

# Homework #4

## Task 1: Collecting Twitter Data

Sample:

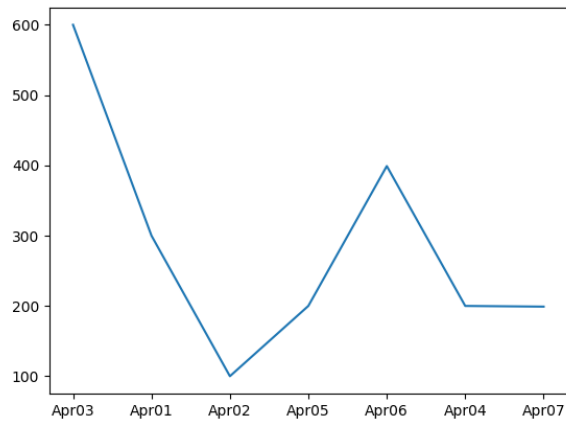
Collect order:

1. set date
2. set keywords = donald trump or bernie sanders
3. set extract count

loop 1-3 until 1000 tweets for each keyword, twitter allowed only 140 requests per 15 minus  
based on 1000 tweets, read each Jason file and store file without RT to Non-RT directory.

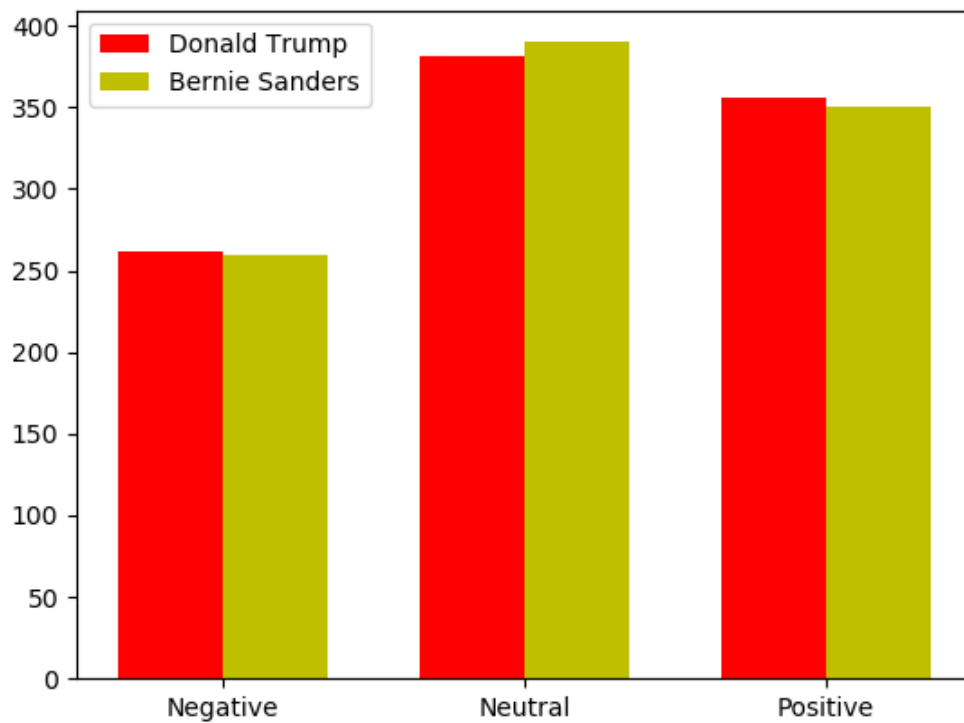
## Task 2: Exploratory Analysis

### 1. time series figure



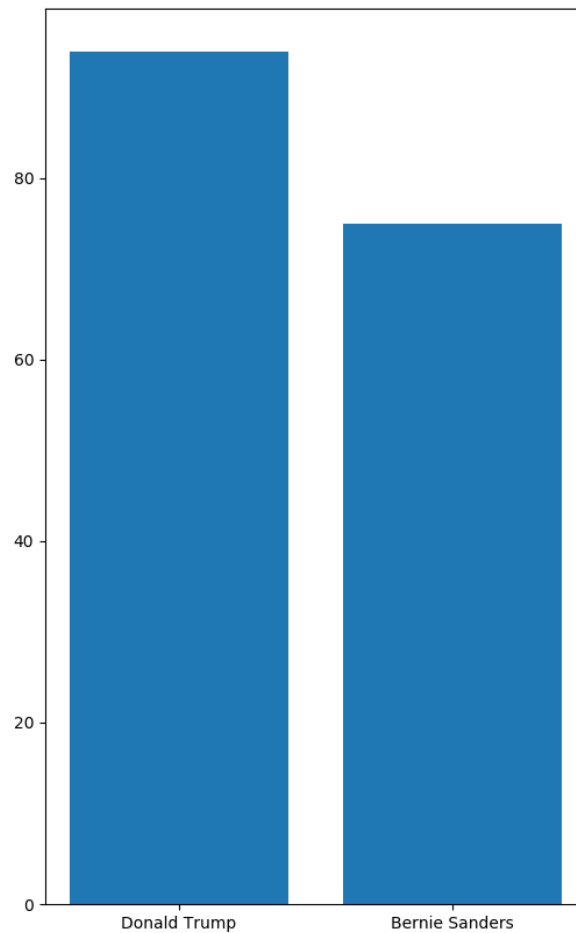
### 2. Trending detects

Extract each tweet (including retweet), break them sentences by sentences. Using textblob Library to score each sentence, the range of score is  $(-1,1)$ . Less than zero means that the sentence is negative and equal to zero is neutral and greater than zero is positive. Average each sentence's score and we can get the score for that tweets. Based on that I Draw following Figure:



3. Attention:

Each tweet's publisher has follower. The more follower a user has, means the more Influence that user has. Which means that if we count the number of those have large follower(Greater than 10k) we can make the conclusion that who( between Bernie Sanders & Donald Trump) draws more attention. Based on that I draw following figure:



Assumption: Donald Trump. He draws more attention than Bernie Sanders and people generally have more positive view to him. (even a little more)