**Objective Questions**

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

**Ans.** Yes, some columns in the given tables does have some missing values. They can be found by running query.

**Syntax:**

Select \* from table\_name

Where column\_name is Null;

**Example in our data:**

SELECT \* FROM customer

WHERE company is Null;

Missing values are found in the following Tables and Columns in the dataset:

**Table name**: customer, **Column Name:** company, state, postal\_code, phone, fax

**Table Name**: track, **Column Name**: composer

To handle these missing values, we can update the value using “Update” statement

**Queries:**

update customer

set company='Unknown

where company is NULL;

update customer

set state='Not Mentioned’

where state is NULL;

update customer

set postal\_code='Unknown’

where postal\_code is NULL;

update customer

set phone='Not Mentioned’

where phone is NULL;

update customer

set fax='Not Mentioned’

where fax is NULL;

update track

set composer='Not Mentioned’

where composer is NULL;

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

**Ans.** **Approach:** To find the top selling track, top artist and most famous genre in USA, we need “country” column from “customer” table, “name” column from “track” table, “quantity” column from “invoice\_line” table, “name” column from “artist” table and “name” from “genre” table. We will perform joins to mentioned tables to get the required data.

**-- Top selling track in USA**

SELECT t.name as track\_name, SUM(il.quantity) AS quantity

FROM customer c

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

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

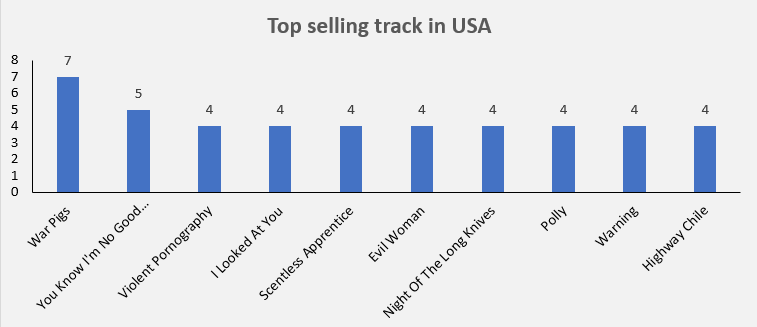
JOIN track t ON il.track\_id = t.track\_id

WHERE c.country = 'USA'

GROUP BY t.name

ORDER BY quantity DESC;

**Observation:** By running above query, we get to know that top selling track in USA is “War Pigs” with highest quantity of sales i.e. 7.



**-- Top Artist in USA**

SELECT art.name as artist\_name, SUM(il.quantity) AS quantity

FROM customer c

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

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

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

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

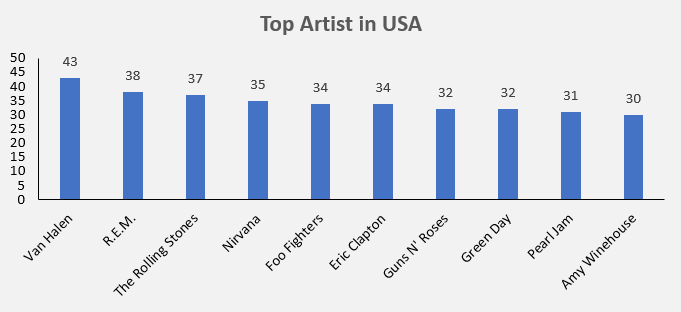
JOIN artist art ON a.artist\_id = art.artist\_id

WHERE c.country = 'USA'

GROUP BY art.name

ORDER BY quantity DESC;

**Observation:** By running above query, we get to know that top artist in USA is “Van Halen” followed by “R.E.M”.



**-- Most famous genre in USA**

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

FROM customer c

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

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

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

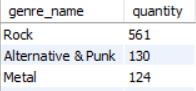
JOIN genre g ON t.genre\_id = g.genre\_id

WHERE c.country = 'USA'

GROUP BY g.name

ORDER BY quantity DESC;

**Observation:** By running above query, we get to know that most famous genre in USA is “Rock”.



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

**Ans.** **Approach:** We do not have “age” and “gender” data in our database, so we will perform customer demographic breakdown on basis of location. Which again we can categorized into “country”, “state” and “city”.

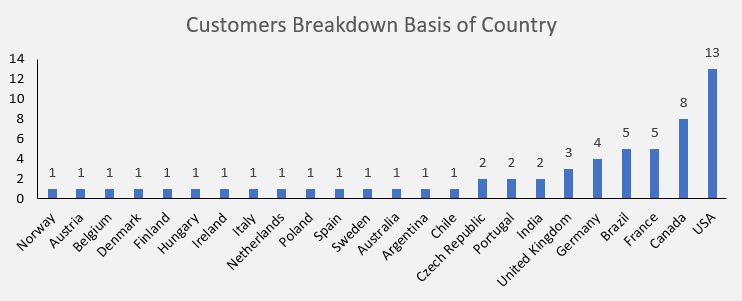
**-- Customer demographic breakdown on the basis of Country**

SELECT country, count(customer\_id) as number\_of\_customers

FROM customer

GROUP BY country

ORDER BY number\_of\_customers;



**-- Customer demographic breakdown on the basis of State**

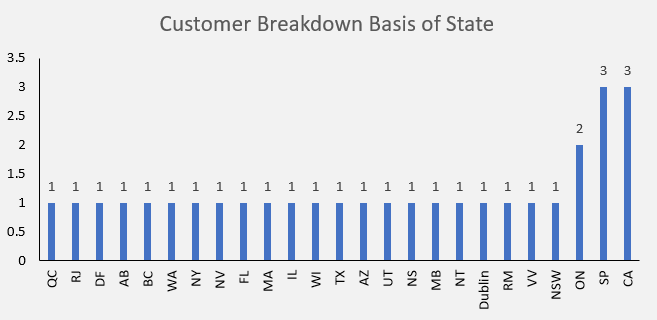
SELECT state, count(customer\_id) as number\_of\_customers

FROM customer

WHERE state != 'Not Mentioned'

GROUP BY state

ORDER BY number\_of\_customers;



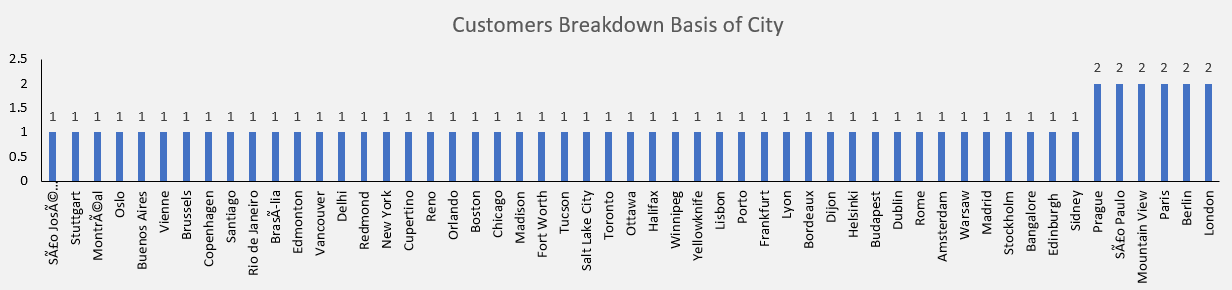
**-- Customer demographic breakdown on the basis of City**

SELECT city, count(customer\_id) as number\_of\_customers

FROM customer

GROUP BY city

ORDER BY number\_of\_customers;



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

**Ans. Approach:** We can directly use “invoice” table to calculate the total revenue and number of invoices for each country, state, and city as it contains data of “billing\_country”, “billing\_state” and “billing\_city”, to calculate total revenue we need to get sum of total for each country, state and city.

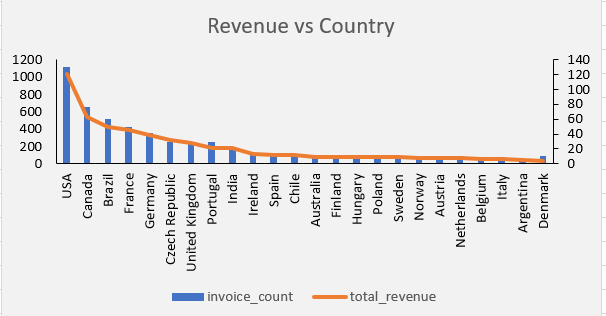
**-- Total revenue and number of invoices for each country**

SELECT billing\_country as country, count(invoice\_id) invoice\_count, sum(total) as total\_revenue

FROM invoice

GROUP BY billing\_country

ORDER BY total\_revenue desc, invoice\_count desc;



**-- Total revenue and number of invoices for each state**

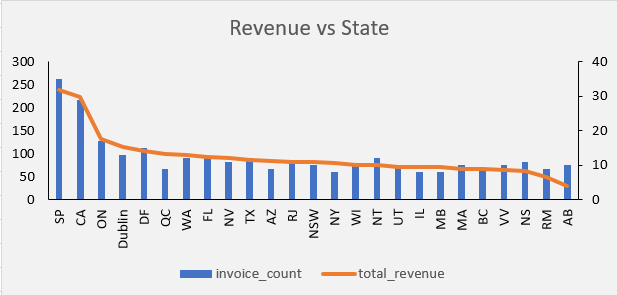
SELECT billing\_state as state, count(invoice\_id) invoice\_count, sum(total) as total\_revenue

FROM invoice

WHERE billing\_state != 'None'

GROUP BY billing\_state

ORDER BY total\_revenue desc, invoice\_count desc;



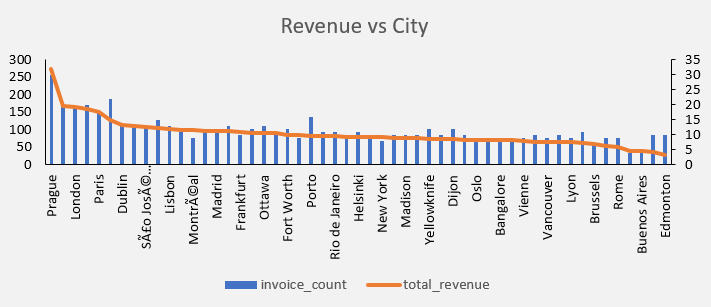
**-- Total revenue and number of invoices for each city**

SELECT billing\_city as city, count(invoice\_id) invoice\_count, sum(total) as total\_revenue

FROM invoice

GROUP BY billing\_city

ORDER BY total\_revenue desc, invoice\_count desc;



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

**Ans. Approach:** We can use CTE to generate each customer total revenue and assign them rank in the basis of decreasing total revenue using rank function using table “invoice”. In the main query we can get customer details like customer\_id, first\_name, last\_name from “customer” table by joining cte with “customer” table and give rank<=5 to get top 5 customers

**-- Top 5 customers by total revenue in each country**

with cte as

(SELECT customer\_id, billing\_country, sum(total) as revenue, rank() over(PARTITION BY billing\_country ORDER BY sum(total) desc) as `rank`

FROM invoice

GROUP BY customer\_id, billing\_country)

SELECT billing\_country as country, c.customer\_id, first\_name, last\_name, revenue, `rank`

FROM cte c

JOIN customer cc on c.customer\_id = cc.customer\_id

WHERE `rank` <=5

**Observation:**

• I have used rank() to represent the top 5 customers by total revenue in each country

• We can see the some counties don’t have much customers to show data for top 5 customers.



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

**Ans. Approach:** We can use “customer”, “invoice” and “track” tables to determine top-selling track for each customer by determine number to times particular track bought by the customer and the rank them.

**-- Top-selling track for each customer**

WITH CustomerTrackSales AS (

SELECT c.customer\_id, c.first\_name, c.last\_name, t.track\_id, t.name AS track\_name,

SUM(il.quantity) AS total\_quantity

FROM customer c

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

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

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name, t.track\_id, t.name),

TopTracks AS (

SELECT customer\_id, first\_name, last\_name, track\_id, track\_name, total\_quantity,

ROW\_NUMBER() OVER (PARTITION BY customer\_id ORDER BY total\_quantity DESC) AS `rank`

FROM CustomerTrackSales)

SELECT customer\_id, first\_name, last\_name, track\_name, total\_quantity

FROM TopTracks

WHERE `rank` = 1

ORDER BY customer\_id;

**Observation:** Maximum quantity purchased by any customer is 2.

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

**Ans. Approach:** For analysing any patterns or trends in customer purchasing behaviour we can extract year from invoice\_date and count the number of invoice\_id to get how many purchase done by customers in each year and calculate average of total to know average spending by customers in each year.

SELECT c.customer\_id, c.first\_name, c.last\_name, YEAR(i.invoice\_date) AS year,

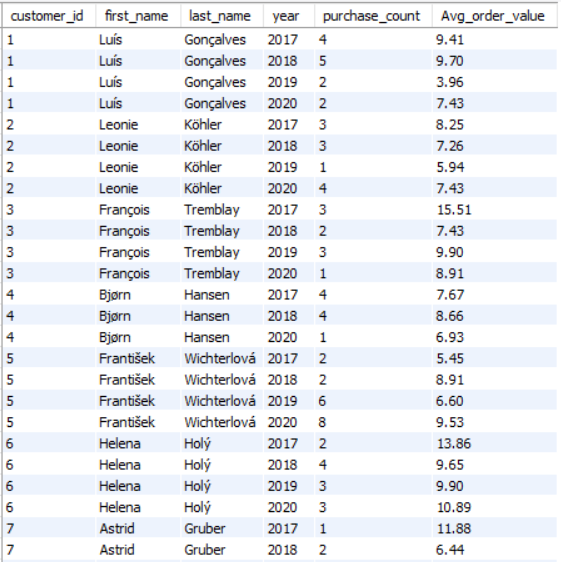
COUNT(i.invoice\_id) AS purchase\_count, round(avg(i.total),2) as Avg\_order\_value

FROM customer c

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name, YEAR(i.invoice\_date)

ORDER BY c.customer\_id, c.first\_name, c.last\_name, YEAR(i.invoice\_date);

**Observation:**

• After considering the above 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 over years.

• Aa well as we can observe that there is a significant decrease in the order value of some customers and increase in order value of some customers over years.

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

**Ans. Approach:** We can check most recent purchase by the customer and check if there is any gap in purchase with interval of a year. The number of customers having gap of 1 year in purchase are churn customers. To get churn rate we have to use below formula:

Churn Rate: (Total count of churn customers / Total customers) \* 100

With CutoffDate AS (

SELECT DATE\_SUB((SELECT MAX(invoice\_date) AS most\_recent\_invoice\_date

FROM invoice), INTERVAL 1 YEAR) AS cutoff\_date),

ChurnCustomers AS (

SELECT c.customer\_id, c.first\_name, c.last\_name, MAX(i.invoice\_date) AS latest\_purchase\_date

FROM customer c

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name

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

SELECT round((SELECT COUNT(\*) FROM ChurnCustomers) / COUNT(\*) \* 100, 2) AS churn\_rate

FROM customer;



**Observation:** From the above query we get to know that churn rate is 1.69, which is low.

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

**Ans. Approach:**  First we will count of each genre in USA and the count of total genre in USA to get the percentage sale of each genre by using formula (genre\_count / total\_count) \* 100.

To get best-selling Artist we will sum the quantity from invoice\_line table for each artist in USA.

-- Best-selling genres and percentage of total sales contributed by each genre in the USA

WITH genre\_counts AS (

SELECT g.name AS genre\_name, COUNT(g.genre\_id) AS genre\_count

FROM customer c

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

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

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

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

WHERE c.country = 'USA'

GROUP BY g.name),

total\_count AS (

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

FROM customer c

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

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

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

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

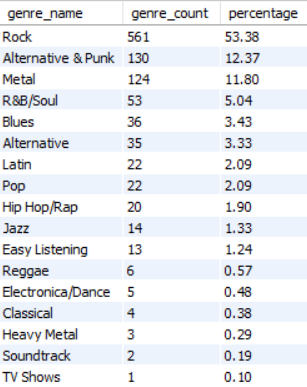
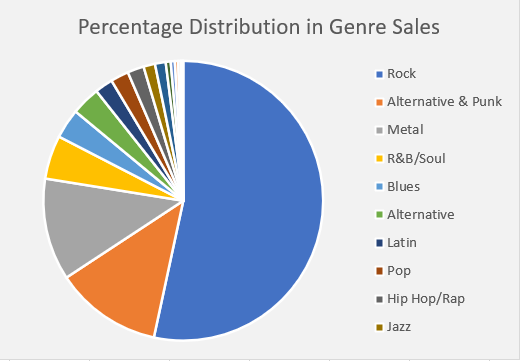
WHERE c.country = 'USA')

SELECT gc.genre\_name, gc.genre\_count, round((gc.genre\_count / tc.total\_count) \* 100,2) AS percentage

FROM genre\_counts gc

CROSS JOIN total\_count tc

ORDER BY gc.genre\_count DESC;

**Observation:** Top selling genre is Rock with a count of 561 and contributed 53.38% of total sales followed by Alternative & Punk genre with count of 130 and 12.37% of contribution in total sales.

-- Best-selling Artist in USA

SELECT art.name as artist\_name, SUM(il.quantity) AS quantity

FROM customer c

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

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

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

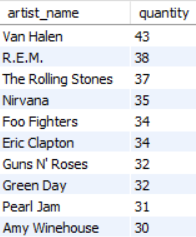
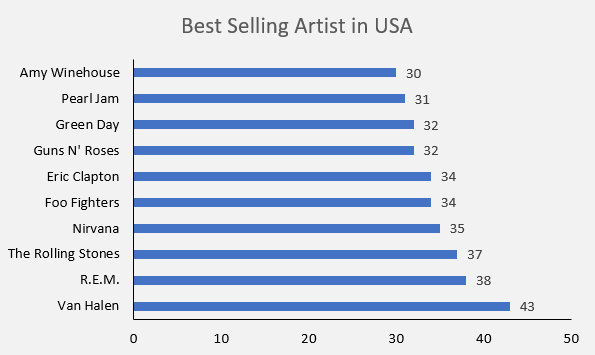
JOIN album a on t.album\_id = a.album\_id

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

WHERE c.country = 'USA'

GROUP BY art.name

ORDER BY quantity DESC;

** **

**Observation:** Best selling Artist in USA is Van Halen with total quantity of 43 followed by R.E.M with 38 quantity.

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

**Ans. Approach:** To get customers who have purchased tracks from at least 3 different genres we need to count distinct genre purchased by each customer and add condition count of distinct genre grater than and equal to 3.

SELECT c.customer\_id, c.first\_name, c.last\_name, count(distinct g.genre\_id) AS genre\_count

FROM customer c

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

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

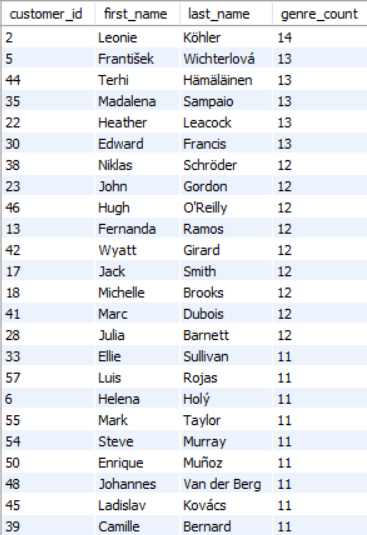
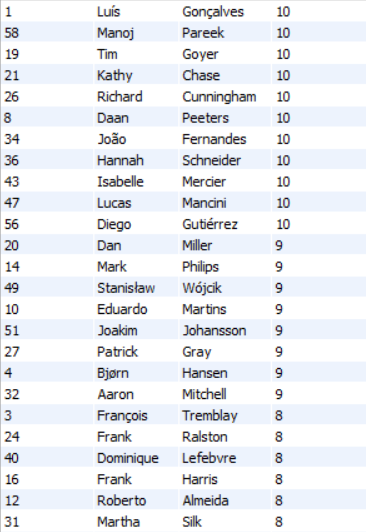
JOIN track t ON il.track\_id = t.track\_id

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name

HAVING count(distinct g.genre\_id) >=3

ORDER BY genre\_count DESC

**Observation:** There are significant number of customers who purchased at least 3 different genres.

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

**Ans. Approach:** To get top sale performance genre we can sum the total from invoice for each genre and rank them up from highest total to least in the USA country.

SELECT g.name, sum(i.total) as genre\_sum, RANK() OVER(ORDER BY sum(i.total) DESC) as `rank`

FROM customer c

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

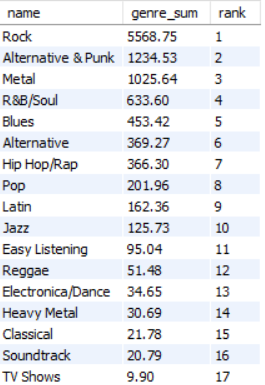
JOIN invoice\_line il on i.invoice\_id = il.invoice\_id

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 g.name



**Observation:** From the above data we can see that Genre Rock stays top in the list with a total sales of $5568.75 followed by Alternative & Punk with total sales of $1234.53

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

**Ans. Approach:** We will create a CTE which will contain customers who have made purchase in last 3

months. In the main query we can find the last purchase date of a customer and check if it is not

present in the CTE table.

WITH last\_3\_month as

(SELECT customer\_id

FROM invoice

WHERE DATE(invoice\_date) BETWEEN '2020-10-01' AND '2020-12-30')

SELECT c.customer\_id, c.first\_name, c.last\_name, max(DATE(invoice\_date)) as last\_purchase\_date

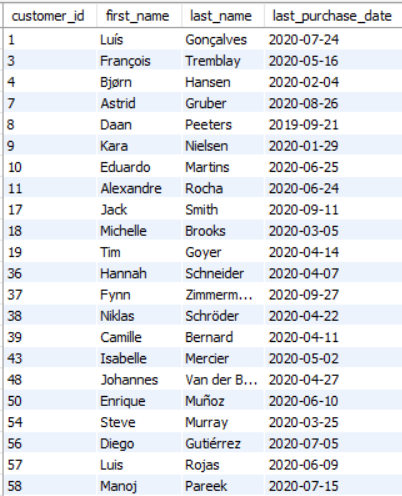
FROM customer c

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

WHERE c.customer\_id not in (select \* from last\_3\_month)

GROUP BY c.customer\_id, c.first\_name, c.last\_name

ORDER BY c.customer\_id;



**Observation:** We can see that there are total 22 customers who have not purchased from last 3 months

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.**

**Ans. Approach:** Based on the genre sales analysis, the three albums that should be prioritized for advertising and promotion in the USA from the new record label should be the top-selling genres.

SELECT g.genre\_id, g.name AS genre\_name, al.album\_id, al.title as album\_title,

SUM(i.total) AS total\_genre\_sales,

DENSE\_RANK() OVER (ORDER BY SUM(i.total) DESC) AS `rank`

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

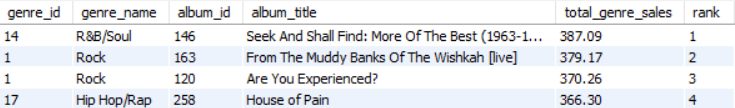
JOIN genre g ON g.genre\_id = t.genre\_id

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;



**Observation:**

• R&B/Soul Genre: As the highest-selling genre in the USA with a total sales of 555.39, a R&B/Soul album would have the strongest potential for mass appeal and high returns.

• Rock Genre: With total sales of 128.7, Rock genre ranks second in performance, making it a strong candidate for promotion.

• Hip Hop/Rap Genre– With sales of 122.76, Hip Hop/Rap is another top-performing genre. Prioritizing a Hip Hop/Rap album could attract a dedicated and loyal fan base, ensuring solid sales.

Based on above top-selling genres, below are the 3 albums that can be prioritised for advertising and promotion in the USA:

• Seek And Shall Find: More Of The Best (1963-1981) (R&B/Soul Genre)

• From The Muddy Banks Of The Wishkah [live] (Rock Genre)

• House of Pain (Hip Hop/Rap Genre)

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

**Ans. Approach:** To get top-selling genres we will sum the total from invoice for each genre in each country to know top-selling genres in countries.

SELECT i.billing\_country, g.name AS genre\_name,

SUM(i.total) AS total\_genre\_sales,

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

FROM invoice i

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

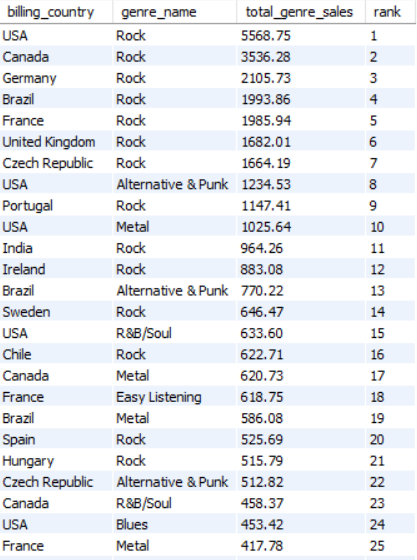
JOIN track t ON t.track\_id = il.track\_id

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

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

GROUP BY i.billing\_country, g.name

ORDER BY total\_genre\_sales DESC;



**Observation:** From the above query we can see that Rock Genre is not only top-selling genre in USA but also top selling genre in other countries like Canada, Germany, Brazil, France, United Kingdom, Czeh Republic, Portugal, India, Ireland Sweden, Chile, Spain, Hungaryand many more.

**3. 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?**

**Ans. Approach:** To get customer purchase behaviour analysis I have created a CTE which contains customers details with their difference in days between first and last purchase date. In another CTE I have distinguished customers in the basis of Long Term if the difference in days between first and last purchase date is greater than or equal to 365 days, else customers are new. In the Main Query I have round off the data to 2 decimal place for understanding.

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;



**Observation:**

• Long-term customers typically have a higher purchase frequency, indicating stronger loyalty.

• A larger basket size for long-term customers suggests they tend to purchase more items per order.

• Long-term customers are likely to spend more in total and have a higher average spending per order, which can be an indicator of trust and satisfaction with the service.

**Retention Strategies:**

• For new customers: Focus on increasing purchase frequency and encouraging larger basket sizes through targeted promotions, discounts, or personalized recommendations.

• For long-term customers: Long-term customers, on average, make around 10 purchases and average spending over 800. Continue nurturing loyalty with rewards programs, personalized offers, or exclusive deals. Their higher spending and frequency show strong engagement, so retention efforts can focus on deepening relationships and encouraging referrals.

**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?**

**Ans. Approach:** To conduct a Product Affinity Analysis in the context of music purchases, we can explore which music genres, artists, or albums are frequently bought together.

By analyzing the affinity between music genres, artists, or albums, businesses can create more personalized and targeted recommendations for individual customers. Offering discounts or promotions on genres, artists, or albums that are often bought together can increase sales.

-- Analyze Genre Affinities:

SELECT g1.name AS genre\_1, g2.name AS genre\_2, COUNT(\*) AS purchase\_count

FROM invoice i

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

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

JOIN genre g1 ON t.genre\_id = g1.genre\_id

JOIN invoice\_line il2 ON i.invoice\_id = il2.invoice\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

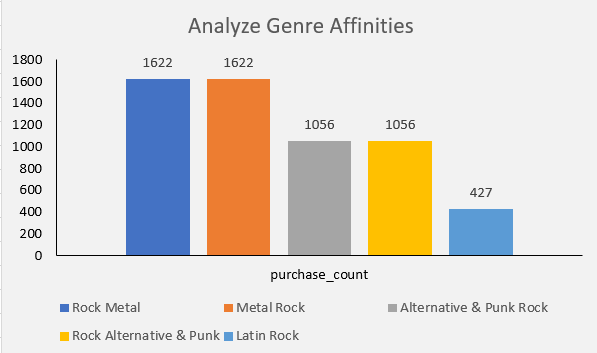
JOIN genre g2 ON t2.genre\_id = g2.genre\_id

WHERE g1.genre\_id != g2.genre\_id

GROUP BY g1.name, g2.name

ORDER BY purchase\_count DESC

LIMIT 5;

**Observation:**

• From the above query we can see that Rock and Metal genres are frequently bought together followed by Alternative & Punk and Rock genres are also preferred by customers they bought together.

-- Analyze Artist Affinities:

SELECT a.name AS artist\_1, a2.name AS artist\_2, COUNT(\*) AS purchase\_count

FROM invoice\_line il

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 invoice\_line il2 ON il.invoice\_id = il2.invoice\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

JOIN album al2 ON t2.album\_id = al2.album\_id

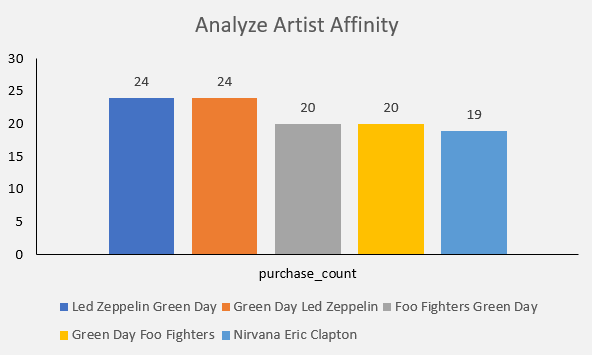
JOIN artist a2 ON al2.artist\_id = a2.artist\_id

WHERE a.artist\_id != a2.artist\_id

GROUP BY a.name, a2.name

ORDER BY purchase\_count DESC

LIMIT 5;

**Observation:**

• From the above query we can see that Led Zeppelin and Green Day artists are frequently bought together also customer bought Foo Fighters with Green Day are also preferred by customers they bought together.

-- Analyze Album Affinities:

SELECT al.title AS album\_1, al2.title AS album\_2, COUNT(\*) AS purchase\_count

FROM invoice\_line il

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

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

JOIN invoice\_line il2 ON il.invoice\_id = il2.invoice\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

JOIN album al2 ON t2.album\_id = al2.album\_id

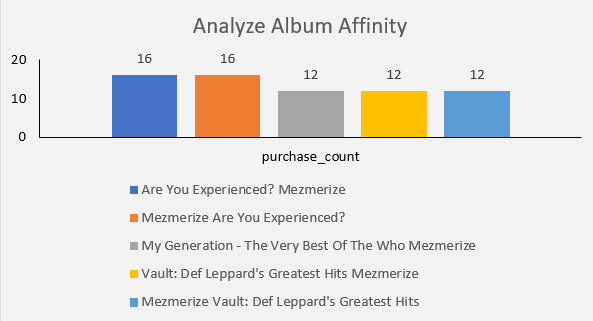
WHERE al.album\_id != al2.album\_id

GROUP BY al.title, al2.title

ORDER BY purchase\_count DESC

LIMIT 5;





**Observation:**

• From the above query we can see that Albums “Are You Experienced?” and “Mezmerize” are frequently bought together also customer bought albums “My Generation – The Very Best Of The Who and Vault: Def Leppard’s Greatest Hits” with “Mezmerize” are also preferred by customers they bought together.

**5. 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?**

**Ans. Approach:** The below query provides comprehensive view of customer activity by region, their purchase frequency, summing total spent, average basket size and calculating the churn rate. It highlights differences in customer purchasing behaviour across various geographic locations.

WITH frequency AS (

SELECT billing\_country, AVG(purchase\_frequency) as avg\_days\_bw\_purchase

FROM (SELECT customer\_id, billing\_country, invoice\_date, next\_date, DATEDIFF(next\_date,invoice\_date) AS purchase\_frequency

FROM (SELECT \*, LEAD(invoice\_date) OVER(PARTITION BY customer\_id) AS next\_date FROM invoice

ORDER BY customer\_id) a) b

GROUP BY billing\_country),

basket\_size AS (

SELECT billing\_country, ROUND(AVG(basket\_size),2) AS avg\_basket\_size

FROM (SELECT customer\_id, billing\_country, i.invoice\_id, COUNT(invoice\_line\_id) AS basket\_size FROM invoice i

JOIN invoice\_line i1 ON i.invoice\_id=i1.invoice\_id

GROUP BY customer\_id, billing\_country, invoice\_id) a

GROUP BY billing\_country),

spending AS(

SELECT billing\_country, SUM(total) AS total\_spent FROM invoice

GROUP BY billing\_country),

last\_purchase as (

SELECT customer\_id, billing\_country, MAX(DATE(invoice\_date)) AS last\_purchase\_date FROM invoice

GROUP BY customer\_id,billing\_country

ORDER BY customer\_id),

active\_or\_inactive as (

SELECT customer\_id, last\_purchase\_date, billing\_country, CASE WHEN last\_purchase\_date > DATE\_SUB('2020-12-31',INTERVAL 12 MONTH) THEN 'Active Customer'

WHEN last\_purchase\_date < DATE\_SUB('2020-12-31',INTERVAL 12 MONTH) THEN 'Churned Customer' END AS customer\_status FROM last\_purchase),

churned\_count AS (

SELECT billing\_country, customer\_status, COUNT(customer\_id) number\_of\_customers FROM active\_or\_inactive

GROUP BY billing\_country, customer\_status),

churn\_rate as (

SELECT billing\_country, ROUND((SUM(CASE WHEN customer\_status='Churned Customer' THEN number\_of\_customers ELSE 0 END)/SUM(number\_of\_customers))\*100,2) AS churn\_rate FROM churned\_count

GROUP BY billing\_country

ORDER BY billing\_country)

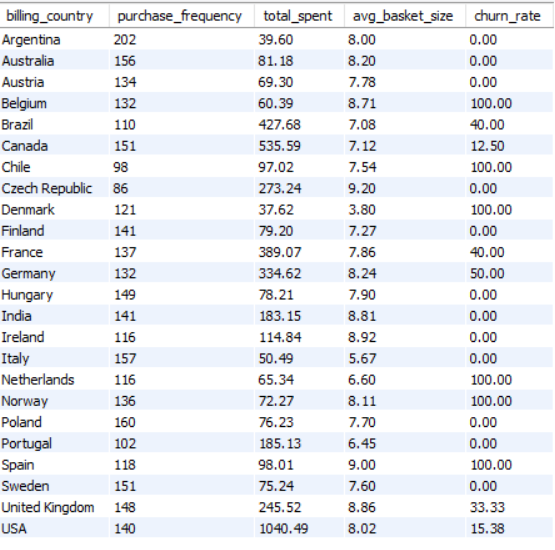
SELECT s.billing\_country, ROUND(avg\_days\_bw\_purchase,0) AS purchase\_frequency, total\_spent, avg\_basket\_size, churn\_rate FROM spending s

JOIN basket\_size b ON s.billing\_country=b.billing\_country

JOIN frequency f ON s.billing\_country=f.billing\_country

JOIN churn\_rate c ON s.billing\_country=c.billing\_country

ORDER BY billing\_country;



**Observation:**

• The purchase frequency is not varied too much among the different countries.

• USA is leading in terms of amount spending, this correlates with the customer engagement and higher average income.

• Due to a less global catalogue of music, there is a huge contrast among the different countries in terms of spending.

• The churn rate is very high in countries like Belgium, Chile, Denmark, Netherland and Norway.

**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?**

**Ans. Approach:** As no data is available for customers age and gender, we can use location in multiple segments as city, state and country. Evaluated high risk customers who have not made purchase in last 365 days and hence have a high chance of churn customers and poses a higher risk of reduced spending.

WITH Customer\_Segments AS (

SELECT c.customer\_id, c.city, c.state, c.country, COUNT(i.invoice\_id) AS purchase\_count, SUM(i.total) AS total\_spending,

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

FROM customer c

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

WHERE c.state != ‘Not Mentioned’

GROUP BY c.customer\_id, c.city, c.state, c.country),

Risk\_Profiles AS (

SELECT cs.customer\_id, cs.city, cs.country, cs.purchase\_count, cs.total\_spending, cs.days\_since\_last\_purchase,

CASE WHEN cs.days\_since\_last\_purchase > 365 THEN 'High Risk' WHEN cs.days\_since\_last\_purchase BETWEEN 180 AND 365 THEN 'Moderate Risk'

ELSE 'Low Risk' END AS churn\_risk

FROM Customer\_Segments cs)

SELECT churn\_risk, COUNT(customer\_id) AS customer\_count, ROUND(AVG(total\_spending),2) AS avg\_spending, ROUND(AVG(purchase\_count),2) AS avg\_purchase\_count

FROM Risk\_Profiles

GROUP BY churn\_risk

ORDER BY FIELD(churn\_risk, 'High Risk', 'Moderate Risk', 'Low Risk');



**Observation:**

Factors Contributing to Churn Risk:

• High Risk Segment: These 30 out of 59 customers haven't made a purchase in over a year, which is a clear indicator of potential churn. They have a average spending ($77.19) and average purchase count (10.33), which indicates higher risk of reduced spending.

• Days Since Last Purchase: Since they haven’t made a purchase in more than a year, they are highly likely to churn unless re-engaged.

**Recommendation:**

• Focusing on re-engaging this high-risk segment by giving them offers, new recommendation we could potentially recover some of these customers and prevent further revenue loss.

**7. 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?**

**Ans. Approach:** To predict Customer Lifetime Value, we can leverage customer data such as tenure, purchase history, and engagement patterns.

• 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 to know which customer has stopped purchasing from a long time.

WITH last\_purchase AS (

SELECT customer\_id, billing\_country, max(DATE(invoice\_date)) AS last\_purchase\_date FROM invoice

GROUP BY customer\_id,billing\_country

ORDER BY customer\_id),

active\_or\_inactive AS (

SELECT customer\_id, last\_purchase\_date, billing\_country, CASE WHEN last\_purchase\_date > DATE\_SUB('2020-12-31', INTERVAL 6 MONTH) THEN 'Active Customer'

WHEN last\_purchase\_date < DATE\_SUB('2020-12-31', INTERVAL 6 MONTH) THEN 'Churned Customer' END AS status FROM last\_purchase),

churned\_cust AS (

SELECT customer\_id FROM active\_or\_inactive

WHERE STATUS = 'Churned Customer'),

active\_cust AS (

SELECT customer\_id FROM active\_or\_inactive

WHERE status = 'Active Customer'),

frequency AS (

SELECT customer\_id, AVG(purchase\_frequency) AS avg\_days\_bw\_purchase

FROM (SELECT customer\_id, invoice\_date, next\_date, DATEDIFF(next\_date, invoice\_date) AS purchase\_frequency

FROM (SELECT \*, LEAD(invoice\_date) OVER(PARTITION BY customer\_id) AS next\_date FROM invoice

ORDER BY customer\_id) a) b

GROUP BY customer\_id),

avg\_days AS (

SELECT 'Active Customer' AS status, ROUND(AVG(avg\_days\_bw\_purchase),2) AS average\_frequency FROM frequency

WHERE customer\_id IN (SELECT \* FROM active\_cust)

UNION

SELECT 'Inactive Customer' as status, ROUND(AVG(avg\_days\_bw\_purchase),2) AS average\_frequency FROM frequency

WHERE customer\_id IN (SELECT \* FROM churned\_cust)),

inactive\_purchase AS (

SELECT \* FROM invoice\_line

WHERE invoice\_id IN (SELECT invoice\_id FROM invoice

WHERE customer\_id IN (SELECT customer\_id FROM churned\_cust))),

inactivity AS (

SELECT ip.track\_id, invoice\_line\_id, genre\_id, album\_id FROM inactive\_purchase ip

JOIN track t on ip.track\_id=t.track\_id),

genrestat AS (

SELECT genre\_id, COUNT(invoice\_line\_id) genre\_sale FROM inactivity

GROUP BY genre\_id

ORDER BY genre\_sale DESC)

SELECT gg.name AS Genre, genre\_sale FROM genrestat g

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

ORDER BY genre\_sale DESC;

-- frequency of inactive customers

WITH last\_purchase AS (

SELECT customer\_id, billing\_country, MAX(DATE(invoice\_date)) AS last\_purchase\_date FROM invoice

GROUP BY customer\_id, billing\_country

ORDER BY customer\_id),

active\_or\_inactive AS (

SELECT customer\_id, last\_purchase\_date, billing\_country, CASE WHEN last\_purchase\_date > DATE\_SUB('2020-12-31', INTERVAL 6 MONTH) THEN 'Active Customer'

WHEN last\_purchase\_date < DATE\_SUB('2020-12-31', INTERVAL 6 MONTH) THEN 'Churned Customer' END AS status FROM last\_purchase),

churned\_cust AS (

SELECT customer\_id FROM active\_or\_inactive

where status='Churned Customer'),

active\_cust AS (

SELECT customer\_id FROM active\_or\_inactive

WHERE STATUS='Active Customer'),

frequency AS (

SELECT customer\_id, AVG(purchase\_frequency) AS avg\_days\_bw\_purchase

FROM (SELECT customer\_id, invoice\_date, next\_date, DATEDIFF(next\_date, invoice\_date) AS purchase\_frequency

FROM (SELECT \*,LEAD(invoice\_date) OVER(PARTITION BY customer\_id) AS next\_date FROM invoice

ORDER BY customer\_id) a) b

GROUP BY customer\_id),

avg\_days AS (

SELECT 'Active Customer' AS status, ROUND(AVG(avg\_days\_bw\_purchase),2) AS average\_frequency FROM frequency

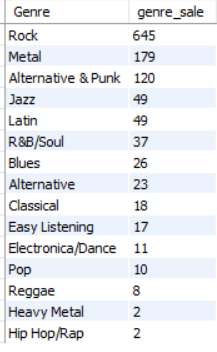
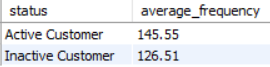
WHERE customer\_id IN (SELECT \* FROM active\_cust)

UNION

SELECT 'Inactive Customer' AS status, ROUND(AVG(avg\_days\_bw\_purchase),2) AS average\_frequency FROM frequency

WHERE customer\_id IN (SELECT \* FROM churned\_cust))

SELECT \* FROM avg\_days;

**Observation:**

• We can the most of the churned customers are from genre Rock and Metal

• Average frequency of day between purchase date of Inactive Customers are less than the Active Customers. This means Inactive Customers were more frequent buyers than Active Customers.

**Retention Strategies for Customers at Risk of Churning:**

• Re-engagement Campaigns: Offer discounts or incentives for customers who haven’t purchased recently.

• Personalized Offers: Offer targeted promotions based on past behaviour. For example, sending personalized email offers that feature products similar to those they previously purchased could revive interest.

**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?**

**Ans. Approach:** If we have data on promotional campaigns, such as discounts, events, or email marketing, we could evaluate the impact of these campaigns on customer acquisition, retention, and overall sales using the following method:

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.

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

**Ans. Approach:** I would follow an approach focused on discovering meaningful insights to inform strategy if the objective and subjective questions weren’t given.

• I will start by understanding the Chinook’s Model, their Market, and their specific goals within the physical music market.

• By reviewing their past sales data, understanding the product offerings, for example: music genres, albums, playlist, artists etc.

• Identifying their key challenges or opportunities, for example how much sales they are expecting to sale for each artist, each genre etc.

• Identifying the performance of artists and their sales etc.

• Which type of customers love different genres, artist by categorizing their ages.

• Finding which region is in top sales for each genre.

• Finding which artist is having highest sales in each region.

• Also comparing which genre is highest in each region so that we can focus more on that genre and supply new albums to each and respective countries.

• Analyzing which all customers are long-term and short-term by seeing their past purchases.

• Based on this analyzing result, I would recommend the changes be made to increase more profit accordingly.

• Also I would suggest to provide new type of genre and new languages which can make a trending sector in some regions.

**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?**

**Ans.** To add a new column named ReleaseYear of type INTEGER to the Albums table, I have used the following SQL ALTER TABLE statement:

ALTER TABLE album

ADD COLUMN ReleaseYear INTEGER;

**11. Chinook is interested in understanding the purchasing behaviour 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.**

**Ans. Approach:** Firstly, calculated the total amount spent and total tracks purchased for each customer, grouped by their country. Then aggregated this data to find:

1. Number of customers per country.

2. Average total spent per customer.

3. Average number of tracks purchased per customer. It uses a WITH clause (CTE) for easier grouping and joins the necessary tables (customer, invoice, and invoice\_line) to gather purchase data.

WITH Customer\_Purchases AS (

SELECT country, c.customer\_id, SUM(i.total) AS total\_spent, COUNT(il.track\_id) 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 country, c.customer\_id)

SELECT country, COUNT(customer\_id) AS number\_of\_customers, ROUND(AVG(total\_spent),2) AS avg\_total\_spent,

ROUND(AVG(total\_tracks),2) AS avg\_tracks\_purchased\_per\_customer

FROM Customer\_Purchases

GROUP BY country

ORDER BY avg\_total\_spent DESC;



**Observation:**

• Top Spenders: Countries like the Czech Republic, Ireland, and Spain have the highest average total spending per customer, with totals more than $1,000 and more than 90 tracks purchased per customer.

• Moderate to Low Spending: The USA has the largest customer base i.e. 13, but the average spending per customer is around $800, much less compared to top spenders. Denmark has the lowest spending, with totals below $200.

**THANK YOU**

*- Ranajoy Dutta*