**Homework#1**

**Task 1:**

**Steps that I have followed to collect twitter data:**

**Step 1:**

Using my twitter account, I collected tweets satisfying the three conditions mentioned in the question.

The language and API that I used to get the tweets were Python and Tweepy respectively.

**Step 2:**

Search function of the API is used to extract tweets.

**Step 3:**

Finally, I stored the result in JSON format.

Below I am pasting the format. The total number of tweets I got was 7100.



Task 2:

Step 1: Converted obtained JSON into required attributes which was later converted into csv file.

Step 2: Sentiment analysis was done in the next step using Python.

Step 3: The procedure was to split the tweet into words and matching them with the positive and negative words from the <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>. and generated the score of the sentiment for each tweet.

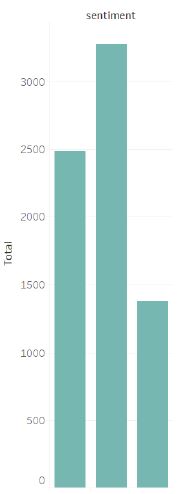
Score <0, tweet is negative.

Score =0, tweet is neutral.

Score >0, tweet is positive.

|  |  |  |
| --- | --- | --- |
| Tweet | Score | Sentiment |
| Donald Trump Unveils Plans for White House https://t.co/eG98S3lT2v #donaldtrump #republicans | 1 | Positive |
| Me when @realDonaldTrump says #DressLikeAWoman #resistance #DonaldTrump #thisisnotnormal #WomensRights https://t.co/nOFF3Xn7u3 | -1 | Negative |
| RT @FoxNews: Here's what President #DonaldTrump did in week two of his presidency. https://t.co/glJLO2tlPA | 0 | Neutral |

Bar chart representing tweets:



First bar: Neutral tweets

Second bar: Positive tweets

Third bar: Negative tweets

**Observation:**

From the figure, it is very clear that majority of the tweets were positive and neutral.

Task 3:

Using the elbow method, I have taken k value as 3.

**Reasons to choose 3 as the k value:**

It is very clear from the below graph that 3 clusters are the best for the analysis.

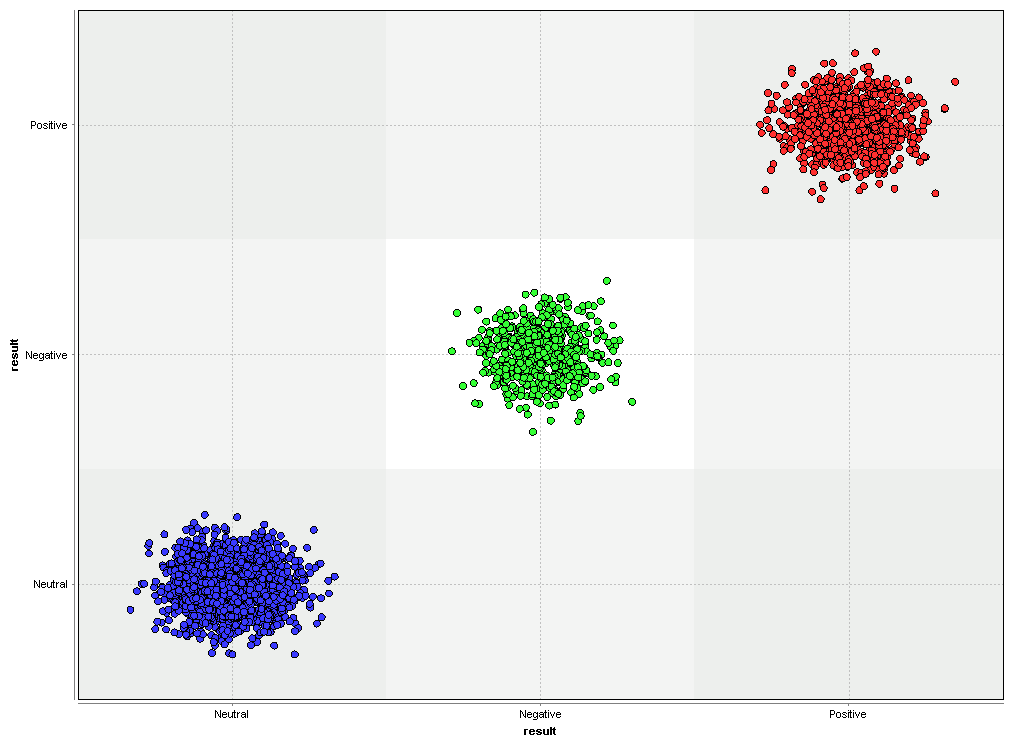
In addition, from the sentiment analysis, I just found the following three sentiments (positive, negative, neutral).

The clustered three parts data is shown below.

**Elbow Graph:**



NOTE: K-means algorithm was used to divide the data into three clusters.



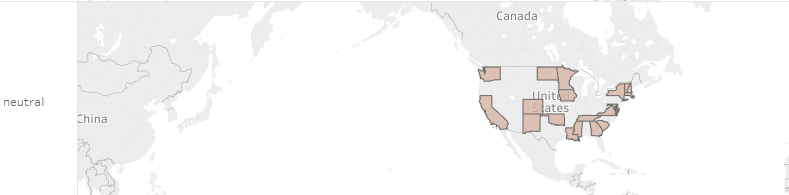
Later, I used tableau to distribute the tweets on map based on the sentiment information obtained.

**Observation:** Majority of the tweets were positive and neutral.

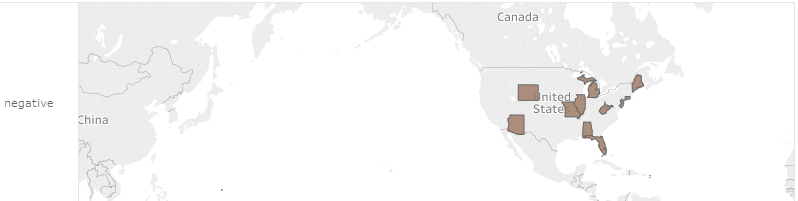
**POSITIVE**



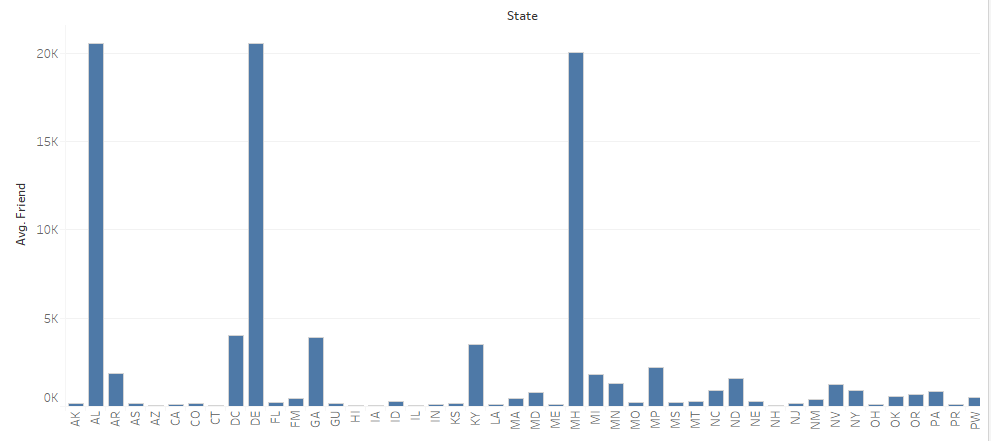
**NEUTRAL**



**NEGATIVE**



**ONE LAST INTERESTING ANALYSIS:**



I figured the average of friends in each state.

**Observations from the figure:**

There are very few states with a big number for average count of friends.

Majority of the states in United States of America, have the average friend count to be less than 5k.