Analyzing Public American Sentiment on Twitter During the First Three Waves of the COVID-19 Pandemic

May 19, 2021

Word Count: 4448

Introduction

A field of study in computer science and linguistics that has garnered more attention in recent years is Natural Language Processing (NLP), which aims to find methods of programming computers so that they are able to comprehend and analyze human language. One particular subfield within NLP is sentiment analysis, a research area aiming to analyze human emotions in text, which has recently risen in popularity. The uses of sentiment analysis are covered by some interesting scientific and commercial areas, such as risk assessment, marketing, and recommender systems. The popular social media platforms of Twitter, Facebook, and Reddit are giant publicly available databases of humans expressing their emotions in text. Users frequently discuss events impacting their daily life and often inject their own opinion within their posts. The position of social media in our society as a virtual space where everyone can voice their experiences and honest opinions whenever they want to make social media the perfect indicator of broader public attitudes.

The novel coronavirus now known as COVID-19 was discovered in December 2019 in Wuhan, China. Since then, it has come to dominate every aspect of the daily lives of Americans, becoming one of the most discussed and researched issues in many fields from economics and politics to biology and psychology. According to the World Health Organization (2021) at the time of writing, there have been more than 150 million cases of COVID-19 and more than 3 million deaths worldwide since the emergence of the virus. From this data, we can see that this is the most severe and impactful virus outbreaks of the past century. While the health impacts of the virus itself are clear, analyzing and getting new insights from data on social media platforms could help health staff and government organizations gauge mental well-being, and understand the emotions and needs of the general population. In this study, tweets have been taken from posts on Twitter, known as tweets, containing specified search keywords related to the COVID-19 pandemic. We can extract these tweets and perform sentiment analysis on the datasets using the Python programming language. Python provides many easy-to-use libraries to process data and creates models for natural language processing. By using a public database, sentiment analysis can be applied to the data available, and a detailed explanation can be produced.

Literature Review

It is already established that the COVID-19 pandemic is having a large impact on mental well-being in the United States. In particular, DePierro et al. of the Icahn School of Medicine at Mount Sinai warn that the effect is especially pronounced among the infected and their family, medical workers, and other essential workers (2020). They predict that COVID-19 will have lasting impacts on mental health demonstrated through higher rates of PTSD, depression, and substance abuse. However, the effects on human mental health of the COVID-19 pandemic are also visible in the general population. For instance, a poll conducted by KFF, a non-profit organization that focuses on national health issues and American stakes in global health policy, shortly after stay-at-home orders became widespread revealed that 47% of respondents who were sheltering in place experienced COVID-related stress or worry (Kirzinger et al., 2020). The immense impact of the pandemic on the mental well-being of the general population exhibits the necessity of research on the changing perceptions within the general public of the pandemic.

Recently, the use of natural language processing of content on social media has become a viable means of surveying the population. In the business world, researchers Jackson et al. from the University of Charleston demonstrate that sentiment analysis of social media can be used to investigate consumer behaviour and build marketing strategies that manage risk more successfully (2020). The paper also concludes that those who used analytics of social media sentiments possessed a measurable advantage over competitors who relied on traditional surveying methods and were even less prone to the risks incurred by the volatility from the COVID-19 pandemic. Similarly, natural language processing on social media has also been able to assess the suicide risk of users without requiring additional time from them. Coppersmith et al. from Qntfy, a technological solutions provider aiming to bridge data science and human behaviour by gathering complex psychological and behavioural data, demonstrate in a paper published in the *Proceedings from the Digital Mental Health Conference - London* the viability of using natural language processing and machine learning to evaluate suicide risk (2018). The researchers were able to detect quantifiable signals around suicide attempts precisely enough to provide utility for healthcare professionals. Although the study acknowledges that the usefulness of such a method is dependent on the individual's presence on social media, it reaffirms that social media posts provide portrayals of mental and emotional states that are accurate enough to be worthwhile in research.

As noted by the previous study, natural language processing of social media depends on users' activity. Therefore, a reduction in social media use among the general population would decrease the accuracy of such a study in reflecting the sentiments of the populace. Fortunately, journalists Ella Koeze and Nathaniel Popper report in an article for the *New York Times* that visits to major social media sites such as Facebook have grown during the COVID-19 pandemic (2020). Based on these results, one can assume that there is greater activity on social media. Therefore, the accuracy of sentiment analysis through natural language processing as a means of surveying the population would not be negatively impacted by the pandemic itself.

An example of research conducted on the effects of the COVID-19 pandemic on mental health is an ongoing study titled “Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: an observational study” by researchers from Harvard Medical School including Daniel M. Low as well as Guillermo Cecchi from IBM’s Thomas J. Watson Research Center (2020). Using textual analysis techniques from natural language processing on posts within specific groups on Reddit such as r/schizophrenia and r/PersonalFinance, the researchers were able to obtain early results that tracked mental health complaints as they arose in real-time and identified themes of interest relevant to the COVID-19 pandemic. As the real-time data was found to have a direct relationship with developments in the pandemic, it can be concluded that language used on social media is an accurate and timely portrayal of attitudes within a population. While Low, Rumker, Talkar, Torous, Cecchi, and Ghosh’s study had a scope focused on the specific area of mental health and thus examined limited areas of the population, other studies successfully used sentiment analysis to examine attitudes of the general population. While Reddit as a platform already organized into communities served the purpose of the previous study well, Twitter is a more efficient platform to study for sentiments within the general population. A study that used sentiment analysis of Twitter to investigate the emotions of a general population was conducted by researchers Patricia P. Iglesias-Sánchez et al. of the University of Malaga in Spain and published in a special issue of the *International Journal of Environmental Research and Public Health* (2020). The researchers focused on changes in emotions within the population in Spain after the instatement of confinement measures in March and April of 2020. Unsurprisingly, they found peaks in the emotions of anger and fear, especially as the pandemic death toll climbed.

The upward trends of fear and anger were reflected in other studies as the novel coronavirus spread in Spring 2020. A study by researchers from the University of Toronto, University of Pittsburgh, and Chinese Academy of Sciences analyzing 1.9 million tweets from January 23 to March 7 found that fear was the dominant element in tweets as users discussed the unknown nature of the virus (Xue et al., 2020). Another study spanning a wider time range from December 2019 to April 2020 by researchers from the German Aerospace Center's Institute of Data Science and the Technical University of Munich comparing sentiments within the different languages and countries in Europe found a deterioration in mood in all countries when their respective governments announced lockdown measures (Kruspe et al., 2020). However, the study detected a recovery in mood in March and April, although not enough to return to pre-pandemic levels.

Most information regarding the evolution of attitudes towards the pandemic on social media aligns with the results of a study published in *JMIR Public Health and Surveillance* by a multinational team of researchers from Nanyang Technological University in Singapore, Tianjin University in China, the University of Lugano in Switzerland, and the University of Melbourne in Australia (Lwin et al., 2020). Their analysis of global tweets related to COVID-19 corroborates the studies previously mentioned, concluding that emotions of fear rose rapidly at the onset of the pandemic but slowly shifted into anger. Much like the previous studies, positive emotions of joy and gratitude appeared to increase near the end of their observation period but the overall effect of the pandemic is detrimental and not yet recovered from.

More recent data on the American outlook on COVID-19 has surfaced in the form of a survey conducted by Justin McCarthy of the polling company Gallup containing data up to August 30, 2020 (2020). It confirms the trend of increasing optimism seen at the end of Lwin et al.'s study and verifies that such emotions also apply to Americans. An increasing proportion of Americans seem to believe that the pandemic situation will improve, which results in less concern and more optimistic views on their future. However, the results do not include attitudes during and following the fall surge in COVID-19 cases, so it is unclear whether the optimistic feelings have persisted.

Method

Demographics are not available for the sample of COVID-19 related tweets, but the collective demographics of Twitter users can be estimated. As of 2020, there are approximately 70 million Twitter users in the United States. According to data gathered by H. Tankovska, a research expert on social media working for Statista, around 61.6% of these users are male while the remaining 38.4% are female (2021). However, a survey conducted by the Pew Research Center in 2019 concluded that women are more active on Twitter with 65% of the most active 10% of Twitter users identifying as female (Wojcik & Hughes, 2019). The same survey also points out discrepancies between active users and the general American population. Twitter users tend to be younger, with a median age of 40 compared to the median of 47 within the general population. Users also tend to be more highly educated, with 42% having completed at least a bachelor's degree compared to 31% among all Americans. Twitter users also have a higher median income than the average American, but the survey concludes that the overall gender, racial, and ethnic makeup of Twitter is reflective of the general adult population (Wojcik & Hughes, 2019). This study includes only data from users who have used COVID-19 related hashtags such as #covid19 and #coronavirus, but specific statistics on this population are unknown.

Data used will be from Rabindra Lamsal's COVID-19 tweets dataset, collected as part of a project for the School of Computer and Systems Sciences for Jawaharlal Nehru University (2020). To filter the data for relevant tweets, only posts containing the COVID-19 related hashtags. (e.g #corona, #coronavirus, #covid, #covid19, #covid-19, #pandemic, #quarantine, #flatteningthecurve) The dataset contains .csv files listing hundreds of tweet IDs, which must be hydrated in order to extract its textual content. This is achieved with Twitter's public API and the public Python library twarc. The tweets are then scanned for geographically identifying content (e.g. New York, Florida) to narrow the scope to tweets originating from the United States. There remains a possibility that some tweets may mention locations associated with the United States without originating from an American resident, but such cases should be rare. Hyperlinks and images are also eliminated from the data as only the textual content is for analysis.

The study focuses on the time period from March 20th, 2020 to March 7th, 2021, which encompasses a year since the declaration of the coronavirus pandemic. This period encapsulates the majority of the first three waves in the United States, during the third of which the US reached its highest ever COVID-19 case numbers. The data will be organized into three more specific time periods that approximately represent one of the three major waves each: March 20, 2020 to June 15, 2020; June 16, 2020 to September 30, 2020; and October 1, 2021 to March 7, 2021. For the purposes of this study, these time periods will represent the first, second, and third waves of the pandemic within the US respectively. Data from each of the three waves will be compared to one another in order to determine whether there have been noticeable changes in public attitudes.

The natural language processing in this study is conducted using the API TextBlob, a Python library for simple textual processing (Loria, 2020). The values representing the positivity of attitudes within tweets are the polarity values of TextBlob's sentiment analysis feature. Tweets are identified with either positive or negative sentiments and ranked on a scale from -1.0 to +1.0, with negative values corresponding to negative sentiments and positive values corresponding to positive sentiments. The magnitude of the number serves as an estimation for the strength and polarity. The data is sorted by the date of the tweet's posting, which allows me to find possible trends in sentiment over time. Although there are hundreds of tweets for each day available in the dataset, the precise number varies from day to day as a result of Twitter activity and geographic filtering. In order for each date to have equal effects on the final result, the mean sentiment score of the tweets of each day is calculated along with the data's standard deviation and then merged together into a single spreadsheet. The mean daily sentiment scores are the values used when determining correlations and trends. During the process of data gathering, I was unable to retrieve enough data for tweets posted on the following dates as the public database used was missing the file or had corrupted data: March 11, 2020; April 12, 2020; and October 27, 2020. As a result, all statistics and conclusions reached in this paper will not include these three dates.

Gathering External Data

All of the data collected will be analyzed through correlation with three statistics collected over the same time periods: daily new COVID-19 cases, daily new COVID-19 deaths, and the Dow Jones Industrial Average at closing. The first two metrics were chosen in order to examine how large of an influence the progression of the pandemic had on public attitudes, if any. Since new positive cases increase and peak at different times than new deaths, both statistics are compared to the data independently. Case count and death count statistics used in this study were retrieved from the dataset of the COVID Tracking Project, a volunteer organization launched from *The Atlantic* that works to collect data on COVID-19 in the United States for news agencies and research projects (The Atlantic, n.d.). Data of the Dow Jones was retrieved from the publicly available historical data on Yahoo! Finance, a part of the Yahoo! Network that provides financial news, data, and stock quotes (Yahoo! Finance, 2021). The DJIA was chosen as the third statistic to correlate with mean sentiments on Twitter as it provides the opportunity to analyze the issue through an economic lens, whereas the other two statistics represent a public health perspective.

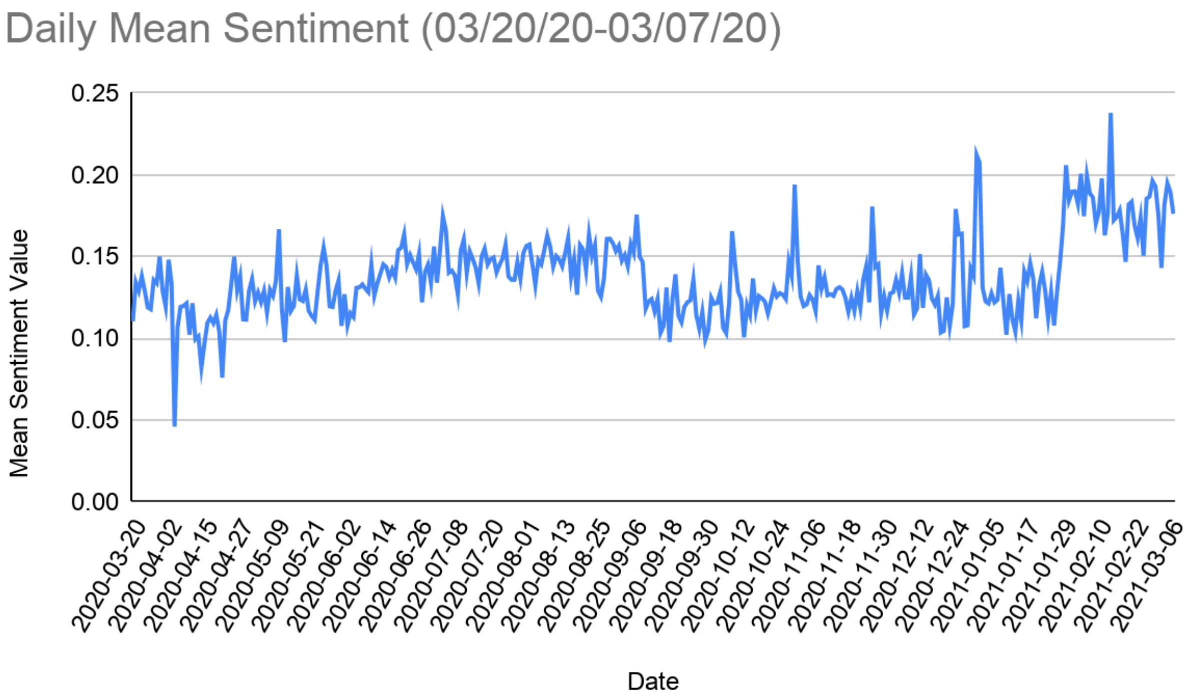
Results

In total, 345,880 tweets were analyzed, with values fluctuating between 0.04 (which occurred on April 4, 2020) and 0.24 (which occurred on February 14, 2021). An average of around 930 tweets per day was analyzed, but the precise number of tweets gathered each day varies widely, with the largest number of tweets gathered in a single day being 2028 on July 4th, 2020, and the lowest number of tweets gathered in a single day being 318 on October 2, 2020. The mean sentiment value of the data collected is 0.1399.

The evolution of mean sentiment over the entire time period from March 20, 2020 to March 7, 2021 is graphed in Figure 1 below. The figure depicts the overall progression of Twitter sentiment over time with the date on the horizontal axis and the mean sentiment value, which can range from -1 to 1, on the vertical axis. The mean sentiment stayed in the positive range for the entirety of the examined period, indicating that sentiments on Twitter remained consistently positive throughout the first year of the pandemic. For visual clarity, Figures 2-7 only exhibit data from either the first and second waves combined or the third wave.

Figure 1

Graph of daily mean Twitter Sentiment



Figures 2 and 3 depict the relationship between mean Twitter sentiment and increases in the US COVID-19 case count. Since not all counties and states report statistics on weekends, the graphed values are the 7-day rolling averages of both Twitter sentiment and increases in positive cases in order to correct for sudden drops and surges in case counts caused by inconsistent reporting. While it is possible to visually determine qualitatively the relationship between the 7 day sentiment average and 7 day positive cases average, correlation allows for the quantification necessary to compare how the relationship evolves over the three waves. The correlation between sentiments of the first, second, and third waves with the increases in positive cases over the same time period is -0.471, 0.213, and -0.371, respectively. In order to determine if the differences in correlational values are statistically significant at an alpha level of 0.05, one can conduct a t-test for correlation. The results of the t-test demonstrate that there is a statistically significant difference between the correlational values of the first and second waves and the second and third waves. This indicates that it is highly unlikely for such a change in correlation to be a result of chance, implying the existence of an underlying cause. However, a t-test is unable to obtain statistical significance when comparing the first and third waves, which evinces that the correlation between Twitter sentiment and COVID-19 case increases did not significantly change between the first and third waves.

Figure 2

Graph of daily mean Twitter Sentiment compared to daily US COVID case increases during the first and second waves

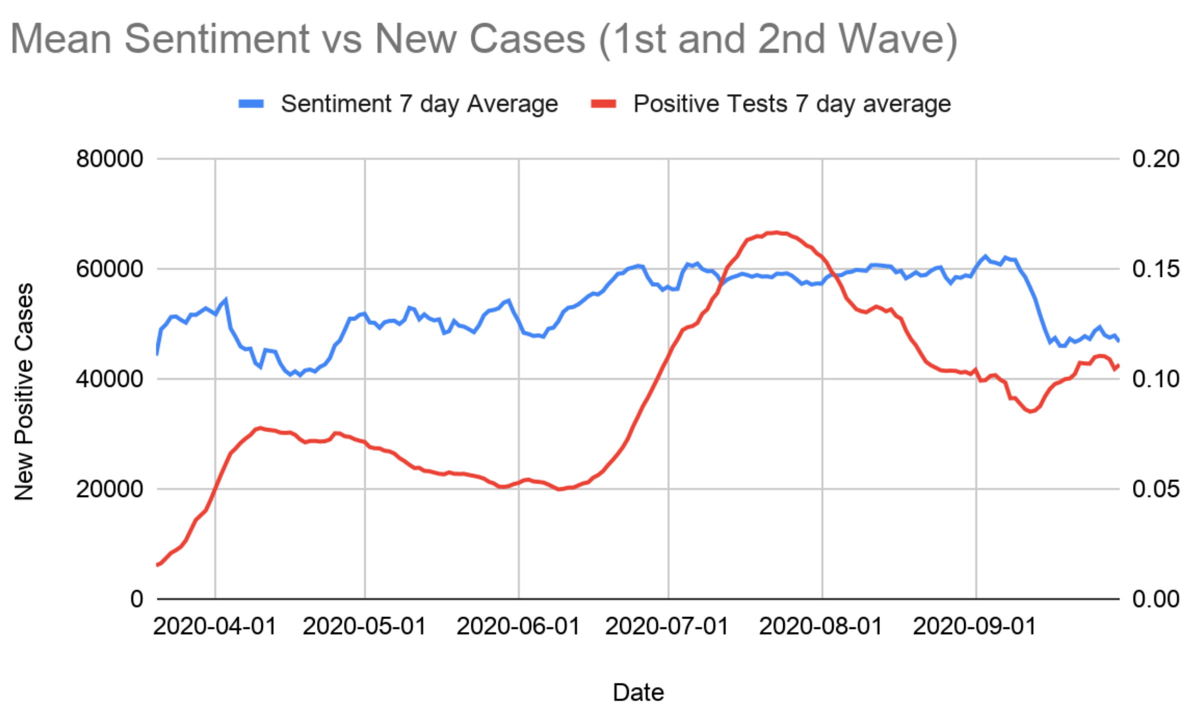
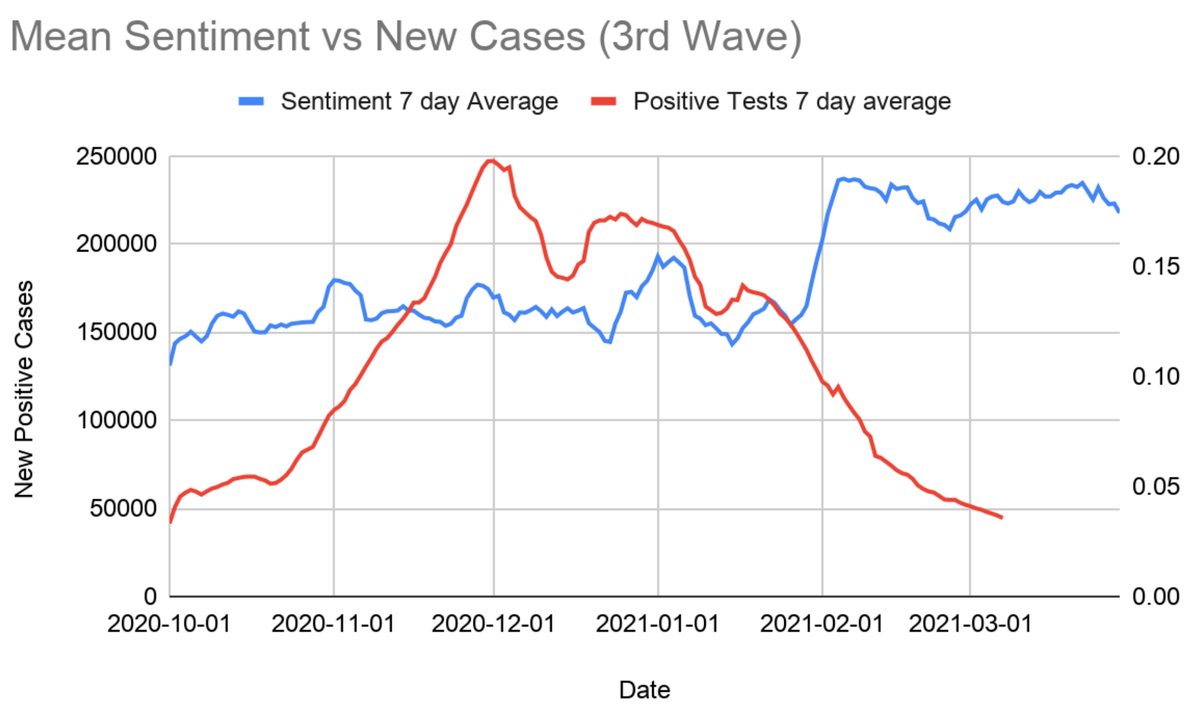


Figure 3

Graph of daily mean Twitter Sentiment compared to daily US COVID case increases during the third waves



In similar fashion to Figures 2 and 3, Figures 4 and 5 depict the relationship between mean Twitter sentiment and increases in US COVID-19 deaths. Once again, the 7-day rolling averages of both values are graphed to account for reporting disparities on weekends. The correlation between sentiments of the first, second, and third waves with the increases in deaths due to COVID-19 over the same time period is -0.541, 0.136, and -0.601, respectively. A t-test conducted for these values unveils results identical to the t-test conducted for the relationship for new positive cases. There is a statistically significant difference between the correlation values of the second wave and both the first and third waves, but there is no significant difference directly between the first and third waves.

Figure 4

Graph of daily mean Twitter Sentiment compared to daily US COVID death increases during the first and second waves

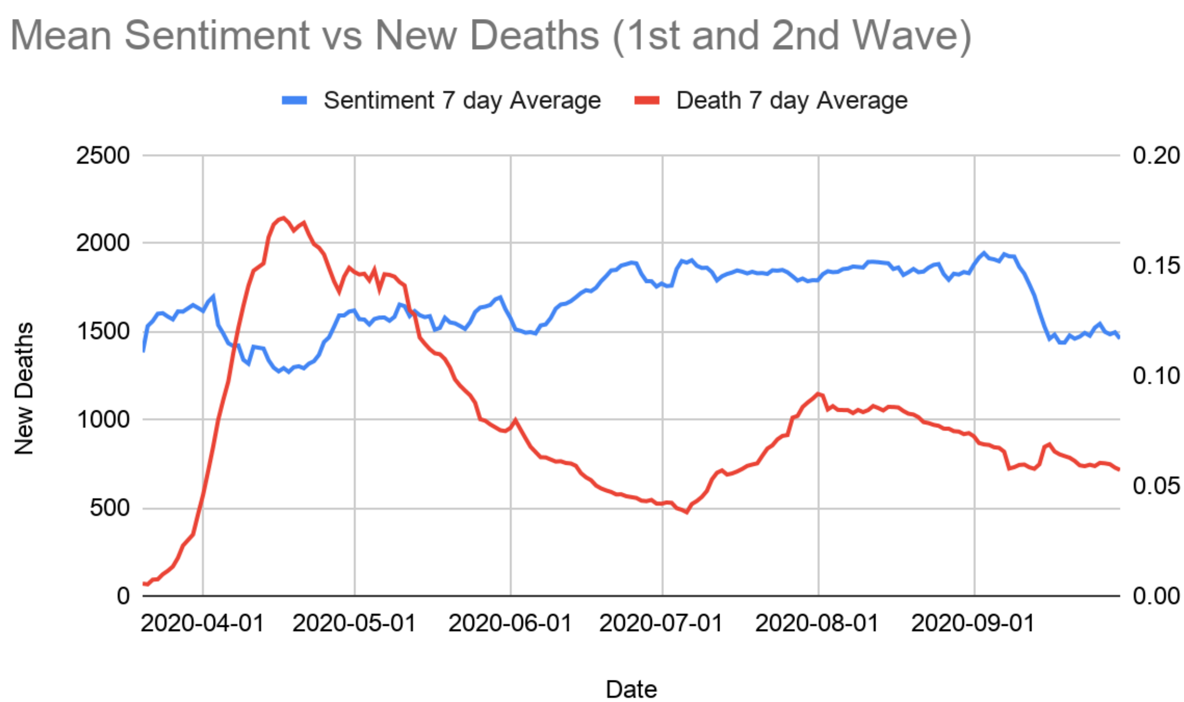
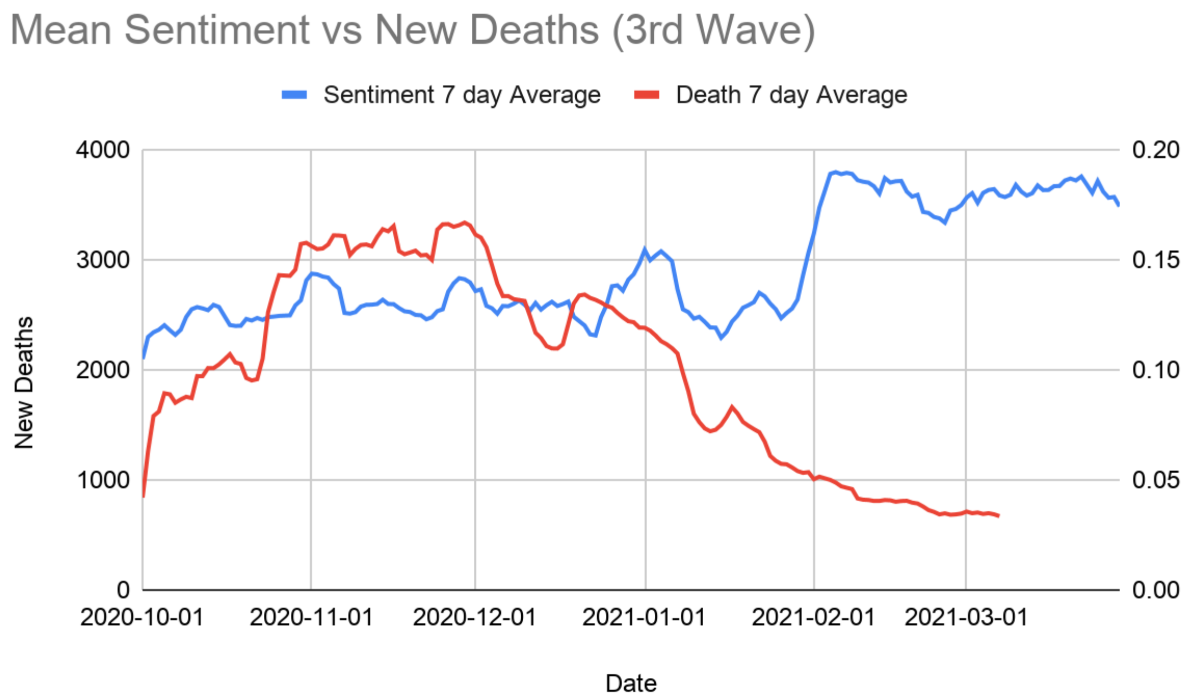


Figure 5

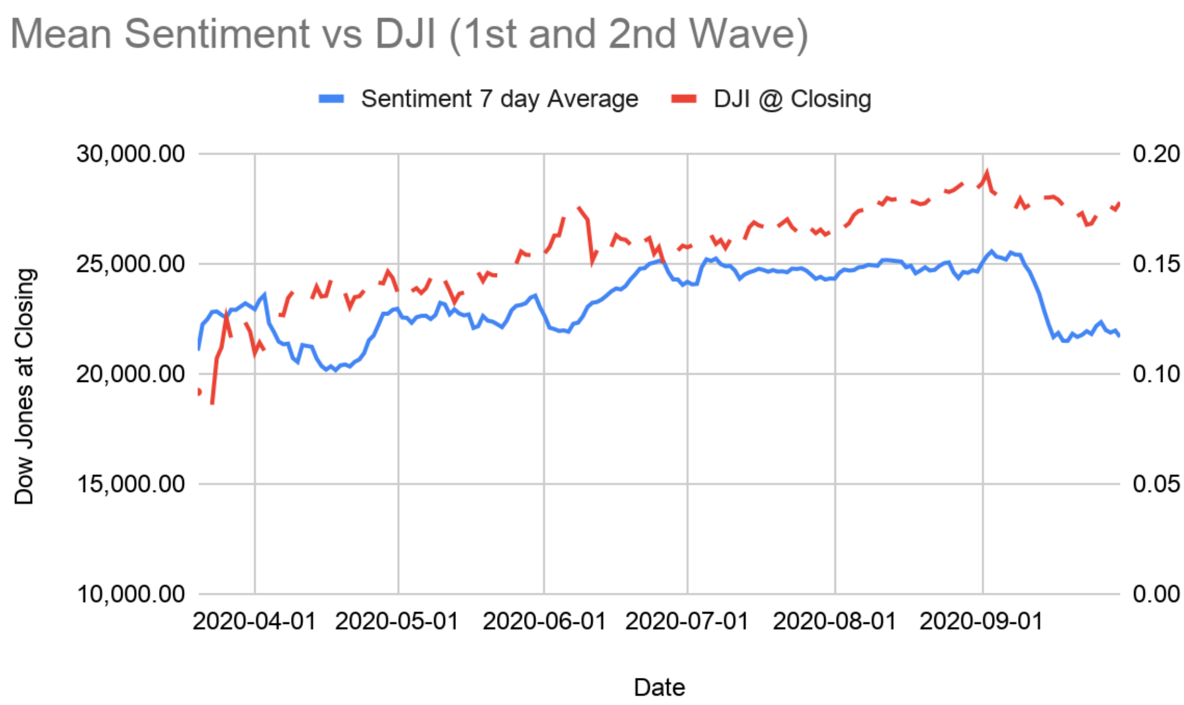
Graph of daily mean Twitter Sentiment compared to daily US COVID death increases during the third waves



Figures 6 and 7 depict the relationship between mean Twitter sentiment and the value of the Dow Jones Industrial Average (DJIA) at closing on days when the stock market is open for public trading. Unlike the previous statistics, there is no reason to take the 7 day average of the DJIA. However, the 7 day average of Twitter sentiment is still used to maintain consistency, as it would be illogical to compare the other two statistics to the 7-day average while using the daily values to calculate correlation with the DJIA. The correlation between sentiments during the first, second, and third waves with the DJIA over the same time period is 0.075, -0.072, and -0.714, respectively. The results of the t-test conducted for these values are dissimilar from results identical of t-tests conducted for the other two statistics; there is no statistically significant difference between the correlation values of the first and second waves, but there is a statistically significant difference between the first and third waves as well as the second and third waves. This indicates that the correlation remained largely unchanged during the first two waves, but underwent a massive shift during the third wave.

Figure 6

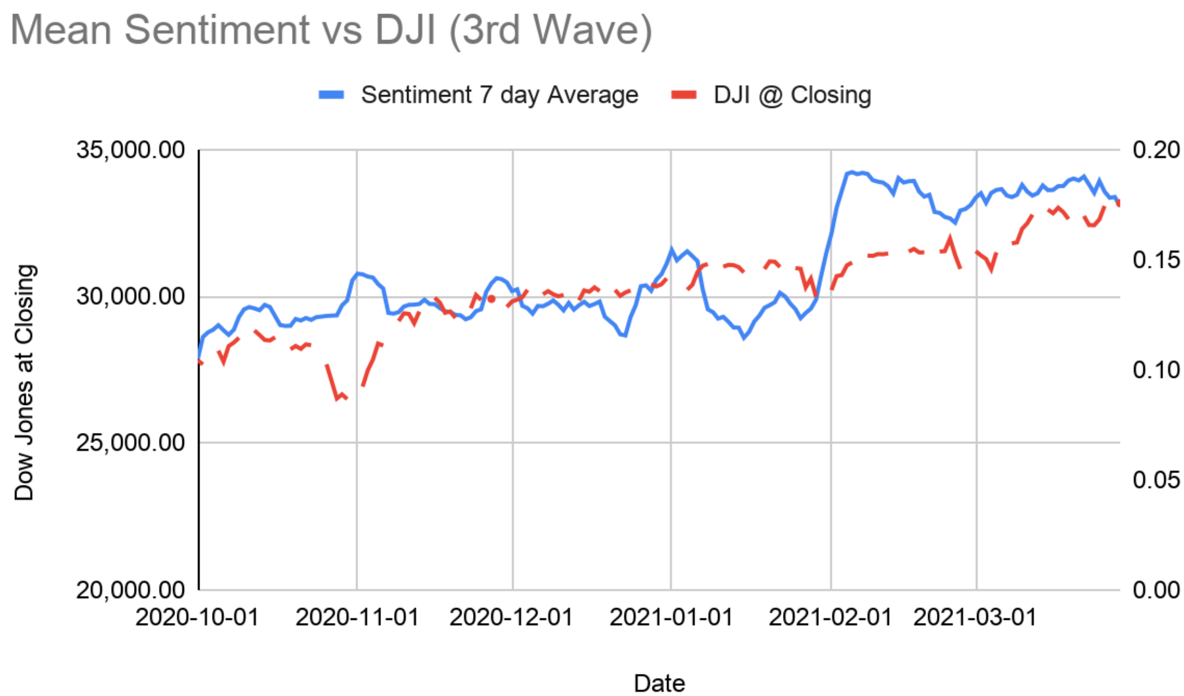
Graph of daily mean Twitter Sentiment compared to the Dow Jones Industrial Average at closing during the first and second waves



Note. Gaps in the DJI data are caused by stock market closures on weekends and holidays.

Figure 7

Graph of daily mean Twitter Sentiment compared to the Dow Jones Industrial Average at closing during the third waves



Note. Gaps in the DJI data are caused by stock market closures on weekends and holidays.

Figure 8

Table summarizing the different correlations between Twitter sentiments and the three statistics

	New Deaths	New Positive Cases	Dow Jones
Correlation during First Wave	-0.5409173359	-0.4709859212	0.07491805406
Correlation during Second Wave	0.1357208164	0.2132224018	-0.07172491027
Correlation during Third Wave	-0.6010065664	-0.3751200456	0.713600626

Discussion

Some results of this study were unsurprising and merely proved existing beliefs. For example, the correlations for virus cases and virus deaths were generally negative as more people being infected and dying arouses negative emotions. The opposite is true for the correlations for Dow Jones as people wish for economic activity to increase. None of the correlations found in this study were exceptionally high (>0.80), which is unsurprising as human emotion is a very complex issue that can be affected by an uncountable number of factors. Therefore, results that have a very high correlation between a statistic and public attitudes should be called into question since there has not yet been an example of a single statistic serving as an accurate predictor of human emotion. New deaths from COVID-19 also had a more negative correlation when compared to new positive cases. This result too is reasonable as deaths are perceived as far more tragic and detrimental to society than someone merely being infected.

The correlations of the Dow Jones Industrial Index proved to be fascinating as there was virtually no correlation between the DJIA and sentiments in COVID-19-related tweets for the first two waves. However, there was a moderately strong positive correlation between the DJIA sentiments in COVID-19-related tweets during the third wave. One possible reason is that the economic recovery has begun and the end of the pandemic is in sight with the rollout of new vaccines, which causes people to concern themselves less with the public health aspect of COVID-19 and more with its economic influences. However, causation cannot be proven in this study so these reasons are mostly speculation and hypotheses that could be tested in future research.

When compared to pre-existing literature, the results of this study support the conclusions drawn by researchers such as Lwin et al., Xue et al., Kruspe et al., and McCarthy as the results demonstrate a strong negative correlation between case counts or deaths and public sentiments early on in Spring 2020. Furthermore, the improvement in public mood hinted at the end of most of the study's research periods and in the poll conducted by McCarthy on behalf of Gallup seem to manifest itself in the form of higher average sentiment scores during the third wave despite the continuing negative correlation. The lack of strong correlations during the second wave could also be interpreted as indifference to the case counts and deaths caused by a lessening of concern for the virus.

At first glance, it is difficult to explain the anomaly in the correlation between public sentiments on Twitter and COVID-19 cases and deaths during the second wave. Since correlation between variables is unable to prove a causal link or reveal underlying factors and causes, one can only speculate on possible reasons behind this phenomenon. One such possible reason is the unique characteristic of the second wave in the US that it appears like a continuation of the first wave as much as it behaves as its own wave. This may cause differing perceptions of the second wave when compared to the first and third waves, which had more generic forms. The precise reason why the public appears to react to the second wave differently than the first and third wave, if they do indeed react differently, can only be determined through future research.

Limitations

This study operated on the assumption that the United States was a geographical region that was homogenized enough in public attitudes that grouping the data into such a large designation would still provide meaningful results. However, it is known that different regions of the US had very different pandemic situations during the same chronological periods (Tahirali, 2021). For example, New York was hit fairly hard by the virus early on but has since fallen past 25th place when it comes to active cases per capita in each state. Nuanced regional differences between states and even cities are not explored in this study as the resources and knowledge necessary to be able to organize more geographically specific information was not available.

As elaborated upon earlier in the methods section, the demographics of active Twitter users are not reflective of the general American population. Some users are also more likely to have their tweet selected for analysis as they may be more active on Twitter and they may add an excessive number of different COVID-19-related hashtags. Since there is no reliable way to retrieve the demographic information of all the authors of the Tweets used in this study, it is unfeasible to attempt to adjust the results of this study to better represent the broader populace.

Possible Areas for Future Research

Since the data gathering process involves filtering tweets for those with pandemic-related hashtags, it is impossible to conduct a similar procedure for sentiments on Twitter prior to the pandemic. Therefore, it is difficult to determine from these results whether public sentiments have recovered from the initial hit of negativity associated with the emergence of the novel coronavirus and subsequent lockdowns instated to curb its spread. Future studies could attempt to compare generalized sentiment data to conclude whether the effects of COVID-19 on public sentiments have disappeared.

The only economic indicator used in this research was the Dow Jones Industrial Index. However, there are other economic indicators such as the price of oil, US Dollar exchange rates, and other stock indicators such as NASDAQ. Future research utilizing a wider variety of economic indicators could help clarify whether the drastic increase in correlation between public sentiments and the DJIA is part of an overall trend. If an overall trend for various statistics is revealed to exist, it could reveal the evolution of COVID-19 from a primarily health issue into an economic one.

References

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights*, *10*, 117822261879286. https://doi.org/10.1177/1178222618792860

DePierro, J., Lowe, S., & Katz, C. (2020). Lessons Learned from 9/11: Mental Health Perspectives on the COVID-19 Pandemic. *Psychiatry Research*, 113024. https://doi.org/10.1016/j.psychres.2020.113024

Iglesias-Sánchez, P. P., Vaccaro Witt, G. F., Cabrera, F. E., & Jambrino-Maldonado, C. (2020). The Contagion of Sentiments during the COVID-19 Pandemic Crisis: The Case of Isolation in Spain. *International Journal of Environmental Research and Public Health*, *17*(16), 5918. https://doi.org/10.3390/ijerph17165918

Jackson, H., Mirza, N., Ryan, H., & Smith, J. (2020, October 12). *The Use of Social Media Analytics to Investigate Consumer Behavior and Build Marketing Strategies. Sentimental Analysis Used to Mitigate Risk Management during the COVID-19 Pandemic*. Papers.ssrn.com. https://ssrn.com/abstract=3710024

Kirzinger, A., Kearney, A., Hamel, L., & Brodie, M. (2020, April 2). *KFF Health Tracking Poll – Early April 2020: The Impact Of Coronavirus On Life In America*. The Henry J. Kaiser Family Foundation. https://www.kff.org/health-reform/report/kff-health-tracking-poll-early-april-2020/

Koeze, E., & Popper, N. (2020, April 7). The Virus Changed the Way We Internet. *The New York Times*. https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html

Kruspe, A., Häberle, M., Kuhn, I., & Zhu, X. X. (2020, July 1). *Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic*. ACLWeb; Association for Computational Linguistics. https://www.aclweb.org/anthology/2020.nlpcovid19-acl.14/

Lamsal, R. (2020). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*. https://doi.org/10.1007/s10489-020-02029-z

Loria, S. (2020). *TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation*. Textblob.readthedocs.io. https://textblob.readthedocs.io/en/dev/index.html

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. (2020). *Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: an observational study*. Psyarxiv.com. https://psyarxiv.com/xvwcy/

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter (Preprint). *JMIR Public Health and Surveillance*. https://doi.org/10.2196/19447

McCarthy, J. (2020, September 9). *Americans Less Negative in Their COVID-19 Outlook*. Gallup.com. https://news.gallup.com/poll/319883/americans-less-negative-covid-outlook.aspx

Salman Aslam Mughal. (2019, January 6). *Twitter by the Numbers (2019): Stats, Demographics & Fun Facts*. Omnicoreagency.com. https://www.omnicoreagency.com/twitter-statistics/

Tahirali, J. (2021, May 11). *Watch the timelapse showing the spread of COVID-19 cases in Canada and the U.S.* CTV News; Bell Media.

https://www.ctvnews.ca/health/coronavirus/watch-the-timelapse-showing-the-spread-of-covid-19-cases-in-canada-and-the-u-s-1.5423360

Tankovska, H. (2021, April 28). *U.S. Twitter user distribution by gender 2020*. Statista. https://www.statista.com/statistics/678794/united-states-twitter-gender-distribution/

The Atlantic. (n.d.). *The COVID Tracking Project*. The COVID Tracking Project. https://covidtracking.com

Wojcik, S., & Hughes, A. (2019, April 24). *Sizing Up Twitter Users*. Pew Research Center: Internet, Science & Tech; Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

World Health Organisation. (2021). *WHO COVID-19 Dashboard*. Covid19.Who.int; World Health Organization. https://covid19.who.int/

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*, *15*(9), e0239441. https://doi.org/10.1371/journal.pone.0239441

Yahoo! Finance. (2021, May 18). *Dow Jones Industrial Average (^DJI) Historical Data - Yahoo Finance*. Finance.yahoo.com. https://finance.yahoo.com/quote/%5EDJI/history/