PROJECT BULL VS. BEAR REPORT

GITHUB REPOSITORY:

https://github.com/Parth01202/Project-Bull-vs-Bear

PROJECT INTRODUCTION:

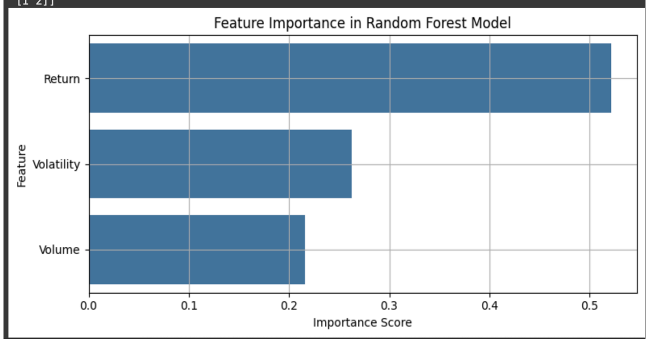
Project Bull vs. Bear aims to identify the performance of different stocks in sectors under bullish and bearish market conditions through historical stock market analysis. The big question we had was: Are we able to predict the performance of industries under bull vs. bear conditions? The importance of this analysis is that it can help minimize risks and boost returns in the market for expected market conditions through resources. The initial hypothesis was that stocks in sectors in healthcare and consumer services exceed others during bear markets, while technology and healthcare technology control bull markets.

DATA:

NASDAQ Data
Sector Information
Data is collected and analyzed from various sources to increase validity and reliability. From Yahoo Finance, datasets containing NASDAQ stocks with attributes data, close, high, low, open, volume, ticket, sector and industry from 2020-2024 are downloaded as CSV files. They are cleaned to ensure there are no missing values and necessary columns can be obtained. Additionally, Alpha Vantage API is utilized to fetch daily stock data such as open, close, and volume data from various companies. Marketstack API is used for its global stock tickers, indices, and historical time series data. The API keys allow stock trends to be analyzed, returns to be calculated and volatility to be assessed.

ML/STATS:

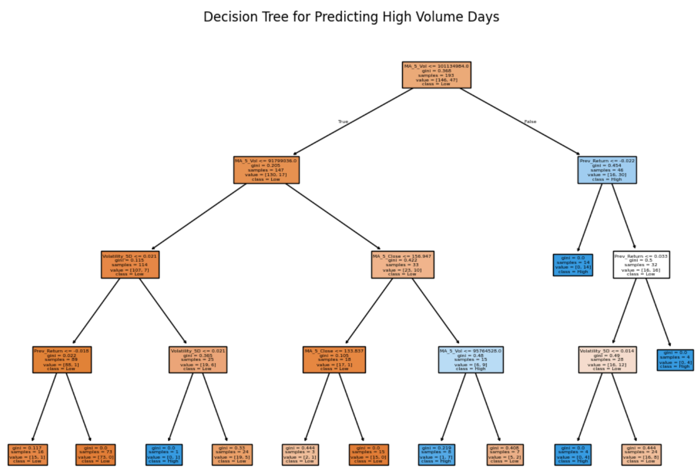
```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         1
           1       1.00      0.67      0.80         3

    accuracy                           0.75         4
   macro avg       0.75      0.83      0.73         4
weighted avg       0.88      0.75      0.77         4

Confusion Matrix:
[[1 0]
 [1 2]]
```



I utilized historical stock prices to develop a model that would forecast the potential increase or decrease of a stock price by the close of the trading day. I engineered features such as return (the magnitude of movement), volatility (the extent of variation), and volume (the quantity of shares moving). With these features, I trained a Random Forest classifier to facilitate the prediction. After conducting tests on the model, I was given a report that detailed the success of the model's performance. Further, I designed a bar chart illustrating which traits performed optimally with respect to prediction. The chart indicated that return and volatility were the most significant factors with respect to the decisions of the model, whereas volume had minimal impact. In general, the model provided useful feedback and identified the most beneficial stock indicators.

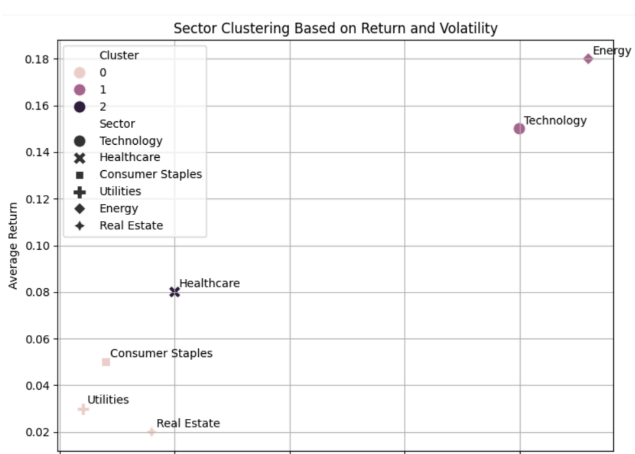
From the screenshot result, the model for Random Forest was 75% correct in that it picked 3 of the 4 test cases accurately. The model correctly labeled one of the sole down-trend, but miscalculated a test case predicting one of the up-trend as a down-trend. Whenever the model had predicted an up-trend, it was correct, but on one occasion when there was an actual up-trend, it failed to recognize it, lowering its recall rate by a minimal amount. The confusion matrix also corroborated this by showing two correct predictions for up-trends and one for a down-trend and one misclassifications. The feature importance plot also claimed that return was the most powerful feature in predicting, followed by volatility and then volume. This means that new price action and trends in price action are more useful to utilize than trading volume in trying to guess the movement of the stock during the day.

```
[[34  1]
 [ 8  6]]
              precision    recall  f1-score   support

           0       0.81      0.97      0.88        35
           1       0.86      0.43      0.57        14

    accuracy                           0.82        49
   macro avg       0.83      0.70      0.73        49
weighted avg       0.82      0.82      0.79        49
```



In an extension of our project's focus on stock performance during unstable periods, I looked into whether short-term price trends can be used to predict days of high-volume trading using a Decision Tree classifier. The plan was to see if return, volatility, and moving averages can predict investor action. The model reached 82% accuracy, with success in identifying the normal-volume days and capturing some of the spikes. The tree also pointed out which one of the indicators was most critical in making those choices, with return and volatility being the most effective e. This ties back into our larger idea by demonstrating that the same price action leading us to suspect direction in the market can also tell us how active investors are participating in bullish and bearish conditions.

While Dipesh used a Random Forest model to demonstrate how volatility and return impact price movement, the Decision Tree complemented that by demonstrating how those very same traits can be used to explain trading volume. The two models together suggest that short-term price patterns don't simply determine where the market goes, they determine how intense people get about trading when markets trend.



For the machine learning analysis, we used the K-Means Clustering method to group different stock market sectors based on how they behaved during the ups and downs from January to June 2020. The goal was to see if we could find any patterns in how different parts of the market reacted to the market crash and early recovery by looking at only their average return and how much their prices moved around. After cleaning the data, we standardized it so that all the sectors could be compared fairly. The K-Means algorithm was then used to divide them into three groups.

The results showed three types of behavior. The group that included Technology and Energy had high returns but also a lot of ups and downs which were riskier due to having high-reward sectors. The second group had Healthcare and Consumer Staples, which showed average returns but much lower risk and volatility. These sectors were stable performers, offering safer investments during volatile periods in the stock market. The last group, which included Real Estate and Utilities, had the lowest returns and lowest volatility, showing they were the least affected by market swings and mainly wanted to keep the invested amount safe. Overall, this showed that sectors tend to follow predictable patterns in times of market stress, and we can use simple data to show those trends.

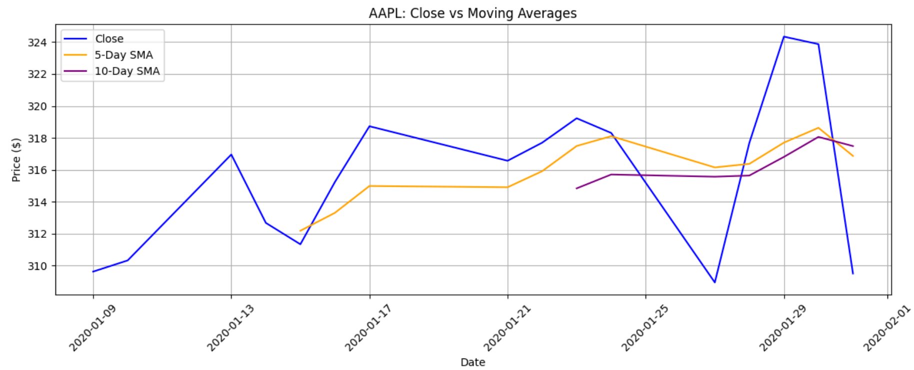
The machine learning analysis aims to cluster NASDAQ stocks based on their volatility profiles and uncovers patterns in the sector behavior. Stocks are grouped by mean and standard deviation of daily volatility and SMOTE is utilized to handle imbalance sector distributions. It shows that there are differences in risk profiles across the sectors such that high-volatility clusters include Technology and Healthcare stocks, while low-volatility clusters include utilities and Energy stocks. It further supports the hypothesis that Healthcare and Consumer Services perform better in bear markets and Technology perform better in bull markets.

The statistical analysis also aims to group stocks based on their mean and standard deviation of daily volatility. Highest sensitivity to market swings can be observed from the Mean_Vol_Mean of Healthcare (7.55%) and Industrials (7.07%), while resilience in bear markets can be observed in Financial Services (3.01%) and Utilities (5.02%). The high volatility range and moderate mean for Technology confirm bull market dominance and Consumer Defensive's low volatility confirms that it is stable in downturns.

Using plain features such as daily return, volatility and the 14 day SMA, I developed a Logistic Regression model that could predict whether the stock would go up or down. The data were obtained from the Alpha Vantage API and specifically for the AAPL stock from January 01, 2024 to March 31, 2024. Using feature engineering, I computed the SMA, volatility, and daily return, and labelled the stock either as bullish or bearish based on which the closing price was greater than the opening price.

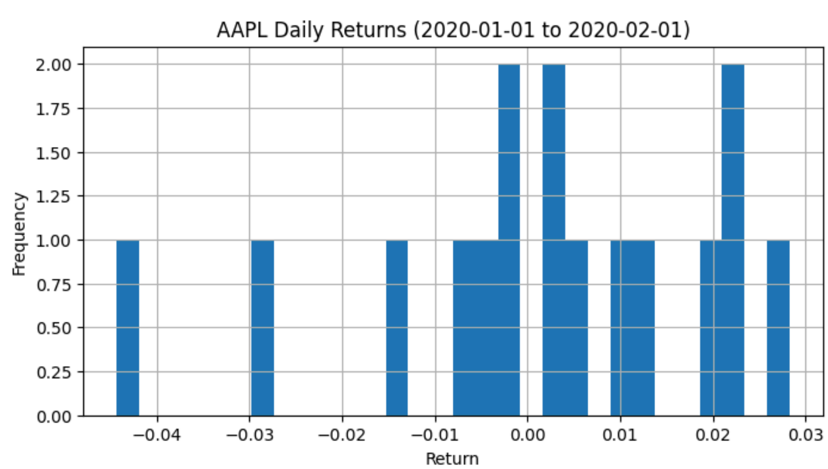
The model achieved an accuracy of around 70% - 80%, performing really well in predicting upward price trends in a span of a given time period. The main features that's driving the model's performance were the Simple Moving Average (SMA) and volatility, both of which were essential for forecasting stock prices. I also added a Random Forest Regressor to predict the closing price using these same features. The model had good results across multiple stocks, with measurable improvements reflected in the mean squared error and root mean squared error. These results were based on 5 stocks and it can work for multiple stocks as well. However, there's still room for improvement— I think more complex models and additional factors could lead to better answers/predictions.

VISUALIZATIONS:

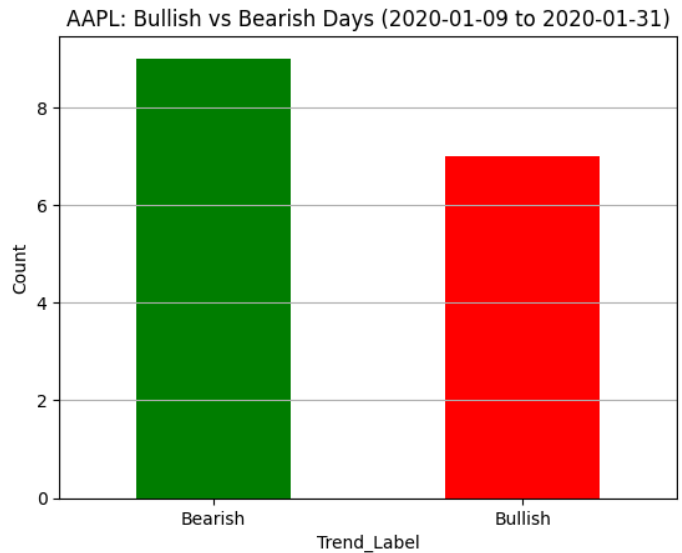


This graph shows how Apple's stock price (AAPL) evolved over time compared to its 5-day and 10-day simple moving averages (SMA). The blue line is the true closing prices, and the orange and purple lines are the 5-day and 10-day averages, respectively. Moving averages smooth out the price fluctuations to reveal underlying trends. Here we can see that the 5-day SMA is closely tracking recent action and that the 10-day SMA is more liquid and reacts behind it.

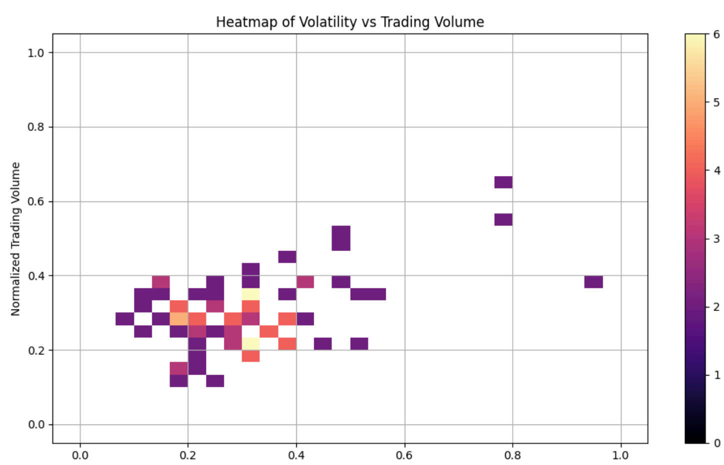
It is preceded by a visible price spike towards the close of the month where the closing price spikes well above both the moving averages, indicating strong short-term momentum. It drops sharply shortly thereafter, however. Since the close price breaks above the SMAs, it will indicate bullish momentum, and breaking below would mean a bear trend. This chart alerts traders to such trend reversals with ease and allow them to make better time management decisions. Overall, the visualization gives a better picture of how short-term trends are connected with longer-term trends.



This histogram charts the distribution of Apple (AAPL) daily returns for the month of January 1 to February 1, 2020. Most of the returns are near zero, reflecting that the stock mostly had small day-to-day moves. Nevertheless, there are various bars further from zero, demonstrating that there were indeed some days when the stock lost or gained more meaningfully. The graph does an effective job of charting overall volatility for the timeframe, giving the immediate sense of how frequently the stock made big moves versus how often it remained stable.



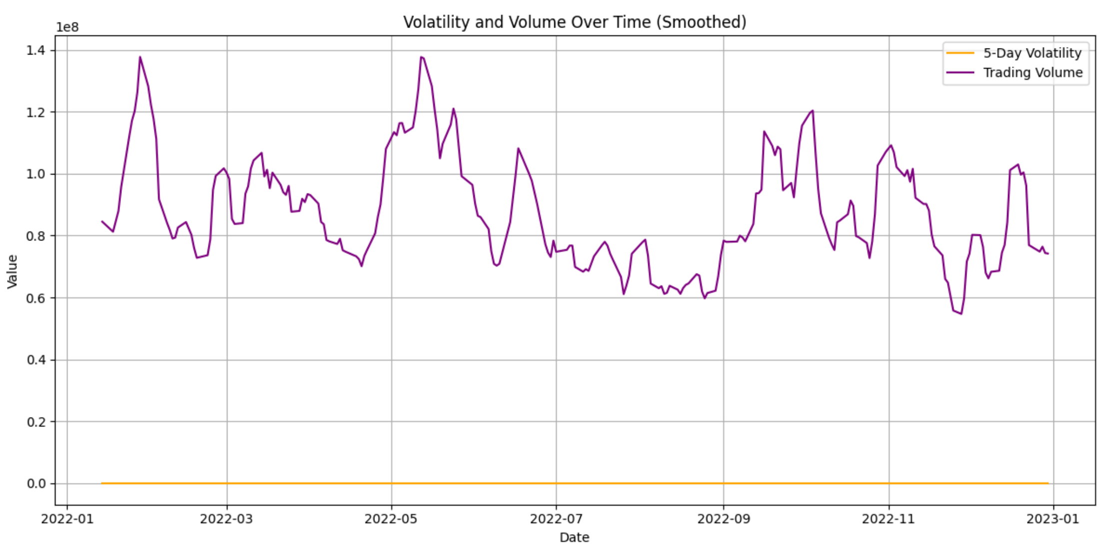
This bar chart emphasizes the number of bearish versus bullish days during Apple (AAPL) trading from January 9 through January 31, 2020. A bearish day is when the stock closed below where it opened, and a bullish day is when it closed above. More bear days (9) than bull days (7) occurred during this period since the stock saw slightly more downward movement in total. This simple visualization helps to obtain the general sentiment of the market for AAPL over the period chosen.

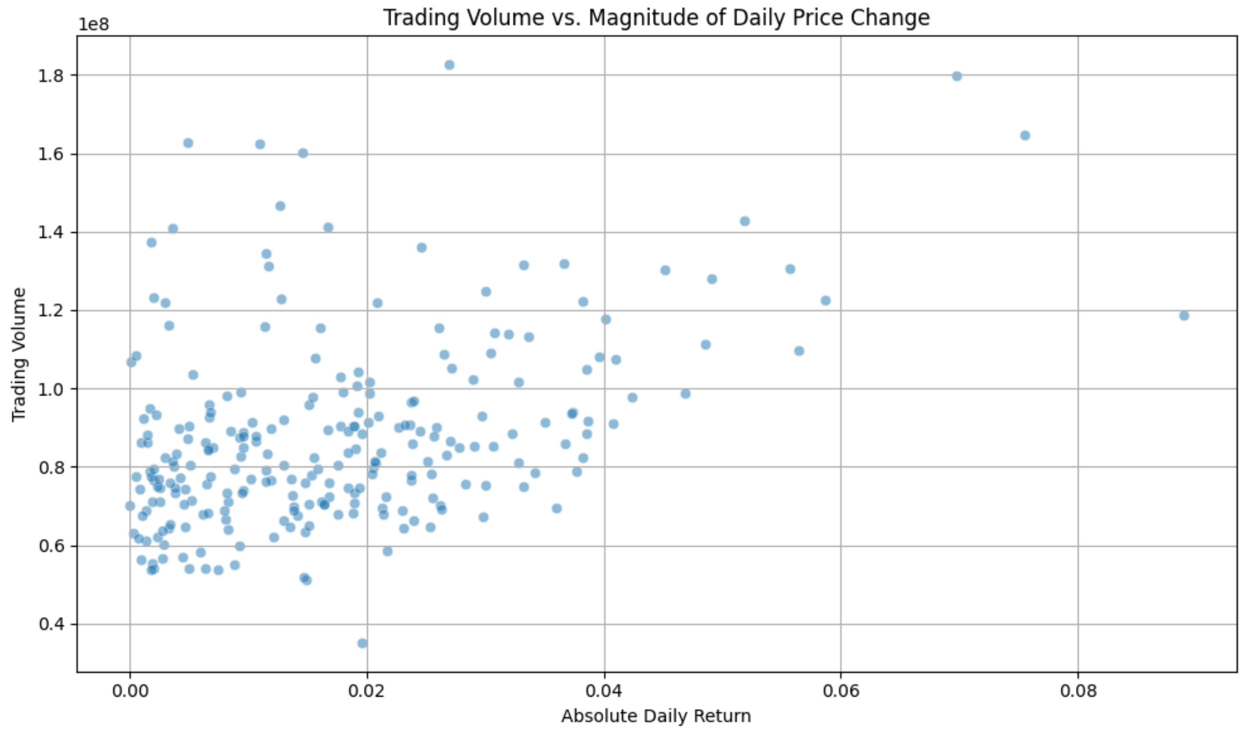


We hypothesized initially that there is a positive correlation between the short-run price volatility of a stock and its volume. Specifically, when the price of a stock exhibits price movements over 5 sequential days, it will attract more attention from the market and consequently generate more volume. With the investigation of this relationship using a heatmap, we aim to see whether this behavior consistently appears in previous trading patterns.

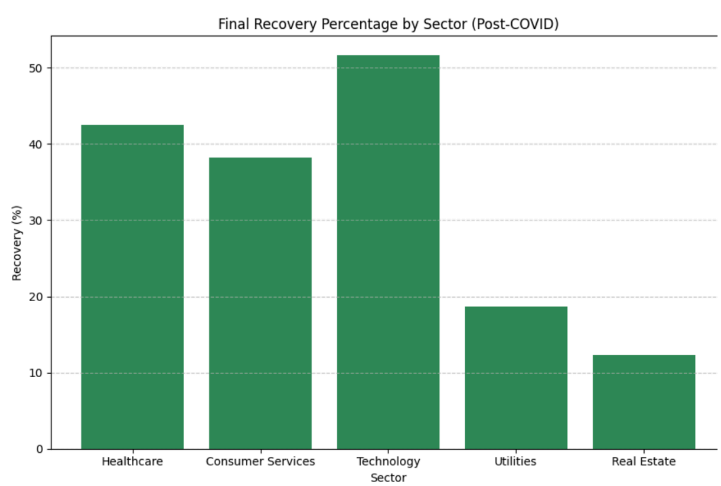
This heatmap essentially shows the correlation between normalized 5-day price volatility and normalized trading volume for AAPL stock in 2022. The densest regions of the plot are in the lower volatility and moderate volume category, indicating that most trading days have relatively stable price action with moderate trading. While there are a few cases of high volatility accompanied by corresponding high volume, distribution as a whole shows only a weak positive relationship. This lends support to our hypothesis to some degree that more volatility will be accompanied by more trading activity but also shows that volume is influenced by more advanced market mechanisms than basic price fluctuations.

To further explore this idea, and to further support our hypothesis, we plotted trading volume and 5-day volatility fluctuations throughout 2022. Initially, volatility appeared to be flat compared to volume due to scale differences. When we re-based the chart so that we were plotting them on different axes, we could notice that there were increases in trading volume corresponding to increases in volatility. This time-based analysis complicates our analysis even further by illustrating that when price action is short-term in nature, trading volume does rise, especially around market-moving events. Although the correlation on each individual occasion is moderate, this is what happens over time and supports our assumption that volatility attracts other market interest and volume short term.

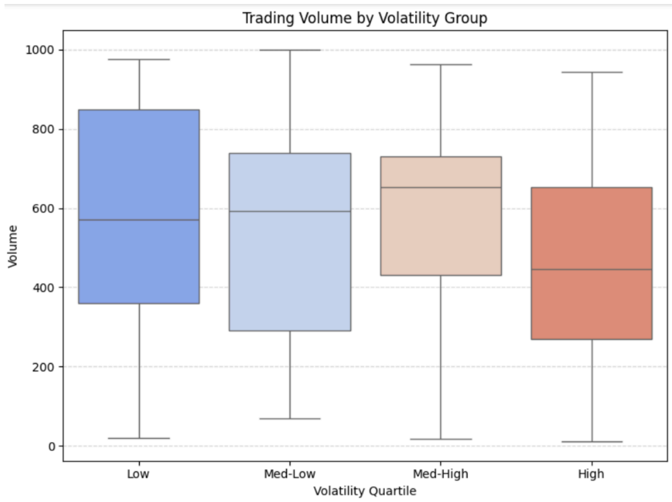




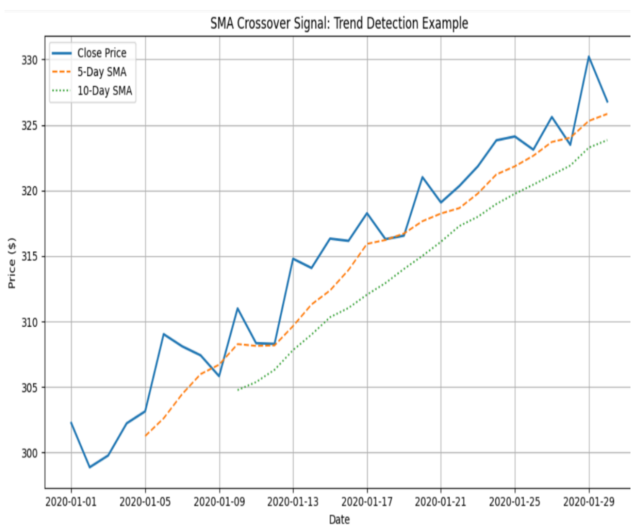
This scatter plot shows the relationship between trading volume and the magnitude of daily price changes of AAPL in 2022. The overall trend is the more significant a price move is, the higher the trading volume is. While most data points cluster in smaller returns and moderate volume, price changes of larger size (over 4%) oftentimes reflect heightened trading activity. This supports the belief that larger market movements attract more investor attention and participation.



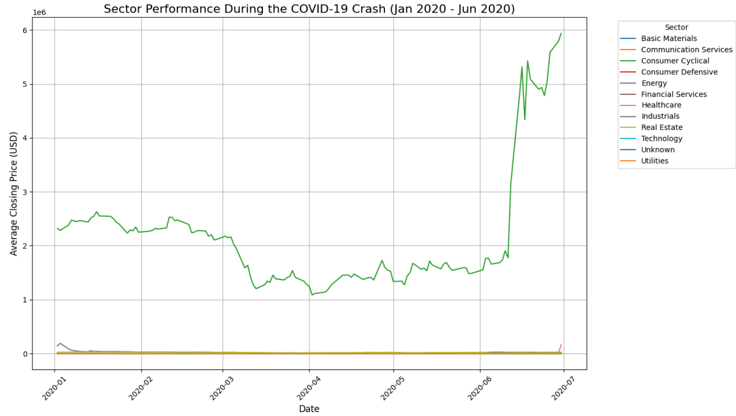
This bar chart visualizes which sectors recovered the most after the market crash. Technology leads in recovery, reinforcing that cyclical sectors were more responsive than defensive ones.



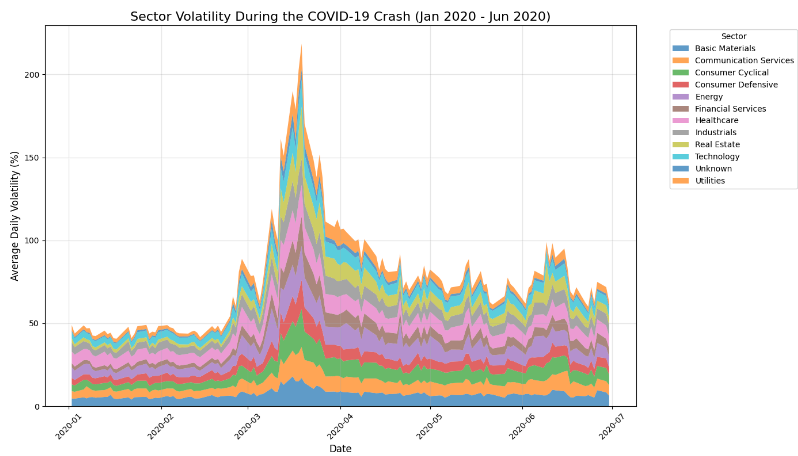
This boxplot confirms the hypothesis that more volatile stocks attract higher trading volumes, a trend highlighted in our findings slide.



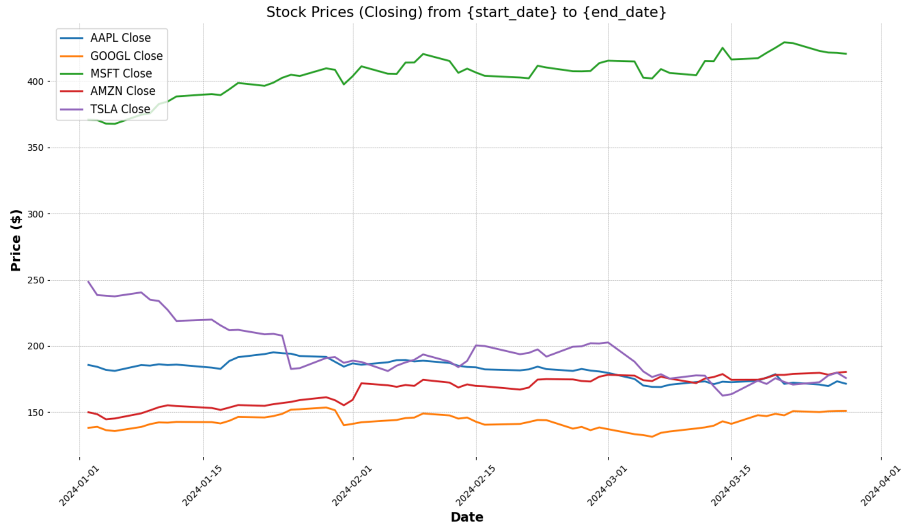
This visualization demonstrates a moving average crossover strategy. Bullish trends can be identified when the short-term SMA crosses above the long-term SMA — a key technique used to analyze trend momentum in our final findings.



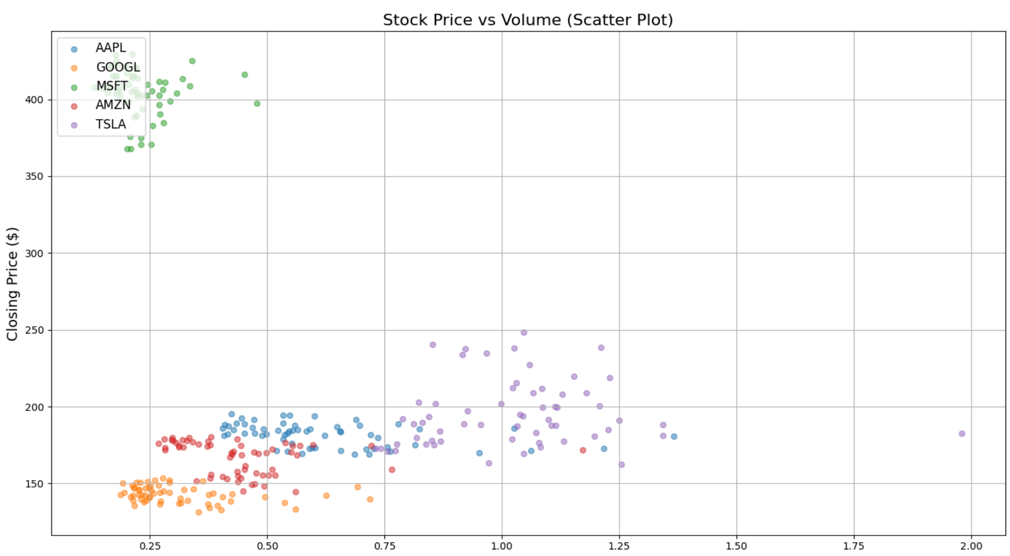
This visual aims to show the average daily closing prices around early COVID-19. It helps us understand how sectors perform during world-wide health crises like COVID-19. The consumer cyclical sector spiked significantly from January 2020 to June 2020. It is likely that it may be the outcome of sudden consumer behavior shifts, rise in e-commerce popularity and government stimulus distribution.



The stacked area plot shows the average daily volatility of NASDAQ stocks by sector during the COVID-19 market crash allowing us to observe sector trends in relation to each other. Daily volatility is calculated as (high - low) / open * 100. The plot highlights volatility spikes and the highest spike can be seen particularly during March 2020, where the pandemic visibly altered many aspects of daily life in reality.



The above graph illustrates the closing stock prices of five significant companies—AAPL, GOOGL, AMZN, MSFT, and TSLA—over a certain time period. Of these, Apple (AAPL) exhibits a particularly steady and rising trend, suggesting robust investor confidence and market resilience. On the other hand, GOOGL, MSFT, and TSLA show steady declines, indicating heightened volatility and possibly a pessimistic outlook for the market. AMZN shows only slight oscillations and appears to be flat. This graphic demonstrates the importance of historical data in assisting analysts and investors in making well-informed decisions and supports the theory that stock price movement and volatility can mirror more general market tendencies, such as bullishness or bearishness.



The scatter plot illustrates the relationship between stock price and trading volume. A concentrated pattern of higher prices with higher volumes may suggest a positive correlation, indicating that increased trading activity might drive up stock prices. If no clear pattern exists, it shows that volume may not significantly impact price movements in this period.

RESULTS / CONCLUSION:

This project allowed us to confirm that we are able to track the performance and behavior of various stock sectors during bull and bear markets. We observed that volatility and return are very useful indicators when predicting stock trends and investor activity. In bear markets, Healthcare and Consumer Staples perform better due to increased stability. However, in bull markets, Technology and Energy sectors perform better due to increased risk. Market behavior analysis and movement forecasting was explored through machine learning models like random forest, decision tree, logistic regression, and k-means clustering. Our findings were supported by various types of visualizations. This analysis can be utilized by investors to understand historical and present stock trends under different market conditions.