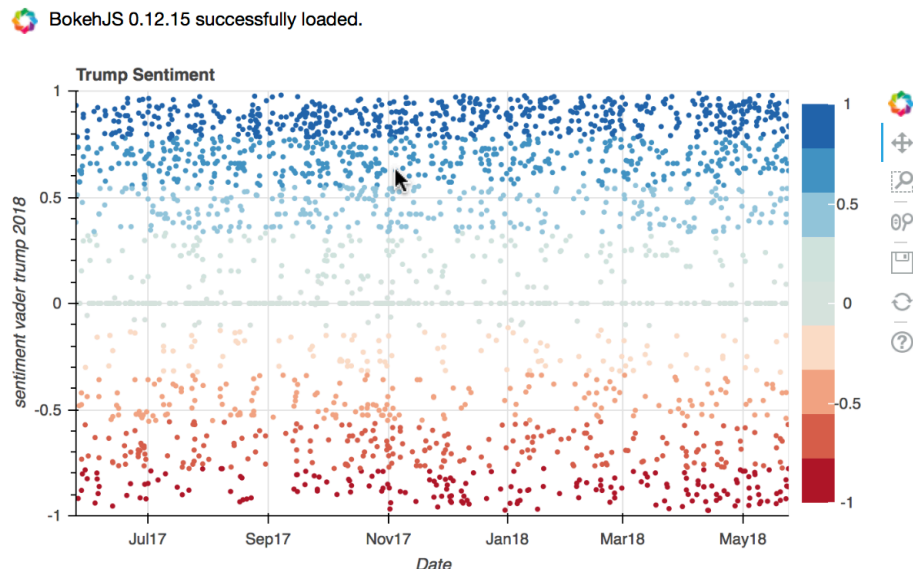


Natural Language Processing: Understanding a Twitter Influencer

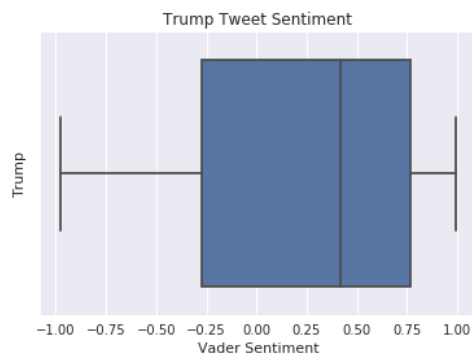
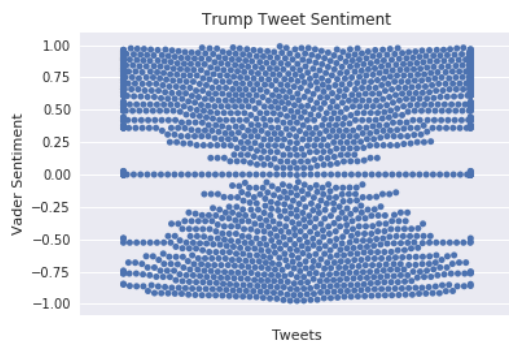
Sometimes a client, a celebrity, or a company can exert an undue amount of influence upon a product, a company, an industry, or a market. We will look at how a company can use Natural Language Processing (NLP) to understand an influencer. The president of the United States is such an influencer. We will look at Donald Trump's tweets to see if we can gain insight into his influence. We will start with looking at the sentiment of his tweets. Knowing whether an influencer is in a good mood today (positive sentiment) can give insight into any interactions with that influencer. We might want to pitch an idea on a good day, or if making an investment decision, we might wait for a better day. Of course, these decisions are all made based on industry knowledge. We look to provide some tools using NLP to assist the decision maker in making these influencer based decisions.

First we need to get some data. We scrap the web to pull in tweets from the last year. We use some Python code developed by Jeremy Cunningham to retrieve these tweets. This saves the tweets for the last year to a csv file. We read these in to a Jupyter Notebook with Pandas. Change a few import setting about delimiters. Not all tweets have text (some have just pictures), so we drop all tweets without text. We reorganize the index. We need to convert the date and time text to datetime objects. We will also clean the tweet text for NLP processing.

We will use the NLTK package, which has some pre-trained algorithms for sentiment recognition. One is the Vader Sentiment analyzer, which we download. We write a function to score each tweet for the overall sentiment based on a pre-trained word lexicon. The algorithm then calculates the overall sentiment based on the word weights for sentiment. It scores on a scale ranging from negative (-1) to neutral, (0) to positive (1). We plot this using Bokeh for an interactive graph. We can use this to explore the data zooming into a specific time span, looking at just positive tweets, or isolating a tweet.



We focus our exploration on the sentiment analysis for the moment. We look at some basic statistics associated and see that overall Trump has a positive tweet feed for the last

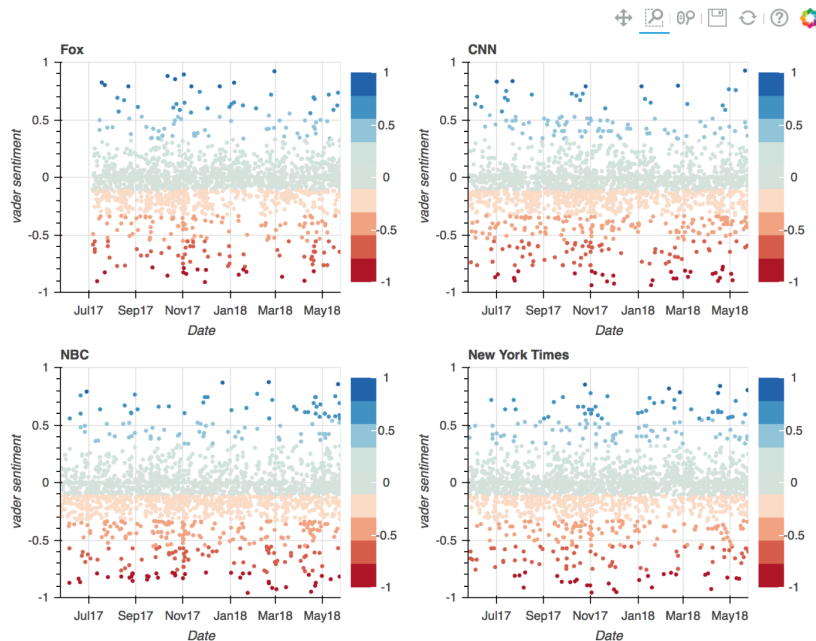


year. The mean of his sentiments is 0.231237 (on a scale from -1 to 1). The median values is 0.4199 not super positive, but positive.

Next, we look at the sentiment of news tweets before trump tweets to see if the news broadcasters influence Trump. The goal is to better understand an influencer, so understanding who influences him could help. We will look at CNN, FOX, NBC, and the New York Times. We will use the same scraper to get csv files for the last year tweets. We clean the data to keep only tweets with text as we did with Trump. We similarly get the sentiment for each news tweet. We need to convert these to float numbers for plotting since the NLTK module returns an object type. Next we call a function to collect all the tweets from the news broadcaster that

BokehJS 0.12.15 successfully loaded.

News Sentiment Before each Trump Tweet



occurred between a trump tweet and the one before it. We calculate the average sentiment of each broadcaster before each Trump Tweet and plot those here with Bokeh.

We see that the overall sentiments are more gathered around the neutral line, and not nearly as strongly negative or positive as Trump tweets. We can once again interactively explore these tweets; zooming in and panning around. Moving on to look at the statistics, we see that the mean sentiment for each broadcaster is slightly negative before each Trump tweet.

	cnn_sentiment_before_tweet		fox_sentiment_before_tweet
count	2146.000000	count	1846.000000
mean	-0.084623	mean	-0.046573
std	0.242029	std	0.233220
min	-0.939300	min	-0.911800
25%	-0.178769	25%	-0.132844
50%	-0.067305	50%	-0.029957
75%	0.008474	75%	0.049975
max	0.927400	max	0.921100

	nbc_sentiment_before_tweet		nyt_sentiment_before_tweet
count	1974.000000	count	2073.000000
mean	-0.092030	mean	-0.052120
std	0.257940	std	0.246312
min	-0.959500	min	-0.956500
25%	-0.198590	25%	-0.148000
50%	-0.082350	50%	-0.033667
75%	0.008679	75%	0.044368
max	0.873300	max	0.851900

How much does the tweets before affect Trump's sentiment? We look at the correlation.

	sentiment_vader_trump_18
sentiment_vader_trump_18	1.000000
fox_sentiment_before_tweet	0.000538
cnn_sentiment_before_tweet	0.066650
nbc_sentiment_before_tweet	-0.001874
nyt_sentiment_before_tweet	0.060123

The numbers don't look encouraging. Investigating the tweets before each trump, we see that the topics vary widely compared to what Trump is tweeting about. We might be able to find a better correlation if we classify tweets by topic. This is what we will attempt next.