

Wallstreetbets during the GME Hype

A data science project

By: Daniel Rosenberger

GitHub: <https://github.com/DanielRbgr>

Version 1.0

xx.yy.2021

Contents

0.	Foreword	3
1.	Introduction.....	3
2.	Getting the data	3
2.1.	The Reddit posts.....	3
2.2.	The Subreddit member count	3
2.3.	Financial Data	3
3.	The Wallstreetbets Subreddit	4
4.	The Short Squeeze Stocks.....	6
5.	The one who started it: DeepFuckingValue	6

0. Foreword

This paper is the result of me trying some data science around the subreddit “Wallstreetbets”, or in short “WSB”. I visit the subreddit regularly and enjoy the posts, even though I wasn’t a member until recently. I started watching all the events around the Short Squeeze from Gamestop and other stocks early on and was fascinated by the events. I wanted to try a little analysis on what the numbers have to say about this incident and thanks to multiple APIs, it was rather easy to get my hands on the relevant data. As a brief warning: I am new in data science, so there are likely some errors and other flaws in the following text as well as unnecessary complicated and inefficient code parts in the source code. This is by no means any scientific paper, it is just a documentation of a private project for people who are interested in the topic.

All the data and the source code I used for this project can be found on my [GitHub](#), so feel free to give it a try as well if you want. I am also open for any feedback, so if you find flaws or have further interesting ideas for the project, feel free to contact me. I hope you find the following pages interesting and amusing.

1. Introduction

2. Getting the data

2.1. The Reddit posts

Since it is not possible to download post older than 24h from the API provided by Reddit itself, the Wallstreetbets posts were downloaded from the [pushshift.io](#) API which allows to download older posts. A python package for the pushshift API and an instruction on how to use it can be found on the [Pushshift GitHub](#) page.

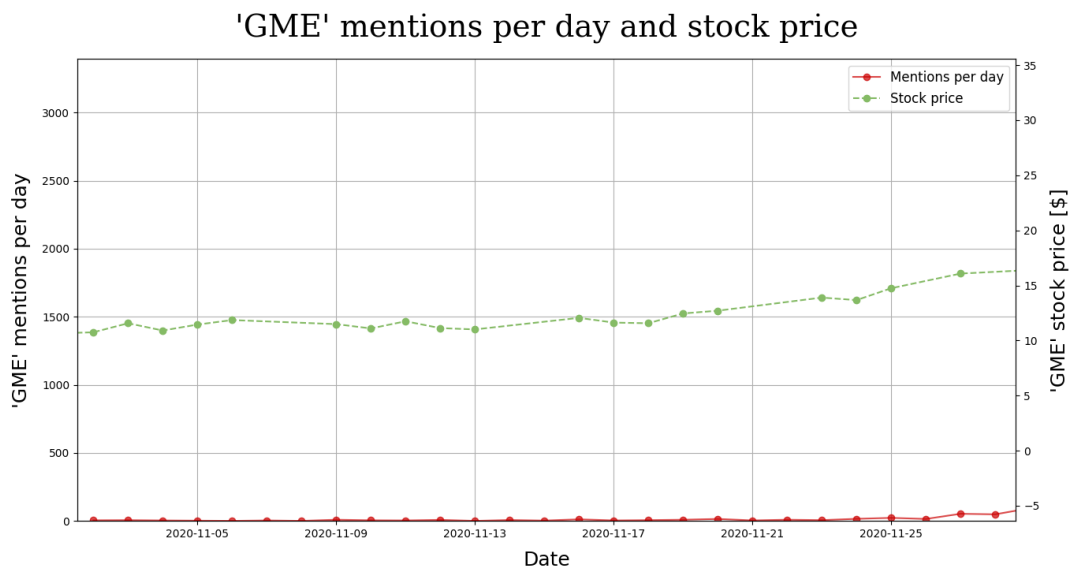
Starting from 1st November 2020, all Wallstreetbets posts were downloaded, each post with the data on creation time (in the utc time format), title, author and the URL. Pushshift also provides information for likes and comments on each posts, sadly these were outdated and thus wouldn’t be reliable, so they were ignored. The data is stored in a CSV file, the CSV is also extendable with never posts, so it is simple to keep the CSV up to date. There has to be mentioned that there is a timeframe of a couple of days where pushshift returns no posts, the last post before the gap being dated to 4th of February 11:47:43pm and the first post after the gap is dated to 8th of February 00:21:21am, meaning there is a gap of about 3 days where no posts are available. Since the timeframe falls into the cooldown phase after the hype, there should be no heavy loss of overall information since an interpolation should solve this problem, nevertheless it is annoying but until now there is no way to achieve these data.

2.2. The Subreddit member count

The member data was downloaded from [subredditstats.com](#), it is not complete though. The missing data mainly appears in two periods of time, the first couple of days in November and the big member growth phase in the second half of January. To complete the data, the linear method of the pandas `interpolate` function was used.

2.3. Financial Data

For the stock price data, the data origins from Yahoo! Finance and is accessed via the *yfinance* API. It's simple to use and is implemented by only a few lines of code. The *yfinance* library with the instructions and download options can be found [here](#). The *yfinance* API will return many different values per stock and day, four different prices: Open, Close, High and Low. All stock prices in the diagrams shown in this documents refer to the Close price. This price was chosen under the assumption, that if the comments and posts on reddit have any influence on a stock price, all this influence would aggregate in the Close price. Another thing to mention is that the chart could sometimes look unsimilar like in the following picture.

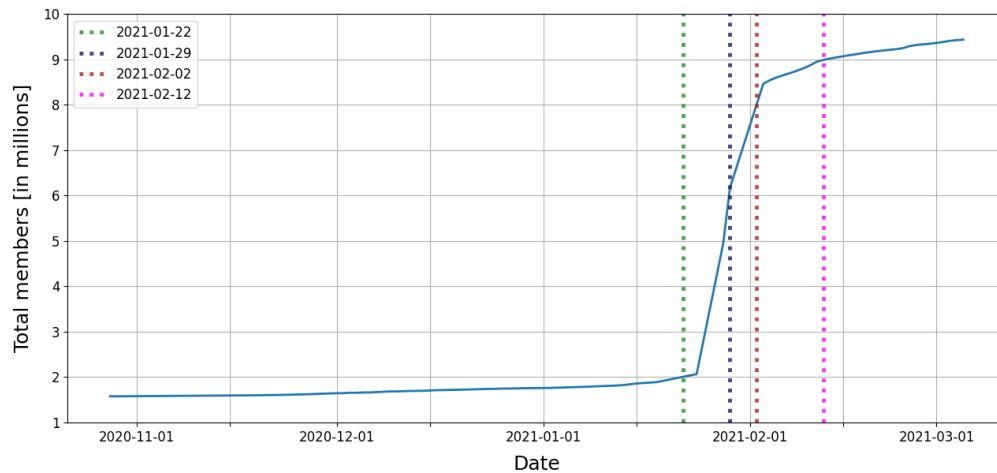


Each dot on a graph represents a data point and it is obvious that the stock price graph is missing some values in comparison to the mentions graph. This gaps in the graph represent the weekend, where the stock markets are closed and thus, there are no new stock prices. There are some singular days missing as well, like at the 26th November 2020 in the picture above, but these are only a few and are due to some missing values from *yfinance*. In all diagrams, the gaps are simply ignored and there are some longer linear connections between the data points before and after the gap. So, if some diagrams look odd, it is most likely due to this little flaw in the stock data.

3. The Wallstreetbets Subreddit

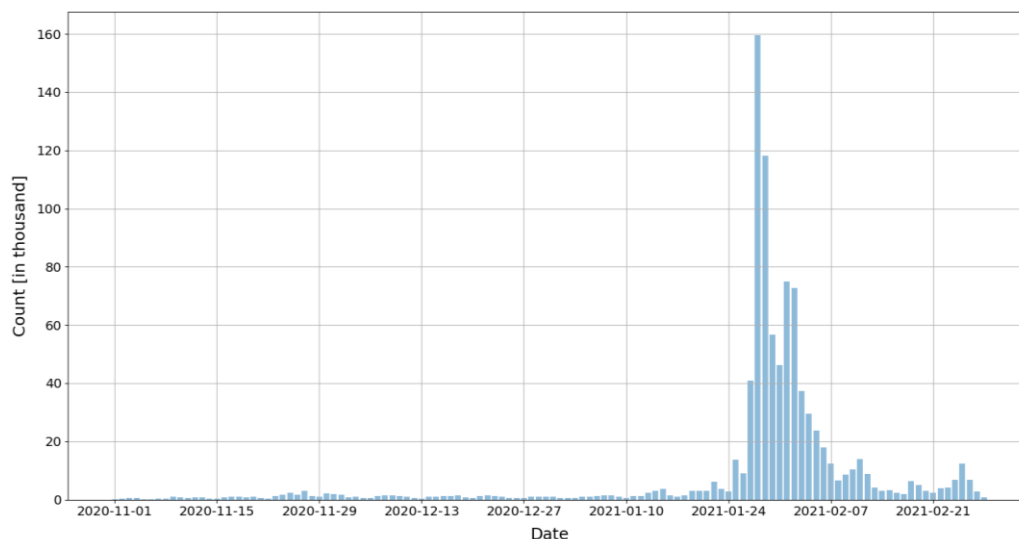
This chapter will give a short overview on what happened to the Wallstreetbets subreddit as a total during the short squeeze. The broad exposure to the media resulted in a massive growth in member counts. On 22nd January 2021, the Subreddit crossed the mark of 2 million members. Then, in a mere week, the member count tripled to 6 million members on 29th January, which equals a growth rate of 16.99% per day, rates every subscription platform on the internet is dreaming of. Interestingly, with the introduction of trade restriction on 28th January, the growth started to slow down, even though the media attention probably skyrocketed since the whole process gained another political and supervisonal dimension. Nevertheless, the Subreddit crossed the 8-million-member mark 4 days later on February 2nd, which means a 7.46% growth rate for this period of time. Another 10 days later, the 9 million members mark was reached, equalling a 1.1% growth rate per day. The overall growth rate from 2 million members to 9 million members equals 7,42%. (Reminder: All the previous mentioned numbers are based on the interpolated data, so the overall magnitude is correct, but the exact values may vary.)

Development of the Wallstreetbets members



It is noticeable that just when the trade restriction was declared on January 28th, and thus the stock price of the hype stocks started to plummet, the member growth started to sink. If these two events are connected to each other, which means that the people got scared by the huge drop in stock prices, is not clear. Maybe the explanation is much simpler than this and by this time, all people willing to join the subreddit were already subscribed. Nevertheless, this growth rates are astonishing and most likely, every internet page with whatever subscribing possibility dreams of this numbers.

Total number of comments



Even bigger than the growth in member was the growth of comments in the subreddit. In late December 2020, the number of comments per day were all in three-digit area. In mid-January, the daily comment counts were in the lower four-digit scope. Then, the numbers skyrocketed. Apart from the little dent around January 23rd till 24th, which was a weekend and therefore the stock markets were closed, the number of comments multiplied by whole numbers each day. And according to the definition of hypes, this dropped nearly as fast, even though the daily comments stayed in the four-digit area. If this increase in platform traffic settles for alle the media attention Reddit received during this time is questionable though. Interestingly enough, the number of

comments rise again at the end of the month, which could be a sign there is some aftermath to follow.

4. The Short Squeeze Stocks

5. The one who started it: DeepFuckingValue

This chapter is about the one that sparked the whole GME short squeeze: the Reddit user [DeepFuckingValue](#). He was the first one posting about Gamestop on Reddit and his [YouTube Channel](#) and received a lot of media attention, which reached it's peak when he had to testify before the United States Congress together with Hedge fund managers and the CEOs of Robinhood and Twitter. His role in the whole process is yet to be resolved. For some, he is a small retail investor who fought the institutional investors victorious. Others view him as a trader who successfully used social media to manipulate a stock price for his own interest. This chapter will examine his investments in GME, how they developed, and which trades he did. Maybe this can help you get a clearer view of his role in the process. Keep in mind, for the sake of objectivity, that the following pages refer to the information he published on Reddit, he may or may not have other, unmentioned positions.

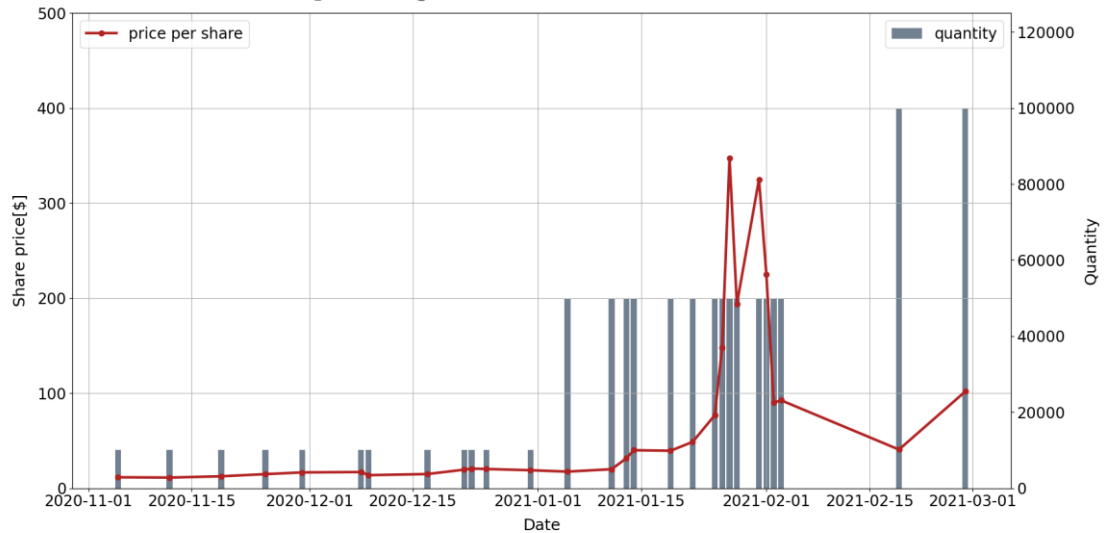
Starting on September 8th, 2019, he started posting regular updates about his GME portfolio on Wallstreetbets, calling them his "GME YOLO updates". Before September 2020, he posted monthly updated, then he posted multiple updates a month and during late January and early February 2021, when the GME rally gathered momentum, he posted an update nearly every day. On February 3rd, he stopped his daily update and went back to a two week rhythm. Since this project examines the time frame starting from November 2020, his older posts won't be displayed here, but you can look them up by simply scrolling down on his reddit page linked above.

In the table below, there is a list with his all his positions including quantity and initial price of all the positions he held at the beginning of November, according to his GME YOLO update from November 5th 2020. He purchased the Call options below between May and August 2020 as follow-up for other Call options he sold because they were about to expire. The first real GME was bought in April 2020, during the next couple of months he stepwise bought more to a final amount of ten thousand shares.

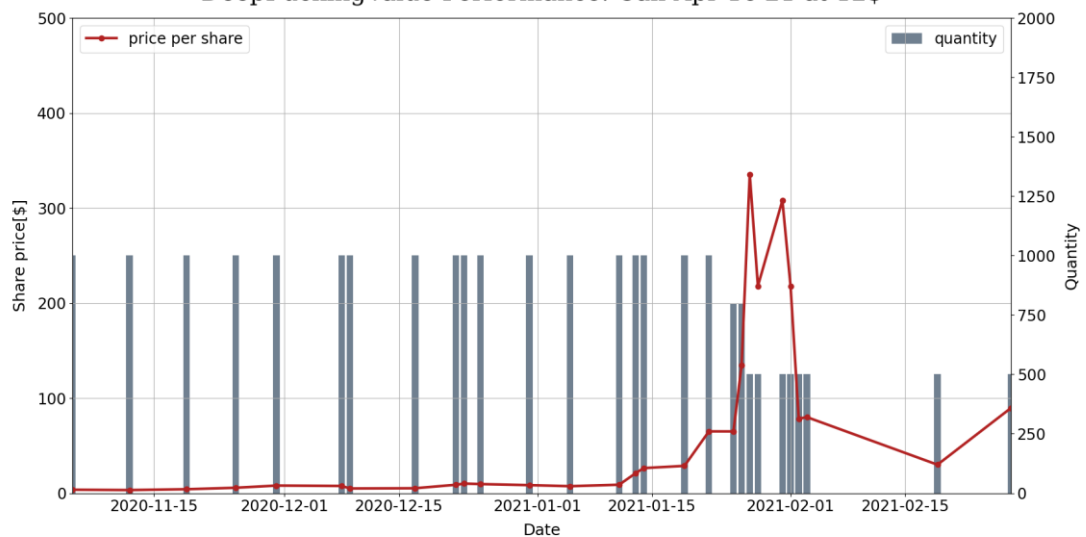
Security paper	Initial price per share [\$]	Initial Quantity	Volume [\$]
GME stock	4.1133	10000	41133
Call: Jan 15'21 at 10\$	0.2	1000	200
Call: Jan 15'21 at 15\$	0.14	1000	140
Call: Jan 15'21 at 17\$	0.15	1000	150
Call: Jan 15'21 at 20\$	0.0726	1000	72.6
Call: Apr 16'21 at 12\$	0.4	1000	400
Total			42096.5

Shown below is the development for every position in his portfolio from above with price per share and quantity of the position. The last four graphs end before February 28th 2021, which is the end of the period examined in this project, meaning he sold his position. He completely sold the positions due in January and reduced the ones due in April by half.

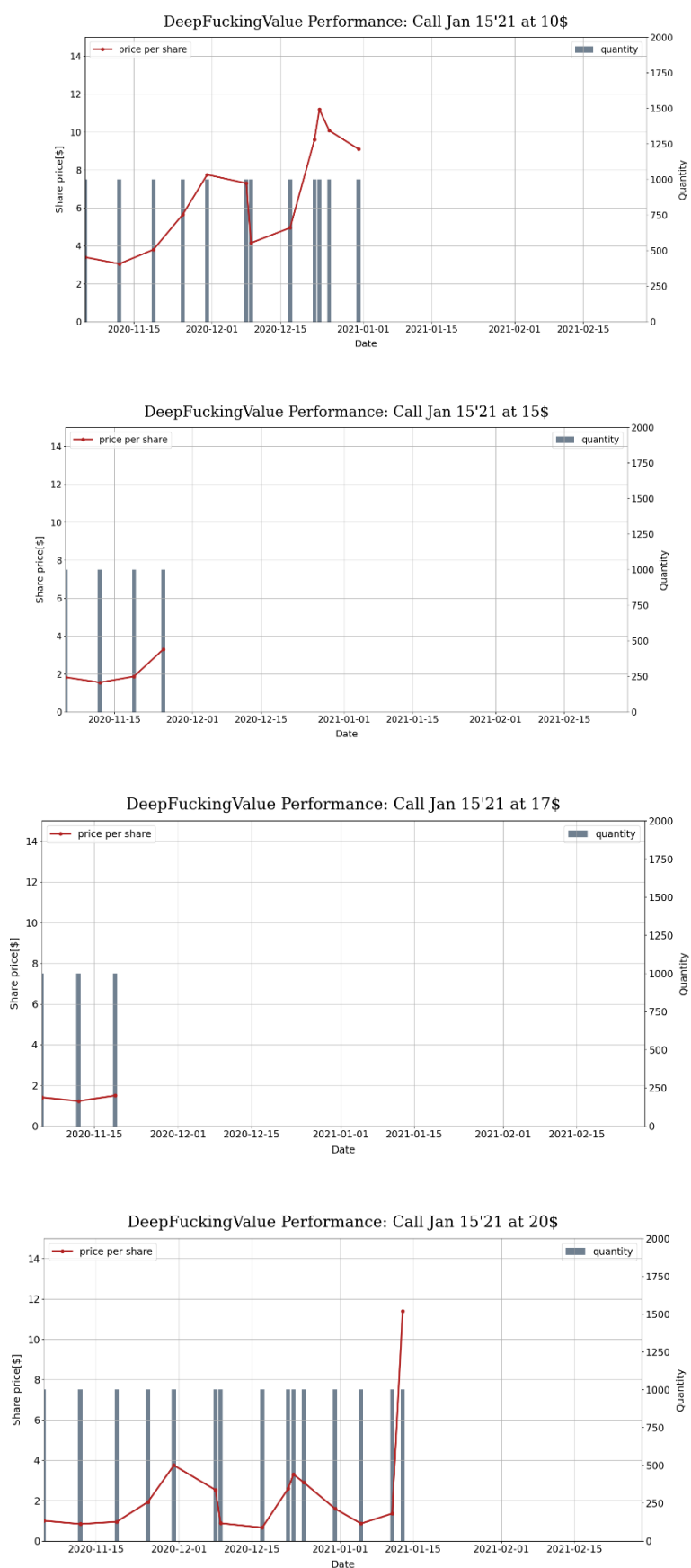
DeepFuckingValue Performance: GME stock



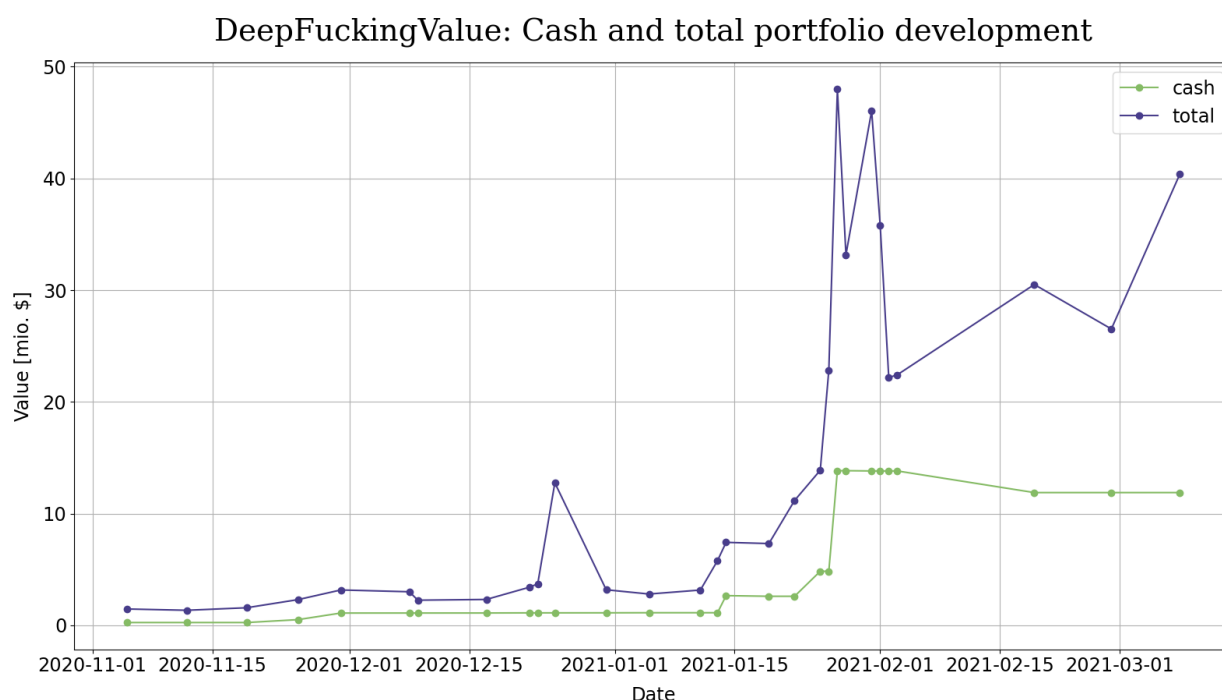
DeepFuckingValue Performance: Call Apr 16'21 at 12\$



For clarity reasons and because they aren't that interesting, these charts are smaller than the ones of the positions he still holds. If you want to have a closer look, the image files are uploaded to GitHub.



You may have noted that he, while he sold his all his due call options and half of his still active options, increased his GME stock position to the tenfold of its initial quantity. He first added four times his initial quantity at the beginning of January and then again doubled his position when the GME stock passed the price peak and dropped to about 40\$. As shown in the picture below, this increase in stock quantity together with the rising GME stock prices at the end of February results in a huge momentum for his portfolio yet again, which is as a result heading back to it's all time high. If the rising price is a sign of another short squeeze or something like this has yet to be evaluated, which will happen in a later version of this project.



The picture above pictures the performance of his cash position and his complete portfolio value (including cash position). For an initial invest of 40k\$, he has a portfolio worth ten of millions of dollars right now. His YOLO bet, like he calls it in his posts, paid off, and this development clearly demonstrates the benefit of “being ahead of the trend”.

Yet, apart from a few thousand dollars from time to time, which could also be portfolio costs, he did not withdraw any of the money in his portfolio. He held on to his position through all ups and downs, losing half of his net worth in a couple of days. It will be interesting to see if the trend at the beginning of march will be just a short High or if it, like some other Redditors claim, is the beginning of the “real short squeeze”.