CS 410 Project Proposal
Group Name: cs-410-team-yoda

1. What are the names and NetIDs of all your team members?
   - Jonathan Birkey, jbirke2@illinois.edu
   - Dang Nguyen, dangn2@illinois.edu
   - Mathew McDade, mmcdade2@illinois.edu
   - Sakshi Maheshwari, smahes20@illinois.edu
2. Who is the captain?
   - Jonathan Birkey
3. What is the type of your project: Is it Data Set Creation or Leaderboard Competition Creation?
   - Data Set Creation
4. If your project is Data Set Creation, what is the novelty of your data set as compared with all the existing data sets?
   - The novelty of our data set comes from the coloration of stock price change and the amount of conversation that stock is receiving on Wall Street Bets. This can help investors predict when there will be massive swings in stock price volatility.
5. Which of the existing data sets is the closet to yours?
   - Our data set will likely be similar to the datasets found on datahub.io/collections/stock-market-data and reddit.com/r/datasets
6. What new task can your new data set be used to evaluate?
   - A user will be able to train algorithms against our dataset to see if their algorithm can predict the probability of a stocks rate of change given the amount of communication that stock is receiving on the subreddit Wall Street Bets.
7. How do you plan to create the data set?
   - We will use reddit's API to scrap data on the Wall Street Bets subreddit for a given timeframe. This will then be curated and formatted in a meaningful way that will be stored as a .txt file. We will then scrape the web for all the stocks in the S&P500 over the same timeframe. This will also be curated and stored with the stocks rate of change for each given day. This is what will be used to test a user's output against.