OPERATION ANALYTICS AND INVESTING METRIC SPIKE

(ADVANCED SQL)

PROJECT DESCRIPTION

In this project we'll do work as a Lead Data Analyst. We'll do Analyze the data and provide valuable insights that can improve the company's operations and understand changes the key metrics.

In this project, The Operational Analytics is a crucial process that involves analyzing a company's end – to – end operations. This analysis helps to identify areas improvement within the company.

As a Data Analyst, we'll work closely with various teams such as operations team, supports team, and marketing to helping them derive the insights from the data.

This Operational Analytics involves with understanding and explaining the sudden changes in key metrics and

also involve with in daily user engagement or a drop in sales.

As a Data Analyst, we'll analyze the data and extract the valuable output from dataset.

APPROACH

Firstly, I'll take approach to upload the dataset in MySQL then create and insert the values in the database by using the DDL (Data Definition Language) , DML (Data Manipulation Language) by using MySQL workbench.

TECH – STACK USED

The MySQL Workbench Community and MySQL and MySQL community server and MySQLWorkbench release for versions 8.0 through 8.0.34.

INSIGHTS

CASE STUDY 1 : Job Data Analysis

In this case study we'll working with the table job_data with the following columns.

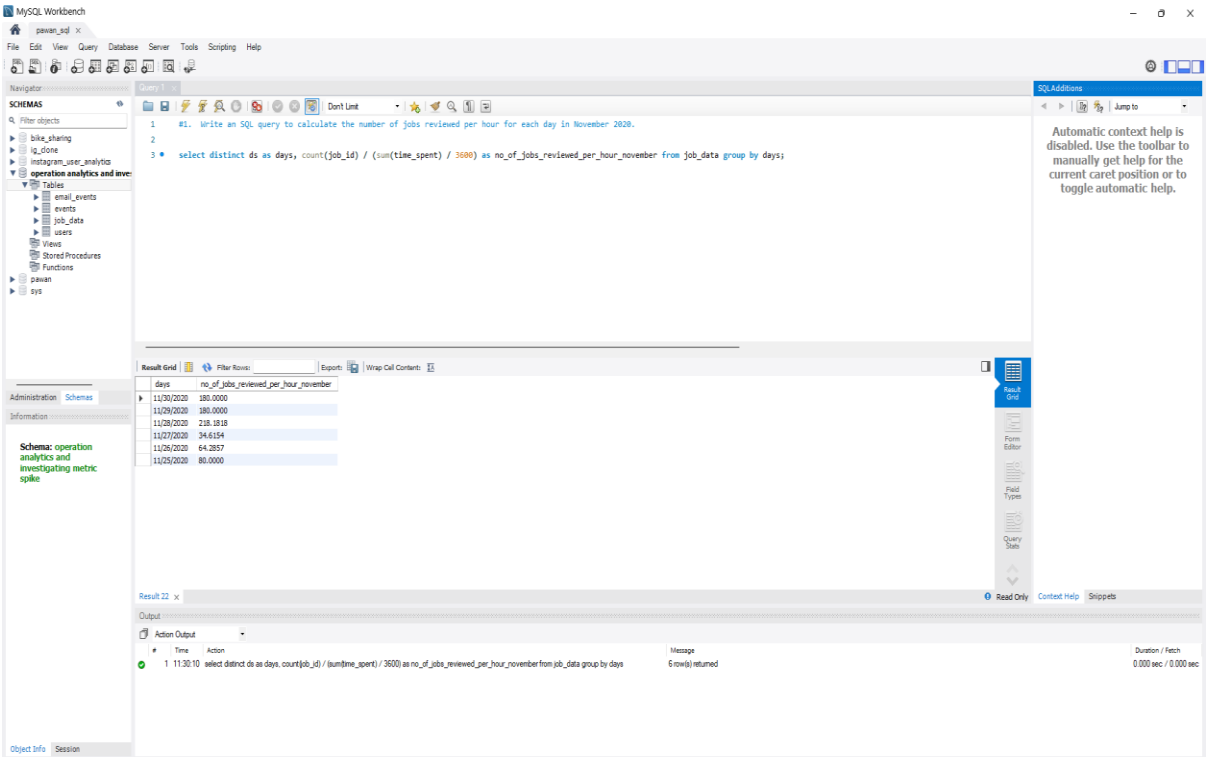
- job_id: Unique identifier of jobs
- actor_id: Unique identifier of actor
- event: The type of event (decision/skip/transfer).
- language: The Language of the content
- time_spent: Time spent to review the job in seconds.
- org: The Organization of the actor
- ds: The date in the format yyyy/mm/dd (stored as text).

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

In this question, we have to calculate the numbers of jobs reviewed per hour for each day in November 2020
Firstly, we select the ds column because we have to extract the output with the help of this column consider ds as a days after this we are using the ratio to find the numbers of jobs

reviewed per hour from November from job_data.

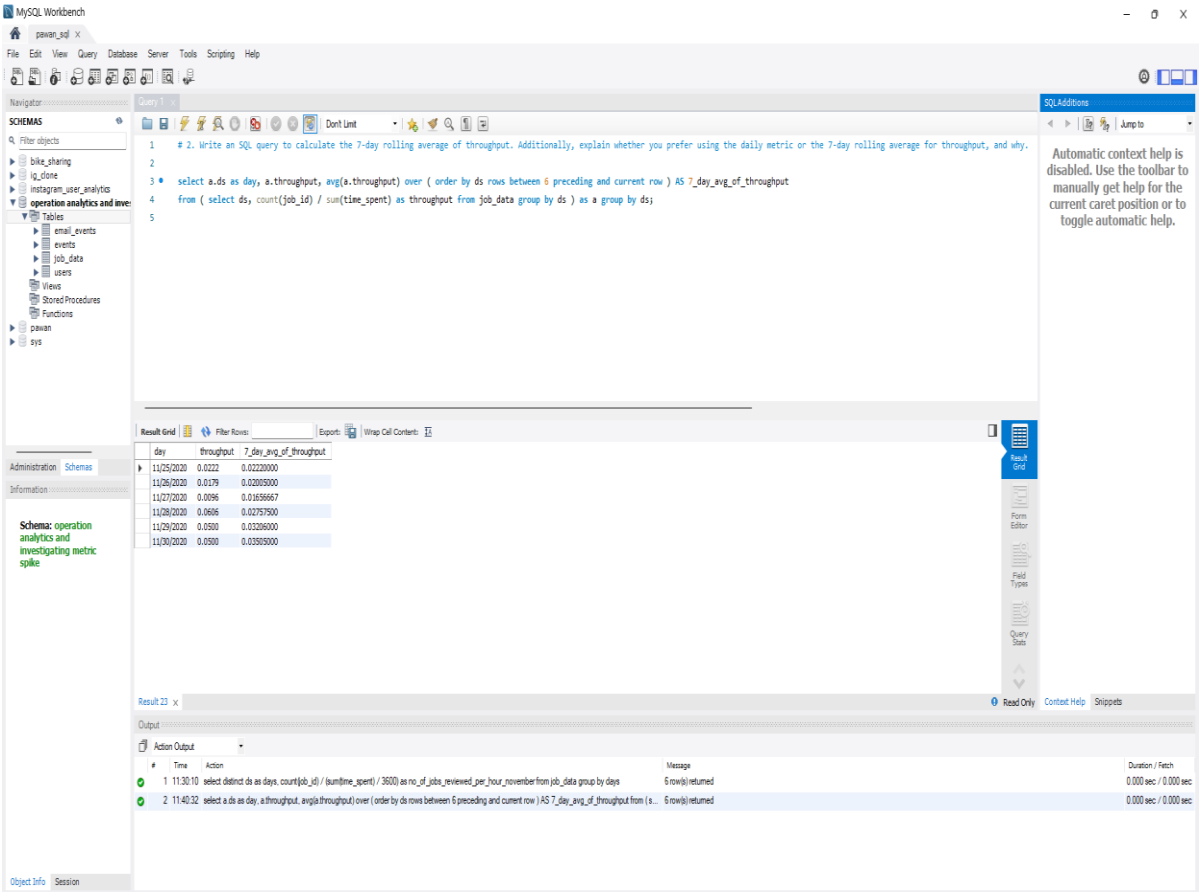
Here's our output :



B. 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.

In this question. We have to calculate the 7-day rolling average of throughput.
So, Firstly we select ds column as a day and **throughput** with alias **a**. In this query we are taking average of **ds** row and **a.throughput** and we're giving name as a **7_day_avg_of_throughput** and we're taking average from the **select ds , count (job_id)/sum(time spent) as throughput from job_data** and we're doing **group by ds**.

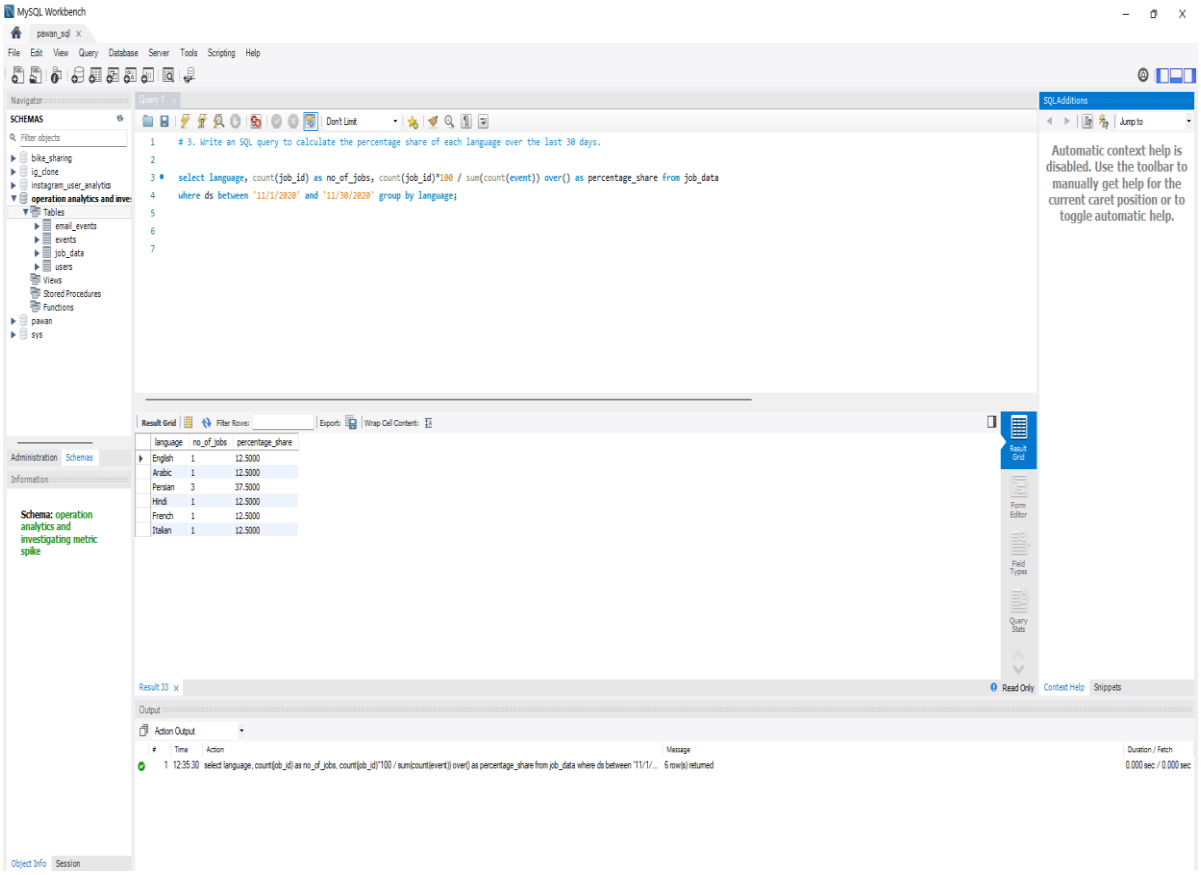
Here's our output :



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

In this question, we have to calculate the percentage share of each language over the last 30 days.
So. Firstly, we select the language and we're taking ratio of count(job_id) as no_of_jobs, count(job_id)*100 / sum(count(event)) OVER() as percentage_share. After this giving some constraints of dates to get required output and using the group by languages.

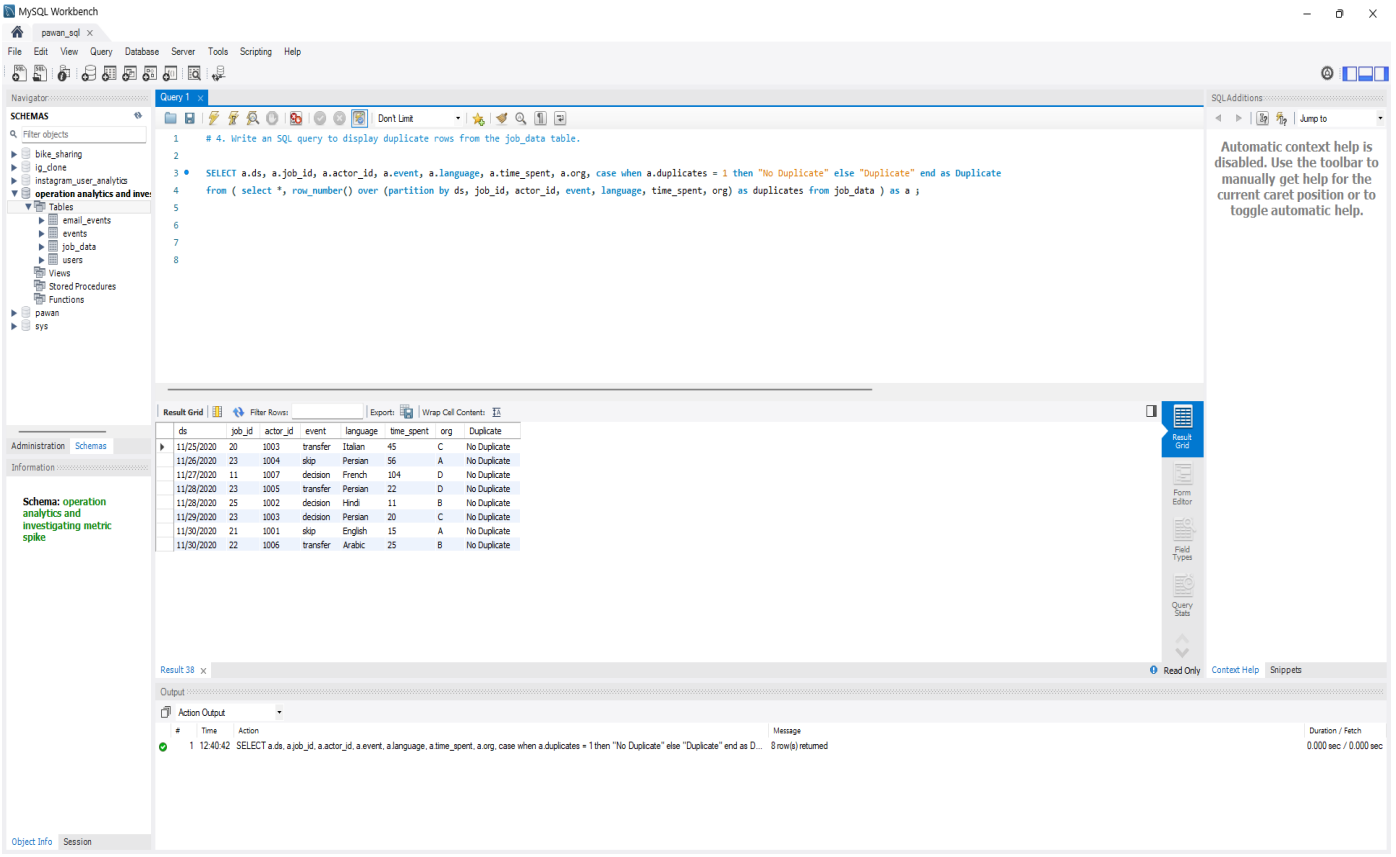
Here's our output :



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

In this question we have to display the duplicate rows from the job_data table.

Firstly, we select all the rows by partitioning by various columns and then we add a duplicate column in the main select query defining whether there are duplicates for that row in the dataset.

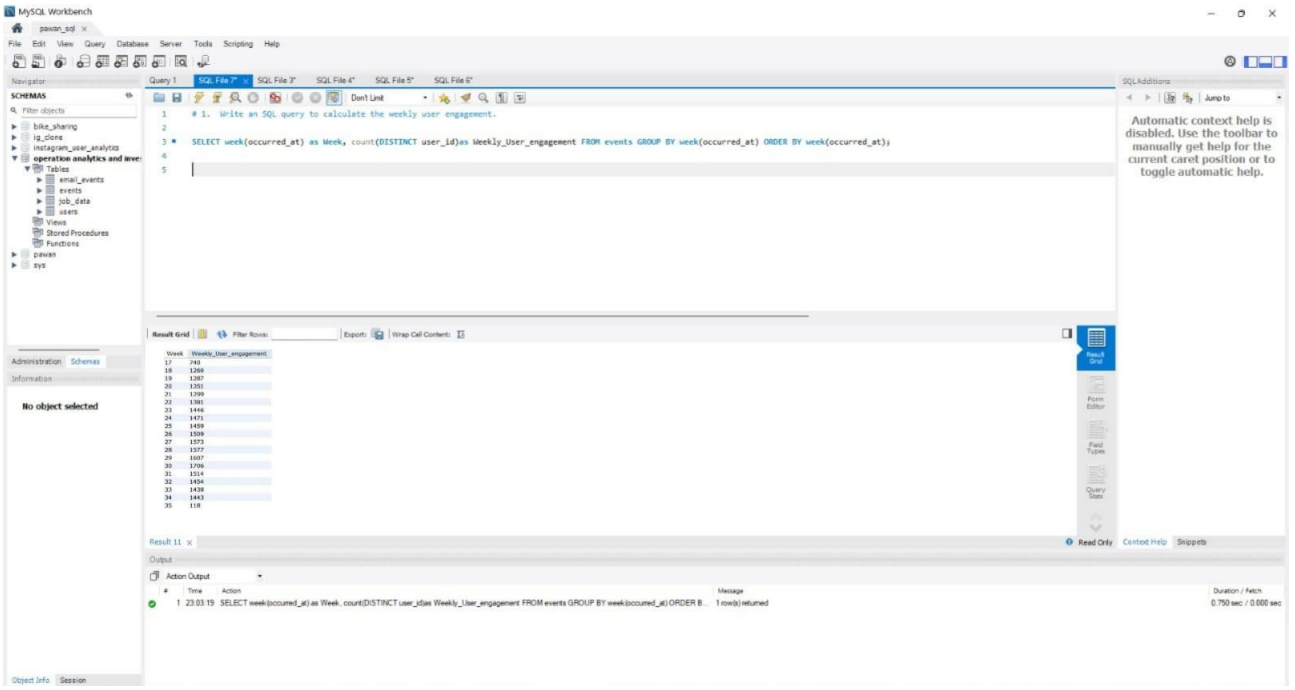


CASE STUDY 2 : Investing Metric Spike

In this case study we're using three tables which are users, events and email_events.

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

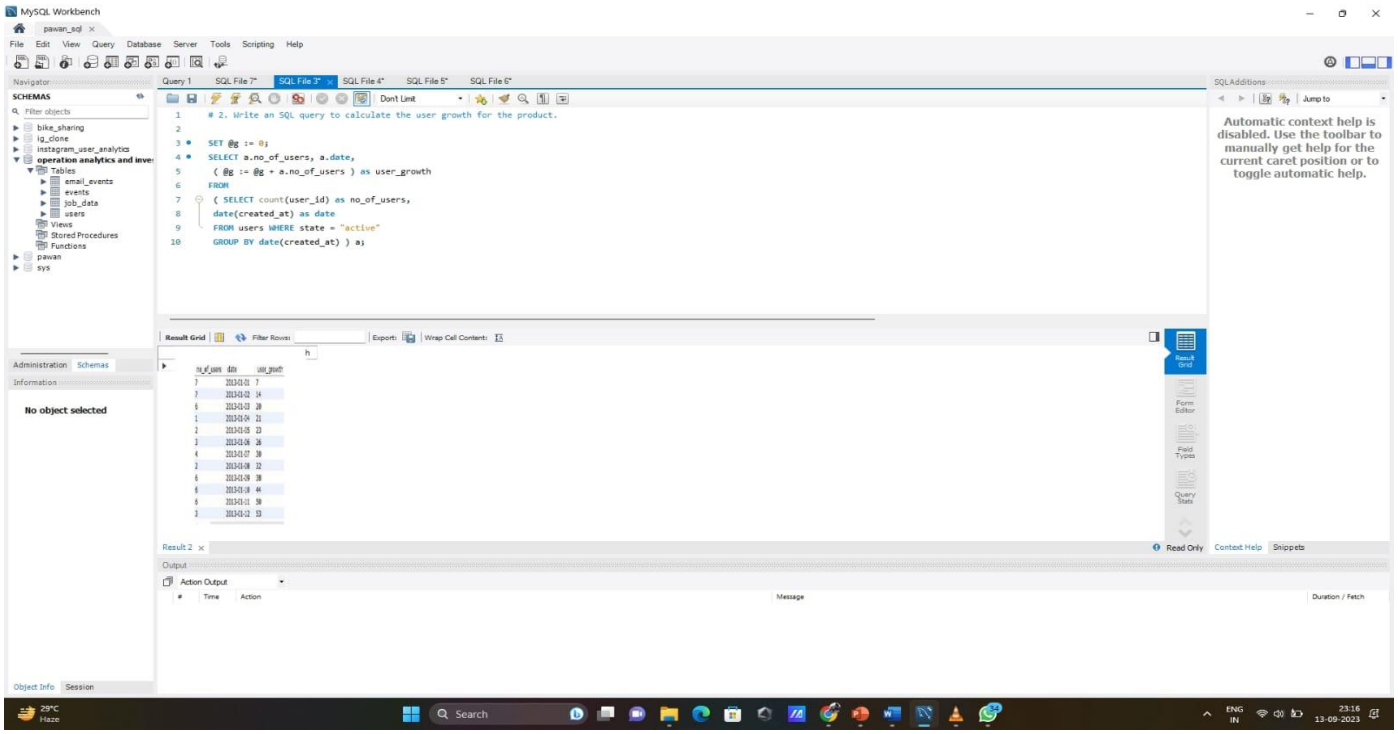
In this question we have to calculate the weekly user engagement from the given data.
In this query, we firstly extract week from the datetime column we have and also we take count of distinct user_id for the output. Finally we group by the extracted week column.



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

In this question, we have to calculate the user growth for the product.

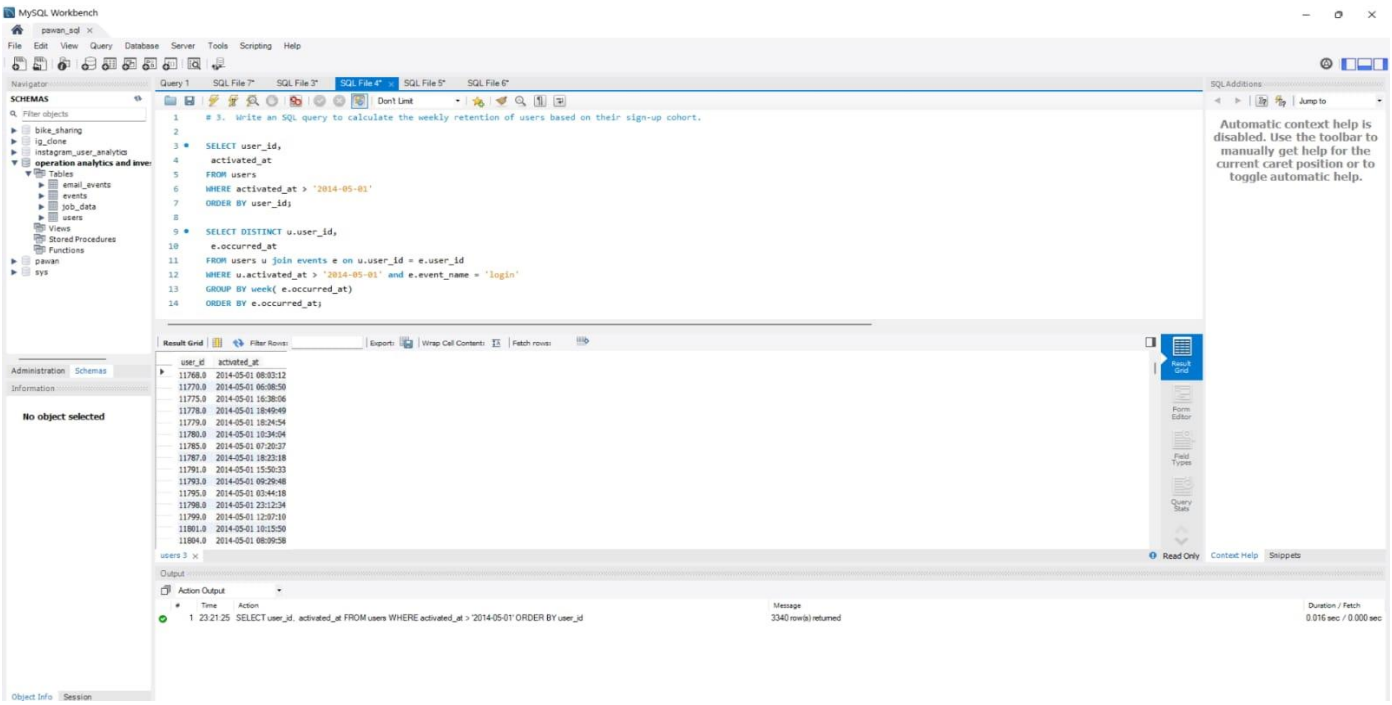
We would use an incremental approach wherein we set a variable to 0 at first and then keep on incrementing that value according to the condition we define according to date and an active status.



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

In this question, we have to calculate the weekly rentention of users based on their sign-up cohort

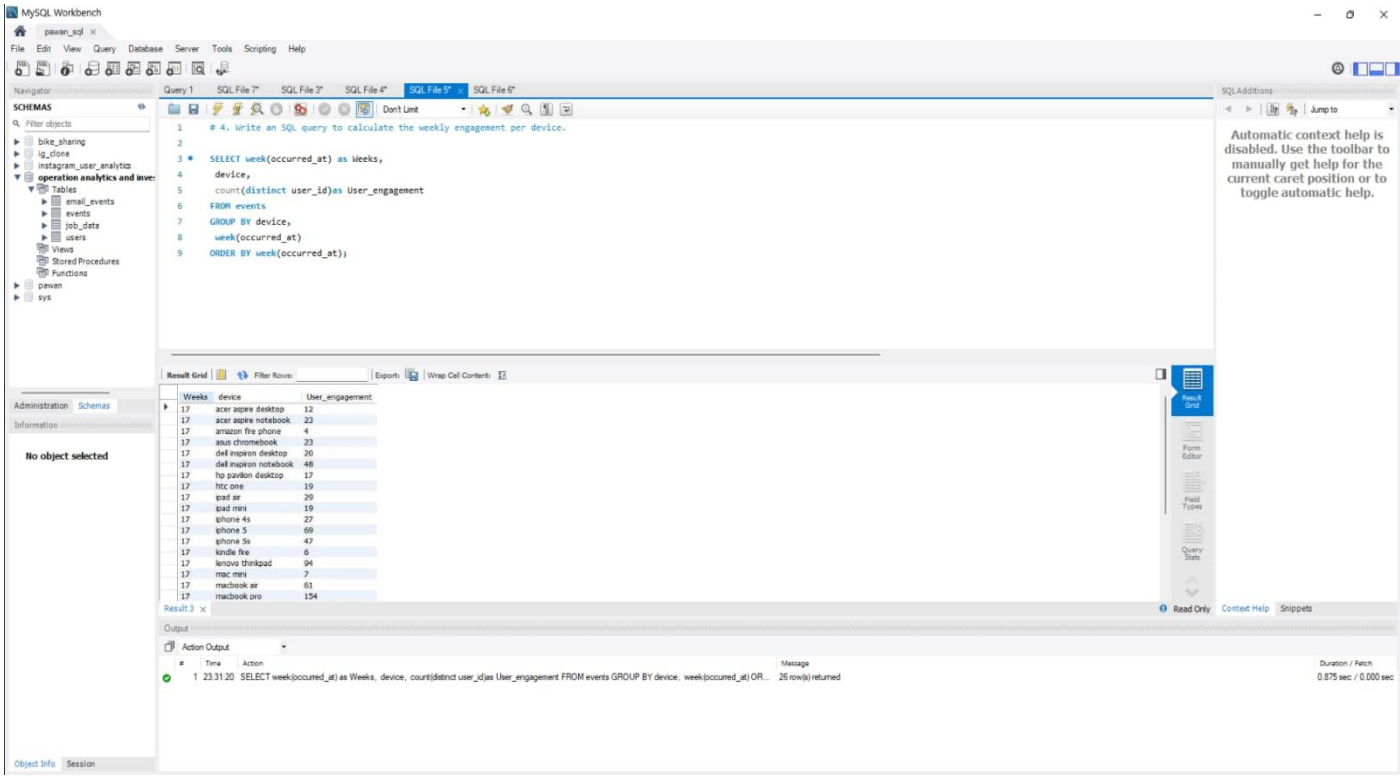
Firstly, We observe from the dataset that the events start from May 2014 onwards so we select all those records where activated date is after the date we mentioned. Then the user activity was extracted from events column for user who signed up after May 2014.



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

In this question, we have to calculate the weekly engagement per device.

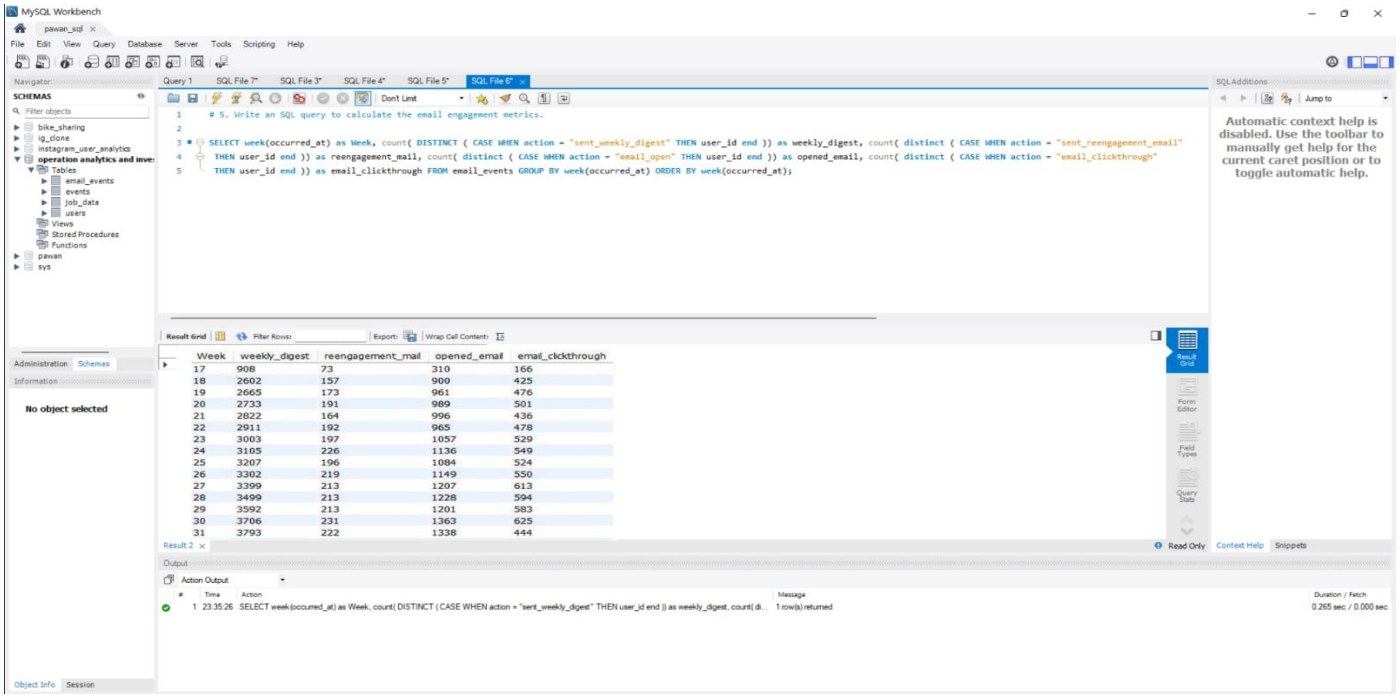
Firstly, for this problem we extract week from the datetime column along with count of distinct user id and we group by these records to get the desired output.



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

In this question, we have to calculate the email engagement metrics.

For this we select count of distinct user ids which contain different email action using case statements according to the week and finally we group by week to get the desired output.



CONCLUSION

- From this project, I learnt how to deal with advanced SQL queries and how to solve and how to apply the functions to extract the valuable and required outcome from the dataset.

- I also learnt from project how the Lead Data Analyst work in this kind of projects and how they deal with it.
- Also, they are dealing and working with various teams such as operationals, support team , marketing team and the Data Analyst helping and giving them valuable insights from the data.
- I gained my advanced SQL skills to analyze the data and improve the company’s operations and their key metrics