Question 1:

1. Find out the total number of users in this dataset

```
hive> select count(*) from fb
. . > ;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. C
onsider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = aghil_20240621232227_6ec7c124-eea4-4cb7-a4d7-0dfbc1aea5df
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_1718910569103_0001, Tracking URL = http://LAPTOP-6NUJLARE:8088/proxy/applic
ation_1718910569103_0001/
Kill Command = C:\hadoop\bin\hadoop.cmd job  -kill job_1718910569103_0001
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. C
onsider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-06-21 23:22:39,668 Stage-1 map = 0%,  reduce = 0%
2024-06-21 23:22:43,820 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.062 sec
2024-06-21 23:22:48,971 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.124 sec
MapReduce Total cumulative CPU time: 2 seconds 124 msec
Ended Job = job_1718910569103_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.124 sec   HDFS Read: 5226012 HDFS Write:
105 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 124 msec
OK
99003
1 row selected (22.91 seconds)
hive>
```

The cool things about hive that even we used a large file we say for example 3 TB, the process will take the same time as a 5MB file.

Question 2:

2. Find out the number of facebook users above the age of 25

```
hive> select count(*) from fb where age > 25;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. C
onsider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = aghil_20240622090847_71717e88-837a-41fa-bcfb-a8ab9d5b5afb
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_1718910569103_0002, Tracking URL = http://LAPTOP-6NUJLARE:8088/proxy/applic
ation_1718910569103_0002/
Kill Command = C:\hadoop\bin\hadoop.cmd job  -kill job_1718910569103_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-06-22 09:08:57,317 Stage-1 map = 0%,  reduce = 0%
2024-06-22 09:09:02,488 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.249 sec
2024-06-22 09:09:06,629 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.529 sec
MapReduce Total cumulative CPU time: 3 seconds 529 msec
Ended Job = job_1718910569103_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.529 sec   HDFS Read: 5226835 HDFS Write:
105 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 529 msec
OK
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. C
onsider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
56676
1 row selected (20.345 seconds)
hive>
```

Question 3:

3. Do male facebook users tend to have more friends, or female users?

- Average for mens :

```
1 row selected (20.345 seconds)
hive> select avg(friends) from fb where gender='male';
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different
xecution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = aghil_20240622092908_ff5a2de2-2813-482a-96b2-f6ee9418b2aa
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_1718910569103_0003, Tracking URL = http://LAPTOP-6NUJLARE:8088/proxy/application_1718910569103_0003/
Kill Command = C:\hadoop\bin\hadoop.cmd job  -kill job_1718910569103_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-06-22 09:29:13,034 Stage-1 map = 0%,  reduce = 0%
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different
xecution engine (i.e. spark, tez) or using Hive 1.X releases.
2024-06-22 09:29:17,167 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.171 sec
2024-06-22 09:29:22,307 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.529 sec
MapReduce Total cumulative CPU time: 3 seconds 529 msec
Ended Job = job_1718910569103_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.529 sec   HDFS Read: 5227605 HDFS Write: 118 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 529 msec
OK
165.03545941885477
1 row selected (15.153 seconds)
```

- Average for female:

```
hive> select avg(friends) from fb where gender='female';
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different e
xecution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = aghil_20240622092936_caba23ed-9e56-48fe-bd0a-ec8dcb64f2bf
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_1718910569103_0004, Tracking URL = http://LAPTOP-6NUJLARE:8088/proxy/application_1718910569103_0004/
Kill Command = C:\hadoop\bin\hadoop.cmd job  -kill job_1718910569103_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-06-22 09:29:40,740 Stage-1 map = 0%,  reduce = 0%
2024-06-22 09:29:44,858 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.92 sec
2024-06-22 09:29:49,990 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.278 sec
MapReduce Total cumulative CPU time: 3 seconds 278 msec
Ended Job = job_1718910569103_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.278 sec   HDFS Read: 5227614 HDFS Write: 118 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 278 msec
OK
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different e
xecution engine (i.e. spark, tez) or using Hive 1.X releases.
241.96994087544095
1 row selected (15.021 seconds)
hive>
```

- Select the avg for all gender in the excel file :

```
hive> select gender, avg(friends) from fb group by gender;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consi
der using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = aghil_20240622093850_d60ae73c-6170-4a72-aef0-15c9dd4d1989
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_1718910569103_0005, Tracking URL = http://LAPTOP-6NUJLARE:8088/proxy/applicatio
n_1718910569103_0005/
Kill Command = C:\hadoop\bin\hadoop.cmd job  -kill job_1718910569103_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-06-22 09:38:55,596 Stage-1 map = 0%,  reduce = 0%
2024-06-22 09:38:59,724 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.436 sec
2024-06-22 09:39:04,873 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.56 sec
MapReduce Total cumulative CPU time: 2 seconds 560 msec
Ended Job = job_1718910569103_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.56 sec   HDFS Read: 5227241 HDFS Write: 195 S
UCCESS
Total MapReduce CPU Time Spent: 2 seconds 560 msec
OK
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consi
der using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
NA 184.41142857142856
female 241.96994087544095
male 165.03545941885477
3 rows selected (15.029 seconds)
```

Question 4:

4. How many likes do young people receive on facebook opposed to older members Find out the count

- Calculat the avg for the young peopel:
  We put the age from 13 years to 25 years.

```
hive> select avg(likes_recd) from fb where age>=13 AND age <=25;
```

200.2870508186264

- Calculat the avg likes for the older people
We put age under 25 years now.

hive> select avg(likes_recd) from fb where age>35;
104.6969792570392

Question 5:

5.Find out the count of facebook users for each birthday month

- hive> select month,count(*) from fb group by month;
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

Question 6:

6.Do young members use mobile phones or computers for facebook browsing?

hive> select avg(mlikes),avg(wlikes) from fb where age >=13 AND age<=25;
Mobile : 123.98981737425284
PC : 55.50010631511801

Question 7:

7. Do adult members use mobile phones or computers for facebook browsing? I

hive> select avg(mlikes),avg(wlikes) from fb where age >=35;
Mobile :94.55878302560441
PC :56.50313679485872