OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

TRAINITY PROJECT #3

Submission by:

ANANAYA LAL (Data Analyst Trainee)

PROJECT DESCRIPTION:

This analysis process involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. Data Analyst works closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

Our goal is to use advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.

This project involves conducting a comprehensive analysis of the end-to-end operations of a company based on two case studies. There were 2 case studies given to work upon. A brief description of both case studies and output required form both of them are given further.

Case Studies

Case Study 1: Job Data Analysis Table name - job_data 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).

Case Study 2: Investigating Metric Spike Table names -

- users: Contains one row per user, with descriptive information about that user's account.
- **events**: Contains one row per event, where an event is an action that a user has taken (e.g., login, messaging, search).
- email_events: Contains events specific to the sending of emails.

Tasks

Case Study 1

- Jobs Reviewed Over Time
- Throughput Analysis
- Language Share Analysis
- Duplicate Rows Detection

Case Study 2

- Weekly User Engagement
- User Growth Analysis
- Weekly Retention Analysis
- Weekly Engagement Per Device
- Email Engagement Analysis

Important Informations

Approach:

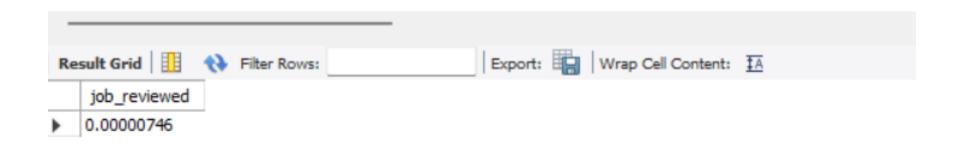
• To conduct a comprehensive analysis of a company's end-to-end operations based on case studies, we would typically follow a structured approach that involves several key steps like, define the objective, collect data, analyse data, identify areas of improvement etc.

Tech-Stack Used:

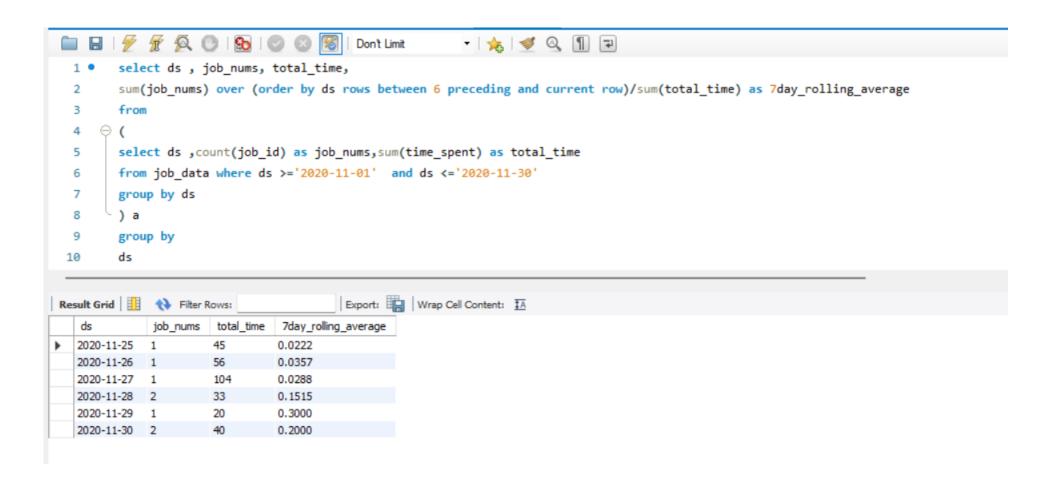
- MySQL
- PostgreSQL
- MS Excel
- MS Word
- Powerpoint Ppt

C1T1. Amount of jobs reviewed over time.

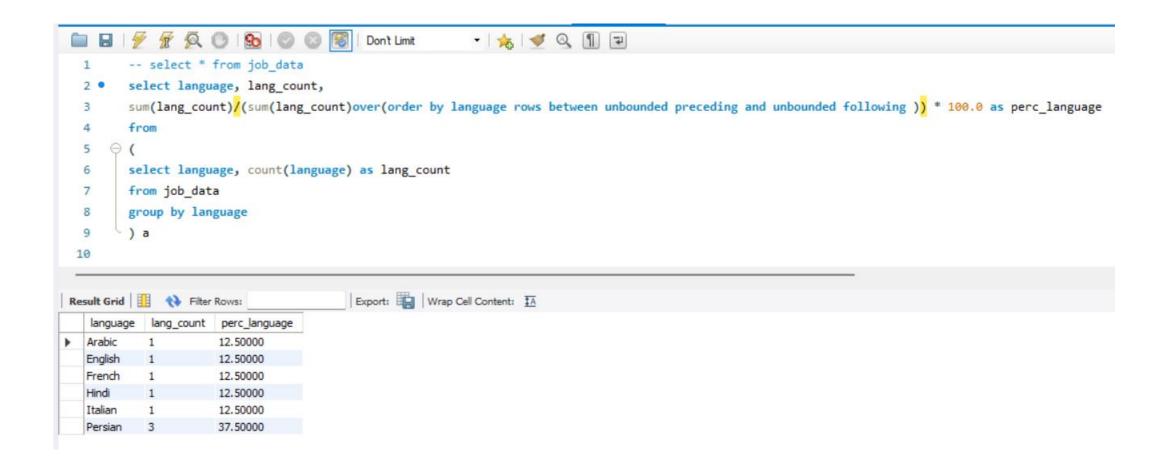
```
1 • use opr_n_ma;
2 • SELECT ((COUNT(job_id))/sum(time_spent)/3600) as job_reviewed
3 FROM job_data
```



C1T2 . No. of events happening per second – to calculate 7 day rolling average.

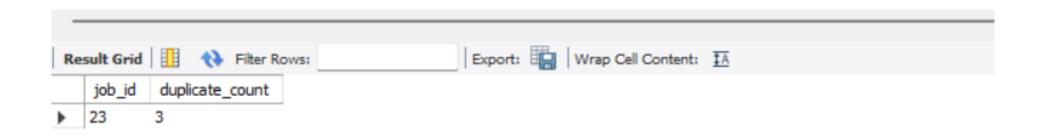


C1T3. Share of each language for different contents.

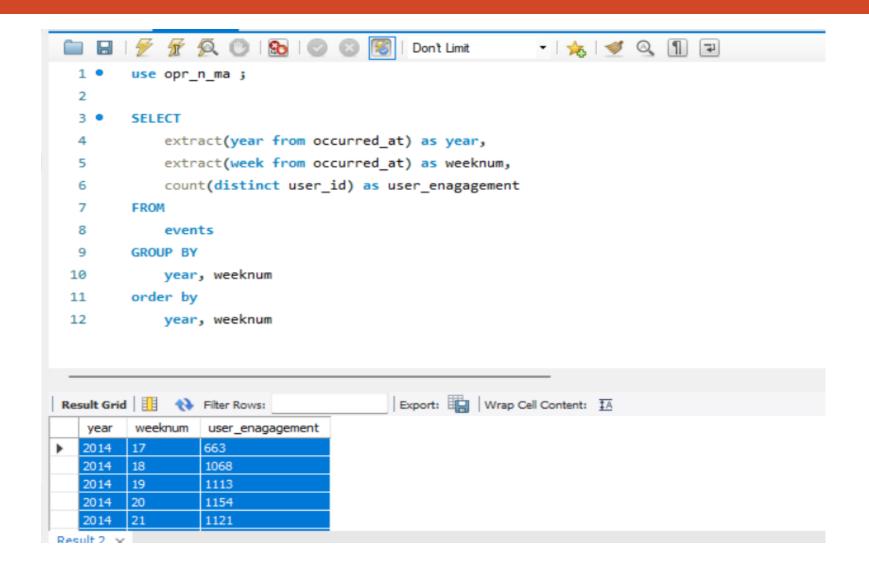


C1T4. Rows that have same values present in them.

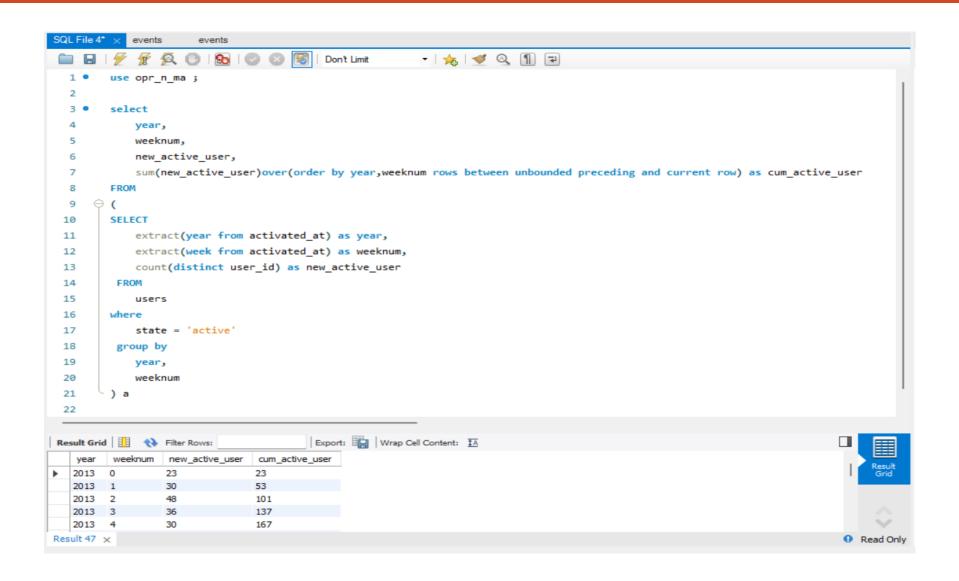
```
1 • select job_id, count(job_id) as duplicate_count
2  from job_data
3  group by job_id
4  having duplicate_count > 1
```



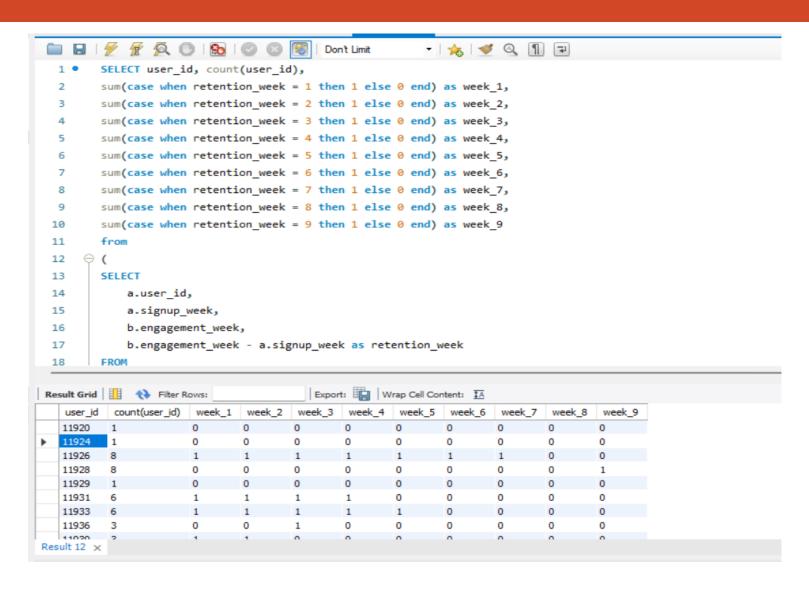
C2T1. To measure the activeness of a user.



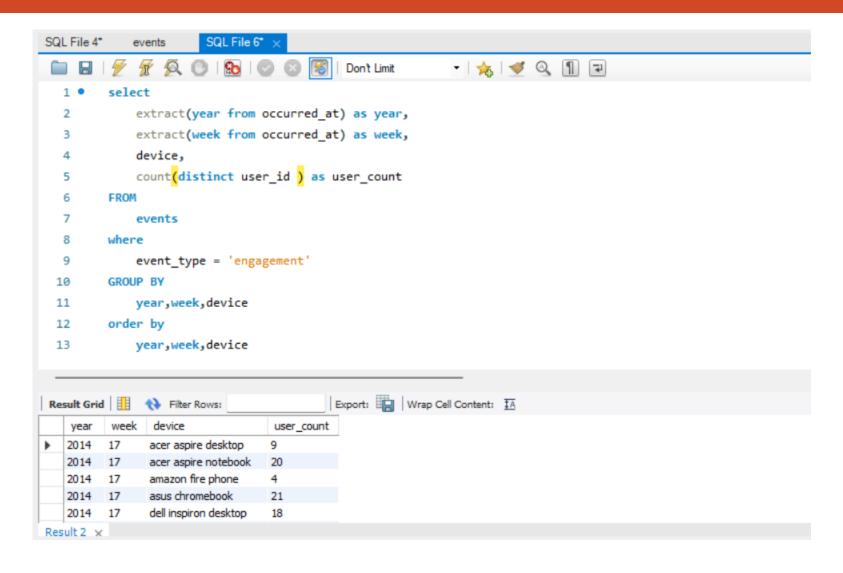
C2T2. Amount of users growing over time for a product – number of active users per week.



C2T3. Users getting retained weekly after signing up for a product.



C2T4. To measure the activeness of a user weekly - Calculate the weekly engagement per device.



C2T5. Users engaging with the email service.

```
SQL File 4*
                       SQL File 6*
                                                           - | 🏡 | 🥩 🔍 👖 📦
                                           Don't Limit
  1 •
         SELECT
         100.0*SUM(case when email cat = 'email open' then 1 else 0 END)/SUM(case when email cat = 'email sent' then 1 else 0 END) as email open rate,
         100.0*SUM(case when email cat = 'email clicked' then 1 else 0 END)/SUM(case when email cat = 'email sent' then 1 else 0 END) as email clicked rate
         FROM
         select *,
  8
         when action IN ('sent_weekly_digest','sent_reengagement_email')
             THEN 'email sent'
  9
         when action IN ('email open')
  10
             THEN 'email open'
  11
         when action IN ('email clickthrough')
  12
             THEN 'email clicked'
  13
        END as email cat
  14
        FROM email events
  15
        ) a
  16
  17
  18
                                         Export: Wrap Cell Content: IA
Result Grid Filter Rows:
    email_open_rate | email_clicked_rate
33.58339
                 14.78989
```

Insights

- 0.00000746 jobs were reviewed over time.
- Highest share of language is of Persian i.e. 37.5 % followed by Arabic, English, French, Hindi and Italian
- Job Id: 23 has 3 duplicate count.
- The highest user engagement was 1154 in week number 20 in 2014.
- The highest weekly engagement as per device has user count 21 by Asus Chromebook in week 17 in 2014.

RESULTS

Results obtained are displayed on each slide respectively.

The project focused on various aspects of advanced SQL and enhanced the skills on the same.