HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 4 (2023)

An Analysis of the Correlation between Stock Price Fluctuations and Social Media Traction

Junze Sun

Iowa City West High School

ABSTRACT

This paper aims to identify the correlation between stock performance and social media traction during 2020. It analyzes data from three different investment-focused communities within the social media platform Reddit. This data is compared to the historical financial data of six stocks across six different industries in the year 2020. The number of posts with the name of the stock within the original post each week is compared to the absolute percent change between market open to close for that week. As a control, data from the same time frame on the SPDR S&P 500 ETF Trust is used for comparison with the experimental data. All of the stock price information was taken from Yahoo! Finance. This study reveals that there is a statistically significant positive correlation between stock price change and number of Reddit posts each week. The observed correlation between stock performance and social media patterns has a significant impact in determining future stock volatility and predicting short-term investor sentiment.

Introduction

From 2015-2022, there has been a 100% increase in the rate of social media usage among young adults, giving social media platforms increasing influence over global events (Vogels et al., 2022). In the stock market, this is no different. The increasing popularity of financial freedom and literacy among young adults and teens has exposed the opportunity for stock market investment and securities trading. A Wells Fargo study revealed that 45% of adolescents ages 13 to 17 became more interested in investing because of social media events in 2021 (Wells Fargo 2021).

Recent events such as the 2021 GameStop (GME) short squeeze demonstrate the clear market-moving potential social media holds. A group of like-minded, Reddit-based investors in the community Wall Street Bets drastically influenced the price of GME stock (Malz 2021). Not only did they receive global media attention, similar events followed in regards to stocks like AMC Entertainment Holdings (AMC) and Bed Bath and Beyond (BBBY). With these clear examples of online communities influencing the stock market, identifying and quantifying the correlation between online media traction and stock price is imperative. This paper will do this through a method of collection/analysis of financial data and will state any patterns or trends observed. The predicted result is that significant changes in stock price coincide with increased social media traction on Reddit.

Methodology

Parameters

All stock and Reddit data was taken from 2020, starting from the first week of January through the last week of December. Data was recorded on a weekly basis to minimize the number of inputs with zero Reddit posts. The year 2020 was chosen because of the increased frequency of social media and internet usage by individuals quarantining within their homes due to the COVID-19 pandemic. In a study conducted by researchers at the University of Connecticut, seventy percent of individuals reported an increase of social media usage during the first wave of the pandemic

(Rosen et al., 2022). As reported by a collaboration between the Financial Industry Regulation Authority Investor Foundation and the National Opinion Research Center at The University of Chicago, 2020 saw a 38% increase in new investors (Lush et al., 2021). As a result, the year 2020 provides robust data as generated by social media users and better represents investor activity.

This paper addresses potential confounding variables as a result of the choice to analyze market data from 2020. There are countless factors that influence stock price, and in 2020, the pandemic caused major fluctuations across the entire market. Social media activity regarding specific stocks may not have as much of an influence on stock performance during this time period when compared to other, less volatile years, but the opportunity regarding the high frequency of social media usage and buying power from new investors is one that cannot be found during any other time. One additional confounding variable is that Reddit posts do not disappear and are permanent. As a result, the number of comments and other such interaction with original posts can change past the historical time frame. However, this should not be a limitation as this paper will analyze the number of posts–which cannot change after 2020–rather than the interactions with each post.

The financial data used in this paper is sourced from Yahoo Finance's historical stock data, which tracks the open, close, adjusted close, high, and low, values per share of any selected stock on market open days. As one of the most used online tools for stock data and analysis, Yahoo Finance provides clear, transferable data that makes calculations and comparisons simple.

The social media platform Reddit was chosen as a source to pull data about social media traction because no other social media platform is likely to have a tangible impact on the stock market. The existing structure of subgroups within the platform allow for communities dedicated to investing. Subreddits are categorized by their purpose and can contain millions of active members who post, comment, share, and interact with other users on the platform. For the purposes of scraping data regarding specific, individual stocks, three Subreddits named r/investing, r/stocks, and r/wallstreetbets were chosen as the sources of social media data. Each Subreddit is populated by millions of active investors. These three were specifically chosen for their different user bases and slight variations in purpose.

The Subreddit WallStreetBets is primarily made of younger, newer investors who seek to learn the ins and outs of trading and understanding the market. This can be observed by the frequent use of company names rather than their tickers. For the majority of stocks, a greater number of posts include the company name rather than the stock ticker in the same time period. New investors have more experience with company names rather than their tickers. In general, these investors are more willing to take higher risk and accept greater volatility. This also makes them more susceptible to being affected by social media trends and impulse decisions. Additionally, WallStreetBets has the most users out of the three, giving it very high buying power. This combination showed its true potential in January 2021 with the GameStop short squeeze. As a result, 2020 was a unique period to analyze. WallStreetBets had both influence on the market as well as a user base of legitimate investors. After the short squeeze, Subreddit was flooded with users seeking only to gamble their money in hopes of extreme returns.

In contrast, the demographic of the Subreddit r/stocks is made up of more experienced investors who wish to retain the value of their investments. While they may not be as influenced by social media activity regarding specific stocks, they represent the greater majority of investors and are more consistent with their stock choices and portfolio diversity. Their experience can be seen through the higher use rate of tickers than the users of WallStreetBets, a testament to their accustomization to using tickers when trading.

The third Subreddit, r/investing, is a medium between r/stocks and r/WallStreetBets. While they actively renounce the extremely high-risk decisions of WallStreetBets investors, they also are less formal than r/stocks and actively joke about their investments.

The control of this research paper is the SPDR S&P 500 ETF Trust (SPY). This is a necessary control that provides a basis of comparison for specific experimental stocks by identifying the differences between the performance of the market as a whole and the performance of individual companies. The stocks analyzed in this paper represent six different industrial sectors of the market to identify whether certain sectors were influenced more or less

by social media traction. The six companies were Walmart Incorporated, Pfizer Incorporated, ExxonMobil Corporation, Ford Motor Company, Nvidia Corporation, and JP Morgan Chase & Co. All six are among the top revenue generators within their respective industries and are represented within the S&P 500. However, two specific companies stand out for the time frame.

Nvidia represents the technology industry in this experiment. As a major player in the chip manufacturing and graphics processing industry, the year 2020 was one of great growth and change. Already a more volatile growth stock, Nvidia is one of the best companies to include to compare change against. Additionally, the user base of Reddit consists of those who would use Nvidia products, such as graphics processing units to play video games and technology for work.

Pfizer, a healthcare company, had similar opportunities to expand and develop because of the COVID-19 pandemic. As a successful developer of the COVID vaccine, Pfizer experienced steady growth as the world began to recover from the pandemic. Similar to Nvidia, this company was perfect for tracking major changes.

Walmart, ExxonMobil, JP Morgan Chase, and Ford represented retail, energy, finance, and manufacturing, respectively. As high market cap companies, they weathered the storm of the initial stages of the COVID-19 pandemic and witnessed growth as recovery began. They retained old customers and acquired new ones as government-provided stimulus checks increased spending power of consumers.

Data Collection

Financial data from Yahoo Finance was retrieved from the "Historical Data" page of each stock. After sorting by week and setting the time frame, the data was downloaded into Excel, where the file was opened once before being copied into a Google Sheet. In order to quantify the online traction of each specific stock, the number of posts per week by Subreddit mentioning the chosen stock was recorded. This was done using two different data scrapers and Google Sheets/Excel to store data.

The first scraper was an online application programming interface (API) tool called Camas Unddit that showed users all the posts in a given subreddit during a specific time frame that included the search query term. By setting the date boundaries to January 1, 2020, and December 31, 2020, each stock name was searched for each subreddit.

The second scraper used was an application called Octoparse that automatically collected data from webpages and exported it to Excel. Pasting the link of the list of posts collected by Camas Unddit into the input of the Octoparse scraper would result in the collection of the title of the post, the URL, the date and time it was posted, the body text, and the user who posted it. This data was downloaded into Excel and then exported into Google Sheets for exporting and formatting.

Data Comparison

In order to structure the data for comparison, all results gathered from Yahoo Finance and the Reddit data scrapers were categorized by the stock they represented. Each sheet containing financial information was modified by changing the formatting of the date to only displaying the day and month. The percent change each week was also calculated by using the formula: the quantity of the open value minus the close value divided by the closing value. Next, the absolute value of this result was attained by simply using the ABS function and the week change percent formula. The resulting value was then formatted into a percent.

After this, three new sheets per stock were created in order to compare the financial information to each Subreddit, separately. A copy of the week change percentage, absolute value of the week change percentage, and the respective week it aligned to were then transferred to all of the new sheets. The date of each entry of the Subreddit data points were also transferred to their respective sheets by Subreddit.

This data was organized into weeks by using the WEEKNUM function from week fifty-three to week one. This resulted in a list the same length as the list of Subreddit data points by day, but in a format that displays which week each entry is a part of. Next to this list, a separate column with each week number was created, one cell each from week fifty-three to week one. In order to count the total number of entries per week, the COUNTIF function was used. By selecting the list of data entries and also each cell of the unique entries, the function returned a total count of each entry by which week it was in.

Equation 1. Correlation function

$$Correl(X,Y) = \frac{\Sigma(x-\underline{x})(y-\underline{y})}{\sqrt{\Sigma(x-\underline{x})^2\Sigma(y-\underline{y})^2}}$$

The comparison of this organized data was done in three ways. The first was the CORREL function as seen in Equation 1, where x and y are the two columns. The lines above the variables represent the average of each respective column. This function returns a value between negative one and positive one depending on the linear correlation of the data entries in two columns of data. By selecting the column with the number of Subreddit entries per week as well as the column with the absolute value percent change a value within the boundaries of negative one and one would be returned. The greater the value, the more direct the linear correlation is between the two columns. This comparison is a direct quantitative measure between the number of reddit posts and the absolute value of percent change in stock price.

The second comparison indicates trend alignment of the Reddit data. The test compares each week's absolute value percent change and the number of Subreddit posts per week to their respective averages throughout the year. If both the percentage and number of posts for each week are above average or both below average, then the data for the week align. If both values are different, where one is above the yearly average and one is below, then the data for the week does not align. The total number of weeks that align as well as the amount that do not are added and compared. If more weeks align than do not align, then the data trends align. While this test is primitive and does not provide a specific quantitative value for trend alignment, it provides a simple analysis in regards to if the data aligns more than it does not.

The third comparison was the creation of graphs that visually compared the number of Subreddit posts per week and the absolute value of the percent change by week. After selecting the two columns representing this data, the graph function was selected. The graph type was changed to a bar-line combination chart. This graph enables a viewer to visually compare the changes in both posts and stock prices. This is most useful to see where maximum and minimum values may coincide and to observe any patterns.

Results

SPY - SPDR S&P 500 ETF Trust

Table 1. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for SPY by Subreddit. The last row displays if there were more correlating results each week for SPY in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	0.1399442284	0.176771024	0.7071952419
Average number of posts per week	11.66037736	15.0754717	66.58333333
More correlating results per week than not?	Yes	Yes	Yes

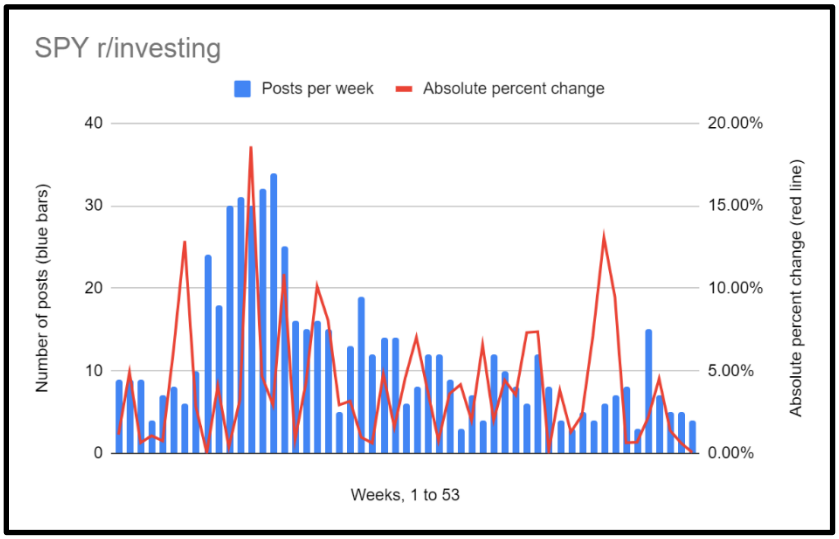

Figure 1. The overlaid graphs of the number of posts per week in 2020 mention SPY in r/investing and the absolute percent change of the share price of the stock each week.

Viewing Figure 1, the graph for r/investing, it can be observed that the week with the greatest number of posts coincided with the week with the greatest absolute percent change in share price. This remains consistent with the visual observations of Figure 2, the graph of r/stocks.

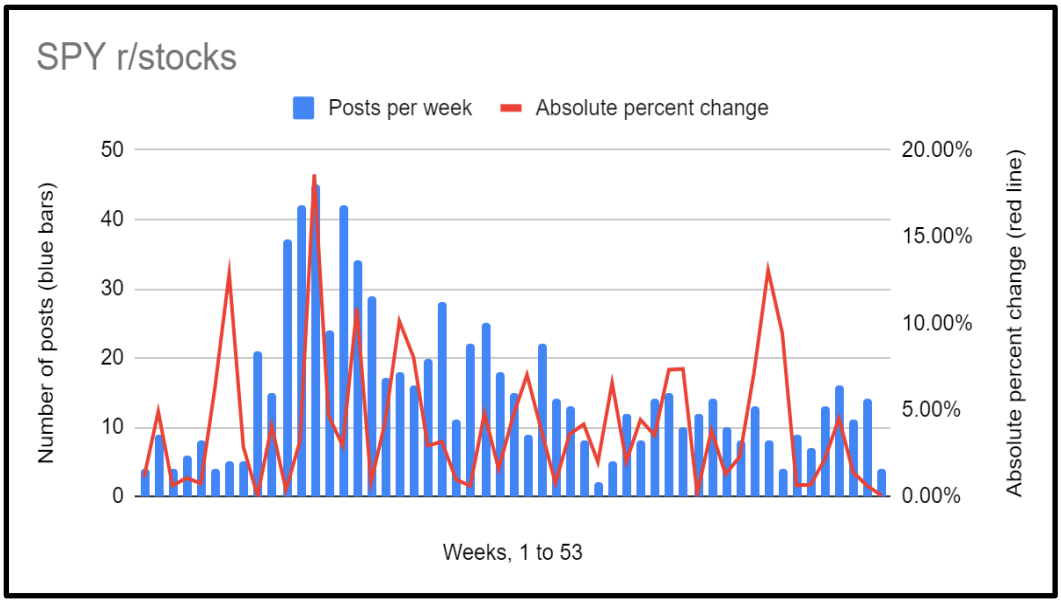

Figure 2. The overlaid graphs of the number of posts per week in 2020 mentioning SPY in r/stocks and the absolute percent change of the share price of the stock each week.

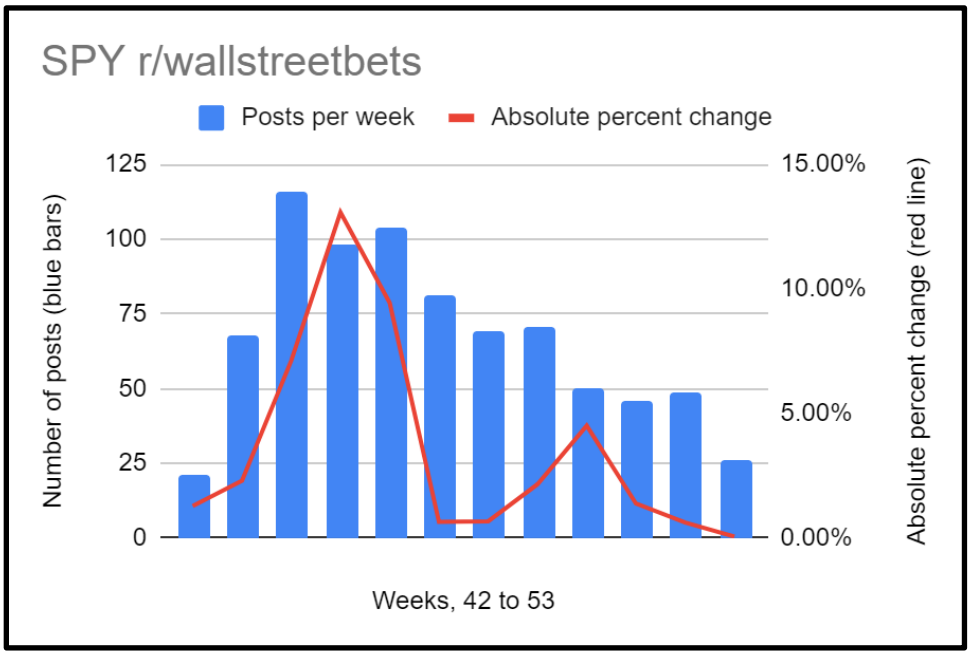

Figure 3. The incomplete overlaid graphs of the number of posts per week in 2020 mentioning SPY in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

To note, the data for the WallStreetBets analysis was corrupted and as such only the time between July 2020 to the end of the year was able to be used. As such, accurately analyzing the data of r/wallstreetbets for SPY is impossible, and Figure 3 is incomplete.

NVDA - Nvidia

Table 2. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for NVDA by Subreddit. The last row displays if there were more correlating results each week for NVDA in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	-0.02063319721	0.1302947956	0.3056926311
Average number of posts per week	2.320754717	5.943396226	7.056603774
More correlating results per week than not?	Yes	Yes	Yes

From figures 4, 5, and 6, the instances where percent change is the highest and the number of posts is greatest coincide or occur within two weeks of each other.

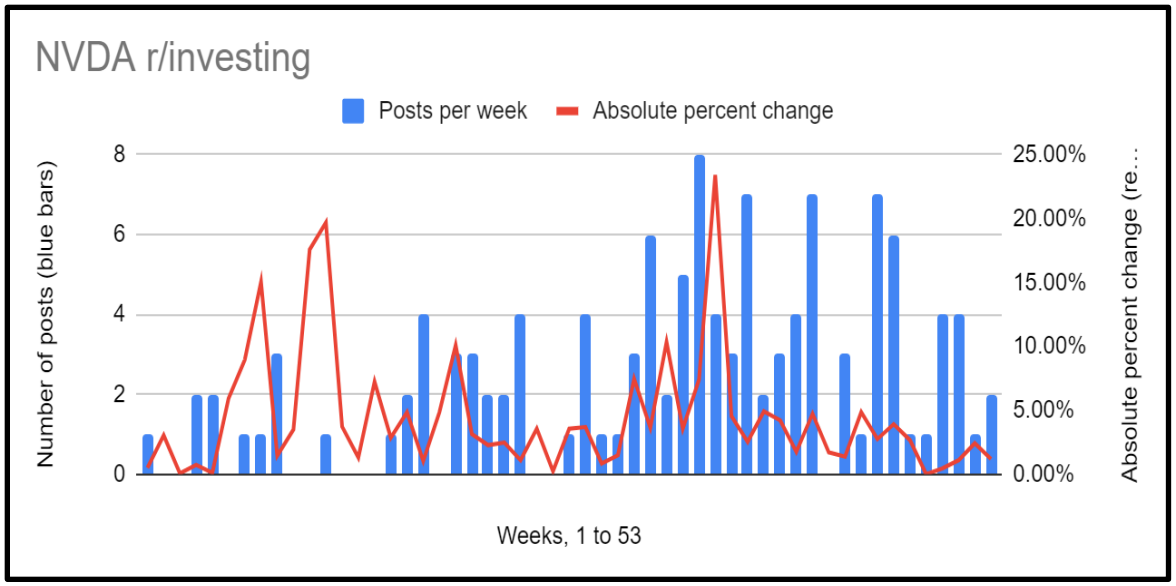

Figure 4. The overlaid graphs of the number of posts per week in 2020 mention NVDA in r/investing and the absolute percent change of the share price of the stock each week.

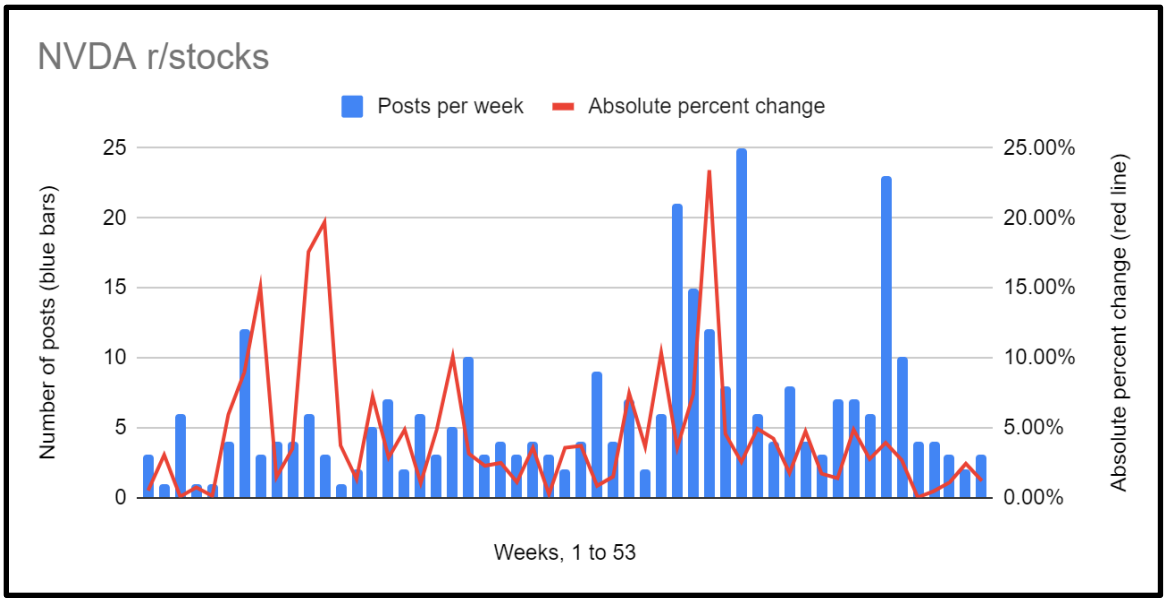

Figure 5. The overlaid graphs of the number of posts per week in 2020 mention NVDA in r/stocks and the absolute percent change of the share price of the stock each week.

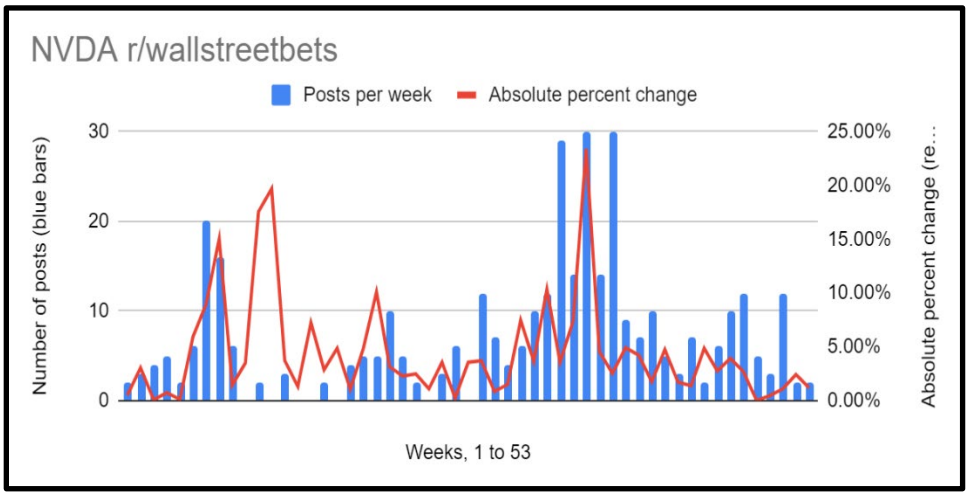

Figure 6. The overlaid graphs of the number of posts per week in 2020 mention NVDA in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

WMT - Walmart

Table 3. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for WMT by Subreddit. The last row displays if there were more correlating results each week for WMT in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	0.1245073183	0.2259080086	0.2952531846
Average number of posts per week	3.037735849	5.641509434	9.301886792
More correlating results per week than not?	Yes	Yes	Yes

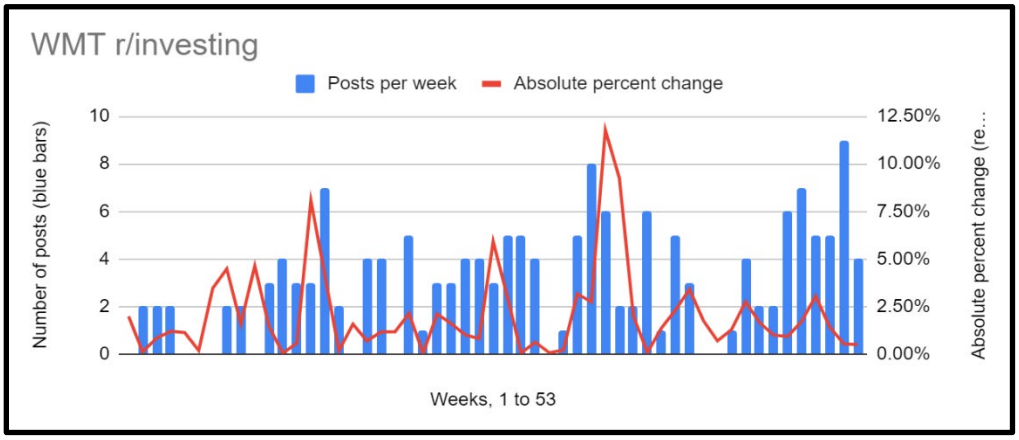

Figure 7. The overlaid graphs of the number of posts per week in 2020 mentioning WMT in r/investing and the absolute percent change of the share price of the stock each week.

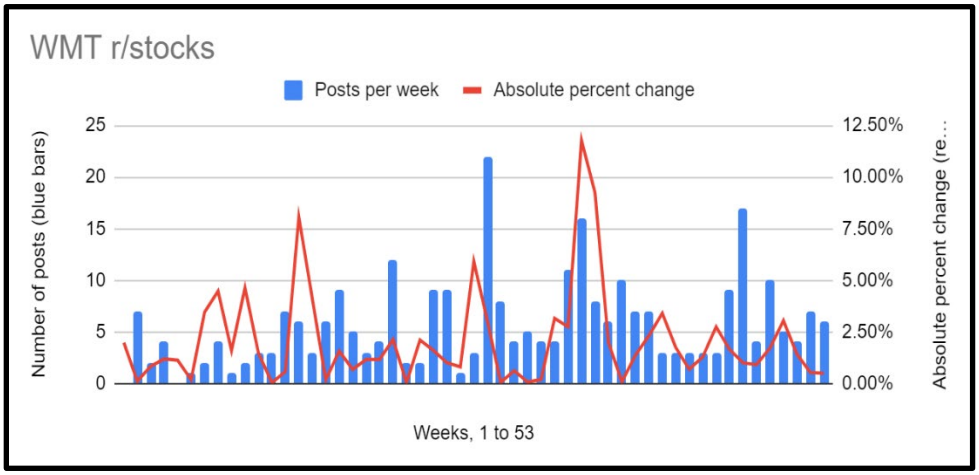

Figure 8. The overlaid graphs of the number of posts per week in 2020 mentioning WMT in r/stocks and the absolute percent change of the share price of the stock each week.

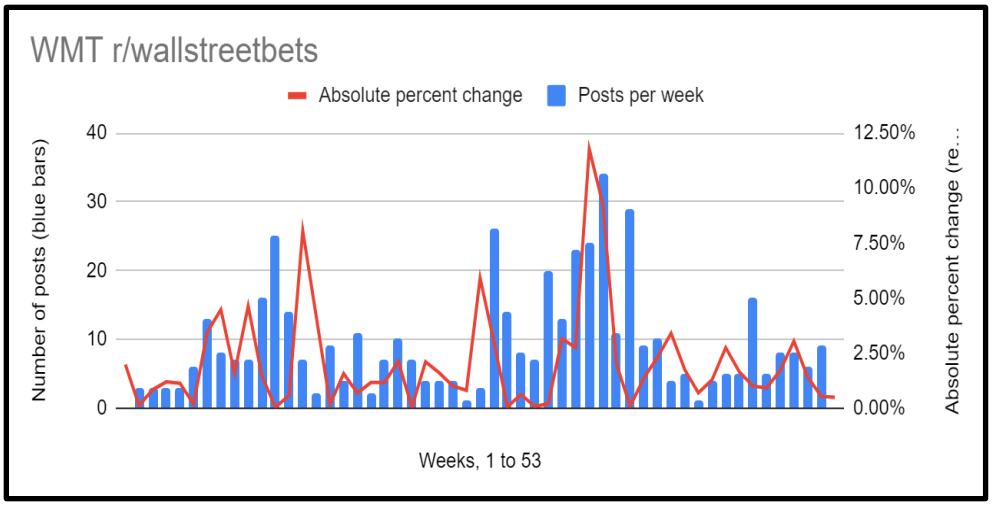

Figure 9. The overlaid graphs of the number of posts per week in 2020 mentioning WMT in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

PFE - Pfizer

Table 4. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for PFE by Subreddit. The last row displays if there were more correlating results each week for PFE in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	Invest	Stock	WSB
Correlation value	0.1076759729	0.2218686261	0.1626731761
Average number of posts per week	2.471698113	6.41509434	9.20754717
More correlating results per week than not?	Yes	Yes	Yes

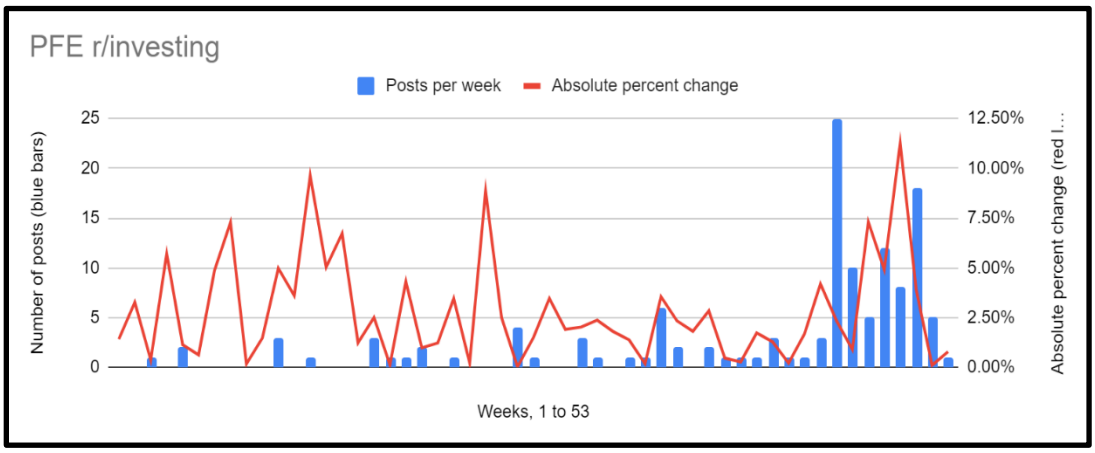

Figure 10. The overlaid graphs of the number of posts per week in 2020 mentioning PFE in r/investing and the absolute percent change of the share price of the stock each week.

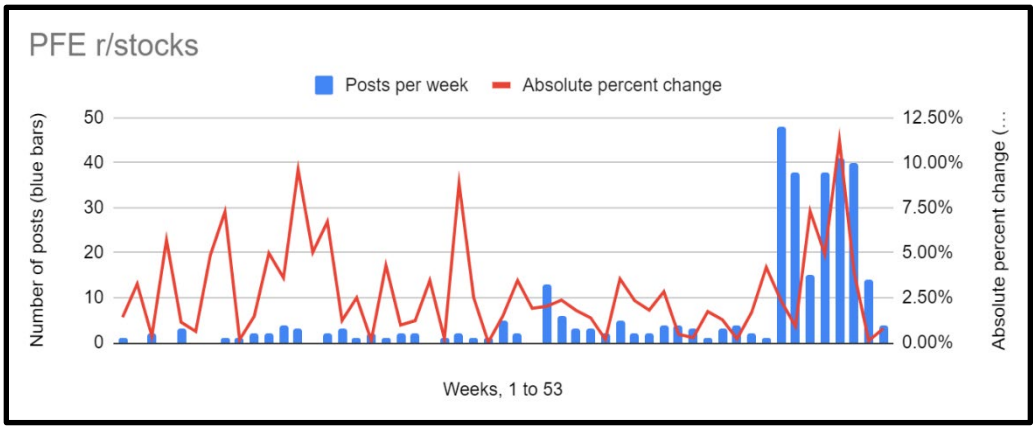

Figure 11. The overlaid graphs of the number of posts per week in 2020 mentioning PFE in r/stocks and the absolute percent change of the share price of the stock each week.

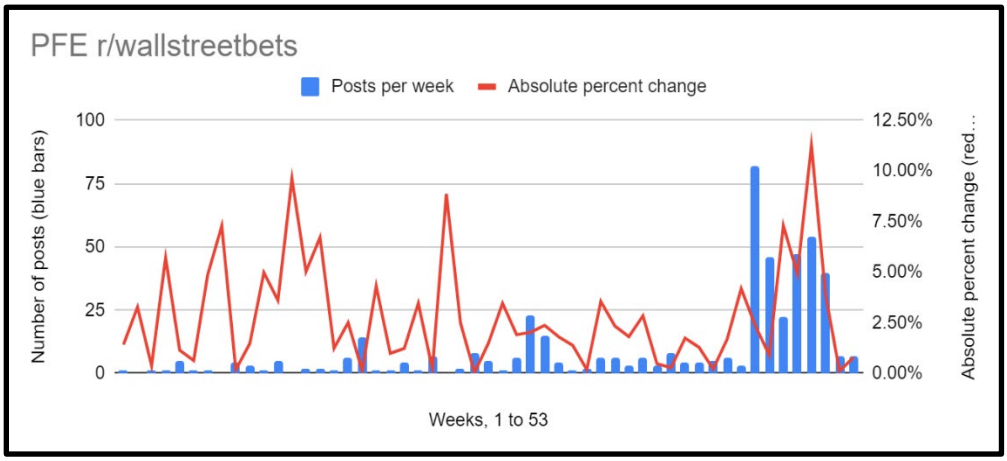

Figure 12. The overlaid graphs of the number of posts per week in 2020 mentioning PFE in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

The inconsistency of posts throughout the year for Pfizer within all Subreddits was likely the result of the COVID-19 Pandemic. As it progressed throughout 2020, large pharmaceutical companies began the development and early testing of vaccines. This would explain the sudden increase in the number of posts mentioning Pfizer towards the end of the year. Once the company gained traction on Reddit, the posts and percent changes began to coincide and correlate more.

XOM - ExxonMobil

Table 5. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for XOM by Subreddit. The last row displays if there were more correlating results each week for XOM in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	0.1043969248	0.121371969	0.1263435499
Average number of posts per week	1.867924528	3.018867925	1.622641509
More correlating results per week than not?	No	Yes	Yes

The lower amount of aligned data is the result of low average post volume mentioning the stock within the r/investing Subreddit. Because of this, the accuracy of the analysis is far less than it would be if there were more posts.

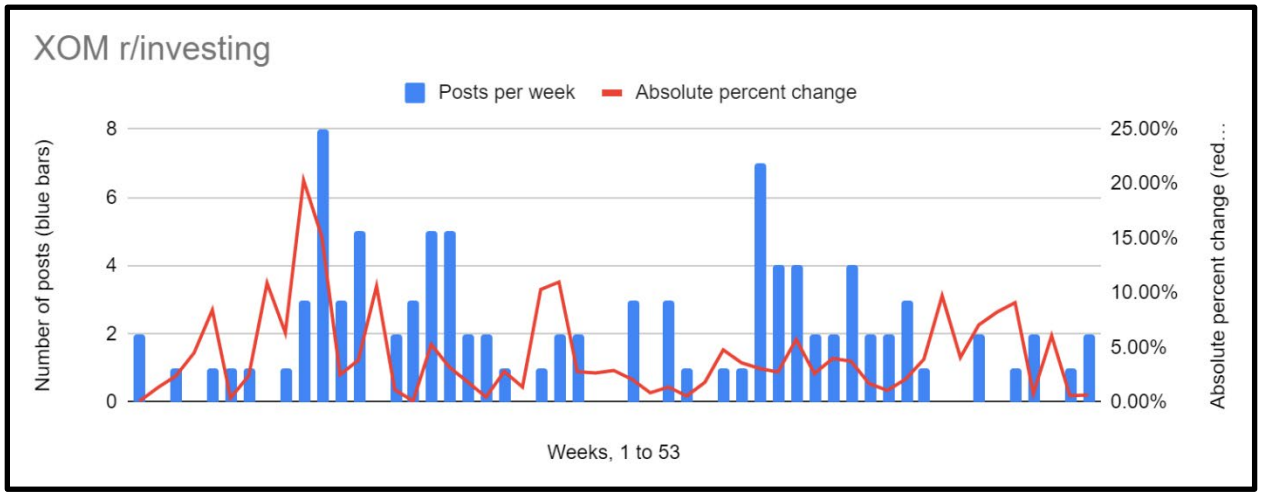

Figure 13. The overlaid graphs of the number of posts per week in 2020 mentioning XOM in r/investing and the absolute percent change of the share price of the stock each week.

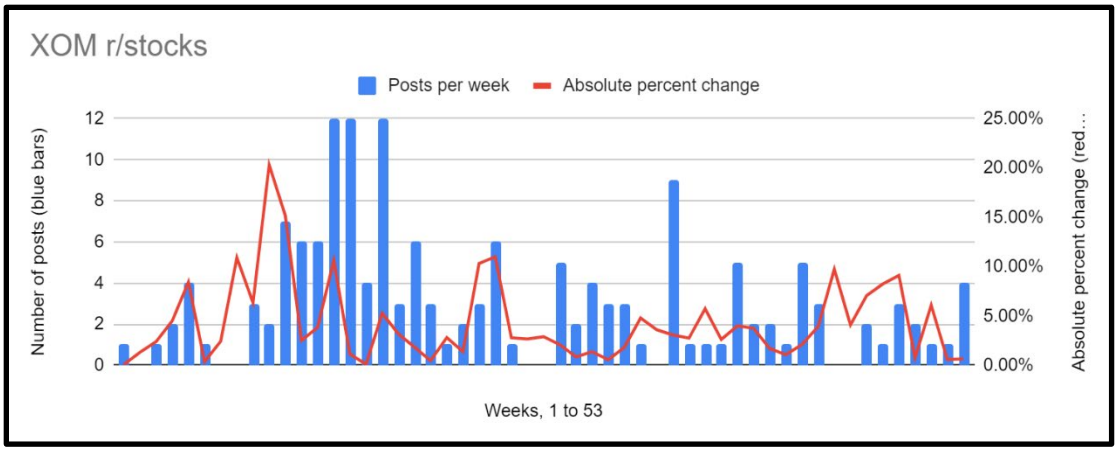

Figure 14. The overlaid graphs of the number of posts per week in 2020 mentioning XOM in r/stocks and the absolute percent change of the share price of the stock each week.

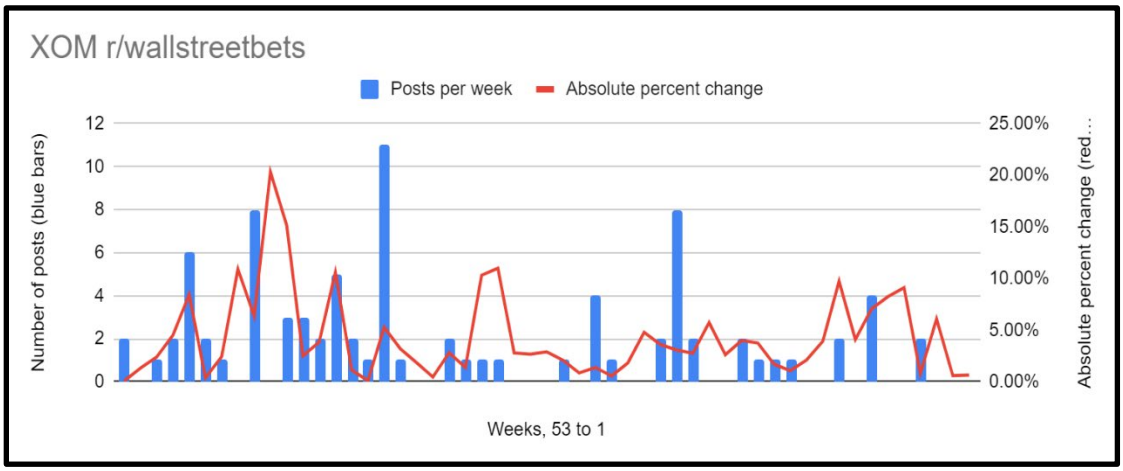

Figure 15. The overlaid graphs of the number of posts per week in 2020 mentioning XOM in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

JPM - JP Morgan Chase

Table 6. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for JPM by Subreddit. The last row displays if there were more correlating results each week for JPM in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	-0.1027386465	0.09174407191	-0.1054564656
Average number of posts per week	1.528301887	4.528301887	5.622641509
More correlating results per week than not?	Yes	Yes	No

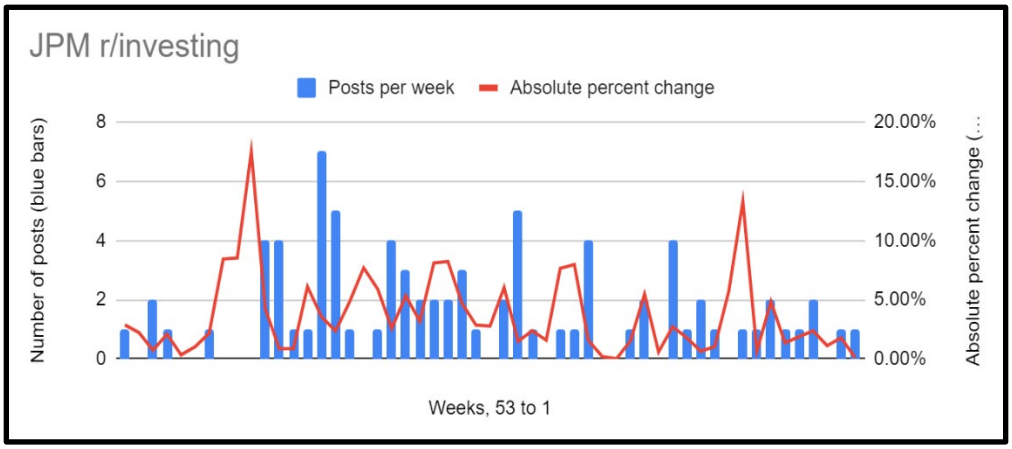

Figure 16. The overlaid graphs of the number of posts per week in 2020 mentioning JPM in r/investing and the absolute percent change of the share price of the stock each week.

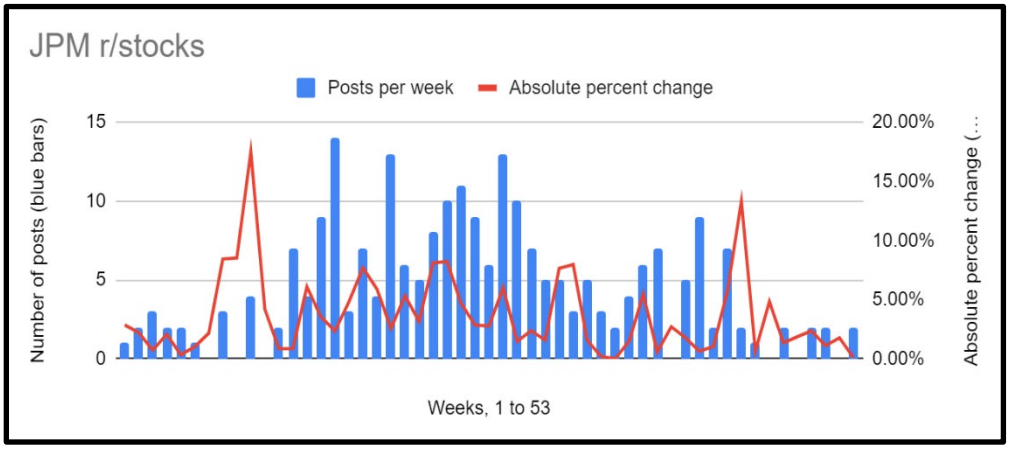

Figure 17. The overlaid graphs of the number of posts per week in 2020 mentioning JPM in r/stocks and the absolute percent change of the share price of the stock each week.

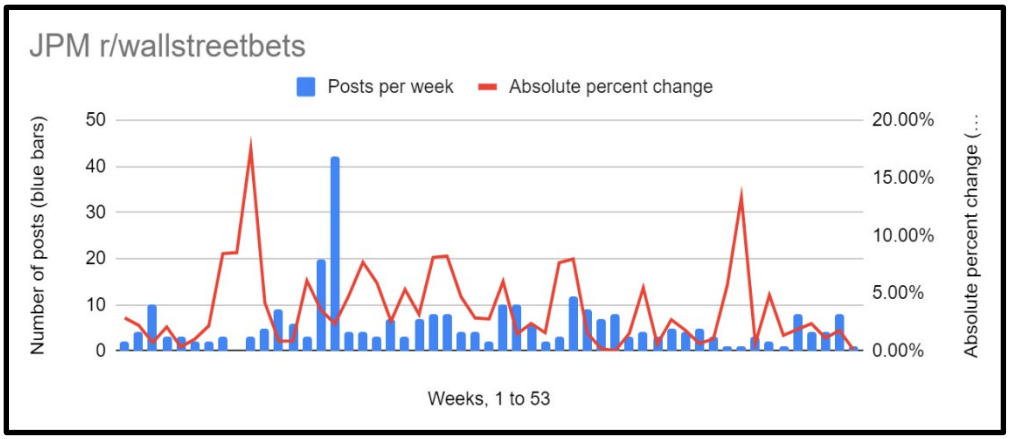

Figure 18. The overlaid graphs of the number of posts per week in 2020 mentioning JPM in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

F - Ford

Table 6. The first two rows in this table display the calculated correlation coefficient using the CORREL function and average number of posts per week for F by Subreddit. The last row displays if there were more correlating results each week for F in each Subreddit. If yes, then there is positive trend alignment. If no, then there is negative trend alignment.

	r/investing	r/stocks	r/wallstreetbets
Correlation value	0.1745632842	0.08897545769	0.3340563898
Average number of posts per week	2.377358491	4.094339623	9.679245283
More correlating results per week than not?	Yes	No	Yes

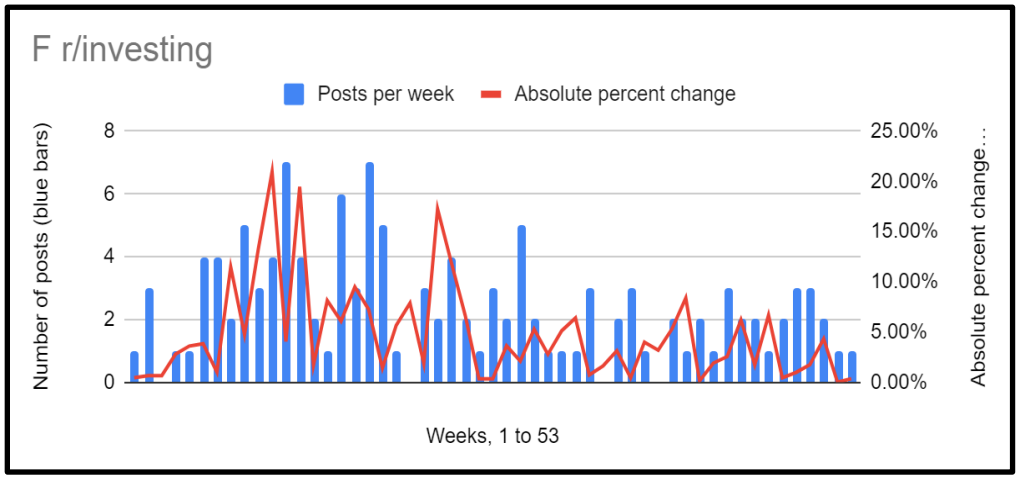

Figure 19. The overlaid graphs of the number of posts per week in 2020 mentioning F in r/investing and the absolute percent change of the share price of the stock each week.

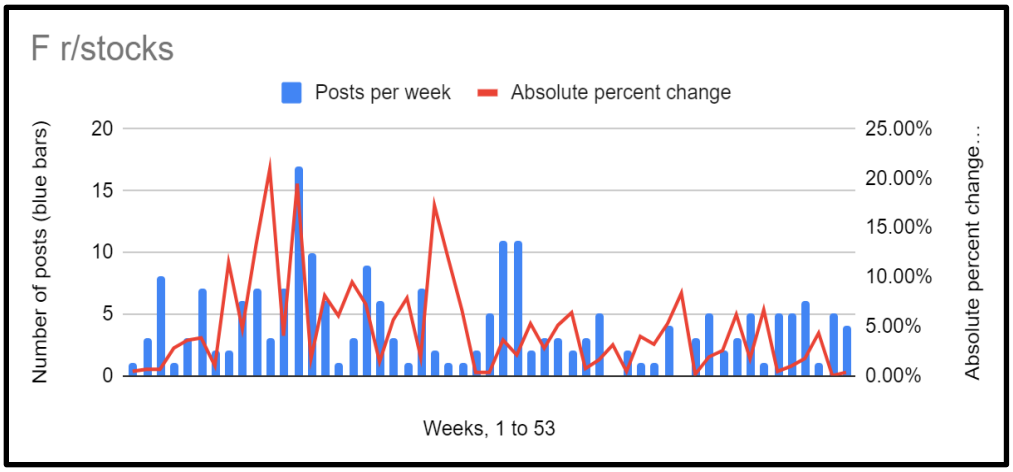

Figure 20. The overlaid graphs of the number of posts per week in 2020 mentioning F in r/stocks and the absolute percent change of the share price of the stock each week.

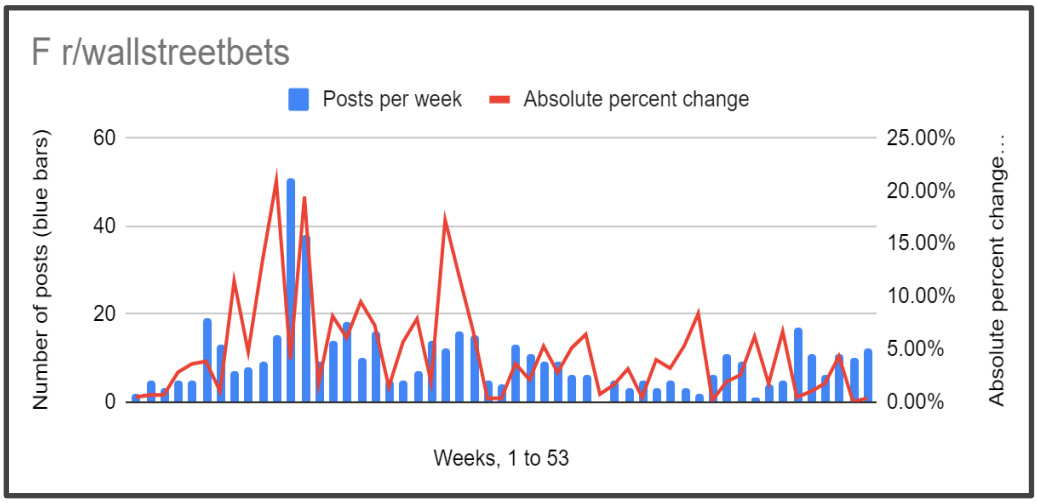

Figure 21. The overlaid graphs of the number of posts per week in 2020 mentioning F in r/wallstreetbets and the absolute percent change of the share price of the stock each week.

Each of the twenty-one graphs had a sudden spike in absolute percent change during the nine to eleven week mark. This coincides with the beginnings of shutdowns due to the COVID-19 Pandemic, and all stocks studied within this paper had major decreases in share price at this time. However, only the graphs of SPY had consistent spikes in Reddit posts that occurred during the same time. This observation must be taken into consideration when conducting correlation comparisons.

Discussion

This investigation reveals that a greater amount of social media traction on Reddit does correlate with greater changes in stock prices. Of the twenty-one intersections of different Subreddits and their corresponding financial data, eighteen had positive correlation values. Eighteen out of the twenty-one also had more posts per week that correlated with the stock price being above or below their respective averages throughout the year. As such, the hypothesis is supported. When comparing the results of each individual stock to the SPY control, certain stocks tend to correlate more. Companies with greater post volume tend to show higher correlation than those with lower post volume. In addition, these

companies are usually brands that consumers interact with more. WMT, NVDA, and F had similar post volume relative to each Subreddit and exhibited higher correlation coefficients. In contrast, JPM, PFE, and XOM had lower post volumes and lower correlation coefficients.

Trends Observed

Two patterns can be visually found within each graph. The first of which is that the maximum in the number of posts per week and the absolute percent change in the stock price occur within two weeks of each other. This pattern is seen with stocks that have greater average post volume and correlation coefficients relative to other companies tested. These include SPY, NVDA, WMT, and F, with only one graph out of twelve, WMT in r/stocks, that did not follow the same trend. If the number of Reddit posts did not peak before the percent change peaked, there was a spike in post volume afterwards within two weeks. This essentially means that a drastic change in a stock's price will always drive social media traction.

However, this does not mean that social media traction is only a result of share price change. Because each of the graphs that meet the earlier requirements follow the same pattern, a drastic change in stock price will happen within two weeks after the week with the highest post volume as long as there was not a spike in stock price change within two weeks before it. This means that a week with extremely high post volume is a potential indicator of share price movement within the following two weeks, provided that the high post volume was not itself driven by a drastic change in share price.

The second pattern can be found within the graphs of PFE, XOM, and JPM. These stocks had lower correlation coefficients and average post volume relative to the other companies tested. Similar to the pattern observed for the more correlating stocks, there seems to be a certain time frame in which these events occur.

The trend observed among these stocks is that the week with the highest post volume and the week with the greatest absolute percent change occur within four to six weeks to each other. For XOM and JPM, the share price changes before the post volume increases. The opposite is true for PFE. Once again, there is one exception to this pattern in the graph of XOM r/investing, where the two events occur within one week of each other. While this trend might seem nearly identical to the trend found in SPY, NVDA, WMT, and F, it is not. This is because there is no consistent result across all nine graphs. Only the graphs of XOM and JPM show a consistent four to six week delay between peak absolute percent change and peak post volume. Additionally, the graphs for PFE show that the week with high post volume occurs before absolute share price change reaches its highest point. As such, while it can be said that the weeks where the highest points in the graphs of post volume and absolute percent change occur within six weeks of each other, there are mixed results when attempting to determine whether social media traction or stock price change occurs first.

As a whole, in every graph the greatest absolute percent change in stock price occurred either within two weeks or between four and six weeks from the time with the greatest post volume on Reddit. In both cases, the stock price change comes before social media traction two-thirds of the time. It is safe to conclude that significant changes in the share price of a stock drives social media traction. However, for stocks with greater correlation–SPY, NVDA, WMT, and F–stock price change could occur within two weeks after the week with the greatest post volume if the peak in post volume was not caused by a drastic change in stock price. This means there potentially could be a way to predict major stock price changes by looking at Reddit post volume.

Impacts

This data shows the relevance of social media and how it represents consumer sentiment towards the market. It reveals that large changes in share price cause individuals to discuss the stock more, increasing the number of people who know about the certain stock. Investors can also use this data to estimate the volatility of a certain stock. If the number of posts regarding a specific stock remains relatively constant throughout a period of time, it is likely to be less volatile

and as such hold less risk. The opposite is also true. If an investor was looking to build a more growth-oriented portfolio, they may consider stocks which have greater fluctuations of social media attention.

Additionally, as mentioned before, significant changes in post volume could potentially be an indicator of a major change in share price. Since the weeks with greatest absolute percent change and post volume have to happen within two weeks of each other in the graphs of SPY, NVDA, WMT, and F, the occurrence of a peak in post volume must result in a major change in share price within two weeks after it. This pattern has fairly impactful ramifications. It could be a significant indicator of when a certain stock might have a ten to twenty-five percent change in value. This is extremely important information for investors and could have a major impact on the short term volatility of a stock. Further research would be necessary to determine if this trend persists with other stocks and other time frames.

Existing Literature

The results of this paper align with the findings of similar research. One Twitter-based study found that online sentiment regarding a certain stock became more or less positive depending on its performance (Coelho et al 2019). Another study found that online rumors had a short-term effect on stock market volatility within the Chinese stock market (Zhang et al., 2022). An additional study has proven that social media sentiment can be used to accurately predict short-term stock price volatility (Wu et al., 2017). This is in agreement with the potential impacts the results of this paper suggest.

Limitations/Next Steps

There is one major issue to be addressed in regards to gathering data within this paper. As this study relied on a third-party API scraper to gather Reddit data, there were limitations in regards to what data could be used. In addition, Reddit itself permanently disabled third-party application programming interface access to the site in the latter half of April 2023 (Fung 2023). This effectively stopped any data being gathered after that point.

In the future, a new method should be used to scrape data. While it would be initially challenging to get around the blocks Reddit has put on such activity, a new scraper would be able to gather more data across a longer period of time within even more Subreddits. In this paper, the data sampled was in weekly intervals within only the time frame of one year. In future research, data should also be assessed at daily frequencies for greater precision. This would allow for more accurate representations of social media trends and the activity of traders. Additionally, multiple companies could be represented in each industry to potentially observe any patterns on the industry level.

The data analysis within this paper could also be enhanced with more professional, precise analysis tools and formulas. Because all analysis occurred in Google Sheets, this paper could only draw conclusions from more primitive functions built into the software. In future research, more advanced methods to quantify correlation should be implemented to provide a deeper analysis of the data.

Conclusion

This experiment shows that social media traction on a platform such as Reddit does correspond to changes in share price of specific stocks. This means that the hypothesis is supported. The majority of the tests ran in this paper prove that social media traction on Reddit is a result of significant change in share price of a stock. Additionally, in certain instances, high post volume could be a potential indicator of a major change in stock price. However, it cannot be proven that increased post volume directly causes these changes in share price. The first result is relevant in that social media traction reflects the attention of traders on the market. This reflection of the market can be used as a qualitative data point about the potential for change a particular stock has. The second observation could be used to predict short-

term stock volatility. The increasing number of investors on social media in the future will only increase the relevance and impact that social media has on the stock market.

Acknowledgements

I would like to thank my research mentor David Tong for guiding me through each and every single part of writing this paper. As I had minimal experience in conducting original research prior to this paper, he was able to share his expertise with me and lead me through the process. I would also like to thank Dr. Daniel Plymire for critical assistance in preparing this paper for publication. Dr. Plymire was instrumental in making my work more professional and truly brought it to another level. This paper would not be possible without the support and knowledge of these individuals, and I am extremely grateful for their help.

References

Camas Unddit [Computer program]. (n.d.). https://camas.unddit.com/

Coelho, J., D'almeida, D., Coyne, S., Gilkerson, N., Mills, K., & Madiraju, P. (2019). Social media and forecasting stock price change. *IEEE 43rd Annual Computer Software and Applications Conference*, *2*, 195-200. https://doi.ieeecomputersociety.org/10.1109/COMPSAC.2019.10206

CORREL function. (n.d.). Microsoft. https://support.microsoft.com/en-au/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92#:~:text=The%20CORREL%20function%20returns%20the,the%20use%20of%20air%20conditioners.

Fung, B. (2023, June 1). *Reddit sparks outrage after a popular app developer said it wants him to pay $20 million a year for data access*. CNN. https://www.cnn.com/2023/06/01/tech/reddit-outrage-data-access-charge/index.html

Lush, M., Fontes, A., Zhu, M., Valdes, O., & Mottola, G. (2021, February). *Investing 2020: New accounts and the people who opened them.* FINRA Foundation. https://www.finrafoundation.org/sites/finrafoundation/files/investing-2020-new-accounts-and-the-people-who-opened-them_1_0.pdf

Malz, A. M. (2021). The gamestop episode: What happened and what does it mean? *Cato Journal.* https://www.cato.org/cato-journal/fall-2021/gamestop-episode-what-happened-what-does-it-mean

Octoparse (Version 8.6.4) [Computer software]. (2016). https://www.octoparse.com/

Smith, S., & O'Hare, A. (2022). Comparing traditional news and social media with stock price movements; Which comes first, the news or the price change? *Journal of Big Data*, *47*(9). https://doi.org/10.1186/s40537-022-00591-6

Social media increases teenage interest in wall street, wells fargo survey finds. (2021, June 2). Wells Fargo. https://newsroom.wf.com/English/news-releases/news-release-details/2021/Social-Media-Increases-Teenage-Interest-in-Wall-Street-Wells-Fargo-Survey-Finds/default.aspx

Vogels, E. A., Gelles-Watnick, R., & Massarat, N. (2022, August 10). *Teens, social media and technology 2022*. Pew Research Center. https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

Wu, X., Wang, X., Ma, S., & Ye, Q. (2017). The influence of social media on stock volatility. *Frontiers of Engineering Management*, *4*(2), 201-211. https://doi.org/10.15302/J-FEM-2017018

Zhang, H., Chen, Y., Rong, W., Wang, J., & Tan, J. (2022). Effect of social media rumors on stock market volatility: A case of data mining in China. *Frontiers in Physics*, *10*. https://doi.org/10.3389/fphy.2022.987799