1. **Project Overview.**

Our project has engaged with the Reddit dataset on Meme Stock: Gamestock. Our guiding research question is *How does emoji usage affect sentiment intensity in meme stock Reddit communities?* With the subquestion *Can the presence or absence of certain emojis in post titles be used to predict sentiment?* Given the current growth of computational tools that perform different types of text analysis in broad datasets, especially sentiment analysis tools, we are interested in reviewing their nuance. ‘Sentiment analysis’ concerns the analysis of human opinion and attitudes that are communicated in written text with respect to entities such as organisations, other people, and events, as well as topics on a broader scale. Online communities such as the one analysed here develop their own symbology through emojis and specific lexicon that require an in-depth contextual analysis to understand. In our research, we have found that the VADER model for sentiment analysis applied to our dataset was unable to grade sentiment correctly especially in the texts using complex community specific emojis such as the fire emoji, and those that interrelated this with sarcasm. Our thesis states that emojis increase intensity and that certain emojis are associated with positive or negative sentiments. Our project concluded that, because of nuances such as sarcasm, a context specific sentiment analysis must be implemented for such an analysis to be successful.

1. **Data Acquisition.**

The dataset is divided in four parts, corresponding to the four main subreddits that were created and used to discuss the Gamestock inversions by those participating in the phenomenon. The first one, r/superstonk consists of 560,125 posts. The second one, r/GME, contains 1,033,236 posts. The third one, r/GMEJungle, has 39,634 posts. The last one, r/DDintoGME offers 5,498 posts. They are all HTML files and include 12 variables (id, title, url, score, author, number of comments, date, flair, negative sentiment, positive sentiment, neutral sentiment, and compound sentiment), and it can be seen how they interact and correlate in the files ending with “features.” The dataset has already implemented sentiment analysis through the VADER software. However, it is unclear in which part of the data such an analysis was conducted. The author explains that he conducted it on “posts”, but whether posts mean comments or titles, remains unclarified. The author also manually adjusted the value of certain emojis on VADER, such as the “gem stone”, “gorilla”, different skin tones of “raising hands”, “rocket”, different versions of “moon”, “open hands”, or the “crayon.”

As a part of data processing, the dataset was divided into a smaller sample consisting of 166 positive, 167 negative and 167 neutral scored titles and the rest of the dataset for each of the subreddits. This smaller dataset is used as a training set for the sentiment predictor and the rest of the dataset is used as the test set. Afterwards data filtering was applied to isolate the titles that included one or more emojis.

1. **Methodology**

The team divided into two groups that performed quantitative and qualitative analyses, respectively.

The first group, conformed by Yulem and Pawel, performed data processing, statistical analysis, and model training to provide a quantitative perspective on the research question. To build a concrete understanding of the dataset, statistical analysis was performed to uncover the statistical patterns and key characteristics of the dataset. This included visualising the overall dataset distribution and analysing the emoji usage across the sentiment categories. These findings were used to investigate the correlation between the specific emojis and the sentiment labels, revealing insights on how the emojis contribute to the sentimental tone.

To answer the subquestion, Bernoulli and Multinomial Naive Bayes model was trained with nominal target: positive, negative and neutral with multi one-hot vectors. The model was trained on the emojis and its sentiment scores of the training set and it predicted the sentiment score of the title from the test set. By assessing the f1-score of these predictions it is possible to conclude on whether the emojis influence the sentiment of the titles.

The second group, formed by Tomas, Rose, Ellie, and Alma, focused each in one of the four datasets. They each performed close reading on a selection of titles containing emojis. The main objective was to compare the VADER score to their own interpretation of the titles, paying special attention to the role of emojis.

1. **Workflow Steps, Challenges, and Solutions**

Initially, our research question was concerned with identifying a correlation between the sentiment of the posts and the stock market. This approach was based on the assumption that a more positive sentiment would be seen in the days of an increase in the market value of the stock, and a negative one the days in which the market value would decrease. We quickly ran into feasibility problems, since the dataset only contains the timestamps of the titles, and not of the comments. Additionally, how the dataset creator ordered the comments from a reddit comment thread was not made clear; new comment chains or different chains originating from one original comment could not be identified. Taking into account the lack of clarity on the sentiment analysis performed on the database, we could not consider our findings relating the correlation between sentiment and market performance valuable, for if the sentiment analysis was performed in the comments we could not exactly compare the stock value to comment sentiment as the comments could be hours, days or weeks later than market values changes. As such the correlation could be expected to decrease significantly. It is important to note that many posts have since been deleted, which made it impossible for us to retrieve further information ourselves.

Since this route of analysis could not be developed further, we decided to focus on sentiment analysis of titles and their related comments. We found that VADER as a tool is now in a category which current literature disproves of; it makes use of a dictionary model that assigns positive and negative scores to specific words and uses the scores assigned to these words to give a grade. Current literature suggests this is not good enough as this way of grading misses out on nuance of sentences and context.

As VADER is a debated tool, given the particular vocabulary of the reddit community around Gamestock, we wondered whether such a tool would assign the same values as would humans, as humans can understand the context of the titles. Because of this, we experimented with manual annotation by attempting to assign precise values to the titles and comparing them to the VADER score. Although we could not be sure that VADER’s sentiment analysis was performed only on the titles, we assumed, for the sake of the analysis, that most of the comments underneath the title would reflect a similar sentiment to the titles of the posts. Nevertheless, we quickly run into new impossibilities. Being six people, it was hard for us to come up with a unified baseline to objectively determine sentiment for the purpose of providing scores to compare with VADER’s scores, and we encountered additional limitations whereby the team members had little to no expertise with respect to manual annotation. Moreover, as humans, it was extremely difficult to decide exactly where to set the manual score where the available values range from +1 to -1 with 3 decimal places.

After deciding to explore the dataset more in depth in order to find new ways of enquiry, we performed an initial close reading of the titles, in which we found a predominant narrative of ‘us vs them.’ There was a prevalent community feeling, in which the individuals in the subreddits were united against the Wall Street Hedge Funds, against short selling, against an unfair market. They felt like they were heroes fighting powerful villains and doing something almost impossible, and our team was intrigued by this community feeling and thinking of how this could link to sentiment analysis. Building on our previous path, we came up with three research questions that we evaluated in terms of relevance and feasibility:

**RQ 1:** *Are certain emojis systematically associated with positive or negative sentiment in meme stock communities?*

**Proposed Approach:**

- Build a frequency table of top-used emojis.

- For each, compute average sentiment scores where it appears.

- Rank by sentiment polarity or variance.

- Use keyword-in-context (KWIC) sampling to examine usage cases.

**Feasibility:** Moderate; requires reliable emoji extraction and some manual review for context.

**Research Aim:** Explore emoji-level sentiment nuance; test if lexicon-based tools are appropriate in Reddit finance talk.

**RQ 2:** *Are emojis used as sentiment amplifiers or replacements in short texts (e.g., titles)?*

**Proposed Approach:**

- Group posts/comments by word count.

- Compare sentiment of very short posts with/without emojis.

- Manually inspect cases where sentiment is conveyed almost entirely via emojis.

- Run TF-IDF or embedding similarity between text-only and emoji-heavy posts.

**Feasibility:** Moderate; needs careful filtering and perhaps manual review.

**Research Aim:** Helps refine sentiment analysis in short, informal text environments (e.g., tweets, memes).

**RQ 3:** *How does emoji usage affect sentiment intensity in meme stock Reddit communities?*

**Approach:**

- Quantify emoji count per post/comment.

- Analyse sentiment scores (compound, pos/neg/neutral) in relation to emoji count.

- Use regression or correlation analysis to see if more emojis = stronger sentiment.

- Compare results across subreddits.

**Feasibility:** High; uses existing data and simple statistical tools.

**Research Aim:** Shows whether emojis intensify or dilute sentiment — useful for improving sentiment models.

After a group discussion, we came up with our current research question: How *does emoji usage affect sentiment intensity in meme stock Reddit communities?* With the subquestion *Can the presence or absence of certain emojis in post titles be used to predict sentiment?* And to answer them we employed the aforementioned methodology.

1. **Ethical Considerations**

The ethical considerations of our project were mostly related to possible biases in the dataset and in the close readers of the project. The dataset may have potentially included biased grades as the author manually assigned scores to emojis that were common in the community. This gave the opportunity for the scores to be more biased depending on the view of this original author. Additionally the author neglected to include key subreddits in his dataset such as r/wallstreetbets which was a key acting subreddit during the gamestop short. As such, it is suggested that users of this dataset engage with it critically. We managed this by considering the included subreddits as a sample of the whole reddit community surrounding the event, not as a complete dataset. The entire project was critical of the dataset and how it had been processed, and we were wary of the biases of the author throughout the project.

Some considerations we made throughout the project included the fact that our own readers are not objective, trained graders; consequently, we came together often to discuss our grading so that we could ensure similar criteria were being applied to form a unified baseline. We discussed examples together to mitigate the worst excesses of our subjective analysis.

1. **Results**

In meme stock subreddits like r/Superstonk, r/GME, r/GMEJungle, r/DDintoGME, emojis function as affective markers that symbolise shared values, tones and emotional intensity. Our findings show that VADER struggles to accurately interpret community-specific language and irony, especially when it comes to emojis. For example, the fire emoji is consistently scored as negative despite its use to express ironic enthusiasm or defiant optimism. This highlights the limitations of using VADER without customisation, as it misses key affective and cultural markers in the language of the community. These results suggest that without adaptation, tools like VADER may misrepresent sentiment in contexts where meaning is shaped by in-group vernacular, sarcasm and performative expression. This is a major limitation of sentiment analysis tools like VADER, as they assign fixed sentiment values to emojis and words through a numerical score representing its emotional tone (positive, negative or neutral). Therefore, they can only analyse sentiment as it is predefined. This is why close reading, a customised sentiment model or community-informed annotation is necessary when performing sentiment analysis. Emoji usage can be a useful predictor of sentiment through a combination of qualitative and quantitative methods.

1. **Documentation and Sustainability**

The original data used in our research is available on: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TUMIPC.

On the github repository of this research the processed dataset is available under the folder "data\_500" and the generated scatter text is available under the "Scattertext\_html" folder.

The "Annotation\_ipynb" folder contains the two versions of the annotation platform and a draft version. Under "graphs" it is possible to find the visualisations that were generated during our initial analysis. These visualisations depict different correlations present in the original dataset. The "Data\_Analysis" folder contains four notebooks, one for each of the subreddits, and it contains statistical analysis on the data from each of the subreddits, model training, prediction generation and classification report.

The project's output does not require active maintenance, as most of the dataset used is readily available through the Harvard Dataverse and the git repository. Using the original dataset and the uploaded data\_500, it is possible to conduct further research by increasing the scale of closing reading and improving the model to provide high f1-score for all four of the subreddits.

1. **Reflection.**

In our research, we explored many directions. Although this ultimately gave us a comprehensive understanding of the dataset and its capabilities, the process could have been more time-efficient if the dataset had a clearer description. As it has been mentioned, there were many shortcomings in the creator’s explanation (i.e. how the sentiment analysis was performed or how much information is now unretrievable). Because VADER is the tool used in the dataset, this was also the tool we used for our sentiment analysis. Future research could compare VADER to more updated softwares for sentiment analysis. Combining quantitative tools like VADER with close reading, or adapting VADER to include community-specific slang and emoji sentiment scores, could make it a more useful baseline for researchers studying sentiment dynamics in similar online meme stock communities and spaces.

Moreover, as humans with the ability to identify and understand the deeper contextual information, we would also have liked to still improve the scores ourselves. For example, we would have liked to create a more comprehensive list of emojis and their meanings, not only interpreting them in their community but giving them a set of positive or negative values that could be used to adapt VADER to this specific community.