

Hive project -1

Name – Trijit Adhikary, Batch – 1

Table creation: -

```
hive (assignment2)> create table agent_login(  
    >     SL int,  
    >     name string,  
    >     date string,  
    >     login string,  
    >     logout string,  
    >     duration string  
    > )  
    > row format delimited  
    > fields terminated by ','  
    > tblproperties ("skip.header.line.count"="1");  
OK  
Time taken: 1.146 seconds
```

```
hive (assignment2)> create table agent_perf(  
    >     SL int,  
    >     date string,  
    >     name string,  
    >     total_chats int,  
    >     ave_response_time string,  
    >     ave_resolution_time string,  
    >     average_rating double,  
    >     total_feedback int  
    > )  
    > row format delimited  
    > fields terminated by ','  
    > tblproperties ("skip.header.line.count"="1");  
OK  
Time taken: 0.361 seconds
```

Loading the data: -

```
hive (assignment2)> load data local inpath '/home/cloudera/Desktop/trijit/hive/assignData/assignment2.0/AgentLoggingReport.csv' into table agent_login;  
Loading data to table assignment2.agent_login  
Table assignment2.agent_login stats: [numFiles=1, totalSize=55351]  
OK  
Time taken: 2.726 seconds  
hive (assignment2)> load data local inpath '/home/cloudera/Desktop/trijit/hive/assignData/assignment2.0/AgentPerformance.csv' into table agent_perf;  
Loading data to table assignment2.agent_perf  
Table assignment2.agent_perf stats: [numFiles=1, totalSize=109853]  
OK  
Time taken: 0.837 seconds
```

List of all agents' names: -

```
hive (assignment2)> select distinct name from agent_login;
```

```

name
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

```

Find out agent average rating: -

```
hive (assignment2)> select name,round(avg(average_rating),3) as avg_rating from agent_perf group by name;
```

name	avg_rating
Abhishek	0.0
Aditya	0.0
Aditya Shinde	1.8
Aditya_iot	2.345
Amersh	0.0
Ameya Jain	2.22
Anirudh	0.645
Ankit Sharma	0.0
Ankitjha	0.267
Anurag Tiwari	0.183
Aravind	2.181
Ashad Nasim	0.167
Ashish	0.0
Ayushi Mishra	3.482
Bharath	2.984
Boktiar Ahmed Bappy	3.568
Chaitra K Hiremath	0.865
Deepranjan Gupta	2.887
Dibyanshu	0.0
Harikrishnan Shaji	2.64
Hitesh Choudhary	0.0
Hrisikesh Neogi	3.136
Hyder Abbas	0.0
Ineuron Intelligence	0.0
Ishawant Kumar	3.543
Jawala Prakash	3.472
Jayant Kumar	1.069
Jaydeep Dixit	3.167
Khushboo Priya	3.704
Madhulika G	3.499
Mahak	0.1
Mahesh Sarade	2.4
Maitry	2.927
Maneesh	0.167
Manjunatha A	3.595

Mithun S	2.359
Mukesh	0.31
Mukesh Rao	0.256
Muskan Garg	0.712
Nandani Gupta	2.924
Nishtha Jain	3.282
Nitin M	0.0
Prabir Kumar Satapathy	2.51
Prateek_iot	2.438
Prerna Singh	3.233
Rishav Dash	1.427
Rohan	0.0
Saif Khan	0.0
Saikumarreddy N	1.98
Samprit	0.0
Sandipan Saha	0.429
Sanjeev Kumar	3.383
Sanjeevan	0.0
Saurabh Shukla	0.556
Shiva Srivastava	0.945
Shivan K	2.841
Shivan S	0.142
Shivananda Sonwane	4.233
Shubham Sharma	3.225
Sowmiya Sivakumar	1.26
Spuri	0.0
Sudhanshu Kumar	0.333
Suraj S Bilgi	0.312
Swati	2.424
Tarun	0.05
Uday Mishra	0.0
Vasanth P	0.0
Vivek	0.501
Wasim	2.4
Zeeshan	2.287

Total working days for each agent: -

```
hive (assignment2)> create view temp as select name,date,count(*) times_loggedIn from agent_login group by name,date;
```

```
hive (assignment2)> select name,count(*) from temp group by name;
```

name	c1
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	7
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
Tarun	1
Wasim	9
Zeeshan	9

Total query that each agent has taken: -

```
hive (assignment2)> select name,sum(total_chats) as total_query_taken from agent_perf group by name;
```

name	total_query_taken
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

Total Feedback that each agent has received: -

```
hive (assignment2)> select name,sum(total_feedback) as total_feedback_recived from agent_perf group by name;
```

name	total_feedback_recived
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

Agent name who have average rating between 3.5 to 4: -

```
hive (assignment2)> select name,avg_ratings
> from (select name,avg(average_rating) as avg_ratings from agent_perf group by name) as exact_avg_ratings
> where avg_ratings between 3.5 and 4;
Query ID = cloudera_20220919074646_f30cf883-9110-4acf-9255-eac3d4c21a84
```

```
name avg_ratings
Boktiar Ahmed Bappy 3.5679999999999999
Ishawant Kumar 3.5433333333333334
Khushboo Priya 3.7036666666666666
Manjunatha A 3.5946666666666665
Time taken: 46.732 seconds, Fetched: 4 row(s)
hive (assignment2)>
```

Agent name who have rating less than 3.5: -

```
hive (assignment2)> select name,avg_ratings
> from (select name,avg(average_rating) as avg_ratings from agent_perf group by name) as exact_avg_ratings
> where avg_ratings < 3.5;
Query ID = cloudera_20220919075454_fa139fa7-536d-4fc8-bc04-5965ff7a637a
```

name	_c1
Abhishek	0.0
Aditya	0.0
Aditya Shinde	1.8
Aditya_iot	2.35
Amersh	0.0
Ameya Jain	2.22
Anirudh	0.65
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
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
Jawala Prakash	3.47
Jayant Kumar	1.07
Jaydeep Dixit	3.17
Madhulika G	3.5
Mahak	0.1
Mahesh Sarade	2.4
Maitry	2.93
Maneesh	0.17
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
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

Agent name who have rating more than 4.5: -

```
hive (assignment2)> select name,round(avg_ratings,2)
> from (select name,avg(average_rating) as avg_ratings from agent_perf group by name) as exact_avg_ratings
> where avg_ratings > 4.5;
```

OK

```
name      _c1
Time taken: 96.74 seconds
hive (assignment2)> |
```

OR

```
hive (assignment2)> select name,max(round(average_rating,2)) from agent_perf where average_rating > 4.5 group by name;
```

name	_c1
Aditya Shinde	5.0
Aditya Iot	5.0
Ameya Jain	5.0
Anirudh	5.0
Ankitjha	5.0
Aravind	5.0
Ayushi Mishra	5.0
Bharath	5.0
Boktiar Ahmed Bappy	5.0
Chaitra K Hiremath	5.0
Deepranjan Gupta	5.0
Harikrishnan Shaji	4.89
Hrisikesh Neogi	4.91
Ishawant Kumar	5.0
Jawala Prakash	5.0
Jayant Kumar	5.0
Jaydeep Dixit	4.91
Khushboo Priya	5.0
Madhulika G	5.0
Mahesh Sarade	4.71
Maitry	5.0
Manjunatha A	5.0
Mithun S	4.8
Mukesh	4.67
Mukesh Rao	5.0
Muskan Garg	5.0
Nandani Gupta	4.8
Nishtha Jain	5.0
Prabir Kumar Satapathy	5.0
Prateek Iot	5.0
Prerna Singh	5.0
Rishav Dash	4.8
Saikumarreddy N	5.0
Sandipan Saha	5.0

Sanjeev Kumar	5.0
Saurabh Shukla	5.0
Shiva Srivastava	5.0
Shivan K	5.0
Shivananda Sonwane	5.0
Shubham Sharma	5.0
Sowmiya Sivakumar	4.75
Sudhanshu Kumar	5.0
Suraj S Bilgi	5.0
Swati	4.78
Vivek	5.0
Wasim	4.94
Zeeshan	4.85

How many feedback agents have received with more than 4.5 average: -?

```
hive (assignment2)> select name,max(round(average_rating,2)) as ratings,sum(total_feedback) as total_feedback_recived from agent_perf w
here average_rating > 4.5 group by name;
```

name	ratings	total_feedback_recived
Aditya Shinde	5.0	73
Aditya Iot	5.0	43
Ameya Jain	5.0	150
Anirudh	5.0	2
Ankitjha	5.0	1
Aravind	5.0	178
Ayushi Mishra	5.0	75
Bharath	5.0	231
Boktiar Ahmed Bappy	5.0	52
Chaitra K Hiremath	5.0	5
Deepranjan Gupta	5.0	83
Harikrishnan Shaji	4.89	80
Hrisikesh Neogi	4.91	183
Ishawant Kumar	5.0	79
Jawala Prakash	5.0	33
Jayant Kumar	5.0	7
Jaydeep Dixit	4.91	179
Khushboo Priya	5.0	134
Madhulika G	5.0	67
Mahesh Sarade	4.71	46
Maitry	5.0	81
Manjunatha A	5.0	132
Mithun S	4.8	93
Mukesh	4.67	17

Mukesh Rao	5.0	1
Muskan Garg	5.0	14
Nandani Gupta	4.8	91
Nishtha Jain	5.0	18
Prabir Kumar Satapathy	5.0	53
Prateek Iot	5.0	54
Prerna Singh	5.0	82
Rishav Dash	4.8	27
Saikumarreddy N	5.0	184
Sandipan Saha	5.0	1
Sanjeev Kumar	5.0	164
Saurabh Shukla	5.0	3
Shiva Srivastava	5.0	2
Shivan K	5.0	130
Shivananda Sonwane	5.0	154
Shubham Sharma	5.0	155
Sowmiya Sivakumar	4.75	16
Sudhanshu Kumar	5.0	2
Suraj S Bilgi	5.0	4
Swati	4.78	103
Vivek	5.0	3
Wasim	4.94	156
Zeeshan	4.85	48

average weekly response time for each agent: -

```
hive (assignment2)> select name,no_of_week,from_unixtime(cast(avg(unix_timestamp(ave_response_time,'HH:mm:ss'))as bigint),'HH:mm:ss')
> from(
> select name, weekofyear(from_unixtime(unix_timestamp(date,'MM/dd/yyyy'),'yyyy-MM-dd')) as no_of_week,ave_respon
se_time from agent_perf
> ) as t
> group by name,no_of_week;
```

name	no_of_week	c2
Abhishek	26	00:00:00
Abhishek	27	00:00:00
Abhishek	28	00:00:00
Abhishek	29	00:00:00
Abhishek	30	00:00:00
Aditya	26	00:00:00
Aditya	27	00:00:00
Aditya	28	00:00:00
Aditya	29	00:00:00
Aditya	30	00:00:00
Aditya Shinde	26	00:00:56
Aditya Shinde	27	00:00:48
Aditya Shinde	28	00:00:57
Aditya Shinde	29	00:00:00
Aditya Shinde	30	00:00:00
Aditya_iot	26	00:00:00
Aditya_iot	27	00:00:21
Aditya_iot	28	00:00:52
Aditya_iot	29	00:00:31
Aditya_iot	30	00:00:51
Amersh	26	00:00:00
Amersh	27	00:00:00
Amersh	28	00:00:00
Amersh	29	00:00:00
Amersh	30	00:00:00

Ameya Jain	26	00:00:00
Ameya Jain	27	00:00:00
Ameya Jain	28	00:00:33
Ameya Jain	29	00:00:30
Ameya Jain	30	00:00:33
Anirudh	26	00:00:00
Anirudh	27	00:01:18
Anirudh	28	00:00:16
Anirudh	29	00:00:00
Anirudh	30	00:00:00
Ankit Sharma	26	00:00:00
Ankit Sharma	27	00:00:00
Ankit Sharma	28	00:00:00
Ankit Sharma	29	00:00:00
Ankit Sharma	30	00:00:00
Ankitjha	26	00:00:00
Ankitjha	27	00:00:00
Ankitjha	28	00:00:00
Ankitjha	29	00:00:00
Ankitjha	30	00:00:23
Anurag Tiwari	26	00:00:00
Anurag Tiwari	27	00:00:37
Anurag Tiwari	28	00:00:00
Anurag Tiwari	29	00:00:00
Anurag Tiwari	30	00:00:00

Aravind	26	00:00:00
Aravind	27	00:00:21
Aravind	28	00:00:35
Aravind	29	00:00:37
Aravind	30	00:00:00
Ashad Nasim	26	00:00:00
Ashad Nasim	27	00:02:46
Ashad Nasim	28	00:00:00
Ashad Nasim	29	00:00:00
Ashad Nasim	30	00:00:00
Ashish	26	00:00:00
Ashish	27	00:00:00
Ashish	28	00:00:00
Ashish	29	00:00:00
Ashish	30	00:00:00
Ayushi Mishra	26	00:00:57
Ayushi Mishra	27	00:00:51
Ayushi Mishra	28	00:00:49
Ayushi Mishra	29	00:01:26
Ayushi Mishra	30	00:00:59
Bharath	26	00:00:28
Bharath	27	00:00:25
Bharath	28	00:00:14
Bharath	29	00:00:45
Bharath	30	00:00:23
Boktiar Ahmed Bappy	26	00:01:01
Boktiar Ahmed Bappy	27	00:00:43
Boktiar Ahmed Bappy	28	00:00:59
Boktiar Ahmed Bappy	29	00:01:07
Boktiar Ahmed Bappy	30	00:01:44
Chaitra K Hiremath	26	00:00:00
Chaitra K Hiremath	27	00:00:00
Chaitra K Hiremath	28	00:00:00
Chaitra K Hiremath	29	00:00:33
Chaitra K Hiremath	30	00:00:38

average weekly resolution time for each agent: -

```
select name,no_of_week,from_unixtime(cast(avg(unix_timestamp(ave_resolution_time,'HH:mm:ss'))as bigint),'HH:mm:ss')
from(
select name, weekofyear(from_unixtime(unix_timestamp(date,'MM/dd/yyyy'),'yyyy-MM-dd')) as no_of_week,ave_resolution_time from agent_perf
) as t
group by name,no_of_week;
```

name	no_of_week	c2
Abhishek	26	00:00:00
Abhishek	27	00:00:00
Abhishek	28	00:00:00
Abhishek	29	00:00:00
Abhishek	30	00:00:00
Aditya	26	00:00:00
Aditya	27	00:00:00
Aditya	28	00:00:00
Aditya	29	00:00:00
Aditya	30	00:00:00
Aditya Shinde	26	00:15:27
Aditya Shinde	27	00:15:04
Aditya Shinde	28	00:22:39
Aditya Shinde	29	00:00:00
Aditya Shinde	30	00:00:00
Aditya_iot	26	00:00:00
Aditya_iot	27	00:05:10
Aditya_iot	28	00:15:11
Aditya_iot	29	00:11:13
Aditya_iot	30	00:12:18
Amersh	26	00:00:00
Amersh	27	00:00:00
Amersh	28	00:00:00
Amersh	29	00:00:00
Amersh	30	00:00:00
Ameya Jain	26	00:00:00
Ameya Jain	27	00:00:16
Ameya Jain	28	00:08:04
Ameya Jain	29	00:08:18
Ameya Jain	30	00:08:00
Anirudh	26	00:00:00
Anirudh	27	00:08:08
Anirudh	28	00:05:07
Anirudh	29	00:00:00
Anirudh	30	00:00:00

Ankit Sharma	26	00:00:00
Ankit Sharma	27	00:00:00
Ankit Sharma	28	00:00:00
Ankit Sharma	29	00:00:00
Ankit Sharma	30	00:00:00
Ankitjha	26	00:00:00
Ankitjha	27	00:03:04
Ankitjha	28	00:00:00
Ankitjha	29	00:00:00
Ankitjha	30	00:00:59
Anurag Tiwari	26	00:00:00
Anurag Tiwari	27	00:05:17
Anurag Tiwari	28	00:00:00
Anurag Tiwari	29	00:00:00
Anurag Tiwari	30	00:00:00
Aravind	26	00:00:00
Aravind	27	00:10:23
Aravind	28	00:14:26
Aravind	29	00:13:28
Aravind	30	00:00:00
Ashad Nasim	26	00:00:00
Ashad Nasim	27	00:01:30
Ashad Nasim	28	00:00:00
Ashad Nasim	29	00:00:00
Ashad Nasim	30	00:00:00
Ashish	26	00:00:00
Ashish	27	00:00:00
Ashish	28	00:00:00
Ashish	29	00:00:00
Ashish	30	00:00:00
Ayushi Mishra	26	00:17:02
Ayushi Mishra	27	00:12:23
Ayushi Mishra	28	00:14:37
Ayushi Mishra	29	00:18:28
Ayushi Mishra	30	00:15:21

Bharath	26	00:06:29
Bharath	27	00:12:22
Bharath	28	00:07:22
Bharath	29	00:09:25
Bharath	30	00:16:45
Boktiar Ahmed Bappy	26	00:15:01
Boktiar Ahmed Bappy	27	00:33:09
Boktiar Ahmed Bappy	28	00:06:49
Boktiar Ahmed Bappy	29	00:11:36
Boktiar Ahmed Bappy	30	00:17:41
Chaitra K Hiremath	26	00:00:00
Chaitra K Hiremath	27	00:00:00
Chaitra K Hiremath	28	00:00:00
Chaitra K Hiremath	29	00:00:43
Chaitra K Hiremath	30	00:06:35
Deepranjan Gupta	26	00:20:08
Deepranjan Gupta	27	00:25:49
Deepranjan Gupta	28	00:16:32
Deepranjan Gupta	29	00:22:00
Deepranjan Gupta	30	00:19:09
Dibyanshu	26	00:00:00
Dibyanshu	27	00:01:46
Dibyanshu	28	00:00:00
Dibyanshu	29	00:00:00
Dibyanshu	30	00:00:00
Harikrishnan Shaji	26	00:00:00
Harikrishnan Shaji	27	00:06:00
Harikrishnan Shaji	28	00:12:43
Harikrishnan Shaji	29	00:17:51
Harikrishnan Shaji	30	00:13:54
Hitesh Choudhary	26	00:00:00
Hitesh Choudhary	27	00:00:00
Hitesh Choudhary	28	00:00:13
Hitesh Choudhary	29	00:00:00
Hitesh Choudhary	30	00:00:00

Find the number of chat on which they have received a feedback: -

```
hive (assignment2)> select name,sum(total_feedback) from agent_perf group by name;
```

name	total_feedback_recived
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

Total contribution hour for each and every agent weekly basis: -

```
import sys

for line in sys.stdin:
    temp = line.strip("\n\r")
    name,date,login,logout= temp.split('\t')
    #date = temp
    month_index = ['test','jan','feb','mar','apr','may','jun','jul','aug','sep','oct','nov','dec']
    #date = '30-Oct-22'
    year = str(2000 + int(date.split('-')[2]))
    month = date.split('-')[1].lower()
    day = str(date.split('-')[0])
    #print(year + '-' + str(month_index.index(month)) + '-' + day)
    formattedDate = year + '-' + str(month_index.index(month)) + '-' + day
    result = '\t'.join([name,formattedDate,login,logout])
    print(result)
```

```
hive (assignment2)> create table agent_login_parse as select transform(name,date,login,logout) using 'python date_formatter.py' as (name
, date, login,logout) from agent_login;
```

```
hive (assignment2)> select name,weekofyear(date) weekNo,
> ((sum((unix_timestamp(logout,'HH:mm:ss') - unix_timestamp(login,'HH:mm:ss')))/60)/60) duration
> from agent_login_parse group by name,weekofyear(date) order by name, weekNo;
```

name	weekno	duration
Aditya Shinde	30	0.03611111111111111
Aditya_iot	29	6.095555555555555
Aditya_iot	30	9.637222222222222
Amersh	30	3.064444444444444
Ameya Jain	29	0.08361111111111111
Ameya Jain	30	17.993055555555554
Ankitjha	30	2.2675
Anurag Tiwari	29	0.26555555555555554
Anurag Tiwari	30	2.5191666666666666
Aravind	29	24.236666666666668
Aravind	30	0.06361111111111112
Ayushi Mishra	29	-54.209444444444444
Ayushi Mishra	30	-51.667222222222222
Bharath	29	24.071666666666665
Bharath	30	24.006111111111111
Boktiar Ahmed Bappy	29	-54.249166666666666
Boktiar Ahmed Bappy	30	-49.479166666666664
Chaitra K Hiremath	29	2.235
Chaitra K Hiremath	30	32.092222222222222
Deepranjan Gupta	29	-71.000833333333333
Deepranjan Gupta	30	-62.716111111111111
Dibyanshu	29	-44.245833333333333
Dibyanshu	30	-47.141388888888888
Harikrishnan Shaji	29	21.438888888888886
Harikrishnan Shaji	30	32.278611111111111
Hrisikesh Neogi	29	2.8936111111111114
Hrisikesh Neogi	30	6.680833333333334
Hyder Abbas	29	0.33611111111111114
Hyder Abbas	30	0.05222222222222222
Ineuron Intelligence	29	1.4488888888888889
Ishawant Kumar	29	-70.276944444444445
Ishawant Kumar	30	-45.939722222222222
Jawala Prakash	29	-71.658888888888889
Jawala Prakash	30	-73.936388888888889
Jaydeep Dixit	29	17.915000000000003
Jaydeep Dixit	30	17.926388888888887
Khushboo Priya	29	21.717499999999998
Khushboo Priya	30	21.844166666666667
Madhulika G	29	25.852222222222224
Madhulika G	30	20.209722222222222
Mahesh Sarade	29	25.483333333333334
Mahesh Sarade	30	-6.558888888888889
Maitry	29	-71.341944444444444
Maitry	30	-17.7125
Manjunatha A	29	18.351666666666667
Manjunatha A	30	22.924722222222222
Mithun S	29	17.380555555555553
Mithun S	30	27.794444444444444
Mukesh	30	8.905555555555557
Muskan Garg	29	3.319444444444444
Muskan Garg	30	14.018611111111111
Nandani Gupta	29	17.334166666666665
Nandani Gupta	30	22.839444444444442
Nishtha Jain	29	22.116666666666667
Nishtha Jain	30	21.7375
Nitin M	29	0.7988888888888889
Prabir Kumar Satapathy	29	17.526111111111111
Prabir Kumar Satapathy	30	15.854722222222222
Prateek_iot	29	7.27
Prateek_iot	30	11.149444444444445
Prerna Singh	29	18.517777777777777
Prerna Singh	30	27.199444444444445
Rishav Dash	29	18.894722222222224
Rishav Dash	30	22.882777777777778
Saikumarreddy N	29	24.981666666666667
Saikumarreddy N	30	18.157500000000002
Sanjeev Kumar	29	19.361944444444444
Sanjeev Kumar	30	25.327222222222225
Saurabh Shukla	29	-7.335277777777778
Shiva Srivastava	29	1.9066666666666667
Shiva Srivastava	30	13.089722222222223
Shivan K	29	16.714444444444446
Shivan K	30	19.391944444444444
Shivananda Sonwane	29	-75.164722222222222
Shivananda Sonwane	30	-67.544444444444444
Shubham Sharma	29	-17.486388888888889
Shubham Sharma	30	-0.7108333333333333
Sowmiya Sivakumar	29	17.066388888888888
Sowmiya Sivakumar	30	27.689722222222223
Sudhanshu Kumar	29	0.45583333333333337
Sudhanshu Kumar	30	-2.2233333333333336
Suraj S Bilgi	30	12.56
Swati	29	18.858888888888888
Swati	30	6.1425
Tarun	26	10.138888888888889
Wasim	29	19.626666666666665
Wasim	30	28.553888888888889
Zeeshan	29	24.428611111111113
Zeeshan	30	24.661944444444444

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: -

```
hive (assignment2)> create table inner_join_result
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> location '/trijit/hive/output'
> as
> select al.name,al.date,al.duration,ap.average_rating from agent_perf ap join agent_login al on (ap.name = al.name);
```

```
[cloudera@quickstart assignment2.0]$ hdfs dfs -ls /trijit/hive/output
Found 1 items
-rwxr-xr-x 1 cloudera supergroup 814260 2022-09-20 22:56 /trijit/hive/output/000000_0
[cloudera@quickstart assignment2.0]$ hdfs dfs -cat /trijit/hive/output/000000_0 > /home/cloudera/Desktop/trijit/hive/assignData/assignment2.0/output.csv
[cloudera@quickstart assignment2.0]$ ls
AgentLoggingReport.csv AgentPerformance.csv output.csv
[cloudera@quickstart assignment2.0]$
```

```
[cloudera@quickstart assignment2.0]$ head -20 output.csv
Prerna Singh,30-Jul-22,01:37:40,4.11
Prerna Singh,29-Jul-22,03:16:37,4.11
Prerna Singh,29-Jul-22,02:12:27,4.11
Prerna Singh,29-Jul-22,00:03:11,4.11
Prerna Singh,27-Jul-22,07:47:29,4.11
Prerna Singh,26-Jul-22,00:02:00,4.11
Prerna Singh,26-Jul-22,00:07:29,4.11
Prerna Singh,26-Jul-22,00:08:30,4.11
Prerna Singh,26-Jul-22,00:01:56,4.11
Prerna Singh,26-Jul-22,05:56:38,4.11
Prerna Singh,25-Jul-22,02:35:22,4.11
Prerna Singh,25-Jul-22,03:22:34,4.11
Prerna Singh,23-Jul-22,06:06:06,4.11
Prerna Singh,22-Jul-22,06:03:51,4.11
Prerna Singh,21-Jul-22,00:03:38,4.11
Prerna Singh,21-Jul-22,00:15:17,4.11
Prerna Singh,20-Jul-22,05:59:11,4.11
Prerna Singh,20-Jul-22,00:03:00,4.11
Nandani Gupta,30-Jul-22,02:26:42,3.14
Nandani Gupta,29-Jul-22,05:21:17,3.14
```

```
hive (assignment2)> select al.name,al.date,al.duration,ap.average_rating from agent_perf ap left join agent_login al on (ap.name = al.name);
```

```
hive (assignment2)> select al.name,al.date,al.duration,ap.average_rating from agent_perf ap right join agent_login al on (ap.name = al.name);
```

Perform partitioning on top of the agent column and then on top of that perform bucketing for each partitioning: -

```
set hive.enforce.bucketing=true
set hive.exec.dynamic.partition.mode=nonstrict
create table name_partition(
  sl int,
  date string,
  login string,
  logout string,
  duration string
)
partition by (name string)
clustered by (date) into 3 buckets;
```

What I have learnt in the live class is,

Partitioning should be done on the less divers columns and bucketing should be done on very divers columns.

Being said that in my opinion if we partition the data by the date and bucket the data by name it will be more efficient.

```
insert into table name_partition partition(name) select sl,date,login,logout,duration, name from agent_login;
```


