PROJECT REPORT

Operation Analytics and Investigating Metric Spike

By

SHIVAM PAL

Contents

1 Project Description 3

1.1 About Project 3

1.2 How I handle the things? 3

1.3 What are the things that I am going to find out through the project? 3

Case Study 1 (Job Data) 3

Case Study 2 (Investigating Metric Spike) 3

2 Approach 4

3 Tech-Stack 12

4 Insights 12

5 Result 13

6 Drive Link 13

1 Project Description

1.1 About Project

Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon.

This kind of analysis is further used to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.

Investigating metric spike is also an important part of operation analytics as being a Data Analyst we must be able to understand or make other teams understand questions like- Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that it's very important to investigate metric spike.

In this project, we are provided with different data sets, tables from which we must derive certain insights out of it and answer the questions asked by different departments.

1.2 How I handle the things?

First thing that we need to do is to ask appropriate question. So, the question asked by stakeholders are not actual question for us, we again need to ask an appropriate question.

Secondly, we use relevant tools to fetch required data from database given to us.

1.3 What are the things that I am going to find out through the project?

Case Study 1 (Job Data)

a) Number of jobs reviewed: Amount of jobs reviewed over time.
b) Throughput: It is the no. of events happening per second.
c) Percentage share of each language: Share of each language for different contents.
d) Duplicate rows: Rows that have the same value present in them.

Case Study 2 (Investigating Metric Spike)

a) User Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service.
b) User Growth: Amount of users growing over time for a product.
c) Weekly Retention: Users getting retained weekly after signing-up for a product.
d) Weekly Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
e) Email Engagement: Users engaging with the email service.

2 Approach

To complete this project the first thing required is to import data from excel (One of the 2 ways can be used to import data in MySQL workbench from excel – By writing query or by using create new schema tab.

How I did? – click me to know

Second crucial task is to ask appropriate question, then project is nothing more than getting access to database and retrieving required data using SQL query.

Case Study 1 (Job Data):

Asked Question	Appropriate Question
• **Number of jobs reviewed:** Amount of jobs reviewed over time.	• Calculate the number of jobs reviewed per hour per day for November 2020?
• **Throughput:** It is the no. of events happening per second.	• Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?
• **Percentage share of each language:** Share of each language for different contents.	• Calculate the percentage share of each language in the last 30 days?
• **Duplicate rows:** Rows that have the same value present in them.	• How will you display duplicates from the table?

Case Study 2 (Investigating Metric Spike):

Asked Question	Appropriate Question
• **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.	• Calculate the weekly user engagement i.e. How many different users engage every nth week?
• **User Growth:** Amount of users growing over time for a product.	• Calculate the user growth for product i.e. How many different users join platform every nth week?
• **Weekly Retention:** Users getting retained weekly after signing-up for a product.	• Calculate the weekly retention of users-sign up cohort?
• **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.	• Calculate the weekly engagement per device i.e. How many different users uses different devices in every nth week?
• **Email Engagement:** Users engaging with the email service	• Calculate the email engagement metrics i.e., email open rate and email click rate.

Case Study 1 (Job Data):

a) *Number of jobs reviewed* - Calculate the number of jobs reviewed per hour per day for November 2020?

MySQL Query:

```
**************************
SELECT
        COUNT(DISTINCT job_id)/(30*24) AS_number_of_jobs_reviewed_per_hour_per_day
FROM
        job_data;
**************************
```

Output:

number_of_jobs_reviewed_per_hour_per_day
0.0083

b) Throughput: - Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

MySQL Query:

```
***********************************
SELECT
        ds,
        jobs_reviewed,
        AVG(jobs_reviewed) OVER(ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT
        ROW) AS throughput_7
FROM
        (SELECT
                ds,
                COUNT( DISTINCT job_id) AS jobs_reviewed
        FROM
                job_data
        GROUP BY
                ds
        ORDER BY
                ds) a;
***********************************
```

Output:

ds	jobs_reviewed	throughput_7
2020-11-25	1	1.0000
2020-11-26	1	1.0000
2020-11-27	1	1.0000
2020-11-28	2	1.2500
2020-11-29	1	1.2000
2020-11-30	2	1.3333

NOTE – For throughput as per me 7-day rolling is preferable because it is a metric that allows us to find trends that would otherwise be hard to detect.

c) Percentage share of each language - Calculate the percentage share of each language in the last 30 days?

MySQL Query:

```
**********
SELECT
        language,
        (num_of_jobs/total_jobs)*100 AS percent_share_of_language
FROM
        (SELECT
                language,
                COUNT(DISTINCT job_id) AS num_of_jobs
        FROM
                job_data
        GROUP BY
                language) a
CROSS JOIN
        (SELECT
                COUNT(DISTINCT job_id) AS total_jobs
        FROM
                job_data) b;
**********
```

Output:

language	percent_share_of_language
Arabic	16.6667
English	16.6667
French	16.6667
Hindi	16.6667
Italian	16.6667
Persian	16.6667

d) Duplicate rows - Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

MySQL Query:

```
**********
SELECT
        *
FROM
        (SELECT
                *,
                ROW_NUMBER() OVER(PARTITION BY job_id) AS row_num
        FROM
                job_data) a
WHERE
        row_num>1;

**********
```

Output:

ds	job_id	actor_id	event	language	time_spent	org	row_num
2020-11-28	23	1005	transfer	Persian	22	D	2
2020-11-26	23	1004	skip	Persian	56	A	3

NOTE – No duplicate found

Case Study 2 (Investigating Metric Spike):

a) User Engagement: Calculate the weekly user engagement?
MySQL Query:

****

```
SELECT
        EXTRACT(WEEK FROM occurred_at) AS weeknum ,
        COUNT(DISTINCT user_id) AS no_of_distinct_user
FROM
        tutorial.yammer_events
GROUP BY
        weeknum;
```

****

Output:

weeknum	no_of_distinct_user
18	791
19	1244
20	1270
21	1341
22	1293
….	….

b) User Growth: Calculate the user growth for product?

MySQL Query:

****

```
SELECT
        year,
        weeknum,
        num_active_user,
        SUM(num_active_user) OVER(ORDER BY year, weeknum ROWS BETWEEN UNBOUNDED
        PRECEDING AND CURRENT ROW) AS cum_active_users
FROM
        (SELECT
                EXTRACT(year from a.activated_at) AS year,
                EXTRACT(week from a.activated_at) AS weeknum,
                COUNT(DISTINCT user_id) AS num_active_user
        FROM
                tutorial.yammer_users a
        WHERE
                state = 'active'
        GROUP BY
                year,
                weeknum
        ORDER BY
                year,
                weeknum) a;
```

****

Output:

year	weeknum	num_active_user	cum_active_user
2013	1	67	67
2013	2	29	96
2013	3	47	143
2013	4	36	179
2013	5	30	209
….	….	….	….

c) Weekly Retention: Calculate the weekly retention of users-sign up cohort?

MySQL Query:

```
**************************************************************

SELECT
        COUNT(user_id),
        SUM(CASE WHEN retention_week = 1 THEN 1 ELSE 0 END) as week_1
FROM
        (SELECT
                a.user_id,
                a.signup_week,
                b.engagement_week,
                b.engagement_week - a.signup_week AS retention_week
        FROM
                ((SELECT
                        DISTINCT user_id,
                        EXTRACT(week FROM occurred_at) AS signup_week
                FROM
                        tutorial.yammer_events
                WHERE
                        event_type = 'signup_flow' AND event_name = 'complete_signup' AND
                        EXTRACT(week from occurred_at) = 18) a
                LEFT JOIN
                        (SELECT
                                DISTINCT user_id,
                                EXTRACT(week FROM occurred_at) AS engagement_week
                        FROM
                                tutorial.yammer_events
                        WHERE
                                event_type = 'engagement') b
                ON
                        a.user_id = b.user_id)
                ORDER BY
                        a.user_id) a;
**********************************************************************
```

Output:

count	Week_1
317	64

d) Weekly Engagement: Calculate the weekly engagement per device?

MySQL Query:

```
**************************************************************

SELECT
        EXTRACT(YEAR FROM occurred_at) AS year,
        EXTRACT(WEEK FROM occurred_at) AS week,
        device,
        COUNT(DISTINCT user_id)
FROM
        tutorial.yammer_events
WHERE
        event_type = 'engagement'
GROUP BY
        1,2,3
ORDER BY
        1,2,3;
**************************************************************************
```

Output:

year	week	device	count
2014	18	acer aspire desktop	10
2014	18	acer aspire notebook	21
2014	18	amazon fire phone	4
2014	18	asus chromebook	23
2014	18	dell inspiron desktop	21
….	….	….	….

e) Email Engagement: Calculate the email engagement metrics?

MySQL Query:

```
**********************************************************

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
                *,
                CASE
                        WHEN action IN ('sent_weekly_digest', 'sent_reengagement_email')
                        THEN 'email_sent'
                        WHEN action IN ('email_open') THEN 'email_open'
                        WHEN action in ('email_clickthrough') THEN 'email_clicked'
                END AS email_cat
        FROM
        tutorial.yammer_emails)
**********************************************************************
```

Output:

email_open_rate	email_clicked_rate
33.5834	14.7899

3 Tech-Stack

Software/Tool Used	Purpose
MS office Professional Plus 2019 (MS Word)	Documentation
MS office Professional Plus 2019 (MS Excel)	For getting logic
MySQL Workbench 8.0 CE	Loading Database and Writing Query
Mode.com	Rechecking answers

4 Insights

- 32nd Week of year 2020 doesn't seem good for platform as significant number of users disengaged.
- The overall growth of platform is good, as the number of new users keep on increasing.
- Mostly used laptop is MacBook pro.
- Mostly used Mobile device is iPhone 5.

5 Result

- This project makes us to understand how Windows Function are used in different SQL problems, the concept of OVER(), PARTITION BY, LEAD and LAG, TIME FUNCTION etc. becomes easy after completing this project.

- This project also makes us to learn how actually insight are drawn from data, to do that an appropriate question is needed to ask.

6 Drive Link

- Folder link
- Case Study 1 SQL Script
- Case Study 2 SQL Script