

# Use of Twitter Sentiment in Predicting Stock Market Movement

## ECON 423

Iqbal Bamrah

Email: [ibamrah@uwaterloo.ca](mailto:ibamrah@uwaterloo.ca)

Student ID: 20682484

---

### ***Abstract***

Predicting the stock market movements is an interesting topic of discussion and is important to understanding how the stock market works. Twitter is a massive social media in which any user can share their opinion on any topics and in this case companies within the stock market. Using this public sentiment on the five companies within the S&P 500 of the stock market, there could potentially be correlation between this sentiment and the movement of the stocks within the market. Understanding this Twitter public sentiment and its relation to the movement within the stock market is an important step in order to understand the strength the public have in the movement of stocks within the market. At the end of the research paper, it is determined that there could be correlation between the movement in the stock prices of Apple, Amazon, Facebook, Intel and Microsoft with the sentiment of the public with the use of Twitter.

*Keywords:* Stock market prediction, Sentiment analysis, Twitter

*JEL Classification:* G14; G17; G41

---

## ***Introduction***

### *Research Question*

The research question that I will be highlighting for the research project is if Tweets pertaining to specific companies be utilized to accurately forecast the stock market movements and volatility regarding the stocks being mentioned on Twitter? I believe Twitter has a tremendous impact on the stock market and the volatility of stocks based on the information that can be presented, and shared on the platform from the millions of users that Twitter has that include many public figures, businesses, etc. My interests in why Twitter can be utilized to predict stock market movements and their volatility are the fact that Twitter has over 300 million users who are able to display their opinions on any topics and how users' moods at any time could potential be a catalyst to the movement of different stock prices within the market. Moreover, I have found that in the past when public figures like Elon Musk voice their opinions on Twitter it has affected the movement of the stock prices of the companies he is involved in and such information can be used more broadly on the general use of Twitter by their users. Consequently, I want to know how this information can be used to predict the movement and volatility of the stock prices in the market.

### *Relevant Research Papers*

I have four research papers that I am going to present that are relevant to the research question I posed regarding if Twitter can be utilized to predict stock market volatility and the general movement of the stock prices within the market. To begin, the first relevant paper which related to the topic at hand that I researched was the use Twitter and how the Twitter mood can predict the stock market. The relevance of this paper in relation to the research

question at hand is that the moods of people on Twitter can be used to project the stock market prices for the future in regard to the Dow Jones and many of the other big stocks within the market. The paper involves the use of specific moods that users had, these parameters were calm, alert, sure, vital, kind, and happy. These moods were found within tweets on Twitter and used to predict the changes in the Dow Jones Index and these topics directly relate to the research question of importance.

Moving on, the next research paper that I analyzed, discussed how opinions from users on Twitter about certain companies has correlation to their stock market volatility. This directly related to the research question at hand in regard to figuring out how the opinions of Twitter users that is shared and interpreted can be utilized to predict the stock market movements and volatility and why Twitter can be used to predict the movement of the stock market.

Moreover, the following research paper that I examined had key information in regard to tweets and historical stock prices in the prediction of stock movement. The research paper relates to the research question posed because in a similar way, the research question posed is to find the correlation between general users' tweets on Twitter with the movement of stock prices within the market. The fundamental problem to be solved revolves around the same topic when considering this research paper and the question at hand.

Furthermore, the final research paper that was reviewed compared daily mood on Twitter to the changes in price of the FTSE 100 to find correlation between the mood of the general public and their willingness to invest by looking at the changes in the FTSE 100 index. Again, with the research topic at hand, this paper directly gives information pertaining to the use of Twitter to changes in the market and how this information regarding these changes was

obtained. To conclude, the four research papers that were analyzed had important information that is necessary to the solution of the research topic at hand.

### *Findings and Contributions of Project*

The empirical data of this Twitter project involved data of five S&P 500 companies which were Apple, Amazon, Facebook, Intel, and Microsoft. The stock data for these companies started from 2021-01-01 and finished on 2021-03-26. The purpose of the data being recent is due to the Twitter data only being able to run a few days into the past for data collection. Also, one problem with the data collection is that with the use of R, the amount of Twitter data that could be collected for Tweets pertaining to each S&P 500 company was minimal in comparison to the number of Tweets there were within Twitter. With this in mind, the findings of the relation between Twitter sentiment through the mentioning of the S&P 500 companies and their respective stock market price did not seem to have any significance based on the fact that R was not able to collect the full data set of Twitter mentions. For four of the five S&P 500 companies of Apple, Amazon, Facebook, and Microsoft, the number of tweets mentioning these four companies was too high for R to completely process all the mentions and this indicated that the correlation was off. However, for Intel, due to there not being many Tweets mentioning Intel, R was able to collect the full data set of Tweets mentioning Intel and this allowed for the regression to be completely processed in terms of stock price of Intel and the complete data of Twitter sentiment of Intel. From the data collected for Intel, the regression created with Twitter sentiment of Intel regressed on to its stock price within the market was statistically significant, and this does affect what the outcome of the other four S&P 500 companies could be if the collection of data from Twitter for those four companies was the full

amount of Twitter mentions of the respective companies. If that was the case, the probability of there being correlation between the Twitter sentiment through mentions and the stock price of the S&P 500 companies would likely be high and presumably statistically significant.

### ***Theory and Models***

The theory and models used to analyze the empirical data from Twitter and the stock market prices was through R utilizing daily returns for stock prices of Apple, Amazon, Facebook, Intel, and Microsoft, and Twitter mentions by users of #apple, #amazon, #facebook, #intel, and #microsoft. Through this data modelling, regression models were created to find significance in the relationship of Twitter sentiment and stock market prices of these S&P 500 companies.

Along with the regression models, line graphs of the forecasted prices and Twitter sentiment is used in order to gain a visual representation of the data within the analysis. Moreover, the proposed model was suitable in analyzing the empirical data because the research question is in essence to find the relationship between Twitter sentiment and the movement of the stock market in order to be able to forecast stock prices through the use of Twitter. With this in mind, the use of a regression model is necessary in order to test whether the relationship between Twitter sentiment and stock price movement is significant and to understand if Twitter can be utilized to understand the stock market more clearly. The motivation of the models is to see if there is a relationship between Twitter sentiment and stock prices within the market. This relationship can be utilized in order to be able to forecast the stock market. The properties of this model are the Twitter mention data from 2021-03-28 to 2020-04-01 and the stock prices from 2021-01-01 to 2021-03-26 of Apple, Amazon, Facebook, Intel, and Microsoft. With this

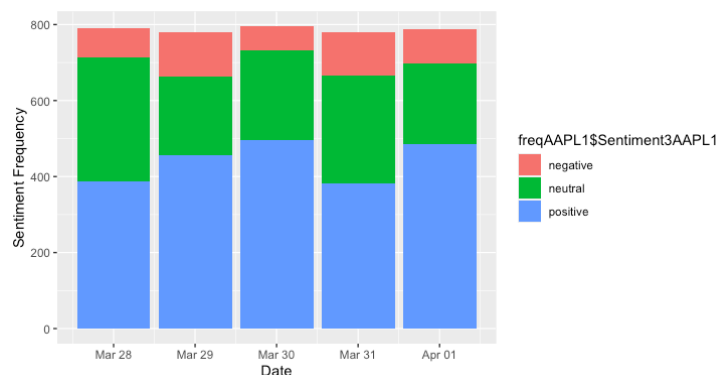
data, the calculations of the daily returns of the stock prices are used to create an arima model to forecast the stock prices for t2021-03-28 to 2021-04-01. Utilizing this forecasted stock price data, and the Twitter sentiment mentions from users on Twitter to create a regression model and graph to understand and find the relationship between Twitter sentiment and the forecasted stock prices of the five S&P 500 companies.

### ***Empirical Data Analysis***

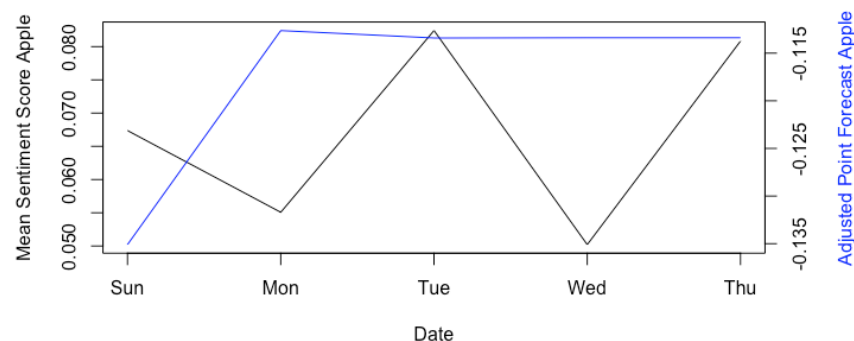
To start, I will go over the analysis of each stock and the Twitter mentions around each stock separately as each company is different and the way the community on Twitter interacts with each company is different and it creates a unique perspective of how each company's stock may react to Twitter sentiment and if there is a relationship there. With the forecasted stock prices, the data is created using arima models of the stock's adjusted close returns as this data is the best indication of the forecasted stock price for each company.

#### ***Apple, AAPL***

The first stock to go over will be Apple by looking at the Twitter sentiment with tweets that mention #apple. With Apple, the first model I created was through the Twitter data and is the sentiment of each tweet from users regarding Apple. From the bar graph we can see the



frequency of the positive, neutral, and negative sentiment from Twitter based on the search for #apple through R. The majority of the tweets can be seen as either positive or neutral and if there is a relationship between this Twitter sentiment and Apple's stock value, this positive and neutral sentiment would directly affect the stock value of Apple. Moving on, looking at the combined line graph of both Twitter sentiment and Apple's forecasted stock value from March 28<sup>th</sup>, 2021 to April 1<sup>st</sup>, 2021, it does look like Apple's predicted stock value does not have much correlation to the Twitter sentiment based on mentions of #apple on Twitter.



This graph shows us the Adjusted Point Forecast of Apple which is the percentage change of Apple's forecasted stock return value over the 5 days in blue and the mean Twitter Sentiment score of Apple for the 5 days in black. Based on this visual diagram, we can see that the forecasted stock value and the Twitter sentiment is not likely correlated based on how the way they change is very different from day to day and they do not move in any correlated way to prove there is significance of a relationship. To further prove they are not correlated, I also regressed the mean Sentiment score of #apple on Twitter on to the forecasted return and based on the estimates and p-value of 0.9695, we can see that there is no relationship between the Twitter sentiment and the forecasted value of Apple's stock.

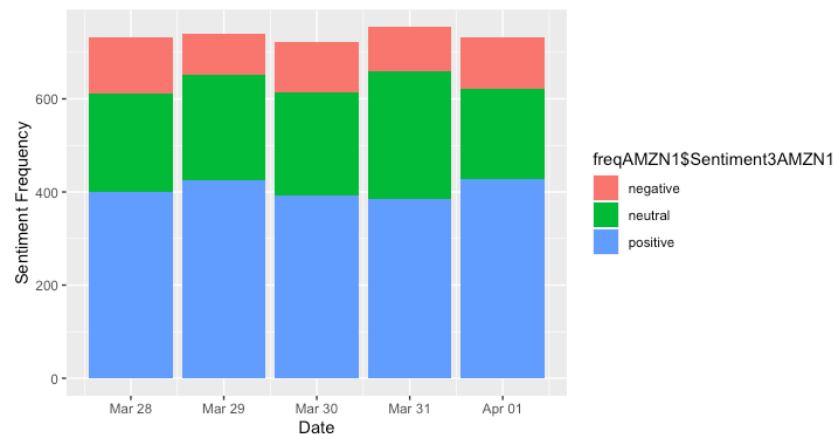
```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.11650    0.02650  -4.396   0.0218 *
MeanSentimentAAPL$secondcolumn -0.01607    0.38722  -0.041   0.9695

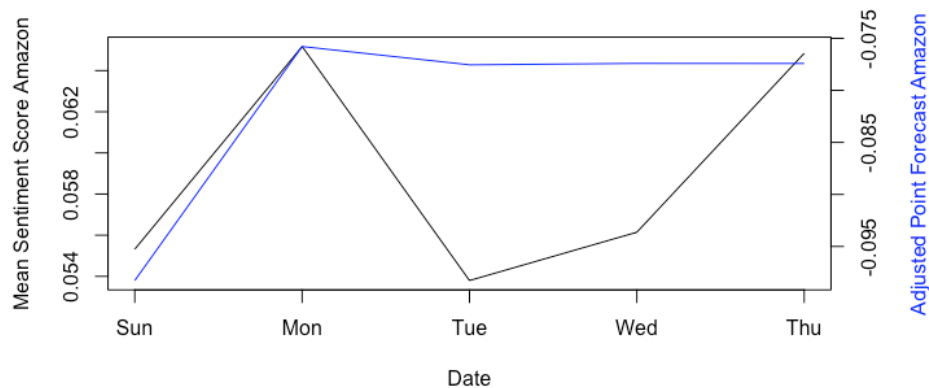
```

## Amazon, AMZN

Moving forward, looking at Amazon's stock returns and the Twitter sentiment around the mention of #amazon on Twitter. The first model created is to look at Twitter's sentiment of the mention of #amazon is and based on the bar graph, we can see that for the 5 days, the sentiment on Twitter comprises of majority positive and neutral mentions of #amazon.



In a way this bar graph is similar to that of the #apple Twitter sentiment bar graph with slight differences which are made up of more or less positive sentiment with each day. Moreover, looking at the line graph of both the Twitter sentiment and the forecasted stock returns of Amazon, it is clear that the first days mean Twitter sentiment and the forecasted Amazon stock



price does align, but after that it does seem to have no correlation, the mean sentiment from Twitter and the forecasted stock price aligned. To further prove this argument that there is no



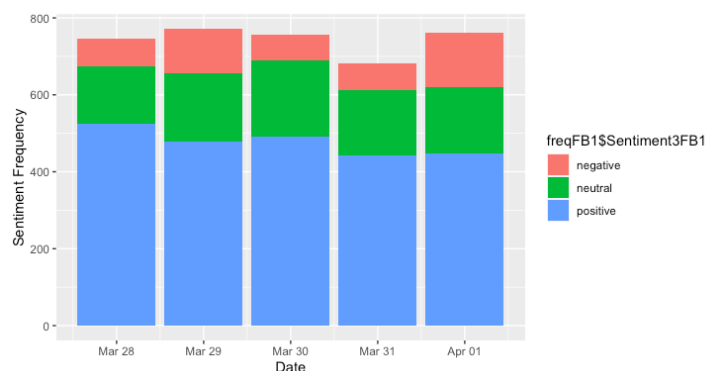
relationship, the regression model of the mean Twitter sentiment for Amazon regressed on the forecasted returns indicates that the p-value of the mean sentiment of #amazon from Twitter is not related to the forecasted stock price of Amazon. From this we can see that the mean

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.1245	0.0537	-2.318	0.103
MeanSentimentAMZN\$secondcolumn	0.7316	0.9061	0.807	0.479

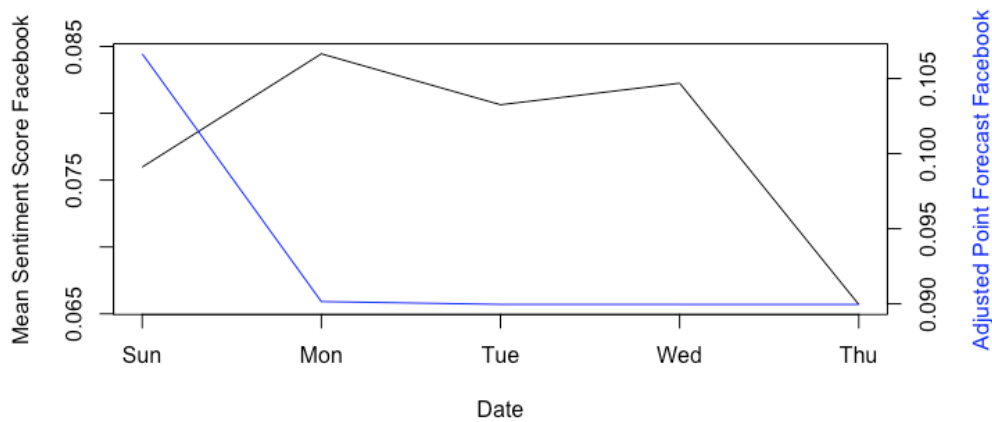
sentiment from Twitter of #amazon had a p-value of 0.479 signifying that there is little to no relationship between the mean Twitter sentiment and the forecasted returns of Amazon.

### *Facebook, FB*

Moreover, looking at Facebook's predicted stock returns from the arima model and its relation to Twitter sentiment. The Twitter sentiment of #facebook on Twitter is much more positive than the sentiment of Apple and Amazon, this may be due to more people tweeting their opinions of Facebook on Twitter in a different light. When people tweet about Amazon or Apple, there is potentially people who are voicing their complaints with something to do with their products and this could be a visual of that. From the diagram below, it shows how much



positive feedback with the mentioning of #facebook there is, and it does indicate that Twitter users are speaking more positively about Facebook in comparison to the other companies.



This graph shows the mean sentiment for each day on Twitter of Facebook, and it is seen as high and stagnant for Monday to Wednesday, and the forecasted stock price of Facebook drops and is stagnant thereafter. This visual representation of the Twitter sentiment and forecasted stock returns of Facebook does seem to have a relationship if we consider the middle three days (Monday to Wednesday) because of how they are both stagnant in their movement, but that does not prove a true relationship. To prove a true relationship, the use of a regression of the mean sentiment of #facebook on Twitter on the forecasted stock returns, there is in fact no relationship between the two variables because of the p-value not being within the 5% threshold. This data does show that the use of #facebook on Twitter does not influence the

Coefficients:

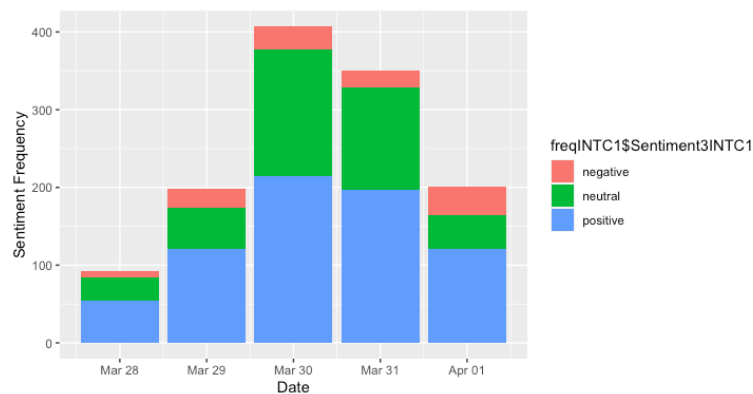
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.10356	0.04467	2.318	0.103
MeanSentimentFB\$secondcolumn	-0.13148	0.57206	-0.230	0.833

change in price forecasted of Facebook within the stock market.

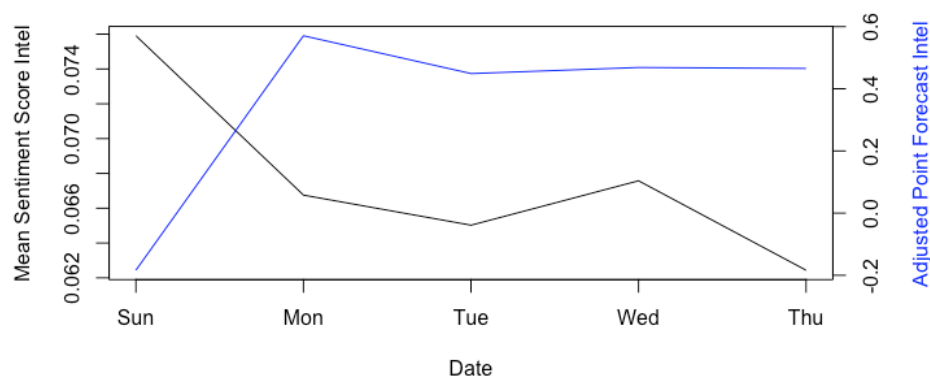
*Intel, INTC*

Moving on, looking at Intel's Twitter sentiment and forecasted stock returns, I think with Twitter's influence, Intel's stock price could be affected because the frequency of tweets that included #intel in them was low in comparison to the other four stocks and this could make the

data much more convincing in proving if there is a relationship with Twitter sentiment of #intel with the forecasted stock market returns of Intel. Looking at the bar graph of the Twitter sentiment, it can be seen that Intel does not get mentioned at the magnitude that Apple or Amazon get mentioned and this does positively affect the effectiveness of the tests conducted. The positive and neutral sentiment is high in comparison to the sentiment of negative



responses which is similar for each company, however with the frequency of tweets being lower, this data has more value in that the probability that this covers most of the tweets regarding Intel for the five days is highly probable. Looking at the line graph with the Twitter sentiment and the forecasted stock returns of Intel it is clear that from Monday on, there seems to be a stagnant relationship based on the visual representation and this may indicate that the Intel stock price changes could have something to do with the Twitter sentiment of Intel. It may be possible that Twitter sentiment with the mention of #intel does have an effect



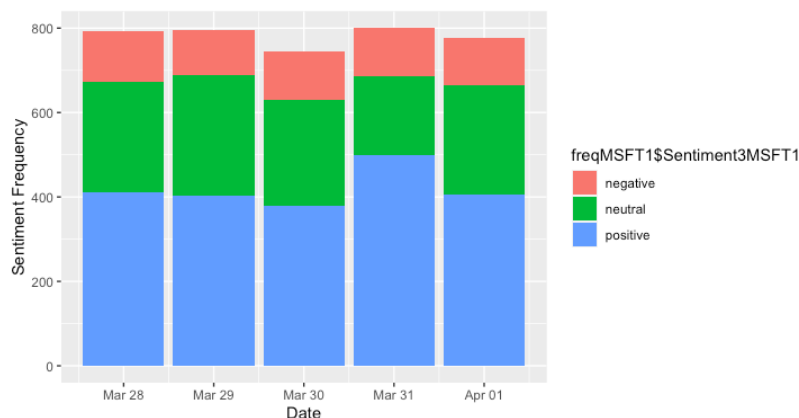
on the stock price of Intel and to prove this the regression of the Twitter sentiment on the forecasted stock returns of Intel is needed. Based on the regression, the mean sentiment on Twitter of Intel is significant at the 5% level based on the p-value. This shows that the Twitter

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.942	1.088	3.623	0.0362 *
MeanSentimentINTC\$secondcolumn	-53.123	16.072	-3.305	0.0456 *

sentiment of Intel does influence the stock market price of Intel and based on that we can prove the relationship.

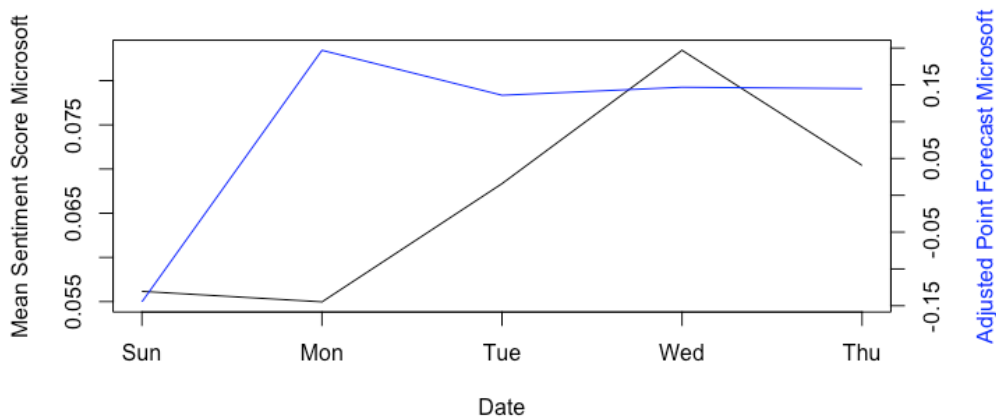
### *Microsoft, MSFT*

Now, looking at Microsoft's stock returns forecast and its relationship to Twitter. Observing the Twitter sentiment of #microsoft and if the sentiment has an effect on the forecasted stock returns of Microsoft. The bar graph below indicates the sentiment on Twitter of #microsoft and



by the data we can see that again, there is a lot of positive and neutral sentiment in comparison to the negative sentiment on Twitter of #microsoft. To tie this into the forecasted stock returns

of Microsoft, the line graph with both the mean Twitter sentiment of #microsoft and the stock returns forecast of Microsoft shows that the Twitter sentiment and the forecasted stock prices



do not have a relationship based on the randomness of the line graph and how both variables do not have any correlation from the visual representation of the data. To prove there is no relationship the use of a regression model with the mean Twitter sentiment regressed onto the forecasted stock returns of Microsoft, it can be seen that the Twitter sentiment has nothing or little to do with the changes in the forecasted stock returns and the value Twitter has in influencing the stock returns is too little to account for in this case.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.2036    0.4210  -0.484   0.662
MeanSentimentMSFT$secondcolumn  4.4922    6.2392   0.720   0.524
  
```

## Conclusion

To conclude, the empirical data that was analyzed used approximately 750 tweets per day over the 5 days to observe if there was a relationship between Twitter sentiment of these companies and their forecasted stock returns. In my opinion, based on the data, it is hard to tell if there could be a relationship between these two variables because the data on Twitter

mentions of these companies that was able to be collected was limited in comparison to the number of tweets there potentially is on each of these companies. With this in mind, the data on each of these companies is still valid and we can see that for the 5 companies within the S&P 500, the only company that proved to have a relationship between Twitter sentiment and the forecasted stock returns was Intel and this was because the data for Intel provided from Twitter was the full length of the Twitter data whereas for the other companies, the data seemed to be limited and incomplete. Thus, it is hard to prove that Twitter sentiment has a relationship with the forecasted stock returns of the companies because the Twitter data is limited, but in my opinion based on the results of Intel and the mention of #intel on Twitter, I could see the possibility of there being a relationship of these variables for the other S&P 500 companies. Based on all this information, I think tweets pertaining to specific companies can be utilized to accurately forecast the stock market movements and volatility regarding the stocks being mentioned on Twitter if there is more data available for R to collect from Twitter. However, because of the limitations, it is impossible to prove this, and I would conclude that tweets from Twitter pertaining to certain companies cannot be proved to have a strong effect on the forecasted movement of the stock market and the volatility of the stocks.

Moreover, there is interesting future research based on this empirical data and analysis because with R the data collection was very limited and with coding software such as Python, there could be potential to gather all the Twitter data pertaining to a specific company and this could give actual results that would prove significant in the grand scheme of the research question. If the data collection was more powerful in collecting the full amount of data available on Twitter, there is interesting future research based on this empirical data and

analysis and it can potentially be used to better understand the stock market and be able to predict the movement of certain stocks that are being mentioned on Twitter.

## **Citations**

- Bollen, J., Mao, H., & Zeng, X.-J. (2010). Twitter mood predicts the stock market. *Pulse of the Nation*, 1-8.
- Nisar, T. M., & Yeung, M. (2017). Twitter as a tool for forecasting stock market movements: A short-window event study. *ScienceDirect*, 101-119.
- Pagolu, V. S., Challa, K. N., Panda, G., & Majhi, B. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *International conference on Signal Processing, Communication, Power and Embedded System*, 1-6.
- Xu, Y., & Cohen, S. B. (2018). Stock Movement Prediction from Tweets and Historical Prices. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 1970-1979.