HW 4 - Exploring Social Networks

Swathi Venkatesh

11/02/2021

Q1

Determine if the friendship paradox holds for a user's Facebook account. (This used to be more interesting when you could more easily download your friend's friends data from Facebook. Facebook now requires each friend to approve this operation, effectively making it impossible.)

HW4-friend-count.csv contains a user's friends' names and number of friends they each have. (You'll need to accept the GitHub Classroom invite to get this datafile.)

Q: What is the mean, standard deviation, and median of the number of friends that the user's friends have?

Answer

```
#!/usr/local/bin/python3
import matplotlib.pyplot as plt
import pandas as pd
f = "HW4-friend-count.csv"
df = pd.read_csv(f,index_col=False, encoding='utf-8')
#remove extra spaces from columns
df.columns =[col.strip() for col in df.columns]
df = df.sort_values(by="FRIENDCOUNT")

#create a list of new users
newuser= []
for x  in range(98):
    c =  str(x)
    newuser.append(c)
# replace this new list to the column user in the df
df.USER = newuser
print('The mean is',df.FRIENDCOUNT.mean())
print('The std deviation is',df.FRIENDCOUNT.std())
print('The median is',df.FRIENDCOUNT.median())
#plot the values
plt.rcParams['figure.figsize'] =20,40
plt.rcParams['font.size'] = 12;
plt.plot(df.USER,df.FRIENDCOUNT,marker='o')
plt.grid(True)
plt.xlabel(xlabel = "friends", fontsize=20)
```

```
26 plt.ylabel(ylabel= "no. of friends", fontsize=20)
27 plt.legend(loc=6,fontsize=25);
28 plt.gca().invert_xaxis()
29 plt.show()
```

Listing 1: frienshipParadox.py

```
Python 3.6.8 Shell
File Edit Shell Debug Options Window Help
Python 3.6.8 (tags/v3.6.8:3c6b436a57, Dec 24 2018, 00:16:47) [MSC v.1916 64 bit
(AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:\Users\SWATHI\Desktop\hw\friendshipParadox.py ==========
The mean is 542.6734693877551
The std deviation is 539.4337385239659
The median is 396.0
```

Figure 1: Mean, standard deviation and median

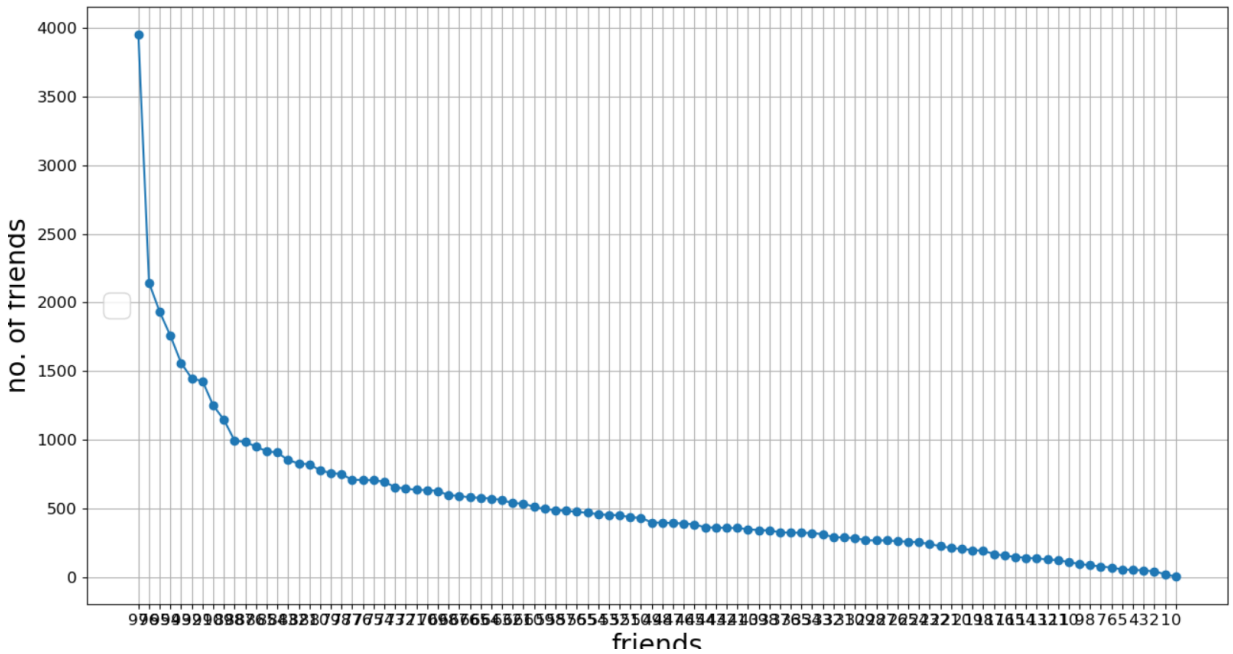

Figure 2: Plot for Q1

Q: Does the friendship paradox hold for this user and their friends on Facebook?

Ans: Yes, friendship paradox hold for this user and their friends on Facebook.

Discussion

I read the csv file using the pandas library. Then, sorted the values of the FriendCount column.

I replaced the user column in dataframe with list values. Using the built dataframe I calculated the mean, standard deviation and median. Finally plotted the graph as shown in Figure 2. As, the friend with the most friends should be on the left side of the graph I used plt.gca().invert_xaxis() for the same.

The mean is 542.6734693877551

The std deviation is 539.4337385239659

The median is 396.0

Q2

Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use followers as the value you measure (i.e., "do your followers have more followers than you?").

If you have less than 50 followers on Twitter, then you can do the analysis for another Twitter account (e.g., my account is weiglemc) and substitute the user you pick for you in the questions below.

Q: What is the mean, standard deviation, and median of the number of followers that your followers have?

Answer

```
#!/usr/local/bin/python3
#import necessary libraries
import pandas as pd
from twarc.client2 import Twarc2
import matplotlib.pyplot as plt
import itertools

def auth():#authorization
    bearer_token="
   AAAAAAAAAAAAAAAAAAAAAJ2AUAEAAAAALRZWrij70cG9YMLEBV1l2uq7aPY%3
   Ddacuw3kcsTgmfyCnftWr8tM98XNOGof3lnxuXCh6W9c6Rlp0CJ"
    tw = Twarc2(bearer_token=bearer_token)
    return tw
def plot_data(df,fwfing):
    plt.rcParams['figure.figsize'] =40,25
    plt.rcParams['font.size'] = 17;
    plt.plot(df.USER,df.FRIENDCOUNT, label="FriendCount",marker='o')
```

```
    plt.grid(True)
    plt.xticks(rotation=90)
    plt.xlabel(xlabel = "friends", fontsize=20)
    plt.ylabel(ylabel= "no. of friends", fontsize=20)
    plt.legend(loc=6,fontsize=25);
    plt.gca().invert_xaxis()
    return plt.show()
def df_replacement(df):
    df = df.sort_values(by="FRIENDCOUNT")
    #create a list of new users
    newUser= []
    try:
        for x  in range(len(df)):
            c = str(x)
            newUser.append(c)
        # replace this new list to the column user in the dataframe
        df.USER = newUser
    except Exception as p:
        print(p)
    return df
def get_number_of_followers(id,i):
    #get user
    user = auth().followers(id)
    try:
        if i == 0:
            fwers_flowing = user.followers_count
    except Exception as p:
        print(p)
    return user.screen_name,fwers_flowing
def getUserFollowers(screen_name):
   users = []
   followersC = []
    #followers

   p= auth().user_lookup(users=n,usernames=True)
   c =1
   for ps in p:
       #parse in 0 to get the followers count
       us, count = get_number_of_followers(ps,0)
       users.append(us)
       followersC.append(count)
       c=c +1
       if(c > 179):
           break;

   prod = pd.DataFrame(list(zip(users,followersC)))
   prod.columns = ["USER", "FRIENDCOUNT"]
```

```
63    #sort and replace the names symbol fn
64    prod = df_replacement(prod)
65    return prod
66 screen_name = "iAchieveODU"
67 try:
68
69    f = getUserFollowers(screen_name)
70    #print the mean, std and median
71    print("The mean is {}".format(f.FRIENDCOUNT.mean()))
72    print("The std deviation is {}".format(f.FRIENDCOUNT.std()))
73    print("The median is {}".format(f.FRIENDCOUNT.median()))
74    #print as a csv file
75    f.to_csv("followers.csv",index=False)
76    #plot the graph
77    plot_data(f,0)
```

Listing 2: twitterFoll.py

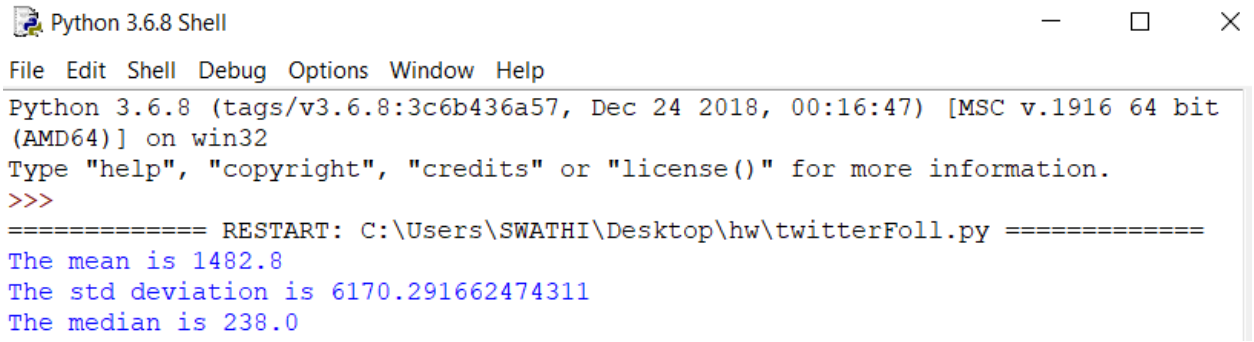

Figure 3: Mean, standard deviation and median

Discussion

I created a function getUserFollowers and used the screen name iAchieveODU and retrieved all followers id.

I passed the user id to get_number_of_followers and this returns the followers count of the associated user id.

I parsed the generated list of users and followers count as pandas dataframe to df_replacement and performed similar operation as Q1 and got the dataframe and found the values of the mean, standard deviation and median.

The mean is 1482.8

The std deviation is 6170.291662474311

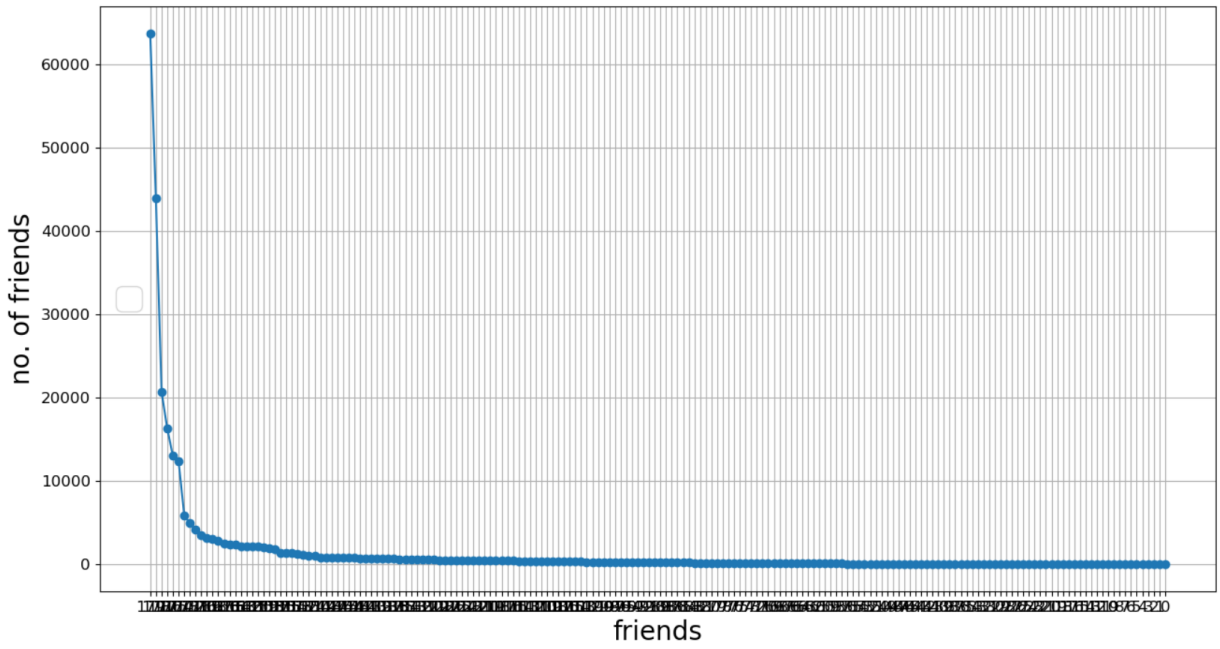

Figure 4: Plot for Q2

The median is 238.0

And then plotted the graph as shown in Figure 4.

Q: Does the friendship paradox hold for you and your followers on Twitter?

Ans. Comparing the mean of iAchieveODU's followers followers count 1482.8 to iAchieveODU's followers count of 176. Those that they follow have a higher followers count than iAChieveODU has.

References

- Twitter, `https://scholarslab.github.io/learn-twarc/06-twarc-command-basics`
- Github , `https://github.com/DocNow/twarc`
- Twitter data , `https://twittercommunity.com/t/downloading-friends-from-a-list-of-users/155024`
- Matplotlib, `https://matplotlib.org/3.2.2/gallery/misc/plotfile_demo_sgskip.html`
- Matplotlib Subplots, `https://matplotlib.org/stable/gallery/subplots_axes_and_figures/invert_axes.html`