##### Project: Operation Analytics and Investigating Metric Spike

**Project Description**

This Project involves analyzing a company's end-to-end operations by working closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

As a data analyst, first I will understand the project requirements and purpose and then I will understand the data which is provided in the project attachment and will perform analysis to get the meaningful insights.

Analysis: involves understanding and explaining sudden changes in key metrics.

For example, in case study 1 results by calculating jobs reviewed per hour for each day shows inconsistency in jobs reviewed per day. With this information operations team can look in to challenges and plans for day-to-day activities to bring the consistency. 7 day rolling averages analysis provides overall performance over time and long-term trends and patterns that can be used for strategic business decisions.

In case study 2 calculation of weekly user engagement showed that there is a raise or increase in user engagement in initial weeks and a steady decline in user engagement after that

**Approach**

1. **Understanding the data**: The first step was to understand the questions given to me in both the case studies and understand structure of the data and created the database “operation\_analytics”. The attachment provided the csv files for 4 tables, one for case study 1 and three for case study 2 including job\_data, users, events and email\_events. I analyzed the data types and values in each column to understand what kind of information each table contains.
2. **Formulating SQL Queries**: Based on your questions, I wrote SQL queries to calculate the metrics you were interested in. This involved using various SQL functions and clauses such as COUNT(), GROUP BY, ORDER BY, DATE\_FORMAT(), WEEK(), INTERVAL, and CASE WHEN. I also used JOIN operations to combine data from different tables when necessary.
3. **Analyzing Data**: Once I had the SQL queries, I executed the queries to retrieve the data from CSV files. For this I used CREATE, LOAD, ALTER and UPDATE queries and added additional functions to my SQL queries like CONCAT(), DATE\_ADD() to get the required data.
4. **Interpreting Results**: After retrieving the data, I interpreted the results to answer the questions of both case studies. For example, I calculated weekly user engagement by counting the number of events per user per week, and I calculated email engagement metrics by counting the number of each action per user.
5. **Providing Insights**: Based on the results of the data analysis, I provided insights that could be useful for decision-making. Throughout this process, my goal was to provide clear, accurate, and helpful insights and answers to questions posed by management team.

**Tech-Stack Used**

The software and versions you used for the project:

MySQL Workbench 8.0 CE – To write SQL queries to analyze Instagram user data and answer questions posed by the management team

MYSQL Server 8.0 – To create and store database

Pandas – to create graph visualisation

**Insights:**

I will provide my insights in the form of answers for the questions posed by management team

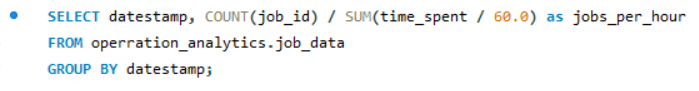
CASE STUDY 1: Job Data Analysis

1. **Jobs Reviewed Over Time:**

* Objective: Calculate the number of jobs reviewed per hour for each day in November 2020.

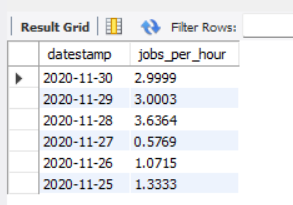
SQL Query:

To calculate the number of jobs reviewed per hour for each day in November 2020, you would first need to calculate the total time spent on jobs each day. Then, divide the number of jobs by the total time spent in hours.

****

Output:

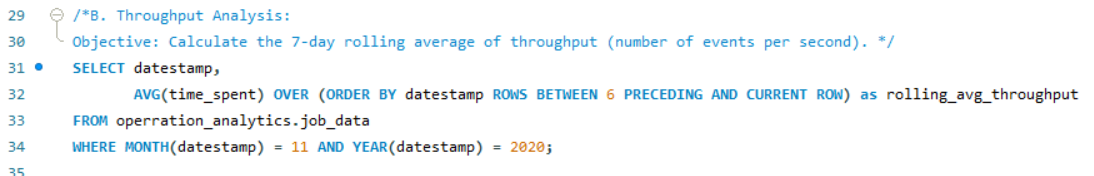
Below is the result for above query which shows number of jobs reviewed per hour for each day.

****

2. **Throughput Analysis:**

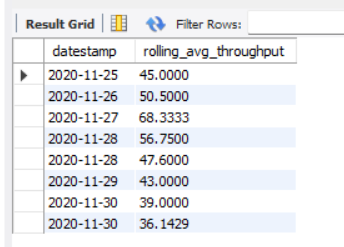
* Objective: Calculate the 7-day rolling average of throughput (number of events per second).

SQL Query:

****

**Output:**

Above query give the average time\_spent for each day and the six preceding days in November 2020. It assumes that time\_spent is a measure of throughput.

****

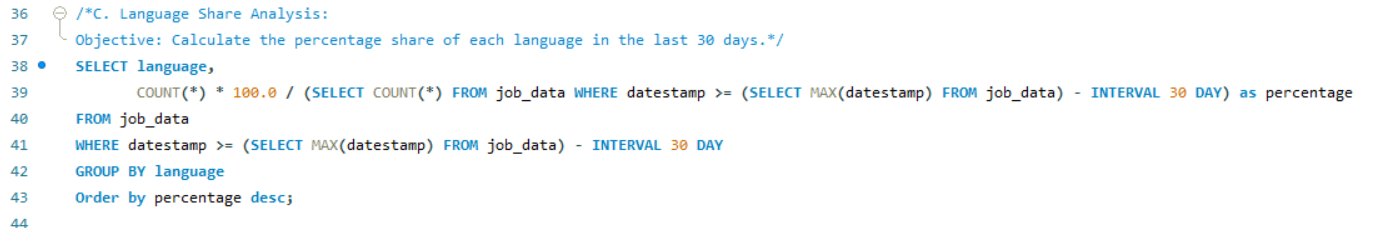
calculating jobs reviewed per hour for each day shows inconsistency in jobs reviewed per day. With this information operations team can look in to challenges and plans for day-to-day activities to bring the consistency. 7 day rolling averages analysis provides overall performance over time and long-term trends and patterns that can be used for strategic business decisions.

So, the choice between these two would depend on whether you’re more interested in short-term operational insights or long-term strategic trends.

**3.** Language Share Analysis:

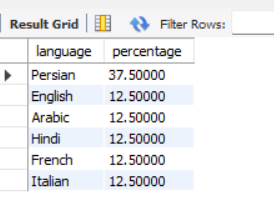
* Objective: Calculate the percentage share of each language in the last 30 days.

**SQL Query:**

****

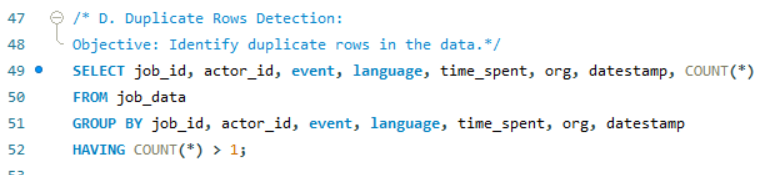
**Output:**Above query shows percentage share of each language in the last 30 days.

Persian language has more share percentage

****

**4. Duplicate Rows Detection:**

* Objective: Identify duplicate rows in the data.

**SQL Query: **

**Output:**

above query gave no result as there is not duplicate rows in the default data provided in the project attachment.

We can create dummy data by adding additional rows to the table by changing some information like year in date column or by copying existing rows then we will get duplicate rows in the results.

CASE STUDY 2: Investigating Metric Spike

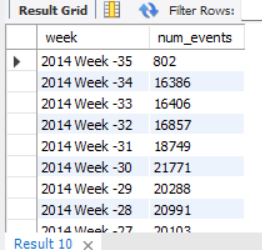
**1. Weekly User Engagement:**

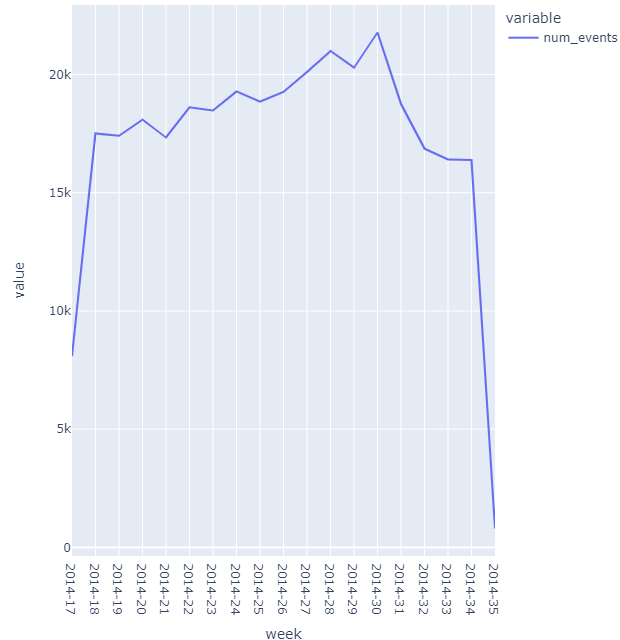
* Objective: Measure the activeness of users on a weekly basis

**SQL Query:**

****

**Output:** As per below result we can see user engagement was steadily increasing but later from last four weeks there is rapid decline in user activity

****

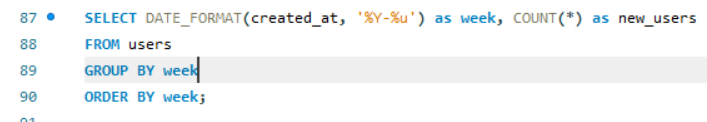
****

**2. User Growth Analysis:**

* Objective: Analyze the growth of users over time for a product.

**SQL Query:**

First calculate the number users registered in each week

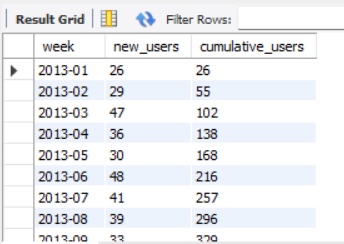


Nextcalculatethe cumulative number of users up to each week to find the growth of users over time

****

**Output:**

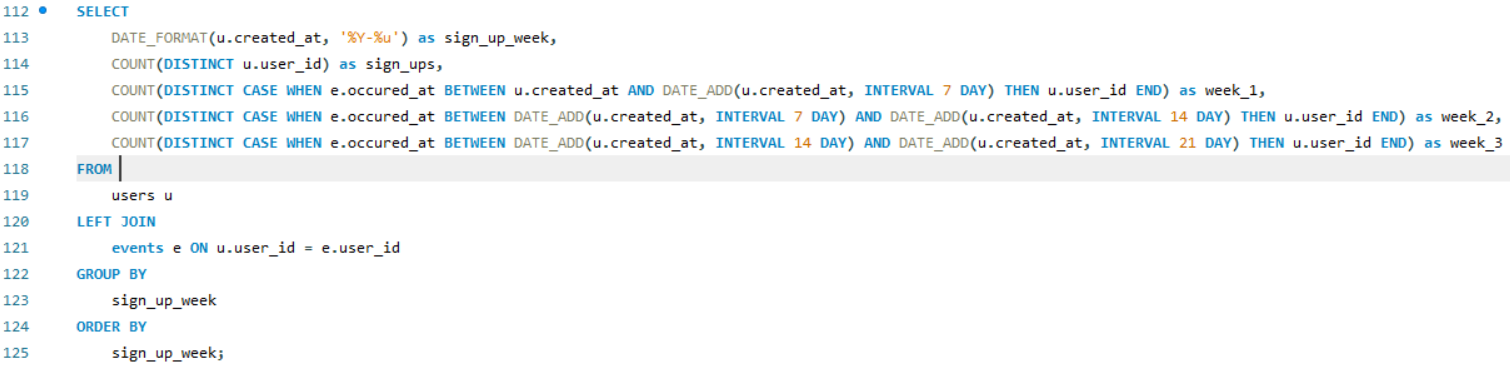
This second query will list the number of new users for each week and also growth of users for each week by adding previous weeks users to the current week



**3. Weekly Retention Analysis:**

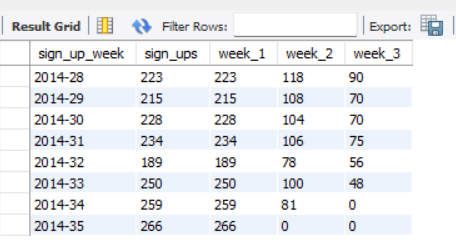
* Objective: Analyze the retention of users on a weekly basis after signing up for a product.

**SQL Query:**

****

**Output:**

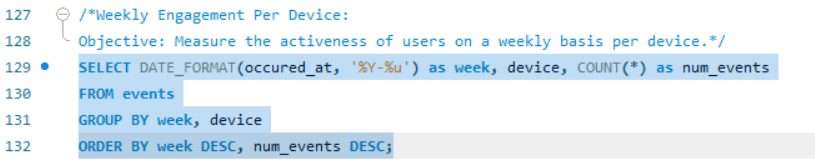
to calculate the weekly retention of users based on their sign-up cohort. Above SQL query calculates the number of users who signed up each week and how many of them had an event in the following weeks

****

**4. Weekly Engagement Per Device:**

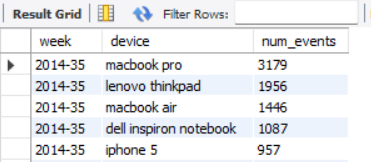
* Objective: Measure the activeness of users on a weekly basis per device.

**SQL Query:**

****

**Output:**

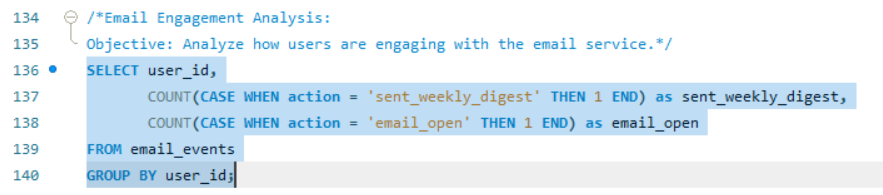
Above query counts the number of events per device per week to find weekly engagement per device

****

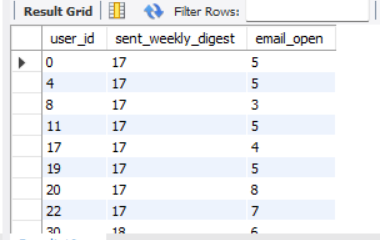
**5. Email Engagement Analysis:**

* Objective: Analyze how users are engaging with the email service.

**SQL Query:**

****

**Output:**

****

**Result**

This project definitely helps me rebrushing my SQL skills and I regained my confidence on using SQL for data analytics. I was relying mostly on Excel and Power BI for data analysis but with this project I feel SQL is easier to use.

In this project based on the above analysis we can find how is the user engagement and new users growth over time, and which language and device most of the users are preferring to interact with the applications. Based on this company can take decisions on the changes or modifications to bring to make the application more user friendly.

Drive Link

Case Study 1

<https://drive.google.com/file/d/1zoSiB2sknV0xB960MIXwkt725wAzkYIO/view?usp=sharing>

Case Study 2

<https://drive.google.com/file/d/1X_2v3Wir3znmo66e0Z8KFL8YwRkMugoj/view?usp=sharing>