Twitter Sentiment Analysis:

President Trump & Immigration

Harlin W. Sanders

Springboard Data Science Career Track

Capstone Project #1

Problem Statement/Challenge

- Can we build a model to answer the following questions:
  - “Do Americans support U.S. President Donald Trump’s position on immigration and, specifically, what many cite as the separation of families as they try to enter/cross the U.S. border?”
  - “Do those opinions vary by geographic region?”

Steps and Processes

- Project Details, Description of the Data, and Data Wrangling
- Exploratory Data Analysis
- Machine Learning and Further Analyses

Data & Data Wrangling

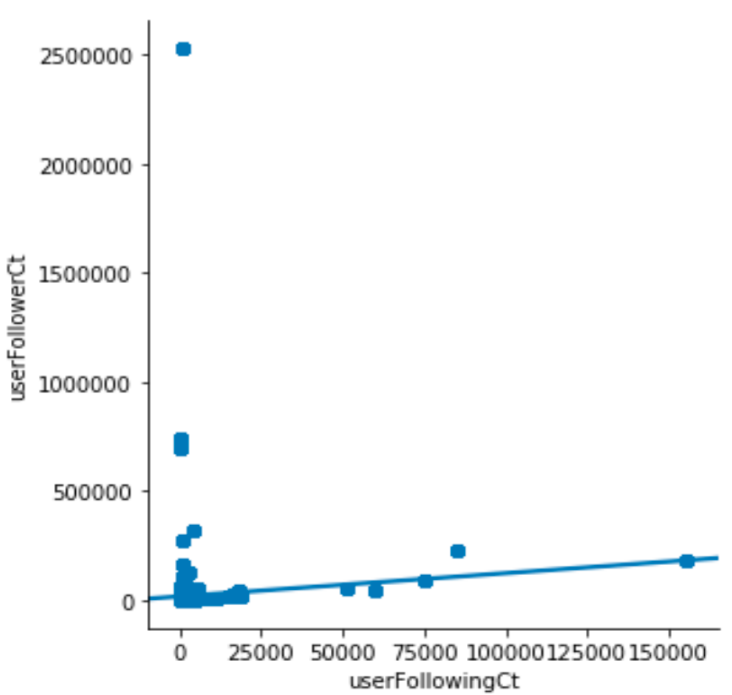
- Standard Twitter API
  - Provided information about the tweet itself and the Twitter user who sent the tweet
    - Premium API provides a more robust search feature but was cost prohibitive for this project
  - Query string with relevant terms
  - *place_id* piece of Twitter metadata for geolocation information

Exploratory Data Analysis

- 49,000 rows and 21 columns
- Features included:
  - tweet ID
  - tweet text
  - number of times the tweet was retweeted
  - source of the tweet (e.g., iPhone, Android, etc.)
  - the date the tweet was created
  - User ID, user description
  - number of followers
  - the number of other Twitter users the sender is currently following.

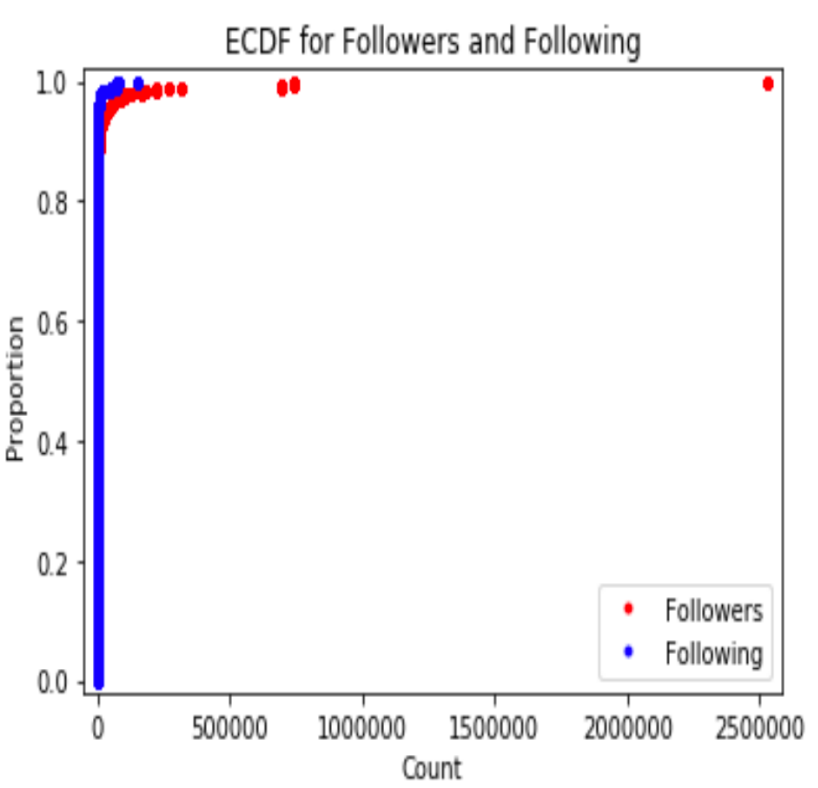
EDA continued…

- The number of Twitter followers is positively correlated with the number being followed (and outliers are very clearly identified).



EDA continued…

- An empirical cumulative distribution function graph highlighting the distribution of the number of user's followers is very similar to the distribution of the number of people they are following.

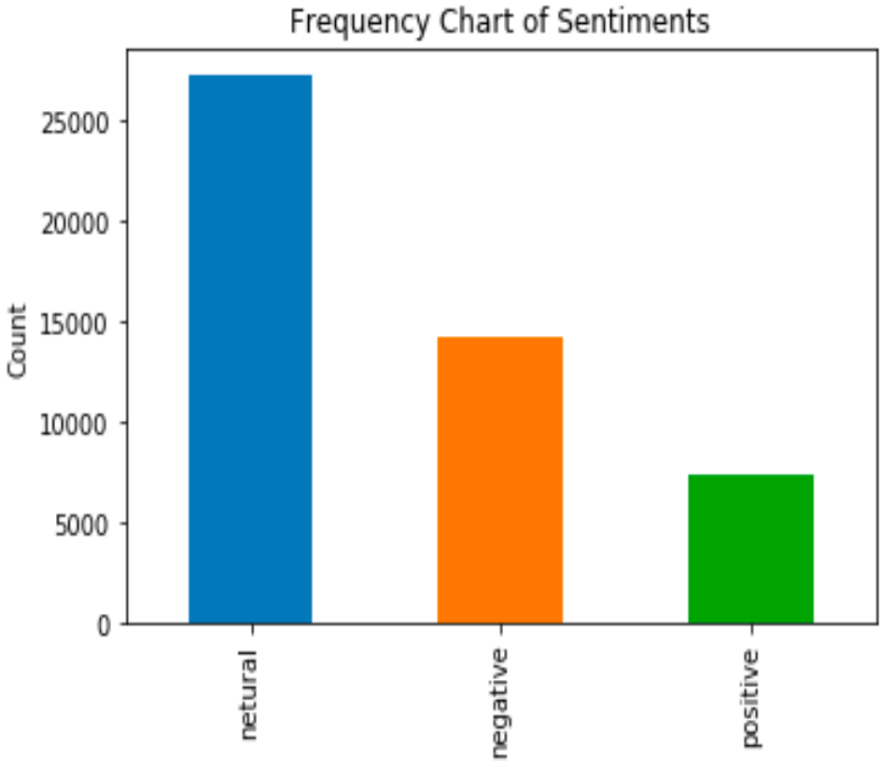


Machine Learning

- Primarily leveraged Bayesian techniques for text analysis
  - Sentiment analysis/classification
    - *textblob* => provides “a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.”

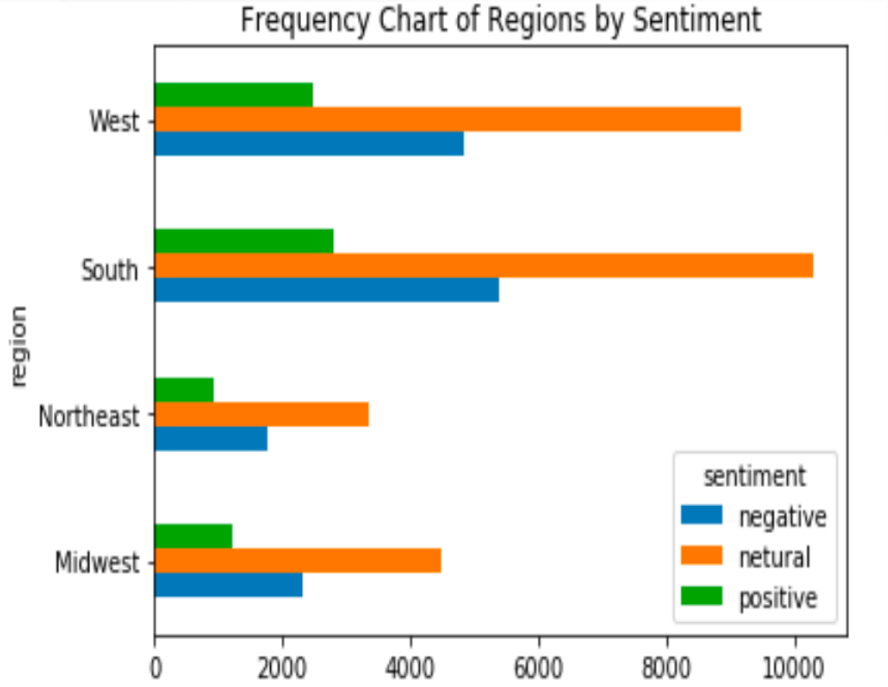
ML continued…

- There were far more “neutral” tweets (55.7%) than “negative” (29.1%) and “positive” (15.1%) tweets.



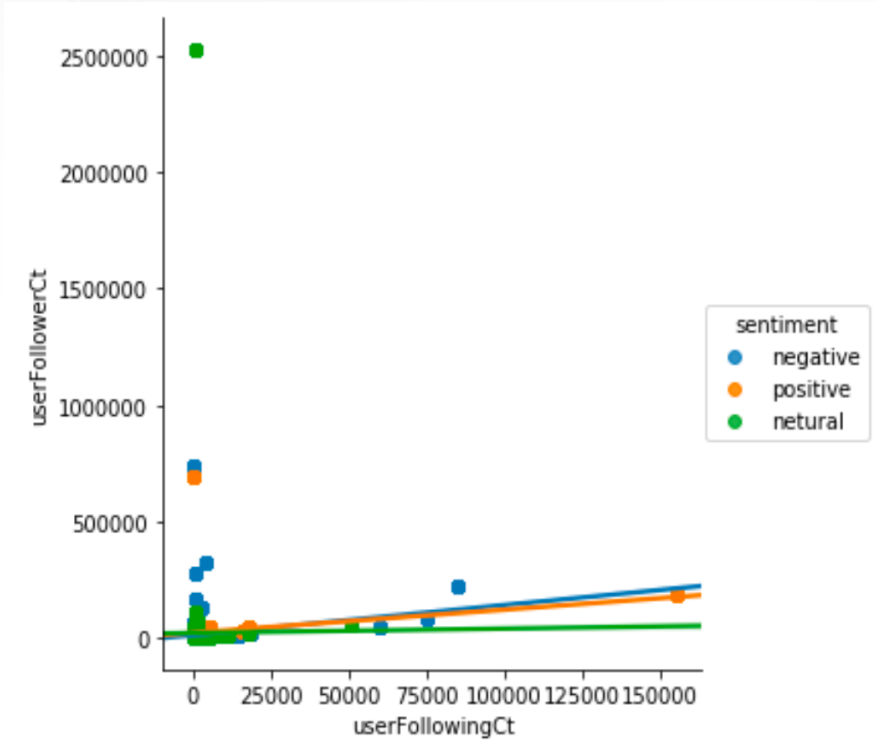
ML continued…

- The West region had the highest percentages of “negative” tweets and the Midwest region had the lowest percentage of “positive” tweets.



ML continued…

- The relationship between number of Twitter followers and number being followed is even stronger among those with “negative” sentiments about the President’s immigration position.



Conclusions

- This model and subsequent analyses can, in fact, be used to answer questions surrounding the decisions and policies of President Donald Trump.
- Of the Twitter users measured in this study, almost twice as many had a negative view than a positive view of the President’s position on immigration.
- Those opinions are equally distributed across the primary geographic regions of the United States; there is little to no difference between the regions.

Recommendations & Further Analyses

- Upgrade Twitter API access to the Premium API
  - Gather historical tweets and
  - Collect more precise geolocation information (i.e., county- and city-bound)