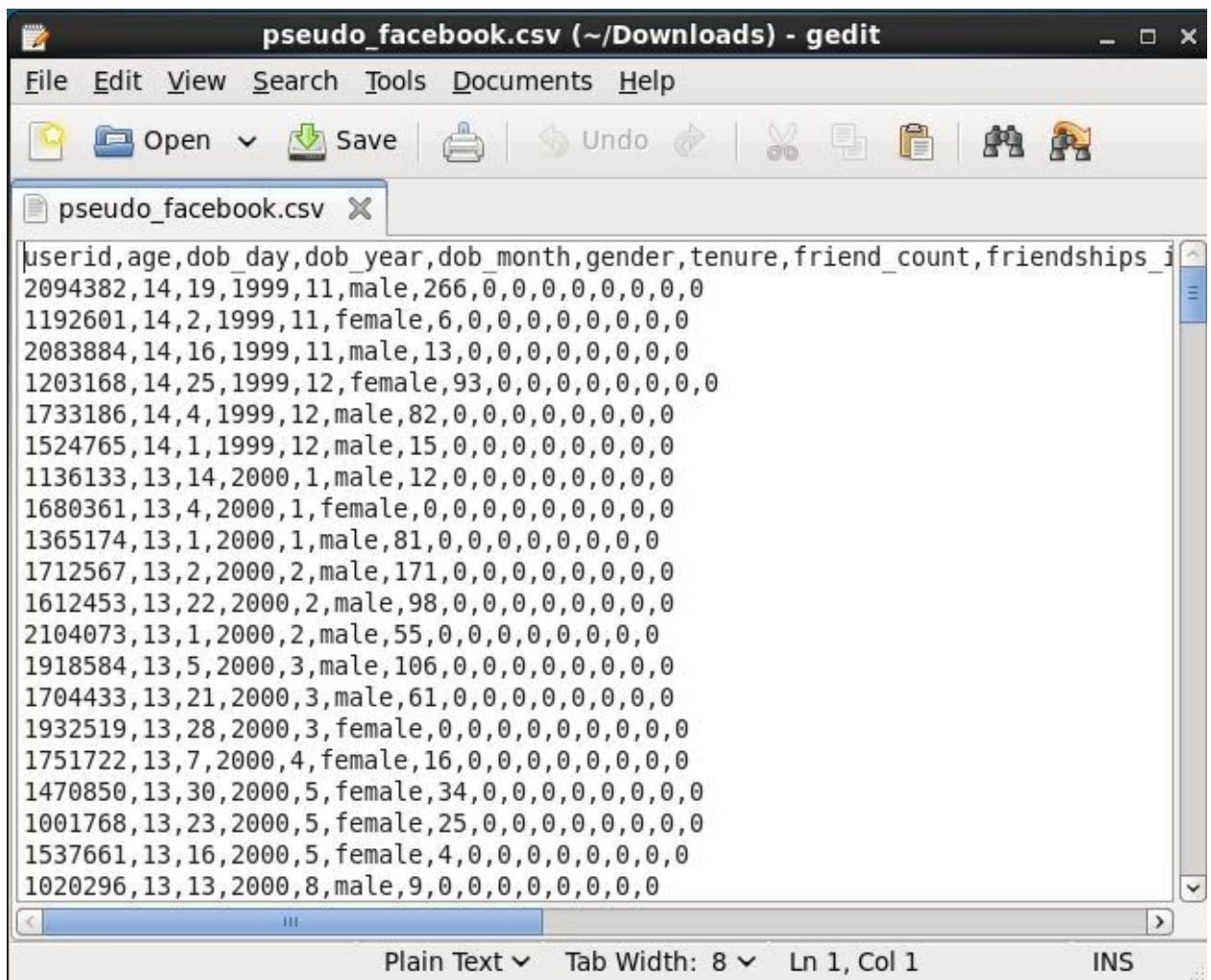
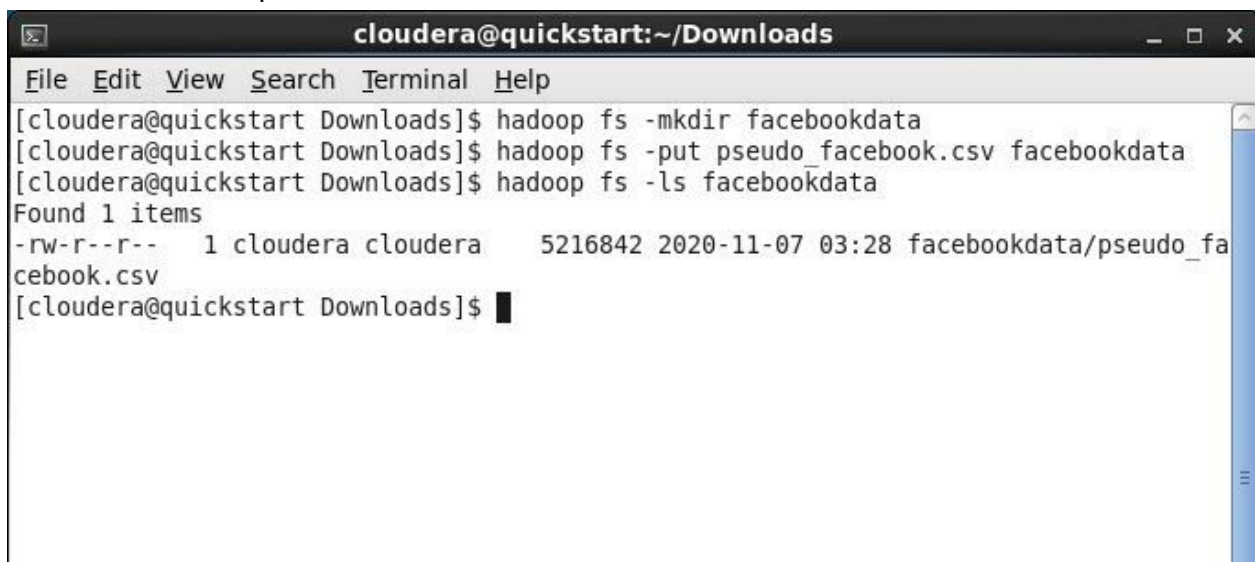


1. Data



```
pseudo_facebook.csv
userid,age,dob_day,dob_year,dob_month,gender,tenure,friend_count,friendships_i
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
1612453,13,22,2000,2,male,98,0,0,0,0,0,0,0,0
2104073,13,1,2000,2,male,55,0,0,0,0,0,0,0,0
1918584,13,5,2000,3,male,106,0,0,0,0,0,0,0,0
1704433,13,21,2000,3,male,61,0,0,0,0,0,0,0,0
1932519,13,28,2000,3,female,0,0,0,0,0,0,0,0,0
1751722,13,7,2000,4,female,16,0,0,0,0,0,0,0,0
1470850,13,30,2000,5,female,34,0,0,0,0,0,0,0,0
1001768,13,23,2000,5,female,25,0,0,0,0,0,0,0,0
1537661,13,16,2000,5,female,4,0,0,0,0,0,0,0,0
1020296,13,13,2000,8,male,9,0,0,0,0,0,0,0,0
```

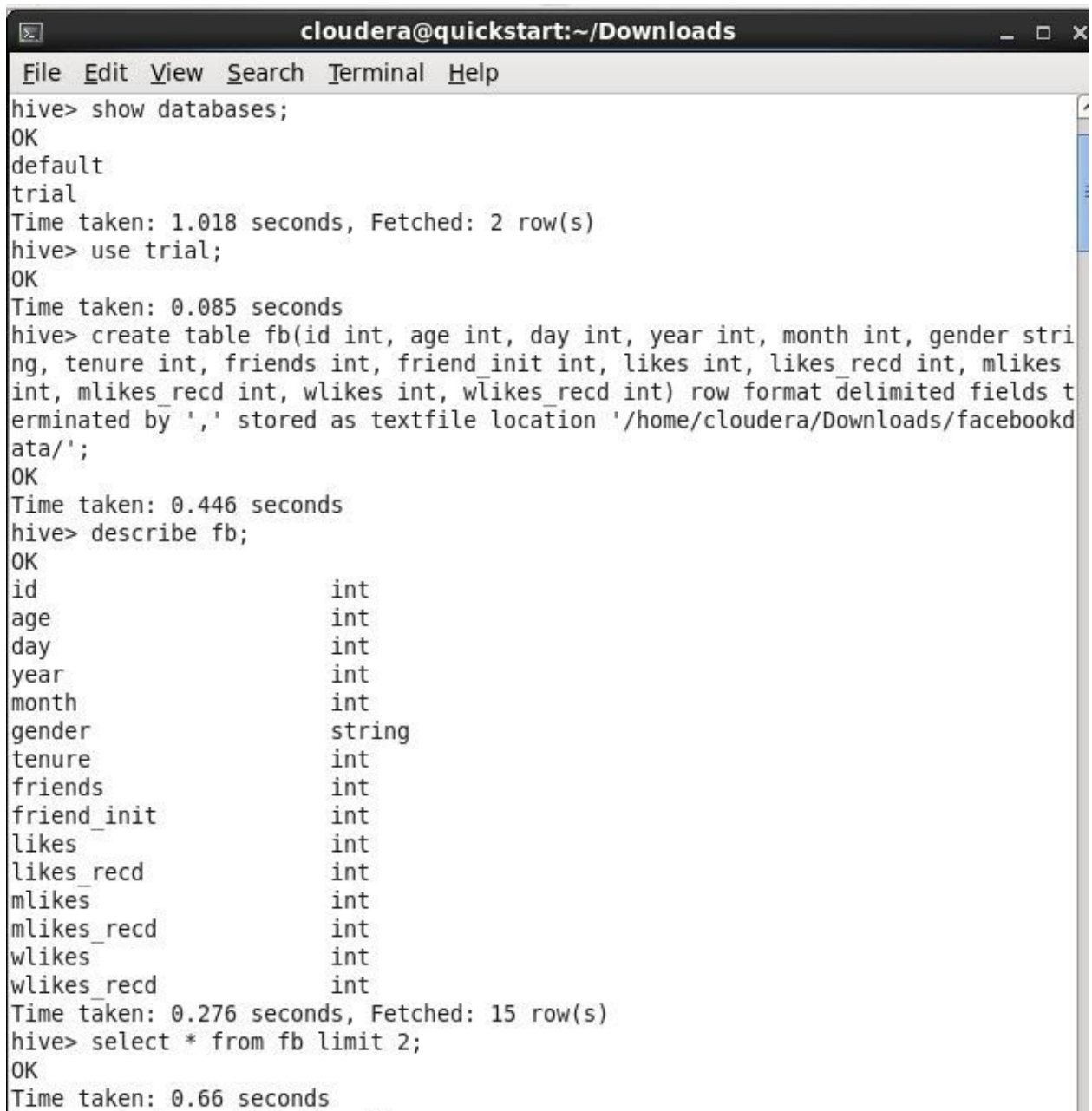
2. Load data in Hadoop:



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ hadoop fs -mkdir facebookdata
[cloudera@quickstart Downloads]$ hadoop fs -put pseudo_facebook.csv facebookdata
[cloudera@quickstart Downloads]$ hadoop fs -ls facebookdata
Found 1 items
-rw-r--r-- 1 cloudera cloudera 5216842 2020-11-07 03:28 facebookdata/pseudo_fa
cebook.csv
[cloudera@quickstart Downloads]$
```

```
hive> load data local inpath '/home/cloudera/Downloads/pseudo_facebook.csv/' into table fb;
Loading data to table trial.fb
Table trial.fb stats: [numFiles=1, totalSize=5216842]
OK
Time taken: 0.782 seconds
```

3. Create database and table in hive to store facebook data:



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> show databases;
OK
default
trial
Time taken: 1.018 seconds, Fetched: 2 row(s)
hive> use trial;
OK
Time taken: 0.085 seconds
hive> create table fb(id int, age int, day int, year int, month int, gender string, tenure int, friends int, friend_init int, likes int, likes_recd int, mlikes int, mlikes_recd int, wlikes int, wlikes_recd int) row format delimited fields terminated by ',' stored as textfile location '/home/cloudera/Downloads/facebookdata/';
OK
Time taken: 0.446 seconds
hive> describe fb;
OK
id                int
age               int
day               int
year              int
month             int
gender            string
tenure            int
friends           int
friend_init       int
likes             int
likes_recd        int
mlikes            int
mlikes_recd       int
wlikes            int
wlikes_recd       int
Time taken: 0.276 seconds, Fetched: 15 row(s)
hive> select * from fb limit 2;
OK
Time taken: 0.66 seconds
```

4. Data from table:

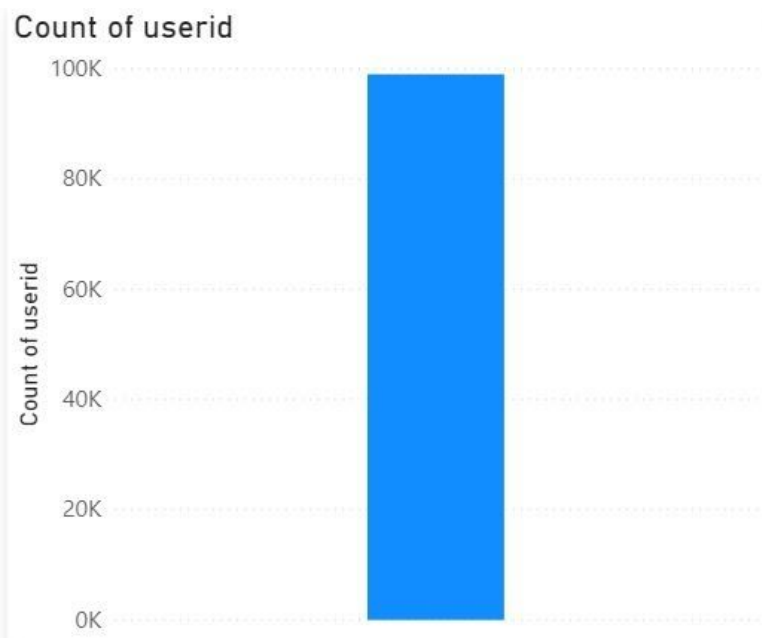
```
hive> select * from fb limit 2;
OK
2094382 14      19      1999    11      male    266      0      0      0      0
0        0        0        0
1192601 14      2       1999    11      female   6        0      0      0      0
0        0        0        0
Time taken: 0.122 seconds, Fetched: 2 row(s)
hive> select count(*) from fb where age>25;
```

```
hive> describe fb;
OK
id                int
age               int
day              int
year             int
month            int
gender            string
tenure            int
friends           int
friend_init       int
likes             int
likes_recd        int
mlikes            int
mlikes_recd       int
wlikes            int
wlikes_recd       int
Time taken: 0.23 seconds, Fetched: 15 row(s)
```

5. Queries:

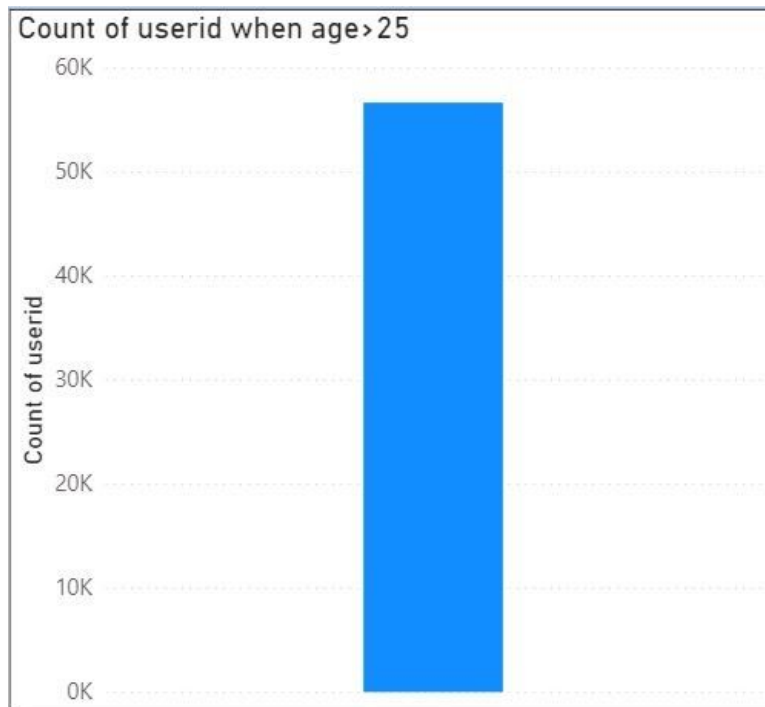
a) Find out the total number of users in this dataset

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> select count(*) from fb;  
Query ID = cloudera_20201130201111_98aa3092-9651-4a74-b935-9c06d61ca6e7  
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_1606795702556_0001, Tracking URL = http://quickstart.cloudera  
:8088/proxy/application_1606795702556_0001/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1606795702556_0001  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2020-11-30 20:12:20,330 Stage-1 map = 0%, reduce = 0%  
2020-11-30 20:12:33,705 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.1 sec  
2020-11-30 20:12:48,506 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.15  
sec  
MapReduce Total cumulative CPU time: 4 seconds 150 msec  
Ended Job = job_1606795702556_0001  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.15 sec HDFS Read: 5224610  
HDFS Write: 6 SUCCESS  
Total MapReduce CPU Time Spent: 4 seconds 150 msec  
OK  
99003  
Time taken: 54.167 seconds, Fetched: 1 row(s)  
hive>
```



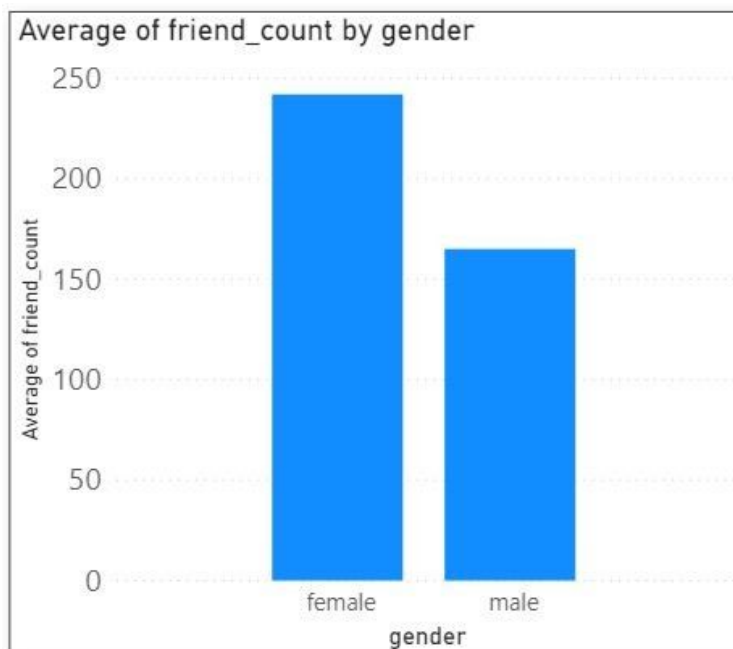
b) Find out the number of Facebook users above the age of 25

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> select count(*) from fb where age>25;
Query ID = cloudera_20201107042121_90397e0b-6302-4eda-95a9-9fc7c661d5ef
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_1604746193410_0006, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1604746193410_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1604746193410_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-07 04:21:20,479 Stage-1 map = 0%, reduce = 0%
2020-11-07 04:21:34,191 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.12 se
c
2020-11-07 04:21:49,060 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.28
sec
MapReduce Total cumulative CPU time: 5 seconds 280 msec
Ended Job = job_1604746193410_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.28 sec HDFS Read: 5225425
HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 280 msec
OK
56676
Time taken: 44.386 seconds, Fetched: 1 row(s)
```

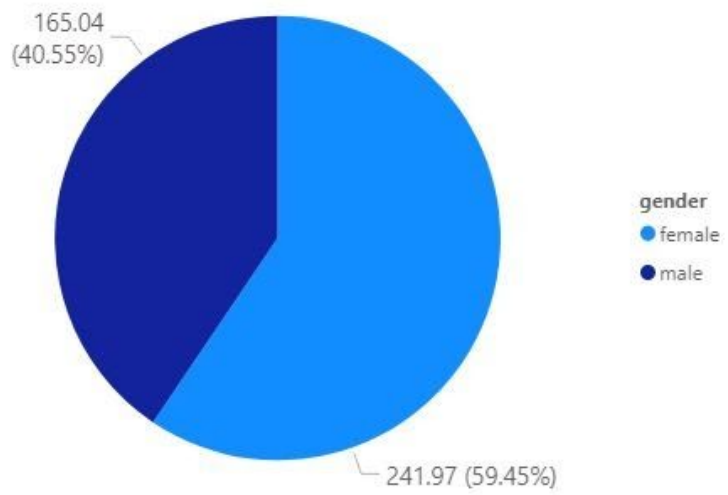


c) Do male facebook users tend to have more friends , or female users ?

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> select gender,avg(friends) from fb group by gender;
Query ID = cloudera_20201107043939_1c7ebbc3-a09e-4819-867a-a91785fa6c3d
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_1604746193410_0008, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1604746193410_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1604746193410_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-07 04:39:58,983 Stage-1 map = 0%, reduce = 0%
2020-11-07 04:40:11,603 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.38 se
c
2020-11-07 04:40:25,286 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.29
sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1604746193410_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.29 sec HDFS Read: 5225334
HDFS Write: 72 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
NA      184.41142857142856
female  241.96994087544095
male    165.03545941885477
Time taken: 41.817 seconds, Fetched: 3 row(s)
```

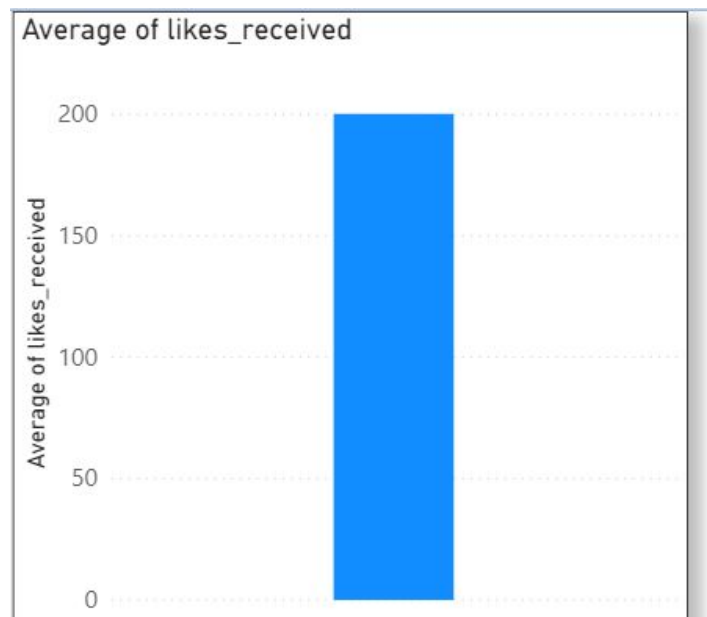


Average of friend_count by gender

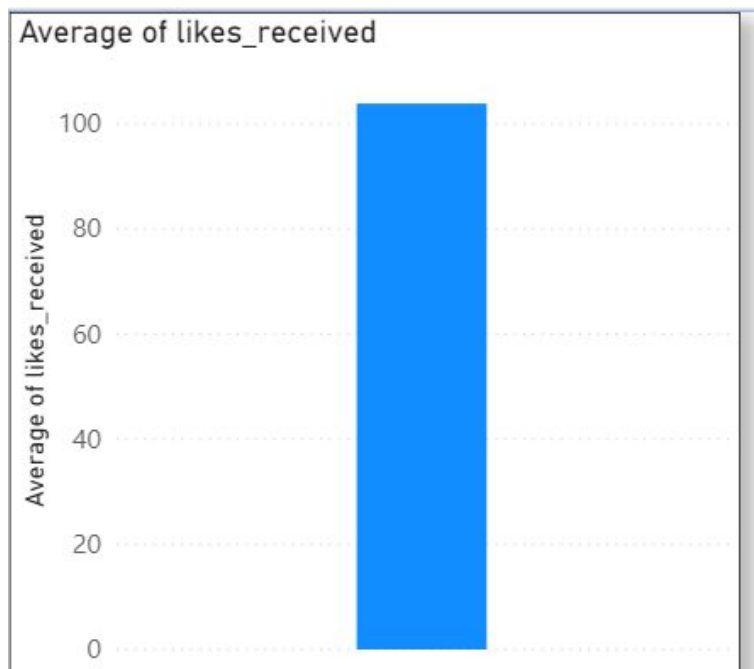


d) How many likes do young people receive on facebook opposed to older members

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> select avg(likes_recd) from fb where age>=13 AND age<=25;
Query ID = cloudera_20201107044141_38760244-9050-4281-8ac9-01bd130a44c9
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_1604746193410_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1604746193410_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1604746193410_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-07 04:41:37,232 Stage-1 map = 0%, reduce = 0%
2020-11-07 04:41:50,632 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.18 sec
2020-11-07 04:42:04,004 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.34 sec
MapReduce Total cumulative CPU time: 5 seconds 340 msec
Ended Job = job_1604746193410_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.34 sec HDFS Read: 5226250
HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 340 msec
OK
200.2870508186264
Time taken: 42.339 seconds, Fetched: 1 row(s)
```

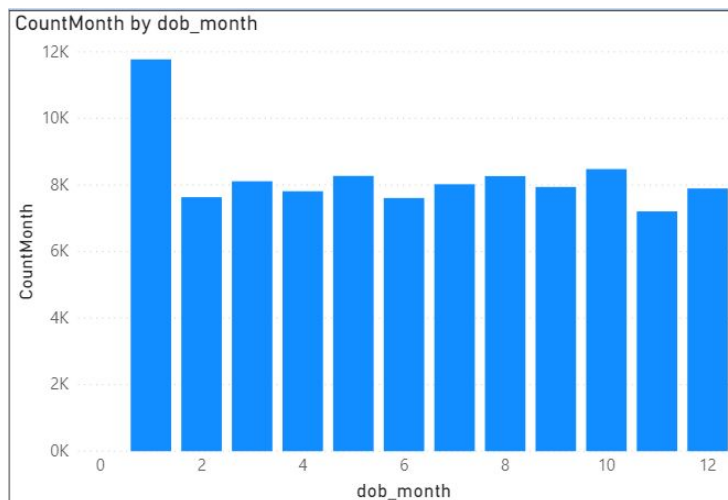



```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> select avg(likes_rec'd) from fb where age>=35;
Query ID = cloudera_20201107044444_a63adbcc-elcc-4143-ab7c-afe48b5fb925
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to 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_1604746193410_0010, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1604746193410_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1604746193410_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-07 04:44:30,772 Stage-1 map = 0%, reduce = 0%
2020-11-07 04:44:44,061 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.88 se
c
2020-11-07 04:44:57,449 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.98
sec
MapReduce Total cumulative CPU time: 4 seconds 980 msec
Ended Job = job_1604746193410_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.98 sec HDFS Read: 5225848
HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 980 msec
OK
103.89021217994491
Time taken: 41.774 seconds, Fetched: 1 row(s)
```



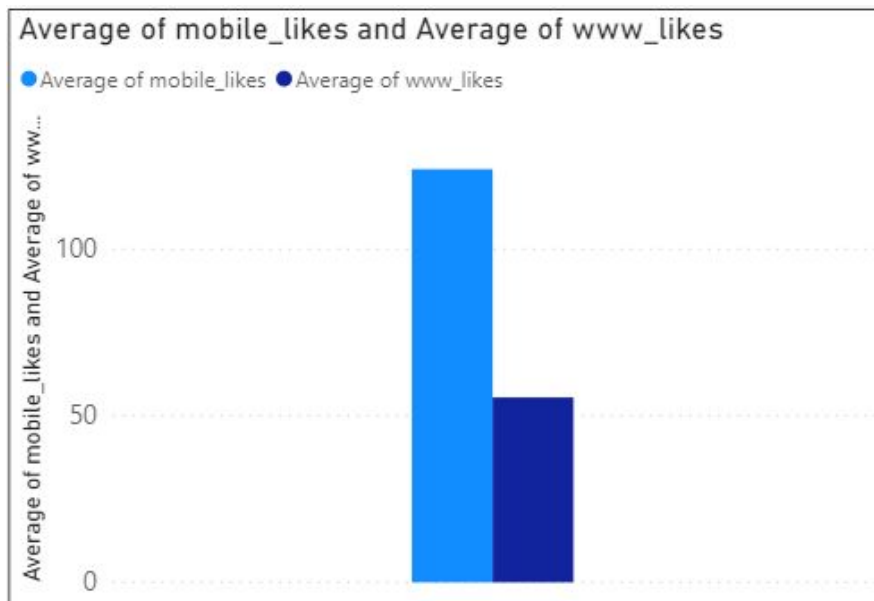
e) Find out the count of facebook users for each birthday month

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
hive> select month,count(*) from fb group by month;
Query ID = cloudera_20201107044545_548611a1-2968-4ad2-8457-56192f1993a4
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_1604746193410_0011, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1604746193410_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1604746193410_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-11-07 04:45:48,205 Stage-1 map = 0%, reduce = 0%
2020-11-07 04:46:00,565 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.23 se
c
2020-11-07 04:46:14,032 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.23
sec
MapReduce Total cumulative CPU time: 4 seconds 230 msec
Ended Job = job_1604746193410_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.23 sec HDFS Read: 5224912
HDFS Write: 88 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 230 msec
OK
1      11772
2      7632
3      8110
4      7810
5      8271
6      7607
7      8021
8      8266
9      7939
10     8476
11     7205
12     7894
Time taken: 41.013 seconds, Fetched: 12 row(s)
```

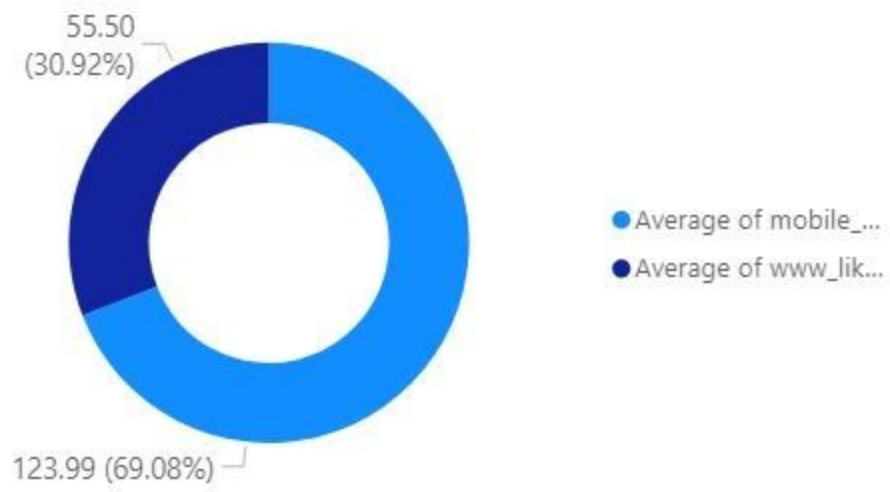


f) Do young members use mobile phones or computers for facebook browsing ?

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> select avg(mlikes),avg(wlikes) from fb where age>=13 AND age<=25;  
Query ID = cloudera_20201130202525_6da93246-0835-4a23-90ea-4ee98b4afb0d  
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_1606795702556_0002, Tracking URL = http://quickstart.cloudera:  
8088/proxy/application_1606795702556_0002/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1606795702556_0002  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2020-11-30 20:25:39,929 Stage-1 map = 0%, reduce = 0%  
2020-11-30 20:25:53,726 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.03 sec  
2020-11-30 20:26:06,208 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.09 s  
ec  
MapReduce Total cumulative CPU time: 5 seconds 90 msec  
Ended Job = job_1606795702556_0002  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.09 sec HDFS Read: 5226872  
HDFS Write: 37 SUCCESS  
Total MapReduce CPU Time Spent: 5 seconds 90 msec  
OK  
123.98981737425284 55.50010631511801  
Time taken: 44.082 seconds, Fetched: 1 row(s)  
hive> █
```

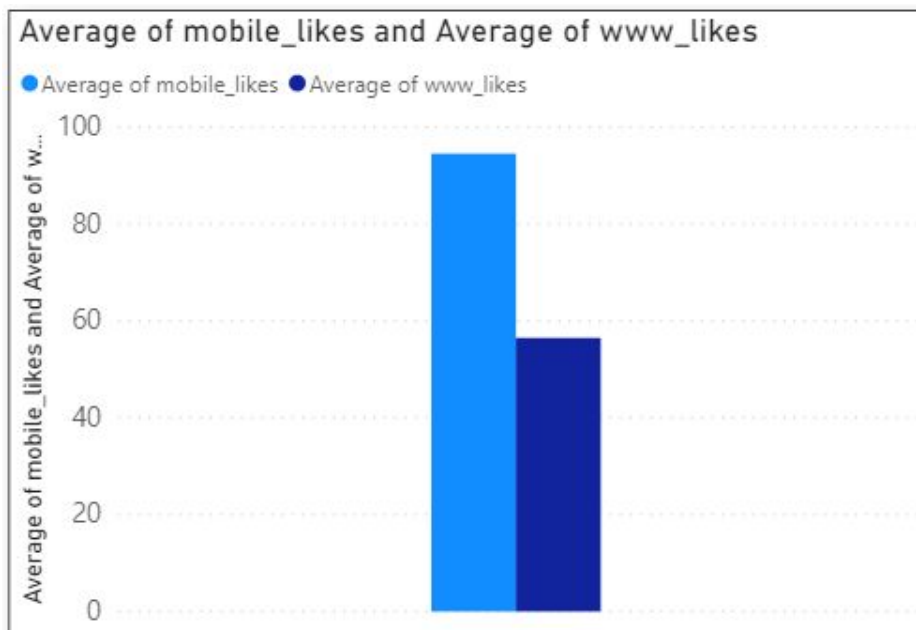


Average of mobile_likes and Average of www_likes



g) Do adult members use mobile phones or computers for facebook browsing ?

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> select avg(mlikes), avg(wlikes) from fb where age>=35;  
Query ID = cloudera_20201130202828_351acdf1-e177-435f-9860-34788b943e2c  
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_1606795702556_0003, Tracking URL = http://quickstart.cloudera:  
8088/proxy/application_1606795702556_0003/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1606795702556_0003  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2020-11-30 20:28:56,365 Stage-1 map = 0%, reduce = 0%  
2020-11-30 20:29:10,458 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.05 sec  
2020-11-30 20:29:24,180 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.04 s  
ec  
MapReduce Total cumulative CPU time: 5 seconds 40 msec  
Ended Job = job_1606795702556_0003  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.04 sec HDFS Read: 5226452  
HDFS Write: 36 SUCCESS  
Total MapReduce CPU Time Spent: 5 seconds 40 msec  
OK  
94.55878302560441      56.50313679485872  
Time taken: 49.884 seconds, Fetched: 1 row(s)  
hive>
```



Average of mobile_likes and Average of www_likes

