Spark Streaming with Twitter and Kafka

A spark streaming application was created that continuously read data from Twitter about the two 2020 presidential candidates Trump and Biden.

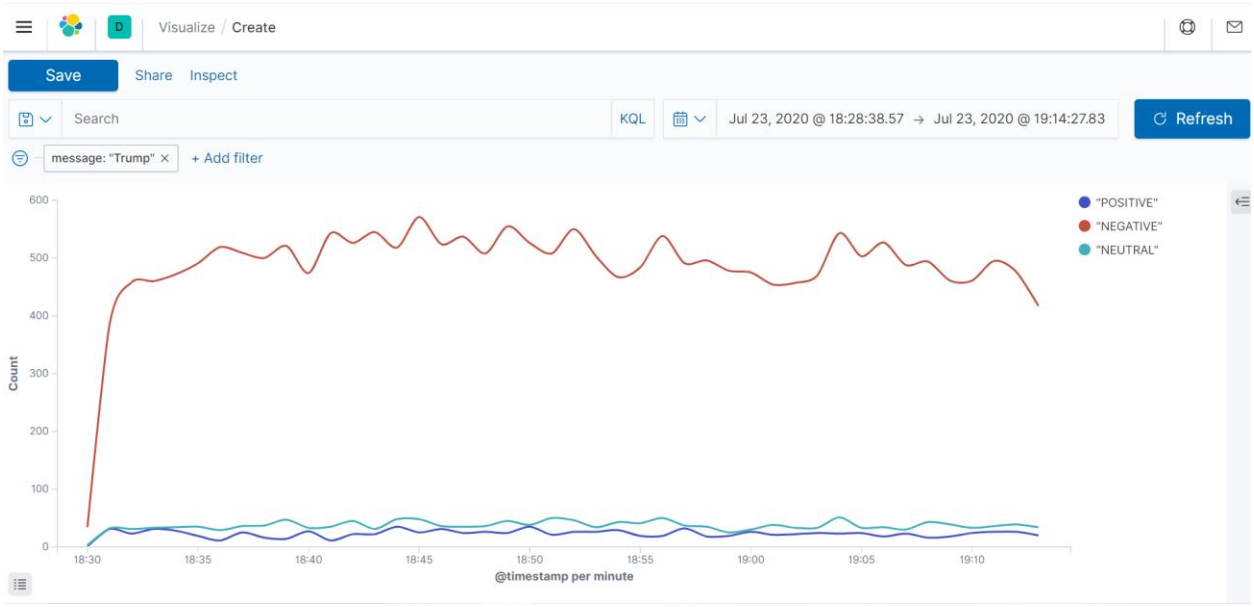
Data was collected for over 2 hours and Stanford NLP models were used for sentiment analysis. For both the 2020 presidential candidates, there were more tweets with negative sentiment than positive sentiment. Over 80% of tweets for both candidates were negative. The topic “Trump” has significantly more tweets compared to the topic “Biden”.

Visualizations

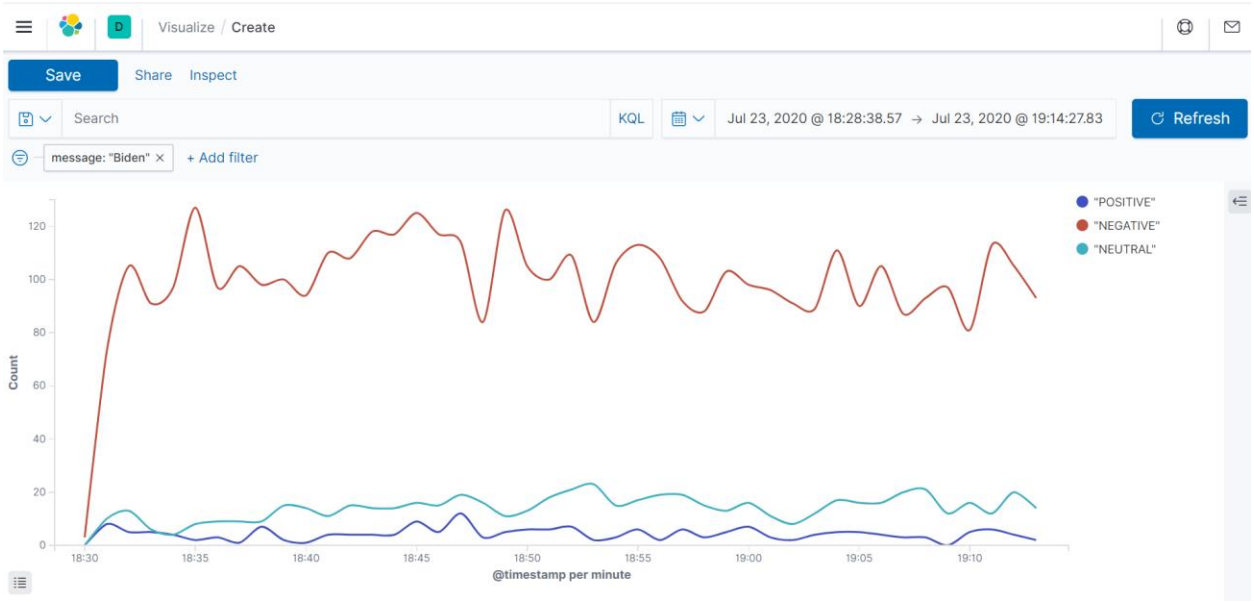
Continuous streaming of tweets regarding Trump and Biden.

Continuous streaming of tweets with positive, negative, and neutral sentiment regarding Trump and Biden.

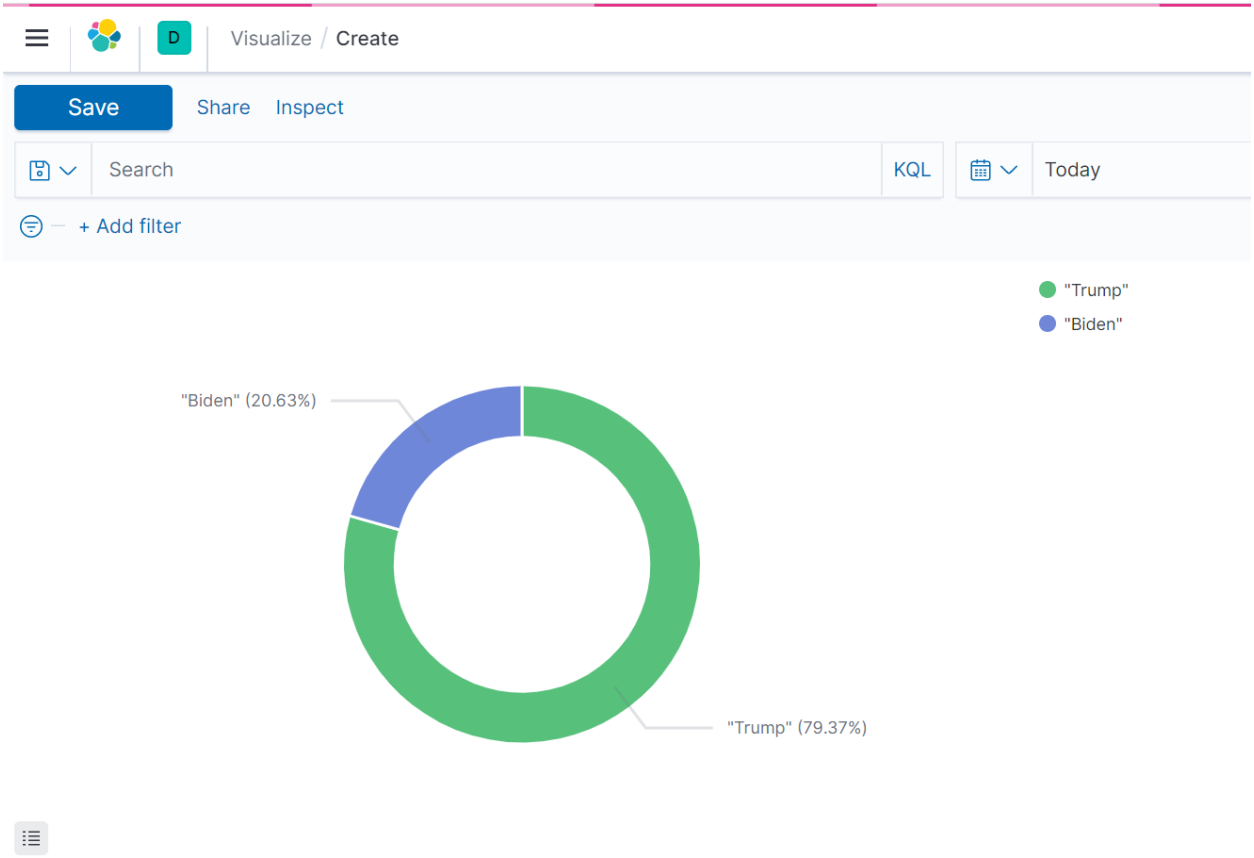
Sentiment variation overtime for “Trump” tweets



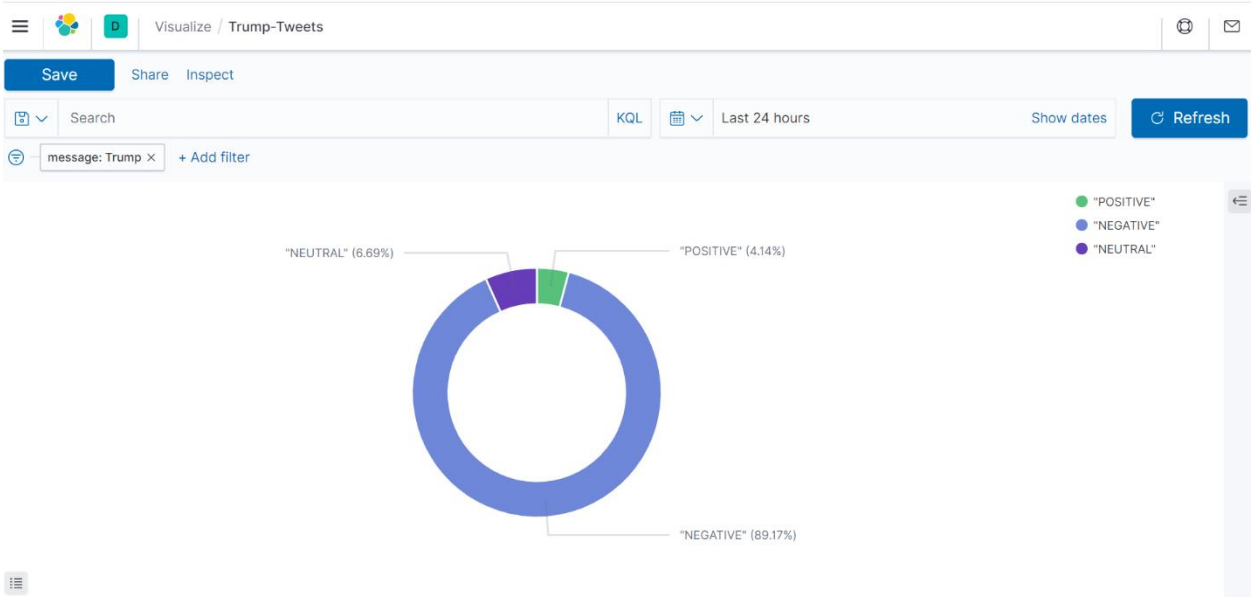
Sentiment variation overtime for “Biden” tweets



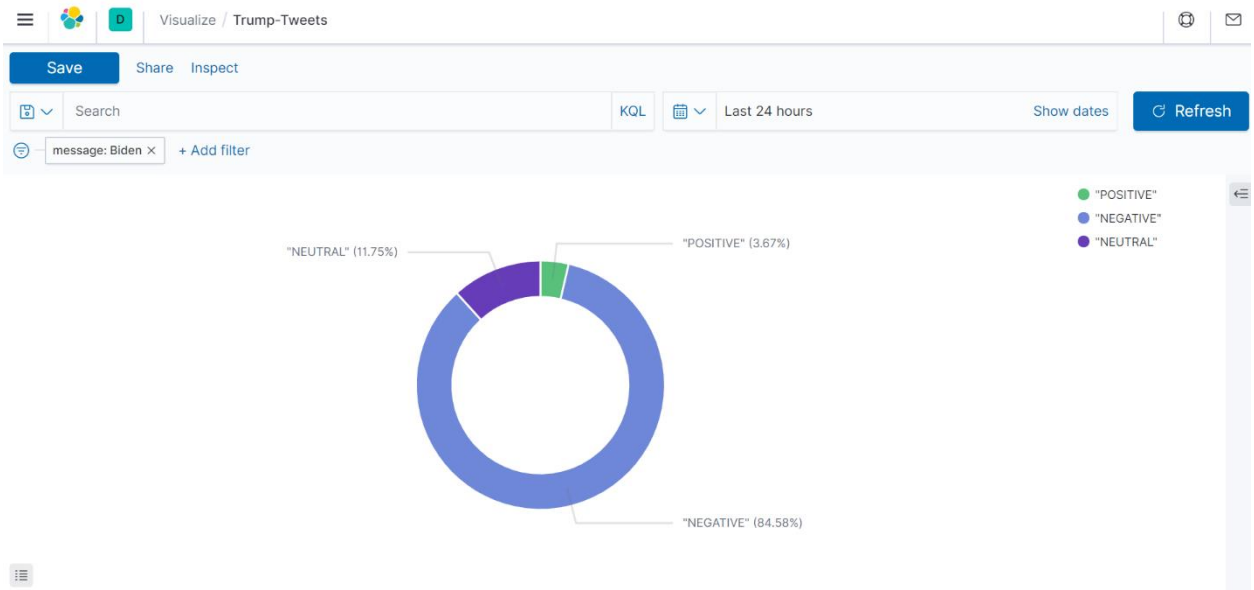
Percentage of “Trump” and “Biden” tweets collected



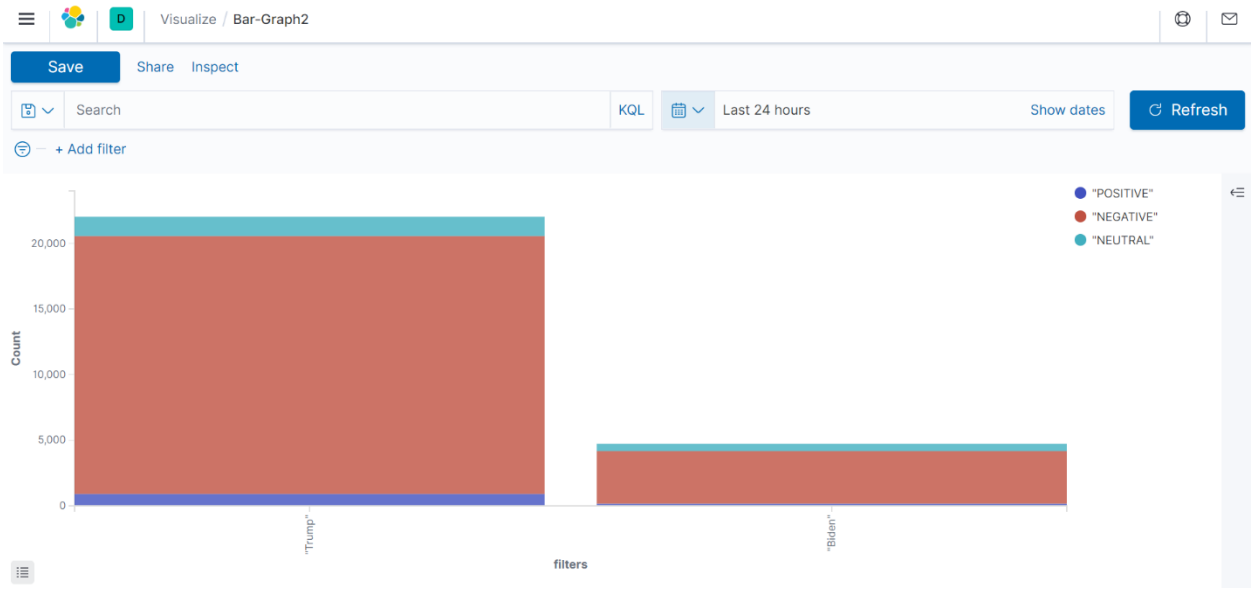
Percentage of positive, negative, and neutral tweets with “Trump” as topic.



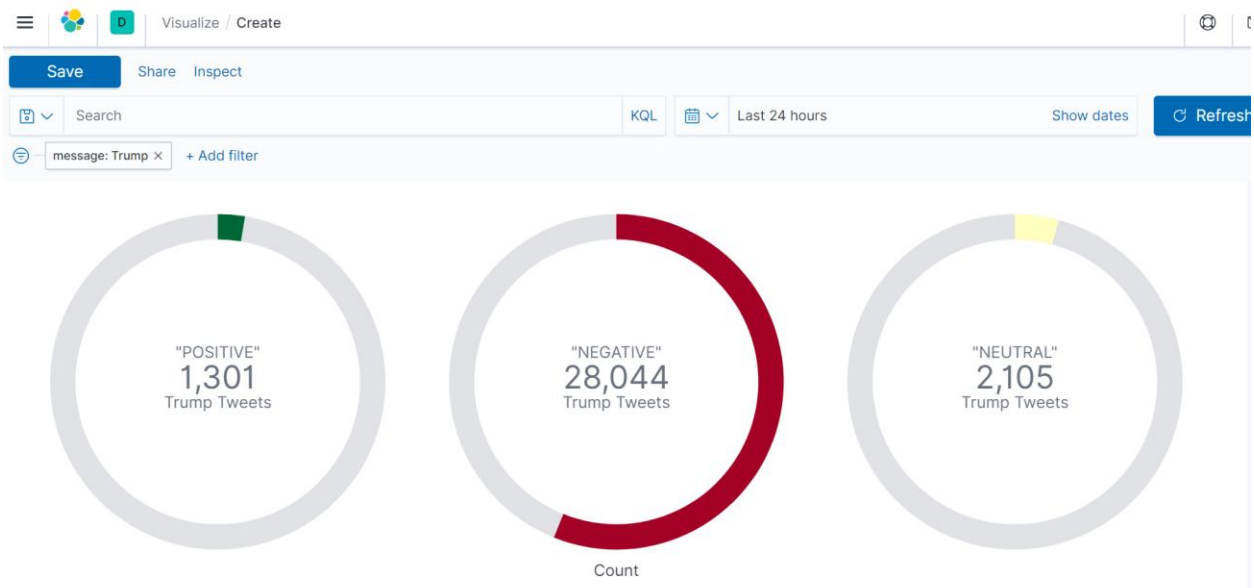
Percentage of positive, negative, and neutral tweets with “Biden” as topic



Comparing number of “Trump” vs “Biden” tweets along with sentiment



Number of positive, negative and neutral tweets collected with topic “Trump”.



Number of positive, negative and neutral tweets collected with topic “Biden”