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

```
# creating table
Hive> Create table AgentLogingReport
(
sr_no int,
Agent string,
Date date,
Login string,
Logout string,
Duration string
)
row format delimited
fields terminated by ','
tblproperties ("skip.header.line.count" = "1");
```

Hive> Create table AgentPerformance

(

sr_no int,

Date date,

Agent_Name string,

Total_charts string,

Avg_Response_Time string,

Avg_Resolution_Time string,

Avg_Rating float,

Total_Feedback int

)

row format delimited

fields terminated by ','

tblproperties ("skip.header.line.count" = "1");

hive> alter table AgentPerformance change Total_charts Total_charts int;

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

Load data local inpath'/home/cloudera/sidd/Challenge/mini _project_1/AgentLogingReport.csv' into table AgentLogingReport;

Load data local inpath'/home/cloudera/sidd/Challenge/mini _project_1/ AgentPerformance.csvinto table AgentPerformance;

```
hive> show databases;
OK
challenge
default
hive_class_b1
practice
Time taken: 2.138 seconds, Fetched: 4 row(s)
hive> use challenge;
OK
Time taken: 0.241 seconds
hive> show tables;
OK
agentlogingreport
agentperformance
air_quality
air_quality2
parking_violations_issued
vw_air_quality
vw_airquality
Time taken: 0.339 seconds, Fetched: 7 row(s)
hive> select * from agentlogingreport limit 5;
OK
1       Shivananda Sonwane      2022-07-30      15:35:29        17:39:39        2:04:10
2       Khushboo Priya  2022-07-30      15:06:59        15:07:16        0:00:17
3       Nandani Gupta   2022-07-30      15:04:24        17:31:07        2:26:42
4       Hrisikesh Neogi 2022-07-30      14:34:29        15:19:35        0:45:06
5       Mukesh  2022-07-30      14:03:15        15:11:52        1:08:36
Time taken: 3.112 seconds, Fetched: 5 row(s)
hive>
```

```
hive> select * from agentperformance limit 5;
OK
1       2022-07-30      Prerna Singh    11      12:00:38 AM     12:04:20 AM     4.11    9
2       2022-07-30      Nandani Gupta   11      12:01:15 AM     12:28:25 AM     3.14    7
3       2022-07-30      Ameya Jain      14      12:00:30 AM     12:11:36 AM     4.55    11
4       2022-07-30      Mahesh Sarade   14      12:01:04 AM     12:15:46 AM     4.71    7
5       2022-07-30      Swati   14      12:01:11 AM     12:16:33 AM     3.67    6
Time taken: 0.453 seconds, Fetched: 5 row(s)
hive>
```

3. List of all agents' names.

Hive> select distinct Agent_Name from AgentPerformance;

Hive> select count(distinct Agent_Name) from AgentPerformance;

```
hive> select distinct Agent_Name from AgentPerformance;
Query ID = cloudera_20220918034242_798df416-6a87-4368-95ee-a6956e49ed71
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 03:42:52,390 Stage-1 map = 0%,  reduce = 0%
2022-09-18 03:43:03,290 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.91 sec
2022-09-18 03:43:12,885 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.18 sec
MapReduce Total cumulative CPU time: 4 seconds 180 msec
Ended Job = job_1663476692610_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.18 sec   HDFS Read: 138418 HDFS Write: 867 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 180 msec
OK
Abhishek
Aditya
Aditya Shinde
Aditya_iot
Amersh
Ameya Jain
Anirudh
Ankit Sharma
Ankitjha
Anurag Tiwari
Aravind
Ashad Nasim
Ashish
Ayushi Mishra
Bharath
Boktiar Ahmed Bappy
Chaitra K Hiremath
Deepranjan Gupta
Dibyanshu
Harikrishnan Shaji
Hitesh Choudhary
Hrisikesh Neogi
```

```
hive> select count(distinct Agent_Name) from AgentPerformance;
Query ID = cloudera_20220918034444_9efe8dfb-3e4d-49d7-8b29-127db991497b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 03:44:28,806 Stage-1 map = 0%,  reduce = 0%
2022-09-18 03:44:37,498 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.06 sec
2022-09-18 03:44:48,214 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.24 sec
MapReduce Total cumulative CPU time: 4 seconds 240 msec
Ended Job = job_1663476692610_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.24 sec   HDFS Read: 139160 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 240 msec
OK
70
Time taken: 32.138 seconds, Fetched: 1 row(s)
hive>
```

4. Find out agent average rating.

Hive> select Agent_name,avg(Avg_Rating) from AgentPerformance group by Agent_name;

```
hive> select Agent_name,avg(Avg_Rating) from AgentPerformance group by Agent_name limt 10;
FAILED: ParseException line 1:76 missing EOF at 'limt' near 'Agent_name'
hive> select Agent_name,avg(Avg_Rating) from AgentPerformance group by Agent_name limit 10;
Query ID = cloudera_20220918060606_5fa77d09-33e6-4617-a528-5af36eb7f7af
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0028, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0028/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0028
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 06:07:06,306 Stage-1 map = 0%,  reduce = 0%
2022-09-18 06:07:20,925 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.98 sec
2022-09-18 06:07:40,017 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.02 sec
MapReduce Total cumulative CPU time: 11 seconds 20 msec
Ended Job = job_1663476692610_0028
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.02 sec   HDFS Read: 139870 HDFS Write: 239 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 20 msec
OK
agent_name      _c1
Abhishek        0.0
Aditya  0.0
Aditya Shinde   1.8003333409627278
Aditya_iot      2.3453333377838135
Amersh  0.0
Ameya Jain      2.21966667175293
Anirudh         0.6449999968210857
Ankit Sharma    0.0
Ankitjha        0.26666666666666666
Anurag Tiwari   0.18333333333333332
Time taken: 62.676 seconds, Fetched: 10 row(s)
hive>
```

5. Total working days for each agents

Hive> select Agent,count(distinct Date) from AgentLogingReport group by Agent;

```
hive> select Agent,count(distinct Date) from AgentLogingReport group by Agent limit 7;
Query ID = cloudera_20220918034646_cb6c8801-4615-45dd-b003-42683840c970
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 03:47:02,640 Stage-1 map = 0%,  reduce = 0%
2022-09-18 03:47:12,199 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.1 sec
2022-09-18 03:47:22,794 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.32 sec
MapReduce Total cumulative CPU time: 4 seconds 320 msec
Ended Job = job_1663476692610_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.32 sec   HDFS Read: 63682 HDFS Write: 89 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 320 msec
OK
Aditya Shinde   1
Aditya_iot      8
Amersh  2
Ameya Jain      7
Ankitjha        2
Anurag Tiwari   10
Aravind 7
Time taken: 33.863 seconds, Fetched: 7 row(s)
hive>
```

6. Total query that each agent have taken

Hive> select Agent_name,sum(total_chats) from AgentPerformance group by Agent_name;

```
hive> select Agent_name,sum(total_chats) from AgentPerformance group by Agent_name limit 7;
Query ID = cloudera_20220918063333_03d9c6df-8f48-4157-b39b-69e9d7958e09
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0032, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0032/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0032
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 06:33:47,242 Stage-1 map = 0%,  reduce = 0%
2022-09-18 06:33:56,809 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.94 sec
2022-09-18 06:34:05,272 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.15 sec
MapReduce Total cumulative CPU time: 4 seconds 150 msec
Ended Job = job_1663476692610_0032
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.15 sec   HDFS Read: 139250 HDFS Write: 93 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 150 msec
OK
agent_name      _c1
Abhishek        0
Aditya  0
Aditya Shinde   277
Aditya_iot      231
Amersh  0
Ameya Jain      322
Anirudh         81
Time taken: 30.185 seconds, Fetched: 7 row(s)
hive>
```

7. Total Feedback that each agent have received

Hive> select Agent_name,sum(Total_Feedback) from AgentPerformance group by Agent_name;

```
hive> select Agent_name,count(Total_Feedback) from AgentPerformance group by Agent limit 7;
FAILED: SemanticException [Error 10004]: Line 1:71 Invalid table alias or column reference 'Agent': (possible column names are: sr_no, date, agent_name, total_charts, avg_response_time, avg
_resolution_time, avg_rating, total_feedback)
hive> select Agent_name,sum(Total_Feedback) from AgentPerformance group by Agent_name limit 7;
Query ID = cloudera_20220918035252_e3d670a7-b290-472a-a0d7-29fc492945c0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 03:53:10,817 Stage-1 map = 0%,  reduce = 0%
2022-09-18 03:53:20,498 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.98 sec
2022-09-18 03:53:32,109 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.37 sec
MapReduce Total cumulative CPU time: 4 seconds 370 msec
Ended Job = job_1663476692610_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.37 sec   HDFS Read: 139404 HDFS Write: 93 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 370 msec
OK
Abhishek        0
Aditya  0
Aditya Shinde   153
Aditya_iot      131
Amersh  0
Ameya Jain      228
Anirudh         39
Time taken: 34.758 seconds, Fetched: 7 row(s)
```

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

Hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating between 3.5 and 4;

```
hive> set hive.cli.print.header = true;
hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating between 3.5 and 4 limit 10;
OK
agent_name      avg_rating
Swati   3.67
Manjunatha A    3.6
Boktiar Ahmed Bappy     4.0
Prateek _iot    3.75
Nandani Gupta   3.79
Jaydeep Dixit   3.95
Mahesh Sarade   3.94
Zeeshan         3.79
Hrisikesh Neogi 3.77
Muskan Garg     4.0
Time taken: 0.176 seconds, Fetched: 10 row(s)
hive>
```

9. Agent name who have rating less than 3.5

Hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating < 3.5;

```
hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating < 3.5 limit 10;
OK
agent_name      avg_rating
Nandani Gupta   3.14
Hitesh Choudhary        0.0
Sanjeevan       0.0
Anirudh         0.0
Shiva Srivastava        0.0
Dibyanshu       0.0
Ashish  0.0
Uday Mishra     0.0
Aditya Shinde   0.0
Jayant Kumar    0.0
Time taken: 0.108 seconds, Fetched: 10 row(s)
hive>
```

10. Agent name who have rating more than 4.5

Hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating > 4.5;

```
hive> select Agent_name,Avg_Rating from AgentPerformance where Avg_Rating > 4.5 limit 10;
OK
agent_name      avg_rating
Ameya Jain      4.55
Mahesh Sarade   4.71
Mukesh  4.62
Saikumarreddy N 5.0
Sanjeev Kumar   5.0
Harikrishnan Shaji      4.57
Sowmiya Sivakumar       4.75
Boktiar Ahmed Bappy     4.75
Shivananda Sonwane      5.0
Ishawant Kumar  4.67
Time taken: 0.09 seconds, Fetched: 10 row(s)
hive>
```

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

Hive> select Agent_name,avg(Total_Feedback) from AgentPerformance having avg(Total_Feedback) > 4.5;

```
hive> select Agent_name,avg(Total_Feedback) from AgentPerformance group by agent_name  having avg(Total_Feedback) > 4.5;
Query ID = cloudera_20220918065858_4fdafaed-4dda-441c-87e8-d046662099ac
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0034, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0034/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0034
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 06:58:31,895 Stage-1 map = 0%,  reduce = 0%
2022-09-18 06:58:40,802 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.34 sec
2022-09-18 06:58:54,752 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.01 sec
MapReduce Total cumulative CPU time: 6 seconds 10 msec
Ended Job = job_1663476692610_0034
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.01 sec   HDFS Read: 140122 HDFS Write: 865 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 10 msec
OK
agent_name      _c1
Aditya Shinde   5.1
Ameya Jain      7.6
Aravind         7.766666666666667
Ayushi Mishra   10.966666666666667
Bharath         8.233333333333333
Boktiar Ahmed Bappy     10.366666666666667
Deepranjan Gupta        10.4
Harikrishnan Shaji      7.7
Hrisikesh Neogi 12.233333333333333
Ishawant Kumar  6.733333333333333
Jawala Prakash  8.333333333333334
```

```
hive> select Agent_name,sum(Total_Feedback)as feedback from AgentPerformance where Avg_Rating > 4.5 group by agent_name limit 7;
Query ID = cloudera_20220918040909_b2c1daed-76b7-4ffc-b94c-2873cb2e3197
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 04:09:29,488 Stage-1 map = 0%,  reduce = 0%
2022-09-18 04:09:38,026 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.38 sec
2022-09-18 04:09:47,535 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.56 sec
MapReduce Total cumulative CPU time: 4 seconds 560 msec
Ended Job = job_1663476692610_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.56 sec   HDFS Read: 140230 HDFS Write: 100 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 560 msec
OK
agent_name      feedback
Aditya Shinde   73
Aditya_iot      43
Ameya Jain      150
Anirudh         2
Ankitjha        1
Aravind         178
Ayushi Mishra   75
Time taken: 27.449 seconds, Fetched: 7 row(s)
hive>
```

12. average weekly response time for each agent.

Hive> select s.agent_name,avg(col1[0]*3600+col1[1]*60+substr(col1[2],1,2))/3600 from(

select agent_name,split(Avg_Response_Time,':') as col1 from AgentPerformance)s group by s.agent_name;

```
hive> select s.agent_name,avg(col1[0]*3600+col1[1]*60+substr(col1[2],1,2))/3600  from(
    > select agent_name,split(Avg_Response_Time,':') as col1  from AgentPerformance )s group by s.agent_name;
Query ID = cloudera_20220918082828_68a4b5c2-58a8-4424-8b12-0af408bdc1a5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0042, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0042/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 08:29:16,401 Stage-1 map = 0%,  reduce = 0%
2022-09-18 08:29:37,919 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 9.15 sec
2022-09-18 08:30:01,911 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.98 sec
MapReduce Total cumulative CPU time: 15 seconds 980 msec
Ended Job = job_1663476692610_0042
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 15.98 sec   HDFS Read: 142891 HDFS Write: 1915 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 980 msec
OK
s.agent_name    _c1
Abhishek        12.0
Aditya  12.0
Aditya Shinde   12.008259259259258
Aditya_iot      12.009435185185186
Amersh  12.0
Ameya Jain      12.00587037037037
Anirudh         12.006046296296297
Ankit Sharma    12.0
Ankitjha        12.001231481481481
Anurag Tiwari   12.002342592592592
```

Hive> select agent_name,avg(Avg_Response_Time)as Avg_Response_Time,weekofyear(Date) as weekly from AgentPerformance group by agent_name,weekofyear(Date);

13. average weekly resolution time for each agents

Hive> select s.agent_name,avg(col1[0]*3600+col1[1]*60+substr(col1[2],1,2))/3600 from(

select agent_name,split(Avg_Resolution_Time,':') as col1 from AgentPerformance)s group by s.agent_name;

```
hive> select s.agent_name,avg(col1[0]*3600+col1[1]*60+substr(col1[2],1,2))/3600  from(
    > select agent_name,split(Avg_Resolution_Time,':') as col1  from AgentPerformance )s group by s.agent_name limit 5;
Query ID = cloudera_20220919014949_1fba2382-1351-4ef7-8186-788eabb58d87
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663564884018_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663564884018_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663564884018_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-19 01:49:44,148 Stage-1 map = 0%,  reduce = 0%
2022-09-19 01:50:04,370 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 8.41 sec
2022-09-19 01:50:28,333 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 16.64 sec
MapReduce Total cumulative CPU time: 16 seconds 640 msec
Ended Job = job_1663564884018_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 16.64 sec   HDFS Read: 143018 HDFS Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 640 msec
OK
s.agent_name    _c1
Abhishek        12.0
Aditya  12.0
Aditya Shinde   12.172398148148147
Aditya_iot      12.163694444444445
Amersh  12.0
Time taken: 68.532 seconds, Fetched: 5 row(s)
hive>
```

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

Hive> select agent_name,sum(Total_chats),Total_Feedback from AgentPerformance where Total_Feedback> 0 group by agent_name,Total_Feedback;

```
hive> select agent_name,sum(Total_charts),Total_Feedback from AgentPerformance where Total_Feedback> 0 group by agent_name,Total_Feedback limit 7;
Query ID = cloudera_20220918043939_e8f9d48d-b98d-4d06-9ebd-d1c0ba66f178
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663476692610_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663476692610_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663476692610_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-18 04:40:03,592 Stage-1 map = 0%,  reduce = 0%
2022-09-18 04:40:10,992 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.5 sec
2022-09-18 04:40:21,582 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.91 sec
MapReduce Total cumulative CPU time: 4 seconds 910 msec
Ended Job = job_1663476692610_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.91 sec   HDFS Read: 140028 HDFS Write: 136 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 910 msec
OK
agent_name      _c1     total_feedback
Aditya Shinde   8       7
Aditya Shinde   17      8
Aditya Shinde   67      9
Aditya Shinde   18      11
Aditya Shinde   27      14
Aditya Shinde   49      15
Aditya Shinde   34      19
Time taken: 29.962 seconds, Fetched: 7 row(s)
hive>
```


15. Total contribution hour for each and every agents weekly basis

Hive> select s.agent,sum(col1[0]*3600+col1[1]*60+col1[2])/3600 timeInHour,s.weekly from(

select agent,split(duration,':') as col1 ,weekofyear(Date) as weekly from AgentLogingReport)s group by s.agent,s.weekly limit 2;

```
hive> select s.agent,sum(coll[0]*3600+coll[1]*60+coll[2])/3600 timeInHour,s.weekly  from(
    > select agent,split(duration,':') as coll ,weekofyear(Date) as weekly from AgentLogingReport )s group by s.agent,s.weekly limit 2;
Query ID = cloudera_20220919012222_af45062d-cd3b-45c4-87fd-722cbf7bd882
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663564884018_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663564884018_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1663564884018_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-19 01:23:14,545 Stage-1 map = 0%,  reduce = 0%
2022-09-19 01:23:42,247 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.58 sec
2022-09-19 01:24:04,125 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.68 sec
MapReduce Total cumulative CPU time: 15 seconds 680 msec
Ended Job = job_1663564884018_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 15.68 sec   HDFS Read: 67252 HDFS Write: 69 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 680 msec
OK
s.agent timeinhour      s.weekly
Aditya Shinde   0.03611111111111111     30
Aditya_iot      6.095277777777778       29
Time taken: 76.566 seconds, Fetched: 2 row(s)
```

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.

Inner join:

hive -e 'select a.agent,a.date,a.Duration,b.Total_chats,b.Total_Feedback from challenge.AgentLogingReport a join challenge.AgentPerformance b on a.agent = b.agent_name' > /home/cloudera/sidd/Challenge/mini_project_1/inner_join.csv;

left join:

hive -e 'select a.agent,a.date,a.Duration,b.Total_chats,b.Total_Feedback from challenge.AgentLogingReport a left join challenge.AgentPerformance b on a.agent = b.agent_name' > /home/cloudera/sidd/Challenge/mini_project_1/left_join.csv;

left join with performance improved due to /*+ streamtable(a) */ hint:

hive -e 'select /*+ streamtable(a) */a.agent,a.date,a.Duration,b.Total_charts,b.Total_Feedback from challenge.AgentLogingReport a left join challenge.AgentPerformance b on a.agent = b.agent_name' > /home/cloudera/sidd/Challenge/mini_project_1/left_join.csv;

Right join:

hive -e 'select a.agent,a.date,a.Duration,b.Total_charts,b.Total_Feedback from challenge.AgentLogingReport a right join challenge.AgentPerformance b on a.agent = b.agent_name' > /home/cloudera/sidd/Challenge/mini_project_1/right_join.csv;

Right join with performance improved due to /*+ streamtable(a) */ hint:

hive -e 'select /*+ streamtable(a) */a.agent,a.date,a.Duration,b.Total_charts,b.Total_Feedback from challenge.AgentLogingReport a right join challenge.AgentPerformance b on a.agent = b.agent_name' > /home/cloudera/sidd/Challenge/mini_project_1/left_join.csv;

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

```
Create table AgentLogingReport_partitioned
(
sr_no int,
Date date,
Login string,
Logout string,
Duration string
)partitioned by (Agent string)
CLUSTERED BY (Date) sorted by (Date) INTO 4 BUCKETS
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';

hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into table AgentLogingReport_partitioned partition(Agent) select
sr_no,Date,Login,Logout,Duration,Agent from AgentLogingReport;
Hive> Create table AgentPerformance_partitioned
(
sr_no int,
Date date,
Total_charts string,
Avg_Response_Time string,
Avg_Resolution_Time string,
Avg_Rating float,
Total_Feedback int
)partitioned by (Agent_name string)
```

CLUSTERED BY (Date) sorted by (Date) INTO 8 BUCKETS

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ',';

Hive> insert into table AgentPerformance_partitioned partition(Agent_name) select sr_no,Date,Total_charts,Avg_Response_Time,Avg_Resolution_Time,Avg_Rating,Total_Feedback,Agent_name from AgentPerformance;