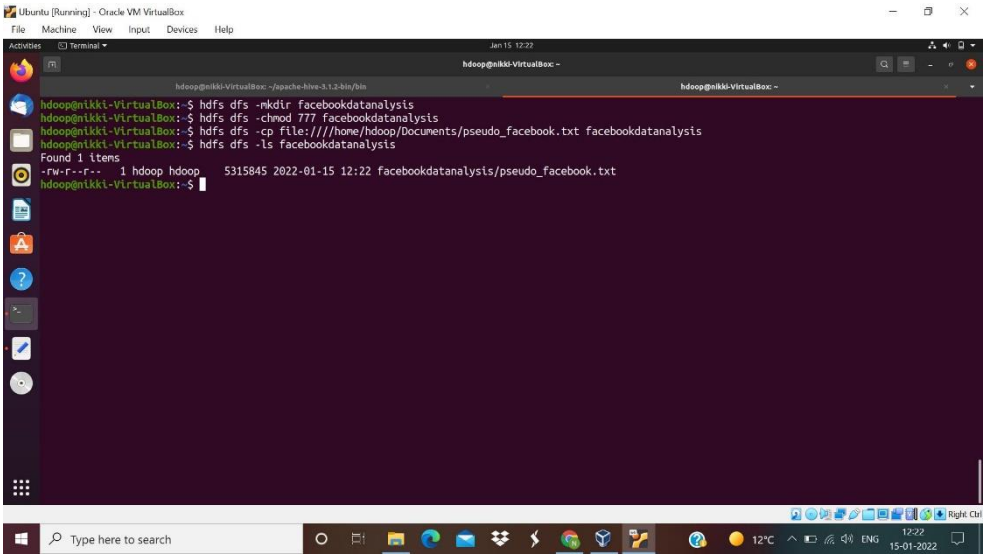
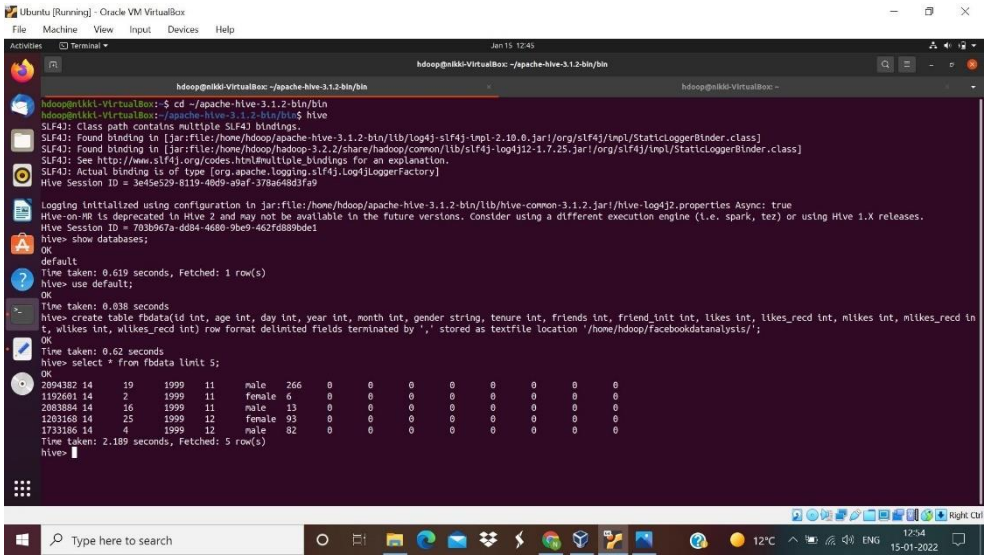
FACEBOOK DATA ANALYSIS

1. Uploading Facebook dataset from my local system into HDFS which is Hadoop Distributed File System.
   Command:
   hdfs dfs -mkdir facebookdatanalysis
   hdfs dfs -chmod 777 facebookdatanalysis
   hdfs dfs -cp file:////home/hdoop/Documents/pseudo_facebook.txt facebookdatanalysis
   hdfs dfs -ls facebookdatanalysis



2. Starting hive in my system
   Creating table inside default database to store the dataset (i.e Facebook data set) inside that table
   Command to create table inside hive:
   create table fbdata(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/hdoop/facebookdatanalysis/';

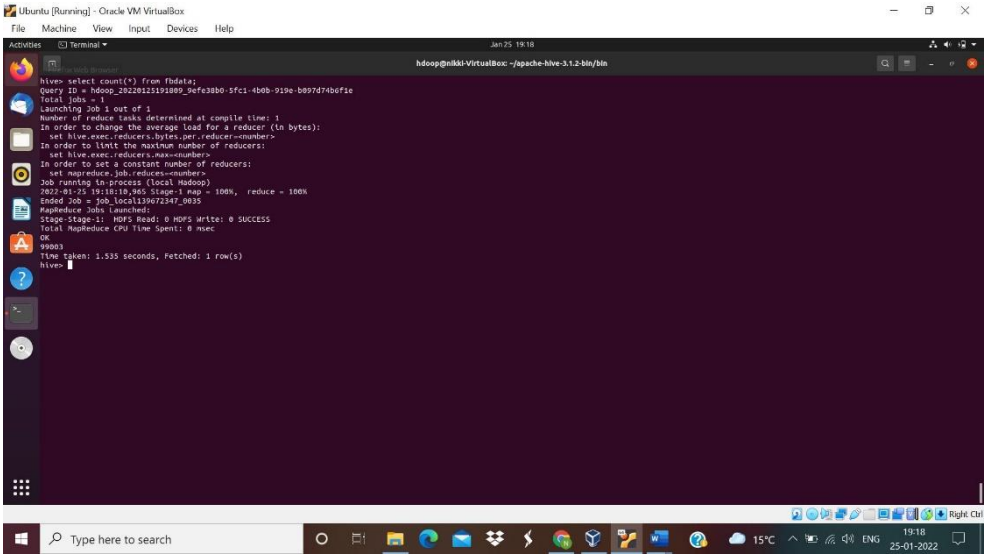


HIVE QUERY

1. **Total number of users in his dataset**

   Query:

   ➔ Select count(*) from fbdata;

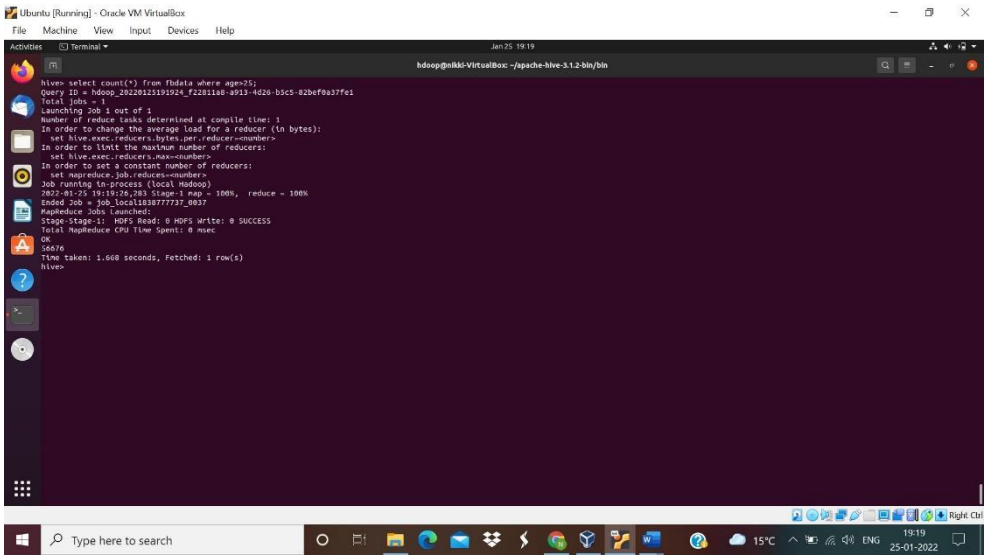


   Result: There are total 99003 users in our dataset

2. **Find out number of Facebook users above the age of 25**

   Query:

   ➔ Select count(*) from fbdata where age>25;



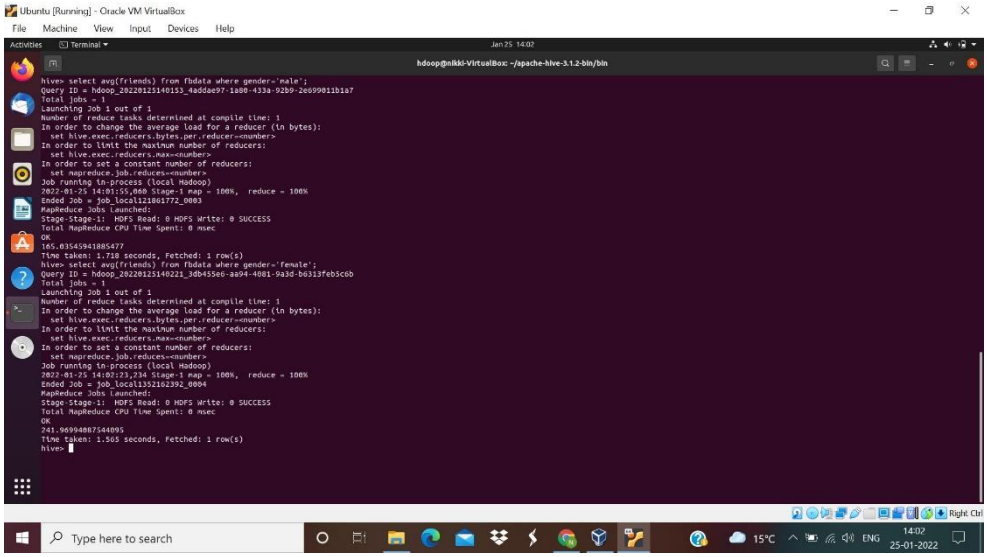
   Result: There are 56676 users that are above the age 25

3. **Do male Facebook users tend to have more friends or female users?**

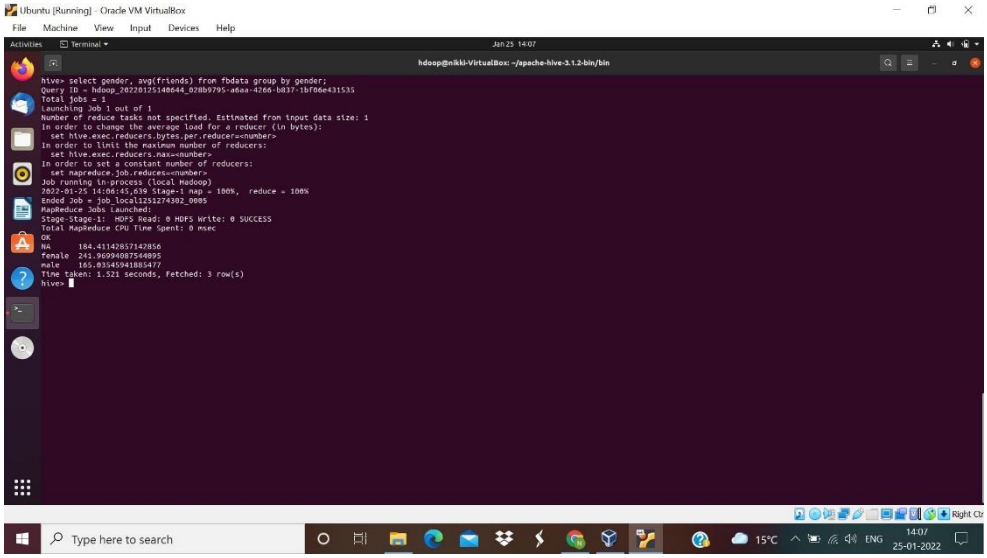
   Query:

   ➔ Select avg(friends) from fbdata where gender='male';

   ➔ Select avg(friends) from fbdata where gender=' female';



➔ select gender, avg(friends) from fbdata group by gender;



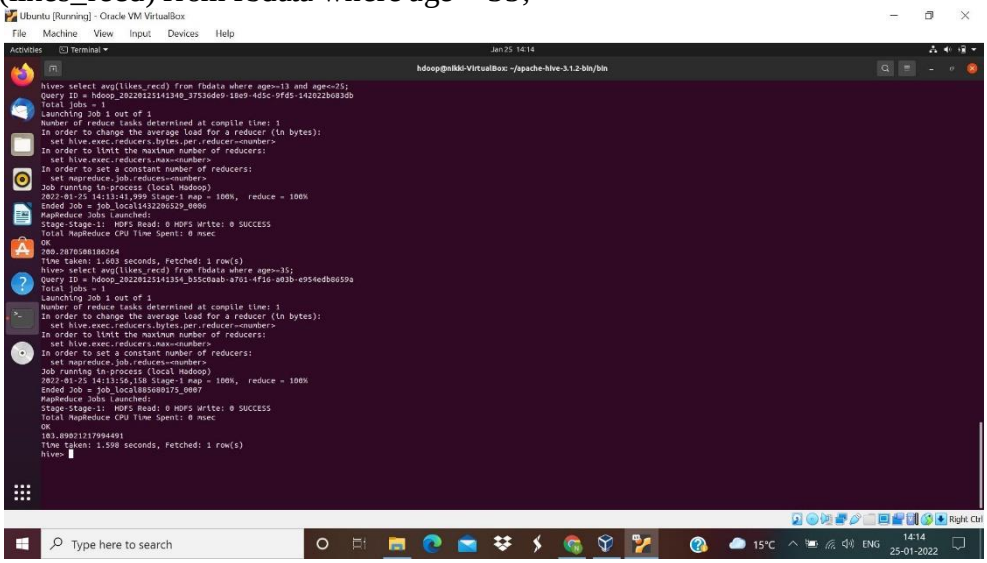
Result: Female user tends to be more active than male user

4. **How many likes do young people receive on Facebook opposed to older members**

Query:

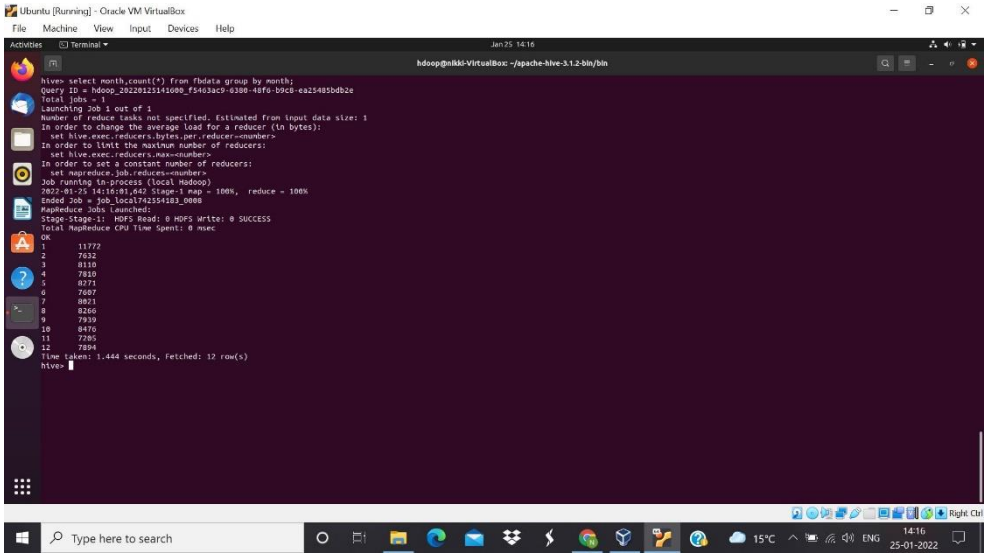
➔ select avg(likes_recd) from fbdata where age>=13 and age<=25;

➔ select avg(likes_recd) from fbdata where age>=35;

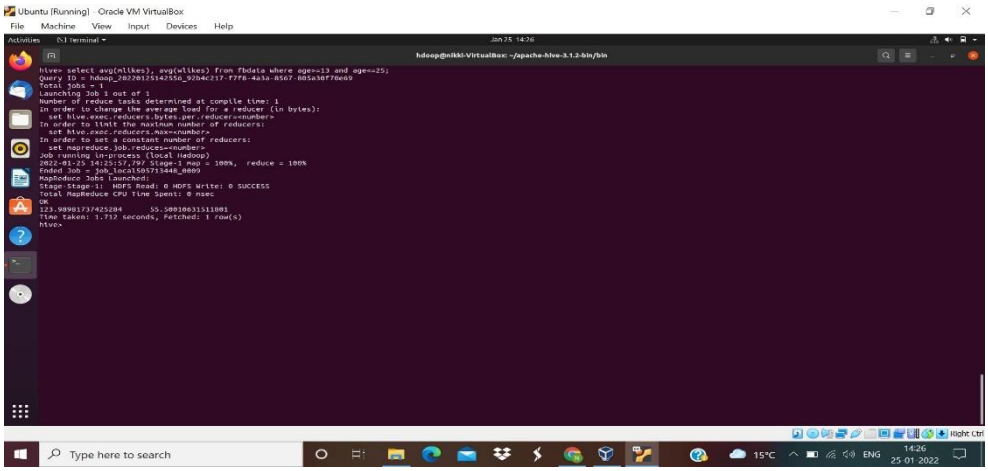


Result: Young people receive more like than older people

5. **Find out the count of Facebook users for each birthday month**
   Query:
   ➔ select month, count(*) from fbdata group by month;

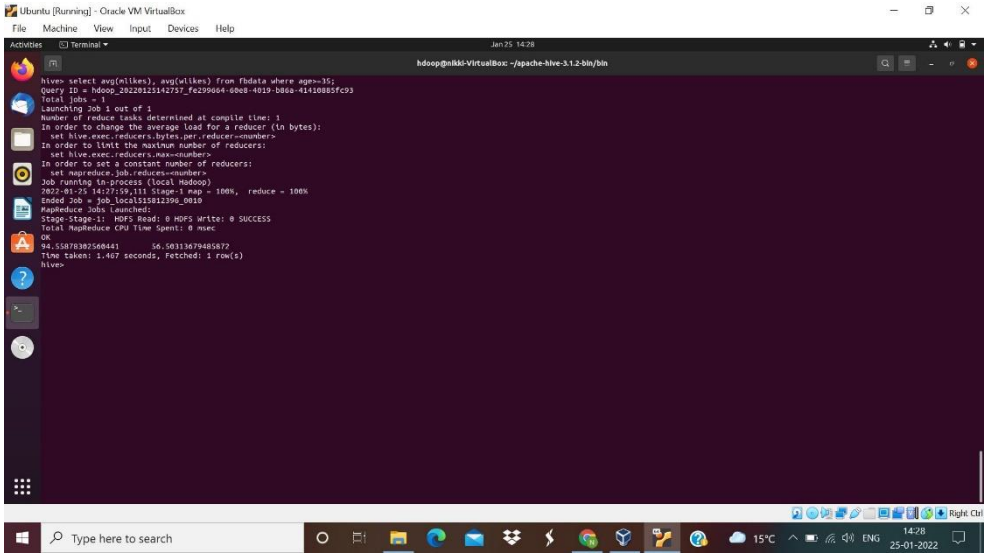


6. **Do young members use mobile phone or computers for Facebook browsing?**
   Query:
   ➔ select avg(mlikes), avg(wlikes) from fbdata where age>=13 and age<=25;



Result: From the above observation young people use mobile phone for Facebook browsing than computers;

7. **Do older members use mobile phone or computer for facebook browsing?**
   Query:
   ➔ select avg(mlikes), avg(wlikes) from fbdata where age>=35;

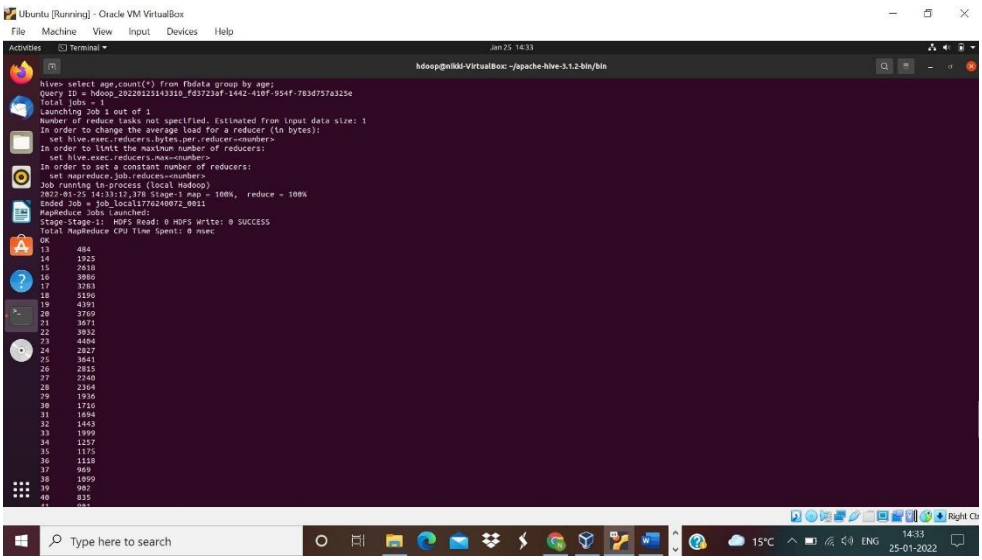


Result: From the above observation older people use mobile phone for Facebook browsing than computers;

8. **Find out number of people in each age group**

Query:

➔ select age, count(*) from fbdata group by age;

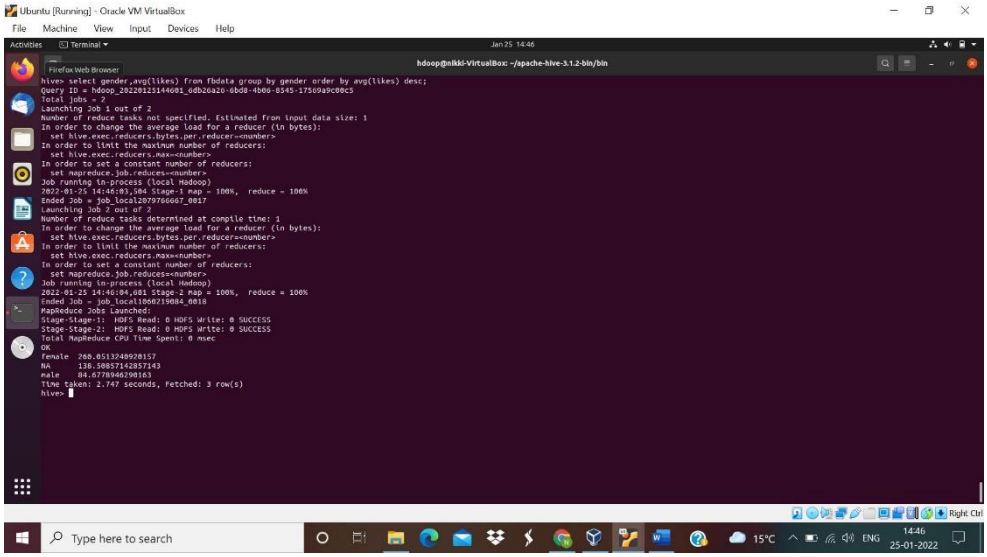


Result: Facebook is most popular between age groups 16 and 26.

9. **Who give more likes male or female?**

Query:

➔ select gender, avg(likes) from fbdata group by gender order by avg(likes) desc;

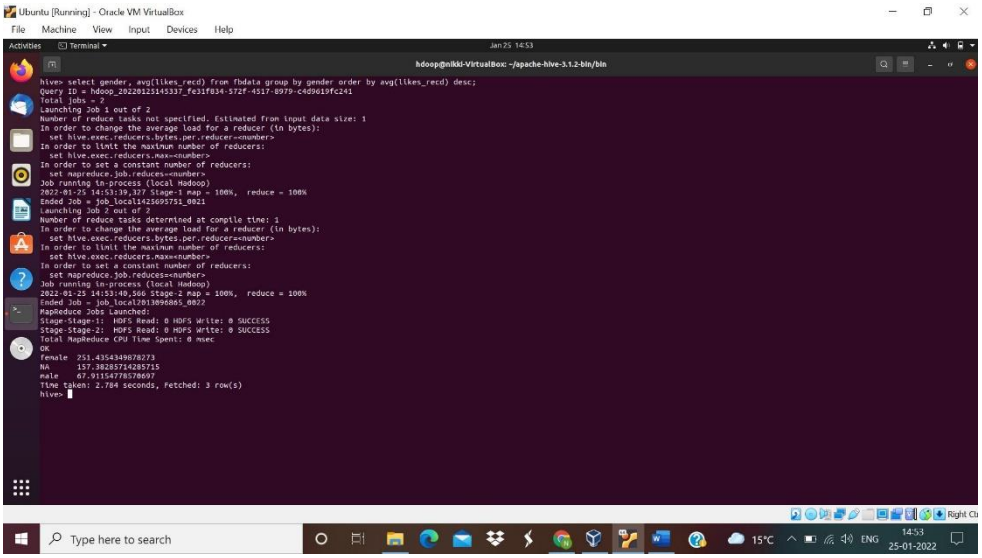


Result: Female give more likes then men.

10. Who receive more likes?

Query:

➔ select gender, avg(likes_recd) from fbdata group by gender order by avg(likes_recd) desc;

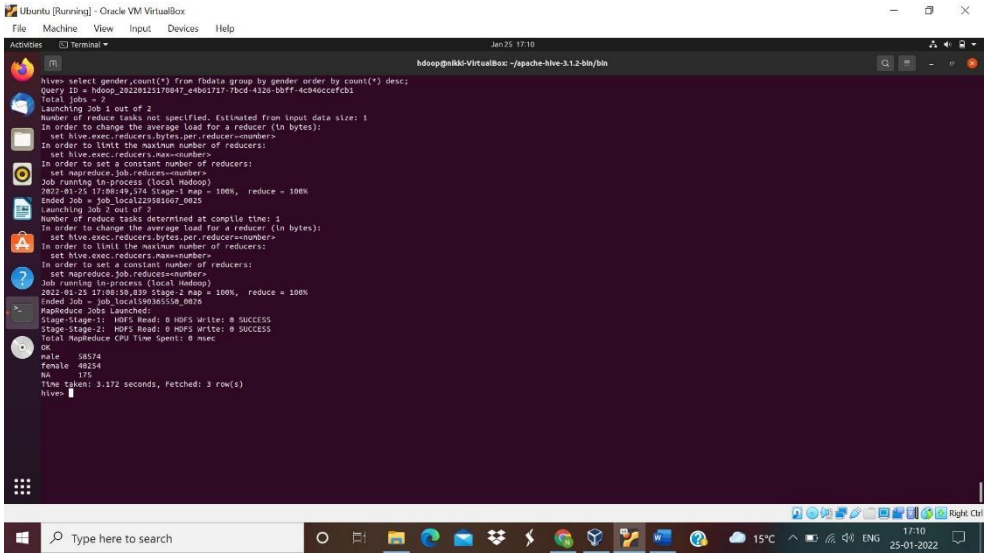


Result: Females receive more likes than men.

11. Gender count

Query:

➔ select gender, count(*) from fbdata group by gender order by count(*) desc;

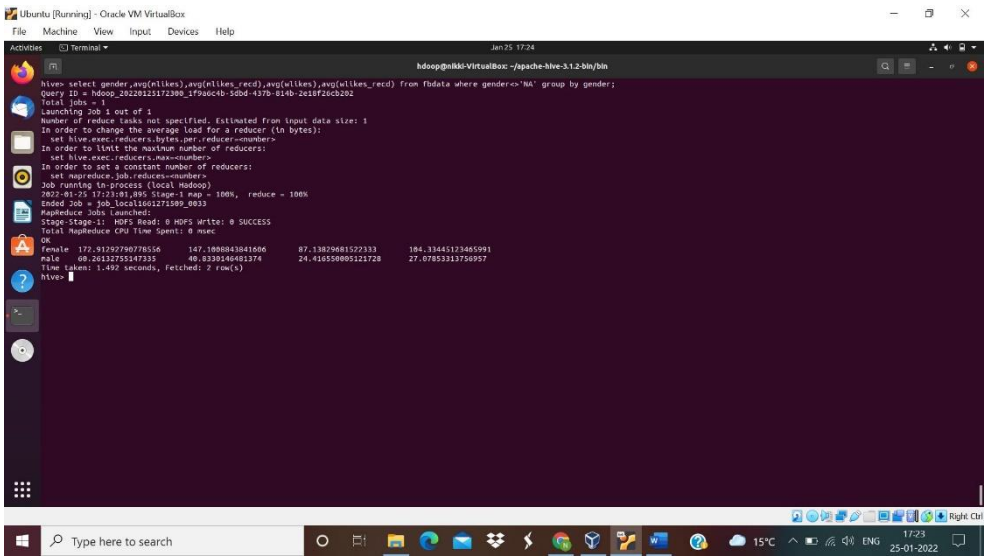


Result: There are more male users than female.

12. Likes split up

Query:

➔ select gender, avg(mlikes), avg(mlikes_recd), avg(wlikes), avg(wlikes_recd) from fbdata where gender<>'NA' group by gender;

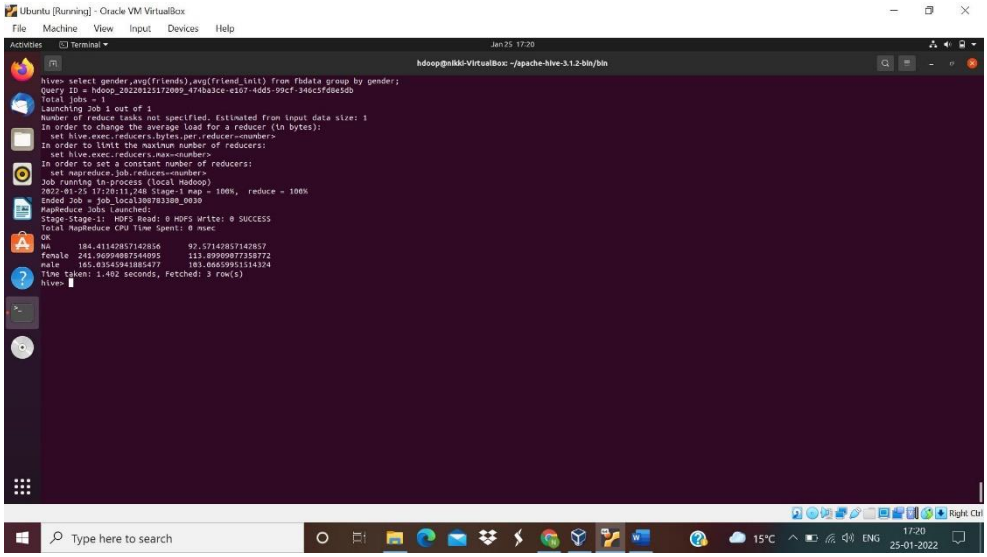


Result: Interesting observation for gender specific interaction with Facebook: women likes are a lot more than man.

13. Friend counts and Friendship initiated

Query:

➔ select gender, avg(friends), avg(friend_init) from fbdata group by gender;



Result: Women have more friends than men on facebook, the friendshios initiated in proportion to friend count are more in case of men than women.