**OBJECTIVES**

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

**Query to check for Duplicates**

select album\_id,title,artist\_id,count(\*) as cnt

from album

group by 1,2,3

having count(\*) > 1;

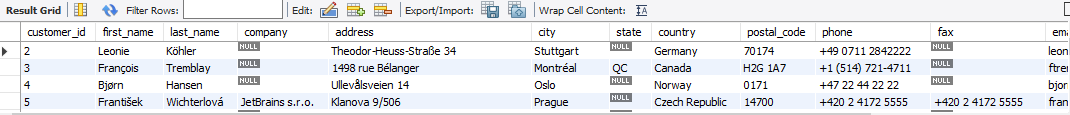
**Query to check for null values:**

select album\_id,title,artist\_id

from album

where album\_id is null or title is null or artist\_id is null;

**OUTPUT:**



**OBSERVATION:**

* + Yes, the customer table, employee table and track table has missing/NULL values in it.
  + No, duplicate values found in any of the given tables
  + To handle missing/NULL values,
* We can replace NULL with a default value
* We can replace missing values with N/A
* If the missing values makes the row unusable, we may also delete them
  + To handle duplicate values,
* To remove duplicate records, we can use distinct to keep only the first occurrence
* Instead of deleting duplicates we can just filter them while writing the query.

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

**QUERY:**

WITH TopTracks AS (

SELECT t.track\_id, t.name AS track\_name, g.name AS genre\_name, SUM(il.quantity) AS total\_units\_sold

FROM invoice\_line il

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

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

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

WHERE i.billing\_country = 'USA'

GROUP BY t.track\_id, t.name, g.name

ORDER BY total\_units\_sold DESC

LIMIT 5

),

TopArtist AS (

SELECT ar.artist\_id, ar.name AS artist\_name, SUM(il.quantity) AS total\_units\_sold

FROM invoice\_line il

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

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

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

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

WHERE i.billing\_country = 'USA'

GROUP BY ar.artist\_id, ar.name

ORDER BY total\_units\_sold DESC

LIMIT 1

),

TopArtistGenre AS (

SELECT DISTINCT ar.artist\_id, g.name AS genre\_name

FROM track t

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

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

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

WHERE ar.artist\_id = (SELECT artist\_id FROM TopArtist)

)

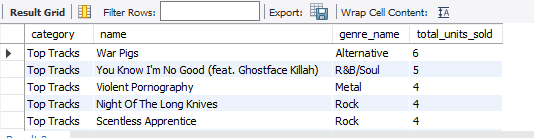
SELECT 'Top Tracks' AS category, track\_name AS name, genre\_name, total\_units\_sold FROM TopTracks

UNION ALL

SELECT 'Top Artist', artist\_name, genre\_name, total\_units\_sold FROM TopArtistGenre

JOIN TopArtist USING (artist\_id);

**OUTPUT:**



**VISUALIZATION:**

**OBSERVATION:**

* Rock is the most popular genre in the USA, based on both top-selling tracks and the best-selling artist.
* Rock appears twice in the top 5 tracks, suggesting it is one of the popular genres in the USA.
* Van Halen dominates the USA market, suggesting a strong fanbase and high album sales.
* The USA music market is diverse, as seen from the mix of Alternative, R&B/Soul, and Metal in the top tracks.
* Rock-based genres (Rock, Metal, Alternative) make up 4 out of 5 top-selling tracks, proving that heavier music styles are widely listened to.
* R&B/Soul is also present, showing that some listeners prefer different styles.

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

**QUERY:**

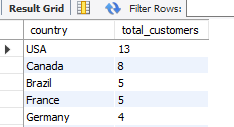
SELECT country, COUNT(customer\_id) AS total\_customers

FROM customer

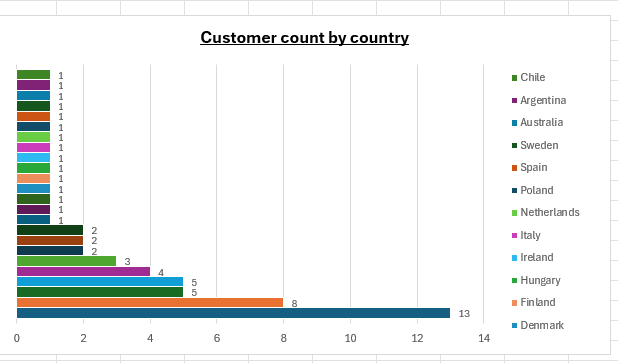
GROUP BY country

ORDER BY total\_customers DESC;

**OUTPUT:**



**VISUALIZATION:**



**OBSERVATION:**

* USA and Canada dominate Chinook’s customer base, making up over 50% of total customers.
* Brazil and France are important secondary markets, showing potential for expansion.
* Europe has multiple small markets, but no single country is a dominant player.
* Presence in India, Australia, and Latin America suggests Chinook could focus on global marketing

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

**QUERY:**

SELECT billing\_country AS country, billing\_state AS state,

billing\_city AS city, COUNT(invoice\_id) AS total\_invoices,

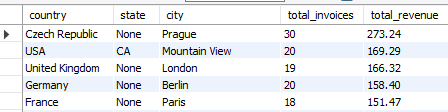
SUM(total) AS total\_revenue

FROM invoice

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

ORDER BY total\_revenue DESC;

**OUTPUT:**

****

**VISUALIZATION:**

**OBSERVATION:**

* The USA is the largest market, contributing 22.1% of total revenue ($1040.49 out of $4709.43)
* Canada and Brazil also perform strongly, with Brazil ranking third despite having fewer customers than Canada.
* European countries like France and Germany are major revenue drivers, showing strong engagement in music purchases.

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

**QUERY:**

WITH CustomerRevenue AS (

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

i.billing\_country AS country,SUM(i.total) AS total\_revenue,

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

FROM customer c

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

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

)

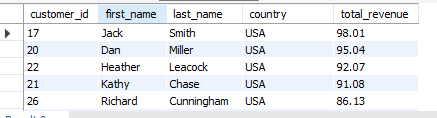
SELECT customer\_id,first\_name,last\_name,country,total\_revenue

FROM CustomerRevenue

WHERE rnk <= 5

ORDER BY country DESC,total\_revenue desc;

**OUTPUT:**

****

**OBSERVATION:**

* Certain customers (possibly from Brazil, Canada, or the USA) seem to have higher total revenue.
* Some countries have a more even revenue distribution, while others have one or two dominant customers.

1. Identify the top-selling track for each customer

**QUERY:**

WITH CustomerTrackSales AS (

SELECT i.Customer\_Id,il.Track\_Id,t.Name AS track\_name,

SUM(il.Quantity) AS total\_quantity\_sold

FROM Invoice\_Line il

JOIN Invoice i ON il.Invoice\_Id = i.Invoice\_Id

JOIN Customer c ON i.Customer\_Id = c.Customer\_Id

JOIN Track t ON il.Track\_Id = t.Track\_Id

GROUP BY i.Customer\_Id,il.Track\_Id, t.Name

), RankedTracks AS (

SELECT Customer\_Id,track\_name,total\_quantity\_sold,

ROW\_NUMBER() OVER (PARTITION BY Customer\_Id ORDER BY total\_quantity\_sold DESC) AS row\_num

FROM CustomerTrackSales

)

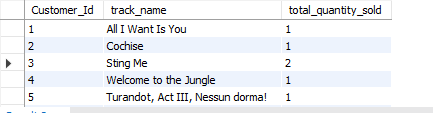
SELECT Customer\_Id, track\_name,total\_quantity\_sold

FROM RankedTracks

WHERE row\_num = 1

ORDER BY customer\_id asc;

**OUTPUT:**

****

**OBSERVATION:**

* The majority of customers have a total quantity sold of 1, meaning they bought different tracks in equal amounts.
* Only a few customers (e.g., ID 3, 10, 12, etc.) have purchased a track more than once.
* No customer has bought the same track more than 2 times
* No single track dominates the list, meaning customer preferences are quite diverse.

1. Are there any patterns or trends in customer purchasing behavior (e.g., frequency of purchases-using invoice id, preferred payment methods, average order value-calculate avg order value)? Monthly purchasing trend, top revenue generating customers

**QUERY:**

WITH PurchaseFrequency AS (

SELECT customer\_id, COUNT(invoice\_id) AS total\_purchases

FROM invoice

GROUP BY customer\_id

),

AvgOrderValue AS (

SELECT customer\_id, ROUND(AVG(total), 2) AS avg\_order\_value

FROM invoice

GROUP BY customer\_id

),

MonthlyTrends AS (

SELECT DATE\_FORMAT(invoice\_date, '%Y-%m') AS purchase\_month,

COUNT(invoice\_id) AS num\_purchases,

SUM(total) AS total\_revenue

FROM invoice

GROUP BY purchase\_month

),

TopRevenueCustomers AS (

SELECT customer\_id, SUM(total) AS total\_revenue

FROM invoice

GROUP BY customer\_id

ORDER BY total\_revenue DESC

LIMIT 10

)

SELECT pf.customer\_id,pf.total\_purchases,aov.avg\_order\_value,

trc.total\_revenue

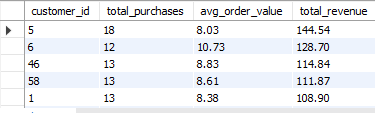
FROM PurchaseFrequency pf

JOIN AvgOrderValue aov ON pf.customer\_id = aov.customer\_id

JOIN TopRevenueCustomers trc ON pf.customer\_id = trc.customer\_id

ORDER BY trc.total\_revenue DESC;

**OUTPUT:**



**OBSERVATION:**

* Frequent buyers spend less per order. Customers 5 and 13 purchase often but have a lower average order value (AOV).
* High AOV customers buy less often. Customers 3 and 6 make fewer purchases but spend more per order.
* Top customers generate most revenue. A few high-value customers contribute a large share of total sales.
* A small group of customers drive sales. Retaining and rewarding them is important for business growth.

1. What is the customer churn rate?

**QUERY:**

WITH LastPurchase AS (

SELECT customer\_id, MAX(invoice\_date) AS last\_purchase\_date

FROM invoice

GROUP BY customer\_id

),

ChurnedCustomers AS (

SELECT COUNT(customer\_id) AS churned\_customers

FROM LastPurchase

WHERE last\_purchase\_date < DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 1 YEAR)

),

TotalCustomers AS (

SELECT COUNT(customer\_id) AS total\_customers FROM customer

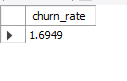
)

SELECT

(c.churned\_customers / t.total\_customers) \* 100 AS churn\_rate

FROM ChurnedCustomers c, TotalCustomers t;

**OUTPUT:**

****

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

**Query to Calculate the Percentage of Total Sales by each Genre in the USA**

WITH GenreSales AS (

SELECT g.genre\_id,g.name AS genre\_name,

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

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

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

WHERE c.country = 'USA'

GROUP BY g.genre\_id, g.name

),

TotalSales AS (

SELECT SUM(total\_sales) AS overall\_sales FROM GenreSales

)

SELECT gs.genre\_name, gs.total\_sales,

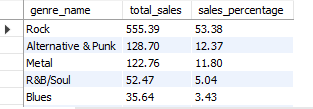
round((gs.total\_sales / ts.overall\_sales) \* 100,2) AS sales\_percentage

FROM GenreSales gs

CROSS JOIN TotalSales ts

ORDER BY gs.total\_sales DESC;

**OUTPUT:**

****

**VISUALIZATION:**

**OBSERVATION:**

* Prioritize Rock – Continue marketing, promoting, and bundling Rock albums as it the top selling genre.
* Leverage Niche Markets – Push Alternative, Punk, and Metal with targeted campaigns.
* Boost Low-Selling Genres – Explore pricing strategies or promotional playlists for minor genres.

**Query to find best-selling artist in USA**

WITH ArtistSales AS (

SELECT ar.artist\_id, ar.name AS artist\_name,

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

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

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

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

WHERE c.country = 'USA'

GROUP BY ar.artist\_id, ar.name

)

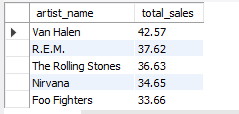
SELECT artist\_name, total\_sales

FROM ArtistSales

ORDER BY total\_sales DESC

LIMIT 5;

**OUTPUT:**



**VISUALIZATION:**

**OBSERVATION:**

* Van Halen has the highest sales contribution, meaning their albums/tracks are the most purchased.
* Both R.E.M. and The Rolling Stones contribute significantly to sales.
* Investigate Nirvana & Foo Fighters Sales – Analyze customer engagement, pricing, and demand.

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

**QUERY:**

WITH CustomerGenreCount AS (

SELECT c.customer\_id,

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

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

FROM customer c

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

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

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

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

GROUP BY c.customer\_id, customer\_name

)

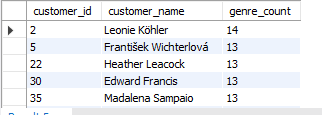
SELECT customer\_id, customer\_name, genre\_count

FROM CustomerGenreCount

WHERE genre\_count >= 3

ORDER BY genre\_count DESC, customer\_id;

**OUTPUT**:



**OBSERVATION:**

* The highest number of genres purchased by a customer is 14 (by Leonie Köhler).
* Many customers have purchased from 10+ different genres, indicating a diverse taste in music.
* Most customers have purchased from 8+ genres, meaning they don’t stick to a single genre.
* Genre count 13 is shared by 5 customers (František Wichterlová, Terhi Hämäläinen, etc.).
* Genre count 12 is shared by 7 customers.

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

**QUERY:**

WITH GenreSales AS (

SELECT g.name AS genre, SUM(il.unit\_price \* il.quantity) AS total\_sales

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

WHERE i.billing\_country = 'USA'

GROUP BY g.name

)

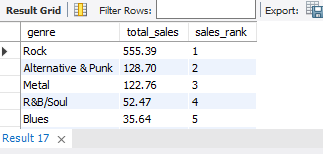
SELECT genre, total\_sales,

RANK() OVER (ORDER BY total\_sales DESC) AS sales\_rank

FROM GenreSales

ORDER BY total\_sales DESC;

**OUTPUT:**



**VISUALIZATION:**

**OBSERVATION:**

* The Rock genre has significantly higher total sales compared to other genres, making it the most popular genre in the USA.
* Target Rock and Metal Audiences – Since these genres are driving sales, the business should focus on promoting and expanding their Rock and Metal offerings.
* More marketing efforts can be made for genres like Hip Hop/Rap and Latin to increase their visibility.
* Since TV Shows and Electronica/Dance have lower sales, bundle deals or promotions could boost their performance.

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

**QUERY:**

WITH LastPurchase AS (

SELECT customer\_id, MAX(invoice\_date) AS last\_purchase\_date

FROM invoice

GROUP BY customer\_id

)

SELECT c.customer\_id, concat(c.first\_name," ",c.last\_name) as cutomer\_name, lp.last\_purchase\_date

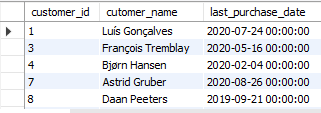
FROM customer c

LEFT JOIN LastPurchase lp ON c.customer\_id = lp.customer\_id

WHERE lp.last\_purchase\_date < DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 3 MONTH)

ORDER BY customer\_id ASC;

**OUTPUT:**

****

**OBSERVATION:**

* Some customers haven’t purchased since 2019, while many stopped in early-to-mid 2020, indicating a drop-off trend.
* Market shifts, competition, pandemic impact, or lack of engagement efforts may have led to inactivity.
* Customers inactive since late 2020 (e.g., Astrid Gruber, Manoj Pareek) might return with targeted promotions.
* Offering discounts, exclusive content, or personalized recommendations could help regain lost customers.

**SUBJECTIVES**

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,genrename,albumname and sum(itotal as totalsales)

**QUERY:**

SELECT g.name AS genre\_name,a.title AS album\_name,

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

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

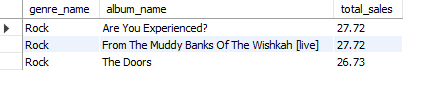
WHERE i.billing\_country = 'USA'

GROUP BY g.name, a.title

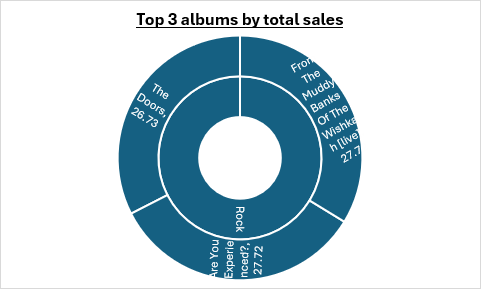
ORDER BY total\_sales DESC

LIMIT 3;

**OUTPUT:**



**VISUALIZATION:**



**INSIGHTS:**

* Rock is the dominant genre in the USA
* All top 3 albums belong to the Rock genre, indicating that Rock music has the highest sales in the US market.
* This suggests a strong customer preference for Rock albums, making it a priority for marketing efforts.
* Since Rock already has a high demand, strategic advertising and promotions can boost these sales further.

**RECOMMENDATIONS:**

* Since Rock is the best-selling genre, marketing should heavily target Rock listeners.
* Use social media ads, influencer collaborations, and genre-based promotions on platforms like Spotify and Apple Music.
* Target existing Rock music fans and past buyers of similar albums.
* Sponsored promotions on music streaming apps will increase visibility and sales.
* Run Instagram & Facebook ads targeting Rock fans in the USA.

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

**QUERY:**

WITH Genre\_Sales AS (

SELECT c.country, g.name AS genre\_name,

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

FROM invoice\_line il

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

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

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

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

WHERE c.country != 'USA'

GROUP BY c.country, g.name

),

Ranked\_Genres AS (

SELECT country, genre\_name, total\_sales,

RANK() OVER (PARTITION BY country ORDER BY total\_sales DESC) AS genre\_rank

FROM Genre\_Sales

)

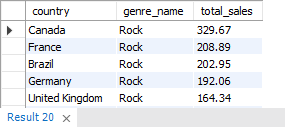
SELECT country, genre\_name, total\_sales

FROM Ranked\_Genres

WHERE genre\_rank = 1

ORDER BY total\_sales DESC;

**OUTPUT:**



**VISUALIZATION:**

**INSIGHTS:**

**Commonalities Across Countries:**

* The top-selling genre in every country except Argentina is Rock.
* This suggests a global appeal for Rock music, spanning continents and cultural backgrounds.
* Canada, France, Germany, the UK, and the Czech Republic show particularly high sales in Rock, indicating a consistent preference in North America and Europe.
* India and Portugal have notable Rock sales, which is interesting given their diverse musical influences.

**Differences Between Countries:**

* Argentina is the Only Country Where Rock is Not the Top-Selling Genre
* Countries like Hungary, Austria, and Poland have relatively smaller Rock sales numbers but still follow the global pattern of Rock dominance.

**RECOMMENDATIONS:**

* Since Alternative & Punk outperforms Rock in Argentina, music stores and streaming platforms should focus on promoting Alternative & Punk artists rather than just Rock.
* Since Canada and the UK are the highest Rock-consuming markets, labels should prioritize early album releases, premium concert access, and exclusive content for these regions.
* Encouraging collaborations between Rock artists and regional musicians can appeal to wider audiences.

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?

**QUERY:**

WITH CustomerCategory AS (

SELECT customer\_id, MIN(invoice\_date) AS first\_purchase\_date,

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

CASE

WHEN MIN(invoice\_date) >= DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 6 MONTH)

THEN 'New Customer'

ELSE 'Long-Term Customer'

END AS customer\_type

FROM invoice

GROUP BY customer\_id

),

PurchaseStats AS (

SELECT i.customer\_id, COUNT(i.invoice\_id) AS total\_purchases,

SUM(i.total) AS total\_spent, AVG(i.total) AS avg\_order\_value,

COUNT(il.track\_id) / COUNT(i.invoice\_id) AS avg\_basket\_size

FROM invoice i

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

GROUP BY i.customer\_id

)

SELECT cc.customer\_type, ROUND(AVG(ps.total\_purchases), 2) AS avg\_purchases,

ROUND(AVG(ps.total\_spent), 2) AS avg\_spent,

ROUND(AVG(ps.avg\_order\_value), 2) AS avg\_order\_value,

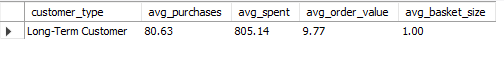
ROUND(AVG(ps.avg\_basket\_size), 2) AS avg\_basket\_size

FROM CustomerCategory cc

JOIN PurchaseStats ps ON cc.customer\_id = ps.customer\_id

GROUP BY cc.customer\_type;

**OUTPUT:**



**VISUALIZATION:**

**INSIGHTS:**

* Long-term customers buy frequently (80+ purchases)
* They spend a lot over time ($805+ total)
* They buy only one item per order (basket size = 1).
* Their average order value is stable ($9.77).
* No new customers in the data.

**RECOMMENDATION:**

* Offer loyalty rewards and exclusive discounts for long-term customers
* Suggest subscription models for recurring purchases.
* Promote bundles and upsell related products.
* Improve New Customer Acquisition: Offer first-time purchase discounts and run targeted ads or email campaigns.

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?

**QUERY:**

WITH

-- Find frequently purchased genre pairs

GenrePairs AS (

SELECT t1.genre\_id AS genre\_1,t2.genre\_id AS genre\_2,

COUNT(\*) AS frequency

FROM invoice\_line il1

JOIN invoice\_line il2

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

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

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

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

WHERE t1.genre\_id < t2.genre\_id -- Avoid duplicate pairs

GROUP BY t1.genre\_id, t2.genre\_id

),

-- Find frequently purchased artist pairs (Corrected)

ArtistPairs AS (

SELECT al1.artist\_id AS artist\_1,al2.artist\_id AS artist\_2,

COUNT(\*) AS frequency

FROM invoice\_line il1

JOIN invoice\_line il2

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

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

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

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

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

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

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

GROUP BY al1.artist\_id, al2.artist\_id

),

-- Find frequently purchased album pairs

AlbumPairs AS (

SELECT t1.album\_id AS album\_1,t2.album\_id AS album\_2,

COUNT(\*) AS frequency

FROM invoice\_line il1

JOIN invoice\_line il2

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

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

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

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

WHERE t1.album\_id < t2.album\_id

GROUP BY t1.album\_id, t2.album\_id

)

-- Final selection combining results

SELECT \* FROM (

-- Top Genre Pairs

SELECT 'Genre' AS category, g1.name AS item\_1,

g2.name AS item\_2, gp.frequency

FROM GenrePairs gp

JOIN genre g1 ON gp.genre\_1 = g1.genre\_id

JOIN genre g2 ON gp.genre\_2 = g2.genre\_id

ORDER BY gp.frequency DESC

LIMIT 5

) AS GenreResults

UNION ALL

SELECT \* FROM (

-- Top Artist Pairs

SELECT 'Artist' AS category,a1.name AS item\_1,a2.name AS item\_2,

ap.frequency

FROM ArtistPairs ap

JOIN artist a1 ON ap.artist\_1 = a1.artist\_id

JOIN artist a2 ON ap.artist\_2 = a2.artist\_id

ORDER BY ap.frequency DESC

LIMIT 5

) AS ArtistResults

UNION ALL

SELECT \* FROM (

-- Top Album Pairs

SELECT 'Album' AS category,al1.title AS item\_1,al2.title AS item\_2,

ap.frequency

FROM AlbumPairs ap

JOIN album al1 ON ap.album\_1 = al1.album\_id

JOIN album al2 ON ap.album\_2 = al2.album\_id

ORDER BY ap.frequency DESC

LIMIT 5

) AS AlbumResults;

**OUTPUT:**



**VISUALIZATION:**

**INSIGHTS:**

* The most frequently purchased genres together include Rock and Metal (1622 purchases), Rock and Alternative & Punk (1056 purchases), and Rock and Latin (427 purchases).
* Metal & Alternative & Punk (315 purchases) frequently appear together, indicating that these audiences have overlapping preferences.
* Led Zeppelin & Green Day (24 purchases), Green Day & Foo Fighters (20 purchases), and Eric Clapton & Nirvana (19 purchases) suggest that classic rock, grunge, and punk fans have shared interests.
* Metallica & Green Day (18 purchases) shows that even heavier rock fans sometimes cross over into punk music.

**Product recommendations and cross-selling initiatives:**

* Customers who purchase Rock albums (e.g., Led Zeppelin, Foo Fighters) can be recommended Metal albums (e.g., Metallica, Iron Maiden) based on frequent co-purchases.
* Fans of Green Day can be recommended Nirvana and Foo Fighters, as they have strong cross-buying behavior.
* If a customer streams or buys a Metal album, suggest Alternative & Punk playlists since these genres frequently cross over.
* Offer discounts on albums frequently purchased together

1. 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?

**QUERY:**

WITH CustomerActivity AS (

-- Determine last purchase date per customer

SELECT c.customer\_id,c.country,COUNT(i.invoice\_id) AS total\_purchases,SUM(i.total) AS total\_spent,MAX(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.country

),

ChurnedCustomers AS (

-- Identify customers who have not purchased in the last 12 months from the latest invoice date

SELECT country, COUNT(customer\_id) AS churned\_customers

FROM CustomerActivity

WHERE last\_purchase\_date < DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 12 MONTH)

GROUP BY country

)

-- Final aggregation: total customers, active customers, and churn rate

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

SUM(ca.total\_purchases) AS total\_transactions,

ROUND(AVG(ca.total\_spent), 2) AS avg\_spending\_per\_customer,

COALESCE(cc.churned\_customers, 0) AS churned\_customers,

ROUND((COALESCE(cc.churned\_customers, 0) / NULLIF(COUNT(ca.customer\_id), 0)) \* 100, 2) AS churn\_rate\_percentage

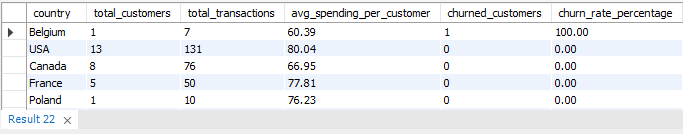
FROM CustomerActivity ca

LEFT JOIN ChurnedCustomers cc ON ca.country = cc.country

GROUP BY ca.country, cc.churned\_customers

ORDER BY churn\_rate\_percentage DESC;

**OUTPUT:**



**VISUALIZATION:**

**INSIGHTS:**

* Very small customer base (only 1 customer) in Belgium, meaning a single churn skews the data.
* USA & Canada: Largest Customer Bases, No Churn
* Countries like France (5 customers), Germany (4 customers), and the UK (3 customers) also show zero churn, indicating stable engagement.
* Countries like Hungary, Austria, Netherlands, Sweden, Finland, etc., all have only 1-2 customers but report zero churn.
* Countries like India, Brazil, Portugal have a higher average spending per customer but still maintain zero churn.

**RECOMMENDATIONS:**

* Localized marketing campaigns or exclusive Belgian artist features could improve retention.
* Invest in local artist partnerships, regionalized playlists, and targeted email marketing.
* Offer bundled purchases or subscription discounts.
* Run targeted promotions to increase awareness and engagement.

1. 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?

**QUERY:**

WITH CustomerActivity AS (

-- Determine each customer's total purchases, total amount spent, and last purchase date

SELECT c.customer\_id, c.country,

COUNT(i.invoice\_id) AS total\_purchases, SUM(i.total) AS total\_spent,

MAX(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.country

),

ChurnRisk AS (

-- Categorize customers into high-risk, medium-risk, and low-risk based on purchase activity

SELECT customer\_id, country, total\_purchases, total\_spent,

last\_purchase\_date,

CASE

WHEN last\_purchase\_date < DATE\_SUB((SELECT MAX(invoice\_date) FROM invoice), INTERVAL 12 MONTH)

THEN 'High Risk' -- No purchase in the last 12 months

WHEN total\_purchases <= 3 OR total\_spent < 50

THEN 'Medium Risk' -- Low purchase frequency or spending

ELSE 'Low Risk' -- Regular customers

END AS risk\_category

FROM CustomerActivity

)

-- Final aggregation: Risk distribution by country

SELECT country,

SUM(CASE WHEN risk\_category = 'High Risk' THEN 1 ELSE 0 END) AS high\_risk\_customers,

SUM(CASE WHEN risk\_category = 'Medium Risk' THEN 1 ELSE 0 END) AS medium\_risk\_customers,

SUM(CASE WHEN risk\_category = 'Low Risk' THEN 1 ELSE 0 END) AS low\_risk\_customers,

COUNT(customer\_id) AS total\_customers,

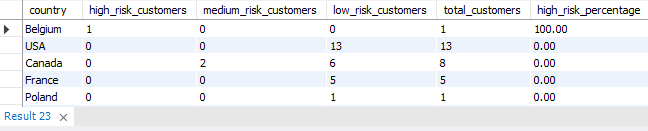
ROUND((SUM(CASE WHEN risk\_category = 'High Risk' THEN 1 ELSE 0 END) / COUNT(customer\_id)) \* 100, 2) AS high\_risk\_percentage

FROM ChurnRisk

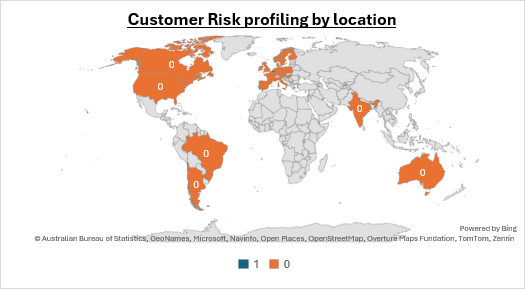
GROUP BY country

ORDER BY high\_risk\_percentage DESC;

**OUTPUT:**



**VISUALIZATION:**



**INSIGHTS:**

* Belgium (100% churn risk) - Only one customer in Belgium, and they haven’t purchased in over a year
* Denmark (1 customer in Medium Risk),Argentina (1 customer in Medium Risk),Canada (2 customers in Medium Risk).These customers have either low purchase frequency or low spending (< $50 total).
* Countries with only one or two customers (e.g., Belgium, Denmark, Argentina) are more volatile. If these customers stop buying, the churn rate spikes.
* Belgium has 100% churn risk – possibly due to a lack of localization, pricing issues, or limited product variety.

**RECOMMENDATIONS:**

***For High-Risk Customers:***

* Re-engage inactive customers with personalized offers, discounts, or email reminders.
* Expand marketing efforts in low-performing regions like Belgium, Denmark, and Argentina.
* Improve customer experience by providing localized content, pricing, and recommendations.

***For Medium-Risk Customers:***

* Encourage higher spending through bundles, discounts, or loyalty programs.
* Analyze customer preferences to recommend relevant music genres or albums.

***For Low-Risk Customers:***

* Upsell and cross-sell based on their purchase history.
* Encourage referrals to acquire more customers in strong-performing regions.

1. 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?

**QUERY:**

WITH CustomerActivity AS (

SELECT c.customer\_id, c.country,

COUNT(i.invoice\_id) AS total\_purchases, SUM(i.total) AS total\_spent,

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

MAX(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.country

),

ChurnedCustomers AS (

SELECT country,

COUNT(customer\_id) AS churned\_customers

FROM CustomerActivity

WHERE last\_purchase\_date < (SELECT MAX(invoice\_date) FROM invoice) - INTERVAL 12 MONTH

GROUP BY country

),

LTV\_Calculation AS (

SELECT ca.country,

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

SUM(ca.total\_spent) / COUNT(ca.customer\_id) AS avg\_revenue\_per\_customer,

SUM(IFNULL(cc.churned\_customers, 0)) AS churned\_customers, -- FIXED: Used SUM() for aggregation

ROUND(1 - (SUM(IFNULL(cc.churned\_customers, 0)) / COUNT(ca.customer\_id)), 2) AS retention\_rate,

IF(ROUND(1 - (SUM(IFNULL(cc.churned\_customers, 0)) / COUNT(ca.customer\_id)), 2) < 1,

1 / (1 - ROUND(1 - (SUM(IFNULL(cc.churned\_customers, 0)) / COUNT(ca.customer\_id)), 2)),

10) AS estimated\_customer\_lifetime

FROM CustomerActivity ca

LEFT JOIN ChurnedCustomers cc ON ca.country = cc.country

GROUP BY ca.country

)

SELECT country, total\_customers,

ROUND(avg\_revenue\_per\_customer, 2) AS avg\_revenue\_per\_customer,

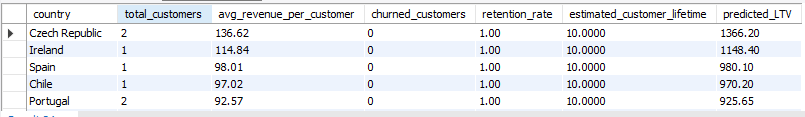
churned\_customers, retention\_rate,estimated\_customer\_lifetime,

ROUND(avg\_revenue\_per\_customer \* estimated\_customer\_lifetime, 2) AS predicted\_LTV

FROM LTV\_Calculation

ORDER BY predicted\_LTV DESC;

**OUTPUT:**



**VISUALIZATION:**

**INSIGHTS:**

* Countries with higher spending customers (avg revenue > $100) tend to retain customers better.
* Countries with small customer bases (1 or 2 customers) are at higher risk of 100% churn.
* Higher spending per customer correlates with longer estimated lifetimes and lower churn.
* USA, Canada, and Brazil have multiple customers and 100% retention, meaning larger customer bases may provide better stability.

**RECOMMENDATIONS:**

* Loyalty & Engagement – Offer discounts, personalized promotions, and re-engagement campaigns for inactive customers.
* Pricing & Upselling – Bundle products, upsell premium options, and adjust pricing for lower-spending regions.
* Customer Experience – Improve onboarding, provide proactive support, and collect feedback for better retention.
* Market Expansion – Invest in ads, referral programs, and subscription models to grow the customer base.

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

**INSIGHTS:**

* Track new vs. existing customers. Compare customer acquisition before and after promotions to see if new buyers increased.
* Measure purchase frequency. Check if existing customers buy more often after receiving discounts or marketing emails.
* Analyze revenue impact. Compare total sales before, during, and after campaigns to see if discounts drive higher revenue.
* Customer retention trends. Track repeat purchases from customers who engaged with past promotions.
* Compare different campaigns. Identify which promotions (discounts, emails, events) have the highest impact on sales and retention.

**RECOMMENDATION:**

* Segment customers for targeted promotions. Offer discounts to first-time buyers and loyalty rewards to repeat customers.
* Optimize email marketing. Personalize emails based on customer behavior to increase engagement.
* Measure long-term impact. Ensure discounts drive real customer retention, not just one-time purchases.
* Use time-based analysis. Compare sales before, during, and after promotions to assess effectiveness.

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

* **Understanding the Business Context** – Identifying the industry, business goals, and challenges to determine how data science can provide insights.
* **Analyzing the Data** – Exploring available datasets, by checking for missing values, and performing exploratory analysis to find trends and correlations.
* **Defining Key Metrics** – Establishing important KPIs like Customer Lifetime Value (LTV), Churn Rate, Retention Rate, and Average Revenue Per Customer.
* **Identifying Use Cases** – Segmenting customers, analyzing churn risk, and assessing the impact of promotions on retention and sales.
* **Developing a Data-Driven Approach** – Using SQL queries and comparing trends across segments.
* **Measuring and Validating** – Evaluating insights, comparing actual vs. predicted outcomes, and providing actionable recommendations.

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?

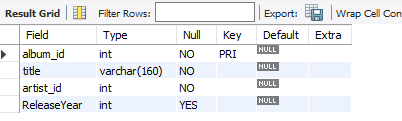
**QUERY:**

ALTER TABLE Album

ADD COLUMN ReleaseYear INTEGER;

DESC Album;

**OUTPUT:**



1. 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.

**QUERY:**

SELECT c.country,

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

ROUND(AVG(totals.total\_amount), 2) AS avg\_amount\_spent\_per\_customer,

ROUND(AVG(totals.total\_tracks), 2) AS avg\_tracks\_purchased\_per\_customer

FROM customer c

LEFT JOIN (

-- Calculating total amount spent & total tracks purchased per customer

SELECT i.customer\_id, SUM(i.total) AS total\_amount,

SUM(il.quantity) AS total\_tracks

FROM invoice i

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

GROUP BY i.customer\_id

) AS totals ON c.customer\_id = totals.customer\_id

GROUP BY c.country

ORDER BY total\_customers DESC;

**OUTPUT:**

