**Objective Questions**

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

**Answer:** **The data provided possess ‘No’ duplicate values although there are NULL values in below tables.**

**Observations:**

1. **No Duplicates Across Tables:  
   All tables (album, artist, customer, etc.) do not have duplicate rows based on checks with SELECT DISTINCT.**
2. **Missing Values Identified:**
   * **customer Table:**
     + **fax: 47 null values**
     + **state: 29 null values**
     + **company: 49 null values**
   * **employee Table:**
     + **reports\_to: 1 null value for employee\_id = 1**
   * **track Table:**
     + **composer: 978 null values**

**Query:**

**show tables;**

**select \* from album;**

**select distinct \* from album;**

**SELECT \* FROM artist;**

**SELECT distinct \* FROM artist;**

**SELECT \* from customer;**

**SELECT distinct \* FROM customer;**

**SELECT COUNT(\*) FROM customer;**

**SELECT \* from employee;**

**SELECT distinct \* FROM employee;**

**SELECT \* FROM invoice;**

**SELECT distinct \* FROM invoice;**

**SELECT \* FROM genre;**

**SELECT distinct \* FROM genre;**

**SELECT \* FROM invoice line;**

**SELECT distinct \* FROM invoice line;**

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

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

**SELECT \* FROM playlist;**

**SELECT distinct \* FROM playlist;**

**SELECT \* FROM playlist track;**

**SELECT distinct \* FROM playlist track;**

**SELECT \* FROM track;**

**SELECT distinct \* FROM track;**

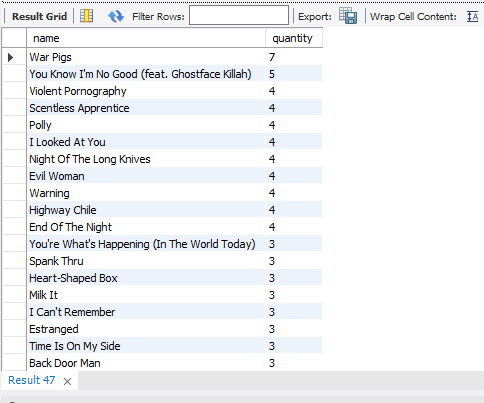
**SELECT COUNT(\*) FROM track;**

**SELECT COUNT(\*) FROM track**

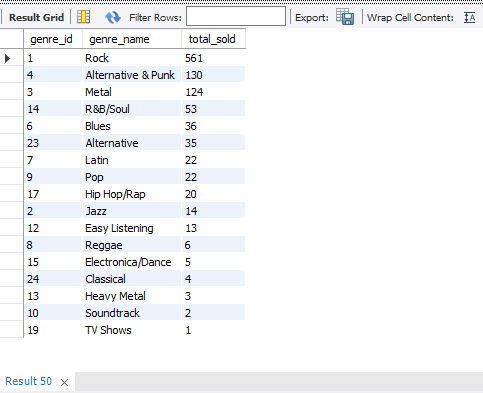
**WHERE composer is NULL;**

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

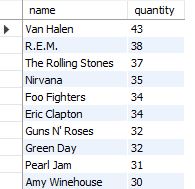
**Answer: Below are the top 10 selling Tracks of the USA, War Pigs by Cake being on top.**



**Below are the top most preferred genre in USA by number of records sold.**



**Below are the top 10 artists in USA:**



**Insights:**

* **Rock leads in US sales.**
* **"War Pigs" by Cake is the best-selling track.**

**Recommendations:**

* **Focus on Rock for inventory management and marketing efforts.**
* **Implement genre-specific promotional strategies.**

**Query:**

**-- Top\_Tracks**

**SELECT t.name, SUM(il.quantity) AS quantity**

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

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

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

**GROUP BY t.name**

**ORDER BY quantity DESC;**

**-- Top\_Genre**

**SELECT**

**g.genre\_id,**

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

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

**FROM**

**invoice\_line il**

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

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

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

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

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

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

**WHERE**

**c.country = 'USA'**

**GROUP BY**

**g.genre\_id, g.name**

**ORDER BY**

**total\_sold DESC;**

**-- Top\_Artist**

**SELECT art.name, SUM(il.quantity) AS quantity**

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

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

**INNER JOIN album a on t.album\_id = a.album\_id INNER JOIN artist art ON a.artist\_id = art.artist\_id**

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

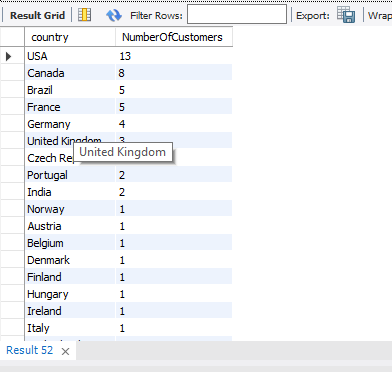
**GROUP BY art.name**

**ORDER BY quantity DESC;**

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

**Answer: Demographic breakdown:**

**On the basis of Country:**



**On the basis of City:**



**Insights:**

* **Customers are distributed across 24 countries.**
* **The United States has the highest number of customers.**

**Recommendations:**

* **Prioritize the US market due to its significant customer base.**
* **Identify growth opportunities in key countries like Canada, Brazil, France, and Germany.**
* **Develop region-specific marketing strategies to better target diverse markets.**

**Query:**

**-- By\_Country**

**SELECT country, COUNT(\*) AS NumberOfCustomers**

**FROM customer**

**GROUP BY country**

**ORDER BY NumberOfCustomers DESC;**

**-- By\_city**

**SELECT city as city, COUNT(customer\_id) AS NumberOfCustomers**

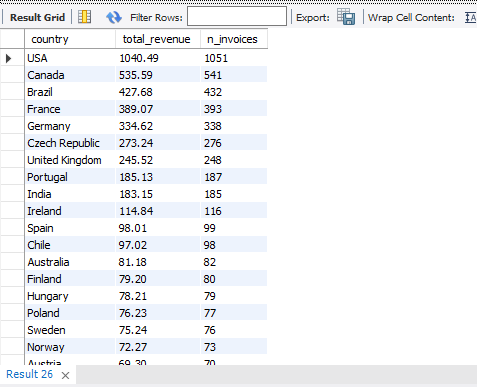
**FROM customer**

**GROUP BY city**

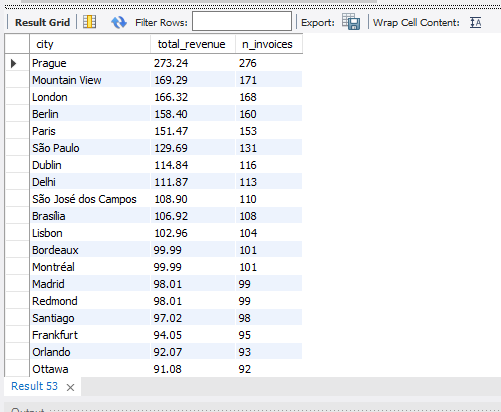
**ORDER BY NumberOfCustomers DESC;**

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

**Answer: Revenue By Country:**



**Revenue By City:**



**Insights:**

* **Prague (Czech Republic) leads in revenue generation.**
* **São Paulo (Brazil) records a high invoice count but comparatively lower revenue.**

**Recommendations:**

* **Study Prague's Model: Explore the factors behind Prague's high revenue per transaction and adapt successful practices for other cities.**
* **Evaluate Location Performance: Review revenue per invoice across all regions to uncover improvement opportunities and optimize pricing policies.**

**Query:**

**-- Revenue vs Country**

**SELECT c.country,**

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

**COUNT(il.invoice\_id) AS n\_invoices**

**FROM invoice I JOIN invoice\_line il ON il.invoice\_id = i.invoice\_id JOIN customer c ON c.customer\_id = i.customer\_id**

**GROUP BY c.country**

**ORDER BY total\_revenue DESC, n\_invoices DESC;**

**-- Revenue vs City**

**SELECT**

**COALESCE(c.city, 'N/A') AS city,SUM(il.unit\_price \* il.quantity) AS total\_revenue, COUNT(il.invoice\_id) AS n\_invoices**

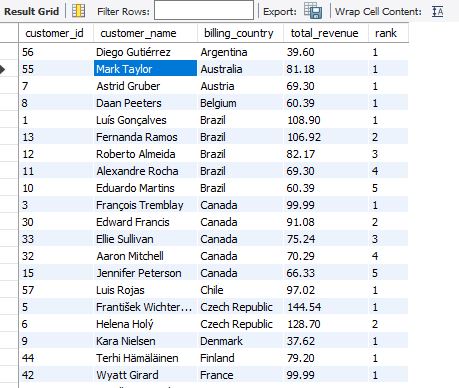
**FROM invoice I JOIN invoice\_line il ON il.invoice\_id = i.invoice\_id JOIN customer c ON c.customer\_id = i.customer\_id**

**GROUP BY city**

**ORDER BY total\_revenue DESC, n\_invoices DESC;**

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

**Answer:**



**Insights:**

1. **Top Performers by Country:**
   * **Czech Republic has the highest-performing customers, with František Wichterlová and Helena Holý generating the highest revenue (144.54 and 128.7 respectively).**
   * **Brazil also shows strong revenue, with multiple top customers including Luís Gonçalves (108.9), Fernanda Ramos (106.92), and others.**
   * **USA shows a variety of top-performing customers, with Jack Smith (98.01), Dan Miller (95.04), and Heather Leacock (92.07) among the top 5 in their country.**
2. **Geographical Trends:**
   * **North America (USA, Canada) appears to have a large number of high-performing customers, with Canada’s François Tremblay (99.99) and Edward Francis (91.08) in the top positions.**
   * **Countries like India, Germany, and France also have notable performers, showing a diverse spread of revenue across the globe.**
3. **Revenue Distribution:**
   * **Argentina, Denmark, Italy, and Norway have lower revenue generation compared to other countries in the list.**
   * **Some countries like Brazil, Czech Republic, and USA exhibit a more concentrated share of high revenue, highlighting potential key markets.**

**Recommendations:**

1. **Focus on High Revenue Countries:**
   * **Target countries like Czech Republic, Brazil, and USA for future expansion and high-value campaigns. The performance in these regions is impressive, and strategies could be tailored to further capitalize on this strength.**
2. **Customer Segmentation:**
   * **Evaluate customers from Argentina, Denmark, and Italy, as their lower revenue suggests potential for growth. Investigate the reasons behind their performance and create localized marketing strategies to increase engagement and sales.**
3. **Leverage Top Performers for Marketing:**
   * **Use high performers in countries like Czech Republic, Germany, and USA as case studies or ambassadors to help expand reach and build trust in similar markets.**

**Query:**

**-- Top 5 customers by total revenue in each country**

**WITH CustomerRevenueRank AS (**

**SELECT**

**c.customer\_id,**

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

**i.billing\_country,**

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

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

**FROM**

**customer c**

**INNER JOIN**

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

**GROUP BY**

**c.customer\_id, i.billing\_country, c.first\_name, c.last\_name**

**)**

**SELECT**

**customer\_id,**

**customer\_name,**

**billing\_country,**

**total\_revenue,**

**`rank`**

**FROM**

**CustomerRevenueRank**

**WHERE**

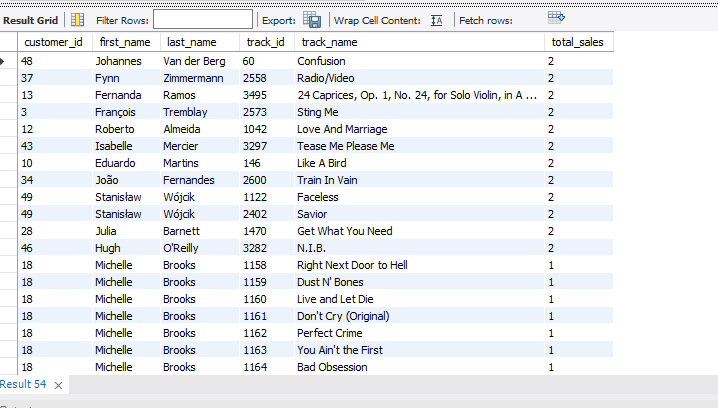
**`rank` <= 5**

**ORDER BY**

**billing\_country ASC, `rank` ASC;**

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

**Answer:** **Below are the top selling tracks for customers:**



**Query:**

**-- 6 Top-selling track for each customer**

**WITH ranked\_tracks AS (**

**SELECT**

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

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

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

**FROMcustomer c**

**JOIN**

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

**JOIN**

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

**JOIN**

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

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

**c.first\_name,c.last\_name,t.track\_id,t.name**

**)**

**SELECT**

**customer\_id,first\_name,last\_name,track\_id,track\_name,total\_sales**

**FROM**

**ranked\_tracks**

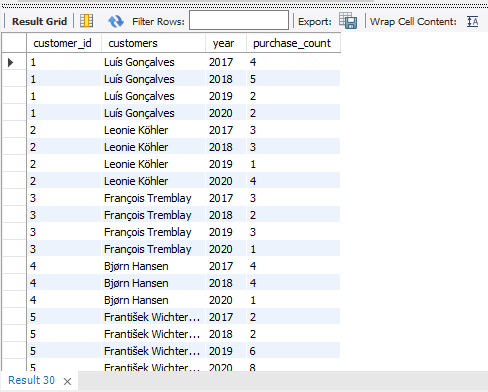
**ORDER BY**

**total\_sales DESC;**

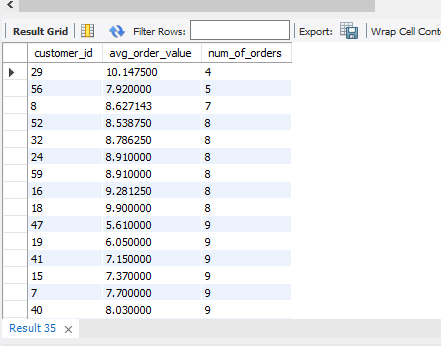
**7.** **Are there any patterns or trends in customer purchasing behaviour (e.g., frequency of purchases, preferred payment methods, average order value)?**

**Answer:**

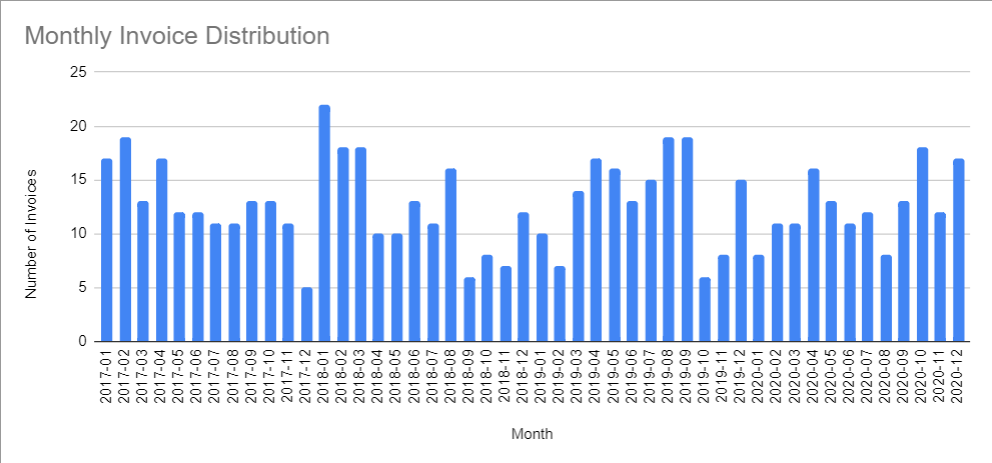
**Trend By Customer Purchase Count:**



**Average Order Value by Customer:**



**Monthly Invoice Distribution from 2017-2020:**



|  |  |
| --- | --- |
| **Row Labels** | **Sum of purchase\_count** |
| 2017 | 154 |
| 2018 | 151 |
| 2019 | 159 |
| 2020 | 150 |
| **Grand Total** | **614** |

**Insight:**

**The total number of purchases over these four years is 614. It is clear that the total purchases are relatively stable, with no significant fluctuations year over year. The year 2019 had the highest purchase count, while 2020 saw a slight decline, although it was still similar to the previous years.**

**Recommendations:**

1. **Identify Seasonal Patterns: If this data is related to a specific product or service, it's worth investigating if there are any seasonal trends or events affecting purchases in these years. For example, was there a particular campaign or event in 2019 that drove higher purchases?**
2. **Increase Engagement in 2020: Given that purchases in 2020 are slightly lower than the previous years, it may be useful to investigate if there were external factors (e.g., economic downturn, competition, or product availability) that contributed to this. Consider strategies such as targeted promotions, advertising, or improving customer loyalty programs to boost purchases.**

**Query:**

**-- 7 Trends in customer purchasing behaviour.**

**SELECT**

**c.customer\_id,**

**CONCAT(c.first\_name, ' ', c.last\_name) as customers,**

**YEAR(i.invoice\_date) AS year,**

**COUNT(i.invoice\_id) AS purchase\_count**

**FROM**

**customer c**

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

**GROUP BY**

**c.customer\_id, customers, YEAR(i.invoice\_date)**

**ORDER BY**

**c.customer\_id, customers, YEAR(i.invoice\_date);**

**-- Avg Order Value By Cust\_ID**

**select customer\_id, avg(total) as avg\_order\_value, count(invoice\_id)as num\_of\_orders**

**from invoice**

**group by customer\_id**

**order by count(invoice\_id),avg(total);**

**-- Invoice Count Monthly**

**select count(invoice\_id) as daily\_invoice\_count, extract(YEAR\_MONTH from invoice\_date), avg(total) as monthly\_avg\_total, sum(total) as monthly\_sum\_total**

**from invoice**

**group by extract(YEAR\_MONTH from invoice\_date)**

**order by extract(YEAR\_MONTH from invoice\_date);**

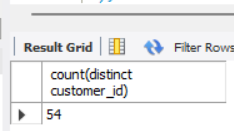
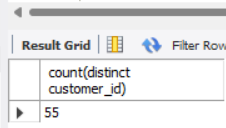
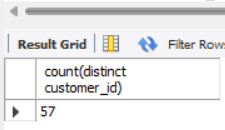
**8.What is the customer churn rate?**

**Answer:**

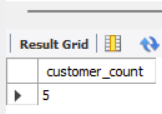
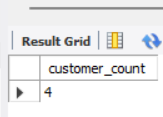
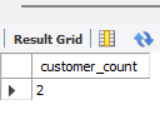
**Summary of Churn and Churn Rate Calculation Steps**

1. **Identify Churned Customers:**
   * **Churned customers are those who made purchases in a given year but did not make purchases in the previous year.**
   * **For example, to identify churned customers in 2018, count customers who made purchases in 2018 but not in 2017.**
2. **Identify Starting Customer Base for Each Year:**
   * **The starting customer base is the number of customers who made purchases in the previous year.**
   * **For example, the starting customer base for 2018 is the count of customers who made purchases in 2017.**
3. **Calculate Churn Rate for Q1:**
   * **The Q1 churn rate measures customers who were active in the first quarter of the year (Jan-Mar) but did not purchase in the last quarter of the previous year (Oct-Dec).**
4. **Calculate Overall Churn Rate for the Year:**
   * **The overall churn rate compares the number of customers in the first three months of the year (Jan-Mar) to those in the last three months (Oct-Dec).**
5. **Repeat for Subsequent Years:**
   * **Perform the same calculations for subsequent years (e.g., 2019 and 2020), comparing the first and last three months to measure churn for each year.**

**Customer Count for Different Years:**

* 

**Churned Customers Count:**

****

**Total Churned Customers = Churned in Year 2018 + Churned in Year 2019 + Churned in Year 2020 = 2 + 4 + 5 = 11**

**(Customers at Start of Year 2018 + Customers at Start of Year 2019 + Customers at Start of Year 2020)/3**

**Churn Rate = 11 \*100/55.33 = 19.88%**

**Insights:**

* **Nearly 20% of customers leave annually, indicating a significant churn rate.**
* **While the overall customer numbers remain stable, customer retention appears to be declining.**

**Recommendations:**

* **Focus on increasing new customer engagement to strengthen long-term retention.**
* **Implement loyalty programs and personalized offers to improve customer retention.**
* **Collect feedback from customers to identify the reasons behind their departure and address them effectively.**

**Query:**

**-- 8 Churn rate calculation**

**-- 1. Count customers who churned in 2018 (customers who made purchases in 2018 but not in 2017)**

**SELECT**

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

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2018-01-01' AND '2018-12-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2017-01-01' AND '2017-12-31'**

**);**

**-- 2. Count customers who churned in 2019 (customers who made purchases in 2019 but not in 2018)**

**SELECT**

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

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2019-01-01' AND '2019-12-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-01-01' AND '2018-12-31'**

**);**

**-- 3. Count customers who churned in 2020 (customers who made purchases in 2020 but not in 2019)**

**SELECT**

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

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2020-01-01' AND '2020-12-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-01-01' AND '2019-12-31'**

**);**

**-- Count total number of customers at the beginning of each year (starting customer base)**

**-- Customers at the beginning of 2018 (purchased in 2017)**

**SELECT**

**COUNT(DISTINCT customer\_id)**

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2017-01-01' AND '2017-12-31';**

**-- Customers at the beginning of 2019 (purchased in 2018)**

**SELECT**

**COUNT(DISTINCT customer\_id)**

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2018-01-01' AND '2018-12-31';**

**-- Customers at the beginning of 2020 (purchased in 2019)**

**SELECT**

**COUNT(DISTINCT customer\_id)**

**FROM**

**invoice**

**WHERE**

**invoice\_date BETWEEN '2019-01-01' AND '2019-12-31';**

**-- Calculate churn rate for the first quarter (Q1) of each year**

**-- Comparing customers who were active in the first 3 months but did not purchase in the last quarter of the previous year**

**WITH Q1\_cte1 AS (**

**SELECT COUNT(DISTINCT customer\_id) AS churn\_count**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-01-01' AND '2018-03-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2017-10-01' AND '2017-12-31'**

**)**

**),**

**Q1\_cte2 AS (**

**SELECT COUNT(DISTINCT customer\_id) AS churn\_count**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-01-01' AND '2019-03-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-10-01' AND '2018-12-31'**

**)**

**),**

**Q1\_cte3 AS (**

**SELECT COUNT(DISTINCT customer\_id) AS churn\_count**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2020-01-01' AND '2020-03-31'**

**AND customer\_id NOT IN (**

**SELECT DISTINCT customer\_id**

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-10-01' AND '2019-12-31'**

**)**

**),**

**Q1\_cte4 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2017-01-01' AND '2017-03-31'**

**),**

**Q1\_cte5 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-01-01' AND '2018-03-31'**

**),**

**Q1\_cte6 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-01-01' AND '2019-03-31'**

**),**

**Q1\_cte7 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2020-01-01' AND '2020-03-31'**

**)**

**SELECT**

**(c1.churn\_count + c2.churn\_count + c3.churn\_count) \* 100 / ((c4.customer\_count + c5.customer\_count + c6.customer\_count + c7.customer\_count) / 4) AS Q1\_Churn**

**FROM**

**Q1\_cte1 AS c1,**

**Q1\_cte2 AS c2,**

**Q1\_cte3 AS c3,**

**Q1\_cte4 AS c4,**

**Q1\_cte5 AS c5,**

**Q1\_cte6 AS c6,**

**Q1\_cte7 AS c7;**

**-- Calculate churn rates based on the first and last three months of each year**

**-- Churn rate in 2017 (January-March vs October-December)**

**WITH first\_three\_months AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2017-01-01' AND '2017-03-31'**

**),**

**last\_three\_months AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2017-10-01' AND '2017-12-31'**

**)**

**SELECT**

**((first3.customer\_count - last3.customer\_count) / first3.customer\_count) \* 100 AS churn\_rate**

**FROM**

**first\_three\_months AS first3,**

**last\_three\_months AS last3;**

**-- Churn rate in 2017: 40.82%**

**-- Repeat for the other years (2018, 2019, 2020) to calculate churn rates.**

**-- Churn rate in 2018 (January-March vs October-December)**

**WITH first\_three\_months AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-01-01' AND '2018-03-31'**

**),**

**last\_three\_months AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2018-10-01' AND '2018-12-31'**

**)**

**SELECT**

**((first3.customer\_count - last3.customer\_count) / first3.customer\_count) \* 100 AS churn\_rate**

**FROM**

**first\_three\_months AS first3,**

**last\_three\_months AS last3;**

**-- Churn rate in 2018: 53.45%**

**-- Churn rate in 2019 (January-March vs October-December)**

**WITH first\_three\_months\_2019 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-01-01' AND '2019-03-31'**

**),**

**last\_three\_months\_2019 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2019-10-01' AND '2019-12-31'**

**)**

**SELECT**

**((first3\_2019.customer\_count - last3\_2019.customer\_count) / first3\_2019.customer\_count) \* 100 AS churn\_rate**

**FROM**

**first\_three\_months\_2019 AS first3\_2019,**

**last\_three\_months\_2019 AS last3\_2019;**

**-- Churn rate in 2019: Calculate based on customer activity in Q1 and Q4**

**-- Churn rate in 2020 (January-March vs October-December)**

**WITH first\_three\_months\_2020 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2020-01-01' AND '2020-03-31'**

**),**

**last\_three\_months\_2020 AS (**

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

**FROM invoice**

**WHERE invoice\_date BETWEEN '2020-10-01' AND '2020-12-31'**

**)**

**SELECT**

**((first3\_2020.customer\_count - last3\_2020.customer\_count) / first3\_2020.customer\_count) \* 100 AS churn\_rate**

**FROM**

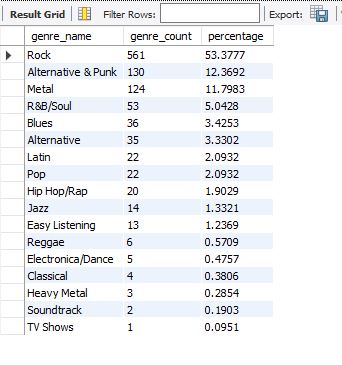
**first\_three\_months\_2020 AS first3\_2020,**

**last\_three\_months\_2020 AS last3\_2020;**

**-- Churn rate in 2020: Calculate based on customer activity in Q1 and Q4**

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

**Answer:**



**Insights:**

* **Dominant Genre: The genre Rock leads with 561 counts, making up 53.38% of the total, dominating the preferences.**
* **Top 3 Genres: Rock, Alternative & Punk, and Metal together contribute more than 77% of the total count, with Rock being the standout genre.**
* **Niche Genres: Genres like Classical, Heavy Metal, Soundtrack, and TV Shows have minimal representation, collectively accounting for only 1.91% of the total count.**

**Recommendations:**

* **Focus on Popular Genres: Given the high percentage of Rock, Alternative & Punk, and Metal, consider promoting or creating content that aligns with these genres to cater to the largest audience.**
* **Target Niche Audiences: Explore opportunities to grow less popular genres like Jazz, Reggae, and Classical by offering tailored experiences or specialized content for their audiences.**
* **Expand Genre Diversity: If looking to diversify, consider experimenting with Electronica/Dance, Pop, or Hip Hop/Rap, which have potential but currently contribute less.**

**Query:**

**WITH genre\_counts AS (**

**SELECT**

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

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

**FROM**

**customer c**

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

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

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

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

**WHERE**

**c.country = 'USA'**

**GROUP BY**

**g.name**

**),**

**total\_count AS (**

**SELECT**

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

**FROM**

**customer c**

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

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

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

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

**WHERE**

**c.country = 'USA'**

**)**

**SELECT**

**gc.genre\_name,**

**gc.genre\_count,**

**(gc.genre\_count \* 100 / tc.total\_count) AS percentage**

**FROM**

**genre\_counts gc**

**CROSS JOIN total\_count tc**

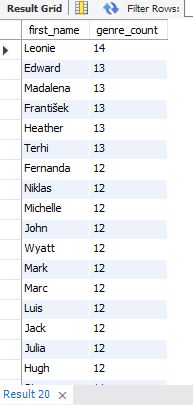
**ORDER BY**

**gc.genre\_count DESC;**

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

**Answer:**

**All 59 customers have purchased records from a minimum of 3 distinct genres, with Leonie Kohler leading as the top buyer.**



**Insights:**

* **Top 5 Individuals: The top 5 individuals with the highest genre counts are Leonie (14), and Edward, Madalena, František, Heather, and Terhi (each with 13). These individuals show strong engagement in the genre.**
* **Consistent Engagement: Many individuals, such as Fernanda, Niklas, Michelle, and others, are closely following with 12 genre counts, indicating a high level of participation across a large group.**
* **Lower Engagement: There is a noticeable drop in engagement starting from Aaron, Eduardo, and Dan, each with 9 counts, and even lower for others further down the list, like Robert with 5 counts.**

**Recommendations:**

* **Recognize Top Contributors: Acknowledge individuals with higher genre counts (e.g., Leonie and those with 13 counts) as they show consistent engagement. Consider incentivizing them to further enhance their involvement.**
* **Increase Engagement for Lower Count Users: Investigate why individuals with lower counts, such as Robert, have fewer interactions. Offering targeted incentives or support could boost their activity.**
* **Analyse Genre Preferences: Understand the type of content or engagement driving high genre counts for specific individuals, especially those in the top ranks, to replicate successful strategies across the board.**

**Query:**

**SELECT**

**c.first\_name, count(distinct g.name) AS genre\_count**

**FROM**

**customer c**

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

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

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

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

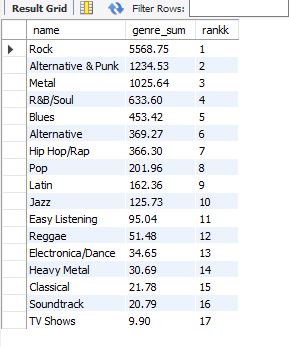
**GROUP BY c.first\_name**

**HAVING count(distinct g.name) >=3**

**ORDER BY genre\_count DESC;**

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

**Answer:**



**Insights:**

* **Top Genre (Rock): Rock leads by a significant margin with a genre sum of 5568.75, holding the top spot.**
* **High Ranking Genres: Alternative & Punk and Metal follow closely, ranking second and third with sums of 1234.53 and 1025.64, respectively.**
* **Lower Ranking Genres: TV Shows has the lowest genre sum at 9.9, with other less popular genres like Classical and Soundtrack also falling at the lower end of the ranking.**

**Recommendations:**

* **Focus on High-Performing Genres: Given the dominance of Rock, Alternative & Punk, and Metal, these genres should be prioritized for further development and engagement strategies.**
* **Explore Niche Genres: Genres such as Classical and Soundtrack are lower in the rankings but could be targeted for niche marketing efforts to grow their customer base.**
* **Diversify Content Offering: To attract a wider audience, consider diversifying the offerings in Pop, Hip Hop/Rap, and Electronica/Dance, which show potential for further growth.**

**Query:**

**select g.name,sum(i.total) as genre\_sum, rank() over(order by sum(i.total) desc) as rankk**

**from customer c inner join invoice i on c.customer\_id = i.customer\_id inner join invoice\_line il on i.invoice\_id = il.invoice\_id inner join track t on il.track\_id = t.track\_id**

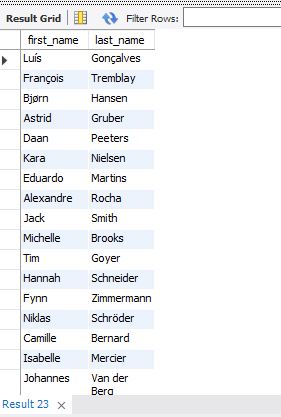
**inner join genre g on t.genre\_id = g.genre\_id**

**where c.country = "USA"**

**group by g.name;**

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

**Answer:**



**Query:**

**select first\_name, last\_name from customer c**

**left join (**

**select \***

**from invoice**

**where invoice\_date > (select max(invoice\_date) from invoice) - interval 3 month) prev\_3\_months**

**on prev\_3\_months.customer\_id = c.customer\_id**

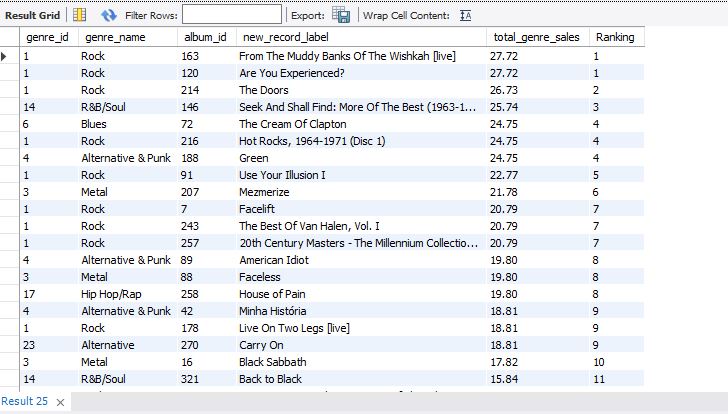
**where invoice\_id is null;**

**Subjective Questions**

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

**Answer:**

**Rock is the most popular genre in the USA based on sales, the top 3 albums from the Rock genre should be prioritized for advertising and promotion.**



**Insights:**

* **Top Genre (Rock): The Rock genre dominates with a strong presence across multiple albums, especially with albums like *"From The Muddy Banks Of The Wishkah [live]"* and *"Are You Experienced?"* both contributing significantly to sales, each with a genre sale of 27.72.**
* **R&B/Soul: R&B/Soul also performs well, with *"Seek And Shall Find: More Of The Best (1963-1981)"* reaching a sales total of 25.74 and securing a ranking of 3.**
* **Alternative & Punk: Alternative & Punk has a notable presence, with albums like *"Green"* and *"American Idiot"* reaching sales totals of 24.75 and 19.8 respectively.**
* **Blues and Metal: Blues and Metal also appear strong in the market, with albums like *"The Cream Of Clapton"* and *"Mezmerize"* earning respectable sales totals, but Rock continues to outperform these genres.**

**Recommendations:**

1. **Prioritize Rock Albums:**
   * **Album 1: *"From The Muddy Banks Of The Wishkah [live]"* (Sales: 27.72) – This album is a top performer and should be heavily promoted, capitalizing on its strong sales and popularity.**
   * **Album 2: *"Are You Experienced?"* (Sales: 27.72) – With equal sales performance to the first album, this should be targeted for advertising efforts aimed at fans of classic rock in the USA.**
   * **Album 3: *"Hot Rocks, 1964-1971 (Disc 1)"* (Sales: 24.75) – Another top performer, this album offers a compelling collection that is likely to resonate well with a diverse rock audience.**
2. **Target R&B/Soul and Alternative & Punk Genres:**
   * **R&B/Soul: Promote *"Seek And Shall Find: More Of The Best (1963-1981)"* (Sales: 25.74) as a key offering in the USA, as it performs well and appeals to a broad audience in this genre.**
   * **Alternative & Punk: Albums like *"Green"* (Sales: 24.75) and *"American Idiot"* (Sales: 19.8) should be considered for targeted campaigns to reach fans of alternative rock and punk music, as these albums show strong performance.**

**By focusing on these high-performing albums from Rock, R&B/Soul, and Alternative & Punk, the advertising and promotional efforts can capitalize on genres with the highest sales potential in the USA.**

**Query:**

**-- 1**

**SELECT**

**g.genre\_id,**

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

**al.album\_id,**

**al.title AS new\_record\_label,**

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

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

**FROM**

**genre g**

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

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

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

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

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

**WHERE**

**c.country = 'USA'**

**GROUP BY**

**g.genre\_id, g.name, al.album\_id,**

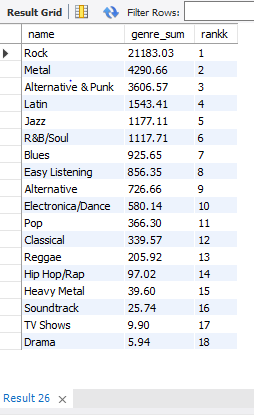
**al.title**

**ORDER BY**

**total\_genre\_sales DESC;**

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

**Answer:**



**Insights:**

* **Rock's Universal Appeal: Rock music remains a dominant genre both in the USA and globally, continuing to capture wide-reaching popularity.**
* **Regional Preferences: There are noticeable differences in genre preferences between the USA and other countries, with Metal, Latin, and Jazz gaining more traction internationally.**
* **Growing Genres Outside the USA: Outside of the US, Metal, Latin, and Jazz are emerging as stronger genres, indicating a shift in musical tastes in global markets.**

**Recommendations:**

* **Focus on Rock Worldwide: Maintain a strong global strategy that emphasizes Rock music, as it consistently performs well across all regions.**
* **Tailored Regional Strategies: Adapt inventory management and promotional efforts to cater to the specific genre preferences of each region.**
* **Promote Metal, Latin, and Jazz Globally: Increase focus on promoting Metal, Latin, and Jazz in markets outside the USA where these genres are gaining popularity.**
* **Localized Marketing Campaigns: Develop region-specific marketing initiatives that resonate with local musical tastes and cultural preferences.**
* **Explore Genre Expansion: Investigate strategies to introduce and promote lesser-known genres both in the US and internationally, expanding the diversity of musical offerings in each market.**

**Query:**

**-- 2**

**SELECT**

**g.name,**

**SUM(i.total) AS genre\_sum,**

**RANK() OVER (ORDER BY SUM(i.total) DESC) AS rankk**

**FROM**

**customer c**

**INNER JOIN**

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

**INNER JOIN**

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

**INNER JOIN**

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

**INNER JOIN**

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

**WHERE**

**c.country != 'USA'**

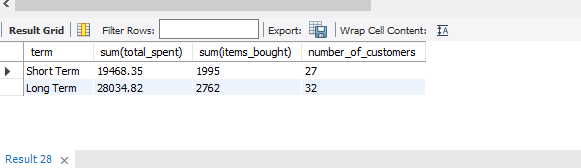
**GROUP BY**

**g.name;**

**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?**

**Answer:**

**Classified customers as Long and Short Term On the basis of amount spend and number of items bought.**



**Insights:**

* **Total Spending: Long-term customers have spent significantly more ($28,034.82) compared to short-term customers ($19,468.35). This suggests that long-term customers are more likely to make larger purchases over time, showing greater overall value to the business.**
* **Items Bought: Long-term customers also buy more items, with a total of 2,762 items purchased, compared to 1,995 by short-term customers. This indicates a higher frequency of purchases and a larger basket size for long-term customers.**
* **Customer Loyalty: The higher spending and item count for long-term customers demonstrate stronger customer loyalty. These customers not only buy more frequently but also spend more per transaction, indicating a deepening relationship with the brand over time.**
* **New vs. Long-term Customers: While short-term customers are still valuable, they tend to purchase fewer items and spend less per transaction. This could suggest that new customers may not yet be fully engaged with the brand or product offerings.**

**Recommendations:**

* **Reward Long-term Customers: Develop targeted loyalty programs for long-term customers, offering exclusive discounts or rewards based on their continued purchasing behaviour. This can further enhance their loyalty and encourage even larger purchases.**
* **Increase Engagement with Short-term Customers: For short-term customers, implement personalized offers or promotions aimed at increasing their basket size and encouraging repeat purchases. This could include bundling offers or loyalty incentives to transition them into long-term customers.**
* **Strengthen Retention Strategies: Focus on customer retention by introducing more engagement points such as personalized email marketing, offering cross-sell or up-sell recommendations based on previous purchases, and nurturing their relationship with the brand to increase their lifetime value.**
* **Incentivize Repeat Purchases: Consider implementing incentives like a points-based system, where customers earn rewards for continued purchases. This could motivate both new and existing customers to spend more and purchase more frequently.**
* **Enhance Customer Experience: Deliver a seamless and personalized shopping experience to keep customers engaged. Providing excellent customer service, tailored recommendations, and easy returns could help in boosting retention and increasing the value of both short-term and long-term customers.**

**Query:**

**-- 3**

**-- CTE to calculate customer-related metrics such as last purchase date, first purchase date, total spent, items bought, frequency, and the number of days a customer has been active.**

**WITH cte AS (**

**SELECT**

**i.customer\_id,**

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

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

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

**SUM(quantity) AS items\_bought,**

**COUNT(i.customer\_id) AS frequency,**

**ABS(TIMESTAMPDIFF(DAY, MAX(invoice\_date), MIN(invoice\_date))) AS customer\_since\_days**

**FROM invoice i**

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

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

**GROUP BY i.customer\_id**

**),**

**-- CTE to classify customers into 'Long Term' or 'Short Term' based on their active days compared to the average customer active days.**

**long\_short\_term AS (**

**SELECT**

**total\_spent,**

**items\_bought,**

**frequency,**

**CASE**

**WHEN customer\_since\_days > (SELECT AVG(customer\_since\_days) AS average\_days FROM cte)**

**THEN 'Long Term'**

**ELSE 'Short Term'**

**END AS term**

**FROM cte**

**)**

**-- Final selection of the sum of total spent, items bought, and the count of customers, grouped by the classification of 'Long Term' or 'Short Term'.**

**SELECT**

**term,**

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

**SUM(items\_bought) AS total\_items\_bought,**

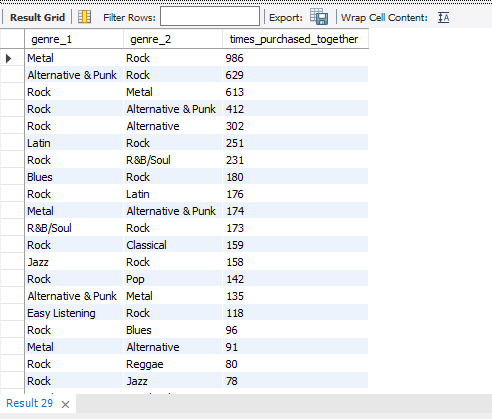
**COUNT(frequency) AS number\_of\_customers**

**FROM long\_short\_term**

**GROUP BY term;**

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

**Answer:**



**Product Affinity Analysis Insights:**

* **Frequent Genre Pairings: The most commonly purchased genre combinations include:**
  + **Rock & Metal (986 times)**
  + **Alternative & Punk & Rock (629 times)**
  + **Rock & Alternative & Punk (412 times)**
  + **Rock & R&B/Soul (231 times)**
  + **Blues & Rock (180 times)**

**These insights reveal that customers who purchase Rock albums tend to buy Metal, Alternative & Punk, and R&B/Soul genres frequently as well.**

* **Cross-Genre Preferences: Customers also show an affinity for pairing genres across a broader spectrum:**
  + **Latin & Rock (251 times)**
  + **Metal & Alternative & Punk (174 times)**
  + **Jazz & Rock (158 times)**
  + **Pop & Rock (142 times)**

**These combinations suggest a strong tendency to purchase genres from both classic and contemporary music styles.**

* **Emerging Pairings: Some lesser-known pairings still show interesting customer behaviour:**
  + **Metal & R&B/Soul (65 times)**
  + **Latin & R&B/Soul (25 times)**
  + **Electronica/Dance & Rock (49 times)**

**These emerging trends may provide opportunities for cross-promotion and expanding genre exposure.**

**Recommendations for Product Recommendations & Cross-Selling:**

1. **Bundle Related Genres: Based on frequently purchased genre pairs, create bundle offers that include both Rock and its frequent pairings, such as:**
   * **Rock & Metal**
   * **Rock & Alternative & Punk**
   * **Rock & R&B/Soul**
   * **Blues & Rock**

**Offering these bundles at a discounted price could increase basket size and customer satisfaction.**

1. **Cross-Sell Genres with Similar Customer Preferences: For customers who buy one genre (e.g., Rock), recommend similar genres that are frequently bought together, such as:**
   * **If a customer buys a Rock album, suggest Metal, Alternative & Punk, or R&B/Soul.**
   * **For Latin listeners, suggest Rock or Blues based on purchasing patterns.**
2. **Targeted Marketing Campaigns: Use the affinity data to create targeted marketing campaigns. For example:**
   * **For Rock fans: Promote new releases from Metal, Alternative & Punk, or R&B/Soul artists alongside Rock albums.**
   * **For Jazz lovers: Promote bundles of Jazz & Blues or Jazz & Rock.**

**This can be done via email campaigns, product recommendations on websites, or personalized offers in mobile apps.**

1. **Explore Cross-Genre Promotions: Since some genres (e.g., Latin & Metal, Electronica/Dance & Rock) appear less often but still show some level of affinity, consider experimenting with cross-genre promotions to introduce customers to new music styles.**
2. **Optimize In-Store or Online Displays: Based on genre affinity, organize physical or digital stores to promote albums from genres often purchased together. For example:**
   * **Create a “Rock & Metal” display section in the store.**
   * **On an online platform, use the “Frequently Bought Together” feature to suggest albums from related genres.**

**By leveraging product affinity data, music retailers can enhance customer satisfaction through personalized experiences, increase sales via cross-selling, and help customers discover new music that aligns with their preferences.**

**Query:**

**-- 4**

**WITH track\_combinations AS (**

**SELECT**

**il1.track\_id AS track\_id\_1,**

**il2.track\_id AS track\_id\_2,**

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

**FROM**

**invoice\_line il1**

**JOIN**

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

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

**GROUP BY**

**il1.track\_id, il2.track\_id**

**),**

**genre\_combinations AS (**

**SELECT**

**t1.genre\_id AS genre\_id\_1,**

**t2.genre\_id AS genre\_id\_2,**

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

**FROM**

**track\_combinations tc**

**JOIN**

**track t1 ON tc.track\_id\_1 = t1.track\_id**

**JOIN**

**track t2 ON tc.track\_id\_2 = t2.track\_id**

**WHERE**

**t1.genre\_id <> t2.genre\_id**

**GROUP BY**

**t1.genre\_id, t2.genre\_id**

**)**

**SELECT**

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

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

**gc.times\_purchased\_together**

**FROM**

**genre\_combinations gc**

**JOIN**

**genre g1 ON gc.genre\_id\_1 = g1.genre\_id**

**JOIN**

**genre g2 ON gc.genre\_id\_2 = g2.genre\_id**

**ORDER BY**

**gc.times\_purchased\_together DESC;**

**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?**

**Answer:**

**Correlation with Local Demographics and Economic Factors:**

1. **Economic Prosperity and Spending on Entertainment:**
   * **In countries with higher GDP per capita and greater disposable income, such as Canada, Ireland, and France, customers generally have more financial freedom to invest in discretionary products, including entertainment. As a result, there is often a higher willingness to spend on digital music and entertainment services. This trend is particularly noticeable in affluent populations who view music as a key component of lifestyle, with higher purchases of premium content like concert tickets, exclusive albums, and subscriptions to music streaming platforms.**
2. **Cultural Influence on Music Consumption:**
   * **Countries that boast a rich musical heritage or a strong local and international music presence, such as Brazil, France, and India, often see higher levels of engagement in music-related activities. These nations have developed robust local music industries and cultural traditions tied to various musical genres. As a result, consumers are not only more likely to purchase local music but also show strong interest in international artists. This high level of engagement boosts both physical and digital music sales, as well as the consumption of diverse music genres, creating a dynamic market for cross-cultural music exchanges.**
3. **Digital Economy and E-commerce Infrastructure:**
   * **Countries with well-developed digital economies and e-commerce infrastructures, such as Ireland, Czech Republic, and Portugal, provide an ideal environment for consumers to engage in online music purchases. The convenience of digital payment systems, high internet penetration, and widespread access to streaming services enables consumers to purchase music seamlessly and frequently. In these regions, customers are more likely to embrace digital platforms, where they can easily purchase, stream, or download music. This trend is fueled by a mature digital landscape, fostering online consumption in both physical and digital formats.**
4. **Affluence and Entertainment Spending:**
   * **Countries with higher GDP per capita and disposable income, such as Canada, Ireland, and France, tend to have a more affluent consumer base that allocates a higher proportion of their income toward entertainment and leisure activities. This includes spending on music and related content like subscriptions to streaming services, concert tickets, and digital music purchases. Consumers in these countries are more likely to engage with both local and international artists, supporting the global music industry and driving spending in both traditional and digital music markets.**

**In summary, economic factors such as higher disposable income, GDP per capita, and advanced digital economies are strongly correlated with greater customer expenditure on music and entertainment. Similarly, countries with vibrant music cultures and strong engagement with both local and international music foster more active purchasing behaviours, contributing to a thriving market for music products and services.**

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

**Answer:**

**To address customer segmentation and risk profiling, we analyse customer count, total revenue, and number of invoices by country. Using these metrics, we can categorize countries into High-Value, Moderate, and High-Churn areas and devise targeted strategies for each group.**

**1. Categorizing Countries**

**A. High-Value Areas**

* **Definition: Countries with significant total revenue and a large customer base. These areas should be prioritized for premium strategies.**
* **Countries:**
  + **USA (286 customers, $1040.49 revenue)**
  + **Canada (187 customers, $535.59 revenue)**
  + **Brazil (47 customers, $427.68 revenue)**
  + **France (205 customers, $389.07 revenue)**
  + **Germany (113 customers, $334.62 revenue)**

**Strategy:**

* **Premium Pricing: Highlight exclusivity and value.**
* **Personalized Marketing: Leverage customer data to create tailored experiences.**
* **Upselling & Cross-Selling: Recommend complementary products or higher-tier options.**

**B. Moderate Areas**

* **Definition: Countries with moderate revenue and a decent customer base. Focus here on strengthening customer loyalty.**
* **Countries:**
  + **Czech Republic (11 customers, $273.24 revenue)**
  + **United Kingdom (159 customers, $245.52 revenue)**
  + **Portugal (69 customers, $185.13 revenue)**
  + **India (117 customers, $183.15 revenue)**
  + **Ireland (46 customers, $114.84 revenue)**

**Strategy:**

* **Loyalty Programs: Reward consistent purchases with points, discounts, or gifts.**
* **Value-for-Money Offers: Introduce bundles or discounts to encourage repeat purchases.**
* **Regular Communication: Use newsletters and updates to maintain engagement.**

**C. High-Churn Areas**

* **Definition: Countries with low revenue per customer or high customer counts but lower total revenue, indicating the potential for churn or reduced spending.**
* **Countries:**
  + **Argentina (56 customers, $39.60 revenue)**
  + **Denmark (9 customers, $37.62 revenue)**
  + **Italy (47 customers, $50.49 revenue)**
  + **Belgium (8 customers, $60.39 revenue)**
  + **Sweden (51 customers, $75.24 revenue)**

**Strategy:**

* **Identify Churn Causes: Conduct surveys to gather feedback on dissatisfaction points.**
* **Retention Campaigns: Offer exclusive discounts or promotions to retain customers.**
* **Improved Customer Service: Train staff and provide faster resolutions to complaints.**

**2. Key Insights**

**High Customer Count with Low Revenue**

* **Countries: Argentina, Italy, Sweden**
* **Risk: These countries show higher customer numbers but lower overall revenue, indicating lower average spending.**
* **Recommendation: Introduce affordable product bundles and reward frequent purchases.**

**High Revenue but Fewer Customers**

* **Countries: Czech Republic, Norway, Austria**
* **Opportunity: Focus on expanding the customer base while maintaining high spending per customer.**
* **Recommendation: Run campaigns to attract new customers and reward referrals.**

**3. Factor Analysis**

**Based on the data:**

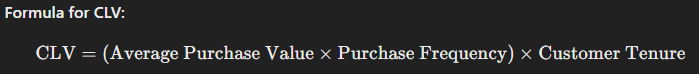
* **Demographics: Countries with younger or lower-income populations might gravitate toward low-value spending patterns.**
* **Economic Conditions: Regions with lower GDP per capita might be more price-sensitive.**
* **Engagement Levels: Countries with low invoice counts suggest underutilized markets.**

**4. Proposed Actions**

| **Segment** | **Focus** | **Tactics** |
| --- | --- | --- |
| **High-Value Areas** | **Maximize value** | **Exclusive product launches, personalized marketing, premium pricing** |
| **Moderate Areas** | **Strengthen engagement** | **Loyalty programs, regular promotions, and enhanced communication** |
| **High-Churn Areas** | **Reduce churn and improve spending** | **Customer feedback surveys, retention discounts, and improved service quality** |

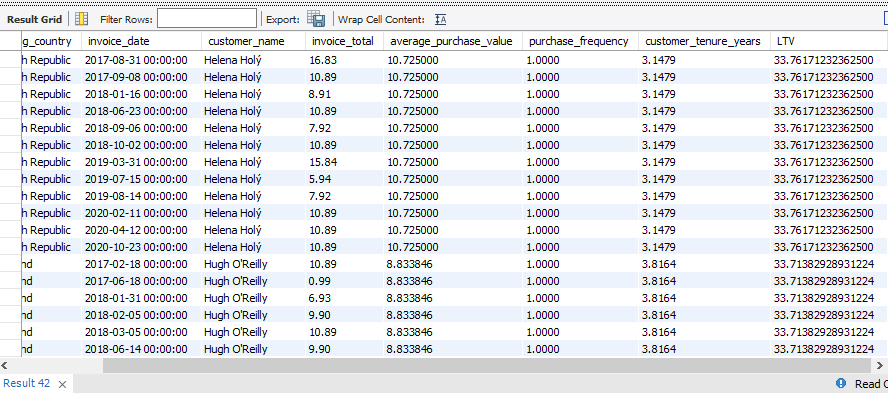
**7.Customer Lifetime Value Modelling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns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

**Answer:**



**Key Factors for Predicting CLV**

1. **Purchase History:**
   * **Customers with higher purchase frequency and larger transaction values tend to have higher CLV.**
2. **Engagement:**
   * **Active participation in loyalty programs and responsiveness to promotions indicate stronger customer relationships.**
3. **Demographics:**
   * **Customers in high-value countries (e.g., USA, Canada, Brazil) typically have higher spending patterns.**
4. **Tenure:**
   * **Longer-tenured customers are more likely to contribute significantly to revenue, but new customers in high-value areas are opportunities for growth.**



**Based on the provided customer data, which includes Customer Lifetime Value (LTV) by country and customer, here are some insights and recommendations:**

**Insights:**

1. **Top Countries by LTV:**
   * **USA has the highest concentration of customers with high LTV values, with notable customers like Dan Miller (31.05), Frank Harris (33.08), Frank Ralston (33.22), and Michelle Brooks (31.38) contributing significantly to the total revenue.**
   * **Canada follows closely behind, with customers like François Tremblay (35.92) and Emma Jones (31.39) showcasing high LTV.**
   * **Other countries with notable LTV figures include France (e.g., Camille Bernard (27.85), Dominique Lefebvre (30.69)) and Brazil (e.g., Alexandre Rocha (18.47), Eduardo Martins (17.03)).**
2. **High Value Customers:**
   * **A few customers have exceptionally high LTV values (e.g., François Tremblay at 35.92, Patrick Gray at 35.30, Wyatt Girard at 35.02, and Helena Holý at 33.76), indicating a strong long-term relationship and likely consistent purchasing habits.**
   * **These customers should be prioritized for loyalty programs, exclusive offers, or personalized marketing to ensure continued engagement and maximize their value.**
3. **Countries with High LTV Customers:**
   * **USA and Canada are strong performers, contributing several high-value customers.**
   * **Other countries like Germany, Brazil, France, and United Kingdom also display promising LTV values across their customer base, showing that LTV is not limited to a few regions but is globally spread.**
4. **Countries with Potential to Improve:**
   * **Some countries have lower LTV values, such as Argentina (e.g., Diego Gutiérrez with 17.53), Denmark (e.g., Kara Nielsen with 11.19), and Netherlands (e.g., Johannes Van der Berg with 18.74). These countries might have customers who either:**
     + **Churned early, or**
     + **Do not engage as frequently.**
   * **In these cases, further investigation into customer retention strategies and engagement programs should be considered.**

**Recommendations:**

1. **High-Value Areas (USA, Canada, France):**
   * **Exclusive Products and Premium Pricing: Consider offering exclusive or premium products to customers in these high-LTV countries.**
   * **Loyalty and Retention Programs: For customers with consistently high LTV, create personalized loyalty programs that reward long-term engagement and high-frequency purchases.**
   * **Targeted Marketing Campaigns: Given the high potential of customers in these regions, implement targeted marketing efforts that highlight the value and benefits of being a loyal customer.**
2. **Moderate-Value Areas (Germany, United Kingdom, Brazil):**
   * **Customer Engagement: Introduce or enhance engagement strategies, such as frequent communication (newsletters, offers), and loyalty programs to increase customer lifetime value.**
   * **Referral Programs: Introduce or strengthen referral programs to leverage existing satisfied customers in these countries to bring in new, high-value customers.**
3. **Low-Value or High-Churn Areas (Argentina, Denmark, Netherlands):**
   * **Churn Analysis and Retention Strategies: Deep dive into churn reasons in these areas. Is it due to customer dissatisfaction, high competition, or a lack of personalized offers?**
   * **Customer Feedback: Gather feedback via satisfaction surveys or direct outreach to understand the barriers to purchase and develop retention strategies.**
   * **Discounts or Offers: Implement discounts, bundled offers, or better customer service to incentivize customers in these regions to return.**
4. **Cross-Country Insights:**
   * **Purchase Frequency: Focus on countries with high LTV but low frequency to improve purchase frequency through personalized campaigns, re-engagement emails, or retargeting ads.**
   * **Tenure and Loyalty: For countries with longer customer tenure and higher LTV, continue nurturing these relationships and reward loyalty through exclusive offers, anniversary deals, or special privileges.**
5. **Global Strategies:**
   * **Invest in High-LTV Customers: Across all countries, focus on customers with high LTV by offering them personalized experiences (e.g., VIP treatment, priority access to new products, etc.).**
   * **Leverage LTV Segmentation: Segment customers based on LTV, identifying the top 10-20% with the highest value and tailoring your marketing and retention strategies specifically for them.**

**Query:**

**-- 7**

**WITH cte AS (**

**SELECT**

**inv.customer\_id,**

**inv.billing\_country,**

**inv.invoice\_date,**

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

**inv.total AS invoice\_total**

**FROM**

**invoice inv**

**LEFT JOIN**

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

**GROUP BY**

**inv.customer\_id, inv.billing\_country, inv.invoice\_date, c.first\_name, c.last\_name, inv.total**

**),**

**cte2 AS (**

**SELECT**

**customer\_id,**

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

**COUNT(invoice\_id) AS total\_invoices,**

**DATEDIFF(MAX(invoice\_date), MIN(invoice\_date)) AS customer\_tenure\_days, -- Customer tenure**

**COUNT(DISTINCT(invoice\_date)) AS active\_periods**

**FROM**

**invoice**

**GROUP BY**

**customer\_id**

**),**

**cte3 AS (**

**SELECT**

**customer\_id,**

**total\_revenue / total\_invoices AS average\_purchase\_value,**

**total\_invoices / active\_periods AS purchase\_frequency,**

**customer\_tenure\_days / 365.0 AS customer\_tenure\_years -- convert tenure into years**

**FROM**

**cte2**

**)**

**SELECT**

**cte.customer\_id,**

**cte.billing\_country,**

**cte.invoice\_date,**

**cte.customer\_name,**

**cte.invoice\_total,**

**cte3.average\_purchase\_value,**

**cte3.purchase\_frequency,**

**cte3.customer\_tenure\_years,**

**cte3.average\_purchase\_value \* cte3.purchase\_frequency \* cte3.customer\_tenure\_years AS LTV**

**FROM**

**cte**

**LEFT JOIN**

**cte3 ON cte.customer\_id = cte3.customer\_id**

**ORDER BY**

**LTV DESC, cte.customer\_name, cte.invoice\_date;**

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

**Answer:**

**Methods to Measure Impact:**

* **Compare Pre-Campaign and Post-Campaign Metrics: Compare the number of new customers acquired before and after the promotional campaign. This can also include measuring website traffic or sign-ups during the campaign to determine if the campaign generated initial interest.**
* **Post-Campaign Surveys: Use surveys or feedback mechanisms to understand if customers are satisfied with the promotional campaign and if it influenced their decision to remain loyal to the brand.**
* **Customer Lifetime Value (CLV): Track the CLV of customers who were acquired or retained through the promotional campaigns and compare it to the overall customer base. If CLV increases after the campaign, it suggests that the campaign positively impacted retention.**
* **Tracking Promotion Redemptions: If the campaign includes specific promotional codes or discount offers, track the redemption rates. A higher redemption rate indicates that the campaign has been effective in driving sales.**

**Measuring Campaign Impact with Data:**

**To properly measure the impact of your campaigns, you will need access to the following data:**

* **Customer Acquisition Data: Number of new customers during and after the campaign.**
* **Customer Retention Data: Repeat purchase rates and retention data for customers who engaged with the campaign.**
* **Sales Data: Total revenue, order values, and conversion rates during the campaign.**
* **Campaign-Specific Metrics: Discount codes, email open rates, event attendance, etc.**

**Recommended Analytical Approaches:**

* **Customer Segmentation: Segment your customer base to understand which segments (age, location, gender, etc.) are most responsive to the campaign and have the highest retention and sales growth. This will help you target future campaigns more effectively.**
* **Multi-Touch Attribution: Track how different touchpoints (e.g., email, ads, events) contribute to customer acquisition and retention. This helps identify which promotional activities had the most significant impact.**

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

**Answer:**

**Without predefined questions, my approach would involve:**

**1. Define Business Objectives & KPIs: Clarify goals like customer acquisition, retention, and sales growth. KPIs include customer count, repeat purchase rate, customer lifetime value (CLV), and ROI.**

**2. Gather Data by Querying: Collect campaign details, customer behaviour (demographics, purchase history), sales data, and retention data before, during, and after campaigns.**

**3. Segment Customers: Segment based on demographics, behaviour, and campaign engagement to understand which groups respond best to promotions.**

**4. Analyse Impact:**

* **Customer Acquisition: Measure new customers and cost per acquisition.**
* **Customer Retention: Track repeat purchase rate and churn, and analyze CLV.**
* **Sales: Compare sales growth, revenue per customer, and sales uplift during the campaign.**

1. **Generate Insights & Recommendations: Identify effective segments, campaign types, and areas for improvement. Suggest budget allocation and strategies for customer re-engagement.**
2. **Monitor & Optimize: Continuously track and optimize campaigns based on real-time data and performance reviews.**

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

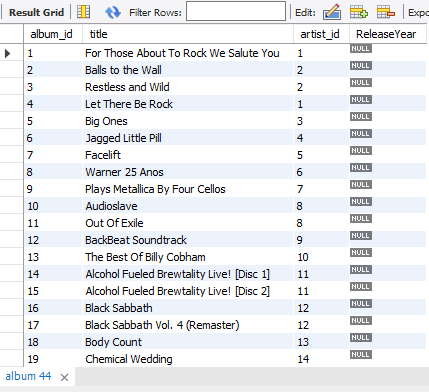
**Answer:**

**We can modify the "Albums" table by adding a new column called "Release Year" with the data type INTEGER to store the release year of each album using the following SQL query:**

**ALTER TABLE Album**

**ADD COLUMN ReleaseYear INTEGER;**

**select \* from album;**



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

**Answer:**

**Query:**

**WITH customer\_tracks AS (**

**SELECT**

**i.customer\_id,**

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

**FROM**

**invoice i**

**JOIN**

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

**GROUP BY**

**i.customer\_id**

**),**

**total\_customer\_spending AS (**

**SELECT**

**c.country,**

**c.customer\_id,**

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

**ct.total\_tracks**

**FROM**

**customer c**

**JOIN**

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

**JOIN**

**customer\_tracks ct ON c.customer\_id = ct.customer\_id**

**GROUP BY**

**c.country, c.customer\_id, ct.total\_tracks**

**)**

**SELECT**

**cs.country,**

**COUNT(DISTINCT cs.customer\_id) AS number\_of\_customers,**

**ROUND(AVG(cs.total\_spent),2) AS average\_amount\_spent\_per\_customer,**

**ROUND(AVG(cs.total\_tracks),2) AS average\_tracks\_purchased\_per\_customer**

**FROM**

**total\_customer\_spending cs**

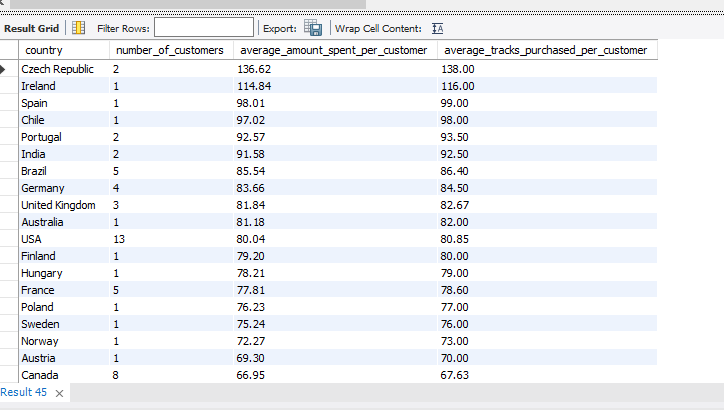
**GROUP BY**

**cs.country**

**ORDER BY**

**average\_amount\_spent\_per\_customer DESC;**

**Output:**



**Insights from the Data:**

1. **Top-Spending Countries:**
   * **Czech Republic has the highest average amount spent per customer (136.62) and the highest average number of tracks purchased per customer (138).**
   * **Other high-spending countries include Ireland (114.84), Spain (98.01), and Chile (97.02), all of which have higher spending than the general average.**
2. **Countries with Lower Spend:**
   * **Denmark has the lowest average amount spent per customer (37.62) and average tracks purchased (38), which may indicate either limited customer base or less interest in purchasing tracks.**
   * **Argentina (39.6) and Italy (50.49) also have relatively low spending and track purchases.**
3. **Countries with Small Customer Bases:**
   * **Several countries have only 1 or 2 customers, such as Spain, Chile, Ireland, Italy, Argentina, and Denmark, which might skew the data due to the small sample size. These countries may need further analysis or targeted marketing to increase customer numbers.**
4. **Consistent Spending Across Larger Customer Bases:**
   * **Countries like USA, Canada, and Brazil have a larger number of customers (5 to 13) but show lower average spend per customer (USA: 80.04, Canada: 66.95, Brazil: 85.54). These may represent larger but more price-sensitive customer bases.**
5. **High Spending with Lower Track Purchases:**
   * **In countries like France (average spend 77.81, average tracks purchased 78.6), customers spend relatively more but purchase fewer tracks, which could indicate higher-value purchases (e.g., premium albums or bundles).**

**Recommendations:**

1. **Focus on High-Value Countries:**
   * **Czech Republic, Ireland, Brazil, and Germany should be prioritized for exclusive offers or loyalty programs since they exhibit high spending and a good number of tracks purchased per customer.**
   * **Personalized campaigns targeting these countries can increase customer retention and further raise average spending.**
2. **Increase Engagement in Low-Spending Countries:**
   * **Denmark, Argentina, and Italy should be analysed further for their low average spending. Consider launching targeted marketing campaigns or providing discounts to incentivize purchases and improve engagement.**
3. **Address Small Customer Bases:**
   * **Countries with only 1 or 2 customers (such as Spain, Chile, Italy, Argentina, and Denmark) may benefit from localized campaigns to increase customer acquisition. Offering entry-level pricing, free trials, or introducing new localized music events could help expand the customer base.**
4. **Customer Loyalty Programs:**
   * **Implement loyalty programs in countries with larger customer bases like USA, Canada, and Brazil to boost retention and spending. Offering discounts, early access to new releases, or exclusive content for repeat customers could increase the number of tracks purchased.**
5. **Targeted Promotions:**
   * **For high-value customers in countries like Czech Republic and France, launch promotions for premium albums, bundles, or limited-edition tracks that align with their higher average spend but lower track purchases. This could drive up the number of tracks purchased while maintaining their high spend levels.**
6. **Leverage Data for Regional Customization:**
   * **Tailor marketing strategies to specific regions by considering cultural preferences, local trends, and past purchase behavior. For example, offering region-specific bundles or event-driven promotions may resonate more with local customers, leading to higher engagement and sales.**