

Smart Big Alpha Model

Financial Technology

Project Team

Chaitany Valluri
Jafar Bakhshaliyev
Nitika Sharma
Rick Nguyen

Fintech and Blockchain
Professor: Eric Zheng

University of Texas at Dallas
Texas 75080

5 October 2021

Abstract

This report contains statistically significant factors that explain the excess returns of five stocks by applying the regression method. The fundamental, technical, and sentimental analyses give information about how these five stocks perform in the market. The project aims to propose a performance of five stocks and report portfolio return according to the chosen trading strategy by analyzing all analyses.

Keywords: big alpha model, regression analysis, trading signal, portfolio

1. Introduction

The project aims to propose a smart big alpha model by applying fundamental, technical, and sentimental analysis. Our model seeks to identify statistically significant factors through statistical methods and regression analysis. The fundamental analysis complemented with technical and sentimental analyses provided comprehensive insights and a valid prediction of how the stocks would perform in the market. To generate an alpha model, we have focused on the stock prices of Tesla, Moderna, AMC Theatres, Cisco, and Netflix and its dependency on factors including, but not limited to, financial statement data and ratios, technical indicators, and social media sentiment for the analysis. Tesla is chosen because it is a growth stock, an electric vehicle manufacturer that is growing each year, and is widely discussed in the media with positive sentiment. Moderna was another company of interest since it is popular nowadays due to the pandemic, vaccination production, and discussions in the media. AMC Theatres was chosen because of its large volatility and Reddit sensitivity. Cisco and Netflix were chosen because their stock prices are affected by Covid-19, the lockdown, and the rising demand for distanced productivity and entertainment. After determining statistically significant factors, we seek to cultivate a working portfolio based on the expected future performance of the stocks by exploiting the insights gained from financial, technical, and sentimental analyses. Lastly, we will analyze the performance and report it in a visual way, including limitations and drawbacks of the model.

For generating an alpha model, we first extracted historical data of stock prices from Yahoo Finance. We found fundamental and technical indicators which include charts that helped predict the future performance of the stocks. In a sentimental analysis, we extracted information from Reddit to determine the general sentiment regarding each company. The analyses are done in R programming language by using libraries such as quantmod, PerformanceAnalytics, TTR, RedditExtractoR, NLP, and others.

2. Analyses

2.1. Fundamental Analysis

In fundamental analysis, we found factors from readily available data such as exchange-traded funds, quarterly financial statements, and quarterly macro-economic data. These factors have been combined and analyzed against each stock's performance. We believe that these factors account for the stock's returns. Financial statement data have been taken from Yahoo Finance and adjusted in Excel to calculate important financial ratios. Macro-economic data have been extracted from FRED with quantmod package of R. Exchange-traded funds were also extracted from Yahoo Finance same as stock prices.

Financial Statement Factors considered for each stock:

- 1) **D.E** - Debt to Equity Ratio - calculates the weight of total debt and financial liabilities against total shareholders' equity.
- 2) **CurrentR** - Current Ratio - measures the company's ability to pay short-term obligations.
- 3) **CashR** - Cash Ratio - measures the company's ability to repay its short-term obligations with cash and cash equivalents.
- 4) **AssetT** - Asset Turnover Ratio - measures how effectively companies are using their assets to generate sales.
- 5) **InventoryT** - Inventory Turnover Ratio - measures how fast a company sells its inventory.
- 6) **ROE** - Return on Equity - measures a corporation's profitability in relation to the stockholder's equity.
- 7) **Rdgrowth** - Research & Development growth - measures how quickly a company has grown its Research and Development expenditures.
- 8) **EPSgrowth** - Earning per share growth - illustrates the growth of earnings per share over time.
- 9) **ROA** - Return on Assets - indicates how well a company utilizes its assets in terms of profitability.
- 10) **Gross.Margin** - Gross Margin - represents each dollar of revenue after subtracting cost of goods sold.
- 11) **Book.Value** - Book Value - represents the total value of the company's assets that shareholders will receive if the company is liquidated.
- 12) **Stockholdergrowth** - Stockholder's Equity growth - measures growth of stockholder's equity over time.

Factors from the balance sheet of the companies generally illustrate the company's financial strengths and weaknesses. Whereas factors from income statements represent whether the business of the companies is profitable. These factors explain how effectively the companies operate, and we believe that these factors could help to explain the stock's returns.

Exchange-Traded Funds Factors and some stock price factors are as follows:

- 1) **XLE** - Energy Select Sector SPDR Fund - includes securities of companies from industries such as oil, gas, and consumable fuels, and energy equipment and services.
- 2) **AWAY** - ETFMG Travel Tech ETF - tracks securities of companies which capture "Travel Technology Business".
- 3) **USAI** - Pacer American Energy Independence ETF - measures portfolio of securities of companies that deal with energy infrastructure activities.

- 4) **GRID** - First Trust NASDAQ Clean Edge Smart Grid Infrastructure Fund - includes companies engaged in electric grid, energy storage and management, and others.
- 5) **LIT** - Global X Lithium & Battery Tech ETF - seeks performance of global companies involved in the lithium industry.
- 6) **PALL** - Aberdeen Standard Physical Palladium Shares ETF - reflects performance of the price of physical palladium.
- 7) **URA** - Global X Uranium ETF - tracks performance of global companies involved in the uranium industry.
- 8) **FAN** - First Trust Global Wind Energy ETF - measures performance of public companies engaged in the wind energy industry.
- 9) **MJO** - MicroSectors Cannabis 2X Leveraged ETNs - represents performance of companies that produce products derived from cannabis for medical use.
- 10) **CURE** - Direxion Daily Healthcare Bull 3X Shares - illustrates the performance of securities of companies from healthcare sectors including area of biotechnology.
- 11) **CTEC** - Global X CleanTech ETF - takes companies which benefited from technology focusing on improving efficiency of renewable energy production into account.
- 12) **HERO** - Global X Video Games & Esports ETF - provides exposure to companies that benefit from increased consumption related to video games and esports.
- 13) **CLOU** - Global X Cloud Computing ETF - represents performance of companies that are engaged in cloud computing technology.
- 14) **GAMR** - Wedbush ETFMG Video Game Tech ETF - tracks performance of securities of companies supporting or utilizing the video gaming industry.
- 15) **Stock Prices:** AAPL, DIS, AMZN, GM, F, TM, GME
- 16) **Index:** GSPC (S&P500)

These exchange-traded funds and stock prices as factors are good candidates for factor analysis as they mostly track performance of industries related to electric, healthcare, biotechnology, video gaming, cloud computing, and energy. Some ETFs factors which are linked to electric, lithium, uranium, and energy industries are generally related to Tesla, while some other factors that are linked to healthcare and biotechnology industries are related to Moderna, and so on. Some stock prices as factors will be helpful to analyze automobile, video-gaming, and other industries. In addition to this, we also added S&P 500 which we thought that stock market index (tracking performance of 500 large companies) will have effect on the stocks. We believe that there will be some factors out of a great number of factors will have explanatory power for the returns of our five stocks.

Macro-Economic Factors are as follows:

- 1) **Unemployment** - Unemployment Rate - represents the number of unemployed as a percentage of the labor force.
- 2) **GDP** - GDP Growth - growth of measure of U.S. output.
- 3) **Inflation_10Y** - the breakeven inflation rate which shows a measure of expected inflation derived from 10-Year Treasury Constant Maturity Securities and 10-Year Treasury Inflation-Indexed Constant Maturity Securities.
- 4) **Interest_10Y** - Interest Rates, Discount Rates for the United States.
- 5) **Bond_HQM** - 10-Year High Quality Market Corporate Bond Spot Rate.
- 6) **Avg_Earnings** - Average Hourly Earnings of All Employees, Total Private.
- 7) **Vehicle_Sales** - Total Vehicle Sales in the world.
- 8) **Avg_Elec_Price** - Average Price - Electricity per Kilowatt-Hour in U.S. City Average.
- 9) **Part_Dealers** - Retail Inventories - Motor Vehicle and Parts Dealers.

It is a well-known concept that some macro-event factors such as GDP, unemployment, inflation, etc. drive stock returns. Furthermore, we also added total cases of Covid-data for the United States to factor analysis.

2.2. Technical Analysis

As a technical analysis, we have done factor and chart analysis separately. From chart analysis, we scrutinized the five chosen stocks and narrowed the time frame to different horizons to seek a detailed picture of their performance.

Technical Analysis Factors are as follows:

- 1) **SMA** - Simple Moving Average - 20-day moving average of prices.
- 2) **EMA** - Exponential Moving Average - 14-day exponential moving average which puts more weight on the most recent data points.
- 3) **BB20_%B** - the location of the price found by Bollinger Bands: it will be greater than 1 if it is above the upper band and it will be less than 0 if it is below the lower band.

- 4) **RSI** - Relative Strength Index - provides technical traders with signals about bullish and bearish price momentum.
- 5) **'MACD – Signal'** - when MACD crosses above the signal line, it triggers a buy signal. Conversely, if MACD crosses below the signal line, it gives a sell signal.

Chart Analysis:

For TSLA, we can say that the stock price will likely increase after 31 August 2021 because the stock price currently does not cross over the upper band of Bollinger Bands and RSI increased to 60 but has not reached 70. Additionally, the MACD line crossed above the signal line. Therefore, every technical indicator prompted a buy signal.

For AMC, the stock price also will likely increase after 31 August 2021 even though the stock price crossed above the upper band of Bollinger Bands since RSI is over 65 but not over 70, and MACD line crossed above the signal line. By considering all information, the technical indicators gave a buy signal. However, technical analysis is not enough for AMC; sentimental analysis is also important. Looking at Bollinger Bands, we see that AMC prices are very much volatile because of Reddit WallStreetBets, media attention, and people's sentiment.

For MRNA, we can say that the stock price will likely decrease after 31 August 2021 because price is in downtrend and does not cross below the lower band according to Bollinger Bands. RSI decreases to 50 and will likely decrease below 40. Additionally, MACD line is below the signal line, all giving signs of a sell signal. However, there is still sentimental analysis about Covid-19 which needs to be analyzed.

For CSCO, we can say that the stock price will likely decrease after 31 August 2021 because price is in downtrend and does not cross below the lower band according to Bollinger Bands. Although the MACD line is above the signal line, RSI decreases from 70. The signal line also increases, which all indicates a sell signal.

For NFLX, we can say that stock price will first increase and then decrease after 31 August 2021 because price is in an uptrend and does not cross above the upper band of Bollinger Bands, but RSI increased to 75. MACD line is above the signal line, and the signal line also increased. Considering all the information, we predict that the stock price will increase momentarily then it will decrease. We need to analyze other factors and choose alternative time horizons to decide to buy or sell.

Figure 1 shows the technical charts for our chosen five stocks. To analyze the performance of stocks, we also made time-horizon narrower and saw a detailed picture of their performance with R package TTR. We only attached the longer time-horizon (two-years) chart that begins from 1 August 2019 to 31 August 2021 which can be analyzed below:

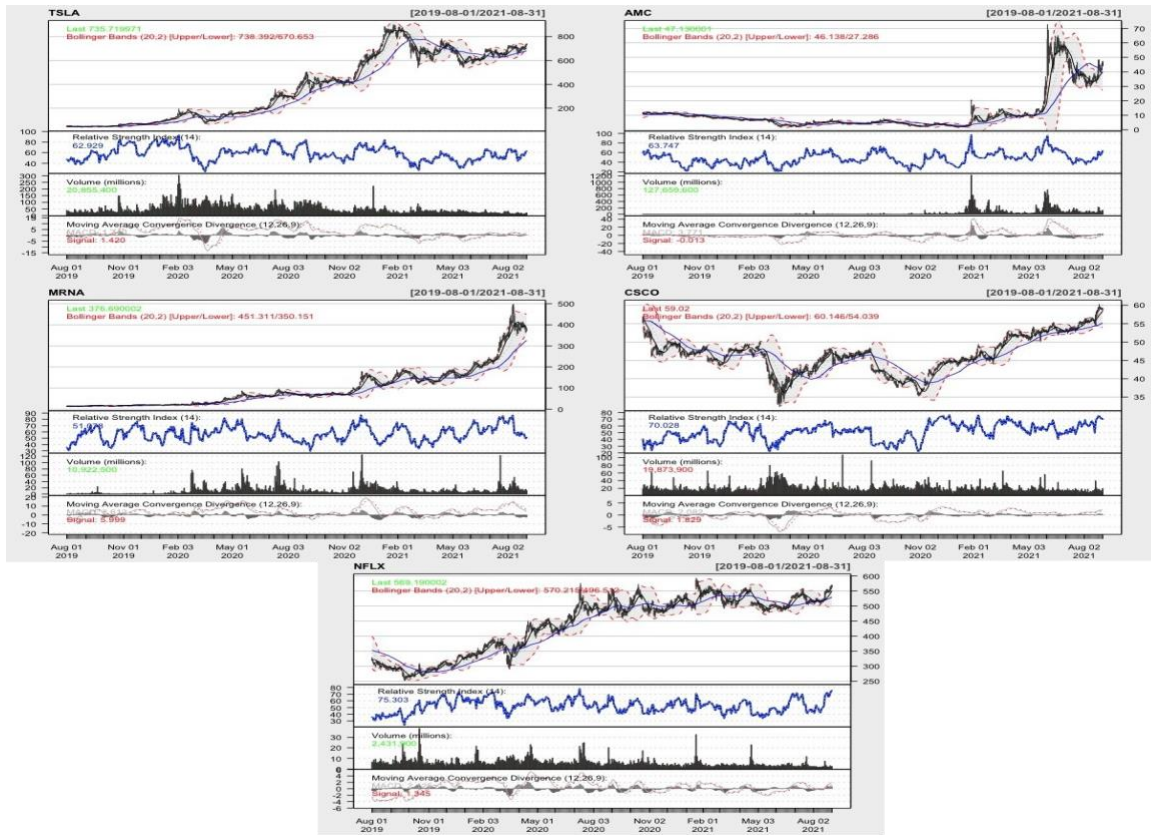


Figure 1: Technical Charts of Stock Prices

2.3. Sentiment Analysis

As a part of the project, we have done sentimental analysis and we extracted sentiment scores for these five stocks from Reddit. By using the RedditExtractoR and NLP packages of R, we got the sentiment scores for the last two - four months and plotted total scores for each date. We will not use these data for our factor analysis because the time horizon is very narrow. In our plot, we drew a red line for y-value (score) being 0 and green line for y-value (score) being 10. It generally shows that if data is below that red line, there has been a net negative sentiment toward the stock on the particular day, but if it is above the green line, it shows a high net positive sentiment on the stocks for the particular day. Our plot and analysis are provided in Figure 2 and below respectively.

For TSLA, we can see that there are more positive impacts when we look at the last month of the plot, from 2021 Aug 01 to 2021 Sep 01. So, it seems that the stock price will increase after 31st August and the analysis indicates a buy signal.

For AMC, we can see that there are more data points above the green line even for the last month, and therefore we can conclude that stock price will likely increase after 31st August 2021.

According to the MRNA plot, although some data are above the red line, near to the end of 31st August 2021, there is more data below the red line. In addition to this, there are generally very few data points above

the green line. We can therefore conclude that this analysis provides a sell signal but should be consolidated with other analyses to be more accurate.

For CSCO, we do not have much information from sentimental analysis for the last month. But looking at some data in July 01, they are mostly below the red line and hover around the red line which gives a possible sell signal.

According to the NFLX plot, there is more data that is either below the red line or near the red line, so we can conclude that this trend gives us a sell signal.

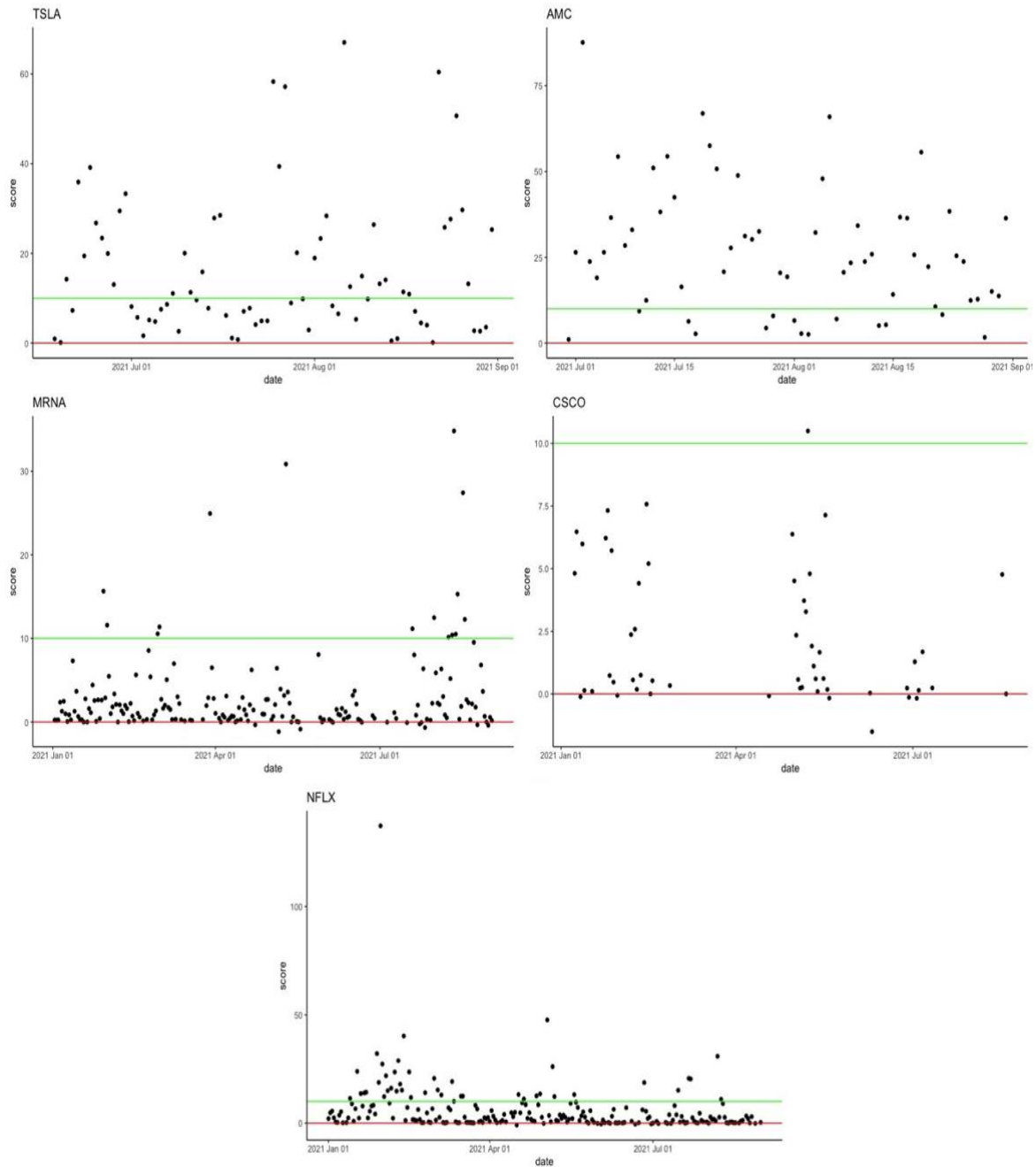


Figure 2: Sentimental Analysis of Stock Prices

3. Performance

To begin our performance, we have done linear regression analysis for five stocks considering the factors from fundamental and technical analyses to get alpha. To do linear regression, we have done some statistical analyses such as Variance Inflation Factor (VIF) and testing of assumptions of linearity by using car and gvlma packages of R. Some variables are not considered due to multicollinearity or missing values. In addition to it, we also considered spurious correlations. Adjusted R-square is high for nearly all regression analysis. The regression analysis and the statistical properties (in Figure 3) are given below:

For TSLA, we cannot get an alpha (intercept) that is statistically significant, so we will consider technical charts and sentimental analysis for our performance of TSLA. For AMC, we got a negative alpha (-2.727) that is statistically significant, and it gave a sell signal. According to the chart, MRNA has a negative statistically significant alpha (-2.144), but we should consider that degrees of freedom are low. For CSCO, we do not have any valuable information about the alpha (intercept). For NFLX, this analysis gives a sell signal because there is a negative alpha (-0.46) that is statistically significant.

```
Coefficients: TSLA
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.403269 0.358320 -1.125 0.2632
D.E 0.002881 0.003491 0.825 0.4112
CashR -0.019408 0.027978 -0.694 0.4895
AssetT 0.162180 0.106843 1.518 0.1323
SMA -0.002782 0.001364 -2.039 0.0441 *
Rdgrwth -0.027994 0.071930 -0.389 0.6980
Gross.Margin 0.016373 0.010816 1.514 0.1333
Stockholdergrowth -0.377905 0.179550 2.116 0.0200 *
XLE -0.199588 0.290409 -0.687 0.4936
GRID -0.101944 0.174413 -0.585 0.5602
LIT 0.039708 0.110178 0.360 0.7193
PALL 0.062916 0.095428 0.659 0.5113
URA -0.117543 0.209822 -0.560 0.5766
FAN 0.211981 0.129241 1.640 0.1042
AAPL 0.113526 0.144089 0.788 0.4327
GM -0.063794 0.148792 -0.429 0.6691
TM -0.102333 0.158055 -0.647 0.5189
GSPC -0.060353 0.460594 -0.131 0.8960
Unemployment 0.013350 0.008707 1.533 0.1285
Inflation_10Y 0.031833 0.035417 0.876 0.3811
Interest_10Y 0.019204 0.015325 1.253 0.2132
Bond_HQM -0.053016 0.024257 -2.186 0.0313 *
Avg.Earnings 1.516574 1.815418 0.835 0.4056
Vehicle_Sales 0.006271 0.008482 0.739 0.4615
Avg.Elec.Price -1.669303 2.205332 -0.757 0.4509
EMA 0.002683 0.001355 1.980 0.0506 .
BB20_3B -0.100367 0.055962 -1.793 0.0760 .
RSI 0.012034 0.004241 2.866 0.0022 ***
MACD-Signal -0.006531 0.004131 -1.581 0.1171
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07398 on 97 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared: 0.8255, Adjusted R-squared: 0.7733
F-statistic: 15.82 on 29 and 97 DF, p-value: < 2.2e-16

Coefficients: MRNA
Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.144e+00 3.085e-01 -6.951 1.26e-06 ***
D.E 3.521e+00 8.503e-01 4.141 0.000556 ***
EPSgrowth -1.750e-01 3.698e-02 -4.732 0.000145 ***
ROE -2.798e+00 5.073e-01 -4.684 0.000162 ***
Book.Value 3.664e+10 8.066e+11 4.543 0.000223 ***
GSPC 4.594e-01 2.905e-01 1.581 0.130268
Inflation_10Y -6.293e-01 1.158e-01 -5.432 3.06e-05 ***
Interest_10Y 2.225e-01 5.116e-02 4.348 0.000346 ***
Bond_HQM 1.407e-01 4.398e-02 3.200 0.004710 ***
SMA 2.189e-03 7.914e-04 2.767 0.012287 *
EMA -5.603e-01 1.469e-01 -3.813 0.001174 **
BB20_3B 3.020e-02 3.530e-03 8.554 6.11e-08 ***
RSI 1.102e-02 4.499e-03 2.449 0.024192 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07077 on 19 degrees of freedom
Multiple R-squared: 0.9511, Adjusted R-squared: 0.9202
F-statistic: 30.8 on 12 and 19 DF, p-value: 6.942e-10

Coefficients: CSCO
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.4618476 0.2474916 -1.866 0.064790 .
D.E -0.1446145 0.0883851 -1.636 0.104767
AssetT 0.2094496 0.3613825 0.580 0.563429
ROE -0.0384083 0.0771010 -0.498 0.619406
Rdgrwth -0.0593554 0.0693890 -0.855 0.394259
Gross.Margin 0.2256164 0.2889630 0.781 0.436672
Stockholdergrowth -0.2105042 0.1489864 -1.413 0.160612
AAPL -0.0636810 0.0482088 -1.321 0.189365
DIS -0.0753686 0.0656415 -1.148 0.253477
AMZN 0.0042783 0.0490962 0.087 0.930723
GSPC 0.7445996 0.1501770 4.958 2.72e-06 ***
Unemployment 0.0013615 0.0026675 0.510 0.610826
Inflation_10Y -0.0108632 0.0127884 -0.849 0.397538
Interest_10Y 0.0090890 0.0110857 0.820 0.414122
Bond_HQM 0.0080163 0.0080978 -0.976 0.339478
SMA 0.2108954 0.6782880 0.311 0.756470
Avg.Earnings 1.2284842 0.9945830 1.235 0.219496
Avg.Elec.Price -0.0002189 0.0005728 -0.382 0.703163
BB20_3B -0.0652600 0.0212488 -3.071 0.002709 **
RSI 0.0046829 0.0005225 8.962 1.21e-14 ***
MACD-Signal 0.0119639 0.0033052 3.620 0.000454 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03232 on 106 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared: 0.8173, Adjusted R-squared: 0.7829
F-statistic: 23.71 on 20 and 106 DF, p-value: < 2.2e-16

Coefficients: NFLX
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.4618476 0.2474916 -1.866 0.064790 .
D.E -0.1446145 0.0883851 -1.636 0.104767
AssetT 0.2094496 0.3613825 0.580 0.563429
ROE -0.0384083 0.0771010 -0.498 0.619406
Rdgrwth -0.0593554 0.0693890 -0.855 0.394259
Gross.Margin 0.2256164 0.2889630 0.781 0.436672
Stockholdergrowth -0.2105042 0.1489864 -1.413 0.160612
AAPL -0.0636810 0.0482088 -1.321 0.189365
DIS -0.0753686 0.0656415 -1.148 0.253477
AMZN 0.0042783 0.0490962 0.087 0.930723
GSPC 0.7445996 0.1501770 4.958 2.72e-06 ***
Unemployment 0.0013615 0.0026675 0.510 0.610826
Inflation_10Y -0.0108632 0.0127884 -0.849 0.397538
Interest_10Y 0.0090890 0.0110857 0.820 0.414122
Bond_HQM 0.0080163 0.0080978 -0.976 0.339478
SMA 0.2108954 0.6782880 0.311 0.756470
Avg.Earnings 1.2284842 0.9945830 1.235 0.219496
Avg.Elec.Price -0.0002189 0.0005728 -0.382 0.703163
BB20_3B -0.0652600 0.0212488 -3.071 0.002709 **
RSI 0.0046829 0.0005225 8.962 1.21e-14 ***
MACD-Signal 0.0119639 0.0033052 3.620 0.000454 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03232 on 106 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared: 0.8173, Adjusted R-squared: 0.7829
F-statistic: 23.71 on 20 and 106 DF, p-value: < 2.2e-16
```

Figure 3: Regression Analysis of Stock Prices

After consolidating the analyses, we have developed a chart which shows the trading signals for each stock below (in Figure 4). For TSLA, a buy trading signal is effective and achieved from technical chart and sentimental analysis. For AMC, although fundamental analysis gives a sell signal, we believe that the technical and sentimental analysis dominates because AMC stock price is volatile due to the dynamic Reddit WallStreetBets sentiment; so, buy signal is effective. For MRNA, all analysis factors indicated a sell signal. For CSCO, we mostly focused on technical analysis as we do not have strong information from fundamental analysis nor have much data from sentimental analysis, which indicated a sell trading signal. For NFLX, all analysis gave a sell signal, but from technical analysis we should wait a few days because price will first increase and then decrease according to RSI and MACD indicators.

<i>Which Analysis Dominates?</i>			
	Fundamental Analysis	Technical Analysis	Sentimental Analysis
TSLA	No Information	Buy	Buy
AMC	Sell	Buy	Buy
MRNA	Sell	Sell	Sell
CSCO	No Information	Sell	Sell (Not strong Info)
NFLX	Sell	Sell	Sell

Figure 4: Trading Signals from Analysis

Our trading strategy according to Figure 4 is that we will allocate weights 1.30, 0.50, -0.30, -0.20, -0.30 for TSLA, AMC, MRNA, CSCO, and NFLX respectively. We are giving more weight to TSLA rather than AMC because AMC is more volatile and stock price is subject to change. We will enter into the trade on 2nd Sept 2021 for the beginning of the month and we will sell at the end of the month (29th September 2021). But for AMC stock, we will give a one - week horizon and sell it on 10th September 2021, because this stock is generally very volatile and choosing short trading horizon for volatile stocks is a good strategy. According to Figure 5, we can say that our strategy is good, and our analysis worked well except for NFLX because NFLX increased though we shorted the stock. Our portfolio return derived by our strategy is 17.44 percent which shows very good performance. This portfolio return can be improved if we do technical charts and web-mining every day because some stocks might be sold before 29th September to achieve higher return and doing these analyses every day may also make NFLX profitable. In conclusion, our trading strategy works well and gives 17.44% return as our portfolio return.

STOCK	WEIGHT	RETURN
TSLA	1.3	0.06679499
AMC	0.5	0.13023882
MRNA	-0.3	-0.047176
CSCO	-0.2	-0.0680329
NFLX	-0.3	0.01785746
Portfolio Return	17.44%	

Figure 5: Portfolio Return

4. Conclusion

Some traders believe that fundamental analysis is the only important factor to trade stocks, while others believe technical analysis is the most important factor. We believe that it is important to consider both fundamental and technical analysis, while also considering sentimental analysis. This project consolidated all analyses such as fundamental, technical, and sentimental to construct a trading strategy. After creating our trading strategy, we reported our performance and a 17.44% return from our portfolio. In conclusion, we successfully determined statistically significant factors, cultivated a working portfolio based on the expected future performance of the stocks by exploiting the insights gained from financial, technical, and sentimental indicators, and successfully tested our big data alpha model to analyze its performance and reliability in generating excess returns.