Chinook Music Data Analysis

**Objectives: -**

* To facilitate data-driven growth, our analysis will focus on key music store trends, including customer demographics, purchasing patterns, and churn rates.
* At Chinook, my objective is to extract valuable insights from music sales data by solving Objective and Subjective questions which are given, and also guiding strategic initiatives in the physical music market.

**Data Description: -**

1. Album Table: -

* album\_id: Unique identifier assigned to each album.
* title: The title or name of the album.
* artist\_id: The ID of the artist associated with the album.

1. Artist Table: -

* artist\_id: Unique identifier assigned to each artist.
* name: The name of the artist.

1. Customer Table: -

* customer\_id: Unique identifier assigned to each customer.
* first\_name: The given name or first name of a customer.
* last\_name: The surname or family name of a customer.
* company: The name of the company associated with a customer.
* address: The street address of a customer's location.
* city: The city where a customer is located.
* state: The state or province where a customer is located.
* country: The country where a customer is located.
* postal\_code: The postal or zip code of a customer's address.
* phone: The phone number of a customer.
* fax: The fax number associated with a customer.
* email: The email address of a customer.
* support\_rep\_id: The employee ID of the support representative assigned to a customer.

1. Employee Table: -
   * employee\_id: Unique identifier assigned to each employee.
   * last\_name: The surname or family name of a employee.
   * first\_name: The given name or first name of a employee.
   * Title: position of the employee within the company
   * reports\_to: employee\_id of the manager
   * birthdate: date of birth of the employee
   * hire\_date: date of employment
   * address: The street address of an employee’s location
   * city: The city where an employee is located.
   * state: The state or province where an employee is located
   * country: The country where an employee is located
   * postal\_code: The postal or zip code of an employee’s address
   * phone: the phone number of an employee
   * fax: the fax number associated with an employee
   * email: The email address of an employee
2. Genre Table Table: -

* genre\_id: Unique identifier assigned to each genre.
* name: The name or description of the genre (e.g., rock, pop, classical).

1. Invoice Table: -

* invoice\_id: Unique identifier assigned to each invoice.
* customer\_id: The customer ID associated with the invoice.
* invoice\_date: The date when the invoice was generated or issued.
* billing\_address: The street address used for billing purposes.
* billing\_city: The city used for billing purposes.
* billing\_state: The state or province used for billing purposes.
* billing\_country: The country used for billing purposes.
* billing\_postal\_code: The postal or zip code used for billing purposes.
* total: The total amount due on the invoice.

1. Invoice\_line Table: -

* invoice\_line\_id: Unique identifier assigned to each line item on an invoice.
* invoice\_id: The invoice ID to which the line item belongs.
* track\_id: The ID of the track or product included in the line item.
* unit\_price: The price per unit for the line item.
* quantity: The quantity of units for the line item.

1. Media\_type Table: -

* media\_type\_id: Unique identifier assigned to each media type.
* name: The name or description of the media type (e.g., MPEG audio file, AAC audio file).

1. Playlist Table: -

* playlist\_id: Unique identifier assigned to each playlist.
* name: The name or title of the playlist.

1. Playlist\_track Table: -

* playlist\_id: The ID of the playlist to which the track belongs.
* track\_id: The ID of the track included in the playlist.

1. Track Table: -

* track\_id: Unique identifier assigned to each track or song.
* name: The title or name of the track.
* album\_id: The ID of the album to which the track belongs.
* media\_type\_id: The ID of the media type associated with the track.
* genre\_id: The ID of the genre associated with the track.
* composer: The name of the composer or artist who composed the track.
* milliseconds: The duration of the track in milliseconds.
* bytes: The file size of the track in bytes.
* unit\_price: The price per unit for the track.

Chinook Music Data Analysis

* To start, we will create the Chinook Music store database by copying and running the code from the provided .sql file.
  + Extra whitespace has been removed from the database queries in the Chinook.sql file.
  + To define table relationships, foreign key constraints were added, and an ERD diagram was generated using MySQL Workbench.

*Foreign Key Costrains: -*

ALTER TABLE `album`

ADD CONSTRAINT `fk\_album\_artist`

FOREIGN KEY (`artist\_id`) REFERENCES `artist`(`artist\_id`);

ALTER TABLE `customer`

ADD CONSTRAINT `fk\_customer\_support\_rep`

FOREIGN KEY (`support\_rep\_id`) REFERENCES `employee`(`employee\_id`);

ALTER TABLE `invoice`

ADD CONSTRAINT `fk\_invoice\_customer`

FOREIGN KEY (`customer\_id`) REFERENCES `customer`(`customer\_id`);

ALTER TABLE `invoice\_line`

ADD CONSTRAINT `fk\_invoice\_line\_invoice`

FOREIGN KEY (`invoice\_id`) REFERENCES `invoice`(`invoice\_id`),

ADD CONSTRAINT `fk\_invoice\_line\_track`

FOREIGN KEY (`track\_id`) REFERENCES `track`(`track\_id`);

ALTER TABLE `track`

ADD CONSTRAINT `fk\_track\_album`

FOREIGN KEY (`album\_id`) REFERENCES `album`(`album\_id`),

ADD CONSTRAINT `fk\_track\_media\_type`

FOREIGN KEY (`media\_type\_id`) REFERENCES `media\_type`(`media\_type\_id`),

ADD CONSTRAINT `fk\_track\_genre`

FOREIGN KEY (`genre\_id`) REFERENCES `genre`(`genre\_id`);

ALTER TABLE `playlist\_track`

ADD CONSTRAINT `fk\_playlist\_track\_playlist`

FOREIGN KEY (`playlist\_id`) REFERENCES `playlist`(`playlist\_id`),

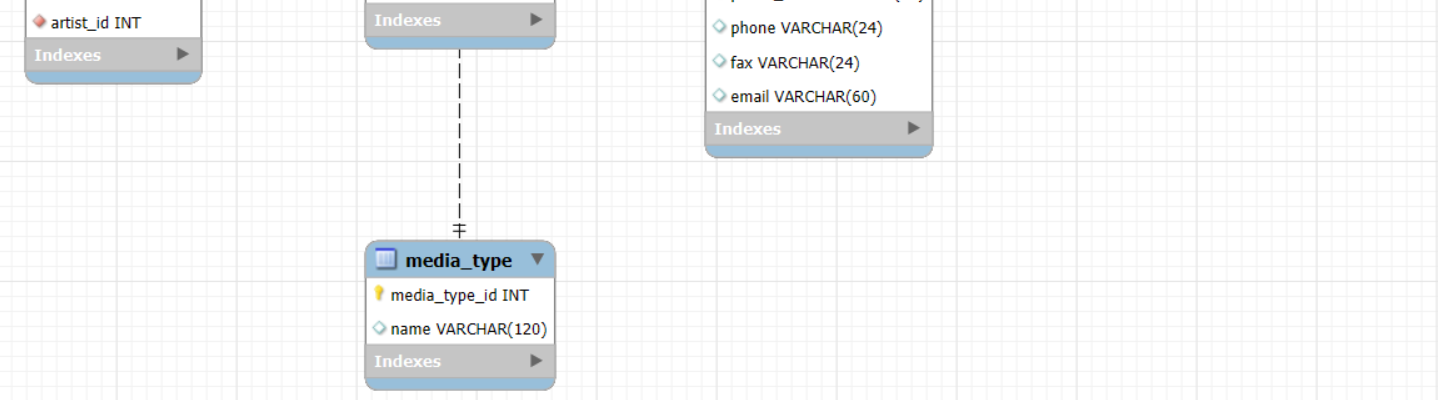
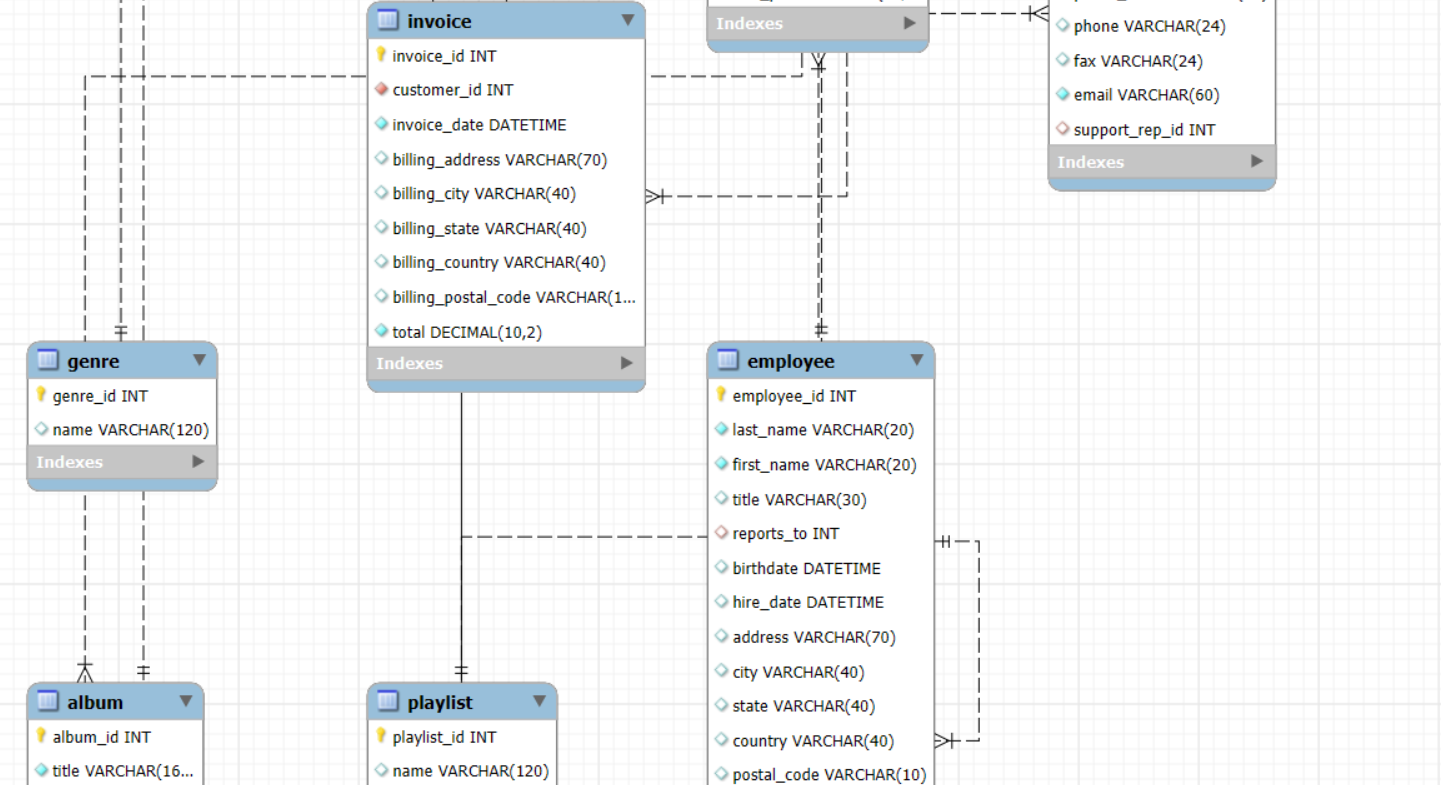
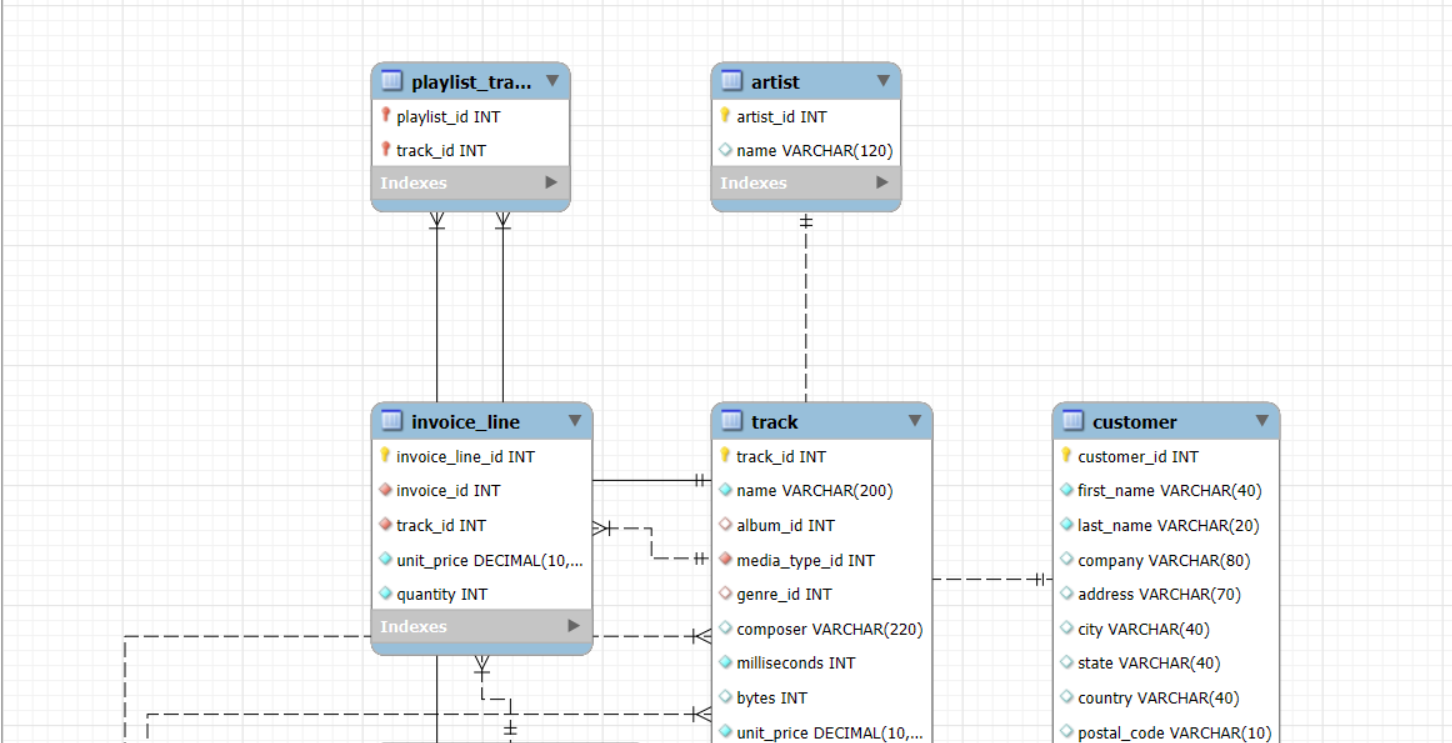
ADD CONSTRAINT `fk\_playlist\_track\_track`

FOREIGN KEY (`track\_id`) REFERENCES `track`(`track\_id`);

ALTER TABLE `employee`

ADD CONSTRAINT `fk\_employee\_reports\_to`

FOREIGN KEY (`reports\_to`) REFERENCES `employee`(`employee\_id`);



Objective Questions

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

**Solution: - *FINDING MISSING VALUES*.**

**Approach: -** By checking for NULL values in specific columns, we can identify, report, and manage missing data within the MySQL table.

* We will use a CASE statement to create a NULL indicator (1 for NULL, 0 for not-NULL) and sum these values to determine the total count of NULLs.

**Query: -** finding missing values for ‘album’ table:

**<Input>** use chinook;

SELECT

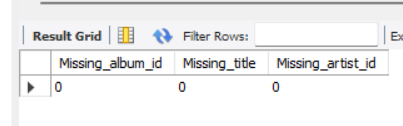
SUM (CASE WHEN album\_id IS NULL THEN 1 ELSE 0 END) AS Missing\_album\_id,

SUM (CASE WHEN title IS NULL THEN 1 ELSE 0 END) AS Missing\_title,

SUM (CASE WHEN artist\_id IS NULL THEN 1 ELSE 0 END) AS Missing\_artist\_id

FROM album;

**<Output>** In this table no missing values found.

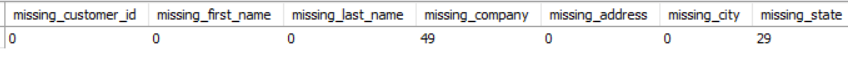


Same we can do for rest of the given tables.

***HANDLING MISSING VALUES*.**

We found some missing values, that table are: - ‘track’, ‘customer’, ‘employee’.

**Output (missing values in ‘customer’ table)**



****

**Approach:** - NULL values can be replaced with a custom string, for instance, 'unknown'.

* To proceed, we must temporarily disable SQL Safe Updates mode and re-enable it upon completion.
* SQL Safe Updates mode is a MySQL setting that requires UPDATE and DELETE operations to be performed with a WHERE clause that uses a key column or a LIMIT clause, so as to limit their impact.

**<Input>**

USE chinook;

SET SQL\_SAFE\_UPDATES = 0;

UPDATE track

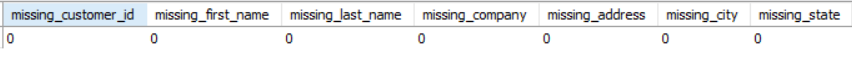
SET composer = 'unknown'

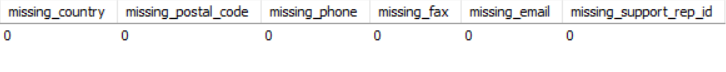
WHERE composer IS NULL;

SET SQL\_SAFE\_UPDATES = 1;

select \* from track

**<Output>**

****

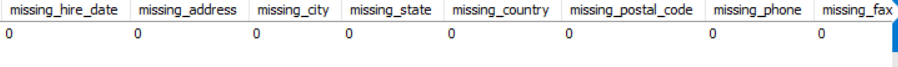
****

Same we can do for rest of the given tables.

**Note: - There are some scenarios where missing values are acceptable, such as in the employee table missing a manager's information in the 'reports\_to' attribute.**

**Example <Output>**

****



* In the Employee table, there is an entry with a NULL in the 'reports\_to' field, but this corresponds to the General Manager, who is the organizational leader and doesn't report to anyone.

***FINDING AND HANDLING DUPLICATES VALUES*.**

**Approach: -** Using a CTE and Window function, we can pinpoint duplicate rows based on unique combinations of values across multiple columns in the MySQL table.

* To focus on duplicates, we'll filter the results based on row numbers that appear more than once.
* No duplicate entries were found in any table.

**Query: -** finding duplicate values for ‘album’ table:

**<Input>** use chinook;

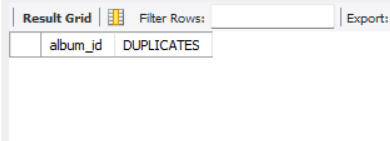
WITH CTE\_album AS (SELECT album\_id,

ROW\_NUMBER() OVER (PARTITION BY title, artist\_id order by album\_id) AS DUPLICATES FROM album)

SELECT \* FROM CTE\_album

WHERE DUPLICATES > 1;

**<Output>**



Same we can do for rest of the given tables.

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

**Solution: -**

**Approach: -** To identify the top-selling tracks, genres, and artists in the US, we can generate a ranked list of the top 10 tracks, incorporating genre and artist information, using CTE and window functions like RANK().

* We can determine the top artists and their dominant genres in the USA by analysing the total sales of their tracks

**Query: -** Top-Selling Tracks Analysis:

**<Input>**

with cte as (select t.name track\_name, ar.name as artist\_name, g.name as genre\_name, sum(il.quantity)as top\_selling\_track from track t

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

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

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

join artist ar on ar.artist\_id = a.artist\_id

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

where i.billing\_country = "USA"

group by t.name, ar.name, g.name ),

cte2 as (select track\_name, artist\_name,genre\_name,top\_selling\_track, rank() over(order by top\_selling\_track desc) as `ranking` from cte)

select track\_name, artist\_name,genre\_name,top\_selling\_track from cte2

where `ranking` <= 10

**Reference:**

**< Output >**

****

* Results:

1. 'War Pigs' - Top-selling track in the USA

2. 'You Know I'm No Good (feat. Ghostface Killah)' - Second-best selling track

3. Rock genre (Alternative Rock) dominates top-selling tracks

**Conclusion:** This analysis highlights the most popular tracks and genres in the US music market, providing valuable insights for music industry stakeholders.

Top-Selling Artists Analysis:

**<Input>** select a.artist\_id, a.name as artist\_name, g.name as genre\_name, sum(il.quantity) as total\_sales FROM invoice AS i

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

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

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

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

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

where i.billing\_country = 'USA'

group by a.artist\_id, a.name, g.name

order by total\_sales desc;

**Reference:** **<Output>**

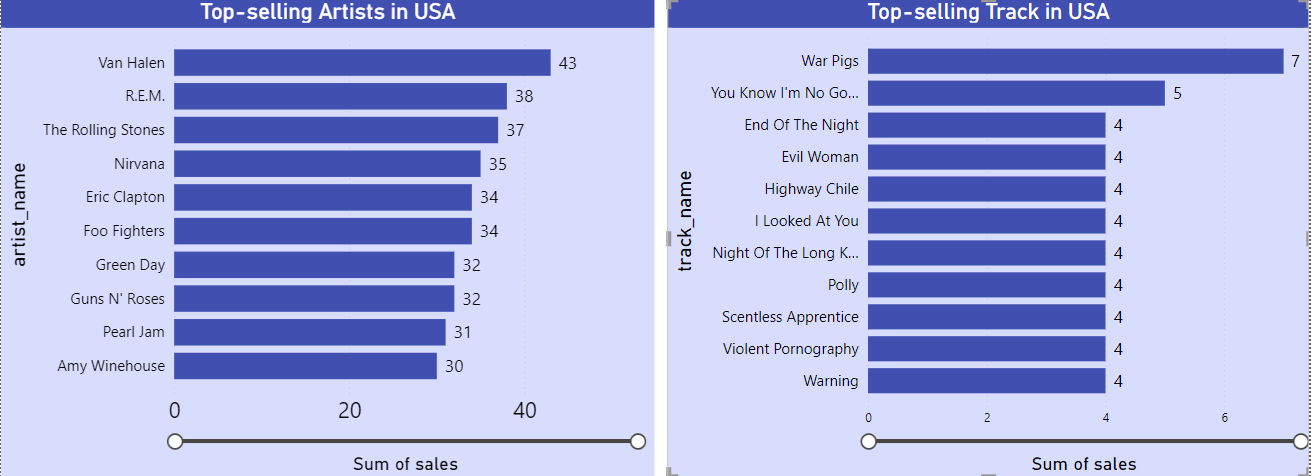


* Findings: (chart insights)

1. Top-selling artist: Van Halen

2. Second-best selling artist: The Rolling Stones

3. Dominant genre: Rock

****

**Conclusion:** This analysis highlights the dominance of Rock music in the USA, with Van Halen and The Rolling Stones emerging as top-selling artists.

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

**Solution: -**

**Approach: -** To understand customer demographics, we'll generate a list of total customers by country, identifying key markets for the music store.

* By calculating customer numbers per country, we can evaluate market reach, pinpoint growth potential, and adapt services to meet local demands.
* **Insights**: - Top 5 Countries by total Chinook Customers

1. USA - 13

2. Canada - 8

3. France & Brazil - 5 (tied)

4. Germany - 4

**Query: -** customers in each country:

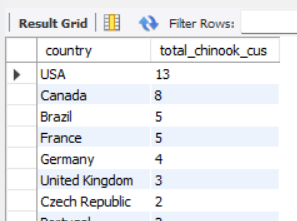
**<Input>** select country, count(customer\_id) as total\_chinook\_cus

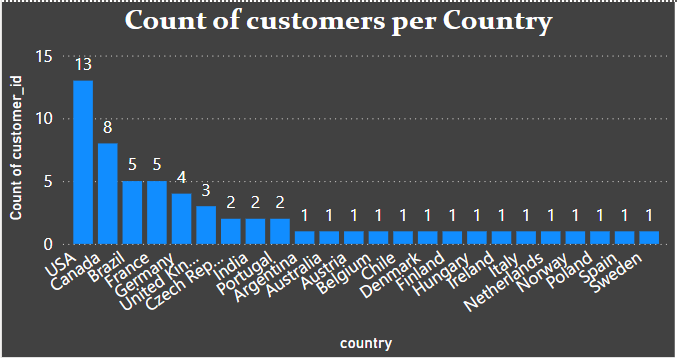
from customer

group by country

order by total\_chinook\_cus desc;

**Reference: <Output>**

****



**Conclusion**: The top five countries with the largest customer bases are the USA, Canada, Brazil, France, and Germany. Maximum chinook customers are available in USA = 13.

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

**Solution: -**

**Approach: -** Using GROUP BY and aggregation, we'll aggregate sales data and invoice counts to derive total revenue and invoice numbers, facilitating analysis by country, state, and city.

* This analysis provides valuable insights into revenue and invoice trends across different geographic locations.

**Query: -**

**<Input>**

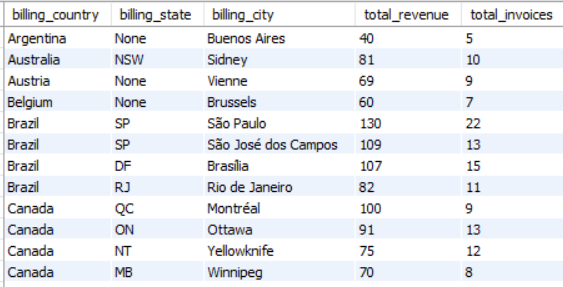
select billing\_country, billing\_state, billing\_city, round(sum(total),0) as total\_revenue, count(invoice\_id) as total\_invoices

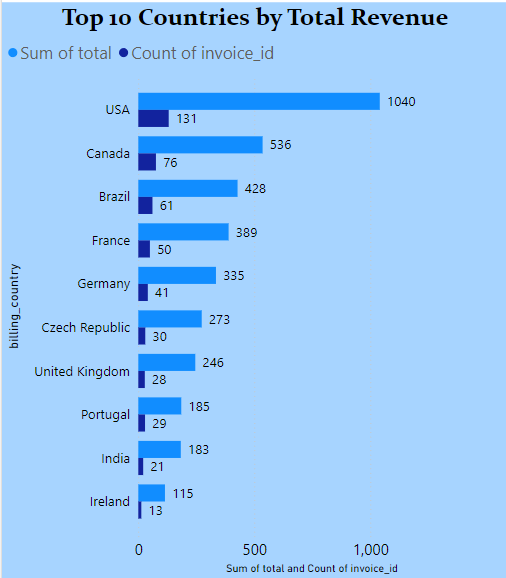
from invoice

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

order by billing\_country, total\_revenue desc;

**Reference: - <Output>**

****

****

* Findings: --Top 5 Countries by Total Revenue:

1. Czech Republic - $273

2. USA - $169

3. United Kingdom - $166

4. Germany - $158

5. France - $151

Top 5 Countries by Number of Invoices:

1. Czech Republic - 30

2. Brazil - 22

3. Germany - 20

4. USA - 20

5. United Kingdom – 1

**Conclusion: -** The Czech Republic emerges as the leading country in terms of total revenue and number of invoices, followed closely by the USA, United Kingdom, Germany, and France.

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

**Solution: -**

**Approach: -** We can leverage window functions, specifically RANK() and PARTITION BY, to identify the top 5 customers by total revenue in each country, achieving a country-wise ranking.

**Query: - <Input>**

with cte as (select c.customer\_id, c.country,

concat(c.first\_name, ' ', c.last\_name) as customer\_name, sum(i.total) as total\_revenue,

rank() over(partition by c.country order by sum(i.total) desc) as `ranking` from customer c

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

group by c.customer\_id, c.country)

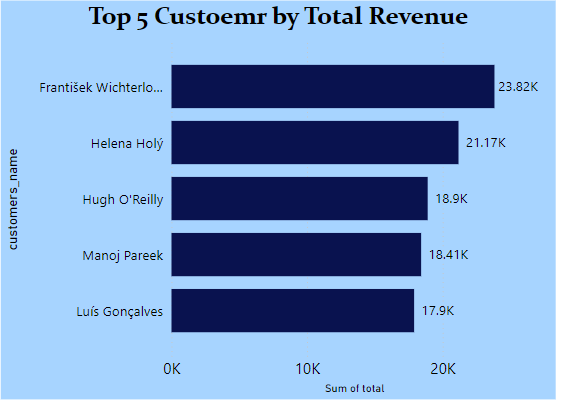
select \* from cte where `ranking` <= 5

**Reference: - <Output>**

****

**Key Insights: -** Top 5 Customers are Frantisek Wichterlova, Helena Holy, Hugh O'Reilly, Manoj Pareek, and Luis Goncalves.

Frantisek Wichterlova Spending High amount as 23.82k.

****

**Conclusion: -** For illustration, Brazil's top 5 customers by total revenue are Luís Gonçalves, Fernanda Ramos, Roberto Almeida, Alexandre Rocha, and Eduardo Martins, demonstrating the country's key revenue contributors. Same as logic implemented in other countries also.

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

**Solution: -**

**Approach: -** To determine top-selling tracks by customer, we'll calculate total sales per customer and then employ ROW\_NUMBER window function for track ranking.

**Query: - <Input>**

with cte3 as

(SELECT c.customer\_id, CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name, t.name as top\_selling\_track, sum(il.quantity) AS total\_sales

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.name ),

cte4 as (select \*, row\_number() over(partition by customer\_id order by total\_sales desc) as ranked

from cte3 )

select customer\_id, customer\_name, top\_selling\_track, total\_sales from cte4

where ranked = 1

order by customer\_id;

**Reference: - <Output>**

****

**Conclusion: -** In conclusion, our analysis has successfully identified the top-selling tracks, including 'Put Your Lights On', 'Cochise', 'Sting Me', 'Welcome to the Jungle', and others, which will enable us to develop targeted marketing strategies, enhance customer engagement, and drive sales growth.

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

**Solution: -**

**Approach: -** To analyse customer purchasing behaviour, we'll write a query to calculate purchase frequency (invoice count), average order value, and segment results by country, city, and customer.

**Query: - <Input>**

with cte\_obj7 as (select c.customer\_id as customer\_id, c.country as country, c.city as city, count(i.invoice\_id) as frequency from customer c

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

group by c.customer\_id, c.country, c.city

order by c.country, c.city, frequency desc), /\*frequency of purchases\*/

cte\_obj7\_2 as (select c.customer\_id as customer\_id, c.country as country, c.city as city, round(avg(i.total),2) as average\_order\_value from customer c

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

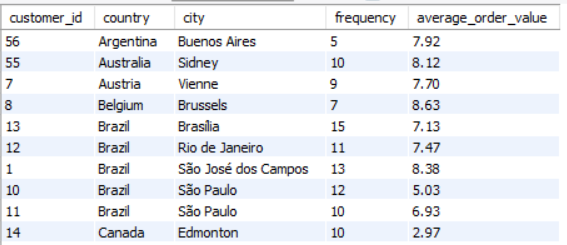
group by c.customer\_id, c.country, c.city

order by c.country, c.city, average\_order\_value desc) /\*average order value\*/

select a1.customer\_id, a1.country , a1.city, a1.frequency, a2.average\_order\_value from cte\_obj7 as a1

join cte\_obj7\_2 as a2 on a1.customer\_id = a2.customer\_id;

**Reference: - <Output>**

****

/\* **relation between three variables frequency, average\_order\_value, and count of customers per country** \*/

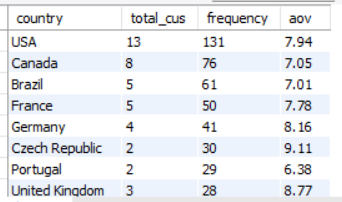
**<Input>**

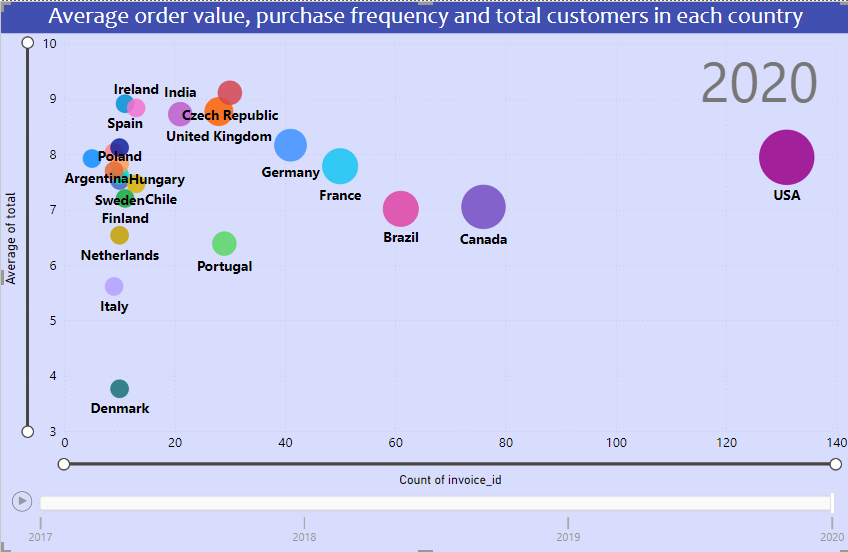
select billing\_country as country, count(distinct customer\_id) as total\_cus, count(invoice\_id) as frequency, round(avg(total), 2) as aov from invoice

group by billing\_country

order by frequency desc , aov desc

**<Output>**

****

****

**Key Insights:** - This provides insights into customer behavior across various regions and can help identify trends, potential opportunities, and areas for improvement.

*Geographic Distribution:*

* Brazil: Here customers buy frequently(15,13,12), but spending varies. São Paulo's low average order value ($5.03) requires attention and can allow product offerings or pricing strategy.
* Canada: Montréal's AOV is $11.11, Edmonton's is $2.97. Edmonton's low spend despite frequent buys suggests price sensitivity.
* USA: Varied purchase frequencies and spend. Mountain View stands out with high frequency (12) and AOV ($9.28), indicating a strong customer base.

*Underperformers:*

* Brazil (São Paulo) and Canada (Edmonton).

*Top Performers:*

* Czech Republic(Prague): it boasts high sales frequency and AOV, a sign of strong market presence.
* Canada(Montréal) shines with high average order value, reflecting effective market strategies.

**Conclusion: -** Targeted marketing is needed in regions with low spend but regular purchases, such as São Paulo and Edmonton, and reward frequent buyers to boost spend.

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

**Solution: -** Customer churn rate is the percentage of customers who stop using a company's product or service during a specific period.

**Approach: -** To analyse customer retention, I applied a **DATE\_SUB** function to identify customers with no purchases in the past 9 months, calculating a churn percentage to inform retention strategies.

**Query: - <Input>**

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

FROM invoice i

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

GROUP BY c.customer\_id),

total\_customers AS (SELECT COUNT(DISTINCT customer\_id) AS total\_customers

FROM customer),

churned AS (SELECT COUNT(DISTINCT customer\_id) AS churned

FROM last\_purchase\_date

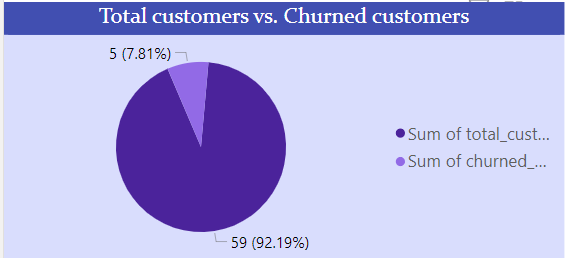
WHERE last\_purchase < DATE\_SUB('2020-12-31', INTERVAL 9 month))

SELECT t.total\_customers, c.churned AS churned\_customer, ROUND((c.churned / t.total\_customers) \* 100, 2) AS churned\_percentage

FROM total\_customers t, churned c;

**Reference: - <Output>**





**Key Insights:** -

* Total customers are 59 from the beginning.
* At the end churned total 5 customers are churned means have not made any purchase from last 9 months
* Churned percentage found 8.47.

**Conclusion: -** The data shows that while customer retention improved in 2018, the increase in customers in 2020 suggests that successful re-engagement and acquisition strategies, likely driven by marketing initiatives and service improvements, were effective.

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

**Solution: -**

**Approach: -** First, we'll break down US sales percentage by each genre, then identify top genres and artists based on sales percentages.

**Query: -** Analyzing the total sales percentage distributed by each genre.

**<Input>** with TotalSales as (select sum(il.quantity) as total\_sale from invoice\_line il

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

where i.billing\_country = 'USA'),

GenreSales as (select g.name as genre\_name, sum(il.quantity) as total\_sale from invoice\_line il

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

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

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

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

join artist ar on a.artist\_id = ar.artist\_id

where i.billing\_country = 'USA'

group by g.name)

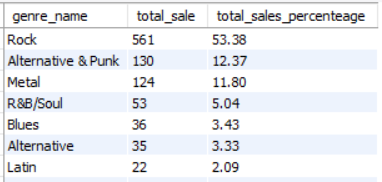
select gs.genre\_name, gs.total\_sale,

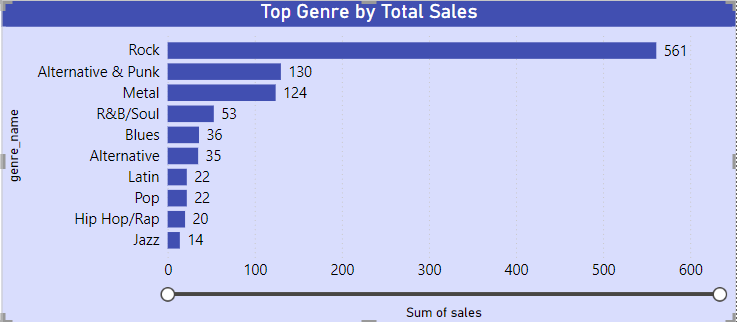
round((gs.total\_sale\*100/ts.total\_sale),2) as total\_sales\_percenteage

from GenreSales gs , TotalSales ts

order by ts.total\_sale desc, total\_sales\_percenteage desc;

**Reference: - <Output>**

****

****

* Calculating the sales percentage for each genre and artist to identify the top genre and the top artist.

**<Input>** with TotalSales as (select sum(il.quantity) as total\_sale from invoice\_line il

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

where i.billing\_country = 'USA'),

GenreSales as (select g.name as genre\_name, ar.name as artist\_name, sum(il.quantity) as total\_sale from invoice\_line il

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

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

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

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

join artist ar on a.artist\_id = ar.artist\_id

where i.billing\_country = 'USA'

group by g.name, ar.name)

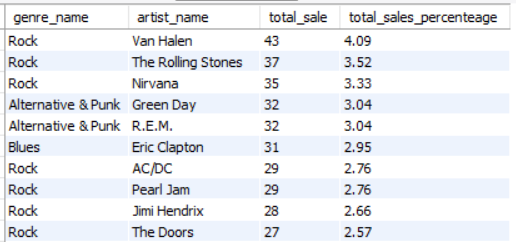
select gs.genre\_name, artist\_name, gs.total\_sale,

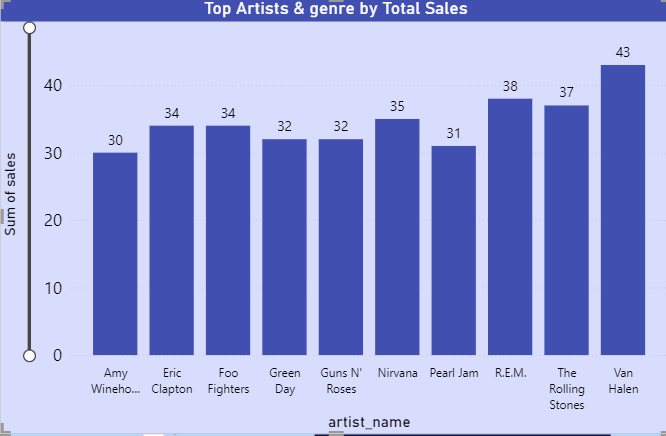
round((gs.total\_sale\*100/ts.total\_sale),2) as total\_sales\_percenteage

from GenreSales gs , TotalSales ts

order by ts.total\_sale desc, total\_sales\_percenteage desc;

**Reference: - <Output>**

****

****

**Conclusion: -** Top genres are ‘Rock’, ‘Alternative& Punk’ , ‘Blues’ and so on Same as top artists are ‘Van Halen’, ‘The Rolling Stones’, ‘Nirvana’, ‘Green Day’, ‘R.E.M.’, ‘AC/DC’ and so on.

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

**Solution: -**

**Approach: -** To determine customers with diverse musical tastes, we utilized an aggregate function to count distinct genre IDs and applied a HAVING clause filter, isolating those with purchases in at least three genres.

**Query: - <Input>**

select c.customer\_id, concat(c.first\_name, " ", c.last\_name) as cus\_name, count(distinct(g.genre\_id)) as total\_genre from track t

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

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

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

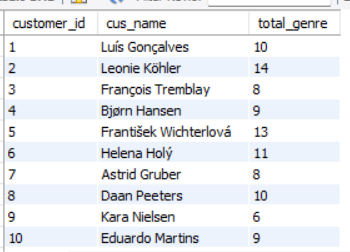
join customer c on c.customer\_id = i.customer\_id

group by c.customer\_id, cus\_name

having count(g.genre\_id) >=3

order by customer\_id asc;

**Reference: - <Output>**

****

**Conclusion: -** The outcome revealed a list of customers characterized by diverse musical preferences, having purchased tracks from at least three distinct genres. These customers are ‘Luís Gonçalves’, ‘Leonie Köhler’, ‘François Tremblay’, ‘Bjørn Hansen’ and so on.

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

**Solution: -**

**Approach: -** To evaluate and rank genres based on their sales performance in the USA, we will utilize the RANK() window function, applying a filter by using having clause to isolate data from the United States.

**Query: - <Input>**

select g.name as Genre, sum(il.quantity) as Total\_Sales,

rank() over(order by sum(il.quantity) desc) as ranking

from track t

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

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

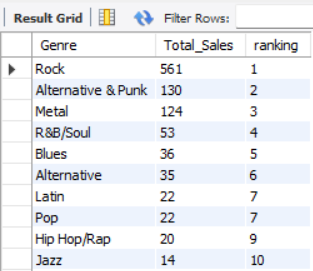
join invoice i on il.invoice\_id = i.invoice\_id

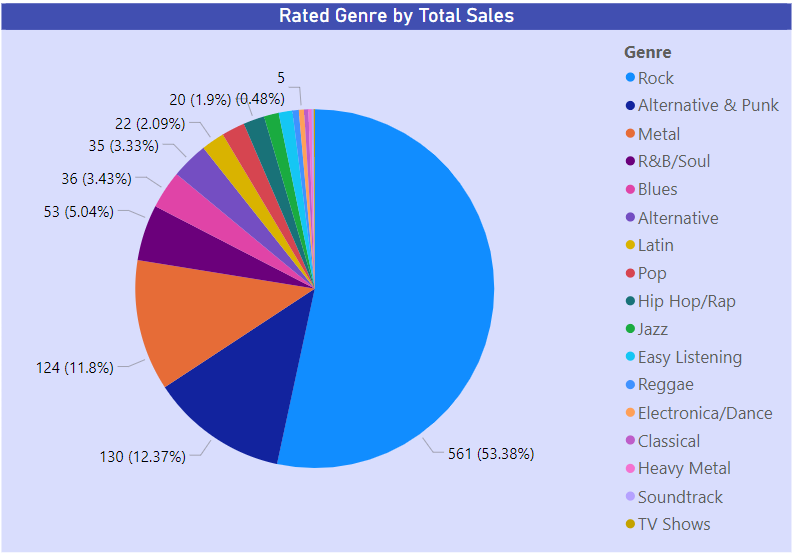
where i.billing\_country = 'USA'

group by g.name

order by ranking

**Reference: - <Output>**

****

****

**Conclusion: -** The outcome revealed a list of top-ranked genres in this list. 'Rock' ranked by 1st means ‘Rock’ is the top most genre after that 'Alternative & Punk', 'Metal', 'R&B/Soul', Blues and so on. ‘Latin’ and ‘Pop’ are the same ranked as 7th .

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

**Solution: -**

**Approach: -** To identify inactive customers, we will employ the DATE\_SUB function to determine customers who have not made a purchase within the last 3 months, retrieving their customer ID, customer name, and most recent purchase date.

**Query: - <Input>**

SELECT c.customer\_id, CONCAT(c.first\_name, " ", c.last\_name) AS customer\_name,

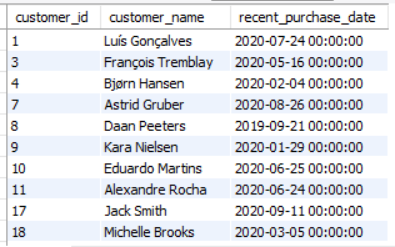
MAX(i.invoice\_date) AS recent\_purchase\_date FROM customer c

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

GROUP BY c.customer\_id, customer\_name

HAVING recent\_purchase\_date < DATE\_SUB('2020-12-31', INTERVAL 3 MONTH);

**Reference: - <Output>**

****

**Conclusion: -** these are the customers who have not made a purchase in the last 3 months for example ‘Luís Gonçalves’, ‘François Tremblay’, ‘Bjørn Hansen’, ‘Astrid Gruber’ and so on.

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.**

**Solution: -**

**Approach: -** To answer this question, we can create list of the albums sold in the USA and sort on the basis of their total sales revenue by genre.

**Query: - <Input>**

select a.album\_id, a.title as album\_name, g.name as genre\_name, sum(il.quantity) as total\_sales from track t

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

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

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

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

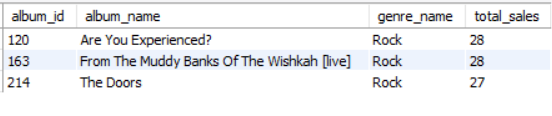
where i.billing\_country = 'USA'

group by a.album\_id, g.name

order by total\_sales desc

limit 3;

**Reference: - <Output>**

****

**Conclusion: -** suggesting these 3 albums should be prioritised for advertising and promotion in the USA based on genre sales analysis (‘Are You Experienced?’, ‘From The Muddy Banks Of The Wishkah [live]’, and ‘The Doors’)

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

**Solution: -**

**Approach: -** Find top 3 genres in non-US countries by sales, then identify common genres across countries.

**Query: -** Find top 3 genres in non-US countries by sales

**<Input>**

with country\_sales as (select c.country, g.name as genre , sum(il.unit\_price \* il.quantity) as total\_sales from track t

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

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

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

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

where c.country <> 'USA'

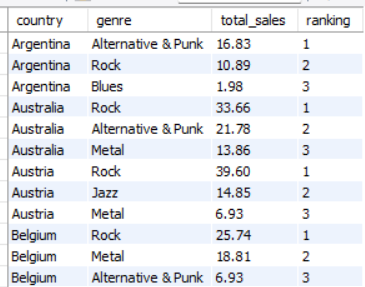
group by c.country, g.genre\_id),

country\_rank as (select \*, row\_number() over(partition by country order by total\_sales desc) as `ranking` from country\_sales)

select \* from country\_rank

where `ranking` <= 3

**Reference: - <Output>**

****

**Query: -** then identify common genres across countries.

**<Input>**

WITH country\_sales AS (SELECT c.country as country, g.name AS genre, SUM(il.unit\_price \* il.quantity) AS total\_sales FROM track t

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

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

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

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

WHERE c.country <> 'USA'

GROUP BY c.country, g.genre\_id)

country\_rank AS (SELECT country, genre, total\_sales,

ROW\_NUMBER() OVER(PARTITION BY country ORDER BY total\_sales DESC) AS ranking

FROM country\_sales),

top\_genres AS (SELECT genre, COUNT(DISTINCT country) as country\_count FROM country\_rank WHERE ranking <= 3 group by genre)

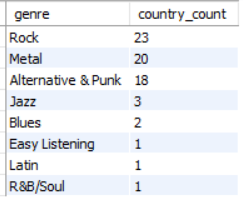
SELECT \*

FROM top\_genres

GROUP BY genre

ORDER BY country\_count DESC;

**Reference: - <Output>**

****

**Key Insights: -**

* Globally, Rock tops genre sales, led by Canada ($329.67M), Brazil ($202.95M), and France ($208.89M).
* Alternative & Punk lags Rock but excels in Argentina, Australia, and Brazil, typically placing second in sales worldwide.
* Metal has a strong global presence, notably in Canada ($71.28M), Brazil ($72.27M), and Germany ($43.56M), often ranking second or third in sales.

**Conclusion: -** In conclusion, Rock music's global dominance presents a prime opportunity for marketing and investment, especially in Canada and Brazil. Meanwhile, Alternative & Punk's growth potential in regions like Argentina opens up new markets. Even niche genres like Jazz and Blues offer targeted marketing possibilities. To succeed, focus on Rock's broad appeal while exploring emerging markets and unique audiences.

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?**

**Solution: -**

**Approach: -** To analyze Customer Purchasing Behavior, we need to calculate the average purchase frequency, total number of customers, average total spending amount, and average basket size.

* To calculate basket size have to sum of sales from invoice line table.

**Query: - <Input>**

WITH CustomerPurchases AS (SELECT c.customer\_id, COUNT(i.invoice\_id) AS frequency, SUM(i.total) AS total\_spent, COUNT(i.invoice\_id) AS num\_invoices,

SUM(il.quantity \* il.unit\_price) AS basket\_size, MIN(i.invoice\_date) AS first\_purchase\_date, max(i.invoice\_date) 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),

CustomerSegments AS (SELECT cp.customer\_id, cp.frequency, cp.total\_spent, cp.basket\_size,

CASE WHEN year(cp.first\_purchase\_date) >= '2020' THEN 'New Customer'

ELSE 'Long-term Customer'

END AS customer\_type

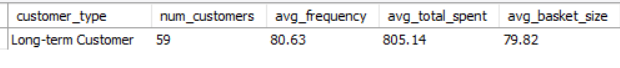
FROM CustomerPurchases cp)

SELECT customer\_type, COUNT(customer\_id) AS num\_customers, round(AVG(frequency),2) AS avg\_frequency, round(AVG(total\_spent),2) AS avg\_total\_spent,

round(AVG(basket\_size),2) AS avg\_basket\_size FROM CustomerSegments

GROUP BY customer\_type;

**Reference: - <Output>**

****

**Key Insights: -**

* Long-term customers buy differently because they trust and love our brand. To keep them loyal, we must deliver quality consistently, serve them exceptionally, and make experiences personal.
* Reward long-term customers with loyalty programs and exclusive offers to strengthen relationships, boost spending, and frequent purchases.
* Since long-term customers have a higher basket size, leverage this insight to implement upselling strategies during the purchasing process for both segments, focusing on complementary products that enhance their experience.
* Make new customers feel valued with a comprehensive onboarding program:
  + Introductory discounts,
  + Tailored advice,
  + Product education.

**Conclusion: -** Convert new customers into loyal buyers by connecting with them through personalized marketing, memorable onboarding experiences, and regular engagement, fostering a lifelong relationship.

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?**

**Solution: -**

**Approach: -** To determine which music genres, artists, or albums are commonly bought together, we can analyze pairs of different items (genres, artists, and music) that customers frequently purchase in conjunction. By examining how often specific combinations occur, we can identify popular groupings.

**Query: - <Input>**

with Affinity\_Analysis as (select i.customer\_id, il.invoice\_id, il.track\_id, t.genre\_id, t.album\_id, a.title as album\_name, ar.name as artist\_name , g.name as genre\_name from track t

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

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

join artist ar on ar.artist\_id = a.artist\_id

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

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

select a1.artist\_name, a1.genre\_name, a1.album\_name, a2.artist\_name, a2.genre\_name, a2.album\_name, count(\*) as frequency from Affinity\_Analysis a1

join Affinity\_Analysis a2 on a1.customer\_id = a2.customer\_id and a1.track\_id < a2.track\_id

where a1.artist\_name <> a2.artist\_name or

a1.genre\_name <> a2.genre\_name or

a1.album\_name <> a2.album\_name

group by a1.artist\_name, a1.genre\_name, a1.album\_name, a2.artist\_name, a2.genre\_name, a2.album\_name

order by frequency desc;

**Reference: - <Output>**

****

**Key Insights: -** Analyzing frequently paired purchases enables the music store to refine product recommendations and cross-selling strategies, boosting sales and customer satisfaction.

* These insights allow us to craft promotions, recommendations, and product bundles that perfectly align with customers' purchasing patterns, driving engagement and sales.
* Customers purchasing Jimi Hendrix's 'Are You Experienced?' album may also appreciate Queen's 'Greatest Hits I', based on their Rock music tastes.
* Cross-Selling: we could also Bundle these two albums or offer a discount on the second when the first is purchased.

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?**

**Solution:**

**Approach: -** To address this question, we need to analyze customer purchasing patterns by calculating total revenue and average order value (AOV) per country, and assess customer churn rates using date subsets to identify customers who have stopped purchasing within the last year.

* also determine the total number of customer statuses per country like whether the customer is a long-term customer, a new customer, or a churned customer so that can easy to analysis the loyalty of the customer.

**Query: - <Input>**

with churned\_cte as (select i.billing\_country, year(i.invoice\_date) as years, max(i.invoice\_date) as last\_purchase,

COUNT(DISTINCT i.customer\_id) AS total\_customers, count(i.invoice\_id) as total\_invoices,

sum(i.total) as total\_revenue, round(avg(i.total),2) as average\_order\_value,

case when max(year(i.invoice\_date)) = 2020 then 'new customer'

else 'long\_term\_customer' end as customer\_type,

case when max(i.invoice\_date) < date\_sub('2020-12-31', interval 1 year) then 'churned customer'

else 'regular customer'

end as customer\_status

from invoice i

group by i.billing\_country, years

order by total\_customers)

select billing\_country as country, years, total\_customers, total\_revenue, average\_order\_value,

count(case when customer\_type = 'long\_term\_customer' then customer\_type end) as total\_long\_term\_customer,

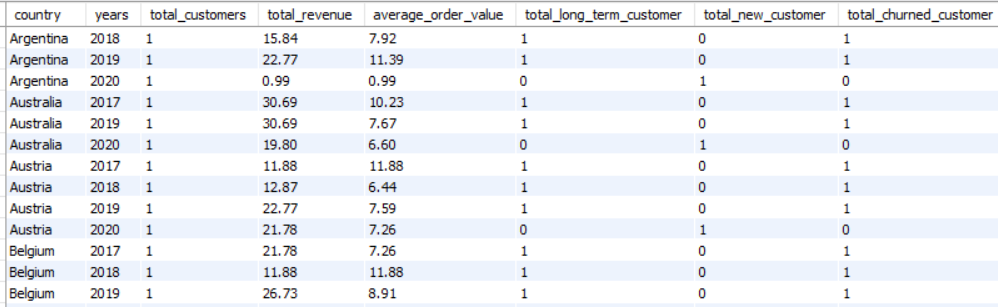
count(case when customer\_type = 'new customer' then customer\_type end) total\_new\_customer,

count(case when customer\_status = 'churned customer' then customer\_type end) total\_churned\_customer

from churned\_cte

group by billing\_country, years, total\_customers, total\_revenue, average\_order\_value;

**Reference: - <Output>**

****

**Key Insights: -** findings are

* Consistently strong revenue performance and customer loyalty in the USA suggest a well-executed marketing approach and deep brand resonance.
* Brazil exhibits high revenue figures in 2017 and 2018 but saw a drop in 2019. However, it rebounded in 2020, suggesting potential market recovery.
* Canada's revenue showed variability, peaking in 2017 but declining thereafter, which may reflect changes in consumer behavior or market competition.
* AOV seems to decrease during downturns in revenue, suggesting that consumers may be opting for lower-priced items during economic challenges.
* Many countries struggle with customer retention, losing one customer for every long-term customer retained each year. whereas Finland and India achieved long-term customer retention until 2020, reflecting successful engagement initiatives.

**Conclusion: -** The analysis highlights the importance of developing effective customer retention strategies, as many markets face high churn rates. Additionally, companies should adapt to economic fluctuations while leveraging successful markets to improve overall revenue and foster long-term customer loyalty.

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?**

**Solution:**

**Approach: -** Segmenting customers by days since last purchase and ranking by average purchase frequency and total spend can provide insights to address this question.Customer can categorize as:

* High risk: - Those customers have not made any purchase from last 1 year. These customers inactive for over a year are at risk of leaving. To keep them, offer personalized deals and targeted reminders to buy again.
* Medium risk: - Customers with a purchase gap between 180-365 days. By monitoring closely and offering timely promotions, can easy to re-engage and retain these customers.
* Low risk: - those customers have not made any purchase in the last 6 month. These customers are engaged, but continued nurturing through great service and relevant offers keeps them loyal.

**Query: - <Input>**

WITH cte AS (SELECT i.customer\_id,i.billing\_country,CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,COUNT(i.invoice\_id) AS frequency, SUM(i.total) AS total\_spent,

DATEDIFF('2020-12-31', MAX(i.invoice\_date)) AS gap\_from\_last\_purchase

FROM invoice i

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

GROUP BY i.customer\_id, i.billing\_country, c.first\_name, c.last\_name),

risk\_cte AS (SELECT \*,

CASE WHEN gap\_from\_last\_purchase > 365 THEN 'High Risk'

WHEN gap\_from\_last\_purchase BETWEEN 180 AND 365 THEN 'Medium Risk'

ELSE 'Low Risk'

END AS risk\_category

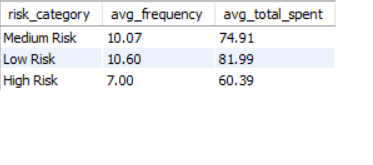
FROM cte)

SELECT risk\_category, AVG(frequency) AS avg\_frequency, AVG(total\_spent) AS avg\_total\_spent

FROM risk\_cte

GROUP BY risk\_category;

**Reference: - <Output>**

****

**Key Insights: -** findings

* average frequency is the same in medium and low-risk customers (10.7).
* high-risk customers have reduced average frequency (7.0) and average total spending as well (60.39).
* medium-risk customers have the highest average total spent (81.99).

**Conclusion: -** High-risk customers lag in engagement and spending, unlike medium and low-risk customers. Tailored marketing and experiences can help convert high-risk customers to consistent buyers, lifting profitability.

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?**

**Solution: -**

**Approach: -** On the basis of tenure, purchase history (total spending and average purchase frequency), and engagement (churned data- which customers have not made any purchase from the last 9 months are considered as churned customers) can categorize customers among loyal, active, and new customers. shows outcome by segment, total number of customers, average of total spending, average of purchase frequency, and total number of churned customers.

**Query: - <Input>**

WITH CustomerLifetimeValue AS (SELECT c.customer\_id, MIN(i.invoice\_date) AS first\_purchase\_date, MAX(i.invoice\_date) AS last\_purchase\_date,

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

DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) AS tenure\_days FROM customer c

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

GROUP BY c.customer\_id ),

CustomerSegments AS (SELECT customer\_id,

CASE WHEN tenure\_days > 1095 THEN 'loyal customers'

WHEN tenure\_days >= 365 AND tenure\_days < 1095 THEN 'moderate customers' ELSE 'new customers'

END AS segment,

purchase\_frequency, total\_spent, tenure\_days FROM CustomerLifetimeValue ),

churned AS ( SELECT i.customer\_id, MIN(i.invoice\_date) AS first\_cus, MAX(i.invoice\_date) AS last\_cus,

CASE WHEN MAX(i.invoice\_date) < '2020-09-30' THEN 'Churned' ELSE 'Active'

END AS customer\_status

FROM invoice i

JOIN CustomerSegments cs ON cs.customer\_id = i.customer\_id

GROUP BY i.customer\_id

ORDER BY i.customer\_id ),

cte AS ( SELECT cs.customer\_id, -- Select the customer\_id from CustomerSegments

cs.segment , cs.purchase\_frequency, cs.total\_spent, cs.tenure\_days, c.first\_cus, c.last\_cus, c.customer\_status FROM CustomerSegments cs

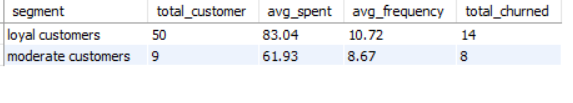
JOIN churned c ON cs.customer\_id = c.customer\_id )

SELECT segment, count(customer\_id) as total\_customer, round(avg(total\_spent),2) as avg\_spent, round(avg(purchase\_frequency),2) as avg\_frequency,

count(case when customer\_status = 'churned' then customer\_status end) as total\_churned FROM cte

group by segment;

**Reference: - <Output>**

****

**Key Insights: -** findings

* Despite having a larger customer base, loyal customers experience higher churn, which may indicate issues with retention even among high-spending customers.
* Loyal Customers have 14 churned and Moderate Customers have 8 churned customers as per 9 months.
* Loyal customers also engage more frequently, indicating a strong relationship with the brand.
* Loyal customers spend significantly more, suggesting that efforts to enhance loyalty could yield higher revenue as $83.04 whereas Moderate Customers $61.93.

**Conclusion: -** By addressing churn and enhancing customer engagement, there is potential for revenue growth and improved customer lifetime value across both segments.

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?**

**Solution:**

**Approach:** - If we were able to obtain that data, we could measure its influence on customer acquisition, retention, and overall sales by linking it to our current customer, invoice, and invoice\_line tables.

Consider a table named 'promotions' with the following structure: -

* promotion\_id: Unique identifier for the promotion
* promotion\_type: Type of promotion (e.g., discount, event, email)
* promotion\_name: Name or description of the promotion
* start\_date: Start date of the promotion
* end\_date: End date of the promotion

Joining this table to customer and invoice tp data analyzes promotion effectiveness.

**Query: - <Input>** SELECT p.promotion\_type, p.promotion\_name,

  COUNT(DISTINCT CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date THEN c.customer\_id ELSE NULL END) AS new\_customers\_acquired,

  COUNT(DISTINCT CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date AND c.customer\_id IN (SELECT customer\_id FROM invoices WHERE invoice\_date < p.start\_date) THEN c.customer\_id ELSE NULL END) AS existing\_customers\_retained,

  SUM(CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date THEN il.unit\_price \* il.quantity ELSE 0 END) AS promotion\_sales

FROM promotions p

LEFT JOIN invoices i ON i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date

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

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

GROUP BY p.promotion\_type, p.promotion\_name;

This query brings together promotions, invoices, customers, and invoice lines to derive metrics and assess promotion performance.

* promotion\_sales: The total sales generated during the promotion period.
* new\_customers\_acquired: The number of new customers acquired during the promotion period.
* existing\_customers\_retained: The number of existing customers who made a purchase during the promotion period.

We can also improve the analysis by adding email marketing and website data to see how promotions perform across different channels.

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

**Solution: -**

**Approach: -** If the objective and subjective questions weren’t given then also, I would refer the same approach.

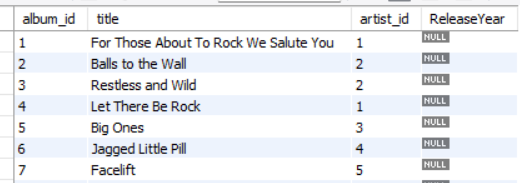
1. **How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?  
   Solution: -**

**Approach: -** Have to use alter function

**Query: - <Input>**

ALTER TABLE album ADD COLUMN ReleaseYear INT;

**Reference: - <Output>**

****

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.**

**Solution: -**

1. **Approach: -** We can analyze the total customer count, the overall amount spent by these customers, and the average number of tracks bought per customer by each country to deliver insights to Chinook.

**Query: - <Input>**

WITH cus\_invoice AS ( SELECT c.country, c.customer\_id, i.total AS invoice\_total, i.invoice\_id FROM customer c

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

track\_count AS (SELECT invoice\_id, COUNT(track\_id) AS track\_count FROM invoice\_line

GROUP BY invoice\_id )

SELECT ci.country AS country, COUNT(DISTINCT ci.customer\_id) AS num\_customers, SUM(ci.invoice\_total) AS total\_amount\_spent,

ROUND(AVG(ci.invoice\_total), 2) AS avg\_amount\_spent\_per\_customer,

ROUND(AVG(it.track\_count), 0) AS avg\_tracks\_purchased\_per\_customer FROM cus\_invoice ci

JOIN track\_count it ON ci.invoice\_id = it.invoice\_id

GROUP BY ci.country

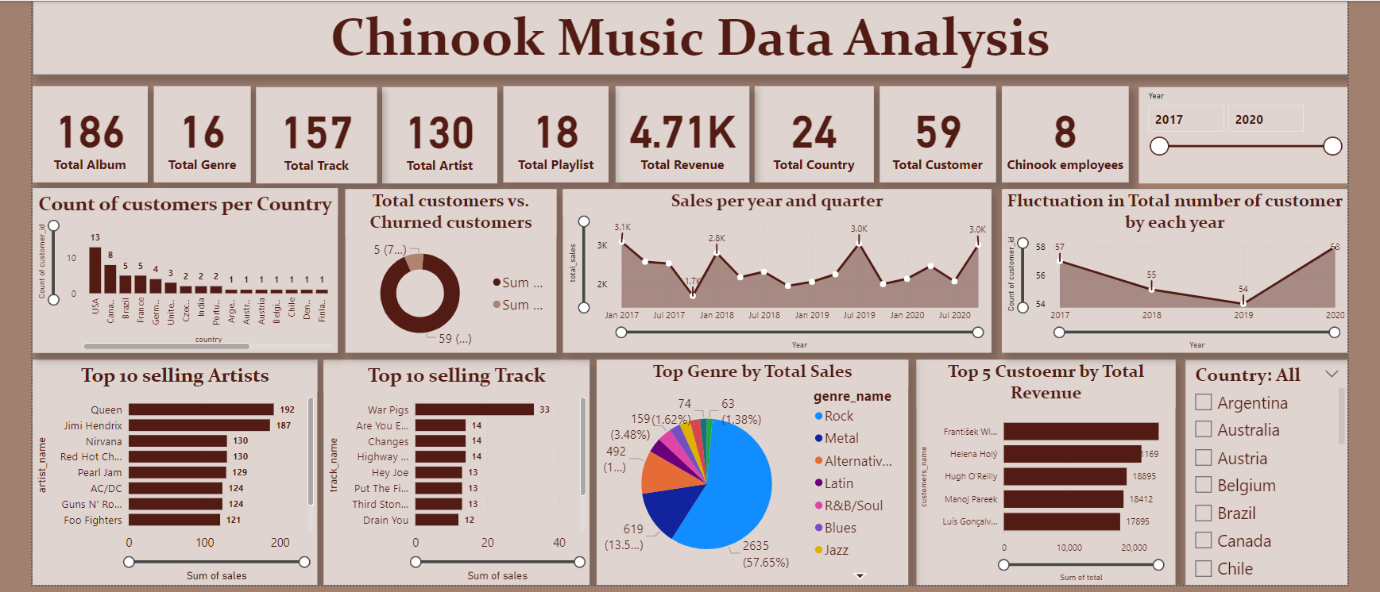
ORDER BY total\_amount\_spent DESC;

**Reference: - <Output>**

****

**Key Insights: -**

* USA is top in number of customers and total amount spent.
* Canada is 2nd in number of customers and total amount spent.
* Canada and Brazil have same average total tracks purchased per customer as 7.
* Brazil and France have same number of customers (5).



# **DASHBOARD**