## Wordcloud, overall sentiment for the day, how positive are the positive tweet

## Top topics being discussed, most engaged locations/states/cities, most engaged users profile, top account mentions, top hashtags related to these hashtags

## Positive and negative terms

## Hive is used to load data from HDFS into hive table. We use serDes here. serDes is serializer — Deserializer tool which help hive to understand the json format of tweet (which is loaded from HDFS)

## Techniques

## The text in the tweets was pre-processed by first converting the text to common case, removing digits, punctuations and special characters. The ‘English’ stop words provided by nltk.corpus, a python package, was used to remove stop words from tweets. Some examples of stop words: 'I', 'me', 'my', 'myself', 'we’,’some’,’such’,’other’. Lastly, stemming operation was performed on the tweets by using nltk.PorterStemmer, a python package.

## Topic analysis, to discover hidden structure in tweets, on entire set of tweets using SKlearn’s Non-Negative Matrix Factorization (NMF) Python package.

## For Sentiment analysis, AFINN and Hu&Li modules were used to capture sentiment scores. Below is the research on each of the following modules.

## Hu&Li Analysis:

## 1. The appearance of an opinion word in a sentence does not necessarily mean that the sentence expresses a positive or negative opinion.

## 2. You will notice many misspelled words in the list. These are not mistakes. They are included as these misspelled words that appear frequently in social media content.

## 3. Hu&Li analysis is in the area of sentiment analysis and opinion mining from social media.

## AFINN Analysis:

## 1. AFINN is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive). The words have been manually labelled by Finn Arup Nielsen in 2009-2011.

## 2. Using a simple word matching we show that the new word list may perform better than ANEW(Affective Norms for English Words)

## We can use sentiwordnet as well

## References

## <http://www.techrepublic.com/article/what-twitter-sentiment-analysis-is-saying-about-the-first-presidential-debate/>

## <http://web.stanford.edu/~jesszhao/files/twitterSentiment.pdf>

## <http://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

## <http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/>

## <https://github.com/riccardotommasini/twitter-sentiment-analysis/tree/master/dataset>

## <https://github.com/ayushoriginal/Sentiment-Analysis-Twitter>

## <https://blog.parsehub.com/mining-tweets-for-sentiment-analysis-part-2/>

## <https://marcobonzanini.com/2015/05/17/mining-twitter-data-with-python-part-6-sentiment-analysis-basics/>

## <http://www.nltk.org/howto/sentiwordnet.html>

## <http://mark-kay.net/2014/08/15/network-graph-of-twitter-followers/>

## <https://www.kaggle.com/c/word2vec-nlp-tutorial/details/part-1-for-beginners-bag-of-words>