'Covid Vaccine' Tweets

Sentiment Analysis with Natural Language Processing

by David Bruce

Content Outline

1 Context

2 Preview Results

3 Data & Methods

4 Recommendations

5 Next Steps

A Global Pandemic

January 20, 2020

First known case of the novel Coronavirus documented in the US

September 22, 2020

US surpasses 200,000 deaths

December 14, 2020

First inoculation given to public of FDA approved Pfizer vaccine

Modeling Results Preview

Multiclass Classification

Model	Train F1	Test F1
Naive Bayes	0.71	0.70
SVM	0.95	0.84
SGD	0.82	0.81
SVM (GridSearchCV)	0.99	0.84

CRISP-DM

An Iterative Process

Webscraping Twitter

Twint

Data Cleaning

Pandas

NLP & EDA

NLTK, SpaCy, VaderSentiment

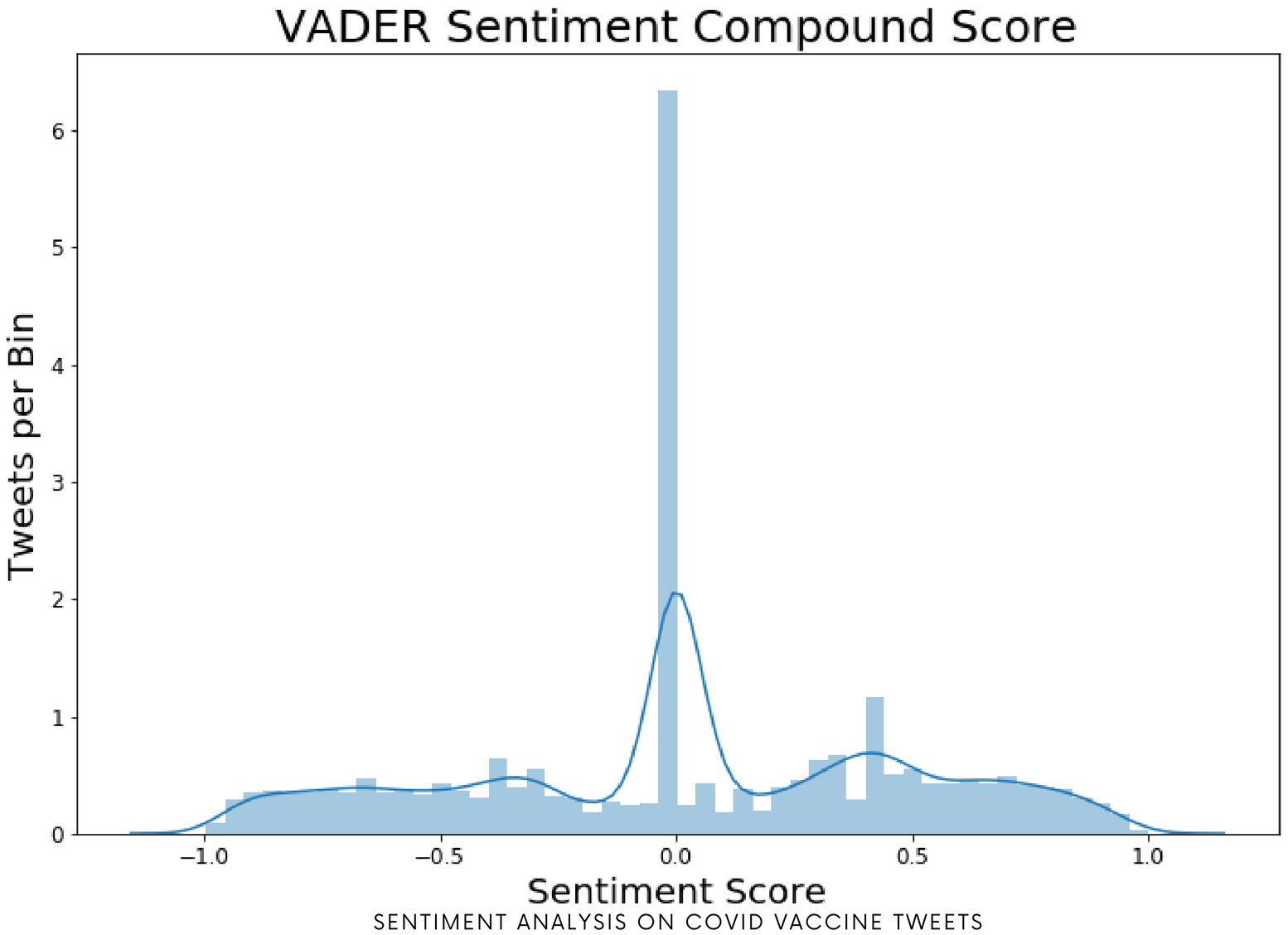
Modeling

Scikit Learn

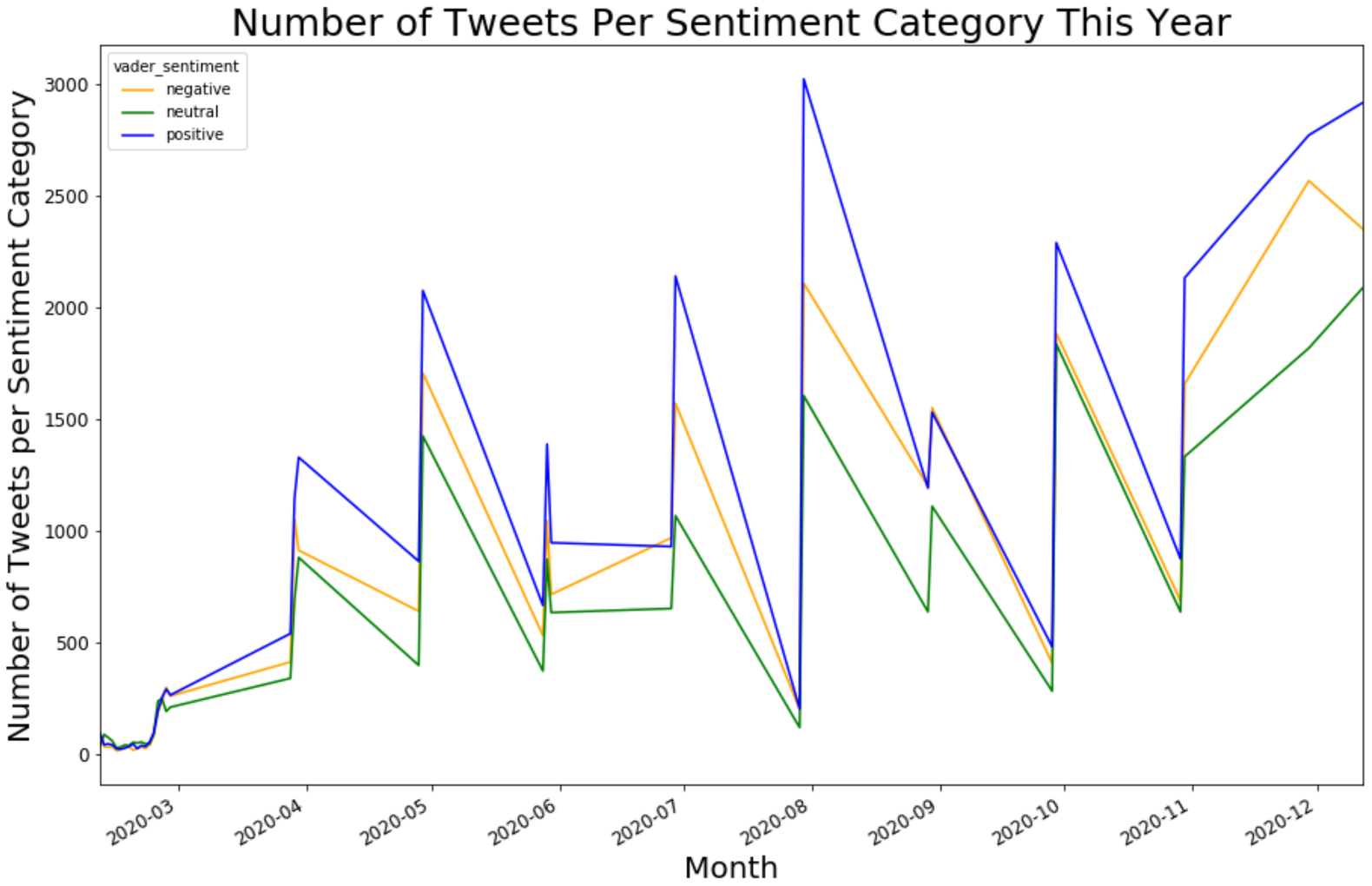
Reiterate

Also used: Matplotlib, Seaborn, WordCloud, and Numpy

DATA ANALYSIS - 1

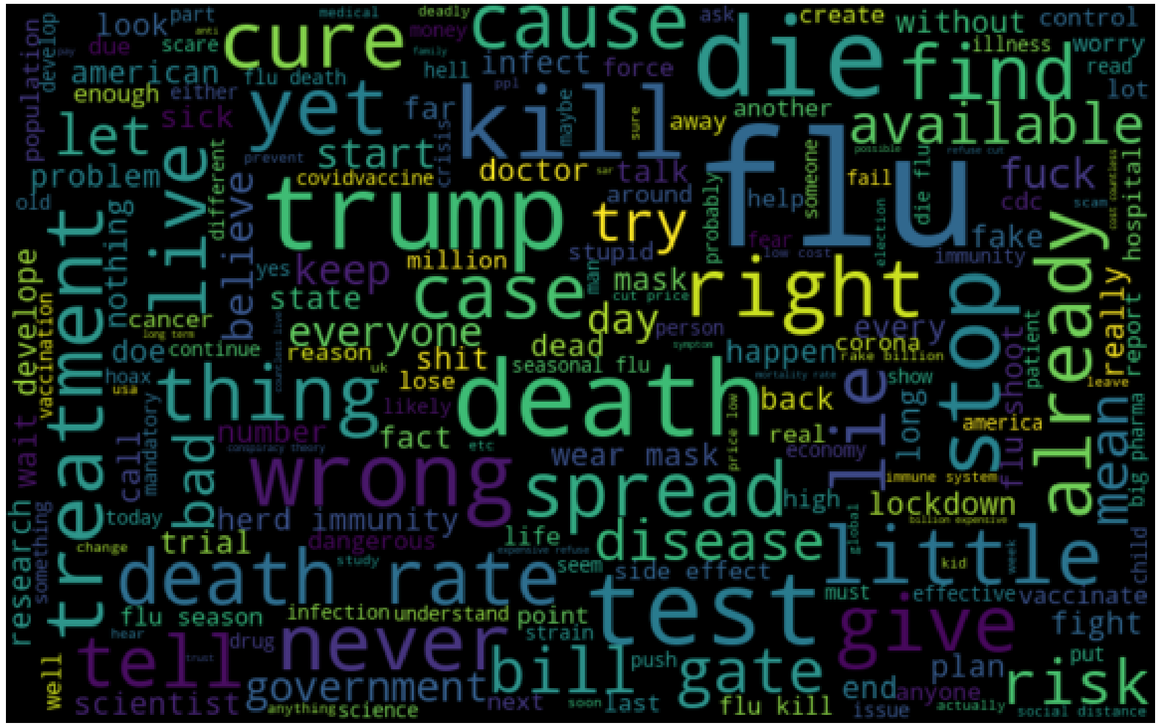


DATA ANALYSIS - 2



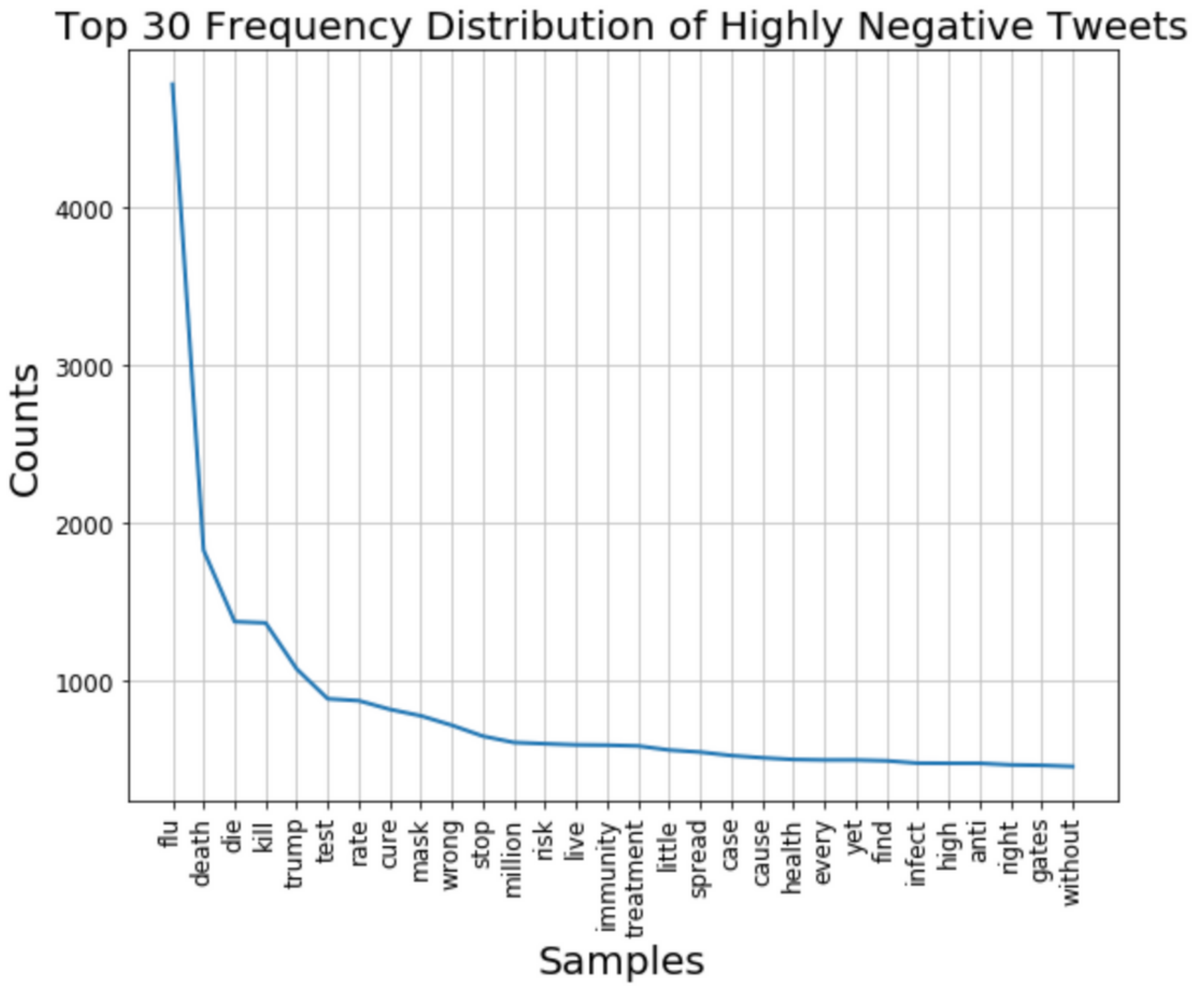
DATA ANALYSIS - 3

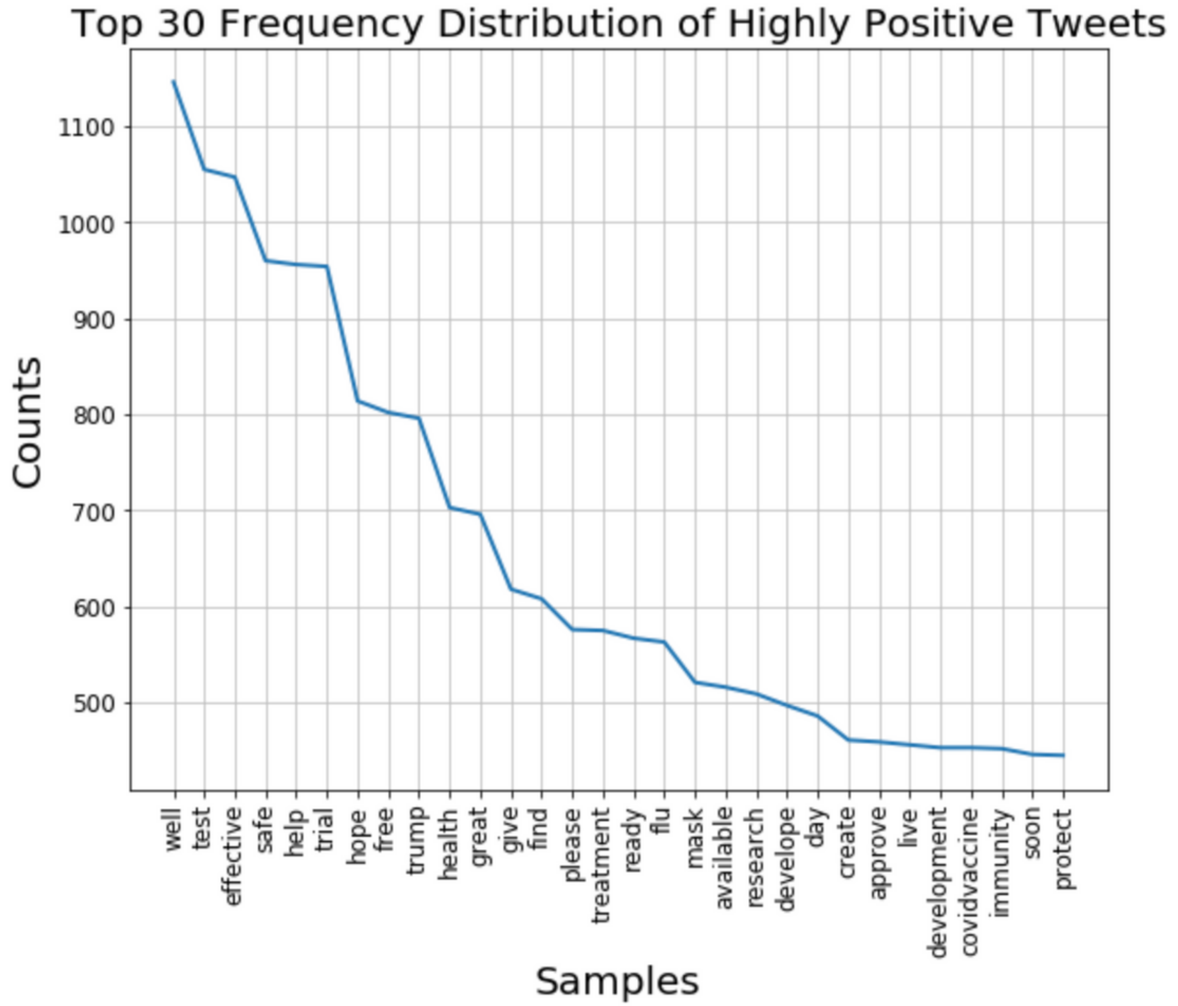
Highly Negative Wordcloud



Highly Positive Wordcloud

DATA ANALYSIS - 4





Next Steps

More Data

Location Based Tweets

LDA Topic Modeling with Gensim

Thank You.

Any Questions?

Email Address
david.bruce14@gmail.com

GitHub
@davidlebruce

Twitter
@dgbruce1