**CHINOOK MUSIC STORE ANALYSIS**

Objective Questions

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

Indeed, there are missing entries in the customer, employee, and track tables that are replaced with "0" for numerical columns and "Unknown" or "None" for categorical fields. No table has duplicate values.

SQL code for identifying duplicates and missing values :

**SELECT \* FROM album;**

* album\_id, title, artist\_id

**SELECT \* FROM artist;**

* artist\_id, name

**SELECT \* FROM customer;**

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

**SELECT \* FROM employee;**

* employee\_id, last\_name, first\_name, title, reports\_to, birthdate, hire\_date, address, city, state, country, postal\_code, phone, fax, email

**SELECT \* FROM genre;**

* genre\_id, name

**SELECT \* FROM invoice;**

* invoice\_id, customer\_id, invoice\_date, billing\_address, billing\_city, billing\_state, billing\_country, billing\_postal\_code, total

**SELECT \* FROM invoice\_line;**

* invoice\_line\_id, invoice\_id, track\_id, unit\_price, quantity

**SELECT \* FROM media\_type;**

* media\_type\_id, name

**SELECT \* FROM playlist;**

* playlist\_id, name

**SELECT \* FROM playlist\_track;**

* playlist\_id, track\_id

**SELECT \* FROM track;**

* track\_id, name, album\_id, media\_type\_id, genre\_id, composer, milliseconds, bytes, unit\_price

Verifying the album table's missing values :

**SELECT \* FROM album**

**WHERE album\_id IS NULL**

**OR title IS NULL**

**OR artist\_id IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT**

**album\_id,**

**title,**

**artist\_id,**

**COUNT(\*) AS totsl\_count**

**FROM album**

**GROUP BY album\_id, title, artist\_id**

**HAVING COUNT(\*) > 1;**  -- no duplicate values found

Verifying for missing values in artist :

**SELECT \* FROM artist WHERE artist\_id IS NULL**

**OR name IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT**

**artist\_id,**

**name,**

**COUNT(\*) AS total\_count**

**FROM artist**

**GROUP BY artist\_id, name**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Verifying for missing values in customer :

**SELECT \* FROM customer**

**WHERE customer\_id IS NULL**

**OR first\_name IS NULL**

**OR last\_name IS NULL**

**OR company IS NULL**

**OR address IS NULL**

**OR city IS NULL**

**OR state IS NULL**

**OR country IS NULL**

**OR postal\_code IS NULL**

**OR phone IS NULL**

**OR fax IS NULL**

**OR email IS NULL**

**OR support\_rep\_id IS NULL;** -- missing values found

Verifying for duplicate values :

**SELECT \*,**

**COUNT(\*) AS total\_count**

**FROM customer**

**GROUP BY customer\_id, email**

**HAVING COUNT(\*) > 1;**  -- no duplicate values found

Verifying missing values for employee :

**SELECT \* FROM employee**

**WHERE employee\_id IS NULL**

**OR last\_name IS NULL**

**OR first\_name IS NULL**

**OR title IS NULL**

**OR reports\_to IS NULL**

**OR birthdate IS NULL**

**OR hire\_date IS NULL**

**OR address IS NULL**

**OR city IS NULL**

**OR state IS NULL**

**OR country IS NULL**

**OR postal\_code IS NULL**

**OR phone IS NULL**

**OR fax IS NULL**

**OR email IS NULL;** -- missing values found

Verifying for duplicate values :

**SELECT \*,**

**COUNT(\*) AS total\_count**

**FROM employee**

**GROUP BY employee\_id, email**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Verifying missing values for genre :

**SELECT \* FROM genre WHERE genre\_id IS NULL OR name IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT \*,**

**COUNT(\*) AS total\_genre\_count**

**FROM genre**

**GROUP BY genre\_id, name**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Verifying missing values for invoice :

**SELECT \* FROM invoice**

**WHERE invoice\_id IS NULL**

**OR customer\_id IS NULL**

**OR invoice\_date IS NULL**

**OR billing\_address IS NULL**

**OR billing\_city IS NULL**

**OR billing\_state IS NULL**

**OR billing\_country IS NULL**

**OR billing\_postal\_code IS NULL**

**OR total IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT \*,**

**COUNT(\*) AS total\_count FROM invoice**

**GROUP BY customer\_id, invoice\_id**

**HAVING COUNT(\*) > 1;**  -- no duplicate values found

Verifying missing values for invoice\_line :

**SELECT \* FROM invoice\_line WHERE invoice\_line\_id IS NULL**

**OR invoice\_id IS NULL OR track\_id IS NULL OR unit\_price IS NULL OR quantity IS NULL;**  -- no missing values found

Verifying for duplicate values :

**SELECT \*, COUNT(\*) AS total\_count FROM invoice\_line**

**GROUP BY invoice\_line\_id**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Verifying missing values for media\_type :

**SELECT \* FROM media\_type WHERE media\_type\_id IS NULL OR name IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT \*, COUNT(\*) AS total\_count FROM media\_type**

**GROUP BY media\_type\_id, name**

**HAVING COUNT(\*) > 1;**

**SELECT \*,**

**COUNT(\*) AS total\_count FROM playlist**

**GROUP BY playlist\_id, name**

**HAVING COUNT(\*) > 1;**

Verifying missing values for playlist :

**SELECT \* FROM playlist WHERE playlist\_id IS NULL OR name IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT \*,**

**COUNT(\*) AS total\_count FROM playlist**

**GROUP BY playlist\_id, name**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Verifying missing values for playlist\_track :

**SELECT \* FROM playlist\_track WHERE playlist\_id IS NULL OR track\_id IS NULL;** -- no missing values found

Verifying for duplicate values :

**SELECT \*, COUNT(\*) AS total\_count FROM playlist\_track**

**GROUP BY playlist\_id, track\_id**

**HAVING COUNT(\*) > 1;**  -- no duplicate values found

Verifying missing values for track :

**SELECT \* FROM track**

**WHERE track\_id is null**

**OR name IS NULL**

**OR album\_id IS NULL**

**OR media\_type\_id IS NULL**

**OR genre\_id IS NULL**

**OR composer IS NULL**

**OR milliseconds IS NULL**

**OR bytes IS NULL**

**OR unit\_price IS NULL;** -- missing values found

Verifying for duplicate values :

**SELECT \*, COUNT(\*) AS total\_count FROM track**

**GROUP BY track\_id, unit\_price**

**HAVING COUNT(\*) > 1;** -- no duplicate values found

Handling Missing Values :

**SET SQL\_SAFE\_UPDATES = 0;**

**UPDATE customer SET company = 'Unknown' WHERE company IS NULL;** -- 49 row(s) affected Rows Matched : 49 Changed : 49 Warnings : 0

**UPDATE customer SET state = 'None' WHERE state IS NULL;** -- 29 row(s) affected Rows Matched : 29 Changed : 29 Warnings : 0

**UPDATE customer SET phone = '+0 000 000 0000' WHERE phone IS NULL;** -- 1 row(s) affected Rows Matched : 1 Changed : 1 Warnings : 0

**UPDATE customer SET fax = '+0 000 000 0000' WHERE fax IS NULL;** -- 47 row(s) affected Rows Matched : 47 Changed : 47 Warnings : 0

**UPDATE customer SET postal\_code = '0' WHERE postal\_code IS NULL;** -- 4 row(s) affected Rows Matched : 4 Changed : 4 Warnings : 0

**UPDATE employee SET reports\_to ='0' WHERE reports\_to IS NULL;** -- 1 row(s) affected Rows Matched : 1 Changed : 1 Warnings : 0

**UPDATE track SET composer = 'Unknown' WHERE composer IS NULL;** -- 978 row(s) affected Rows Matched : 978 Changed : 978 Warnings : 0

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

We want 3 things together here :

* Top-selling tracks in the USA.
* Top-selling artist in the USA.
* Their most popular genres.

Query for the Analysis :

**SELECT**

**t.name AS top\_selling\_track,**

**ar.name AS top\_artist,**

**g.name AS top\_genre,**

**SUM(il.quantity) AS total\_copies\_sold,**

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

**FROM invoice\_line il**

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

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

**JOIN artist ar ON a.artist\_id = ar.artist\_id**

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

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

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

**GROUP BY t.track\_id, t.name, ar.name, g.name**

**ORDER BY total\_revenue DESC**

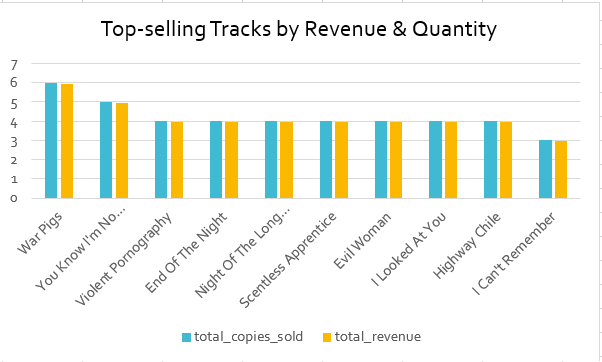
**LIMIT 10;**

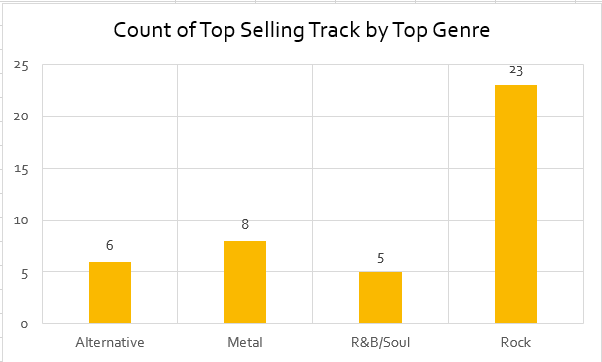
Output → A single results table showing :

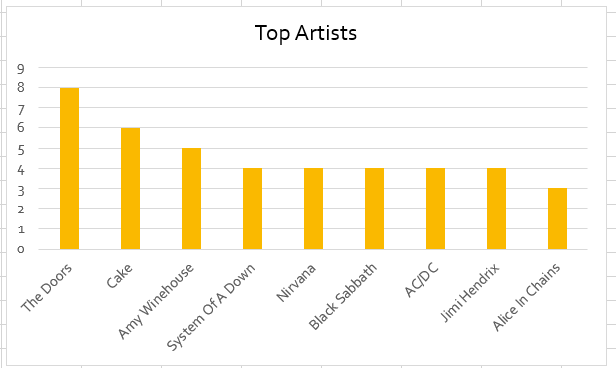
Top\_selling\_track | Top\_artist | Top\_genre | total\_copies\_sold | total\_revenue

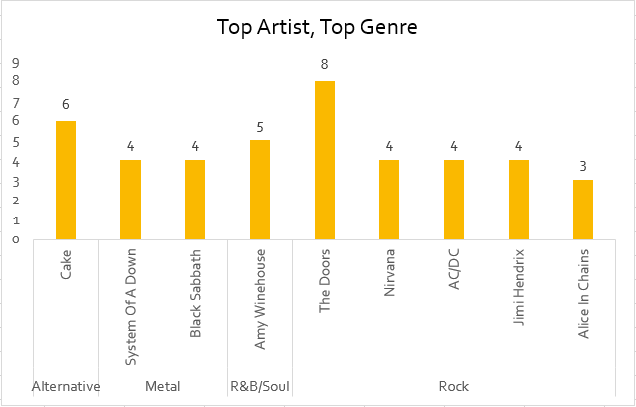
This table allows easy comparison between popularity and profitability.

****

****

****

****

****

**Tables Used :** invoice\_line, track, invoice, album, artist, genre

## **Approach :**

**Objective**

* Identify the most popular music in the USA based on track, artist, and genre.
* Measure popularity using both revenue (sales value) and volume (number of copies sold) to capture two perspectives:  
  + Revenue shows which tracks/artists earned the most money.
  + Volume shows which tracks were most widely purchased, even if cheaper.

**Method**

* Used the following tables:  
  + track, album, artist, genre → to get track names, artist names, and genre.
  + invoice and invoice\_line → to calculate sales data and filter only customers from the USA.
* Aggregated the sales data using:  
  + SUM(il.quantity) → total copies sold.
  + SUM(il.quantity \* il.unit\_price) → total revenue generated.
* Grouped by track, artist, and genre to combine the results into a single table.
* Ranked the results using ORDER BY and displayed the Top 10.

When comparing **revenue vs. volume**, we notice that:

* Some tracks earned high revenue even with fewer copies sold (due to higher price).
* Other tracks achieved popularity through larger sales volume, even if the revenue was slightly lower.

**Insights:**

· **Rock** is the most dominant genre in the USA with the highest number of sales.

· Artists like **Amy Winehouse, AC/DC, System of a Dawn, Cake** are the top artist in the United States with the highest number of sales, who consistently rank among the top performers.

· **War Pigs** and **You Know I’m No Good** are the most popular tracks in the United States with the highest total sales.

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

Query to break down customers country wise :

**SELECT**

**country,**

**COUNT(\*) AS total\_customers**

**FROM customer**

**GROUP BY country**

**ORDER BY total\_customers DESC;**

Output → A single table showing :



Query to break down customers city wise :

**SELECT**

**country,**

**city,**

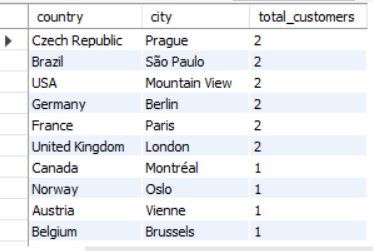
**COUNT(\*) AS total\_customers**

**FROM customer**

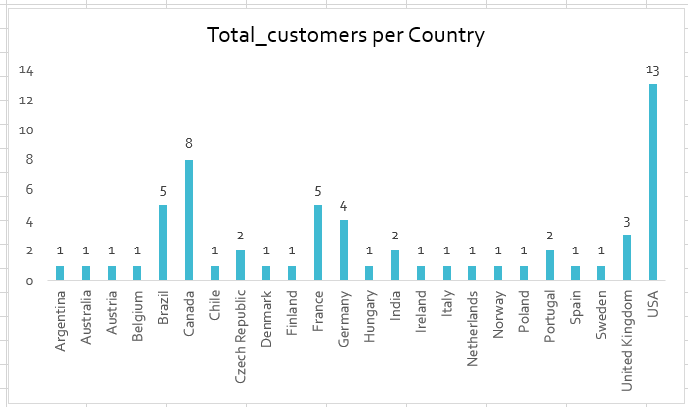
**GROUP BY country, city**

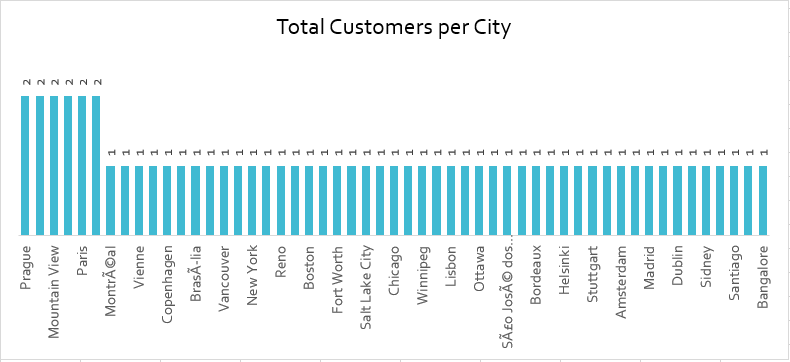
**ORDER BY total\_customers DESC;**

Output → A single table showing :

****

**Visualisation :**

****

****

**Approach:**

The queries were used to retrieve the total number of customers country wise and city wise, using aggregation concept.

**Insights:**

· **USA** has the highest number of total customers, next follows **Canada**

· Some cities like **Berlin**, **Mountain View**, **Paris**  etc. have the highest of two customers each.

NOTE: There was no data for age and gender.

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

Query for total revenue and no. of invoices :

**SELECT**

**billing\_country AS country,**

**billing\_state AS state,**

**billing\_city AS city,**

**COUNT(invoice\_id) AS number\_of\_invoices,**

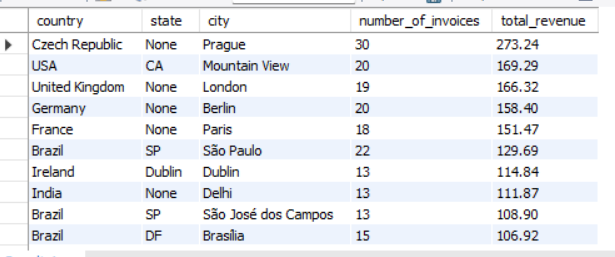
**SUM(total) AS total\_revenue**

**FROM invoice**

**GROUP BY billing\_country, billing\_state, billing\_city**

**ORDER BY total\_revenue DESC;**

Output → A single table showing :



**Approach:**

* I have used the invoice table since it contains both sales (total) and location information.
* Grouped invoices by country, state, and city.
* Calculated:  
  + Total revenue (SUM(total))
  + Number of invoices (COUNT(invoice\_id))
* Sorted results by revenue to highlight the top-performing locations.

This helps identify which regions contribute most to Chinook’s sales.

**Insights :**

* The **Czech Republic** generated the highest revenue (273.24), followed by the **USA** (169.29).
* At the city level, **Prague** and **Mountain View** contributed the most invoices (30 and 20 respectively).
* This suggests that while the USA is a strong market, European cities like Prague are driving significant sales in terms of revenue.

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

Query for top 5 customer by total revenue (country wise) :

**WITH cte1 AS (**

**SELECT**

**c.country,**

**c.first\_name,**

**c.last\_name,**

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

**FROM customer c**

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

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

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

**GROUP BY c.country, c.first\_name, c.last\_name**

**),**

**cte2 AS (**

**SELECT**

**country,**

**first\_name,**

**last\_name,**

**total\_revenue,**

**RANK() OVER (PARTITION BY country ORDER BY total\_revenue DESC) AS rnk**

**FROM cte1**

**)**

**SELECT**

**country,**

**concat(first\_name, ' ', last\_name) as customer\_name,**

**total\_revenue**

**FROM cte2**

**WHERE rnk <= 5**

**ORDER BY country, rnk;**

Output → A single table showing :



**Approach:**

Step 1 – Calculate Total Revenue per Customer

* First, I calculated the total revenue generated by each customer.
* I used INNER JOIN to ensure that only customers who have made **at least one purchase** are considered. Using a LEFT JOIN would include customers with **no purchases**, resulting in zero revenue values.  
  Since the goal was to find the **top revenue-generating customers**, INNER JOIN ensures that only **actual revenue data** is included, keeping the results relevant and clean.
* Revenue is computed by multiplying the unit\_price of each track by the quantity purchased (invoice\_line.quantity).
* I grouped the data by country and customer name to get total revenue for each customer.

Step 2 – Rank Customers within Each Country

* Next, I assigned a rank to each customer within their country, based on their total revenue in descending order.
* I used the SQL RANK() function so that customers with the same revenue receive the same rank.

Step 3 – Select Top 5 Customers per Country

* After ranking, I filtered the customers to include only the **top 5 per country**.
* Finally, I formatted the output to show the country, full customer name, and total revenue.

**Insights :**

* High spending customers exists across all five countries exceeding 60.
* Top customers in Brazil spends the most
* USA showed more balanced Distribution

1. Identify the top-selling track for each customer

Query used to find the top-selling track for each customer was:

**WITH cus\_track\_rank AS (**

**SELECT**

**t.track\_id,**

**t.name AS track\_name,**

**c.customer\_id,**

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

**SUM(il.quantity) AS total\_quantity,**

**ROW\_NUMBER() OVER(PARTITION BY c.customer\_id ORDER BY SUM(il.quantity) DESC) AS row\_rnk**

**FROM track t**

**JOIN invoice\_line AS il**

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

**JOIN invoice AS i**

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

**JOIN customer AS c**

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

**GROUP BY c.customer\_id, customer\_name, t.track\_id, t.name**

**)**

**SELECT customer\_id, customer\_name, track\_id, track\_name, total\_quantity**

**FROM cus\_track\_rank**

**WHERE row\_rnk = 1**

**ORDER BY total\_quantity desc;**

Output → A single table showing :



**Approach:**

Step 1 – Calculate Total Quantity per Track

* For each customer, I used SUM() on quantity purchased for each track.
* This gives the total units bought per track by each customer.

Step 2 – Rank Tracks per Customer

* Use the SQL ROW\_NUMBER() function to rank tracks for each customer based on total quantity purchased in descending order. Also, I used ROW\_NUMBER() because it ensuresunique rank per customer, and INNER JOIN ensures only **customers who bought tracks** are considered, making results relevant.
* The top track gets rank 1.

Step 3 – Select Top Track per Customer

* Filter only the rows with track\_rank = 1 to get the top-selling track for each customer.

**Insights :**

The top selling tracks are Love and Marraige, 24 Caprices Op. 1…, Sting Me, Get What You Need etc.

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

Since my dataset doesn’t have payment methods, so, I only need to focus on :

* Frequency of Purchases (how many times each customer buys)
* Order Value (how much money they spend per order, on average)

Query for calculating Frequency of Purchase :

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,**

**COUNT(i.invoice\_id) AS total\_orders**

**FROM customer AS c**

**LEFT JOIN invoice AS i**

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

**GROUP BY c.customer\_id, customer\_name**

**ORDER BY total\_orders DESC;**

Output → A single table showing :

****

Query for calculating Average Order Value :

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,**

**COUNT(i.invoice\_id) AS total\_orders,**

**ROUND(SUM(i.total),2) AS total\_spent,**

**ROUND(AVG(i.total),2) AS avg\_order\_value**

**FROM customer AS c**

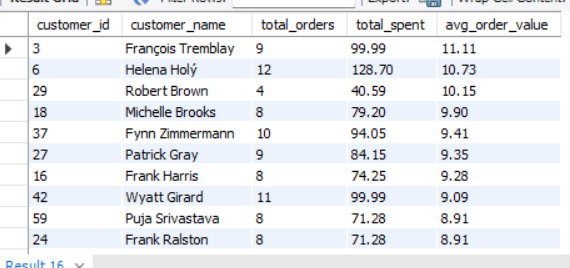
**JOIN invoice AS i**

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

**GROUP BY c.customer\_id, customer\_name**

**ORDER BY avg\_order\_value DESC;**

Output → A single table showing :



**Insights:**

* From the analysis, we can see that customers generally spend between **$9 and $11 per order** on average.
* Some customers, with id like 6, 3,, not only purchase frequently but also spend slightly more per order, making them high-value customers.
* Others, like id 29, purchase less often but maintain a steady order value.
* Overall, this shows that while the average order size is fairly consistent, customer loyalty (number of orders) is the main driver of total revenue.

Note: Since there is no column for payment methods, we cannot identify trends about preferred payment methods

1. What is the customer churn rate?

**Churn Rate** = (Number of customers lost during a period / Number of customers at the start of the period) x 100

Query for calculating Customer Churn Rate :

**WITH last\_purchase AS (**

**SELECT**

**customer\_id,**

**MAX(invoice\_date) AS last\_order\_date**

**FROM invoice**

**GROUP BY customer\_id**

**),**

**churned\_customers AS (**

**SELECT**

**COUNT(customer\_id) AS churned\_count**

**FROM last\_purchase**

**WHERE last\_order\_date < DATE\_SUB('2020-12-30', INTERVAL 6 MONTH)**

**),**

**total\_customers AS (**

**SELECT COUNT(DISTINCT customer\_id) AS total\_count FROM customer**

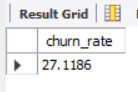
**)**

**SELECT**

**(c.churned\_count / t.total\_count) \* 100 AS churn\_rate**

**FROM churned\_customers c, total\_customers t;**

Output → A single table showing :



**Approach:**

* Identified each customer’s **most recent purchase date**.
* Defined churned customers as those who **haven’t purchased in the last 6 months before Dec 30, 2020**.
* Counted churned customers and divided by the total number of customers to calculate the churn rate.

**Insights :**

* The churn rate is 27.12%, meaning that more than 1 in 4 customers have stopped purchasing in the last 6 months.
* This indicates a **significant portion of customers are becoming inactive**, highlighting the need for customer retention strategies such as targeted promotions, loyalty programs, or personalized engagement.

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

Query for calculating the total Sales by each genre (USA), best-selling genres and artists :

**WITH genre\_sales AS (**

**SELECT**

**g.name AS genre\_name,**

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

**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 g.name**

**),**

**artist\_sales AS (**

**SELECT**

**ar.name AS artist\_name,**

**g.name AS genre\_name,**

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

**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 ar.name, g.name**

**)**

**SELECT**

**gs.genre\_name,**

**ROUND(gs.genre\_revenue \* 100 / (SELECT SUM(genre\_revenue) FROM genre\_sales), 2) AS genre\_percentage,**

**(SELECT artist\_name**

**FROM artist\_sales a**

**WHERE a.genre\_name = gs.genre\_name**

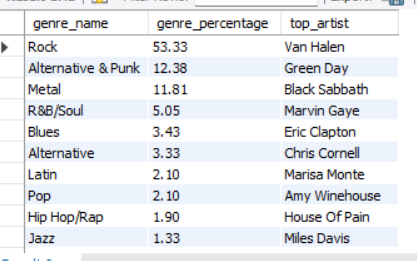
**ORDER BY a.artist\_revenue DESC**

**LIMIT 1) AS top\_artist**

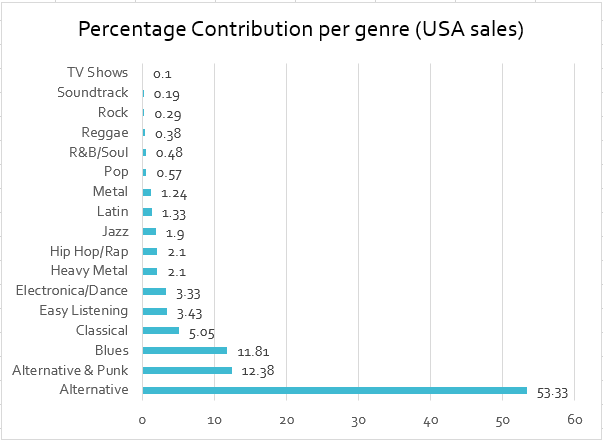
**FROM genre\_sales gs**

**ORDER BY genre\_percentage DESC;**

Output → A single table showing :



**Visualisation** :



**Approach:**

* Filtered data for USA customers only.
* Created one CTE (genre\_sales) to calculate the total revenue for each genre.
* Created another CTE (artist\_sales) to calculate total revenue by artist within each genre.
* Calculated each genre’s percentage contribution to total US sales.
* Identified the top-selling artist for every genre based on highest revenue.

**Insights :**

* Rock contributes to 53% of the entire sales in the USA, followed by Alternative & Punk and Metal which contribute to 12% each.
* Top 5 best - selling artists in the USA are Van Halen, Green Day, Black Sabbath, Marvin Gaye, Eric Calpton.
* Rock is the best-selling genre followed by Alternative & Punk, metal etc..

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

Query for calculating customer’s at least 3 different genre count :

**WITH customer\_genres AS (**

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,**

**COUNT(DISTINCT g.genre\_id) AS genre\_count**

**FROM customer c**

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

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

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

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

**GROUP BY c.customer\_id, customer\_name**

**)**

**SELECT**

**customer\_id,**

**customer\_name,**

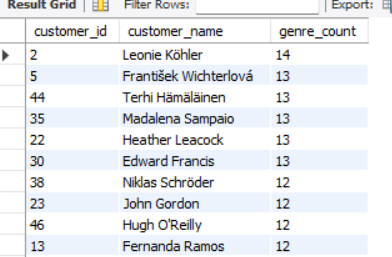
**genre\_count**

**FROM customer\_genres**

**WHERE genre\_count >= 3**

**ORDER BY genre\_count DESC;**

Output → A single table showing :



**Approach:**

* Joined customers with their invoices, invoice lines, and tracks to capture purchases.
* Counted the number of distinct genres purchased by each customer.
* Selected only those customers who bought tracks from 3 or more different genres.

**Insights :**

* A segment of customers shows diverse music preferences, purchasing from 3+ genres.
* These customers may represent loyal and exploratory buyers, making them valuable for cross-selling and promotional campaigns.

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

Query for calculating customer’s at least 3 different genre count :

**SELECT**

**g.name AS genre\_name,**

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

**RANK() OVER (ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS sales\_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**

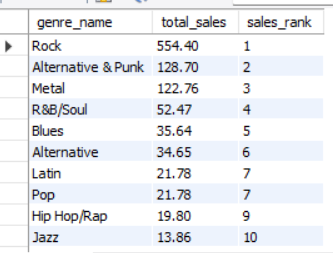
**JOIN genre g ON t.genre\_id = g.genre\_id**

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

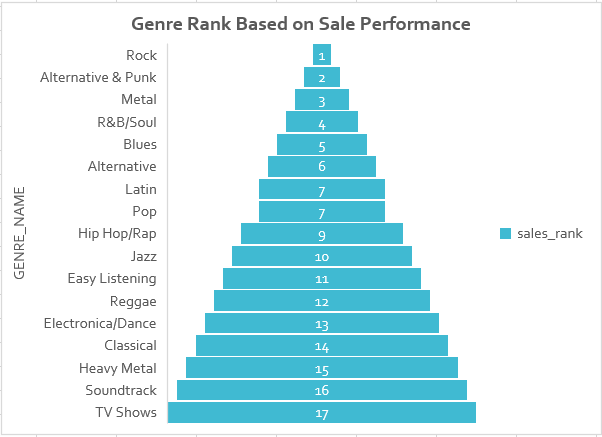
**GROUP BY g.name**

**ORDER BY total\_sales DESC;**

Output → A single table showing :

****

**Visualisation :**

****

**Insights :**

* Rock, Metal and Alternative &Punk are the top three genres in the United States

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

Query for calculating customers who have not made a purchase in the last 3 months :

**WITH inactive\_customers AS (**

**SELECT**

**c.customer\_id,**

**c.first\_name,**

**c.last\_name,**

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

**SUM(i.total) AS total\_revenue**

**FROM customer c**

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

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

**HAVING MAX(i.invoice\_date) < DATE\_SUB(CURDATE(), INTERVAL 3 MONTH)**

**)**

**SELECT \***

**FROM inactive\_customers;**

Output → A single table showing :



· **Insights :**

* The customers in the output table above have not made a purchase in the past 3 months.

Subjective Questions

1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.

Query :

**WITH RecommendedAlbums AS (**

**SELECT**

**al.title AS album\_name,**

**a.name AS artist\_name,**

**g.name AS genre\_name,**

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

**SUM(il.quantity) AS total\_quantity,**

**DENSE\_RANK() OVER(ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS sales\_rank**

**FROM customer c**

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

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

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

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

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

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

**WHERE c.country = 'USA'**

**GROUP BY al.title, a.name, g.name**

**)**

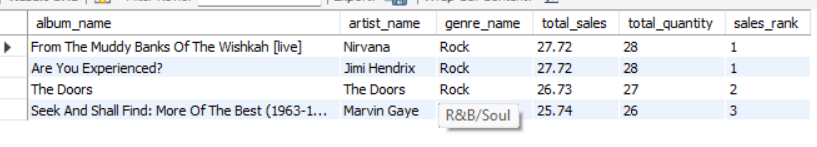
**SELECT \***

**FROM RecommendedAlbums**

**WHERE sales\_rank <= 3**

**ORDER BY total\_sales DESC;**

Output → A single table showing :



**Tables Used** : Customer , invoice , invoice\_line , track , album, artist,genre

**Concept** : CTE , Joins , Aggregation, Group BY , Order By

**Insights :**

* · Albums to Prioritize for promotion based on unit sales , revenue

### **Recommended Albums for Promotion in the USA :**

Based on the genre sales analysis in the USA, the following albums are recommended for prioritization in advertising and promotion due to their strong unit sales and revenue performance:

1. *From The Muddy Banks Of The Wishkah [Live]* and *Are You Experienced?*
2. *The Doors*
3. *Seek And Shall Find: More Of The Best (1963-1981)*

**Insights :**

* All three albums belong to the **Rock genre**, which is the top-selling genre in the USA. For marketing strategies, these albums should be the focus, leveraging both digital and physical channels.
* Also, the associated artists — **Nirvana, Jimi Hendrix, The Doors, and Marvin Gaye** — have a longstanding fanbase. This makes them ideal candidates for **physical formats such as CDs**, as well as targeted promotions to their existing audience, ensuring higher engagement and sales potential.

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

Query :

**SELECT**

**i.billing\_country,**

**g.name AS genre\_name,**

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

**FROM invoice\_line il**

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

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

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

**WHERE i.billing\_country != 'USA'**

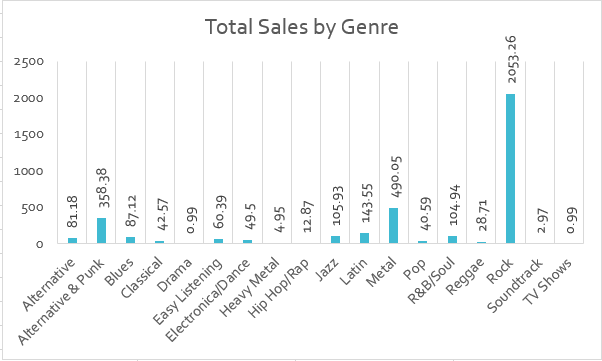
**GROUP BY i.billing\_country, g.name**

**ORDER BY i.billing\_country, total\_sales DESC;**

Output → A single table showing :



**Visualisation** :



**Insights :**

Similarities:

* The top 3 genres in the USA and outside the USA are Rock, Metal, and Alternative & Punk.
* Rock is the most popular genre in both regions, with a clear lead in sales.
* Several genres appear in the top 10 in both regions, such as Latin, Jazz, R&B/Soul, and Blues.

Differences:

* Metal ranks 2nd outside the USA, while Alternative & Punk ranks 2nd in the USA.
* R&B/Soul and Blues sell more in the USA, while Latin and Jazz sell more outside the USA.
* Hip Hop is in the top 10 in the USA but not outside the USA.

**Recommendation :**

* Promote Rock, Metal, and Alternative & Punk albums and tracks both in the USA and internationally.
* Outside the USA, give extra focus to Latin and Jazz genres for advertising and promotion.

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?

Query :

**WITH CustomerMetrics AS (**

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,**

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

**MAX(DATE(i.invoice\_date)) AS last\_purchase\_date,**

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

**ROUND(AVG(il.quantity), 0) AS avg\_basket\_size,**

**ROUND(AVG(i.total), 2) AS avg\_spending\_amount,**

**DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) AS tenure\_days,**

**CASE**

**WHEN DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) > 1000 THEN 'Long Term'**

**ELSE 'New'**

**END AS customer\_type**

**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, customer\_name**

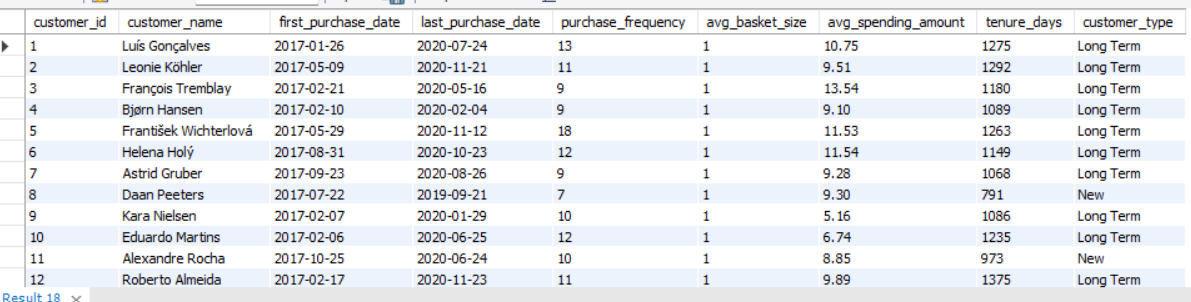
**)**

**SELECT \***

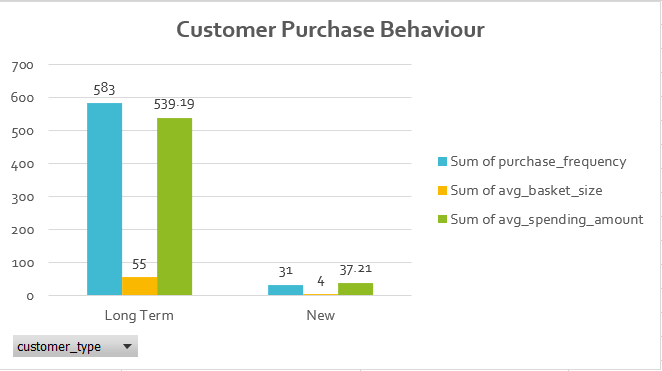
**FROM CustomerMetrics**

**ORDER BY customer\_id;**

Output → A single table showing :



**Visualisation** :



**Tables Used** : Customer, Invoice, Invoice\_line

**Approach** :

* I calculate first and last purchase dates per customer to determine tenure.
* Then compute purchase frequency, average basket size, and average spending for each customer.
* Classified customers into Long Term and New based on tenure.
* Compared the metrics for both customer types to analyze purchasing behavior and loyalty.

## **Insights :**

* Long-term customers buy more frequently and spend more overall.
* New customers have lower purchase frequency and smaller spending amounts.
* Most customers have a basket size of 1 (likely buying single tracks per invoice).
* Long-term customers show strong loyalty, while new customers are still building their purchase habits.

## **Recommendation :**

* Reward long-term customers with loyalty programs, discounts, or early access to new releases.
* Encourage new customers to buy more frequently through personalized offers and promotions.
* Consider bundle deals or cross-selling to increase basket size.
* Monitor customer tenure to identify at-risk customers and run retention campaigns.

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?

Query :

**SELECT**

**g1.name AS genre\_1,**

**g2.name AS genre\_2,**

**COUNT(\*) AS times\_bought\_together**

**FROM invoice\_line AS il1**

**JOIN invoice\_line AS il2**

**ON il1.invoice\_id = il2.invoice\_id**

**AND il1.track\_id <> il2.track\_id**

**JOIN track AS t1**

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

**JOIN track AS t2**

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

**JOIN genre AS g1**

**ON t1.genre\_id = g1.genre\_id**

**JOIN genre AS g2**

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

**GROUP BY genre\_1, genre\_2**

**ORDER BY times\_bought\_together DESC**

**LIMIT 10;**

**-- Artist Co-Purchases**

**SELECT**

**ar1.name AS artist\_1,**

**ar2.name AS artist\_2,**

**COUNT(\*) AS times\_bought\_together**

**FROM invoice\_line AS il1**

**JOIN invoice\_line AS il2**

**ON il1.invoice\_id = il2.invoice\_id**

**AND il1.track\_id <> il2.track\_id**

**JOIN track AS t1**

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

**JOIN track AS t2**

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

**JOIN album AS al1**

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

**JOIN album AS al2**

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

**JOIN artist AS ar1**

**ON al1.artist\_id = ar1.artist\_id**

**JOIN artist AS ar2**

**ON al2.artist\_id = ar2.artist\_id**

**GROUP BY artist\_1, artist\_2**

**ORDER BY times\_bought\_together DESC**

**LIMIT 10;**

**-- Album Co-Purchases**

**SELECT**

**al1.title AS album\_1,**

**al2.title AS album\_2,**

**COUNT(\*) AS times\_bought\_together**

**FROM invoice\_line AS il1**

**JOIN invoice\_line AS il2**

**ON il1.invoice\_id = il2.invoice\_id**

**AND il1.track\_id <> il2.track\_id**

**JOIN track AS t1**

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

**JOIN track AS t2**

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

**JOIN album AS al1**

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

**JOIN album AS al2**

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

**GROUP BY album\_1, album\_2**

**ORDER BY times\_bought\_together DESC**

**LIMIT 10;**

Output → A single table showing : (Genre Co-Purchases)

****

(Artist Co-Purchases)



(Album Co-Purchases)



**Tables Used** :

* invoice\_line → contains details of purchased tracks.
* track → maps each track to its album and genre.
* album → links tracks to albums.
* artist → links albums to artists.
* genre → categorizes tracks into genres.

**Approach** :

Identify items purchased together by checking multiple tracks in the same invoice.

Analyze affinity at three levels:

* **Genre → Genre** (which genres co-occur in purchases)
* **Artist → Artist** (which artists are purchased together)
* **Album → Album** (which albums are purchased together)
* I used COUNT(\*) to calculate how many times pairs of genres/artists/albums were bought together.
* Then use ORDER BY times\_bought\_together DESC LIMIT 10 to bring back the top results.

### **Insights :**

* **Rock with Rock** is the most frequent co-purchase, far ahead of other combinations.
* **Metal with Metal** and **Alternative & Punk with Alternative & Punk** also appear but with lower frequency.
* Rock is often paired with **Alternative & Punk** and **Metal**, showing cross-genre overlap.
* For artists and albums, the top co-purchases are mostly **within the same artist or same album**.

### **Recommendations :**

* Focus cross-selling on **Rock**, offering bundles, discounts, and recommendations of new Rock tracks/artists.
* For **Metal** and **Alternative & Punk**, suggest more tracks in the same genre and incentivize with offers.
* Use Rock’s overlap with **Alternative & Punk** and **Metal** to design **cross-genre promotions** (e.g., Rock + Punk bundles).
* If a customer buys multiple tracks from the **same artist or album**, recommend the **artist’s other albums** or upsell the **entire album**.

1. Regional Market Analysis: Do customer purchasing behaviors and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

Query :

**WITH last\_purchase AS (**

**SELECT**

**customer\_id,**

**billing\_country AS region,**

**MAX(invoice\_date) AS last\_order\_date**

**FROM invoice**

**GROUP BY customer\_id, billing\_country**

**),**

**dataset\_end AS (**

**SELECT MAX(invoice\_date) AS max\_date FROM invoice**

**)**

**SELECT**

**l.region,**

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

**SUM(CASE WHEN l.last\_order\_date < DATE\_SUB(d.max\_date, INTERVAL 6 MONTH) THEN 1 ELSE 0 END) AS churned\_customers,**

**ROUND(**

**SUM(CASE WHEN l.last\_order\_date < DATE\_SUB(d.max\_date, INTERVAL 6 MONTH) THEN 1 ELSE 0 END) \* 100.0**

**/ COUNT(DISTINCT l.customer\_id),**

**2) AS churn\_rate**

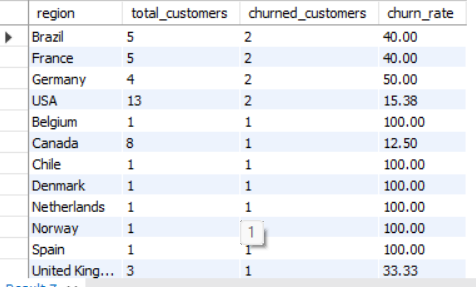
**FROM last\_purchase l**

**CROSS JOIN dataset\_end d**

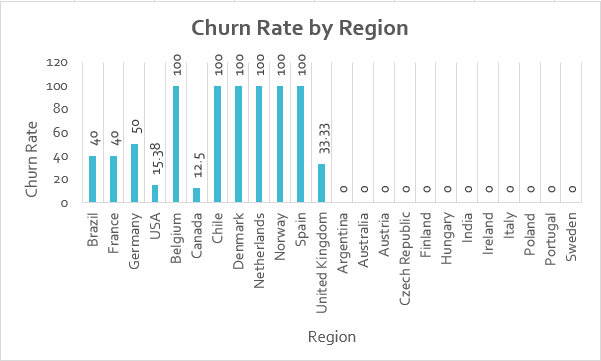
**GROUP BY l.region**

**ORDER BY churned\_customers DESC;**

Output → A single table showing :



**Visualisation** :



### **Approach :**

We analyzed customer behavior by country using two queries:

* To calculate total customers, orders, revenue, average order value, and purchase frequency.
* To calculate churn rate (customers who haven’t purchased in the last 6 months).

### **Insights :**

* USA has the highest number of customers, orders, and revenue. Rock music’s popularity could be a big reason.
* Canada comes second in customers, orders, and revenue. English language and developed economy may explain this.
* Czech Republic has the highest average order value (9.11) and purchase frequency (15 orders per customer). This suggests people there spend more on music.
* Countries like Spain, Norway, and Netherlands show a 100% churn rate, but each has only 1 customer, so this isn’t meaningful.
* Among big markets:  
  + **USA churn rate = 15.38%**
  + **Canada churn rate = 12.50%**
* **Germany** shows a churn rate of 50% (2 of 4 customers stopped buying).
* **France and Brazil** each have a churn rate of 40% (2 of 5 customers churned). This may be due to less interest in English-language music.

Note: Churned customers = customers whose last purchase was more than 6 months before the latest purchase date in the dataset.

### **Recommendations :**

* For countries with **only 1 customer**, efforts should focus on **customer acquisition** (e.g., referral programs, promotions).
* For countries with **high churn rates** (Germany, Brazil, France), more effort is needed to **retain existing customers**, such as targeted offers, local-language marketing, or promoting genres popular in those regions.
* In **strong markets like USA and Canada**, keep building on the existing fanbase by cross-selling Rock and related genres.

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?

Query :

**WITH customer\_summary AS (**

**SELECT**

**c.customer\_id,**

**c.country,**

**COUNT(i.invoice\_id) AS total\_orders,**

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

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

**DATEDIFF(CURRENT\_DATE, MAX(i.invoice\_date)) AS days\_since\_last\_purchase**

**FROM customer c**

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

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

**GROUP BY c.customer\_id, c.country**

**),**

**customer\_risk AS (**

**SELECT**

**customer\_id,**

**country,**

**total\_orders,**

**total\_spent,**

**last\_purchase\_date,**

**days\_since\_last\_purchase,**

**CASE**

**WHEN days\_since\_last\_purchase > 180 THEN 'High Risk'**

**WHEN days\_since\_last\_purchase BETWEEN 90 AND 180 THEN 'Medium Risk'**

**ELSE 'Low Risk'**

**END AS churn\_risk,**

**CASE**

**WHEN total\_spent < 100 THEN 'Low Value'**

**WHEN total\_spent BETWEEN 100 AND 500 THEN 'Medium Value'**

**ELSE 'High Value'**

**END AS value\_segment**

**FROM customer\_summary**

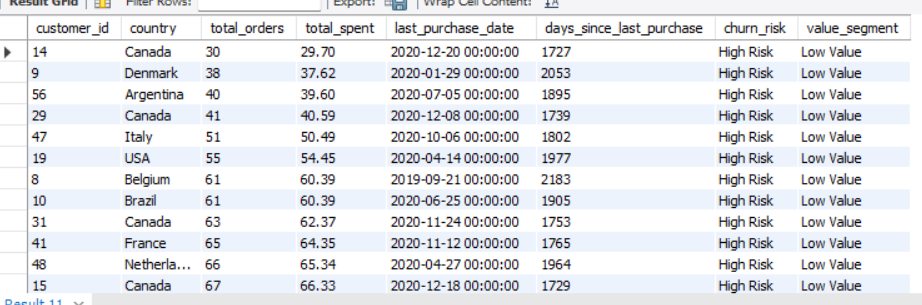
**)**

**SELECT \***

**FROM customer\_risk**

**ORDER BY churn\_risk DESC, total\_spent ASC;**

Output → A single table showing :



#### **Tables Used :**

* customer – Contains customer information including customer\_id, country, and basic demographics.
* invoice – Contains invoice information including invoice\_id, customer\_id, and invoice\_date.
* **invoice\_line** – Contains details of purchased items including invoice\_id, track\_id, unit\_price, and quantity.

#### **Approach :**

* Data Aggregation:  
  + Calculated total orders (COUNT(invoice\_id)), total spending (SUM(unit\_price \* quantity)), last purchase date (MAX(invoice\_date)), and days since last purchase (DATEDIFF(CURRENT\_DATE, last\_purchase\_date)) for each customer.
* Risk Segmentation:  
  + Customers were segmented into High Risk, Medium Risk, and Low Risk based on inactivity (days\_since\_last\_purchase).
  + Customers were also segmented into Value Segments (Low, Medium, High) based on total spending.
* Final Output:  
  + Combined risk and value segments to identify customers likely to churn or provide low revenue.

#### **Insights :**

* Majority of customers are classified as **High Risk** due to long inactivity periods (most last purchases over 1700 days ago).
* Even customers with frequent orders have **Low Value**, indicating low spending per order.
* Countries like Canada, Denmark, Argentina, Italy, USA, and Belgium have several high-risk, low-value customers.

#### **Recommendations :**

* Customer Retention Campaigns:  
  + Target High Risk / Low Value customers with special promotions, discounts, or engagement emails.
* Loyalty Programs:  
  + Encourage repeat purchases with reward points or exclusive deals for high-value customers.
* Further Analysis:  
  + Investigate reasons for low spending despite frequent purchases (e.g., pricing, product offerings).
  + If possible, enrich data with demographic info to improve profiling.

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?

Query :

**WITH CustomerMetrics AS (**

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) AS customer\_name,**

**c.country,**

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

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

**DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) + 1 AS tenure\_days,**

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

**SUM(i.total) AS total\_spent,**

**ROUND(SUM(i.total) / COUNT(i.invoice\_id), 2) AS avg\_order\_value,**

**DATEDIFF(CURRENT\_DATE, MAX(i.invoice\_date)) AS days\_since\_last\_purchase**

**FROM customer c**

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

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

**)**

**SELECT**

**customer\_id,**

**customer\_name,**

**country,**

**tenure\_days,**

**purchase\_frequency,**

**total\_spent,**

**avg\_order\_value,**

**days\_since\_last\_purchase,**

**CASE**

**WHEN days\_since\_last\_purchase > 180 THEN 'Churned'**

**WHEN days\_since\_last\_purchase > 30 THEN 'At-Risk'**

**ELSE 'Active'**

**END AS customer\_status,**

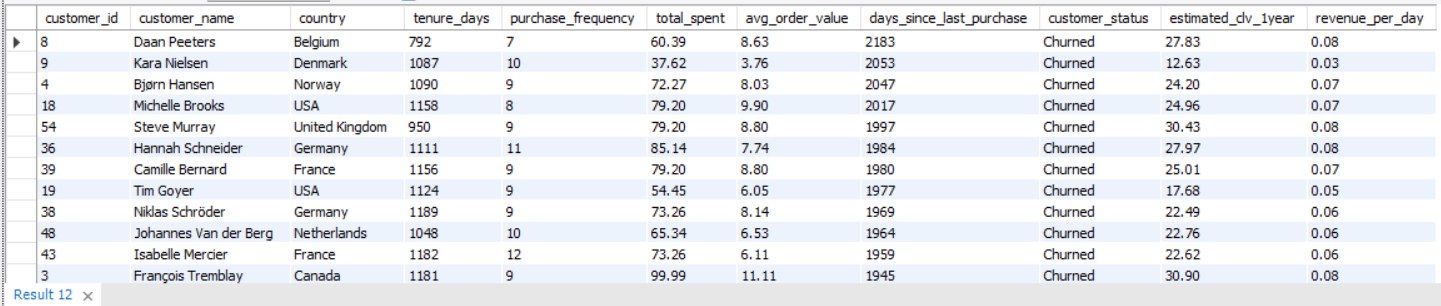
**ROUND((total\_spent / tenure\_days) \* 365, 2) AS estimated\_clv\_1year,**

**ROUND(total\_spent / tenure\_days, 2) AS revenue\_per\_day**

**FROM CustomerMetrics**

**ORDER BY days\_since\_last\_purchase DESC;**

Output → A single table showing :



#### **Approach :**

Step-by-step approach to answer the question :

1. Extract customer-level metrics :  
   * Calculated each customer’s total spend, purchase frequency, average order value, tenure, and recency (days since last purchase).
   * This forms the basis for CLV estimation and churn analysis.
2. Segment customers based on behavior :  
   * Active: purchased within last 30 days
   * At-Risk: purchased 31–180 days ago
   * Churned: no purchase in last 180+ days
3. Estimate Customer Lifetime Value (CLV) :  
   * Using historical purchase behavior, computed revenue per day and projected estimated CLV for 1 year.
4. Analyze patterns among churned customers :  
   * Checked demographics (country), purchase frequency, and tenure to identify common characteristics.

#### **Tables Used :**

#### Customer table: Provides demographic and unique customer identifiers.

#### Invoice table: Contains all purchase transactions needed to compute spend, frequency, and tenure.

#### Together, these tables allow us to measure customer behavior and estimate lifetime value.

#### **Insights :**

* + Customers with **high total spend and frequent purchases** have the **highest estimated CLV** → these are high-value customers.
  + Customers flagged as **Churned** tend to:
* Have **long gaps since last purchase** (high recency)
* Often come from specific countries (geographical pattern)
* Show **low purchase frequency** and/or **short tenure**
  + Customers with **short tenure but high purchases** represent an opportunity to **convert into loyal customers** through engagement campaigns.

#### **Recommendation** :

#### Target high-value customers :

#### Offer loyalty programs, VIP perks, or early access to encourage continued purchases.

#### Re-engage churned or at-risk customers :

#### Send promotional offers, discounts, or targeted campaigns.

#### Focus on countries or segments identified as high-risk in churn analysis.

#### Promote retention among promising short-tenure customers :

#### Identify customers with high purchase activity early and engage them to boost loyalty.

#### Use product/genre insights (optional next step) :

#### Recommend cross-selling or personalized promotions based on prior purchases.

#### 

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?

To understand how promotions like **discounts**, **events**, or **email campaigns** affect a business, I can focus on three main areas: attracting new customers, keeping existing ones, and increasing sales.

* Customer Acquisition : We can check how many new customers made their first purchase during the promotion and compare it to the usual number of new customers. This shows whether the promotion helped bring in more people to the business.
* Customer Retention : We can also look at existing customers to see if they bought more often, made bigger purchases, or stayed active for a longer time after the promotion. This tells us whether the promotion helped keep current customers engaged.
* Sales Impact : We can measure total sales, average order value, and extra revenue during the promotion period. We can also track how many people used discount codes or participated in events, which helps us see how effective the promotion was.

By analyzing these areas, businesses can understand which promotions work best, decide where to invest marketing efforts, and create better strategies to bring in new customers and retain loyal ones.

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

### My approach if no specific questions are given

If there were no specific questions provided, I would approach the problem in the following steps:

**Step 1: Understand the business and the data**

* Identify the main goals of the company, such as **growing revenue, retaining customers, and increasing engagement**.
* Review the available data, including customer details, purchase history, product information, and regional data.

**Step 2: Explore and clean the data**

* Handle **missing values or duplicates** to ensure accuracy.
* Explore customer behavior, purchase patterns, and product trends to understand how people are interacting with the business.

**Step 3: Identify key metrics**

* Define important metrics like:  
  + **Top genres, artists, and albums** by sales or popularity
  + **Customer behavior**: churn rate, acquisition rate, lifetime value, average order value
  + **Regional or geographical trends** in purchases

**Step 4: Analyze the data**

I would have also developed various business questions like: -

* Which are locations from where maximum sale is coming?
* Which tracks are the best-selling in each location?
* Which are the most popular artists and albums in different regions?
* How is sales distributed among different countries?
* Which genre is most liked in each country?

And also look over the Trends in the purchases of tracks like: -

* What is the total revenue generated over a period of time?
* How is the invoice total of each customer changed with time?
* What is the trend of a particular genre in a particular country?

Then I would have summarized all the insights and presented the data.

* I’d use SQL, Excel, or other tools to analyze **sales by genre, artist, country, and customer segment**.
* Perform **affinity analysis** to find patterns in purchases for **cross-selling, bundle deals,** or identify high-performing areas as well as underperforming ones.

**Step 5: Promote engagement and frequent purchases**

* Encourage customers to buy more often by **reducing time between purchases** using incentives like point systems, discounts, or friendly engagement messages.
* Suggest promotions for the **most popular genres in each region**, and recommend tracks and albums that are likely to sell well.
* Use **cross-selling** strategies and bundle offers to increase average order value and total purchases.

**Step 6: Retain and win back customers**

* Offer **exclusive deals and discounts** to churned customers.
* Run **retention campaigns** for at-risk customers.
* Maintain **loyalty programs** for active and high-value customers.

**Step 7: Visualize and recommend strategies**

* Create charts showing top genres by revenue, churn rates by region, or average spend per customer.
* Based on insights, propose actions like:  
  + Promoting top genres region-wise
  + Offering discounts or bundles to boost sales
  + Running targeted campaigns to retain or win back customers
  + Using loyalty programs to reward active buyers

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?

Query :

**ALTER TABLE album**

**ADD COLUMN release\_year INT(4);**

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 an SQL query to provide this information.

Query :

**WITH CustomerPurchases AS (**

**SELECT**

**c.customer\_id,**

**c.country,**

**SUM(i.total) AS total\_spent,**

**SUM(il.quantity) AS total\_tracks**

**FROM customer c**

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

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

**GROUP BY c.customer\_id, c.country**

**)**

**SELECT**

**country,**

**COUNT(customer\_id) AS num\_customers,**

**ROUND(AVG(total\_spent), 2) AS avg\_total\_spent,**

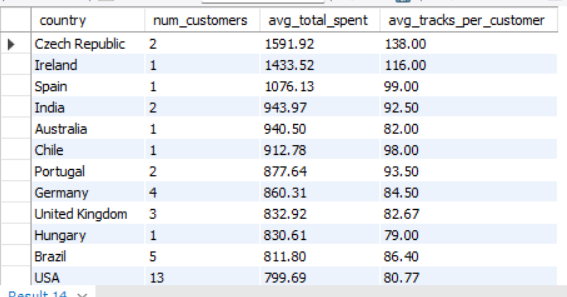
**ROUND(AVG(total\_tracks), 2) AS avg\_tracks\_per\_customer**

**FROM CustomerPurchases**

**GROUP BY country**

**ORDER BY avg\_total\_spent DESC;**

Output → A single table showing :



**Tables Used** : Customer, Invoice, Invoice\_line

## **Approach :**

* Understand the objective :  
  + Analyze purchasing behavior of customers based on geographical location.
  + Calculate metrics like average total spent, number of customers, and average tracks purchased per customer.
* Aggregate customer-level data :  
  + For each customer, calculate total amount spent and total number of tracks purchased.
* Aggregate country-level data :  
  + For each country, calculate:  
    - Total number of customers
    - Average total spent per customer
    - Average tracks purchased per customer
* Sort and analyze :  
  + Sort countries by average spending to identify high-value regions.

#### **Insights :**

* Customers in Czech Republic, Ireland, and Spain have the highest average spending per customer.
* The largest portion of revenue comes from the USA.
* India and Brazil are showing growth in customer purchases.
* Customers in Argentina, Denmark, and Italy spend the least on average.

**Recommendations** :

* Offer premium deals or early access to top-spending customers to reward loyalty.
* Maintain loyalty programs in the USA to keep high-value customers engaged.
* Run targeted local promotions in growing markets like India and Brazil.
* Reach out to low-spending countries (Argentina, Denmark, Italy) with special offers or surveys to understand customer needs and improve engagement.

Documented By: Tuhina Ghosh

THE END