OBJECTIVE QUESTIONS

***OBJECTIVE QUESTIONS***

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

**Answer :** Yes, some tables in the dataset have missing (NULL) values. To handle them, I first identified the columns containing NULLs by executing SQL queries on each table. Here’s an example query for the customer table:

**select**

**sum(case when customer\_id is null then 1 else 0 end) as column1\_nulls,**

**sum(case when first\_name is null then 1 else 0 end) as column2\_nulls,**

**sum(case when last\_name is null then 1 else 0 end) as column3\_nulls,**

**sum(case when company is null then 1 else 0 end) as column4\_nulls,**

**sum(case when address is null then 1 else 0 end) as column5\_nulls,**

**sum(case when city is null then 1 else 0 end) as column6\_nulls,**

**sum(case when state is null then 1 else 0 end) as column7\_nulls,**

**sum(case when country is null then 1 else 0 end) as column8\_nulls,**

**sum(case when postal\_code is null then 1 else 0 end) as column9\_nulls,**

**sum(case when phone is null then 1 else 0 end) as column10\_nulls,**

**sum(case when fax is null then 1 else 0 end) as column11\_nulls,**

**sum(case when email is null then 1 else 0 end) as column12\_nulls,**

**sum(case when support\_rep\_id is null then 1 else 0 end) as column13\_nulls**

**from customer;**

This query identifies the number of NULL values in each column.

After identifying NULL values, I handled them using the UPDATE statement, replacing missing values with appropriate defaults:

**update customer**

**set last\_name = coalesce(last\_name, 'Not Available'),**

**state = coalesce(state, 'Not Available'),**

**postal\_code = coalesce(postal\_code, 'N/A'),**

**phone = coalesce(phone, 'Not Available'),**

**fax = coalesce(fax, 'Not Available')**

**where last\_name is null or state is null or postal\_code is null**

**or phone is null or fax is null;**

Yes, the dataset contains duplicate values, but they play a significant role in maintaining relationships across tables. However, to **identify duplicates**, the following query is used:  **select first\_name, last\_name, company, city, state, country, postal\_code,**

**phone, fax, email, support\_rep\_id, count(\*) as count**

**from customer**

**group by first\_name, last\_name, company, city, state, country, postal\_code, phone, fax, email, support\_rep\_id**

**having count(\*) > 1;**

This query helps detect duplicate records based on all key attributes.

To **remove duplicates while keeping one record**, the following approach is used:

**with remove\_duplicates as**

**(select customer\_id,**

**row\_number() over (partition by first\_name, last\_name, company, city, state, country, postal\_code, phone, fax, email, support\_rep\_id**

**order by customer\_id) as row\_num**

**from customer)**

**delete from customer**

**where customer\_id in (select customer\_id from remove\_duplicates**

**where row\_num > 1);**

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

**Answer :** To determine the top-selling tracks and top artists in the USA, I first analyzed the invoice data to identify the tracks with the highest sales volume. The invoice and invoice\_line tables contain purchase data, while the track, album, and artist tables provide information about songs and their respective artists.

1. Identifying the Top-Selling Tracks in the USA:
   * The invoice table contains purchase transactions, so I filtered the data to include only records where billing\_country = 'USA'.
   * Using the invoice\_line table, I counted the number of times each track was purchased in the USA.
   * I ranked the tracks based on their total purchase count.
2. Finding the Top-Selling Artist:
   * Since albums contain multiple tracks, I linked each track to its respective album and artist.
   * By aggregating the sales data at the artist level, I identified the most popular artist in the USA.
3. Identifying the Most Famous Genres:
   * Each track belongs to a specific genre, which is stored in the genre table.
   * By grouping the sales data by genre, I determined which genres were the most purchased in the USA.

The following query was used to achieve this analysis :

**with top\_in\_USA as**

**(select l.track\_id, i.billing\_country, t.name as track\_title,**

**a.title as album\_title, art.name as artist\_name, g.name as genre\_name**

**from invoice i**

**join invoice\_line l**

**on i.invoice\_id = l.invoice\_id**

**join track t**

**on l.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**

**join genre g**

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

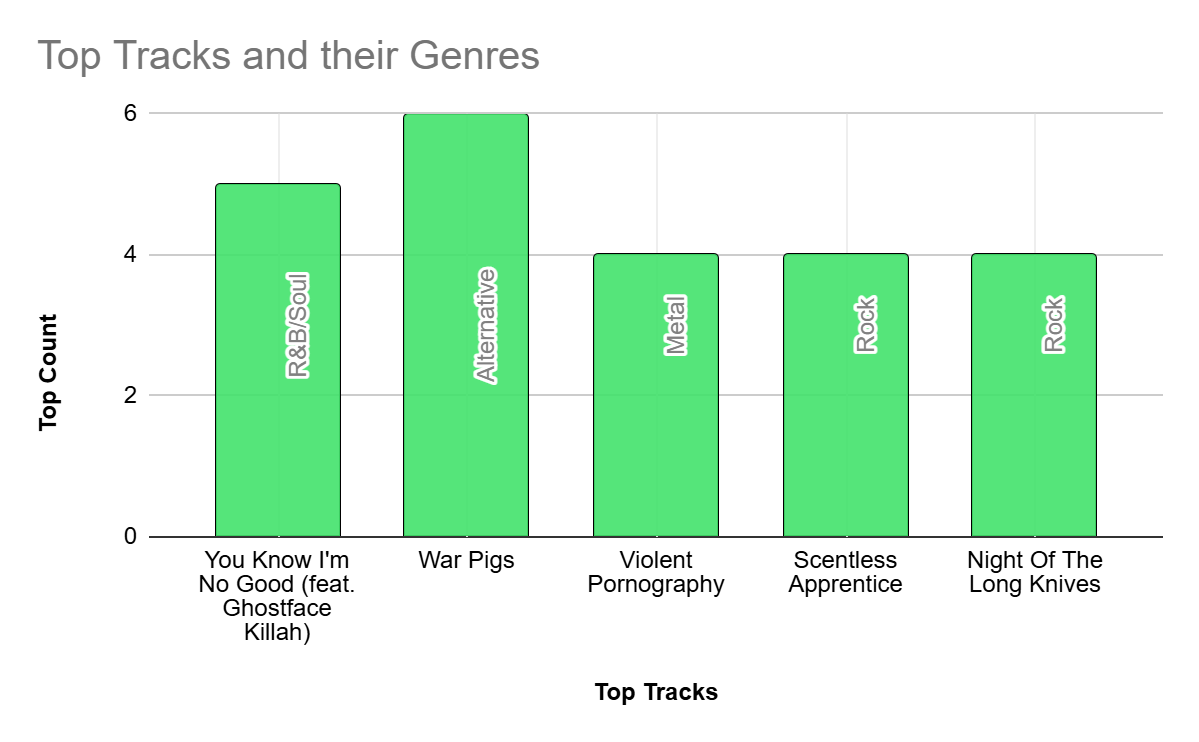
**where billing\_country = 'USA')**

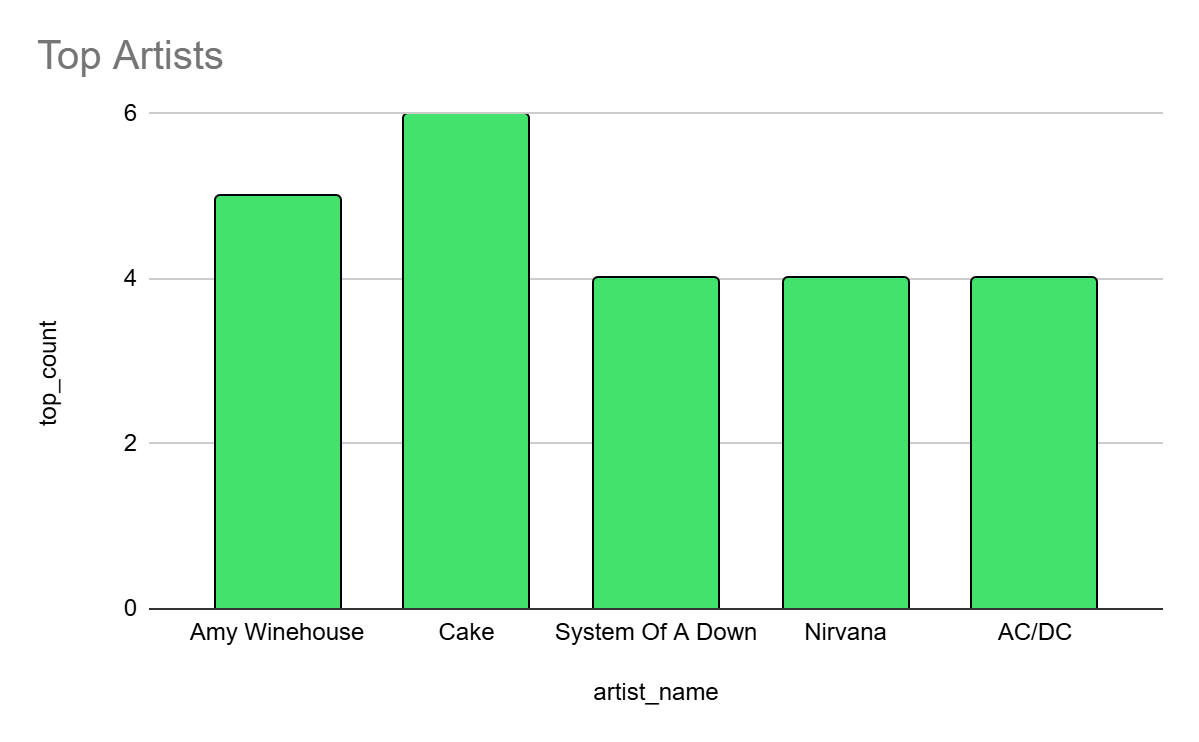
**select track\_id, track\_title, artist\_name, genre\_name, count(\*) as top\_count from top\_in\_USA**

**group by track\_id, track\_title, artist\_name, genre\_name**

**order by top\_count desc**

**limit 5;**

****

****

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

**Answer :** To analyze the demographic breakdown of Chinook’s customer base, I focused on location-based distribution since the dataset does not contain explicit columns for age or gender.

#### 1. Country-wise customer distribution:

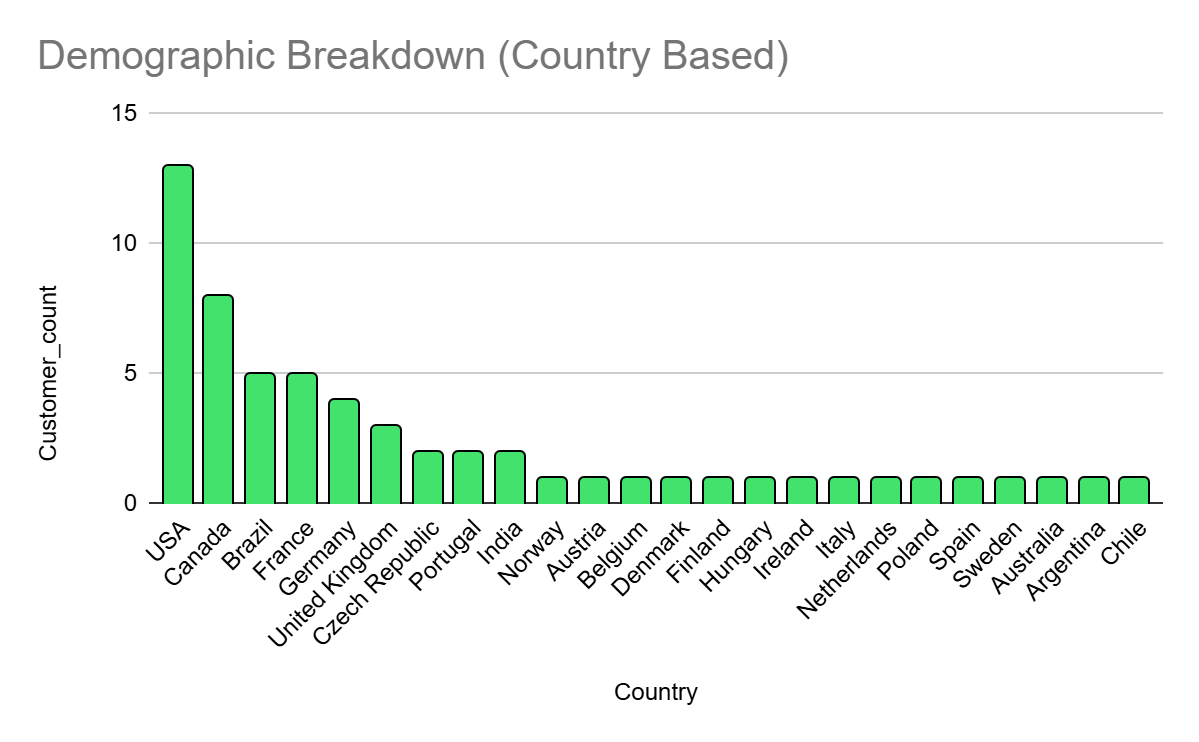
* To get an overview of how customers are distributed across different countries, I used the customer table.
* The query groups customers by country and counts the number of customers in each.
* The results are ordered in descending order of customer count to highlight the countries with the highest number of customers.

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

**from customer**

**group by country**

**order by count(\*) desc**

****

#### 2. Detailed location-wise breakdown (country, state, city):

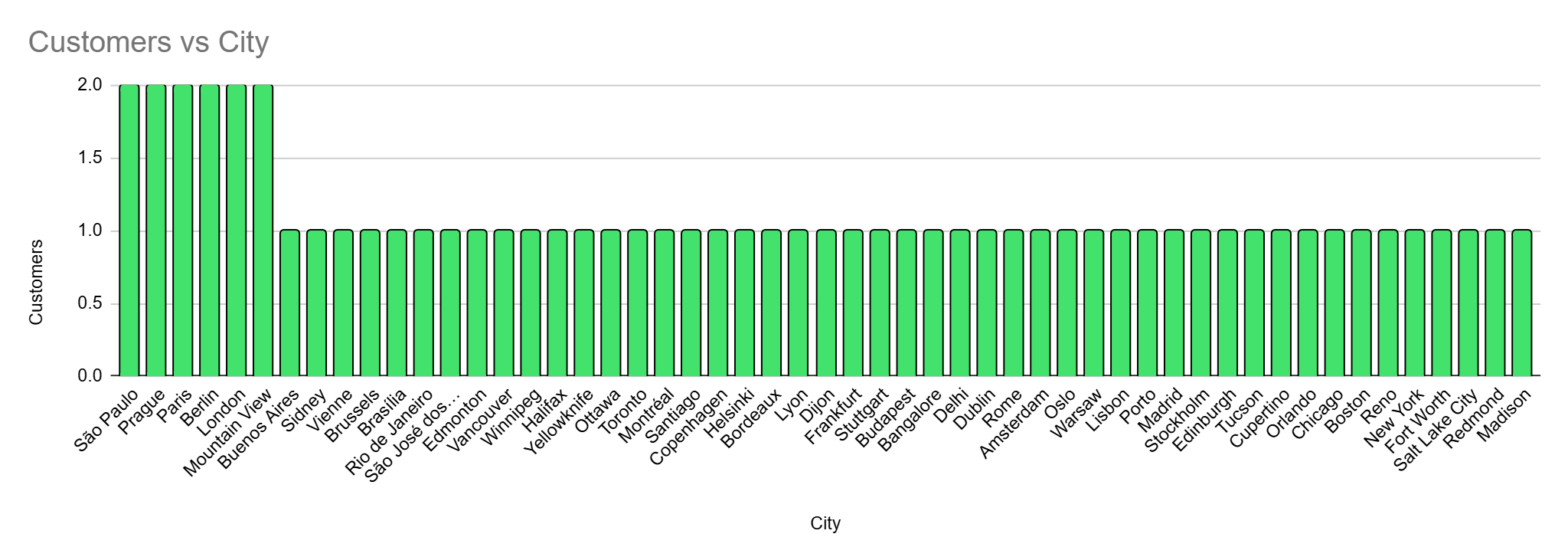
* Since the dataset provides customer location details at multiple levels (country, state, and city), I further grouped the data by these attributes.
* This helped in identifying the concentration of customers within specific regions.
* The results are ordered primarily by customer count and then by country and state for better readability.

**SELECT country, state, city, count(\*) as cust\_count**

**from customer**

**group by country, state, city**

**order by count(\*) desc, country, state;**



#### 3. Missing age and gender data:

* The Chinook database does not contain columns for age or gender, which means customer segmentation based on these attributes is not possible.
* If such data were available, we could analyze age distribution (e.g., grouping by age ranges) and gender-based purchasing behavior (e.g., gender-wise revenue contribution).

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

**Answer :** To analyze the total revenue and invoice count, I used the invoice table and grouped the data at different levels: country, state, and city. This helps in identifying which locations contribute the most to revenue generation.

#### 1. Total revenue and invoices by country:

* This query aggregates total revenue (sum(total)) and invoice count (count(\*)) per country.
* The results are sorted by total revenue and then by invoice count in descending order.

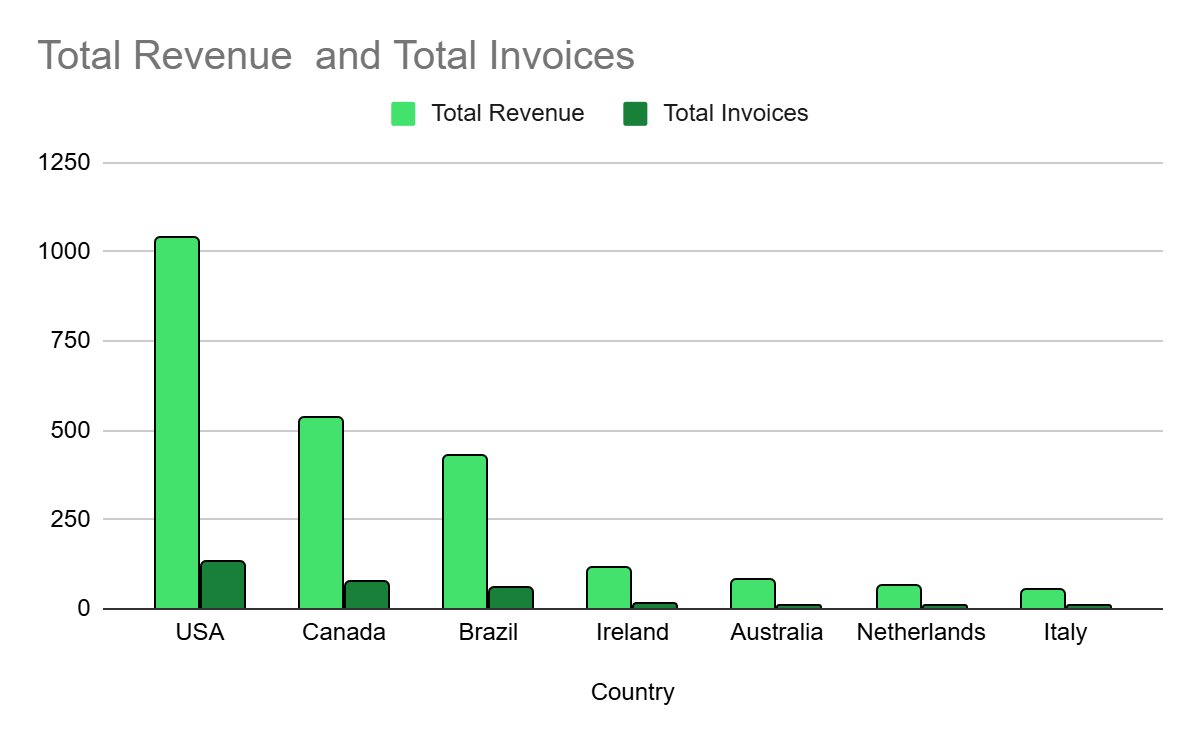
**select billing\_country, sum(total) as total\_revenue,**

**count(\*) as total\_invoice**

**from invoice**

**group by billing\_country**

**order by total\_revenue desc, total\_invoice desc;**

**

#### 2. Total revenue and invoices by state:

* Since some countries have multiple states, this query provides a more detailed breakdown.
* The revenue and invoice count are grouped at the **state level** within each country.

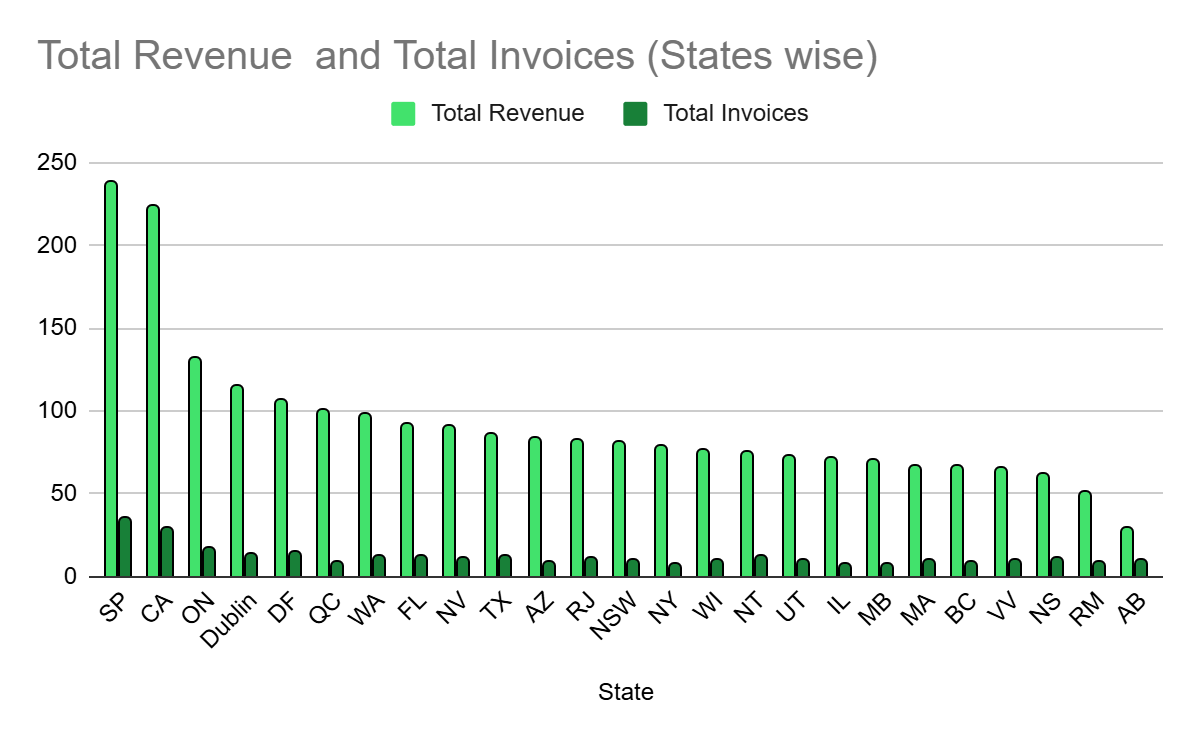
**select billing\_state, sum(total) as total\_revenue,**

**count(\*) as total\_invoice**

**from invoice**

**group by billing\_state**

**order by total\_revenue desc, total\_invoice desc;**



#### 3. Total revenue and invoices by city:

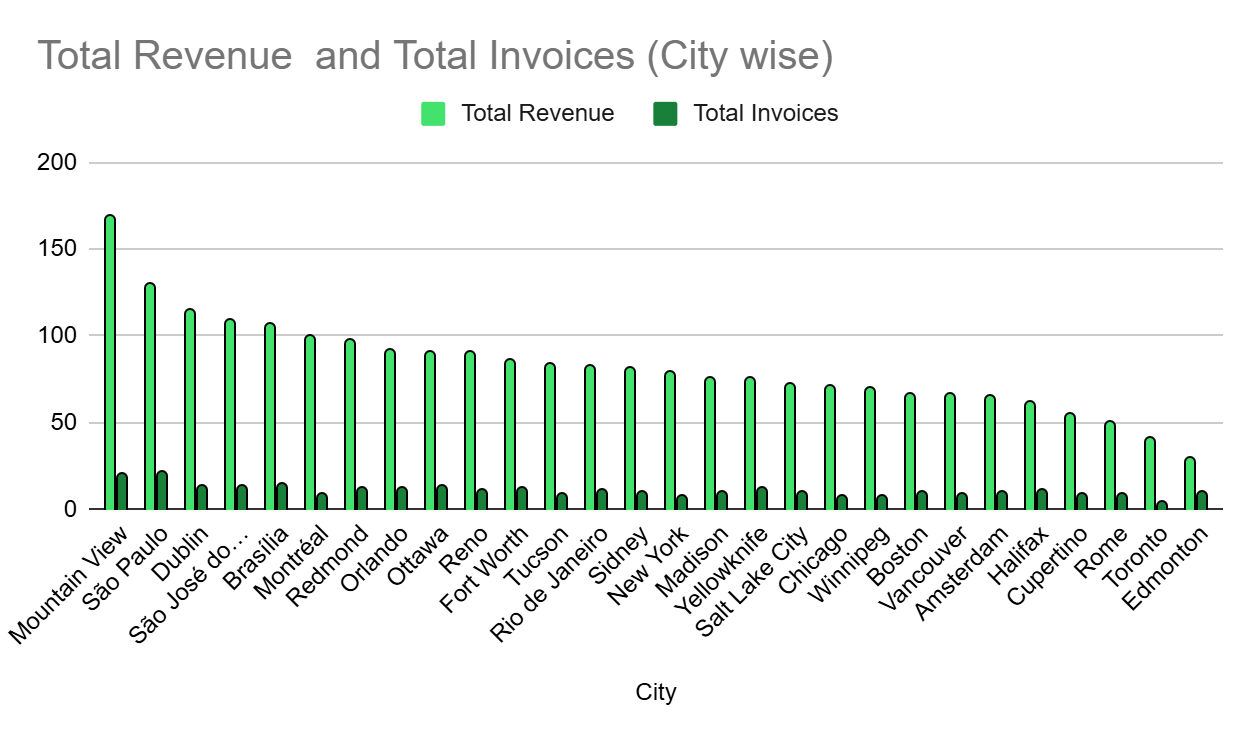
* This query breaks down the revenue further to the city level to analyze which cities generate the highest revenue and invoice count

**select billing\_city, sum(total) as total\_revenue, count(\*) as total\_invoice**

**from invoice**

**group by billing\_city**

**order by total\_revenue desc, total\_invoice desc**



#### 4. Combined revenue and invoices analysis (country, state, city):

* To get a detailed hierarchy, I combined country, state, and city in a single query.
* This allows us to pinpoint the exact locations contributing to revenue growth.

**select billing\_country, billing\_state, billing\_city,**

**sum(total) as total\_revenue, count(\*) as total\_invoice**

**from invoice**

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

**order by total\_revenue desc, total\_invoice desc;**

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

**Answer :** To determine the top 5 customers in each country, we calculated the total revenue generated by each customer and ranked them within their respective countries. The dense\_rank() function was used to assign rankings, ensuring that customers with the same revenue received the same rank. Finally, we filtered the results to include only the top 5 customers per country. The query :

**with top5\_cust as**

**(select i.customer\_id as cust\_id,**

**concat(c.first\_name, ' ', c.last\_name) as name,**

**i.billing\_country as country, sum(total) as total\_revenue,**

**dense\_rank() over (partition by i.billing\_country order by sum(total) desc) as ranking**

**from invoice i**

**join customer c**

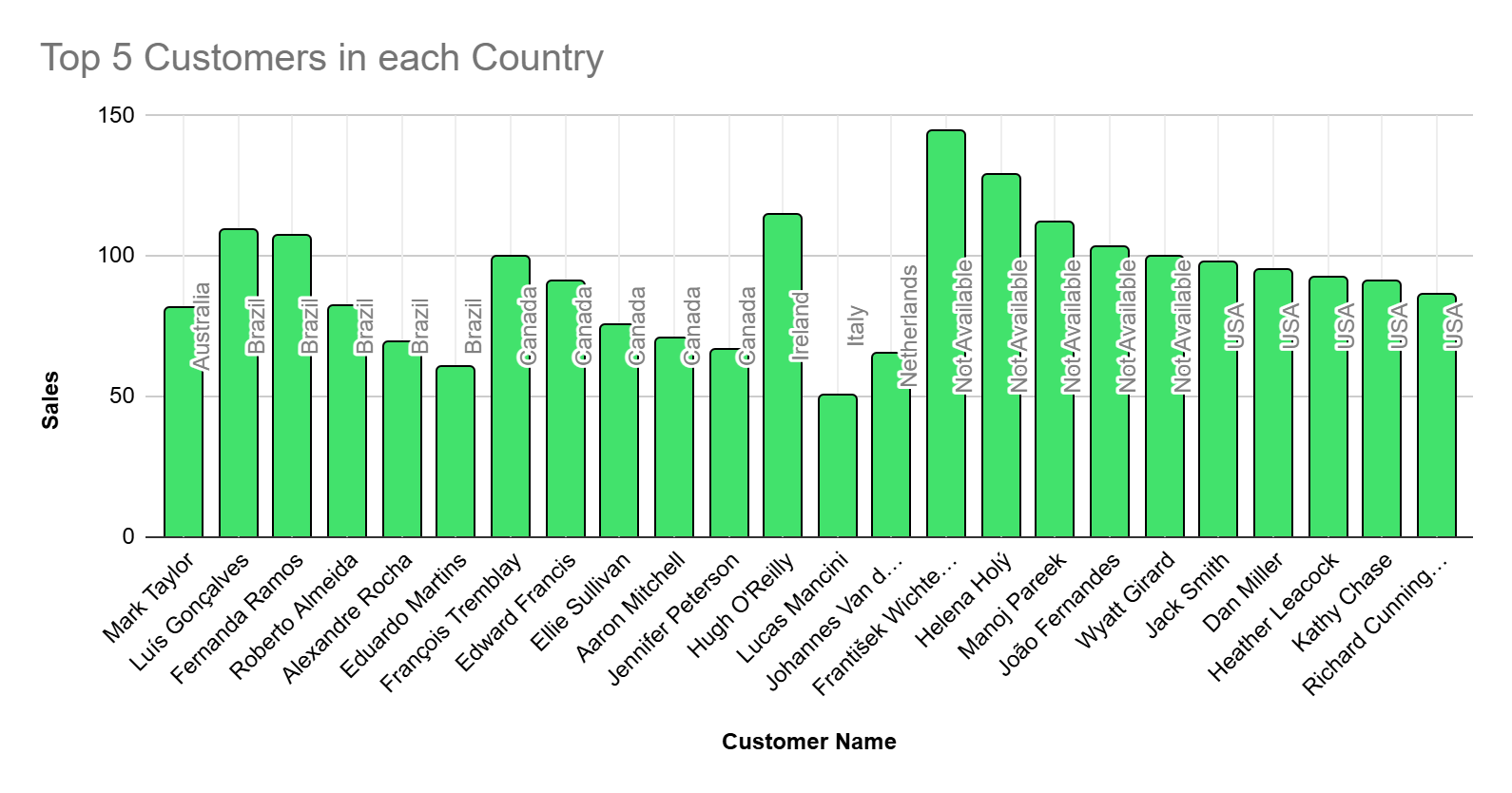
**on i.customer\_id = c.customer\_id**

**group by i.customer\_id, name, I.billing\_country**

**order by country, total\_revenue desc)**

**select cust\_id, name, country, total\_revenue from top5\_cust**

**where ranking <= 5;**

**

This query helps identify the highest-spending customers in each country, which can be useful for targeted marketing and customer retention strategies.

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

**Answer :** To find the most purchased track for each customer, we first counted the number of times each track was bought by a customer. Using the row\_number() function, we ranked the tracks in descending order of purchase count for every customer. Finally, we selected only the top-ranked track for each customer, ensuring that if multiple tracks had the same count, they were ordered alphabetically.

**with tracksales as**

**(select c.customer\_id, concat(c.first\_name,' ', c.last\_name) as name,**

**il.track\_id, t.name as track\_name,**

**count(il.track\_id) as purchase\_count,**

**row\_number() over**

**(partition by c.customer\_id order by count(il.track\_id) desc, t.name asc)**

**as rank\_order**

**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**

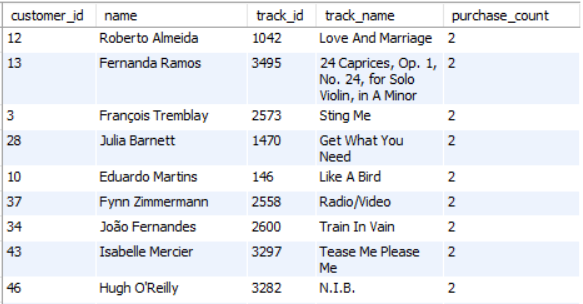
**group by c.customer\_id, c.first\_name, c.last\_name, il.track\_id, t.name)**

**select customer\_id, name, track\_id, track\_name, purchase\_count**

**from tracksales**

**where rank\_order = 1**

**order by purchase\_count desc;**

*Sample Output :*

This query helps in understanding customer preferences by identifying their most frequently purchased track. It can be useful for recommending similar music or creating personalized playlists based on customer interests.

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

**Answer :** To analyze customer purchasing behavior, we examined three key aspects: total orders and average purchase frequency, average order value, and the average time between purchases.

Total Orders and Average Purchase Days per Customer:

* + We counted the total purchases made by each customer.
  + Using the lead() function, we calculated the time gap between consecutive purchases for each customer.
  + The average gap in days was then computed to understand how frequently customers return to buy more music.

Query to achieve the above tasks is :

**with purchase\_gaps as**

**(select customer\_id, invoice\_date,**

**lead(invoice\_date) over (partition by customer\_id order by invoice\_date) as next\_purchase\_date**

**from invoice)**

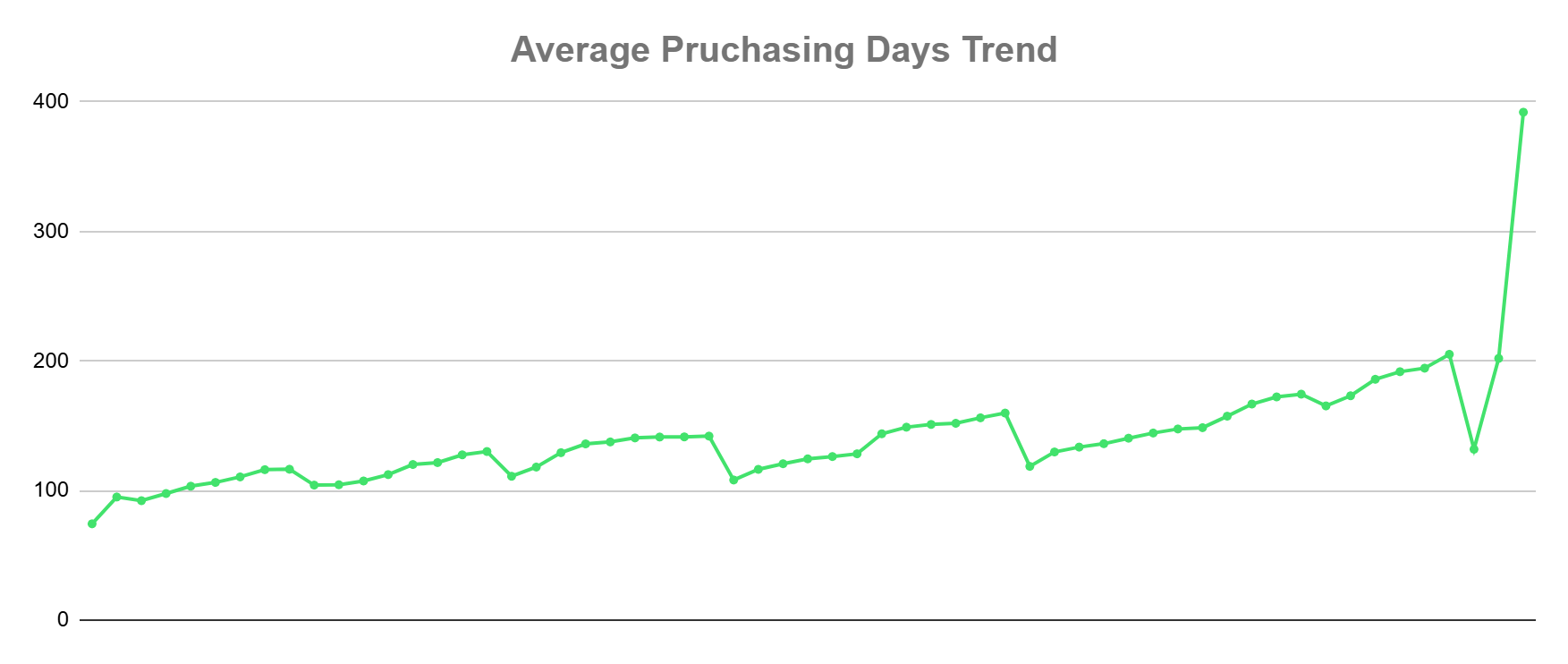
**select customer\_id, count(invoice\_date) as total\_purchases,**

**round(avg(datediff(next\_purchase\_date, invoice\_date)), 2) as avg\_days\_between\_purchase\_**

**from purchase\_gaps**

**group by customer\_id**

**order by total\_purchases desc;**

**

*Average Order Value:*

* We calculated the average amount spent per order (avg\_order\_value).
* Additionally, we summed the total revenue from each customer to identify high-value customers.

**select customer\_id,**

**round(avg(total), 2) as avg\_order\_value,**

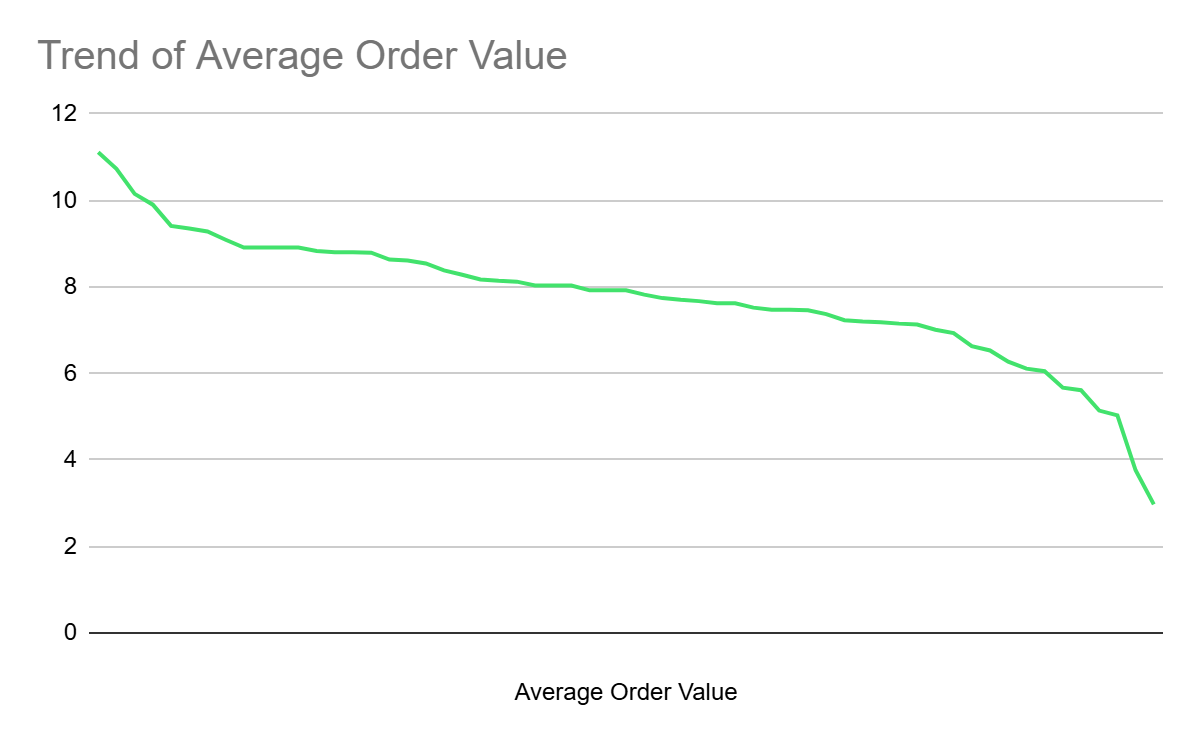
**sum(total) as total\_spent,**

**count(invoice\_id) as total\_orders**

**from invoice**

**group by customer\_id**

**order by avg\_order\_value desc;**

**

We couldn't analyze customer preferences for payment methods because the dataset does not contain any relevant columns related to transaction types, such as credit card, debit card, or digital wallet usage. Without this data, we are unable to determine which payment methods customers prefer or if there are any patterns in their choices.

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

**Answer :** Churn rate helps measure how many customers stop purchasing over time. This query tracks monthly active customers and calculates how many stopped making purchases compared to the previous month.

To analyze churn, we first identify the number of unique customers making purchases each month. We then compare this with the previous month's customer count using the lag() function to determine how many customers did not return. The difference between the two gives us the number of churned customers. Finally, we calculate the churn rate as a percentage of the previous month's active customer base, helping us understand fluctuations in customer retention.

**with monthly\_active\_customers as**

**(select date\_format(str\_to\_date(invoice\_date, '%Y-%m-%d'), '%Y-%m') as month\_year, count(distinct customer\_id) as active\_customers**

**from invoice**

**group by month\_year),**

**churn\_analysis as**

**(select month\_year, active\_customers,**

**lag(active\_customers) over (order by month\_year) as prev\_month\_customers,**

**lag(active\_customers) over (order by month\_year) - active\_customers as churned\_customers**

**from monthly\_active\_customers)**

**select month\_year, active\_customers, prev\_month\_customers,**

**case**

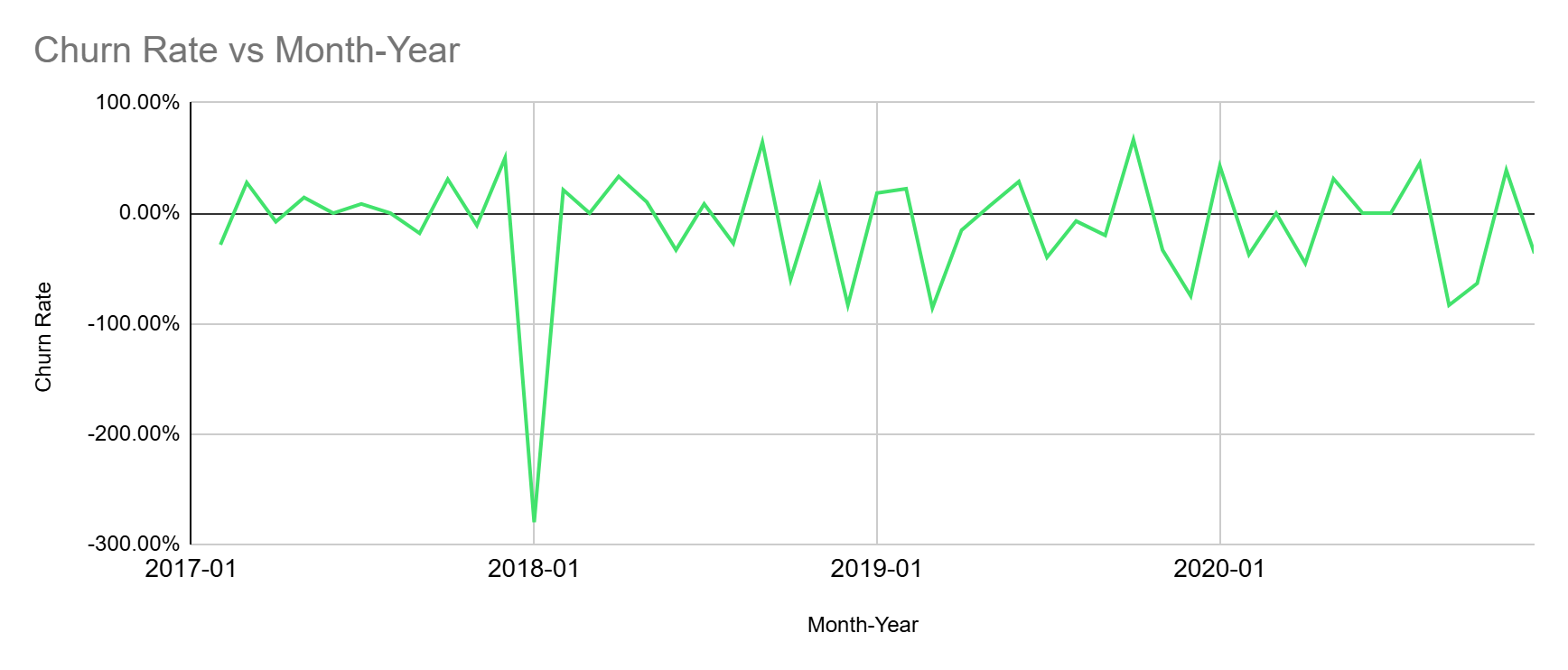
**when churned\_customers > 0 then concat('+', round((churned\_customers / prev\_month\_customers) \* 100, 2), '%')**

**else concat(round((churned\_customers / prev\_month\_customers) \* 100, 2), '%')**

**end as churn\_rate**

**from churn\_analysis**

**order by month\_year;**

**

Tracking churn rate over time provides insights into customer retention and engagement. A high churn rate may indicate dissatisfaction or a lack of incentives to return, while a stable or decreasing churn rate suggests strong customer loyalty. Identifying periods of increased churn can help businesses strategize targeted promotions, personalized outreach, or loyalty programs to improve retention. Understanding these trends enables better decision-making to sustain long-term customer relationships.

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

**Answer :** This analysis focuses on identifying the top-selling artists and genres in the USA based on total revenue.

The **first query** calculates total sales for each artist by summing up the invoice amounts associated with their tracks. It then determines their percentage contribution to the total sales in the country and ranks the top five artists.

The **second query** follows a similar approach but groups sales by genre instead of artists. It calculates the total revenue generated by each genre and its percentage contribution to overall sales in the USA.

These queries provide a breakdown of which artists and genres dominate the U.S. music market.

*Total Sales For Each Artist*

**with artist\_sales as**

**(select g.name as genre\_name, ar.name as artist\_name,**

**sum(i.total) as total\_sales**

**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 album al**

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

**join artist ar**

**on al.artist\_id = ar.artist\_id**

**join genre g**

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

**where i.billing\_country = 'USA'**

**group by genre\_name, artist\_name),**

**total\_sales as**

**(select sum(total\_sales) as usa\_total\_sales from artist\_sales)**

**select ars.artist\_name, ars.total\_sales,**

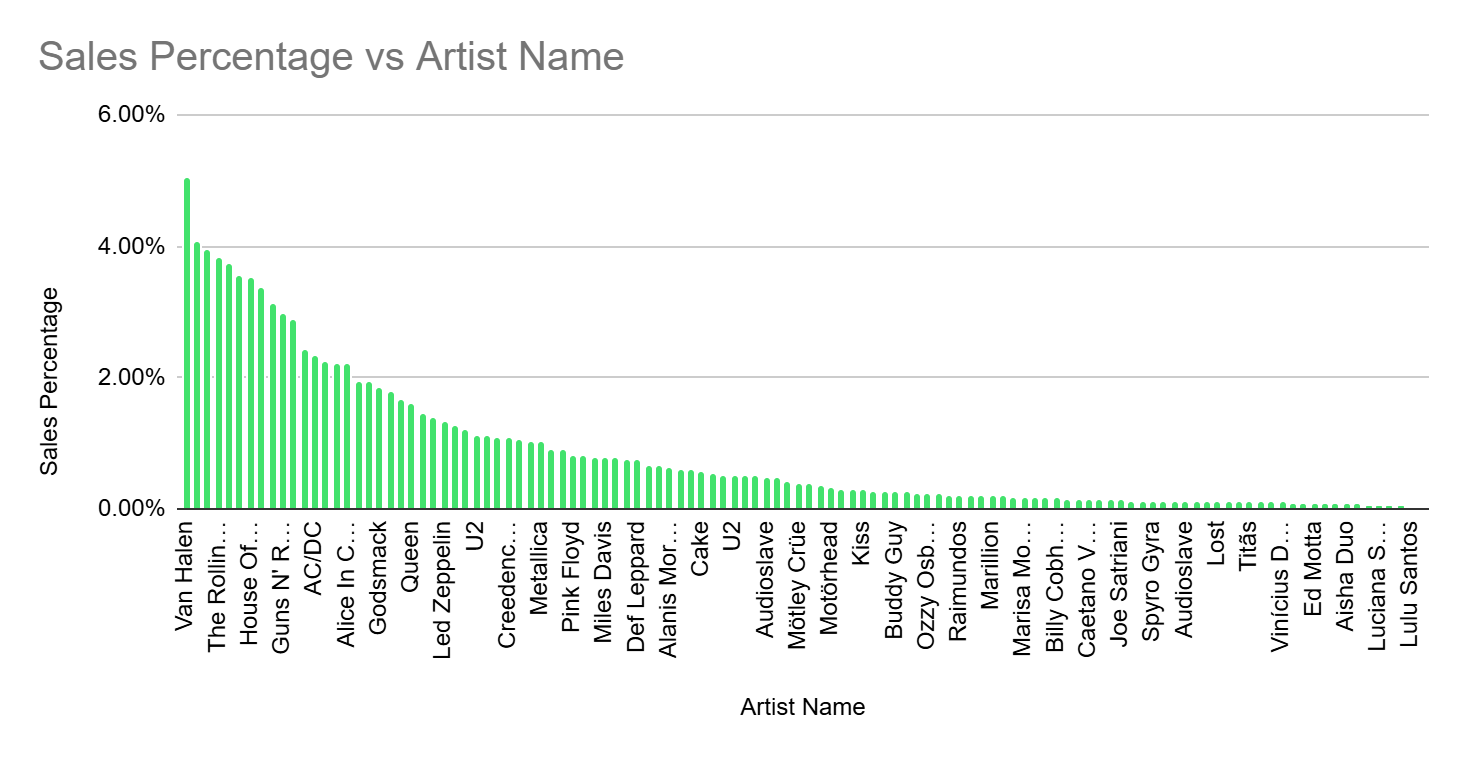
**concat(round((ars.total\_sales / ts.usa\_total\_sales) \* 100, 2),'%') as sales\_contribution**

**from artist\_sales ars**

**join total\_sales ts**

**order by ars.total\_sales desc**

**limit 5;**

**

*Sales contribution by genre in USA*

**with genre\_sales as**

**(select g.name as genre\_name, sum(i.total) as total\_sales**

**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 g**

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

**where i.billing\_country = 'USA'**

**group by genre\_name),**

**total\_sales as**

**(select sum(total\_sales) as usa\_total\_sales from genre\_sales)**

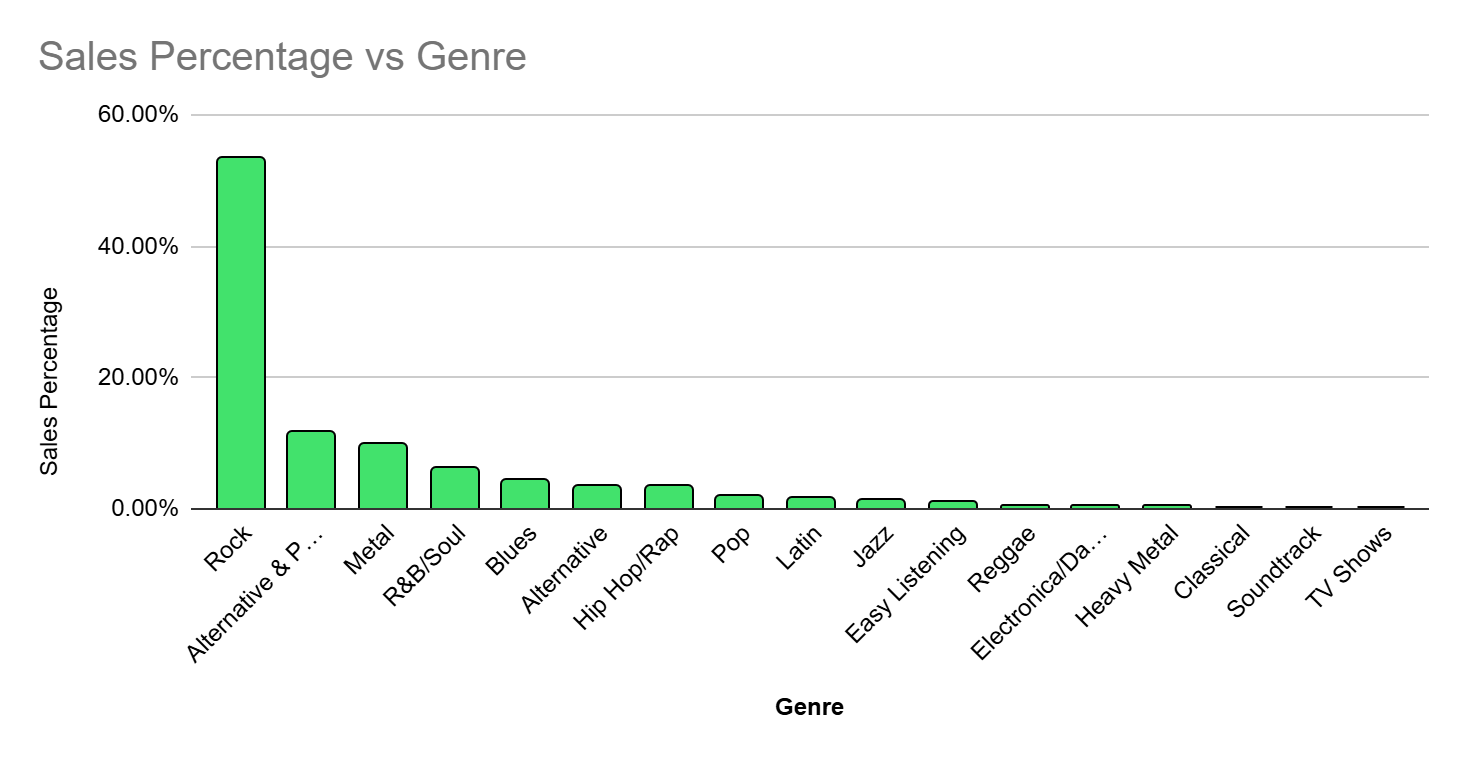
**select gs.genre\_name, gs.total\_sales,**

**concat(round((gs.total\_sales / ts.usa\_total\_sales) \* 100, 2),'%') as sales\_contribution**

**from genre\_sales gs**

**join total\_sales ts**

**order by total\_sales desc, sales\_contribution desc;**

**

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

**Answer :** To understand customer preferences, we analyzed how many different genres each customer has purchased music from. By counting the distinct genres associated with each customer’s purchases, we identified those who have bought from at least three different genres. This helps in recognizing customers with diverse music tastes, which could be useful for targeted marketing and personalized recommendations. The query to achieve this is :

**with customer\_genre\_count as**

**(select i.customer\_id, count(distinct g.genre\_id) as genre\_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 g**

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

**group by i.customer\_id)**

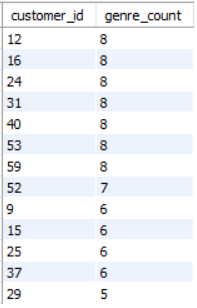
**select customer\_id, genre\_count**

**from customer\_genre\_count**

**where genre\_count >= 3**

**order by genre\_count desc, customer\_id;**

*Sample Output :*



Our analysis revealed that many customers engage with a variety of music genres, purchasing from at least three distinct categories. This suggests that a significant portion of the customer base has diverse musical preferences rather than sticking to a single genre. On the other hand, some customers show strong loyalty to specific genres, indicating a focused listening habit. Recognizing these patterns can be valuable for businesses looking to offer personalized recommendations, cross-genre promotions, or curated playlists. Customers with broader genre interests can be ideal targets for bundled music deals, while those with specific preferences may respond well to exclusive content or artist-based promotions.

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

**Answer :** This query ranks music genres based on total sales in the USA. By aggregating revenue for each genre and applying a ranking function, we identify which genres generate the highest revenue. This insight helps in understanding market trends and customer preferences within the country.

**select g.genre\_id, g.name as genre\_name, sum(i.total) as total\_sales,**

**rank() over (order by sum(i.total) desc) as genre\_rank**

**from genre g**

**join track t**

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

**join invoice\_line il**

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

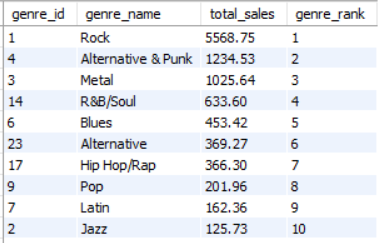
**join invoice i**

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

**where i.billing\_country = 'USA'**

**group by g.genre\_id, g.name;**

*Sample Output :*



The analysis highlights the most popular music genres in the USA based on revenue. The ranking provides a clear picture of which genres drive the most sales, which can be valuable for record labels, artists, and music platforms looking to optimize their offerings. The top-ranked genres likely dominate customer interest, while lower-ranked ones may represent niche markets with growth potential. Understanding this distribution allows businesses to tailor marketing campaigns, curate playlists, and allocate resources to maximize revenue.

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

**Answer :** This query identifies customers who have not made a purchase in the last three months. Tracking inactive customers helps businesses understand retention trends and re-engage them with targeted offers. The approach I used to achieve this task is :

1. Identified the latest purchase date for each customer using MAX(invoice\_date).
2. Calculated the inactivity cutoff by subtracting three months from the latest recorded transaction in the dataset.
3. Filtered customers whose last purchase date falls before this cutoff.
4. Sorted results to list the longest inactive customers first.

**with customer\_latest\_purchase as**

**(select c.customer\_id, concat(c.first\_name, ' ', c.last\_name) as customer\_name,**

**date(max(i.invoice\_date)) as last\_purchase\_date**

**from customer c**

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

**group by c.customer\_id, customer\_name)**

**select clp.customer\_id, clp.customer\_name, clp.last\_purchase\_date**

**from customer\_latest\_purchase clp**

**where clp.last\_purchase\_date <= date\_sub((select max(invoice\_date) from invoice), interval 3 month)**

**order by clp.last\_purchase\_date asc;**

#### *Sample Output :*

#### Instead of assuming today's date, we used MAX(invoice\_date) from the dataset to ensure the analysis is based on the most recent recorded transactions. This prevents incorrect classifications due to outdated data in this historical dataset.

The results show which customers have been inactive for over three months. Businesses can use this to re-engage them through promotions or understand why they stopped purchasing.

SUBJECTIVE QUESTIONS

***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.

**Answer :** To decide which three albums should be prioritized for promotion in the USA, we looked at which music genres are selling the most. Since we don’t have direct data on record labels, the best way to approach this was to see which genres are performing well and then find the top albums within those genres. If a genre is already popular, albums from that genre have a better chance of doing well with the right marketing push.

We started by identifying the highest-selling genres in the USA based on total revenue. Once we had that, we dug deeper to see which albums contributed the most to those sales. The three albums with the highest sales numbers stood out as the best choices for promotion. Since they’re already in demand, running targeted ads or special offers on them would likely boost sales even further. The query to find the required result is :

**with genre\_sales as**

**(select g.genre\_id, g.name as genre\_name, sum(i.total) as total\_sales**

**from genre g**

**join track t 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.genre\_id, g.name**

**order by total\_sales desc**

**limit 3),**

**top\_albums as**

**(select al.album\_id, al.title as album\_name, g.name as genre\_name,**

**sum(il.unit\_price \* il.quantity) as album\_sales**

**from track t**

**join album al**

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

**join genre g**

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

**join invoice\_line il**

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

**join invoice i**

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

**where i.billing\_country = 'USA'**

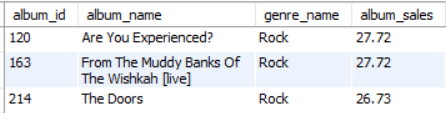
**and g.genre\_id in (select genre\_id from genre\_sales)**

**group by al.album\_id, al.title, g.name**

**order by album\_sales desc)**

**select \* from top\_albums;**

*Output :*



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

**Answer :** To determine the top-selling genres in different countries (excluding the USA), we analyzed total sales per genre across various regions. We first linked invoices with invoice lines to track purchases. From there, we joined the track and genre tables to identify which genres were being purchased. The total revenue generated by each genre was then summed up for every country, and we used the RANK() function to assign rankings within each country based on sales.

From the analysis, we observed that different regions tend to have varied music preferences. Some genres remain consistently popular across multiple countries, while others show more localized appeal. For instance, ***Rock, Metal*** and ***Pop*** frequently dominate in high-sales countries, but there are unique cases where regional genres gain traction. Understanding these patterns helps in tailoring marketing efforts based on regional preferences, ensuring better-targeted promotions and sales strategies. Query used for this analysis :

**with top\_genres as**

**(select i.billing\_country, g.genre\_id, g.name as genre\_name, sum(i.total) as total\_sales,**

**rank() over (partition by i.billing\_country order by sum(i.total) desc) as genre\_rank**

**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 g**

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

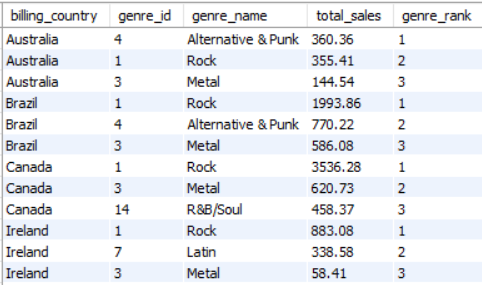
**where i.billing\_country <> 'USA'**

**group by i.billing\_country, g.genre\_id, g.name)**

**select \* from top\_genres**

**where genre\_rank <= 3**

This query gives me the top 3 genres for each country that helps further analyse the most popular genres.

*Sample Output :*

**3.** 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?

**Answer :** To understand how long-term customers and new customers behave differently in terms of purchases, we broke down the analysis into a few key steps:

### Calculating Customer Purchase Behavior

We first determined each customer’s purchase frequency, meaning how many separate purchases (invoices) they have made.  
We also calculated their total spending (sum of all purchases) and their average order value (total spent divided by the number of purchases).  
Additionally, we measured customer tenure, which is the duration between a customer’s first and last recorded purchase. This helps in distinguishing long-term customers based on their time on the platform.

### Classifying Customers into Two Categories

To define long-term customers, we set two conditions:

1. Any customer whose purchase frequency is higher than the average was labeled as a long-term customer.
2. Customers with a lower-than-average purchase frequency were categorized as new customers.  
   Instead of just considering frequency, tenure can also be used to refine the classification, ensuring that customers with both high frequency and longer presence on the platform are truly long-term.

### Comparing the Two Customer Groups

We then calculated key metrics for both groups:

* Average Purchase Frequency – How often each group places orders.
* Average Total Spending – The overall amount spent per customer.
* Average Order Value – The average spending per order.
* Average Customer Tenure – The average duration customers stay on the platform.

This helped us understand how their behaviors differ and what factors contribute to customer loyalty and spending habits.

### *Query to Derive These Results :*

**with customer\_purchase\_stats as**

**(select i.customer\_id,**

**min(i.invoice\_date) as first\_purchase\_date,**

**count(distinct i.invoice\_id) as total\_orders,**

**sum(i.total) as total\_revenue,**

**avg(i.total) as avg\_order\_value**

**from invoice i**

**group by i.customer\_id),**

**customer\_type as**

**(select cps.customer\_id, cps.total\_orders, cps.total\_revenue,**

**cps.avg\_order\_value,**

**round(datediff((select max(invoice\_date) from invoice),**

**cps.first\_purchase\_date) / 365, 2) as tenure\_years,**

**case**

**when round(datediff((select max(invoice\_date) from invoice),**

**cps.first\_purchase\_date) / 365, 2) > 1 and cps.total\_orders > (select avg(total\_orders) from**

**customer\_purchase\_stats)**

**then 'Long-Term Customer'**

**else 'New Customer'**

**end as customer\_category**

**from customer\_purchase\_stats cps)**

**select customer\_category,**

**round(avg(total\_orders), 2) as avg\_orders\_per\_customer,**

**round(avg(total\_revenue), 2) as avg\_total\_revenue\_per\_customer,**

**round(avg(avg\_order\_value), 2) as avg\_order\_value,**

**round(avg(tenure\_years), 2) as avg\_tenure\_years**

**from customer\_type**

**group by customer\_category;**

### *Chart*

### 

### Insights That Can Be Drawn

* Long-term customers purchase more frequently and have a higher total spend compared to new customers.
* However, average order value remains similar for both groups, meaning that while long-term customers spend more overall, they do not necessarily spend more per transaction.
* This suggests that customer retention and encouraging repeat purchases play a bigger role in revenue growth than trying to increase the amount spent per transaction.
* The tenure metric further highlights that customers who stay longer tend to develop a habit of purchasing more frequently over time.

### Business Implementations

* Loyalty programs: Reward frequent buyers with discounts, points, or exclusive offers to encourage retention.
* Personalized marketing: Use data insights to target new customers with incentives to make repeat purchases.
* Subscription or membership models: Offer exclusive benefits to keep long-term customers engaged.
* Customer engagement strategies: Regular follow-ups, personalized recommendations, and seasonal promotions can help convert new customers into long-term buyers.
* Since order values remain consistent, the best way to increase revenue is by encouraging repeat purchases rather than increasing per-order spending.

By analyzing these patterns, businesses can build data-driven customer retention strategies that maximize revenue and foster long-term customer relationships.

**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?

**Answer :** Product affinity analysis is a technique used to understand relationships between products that are frequently purchased together. By examining purchase patterns, businesses can uncover valuable insights into consumer behavior, enabling them to optimize marketing strategies, improve inventory management, and enhance customer recommendations.

My analysis focuses on identifying albums that are often bought together within the same transaction. This is achieved by pairing albums from the same invoice and counting how often these pairs appear across multiple purchases. By incorporating attributes such as genre, artist, and album, the analysis provides a deeper understanding of how customer preferences align and which types of music tend to be purchased together.

#### *Query to Identify Frequently Purchased Album Pairs*

**with purchase\_pairs as**

**(select il1.invoice\_id,**

**t1.genre\_id as genre\_1, t2.genre\_id as genre\_2,**

**al1.artist\_id as artist\_1, al2.artist\_id as artist\_2,**

**al1.album\_id as album\_1, al2.album\_id as album\_2**

**from invoice\_line il1**

**join invoice\_line il2**

**on il1.invoice\_id = il2.invoice\_id and il1.track\_id < il2.track\_id**

**join track t1**

**on il1.track\_id = t1.track\_id**

**join track t2**

**on il2.track\_id = t2.track\_id**

**join album al1**

**on t1.album\_id = al1.album\_id**

**join album al2**

**on t2.album\_id = al2.album\_id**

**where al1.album\_id != al2.album\_id),**

**paired\_purchases as**

**(select genre\_1, genre\_2, artist\_1, artist\_2, album\_1, album\_2,**

**count(\*) as times\_purchased\_together**

**from purchase\_pairs**

**group by genre\_1, genre\_2, artist\_1, artist\_2, album\_1, album\_2)**

**select g1.name as genre\_1, g2.name as genre\_2,**

**a1.name as artist\_1, a2.name as artist\_2,**

**al1.title as album\_1, al2.title as album\_2,**

**times\_purchased\_together**

**from paired\_purchases pp**

**join genre g1**

**on pp.genre\_1 = g1.genre\_id**

**join genre g2**

**on pp.genre\_2 = g2.genre\_id**

**join artist a1**

**on pp.artist\_1 = a1.artist\_id**

**join artist a2**

**on pp.artist\_2 = a2.artist\_id**

**join album al1**

**on pp.album\_1 = al1.album\_id**

**join album al2**

**on pp.album\_2 = al2.album\_id**

**order by times\_purchased\_together desc**

**limit 10;**

*Sample Output :   
  
*

One key application of this analysis is in product recommendations. Streaming services, online music stores, and retail businesses use similar techniques to suggest relevant albums to users. For example, if a significant number of customers purchase both a *Rock* album by *Queen* and a *Metal* album by *Metallica*, the system can recommend Metallica to a customer who recently bought Queen’s album. This increases engagement and encourages further purchases by leveraging customer preferences.

Another major benefit is in cross-selling and bundling strategies. Retailers can design promotional campaigns based on the insights derived from purchase patterns. If *Jazz* albums by *Miles Davis* are often bought alongside *Blues* albums by *Eric Clapton*, a store might offer special bundle discounts or curate themed collections that encourage customers to buy both. These strategies help maximize revenue while enhancing the shopping experience for consumers.

From an inventory management perspective, businesses can ensure that frequently paired albums are stocked together. This is particularly useful in physical retail stores, where placing complementary products close to each other can drive impulse purchases. Online platforms can replicate this approach by displaying “Frequently Bought Together” suggestions on product pages, improving discoverability.

Beyond music, product affinity analysis is widely applied in various industries, including e-commerce, grocery retail, and digital streaming platforms. Amazon, for instance, utilizes similar methods to recommend books, electronics, or fashion items based on past purchases. Similarly, food delivery apps analyze dish pairings to suggest complementary items, such as beverages or desserts, when placing an order.

By leveraging data-driven insights from purchase behavior, businesses can refine their marketing efforts, personalize customer interactions, and ultimately drive higher sales. This approach not only benefits companies by increasing revenue but also improves customer satisfaction by offering tailored recommendations that align with individual preferences.

**5.** 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?

**Answer :** Customer purchasing habits and churn rates are not the same everywhere. They change based on factors like location, economy, culture, and local competition. Some areas might have loyal customers who stick with a brand, while others might see frequent switching between brands. Understanding these differences can help businesses improve customer retention and make better marketing decisions.

One big reason for this variation is consumer preference. People in different places have different tastes. For example, in tech-driven cities, customers might prefer online purchases, while in smaller towns, physical stores might still be more popular. Spending power also matters. A high-income city might see more frequent purchases, while budget-conscious areas might have customers who shop less often or wait for discounts.

At the same time, churn rates (the rate at which customers stop buying) also change by region. In competitive areas with many alternatives, customers may switch brands often. But in places where choices are limited, customers are more likely to stick with familiar brands. The quality of service and customer experience is another major factor. If delivery is fast, service is excellent, and after-sales support is strong, people are more likely to remain loyal.

Economic conditions can also affect churn. If a region is facing a financial downturn or job losses, people may cut down on spending, leading to higher churn. Seasonal trends also play a role. Customers may buy more during festivals and holidays but reduce spending afterward. Similarly, urban customers tend to have more choices and may switch brands frequently, while in rural areas, limited options often lead to stronger brand loyalty.

To better understand these trends, businesses can analyze monthly churn rates by region. A structured query can help track how many customers are lost each month across different locations. The query that identifies churned customers, those who made a purchase last month but not in the current month across different countries :

**with monthly\_active as**

**(select customer\_id, billing\_country,**

**date\_format(invoice\_date, '%Y-%m') as purchase\_month**

**from invoice**

**group by customer\_id, billing\_country, purchase\_month),**

**customer\_counts as**

**(select billing\_country, purchase\_month,**

**count(distinct customer\_id) as active\_customers**

**from monthly\_active**

**group by billing\_country, purchase\_month),**

**churn\_analysis as**

**(select billing\_country, purchase\_month, active\_customers,**

**lag(active\_customers) over (partition by billing\_country order by**

**purchase\_month) as prev\_month\_customers**

**from customer\_counts)**

**select billing\_country, purchase\_month,**

**sum(active\_customers) as total\_active\_customers,**

**sum(case when prev\_month\_customers is null then 0**

**else prev\_month\_customers - active\_customers**

**end) as total\_churned\_customers,**

**concat(case when sum(prev\_month\_customers - active\_customers) > 0**

**then '+' else '' end,**

**round((sum(prev\_month\_customers - active\_customers) \* 100.0 /**

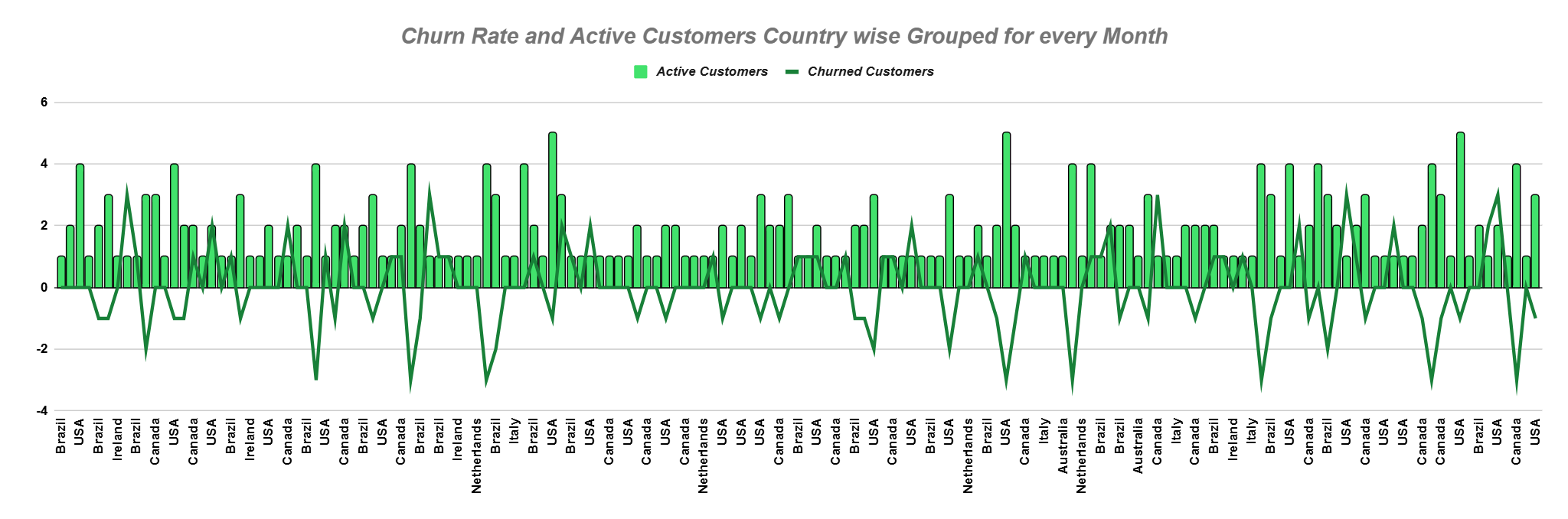
**nullif(sum(prev\_month\_customers), 0)),**

**2),' %') as churn\_rate**

**from churn\_analysis**

**group by billing\_country, purchase\_month**

**order by purchase\_month asc, billing\_country asc;**



By studying the results, businesses can see which regions have the highest churn rates and investigate the possible causes. If one region has a much higher churn rate, it could mean there is a problem with pricing, service quality, or local competition. In such cases, businesses can use strategies like localized discounts, improved customer service, or better loyalty programs to retain customers.

There is also a connection between churn rates and local demographics or economic conditions. Younger customers may prefer digital experiences and might switch brands more easily, requiring businesses to invest in personalized online marketing. In economically weaker regions, flexible payment options and discounts can help retain customers.

By keeping track of regional purchasing behaviors and churn trends, businesses can adjust their strategies to fit local needs. A region-specific approach can help improve retention, increase revenue, and create stronger customer relationships across different markets.

**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?

**Answer :**  Customer risk profiling helps in identifying which customer segments are more likely to churn or reduce their spending over time. While age and gender are often important factors in such analyses, our dataset did not include these attributes. Therefore, our analysis primarily focuses on **location, purchase history, and spending patterns** to determine risk levels.

One of the key aspects of this profiling is tracking **monthly spending trends**. Customers who show a significant drop in spending compared to their previous month are categorized as **High Risk**, indicating a likelihood of reduced engagement or potential churn. Those whose spending remains unchanged are classified as **Moderate Risk**, while those with an increasing trend fall into the **Low Risk** category.

To achieve this, we structured our query to:

* Calculate each customer's **total spending, purchase frequency, and average time between purchases**.
* Compare **current and past months’ spending** to track fluctuations.
* Categorize customers into **risk levels** based on their spending behavior.

**with customer\_spending as**

**(select c.customer\_id, c.country,**

**date\_format(i.invoice\_date, '%Y-%m-01') as purchase\_month,**

**count(distinct i.invoice\_id) as purchase\_count,**

**sum(i.total) as total\_spent**

**from customer c**

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

**group by c.customer\_id, c.country, purchase\_month),**

**purchase\_intervals as**

**(select customer\_id, invoice\_date,**

**lag(invoice\_date) over (partition by customer\_id order by invoice\_date) as**

**prev\_purchase\_date**

**from invoice),**

**customer\_avg\_purchase as**

**(select customer\_id,**

**round(avg(datediff(invoice\_date, prev\_purchase\_date)), 2) as**

**avg\_purchase\_days**

**from purchase\_intervals**

**where prev\_purchase\_date is not null**

**group by customer\_id),**

**spending\_with\_lag as**

**(select cs.\*, cap.avg\_purchase\_days,**

**lag(cs.total\_spent) over (partition by cs.customer\_id order by**

**cs.purchase\_month) as prev\_month\_spent**

**from customer\_spending cs**

**left join customer\_avg\_purchase cap**

**on cs.customer\_id = cap.customer\_id),**

**spending\_risk as**

**(select customer\_id, country, purchase\_month, purchase\_count, total\_spent,**

**avg\_purchase\_days, prev\_month\_spent,**

**case**

**when total\_spent < prev\_month\_spent then 'High Risk'**

**when total\_spent = prev\_month\_spent then 'Moderate Risk'**

**else 'Low Risk'**

**end as risk\_category**

**from spending\_with\_lag)**

**select country, purchase\_month, risk\_category,**

**count(customer\_id) as customer\_count,**

**round(avg(purchase\_count), 2) as avg\_purchases,**

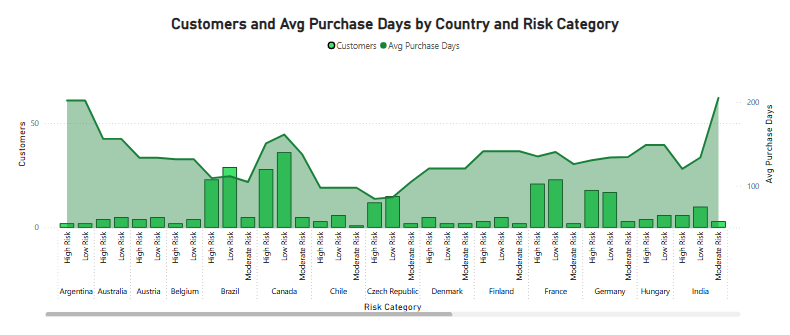
**round(avg(total\_spent), 2) as avg\_spent,**

**round(avg(avg\_purchase\_days), 2) as avg\_purchase\_days**

**from spending\_risk**

**group by purchase\_month, risk\_category, country**

**order by purchase\_month asc, risk\_category asc;**



**T**he results are then aggregated by **country and purchase month**, ensuring a structured and sequential representation of different geographic regions. This allows us to observe spending trends at a regional level, helping businesses understand where the risk is higher. Factors such as **local economic conditions, store accessibility, and competitive alternatives** could also play a role in why certain regions experience more fluctuations in spending.

By focusing on **historical purchase data**, this analysis provides valuable insights into customer retention challenges, enabling businesses to take proactive measures like **targeted promotions, loyalty programs, or personalized offers** to retain at-risk customers.

**7.** Customer Lifetime Value Modeling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns among customers who have stopped purchasing?

**Answer :** Customer Lifetime Value (CLV) is a key metric that helps businesses estimate the total revenue a customer will generate throughout their relationship with the company. It is crucial for making informed decisions on marketing, pricing, and customer retention strategies.

To predict CLV, businesses analyze customer data, focusing on three main aspects:

* **Tenure** – The length of time a customer has been active.
* **Purchase History** – How often they buy and how much they spend.
* **Engagement** – Their level of interaction with the company (e.g., website visits, email interactions, loyalty program usage).

By analyzing these factors, businesses can group customers into different segments, identify high-value customers, and design personalized marketing efforts.

#### **How We Calculated CLV**

To accurately predict CLV, we examined key customer behaviors such as:

* **Total Spend** – The sum of all purchases made by a customer.
* **Average Order Value (AOV)** – The typical amount spent per order.
* **Purchase Frequency** – How often a customer places an order.
* **Recency of Purchase** – The time since their last order.

Using this data, we identified patterns in customer spending habits and estimated their future value to the business.

#### **Churn Rate Calculation Using Lag Function**

To understand customer churn, we tracked how many customers stopped purchasing over time. We defined churned customers as those who have not made a purchase in over **180 days**.

We used the **LAG function** to compare a customer’s last purchase date with previous months, allowing us to observe changes in activity and detect churn patterns. The churn rate was then calculated on a **monthly basis** to track trends effectively.Below is the SQL query used to calculate the **monthly churn rate using the LAG function**:

**with data\_boundaries as**

**(select max(invoice\_date) as latest\_date from invoice),**

**customer\_spending as**

**(select c.customer\_id, c.country,**

**sum(i.total) as total\_spend,**

**count(i.invoice\_id) as total\_orders,**

**avg(i.total) as avg\_order\_value,**

**min(i.invoice\_date) as first\_order\_date,**

**max(i.invoice\_date) as last\_order\_date,**

**timestampdiff(month, min(i.invoice\_date),**

**(select latest\_date from data\_boundaries)) as tenure\_months**

**from customer c**

**join invoice i**

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

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

**customer\_churn as**

**(select customer\_id, country,**

**lag(last\_order\_date) over (partition by customer\_id order by last\_order\_date) as prev\_order\_date,**

**last\_order\_date**

**from customer\_spending),**

**churn\_analysis as**

**(select country,**

**count(case when datediff((select latest\_date from data\_boundaries), last\_order\_date) > 180**

**then customer\_id end) as churned\_customers,**

**count(customer\_id) as total\_customers,**

**round(avg(tenure\_months),1) as avg\_tenure,**

**round(avg(total\_spend),2) as avg\_lifetime\_spend**

**from customer\_spending**

**group by country),**

**clv\_estimation as**

**(select country,**

**avg\_lifetime\_spend / nullif((churned\_customers / total\_customers),0) as estimated\_clv**

**from churn\_analysis)**

**select c.country, c.total\_customers,**

**c.churned\_customers, c.avg\_tenure, c.avg\_lifetime\_spend,**

**round(100 \* c.churned\_customers / nullif(c.total\_customers,0),2) as churn\_rate,**

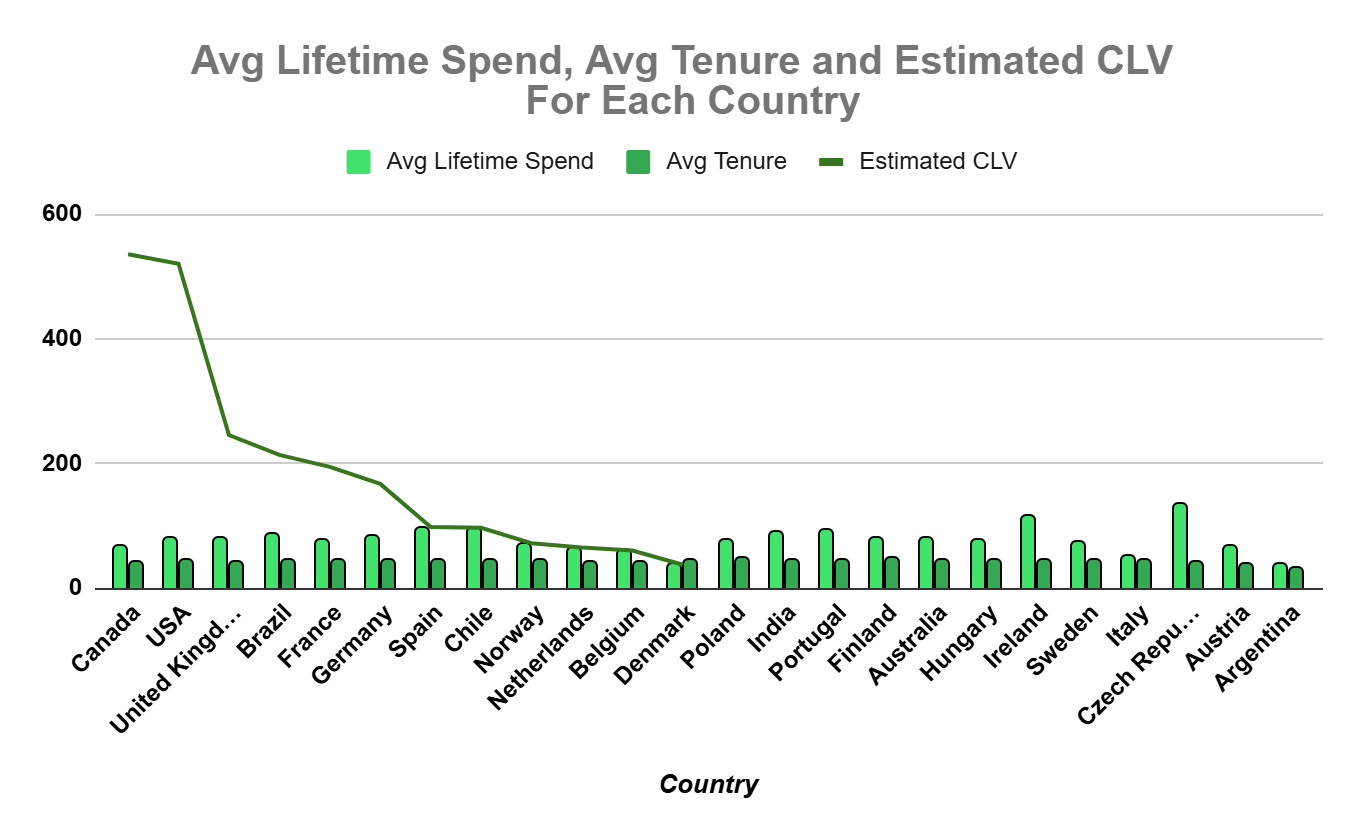
**round(e.estimated\_clv,2) as estimated\_clv**

**from churn\_analysis c**

**left join clv\_estimation e**

**on c.country = e.country**

**order by estimated\_clv desc;**

****

#### **Common Characteristics of Churned Customers**

After analyzing customer behavior, we observed some key patterns among those who stopped purchasing:

* **Low Purchase Frequency** – Customers who made infrequent purchases were more likely to churn.
* **Lower Average Order Value (AOV)** – Those who spent less per order tended to drop off faster.
* **Shorter Tenure** – Newer customers had a higher likelihood of churning compared to long-term customers.
* **Regional Differences** – Some locations had higher churn rates, possibly due to pricing, competition, or service issues.
* **Engagement Levels** – Customers who engaged with loyalty programs, discounts, or personalized recommendations had **lower churn rates**.
* **Estimated CLV :**  For the countries having churn rate increased over time, there’s no estimation of CLV for them.

#### **How Cinook Can Use These Insights**

By leveraging CLV modeling and churn analysis, businesses can take targeted actions to **reduce customer attrition and maximize revenue**. Some effective strategies include:

* **Personalized Retention Campaigns** – Offering discounts, exclusive deals, or loyalty rewards to high-risk customers.
* **Reactivation Strategies** – Sending re-engagement emails or special offers to customers who haven't purchased recently.
* **Segmentation-Based Marketing** – Using CLV data to focus marketing efforts on high-value customer segments.
* **Improving Customer Experience** – Enhancing product offerings, pricing strategies, and customer service to retain customers.

By implementing these strategies, businesses can **reduce churn, increase customer lifetime value, and improve overall profitability**.

**8.** If data on promotional campaigns (discounts, events, email marketing) is available, how could you measure their impact on customer acquisition, retention, and overall sales?

**Answer :** Promotional campaigns, such as discounts, events, and email marketing, help businesses attract new customers, retain existing ones, and boost sales. To measure their effectiveness, it’s important to track key metrics before, during, and after the campaign.

#### **1. Customer Acquisition**

A well-executed campaign should bring in new customers who wouldn’t have purchased otherwise. The impact on acquisition can be measured by:

* **New Customer Count** – Tracking the number of first-time buyers during and after the promotion.
* **Website Traffic & Sign-ups** – An increase in visits or registrations indicates that the campaign is generating interest.
* **Conversion Rate** – If more visitors are turning into buyers, it suggests the promotion is effective in driving purchases.

However, it’s essential to check if these new customers stick around or if they only came for the discount.

#### **2. Customer Retention**

The real success of a promotion isn’t just in acquiring customers but in keeping them engaged. To assess retention, businesses can look at:

* **Repeat Purchase Rate** – How many customers make another purchase after the promotion?
* **Time Between Purchases** – Are customers coming back sooner than they normally would?
* **Engagement Post-Campaign** – Do customers continue interacting with emails, social media, or loyalty programs?

If retention rates are low, it might mean that customers were only interested in the one-time offer rather than the brand itself.

#### **3. Impact on Sales and Revenue**

While a promotion might boost sales, it’s important to ensure it doesn’t hurt profitability. Key metrics include:

* **Total Sales Volume** – Did the number of transactions increase significantly?
* **Average Order Value (AOV)** – Did customers buy only discounted items, or did they also purchase full-priced products?
* **Profit Margins** – Was the revenue increase enough to justify the discount or event costs?

A common mistake is running heavy discounts that bring in sales but lower overall profits. The goal should be to increase revenue without sacrificing long-term sustainability.

#### **4. Using Test and Control Groups**

One of the best ways to measure the true impact of a campaign is by using test and control groups:

* **Test Group** – Customers who received the promotion.
* **Control Group** – Customers who did not receive the promotion.

By comparing the behavior of both groups, businesses can determine if the promotion directly influenced sales and customer behavior. If the test group shows a higher retention rate or increased spending, it proves the campaign had a meaningful impact.

#### **5. Long-Term Effects**

Some campaigns create short-term spikes in sales but fail to drive lasting customer loyalty. To evaluate long-term success, businesses should analyze:

* **Customer Lifetime Value (CLV)** – Are customers acquired through the promotion more valuable over time?
* **Post-Campaign Drop-off** – Did sales drop sharply after the promotion ended?
* **Brand Perception** – Did the campaign enhance the brand’s image, or did it attract only deal-seekers?

If a campaign leads to sustained growth in engagement and repeat purchases, it can be considered a long-term success. However, if sales decline immediately after the promotion ends, it may indicate that customers were only motivated by the temporary offer.

By tracking these factors, businesses can optimize future marketing strategies, ensuring that promotions not only boost short-term sales but also contribute to sustainable growth.

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

**Answer :** If I were given a Chinook dataset without any specific objectives, I would start by understanding how the business works. Since Chinook is a digital music store or streaming service, I’d first look at how it makes money. It could be through single-track purchases, album sales, subscriptions, or even ads. I’d also try to figure out what makes a customer valuable. Are they the ones who spend the most, buy frequently, or engage with music recommendations? At the same time, I’d think about possible challenges the business might be facing, like low customer retention, declining sales, or marketing efforts that don’t seem to be working.

Once I have a good understanding of the business, I’d dive into the dataset to see what kind of information is available. I’d check for things like customer details, purchase history, sales trends, and marketing data. Before doing any real analysis, I’d clean the data by fixing any missing values, inconsistencies, or errors to make sure everything is accurate. After that, I’d define key business metrics to focus on. Some of the most important ones would be customer retention, sales trends, and how much revenue different types of customers contribute. I’d also ask some important questions, like whether people who buy full albums spend more over time compared to those who only buy single tracks, or if personalized recommendations actually lead to more purchases.

To dig deeper, I’d analyze customer behavior by grouping people based on their spending habits. I’d check if some users are regular buyers while others only make a one-time purchase. I’d also look for patterns in buying trends, like whether people who buy rock albums also tend to buy pop songs or if certain times of the year have higher sales. If marketing data is available, I’d run A/B tests to see which types of promotions actually drive sales and engagement.

After gathering insights, I’d focus on turning them into actionable recommendations. If customer retention is low, Chinook could introduce loyalty rewards or exclusive content to keep users engaged. If certain marketing campaigns are working well, the company could invest more in similar strategies. Optimizing product recommendations by promoting albums or bundles that customers are more likely to buy could also help boost sales. If sales are higher during certain seasons, running targeted promotions during those times could be a smart move.

Finally, to make sure these strategies continue to work, I’d set up real-time dashboards to track key performance indicators. A feedback loop would also be helpful, where data is continuously analyzed to improve decision-making over time. By following this approach of understanding the business, analyzing the data, identifying trends, and making data-driven recommendations, I’d be able to provide valuable insights that could help Chinook grow.

**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?

**Answer 10 :** To add a new column named **"ReleaseYear"** of type **INTEGER** in the **Albums** table, I would use the ALTER TABLE statement. Since we are only adding a new column and not modifying any existing data, this is a straightforward operation.

Here’s the SQL query to accomplish this :

**alter table album**

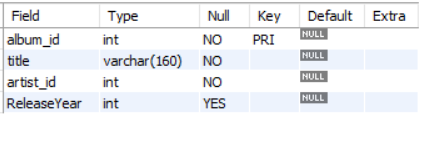
**add column ReleaseYear integer;**

This command updates the **Albums** table by introducing a new column named **ReleaseYear**, which will store integer values representing the year each album was released. Since we haven’t specified any constraints like NOT NULL or DEFAULT, the column will initially contain NULL values for existing records until they are updated.

After running the above command, I checked the album table to check if the query executed correctly or not. For this, I used DESCRIBE TABLE.

**describe album;**

By the help of this execution, I got a description of the album table mentioned below. Hence, making sure that the altering of table happened successfully. After adding the new column, this is the table description :



**11.** Chinook is interested in understanding the purchasing behavior of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write an SQL query to provide this information.

**Answer :** To analyze customer purchasing behavior based on geographical location, the **invoice** table serves as the primary source, as it contains details on transactions, including the **billing country** for each purchase. This enables the categorization and comparison of customer spending patterns across different locations.

### **1. Identifying Customer Distribution**

Understanding customer distribution across countries requires counting distinct customer\_id values from the invoice table. This helps determine where the majority of purchases originate and highlights key markets.

### **2. Measuring Total Revenue Per Country**

The total revenue generated from each country is calculated by summing the total column from the invoice table. This provides insights into the financial contribution of customers from different locations.

### **3. Analyzing Purchasing Patterns**

Purchasing habits can be examined by evaluating the total number of tracks bought per country. The **invoice\_line** table, which records individual track purchases, contains the quantity column that represents the number of tracks purchased per invoice. Summing this value for each country helps in understanding music consumption trends.

### **4. Calculating Key Metrics**

To provide actionable insights, the following key metrics are calculated:

* **Average total spending per customer** – Helps in identifying the purchasing power of customers from different countries.
* **Average number of tracks purchased per customer** – Indicates whether customers from certain locations tend to buy more music.

**with CountryStats as**

**(select i.billing\_country as country,**

**count(distinct i.customer\_id) as customer\_count,**

**sum(i.total) as total\_amount\_spent,**

**sum(il.quantity) as total\_tracks\_purchased**

**from invoice i**

**join invoice\_line il**

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

**group by i.billing\_country),**

**FinalStats as**

**(select country,**

**customer\_count,**

**round(total\_amount\_spent / customer\_count, 2) as avg\_total\_spending,**

**round(total\_tracks\_purchased / customer\_count, 2) as**

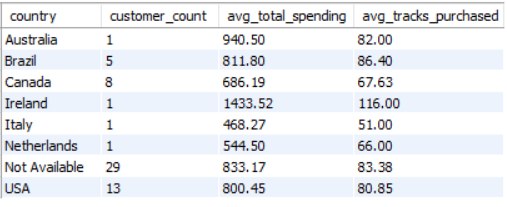
**avg\_tracks\_purchased**

**from CountryStats)**

**select \* from FinalStats**

**order by country, avg\_total\_spending desc;**

*Output :*

**

### **6. Interpreting the Data**

The results can be organized by **average spending per customer** in descending order to highlight high-value markets. This analysis enables Chinook to make data-driven decisions regarding targeted marketing strategies and sales optimization in different geographical regions.