



# Be Smart Twitter

Nicholas Montoya

Ziyang Yang

Zihan Zhou

# Description

As we all know, Trump (Donald J. Trump) is such a controversial president, and he might be the first president who is addicted to social media, especially Twitter. We want to know more about President Trump and figure out the impact he has on other users. Using Twitter, we are going to explore the correlations between news source statements, Trump's tweets, and other users. Also, we hope to answer the following **questions**:

- What is the sentiment felt towards various news sources?
- How does this change after releasing an article or headline regarding the president?
- How does this change after the president mentions a specific news source or the news in general?
- How do some common phrases ("fake news", "failing New York Times", etc.) change user responses?

In general, we would like to see how much statistical change can be attributed to a tweet from the president.

# Prior Work

Because of Twitter's easy-to-use API and President Trump's tendency to tweet about controversial topics, there have been a few similar works.

This link(<https://www.fxcm.com/insights/president-trumps-twitter-impact-forex-markets-stocks/>) provides an article showing how these tweets have impacted the stock market. We chose to provide this link because it has a references section at the bottom with more relevant links to articles showing connections between tweets and world economies.

On a more scientific note, the University of Toronto has a sample project in which they performed sentiment analysis on the president's tweets to differentiate which tweets were written by staff members. This is similar to what we are doing except instead of looking at the POTUS sentiment, we want to see what everyone else was feeling.

# Datasets

A lot of our data will come from Twitter. While President Trump doesn't have millions of tweets, he has thousands. And each of those tweets garners thousands more responses for us to analyze. Due to this (and rate limiting with the Twitter API) we have downloaded 10,005 tweets, 2000 responses per POTUS tweet, 5 POTUS tweets. Because we can get most of our data from an API request, we are only planning on downloading small enough sets to test on for data cleaning, preparation, etc.

We are looking at The New York Times, CNN, and Fox for news. As most of them post links to articles regarding the president on Twitter, that may be enough to mark it as a time to analyze. However, if we experience difficulty figuring out if a particular story is related to the president we will use their API's to accurately find out.

## Links:

<https://developer.twitter.com/en/docs>

<https://developer.nytimes.com/>

<https://developer.cnn.com/docs/read/api>

<https://newsapi.org/s/fox-news-api>

# Proposed work

**Data Cleaning:** Remove the random messages(common) which do not make sense at all, and the commons which only contain emojis. Also, remove Twitter specific data such as usernames, links, irrelevant hashtags.

**Data Integration:** Integrate news articles and dates from the 3 non-Twitter API's in order to determine relevant dates and if the article is relevant.

To give more of an insight, we see the life cycle of one scenario being:

- Locate an article from a major news source regarding the president.
- Determine user sentiment before Pres. Trump's response
- Determine user sentiment after

And on a higher level, we'll track the overall shift in sentiment for a news source over a long period vs. POTUS tweets about that news source.

## List of Tools

- GitHub
- Gdelt (if needed)
- NY Times API
- Python resources:
- Tweepy
- NumPy
- SciPy
- Anaconda
- Matplotlib



# Evaluation

We can evaluate our results using statistical comparison and finding out if the correlation coefficient is sufficient enough to indicate genuine results.