Executive Summary
October 29, 2021

Executive Summary

For our project, we plan to utilize the skills that we have developed over the previous eleven weeks to create a machine learning model that predicts whether a stock will outperform the S&P 500 on a given day. We plan to use approximately fourteen years of daily stock price data for six companies in the finance and insurance industry. Our companies consist of two large, two medium, and two small companies based on market capitalization: J.P. Morgan, Goldman Sachs, Discover Financial Services, Synovus, Bank of Hawaii, and New York Community Bank. After accumulating and cleaning our data, we plan to run a logistic regression using the daily stock price data (high, low, close, open, volume, and percent change of stock price from previous day) to determine whether a stock's percent increase is higher or lower than the percent increase in the value of the S&P 500 index.

Exploratory Questions

1. Which subcategory of the Finance and Insurance industry leads in sales/revenue?
2. How accurately can we predict whether a stock price will perform better or worse than the S&P 500?
3. How is a company's market capitalization correlated with stock price changes? Do larger companies tend to perform better or worse compared to the S&P 500 than small or mid-cap stocks?
4. How do major events such as the Great Recession, Coronavirus Pandemic, and presidential elections impact stock performance compared to the S&P 500 performance? How do expectations affect these outcomes?
5. How do companies perform compared to one another in the stock market? Are there any common characteristics between the top performing stocks?
6. How do quarterly financial measures impact the performance of our model's predictions? How do our results change from merely using stock data?