STEM Fellowship Big Data Challenge 2021

What Are The Pandemic Events That Have Most Affected Public Sentiment Towards the Pfizer, Moderna and AstraZeneca COVID-19 Vaccines?

Muskaan Kaur Bajwa[1], Angelica Ramoutar[2], and Neha Purakan[2]

[1]McMaster University
[2]University of Toronto Mississauga

July 4, 2021

Abstract

Vaccination of the general public is a key step towards reducing the spread of COVID-19 and achieving herd immunity [1]. However, vaccine hesitancy is one of the greatest hindrances to mass vaccination campaigns. As such, being able to understand the factors that affect vaccine sentiment amongst the public is crucial to being able to manage the spread of COVID-19. Previous studies have highlighted the relationship between tweet sentiment and public sentiment, and have also found that stock prices are reflective of public opinion towards a particular company. Since Pfizer, Moderna and AstraZeneca are the primary vaccine producers, tweets relating to the vaccines produced by these companies as well as the stock prices of these companies are likely good indicators of public sentiment towards the COVID-19 vaccines. To identify and quantify the influence of pandemic events that have affected people's sentiments towards the Pfizer, Moderna, and AstraZeneca vaccines the most, Twitter sentiment analysis as well as a stock price analysis for these companies was performed. A data set of 4,093,081 tweets authored by 1,708,892 unique users which contained key phrases or hashtags related to the AstraZeneca, Pfizer and Moderna vaccines was compiled from February 24th, 2020, to May 28th, 2021, using the Twint Python tool. Sentiment analysis was conducted using the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool[2]. The stock price data was extracted from Yahoo Finance over the same period of time. Timelines of influencing events were constructed by identifying days for which a significant change in stock price or tweet sentiments occurred and then matching those days to significant events related to the pandemic. Our results suggest that news relating to vaccine clinical trials, vaccine distribution, the development of COVID-19 therapeutics, government public health guidelines, government funding of vaccine development and the distribution and eligibility guidelines for the vaccines and PPE distribution are factors that likely influence change in the public's positive, negative and neutral opinions on the vaccines. Furthermore, our analysis indicates that, to combat the spread of misinformation, public health units should devote more resources to ensuring that information about vaccine eligibility and vaccine distribution is accessible to the public.

Keywords
Twitter, VADER Sentiment Analysis, Stock Price Analysis, COVID-19, COVID-19 Vaccines, Pfizer Inc., AstraZeneca Plc., Moderna Inc.

1 Introduction

On January 7th 2020, the novel coronavirus SARS-CoV2 was identified as causing the pneumonia-like illness that is now known as COVID-19. [3] Since then, at the time of writing, there have been 182,582,291 confirmed cases worldwide and 3,954,621 confirmed deaths due to COVID-19, with cases and deaths reported from more than 180 countries all over the world. [4].

Since the start of the pandemic, there has been an unprecedented multilateral effort to develop COVID-19 vaccines aimed at reducing the transmission of SARS-COV2 and reducing the severity of illness due to infection. [5] Most notably, vaccine trials for the Pfizer, Moderna and AstraZeneca vaccines have been able to demonstrate some of the highest degrees of efficacy in preventing infection from COVID-19 amongst all COVID-19 vaccines available today. [6] [7] [8]

Despite this success in developing effective vaccines, the progress of mass vaccination campaigns, which are necessary to achieve herd immunity, is dependent upon vaccine acceptance amongst the public. Thus, being able to understand the factors that affect vaccine sentiment amongst the public is crucial to being able to manage the spread of COVID-19. [9]

Numerous previous studies have sought to characterize vaccine sentiments through the analysis of tweets, suggesting that tweet sentiment is a good metric for public sentiment. Yousefinaghani et. al [10] conducted a study in which tweet sentiments were used to analyze and make conclusions about changes in COVID-19 vaccine sentiment over time amongst the general public. The study was also able to suggest reasons (i.e. events) that might have influenced these changes. A limitation of this paper was that it did not offer any quantitative information about these factors nor did it characterize them. In other words, there were no common themes highlighted amongst the influencing events. Another study conducted by D'Andrea et. al [11] also correlated changes in the sentiment of Italian tweets to the occurrence of specific influential pandemic events in Italy, and acknowledged shared attributes among the events, citing that most events were related to parliament announcements and vaccination mandates. However, like the study conducted by Yousefinaghani et. al, D'Andrea et. al did not offer quantitative results as to the relative influence of those event categories. Additionally, previous studies have examined how pandemic events have affected stock prices. One such study [12] conducted a sentiment analysis of tweets related to stock prices that were created before and after the government-mandated lockdowns. The analysis showed that both stock prices and sentiment were negatively impacted by the restrictions due to the pandemic. Thus, stock prices could be reflective of the general sentiment towards the pandemic. More specifically, a previous study conducted by Rui et al. [13] demonstrated that the emotional index of comments left on stock forums on the Shanghai Composite Index was strongly correlated with trends in stock price change. The authors of this study state that emotional metrics related to stock price trends may be used as a tool for the prediction of stock price trends, suggesting that stock prices are a reflection of public opinion.

The aim of our study is to fill in the aforementioned gaps in research relating to the identification of factors influencing vaccine sentiment. That is, this study seeks to identify which categories of pandemic events have had the most influence on public sentiment towards the AstraZeneca, Moderna and Pfizer vaccines, and quantify their relative influences. Changes in the sentiments of tweets relating to the vaccines and changes in stock prices of the aforementioned companies will be taken as proxy measurements for public sentiments towards the vaccines. The results of our study will also help identify potential sources of misinformation (ie. the reasons why some events have triggered a significant change in the sentiments towards the vaccines) so that they may be addressed going forward as the fight against the spread of COVID-19 continues.

Twitter is known as a micro-blogging platform that allows users to express their opinion, engage with other users and discuss current events. The Twitter platform emphasizes the use of short sentences while still allowing users to share videos, photos, and hyperlinks. It is also considered to be one of the top ten most user-friendly social media platforms. Tweets are representative of a diverse population of users that express different opinions since Twitter connects individuals from around the world and allows users to share different types of information ranging from educational content to government-related topics. [14] As of April 2021, the majority of Twitter users are in the 13 to 50 years age group. However, some users are over the age of 50. [15] Due to the breadth of this range, sentiment analysis of Twitter content is likely to be at least somewhat reflective of the sentiments of the general public. Twitter is valuable in the analysis of public sentiments because of the abundance of quantifiable information (in the format of text) it provides as compared to other more picture-based platforms such as Instagram or Facebook. Furthermore, extracting data from Twitter is convenient for analysis, since there are many open-source software available for tweet scraping.

Analyzing the stock price change for AstraZeneca, Moderna, and Pfizer is important because, as stated earlier, stock price is an indication of public opinion towards the companies. Since the most relevant products from these companies right now are their vaccines, changes in the stock price of these companies are likely to be a reflection of public sentiment towards the vaccines. The way the media portrays news events and the way news articles are titled has been found to have an effect on public opinions and thus can affect stock prices. [12] As such, it is appropriate to attempt to associate news events that are described by news articles as influencing factors on the stock price of the aforementioned companies. Additionally, considering stock price as an indicator of public sentiment will increase the 'sample size' of the people whose sentiments we are studying. This is because the stock price of these companies is probably dependent on the opinions

of more people than the number of people we are studying through our Twitter analysis. With approximately 14% of US households being directly invested in the stock market (about 17 million households) [16], there are likely more people who invest in these stocks than there are in our tweet dataset (1,708,892), meaning that considering stocks as an indicator of sentiment will likely result in more unique opinions being considered in our study.

2 Materials & Methods

2.1 Collecting Data

Selecting a Data Range: February 24th 2020 was selected as the 'start date' for tweet collection because it is the day before the United States Center for Disease Control (CDC) issued a press release identifying the impending pandemic status of COVID-19. This was the first time that a large regulatory organization recognized the threat of the virus. [17] The 'end date' was the last possible date for which data collection could be done given the time constraints of this competition. Stock data was analyzed for the same time period.

Moderna, Pfizer and AstraZeneca Company Stock Data Collection: The stock data for Moderna Inc., Pfizer Inc. and AstraZenca Plc. were collected using Yahoo Finance. Only the daily closing prices were analyzed over the aforementioned time period. This data is summarized in Figure 2

Tweet Scraping: A total of 4,093,317 tweets were scraped from Twitter with the Twint Python package [18]. English tweets containing a combination of keywords 'AstraZeneca', 'Pfizer' or 'Moderna', or the hashtags '#astrazeneca', '#pfizer', '#moderna', '#astrazenecavaccine', '#pfizervaccine', or '#modernavaccine' were collected from February 24, 2020, to May 28th, 2021. There are 1,708,892 unique users that authored tweets in the data set.

2.2 Data Cleaning and Wrangling

Data manipulation was performed using Excel, Google Sheets and Python (Jupyter Notebook).

Stock Data: Daily stock price changes were calculated for each day (except for the first day - February 24th, 2020) by finding the difference between the closing price for each day and the closing price from the day before. For example, the closing prices for Moderna Inc. on February 25th, 2020 and February 26th, 2020 were \$23.76 and \$29.16 respectively. Thus, the stock price change for February 26th, 2020 was calculated as \$29.16 - \$23.76 = +\$5.40.
. Furthermore, the data was then separated by trend (increase/decrease): a separate data set containing dates and associated stock metrics (i.e. daily stock price, daily stock price change) was created for each of the following *Trend and Metric* categories:

- Days for which an increase (relative to the previous day) in the AstraZeneca Plc. stock price was observed
- Days for which a decrease (relative to the previous day) in the AstraZeneca Plc. stock price was observed
- Days for which an increase (relative to the previous day) in the Moderna Inc. stock price was observed
- Days for which a decrease (relative to the previous day) in the Moderna Inc. stock price was observed
- Days for which an increase (relative to the previous day) in the Pfizer Inc. stock price was observed
- Days for which a decrease (relative to the previous day) in the AstraZeneca Plc. stock price was observed

The segregation of the data was done to determine if certain factors (i.e. news events) had a greater influence on the rises of the stock prices versus their declines.

Sentiment Data: Tweets were cleaned before VADER analysis. User tags (for example, '@pfizer'), hyperlinks, and emojis were removed from tweet text. 236 tweets that were missing date attributes were removed from the data set. VADER sentiment analysis was conducted for all 4,093,081 remaining tweets. VADER was selected for sentiment analysis because it is a tool that was specifically trained on social media data [2]. Tweets with a compound score of less than zero, equal to zero, and more than zero were categorized as having negative, neutral, and positive sentiment respectively. For each day from February 24th, 2020 to May 28th, 2021, the count and proportion of total tweets associated with positive, negative and neutral sentiments were both calculated. The daily count and proportion of positive tweets for February 25th, 2020, for example, was found by counting the number of tweets (610) and then dividing the count by the total number of tweets collected for that day (610/1440 = 42%). These data are summarized in Figure 4 and Figure 5. Furthermore, the daily *change* in the proportion of total tweets for each sentiment was also

tabulated for each day by taking the difference between the percent of tweets associated with a particular sentiment for a given day and the percent of total tweets with that same sentiment for the previous day. For example, the change percent of positive tweets for February 26th was 42% - 53% = -11%. This process was repeated for every sentiment. Next, the data was organized by the trend: a separate data set containing dates and associated daily sentiment metrics (i.e. daily count of tweets of each sentiment, the daily proportion of total tweets associated with each sentiment, change (relative to previous day) in the daily proportion of total tweets associated for each sentiment) was created for each of the following *Trend and Metric* categories:

- Days for which an increase (relative to the previous day) in the proportion of positive tweets was observed
- Days for which a decrease (relative to the previous day) in proportion of positive tweets was observed
- Days for which an increase (relative to the previous day) in the proportion of negative tweets was observed
- Days for which a decrease (relative to the previous day) in the proportion of negative tweets was observed
- Days for which an increase (relative to the previous day) in the proportion of neutral tweets was observed
- Days for which a decrease (relative to the previous day) in the proportion of neutral tweets was observed

Data segregation was done so that the influence of factors (i.e. pandemic news events) affecting sentiments may be characterized with greater granularity. We may specifically isolate news events that were important in mitigating the spread of misinformation (i.e. factors influencing the decrease in negative sentiment and increase in positive sentiment), which may be distinct from news events that were the most polarizing (i.e. news events that coincided with decreases in neutral sentiment).

2.3 Statistical Analysis

Identifying Influential Dates (Outlier Identification): Dates of significant change in sentiment and stock price were identified as the outlier data points from each of the *Trend and Metric* categories identified above. Boxplots were created for each category using Plotly (an open-source graphing library) [19] and are shown in Figure 3 and Figure 6. Descriptive statistics from these plots (Figure 11 and Figure 12) were used to identify outlier dates in each category. Boxplots were deemed as being the appropriate tool for looking at the spread of the data and identifying outliers in each category because the interquartile range (IQR) from the boxplots is not affected by outliers whereas a normal range calculation would get affected by outliers present in the data. Furthermore, the boxplot method was deemed as an appropriate analysis tool because the sample size (shown as the variable *count* in Figure 11 and Figure 12) of dates in each of the *Trend and Metric* categories was greater than twenty, which is the minimum number of sample points necessary for using a boxplot to properly represent data [20]. The exclusive quartile method was used for all box plots as it applies to data sets containing either an even or odd number of data points. This method is also more suitable for large samples sizes such as the ones used in this paper. All influential dates, once identified, were aggregated into one data set.

2.4 Identifying Pandemic Events Influential to Vaccine Company Stock Price and Vaccine Sentiments

Identifying news events coinciding with influential dates: For each influential date, a separate MonkeyLearn [21] word cloud (example: Figure 13) with the top 50 'relevant' words was generated from all tweet text produced on that day. Each word cloud was analyzed for indicators of what the dominant news event was for that day. For each day, a selection of the top relevant words were entered together as a query into Google with the search results restricted in time to the date being analyzed. The news result that encompassed the most amount of the word cloud terms was then determined to be the dominant news item for that day.

Categorizing News Events Coinciding with Influential Dates: Common themes amongst all influential news events were identified and news event categories were created. All influential news events were classified into one of the following categories:

- Alternative Drug News Events are events relating to the production or distribution of COVID-19 Therapeutics (Remdesivir, monoclonal antibody treatments, anti-viral compounds, etc.)
- Vaccine Clinical Trial/Efficacy News Events are events relating to the release of data from clinical trials in which efficacy of

COVID-19 vaccines was tested experimentally

- Economic News Events News are events relating to government funding towards vaccine development
- Government Guideline News Events are events relating to quarantine, lockdown, mask mandates, social distancing guidelines, government-imposed vaccine dosage interval guidelines.
- Side Effects/Vaccine Reaction News Events are events relating to side effects or reactions documented in clinical trials and general population after vaccine administration
- Vaccine Company News Events are events relating to business dealings and company updates unrelated to vaccine clinical trials
- Vaccine Distribution/Eligibility and PPE Distribution Events are evens relating to the procurement, distribution, and accessibility of vaccines and PPE for the general public.

There were some shared significant dates between each trend and metric category. However, it was ensured that each unique date from February 24th 2020 to May 28th 2021 was only associated with one main news category event determined by the word cloud of tweet content from that day. All significant events and their associated news items are summarized in Table 1. Although news items chosen for the days in which significant changes in stock prices were observed were chosen based on word cloud analysis of tweet data, previous studies have demonstrated that news items have significant impacts on stock prices [12], so associating the news items with these significant days of change for the stocks is still relevant.

Determining Distribution of News Events For Each Trend and Metric Category: Significant dates of change for each stock and sentiment type along with their associated news events were plotted, as shown in Figure 7. For each trend and metric category, the number of events corresponding to each of the news event categories described above was tabulated. The results from this step are summarized in Figure 8. Furthermore, the relative percent proportion of significant dates that each news category was associated with was calculated. This is summarized in Figure 9.

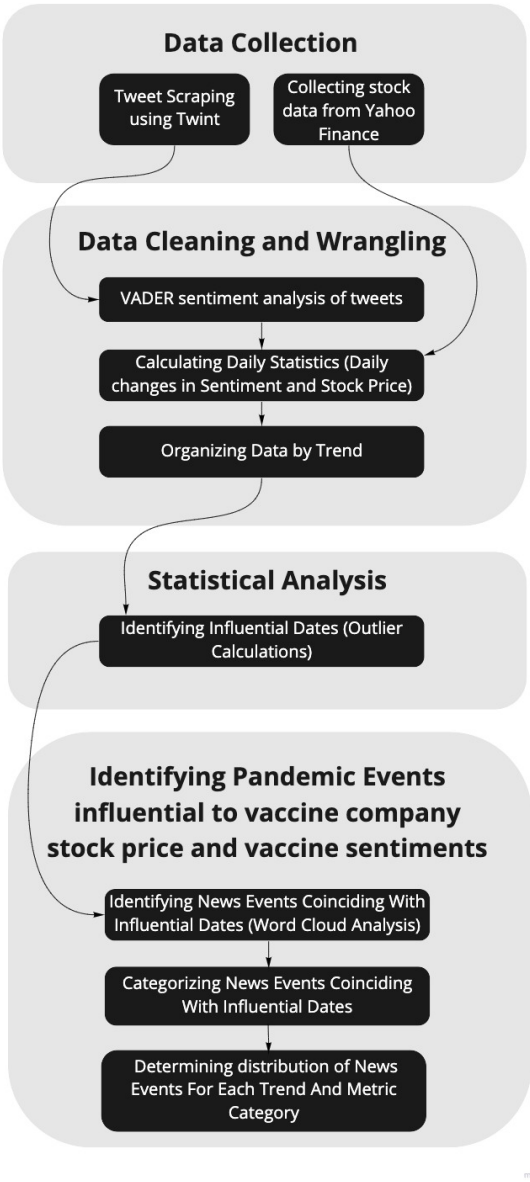

Figure 1: Summary of Project Workflow

3 Results

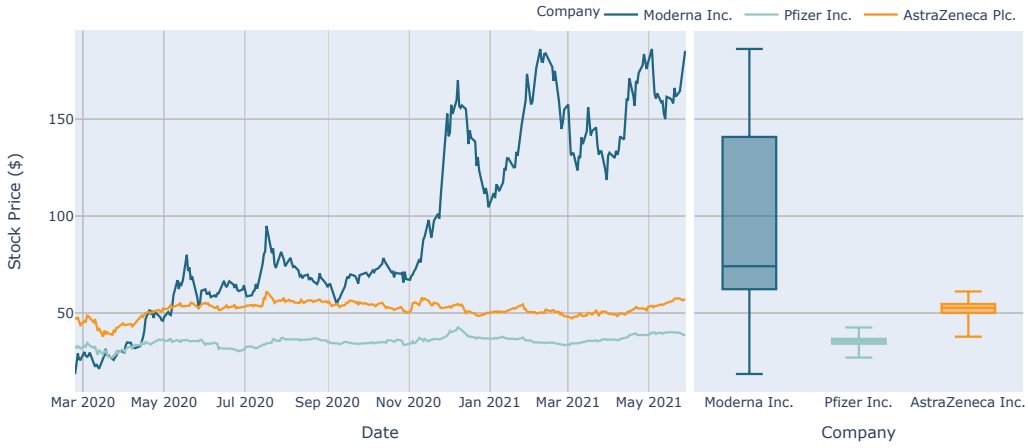

Figure 2: Daily stock price of Moderna (NASDAQ: MRNA), AstraZeneca (LON: AZN) and Pfizer (NYSE: PFE). Data range = February 24, 2020 to May 28 2021

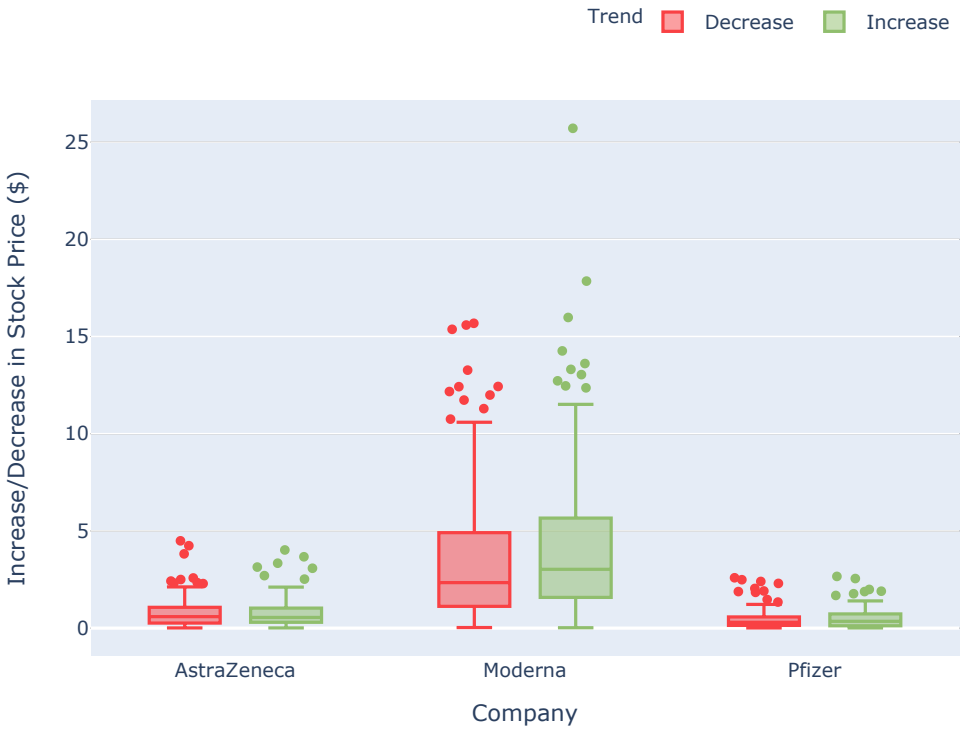

Figure 3: (above) Distribution of changes in daily stock price. Change in daily stock price for all days is defined as the difference between the closing price of stock on a particular day and the closing price of the previous day. The outlier points shown are the 'significant days of change' in stock price for each company

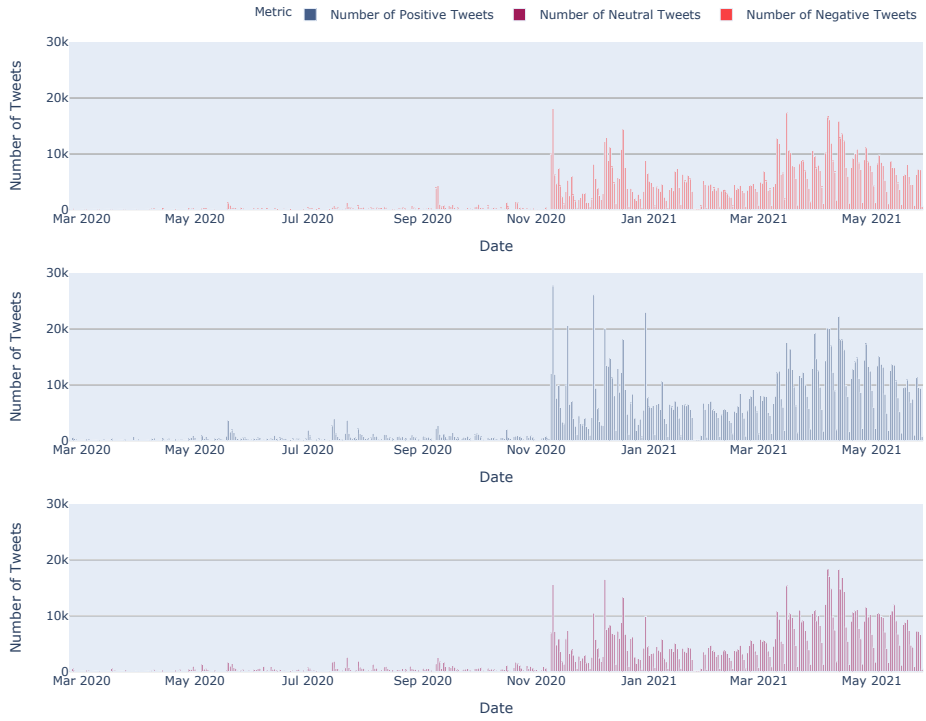

Figure 4: (above) Daily count of all tweets belonging to a particular sentiment. Data range = [February 24 2020 to May 28 2021] with the exception of the following dates: [January 24 2021 to January 28 2021] If data is not showing in pdf viewer, please zoom in.

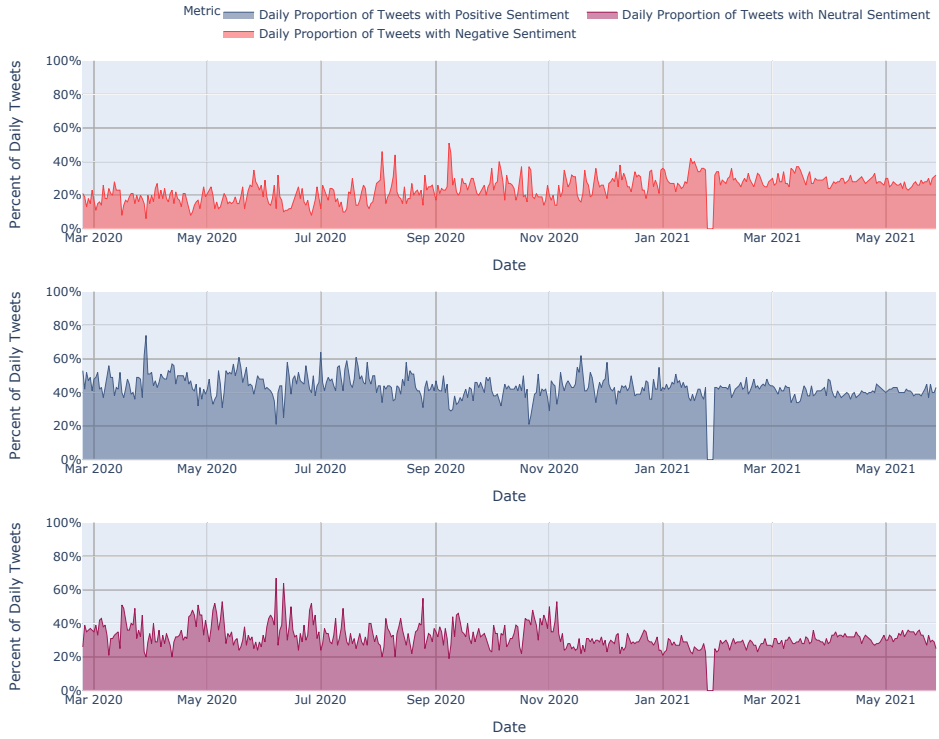

Figure 5: (above) Daily proportions of all tweets belonging to a particular sentiment. Data range = [February 24 2020 to May 28 2021] with the exception of the following dates: [January 24 2021 to January 28 2021].

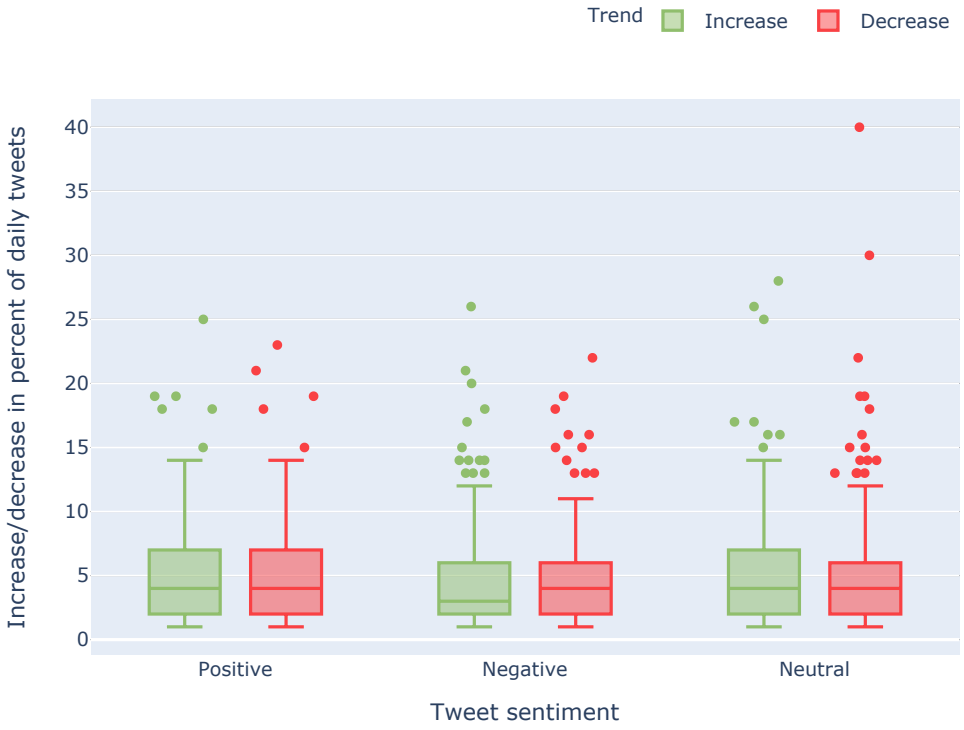

Figure 6: Distribution of daily changes in the proportion of tweets of each sentiment. Daily changes in sentiment proportion for all days is defined as the difference between the percent proportion of all tweets belonging to a particular sentiment on a particular day and that of the previous day. The outliers correspond to days of significant change in the proportion of tweets belonging to each sentiment.

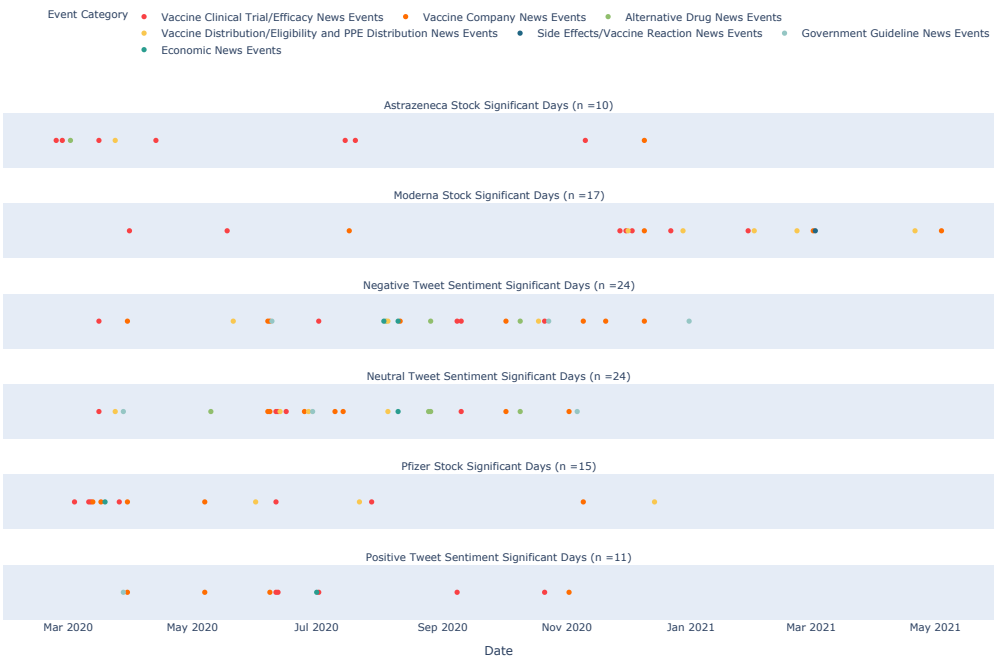

Figure 7: Distribution of days of significant change in stock price and proportion of sentiments over time. Each day is also associated with the dominant news item coinciding with that day. Each point is an outlier point from either Figure 3 or Figure 6. Aggregated event timeline is shown in Table 1

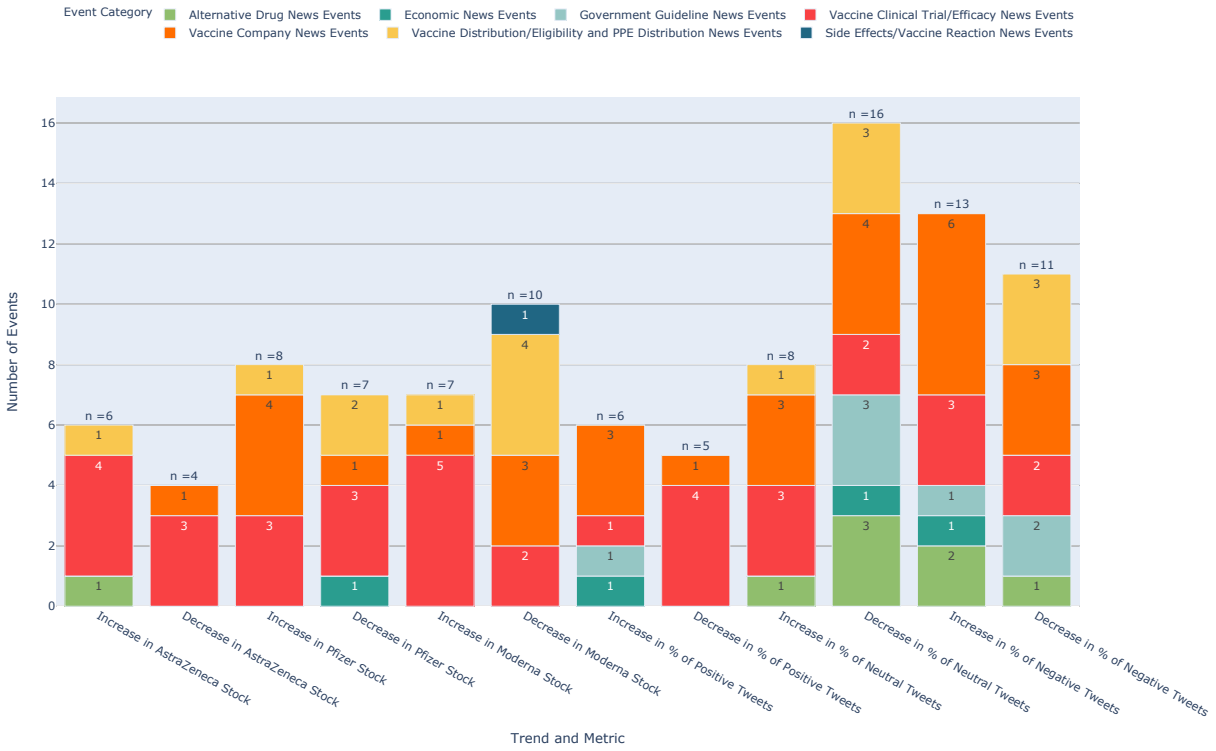

Figure 8: Distribution of news items coinciding with significant days of change for all metrics analyzed.

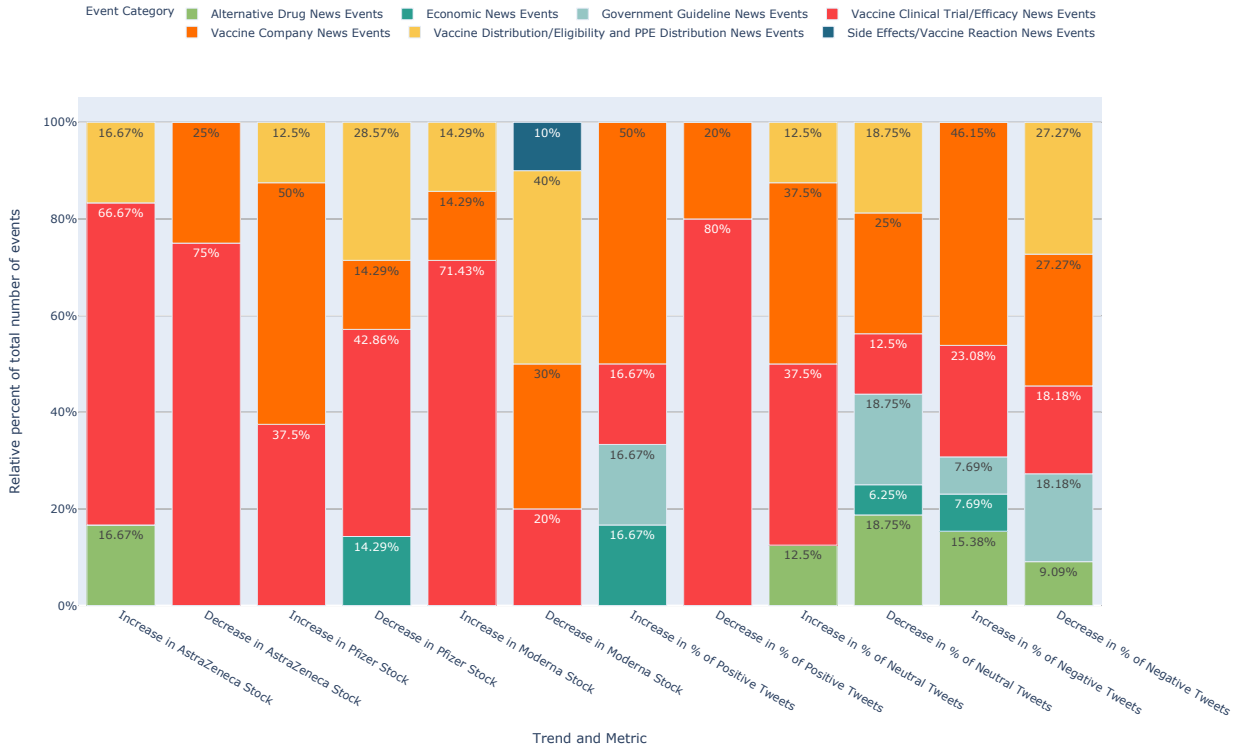

Figure 9: Relative proportion of news items for each metric analyzed.

4 Discussion

Figure 2 shows the change in the stock price of Pfizer Inc., Moderna Inc., and AstraZeneca Plc. over time. The stock price for Pfizer and AstraZeneca stays relatively stable for the entire timeline. However, the stock price for Moderna starts with having minor fluctuations between the March 2020 to October 2020 time period. After October 2020, the Moderna stock price fluctuates considerably and this can be seen through the large dips and increases seen in the figure. This period of volatility coincides with fluctuations seen in the number of positive and negative tweets after October 2020 in Figure 4, suggesting that the developments of the Moderna vaccine are correlated to the public perception of COVID-19 vaccines in general. The stock price of Moderna increased over time, suggesting that the company's efforts to develop its COVID-19 vaccine could have influenced the demand for the company's stocks. The box plot in the right panel of Figure 2 shows the distribution of daily stock prices for Pfizer Inc., Moderna Inc., and AstraZeneca Plc.. Moderna has the highest spread of stock price data, followed by AstraZeneca which is then followed by the Pfizer stock. The vaccine development efforts have a more pronounced effect on Moderna's stock price because it is a relatively smaller company than Pfizer and AstraZeneca, and it has never introduced a commercial drug into the market.[22] Pfizer on the other hand, does not heavily depend on its vaccine program as it generates revenue from a wide variety of well-established products, providing overall stability to its stock. [23] [24]

Figure 3 shows the distribution of the daily *changes* in the stock price for each company. The change is calculated for every single day by subtracting the closing stock price from the previous day from the closing stock price of the present day. The outliers in this figure represent outlier days for each company. These outlier days are the days in which a significant change in the stock price was observed (statistical outlier) as a result of a significant event occurring. The 'decrease' trend represents negative changes where the stock price decreased relative to the previous day. The 'increase' trend represents positive changes where the stock price increased relative to the previous day. The 'decrease' and 'increase' trends have their own boxplots as this allows for the identification of factors that can potentially influence the confidence people have in the stocks- and by extension- the companies and their products (the vaccines). In this figure, the changes in the Moderna stock price- both for the 'increase trend' and the 'decrease' trend- have greater variation when compared to Pfizer and AstraZeneca's spread of stock price changes. This is further confirmed by comparing the standard deviation of stock price changes between the three companies. Moderna has the highest standard deviation values- $3.62 for the decrease trend and $3.94 for the increase trend in Figure 10 - indicating greater stock price change variability. The category with the least number of significant days of change (least number of outliers) is the 'decrease' in the AstraZeneca stock, suggesting that the AstraZeneca stock is least susceptible to negative change. One possible reasoning for this could be that AstraZeneca most likely will not make a lot of profits from their Covid-19 vaccine because of their pledge to not gain any profit from vaccine sales during the pandemic. The production of the AstraZeneca vaccine also does not employ any novel technologies, unlike the Pfizer and Moderna messenger RNA technology. AstraZeneca is well known for its other well-established products related to oncology and the company's oncology portfolio led to a 24 percent increase in revenue in 2020. Therefore, investors are more likely to get swayed by news relating to AstraZeneca's well-established, blockbuster products like Tagrisso (used to treat non-small-cell lung carcinomas) and not as much by news developments related to their Covid-19 vaccine. [25]

Figure 4 shows the number of positive, negative, and neutral tweets over time. For all sentiment types, the number of daily tweets between the April 2020 to October 2020 time period stays below 10,000. Within this time period, there also aren't any major fluctuations in the number of positive, negative and neutral tweets. In contrast, the time period after October 2020 shows large fluctuations in the number of positive, negative and neutral tweets. The positive tweets show some of the biggest spikes in tweet numbers. These large fluctuations in the number of positive, negative and neutral tweets could be attributed to the release of phase three clinical trial results of all three vaccine candidates. The phase three clinical trial results for Pfizer Inc., Moderna Inc., and AstraZeneca Plc. were released on November 18, 2020, November 16, 2020, and December 8, 2020 respectively. [26][27][28] There is a visible gap in the data around January 2021- this is because of errors in scraping tweets for the dates of January 24th 2021 to January 28th 2021, resulting in no tweet data being collected for this time period. This could affect the identification of outlier significant dates but since only four days are missing out of 459 days, this gap is minuscule and should not affect the general trends of the graph.

Figure 5 depicts the proportion of tweets associated with each sentiment over the period of February 24th 2020 through May 28th 2021, with the exclusion of January 24th 2021 to January 28th, 2021 as mentioned previously. The

percent of daily tweets with a negative sentiment steadily increases over time. Additionally, significant spikes in the negative sentiment occurred around September 2020. These spikes correspond to the time period in which news events relating to the vaccine clinical trial news were observed. However, since the efficacy of the vaccines was reportedly high (which is news that we would expect people would view positively) there could be other factors that could have contributed to this spike in negative sentiment. For instance, this spike could be attributed to hesitation about the vaccine due to a perceived rush in the testing and trials, which is a sentiment would be communicated more around the time of the release of clinical trial data. Additionally, pre-clinical studies from AstraZeneca published in May 2020 were criticized as being 'overly optimistic'- thus, news about clinical trials from other companies that came out during this time in September 2020 could have also been perceived similarly and negatively.[29][30] The daily proportion of positive tweets varies over time, ranging from as low as 20% to as high as 70%. Specifically, the daily proportion of tweets with positive sentiments decreased overtime- at the beginning of 2020, the percent of positive tweets was almost always above 40%, whereas in 2021, the percent of positive tweets was just about 40%. This may be because people are not responding to events positively and not necessarily because no positive events are occurring. Compounded negative effects (prolonged lockdowns, lifestyle changes, economic hardships, job loss, lack of social interaction) due to the pandemic is could be one of the reasons for this trend. How the media has reported on the previously mentioned negative effects could have contributed to the increase in negative sentiment. For the daily proportion of tweets with the neutral sentiment over time, there is a steady decrease. There is a significant spike in the neutral sentiment around June 2020, which correlates with the identified news event that the National Institutes for Health (NIH) claimed joint ownership of Moderna's coronavirus vaccine. The NIH is the central agency of the United States government which is responsible for biomedical and public health research and is currently an essential part of the United States Department of Health and Human Services. Government endorsement of the Moderna vaccine may have caused an increase in neutrality (which implies a decrease in the amount of opinionated tweets) towards the COVID-19 vaccines in general because the approval of an authoritative body may have shifted the conversation about the vaccines to a more factual basis. Furthermore, this spike in the neutral sentiment could be due to the fact that, on this day, more people- rather than authoring their own original tweet with no opinion to it- were simply sharing the news headline, which itself would be of neutral sentiment. This makes sense because government endorsement of the vaccine is a significant event that would likely be highly discussed.

The variability in the daily proportion of all three sentiments is high in the 2020 months and low in the 2021 months. This could be due to the fact that, as time goes on and more data is introduced to the public, and they will be able to form a final judgement about the vaccines, and not be so hesitant in their opinions.

Furthermore, there are visually significant spikes in all three sentiment graphs that do not correspond to significant days identified by outlier calculations, which indicates that there may be significant, long-term, or repeating influencing factors (i.e. pandemic events) that are not being considered in our analysis due to our threshold of what we consider as a 'significant event' being high.

Figure 6 is showing the distribution of daily tweet sentiment changes separated by trend- 'increase' or 'decrease'. For the positive tweet sentiment 'increase', it can be seen that the outliers are spread apart, and the highest outlier corresponds to a 25% increase. For the positive sentiment 'decrease' trend, the outliers are again spaced out and they do not cover a wide range of the plot as they reach up to only about 22%. For negative tweet sentiment increase, the outliers are clustered at the beginning and range from below 15% to 26% and begin to spread out. For negative sentiment decrease, the outliers range from below 15% to 23%. For neutral increase, the outliers begin at 15% and have a longer range to 27%. As well, for neutral decrease, the outliers start from below 15% and show an increase in range to 40%. The distribution of changes between sentiments is similar. This is indicated by the similarities in the appearance of the boxes themselves.

Figure 7 shows the distribution of days of significant change for stocks and/or sentiments with their associated news event categories over a time period from February 24th 2020 through May 28th 2021. The vaccine distribution/eligibility and PPE distribution news category, vaccine company news events category and vaccine clinical trial/efficacy news category all contain influential dates that occurred in both 2020 and 2021. However, the vaccine distribution/eligibility and PPE distribution news category had more number of influential dates occurring in 2021 (3 dates) than the other two categories (vaccine clinical trial/efficacy news category having one event in 2021 and vaccine company news events having two dates in 2021), suggesting that this category has the most influence on public sentiment in 2021. The dates for 2020 in this category are associated with the procure-

ment of potential vaccines and PPE, and the making of deals between governments and vaccine companies. In contrast, the dates in 2021 have more to do with the actual shipment of vaccines. The one and only event in the side effects/vaccine reaction news events category also occurred in 2021. Results from an online survey conducted by Verywell Health [31] suggest that vaccine reaction concerns amongst the public are outweighed by fear of COVID-19 infection, which supports our conclusions from the analysis in this figure that news about side effects or reactions from the vaccine do not have a lot of influence on public sentiment towards the vaccines. Significant days coinciding with news about the development of alternative coronavirus drugs, the imposition of government public health guidelines, and government funding of vaccine development (economic news events) are exclusively concentrated in 2020. This suggests that the influence of these types of events are limited in 2021. Days of significant change for Moderna stock are concentrated in 2021, with not many significant dates for the stock occurring in 2020. Although phase 3 clinical trial results for the Moderna vaccine were announced in November of 2020 [8], the concentration of significant dates for this stock in 2021 could be due to the fact that the majority of the actual procurement and distribution of the Moderna vaccine occurred for most countries at the end of 2020 and throughout 2021. [32] Since Moderna profits from the sale of its vaccines [33], it makes sense that significant changes in the stock price would ensue from this activity in 2021. A potential reason why this trend is not observed for AstraZeneca and Pfizer is that, as discussed previously, the stock prices of these companies are less dependent on the sale of their COVID-19 vaccines. For all other sentiment metric categories, significant days of change are concentrated in 2020, with little to no significant days occurring in 2021. This agrees with the trend observed in Figure 5, in which the variability of the daily proportions of each sentiment was the highest in 2020 and the lowest in 2021. The metrics with the most number of significant days of change are the negative tweet sentiment and the neutral tweet sentiment. This suggests that the neutral and negative sentiments were subject the most to change. Considering that previous studies have found an association between negative sentiment of tweets relating to COVID-19 vaccines and vaccine misinformation [10], this observation suggests that the spread of misinformation is manageable, since the negative sentiment has a demonstrated potential for significant change.

Figure 8 shows the numerical distribution of events associated with significant days of change for each trend and metric category. For example, out of the four significant days of change for the decrease in AstraZeneca stock, three of those days coincide with a news item related to (as determined by the daily word clouds) vaccine clinical trials, and one of those days is associated with a news item related to the business dealings of the vaccine companies. In this figure, the trend and metric category with the most number of significant days of change (most number of outliers) is the decrease in neutral tweets. This means that the neutral sentiments towards the vaccine were changing the most. The trend and metric category with the most number of event categories (that is, the most number of influencing factors) is the decrease in the percent of neutral tweets. This suggests that a variety of factors contribute to people making up their minds about the vaccine.

Figure 9 is very similar to Figure 8 except for the fact that the area taken up by each colour bar represents the proportion of that event category measured as a percentage. Decimals were rounded to two digits to show equivalence between certain categories as rounding to the nearest whole percent would have skewed the data in favour of one or more categories. If it is assumed that the recurrence of a particular category is associated with its influence, then this means that the area taken up by each colour is representative of the proportion of its influence. For example, the trend and metric column 'Decrease in percentage of negative tweets' shows that 27.27% of the significant days of change for that column coincide with a day where the dominant news event was related to vaccine distribution and eligibility. Because this news category represents a large proportion of the column, we can postulate that vaccine distribution and eligibility news are correlated to decreases in the percentage of negative tweets. This indicates that this kind of news has the capability of decreasing the percent of negative tweets. Previous research has highlighted the importance of identifying negative sentiments on social media to reduce the spread of misinformation.[10] [34]. This correlation was demonstrated in Yousefinaghani et al. through a word cloud analysis conducted for positive and negative tweets relating to COVID-19 vaccines.[10] In this study, the word cloud analysis for negative tweets had a lot of phrases and keywords related to misinformation. Some examples of such keywords include 'conspiracy', 'scammed', 'fool', 'blinding', and 'misrepresents'.[10] Due to this correlation between negative sentiment tweets and fake news, previous research has observed increased vaccine hesitancy amongst the general population and this makes negative sentiment tweets an effective target for reducing the impact of the spread of misinformation. [10] Therefore, the types of information conveyed in the news

categories shown in the 'decrease in percent negative tweets' column- that is, information about vaccine distribution and eligibility, news about the business dealings of vaccine companies, news about data released from vaccine clinical trials, news about government public health guidelines, and news about the development of alternative coronavirus drugs are effective in combating the spread of misinformation. This column is also one of the most diverse in its distribution of news event categories suggesting that there is a lot of different information and knowledge that can help combat the spread of misinformation. The news categories present in the column 'Increase in Moderna Stock' could also be indicative of the types of information that could help combat the spread of misinformation. As discussed previously, changes in the price of this stock closely align with changes in the number of tweets of each sentiment over time, suggesting that Moderna stock is the most reflective of public opinion towards the vaccines out of all other stocks. The categories of information in this column are all shared with categories of information present in the 'decrease in negative sentiment' column, reinforcing that information about vaccine distribution and eligibility, vaccine clinical trials and vaccine company business dealings are all effective in increasing public confidence in the vaccines and decreasing vaccine misinformation. Another column that could yield insight into managing the spread of vaccine misinformation is the 'increase in % positive tweets'. Three of the news event categories present in this column are shared with the 'decrease in % negative tweets' column- news about vaccine company business dealings, vaccine clinical trials, and government public health guidelines. Two of the news event categories present in this column are shared with the 'increase in Moderna stock' column- news about vaccine company business dealings and vaccine clinical trials. This observation suggests the importance of news about vaccine company business dealings and vaccine clinical trials for boosting public confidence about the vaccines and limiting the spread of misinformation. Although the news related to vaccine clinical trials and vaccine company business dealings appear to have more influence (i.e. collectively comprise the majority of the area in each column), as discussed previously, the influence of these news events is reduced in comparison to the influence of news about vaccine distribution and eligibility in 2021. The distribution of events in the categories 'increase in percent of positive tweets' and 'decrease in percent of negative tweets' is at odds with the findings of a previous study investigating public sentiments towards COVID-19 vaccines amongst Italians [11]. In this study, it was found that increases in the proportion of positive tweets and the decrease in the proportion of negative tweets were associated with news items relating to government-mandated vaccination and an increase in coronavirus transmission. While it is clear that news relating to government guidelines is a contributing factor to changes in the 'increase in percent of positive tweets' and 'decrease in percent of negative tweets' (which agrees with the previous findings), we do not see any influence from news relating to the transmission of coronavirus. This discrepancy is likely due to the fact that vaccine sentiment is contingent on the local response towards the pandemic, which varies from nation to nation. This explanation also agrees with the findings of Yousefinaghani et. al [10]. Furthermore, factors that impact vaccine sentiment are likely different between English-speaking and non-English-speaking countries such as Italy. Consideration of vaccine sentiment amongst people from non-English-speaking countries has not been considered in our study because we have only analyzed English tweets. The news event categories present in the 'decrease in % neutral tweets' column- that is, news about vaccine company business dealings, vaccine clinical trials, government public health guidelines, government funding of vaccine development (economic news), and the development of alternative coronavirus therapeutics- could be representative of information that is the most useful in helping people make judgements about the vaccines.

Conclusions

News events (and thus information) that are potentially influential to public sentiment toward the AstraZeneca, Moderna, and Pfizer vaccines and these companies' stocks were organized into the following identified categories: news about the development of Covid-19 therapeutics, government funding of vaccine development (economic news), government and public health guidelines, vaccine clinical trials, vaccine company business dealings, vaccine distribution and eligibility and PPE distribution, and vaccine reactions/side effects.

The sentiment metrics with the most number of significant days of change are negative tweet sentiment and neutral tweet sentiment. This suggests that the neutral and negative sentiments towards the vaccines are subject to the most change. Considering that previous studies have demonstrated a link between negative sentiments in tweets related to COVID-19 vaccines and vaccine misinformation [10], this observation suggests that the spread of misinformation is manageable, since the negative sentiment has a demonstrated potential for significant change. Similarly, Figure 9 showed that information that

is potentially important to share with the public to mitigate the spread of misinformation includes information about the distribution and eligibility requirements for the vaccines, information about the business dealings of the vaccine companies, information about vaccine clinical trials, information about government and public health guidelines, information about government endorsements of the vaccines, and information about the development of COVID-19 therapeutics.

Furthermore, this same figure suggests that information that is correlated with an increased spread of misinformation (increase in % of negative sentiment) includes news about the business dealings of the vaccine companies, vaccine clinical trials, government and public health guidelines, vaccine reactions/side effects, and the development of coronavirus therapeutics. Based on these observations, it seems that news about vaccine company business dealings has the power to both increase and decrease the spread of misinformation. Upon comparison of the news headlines associated with days in which a decrease in the negative sentiment was observed with the news headlines associated with days in which an increase in the negative sentiment was observed, there doesn't seem to be any distinguishing characteristics between the two groups of news headlines. Thus, this conflicting result could be explained by shortcomings of our methodology- the most dominant news items were determined based on word cloud analysis, and it is possible that the news items identified associated with each influential day are not the actual news item that is affecting the observed changes in sentiment. More research into the effect of information about vaccine business dealings on public sentiment towards COVID-19 vaccines is warranted. The same conflict exists for news items related to vaccine clinical trials, government public health guidelines, and the development of alternative COVID-19 therapeutics. Another possible explanation for this observation is that these news items are more controversial, and are perceived both negatively and positively. Our results indicate that the only category of information that is uniquely associated with the *decrease* in the percent of negative tweets (as compared to the news events associated with the increase in the percent of negative tweets) is news about the distribution and eligibility requirements of the COVID-19 vaccines and the distribution of PPE. This suggests that to combat the spread of misinformation about vaccines, public health units should invest more resources into informing the public about opportunities to get vaccinated as soon as possible and making sure that this information is accessible. Efforts made by online groups like Vaccine Hunters Canada have greatly facilitated the spread and accessibility of information relating to vaccine distribution and eligibility amongst the public, and have also been credited with decreasing vaccine hesitancy amongst the public. [35] Another piece of evidence from our results suggesting that increased effort should be put into making information about vaccine distribution and eligibility more accessible and better communicated is the fact that more significant dates of change associated with news events belonging to this event category were observed in 2021 than for any other news event category (shown in Figure 7). Thus, although multiple categories of information were identified as having the potential to decrease vaccine misinformation and increase public confidence in the vaccines, better communication of information about the distribution and eligibility requirements of the vaccines would likely be *most* effective in combating the spread of vaccine misinformation.

There are potentially limiting factors in the validity and generality of our conclusions. One simplification that our study makes is that the effect of news items is the strongest on the day that news articles come out about that event. Furthermore, we considered the duration of the impact that an event has to be 24 hours since we are associating these events to daily changes in stock price or sentiment. In reality, it is more than likely that the impacts of the chosen events on the opinions of the public last longer than this. Additionally, we made the simplification that only a single news event could be associated with a particular day of significant change in stock or sentiment. In reality, it could be multiple news items that influences public opinion towards a vaccine in a combinatorial fashion.
Another simplification that our study makes is that the sentiment of tweets calculated by VADER analysis is truly reflective of the tweets' original sentiments. In practice, natural language processing tool kits such as VADER do not perform well when it comes to attributing sentiment to language that utilizes complex ideas such as figures of speech, irony, and sarcasm, and the resolution of these features with sentiment attribution is still a developing area of research. [36] As such, there could be discrepancies in our data for the measured daily proportion of tweets with a particular sentiment and the actual daily proportion of tweets with a particular sentiment.
Furthermore, the results of our study are heavily dependent upon the segregation of events coinciding on significant days of change into categories with similar characteristics. In order to capture the reality that a news item could be related to multiple categories, future similar studies might utilize a 'tagging' system for identifying influential categories of information,

in which news events may be associated with multiple factors related to the response to the COVID-19 pandemic and vaccine development. Furthermore, more research is warranted into the best way to characterize pandemic news events.

One strength of our study is that it offers quantitative information about the relative strengths of influence that certain categories of information have. This has allowed us to directly make conclusions about the importance of specific news events. Another strength in our methodology is that our determination of the dominant news event for each day of significant change in either stock or sentiment is directly based on the content of the discussions occurring on Twitter. Previous studies that attributed news events to significant changes in Twitter sentiment toward hydroxychlorquine such as that conducted by Chan et al. [37] identified important news events based on the number of search results yielded by a Google News search with a set of pre-identified key terms. This method is less effective because the news items identified may not necessarily reflect the topics being discussed on a particular day by actual Twitter users. Furthermore, the generality of our conclusions are likely justified by the fact that our data set includes the opinions of over one million unique Twitter users.

Further research directions include evaluating the effectiveness of different methods (such as the method used in our study and the proposed 'tagging' method) of organizing news events into discrete categories based on common characteristics. Furthermore, perfecting the method of identifying the descriptive categories themselves given a set of news items- specifically in finding the balance between striving for specificity in the categories and identifying common themes in the articles- is of great importance because it would allow us to make better general conclusions about the types of influential information that need to be communicated better to the public in order to mitigate the spread of vaccine misinformation.

Acknowledgements

Our group would like to thank STEM Fellowship for the incredible opportunity to develop our data science and data visualization skills.

Data and Code Availability

The tweet data collected using Twint and all scripts used to analyze data can be found in the following GitHub repository: https://github.com/n-purakan/SF-BDC-2021-Team-3-Project. Furthermore, url links to all news articles used to associate news events to significant days of change for stocks and sentiments can be found here: https://docs.google.com/spreadsheets/d/1hFgHgLMXzvZpX9GcSMibP-x09pSNE8ENiMfNc1GM9II/edit?usp=sharing

References

[1] Coronavirus disease (covid-19): Herd immunity, lockdowns and covid-19. https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19. (Accessed on 06/28/2021).

[2] C J Hutto and Eric Gilbert. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Technical Report 1, May 2014.

[3] Coronavirus Outbreak Timeline Fast Facts - CNN. https://www.cnn.com/2020/02/06/health/wuhan-coronavirus-timeline-fast-facts/index.html.

[4] World Covid-19 tracker: Latest cases and deaths by country. https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/.

[5] The race for the COVID-19 vaccine: A story of innovation and collaboration | CF - Carnall Farrar. https://www.carnallfarrar.com/articles/the-race-for-the-covid-19-vaccine-a-story-of-innovation-and-collaboration/.

[6] Oxford AstraZeneca Covid vaccine has up to 90% efficacy, data reveals | Coronavirus | The Guardian. https://www.theguardian.com/society/2020/nov/23/astrazeneca-says-its-coronavirus-vaccine-has-70-per-cent-efficacy-covid-oxford-university.

[7] Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study | pf-pfizeruscom. https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against.

[8] Moderna Announces Primary Efficacy Analysis in Phase 3 COVE Study for Its COVID-19 Vaccine Candidate and Filing Today with U.S. FDA for Emergency Use Authorization | Moderna, Inc. https://investors.modernatx.com/news-releases/news-release-details/

moderna-announces-primary-efficacy-analysis-phase-3-cove-study.

[9] Rebecca A. Ferrer and Erin M. Ellis. Moving beyond categorization to understand affective influences on real world health decisions. *Social and Personality Psychology Compass*, 13(11):e12502, nov 2019.

[10] Samira Yousefinaghani, Rozita Dara, Samira Mubareka, Andrew Papadopoulos, and Shayan Sharif. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 108:256–262, jul 2021.

[11] Eleonora D'Andrea, Pietro Ducange, Alessio Bechini, Alessandro Renda, and Francesco Marcelloni. Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116:209–226, feb 2019.

[12] Prajwal Eachempati, Praveen Ranjan Srivastava, and Prabin Kumar Panigrahi. Sentiment Analysis of COVID-19 Pandemic on the Stock Market. *American Business Review*, 24(1):141–165, may 2021.

[13] Rui Li, Dian Zheng Fu, and Zeyu Zheng. An analysis of the correlation between internet public opinion and stock market. In *Proceedings - 2017 4th International Conference on Information Science and Control Engineering, ICISCE 2017*, pages 150–153. Institute of Electrical and Electronics Engineers Inc., nov 2017.

[14] Zulfadzli Drus and Haliyana Khalid. Sentiment analysis in social media and its application: Systematic literature review. In *Procedia Computer Science*, volume 161, pages 707–714. Elsevier B.V., jan 2019.

[15] Global Twitter user age distribution 2021. https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/.

[16] More than half of U.S. households have investment in stock market | Pew Research Center. https://www.pewresearch.org/fact-tank/2020/03/25/more-than-half-of-u-s-households-have-some-investment-in-the-stock-market/.

[17] Center for disease control warns that covid-19 is likely headed toward pandemic stage, could affect us schools, businesses. https://www.ajmc.com/view/cdc-warns-that-covid19-is-likely-headed-toward-pandemic-stage-could-affect-us-schools-businesses. (Accessed on 06/30/2021).

[18] twintproject/twint: An advanced twitter scraping & osint tool written in python that doesn't use twitter's api, allowing you to scrape a user's followers, following, tweets and more while evading most api limitations. https://github.com/twintproject/twint. (Accessed on 06/30/2021).

[19] Plotly python graphing library | python | plotly. https://plotly.com/python/. (Accessed on 06/30/2021).

[20] Data considerations for Boxplot - Minitab Express. https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/graphs/boxplot/before-you-start/data-considerations/.

[21] Create a word cloud with a free online word cloud generator. https://monkeylearn.com/blog/create-a-word-cloud/. (Accessed on 06/30/2021).

[22] Pfizer and Moderna: How 2 very different companies developed a COVID vaccine - ABC News. https://abcnews.go.com/Health/pfizer-moderna-companies-developed-covid-vaccine/story?id=74291125.

[23] What's Happening With Moderna Stock? https://www.forbes.com/sites/greatspeculations/2021/01/12/whats-happening-with-moderna-stock/?sh=69e14cec3c3d.

[24] Our Products | Pfizer Canada. https://www.pfizer.ca/our-products.

[25] Forget the Vaccine. Here Are the Real Reasons AstraZeneca Stock Is a Buy | The Motley Fool. https://www.fool.com/investing/2020/12/16/forget-the-vaccine-here-are-the-real-reasons-astra/.

[26] AZD1222 Oxford Phase III trials interim analysis results published in The Lancet. https://www.astrazeneca.com/media-centre/press-releases/2020/azd1222-oxford-phase-iii-trials-interim-analysis-results-published-in-the-lancet.html.

[27] Pfizer and BioNTech Conclude Phase 3 Study of COVID-19 Vaccine Candidate, Meeting All Primary Efficacy Endpoints | pfpfizeruscom. https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-conclude-phase-3-study-covid-19-vaccine.

[28] Promising Interim Results from Clinical Trial of NIH-Moderna COVID-19 Vaccine | National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/promising-interim-results-clinical-trial-nih-moderna-covid-19-vaccine.

[29] A timeline of the Oxford-AstraZeneca Covid-19 vaccine trials. https://hildabastian.net/index.php/covid-19/100-a-timeline-of-the-oxford-astrazeneca-covid-19-vaccine-trials.

[30] Criticism of the Oxford Coronavirus Vaccine | In the Pipeline. https://blogs.sciencemag.org/pipeline/archives/2020/05/18/criticism-of-the-oxford-coronavirus-vaccine.

[31] Fear of COVID-19 Outweighs Fear of Vaccine Side Effects. https://www.verywellhealth.com/vaccine-sentiment-tracker-covid-symptoms-vs-vaccine-side-effects-5116655.

[32] First shipments of Moderna's COVID-19 vaccine begin distribution in the U.S. - CBS News. https://www.cbsnews.com/news/moderna-covid-vaccine-shipments-begin-united-states/.

[33] From Pfizer to Moderna: who's making billions from Covid-19 vaccines? | Pharmaceuticals industry | The Guardian. https://www.theguardian.com/business/2021/mar/06/from-pfizer-to-moderna-whos-making-billions-from-covid-vaccines.

[34] Sho Tsugawa and Hiroyuki Ohsaki. Negative messages spread rapidly and widely on social media. In *COSN 2015 - Proceedings of the 2015 ACM Conference on Online Social Networks*, pages 151–160. Association for Computing Machinery, Inc, nov 2015.

[35] How COVID 'vaccine hunters' help Canadians find appointments | Healthing.ca.

[36] Martin Sykora, Suzanne Elayan, and Thomas W. Jackson. A qualitative analysis of sarcasm, irony and related hashtags on Twitter. *Big Data and Society*, 7(2), dec 2020.

[37] Emily Chan, Ginah Choi, Kendrew Wong, and Shirley Zeng. An Analysis of the COVID-19 Infodemic: The Extensive and Disproportionate Impact of Public Figures on Public Behaviour and Sentiment Towards Hydroxychloroquine. *Stem Fellowship Journal*, page 16, 2020.

Appendices

Appendix 1: Timeline of days for which a significant change in stock prices and/ or tweet sentiment proportions was observed

Date	Metric	Trend	Event Category	Coinciding News Item
24-Feb-20	Astrazeneca Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Moderna Ships mRNA Vaccine Against Novel Coronavirus (mRNA-1273) for Phase 1 Study
27-Feb-20	Astrazeneca Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Gilead and Moderna lead on coronavirus treatments
2-Mar-20	Astrazeneca Stock Significant Days	Increase	Alternative Drug News Events	Pfizer identified some antiviral compounds with potential as coronavirus treatments
4-Mar-20	Pfizer Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer's R&D Chief On Why The Healthcare Industry Must Collaborate To Beat COVID-19 Coronavirus
11-Mar-20	Pfizer Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Researchers rush to test coronavirus vaccine in people without knowing how well it works in animals
12-Mar-20	Pfizer Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Early concerns raised over levels of intact mRNA in Pfizer vaccine
13-Mar-20	Pfizer Stock Significant Days	Increase	Vaccine Company News Events	Coronavirus: Canadian company announces COVID-19 vaccine candidate
16-Mar-20	Astrazeneca Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Clinical Trial of COVID-19 Vaccine Begins in Seattle
16-Mar-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Clinical Trial of COVID-19 Vaccine Begins in Seattle
16-Mar-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Clinical Trial of COVID-19 Vaccine Begins in Seattle
17-Mar-20	Pfizer Stock Significant Days	Increase	Vaccine Company News Events	Pfizer and BioNTech to Co-develop Potential COVID-19 Vaccine

19-Mar-20	Pfizer Stock Significant Days	Decrease	Economic News Events	Did BioNTech Stock Rise Too Far, Too Fast, on Coronavirus Vaccine News?
24-Mar-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	AstraZeneca donates nine million face masks to support staff fighting Covid-19
24-Mar-20	Astrazeneca Stock Significant Days	Increase	Vaccine Distribution/Eligibility and PPE Distribution	AstraZeneca donates nine million face masks to support staff fighting Covid-19
26-Mar-20	Pfizer Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer reports safety data of azithromycin in Covid-19 trial
28-Mar-20	Neutral Tweet Sentiment Significant Days	Decrease	Government Guideline News Events	Merkel thanks compliant Germans, shutdown to continue until at least April 20
28-Mar-20	Positive Tweet Sentiment Significant Days	Increase	Government Guideline News Events	Merkel thanks compliant Germans, shutdown to continue until at least April 20
30-Mar-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	J&J, Moderna sign deals with U.S. to produce huge quantity of possible coronavirus vaccines
30-Mar-20	Pfizer Stock Significant Days	Increase	Vaccine Company News Events	J&J, Moderna sign deals with U.S. to produce huge quantity of possible coronavirus vaccines
30-Mar-20	Positive Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	J&J, Moderna sign deals with U.S. to produce huge quantity of possible coronavirus vaccines
31-Mar-20	Moderna Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Pfizer says its COVID-19 vaccine protects younger teens
13-Apr-20	Astrazeneca Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	EXCLUSIVE: AstraZeneca's Calquence Shows Early Promise For COVID-19 Patients
7-May-20	Pfizer Stock Significant Days	Decrease	Vaccine Company News Events	Moderna shares surge after FDA approves coronavirus vaccine for phase 2 trial

7-May-20	Positive Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Moderna shares surge after FDA approves coronavirus vaccine for phase 2 trial
10-May-20	Neutral Tweet Sentiment Significant Days	Decrease	Alternative Drug News Events	Gilead in talks to expand global supply of COVID-19 drug remdesivir
18-May-20	Moderna Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna Announces Positive Interim Phase 1 Data for its mRNA Vaccine (mRNA-1273) Against Novel Coronavirus
21-May-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	US secures 300 million doses of potential AstraZeneca coronavirus vaccine
1-Jun-20	Pfizer Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Anthony Fauci on Covid-19 reopenings, vaccines, and moving at "warp speed"
7-Jun-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	Report: AstraZeneca Contacted Gilead For Potential Merger
7-Jun-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Report: AstraZeneca Contacted Gilead For Potential Merger
8-Jun-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	AstraZeneca and Gilead reportedly talked about coronavirus mega merger
8-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	AstraZeneca and Gilead reportedly talked about coronavirus mega merger
8-Jun-20	Positive Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	AstraZeneca and Gilead reportedly talked about coronavirus mega merger
9-Jun-20	Negative Tweet Sentiment Significant Days	Decrease	Government Guideline News Events	Race-based COVID-19 data collection should be mandatory, says City of Vancouver committee
11-Jun-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna Vaccine Trials Moving Fast With Final Study in July
11-Jun-20	Pfizer Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Moderna Vaccine Trials Moving Fast With Final Study in July

11-Jun-20	Positive Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Moderna Vaccine Trials Moving Fast With Final Study in July
12-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Moderna COVID-19 vaccine appears to clear safety hurdle in mouse study
12-Jun-20	Positive Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna COVID-19 vaccine appears to clear safety hurdle in mouse study
13-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	AstraZeneca agrees to supply Europe with 400 million doses of COVID-19 vaccine
16-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	AstraZeneca COVID-19 vaccine likely to protect for a year -CEO
25-Jun-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	The NIH claims joint ownership of Moderna's coronavirus vaccine
27-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Canada pledges $300 million to address humanitarian concerns of COVID-19 abroad
29-Jun-20	Neutral Tweet Sentiment Significant Days	Decrease	Government Guideline News Events	The Reason Face Mask Guidelines Keep Changing
1-Jul-20	Positive Tweet Sentiment Significant Days	Increase	Economic News Events	Pfizer stock jumps after it reports positive data in early stage coronavirus vaccine trial
2-Jul-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer/BioNTech COVID-19 vaccine shows promise - virus neutralizing antibodies achieved
2-Jul-20	Positive Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Pfizer/BioNTech COVID-19 vaccine shows promise - virus neutralizing antibodies achieved
10-Jul-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	German Biotech Sees Its Coronavirus Vaccine Ready for Approval by December

14-Jul-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	AstraZeneca ties up with IQVIA to speed up studies of COVID-19 vaccine candidate
15-Jul-20	Astrazeneca Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna stock surges after company announces promising coronavirus vaccine trial data
17-Jul-20	Moderna Stock Significant Days	Increase	Vaccine Company News Events	Exclusive: EU in talks with Moderna, BioNtech, CureVac to secure possible COVID vaccines
20-Jul-20	Astrazeneca Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Studies provide glimpse at efficacy of Covid-19 vaccines from Oxford-AstraZeneca and CanSino
22-Jul-20	Pfizer Stock Significant Days	Increase	Vaccine Distribution/Eligibility and PPE Distribution	U.S. inks coronavirus trial vaccine deal with Pfizer for first 100M doses
28-Jul-20	Pfizer Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer and BioNTech begin large-scale trial of coronavirus vaccine in the United States
3-Aug-20	Negative Tweet Sentiment Significant Days	Increase	Economic News Events	Sickened by COVID-19, low-wage workers lose jobs. Others are denied paid leave.
4-Aug-20	Negative Tweet Sentiment Significant Days	Decrease	Alternative Drug News Events	FDA Approves Abiomed heart pump to treat Covid-19 patients
5-Aug-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Canada signs deals with Pfizer, Moderna for experimental COVID-19 vaccines
5-Aug-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Distribution/Eligibility and PPE Distribution	Canada signs deals with Pfizer, Moderna for experimental COVID-19 vaccines
10-Aug-20	Negative Tweet Sentiment Significant Days	Increase	Economic News Events	How reborn pharma giant AstraZeneca is taking the lead against COVID-19
10-Aug-20	Neutral Tweet Sentiment Significant Days	Decrease	Economic News Events	How reborn pharma giant AstraZeneca is taking the lead against COVID-20

11-Aug-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	Trump says U.S. has reached deal with Moderna for 100 million doses of coronavirus vaccine
25-Aug-20	Neutral Tweet Sentiment Significant Days	Increase	Alternative Drug News Events	Monoclonal antibodies may be a better choice than convalescent plasma for treating COVID-19
26-Aug-20	Negative Tweet Sentiment Significant Days	Increase	Alternative Drug News Events	Existing drugs like statins may be promising COVID-19 treatments
26-Aug-20	Neutral Tweet Sentiment Significant Days	Decrease	Alternative Drug News Events	Existing drugs like statins may be promising COVID-19 treatments
8-Sep-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	The importance of Phase 3 trials for a coronavirus vaccine, and why it can't be rushed
8-Sep-20	Positive Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	The importance of Phase 3 trials for a coronavirus vaccine, and why it can't be rushed
10-Sep-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Pfizer and BioNTech Announce Data from Preclinical Studies of mRNA-based Vaccine Candidate Against COVID-19
10-Sep-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer and BioNTech Announce Data from Preclinical Studies of mRNA-based Vaccine Candidate Against COVID-20
2-Oct-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	Covid-19 vaccine development won't be affected by politics, Pfizer CEO says
2-Oct-20	Neutral Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Covid-19 vaccine development won't be affected by politics, Pfizer CEO says
9-Oct-20	Negative Tweet Sentiment Significant Days	Increase	Alternative Drug News Events	Trump's COVID-19 treatment developed using cells originally drawn from fetal tissue
9-Oct-20	Neutral Tweet Sentiment Significant Days	Decrease	Alternative Drug News Events	Trump's COVID-19 treatment developed using cells originally drawn from fetal tissue

18-Oct-20	Negative Tweet Sentiment Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Hope of Covid-19 vaccine deployment in UK by New Year: Report
21-Oct-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Unprecedented vaccine trials on track to begin delivering results
21-Oct-20	Positive Tweet Sentiment Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	Unprecedented vaccine trials on track to begin delivering results
23-Oct-20	Negative Tweet Sentiment Significant Days	Decrease	Government Guideline News Events	California gives green light to tattoo parlors, nail salons and massage studios to resume indoor operations
2-Nov-20	Neutral Tweet Sentiment Significant Days	Decrease	Vaccine Company News Events	Pfizer To Bypass US Government System For COVID Vaccine Distribution
2-Nov-20	Positive Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Pfizer To Bypass US Government System For COVID Vaccine Distribution
6-Nov-20	Neutral Tweet Sentiment Significant Days	Decrease	Government Guideline News Events	Coronavirus Roundup: CDC Developing App for Tracking Vaccine Side Effects; Biden Plans to Create His Own COVID Task Force If Elected
9-Nov-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	So is Pfizer part of Operation Warp Speed or not? Yes and no.
9-Nov-20	Pfizer Stock Significant Days	Increase	Vaccine Company News Events	So is Pfizer part of Operation Warp Speed or not? Yes and no.
10-Nov-20	Astrazeneca Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Vaccine Race Won,Äôt Decide AstraZeneca's Future
20-Nov-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Trump accuses Pfizer of delaying vaccine announcement until after election
27-Nov-20	Moderna Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Nanomedicine and the COVID-19 vaccines

30-Nov-20	Moderna Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna Announces Primary Efficacy Analysis in Phase 3 COVE Study for Its COVID-19 Vaccine Candidate and Filing Today with U.S. FDA for Emergency Use Authorization
1-Dec-20	Moderna Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	COVID-19 vaccine sprint as Pfizer-BioNTech, Moderna seek emergency EU approval
3-Dec-20	Moderna Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Moderna vaccine confers at least 3 months immunity: study
9-Dec-20	Astrazeneca Stock Significant Days	Decrease	Vaccine Company News Events	Here's What Could Make Moderna's COVID Vaccine a Bigger Winner Than Pfizer's
9-Dec-20	Moderna Stock Significant Days	Decrease	Vaccine Company News Events	Here's What Could Make Moderna's COVID Vaccine a Bigger Winner Than Pfizer's
9-Dec-20	Negative Tweet Sentiment Significant Days	Increase	Vaccine Company News Events	Here's What Could Make Moderna's COVID Vaccine a Bigger Winner Than Pfizer's
14-Dec-20	Pfizer Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	A 'constant flow' of vaccine: Pfizer's Covid-19 shots begin massive rollout
22-Dec-20	Moderna Stock Significant Days	Decrease	Vaccine Clinical Trial/Efficacy News Events	The Moderna Vaccine Antibodies May Not Last As Long As We Hoped
28-Dec-20	Moderna Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Pfizer Backs Two-Dose Covid Vaccine Schedule after U.K Shift
31-Dec-20	Negative Tweet Sentiment Significant Days	Increase	Government Guideline News Events	Pfizer Backs Two-Dose Shot Schedule as U.K. Spaces Out Shots
29-Jan-21	Moderna Stock Significant Days	Increase	Vaccine Clinical Trial/Efficacy News Events	Pfizer vaccine label changes could force Canada into global syringe race
1-Feb-21	Moderna Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Moderna Increases COVID-19 Vaccine Shipments While Pfizer Lags Behind

22-Feb-21	Moderna Stock Significant Days	Decrease	Vaccine Distribution/Eligibility and PPE Distribution	Canada will receive its largest COVID-19 vaccine shipment to date this week
2-Mar-21	Moderna Stock Significant Days	Decrease	Vaccine Company News Events	EMA, Health Canada publish Moderna data package
3-Mar-21	Moderna Stock Significant Days	Decrease	Side Effects/Vaccine Reaction News Events	Good news about the 'Moderna arm' vaccine rash, it's no worse with second shot
21-Apr-21	Moderna Stock Significant Days	Increase	Vaccine Distribution/Eligibility and PPE Distribution	Covid: What is happening with the EU vaccine rollout?
4-May-21	Moderna Stock Significant Days	Decrease	Vaccine Company News Events	Pfizer-BioNTech vaccine not troubled by variants so far, says CEO

Appendix 2: Descriptive Statistics for Boxplots

AstraZeneca Stock Price		Moderna Stock Price		Pfizer Stock Price	
count	320.000000	count	320.000000	count	320.000000
mean	51.950625	mean	95.501000	mean	35.369344
std	3.887973	std	47.567427	std	2.589325
min	37.790000	min	18.590000	min	27.030000
25%	50.045000	25%	62.307500	25%	34.192500
50%	52.680000	50%	74.100000	50%	35.520000
75%	54.625000	75%	140.582500	75%	36.700000
max	61.100000	max	186.020000	max	42.560000

Figure 10: (above) Distribution of daily stock prices of Moderna Inc., Pfizer Inc. and AstraZeneca Plc.

		Change in Price							
		count	mean	std	min	25%	50%	75%	max
Company	Trend								
AstraZeneca	Decrease	151.0	0.788742	0.768935	0.01	0.265	0.590	1.0200	4.49
	Increase	167.0	0.752695	0.705470	0.01	0.305	0.550	1.0100	4.02
Moderna	Decrease	150.0	3.619867	3.622115	0.03	1.130	2.340	4.8825	15.68
	Increase	170.0	4.175059	3.944686	0.02	1.580	3.025	5.6300	25.71
Pfizer	Decrease	161.0	0.451863	0.495315	0.01	0.140	0.300	0.5700	2.59

Figure 11: Distribution of daily changes in stock prices (change in closing price compared to previous day) of Moderna Inc., Pfizer Inc. and AstraZeneca Plc. separated by change trend

		Percent Change							
		count	mean	std	min	25%	50%	75%	max
Sentiment Type	Trend								
Negative	Decrease	200.0	4.720000	3.712210	1.0	2.0	4.0	6.0	22.0
	Increase	208.0	4.591346	4.149538	1.0	2.0	3.0	6.0	26.0
Neutral	Decrease	205.0	5.243902	5.061293	1.0	2.0	4.0	6.0	40.0
	Increase	202.0	5.316832	4.624036	1.0	2.0	4.0	7.0	28.0
Positive	Decrease	205.0	4.980488	4.172695	1.0	2.0	4.0	7.0	23.0

Figure 12: Distribution of daily changes in percent proportion of daily tweets (change in percent of all tweets associated with a particular sentiment as compared to previous day) associated with positive, negative and neutral sentiment separated by change trend

Appendix 3: Daily Tweets Word Cloud Example

Figure 13: Example of word cloud generated by MonkeyLearn that was used to determine the dominant news event