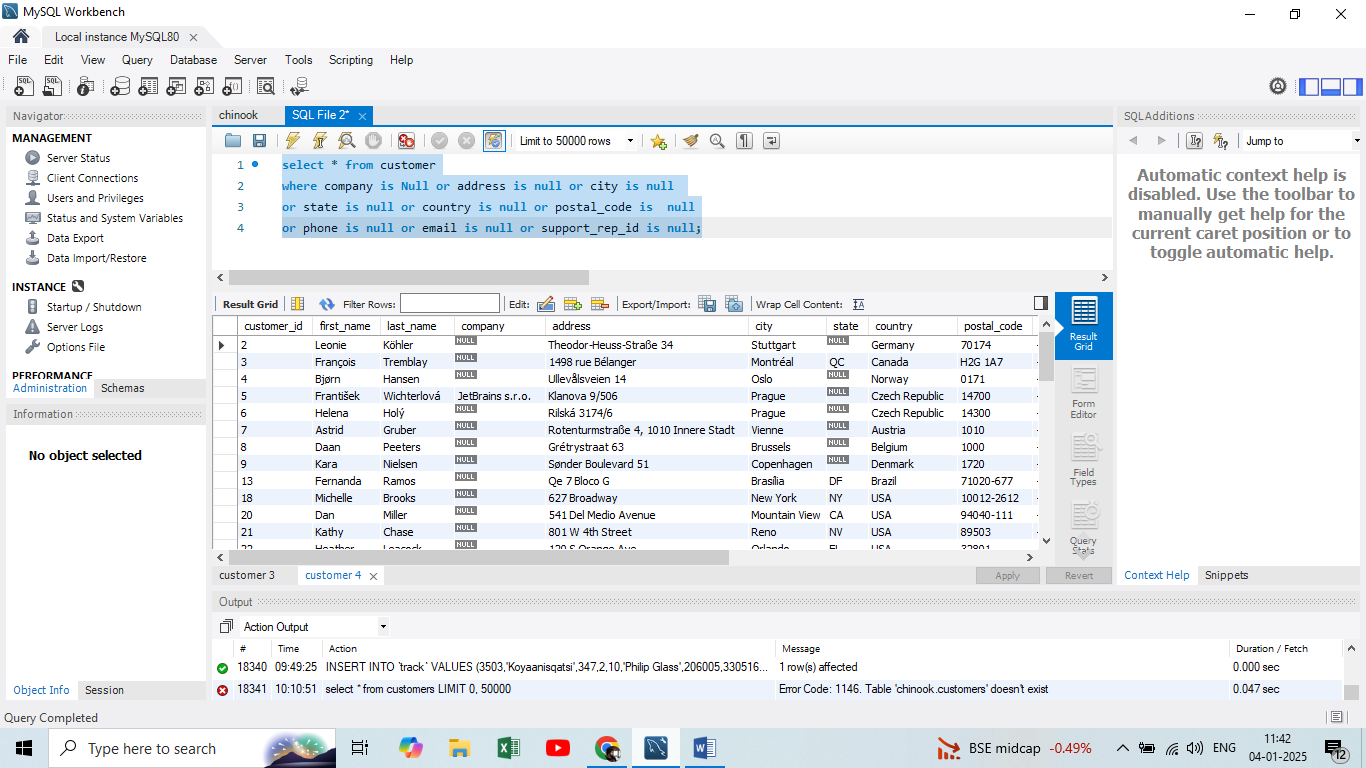
**CHINOOK MUSIC DATA ANALYSIS**

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

Ans: Missing values or null values can be checked by select statement and IS NULL condition in where column. While retrieving null values can be addressed by coalesce clause. Example: To find null values in customer table of chinook database. (Count(\*) can also be used instead of \* in select statement)



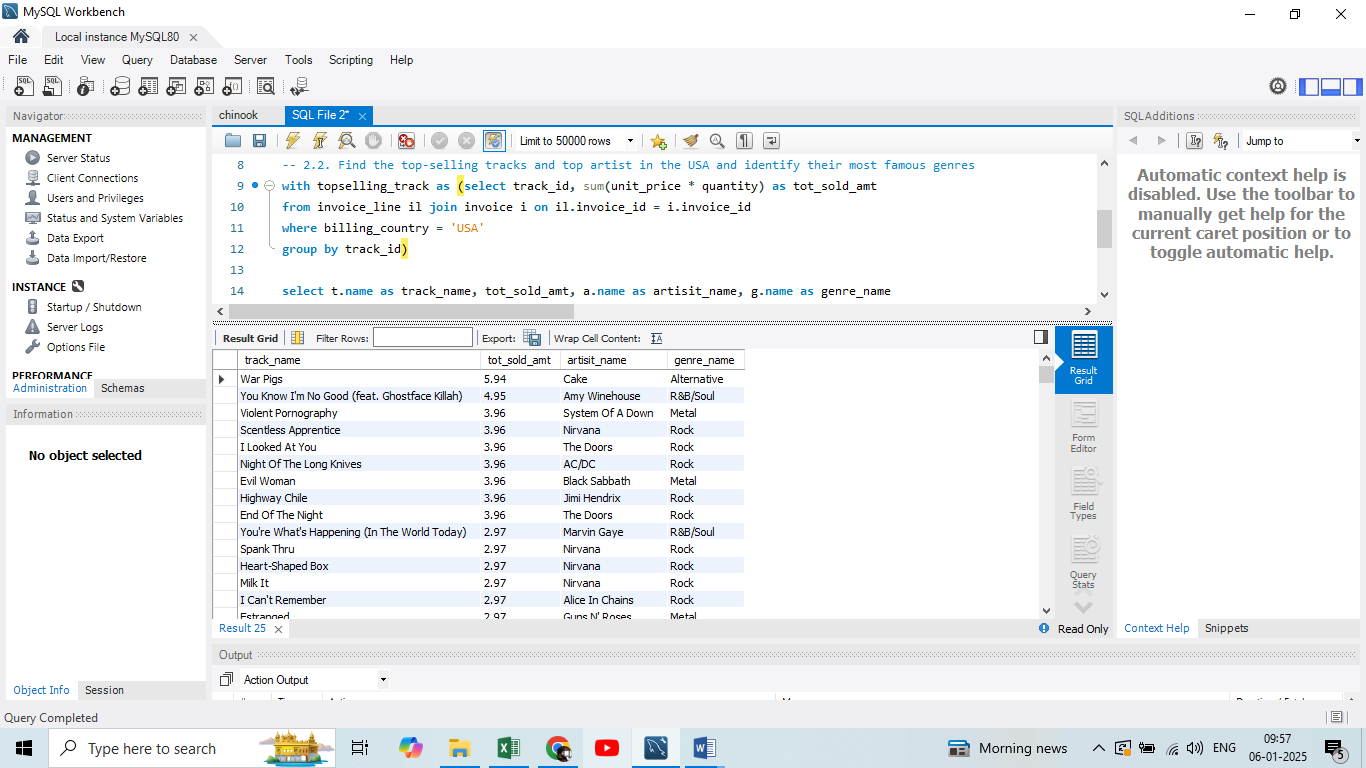
But null values can be permanently replaced by blank, average, mode etc. based on requirement by UPDATE statements.

Example:

Update customer set company = ‘default()’ where company is Null.

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

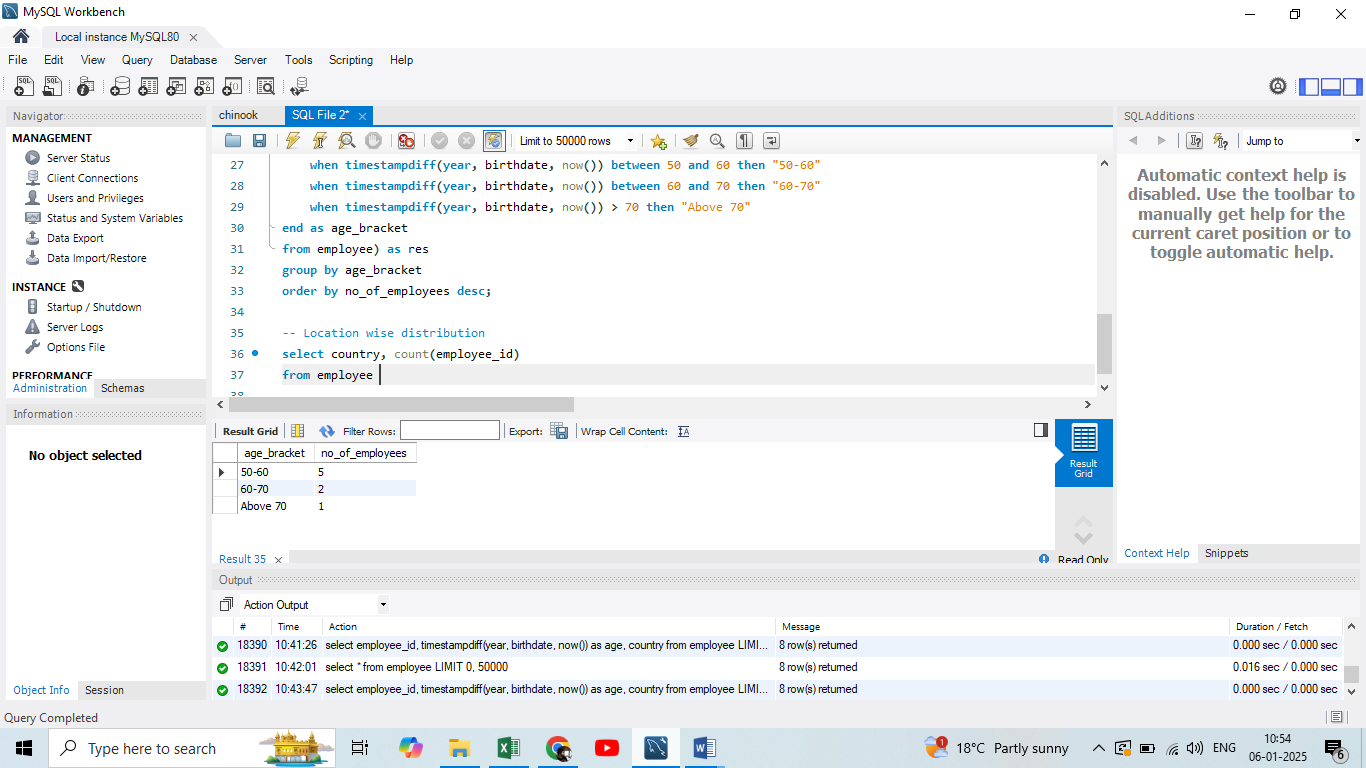
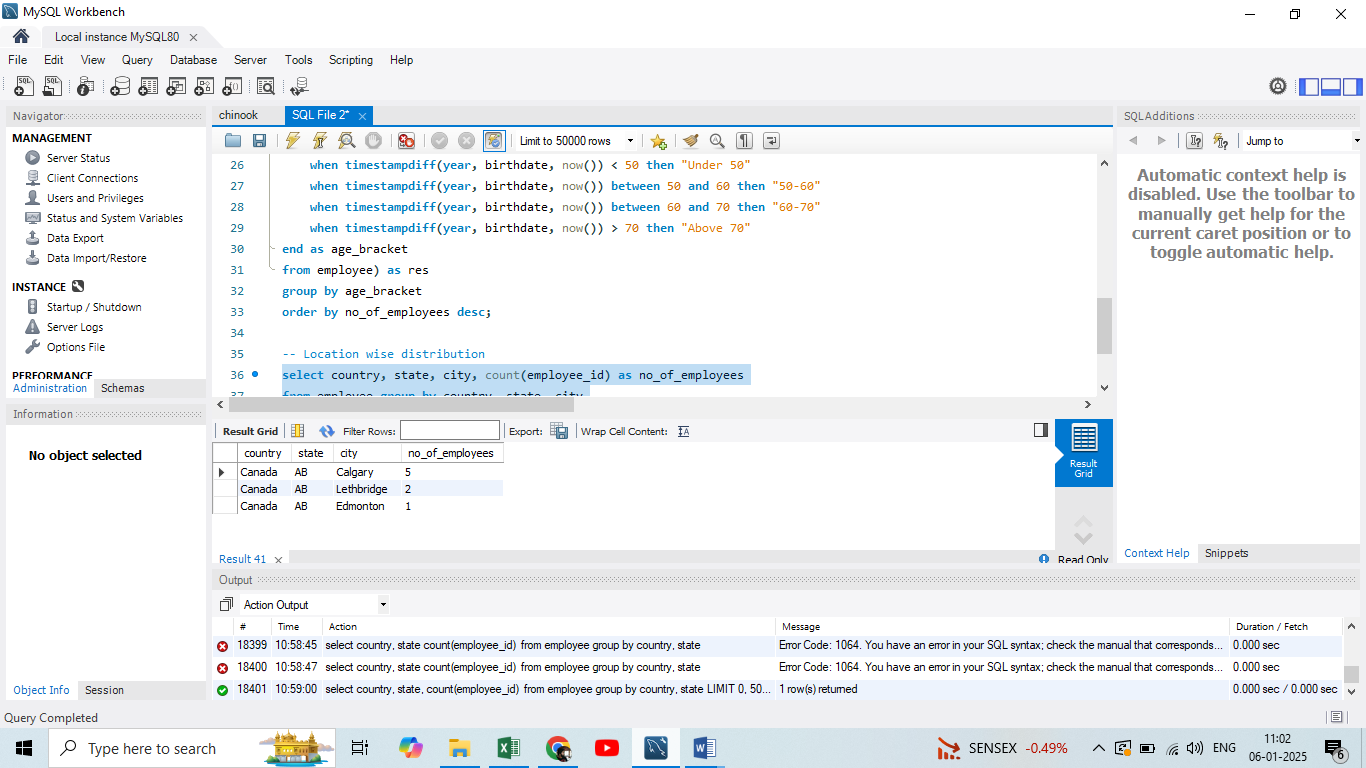
Ans: Top selling tracks in USA by their total selling amount is calculated from invoice and invoice\_line tables and then it is joined with track, album, artist, genre table to retrieve respective artist and genre and track names.

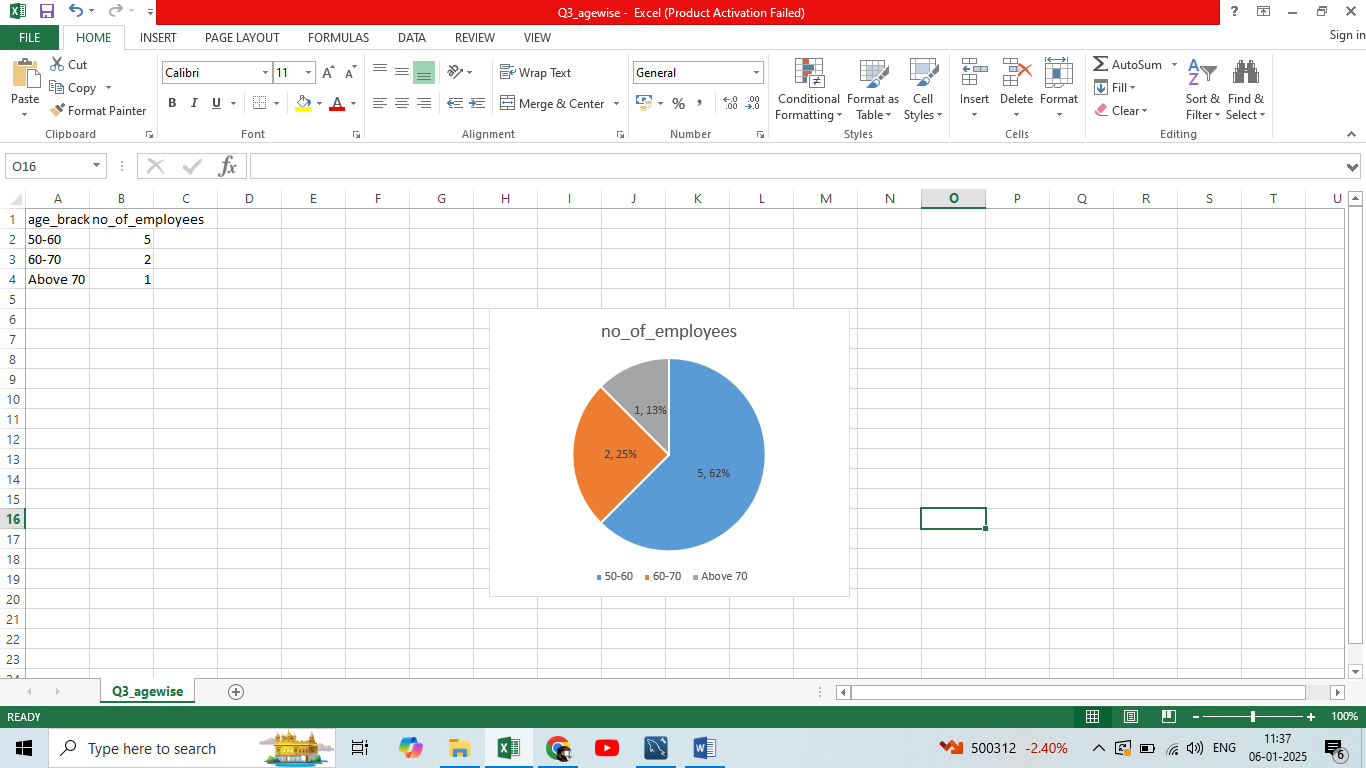
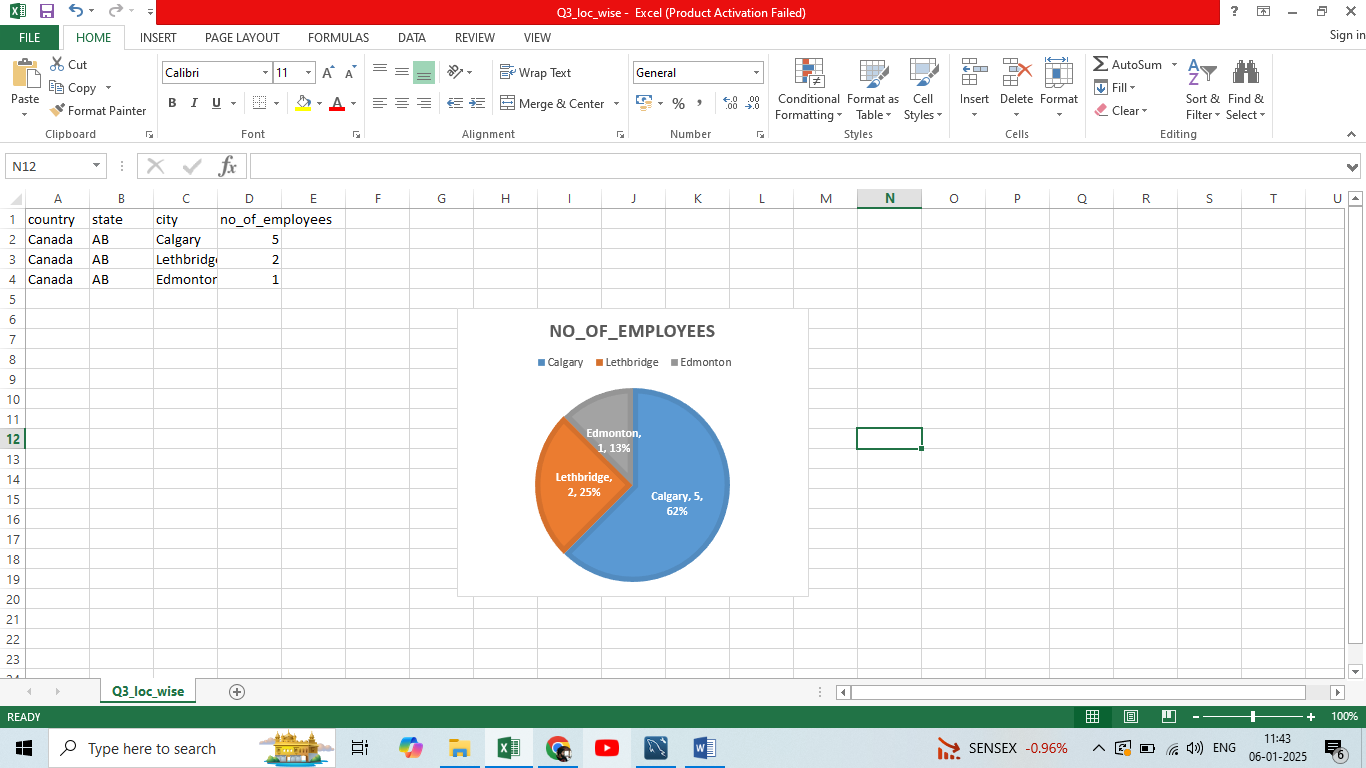


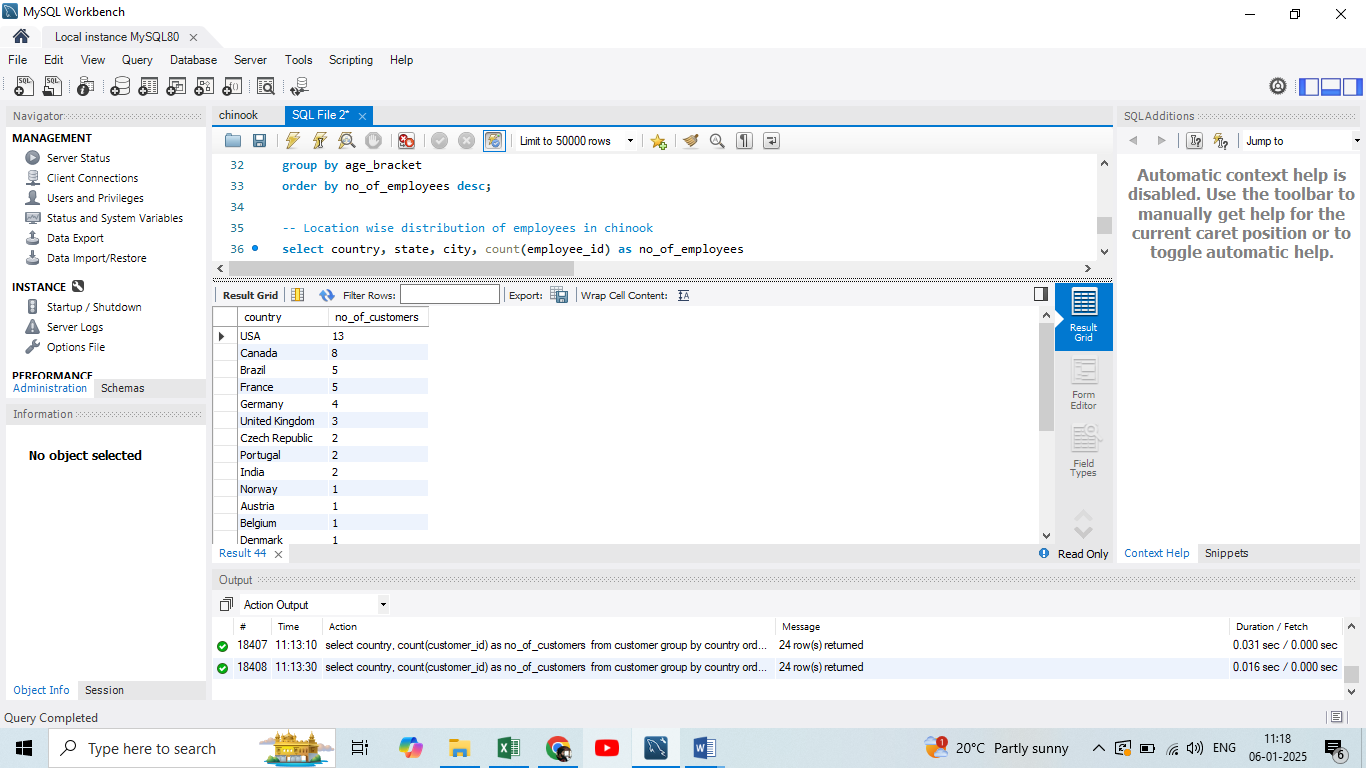
From the above inference top sold track in USA is “War Pigs”, the artist is “Cake” and genre is “Alternative”.

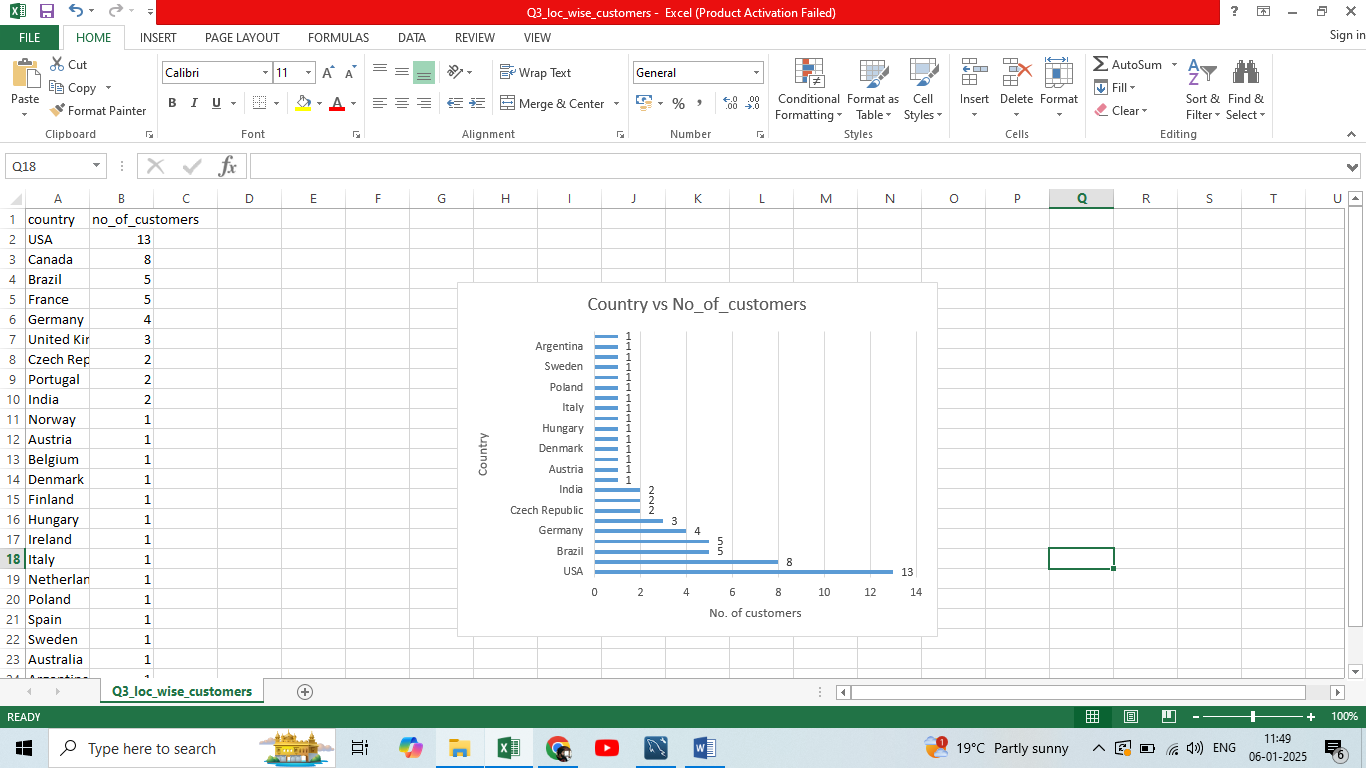
1. What is the customer demographic breakdown (age, gender, location) of Chinook's customer base?

Ans:

 Above information is a demographic distribution of employees in chinook. There are 8 employees in total and all of them are from Canada and 5 of them in 50 to 60 age group.



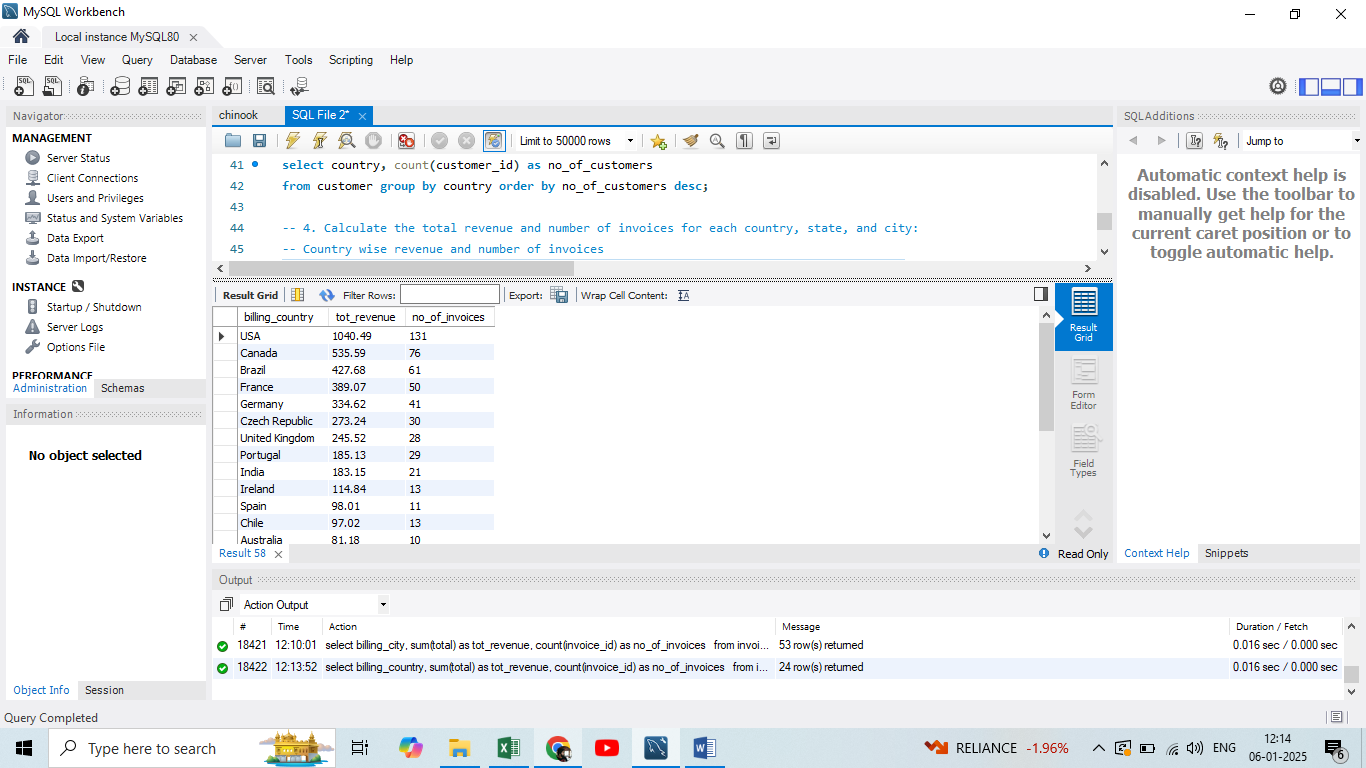


Above table is distribution of customers in chinook database and maximum number of customers are from USA followed by Canada.

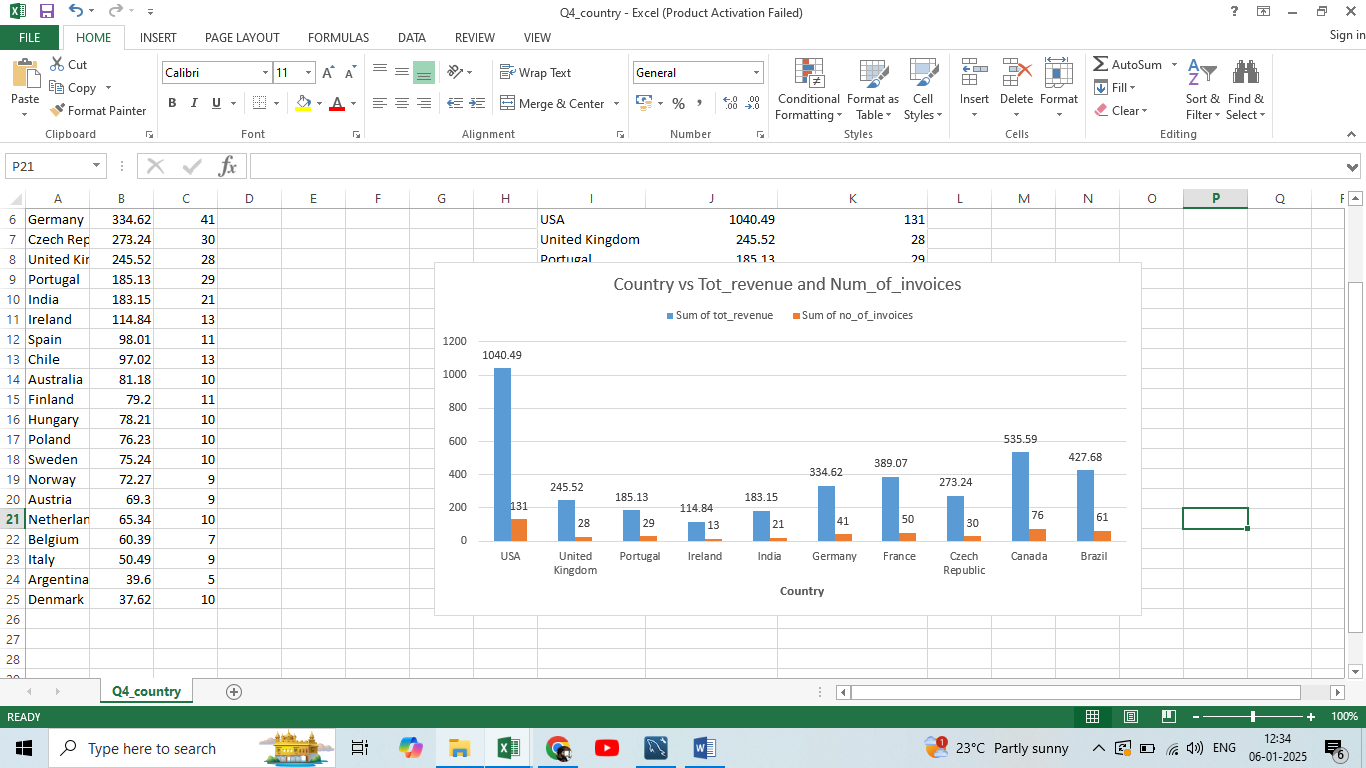
1. Calculate the total revenue and number of invoices for each country, state, and city:

Ans:

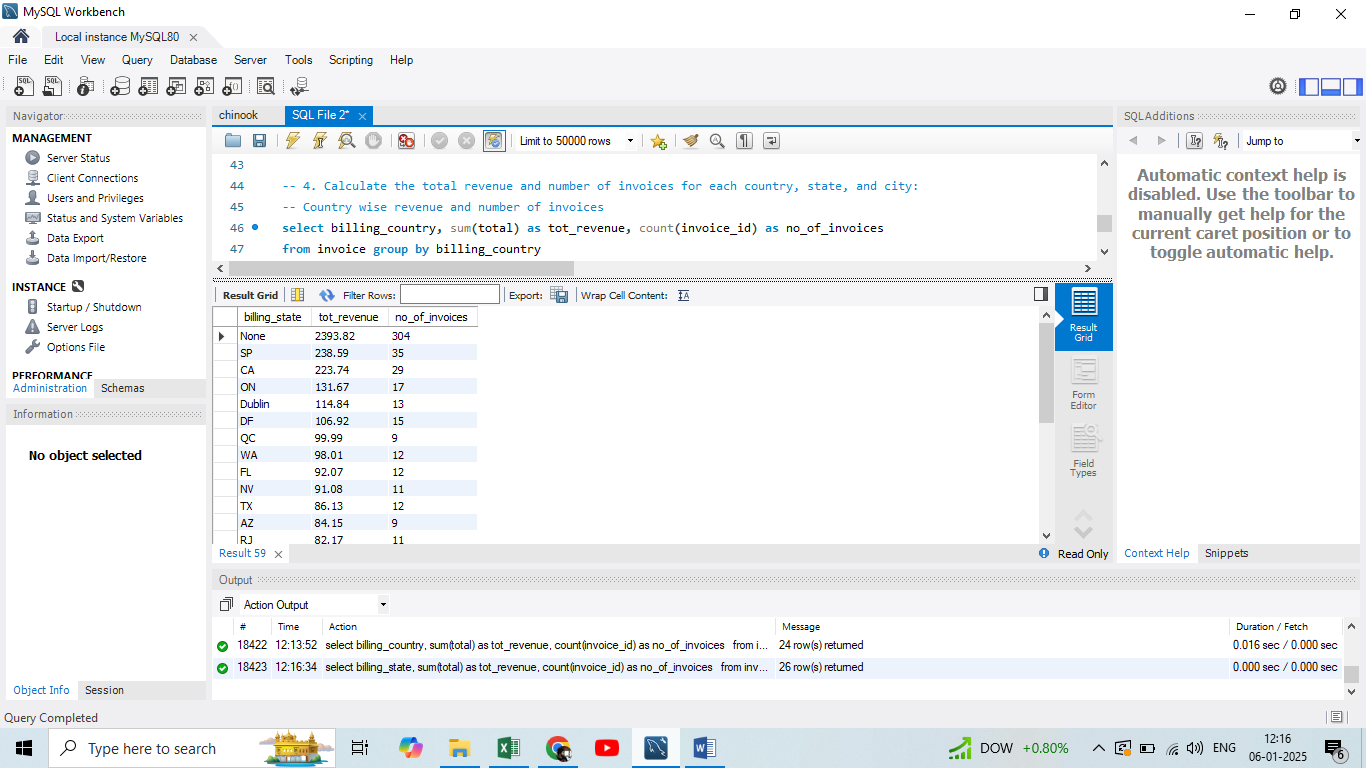
Country wise analysis of total revenue and number of invoices:



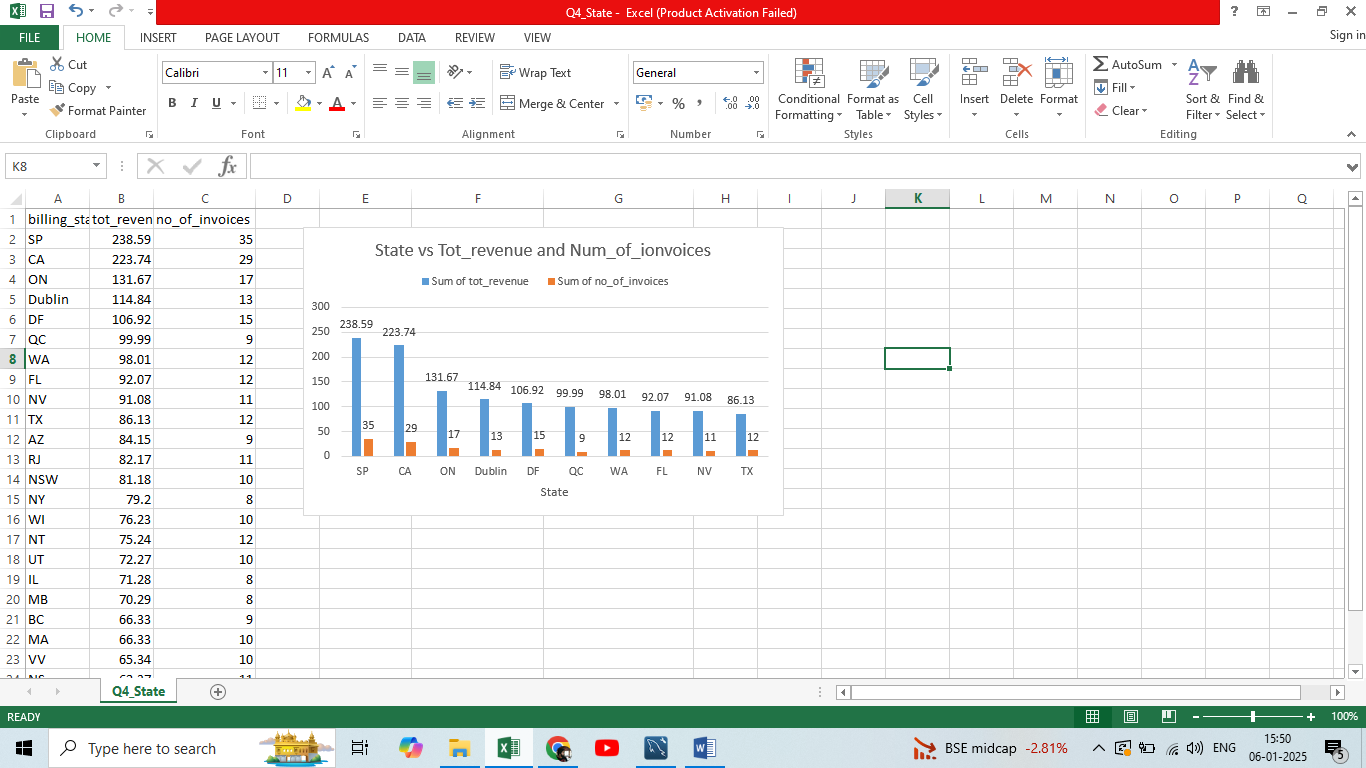
Visualisation of top 10 country’s revenue and number of invoices:



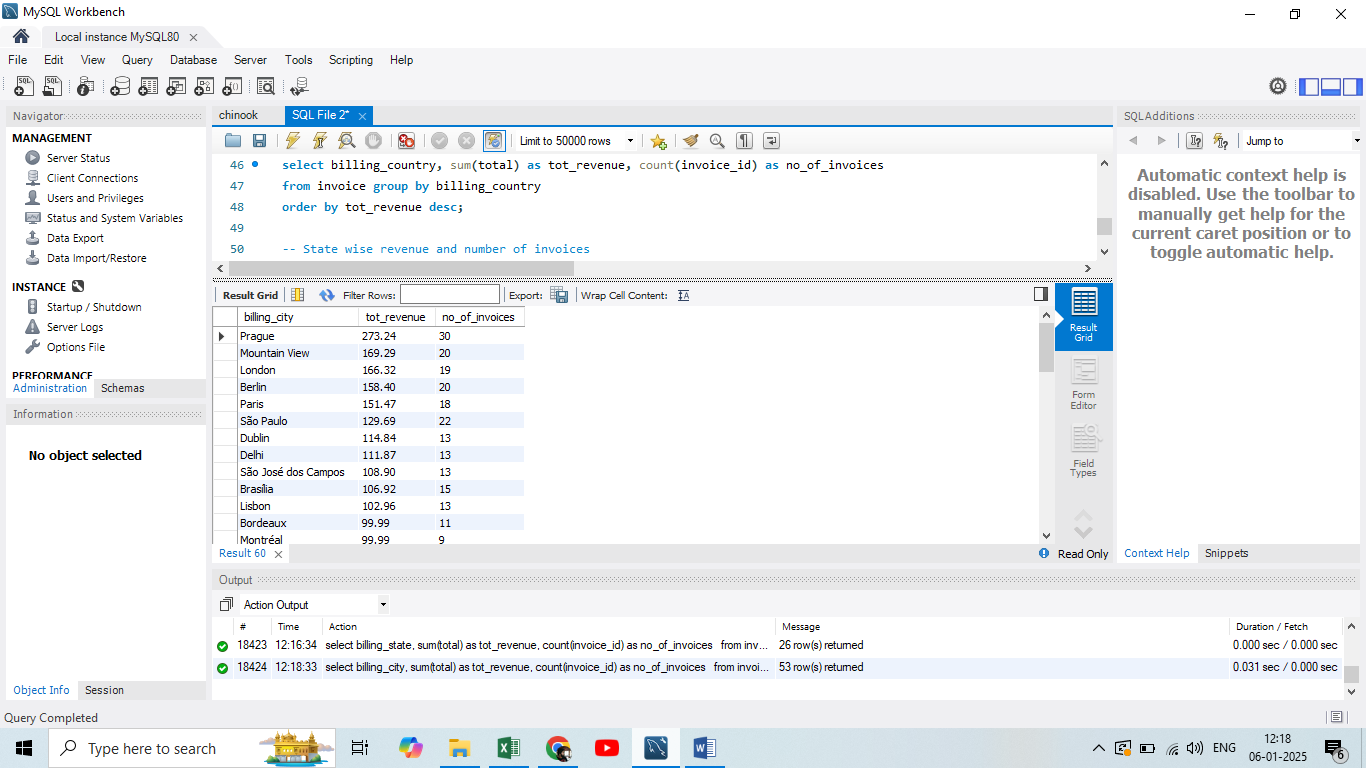
State wise analysis of total revenue and number of invoices:



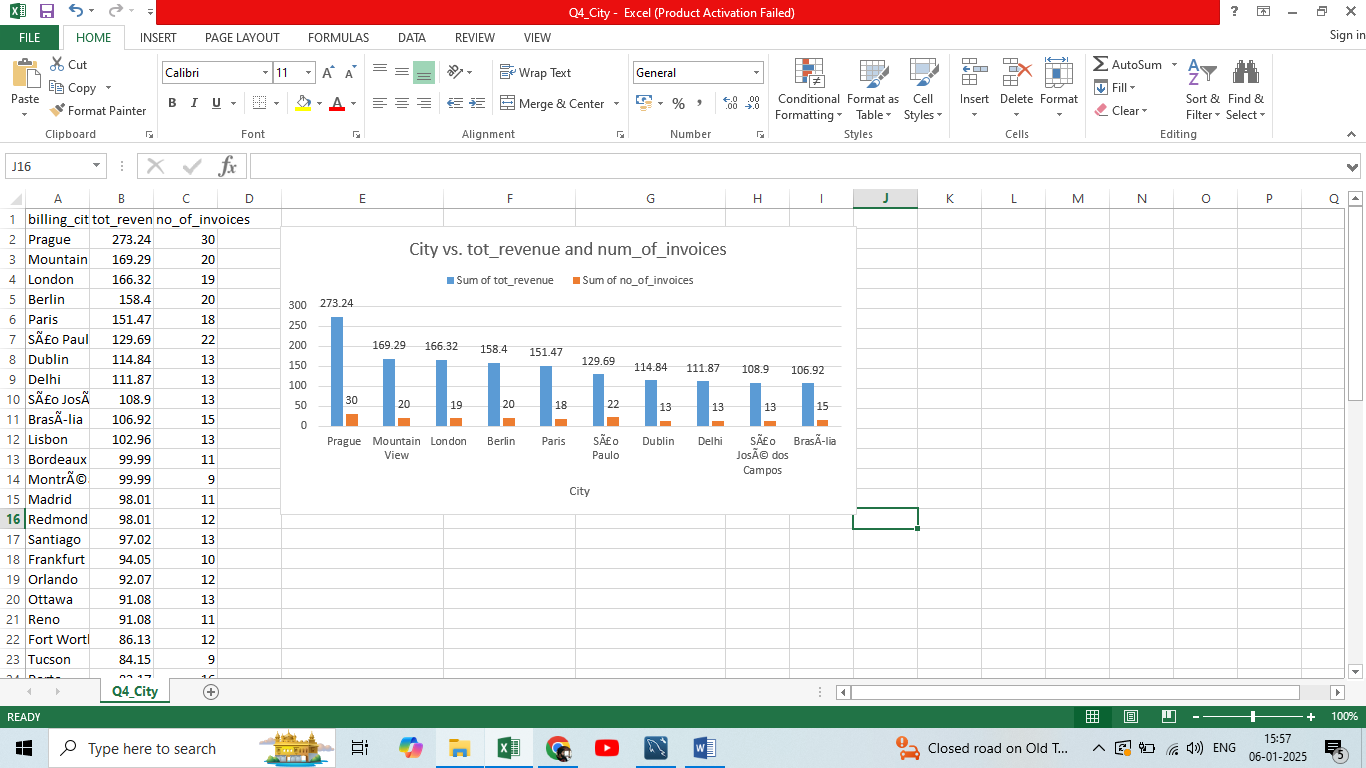
Visualisation of top 10 States’ revenue and number of invoices:



Country wise analysis of total revenue and number of invoices:



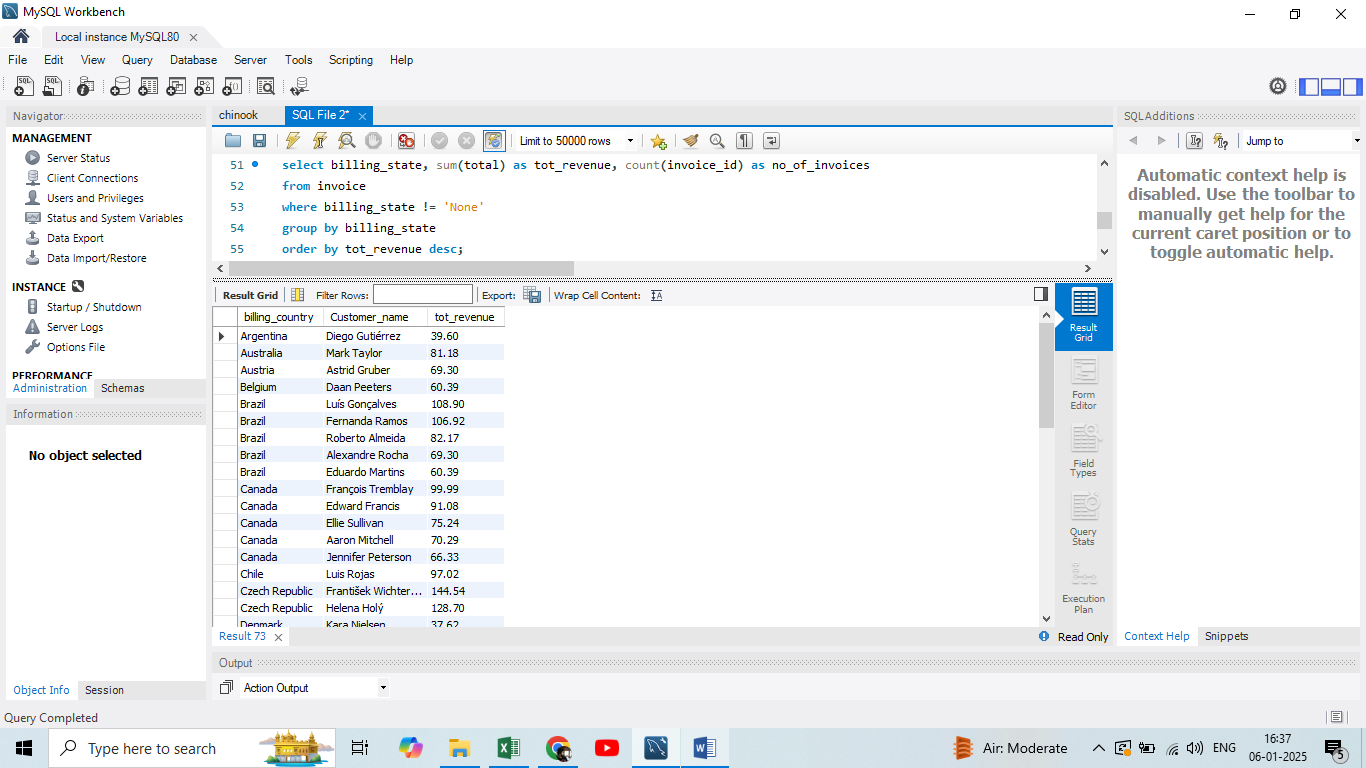
Visualisation of top 10 City’s revenue and number of invoices:



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

Ans:

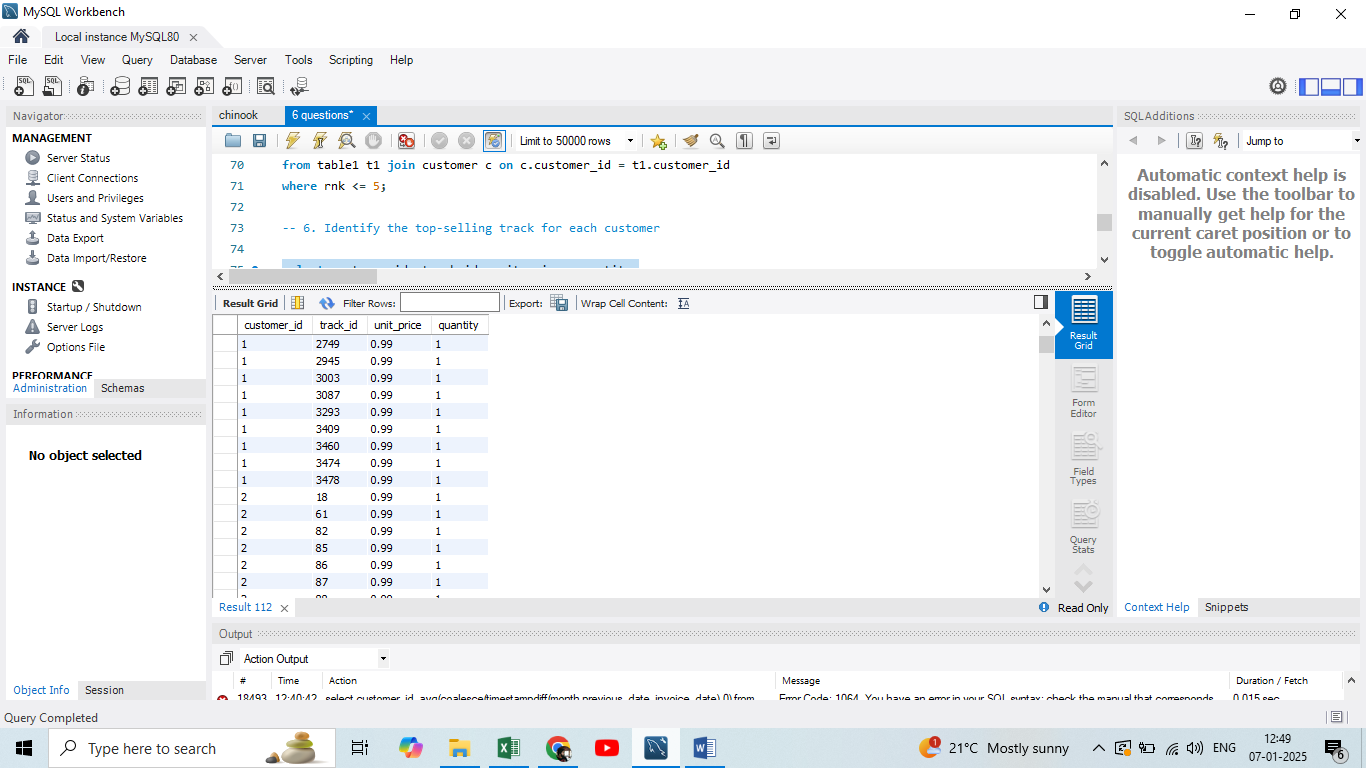
Top 5 customers in each country is retrieved by joining customer and invoice table and also using rank function.



1. Identify the top-selling track for each customer

Ans:

By joining invoice and invoice\_line table all the tracks purchased by each customer is retrieved.



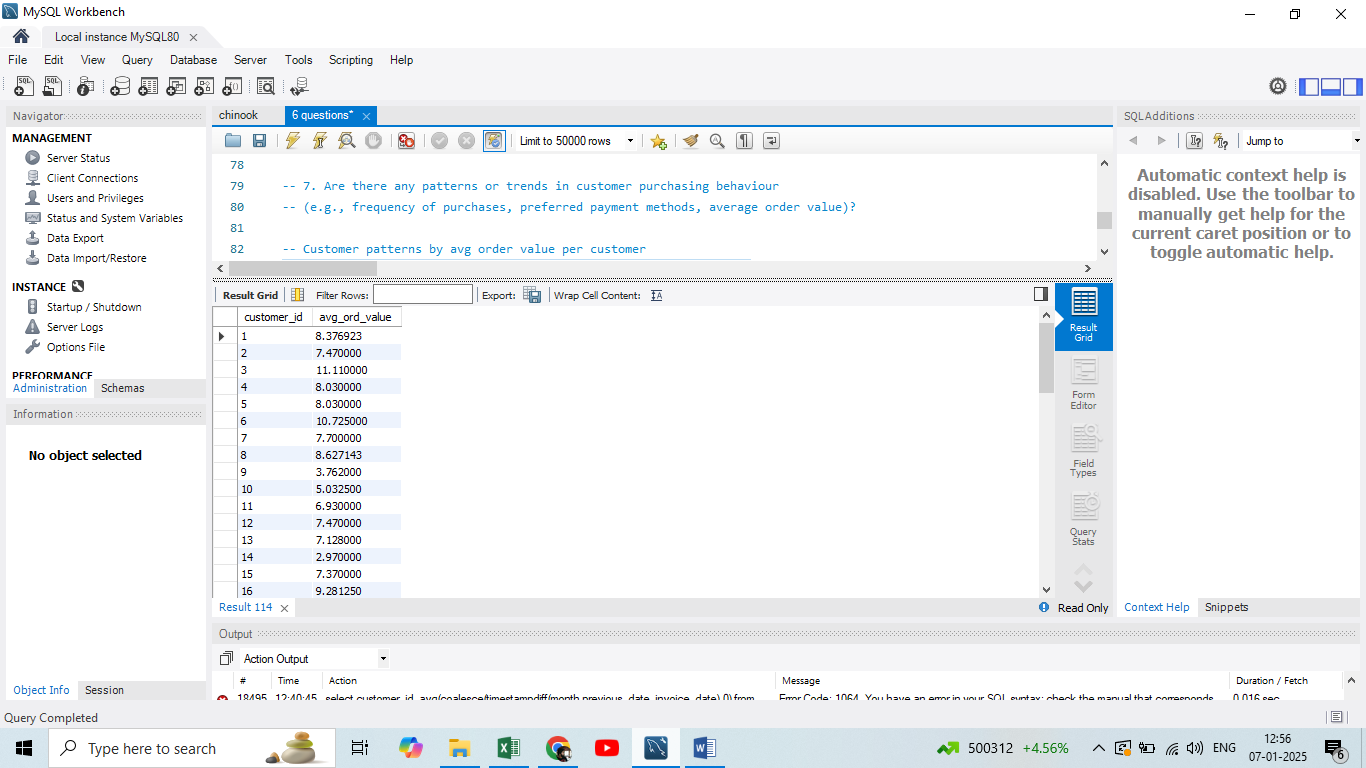
From the above output it is evident that one customer has purchased many tracks but one track only once and unit\_cost of all tracks and quantity of all tracks being same top selling track for each customer cannot be determined. The list of tracks purchased for each customer can be retrieved.

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

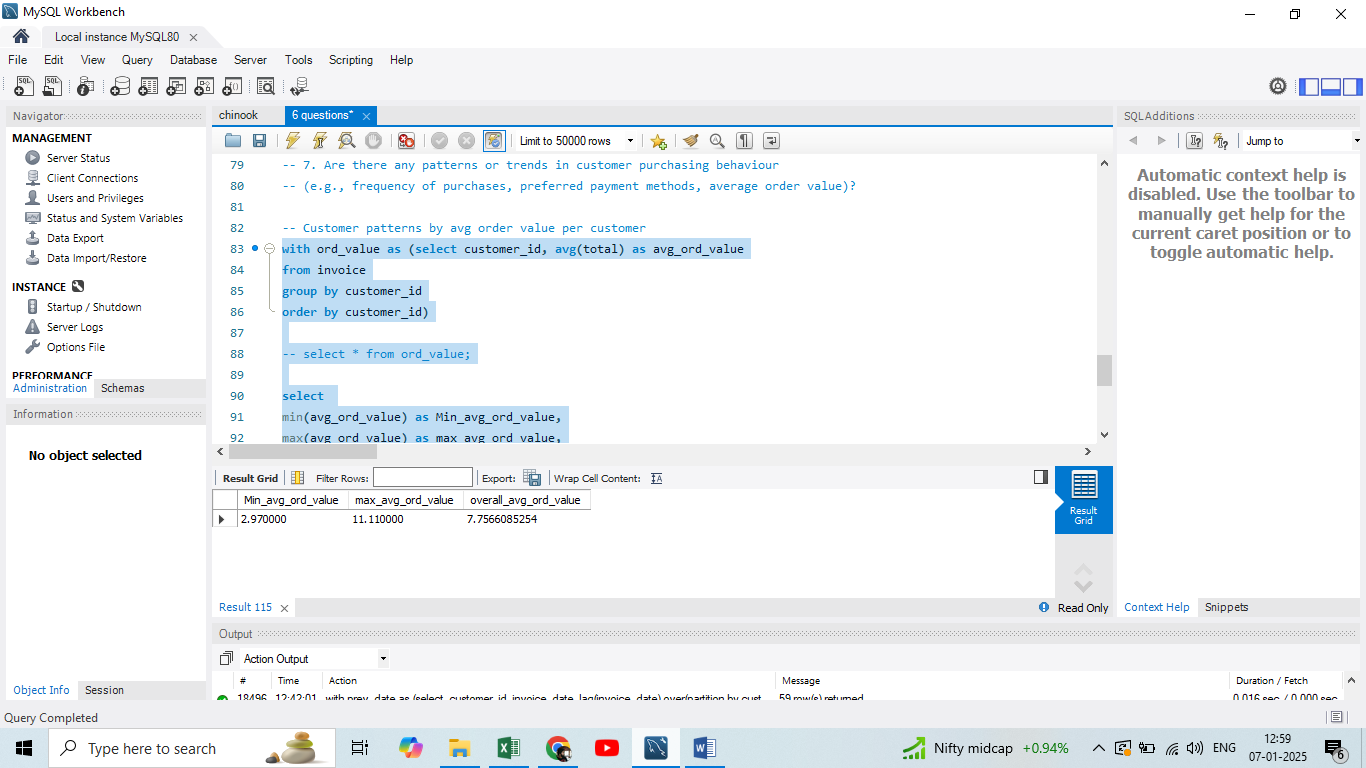
Ans:

Average order value analysis for chinook customers:

Average invoice value for each customer is retrieved from invoice table.

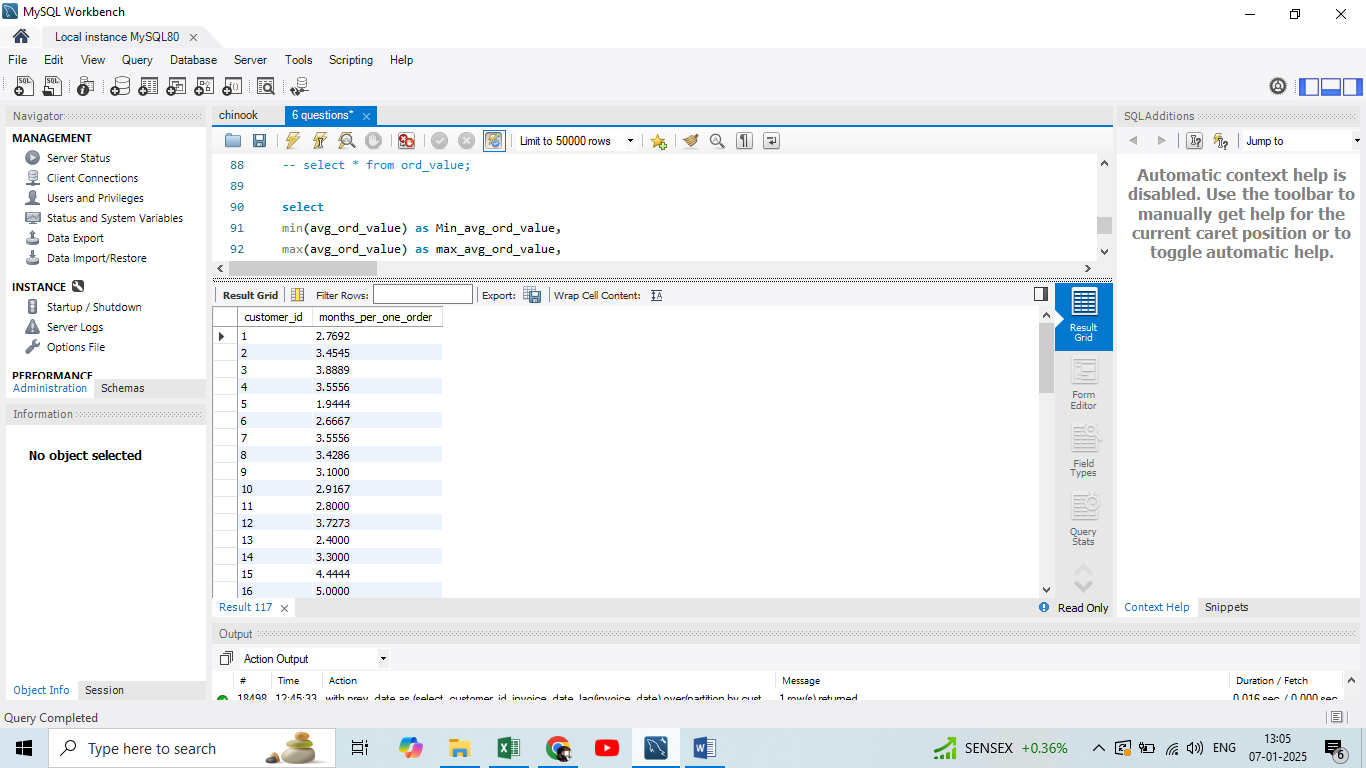


From the above table we can also retrieve the minimum, maximum and overall average value per invoice for entire customer base.



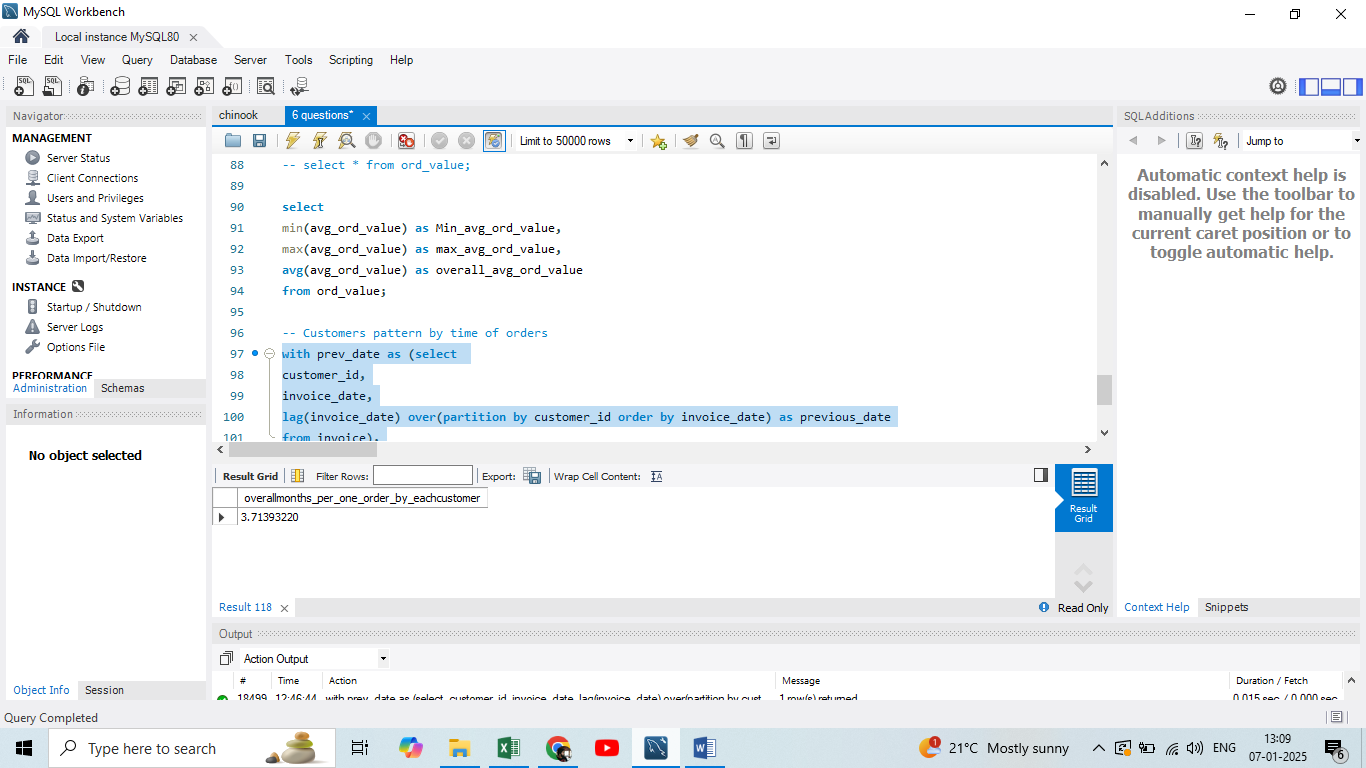
So, for Chinook company average amount per invoice per customer is 7.75

Invoice date analysis of chinook customers:



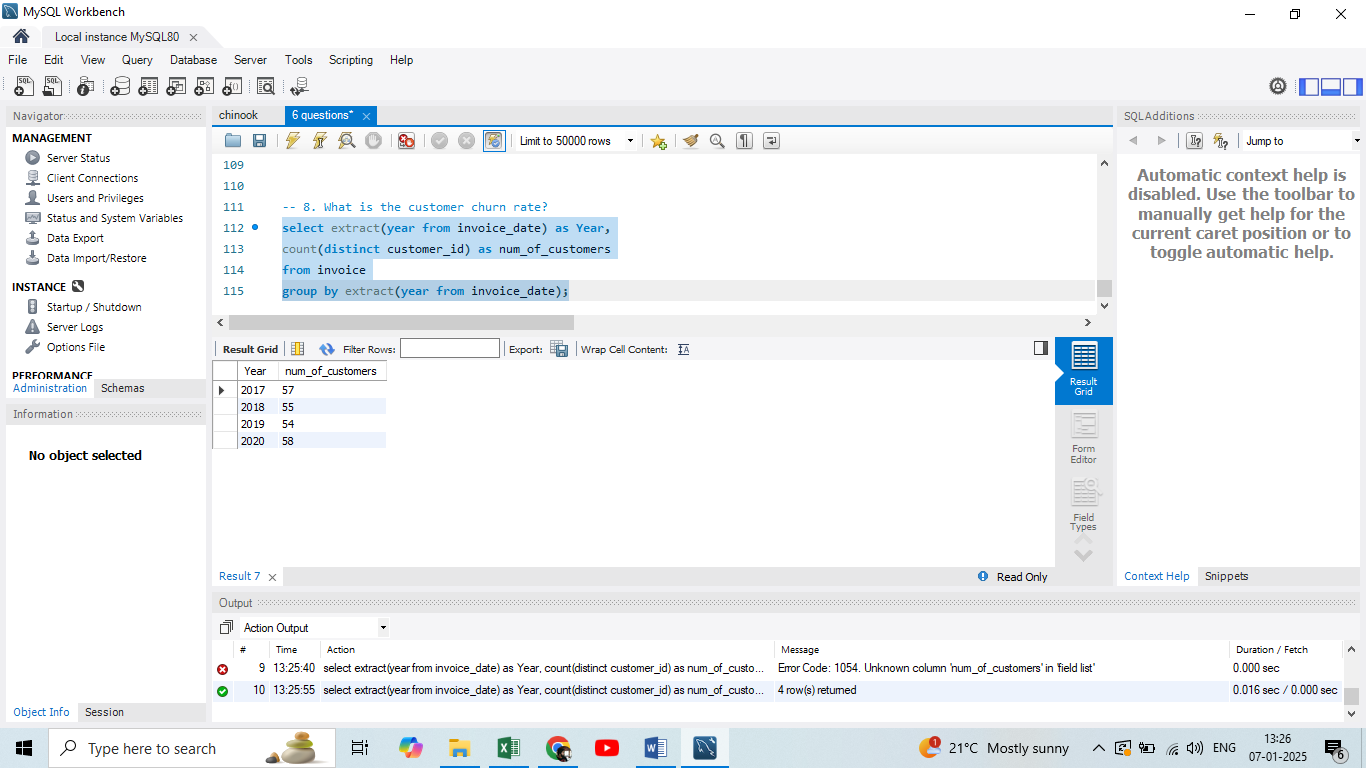
From the above figure it means that customer1 is placing one order per 2.7 months, customer2 in 3.4 months etc.

Likewise for chinook company every customers on an average are placing one order for 3.72 months.

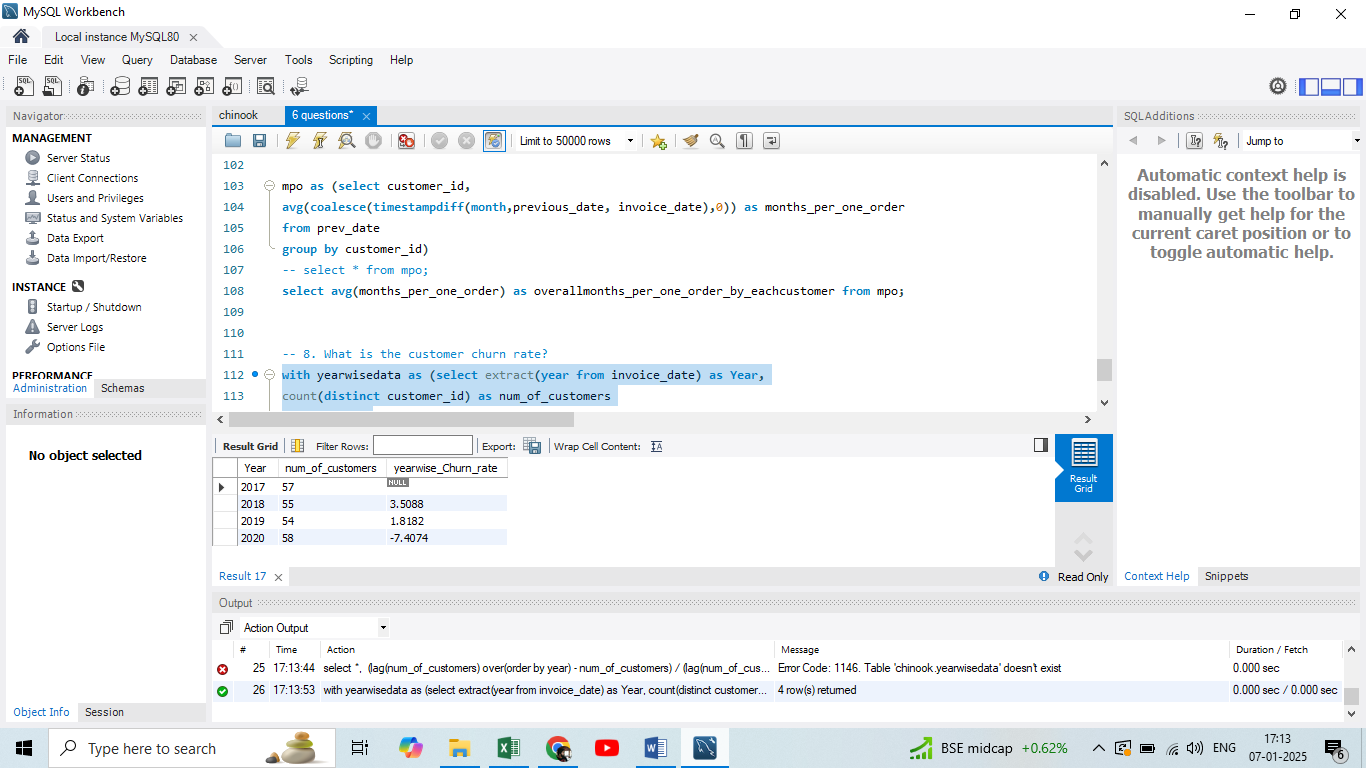


1. What is the customer churn rate?

Ans: Below is the number of customers purchasing tracks from chinook in a year.



So, from the above table year on year churn rate can be calculated as (lost\_customers / initial\_num\_of\_customers) \* 100



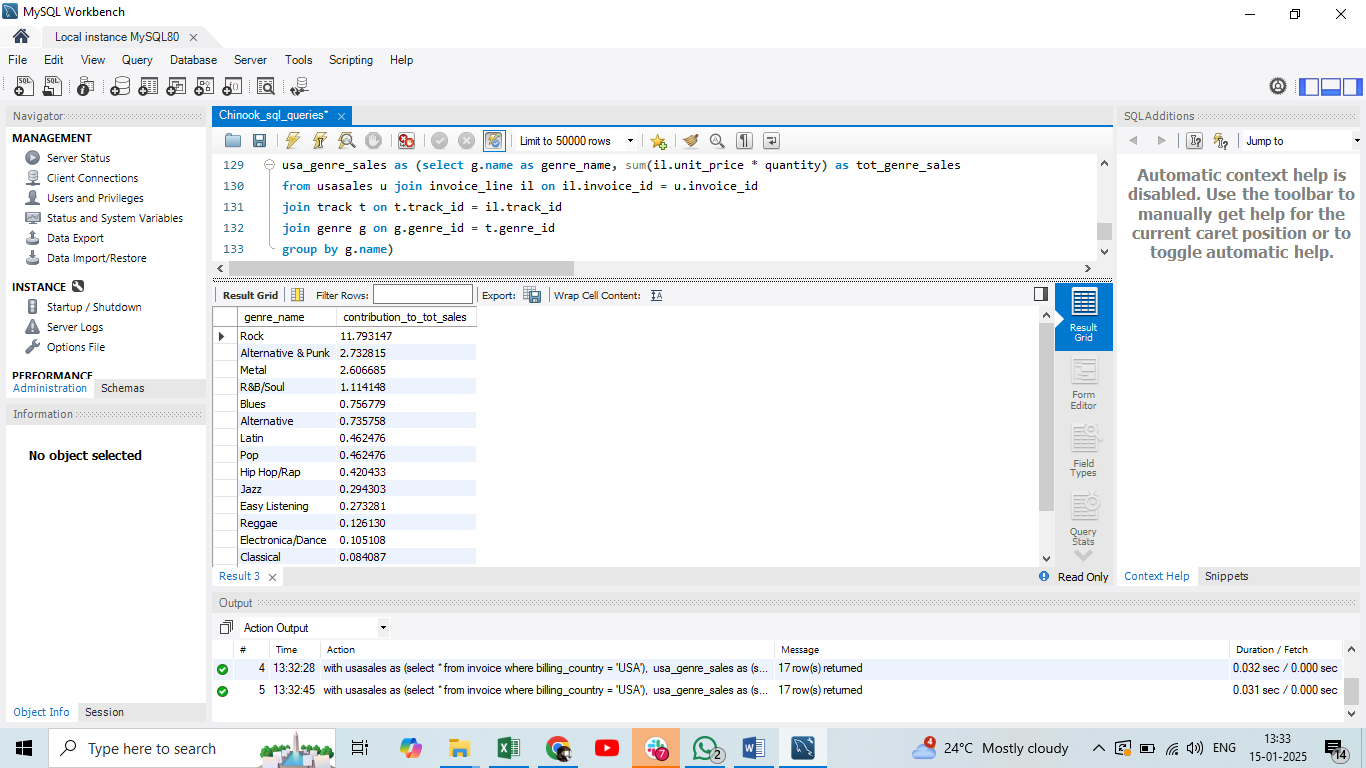
Positive churn rate is bad for company it means it has lost few customers and if churn rate is negative churn rate means it has gained customers than its previous year.

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

Ans: First USA records are retrieved from Invoice table and then it is linked with track and genre table to retrieve the genres in USA.

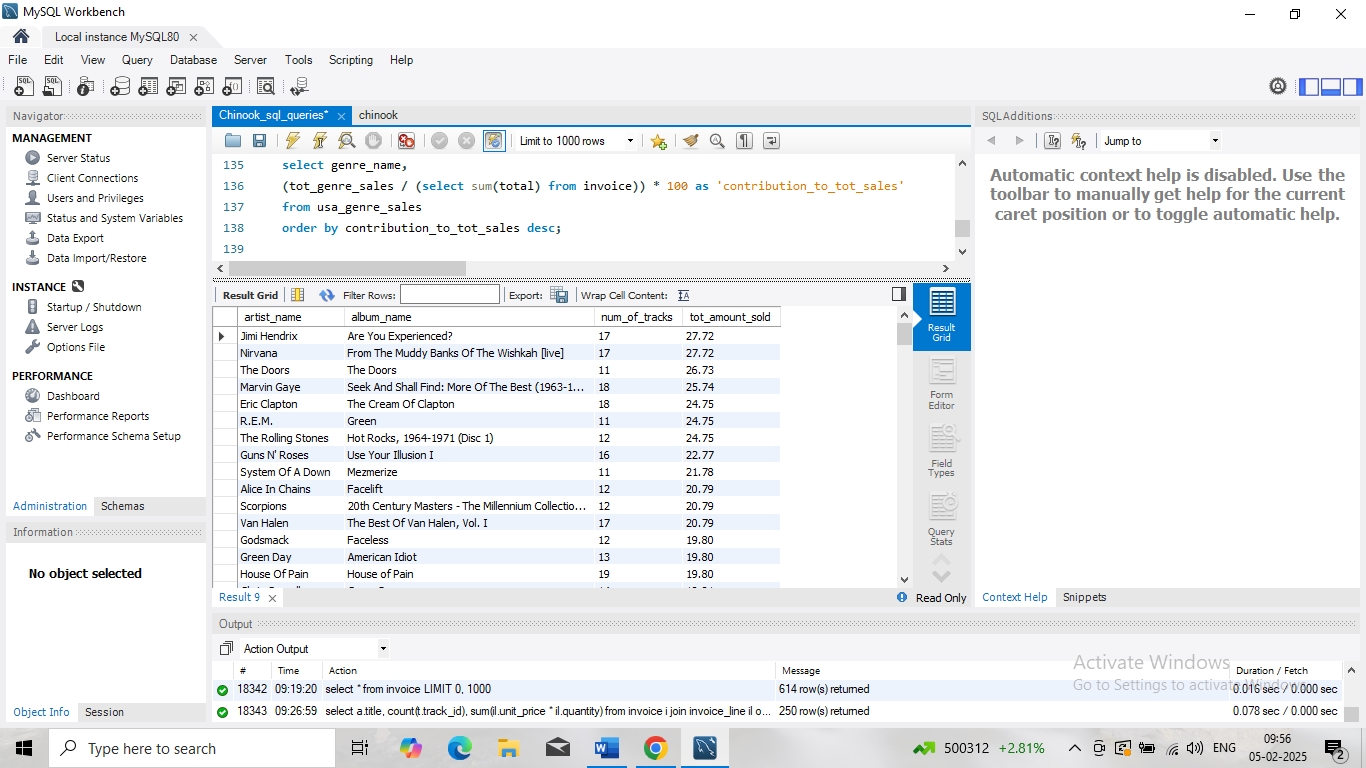
Contribution of each genre is calculated by:

Sales\_of\_genre / Total\_sales \* 100.



So, from above data we can conclude that ‘Rock’ is the popular genre in USA as it is the highest contributor to total sales.

Top selling artist in USA:

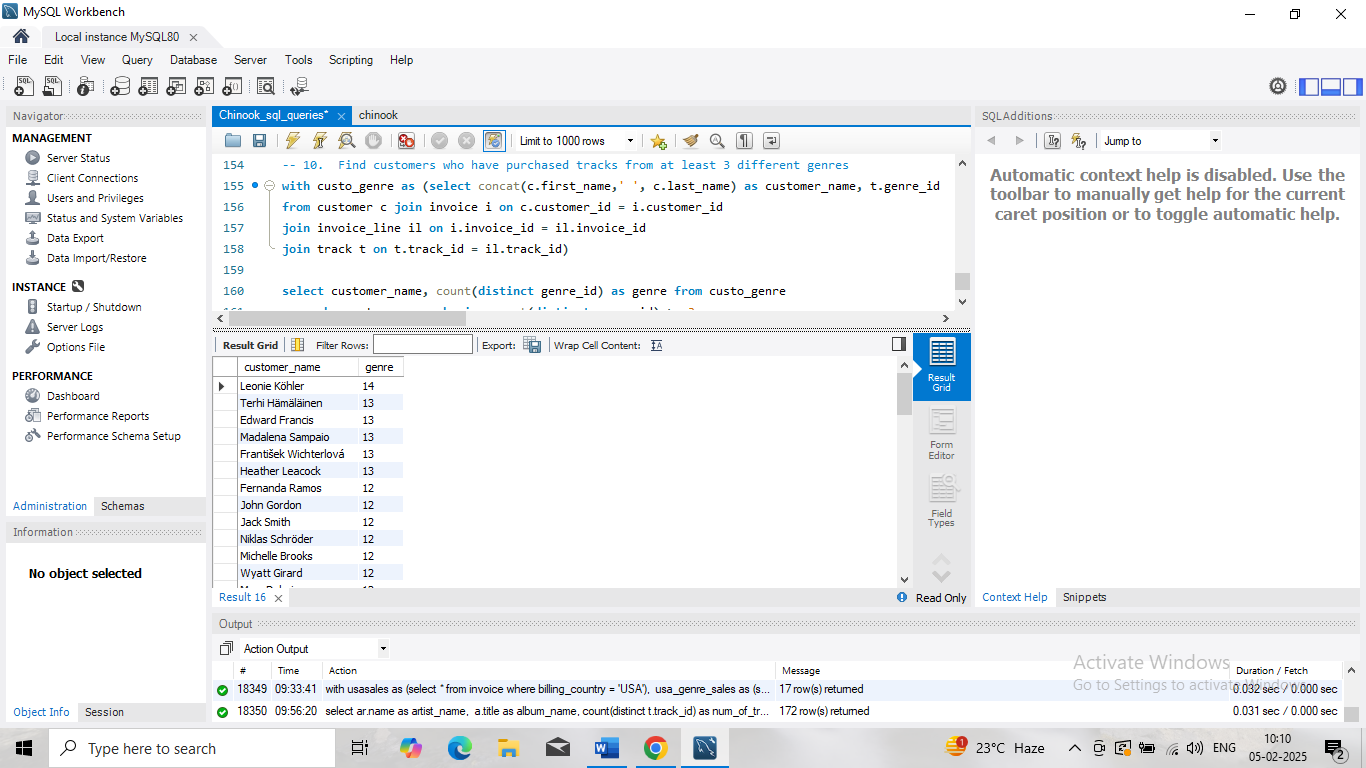


Top selling album in USA is retrieved and then the corresponding artist of the album is retrieved. So, Jimi Hendrix and Nirvana are top selling artists in USA.

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

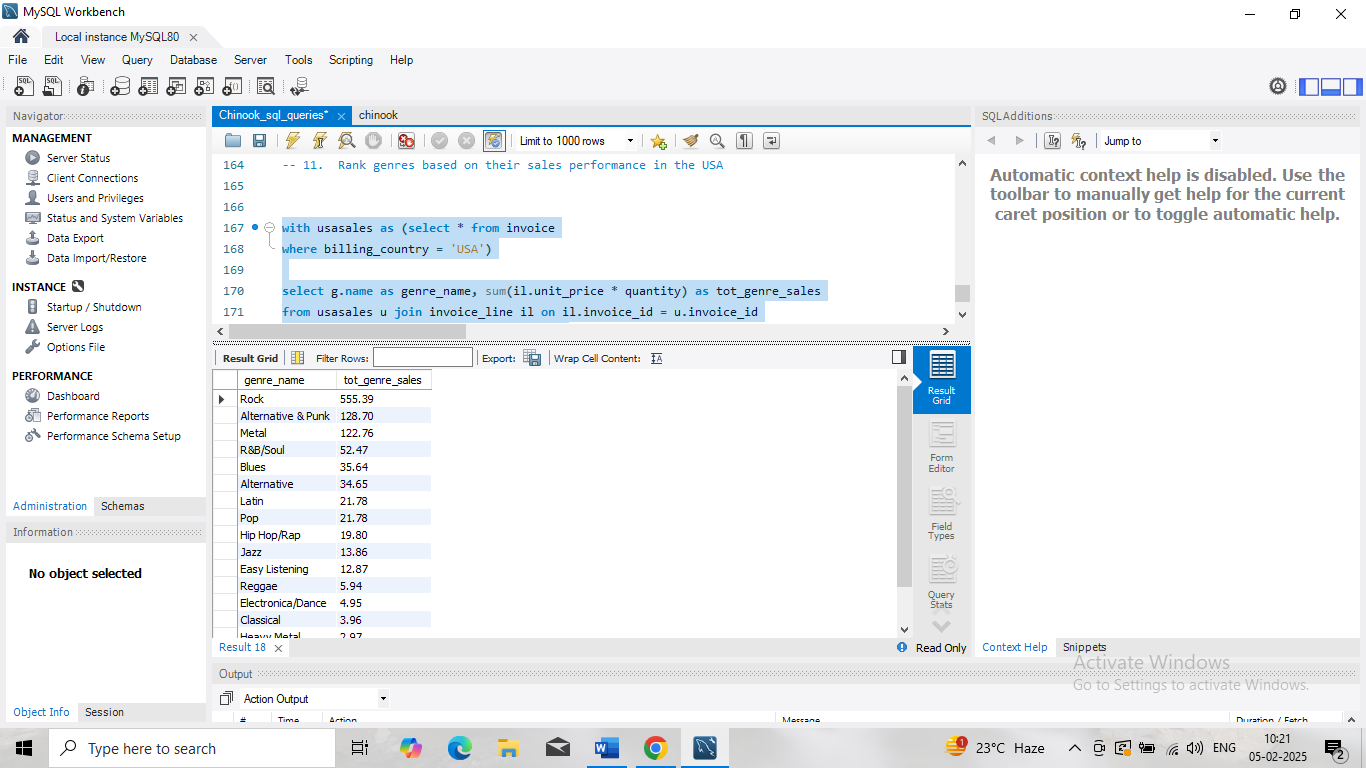
Ans:

customer, invoice, invoice\_line, track tables are joined together to retrieve customers and their count of genres in which tracks are purchased.

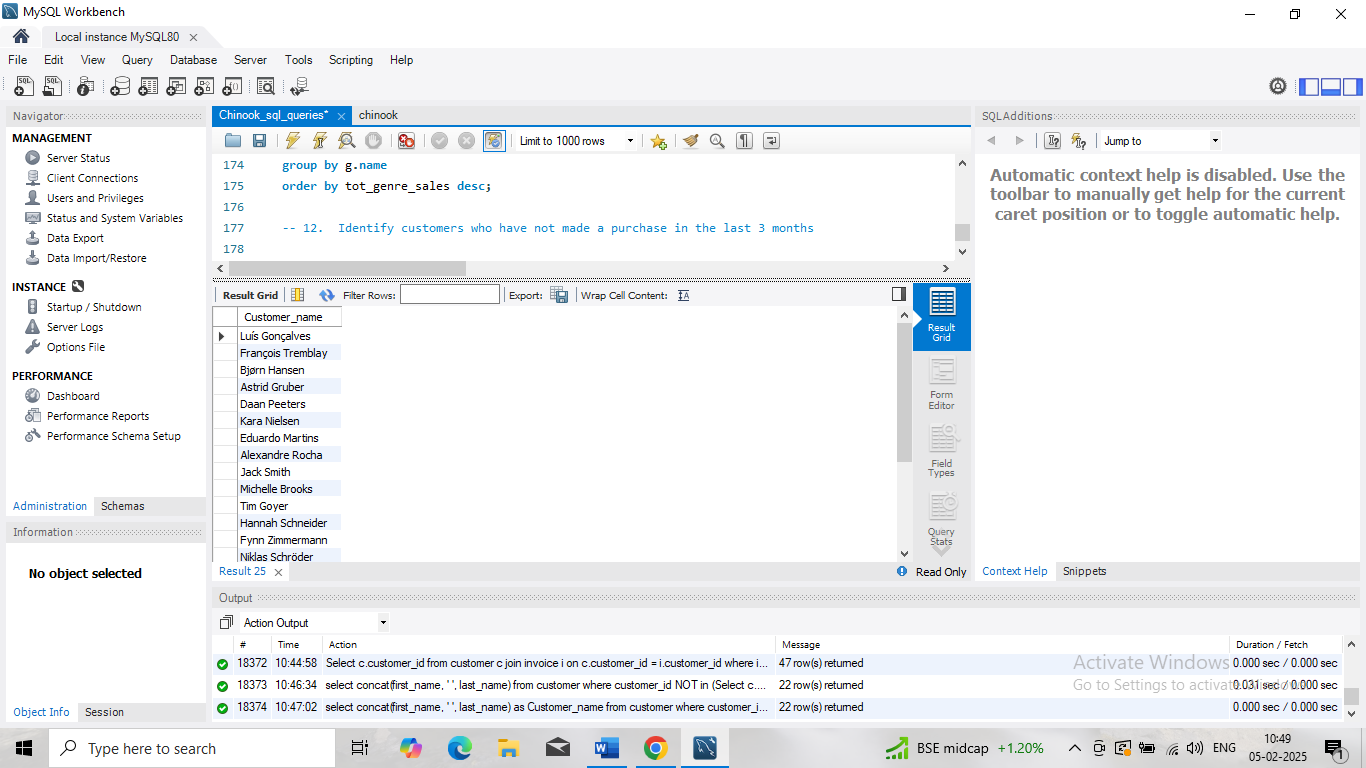


1. Rank genres based on their sales performance in the USA

Ans: Best sold genre in USA is Rock followed by Alternative & Punk and then Metal.



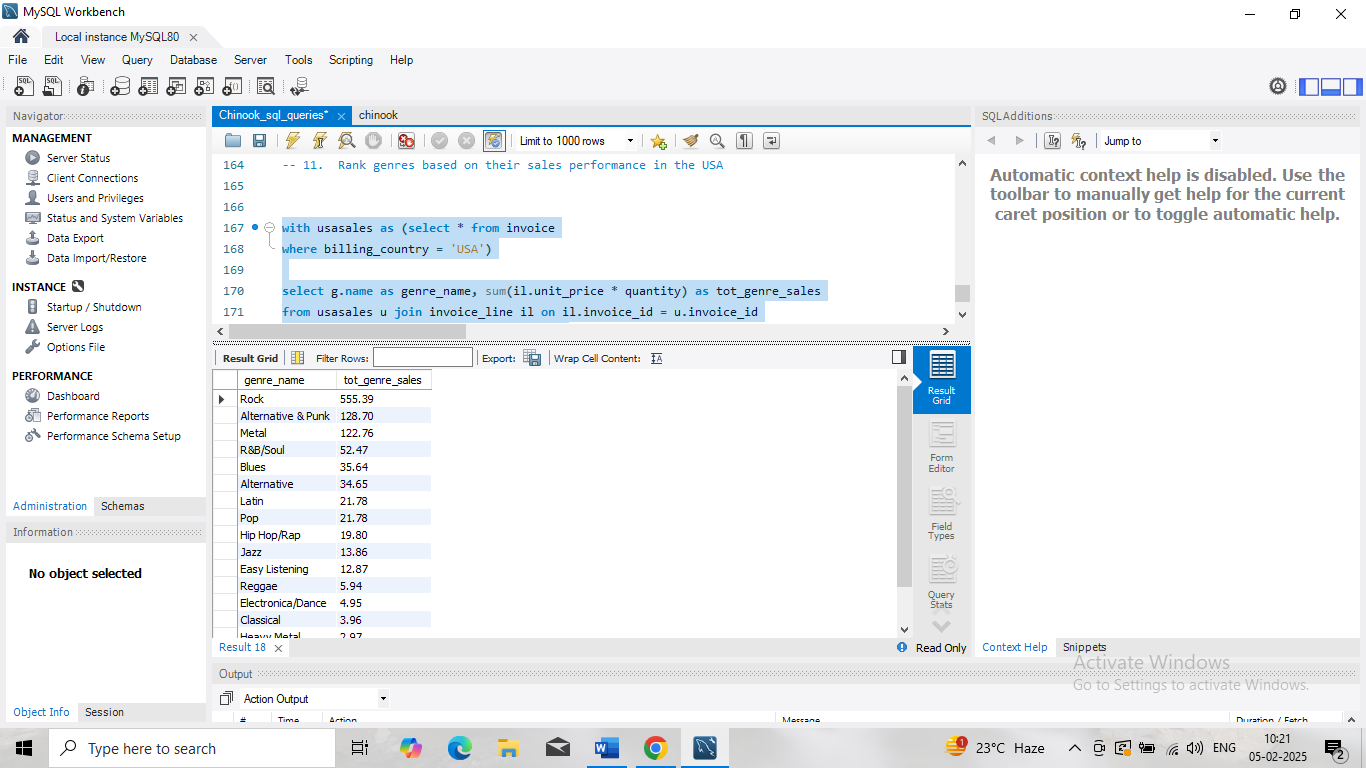
1. Identify customers who have not made a purchase in the last 3 months Ans: First customers who have purchased in last 3 month were retrieved and then they are eliminated from full list of customers.



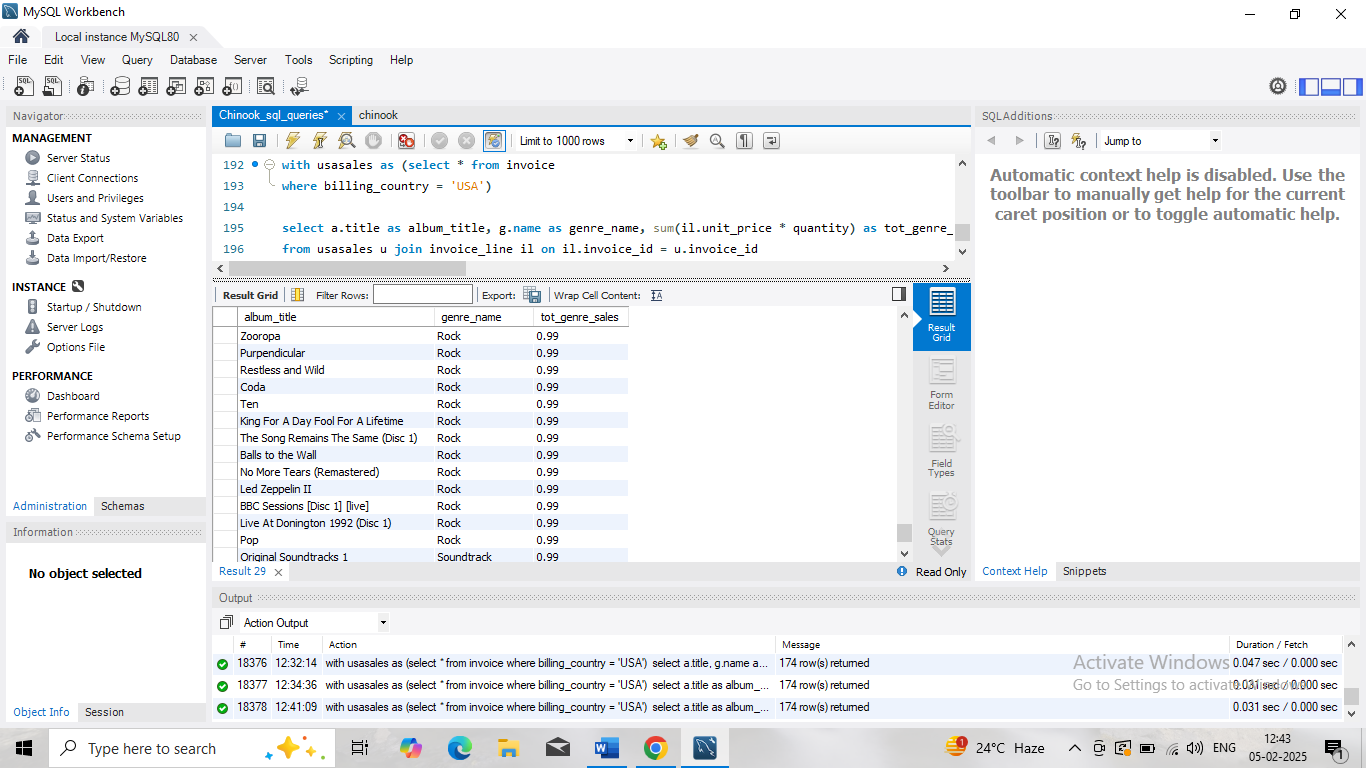
Subjective Questions

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.

Ans:



The above picture shows the top selling genres in USA. Rock, Alternative & Punk, Metal are three top selling genres.



But the above data shows that despite USA people’s affinity towards Rock genre there are many albums in rock genre which are sold very less in USA.

So, any of those albums which are selling in low numbers despite being in rock genre must be advertised and promoted.

I recommend ‘King for a day fool for a lifetime’, ‘Restless and wild’, ‘No more tears’ albums for advertisement and promotion in USA.

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

Ans:



The above figure shows the top selling genres in rest of the countries.

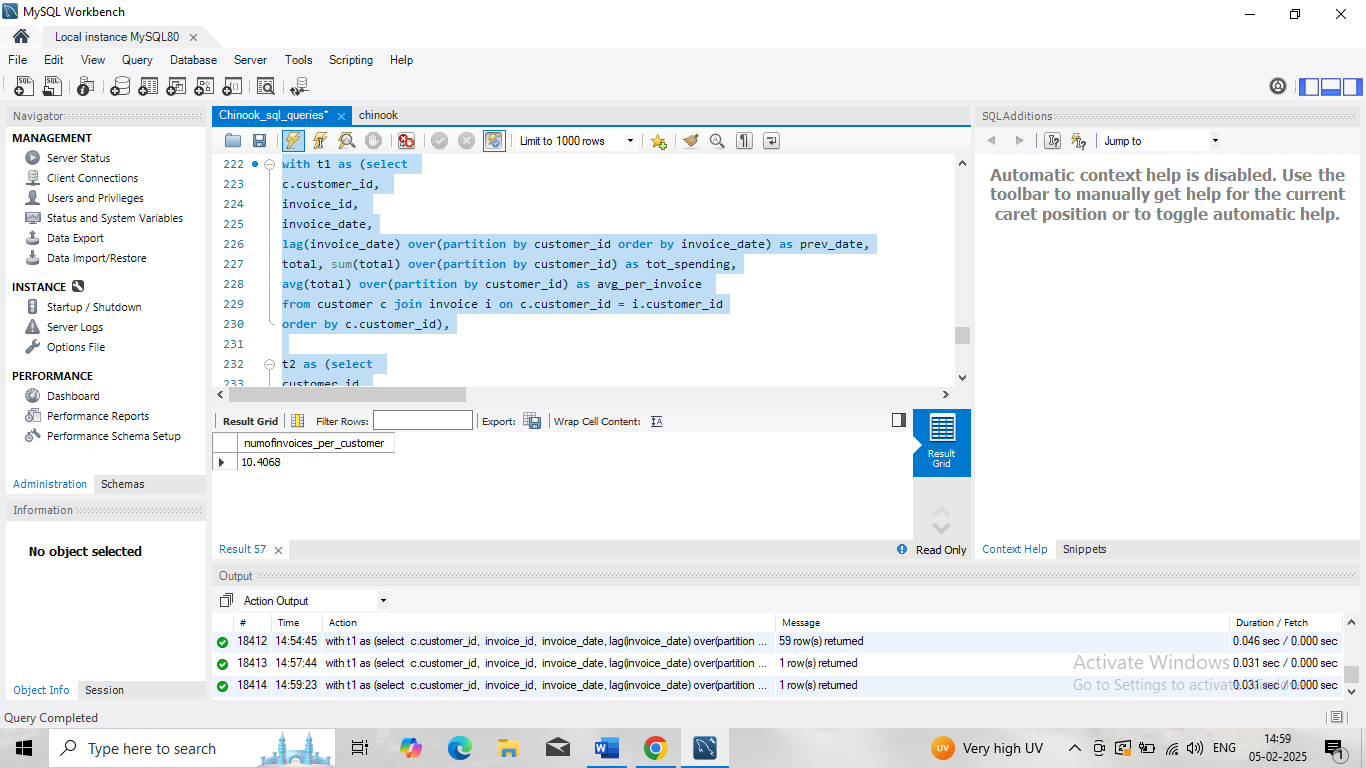
Commonality between USA and rest of the countries genres by sales is Rock stands out to be tope selling genre in both USA and other countries. In fact, top 3 genres are same for both USA and other countries.

Difference is in USA Alternative & Punk and metal are a second and third most sold genre, but in other countries Metal takes the 2nd position and Alternative & Punk takes 3rd position.

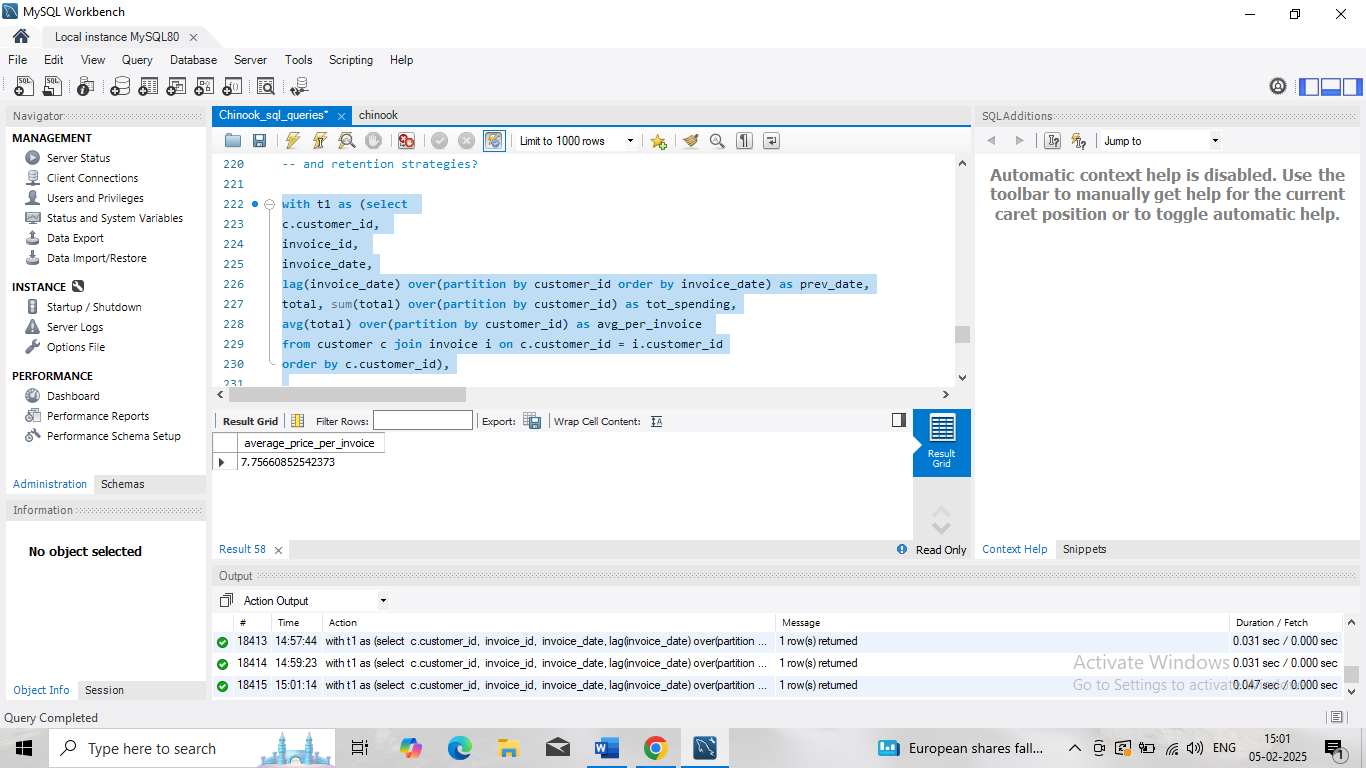
1. Customer Purchasing Behaviour Analysis: How do the purchasing habits (frequency, basket size, spending amount) of long-term customers differ from those of new customers? What insights can these patterns provide about customer loyalty and retention strategies?

Ans:

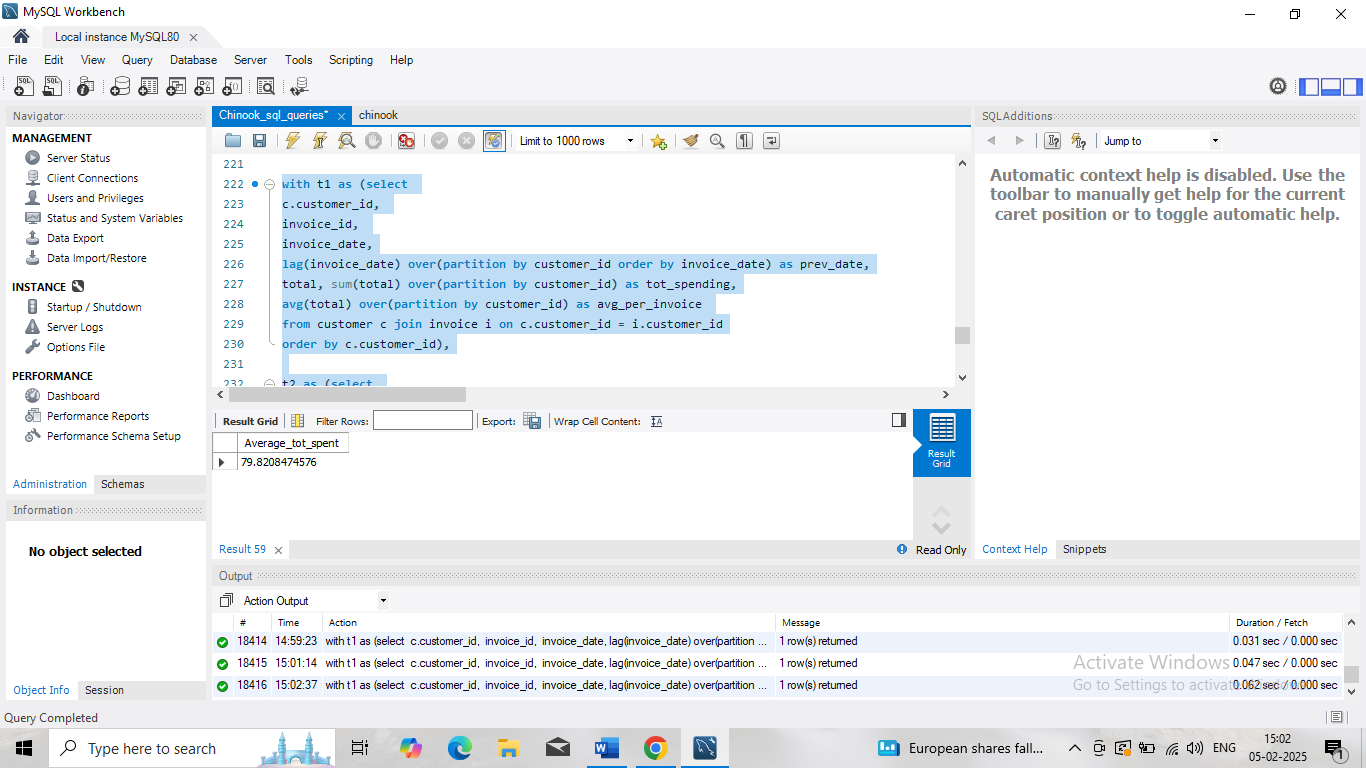
Average number of invoices per customer is 10. It means that customers are purchasing through different invoices and are turning back frequently.



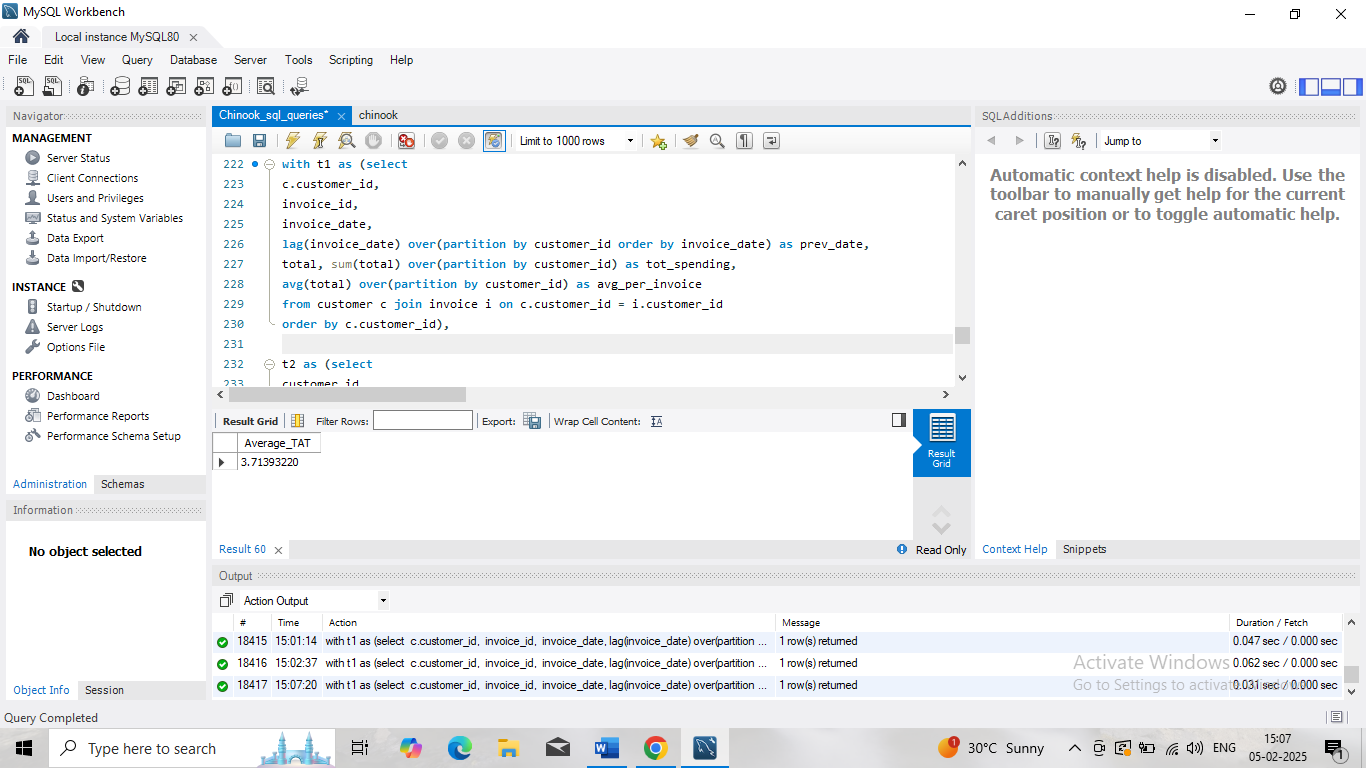
Average amount per invoice is $7.75. On an average this amount is spent by customer on his each invoice.



Each customer through all his invoices, spends around $79.

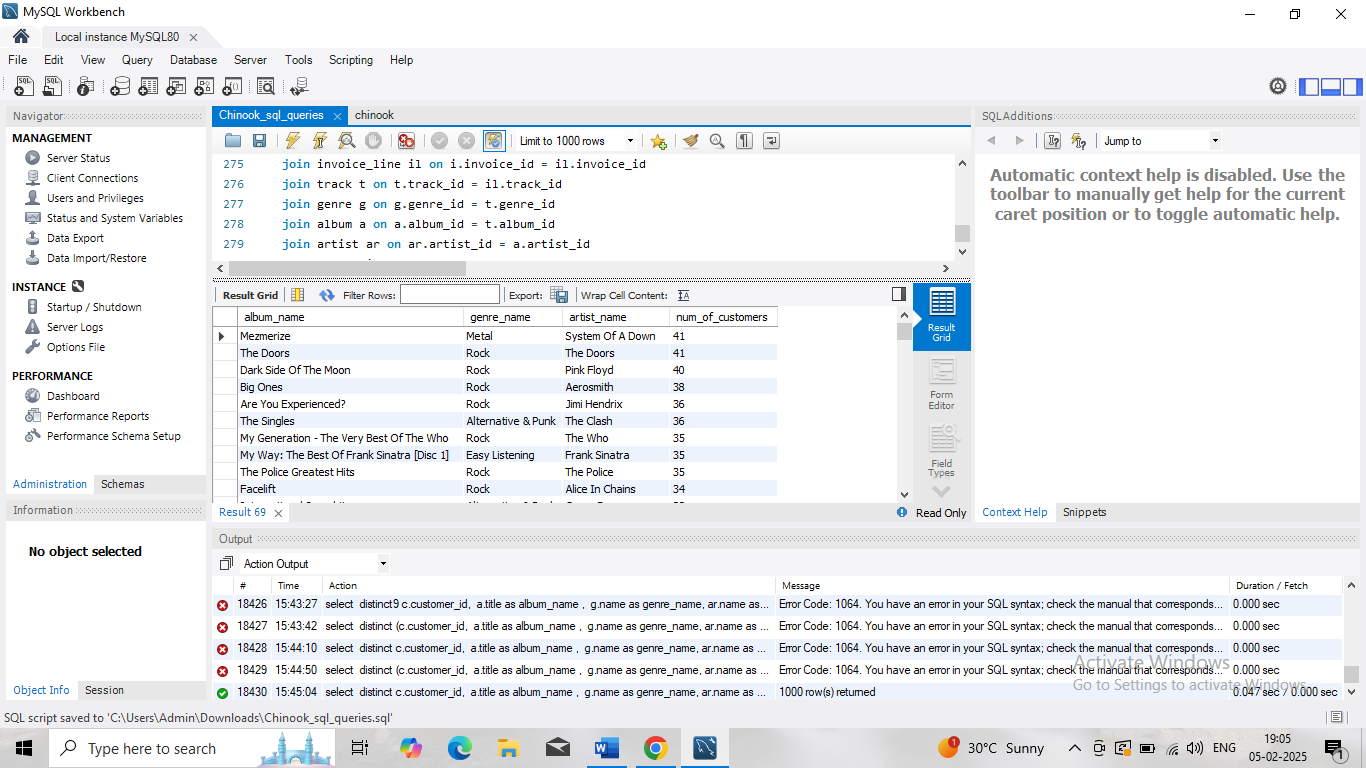


Each customer is turning back to purchase. But they are not coming back in any particular pattern. So, if average turn around time is calculated it is 3.7 months. It means that each customer is coming back at least in 3.7 months.



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?

Ans:



Above data shows the combination of album, genre and artist and the number of customers who are purchasing it.

From the data the further product recommendation can be made like which artist’s which genre is frequently purchased.

It will also help in advertising and promoting artists and genres.

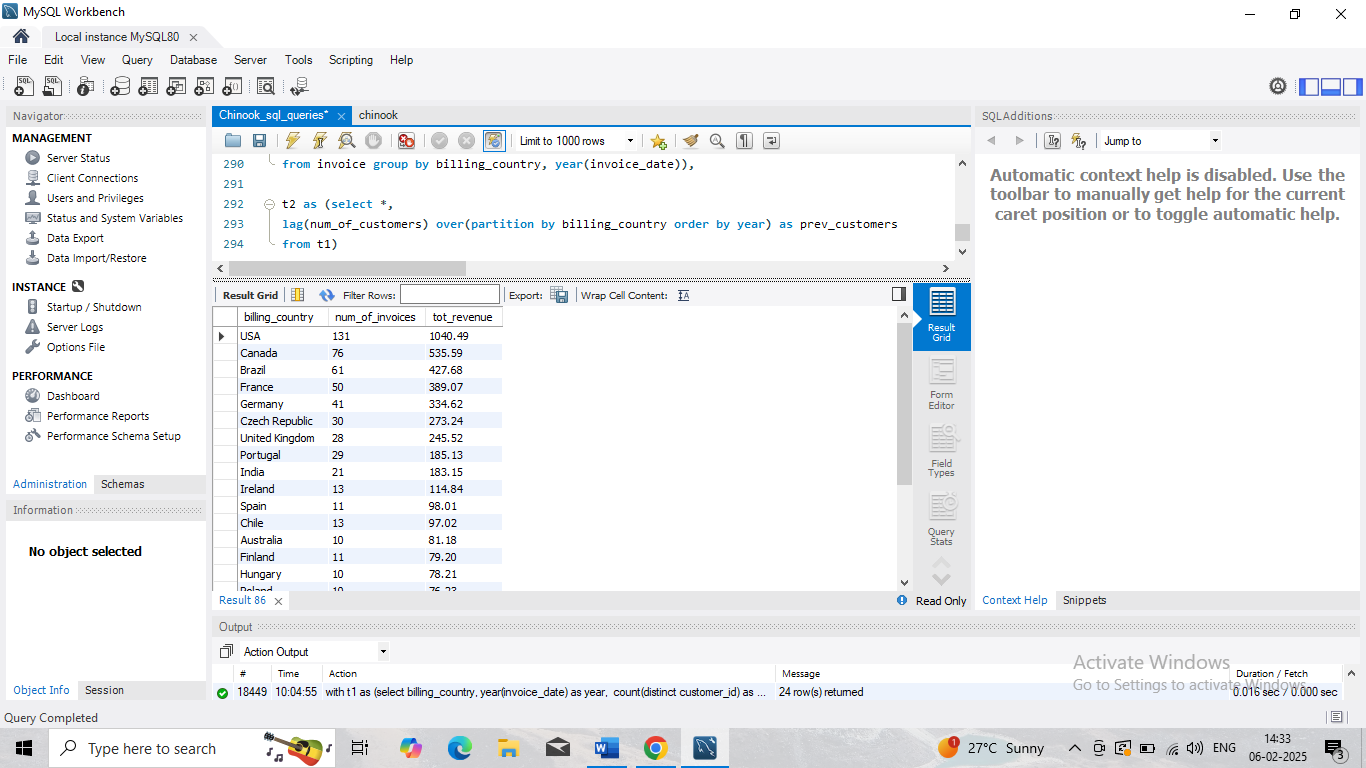
1. Regional Market Analysis: Do customer purchasing behaviours and churn rates vary across different geographic regions or store locations? How might these correlates with local demographic or economic factors?

Ans:



Churn rate will be different for different regions, negative churn rate is good for a company, that means in India, France, Canada and USA number of customers are increasing. If it is zero means customers are neither increasing nor decreasing example for this case is Australia, Austria etc.

If the churn rate is positive then that means company is losing its customers.



Purchasing behaviour depends on purchasing power of a people in the country. USA stands first in revenue yielding country and reason is obvious that it is a developed country. Above table shows country wise revenue in descending order.

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?

Ans:

If country wise data is taken to analyse customer churn rate, fortunately chinook do not have a positive churn rate in any location.

In fact, some countries have shown a negative churn rate that means in these countries number of customers are increasing.



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?

Ans:

Customer Lifetime Value (CLV) modelling helps businesses understand the long-term financial value of their customers, which can be crucial for designing targeted marketing strategies and loyalty programs. By leveraging customer data such as **tenure, purchase history**, and **engagement** metrics, you can predict CLV for different customer segments and develop insights into behaviours that drive customer retention or churn.

This is a simple model that calculates CLV based on past purchasing behaviour. You look at how much each customer has spent in the past and assume they will continue purchasing at the same rate.

CLV=

Average Purchase Value×Purchase Frequency×Customer Lifespan.

* + Average Purchase Value: Total revenue generated by a customer divided by the number of orders.
  + Purchase Frequency: The number of purchases a customer makes in a specific period.
  + Customer Lifespan: The number of years a customer is expected to continue purchasing.

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?

Ans:

To measure the impact of promotional campaigns (discounts, events, email marketing) on customer acquisition, retention, and overall sales, we can use a variety of methods and metrics.

Some of them are:

* + - 1. **New Customer Count:** Track the number of new customers who make their first purchase during or after the campaign.
      2. **Customer Acquisition Cost (CAC):** Divide the total cost of the campaign by the number of new customers acquired. This helps assess the efficiency of the campaign in attracting new customers.
      3. **Conversion Rate**: Measure the conversion rate of individuals exposed to the campaign (email open rates, event attendance, clicks on ads) and how many actually became paying customers.

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

Ans:

If objective and subjective questions were not given then my approach would be:

1. First I would analyse all the tables and then understand the primary key and foreign key in the data set

2. In my understanding Invoice table is very important because it is fact table. I would have analysed the sale amount by each country, then average spending by each customer and number of tracks purchased by each customer etc.

3. I would have retrieved popular artists, genres and albums in different region to help further promotion and advertisement.

4. I would have analysed churn rate of company as a whole and also according to the region.

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?

Ans:

Structure of the table can be altered by Data Definition language of MySql. Sql Query to add a column ‘ReleaseYear’ to Albums table is: ALTER TABLE Albums add ‘ReleaseYear’ INT;

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.

Ans:

