

Assignment 7 Extra Credit: Twitter Analysis

Kathleen Hablutzal, CSC 111

2 November 2019

I decided to look at two different aspects of this Twitter feed: Democratic candidate mentions and trending retweets.

Part 1: Democratic Candidate Mentions

Building on the word search code from the assignment, I searched the text of the Twitter feed for mentions of Democratic presidential candidates, such that I could compare frequency of mentions to polling data. Rather than searching for one name, I adapted the code to first search for each candidate's full name and then search by last name if no full name was found within the line.

```
# count of full name within line
tally = line.count(name[0] + ' ' + name[1])
# if full name not in line, try adding counts of first and last
# names separately
if tally == 0:
    # add last name
    tally += line.count(name[1])
```

Accounting for First Name-Only Mentions

Some candidates, such as Sanders and Buttigeig, are commonly referenced by only their first names. When I did not include first name-only mentions, Buttigeig never appeared and Sanders only appeared eight times. When I added a feature to also search by first name-only, Buttigeig appeared 39 times and Sanders appeared 20 times - over twice as many mentions. However, names such as *Michael* Bennet, *Steve* Bullock, and *John* Delaney also received enormous boosts in their counts due to their extremely common first names. This shell output shows counts of mentions for each candidate, the first run counting first names (except for Joe Sestak) and the second run excluding all first name-only mentions:

```
===== RESTART: /Users/krh/Desktop/CSC 111/Assignments/TwitterExtra.py =====
{'Bennet': 22, 'Booker': 4, 'Bullock': 55, 'Buttigeig': 39, 'Castro': 1, 'Delaney': 73, 'Gabbard': 17, 'Harris': 35, 'Klobuchar': 15, 'Messam': 16, 'Sanders': 20, 'Steyer': 45, 'Warren': 23, 'Williamson': 0, 'Yang': 22, 'Sestak': 0, 'Biden': 190}
>>>
===== RESTART: /Users/krh/Desktop/CSC 111/Assignments/TwitterExtra.py =====
{'Bennet': 3, 'Booker': 1, 'Bullock': 0, 'Buttigeig': 0, 'Castro': 0, 'Delaney': 1, 'Gabbard': 8, 'Harris': 24, 'Klobuchar': 4, 'Messam': 0, 'Sanders': 8, 'Steyer': 1, 'Warren': 23, 'Williamson': 0, 'Yang': 14, 'Sestak': 0, 'Biden': 67}
```

I decided to exclude a select few common first names (Michael, Steve, John, Andrew, Amy, and Tom) but allow first name-only mentions for all other candidates. I left 'Pete' because Buttigeig is never mentioned by his last name, and I left 'Joe' as a common reference for Biden despite potential confounding uses such as 'Morning Joe'. These selective exclusions introduce more of my human bias into the experiment, but they seem necessary to achieve reasonable results.

Other Small Adjustments

The Joes: Since Joe Biden polls much higher than Joe Sestak, I counted any first name-only mentions of 'Joe' toward Biden.

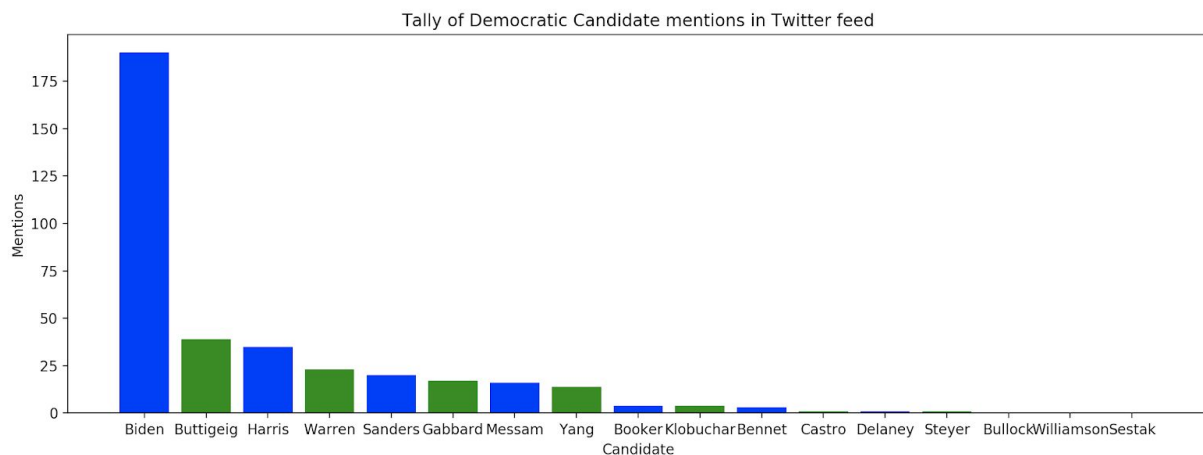
Julián Castro: I replaced all 'á's with the acute accent with a plain 'a', such that 'Julián' and 'Julian' both count as mentions of Castro.

Plotting Results

I stored tallies of candidate mentions in a dictionary. To plot these counts, I re-used the given code for iterating through keys and values in reverse order and constructed two parallel lists of names and their tallies:

```
names = []
tallies = []
# lists of names and tallies in popularity order
for key, value in sorted(d.items(), reverse = True, key=lambda item: item[1]):
    names += [key]
    tallies += [value]
```

Then, I used the matplotlib library to create a bar chart of mentions for each candidate:



Analysis of Results

I decided to compare popularity within this Twitter feed to the recent national polling averages as reported by [Real Clear Politics](#):

Polling Data															
Poll	Date	Biden	Warren	Sanders	Buttigieg	Harris	Yang	Klobuchar	O'Rourke	Booker	Gabbard	Steyer	Castro	Bennet	Spread
RCP Average	10/17 - 10/31	27.6	20.4	17.0	7.1	4.7	2.6	2.0	2.0	1.9	1.7	1.0	0.6	0.5	Biden +7.2

While Biden certainly leads in the polls, he dwarfs all other candidates by mentions within the Twitter stream. Part of this surge could be due to a general frequent occurrence of the common name “Joe”, or the trend of another string containing “Joe” (as observed in Part 2). However, considering the subject of the ongoing impeachment inquiry, where Trump asked Ukraine to investigate Biden, Biden and Trump have appeared together frequently in the headlines. Thus, since the tweets were filtered on “Trump,” we would expect this high frequency of Biden mentions.

I was personally surprised that Buttigeig and Harris appeared more frequently than Warren and Sanders. I generally assumed that candidates with higher polling numbers would also have greater presence on Twitter. However, two campaign differences may account for the greater association of Buttigeig’s and Harris’s names with the term Trump: (1) A more active social media presence, and (2) an active campaign strategy of criticizing Trump. Both could be strategic moves to try to raise polling numbers versus the big top three candidates. In addition, these same strategies could explain the disproportionate mentions of Gabbard and Messam.

Finally, I would like to note that all of Buttigeig’s mentions come from “Pete” rather than “Buttigeig.” This is likely because he campaigns as “Mayor Pete” and because “Buttigeig” is much harder to spell.

Part 2: Trending Retweets

This was a shorter analysis, where I adapted the hashtag counting code to find the ten most retweeted tweets within the given Twitter stream.

These were:

1. @AOC: Whats worse: since many of the flash mob already sat on the committees, they *knew* how serious a breach it was to bring devices **(57 times)**
2. @matthewamiller: The AG flew to Italy twice to personally investigate a right wing conspiracy theory the Italian government has now conf **(51 times)**
3. @algore: Reminder: Despite President Trumps best efforts, the U.S. cannot officially withdraw from the #ParisAgreement until one day af **(45 times)**
4. @realDonaldTrump: Dont hurt them please! <https://t.co/cKV2HIJQb> **(38 times)**
5. @JoyceWhiteVance: It seems absolutely inevitable that the end of this tawdry charade will involve confirmation that it has always been a **(34 times)**
6. @GOPChairwoman: Nancy Pelosi said: None of us came to Congress to impeach a president. **(33 times)**
7. @realDonaldTrump: Just a continuation of the Witch Hunt! <https://t.co/TTCXcg0jdU> **(31 times)**
8. @funder: "Good thing Colorado now offers free full day kindergarten so our kids can learn basic geography, Colorado Governor Jared Poli **(29 times)**
9. @Morning_Joe: "I think we all will be absolutely fascinated when we finally figure out what Vladimir Putin has on Donald Trump." --@JoeN **(26 times)**
10. @MollyJongFast: Humanitarian atrocities perpetrated on Americas loyalest allies in the Middle East, all because Donald J Trump gave Erd **(26 times)**

A few observations:

- Many of the tweets are cut off at the end. The text data could have been sliced down to 140 characters, even though Twitter now allows up to 280 characters in tweets.
- The '@Morning_Joe' tag at number nine augments the tally of Joe Biden mentions for Part 1 of my analysis.
- Multiple trending retweets seem to pertain to the impeachment inquiry. Others seem to pertain to the Colorado border and the Paris Climate Accord, due to a speech the previous day where Trump mentioned a wall in Colorado (which does not border Mexico) and defended his plan to pull out of the Paris Climate Accord. Thus, these tweets capture multiple important moments of the political news from October 24, 2019.