SUNY PLATTSBURGH

Social Media Analytics

MSA 575 - Final Report

Evaluating the Impact of Social Media on Stock Prices due to Coronavirus Pandemic

Authors:
Danish Rai
Bridget Hanley
Siddant Mehta

Supervisor:
Dr. Zareef Mohammed

May 12, 2020

1 Introduction

Recently there has been a lot of chatter about the economy online due to the pandemic. There has been a lot of controversy and debate about whether the country should move forward with reopening businesses and allowing Americans to get back to "normal". Nearly 36 million Americans have applied for unemployment in the last two months, and many are struggling to even receive the benefits [2]. Some experts are saying that the decision to open the economy is more than just a debate of life versus death. We should also be considering the impact that a major depression could have on our quality of life, and how that can lead to deaths as well [3]. A survey published by Harvard shows that the majority of Americans, 93% to be specific, do not think the country and its economy should be reopened anytime soon [4]. With so many Americans out of jobs and businesses being closed, there is less activity in the stock market which is causing a lot of volatility in the market.

2 Objective

We wanted to explore if how Americans are feeling about the economy right now is actually reflective of what the stock market is doing. If we can draw a significant correlation or relationship between the two, we may be able to gain better insight on the overall impact that this pandemic may have on the economy. Overall, our objective is to evaluate the impact of social media on stock prices due to the Corona virus pandemic.

3 Data Collection and Preparation

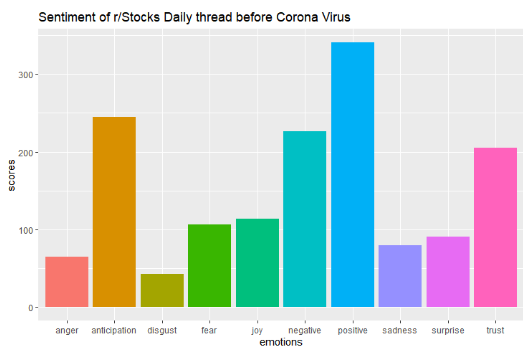
For our analysis, we are looking different time periods of the pandemic to see its effects on the economy. First, we will look at the "before and after" time period of before the pandemic really hit America, and then after. For this, we chose a cut off date of January 31st because that is the date that Trump closed the border to travel from China, and the first case appered in the United states just 11 days before that. For this before and after, we pulled data from the r/Stocks subreddit since there is a daily discussion thread on the subreddit about the stock market and what investors are doing. January 31st was a Friday, so we pulled the comments from every Friday discussion thread for every week between January 31st to May 8Th. We then scraped data in the same manner, just moving backwards from January 31st for the same amount of weeks, so we have dating back every Friday to October 11Th, 2019. For the purpose of analyzing the sentiment of how people feel during the Pandemic we looked at more recent data from Twitter. The data was scrapped using the keywords such as "US Stock Market", "US AND Recession", "US AND Economy", "SPX OR SP500", "DJI OR Dow Jones", and "NDX OR NASDAQ". In addition, the closing index prices for SP 500, Dow Jones and NASDAQ were acquired from Yahoo Finance[1].

We calculated percentage change in index values for each of the indexes for correlation purposes. Furthermore, the data was filtered to only include the weekdays so that the sentiments could be evaluated with the actual changes in the stock market. For our study in relation to recent sentiment analysis we are only focusing on data between April 27Th to May 8Th. Lastly, we analyze the primary companies associated with each of the indexes between the same dates of April 27Th and May 8Th. Companies of our choice where Amazon, Walt Disney and Exxon Mobil Corporation. Amazon has index membership to NASDAQ and SP 500. Walt Disney's and Exxon Mobil's index membership is associated with Dow Jones and SP 500. We utilized Twitter to scrape data for each of the companies. Some of the keywords in our query were "stocks", "market", and "coronavirus" in association to each of the companies stated above. We were able to acquire approximately 10,000 observations for each company to run our analysis. In addition to scraping Twitter data about these companies, we gathered data on the stock prices from each of the companies for the corresponding dates so we could determine if there is an association between the sentiments and the stocks activity.

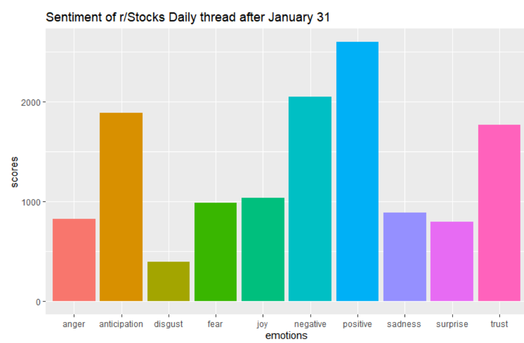
4 Analysis

4.1 Before and After COVID-19

To give an idea of the amount of interaction that the r/Stocks subreddit daily discussion threads have, we first looked at some summary statistics of both data sets. For the before data set, we found a total of 2,738 comments on 16 posts, with an average of 17 comments on every post. The after data set had 305,366 comments with an average of 160 comments on every thread. From just this just basic information, we can see that there has been a lot more interaction after our decided cut off point. Figure 23 below shows the sentiment of the discussion threads before and after January 31st. The sentiment between the two time periods seems to be very similar, but there is a large increase in the negative sentiment after the cut off date. There is also an increase in the sadness as well as fear emotions. However, despite the concern about the economy that many people are expressing, in both cases the entire discussion seems to overall be very positive. This should be taken lightly though, since the method used to created these sentiments does not look at the context of the word, it only looks at the individual words themselves. Therefore, this analysis may not be the most reliable.



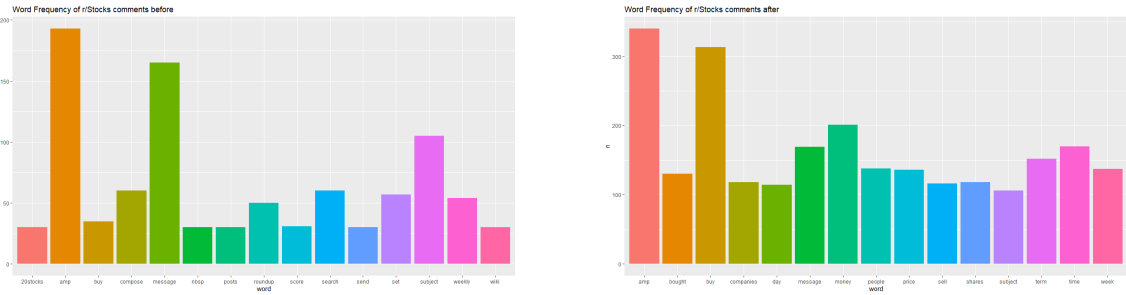
(a) Before (Oct11Th to Jan 31st)



(b) After (Jan 31st to May 8Th)

Figure 1: NRC Sentiment for Before and After Jan 31st

After running this sentiment analysis, we wanted to get a better idea of what people are actually discussing on these threads, so we did some text analysis to get the top 15 most frequently occurring words. The graphs of this are show below in figure 2. The most frequency word between both of the time periods is AMP, which is actually the ticker for Ameriprise Financial. We did look into this and we found that the stock actually had a dramatic decrease between mid February to mid March, but has since began to slowly rise again. In the graph on the right hand side "buy" is the second most frequent word, but we cannot exactly determine the manner in which the word buy is being used, especially from the NRC analysis due to its limitations.



(a) Before (Oct11Th to Jan 31st)

(b) After (Jan 31st to May 8Th)

Figure 2: Top 15 Most Frequent Words in the Daily Discussion Threads

Due to this limitation, we ran another sentiment analysis using the sentimentR package to give us a better idea. Figure 3 is showing the trend of the sentiments before and after our cut off date, with the purple line in the middle dividing the two time periods. The range of these sentiment values is from -1 to 1, so while this graph seems very volatile in the first half, by looking at the scale on the left hand side you will see the movement is no as drastic as one would initially think. Before January 31st, the sentiments do seem to be more susceptible to sharp movements, but really they're moving from a more neutral score up towards a more positive neutral score and at no point does the sentiment drop to be negative. After January 31st, the sentiment seems to level out, only decreasing slightly and does not have many sharp increases or decreases. This can suggest that people, while they are still feeling somewhat optimistic about the market, are being more cautious than they might be in their investments. This could also suggest that the discussions with the word "buy" in them after January 31st include themes about whether or not now is a good time to invest in certain companies, or possibly hold off on buying based on the cautious sentiments.

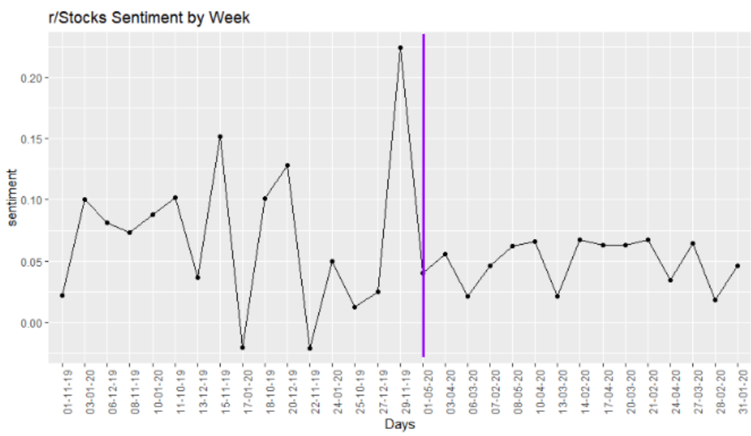


Figure 3: Sentiment of Discussion Threads Over the Weeks

After getting the trend of the sentiments over our specified time period, we wanted to compare the trend to the trends of the indexes that we will be using later in our analysis. The trends of these indexes are seen below in figure 5. Again, the purple line in the middle of the graphs is showing you the cut off date. The trend of these graphs seem to follow a similar trend to the graph in figure 3. At the beginning of the coronavirus outbreak in the United States, it seems as if these indexes were still growing slightly but then dropped off around mid February. At that point in time, we did not have a stay at home order issues in the U.S. but the number of cases were growing and we reported our first death shortly after on February 29th [5]. So as the sentiments on the r/Stocks discussion threads start to level out and people become more cautious of their investments and what not, these indexes seem to follow more of a downward trend, but still similar.

(a) Trend of the NDX index

(b) Trend of the SP500 index

Figure 4

(a) Trend of the DIJ index

Figure 5: Trend of Indexes From October 2019 to May 2020

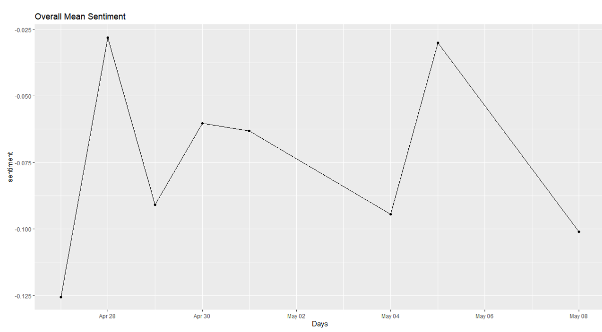
From all of this we can conclude that as the stock market takes a hit during the pandemic, people overall seem to be feeling fairly neutral about the market. However, this conclusion does come with some limitations. In addition to the limitation discussed above with NRC, we do have a very limited scope of information here since we are only looking at one specific subreddit page. In comparison to the amount of posts and data that are out there on the internet about this topic right now, this is a very niche part of the internet and has left us with a very narrow view that doesn't necessiarly translate to how the general public is feeling. Also, it could be assumed that people who are posting on such a subreddit have more experience in the financial world and therefore are more knowledgeable about this topic than the average American, althought there is no way we can know this for sure.

4.2 During COVID-19

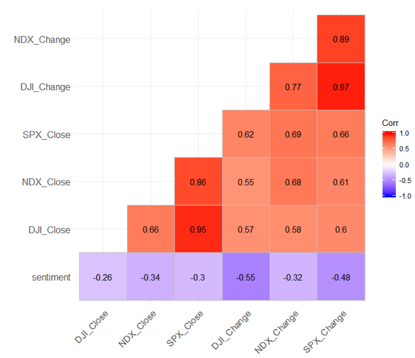
The data set scrapped from Twitter in relation to the stock market and overall index had a total of 10862 observation for overall stock market, 3396 for SP 500 index, 4439 for Dow Jones and 3907 for NASDAQ. There were over 90 different variables in each of the data sets, however, the two important variables we wanted to focus on were the text in tweets and the day it was posted. The text was used to do a sentiment analysis using the SentimentR package on all four data sets. We later merged the data set with index prices and percentage changes in all four data set to analyse its correlation with the sentiments. Furthermore, we used the Sentiment NRC package to look at the words used in the text and categorize the sentiment in different words. However, NRC comes with a limitation as it only focuses on each word rather than the context of the entire text. We later generated frequency plots to observe what were the most frequent words used.

Sentiment and Correlation Analysis

Sentiment R generates values between -1 and 1 for each sentences for each tweet which is refereed to as element_id. We grouped our results by element id to get a mean sentiment of the entire tweet. We later grouped by created at which gave us a mean sentiment for each day for each of our data sets. A correlation analysis was done between the mean sentiment and the index prices with percentage change.



(a) Mean Sentiment

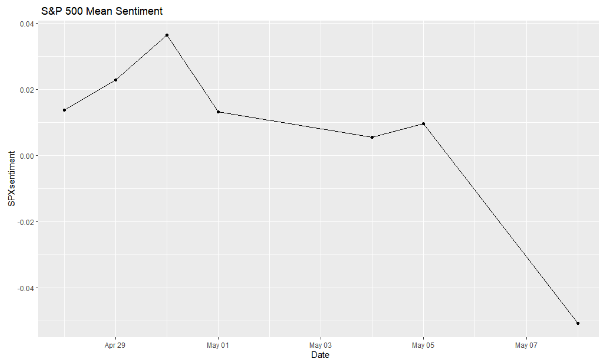


(b) Correlation Plot

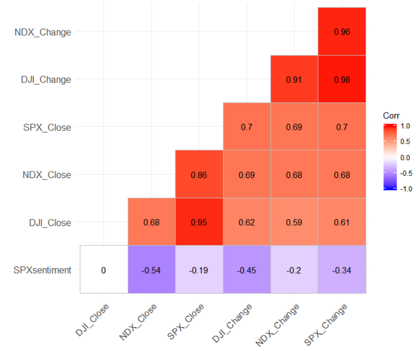
Figure 6: Tweets & Index Prices between (April 27Th to May 08Th)

The 6 (a) shows that overall sentiment has been negative during the 2 weeks. There is some fluctuations in it but it mostly says below zero. We also see that it is much closer to zero then it is to negative one which indicates that the tweets have been negative in majority but there is a significant amount of positive tweets as well. The correlation plot in 6 (b) shows a strong correlation between SPX and DJI index, and SPX and NDX index. In terms of the sentiment of the overall stock market and stock prices there wasn't a presence of significant correlation as the maximum correlation was about 0.55. We tested the significance of the correlation using t-test and we see that for negative correlation of 0.55, p-value was 0.1624 which shows that it is relatively insignificant. We got very high p-values for rest of the correlation, showing insignificance.

When we focused on a specific index such as the SP 500 in figure 6 (a), we see that the mean sentiment fluctuates from positive -0.04 to 0.04. Even though we see it changing, it says neutral as it is very close to zero. In addition, we see that the sentiment and the index value have no significant correlation in figure 7 (b). The maximum correlation is between the SP 500 sentiment and NASDAQ index of -0.54. We tested the correlation for significance through the t-test. We found the p-value to be 0.2142 which shows the insignificance of the correlation.

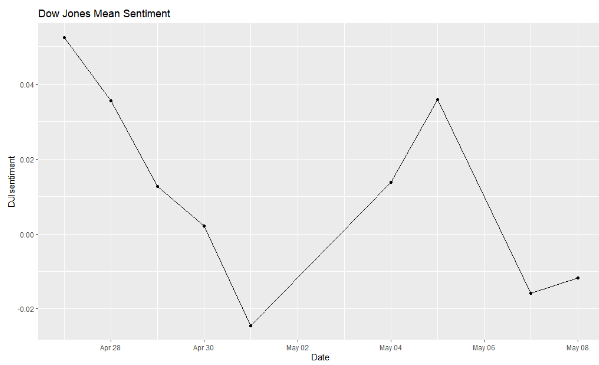


(a) Mean Sentiment

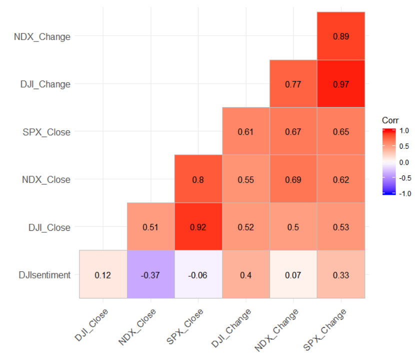


(b) Correlation Plot

Figure 7: SPX realted Tweets & Index Prices between (April 27Th to May 08Th)

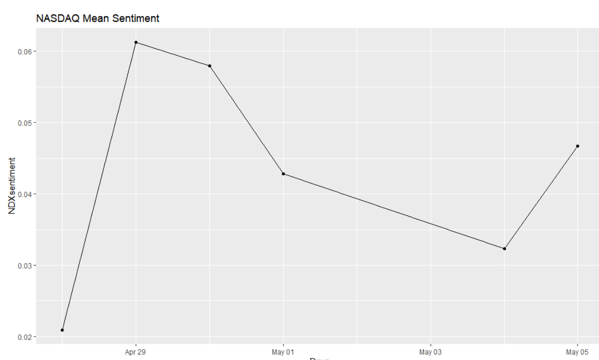


(a) Mean Sentiment

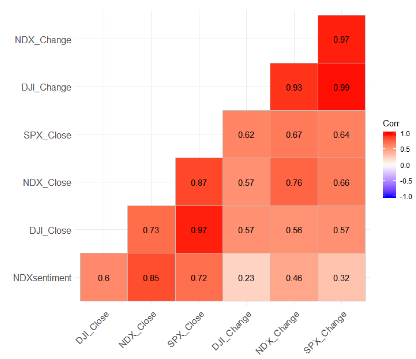


(b) Correlation Plot

Figure 8: DJI realted Tweets & Index Prices between (April 27Th to May 08Th)



(a) Mean Sentiment



(b) Correlation Plot

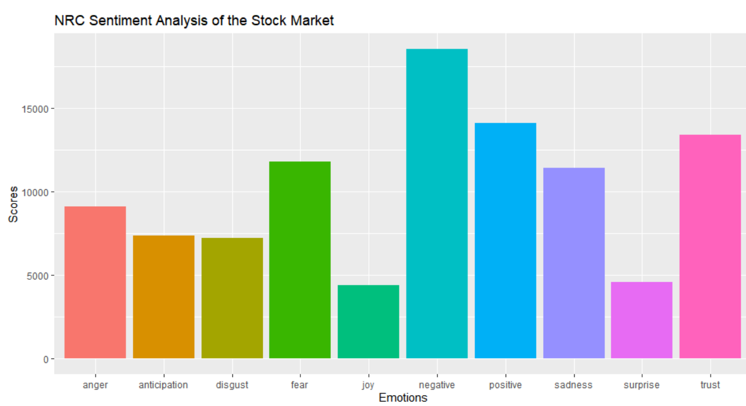
Figure 9: NDX realted Tweets & Index Prices between (April 27Th to May 08Th)

Moreover, for Dow Jones, the sentiment plot shows a downturn in the mean sentiment but it is not significant as it stays close to zero. This again reflects the discussions on Twitter have been both positive and negative. We later analyze if there is a significant correlation between the sentiment and the three indexes. The figure 8 (b) shows that no index or index change has a correlation above 0.5. We analyzed the DJI sentiment and DJI index which has a correlation of 0.4. We tested the significance of the correlation using t-test and found its p-value to be 0.283 which is very high. This suggests that there is no significant correlation between the sentiment of DJI on Twitter with the index value.

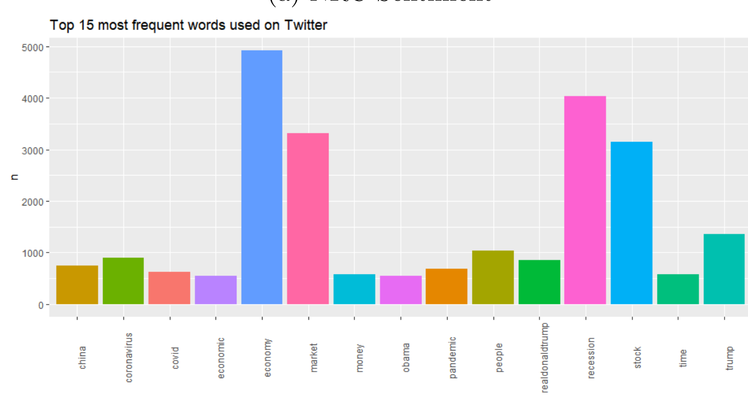
Lastly, we analyze the sentiment plot for NASDAQ in figure 9 (a) which shows that the mean sentiment has been positive throughout with high volatility. When analyzed for correlation in figure 9 (b), we see that there is a significant correlation between the sentiment and the NDX Index of 0.85. We also see a strong correlation of 0.72 between NDX sentiment and SPX index. We tested the significance of the two pairs using t-test which shows a p-value of 0.03117 and t = 3.251 for NDX sentiment and NDX close and a p-value of 0.1033 and t = 2.1029 for NDX Sentiment and SPX index. Our alpha was 0.05, therefore, NDX snetiment and closing index have a significant positive correlation.

NRC Sentiment and Frequency

We used the NRC sentiment package to categorize the sentiments and used the frequency plot to see what were the most used words for overall stock market and each index. First we analyze the overall stock market in figure 10 (a) and see that the majority of the sentiments were negative with a high fear, sadness and anger. We also see a significant portion of positive sentiments with a high trust. The figure 10 (b) helps explain better what the negative sentiments were about. We see that the most frequent word was economy, market, recession and stock. This reflects that people are not very positive about the future of the economy and we see more people taking about recession.

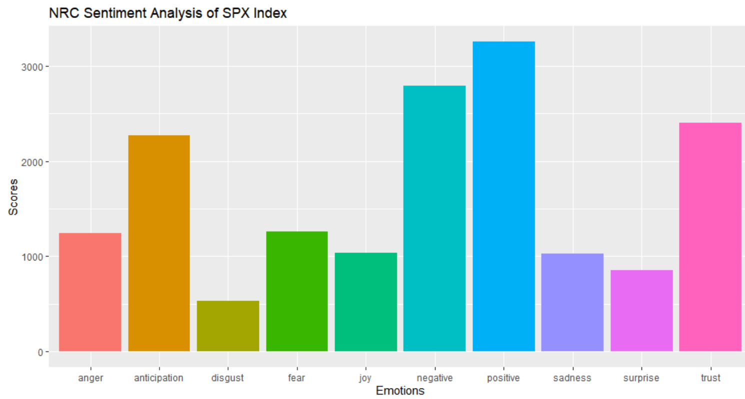


(a) NRC Sentiment

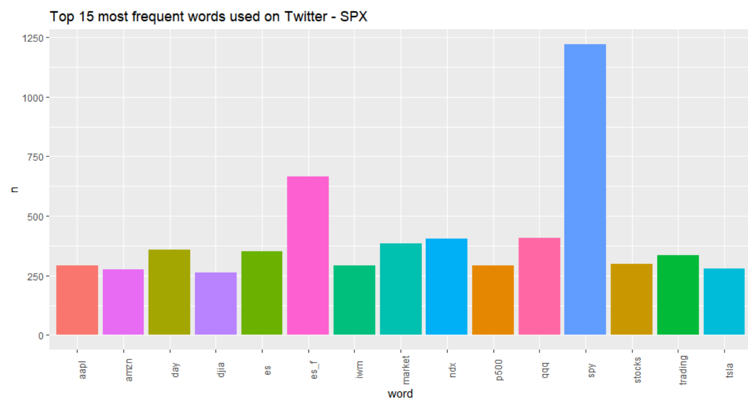


(b) Top 15 most frequent words

Figure 10



(a) NRC Sentiment



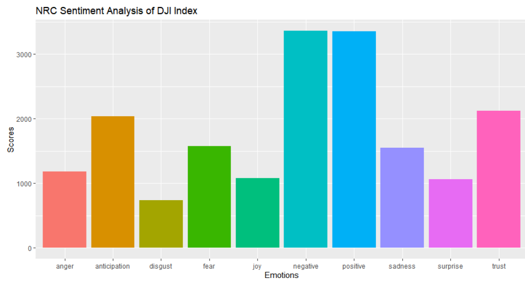
(b) Top 15 most frequent words

Figure 11: S&P 500 NRC Sentiment and word frequency plot

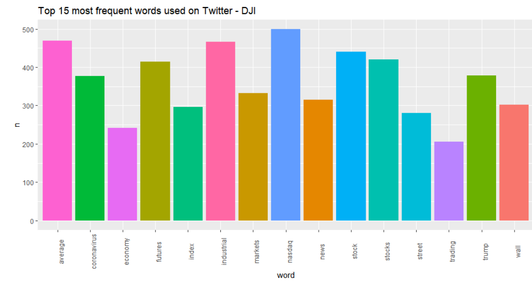
The analysis for SP 500 in figure 11 (a) shows positive sentiments as a majority with a high number of trust and anticipation sentiments. We also see a significant position of negative sentiments with high fear and anger. We further analyzed the sentiments by looking at the frequency plot in figure 11 (b) which shows spy as the most frequent word used. Upon research, we found that spy is is an exchange-traded fund (ETF) which trades on the NYSE Arca under the symbol, SPDR. SPDR is an acronym for the Standard Poor's Depositary Receipts, the former name of the ETF. It is designed to track the SP 500 stock market index. This fund is the largest ETF in the world with top holdings in Microsoft, Apple, Amazon, Facebook. Es_f, the second most frequent word, is a stock market index futures contract traded on the Chicago Mercantile Exchange's Globex electronic trading platform. We see other futures and ETF indexes which shows that people have been talking positively about these ETFs and SP overall.

Taking a closer look at the Dow Jones' NRC sentiment in figure 12 (a), we see that interestingly the sentiments are neutral as positive and negative sentiments have the same score. We see that there is same number of trust and anticipation, sadness and fear, and joy and surprise. Moreover, the frequency plot in figure 12 (b), we see that the words such as industrial, coronavirus, stocks, trump and markets. This shows that there have been a lot of discussion on what the President Trump says and the impact of coronavirus on the stock market.

Lastly, we analyze the NRC sentiment for Nasdaq in 13 (a) and see a that there were majority of positive sentiments with low number of anger and sadness. In figure 13 (b),

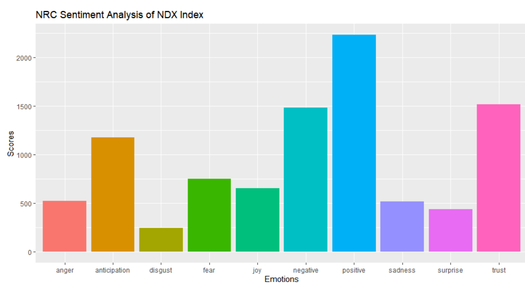


(a) NRC Sentiment

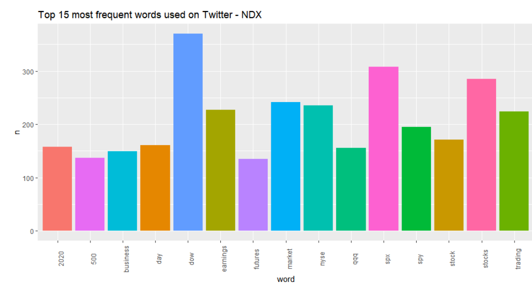


(b) Top 15 most frequent words

Figure 12: Dow Jones NRC Sentiment and word frequency plot



(a) NRC Sentiment



(b) Top 15 most frequent words

Figure 13: NASDAQ NRC Sentiment and word frequency plot

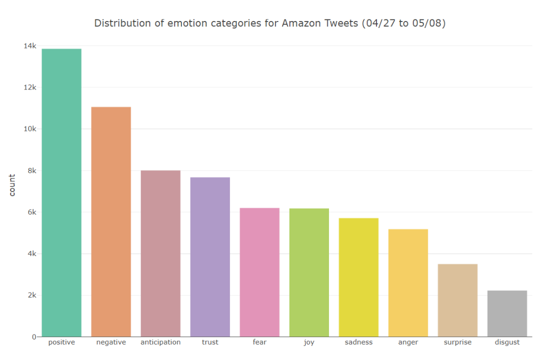
dow is the most frequent word used showing connection between the indexes. This might be because people usually discuss all three sentiments together. We see spy and qqq used frequently in Nasdaq as well in addition to futures. This shows that anytime ETFs and futures were used, the sentiment has been mostly positive.

4.3 Analysis of Primary Stocks Within Their Index

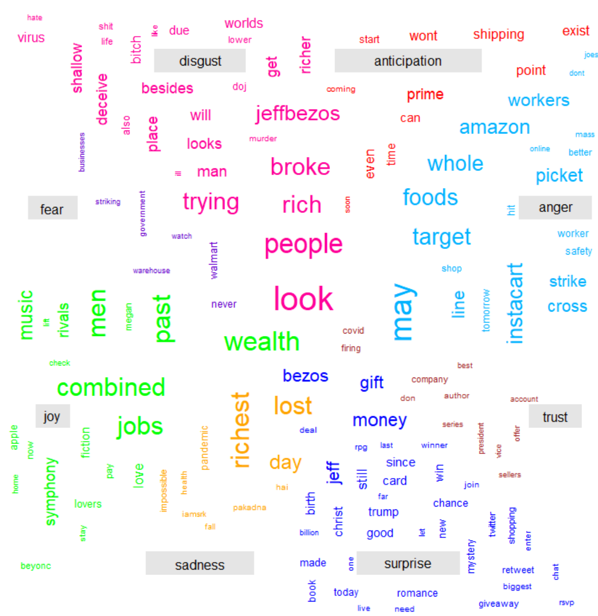
We now focus on the model development for Amazon, Walt Disney and Exxon Mobil Corporation. For each of these companies, we want to derive the sentiments associated to the data we derived from Twitter. The sentiments will provide us a general understanding of how the public is reacting to these companies given our recent circumstances. These companies highly rely on the public overview as it has a great impact on the organization as a whole. With that being said, Our goal is to develop a model that can portray an association between each companies closing stock price and their mean sentiment obtained from Twitter. In order to calculate the mean sentiment, we will be utilizing lexicon "afinn" and associated it with the calculated z-score of the closing stock prices. Z-score is a very useful statistic because it allows us to calculate the probability of a score occurring within a normal distribution and enables us to compare it with the mean sentiment. We went through a vigorous data cleaning process so our analysis can provide us accurate results. As previously mentioned, our study will specifically focus on the dates between April 27th, 2020 and May 8th, 2020.

AMAZON

The Covid-19 outbreak has led to many retailers shutting down store operations and pushing consumers to their websites. As a result, even more consumers are flocking to Amazon for essential and non-essential items. Our analysis focused on deriving the sentiments associated to tweets that were posted on each date. We start of by performed sentiment NRC to obtain the emotions associated with the data set as a whole. Figure 14(a) presents the distribution of emotions by category for Amazon tweets. As you can see, our results indicate that Amazon related tweets had mainly positive output amongst its peers. Approximately 11000 observations associated with negative emotions. There seems to be a lot of anticipation and trust associated with Amazon during this pandemic. Public also shows fear, joy, sadness, anger, surprise and disgust on the lower bound portion of this chart.



(a) Distribution of Emotions by Category for Amazon's Tweets



(b) Amazon's Word Cloud

Figure 14

After taking a look at the distribution of emotions presented above, we want to understand what are some of the most frequent words that are associated with each emotion. Figure 14(b) represents a word cloud of the most frequently used words in association to the distribution of emotions by category. We can see that Jeff Bezos is mentioned in numerous categories. What is interesting to see, the emotion " anger" preferences workers and their safety. This has been an ongoing topic on social media where the public has expressed Amazons inadequacy in providing their workers with a safe environment. The most frequent words associated with " anticipation" reference Amazon Prime along with their shipments. Ultimately, this word cloud is a great illustration behind the most commonly used words in association to the distribution of emotions that we were able to drive through sentiment NRC.

After obtaining a general understanding behind the emotions along with the realization of the most frequently used words, we perform further sentiment analysis to obtain the mean sentiments along with its frequency for the dates associated with our study. The charts below represents the mean sentiment score along with their frequency(positive, negative and neutral).

	date	meanSentiment
1	2020-04-27	-0.469603524229075
2	2020-04-28	0.332050048123195
3	2020-04-29	0.567332754126846
4	2020-04-30	0.708791208791209
5	2020-05-01	1.13802083333333
6	2020-05-04	0.675289919714541
7	2020-05-05	0.685714285714286
8	2020-05-06	0.306688417618271
9	2020-05-07	0.22213181448332
10	2020-05-08	0.217938630999213

(a) Mean Sentiment

	date	negative	neutral	positive
1	2020-04-27	352	478	305
2	2020-04-28	321	312	406
3	2020-04-29	254	452	445
4	2020-04-30	413	294	385
5	2020-05-01	271	299	582
6	2020-05-04	303	370	448
7	2020-05-05	188	311	341
8	2020-05-06	324	451	451
9	2020-05-07	281	475	473
10	2020-05-08	115	918	238

(b) Sentiment Frequency

Figure 15

There are 2822 sentiments that are categorized as negative, 4360 sentiments that are categorized as neutral, and 4074 sentiments that are categorized as positive within the dates of our study. The results that we have obtained are associated with the method "afinn" for sentiment analysis. The chart below represents the distribution of sentiments from April 27th, 2020 and May 8th, 2020. As you can see, May 1st provided us with the highest frequency of positive sentiments compared to May 8th where it provided us the lowest. The sentiment frequency fluctuates day by day and ultimately provides us with an understanding of the public reaction to Amazon's presents during this pandemic.

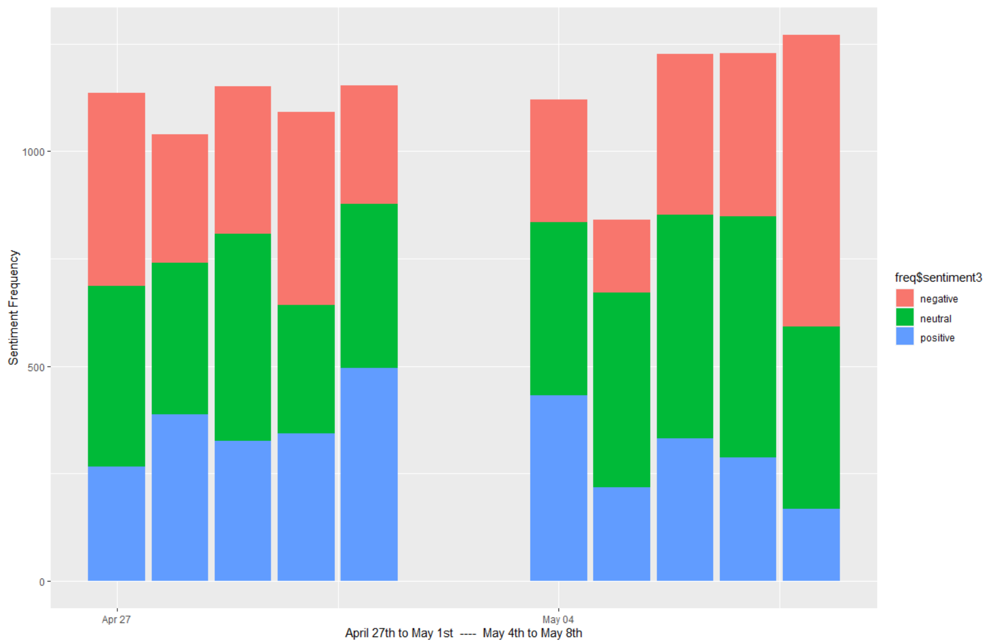


Figure 16: Sentiment Frequency Graph

Finally, we utilize the results received from the meat sentiments and calculate the Z score of Amazon's closing stock price to show association between the two results. The two graphs below represent the association between the mean sentiments score along with Amazon's closing stock prices. Our analysis was able to provide us with excellent results. As you can see, there is a significant association between Amazon's closing stock prices to the mean sentiment obtained through Twitter. Our results seem to indicate a predominant inverse correlation between the two analyses. With the exception of April 28th to April 30th, as the stock price increases, the mean sentiment score decreases and vice versa.

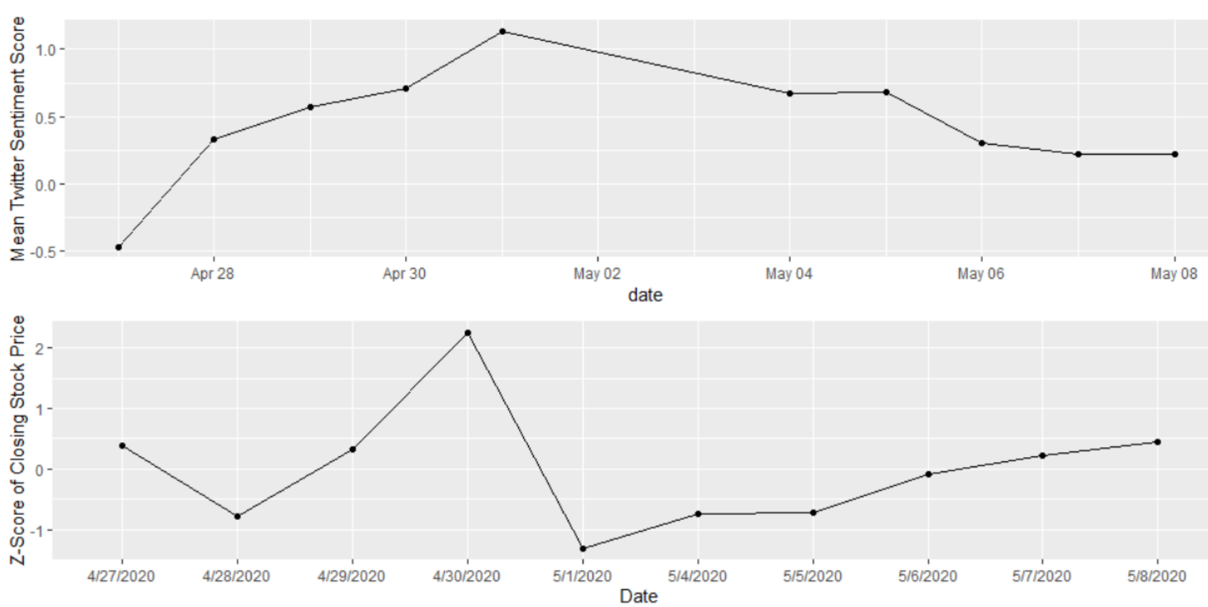
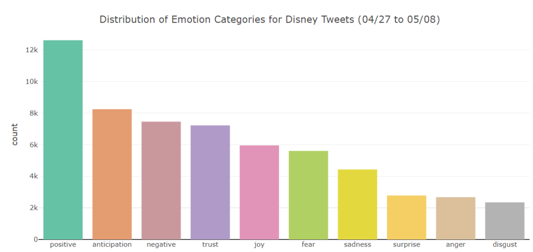


Figure 17: Mean Sentiment Score by Amazon's Closing Stock Price

THE WALT DISNEY COMPANY

We now run our analysis on Walt Disney. We utilize the same methodology in terms of developing a model similar to Amazon. Disney's massive empire has been hit harder than most media companies due to the coronavirus pandemic. With operations beyond producing and distributing film and television, including its theme parks and consumer goods, Disney is seeing a significant disruption as people practice social distancing. We start off by performing sentiment NRC to obtain the emotions associated with the data set as a whole. Figure 18(a) presents the distribution of emotions by category for Walt Disney's tweets. As you can see, our results indicate that Disney's related tweets had mainly positive outputs amongst its peers. Public also shows fear, joy, sadness, anger, surprise and disgust on the lower bound portion of this chart. Taking a look at the word cloud, we can see that the words "shanghai" and "disneyland" have had high frequencies. This is very interesting to see because Disney made the decision to close all of its theme parks around the world in mid-March after it closed its parks in Shanghai this year. During Disney's second-quarter earnings call, management estimated closing Shanghai's park for two months will result in 145 million dollars loss of operating profit. This has been an ongoing topic in social media.



(a) Distribution of Emotions by Category for Walt Disney's Tweets

(b) Walt Disney's Word Cloud

Figure 18

After obtaining a general understanding behind the emotions along with the realization of the most frequently used words, we perform sentiment analysis to obtain the mean sentiments along with its frequency for the dates associated with our study. Figure 19 represents the mean sentiment score along with their frequency. There are 2339 sentiments that are categorized as negative, 4688 sentiments that are categorized as neutral, and 4351 sentiments that are categorized as positive within the dates of our study. The results that we have obtained are associated with the method "afinn" for sentiment analysis. The chart below represents the distribution of sentiments from April 27th, 2020 and May 8th, 2020. As you can see, April 29th provided us with the highest frequency of positive sentiments compared to May 5th where it provided us the lowest. The sentiment frequency fluctuates day by day and ultimately provides us with an understanding of how the public reacts to Disney's presents during this pandemic.

	date1	meanSentiment
1	2020-04-27	0.568609865470852
2	2020-04-28	-0.443396226415094
3	2020-04-29	5.1852207293666
4	2020-04-30	3.33239962651727
5	2020-05-01	1.49830508474576
6	2020-05-04	0.391030684500393
7	2020-05-05	-0.140350877192982
8	2020-05-06	0.406303236797274
9	2020-05-07	0.803071364046974
10	2020-05-08	0.358608385370205

(a) Mean Sentiment

	date1	negative	neutral	positive
1	2020-04-27	253	405	457
2	2020-04-28	524	357	285
3	2020-04-29	64	180	798
4	2020-04-30	80	364	627
5	2020-05-01	102	639	439
6	2020-05-04	356	452	463
7	2020-05-05	446	508	186
8	2020-05-06	169	633	372
9	2020-05-07	131	599	377
10	2020-05-08	214	551	356

(b) Sentiment Frequency

Figure 19

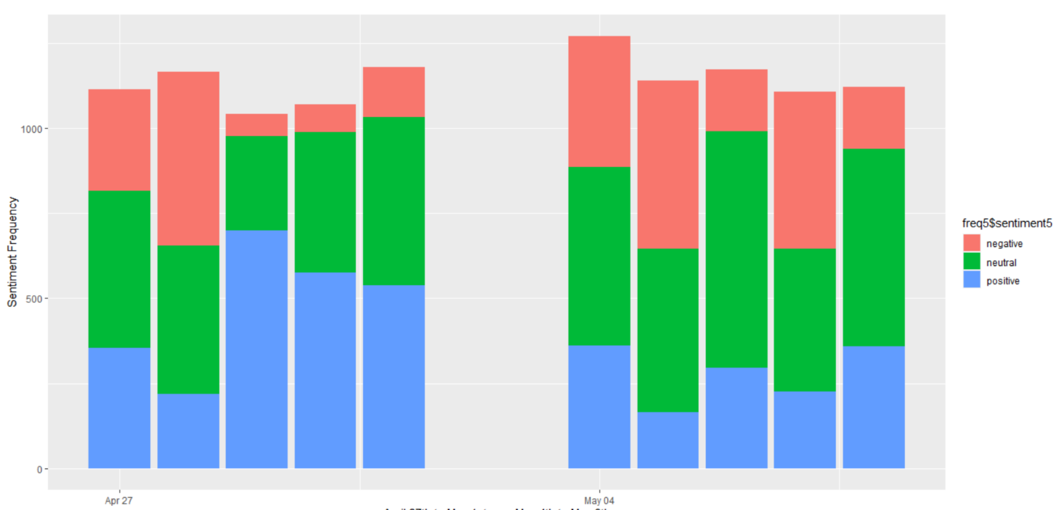


Figure 20: Sentiment Frequency Graph

Finally, we utilize the results received from the meat sentiments and calculate the Z score of Walt Disney's closing stock price to show association between the two results. The graph below represents the association between the mean sentiment score along with Walt Disney's closing stock prices. The results that we obtained from our analysis were very satisfactory. As you can see, there is a significant association between Walt Disney's closing stock prices to the mean sentiment obtained through Twitter. Unlike the analysis we ran for Amazon, Walt Disney's analysis seems to have a direct association between the mean sentiment score and their closing stock prices.

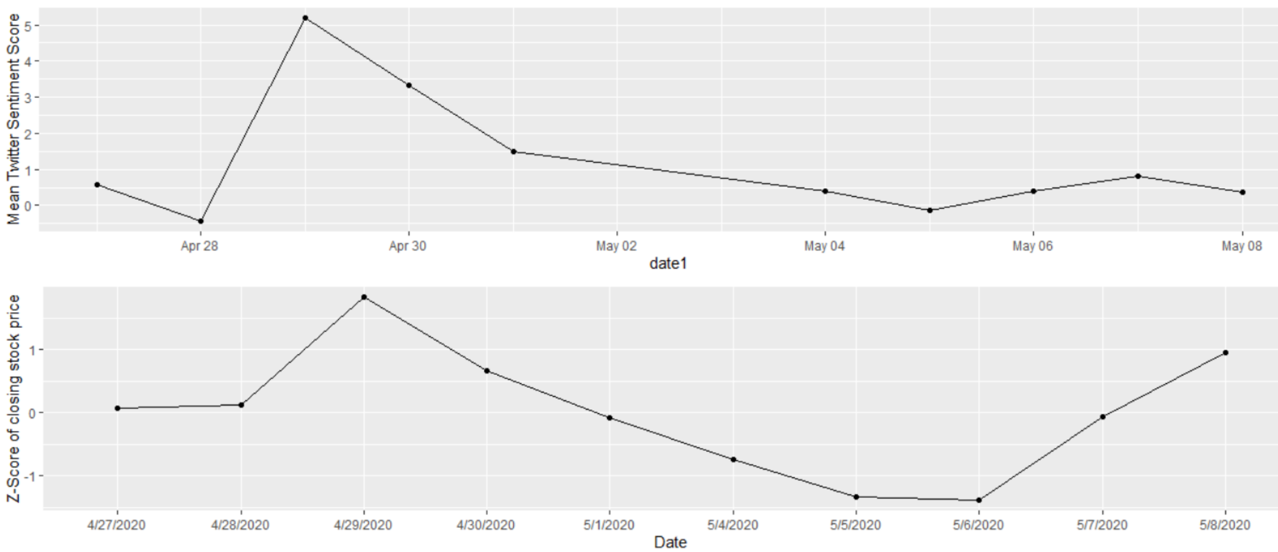
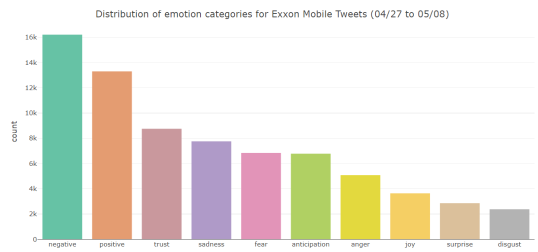


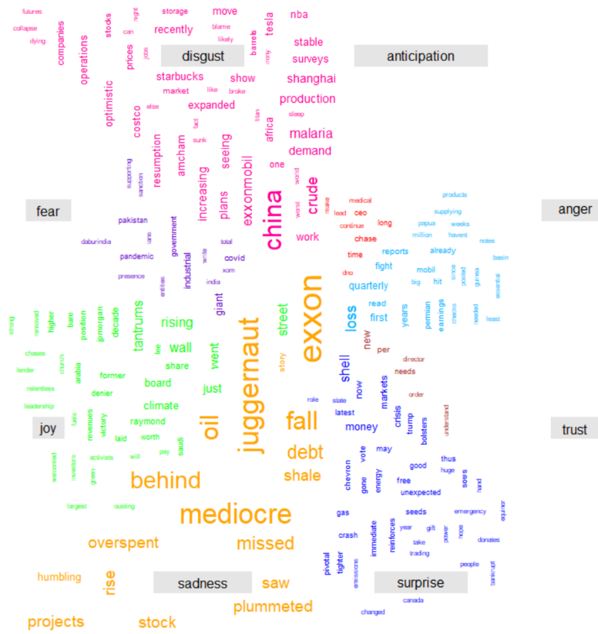
Figure 21: Mean Sentiment Score by Walt Disney's Closing Stock Price

EXXON MOBIL CORPORATION

We now run our final analysis on Exxon Mobil by utilizing the same methodology in terms of model development as before. Exxon Mobil has cut its 2020 capital spending by 30 percent as global demand for oil is sapped by the coronavirus. We start off by performing sentiment NRC to obtain the emotions associated with the data set as a whole. The chart below presents the distribution of emotions by category for Exxon Mobil's tweets. As you can see, there are predominantly negative emotions associated with Exxon Mobil's recent activities on social media. Approximately 13000 observations associated with positive peer overview. Lower bound emotions consist of sadness, fear, anticipation, anger, joy, surprise and disgust. Taking a look at the word cloud, we see that the most frequent words can provide us with a vague understanding of the most frequently talked about topics. There seems to be a big concentration on Exxon Mobil's debt along with the recent changes in oil prices that have occurred in the market. Ultimately, this word cloud is a great illustration behind the most commonly used words in association to the distribution of emotions that we were able to drive through sentiment NRC.



(a) Distribution of Emotions by Category for Exxon Mobil's Tweets



(b) Word Cloud

Figure 22

After obtaining a general understanding behind the emotions along with the realization of the most frequently used words, we perform sentiment analysis to obtain the mean sentiments along with its frequency for the dates associated with our study. The charts below represent the mean sentiment score along with their frequency. There are 4745 sentiments that are categorized as negative, 4451 sentiments that are categorized as neutral, and 3541 sentiments that are categorized as positive within the dates of our study. The results that we have obtained are associated with the method "afinn" for sentiment analysis. The chart below represents the distribution of sentiments from April 27th, 2020 and May 8th, 2020. As you can see, The sentiment frequency fluctuates day by day and ultimately provides us with an understanding of the public reacting to Exxon Mobil's presents during this pandemic.

	date2	meanSentiment
1	2020-04-27	-0.418552036199095
2	2020-04-28	0.600896860986547
3	2020-04-29	0.32887189292543
4	2020-04-30	-1.51652173913043
5	2020-05-01	-1.77502944640754
6	2020-05-04	-0.157291666666667
7	2020-05-05	0.0365088419851683
8	2020-05-06	-0.281481481481481
9	2020-05-07	0.166904422253923
10	2020-05-08	-0.270119521912351

(a) Mean Sentiment

	date2	negative	neutral	positive
1	2020-04-27	164	111	167
2	2020-04-28	127	172	370
3	2020-04-29	140	134	249
4	2020-04-30	962	492	271
5	2020-05-01	967	464	267
6	2020-05-04	635	690	595
7	2020-05-05	478	732	543
8	2020-05-06	466	555	329
9	2020-05-07	342	639	421
10	2020-05-08	464	462	329

(b) Sentiment Frequency

Figure 23

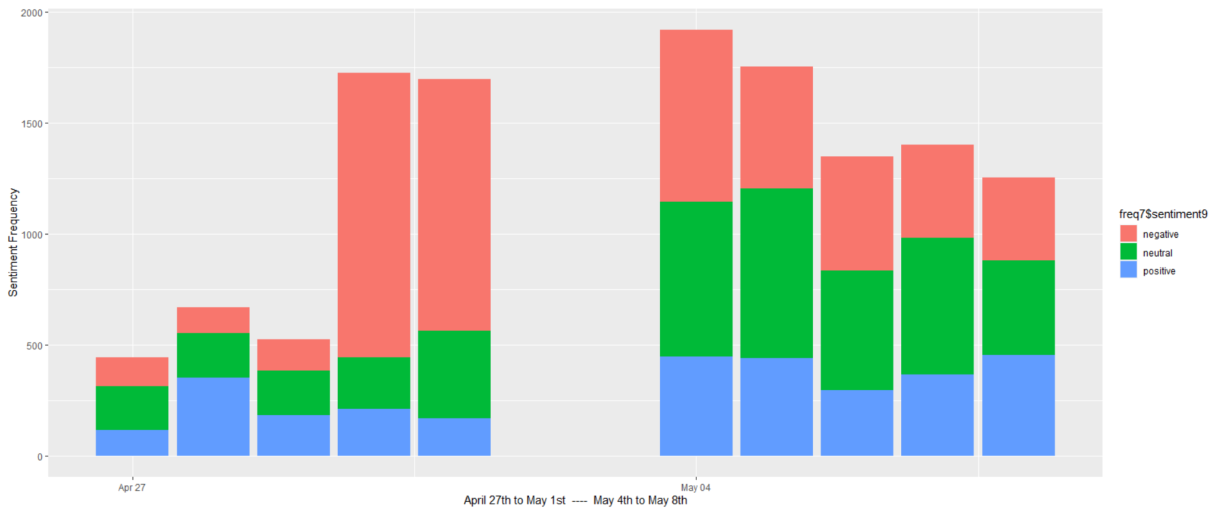


Figure 24: Sentiment Frequency Graph

Finally, we utilize the results received from the mean sentiments and calculate the Z score of Exxon Mobil's closing stock price to show association between the two results. The graph below represents the association between the mean sentiment score along with Exxon Mobil's closing stock prices. The results obtained from our analysis we're not as promising as Amazon's or Walt Disney's. Although we can see some association between the mean sentiment score along with Exxon Mobil's stock prices, there seems to be some discrepancy in our models accuracy. Taking a look at the Sentiment Frequency Graph, we can see that the frequencies fluctuate drastically from April 29th to April 30th. One of the reasons our model shows discrepancy can be due to the uneven frequency presented in our analysis. We can conclude by stating that a larger sample of data should be obtained for this model to present more accurate results.

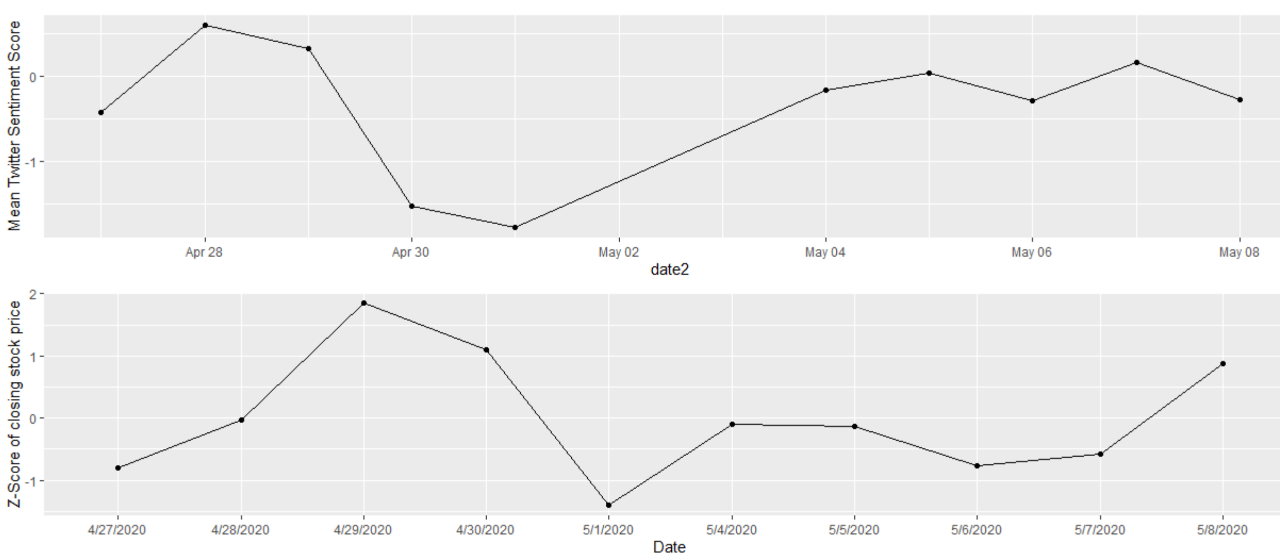


Figure 25: Mean Sentiment Score by Exxon Mobil's Closing Stock Price

5 Limitations

Our study comes with few limitations. Firstly, the data extracted from Twitter only has 1000 tweets from each time it was scrapped. The limitation to the RTweet package is that it only provides recent tweets, therefore, we could only study the two weeks between April 27 and May 08. To better understand the association between the sentiment and stock market, more observation would me optimal due to the volatility in the stock market. In addition, we chose Twitter and Reddit as the two social media platform to analyze and scrape data from. However, a lot of users might use Facebook, blog discussions, news article discussion, or stocktwits to express their sentiments about the economy and stock market. Moreover, our scrapping technique only includes tweets and comments in English. This excludes comments or tweets made in other languages such as Spanish, second largest spoken language in the US. This is a significant limitation to our study as the output of out study could be completely different otherwise.

Secondly, the sentiment techniques used come with limitations itself. For instance, NRC sentiment The words in the NRC lexicon are often duplicated. The words in the NRC lexicon are often duplicated. This duplication often results in difficulties during sentiment analysis. Moreover, since each word is scored in isolation, it can't process modifiers. This means firstly that intensifiers have no effect, so that adding "very" or "really" won't change the valence. On the other hand, the sentence "Stock market is great" has exactly the same positive valence as "Stock market is really good".

6 Conclusion

Our analysis, even with limitation, shows some correlation between the comments and tweets that are put out on the social media about the stock market. Overall, people have been very indifferent about the stock market as Coronavirus has left people in the state of uncertainty. Our before and after Covid-19 analysis shows that the market has been on a negative turn and the same reflects in the sentiment. The market, however, is backed by the US government with a $1.5 trillion investment. This explains the indifferent sentiment as people are unsure how long the US government can keep bailing the market out. The Twitter analysis further provides evidence that there is some association between the sentiments on Twitter. For instance, sentiment for Nasdaq and the index values have a significant positive correlation. When we further analyze individual stocks, we see a significant association between the stock sentiments and the price. With a bigger data set, our analysis could be further expanded to achieve results with higher significance.

References

[1] Yahoo Finance
https://finance.yahoo.com

[2] 'Rolling Shock' as Job Losses Mount Even With Reopenings
https://www.nytimes.com/2020/05/14/business/economy/coronavirus-unemployment-claims.html

[3] 'Will some people be affected badly? Yes.' As Trump says U.S. must reopen soon, question hangs in the air: Can the economy be saved without sacrificing lives?
https://www.marketwatch.com/story/how-do-you-choose-between-economic-deaths-of-despair-and-coronavirus-victims-economists-lawmakers-grapple-with-a-moral-conundrum-2020-03-26

[4] And the survey says, 'keep it closed'
https://news.harvard.edu/gazette/story/2020/04/u-s-public-firmly-opposed-to-reopening-the-economy-immediately/

[5] How the Coronavirus Pandemic Unfolded: a Timeline
https://www.nytimes.com/article/coronavirus-timeline.html