Team 2 (Stocks/Social Media)
Progress Report 2
March 25, 2022

I. Description of the MVP

Our project should include:

- A stock trend prediction model (e.g. up, down, same over the next 3 days) with an accuracy greater than 60% on random data.
- Twitter scraping tool that uses hashtags to gather tweets about certain companies.
- A sentimental analysis model that will parse twitter messages and assign a positive, negative, or neutral vector to the words comprising the tweet. We will probably use various provided metrics online for our sentiment analysis model and see which metrics have better trends and predictions with certain companies, and tie this in with the performance of that company's stock.
- Stock prices can be visualized easily (with many view options) with a dashboard.
- Models' performance should be possible to visualize somehow as well.
- Helpful stock metrics with Stock KPI's. KPI stands for key performance indicator and it is a metric to understand how well stock is performing against the market. The KPIs we will use are: CAGR, Volatility, Sharpe Ratio, Sortino Ratio. CAGR stands for Compounding Annual Growth Rate and the purpose is to measure the growth of a stock/portfolio of stocks. Volatility is a measure of the riskiness of a stock/ portfolio of stocks. The Sharpe Ratio is a metric to measure return compared to risk. Sortino is a metric to measure return having been risk adjusted.
- Combine all data analysis subparts (predict stock trend, sentimental, KPI metrics) to create a single decision model that outputs the current state of a stock (buy, hold, sell, etc. and perhaps also a score).

II. Current progress

- Created a basic dashboard that displays historical price data for a ticker of our choice across different periods.
- Cleaned and formatted raw price data from Yahoo Finance API to be used with a recurrent neural network. (convert data from a table where each row represents a day data to a table where each row represents multiple days data along with a label - trend of stock price over the next few days)
- We found a Python package called Tweepy that accesses the Twitter API via hashtags. We will also be using another Python package called Flair that is a pre-trained embedding-based language model that we can use to assign vectors for the Twitter tweets we scrape and determine the user's sentiment. Finally, a Twitter developer account has been made and some test pulls have been done.
- The algorithms and strategies to create the KPI have been developed.

Team 2 (Stocks/Social Media)
Progress Report 2
March 25, 2022

Problems we anticipate:

- Sentiment analysis: covering edge cases as well as negation statements and words (ex. “Not”, “can’t”, etc.) and basing a sentiment analysis score off of those.
- We have to worry about finding a balance of storing data we scrape in our repo and file management/clutter; we don’t want too many files in one place but we also need some place to store and easily access any files we do decide to store.
- KPI’s are not the only way to determine stock performance in the market
- Prediction accuracy
- We might run into issues with matching the performance trend of different stocks to the actual performance. There are many external factors that cannot be perfectly modeled that affect the performance of a stock.

III. Anticipated relative contributions

Presentable	Goals	Student	Workshare
Dashboard with Tabs comparing prices and querying Api and data and etc	**Displaying data and modeling visualization (stock prices, models’ performances)**	Keith	80%
		Hoang	20%
Cleaned data for Recurrent Neural Network + RNN model to predict stocks with clear results based on multiple metrics	**Predicting short-term stock trend from historical data**	Hoang	80%
		Keith	20%
Clean, documented code of API scraper + Database of Tweets of Companies	**Scraping/Cleaning Twitter tweets based on # of different companies found by the stock team.**	Andrew	50%
		Zach	50%

Team 2 (Stocks/Social Media)
Progress Report 2
March 25, 2022

Dashboard with tabs comparing sentiment analysis metrics, and how accurate they are when measuring/predicting stock of a company	**Implementing sentiment analysis model on Twitter tweets**	Andrew	50%
		Zach	50%
		Zach	50%
		Andrew	50%
Various Data Charts/Tables that show Stock KPIs (Key Performance Indicators)	**Finding Stock KPIs (Key Performance Indicators)**	John	100%
Documentation (API, MKDocs)		Everyone	20% each