This is a real time dataset of the ineuron technical consultant team. You have to perform hive analysis on this given dataset.

Download Dataset 1 - https://drive.google.com/file/d/1WrG-9qv6atP-W3P\_-gYln1hHyFKRKMHP/view

Download Dataset 2 - https://drive.google.com/file/d/1-JIPCZ34dyN6k9CqJa-Y8yxIGq6vTVXU/view

Note: both files are csv files.

1. Create a schema based on the given dataset

Ans. CREATE schema data;

2. Dump the data inside the hdfs in the given schema location.

Ans.

INSERT OVERWRITE DIRECTORY '/path/to/hdfs/location'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

SELECT \*FROM Agent;

3. List of all agents' names.

Ans. select agent\_name from agent\_csv

4. Find out agent average rating.

Ans.SELECT agent\_name,avg(average\_rating) from agent1

GROUP BY agent\_name

5. Total working days for each agents ( not complete)

Ans. SELECT agent\_name, SUM(DATE(end\_date, start\_date)) AS total\_working\_days

FROM agent\_csv

GROUP BY agent\_name;

6. Total query that each agent have taken

Ans. SELECT agent\_name, COUNT(\*) AS num\_queries

FROM agent\_csv

GROUP BY agent\_name;

7. Total Feedback that each agent have received

Ans. SELECT agent\_name,sum(total\_feedback ) FROM agent\_csv GROUP BY agent\_name;

8. Agent name who have average rating between 3.5 to 4

Ans.select agent\_name from agent\_csv where average\_rating between (3.5,4)

GROUP BY agent\_name;

9. Agent name who have rating less than 3.5

Ans.

SELECT agent\_name

FROM agent\_csv

WHERE rating < 3.5

GROUP BY agent\_name;

10. Agent name who have rating more than 4.5

Ans. SELECT agent\_name

FROM agent\_csv

WHERE rating > 4.5

GROUP BY agent\_name;

11. How many feedback agents have received more than 4.5 average

Ans. SELECT agent\_name, sum(Total Feedback)

FROM agent\_csv

WHERE Average rating > 4.5

Group BY agent\_name;

12. average weekly response time for each agent

Ans. SELECT agent\_name,weekofyear(date),

avg(unix\_timestamp(average\_response\_time,'HH:mm:ss')) AS avg\_response\_time

FROM agent\_csv

GROUP BY agent\_name, weekofyear(date)

13. average weekly resolution time for each agents

Ans. SELECT agent\_name,weekofyear(date),

avg(unix\_timestamp(average\_resoiution\_time,'HH:mm:ss')) AS avg\_resolution\_time

FROM agent\_csv

GROUP BY agent\_name, weekofyear(date)

14. Find the number of chat on which they have received a feedback

Ans.select count(total\_chats) from agent

where total\_feedback!=0

15. Total contribution for each and every agents weekly basis

Ans. SELECT agent,weekofyear(date),

sum(unix\_timestamp(Duration,'HH:mm:ss')) AS Total\_Duration

FROM agent\_login\_report

GROUP BY agent, weekofyear(date)

16. Perform inner join, left join and right join based on the agent column and after joining the table export that data into your local system.

Ans. insert overwrite local directory '/my/local/filesystem/directory/path'

Select \* from AgentLoggingReport alr join AgentPerformance ap on alr.Agent=ap.Agent Name

insert overwrite local directory '/my/local/filesystem/directory/path'

Select \* from AgentLoggingReport alr left join AgentPerformance ap on alr.Agent=ap.Agent Name

insert overwrite local directory '/my/local/filesystem/directory/path'

Select \* from AgentLoggingReport alr right join AgentPerformance ap on alr.Agent=ap.Agent Name

17. Perform partitioning on top of the agent column and then on top of that perform bucketing for each partitioning.

Ans. create table AgentReport (Date Timestamp,Login Time timestamp,Logout Time timestamp,duration timestamp) PARTITIONED BY(Agent string);

set hive.exec.dynamic.partition.mode=nonstrict

INSERT OVERWRITE TABLE AgentReport PARTITION(Agent)

SELECT Date,Duration,Login Time,Logout time from AgentReport;