Objective Questions

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

There are two primary ways to replace NULL values with specific values.

**Using the COALESCE() Function**  
The COALESCE() function helps handle NULL values by providing a default value.

Its syntax is: COALESCE(column\_name, 'default\_value')

**Using the UPDATE Statement**  
The UPDATE statement allows us to modify table values based on a specific condition.

The syntax is: UPDATE table\_name SET column1 = value1 WHERE condition;

We have NULL values in various columns across different tables, such as:

* **Customer table: company, state, phone, fax**
* **Track table: composer**

To handle NULL values:

* **Numeric fields can be replaced with 0.**
* **Text fields can be replaced with placeholders like 'None', 'Unknown', etc.**

**Example:**  
If the composer column in the Track table contains NULL values, they can be replaced with 'Unknown'. Similarly, other NULL columns can be handled using appropriate default values.

|  |  |  |
| --- | --- | --- |
| **Tables with NULL values** | **Column Name / Attribute** | **Value to be replaced** |
| Customer | Company | ‘Unknown’ |
| State | ‘None’ |
| Phone | ‘+0 000 000 0000’ |
| Fax | ‘+0 000 000 0000’ |
| Track | Composer | ‘Unknown’ |

|  |  |  |
| --- | --- | --- |
| Sr.No | **Using COALESCE** | **Using UPDATE** |
| 1 | SELECT COALESCE(company,’Unknown’) FROM customers  WHERE company IS NULL; | UPDATE customer  SET company = 'Unknown' WHERE company IS NULL; -- 49 row(s) affected |
| 2 | SELECT COALESCE(state,’None’)  FROM customers  WHERE company IS NULL; | UPDATE customer SET state = 'None' WHERE state IS NULL;  -- 29 row(s) affected |
| 3 | SELECT COALESCE(phone, '+0 000 000 0000')  FROM customers WHERE phone IS NULL; | UPDATE customer  SET phone = '+0 000 000 0000' WHERE phone IS NULL;  -- 1 row(s) affected |

|  |  |  |
| --- | --- | --- |
| 4 | SELECT COALESCE(fax, '+0 000 000 0000')  FROM customers WHERE fax IS NULL; | UPDATE customer  SET fax = '+0 000 000 0000' WHERE fax IS NULL;  -- 47 row(s) affected |
| 5 | SELECT COALESCE(company,’Unknown’) FROM customers  WHERE company IS NULL; | UPDATE track  SET composer = 'Unknown' WHERE composer IS NULL;  -- 978 row(s) affected |

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

**Solution:**

**Key Concepts Used:** Common Table Expressions (CTE), Aggregate Functions, GROUP BY, Joins, Sorting (ORDER BY).

**Tables Referenced:** invoice, invoice\_line, track, album, artist, genre.

**Query:**

WITH topSellingTracknArtist AS (

SELECT t.name AS track\_name, a.name AS artist\_name, g.name AS genre\_name, SUM(i.total) AS total\_sales,

RANK() OVER(ORDER BY SUM(i.total) 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 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 i.billing\_country = 'USA'

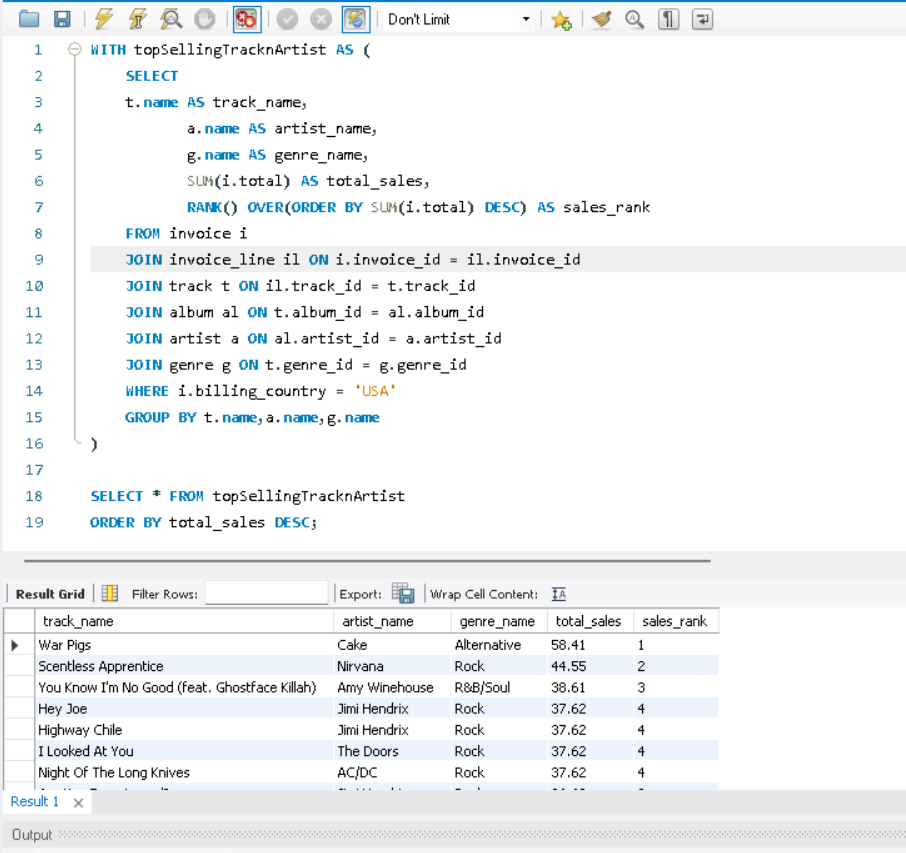
GROUP BY t.name,a.name,g.name

)

SELECT \* FROM topSellingTracknArtist

ORDER BY total\_sales DESC;

**Result: (783 rows returned as output)**



**Explanation**

Step 1: TopTracks CTE

This part calculates the **top-selling track**, **top artist**, and **top genre** by **total sales** (unit price \* quantity) from the track, invoice\_line, invoice, album, artist, and genre tables.

Explanation:

* **JOINs**: Combines data from multiple tables: track, invoice\_line, invoice, album, artist, and genre.
  + invoice\_line gives us the quantity of each track sold.
  + invoice is used to filter the records by billing\_country = 'USA'.
  + album, artist, and genre provide the names of the album, artist, and genre respectively.
* **SUM(il.quantity \* t.unit\_price)**: Calculates the total sales for each track by multiplying the quantity by the unit price of each track.
* **GROUP BY**: Groups the results by track name, artist name, and genre name to compute total sales for each combination.
* **WHERE**: Filters records to only include sales in the USA (i.billing\_country = 'USA').

Step 2: TopGenres CTE

This part calculates the top genre based on the total quantity sold in the USA.

Explanation:

* **JOINs**: Combines the track, invoice\_line, invoice, and genre tables.
  + invoice\_line provides the quantity sold for each track.
  + invoice filters the records to those in the USA.
  + genre provides the genre names.
* **SUM(il.quantity)**: Sums the quantity of each track sold, which represents the popularity of each genre.
* **GROUP BY**: Groups the results by genre name to calculate total quantities sold for each genre.

Step 3: Final Selection

Once the CTEs are defined, the main queries select the top-selling track, artist, and genre.

Explanation

* **ORDER BY TotalSales DESC**: Orders the result by TotalSales in descending order to get the highest-selling track.
* **LIMIT 10**: Restricts the result to only the top 10 -selling track, artist, and genre.

**Result:**

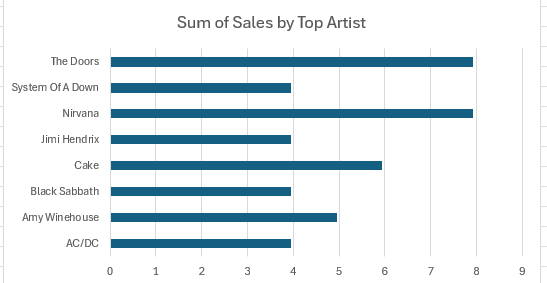
1. **Top Selling Track, Artist, and Genre**: The first query gives us the top-selling track, artist, and genre by total sales in the USA.
2. **Top Genre**: The second query returns the most famous genre by the total quantity sold in the USA.

**Summary:**

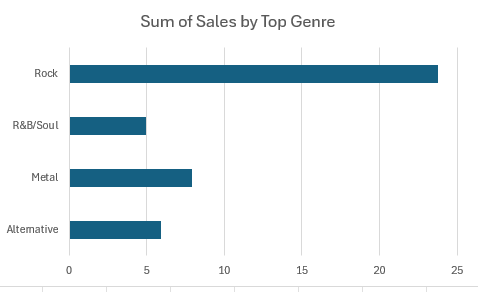
* The **CTEs** break down the problem into manageable pieces: one for finding the top-selling track/artist/genre and another for finding the most famous genre.
* The use of JOIN allows for combining multiple related tables and aggregating sales/quantities across tracks, genres, and artists.

**Visualization -**

**Sum of sales by top artist:**

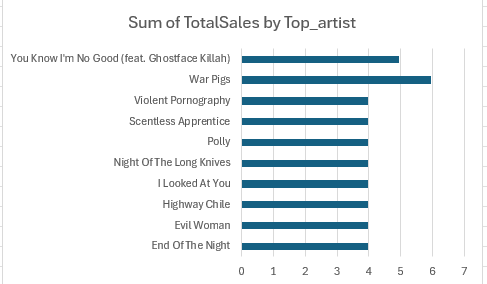


**Sum of Sales by genre:**



.

**Sum of Sales by to selling track:**



**Output:**

****

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

**Solution:**

* Concepts used: Aggregate Functions, GROUP BY, Sorting(ORDER BY)
* Tables used: customer

**Query:**

SELECT

country,

COALESCE(state,'None') AS state,

city,

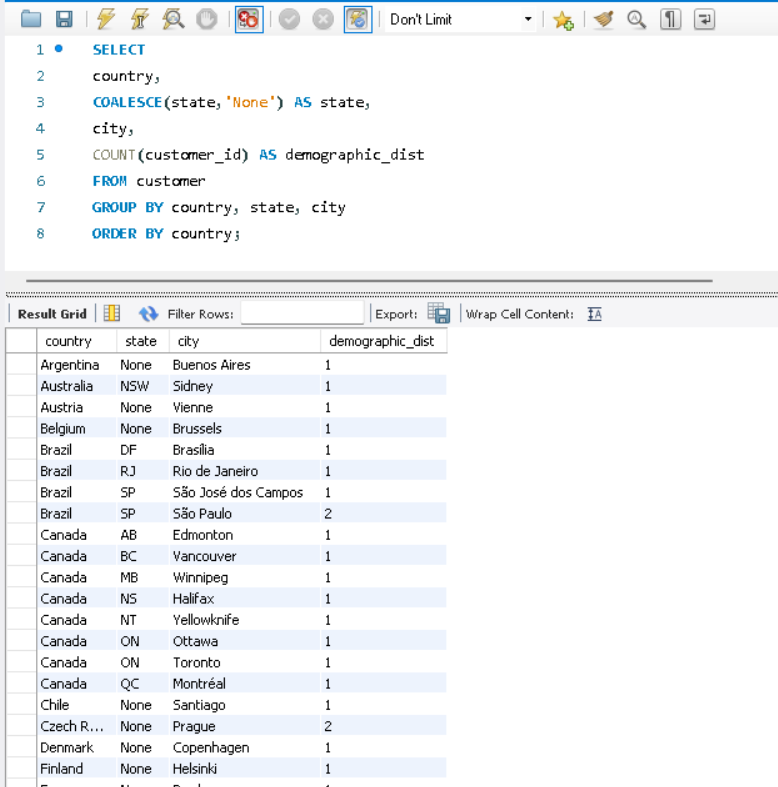
COUNT(customer\_id) AS demographic\_dist

FROM customer

GROUP BY country, state, city

ORDER BY country;

**Result: (53 Rows Returned)**

****

**Explanation**:

This query provides a demographic breakdown of customers based on their location (country) from the Chinook database's `Customer` table. Below is a detailed explanation:

1. SELECT Distinct country, COUNT(\*) AS Count\_of\_customers\*\*

`DISTINCT country` ensures that only unique countries are considered in the result. `COUNT(\*) AS Count\_of\_customers` counts the total number of customers for each country. The result is labeled as `Count\_of\_customers` for clarity.

2. FROM Customer

The query retrieves data from the `Customer` table, which contains customer-related information.

3. GROUP BY country

Groups all rows in the `Customer` table by the unique values in the `country` column. Within each group, the `COUNT(\*)` function calculates the number of customers from that specific country.

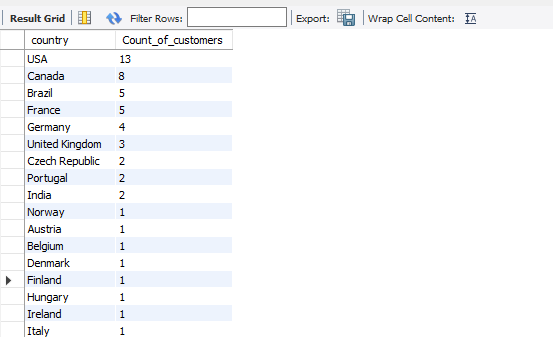
4. ORDER BY COUNT(\*) DESC

Sorts the results in descending order based on the number of customers in each country. Countries with the highest number of customers appear at the top of the result.

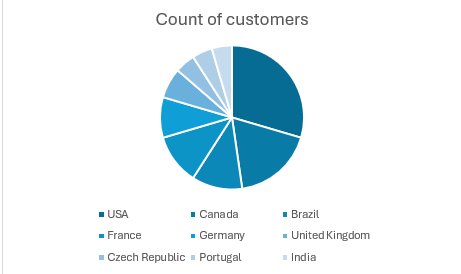
The query provides the geographic breakdown of Chinook's customer base by counting how many customers are from each country. It highlights which countries have the most customers, helping to identify key markets or regions where Chinook's business is strongest.

This query focuses only on location (country) as the demographic factor. Additional demographic attributes like age or gender are not included in the `Customer` table by default in Chinook and would require joining with other data if available.

**Output**



**Visualization-**

****

**Analysis-**

* + USA and Canada have the largest share in the customer base of Chinook
  + Only 10 countries have more than 1 customer and the rest have only 1

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

**Solution:**

* Concepts used: Aggregate Functions, GROUP BY, Sorting (ORDER BY)
* Table used: invoice

**Query:**

SELECT

billing\_country,

billing\_state,

billing\_city,

SUM(total) AS total\_revenue,

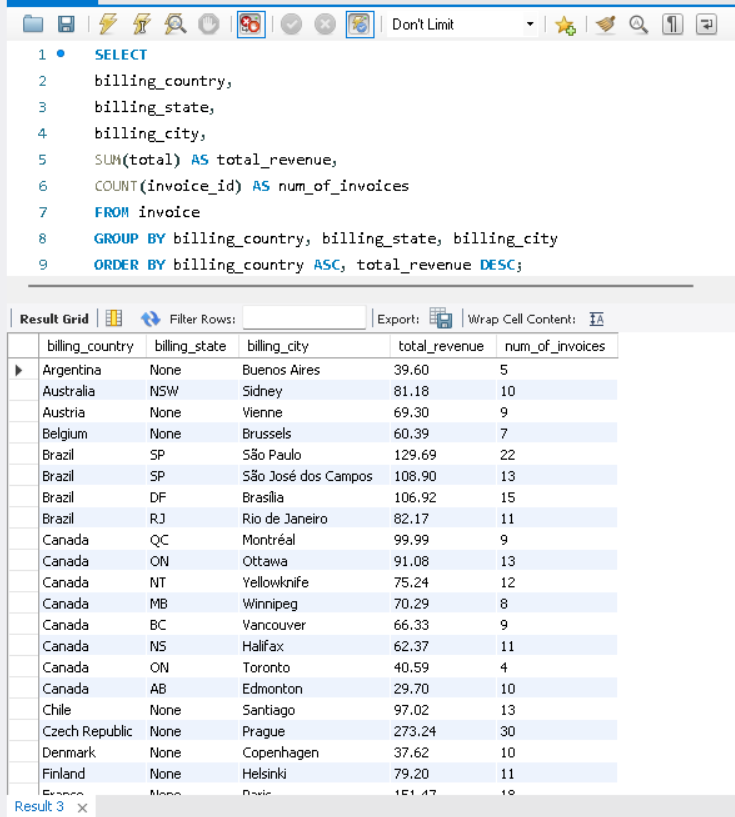
COUNT(invoice\_id) AS num\_of\_invoices

FROM invoice

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

ORDER BY billing\_country ASC, total\_revenue DESC;

**Result:**

****

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

**Solution:**

* Concepts used: CTE, Joins, GROUP BY, Aggregate Functions, Sorting (ORDER BY)
* Tables used: customer, invoice

**Query:**

WITH Top5CustomersCountryWise AS (

SELECT

c.country,

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

SUM(i.total) AS total\_revenue,

RANK()

OVER (

PARTITION BY c.country

ORDER BY SUM(i.total) DESC

) AS countrywiseRank

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

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

)

SELECT

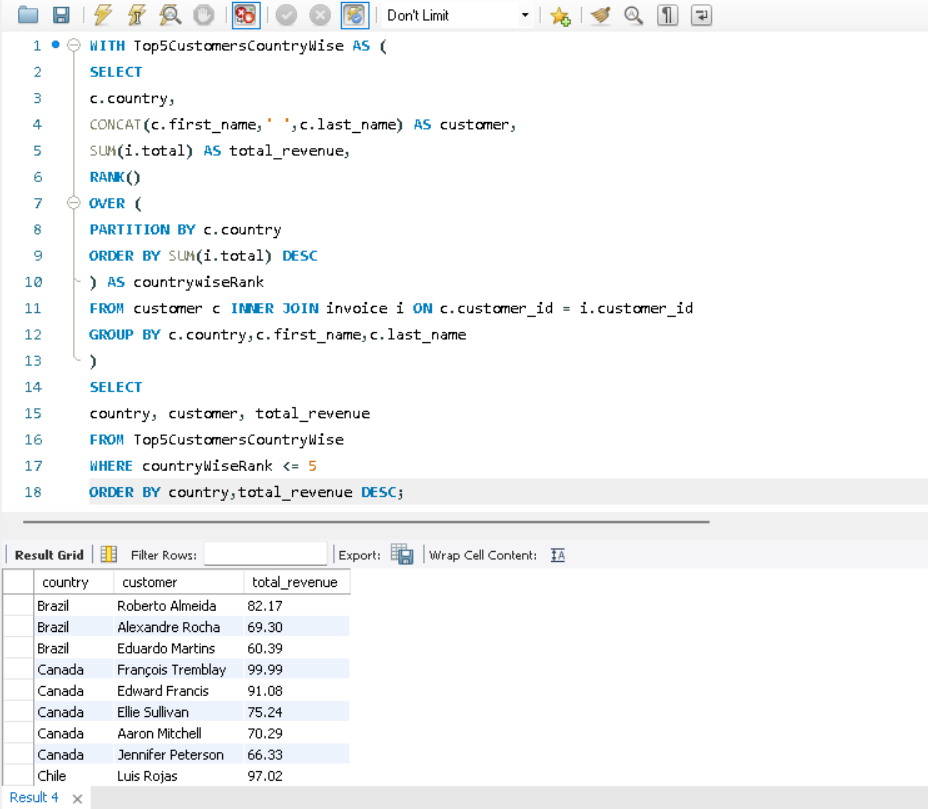
country, customer, total\_revenue

FROM Top5CustomersCountryWise

WHERE countryWiseRank <= 5

ORDER BY country,total\_revenue DESC;

**Result: (48 Rows Returned)**

****

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

**Solution:**

* Concepts used: CTE, Joins, GROUP BY, Aggregate Functions (SUM)
* Tables used: customer, invoice, invoice\_line, track

**Query:**

WITH CustomerTrackSales AS (

SELECT

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

SUM(il.quantity) AS total\_quantity, SUM(i.total) AS total\_sales,

ROW\_NUMBER()

OVER(

PARTITION BY c.customer\_id

ORDER BY SUM(i.total) DESC

) AS sales\_rank

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

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

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

)

SELECT

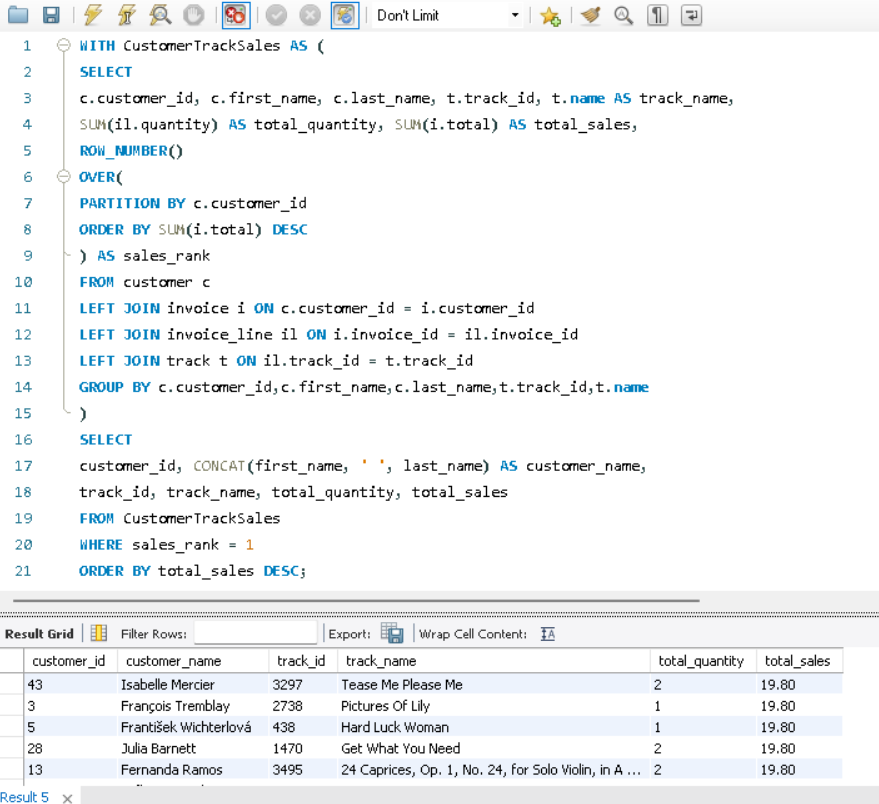
customer\_id, CONCAT(first\_name, ' ', last\_name) AS customer\_name,

track\_id, track\_name, total\_quantity, total\_sales

FROM CustomerTrackSales

WHERE sales\_rank = 1

ORDER BY total\_sales DESC;

**Result: (59 Rows Returned)**

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

**Solution:**

**1.Frequency of Purchases-**

* Concepts used:

CTE, Joins, GROUP BY, Aggregate & DATE Functions, Sorting

* Tables used: customer, invoice

**Query:**

WITH PurchaseFrequency AS (

SELECT

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

COUNT(i.invoice\_id) AS total\_purchases,

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

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

ROUND(

DATEDIFF(MAX(DATE(i.invoice\_date)),MIN(DATE(i.invoice\_date))) /

COALESCE(COUNT(i.invoice\_id)-1, 0), 0) AS avg\_days\_bet\_purchases

FROM customer c

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

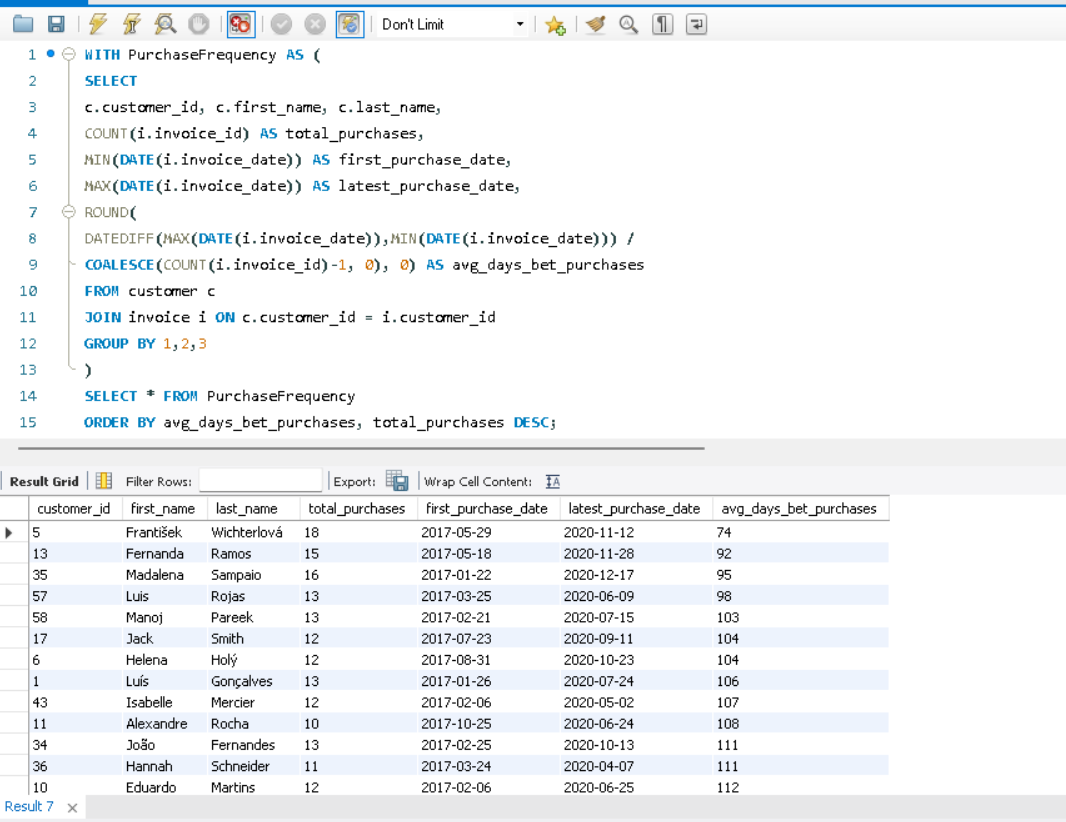
GROUP BY 1,2,3

)

SELECT \* FROM PurchaseFrequency

ORDER BY avg\_days\_bet\_purchases, total\_purchases DESC;

**Result: (59 Rows Returned)**

****

**2.Average Order Value-**

* Concepts Used: CTE, Aggregate Functions, GROUP BY, Sorting (ORDER BY)
* Tables used: customer, invoice

**Query:**

WITH CustomerPurchases AS (

SELECT

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

SUM(i.total) AS total\_order\_value,

COUNT(i.invoice\_id) AS total\_purchases,

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

FROM customer c

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

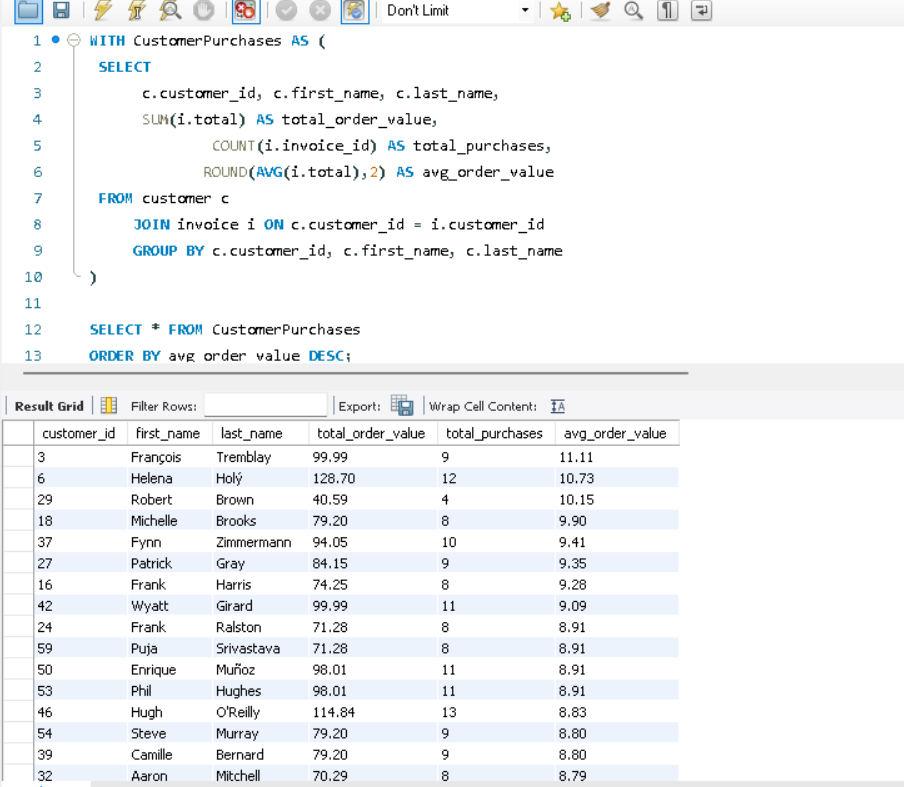
GROUP BY c.customer\_id, c.first\_name, c.last\_name

)

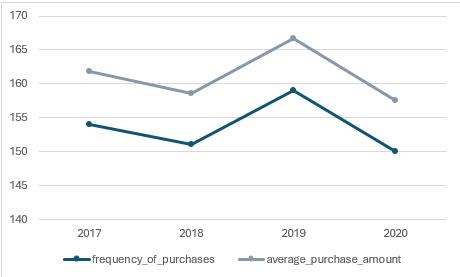
SELECT \* FROM CustomerPurchases

ORDER BY avg\_order\_value DESC;

**Result: ( 59 Row Returned )**

****

**Visualization:**

****

**Analysis**

Steady Purchase Frequency:

* The frequency of purchases remains relatively consistent across the years, ranging between 150 and 159, indicating stable customer engagement over time.

Slight Decline in Average Purchase Amount:

* The average purchase amount shows a slight downward trend from $7.8 in 2017 to $7.59 in 2020, which may indicate:
  + A change in purchasing behavior (e.g., smaller purchases).
  + Discounts or price adjustments on products.

2019 Performance:

* 2019 saw the highest frequency of purchases (159), paired with a slight recovery in the average purchase amount to $7.68 after a decline in 2018.

2020 Impact:

* The slight dip in both purchase frequency and average purchase amount in 2020 may suggest external factors like market conditions or changes in customer spending habit

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

**Solution:**

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

In this case, I have considered a customer to be churned if they have not made any purchase

for >180 days between the last purchase date and the second last purchase date.

* Concepts Used: CTE, Joins, Aggregate Functions, Window Functions, Date Functions
* Tables used: customer, invoice

**Query:**

WITH PreviousCustomerPurchases AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

DATE(i.invoice\_date) AS invoice\_date,

LEAD(DATE(i.invoice\_date)) OVER(PARTITION BY c.customer\_id ORDER BY invoice\_date

DESC) AS prev\_purchase

FROM customer c

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

),

PrevPurchaseRank AS (

SELECT

\*,

ROW\_NUMBER() OVER(PARTITION BY customer\_id ORDER BY prev\_purchase DESC)

AS prev\_purchase\_rn

FROM PreviousCustomerPurchases

),

PreviousPurchaseDate AS (

SELECT

\*,DATEDIFF(invoice\_date,prev\_purchase) AS days\_since\_last\_purchase

FROM PrevPurchaseRank

WHERE prev\_purchase\_rn = 1

AND DATEDIFF(invoice\_date,prev\_purchase) > 180

ORDER BY days\_since\_last\_purchase DESC

)

SELECT

COUNT(pp.customer\_id) AS churned\_customers,

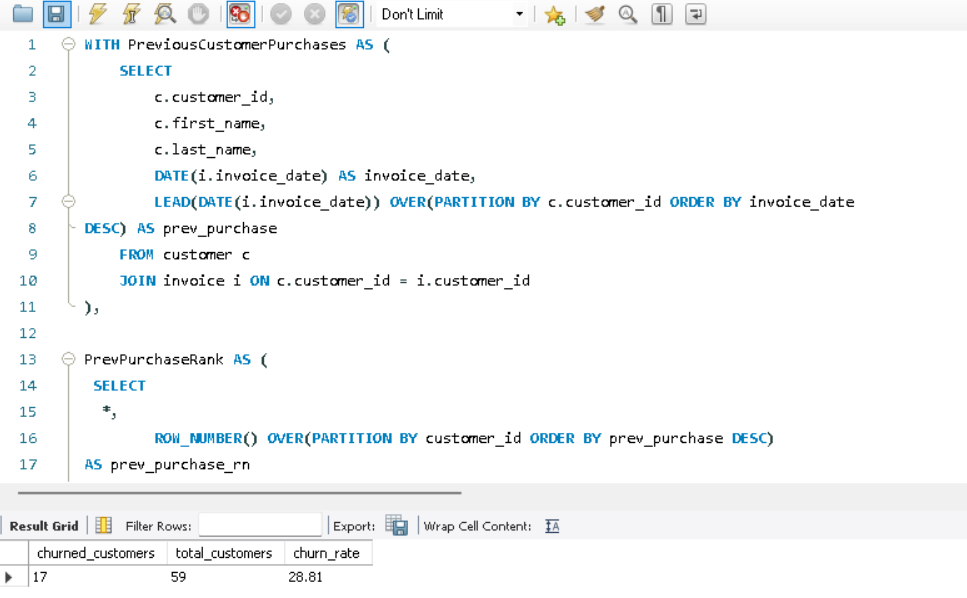
COUNT(c.customer\_id) AS total\_customers,

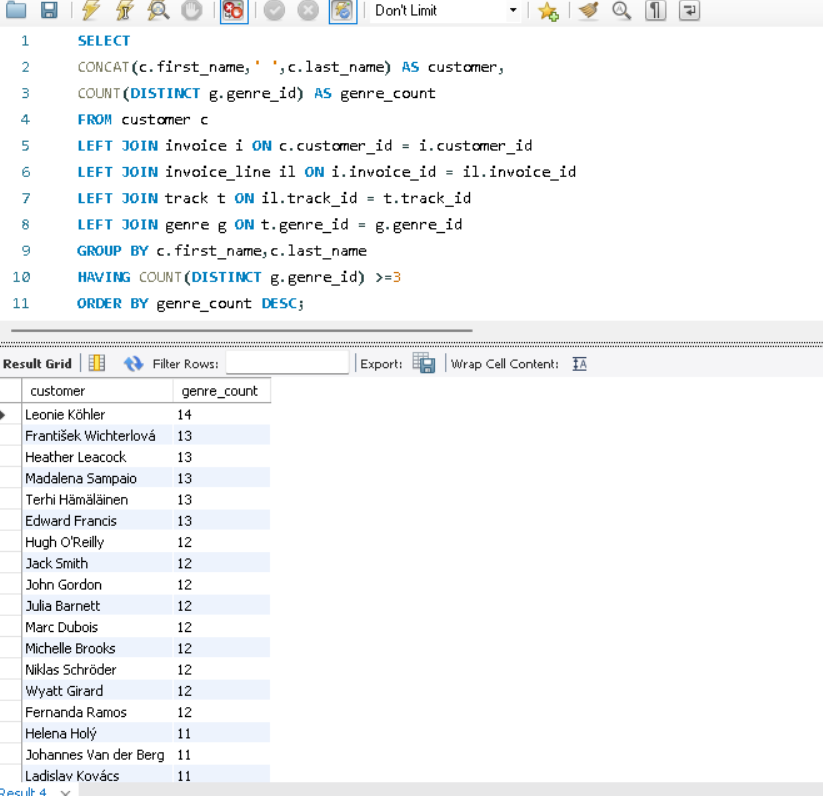
ROUND((COUNT(pp.customer\_id) \* 100) / COUNT(c.customer\_id), 2) AS churn\_rate

FROM customer c

LEFT JOIN PreviousPurchaseDate pp ON c.customer\_id = pp.customer\_id;

**Result:**

****

****

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

**1. Percentage of total sales contributed by each genre in the USA:**

* Concepts Used: CTE, Joins, Aggregate Functions, GROUP BY, Sorting (ORDER BY)
* Tables used: genre, track, invoice, invoice\_line, album, artist

**Query:**

WITH SalesGenreRankUSA AS (

SELECT

g.name AS genre,

ar.name AS artist,

SUM(i.total) AS genre\_sales,

DENSE\_RANK()

OVER(

PARTITION BY g.name

ORDER BY SUM(il.unit\_price \* il.quantity) DESC

) AS genre\_rank

FROM genre g

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

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

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

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

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

WHERE i.billing\_country = 'USA'

GROUP BY 1,2

),

TotalSalesUSA AS (

SELECT

SUM(i.total) AS total\_sales

FROM invoice\_line il

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

WHERE i.billing\_country = 'USA'

)

SELECT

\*,

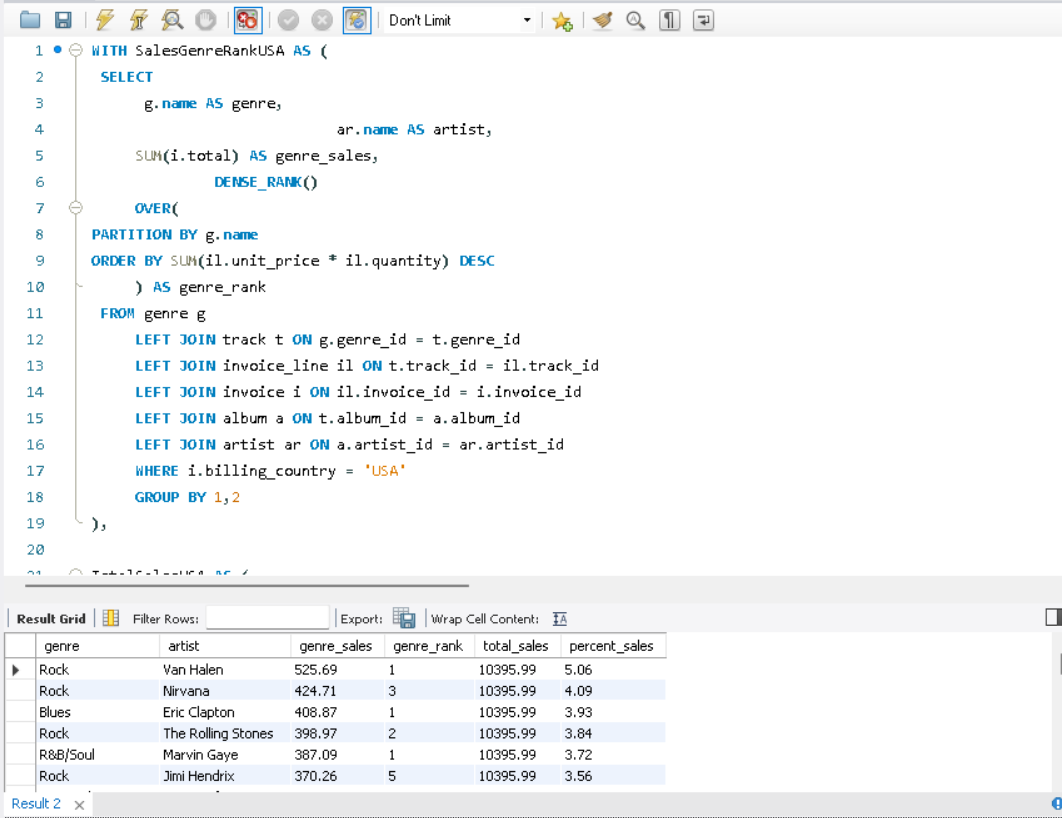
ROUND((s.genre\_sales / t.total\_sales)\* 100,2) AS percent\_sales

FROM SalesGenreRankUSA s

JOIN TotalSalesUSA t

ORDER BY s.genre\_sales DESC, s.genre ASC;

**Result: ( 59 Rows Returned )**

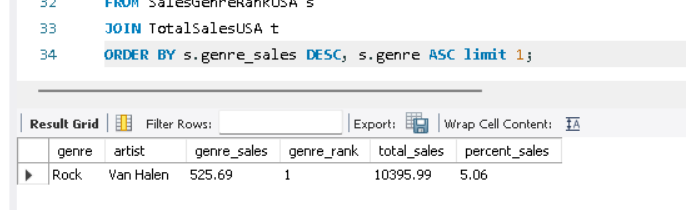
****

**2. Best Selling Genre and Artist**

To identify the best selling genre and artist, we have to include LIMIT 1 at the end of order by in the last query itself which is

**ORDER BY s.genre\_sales DESC, s.genre ASC LIMIT 1;**

We will get the following result:



**Visualization-**

**Analysis-**

* **Top Artist**: Van Halen leads the chart with the highest value (525.69), making them the most significant rock artist in the dataset.
* **Other Major Artists**: Nirvana (424.71), The Rolling Stones (398.97), and Jimi Hendrix (370.26) follow closely, indicating their strong influence in the rock genre.
* **Lower-Ranked Artists**: Scorpions (199.98) and Alice In Chains (229.68) have the smallest shares, showing relatively lower recognition or impact compared to others.
* **Even Distribution**: While Van Halen dominates, most other artists have values ranging between 200-400, indicating a fairly even spread of popularity among them.

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

**Solution:**

* Concepts used: Joins, GROUP BY, HAVING, Sorting (ORDER BY)
* Tables used: customer, invoice, invoice\_line, track, genre

**Query:**

SELECT

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

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

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

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

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

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

HAVING COUNT(DISTINCT g.genre\_id) >=3

ORDER BY genre\_count DESC**;**

**Result: (59 Rows Returned)**

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

**Solution:**

* Concepts used: CTE, Joins, GROUP BY, Window Functions (DENSE\_RANK)
* Tables used: genre, track, invoice\_line, invoice

**Query:**

WITH SalesWiseGenreRank AS (

SELECT

g.name AS genre,

SUM(i.total) AS total\_sales,

DENSE\_RANK() OVER (ORDER BY SUM(i.total) DESC) AS genre\_rank

FROM genre g

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

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

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

WHERE i.billing\_country = 'USA'

GROUP BY g.name

)

SELECT

genre,

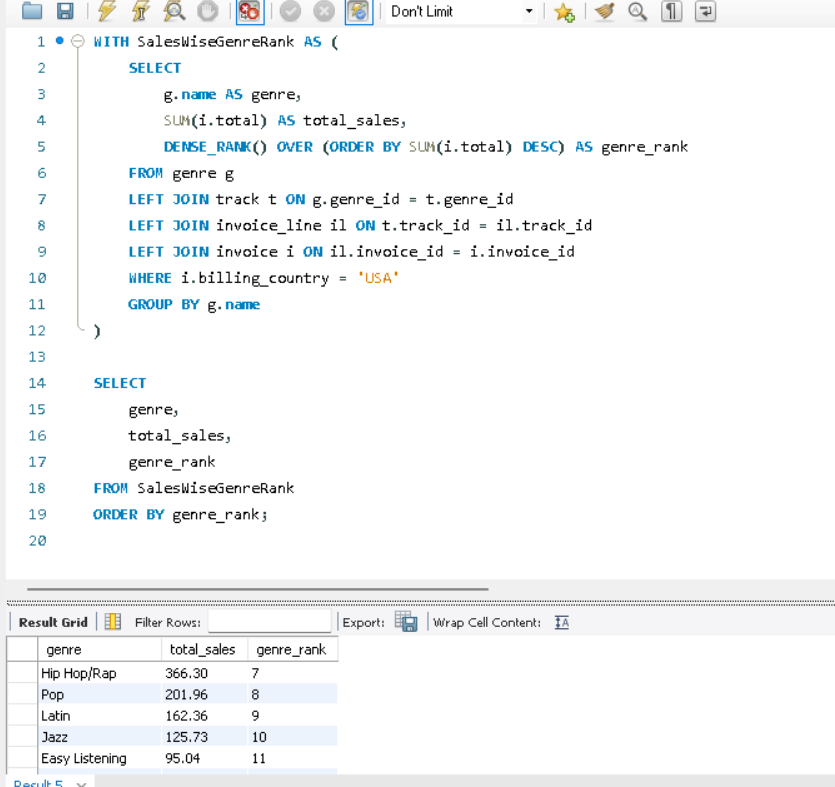
total\_sales,

genre\_rank

FROM SalesWiseGenreRank

ORDER BY genre\_rank;

**Result: (17 Rows Returned)**

****

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

**Solution:**

* Concepts Used: CTE, Joins, Aggregate Functions, GROUP BY, Sorting (ORDER BY)
* Tables used: customer, invoice

**Query:**

WITH CustomerLastPurchase AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

MAX(DATE(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, c.first\_name, c.last\_name

),

CustomerPurchases AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

DATE(i.invoice\_date) AS invoice\_date

FROM customer c

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

)

SELECT

clp.customer\_id,

clp.first\_name,

clp.last\_name,

clp.last\_purchase\_date

FROM CustomerLastPurchase clp

LEFT JOIN CustomerPurchases cp ON clp.customer\_id = cp.customer\_id

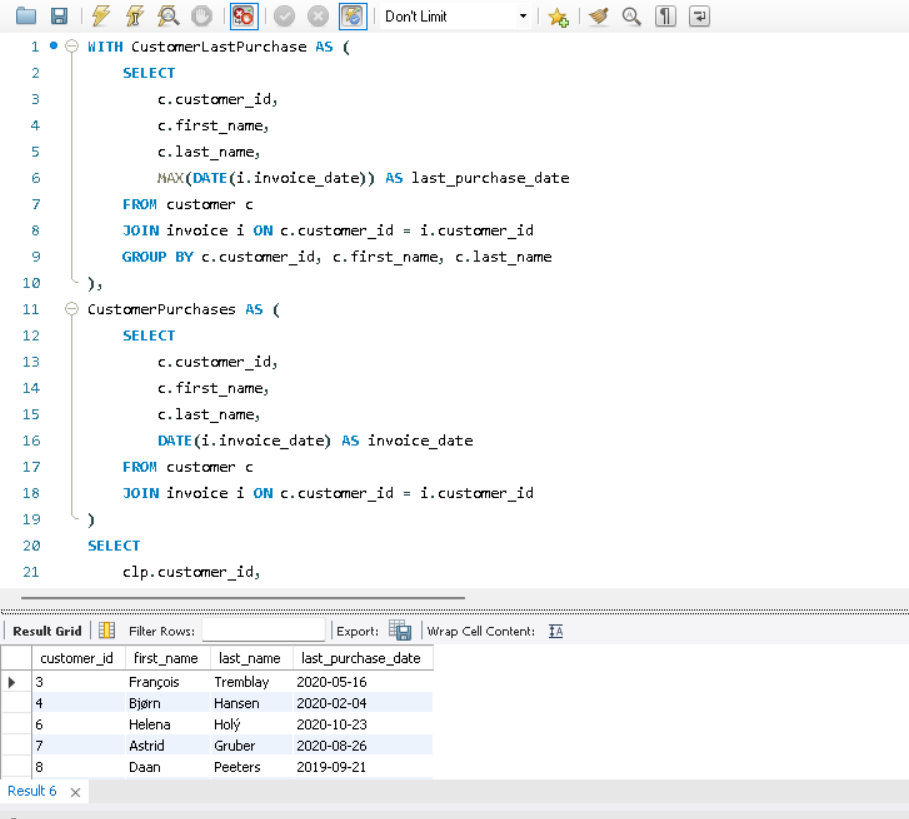
AND cp.invoice\_date BETWEEN clp.last\_purchase\_date - INTERVAL 3 MONTH AND

clp.last\_purchase\_date - INTERVAL 1 DAY

WHERE cp.invoice\_date IS NULL

ORDER BY clp.customer\_id;

**Result: (35 Rows Returned)**

****

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.**

Based on the **genre sales analysis** from the bar chart, the **top three**

**albums** that should be prioritized for **advertising and promotion in the USA** are:

**Top 3 Recommended Albums for Promotion:**

1. **Seek And Shall Find: More Of The Best (1963-1981)**
   * **Genre:** R&B/Soul
   * **Sales Value:** **387.09** (Highest)
   * **Reason:** This album leads in sales, indicating strong demand for R&B/Soul in the USA.
2. **From The Muddy Banks Of The Wishkah [live]**
   * **Genre:** Rock
   * **Sales Value:** **379.17**
   * **Reason:** Rock is a dominant genre, and live albums often attract dedicated fan bases.
3. **Are You Experienced?**
   * **Genre:** Rock
   * **Sales Value:** **370.26**
   * **Reason:** A classic rock album with high sales potential, making it a strong candidate for promotion.

**Justification for Promotion Strategy:**

* **Rock is the leading genre** in terms of sales representation, making it a safe investment.
* **R&B/Soul album has the highest sales**, indicating a strong audience demand.
* **Promoting these albums in the USA** can drive engagement, as these genres have a solid listener base.

**Chart has been created by exporting the SQL output into a excel file:**

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

**Top-Selling Genres (Outside the USA)**

1. **Rock** - Highest-selling genre with **21,183.03** in sales.
2. **Metal** - Second highest, with **4,290.66** in sales.
3. **Alternative & Punk** - Close to Metal, with **3,606.57** in sales.
4. **Latin** - Least among the top genres, with **1,543.41** in sales.
5. **Grand Total** - **31,800.78**, summing up all genres' sales outside the USA.

**Commonalities and Differences**

* **Commonality:** Rock remains a dominant genre outside the USA, similar to its global popularity.
* **Difference:** Metal and Alternative & Punk show notable sales, whereas genres like Pop or Hip-Hop (which are typically strong in the USA) are missing.
* **Latin music**, while popular in specific regions like South America and Spain, has significantly lower sales compared to Rock.

**Comparison of Top-Selling Genres: USA vs. Other Countries**

**Observations from the USA Sales Rank Chart:**

1. **Rock** is the dominant genre in the USA, with a **sales rank of 412**, similar to its strong presence outside the USA.
2. **Alternative & Punk** (95) and **Metal** (87) are also popular, consistent with international trends.
3. **R&B/Soul (36)** and **Blues (27)** appear in the USA but were absent from the previous chart (non-USA sales).

**Key Comparisons: USA vs. Other Countries**

| **Genre** | **Sales (Outside USA)** | **Sales Rank (USA)** |
| --- | --- | --- |
| **Rock** | 21,183.03 | 412 |
| **Metal** | 4,290.66 | 87 |
| **Alternative & Punk** | 3,606.57 | 95 |
| **Latin** | 1,543.41 | Not in USA Chart |
| **R&B/Soul** | Not in Non-USA Chart | 36 |
| **Blues** | Not in Non-USA Chart | 27 |

**Commonalities:**

* **Rock** leads in both the USA and international markets.
* **Metal** and **Alternative & Punk** have a strong presence in both markets.

**Differences:**

* **Latin music** is among the top-selling genres outside the USA but is absent from the USA ranking.
* **R&B/Soul and Blues** appear in the USA's top-selling genres but are missing from the international rankings.

1. **Customer Purchasing Behaviour Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?**

### 1. **Purchase Frequency (Long-Term vs. New Customers)**

* Long-term customers have a significantly higher purchase frequency (583 purchases) compared to new customers (31 purchases).
* This indicates that customer retention strategies are effective in encouraging repeat purchases.

### 2. **Spending Amount (Long-Term vs. New Customers)**

* The average spending amount of long-term customers (9.80) is slightly higher than that of new customers (9.30).
* While the difference is not substantial, it suggests that long-term customers may be more comfortable making consistent purchases.

### **3. Top Customers' Purchasing Patterns**

* The top 17 customers show varying purchase frequencies and spending amounts.
* Some customers have a high frequency but moderate spending, while others show consistent spending habits.
* Identifying high-value customers with frequent and high spending can help in targeted loyalty programs.

### **Insights for Customer Retention Strategies-**

* **Encourage Repeat Purchases**: Since long-term customers buy more frequently, offering loyalty rewards, discounts, or exclusive deals can further boost retention.
* **Enhance New Customer Engagement**: The low purchase frequency among new customers suggests a need for better onboarding, first-time discounts, or personalized recommendations.
* **Segment High-Value Customers**: Focusing on the top spenders and frequent buyers with personalized promotions can maximize revenue.
* **Improve Basket Size & Spending**: While long-term customers spend more, increasing the average transaction value through bundle offers or upselling can drive higher revenue.

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?**

### **Product Affinity Insights**

#### **1. Genre-Based Affinity**

* **Rock + Metal + Alternative & Punk** appear frequently together, meaning customers who buy Rock albums may also be interested in Metal and Punk collections.
* **Latin & Blues** appear as secondary choices; these could be bundled with similar genre playlists or albums.

#### **2. Cross-Selling Strategies**

* **Bundles & Discounts**: Offer Rock & Metal combo packs or discount vouchers when purchasing within these genres.
* **Recommendation System**: Suggest "Customers who bought Rock albums also purchased Alternative & Punk albums."
* **Personalized Playlists**: For subscription services, curated playlists combining these genres can enhance engagement.

#### **3. Artist/Album Affinity (Requires More Data)**

* If data on specific **artists or albums** is available, we could analyze which artists are often purchased together.
* Example: If **Led Zeppelin and Metallica albums** are often bought together, recommend Metallica albums to Led Zeppelin buyers.

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?**

### **Regional Market Analysis: Customer Purchasing & Churn Trends**

This analysis aims to understand whether **customer purchasing behaviors and churn rates vary across geographic regions** and how they might relate to **local demographics and economic factors**.

### **Key Observations from the Chart**

1. **High Churn Rates in Some Countries**
   * **100% churn rate** is observed in multiple countries, including Argentina, Belgium, Chile, Hungary, Italy, Netherlands, and Norway.
   * **High churn rates (50% and above)** are seen in countries like Germany, France, and the Czech Republic.
   * **USA has a lower churn rate (15.38%)** despite having the highest number of total customers (13).
   * **Brazil has an unusual churn rate of 37.5%, indicating partial customer retention.**
2. **Customer Distribution Across Countries**
   * The **USA has the largest customer base (13 customers)**, followed by **Brazil (8 customers)**.
   * Other countries have **only 1-5 customers**, indicating smaller markets.

### **Potential Correlations & Insights**

#### **1. Economic & Demographic Factors**

* Countries with **weaker economies or fluctuating currencies** (e.g., Argentina, Hungary) may have **higher churn rates** due to affordability issues.
* Wealthier countries (e.g., USA, UK) tend to have more stable customers, indicating higher retention.

#### **2. Local Market Penetration & Brand Awareness**

* Countries with a **higher total number of customers** but **lower churn rates** (USA, Brazil) suggest better market presence and customer engagement.
* Countries with **only 1 or 2 total customers but 100% churn** might indicate **low product adoption or lack of brand awareness**.

#### **3. Possible Store or Subscription Model Influence**

* If this dataset is from a **subscription-based service**, high churn might be due to a lack of localized content or pricing mismatches.
* If it's from **physical stores**, regional competition and availability might affect customer behavior.

### **Strategic Recommendations**

1. **Customer Retention Programs in High-Churn Countries**
   * **Localized promotions & discounts** to retain customers in **high-churn regions**.
   * **Improve customer engagement** via targeted marketing campaigns in underperforming markets.
2. **Market Expansion in Low-Churn, High-Customer Regions**
   * **USA & Brazil have the strongest customer bases** → invest in marketing & expansion in these regions.
   * Offer **localized services, better payment options, or exclusive content** in regions with potential.
3. **Conduct Further Analysis on Product Preferences by Region**
   * Understand **which genres or artists are popular in specific countries**.
   * Compare churn trends with **customer purchasing behavior** to optimize **product offerings**.
4. **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?**

### **Customer Risk Profiling: Identifying High-Churn Segments**

This analysis aims to determine **which customer segments are more likely to churn** or **reduce their spending**, based on factors such as **location, demographics (age, gender), and purchase history**.

### **Key Insights from Churn Rate by Country Chart**

* **High-Risk Regions:**
  + Countries with **100% churn rate**: Italy, Norway, Chile, Netherlands, Belgium, Hungary, Argentina.
  + **50% churn rate**: Czech Republic, Germany.
  + These regions may indicate high dissatisfaction, lack of engagement, or economic constraints.
* **Low-Risk Regions:**
  + Countries with **0% churn rate**: Finland, Australia, Sweden, India, Brazil, UK, Poland, Denmark, Portugal, Austria, Spain.
  + These regions may have strong customer retention and engagement.
* **Moderate-Risk Regions:**
  + **USA (15.38%)** and **Canada (37.5%)** show churn but at manageable levels.

### **Factors Contributing to Churn Risk**

1. **Demographics (Age & Gender)**
   * **Younger customers** tend to have lower brand loyalty and switch frequently.
   * **Older customers** may churn due to lack of digital engagement.
   * Gender-based preferences in spending and product preferences might play a role.
2. **Economic & Regional Factors**
   * **High-churn regions** may experience economic instability, affecting spending habits.
   * **Wealthier regions with low churn** suggest higher customer satisfaction.
3. **Purchase History & Behavior**
   * **Customers who purchase infrequently** or have **low-value transactions** are more likely to churn.
   * **Subscription cancellations or long gaps between purchases** indicate churn risk.
4. **Engagement & Customer Experience**
   * Poor customer service, lack of personalization, or irrelevant product recommendations drive churn.
5. **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?**

### **Customer Lifetime Value (CLV) Modeling - Insights from Charts**

1. **Identifying At-Risk Customers** **(📊 Days Since Last Purchase Chart)**
   * Customers with high days since the last purchase (e.g., **Daan Peeters: 1990 days**) may have churned.
   * Common pattern: These customers have **lower purchase frequency**, indicating disengagement.
2. **High-Value Customers & Engagement (📊 Purchase Frequency Chart)**
   * Customers like **Niklas Schräder and Hannah Schneider** purchase frequently, making them valuable for **loyalty programs**.
   * Targeting them with **VIP discounts, early access sales, or referral bonuses** can boost retention.
3. **Segmenting Customers for CLV Prediction** (📊 Both Charts)
   * **High-frequency, low days since last purchase** → Loyal customers; prioritize retention offers.
   * **Low-frequency, high days since last purchase** → Churn-risk customers; require re-engagement campaigns.
4. **Predicting Future Value Using Purchase Patterns** (📊 Both Charts)
   * Customers with **consistent purchase behavior over time** tend to have a **higher lifetime value (LTV)**.
   * A **drop in purchase frequency or increased days since last purchase** signals a decline in CLV.
5. **Targeted Marketing & Loyalty Strategies**
   * Use **personalized discounts** for low-frequency customers to boost engagement.
   * Implement **win-back email campaigns** for customers showing churn behavior.
   * Offer **tiered loyalty programs** to reward frequent purchasers and increase retention.
6. **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 evaluate the effectiveness of promotional campaigns, compare key metrics such as sales, customer acquisition, and retention before, during, and after the campaign.**

1. **Customer Acquisition** – Analyze the number of new customers gained during the campaign and assess how promotions influenced their initial purchases.
2. **Retention Analysis** – Track repeat purchase behavior and churn rates among customers who engaged with the promotion to determine its long-term impact.
3. **Sales Performance** – Compare overall revenue, average order value, and purchase frequency during the campaign period against baseline trends.
4. **Customer Segmentation** – Identify which customer groups responded most positively to promotions to refine future targeting strategies.
5. **Engagement Insights** – Assess customer interactions, such as email open rates or discount redemption rates, to measure promotional effectiveness beyond just sales figures.
6. **How would you approach this problem, if the objective and subjective questions weren't given?**

**If the objective and subjective questions weren’t provided, I would begin with a broad exploration of the dataset to uncover patterns and insights related to customer behavior, sales trends, and promotional impact. My approach would include the following steps:**

1. **Understanding Business Objectives** – I would first clarify key business goals, such as improving customer retention, increasing sales, or identifying high-value customer segments, to ensure the analysis is aligned with strategic needs.
2. **Data Exploration and Cleaning** – I would perform data cleaning to address missing values or inconsistencies, followed by exploratory analysis to understand data distribution, trends, and potential anomalies.
3. **Defining Key Metrics and Customer Segments** – Establishing relevant KPIs like customer acquisition rate, churn rate, customer lifetime value (CLV), average order value, and campaign ROI would be crucial. I would also segment customers based on demographics, geography, and purchasing patterns to identify meaningful trends.
4. **Conducting Analysis and Modeling**:
   * **Churn Analysis** – Identifying factors influencing customer churn, such as purchase frequency and engagement with past promotions.
   * **Campaign Impact Assessment** – Comparing pre- and post-campaign data to measure changes in customer acquisition, retention, and overall sales performance.
5. **Deriving Actionable Insights** – Based on the findings, I would provide data-driven recommendations to enhance customer retention, target high-value segments, and refine marketing strategies for better ROI.
6. **How can you alter the "Albums" table to add a new column named "NewReleaseYear" of type INTEGER to store the release year of each album?**

We can make use of the ALTER statement to add a new column to a table.

The syntax is as follows:

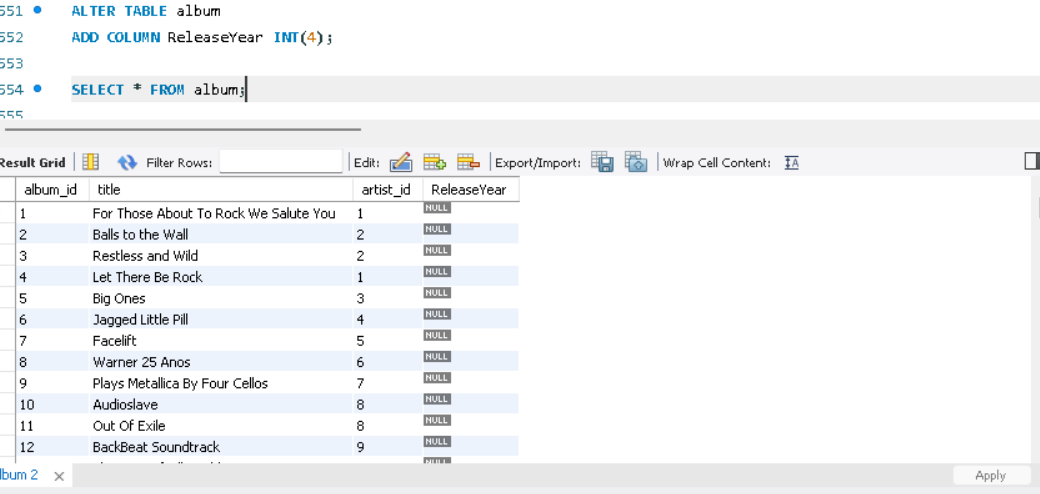
ALTER TABLE table\_name ADD COLUMN column\_name datatype;

To add the column named “NewReleaseYear” with INTEGER dataype to the album table, the following query can be used.

ALTER TABLE album ADD COLUMN NewReleaseYear INT(4);

SELECT \* FROM album;

**Output:**



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

SELECT

c.country,

ROUND(AVG(track\_count)) AS average\_tracks\_per\_customer,

SUM(i.total) AS total\_spent,

COUNT(DISTINCT c.customer\_id) AS no\_of\_customers,

ROUND(SUM(i.total)/ COUNT(DISTINCT c.customer\_id),2) AS avg\_total\_spent

FROM customer c

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

JOIN (

SELECT

invoice\_id,

COUNT(track\_id) AS track\_count

FROM invoice\_line

GROUP BY invoice\_id

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

GROUP BY c.country

ORDER BY avg\_total\_spent DESC;

**Output:**

