Final Project Report – US Election 2020 Sentiment Analysis

* 1. **Introduction:**

Recently, US 2020 Elections were held, and the main contestants were Joe Biden from Democratic party and Donald Trump from Republican Party. I wanted to do an Analysis on Twitter data to get result of who will be the possible winner of US 2020 Elections.

* 1. **Problem Statement:**

1. Is Trump liked more or Biden?
2. Trump vs Biden Tweet count in India?
3. Trump or Biden, who has more positive tweets?
   1. **Related Work:**

* 1. **Methods:**

Tools used –

* + - I worked with Python Programming Language in IDE Google Colab. I used Pandas for Reading the Data set and also for Data Transformation, NumPy also for data Transformation. Matplotlib for Data Visualizations.

Load Data –

* + - First, I loaded the Two data sets “Trump Data”, “Biden Data” into two different data frames.

Data Transformation –

* + - I drop () function to drop few columns in the data set as they were not useful for my analysis.
    - I used loc [] method to get only the data of Trump in Trump Dataset and Biden in Biden Data set.
    - I used Concat () method to mix both the data frames into one data frame.
    - I used Sort\_values () method to sort the data frame on the column “created\_at” in ascending order.

Data Visualization –

* + - Xtick () function is used in bar plot to keep labels in x axis. Here I used Xticks for labeling Donald Trump and Biden on x-axis.
    - Used Bar plot from Matplotlib for this.
    - Here, I used default blue color as we don’t have many variables to compare to.

Chart, bar chart

Description automatically generated

* + - Tweets count of each candidate is visualized through pie chart. Here I defined a function to calculate the pie chart with input of country name.



Chart, pie chart

Description automatically generated

* + - Orange and Blue colors differentiate each candidate in pie chart.

Data Cleaning:

* + - In order to do Sentiment Analysis we have to remove the noise in the tweets like punctuation, numbers, stop words, spaces and convert them to lower case.

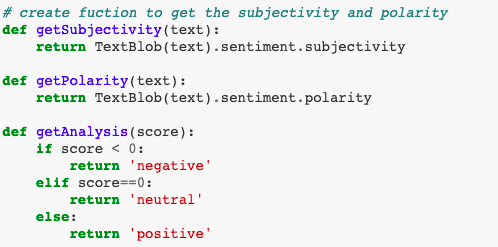
Graphical user interface, text, application, email

Description automatically generated

* + - Dropped the Null Values by function dropna()

Sentiment Analysis:

* + - I have created functions to calculate the subjectivity score, polarity score and analysis based on score to find the sentiment of the tweets.
    - Here I had used NLTK, TextBlob for finding the scores.

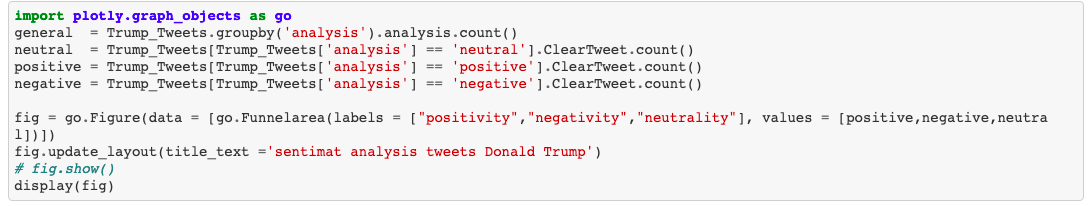


* + - Imported Counter from Collections library to count the values of each sentiment positive, negative and neutral.

Text

Description automatically generated

* + - I imported graph\_objects from plotly library to make a funnel area chart.



* + - Funnel Area is used to show data in different stages.

Chart, funnel chart

Description automatically generated

Chart, funnel chart

Description automatically generated

* + - We can clearly see from the chart that Trump has 39.8%, Joe Biden has 45.5% positive tweets.
  1. **Results:**
  2. **Discussion:**

What has the audience learned about visualization from your work?

* 1. **Future Work:**
  2. **References:**
  3. <https://medium.com/swlh/data-visualization-us-election-2020-a3d3531b611c>