NLP and Campaign Donation Analysis

By Trevor McElhaney

Problem Statement

- The objective of this analysis was to determine if Trump's Twitter posts can be used to predict his campaign donation amounts.
  - Trump Twitter Archive
  - Federal Election Committee archive

Background:

- More than 90 percent of candidates who spend the most win. From 2000 through 2016, there was only one election cycle where that wasn't true: 2010.
- Trump has raised more money for the 2020 Presidential Election than any other president sence Reagan.

More Americans say they donate to political candidates and parties

% of U.S. adults who say they contributed to a candidate, political party or other group that supported or opposed candidates over the last 12 months

Candidates

Parties

Other groups

6
6
4

12
9
5

1992 1996 2000 2004 2008 2012 2016
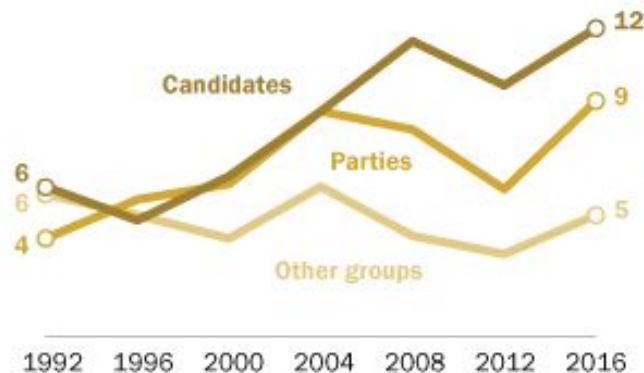
Source: American National Election Studies: Time Series Studies from 1992-2016. Post-election questionnaires fielded in the two months following each election.

PEW RESEARCH CENTER

Background:

- President Obama aka "First social-media President"
  - He was the first politician I dealt with who understood that the technology was a given and that it could be used in new ways." Marc Andreessen(founder of Netscape)
- Raised $722 million in 2012
- Trump Tweets ~11 times per day
  - Won the 2016 Presidential Election

President Obama ✔
@POTUS44

Thank you for everything. My last ask is the same as my first. I'm asking you to believe—not in my ability to create change, but in yours.

♡ 1.77M   10:52 PM - Jan 10, 2017

💬 832K people are talking about this

Donald J. Trump ✔
@realDonaldTrump

It's freezing and snowing in New York--we need global warming!

RETWEETS   LIKES
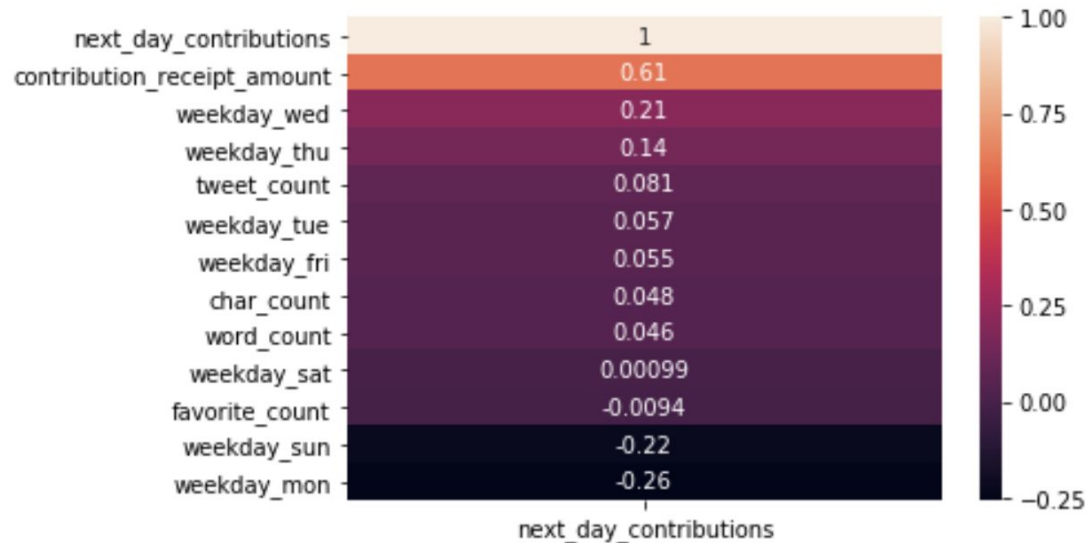4,223   2,391

11:24 AM - 7 Nov 2012
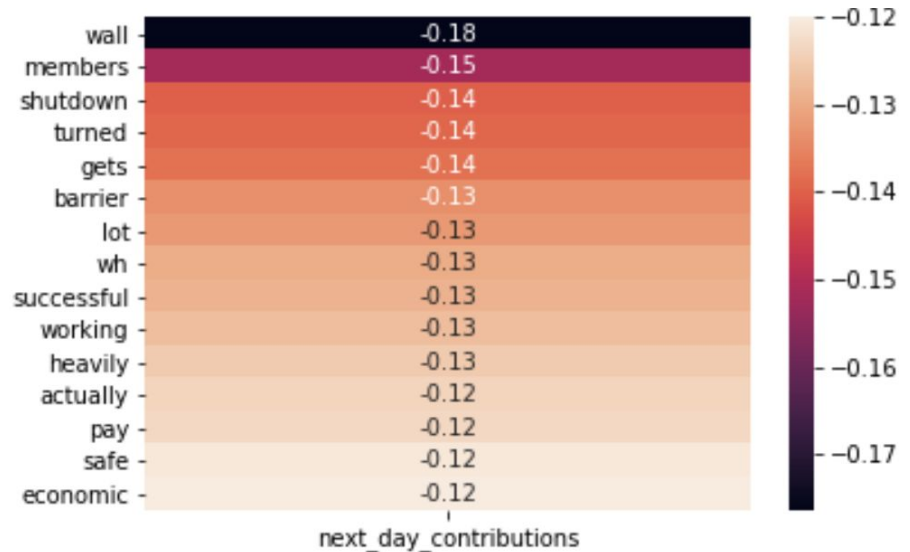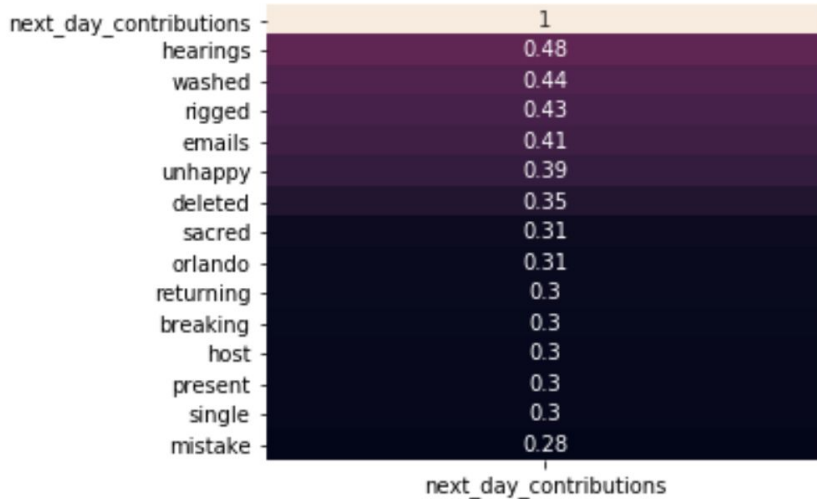
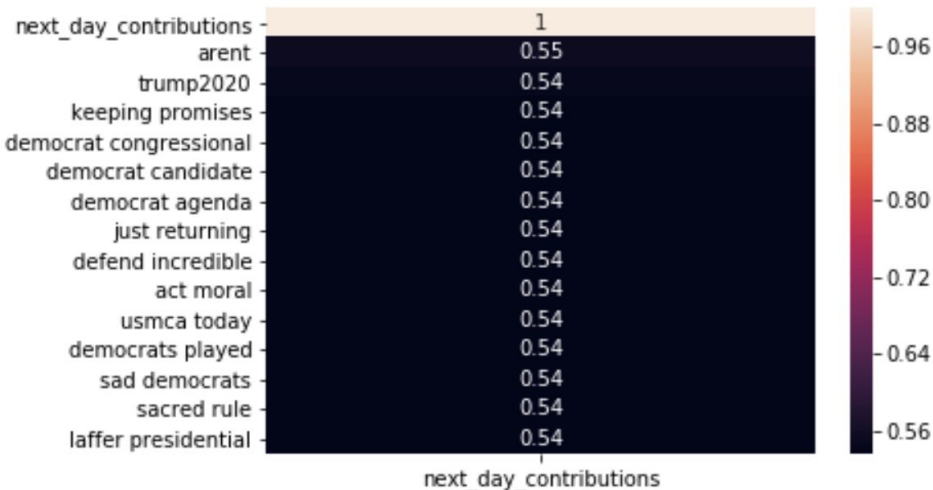↩ 1.7K   ↻ 4.2K   ♡ 2.4K

Exploratory Data Analysis

- Trump averaged $150k per day in individual donations.
  - June 18-19th he raised around $2 million.

# Top words correlated with donations

TF-IDF Analysis

	max	mean	sum
nan	1.000000	0.005525	1.000000
dday75thanniversary	0.374746	0.002327	0.421171
mexico	0.340680	0.014553	2.634177
interior	0.335228	0.001852	0.335228
secretary interior	0.335228	0.001852	0.335228
china	0.317833	0.013504	2.444293
walls	0.286680	0.003417	0.618407
1994	0.283713	0.001567	0.283713
america great	0.278124	0.007354	1.331126
good morning	0.276659	0.001529	0.276659
exoneration america	0.276659	0.001529	0.276659
day make	0.276659	0.001529	0.276659
obstruction complete	0.276659	0.001529	0.276659
morning great	0.276659	0.001529	0.276659
golf club	0.271337	0.001499	0.271337

Heatmap (left):

	next_day_contributions
next_day_contributions	1
arent	0.55
trump2020	0.54
keeping promises	0.54
democrat congressional	0.54
democrat candidate	0.54
democrat agenda	0.54
just returning	0.54
defend incredible	0.54
act moral	0.54
usmca today	0.54
democrats played	0.54
sad democrats	0.54
sacred rule	0.54
laffer presidential	0.54

Sentiment Analysis

	compound	neg	neu	pos
count	181.000000	181.000000	181.000000	181.000000
mean	0.386555	0.111779	0.717983	0.159193

- 11% negative
- 72% neutral
- 16% positive
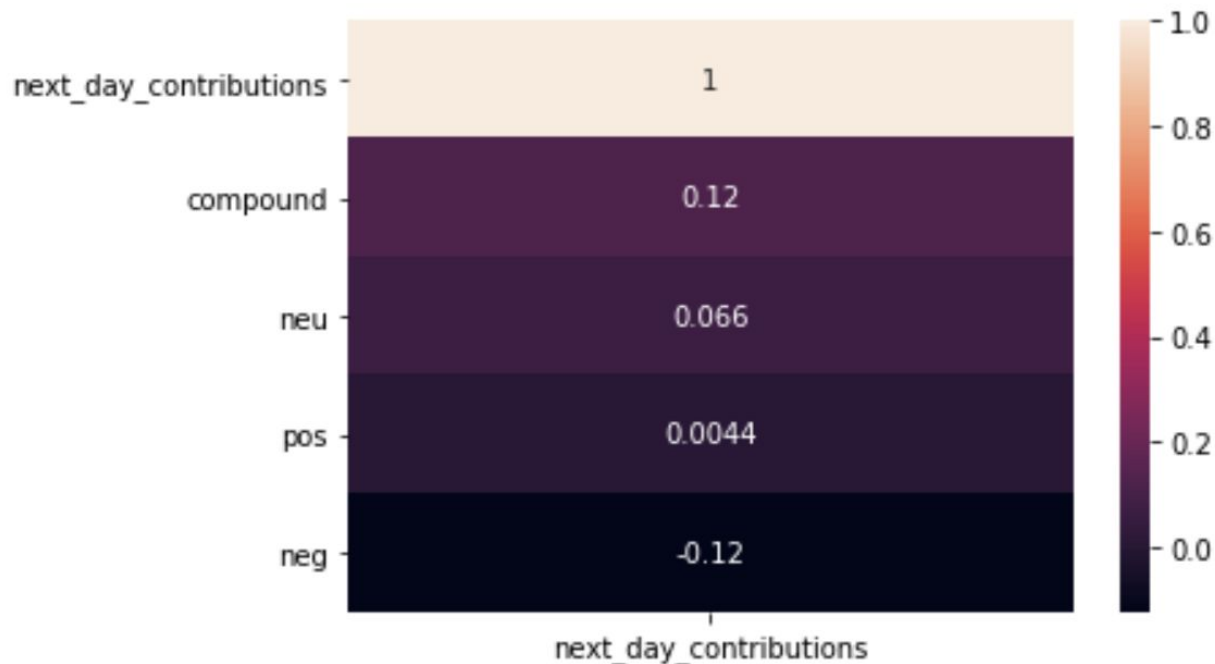
LSA Analysis

Explained variance by component
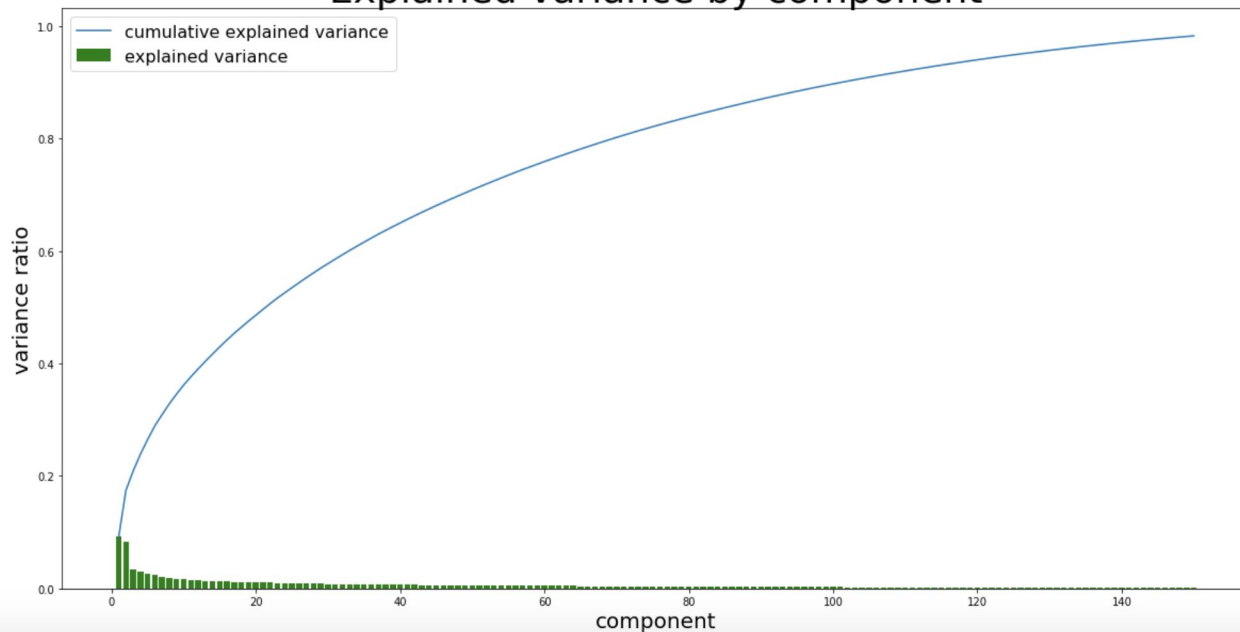
	component_1
great	0.349794
amp	0.336806
china	0.192415
people	0.189924
president	0.180302
democrats	0.178627
border	0.174216
just	0.149025
country	0.147425
news	0.121242

	component_2
china	0.562738
tariffs	0.270058
billion	0.266345
dollars	0.257473
products	0.195266
farmers	0.174437
border	-0.164987
democrats	-0.134907
people	-0.134774
deal	0.120235

Time Series Analysis
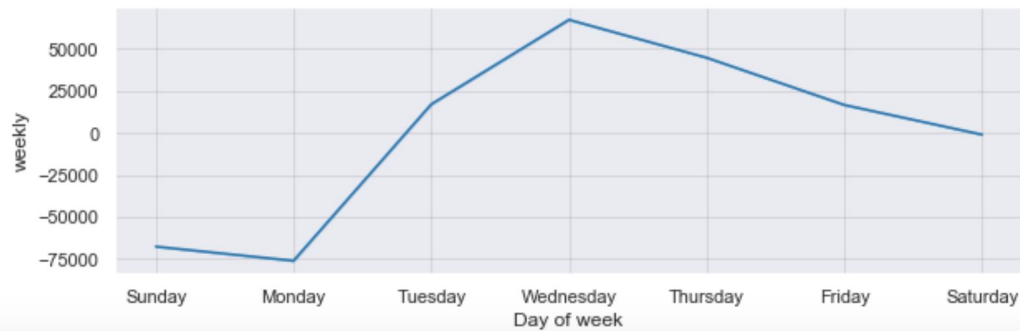
Predictive Models:

- Naive - 51.0% accuracy
- Logistic Regression
  - C=0.001, 'l2'
  - train-53%, val-47%
- Random Forest
  - Max depth-4, n_estimators-150
  - train-62.4%, val-62.2%

- Gaussian Naive Bayes
  - train-57.9%, val- 46.7%
- Logistic Regression
  - C=0.001, 'l2'
  - train-53%, val-47%
- XGBoost
  - Val- 64.4%

Conclusions:

- Overall, Trump tweets ≒ Donations
- There is small correlation
  - Reactionary?
- Further Study
  - Gather more data
  - Run RNN