HIVE ASSIGNMENT 1

Table of Contents

DOWNLOAD AGENT DATA
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

Downloaded CSV Files
- AgentLogingReport.csv
- AgentPerformance.csv

CREATE A SCHEMA BASED ON THE GIVEN DATASET
> hive;
> use hive_sales_data_v1;
> use hive.cli.print.header=true;

CREATE TABLE FOR AGENTLOGING

```
create table agent_loging_report_data_csv
(
SERIAL_NO int,
AGENT string,
DATE string,
LOGIN_TIME string,
LOGOUT_TIME string,
DURATION string
)
row format delimited
fields terminated by ','
stored as textfile
tblproperties("skip.header.line.count"="1")
;
```

CREATE TABLE FOR AGENTPERFORMANCE

```
create table agent_performance_data_csv
(
SERIAL_NO int,
Date string,
AGENT_NAME string,
TOT_CHAT int,
AVG_RESPONSE_TIME string,
AVG_RESOLUTION_TIME string,
AVG_RATING float,
TOT_FEEDBACK int
)
row format delimited
fields terminated by ','
stored as textfile
tblproperties("skip.header.line.count"="1")
;
```

DUMP THE DATA INSIDE THE HDFS IN THE GIVEN SCHEMA LOCATION
LOAD DATA INTO TABLE "agent_loging_report_data_csv"
load data local inpath '/home/cloudera/sourav/Data/AgentLogingReport.csv' into table
agent_loging_report_data_csv

LOAD DATA INTO TABLE "agent_performance_data_csv"
load data local inpath '/home/cloudera/sourav/Data/AgentPerformance.csv' into table
agent_performance_data_csv

LIST OF ALL AGENTS' NAMES

```
with aperf as (
select DISTINCT AGENT_NAME
from agent_performance_data_csv),
alog as (
select DISTINCT AGENT
from agent_loging_report_data_csv),
all_agent as (
select aperf.AGENT_NAME AgentName
from aperf where lower(trim(aperf.AGENT_NAME)) NOT IN (
select lower(trim(alog.AGENT))
from alog)
union all
select alog.AGENT AgentName from alog)
select AgentNM as AgentName from all_agent
;
```

OUTPUT
AgentName
Aditya Shinde
Aditya_iot
Amersh
Ameya Jain
Ankitjha
Anurag Tiwari
Aravind
Ayushi Mishra
Bharath
Boktiar Ahmed Bappy
Chaitra K Hiremath
Deepranjan Gupta
Dibyanshu
Harikrishnan Shaji
Hrisikesh Neogi
Hyder Abbas
Ineuron Intelligence
Ishawant Kumar
Jawala Prakash
Jaydeep Dixit
Khushboo Priya
Madhulika G
Mahesh Sarade
Maitry
Manjunatha A
Mithun S
Mukesh
Muskan Garg
Nandani Gupta
Nishtha Jain
Nitin M

```
Prabir Kumar Satapathy
Prateek _iot
Prerna Singh
Rishav Dash
Saikumarreddy N
Sanjeev Kumar
Saurabh Shukla
Shiva Srivastava
Shivan K
Shivananda Sonwane
Shubham Sharma
Sowmiya Sivakumar
Sudhanshu Kumar
Suraj S Bilgi
Swati
Tarun
Wasim
Zeeshan
Abhishek
Aditya
Anirudh
Ankit Sharma
Ashad Nasim
Ashish
Hitesh Choudhary
Jayant Kumar
Mahak
Maneesh
Mukesh Rao
Rohan
Saif Khan
Samprit
Sandipan Saha
Sanjeevan
Shivan_S
Spuri
Uday Mishra
Vasanth P
Vivek
Time taken: 163.627 seconds, Fetched: 70 row(s)
```

FIND OUT AGENT AVERAGE RATING

select agent_name, round(avg(avg_rating), 2) average_rating
from agent_performance_data_csv group by agent_name;

OUTPUT	
agent_name	**average_rating**
Abhishek	0.0
Aditya	0.0
Aditya Shinde	1.8

Name	Col2	Col3
Aditya_iot	2.35	
Amersh		0.0
Ameya Jain	2.22	
Anirudh	0.64	
Ankit Sharma	0.0	
Ankitjha	0.27	
Anurag Tiwari	0.18	
Aravind	2.18	
Ashad Nasim	0.17	
Ashish	0.0	
Ayushi Mishra	3.48	
Bharath	2.98	
Boktiar Ahmed Bappy	3.57	
Chaitra K Hiremath	0.86	
Deepranjan Gupta	2.89	
Dibyanshu	0.0	
Harikrishnan Shaji	2.64	
Hitesh Choudhary	0.0	
Hrisikesh Neogi	3.14	
Hyder Abbas	0.0	
Ineuron Intelligence	0.0	
Ishawant Kumar		3.54
Jawala Prakash	3.47	
Jayant Kumar	1.07	
Jaydeep Dixit	3.17	
Khushboo Priya	3.7	
Madhulika G	3.5	
Mahak	0.1	
Mahesh Sarade	2.4	
Maitry	2.93	
Maneesh	0.17	
Manjunatha A	3.59	
Mithun S	2.36	
Mukesh		0.31
Mukesh Rao	0.26	
Muskan Garg	0.71	
Nandani Gupta	2.92	
Nishtha Jain	3.28	
Nitin M	0.0	
Prabir Kumar Satapathy	2.51	
Prateek _iot	2.44	
Prerna Singh	3.23	
Rishav Dash	1.43	
Rohan	0.0	
Saif Khan	0.0	
Saikumarreddy N		1.98
Samprit	0.0	
Sandipan Saha	0.43	
Sanjeev Kumar	3.38	
Sanjeevan	0.0	
Saurabh Shukla	0.56	
Shiva Srivastava	0.94	

Shivan K	2.84	
Shivan_S	0.14	
Shivananda Sonwane	4.23	
Shubham Sharma		3.23
Sowmiya Sivakumar	1.26	
Spuri	0.0	
Sudhanshu Kumar		0.33
Suraj S Bilgi	0.31	
Swati	2.42	
Tarun	0.05	
Uday Mishra	0.0	
Vasanth P	0.0	
Vivek	0.5	
Wasim	2.4	
Zeeshan	2.29	

Time taken: 26.776 seconds, Fetched: 70 row(s)

TOTAL WORKING DAYS FOR EACH AGENTS

```
with duration_in_sec as (
select agent,
date,
case
when duration  like '%:%:%' then
split(duration, ':')[0] * 3600 +
split(duration, ':')[1] * 60 +
split(duration, ':')[2]
else
duration
end as duration_sec
from agent_loging_report_data_csv),
valid_working_days as (
select agent, date, round(sum(duration_sec),0)
from duration_in_sec
group by agent, date
having round(sum(duration_sec),0) <= 86400
and round(sum(duration_sec),0) > 0)
select agent, count(date) Total_working_days
from valid_working_days
group by agent;
```

OUTPUT	
agent	**total_working_days**
Aditya Shinde	1
Aditya_iot	8
Amersh	2
Ameya Jain	7
Ankitjha	2
Anurag Tiwari	10
Aravind	7
Ayushi Mishra	9

```
Bharath                      8
Boktiar Ahmed Bappy          9
Chaitra K Hiremath           7
Deepranjan Gupta             10
Dibyanshu                    9
Harikrishnan Shaji           9
Hrisikesh Neogi              9
Hyder Abbas                  2
Ineuron Intelligence         1
Ishawant Kumar               11
Jawala Prakash               9
Jaydeep Dixit                6
Khushboo Priya               8
Madhulika G                  8
Mahesh Sarade                8
Maitry                       5
Manjunatha A                 7
Mithun S                     8
Mukesh                       2
Muskan Garg                  6
Nandani Gupta                9
Nishtha Jain                 8
Nitin M                      1
Prabir Kumar Satapathy       7
Prateek _iot                 11
Prerna Singh                 9
Rishav Dash                  7
Saikumarreddy N              7
Sanjeev Kumar                9
Saurabh Shukla               4
Shiva Srivastava             8
Shivan K                     8
Shivananda Sonwane           10
Shubham Sharma               11
Sowmiya Sivakumar            8
Sudhanshu Kumar              6
Suraj S Bilgi                2
Swati                        4
Wasim                        9
Zeeshan                      9
Time taken: 22.678 seconds, Fetched: 48 row(s)
```

/*
COUPLE OF OUTLIERS WITH DURATION MORE THAN 24 HOURS A DAY. HENCE, WE REJECTED 2 DATES FOR
COUPLE OF AGENTS "Jaydeep Dixit" AND "Tarun"
*/

agent	date	high_duration
Jaydeep Dixit	24-Jul-22	107192.0
Tarun	01-Jul-22	1850900.0

TOTAL QUERY THAT EACH AGENT HAVE TAKEN

```
select agent_name, sum(tot_chat ) total_query
from agent_performance_data_csv
group by agent_name;
```

OUTPUT	
agent_name	**total_query**
Abhishek	0
Aditya	0
Aditya Shinde	277
Aditya_iot	231
Amersh	0
Ameya Jain	322
Anirudh	81
Ankit Sharma	0
Ankitjha	5
Anurag Tiwari	4
Aravind	366
Ashad Nasim	18
Ashish	0
Ayushi Mishra	514
Bharath	369
Boktiar Ahmed Bappy	452
Chaitra K Hiremath	64
Deepranjan Gupta	493
Dibyanshu	1
Harikrishnan Shaji	381
Hitesh Choudhary	1
Hrisikesh Neogi	578
Hyder Abbas	0
Ineuron Intelligence	0
Ishawant Kumar	338
Jawala Prakash	439
Jayant Kumar	127
Jaydeep Dixit	512
Khushboo Priya	446
Madhulika G	469
Mahak	7
Mahesh Sarade	364
Maitry	542
Maneesh	4
Manjunatha A	413
Mithun S	503
Mukesh	19
Mukesh Rao	5
Muskan Garg	56
Nandani Gupta	560
Nishtha Jain	373
Nitin M	0
Prabir Kumar Satapathy	299
Prateek _iot	190

Prerna Singh	401
Rishav Dash	409
Rohan	0
Saif Khan	0
Saikumarreddy N	364
Samprit	1
Sandipan Saha	30
Sanjeev Kumar	507
Sanjeevan	0
Saurabh Shukla	16
Shiva Srivastava	53
Shivan K	357
Shivan_S	7
Shivananda Sonwane	441
Shubham Sharma	510
Sowmiya Sivakumar	206
Spuri	0
Sudhanshu Kumar	2
Suraj S Bilgi	28
Swati	524
Tarun	22
Uday Mishra	0
Vasanth P	0
Vivek	44
Wasim	433
Zeeshan	542

Time taken: 21.644 seconds, Fetched: 70 row(s)

TOTAL FEEDBACK THAT EACH AGENT HAVE RECEIVED

select agent_name, sum(tot_feedback) total_feedback
from agent_performance_data_csv
group by agent_name limit 10

OUTPUT

agent_name	total_feedback
Abhishek	0
Aditya	0
Aditya Shinde	153
Aditya_iot	131
Amersh	0
Ameya Jain	228
Anirudh	39
Ankit Sharma	0
Ankitjha	3
Anurag Tiwari	3
Aravind	233
Ashad Nasim	9
Ashish	0
Ayushi Mishra	329
Bharath	247
Boktiar Ahmed Bappy	311

Chaitra K Hiremath	37
Deepranjan Gupta	312
Dibyanshu	0
Harikrishnan Shaji	231
Hitesh Choudhary	0
Hrisikesh Neogi	367
Hyder Abbas	0
Ineuron Intelligence	0
Ishawant Kumar	202
Jawala Prakash	250
Jayant Kumar	70
Jaydeep Dixit	305
Khushboo Priya	289
Madhulika G	281
Mahak	5
Mahesh Sarade	216
Maitry	347
Maneesh	3
Manjunatha A	254
Mithun S	364
Mukesh	17
Mukesh Rao	5
Muskan Garg	37
Nandani Gupta	308
Nishtha Jain	257
Nitin M	0
Prabir Kumar Satapathy	222
Prateek _iot	107
Prerna Singh	235
Rishav Dash	264
Rohan	0
Saif Khan	0
Saikumarreddy N	290
Samprit	0
Sandipan Saha	18
Sanjeev Kumar	311
Sanjeevan	0
Saurabh Shukla	8
Shiva Srivastava	46
Shivan K	243
Shivan_S	4
Shivananda Sonwane	263
Shubham Sharma	300
Sowmiya Sivakumar	141
Spuri	0
Sudhanshu Kumar	2
Suraj S Bilgi	15
Swati	302
Tarun	6
Uday Mishra	0
Vasanth P	0
Vivek	20

Wasim	284
Zeeshan	335
Time taken: 20.494 seconds, Fetched: 70 row(s)	

AGENT NAMES WHO HAVE AVERAGE RATING BETWEEN 3.5 TO 4

select agent_name, round(avg(avg_rating),1) avg_rating
from agent_performance_data_csv
group by agent_name
having avg(avg_rating) >= 3.5 and avg(avg_rating) <= 4;

OUTPUT	
agent_name	**avg_rating**
Boktiar Ahmed Bappy	3.6
Ishawant Kumar	3.5
Khushboo Priya	3.7
Manjunatha A	3.6
Time taken: 21.813 seconds, Fetch: 4 row(s)	

AGENT NAMES WHO HAVE AVERAGE RATING LESS THAN 3.5

select agent_name, round(avg(avg_rating),1) avg_rating
from agent_performance_data_csv
group by agent_name
having avg(avg_rating) < 3.5;

OUTPUT	
agent_name	**avg_rating**
Abhishek	0.0
Aditya	0.0
Aditya Shinde	1.8
Aditya_iot	2.3
Amersh	0.0
Ameya Jain	2.2
Anirudh	0.6
Ankit Sharma	0.0
Ankitjha	0.3
Anurag Tiwari	0.2
Aravind	2.2
Ashad Nasim	0.2
Ashish	0.0
Ayushi Mishra	3.5
Bharath	3.0
Chaitra K Hiremath	0.9
Deepranjan Gupta	2.9
Dibyanshu	0.0
Harikrishnan Shaji	2.6
Hitesh Choudhary	0.0
Hrisikesh Neogi	3.1
Hyder Abbas	0.0
Ineuron Intelligence	0.0
Jawala Prakash	3.5

Jayant Kumar	1.1
Jaydeep Dixit	3.2
Madhulika G	3.5
Mahak	0.1
Mahesh Sarade	2.4
Maitry	2.9
Maneesh	0.2
Mithun S	2.4
Mukesh	0.3
Mukesh Rao	0.3
Muskan Garg	0.7
Nandani Gupta	2.9
Nishtha Jain	3.3
Nitin M	0.0
Prabir Kumar Satapathy	2.5
Prateek _iot	2.4
Prerna Singh	3.2
Rishav Dash	1.4
Rohan	0.0
Saif Khan	0.0
Saikumarreddy N	2.0
Samprit	0.0
Sandipan Saha	0.4
Sanjeev Kumar	3.4
Sanjeevan	0.0
Saurabh Shukla	0.6
Shiva Srivastava	0.9
Shivan K	2.8
Shivan_S	0.1
Shubham Sharma	3.2
Sowmiya Sivakumar	1.3
Spuri	0.0
Sudhanshu Kumar	0.3
Suraj S Bilgi	0.3
Swati	2.4
Tarun	0.1
Uday Mishra	0.0
Vasanth P	0.0
Vivek	0.5
Wasim	2.4
Zeeshan	2.3
Time taken: 21.605 seconds, Fetched: 65 row(s)	

AGENT NAMES WHO HAVE AVERAGE RATING MORE THAN 4.5