CS532S19: Assignment #4

Due on Sunday, March 09, 2019

Nwala, Alexander C

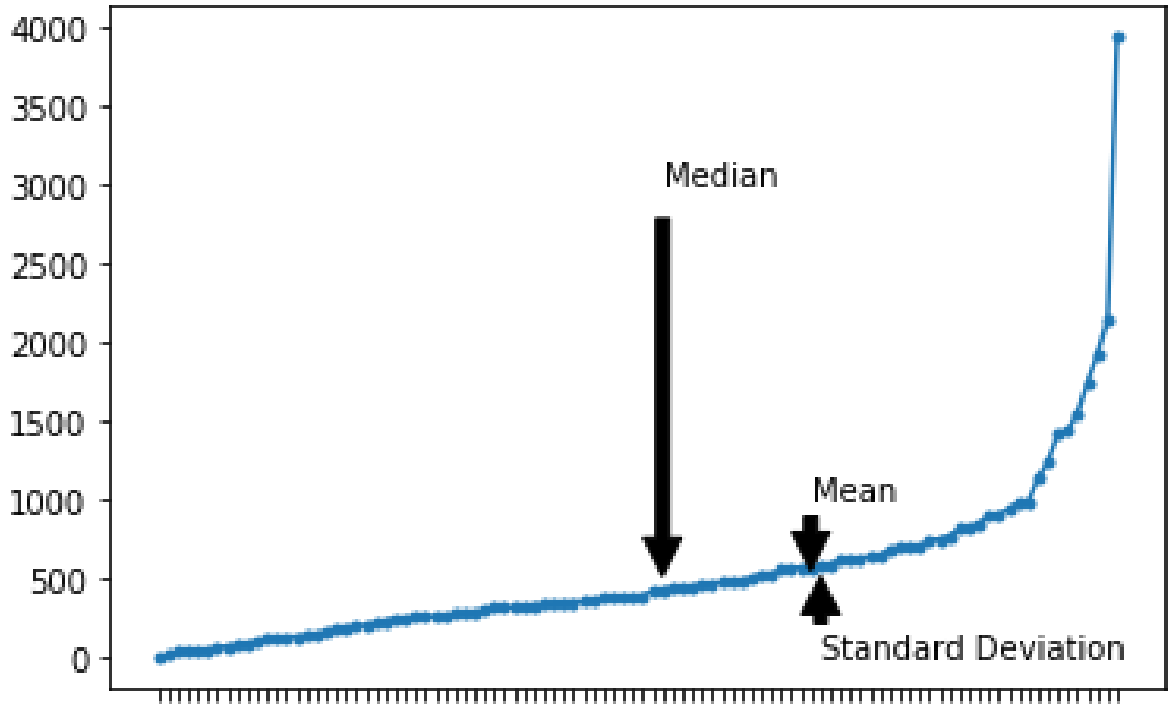
Giridharan Ganeshkumar

Question 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 dont need to be labeled on the x-axis: just fn. Do include me in the graph and label me accordingly.

Facebook Friends

1. Mean :542.6734693877551
2. Median :396.0
3. Standard Deviation :539.4337385239659



1. As shown in the below code, we draw a graph using Matplotlib. The graph is shown above we also mark the mean, median and standard deviation.
2. we make use of the statistics library to calculate the mean, median and standard deviation with mean(), median() and strdev() respectively.
3. we make use of the annotate method to mark the mean, median and standard deviation on the graph.

Listing 1: Python Script

```
import matplotlib
import matplotlib.pyplot as plt
import csv
from sklearn import linear_model
import seaborn as sns
```

Question 1 continued on next page...

```
import numpy as np
from statistics import *
from adjustText import adjust_text

friedsCounter = 0
datax, datay = [],[]

with open("C:\\acnwala_twitter_follows_count.csv") as acnwalafriends:
    reader_acnwalafriends = csv.reader(acnwalafriends, delimiter=',')
    next(reader_acnwalafriends)
    for row in reader_acnwalafriends:
        friedsCounter = friedsCounter + 1
        datax.append("f" + str(friedsCounter))
        datay.append(int(row[1]))


datay.sort()
print("Mean :" + str(mean(datay)))
print("Median :" + str(median(datay)))
print("Standard Deviation :" + str(stdev(datay)))
ax = plt.axes()
ax.plot(datax, datay,'.-')
ax.xaxis.set_major_formatter(plt.NullFormatter())
ax.annotate('Standard Deviation', xy=(83,2620.146952509964),xytext=(83,
    10000),arrowprops=dict(facecolor='black', shrink=0.05))
ax.annotate('Mean', xy=(80,1198.5054945054944),xytext=(80, 12500),
    arrowprops=dict(facecolor='black', shrink=0.05))
ax.annotate('Median', xy=(47,450),xytext=(47, 2500),arrowprops=dict(
    facecolor='black', shrink=0.05))
```

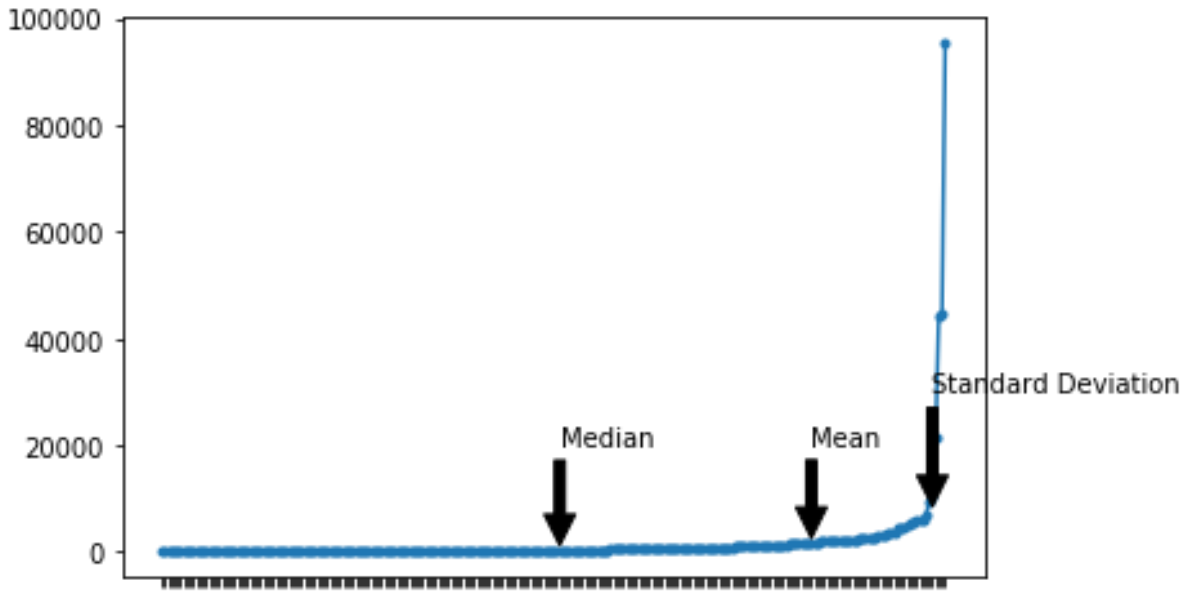
Question 2

Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use followers as value you measure. Generate the same graph as in question 1, and calcuate the same mean, standard deviation, and median values

Twitter Followers

1. Mean :1198.5054945054944
2. Median :450
3. Standard Deviation :2620.146952509964

Question 2 continued on next page...



1. We make use of the tweepy library to get the followers, for this problem we start by authenticating with the api and then use Cursor with api.followers and pass screen name as variable. Then get the followers count.
2. As shown in the below code, we draw a graph using Matplotlib. The graph is shown above we also mark the mean, median and standard deviation.
3. we make use of the statistics library to calculate the mean, median and standard deviation with mean, median and strdev respectively.
4. we make use of the annotate to mark the mean, median and standard deviation on the graph.

Listing 2: Python Script

```
import tweepy
from tweepy import OAuthHandler
import json
import urllib
import csv
import time

auth = tweepy.OAuthHandler('sXBU4507CXpRMqo229oBxDzPP', '
    U9HiwDJVtLz57vpb1jRbfr5gVlrMRARf2YEjeD9klE6Sjiyayq')
auth.set_access_token('755808986977501184-5U4n5l6ioaak7Ja5Lb0n8ZgADCCnQaW',
     'xj1PjwholoS0e4eeSTSlb8QdlgJoYhxKhlxNlDeeLO5G4')
api = tweepy.API(auth)

users = tweepy.Cursor(api.followers, screen_name="acnwala").items()
with open("C:\\acnwala_twitter.csv", "w", newline='') as acnwala_twitter:
    writer = csv.writer(acnwala_twitter)
    currentData = ['UserName','FollowersCount']
    writer.writerow(currentData)
    while True:
        try:
            user = next(users)
        except tweepy.TweepError:
```

Question 2 continued on next page...

```
            print(tweepy.TweepError)
            time.sleep(60*15)
            user = next(users)
        except StopIteration:
            break
        currentData = [user.screen_name,user.followers_count]
        writer.writerow(currentData)
print('Completed.')
```

Question 3

Repeat question 1, but with your or a specified LinkedIn profile

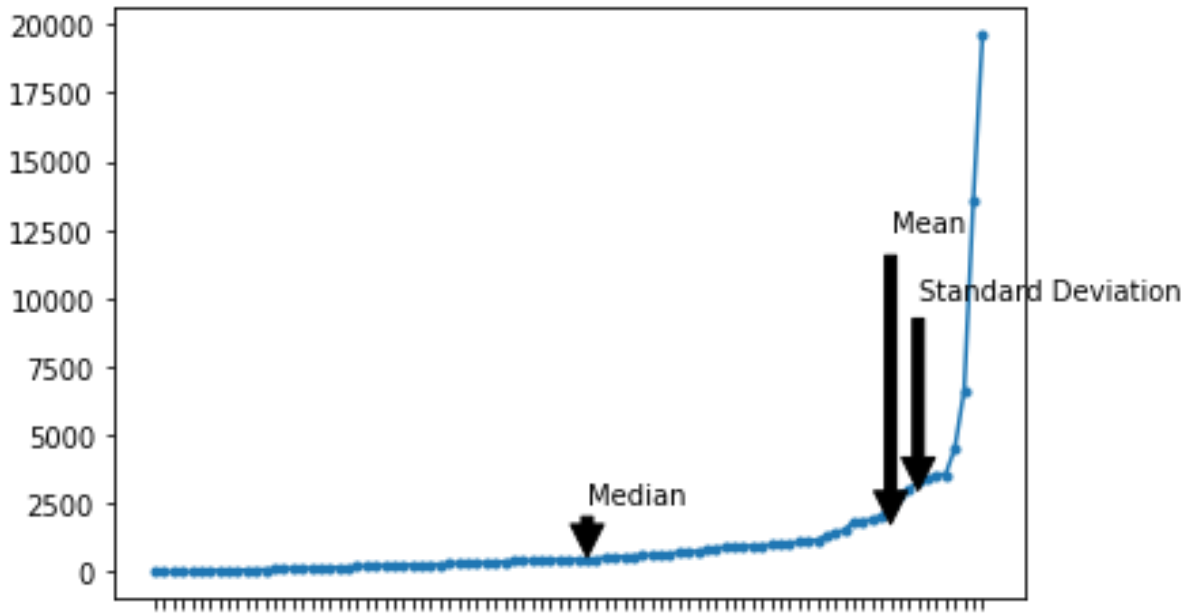
1. I did attempt this question with linkedin library but unfortunately couldnt get the api to authenticate due to python version that i was using, it failed with except (requests.ConnectionError, requests.HTTPError), error.

Question 4

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

Twitter Following

1. Mean :1703.9514170040486
2. Median :277
3. Standard Deviation :7432.790795237021



Question 4 continued on next page...

1. We make use of the tweepy library to get the following, for this problem we start by authenticating with the api and then use Cursor with api.friends and pass screen name as variable. Then get the firends count.
2. As shown in the below code, we draw a graph using Matplotlib. The graph is shown above we also mark the mean, median and standard deviation.
3. we make use of the statistics library to calculate the mean, median and standard deviation with mean, median and strdev respectively.
4. we make use of the annotate to mark the mean, median and standard deviation on the graph.

Listing 3: Python Script

```
import tweepy
from tweepy import OAuthHandler
import json
import urllib
import csv
import time

auth = tweepy.OAuthHandler('sXBU4507CXpRMqo229oBxDzPP', '
    U9HiwDJVtLz57vpb1jRbfr5gVlrMRARf2YEjeD9klE6Sjiyayq')
auth.set_access_token('755808986977501184-5U4n5l6ioaak7Ja5Lb0n8ZgADCCnQaW',
     'xj1PjwholoS0e4eeSTSlb8QdIgJoYhxKhlxNIDeeLO5G4')
api = tweepy.API(auth)

users = tweepy.Cursor(api.friends, screen_name="acnwala").items()
with open("C:\\acnwala_twitter_follows_count.csv", "w", newline='') as
    acnwala_twitter:
    writer = csv.writer(acnwala_twitter)
    currentData = ['UserName','FollowsCount']
    writer.writerow(currentData)
    while True:
        try:
            user = next(users)
            #users_current = tweepy.Cursor(api.friends, screen_name=user.
                screen_name).items()
        except tweepy.TweepError:
            print(tweepy.TweepError)
            time.sleep(60*15)
            user = next(users)
        except StopIteration:
            break
        currentData = [user.screen_name,user.friends_count]
        writer.writerow(currentData)
print('Completed.')
```