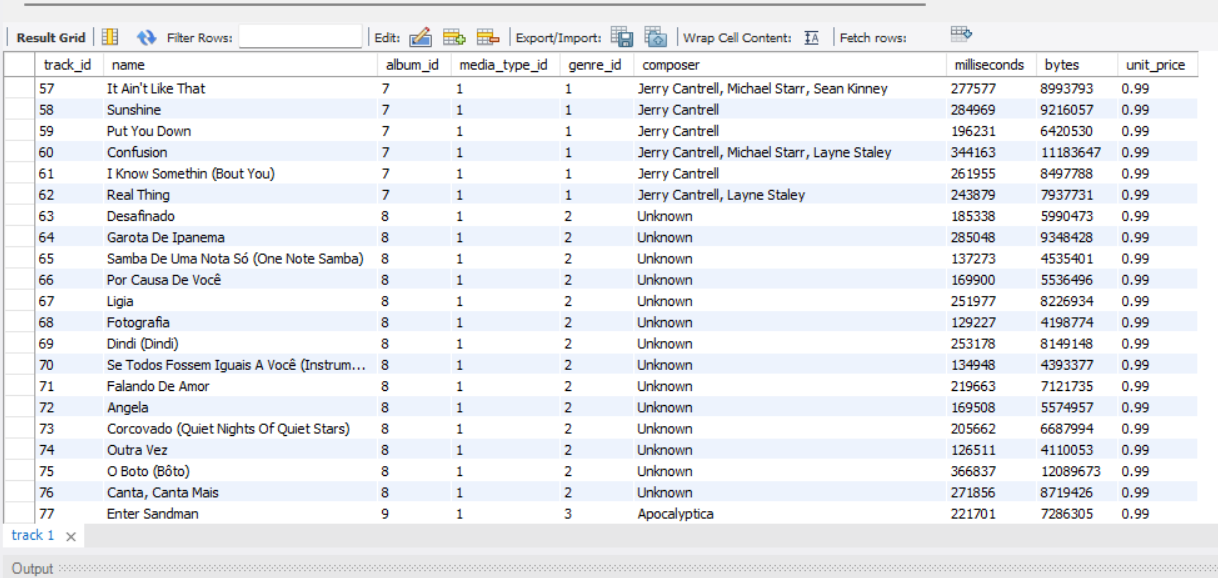
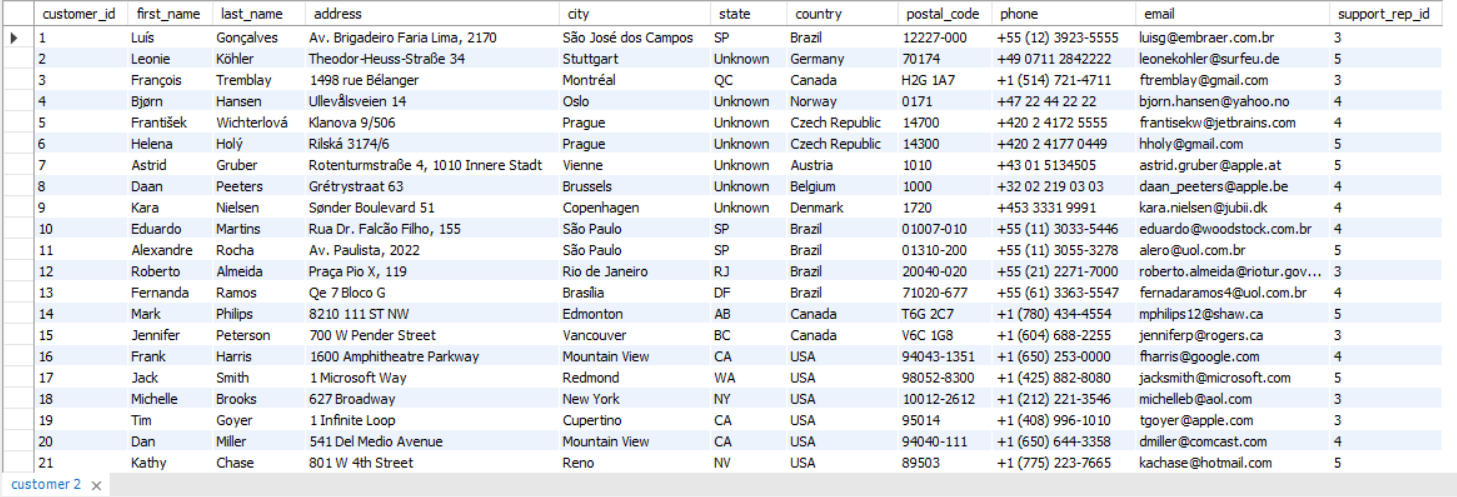
Objective Questions

1. Does any table have missing values or duplicates? If yes how would you handle it?  
     
   Answer:  
   **Code:**  
   update track set composer='Unknown' where composer is null;  
     
     
     
   alter table customer drop column company;

alter table customer drop column fax;

UPDATE customer   
SET   
State = COALESCE(State, 'Unknown'),   
Postal\_Code = COALESCE(Postal\_Code, '00000'),

Phone = COALESCE(Phone, 'Not Provided');  
  
  
  
 Explanation: Yes, two tables Customer and Track are having null values.  
  
a. Track has Composer column which contains null value so replaced it with the ‘Unknown’ value   
b. Customer table have 5 column with null values.

|  |  |
| --- | --- |
| **Column Name** | **Null Count** |
| company | 49 |
| Fax | 47 |
| State | 29 |
| Postal\_Code | 4 |
| Phone | 1 |

* I have dropped the Columns **Company** and **Fax** as it has maximum no of null values.
* For the remaining columns I have updated with below default values:  
   State --> Unknown

Postal\_Code --> 0

Phone --> Not Provided

1. Find the top-selling tracks and top artist in the USA and identify their most famous genres.  
   Answer:  
   **Code :**with top\_selling\_track as(

select track\_id, total\_price from

(select

track\_id, sum(unit\_price\*quantity) as total\_price,

dense\_rank() over (order by sum(unit\_price\*quantity) desc) as rnk

from

invoice i

join

invoice\_line il

on

i.invoice\_id = il.invoice\_id

where billing\_country='USA'

group by 1) a where rnk=1)

select

t.name as track\_name, g.name as genre\_name, ar.name as artist\_name, c.total\_price

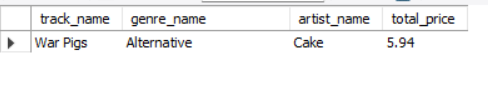
from

track t

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

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

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

join top\_selling\_track c on c.track\_id=t.track\_id;  
  
  
Answer: The top-selling track in the USA is "War Pigs", which falls under the Alternative genre. This track is performed by the artist Cake, and it has generated a total revenue of $5.94.

This insight highlights the popularity of the Alternative genre among US customers and suggests that Cake's music has a strong fan base in this region. Businesses can leverage this information to optimize playlist suggestions, and design genre-based promotions for better engagement.

1. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?  
   Answer:  
   **Code:**  
   SELECT Country, COUNT(\*) AS Customer\_Count

FROM Customer

GROUP BY Country

ORDER BY 2 DESC;   
  
  
  
The Chinook customer base is spread across multiple countries, with the highest concentration in the USA (13 customers), followed by Canada (8 customers) and Brazil & France (5 customers each). Other notable markets include Germany (4 customers) and the United Kingdom (3 customers). The remaining customers are distributed across multiple countries, including Czech Republic, Portugal, India, Norway, Austria, Belgium, and more, each with 1 or 2 customers.

This data suggests that Chinook's strongest markets are in North America (USA & Canada), with a presence in Europe, South America, and other regions. Understanding regional purchasing patterns and customer preferences could help in tailoring marketing strategies and expanding customer engagement in key markets.

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

i.billing\_country as country,

i.billing\_state as state,

i.billing\_city as city,

count(i.invoice\_id) as invoice\_count,

sum(unit\_price\*quantity) as revenue

from

invoice i

join

invoice\_line il

on

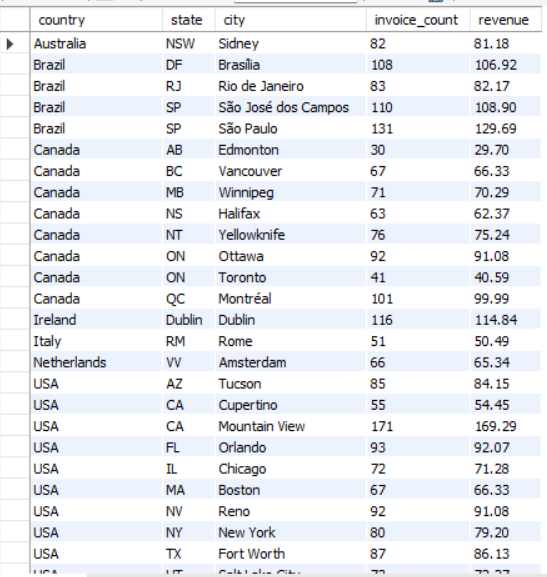
i.invoice\_id=il.invoice\_id

where billing\_state<>'None'

group by

country,state,city

order by

country,state,city;  
  
  
  
  
  
The total revenue and invoice count for each country, state, and city reveal key insights into Chinook's market performance:

* USA and Brazil are among the top-performing countries in terms of revenue and transaction volume.
* In the USA, Mountain View (CA) leads with 171 invoices totalling $169.29, followed by Redmond (WA) with 99 invoices ($98.01) and Fort Worth (TX) with 87 invoices ($86.13).
* Brazil's key cities include São Paulo (131 invoices, $129.69) and São José dos Campos (110 invoices, $108.90).
* Canada shows strong customer activity, with Montréal (101 invoices, $99.99) and Ottawa (92 invoices, $91.08) being the top revenue-generating cities.
* In Europe, Dublin (Ireland) leads with 116 invoices ($114.84), followed by Amsterdam (Netherlands) with 66 invoices ($65.34) and Rome (Italy) with 51 invoices ($50.49).
* Other cities with notable revenue contributions include Sydney (Australia), Edmonton (Canada), and New York (USA).

1. Find the top 5 customers by total revenue in each country.  
   Answer:  
   **Code:**with cte as(

select

c.country,c.customer\_id,

concat(c.first\_name,' ', c.last\_name) as customer\_name,

sum(il.unit\_price\*il.quantity) as revenue,

dense\_rank() over (partition by c.country order by sum(il.unit\_price\*il.quantity) desc) as rnk

from

customer c

join

invoice i

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

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

group by 1,2,3)

select

country,

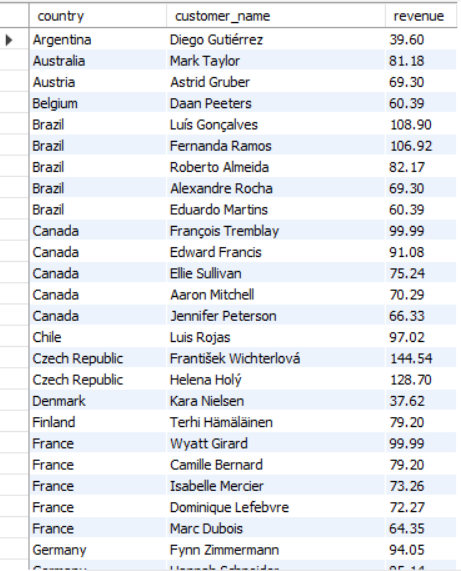
customer\_name,

revenue

from

cte

where rnk<6

order by country, rnk;  
  
  
  
This analysis identifies the highest-grossing customers in each country, showcasing their contribution to total revenue.  
Key Findings:

* Czech Republic has the highest revenue-generating customer, František Wichterlová with $144.54.
* India's top customer, Manoj Pareek, has $111.87, showing strong purchasing power in the region.
* Brazil has multiple high-revenue customers, with Luís Gonçalves ($108.90) and Fernanda Ramos ($106.92) leading.
* In Canada, François Tremblay ($99.99) and Edward Francis ($91.08) are the top revenue contributors.
* The USA's top customer, Jack Smith ($98.01), is closely followed by Dan Miller ($95.04) and Heather Leacock ($92.07).
* Countries with a single top customer include Denmark (Kara Nielsen, $37.62), Argentina (Diego Gutiérrez, $39.60), and Italy (Lucas Mancini, $50.49), indicating a smaller customer base.

1. Identify the top-selling track for each customer.  
   Answer:  
   **Code:**  
   with sales\_track as(

select

c.customer\_id,concat (c.first\_name,' ', c.last\_name) as customer\_name,

t.track\_id, t.name as track\_tname, count(il.track\_id) as purchase\_count,

RANK() OVER (PARTITION BY c.Customer\_ID ORDER BY COUNT(il.Track\_ID) DESC, t.track\_id ASC) AS rnk

from

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

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

join track t on il.track\_id=t.track\_id

group by 1,2,3,4)

select

customer\_id,

customer\_name,

track\_id,

track\_tname,

purchase\_count from sales\_track

where rnk=1

order by customer\_id ;

  
  
The top-selling track for each customer has been identified based on the highest sales count. Each customer has a unique top-selling track. Each customer’s top-selling track has been listed in the output as mentioned in the above screenshot.

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

customer\_id,

avg(unit\_price\*quantity) as avg\_order\_value,

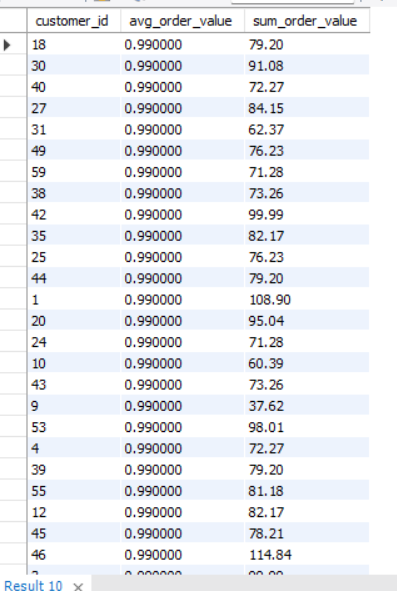
sum(unit\_price\*quantity) as sum\_order\_value

from

invoice i

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

group by 1

order by 2 desc;  


WITH PurchaseDifferences AS (  
SELECT   
customer\_id,

invoice\_id,

invoice\_date,

TIMESTAMPDIFF(DAY,

LAG(invoice\_date) OVER (PARTITION BY customer\_id ORDER BY invoice\_date),

invoice\_date) AS days\_between\_purchases

FROM

invoice

)

SELECT   
customer\_id,

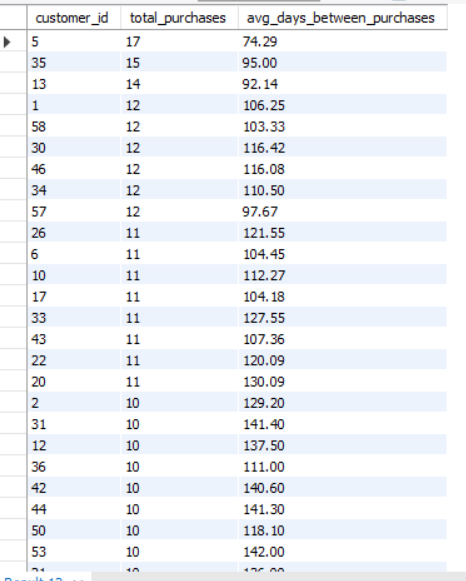
COUNT(invoice\_id) AS total\_purchases,

ROUND(AVG(days\_between\_purchases), 2) AS avg\_days\_between\_purchases

FROM PurchaseDifferences

WHERE days\_between\_purchases IS NOT NULL

GROUP BY customer\_id

ORDER BY total\_purchases DESC;  
  
  
  
  
  
  
Answer: Based on the data, the following patterns and trends can be observed in customer purchasing behaviour:

1. Frequency of Purchases:
   * Some customers, such as customer ID 5 (17 purchases) and 35 (15 purchases), make frequent purchases, while others, such as customer ID 29 (3 purchases), have a much lower purchase frequency.
   * The average days between purchases vary widely, with some customers making purchases every 74–130 days (e.g., customers 5, 35, and 13), while others wait over 190–390 days (e.g., customers 52, 59, and 29).
2. Average Order Value:
   * Most customers have an average order value of $0.99, suggesting that they generally purchase individual items rather than bundles or higher-priced products.
   * The total order value varies significantly, with some customers like customer 5 ($144.54) and customer 6 ($128.70) spending more than others like customer 14 ($29.70).
3. Spending Patterns:
   * Higher total purchase values correlate with more frequent purchases (e.g., customers 5, 6, and 46).
   * Customers with fewer purchases tend to have higher gaps between purchases, indicating they might be occasional buyers rather than regular ones.
4. Possible Insights:
   * Customers who purchase frequently (shorter average days between purchases) may be loyal customers who continuously buy new music.
   * Customers with a high sum\_order\_value but a low purchase frequency might be making bulk purchases in a single order.
   * There could be an opportunity to introduce bundle deals or subscription models for high-frequency buyers to increase overall revenue.

Overall, the purchasing behaviour suggests two key customer segments:

1. Frequent buyers with smaller, regular purchases.
2. Occasional buyers making larger purchases but with long intervals between orders.

These insights could be used to tailor marketing strategies, such as offering discounts to high-frequency buyers or targeting occasional buyers with promotions to encourage more frequent purchases.

1. What is the customer churn rate?  
   Answer:   
   **Code:**  
   with cte as (

select year(invoice\_date) as Year , count(DISTINCT customer\_id) as CustomersAtStart

from invoice

group by 1)

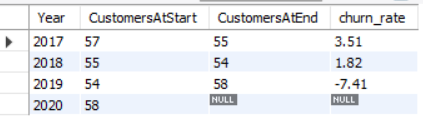
select

Year,

CustomersAtStart,

lead(CustomersAtStart) over (order by Year) as CustomersAtEnd,

round(((CustomersAtStart- lead(CustomersAtStart) over (order by Year))/CustomersAtStart)\*100,2) as churn\_rate

from cte;  
  
 

### **Key Observations:**

1. **Decreasing Churn Rate (2017-2018):**
   * The churn rate declined from **3.51% in 2017 to 1.82% in 2018**, indicating improved customer retention.
2. **Negative Churn in 2019 (-7.41%):**
   * This suggests that more customers joined than left, possibly due to successful marketing, improved services, or customer reactivation.

### **Business Implications:**

* The decline in churn from 2017-2018 shows **better customer retention efforts**.
* The negative churn in 2019 indicates **growth** and a potential increase in brand loyalty.

To further reduce churn, the business could consider:

* Offering loyalty programs or personalized promotions.
* Engaging inactive customers through targeted campaigns.
* Analysing feedback from churned customers to address pain points.

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

with cte as(

select g.name as genre\_name, sum(il.unit\_price\*quantity) as sales

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

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

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

where billing\_country = 'USA'

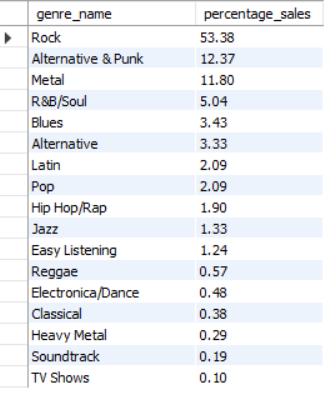
group by g.name)

, totl\_sale\_amount as(

select sum(sales) total\_sales from cte)

select genre\_name, round((sales/total\_sales)\*100,2) as percentage\_sales

from cte, totl\_sale\_amount

order by 2 desc;  
  
  
  
  
**-- part 2: the best-selling genres and artists worldwide.**

**-- (a) best-selling genres**

with genre\_ranking as(

select

g.name as genre\_name,

sum(il.unit\_price\*quantity) as sales,

dense\_rank() over (order by sum(il.unit\_price\*quantity) desc) as rnk

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

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

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

group by g.name)

select \* from genre\_ranking where rnk=1;



**-- b. top artists**

with artist\_ranking as(

select

ar.artist\_id,

ar.name as artist\_name,

sum(il.unit\_price\*quantity) as sales,

dense\_rank() over (order by sum(il.unit\_price\*quantity) desc) as rnk

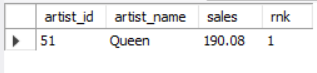
from invoice\_line il

join track t on il.track\_id=t.track\_id

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

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

group by 1,2)

select \* from artist\_ranking where rnk=1;  
  
  
Answer:

Key Insights:

* More than half of total sales come from Rock, making it the most popular genre.
* Alternative & Punk (12.37%) and Metal (11.80%) along with Rock, account for over 75% of total sales, highlighting a strong preference for guitar-driven music.
* Niche and Lower-Contributing Genres
* Jazz (1.33%), Easy Listening (1.24%), and Reggae (0.57%) have smaller but dedicated audiences.
* Classical (0.38%), Heavy Metal (0.29%), and Soundtrack (0.19%) have limited sales.
* TV Shows (0.10%) contribute the least to total sales.

### **Business Implications:**

* **Marketing efforts should prioritize Rock, Alternative, and Metal**, as they dominate sales.
* **Expanding into emerging genres like Latin and Hip Hop/Rap** could open new revenue streams.
* **Niche genres like Jazz and Classical should be targeted towards specific audiences** through curated playlists and personalized recommendations.

This data provides valuable insight into customer preferences and can guide **inventory planning, targeted marketing, and promotional strategies** to maximize revenue.

**2nd part:** Best-Selling Genres and Artists Worldwide

**(a) Best-Selling Genres:**

* **Rock** is the **best-selling genre,** with total sales of **2,608.65** units worldwide.
* This dominance suggests a strong global fan base, a rich catalog of classic and contemporary rock music, and consistent demand across different regions.

(b) Best-Selling Artists Worldwide**:**

* Queen ranks as the #1 best-selling artist, with total sales of 190.08 units. Queen's lasting popularity can be attributed to their iconic hits, legendary performances.

1. Find customers who have purchased tracks from at least 3 different genres.  
   Answer:  
   **Code:**  
   select

c.customer\_id,

concat(first\_name,' ', last\_name) as customer\_name,

count(distinct g.genre\_id) as genre\_count

from

customer 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

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

group by 1,2

having count(distinct g.genre\_id)>=3

### order by genre\_count desc; The customers have demonstrated diverse musical tastes by purchasing tracks from at least **three different genres**. **Insights:**

* **Leonie Köhler** (Customer ID: 2) has purchased tracks from **14 different genres,** making her the most diverse music listener.
* Other top customers include **František Wichterlová (13 genres), Terhi Hämäläinen (13 genres), and Madalena Sampaio (13 genres).**
* The table provides a **list of 20 customers,** all of whom have purchased tracks from **at least 10 different genres.**

### **Possible Use Cases:**

* **Customer Personalization:** These customers might enjoy **a diverse recommendation system** to match their varied taste.
* **Marketing Strategies:** Special offers or promotions on **multi-genre playlists** can be targeted at such users.
* **Music Store Analysis**: Identifying customers with broad genre interests can help **optimize inventory or playlist curation.**

1. Rank genres based on their sales performance in the USA.  
   Answer:  
   **Code:**  
   select

g.name as genre\_name,

sum(il.unit\_price\*quantity) as sales,

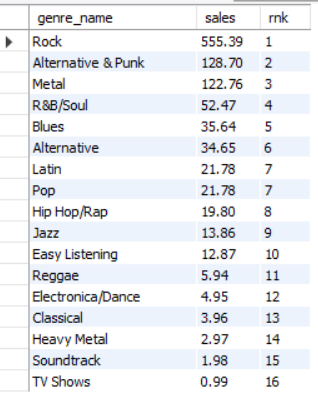
dense\_rank() over (order by sum(il.unit\_price\*quantity) desc) as rnk

from

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

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

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

where i.billing\_country = 'USA'  
group by g.name;  
  
  
  
  
The **top-performing music genres** in the **USA**, based on total sales, are ranked.  
**Insights Based on Sales Ranking**

1. **Rock is the Most Popular Genre** – With the highest sales, it indicates that Rock music has the strongest demand among customers in the USA.
2. **Alternative & Punk and Metal** – These genres take the **2nd and 3rd** positions, showing strong preference among listeners.
3. **R&B/Soul and Blues –** These genres still have notable sales but rank **4th and 5th,** showing a significant drop compared to Rock.
4. **Classical and Heavy Metal** – These genres have the lowest sales, indicating they are niche markets in the USA.
5. **TV Shows and Soundtracks Have the Least Sales** – This suggests that customers in the USA primarily purchase music rather than media-related tracks.
6. Identify customers who have not made a purchase in the last 3 months.  
   Answer:  
   **Code:**select   
    customer\_id,   
    concat(first\_name,' ',last\_name) as customer\_name   
   from   
    customer   
   where customer\_id   
   not in (  
    SELECT DISTINCT customer\_id  
    FROM invoice  
    WHERE invoice\_date >= DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 3 MONTH)  
   )order by customer\_id;  
     
   Answer: The above mentioned **22 customers** have not made any purchases in the last three months, indicating **potential churn risk** or decreased engagement:

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:  
   **Code:**   
   with genre\_analysis as (

select

g.genre\_id,g.name as genre\_name,

sum(il.unit\_price\*quantity) as sales

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

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

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

where billing\_country='USA'

group by g.genre\_id,g.name

order by sales desc

limit 3)  
  
 -- Step 2: Identify top-selling albums from the new record label

, AlbumSales AS (

SELECT

a.album\_id,

a.title AS album\_name,

g.name AS genre\_name,

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

dense\_rank() over (order by SUM(il.unit\_price \* il.quantity) desc) as rnk

FROM invoice i

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

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

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

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

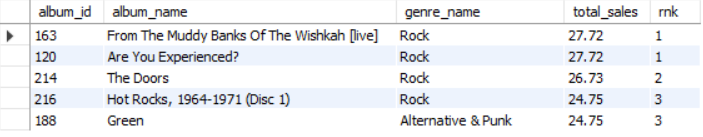
WHERE i.billing\_country = 'USA'

AND g.genre\_id IN (SELECT genre\_id FROM genre\_analysis) -- Filter albums in top genres

GROUP BY a.album\_id, a.title, g.name

ORDER BY total\_sales DESC

)

SELECT \* FROM AlbumSales where rnk<4;  
  
  
  
  
 **Approach:**

1. Identify the Top-Selling Genres in the USA:
   * We first determine the most profitable music genres in the USA by summing up total sales (unit price \* quantity) per genre.
   * The top 3 genres are then selected to focus on high-performing categories.
2. Identify Top-Selling Albums Within These Genres:
   * We filter albums belonging to the top-performing genres.
   * Sales figures are calculated for each album in the USA.
   * Albums are ranked using DENSE\_RANK() to highlight the best-performing ones.
3. Select the Top 3 Albums for Advertising & Promotion:
   * From the ranked albums, the top 3 are selected for prioritization in advertising and promotion.

**Insights:**

* Rock dominates the US market, with three of the top four albums belonging to this genre.
* Alternative & Punk also performs well, securing a spot in the top 3 albums.
* Rock music has strong audience demand and revenue potential in the USA, making it the most lucrative genre for promotion.

**Recommendations:**

1. Prioritize Marketing & Advertising for Rock Albums
   * Since Rock has the highest sales, allocate a larger portion of the marketing budget to Rock albums.
   * Invest in digital and traditional advertising campaigns targeting Rock music listeners.
2. Feature the Top-Selling Albums in Promotions
   * Focus on albums *From The Muddy Banks Of The Wishkah [live]*, *Are You Experienced?*, and *The Doors* in ad campaigns.
   * Offer exclusive promotions, limited-edition releases, and live event tie-ins.
3. Leverage Streaming & Social Media for Maximum Reach
   * Partner with music streaming platforms (Spotify, Apple Music) to feature these albums in curated playlists.
   * Use social media campaigns and influencer collaborations to increase awareness and engagement.
4. Expand Alternative & Punk Promotion
   * The album *Green* has strong sales within the Alternative & Punk genre.
   * A secondary marketing campaign targeting Alternative Rock listeners can drive additional sales.
5. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

**Code:**   
with top\_genre\_in\_country as(

select

i.billing\_country as country,

g.name as genre\_name,

sum(il.unit\_price\*quantity) as sales,

dense\_rank() over (partition by billing\_country order by sum(il.unit\_price\*quantity) desc) as rnk

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

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

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

where billing\_country <> 'USA'

group by i.billing\_country,g.genre\_id,g.name )

select

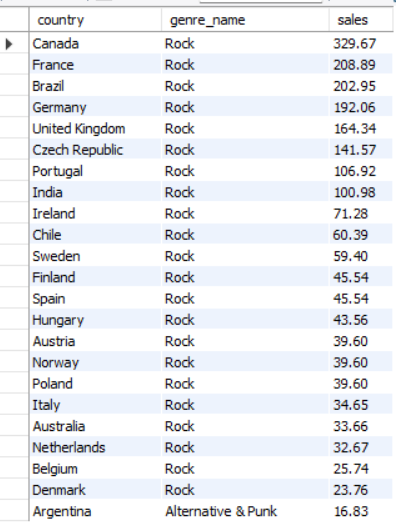
country,

genre\_name,

sales

from top\_genre\_in\_country

where rnk=1

order by sales desc;  
  
  
  
 **Approach:**

1. Identify the Top-Selling Genre in Each Country (Excluding the USA)
   * We calculate the total sales for each genre by country.
   * We use DENSE\_RANK() to rank genres within each country based on sales.
   * Only the top genre per country (where rnk = 1) is selected for further analysis.
2. Analyze the Most Popular Genre Across Countries
   * We sort the data by total sales to determine which genre dominates globally (outside the USA).
   * We identify the most consistent genre preference across multiple countries.

**Insights:**

* **Rock is the dominant genre in nearly every country outside the USA, just like in the USA. This suggests a global preference for Rock music across diverse regions.**
* **Canada, France, Brazil, and Germany have the highest Rock sales outside the USA, showing particularly strong Rock fan bases.**
* **Even in non-English-speaking countries like Czech Republic, Portugal, India, and Spain, Rock still tops the charts, indicating its universal appeal.**

**Recommendations:** Rock is the best-selling genre worldwide, except in Argentina. This trend supports continued investment in Rock music promotion globally, with some focus on Alternative & Punk in specific markets like Argentina.

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

**Code:**with CustomerTenure AS (

SELECT

c.customer\_id,

c.first\_name,

c.last\_name,

c.country,

MIN(i.invoice\_date) AS first\_purchase\_date, -- First purchase date

COUNT(DISTINCT i.invoice\_id) AS total\_purchases, -- Purchase frequency

SUM(il.unit\_price\*quantity) AS total\_spent, -- Total spending

ROUND(AVG(il.unit\_price\*quantity), 2) AS avg\_spent\_per\_purchase, -- Average order value

ROUND(AVG(track\_count), 0) AS avg\_tracks\_per\_invoice, -- Basket size

CASE

WHEN MIN(i.invoice\_date) <= DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 3 YEAR)

THEN 'Long-Term'

ELSE 'New'

END AS customer\_type -- Classify customers as Long-Term or New

FROM customer c

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

JOIN (

SELECT

invoice\_id,

COUNT(track\_id) AS track\_count

FROM invoice\_line

GROUP BY invoice\_id

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

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

GROUP BY c.customer\_id

)

SELECT

customer\_type,

COUNT(customer\_id) AS total\_customers,

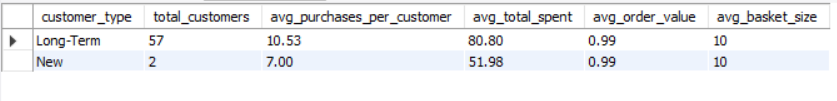
ROUND(AVG(total\_purchases), 2) AS avg\_purchases\_per\_customer,

ROUND(AVG(total\_spent), 2) AS avg\_total\_spent,

ROUND(AVG(avg\_spent\_per\_purchase), 2) AS avg\_order\_value,

ROUND(AVG(avg\_tracks\_per\_invoice), 0) AS avg\_basket\_size

FROM CustomerTenure

GROUP BY customer\_type;  
  


**Approach:**

1. **Classifying Customers Based on Tenure**
   * Customers who made their first purchase more than 3 years ago are classified as Long-Term Customers.
   * Customers who made their first purchase within the last 3 years are classified as New Customers.
2. **Analysing Purchasing Behaviour Metrics**
   * Total Purchases: Frequency of orders placed.
   * Total Spent: Total revenue contribution per customer group.
   * Average Order Value (AOV): The amount spent per order.
   * Basket Size: The number of tracks purchased per invoice.
3. **Comparing Long-Term vs. New Customers**
   * Aggregate statistics are computed for each customer group to identify behavioural differences.

### **Insights:**

* **Long-term customers** (57 total) make **more frequent purchases** (10.53 on average) and spend **more overall** ($80.80 per customer).
* **New customers** (2 total) make **fewer purchases** (7 on average) and spend **less overall** ($51.98 per customer).
* **Basket size and order value remain consistent** across both groups, meaning spending habits are stable, but **engagement differs.**

### **Retention Strategies to Boost Loyalty:**

* Loyalty Programs: Offer discounts, points, or exclusive deals for repeat purchases.
* Personalized Engagement: Recommend tailored music collections or new releases based on past purchases.
* Reactivation Campaigns: Target new customers with limited-time offers to encourage repeat purchases.
* Subscription Model: Consider offering a VIP membership with exclusive content or discounts to incentivize long-term spending.

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

**Code:   
-- genre**

SELECT

g1.name AS genre\_1,

g2.name AS genre\_2,

COUNT(\*) AS purchase\_count

FROM invoice\_line il1

JOIN track t1 ON il1.track\_id = t1.track\_id

JOIN genre g1 ON t1.genre\_id = g1.genre\_id

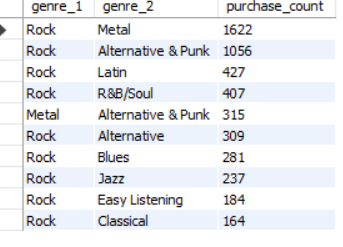
JOIN invoice\_line il2 ON il1.invoice\_id = il2.invoice\_id AND il1.track\_id <> il2.track\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

JOIN genre g2 ON t2.genre\_id = g2.genre\_id

WHERE g1.genre\_id < g2.genre\_id -- Avoid duplicate pairs

GROUP BY genre\_1, genre\_2

ORDER BY purchase\_count DESC  
LIMIT 10;  
  
  
**Approach:**

This query identifies the **top 10 most commonly purchased genre pairings** by analyzing customer purchase patterns from invoices.

1. **Self-Joining the** invoice\_line **Table:**
   * The query joins invoice\_line (il1 and il2) on the same invoice\_id, meaning we are looking at purchases **within the same invoice** (i.e., items bought together).
   * The condition il1.track\_id <> il2.track\_id ensures that different tracks from the same invoice are considered.
2. **Mapping Tracks to Genres:**
   * Each track is mapped to a **genre** (g1 for il1 and g2 for il2).
   * The condition g1.genre\_id < g2.genre\_id eliminates duplicate pairings (e.g., avoiding both **(Rock, Pop)** and **(Pop, Rock)** being counted separately).
3. **Grouping & Counting:**
   * The query groups by **genre pairs** and counts the number of invoices where the genre combination appears together.
   * The results are **sorted in descending order** by purchase\_count to identify the **most frequently purchased genre combinations.**

**Key Findings:**

* Rock music is the most frequently paired genre, appearing alongside Metal (1,622 purchases), Alternative & Punk (1,056), and Latin (427).
* Metal and Alternative & Punk also show strong connections (315 purchases together).
* Cross-genre purchases suggest that customers have diverse music tastes and are open to exploring similar genres.

**-- artist**

SELECT

a1.name AS artist\_1,

a2.name AS artist\_2,

COUNT(\*) AS purchase\_count

FROM invoice\_line il1

JOIN track t1 ON il1.track\_id = t1.track\_id

JOIN album al1 ON t1.album\_id = al1.album\_id

JOIN artist a1 ON al1.artist\_id = a1.artist\_id

JOIN invoice\_line il2 ON il1.invoice\_id = il2.invoice\_id AND il1.track\_id <> il2.track\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

JOIN album al2 ON t2.album\_id = al2.album\_id

JOIN artist a2 ON al2.artist\_id = a2.artist\_id

WHERE a1.artist\_id < a2.artist\_id -- Avoid duplicate pairs

GROUP BY artist\_1, artist\_2

ORDER BY purchase\_count DESC  
LIMIT 10;  
  
  
  
**Approach:**

This query analyzes the **top 10 most frequently purchased artist pairings**, showing which artists’ music is often bought together in the same order.

1. **Self-Joining the** invoice\_line **Table:**
   * The query joins invoice\_line (il1 and il2) on the same invoice\_id, ensuring we analyze **tracks purchased together in the same transaction**.
   * The condition il1.track\_id <> il2.track\_id ensures only different tracks within an invoice are considered.
2. **Mapping Tracks to Artists:**
   * Each track is linked to an **album** (al1 for il1 and al2 for il2).
   * Each album is linked to an **artist** (a1 for il1 and a2 for il2).
   * The condition a1.artist\_id < a2.artist\_id eliminates duplicate pairings (avoiding both **(Led Zeppelin, Green Day)** and **(Green Day, Led Zeppelin)** being counted separately).
3. **Grouping & Counting:**
   * The query **groups by artist pairs** and counts the number of transactions in which they were bought together.
   * The results are sorted in **descending order** of purchase\_count, showing the **most frequently paired artists.**

**Insights:**

* Led Zeppelin & Green Day were bought together 24 times, indicating a strong correlation between their fan bases.
* Green Day appears in 4 of the top 10 pairs, showing its music is often purchased alongside multiple artists.
* Classic rock bands (e.g., Led Zeppelin, Queen, The Rolling Stones) are frequently paired with punk/alternative bands (e.g., Green Day, Nirvana, Foo Fighters)—suggesting that fans of classic rock often explore newer rock subgenres.

**-- album**

SELECT

al1.title AS album\_1,

al2.title AS album\_2,

COUNT(\*) AS purchase\_count

FROM invoice\_line il1

JOIN track t1 ON il1.track\_id = t1.track\_id

JOIN album al1 ON t1.album\_id = al1.album\_id

JOIN invoice\_line il2 ON il1.invoice\_id = il2.invoice\_id AND il1.track\_id <> il2.track\_id

JOIN track t2 ON il2.track\_id = t2.track\_id

JOIN album al2 ON t2.album\_id = al2.album\_id

WHERE al1.album\_id < al2.album\_id -- Avoid duplicate pairs

GROUP BY album\_1, album\_2

ORDER BY purchase\_count DESC

LIMIT 10;  
  


**Approach:**

* Identifies top 10 most frequently purchased album pairs using self-join on invoice\_line.
* Ensures unique pairings (al1.album\_id < al2.album\_id) and counts occurrences.

**Insights:**

1. "Mezmerize" dominates pairings, appearing in 5 of the top 10.
2. Classic Rock & Greatest Hits albums are frequently bought together (e.g., *Are You Experienced?* + *Vault*).
3. Cross-genre appeal—alternative and classic rock albums are often purchased together.

**Recommendations:**

* Create Rock & Metal album bundles or offer discounts when purchasing from both genres.
* If a customer buys Rock albums, suggest Metal or Alternative & Punk albums based on purchase patterns.
* Run email or in-app promotions featuring deals on related genres.
* Offer limited-time discounts for customers who have previously bought from one genre (e.g., Rock buyers get a Metal album at 20% off).
* Offer **discounted bundles** featuring popular artist pairings (e.g., **Led Zeppelin + Green Day, Nirvana + The Rolling Stones).**
* Advertise **concert collaborations** or feature artist playlists based on pairing trends.
* Offer Bundles & Discounts on frequently paired albums to boost sales.
* Implement Personalized Suggestions at checkout based on purchase patterns.
* Run Targeted Ads focusing on classic rock and alternative rock fans.

1. 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?  
   **Code:**

**-- purchasing behaviours across different geographic regions**

select

country,

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

count(invoice\_id) as total\_purchases,

ROUND(COUNT(i.invoice\_id) / COUNT(DISTINCT c.customer\_id), 2) AS avg\_purchases\_per\_customer

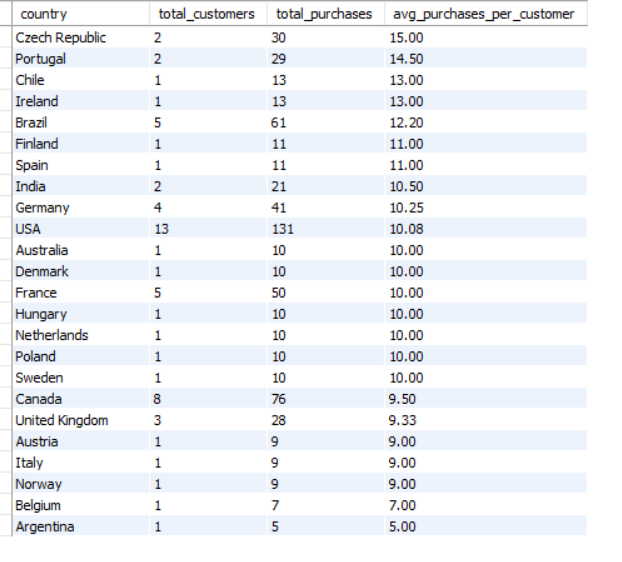
from

customer c

join

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

group by country

order by avg\_purchases\_per\_customer desc;  
  
  
**Approach:**

* Group customers by country and count unique customers.
* Calculate total purchases and derive average purchases per customer.
* Rank countries by engagement to identify high-retention markets.

**-- churn rates across different geographic regions**

WITH LatestDate AS (

SELECT MAX(invoice\_date) AS max\_invoice\_date FROM invoice

),

CustomerActivity AS (

SELECT

c.customer\_id,

c.country,

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

FROM customer c

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

GROUP BY c.customer\_id, c.country

)

SELECT

ca.country,

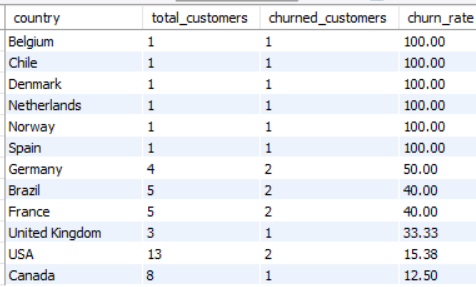
COUNT(DISTINCT ca.customer\_id) AS total\_customers,

COUNT(DISTINCT CASE WHEN ca.last\_purchase\_date <= DATE\_SUB((SELECT max\_invoice\_date FROM LatestDate), INTERVAL 6 MONTH) THEN ca.customer\_id END) AS churned\_customers,

ROUND(100 \* COUNT(DISTINCT CASE WHEN ca.last\_purchase\_date <= DATE\_SUB((SELECT max\_invoice\_date FROM LatestDate), INTERVAL 6 MONTH) THEN ca.customer\_id END) / COUNT(DISTINCT ca.customer\_id), 2) AS churn\_rate

FROM CustomerActivity ca

GROUP BY ca.country HAVING churn\_rate>0

ORDER BY churn\_rate DESC;  
  


**Approach:**

* Identify the latest purchase date from the invoice table.
* Determine each customer’s last purchase date and group by country.
* Count total customers and those inactive for 6+ months (churned customers).
* Calculate churn rate as a percentage of total customers.
* Rank countries by churn rate to highlight high-risk markets.

**Purchasing behaviour varies by region  
Insights:**

* Czech Republic & Portugal have the highest average purchases per customer (15 & 14.5, respectively), suggesting a strong customer engagement.
* Brazil (12.2) and France (10) also show solid purchasing activity, possibly due to market familiarity with the platform.
* Lower engagement in Argentina (5 purchases per customer) suggests weaker market penetration.

**Churn rates across different geographic regions**  
**Insights:**

* High-Churn Countries (100%)
* Belgium, Chile, Denmark, Netherlands, Norway, and Spain all have 100% churn rates. Every customer in these regions stopped purchasing.
* Likely causes: Lack of localized content, limited engagement, or low brand presence.
* Moderate-Churn Countries (33–50%)
* Germany (50%), Brazil (40%), France (40%), and the UK (33.33%) show mid-range churn.
* Possible factors: Less frequent promotions, weaker customer retention strategies, or market competition.
* Low-Churn Markets (≤15%)
* USA (15.38%) and Canada (12.5%) maintain strong customer retention.
* These markets likely have better engagement, promotional efforts, and subscription models to keep customers active.

**Recommendations:**

* Re-engagement campaigns in high-churn markets (Belgium, Chile, etc.) using localized promotions, discounts, or curated playlists.
* Strengthen loyalty programs in mid-churn countries (Germany, Brazil and France) by offering bundled music packages or cross-selling with related genres.
* **Leverage Low-Churn Success in USA & Canada** to **implement similar customer retention strategies globally.**

#### 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: **Code:** with Customer\_risk\_profiling as( select country, c.customer\_id, max(invoice\_date) as last\_purchase\_date, count(i.invoice\_id) as total\_purchases, sum(quantity\*unit\_price) as total\_spent from customer c join invoice i on c.customer\_id=i.customer\_id join invoice\_line il on il.invoice\_id=i.invoice\_id group by c.customer\_id, country)

#### select country, count(customer\_id) as total\_customers, round(avg(total\_purchases),0) as avg\_purchase\_count, round(avg(total\_spent),2) as avg\_total\_spent, count(case when last\_purchase\_date<= DATE\_SUB((select max(invoice\_date) from invoice),interval 6 month) then 1 end) as at\_risk\_customers, case when COUNT(CASE WHEN last\_purchase\_date <= DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 6 MONTH) THEN 1 END) > (COUNT(customer\_id) \* 0.5) then 'High Risk' else 'Low Risk' end as risk\_category from Customer\_risk\_profiling group by country having at\_risk\_customers <> 0 order by risk\_category, at\_risk\_customers desc;

**Approach:**

* Calculate each customer’s last purchase date, total purchases, and total spending.
* Categorize customers "At-Risk" if they haven’t purchased in the last 6 months.
* Calculate total customers, average purchases, and spending per country.
* Assign a risk category:
  + High Risk: More than 50% of customers are at risk.
  + Low Risk: Less than 50% at risk.
* Rank countries based on risk level and at-risk customers.

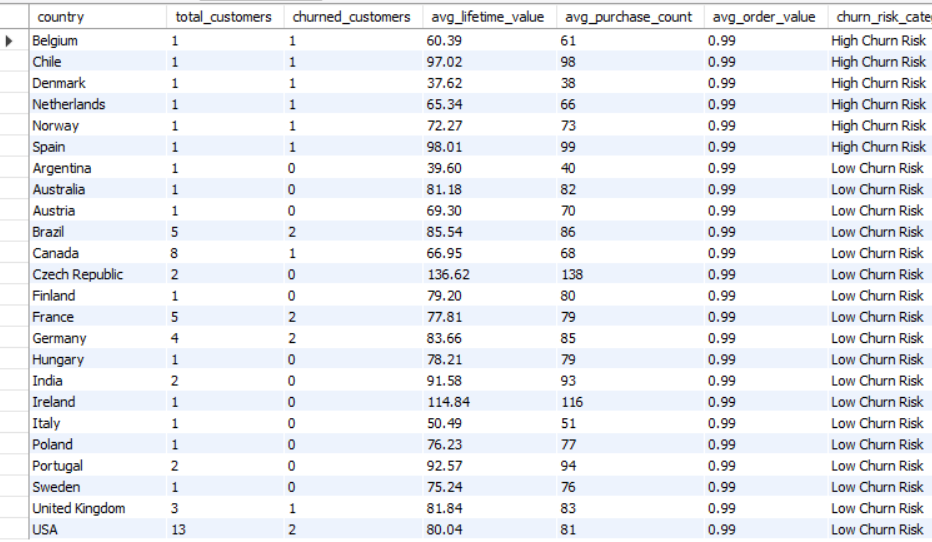
**Key Insights:**

1. **High-Risk Customers (Likely to Churn or Reduce Spending)**
   * Countries: **Denmark, Norway, Spain, Chile, Netherlands, Belgium**
   * **Risk Factors:**
     + **Low total customer count (1 per country)** → less market penetration.
     + **High purchase frequency but relatively low total spend** → Indicates a drop in purchasing behaviour or dissatisfaction.
2. **Low-Risk Customers ( More Stable & Engaged)**
   * Countries: **USA, France, Germany, Brazil, Canada, United Kingdom**
   * **Risk Factors:**
     + Higher total spend (~80-86) and frequent purchases (~68-86 purchases per customer).
     + **Lower churn rates (≤15%)** indicate **better retention strategies** in place.

**Factors Contributing to Churn Risk:**

* **Market Awareness & Localized Offerings:** New or smaller markets (e.g., Denmark, Belgium) may lack localized promotions.
* **Economic Conditions & Affordability:** Some regions might have **price sensitivity issues** affecting long-term engagement.
* **Product Fit & Personalization:** If **music preferences aren't met**, customers may reduce spending or switch platforms.
* **Competitive Market Landscape:** Countries with **strong local music streaming competitors** (e.g., Europe) may see **higher churn**.  
    
    
  **Recommendations:**
* Target High-Risk Countries: Denmark, Norway, Spain, and Chile have 100% at-risk customers. Launch reactivation campaigns (e.g., discounts, personalized offers).
* Loyalty Programs: Low-risk countries (USA, France and Brazil) still have at-risk customers. Implement rewards programs to boost engagement.
* Behaviour-Based Retention: Analyze spending patterns to identify triggers for churn and offer early engagement incentives.

### 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? **Code:** WITH CustomerMetrics AS ( SELECT c.customer\_id, c.country, MIN(i.invoice\_date) AS first\_purchase\_date, -- Customer tenure start MAX(i.invoice\_date) AS last\_purchase\_date, -- Last purchase date COUNT(i.invoice\_id) AS total\_purchases, -- Purchase frequency SUM(quantity\*unit\_price) AS total\_spent, -- Total revenue from customer ROUND(AVG(quantity\*unit\_price), 2) AS avg\_order\_value, -- Average purchase amount (SELECT MAX(invoice\_date) FROM invoice) AS max\_invoice\_date, CASE WHEN MAX(i.invoice\_date) <= DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 6 MONTH) THEN 'Churned' ELSE 'Active' END AS customer\_status -- Identify churned vs active customers FROM customer c JOIN invoice i ON c.customer\_id = i.customer\_id join invoice\_line il on il.invoice\_id=i.invoice\_id GROUP BY c.customer\_id, c.country )

SELECT   
 country,  
 COUNT(customer\_id) AS total\_customers,  
 COUNT(CASE WHEN customer\_status = 'Churned' THEN 1 END) AS churned\_customers,  
 ROUND(AVG(total\_spent), 2) AS avg\_lifetime\_value,  
 ROUND(AVG(total\_purchases), 0) AS avg\_purchase\_count,  
 ROUND(AVG(avg\_order\_value), 2) AS avg\_order\_value,  
 CASE   
 WHEN COUNT(CASE WHEN customer\_status = 'Churned' THEN 1 END) > (COUNT(customer\_id) \* 0.5) THEN 'High Churn Risk'  
 ELSE 'Low Churn Risk'  
 END AS churn\_risk\_category  
FROM CustomerMetrics  
GROUP BY country;  
order by churn\_risk\_category, country;  


**Approach:**

* Identify first & last purchase dates per customer.
* Classify as Active (recent buyers) or Churned (inactive >6 months).
* Calculate customer count, churn rate, avg lifetime value and purchase frequency.
* Categorize High vs. Low Churn Risk (≥50% churned).

### **Observations & Key Insights: 1️. High Churn Risk Countries**

* **Belgium, Chile, Denmark, Netherlands, Norway, Spain** all have a **100% churn rate**.
* Customers in these countries had **relatively lower lifetime value (CLV)** compared to other regions.
* **Common purchase pattern:**
  + The **average purchase count is high** (above 60 purchases per customer).
  + However, the **average order value is low** ($0.99 per order).
  + This suggests that **these customers made frequent small purchases but eventually stopped buying**.

### **2️. Low Churn Risk Countries**

* Countries like **Czech Republic (highest CLV - 136.62), Ireland (114.84), India (91.58), and Portugal (92.57)** have:
  + Higher **customer lifetime value (CLV)**.
  + Higher **average purchase count**.
  + Lower **churn rates**.

### **3️. USA & Major Markets (Canada, UK, Germany, Brazil, France)**

* The USA, Canada, UK, Germany, and Brazil **have relatively stable customers** but **some churn risk (12-40%)**.
* Their **CLV is around $80, with an average of 81 purchases per customer**.
* Churn **is not as extreme as in smaller markets, but still notable**.  
    
  **Recommendations:**
* Re-engage High-Churn Countries (Belgium, Chile, etc.) with discounts & loyalty perks.
* Retain Low-Churn Markets with bundles & exclusive offers.
* Proactive campaigns for Brazil, USA, and France to prevent rising churn.

## 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: To evaluate how promotional campaigns (discounts, events, email marketing) influence customer behaviour, you can analyse three key areas: 1.Customer Acquisition – How many new customers were gained. 2️. Customer Retention – Did existing customers continue purchasing? 3. Overall Sales Impact – Did total revenue increase. Insights from Promotional Campaign Analysis

1️. I**f customer acquisition increased,** promotions successfully **attracted new buyers**.  
2️. **If retention improved,** campaigns helped keep **existing customers engaged**.  
3️. **If total sales increased,** discounts/events **boosted revenue** effectively.  
4️. **If churn increased post-campaign,** it may suggest **customers only buy during promotions** → Need a loyalty strategy!

1. How would you approach this problem, if the objective and subjective questions weren't given?  
   Answer: If the objective and subjective questions weren’t given, I would approach the problem systematically using **exploratory data analysis (EDA) and business goals** to identify key insights.  
     
   **Understanding the Business Context**

* **What are the business goals? (e.g., increase revenue, reduce churn and improve customer retention)**
* **Who are the key stakeholders? (Marketing, Sales, Product Teams)**
* **What business challenges are we trying to solve?**

**Exploring the Data (EDA):**

* **Customer Data** (demographics, location, tenure)
* **Sales Data** (invoices, purchases, total revenue)
* **Product Data** (genres, albums, artists)

**Identifying Key Business Metrics:**

* Customer Acquisition Rate
* Retention & Churn Rate
* Total Revenue
* Average Order Value (AOV)
* Which genres/albums sell together?

1. 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?  
   **Code:**  
   alter table album add column ReleaseYear INTEGER;  
     
     
     
   Answer: To **add a new column** named ReleaseYear of type INTEGER to the Albums table used Alter clause.
2. 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.  
   **Code:**  
   select

billing\_country as country,

count(distinct i.customer\_id) as total\_customers,

round(avg(total\_spent),2) as avg\_total\_spent\_per\_customer,

round(avg(track\_count),0) as avg\_tracks\_per\_customer

from invoice i

join (

select

customer\_id,

sum(il.unit\_price\*quantity) as total\_spent,

count(t.track\_id) AS track\_count

from

invoice i

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

join track t on t.track\_id = il.track\_id

group by 1

) t on i.customer\_id=t.customer\_id

group by billing\_country

### order by avg\_total\_spent\_per\_customer desc; **Approach:**

* Calculate **total spent & track count per customer**.
* Aggregate **average total spent & tracks per customer by country**.
* Rank countries by **spending behaviour** to identify high-value markets.

### **Recommendations:**

* **Target high-spending markets** (Czech Republic, Ireland and Spain) with **premium offers.**
* **Encourage engagement** in **low-spending markets** (Denmark, Argentina) with **discounts & curated bundles.**
* **Enhance customer retention** in mid-tier markets (USA, Brazil and UK) with **personalized recommendations.**