**Project Title**: Stock Analysis Utilizing Tweet Sentiment and Historical Correlation

**Team Members**: Andrew Kling, Joel Boyette, Charles Gaskins, TJ Zientek

**Project Description**: Use data from the SEC, twitter and the NYSE to track tweet sentiment and historical correlation regarding the top 50 companies in the NYSE by trade volume.

**Questions to Answer**:

1. Have any sectors of the S&P 100 significantly outperformed others or the average in the past 5 years?
2. Do S&P 100 companies tweet more or less than rest of S&P?
   1. Which (if any) S&P companies interact with twitter users (respond to comments / retweet other users)?
3. How do mergers and acquisitions affect stock price in the S&P 100?
4. What is the optimal amount of time to stay invested in a S&P 100 company? And on average when was the best time purchase and sell shares for S&P 100 companies?
5. Can we identify good predictors for stock price volume trade? Day of week? Quarterly earnings calls? Cyclical/Seasonal relation?

**Data Sources**:

* Historical stock data from kaggle
* Twitter API using tweepy library
* IEX Stock API
* SEC Dataset -Edgar (need to update with TJ’s dataset)

**Tasks**:

* Charles will focus on question 1
* Joel will focus on question 2
* TJ will focus on question 3
* Andrew will focus on question 4&5

BACKUP QUESTIONS TO ANSWER:

* Using historical Kaggle dataset:
  + What is the average amount of time for maximum profit?
    - Aka is time in the market more profitable than timing the market?
  + How many “false peaks” occurred during this maximum profit period?
    - False peaks would be where the stock price was going up and then started going down. Could count the number of new maxes before reaching the final one
    - Story: if large amount of peaks before final peak then this is more attributed to luck than anything
    - Correlate to SEC data?
  + Maximum change day over day and does this correlate with SEC data for that stock?
  + If we invested with a buy and hold methodology vs invested using dollar cost averaging or sliding window average is there a statistically valid difference in return? Certain time windows when this works better or worse?