

THESIS WORK

Túú András 2022



The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

Tűű András

BA in Business and
Management

2022

Thesis Supervisor: Tamásné Vőneki Zsuzsanna

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

I. Introduction	1
1.1 What are the “meme stocks”	1
1.2 Aim of the thesis	1
II. Theoretical background.....	3
2.1 Timeline	3
2.1.2 The second phase:	4
2.1.3 The third phase:.....	4
2.2 Stockmarket	6
2.2.3 Economic effects of COVID-19	6
2.2.4 Returns in largest indices	7
2.2.5 Market sentiment in late 2020-2021	8
2.3 Stock Market trading.....	10
2.3.1 Introduction to stock market trading.....	10
2.3.2 Options trading.....	10
2.3.3 Gamma squeeze	10
2.3.4 Short squeeze	11
III. Methodological background	13
3.2 Introduction to the analysis of the Reddit thread	13
3.2.1 API and web scraping	13
3.2.2 Sentiment analysis	14
3.2.3 Regression analysis methodology	16
3.2.4 Analysis of descriptive statistics of variables	17
3.3.4.1 Descriptive analysis of original data.....	17
3.3.4.1.1 Text2Emotion descriptive statistics	17
3.3.4.1.2 VADER descriptive statistics	18
3.3.4.2 Descriptive analysis of modified data.....	18
3.3.4.2.1 Text2Emotion descriptive statistics	18
3.3.4.2.2 VADER descriptive statistics	19
3.3.4.3 Financial and other data descriptive analysis	19
3.2.5 Hypotheses	21
IV. Analysis	22
4.1 Fundamental analysis	22
4.1.1 Gamestop (GME).....	22
4.1.1.1 Introduction.....	22
4.1.1.2 Sales Breakdown by Product Category	22
4.1.1.3 Operating margin	23
4.1.1.4 The turnaround plan.....	24

4.1.1.5 Bottom line	24
4.1.1.6 Is the price justified?	24
4.1.2 AMC	25
4.1.2.1 Introduction.....	25
4.1.2.2 Revenue	25
4.1.2.3 Debt.....	25
4.1.2.4 Valuation.....	25
4.1.2.5 Bottom line	26
4.2 Analysis of Sentiment	28
4.2.1 VADER(Original).....	28
4.2.2 VADER(modified).....	29
4.2.3 Text2Emotion (Original)	30
4.2.4 Text2Emotion (Modified).....	31
4.3 GME analysis.....	33
4.3.1 Volume, short interest, mentions	33
4.3.2 Sentiment regression analysis	34
4.3.2.1 VADER sentiment regression analysis.....	34
4.3.2.2 T2E sentiment regression analysis.....	35
4.4 AMC analysis.....	36
4.4.1 Volume, short interest, mentions	36
4.4.2 Sentiment regression analysis	37
4.4.2.1 VADER sentiment regression analysis.....	37
4.4.2.2 T2E sentiment regression analysis.....	38
4.5 Hypothesis testing	39
4.6 Bidirectional relationships	40
4.7 Limitations of the research.....	40
V. Summary	41
VI. References:	44
VII. List of figures	46
VIII. Appendix.....	48
8.1 Descriptive statistics for original data.....	48
8.2 Descriptive statistics for modified data.....	48

I. Introduction

1.1 What are the “meme stocks”

The topic of this thesis is the internet and stockmarket phenomenon of 2021, the so-called meme stocks. I have been and am closely connected to finance, new trends and technologies that are relatively new to the market and experience large increases or decreases. For example different cryptocurrencies and blockchain technologies, also the center point of my analysis the Gamestop and AMC stocks have experienced rarely seen concepts and movements.

At the start of 2021 several stocks emerged with extreme volatility and multiplied their value in days, these so-called “meme stocks” were Gamestop (GME), AMC (AMC), Blackberry (BB), and Nokia (NOK). These corporations are from various industries but were all picked up by a group on Reddit, an online forum called Wallstreetbets. The idea behind the group buying these stocks was the short-squeeze, driving up the value of overly shorted stocks and this way forcing the shorts to close. All of these companies were shorted for various reasons, bad financials, outdated business models (Gamestop), lack of innovation in a rapidly advancing technological sector (Blackberry, Nokia), or the impact of the global COVID-19 pandemic in the case of AMC. The biggest phenomenon out of these was the Gamestop and AMC stock, which I will analyze much deeper in the following parts of my thesis work.

1.2 Aim of the thesis

I will try to find out in my analysis, the part that the online platform Reddit and the retail investors using the platform took in the events. While, I will approach the events from various angles, including a fundamental analysis conducted on both of these stocks. Trying to find the reason that the stock prices increased in the early months of 2021.

I will further analyze the stocks mentioned, based on their performance during the pandemic, compared to the timeframe’s stock market movements. For a better understanding of these events the

Furthermore, I will conduct a deeper analysis of the Reddit thread and the sentiment of the wallstreetbets group using a variety of computer software during my search. I will use web scraping and different text and sentence analyzing packages, that can extract the sentiment on a large scale of the posts. I will analyze the possible relationship between the sentiment of the Wallstreetbets thread with the price of both the GME and AMC share prices.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

My research questions are the following:

- Did the two most popular meme stocks experience these extreme movements due to fundamental value or due to a short-squeeze?
- Did the sentiment of the Reddit thread Wallstreetbets influence the rise and fall of the stock prices?

In the following theoretical part, I will analyze the financial environment, and the current trends in global stock markets focused on the influences of the global pandemic.

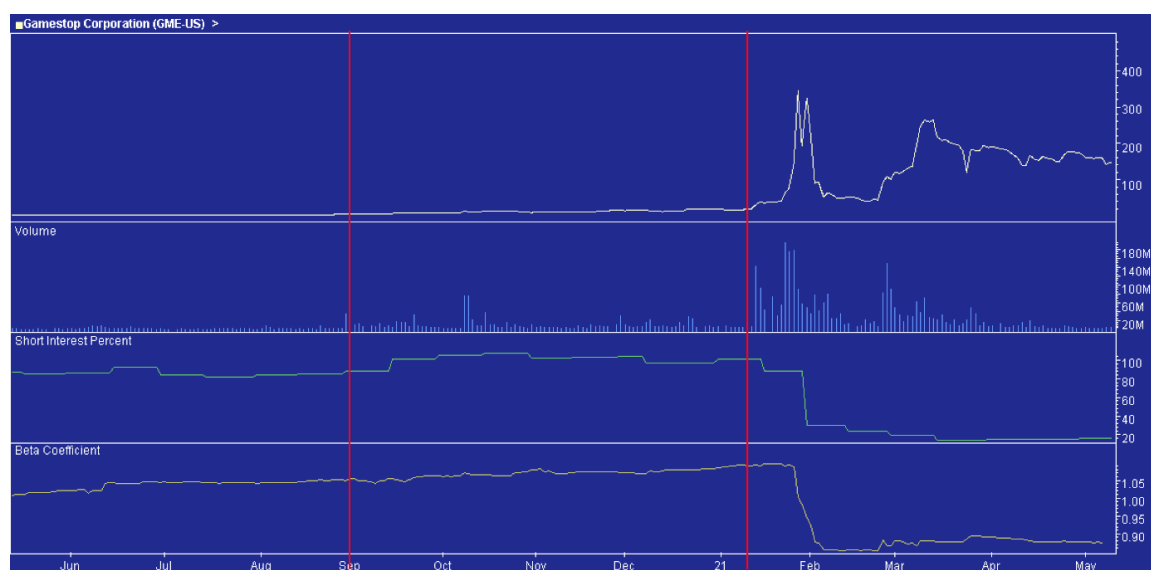
II. Theoretical background

2.1 Timeline

The timeline of the topic takes up approximately $\frac{3}{4}$ of a year starting from August/September. The timeline could be separated into 3 different parts. The first part is the beginning, with a slow, but steady increase. The first “squeeze” meaning the period in January when the stock moved with extreme volatility, the majority of the short positions were closed at the end of this stage. And the third phase was with high volatility in March, followed by a steady decline.

2.1.1 The first phase:

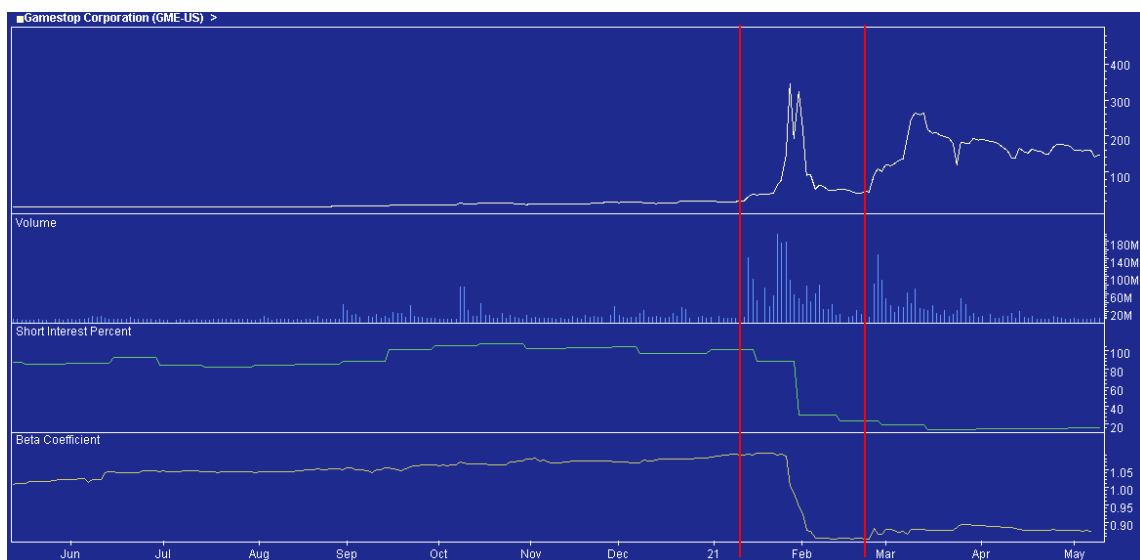
As it is visible on the 1. Figure, at the end of August the stock price for Gamestop started steadily growing starting from 5 USD per share, reaching 10,5 USD by the end of September, as a result of the earnings report on 09/09/2021, reporting that the Q2 loss of the company was 1,71 USD per share in contrast to the Q1 report when the loss per share was 2,57 USD. In October the share price stagnated, being traded at 10,47 USD at the end of October. At the end of November, the stock closed at 16,56 USD. In December the company published the Q3 earnings report, with a 0,29 USD loss per share, the stock price closed at 18,8 USD at the end of December. The stock got picked up by Redditors in mid-January, as a member of the forum said that the stock is underpriced and has a very high short interest. The short interest over the period increased from 89% at the end of August to 102% on January 13th.



1. Figure Gamestop price, volume, short interest, Beta chart 2021 January-2021 March Source: Thomson One

2.1.2 The second phase:

As it is visible on the 2. Figure, on January 13th the price started increasing rapidly with 70 million volume, which is 9 times the average volume of previous months, moving from 20,42 USD at market open to 31,4 USD at the close. From the 13th of January to the 21st of January the stock price increased to 43,03 USD. From the 21st of January to the 28th of January was the highest increase to the 483 USD peak on the 28th, this stage had the highest volume averaging around 150 million a day. On the 28th of January, most Brokers stopped allowing their customers to buy Gamestop, just allowing investors to sell the stock, also the short interest dropped at this point from 88% to 30% on the 29th of January. From the 29th of January to the 23rd of February the stock price decreased to 45 USD, with a high average volume.



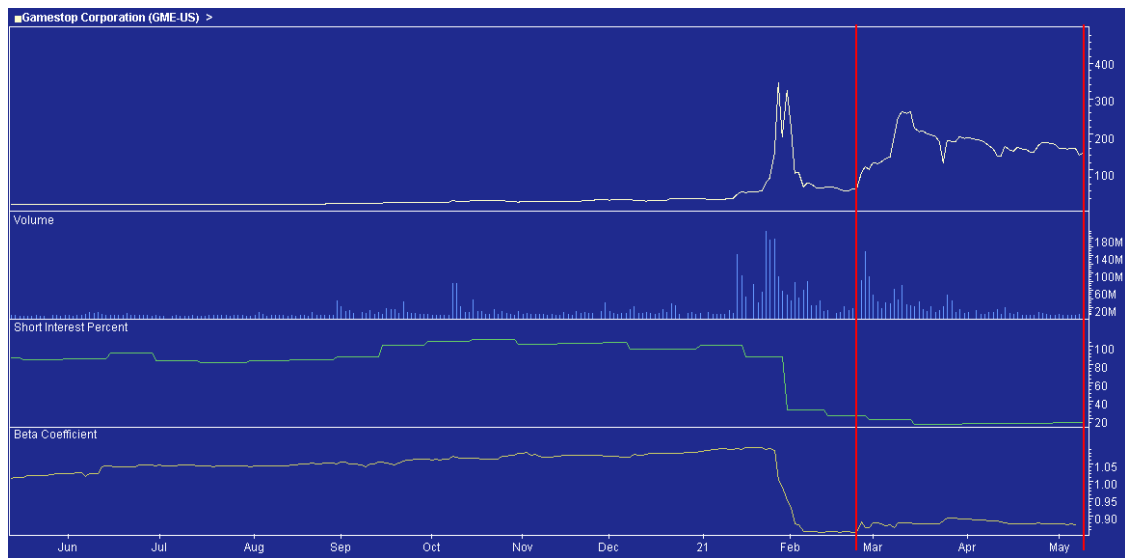
2. Figure Gamestop price, volume, short interest, Beta chart 2021 January-2021 March Source: Thomson One

2.1.3 The third phase:

As it is visible on the 3. Figure, from the 24th of February the stock price increased from 45 USD to a new high of 348,5 USD on the 10th of March, with volumes of around and over 100 million shares a day. On the 23rd of February, the CFO of Gamestop resigned, and changes in the whole structure and business model of the company were announced. On the 23rd of March, the Q4 earnings report was published reporting, that Adjusted Earnings per share is 1,34 USD. After the 10th of March, the stock price is slowly declining closing on the 24th of March at 120,34 USD. After the earnings and transformation report and the personnel changes in management, the stock soared to 194 USD on the 30th of March. From then till the 12th of May the price of the stock declined

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

to 146 USD with volumes of around 5-6 million shares per day. The short interest during this phase decreased from 23,6% to 14,5%.

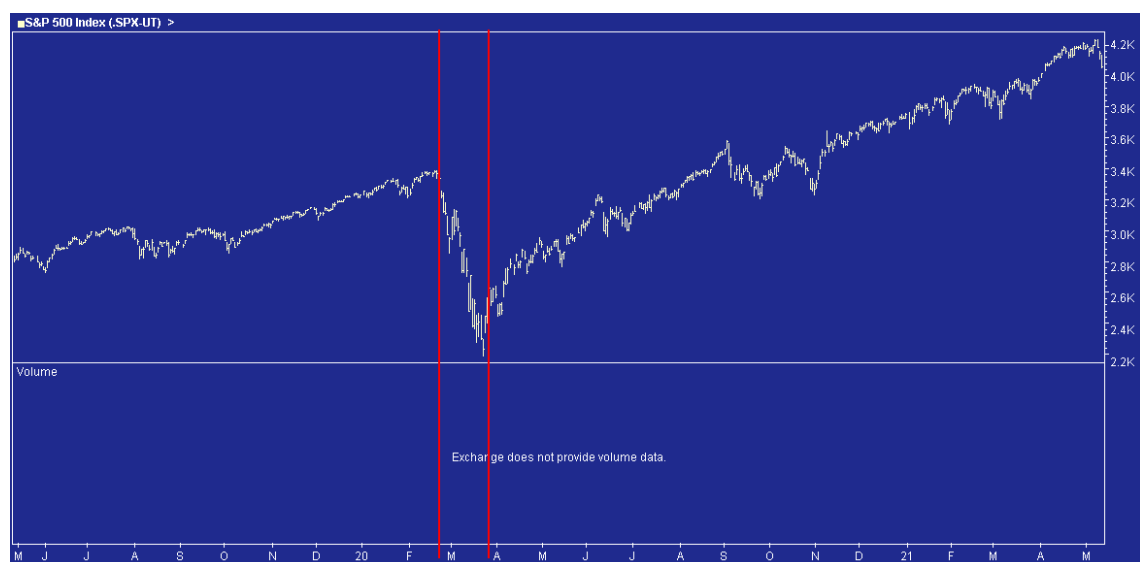


3. Figure Gamestop price, volume, short interest, Beta chart 2021 March-2021 May Source: Thomson One

2.2 Stockmarket

2.2.2 Current situation in the stock market

The current situation on the stock market, especially the U.S. stock market was influenced and created mostly by the COVID-19 pandemic. The pandemic hit the whole world in late winter, early spring of 2020. The coronavirus pandemic spread very fast like no other pandemic has hit in recent times (Ebola, Swine flu). There is no example of an outbreak in recent times, that can be compared in magnitude and speed to the COVID-19 pandemic. The uncertainty and losses due to the setback the pandemic caused, hitting all of the world's stock markets, the S&P 500 dropped 32% from the 24th of February to the 23rd of March, from 3226 USD to 2191 USD, the period is marked on Figure 4.



4. Figure S&P 500 chart February 24th to March 23rd Source: Thomson One

2.2.3 Economic effects of COVID-19

The pandemic induced both a supply and demand shock, as mentioned, by (Brinca et al., 2020), the supply shock is anything that reduces the capacity of the economy to produce goods and services at given prices. The demand shock reduces both the ability of the consumers to buy goods and services at given price points and their willingness to do so. Many businesses were forced by the government to shut down leaving people jobless, creating a supply shock, because factories, restaurants, and hotels were shut down and could not produce the goods and services. On the other hand, these jobless people were less likely to consume goods and services after being out of their job, and having no income.

Four main reasons causing decrease in the economic output of the world can be associated with COVID-19., according to the article by (Brinca et al., 2020).

The first of them is lockdowns, most people are working from home during the pandemic. In the production sector, the absence of workers results in a labor shortage, followed by less demand for capital ultimately leading to a cut down in production decreasing GDP.

The second reason behind the decline of economic output is the decline of net export, there are multiple reasons behind this, cross-border transportation became longer and this way more expensive. Furthermore, the producers are not able to run the factories because of the labor shortage caused by lockdowns, and the raw material suppliers are also having problems. The supply chains of most companies were disrupted, due to COVID-related reasons.

The third cause is the fall of tourism, hospitality, and leisure sectors, this impacts heavily both the companies and workers in these industries, but also the government, especially in countries and regions that rely heavily on tourism.

The fourth part is the demand shock that is causing the MCP, marginal propensity to consume to decrease, this way decreasing GDP. Most people are saving more caused of uncertainty and cutting down on non-essential products, also the jobless people do not have the income to consume.

2.2.4 Returns in largest indices

In the research from Chowdhury, Khan, and Dhar (2021), we can see in Figure 5, the changes in the mean returns of the sample indices, after the pandemic hit.

The mean return decreased the most for Japan (NIKKEI 225), that has an 827,83 % decrease, the second is the United Kingdom (FTSE 100 UK) decreased by 156,93 %, the third is China (SSE) by 123,39%, the fourth is Germany (DAXI) by 104,38 %. Under 100 %, but also huge decreases were also in India (NIFTY 50), by 89,17%, and in Spain (IBEX) by 86,55%. The United States (S&P 500) also decreased by 85,62%.

The only market that had an increase in mean average return out of the 12 is the Singaporean by 51,07%.

The standard deviation also tells us, that volatility increased in all the countries, except for the United States and the United Kingdom.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

	Pre-event		Post-event		Change	
	Mean	SD	Mean	SD	Mean (%)	SD
S&P 500 US	-0.0044	0.01559	-0.0006	0.01094	-85.62	-0.0046
FTSE 100 UK	-0.0039	0.02423	0.0022	0.00475	-156.93	-0.0195
FTSE ITALIA	-0.011	-0.0022	-0.0022	0.01257	-80.40	0.0147
NIKKEI 225	0.00033	-0.0024	-0.0024	0.00944	-827.83	0.0118
STI	0.00033	0.00049	0.00049	0.00557	51.07	0.0051
NIFTY 50	-0.009	-0.001	-0.001	0.00598	-89.17	0.0069
DAXI	-0.0101	0.00044	0.00044	0.00996	-104.38	0.0095
IBEX	-0.0105	-0.0014	-0.0014	0.00871	-86.55	0.0101
HANG SENG	-0.0043	-0.0014	-0.0014	0.01157	-66.51	0.0130
SSE	-0.0018	0.00042	0.00042	0.01834	-123.39	0.0179
CAC40	-0.0112	-0.0002	-0.0057	0.0211	-49.20	0.0213
SA40	-0.0004	-0.0057	-0.0002	0.00494	-52.27	0.0107

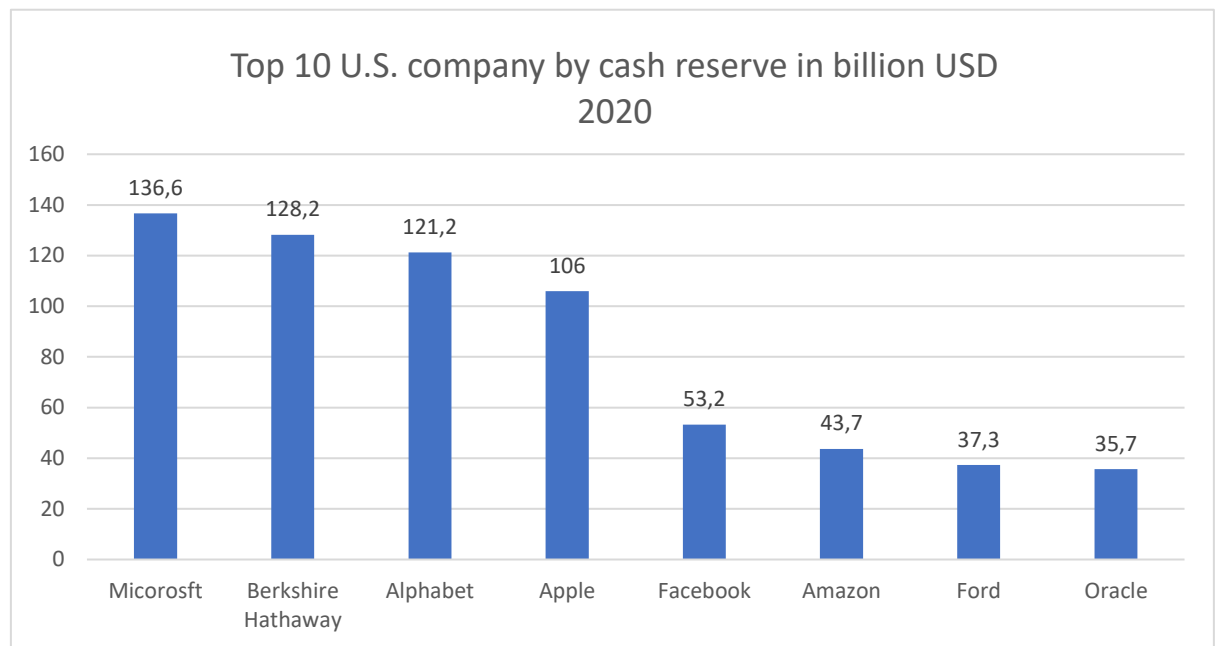
5. Figure Returns in largest indices during the Covid-19 pandemic Source: Chowdhury, Khan, and Dhar (2021)

2.2.5 Market sentiment in late 2020-2021

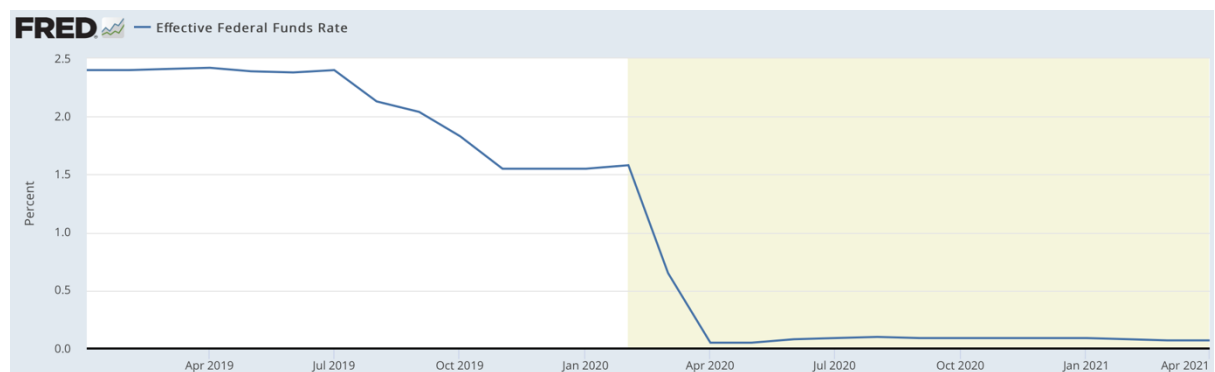
Comparing the market crash of 2020 to the most recent market crash of 2008, we can see that the situation is different. The 2008 market crash and financial crisis, had much longer economic effects, and reduced corporate profits for longer. The 2020 market crash was more of a panic selling, the economic and corporate profit setbacks are short-term, 1-2 years. High-quality companies had the cheapest valuation on equities in a long-time. Furthermore, the interest rates were low and will be lowered, by governments to stimulate the economy, this is illustrated on the 7. Figure. This makes it cheaper and easier for corporations to borrow, therefore corporations will borrow and invest their money generating economic growth.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

Corporations were also at the end of an 11-year-long bullish period, they built up cash reserves, this is shown in Figure 6, that can be used to survive for a longer period, the setbacks the following companies discovered were the production and supply chain setbacks due to the lockdown and border control. Also in the technology sector, the chip shortage caused by the effects of the coronavirus and the high demand slowed the growth, that was experienced in the last 11 years.



6. Figure U.S. corporations with largest cash reserves Source: Thomson One



7. Figure FED interest rate in the U.S. Source: FRED Economic Data

2.3 Stock Market trading

2.3.1 Introduction to stock market trading

The following part aims to introduce key trading aspects, that are connected to the Gamestop saga. The following 3 concepts, all took a part in the rapid increase in the share price. I will give a broad introduction to options trading, options were widely traded among investors on the wallstreetbets subgroup, probably due to the high risk, high reward nature of the options. The gamma squeeze is closely connected to the options and causes upward pressure on the underlying share price. Lastly, I will give an overview of short-squeezes, the causes of a short squeeze, and compare the chart of Gamestop to the Volkswagen short-squeeze of 2008.

2.3.2 Options trading

Options are contracts that can be both traded and executed between investors, 1 contract, means that 100 stocks can be bought or sold of the underlying stock.

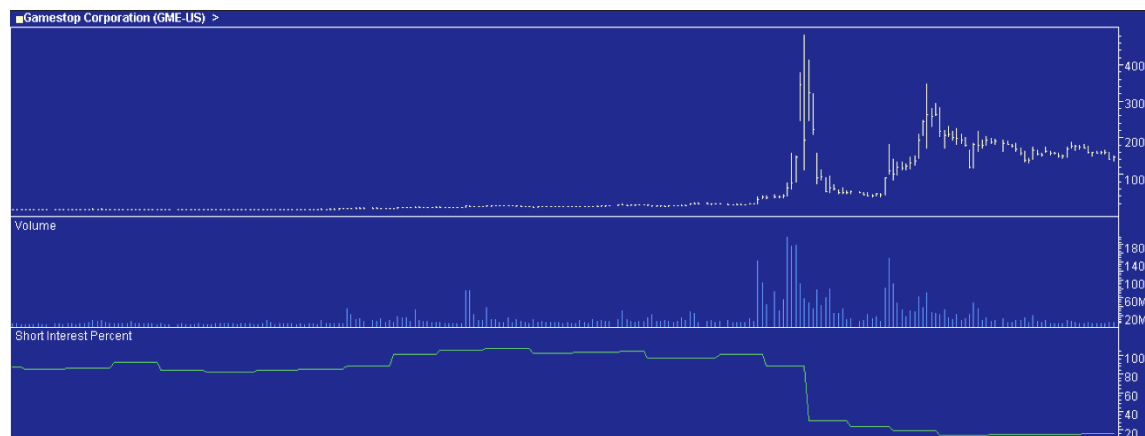
These options all have an expiration date and can fall into 4 categories, Long Call, Long Put, Short Call, and Short Put. American options can be executed at any given time between the purchase and expiration date, European options only can be executed at the expiration date. In the analysis, the long call options will be the most important when analyzing the meme stock and the short squeeze.

2.3.3 Gamma squeeze

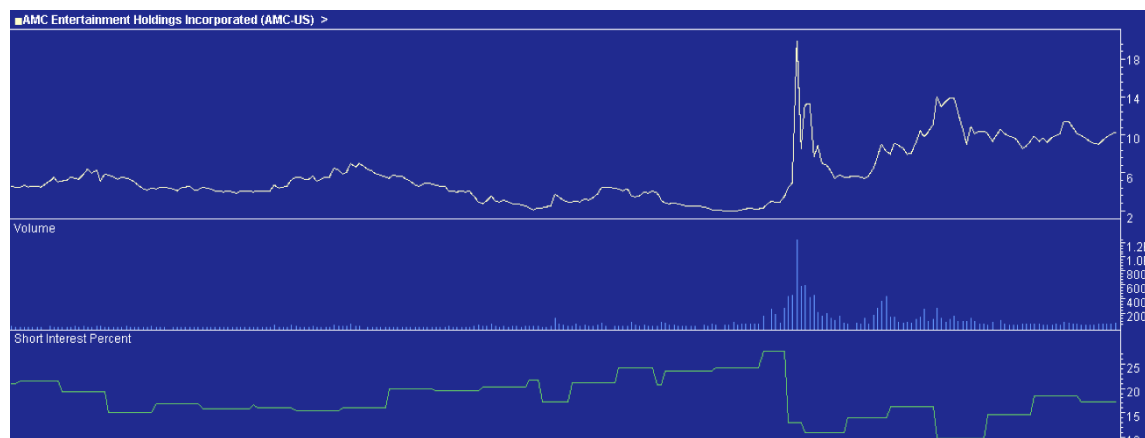
A Gamma squeeze happens when investors assume that a stock price will rise in the future, and start buying large amounts of short-dated call options in the stock. This puts the institutional investors that are selling the options in a short position. Institutional investors who are selling these options need to hedge their positions, this way need to buy stocks. Since, if the call options are exercised they need to be able to give these stocks at the given price point to those who hold the call options. This increased demand drives up the stock price, quickly in a short amount of time, and as more of these call options are sold the stock price goes even higher.

2.3.4 Short squeeze

The short squeeze is connected to the gamma squeeze, this is a strategy if institutional investors or a group of investors, buy and hold the stock to drive stock prices up. It is very effective when the short interest is already high, this was the case with the Volkswagen stock in 2007, and with the Gamestop stock in 2020/2021. Short squeezes are not common, it could be executed in the case of Gamestop and less likely in the case of AMC, because at the peak there was over a 100% short interest in Gamestop and around 30% in AMC., as it can be seen on Figure 8 and 9.



8. Figure Gamestop stock price, volume, and short interest Source: Thomson One

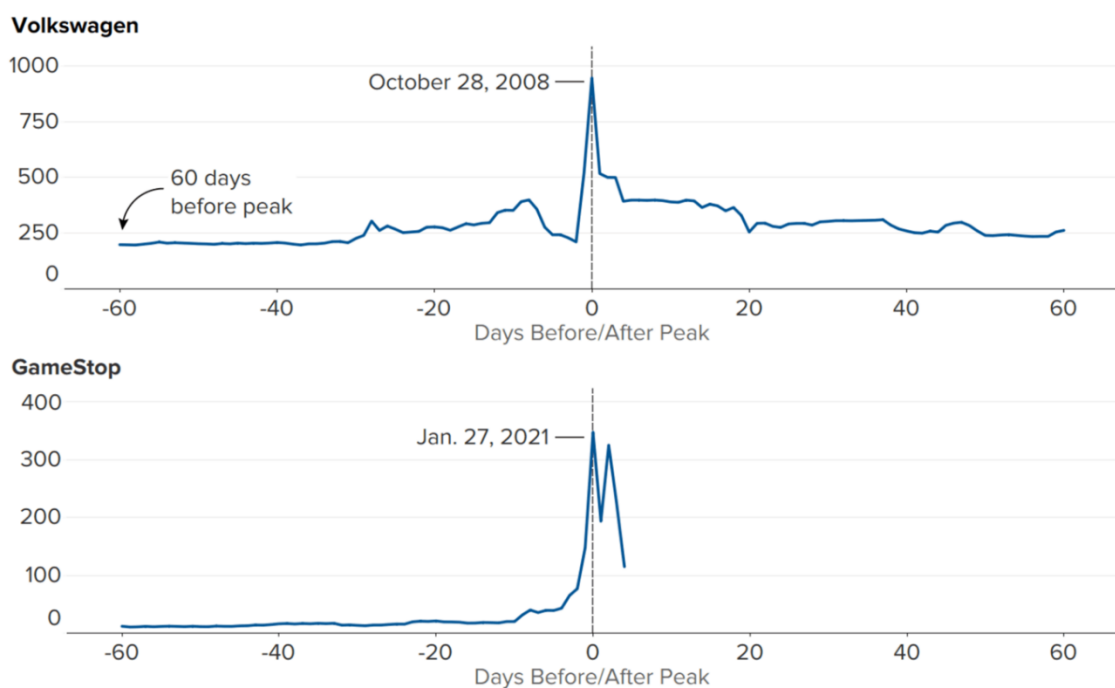


9. Figure AMC stock price, volume, and short interest Source: Thomson One

Furthermore, stock squeezes usually occur in stocks with either small market capitalization and float or with stocks that have a huge amount of short interest. Short squeezes can be facilitated by a short time to expiration call options. The stock prices will rise, if investors in a short position close their position, buying stocks, this way there will be higher demand and the price will go up. The price increase will cause other short sellers to close their positions this way creating a chain reaction and driving prices higher.

In Figures 8 and 9, it is visible that short interest falls when prices go up, the short squeeze starting a chain reaction after hitting the price point where investors close their positions to minimize losses, or because of a margin call.

The most important short squeeze in recent years was the short squeeze of the Volkswagen stock (VW), which happened in the middle of the 2008 financial crisis, Volkswagen has already had debt piling up, and the financial crisis lowered the demand for cars. However in 2008 Porsche decided that they want to increase their position in VW, which was at 31% in March of 2008, and did so by acquiring 74,1% of the company by October 2008. Porsche this way controls 74,1% of the company, and the second-largest stock owner was Lower Saxony holding 20% of the stocks. The stock available to buy on the open market this way decreased to 6%, around 12% of VW's shares were sold short at the time, creating a large inequality in supply and demand, with the price skyrocketing from 210,85EUR to 1000EUR in less than two days, as it can be seen on Figure 10, driving the company to be the largest company by market capitalization. After the peak, the stock price plunged by 58%, similar to the GME stock. This stimulus, that start the short-squeeze to happen was Porsche at the time, at GME, it was retail investors driving up the price and speculating on the short-squeeze. It is also worth noticing, that GME had a much bigger amount of stocks shorted, than VW had, this way a short squeeze was easier to happen.



10. Figure Volkswagen and Gamestop stockprice at the short squeeze Source: Internationalbanker, 2021

III. Methodological background

3.2 Introduction to the analysis of the Reddit thread

In the following part, I will explain the methodology I have used, starting with the collection of the data, following with the structuring and further analysis of the data. Afterward, I will perform a statistical analysis and conclude my thoughts regarding the whole topic, while looking at the fundamentals mentioned before. The next part will aim to find correlation if there is any between the mentions and the share price, trading volume, and short interest.

3.2.1 API and web scraping

In order to analyze these stocks further and connect them to the Reddit thread wallstreetbets, I have used a web scraping program to collect the posts mentioning the GME, AMC ticker, or both. The program that I have used collected all the data and structured it into excel, from 2020.08.31. to 2021.05.29.

The program is relatively simple, it is a computer program written in Python to extract data from web spaces, and only does this when it is allowed. In my case it is allowed since it is legal to web scrape on Reddit, the program automatically searches the posts that mention the given text, in my case the GME and AMC ticker, in the given timeframe and outputs it to an excel file, where they can be analyzed further. This data contains the posts in text format, the data, and an exact time, regarding the sharing of the post. I choose this method, since it is the easiest to perform data collection software, the code for the program is in Python and requires only a domain, as to where do you need the data to be extracted from, what data do you need from that webpage, and universal timecodes, regarding the length of the search, in my case it was from 2020.08.31. to 2021.05.29. Also, a large part of the data did contain both AMC and GME, so I have used the data where both of these companies were mentioned to analyze them compared to the share prices of the two stocks. I have specified this data as mentions in the future and structured it daily, however, it is important to mention that it was also necessary for my research to give meaning to the approximately 170000 individual posts since I cannot derive any correlation between the sheer amount of mentions and the stock price of the company. The posts needed to be analyzed in their sentiment, extracting the emotions the given sentence carries. The next part is concerned with the methods I have used for extracting this sentiment from these posts.

3.2.2 Sentiment analysis

Regarding sentiment analysis after carefully analyzing options, I have found two possible options, the first of them being manually deciding on the positivity negativity, or neutrality of each sentence, this option would have been the most accurate, however, given the magnitude of my research it was not reasonable to go through 170000 sentences and decide individually about the sentiment of each sentence. The other option explained in detail below was to use software to extract the sentiment of these sentences, using preexisting vocabularies and sentence analyzing tools, combined with machine learning. I have found two software that are applicable for my research, which were also used to conduct research, to analyze the relationship between the sentiment of Reddit posts and GME share price. (Long et al., 2021).

The first of them is the VADER sentiment analysis tool, the second one is the Text2emotion sentiment analyzing tool.

These tools use premade dictionaries and machine learning, supervised learning to determine different feelings and sentiments out of texts.

Firstly, regarding the VADER sentiment analysis tool, the name stands for Valence Aware Dictionary and Sentiment Reasoner, it is essentially a lexicon for words, and also a rule-based sentiment analyzing instrument that is made to analyze the web-based media. VADER utilizes a mix of words that are marked by their semantic direction positive and negative. The program is not only capable of telling if the word is positive or negative, but it also tells how positive or negative the word is, by giving us a value between 0 and 1 in order to measure the positivity or negativity. The VADER tool examines the text and verifies whether any of the words can be found in the lexicon. For example, the words good, and amazing have positive values up to +1, amazing is regarded as a more positive word, and thus has a higher positive score. Similarly, on the other end of the spectrum, negative words, if verified with the dictionary will give back a negative value up to -1. For example, bad and terrible both have negative values, however terrible is a stronger negative word, resulting in a value closer to -1. Other words, that are not in the dictionary are determined as neutral values and have a value of 0. VADER produces 4 measurements after completing these tests, the positivity, neutrality, and negativity score, while also calculating a compound score, which is a total score for the sentence.

Secondly, the Text2Emotion package was also developed to analyze large amounts of text data automatically, with the help of dictionaries. The key difference compared to VADER is that Text2Emotion is a more complex package, capable of

verifying 5 different emotions in the text, and presenting numerical values for them, similarly to VADER, however, Text2Emotion gives values between 0 and 1 for each of the 5 main emotions: happy, angry, sad, surprise and fear. It is also important to mention that this package can identify emotions from the emojis used in the sentence.

These packages are used in the finance, social media, and e-commerce industry just to name a few. Extracting values from text is a very important cornerstone in the future of analyzing sentiment on the stock market, public opinion about brands, products, or government policies.

Even though these packages are very complex and give mostly accurate output regarding the emotions of sentences, these softwares also have their limitations. Especially in my research, in the Wallstreetbets group on Reddit, there are a large number of comments, that use slang words, that are exclusively used on this subreddit, fortunately, these keywords and phrases signaling sentiment can be found in a Reddit Wallstreetbets post. I have used a table containing these words and phrases, from (Long et al., 2021). The Glossary mentioned has all the keywords, and phrases and their meaning. This Glossary can be found here: (https://www.reddit.com/r/wallstreetbets/comments/17fr21/basic_guide_to_wallstreetbets_culture_for/).

In order to factor in these slang words, I have filtered the output of the two packages for words and phrases that are associated with 4 signals and emotions: bullish, bearish, buy and hold. Bullish and buy being positive, while hold being neutral, rather positive than negative, and bearish being negative. I have manually changed the emotion output values for the posts that did contain these words, contractions, or phrases.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

All in all, I have used the packages to extract sentiment from the data, but I have also edited the output manually after the tests. I will analyze both the original data, that is the exact output of the two packages. I will analyze the sentiment analysis results, with both the GME and AMC financial data respectively. I will also analyze the data, I have modified, using keywords and phrases, that can be found in Figure 11., in order to receive a more accurate output. The exact phrases and their meaning, based on which I have edited the data can be seen in the table just after, to clarify, this table is from the analysis of Chen Long (Long et al., 2021).

Table 2: Reddit - r/WallStreetBets lexicon

Word/Phrase/Hashtag	Signal	Interpretation
I just like the stock	bullish	Used to justify purchasing the stock or holding it
We like the stock	bullish	Used to justify purchasing the stock or holding it
To the moon	bullish	Used to justify purchasing the stock or holding it
Mooning	bullish	The stock goes up
When Lambo	bullish	Used to say when will you get ridiculously rich off of the stocks
Apes	bullish	Retail investors bullish on heavily-shorted stocks
Tendies	bullish	Refers to financial gains in the stock market
Long as a python	bullish	Being very bullish with no regard to risk
Diamond Hands	hold	Used to encourage community to not sell the stock
Paper Hands	hold	Used as an offence to discourage community from selling the stock
Hold the line	hold	Holding on to a stock
Bagholding/Bagholder	hold	holds stock that's fallen in value and keeps holding
HODL	hold	hold on the stocks
YOLO	buy	"You only live once" used to encourage to invest more in the stock
Buy High Sell low	buy	Used to encourage to invest more in the stock
Buy the Dip/BTFD	buy	Used to encourage to buy the stock has dropped in price
DD/Double Down	buy	Used to mean that the stock is promising
Rocket Ships	buy	Used to mean the stock mentioned will have a dramatic gain
Guh	bearish	Used when share price fell caused losses
Drilling	bearish	The stock price goes down
TDM/Threatening downward movement	bearish	Used to encourage to sell any positions
FUD	bearish	Short form for 'fear, uncertainty and doubt'
Bear Gang/Gay Bears	bearish	People who gets happy when stocks go down
Bull Gang/Big Dongus Crew	bullish	People who gets happy when stocks go up
Kang Gang	hold	People who gets happy when stocks go way up and then way down
Theta Gang	hold	People who gets happy when stocks go sideways

Note: For more detailed r/WallStreetBets glossary follow <https://www.reddit.com/r/wallstreetbets/wiki/glossary>

11. Figure Key phrases from Reddit Wallstreetbets Source: Long, C., Lucey, B. M., & Yarovaya, L. (2021). "I just like the stock" versus "Fear and loathing on main street"

3.3.3 Regression analysis methodology

I have chosen to perform a multiple linear regression analysis, in order to find correlation between the variables. I have chosen the linear regression analysis since it is the simplest and one of the most efficient ways to describe relationships between the variables. For the dependent variable, I had the GME and AMC average daily price, respectively, and for the independent variables, I have used both the values concerned with the sentiment, and the values that are not sentiment-based.

Also, I have separated the variables into 4 different groups. The linear regression analysis was performed for both the GME and the AMC data. Furthermore, it was categorized into 2 categories each, the first is not connected to the sentiment analysis

output, these variables are the following: Daily volume, Short interest, and the number of mentions each day. The second category is concerned with the different outputs of the sentiment analysis packages, and their relationship, with the average daily price of the GME and AMC stock.

As it will be clearly visible in the analysis part, the linear regression analyses concerned with the volume of short interest and mentions are statistically significant. On the other hand, some of the sentiment-based regression analyses are not significant at a low confidence level. This can be caused by the data not being large enough, or if there is no relationship between the dependent and independent variables.

3.3.4 Analysis of descriptive statistics of variables

In this part, I will analyze the descriptive statistics of the variables for both the original and the modified dataset. The two tables regarding the descriptive statistics of the two datasets are in the appendix.

3.3.4.1 Descriptive analysis of original data

Firstly, the descriptive statistics for all the variables of the original data can be seen in Figure 30. The n value represents the number of days I have analyzed in my thesis and for all variables $n=242$.

3.3.4.1.1 Text2Emotion descriptive statistics

The first 5 rows of Figure 30. are the descriptive statistics values of the Text2Emotion analysis, the minimum values, and maximum values except for the emotion sad, take up all the possible values between 0 and 1, while sad is the only one that's maximum value is only 0,5. The mean of the emotion angry is 0,075 while the standard deviation is 0,103. Also, the emotion angry has a positive skewness, the individual entries for this emotion will be skewed to the right. The mean of the emotion happy is 0,072 while the standard deviation is 0,104. Also, the emotion happy has a positive skewness, the individual entries for this emotion will be skewed to the right. . The mean of the emotion fear is 0,22 while the standard deviation is 0,121. Also, the emotion fear has a positive skewness, the individual entries for this emotion will be skewed to the right. The mean of the emotion sad is 0,11 while the standard deviation is 0,76. Also the emotion sad has a positive skewness, the individual entries for this emotion will be skewed to the right. The mean of the emotion surprise is 0,1 while the standard

deviation is 0,84. Also, the emotion surprise has a positive skewness, the individual entries for this emotion will be skewed to the right.

3.3.4.1.2 VADER descriptive statistics

In the following part I will analyze the values of the VADER sentiment analysis output, the n values did not change, so $n=242$. In Figure 30. it is visible, that the positive values have a minimum close to 0, and the maximum value is 0,33, while the mean value is 0,08 and the standard deviation is 0,048. Also, the sentiment positive has a positive skewness, the individual entries for this emotion will be skewed to the right. The neutral values have a minimum close to 0,541, and the maximum value is 1, while the mean value is 0,87 and the standard deviation is 0,057. Also the sentiment neutral has a negative skewness, the individual entries for this emotion will be skewed to the left. The negative values have a minimum close to 0, and the maximum value is 0,46, while the mean value is 0,46 and the standard deviation is 0,044. Also, the sentiment negative has a positive skewness, the individual entries for this emotion will be skewed to the right. The compound values have a minimum close to -0,527, and the maximum value is 0,54, while the mean value is 0,5 and the standard deviation is 0,1. Also, the compound value has a negative skewness, the individual entries for this emotion will be skewed to the left.

3.3.4.2 Descriptive analysis of modified data

Firstly, the descriptive statistics for all the variables of the modified data can be seen in Figure 31. The n value represents the number of days I have analyzed in my thesis and for all variables $n=242$.

3.3.4.2.1 Text2Emotion descriptive statistics

The first 5 rows of Figure 31. are the descriptive statistics values of the Text2Emotion analysis, the minimum values, and maximum values except for the emotion sad, take up all the possible values between 0 and 1, while sad is the only one that's maximum value is only 0,4. The mean of the emotion angry is 0,066 while the standard deviation is 0,1. Also, the emotion angry has a positive skewness, the individual entries for this emotion will be skewed to the right. The mean of the emotion happy is 0,25 while the standard deviation is 0,16. Also, the emotion happy has a positive skewness, the individual entries for this emotion will be skewed to the right. . The mean of the emotion fear is 0,17 while the standard deviation is 0,109. Also, the emotion fear has a positive skewness, the individual entries for this emotion will be skewed to the right. The mean of the emotion sad is 0,09 while the standard deviation is 0,66. Also the emotion sad has a positive skewness, the individual entries for this emotion will be

skewed to the right. The mean of the emotion surprise is 0,08 while the standard deviation is 0,79. Also, the emotion surprise has a positive skewness, the individual entries for this emotion will be skewed to the right.

3.3.4.2.2 VADER descriptive statistics

In the following part I will analyze the values of the VADER sentiment analysis output, the n values did not change, so n=242. In Figure 31. it is visible, that the positive values have a minimum close to 0, and the maximum value is 1, while the mean value is 0,25 and the standard deviation is 0,14. Also, the sentiment positive has a positive skewness, the individual entries for this emotion will be skewed to the right. The neutral values have a minimum close to 0, and the maximum value is 1, while the mean value is 0,7 and the standard deviation is 0,14. Also the sentiment neutral has a negative skewness, the individual entries for this emotion will be skewed to the left.

The negative values have a minimum close to 0, and the maximum value is 0,46, while the mean value is 0,4 and the standard deviation is 0,047. Also, the sentiment negative has a positive skewness, the individual entries for this emotion will be skewed to the right. The compound values have a minimum close to -0,527, and the maximum value is 1, while the mean value is 0,221 and the standard deviation is 0,17. Also, the compound value has a positive skewness, the individual entries for this emotion will be skewed to the right.

3.3.4.3 Financial and other data descriptive analysis

In the following part I will analyze the price, the volume traded, short interest, and number of mentions for both AMC and GME. The n value remains 242 for these variables as well. Since this data has not been modified the analysis is for both the original and modified data

The GME average daily price has a minimum of 6,2 and a maximum of 352,4 over the period. The mean of the average daily price is close to 80, while the standard deviation is 86, the variable also has a positive skewness, and the individual entries for this variable will be skewed to the right. The GME daily volume has a minimum of 2478168 and a maximum of 197157946 over the period. The mean of the daily volume is close to 22153850, while the standard deviation is 33608498, the variable also has a positive skewness, the individual entries for this variable will be skewed to the right.

The AMC average daily price has a minimum of 2 and a maximum of 29 over the period. The mean of the average daily price is close to 6,6, while the standard deviation is 4,4, the variable also has a positive skewness, the individual entries for this variable

will be skewed to the right. The AMC daily volume has a minimum of 2090061 and a maximum of 1253253552 over the period. The mean of the daily volume is close to 86477559, while the standard deviation is 150004839, the variable also has a positive skewness, the individual entries for this variable will be skewed to the right.

The minimum statistic for the GME short interest is 16,5, while the maximum is 115,3. The mean value is close to 72, while the standard deviation is 43,3, the variable also has a negative skewness, the individual entries for this variable will be skewed to the left. The minimum statistic for the AMC short interest is 3,5, while the maximum is 20. The mean value is close to 11,7, while the standard deviation is 5,3, the variable also has a positive skewness, the individual entries for this variable will be skewed to the right.

The daily mentions of both AMC and GME have a minimum of 1 and a maximum of 27472 the mean number of mentions is close to 640 while the standard deviation is 2586 the variable also has a positive skewness, the individual entries for this variable will be skewed to the right.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

3.3.5 Hypotheses

Firstly, in order to start the analysis, I will formulate two different hypotheses, and I will derive conclusions, based on the given data and, by using regression analysis. The statistical analysis will be conducted in 4 ways in order to capture the relationships between the variables. I will analyze the original data for the AMC volume, short interest, and average daily share price, I will also analyze the GME share in the same way, with the same indicators. Additionally, I will also perform the same statistical analysis on the modified data set, both with the GME and AMC indicators.

H1A: The tone of Reddit's sentiments affects daily average GME stock prices

The focus of this hypothesis is to derive, whether the sentiment of Wallstreetbets posts had a relationship with the daily average stock price and the trading volume of GME shares.

H1B: The tone of Reddit's sentiments affects daily average AMC stock prices

The focus of this hypothesis is to derive, whether the sentiment of Wallstreetbets posts had a relationship with the daily average stock price and the trading volume of AMC shares.

H2A: GME short interest affects daily average GME stock prices

This hypothesis focuses on, whether there was a short-squeeze and how much it has affected the GME share price and trading volume.

H2B: AMC short interest affects daily average AMC stock prices

This hypothesis focuses on, whether there was a short-squeeze and how much it has affected the AMC share price and trading volume.

IV. Analysis

4.1 Fundamental analysis

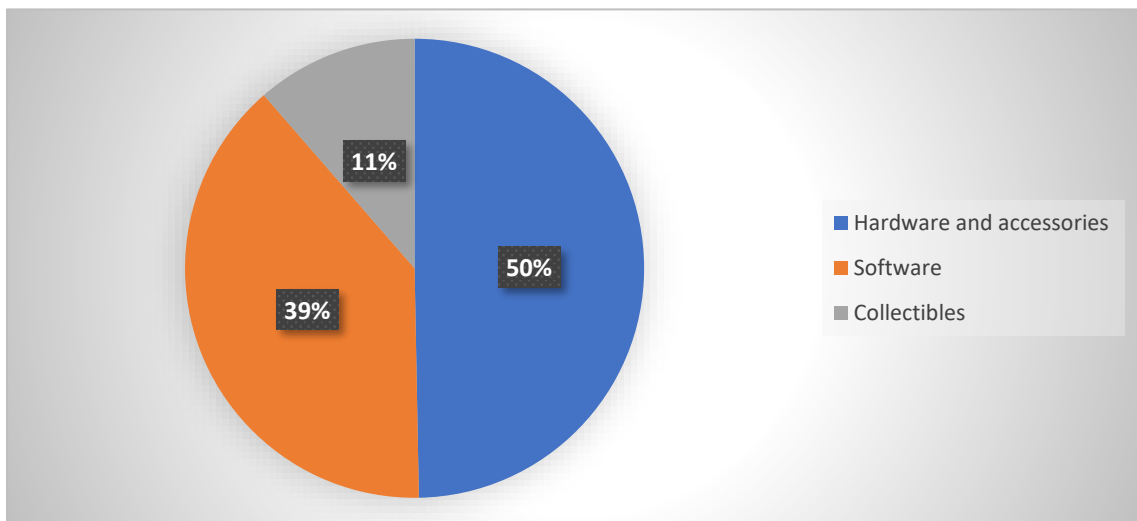
4.1.1 Gamestop (GME)

4.1.1.1 Introduction

Firstly, in order to rule out the possibility of the GME stock being a good trade based on fundamentals, I will conduct an analysis of the key financials. For this, I will look at the following data. Gamestop is in the specialty retail sector, selling hardware, such as video game consoles (i.e. XBOX, Playstation,). The other two categories of sales are video games (software) and also collectibles (video game statues, trading cards). For the Fiscal year ending January 2020. Starting with the Sales Breakdown by Product Category.

The fundamental analysis has been conducted based on the analysis by (Fundamental Diagnosis, 2021).

4.1.1.2 Sales Breakdown by Product Category



12. Figure Gamestop sales breakdown by product category Source: Research, Gamestop yearly reports

In the 12th Figure, it is visible, that 50% of the sales are from hardware and accessories, 39% are from software, and 11% are from collectibles. Looking at the sales of the company in the past, the company is constantly declining from 2018 to 2019, by approximately 22%, and from 2019 to 2020, by 21%. The three main product categories' ratios also change, as it is illustrated in Figure 13, with hardware and accessories accounting for a large part of sales, software declining, and collectibles increasing. The main reason for the decline is online shops, which are the most relevant with software,

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

resulting in the largest decline in that segment.

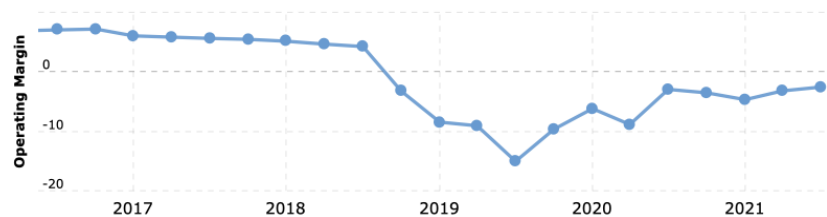
	Fiscal Year 2020		Fiscal Year 2019		Fiscal Year 2018	
	Net Sales	Percent of Net Sales	Net Sales	Percent of Net Sales	Net Sales	Percent of Net Sales
Hardware and accessories	\$ 2,530.8	49.7 %	\$ 2,722.2	42.1 %	\$ 3,717.8	44.9 %
Software	1,979.1	38.9	3,006.3	46.5	3,856.5	46.5
Collectibles	579.9	11.4	737.5	11.4	711.0	8.6
Total	\$ 5,089.8	100.0 %	\$ 6,466.0	100.0 %	\$ 8,285.3	100.0 %

13. Figure Gamestop sales breakdown by product category from 2018 to 2020 Source: Fundamental Diagnosis, 2021

4.1.1.3 Operating margin

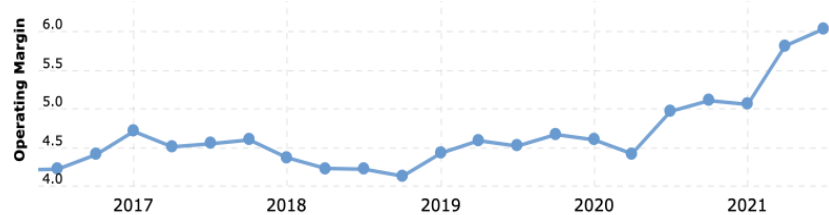
Operating margin is a way to measure the efficiency of a business, dividing operating income by sales. The strong decline in sales caused Gamestop's operating margin to decline. The company was not able to reduce costs quickly enough to keep up with the strong decline in sales. Comparing the operating margin of Gamestop to operating margins in the retail industry (Walmart(Figure 15) and Best Buy(Figure 16.)) It is visible that, larger competitors, can achieve operating margins of 4-5%, which was the case for Gamestop as well prior to the decline starting in 2019.

Gamestop



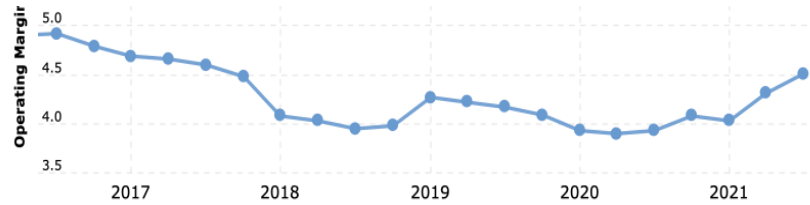
14. Figure Gamestop operating margin Source: Fundamental Diagnosis, 2021

Best Buy



15. Figure Best Buy operating margin Source: Fundamental Diagnosis, 2021

Walmart



16. Figure Walmart operating margin Source: Fundamental Diagnosis, 2021

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

4.1.1.4 The turnaround plan

It is also important to mention, that in 2018 Gamestop made efforts to create a new business plan, that transforms the business model to open an eCommerce store. The company said in 2019, that this store would be generating 1 billion dollars in sales, this amount under current circumstances would account for 20% of the sales of the whole business. However there are difficulties in this scenario, the total market for video game products is approximately 150 billion USD and it is expected to grow in the next 5 years. Gamestop is focused on the console market, with a size of approximately 20 billion USD, 13% of the total market. While the market is expected to grow, the main areas for growth are cloud gaming (Stadia), gaming on mobile devices, and eSports getting even more recognition. Unfortunately, most of the eSport industry is played on PCs, a market that Gamestop is not a relevant player in. Also, the market for online video game stores has very fierce competition, such as Amazon, and Steam. The manufacturers also have their online stores, such as the PlayStation Store from Sony, or Origin from Electronic Arts.

4.1.1.5 Bottom line

The main problem for Gamestop is the decline in sales. The decline in sales can be caused by several factors, such as the declining demand for physical videogames, videogames are purchased on online platforms instead. This is enabled by cheaper prices, cheaper storage, faster internet connection, and this option being more convenient for the customer than going to a store and purchasing a physical copy. The hardware market for video game consoles is also cyclical, with 2019 being the last year of the previous generation gaming consoles (Xbox One, Playstation 4), with the new generation (Xbox Series X, Playstation 5) being released in 2020. There is much less demand for gaming consoles in the last year of their 6-7 year life cycle. Also because of the pandemic physical stores were shut down, and not visited as much by customers, this did not help Gamestop either and made more people switch to purchasing digital copies of games.

4.1.1.6 Is the price justified?

Based on analyses from multiple analysts, (W.S. Journal, 2021) even in the case of a widely successful turnaround calculating with an 8% hurdle rate, and the annual cash flows. In a very optimistic case, the share should not be worth more than 60 USD. Based on the rating of 4 analysts with the highest target price is 190 USD and the lowest at 20 USD, the average target price for a share is 71 USD, which is similar to the calculations mentioned above.

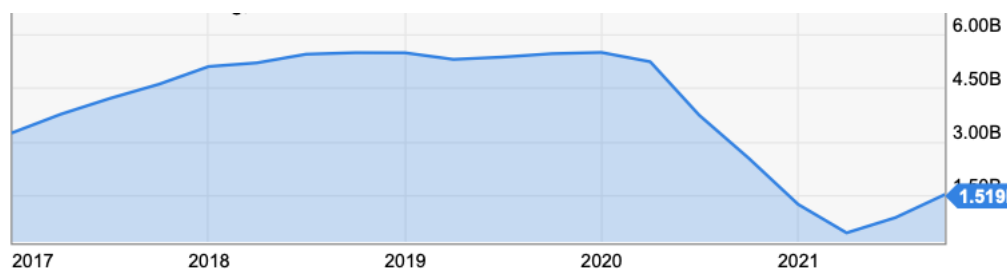
4.1.2 AMC

4.1.2.1 Introduction

AMC was hit hard by the pandemic since cinemas were closed down. Other players in the movies and entertainment industry, those who focus on streaming services thrived on the contrary. Furthermore, even before the pandemic, AMC is part of a slowly dying industry, more and more consumers prefer streaming and watching movies at home, instead of going to cinemas. My fundamental analysis is based on the analysis conducted by (Vashi, 2021).

4.1.2.2 Revenue

Before the pandemic hit, the revenue of AMC was approximately 5 billion USD from 2018-to 2021. In 2020, after the pandemic hit, revenue dropped down to around 1,2 billion USD. Although we can see an increase in revenue in the second part of 2021, in Figure 17, more people choose to stream movies instead, and even content makers, such as Disney are moving to an online streaming platform.



17. Figure AMC Revenue Source: (Vashi, 2021)

4.1.2.3 Debt

Also in order to survive the pandemic, AMC has raised debt and equity capital in order to survive. This means that the company has 5 times the shares outstanding compared to pre-COVID levels. The company has 5,5 billion USD worth of debt, and with already relatively low-profit margins will become even lower with increased interest costs. Also regarding COVID, it is important to mention that the pandemic comes back stronger cyclically with more advanced variations, so another shock, like closing down all cinemas for a month has a real possibility.

4.1.2.4 Valuation

Vashi, 2021 valued the company with the L.A. Stevens Valuation model. Using a DCF model with free cash flow discounted by the cost of capital, and with the model accounting for the effects of change in outstanding shares. The data in the valuation are assumptions of (Vashi, 2021):

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

Forward 12-month revenue [A]	\$4 billion
Potential Free Cash Flow Margin (realistic) [B]	2%
Average diluted shares outstanding [C]	~513 million
Free cash flow per share [$D = (A * B) / C$]	\$0.155
Free cash flow per share growth rate (generous estimate)	5%
Terminal growth rate	2%
Years of elevated growth	10
Total years to stimulate	100
Discount Rate (Our "Next Best Alternative")	9.8%

18. Figure AMC Discounted Cash Flow analysis input estimations Source: Vashi, 2021

Using the assumed variables, illustrated in Figure 18 in the model, (Vashi, 2021) found that the Present Value of an AMC stock is 2,52 USD. This means that using these realistic assumptions the share is massively overpriced, and is 1290% overvalued.

4.1.2.5 Bottom line

I can conclude that there is no fundamental reason why the stock would have good value, therefore it is down to technicals, such as short interest and sentiment-driven pricing, as we can see in the case of GME.

Here's what AMC's President and CEO Adam Aron [had to say](#) on his company's retail shareholding:

"The number of investors who want to own a part of AMC continues to increase and now stands at approximately 4.1 million. More than 80% of AMC shares are held by a broad base of retail investors with an average holding of around 120 shares. Some hold more and some hold less, however, each and every shareholder is important to AMC. Each shareholder has a critical role to play in

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

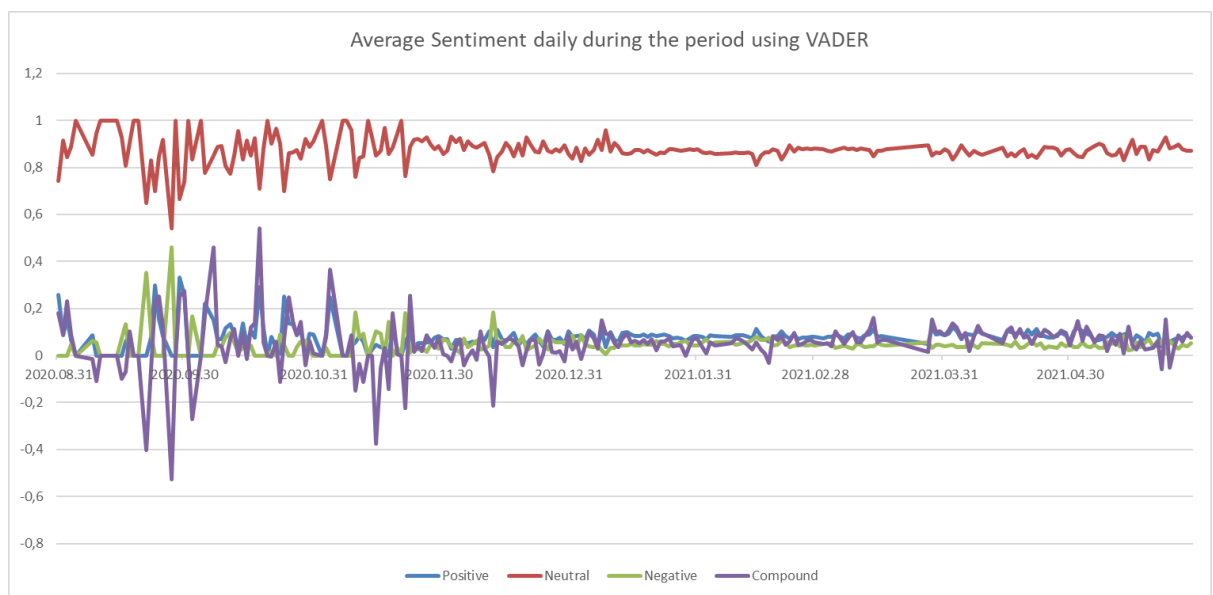
AMC's future by having their voice heard by voting at our upcoming Shareholder Meeting. By voting in favor of the proposals, together we can help position AMC, in its 101st year of business, for continued success over the next century."

With approximately 4,1 million retail shareholders, who are buying based on sentiment, and with the short interest of approximately 16,5% another short squeeze cannot be ruled out. However, in this strategy, it is essential to sell the stock before the prices drop back in the long run to the real fundamental valuation of approximately 2,5 USD.

4.2 Analysis of Sentiment

4.2.1 VADER(Original)

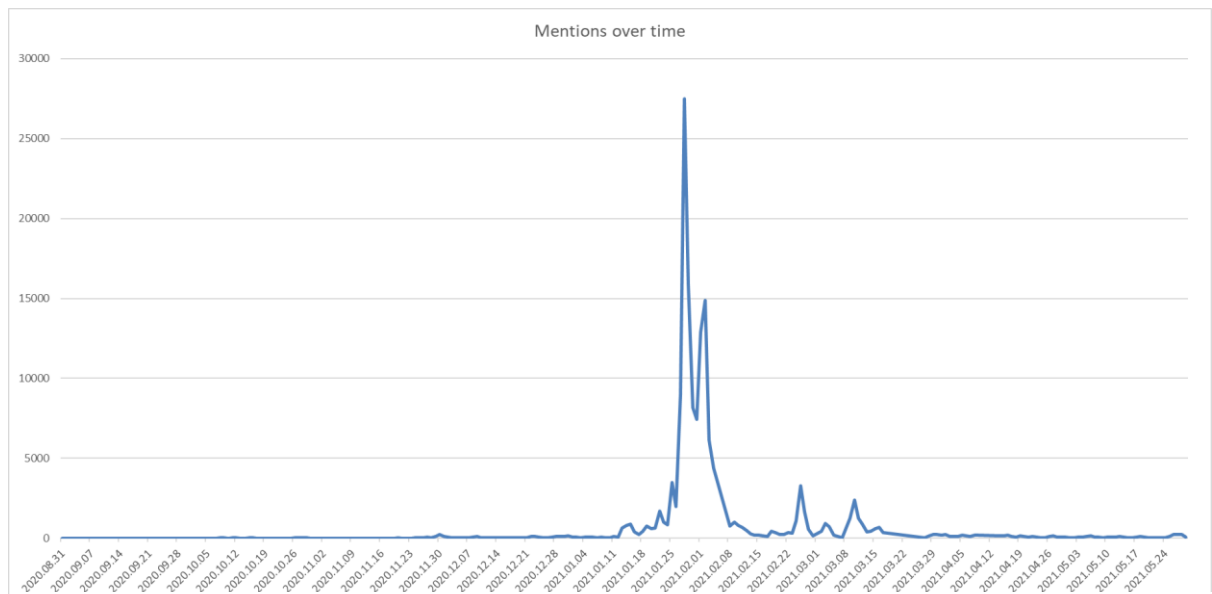
Firstly, I will start by analyzing the original daily sentiment output of the VADER method. The results are illustrated in Figure 19. The neutral sentiment is dominant in the subreddit in the time period, this can be a consequence of the package not being trained for the specific subreddit, where there are a lot of slang words and phrases, contractions used, this way the program gives out a neutral rating because it cannot decide on the positivity or negativity of these words.



19. Figure Average daily sentiment using the VADER package and the original data Source: Túű András

Secondly, the negative and positive values are similarly distributed, with values around 0 mostly over the period. It is important to mention that contrary to the expectations, the compound and positive and negative sentiment had a larger motion from late august to mid-November, which is the result of the small amount of data. There were far fewer mentions in that time period than in the later stage from the end of December onwards.

Lastly, the compound value, which gives an overall value for sentiment is moving in a similar fashion to the positive values after the end of December. The chart of the original data shows, that there is no real spike in positivity, that was expected in February, also the spikes of both the positive, negative, and compound values can be associated, with the limited data for the early periods. It can be seen in Figure 20, that compared to the spikes in mentions data is concentrated around January and February.



20. Figure Average daily mentions Source: Túú András

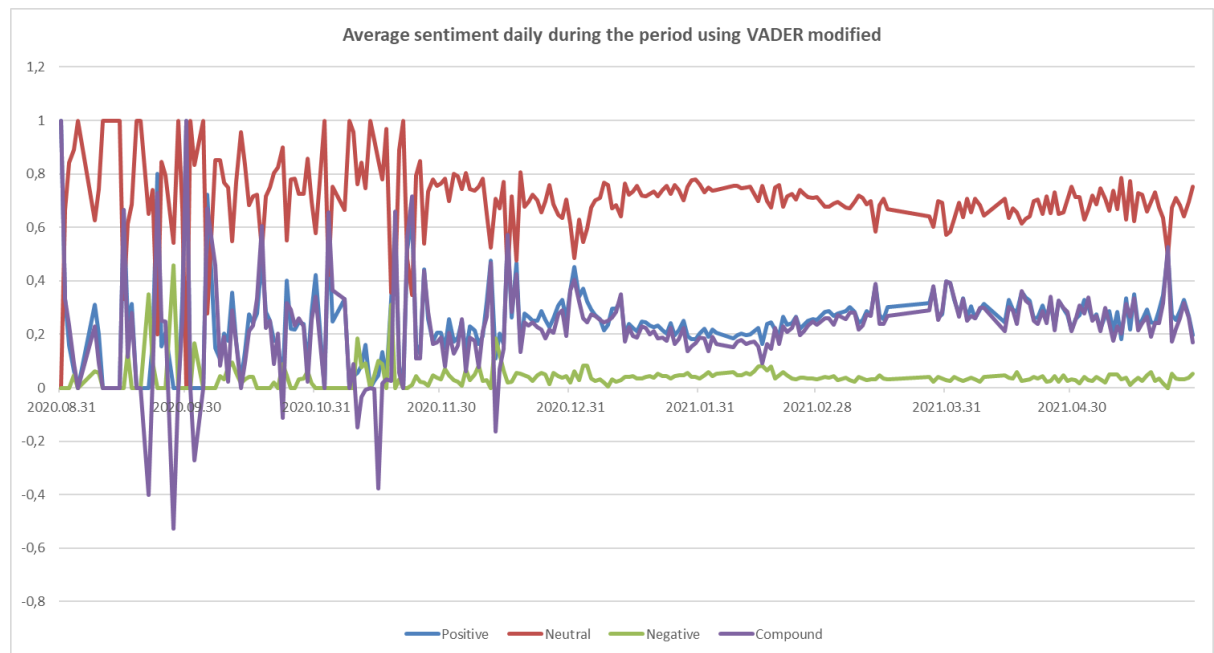
4.2.2 VADER(modified)

On the 21. Figure of the modified data, using also VADER analysis, there are more trends visible. The neutral value is dominant in this analysis as well. However, the positive sentiment is stronger here, while the negative sentiment is very similar to the original data.

The fluctuation of the values before December can also be seen here, which I have mentioned earlier due to the lack of data from that time period. By lack of data I mean, that the stocks were not that popular back then, hence there are not many mentions of them. The compound value similarly to the original unmodified data follows the line of the positive sentiment closely.

To sum it up, the modification of the data, which was aimed at getting a more accurate sentiment of the posts, was successful. The natural values, due to the slang words, and phrases are reduced after the modification to the data, resulting in more accurate statistical results in the next parts. This was just a limitation of the VADER package in this niche.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.



21. Figure Average daily sentiment using the VADER package and the modified data Source: Túű András

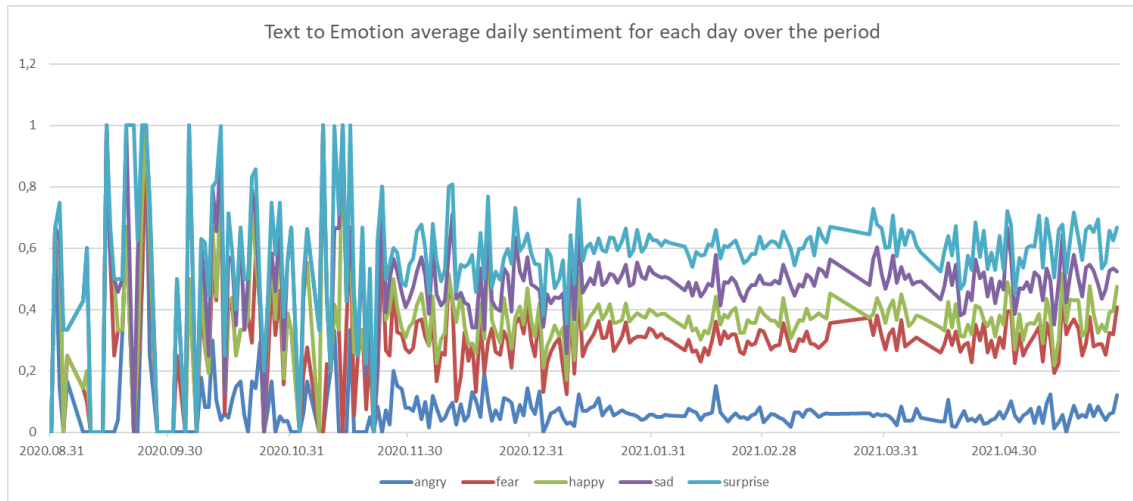
4.2.3 Text2Emotion (Original)

The other more complex text2emotion package has several differences, compared to the simpler VADER package. Meaning that it can have an output of 5 different emotions and can also recognize emojis, which comes in handy when dealing with online posts. The results are illustrated on the 22. Figure.

The dominant sentiment in the case of this analysis is surprise, which is generally a positive emotion, mostly associated with happiness. The other sentiments, sadness, happiness, and fear come after that. Sad is the second most common sentiment, which can be associated, with the losses the retail investors earned on options and long positions. The happy emotion is associated, with the strong upwards motion of the stocks. The angry emotion is caused by the frustration of the halting of trading and the hate towards the hedge funds shorting the shares, while the fear is associated with the stock price tanking, and the positions that have caused huge losses. The huge fluctuations in emotions in the early months of the period can also be credited to the limited amount of data. The different emotions follow a similar pattern over the period after the end of December, also a spike in all emotions is visible at the time when the share price jumped to its highest value.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

It is important to mention that this is also the unedited data, and has its own limitations of the text2emotion package, however, this package has managed to extract more emotions, even unedited, this can be associated, with the complexity of the package, and the ability to recognize emojis.



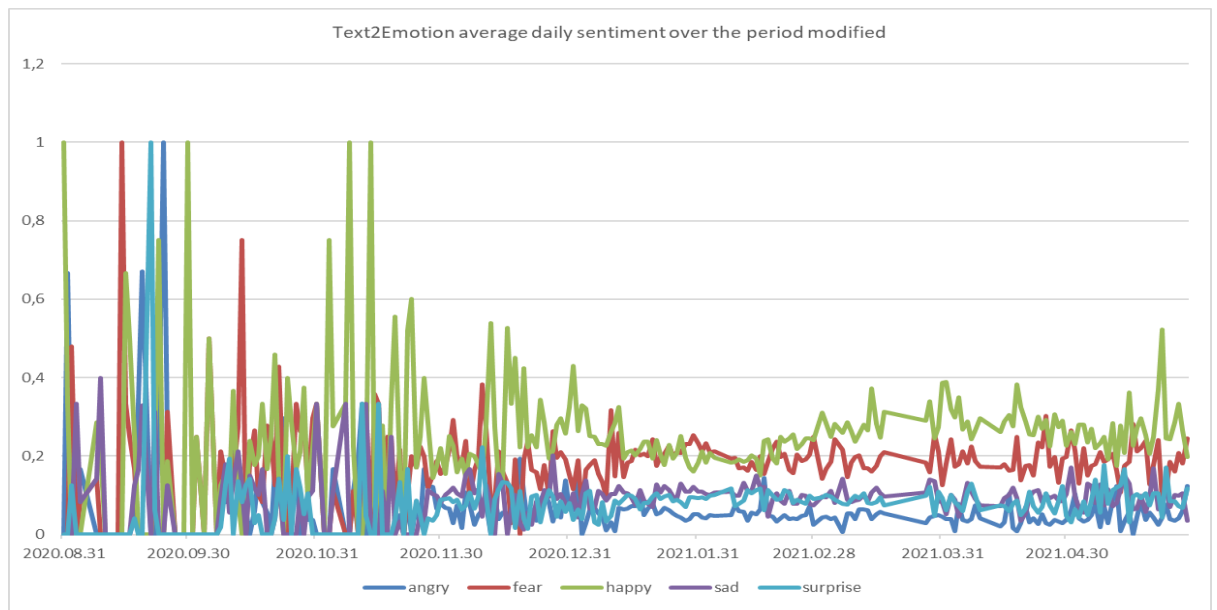
22. Figure Average daily sentiment using the T2E package and the original data Source: Túű András

4.2.4 Text2Emotion (Modified)

After the manual modification of the data, there are changes in terms of dominant sentiment in the subreddit. The changes are illustrated in Figure 23. The dominant sentiment was happy, with the second strongest sentiment being fear. Sadness, fear, and anger have also decreased largely after the modification of the data. The key difference, between after and before the modification was the reduction in surprise and sadness, while anger and fear stayed relatively at the same levels. And happiness as sentiment increased, reaching higher levels than before.

The lack of data in the early stages is still visible, after that the different emotions similar to the other 3 charts become less volatile, due to the increase in the number of posts. Also, another important difference between the modified and original data is the stronger increasing trend, regarding the different emotions after the 28th of February.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.



23. Figure Average daily sentiment using the T2E package and the modified data Source: Túű András

4.3 GME analysis

4.3.1 Volume, short interest, mentions

In the following part, I will analyze the regressions of the GME average daily price and its correlation with the independent variables GME daily traded volume, the short interest in GME, and the daily mentions of the GME ticker on the wallstreetbets subreddit. Furthermore, I have conducted the analysis, with both the original GME data and the modified one, even though I have not modified the data being analyzed here, this will be important in the analysis of the sentiment regressions. The results are presented in the 24. Figure.

	GME original		GME modified	
	Coefficients	Significance	Coefficients	Significance
GME daily volume	0,101	0,115	0,101	0,115
GME Short Interest	-0,808	0,001	-0,808	0,001
GME mentions	0,374	0,001	0,374	0,001

24. Figure GME volume, short interest, mentions regression analysis output Source: Tűű András

Firstly, there is a weak positive correlation between the Daily volume of GME and the Average daily price of GME, the correlation coefficient is around 0,1, but also has a significance level over 0,1, meaning that there is an 11,5% chance that the correlation is a type 1 error, and the relationship could have occurred by chance. However, the 0,101 Correlation coefficient means that if there isn't a Type 1 error then there is a weak positive relationship between the average price of GME and the daily volume. This means that the more transactions during the day are more likely the transactions to be buying the stock, driving the price upwards.

Secondly, both the Short interest and the GME mentions have very low significance levels. This way these relationships have a very low chance, less than 1% chance that the relationship is by chance, and a type 1 error. The short interest correlation coefficient tells me that there is a strong negative relationship between the short interest and the share price. This means that the share prices increases, when the short interest decreases, this is due to in order to close shorts the investor needs to buy the shares back, this way driving the price up.

Lastly, there is also a moderate positive correlation between the amount of GME posts and the share price. This can either mean that the share price is up because of the more mentions, or the more mentions can be a result of the share prices going up, and this

relationship is very likely to be bidirectional. I will analyze this further at later stages of my research.

4.3.2 Sentiment regression analysis

In the following part, I will analyze the potential relationships between the different sentiments, and emotions and the GME average daily share price. It is very important to mention that this part will be categorized into subgroups since due to the manual modifications conducted on the data, the regression analysis differs in values between the two data sets.

4.3.2.1 VADER sentiment regression analysis

Firstly, I am going to start off with the analysis of the VADER sentiments, these are illustrated in Figure 25. The different sentiments are positive, neutral, negative, and compound. The compound variable is a single variable created to generate an overall sentiment based on the other 3 values.

	GME original		GME modified	
	Coefficients	Significance	Coefficients	Significance
Positive	0,78	0,229	0,119	0,065
Neutral	-0,53	0,408	-0,99	0,126
Negative	-0,16	0,806	-0,67	0,301
Compound	0,154	0,017	0,153	0,017

25. Figure VADER sentiment regression analysis outputs for GME Source: Túll András

The positive, neutral and negative sentiment in the original data has very high significance levels, and thus has a high risk of producing a type 1 error, producing a nonexisting relationship due to chance. These values are statistically insignificant, however, it is visible, that the coefficient for the positive sentiment is associated strongly and positively, with the GME average price. This means, that if the relationship is real, then the increase in the positive sentiment across the comments results in an increase in the share price. The neutral sentiment has a moderately negative relationship with the share price, while there is a weak negative relationship between the negative sentiment and the share price, meaning that with the increase of these sentiments the share price will decrease. It is important to mention that these variables have a high possibility of having a bidirectional relationship with the average daily GME share price. Which will be further analyzed in the later stages of my thesis. On the contrary, the compound value is statistically significant and has a weak positive relationship with the average daily GME share price. From this, we can derive that in the case of the original data, the

compound variable is positively associated with the movement of the share price and that the aggregator value of the 3 emotions is positively correlated. The average sentiment is positively correlated with the average share price. The possibility of a bidirectional relationship is also important to mention in the case of the compound variable since its components have a possibility of having a bidirectional relationship.

Secondly, the interpretation needs to be repeated for the modified data, after which I will also compare my findings.

The positive sentiment is statistically significant at the 90% confidence interval, and there is a weak positive relationship between the share price and the positive sentiment. The neutral and negative values have a lower significance level, but these variables are still statistically insignificant. The neutral sentiment has a significance level close to the 90% confidence interval, and has a negative relationship with the average share price, while the negative sentiment, has a statistically insignificant confidence level of about 70%, and has a weak negative relationship with the share price. The compound value is almost the same as the value of the original data, with a very similar significance level. It is also important to mention that the potential bidirectional relationship is also a possibility for the modified data as well.

4.3.2.2 T2E sentiment regression analysis

In the following part of the interpretation of the statistical analysis, I will analyze the Text to Emotion sentiment's relationship with the average price of GME.

	GME original		GME modified	
	Coefficients	Significance	Coefficients	Significance
angry	-0,148	0,022	-0,148	0,021
fear	0,195	0,002	0,118	0,066
happy	-0,17	0,798	0,065	0,317
sad	0,123	0,056	0,1	0,122
surprise	0,174	0,007	0,121	0,061

26. Figure T2E sentiment regression analysis outputs for GME Source: Túű András

As it can also be seen in Figure 26, out of the 5 different variables, 4 are statistically significant, while happiness is the only one that is insignificant. The angry and happy sentiment has a weak negative relationship with the share price in the original data. While fear, sadness, and surprise have a weak positive relationship with the average daily share price. These results are very unexpected since fear and sadness are not usually associated with an increase, but rather with a decrease. The modified data has only 3

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

statistically significant variables, fear, anger, and surprise. Anger has a weak negative relationship, while all the other variables have a weak positive relationship.

4.4 AMC analysis

4.4.1 Volume, short interest, mentions

In the following part, I will analyze the regressions of the AMC average daily price and its correlation with the independent variables AMC daily traded volume, the short interest in AMC, and the daily mentions of the AMC ticker on the wallstreetbets subreddit. Furthermore, I have conducted the analysis, with both the original AMC data and the modified one, even though I have not modified the data being analyzed here, this will be important in the analysis of the sentiment regressions. The 27. Figure represents my findings.

	AMC original		AMC modified	
	Coefficients	Significance	Coefficients	Significance
AMC daily volume	0,597	0,001	0,597	0,001
AMC Short Interest	0,614	0,001	0,614	0,001
AMC mentions	0,237	0,001	0,237	0,001

27. Figure AMC volume, short interest, mentions regression analysis output Source: Túű András

Firstly, there is a moderate positive correlation between the Daily volume of AMC and the Average daily price of AMC, the correlation coefficient is around 0,6. This means that the more transactions during the day are more likely the transactions to be buying the stock, driving the price upwards.

Secondly, all three variables have very low significance levels. This way these relationships have a very low chance of less than 1% to produce a type 1 error. The short interest correlation coefficient tells me that there is a strong positive relationship between the short interest and the share price. This is the opposite of what I could see with the analysis of the GME regressions. The increase in short interest means that with the increase of short positions the price increases and has a moderate positive correlation.

Lastly, there is also a moderate positive correlation between the amount of AMC posts and the share price. This can either mean that the share price is up because of the more mentions, or the more mentions can be a result of the share prices going up, and this relationship is very likely to be bidirectional. I will analyze this further at later stages of my research.

4.4.2 Sentiment regression analysis

4.4.2.1 VADER sentiment regression analysis

Firstly, I am going to start off with the analysis of the VADER sentiments, these are the first 4 rows, with their respective values, the different sentiments are positive, neutral, negative, and compound. The compound variable is a single variable created to generate an overall sentiment based on the other 3 values. The results are visible in Figure 28.

	AMC original		AMC modified	
	Coefficients	Significance	Coefficients	Significance
Positive	0,046	0,472	0,084	0,191
Neutral	-0,38	0,553	-0,071	0,269
Negative	-0,002	0,981	-0,043	0,501
Compound	0,106	0,099	0,109	0,09

28. Figure VADER sentiment regression analysis outputs for AMC Source: Túú András

The positive, neutral and negative sentiment in the original data has very high significance levels, and thus has a high risk of producing a type 1 error, producing a nonexisting relationship due to chance. These values are statistically insignificant, however, it is visible, that the coefficient for the positive sentiment is associated weakly and positively, with the AMC average price. This means, that if the relationship is real, then the increase in the positive sentiment across the comments results in an increase in the share price. The neutral sentiment has a moderately negative relationship with the share price, while there is a weak negative relationship between the negative sentiment and the share price, meaning that with the increase of these sentiments the share price will decrease. It is important to mention that these variables have a high possibility of having a bidirectional relationship with the average daily AMC share price. Which will be further analyzed in the later stages of my thesis.

On the contrary, the compound value is statistically significant and has a weak positive relationship with the average daily AMC share price. From this, we can derive that in the case of the original data, the compound variable is positively associated with the movement of the share price and that the aggregator value of the 3 emotions is positively correlated. The average sentiment is positively correlated with the average share price. The possibility of a bidirectional relationship is also important to mention in the case of the compound variable since its components have a possibility of having a bidirectional relationship.

As for the modified data, the positive, negative, and neutral data have very similar coefficients to the ones in the original data, and even though the significance values are lower, these values are still too high to be able to say that there is a relationship between the values and the average AMC share price.

The compound value is almost the same as the value of the original data, with a very similar significance level. It is also important to mention that the potential bidirectional relationship is also a possibility for the modified data as well.

4.4.2.2 T2E sentiment regression analysis

In the following part of the interpretation of the statistical analysis, I will analyze the Text to Emotion sentiment's relationship with the average price of AMC. Out of the 5 different variables, 2 are statistically significant, fear and surprise being the only statistically significant. The results can be seen in Figure 29.

	AMC original		AMC modified	
	Coefficients	Significance	Coefficients	Significance
angry	-0,09	0,165	-0,086	0,18
fear	0,154	0,016	0,09	0,164
happy	-0,047	0,462	0,031	0,634
sad	0,093	0,151	0,073	0,256
surprise	0,133	0,038	0,095	0,139

29. Figure T2E sentiment regression analysis outputs for AMC Source: Túű András

The angry and happy sentiment has a weak negative relationship with the share price in the original data. While fear, sadness, and surprise have a weak positive relationship with the average daily share price. These results are very unexpected since fear and sadness are not usually associated with an increase, but rather with a decrease. The modified data has no statistically significant relationships, and the coefficients of the variables except for happiness have stayed in similar territories.

4.5 Hypothesis testing

H1A: Based on the analysis of the relationship of the compound value in both the original and the modified data, the correlation coefficient indicated a weak positive relationship. I can accept the hypothesis based on the compound data being statistically significant, and say that there is a weak positive relationship between the GME average stock price and the compound score, in both the original and the modified data tests. As for the Text to Emotion coefficients, the data indicates, that in the original dataset there are 4 out of 5 variables that have a statistically significant relationship with the average daily price of GME.

H1B: Based on the analysis of the compound value in both the original and modified data, the correlation coefficient indicates a weak positive relationship, I can accept the hypothesis based on the compound value having a statistically significant relationship with the AMC average daily share price. Only two of the Text to Emotion sentiments were statistically significant, with fear and surprise both having weak positive relationships with the average share price. It is worth mentioning that in both the original and modified datasets, the compound score had a very similar relationship and significance levels.

H2A: The GME share price has a statistically significant, strong negative relationship with the Short Interest of GME, with a confidence level of less than 1%. Based on this I can accept the hypothesis that the GME short interest affected average daily GME prices. The hypothesis stated that a short-squeeze has happened and the strong negative relationship confirms that, the expectation during a short-squeeze is an increase in the price of the instrument and a decrease in the short interest as investors with short positions, buy back the shares for their lenders.

H2B: There is a moderate positive relationship between the average price of the AMC stock and the AMC short-interest, however, a short squeeze is based on the price movement due to the decrease in short interest, and simultaneously large amounts of investors buying the share resulting the share price to increase.

4.6 Bidirectional relationships

In the following part, I will discuss the potential bidirectional relationship mentioned during the analysis of the regression of sentiments and the stock price of both GME and AMC.

In a research that has been conducted by (Long et al., 2021), the researchers also analyzed the sentiment of the Reddit thread, with the same emotion extracting tools and financial data, and even though in the research the data was not modified, the potential bidirectional relationship was also concerning the authors. Thus, they conducted a Granger causality test, analyzing whether the relationship between the price of GME and sentiment has been an outcome of the changes in the price of Gamestop. Essentially, meaning that the posts on Reddit are a result of the movement of the share price. Thus the people writing these posts are just content creators and are not sharing their original thoughts on the stock price and company, rather sharing on the forum what happened that day. The results of this Granger causality test were used to determine, whether this relationship was in fact in the other direction. The only relationship, that was bidirectional according to the research conducted was between the intraday share price and the emotion happy. Furthermore, the relationship between happy and the share price has been concluded statistically insignificant in 3 of the 4 regression analyses. And will be excluded from the conclusion of the analysis because of this relationship.

4.7 Limitations of the research

I would like to point out that similarly, to the research of (Long et al., 2021), there are limitations and assumptions in the case of my research as well. I have not distinguished between, the posts that were written by non-investor discussion participants, and investing participants. Also, while I have modified the data to have a more accurate model, the limitations of the packages used in my research are apparent, and there is a high possibility, that the extracted sentiment from each post is not completely accurate. Furthermore, I made my research under the assumption, that all individual posts are genuine and there are no bots, in the discussion.

V. Summary

In conclusion, the aim of my thesis work was to find out, the reasons behind the extreme movements in the prices of two, „meme stocks“. The two main questions of my thesis were, whether there was fundamental value or a short squeeze that was causing the sudden increase in share prices, or whether the sentiment of the Reddit thread wallstreetbets and retail investors took a place in inflating the share price. At the start, I have concluded the timeline, in which I have analyzed the Gamestop and AMC share prices, and short interest and volume traded respectively.

I have further analyzed the influences of the COVID-19 pandemic during the time period, which was the largest influence at the time, having serious economical consequences, thus influencing the movements on stock markets all around the world. I have also analyzed the GME price movements further, with the recent large short-squeeze causing price movements of the Volkswagen share back in 2008.

As for the methodology I have used to further analyze the two most important shares, In the fundamental analysis of both companies, the results were, that is the absolutely ideal and best-case scenarios, the shares are very overpriced. I have concluded after the fundamental analysis that there are no underlying, overlooked reasons that give reason to believe that the companies have intrinsic value.

In the second part of the analysis, regarding the sentiment of the Reddit subgroup Wallstreetbets, I extracted the data using web scraping and collected a dataset of around 170000 individual posts, after excluding duplicate messages. I further separated the data into subgroups, to have a more accurate dataset I have cleaned the data using phrases and slang words that are used in the Reddit subgroup. I have analyzed the sentiment using python packages that are used to extract sentiment from sentences. The dominant sentiment using the VADER package on both the original and modified data was neutral, while a key difference can be seen here between the two charts, the positive and compound values have a bigger value in the modified data. In the analysis using the Text2Emotion package, the dominant emotion was surprise in the original dataset, while sadness, fear, and happiness were also significant. It is interesting to see that contrary to my expectations fear was a more frequent emotion, than happy, while it is also important to mention that the emotion angry was almost insignificant in both the original and the modified dataset. I expected the angry emotion to be more frequent due to the hate against hedge funds, that shorted the stocks and Robinhood, which halted trading.

In the GME analysis part, I have found that the increase in mentions on Reddit resulted in a positive change in the average share price, also, the decrease of the short-interest resulted in the increase in the average share price, thus for the GME share, I can validate, that there was indeed a short squeeze, that drove up the share price.

In the case of the VADER analysis, the compound sentiment output had a statistically significant, positive relationship with the average share price, in both the original and the modified data, while the other sentiment outputs were either statistically insignificant or excluded from the analysis due to the bidirectional relationship. The Text2Emotion analysis concluded controversial results in both the original and the modified data since fear and sadness had a positive relationship with the share price, which was unexpected and can be a result of the complexity of the package, which in this difficult environment, with slang words and phrases is not that accurate. This could be caused by the sentiment having no influence on the share price, but since the VADER analysis gave the statistically significant positive compound value, I will conclude, that there is a relationship between the sentiment and the share price, while because of the difficulties in the dataset mentioned above, the Text2Emotion package did not result in accurate outputs.

In the AMC analysis part, I have found that with the increase in volume and mentions the share price increases, while the Short Interest has a positive relationship, with the share price, thus I reject the hypothesis, that there was a short squeeze in the case of AMC. However, with the positive relationship between the compound value and the share price, it is very likely that the sentiment-driven price increase is responsible for the increase in the price of the AMC share. In the Text2Emotion analysis, the values are very similar to those in the case of GME.

As for the data I have used, I have used the posts where both GME and AMC are mentioned, thus the sentiment analysis inputs are the same. However all the other financial data is different, this can give me a picture of how much the sentiment has driven up each share price during the period, while in the case of GME there was indeed a short-squeeze, in the case of AMC there was not. Following this idea the difference between the share increase in share prices can give me a picture about how much sentiment drove up share prices. I also have to say, that the given share price is based on multiple factors, and that this gives no proof that the difference in increase of the AMC share price and the GME share price, is the share price increase caused by the short squeeze.

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

To sum it up, in the case of GME there was indeed a short-squeeze, while in the case of AMC the share price, I cannot conclude that the share price was driven up, by the decrease in short interest. In both cases based on the VADER sentiment analysis, I can conclude, that the overall sentiment had a statistically significant relationship, with both of the average share prices.

VI. References:

- AMC Entertainment Holdings, Inc. Announces Shareholder Count. (2021, June 9). AMC Theatres. Retrieved November 10, 2021, from <https://investor.amctheatres.com/newsroom/news-details/2021/AMC-Entertainment-Holdings-Inc.-Announces-Shareholder-Count/default.aspx>
- Brian, S. (2020, January 28). Coronavirus stock market crash may have created a once in a lifetime buying opportunity: strategist. Retrieved May 14, 2021, from <https://finance.yahoo.com/news/coronavirus-stock-market-crash-of-2020-may-have-created-a-once-in-a-lifetime-buying-opportunity-strategist-173608909.html>
- Brinca, P., Duarte, JB, & E Castro, MF (2020). Is the COVID-19 Pandemic a Supply or a Demand Shock? *Economic Synopses*, 2020 (31). <https://doi.org/10.20955/es.2020.31>
- Chohan, Usman W. and Chohan, Usman W., Counter-Hegemonic Finance: The Gamestop Short Squeeze (January 28, 2021). Available at SSRN: <https://ssrn.com/abstract=3775127> or <http://dx.doi.org/10.2139/ssrn.3775127>
- Chowdhury, E. K., Khan, I. I., & Dhar, B. K. (2021). Catastrophic impact of Covid-19 on the global stock markets and economic activities. *Business and Society Review*. Published. <https://doi.org/10.1111/basr.12219>
- David, S., Inácio Jr., C., & Tenreiro Machado, J. A. (2021). The recovery of global stock markets indices after impacts due to pandemics. *Research in International Business and Finance*, 55, 101335. <https://doi.org/10.1016/j.ribaf.2020.101335>
- FRED Economic Data. (2021, May 3). Retrieved May 14, 2021, from <https://fred.stlouisfed.org/series/FEDFUNDS>
- Fundamental Diagnosis (2021b, februari 2). Examining The Fundamental Value Of GameStop. SeekingAlpha. Geraadpleegd op 28 april 2022, van <https://seekingalpha.com/article/4402732-examining-fundamental-value-of-gamestop>
- Hasso, T., Müller, D., Pelster, M., & Warkulat, S. (2022). Who participated in the GameStop frenzy? Evidence from brokerage accounts. *Finance Research Letters*, 45, 102140. <https://doi.org/10.1016/j.frl.2021.102140>
- Internationalbanker. (2021, September 29). The Volkswagen Short Squeeze (2008). Retrieved October 25, 2021, from <https://internationalbanker.com/history-of-financial-crises/the-volkswagen-short-squeeze-2008/>
- Jones, Charles M. and Reed, Adam V. and Waller, William, When Brokerages Restrict Retail Investors, Does the Game Stop? (November 18, 2021). Columbia Business School Research Paper Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3804446> or <http://dx.doi.org/10.2139/ssrn.3804446>
- Lake, R. (2021, March 8). What Is a Gamma Squeeze? Retrieved May 14, 2021, from <https://smartasset.com/investing/gamma-squeeze>
- Li, Y. (2021, February 2). As GameStop plunges, Volkswagen's 2008 short squeeze gives an idea of how painful it will get. CNBC. Retrieved October 25, 2021, from <https://www.cnbc.com/2021/02/02/as-gamestop-plunges-volkswagens-2008-short-squeeze-gives-an-idea-of-how-painful-it-will-get.html>
- Long, C., Lucey, B. M., & Yarovaya, L. (2021). "I just like the stock" versus "Fear and loathing on main street" : The role of reddit sentiment in the GameStop short squeeze. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3822315>
- Lyócsa, T., Baumöhl, E., & Výrost, T. (2022). YOLO trading: Riding with the herd during the GameStop episode. *Finance Research Letters*, 46, 102359. <https://doi.org/10.1016/j.frl.2021.102359>

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

McIntyre, D. A. (2020, March 28). The 10 American Companies With the Most Cash. Retrieved May 14, 2021, from <https://247wallst.com/investing/2020/03/28/the-10-american-companies-with-the-most-cash/2/>

McKibbin, W., & Fernando, R. (2020). The economic impact of COVID-19. *Economics in the Time of COVID-19*, 45.

Pedro Brinca, Joao B. Duarte, and Miguel Faria e Castro, "Is the COVID-19 Pandemic a Supply or a Demand Shock?," *Economic Synopses*, No. 31, 2020. <https://doi.org/10.20955/es.2020.31>

Short Squeeze. (2021, January 28). Retrieved May 14, 2021, from <https://www.investopedia.com/terms/s/shortsqueeze.asp>

Thomson ONE.com. (2021) *Thomson ONE.com. Thomson Reuters*. [Online]. Available at: <http://www.thomsonone.com/> (Accessed: May 14, 2021)

Thorbecke, C. (2021, February 13). GameStop timeline: A closer look at the saga that upended Wall Street. Retrieved May 12, 2021, from <https://abcnews.go.com/Business/gamestop-timeline-closer-saga-upended-wall-street/story?id=75617315>

Umar, Z., Gubareva, M., Yousaf, I., & Ali, S. (2021). A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance*, 30, 100501. <https://doi.org/10.1016/j.jbef.2021.100501>

Vashi, A. (2021, September 18). *AMC: An Objective Assessment*. Seeking Alpha. Retrieved November 10, 2021, from <https://seekingalpha.com/article/4450228-amc-stock-an-objective-assessment>
Viceni Investing. (2021, February 2). *Examining The Fundamental Value Of GameStop*. SeekingAlpha. Retrieved November 2, 2021, from <https://seekingalpha.com/article/4402732-examining-fundamental-value-of-gamestop>

W. S. Journal, (2021, November 1). *GME / GameStop Corp. C1 A Analyst Estimates & Rating – WSJ*. GameStop Corp. C1 A. Retrieved November 1, 2021, from <https://www.wsj.com/market-data/quotes/GME.EXA/research-ratings>

VII. List of figures

1. Figure Gamestop price, volume, short interest, Beta chart 2021 January-2021 March Source: Thomson One	3
2. Figure Gamestop price, volume, short interest, Beta chart 2021 January-2021 March Source: Thomson One	4
3. Figure Gamestop price, volume, short interest, Beta chart 2021 March-2021 May Source: Thomson One	5
4. Figure S&P 500 chart February 24th to March 23rd Source: Thomson One	6
5. Figure Returns in largest indices during the Covid-19 pandemic Source: Chowdhury, Khan, and Dhar (2021).....	8
6. Figure U.S. corporations with largest cash reserves Source: Thomson One.....	9
7. Figure FED interest rate in the U.S. Source: FRED Economic Data	9
8. Figure Gamestop stock price, volume, and short interest Source: Thomson One.....	11
9. Figure AMC stock price, volume, and short interest Source: Thomson One	11
10. Figure Volkswagen and Gamestop stockprice at the short squeeze Source: Internationalbanker, 2021	12
11. Figure Key phrases from Reddit Wallstreetbets Source: Long, C., Lucey, B. M., & Yarovaya, L. (2021). "I just like the stock" versus "Fear and loathing on main street"	16
12. Figure Gamestop sales breakdown by product category Source: Research, Gamestop yearly reports	22
13. Figure Gamestop sales breakdown by product category from 2018 to 2020 Source: Fundamental Diagnosis, 2021.....	23
14. Figure Gamestop operating margin Source: Fundamental Diagnosis, 2021	23
15. Figure Best Buy operating margin Source: Fundamental Diagnosis, 2021	23
16. Figure Walmart operating margin Source: Fundamental Diagnosis, 2021.....	23
17. Figure AMC Revenue Source: (Vashi, 2021)	25
18. Figure AMC Discounted Cash Flow analysis input estimations Source: Vashi, 2021	26
19. Figure Average daily sentiment using the VADER package and the original data Source: Túú András.....	28
20. Figure Average daily mentions Source: Túú András.....	29
21. Figure Average daily sentiment using the VADER package and the modified data Source: Túú András	30
22. Figure Average daily sentiment using the T2E package and the original data Source: Túú András.....	31
23. Figure Average daily sentiment using the T2E package and the modified data Source: Túú András	32
24. Figure GME volume, short interest, mentions regression analysis output Source: Túú András	33
25. Figure VADER sentiment regression analysis outputs for GME Source: Túú András	34
26. Figure T2E sentiment regression analysis outputs for GME Source: Túú András..	35
27. Figure AMC volume, short interest, mentions regression analysis output Source: Túú András	36
28. Figure VADER sentiment regression analysis outputs for AMC Source: Túú András.....	37
29. Figure T2E sentiment regression analysis outputs for AMC Source: Túú András .	38
30. Figure Descriptive statistics for variables of the original data Source:Túú András, SPSS.....	48

The idea behind the "Meme stocks, their rise, and connection, with the Reddit thread.

31. Figure Descriptive statistics for variables of the original data Source:Túú András, SPSS.....	48
--	----

VIII. Appendix

8.1 Descriptive statistics for original data

Descriptive Statistics							
	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Skewness Statistic	Std. Error
angry	242	,00000000000	1,00000000000	,07526537846	,10312064200	5,179	,156
happy	242	,00000000000	1,00000000000	,07249974538	,10440976582	6,287	,156
fear	242	,00000000000	1,00000000000	,22030187281	,12219205023	1,166	,156
sad	242	,00000000000	,50000000000	,11196921952	,07634465695	1,222	,156
surprise	242	,00000000000	1,00000000000	,09559506215	,08440070115	5,392	,156
Positive	242	,00000100000	,33300100000	,07906772192	,04788096837	1,957	,156
Neutral	242	,54100100000	1,00000100000	,87498083631	,05751815095	-1,147	,156
Negative	242	,00000100000	,45900100000	,04587519408	,04485938444	4,968	,156
Compound	242	-,52669900000	,54173433333	,05082137690	,10055316932	-,787	,156
GME avg p	242	6,17500000000	352,355000000	79,945048554	86,076202514	1,044	,156
GME vol	242	2478168	197157946	22153850,09	33608498,104	3,365	,156
AMC avg p	242	1,98500000000	28,965000000	6,5790409091	4,3724190793	1,710	,156
AMC volume	242	2090061	1253253552	86477559,81	150004839,27	3,780	,156
GME SI	242	16,50	115,30	71,6394	43,28201	-,293	,156
AMC SI	242	3,46000000000	19,940000000	11,650330579	5,3010799704	,129	,156
mentions	242	1	27472	639,34	2586,294	7,046	,156
Valid N (listwise)	242						

30. Figure Descriptive statistics for variables of the original data Source:Tűű András, SPSS

8.2 Descriptive statistics for modified data

Descriptive Statistics							
	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Skewness Statistic	Std. Error
angry	242	,00000000000	1,00000000000	,06648106984	,09992909143	5,603	,156
happy	242	,00000000000	1,00000000000	,25052335052	,15801953954	1,936	,156
fear	242	,00000000000	1,00000000000	,17731318218	,10949920193	2,248	,156
sad	242	,00000000000	,40000000000	,08931195733	,06598627147	1,418	,156
surprise	242	,00000000000	1,00000000000	,07599456889	,07911444028	6,882	,156
Positive	242	,00000100000	1,00000000000	,25107086495	,14178197983	1,612	,156
Neutral	242	,00000000000	1,00000100000	,70853799655	,13862839667	-1,271	,156
Negative	242	,00000000000	,45900100000	,04037288645	,04758042260	5,077	,156
Compound	242	-,52669900000	1,00000000000	,22196912606	,17286295381	,313	,156
GME avg p	242	6,17500000000	352,355000000	79,945048554	86,076202514	1,044	,156
GME vol	242	2478168,0000	197157946,00	22153850,091	33608498,104	3,365	,156
AMC avg p	242	1,98500000000	28,965000000	6,5790409091	4,3724190793	1,710	,156
AMC volume	242	2090061,0000	1253253552,0	86477559,810	150004839,27	3,780	,156
GME SI	242	16,50	115,30	71,6394	43,28201	-,293	,156
AMC SI	242	3,46	19,94	11,6503	5,30108	,129	,156
mentions	242	1	27472	639,34	2586,294	7,046	,156
Valid N (listwise)	242						

31. Figure Descriptive statistics for variables of the original data Source:Tűű András, SPSS