Case Study 1 : Job Data Analysis

A)Jobs Revived Over Time

```
Query: select count(distinct job id)/(30*24)
as jobs per day
from job data
where monthname(cast(ds as date))='November';
output:
  26
        #A)no of jobs reviewd
  28 • select count(distinct job_id)/(30*24)
       as jobs_per_day
  29
       from job_data
  30
       where monthname(cast(ds as date))='November';
Export: Wrap Cell Content: IA
   jobs_per_day
 0.0083
```

B) Throughput Analysis

```
Query: with query1 as

(select ds, count(distinct event) as total_events

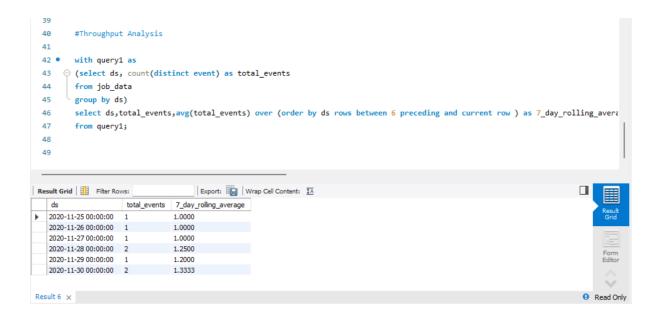
from job_data

group by ds)

select ds,total_events,avg(total_events) over (order by ds rows between 6 preceding and current row ) as 7_day_rolling_average

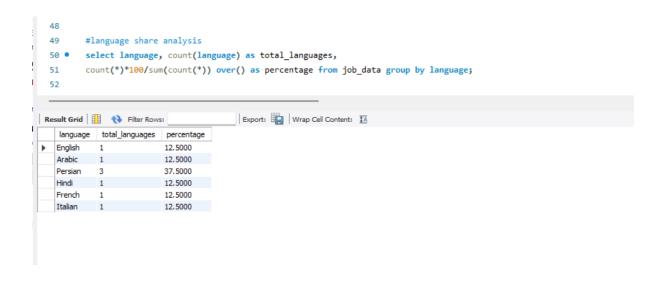
from query1;
```

Reason: Since 7 day rolling average provides better understanding of trends over time, as it provides a longer term perspective comperatively.



C)Language Share Analysis

Query: select language, count(language) as total_languages,
count(*)*100/sum(count(*)) over() as percentage from job_data group by language;



D)Duplicate rows detection

Query: with query1 as(select * , row_number() over (partition by job_id) as row_Num
from job_data)

select * from query1 where row_num>1;



Project details:

<u>Description</u>: The purpose of this project is to identify no of jobs reviewed, to identify the duplicate records present in the data given and so on.

Approach: As a first step I went through the project to identify the topics that I need to implement in the project. later I learnt those concepts and implemented in the project.

Tech-stack used: MySQL workbench 8.0

Reason : I have used MySQL workbench. Reason why I have selected the mentioned tech stack is it is cross platform compatible open source RDBMS.

<u>Insights</u>: The insights we can draw from this project is to implement & day rolling average concept, number of jobs reviewed and so on

<u>Result</u>: I was able to extract the required data as per the project given. I gained knowledge in new concepts like calculating throughput, 7 day rolling average and so on.