Module 3: Assignment – 3

Creating Table And Schema

```
hive> create table fbdata
    > (userid int, age int, dob_day int, dob_year int, dob_month int, gender string, tenure int,
    > friend_count int, friendships_initiated int, likes int, likes_received int, mobile_likes int,
    > mobile_likes_received int, www_likes int, www_likes_received int)
    > row format delimited fields terminated by ',' stored as textfile
    > tblproperties("skip.header.line.count"="1");
OK
Time taken: 1.338 seconds
```

Loading data into table

```
hive> load data local inpath '/home/bitnami/pseudo_facebook.csv' into table fbdata;
Loading data to table aug.fbdata
OK
Time taken: 1.322 seconds
```

Viewing first few records

```
hive> select * from fbdata limit 10;
OK
2094382 14      19      1999    11      male    266     0       0       0       0       0       0       0       0
1192601 14      2       1999    11      female  6       0       0       0       0       0       0       0       0
2083884 14      16      1999    11      male    13      0       0       0       0       0       0       0       0
1203168 14      25      1999    12      female  93      0       0       0       0       0       0       0       0
1733186 14      4       1999    12      male    82      0       0       0       0       0       0       0       0
1524765 14      1       1999    12      male    15      0       0       0       0       0       0       0       0
1136133 13      14      2000    1       male    12      0       0       0       0       0       0       0       0
1680361 13      4       2000    1       female  0       0       0       0       0       0       0       0       0
1365174 13      1       2000    1       male    81      0       0       0       0       0       0       0       0
1712567 13      2       2000    2       male    171     0       0       0       0       0       0       0       0
Time taken: 0.253 seconds, Fetched: 10 row(s)
```

1. What is the average age of a social media account user?

```
hive> select avg(age) from fbdata;
Automatically selecting local only mode for query
Query ID = root_20211104135541_db10640c-1512-418e-8b1e-86ef2d1ec028
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>
Job running in-process (local Hadoop)
2021-11-04 13:55:43,287 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1684230484_0073
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1165706646 HDFS Write: 25639333 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
37.28022383160106
Time taken: 1.79 seconds, Fetched: 1 row(s)
```

2. Does the social media platform have a higher number of male users, female users, or gender undisclosed users?

```
hive> select gender, count(gender) from fbdata group by gender;
Automatically selecting local only mode for query
Query ID = root_20211104135830_f17d090a-74bd-4248-98d8-da5c4dd1144f
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>
Job running in-process (local Hadoop)
2021-11-04 13:58:32,102 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local95631398_0076
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1197009512 HDFS Write: 25640152 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
NA      175
female  40254
male    58574
Time taken: 2.067 seconds, Fetched: 3 row(s)
```

From above statistical analysis, it can be inferred that:
the social media platform has a higher number of “Male” users.

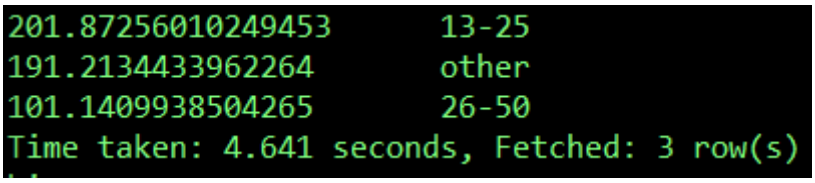
3. In male users, on average, does the age demographic of 13–25 have more, or less, friends than the demographic of 26–50? Assess this with appropriate statistical reasoning

```
hive> select avg(friend_count) as avg_frnd_count,
    > case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=50 then '26-50'
    > else 'other'
    > end as age_group
    > from fbdata
    > where gender='male'
    > group by case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=50 then '26-50'
    > else 'other'
    > end
    > sort by avg_frnd_count desc
    > limit 3;
```

Running Map Reduce Job to view output:

```
Automatically selecting local only mode for query
Query ID = root_20211104145316_dc2b821b-d538-4624-99c9-b306e085ce83
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 14:53:18,062 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1588148921_0109
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 14:53:19,894 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local241999394_0110
Launching Job 3 out of 3
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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 14:53:21,342 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local1360015663_0111
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1407936284 HDFS Write: 27883903 SUCCESS
Stage-Stage-2:  HDFS Read: 1407946157 HDFS Write: 27896733 SUCCESS
Stage-Stage-3:  HDFS Read: 1407960029 HDFS Write: 27911270 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:



From above statistical analysis, it can be inferred that “Male” users of age group 13-25, on average have more friends than “Male” users of age group 26-50.

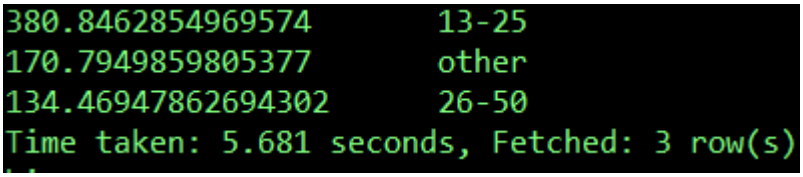
4. In female users, on average, does the age demographic of 13–25 have more, or less, friends than the demographic of 26–50? Assess this with appropriate statistical reasoning

```
hive> select avg(friend_count) as avg_frnd_count,
    > case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=50 then '26-50'
    > else 'other'
    > end as age_group
    > from fbdata
    > where gender='female'
    > group by case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=50 then '26-50'
    > else 'other'
    > end
    > sort by avg_frnd_count desc
    > limit 3;
```

Running Map Reduce Job to view output:

```
Automatically selecting local only mode for query
Query ID = root_20211104145711_1c6cf6b0-d9d9-4402-924f-7480c652eed3
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 14:57:12,694 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local665527973_0112
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 14:57:14,533 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local145109624_0113
Launching Job 3 out of 3
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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 14:57:16,803 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local417200130_0114
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1418398448 HDFS Write: 27911655 SUCCESS
Stage-Stage-2:  HDFS Read: 1418408321 HDFS Write: 27924485 SUCCESS
Stage-Stage-3:  HDFS Read: 1418422193 HDFS Write: 27939022 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:



From above statistical analysis, it can be inferred that “Female” users of age group 13-25, on average have more friends than “Female” users of age group 26-50.

5. Which gender is more likely to send out a higher number of friend requests on average?

```
hive> select gender, avg(friendships_initiated) as requests_sent
    > from fbdata
    > group by gender
    > sort by requests_sent desc
    > limit 2;
```

Running Map Reduce Job to generate output:

```
Automatically selecting local only mode for query
Query ID = root_20211104150453_55ca4308-437c-45d7-b556-284c9d6a953e
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 15:04:54,432 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local443781338_0116
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 15:04:56,288 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local1207018758_0117
Launching Job 3 out of 3
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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 15:04:58,544 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local1804736902_0118
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1439303140 HDFS Write: 27939714 SUCCESS
Stage-Stage-2:  HDFS Read: 1439313245 HDFS Write: 27952840 SUCCESS
Stage-Stage-3:  HDFS Read: 1439327134 HDFS Write: 27967293 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:

```
female  113.89909077358772
male    103.06659951514324
Time taken: 5.511 seconds, Fetched: 2 row(s)
```

From above statistical analysis, it can be inferred that users of gender “Female” are more likely to send friend requests on average compared to “Male” users.

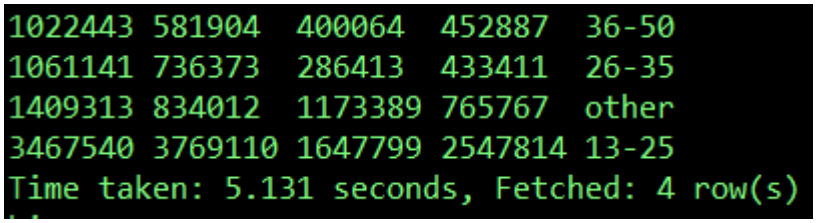
6. With the age demographics of 13–25, 26–35, and 36–50 as focal points, evaluate the comparison between mobile application usage and web browser usage when accessing the social media website. Use this to determine if mobile phones have indeed taken over the digital market space.

```
hive> select sum(mobile_likes) as mob_likes,
    > sum(mobile_likes_received) as mob_likes_rec,
    > sum(www_likes) as browser_likes,
    > sum(www_likes_received) browser_likes_rec,
    > case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=35 then '26-35'
    > when age >=36 and age <=50 then '36-50'
    > else 'other'
    > end as age_group
    > from fbdata
    > where gender='female'
    > group by case
    > when age >=13 and age <=25 then '13-25'
    > when age >=26 and age <=35 then '26-35'
    > when age >=36 and age <=50 then '36-50'
    > else 'other'
    > end
    > sort by mob_likes,browser_likes,mob_likes_rec,browser_likes_rec desc
    > limit 50;
```

Running Map Reduce Job to generate output:

```
Automatically selecting local only mode for query
Query ID = root_20211104151919_cf91babf-af38-4437-97af-517c74497a12
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 15:19:20,907 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1867722967_0124
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 15:19:22,380 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local748279575_0125
Launching Job 3 out of 3
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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 15:19:24,652 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local1730571489_0126
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1481101996 HDFS Write: 28001756 SUCCESS
Stage-Stage-2:  HDFS Read: 1481113689 HDFS Write: 28017132 SUCCESS
Stage-Stage-3:  HDFS Read: 1481130153 HDFS Write: 28034327 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:



From above statistical analysis, it can be inferred that users of every age group bet it 13-25 or 26-35 or 36-50 or other, are all making use of mobile application to browse and interact on social media platform.

It is safe to say that use of mobile platforms for social media has indeed taken over digital marketing space.