Metrics of Candidate Tweets in Relation to Polling

Introduction

In recent years, we have seen a massive shift to digital campaigning in political primaries in the United States. It was a vital part of the 2016 primaries and is shaping to be the same in 2020 as Americans' main source of news is now social media (Shearer). In particular, we have seen the outsized influence that Twitter can play in presidential campaigns in relation to other social media sites. This is in large part due to the fact that Twitter has the ability to shape the traditional media narrative. Donald Trump's tweets garnered massive airtime on cable news channels during the 2016 presidential election and were able to reach massive audiences in the process (Francia).

However, despite the fact that Twitter is now such a prominent medium for campaigns to reach broad audiences, there is still little research about the role of Twitter in campaigns, compared to the extent of research available on more traditional political communication. This especially holds true for political communication dependent on a candidate's polling numbers. In traditional political communication, there is an extensive literature detailing how underdog candidates employ different tactics during campaigning. Mainly, underdog candidates often attack the leading candidates in traditional political communication in order to boost themselves among the electorate (Chou and Lien). While there have been a few researchers who have explored the relationship between tweets and election results in other countries, namely Karlsen and Enjolras, and Kruikemeier in Norway and Denmark respectively, there has been very little research on the topic in the United States, despite the great public attention that social media and its relationship to campaigning has garnered in recent years.

The central question that this paper will be exploring is whether there is a correlation between the polling of a candidate and the sentiment of that candidate's tweets and the percentage of their tweets mentioning other candidates or prominent political figures. Analyzing these relationships could help political scientists understand how modern campaigns are employing Twitter and social media to effectively campaign past a "one size fits all" approach for political communication. Going further, if social media sites were to better understand how these algorithms are being employed by campaigns, they could tweak their algorithms in order to even the playing field or increase political awareness among the general public. Finally, this research will be of interest to candidates and campaigns themselves as they attempt to better understand the competition and how to optimize their own campaign operations in order to engage with the electorate.

Initial results from this analysis support my hypothesis that there is no significant correlation between the features of a candidate's tweet and their position in the polls. My tests indicate that there is no significant relationship between the sentiment of a candidate's tweets and their position in the polls. No matter where a candidate is polling, they tend to tweet with a neutral sentiment on average. Similarly, we do not see a strong correlation between a candidate's polling position and their mentions of another democratic candidate or prominent republican. Regardless of their polling, candidates most often do not mention any other prominent figures in their tweets, and it is only the leading candidate that tends to deviate from this trend to mention the GOP more. However, even this deviation in the data is not significant.

Background

As previously mentioned, I was unable to uncover any previous research that looked at the relationship between the sentiment of and mentions in tweets of candidates and their position in the polls in the US. However, Karlsen and Enjolras, and Kruikemeier conducted similar research in other countries and discovered a correlation between the tweets and their correlation in the polls. Karlsen and Enjolras conducted the analysis most similar to my work. They looked at the tweets of candidates in Norway and in one aspect of their research, compared different aspect of a candidate's tweets with that candidate's respective reach and influence on Twitter. They ultimately found that candidates who tweet more often are more likely to be influential on Twitter, a term encompassing the number of engagements and views of a Twitter user, as well as the number of followers of that user. They also did a quick analysis of the topics that candidates were tweeting about based on whether they were influential or not and found the results shown in Figure 1 below. In their analysis, Twitter influentials amount to the frontrunners in the Norwegian election, and they found that Twitter influentials often tweet less about personal anecdotes and more about policy, which tends to be more neutral in sentiment. They also tend to retweet less than other candidates, a variable that I was unable to test for in my analysis. Finally, they looked at the number of mentions of another Twitter user that candidates made and found that there does not seem to be a significant difference between influentials and non-influentials using mentions (Karlsen and Enjolras 349-352). My analysis will only cover mentions of other Democratic candidates in 2020, and although in a different country, where political communication tactics may vary largely across borders, it will be interesting to see how sentiment and the mention of other candidates or prominent figures for American candidates compares to that of Norwegian candidates.

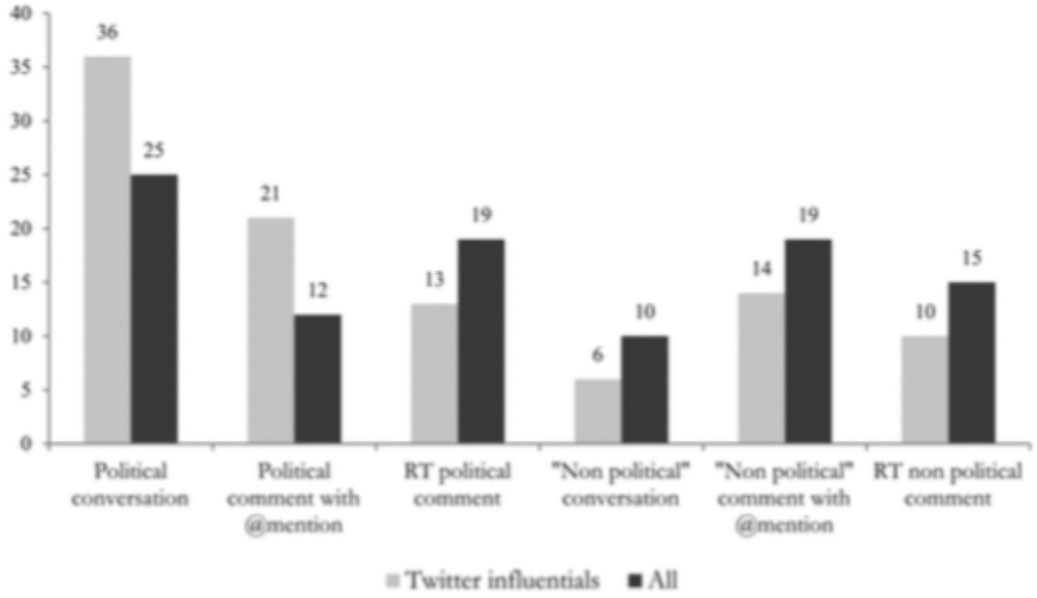


Figure 1: Analysis of Twitter Influentials vs. All Twitter Users from the research of Karlsen and Enjolras. The y-axis measures the percent of their tweets containing the variable along the x-axis. (Karlsen and Enjolras 351)

Kruikemeier's research was similar in the sense that it looked at the topics that Dutch candidates were tweeting about, analyzing how these topics compared to their results in the election. Kruikemeier discovered that there is a correlation between the topics that candidates are tweeting about and how they performed in the election. For instance, Kruikemeier discovered that candidates who tweet about their personal life tended to perform worse than those who did not, a finding that is in line with the results of Karlsen and Enjolras and again is interesting considering the fact that tweets about personal life tend to have a greater variance in sentiment (Kruikemeier). However, a European population's response to certain messaging could vary drastically from an American population's response given the different institutional and social factors at play in each country. Additionally, Kruikemeier noted that not all of the candidates in his election's study were on Twitter, which is drastically different from the political primaries in

2020, where most candidates are using Twitter and other social media sites as one of the primary ways to communicate with voters.

Karlsen and Enjolras were using the Twitter influence of candidates as a dependent variable, which itself, correlated pretty strongly with poll results, and Kruikemeier used election results as the dependent variable in their analysis, using the topics of tweets as the main measure of a candidate's tweets. However, there is currently no existing research that looks at similar a similar dependent variable, such as how a candidate is performing or will perform in the election, and how that variable compares with the sentiment of a candidate's tweets.

There have been some analyses of sentiment on political Twitter. Ceron et al. looked at the Facebook posts of the general public in Italy and France and analyzed these posts for sentiment, finding that there is a correlation between sentiment of social media posts among the general public and a candidate's performance in the polls (Ceron et al.). This shows that controlling the narrative on social media may be correlated with a positive performance in the polls. However, this does not directly correlate with our research question, given that we want to determine if there is a correlation between the tweets of candidates and the election results, not the tweets of the public and election results. Other work, most prominently that done by Bermingham and Smeaton as well as Chung and Mustafaraj, has also looked to analyze general Twitter sentiment in comparison with election results. In fact, the literature on that topic is much more extensive, but does not comment on the political communication tactics that candidates are employing in relation to their standing in the polls (Bermingham and Smeaton; Chung and Mustafaraj). Therefore, there is research that comments on the topics that candidates are tweeting about, which often gives us insight into the sentiment of a candidate's tweets, and also research

about the sentiment of Twitter in general about a candidate. However, there is no outstanding research that looks at the sentiment of a candidate's tweets directly and its correlation with their positioning in the polls. Also, existing research does show that candidates who are in the lead tend to retweet less and mention others less in their tweets. However, this research does not scope itself to mentions of other candidates, in order to see if classical political communication trends of underdogs attacking frontrunners holds on Twitter. Additionally, none of these studies were conducted on a US population, so results could vary drastically given the societal and institutional factors at play.

I hypothesize that there will be no significant correlation between the polling of a candidate and the sentiment of their tweets and the percentage of their tweets mentioning other candidates or prominent political figures. I do not think this trends seen in other countries will hold in the United States. I think that the theatrical aspect of American politics and the lack of separation between entertainment and politics lends itself to American candidates seeking to go viral more regardless of their current polling numbers, which would lead to candidates all tweeting in a similar vein. Therefore, I do not believe that there will be a significant relationship between the sentiment of tweets nor the mentions of other candidates or prominent figures despite research indicating that there is a significant difference in classical political communication (Chou and Lien). In order to go viral in a positive way and gain name recognition and favorability, candidates will attempt to employ the same tactics that evoke emotions that go viral more often. While it is incredibly difficult to predict virality, research has shown that certain emotions are present more often in viral content (Berger and Milkman). Therefore, I would expect candidates to attempt to employ those emotions and go viral

regardless of their position in the polls. I would expect some slight variation in all of the features of a tweet that we are mentioning since each candidate has a unique personality that they are attempting to portray in order to distinguish themselves from other candidates. However, I do not expect the sentiment and mentions of their tweets to have a significant difference from the tweets of other candidates.

There are a slew of factors that could provide evidence to reject this hypothesis. While I am assuming that candidates have the same potential to go viral on a platform like Twitter, this might not be the case as candidates who are performing better in the polls are more likely to have more Twitter followers and thus generate higher engagement with their tweets, leading to more viral tweets. This could force underdog candidates to again attempt communication strategies from more classical methods of political communication and spend more time attacking leading candidates in order to gain more traction. Additionally, it may be that certain candidates are not targeting Twitter's main user base, and therefore care less about going viral. Thus, they may tweet more about policy or their positions on issues in a neutral or slightly positive tone instead of tweeting about topics that evoke emotions that tend to go viral. Similarly, candidates who are older and more highly connected to the establishment may shy away from using Twitter in the way that a younger candidate would in order to avoid seeming inauthentic like Joe Biden was perceived to be when joining SnapChat (Singh). This could cause some variance in the features of their tweets as well.

Data/Methods

In order to conduct this analysis, both polling data and Twitter data was needed. First, the polling data used to determine how a candidate is polling was retrieved by the RealClearPolitics (RCP) Python API. The poll data was scraped on November 17, 2019 and contained all of the polls in the RealClearPolitics database from September 2, 2019 to November 16, 2019, including the RCP average for November 16. The only candidates listed in these polls are the candidates that were still in the race as of November 17, 2019. The Twitter data being analyzed is a collection of all tweets with 30 words or greater between the dates of September 2, 2019 and November 23, 2019 by the candidates in Table 1 of the Appendix. Due to the fact that both Michael Bloomberg and Tom Steyer were late entries into the presidential primary, they are excluded from this analysis although they showed up in the polling numbers. Again, due to a lack of accessible polling data history, candidates who had dropped out of the race prior to November 16 are not included in the data, despite the fact that their tweets could provide valuable insight into the correlation between a candidate's tweets and how they are polling. Including both candidates that dropped out of the race early and those that entered the race late would have changed our results most likely since it would have mostly likely given us a broader understanding of candidates instead of just the candidates who follow the standard timeline for presidential candidacies. Additionally, looking at tweets and polls over a longer stretch of time and conducting a time-series analysis would have provided us with more information about how the communication styles of campaigns change over time as their respective campaign strategies begin to succeed or fail. Tweets with less than 15 words were filtered out of this analysis due to the fact that sentiment analysis is incredibly difficult and inaccurate on shorter text documents. In order to increase the quality of results, only tweets with more than 15 words were chosen for

this analysis. Had all tweets by the candidates been included, it would have made the accuracy of the results more difficult but if coded accurately, could drastically improve the quality of our responses since shorter tweets are often the ones that garner the most engagement and candidates could show different emotional tones in shorter tweets as opposed to longer tweets. Additionally, retweets are excluded since it is not the candidate who is speaking these words. Although the candidate is endorsing the message most often, it is not a message directly from the candidate, so retweets are also excluded from this analysis.

The sentiment of tweets was calculated by using the TextBlob python package (https://textblob.readthedocs.io/en/dev/). The package offers a built-in sentiment analysis tool that returns a sentiment score on a continuous scale between -1 and 1, with -1 being the most negative and 1 being the most positive. This score was calculated for each individual tweet. Some example tweets and their sentiment scores can be seen in the Appendix.

The mentions of other candidates was stored as a categorical variable where a tweet mentioning another Democratic candidate received a category 1 classification, a tweet mentioning a prominent Republican figure received a category 2 classification, a tweet mentioning both received a category 3 classification, and a tweet mentioning neither received a category 0 classification. The tweet was determined to have mentioned one of these persons if it contained one of 62 boolean search terms, which can also be found in the Appendix.

Once generating these variables, all of the tweets were placed into a dataset, where each row was a tweet containing the numerical sentiment score and the categorical mentions score. Then, for each tweet, the position in the most recent poll for the candidate sending the tweet was added as a response variable for each tweet. A regression was run exploring the correlation

between the sentiment of a candidate's tweets and their position in the polls. This process was repeated to run a regression exploring the correlation between the mentions of other candidates and prominent figures in a tweet and the candidate's polling position.

Running these regressions will help to shed light on any correlation between these features and will also return to us a p-value that will allow us to determine the significance of these results much more quantitatively than a blind eye test. Additionally, since there are over 4000 tweets that are being considered, a regression analysis will be useful in explaining the relationships for all tweets instead of just the averages of tweets.

In order to conduct this analysis, the mentions variable had to be coded on a scale of 0 to 3, with a tweet having a mentions value of 0 indicating that a tweet made no mention of other democratic candidates or high profile GOP members. A value of 1 indicated that a tweet mentioned a democratic candidate but not a GOP figure. A tweet received a mentions value of 2 if a GOP member was mentioned but no democratic candidate was mentioned and a score of 3 if both groups were mentioned in a tweet.

There are a few limitations to conducting the analysis in this way. First, an out of the box sentiment analysis package on short text documents is not incredibly reliable. Filtering for tweets that, when preprocessed, contain 15 words or more helps to make these results more accurate but there can still be a fair amount of variance in the results. Additionally, the list of boolean search terms that determine whether another candidate or high profile GOP member is mentioned may not be a complete list. For instance, if a candidate was responding to a tweet about Joe Biden and referred to him as Joe, my check would not pick up on that tweet's mention, since Joe is a fairly common name, and I did not want to include tweets about constituents mentioned that are named

Joe. The same goes for subtweeting. For instance, following Hillary Clinton's subtle jab at Tulsi Gabbard that she may be working towards the interest of a foreign government, a candidate tweeting "I am not working for a foreign government, so I have that going for me" would not be directly mentioning Rep. Gabbard, but would be mentioning her in subcontext (Beauchamp). We also have no way to account for these sort of mentions. Finally, a regression analysis is a strong tool to discover correlation between multiple explanatory variables and a response variable. However, the results can often be misleading due to confounding variables. Therefore, while not a limitation, I caution readers that the results determined from the regression analysis only indicates that there is no correlation effect between these variables and does not comment on causation.

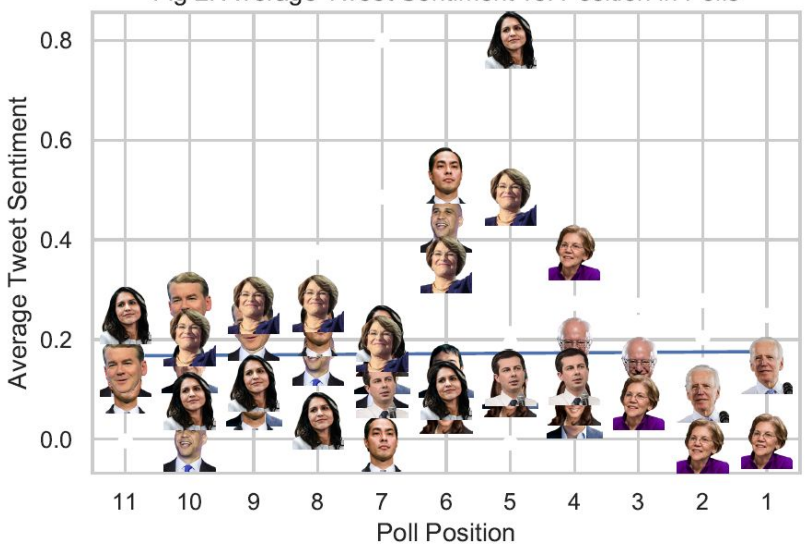
Results

My hypothesis was that there would be no correlation between a candidate's polling position and the features of that candidate's tweets, such as sentiment and the mentions of other prominent politicians. If this hypothesis was supported for sentiment analysis, we would both fail to see any discernible pattern in our graph and additionally, our regression analysis would return a p-value above our significance threshold of 0.05. If my hypothesis held for there not being a correlation between mentions in a candidate's tweets and their polling position, we would fail to see a significant difference in the distributions of mention values across the different polling positions. In our regression analysis, this would translate to again getting a p-value of greater than 0.05 for our coefficient. When running a regression analysis with both variables together,

our hypothesis would be supported if we failed to see a p-value of greater than 0.05 for any of our coefficients, whether it be the interaction effect or any of the individual variables.

We see from the results of our graphs and regression analyses that our hypothesis holds. Looking first at our graphs for sentiment analysis, we see in Figure 2 that there is no obvious correlation between sentiment analysis and where the candidate is ranking in the current polls. This graph looks simply at when a candidate is polling at a certain position, the average sentiment of their tweet. We see that Tulsi Gabbard is an outlier when ranking 5th in the polls. However, this is an outlier, given the rest of the data, and we see that there is almost no correlation between the average tweet sentiment when a candidate is polling at a position and their ranking in the polls. Running a regression analysis on the below data returns a correlation coefficient of 0.00086 with a p-value of 0.9099 indicating that we have very little evidence to reject our null hypothesis.

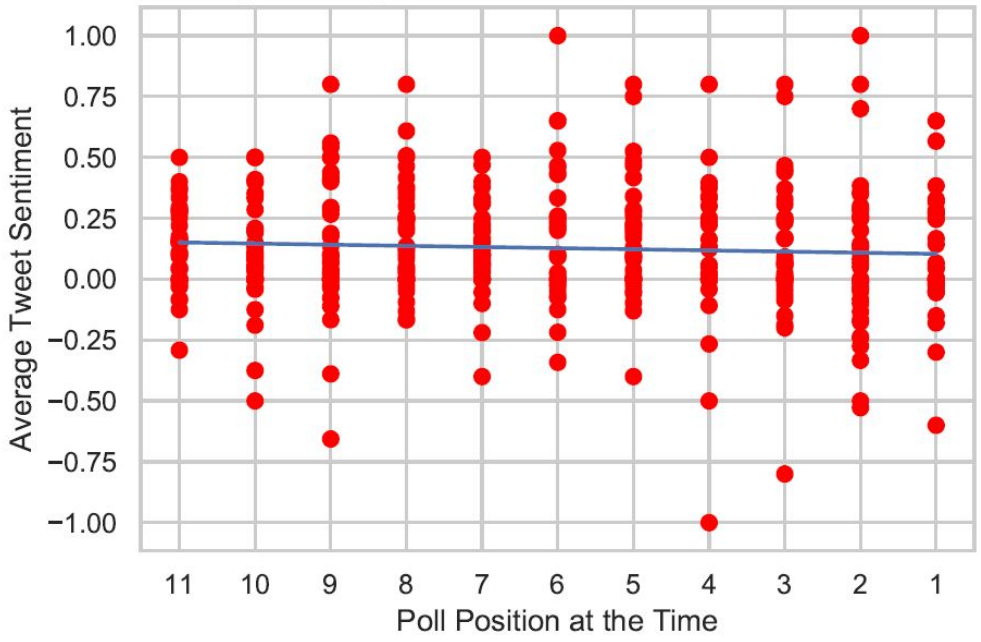
Fig 2: Average Tweet Sentiment vs. Position in Polls



Each image represents the average tweet sentiment of a candidate's tweets dependent on their poll position from September 13 and November 23. The poll position of a candidate being represented above is the most recent poll before the window of time starts. Tweet sentiment is calculated using the TextBlob Python package on a continuous scale from -1 (negative) to 1 (positive).

Looking at the correlation between sentiment of tweets and a candidate's position in the polls from a different angle, we see that the same lack of evidence for a correlation holds. When we consider the average tweet sentiment for a candidate following a single poll, so that each time a candidate is polled at a position, whether or not they were polled at that position before, we plot it as a new data point, we see in Figure 3 that the lack of correlation holds. We see a much larger range in average sentiment because there were some instances where a candidate would only tweet once between polls being released and could therefore achieve extreme values for the average sentiment of their tweets between polls. Running a simple regression analysis on the data below reveals that the data does not reject our hypothesis since its slight correlation of -0.0047 has a corresponding p-value of 0.231 indicating that this relationship is not significant.

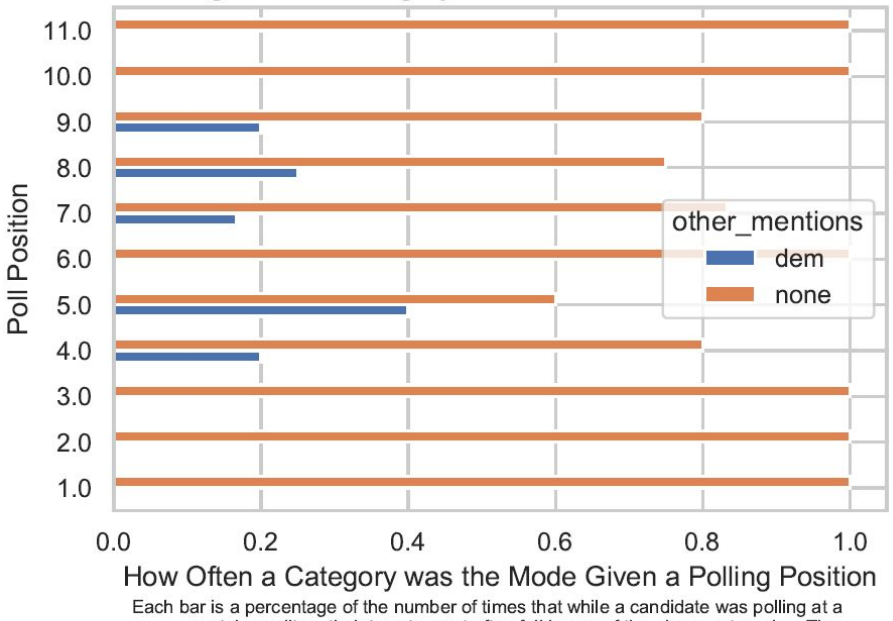
Fig 3: Average Tweet Sentiment vs. Position in Polls



Each image represents the average tweet sentiment of a candidate's tweets between polls from September 13 and November 23. The poll position of a candidate being represented above is the most recent poll before the window of time starts. Tweet sentiment is calculated using the TextBlob Python package on a continuous scale from -1 (negative) to 1 (positive).

Now looking at the same groupings of data for the mentions of others that a candidate has in their tweets, we see that again, there is no significant correlation between the mentions of prominent politicians that they make and their position in the polls. Looking back at when a candidate is polling at a certain position, similar to Figure 1's grouping of data, we see that there do not seem to be any significant trends in the data. In Figure 4, we can see the percentage of candidates who, after each poll placing them at a new ranking, most often had their tweets fall in one of the following 4 categories: no mentions of prominent politicians, only a mention of another Democratic candidate, only a mention of a major GOP figure, or a mention of both another Democratic candidate and a major GOP figure. There were no instances in which candidates while polling at a certain position, had mostly tweets mentioning GOP members or both GOP members and Democrats, which does not come as too much of a shock as most candidates would want to focus on themselves. It is a little interesting that candidates polling at the middle tend to mention other candidates. This may be due to the same theories that underdog candidates tend to attack leading candidates in more classical political communication in an attempt to gain traction in the polls (Chou and Lien). However, running a regression analysis again, we see that there is not evidence to reject our null hypothesis, so we cannot determine that there is a significant relationship between the mentions of prominent figures in a candidate's tweets and their position in the polls. Our regression analysis coefficient for the mentions of others in a tweet while looking only at a candidate's position has a p-value of 0.7176, far above our threshold of 0.05.

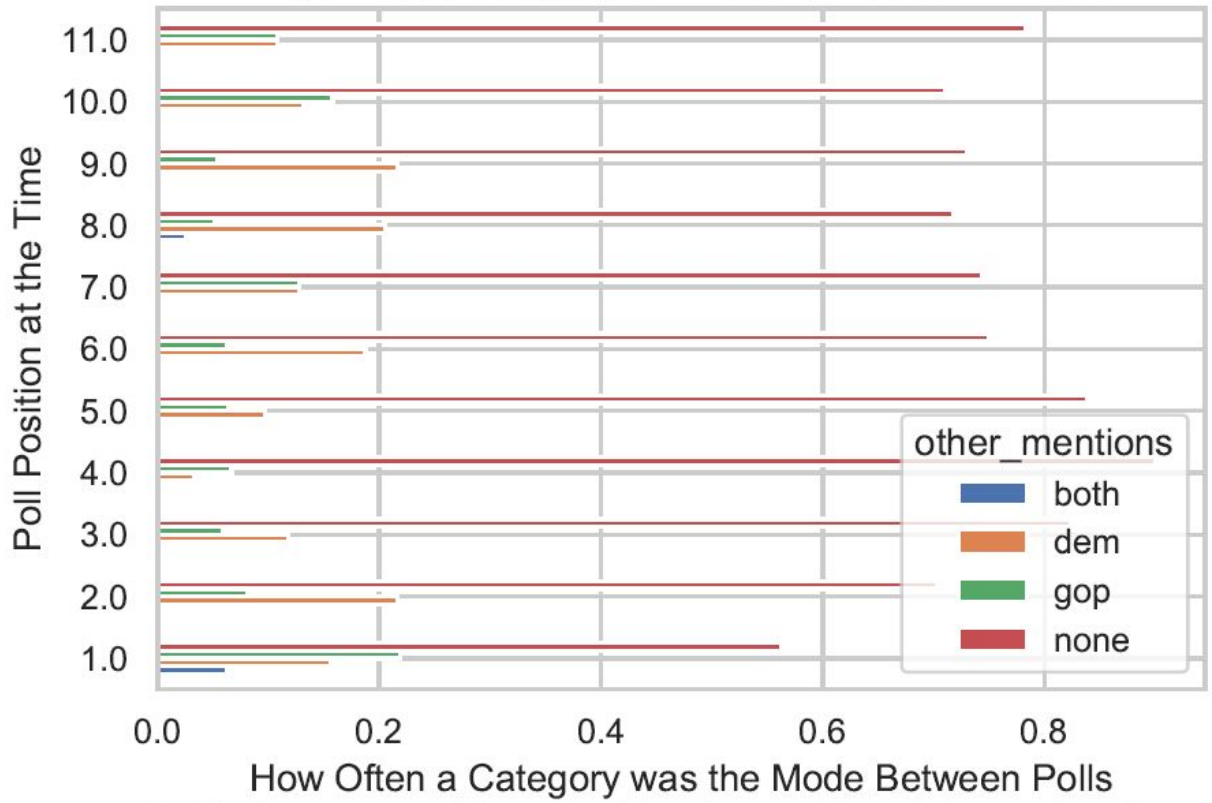
Fig 4: Mode Category of Tweet vs. Position in Polls



Each bar is a percentage of the number of times that while a candidate was polling at a certain posiiton, their tweets most often fell in one of the above categories. The poll position of a candidate being represented above is the most recent poll before the window of time starts. The boolean search terms used to generate the category classifications can be found in the appendix

Looking at grouping the data so that we look at the mode category of tweets for a candidate after each new poll, we again see a lack of a significant relationship between the mentions a candidate makes of prominent figures and their polling. This time, running our regression analysis returns a p-value of 0.9454, and again we fail to gather any significant relationship to reject our null hypothesis. However, we do see in Figure 5 that when candidates are polling in first, they often tend to mention GOP figures more and mention no one in their tweets less often, which may be an indication of them becoming more confident that they will face Donald Trump in the general election and mentioning him by name more often. However, it could also be an indication of the fact that a candidate who can beat Trump is important to voters this election cycle, and thus, the candidates who talk most about Trump are polling highest.

Fig 5: Mode Category of Tweet vs. Position in Polls



Each bar is a percentage of the number of times that a candidate polled at a certain posiiton and before the next poll, their tweets most often fell in one of the above categories. The poll position of a candidate being represented above is the most recent poll before the window of time starts. The boolean search terms used to generate the category classifications can be found in the appendix

Discussion

The findings in this paper support our hypothesis that there is no correlation between the sentiment of a candidate's tweets and their position in the polls, nor the mentions of other prominent politicians that a candidate includes in their tweets and their position in the polls. Plotting the relationships between these variables indicated that there is no relationship between the sentiment of a candidate's tweets and their position in the polls. While the bar graphs for the categories that candidates most tweeted about after being ranked by a new poll indicated that

there may be some interesting trends there, the graphs seemed to confirm for the most part that there was no relationship between a candidate's position in the polls and their mentioning of others in their tweets. Running regression analyses on these datasets confirmed that there was no statistically significant relationship between the features of a candidate's tweets and their poll rankings. We saw that our p-values for significance were consistently greater than 0.05 and this helped cement our initial inclinations with statistical evidence as well.

Most of this analysis was designed to be reproducible, using open source Python packages and lacking human interpretation wherever possible. Hopefully, this implies that the analysis could be redone by other researchers and they would be able to achieve the same results. However, if this study was conducted over a different time frame, or on a different election, the results may vary wildly. Prior research indicated that in some European countries, that the tweets of candidates vary with their performance in the election or their popularity (Karlsen and Enjolras; Kruikemeier). Similarly, a study conducted later in the same election that I studied may return different results as candidates who are failing to gain traction adopt more extreme measures, and if there is a single candidate who is leading the pack towards the end of the race, they may begin to shift their focus towards their general election competitor and less so on their fellow democratic candidates. On that same note, a similar study conducted on a republican primary may return different results as voter bases, information bubbles, and party dynamics could play a large role in messaging on Twitter.

Tweets containing less than 15 words were removed from this study as well in order to improve sentiment analysis results but keep enough tweets to generate meaningful insights. Having a different threshold for the number of words in a tweet may also return different results

that could prove insightful. Additionally, as mentioned earlier, this study omitted both early dropouts and latecomers to the primary. Therefore, valuable information could be lost about the tactics that these candidates tend to employ on Twitter. For instance, Beto O'Rourke had a very interesting role on Twitter as a pro-gun control candidate who was incredibly impassioned following the mass shooting in El Paso. The omission of his unique tweets from this study may have prevented a valuable insight from being made. Finally, if the study is conducted in another year, a researcher might find different results as the Trump presidency has lent itself to many unique situations. If a study was conducted in a year that seemed to be less quiet and more normative, a researcher might find either more normative results, or even more extreme results from candidates who are trying all they can to be noticed by a disengaged general public.

In future research, I would recommend implementing topic modelling on Tweets as well. It was omitted from this study since topic modelling on short text documents is incredibly difficult and I do not have that skill yet, but it would be interesting to see if the trends that Karlsen and Enjolras, and Kruikemeier found in other countries maintained here in the United States. Also, I would recommend attempting different ways of exploring the relationship between the features of a tweet and the candidate's position in the polls. For instance, both clustering algorithms and SVM algorithms might be an interesting way to go about exploring this relationship. I think it would also be incredibly interesting to test different populations as well. For instance, reproducing the study with all candidates who ran in the 2020 democratic primary, including those who dropped out early and those who joined late would be a great place to expand on this research. Perhaps most interesting would be reproducing this study on republican primary candidates. This could be done on candidates from the 2016 election cycle. Although an

odd year, and an especially odd primary, for presidential elections in the United States, it would be an incredibly interesting analysis that should not be avoided because it was an odd year in US politics.

Appendix

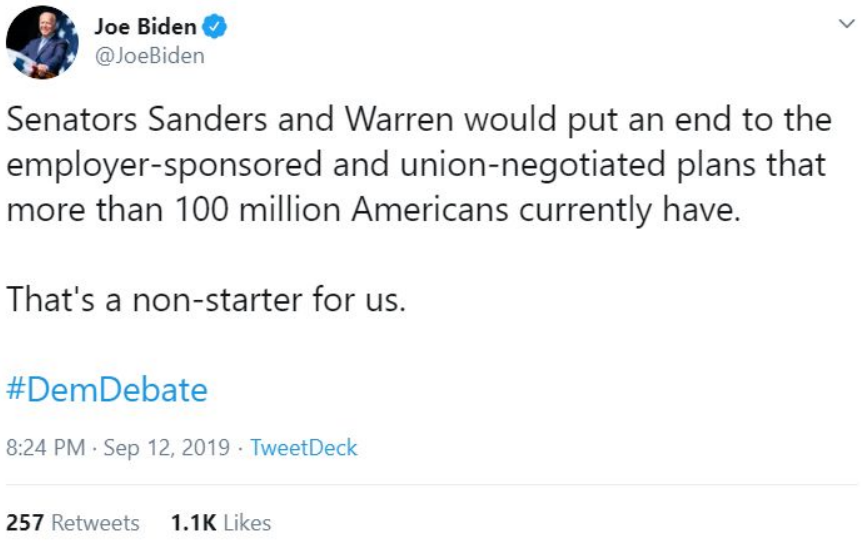
Table 1 (Candidates Under Consideration):

Candidate Name	Candidate Twitter Handle
Andrew Yang	@AndrewYang
John Delaney	@JohnDelaney
Elizabeth Warren	@ewarren
Beto O'Rourke	@BetoORourke
Pete Buttigieg	@PeteButtigieg
Bernie Sanders	@BernieSanders
Steve Bullock	@GovernorBullock
Michael Bennet	@MichaelBennet
Amy Klobuchar	@amyklobuchar
Kamala Harris	@KamalaHarris
Tulsi Gabbard	@TulsiGabbard
Marriane Williamson	@marwilliamson
Julian Castro	@JulianCastro
Cory Booker	@CoryBooker
Joe Biden	@JoeBiden
Jay Inslee	@JayInslee
John Hickenlooper	@Hickenlooper
Eric Swalwell	@EricSwallwell
Bill De Blasio	@BillDeBlasio
Kirsten Gillibrand	@SenGillibrand
Seth Moulton	@sethmoulton
Tim Ryan	@TimRyan

Table 2 (Boolean Search Terms for Mentions Classification):

Democratic Search Terms	'bennet', 'biden', 'bloomberg', 'booker', 'buttigieg', 'castro', 'delaney', 'gabbard', 'klobuchar', 'patrick', 'sanders', 'steyer', 'warren', 'williamson', 'yang', 'bullock', 'de blasio', 'gillibrand', 'harris', 'hickenlooper', 'inslee', 'messam', 'moulton', 'ojeda', 'o'rourke', 'ryan', 'sestak', 'swalwell', 'beto', 'kamala', 'julian', 'tulsi', 'bernie', 'kirsten', 'mayor pete', '@andrewyang', '@johndelaney', '@ewarren', '@betoorourke', '@petebuttigieg', '@berniesanders', '@governorbullock', '@michaelbennet', '@amyklobuchar', '@kamalaharris', '@tulsigabbard', '@marwilliamson', '@juliancastro', '@corybooker', '@joebiden', '@jayinslee', '@hickenlooper', '@ericswallwell', '@billdeblasio', '@sengillibrand', '@sethmoulton', '@timryan'
GOP Search Terms	'trump', 'mitch', 'mcconnell', '@realdonaldtrump', '@senatemajldr'

Figure 6 (Example Tweets and Sentiment Scores)
6a. (Biden)



Sentiment score: 0.00

6b. (Warren)

Sentiment Score: 1.00

6c. (Buttigieg)

Sentiment score: -1.00

6d. (Yang)

Sentiment score: 0.30833

All Code completed in this analysis can be found at:

https://github.com/pmohamma/political_tweet_analysis

Works Cited

Beauchamp, Zack. "The Hillary Clinton-Tulsi Gabbard Feud, Explained." *Vox*, 22 Oct. 2019. *Vox*, www.vox.com/policy-and-politics/2019/10/22/20924999/hillary-clinton-tulsi-gabbard-fight-explained. Accessed 11 Dec. 2019.

Berger, Jonah, and Katherine L. Milkman. "What Makes Online Content Viral?" *Journal of Marketing Research*, 1 Apr. 2012, doi.org/10.1509%2Fjmr.10.0353. Accessed 11 Dec. 2019.

Bermingham, Adam, and Alan Smeaton. "On Using Twitter to Monitor Political Sentiment and Predict Election Results." *ACL Anthology*, vols. Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), 2011, pp. 2-10, www.aclweb.org/anthology/W11-3702. Accessed 11 Dec. 2019.

Biden, Joe. "Senators Sanders and Warren would put an end to the employer-sponsored and union-negotiated plans that more than 100 million Americans currently have. That's a non-starter for us. #DemDebate." *Twitter*, 12 Sept. 2019, 8:24 p.m., twitter.com/JoeBiden/status/1172305094802432000.

Buttigieg, Pete. "A 16-year-old girl. A 14-year-old boy. My heart breaks for their families, the Santa Clarita community, and our country. America is sacrificing a

generation to the devastating effects of gun violence because our leaders in Washington lack the will to save lives. We must act." *Twitter*, 14 Nov. 2019, 9:56 p.m., twitter.com/PeteButtigieg/status/1195173800779886592.

Ceron, Andrea, et al. "Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens' Political Preferences with an Application to Italy

2

and France." New Media & Society, vol. 16, no. 2, 4 Apr. 2013, pp. 340–358., doi:10.1177/1461444813480466.

Chou, H.-Y., and N.-H. Lien. "The Effects of Appeal Types and Candidates' Poll Rankings in Negative Political Advertising.'" *Asian Journal of Communication*, 18 Oct. 2012, doi.org/10.1080/01292986.2012.731607. Accessed 11 Dec. 2019.

Chung, Jessica Elan, and Eni Mustafaraj. "Can Collective Sentiment Expressed on Twitter Predict Political Elections?" *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 4 Aug. 2011, pp. 1770-71, www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewPaper/3549. Accessed 11 Dec. 2019.

Francia, Peter L. "Free Media and Twitter in the 2016 Presidential Election: The Unconventional Campaign of Donald Trump." *Social Science Computer Review*, 27 Sept. 2017, doi.org/10.1177%2F0894439317730302. Accessed 11 Dec. 2019.

Karlsen, Rune, and Bernard Enjolras. "Styles of Social Media Campaigning and Influence in a Hybrid Political Communication System: Linking Candidate Survey Data with Twitter Data." *The International Journal of Press/Politics*, 27 Apr. 2016, doi.org/10.1177%2F1940161216645335. Accessed 11 Dec. 2019.

Kruikemeier, Sanne. "How Political Candidates Use Twitter and the Impact on Votes." Computers in Human Behavior, vol. 34, May 2014, pp. 131–139., doi:10.1016/j.chb.2014.01.025.

Shearer, Elisa. *Social Media Outpaces Print Newspapers in the U.S. as a News Source*. 10 Dec. 2018. *Pew Research Center*, www.pewresearch.org/fact-tank/2018/12/10/social-media-

outpaces-print-newspapers-in-the-u-s-as-a-news-source/. Accessed 11 Dec. 2019.

Singh, Maanvi. " 'I'm on Snapchat!': Joe Biden's 'exciting' Announcement Draws Snark." *The Gaurdian* [San Francisco], 31 July 2019. *The Guardian*,

www.theguardian.com/us-news/2019/jul/31/biden-snapchat-democrats-2020-social-media. Accessed 11 Dec. 2019.

Warren, Elizabeth. "I had a wonderful time meeting Lily and her friend Jenny. Lily donated just $10 and won the @TeamWarren contest to join me at tonight's #DemDebate in Atlanta! Thank you both for being a part of this grassroots movement." *Twitter*, 20 Nov. 2019, 7:50 p.m., twitter.com/ewarren/status/1197316212776144898.

Yang, Andrew. "Great speaking to the National Action Network this morning - we must further Martin Luther King's vision of an economically just society. 'Capitalism forgets that life is social.'" *Twitter*, 21 Nov. 2019, 10:33 a.m., twitter.com/AndrewYang/status/1197538586528227328.