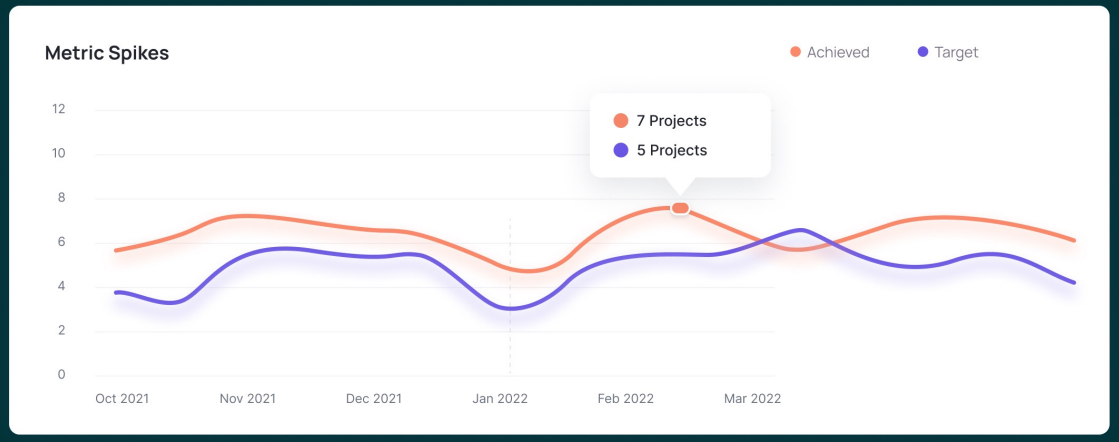
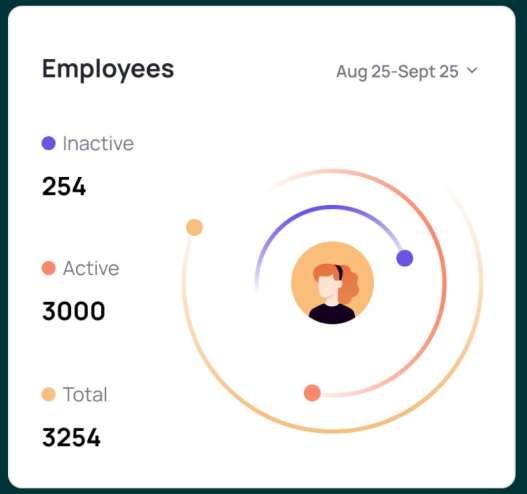
Operation Analytics and Investigating Metric Spike

Trainity Project by Mishree Bagdai

trainity

Operation Analytics & Investigating metric spike case study





Project Description

- Operation analytics is the analysis performed for a company's whole end-to-end operations. This helps the business identify the areas where it needs to make improvements. You collaborate closely with the operations team, the support team, the marketing team, etc. and assist them in drawing conclusions from the data they gather.
- Being one of the most crucial components of a business, this form of analysis is also utilised to forecast the general upward or downward trend in a company's fortune. Better automation, improved communication among cross-functional teams, and more efficient workflows are the results.
- Investigating metric spikes is crucial for operation analytics since a data analyst must be able to comprehend questions or help other teams grasp them.

Project Description

- We will analyse and extract the following information from the data provided.
- Number of jobs reviewed
- Throughput
- Percentage share of each language
- Duplicate rows
- User Engagement
- User Growth
- Weekly Retention
- Weekly Engagement
- Email Engagement

Approach

- Before starting my project, I devoted time to go through the relevant information given on the website. I downloaded the resources and the tables and converted them into proper format if required. Once I had the csv files I created the dataset on mySQL workbench wherein I named my dataset(schema) as 'operations_analytics' and then imported the excel tables on the same and renamed them accordingly. I understood the different columns and their importance in the tables and then started working towards the first case study, in order to derive the required insights.

Tech-Stack Used

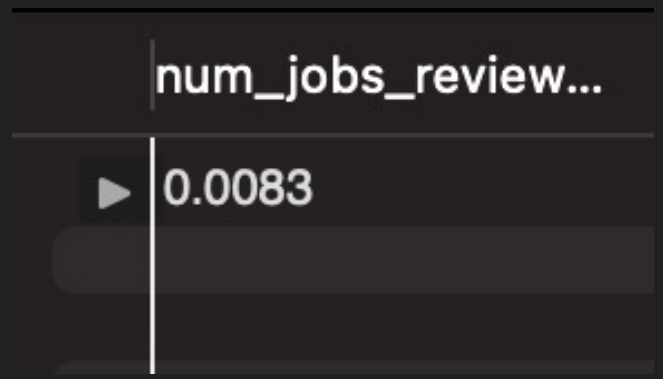
- For the following project I use mySQL workbench on my macOS computer. Along with that I used Microsoft power point to create the presentation and google drive to submit the same.

SQL queries execution – Case Study 1(Job Data)

A. Number of jobs reviewed: Amount of jobs reviewed over time.

task: Calculate the number of jobs reviewed per hour per day for November 2020?

```
use operations_analytics;
select count(distinct job_id)/(30*24) as num_jobs_reviewed
from jobs_data
where
ds between '2020-11-01' and '2020-11-30';
```



num_jobs_review...
0.0083

SQL queries execution – Case Study 1(Job Data)

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

Task: Let's say the above metric is called throughput. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

```
select count(event)/sum(time_spent) as weekly_throughput
from operations_analytics.jobs_data;
```

weekly_throughput
0.0268

```
select ds, jobs_reviewed,
avg(jobs_reviewed)over(order by ds rows between 6 preceding and current row)
as throughput_7_rolling_avg
from
(
select ds, count(distinct job_id) as jobs_reviewed
From operations_analytics.jobs_data
where ds between '2020-11-01' and '2020-11-30'
group by ds
order by ds
)a;
```

ds	jobs_reviewed	throughput_7_rolling_a...
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

SQL queries execution – Case Study 1(Job Data)

C. Percentage share of each language: Share of each language for different contents.

Task: Calculate the percentage share of each language in the last 30 days?

```
select language, count(language)/8*100 as percentage_share
from operations_analytics.jobs_data
group by language;
```

language	percentage_share
English	12.5000
Arabic	12.5000
Persian	37.5000
Hindi	12.5000
French	12.5000
Italian	12.5000

SQL queries execution – Case Study 1(Job Data)

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

Task: Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

```
select * from operations_analytics.jobs_data
where (actor_id) in (
select actor_id from operations_analytics.jobs_data group by actor_id
having count(*) > 1
);
```

ds	job_id	actor_id	event	language	time_spent	org
2020-11-29	23	1003	decision	Persian	20	C
2020-11-25	20	1003	transfer	Italian	45	C

SQL queries execution – 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.

Task: Calculate the weekly user engagement?

```
SELECT
    WEEK(occurred_at, 3) AS week_number,
    MIN(occurred_at) AS week_start_date,
    MAX(occurred_at) AS week_end_date,
    COUNT(*) AS weekly_user_engagement
FROM operations_analytics.events
where event_type = 'engagement'
GROUP BY WEEK(occurred_at, 3);
```

week_number	week_start_date	week_end_date	weekly_user_engagement
18	2014-05-01 02:27:15	2014-05-04 23:39:10	8709
19	2014-05-05 00:01:57	2014-05-11 23:54:44	17532
20	2014-05-12 00:58:46	2014-05-18 23:59:31	17047
21	2014-05-19 00:00:04	2014-05-25 22:38:55	17890
22	2014-05-26 00:02:00	2014-06-01 22:05:02	17193
24	2014-06-09 00:31:36	2014-06-15 23:59:49	18233
23	2014-06-02 01:41:21	2014-06-08 23:28:57	18608
25	2014-06-16 00:00:06	2014-06-22 22:38:05	18976
26	2014-06-23 00:35:35	2014-06-29 23:36:27	18859
29	2014-07-14 00:31:37	2014-07-20 23:59:36	20723
27	2014-06-30 00:10:34	2014-07-06 21:09:06	18959
31	2014-07-28 00:00:43	2014-08-03 23:50:07	21472
30	2014-07-21 00:43:08	2014-07-27 23:59:57	20132
28	2014-07-07 00:35:26	2014-07-13 23:33:51	19965
32	2014-08-04 00:17:09	2014-08-10 21:36:22	18341
33	2014-08-11 00:49:27	2014-08-17 23:05:04	16612
34	2014-08-18 01:03:15	2014-08-24 23:29:24	16158
35	2014-08-25 00:47:36	2014-08-31 23:03:39	16166

SQL queries execution – Case Study 2(Investigating metric spike)

B. User Growth: Amount of users growing over time for a product.

Task: Calculate the user growth for product?

```
select year, num_week, num_active_users,
sum(num_active_users) over(order by year, num_week rows between unbounded
preceding and current row)
as cumm_active_users
from
(select
 extract(year from a.activated_at) as year,
 extract(week from a.activated_at)as num_week,
 count(distinct user_id) as num_active_users
from operations_analytics.users a
where state='active'
group by year, num_week
order by year, num_week
)a;
```

year	num_week	num_active_users	cumm_active_users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372
2013	10	32	404
2013	11	31	435
2013	12	33	468
2013	13	39	507
2013	14	35	542
2013	15	43	585
2013	16	46	631
2013	17	49	680
2013	18	44	724
2013	19	57	781
2013	20	39	820
2013	21	49	869

Result 16

SQL queries execution – Case Study 2(Investigating metric spike)

C. Weekly Retention: Users getting retained weekly after signing-up for a product.

Task: Calculate the weekly retention of users-sign up cohort?

```
select count(user_id),
 sum(case when retention_week = 1 then 1 else 0 end) as
per_week_retention
from
(
select a.user_id,
 a.sign_up_week,
 b.engagement_week,
 b.engagement_week - a.sign_up_week as retention_week
from
(
(select distinct user_id, extract(week from occured_at) as sign_up_week
from operations_analytics.events
where event_type = 'signup_flow'
and event_name = 'complete_signup'
and extract(week from occured_at)=18)a
left join
(select distinct user_id, extract(week from occured_at) as engagement_week
from operations_analytocs.events
where event_type = 'engagement')b
on a.user_id = b.user_id
)
group by user_id
order by user_id;
```

SQL queries execution – Case Study 2(Investigating metric spike)

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

Task: Calculate the weekly engagement per device?

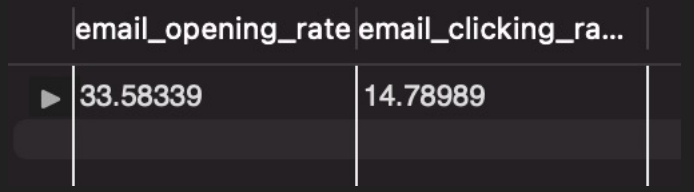
```
select
extract(year from occurred_at) as year_num,
extract(week from occurred_at) as week_num,
device,
count(distinct user_id) as no_of_users
from operations_analytics.events
where event_type = 'engagement'
group by 1,2,3
order by 1,2,3;
```

	year_num	week_num	device	no_of_users
	2014	17	lenovo thinkpad	86
	2014	17	mac mini	6
	2014	17	macbook air	54
	2014	17	macbook pro	143
	2014	17	nexus 10	16
	2014	17	nexus 5	40
	2014	17	nexus 7	18
	2014	17	nokia lumia 635	17
	2014	17	samsumg galaxy tablet	8
	2014	17	samsung galaxy note	7
	2014	17	samsung galaxy s4	52
	2014	17	windows surface	10
	2014	18	acer aspire desktop	26
	2014	18	acer aspire notebook	33
	2014	18	amazon fire phone	9
	2014	18	asus chromebook	42
	2014	18	dell inspiron desktop	58
	2014	18	dell inspiron notebook	77
	2014	18	hp pavilion desktop	37
	2014	18	htc one	19
	2014	18	ipad air	52

SQL queries execution – Case Study 2(Investigating metric spike)

E. Email Engagement: Users engaging with the email service.

Task: Calculate the email engagement metrics?



email_opening_rate	email_clicking_ra...
33.58339	14.78989

Insights and Result

- Case Study 1

- The number of jobs reviewed per hour per day is 0.083
- We prefer to use the 7-day rolling average as it includes all the 7 days instead of daily average.
- The percentage share of each language other than that of Persian is 12.5% and Persian language has the highest share of 37.5%
- There are two duplicate rows if we check the actor_id column otherwise all rows are unique as such

Insights and Result

- Case Study 2

- From week 18 until week 31, when it peaked, the weekly engagement of users grew. After that, it began to decline. This indicates that some customers have not found the product or service to be very high-quality in recent weeks.
- From the first week of 2013 through the 35th week of 2014, there were 9381 active users overall.
- MacBook and iPhone users had the highest overall weekly involvement per device used.
- Around 34% of emails are opened, and 15% of emails are clicked. Users are actively using the email service, which is beneficial for the business's growth.

Insights and Result

After the successful completion of the project, I was able to learn and implement several new functions and derive the required insights. I feel I have a better grasp over the SQL language and how to write efficient queries to extract relevant information. I was able to understand the spike in the metric and changes and what might lead to such fluctuations. I discovered how the business identifies several business-related areas for improvement. I learned about looking into metric spikes (the causes of booms and dips).

Thank you