INSTAGRAM USER ANALYTICS

ADVANCED SQL

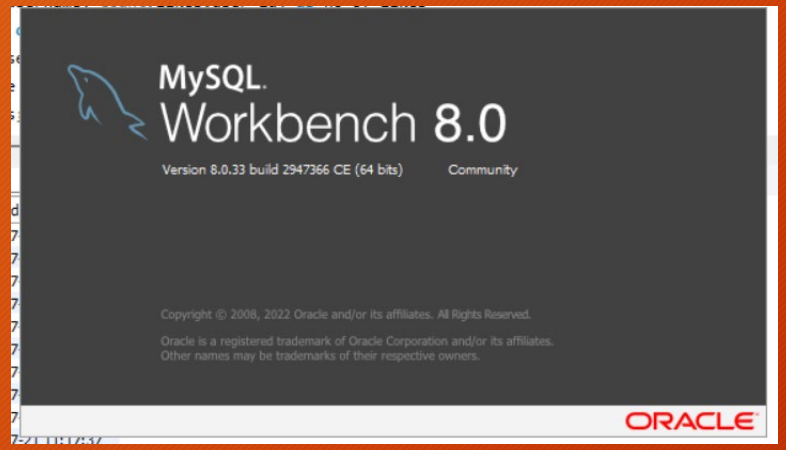
Project by: KIRAN KUMAR CHELLI

Project Description:

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.

Approach: As per the dataset given, I have executed the queries. To find out solution for each question I wrote queries. As the above 4 tables contains the data according to their names, which can be used in our opernation analytics.

Tech Stack Used: In this analytics process I used



to write queries. As per the tutorials I downloaded and installed it.

Case Study 1 (Job Data):

A. Number of jobs reviewed: Amount of jobs reviewed over time.
My task: Calculate the number of jobs reviewed per hour per day for November 2020?

```
Select
count(distinct job_id)/(30*24) as num_jobs_reviewed
from job_data
where ds between '2020-11-01' and '2020-11-30';
```

B. **Throughput Analysis:** It is the no. of events happening per second.
My task: Let's say the above metric is called throughput. Calculate 7 day rolling average ofthroughput? For throughput, do you prefer daily metric or 7-day rolling and why?

```
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 job_data where ds between '2020-11-01' and '2020-11-30
group by ds
order by ds
)a;
```

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

```
select language, num_jobs,
100.0% num_jobs/total_jobs as pct_share_jobs
From
(
select language, count(distinct job_id) as num_jobs
from job_data
group by language
)a
cross join
(
select count(distinct job_id) as total_jobs
from job_data
)b;
```

D. Duplicate rows: Rows that have the same value present in them.
My task: Let's say you see some duplicate rows in the data. How will you display duplicatesfrom the table?

```
select * from
(
select *,row_number()over (partition by job_id) as rownum
from job_data
)a
where rownum>1;
```

Result

At starting stage this project was bit complicated to me. But I practiced again and again in SQL for better understanding the queries to get the desired result. Thank you for giving me this opportunity to explore myself to understand the Data Analytics.

Thank You