**Assumptions:**

* Information above price units is not give so I just mentioned revenue like xx units.
* Analysis is done on 2020-12-30

**Objective Questions**

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

There are a total 11 tables.

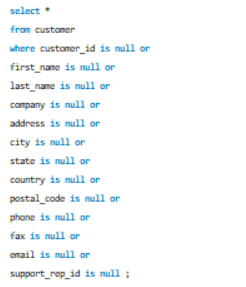
Each tables missing values and duplicates rows are checked and handled.

For example :-

In customer table

* **Missing values**

Query used

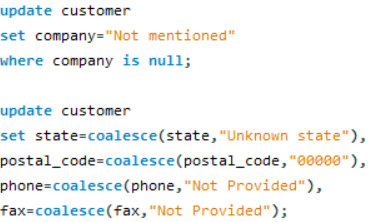
****

Result:

****

There are missing values in company, state, postal\_code, fax and phone columns. As they are text column these missing values are handled by using some default value.

* Handling missing values



This can be done using two methods that I mentioned above using update.

* Duplicate values



customer\_id is primary key for the customer table and a primary key should be repeated. so, duplicates are checked using primary key.

Result:



No duplicate values are found.

For other tables using similar ways as above missing values and duplicates are handled

* album : No missing values and no duplicate rows.
* artist : No missing values and no duplicate rows.
* employee : There is a missing value in reports\_to column it is handled using mode value and there are no duplicate rows.
* genre : No missing values and no duplicate rows.
* invoice : No missing value and no duplicate rows.
* invoice\_line : No missing value and no duplicate rows.
* media\_type : No missing value and no duplicate rows.
* playlist : No missing value and no duplicate rows.
* playlist\_track : No missing value and no duplicate rows.
* track : There are missing values in composer column and they are replaced using a default values and no duplicate rows.

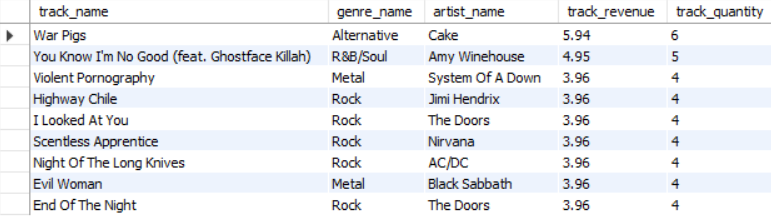
1. **Find the top-selling tracks and top artist in the USA and identify their most famous genres.**

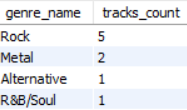
First top selling tracks and their genre are artist can get using the below query

Top selling tracks in USA :



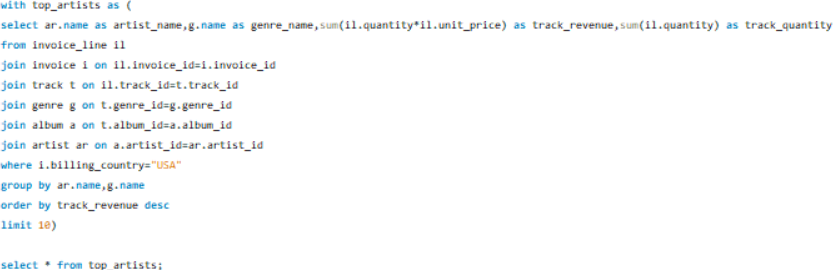
The top 9 selling tracks in USA and their genre and artist are :

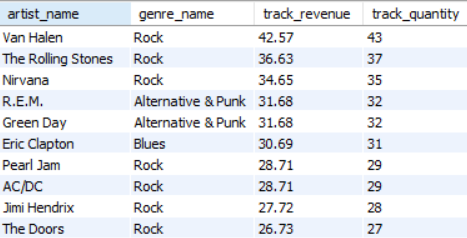




* The top selling track in USA is “War Pigs” with revenue and its genre is Alternative.
* The most famous genre of top selling tracks is Rock and around 5 out of 9 are in rock genre.
* As we can see the revenue difference between top selling tracks is not that big so we can say the tracks related to Rock genre contribute a lot of revenue.

Next top artist by revenue and their genre are found out by below query







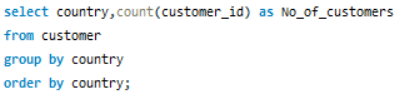
* The top artist in USA is Van Halen with a revenue of 42.57 units and his tracks belongs to Rock genre followed by The rolling stones with a revenue of 36.63 units and Nirvana with revenue of 34.65 units and also these two belongs to the rock genre.
* Almost 7 of 10 top artist produce tracks related to Rock genr

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

The customers age and gender or others things are not mentioned in the data.

Only location is provided with contact information

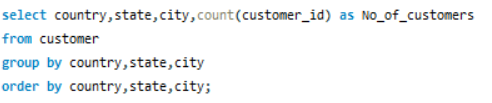
For each country no of customers is calculated using following query

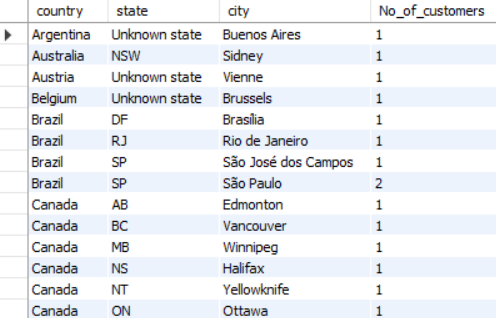
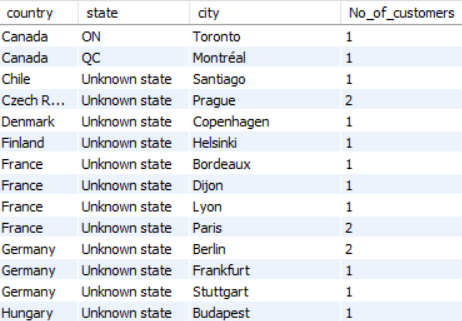


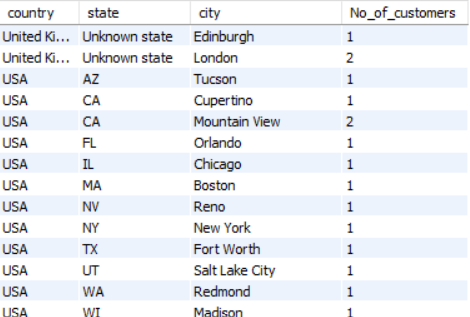
 

* From above , USA has the most no of customers followed by Canada,Brazil and France

And also for each city,state and country customer breakdown is calculated using below query



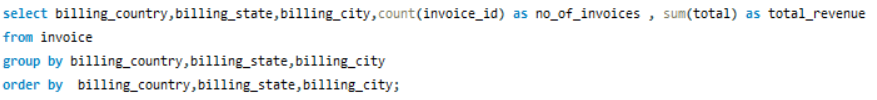
 

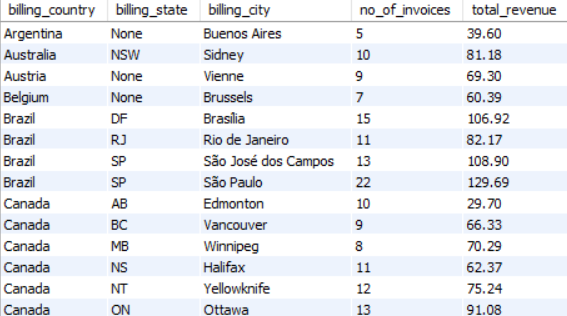
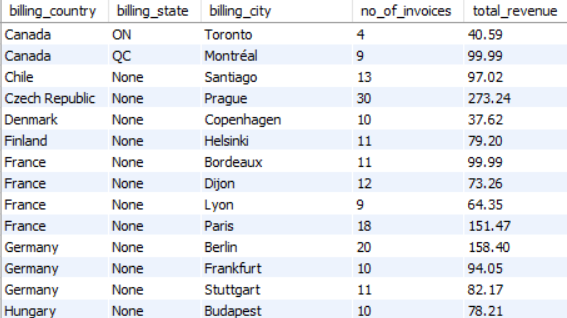
* There are some missing values of States in the original data . So,states analayis is not done.
* Coming the cities almost all the cities have same number of customers.

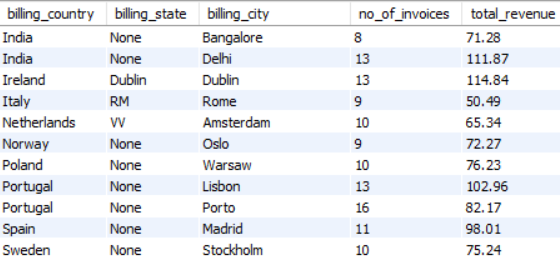
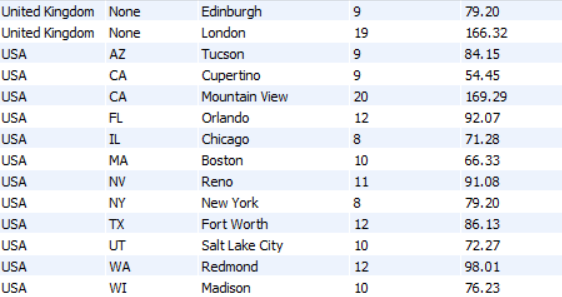
1. **Calculate the total revenue and number of invoices for each country, state, and city:**

For each city, state and country total revenue and invoices are calculated using below query



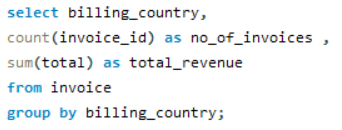
For each country, city and state the total revenue and number of invoices are as :-

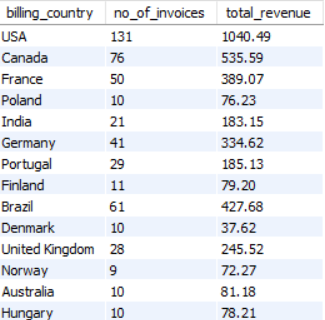
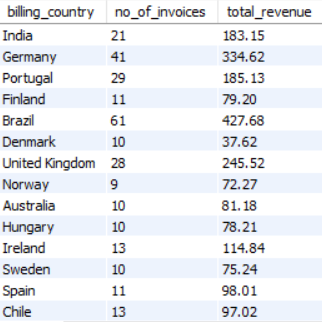
 

The above analysis doesn’t give any great insignts.

So for each country seperatly revenue and numbers of invoices are calculated using queries



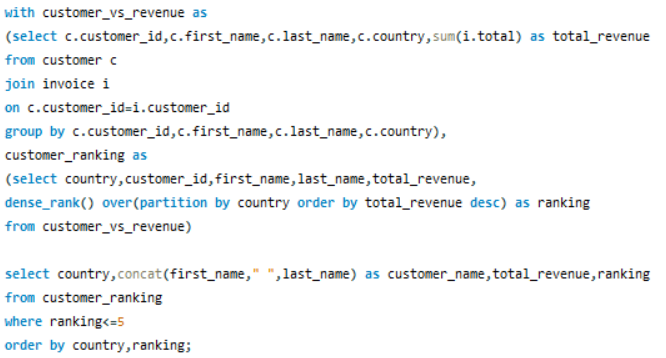
 

* USA has the highest revenue of 1040.49 units with greatest margin and highest no of invoices 131. Followed by Canada with revenue and Brazil around revenue of 400-500 and total invoices of 60-80.

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

Top 5 customers for each country by revenue find out using the following query



1. Argentina:



Argentina has only one customer named Diego and their revenue is around 39 units.

1. Australia



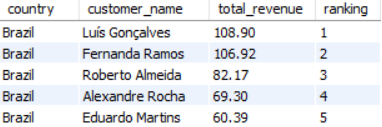
Australia also has one customer named Mark and their revenue is around 81 units.

1. Austria and Belgium



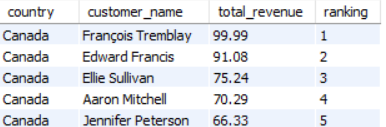
Austria and Belgium also has one customer named Astrid and Daan and their revenue is around 69 and 60 units respectively.

1. Brazil



Brazil top customers are Luis and Fernanda revenue above 100 units and top and least customers revenue varies by around 50 units

1. Canada



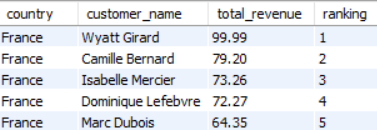
Canada’s top customers are Francois and Edward with revenue above 90 units. There is no big difference between revenues of customers in Canada. The top and low revenue customer revenues only varies by 30 units.

1. Chile, Czech Republic, Denmark and Finland



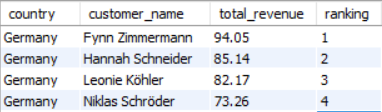
* Chile has only one customer Luis with revenue around 90 units.
* Czech has two customers Helena and Frantisek with revenue above 100 units.
* Denmark has only one customer Kara with revenue of 37 units which is low compared to the above-mentioned customers.
* Finland also has one customer Terhi with revenue around 80 units.

1. France



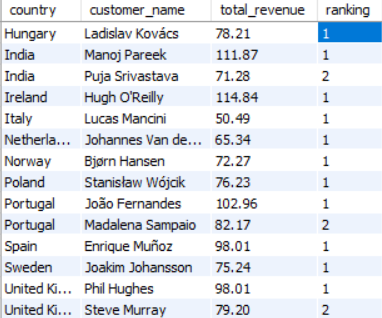
France top customer Wyatt has revenue around 100 units. And difference between top and last customers in around 35 units.

1. Germany



Germany top customer Fynn has revenue around 95 units. And the difference between revenues is not that big.

1. Hungary, India, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden and UK



* Hungary has only one customer with revenue around 78 units.
* India has two customers. Top customer has revenue around 111 units and next 71 units with a big difference.
* Italy has one customer with a low revenue around 50 units.
* Netherlands, Norway, Poland and Sweden has one customer with revenue around 70-80 units.
* Portugal has two customers and revenue of 102 and 82 and both customers has big difference
* Spain has one customer with revenue around 98 units
* UK has two customers and difference between their revenue is not that high and top customer has revenue around 98 units.

1. USA



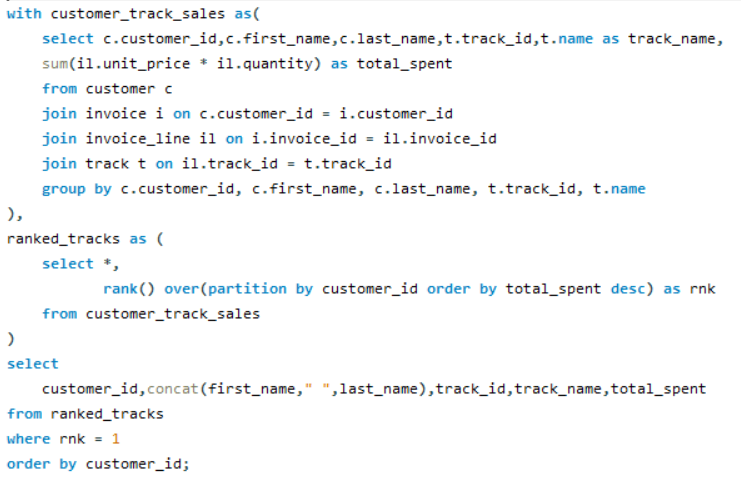
USA is the top revenue generating country has the top customers revenue is around

90 units

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

Based on revenue each customer top selling track can be found out using below query

And rank() function is used to get the top selling track for each customer



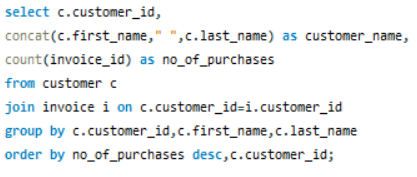
* Displaying output is some much hectic because almost all customers has more than one top selling track.
* The above query gives the required output.
* Most tracks sold for **0.99** units , with a few exceptions where tracks were sold for a higher price, such as **1.98** units.

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

Patterns or trends in customer purchasing behaviour gives some valuable insights

Frequency of purchase of Customers

For each customer their purchase frequency is calculated using following query



* From the output, we can see that František has the highest purchase count, with a total of 18.
* And next top frequency is 16 and 15
* Almost all customers falls in frequency of 8-13
* Daan and Robert has least.

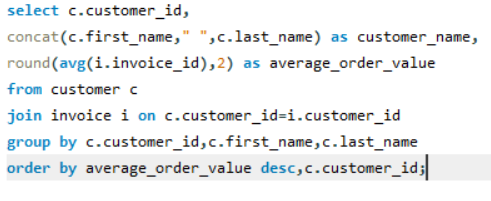
Preferred payment method

There is not data related to payment method

If its given similar to above query most frequent payment method for customers can be calculated

Average Order value

For each customer average order value is calculated using following query

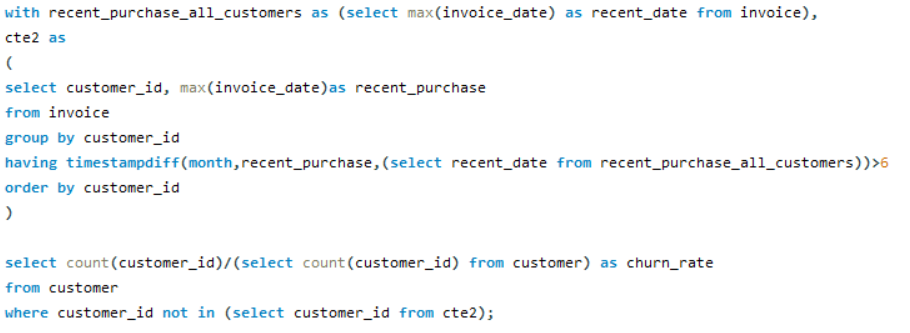


* Average order value of customer varies from 400 to 182.
* John Gordon has highest average order values .
* Most of the customer has average order value around 250-350.

1. **What is the customer churn rate?**

As per my analysis the customers who are churned are those who didn’t place any order in last 6 months.

Below query gives the customer churn rate

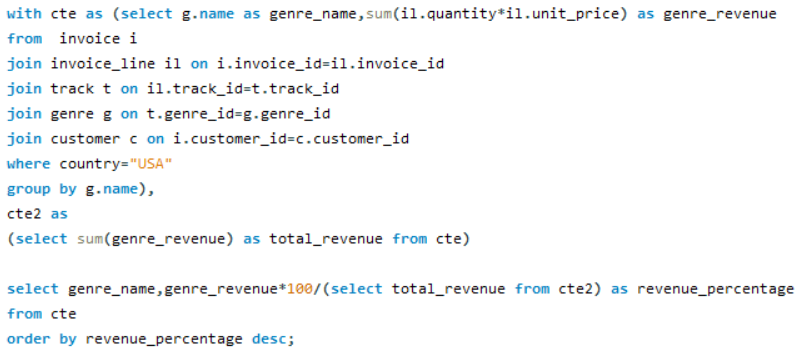


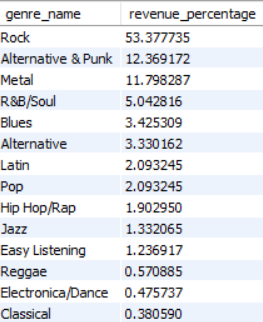


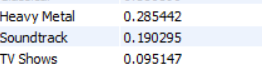
* I got churn rate of 0.7966 which is high because it states around 80% (47 customers) of the customers didn’t place their orders in last 6 months.
* This shouldn’t be ignored. This may lead to customers drop.

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

The percentage of total sales contributed by each genre in USA is calculated using below query







* Almost half the revenue generated by the Rock genre followed by Alternative & Punk 12% and Metal 11% which covers around 80% of revenue

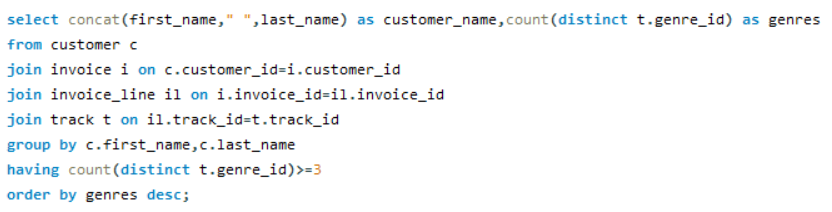
Top artists and their genre by revenue can be done using below query

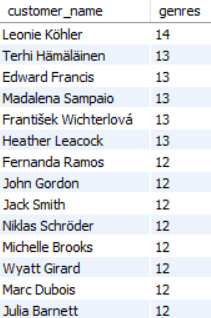
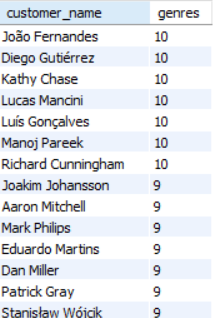


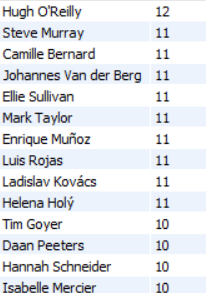


* Except one all the top artist in USA belongs to the genre Rock, Metal and Alternative & Punk
* Top 10 artist in USA are mentioned above

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





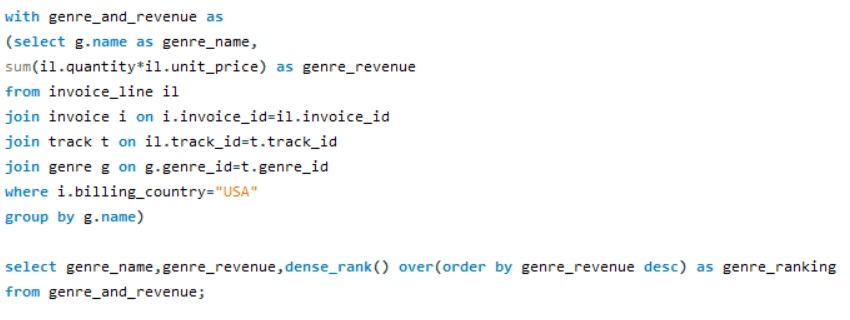
* All 59 customers that are mentioned in the data have purchased the tracks at least from 3 different genres.
* There are total of 25 different genres and around 59% of customers have purchased 10 and more different genres.

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

The sales performance of all genres in USA are calculated using below query.

Used dense\_rank() function because some genre may have same revenue and the next genres rank cannot be skipped.

Query used : -





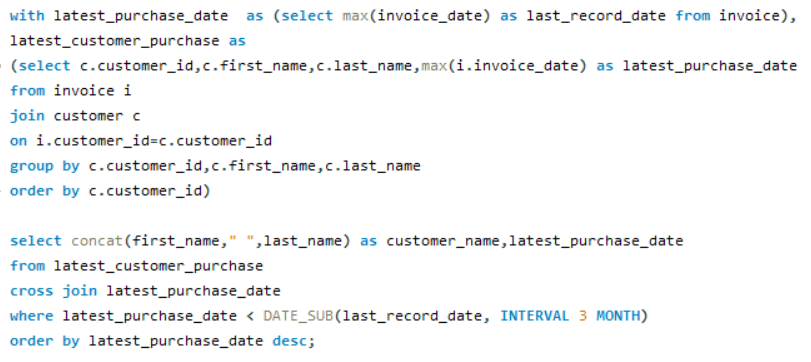


* The top genre by sales is “Rock” with revenue of 555.39 units with a significant margin followed by Alternative & Punk with revenue of 128.70 units and Metal with revenue of 122.76 units.
* R&B/Soul, Blues, Alternative, Latin, Pop, Hip Hop/Rap has moderate level of revenue around 52-20 units.
* And remaining all genres has very less revenue below 20 units where TV shows and Soundtrack genre the least sales.

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

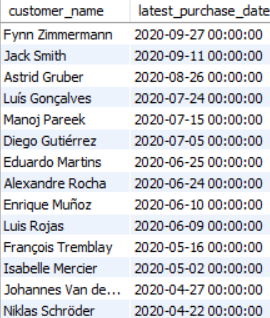
The customer who doesn’t made a purchase in the last 3 months are calculated with reference of last record date of all customer purchases.

Query used: -



Customer vs their last purchase date who does not made any purchase since

last 3 months are as

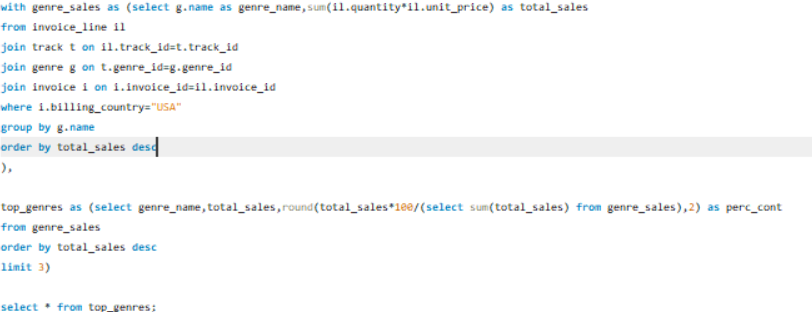
* Fynn and Jack last purchase is in September these customers are mostly likely to make the orders again .
* Around 10 customers the last purchase date is in last 6 months span so the customers in this span has possiblities that they make their order again.
* Remaining are customers are in high risk span these customers may or may not place their order again

**Subjective Questions**

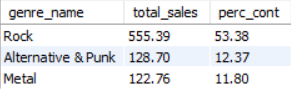
1. **Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.**

Based on the genre sales top 3 album which should be prioritised for advertising and promotions should be selected.

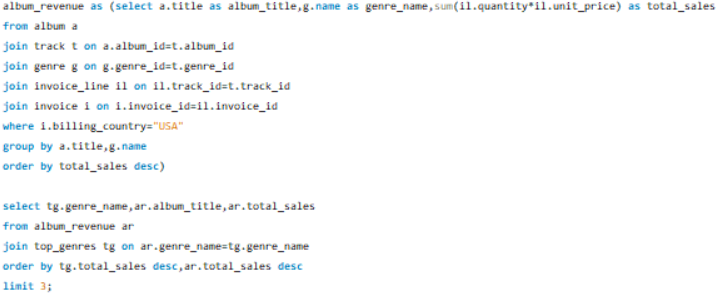
First the top 3 genres in USA can be found out by using below query

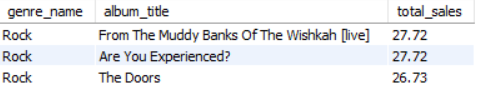


and the top 3 genres are Rock, Alternative & Punk and Metal



And next top 3 albums by revenue which belongs this genre is found out by below query





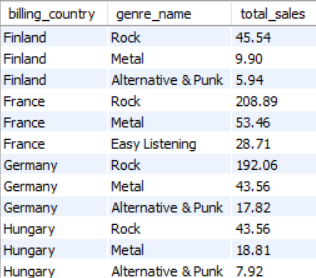
* The top 3 albums my sales are belong to Rock genre
* From the Muddy Banks Of the Wishkah
* Are you experienced?
* The Doors
* The top 3 albums are of Rock genre these three albums should be prioritised first from promotions and advertising.
* We can also promote these albums related to rock genre in USA to increase the revenue because people prioritising the albums related to this genre

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

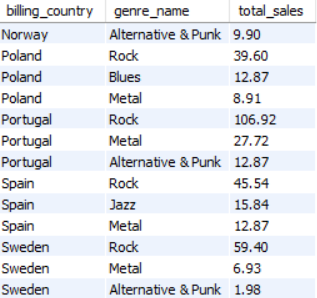
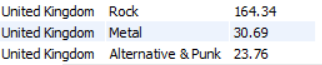
For countries other than USA top 3 genres by revenue is analysing is done using the query below



This chart is the countries count with has these genres as Top 3

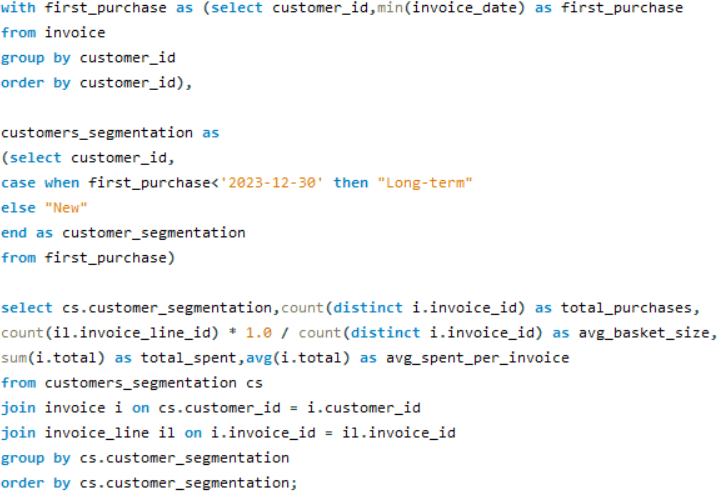
 

* All the countries has Rock genre as one of their top 3 genre.
* Except Argentia all countries has Rock their no 1 genre and countries like Canada ,Brazil, UK,Portugal,India,Germany, Czech has a very great revenue in Rock genre.
* After Rock, most common top 3 genres are Alternative & Punk and Metal.
* Some countries like Austria, Denmark, Spain has Jazz in place of Alternative & Punk as their top 3.
* Argentina and Poland has blues genre as one of their top 3 and Ireland has latin.
* Genrellay to increase revenue of store Rock, Alternative & Punk, Metal,Jazz and Blues should be our main focus genre in countries other than USA

1. **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?**

First the customers are categorised based on their first purchase whether they are new customer or Long term customer

Next for new customers and Long term customers total no of purchases, total spent, average buck size and average spent per invoice analysis is done





* In the given data there are no new customers all the customers present made there first purchase like 1-3 years ago.
* Generally, the long-term customers total purchases and total spent would be high than new customers.
* The average bucket size and average spent per invoice is good.

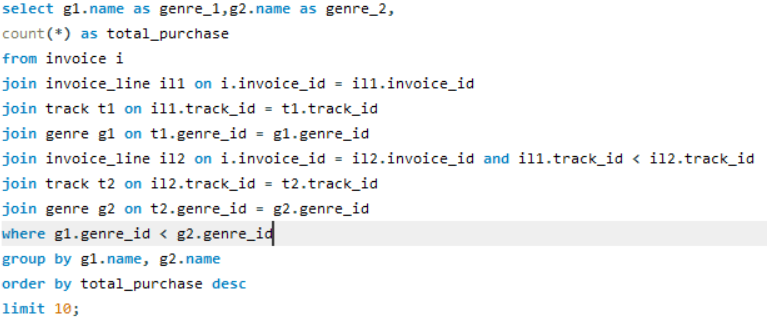
Suggestions:-

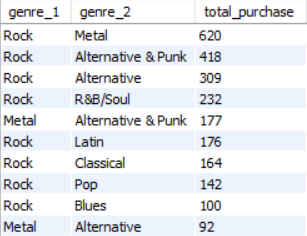
* Currently the long term customers has very good record and they are loyal. To maintain their loyalty the store can provide special discounts or reward points.
* Use their behaviour to build customer personas for targeting similar users.
* And to attract new customer advertising should be done and also the store can provide special offers these customers.
* For new customers Trigger follow-up emails or offers shortly after the first purchase.
* Offer **time-limited discounts** to encourage the second purchase.
* Improve onboarding with **personalized recommendations.**

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?**

Genre, Artist and album affinities are analysed separately

Genre Affinity analysis



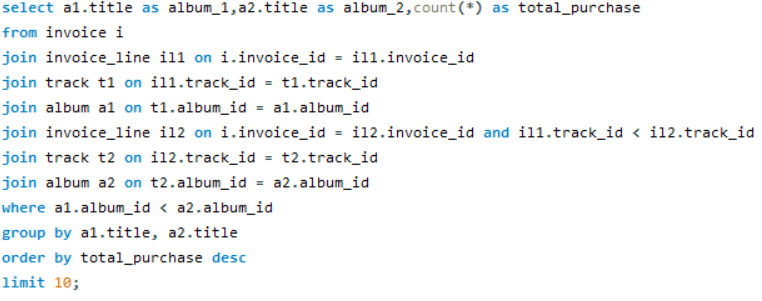


* From above we can see that 8 out of 10 top pairs has Rock as one of the genre this says that the Rock genre fans often explore other genres, making it a strong anchor for cross-promotions.
* And top purchases pair has Metal and Alternative & Punk and Alternative along with Rock with a great no of total purchases.
* R&B/Soul, Classical, Blues, Latin also appear, though with lower frequencies this is showing Rock listeners have a diverse taste profile.

Suggestions:

* When the customer purchases any track related to Rock genre we can recommend Metal, Alternative & Punk and Alternative as follow up suggestion with a discount and special offers like buy 2 rock track and get 1 metal track free.
* Run targeted campaigns for Rock listeners highlighting **crossover appeal.**

Album Affinity analysis





* As we can 6 out of 10 album pairs has Mezmerize album
* In top purchase along with Mezmerize, there are album like

Are you Experienced,

Vault: Def Leppard’s Greatest Hits,

My Generation – The Very Best of The Who and

The Singles

Suggestions

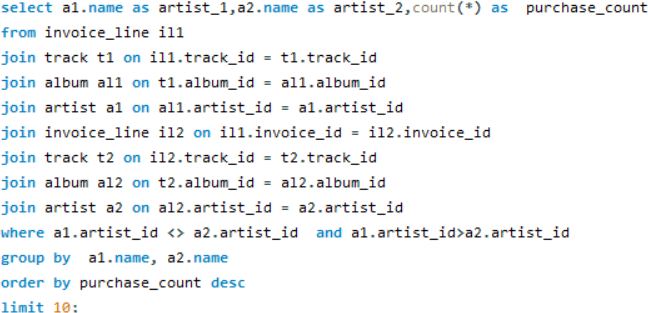
* + Create special offers or bundles featuring "Mezmerize" with:

Are You Experienced?

Vault: Def Leppard’s Greatest Hits

My Generation – The Very Best of The Who

* + Almost whenever customers purchases a album the above mentioned album should be present In next follow up suggestions. Like the customers who purchased Mezmerize also liked……
  + And also recommend the same genre albums.





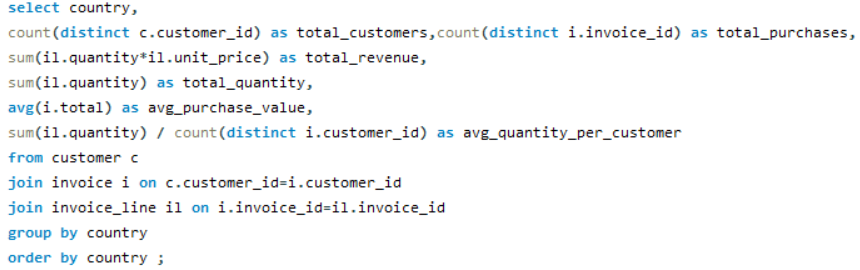
* + Green Day is the most commonly co-purchased artist. Co-purchased with Led Zeppelin (24), Foo Fighters (20), Metallica (18), and Queen (17), Guns N’ Roses (16).
  + Niravana, Queens, Guns N’ Roses also appeared twice in top 10.

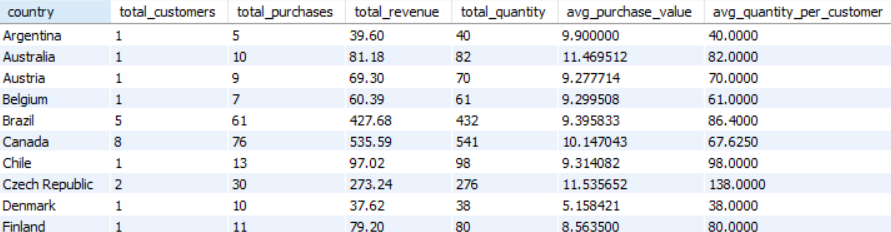
Suggestions

* + Recommend artists like Led Zeppelin, Foo Fighters, Queen, or Metallica to Green Day fans.
  + For Nirvana fans, Suggest Eric Clapton, Rolling Stones, or Amy Winehouse for mood-oriented cross-genre recommendations.
  + The artist pair with great number of purchases provide offers,
  + Also create playlist with most famous artists for big revenue.

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

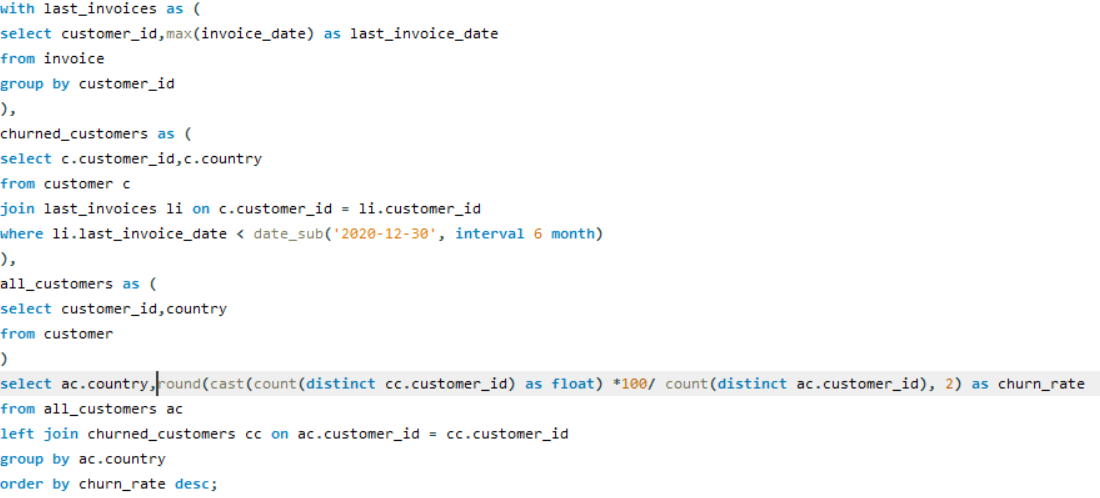
First customer purchasing behaviour for each country is analysed

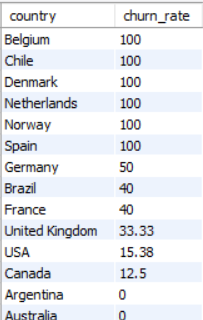






Next for each country customer churn is calculated on basis of churned customers are the costumers who didn’t made any purchase in last 6 months





Insights and suggestions

100% churn rate countries

* + Countries like Belgium, Chile, Denmark, Netherlands, Norway, and Spain have a 100% churn rate and the no of customer from these countries is only one that is the only customer is not making any purchases since last 6 months and the revenue and purchases generated by these customers is moderate.
  + These may be low-volume markets with only one-time or trial customers and also Lack of local engagement or weak customer retention strategies.
  + For countries with 100% churn, conduct surveys or interviews to identify barriers (e.g. language, most famous genre) and launch re-engagement campaigns or localized promotions.
  + Countries with minimal purchases and full churn may not be worth further investment without major changes.

40-50% churn rate

* Germany, Brazil and France has churn rate between 40-50% and no of customers is 4-5 and their no of purchases are good and also revenue generated by these customers is also great.
* But 40-50% churn rate says almost half of the customers didn’t make any purchase in last 6 months.
* Maybe its because of the cost factor or the tracks from their favourite genre are so less.
* These countries need to be concentrated and look for their requirements to improve revenue.

<40 churn rate

* USA, Canada, United Kingdom show low churn with high purchases and revenue:

USA: churn 15.38%, 1040+ in revenue. (13 customers)

UK: churn 33.33%, 245.52 in revenue. (3 customers)

Canada: churn 12.5%, 535.59 in revenue. (8 customers)

These countries revenue generated is really high the churned customer are 1-2 but in UK only three customers.

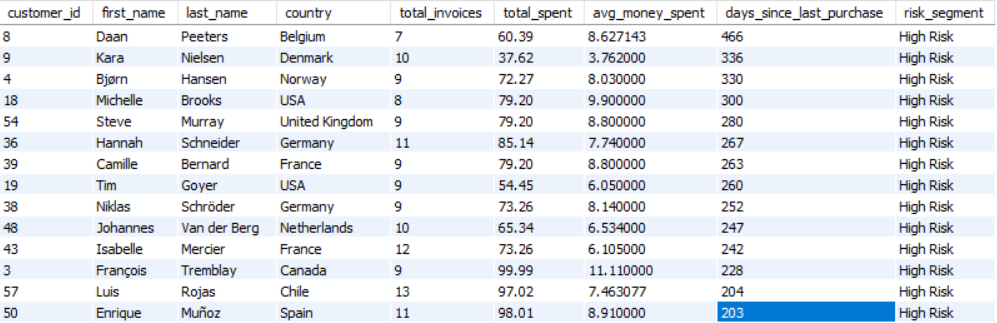
* So in UK marketing should be done to get customers.
* Even USA and Canada has high no of customers but it’s not too great so for increasing customers effort should be done in these countries.
* Portugal, India and Czech has churn rate of 0% and has moderate revenue and purchases but the no of customers is less
* Sweden, Poland, Italy, Hungary, Ireland, Finland, Austria, Australia, Argentina has 0 churn rate and low revenue and purchases and customers.
* In the above countries conduct surveys or interviews to identify barriers and launch re-engagement campaigns or localized promotions to increase market and to gain customers.

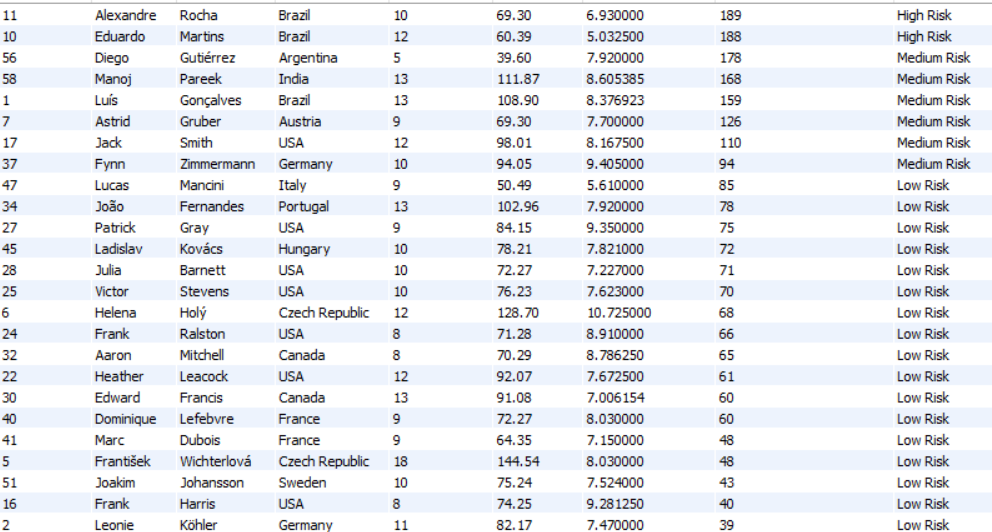
1. **Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?**

In the data, information about age and gender is not given. If they present analysis is done similar as below

Based on location and purchase history the customers who are most likely to churn and higher risk of reduced spending.









* High Risk customers are those how has longer days since their last purchase . customer Daan has days since last purchase as 466 days and he belongs to High Risk customers.
* Some customers with low total spending (e.g., ID 9, $37.62) still fall into "High Risk" due to prolonged inactivity.
* High spenders can also be at high risk because of the their days since their last purchase maybe too high.
* Countries like USA, Germany, France, and Brazil have customers across all segments, indicating mixed behaviour.
* Some smaller markets (e.g., Belgium, Chile, Norway) have customers in high-risk segments, hinting at possible market-specific engagement issues.
* Some high-risk customers have 10+ invoices, e.g., ID 43 (France), ID 3 (Canada), showing they were loyal before churning.

Suggestions:

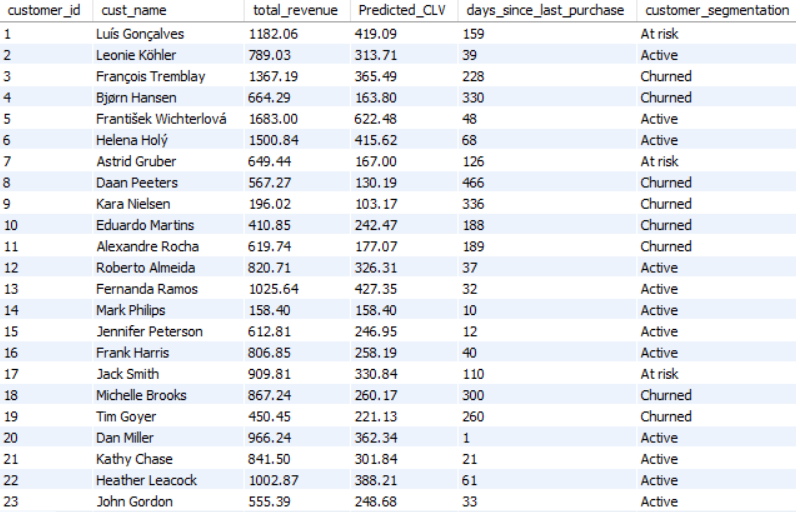
* For customers having high average spend and high days since their last purchase is high re-engagement campaign should be arranged.
* Localized reactivation campaigns for high-risk regions (Belgium, Norway, Chile).
* Offer time-sensitive discounts or personalized playlists to encourage repeat purchases.
* Low spenders and high days since last purchase can be ignored or special discounts and offers can be provided for them.

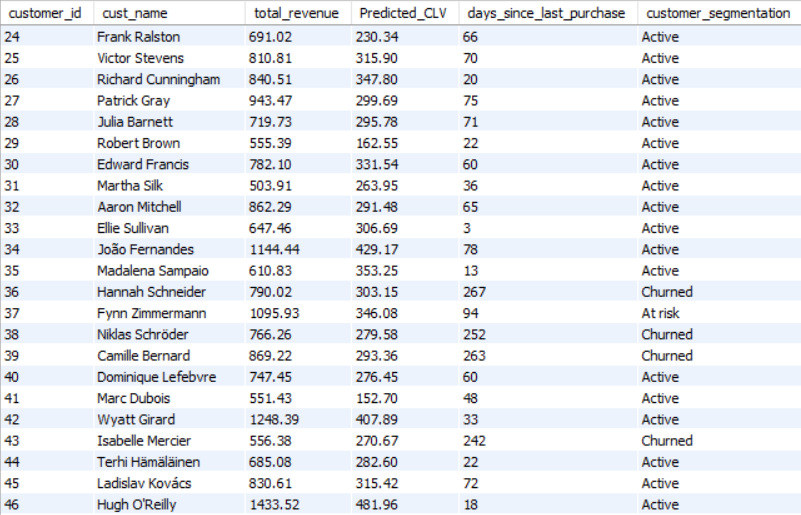
1. **Customer Lifetime Value Modelling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns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

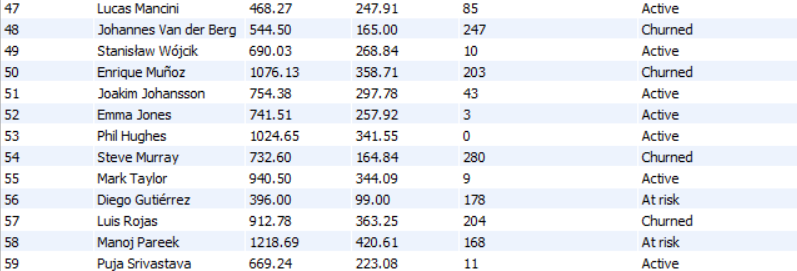
For the customers lifetime value is predicted on basis of their tenure years,

No of purchases and their average order values and these customers are categorized on basis of the days since their last purchase whether they are active churned or at risk









* For all the customers their actual revenue is greater than the predicted value which implies the expected revenue is greater than actual.
* The customers who produce great revenue than expected and belongs to the churned customers should be concentrated and need to arrange discounts and special offers for the loyalist.
* This helps to identify high value segments for marketing and loyalty programs while low value segments can be engaged with good offers to increase profit

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?**

* Similar to the above analysis I will identify the key attributes in discounts, events, email and marketing tables and clean data.
* Split customers and sales data into two groups: those exposed to promotions and those who weren’t.
* And next I would measure new customer sign-ups or first-time purchases during the campaign period.
* Demographic trends of new customers
* And next I would analyse which country customers are acquired after promotions this gives where further promotions and marketing need to be done.
* Compare total sales, average order value, and basket size during promotional periods with non-promotional periods.
* From discount table, I would analyse the discounts and revenue like where discounts provided is more and revenue is great and vice versa to make changes.
* Next I would check the marketing in less performing countries and where improvising should be done .
* Analyze the retention rate by comparing the number of repeat purchases made by customers during and after promotional periods.

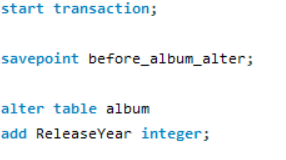
1. **How would you approach this problem, if the objective and subjective questions weren't given?**

* My approach would be almost similar.
* Firstly, I would understand the data, key entities customers, invoices, tracks, albums, artists, and genres and their attributes.
* Next cleanings the data which includes identifying null values, replacing null values and check duplicates if present.
* For each country, I will do analysis of revenue, no of purchases, average order value of customer and I will check which countries should be concentrated and which need to be ignored.
* Next, I would go with analysis of genres, album and artists for each country.
* Above mentioned step help for customers recommendations for each country.
* And next I would see the customers who are inactive and make good recommendation based on their previous purchase and also what are famous in their country.
* Identify the long-term customers and making special offers.
* Next, I would look for product affinities and cross-sell opportunities.
* Identify growth opportunities, optimize pricing, and develop targeted marketing strategies.
* Summarize findings with visualizations and provide actionable recommendations.

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

Adding a new column can be done using alter command

I also used TCL commands because we do not want the release year column anymore we can rollback to the original state



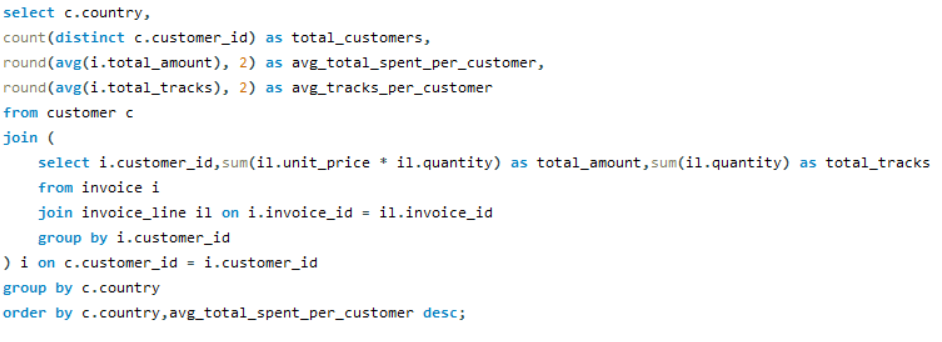
After inserting new column values can be inserted using

insert into album(year)

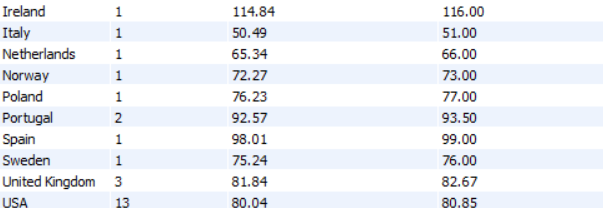
values (),……..;

1. **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.**

For each country no of customers, average amount spent per customers and average tracks purchased per customer is calculated using below query







* Countries like Czech Republic, Ireland, Spain, Chile have a high average total spent and average track purchased per customer but with 1-2 customers.
* Argentina, Denmark and Italy market is not that good. Survey’s need to be done to improve.
* Remaining all countries are good.