**Learners have to come up with a Report to support the answers to the following questions and suggestions**

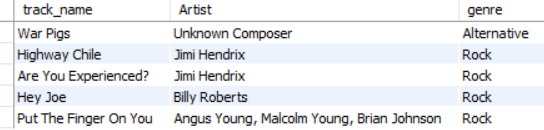
Objective Questions

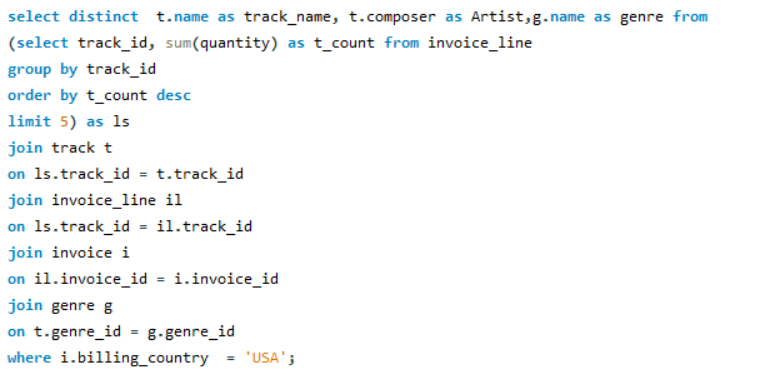
1. Does any table have missing values or duplicates? If yes how would you handle it?

* For checking duplicates, I would run this query for every table.
* Let’s consider the table ‘Customer’, here customer\_id is primary key so it should not repeat.  
    
  **“select count (\*) from customer group by customer\_id having Count (\*)>1;”**

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

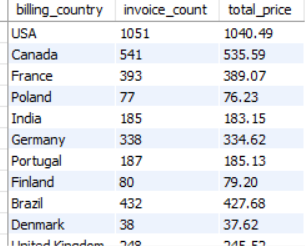
The top selling tracks and artists and famous genres in USA are



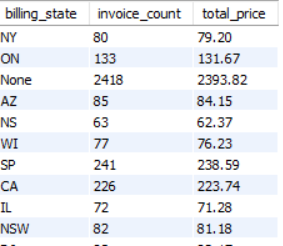
Code: 

1. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?
2. Calculate the total revenue and number of invoices for each country, state, and city:

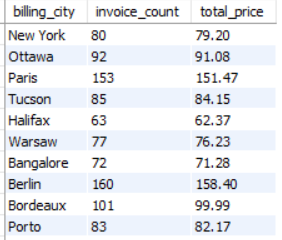
Revenue of Each Country



Revenue of Each State

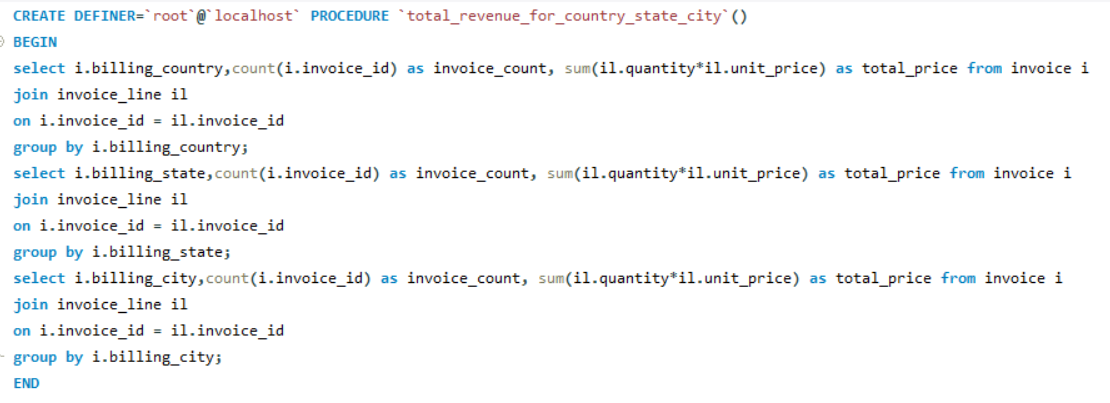


Revenue of Each City



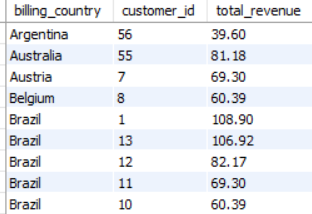
Created a procedure where it shows revenue by country,city and state

Code:

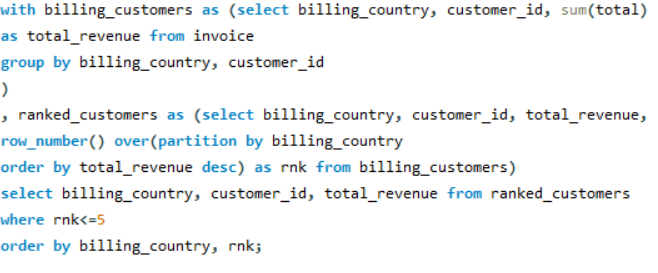


1. Find the top 5 customers by total revenue in each country

Top 5 customers of each country



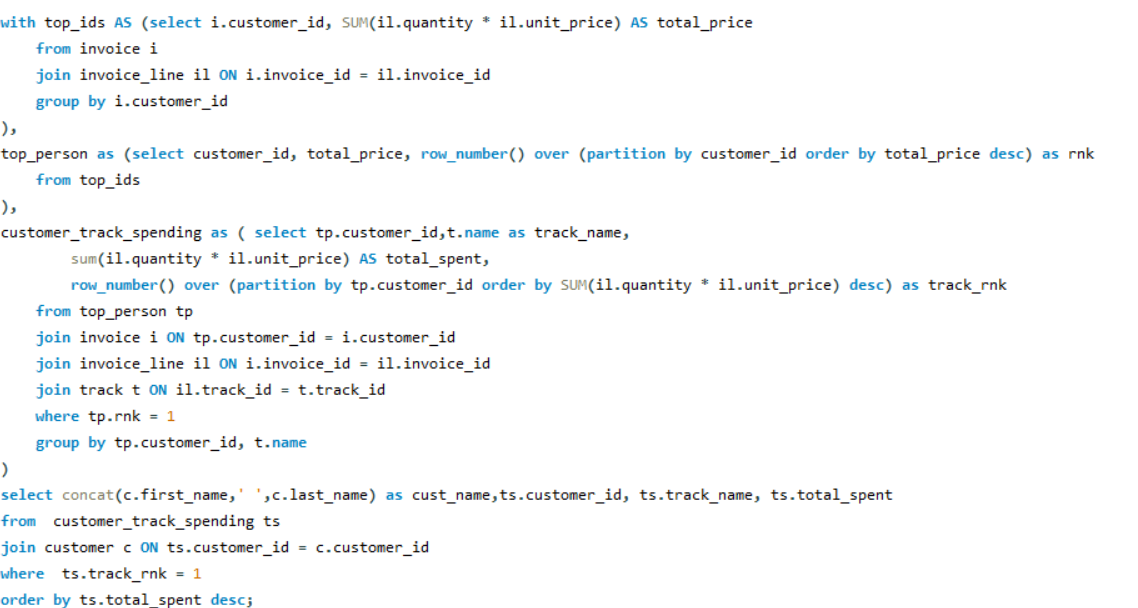
Code:



1. Identify the top-selling track for each customer

****

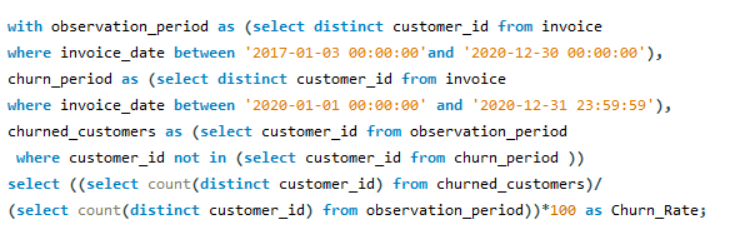
Code:



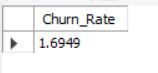
1. What is the customer churn rate?

* The customer churn rate is 1.6949
* Formula for churn rate is (churned\_customers/total\_customers) \*100

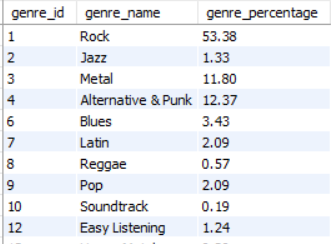
Code:



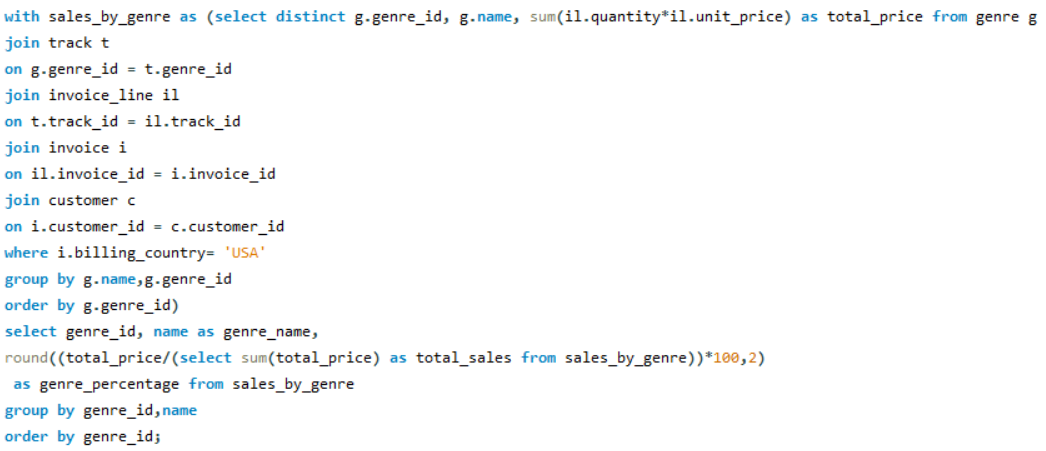
Output:



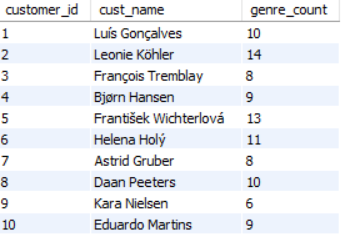
1. Calculate the percentage of total sales contributed by each genre in the USA and identify the best-selling genres and artists.



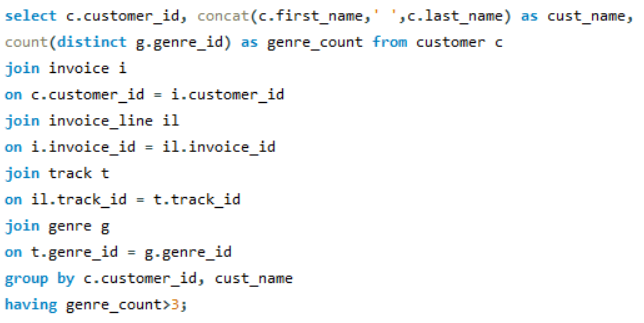
Code:



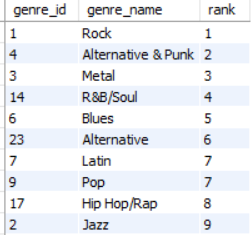
1. Find customers who have purchased tracks from at least 3 different genres



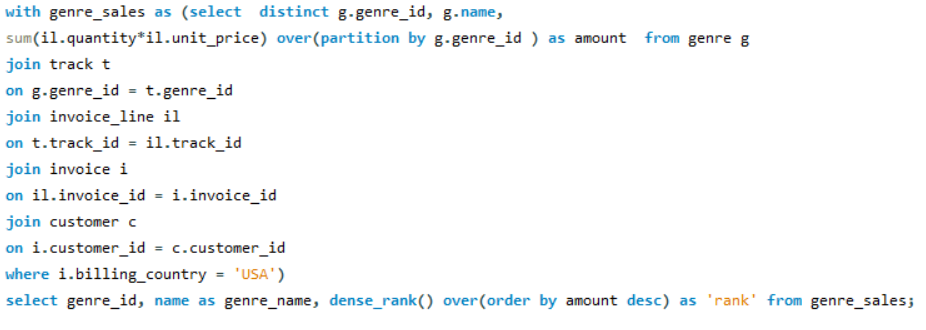
Code:



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

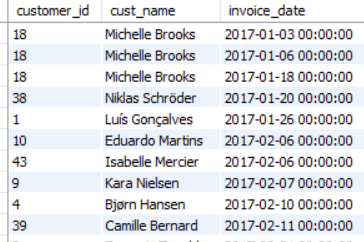


Code:

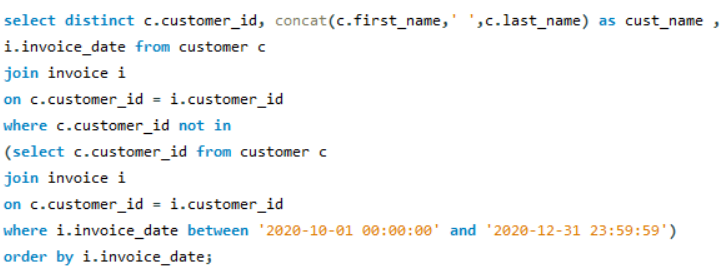


1. Identify customers who have not made a purchase in the last 3 months

This is the list of customers that have not made a purchase from Oct 2020 to Dec 2020



Code:

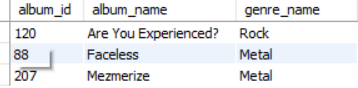


Subjective Questions

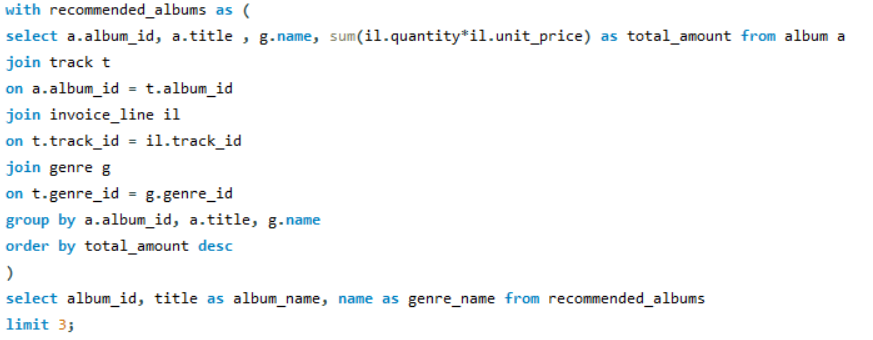
1. Recommend the three albums from the new record label that should be prioritised for advertising and promotion in the USA based on genre sales analysis.

* First, I found out the price that is spent on each album and their quantity.
* Then sorted them in descending order of the price spent
* And then put the limit as 3 which shows the top 3 albums which have the maximum spending.

The three albums that should be prioritised for advertising are

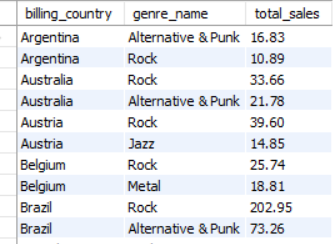


Code:

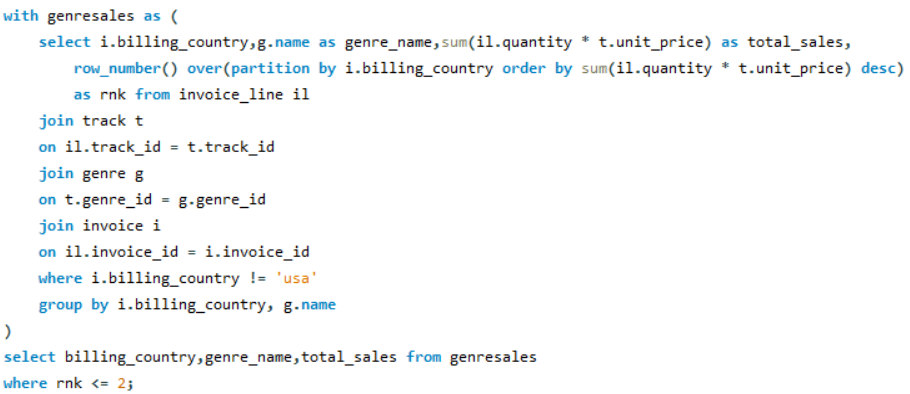


1. Determine the top-selling genres in countries other than the USA and identify any commonalities or differences.

* Calculated the selling price for all genres.
* Used the PARTITION BY clause to divide the data by country and sorted the genres in descending order of sales.
* Assigned a rank to each genre within each country using the ROW\_NUMBER function.
* Grouped the data by country and applied a filter to exclude the USA using the NOT operator.
* Included only rows where the ROW\_NUMBER value was less than 2.



Code:



Similarities:

* Popularity of Global Genres: Genres like Pop and Rock are widely favored due to their universal appeal and global artist dominance.
* Influence of Streaming Platforms: Services like Spotify and Apple Music promote global hits, creating consistent trends across countries.
* Youth-Oriented Genres: Genres like Hip-Hop/Rap are popular among younger demographics in multiple regions.

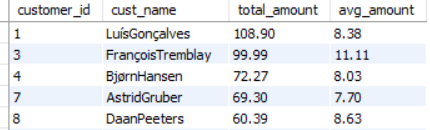
Differences:

* Regional preferences differ, such as Reggaeton in Latin America, K-Pop in Asia, and French Chanson in Europe.
* Countries like Germany excel in Classical and Techno, while India sees Bollywood and regional folk music dominate.
* Wealthier nations may focus on niche or premium formats (e.g., Jazz vinyl in Japan), while emerging markets prioritize streaming or local content.

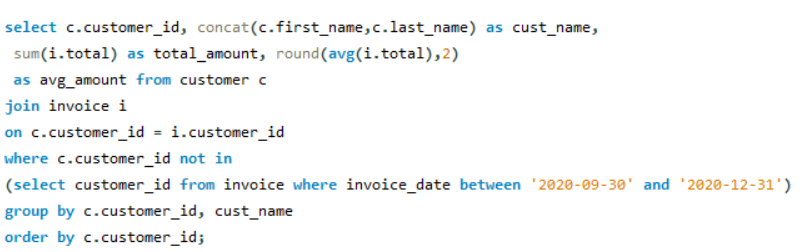
1. Customer Purchasing Behaviour Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?

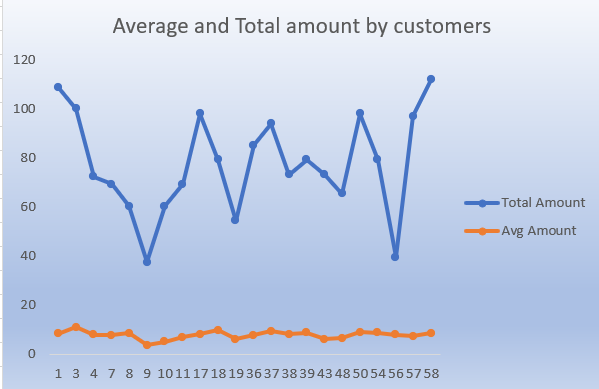
* Here I have divided customers into old and new customers.
* New customers are those who have made a purchase in the year 2020 and old are those who have made purchase in the year before 2020
* The average amount spent by new customer is around 8 to 10 and of old customer is 8 to 11. So here we can see the average amount spent by old customer is a bit high than new customer.

Old Customer

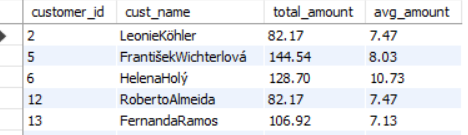


Code:

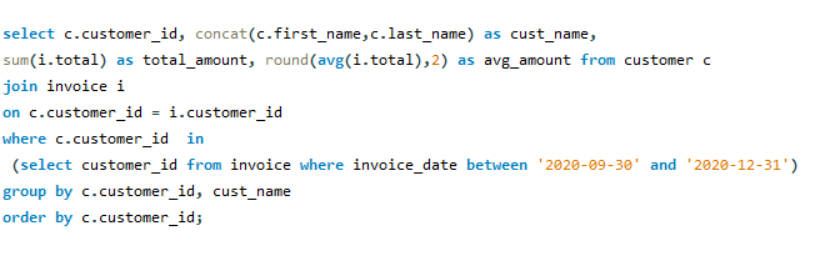


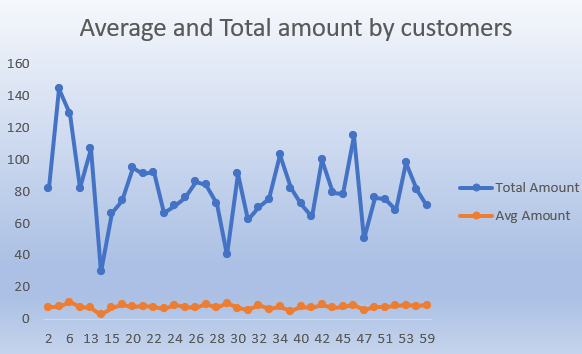


New Customer



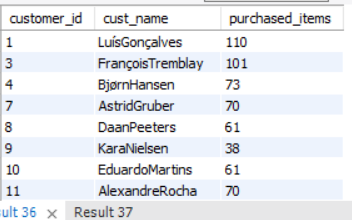
Code:



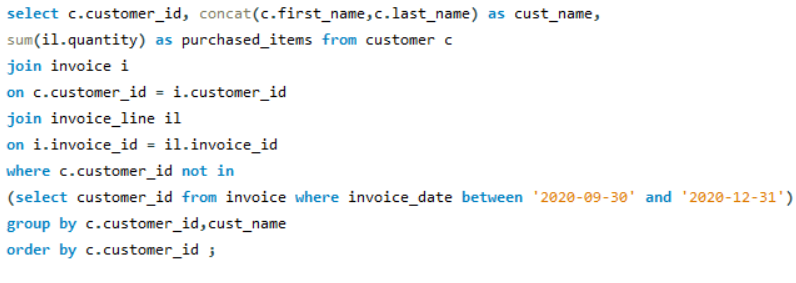


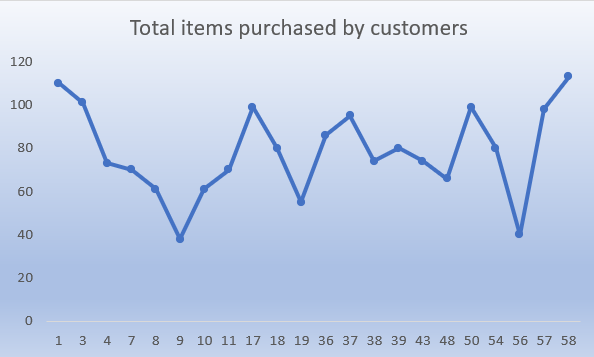
* The no of purchases of new customer is 50 to 140 and of old customer is 5 to 120. So here we can conclude that the no of purchases of new customers is quite higher than old customers

Old customer

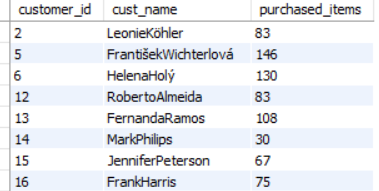


Code:



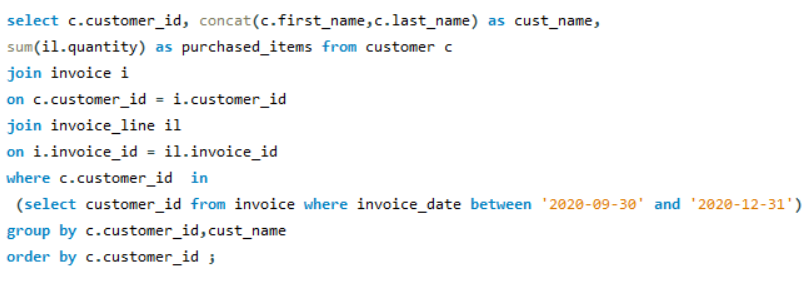


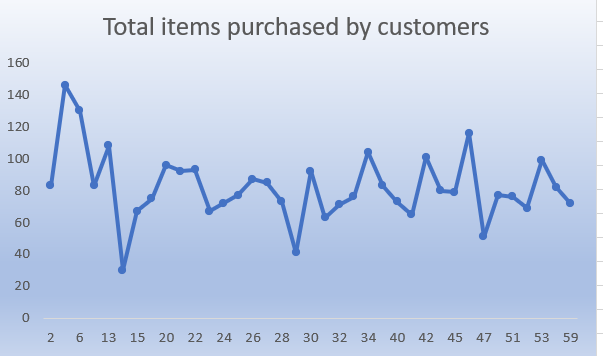
New Customer



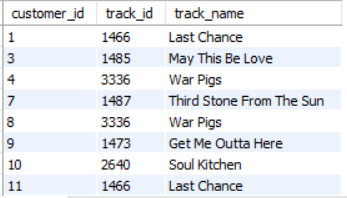
* The top tracks of new customers are “War Pigs”, “Dead and Broken” and “Old School Hollywood” and of old customers are
* “War Pigs”, “Dead and Broken” and “Are you Experienced”

Code:

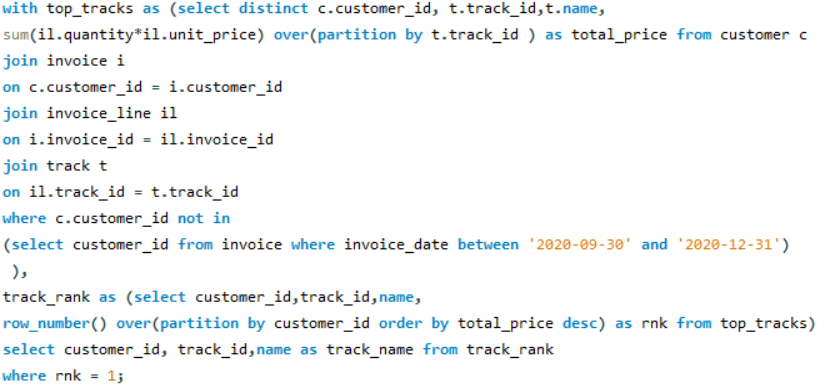




Old Customer



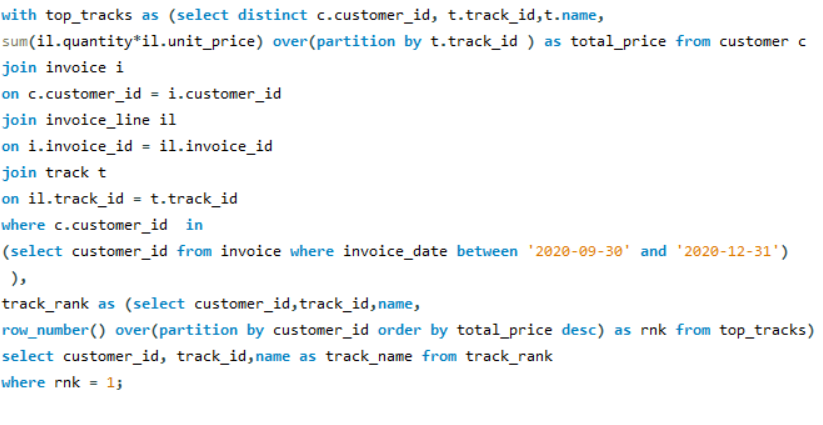
Code:



New Customer

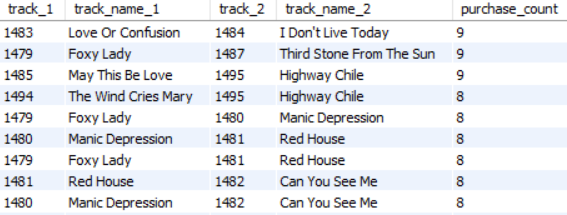


Code:

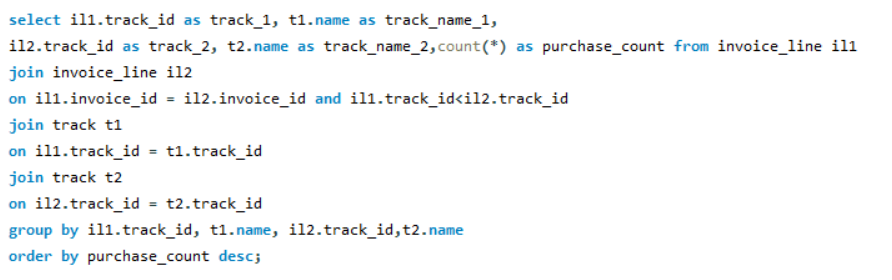


1. Product Affinity Analysis: Which music genres, artists, or albums are frequently purchased together by customers? How can this information guide product recommendations and cross-selling initiatives?

* Select the columns name, track\_id, and count of purchases as purchase count
* Joined the invoice\_line table with itself
* Then joined with Track table
* Then grouped the data by track\_id and track\_name
* Sorted the data in descending order of purchase count.



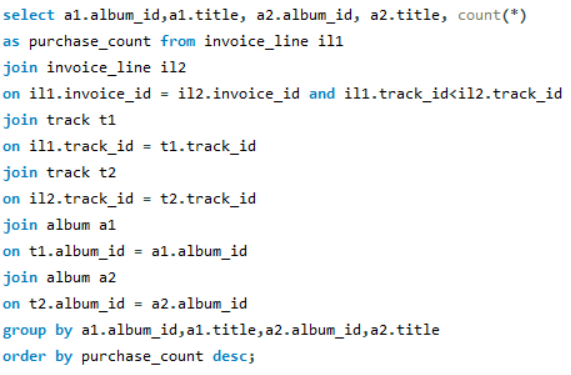
Code:



Albums that are purchased together frequently



Code:

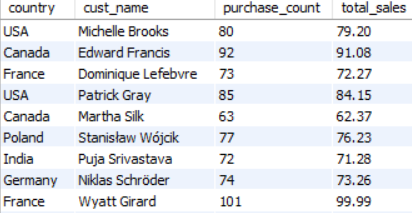


After getting this information it can help in product recommendations as follows

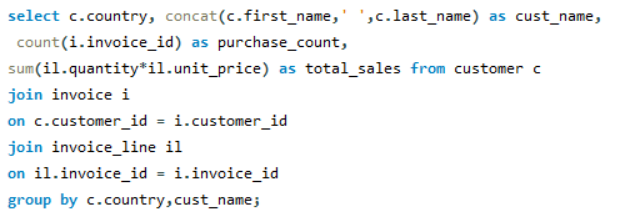
* As we know which are the tracks and albums that are most frequently purchased together so in future you should ensure that both the tracks and albums are available.
* If one customer buys a track1 so you can recommend him track 2 as it is frequently bought with track1
* If some person has a long history of purchasing particular 2 tracks frequently so you can recommend him another two tracks that are purchased frequently.

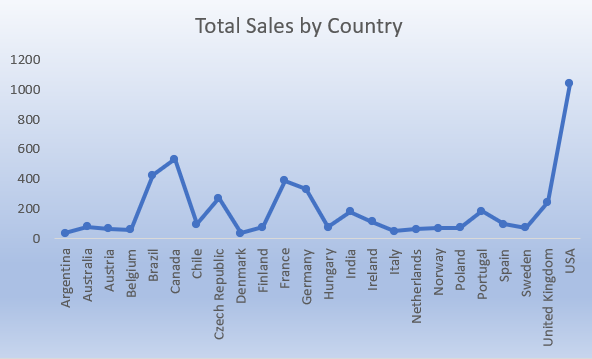
1. Regional Market Analysis: Do customer purchasing behaviours and churn rates vary across different geographic regions or store locations? How might these correlate with local demographic or economic factors?

Sales and Purchases by Region

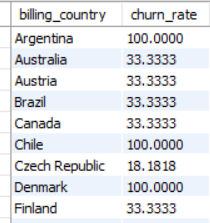


Code:

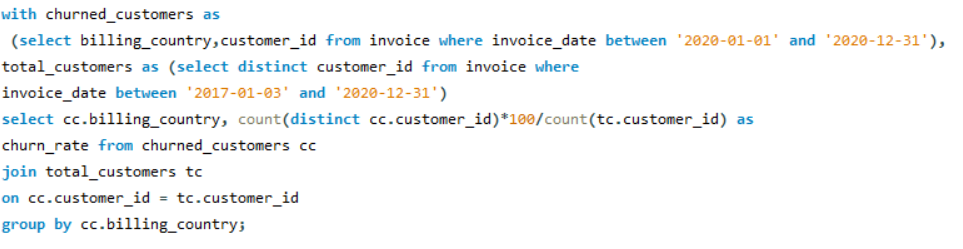


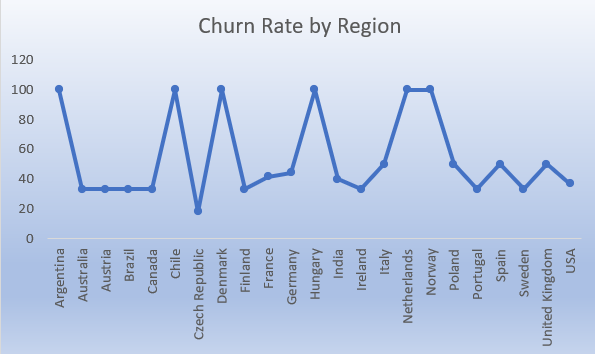


Churn Rate by Region



Code:



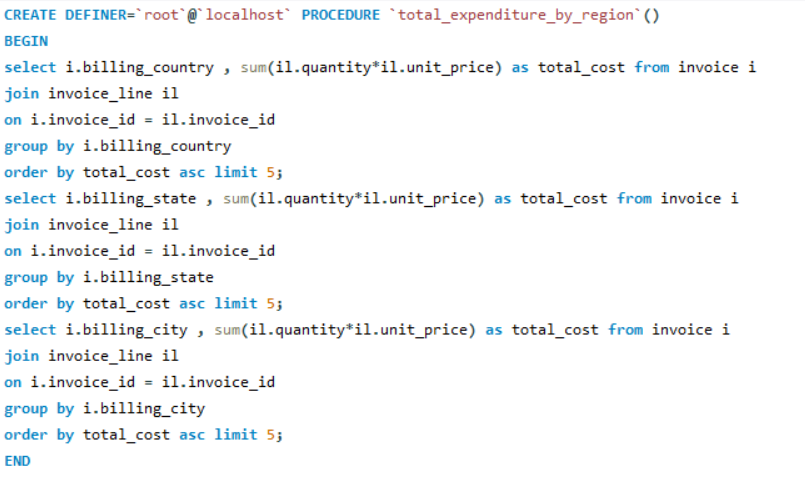


* Some regions have higher sales and purchases which shows strong customer activity.
* Others have lower sales because of fewer resources or less demand.
* Regions with higher churn rates suggest customers are not returning, because of poor service or local competition.
* High sales regions are wealthier or urban, while high-churn areas may face economic struggles or lack good service.
* Improve service and offer discounts in high-churn areas. In regions with strong sales, introduce loyalty programs or premium products to boost revenue.

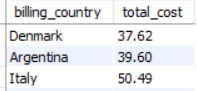
1. Customer Risk Profiling: Based on customer profiles (age, gender, location, purchase history), which customer segments are more likely to churn or pose a higher risk of reduced spending? What factors contribute to this risk?

* Created a procedure where I can store total expenditure by region which includes sales by country, state and city

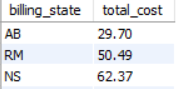
Code:



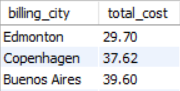
* Calculated the total amount spent by each customer in every country, city and state.
* Country wise these are the top 3 countries where there is higher risk of reduced spending



* State wise these are the top 3 states where there is higher risk of reduced spending

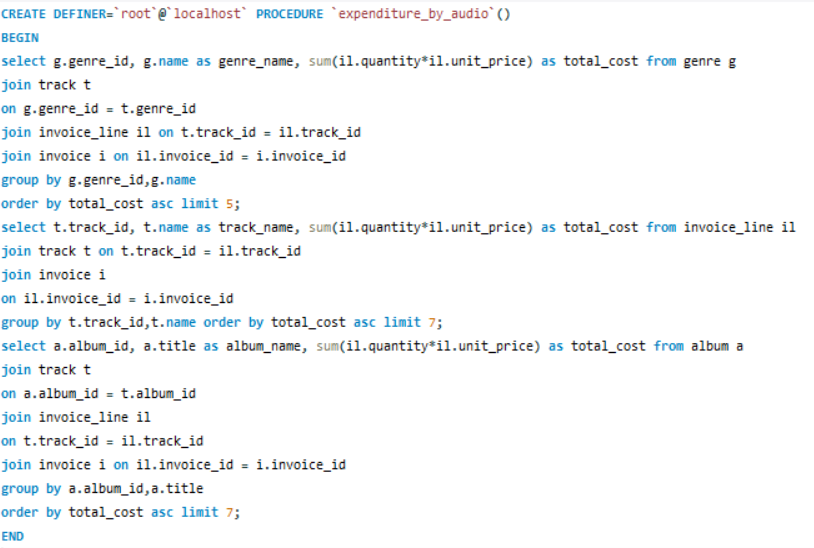


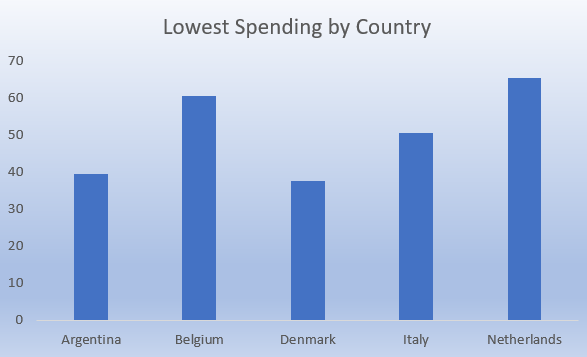
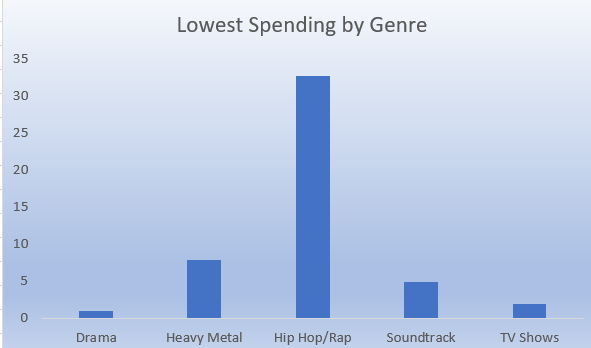
* City wise these are the top 3 cities where there is higher risk of reduced spending.



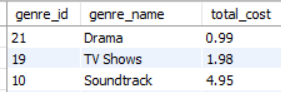
* Created another procedure where it shows total expenditure by genres, albums and tracks

Code:

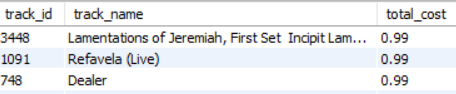


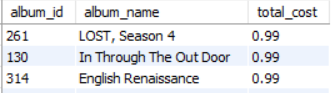
* Calculated the total cost spent on genres, albums and tracks by each customer.
* Genre wise these are the top 3 genres where there is higher risk of reduced spending.



* Album wise these are the top 3 albums where there is higher risk of reduced spending

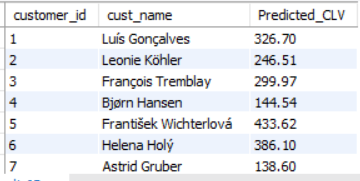


* Track wise these are the top 3 tracks where there is higher risk of reduced spending

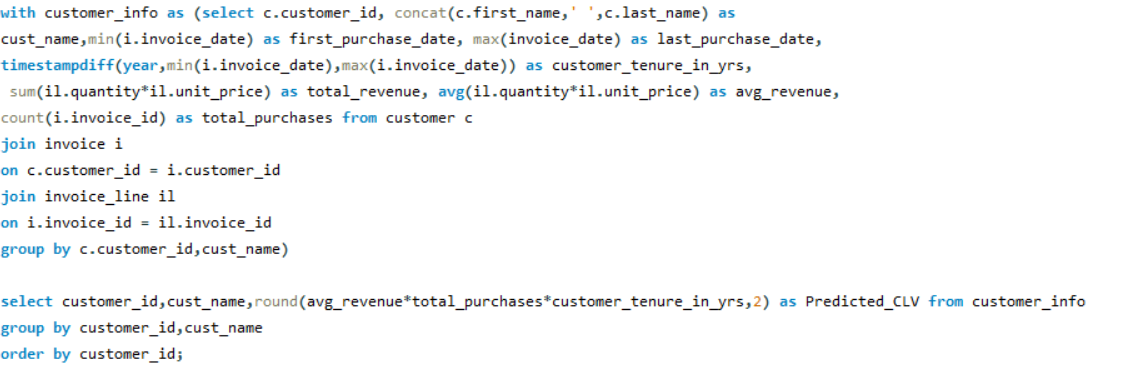


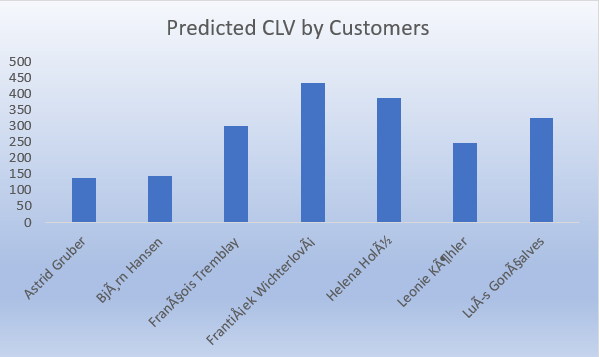
1. Customer Lifetime Value Modelling: How can you leverage customer data (tenure, purchase history, engagement) to predict the lifetime value of different customer segments? This could inform targeted marketing and loyalty program strategies. Can you observe any common characteristics or purchase patterns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

* Customer data like customer’s tenure, purchase history and engagement can be used to calculate predicted customer lifetime value and customer loyalty with time.
* 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

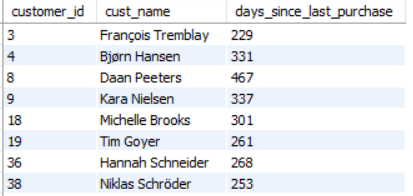
Below is the table of customer with their predicted CLV  


Code:

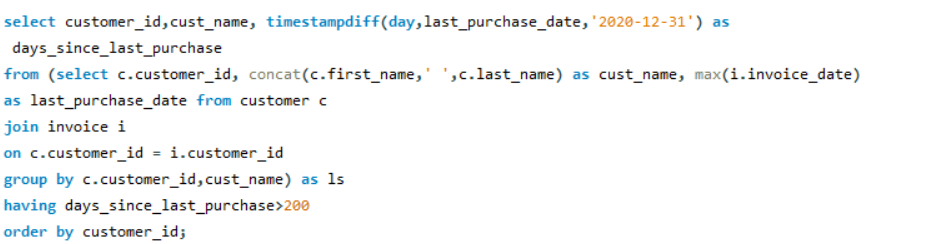


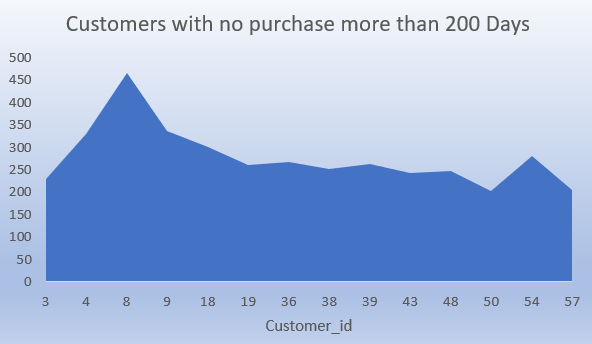


* Below is the table of customers who have not purchased from the last 200 days.



Code:





* From the above table which consists of customers who have not purchased since 200 and more days.
* Most inactive customers fall in the range of 243 to 467 days since their last purchase.
* These customers are not new but not very loyal either, they had some past activity but stopped due to a lack of offers or discounts.
* These customers are not entirely new but also not most loyal.

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

If the data on promotional campaigns is available then I would do the following

* Look at how many new customers joined during the campaign compared to before it  
  Analyze the increase in new customer acquisitions during the campaign period compared to the weeks or months leading up to it. This can reveal whether the campaign attracted fresh attention to your brand. Key metrics include:
* Number of new customer sign-ups or registrations.
* Demographic trends of new customers (e.g., age, location).
* Customer acquisition rate during and before the campaign.
* Check if customers keep buying after the campaign ends to see if it helped keep them loyal  
  Evaluate the retention and purchase frequency of customers who joined during the campaign period. This helps assess whether the campaign resulted in long-term customer engagement or merely a short-term spike. Metrics to track include:
* Repeat purchase rate.
* Average order value (AOV) over time.
* Customer Lifetime Value (CLV) comparison between campaign-acquired customers and pre-existing customers.
* Compare sales numbers before and after the campaign to see if it boosted revenue  
  Examine overall revenue trends during the campaign and the subsequent period. This provides insight into the financial impact of the campaign. Consider:
* Total revenue increase during and after the campaign.
* Conduct a comparative analysis to determine the campaign's effectiveness. Use control groups or segment customers into those exposed to the campaign and those who weren’t. Key areas to evaluate:
* Conversion rates among exposed vs. unexposed customers.
* Difference in average spend per customer.
* Retention rates and customer satisfaction levels.

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

If objective and subjective questions weren’t given I

would go by this approach:

* Start by identifying and defining Key Performance Indicators (KPIs) to quantify campaign performance effectively. These metrics provide measurable insights into the campaign's impact. For instance:
* Customer Retention Rate: Measures the percentage of customers who continue to make purchases after the campaign.
* Sales Growth: Tracks the increase in sales volume or revenue compared to pre-campaign levels.
* Total Sales and Total Profit: Evaluates the financial outcomes to assess overall profitability.
* Additional metrics like average purchase value or campaign-specific ROI can be included for deeper analysis.
* Analyse sales, customer behaviour, and engagement before and after the campaign to identify changes in performance  
  Conduct a comparative analysis of sales trends, customer behavior, and engagement metrics to assess the campaign's impact. Key focus areas include:
* Sales trends: Compare revenue, transaction volume, and basket sizes pre- and post-campaign.
* Customer behavior: Assess changes in frequency of purchases, product preferences, or purchase channels.
* Engagement: Monitor shifts in interaction rates, such as website traffic, social media engagement, or email open rates.
* Monitor customer feedback, ratings, and purchase behavior related to specific promotions to gauge effectiveness  
  Collect and evaluate customer responses to the campaign to understand its appeal and effectiveness. Areas to consider:
* Customer Feedback: Use surveys or reviews to gauge satisfaction with campaign promotions.
* Ratings: Analyze product or service ratings received during the campaign.

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?

* Here we will use **Alter** command to change the structure of the table and then we use **Add** command to add a new column in the table.
* Given below is the syntax

Alter table Albums

Add ReleaseYear int;

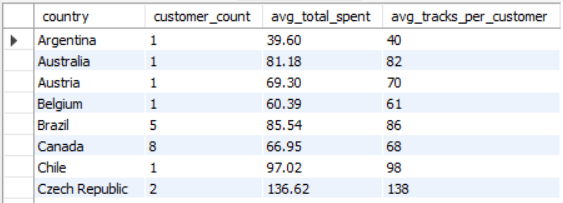
* We can also use this command

Alter table Albums

Add column ReleaseYear int;

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.

* Here I would first write a subquery which will find total amount, number of purchases made and average of total amount spent by each customer.
* And then using this subquery I would write an outer query which consists of country column and it will show the count of customers, average of total amount spent and average of tracks listened by each customer.
* After finding these values I will group them by country which will show count of customers, average of total amount according to country



Code:

