**Operation Analytics and Investigating Metric Spike**

## (project-3)

By- Abhishek Adyani

**Project description:**

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales.

In this project, with various datasets and tables, the task is to derive insights from the given data to answer questions posed by different departments within the company.

**Approach**:

As we have different datasets, I have first studied them thoroughly and understand the terminologies that are being used in it, then after checking for any null values and cleaned the data if needed. I solved the different questions using my SQL knowledge to derive different insights for the management team.

**Tech-stack used:**

I have used MYSQL workbench 8.0.36 community version by connecting to my local server with the given database, I chose this software as it is very user friendly and a great tool to analyze the data

**Execution**:

**Case study 1 : Job Data Analysis**

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

Query code used:

SELECT

ds AS date,

ROUND((COUNT(job\_id) / SUM(time\_spent) \* 3600, 2) AS job\_review\_per\_hour

FROM

job\_data

WHERE

ds BETWEEN '2020-11-01' AND '2020-11-30'

GROUP BY ds

ORDER BY ds;

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

Query code used:

SELECT

ROUND((COUNT(event) / SUM(time\_spent)), 2) AS weekly\_throughput

FROM

job\_data;

and for daily throughput :

SELECT

ds AS date,

ROUND((COUNT(event) / SUM(time\_spent)), 2) AS daily\_throughput

FROM

job\_data

GROUP BY ds;

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

Query code used:

SELECT

language,

COUNT(language) AS language\_used,

(COUNT(language) / (SELECT COUNT(\*) FROM job\_data)) \* 100 AS perc\_share\_language

FROM

job\_data

WHERE

ds BETWEEN '2020-11-01' AND '2020-11-30'

GROUP BY language

ORDER BY language DESC;

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

Query code used:

SELECT \*

FROM

(

SELECT \*,

row\_number( )over(partition by job\_id) as rownum

FROM job\_data

)a

WHERE rownum>1

union

SELECT \* from

(

SELECT \*,

row\_number( )over(partition by actor\_id) as rownum

FROM job\_data

)a

WHERE rownum>1;

Case study 2 : Investigating Metric Spike

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

Query code used:

SELECT

EXTRACT(WEEK FROM occurred\_at) AS week\_num,

COUNT(DISTINCT user\_id) AS users\_engage

FROM

events

WHERE

event\_type = 'engagement'

GROUP BY week\_num

ORDER BY week\_num;

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

Query code used:

SELECT quarter\_num, year, active\_users,

SUM(active\_users) OVER (ORDER BY year, quarter\_num rows between unbounded preceding and current row) AS growth\_of\_users

FROM (

SELECT EXTRACT(QUARTER FROM activated\_at) AS quarter\_num,

EXTRACT (YEAR FROM activated\_at) AS year,

EXTRACT (DISTINCT user\_id) AS active\_users

FROM users

GROUP BY quarter\_num, year

ORDER BY quarter\_num, year

)AS a;

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

Query code used:

SELECT

EXTRACT(WEEK FROM occurred\_at) AS week\_num,

COUNT(DISTINCT user\_id) AS users\_retained

FROM

events

WHERE

event\_name = 'complete\_signup'

AND event\_type = 'signup\_flow'

GROUP BY week\_num

ORDER BY week\_num;

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

Query code used:

SELECT

device,

EXTRACT(WEEK FROM occurred\_at) AS week\_num,

COUNT(DISTINCT user\_id) AS users\_engage

FROM

events

WHERE

event\_type = 'engagement'

GROUP BY week\_num , device

ORDER BY week\_num;

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

Query code used:

SELECT

(SUM(CASE WHEN email\_category = 'opened' THEN 1 ELSE 0 END) / SUM(CASE WHEN email\_category = 'email\_sent' THEN 1 ELSE 0 END)) \* 100 AS percentage\_email\_opened ,

(SUM(CASE WHEN email\_category = 'clickthrough' THEN 1 ELSE 0 END) / SUM(CASE WHEN email\_category = 'email\_sent' THEN 1 ELSE 0 END)) \* 100 AS percentage\_email\_clickthrough

FROM

(SELECT \*,

CASE

WHEN action IN ('email\_clickthrough') THEN ('clickthrough')

WHEN action IN ('email\_open') THEN ('opened')

WHEN action IN ('sent\_weekly\_digest' , 'sent\_reengagement\_email') THEN ('email\_sent')

END AS email\_category

FROM

email\_events) a;

**Insights:**

**Case study 1 :**

* Number of jobs reviewed per hour for each day in November 2020 are varied from **64.29** jobs/hour (lowest) on **26-11-2020** to **180** jobs/hour (highest) on **30-11-2020**.
* I used 7-day rolling throughput or weekly throughput as it directly gives us the insight on day 1st to day 7th. 7-day rolling throughput average was **0.03**
* Percentage share of each language for the past 30days are Persian (**37.5%**) being most used and every other language is **12.5%**
* There are 3 duplicate 2 from job\_id and 1 from actor\_id when we partition it by job\_id, but if we look the overall data it’s all unique.

**Case study 2:**

* In the weekly user engagement we saw a trend of increasing user engagement from 18th week to 30th week, and saw a gradual decrease from 31st week to 31st week, the highest number of user engagement was on 30th week being **1467.**
* There are total of **9381** number of active users from 1st quarter of 2013 to 3rd quarter of 2014. The highest growth rate of users was on 2nd quarter of 2014.
* From weekly retention analysis, the number of users on a weekly basis after signing up for a product retained **238 on 30th week** being the highest and the lowest went to **18 on 35th week.**
* As from the weekly engagement per device we saw a trend where users who were using **MacBook pro** and **Lenovo ThinkPad** were engaging more than others.
* The opening rate for email is around **34%** and the clickthrough rate is around **15%**. Means **15%** of the customers are engaging with the company through emails and **34%** of the customers are only opening the emails.

**Results:**

In this project, I learned how to perform operational analysis through my SQL knowledge where I used a bunch of different concepts like window functions, nested queries, aggerate functions etc. It helped me to vast my SQL knowledge. This project helped me know what kind of right questions to ask under different circumstances. It helped me gain knowledge on investigating metric spike which is the key concept of operational analysis.