**CMSC 122 Project Proposal**
Team: Fi-Sci
Hannah Ni, Hannah Walter, Lin Su

**GOAL:**
Given the ticker or name of a company, our program will scrape from several news sources such as New York Times, The Guardian, USA Today, etc. to gather all the words associated with news articles about that company. We will then highlight the most common words, and using an existing sentiment lexicon, give an overall positive or negative score for news about the company. We will also draw from social media such as Twitter to track sentiments about the company and assign a social media score as well. Lastly, we will pull the stock data from Yahoo finance to display alongside the qualitative components we are scraping. The hope is to provide a well-rounded view on current sentiments about the company and how that might affect future stock movements. To provide significant analysis, we would only analyze the companies that have a sufficient amount of news coverage and social media discussions recently (e.g. Apple).

**DATA SOURCES:**
1. News sites to scrape
   - New York Times
   - USA Today
   - Huffington Post
   - Forbes
2. Social media sites
   - Twitter API
3. Quantitative Stock Data
   - Yahoo finance
4. Sentiment Analysis Lexicons
   - Bing Lu, Mingqin Hu sentiment lexicon
   - Many options

**POSSIBLE SENTIMENT ANALYSIS OPTIONS:**
Given time, we would love to analyze articles and tweets about a company using different sentiment analysis techniques. Some examples are the popular Naive Bayes sentiment classifier, as well as more advanced ones like Support Vector Machines.

**MAIN FUNCTION:**
- Input:
  - Name of Company, Ticker, Starting Date
- Output:
  - Top 10 words associated with a company
  - Social media sentiment score, from Twitter
  - News article sentiment score
  - Historical back-testing of our sentiment analysis showing strength of correlation between sentiment and stocks, along with recommendation to buy or sell