Operation Analytics and Investigating Metric Spike

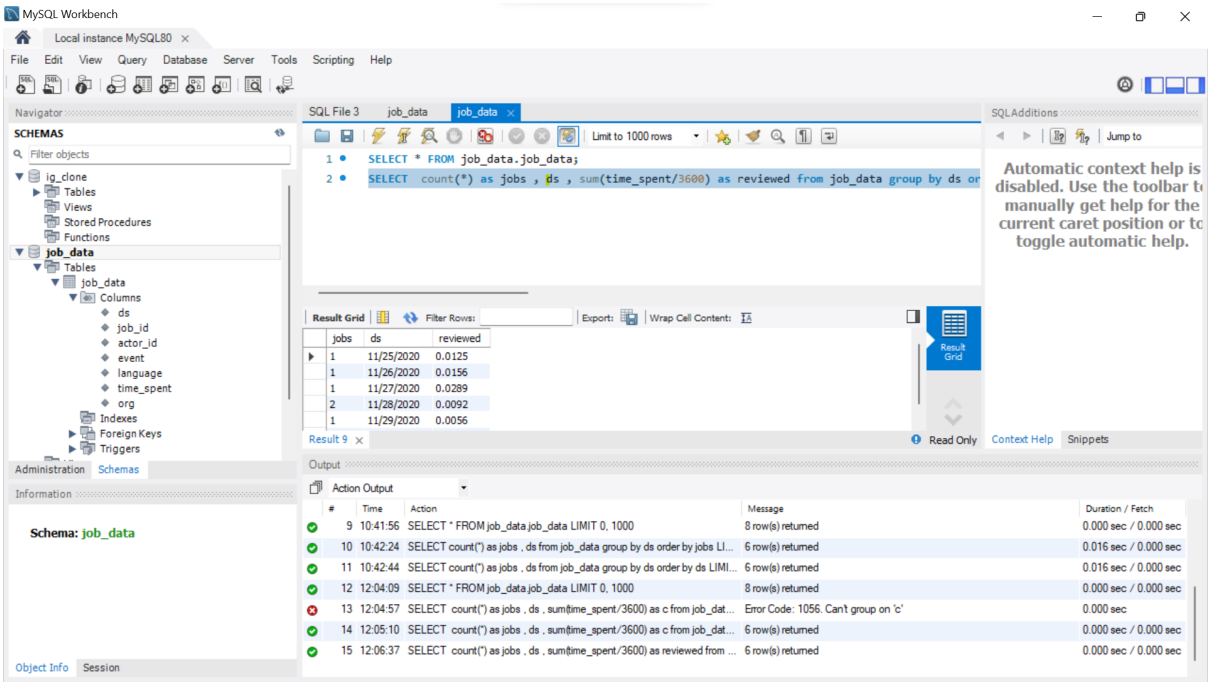
This is the third project as a part of training.This project requires an intermediate knowledge of SQL so doing this project is very much beneficial as it takes our knowledge to the next level. It has two case studies of which case study 1 is a job data analysis with four queries and case study 2 is investigating metric spike with five queries. Answering these queries will gain us more experience and confidence in SQL. There will be some queries in which we get stuck but it is all a part of self learning because to clear these doubts we research on the internet leading to create a strong foundation on basic concepts which is far better than spoon feeding.

Approach

First we need to download all the datasets provided. There are two case studies. Case study1 contains a small dataset so I imported it into mysql using import wizard, but the dataset of case study 2 is very large which will consume a lot of time if we import it using import wizard so import it by following the instructions provided in the given video.

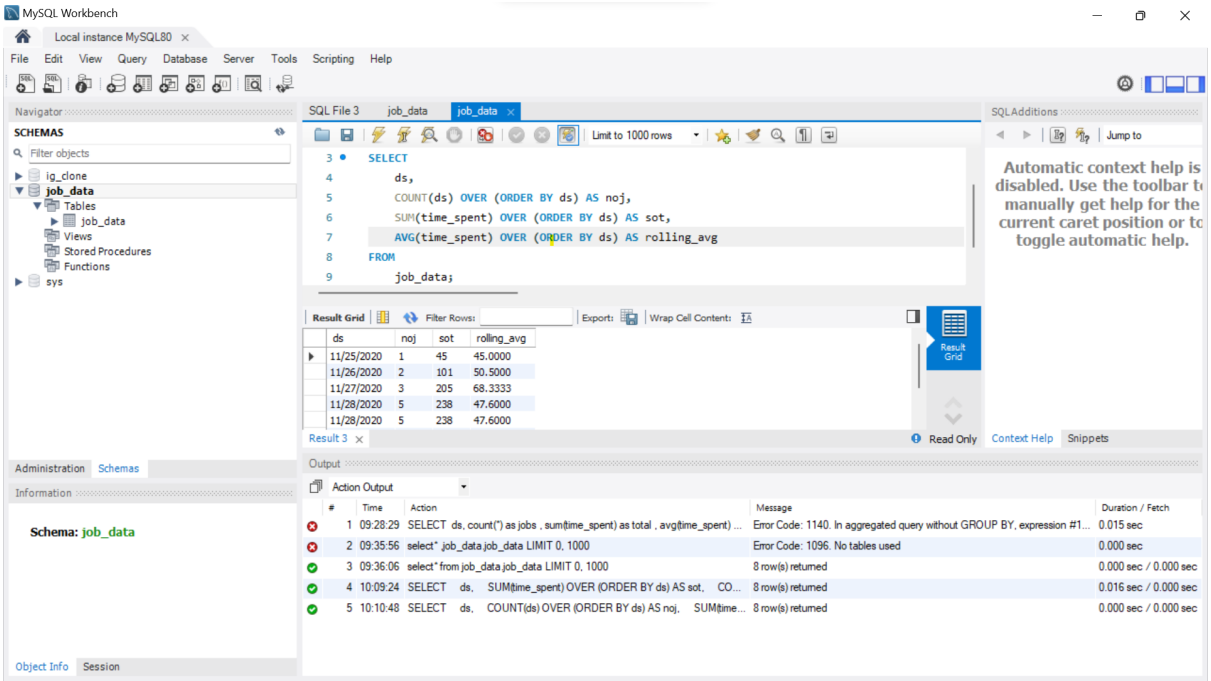
Case study 1:

1. Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.



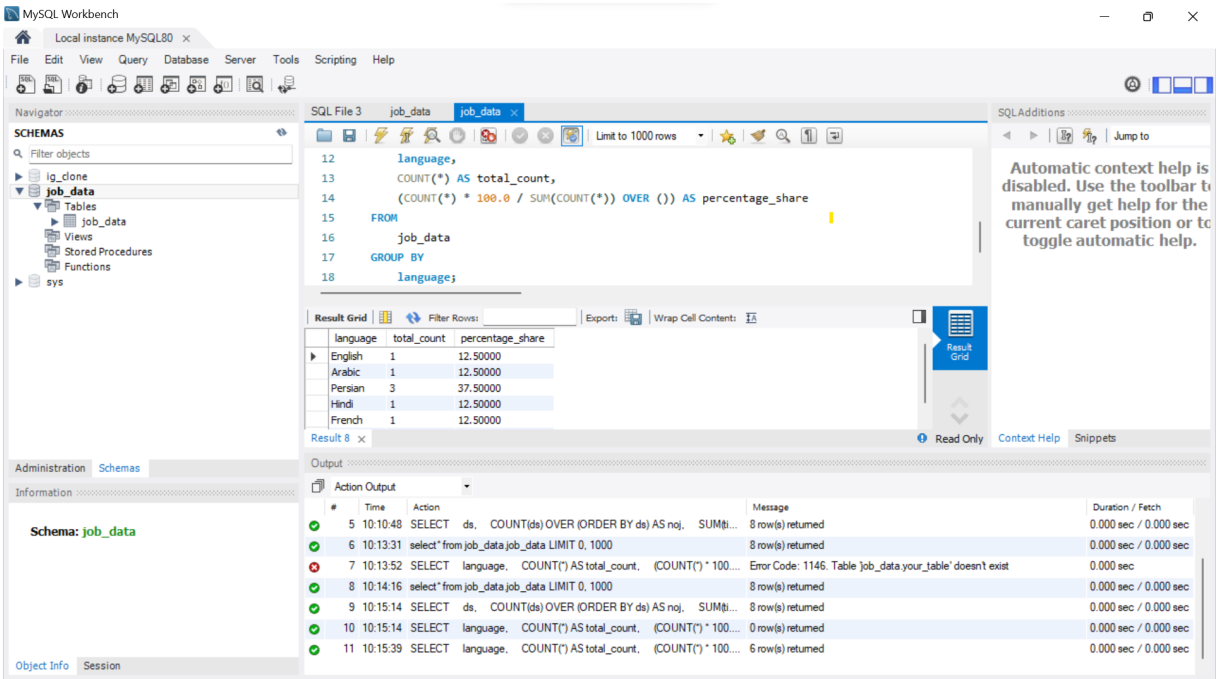
Here, I need to calculate the number of jobs that are reviewed per hour for each day. Since the dataset is in November ,I did not specify the month in the query. I counted the number of jobs and I used sum() on the time_spent and divided it by 3600 to get the review per hour since the time_spent is given in seconds and group and order it by the date.

2.Write an SQL query to calculate the 7-day rolling average of throughput. Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why.



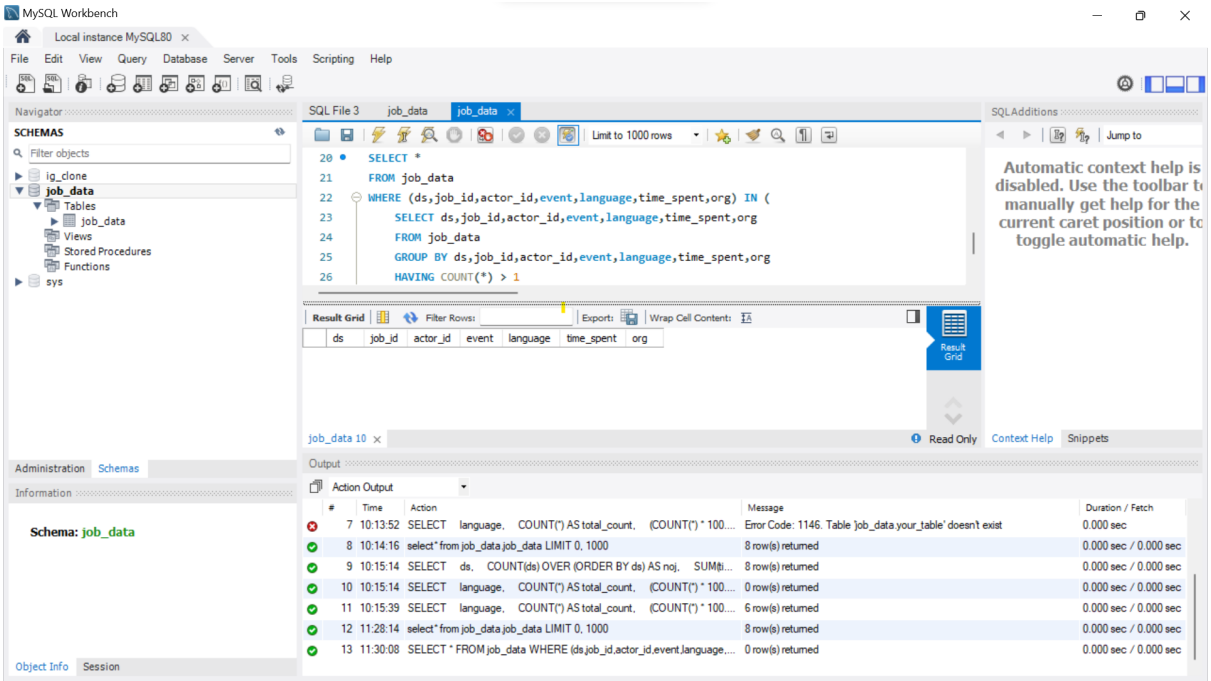
Here, I am required to calculate the 7-day rolling average throughput but in the dataset only some days are given. I counted the number of jobs by counting the dates , I performed sum() on the time_spent and also calculated the average . I used the window function in all these rather than group by.
I prefer using the daily metric for the given date but if the dates are consistent and if the dataset is long ,I would prefer a 7-day average throughput.

3.Write an SQL query to calculate the percentage share of each language over the last 30 days.



Here, I have to calculate the percentage share of each language so I selected the language and counted each of them and then i calculated the percentage share by calculating the count*100/sum of total count. I did not specify the calculation for the last 30 days as the dataset does not contain much data.

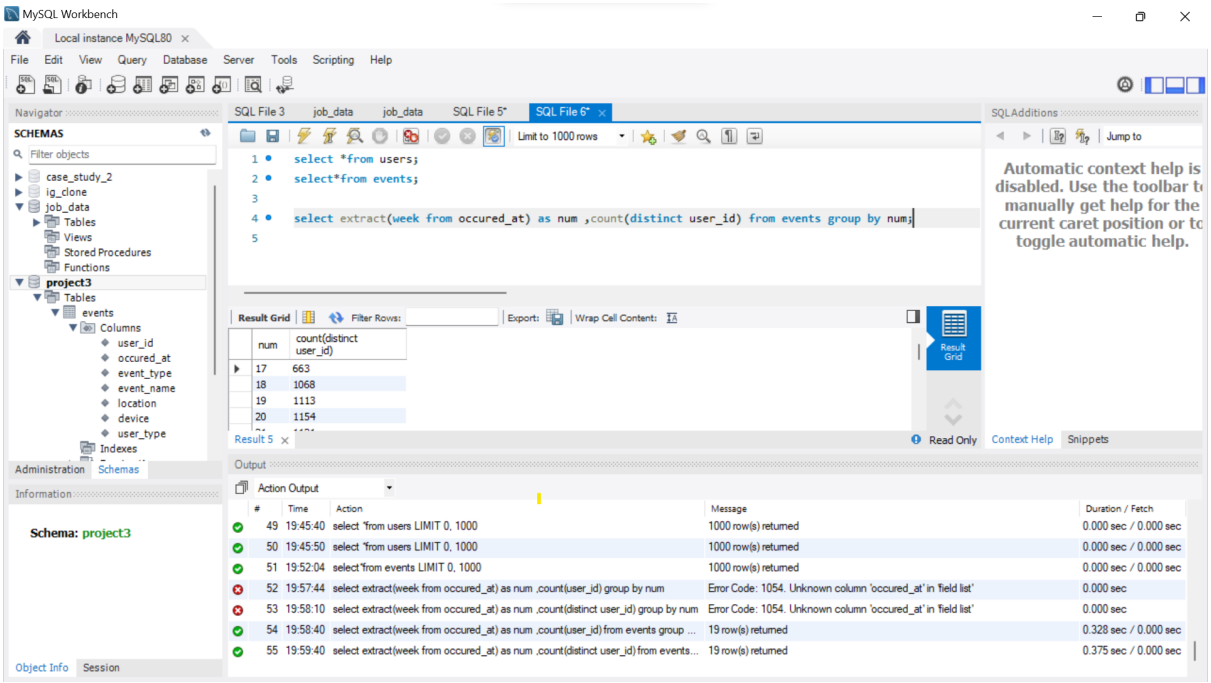
4.Write an SQL query to display duplicate rows from the job_data table.



Here, I need to find out the duplicate rows. One way to find out is to check for the rows of which all the columns are repeating with any previous rows. To do that I selected all the rows from the job_data which are repeating by giving a condition to select those which have the count of having the same column greater than one.

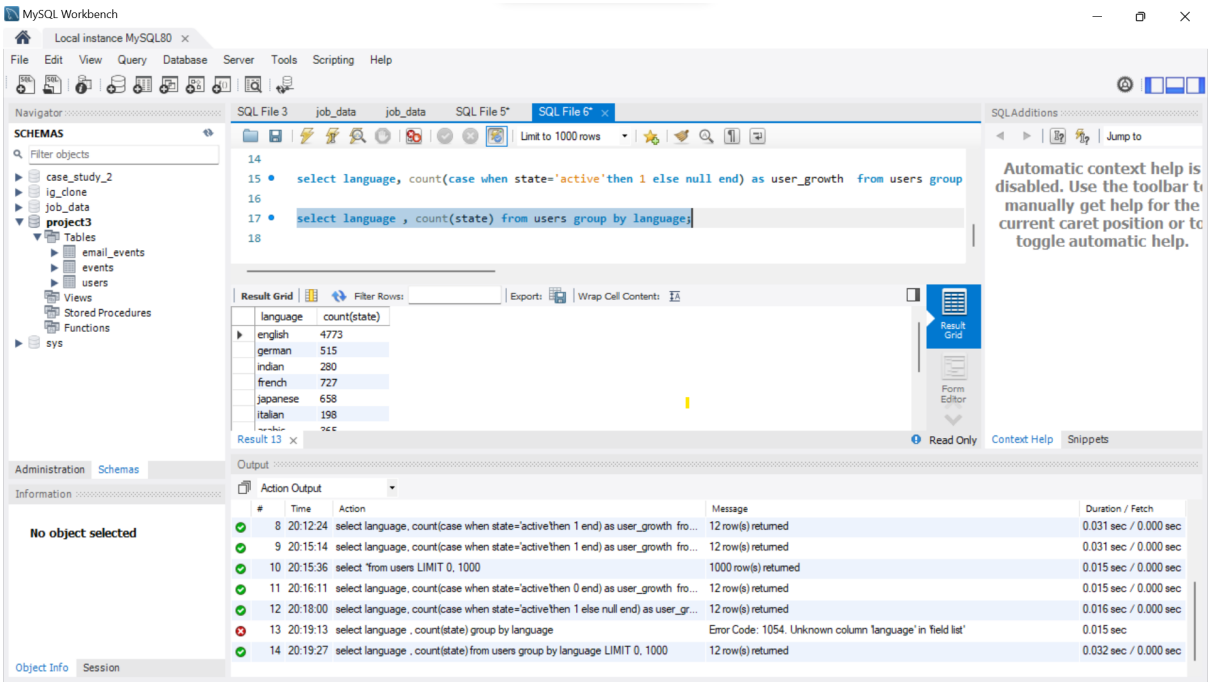
Case study 2:

1. Write an SQL query to calculate the weekly user engagement.



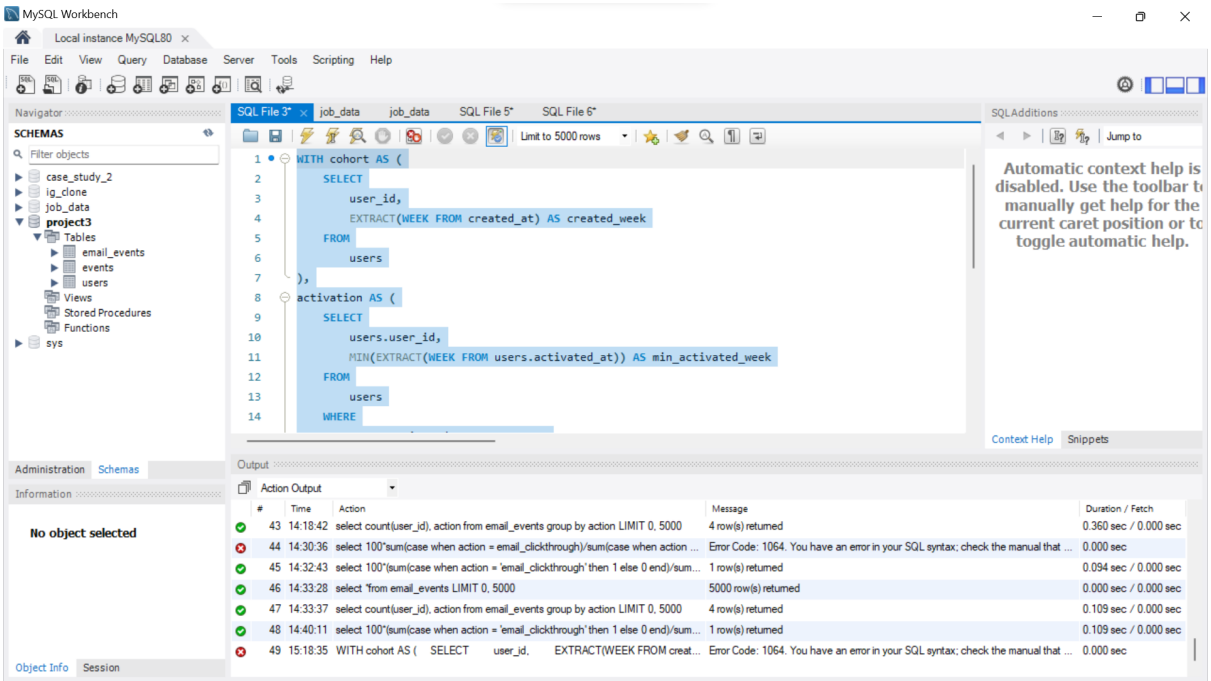
Here, I am required to calculate the weekly user engagement , so I used extract() to get the weeks and count the number of user_id of that particular week . This way I got the number of users present in the event for each week.

2.Write an SQL query to calculate the user growth for the product.



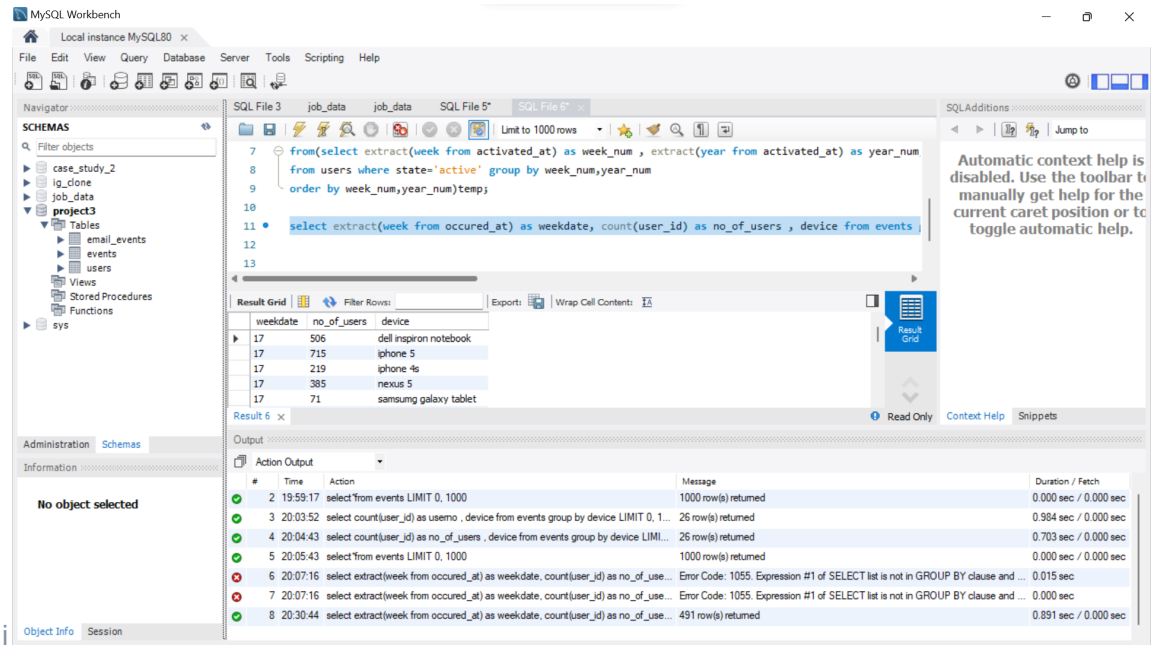
Here, I am required to calculate the user growth of the product. So for that I counted all the active states per language and presented it in the result grid. However , I am not sure if this is the required result.

3.Write an SQL query to calculate the weekly retention of users based on their sign-up cohort.



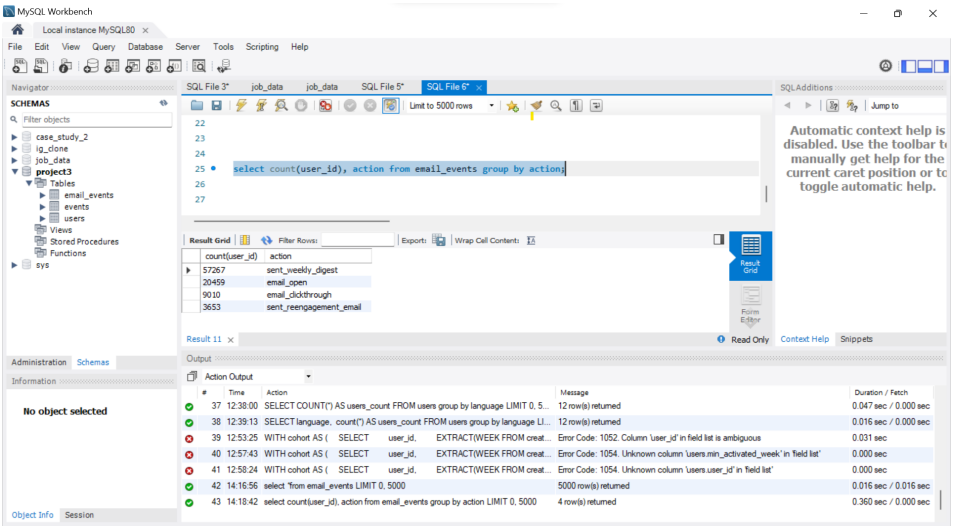
Here, I am required to calculate the weekly retention of users.I tried different ways but i was not getting the desired result.

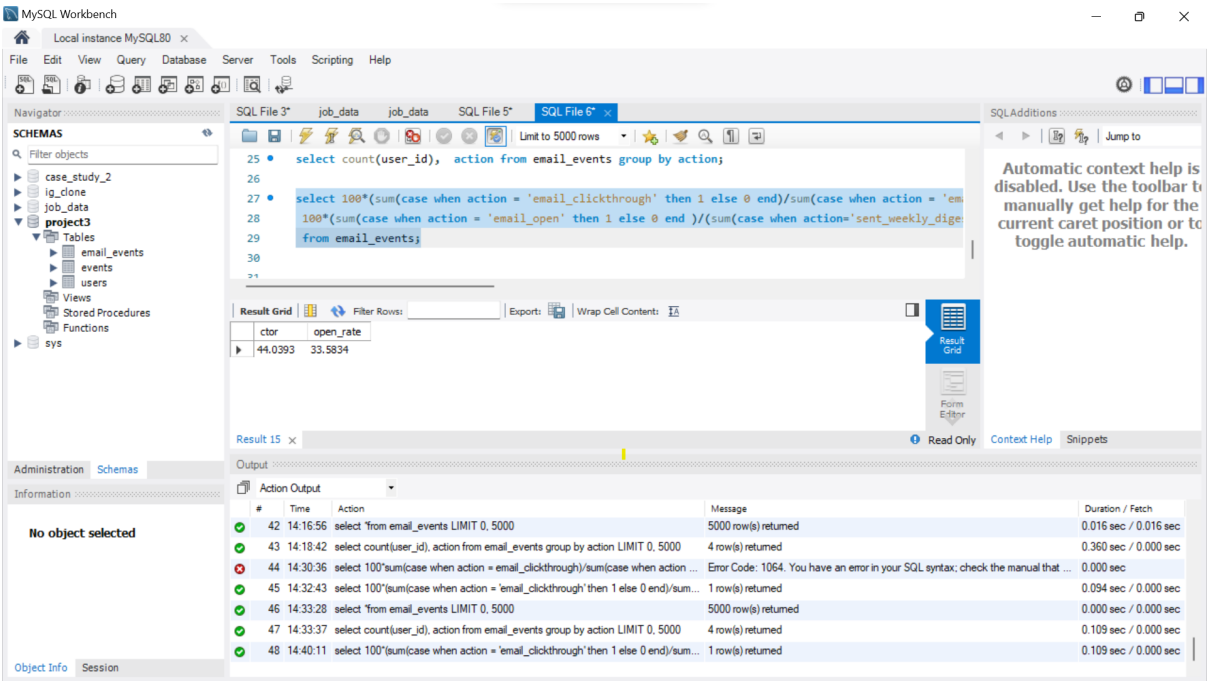
4.Write an SQL query to calculate the weekly engagement per device.



Here, I have to calculate the weekly engagement per device. So to get the week, I first extracted the week using extract() from occured_at and I counted the user_id and group it by week to get the desired result.

5.Write an SQL query to calculate the email engagement metrics.





Here, I am required to calculate the email engagement metrics.So, first I check the different actions present in the database. THen i calculated the open rate and CTOR by applying the formula. CTOR is the click-to-open-rate which is the rate of number of clicks in the email out of the total emails opened.To calculate CTOR we use,
Email clicks/email open *100.
Open rate is the rate of number of number of email open with respect to the number of emails sent.To calculate open rate we use,
Email open/email sent *100.

Tech-Stack Used:

To do this project I used the mysql workbench 8.0 CE as per requirements which was already downloaded for the previous projects .. It is a visual tool for databases , developers etc. using mysql workbench adds a lot of advantages due its features. It provides user friendly graphical interfaces, it allows designing databases , it helps to plan and organise your database structures including the tables, columns, rows etc, it also allows to directly write the query and executes it simultaneously displaying the indication of errors whenever there is one and many more.another main advantage is that it is an open source software so the licensing cost is not required.

Insights:

I gained several insights from doing this project. This project includes 2 case studies which are different from each other. Case study 1 consists of job_data from which I learned to calculate the rolling average and percentage share. Case study 2 consists of data to investigate the metric spike. Here, I was introduced to many new terms like retention rate, cohort, engagement metrics etc. From doing email engagement metrics , I learned open rate and CTOR.I learned a lot from this project. Most importantly I learned to deal with large datasets.

Result:

Doing this project made me realise that I have a lot more to learn even though I have grasped the basics. Each project will help me to learn more and more.I learned to implement mathematical calculations in the query. I also learned to import large datasets with ease.
I learned different concepts . Learning by doing is really beneficial to understand the basic concepts to the core. This project really helps me to increase my familiarity with mysql .