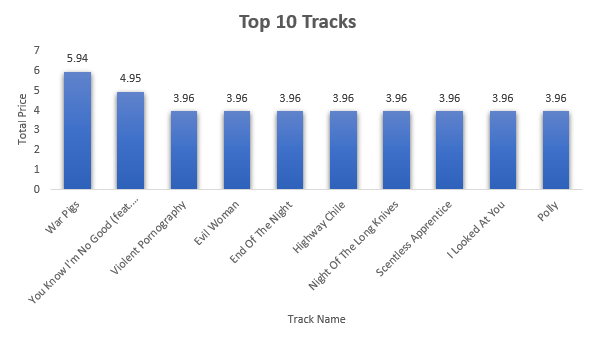
Objective Questions

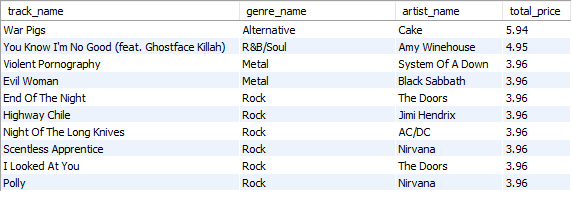
1. Does any table have missing values or duplicates? If yes, how would you handle it?

There are some tables which have null values in columns like customer, employee, invoice, track tables. Those columns were not meaningful to the analysis so I dropped them.

There were duplicates in playlist table, so I remove them using alter clause. Data cleaning, I have done in data\_cleaning.sql file.

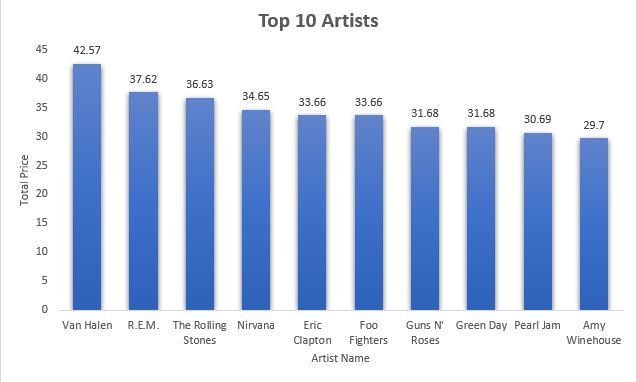
1. Find the top-selling tracks and top artist in the USA and identify their most famous genres.

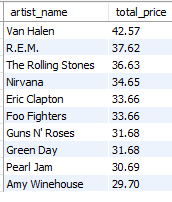
Top 10 selling Tracks



These are Top 10 selling tracks in USA. Most famous is “War Pigs”

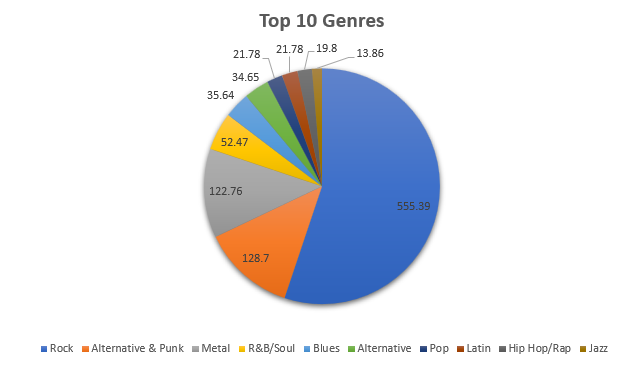
Top 10 Artists

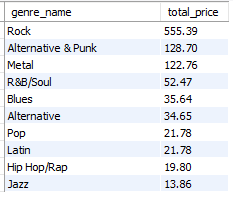




These are top artists in USA and most famous is Van Halen

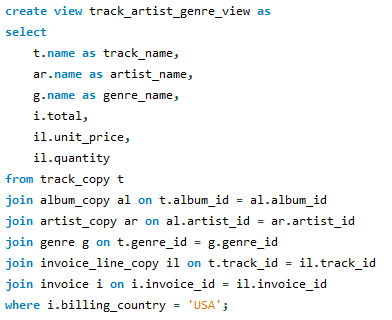
Top 10 Genres

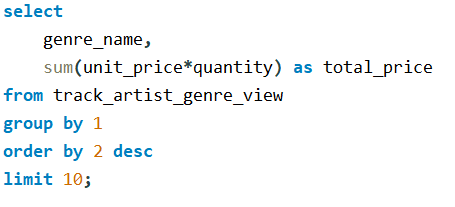
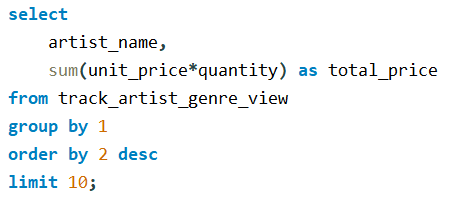
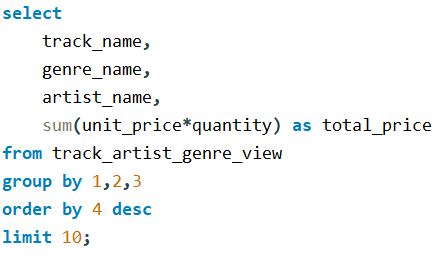




These are most famous genres in USA and Rock is most famous among them

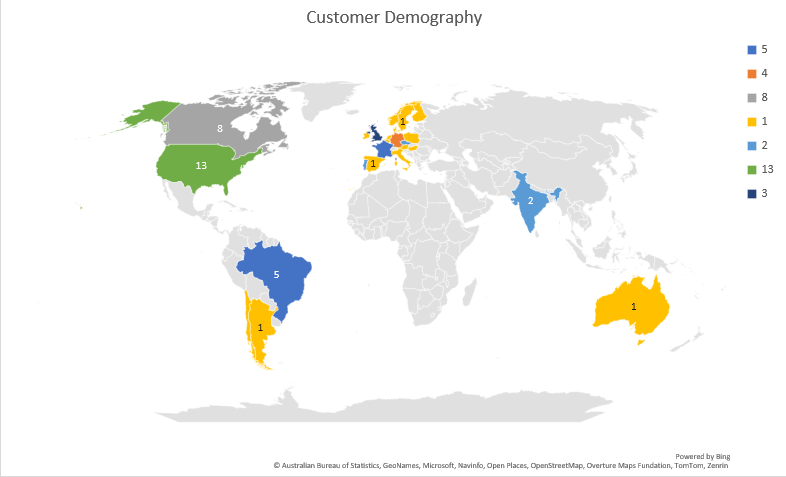
Queries performed

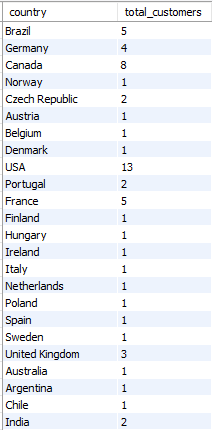




1. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?

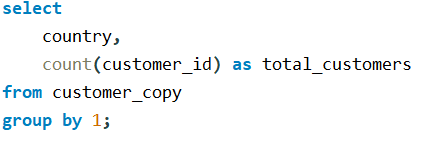
Demography is shown according to total customers country wise



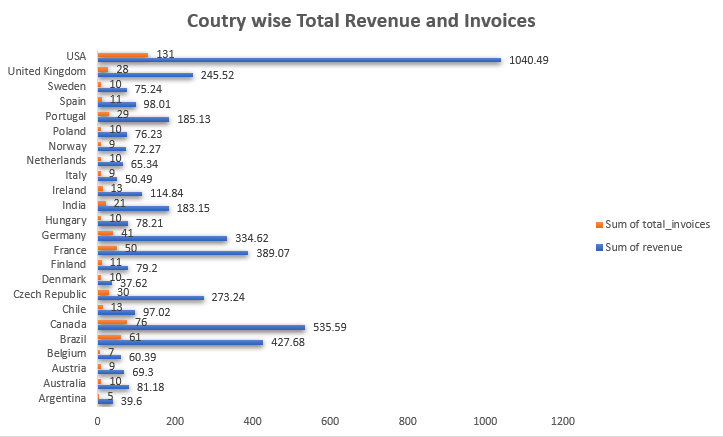


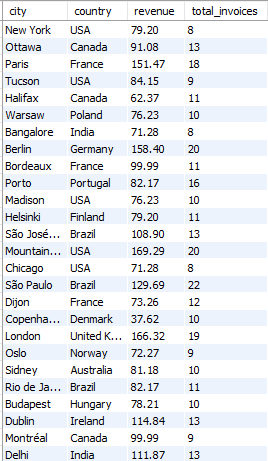
USA has most customers buying the Physical copies.

Queries performed



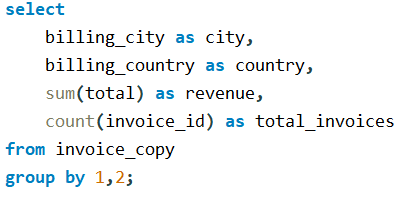
1. Calculate the total revenue and number of invoices for each country, state, and city:





From the above data we can see that USA has most number of Revenue generated and most number of invoices.

Queries performed



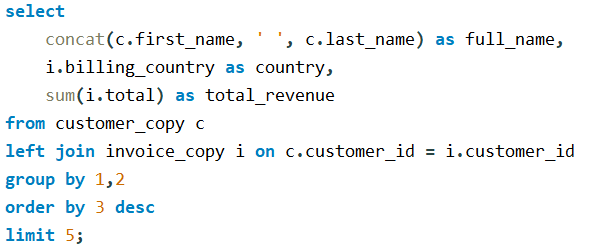
1. Find the top 5 customers by total revenue in each country



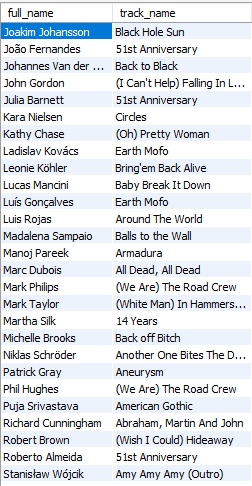
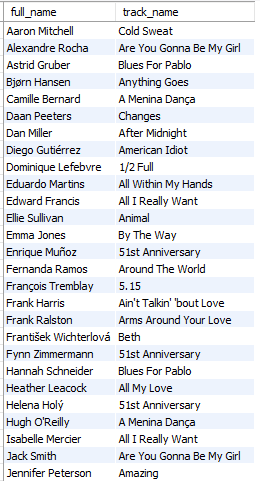


Top 2 customers are from Czech Republic but it is very interesting to identify that USA has generated most revenue but none of its people present in this top five

Queries Performed

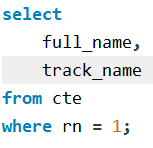
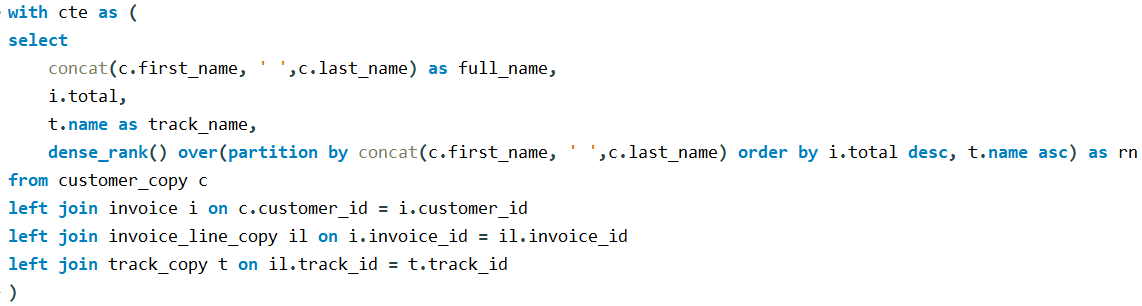


1. Identify the top-selling track for each customer

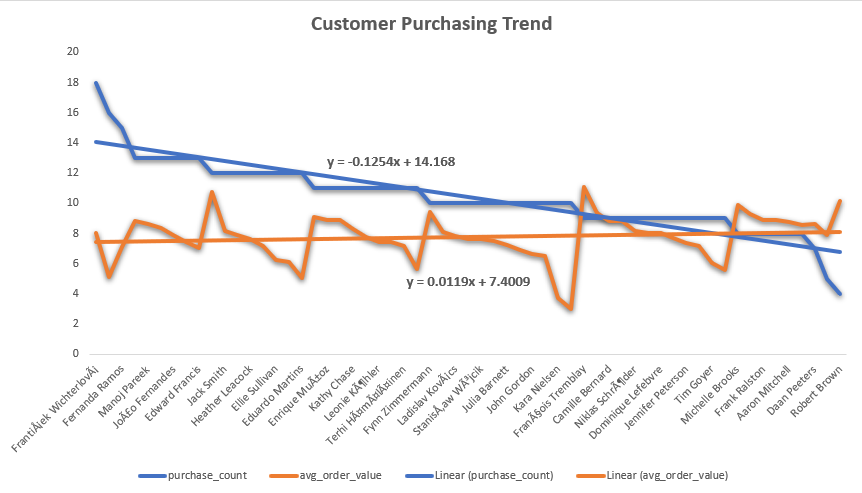


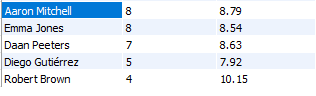
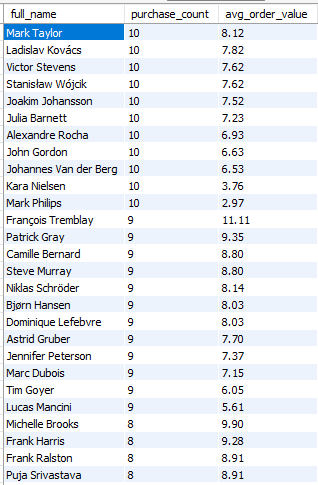
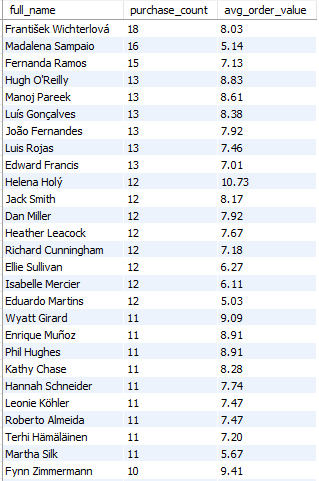
This shows the top selling track for each customer and ranking is done on the basis of total and name of the track.

Queries Performed



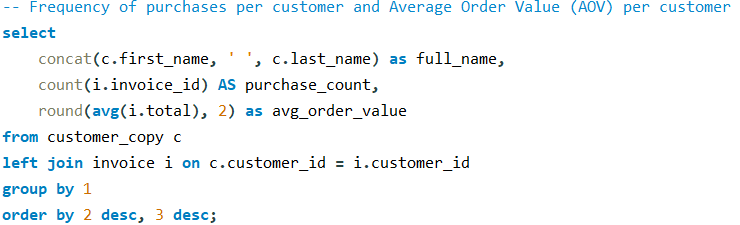
1. Are there any patterns or trends in customer purchasing behaviour (e.g., frequency of purchases, preferred payment methods, average order value)?





From this trendlines we can understand that in general, when purchase count decreases the purchasing power i.e. average order value increases

Queries Performed



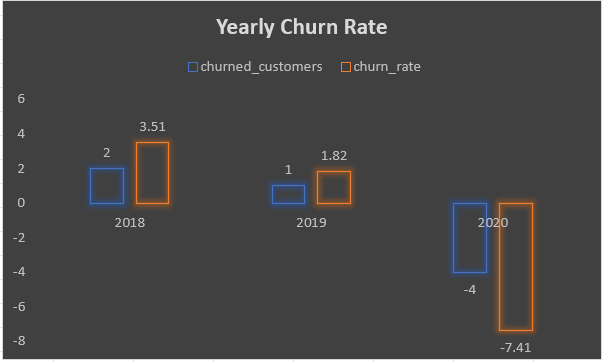
1. What is the customer churn rate?

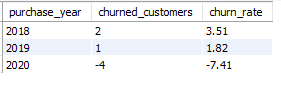
The **Customer Churn Rate** is the percentage of customers who have stopped making purchases within a given period (e.g., the last 12 months). It helps in understanding customer retention and loyalty.

Formula is   
Churn Rate =

(Number of customers who haven’t purchased in the last year) /

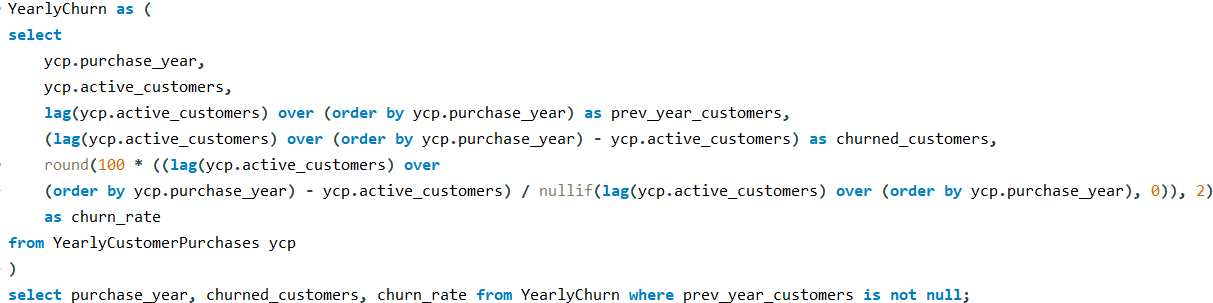
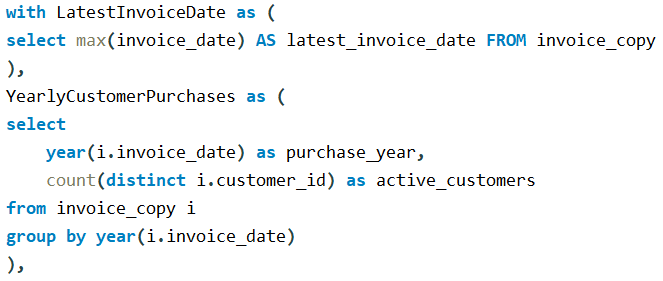
(Total Number of Customers) \* 100



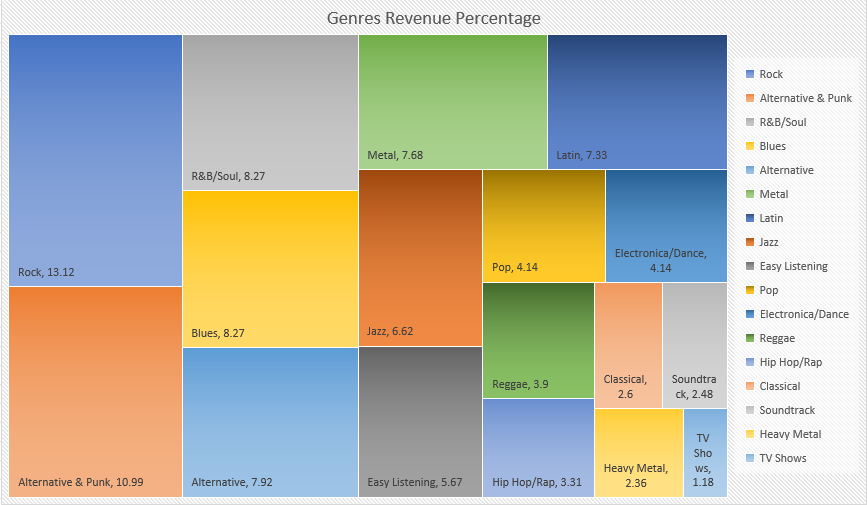


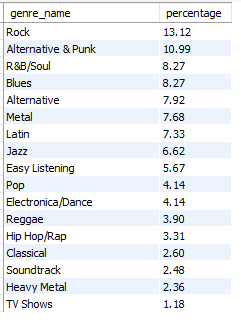
This trend in churn rate shows that company is very much able to hold on its customers. That -4 value in 2020 is the clear indication of that

Queries Performed are



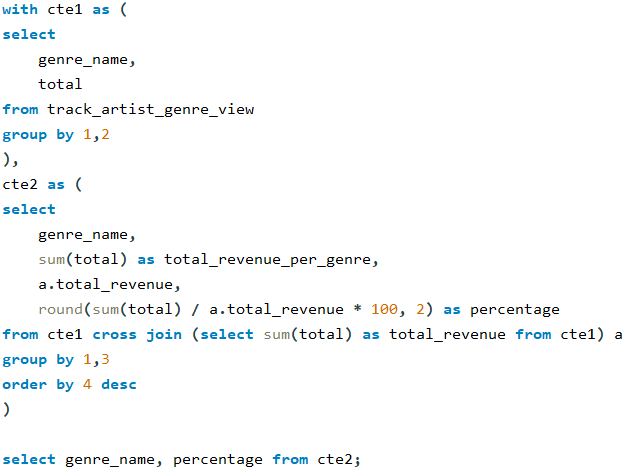
1. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.



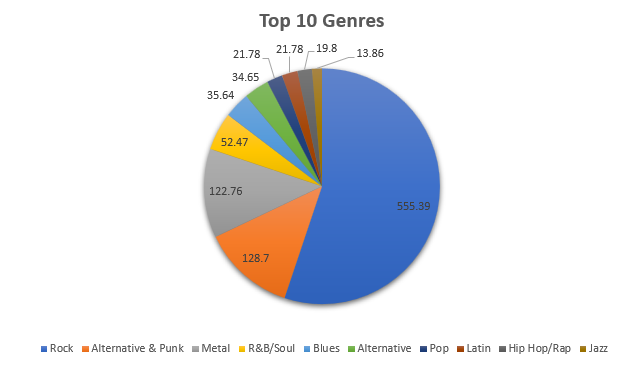
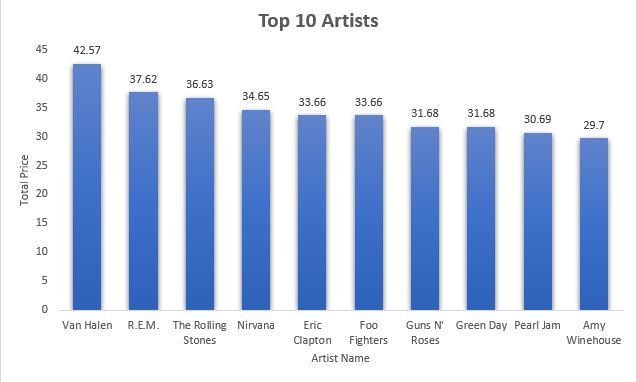


The above chart shows the most popular genres in USA with there percentage. With “Van Halen” being the most sold artists, the person is not from Rock Genre background.

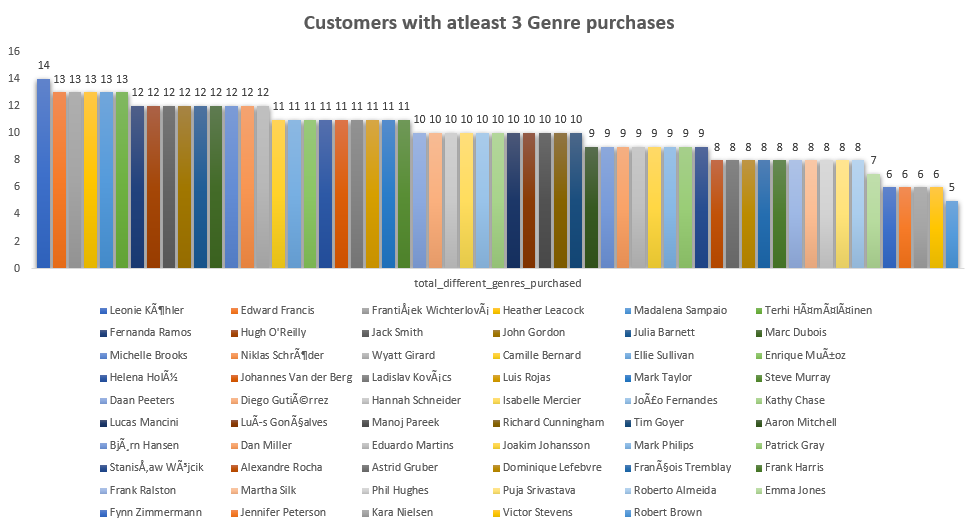
Queries performed

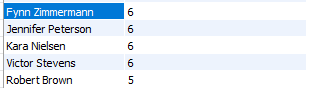
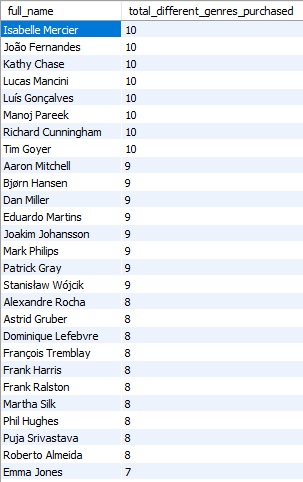
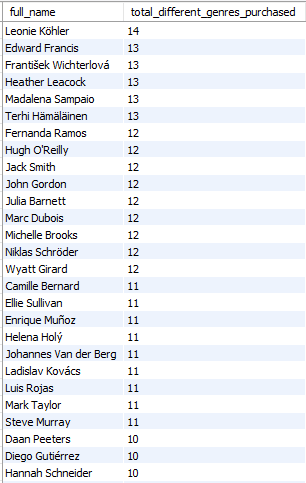


Best Selling Genres and Artists



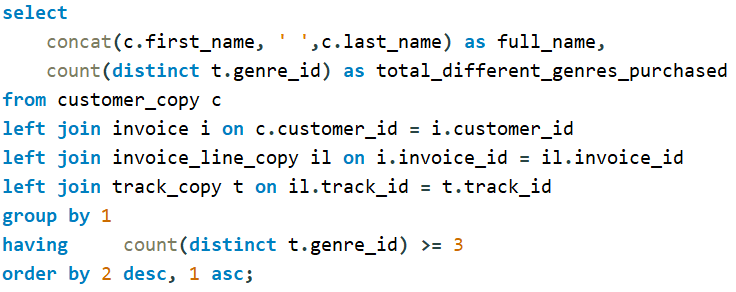
1. Find customers who have purchased tracks from at least 3 different genres



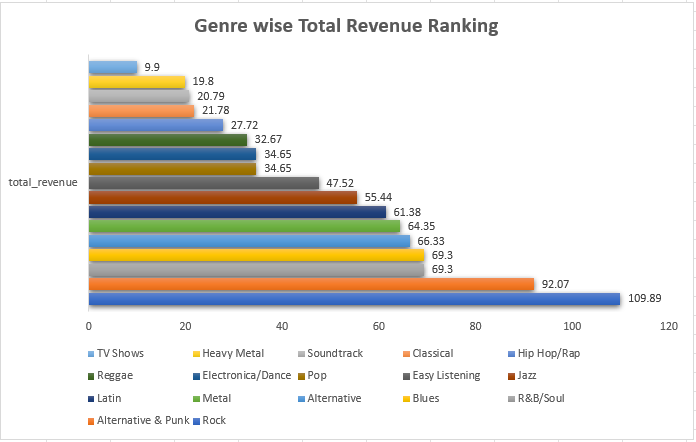


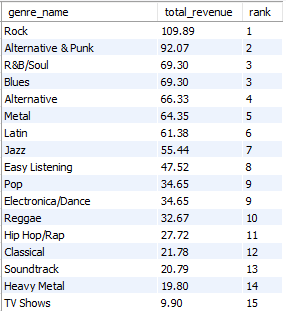
“Leonie Kohler” is the only person who purchased 14 different genre titles.

Queries performed



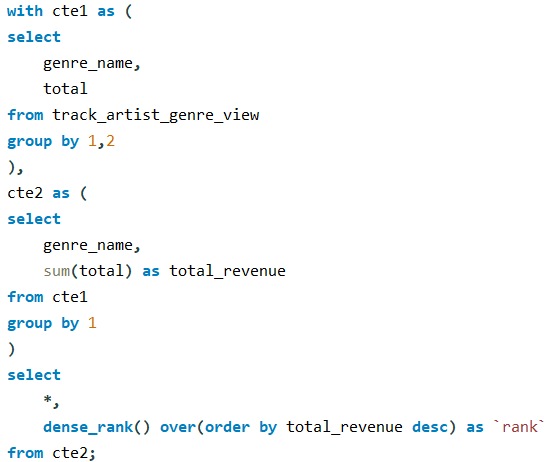
1. Rank genres based on their sales performance in the USA





“Rock” is most sold genre is USA with total revenue of 109.89

Queries Performed

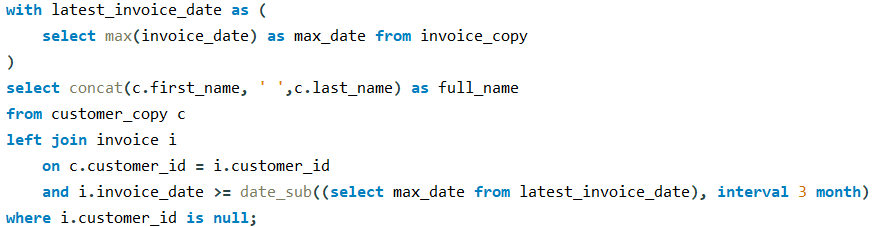


1. Identify customers who have not made a purchase in the last 3 months.

These are the full names of Customers who haven’t purchased anything in last three months in respect of earliest invoice date in whole data



Queries performed are

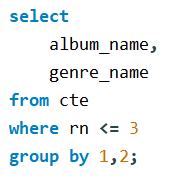
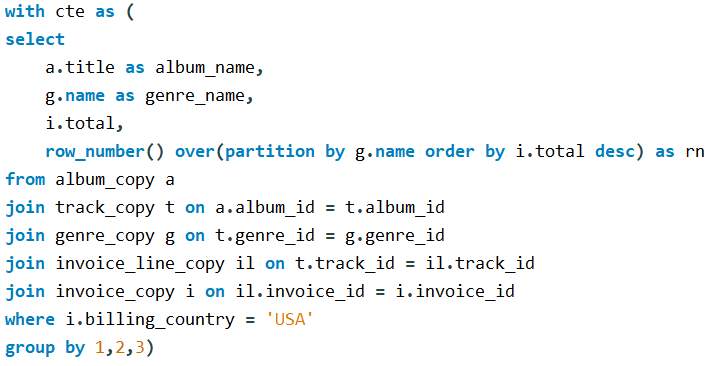


Subjective Questions

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.

These are the Albums Genre wise who should be prioritised for advertising and Promotion



Queries performed are  
 

Approach: –

* Ranking is done on the basis of revenue generated by titles in each genre. Top 3 are picked in each of them as you can see in the where clause.
* row\_number() ranking function is used because many titles will get same ranking so getting only 3 is the important thing.
* Since in question it is asked for USA then appropriate where clause is used on billing\_country in the CTE created.
* Partition is done on genre name and ranking is done based on total of that genre.

Insights: –

* There are certain Genres like “TV Shows”, “Hip Hop” etc. where only one title is present.
* Since the data is till 2020, so by then also world didn’t move from the original/traditional genres like Rock, Blues, Pop etc. There popularity is still intact and always will be.

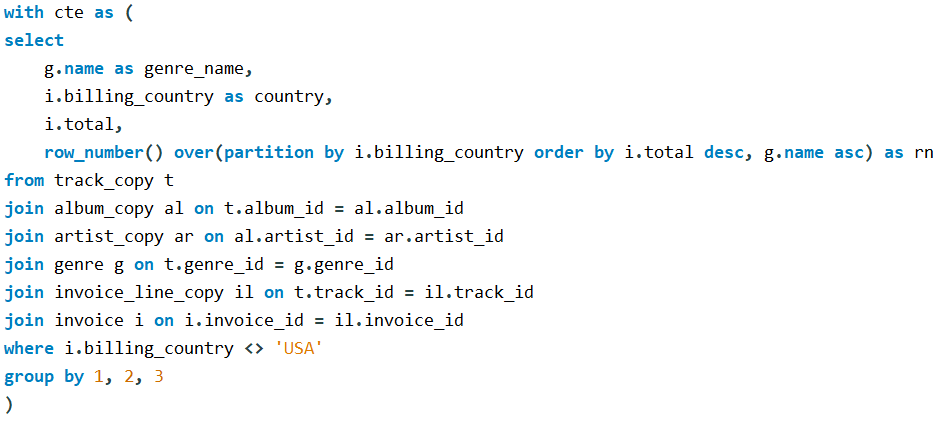
Recommendation: –

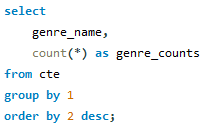
* These are the titles which help company to set up their events and generate revenues for the future.
* The artists who created these albums are most important person as tying up with them and hosting events in which they are chief guest, public will come surely.
* They also need to focus which genres are more popular so that more focus can be given to those.

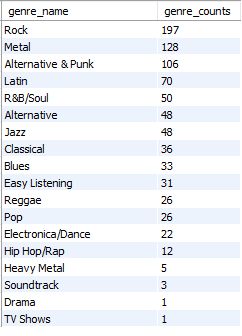
1. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

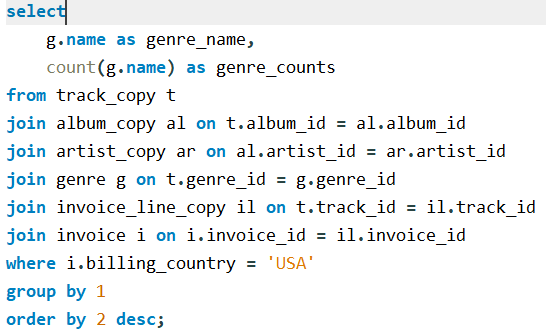
To find the top selling Genres from each country outside USA,

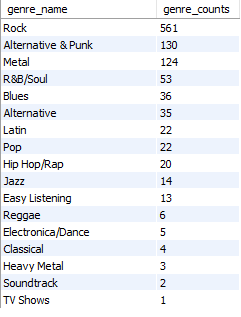
Queries performed are

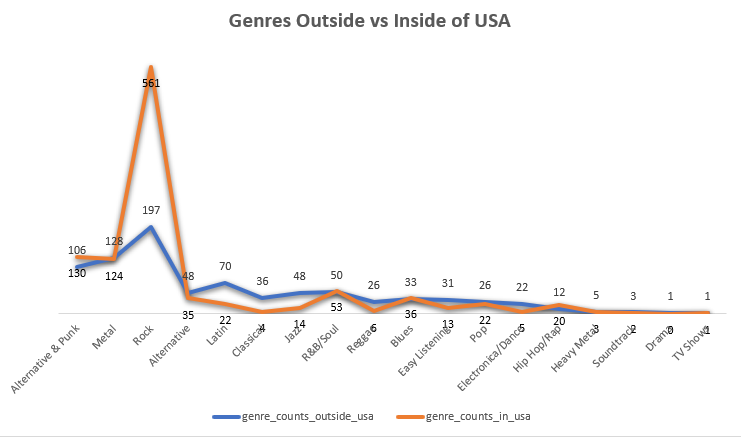












We can see from the graph that Rock is most popular and the trend remains almost similar after R&B/Soul Genre.

Approach: –

* Both queries performed are almost same with just one logic change, either country is equal to USA or not equal to USA.
* In first query condition is put on that country is not USA and I put that into a CTE.
* In second query, for comparison I just changed the billing\_country = USA logic, else everything is same.

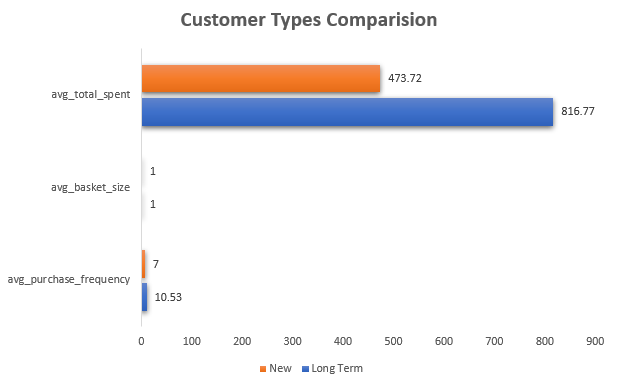
Insights: –

* Rock is being the most successful genre all over the world with the count of 197 albums/tracks bought while TV shows being the least with the count of just 1.
* In USA, people bought 561 copies in Rock genre which more than total of other countries making it most famous worldwide.
* In top three genres all over the world, there is Rock, Metal and Alternative & Punk.

Recommendations: –

* These values will help company to devise campaign strategies according to region they are going to promote the albums or tracks.
* Genre having above average copies sold will be crucial for the company as they are the ones generating money for company. So, Company needs focus on them.
* Creating relationships and brand ambassadors with Highly Rated artists in these genres will very much help in events that company will organise.

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?

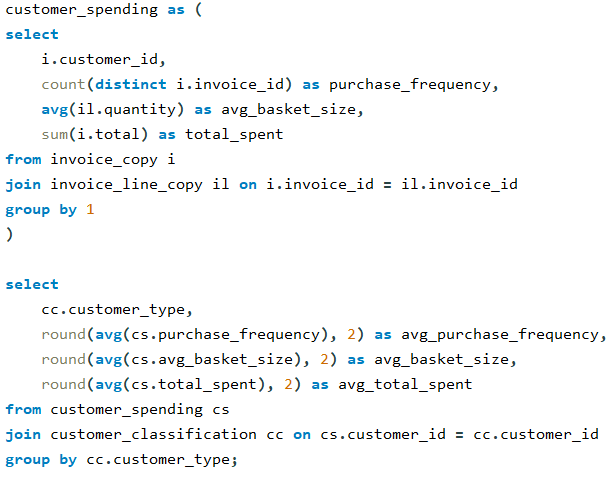
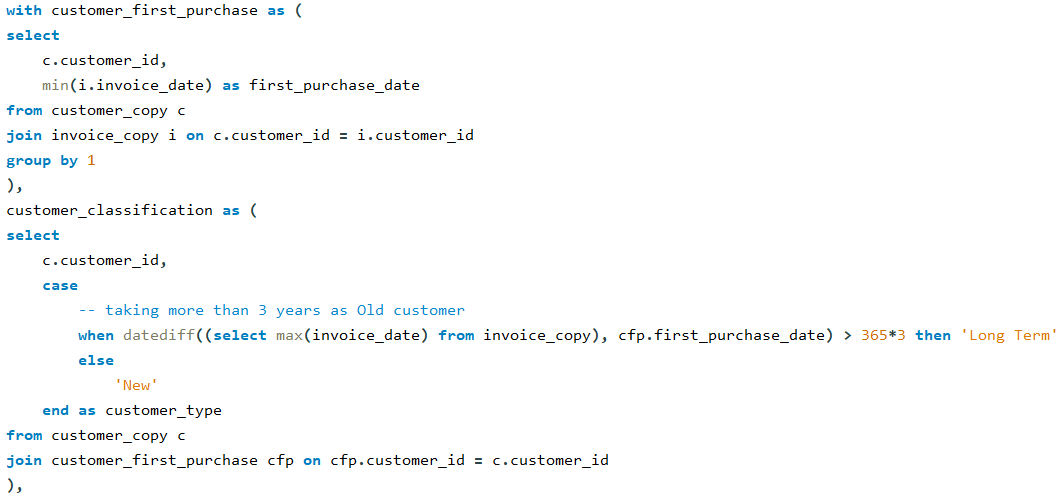


For long term customers which I have taken as more than 3 years since their last purchase, Average total spent and purchase frequency is higher, as there were more long-term customers than new customers.

This shows that company is able to retain long term customers.

Also, Average Basket size is same for both indicating that neither long term nor new customers bought tracks in bulk.

Queries performed are





Approach: –

* To categorize customers into long term and new I used the period of 3 years from last purchase of each customer.
* 3 CTEs with names customer\_first\_purchase, customer\_classification and customer\_spending created and then joining is done to get data in one place.

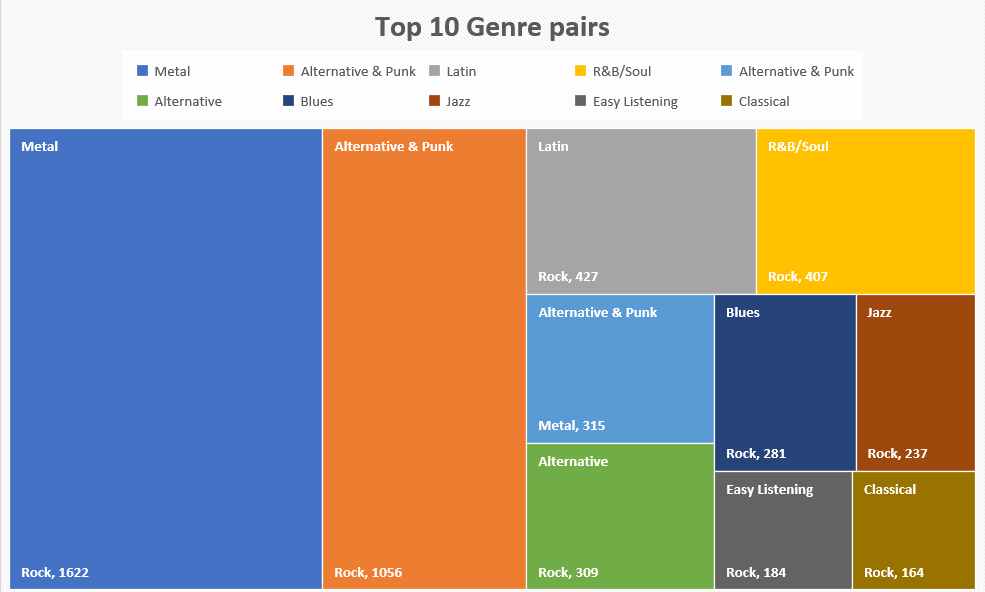
Insights: –

* There is no change in average basket size as in the data most tracks/albums were bought in quantity of 1.
* Long term customers are generating more revenue which is good sign. The problem is having low average basket size implying customer is looking for some different media or does not want to buy many physical copies.

Recommendations: –

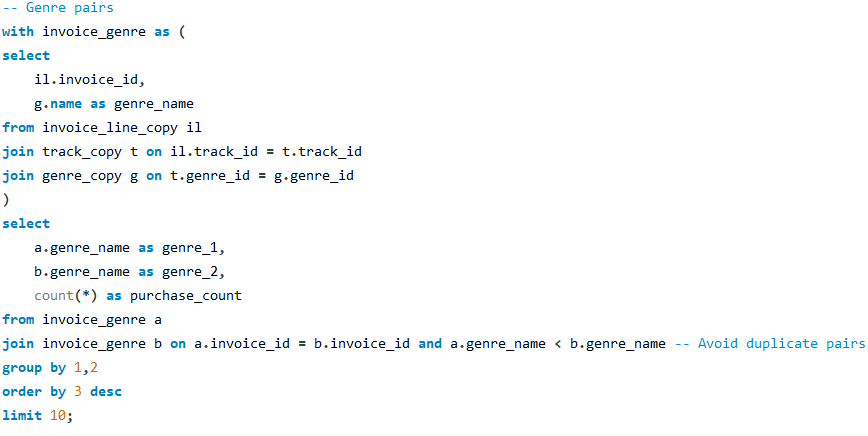
* Company need to devise some strategy to increase customer acquisition either by promoting their best sellers or starting new venture like going online.

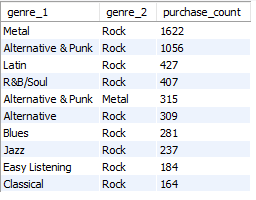
1. Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?

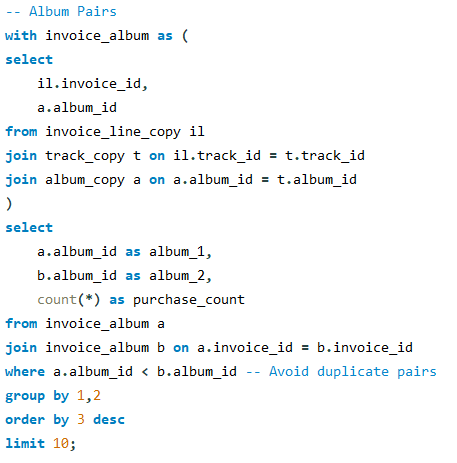


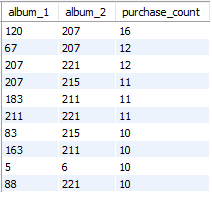
These are the Top 10 Genre and album pairs which were bought together. The high selling counts indicate that people love to listen these pairs together and this will help Chinook employees to recommend customers best albums and Genres and also they can put sale on these individual Genres and Albums.

Queries performed are









Approach: –

* To create unique genre and album pairs I used self-join and one condition in on clause which gives uniqueness to pairs.
* Both the Genre and Album queries are similar difference being the joined tables but the track table is common joined table in both.
* The condition of a.genre\_name < b.genre\_name or a.album\_id < b.album\_id helps to avoid duplicates as well as sorts the data.

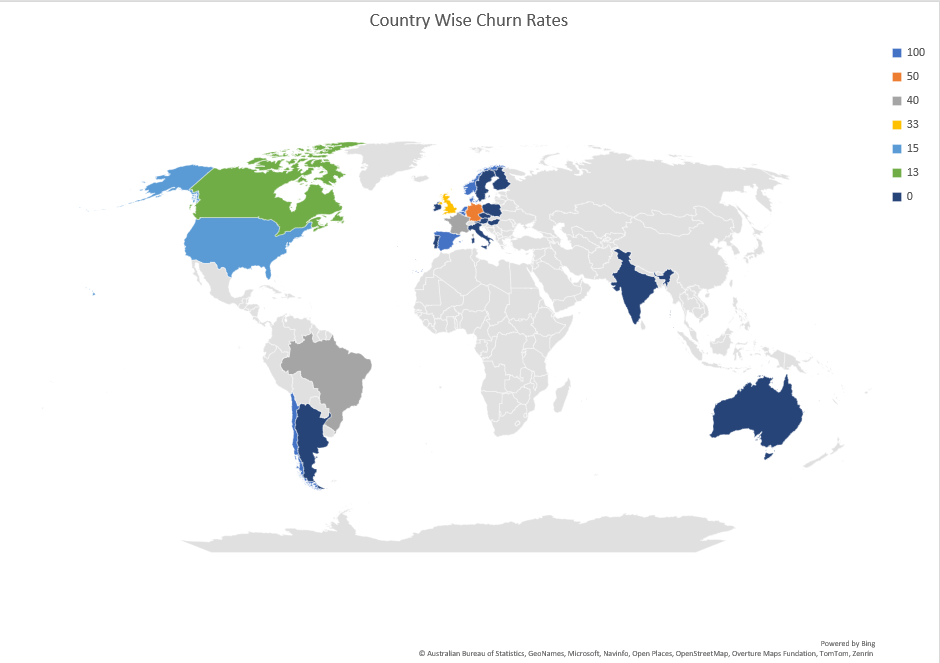
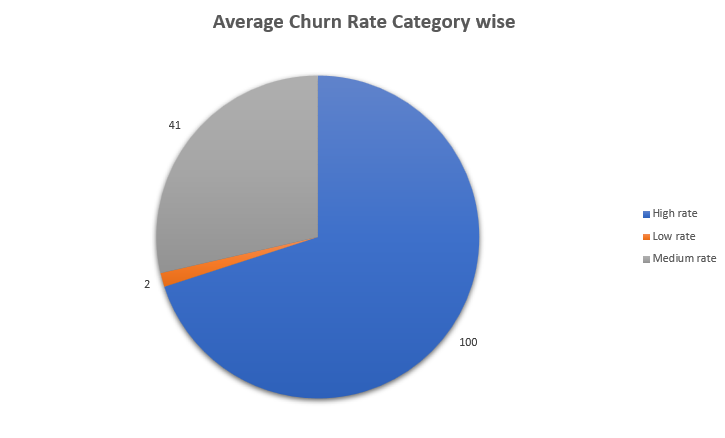
Insights: –

* In top pairs “Rock” genre is paired most of the time signifying that this genre is most bought.
* Similarly, album ids like 207, 211, and 215 are most paired.
* Metal with Rock paired is most bought with 1622 counts. 10th is Classical with Rock paired having 164 counts.
* Album ids 120 and 207 are most bought in pair with 16 counts.

Recommendations: –

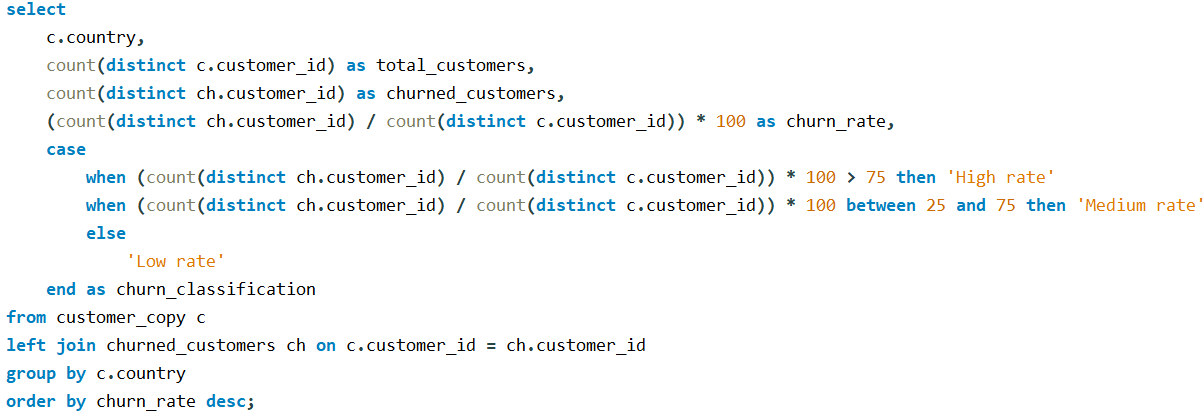
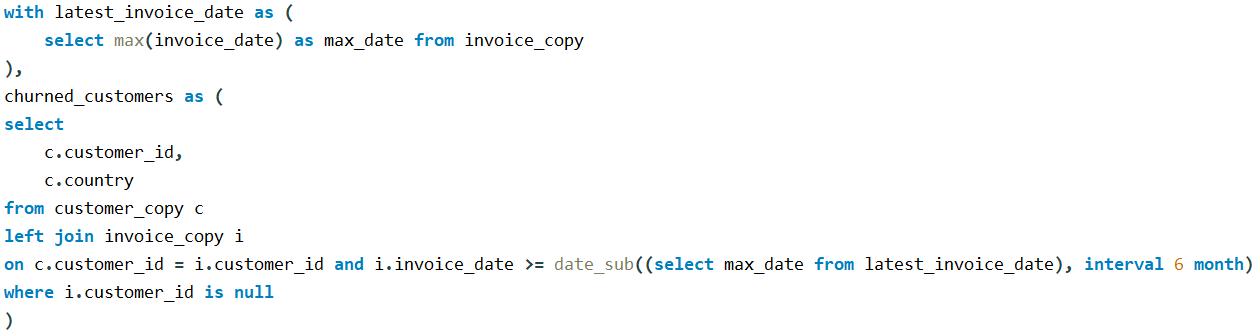
* These pairings will help company to decide discount offers and which album/genre to promote.
* Since “Rock” is the most famous genre around the world, company tie up with Rock Artists and stage special events for these top Genres to promote them more around the world.
* They could also start Rock Music production as a side venture. This can help in more inflow of customers

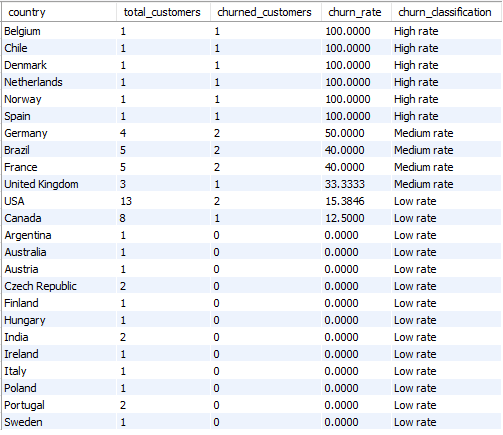
1. Regional Market Analysis: Do customer purchasing behaviours and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

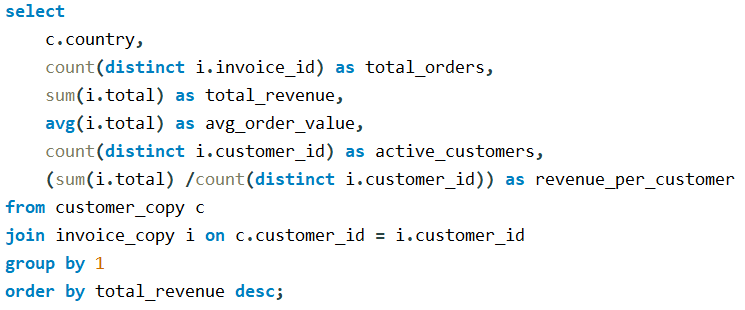


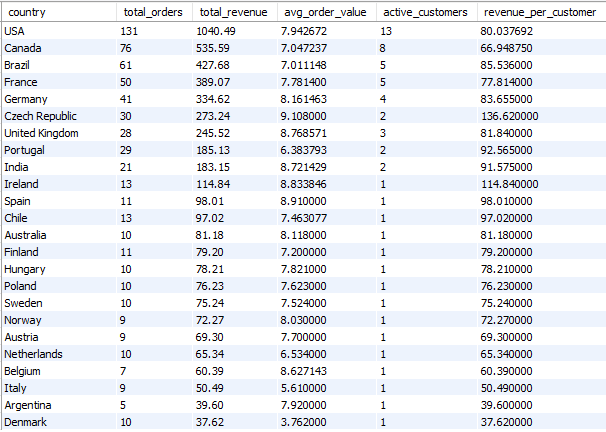
The churn rates here are calculated in period of 6 months starting from earliest invoice date to latest date in whole data. As we can see that 0% churn rate is in many countries like USA, Canada, India etc. These countries have strong economic and financial sectors which allows people of these countries to buy things rather easily. There can also be a factor of love of music and rate at which music is produced in a country.

Queries performed are









Approach: –

* Initially I calculated country wise churn rates in period of 6 months using the churned\_customers cte.
* Then Using the Case Statement I categorized the countries in to High, Medium and Low Rate.
* Then to see how company is doing in each country I calculated revenue, count of sales, revenue per customer, active customers and Total orders.

Insights: –

* All countries with either Low or Medium risk, there revenue values show why they are categorized in that. Meaning most of these companies have good revenue rate and good sales, so both the output tables complement their results for each other.
* There are some anomalies like Argentina where revenue generated is less but it is under low risk. This may be due the Economy state or technological advancement of the countries where revenue generated is less but comes under low churn rate category.

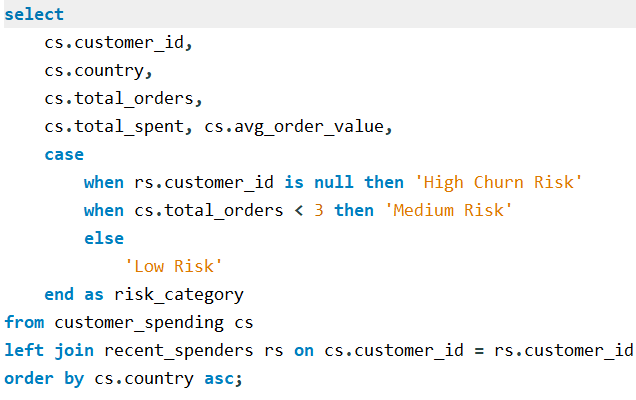
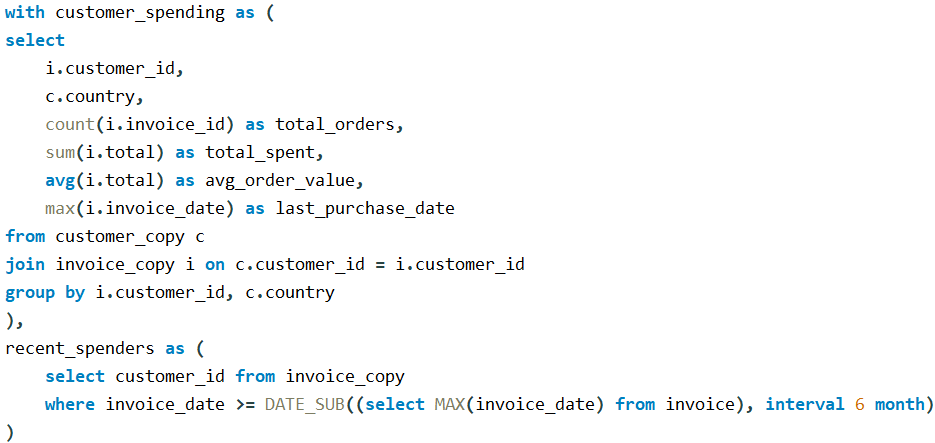
Recommendations: –

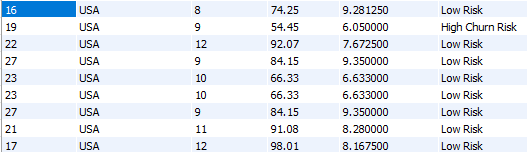
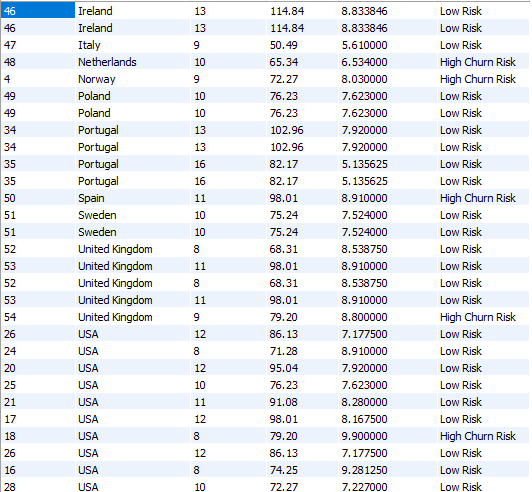
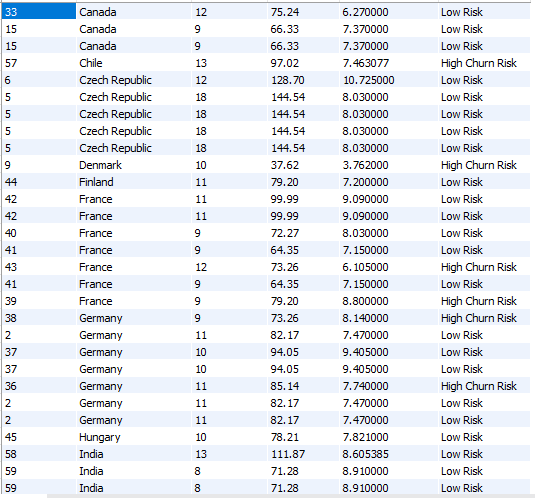
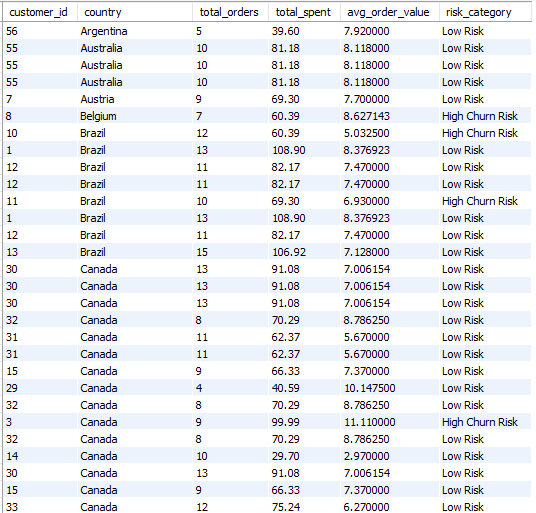
* Company needs to see whether their future in High-risk countries is good or bad as per the revues they are generating there. Since both the output tables comply with each other mostly so it is better to cease their business in High-risk countries like Belgium, Chile etc.
* They also need to devise strategies to not making the Medium Rate countries convert to high. For that Campaigns and hosting events are the best option.
* For Low rate countries, they need to expand their business idea from physical to online entertainment, to generate more revenue.

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?

Since the Gender, Age like data is not in customers table, so we can only profile on the basis of Country from which customer belongs.

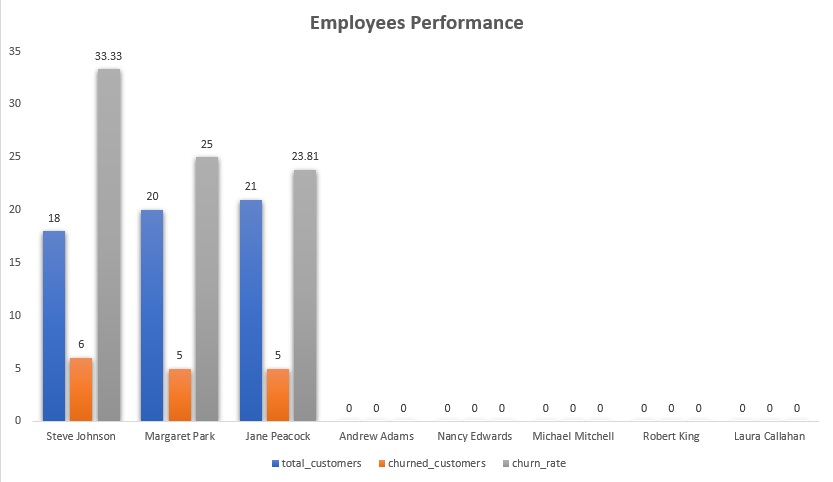
Queries performed are



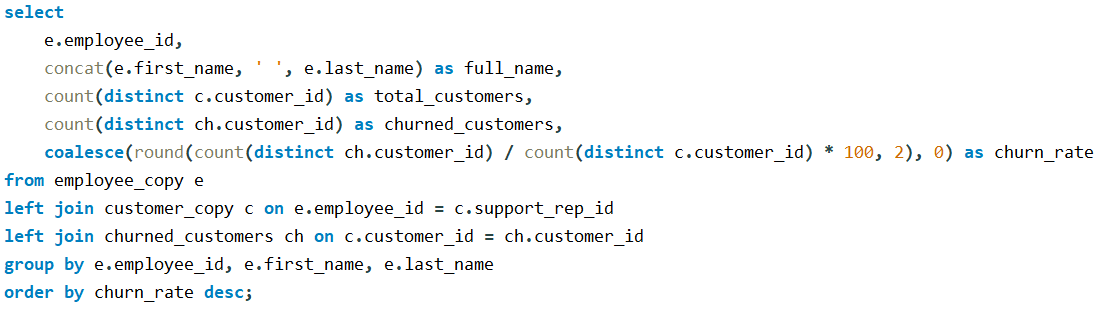
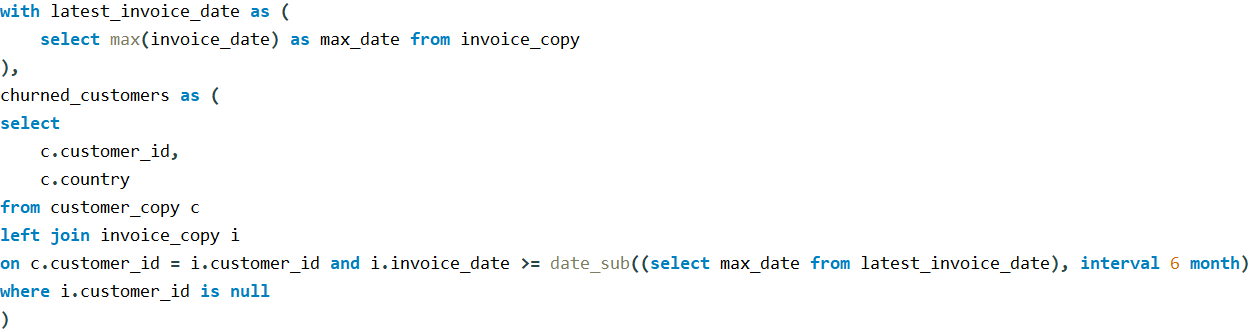


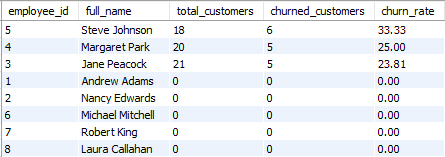
Above queries performed, gives you the idea which customers are at high churn risk and which ones are at low. They are also segregated between various countries to also see how business is doing in various regions. This helps in Customer Risk Profiling.

We can also do the profiling on the basis of services provided by employees, like which employees are causing the most churned rates, some action need to be taken on them.



Queries performed are





As you can see from the chart that only 3 employees are handling

all customers, this shows that what level of services company is providing.

Approach: –

* There are 2 parts to solve this question. In first one, churn risk is being calculated on the basis Recent spending of customers. The time period chosen is 6 months.
* In second I used Case statement to categorize customers as High Risk, Medium Risk and Low risk on the basis of conditions like customer id is null in recent\_spending cte then it’s high risk. If total orders are less than 3 the Medium Risk, others low Risk.

Insights: –

* There is no customer with Medium Risk category. It is good because Low risk people are more than high risk so company only needs to focus on the extremes nothing in middle of it.
* This categorization is made possible due to the time period I have taken. If it was for 1 year or 3 months, we might see the Medium Risk Category people.
* Total 8 employee’s data I have, in that also only 3 are working as the chart depicts. So, work management is very bad.
* Employee Steve Johnson has the highest Churn rate with lowest customers he is currently handling. This is a point of concern in respect of Employee’s performance

Recommendation: –

* Focus more on high-risk customers, try to get in contact with them and get their reviews on how we can satisfy your needs more accurately.
* Work on the management of employees, don’t give only certain employees so much work, it should be equally divided and for that company can establish mentorship programs.
* Need of monitoring of Employees performances to ensure better services to customers.

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?

**Customer Lifetime Value (CLV)** is a metric that estimates the total net profit a company can generate from a customer over their entire relationship with the business.

It can be predicted using this formula  
**CLV =**

**Average Order Value × Purchase Frequency × Customer Lifespan**

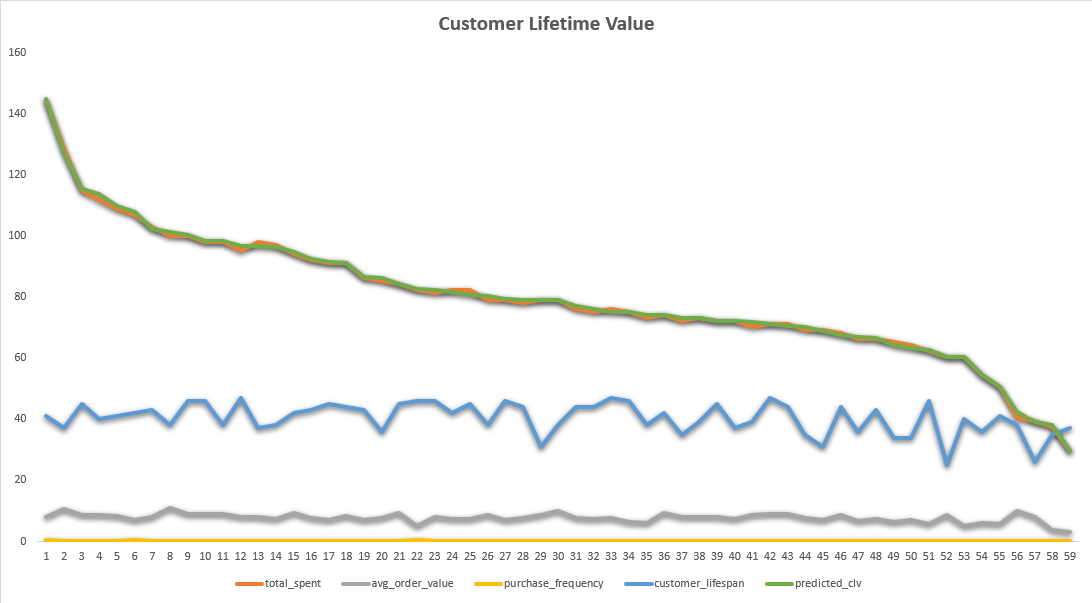
Where each term is

**AOV** = Total Revenue / Total Number of Orders

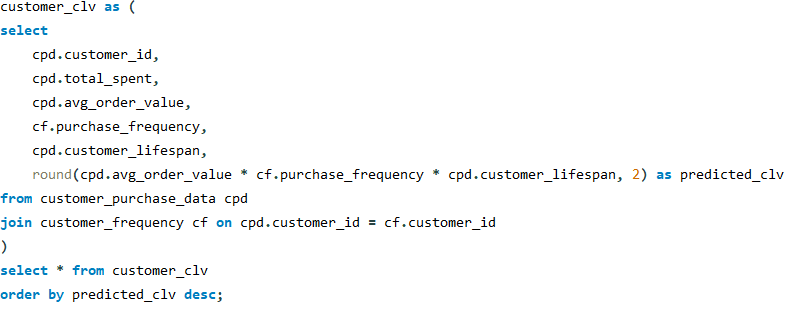
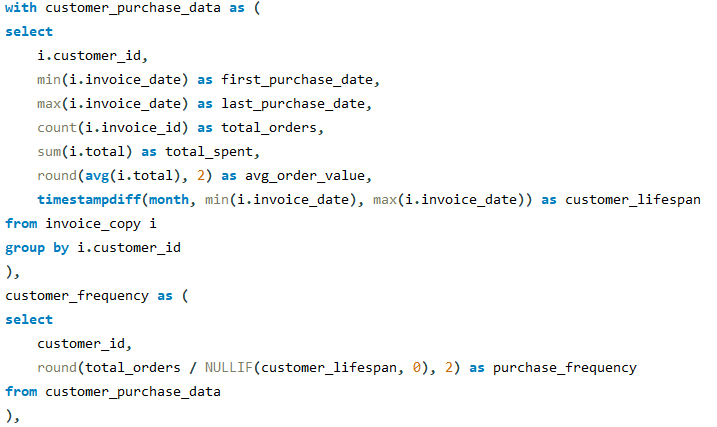
**Purchase Frequency** = Total Orders / Total Active Months

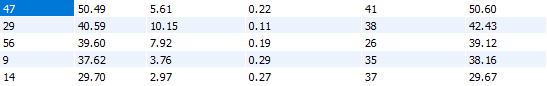
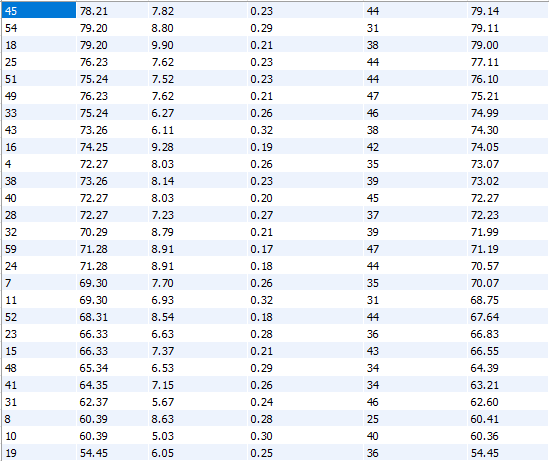
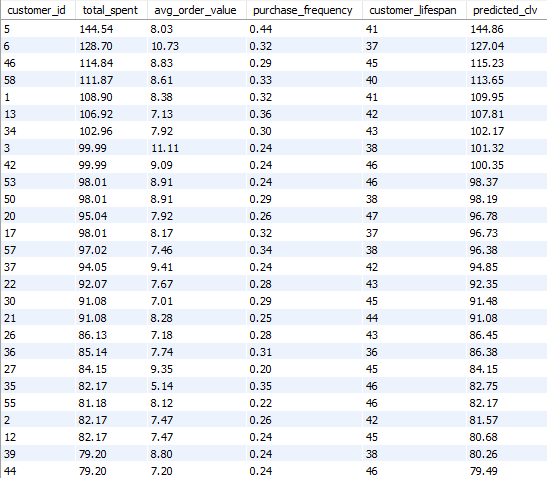
**Customer Lifespan** = Number of Months Between First & Last Purchase.

A higher CLV indicates that **a customer is generating more revenue for the business over time**, which is generally viewed as more desirable. This can mean that customers are purchasing more frequently or spending more per transaction, leading to increased profitability.



Queries performed to calculate this





From the line graph we can see that, total spent by each customer is almost similar to predicted CLV value signifying that formula is correct and data is reliable for analysis.

Approach: -

* For understanding more about this type of analysis I used Google and ChatGPT to get information on it and the formula I got is from Google only.
* On the basis of the formula, I formulated my query by creating 3 CTEs.
* Initially I summarised customer purchase data, then I calculated the purchase frequency and finally I calculated predicted CLV for each customer. Generally, its value is almost similar to total spent by customers and from the above graph we can see that clearly.
* Customer Lifespan is calculated in months using timestampdiff function and purchase frequency is Total Order / Customer Lifespan.

Insights: –

* CLV is almost similar to Total spent value which ensures data credibility.
* But the curve is decreasing curve and in between slope is constant but initially and in later stage slope falls sharply.
* So, for these customers where slope is falling sharply, these customers are losing interest in the company.
* Highest CLV is 144.86 and lowest is 29.67.
* Customer purchase frequency is below 0.5 for all which is bad sign.
* Average order value is almost constant but the Customer Lifespan curve showing more ups and downs toward later stages

Recommendations: –

* Company needs to take care of those customers where slope is declining sharply.
* Also need to keep in mind about the customer lifespan which is fluctuating many times. Campaigning and events are the ways to keep customers on their sides.

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?

First, impact on customer acquisition

* Promotional campaigns like Events, marketing and most importantly putting up discounts on Most liked albums and tracks will give large boost to Company’s revenue as they will be able to acquire new customers.

Second, impact on customer retention

* It is very important for these Retailing companies to retain their customers not only for their revenue purposes but also for future growthy of company in various parts of world. Lower the churn rate, meaning more people being retained and these campaigns are crucial for this.

Lastly the overall sales are direct results of how the company performed in customer acquisition and retention, they need to do good in at least in retention because if leaving rate becomes more than acquiring rate sales will decrease. Retaining old customers shows that company is being loyal, attentive and listening to what customer wants and what they seek.

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

Since the objective is related to increase sales in physical music market, my first opinion will be making it an online platform because in this 21st century everything is online, no one have that time to go and buy albums from store.

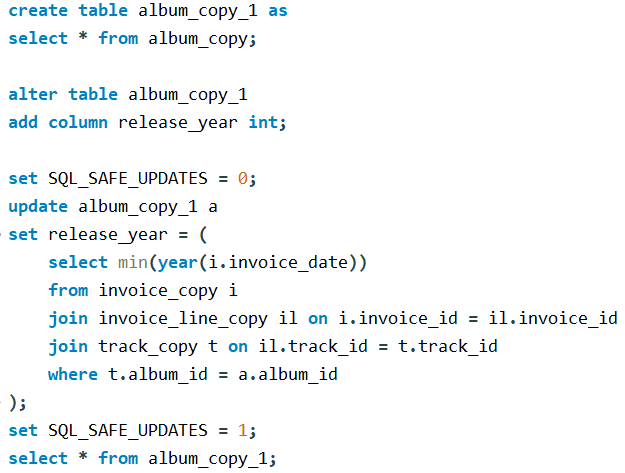
Even if the company insist on still doing most of its business offline but able to understand today’s online market need for the business growth, here are some areas I’ll analyse

* What are the Average sales of company and in which parts of the world it is doing great. From that I’ll be able to suggest that in this region you need to expand and, in these regions, you need to shut down.
* What is company’s churn rate or retention rate of customers if they are in bad numbers, company need to promote their best albums, start giving discounts and hold events and if possible, tie up with some best music artist to come up in their events.
* How well the company employees are doing, are they able to provide good service to customers because this will directly impact churn rate.
* I will also analyse the sales whether the customers are buying in bulk or the numbers are down. Here providing online services matter because buying albums is more costly than taking yearly music subscription to the online music platform. For increasing numbers same campaigning, organizing events, giving discounts and other things will increase sales.

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?

In this question I’m assuming that release year of each album as the year in which each album id bought earliest, just to put some value in that column.

Queries performed are

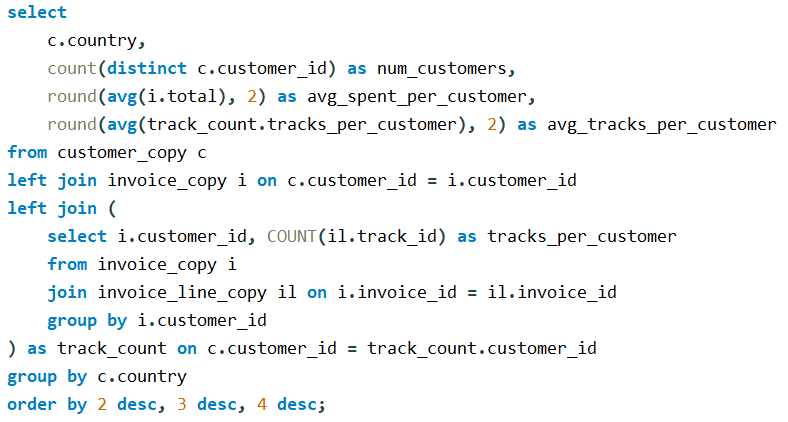


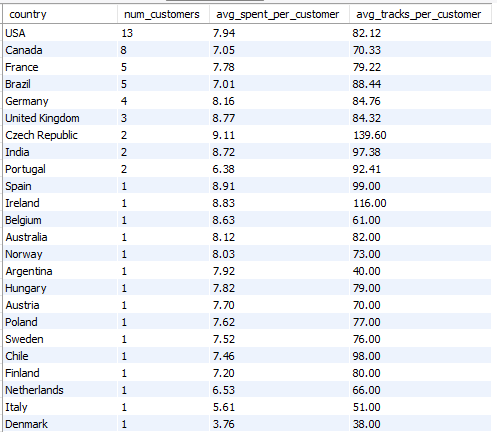
Approach: –

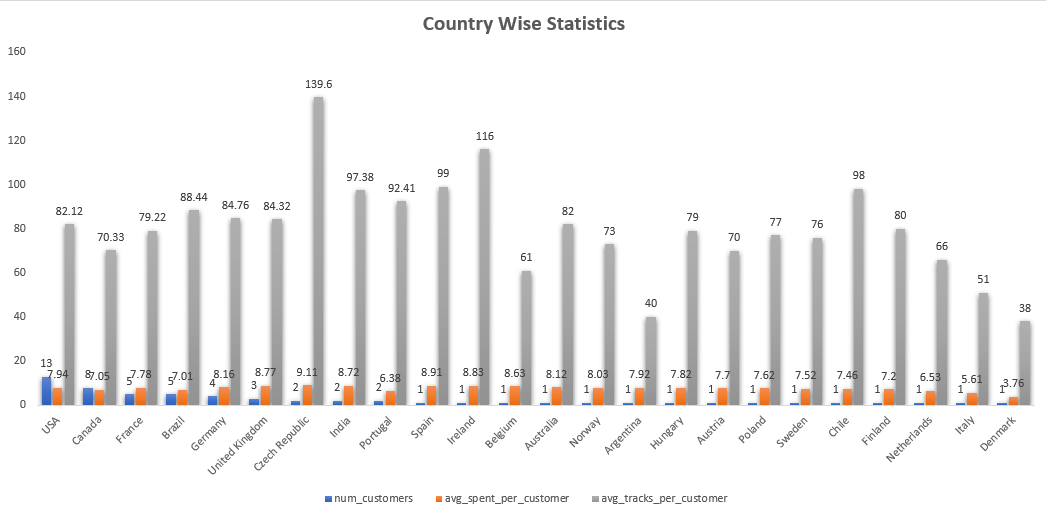
* The assumption that I have taken above is just to put some values in new column instead of null. For that I used subquery to get the earliest bought date of albums and before that I made a copy of the copied table in order to preserve my cleaned data.
* First, I created the release\_year column using alter clause.
* Then using the update clause I set the release year values to the earliest bought date of the albums which is calculated in the subquery.

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.

The SQL Query performed







Approach: –

* To get the average tracks per customer, created a temporary table named as track\_count in which I joined invoice and invoice\_line to get the customer\_id and tracks per customer mapping.
* Then performed several left joins with customer and invoice tables and calculated country wise total customers, average spent per customer and average tracks per customer and finally grouped by country

Insights: –

* USA has most customers with the count of 13
* Czech Republic has most spending (average spend = 9.11) and most average number of tracks bought which is 139.6.
* India is also not far from the race but it’s the only country with Czech Republic representing Asia.

Recommendations: –

* Company needs to do sentimental analysis for love of music in various parts of world as they are selling music only.
* Music in my view is only commodity which is loved by everyone and in this online era people are still buying it may be for collection purposes or maybe they are cheap or whatever the reasons are. For a company is its customers happiness what matters so they need to focus in the regions where they are actually getting revenue, sales and low churn rate.