



Stock Prediction Analysis

Based on Seasonal Financial Report

Ruocheng Jiang

Data Source

The logo for finviz, featuring the word "finviz" in a dark blue, serif font. A small blue dot is positioned above the letter 'i'.

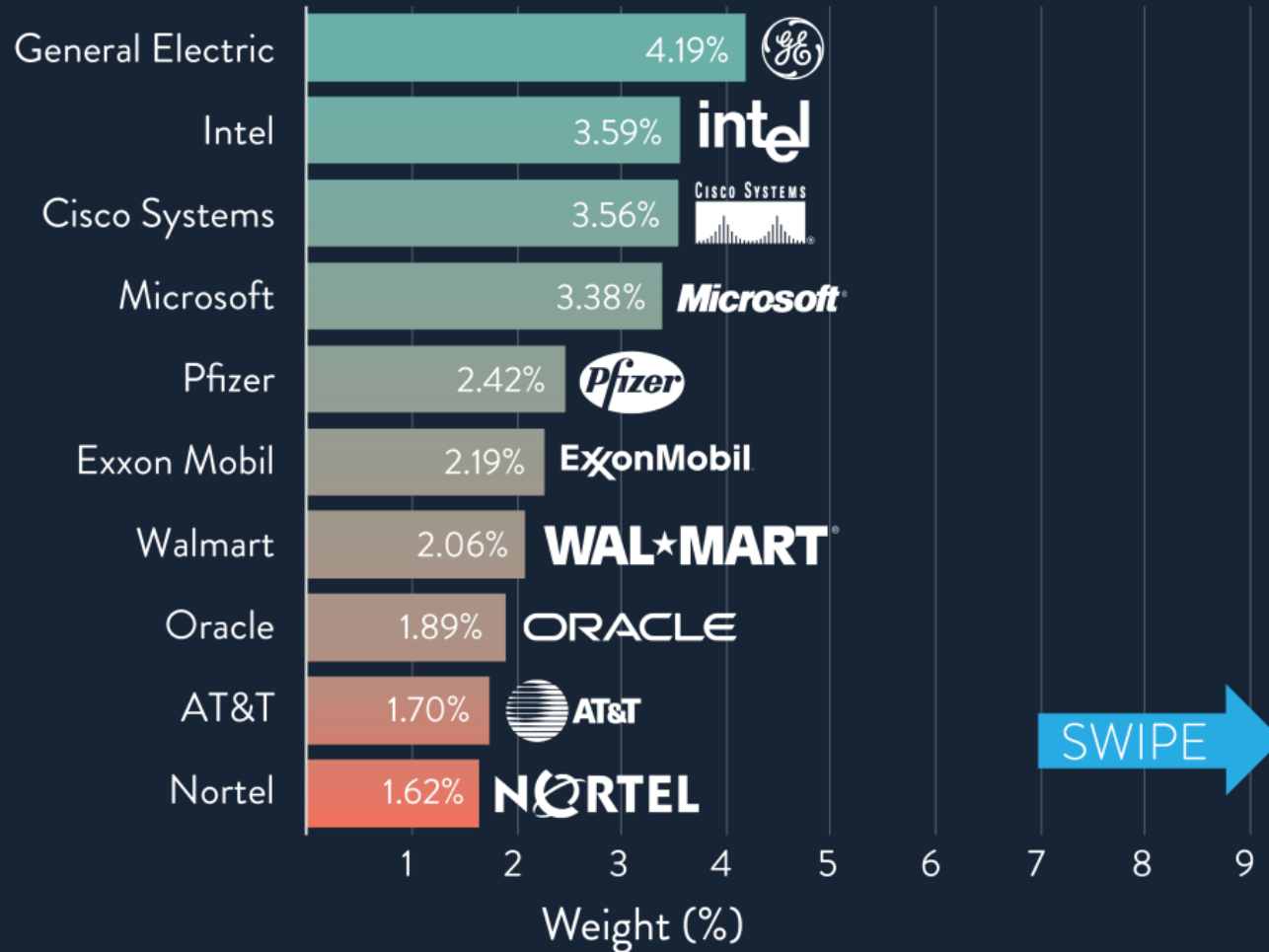
<https://finviz.com/>

The logo for Yahoo! Finance, featuring the word "yahoo!" in a bold, purple, sans-serif font, with the word "finance" in a smaller, purple, sans-serif font below it.

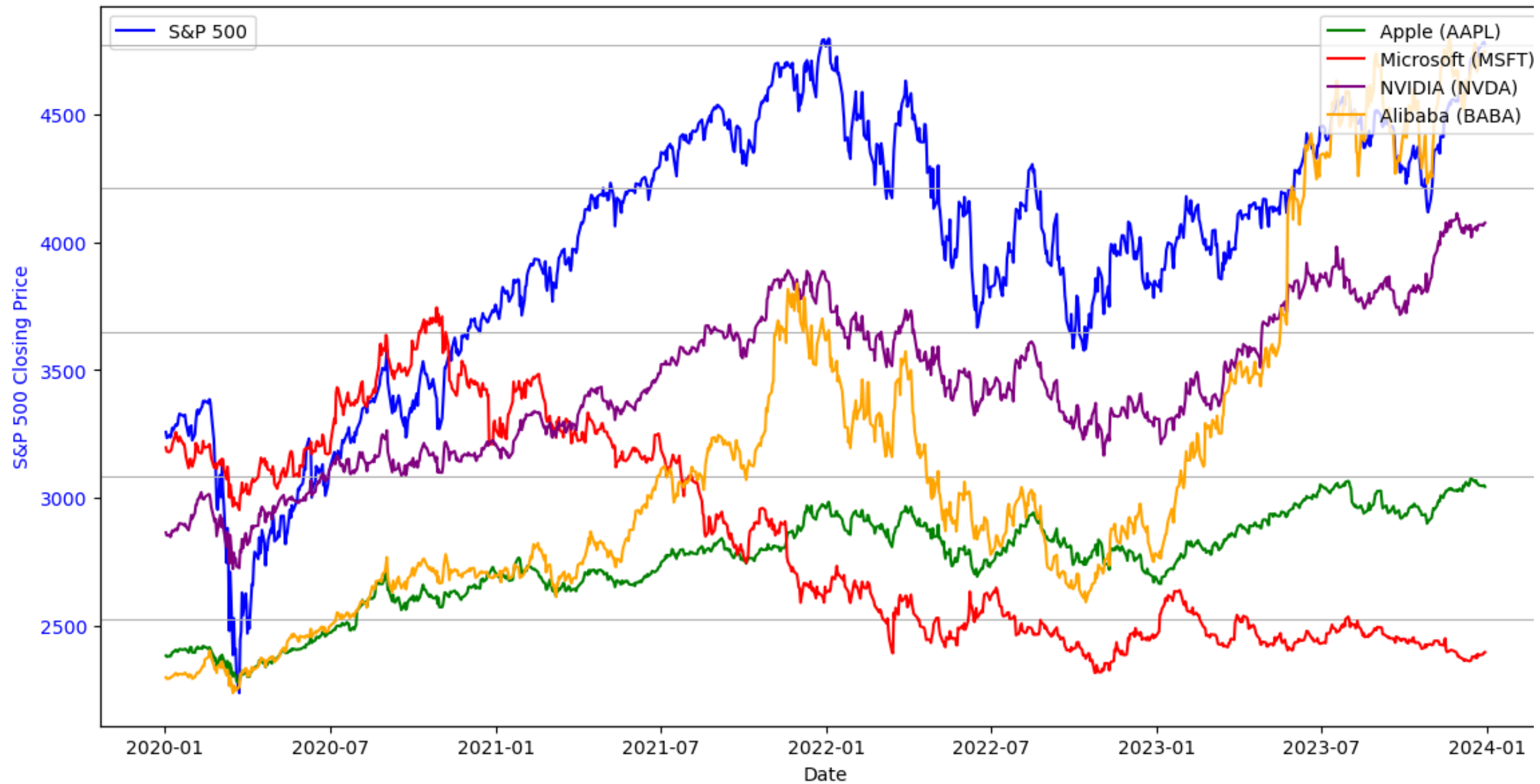
<https://finance.yahoo.com/>

TOP S&P 500 STOCKS

Here's what the list looked like in **June 2000**.
How many are still in the top 10 today?



Technology Stock Prices Over Time with Dual Y-Axis



Seasonal Changes?

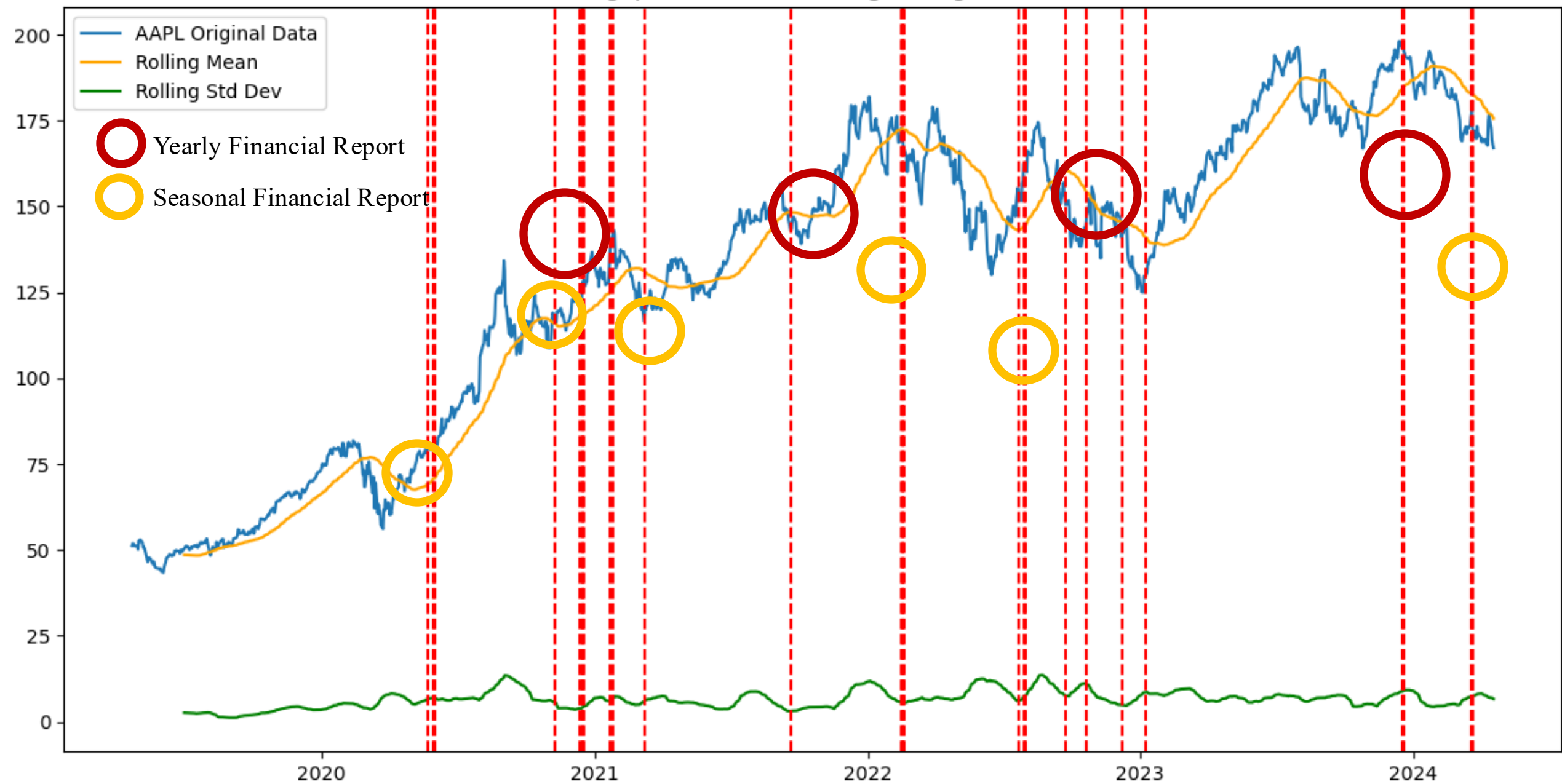
Changepoint Detection Analysis ...

Changepoint Detection using Rolling Statistics



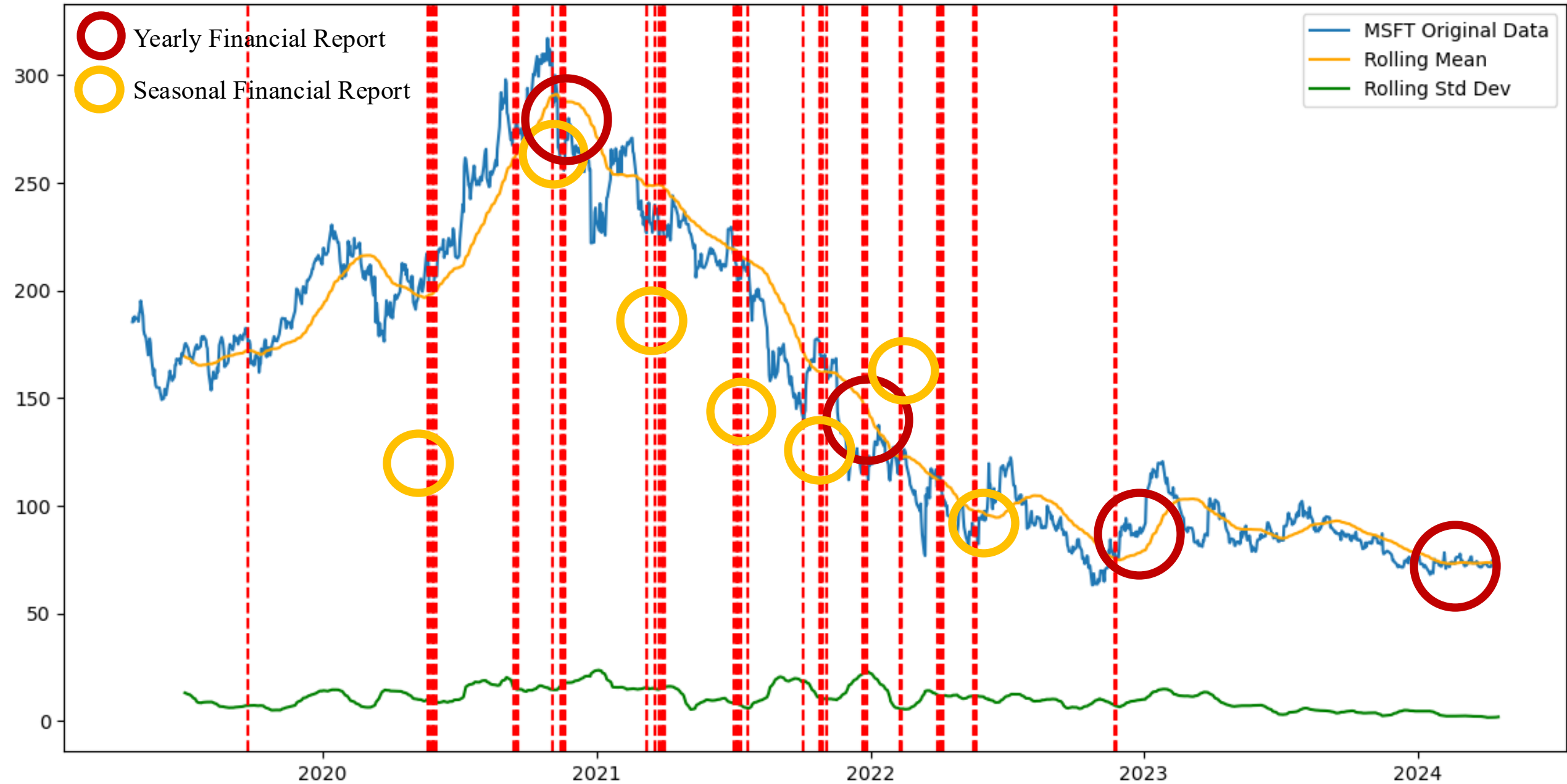
AAPL Stock Price

Changepoint Detection using Rolling Statistics

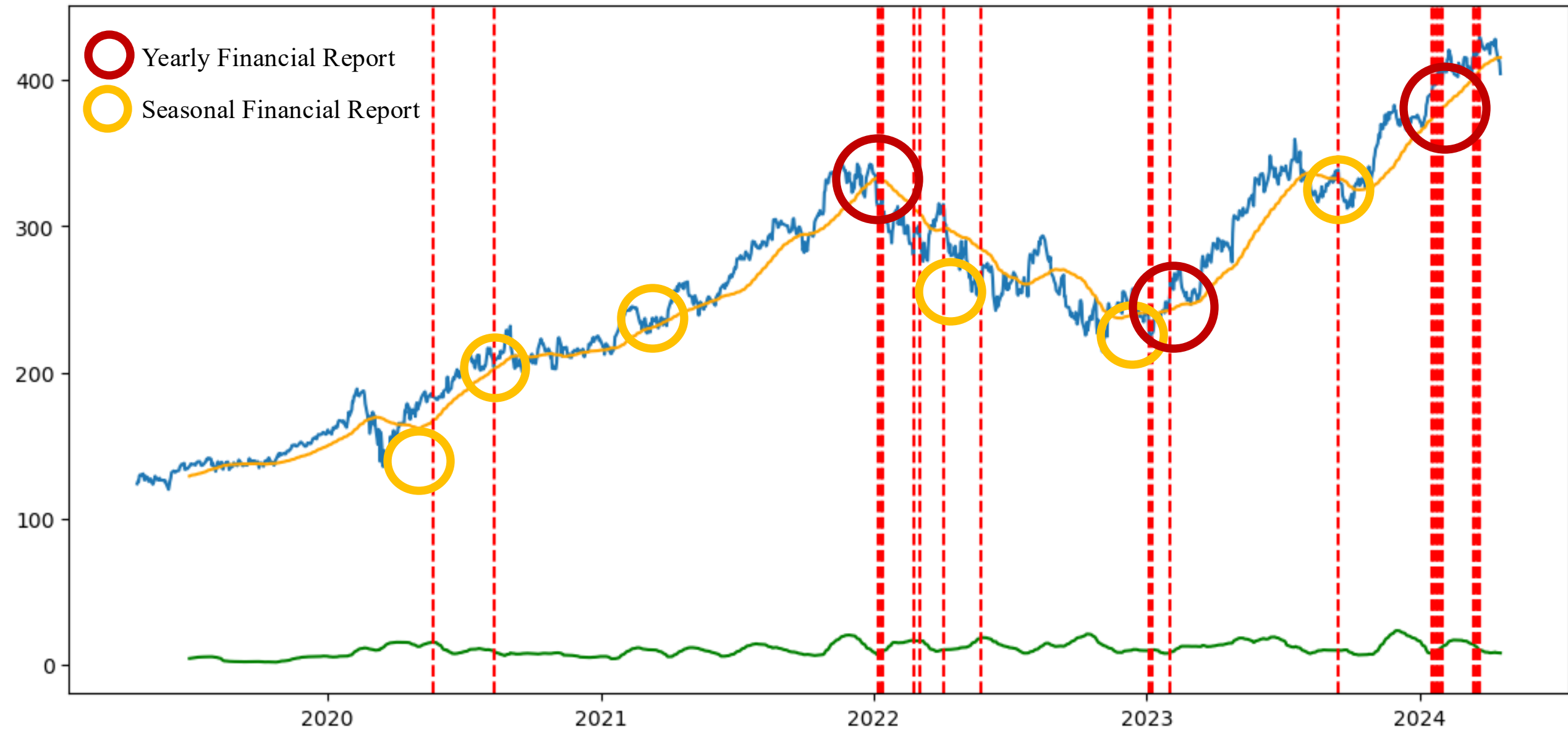


MSFT Stock Price

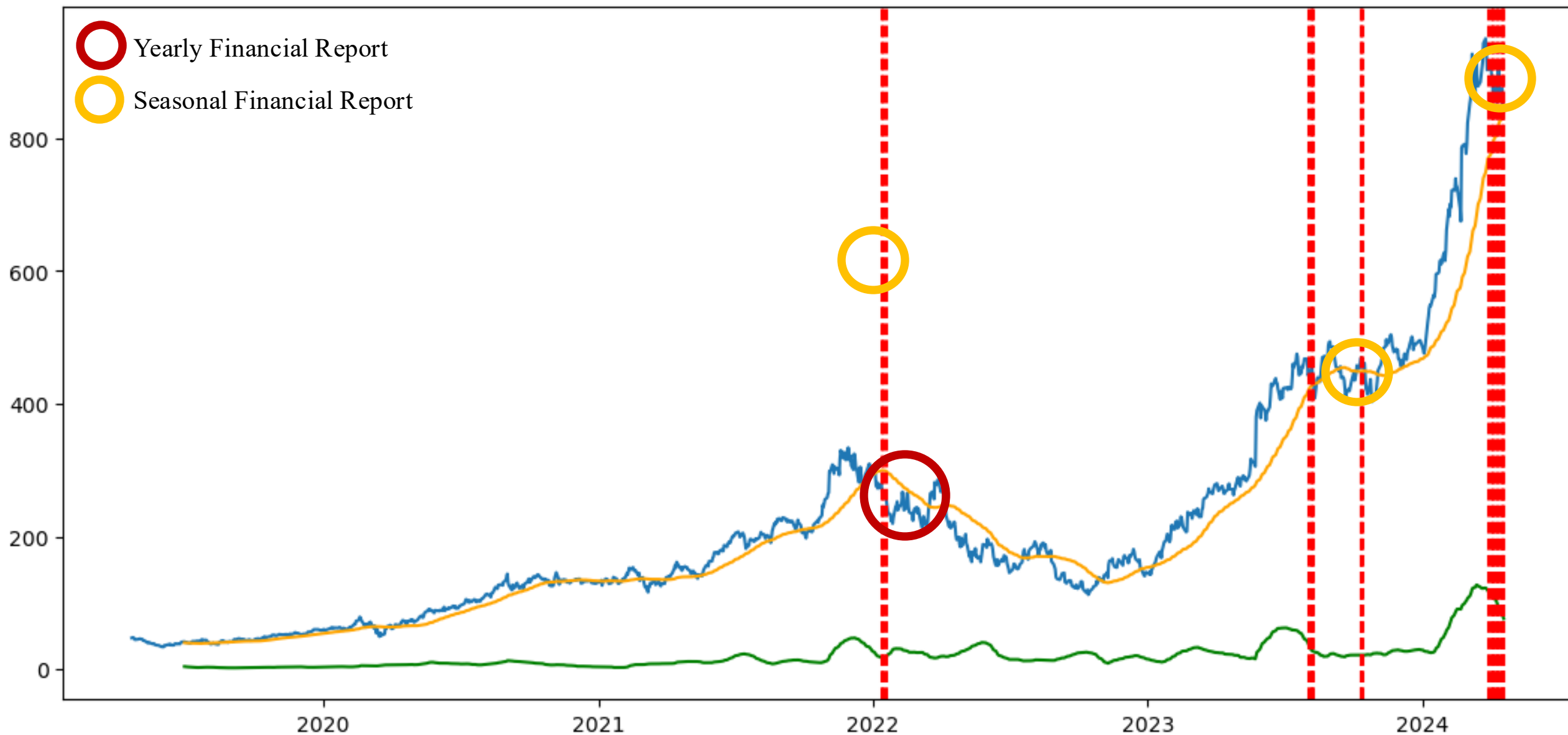
Changepoint Detection using Rolling Statistics



NVDA Stock Price



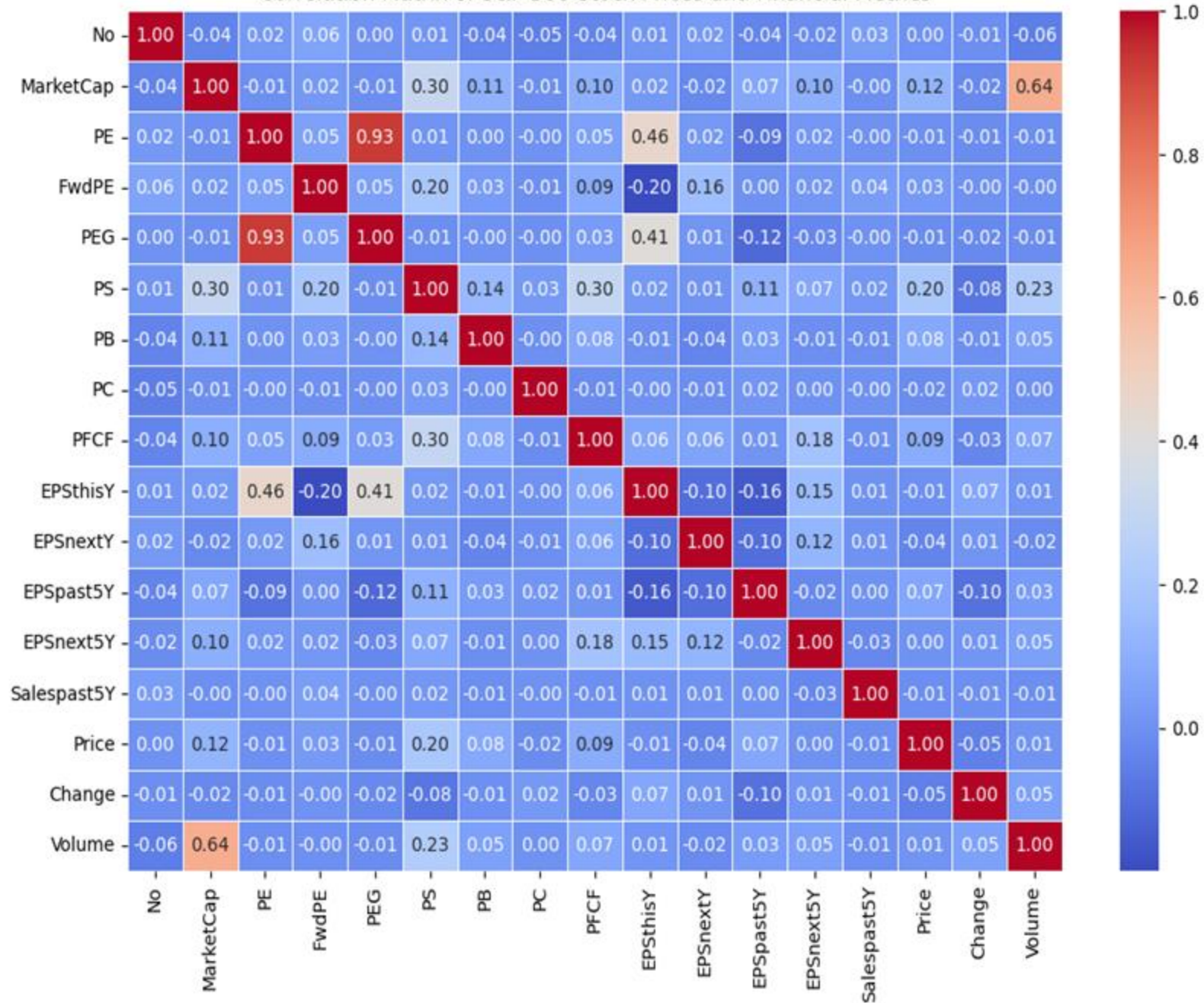
BABA Stock Price



What Financial Factors?

Let's see what correlation analysis says...

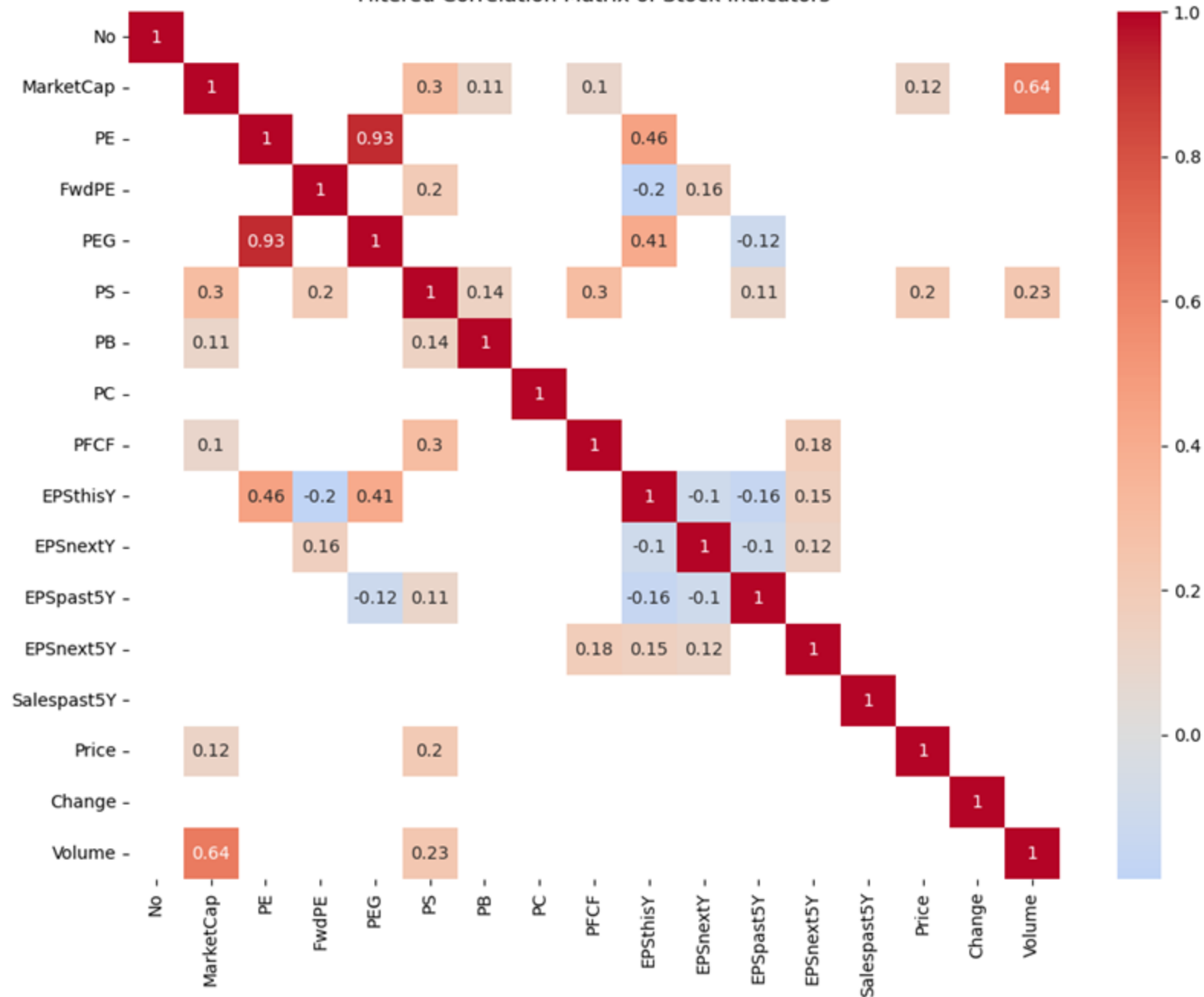
Correlation Matrix of S&P 500 Stock Prices and Financial Metrics



Heatmap of SP500

Too complicated?
Increase the thresholds

Filtered Correlation Matrix of Stock Indicators



Heatmap of SP500

PE and PEG (0.93)

Volume and MarketCap (0.64)

EPSthisY and PE (0.46)

EPSthisY and PEG (0.41)

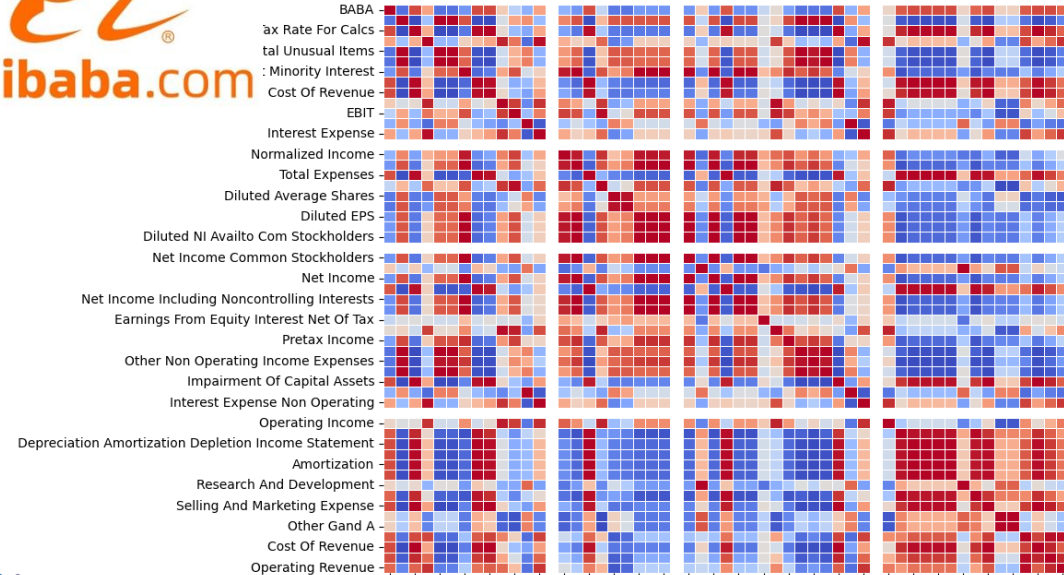
PFCF and PS (0.3)

Price Related Factors:

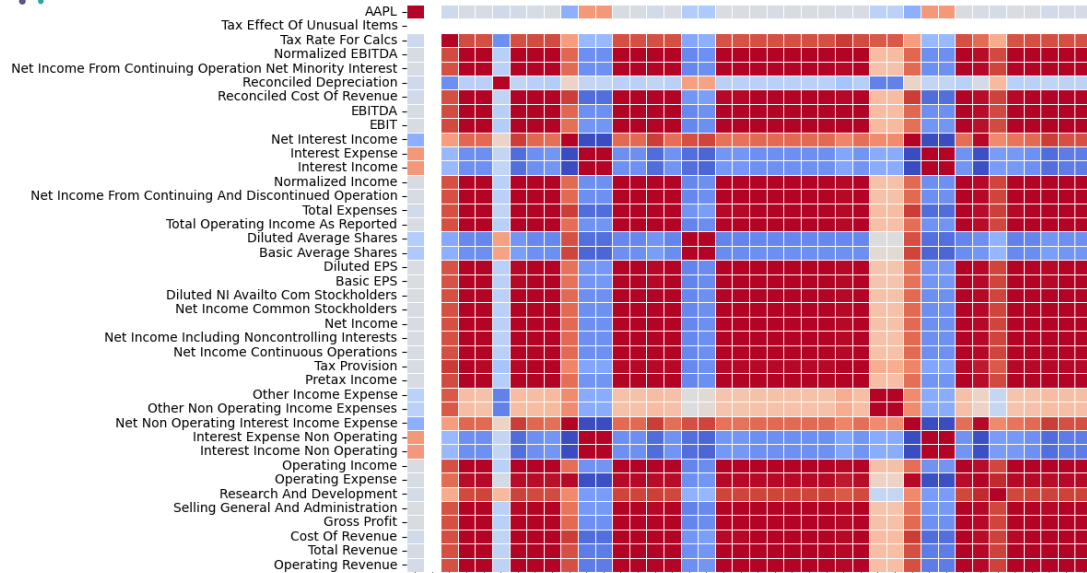
PS Ratio (0.20) = Stock Price / Sales



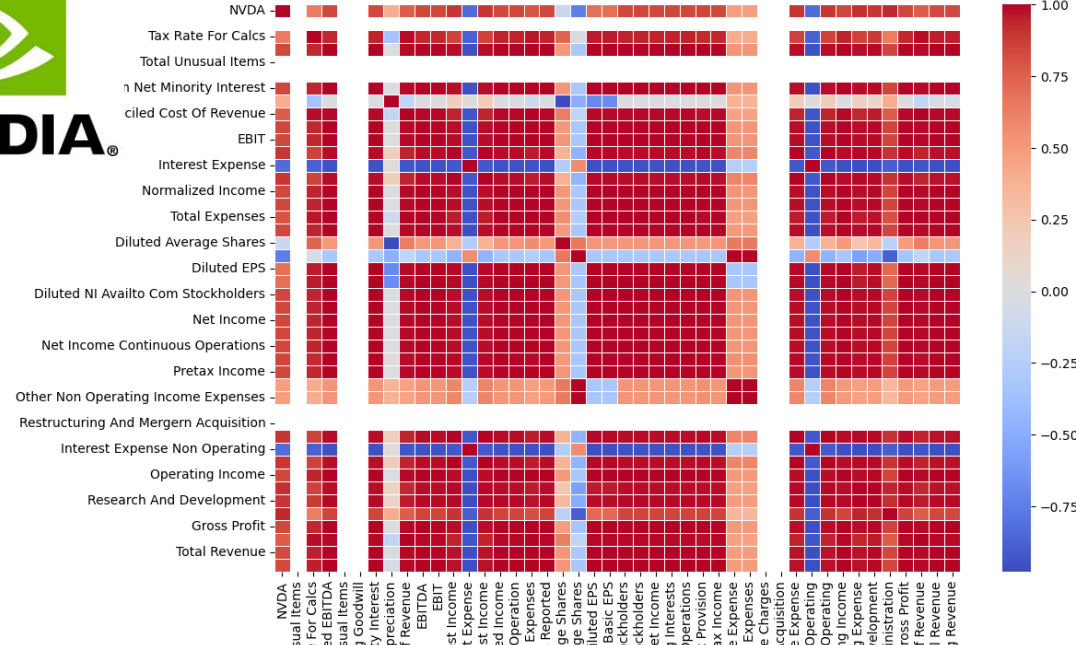
Correlation Matrix of AAPL Stock Prices and Financial Metrics



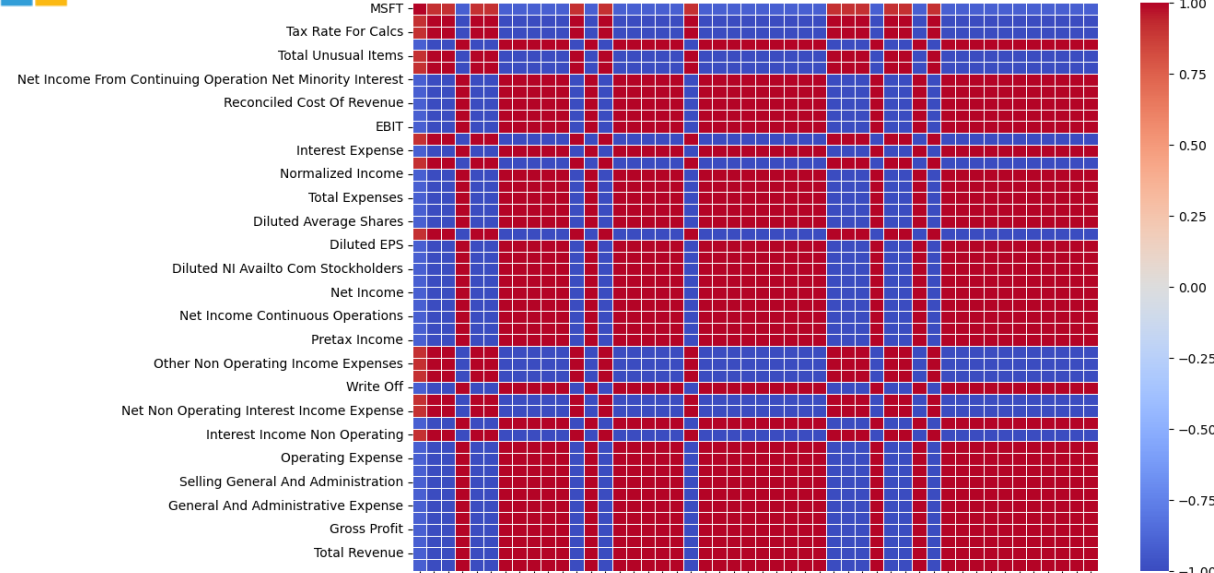
Correlation Matrix of AAPL Stock Prices and Financial Metrics



Correlation Matrix of AAPL Stock Prices and Financial Metrics



Correlation Matrix of MSFT Stock Prices and Financial Metrics



Financial Factors be Added to Model

- 'Interest Income Non Operating'
- 'Interest Income'
- 'Interest Expense Non Operating'
- 'Interest Expense'
- 'Basic EPS'
- 'Net Interest Income'
- 'Net Non Operating Interest Income Expense'
- 'Basic Average Shares'
- 'Diluted Average Shares'
- 'Net Income'
- 'Operating Income'

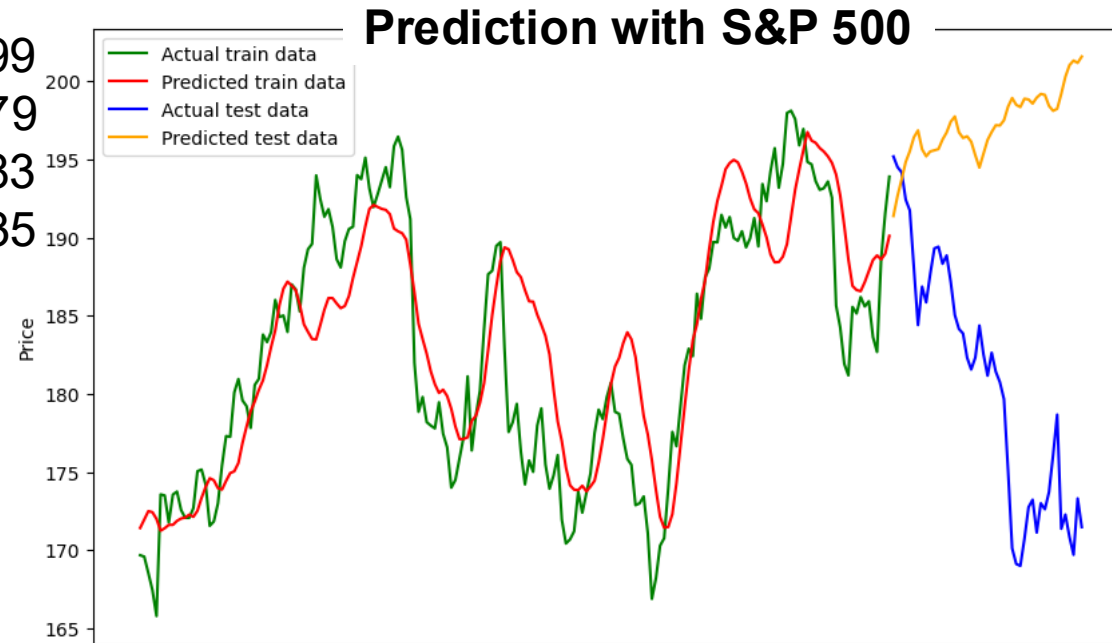
Long Short-Term Memory (LSTM) Model

- **Normalization:** Using MinMaxScaler to scales the data to a fixed range of 0 to 1 (speed & performance)
- **Splitting Dataset:** 80% Training, 20% Testing
- **Model Architecture:** Sequential architecture consisting of bidirectional LSTM layers (past and future, accuracy)
- **Evaluation:** Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared value, quantify the model's performance in terms of error magnitude and prediction accuracy

Now Let's See the Result...

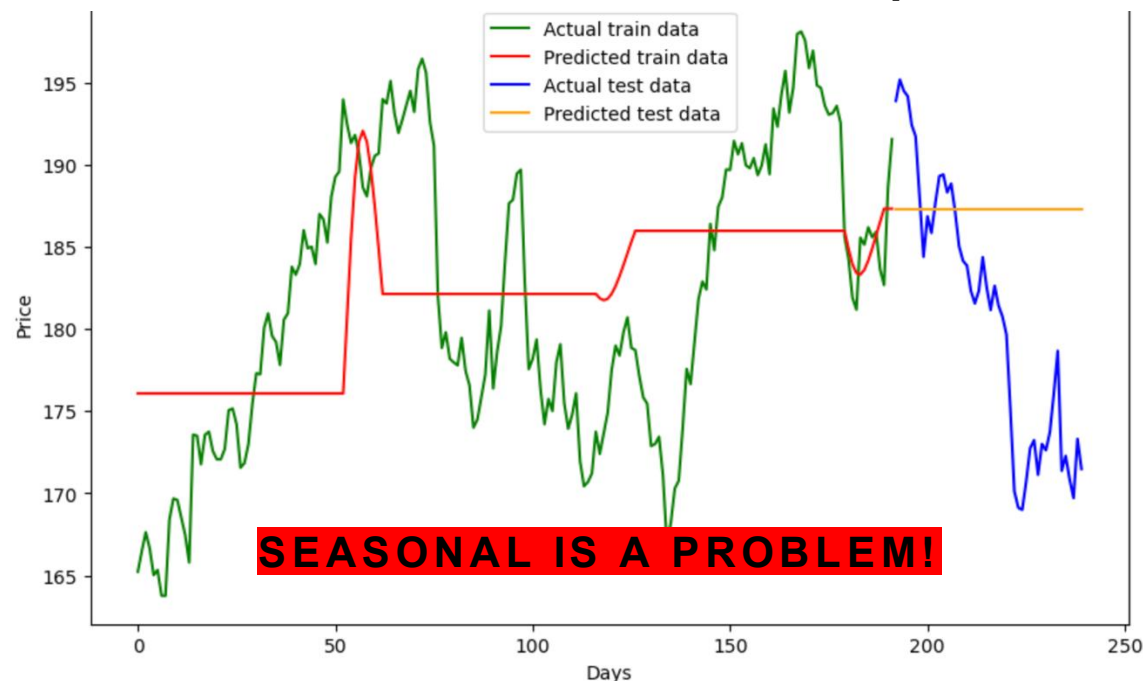
Long Short-Term Memory (LSTM) Models

Mean Squared Error: 32399
Root Mean Squared Error: 179
Mean Absolute Error: 179.83
R-squared: -534.35



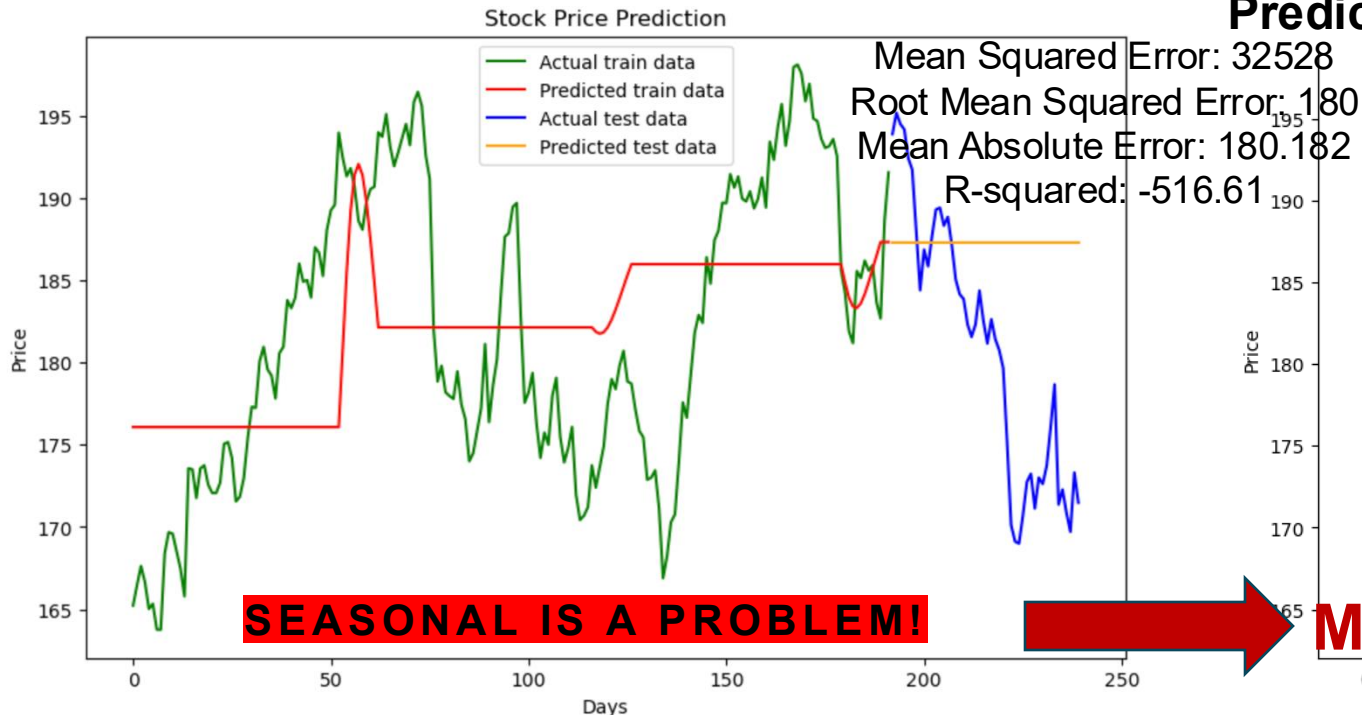
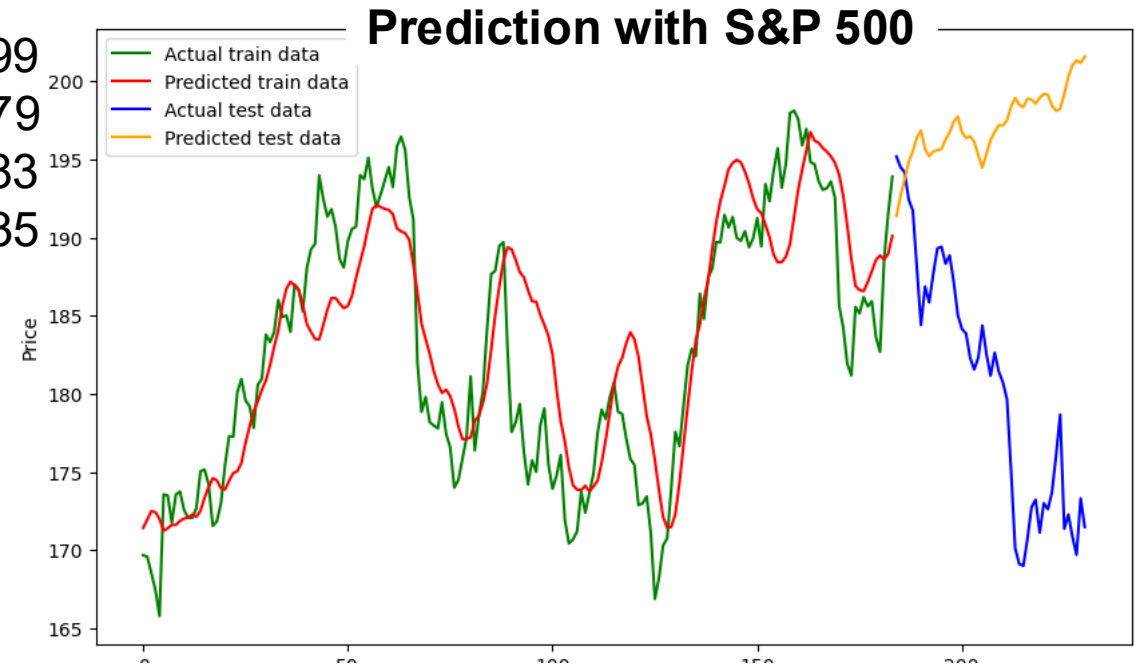
Mean Squared Error: 32528
Root Mean Squared Error: 180
Mean Absolute Error: 180.182
R-squared: -516.61

Prediction with Seasonal Financial Report Factors



Mean Squared Error: 32399
Root Mean Squared Error: 179
Mean Absolute Error: 179.83
R-squared: -534.35

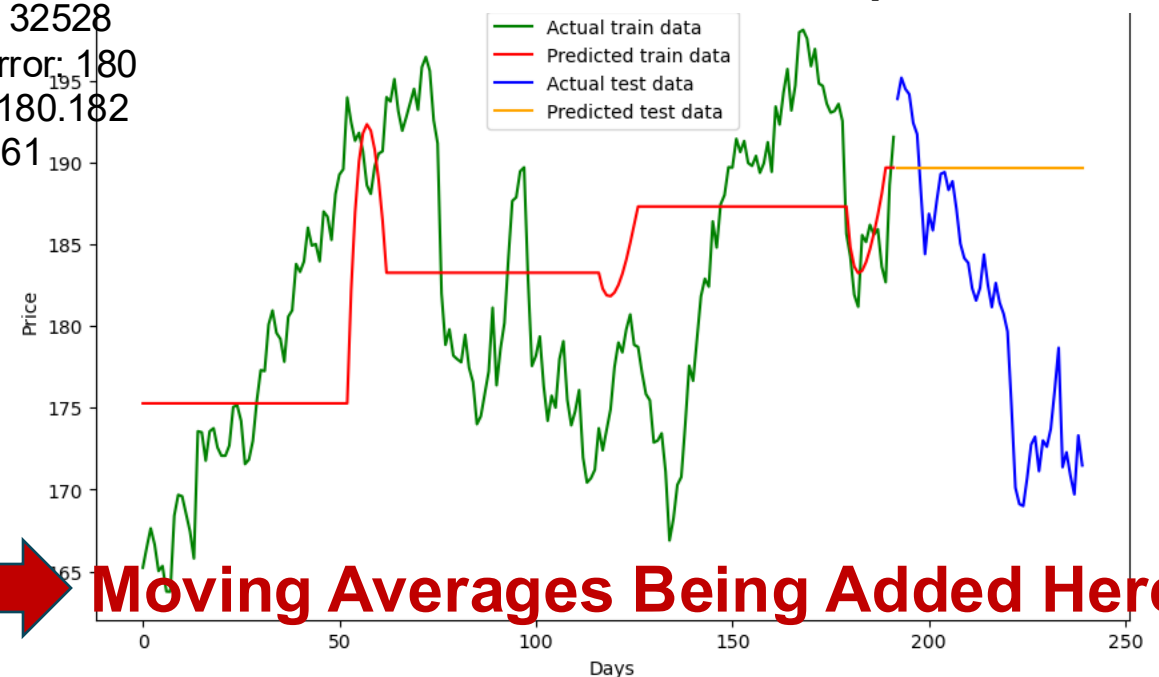
Long Short-Term Memory (LSTM) Models



SEASONAL IS A PROBLEM!

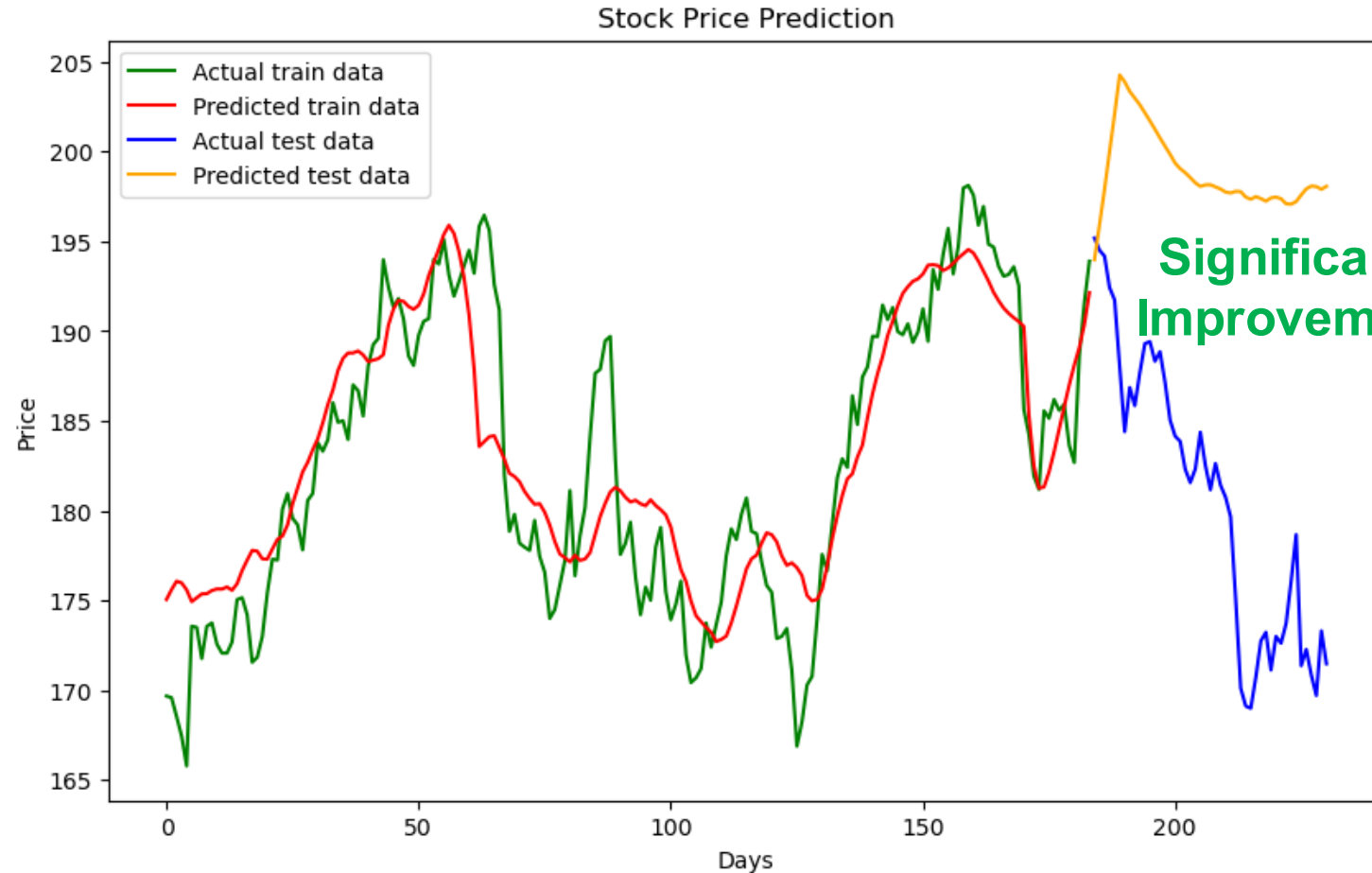


Prediction with Seasonal Financial Report Factors



Moving Averages Being Added Here

LSTM Time Series Model



LSTM Time Series Model enhanced by combining stock prices of AAPL, S&P 500, AAPL's financial report datasets.

Mean Squared Error: 1233

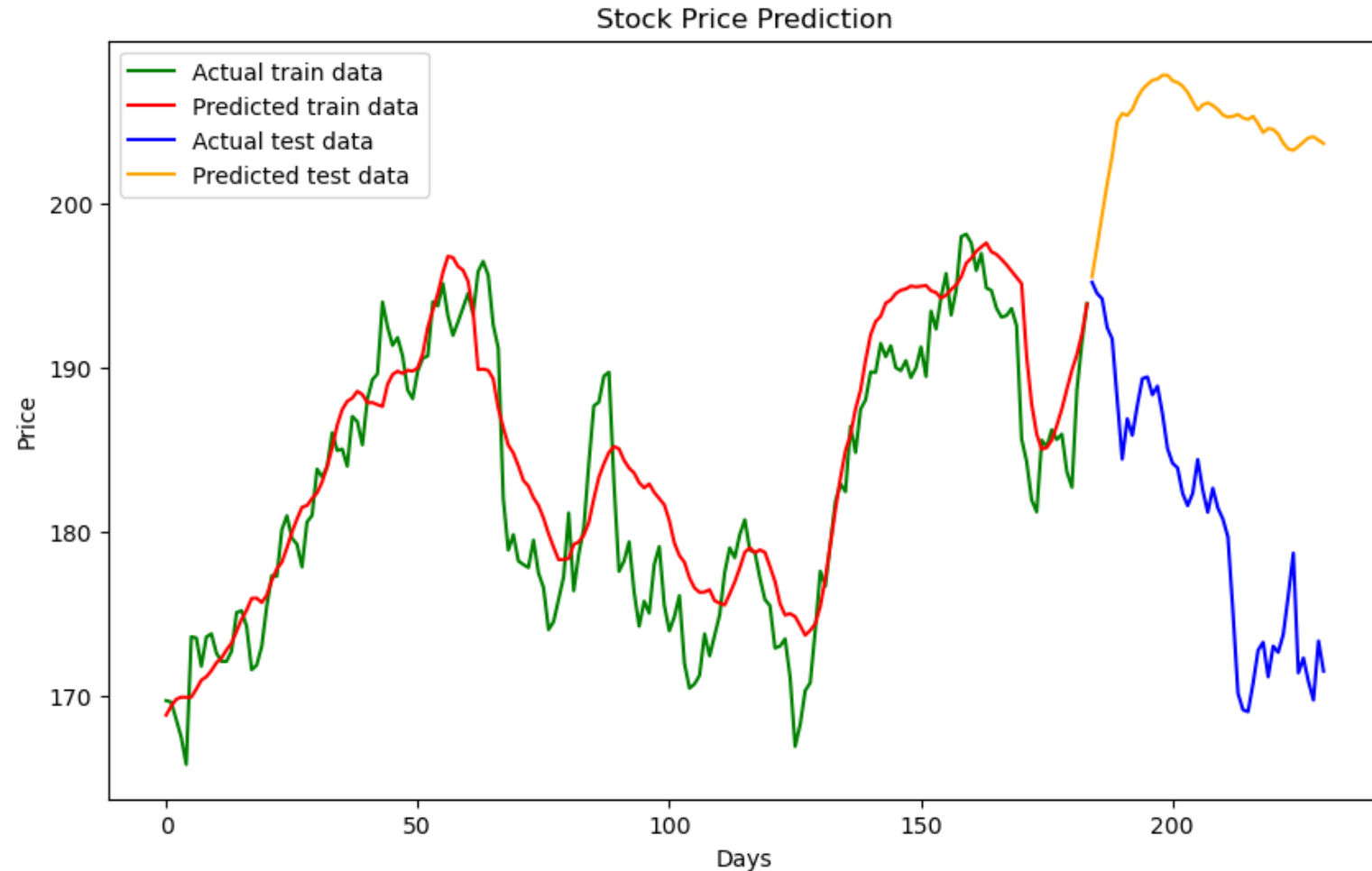
Root Mean Squared Error: 35

Mean Absolute Error: 33

R-squared: -20

- Significant improvements in prediction accuracy
- Capture broader market trends
- Include reporting cycles that influence stock prices

LSTM Time Series Model



LSTM Time Series Model enhanced by incorporating stock prices from MSFT, NVDA, BABA

Mean Squared Error: 479.72
Root Mean Squared Error: 21.90
Mean Absolute Error: 20.24
R-squared: -6.92

- noticeable improvement in prediction accuracy
- more market-related data provided
- capture the complex dynamics of the stock market better

Further Improvements

- **More Data Dimensions:** trading volume, macroeconomic indicators, industry news
- **Regularization:** to prevent overfitting, incorporating L1 or L2 regularization.
- **Longer Period Dataset:** analyzing longer historical data

Application

- **Risk Management:** Estimate potential price volatility and risk
- **Timing for Buying/Selling:** e.g. buying when a price increase is predicted and selling at predicted peaks.
- **Portfolio Diversification:** Analyze correlations between different stocks or assets with the model to build a diversified investment portfolio.
- **Long-term Investment Strategies:** Choose stocks with high long-term growth potential based on the model's predictions of long-term trends.

Thank You So Much.

Any Questions?