Nilesh Nayak

Pune

**Operation Analytics and Investigating Metric Spike**

**Project 3**

**Project Description:**

**Operation Analytics**

Operations analytics is the procedure of gathering, examining, and deriving meaning from data in order to acquire insights into many operational facets of a company or organization. To enhance processes, increase efficiency, and make wise judgments, this method entails using data from several sources. It frequently entails the application of numerous analytical approaches, including statistical analysis, data mining, predictive modelling, and machine learning, to find patterns, trends, and opportunities for process improvement. Operational analytics is a subset of business analytics that focuses on immediate action.

Typically, data for an operations analytics project would be gathered from sources such as production systems, supply chains, customer interactions, and others. You can find bottlenecks, inefficiencies, and locations where resources can be better deployed by examining this data. This can result in better decision-making, lower costs, higher customer happiness, and overall corporate success.

**Metric Spike Investigation**

Metric spike investigation is the act of studying and comprehending unexpected and significant increases (or spikes) in specific metrics or key performance indicators (KPIs). These surges can occur in a variety of areas, including website traffic, sales, client queries, and any other observable statistic.

The investigation involves several steps:

* **Detection**: Identifying the metric that has experienced a sudden spike.
* **Isolation**: Determining the time and context in which the spike occurred, as well as the affected segments or areas.
* **Analysis**: Investigating the potential causes of the spike, which could include marketing campaigns, external events, technical issues, or changes in user behaviour.
* **Validation**: Confirming the accuracy of the data and the legitimacy of the spike. Sometimes, data anomalies or errors can lead to false spikes.
* **Action**: Taking appropriate actions based on the findings. This might involve optimizing resources to meet increased demand, fixing technical issues, or capitalizing on the opportunity presented by the spike.

The above project “Operations analytics and Investigation of Metric Spike” is having four datasets like job\_data, users, events, email\_events. Those data set is used to get answers of the questions for two case studies of job data analysis and metric spike.

First Case study is all about total job reviewed over time, language share analysis, duplicate row detection and throughput analysis.

Second Case study is about user growth analysis, weekly user engagement, user retention analysis, weekly engagement of device, email engagement.

**Approach:**

* Import Data: Data is given in CSV format and it is imported into SQL workbench using SQL queries;
* Understanding data: Database named ‘**project3**’ is given in tabular format having total 7 tables with table names given below:

|  |  |  |
| --- | --- | --- |
| Table Name | Number of Rows | Number of Columns |
| job\_data | 8 | 7 |
| users | 9381 | 6 |
| events | 325255 | 7 |
| email\_events | 90389 | 4 |

* Understanding of data-types and table schemas;
* Converting the date column to DATE type as it is read like string while importing the datasets;
* Recognition and understanding of Referential-Integrity-Constraint between tables;
* Formation of modular queries using SQL techniques;
* Merging of two or more queries to get actual answers for the given task;
* Capturing results/output.

**Tech-Stack Used:**

I have executed the query on MY SQL workbench installed on windows 10 operating system, more details are given below:

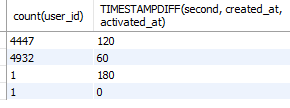
|  |  |
| --- | --- |
| Software Details | |
| Name: | Local instance MySQL80 |
| Host: | localhost |
| Port: | 3306 |
| Login User: | root |
| Current User: | root@localhost |
| SSL cipher: | SSL not used |
| Server | |
| Product: | MySQL Community Server - GPL |
| Version: | 8.0.34 |
| Connector | |
| Version: | C++ 8.1.0 |

I choose this MY SQL workbench because of the following reasons:

* It is available for my operating system;
* It takes less memory to install in the system;
* It allows access to data directly;
* Easy to understand and intuitive GUI;
* I can analyse multiple table at once;
* It provides cross-platform support;

**Insights:**

* There are 3 jobs (“skip”, “transfer”, “decision”) which is available in only 6 languages throughout 4 organisations.
* Users registered for the app is from 12 languages.
* Users activated within 180 seconds after the registration.



* All users are active. The state column of users table contains some Unicode character as it shows the length 7 instead of 6.
* There were total 18 unique events has been organised for engagement and signup\_flow of the users, 3 types of users participated in these events from 47 different locations and they used 26 types of devices to interact in the events.
* There were total 4 types of email-events (“sent\_weekly\_digest”, “email\_open”, “email\_clickthrough”, “sent\_reengagement\_email”) to track engagement of the users for different events.

**Result:**

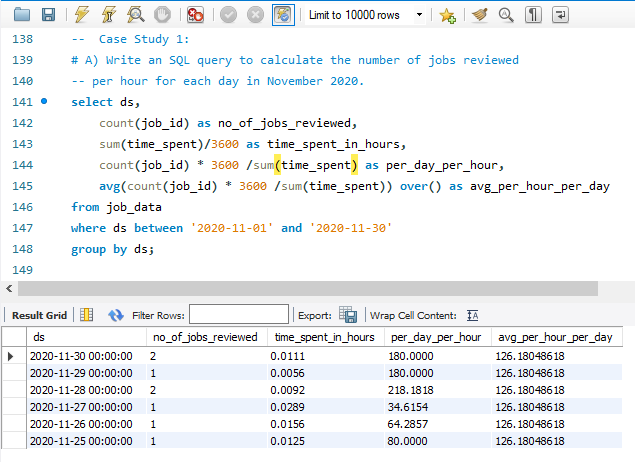
**Case Study 1: Job Data Analysis**

**Jobs Reviewed Over Time:** Calculate the number of jobs reviewed per hour for each day in November 2020.

**Task:**

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

**Query and Output:**

****

**Insights and Interpretations:**

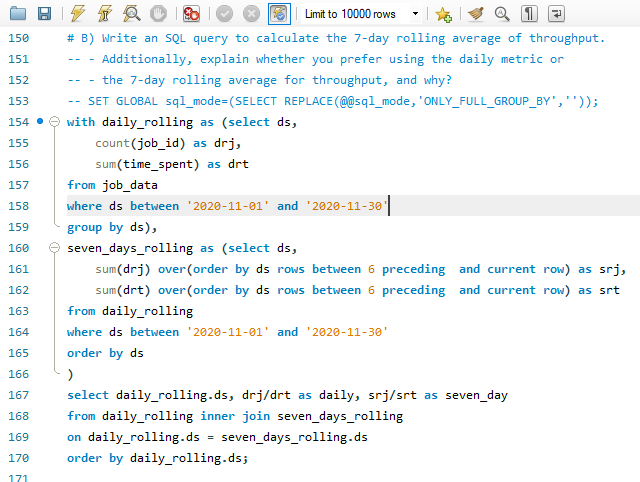
* Minimum Job reviewed per hour on 27th November and maximum job reviewed per hour on 30th and 29th November.
* Average number of job reviewed per hour per day is around 126.

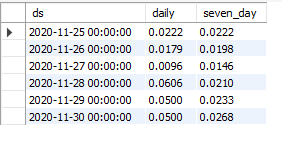
**Throughput Analysis:** Calculate the 7-day rolling average of throughput (number of events per second).

**Task:**

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.

**Query and Output:**

****

****

**Insights and Interpretations:**

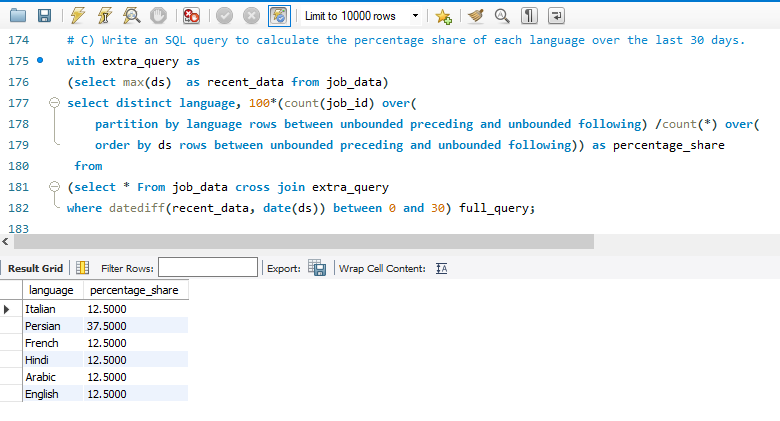
* It can be observed that the seven-day rolling average is better than daily rolling stats for initial days and daily rolling stats increases in later days.
* We need to focus on 7-day rolling average as it mitigates the offset time (more in daily stats).

**Language Share Analysis:** Calculate the percentage share of each language in the last 30 days.

**Task:**

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

**Query and Output:**

****

**Insights and Interpretations:**

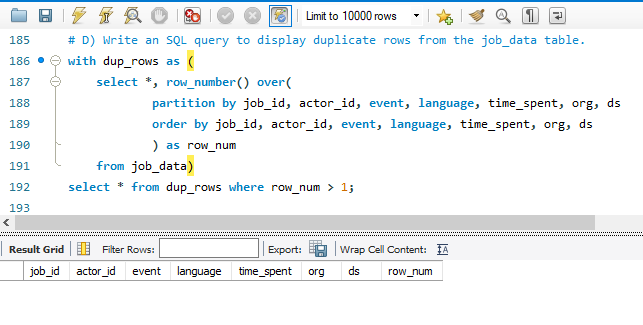
* It can be observed that the Persian language is having more share (37.5%) for job review and other five languages is having equal share.

**Duplicate Rows Detection:** Identify duplicate rows in the data.

**Task:**

Write an SQL query to display duplicate rows from the job\_data table.

**Query and Output:**

****

**Insights and Interpretations:**

* There are no any duplicate rows in the job\_data table.

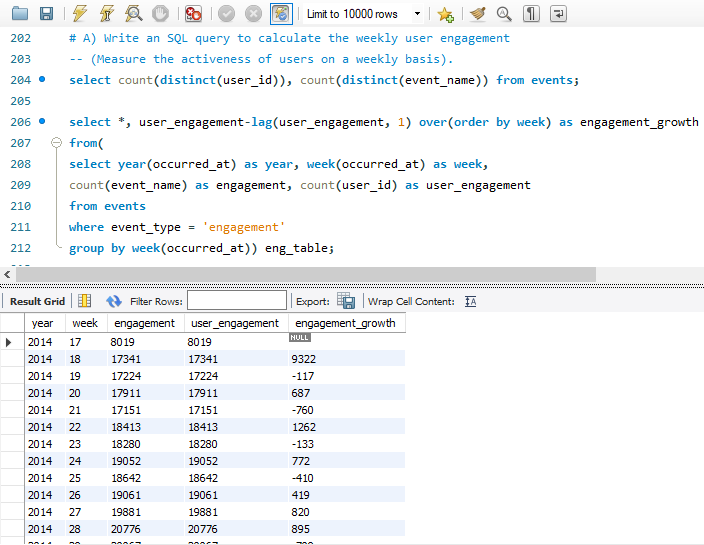
**Case Study 2: Investigating Metric Spike**

**Weekly User Engagement:** Measure the activeness of users on a weekly basis.

**Task:**

Write an SQL query to calculate the weekly user engagement.

**Query and Output:**

****

**Insights and Interpretations:**

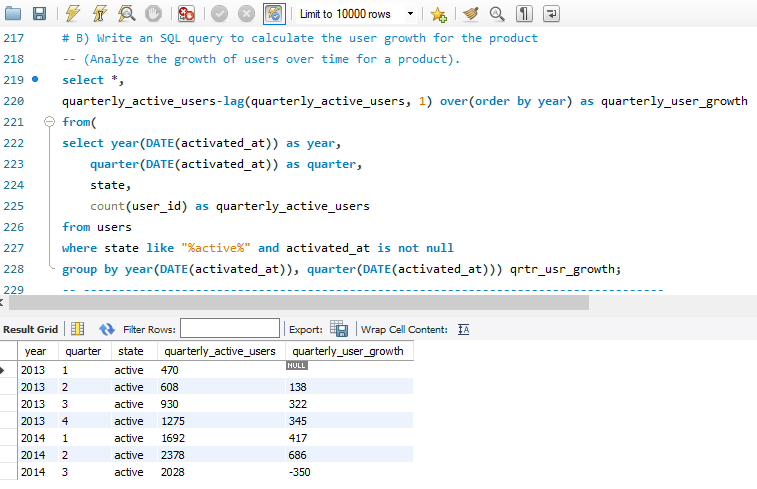
* The user engagement/activeness is fluctuating in initial weeks but later stabilizes.

**User Growth Analysis:** Analyze the growth of users over time for a product.

**Task:**

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

**Query and Output:**

****

**Insights and Interpretations:**

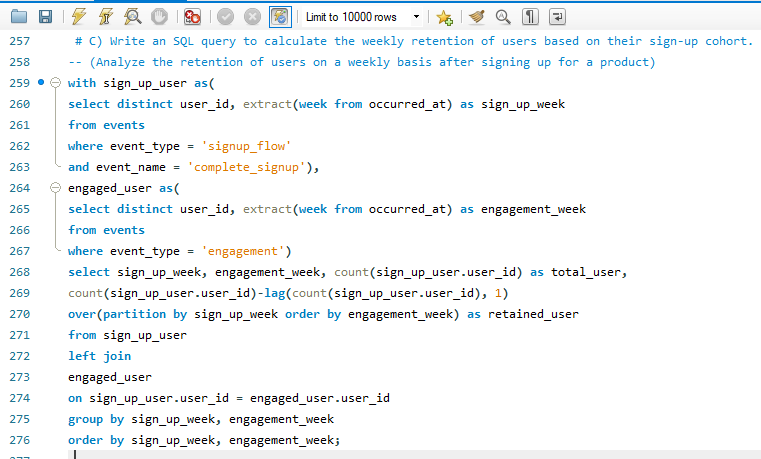
* In the year 2013 the user base is growing for each quarter but it reduced in 3rd quarter of 2014.
* Quarter 2 of 2014 records maximum growth of users.

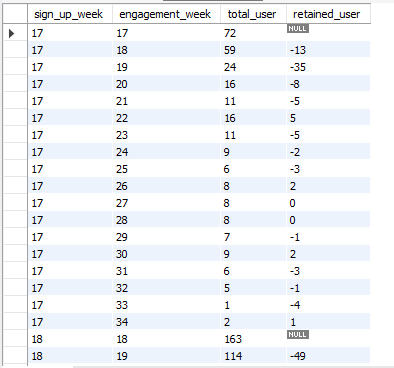
**Weekly Retention Analysis:** Analyze the retention of users on a weekly basis after signing up for a product.

**Task:**

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

**Query and Output:**

****

****

**Insights and Interpretations:**

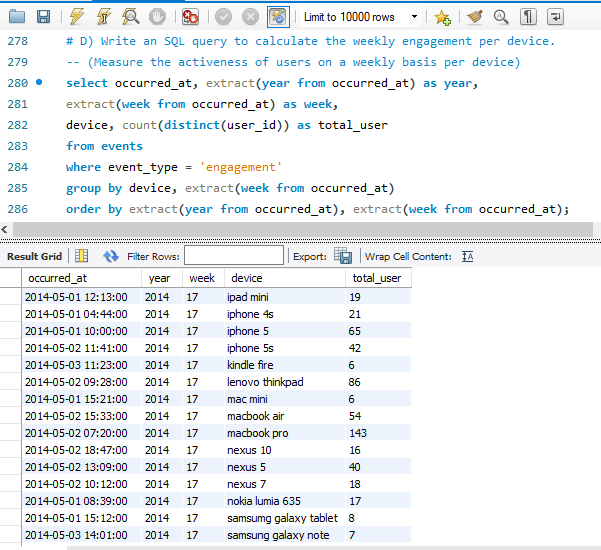
* Total 179 rows have been reflected and it clearly shows that there are negative trends in user retention for weekly cohort. Though very less number of times user retention is positive.

**Weekly Engagement Per Device:** Measure the activeness of users on a weekly basis per device.

**Task:**

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

**Query and Output:**

****

**Insights and Interpretations:**

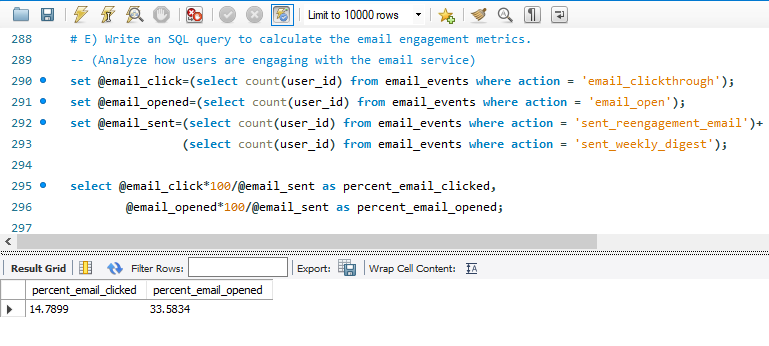
* Total 491 rows have been reflected and it is observed that each week the macbook pro user was maximum to interact in the events. It seems macbook pro is very popular among users.

**Email Engagement Analysis:** Analyze how users are engaging with the email service.

**Task:**

Write an SQL query to calculate the email engagement metrics.

**Query and Output:**

****

**Insights and Interpretations:**

* Total 60920 email was sent out of which 9010 (14.7899%) users clicked the sent email and 20459 (33.5834%) users opened the email.
* It is good strategy to send email to users for promotional activities as 33.6% users clicks and reads the email-content.

**Conclusion:**

The "Two case studies" of the project is carried out utilizing SQL queries on MYSQL workbench. This initiative has tracked users' insights and examined their engagements with the app.

By accomplishing this assignment, I learned about SQL advanced queries and query optimization. I thoroughly comprehended the data and learned how to link it using a join query. I investigated strategies for working with SQL's date type capability. I've learned about window functions and common table expressions(CTE). I understood about the cohort analysis to find the retention of the users thorough the time given or size based or segment based.

**Drive Link:**

https://drive.google.com/drive/folders/1vSQW5fLT2B8L5paGd2faZkN1NX9mRW\_W?usp=sharing