Operation Analytics and Investigation Metric Spike Using SQL

Case-1 : Job Data Analysis

Project Description

The primary objective of this project was to analyze job data to derive actionable insights and understand trends related to job reviews. The dataset includes information on job reviews performed by various actors, including details such as the event type, language of content, time spent reviewing, and more. The analysis aimed to address several key areas:

1. **Jobs Reviewed Over Time:** To determine the number of jobs reviewed per hour for each day in November 2020.
2. **Throughput Analysis:** To compute the 7-day rolling average of throughput (events per second).
3. **Language Share Analysis:** To calculate the percentage share of each language in the dataset over the last 30 days.
4. **Duplicate Rows Detection:** To identify and address any duplicate entries in the dataset.

Approach

1. **Data Download and Extraction**: Data downloaded from given link and extract from zip.
2. **Data Cleaning:** Checked for duplicates and inconsistencies in the data, ensuring that the dataset is accurate and reliable for analysis.
3. **Data import in Mysql :** After cleaning , imported data in mysql workbench from csv file. And changed datatype (varchar to date).
4. **Analysis Execution:**
    - **Jobs Reviewed Over Time:** Calculated the number of jobs reviewed per hour for each day in November 2020 by aggregating data and computing the ratio of jobs reviewed to the time spent.
    - **Throughput Analysis:** Computed the daily throughput and applied a 7-day rolling average to smooth out fluctuations and identify trends.
    - **Language Share Analysis:** Determined the percentage share of each language by aggregating language counts and calculating their proportions relative to the total.
    - **Duplicate Rows Detection:** Identified actor IDs with duplicate entries and reviewed these instances to address any data quality issues.

Tech-Stack Used

- **MySQL Workbench (Version 8.0):** Utilized for executing SQL queries, managing database schema, and performing data analysis. MySQL Workbench was essential for querying the job_data table and deriving insights from the data.

Insights

1. **Jobs Reviewed Over Time:**
    - Peak review activity was observed on November 28, 2020, with a notable increase in jobs reviewed per hour.
    - Lower review activity was recorded on November 27, 2020, suggesting potential operational or staffing issues on that day.
2. **Throughput Analysis:**
    - Daily throughput varied, with November 28 showing the highest rate of 0.06 events per second.
    - The 7-day rolling average indicated a stable trend with a slight upward movement towards the end of the period.
3. **Language Share Analysis:**
    - Persian was the most dominant language with a 37.50% share, while other languages (English, Arabic, Hindi, French, Italian) each had an equal share of 12.50%.
4. **Duplicate Rows Detection:**
    - Identified duplicate entries for actor ID 1003, indicating the need for further investigation and data cleaning to ensure accuracy.

Result

The analysis provided valuable insights into job review patterns, language distribution, and data quality. The key achievements include:

- Understanding daily and rolling throughput trends, which aids in resource allocation and operational planning.
- Identifying and addressing duplicate entries, improving data accuracy.
- Gaining insights into language distribution, allowing for better focus on languages with higher representation.

These findings contribute to a better understanding of job review dynamics and support informed decision-making for improving processes and resource management.

Case 2: Investigating Metric Spike

Project Description

The objective of this project is to analyse email events data to evaluate the performance of email campaigns. The focus is on understanding the effectiveness of emails by calculating the open and click rates. This analysis helps in assessing how engaging and successful the email communications are.

Approach

To achieve the analysis, the following approach was taken:

1. **Data Download and Extraction**: Data downloaded from given link and extract from zip.
2. **Data Cleaning:** Checked for duplicates and inconsistencies in the data, ensuring that the dataset is accurate and reliable for analysis.
3. **Data import in Mysql :** After cleaning , imported data in mysql workbench from csv file. And changed datatype (varchar to datetime) for all 3 tables (users, events and email_events).
4. **Query**: Wrote query for each task.

   Computed the open rate and click rate using SQL queries.

   - Open rate = (Number of email opens / Number of emails sent) * 100
   - Click rate = (Number of clicks / Number of emails sent) * 100

Tech-Stack Used

- **MySQL**: Used for querying the database and performing calculations. The version used was MySQL 8.0.
- **MySQL Workbench**: Utilized for writing and executing SQL queries, and visualizing the data.

Insights

- **Open Rate**: 33.58%
  - A significant portion of recipients engaged with the email by opening it, indicating effective subject lines or email timing.
- **Click Rate**: 14.79%
  - While the open rate is fairly good, the click rate suggests room for improvement in terms of content engagement or call-to-action effectiveness.

Result

The analysis provided valuable insights into the performance of email campaigns. The high open rate indicates good engagement, while the click rate suggests that while many recipients open the emails, fewer take further action. These insights can guide future improvements in email content and strategies to increase both open and click rates.