Reddit's Influence on Stock Price

Section 1 Group 6

Khristion Lambert, Shijun Wei, Zhe Wang

Executive Summary

This report provides a comprehensive breakdown and analysis of the ripple effect that the Wallstreetbets subreddit caused in the stock market. The research primarily focuses on how Gamestop (GME) along with many other companies were engaging in a short sell of their stock and after many people on reddit caught wind of this they began to all buy into this market (short squeeze), coincidentally pumping the price of stocks to astronomical high values. To uncover what truly happened during time, research was conducted. In order to create some insights into the data we conducted several analyses via web scraping of the Wallstreetbets subreddit and Yahoo finance, sentiment analysis, and correlation/comparison techniques to establish relationships between comments featured in Wallstreetbets and the behavior of stock prices daily from January 1, 2020 - March 14, 2020. Results of the data analysis showed very shocking positive relationships between the number of comments mentioned and the price of the stocks as well as the sentiment of each comment and the stock performance. With our pipeline, people can collect the comments from Reddit, analyze the sentiment of the comments, and use the results as a guide of their trading actions.

Introduction

In an era where technology and social media hold an integral part of our society we could understand the potential influence that comes from it. An open discussion forum known as Reddit has quickly dominated the early part of 2021. Most users on Reddit have the ability to submit content such as links, text posts, and images that are voted on by other

members based on how good or bad they are. Part of the architecture that is involved on the website is the element of communities based on certain subject areas. These are referred to as "subreddits" and some topics that cover these subreddits are things like sports, finance, technology, education, etc. A finance focused subreddit, Wallstreetbets, has recently been gaining a lot of traction. In January 2021 members of this channel engaged in a strategy called a short squeeze that American video game retailer Gamestop (GME) and other securities fell victim to. At this time almost 140% of Gamestop's public shares had been short sold, and the rush to buy shares that covered these positions rose significantly causing the price to rise ever further. To put it in perspective at its peak, January 28, the short squeeze caused Gamestop's stock price to reach a pre-market value of over $500 per share, nearly 30 times the $17.25 share price at the beginning of the month. During this time the price of other heavily shorted securities also increased. Due to this unforeseen circumstance, people on Reddit and other social media platforms went crazy, information spread like wildfire causing many old and new investors to join in on the stock market game. As analysts we wanted to see if there were general trends within the sentiment of comments being talked about and the volume of stock purchases, as well as other significant variables relevant to this phenomenon.

Exploratory Data Analysis

To better understand the data we collected using web-scraping from the Internet, we did EDA initially. The time range we chose is when the GME and other popular stocks on Reddit had short squeezes. We chose numeric variables of comments as well as the

stock performance in that period and tried to find correlations between each other. For the comments, the total number of comments, the total score of comments are used.
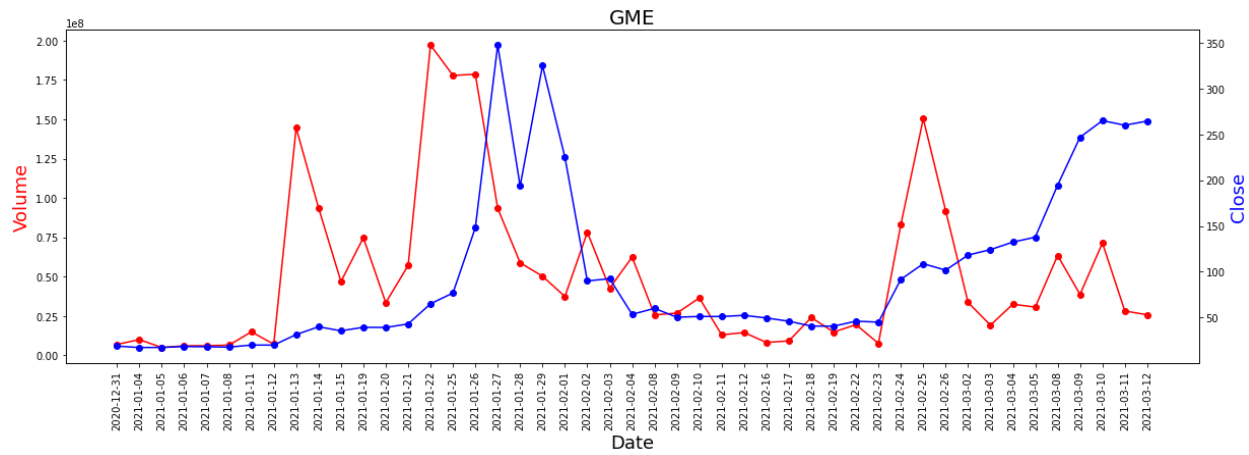

Figure.1 The comparison between closing price and trading volume of GME.

For the stock performance, we chose the close price and trading volume of each day. As we can see in Figure 1,for the stock GME, the blue line represents the closing price and the red line represents the trading volume since Jan. 1st 2021. The performance of stock can be divided into two rounds. The first round is from Jan. 1st to Feb. 4th, there is a massive short squeeze. The short squeeze breakout happened on Jan.12th. The top is between Jan. 26th to Feb.1st where the stock price is rocket high. The end of the short squeeze is between Feb 2ed to Feb 4th where the stock price drops dramatically and back to the bottom. After about 2 weeks，the stock price popped again for a second round from Feb 23th until the most recent trading day. The surge of trading volume happened before the surge of stock price for both rounds.
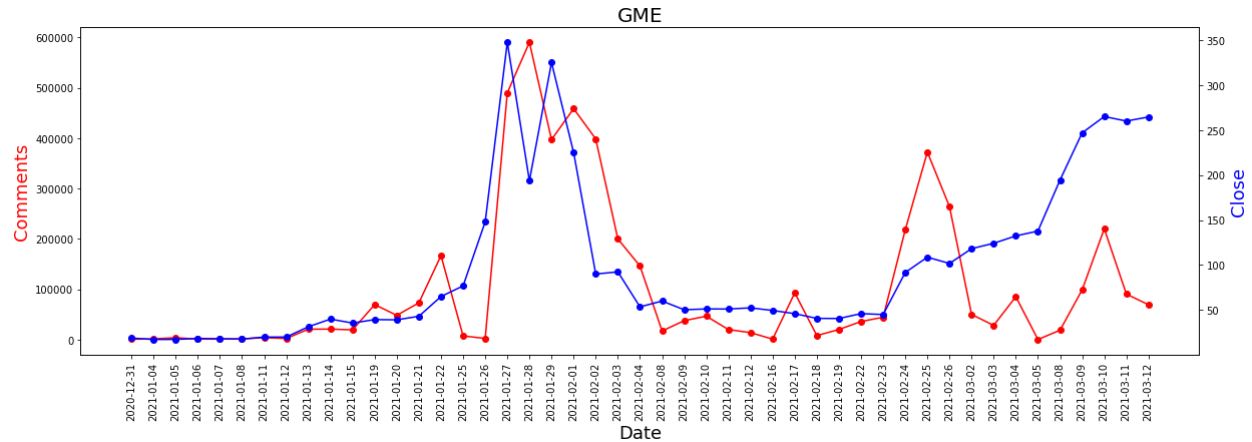
Figure.2 The comparison between total comments and closing price of GME.

As demonstrated in Figure.2, the total comments and the stock price showed a spectacular synchronism. It is really hard to believe that the behaviors of the individual investors on reddit has a lot of connection with the stock price performance. Even during the beginning of the second round, the total comments and the stock price increased in the same direction. But we cannot make a conclusion about the causality between the performance of stock price and the comments on Reddit.
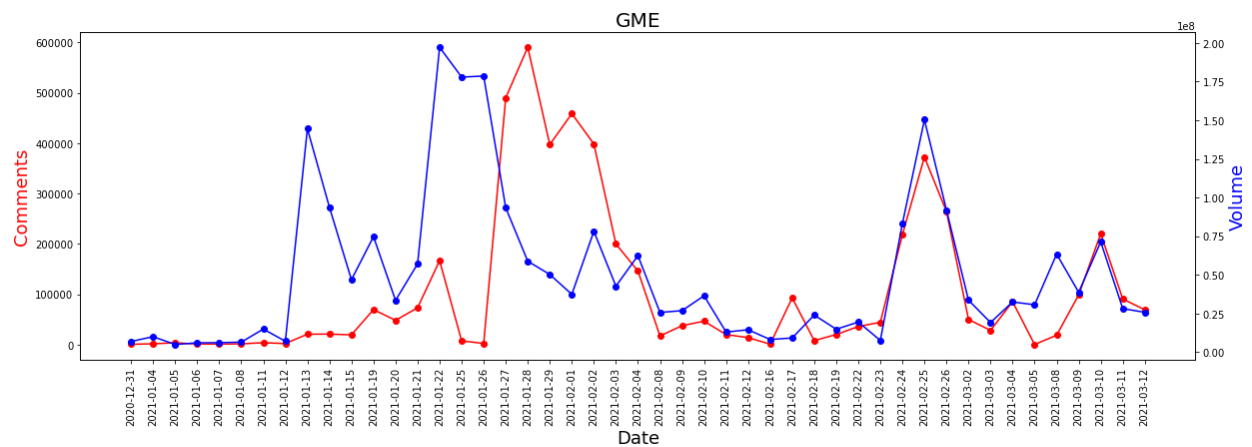

Figure.3 The comparison between total comments and trading volume of GME.

In FIgure.3, the comments and the trading volume shows two patterns for each round. In the first round, the trading volume exploded before the surge of the comments on Reddit. While in the second round, the comments and the trading volume are synchronized.
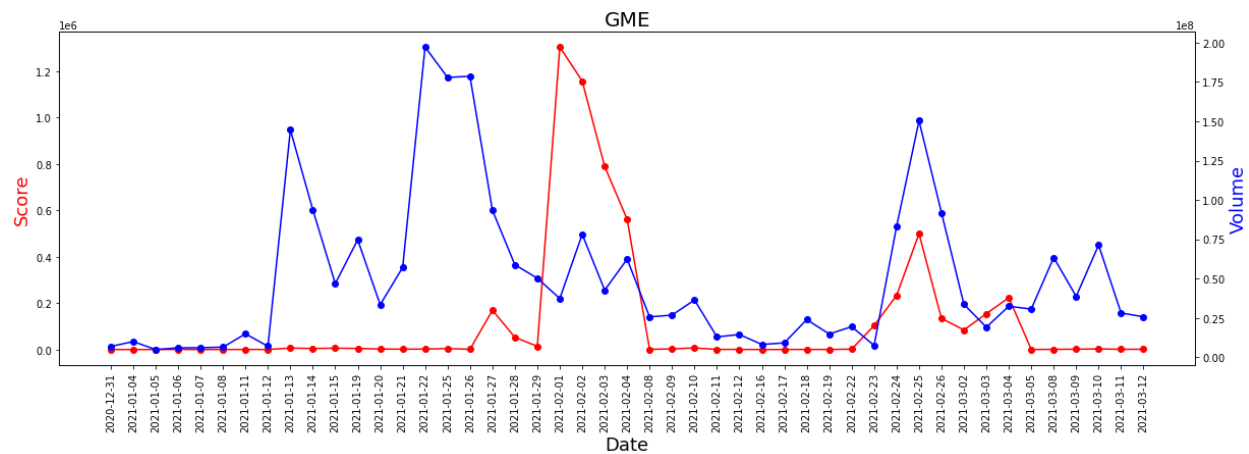

Figure.4 The comparison between total score of the comments and closing price of GME.

In Figure.4 , the correlation between the total score of comments and the closing price are also compared, but the correlation is weaker than the total comments.

Wordcloud

We used all the collected comments and built a word cloud. As shown in Figure 5, the result is expected. Most members in the wallstreetbets are optimistic about the performance of GME. What is noticeable is that the AMC is frequently mentioned in the comments of GME.

Figure 5. The word cloud of all the comments about GME

Sentiment Analysis

In the next step, we took a deeper look into the comments from Wallstreetbets that reflect what people think of GME over time. To do this, we decided to incorporate natural language processing techniques on those comments. In general we want to explore whether people's emotional fluctuations, or the changes of their attitude towards GME will have correlation with performances of GME stock. To interpret this, our assumption is that while people are optimistic about GME, the stock should perform better with an increase in price; on the other hand, if people mainly post negative comments on reddit, it should impact the performance and decrease its price.

The method we used is called *SentimentIntensityAnalyzer* from the nltk package. The package allowed us to quickly measure the compound sentiment score of each comment in each day and classify them to be positive if the score is larger than 0, negative if the

score is less than 0, and neutral otherwise. After running the analyzer we obtained a table with each comment labelled a sentiment score. (Figure. 8)

	Title	Publish Date	sentiments	compound
0	gme to 420 69 but only if we make it happen dd...	2020-12-31	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	0.0000
1	what would make gme shorts win	2020-12-31	{'neg': 0.0, 'neu': 0.568, 'pos': 0.432, 'comp...	0.5859
2	not sure how reliablerandom comment in cohens ...	2020-12-31	{'neg': 0.305, 'neu': 0.564, 'pos': 0.131, 'co...	-0.6521
3	how have we been so fucking blind gme is liter...	2020-12-31	{'neg': 0.229, 'neu': 0.771, 'pos': 0.0, 'comp...	-0.5059
4	gme is the rockets	2020-12-31	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	0.0000

Figure.6 The table with compound sentiment scores for each comment

The first analysis we did was the comparison of the number of positive and negative comments on a daily basis. Through this chart, we wanted to observe whether the number of positive comments would largely exceed the number of negative comments when GME revealed a good performance. However, after we plot the time series plot of the ratio comparison between positive and negative comments, we could not observe a clear pattern. (Figure. 9) This might be caused by the changing number of neutral comments that the algorithm detected and dropped from our table.
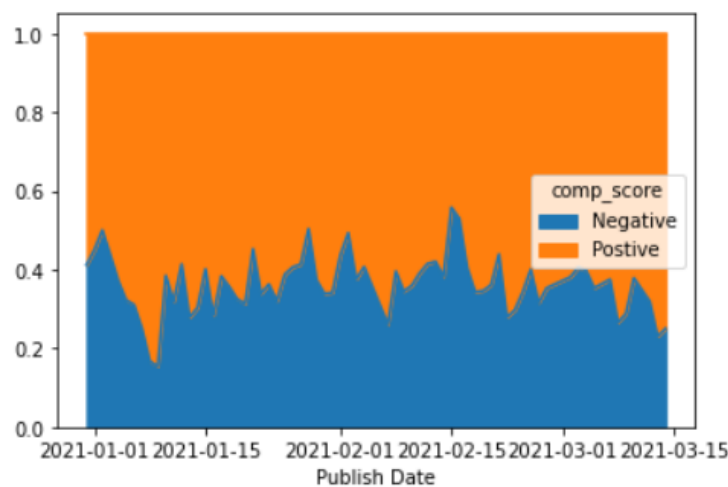

Figure.7 The ratio comparison between number of positive vs negative comments

Our second analysis focused on comparing the fluctuation of emotions of people with GME stock performances. Different people have different thoughts about GME in one day, so we choose to calculate the sum of compound sentiment in each day to present the general sentiments via grouping by dates. The first time series graph we plotted is the comparison between GME stock volume and compound sentiments.(Figure. 9) This figure demonstrates an approximately matching trend for people's excitement with GME continuously rising trading volumes. We can observe a delayed reaction of compound to GME volume change. In general it shows a good indication of the existence of correlation between those two trends.
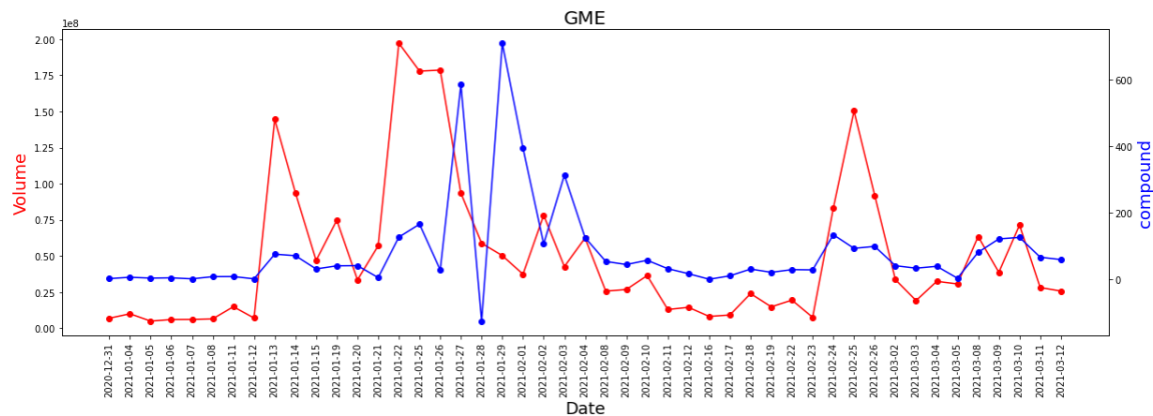

Figure.8 The comparison between trading volume of GME and people's sentiments

Another time series graph we plotted is trading's close price versus comments sentiment score. In this graph we exchanged stock volume to stock close price and observed an incredibly matching trend between them. (Figure. 10) From those graphs we can recognize an existing correlation between the performance of GME and sentiment of

people from the past three months by eyeballing. So our model indicates that the emotion of comments from people on Reddit can partially reflect the real performance of GME stocks. To achieve business value, this model can be used as a data-driven method to guide people to buy or sell stocks based on sentiments of comments on Wallstreetbets.
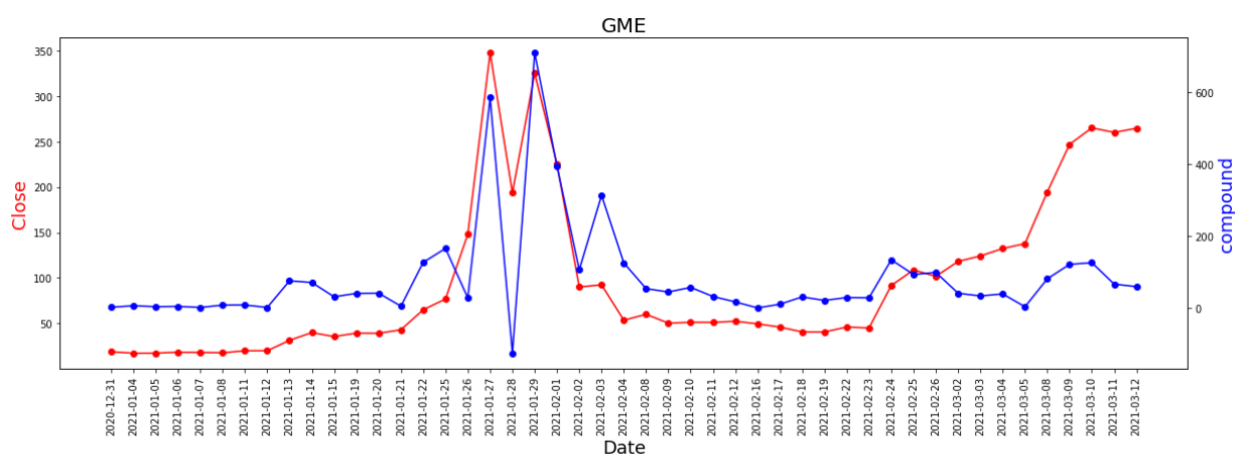

Figure.9 The comparison between trading price of GME and people's sentiments

Conclusion

Based on the figures shown above, we made some speculations. In the first round, there were very few investors who believed that there would be a short squeeze happen and their beliefs paid back. After the stock price surged, GME became popular among Reddit which aroused more people to buy. When the stock plunged, the number of comments plunged as well. Since people who followed others to buy were reluctant to accept the truth. In the second round, the comments show amazing synchron with the trading volume which suggests that it is started by the individual investors on Reddit. No wonder there are some media called: it is the triumph of the mob.

Our model can be a great indicator to investors since the influence of Reddit has been increased due to this special short squeeze. No matter if you are an individual investor or

institutional investor，keeping an eye on the comments in Reddit is a must from now on especially when your positions are frequently mentioned in Reddit. Our model can also be an important reference for trading as well.