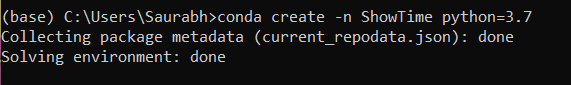
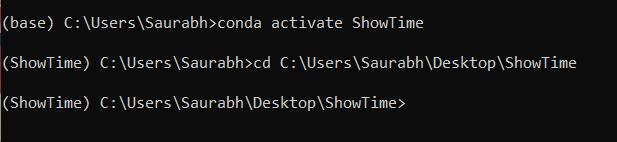
**INTRODUCTION**

I will be using Jupyter Notebook by creating an environment in Anaconda prompt, to apply Exploratory data analysis and perform all the questions given in the assignment. I am using the following python libraries.

Following is the code snippet to crate the anaconda environment:

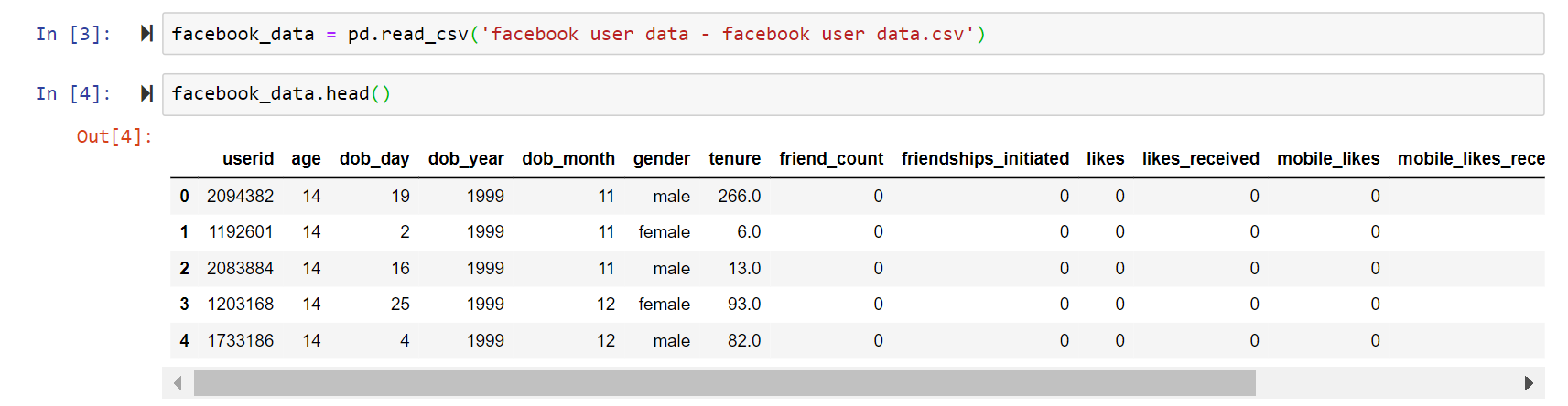




**1) Load the data and impute missing values**

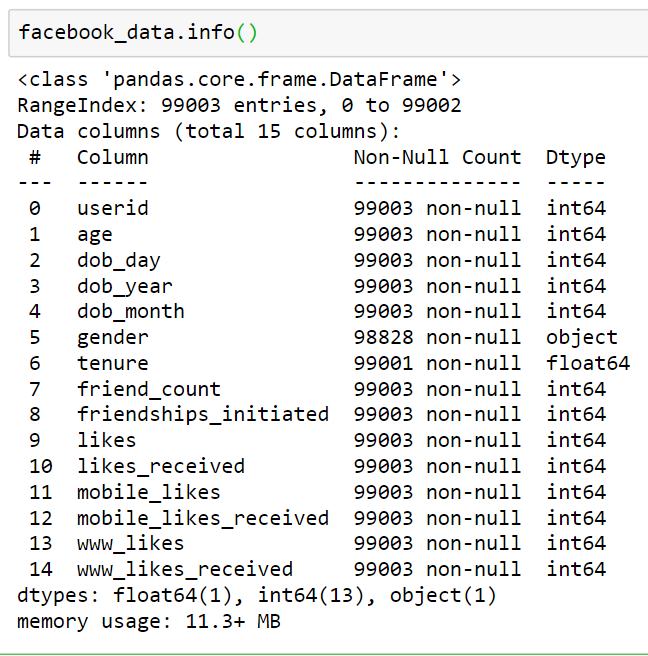
1. **Loading the data**

I downloaded the data from the given link in csv form. With the help of pandas I loaded the data to jupyter notebook.



1. **Taking insights from data**

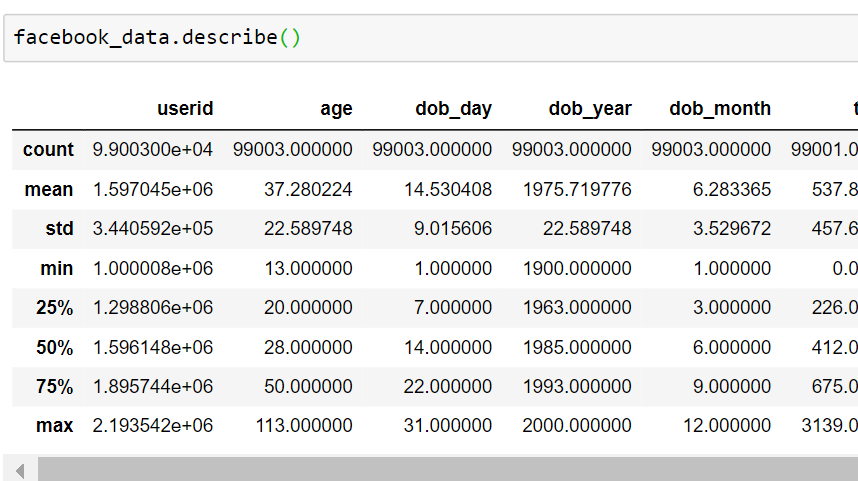
Using info() to know about the data types and null values existing in the dataset.



We can observe following points :

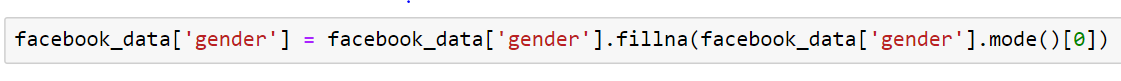
* Feature ‘tenure’ has float64 datatype , ‘gender’ has object data type while all other features have int64 datatype.
* Feature gender has 175 null values and feature tenure has 2 null values.

Some more insights:

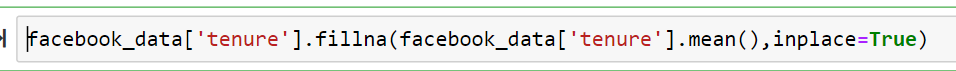


1. **Data Cleaning**

The gender feature with data type object has categorical values as male and female. We will use mode to impute null values here.

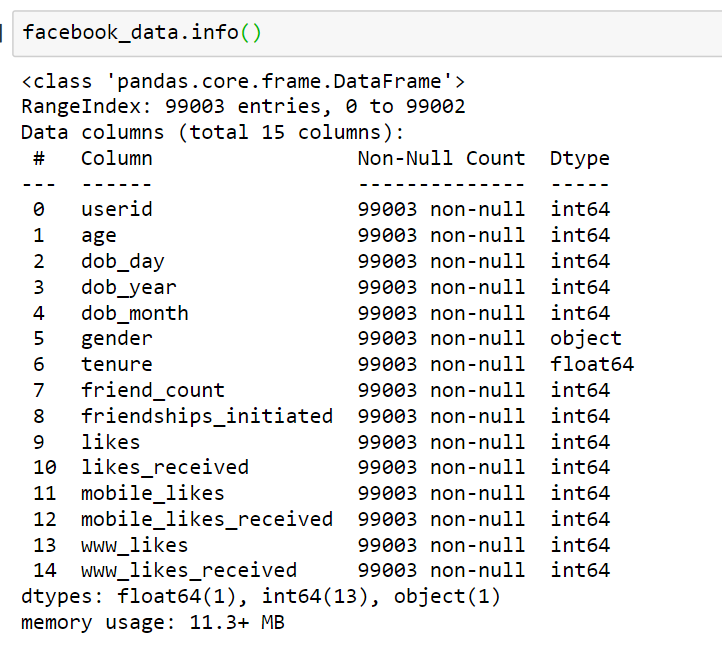


Now to impute tenure we will use mean as it has numerical data.



We use mode/median to fill the null data because it gives better results when we are training the machine learning model, else the model will falsely interpret some values.

Now we can see there are no null values

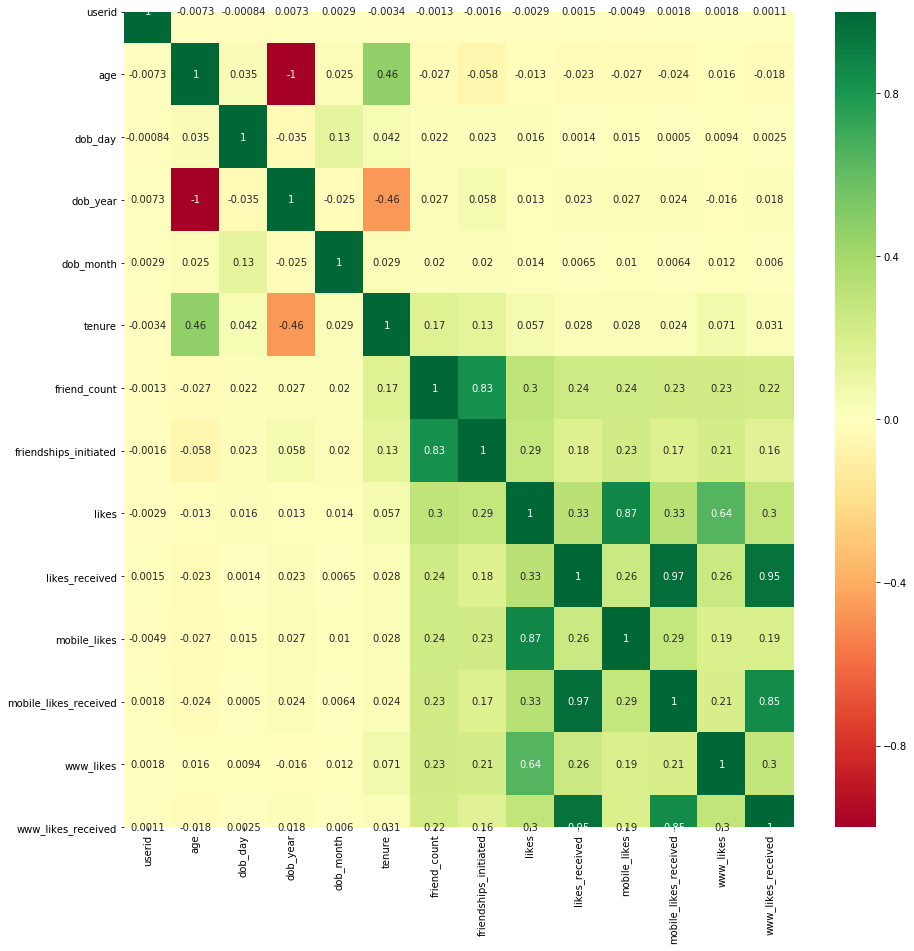


**2) Plot heatmap / correlation matrix on all the columns.**

I will here make a correlation matrix using corr() and then use seaborn to plot the heatmap



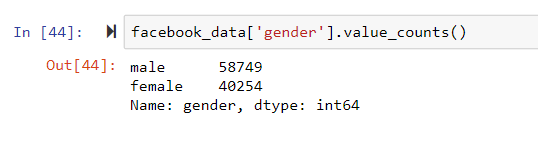
This will give the following heat map kind of heat map

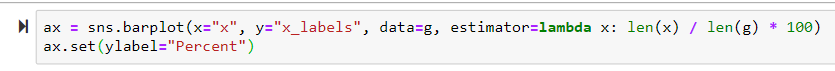


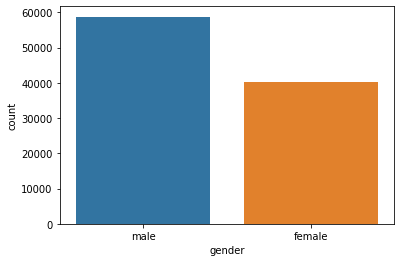
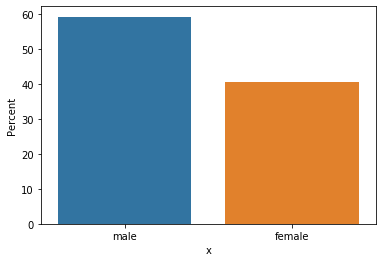
Here we can see the green boxes show the features with positive correlation and the red show that the features that have negative correlation.

**3) Analysis based on gender of the users**

**What is composition of male and female users?**



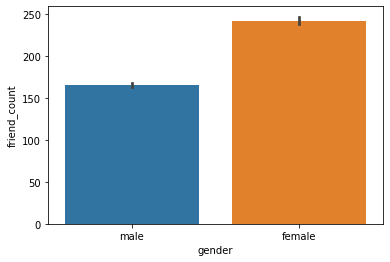
Lets plot the percentage composition using count plot and bar plot from seaborn library.

**Which category of gender has more friends?**

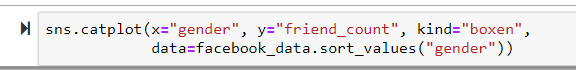
I will visualize this with bar plot



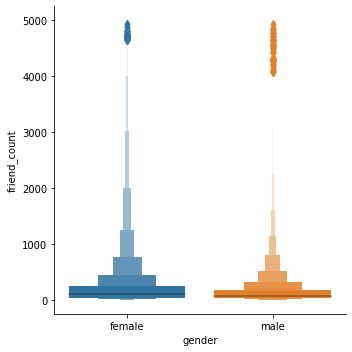


We can see that on an average females have more friends than males.

Now I will visualize this with boxen plot in catplot to get more insights.



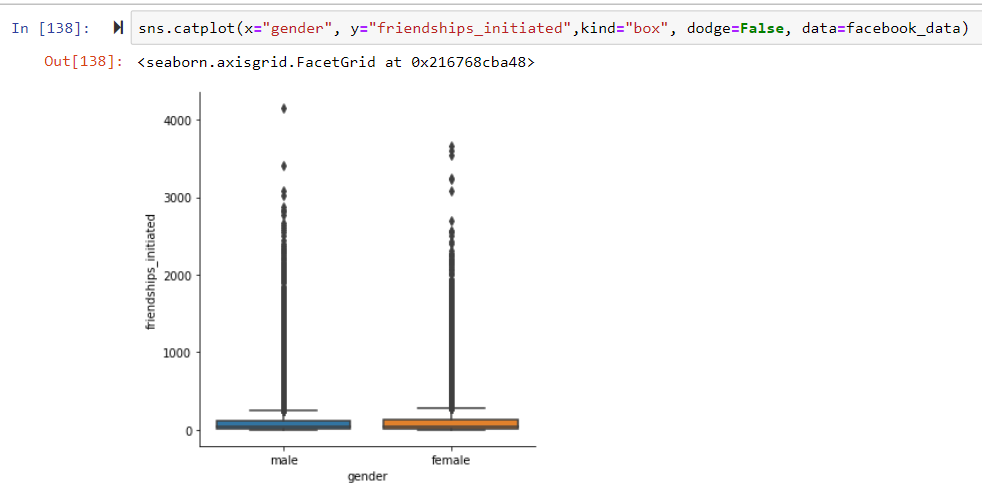
The ouput to this is given below:



This tells us more insights like more no of females have friend\_count around in 500-1000 range than males.

**Which category of gender initiated more friendships?**

I will use catplot with kind box as well as bar plot to visualize this



The horizontal lines tell us about 25,50,75 percentile values

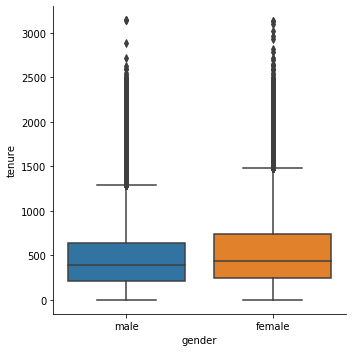


This gives the answer to the question females have initiated more friendships than male.

**What is the distribution of tenure across different categories of gender?**

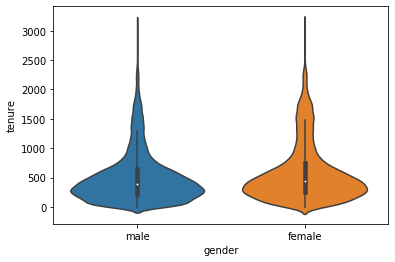
I will use catplots to see distribution of tenure across the gender:





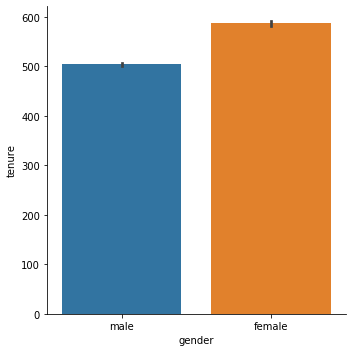
This shows the 0,25,50,75 percentile as well as mean values across the tenure based on gender





This gives distribution based on tenure values and gender

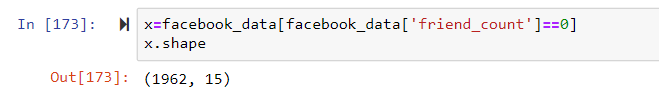




Finally this bar graph shows the average values of tenure based on gender

**4) Analysis based on the least active users on Facebook**

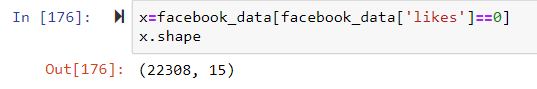
**● How many users have no friends?**



Here I applied the condition facebook\_data['friend\_count']==0 to get the rows with 0 friend\_counts.

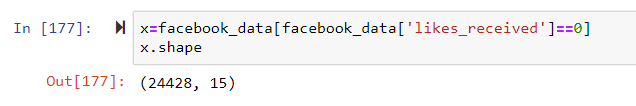
Shape[0] shows the number of rows which gives that 1962 users have 0 friends.

**● How many users did not like any posts?**



22308 users have not liked any post.

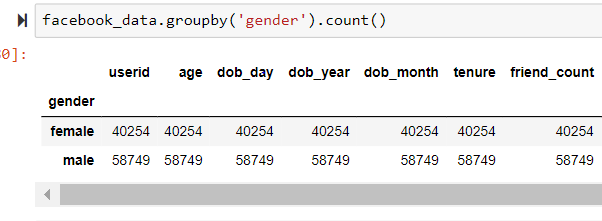
**● How many users did not receive any likes?**



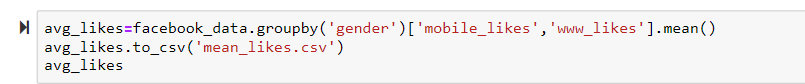
24428 users have not received any likes.

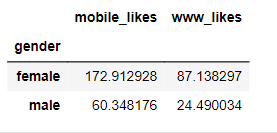
# 5) Analysis based on the user accessibility (Mobile Devices vs. Web Devices)

Lets first explore some data based on gender



**● What is the average number of posts liked by users (based on gender) through web vs. mobile devices?**



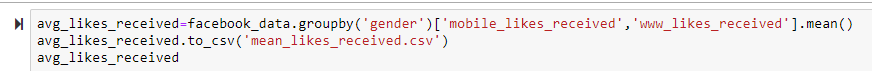


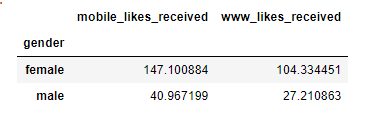
I have saved this dataset as mean\_likes.csv

I grouped the data using gropuby() based on gender and the took the average of ‘mobile\_likes’ and ‘www\_likes’ columns to get the numbers specified in the question.

This shows the average likes of the user from mobile vs from web based on gender.

**● What is the average number of likes received by users (based on gender) through web vs. mobile devices?**





I have saved this file as mean\_likes\_received.csv

This shows the average likes received by the user from mobile vs from web based on gender