INFO 5082

ROHAN RAJ MALAKPETA

11377126

**Sentiment Analysis on Presidential Elections 2020 Twitter Data.**

**INTRODUCTION**

Sentiment analysis is utilized by various information investigation organizations on an assortment of subjects. A portion of the well-known business sectors where it is utilized are:

Business: The advertising group of numerous organizations use it to make business techniques, to see how clients see their item and to comprehend client conduct to improve deals.

Political issues: In the political field, it is utilized to monitor districts where the up-and-comer is positive and work towards locales where the applicant isn't ideal to improve their odds in a political race.

The main role of this project is to comprehend the nature of the political talk that occurred on Twitter during the decisions as far as conclusion, as often as possible referenced terms, and mainstream tweets/retweets. Client credits, for example, the quantity of adherents and companions, span of action on Twitter, number of messages to date and notoriety were additionally used in such manner.

Sentiment analysis is a method of dissecting popular assessment on an issue. The official political decision in the United States is a hot issue that will influence different parts of the world. The objective of this analysis is to gauge the result of the US official political decision and to contrast these outcomes and the real aftereffects of the surveys. The sentiment analysis utilized in this examination is the dictionary-based sentiment analysis. The strategy utilized in this exploration is information assortment, information pre-processing, information planning and sentiment analysis. The information in this examination was acquired from Twitter required multi-week before the United States official political race was held. The model utilized in this exploration is VADER sentiment analysis. The information cleaning component in this investigation utilizes a strategy in text mining, where the information is first cleaned of different things that are not viewed as significant in the analysis.

**STATEMENT PROBLEM**

The methodology comprises of following the steps like usage of search terms like “Trump”, “JoeBiden” and “Election2020” to gather the Twitter data. Data Cleaning and Extraction, sentiment tagging and classification of gathered tweets. Importing data in MySQL database to perform exploratory data analysis. Finally, I would like to develop a user behaviour model and formulate the hypotheses.

**REVIEW OF LITERATURE**

Interest in mining sentiment and assessment in texts has risen consistently throughout the most recent decade principally due to the expanded accessibility of data and closely-held conviction messages. All in all, sentiment analysis is utilized to make predictions or measures in different fields like the financial exchange, governmental issues, and surprisingly friendly developments. old investigations of legislative issues on social networks have wound up lacking or in any event, lacking data since they can just take a little example. In this examination we will make an analysis of how precise the consequences of this investigation are and contrast them and the choice that has been acquired. This sentiment analysis research is centered around assembling sociology training with specific quantitative philosophy: our methodology joins educated and correlative real-time information assortment and predictive sentiment modeling, and comprehension using Twitter of the social and political practices at work.

**OBJECTIVES OF THE STUDY**

The main objective of the project is to collect the tweets and to ascertain if there is a correlation between the sentiment of users on Twitter. I would like to investigate whether can we predict the election outcome of each state, can we predict the candidate from tweet text only.

**DATA COLLECTION**

Tweets are collected using the Twitter API I have created Twitter developer account and imported the install Tweepy.

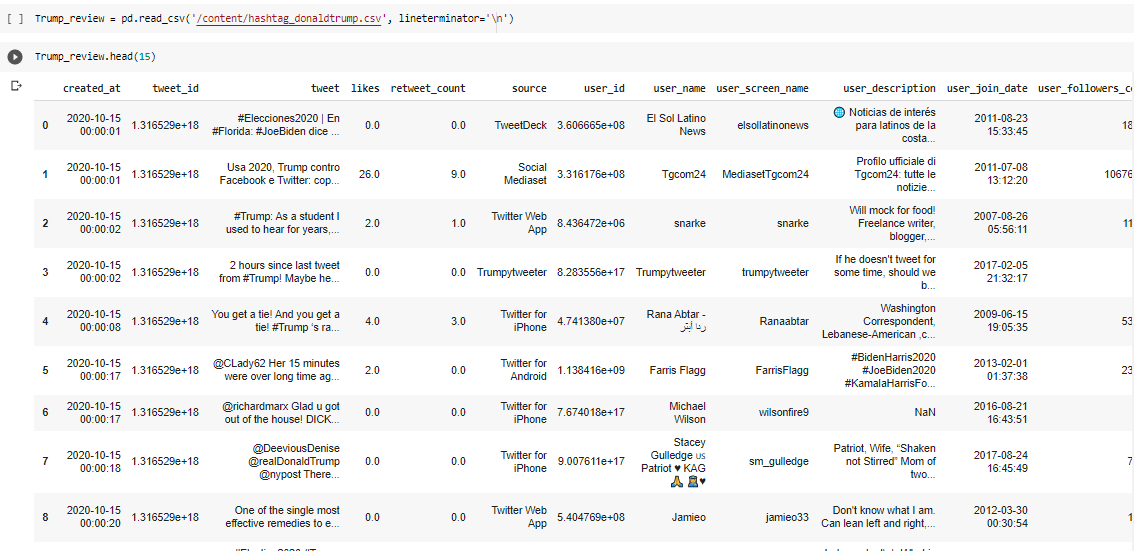
The tweets were collected from the timeframe October 15th 2020 and November 4th 2020. The columns include created\_at, tweet\_id, tweet, likes, retweet\_count, source, user\_id, user\_name, user\_screen\_name, user\_description, user\_join\_date, user\_folllowers\_count, user\_location, latitude, longitude, city, country, state, state\_code and collected\_at.

The dataset has been obtained from Kaggle and I have also extracted the tweets using the Twitter API account. The total tweets collected were 355,000. The total attributes are ten in both the datasets.

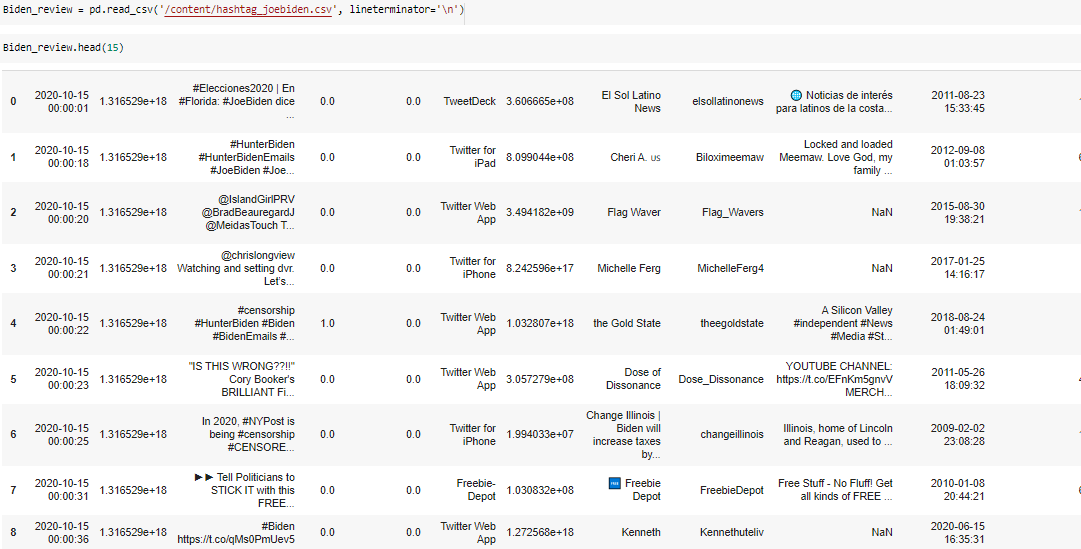
<https://www.kaggle.com/manchunhui/us-election-2020-tweets>

**EXPLORATORY DATA ANALYSIS**

**Importing the Trump tweets dataset**

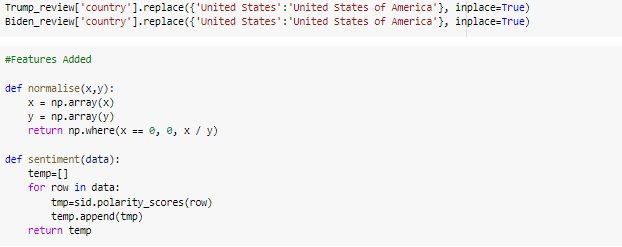


**Importing the Biden dataset**

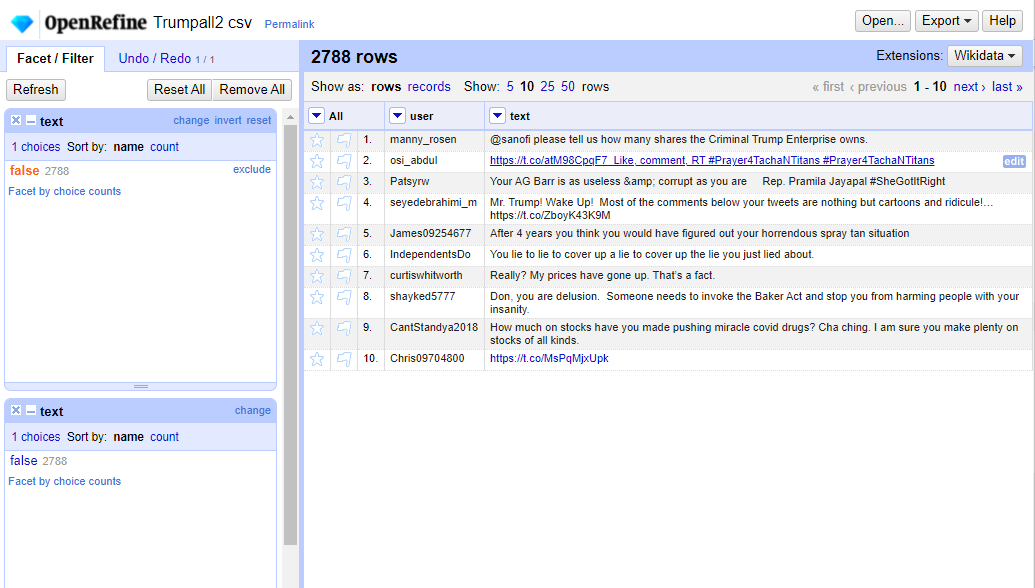


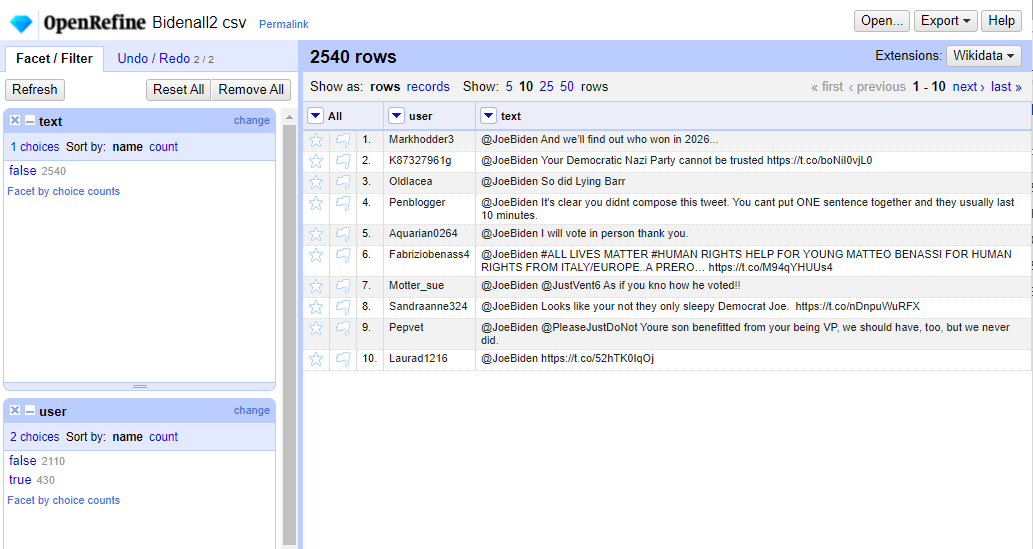
**Data Cleaning**

There are some null values and I have added features like polarity to test the tweets nature like which are positive negative and neutral.



I have used Open Refine tool to clean the data and observed that there are some null values associated with the columns users and text.

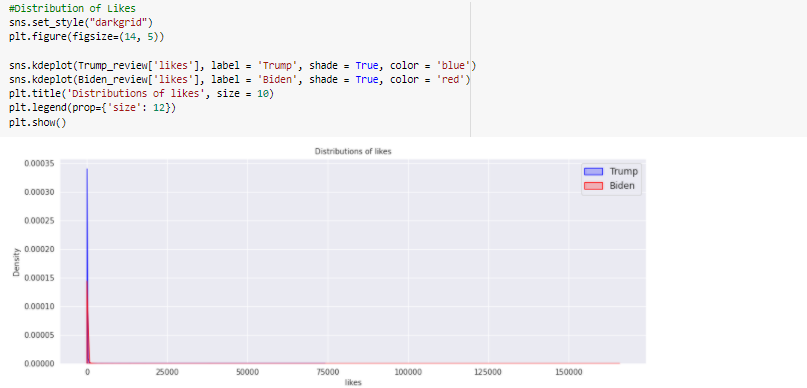


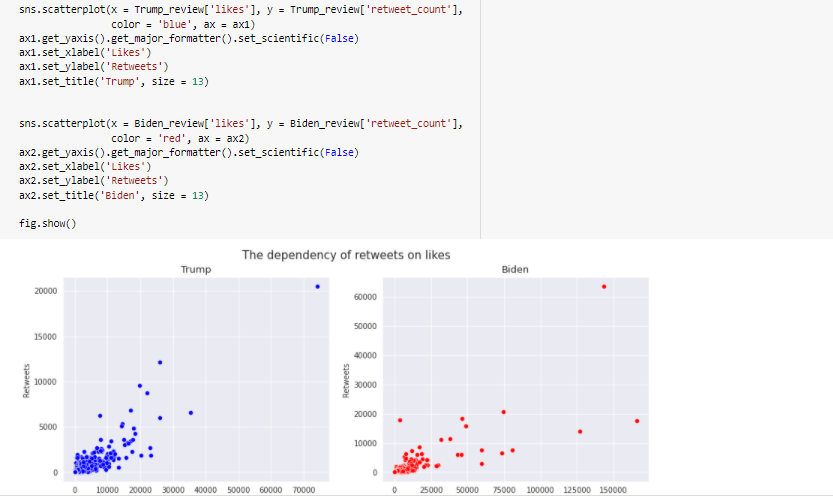


Trump tweets were more in count when compared to Biden.

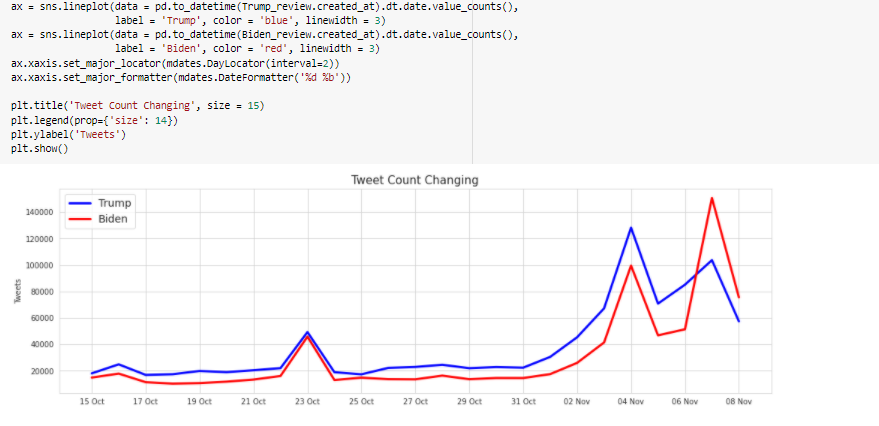


The distribution of likes were more for Trump than compared with Biden.





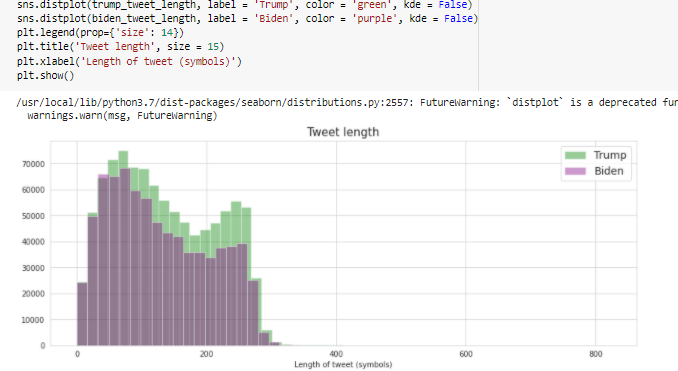
There is a very strong positive correlation between tweets and retweets.



we could see that the tweet count has been drastically increasing for Biden when compared to Trump.

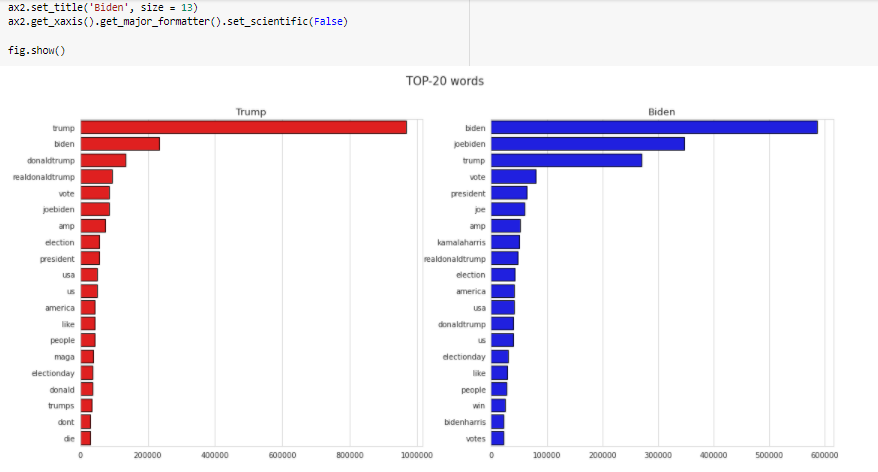
**Data Analytics**

**The length of Tweets**



The distribution of length of tweets were mostly same for both the candidates.

**Finding the top 20 words**

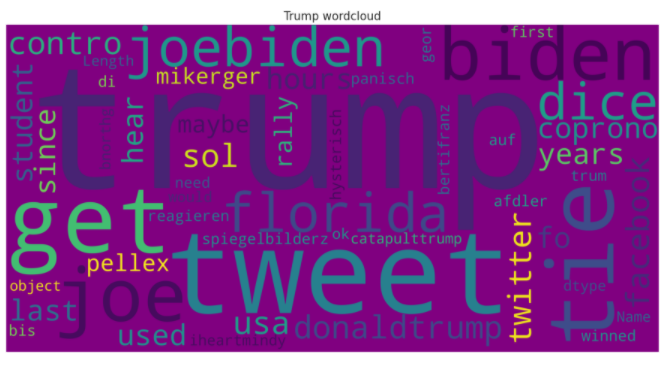


Most frequent words in the Trump tweets are vote, election, president, people, Election day and there are some specific words like MAGA or die.

The tweets dedicated to Biden are win, Kamala Harris, BidenHarris.

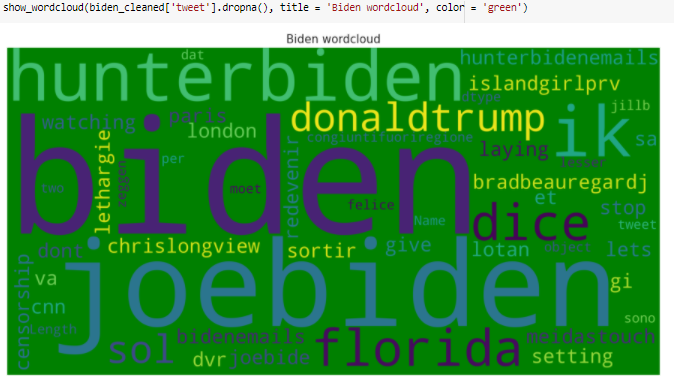
**Trump Tweets Word Cloud**

The most frequent words in the trump tweets are trump, get, tweet, florida, twitter.



**Biden Tweets Word Cloud**

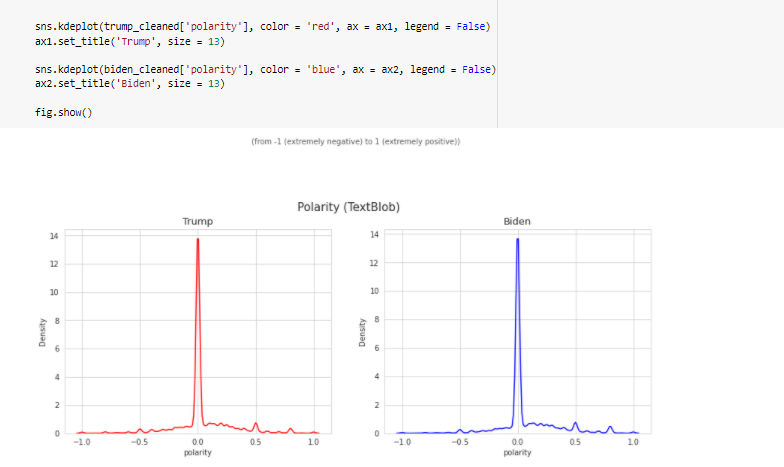
The most frequent words in the Biden Tweets are hunterbiden, joebiden, florida, donaldtrump, dice.



**Data Visualization & Results**

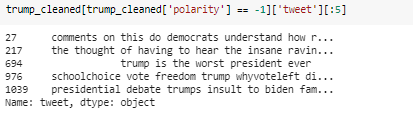
**Sentimental Analysis using TextBlob**

TextBlob is a python library for processing textual data, it provides simple API with polarity that ranges from -1 to +1 and will tell whether the text has negative sentiments or positive sentiments.

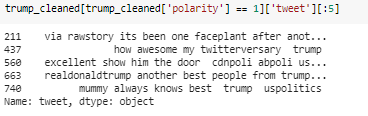


**Classification of Tweets**

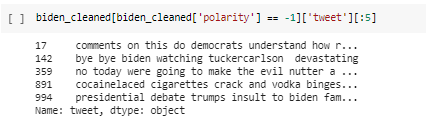
**Trump Negative Tweets**



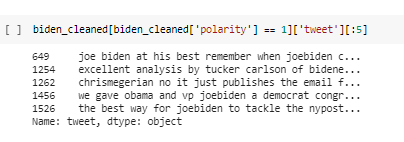
**Trump Positive Tweets**



**Biden Negative Tweets**



**Biden Positive Tweets**

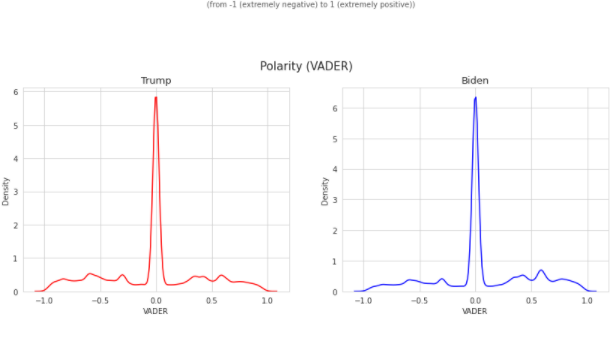


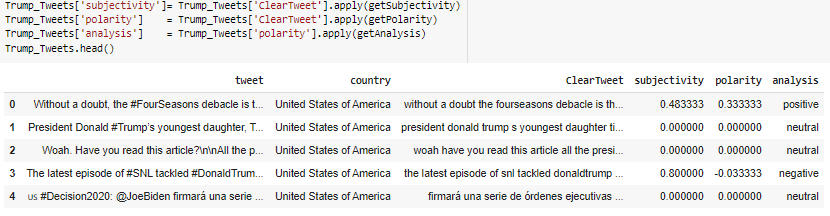
**Sentimental Analysis using Vader Analyzer**

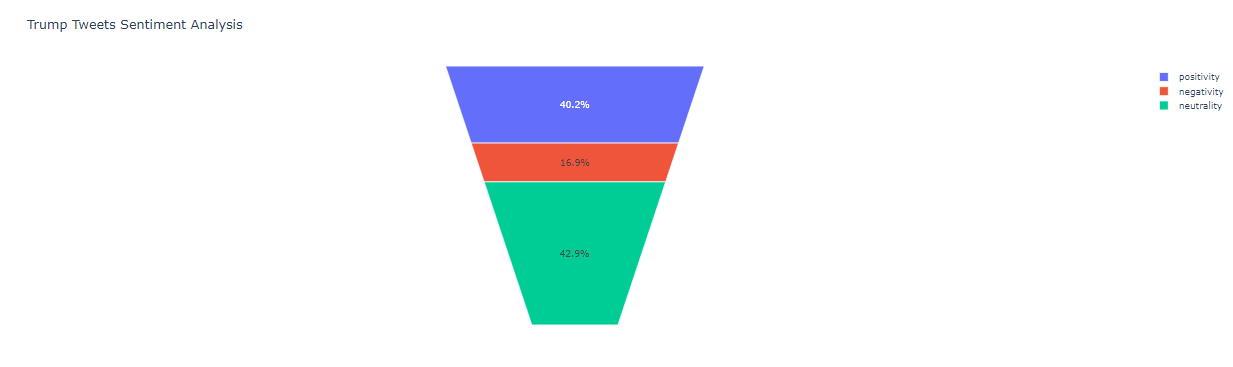
Vader is valence aware dictionary for sentiment reasoning it is a model that is used for text sentiment analysis and which is too sensitive to both polarity and intensity of emotion.

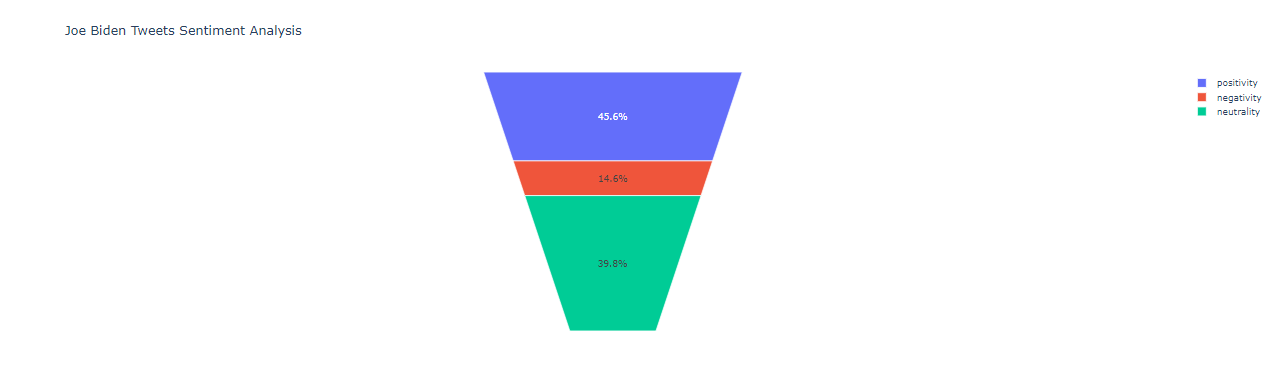
Vader sentimental analysis relies on dictionary that maps lexical features to emotion intensities that are known as sentiment scores. VADER sentiment analysis depends on a word reference that maps lexical highlights to feeling forces known as sentiment scores.

The estimation score of a book can be gotten by summarizing the power of each word in the content. The sentimental analysis of VADER depends on a word reference planning lexical attributes to passionate powers known as estimation scores.









With tweets from 40 unique dialects, yet with an enormous extent of tweets in English and beginning from the US, there is interest in US elections from various nations on the planet. The estimation examination was directed uniquely on information from the "United States of America" that had geo-information to attempt to build up the conclusion in every dataset and thus each official applicant. A huge greater part of states was moving to a "Positive" sentiment score for the majority rule chosen one from the beforehand more "Neutral" sentiment while dissecting sentiment at the state level as we moved toward the political race date. Though a few states with the Republican applicant are presently moderately Neutral. When seeing the sentiment examination from an information viewpoint, this example is correctable.

**CONCLUSION**

From the exploration above, it is obvious that the use of assumption investigation in dissecting the aftereffects of the 2020 US presidential election has high precision, the use of slant examination is additionally amazing in breaking down any information that has information as text or sentences which obviously requires an examination in it, in this exploration information We get it from Twitter, perhaps for additional examination it tends to be finished using information as overviews or other online media stages. The outcomes we get, obviously, can in any case be improved once more, contrasted with neutral, positive, and negative sentiments, it is workable for additional exploration to give a more explicit gathering of sentiments. The utilization of TextBlob and VADER in this examination can be supposed to be very exact and can deliver great perception, even though obviously some further upgrades are required. In any case, as an end, this Sentiment analysis research is required to give adequate support to per users to do comparable examination, to help political decision members freely break down their political decision results, and maybe the two possibility to discover what remarks or feelings from their electors.

**Bibliography:**

1. Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2017). Sentiment based analysis of tweets during the US presidential elections. Proceedings of the 18th Annual International Conference on Digital Government Research. <https://doi.org/10.1145/3085228.3085285>.
2. Calderon, N. A., Fisher, B., Hemsley, J., Ceskavich, B., Jansen, G., Marciano, R., & Lemieux, V. L. (2015). Mixed-initiative social media analytics at the World Bank: Observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. 2015 IEEE International Conference on Big Data (Big Data). <https://doi.org/10.1109/bigdata.2015.7363939>.
3. [3] Mergel, I. (2012). “Connecting to Congress”: The use of Twitter by members of Congress. Zeitschrift für Politikberatung, 5(3), 108-114. https://doi.org/10.5771/1865-4789-2012-3-108.
4. [4] Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2015). Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter. Journal of the Association for Information Science and Technology, 67(1), 232-238. https://doi.org/10.1002/asi.23456.
5. [5] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544-2558. https://doi.org/10.1002/asi.21416.
6. [6] Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing, 9(6), 811-824. https://doi.org/10.1109/tdsc.2012.75.

**Bibliography**

1. Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2017). Sentiment based analysis of tweets during the US presidential elections. Proceedings of the 18th Annual International Conference on Digital Government Research. <https://doi.org/10.1145/3085228.3085285>.
2. Calderon, N. A., Fisher, B., Hemsley, J., Ceskavich, B., Jansen, G., Marciano, R., & Lemieux, V. L. (2015). Mixed-initiative social media analytics at the World Bank: Observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. 2015 IEEE International Conference on Big Data (Big Data). <https://doi.org/10.1109/bigdata.2015.7363939>.
3. Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2015). Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter. Journal of the Association for Information Science and Technology, 67(1), 232-238. <https://doi.org/10.1002/asi.23456>.
4. Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing, 9(6), 811-824. <https://doi.org/10.1109/tdsc.2012.75>.
5. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544-2558. <https://doi.org/10.1002/asi.21416>.
6. Mergel, I. (2012). “Connecting to Congress”: The use of Twitter by members of Congress. Zeitschrift für Politikberatung, 5(3), 108-114. <https://doi.org/10.5771/1865-4789-2012-3-108>.