Objective Questions

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

There are two main methods to replace NULL values with specific values in a database:

* **Using the COALESCE() Function**

The COALESCE() function returns the first non-NULL value from a list. It is commonly used to provide a default value when a column contains NULL.

**Syntax:** COALESCE(column\_name, 'default\_value')

* **Using the UPDATE Statement**

The UPDATE statement is used to change existing values in a table, including replacing NULLs based on specified conditions.

**Syntax:** UPDATE table\_name SET column\_name = value WHERE condition;

* **Handling NULL Values in Specific Tables:**

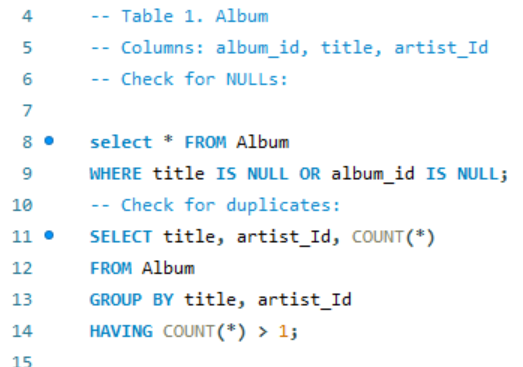
NULL values may be found in several columns across different tables, such as:

1. **Customer Table:** company, state, phone, fax
2. **Track Table:** composer

* **Replacement Strategy:**

1. For **numeric fields**, replace NULLs with 0.
2. For **text fields**, use placeholders such as 'None', 'Unknown', or 'Not Provided'.
3. This approach ensures data consistency and improves the accuracy of analysis.

The SQL query for one table is provided below. Please refer to the attached SQL file for the query related to the other table.



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

**Explanation:**

* Create a CTE named top\_selling\_tracks\_artists to simplify and structure the main query. It holds pre-processed data including track names, artist names, genres, sales totals, and ranks.
* Joins multiple related tables: invoices, invoice lines, tracks, albums, artists, and genres. This brings together all necessary info to link sales data with specific music tracks and artists.
* Applies a filter to include only U.S. sales using WHERE i.billing\_country = 'USA'. This ensures the data reflects only purchases made by customers in the United States.
* Groups the data by track, artist, and genre, then sums up total sales per track. It also uses RANK() to assign a sales rank based on total sales in descending order.
* Final SELECT fetches top 10 tracks from the CTE, ordered by total sales and track name. This gives a clear list of the most popular tracks in the USA by revenue.

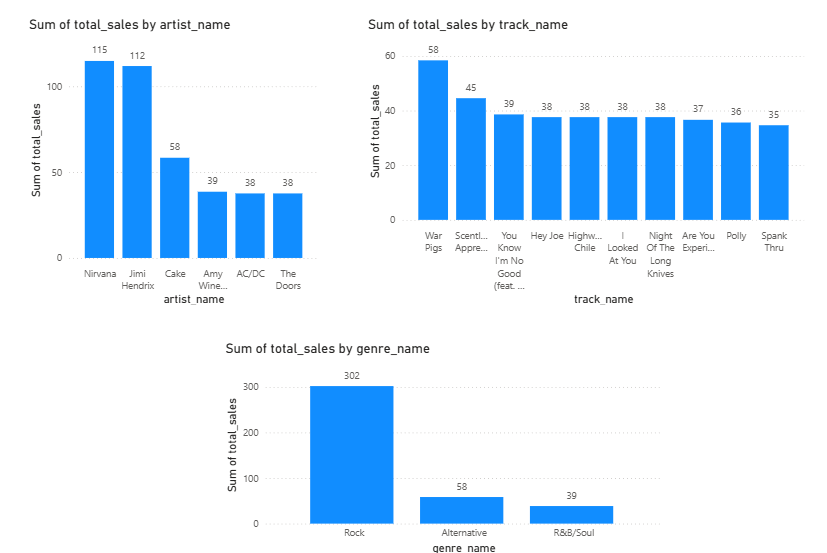
**Insights:**

* **"War Pigs" by Cake** is the top-selling track, with the highest total sales of **58.41**, in the **Alternative** genre.
* The **Rock genre dominates** the list, with **8 out of 10** tracks belonging to it.
* **Nirvana** appears **three times**, showing they are a consistently high-selling artist in the USA.
* **Jimi Hendrix** also has **three tracks** in the top 10, tying with Nirvana in frequency.
* There's only **one R&B/Soul track** ("You Know I'm No Good") by **Amy Winehouse**, ranked 3rd, showing some genre diversity.

Refer the query result and visualizations below :







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

**Explanation: customer demographic breakdown by country, state and city**

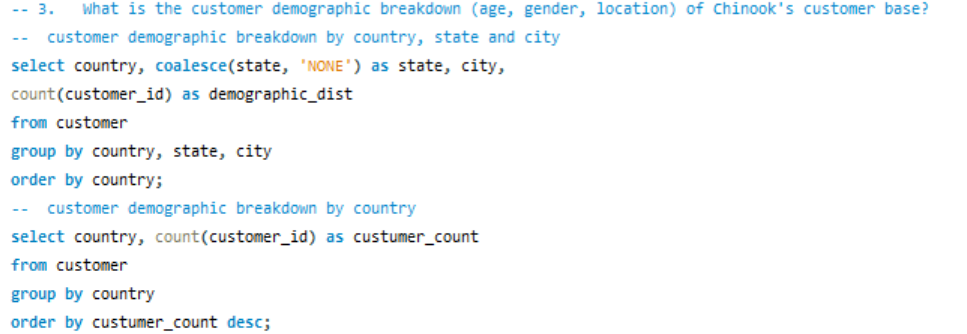
* Select Columns: The query selects country, state (with NULL replaced by 'NONE'), and city from the customer table.
* Count of Customers: It counts the number of unique customer\_id values in each combination of country, state, and city.
* Grouping & Ordering: The results are grouped by country, state, and city, and ordered by country.

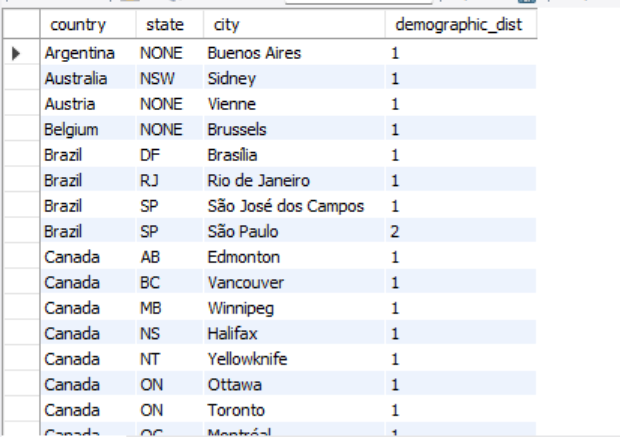
**Explanation: customer demographic breakdown by country**

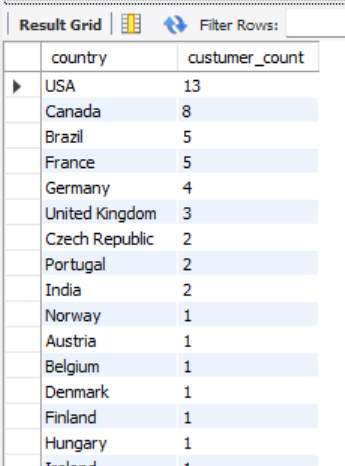
* Select Columns: The query selects country from the customer table.
* Count of Customers: It counts the number of unique customer\_id values in each country
* Grouping & Ordering: The results are grouped by country and order by customer count in descending order

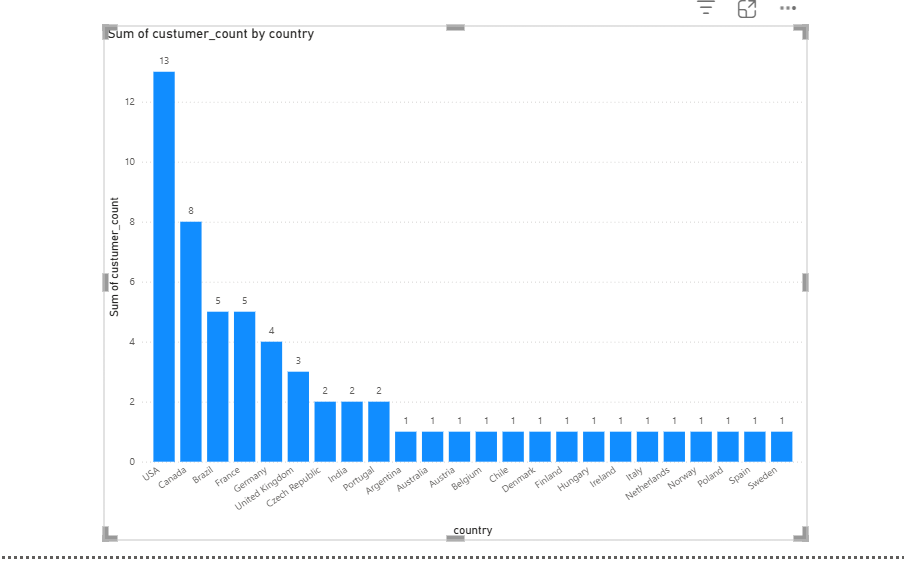
**Insight:**

* The customer demographic breakdown based on location is highly diversified.
* The USA has the highest representation with 13 entries, followed by Canada with 8 and Brazil with 5.
* Most countries in the list have minimal representation, with only one entry each from several nations.









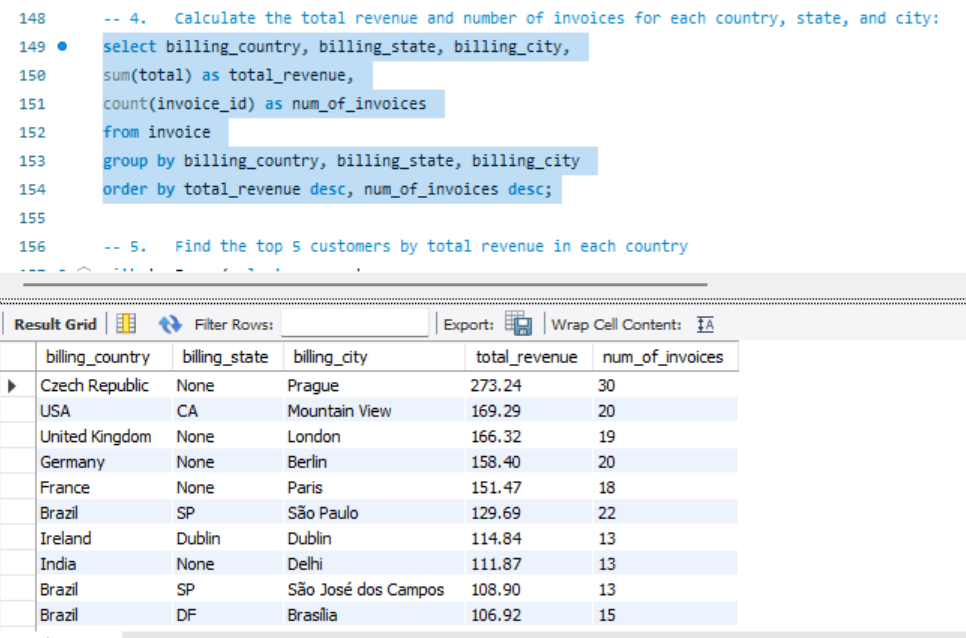
1. Calculate the total revenue and number of invoices for each country, state, and city:

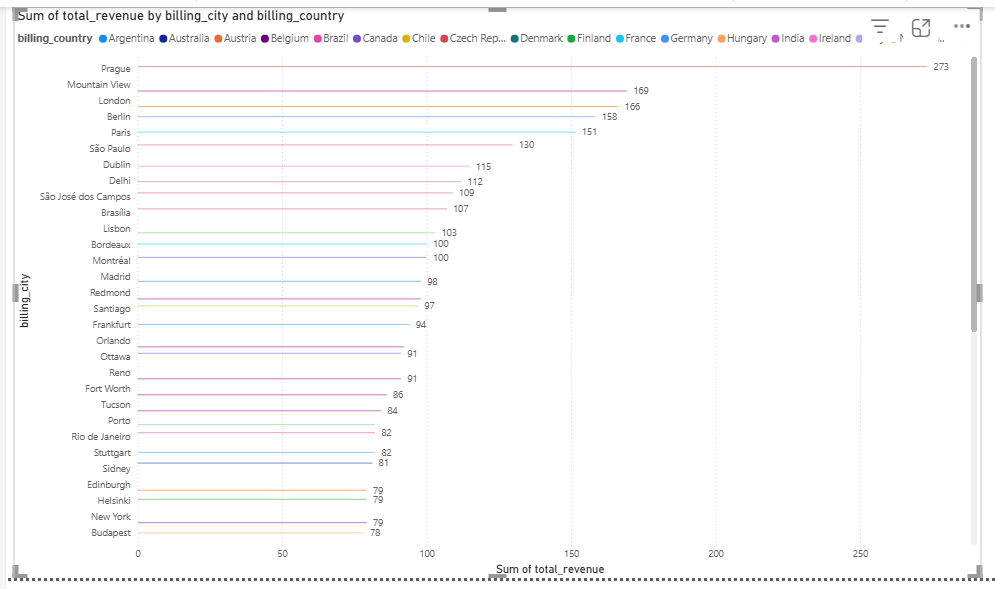
**Explanation:**

* Selected Data: The query retrieves billing\_country, billing\_state, and billing\_city to analyze revenue sources. It also calculates total\_revenue and counts the number of invoices per location.
* Revenue Calculation: SUM(total) aggregates the total sales revenue for each city-level billing location. This helps identify high-revenue geographic areas.
* Invoice Volume: COUNT(invoice\_id) gives the number of transactions in each location.  
  This shows both high-frequency and high-value billing locations.
* Grouping & Sorting: Data is grouped by country, state, and city to aggregate at a local level. Results are sorted by total\_revenue and then by num\_of\_invoices in descending order.

**Insight:**

* **Top Revenue Cities**: Prague, Mountain View, and London lead in total revenue, with Prague generating the highest at $273.24 across 30 invoices. These cities show strong customer engagement and high-value transactions.
* **Diverse Geographic Distribution**: High-revenue locations span Europe, North America, South America, and Asia. This indicates a globally distributed customer base contributing to revenue.
* **Brazil and USA Are Key Markets**: Both countries have multiple cities listed with moderate to high revenue and invoice counts. São Paulo and various U.S. cities show consistent sales activity.
* **Moderate Revenue, Frequent Transactions**: Cities like Porto, Frankfurt, and Bordeaux show lower revenue per invoice but maintain solid invoice volume. This suggests repeat business or smaller purchases across those regions.





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

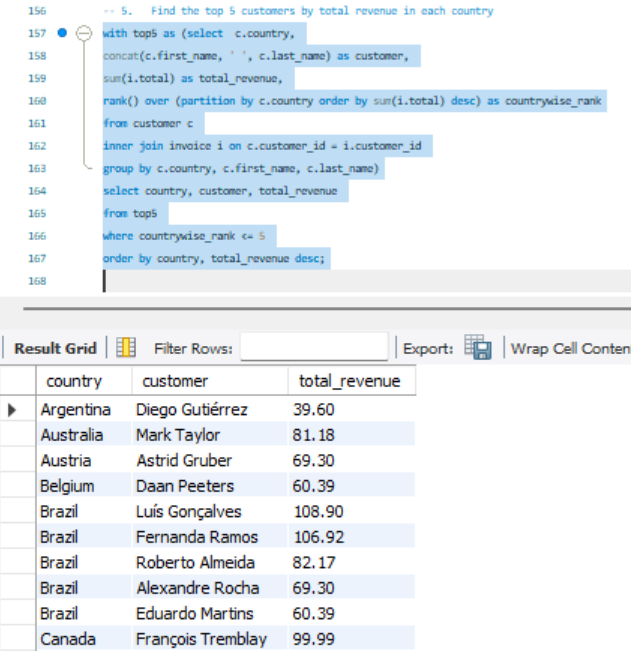
**Explanation:**

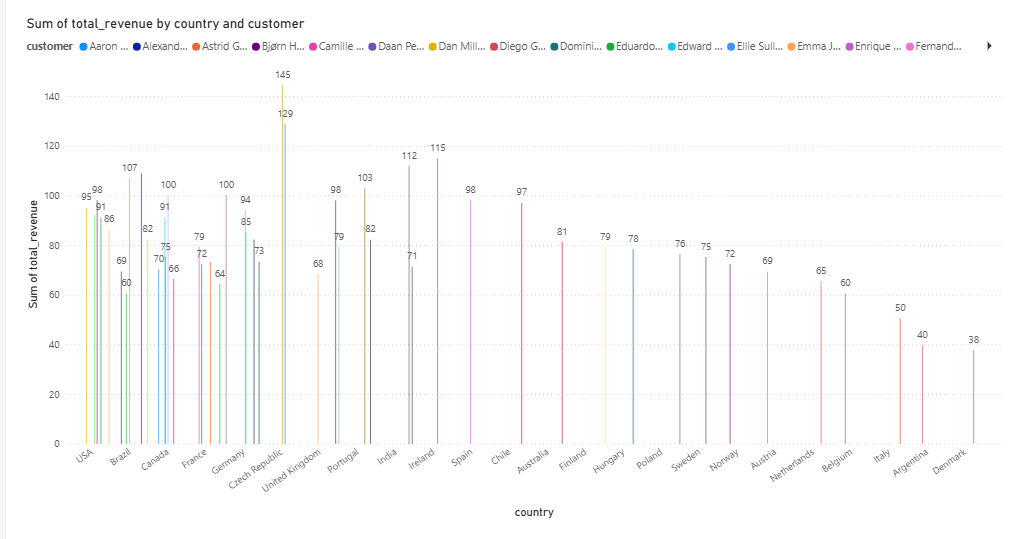
* CTE Creation (top5): The WITH clause defines a Common Table Expression (CTE) named top5. It calculates total revenue per customer and ranks them within each country.
* Joining Tables: The query joins customer and invoice tables using customer\_id.  
  This enables aggregation of total revenue for each customer.
* Ranking Customers: RANK() assigns a rank to each customer within their country based on revenue. Customers with the highest revenue receive rank 1.
* Final Selection: The outer query selects only the top 5 ranked customers per country.  
  The result is sorted by country and descending revenue for clear comparison.

**Insight:**

* **High-Spending Customers**: Czech Republic, Brazil, and India have the highest top-customer revenues, led by František Wichterlová ($144.54). This indicates strong individual customer value in these regions.
* **Consistent Revenue in Brazil and Canada**: Both countries show all five top spots filled, with customers contributing between $60 and $100+. This suggests a broad base of engaged, high-value customers.
* **Europe Dominates Diversity**: Many European countries like France, Germany, and Portugal feature prominently with varying revenue levels. These regions reflect a well-distributed customer base.

**USA and UK Have Strong Mid-Range Earners**: Top U.S. and U.K. customers earn close to $100 each, showing a competitive mid-tier market. However, they don’t top the global chart, indicating room for growth.





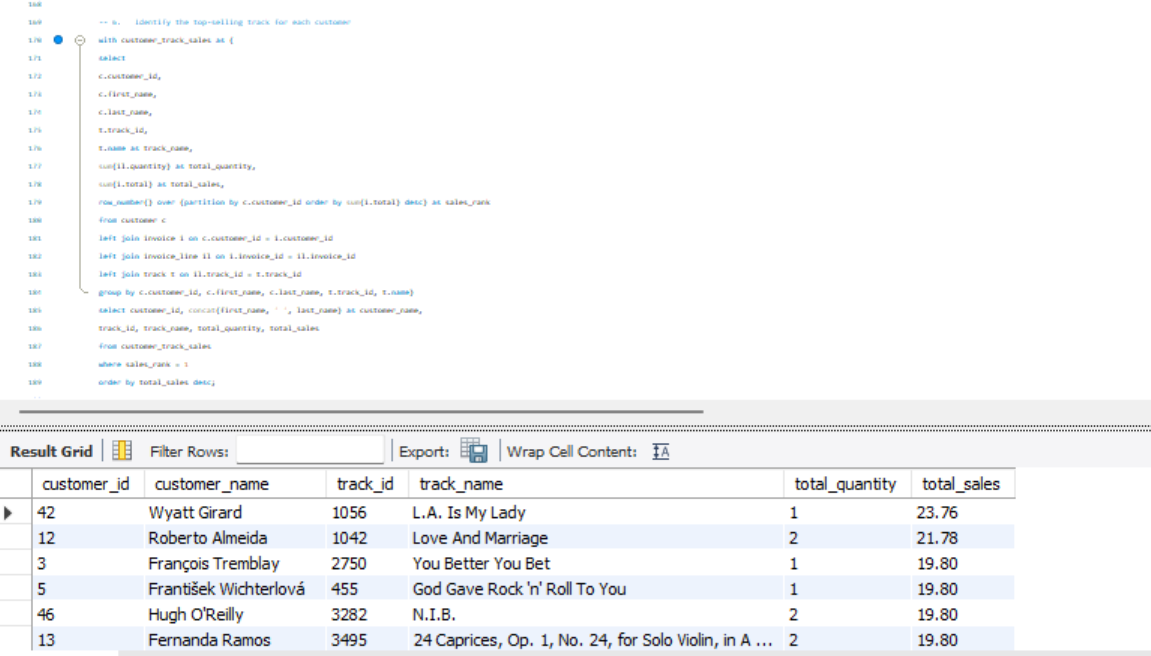
1. Identify the top-selling track for each customer

**Explanation:**

* CTE for Sales per Track: The customer\_track\_sales CTE calculates total quantity and sales per track for each customer. It also ranks tracks by total sales per customer using ROW\_NUMBER().
* Multiple Table Joins: The query joins customer, invoice, invoice\_line, and track tables.  
  This links each customer to the tracks they purchased and their corresponding sales.
* Ranking Top Tracks: ROW\_NUMBER() is used to assign a rank to each track purchased by a customer. Only the highest-selling track per customer is kept using sales\_rank = 1.
* Final Output: The outer query retrieves the top track (by revenue) per customer along with total quantity and sales. Results are sorted in descending order of total sales to highlight top-performing purchases.

**Insight**:

* Several customers had top purchases over $19, with *Wyatt Girard's* leading track generating **$23.76**.
* Many customers' favorite tracks had exactly **1 unit sold**, with high unit value driving the total sales.
* Tracks like “**Highway Chile**” and “**Earth Mofo**” appeared as favorites for **multiple customers**, showing repeated popularity.



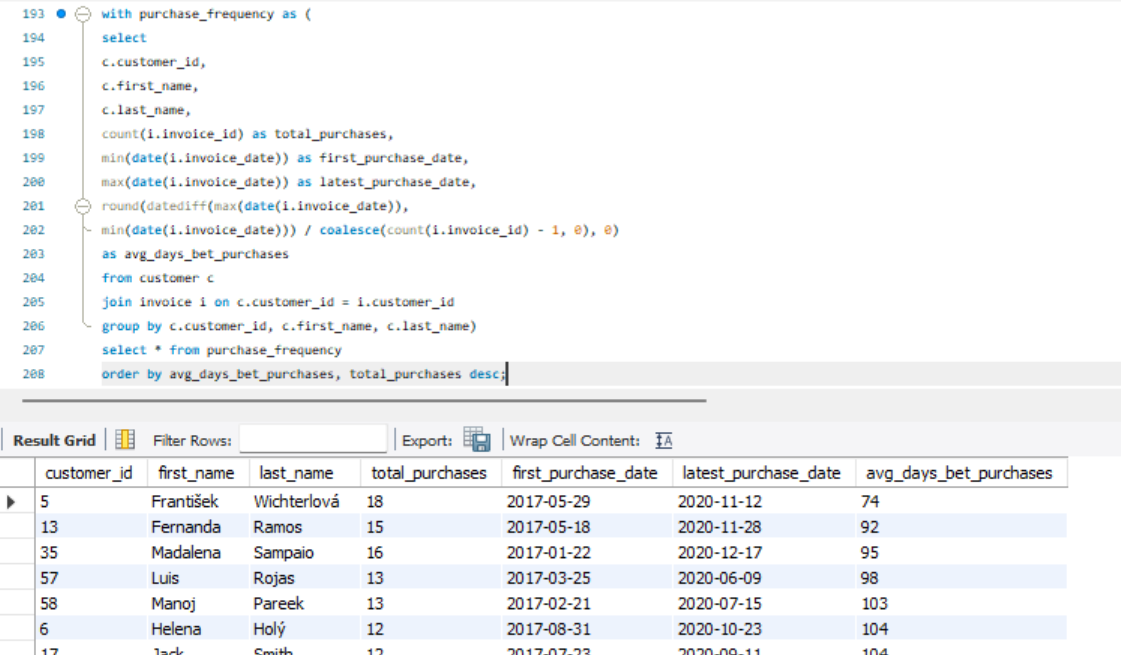
1. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases, preferred payment methods, average order value)?

**Explanation:**

* The subquery calculates total purchases, first and latest purchase dates, and average days between purchases for each customer.
* It joins the customer and invoice tables on customer\_id.
* Coalesce is used to avoid division by zero when calculating average days between purchases.
* The main query selects all results from the subquery.
* Results are ordered by average days between purchases (ascending) and total purchases (descending).

**Insight:**

* Customers with fewer purchases (e.g., 4–9 total) tend to have significantly higher average days between purchases, indicating less frequent buying behavior.
* Customers with more purchases (e.g., 13–18 total) have lower average days between purchases, showing more regular buying habits.
* The range of average days between purchases spans from 74 days to 392 days, highlighting a diverse set of customer buying patterns.



1. What is the customer churn rate?

with previous\_customer\_purchases as

(

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

date(i.invoice\_date) as invoice\_date,

lead(date(i.invoice\_date)) over(partition by c.customer\_id order by invoice\_date desc) as prev\_purchase

from customer c

join invoice i on c.customer\_id = i.customer\_id

),

prev\_purchase\_rank as

(

select \*,

row\_number() over(partition by customer\_id order by prev\_purchase desc) as prev\_purchase\_rn

from previous\_customer\_purchases

),

previous\_purchase\_date as

( select \*,

datediff(invoice\_date, prev\_purchase) as days\_since\_last\_purchase

from prev\_purchase\_rank

where prev\_purchase\_rn = 1

and datediff(invoice\_date, prev\_purchase) > 180

order by days\_since\_last\_purchase desc )

select

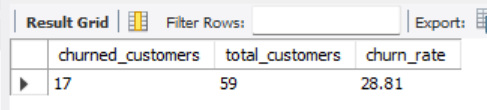
count(pp.customer\_id) as churned\_customers,

count(c.customer\_id) as total\_customers,

round((count(pp.customer\_id) \* 100) / count(c.customer\_id), 2) as churn\_rate

from customer c

left join previous\_purchase\_date pp on c.customer\_id = pp.customer\_id;



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

For this calculation, a customer is considered churned if they have not made any purchase for more than 6 month (Assuming to 180 days) between their last and second-last purchase dates.

**Explanation:**

* previous\_customer\_purchases: This part collects each customer's purchase history, including their last purchase date (invoice\_date) and the date of their previous purchase (prev\_purchase). It uses the LEAD function to get the previous purchase date for each customer, ordered by the most recent purchase.
* prev\_purchase\_rank: This part assigns a rank (prev\_purchase\_rn) to each customer's most recent purchase based on their prev\_purchase date, so we can identify the most recent pair of purchases for each customer.
* previous\_purchase\_date: This step calculates the number of days since each customer's last purchase (days\_since\_last\_purchase). It filters out customers who have not made a purchase in over 180 days between their last two purchases, marking them as "churned." These customers are ordered by the number of days since their last purchase.
* Final Calculation: The final query calculates the churn rate by counting how many customers are considered churned (those who haven't made a purchase in more than 180 days) and dividing that by the total number of customers. The result is multiplied by 100 to get the churn rate as a percentage.

**Insight:**

* Churn Rate: The churn rate of 28.81% indicates that approximately 29% of your customer base has stopped using the product.
* Customer Retention: Since about 71% of your customers are still active, the retention is relatively good, but there's room for improvement.

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

**Explanation:**

* Genre and Artist Sales Calculation: The query calculates the total sales for each artist within their genre in the USA, using data from multiple tables related to genres, tracks, invoices, and artists.
* Ranking Artists: It ranks artists within each genre based on their total sales, with the highest-selling artist ranked first.
* Total Sales in the USA: It calculates the overall total sales in the USA, considering all genres and artists.
* Final Results: The query shows each genre, artist, their total sales, their rank within the genre, and the percentage of total USA sales that each artist’s sales represent. Results are sorted by sales and genre.

**Insight:**

* **Rock** is the dominant genre in the USA, **5.05%** of percentage sales, with **Van Halen** as the top-selling artist.

with sales\_genre\_rank\_usa as

( select

g.name as genre,

ar.name as artist,

sum(i.total) as genre\_sales,

dense\_rank() over(partition by g.name order by sum(il.unit\_price \* il.quantity) desc) as genre\_rank

from genre g

left join track t on g.genre\_id = t.genre\_id

left join invoice\_line il on t.track\_id = il.track\_id

left join invoice i on il.invoice\_id = i.invoice\_id

left join album a on t.album\_id = a.album\_id

left join artist ar on a.artist\_id = ar.artist\_id

where i.billing\_country = 'USA'

group by g.name, ar.name

),

total\_sales\_usa as

(select

sum(i.total) as total\_sales

from invoice\_line il

left join invoice i on il.invoice\_id = i.invoice\_id

where i.billing\_country = 'USA'

)

select

s.genre,

s.artist,

s.genre\_sales,

t.total\_sales,

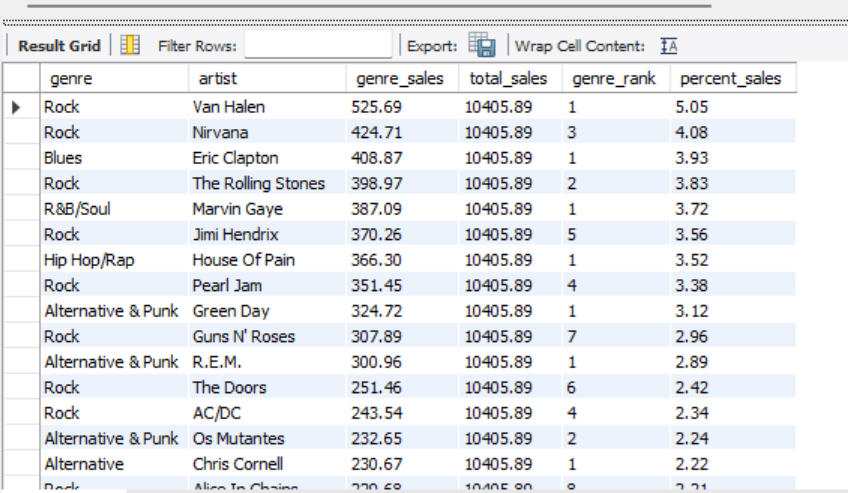
s.genre\_rank,

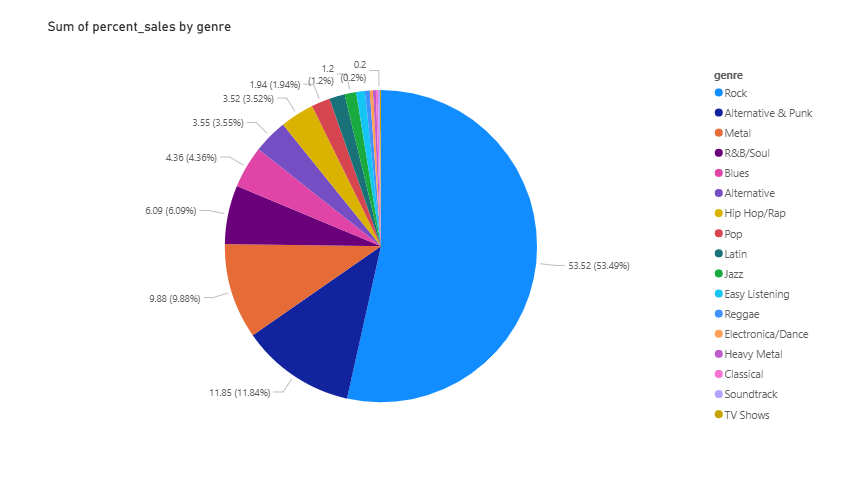
round((s.genre\_sales / t.total\_sales) \* 100, 2) as percent\_sales

from sales\_genre\_rank\_usa s

join total\_sales\_usa t

order by s.genre\_sales desc, s.genre asc;





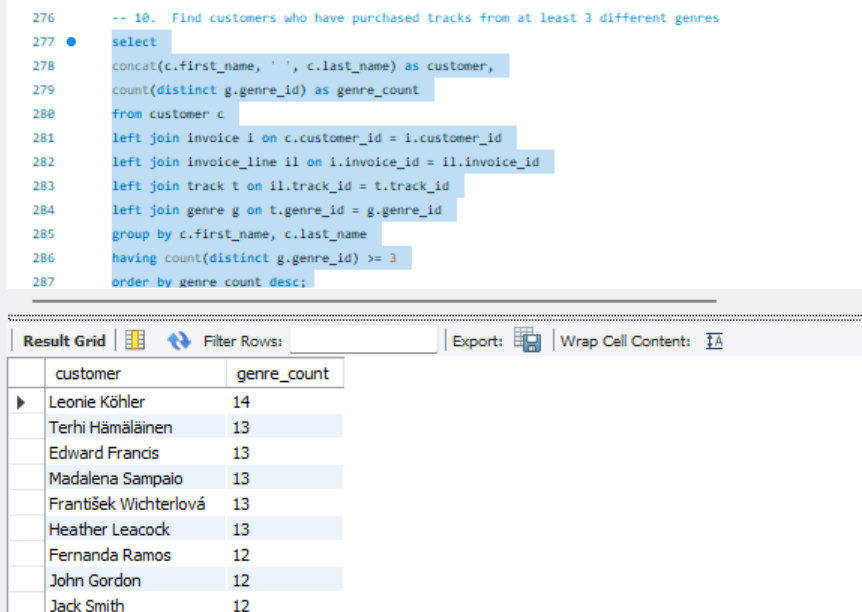
1. Find customers who have purchased tracks from at least 3 different genres

**Explanation:**

* **Joins customer purchases with music data**: It combines customer information with their invoice, tracks they bought, and the genres of those tracks.
* **Counts unique genres per customer**: For each customer, it counts how many different genres they have purchased music from.
* **Filters for diverse listeners**: It only shows customers who have bought music from **3 or more different genres**.
* **Order by diversity**: The results are sorted to show customers with the **most diverse music tastes** (i.e., highest genre count) at the top.

**Insight:**

* Leonie Köhler is the most musically diverse customer, purchasing from 14 different genres, followed closely by several others with 13 genres.
* A large group of customers (over 40) explored music from 8 or more genres, showing a wide variety of musical tastes among top listeners.



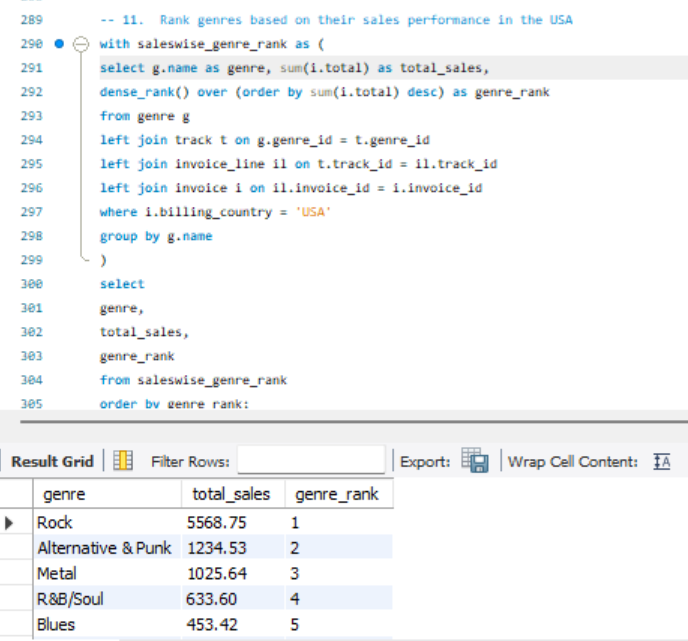
1. Rank genres based on their sales performance in the USA

**Explanation:**

* Analyzes music sales by genre: It looks at total music sales in the USA, grouped by each music genre.
* Ranks genres by sales: Genres are ranked using dense\_rank(), so genres with the highest sales get the top rank (1st, 2nd, etc.).
* Uses multiple joins: It connects genre info with tracks, invoice lines, and invoices to calculate the total sales per genre.
* Shows ranked results: The final output lists genres with their total USA sales and sales rank, ordered by rank from highest to lowest.

**Insight:**

* Rock dominates music sales in the USA with a total of $5568.75, significantly outperforming all other genres.
* Genres like Alternative & Punk and Metal follow, while TV Shows and Soundtrack genres have the lowest sales, indicating limited popularity.



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

**Explanation:**

* Identifies the most recent purchase: It first creates a list of customers with their last purchase date by finding the most recent invoice for each customer.
* Tracks all customer purchases: It then creates another list showing each customer’s individual purchases along with the invoice date for each.
* Looks for customers who haven't bought recently: The query identifies customers who haven’t made a purchase in the last 3 months (from their last purchase date).
* Returns those customers: It lists customers who have not made a purchase during that 3-month period, ordered by their customer ID.

with customer\_last\_purchase as

(

select

c.customer\_id,

c.first\_name,

c.last\_name,

max(date(i.invoice\_date)) as last\_purchase\_date

from customer c

join invoice i on c.customer\_id = i.customer\_id

group by c.customer\_id, c.first\_name, c.last\_name),

customer\_purchases as

(

select

c.customer\_id,

c.first\_name,

c.last\_name,

date(i.invoice\_date) as invoice\_date

from customer c

join invoice i on c.customer\_id = i.customer\_id

)

select clp.customer\_id,

clp.first\_name,

clp.last\_name,

clp.last\_purchase\_date

from customer\_last\_purchase clp

left join customer\_purchases cp on clp.customer\_id = cp.customer\_id

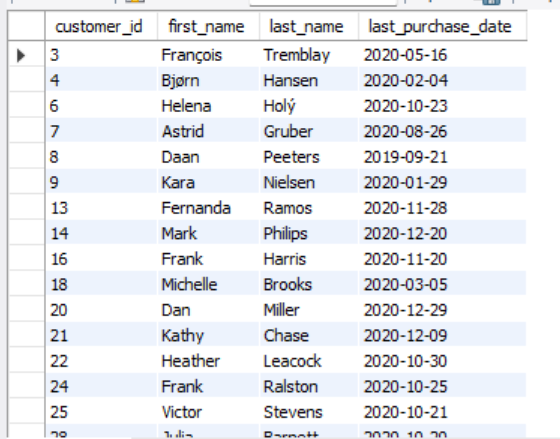
and cp.invoice\_date

between clp.last\_purchase\_date - interval 3 month

and clp.last\_purchase\_date - interval 1 day

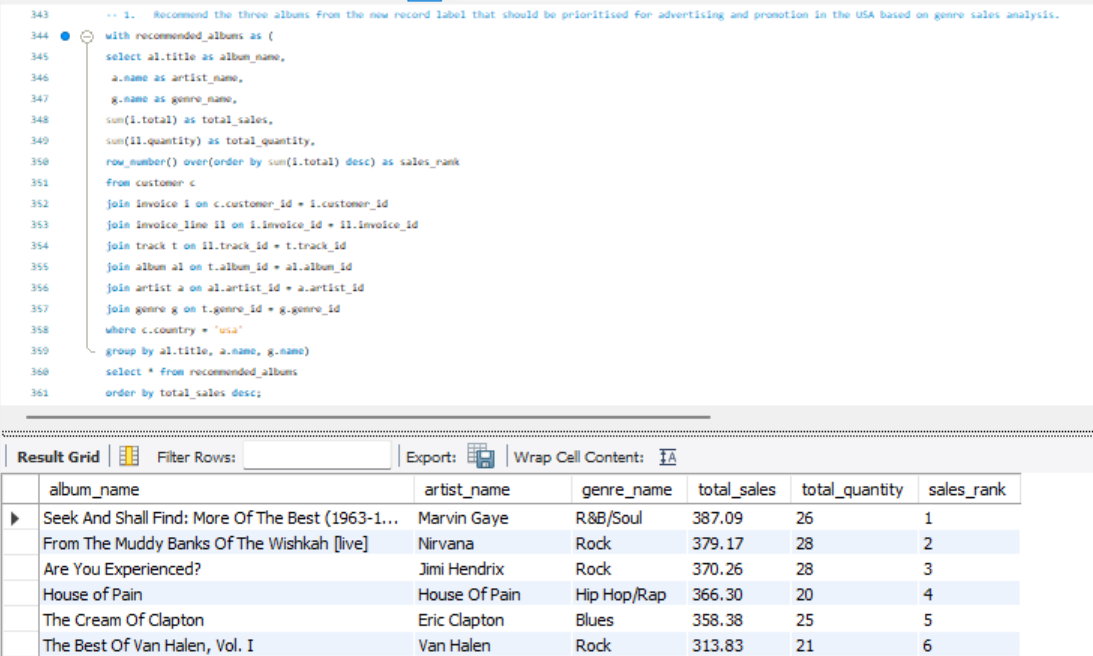
where cp.invoice\_date is null

order by clp.customer\_id;



Subjective Questions

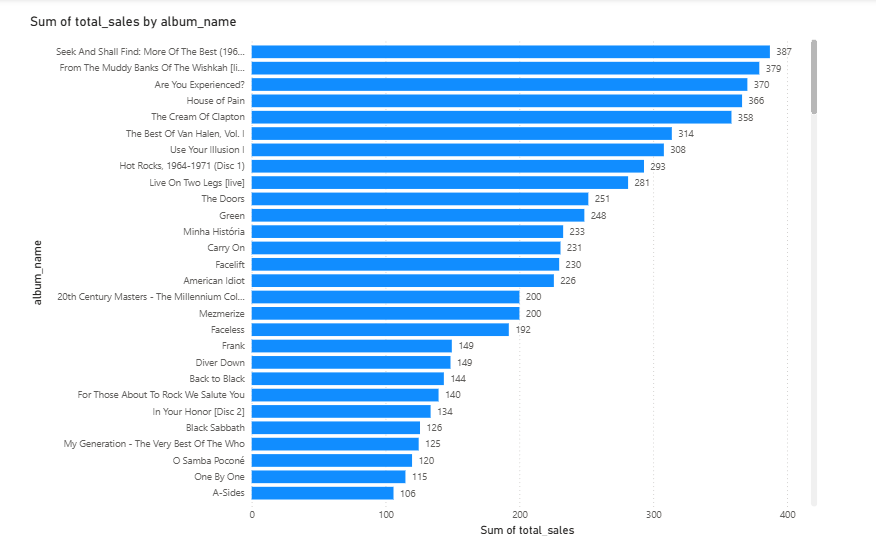
1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.



Top 3 Recommended Albums for Promotion in the USA (Based on Genre Sales Analysis):

Based on the bar chart analysis of genre-wise sales, the following three albums are the most suitable for advertising and promotional focus in the USA:

1. **Seek And Shall Find: More Of The Best (1963–1981)**
   * Genre: R&B/Soul
   * Sales Value: 387.09 (Highest)
   * Justification: This album tops the sales chart, highlighting a strong market demand for R&B/Soul in the USA.
2. **From The Muddy Banks Of The Wishkah [Live]**
   * Genre: Rock
   * Sales Value: 379.17
   * Justification: With Rock being a leading genre, this live album is likely to attract a loyal fanbase and drive promotional success.
3. **Are You Experienced?**
   * Genre: Rock
   * Sales Value: 370.26
   * Justification: A classic rock album with proven sales potential, making it an ideal candidate for focused marketing efforts.
4. Rock leads in overall sales, making it a reliable and low-risk genre for promotional investment.
5. The R&B/Soul album shows the highest individual sales, reflecting strong audience demand and market potential.
6. Focusing promotions on these albums in the USA is likely to boost engagement, given the established listener base for both genres.



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

**Top Genres Outside the USA:**

* Rock is the BIG winner: It sells way more than any other type of music.
* Metal and Alternative/Punk are also popular: They come in second and third place.

**Top Genres in the USA:**

* Rock is still number one: Just like in other countries, it sells the most.
* Alternative/Punk is second: More people buy this than Metal in the USA.
* Metal is third: It's still liked, but not as much as Alternative/Punk in the USA.

**What's the same?**

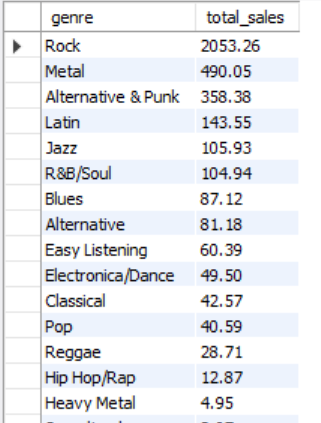
* Rock is king everywhere: It's the top choice in both the USA and other countries.
* Metal and Alternative/Punk are liked in both places: They are in the top three for both.

**What's different?**

* Rock is MUCH bigger outside the USA: People in other countries buy a lot more Rock music compared to people in the USA, based on this data.
* Metal vs. Alternative/Punk: Outside the USA, more people buy Metal than Alternative/Punk. But in the USA, it's the other way around.
* Overall sales are higher outside the USA for the top genres: It looks like people in other countries buy more of these top kinds of music in general.

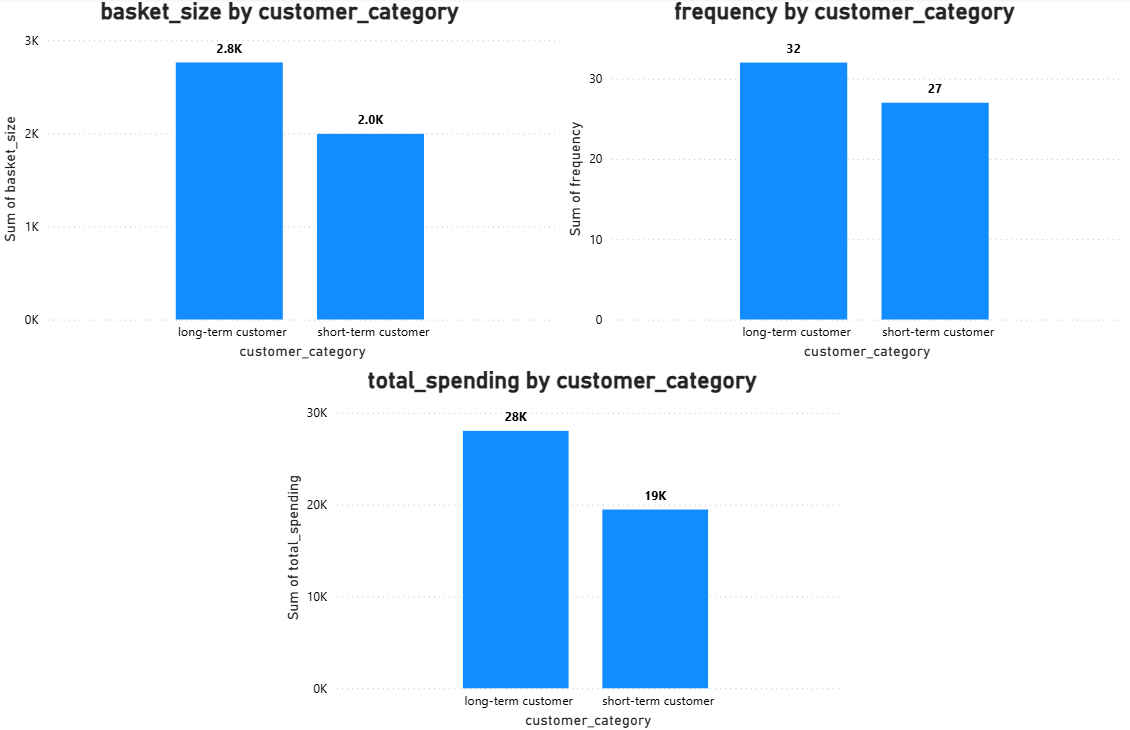
In simple terms: Rock music is loved worldwide, but it's especially popular outside the USA in this snapshot. While Metal and Alternative/Punk are liked in both places, their popularity compared to each other is a bit different between the USA and other countries.







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

****

**with cte as**

**(select i.customer\_id,**

**max(invoice\_date) as max\_invoice\_date,**

**min(invoice\_date) as min\_invoice\_date,**

**abs(timestampdiff(month,min(invoice\_date),**

**max(invoice\_date))) as time\_for\_each\_customer,**

**sum(total) as total\_sales,**

**sum(quantity) as total\_items,**

**count(invoice\_date) as invoice\_frequency**

**from invoice i**

**left join customer c on i.customer\_id = c.customer\_id**

**left join invoice\_line il on i.invoice\_id = il.invoice\_id**

**group by i.customer\_id**

**order by time\_for\_each\_customer desc**

**),**

**average\_time as**

**(**

**select avg(time\_for\_each\_customer) as average\_customer\_lifetime from cte**

**),**

**categorization as**

**(**

**select**

**cte.customer\_id,**

**cte.max\_invoice\_date,**

**cte.min\_invoice\_date,**

**cte.time\_for\_each\_customer,**

**cte.total\_sales,**

**cte.total\_items,**

**cte.invoice\_frequency,**

**case**

**when cte.time\_for\_each\_customer > (select average\_customer\_lifetime from average\_time)**

**then "long-term customer"**

**else "short-term customer"end as customer\_category**

**from cte**

**)**

**select**

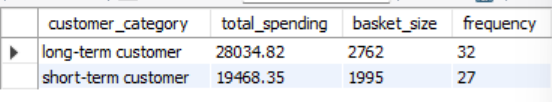
**customer\_category,**

**sum(total\_sales) as total\_spending,**

**sum(total\_items) as basket\_size, count(invoice\_frequency) as frequency**

**from categorization**

**group by customer\_category;**

****

Purchasing Habits of Different Customer Groups:

**Frequency (How often they buy):**

* Long-term customers: Buy more often (32 times on average).
* Short-term customers: Buy less often (27 times on average).

**Basket Size (How much they buy in one go):**

* Long-term customers: Buy a larger amount per purchase (average basket size of 2.8K).
* Short-term customers: Buy a smaller amount per purchase (average basket size of 2.0K).

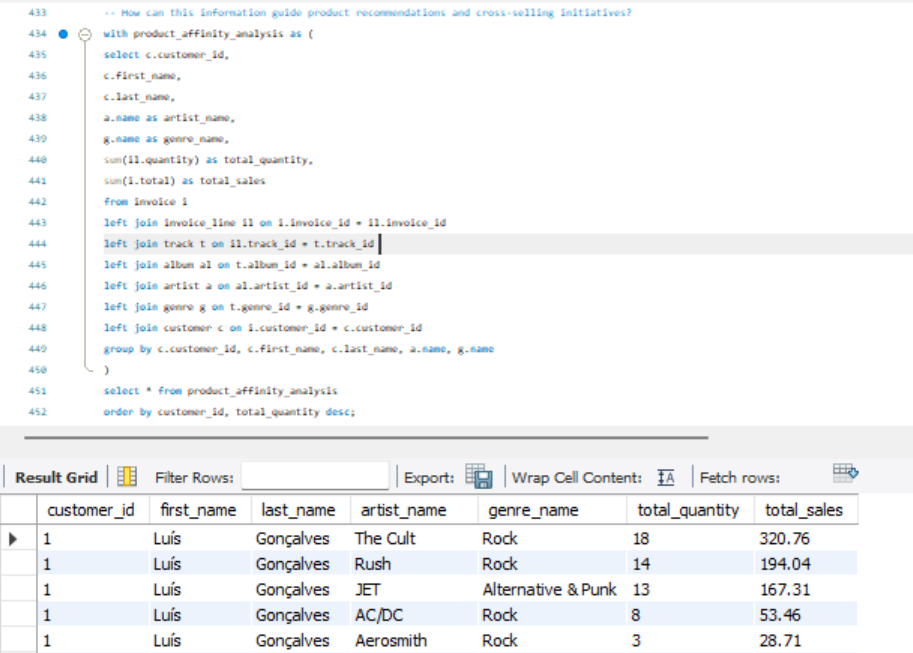
**Spending Amount (Total money spent):**

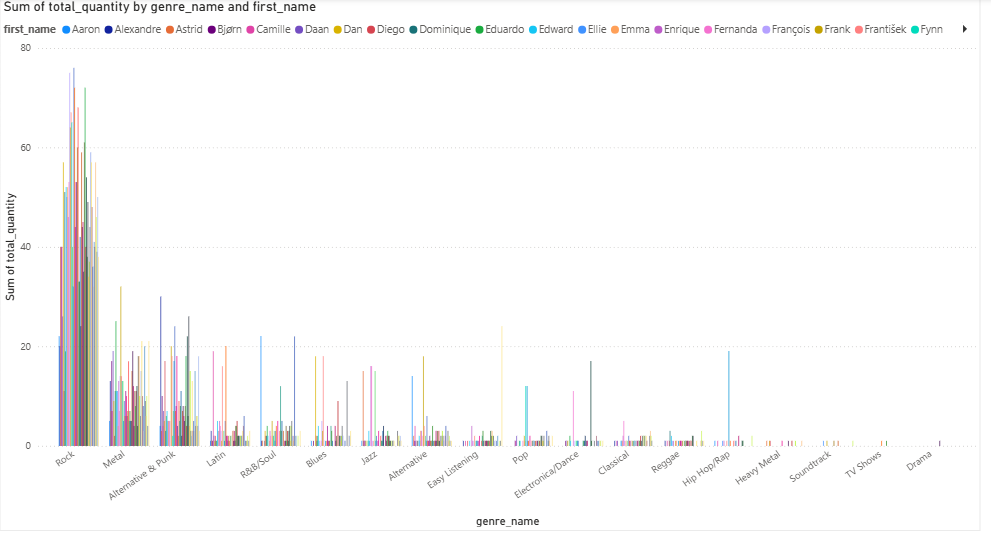
* Long-term customers: Spend significantly more in total (28K).
* Short-term customers: Spend less in total (19K).

**These patterns show us:**

* Loyal customers are more valuable: They buy more often and spend more per purchase, leading to much higher total spending.
* Keeping customers is profitable: Longer relationships mean more frequent purchases and bigger spending over time.
* Stronger relationships lead to bigger purchases: Long-term customers might trust the brand more and buy more at once.
* New customers are an opportunity: While they spend less now, engaging them could turn them into valuable long-term buyers.

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?





**Genre-Based Preferences:**

* Rock, Metal, and Alternative & Punk are often bought together, so fans of one genre may enjoy the others too.
* Latin and Blues show up as secondary choices and can be grouped into themed playlists or bundles.

**Cross-Selling Ideas:**

* Bundles & Discounts: Create combo offers like Rock + Metal album packs or give discounts for buying within related genres.
* Recommendations: Show suggestions like People who bought Rock albums also bought Punk and Alternative albums.

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

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

The above visuals show how customer purchasing and churn vary across regions, potentially linked to local demographics and economies.

**Key Observations:**

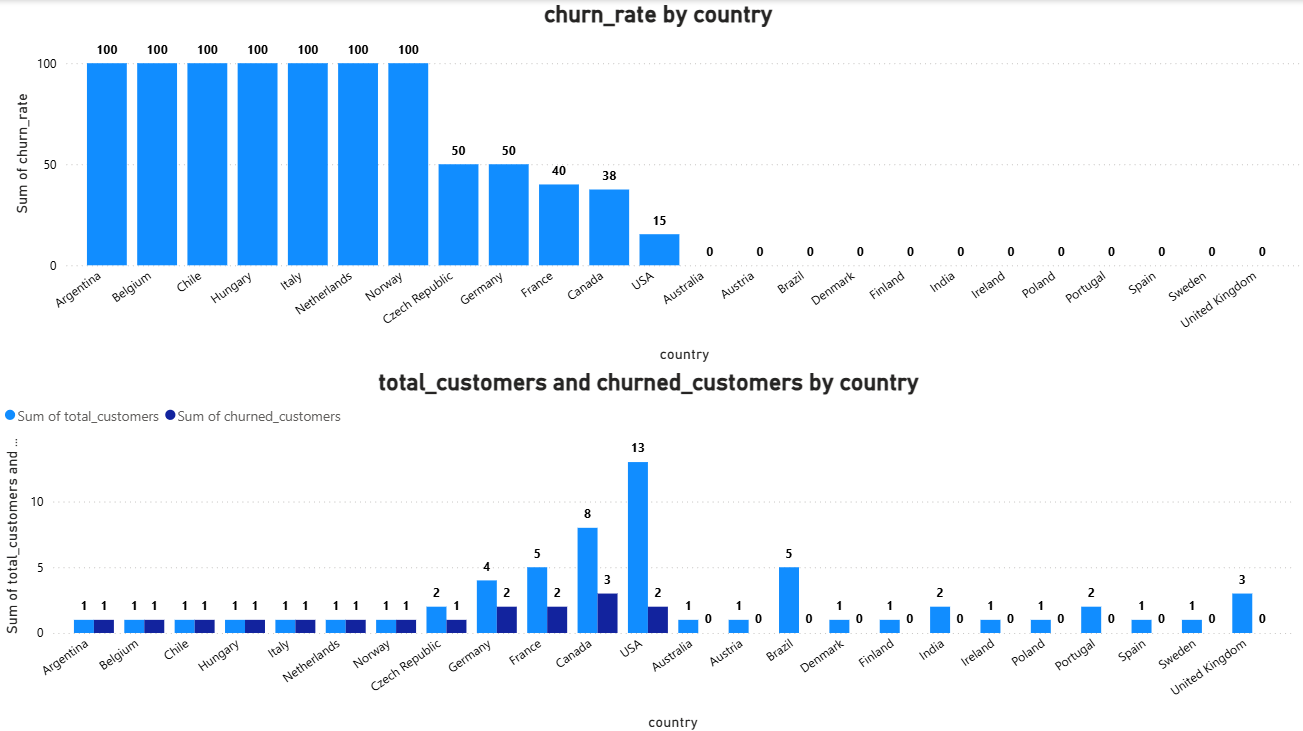
* High Churn: Several countries (Argentina, Belgium, Chile, Hungary, Italy, Netherlands, Norway) show 100% churn. Germany, France, and Czech Republic have churn above 50%. The USA has lower churn (15.38%) despite the largest customer base (13). Brazil's 37.5% churn indicates partial retention.
* Customer Distribution: The USA has the most customers (13), followed by Brazil (8). Other countries have 1-5 customers.

**Potential Insights:**

* Economic Factors: Weaker economies may correlate with higher churn (e.g., Argentina, Hungary). Wealthier nations (e.g., USA, UK) show more stable customers.
* Market Penetration: Higher customer numbers and lower churn (USA, Brazil) suggest strong market presence. Low customer numbers and 100% churn may indicate poor adoption or awareness.
* Business Model Influence: High churn in subscription services could stem from localized content or pricing issues. For physical stores, regional competition and availability might be factors.

**Strategic Recommendations:**

* Retention in High-Churn Areas: Implement localized promotions and improve engagement through targeted marketing.
* Expansion in Strong Regions: Invest in marketing and expansion in the USA and Brazil. Offer localized services and content in promising regions.
* Product Preference Analysis: Analyze regional product popularity and correlate churn with purchasing behavior to optimize offerings.

****

**with previouscustomerpurchases as**

**(**

**select c.country,**

**c.customer\_id,**

**c.first\_name,**

**c.last\_name,**

**date(i.invoice\_date) as invoice\_date,**

**lead(date(i.invoice\_date)) over(partition by c.customer\_id order by invoice\_date desc) as prev\_purchase**

**from customer c**

**join invoice i on c.customer\_id = i.customer\_id**

**),**

**prevpurchaserank as**

**(select \*,**

**row\_number() over(partition by customer\_id order by prev\_purchase desc) as prev\_purchase\_rn**

**from previouscustomerpurchases**

**),**

**previouspurchasedate as**

**(**

**select \*,**

**datediff(invoice\_date,prev\_purchase) as days\_since\_last\_purchase**

**from prevpurchaserank**

**where prev\_purchase\_rn = 1**

**and datediff(invoice\_date,prev\_purchase) > 180**

**order by days\_since\_last\_purchase desc**

**)**

**select c.country,**

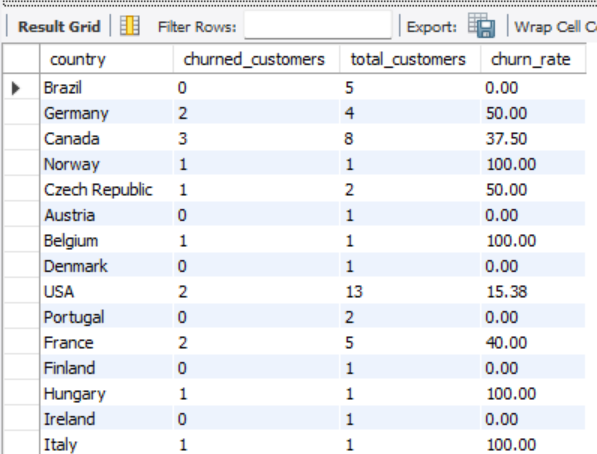
**count(pp.customer\_id) as churned\_customers,**

**count(c.customer\_id) as total\_customers,**

**round((count(pp.customer\_id) \* 100) / count(c.customer\_id), 2) as churn\_rate**

**from customer c left join previouspurchasedate pp on c.customer\_id = pp.customer\_id**

**group by c.country;**

****

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?

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

* High-Risk Regions (100% Churn): Italy, Norway, Chile, Netherlands, Belgium, Hungary, Argentina. (50% Churn): Czech Republic, Germany. These areas suggest high dissatisfaction, low engagement, or economic issues.
* Low-Risk Regions (0% Churn): Finland, Australia, Sweden, India, Brazil, UK, Poland, Denmark, Portugal, Austria, Spain. These indicate strong retention.
* Moderate-Risk Regions: USA (15.38%) and Canada (37.5%) show manageable churn.

**Factors Contributing to Churn Risk:**

* Demographics: Younger customers show lower loyalty. Older customers may churn due to digital engagement issues. Gender influences spending and product preferences.
* Economic & Regional Factors: High-churn regions may face economic instability. Wealthier regions correlate with low churn and higher satisfaction.
* Purchase History: Infrequent or low-value purchases increase churn risk. Subscription cancellations or long purchase gaps are indicators.
* Engagement & Customer Experience: Poor service, lack of personalization, and irrelevant recommendations drive churn.

Note: The customer table is missing age and gender columns, limiting the ability to analyze the demographic breakdown of customers in detail.



with previouscustomerpurchases as

(

select c.country,

c.customer\_id,

c.first\_name,

c.last\_name,

date(i.invoice\_date) as invoice\_date,

lead(date(i.invoice\_date)) over(partition by c.customer\_id order by invoice\_date desc) as prev\_purchase

from customer c

join invoice i on c.customer\_id = i.customer\_id

),

prevpurchaserank as

(

select \*,

row\_number() over(partition by customer\_id order by prev\_purchase desc) as prev\_purchase\_rn

from previouscustomerpurchases

),

previouspurchasedate as (

select \*,

datediff(invoice\_date,prev\_purchase) as days\_since\_last\_purchase

from prevpurchaserank

where prev\_purchase\_rn = 1

and datediff(invoice\_date,prev\_purchase) > 180

order by days\_since\_last\_purchase desc

)

select c.country,

count(pp.customer\_id) as churned\_customers,

count(c.customer\_id) as total\_customers,

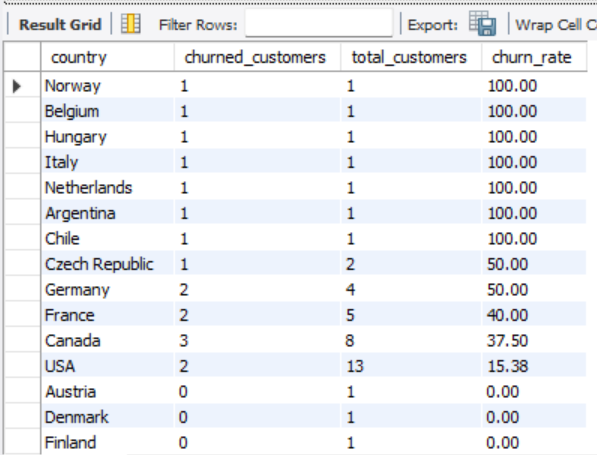
round((count(pp.customer\_id) \* 100) / count(c.customer\_id), 2) as churn\_rate

from customer c

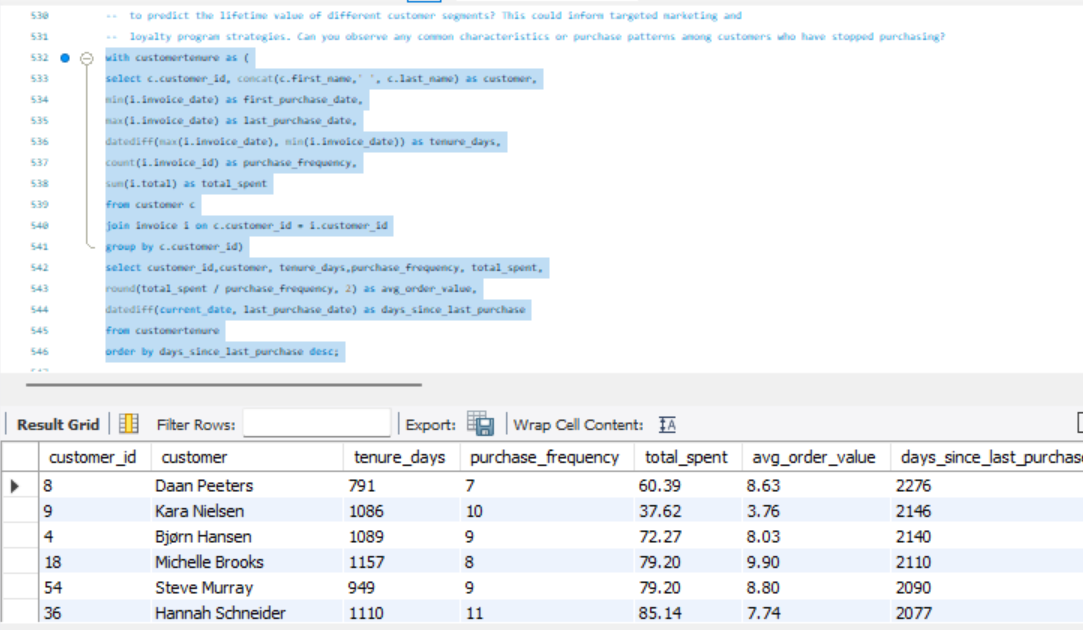
left join previouspurchasedate pp on c.customer\_id = pp.customer\_id

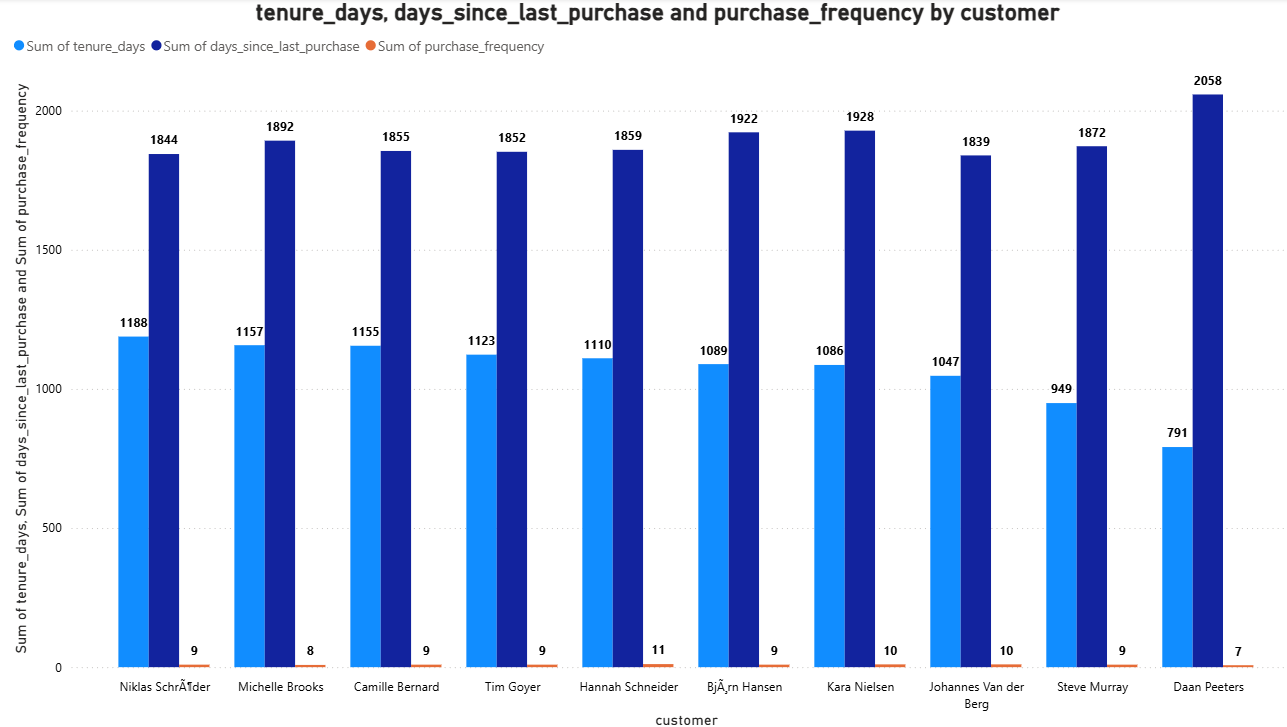
group by c.country

order by churn\_rate desc, total\_customers asc;



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?





* **High Tenure with Varied Recency:** Several customers, such as Niklas Schräder, Michelle Brooks, and Camille Bernard, have a long tenure (dark blue bars around 1100-1200 days) but also a relatively long time since their last purchase (light blue bars around 1800-1900 days). This suggests they were once active but might be at risk of churning.
* **Recent Purchases but Shorter Tenure:** In contrast, customers like Daan Peeters show a much shorter tenure (around 791 days) but a more recent last purchase (2058 days is likely an anomaly or a misinterpretation of the data, as it's longer than the tenure). This could indicate a newly acquired customer who made a purchase recently.
* **Low Purchase Frequency Across the Board:** The purchase frequency (red dots) is consistently low for all displayed customers, generally staying below 11. This indicates that most customers make very few purchases over their entire tenure with the business.
* **Potential Inactive Long-Term Customers:** Customers like Tim Goyer, Hannah Schneider, and Bjørn Hansen exhibit a long tenure (dark blue around 1100 days) and a significant time since their last purchase (light blue around 1850 days), coupled with a low purchase frequency (around 9-11). They represent a segment of long-term customers who haven't engaged recently.
* **Outlier in Last Purchase Data:** The "days since last purchase" value for Daan Peeters (2058 days) is notably higher than his tenure (791 days) and all other customers. This is likely an error in the data or requires further investigation to understand the actual purchase recency for this customer.

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?

* **Compare Before, During, and After:** To see how well promotions worked, look at key numbers (like sales and customer counts) before, while the promotion was happening, and after it ended.
* **Track New Customers:** Count how many new customers joined during the campaign and figure out if the discounts, events, or emails brought them in.
* **Analyze Repeat Business:** See if customers who used the promotion came back to buy again. Also, check if fewer of those customers stopped buying altogether (churn).
* **Measure Sales Changes:** Look at the total money made, how much people spent on average, and how often they bought during the promotion compared to normal.
* **Understand Engagement:** Check how people interacted with the promotions, like opening emails or using discount codes. This helps understand what caught their attention.

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

If the objective and subjective questions weren't provided, I would begin by reviewing the project’s aim and identifying key performance indicators aligned with the objective. Based on this I would develop a relevant questions.

1. **Business Understanding First:**

* Begin by thoroughly understanding the organization's main goals and operational context.
* Identify potential business challenges and areas where data analysis could provide value.
* This foundational knowledge ensures the analysis is relevant and strategically aligned.

1. **Initial Data Exploration:**

* Conduct a broad overview of the dataset to understand its structure and contents.
* Identify data types, potential quality issues (missing values, inconsistencies), and initial trends.
* This step helps in getting familiar with the data landscape and potential avenues for deeper analysis.

1. **Define Key Metrics:**

* Propose relevant Key Performance Indicators (KPIs) that can measure progress towards business objectives.
* Examples include customer churn rate, sales growth, or website conversion rates.
* These metrics provide a quantifiable way to assess performance and the impact of potential recommendations.

1. **Formulate Key Questions:**

* Develop specific, answerable questions about the data that relate to the business goals and initial observations.
* Examples include "What are the primary drivers of customer churn?" or "Which marketing channels yield the highest ROI?".
* These questions guide the focused analysis and ensure it addresses pertinent business inquiries.

1. **Perform Focused Analysis:**

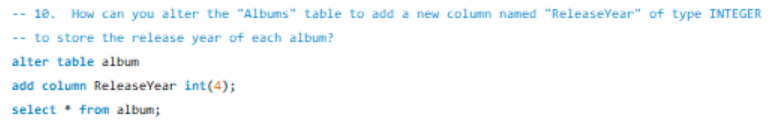
* Apply appropriate analytical techniques (e.g., segmentation, regression, time series analysis) to answer the formulated questions.
* Explore relationships between variables and identify statistically significant patterns or trends.
* This stage involves in-depth investigation to extract meaningful insights from the data.

1. **Synthesize Insights & Recommendations:**

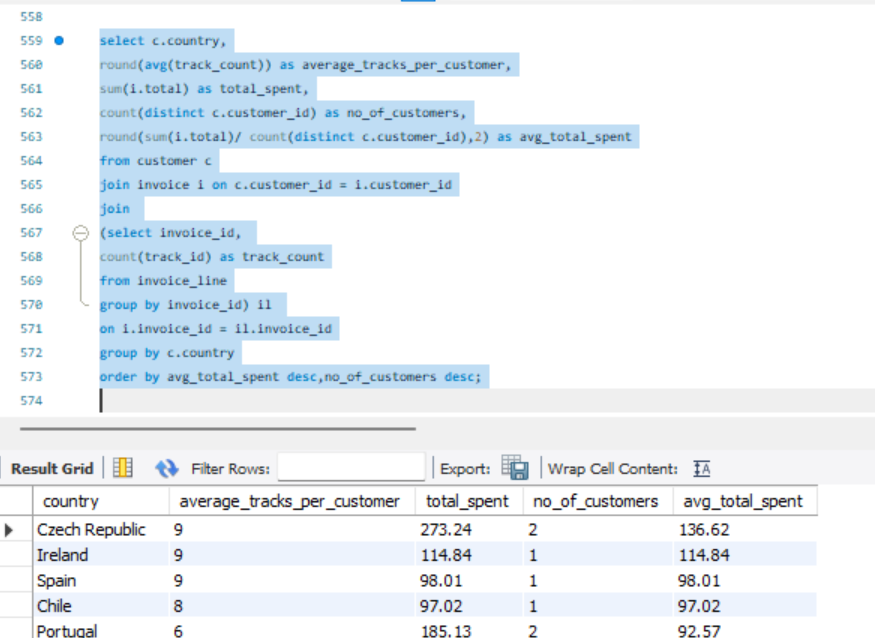
* Combine the findings from the analysis into clear, concise, and actionable insights.
* Translate these insights into practical recommendations that the business can implement.
* Communicate these findings effectively with supporting data and visualizations to facilitate understanding and adoption.

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?

We can make use the ALTER TABLE and ADD COLUMN command to the Album table



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.



**Explanation:**

* **Aggregates by Country:** It groups all the data based on the country of the customer.
* **Calculates Averages and Sums:** For each country, it computes the average number of tracks bought, the total money spent, and the total number of customers.
* **Finds Average Spending:** It also calculates the average total amount spent by each customer within each country.
* **Orders the Results:** The final output is sorted to show countries with the highest average spending per customer first, and then by the number of customers in descending order.

**Insight:**

* **Czech Republic leads in average spending:** Despite only having two customers, the Czech Republic exhibits the highest average total spent per customer (136.62), indicating potentially high-value individuals. This suggests focusing on understanding and potentially nurturing these customers.
* **USA has the most customers with moderate spending:** The USA has the largest number of customers (13), but their average spending (80.04) is mid-range. This implies a broad customer base with consistent, but not exceptionally high, individual spending.
* **Single-customer countries show varied spending:** Numerous countries have only one customer, with their total and average spending being the same. This highlights the need for caution when analyzing trends in these regions due to the small sample size.
* **Track purchase average is fairly stable:** Across most countries, the average number of tracks purchased per customer remains relatively consistent (around 7-9). This suggests a general pattern in the quantity of music acquired per individual buyer.