI-generated content may be incorrect.

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

**Query** WITH CustomerRevenue AS (

SELECT

c.customer\_id,

c.country,

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

FROM customer c

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

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

GROUP BY c.customer\_id, c.country

)

SELECT

customer\_id,

country,

total\_revenue

FROM (

SELECT

cr.customer\_id,

cr.country,

cr.total\_revenue,

ROW\_NUMBER() OVER (PARTITION BY cr.country ORDER BY cr.total\_revenue DESC) AS row\_num

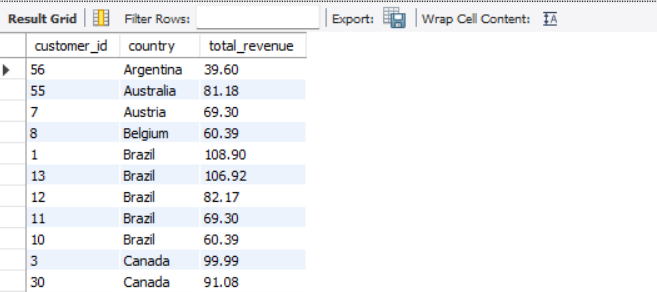
FROM CustomerRevenue cr

) AS RankedCustomers

WHERE row\_num <= 5

ORDER BY country, total\_revenue DESC;

**Result**

****

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

**Query**Select concat(c.first\_name, " " , c.last\_name) as customer\_name,

t.name as track\_name, SUM(il.quantity) as total\_sales

from customer c

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

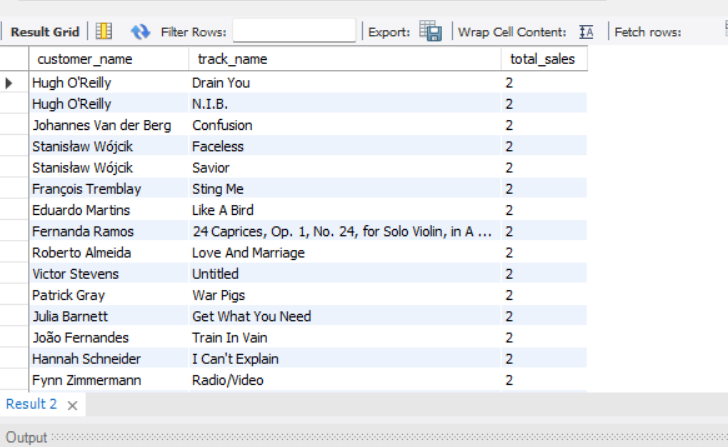
left join invoice\_line il on il.invoice\_id = i.invoice\_id

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

group by customer\_name, track\_name

order by total\_sales desc;

**Result**

****

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

c.customer\_id,

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

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

i.billing\_country,

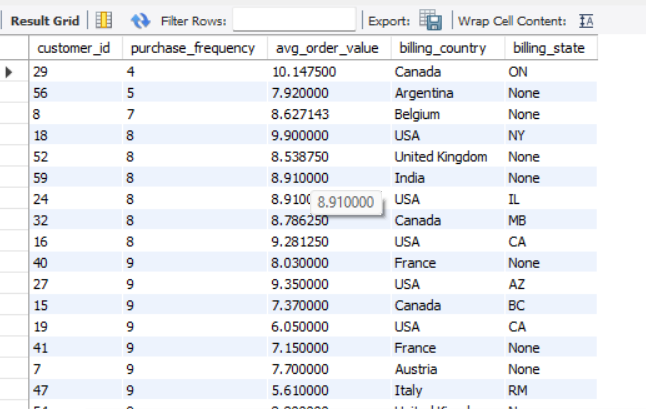
i.billing\_state

FROM customer c

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

GROUP BY c.customer\_id, i.billing\_country, i.billing\_state

ORDER BY purchase\_frequency ASC

LIMIT 1000;  
  
**Result  
**The average order a customer takes is around 8-9 with the least one is 4

A graph of blue and yellow lines

AI-generated content may be incorrect.

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

**Query**select count(distinct customer\_id) as customer\_count from invoice

where invoice\_date between '2018-01-01' and '2018-12-31' and customer\_id not in

(select distinct customer\_id from invoice

where invoice\_date between '2017-01-01' and '2017-12-31');

-- customers churned in 2018

select count(distinct customer\_id) as customer\_count from invoice

where invoice\_date between '2019-01-01' and '2019-12-31' and customer\_id not in

(select distinct customer\_id from invoice

where invoice\_date between '2018-01-01' and '2018-12-31');

-- customers churned in 2019

select count(distinct customer\_id) as customer\_count from invoice

where invoice\_date between '2020-01-01' and '2020-12-31' and customer\_id not in

(select distinct customer\_id from invoice

where invoice\_date between '2019-01-01' and '2019-12-31');

-- customers churned in 2020

select count(distinct customer\_id) from invoice

where invoice\_date between '2017-01-01' and '2017-12-31';

-- customers at the starting of 2018

select count(distinct customer\_id) from invoice

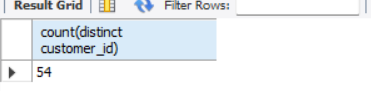
where invoice\_date between '2018-01-01' and '2018-12-31';

-- customers at the starting of 2019

select count(distinct customer\_id) from invoice

where invoice\_date between '2019-01-01' and '2019-12-31';

-- customers at the starting of 2020



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

SELECT

g.name AS genre\_name,

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

FROM invoice\_line il

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

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

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

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

WHERE c.country = 'USA'

GROUP BY g.genre\_id

),

TotalSales AS (

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

FROM invoice\_line il

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

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

WHERE c.country = 'USA'

)

SELECT

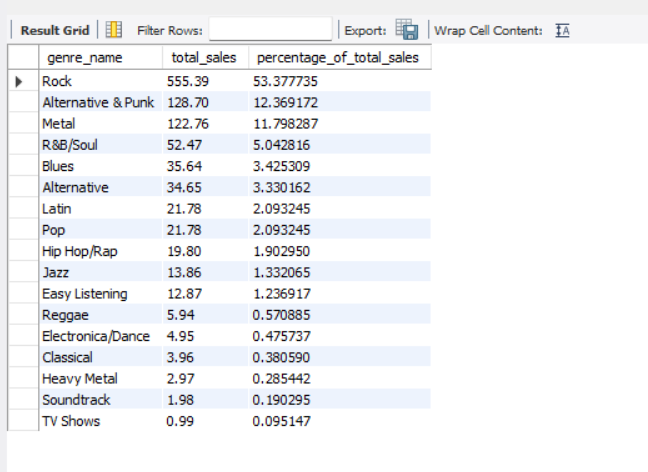
gs.genre\_name,

gs.total\_sales,

(gs.total\_sales / ts.total\_sales) \* 100 AS percentage\_of\_total\_sales

FROM GenreSales gs, TotalSales ts

ORDER BY gs.total\_sales DESC;

**Result**  


A colorful circle with text

AI-generated content may be incorrect.

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

c.customer\_id,

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

FROM customer c

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

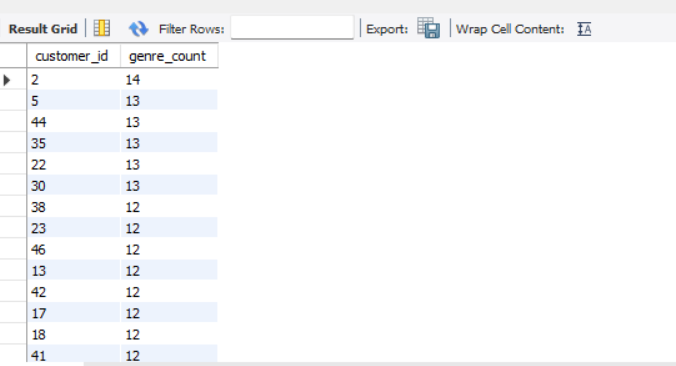
JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id

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

GROUP BY c.customer\_id

HAVING genre\_count >= 3

order by genre\_count desc;

**Result  
  
**

A graph of a number of blue bars

AI-generated content may be incorrect.

The top 5 genre count giving customer id’s are 2,5,44,35,22 with genre count of 14,13,13,13,13 respectively

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

g.name AS genre\_name,

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

FROM invoice\_line il

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

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

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

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

WHERE c.country = 'USA'

GROUP BY g.genre\_id

ORDER BY total\_sales DESC;

**Result**

A screenshot of a computer

AI-generated content may be incorrect.

* Top Genre (Rock): Rock leads by a significant margin with a genre sum of 555.39, holding the top spot.
* High Ranking Genres: Alternative & Punk and Metal follow closely, ranking second and third with sums of 128.70 and 122.76, respectively.

Lower Ranking Genres: TV Shows has the lowest genre sum at 0.99, with other less popular genres like Classical and Soundtrack also falling at the lower end of the ranking

**12.Identify customers who have not made a purchase in the last 3 months  
  
Query**select c.customer\_id, concat(c.first\_name, " " , c.last\_name) as customer\_name, max(i.invoice\_date) as last\_purchase\_date

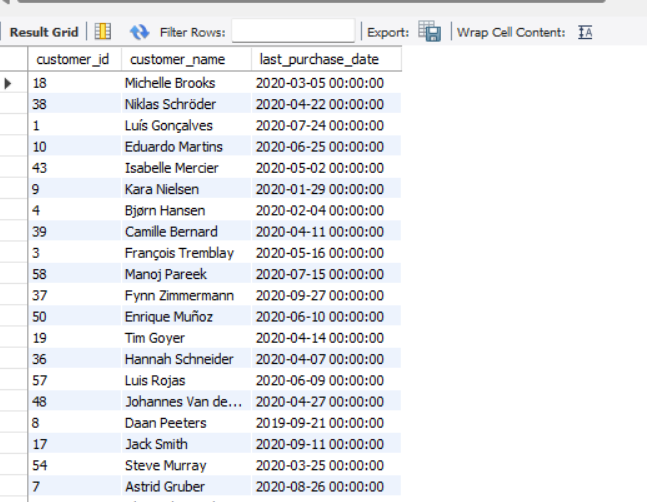
from customer c

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

group by c.customer\_id

having last\_purchase\_date < DATE\_SUB((Select max(invoice\_date) from invoice), interval 3 month);

**Result**

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

**SUBJECTIVE QUESTIONS  
  
  
  
1.Recommend the three albums from the new record label that should be prioritized for advertising and promotion in the USA based on genre sales analysis.  
  
Query**SELECT a.album\_id, a.title, g.name AS genre, SUM(il.quantity \* il.unit\_price) AS total\_sales

FROM album a

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

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

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

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

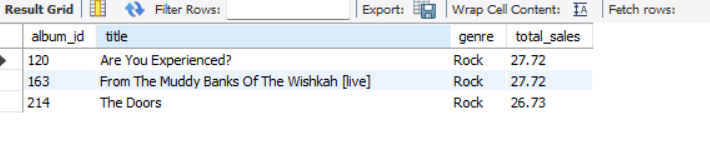
JOIN customer c ON i.customer\_id = c.customer\_id

WHERE c.country = 'USA'

GROUP BY a.album\_id, a.title, g.name

ORDER BY total\_sales DESC

LIMIT 3**;**

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

* Are You Experienced?
* From The Muddy Banks Of The Wishkah[live]
* The Doors

2.**Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.  
  
Query**SELECT c.country, g.name AS genre, SUM(il.quantity \* il.unit\_price) AS total\_sales

FROM invoice\_line il

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

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

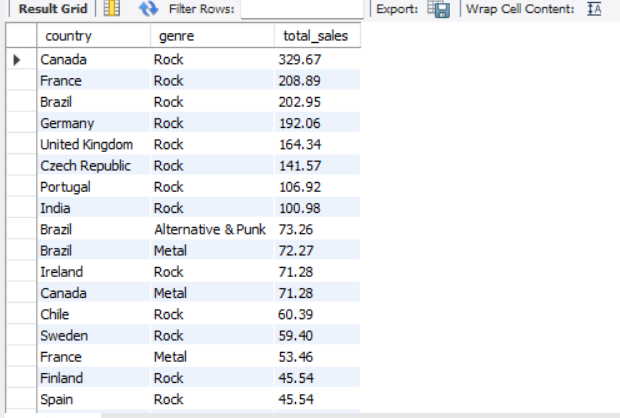
JOIN track t ON il.track\_id = t.track\_id

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

WHERE c.country != 'USA'

GROUP BY c.country, g.name

ORDER BY total\_sales DESC;

**Result  
  
**The commonality between the data regarding USA and rest of the countries is that the ROCK genre is the top genre in all countries. The 2nd and 3rd position is taken by METAL & ALTERNATIVE & PUNK genre among all the countries.  
  
**3.Customer Purchasing Behavior Analysis: How do the purchasing habits (frequency, basket size, and spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?  
  
Query**with month\_difference as(Select i.customer\_id, max(invoice\_date), min(invoice\_date),

ABS(TIMESTAMPDIFF(MONTH, max(invoice\_date), min(invoice\_date))) time\_for\_each\_customer,

sum(total) sales, sum(quantity) items, count(invoice\_date) frequency

from invoice i

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

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

group by i.customer\_id

order by time\_for\_each\_customer desc),

average\_time as(Select avg(time\_for\_each\_customer) as average from month\_difference),

customer\_category as(Select \*,

Case

when time\_for\_each\_customer > (Select average from average\_time) then "Long-term Customer"

else "New Customer"

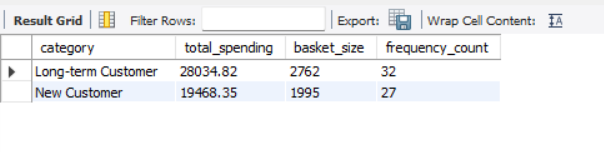
end as category

from month\_difference)

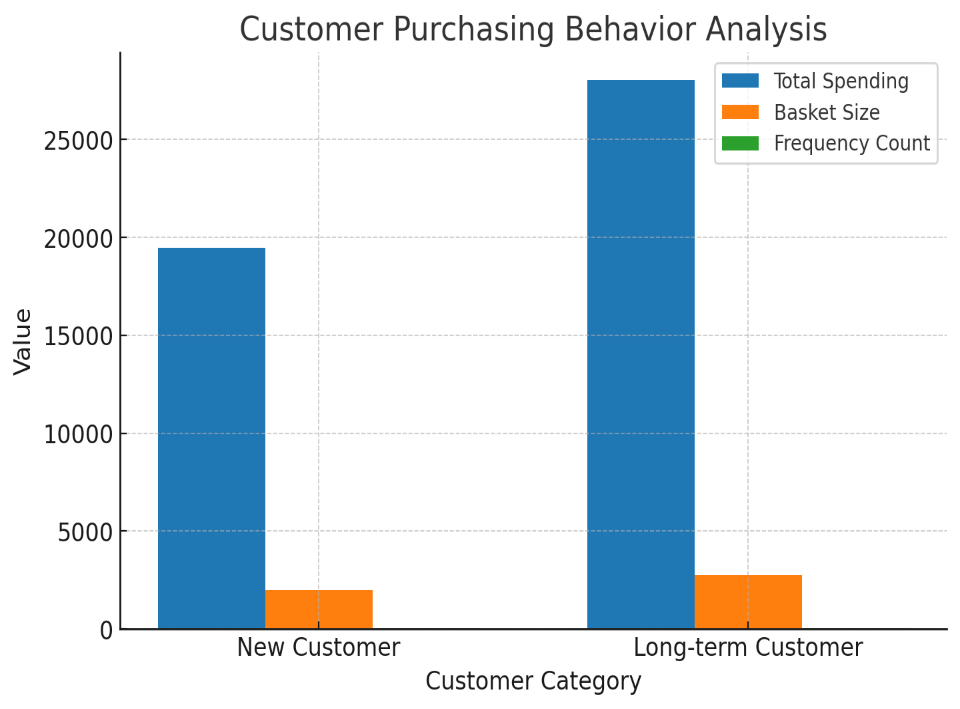
Select category, sum(sales) total\_spending, sum(items) basket\_size, count(frequency) frequency\_count

from customer\_category  
group by category**;**

**Result**

**  
Insights-** From above output it can be seen that the long-term customers are having more spending amount and basket size than new customers.

**Recommendations -**

* It shows that customer loyalty plays an important role in increasing the revenue of the company because the long-term customers tend to buy more than the new customers.
* The company should focus on the retention rate of the customers so as to increase the sales over time.
* Company can introduce discounts to new customers so that they can spend frequently.  
    
  

**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?  
  
Query**

SELECT

g.name AS genre,

a.title AS album,

COUNT(DISTINCT i.invoice\_id) AS times\_purchased\_together,

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

FROM invoice\_line il

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

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

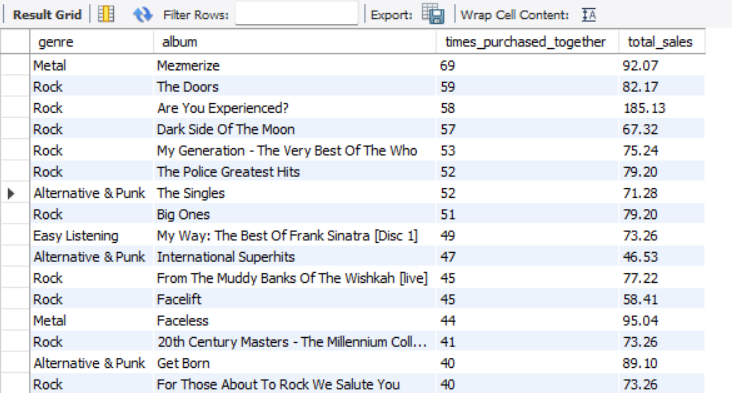
JOIN album a ON t.album\_id = a.album\_id

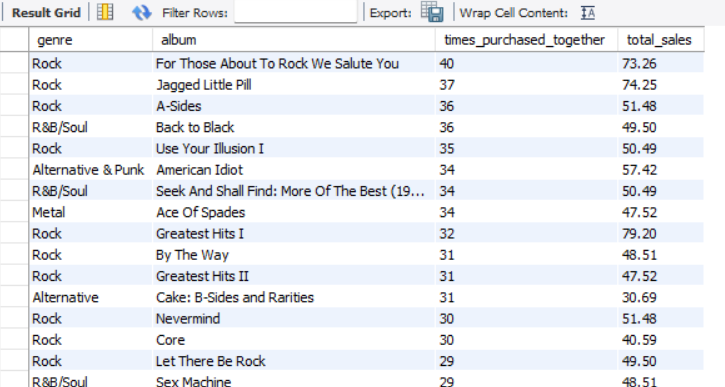
JOIN genre g ON t.genre\_id = g.genre\_id

GROUP BY g.name, a.title

ORDER BY times\_purchased\_together DESC, total\_sales DESC;

**Result**

****

****  (complete data can be seen in descending form while running the query) **1.Frequently Purchased Together:**

* + If customers who buy *Mezmerize* (Metal) also often purchase *Ace of Spades* (Metal), these albums have a strong affinity.

1. **Genre-Based Affinity:**
   * Customers who purchase *Dark Side of the Moon* (Rock) might also buy *The Doors* or *Are You Experienced?* (Rock) due to genre similarities.
2. **Artist-Based Affinity:**
   * A customer buying *The Best of The Who* could be recommended *My Generation - The Very Best of The Who* based on their affinity for the artist.

### **Recommendations:**

* Cross-Selling: Suggest albums from the same genre or artist based on what customers have already purchased.
* Personalized Suggestions: Offer complementary albums or artists to increase purchases, e.g., "If you like *Mezmerize*, check out *Ace of Spades*."

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?**

**Query**

WITH CustomerPurchases AS (

SELECT

c.customer\_id,

c.city,

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

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

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

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

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, c.city

),

ChurnAnalysis AS (

SELECT

cp.customer\_id,

cp.city,

CASE

WHEN DATEDIFF(CURDATE(), MAX(i.invoice\_date)) > 90 THEN 'Churned'

ELSE 'Active'

END AS churn\_status

FROM CustomerPurchases cp

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

GROUP BY cp.customer\_id, cp.city

)

SELECT

ca.city,

COUNT(DISTINCT ca.customer\_id) AS total\_customers,

SUM(CASE WHEN ca.churn\_status = 'Churned' THEN 1 ELSE 0 END) AS churned\_customers,

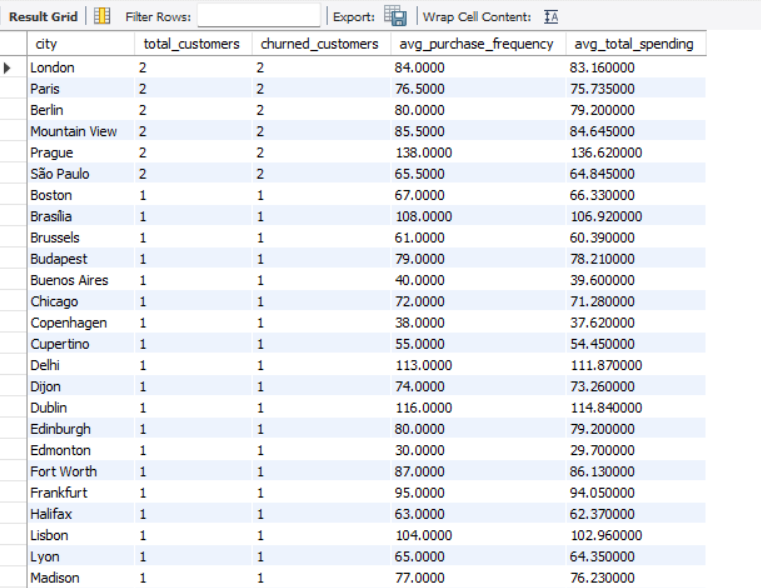
AVG(cp.purchase\_frequency) AS avg\_purchase\_frequency,

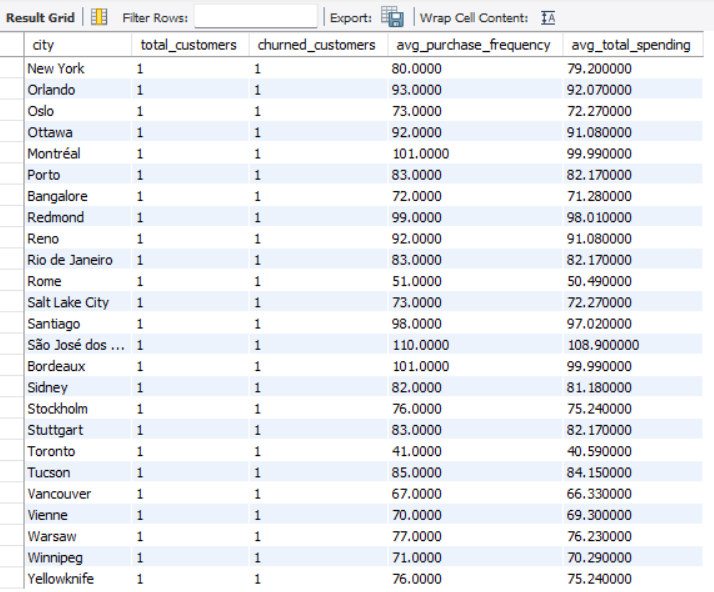
AVG(cp.total\_spending) AS avg\_total\_spending

FROM ChurnAnalysis ca

JOIN CustomerPurchases cp ON ca.customer\_id = cp.customer\_id

GROUP BY ca.city

ORDER BY churned\_customers DESC;  
  
**Result**  




1. **High-value Markets:**
   * Cities like *London*, *Mountain View*, and *Prague* have higher average spending and frequency.
   * Focus on premium products, loyalty programs, and exclusive offers for these regions.
2. **Low-value Markets:**
   * Cities like *Buenos Aires* and *São Paulo* show lower average spend.
   * Offer budget-friendly products, target these areas with specific **promotions.**
3. **Churn Insights:**
   * Higher churn rates in cities like *Santiago*, *New York*, and *Bangalore*.
   * Improve customer retention through personalized outreach, better service, and engagement strategies.
4. **Cross-selling and Targeted Marketing:**
   * Use purchasing behaviors and data from top cities to create cross-sell opportunities.
   * Focus on promoting complementary products in cities with higher spending (like *London* and *Prague*).
5. **Growth Potential in Emerging Markets:**
   * Invest in cities with lower current spend but potential for growth, like *São Paulo*, *Buenos Aires*, and *Bangalore*.

**Recommendations:**

* Premium Markets: Tailor high-end product offerings, loyalty rewards, and exclusive promotions.
* Budget-friendly Strategies: Offer affordable options in low-spending markets and target them with localized promotions.
* Churn Reduction: Work on improving customer service and engagement to reduce churn in regions with high rates.
* Cross-sell: Leverage product affinities to increase sales in key cities.
* Future Growth: Keep an eye on emerging regions for future investment opportunities

A graph of blue lines with black text

AI-generated content may be incorrect.

.

**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?  
  
Query**SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

c.city,

c.state,

c.country,

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

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

DATEDIFF(CURDATE(), MAX(i.invoice\_date)) AS days\_since\_last\_purchase,

CASE

WHEN SUM(il.quantity \* il.unit\_price) < 100 THEN 'Low Spender'

WHEN COUNT(i.invoice\_id) < 5 THEN 'Infrequent Shopper'

WHEN DATEDIFF(CURDATE(), MAX(i.invoice\_date)) > 90 THEN 'At Risk'

ELSE 'Active'

END AS risk\_profile

FROM customer c

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

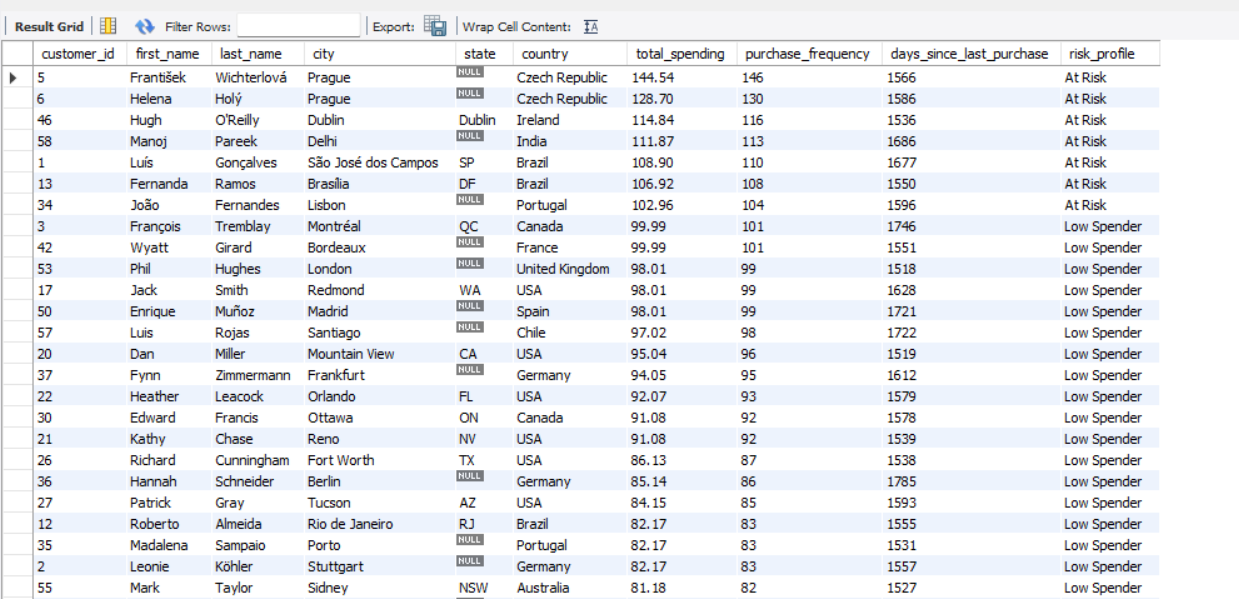
JOIN invoice\_line il ON i.invoice\_id = il.invoice\_id

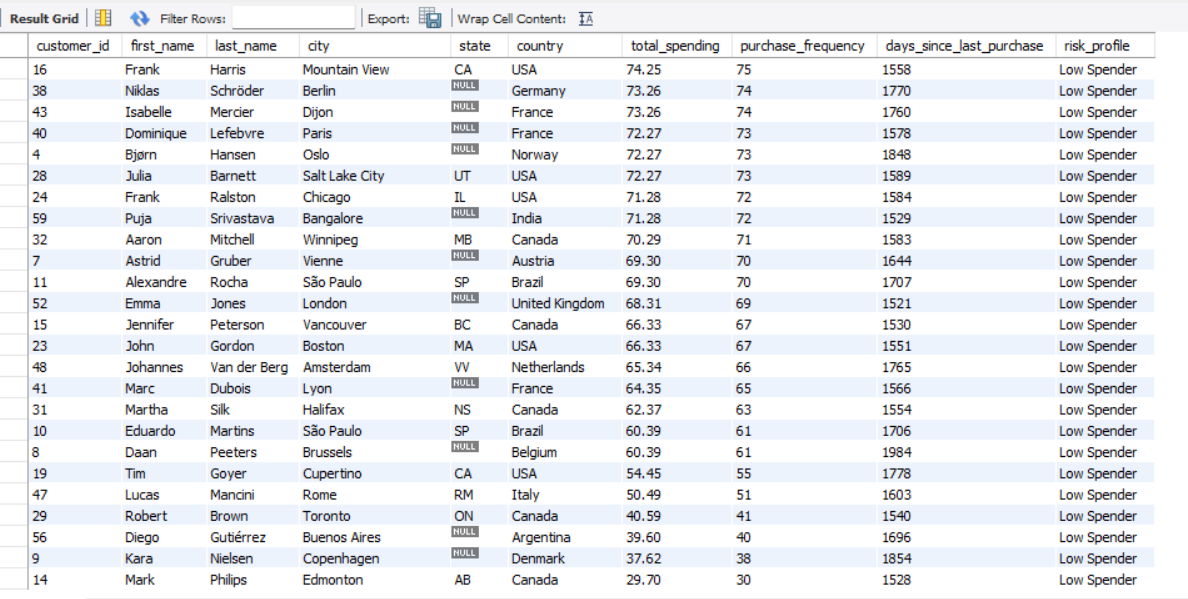
GROUP BY c.customer\_id

HAVING risk\_profile IN ('At Risk', 'Low Spender', 'Infrequent Shopper')

ORDER BY total\_spending DESC;

**Result**



****

customers with low spending, infrequent purchases, and high days since the last purchase are at the highest risk of churn and reduced spending.  
  
  
**Key Factors Contributing to Risk:**

1. **Low Spending**: Customers with lower total spending or infrequent purchases are more likely to be at risk.
2. **High Days Since Last Purchase**: Longer gaps since the last purchase can indicate disengagement, increasing the likelihood of churn.
3. **Purchase Frequency**: A lower frequency of purchases suggests reduced customer loyalty and engagement.
4. **Location**: In some regions, customers may have lower overall spending habits or fewer repeat purchases, contributing to higher risk.
5. **Risk Profile**: Customers marked as "At Risk" have a combination of the above factors and need special attention to avoid churn.

### **Segments Likely to Churn:**

* **At Risk Profile**: Customers with a low spending history, a high number of days since their last purchase, and infrequent purchases (e.g., customers from cities like Prague, São Paulo, or Delhi).
* **Low Spender Segment**: Customers consistently making smaller purchases over time are more likely to reduce spending further, especially if they are not regularly engaged.

### **Contributing Factors:**

* Lack of recent engagement (high days since last purchase).
* Lower total spending combined with low purchase frequency.
* Inconsistent engagement patterns, particularly for international customers.

A graph with different colored lines

AI-generated content may be incorrect.

**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?**

**Query**WITH CustomerLifetime AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

c.city,

c.state,

c.country,

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

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

DATEDIFF(CURDATE(), MIN(i.invoice\_date)) AS customer\_tenure, -- Customer tenure in days

DATEDIFF(CURDATE(), MAX(i.invoice\_date)) AS days\_since\_last\_purchase,

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

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

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, c.city, c.state, c.country -- Include city, state, country here

),

CLVPrediction AS (

SELECT

customer\_id,

total\_spending,

purchase\_frequency,

customer\_tenure,

(total\_spending / purchase\_frequency) AS avg\_purchase\_value, -- Average value per purchase

(customer\_tenure / 365) AS customer\_lifespan\_years, -- Approximate customer lifespan in years

CASE

WHEN days\_since\_last\_purchase > 90 THEN 'At Risk'

ELSE 'Active'

END AS churn\_status,

city,

state,

country

FROM CustomerLifetime

)

SELECT

customer\_id,

avg\_purchase\_value \* purchase\_frequency \* customer\_lifespan\_years AS predicted\_clv, -- CLV prediction

churn\_status,

city,

state,

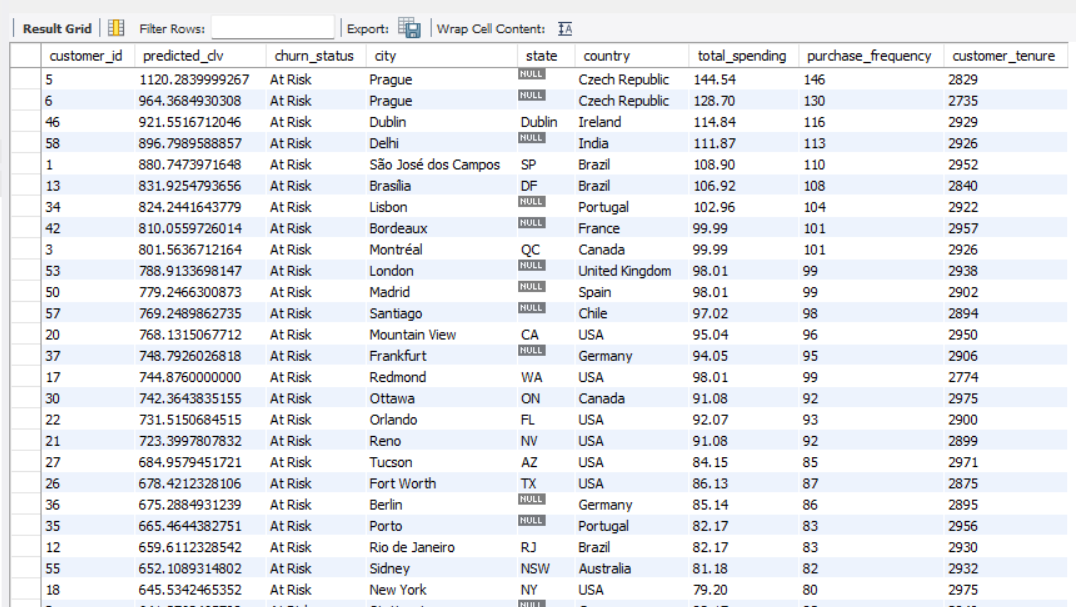
country,

total\_spending,

purchase\_frequency,

customer\_tenure

FROM CLVPrediction

ORDER BY predicted\_clv DESC;  
  
  
**Result**  


****

**Customers Who Have Stopped Purchasing:**

* **Declining Purchase Frequency**: Many customers reduce purchases before stopping completely. E.g., customers with low purchase frequency (like customer 9, with just 37.62 in spending) are at risk.
* **Low Spending**: Customers with a significant drop in total spending, like customer 14 (spending just $29.70), may signal they’ve stopped purchasing.
* **Inactivity**: Customers not purchasing recently (e.g., customers like 14 who have been inactive for months) are likely to churn.
* **Engagement Drop**: Customers with poor engagement, such as those who don’t open emails or engage with promotions, are more likely to stop purchasing.

**Targeted Strategies:**

1. **Loyalty Programs**: Reward long-term customers with offers, especially those with a long tenure but low recent purchases.
2. **Re-engagement Campaigns**: For inactive customers, personalized offers (e.g., discounts, reminders) can bring them back.
3. **Segmented Offers**: High-value customers can receive exclusive rewards, while low-value customers might be incentivized to spend more with targeted promotions.  
     
     
   **8.If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?**

**Answer-** If the data on promotional campaigns was available, I would have used it to analyse its impact on following factors:

1. **Customer Acquisition**

* The increase in the number of customers with time.
* Number of people participating in the events held in different locations.

1. **Retention**

* Number of old customers attending the event and then purchasing the track again after long time.
* Number of customers getting a discount.

1. **Sales**

* Increase in the sales due to promotional campaigns.
* Trends to analyse which promotional campaign was the best.
* New customers being generated due to discounts.

**Conclusion:**

By systematically tracking and analysing these metrics, Chinook can determine the effectiveness of promotional campaigns, identify areas for improvement, and optimize strategies for customer acquisition, retention, and revenue growth.

A graph of a number of blue lines

AI-generated content may be incorrect.

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

**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?**

**Query**

Alter table album

Add column release\_year INT DEFAULT 0;

DESC album;

**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.**

**Query**

SELECT

customer\_summary.country,

COUNT(DISTINCT customer\_summary.customer\_id) AS num\_customers,

AVG(customer\_summary.total\_spent) AS avg\_total\_spent,

AVG(customer\_summary.total\_tracks) AS avg\_tracks\_purchased

FROM (

SELECT

c.customer\_id,

c.country,

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

SUM(il.quantity) AS total\_tracks

FROM customer c

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

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

GROUP BY c.customer\_id, c.country

) AS customer\_summary

GROUP BY customer\_summary.country;

The average total amount spent per customer seems to be in the range of 8 to 12 approximately with an outlier of 5.158 in Denmark.The number of customers are very few (single digits) in most of the countries.

USA is the country with highest number of customers.

**Result**

