

**Objective**

* To help the music store analyse key trends like Customer demographics, purchase behaviour analysis, churn rate, etc. in order to achieve sustainable growth the near-to-mid-term
* As a data analyst at Chinook, the objective is to analyze music record sales data to gain insights and make recommendations for the company's strategy in the physical music market.

**Question Types**

* Objective Questions: Identifies key customer trends at the chinook music store
* Subjective Questions: Deeper analysis to suggest / make recommendations regarding strategies to chart out future growth

**Data description (tables and its columns)**

1. customer:
   * customer\_id: Unique identifier assigned to each customer.
   * first\_name: The given name or first name of a customer.
   * last\_name: The surname or family name of a customer.
   * company: The name of the company associated with a customer.
   * address: The street address of a customer's location.
   * city: The city where a customer is located.
   * state: The state or province where a customer is located.
   * country: The country where a customer is located.
   * postal\_code: The postal or zip code of a customer's address.
   * phone: The phone number of a customer.
   * fax: The fax number associated with a customer.
   * email: The email address of a customer.
   * support\_rep\_id: The employee ID of the support representative assigned to a customer.
2. employee:
   * employee\_id: Unique identifier assigned to each employee.
   * last\_name: The surname or family name of a employee.
   * first\_name: The given name or first name of a employee.
   * Title: position of the employee within the company
   * reports\_to: employee\_id of the manager
   * birthdate: date of birth of the employee
   * hire\_date: date of employment
   * address: The street address of an employee’s location
   * city: The city where an employee is located.
   * state: The state or province where an employee is located
   * country: The country where an employee is located
   * postal\_code: The postal or zip code of an employee’s address
   * phone: the phone number of an employee
   * fax: the fax number associated with an employee
   * email: The email address of an employee
3. invoice:
   * invoice\_id: Unique identifier assigned to each invoice.
   * customer\_id: The customer ID associated with the invoice.
   * invoice\_date: The date when the invoice was generated or issued.
   * billing\_address: The street address used for billing purposes.
   * billing\_city: The city used for billing purposes.
   * billing\_state: The state or province used for billing purposes.
   * billing\_country: The country used for billing purposes.
   * billing\_postal\_code: The postal or zip code used for billing purposes.
   * total: The total amount due on the invoice.
4. invoice\_line**:**
   * invoice\_line\_id: Unique identifier assigned to each line item on an invoice.
   * invoice\_id: The invoice ID to which the line item belongs.
   * track\_id: The ID of the track or product included in the line item.
   * unit\_price: The price per unit for the line item.
   * quantity: The quantity of units for the line item.
5. playlist:
   * playlist\_id: Unique identifier assigned to each playlist.
   * name: The name or title of the playlist.
6. playlist\_track**:** 
   * playlist\_id: The ID of the playlist to which the track belongs.
   * track\_id: The ID of the track included in the playlist.
7. track:
   * track\_id: Unique identifier assigned to each track or song.
   * name: The title or name of the track.
   * album\_id: The ID of the album to which the track belongs.
   * media\_type\_id: The ID of the media type associated with the track.
   * genre\_id: The ID of the genre associated with the track.
   * composer: The name of the composer or artist who composed the track.
   * milliseconds: The duration of the track in milliseconds.
   * bytes: The file size of the track in bytes.
   * unit\_price: The price per unit for the track.
8. album:
   * album\_id: Unique identifier assigned to each album.
   * title: The title or name of the album.
   * artist\_id: The ID of the artist associated with the album.
9. artist:
   * artist\_id: Unique identifier assigned to each artist.
   * name: The name of the artist.
10. media\_type:
    * media\_type\_id: Unique identifier assigned to each media type.
    * name: The name or description of the media type (e.g., MPEG audio file, AAC audio file).
11. genre:
    * genre\_id: Unique identifier assigned to each genre.
    * name: The name or description of the genre (e.g., rock, pop, classical).



* Before we start on working on the Chinook Music store database, we will create the said database using the provided .sql file by copying the code and running all the queries.
  + Extra spaces in the database query in chinook.sql files were removed
  + Foreign key constraints were added to join tables and create ERD diagram on MySQL workbench

-- Adding foreign key constraints

ALTER TABLE `album`

ADD CONSTRAINT `fk\_album\_artist`

FOREIGN KEY (`artist\_id`) REFERENCES `artist`(`artist\_id`);

ALTER TABLE `customer`

ADD CONSTRAINT `fk\_customer\_support\_rep`

FOREIGN KEY (`support\_rep\_id`) REFERENCES `employee`(`employee\_id`);

ALTER TABLE `invoice`

ADD CONSTRAINT `fk\_invoice\_customer`

FOREIGN KEY (`customer\_id`) REFERENCES `customer`(`customer\_id`);

ALTER TABLE `invoice\_line`

ADD CONSTRAINT `fk\_invoice\_line\_invoice`

FOREIGN KEY (`invoice\_id`) REFERENCES `invoice`(`invoice\_id`),

ADD CONSTRAINT `fk\_invoice\_line\_track`

FOREIGN KEY (`track\_id`) REFERENCES `track`(`track\_id`);

ALTER TABLE `track`

ADD CONSTRAINT `fk\_track\_album`

FOREIGN KEY (`album\_id`) REFERENCES `album`(`album\_id`),

ADD CONSTRAINT `fk\_track\_media\_type`

FOREIGN KEY (`media\_type\_id`) REFERENCES `media\_type`(`media\_type\_id`),

ADD CONSTRAINT `fk\_track\_genre`

FOREIGN KEY (`genre\_id`) REFERENCES `genre`(`genre\_id`);

ALTER TABLE `playlist\_track`

ADD CONSTRAINT `fk\_playlist\_track\_playlist`

FOREIGN KEY (`playlist\_id`) REFERENCES `playlist`(`playlist\_id`),

ADD CONSTRAINT `fk\_playlist\_track\_track`

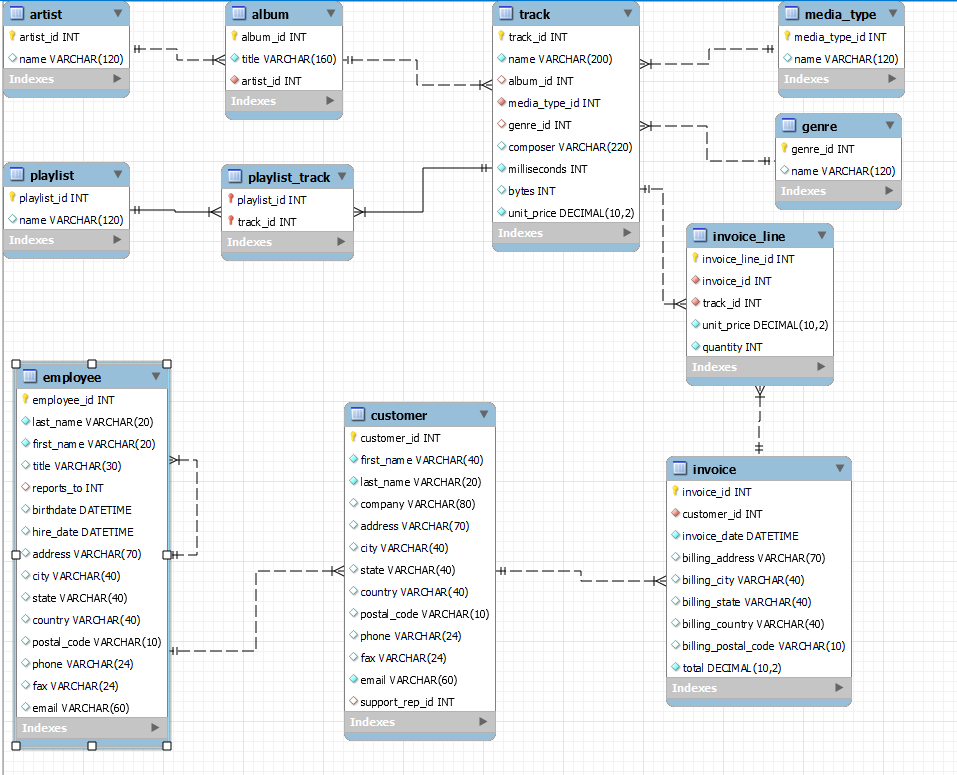
FOREIGN KEY (`track\_id`) REFERENCES `track`(`track\_id`);

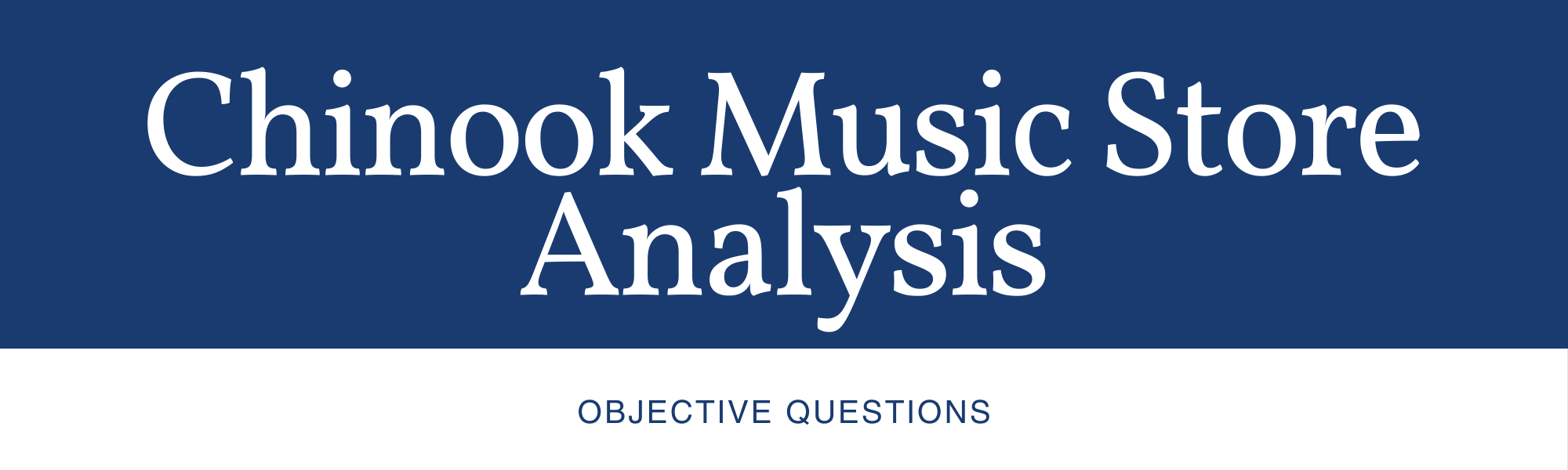
ALTER TABLE `employee`

ADD CONSTRAINT `fk\_employee\_reports\_to`

FOREIGN KEY (`reports\_to`) REFERENCES `employee`(`employee\_id`);

**DATABASE SCHEMA**





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

IDENTIFYING MISSING VALUES

**Approach:** to identify and handle missing values, we can check for missing values (NULLs) in specific columns of a MySQL table.

* + We can use a CASE statement to count the number of Null values (assigned as 1, and 0 if not null) and then do a SUM of these total null values.
  + The output would return number of rows within each of the column having null values.

**Example Input**

**-- To check for missing values in the album table**

SELECT

SUM(CASE WHEN album\_id IS NULL THEN 1 ELSE 0 END) AS Missing\_album\_id,

SUM(CASE WHEN title IS NULL THEN 1 ELSE 0 END) AS Missing\_title,

SUM(CASE WHEN artist\_id IS NULL THEN 1 ELSE 0 END) AS Missing\_artist\_id

FROM album;

**Output (no null values)**

ce277b92dca7f2e903f58f1644ec2ffe.png

**HANDLING MISSING VALUES**

**Approach**: In case a table comes out with a missing value, for example the following output in the customer table

**Output** **(null values)**

f891ca52d14d5f05cf47d8026e9a2b29.png

* + one missing phone number found,
    - **we can replace null value to a string, for example, 'unknown'**
    - For this we would first need to disable the SQL Safe Updates mode, and then disable it once done.
    - SQL Safe Updates mode is a MySQL setting that requires UPDATE and DELETE operations to be performed with a WHERE clause that uses a key column or a LIMIT clause, so as to limit their impact.

**Input**

SET SQL\_SAFE\_UPDATES = 0;

UPDATE customer

SET phone = 'unknown'

WHERE phone IS NULL;

SET SQL\_SAFE\_UPDATES = 1;

**Note: In some cases, handling missing values is not required. For example, one employee In the employee table does not have a manager.**

**Output**

* + d685d1506e5ed9803dccc0ed80a573b5.pngEmployee table has one record with a missing ‘reports\_to’ row. However, that is a General Manager level employee, who must be heading all the other employees. Hence, it may be ignored

**IDENTIFYING AND HANDLING DUPLICATES**

**Approach:** to identify and handle duplicate values, we can create a CTE and use a Window function to identify rows based on a combination of values in different columns of a MySQL table.

* + Later we can filter the results on the basis of any row number that is coming more than once
  + **RESULTS: NO DUPLICATE VALUES WERE FOUND IN ANY OF THE TABLE**

**-- 1. Albums Table**

**-- Identify Duplicates:**

**Input**

WITH CTE\_Albums AS (

SELECT

album\_id,

ROW\_NUMBER() OVER (PARTITION BY title, artist\_id ORDER BY album\_id) AS RowNum

FROM album

)

SELECT \*

FROM CTE\_Albums

WHERE RowNum > 1;

**Output**

92c7e4a639dcd21289d11dc6af991923.png

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

**Approach:** To find the top selling tracks, we can list the top 10 tracks sold in the USA, based on sales in the USA. By associating genre name column in the output, we can also find the top genre associated with these tracks in the USA

* + We can also find the top selling artists and their most famous genres in the USA by calculating the total sales of their tracks
  + This would help us in determining the most popular artists and genres in the USA that the music store can prioritize by means of discounts and strategic bundles to increase their revenue

**Input (top selling tracks)**

SELECT

t.name AS TrackName, ar.name AS ArtistName, g.name AS GenreName,

SUM(il.quantity) AS TotalSales

FROM invoice\_line il

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

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

WHERE i.billing\_country = 'USA'

GROUP BY t.name, ar.name, g.name

ORDER BY TotalSales DESC

LIMIT 10;

**Output**



**Insights:** Top tracks by Sales in USA include War Pigs, You Know I’m No Good (ft. Ghostface Killah), Violent Pornography, etc. Most of these popular tracks are from the ‘Rock’ Genre, but none in top 3.

**Input**

SELECT

a.artist\_id, a.name AS artist\_name, g.name AS genre\_name,

SUM(il.quantity) AS TotalSales

FROM

invoice AS i

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

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

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

JOIN artist AS a ON al.artist\_id = a.artist\_id

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

WHERE

i.billing\_country = 'USA'

GROUP BY

a.artist\_id, a.Name, g.Name

ORDER BY

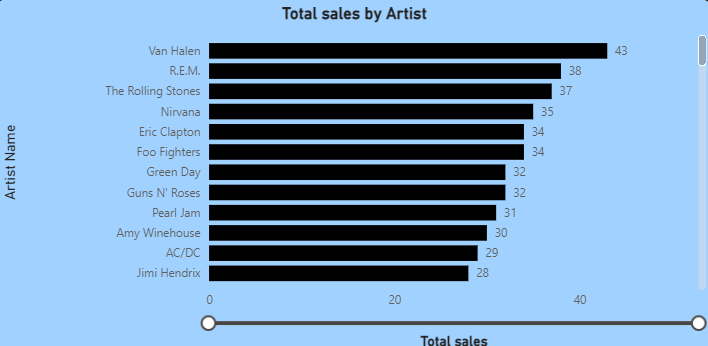
TotalSales DESC

LIMIT 10;

**Output**



**Visualization**



**Insights:** Top-selling artist in the USA are mostly from the ‘Rock’ Genre, including ‘Van Halen’, ‘The Rolling Stones’ & ‘Nirvana’

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

**Approach:** To get an idea about the demographic breakdown from the given data, we can list the total number of customers per country to see the largest markets for the music store

* + This would help us to prioritize larger markets to increase the revenue and also to see which markets could be explored for future growth

**Insights:** Countries with top customers are USA, Canada, Brazil, France & Germany

* + North America is the largest continental market with Europe at second place

**Input**

SELECT c.country, COUNT(\*) AS CustomerCount

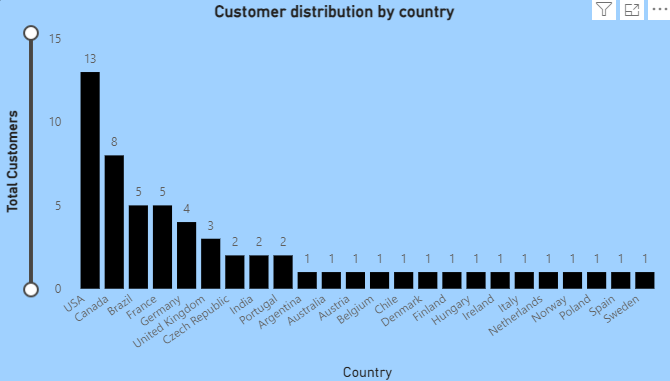
FROM customer AS c

GROUP BY c.country

ORDER BY CustomerCount DESC;

**Output**

**Visualization**

****

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

**Approach:** By getting a breakdown of the total revenue and number of invoices, we can get an idea about the average cost per invoice in each country and city to see where the music store is generating more per customer revenue from

* + This would help us to prioritize larger markets to increase the revenue and also to see which markets could be explored for future growth

**Insights:** Countries with top customers are USA, Canada, Brazil, France & Germany

**Input**

SELECT

i.billing\_country, i.billing\_state, i.billing\_city, SUM(i.total) AS TotalRevenue,

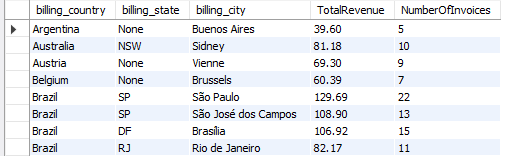
COUNT(i.invoice\_id) AS NumberOfInvoices

FROM invoice i

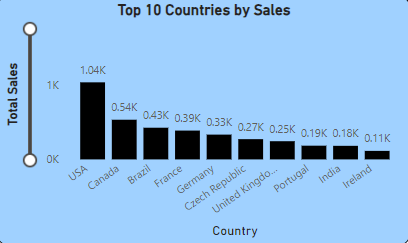
GROUP BY i.billing\_country, i.billing\_state, i.billing\_city

ORDER BY i.billing\_country, SUM(i.total) DESC;

**Output format**



**Visualization**



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

**Approach:** To identify top 5 customers in each country we can rank all the customers by their country and order on the basis of the total revenue that they had generated.

**Input**

WITH CustomerRevenue AS (

SELECT c.customer\_id, c.first\_name, c.last\_name, c.country, SUM(i.total) AS TotalRevenue

FROM customer c

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name, c.country

)

SELECT \*

FROM (

SELECT customer\_id, first\_name, last\_name, country, TotalRevenue,

ROW\_NUMBER() OVER (PARTITION BY country ORDER BY TotalRevenue DESC) AS `rank`

FROM CustomerRevenue

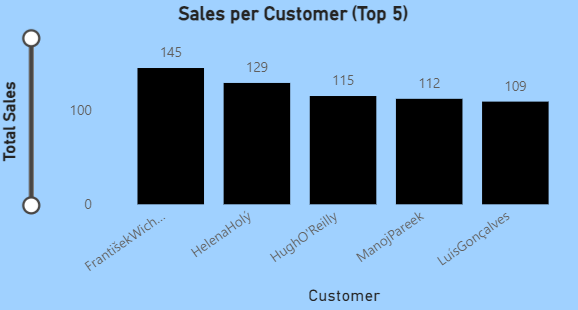
) AS ranked

WHERE `rank`<= 5;

**Output (format)**



**Visualization**

****

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

**Approach:** Since it is likely that most customers would not buy any track more than once, we can calculate the total sales of each track per customer, rank the tracks for each customer based on their total sales and then select the top-ranked track for each customer. This can help the music store by

* + Understanding Customer Preferences to help in tailoring marketing efforts to individual tastes, such as recommending similar tracks or albums.
  + Customer Retention: Knowing what tracks customers prefer can help in creating personalized playlists or special offers, which can enhance customer satisfaction and loyalty.
  + Targeted Marketing: by means of personalized recommendations to customers and to create targeted promotional campaigns, increasing the likelihood of additional purchases.
  + Enhancing User Experience and create competitive advantage.

**Input**

WITH CustomerTrackSales AS (

SELECT

c.customer\_id, (c.first\_name, " ", c.last\_name) AS customer\_name,

t.name AS track\_name, (il.quantity) AS total\_sales

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

GROUP BY

c.customer\_id, c.first\_name, c.last\_name, t.name

),

RankedTracks AS (

SELECT customer\_id, customer\_name, track\_name, total\_sales,

ROW\_NUMBER() OVER (PARTITION BY customer\_id ORDER BY total\_sales DESC) AS `rank`

FROM CustomerTrackSales

)

SELECT

customer\_id, customer\_name, track\_name AS top\_track, total\_sales

FROM RankedTracks

WHERE `rank` = 1

ORDER BY customer\_id;

**Output**



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

**Approach**: To derive trends and patterns from the customer purchases, we can find the frequency of purchases and average purchase value to a country/city level

**Insights**:

* + Different countries show varied Purchase Patterns Across Different Regions. For example, within USA, one of the customer in Boston made 10 purchases with an average order value of $6.63; in Mountain View, a customer made 12 purchases with an average order value of $7.92.
  + Regional Trends: In North America, USA and Canada show significant diversity in purchasing patterns, with some cities having high purchase frequencies but lower average order values (e.g., Edmonton) and others having both high frequencies and high values (e.g., Montréal).
    - In Europe, France, Germany, and the UK have cities where customers tend to have high average order values. For instance, Frankfurt and Paris have notable high average order values.
    - Central Europe (Czech Republic) also shows high purchase frequencies with high average order values.
    - In South America, Brazil shows varied customer behaviors with some cities like São José dos Campos having high average order values and others like São Paulo showing both high and low values.
    - In Asia and Oceania, India and Australia show consistent purchase behaviours with relatively high average order values.

**Input**

SELECT

c.country,

c.city,

c.customer\_id,

COUNT(i.invoice\_id) AS NumberOfPurchases,

ROUND(AVG(i.total), 2) AS AverageOrderValue

FROM

customer c

JOIN

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

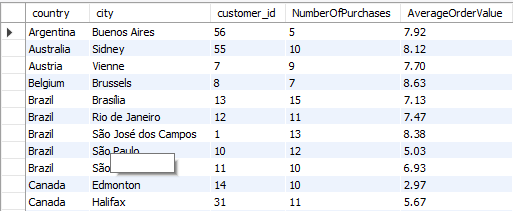
GROUP BY

c.country, c.city, c.customer\_id

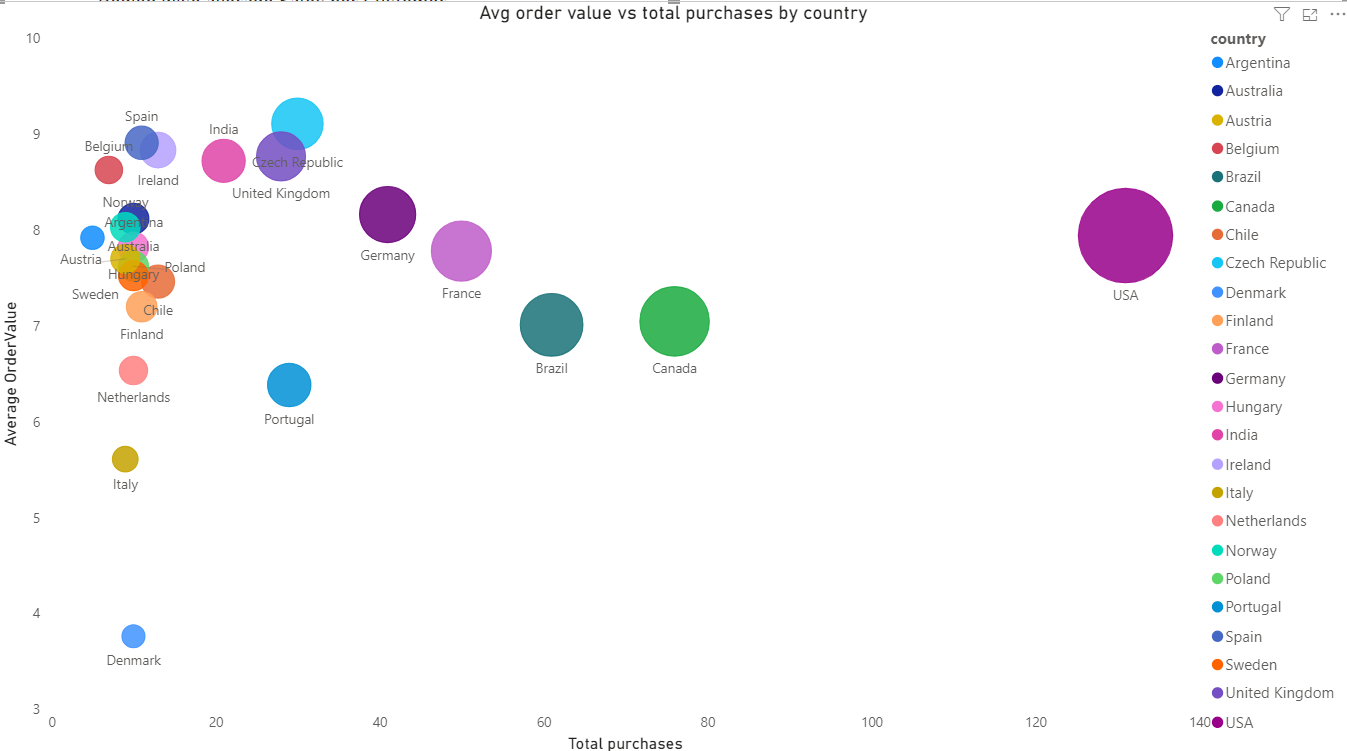
ORDER BY

c.country, c.city, NumberOfPurchases DESC;

**Output format**



**Visualization**



1. **What is the customer churn rate?** 
   * **Approach**: Any customers who has not made any purchase is considered to have been churned i.e. they ceased to have any relationship with the music store.
   * **Insights**: Approximately 27% of the music store’s customers have churned in the past 6 months

**Input**

WITH LastPurchase AS (

SELECT c.customer\_id, 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

),

TotalCustomers AS (

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

FROM customer

),

ChurnedCustomers AS (

SELECT COUNT(DISTINCT customer\_id) AS churned\_customers

FROM LastPurchase

WHERE last\_purchase\_date < DATE\_SUB('2020-12-31', INTERVAL 6 MONTH)

)

SELECT

tc.total\_customers,

cc.churned\_customers,

ROUND((cc.churned\_customers / tc.total\_customers) \* 100,2) AS churned\_percentage

FROM

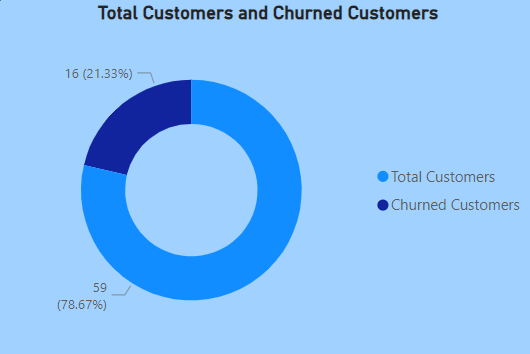
TotalCustomers tc,

ChurnedCustomers cc;

**Output**



**Visualization**



1. **Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.** 
   * **Approach**: We can identify the best-selling genres (in USA) at the music store by finding out the total percentage of sales contributed by each of these genres. Additionally, we can also map the genres to the artists to find best-selling artists.
   * **Insights**: Almost half (~53%) of the sales in the USA are contributed by the rock genre, with Rock artists like Van Halen, The Rolling Stones & Nirvana leading the pack by amount by being the 3 top-selling artists contributing almost ~11% of the sales within the USA.
     + This information would be really helpful for the music store as they can enhance the user-experience of their customers within their largest market by prominently displaying rock genre in their physical store, and also by highlighting this on the main page of their website
     + The store can also accommodate this information into their loyalty programs **(early access to rock music for their premium members)** and to make personalized recommendations.

**Input**

SELECT g.name AS Genre, SUM(il.quantity) AS TotalSales,

ROUND((SUM(il.quantity) \* 100.0 / (SELECT SUM(il2.quantity)

FROM invoice\_line il2

JOIN invoice i2 ON il2.invoice\_id = i2.invoice\_id

WHERE i2.billing\_country = 'USA')),2) AS PercentageOfTotalSales

FROM invoice\_line il

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

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

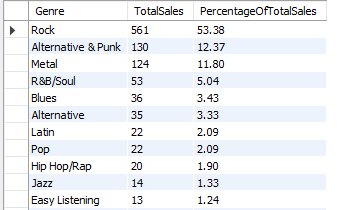
JOIN invoice i ON il.invoice\_id = i.invoice\_id

WHERE i.billing\_country = 'USA'

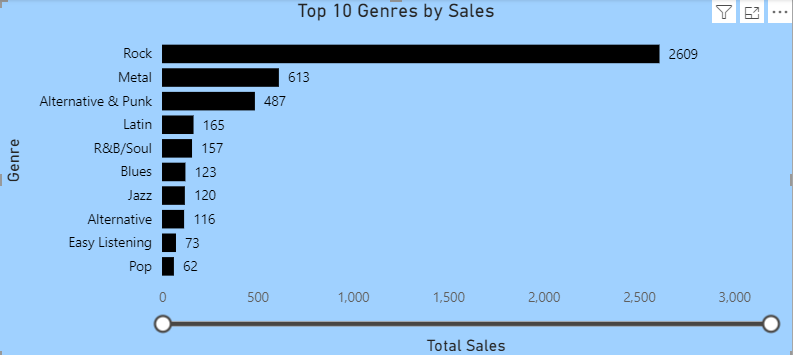
GROUP BY g.name

ORDER BY TotalSales DESC;

**Output format**



**Visualization**



**Input**

WITH GenreSales AS (

SELECT

g.name AS Genre, ar.name AS Artist, SUM(il.quantity) AS TotalSales,

ROUND((SUM(il.quantity) \* 100.0 /

(SELECT SUM(il2.quantity)

FROM invoice\_line il2

JOIN invoice i2 ON il2.invoice\_id = i2.invoice\_id

WHERE i2.billing\_country = 'USA')), 2) AS PercentageOfTotalSales

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

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

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

WHERE i.billing\_country = 'USA'

GROUP BY

g.name, ar.name

)

SELECT

Genre,

Artist,

TotalSales,

PercentageOfTotalSales

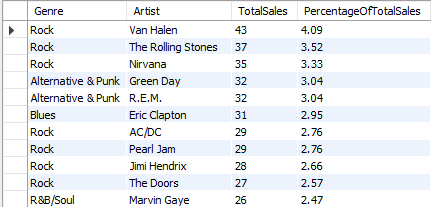
FROM

GenreSales

ORDER BY

TotalSales DESC;

**Output Format**



1. **Find customers who have purchased tracks from at least 3 different genres** 
   * **Approach**: We can identify the customers who have purchased at least 3 different genres of music by counting the distinct genre IDs for each customer
   * **Insights**: All the customers have purchased more than 3 different genres over time suggesting diverse music interests and showing a high level of engagement.
     + Customers with purchase of more than 10 genres can be classified as eclectic listeners and should be given bundling opportunities, i.e. discounts on buying more than albums of different genres and should be given additional rewarding points in the loyalty program.

**Input**

SELECT c.customer\_id, CONCAT(c.first\_name," ", c.last\_name) AS customer\_name, COUNT(DISTINCT g.genre\_id) AS GenreCount

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

HAVING GenreCount >= 3;

**Output format**



1. **Rank genres based on their sales performance in the USA** 
   * **Approach**: We can identify the top selling genres in the USA by by means of their sales
   * **Insights**: Rock is the highest revenue generating genre inside the USA, followed by Alternative & Punk and Metal.

**Input**

SELECT g.name AS Genre, SUM(il.quantity) AS TotalSales,

RANK() OVER (ORDER BY SUM(il.quantity) DESC) AS GenreRank

FROM invoice\_line il

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

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

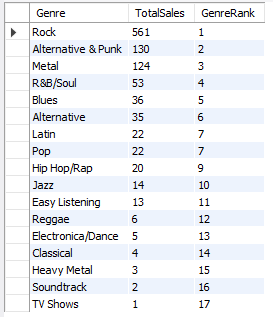
JOIN invoice i ON il.invoice\_id = i.invoice\_id

WHERE i.billing\_country = 'USA'

GROUP BY g.name

ORDER BY GenreRank

**Output**



1. **Identify customers who have not made a purchase in the last 3 months** 
   * **Approach:** Given data for invoices is up until the end of 2020, and hence the last date of 2020 was taken as a measure for evaluation for finding out customers who have not made a purchase in the last 3 months
   * **Insights**: A total of 22 customers, i.e. ~37% of customers has not made any purchase for the past 3 months and can be deemed as ‘inactive’. Such customers should be enticed with offers on buying new albums/music as cost of retaining any customer is lesser than gaining a new one.

**Input**

SELECT c.customer\_id, CONCAT(c.first\_name," ", c.last\_name) AS customer\_name, MAX(i.invoice\_date) AS LastPurchaseDate

FROM customer c

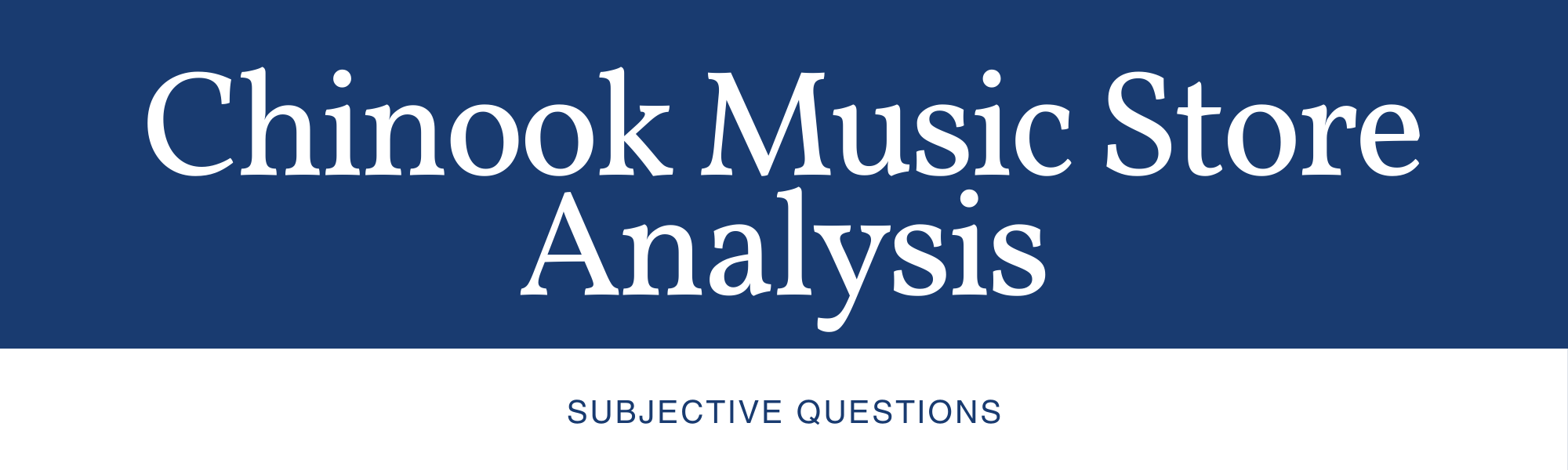
JOIN invoice i ON c.customer\_id = i.customer\_id

GROUP BY c.customer\_id

HAVING LastPurchaseDate < DATE\_SUB(‘2020-12-31’, INTERVAL 3 MONTH);

**Output format**





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.

* **Approach:** Albums sold in the USA are grouped and ranked by their genre and top 3 albums in each genre are listed on the basis of their total sales revenue.
  + **Alternatively, we can also list only the albums sold in the USA on the basis of their total sales revenue**
* **Insights**: Top-Selling Genres and Albums include
  + Rock: Are You Experienced? and From The Muddy Banks Of The Wishkah [live] are the top-selling albums, each with sales of 27.72.
  + R&B/Soul: Seek And Shall Find: More Of The Best (1963-1981) stands out on top with 25.74 in sales.
  + Alternative & Punk: Green leads with 24.75 in sales. American Idiot and Minha História are notable with 19.80 and 18.81 in sales, respectively.
* **Insights**: These albums and genre represents a potential for targeted advertising especially in case of the dominance of Rock genre where a special focus with targeted advertising accompanied with merchandise could yield significant returns.
  + Emerging Genres: Genres like Alternative & Punk and R&B/Soul also show strong sales figures. Campaigns promoting popular albums in these genres could also attract dedicated fanbases and boost sales further in the USA.

**Input**

WITH album\_sales AS (

SELECT

a.title AS album,

g.name AS genre,

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

FROM invoice\_line il

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

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 a.title, g.name

),

ranked\_album\_sales AS (

SELECT album, genre, total\_sales, DENSE\_RANK() OVER (PARTITION BY genre ORDER BY total\_sales DESC) AS `rank`

FROM album\_sales

)

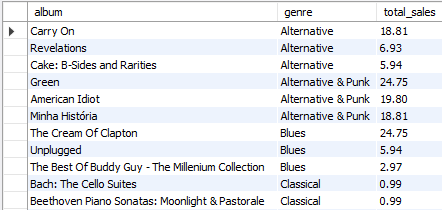
SELECT album, genre, total\_sales

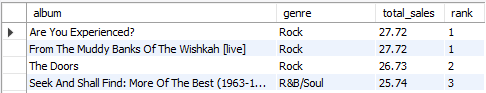
FROM ranked\_album\_sales

WHERE `rank` <= 3

ORDER BY genre, `rank`, album;

**Output (format)**

****

**Output (for just top 3 albums on the basis of their sales)**

**Visualization**



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

**Insights:** Rock is the most common top-selling genre in countries other than the USA, followed by Metal and Alternative & Punk. Easy Listening, Latin & R&B/Soul are the least popular selling genres.

* **Approach:** Top selling genres across countries other than the USA could be identified by finding the total sales revenue generated by each album sold in different countries
* **Insights**: Top-Selling Genres outside the USA include Rock, Metal and Alternative Punk representing approx. 88% of the total sales revenue.
  + Jazz, Blues, Easy Listening, Latin, and R&B/Soul have relatively fewer appearances in the top 3 lists. This indicates that while these genres have dedicated listeners, their appeal may be more localized or niche. A detailed country wise analysis can be carried out to find individual country level preferences of customers.
    - Latin is the second highest sold genre by sales in Ireland and R&B appears in Netherlands only, indicating a potential niche market for Latin music in non-traditional regions.
    - Such country-specific niche markets could also be targeted to expand revenue in each of the country
      * However, discrepancy due to sample size could not be discredited

**\*\* Most popular genres combined outside USA**

**Input**

WITH TopGenresByCountry AS (

SELECT c.country, g.name AS genre,

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

ROW\_NUMBER() OVER (PARTITION BY c.country ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS genre\_rank

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

),

Top3GenresByCountry AS (

SELECT country, genre, total\_sales

FROM TopGenresByCountry

WHERE genre\_rank <= 3

),

GenreFrequency AS (

SELECT genre, COUNT(\*) AS frequency

FROM Top3GenresByCountry

GROUP BY genre

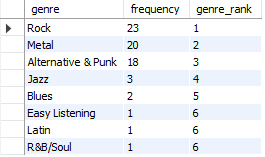
)

SELECT genre, frequency, RANK() OVER (ORDER BY frequency DESC) AS genre\_rank

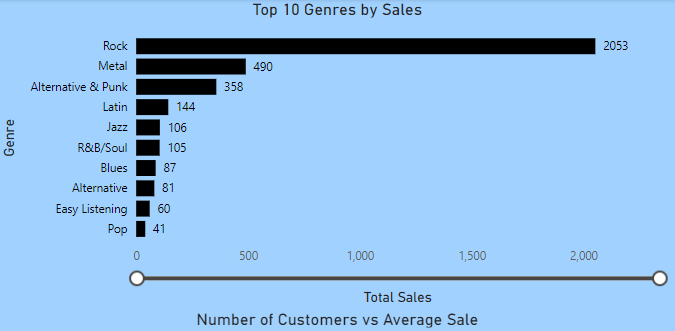
FROM GenreFrequency

ORDER BY genre\_rank;

**Output**



**Visualization**



**\*\* Country specific popularity of genres outside USA**

**Input**

WITH TopGenresByCountry AS (

SELECT c.country, g.name AS genre,

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

ROW\_NUMBER() OVER (PARTITION BY c.country ORDER BY SUM(il.unit\_price \* il.quantity) DESC) AS genre\_rank

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

)

, Top3GenresByCountry AS (

SELECT country, genre, total\_sales

FROM TopGenresByCountry

WHERE genre\_rank <= 3

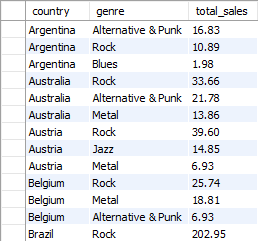
)

SELECT country, genre, total\_sales

FROM Top3GenresByCountry

ORDER BY country, total\_sales DESC;

**Output format**

****

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?**

**Approach:** Any customers that made their first purchase in 2020 (the year of which the latest data is available) were classified as ‘New customers’ and the rest were classified as ‘Long-term customers’.

**Insights**: All customers were onboarded pre-2020 and hence, all customers have been with the Chinook music store for some time.

* Loyalty/retention rate of customers is high, as all of them came back for a purchase with a total purchase frequency of greater than or equal to 5
  + Average spending increased too with the frequency of purchases
  + Such customers should be given benefits by means of loyalty programs and personalized offers
* However, the music store is struggling to on-board any new customers since at least the past year
* Current customers could be roped in by means referral programs for discounts for on-boarding new customers

**Input**

WITH customer\_purchases AS (

SELECT

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

MIN(i.invoice\_date) AS first\_purchase\_date, COUNT(i.invoice\_id) AS purchase\_frequency, ROUND(AVG(i.total),2) AS average\_spending, SUM(i.total) AS total\_spending

FROM customer c

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name

)

SELECT

customer\_id, CONCAT(first\_name,“ ”, last\_name) AS name,

purchase\_frequency, average\_spending, total\_spending,

CASE

WHEN YEAR(first\_purchase\_date) = 2020 THEN 'New Customer'

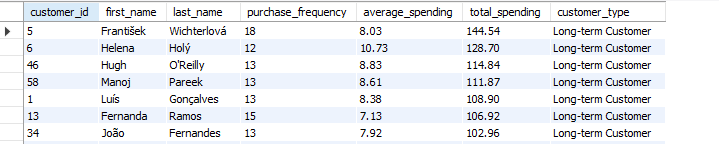
ELSE 'Long-term Customer'

END AS customer\_type

FROM customer\_purchases

ORDER BY total\_spending DESC, customer\_type;

**Output**



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?** 
   * **Approach:** To identify which music genres, artists, or albums are frequently purchased together by customers, we can find pairs of different items (genres, artists & music) that are purchased together and check the amount of time that a particular combination was repeated.
   * **Insights:** By leveraging the data on frequently paired purchases of genres, artists or albums, the music store can enhance their product recommendations and cross-selling initiatives, leading to increased sales and customer satisfaction.
     + These insights can help create targeted promotions, personalized recommendations, and strategic product bundles that align with customers' purchasing patterns
     + For example, if a customer is into Rock, and is buying Jimi Hendrix’s album "Are You Experienced?", we can also recommend then Queen’s "Greatest Hits I".
     + **Cross-Selling**: we could also Bundle these two albums or offer a discount on the second when the first is purchased.

**Input**

WITH customer\_purchases AS (

SELECT

il.invoice\_id, il.track\_id, t.album\_id, t.genre\_id,

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

ar.name AS artist\_name, i.customer\_id

FROM invoice\_line il

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

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

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

)

SELECT

p1.genre\_name AS genre1, p1.artist\_name AS artist1, p1.album\_title AS album1,

p2.genre\_name AS genre2, p2.artist\_name AS artist2, p2.album\_title AS album2,

COUNT(\*) AS frequency

FROM customer\_purchases p1

JOIN customer\_purchases p2 ON p1.customer\_id = p2.customer\_id

AND p1.track\_id < p2.track\_id

AND (p1.genre\_name != p2.genre\_name OR p1.artist\_name != p2.artist\_name OR p1.album\_title != p2.album\_title)

GROUP BY p1.genre\_name, p1.artist\_name, p1.album\_title, p2.genre\_name, p2.artist\_name, p2.album\_title

ORDER BY frequency DESC

LIMIT 10;

**Output**



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?** 
   * **Assumptions:** To do a regional market analysis, we can identify the total size of the market per country, total sales per country, average order value per country and the number of churned customers.
   * **Insights:** Several countries have **churned customers**, indicating that churn is a global issue rather than confined to specific regions.
   * Notable countries with higher churned rates include Brazil, France and Germany
   * Most countries have just a few customers (1) hence it is making it challenging to draw statistically significant conclusions for these regions.
   * Average order value is relatively consistent across countries indicating a similar spending behaviour irrespective of the region
   * All customers in this data set are long-term customers, as no new customers are identified., suggesting a mature customer base with potential for loyalty programs to be effective in retention.

**Input**

WITH customer\_dates AS (

SELECT

c.customer\_id, c.country, MIN(i.invoice\_date) AS first\_purchase\_date, MAX(i.invoice\_date) AS last\_purchase\_date,

DATEDIFF(MAX(i.invoice\_date), MIN(i.invoice\_date)) AS tenure,

DATEDIFF('2020-12-31', MAX(i.invoice\_date)) AS days\_since\_last\_purchase

FROM customer c

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

GROUP BY c.customer\_id, c.country

),

customer\_classification AS (

SELECT

customer\_id, country, tenure, days\_since\_last\_purchase,

CASE

WHEN YEAR(first\_purchase\_date) = 2020 THEN 'New Customer'

ELSE 'Long-term Customer'

END AS customer\_type,

CASE

WHEN last\_purchase\_date < DATE\_SUB('2020-12-31', INTERVAL 6 MONTH) THEN 'Churned Customer'

ELSE 'Active Customer'

END AS customer\_status

FROM customer\_dates

)

SELECT

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

SUM(i.total) AS total\_sales,

ROUND(AVG(i.total),2) AS average\_order\_value,

COUNT(DISTINCT CASE WHEN cc.customer\_type = 'Long-term Customer' THEN c.customer\_id END) AS long\_term\_customers,

COUNT(DISTINCT CASE WHEN cc.customer\_type = 'New Customer' THEN c.customer\_id END) AS new\_customers,

COUNT(DISTINCT CASE WHEN cc.customer\_status = 'Churned Customer' THEN c.customer\_id END) AS churned\_customers

FROM customer c

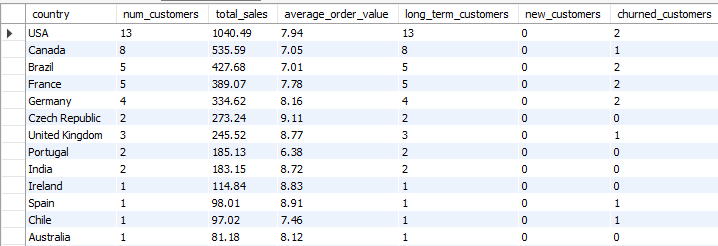
JOIN invoice i ON c.customer\_id = i.customer\_id

JOIN customer\_classification cc ON c.customer\_id = cc.customer\_id

GROUP BY c.country

ORDER BY total\_sales DESC;

**Output format**



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?
   * **Assumptions:** On the basis of time since last purchase date, we can identify the customers that are likely to churn, and also, we can classify them as following:
     + High-Risk Customers: Customers with a long period since their last purchase (>180 days) are likely to churn or reduce spending. Strategies might include targeted re-engagement campaigns or personalized offers to incentivize repeat purchases.
     + Medium-Risk Customers: These customers may show some engagement with some required efforts. Monitoring their behavior closely and providing timely promotions could help retain them.
     + Low-Risk Customers: Recent purchasers (<=90 days since last purchase) are more engaged and may require less immediate attention. However, continued nurturing through excellent service and relevant offers is crucial to retaining their loyalty.
   * **Insights:** High risk customers show a reduced average spending.
     + Also, the average number of purchases done per person in all three risk categories show purchase frequency decreasing marginally in medium and high-risk customer (High = 10, Medium = 10, Low = 11)

**Input (for identifying customers and risk frequency)**

SELECT

c.customer\_id,

CONCAT(c.first\_name," ",c.last\_name) AS name, c.country,

COUNT(i.invoice\_id) AS purchase\_frequency, SUM(i.total) AS total\_spending, DATEDIFF('2020-12-31', MAX(i.invoice\_date)) AS days\_since\_last\_purchase,

CASE

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) > 180 THEN 'High Risk'

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) BETWEEN 90 AND 180 THEN 'Medium Risk'

ELSE 'Low Risk'

END AS risk\_category,

CASE

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) > 180 THEN 1

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) BETWEEN 90 AND 180 THEN 2

ELSE 3

END AS risk\_order

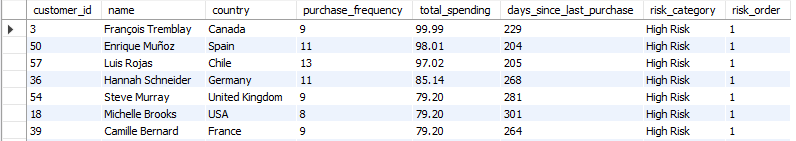
FROM customer c

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

GROUP BY c.customer\_id, c.first\_name, c.last\_name, c.country

ORDER BY risk\_order, total\_spending DESC;

**Output**



**Input (for identifying avg purchase and spend across risk categories)**

SELECT

risk\_category,

ROUND(AVG(total\_spending),2) AS avg\_spending,

ROUND(AVG(purchase\_frequency),0) AS avg\_purchase\_per\_person

FROM (

SELECT

c.customer\_id,

CONCAT(c.first\_name, " ", c.last\_name) AS name,

c.country,

COUNT(i.invoice\_id) AS purchase\_frequency,

SUM(i.total) AS total\_spending,

DATEDIFF('2020-12-31', MAX(i.invoice\_date)) AS days\_since\_last\_purchase,

CASE

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) > 180 THEN 'High Risk'

WHEN DATEDIFF('2020-12-31', MAX(i.invoice\_date)) BETWEEN 90 AND 180 THEN 'Medium Risk'

ELSE 'Low Risk'

END AS risk\_category

FROM customer c

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

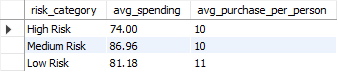
GROUP BY c.customer\_id, c.first\_name, c.last\_name, c.country

) AS customer\_risk\_profile

GROUP BY risk\_category

ORDER BY FIELD(risk\_category, 'High Risk', 'Medium Risk', 'Low Risk');

**Output**

****

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?** 
   * **Approach:** To predict the customer lifetime value (CLV) and analyze characteristics of customers who have stopped purchasing, we can
     + provide a breakdown of customers based on their tenure, helping to understand different stages of customer loyalty.
     + Average Spending and Orders: Indicates how spending and purchasing behaviour vary across customer segments.
     + Churned Customers: Identifies how many customers have stopped purchasing, potentially indicating characteristics or behaviours associated with churn.
   * **Insights:** Most of the customers have been associated with the music store since the beginning, however, a significant number have started to churn.
     + Loyal customers are the highest spenders; hence it could yield significant returns to keep them associated with the music store by means of loyalty programs, personalized offers and regular engagement by means of early access to concerts and events

**Input**

WITH customer\_clv AS (SELECT c.customer\_id, c.first\_name, c.last\_name, MIN(i.invoice\_date) AS first\_purchase\_date, COUNT(DISTINCT i.invoice\_id) AS num\_orders, SUM(il.unit\_price \* il.quantity) AS total\_spending, DATEDIFF('2020-12-31', MIN(i.invoice\_date)) AS customer\_tenure\_days

FROM customer c

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

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

GROUP BY c.customer\_id

)

SELECT CASE

WHEN customer\_tenure\_days < 365 THEN 'New Customer'

WHEN customer\_tenure\_days >= 365 AND customer\_tenure\_days < 1095 THEN 'Existing Customer'

ELSE 'Loyal Customer'

END AS customer\_segment,

COUNT(\*) AS num\_customers, ROUND(AVG(total\_spending), 2) AS avg\_spending, ROUND(AVG(num\_orders), 2) AS avg\_orders

FROM customer\_clv

GROUP BY customer\_segment

ORDER BY avg\_spending DESC;

**Output**



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?**

* If such data were available, we could measure its impact on customer acquisition, retention, and overall sales by joining it with the existing customer, invoice, and invoice\_line tables. For example, let's assume we have a table called promotions with the following columns:

· promotion\_id: Unique identifier for the promotion

· promotion\_type: Type of promotion (e.g., discount, event, email)

· promotion\_name: Name or description of the promotion

· start\_date: Start date of the promotion

· end\_date: End date of the promotion

We could then join this table with the invoice and customer data to analyze the impact of promotions on various metrics:

**Input**

SELECT

  p.promotion\_type, p.promotion\_name,

  COUNT(DISTINCT CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date THEN c.customer\_id ELSE NULL END) AS new\_customers\_acquired,

  COUNT(DISTINCT CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date AND c.customer\_id IN (SELECT customer\_id FROM invoices WHERE invoice\_date < p.start\_date) THEN c.customer\_id ELSE NULL END) AS existing\_customers\_retained,

  SUM(CASE WHEN i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date THEN il.unit\_price \* il.quantity ELSE 0 END) AS promotion\_sales

FROM promotions p

LEFT JOIN invoices i ON i.invoice\_date >= p.start\_date AND i.invoice\_date <= p.end\_date

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

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

GROUP BY p.promotion\_type, p.promotion\_name;

This query joins the promotions table with the invoices, customers, and invoice\_lines tables, and calculates the following metrics for each promotion:

* new\_customers\_acquired: The number of new customers acquired during the promotion period
* existing\_customers\_retained: The number of existing customers who made a purchase during the promotion period
* promotion\_sales: The total sales generated during the promotion period
* By analyzing these metrics across different promotion types and campaigns, we can evaluate their effectiveness in driving customer acquisition, retention, and overall sales performance.
* Additionally, we could further enhance the analysis by incorporating other relevant data sources, such as email marketing metrics (e.g., open rates, click-through rates) or website analytics, to gain a more comprehensive understanding of the promotional impact across different channels.

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

* Define the business objectives.
  + To find a sustainable growth path for the music store
* Identify key metrics (sales, customer acquisition, retention).
  + Total sales for understanding the company’s financial health
  + Sales by Genre, Album, Country to determine what kind of music and what geographical location is generating the most revenue
  + Sales by artist and album to measure revenue by artist and defining artist-specific promotional strategies to retain and attract customers
  + Monthly/Quarterly sales to identify any seasonal trends (holidays or festivals)
* Identifying key markets
  + Largest country and continental markets and find out KPIs leading the revenue charts in those markets
  + Significant up-and-coming markets with strong growth potential
    - Significant average revenue per customer
    - Strong retention rate
* Conduct analysis to find trends and patterns, including:
  + Customers lifetime value for forecasting revenue to be generated from the over lifetime and to retain loyal customers using loyalty programs or discounts or early access to some artists or albums
  + Purchase frequency to understand customer engagement
  + Customer retention rate to track loyalty and identify customers who make repeat purchases
  + Churn rate to determine the percentage of customers who did not make a repeat purchase over a period of time
  + Whether the average basket price of churned customers greater or lower than average
* Find two or more products that are bought together to give suggestions for next purchase based on one purchase based on other or to club the two products with a discount offer
  + This is especially beneficial when trying to decide what albums or tracks are to be suggested for giving recommendations to customers

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?**

**Input**

ALTER TABLE album ADD COLUMN ReleaseYear INTEGER;

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.** 
   * **Approach:** We can group the total numbers of customers, total amount spent by such customers, and the average number of tracks purchased per customer by the each country to give insights to Chinook
     + **Insights:** Most of the customers have been associated with the music store since the beginning, however, a significant number have started to churn.

* USA leads as the top spending country, followed by Canada and Brazil
* Czech Republic (9.11), United Kingdom (8.77), and India (8.72) have the highest average amount spent per customer, suggesting higher individual engagement or purchasing power in these regions
  + Countries like Czech Republic, United Kingdom, India, Ireland, Spain, Belgium show higher average tracks purchased per customer (9), indicating higher music consumption per customer.
    - Denmark has the lowest average track purchased per customer, suggesting lower customers engagement
  + Such markets can be explored for future expansion

**Input**

SELECT

c.country AS country,

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

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

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

ROUND(AVG(il.track\_count),0) AS avg\_tracks\_purchased\_per\_customer

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

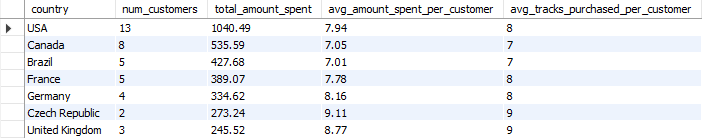
GROUP BY invoice\_id

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

GROUP BY c.country

ORDER BY total\_amount\_spent DESC;

**Output**



DASHBOARD

