**Objective Questions**

**Question 1:** **Does any tables have missing values or duplicates? If yes, how would you handle it?**

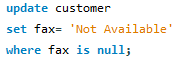
Answer:

There are no duplicates present in the given data, although there are missing values present in the 3 tables of the provided database i.e., Customer, Employee, Track.

In the table customer there are various missing values in the following columns:

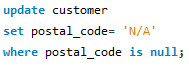
**Fax:** Fax column has missing values but as the column is not essential for the analysis but still the missing values can be filled with ‘Not Available’ as for the contact information phone number is present in the database.

It can be done by using the following query:

****

**Postal code:** Postal code column has few missing values as the column is not relevant for the analysis the missing values can be filled with ’N/A’.

It can be done using the query:

****

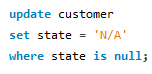
**Company:** Company column has missing values and it is not relevant for the analysis so the missing values can be filled as ‘Not Available’.

It can be done using the query:



**State**: State column has missing values which can filled as ‘N/A’.

It can be done using the following query:



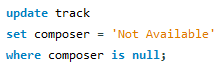
In the table Employee there are missing values:

**reports to**: in the reports to column there is null value for the employee id – 1 as employee with employee id 1 is general manager and reports to no one so that does not affect the analysis and the value can be remained as null.

In the table Track there are missing values:

**Composer**: In the composer column there are missing values and the column composer is not required for the analysis so the missing values can be filled with ‘Not Available ‘.

It can be done by using the query:



MySQL query:

update customer

set fax= 'Not Available'

where fax is null;

update customer

set postal\_code= 'N/A'

where postal\_code is null;

update customer

set company = 'Not Available'

where company is null;

update customer

set state = 'N/A'

where state is null;

update track

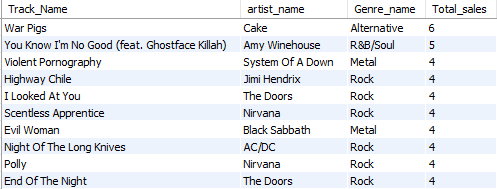
set composer = 'Not Available'

where composer is null;

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

**Answer:**

**Top 10 most selling tracks and their artist along with the genre:**

****

**Top 10 Genres highest selling genres in USA:**

****

**MySQL Query:**

**-- Top 10 tracks in USA**

**select t.name as Track\_Name,a.name as artist\_name,**

**g.name as Genre\_name ,sum(il.quantity) as Total\_sales**

**from invoice\_line il**

**inner join invoice i**

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

**inner join track t**

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

**inner join customer c**

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

**inner join album al**

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

**inner join artist a on al.artist\_id = a.artist\_id**

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

**where country= 'USA'**

**group by t.name, a.name, g.name**

**order by Total\_sales desc**

**limit 10;**

**-- Top 10 genres in USA**

**select g.name as Top\_Genres**

**from track t**

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

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

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

**where i.billing\_country = 'USA'**

**group by g.name**

**order by sum(il.quantity) desc**

**limit 10;**

**Insights:**

* **War Pigs** by the artist **Cake** is the most selling track in USA
* **Rock** is the highest selling genre in USA

**Recommendation:**

* **Rock** is the most selling genre so for marketing rock genre should be preferred over others because of the popularity
* Using the analysis, we can use the top genres for marketing which will help in boosting the revenue

**Visualization:**

**Question 3:** **What is the customer demographic breakdown (age, gender, location) of Chinook’s customer database.**

**Answer:** Chinook music stores has the franchises in 24 different countries

i.e., USA, United Kingdom, Sweden, Spain, Portugal, Poland, Norway, Netherlands, Italy, Ireland, India, Hungary, Germany, France, Finland, Denmark, Czech Republic, Chile, Canada, Brazil, Belgium, Austria, Australia, Argentina.

Top 10 countries on the basis of customer count:



**MySQL** **query**:

select distinct(country)

from customer;

select count(distinct(country)) as country\_count

from customer;

select country, count(\*) as customer\_count

from customer

group by country

order by customer\_count desc

limit 10;

**Insights**:

* Chinook music store has franchises in 24 different countries.
* USA has the highest customer count among all the countries i.e., 13.

**Recommendation**:

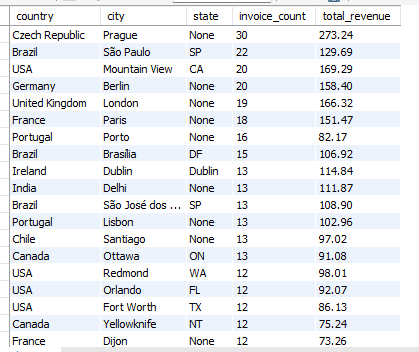
* USA should be the targeted country because of the high customer count and better market opportunities.
* USA, Canada, Brazil, France, Germany, United Kingdom should also be the priority as there are the growth opportunities in these areas.

**Visualization:**

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

**Answer:**

**Total revenue and count of invoice for each country, state and city are:**



**MySQL query:**

**select billing\_country as country, billing\_city as city, billing\_state as state,**

**count(invoice\_id) as invoice\_count, sum(total) as total\_revenue**

**from invoice**

**group by billing\_country, billing\_city, billing\_state**

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

**Insights:**

* In country **Czech Republic,** state **Prague** has the highest count of invoices and total revenue.
* Locations with high invoice counts but low revenue may indicate frequent small transactions.
* Locations with low invoice counts but high revenue may indicate premium or high-value customers.

**Recommendation:**

* By investing in high revenue locations, we can maximize our revenue and the services can tweaked for retaining the high-value customer.
* By conducting the market research in low performing areas, we can customize our marketing strategies for getting results.

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

**Answer**:

**Top 5 customers from each country on the basis of total revenue:**



**MySQL query:**

**with cte as**

**(select customer\_id, sum(total) as total\_revenue**

**from invoice**

**group by customer\_id),**

**ranked\_customers as**

**(select c.first\_name, c.last\_name, cte.total\_revenue,**

**dense\_rank() over(partition by c.country order by cte.total\_revenue desc) as ranking**

**from cte**

**left join customer c**

**on cte.customer\_id = c.customer\_id)**

**select first\_name, last\_name, total\_revenue, ranking**

**from ranked\_customers**

**where ranking <=5**

**order by ranking asc;**

**Insights:**

* We can see our top 5 customers from each country.

**Recommendation:**

* Consider offering personalized services or loyalty programs to high-revenue customers.
* Focus marketing efforts in countries where the top 5 customers have the highest spending.
* Use these regions as case studies for replicating success in other countries.
* Understand why customers in certain countries generate lower revenue. Look at factors like pricing, product availability, and customer behaviour.

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

**Answer:**

**Top-selling of each customer:**

****

**MySQL query:**

**with cte as**

**(select c.customer\_id, c.first\_name, c.last\_name, t.name as track\_name, sum(i.total) as total\_revenue**

**from customer c**

**inner join invoice i**

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

**inner join invoice\_line il**

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

**inner join track t**

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

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

**ranked\_customer as**

**(select customer\_id, first\_name, last\_name, track\_name, total\_revenue,**

**row\_number() over(partition by customer\_id order by total\_revenue desc) as ranking**

**from cte**

**order by total\_revenue desc)**

**select customer\_id, first\_name, last\_name, track\_name as top\_track**

**from ranked\_customer**

**where ranking = 1**

**order by customer\_id asc;**

**Insights:**

* The query identifies each customer's most-purchased track based on total revenue, highlighting high-value customers and their preferences.
* By analysing the most-purchased tracks across customers, patterns may emerge about popular genres, albums, or artists preferred by different customer groups.
* Tracks associated with higher revenue for individual customers may indicate premium pricing, frequent purchases, or both.
* Conversely, tracks with lower revenue among top picks may suggest frequent purchases of lower-priced items.

**Recommendation:**

* Group customers based on total revenue and track preferences to tailor engagement strategies.
* Provide loyalty points, discounts, or free track downloads for high-spending customers to incentivize continued purchases.
* Use aggregated data on top tracks to identify emerging trends in customer preferences by genre or artist.

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

**Answer:**

**We can analyze the frequency of the purchases across the years.**

**We can analyze the average order value of the orders placed in months across the years.**

**MySQL query:**

**select count(invoice\_id) as invoice\_count,**

**date\_format(invoice\_date,'%Y-%m') as date,**

**avg(total) as monthly\_avg\_order\_value**

**from invoice**

**group by date\_format(invoice\_date,'%Y-%m')**

**order by date\_format(invoice\_date,'%Y-%m');**

**Insights:**

* **Frequency of Purchases:**
* Peaks in **January** **2018** (**22**), **February** **2019** and **August** **2019** (**19**), and **October** **2020** (**18**).
* Significant drop in **December 2017 (5 invoices)** and **September 2018 (6)**.
* Months like **April 2019 (17)** and **April 2020 (16)** suggest steady performance during that period.
* Low activity in specific months like **November 2018** (**7**) and **November 2019** (**8**) could reflect seasonal dips.
* **Average order value:**
* **Trends:**
* Peaks and troughs are visible across months, e.g., higher averages in July 2017, June 2020, and October 2018.
* December values seem inconsistent, with lows in 2017 and 2020, but higher averages in other years.
* **Yearly Comparison:**
* There are fluctuations in the average order value across the years.
* June 2020 has a notable spike to 10.98, the highest monthly average.
* **Fluctuations:**
* Some months, like January 2020 (5.445), indicate lower engagement compared to January in other years.
* The average tends to dip in months like October-November, except for November 2019 and November 2020.

**Recommendation:**

* **Frequency of Purchases:**
* Plan promotions and campaigns in **January**, **February**, and **August**, as these months show consistently high engagement.
* Analyse reasons for low invoice counts in months like **December 2017** and **September 2018**. Possible causes could include seasonal preferences or ineffective campaigns.
* For months with irregular patterns (e.g., **December**), introduce targeted strategies to boost engagement, such as holiday-themed promotions or discounts.
* Months like **April** have shown consistent performance, making them good candidates for retaining existing customers and introducing loyalty programs.
* **Average order value:**
* Months with higher averages, such as **June 2020**, suggest opportunities for promotional activities or replicating strategies used during these periods.
* Months like December 2017 (5.742) or January 2020 (5.445) could indicate factors like reduced customer activity or inadequate campaigns.
* Build targeted campaigns around months consistently showing high averages (e.g., **June, July, or October**).

**Question 8: What is the customer churn rate?**

**Answer: Churn rate is the percentage of customer who stopped using the services at the specific time period which can be observed using the following formula:**

**Number of churned customers/ total number of customers \* 100**

****

**i.e., (16/59) \* 100 which gives the churned percentage of 27.12**

**MySQL query:**

**with last\_purchase as**

**(select c.customer\_id, max(i.invoice\_date) as last\_purchase\_date**

**from customer c**

**inner join invoice i**

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

**group by c.customer\_id**

**),**

**total\_customer\_count as**

**(select count(distinct customer\_id) AS total\_customers**

**from customer**

**),**

**ChurnedCustomers as**

**(select count(distinct customer\_id) as churned\_customers**

**from last\_purchase**

**where last\_purchase\_date < date\_sub('2020-12-31', interval 6 month)**

**)**

**select t.total\_customers, c.churned\_customers,**

**round((c.churned\_customers / t.total\_customers) \* 100,2) as churned\_percentage**

**from total\_customer\_count t, ChurnedCustomers c;**

**Insights:**

* The churn rate is 27.12% which is high and indicates there is high loss customers within the last years.
* Losing 27% of customers can significantly impact revenue, as retaining an existing customer is often more cost-effective than acquiring a new one.
* A high churn rate suggests potential challenges in customer engagement or satisfaction.
* It could indicate a lack of loyalty or insufficient efforts to keep customers returning.
* Possible causes could be Weak customer loyalty programs or incentives.

**Recommendation:**

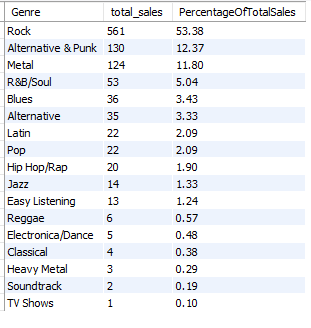
* Email Campaigns: Regularly engage customers with personalized emails highlighting new releases, recommendations, and exclusive discounts.
* Feedback Collection: Use surveys or feedback forms to understand why customers leave and what improvements they seek.
* Introduce a points-based loyalty program where customers earn rewards for purchases, encouraging repeat business.
* Introduce more diverse genres or exclusive content to attract and retain customers with varied tastes.

**Visualization:**

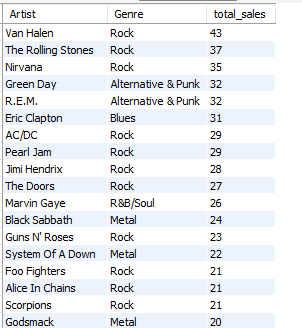
**Question 9: Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.**

**Answer:**

**Total sales contributed by each genre in USA:**

****

**Best-selling genres and artists in USA:**

****

**MySQL query:**

**-- percentage of total sales contributed by each genre**

**select g.name as Genre, sum(il.quantity) as total\_sales,**

**round((sum(il.quantity) \* 100.0 / (select sum(il2.quantity)**

**from invoice\_line il2**

**inner join invoice i2**

**on il2.invoice\_id = i2.invoice\_id**

**where i2.billing\_country = 'USA')),2) as PercentageOfTotalSales**

**from invoice\_line il**

**inner join track t**

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

**inner join genre g**

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

**inner join invoice i**

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

**where i.billing\_country = 'USA'**

**group by g.name**

**order by total\_sales desc;**

**-- best-selling artists and genre in USA:**

**with cte as**

**(select g.name as Genre, ar.name as Artist,**

**sum(il.quantity) as total\_sales**

**from invoice\_line il**

**inner join track t**

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

**inner join genre g**

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

**inner join album al**

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

**inner join artist ar**

**on al.artist\_id = ar.artist\_id**

**inner join invoice i**

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

**where i.billing\_country = 'USA'**

**group by g.name, ar.name)**

**SELECT Artist, Genre, total\_sales**

**FROM cte**

**ORDER BY total\_sales DESC;**

**Insights:**

* **Rock** is the top-performing genre with significant sales, driven by iconic artists like **Van Halen (43 sales)**, **The Rolling Stones (37 sales)**, and **Nirvana (35 sales)**. The strong customer preference for Rock suggests that it should remain a primary focus for promotional activities.
* Several other prominent Rock artists such as **AC/DC**, **Pearl Jam**, and **Jimi Hendrix** also contribute significantly to the overall sales, indicating sustained engagement with Rock music across a broad audience.
* Genres like **Alternative & Punk** (e.g., **Green Day (32 sales)** and **R.E.M. (32 sales)**) and **Metal** (e.g., **Black Sabbath (24 sales)**, **System of a Down (22 sales)**) contribute notably, although their audiences may be more niche compared to Rock. These genres have potential for growth with the right marketing strategies.
* **R&B/Soul** and **Hip Hop/Rap**, with artists like **Marvin Gaye (26 sales)** and **House of Pain (20 sales)**, also show steady sales, highlighting their appeal in specific demographic segments.
* **Classical** genres, represented by artists like **Luciano Pavarotti (1 sale)**, show limited engagement and may benefit from targeted niche marketing.

**Recommendation:**

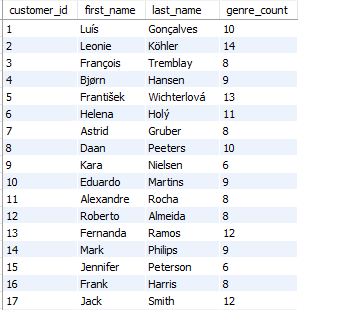
* Rock is the top-performing genre with significant sales, driven by iconic artists like Van Halen (43 sales), The Rolling Stones (37 sales), and Nirvana (35 sales). The strong customer preference for Rock suggests that it should remain a primary focus for promotional activities.
* **Targeted Campaigns**: Launch promotional campaigns around Alternative & Punk and Metal genres, especially highlighting key artists like **Green Day**, **R.E.M.**, **Black Sabbath**, and **System of a Down**.
* **Blues, R&B/Soul, and Hip Hop/Rap**: While these genres are performing moderately, it would be valuable to increase exposure through themed playlists (e.g., "Blues Legends", "R&B Classics") or collaborations with popular artists in these genres.
* **Genre-Specific Content**: Build genre-specific playlists that cater to the tastes of customers interested in these genres, incorporating emerging artists to keep the content fresh.
* Bundle albums or songs from genres with high engagement (e.g., Rock and Alternative & Punk) with emerging genres to boost sales and introduce customers to new music.

Visualization:

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

**Answer:**

**Customers who have purchased tracks from at least 3 different genres.**

****

**MySQL query:**

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

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

**FROM customer c**

**inner join invoice i**

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

**inner join invoice\_line il**

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

**inner join track t**

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

**inner join genre g**

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

**group by c.customer\_id**

**having genre\_count >= 3**

**order by c.customer\_id asc;**

**Insights:**

* There are 59 customers in total who have purchased tracks from 3 or more different genres.
* The number of genres varies from 5 to 14 genres per customer, with some customers engaging with a broad range of music styles.
* Customers who engage with a wide variety of genres are likely to be more engaged with the platform, making them important for retention strategies.

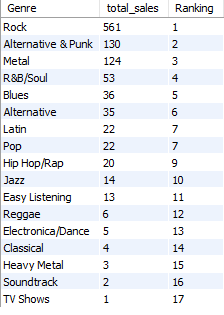
**Recommendation:**

* Offering promotions or discounts targeted at customers who engage with a wide range of genres (like those who purchase from 10+ genres). You can personalize marketing campaigns based on their broad tastes, encouraging them to try new releases in genres they haven’t explored yet.
* For customers purchasing from fewer genres, offering curated recommendations or genre-specific bundles to increase engagement with other genres they might enjoy.
* For customers who purchase from fewer genres (e.g., 5 genres), it may be beneficial to encourage them to try new genres. Sending targeted emails or creating pop-up recommendations on the platform based on their historical preferences more genres.

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

**Answer:**

**Genres ranked on the basis of their sales in USA:**

****

**MySQL query:**

**select g.name as Genre, sum(il.quantity) as total\_sales,**

**rank() over (ORDER BY sum(il.quantity) DESC) as Ranking**

**from invoice\_line il**

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

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

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

**where i.billing\_country = 'USA'**

**group by g.name**

**order by Ranking;**

**Insights:**

* **Rock** is by far the best-selling genre in the USA.
* **Alternative & Punk** and **Metal** are also performing well in the market. These genres collectively represent a large portion of the total sales, and they show potential for further growth.
* **R&B/Soul** and **Blues** are also notable genres, contributing to a substantial share of the sales.
* **Classical** and **TV Shows** are the least popular genres in terms of sales
* Genres like **Latin**, **Pop**, and **Hip Hop/Rap** have relatively low sales, with each contributing around 2-3% to the total sales. While not among the top genres, these still have a dedicated following.

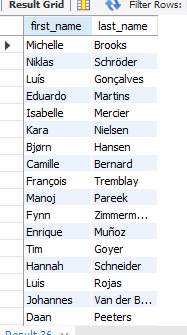
**Recommendation:**

* Since **Rock** and **Alternative & Punk** are leading the sales, continue to invest in and promote these genres for generating more revenue.
* While **R&B/Soul**, **Blues**, and **Metal** are doing relatively well, they can benefit from a targeted marketing strategy to expand their audience.
* **Classical** and **Soundtrack** genres have a small share of the market. Marketing campaigns could help to generate more interest and sales in these genres.

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

**Answer:**

**Customers who have not made any purchases in last 3 months:**

****

**MySQL query:**

**select c.first\_name, c.last\_name**

**from customer c**

**inner join invoice i**

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

**group by c.first\_name, c.last\_name**

**having max(i.invoice\_date) < DATE\_SUB('2020-12-31', interval 3 month);**

**Insights:**

* These customers have been inactive for a significant period (more than 3 months), which could indicate a potential churn or reduced engagement.

**Recommendation:**

* **Personalized campaigns** could be used to encourage these customers to return.
* Offer **discounts** or **promotions** based on their past purchasing history, especially if they are high-value customers.
* Reach out to this segment with a **short survey** to understand why they haven't made purchases recently. Their feedback could provide valuable insights into potential issues with the product offering, pricing, or customer experience.
* Implementing or promoting **loyalty rewards** for return purchases can incentivize these customers to come back.

**Subjective Questions**

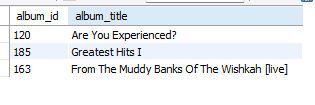
**Question 1: Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.**

**Answer:**

**Since there is no mention of the record label “New Record Label” in the dataset.**

So, as the Rock genre was the best-selling genre from the previous genre analysis,

Following are the top 3 best-selling albums from rock genre which should be prioritized for advertising and promotion:



MySQL query:

with cte as

(select a.album\_id, a.title as album\_title,

SUM(i.total) as total\_revenue

from album a

inner join track t

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

inner join invoice\_line il

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

inner join invoice i

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

where t.genre\_id = 1

GROUP BY a.album\_id

ORDER BY total\_revenue DESC)

select album\_id,album\_title

from cte

limit 3;

Insights:

* "Greatest Hits I", "From the Muddy Banks of The Wishkah [live]" and "From the Muddy Banks of The Wishkah [live]" are the top selling albums from rock genre
* Rock genre is the best-selling genre which was obtained using the previous genre analysis.

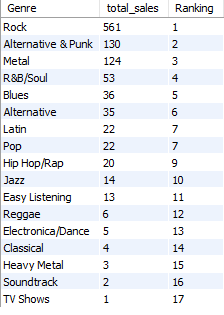
Recommendation:

* These albums are the best-selling albums so they should be promoted and advertised as they are already well established in the market and promoting them will generate more revenue.
* Promoting these albums would also attract new customers and can help in expanding the customer count and will help in maximizing the profits.

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

**Answer:**

**Top 10 best-selling genre from countries other than USA are:**

** **

**MySQL query:**

**with cte as**

**(select g.name as Genre**

**from track t**

**inner join invoice\_line il**

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

**inner join invoice i**

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

**inner join genre g**

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

**where i.billing\_country != 'USA'**

**group by g.name**

**order by sum(il.quantity) desc**

**limit 10)**

**select Genre from cte;**

**Insights:**

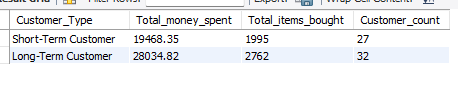
* Rock continues to dominate worldwide, making it the top-selling genre in many countries.
* Genre like Latin and Blues are more popular in other regions which could be because of the distinct musical tradition
* Outside the USA Metal, Latin, and jazz show stronger performance.

**Recommendation:**

* **Global Strategy**: Maintain a strong global focus on Rock music, as it continues to perform well internationally.
* **Regional Adaptation**: Tailor inventory management and promotions to suit local musical preferences and trends.
* **Expand Beyond the US**: Increase the promotion of Metal, Latin, and jazz music in markets outside the USA, where these genres have strong followings.
* **Targeted Marketing**: Develop marketing campaigns that cater to regional tastes, addressing the unique music preferences of each market.
* **Genre Diversification**: Explore opportunities to introduce and promote less popular genres both in the US and internationally, fostering greater diversity in the music offering.

**Question 3: Customer Purchasing Behaviour Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?**

**Answer:**

****

**MySQL query:**

**with cte as**

**(select i.customer\_id, max(invoice\_date) as last\_purchase\_date,**

**min(invoice\_date) as first\_purchase\_date,**

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

**sum(quantity) as items\_bought,**

**count(i.customer\_id) as frequency,**

**abs(timestampdiff(day, max(invoice\_date), min(invoice\_date))) as customer\_since\_days**

**from invoice i**

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

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

**group by i.customer\_id),**

**avg\_days AS**

**(select avg(customer\_since\_days) as average\_days**

**from cte),**

**tenure as**

**(SELECT total\_spent, items\_bought, frequency,**

**case**

**when customer\_since\_days> (SELECT average\_days FROM avg\_days)**

**then 'Long-Term Customer'**

**else 'Short-Term Customer'**

**end Customer\_Type**

**from cte)**

**select Customer\_Type, sum(total\_spent) as Total\_money\_spent,**

**sum(items\_bought) as Total\_items\_bought,**

**count(frequency) as Customer\_count**

**from tenure**

**group by Customer\_Type;**

**Insights:**

* Long-Term Customers have spent significantly more (28034.82) compared to Short-Term Customers (19468.35). This suggests that long-term customers are more loyal and likely to make larger or more frequent purchases.
* Long-term customers have also purchased more items (2762) compared to short-term customers (1995). This further emphasizes that long-term customers engage more with the platform, suggesting a higher degree of satisfaction and loyalty.
* There are more Short-Term Customers (27) compared to Long-Term Customers (32). This indicates that the business has a relatively larger pool of customers who are in the early stages of their buying journey.

**Recommendation:**

* Since long-term customers contribute significantly more in terms of total revenue and items bought, it’s essential to prioritize their retention. Consider offering exclusive deals, loyalty programs, and personalized services to keep them engaged.
* Target short-term customers with tailored marketing strategies. You can use discounts, personalized recommendations, or incentives for repeat purchases to encourage them to become more engaged with your business.
* The relatively low customer count for short-term customers (27 vs. 32) suggests that the business may not be fully capitalizing on potential customer lifetime value. More marketing efforts should be focused on driving repeat purchases and building long-term relationships with customers.
* Develop separate marketing campaigns for short-term and long-term customers. Short-term customers might benefit from incentive-based campaigns to encourage them to spend more or purchase frequently. Long-term customers might appreciate more reward-based campaigns that recognize their loyalty.

Visualization:

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

**Answer:**

**Following are the Genres, Artists and Albums which are frequently purchased together by the customer:**

****

**MySQL query:**

**with cte 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**

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

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

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

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

**inner join invoice i on il.invoice\_id = i.invoice\_id)**

**select c.genre\_name as Genre1, c.artist\_name as Artist1,**

**c.album\_title as Album1,ct.genre\_name as Genre2,**

**ct.artist\_name as Artist2, ct.album\_title as Album2, COUNT(\*) as Frequency**

**from cte c**

**inner join cte ct**

**on c.customer\_id = ct.customer\_id**

**and c.track\_id < ct.track\_id**

**and (c.genre\_name != ct.genre\_name or c.artist\_name != ct.artist\_name or c.album\_title != ct.album\_title)**

**group by c.genre\_name, c.artist\_name, c.album\_title, ct.genre\_name, ct.artist\_name, ct.album\_title**

**order by frequency desc**

**limit 10;**

**Insights:**

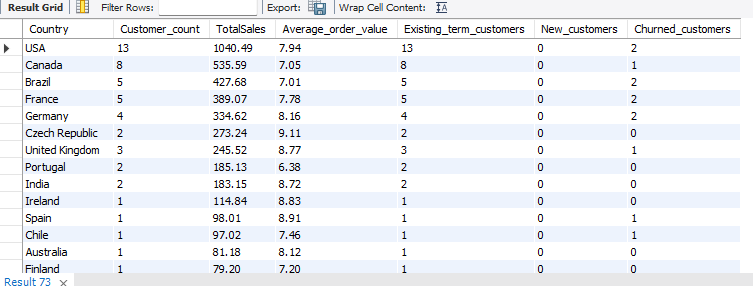
* "Are You Experienced?" by Jimi Hendrix is a recurring album in the top co-purchase combinations. This suggests that rock music fans tend to purchase multiple classic rock albums, such as those from Queen and The Who.
* The prominence of rock genres, with albums like "Greatest Hits I" by Queen, "My Generation - The Very Best of The Who", and "Greatest Kiss" by Kiss, indicates a high level of customer interest in rock music and its legacy artists.
* Rock and Blues: Albums like "The Cream of Clapton" by Eric Clapton and "In Your Honora [Disc 2]" by Foo Fighters show that rock and blues music often overlap in customer purchases.
* The pairing of "Are You Experienced?" by Jimi Hendrix and "Miles Ahead" by Miles Davis highlights that rock and jazz are often bought together and shows cross-genre purchases.
* "Get Born" by JET and "Are You Experienced?" by Jimi Hendrix reflect the intersection between alternative punk and rock genres.
* There is a mixture of genres in many co-purchases, such as Easy Listening (Frank Sinatra) with Metal (God smack) or Rock (Foo Fighters). This suggests that customers who enjoy one genre might also be inclined to purchase albums from diverse genres, especially those with similar musical qualities or themes.

**Recommendation:**

* Given that certain albums like "Are You Experienced?" by Jimi Hendrix and "Greatest Hits I" by Queen appear frequently in co-purchases, create bundles featuring these albums together. Consider marketing them as "Classic Rock Essentials" or "Rock Legends Collection" to attract customers who enjoy both.
* Promote albums across complementary genres. For instance, offer discounts for purchasing Rock albums like "Are You Experienced?" and Blues albums like "The Cream of Clapton" together, encouraging cross-genre exploration.
* Create special offers that pair Alternative & Punk albums like "Get Born" by JET with rock staples, targeting fans who appreciate both genres.
* Use these co-purchase insights for personalized product recommendations. For example, if a customer buys a rock album, suggesting them to also try an album from Jazz or Blues, as seen in the co-purchase patterns.
* By focusing on these common co-purchase patterns, we can optimize product placements, create compelling cross-sell offers, and enhance the overall customer experience through targeted marketing campaigns.

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

**Answer:**

****

**MySQL query:**

**with cte 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**

**inner 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 cte)**

**select c.country as Country,**

**count(distinct c.customer\_id) as Customer\_count,**

**sum(i.total) as TotalSales,**

**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 Existing\_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**

**inner join invoice i**

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

**inner join customer\_classification cc**

**on c.customer\_id = cc.customer\_id**

**group by c.country**

**order by TotalSales desc;**

**Insights:**

* Customer churn is a global issue, affecting multiple countries rather than being limited to specific regions.
* Countries with notably higher churn rates include Brazil, France, and Germany.
* In many countries, the customer base is very small making it difficult to draw statistically significant conclusions for these regions.
* Average order value is relatively consistent across all countries, indicating similar spending behaviour regardless of the region.
* The dataset comprises only long-term customers, with no new customers identified, suggesting a mature customer base. This presents an opportunity to implement loyalty programs aimed at enhancing customer retention.

**Recommendations:**

* For countries with higher churn rates (e.g., Brazil, France, Germany), implement tailored loyalty programs or re-engagement campaigns to address customer attrition.
* Focus on localized marketing and promotional efforts in regions with a small customer base to drive deeper engagement and potentially grow the market.
* Invest in strategies to attract new customers globally to complement the existing mature customer base.
* In countries with a single customer, monitor trends carefully to identify opportunities for expansion or deeper insights.
* Segment customers based on spending behaviour and churn likelihood to personalize communication and optimize retention initiatives.

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

**Answer:**

**Customer Risk Profile:**

****

**MySQL query:**

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

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

**sum(i.total) as Total\_money\_spent,**

**datediff('2020-12-31',max(i.invoice\_date)) as Day\_count\_since\_last\_purchase,**

**case**

**when DATEDIFF('2020-12-31', MAX(i.invoice\_date)) > 180 then 'High'**

**when DATEDIFF('2020-12-31', MAX(i.invoice\_date)) between 90 and 180 then 'Medium'**

**else 'Low'**

**end AS Risk\_level**

**from customer c**

**inner 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 Total\_money\_spent desc;**

**Insights:**

* Customers are categorized into 3 levels on the basis of their inactivity.
* High level (more than 180 days of inactivity), Low level (less than 90 days of inactivity) and medium level (between 90 and 180 days)
* Purchase history of the customer can also be analysed and the trend of churn rate can be analysed using the data and appropriate strategies can be applied for retaining the customers.

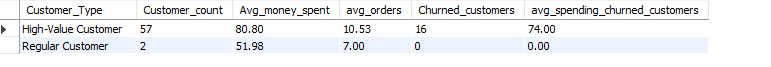
**Recommendations:**

* Promotional campaigns should be applied for the customers with the risk level of medium and high.
* Customers based on churn likelihood to personalize communication and optimize retention initiatives.
* Personalized emails and promotional campaigns should be used for retaining the customer.
* Customer at the low risk level should be provided with loyalty program so that they can be retained and there would no risk of churn in future.

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

**Answer:**

**Customer Lifetime Value Modelling:**

****

**MySQL query:**

**with cte 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,**

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

**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**

**),**

**customer\_status as**

**(select customer\_id,**

**case**

**when max(i.invoice\_date) < date\_sub('2020-12-31', interval 6 month) then 'Churned Customer'**

**else 'Active Customer'**

**end as purchase\_status**

**from invoice i**

**group by customer\_id**

**),**

**clv\_status AS**

**(select c.customer\_id, c.first\_name, c.last\_name, c.first\_purchase\_date, c.num\_orders,**

**c.total\_spending, c.customer\_tenure\_days, c.last\_purchase\_date, cs.purchase\_status**

**from cte c**

**left join customer\_status cs**

**ON c.customer\_id = cs.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 'High-Value Customer'**

**END AS Customer\_Type,**

**count(\*) as Customer\_count,**

**round(avg(total\_spending), 2) AS Avg\_money\_spent,**

**round(avg(num\_orders), 2) AS avg\_orders,**

**sum(case when purchase\_status = 'Churned Customer' then 1 else 0 end) as Churned\_customers,**

**round(avg(case when purchase\_status = 'Churned Customer' then total\_spending else null end), 2) as avg\_spending\_churned\_customers**

**from clv\_status**

**GROUP BY Customer\_Type**

**ORDER BY Avg\_money\_spent DESC;**

**Insights:**

* Customers are categorized on the basis of their tenure i.e., High-Value Customer, Regular Customer and New Customer
* High-Value Customer:
* This group represents 57 customers with an average spending of 80.80 per customer and an average of 10.53 orders.
* There are 16 churned customers in this segment, with an average spending of 74.00, indicating that even among the churned customers, spending is significant.
* Retaining these customers should be prioritized, as they contribute heavily to revenue.
* Regular Customer:
* This group includes only **2 customers** who spend **51.98** on average, with an average of **7 orders**.
* There are no churned customers in this segment, which may indicate either effective engagement strategies or low churn due to small sample size.
* While this segment is small, nurturing it can lead to higher spending and potentially convert them into high-value customers.

**Recommendations:**

* **High-Value Customers** should be the top-most priority.
* Implement targeted loyalty programs or exclusive benefits to encourage retention and reduce churn in this critical segment.
* Offer personalized incentives, such as discounts on frequently purchased items or VIP-only sales, to re-engage churned customers.
* Invest in marketing campaigns to expand the regular customer base. This could include upselling and cross-selling products related to their purchase history.
* Use churn indicators like reduced purchase frequency or spending to proactively address customer disengagement.
* Conduct surveys or collect feedback from churned high-value customers to identify specific pain points and improve services.
* Use historical purchase data and customer behaviour to predict churn risk and take preventive measures.

**Visualization:**

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

**Answer:**

If there was data on promotional campaigns (discounts, events, email marketing) then following would be the approach for measuring the impact on customer acquisition, retention, and overall sales.

Analysis would be conducted to find the growth patterns in the business which will be one the basis of following criteria:

**Customer Acquisition:**

* We can observe the growth in the number of customers (new customers) after the promotional campaigns.
* New customers further can be categorized on the basis of the type of promotional campaign which attracted the customers which can be discounts, events or email marketing.

**Retention:**

* **Retention rate:** We can observe the percentage of existing customers who made the repeated purchases after the start of the promotional campaigns.
* **Churn Rate:** We can observe the percentage of customers who stopped purchasing during after the start of the promotional campaigns.

**Overall Sales:**

* **Growth in Sales:** The growth in the revenue generated after the promotional campaigns can be observed by comparing the revenue generated with the revenue from the sales before the campaign.
* **Purchase frequency:** Difference between the frequency of customers purchases after the campaigns can be observed.

**Insights:**

**Customer Acquisition:**

* **New Customer Growth:**
* The number of customers acquired during or after the promotional campaigns reflects campaign effectiveness.
* **Attribution by Campaign Type:**
* By categorizing new customers based on the type of promotional campaign (e.g., discounts, events, email marketing), we can identify which strategy drove the most acquisitions.

**Retention Rate:**

* **Retention Rate:**
* The percentage of existing customers who made repeat purchases post-campaign provides insights into how well the campaigns kept customers engaged.
* **Churn Rate:**
* Tracking the percentage of customers who stopped purchasing during or after the campaigns highlights areas for improvement in sustaining interest.

**Overall Sales:**

* **Revenue Growth**
  + Comparing the revenue generated before and after the promotional campaigns demonstrates the financial impact of the campaigns.
* **Purchase Frequency**
  + Observing any changes in how frequently customers make purchases post-campaign can indicate increased engagement and interest.

Recommendation:

* **Targeted Campaign Design**
  + Focus on the type of campaign (discounts, events, or email marketing) that drives the highest number of new customers. For instance, if discounts attract the most customers, allocate more resources to offering compelling deals.
* **Campaign Personalization**
  + Use customer demographics and purchasing history to design personalized campaigns. This could include targeted email marketing or localized events that appeal to specific groups.
* **Loyalty Programs**
* Introduce or enhance loyalty programs to reward repeat purchases and reduce churn. Offering exclusive benefits to returning customers can help boost retention rates.
* **Follow-Up Campaigns**
* Re-engage customers at risk of churning by identifying inactive users and targeting them with tailored offers or reminders.

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

**Answer:**

**If the guidelines were not provided following would be the approach that I would take to conduct the analysis:**

**Understanding the Data and Data Cleaning:**

Firstly, I would have observed the data and then cleaned the data i.e., removed duplicates, handling null values which would help in avoiding the errors in the analysis and will help in conducting better analysis.

**Data Analysis:**

Conducting the general analysis which would help in understanding the business, analysis would have been conducted on the following topics:

**Customer Information:**

* Total customer count and their information would help in understanding the demographic distribution of the customers.
* Categorizing customers on the basis of their countries which help in understanding the market on the basis of customer distribution.

**Geographic Insights:**

* Total Sales generated in different countries which would help in understanding the countries which generates the most and least revenue.

**Revenue and Sales Analysis:**

* Sales generated by different Artists, Albums and Genres which would help in analysing the most popular artist, album and genre which help in customising the inventory for generating more revenue.

**Customer Segmentation:**

* Customers would be divided into categories on the basis of number of purchases made by the customers.
* It will help in understanding the customers and specialised campaigns can be conducted for the customers.

**Trend Analysis:**

* Monthly or yearly revenue trend can be observed which help in understanding whether the business is in profit or loss and appropriate measures can be taken on the basis of the trend analysis.

**Churn rate:**

* Churn rate of the customers will help in understanding the number of customers who are no longer making a purchase.
* It will help in understanding the countries with highest churn rate.
* Appropriate measures can be taken to increase the retention rate and minimize the churn rate.
* It will help in retaining the customers and will eventually also help in gaining new customers.

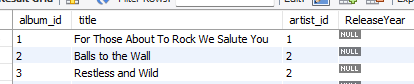
Visualization:

* Charts can be made for the visualization of the data which will help in understanding the insights in more efficient way.

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

**Answer:**





**MySQL query:**

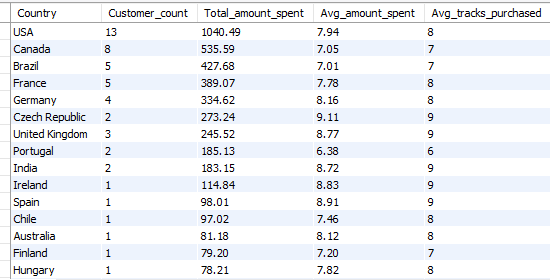
**alter table Album add ReleaseYear int;**

**select \* from Album;**

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

**Answer:**

**Purchasing behaviour of customers based on their geographical location:**

****

**MySQL query:**

**with cte as**

**(select invoice\_id, count(track\_id) as track\_count**

**from invoice\_line**

**group by invoice\_id)**

**select c.country as Country,**

**count(distinct c.customer\_id) as Customer\_count,**

**sum(i.total) as Total\_amount\_spent,**

**round(avg(i.total),2) as Avg\_amount\_spent,**

**round(avg(cte.track\_count),0) as Avg\_tracks\_purchased**

**from customer c**

**inner join invoice i**

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

**inner join cte**

**on cte.invoice\_id = i.invoice\_id**

**GROUP BY c.country**

**order by total\_amount\_spent desc;**

**Insights:**

* USA contributes the highest to total revenue with 13 customers spending 1,040.49 collectively, indicating a strong customer base.
* Other key markets include Canada, Brazil, and France, which have moderate customer counts but still generate significant revenue.
* The average amount spent per invoice is relatively consistent across countries, ranging between 7 and 8, suggesting uniform pricing or similar customer purchasing patterns globally.
* Smaller markets like Germany, Spain, and Italy have fewer customers but maintain comparable average spending per invoice, highlighting potential for expansion.
* Countries with a single customer (e.g., Ireland, Argentina) still show promising spending behaviour, warranting exploration for growth opportunities.

**Recommendation:**

**Maximize Revenue from High-Value Markets**

* Focus marketing and promotional campaigns in **USA**, **Canada**, and **Brazil**, where the largest revenues are being generated.
* Offer loyalty programs or subscription plans to retain customers in these markets.

**Expand in Smaller Markets**

* Explore targeted campaigns in underpenetrated regions like **Germany** and **Spain**, leveraging data to identify local music preferences.
* Invest in building customer acquisition strategies for single-customer regions with high spending, such as **Ireland** and **Argentina**.

**Leverage Data for Personalization**

* Analyse genre and artist preferences in top markets to create personalized recommendations, enhancing user experience and increasing sales.