



ASSIGNMENT 4

CS 432 Web Science

Mackenzie Kerchner

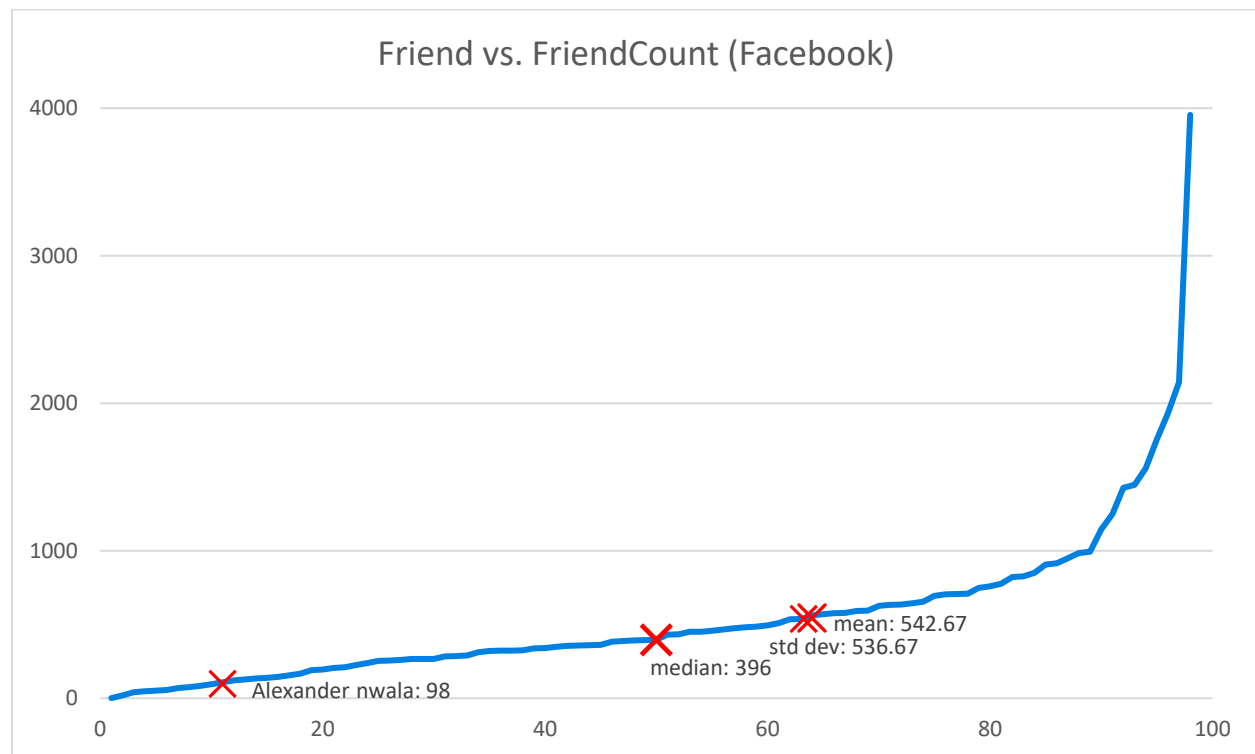
Contents

Problem 1	2
Problem 2	3
Problem 4	5

Problem 1

1. Determine if the friendship paradox holds for my Facebook account. Compute the mean, standard deviation, and median of the number of friends that my friends have. Create a graph of the number of friends (y-axis) and the friends (x-axis) themselves, sorted by number of friends (y-axis). (The friends don't need to be labeled on the x-axis: just f1, f2, f3, ... fn.) Do include me in the graph and label me accordingly.

Solution



USER	"FRIENDCOUNT"
Sharon Owens	1
Osakpolo Udeh	20
Chukwuemeka Udeh	40
Ernst Wilson	48
Eric Akers	51
Sarah-hope Tebira	55
Anne Garland	68
Kingdom Nwala	76
Tilman KÄhner	85
Joel Perry	94
Ramesh Raj	108
Lovelyn Chigozilem Godknows	122
Shelby Thomas	127
Dennis Tuburu	135
Alex Dohrn	139
Melody Elenwo	146
Victor Nwala	155
Ogbondah Anele	167
Joy Chinwe Nalley	191
Jose Antonio Olvera CaÃtizes	195
Peace Nnenda	205
Fortune Tall Essien	210

I plugged the list of friends into Microsoft excel and sorted lowest to highest. Then I added in Professor Nwala himself as a standalone data series. Then I used Excel formula functions to calculate the mean, median, and standard deviation for the list of friends. I used the population version of the std. deviation formula since the list of friends is its own population and not a sample of a larger one. The friend paradox does in fact hold for Professor Nwala's Facebook account, as he has fewer friends than his friends do.

mean:	std dev:	median:
542.67	536.67	396

Problem 2

2. Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use "followers" as value you measure (i.e., "do your followers have more followers than you?"). Generate the same graph as in question #1, and calculate the same mean, standard deviation, and median values. If you do not have followers on Twitter (or don't have more than 50), then use my twitter account "acnwala".

Solution

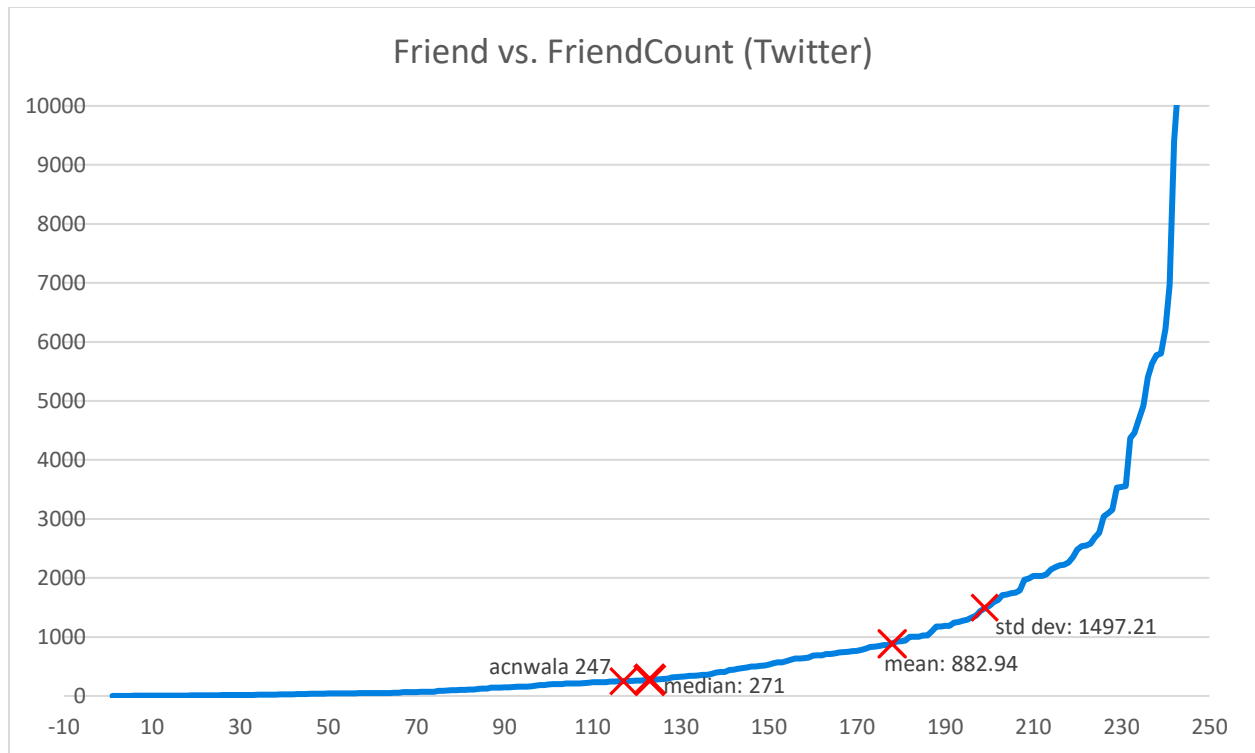
```
from tweepy import OAuthHandler
import tweepy
import time
import twitter_cred

auth = OAuthHandler(twitter_cred.CONSUMER_KEY, twitter_cred.CONSUMER_SECRET)
auth.set_access_token(twitter_cred.ACCESS_TOKEN, twitter_cred.ACCESS_TOKEN_SECRET)
api = tweepy.API(auth)

primary_friend_count = 0
friend_list = []
filename = "followerlist.csv"
screenName = "acnwala"
with open(filename, 'w'):
    pass # clear past entries into file
for friend in tweepy.Cursor(api.followers, screen_name=screenName, count=200).items():
    try:
        primary_friend_count += 1
        output = friend.screen_name + "," + str(friend.followers_count)
        print output
        with open(filename, "a") as outfile:
            outfile.write('%s\n' % output)
    except tweepy.TweepError:
        print "rate limit reached, please wait for reset"
        time.sleep(901) # limit is 5000 requests per 15 minutes
    except:
        print "unknown error"
output = screenName + "," + str(primary_friend_count)
print output
with open(filename, "a") as outfile:
    outfile.write('%s\n' % output)
```

aminabah	0		
UdochiNw	0		
nwalpin	0		
sahelno	2		
BabithaBc	2		
suleiman	3		
IRI_2019	3		
jsayedi12	3		
janesodea	3		
SirigerePr	4		
RyanOkta	4	knowtheory	4456
AijazAh20	4	ianmilligan1	4690
Phoneriur	5	meangrape	4919
amnsaleh	6	ResearchBuzz	5401
mowdydo	6	EbenezarWikina	5632
AntonJRas	6	edsu	5777
akkshirsag	6	silvertje	5805
SaligariTw	7	NLM_DIMRC	6221
DanMilan	8	ahidamzair	6968
JanaeJack	9	james3neal	9403
Okereke	9	hrbrmstr	10461
network2	10	NiekoGloss	21542
epicurean	12	zittrain	44419
Vendett2	12	blossomozurumba	44836
Mohamm	13	thepamilerin	95590

I used professor Nwala's twitter as I have no followers. I used the documentation located at [1] to refresh on how the cursor worked, my original method to generate a follower list made it difficult to determine when the end of the list occurred. I used the documentation for twitter user objects located here [2] to get follower counts rather than the one Professor Nwala provided, I prefer object.information to a get function. Once all the data was gathered and outputted to a .csv file, I opened it up in excel, sorted it by smallest to largest, and moved it to a copy of the excel file for part 1. This was much easier than remaking the entire graph from scratch. There was a problem with the data in that there were 4 enormous outliers at the end of the list that made the graph unreadable. I opted to cut out the 4 users with the largest number of friends and adjusted the calculations and graph accordingly. The graph now has the x axis bound to 250, and the y axis bound to 10,000. The friendship paradox holds for Professor Nwala's Twitter account as well.



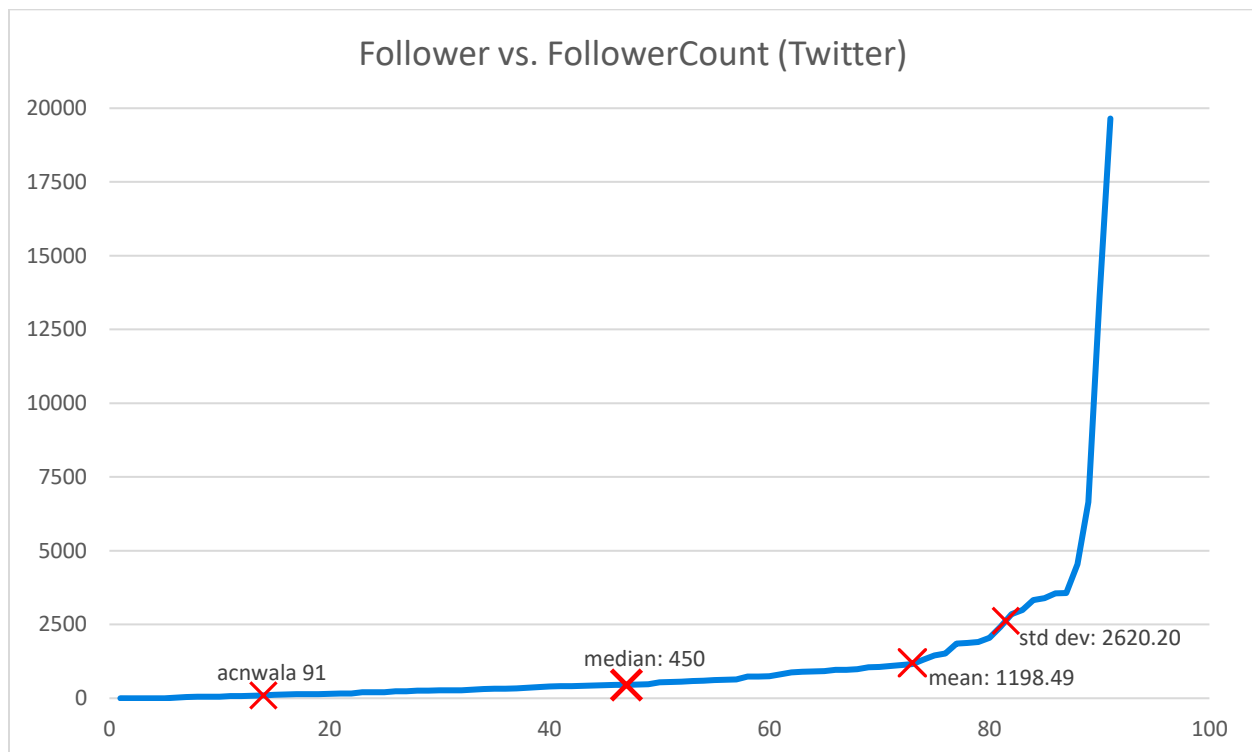
Problem 4

Repeat question #2, but change "followers" to "following"? In other words, are the people I am following following more people?

Solution

```
filename = "followinglist.csv"
for friend in tweepy.Cursor(api.friends, screen_name=screenName, count=200).items():
    output = friend.screen_name + "," + str(friend.friends_count)
output = screenName + "," + str(api.get_user(screenName).friends_count)
```

By changing 4 lines of code (Tweepy is radical) I was able to grab users following rather than followers. No outliers this time, same strategy and similar results to problem 2.



References

- [1] <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object>
- [2] http://docs.tweepy.org/en/v3.5.0/cursor_tutorial.html