Graphical user interface, website

Description automatically generated

## Scenario

You have been recruited to a hedge fund. They are concerned that there are new market risks that are currently unforeseen to their investment managers. For example, the rise of the retail investor through applications like RobinHood have invited millions of amateur investors to the market. The additional rise of amateur, unsophisticated investors was not thought of as a risk until many of these investors organized on a Reddit Forum called WallStreetBets. On this forum information and investment ideas are shared.

In early 2021, the forum investors caused a sudden rise in Gamestop’s stock. Many investors there pushed into gamestop causing many brokerages to halt or limit trading. This sudden rise cost professional investors, particularly “short sellers” to lose billions.

Chart, histogram

Description automatically generated

The sudden rise from $20 to $350 within a month.

## Contextual Information

### Short Selling

A “short sell” is when an investor borrows an equity, sells it immediately with the promise to pay it back at a future date. As the future date nears, the short seller must purchase the share to give back to the lender. It is a successful investing strategy when the share price goes down between the borrowing and repurchasing of the stock. For example, an investor would borrow Company A when it is trading at $100, because they believe the stock will be going down in the future. The investor immediately sells the share at $100. 30 days later, the stock may be trading at $75 so the investor purchases the stock at $75 to return to the original lender. In doing so, the investor’s profit is $25 ($100-$75). However, in that timeframe if the stock price goes up, the investor must still return the stock. For example, if 30 days later, the stock is trading at $125, the investor must purchase it to return and as a result lost $25 ($100-$125). Short selling can be very lucrative, and some would say serves a purpose to identify mispriced stocks. Interestingly a short sell investment is asynchronous because the amount of money a short seller could lose is theoretically infinite versus the amount they could gain. A loss can be of any value because a stock can go upwards infinitely yet a profit is limited by the amount a stock can go down. Given this situation, not every investment firm participates in short selling.

### Short Squeezes & the Wall Street Investor Risk

The forum’s millions of readers did not appreciate the “short” investments of a favorite stock called Gamestop, among others. Gamestop is an international retailer focused on gaming. As of February 2021, Gamestop operates 5509 stores in 14 countries selling consumer electronics, collectibles and electronics including new and used merchandise. As a result of Wall Street firms shorting this community favorite equity, many forum authors advocated for purchasing and boosting the Gamestop stock to create a “short squeeze”. A successful short squeeze would cost the professional investors billions of dollars. It’s called a “squeeze” because a sharp jump in the shorted stock will cause the short investor to have to purchase the stock to limit their losses ahead of when they need to return the asset. They are “squeezed” by the rise ensuring their loss and an impending deadline to return the asset. Overall, it is illegal to manipulate the market with non-public information. However, given this is a public forum with access for all, the organization of the forum readers to boost a stock price is not illegal manipulation because anyone including the professional investors could have read the information there.

**As a result, you have been asked to examine the text and look for any early indication that gamestop was gaining traction and that it was poised to move. This “signal” could help the hedge fund monitor this and other public forums to avoid future losses like this.**

## Project Deliverables

Your goal is to identify in historical data the posts related to gamestop. Then using text mining approaches identify how soon an investment firm could have realized that the change in attitude towards gamestop was going to affect the stock price. You need to present an “investment thesis” to the hedge fund stating what signal you identified in the text and how it was created. The thesis being that it would work for Gamestop historically and that the same or similar signal using text mining approaches could be used to monitor the forum for the next significant market impact.

## Data

The data comes directly from the Reddit forum API for the subreddit, WallStreetBets and covers a timeframe leading up to the gamestop stock movement. The data includes the post, and follow-up comments to the post along with meta data.

Text Data:

` CASE\_gme.csv`

Additionally the share price for GME is included as a secondary data set for the timeframe relating to the forum data. You may not need this data but it is provided for context.

GME Stock Data:

` gme.csv`

## Data Dictionary - Text

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Example |
| id | Unique Identifier in the data set. This is the second observation in the dataset. | 2 |
| structure | The relationship to the original ID for a comment. This observation is the first "\_1" comment related to the first "1" observation | 1\_1 |
| post\_date | Original Posting Date | 1/23/21 |
| comm\_date | Original Commenting Date | 1/23/21 |
| upvote\_prop | The proportion of upvote to downvote on the forum | 0.92 |
| post\_score | A return from the Reddit API scoring a post for popularity including views not just up/down votes | 21497 |
| comment | The text of a comment. This is the more interesting column of text because multiple comments will be under a single post title/post text. | Barbarians at the gate… |
| title | The text of a post title. The title will repeat for each comment in the thread. | Boomer Old Money… |
| post\_text | The original post text underneath the title. It will repeat for each comment in the thread. | Alert dd -https://www.bloomberg.com/… |
| link | URL for the original thread | https://www.reddit.com/r/wallstreetbets... |
| post\_date\_yr | Year of the post | 2021 |
| post\_date\_month | Month of the post | 1 |
| post\_date\_day | Day of the month post | 23 |
| post\_date\_weekday | weekday integer of the post | 1 |
| comm\_date\_yr | Year of the comment | 2021 |
| comm\_date\_month | Month of the comment | 1 |
| comm\_date\_day | Day of the month comment | 23 |
| comm\_date\_weekday | Weekday integer of the comment | 1 |

## Data Dictionary – Stock

The trading day, opening stock price, high reached in the trading session, low in the trading session, closing price and volume for the equity. The adjusted column represents after trading session adjustments resulting from company and institutional investor actions. If used in this case, the “close” price or volume is likely the columns of note. Example data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| date | GME.Open | GME.High | GME.Low | GME.Close | GME.Volume | GME.Adjusted |
| 5/31/19 | 7.47 | 7.66 | 7.32 | 7.58 | 3983100 | 7.58 |
| 6/3/19 | 7.58 | 7.65 | 7.42 | 7.47 | 5229100 | 7.47 |
| 6/4/19 | 7.58 | 7.85 | 7.52 | 7.82 | 10430100 | 7.82 |
| 6/5/19 | 5.49 | 5.59 | 4.71 | 5.04 | 39354200 | 5.04 |

## Criteria for Success

## **Organization of content**– Logical ordering of ideas, modeling artifacts, applicable visualizations in slides

## **Organization of code**- R Code is well organized, concise, and free from error

## **Text mining process** – Recognize the type of data mining problem, adherence to established main data and text mining steps.

## **Completeness** – Understood impact, and mined the data for relevant insights/recommendations