Operations Analytics

Submitted by – Akhilesh Chaturvedi
Date: 13 Dec 2022

Project Description and Approach

Operations Analytics is important for getting the glimpse, as how businesses understand their users by capturing different types of data.
In Operations Analytics overall functioning of a company is investigated for taking better performance decision in future. It's like an offset removal system for any company to find continuous operational efficiency.

Approach

The Operations data for performing a review work by different actors in different languages was collected. The Dataset was uploaded to MySQL Workbench, which is having columns like date, job_id, actor_id, event, language, time_spent and organisation etc. After that, as per Business requirement, the data was analysed using SQL queries, to find solution for respective problems accordingly.

Tech-Stack Used: MySQL Workbench 8

Problem Statements

Case Study 1 (Job Data)

Q: A- Number of jobs reviewed is equal to number of jobs reviewed over time. Calculate the number of jobs reviewed per hour per day for November 2020?

```
5  select sum(hours_spent)/sum(jobs_per_day) as jobs_reviewed_per_hour_per_day
6  from
7  (select count(job_id) as jobs_per_day, sum(time_spent/3600) as hours_spent
8  from table1
9  where ds between '2020-11-1' and '2020-11-30'
10 group by ds) a;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

jobs_reviewed_per_hour_per_day
0.01036250

Q: B- 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?

```
15 •  select ds, num_rev,
16        sum(num_rev) over (partition by ds order by ds rows between 6 preceding and current row) / sum(total_time)
17        over (partition by ds order by ds rows between 6 preceding and current row) as thoughput_7day
18    from (select ds,
19        count(job_id) as num_rev,
20        sum(time_spent) as total_time
21    from table1
22    group by ds ) ab;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

ds	num_rev	thoughput_7day
2020-11-25	1	0.0222
2020-11-26	1	0.0179
2020-11-27	1	0.0096
2020-11-28	2	0.0606
2020-11-29	1	0.0500
2020-11-30	2	0.0500

Q: C- Percentage share of each language: Share of each language for different contents. Calculate the percentage share of each language in the last 30 days?

```
31 •  select lang,
32    sum(count(lang)) over (partition by lang order by lang rows between unbounded preceding and unbounded following)
33    as lang_occ,
34    sum(count(lang)) over () as lang_tot,
35    (100*sum(count(lang)) over (partition by lang order by lang rows between unbounded preceding and unbounded following) / sum(count(lang))
36    over () )as percentage
37
38    from table1
39    group by lang;
40
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

lang	lang_occ	lang_tot	percentage
Arabic	1	8	12.5000
English	1	8	12.5000
French	1	8	12.5000
Hindi	1	8	12.5000
Italian	1	8	12.5000
Persion	3	8	37.5000

Q: D- Duplicate rows: Rows that have the same value present in them?

```
61  select *, rownum
62  from (
63      SELECT *,
64      ROW_NUMBER() OVER (partition by job_id order by job_id) AS rownum
65      FROM table1) a
66  where rownum > 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

job_id	actor_id	event_sa	lang	time_spent	org	ds	rownum	rownum
23	1004	skip	Persion	56	A	2020-11-26	2	2
23	1005	transfer	Persion	22	D	2020-11-28	3	3