**Operation Analytics and Investigating Metric Spike**

**Description**

Operation analysis, or performing end-to-end operations for business growth and identifying opportunities for improvement, is the foundation of the project. The following duties will be located by me. [I] The hourly and daily review rates of positions for November 2020. [II] The throughput average for the past 7 days. [III]The percentage of each language's usage during the last 30 days. [IV] Displaying Duplicates. [V] Weekly user participation. [VI] The product's user growth. [VII] The cohort of newly registered users' weekly retention. [VIII] To gauge a user's level of activity. [IX] The metrics for email engagement, item from the database provided.

**Contents**

1. Project Description

2. Approach

3. Case Study 1

4. Case Study 2

**Project Description**

* This project is about conducting operation analytics for two separate cases: job data and investigating metric spikes. The job data case involves analyzing the data from a table called "job\_data" which contains information about the jobs reviewed by different actors.
* The second case involves investigating metric spikes by analyzing data from two tables: "users" and "events." The "users" table contains descriptive information about each user's account, while the "events" table contains information about the actions taken by each user.

**Approach**

The approach for this project is to first create a database and then the necessary tables using the structure and links provided. Then, SQL will be used to perform the analysis and answer the questions listed for each case. The answers will be presented in a report to the leadership team.

1. **Case Study 1: Job data**

* Number of jobs reviewed
* Throughput
* Percentage share of each language
* Duplicate rows

1. **Number of jobs reviewed:** Number of jobs reviewed over time.

|  |  |
| --- | --- |
| **date** | **jobs\_reviewed** |
| 2020-11-30 | 2 |
| 2020-11-29 | 1 |
| 2020-11-28 | 2 |
| 2020-11-27 | 1 |
| 2020-11-26 | 1 |
| 2020-11-25 | 1 |
| 2020-11-01 | 2 |
| 2020-11-02 | 1 |
| 2020-11-13 | 1 |
| 2020-11-24 | 1 |
| 2020-11-05 | 1 |
| 2020-11-06 | 1 |
| 2020-11-07 | 1 |
| 2020-11-08 | 2 |
| 2020-11-09 | 1 |
| 2020-11-10 | 1 |

1. **Throughput:**It is the no. of events happening per second.

|  |  |
| --- | --- |
| ds | avg\_throughput |
| 2020-11-01 00:00:00 | 2.0000 |
| 2020-11-02 00:00:00 | 1.5000 |
| 2020-11-05 00:00:00 | 1.3333 |
| 2020-11-06 00:00:00 | 1.2500 |
| 2020-11-07 00:00:00 | 1.2000 |
| 2020-11-08 00:00:00 | 1.3333 |
| 2020-11-09 00:00:00 | 1.2857 |
| 2020-11-10 00:00:00 | 1.1429 |
| 2020-11-13 00:00:00 | 1.1429 |
| 2020-11-24 00:00:00 | 1.1429 |
| 2020-11-25 00:00:00 | 1.1429 |
| 2020-11-26 00:00:00 | 1.1429 |
| 2020-11-27 00:00:00 | 1.0000 |
| 2020-11-28 00:00:00 | 1.1429 |
| 2020-11-29 00:00:00 | 1.1429 |
| 2020-11-30 00:00:00 | 1.2857 |

I prefer 7-day rolling average for throughput because it provides a better understanding of the average event rate over a week's time, rather than just for one day. This allows us to identify trends and patterns in the data that might not be immediately visible from daily metrics.

1. **Percentage share of each language**: Share of each language for different contents.

|  |  |  |
| --- | --- | --- |
| language | time\_spent | percentage |
| English | 15 | 2.16 |
| Arabic | 25 | 3.61 |
| Persian | 98 | 14.14 |
| Hindi | 36 | 5.19 |
| French | 104 | 15.01 |
| Italian | 45 | 6.49 |
| Spanish | 13 | 1.88 |
| German | 20 | 2.89 |
| Russian | 84 | 12.12 |
| Korean | 129 | 18.61 |
| Chinese | 14 | 2.02 |
| Dutch | 60 | 8.66 |
| Turkish | 27 | 3.90 |
| Swedish | 23 | 3.32 |

1. **Duplicate rows:** Rows that have the same value present in them.

|  |  |
| --- | --- |
| job\_id | duplicate |
| 23 | 3 |

1. **Case Study 2: Investigating metric spike**

* User Engagement
* User Growth
* Weekly Retention
* Weekly Engagement
* Email Engagement

1. **User Engagement:**To measure the activeness of a user. Measuring if the user finds quality in a product/service.



1. **User Growth:**Number of users growing over time for a product.

Data file has been attached below

[**user growth for product.csv**](user%20growth%20for%20product.csv)

1. **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly**.**

Data file has been attached below

[**engagement by device over time period.csv**](engagement%20by%20device%20over%20time%20period.csv)

1. **Email Engagement:** Users engaging with the email service.

Data file has been attached below

[**email engagement metrics.csv**](email%20engagement%20metrics.csv)