|  |  |
| --- | --- |
| Musical instrument,Light,String instrument,String instrument,Guitar,Guitar accessory,Musical instrument accessory,Music,Lighting,Yellow  Chinook Music Analysis  Swathi N S | Chinook Music Services is a global leader in the physical music records industry, dedicated to delivering exceptional musical entertainment around the world. With an extensive network of outlets and services, Chinook offers a rich variety of musical genres and albums, ensuring that every customer discovers their perfect rhythm. By connecting people through the universal language of music, Chinook has earned a reputation for quality, diversity, and global reach—making it a trusted choice for music lovers everywhere. |

**Objective Questions**

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

Yes, some tables have missing values.

**Approach:**

* To find missing values, a check for NULL values was done in key tables (Employee, Customer, and Track) using SQL queries with the WHERE condition.
* The query used:

SELECT \*

FROM table\_name

WHERE column\_name IS NULL;

By applying the above query for each key table, the following tables were found to have columns with missing values within them

* Customer
* Employee
* Track

**Customer Table:**

Missing values in specific columns were handled using the COALESCE function by replacing NULL values with meaningful placeholders**: "N/A"** for the Company, Postal Code, and Fax columns, **"Unknown"** for the State column, and **"Not Provided"** for the Phone column.

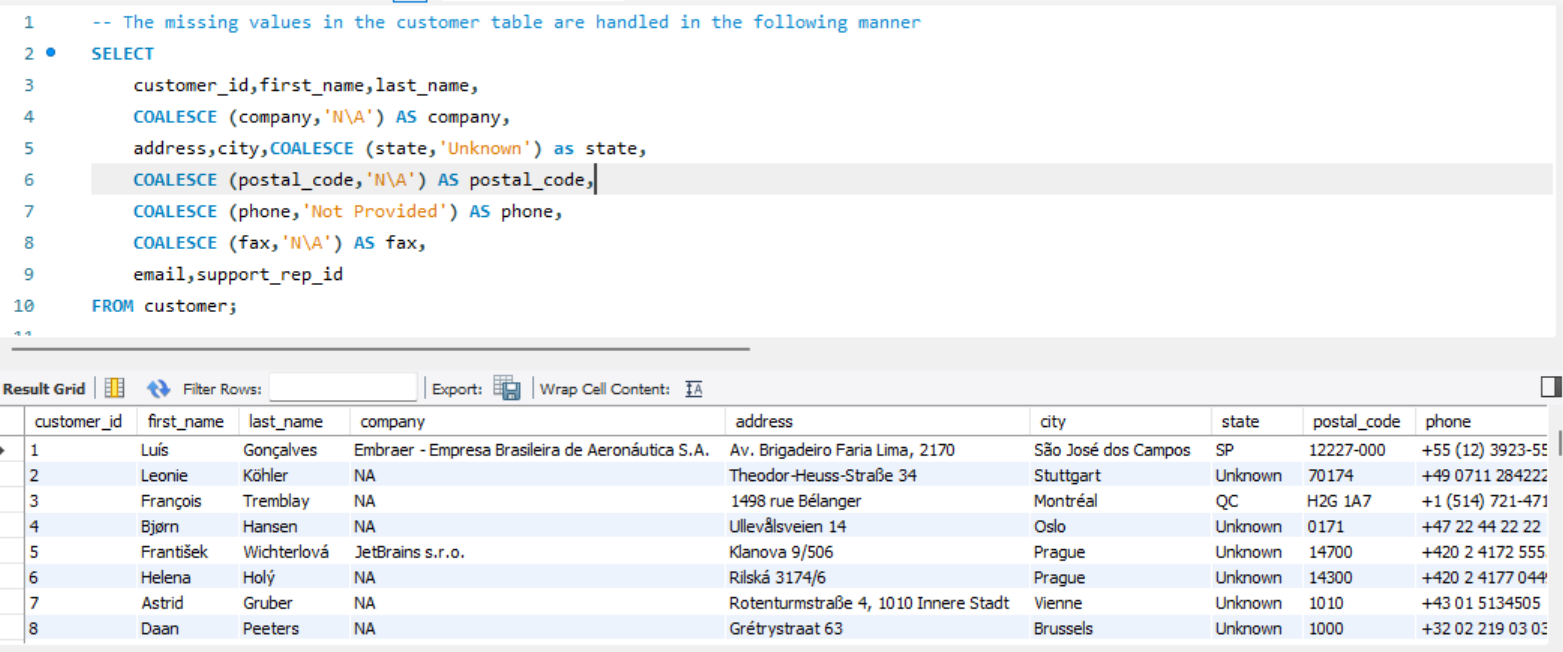
**Employee Table:**

The "reports\_to" column contained NULL values, which were replaced with **"None"** using the COALESCE function.

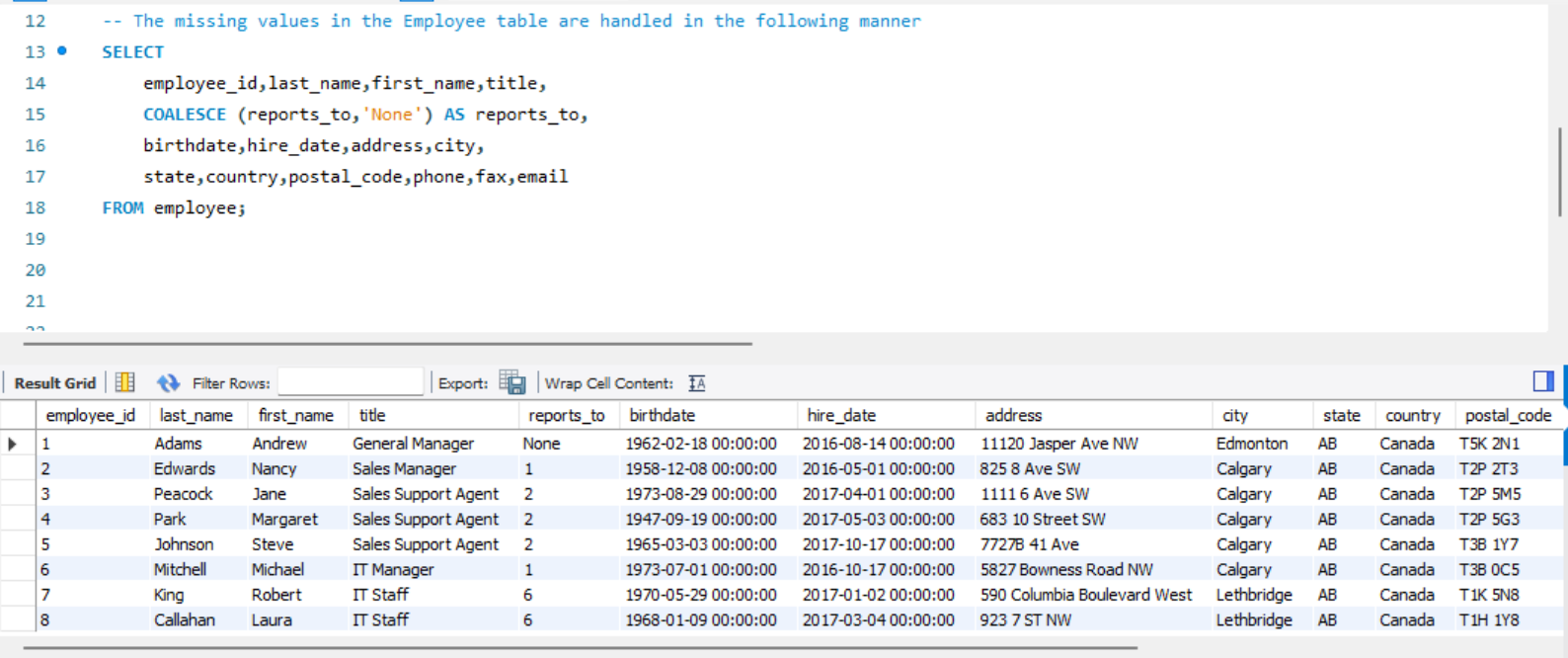
**Track Table:**

The "composer" column had missing data, and NULL values were substituted with "**No Composer"** to maintain consistency.

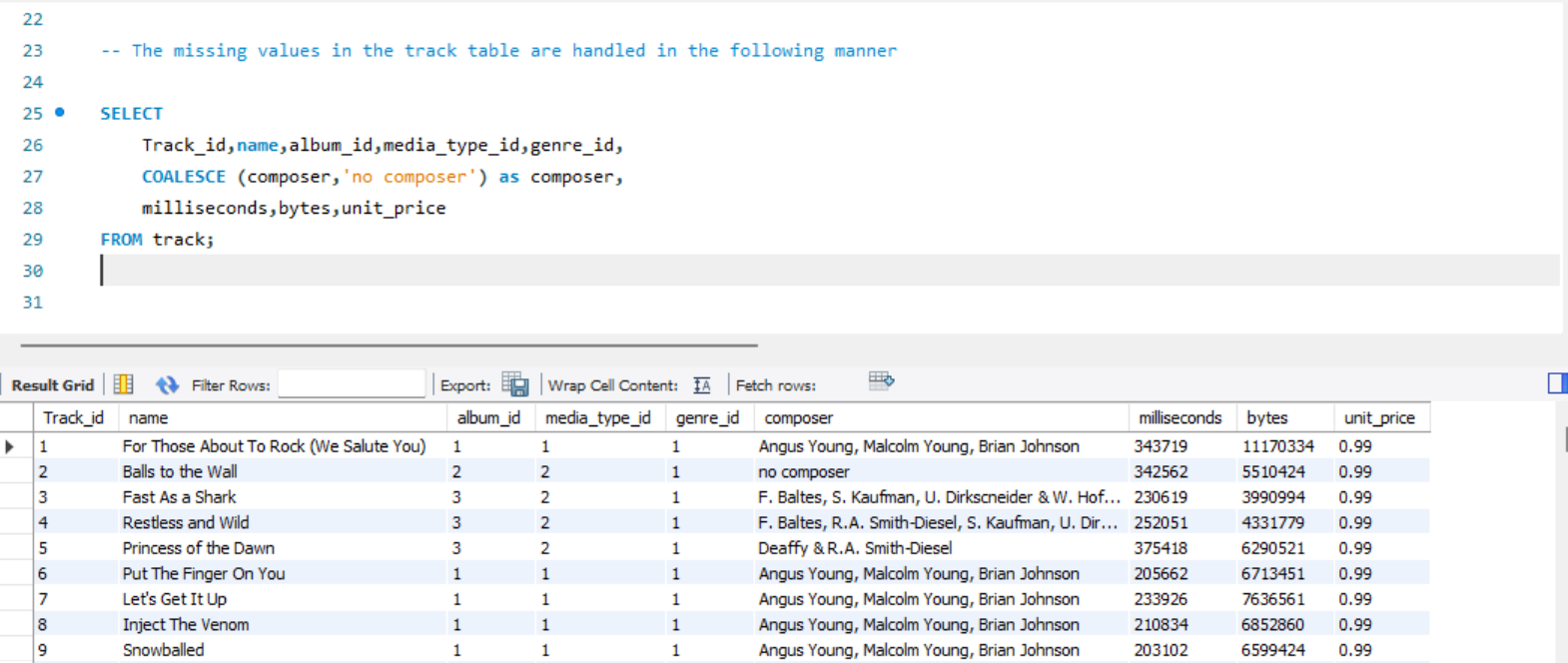
**Customers Table**



**Employees table**



**Tracks table**



* To check for duplicates, a query was used to count how many times each row (or specific columns) appeared in a table. If any count was more than one, it meant duplicates were present.
* To handle duplicates, different methods can be used:
  + Deleting extra copies and keeping only one.
  + Marking duplicate rows with a flag.
  + Creating a new table with only unique records.
* The GROUP BY clause was used on important columns along with HAVING COUNT (\*) > 1 to find duplicates. This helped identify rows that appeared more than once.
* If duplicates were found, the ROW\_NUMBER () function could be used to give each row a unique number within its group. Then:

Rows with ROW\_NUMBER () > 1 could be deleted to keep only one.

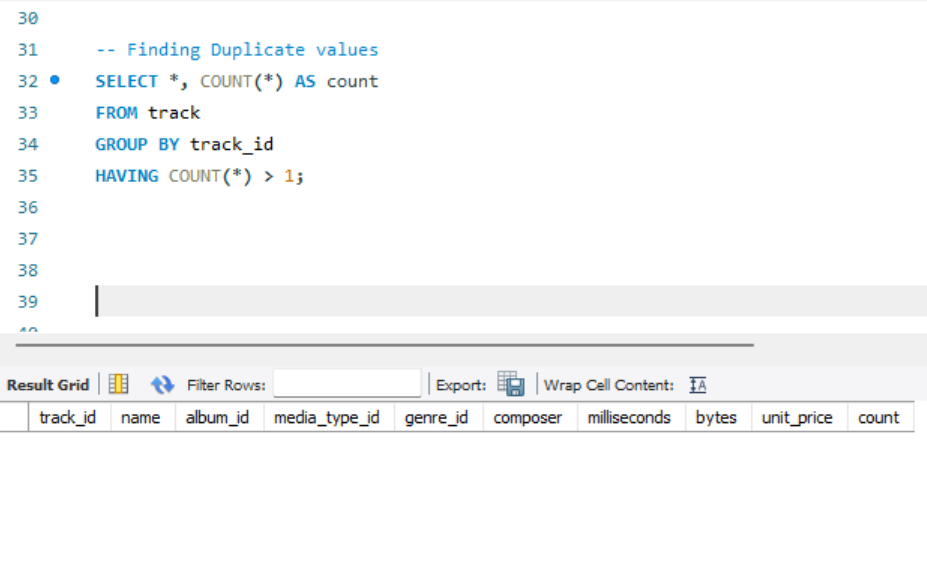
* Query Used:

SELECT \*, COUNT (\*) AS count

FROM track

GROUP BY track\_id

HAVING COUNT (\*) > 1;

****

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

**Approach:**

* The goal was to find the top-selling tracks and artists in the USA along with their most famous genres. To do this, I first checked the database structure to find the necessary tables.
* The Invoice and Track tables had sales details.
* The Artist and Genre tables contained artist and genre information.
* Since there was no direct link between Invoice and Track, I used the Invoice\_Line table to connect them.
* To link Track with Artist, I used the Album table.

**Solution:**

* I used INNER JOIN to connect multiple tables: Track, Invoice\_Line, Invoice, Genre, Album, and Artist.
* Aggregate functions like SUM were used to calculate total sales.
* GROUP BY and ORDER BY were applied to organize the results.
* The WHERE clause filtered data for the USA.
* Below is the query for the same.

SELECT

t.name AS Track\_Name,

SUM(i.total) AS Total\_Sales,

ar.name AS Artist\_Name,

g.name AS Genre

FROM track t

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

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

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

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

JOIN artist ar ON am.artist\_id=ar.artist\_id

WHERE i.billing\_country = "USA"

GROUP BY t.name,ar.name,g.name

ORDER BY Total\_Sales DESC;



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

**Approach:**

The question required an analysis of customer demographics based on age, gender, and location. However, since there were no columns for age and gender in the available tables, I focused on analyzing customer data based on location. To make the analysis more insightful, I included additional relevant columns. The key tables I used were Track, Invoice\_Line, Invoice, Genre, Album, and Artist.

**Solution:**

This SQL query provides a summary of the **customer base and total sales** by country from the **Chinook** database's invoice table. It calculates the **number of unique customers** in each country using COUNT (DISTINCT customer\_id) and the **total revenue** generated from those customers using SUM (total). The results are grouped by billing\_country, which represents where each customer was billed. Finally, the data is sorted in descending order by Total\_Sales, so the countries that generated the highest revenue appear at the top. This query is useful for identifying top-performing markets and understanding where the majority of customers and sales are coming from globally.

Below is the SQL query that implements this approach:

SELECT billing\_country,

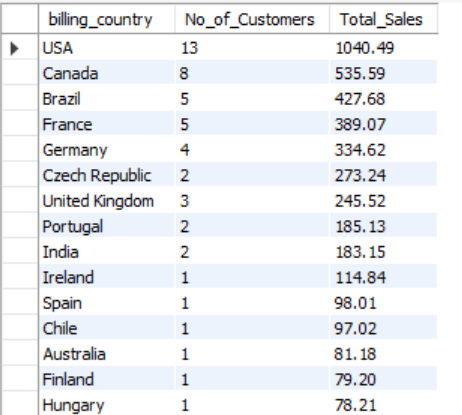
COUNT (DISTINCT customer\_id) AS No\_of\_Customers,

SUM (total) AS Total\_Sales

FROM invoice

GROUP BY billing\_country

ORDER BY Total\_Sales DESC;



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

**Approach:**

The goal was to calculate the total revenue and the number of invoices for each country, state, and city. After reviewing the data, it was clear that all the required details were available in the Invoice table, so there was no need to join with other tables. The method involved grouping the invoices at different levels—first by country, then by state within each country, and finally by city within each state. Aggregate functions were then applied to get the required numbers.

**Solution:**

To make the analysis easier to understand, column names such as billing\_country, billing\_state, and billing\_city were given aliases. This improved the readability of the output. The SUM function was used to calculate the total revenue, while the COUNT function was used to determine the number of invoices. The GROUP BY clause helped organize the data at each level (country → state → city). Finally, the ORDER BY clause arranged the results in a clear and logical order, making them easy to interpret.

Below is the SQL query that implements this approach.

SELECT

billing\_country AS Country,

billing\_state AS State,

billing\_city AS City,

SUM (total) AS Total\_Revenue,

COUNT (invoice\_id) AS No\_of\_Invoices

FROM invoice

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

ORDER BY billing\_country,billing\_state,billing\_city;



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

**Approach:**

The task requires identifying the top 5 customers based on total revenue in each country. To achieve this, I first analyzed the database schema and identified the relevant tables. The **Customers** table contains customer details, while the **Invoice** table holds revenue-related data. Since both tables share a common **Customer\_id**, I used this as the key to link them.

**Solution:**

To extract the required information, I performed a **JOIN** between the **Customers** and **Invoice** tables. I utilized **CONCAT** to combine names for better readability, used **aliases** for clarity, and applied **SUM** to calculate total revenue. To organize the data, I implemented **GROUP BY** and **ORDER BY**. Additionally, I applied **RANK ()** to assign rankings based on total revenue within each country. Finally, I wrapped the query inside a **CTE** to enhance readability and structure.

Here’s the final query:

WITH CountryRanking AS (

SELECT

i.billing\_country AS Country,

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

SUM(i.total) AS TotalRevenue,

RANK () OVER (PARTITION BY i.billing\_country ORDER BY SUM(i.total) DESC) AS rnk

FROM customer c

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

GROUP BY CustomerName,i.billing\_country

ORDER BY country,TotalRevenue DESC

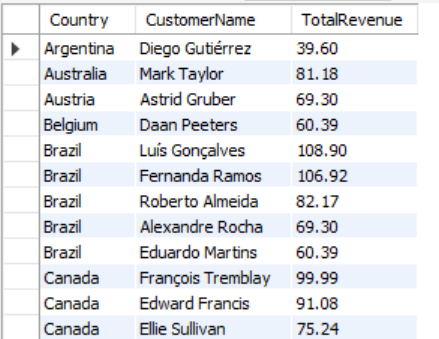
)

SELECT

Country,CustomerName,TotalRevenue

FROM CountryRanking

WHERE rnk <=5;



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

**Approach:**

To identify the top-selling tracks for each customer, I first examined the database structure to determine which tables contained the necessary information. I found that customer details are stored in the **Customers** table, while track details are in the **Track** table. Since there is no direct link between these tables, I used the **Invoice** and **Invoice\_Line** tables to establish a connection.

**Solution:**

To retrieve the required data, I combined four tables: **Customers, Invoice, Invoice\_Line,** and **Track** using joins. I used SQL functions like **CONCAT** to format the output, **Aliases** for clarity, and **COUNT** to determine the most purchased tracks. To organize the results, I applied **GROUP BY** and **ORDER BY** clauses. Additionally, I used the **RANK** function to rank the top-selling tracks for each customer. To improve readability, I structured the query using a **Common Table Expression (CTE)**.

Here’s the final query:

WITH Top\_Tracks AS (

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

t.name AS TrackName,

SUM(i.total) AS TotalSales,

RANK () OVER (PARTITION BY CONCAT (c.first\_name,' ',c.last\_name) ORDER BY SUM(i.total) DESC, t.name ASC) AS rnk

FROM customer c

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

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

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

GROUP BY CONCAT (c.first\_name,' ',c.last\_name), t.name

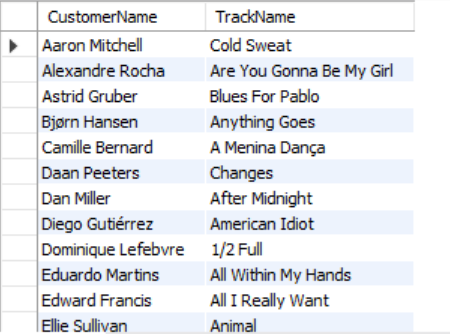
ORDER BY CustomerName

)

SELECT CustomerName,TrackName

FROM Top\_Tracks

WHERE rnk = 1;



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

**Approach:**

The objective was to analyze customer purchasing behavior by examining purchase frequency and average order value. Since payment method details were not available in the database, the analysis focused on these two factors.

Purchase frequency was calculated by counting the number of invoices per customer within specific timeframes. Average order value was determined by dividing the total invoice amount by the number of invoices for each customer. These insights provided a better understanding of how often customers made purchases and their typical spending per transaction.

**Solution:**

The **MONTH (invoice\_date)** function extracts the month from each invoice date, allowing for monthly analysis. The **COUNT (DISTINCT invoice\_id)** function counts the number of unique invoices in each month, representing the total number of purchases made. To measure spending behavior, the **AVG (total)** function calculates the average order value, which is then rounded to two decimal places using the **ROUND** function for better readability. The **GROUP BY Month** clause organizes the data by month, ensuring that purchases and average order values are calculated separately for each period. Finally, the **ORDER BY Month** clause arranges the results in chronological order, making it easier to analyze trends over time. The findings revealed seasonal shopping patterns, with **November**—despite being a festive month—recording the lowest number of purchases, while **April** had the highest. In terms of spending behavior, **May** had the lowest average order value, whereas **March** recorded the highest.

Below is the SQL query that implements this approach.

SELECT

MONTH (invoice\_date) as Month,

COUNT (DISTINCT invoice\_id) as Purchase\_num,

ROUND (AVG (total),2) as AvgOrderValue

FROM invoice

GROUP BY month

ORDER BY Month;



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

**Approach:**

The first step to solving the question is to understand what **churn rate** means.

Churn rate is the percentage of customers who **stop using** a service or product within a certain time. It’s like measuring how many people **leave** after signing up. Businesses, especially those with subscriptions, use this to see how well they’re keeping their customers.



**Solution:**

SELECT

((COUNT (DISTINCT CASE WHEN invoice\_date BETWEEN '2017-01-01' AND '2017-03-31'

THEN customer\_id END)

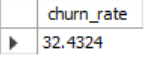
- COUNT (DISTINCT CASE WHEN invoice\_date BETWEEN '2020-11-01' AND '2020-12-31'

THEN customer\_id END))

/ COUNT (DISTINCT CASE WHEN invoice\_date BETWEEN '2017-01-01' AND '2017-03-31'

THEN customer\_id END)) \* 100 AS churn\_rate

FROM invoice;

****

The company's customer churn rate is **32.432%**, based on the total number of customers.

**Customer churn rate** refers to the percentage of customers who stop using a company's services over a given period.

* The **total number of customers at the start** is assumed to be equal to the number of customers who joined in the **first three months**.
* The **churn rate** is calculated based on the number of customers who left during the **last two months**.

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

**Approach:**

The goal was to find out how much each music genre contributed to total sales in the USA and identify the top-selling genres and artists.

To start, I looked at the database structure and found the necessary tables. The **Invoice** table contains sales and country details, while genre and artist information is stored in their respective tables. To connect all the necessary data, I also used tables like **Track, Invoice\_Line, and Album**.

**Solution:**

I used **INNER JOIN** to link multiple tables, including **Track, Invoice\_Line, Invoice, Genre, Album, and Artist**. To process the data effectively, I used:

* **Aliases** to make the query easier to read
* **Aggregate functions** like **SUM** to calculate total sales
* **ROUND** to format the results neatly
* **RANK** to rank the genres and artists based on sales
* **GROUP BY and ORDER BY** to organize the results properly
* **WHERE** clause to filter for sales in the USA

WITH Genre\_Rank AS (

SELECT

ar.name AS ArtistName, g.name AS Genre, SUM(i.total) as TotalSales,

ROUND(((SUM(i.total))/ (SELECT SUM(total) FROM invoice)) \*100,2) AS PercentageSales,

RANK() OVER(PARTITION BY g.name ORDER BY SUM(i.total) DESC) AS rnk

FROM track t

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

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

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

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

JOIN artist ar ON am.artist\_id=ar.artist\_id

WHERE i.billing\_country = "USA"

GROUP BY ar.name,g.name

)

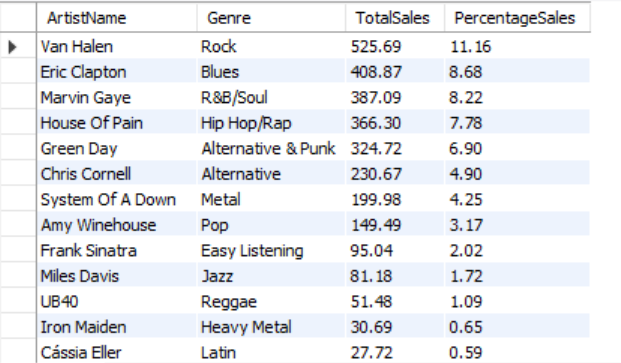
SELECT

ArtistName,Genre,TotalSales,PercentageSales

FROM Genre\_Rank

WHERE rnk = 1

ORDER BY PercentageSales DESC;



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

**Approach:**

To identify customers who purchased music from at least three different genres, it was necessary to connect the **Customer** and **Track** tables. Since there was no direct link between them, intermediary tables provided the required connection.

* The **Invoice\_Line** table linked tracks to invoices.
* The **Invoice** table connected invoices to customers.

By first joining **Customer** with **Invoice**, then **Invoice** with **Invoice\_Line**, and finally **Invoice\_Line** with **Track**, a relationship between customers and the genres they purchased was established. This allowed for counting the number of unique genres each customer bought.

**Solution:**

To enhance clarity, the **CONCAT** function combined the first and last names of customers, creating the alias **Customer\_Name**.

* The **COUNT (DISTINCT genre\_id)** function calculated the number of unique genres purchased, labeled as **Number**.
* The **GROUP BY customer\_id** clause ensured that results were aggregated per customer.
* To filter for customers who purchased from at least three different genres, the **HAVING COUNT (DISTINCT genre\_id) >= 3** condition was applied.
* Finally, the **ORDER BY customer\_id** clause sorted the output for better readability

SELECT

c.Customer\_id,

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

COUNT (distinct t.genre\_id) AS Number

FROM track t

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

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

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

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

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

ORDER BY c.customer\_id;



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

**Approach:**

To rank music genres by sales in the US, data from multiple tables was required. The **Genre** table contained genre names, while sales and country details were stored separately in the **Invoice** table. Since no direct connection existed between these tables, two intermediate tables were necessary:

* **Invoice\_Line**, which linked invoices to tracks
* **Track**, which linked tracks to genres

By joining **Genre → Track → Invoice\_Line → Invoice**, a complete path from genre to sales data was created, making it possible to analyze genre performance in the US.

**Solution:**

To calculate and rank genre sales, four tables were joined: **Genre, Track, Invoice\_Line,** and **Invoice**. Table aliases were used for clarity and simplicity.

* The **SUM ()** function was applied to compute total sales for each genre.
* A **WHERE** clause filtered the data to include only US sales.
* The **RANK ()** window function assigned ranks based on total sales.
* A **GROUP BY** statement grouped sales data by genre before ranking.

This method provided a structured view of genre performance in the US market.

SELECT

c.Customer\_id,

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

COUNT (distinct t.genre\_id) AS Number

FROM track t

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

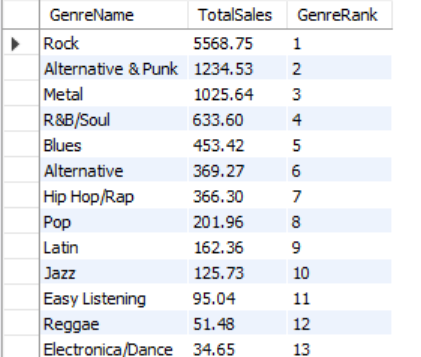
JOIN invoice i ON i.invoice\_id=il.invoice\_id

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

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

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

ORDER BY c.customer\_id;



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

**Approach:**

The goal is to find customers who have not made any purchases in the last three months. There are two tables:

* **Customer table** – contains customer details.
* **Invoice table** – stores invoice dates.

Both tables share a common **Customer\_id**, which makes it easy to connect them. Since only these two tables are needed, no extra data sources are required. The focus is on finding customers whose most recent invoice is older than three months from today.

**Solution:**

To identify inactive customers, the query selects **Customer\_id, First\_name, and Last\_name** from the **Customer table**. A subquery using **NOT EXISTS** helps remove customers who made purchases in the last three months. This subquery checks the **Invoice table** for any recent transactions.

The **WHERE clause** ensures that only customers without an invoice in the last three months appear in the result. The connection between the **Customer table** and the **Invoice table** happens through **Customer\_id**. The **DATE\_SUB** function calculates the date three months before today, helping to define the period of inactivity.

SELECT c.Customer\_id, c.First\_name, c.Last\_name

FROM customer c

WHERE NOT EXISTS (

SELECT 1

FROM invoice i

WHERE i.customer\_id = c.customer\_id

AND i.invoice\_date >= date\_add(curdate(), interval -3 month)

);

****

**Subjective Questions**

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

**Approach:**

The goal was to find three albums from the new record label that should be advertised and promoted in the USA, based on genre sales. To solve this, different tables were used:

* **Artist table** for artist details
* **Genre table** for genre information
* **Invoice table** for sales data
* **Track and Invoice\_Line tables** to connect everything

To make the results clear and readable, column names were given short and meaningful aliases, such as **AlbumName** and **Genre**. The total sales for each genre were calculated using the **SUM** function, and the data was grouped properly using **GROUP BY**. A **WHERE** clause was applied to focus on relevant data.

To identify the three least popular albums, the results were sorted in ascending order. Finally, the output was limited to just three albums.

SELECT am.title AS AlbumName,

g.name AS Genre,

SUM(i.total) AS TotalSales

FROM track t

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

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

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

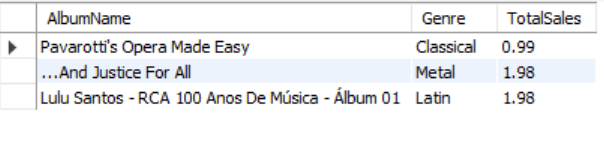
JOIN album am ON t.album\_id = am.album\_id

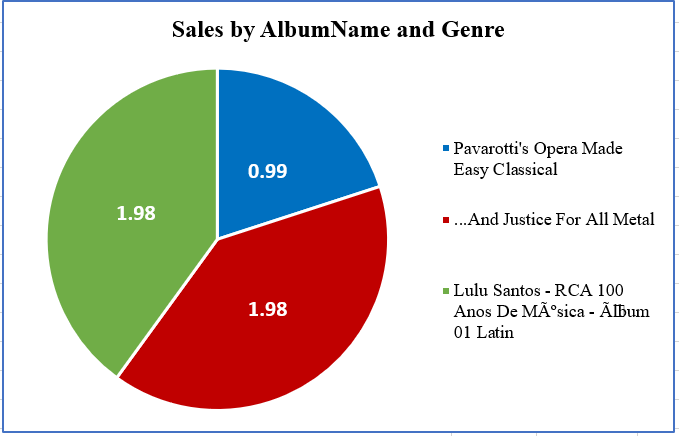
WHERE i.billing\_country = "USA"

GROUP BY am.title, g.name

ORDER BY TotalSales

limit 3;





**Insights:**

* **Sales Distribution:**
  + The album **"Pavarotti's Opera Made Easy" (Classical)** has the lowest sales (0.99).
  + **"...And Justice For All" (Metal)** and **"Lulu Santos - RCA 100 Anos De Música - Álbum 01" (Latin)** both have the highest sales (1.98 each).
* **Genre Performance:**
  + The **Metal and Latin genres** have performed equally well in terms of sales.
  + The **Classical genre** has significantly lower sales compared to the other two.

**Recommendations:**

* **Prioritize Advertising for High-Selling Albums:**
  + Since **Metal and Latin albums** have higher sales, these genres should be the focus for advertising and promotions to maximize revenue.
* **Boost Sales for Classical Music:**
  + The **Classical album has low sales**, indicating a need for better marketing strategies, promotions, or targeting a more specific audience.
  + Special discounts, bundling, or collaborations with streaming platforms could help increase its popularity.
* **Target Audience Analysis:**
  + Further analysis of customer demographics can help in identifying the right audience for each genre.
  + If the Latin and Metal genres have strong demand, consider expanding the catalog in these categories.

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

**Approach:**

To find the top-selling music genres outside the USA, data from different tables was combined. The **Genre** table provided genre names, while the **Invoice** table contained sales and country details. Two other tables, **Track** and **Invoice\_Line**, helped connect them. The **Track** table linked each song to its genre, and the **Invoice\_Line** table linked songs to invoices.

With all four tables joined, sales data for each genre in different countries was calculated. Column aliases, such as **Genre\_Name**, made the results easier to read. The **SUM (total)** function was used to find the total sales for each genre. The **GROUP BY genre\_name, billing\_country** clause grouped the sales by genre for each country. To focus only on international sales, the **WHERE billing\_country <> 'USA'** condition was applied. Finally, the results were sorted in descending order using **ORDER BY total\_sales DESC** to show the most popular genres in each country.

This method provided a clear way to compare which music genres performed best in international markets.

SELECT

g.name as GenreName,

SUM(i.total) as TotalSales

FROM track t

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

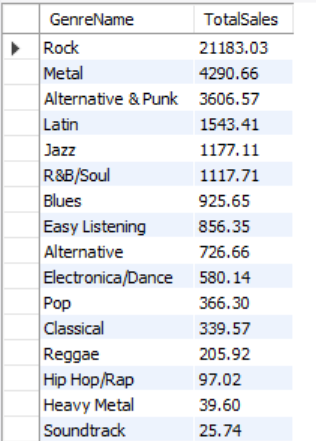
JOIN invoice i ON i.invoice\_id = il.invoice\_id

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

WHERE i.billing\_country <> "USA"

GROUP BY g.name

ORDER BY TotalSales DESC;



**Insights:**

The chart shows that **Rock** is the most popular genre outside the USA, with significantly higher sales compared to other genres. This suggests that Rock has a strong international fan base and drives the majority of music sales in these markets.

**Metal** and **Alternative & Punk** have moderate sales, indicating that while these genres are not as dominant as Rock, they still have a dedicated audience. These genres may have potential for further growth with the right promotional strategies.

On the other hand, **Latin** and **Jazz** have the lowest sales among the top five genres. This suggests that these genres have limited demand in international markets compared to others. Their lower sales numbers indicate either a smaller target audience or less market penetration.

**Recommendations:**

To maximize sales in international markets, a greater focus should be placed on **Rock**, as it significantly outperforms all other genres. Marketing efforts, promotions, and distribution strategies should prioritize this genre to maintain and further increase its sales.

For **Metal** and **Alternative & Punk**, there is potential for growth. Expanding promotional activities, artist collaborations, and regional marketing strategies could help increase their reach and sales in different countries.

Since **Latin** and **Jazz** have the lowest sales, a reassessment of investment in these genres is needed. Strategies could include targeting niche markets where they may perform better or shifting resources towards more profitable genres.

Additionally, a **country-specific analysis** could provide deeper insights into where each genre performs best. This would allow for more tailored strategies, ensuring that music sales align with audience preferences in different regions.

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

**Approach:**

The approach begins by extracting customer purchase history, including customer IDs, first and last purchase dates, total spending, number of items bought, and purchase frequency. To determine customer longevity, the time span between their first and last purchases is calculated in months. Next, the average purchase duration across all customers is computed to serve as a benchmark for categorization. Customers are then classified as "Long-term" if their purchase duration exceeds the average and "Short-term" otherwise. Finally, spending habits, basket sizes, and purchase frequencies of both groups are analyzed to identify patterns in customer retention and loyalty.

WITH Details AS (

SELECT

i.customer\_id,

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

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

TIMESTAMPDIFF (MONTH, MIN(i.invoice\_date), MAX(i.invoice\_date)) AS CustomerDuration,

SUM(i.total) AS TotalSpending,

SUM (il.quantity) AS BasketSize,

COUNT(i.invoice\_id) AS Frequency

FROM invoice i

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

GROUP BY i.customer\_id

),

average\_duration AS (

SELECT AVG(CustomerDuration) AS AvgDuration FROM Details

)

SELECT

CASE

WHEN CustomerDuration > (SELECT AvgDuration FROM average\_duration)

THEN 'Long term Customer'

ELSE 'Short term Customer'

END AS category,

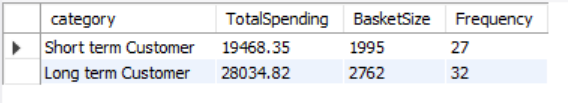
SUM(TotalSpending) AS TotalSpending,

SUM(BasketSize) AS BasketSize,

COUNT(Frequency) AS Frequency

FROM Details

GROUP BY category;



**Insights:**

* **Higher Spending by Long-Term Customers** – Long-term customers have a total spending of **28,034.82,** which is significantly higher than short-term customers (**19,468.35**). This indicates that customers who stay longer tend to contribute more revenue over time.
* **Bigger Basket Size for Long-Term Customers** – The average number of items purchased per transaction is **2,762** for long-term customers, compared to **1,995** for short-term ones. This suggests that loyal customers buy more products per order.
* **Slightly Higher Purchase Frequency** – Long-term customers have a **purchase frequency of 32**, while short-term customers have **27**. While this difference is small, it shows that retaining customers slightly increases their purchase frequency.

**Recommendations for Customer Retention & Growth:**

* **Encourage Repeat Purchases for Short-Term Customers**
  + Offer personalized promotions, discounts, or loyalty rewards to encourage short-term customers to make repeat purchases.
  + Send follow-up emails with targeted product recommendations based on their initial purchases.
* **Enhance Customer Loyalty Programs**
  + Implement tiered loyalty programs where customers unlock better rewards the longer, they stay.
  + Provide exclusive deals or early access to products for long-term customers.
* **Increase Basket Size with Cross-Selling & Upselling**
  + Use bundling strategies to encourage customers to buy complementary products.
  + Offer discounts for bulk purchases or suggest higher-value alternatives during checkout.
* **Analyze Drop-off Points for Short-Term Customers**
  + Identify at what stage short-term customers stop purchasing.
  + Improve customer engagement through better onboarding, customer support, or reminders.

By focusing on these strategies, businesses can increase retention, improve revenue from short-term customers, and maximize the value of long-term customers.

**4. Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?**

**Approach:**

To find which music items are often bought together, the query looks through connected tables in the Chinook database—like tracks, albums, genres, and artists. It links tracks to invoices and checks if a customer bought songs from different genres in one order. If so, it counts how often certain album and artist combinations appear with each genre. By grouping and sorting these combinations, the query shows popular pairings. This helps suggest what albums or artists might interest someone based on what they’ve already bought, making it easier to offer smart recommendations or create music bundles that people are likely to enjoy.

SELECT

g.name AS Purchased\_Genre,

al.title AS Recommended\_Album,

ar.name AS Recommended\_Artist,

COUNT(DISTINCT il.invoice\_id) AS Number\_of\_Copurchases

FROM invoice\_line il

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

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

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

JOIN artist ar ON al.artist\_id = ar.artist\_id

WHERE EXISTS (

SELECT 1

FROM invoice\_line il\_inner

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

WHERE il\_inner.invoice\_id = il.invoice\_id

AND t\_inner.genre\_id <> t.genre\_id

)

GROUP BY Purchased\_Genre, Recommended\_Album, Recommended\_Artist

ORDER BY Purchased\_Genre, Number\_of\_Copurchases DESC;



**Insights:**

* The Singles (50) and International Superhits (47) are top co-purchased albums.
* Green Day (78 total) and Audioslave appear multiple times, showing strong fan loyalty.
* Some albums have low copurchases (e.g., Out of Exile - 10).

**Recommendations:**

* Promote high-performing albums.
* Cross-sell Alternative & Punk.
* Boost lower-selling albums with ads/discounts.

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

**Approach:**

Analyzing regional variations in customer purchasing behavior and churn rates involves integrating data from the **Customer** and **Invoice** tables. The **Customer** table provides location details (country, state, city), while the **Invoice** table contains purchase history, including invoice dates and amounts. By joining these tables on **customer\_id**, customer location is linked to purchase activity.

The query below calculates the **churn rate by country** by grouping invoices by **billing\_country** and counting distinct customers. The **LAG** function determines the year-over-year change in customer counts, and churn rate is calculated as the percentage change. This analysis identifies churn trends across different countries.

SELECT

billing\_country AS Country,

COUNT (DISTINCT customer\_id) AS Num\_customers,

ROUND (COALESCE ((COUNT (DISTINCT customer\_id) - LAG (COUNT(DISTINCT customer\_id))

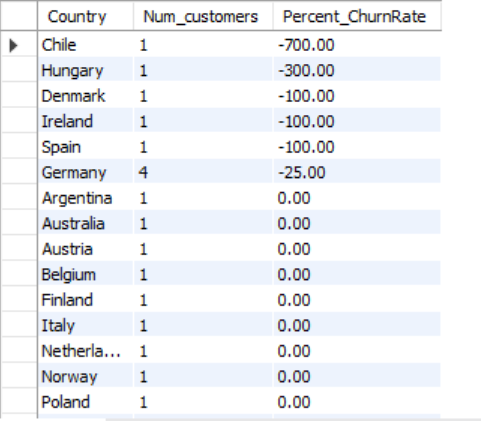
OVER (ORDER BY billing\_country)) / COUNT (DISTINCT customer\_id) \* 100, 0), 2)

AS Percent\_ChurnRate

FROM invoice

GROUP BY billing\_country

ORDER BY Percent\_ChurnRate;



**Insights:**

* **Chile** has the highest churn rate, followed by **Hungary, Denmark, and Ireland**, indicating a significant customer loss.
* **Germany, Argentina, and Austria** show stable churn rates, suggesting customer retention.
* **France, Brazil, and the USA** have **positive churn rates**, indicating growth in customer numbers.
* European countries, especially **Denmark, Spain, and Ireland**, exhibit higher churn, highlighting regional dissatisfaction.

**Recommendations:**

* Investigate high churn regions (**Chile, Hungary**) to identify issues like pricing, service, or competition.
* Enhance retention strategies, including targeted promotions and loyalty programs.
* Expand marketing in growing markets (**France, Brazil**) to sustain positive trends.

**6. Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?**

**Approach**

Customer risk profiling involves analyzing **location, purchase history, and spending behavior** to identify high-churn segments. The query retrieves **customer spending, order frequency, and billing country** by joining the **Invoice** and **Customer** tables. To assess churn risk, segment customers by **billing\_country** and **totalspending** to detect low-revenue regions. Analyzing **numoforders** helps identify declining purchase activity, while **invoice\_date** highlights inactive customers. Incorporating demographics like **age and gender** enhances risk assessment. Comparing churn trends with spending behavior predicts high-risk customers, enabling targeted retention strategies.

SELECT

i.customer\_id,

CONCAT (first\_name, " ", last\_name) AS customer\_name,

billing\_country,

invoice\_date,

SUM (total) AS totalspending,

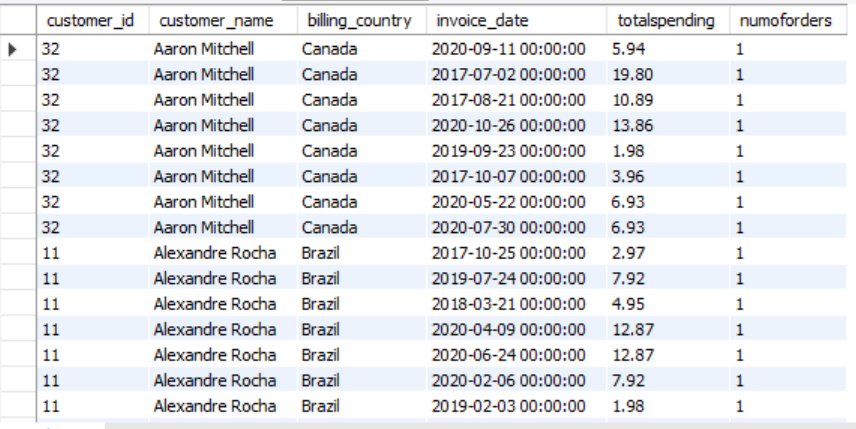
COUNT (invoice\_id) AS numoforders

FROM invoice i

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

GROUP BY i.customer\_id,customer\_name,billing\_country,invoice\_date

ORDER BY customer\_name;



**Insights:**

* **Customer Spending Trends**: Certain customers have significantly higher total spending, while others have lower spending, indicating varied purchasing behavior.
* **Order Frequency**: Some customers make frequent purchases, while others have only a few orders, suggesting different engagement levels.
* **Geographic Impact**: Customers from certain billing countries may exhibit higher spending or order frequency, helping identify key markets.
* **Potential Churn Risk**: Customers with older last invoice dates and fewer orders may be at risk of churning.

**Recommendations:**

* **Target high-spending customers** with loyalty programs to maintain engagement.
* **Re-engage low-order customers** through personalized offers and promotions.
* **Analyze country-wise purchasing trends** to optimize regional marketing strategies.
* **Identify inactive customers** and send retention-focused emails or discounts.

**Insights:**

* After analyzing the sales per month, it is evident that **countries with high spending and frequent orders continue to grow**, while others remain **stagnant or show signs of decline.**
* The chart reflects a **gradual decline in overall sales during the second-to-last quarter** of the year.
* However, a **positive trend is seen in the last quarter**, with a slight rebound in December sales, indicating **potential for recovery.**

**Recommendations:**

* Launch **targeted promotional campaigns** in countries showing stagnant or declining performance to **reduce churn** and **revive customer engagement**.
* Focus on **retaining existing high-value customers** while activating low-frequency buyers through **personalized offers** and **seasonal discounts**.
* Leverage the year-end momentum seen in December by **expanding marketing efforts earlier in Q4**, especially in underperforming regions.
* Use **customer segmentation** (by country, spend level, and order frequency) to better tailor marketing and improve ROI

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

**Approach**

Modeling **Customer Lifetime Value (CLTV)** requires analyzing **customer tenure, purchase history, and engagement** to predict future spending. The **Invoice** table provides **invoice dates and total amounts**, essential for determining **purchase frequency, spending patterns, and customer lifespan**. By grouping invoices by **customer\_id and billing\_country**, key customer-level insights are extracted.

The query calculates **Tenure\_in\_months** (customer lifespan), **Number\_of\_Purchases** (purchase frequency), and **Total\_Spent** (total spending). Customers can be segmented based on **tenure (new, mid-term, long-term), purchase frequency (frequent, occasional, infrequent), and spending (high, medium, low)**. Identifying early churn cases (zero or short tenure) helps uncover retention challenges and refine marketing strategies.

SELECT

customer\_id,

billing\_country AS country,

TIMESTAMPDIFF(MONTH, MIN(invoice\_date), MAX(invoice\_date)) AS customer\_tenure\_months,

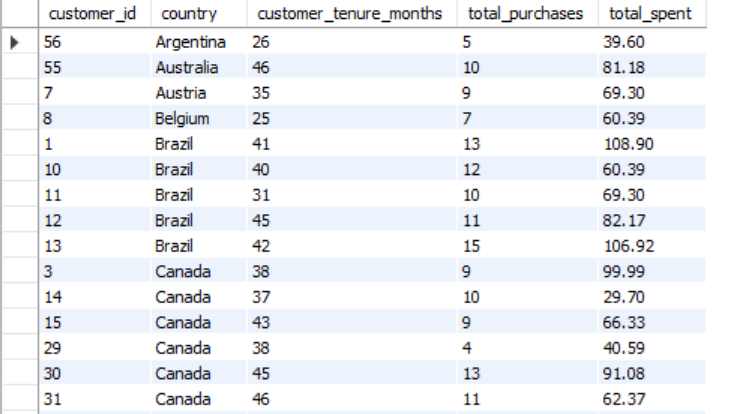
COUNT(DISTINCT invoice\_id) AS total\_purchases,

SUM(total) AS total\_spent

FROM invoice

GROUP BY customer\_id, billing\_country

ORDER BY country, customer\_id;



**Insights:**

* **USA has the highest total customer tenure**, indicating a strong and long-term customer base.
* **Canada, France, and Brazil also have relatively high tenure**, suggesting stable retention in these regions.
* **Countries like Argentina, Chile, Hungary, and Norway have significantly lower tenure**, indicating potential customer churn or limited engagement.
* **European markets show moderate tenure**, suggesting consistent but not highly engaged customers.
* **Countries with lower tenure might have a higher churn rate or newer customer bases.**

**Recommendations:**

* **Strengthen retention efforts in low-tenure countries** by offering loyalty programs, discounts, and personalized recommendations.
* **Analyze high-tenure countries** to identify best practices that can be applied globally.
* **Improve engagement in low-tenure regions** with targeted email campaigns, promotions, or customer service enhancements.
* **Conduct a churn analysis in low-tenure countries** to understand drop-off reasons and optimize marketing strategies.

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

**Impact on New Customer Growth**

* Track the upward trend in new customer registrations and sign-ups over time.
* Assess the engagement levels in events conducted across different regions.
* Examine the rise in email campaign click-through rates following promotional outreach.

**Influence on Customer Loyalty**

* Identify returning customers who re-engage with events or make purchases after a period of inactivity.
* Evaluate how discount offers contribute to customer loyalty and encourage repeat transactions.

**Contribution to Revenue and Sales Dynamics**

* Measure the boost in overall sales attributed to promotional efforts.
* Pinpoint which marketing campaigns generated the highest sales impact.
* Study the correlation between discounts and the onboarding of first-time buyers.

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

**Clarify the Core Business Intent**

* Begin by pinpointing the central aim behind the analysis. Is the business trying to boost sales, attract new customers, retain existing ones, or a combination of these?
* Align the analytical approach to support strategic priorities like revenue acceleration, customer loyalty, or market expansion.

**Evaluate Data Readiness and Relevance**

* Conduct an audit of available datasets, ensuring access to information related to marketing efforts, customer activity, and sales transactions.
* Focus on critical promotional data such as:
  + **Campaign specifics**: Start/end dates, discount percentages, and targeted segments.
  + **Customer engagement signals**: Email open and click rates, event attendance.
  + **Sales behavior**: Frequency of purchases, transaction amounts, time between purchases.
  + **Customer profiles**: Age, location, buying preferences, and past engagement.

**Categorize and Classify the Audience**

* Group customers based on behavioral patterns and lifecycle stage:
  + First-time buyers reached through promos
  + Repeat purchasers incentivized via loyalty rewards
  + Lapsed customers re-engaged through special campaigns
* Study how each group responded to marketing tactics to tailor insights per segment.

**Define Success Indicators**

* **New Customer Acquisition**: Number of fresh customers gained during specific campaigns.
* **Retention Metrics**: Measure repurchase frequency, customer lifetime value (CLV), and churn post-campaign.
* **Sales Impact Analysis**: Track total revenue, average order size, and shifts in purchasing trends.
* **Engagement Performance**: Examine key indicators like open rate, CTR (click-through rate), and actual conversions.

**Conduct a Multi-Level Data Analysis**

* **Temporal Analysis**: Observe how metrics fluctuate across pre-, during-, and post-campaign timeframes.
* **Cohort Insights**: Study customer behavior based on the time they entered the funnel or made their first purchase.
* **Controlled Testing**: Leverage A/B testing to compare results between campaign recipients and non-recipients.

**Detect High-Value Segments**

* Identify which customer groups responded most positively to specific campaigns.
* Use insights to design more precise and effective promotions for each segment moving forward.

**Develop Visual Narratives for Stakeholders**

* Create dashboards and graphs to visually depict trends, patterns, and key metrics like retention rate, customer churn, or sales spikes.
* Summarize impactful insights—what worked, what didn’t, and how each campaign influenced specific customer behaviors.

**Inform Strategic Optimization**

* Use analysis outcomes to propose smarter campaign strategies:
  + **Target more accurately**: Focus efforts on the most promising customer groups.
  + **Time strategically**: Launch promotions when customers are most likely to convert.
  + **Channel prioritization**: Allocate more budget to effective channels (e.g., email, SMS, social ads).

**Adopt a Continuous Improvement Mindset**

* Make campaign evaluation an ongoing cycle. Test, learn, and iterate by adjusting promotional tactics based on observed outcomes and behavioral changes.
* Set up regular feedback mechanisms to ensure marketing strategies evolve with shifting customer expectations.

**Conclusion:**  
When specific objective or subjective questions aren’t provided, my approach would center on grasping the core business objectives, assessing the data landscape, and identifying the right performance indicators to evaluate promotional success. By examining customer behavior patterns and tracking key metrics over time, I can generate meaningful insights that inform smarter, data-driven promotional strategies for continuous improvement.

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

To modify the "Albums" table by adding a new column for storing the release year of each album, the ALTER TABLE statement can be used. Specifically, the command ALTER TABLE Albums ADD COLUMN ReleaseYear INTEGER; will add a new column named "ReleaseYear" with the data type INTEGER. This allows the table to store the year each album was released, enhancing the dataset with temporal information that can be useful for sorting, filtering, or analysis purposes.

ALTER TABLE Album

ADD COLUMN ReleaseYear INTEGER;

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

**Approach:**

To analyze customer purchasing behavior by geographical location in the Chinook database, data from the Customer, Invoice, and Invoice\_Line tables is utilized. The Customer table provides location details, while Invoice and Invoice\_Line contain purchase and track-level data. A Common Table Expression (CTE) named customer\_purchases is used to calculate total spending and the number of tracks purchased per customer by joining these tables. The final query groups result by country, calculating the number of customers, average spending, and average tracks purchased per customer. This method offers insights into regional purchasing trends and customer engagement across various countries.

WITH cp AS (

SELECT

cu.country,

cu.customer\_id,

SUM(inv.total) AS total\_spent,

COUNT(DISTINCT il.track\_id) AS tracks\_purchased

FROM customer cu

LEFT JOIN invoice inv ON cu.customer\_id = inv.customer\_id

LEFT JOIN invoice\_line il ON inv.invoice\_id = il.invoice\_id

GROUP BY cu.country, cu.customer\_id

)

SELECT

cp.country AS Country,

COUNT(DISTINCT cp.customer\_id) AS No\_of\_Customers,

ROUND(AVG(cp.total\_spent), 2) AS Avg\_Spent,

ROUND(AVG(cp.tracks\_purchased), 2) AS Avg\_Tracks\_Purchased

FROM cp

GROUP BY cp.country

ORDER BY cp.country;



**Insights:**

* **High-Spending Countries:**
  + The **Czech Republic** and **Ireland** show the highest average spending (red bars). These countries may have customers who purchase more expensive items or make frequent purchases.
  + **India, Australia, and Canada** also exhibit high spending patterns compared to other regions.
* **Number of Customers:**
  + **USA and Canada** have a relatively high number of customers (blue bars), indicating strong customer engagement.
  + **Argentina, Australia, and Spain** also show a decent customer base.
* **Track Purchases:**
  + The green bars, representing the average number of tracks purchased, are relatively small across all countries, suggesting that customers tend to buy a limited number of tracks per transaction.
  + Countries with high spending do not necessarily have higher track purchases, meaning purchases might be driven by factors other than volume (e.g., higher-priced items or subscriptions).
* **Countries with Lower Engagement:**
  + **Chile, Denmark, Netherlands, and Norway** have low spending and fewer customers, indicating a weaker market presence.

**Recommendations:**

* **Market Expansion Strategies:**
  + Since **Czech Republic and Ireland** have high spending, increasing marketing efforts and exclusive promotions in these regions could further boost revenue.
  + **Countries with lower customer bases but decent spending (e.g., India, Australia)** could benefit from targeted campaigns to attract more users.
* **Promote Track Sales:**
  + The low **Avg\_Tracks\_Purchased** suggests that customers buy fewer tracks per transaction. Offering **bundle deals, discounts, or subscriptions** might increase track purchases.
* **Optimize for Low-Engagement Countries:**
  + **Chile, Denmark, and Norway** show weak customer numbers and spending. Localized marketing strategies, discounts, or partnerships might help improve engagement in these regions.
* **Customer Retention & Loyalty Programs:**
  + Since high-spending customers drive revenue, launching **loyalty programs, exclusive discounts, or personalized recommendations** can help maintain engagement and increase lifetime value.