Team Red Pandas - WallStreetBets Tracker

ITSC 3155 Final Project Report

Prepared by:

Jonathan Mendenhall

Ian Zheng

Charlotte Hill

Daniel Tyree

Date:

May 2, 2021

Table of Contents

Table of Contents **2**

Introduction **4**
- Product Vision 4
- Customer Description 4
- Project Scope and Objectives 5
- Ethical concerns 5

Project Resources **6**
- Group Members 6
- Data 6
- Hardware and Software Resources 6
- Special Resources 6

Plan **7**
- Timeline Chart - Milestones 7
- Timeline Chart - Tasks 8
- Milestone Descriptions 9
- Task Descriptions 9
- Resource Table 12

System Design **14**
- Use Case Diagram 14
- DFD diagram 15
- User Stories 15
- Feature List 16
- Storyboard 17

User Tests **18**
- Test procedure 18
- User Test and Results 18
- Conclusion 18

Lessons Learned **19**

Future work **20**

Appendices **21**
- Sketches 21
  - Jonathan Mendenhall’s Sketch 22
  - Daniel Tyree’s Sketch 23
  - Ian Zheng’s Sketch 24
  - Charlotte Hill’s Sketch 25

Combined Features Sketch 26
Software Repository 27
WBS tool 27

Installation **27**

Running **27**

Final Product Screenshots **29**

Introduction

Investing in the stock market can be quite intimidating for most people. Many are not sure what stocks they should invest in and are worried that investments they make may result in lost money due to a poor investment decision. While long-term investments are always a good way to slowly build and expand an investment account, short-term investments also offer the chance to take advantage of certain market phenomena and create a possibility for dramatically increasing investment wealth. Because the social media age is becoming more ingrained in everyday life, it is creating an environment to foster this kind of market phenomenon.

As seen during January 2021, the WallStreetBets subreddit has a strong ability to cause inflation in the price of stocks in their discussions, and knowing about this early in the rise could help someone have a huge return on investment as long as they get out before the price of the stock drops again. An app that lets someone see what the currently trending stocks in WallStreetBets are, could help the individual jump in on a “moonshot” early for a low price and then sell their position when they think the stock price is at or around its high and starting to go back down.

Product Vision

For anyone interested in taking advantage of bullish stock behavior caused by mass discussions on Reddit, WallStreetBets Tracker is a web app that analyzes Reddit posts to find trends in different stocks, allowing users to take advantage of stock price increases due to mass market activity. Unlike Yahoo Finance or other market-viewing tools that leave all of the judgment to the user, our product allows users to view trends in “hype” around various stocks that have previously led to large increases in stock price.

Customer Description

This product is designed for:

- Anyone interested in individual investment in the stock market
- People who want to take advantage of the “hype waves” that can happen because of WallStreetBets posts and don’t mind the risk
- Individuals who are new to the stock market looking to find trends in specific “hyped” up stocks.
- People who are interested in being connected to a broker and or to find a list of sources regarding trading.

Project Scope and Objectives

This project will allow stocks mentioned in Reddit/WallStreetBets posts to be counted and displayed in user-friendly and intuitively understandable graphs and charts. This will allow any trends in stocks being written about overtime to be easily viewable. Users will be able to see what stocks are becoming popular and have the potential to find a stock they would like to invest in. In summary, the project will:

- View trends in stocks mentioned over time across r/wallstreetbets posts.
- View currently hyped stocks so users can choose to open a position early on in the potential wave created by exponential discussion.
- Find trends in stocks mentioned with specific phrases.

Ethical concerns

Due to the fact that WallStreetBets Tracker is dealing with data analysis of the stock market and user-based preferences of stocks, it has the possibility to create unintended market manipulation and encourage individuals to invest in a potentially high-risk stock. As such, we have included a disclaimer on our web page at the bottom that reads with the following:

- We do not provide personal investment advice, and We are not a qualified licensed investment advisor. We will not and cannot be held liable for any actions you take as a result of anything you read in the WSB Trackers.
- All information found here, including any ideas, opinions, views, predictions, forecasts, commentaries, suggestions, or stock picks, expressed or implied herein, are for informational, entertainment, or educational purposes only and should not be construed as personal investment advice. While the information provided is believed to be accurate, it may include errors or inaccuracies.
- Conduct your own due diligence, or consult a licensed financial advisor or broker before making any and all investment decisions. Any investments, trades, speculations, or decisions made on the basis of any information found on this site, expressed or implied herein, are committed at your own risk, financial or otherwise.

Furthermore, WSB Tracker conducts data analytics only on posts from the Reddit/WallStreetBets postings and has no direct interaction with the data and does not artificially inflate the number of postings of stock to potentially create a higher demand for the stock in order to cause any form of market manipulation.

Project Resources

Group Members

1. Jonathon Mendenall - JM
2. Daniel Tyree - DT
3. Ian Zheng - IZ
4. Charlotte Hill - CH

Data

The primary dataset our application uses is a dataset of all posts on the WallStreetBets subreddit. The dataset includes eight columns, but our application only reads from the following columns in order to enable the desired functionality:

- *body*: the content of the post on the subreddit.
- *timestamp*: the date and time when the post was created.

The dataset can be accessed with the following link.
https://www.kaggle.com/gpreda/reddit-wallstreetsbets-posts

Hardware and Software Resources

The only software resources required for this project are Python, and PyCharm for development. Our web application is built primarily using the following libraries:

- pandas: loading and transforming datasets
- dash: create an interactive web app to visualize the data

No hardware resource requirements are needed for this project.

For instructions of how to install and run WallStreetBets Tracker please see Appendices Run and Installation on page 27.

Special Resources

In addition to the Reddit posts dataset, our app requires a list of stock ticker symbols to search for in all the Reddit posts, so listings of stock tickers and company information on various stock exchanges are included as necessary datasets. Most relevant to our app, the ticker listings include the following information:

- *symbol*: the ticker symbol of the stock for a given company that will be searched for in the messages dataset

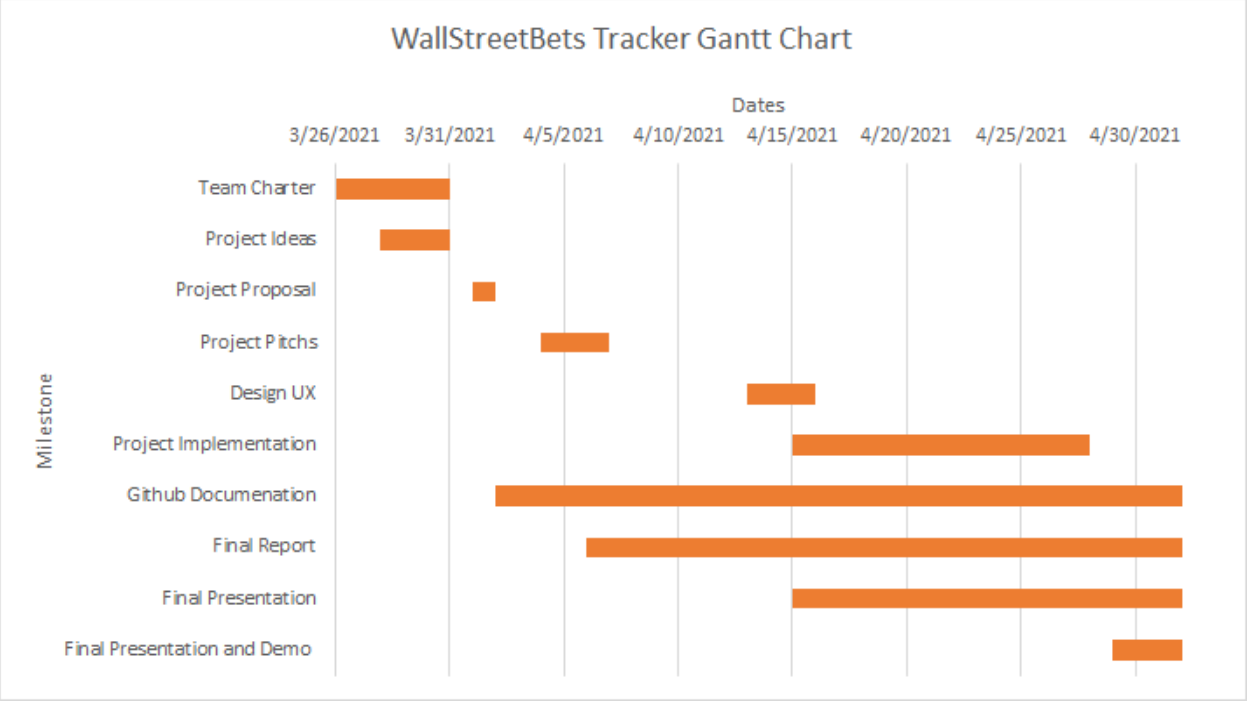
- *name*: the name of a company
- *sector*: a categorical classification of the field a company is related to, so analysis of the top N companies in a given field can be accomplished.

The listings include much more information like the last closing price, net change, and percent change of the stock, the market capitalization, country, IPO year, volume but these are not necessary for our application. The stock listings also included another company classification column, *industry*, that is more specific than the *sector*, and thus created more groups with fewer companies each, so we decided against using it when allowing the user to filter data as it creates more difficulty and is less useful.

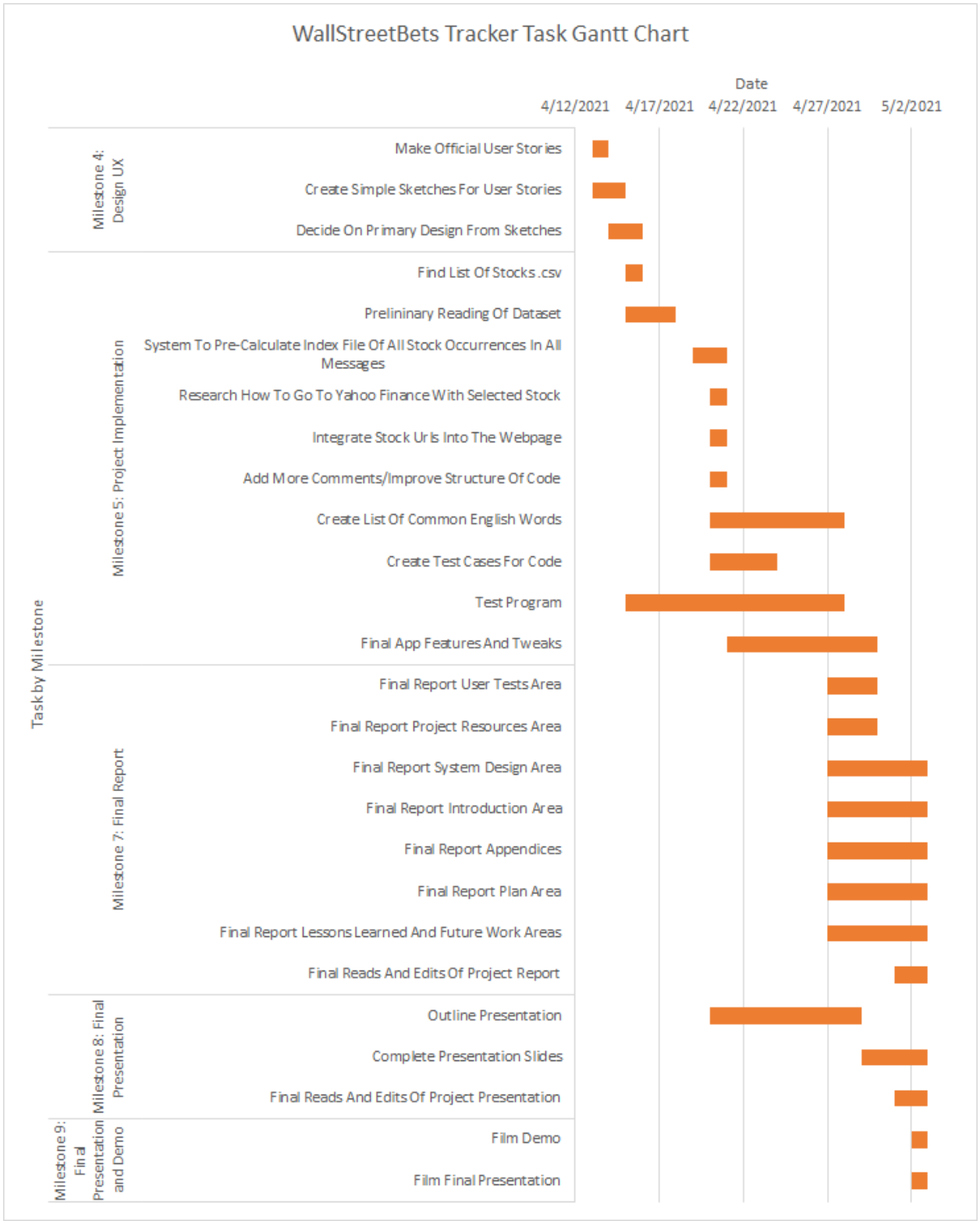
The datasets can be accessed with the following link.
https://www.nasdaq.com/market-activity/stocks/screener

Plan

Timeline Chart - Milestones



Timeline Chart - Tasks



Milestone Descriptions

- Milestone 0 - Team Charter: The basis of how the team will interact with each other, how aspects of the project will be handled, what the expectations are for how the project will be handled, and any dedicated roles of group members.
- Milestone 1 - Project Ideas: A set of four potential project ideas, one from each team member, based on a dataset that may offer a useful application.
- Milestone 2 - Project Proposal: An explanation of why the chosen project idea is important, along with a product vision that describes the purpose of the project, who could benefit from the project, and the scope of the project.
- Milestone 3 - Project Pitches: A pitch from each team member about how they view the problem the project will solve and the potential benefits to any user.
- Milestone 4 - Design UX: Design the structure of the program web page.
- Milestone 5 - Project Implementation: Source code of the project with any instructions needed for install and execution.
- Milestone 6 - GitHub Documentation: All documentation involved with creating the project.
- Milestone 7 - Final Report: A full report of what the project is about, the objectives, resources, a timeline, milestone descriptions, diagrams, storyboard, testing, and etc.
- Milestone 8 - Final Presentation: Final slide presentation that needs to be a summary of the final report.
- Milestone 9 - Final Presentation and Demo: Recording of presentation and a working project.

Task Descriptions

Milestones 0 through 3 and milestone 6 were stand alone milestones that were not broken into tasks.

Milestone 4 - Design UX

- Task 1 - Make Official User Stories: Create three or more user stories for how a user could use a product like how our idea would be used.
- Task 2 - Create Simple Sketches for User Stories: Using the user stories, create a simple but as fully developed idea of what the webpage could look like.
- Task 3 - Decide on Primary Design from Sketches: Discuss and and decide on what elements from each member's sketch would be best suited to create a simple, streamlined webpage for the project idea.

Milestone 5 - Project Implementation

- Task 1 - Preliminary Reading of Dataset: Read the Reddit posts from the dataset and begin to work with the data to count the number of occurrences for each stock symbol.
- Task 2 - Find List of Stock .csv: Find a source that offers a dataset containing the currently held stocks, along with their corresponding company/industry/ticker information, in the NASDAQ and NYSE.

- Task 3 - System to Pre-Calculate Index File of All Stock Occurrences in All Messages: Create an algorithm that will use the stock information to form an index that will hold the needed information about each stock and the number of occurrences in the Reddit posts.
- Task 4 - Research How to go to Yahoo Finance with Selected Stock: Research how to be able to take the user to a specific stock's Yahoo Finance page instead of taking the user to the main Yahoo Finance page when they have selected a specific stock.
- Task 5 - Integrate Stock URLs Into the Webpage: Based on the research of how to formulate the URL to a specific stock's Yahoo Finance page, implement the functionality into the program.
- Task 6 - Add More Comments/Improve Structure of Code: Ensure that there are the appropriate comments describing what the code is doing and the purpose behind it. Ensure that the structure of the code reduces any excess code and minimizes any bad "smells".
- Task 7 - Create List of Common English Words: Research the most common written words in the English language and create a list to use to filter them from the Reddit dataset as it is read into the program.
- Task 8 - Create Test Cases for Code: Making test cases for parts of the code to ensure functions and algorithms perform correctly across possible inputs, yielding the expected outputs
- Task 9 - Test Program: With our partner team, Team CIS_Vikings, review the current iteration of the program and discuss any problems or features that need to be addressed.
- Task 10 - Final App Features and Tweaks: Small remaining adjustments to the code to ensure that the webpage and program are intuitive to use.

Milestone 7 - Final Report

- Task 1 - Final Report User Tests Area: An explanation of how the user can interact with the project, a summary of the user testing with our partner team, Team CIS_Vikings, and any final conclusions drawn from their feedback.
- Task 2 - Final Report Project Resources Area: Identifies the team of this project, what the dataset is about, any software/hardware requirements to run the project locally, and any special resources that the project used in addition to the main dataset.
- Task 3 - Final Report System Design Area: Describes the underlying structure of the program and how the program and datasets interact. It also includes the user stories that the project is based on, the feature list and a storyboard about how a user could interact with the webpage.
- Task 4 - Final Report Introduction Area: Using the proposal as a basis, review and make any necessary changes to any of the introduction subsections to present a better overall introduction to the project.
- Task 5 - Final Report Appendices: This section contains the design sketches for the UX, the repository for the entire project, the instructions of how to install and run the program, the WBS tool used with a description of how it was used, and final screenshots of the webpage as it should look upon running the program.

- Task 6 - Final Report Plan Area: A set of Gantt charts used to visualize how time was spent for each milestone/task, descriptions of each milestone and task, and who contributed predominantly in tasks.
- Task 7 - Final Report Lessons Learned and Future Works Area: An explanation of what issues the team encountered while working on the project and how they were resolved, along with how the team could envision further improvements that could be made to the program so that it would have more features and functionality.
- Task 8 - Final Reads and Edits of Project Report: A series of final reads and edits of the final report to further refine the content.

Milestone 8 - Final Presentation

- Task 1 - Outline Presentation: A basic outline of the presentation based on suggestions from the instructor.
- Task 2 - Complete Presentation Slides: Tailoring the presentation slides to present the most important information about the project and program.
- Task 3 - Final Reads and Edits of Project Presentation: A series of final reads and edits of the presentation to further refine the content and present any ideas of how to effectively communicate the information to the viewer without overwhelming them with too much information.

Milestone 9 - Final Presentation and Demo

- Task 1 - Film Demo: Film/recording of the demonstration of the webpage and how the user can interact with it and an explanation of its features.
- Task 2 - Film Final Presentation: Film/recording of the presentation of the project, including the demonstration of the program, with an explanation of why the project is important, what its purpose is, how the program is designed, what enhancements could be made, and what was learned during the course of the project.

Resource Table

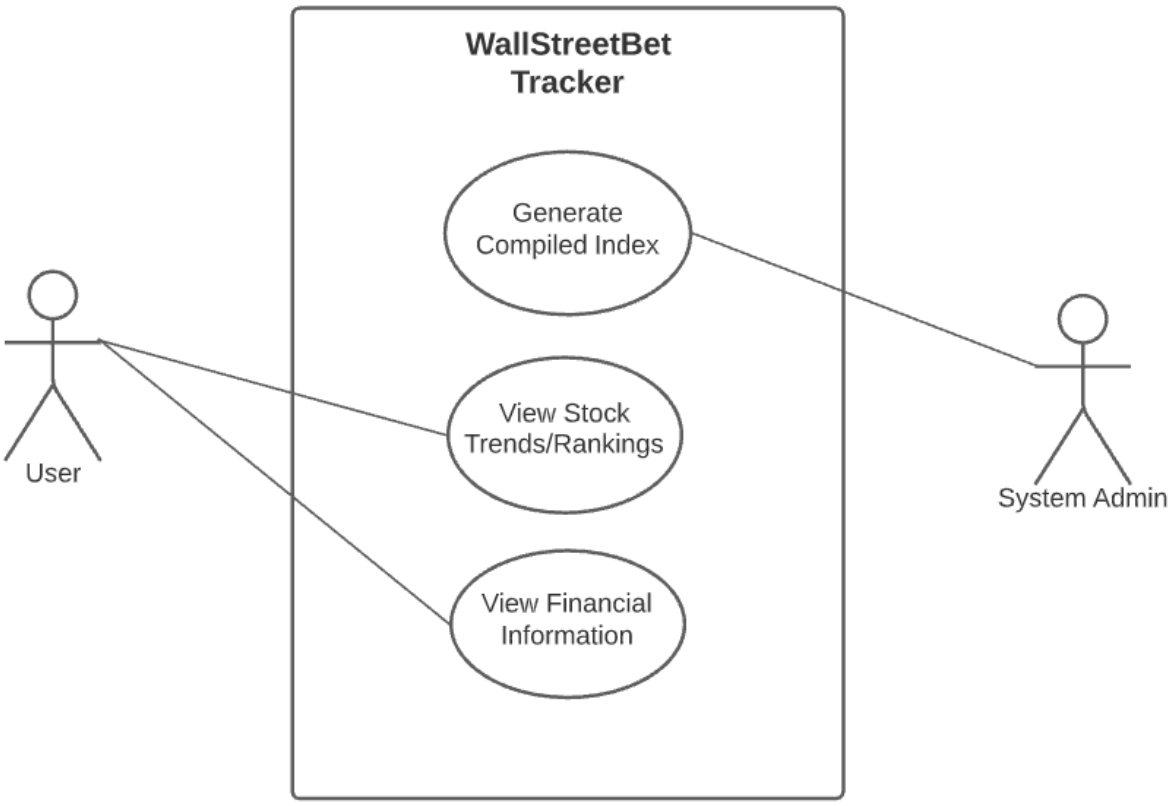
All group members contributed to every milestone in some way.

Task	People
Make Official User Stories	All
Create Simple Sketches for User Stories	All
Decide on Primary Design from Sketches	All
Preliminary Reading of Dataset	JM
Find List of Stock .csv	IZ
System to Pre-Calculate Index File of All Stock Occurrences in All Messages	JM, CH, IZ
Research How to go to Yahoo Finance with Selected Stock	CH
Integrate Stock URLs Into the Webpage	JM
Add More Comments/Improve Structure of Code	JM
Create List of Common English Words	CH
Create Test Cases for Code	JM
Test Program	JM
Final App Features and Tweaks	JM
Final Report User Tests Area	JM
Final Report Project Resources Area	JM, CH
Final Report System Design Area	IZ, DT, JM
Final Report Introduction Area	CH
Final Report Appendices	JM
Final Report Plan Area	CH
Final Report Lessons Learned and Future Works Area	JM, CH

Final Reads and Edits of Project Report	All
Outline Presentation	DT
Complete Presentation Slides	All
Final Reads and Edits of Project Presentation	All
Film Demo	IZ
Film Final Presentation	DT

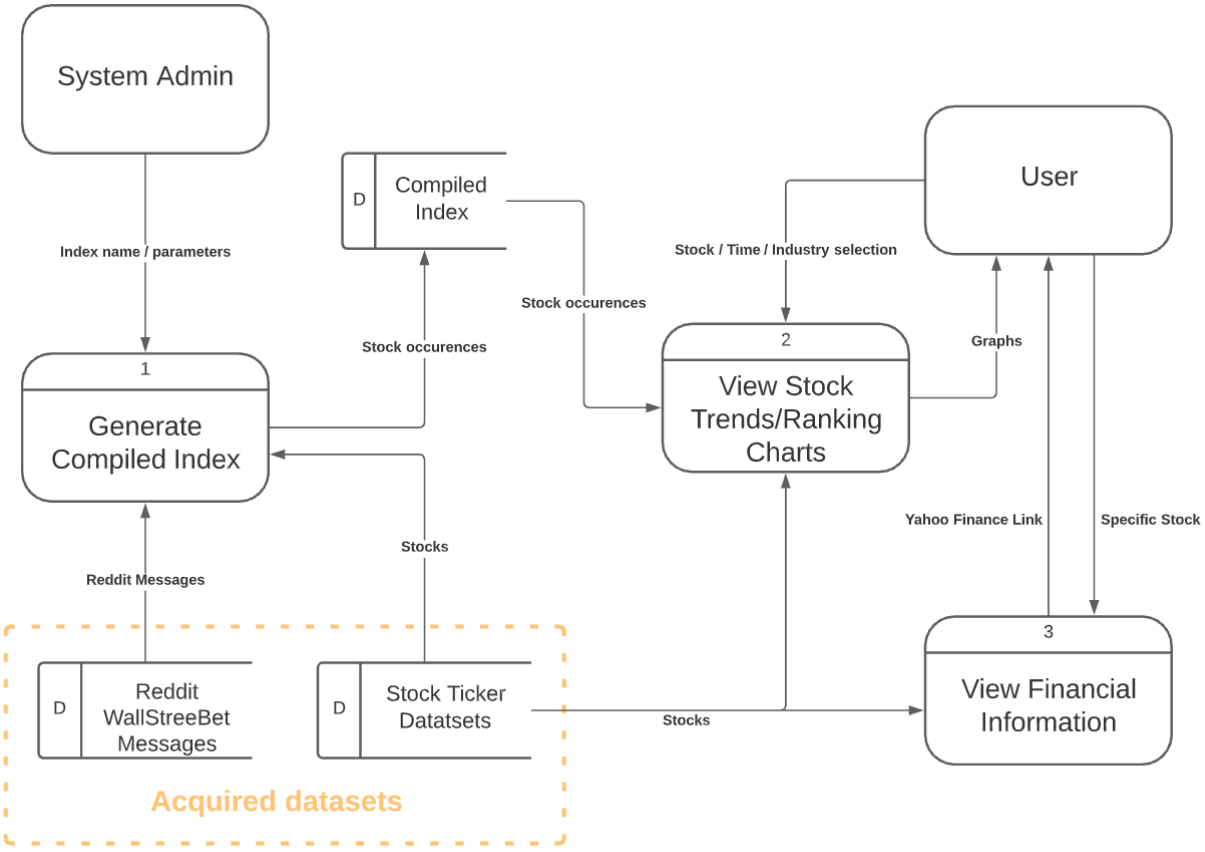
System Design

Use Case Diagram



DFD diagram

The following Data-Flow-Diagram shows what information flows between various parts of the application, including the actors, processes, and datastores. Datastores that are the source datasets are indicated by the *Acquired datasets* section. For diagram purposes, the compiled index is considered a datastore but is more accurately described as a pre-processed dataset that is generated from the original Reddit messages dataset.



User Stories

The main goal of this web application is to save time and effort for users by allowing them to filter stocks in the WallStreetBets Reddit discussion forum with just a few clicks. The simplicity of the interface, ease of use, and intuitive navigation are all priorities for our development team. The smooth process would ensure that the application experience remains valid and that users return.

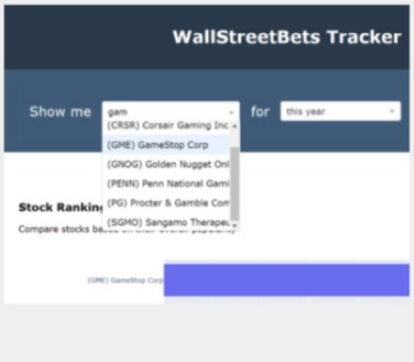
- As a user, I want to be able to see the currently trending stocks mentioned by people.
- As a user, I want to be able to view the trend in a specific stock over time.
- As a user, I want to be able to view real financial data for stock trends.

Feature List

Features	Explanation
Stock Mentions	Users are able to select which stock they would like to view as well as the duration of time to view a graph illustrating the amount of mentions per stock.
Trends Overtime	Users are able to select a specific stock and view a graph illustrating a stock's trends overtime.
Real Financial Information	Users are able to view real financial statistics via Yahoo Finance.

Storyboard

From the home page, users are presented with drop down menus that allow them to select which stocks they would like to view. The first drop down menu selects which company they would like to view, the second selects the time frame and the third identifies the industry.



From the first drop down menu, the user has selected that they would like to view GameStop Corp (GME).

From the second drop down menu the user has selected that they would like to view GameStop trends from this year.

The third and final drop down menu shows which industry the user would like to view. As a user you are able to select options from each menu or only the ones necessary for the preferred search.

Based on the selections in the menus, the user is shown a graph of the selected stock with a time that identifies how often the stock was mentioned on the /rWallstreetBets trading forum on Reddit.

Users are also presented with an opportunity to view real, statistical financial information from a credible location, this being Yahoo Finance.

Users will find that on most of the application pages, we have posted a disclaimer regarding the information that we post. We like to remind users that any information that is found and or used on our application is not real financial advice. We appreciate the usage however we are not to blame for any loss of funds in any way.

User Tests

Test procedure

The CIS Vikings group tested our application based on the following user stories which cover the functional requirements of our application.

User Test and Results

#	User Story	Test Results
1	As a user, I want to be able to see the currently trending stocks mentioned by people.	Tester performed the action quickly and intuitively.
2	As a user, I want to be able to view the trend in a specific stock over time.	Tester performed the action quickly and intuitively.
3	As a user, I want to be able to view real financial data for stock trends.	Tester performed the action quickly and intuitively.

Conclusion

After giving the testing team a short walkthrough of our application's functionality, we read our user stories and allowed a member of the other team to guide one of us through navigating the app to execute each action. All user stories were executed intuitively. The only notable event was when executing user story 1. Seeing the currently trending stocks involves looking at the most recent data, and our app allows looking at the trends for *this week*, the *past 30 days*, *past 90 days*, *past year*, and *all time*. The tester chose *this week* as the date range to view current trends, and this would be correct, but our dataset does not actually extend to the range he selected, so he saw no data. This was corrected quickly as we told him our dataset did not extend this far and he switched his selection to the *past 30 days* and was able to see the trend data. The testing teams commented that the app was very easy to use and they felt the controls were intuitive and interacting with the app was tuned and streamlined.

Lessons Learned

Creating a web application that handles a large amount of data is challenging and requires special effort into designing systems to reduce overhead and ensure the application is efficient. Pretty early on, we realized searching through the entire dataset repeatedly while the application was running would be computationally expensive, and lead to a poor user experience as the users in our audience want data as fast as possible to be able to make informed decisions. Furthermore, because this app is not meant for serious investment advice, people would likely not want to use the app as the perceived value of the data provided would be less than the cost of waiting a long time.

We chose to implement an indexing system for improving the performance and efficiency of our app. The system effectively acts like a cache that holds the number of occurrences of all stocks by date, rather than the original dataset that holds all messages. Our indexing process can be run once on the entire messages dataset, and the results are saved in a new dataset that can then be handled by the application practically instantly. By implementing this indexing process, we realized how useful and prevalent caching systems are in production applications.

In the process of working with the indexing system, it became apparent that we were obtaining false positives while counting the number of times a stock was mentioned. Originally while reading in the Reddit dataset, the program was configured to capitalize all letters and turn any special characters into whitespaces. While turning any special characters into whitespaces enabled the program to count instances where posters were using a special character such as $ or # immediately before a stock's symbol, capitalizing all of the lowercase letters allowed any words that also happened to be a stock symbol to be counted as if the stock had been mentioned. While we reviewed and researched this situation and discussed it in depth, we came to the conclusion that by instead of capitalizing all lowercase letters, the program would be better served by discarding any word being read from a post that contained a lowercase letter. During our review of several hundred messages in the data set, it was found that an overwhelming majority of the posters always referred to stocks by their symbol or their company name, with the later having dramatically less occurrences than the former.

While discussing possible false positives that could occur from common written words in the posts and their correlation to stock symbols, it was decided that while there are most likely instances of posters using all capital letters in at least part of their post that would not be referring to any stock symbols but simply being used to express an opinion or feeling, this only occurred under rare circumstances and the false positives that would be caused from this would be of negligible value to the overall instances of a stock symbol occurrence. Therefore, we chose to keep these instances counting as occurrences of stock symbols. However, we did find an instance of one stock symbol that we chose to disallow from the indexing system. The capital letter A is also considered a word and is permitted to use as the starting word of a sentence. Given that the word A is an extremely common written word, our team decided that any occurrence of the letter A would not be considered an occurrence of the stock symbol A.

Future work

Our team believes that WallStreetBets Tracker has the potential to offer a much more comprehensive analysis of the posts on the Reddit/WallStreetBets forum. An ability to show the financial data of a stock in the tracker would enable the information to be more tailored to the application's target audience by allowing a simpler and more streamlined display of necessary information about the stock and the company it represents. We would also like to see about expanding the tracker to index funds like the Dow Jones Industrial Average and the S&P 500, along with foreign stock exchanges and other, less popular American exchanges as a future possibility as well.

Further refinement of the algorithm searching for stock mentions in the Reddit posts would also offer more accurate results and allow for stocks with unique symbols to be counted when they are not currently. Bringing the concept of searching with more refinement to the user side of the application would allow for users to search the Reddit posts for specific wording being used instead of just a specific stock. This is not a current possibility with the indexing system being used to hold the stock symbol occurrence information. In summary, in the future we believe that the WallStreetBets Tracker could be expanded to include the following abilities:

- Allow search stock trends when a specific phrase is mentioned. For example, occurrences of stocks in messages that contain “to the moon”
- Show financial data directly on the app sourced from a well-established system like Yahoo Finance or Google Finance
- Able to automatically scrape most current discussions from WallStreetBets forum for most up to date chart displays
- More advanced search algorithm possibly using machine learning for more accurate results
- Expansion into index funds like the Dow Jones Industrial Average, S&P 500, etc...

Appendices

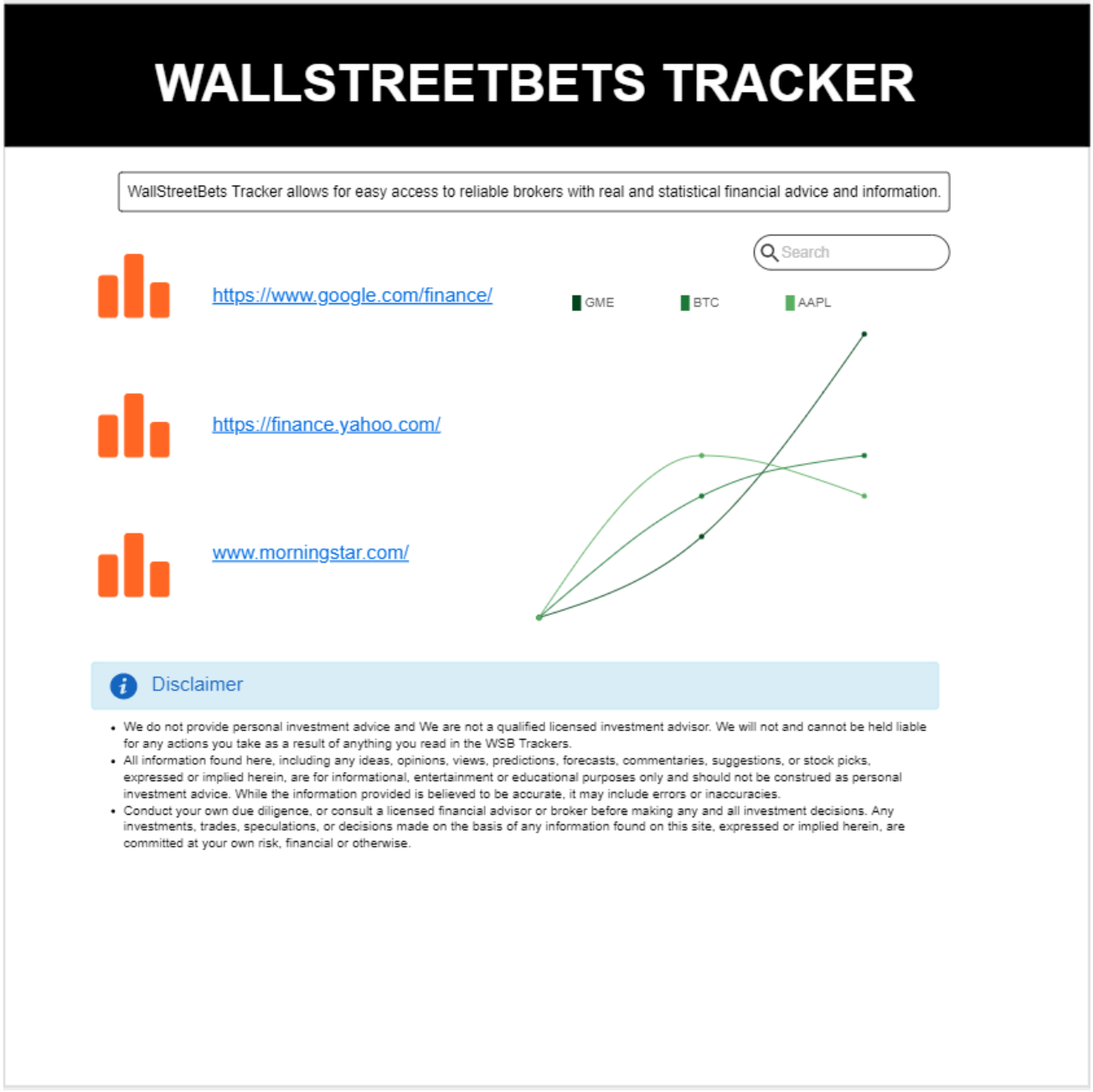
Sketches

These sketches show each of our individual ideas from the start of the process of how the app could look in the end. After we each made these sketches, we came together as a group and talked about each sketch and the good things we liked about them, and incorporated these ideas into the final group sketch. As development progressed through the project, we refined the interface, adding elements to make the app more appealing and helpful to the user. As such, these do not reflect the final app, only the inspirations behind it.

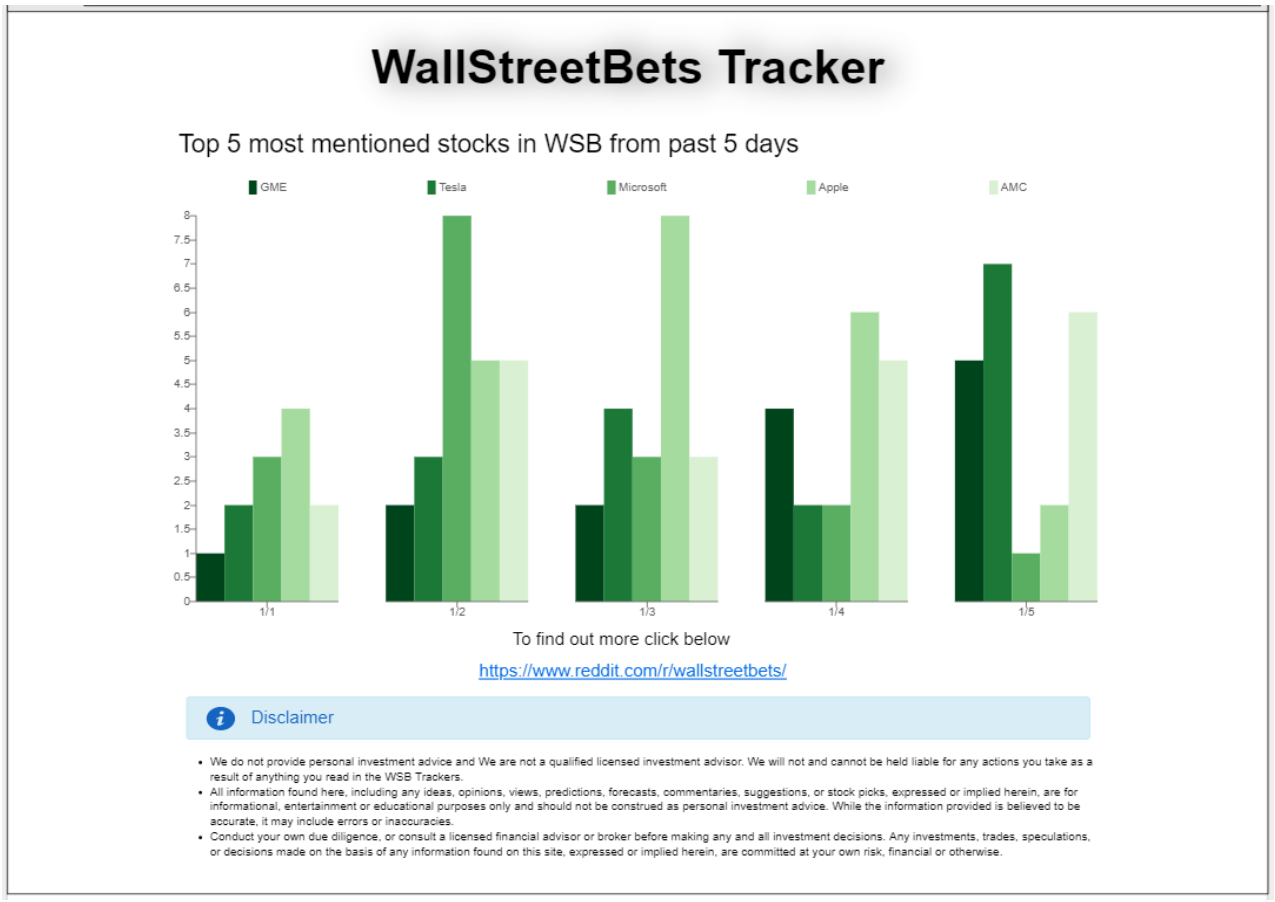
The original sketches can be found on Moqup with the following link:
https://app.moqups.com/kDcmLX2Jj5/view/page/ad64222d5

Jonathan Mendenhall's Sketch

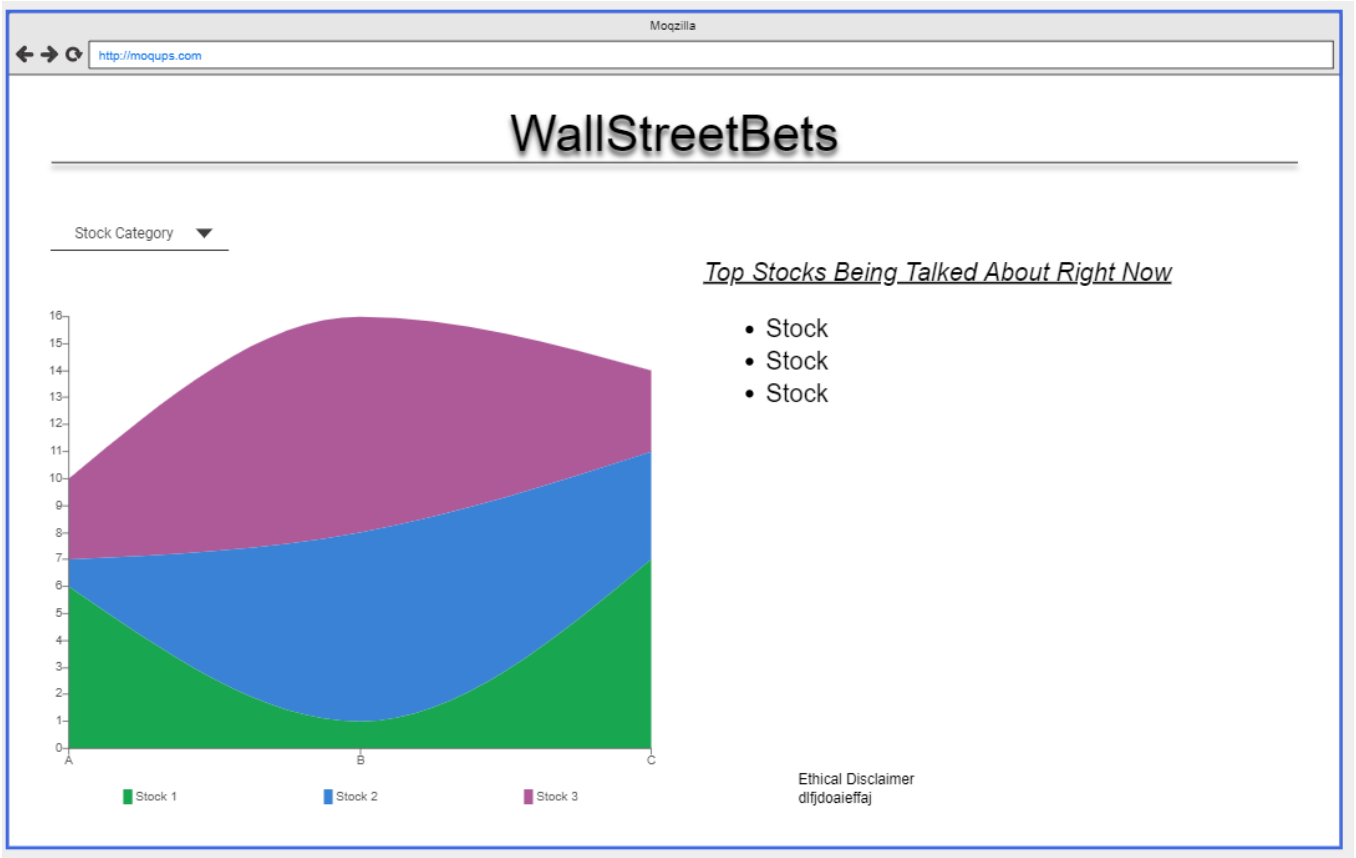
Daniel Tyree's Sketch



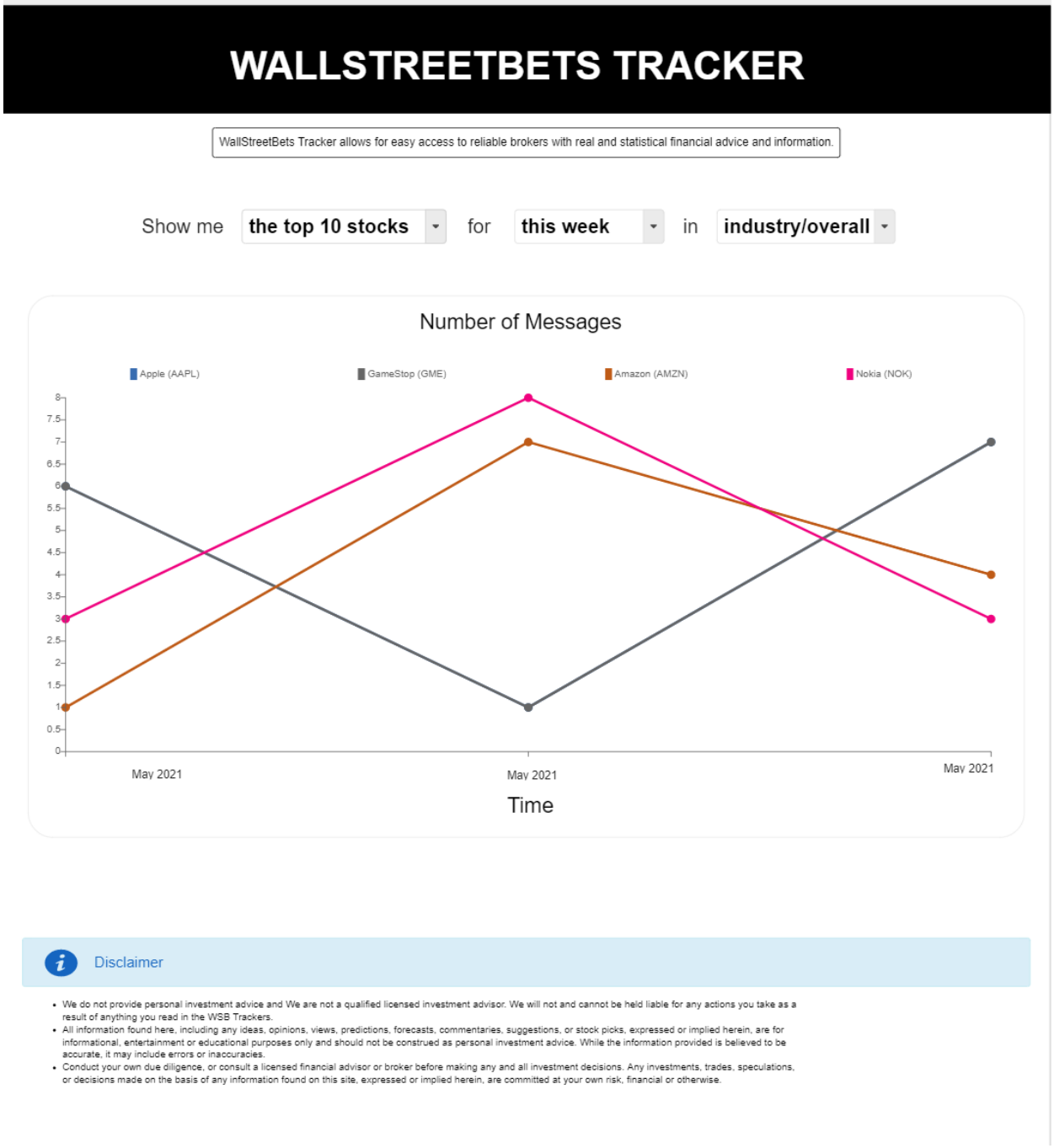
Ian Zheng's Sketch



Charlotte Hill's Sketch



Combined Features Sketch



Software Repository

https://github.com/hcy0227/ITSC-3155-Team-Red-Pandas-FinalProject

WBS tool

Our team chose to use Trello as a way to break down milestones into tasks with descriptions for what most of the tasks were and what were required of them. Additionally, we used it to both assign tasks to certain members and to keep track of who was working on what in the areas that were not assigned. Our team also found it to be very useful to leave comments on a task we were working on independently to identify anything that had come up during the work, to address any issue, or to hold information that may be needed in the future. We also used the checklist feature of the tasks to keep track of requirements or sections of a task if it was needed. The following link is for our group Trello board for the project, created by Ian Zheng:
Final Project Workflow | Trello

Installation

A zip file containing the GitHub repository and the virtual environment for running the app is provided along with this submission. The repository contains documentation of the app, its source code, and the data the app needs to run. It is best practice not to include virtual environments or workspace settings inside of repositories because paths can vary across different computers or users who might use the code. In order for the project to be set up to run, the zip file simply needs to be extracted to some place.

If the virtual environment does not work for some reason, then simply ensuring the system's python3 environment has the necessary packages can be done using the following command inside of the root directory of the git repository.

```
python3 -m pip -r requirements.txt
```

The executable name for python can vary across systems, but on your own system, it should be known whether it's `python`, `py`, `python3`, or something else

Running

Open a terminal inside of the `ITSC-3155-Team-Red-Pandas-FinalProject/src` directory and run the following command to start the app:

```
./"../../venv/Scripts/python.exe" app.py
```

This will run the main app.py file using the virtual environment's python interpreter rather than the system's default python interpreter as would be the case if simply running `python app.py`

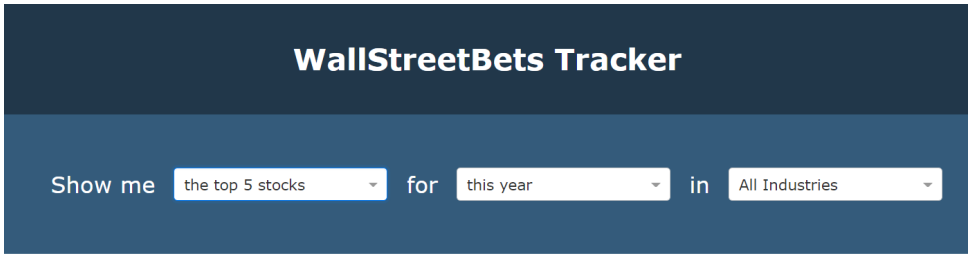
It is **imperative** that the app be run with `ITSC-3155-Team-Red-Pandas-FinalProject/src` as the working directory so the relative location of the data directory is found properly. If the following error occurs when the app starts up, then you are not running the app from the correct working directory, and need to navigate to the directory using the `cd` command.

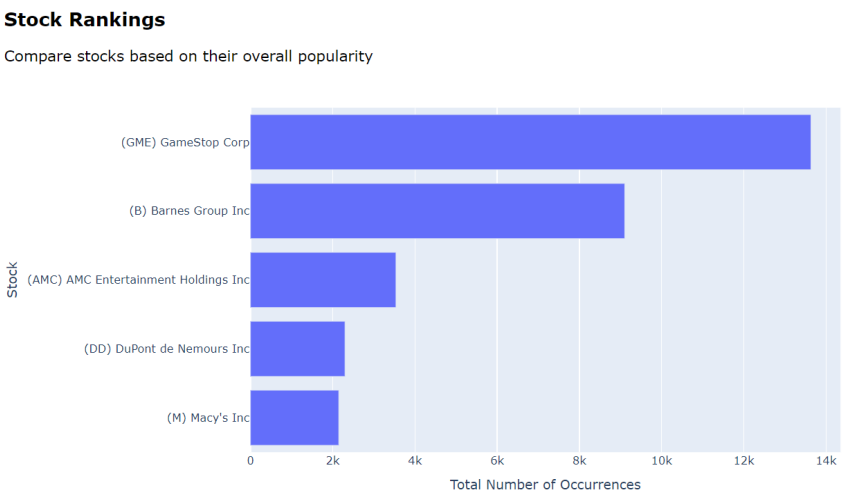
```
FileNotFoundError: [Errno 2] No such file or directory:
'../data/NYSE_stock_tickers.csv'
```

When the app is running and no errors occur in the terminal, the output should display like below, and the app can be viewed in a web browser following the link shown: http://127.0.0.1:8050/

```
Dash is running on http://127.0.0.1:8050/
* Serving Flask app "WallStreetBets Tracker" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production
deployment.                                          Use a production WSGI
server instead.
* Debug mode: on
```

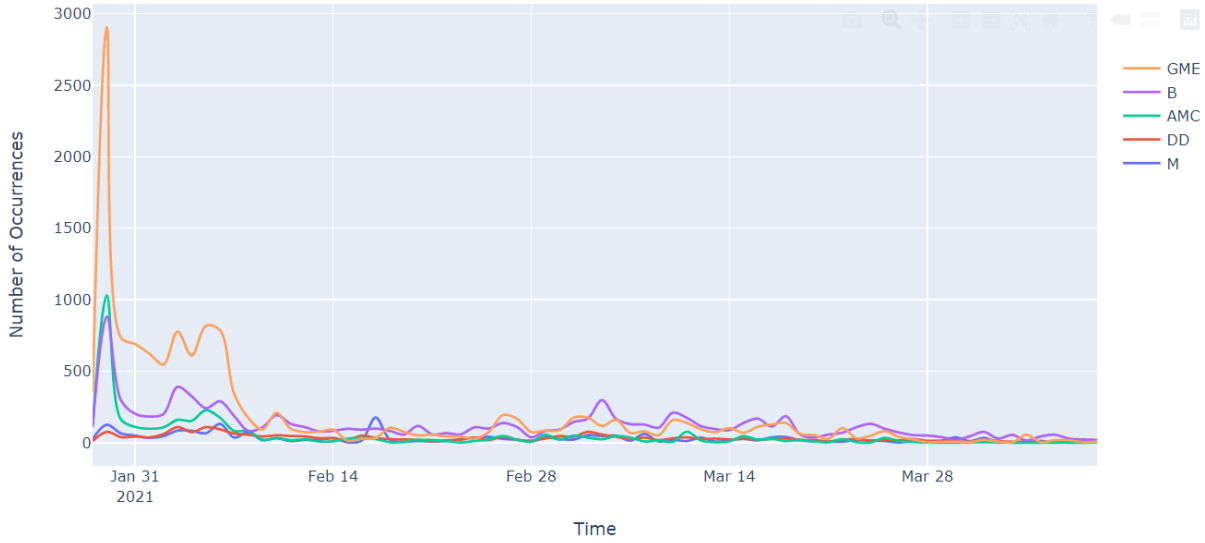
Final Product Screenshots





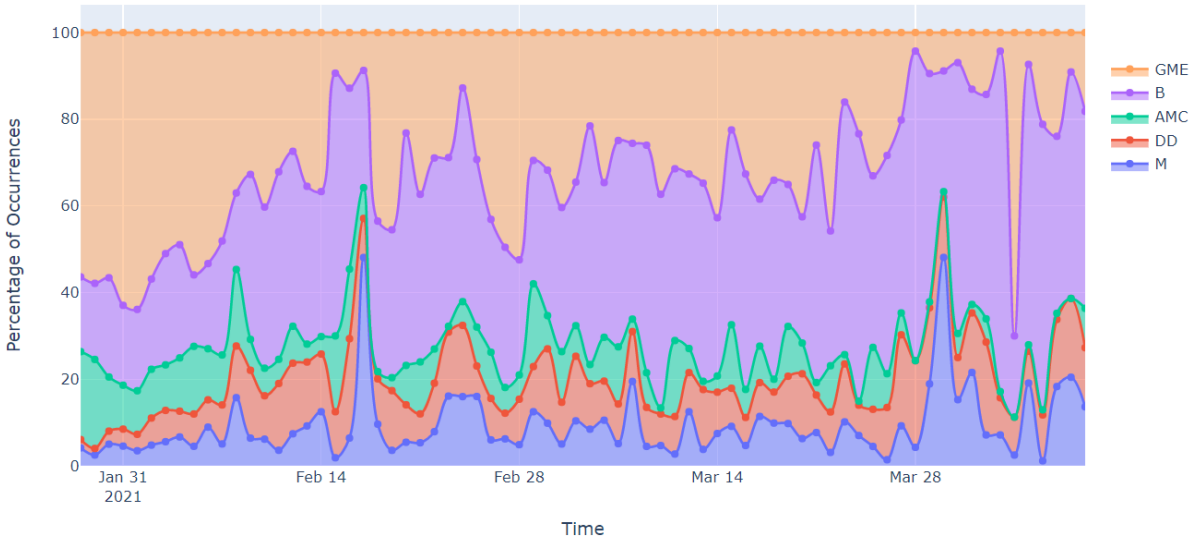
Stock Symbol Frequency

Understand the trends in people mentioning specific stocks over time



Relative Stock Symbol Frequency

See how stocks compare to others mentioned on the same day



Financial Information

View each stock's price trend on Yahoo Finance

1	GME	GameStop Corp	2	B	Barnes Group Inc
3	AMC	AMC Entertainment Holdings Inc	4	DD	DuPont de Nemours Inc
5	M	Macy's Inc			

Disclaimer

We do not provide personal investment advice and We are not a qualified licensed investment adviser. We will not and cannot be held liable for any actions you take as a result of anything you read in the WSB Trackers.

All information found here, including any ideas, opinions, views, predictions, forecasts, commentaries, suggestions, or stock picks, expressed or implied herein, are for informational, entertainment or educational purposes only and should not be construed as personal investment advice. While the information provided is believed to be accurate, it may include errors or inaccuracies.

Conduct your own due diligence, or consult a licensed financial adviser or broker before making any and all investment decisions. Any investments, tr speculations, or decisions made on the basis of any information found on this site, expressed or implied herein, are committed at your own risk, finan otherwise.