Team Spicy Chicken Stock

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
   a. [**Captain**] Danielle Yang (dy6) (Captain)
   b. Eric McCarthy (ericm4)
   c. Sabelle Huang (sabelle2)

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class
   a. We want to create a Chrome Extension that analyzes social media sentiment over various stock tickers. This will allow users to see how stock price movement is correlated with social media sentiment from various platforms, such as Twitter, Reddit (r/wallstreetbets, r/stocks, etc.), etc. This project was inspired from the craze during early 2021 where the 100x increase of stock price in Gamestop's stock (GME) was largely attributed to the subreddit known as r/wallstreetsbets.
   b. This relates to the class because we need to scrape various social media websites for data and then process and analyze the data for the user's needs.

3. Briefly describe any datasets, algorithms or techniques you plan to use:
   We plan to scrape data from Reddit, specifically stock-related subreddits such as
   - r/wallstreetbets
   - r/investing
   - r/robinhood
   - r/stocks
   - r/dogecoin
   - r/news

   Data Scale: On r/wallstreetbets, the most popular of the above stock subreddits, there typically 10,000 new comments per day (Source: https://subredditstats.com/r/wallstreetbets).

   If time permits, we may also analyze sentiment from other social media platforms such as Twitter.

4. How will you demonstrate that your approach will work as expected?
   a. We can start by analyzing stock sentiment for one stock and one subreddit over a short period of time. If this works well, we will increase our scope until we cover multiple stocks over longer periods of time.

5. Which programming language do you plan to use?
   a. Python, HTML/CSS, JavaScript
   Libraries:
   b. Tableau or D3 (for graphs)

      c. Selenium (for scraping)

      d. NLTK Vader (Sentiment analysis)

6. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.
   a. Code for scraping data (~15 hr)
   b. Scraping data from social media posts (~5 hr)
   c. Implement Sentiment Analysis Algorithm (~5 hr)
   d. Consolidating and displaying data in a meaningful way (~15 hr)
      i. Come up with heuristics to plot the data (~10 hr)
   e. Creating chrome extension (~10 hr)

At the final stage of your project, you need to deliver the following:
- Your documented source code.
- A demo that shows your implementation actually works. If you are improving a function, compare your results to the previously available function. If your implementation works better, show it off. If not, discuss why.