# OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE. Advanced SQL Metric Spikes Achieved Target Inactive 254 Active

**Analysis** 

3000

Total3254

# **INTRODUCTION**

This project focuses on analyzing the data which is provided by company. My task is to derive insights and answer the questions asked by different departments. So that these insights are then used by ops team, support team, marketing team, etc to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between crossfunctional teams, and more effective workflows

In case study 1 there is job\_data table while in case study 2 there are users, events and email\_events tables.

#### WHAT IS OPERATION ANALYTICS?

Operational analytics focuses on measuring the existing and real-time operations of the company so that the company can monitor their day-to-day operations basis which they can take the necessary actions to improve customer satisfaction and bottom line.

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. Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company's fortune.

# Here are few examples:

- **Ops**: Developers can use real-time data to look at how customers are using their products and make changes on the fly.
- **Marketing**: Businesses can optimize user engagement in real-time by using operational analytics to make personalized recommendations.

# **PROJECT APPROACH**

This project is developed using SQL Workbench. First I need to create database by using dataset file which was provided by the company. Next step load the data into SQL Workbench then performed analysis and find the information that will the help the ops team, support team, marketing team, etc to 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.

# **TECH-STACK USED**

MySQL Workbench is a visual editor that unifies data modelling, SQL development and database administration in one interface. It allows you to visually design, generate and manage databases.



MySQL Workbench in widely used to handle structured data. It is an opensource Relational Database Management System (RDBMS) developed by Oracle Corporation, Sun Microsystems that uses Structured Query language (SQL) to interact with databases.

MySQL Workbench offers database migration options, making it easier to move data to and from the Microsoft SQL Server, Microsoft Access and other RDBMS tables.

# **KEY FUNCTIONALITIES OF MYSQL WORKBENCH**

- Visual SQL Editor: MySQL Workbench is equipped with a visual SQL editor where developers can build, edit, and run queries. What's great about this is that it allows you to preview your changes before applying them.
- Database Administration: Aside from providing you with SQL editing tools, MySQL Workbench also comes with a database administration suite. This makes it easy for you to audit your databases, configure servers, and view logs.
- **Performance Monitoring**: MySQL Workbench gives users a dashboard where they can view the status of their queries, client timing, network latency, and index usage. This allows for simpler identification of possible ways to optimize SQL performance.

# **INSIGHTS:**

## **CASE STUDY 1: JOB DATA ANALYSIS**

## A.) Jobs Reviewed Over Time:

Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.

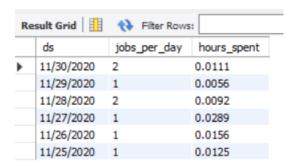
#### **QUERY:**

select ds, count(job\_id) as jobs\_per\_day, sum(time\_spent)/3600 as hours\_spent

from job\_data

where ds>='11/01/2020' and ds<='11/30/2020'

group by ds;



In this I have used sum and comparison operators in the query to get the result.

#### **B.) Throughput Analysis:**

Write an SQL query to calculate the 7-day rolling average of throughput. Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why.

#### **QUERY:**

select round(count(event)/sum(time\_spent),2) as "Weekly Throughput"

from job\_data;



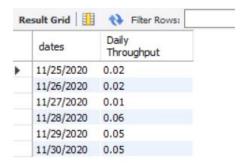
Here, the weekly throughput is 0.03.

select ds as dates, round(count(event)/sum(time\_spent),2)as "Daily Throughput"

from job\_data

group by ds

order by ds;



On date 11/28/2020 the throughput is highest 0.06.

Metrics will always go up and down on a weekly and daily basis. You'll get members faster every day or minute if you want. As a result, rolling metrics are superb at showing if your metrics are trending up or down on a daily level.

# **C.)** Language Share Analysis:

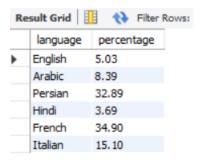
Write an SQL query to calculate the percentage share of each language over the last 30 days.

#### **QUERY:**

select language, round( (sum(time\_spent)/(select sum(time\_spent) from job\_data)) \* 100 ,
2) as percentage

from job\_data

group by language;



French language has highest percentage share with 34.90% total over the last 30 days.

# **D.) Duplicate Rows Detection:**

Write an SQL query to display duplicate rows from the job\_data table.

## **QUERY:**

select actor\_id, count(\*) as duplicates

from job\_data

group by actor\_id

having count(\*)>1;



Actor ID 1003 has duplicates rows.

#### **CASE STUDY 2: INVESTIGATING METRIC SPIKE**

# A.) Weekly User Engagement:

Write an SQL query to calculate the weekly user engagement.

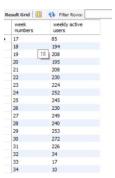
# **QUERY:**

select extract(week from occurred\_at) as "week numbers", count(distinct user\_id) as "weekly active users"

from events

where event\_type = 'engagement'

#### group by 1;



# **B.) User Growth Analysis:**

Write an SQL query to calculate the user growth for the product.

# **QUERY:**

select registration\_month, new\_users, cumulative\_users,
case when lag(cumulative\_users) over (order by registration\_month) = 0 then null
else round((new\_users / lag(cumulative\_users) over (order by registration\_month)) \* 100,
2)

end as growth\_in\_percentage

from (

select date\_format(created\_at, '%y-%m') as registration\_month, count(distinct user\_id) as new\_users,

@cumulative\_users := @cumulative\_users + count(distinct user\_id) as cumulative\_users

from users

cross join (select @cumulative\_users := 0) as init

group by registration\_month ) growth\_data

order by registration\_month;

Re	esult Grid	Filter Rows:		Export: Wrap Cell Co
	registration_month	new_users	cumulative_users	growth_in_percentage
	2013-01	160	160	HULL
	2013-02	160	320	100
	2013-03	150	470	46.88
	2013-04	181	651	38.51
	2013-05	214	865	32.87
	2013-06	213	1078	24.62
	2013-07	284	1362	26.35
	2013-08	316	1678	23.2
	2013-09	330	2008	19.67
	2013-10	390	2398	19.42
	2013-11	399	2797	16.64
	2013-12	486	3283	17.38
	2014-01	552	3835	16.81
	2014-02	525	4360	13.69
	2014-03	615	4975	14.11
	2014-04	726	5701	14.59
	2014-05	779	6480	13.66
	2014-06	873	7353	13.47
	2014-07	997	8350	13.56
	2014-08	1031	9381	12.35

# **C.) Weekly Retention Analysis:**

Write an SQL query to calculate the weekly retention of users based on their sign-up cohort.

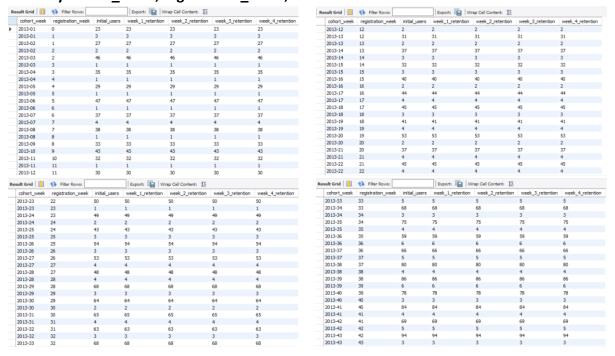
#### **QUERY:**

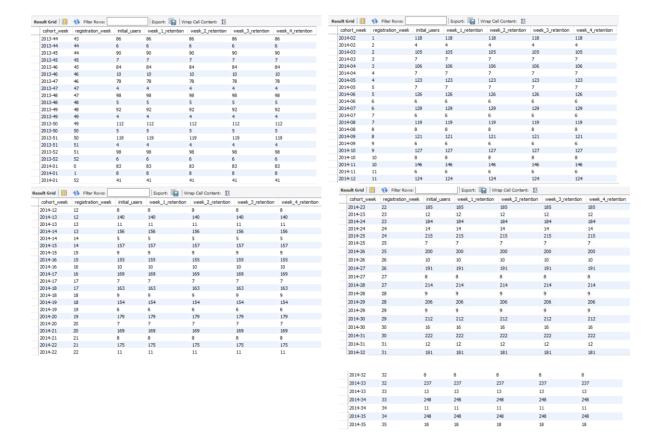
select cohort\_week, registration\_week, count(distinct user\_id) as initial\_users, sum(week\_1\_activity) as week\_1\_retention, sum(week\_2\_activity) as week\_2\_retention, sum(week\_3\_activity) as week\_3\_retention, sum(week\_4\_activity) as week\_4\_retention from ( select user\_id, date\_format(created\_at, '%x-%v') as cohort\_week, week(created\_at) as registration\_week,

max(date(activated\_at) <= date\_add(created\_at, interval 1 week)) as week\_1\_activity,
max(date(activated\_at) <= date\_add(created\_at, interval 2 week)) as week\_2\_activity,
max(date(activated\_at) <= date\_add(created\_at, interval 3 week)) as week\_3\_activity,
max(date(activated\_at) <= date\_add(created\_at, interval 4 week)) as week\_4\_activity
from users</pre>

group by cohort\_week, registration\_week, user\_id ) cohort\_data group by cohort\_week, registration\_week

order by cohort\_week, registration\_week;





## D.) Weekly Engagement Per Device:

#### 4. Write an SQL query to calculate the weekly engagement per device.

#### **QUERY:**

select extract(week from occurred\_at) as "week numbers",

count(distinct case when device in('dell inspiron notebook') then user\_id else null end) as "dell inspiron notebook",

count(distinct case when device in('iphone 5') then user\_id else null end) as "iphone 5", count(distinct case when device in('iphone 4s') then user\_id else null end) as "iphone 4s", count(distinct case when device in('windows surface') then user\_id else null end) as "windows surface",

count(distinct case when device in('mackbook\_air') then user\_id else null end)as "macbook air",

count(distinct case when device in('iphone 5s') then user\_id else null end) as "iphone 5s",

count(distinct case when device in('macbook pro') then user\_id else null end) "mackbook pro" ,

count(distinct case when device in('kindle fire') then user\_id else null end) as "kindle fire", count(distinct case when device in('ipad mini') then user\_id else null end) as "ipad mini", count(distinct case when device in('nexus 7') then user\_id else null end) as "nexus 7", count(distinct case when device in('nexus 5') then user\_id else null end) as "nexus 5", count(distinct case when device in('samsung galaxy s4') then user\_id else null end) as "samsung galaxy s4",

count(distinct case when device in('lenovo thinkpad') then user\_id else null end) as "lenovo thinkpad",

count(distinct case when device in('samsung galaxy tablet') then user\_id else null end) as "samsung galaxy tablet",

count(distinct case when device in('acer aspire notebook') then user\_id else null end) as "acer aspire notebook",

count(distinct case when device in('asus chromebook') then user\_id else null end) as "asus chromebook",

count(distinct case when device in('htc one') then user\_id else null end) as "htc one", count(distinct case when device in('nokia lumia 635') then user\_id else null end) as "nokia lumia 635",

count(distinct case when device in('samsung galaxy note') then user\_id else null end) as "samsung galaxy note",

count(distinct case when device in('acer aspire desktop') then user\_id else null end) as "acer aspire desktop",

count(distinct case when device in('mac mini') then user\_id else null end) as "mac mini", count(distinct case when device in('hp pavilion desktop') then user\_id else null end) as "hp pavilion desktop",

count(distinct case when device in('dell inspiron desktop') then user\_id else null end) as "dell inspiron desktop",

count(distinct case when device in('ipad air') then user\_id else null end) as "ipad air",

count(distinct case when device in('amazon fire phone') then user\_id else null end) as "amazon fire phone",

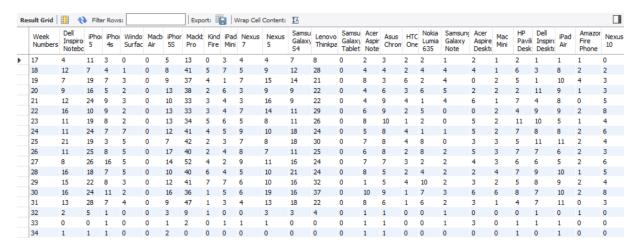
count(distinct case when device in('nexus 10') then user\_id else null end) as "nexus 10"

from events

where event\_type ='engagement'

group by 1

order by 1;



#### E.) Email Engagement Analysis:

# 5. Write an SQL query to calculate the email engagement metrics.

#### **QUERY:**

select week, round((weekly\_digest/total\*100),2) as "weekly digest rane",
round((email\_opens/total\* 100),2) as "email open rate",
round((email\_clickthroughs/total\* 100),2) as "email clickthrough rate",
round((reengagement\_emails/total\* 100),2) as "reengagement email rate"
from ( select extract(week from occurred\_at) as week,
count(case when action = 'sent\_weekly\_digest' then user\_id else null end) as weekly\_digest,
count(case when action = 'email\_open' then user\_id else null end) as email\_opens,

count(case when action = 'email\_clickthrough' then user\_id else null end) as email\_clickthroughs,

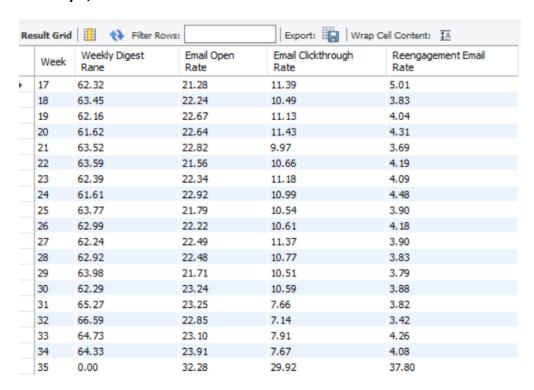
count(case when action = 'sent\_reengagement\_email' then user\_id else null end) as reengagement\_emails,

count(user\_id) as total from email\_events

group by 1) sub

group by 1

order by 1;



# **RESULT**

**How this project helped me:** This project helps me to understand the importance of operation analytics. Through this project I am able to understand how the companies use metric spike as a secret weapon. With an informed and proactive approach, they can leverage insights to make databacked decisions that optimize their strategy and boost ROI.

Challenges that I faced in this project: The challenge here is that the data in case study 2 is very huge, as the huge amount of data SQL Workbench is very slow to import it. To tackle this situation I have to use LOAD DATA statements. Now, there is another problem arises in the column user\_type in events table that has datatype int which is stopping the process of importing. First I need to change its datatype to text then restart the process of loading the data into events table.

# **CONCLUSION:**

Operational Analytics tackles the problem by synchronizing real time data. Operational Analytics has the capability to aggregate data from multiple data sources into a cumulative, organized, actionable solution capable of delivering analytical models in real-time to create individual customer profiles and a holistic view of operations for a company. This guarantees that your operational routines und systems are used efficiently. Whenever utilized correctly, operational analytics can achieve a significant positive effect on our general public and world everywhere and increment the general efficiency of specific areas.