Operation Analytics and Investigating Metric Spike

Project description

This project focuses on operational analytics which is a crucial process in analysing a company's end to end operations. The goal is to identify areas where the company needs to make the improvements. One key aspect of the operational analytics used in this project is metric spike investigation, which involves understanding and identifying sudden changes in key metrics such as a decline in daily user engagement or a drop in sales.as the lead data analyst in this project at a company like Microsoft, I used advanced SQL skills to analyse dataset and provide valuable insights that help improve the company's operations. This project mainly involves 2 case studies.

First case study focuses on job data analysis. It includes analysing the number of jobs reviewed per hour for each day, calculating 7 days of rolling average throughput, determining percentage share of each language in last 30 days and identifying duplicate rows in the data

The second case study focuses on analysing user activity.it involves measuring the activeness of users on a weekly basis, tracking user growth over time for a product, analysing user retention on a weekly basis after signing up for a product, measuring user engagement per device and calculating email engagement metrics

Approach

Firstly, I created a database called operation analytics and then created necessary tables for the project. Case study 1 involves only one table job_data, while case study 2 involves 3 tables that is email_events, events and users. All these tables were included in the database. After creating the tables, I imported large dataset into each one. I then executed SQL query to analyse the job data and investigate metric spikes. This analysis provided valuable insights that could help improve the company's operations.

Tech- Stack Used

I used My SQL Workbench version 8.0.40 to execute SQL query for analysing the dataset. Because it has a user-friendly interface that makes managing databases easy and efficient.

Project insights

Below are the insights gained from case study 1 and case study 2

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

```
select ds,
round((count(job_id) / (sum(time_spent) / 3600)),2)
as “job viewd per hour for each day”
from job_data
group by ds
```

Result

Below are the number of jobs reviewed per hour for each day

date	jobs viewd per hour per day
2020-11-30	180
2020-11-29	180
2020-11-28	218
2020-11-27	35
2020-11-26	64
2020-11-25	80

2.Throughput Analysis

Calculate the 7-day rolling average of throughput (number of events per second). Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why

Query

```
with daily throughput as
(
select ds,
count(job_id) / sum(time_spent) as daily_throughput
from job_data
group by ds
)
select ds, daily_throughput,
round(avg(daily_throughput)
over (order by ds rows between 6 preceding and current row),4)
as rolling_7_day_throughput
from daily_throughput
order by ds
```

Result

ds	daily_throughput	rolling_7_day_throughput
2020-11-25	0.0222	0.0222
2020-11-26	0.0179	0.0201
2020-11-27	0.0096	0.0166
2020-11-28	0.0606	0.0276
2020-11-29	0.0500	0.0321
2020-11-30	0.0500	0.0351

After analysing the result, I prefer 7 day rolling average over the daily metric. Because it provides a clearer picture of overall trends without the noise of daily ups and downs.

3.Language share Analysis

Calculate the percentage share of each language

Query

```
select language,
count(language) as language_count,
round((count(language) * 100) / (select count (*) from job_data),2)
as percentage
from job_data
group by language
order by percentage desc
```

Result

The percentage share of each language is listed below.

language	language_count	percentage
Persian	3	37.50
English	1	12.50
Arabic	1	12.50
Hindi	1	12.50
French	1	12.50
Italian	1	12.50

4. Duplicate rows analysis

Identify duplicate rows in the data.

Query

```
select * from job_data where job_id in
(
select job_id from job_data
group by job_id
having count(*)  >  1
)
```

Result

ds	job_id	actor_id	event	language	time_spent	org
2020-11-29	23	1003	decision	Persian	20	C
2020-11-28	23	1005	transfer	Persian	22	D
2020-11-26	23	1004	skip	Persian	56	A

Case study 2 : Investigating_Metric_Spike

1.Weekly user engagement

Calculate the weekly user engagement.

Query

```
select  extract(week from occurred_at)  as  weeks,
count(distinct job_id) as “number of users”
from events
where event_type= ‘ engagement ’
group by weeks
```

Result

weeks	number of users
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104

2. User growth analysis

Analyse the growth of users over time for a product

Query

```
select date_format(created_at, “%Y - %M”) as “year and month”,
count(user_id) as “total number of users”
from users
group by date_format(created_at, “%Y - %M”)
order by “year and month”
```

Result

The growth of users based on the year and month time period is given below.

year and month	total number of users
2013 - May	214
2013 - June	213
2013 - July	284
2013 - August	316
2013 - September	330
2013 - October	390
2013 - November	399
2013 - December	486
2014 - January	552
2014 - February	525
2014 - March	615
2014 - April	726
2014 - May	779
2014 - June	873
2014 - July	997
2014 - August	1031

3. Weekly retention analysis

Analyse the retention of users on a weekly basis after signing up for a product.it means calculate the weekly retention of users based on their sign-up cohort

Query

```
Select extract(week from occurred_at)  as  weeks,
count(distinct user_id)  as  “number of users”
from events
where event_type= ‘ signup_flow’ and event_name= ’complete_signup’
group by weeks
```

Result

weeks	number of users
17	72
18	163
19	185
20	176
21	183
22	196
23	196
24	229
25	207
26	201
27	222
28	215
29	221
30	238
31	193
32	245
33	261
34	259
35	18

4. Weekly engagement per device

Measure the activeness of users on a weekly basis per device.

Query

```
select  extract(week from occurred_at) as weeks,
count(distinct user_id) as “number of users”,
device  from events
where event_type= ‘engagement’
group by device, weeks
order by weeks
```

Result

The following table provides a sample of the project results. The complete output contains, which are too extensive to display on a single page, so only a subset is shown here for reference

weeks	number of users	device
17	9	acer aspire desktop
17	20	acer aspire notebook
17	4	amazon fire phone
17	21	asus chromebook
17	18	dell inspiron desktop
17	46	dell inspiron notebook
17	14	hp pavilion desktop
17	16	htc one
17	27	ipad air
17	19	ipad mini
17	21	iphone 4s
17	65	iphone 5
17	42	iphone 5s
17	6	kindle fire
17	86	lenovo thinkpad
17	6	mac mini

5. Email engagement analysis

Calculate the email engagement metrics.

Query

```
Select
(sum(case when action= ‘email_open’ then 1 else 0 end) /
sum(case when action = ‘sent_weekly_digest’ or
action= ‘sent_reengagment_email’ then 1 else 0 end)) * 100 as “open rate”


(sum(case when action= ‘email_clickthrough’ then 1 else 0 end) /
 sum(case when action = ‘sent_weekly_digest’ or
action=‘sent_reengagment_email’ then 1 else 0 end)) *100
as “click through rate”
from email_events
```

Result

Two email engagement metrics are available here open rate and click through rate.

open rate	click through rate
33.5834	14.7899

Conclusion

Through this project I gained valuable knowledge in operational analytics and investigating metric spike. I learned how to import large excel files into SQL, which streamlining data processing and analysis. By examining the datasets, I derived actionable insights such as the number of jobs reviewed per hour for each day, the 7 days rolling average of throughput, language percentage shares and duplicate rows in the data. I also analysed user activeness on a weekly basis, tracked user growth trends over time, evaluated user retention after sign-up and measured engagement across devices and email metrics. These findings significantly enhanced my understanding of data driven strategies and decision making

Drive link

https://drive.google.com/file/d/1jJVqJIqBN4gQOXZEAK9TcffZLN7WcpuD/view?usp=sharing