***Objective Questions***

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

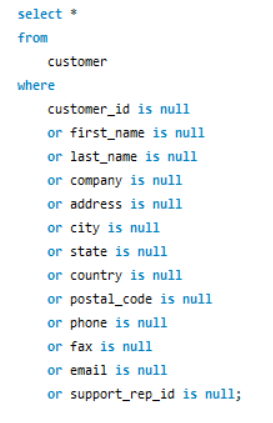
Ans:

**Approach:**

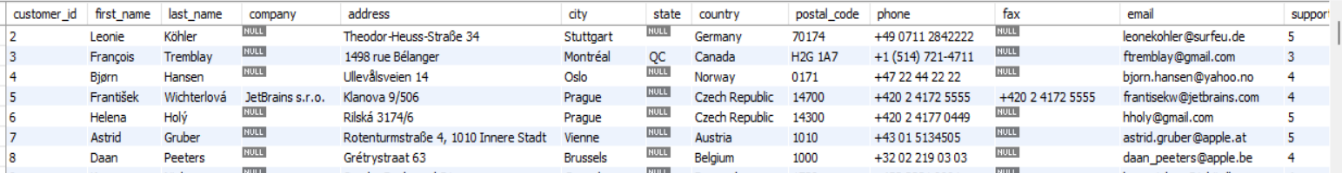
* To identify whether contains missing value, we used SQL queries with **‘is null’** and aggregation functions.
* Missing values can be managed using NULL handling functions like **coalesce().**

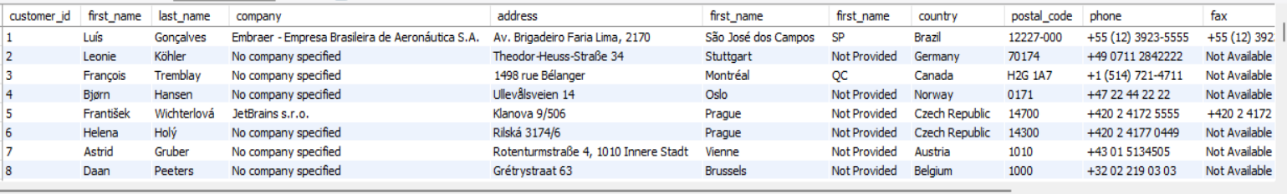
1. **Customer Table:**

* We inspected each column in the customer table using a query with multiple **‘is null’** checks.
* The output showed that several rows have **null** values in columns.
* Instead of deleting these records, we replaced **nulls** with **default text values** using the **coalesce() function.**
* Output shows that column **‘company’, ‘address’, ’city’ ,’state’ ,’postal code’ , ’phone’**and **‘fax’** contains **null** values.

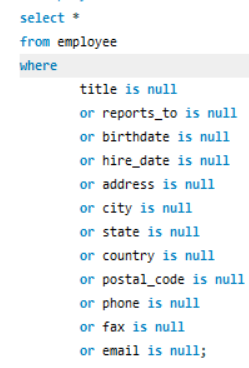
**Output:**

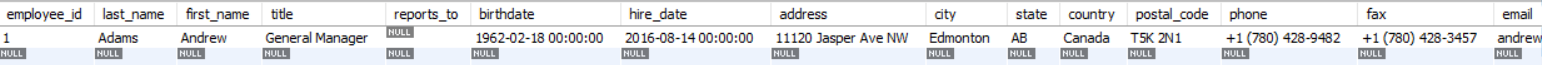
****

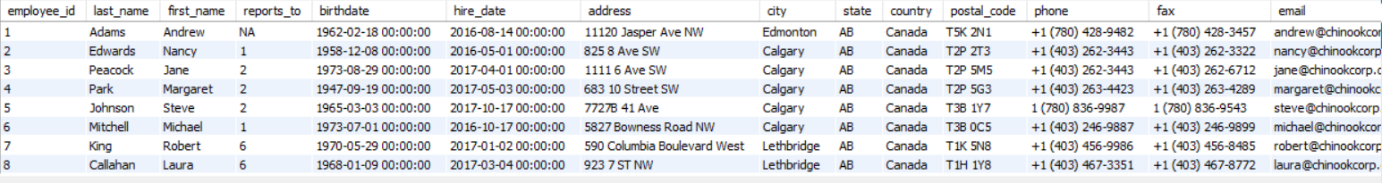


1. **Employee Table:**

* Checked for **null** values.
* Output showed only one row and in a one column **‘reports\_to’** has a null value.
* We handle the missing or we can say null value using **coalesce() function** and passing a default message **NA.**

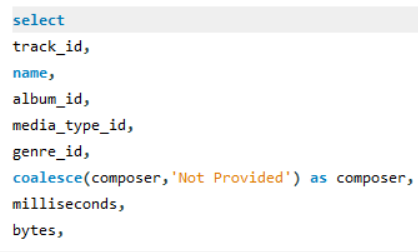
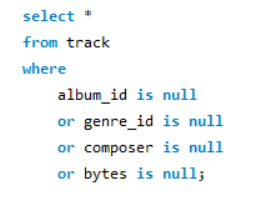
** **

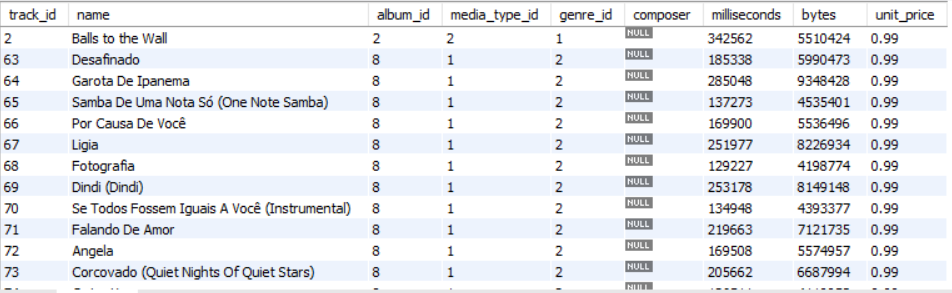
** Output:**

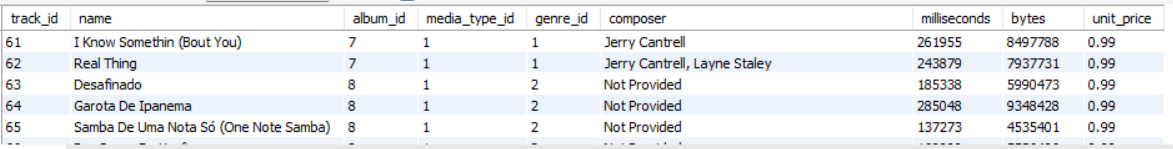
****

**3. Track:**

* Checked the **null values** or **missing values** from Track table.
* We inspected each column using **is null**.
* Using **coalesce() function** replace the null values and passing a default message**.**



**Output: **

****

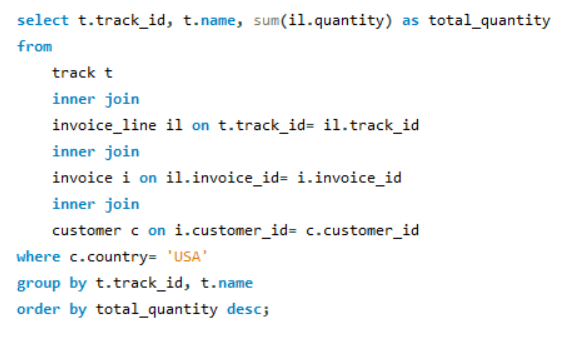
**2. Find the top-selling tracks and top artist in the USA and identify their most famous genres.**

**Ans:**

**Top selling tracks:**

**Approach:**

* We use join function(inner join), aggregate function(sum), filter function, group by and order by to get top selling tracks.



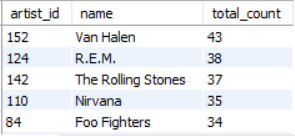
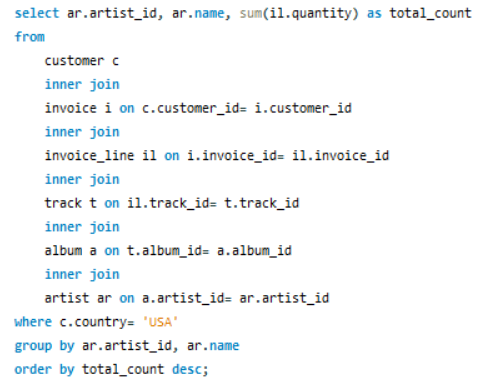
****

**Output:**

* The output display the top-selling tracks in the USA based on total quantity sold. In output table shows the track\_id, name(track name), and total\_quantity.
* And the output shows arrange in descending order according to total quantity.

**Top Artist**:

* We use join function(inner join), aggregate function(sum), filter function, group by and order by to get top selling tracks.

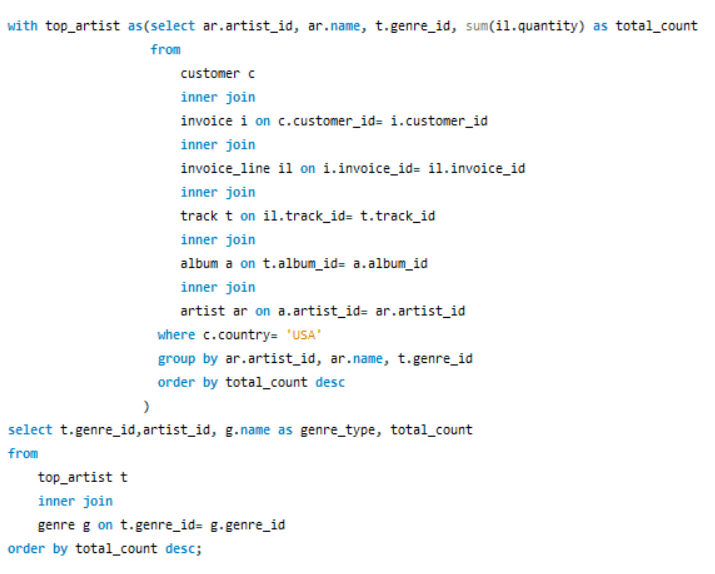
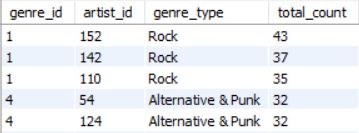


**Output:**

* The output display the top artist tracks in the USA based on total quantity sold. In output table shows the artist\_id, name(track name), and total\_count.
* And the output shows arrange in descending order according to count.

**Famous Genre:**

* We use CTE(common table expression), aggregate, join function, filter function, order by and group by to get the top famous genre.

****

**Output:**

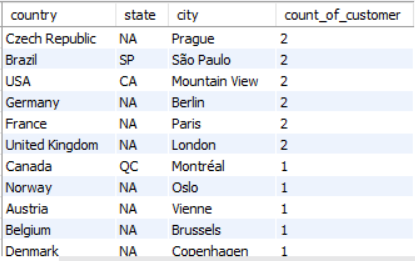
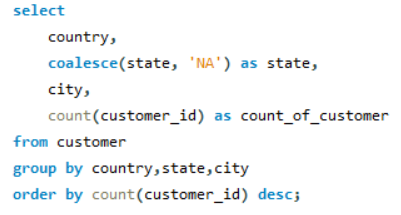
* The output display the top genre in the USA based on total quantity sold. In output table shows the genre\_id, artist\_id, genre\_type, total\_count.

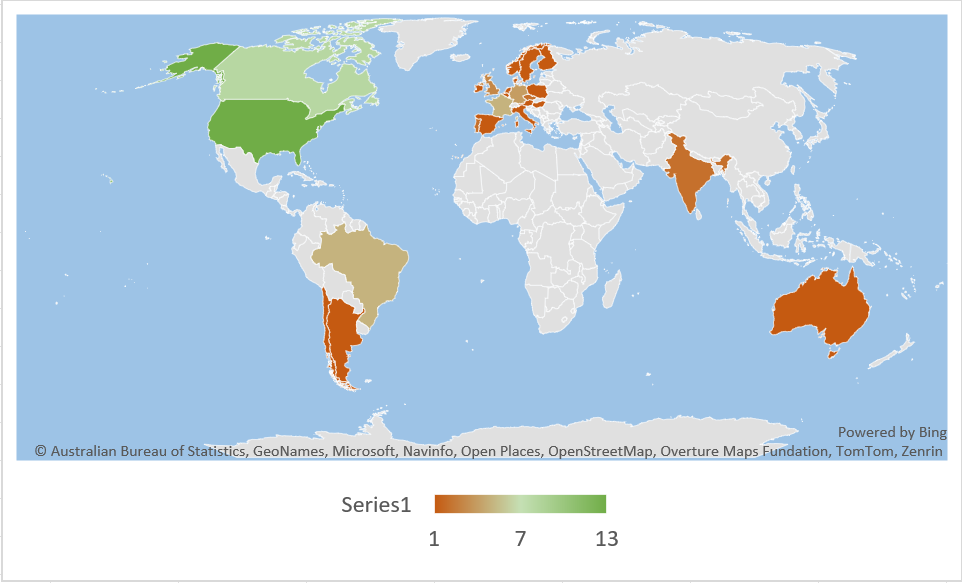
**3. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?**

**Ans:**

**Approach:**

* Missing values in the state column were handled using coalesce(state, 'NA') to avoid grouping issues
* We used group by on country, state, and city fields to count how many customers are located in each region.
* The count(customer\_id) function was applied to find the total number of customers in each city
* The results were ordered by customer count in descending order to highlight the most populated regions.



****

**Output:**

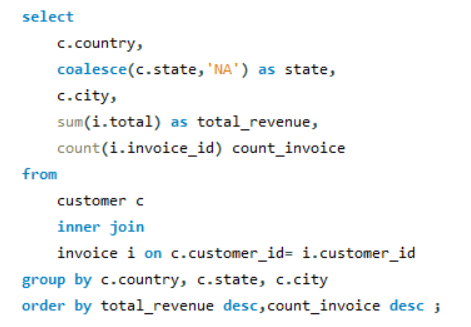
* We analyzed the customer distribution by **grouping data based** on country, state, and city using aggregation functions, handled null values with **coalesce**, and identified the regions with the highest customer concentrations.
* The dataset does not contain age or gender information, so the demographic breakdown is limited to location-based insights only.

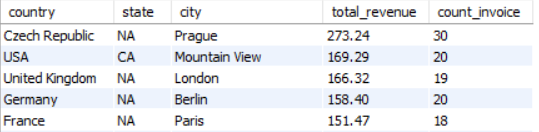
**4. Calculate the total revenue and number of invoices for each country, state, and city:**

**Ans:**

**Approach:**

* Performed an inner join between the customer and invoice tables to link customer location with their invoice data.
* Aggregated total revenue and invoice count per city using **sum** and **count**, while handling missing state values using coalesce.





**Output:**

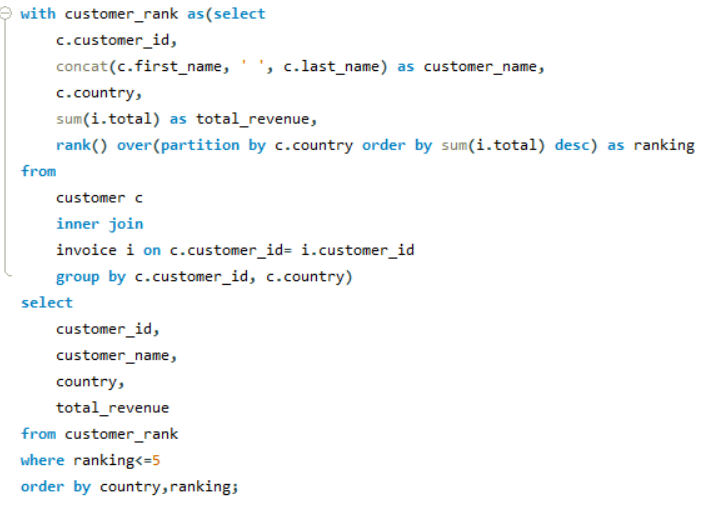
* Joined customer and invoice tables to calculate total revenue and invoice count based on location details.
* Grouped the results by country, state, and city, and ordered them to identify top-performing regions by revenue.
* Handled missing state values using the coalesce function to ensure accurate grouping and consistent output.

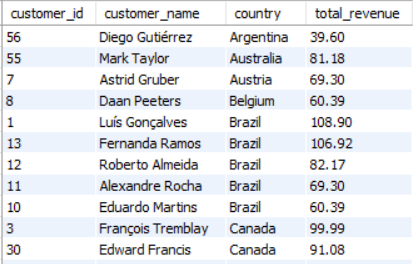
**5. Find the top 5 customers by total revenue in each country.**

**Ans:**

**Approach:**

* Performed an inner join between the customer and invoice tables to link each customer with their transactions.
* Used sum() to calculate total revenue generated by each customer.
* Applied the rank() window function to rank customers within each country based on their total revenue in descending order.
* Filtered the result to return only the top 5 customers per country and sorted the output by country and rank.





**Output:**

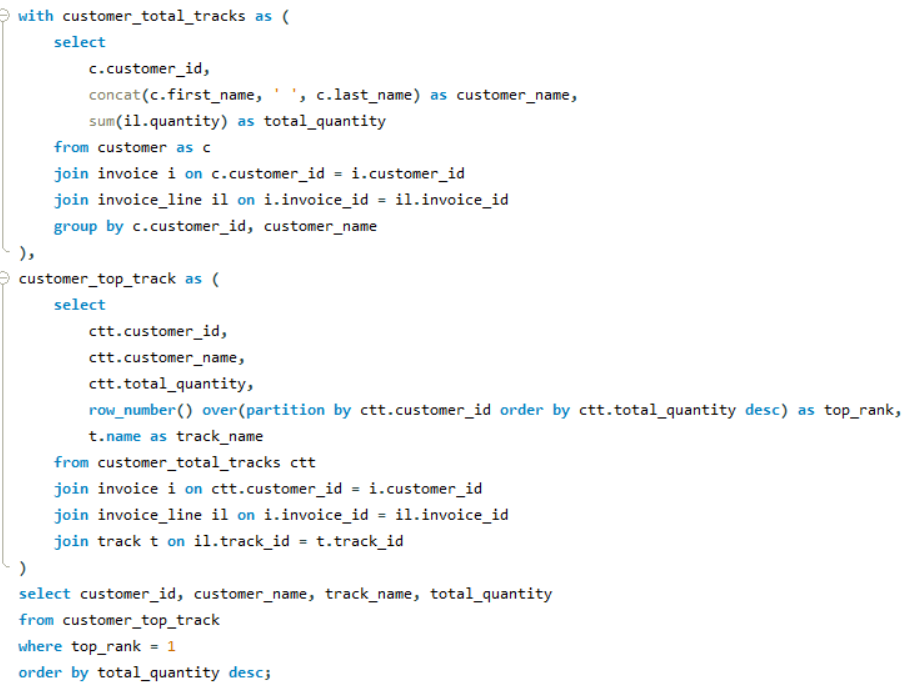
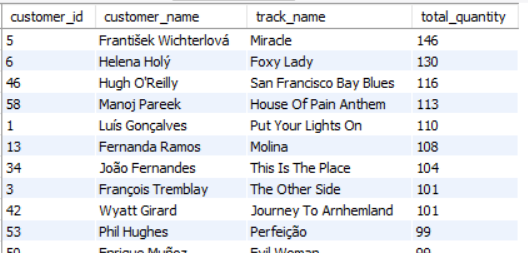
* The result displays the top 5 customers by revenue from each country.
* Customers are listed along with their names, countries, and total revenue generated.
* The data is sorted first by country and then by revenue rank within each country.
* This helps identify the most valuable customers per region for targeted business insights.

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

**Ans:**

**Approach:**

* Data Integration: Join customer, invoice, and invoice\_line tables to collect purchase data including quantities and prices.
* CTE Usage: Use a single CTE to aggregate total tracks and spend per customer by country.
* Result Analysis: Group by country, calculate averages, and sort by spend to highlight top regions.

****

**Output:**

* The result summary presents an overview of purchasing behaviour across various regions, highlighting the total number of customers and tracks purchased per region.
* It includes average metrics for tracks and spending per customer, with results sorted to emphasize regions with the highest average spend, aiding strategic insights.

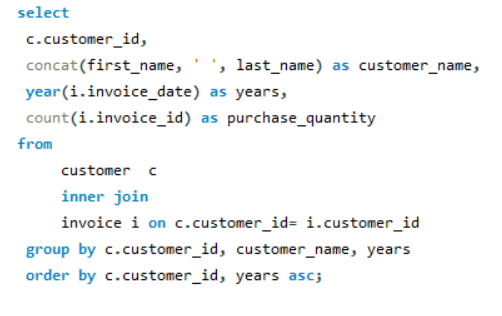
**7. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?**

**Ans:**

**Approach:**

**Analyse yearly customer purchase activity using SQL joins and grouping.**

* **Year wise purchase count:**
* Performed an inner join between the customer and invoice tables to extract customer\_id, full name, and invoice year. Applied count(invoice\_id) to compute total purchases per customer each year. Grouped by customer\_id, customer\_name, and year, and sorted results in ascending order.



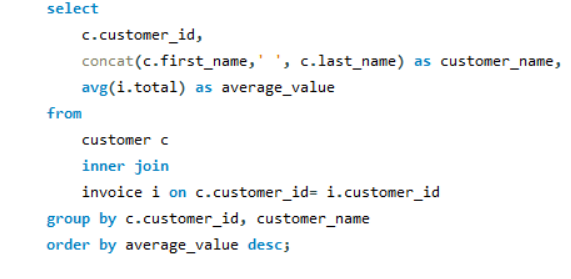
**Output:**

* Each customer has purchase records from 2017 to 2020 with varying yearly quantities.
* Overall purchase quantity shows a decreasing trend toward 2020 across all customers.

**Top Customers by Average Invoice Value:**

**Approach:**

* Join the customer and invoice tables using the customer\_id key.
* Concatenate first and last names to get the full customer name.
* Calculate the average invoice amount (avg(i.total)) per customer.
* Group the results by customer and sort them in descending order of average invoice value.



**Output:**

* The query returns each customer's ID, full name, and their average invoice value.
* Customers are listed in descending order based on their average spending.

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

**Ans:**

**Approach to Calculate Churn Rate (1-Year Interval):**

* **Identify purchase years per customer**Use the invoice table to extract the year of each transaction and group them by customer. This gives you a list of active years per customer.
* **Find the latest consecutive years**Determine the most recent pair of back-to-back years where customer activity occurred. This focuses the churn analysis on the latest available period.
* **Check customer activity across years**  
  Create two customer groups one active in the earlier year and another in the following year. Then compare them to find out who did not return in the next year.
* **Calculate churn rate**  
  Divide the number of customers who left like who active in the first year but not the next by the total number of customers in the first year. Multiply by 100 to get the churn rate percentage.

with invoice\_years as(select

customer\_id,

year(invoice\_date) as year

from

invoice

group by customer\_id, year

),

year\_pairs as (select distinct

prev.year as prev\_year,

next.year as next\_year

from invoice\_years prev

join invoice\_years next on next.year = prev.year + 1

),

latest\_pair as (select \*

from

year\_pairs

order by prev\_year desc

limit 1

),

active\_prev\_year as (select customer\_id

from invoice\_years

where year = (select prev\_year from latest\_pair)

),

active\_next\_year as (select

customer\_id

from invoice\_years

where year = (select next\_year from latest\_pair)

),

churned\_customers as (select

customer\_id

from active\_prev\_year

**** where customer\_id not in (select customer\_id from active\_next\_year)

)

select

(select count(\*) from churned\_customers) \* 100.0 /

(select count(\*) from active\_prev\_year) as churn\_rate\_percent;

**Output:**

* The churn rate of 1.69% is indicate that the company losses 1.69% customer over a periods(year or month). Generally, a lower churn rate is better, suggesting that a business is effectively retaining its customers.

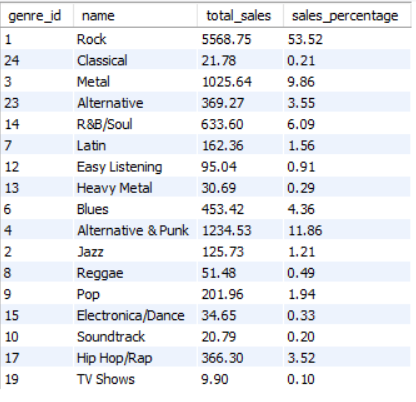
**9. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.**

**Ans:**

**Genre-wise sales percentage analysis for USA customers:**

**Approach:**

* **Data Preparation through Joins:**  
  Multiple tables (customer, invoice, invoice\_line, track, genre) are joined to connect each customer’s purchase to the corresponding music genre. This helps trace the complete path from customer to genre-level sales.
* **Filter by Target Region:**  
  Only customers from the USA are considered using a where clause. This ensures the analysis focuses on regional buying behavior and preferences.
* **Genre-level Sales Aggregation:**  
  After filtering, sales (total) are aggregated by genre\_id and genre\_name using sum() and group by. This gives total sales for each genre within the USA.
* **Sales Contribution Calculation:**  
  A separate subquery calculates total sales across all genres. Each genre's sales are then divided by this total to find the percentage contribution, rounded to two decimal places.

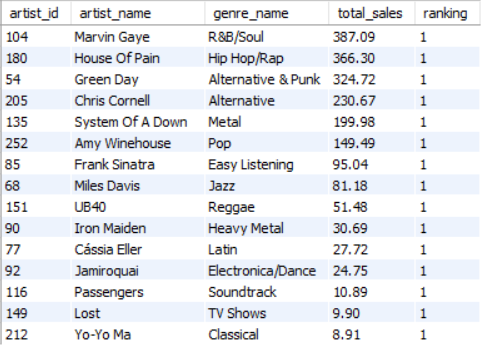
****

**Top selling artists by genre in the USA:**

**Approach:**

* **Join Necessary Tables:**Combine customer, invoice, invoice\_line, track, genre, album, and artist tables to access relevant sales, track, genre, and artist information.
* **Filter by Country:**  
  Focus only on customers from the USA to analyse domestic music consumption patterns.
* **Aggregate Sales:**  
  Calculate total sales for each artist within every genre using sum(i.total) and group the results by artist and genre.
* **Rank Artists within Each Genre:**  
  Use dense\_rank() to assign a rank based on sales within each genre, allowing identification of top artists per genre.
* **Select Top Artists per Genre:**  
  Filter results to show only those artists who ranked 1st in their genre, and sort the final list by total sales in descending order.



****

**Output:**

The result shows the top-selling artist for each genre in the USA, ranked by their total sales, with all artists holding the number one position within their respective genres.

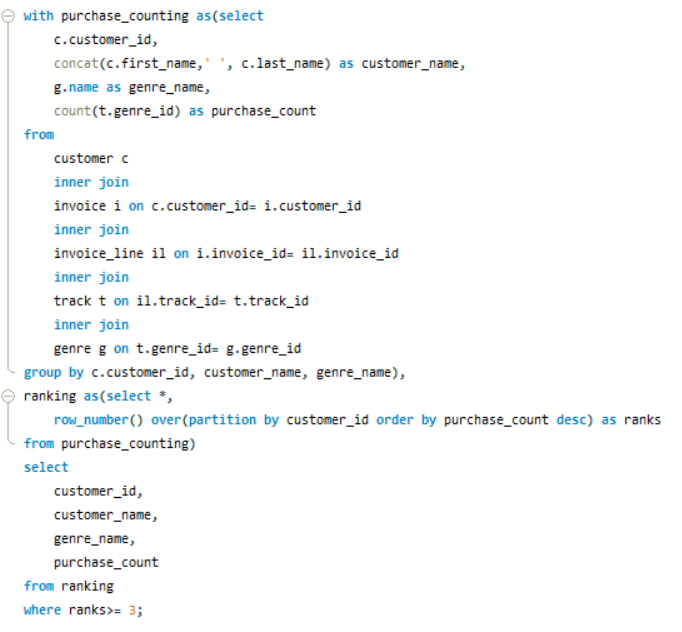
**10. Find customers who have purchased tracks from at least 3 different genres.**

**Ans:**

**Customer who purchase at least 3 genre:**

**Approach:**

* Counted how many different genres each customer purchased tracks from.
* Used row\_number() to rank genres by purchase frequency for each customer
* Selected only those customers who have purchased from 3 or more genres, filtering rows where the rank is greater than or equal to 3.



**Output:**

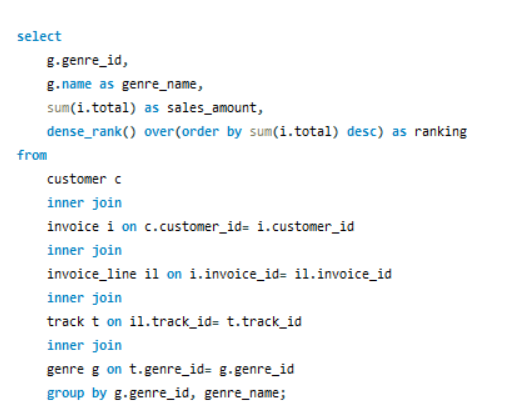
The output displays customers who have purchased tracks from three or more distinct music genres. Each record includes the customer ID, name, the genre they purchased from, and the total number of purchases within that genre. This helps identify customers with diverse musical preferences.

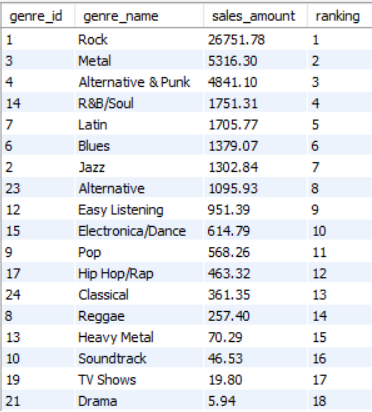
**11. Rank genres based on their sales performance in the USA**

**Ans:**

**Approach:**

* Join multiple tables including customer, invoice, invoice\_line, track, and genre to connect customer purchases with music genres.
* Aggregate sales by summing up the total invoice amounts (i.total) for each genre.
* Rank genres using dense\_rank() based on total sales in descending order to identify the top-performing genres.
* Group the result by genre ID and genre name to show total sales and rank for each genre.



****

**Result:**

The analysis reveals that Rock is the highest-grossing genre, followed by Metal and Alternative & Punk in terms of total sales revenue. Less popular genres like TV Shows and Drama show minimal sales, indicating lower customer preference. This ranking helps identify customer interests and genre profitability.

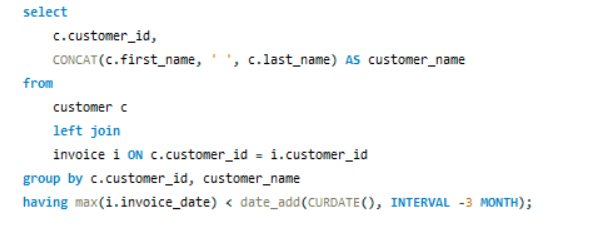
**12.Identify customers who have not made a purchase in the last 3 months**

**Ans:**

**Approach:**

* **Join Customer with Invoice Table**  
  Used a left join to connect each customer with their purchase records, ensuring customers with no recent activity are still included.
* **Group by Customer**  
  Grouped the data by customer ID and full name to evaluate purchases at the individual level.
* **Calculate Latest Purchase Date**  
  Used max(invoice\_date) to determine the most recent purchase date for each customer.
* **Filter Inactive Customers**  
  Applied a condition to return only those customers whose last purchase was more than 3 months ago from the current date.





**Result:**

A list of customers was identified who have not made any purchases in the last 3 months, indicating potential churn or inactivity.

***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.**

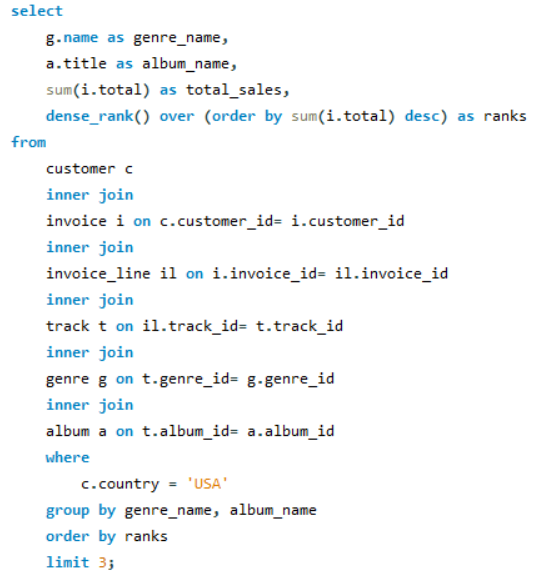
**Ans:**

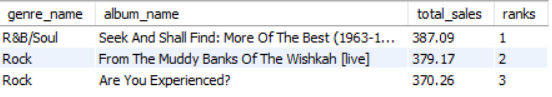
**Objectives:**

To identify the top 3 albums based on total sales in the USA, grouped by genre and album, to help the new record label prioritize which albums to promote.

**Approach:**

* **Data Source and Filtering**
* Queried sales data by joining tables: customer, invoice, invoice\_line, track, genre, and album.
* Applied a filter to only include **customers from the USA** (c.country = 'USA').
* **Sales Aggregation and Ranking**
* Calculated total sales per **genre and album** using sum(i.total).
* Ranked albums using dense\_rank() based on total sales in **descending** order (highest sales ranked first).
* **Selection of Top Albums:**
* Grouped by genre\_name and album\_name, then selected the top 3 albums using LIMIT .





**Insights:**

The sales analysis based on US customer purchases reveals that albums from the R&B/Soul and Rock genres are performing exceptionally well. The album *“Seek And Shall Find: More Of The Best”* from the R&B/Soul genre has achieved the highest sales, followed closely by two Rock albums – *“From The Muddy Banks Of The Wishkah [live]”* and *“Are You Experienced?”*. This indicates that these genres have a strong listener base in the USA, with consistent demand reflected in their sales volume and ranking.

**Recommendation:**

To maximize impact in the US market, it is recommended to prioritize the promotion of these three top-selling albums. Marketing campaigns should focus on these genres, particularly R&B/Soul and Rock, as they have already shown proven engagement. By allocating resources to boost the visibility of these high-performing albums, the new record label can effectively capture the interest of a receptive audience and drive further growth in sales.

**Final Suggestion:**

The record label should prioritize advertising the top three albums listed above, focusing especially on R&B/Soul and Rock genres, which are performing well in the USA market based on sales data.

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

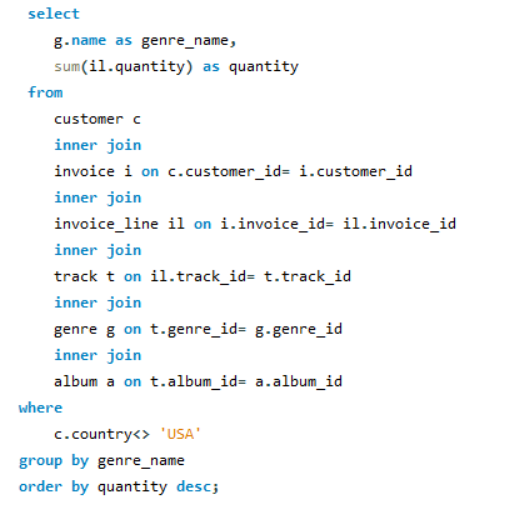
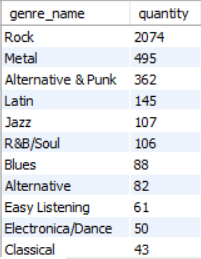
**Ans:**

**Objective:**

The goal was to identify which music genres are most popular among customers outside the USA, using sales data. This helps in understanding regional preferences and planning targeted marketing or inventory strategies.

**Approach:**

* **Data Selection and Filtering**:  
  We used data from multiple related tables—customer, invoice, invoice\_line, track, genre, and album—by performing **inner joins**. The filter c.country <> 'USA' was applied to **exclude U.S. customers**, focusing only on international sales.
* **Aggregation Logic**:  
  The query grouped data by genre\_name and summed up the quantity of tracks sold using SUM(il.quantity). This aggregation allowed us to calculate **total sales per genre** from international customers.
* **Sorting for Top Performers**:  
  The result was ordered in **descending order** of total quantity sold (order by quantity desc) to highlight the **most popular genres** among international audiences.
* **Interpretation of Output**:  
  The final output shows which genres are in high demand outside the US. This insight can be used for **strategic decisions** like recommending more Rock or Metal tracks to global users, or exploring why certain genres underperform in non-U.S. markets.



**Insights:**

Based on the data from countries other than the USA, Rock is the most purchased genre with (2,074 units), followed by Metal (495 units) and Alternative & Punk (362 units). These top three genres show a clear preference for high-energy, guitar-based music among international customers. On the other hand, genres like Classical (43 units), Electronica/Dance (50 units), and Easy Listening (61 units) have the lowest sales, indicating relatively less interest outside the USA.

**Recommendations:**

* **Prioritize Rock Genre:** Since Rock has emerged as the top-performing genre across multiple international markets, promotional campaigns should emphasize Rock albums to capture wider audience interest.
* **Diversify Genre Promotions:** Alongside Rock, include genres like Pop, Latin, and Metal in promotional efforts, as these genres also show strong performance and can help tap into varied listener preferences.
* **Optimize Ad Targeting by Genre Popularity:** Use genre-level sales data to guide digital ad placement strategies, ensuring that each market receives content aligned with its top genre preferences.
* **Align Album Releases with Genre Demand:** Schedule and promote album releases in line with the most favoured genres in each region to improve engagement and return on investment.

**Final Suggestion:**

To maximize sales and global reach, the label should strategically focus its marketing efforts on genres that consistently perform well across different regions. By leveraging data-driven insights on genre popularity, the label can optimize promotional activities, tailor content to audience preferences, and increase the effectiveness of its campaigns—ultimately leading to better market penetration and higher revenue.

**3. 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?**

**Ans:**

**Objective:**

We analyse customer purchase habits by segmenting them into 'recent' (less than 1 year) and 'long-term' (1 year or more) based on their purchase history. For each segment, we calculated average purchase frequency, basket size (total products per customer), spending amount, and average order value. Long-term customers generally showed higher purchase frequency and total spending, indicating stronger engagement. Recent customers, on the other hand, had slightly smaller basket sizes and lower overall spending. This analysis helps identify loyal customers and tailor marketing strategies accordingly.

**Approach:**

**Join Core Tables:**

* Joined the customer, invoice, and invoice\_line tables to collect transactional data, linking customer\_ids to purchases and product quantities.

**Calculate Key Metrics per Customer:**

* Aggregated data at the customer level to compute:
  + Number of purchases (purchase\_quantity)
  + Total products bought (total\_product\_purchase)
  + Total amount spent (total\_spent)
  + Average order value (avg\_spent\_per\_order)
  + Customer lifetime in days (customer\_lifetime\_days)

**Classify Customers:**

* Created a segment using a case statement:
  + Customers with a lifetime under 365 days labelled as 'recent'
  + Others labelled as 'long-term'

**Create Customer Segmentation Table:**

* Stored all metrics along with the customer category into a CTE named customer\_segment for clean modular analysis.

**Aggregate by Customer Category:**

* Grouped the data by customer\_category to analyse segment-level behaviour.

**Calculate Segment-Level Averages:**

* Computed:
  + Average purchase frequency
  + Average basket size
  + Average total spending
  + Average order value

**Formatted Output:**

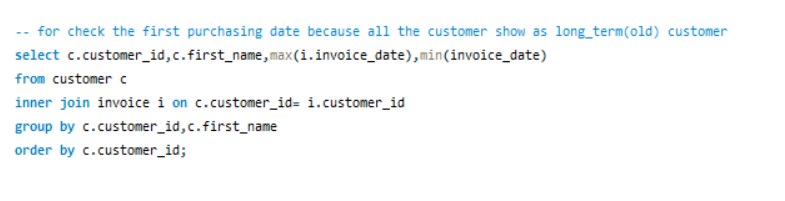
* Used the round() function to keep the result readable and suitable for reporting dashboards or business insights.

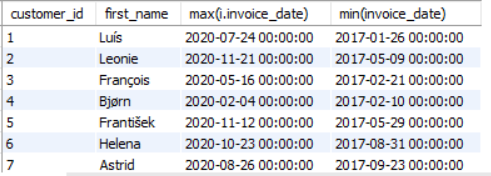


****

**Ensuring Accurate Lifetime Days Calculation Based on Initial Purchase:**

To accurately segment customers, check each customer's first purchase date using min(invoice\_date) instead of relying solely on the full date range. This ensures the customer\_lifetime\_days reflects the time since their first purchase, not just the dataset range.





**Insights:**

* **Only Long-Term Customers Identified:**  
  The dataset currently contains only long-term customers (with a lifetime over 1 year); no recent customers were found, indicating either missing new customer data or outdated segmentation logic.
* **High Engagement from Long-Term Customers:**  
  Long-term customers show high purchase frequency and strong basket sizes, suggesting consistent engagement and brand loyalty.
* **Steady Revenue Contribution:**  
  These customers also contribute significantly to total revenue, although their average order value remains modest, indicating frequent but smaller purchases.
* **Implications for Loyalty Strategy:**  
  The presence of only long-term customers reflects a loyal customer base, but also highlights a **lack of new customer acquisition**, which may pose long-term growth risks.
* **Retention vs. Acquisition Balance:**  
  The business should continue nurturing existing customers while focusing on strategies to **attract and convert new users** to maintain sustainable growth.

**Result Summary:**

* **Customer Category:** Long-Term
* **Average Purchase Frequency:** 80.63
* **Average Basket Size:** 80.63 products
* **Average Total Spending per Customer:** ₹805.14
* **Average Order Value:** ₹9.77

**Recommendation:**

**No New Customers Identified:**

* The current dataset shows only long-term customers, indicating that no new (recent) customers have been recorded in the data within the last 12 months.

**Review Customer Segmentation Logic:**

* Recalculate customer\_lifetime\_days using each customer’s first purchase date min(invoice\_date) and compare it with the dataset’s maximum invoice date to correctly identify recent customers.

**Check Data Recency:**

* Ensure the dataset includes recent transactions. If the data is outdated, update it to include more recent customer activity.

**Introduce Time-Based Filters:**

* Use a consistent reference date (e.g., max(invoice\_date)) instead of datediff(max, min) across all customers to avoid skewed segmentation.

**Enable Comparative Analysis:**

* Once new customers are identified correctly, analyse and compare their purchasing behaviour with long-term customers for better business decisions.

**Investigate Acquisition Strategy:**

* If genuinely no new customers exist, the business should review its customer acquisition efforts and explore new channels or promotions.

**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?**

**Ans:**

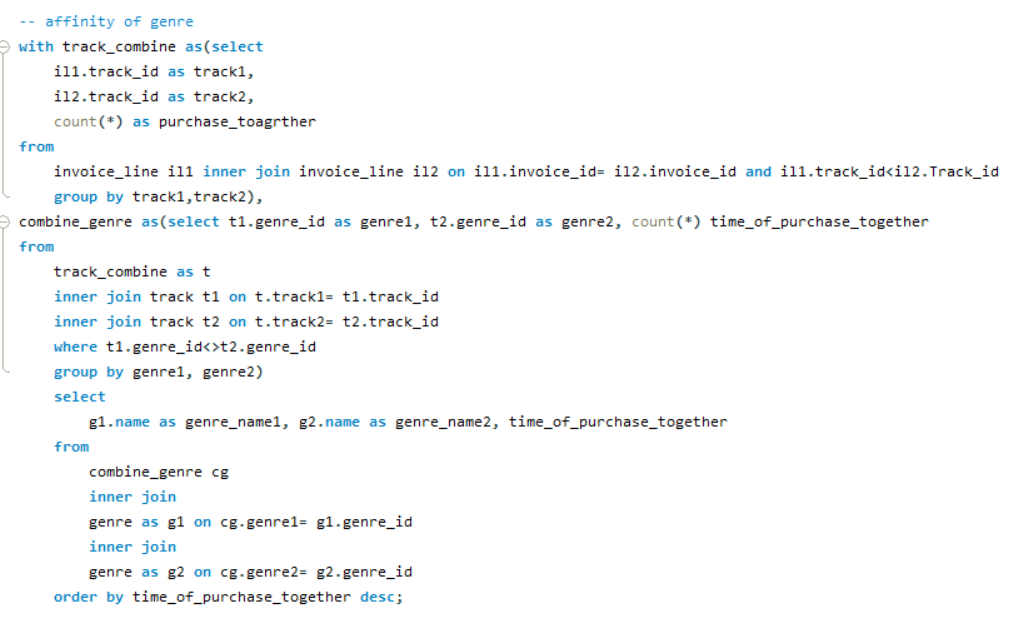
**Objective:**

The objective of the SQL queries is to analyse the purchasing patterns of tracks, albums, and genres to identify combinations frequently bought together, providing insights into customer preferences and affinities among artists, albums, and genres.

**Approach:**

* The SQL queries analyse affinity by combining data from invoice\_line, track, album, and artist or genre tables using inner joins to identify pairs of tracks, albums, or genres purchased together.
* The results are grouped by the combined pairs, counted for frequency (time\_of\_purchase\_together), and ordered by this count to highlight the most frequently co-purchased combinations.

**Affinity of genre:**

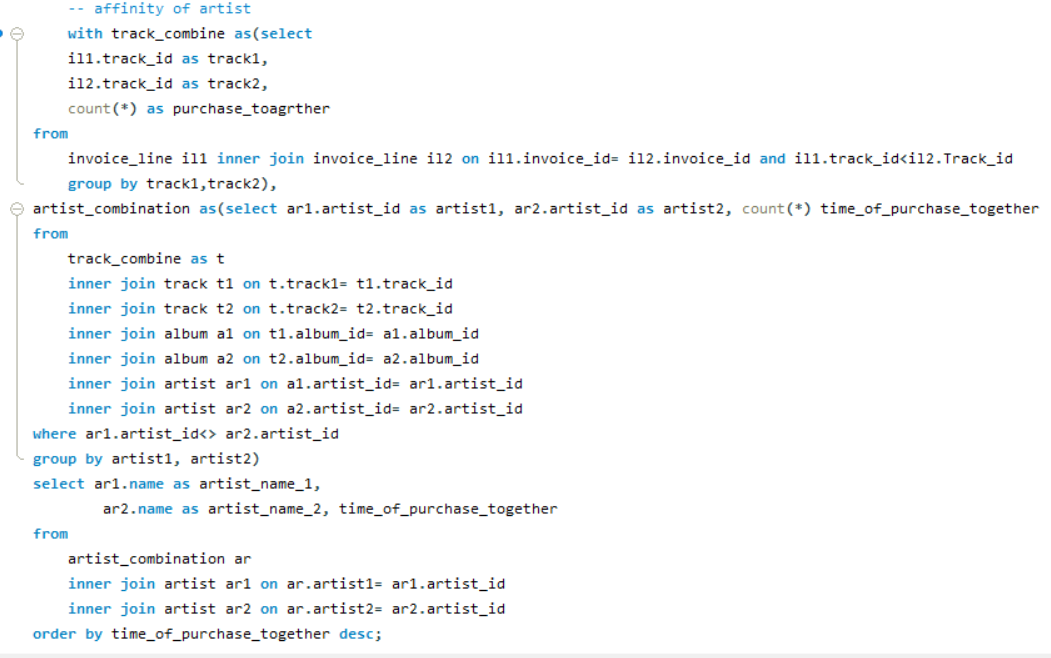
****

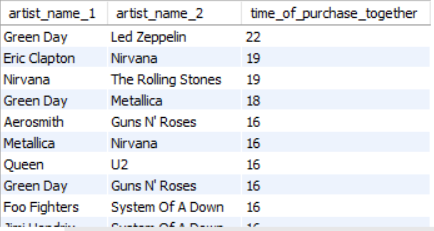
**Insights:**

* The query identifies pairs of genres frequently purchased together, offering insights into customer music preferences and potential genre combinations for marketing or recommendations.
* The time\_of\_purchase\_together count highlights the popularity of specific genre pairings, with higher counts indicating stronger affinities.
* The use of inner joins ensures only valid combinations from the same invoices are considered, providing accurate data on simultaneous purchases.
* Ordering by time\_of\_purchase\_together descending allows businesses to prioritize the most common genre pairings for targeted promotions or playlist suggestions.
* The inclusion of genre names enhances the usability of the results, making it easier to interpret and act on the affinity data.

**Result Summary:**

* The highest affinity is between "Metal" and "Rock" with 986 co-purchases.
* "Alternative & Punk" and "Rock" follow with 629 co-purchases.
* Other notable combinations include "Rock" with "Metal" (613), "Alternative" (302), and "Alternative & Punk" (412).
* Lesser affinities are observed with "Rock" and "R&B/Soul" (231), "Blues" and "Rock" (180), and "Rock" and "Latin" (176).
* The data highlights "Rock" as a central genre with strong connections to multiple other genres.

**Affinity of artist:**

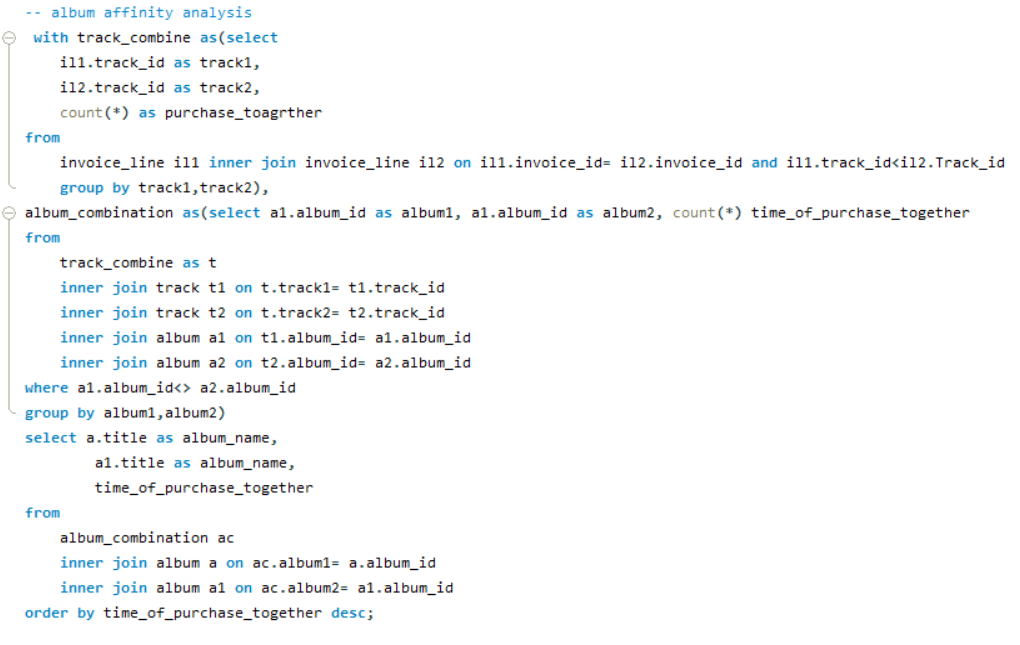
****

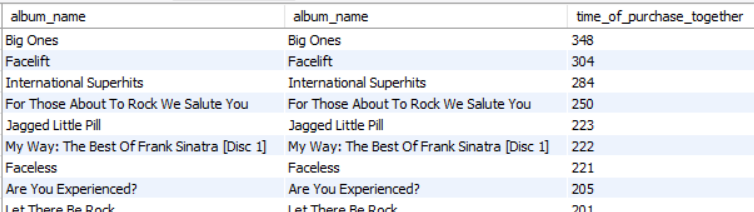
**Insights:**

* The query identifies artist combinations frequently purchased together, revealing customer preferences for specific artist pairings.
* Inner joins with track, album, and artist tables ensure accurate linking of tracks to their respective artists, providing reliable affinity data.
* The time\_of\_purchase\_together count measures the frequency of co-purchases, highlighting popular artist collaborations or preferences.
* Ordering by time\_of\_purchase\_together descending prioritizes the most frequently bought artist pairs, aiding in targeted marketing or recommendation strategies.
* The inclusion of artist names enhances the interpretability of the results for practical business applications.

**Result Summary:**

* The result summary highlights the affinity between various artists based on their co-purchase frequencies.
* It identifies popular artist pairings, with some combinations showing stronger connections than others.
* The data is ordered to emphasize the most frequently purchased pairs, reflecting customer preferences.
* The inclusion of artist names improves the usability of the results for further analysis or marketing purposes.

**Album affinity analysis:**

****

**Insights:**

* The query analyzes album affinities by identifying pairs of albums frequently purchased together, offering insights into customer music preferences.
* Inner joins with track and album tables ensure accurate linking of tracks to their respective albums, providing reliable co-purchase data.
* The time\_of\_purchase\_together count measures the frequency of album combinations, highlighting popular pairings.
* Ordering by time\_of\_purchase\_together descending prioritizes the most commonly bought album pairs for targeted marketing or recommendation strategies.
* The inclusion of album titles enhances the interpretability of the results, making them actionable for business purposes.

**Result Summary:**

* The result summary showcases the affinity between various albums based on their co-purchase frequencies, with "Big Ones" paired with itself at 348 times, indicating strong individual album popularity.
* Other notable pairings include "Facelift" with itself at 304 times, "International Superhits" with itself at 284 times, and "For Those About To Rock We Salute You" with itself at 250 times.
* Additional combinations such as "Jagged Little Pill" with itself at 223 times, "My Way: The Best Of Frank Sinatra [Disc 1]" with itself at 222 times, and "Faceless" with itself at 221 times show consistent co-purchase trends.
* Lesser affinities are observed with "Are You Experienced?" paired with itself at 205 times and "I Am The New Rock" with itself at 201 times.
* The inclusion of album names enhances the usability of the results for further analysis or marketing strategies.

**Recommendation:**

* **Artist Affinity**: Focus marketing campaigns on popular artist pairs like those with high co-purchase frequencies (e.g., Green Day and Led Zeppelin, Nirvana and The Rolling Stones). Create bundled offers or playlists featuring these artists to leverage their combined appeal.
* **Album Affinity:** Promote albums with strong self-pairing frequencies (e.g., "Big Ones" at 348, "Facelift" at 304) through targeted discounts or loyalty rewards. Suggest these albums as must-haves in customer recommendations to boost individual sales.
* **Genre Affinity:** Highlight genre combinations with high co-purchase counts (e.g., Rock with Metal, Alternative & Punk) in curated playlists or genre-specific promotions to enhance cross-genre engagement and increase overall sales.

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

**Ans:**

**Customer purchasing behaviour by region:**

**Objective:**

To analyse customer purchasing patterns by region (country, state, city), focusing on total customers, total purchases, average order value, total spending, and purchase frequency. This helps identify high-performing regions and understand customer behaviour.

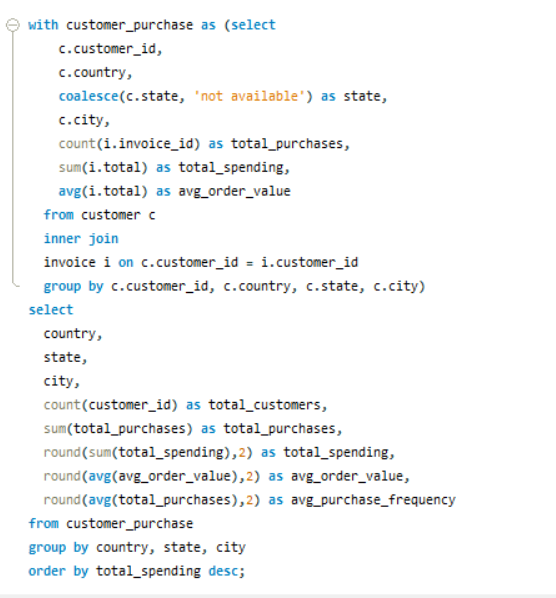
**Approach:**

* Created a CTE (customer\_purchase) to calculate each customer's total purchases, total spending, and average order value.
* Grouped results by country, state, and city to compute:

Total number of customers per region.

Overall and average purchase metrics (spending, frequency, order value).

* Used aggregation functions to summarize regional trends.



****

**Insights:**

* Prague (Czech Republic) has the highest total spending, suggesting highly active or high-value customers.
* Dublin (Ireland), despite having only 1 customer, shows high average spending and frequency—indicating strong engagement.
* Cities like São Paulo and Mountain View have consistent average order values but lower purchase frequency.
* Some states are listed as "not available", indicating incomplete or missing data in the customer address fields.

**Result Summary:**

The analysis highlights that cities like Prague and Mountain View have highly engaged customers with strong spending behavior. Other cities such as London and Berlin show steady performance. Some locations with fewer customers, like Dublin, still show high average engagement. Missing state data in many records limits deeper regional insights.

**Recommendation:**

* Prioritize Prague for retention and upselling campaigns—high-value customers exist there.
* Investigate Dublin’s single high-engagement customer—may indicate niche opportunity.
* Clean up address data (e.g., missing states) for better segmentation.
* Use these insights to guide regional marketing strategies and personalized outreach based on average spending and frequency.
* Consider incentivizing cities with medium engagement (like Berlin and Paris) to increase frequency or order value.

**Churn Rate by Region:**

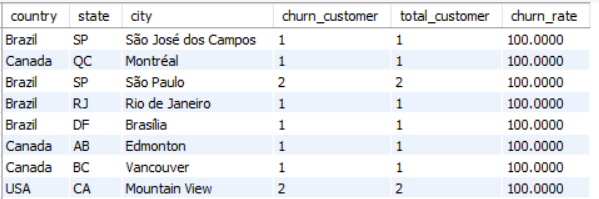
**Objective:**

To identify the churn rate of customers across different regions by analysing their most recent purchase dates and determining where customer retention is weakest.

**Approach:**

**Used two CTEs:**

* The first CTE (region\_churn\_rate) calculates the most recent purchase date for each customer by region.
* The second CTE (churn\_customer) filters out customers who haven't purchased anything in over a year.
* ****Finally, we joined the churn data with customer data to compute the churn rate as a percentage of total customers per region.

****

**Insights:**

* Every city analysed in the dataset has a churn rate of 100%.
* This indicates that in all these regions, customers have not made any purchases in the last year.
* The churn is consistent across countries like Brazil, Canada, and the USA, which may point to broader customer engagement or product usage issue.

**Result Summary:**

The churn analysis reveals that customers from several cities across different countries have become inactive for over a year, resulting in a 100% churn rate in those locations.

**Recommendation:**

* Investigate the reasons behind the complete churn in these cities—whether due to lack of marketing, poor service experience, or regional product fit.
* Re-engagement campaigns and customer feedback collection should be prioritized in these regions.
* Evaluate onboarding and retention strategies to prevent similar trends in other areas.

**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?**

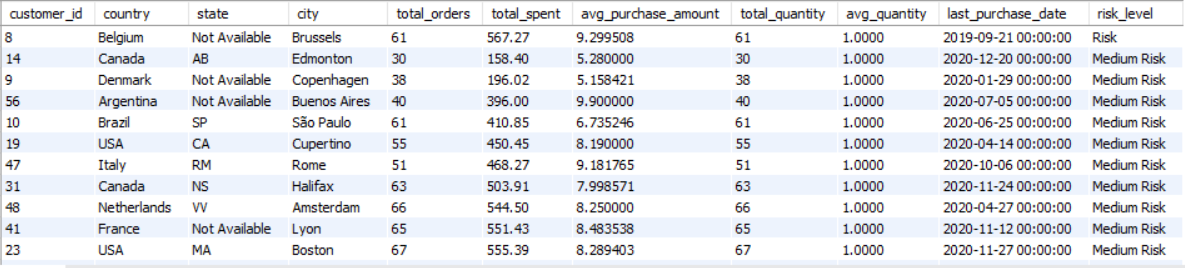
**Ans:**

**Objective:**

To identify and categorize customers based on their purchasing behaviour to assess their risk level of churn or reduced engagement. This allows for proactive strategies to retain valuable customers.

**Approach:**

* Used two CTEs: one to capture the latest invoice date, and another to summarize customer purchases.
* Calculated key metrics such as total orders, total spent, average purchase amount, and average quantity.
* Classified customers into three risk levels (Risk, Medium Risk, Low Risk) based on:
  + - * + How long it has been since their last purchase and their total purchase.

****

**Insights**

* Customers label as Risk have not made a purchase in over a year, showing signs of potential churn.
* The Medium Risk group includes customers with moderate spending but recent activity. They may need engagement to maintain loyalty.
* Customers in the Low Risk category consistently engage and spend more, indicating high loyalty and satisfaction.

**Result Summary:**

The analysis effectively segments customers based on recent behavior and purchase value. Those marked as high risk had no recent purchases and relatively lower spending patterns. Meanwhile, medium-risk customers have shown some engagement but with limited purchase activity. The majority of consistent spenders fall into the low-risk category.

**Recommendations:**

* **For Risk customers:** Initiate re-engagement campaigns (e.g., discount offers, personalized emails).
* **For Medium Risk customers:** Offer loyalty programs or incentives to increase purchase frequency.
* **For Low Risk customers:** Maintain their satisfaction with exclusive benefits or early access to new products.

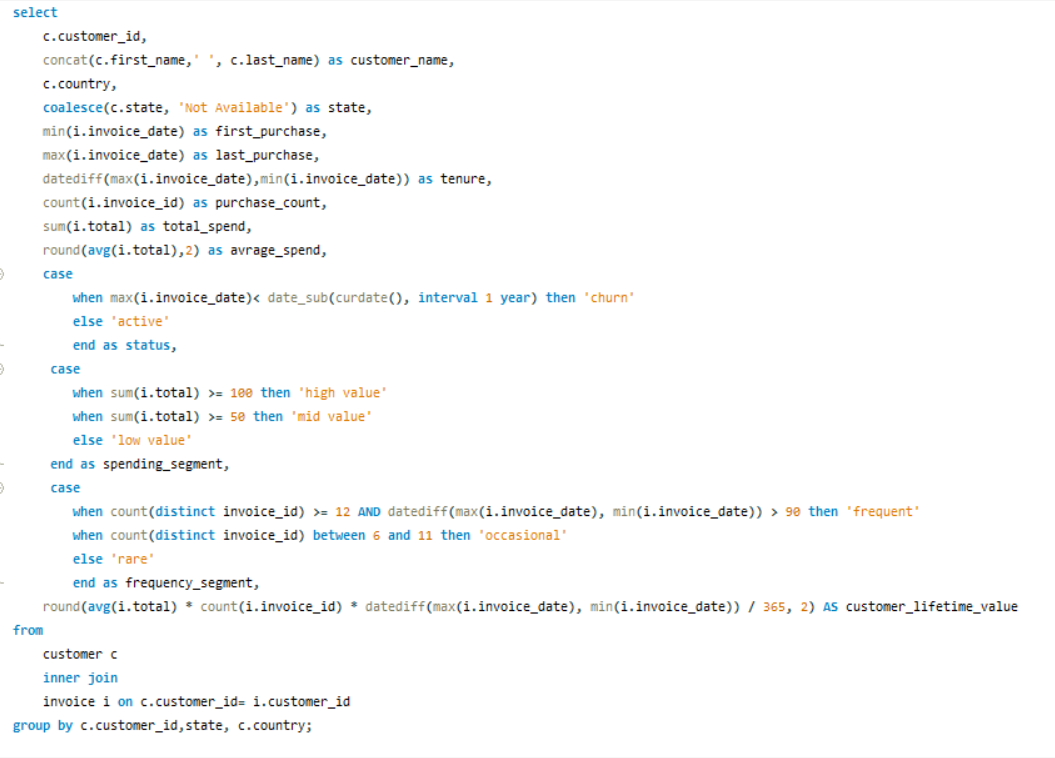
**7.Customer Lifetime Value Modeling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns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

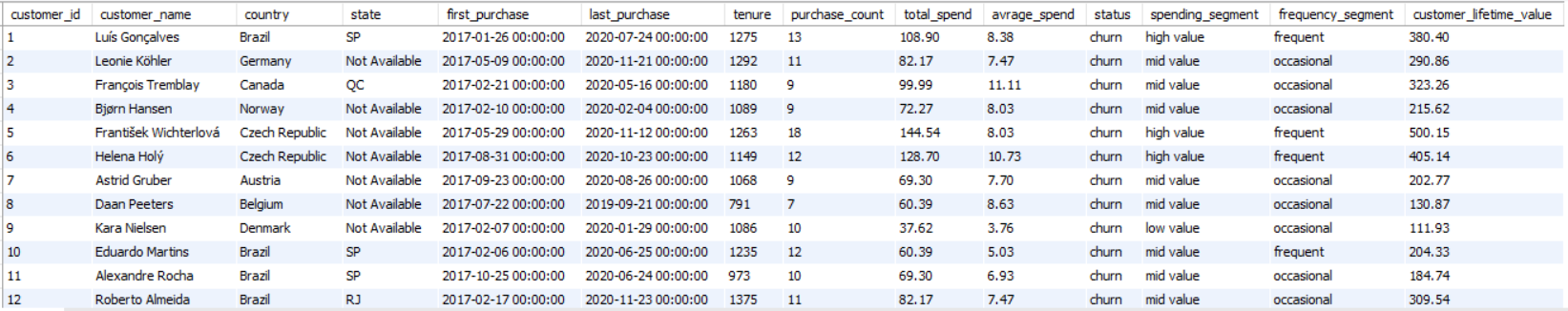
**Ans:**

**Objective:**

The objective is to analyse customer purchase data to model Customer Lifetime Value by leveraging tenure, purchase history, and engagement metrics. This will enable the identification of customer segments, prediction of future value, and development of targeted marketing and loyalty program strategies to enhance retention and maximize revenue.

**Approach:**

1. **Data Collection and Aggregation:** 
   * Extract customer details (customer\_id, customer\_name, country, state, city) and invoice data (dates, totals) using an inner join between the customer and invoice tables.
   * Compute key metrics: first\_purchase min(invoice date), last\_purchase max (invoice date), tenure (datediff of last and first purchase), purchase\_count (count of invoices), total\_spend (sum of invoice totals), and average\_spend (rounded avg of totals).
2. **Segmentation:** 
   * Categorize status as 'churn' (if last purchase > 1 year ago) or 'active' using a case statement.
   * Segment spending\_segment into 'high value' (>100), 'mid value' (50-100), or 'low value' (<50) based on total spend.
   * Classify frequency\_segment as 'frequent' (>12 purchases and tenure > 90 days), 'occasional' (6-11 purchases), or 'rare' (<6 purchases) using case statements.
3. **CLV Calculation:** 
   * Calculate customer\_lifetime\_value as the product of average\_spend, purchase\_count, and tenure (in days) divided by 365, rounded to 2 decimal places.
   * Use this as a baseline for predictive modeling with advanced techniques (e.g., regression) to forecast future value.
4. **Analysis and Strategy Development:** 
   * Identify patterns in churned customers (e.g., tenure, spending, frequency) to inform re-engagement tactics.
   * Develop targeted marketing and loyalty programs based on segments (e.g., high value frequent customers receive exclusive offers, mid/low value occasional customers get re-engagement campaigns).
   * ****Continuously update the model with new data and refine segments based on emerging trends.

****

**Insights:**

The analysis reveals a consistent pattern of inactivity among all customers, indicating a widespread churn issue that requires urgent attention.

Customer engagement varies across different spending and purchase frequency levels, suggesting diverse reasons for disengagement that need segmented exploration.

The calculated Customer Lifetime Value (CLV) highlights varying contributions, providing a basis for prioritizing retention efforts on higher-value customers.

The lack of recent purchases across multiple regions, with some missing location details, points to potential data gaps or uniform market challenges affecting retention.

Tenure and last purchase trends suggest that even long-term customers may lose interest over time, emphasizing the need for proactive re-engagement strategies.

**Result:**

* The analysis provides an overview of customer profiles across various regions, with some location details missing.
* Purchase metrics reveal a range of engagement durations, transaction frequencies, and spending levels among customers.
* Customers are segmented based on their spending habits and purchase frequency, identifying distinct behavioural groups.
* Customer Lifetime Value (CLV) is calculated to assess the overall value contributed by each customer over their engagement period.
* All customers are currently inactive, with no recent purchases, suggesting a uniform churn pattern that may indicate broader market or retention issues**.**

**Informing Targeted Marketing and Loyalty Program Strategies:**

The analysis of customer data, including tenure, purchase history, and engagement metrics, can inform targeted marketing and loyalty program strategies by:

* **Segment-Specific Campaigns**: Tailoring offers based on spending and frequency segments (e.g., high-value frequent customers receive premium rewards, while mid-value occasional customers get re-engagement incentives).
* **Retention Focus**: Using tenure and last purchase dates to identify at-risk customers for proactive outreach, such as personalized discounts or reminders.
* **Loyalty Enhancements**: Designing loyalty programs that reward consistent spending and frequent purchases, encouraging continued engagement from valuable segments.
* **Customer lifetime value Optimization**: Leveraging customer lifetime value insights to prioritize high-value customers for exclusive benefits, maximizing long-term revenue.

**Common Characteristics or Purchase Patterns Among Customers Who Have Stopped Purchasing:**

Observations of customers who have stopped purchasing (all classified as 'churn') include.

* **Inactivity Period:** A consistent lack of recent purchases, suggesting a significant time gap since their last transaction, which may indicate disengagement or external factors.
* **Varied Engagement Levels:** Customers exhibit a range of purchase frequencies and tenures, with no single pattern dominating, implying churn affects both frequent and occasional buyers.
* **Diverse Spending:** Churn occurs across different spending levels, indicating that high, mid, and low spenders are equally susceptible, pointing to factors beyond monetary contribution.
* **Geographical Spread:** Inactivity is observed across multiple regions, with some location details missing, suggesting churn may not be region-specific but could relate to data gaps or uniform market conditions.
* **Potential Decline**: Longer engagement periods with recent inactivity might suggest a gradual loss of interest or satisfaction, warranting further investigation into retention triggers.

**Recommendation:**

* **Targeted Marketing:** Focus on re-engaging inactive customers with personalized campaigns, such as offering discounts on genres or artists they previously enjoyed, to recapture their interest.
* **Loyalty Programs:** Introduce tiered rewards for frequent spenders, encouraging continued purchases with exclusive content or early access to new releases.
* **Retention Strategies:** Identify at-risk customers early by monitoring purchase gaps and tenure, then offer tailored incentives like bundle deals to prevent churn.
* **Segmented Offers:** Prioritize high-value segments with premium loyalty benefits, while using low-cost promotions to boost engagement among lower-value customers.

**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?**

**Ans:**

**Approach:**

To understand how promotional campaigns (such as discounts, events, or email outreach) influence customer growth, loyalty, and sales performance, a structured analysis can be followed:

**Customer Acquisition**

* Monitor how many new customers joined during promotional activities.
* Compare these numbers with periods without any promotions to measure how well the campaign attracted fresh customers.

**Retention & Repeat Purchases**

* Observe if customers continued buying after the campaign ended.
* Evaluate this against periods without promotions to see if the campaign influenced long-term customer engagement.

**Sales Uplift**

* Measure the increase in revenue by comparing sales during the campaign with usual sales from non-promotional times.
* Consider seasonal trends or external events that might have affected the results.

**Segmented Response**

* Analyse how various customer groups (such as based on location, age, or spending habits) reacted to the promotion.
* Find out which segments responded better so that future campaigns can be more focused and effective.

Each of these evaluations provides a thorough perspective on how well promotional activities shape customer actions and financial results.

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

**Ans:**

* **Data Cleanup:** Start by checking and fixing the dataset, handling missing entries and removing duplicates for better accuracy.
* **Key Metrics Check:** Look at important numbers like how often customers buy, churn rates, and sales results to get a clear picture.
* **Asking Questions:** Come up with simple questions from the data, like which albums or tracks are most popular, top-selling genres, or buying differences by country.

1.Which albums or tracks are the most popular among customers?

2.What genres are leading in sales figures across the dataset?

3.How do buying patterns differ based on customers' countries?

* **Customer Groups:** Split customers into categories based on age, past purchases, and activity to understand their habits better.
* **Churn Look:** Find out why customers leave, focusing on at-risk groups and possible reasons for their exit.
* **Country Differences:** Compare what customers like, how they spend, and which music is popular in different countries.
* **Sales Trends:** Check which albums or tracks sell the most, looking at prices, promotions, and who’s buying them.

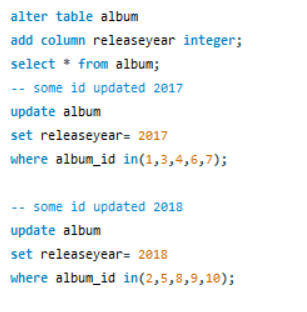
**10. How can you alter the "Albums" table to add a new column named "ReleaseYear" of type INTEGER to store the release year of each album?**

**Ans:**

**Approach:**

* **Table Modification**: Use the **alter table** statement to add a new ReleaseYear column with an **integer** data type to the album table, enabling the storage of release year information.
* **Data Update**: Employ the **update** function to modify the ReleaseYear column, setting specific year values for selected records based on album\_id conditions.

****



**Result Summary:**

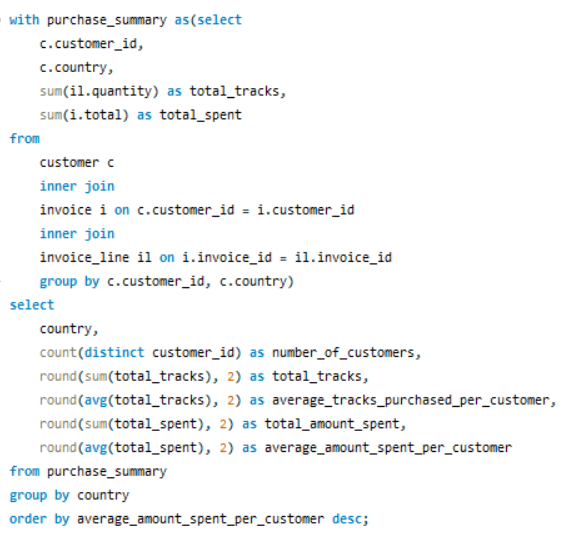
* The **alter table** statement successfully added a new column named **releaseyear** with an integer data type to the album table, enabling the storage of release year information for each album.
* The initial **select \* from album** query confirmed the addition of the ReleaseYear column, though it initially contained no values.
* The **update** statement for albums with album\_id in (1, 3, 4, 6, 7) set their **releaseyear** to 2017, effectively assigning a release year to this group.
* The subsequent **update** statement for albums with album\_id in (2, 5, 8, 9, 10) set their **releaseyear** to 2018, completing the year assignment for the remaining albums.
* The process ensured that all specified albums now have their **releaseyear** values updated based on their respective album\_id groups.

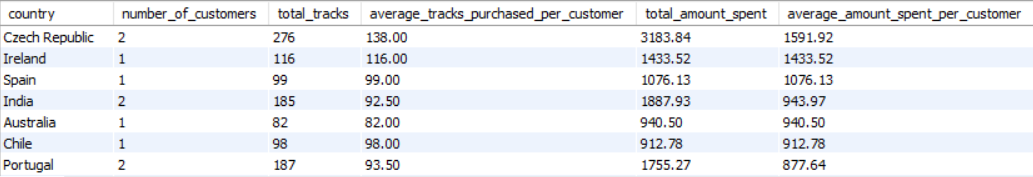
**11. Chinook is interested in understanding the purchasing behavior of customers based on their geographical location. They want to know the average total amount spent by customers from each country, along with the number of customers and the average number of tracks purchased per customer. Write an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

**Ans:**

**Approach:**

* **Data Preparation:** Start by joining customer, invoice, and invoice\_line tables to gather purchase data, including customer\_id, country, quantity, and unit\_price, ensuring a solid foundation for analysis.
* **CTE and Aggregation:** Use a single CTE (purchase\_summary) to calculate total\_tracks and total\_spent per customer, then aggregate by country to derive metrics like total tracks, average tracks per customer, and average spend.
* **Result Organization:** Sort the output by average\_amount\_spent\_per\_customer in descending order to prioritize high-spending regions, with results validated for accuracy.





**Insights:**

* The analysis highlights differences in customer engagement and spending across various regions, pointing to opportunities for tailored marketing approaches.
* The single CTE approach streamlines purchase data aggregation, offering valuable insights into average customer behaviour that can inform resource and strategy decisions.
* Sorting by average spend per customer identifies leading regions, enabling a focused strategy to enhance revenue potential.

**Result Summary:**

The result summary showcases the aggregated purchasing behavior across different regions, reflecting variations in customer engagement and spending patterns. It highlights the number of customers and total tracks purchased per region, alongside average metrics for tracks and amount spent per customer. The results are organized to emphasize regions with the highest average spending, providing a clear basis for strategic decision-making.