# Time Series Analysis and Modeling of FAANG Stock Prices

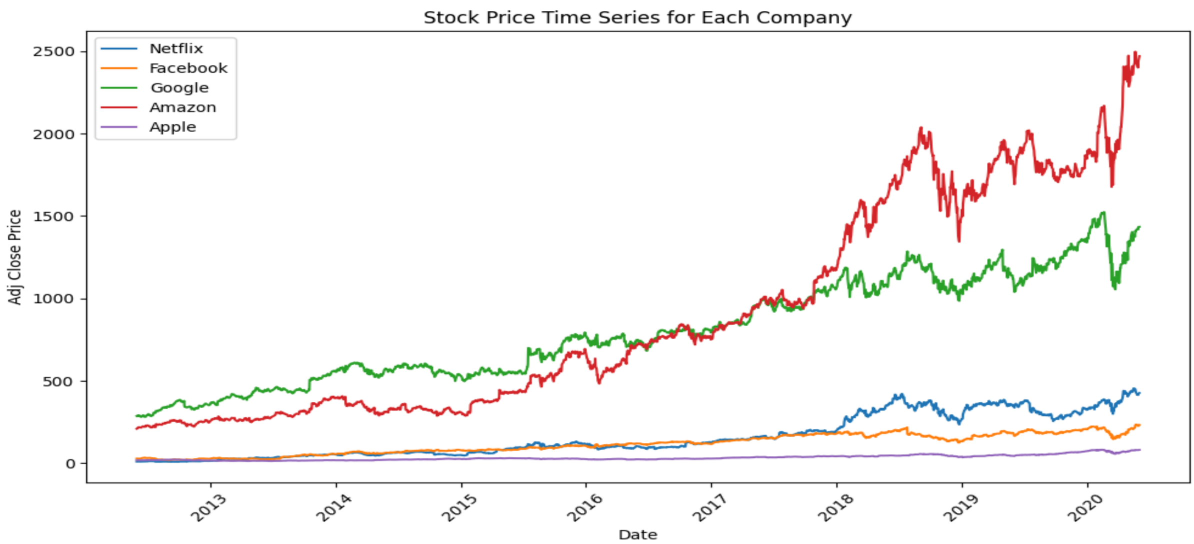
## Introduction

Stock price modelling plays a pivotal role in understanding financial market trends, analysing investment opportunities, and predicting future stock behaviour. In this project, we delve into the historical adjusted closing prices of five major technology companies—Facebook (now Meta), Amazon, Apple, Netflix, and Google (FAANG)—from June 2012 to June 2020.

Our analysis focuses on identifying patterns, trends, and interdependencies within these time series data. By leveraging statistical and machine learning techniques, we aim to:

- Examine the inherent trends, seasonality, and stationarity of each stock.
- Explore causal relationships between companies using Granger causality tests.
- Apply both univariate (ARIMA) and multivariate (Vector AutoRegression) time series models to predict stock prices.
- Compare the performance of these models to understand their strengths and limitations in capturing stock price dynamics.

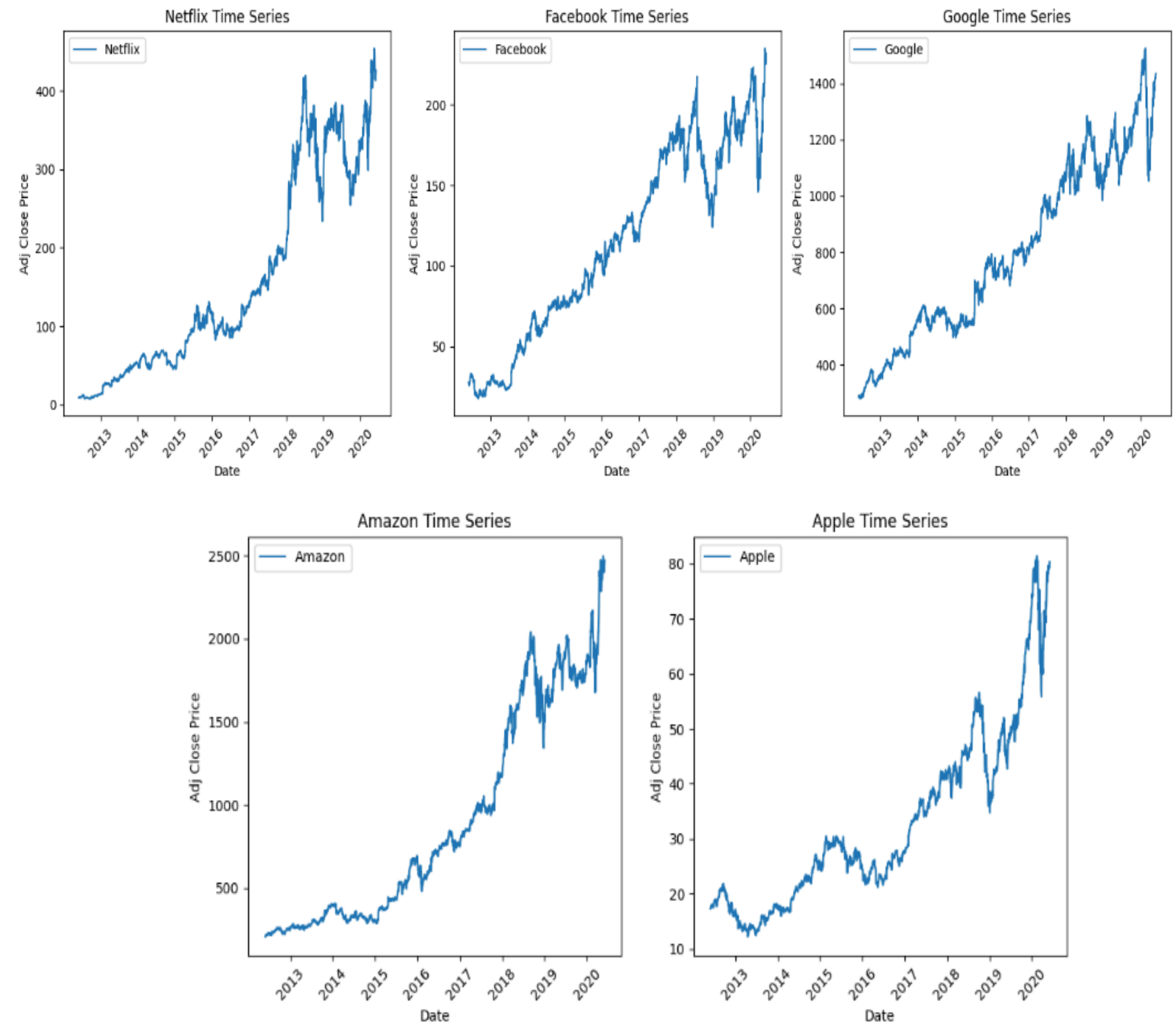
This study provides valuable insights into the interconnected behaviour of FAANG stocks and lays the foundation for advanced modelling techniques in financial analysis.



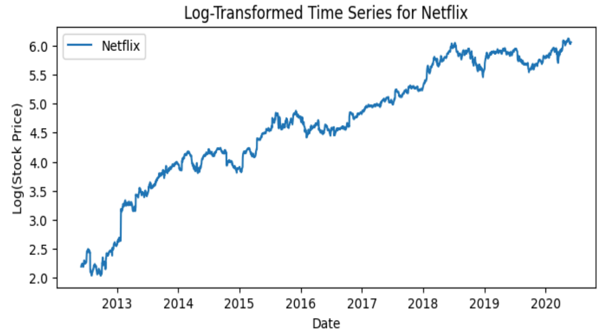
# Exponential Growth and Log Transformation

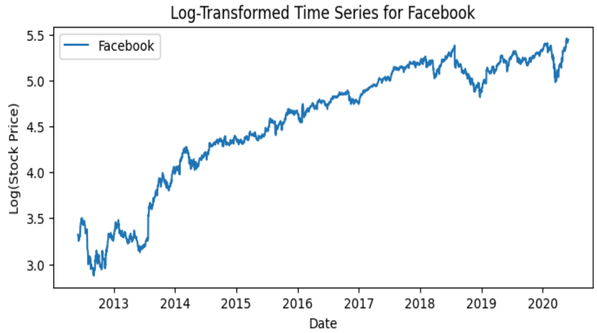
Initially, we observed that the stock price data exhibited exponential growth, making direct modelling ineffective due to the non-linear scale.
To stabilise the variance, we applied a logarithmic transformation to the data.

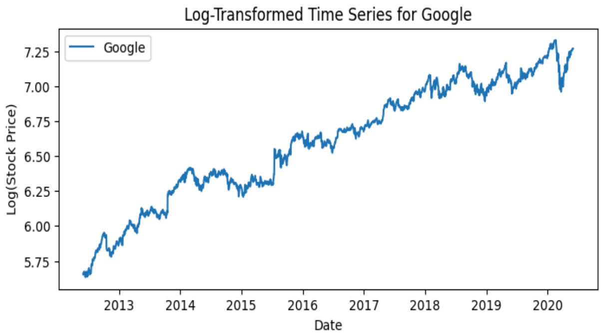
Visualisation before log transformation:

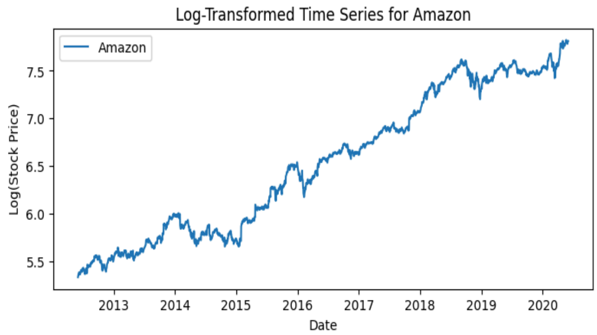


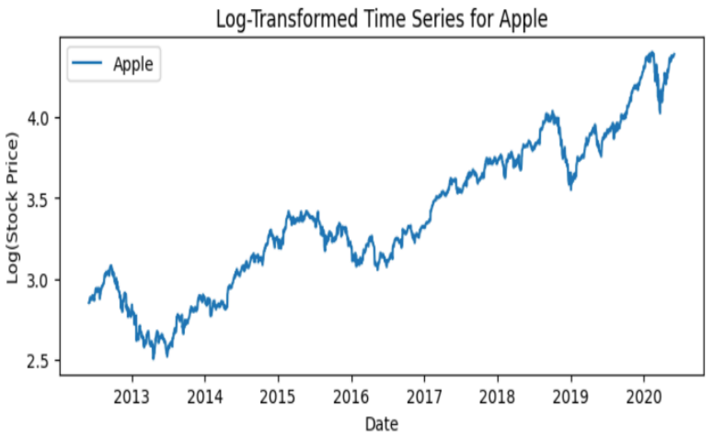
Visualisation after log transformation:



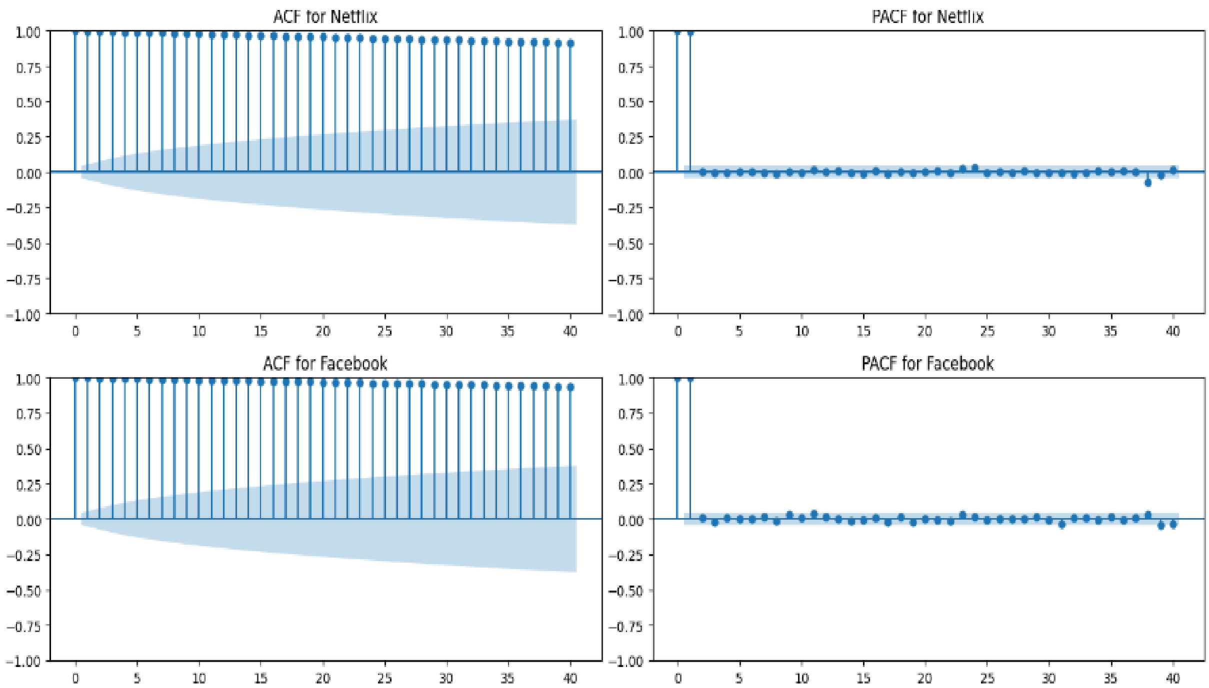
# Non-Stationarity and Differencing

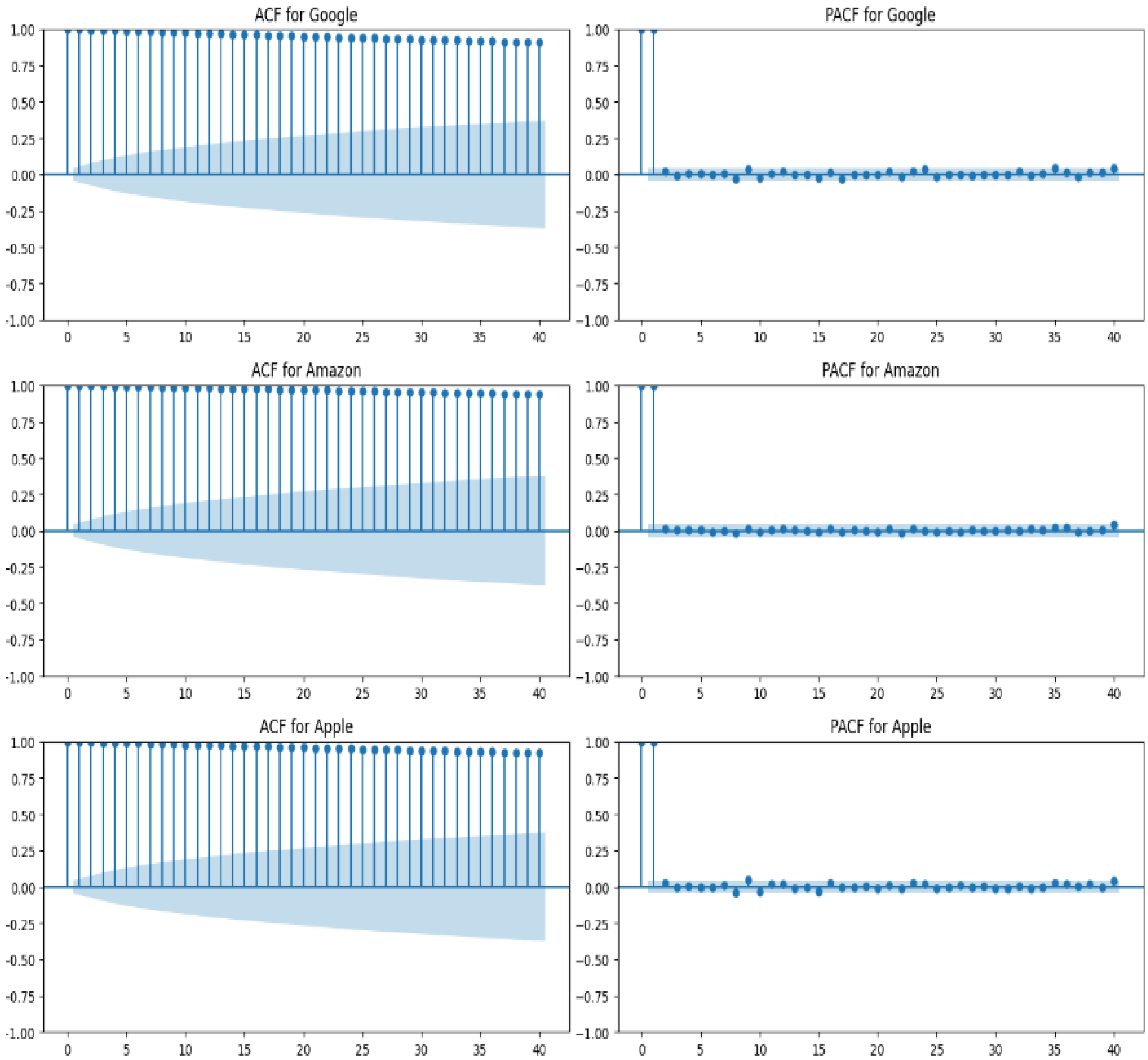
The augmented Dickey-Fuller (ADF) test confirmed that the data was non-stationary. Since stationarity is essential for time series modelling, we applied differencing to remove trends and make the series stationary.

ADF test results on the data after log transformation:

```
Netflix: ADF Statistic = -1.9996, p-value = 0.2868
Netflix is not stationary (p >= 0.05)

Facebook: ADF Statistic = -1.1448, p-value = 0.6969
Facebook is not stationary (p >= 0.05)

Google: ADF Statistic = -1.7340, p-value = 0.4137
Google is not stationary (p >= 0.05)

Amazon: ADF Statistic = -0.3271, p-value = 0.9216
Amazon is not stationary (p >= 0.05)

Apple: ADF Statistic = 0.1518, p-value = 0.9694
Apple is not stationary (p >= 0.05)
```

ACF and PACF plots before differencing:

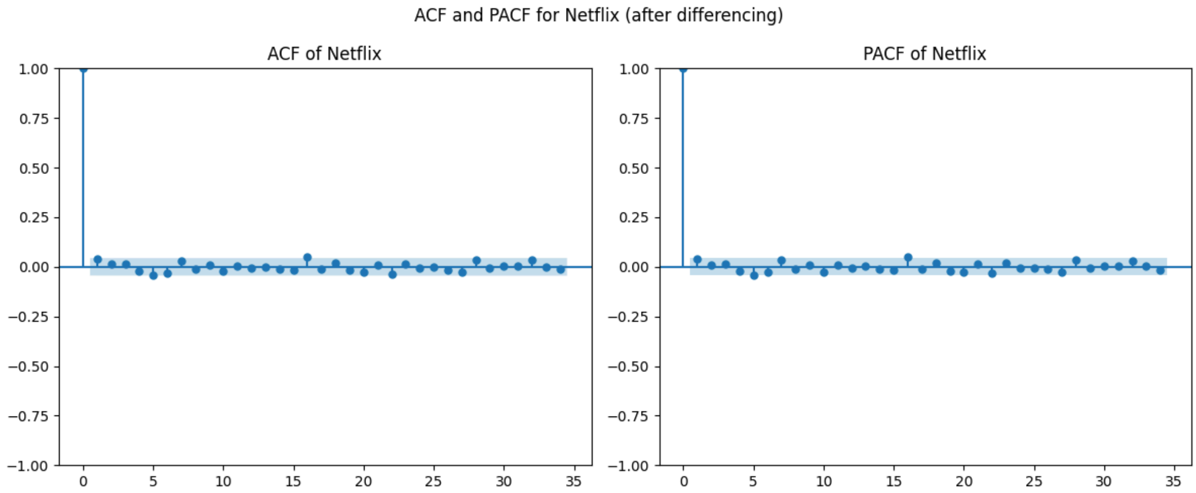


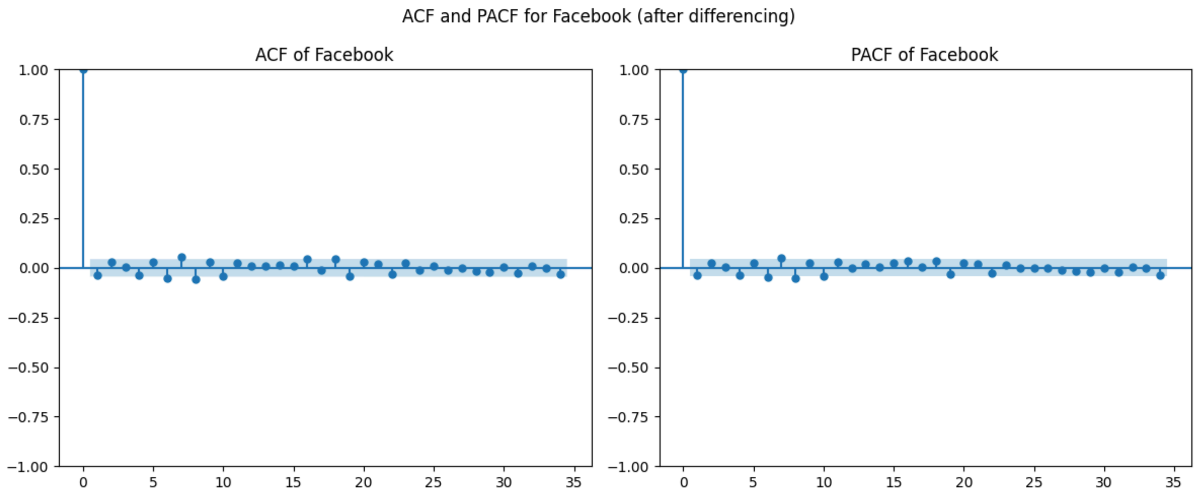


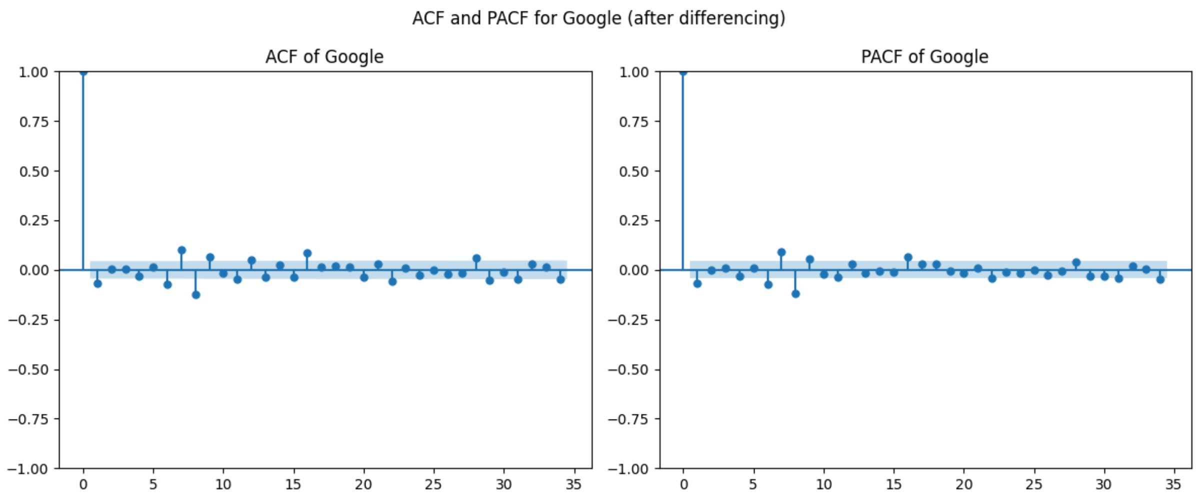

ADF test results on the data after differencing once:

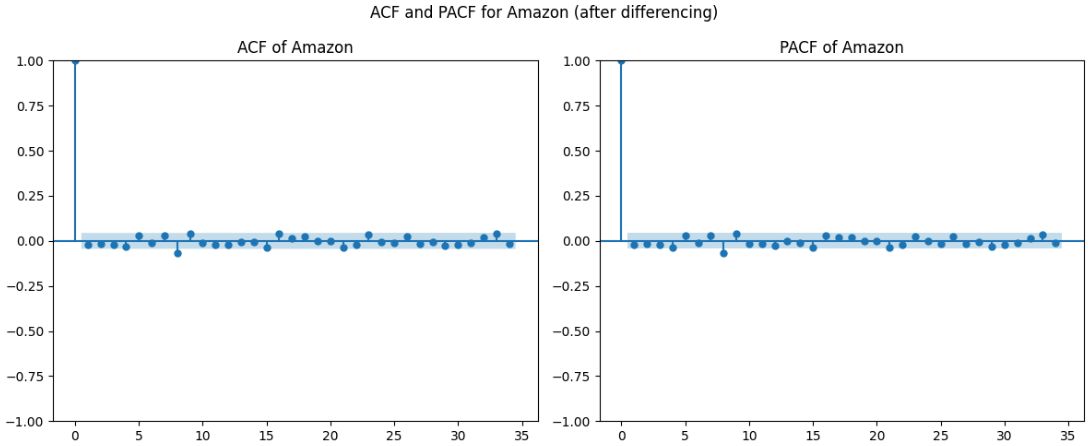
```
Netflix: ADF Statistic = -43.1191, p-value = 0.0000
Netflix is stationary (p < 0.05)

Facebook: ADF Statistic = -14.8660, p-value = 0.0000
Facebook is stationary (p < 0.05)

Google: ADF Statistic = -11.1062, p-value = 0.0000
Google is stationary (p < 0.05)

Amazon: ADF Statistic = -15.0453, p-value = 0.0000
Amazon is stationary (p < 0.05)
```

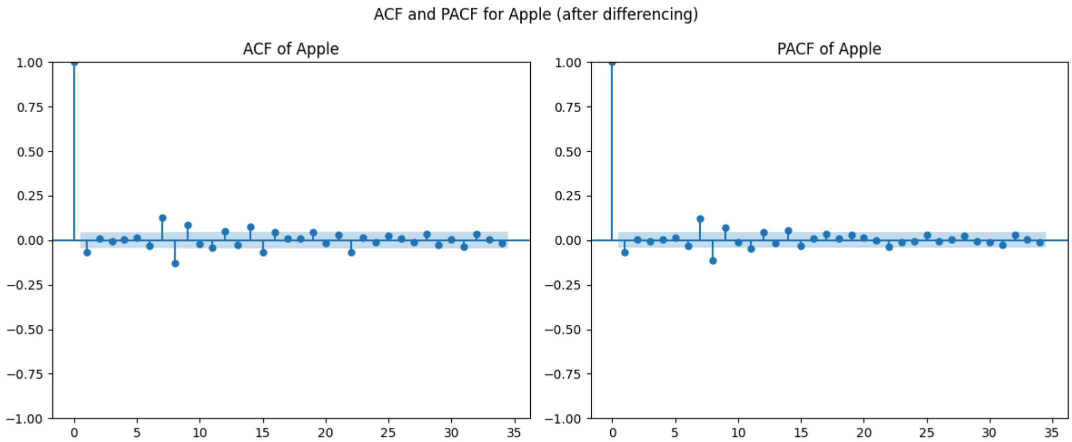
ACF and PACF plots after differencing once:



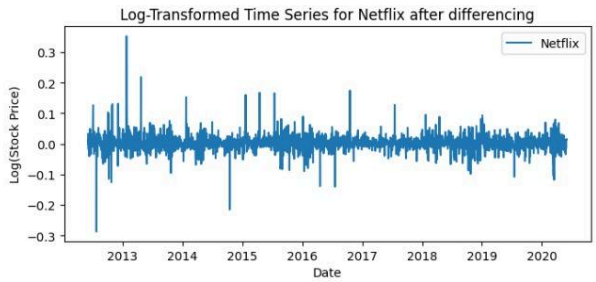



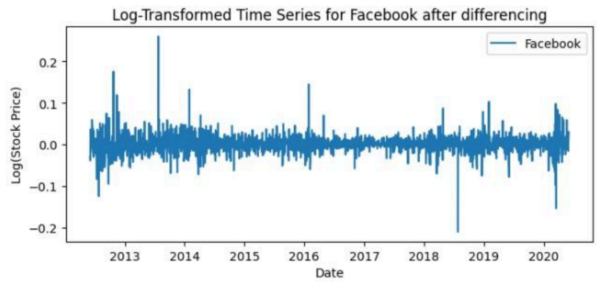



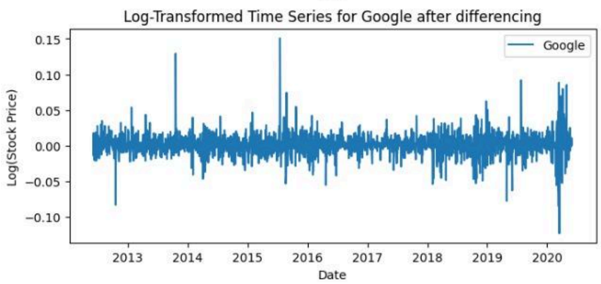


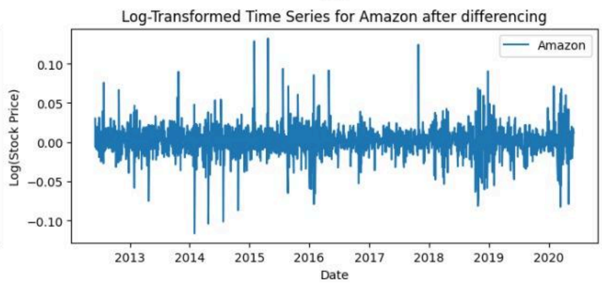



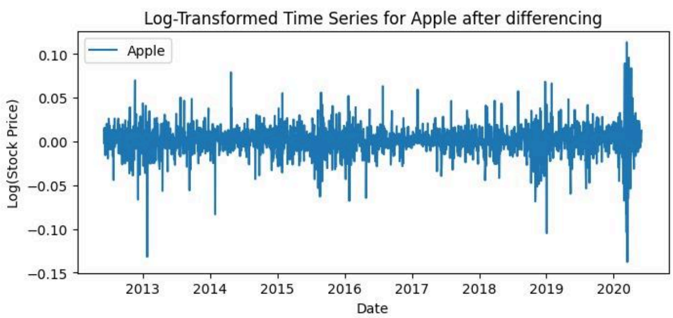

# Visualisation after differencing:



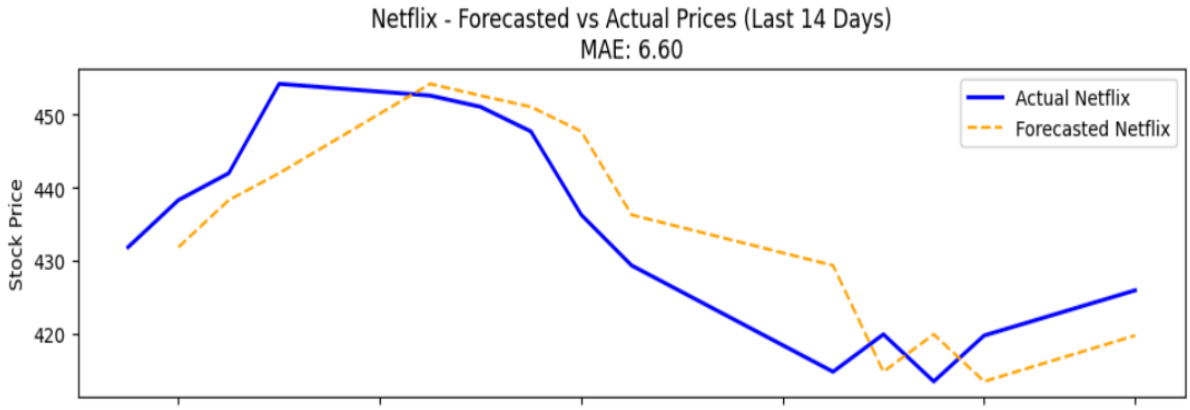



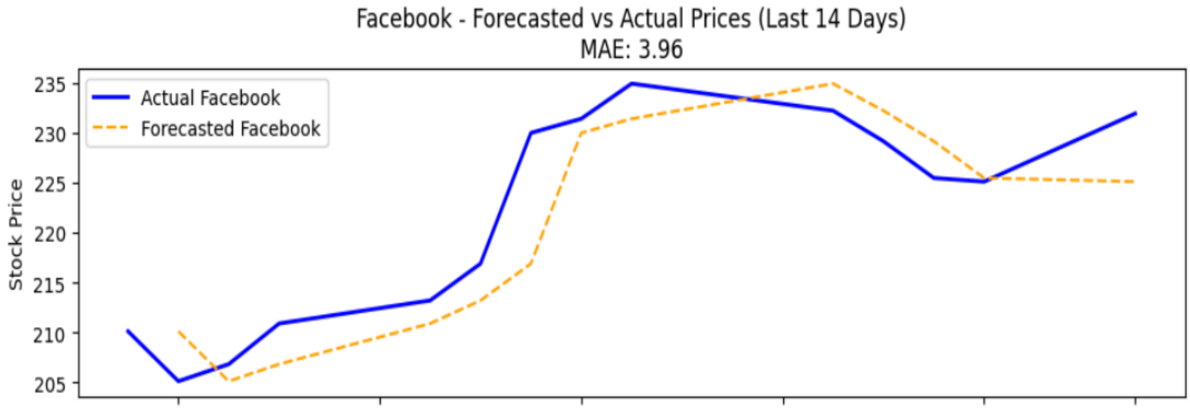
# Fitting Appropriate Models to Each Company:

## Random Walk

Following the initial analysis and differencing of the data, we observed distinct cutoffs in both the ACF and PACF plots across each stock series. These cutoffs suggested that the differenced series exhibited minimal autocorrelation beyond lag 1, implying that past values provided limited predictive power for future values beyond a single step. Given this structure, the random walk model was selected as a simple and robust baseline approach.

The Random Walk model's predictions were directly based on the last observed value before the test period for each stock. The Mean Absolute Error (MAE) for each stock series is presented below, alongside actual vs. forecasted values.




# SARIMA

Our initial analysis indicated the presence of seasonality in the dataset, as confirmed by the Kruskal-Wallis test. The PACF plot displayed a clear cutoff, suggesting the appropriateness of including seasonal parameters in the SARIMA model. Based on these observations, we conducted a grid search to optimise the SARIMA model parameters, incorporating seasonal elements.

```
Netflix: Kruskal-Wallis H-statistic = 28.4095, p-value = 0.0028
Significant seasonality detected for Netflix (p < 0.05).

Facebook: Kruskal-Wallis H-statistic = 16.9799, p-value = 0.1085
No significant seasonality detected for Facebook (p >= 0.05).

Google: Kruskal-Wallis H-statistic = 25.6691, p-value = 0.0073
Significant seasonality detected for Google (p < 0.05).

Amazon: Kruskal-Wallis H-statistic = 19.7492, p-value = 0.0489
Significant seasonality detected for Amazon (p < 0.05).

Apple: Kruskal-Wallis H-statistic = 16.9559, p-value = 0.1092
No significant seasonality detected for Apple (p >= 0.05).
```

SARIMA models capture the autoregressive (AR), differencing (I), and moving average (MA) components, as well as seasonal patterns if present. For each stock, we performed a grid search over parameters (p,d,q), with d fixed at 0 or 1. The model that minimised the Mean Absolute Error (MAE) during the test period was selected as the best for each stock series.

The parameter combinations considered were:

- **p (AR Order):** [0, 1, 2]
- **d (Differencing):** [0, 1]
- **q (MA Order):** [0, 1, 2]

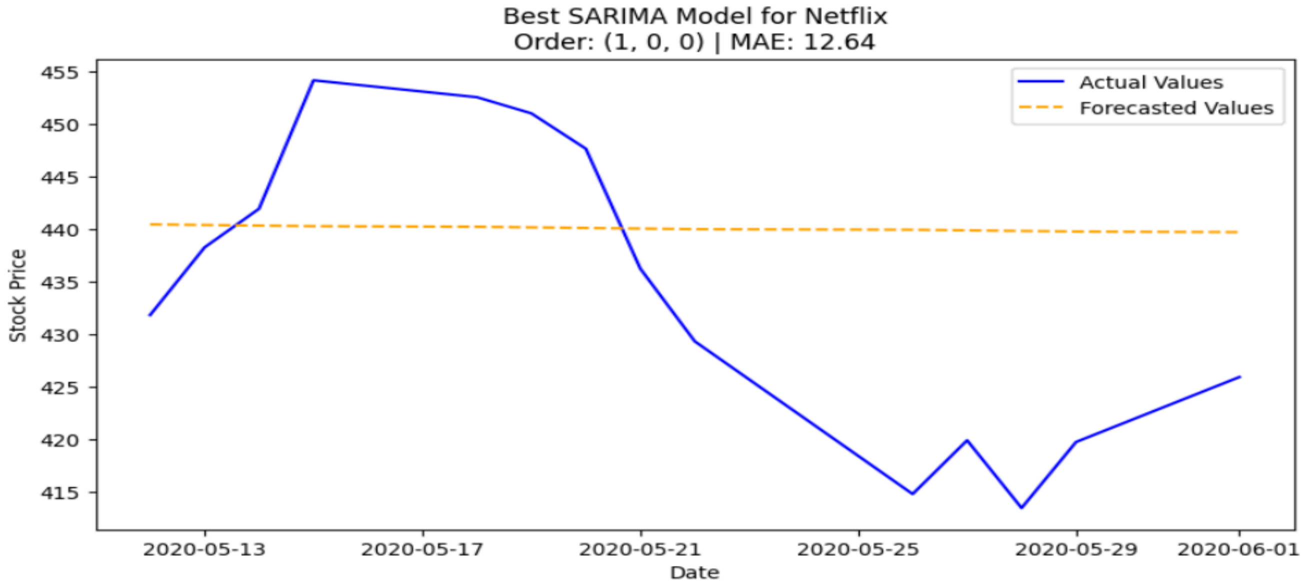
The Random Walk model assumes that future stock prices are best predicted by the most recent price, implying no mean reversion or predictable trend. This model is widely used in financial markets due to its simplicity and is expressed as:

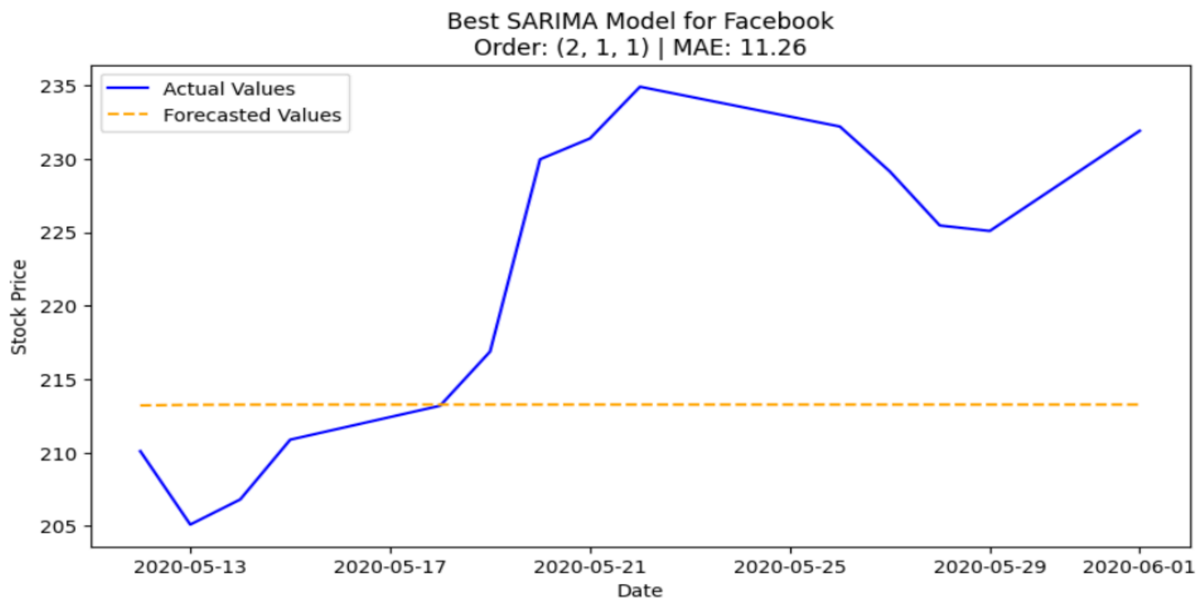
$$Y_t = Y_{t-1} + \varepsilon_t$$

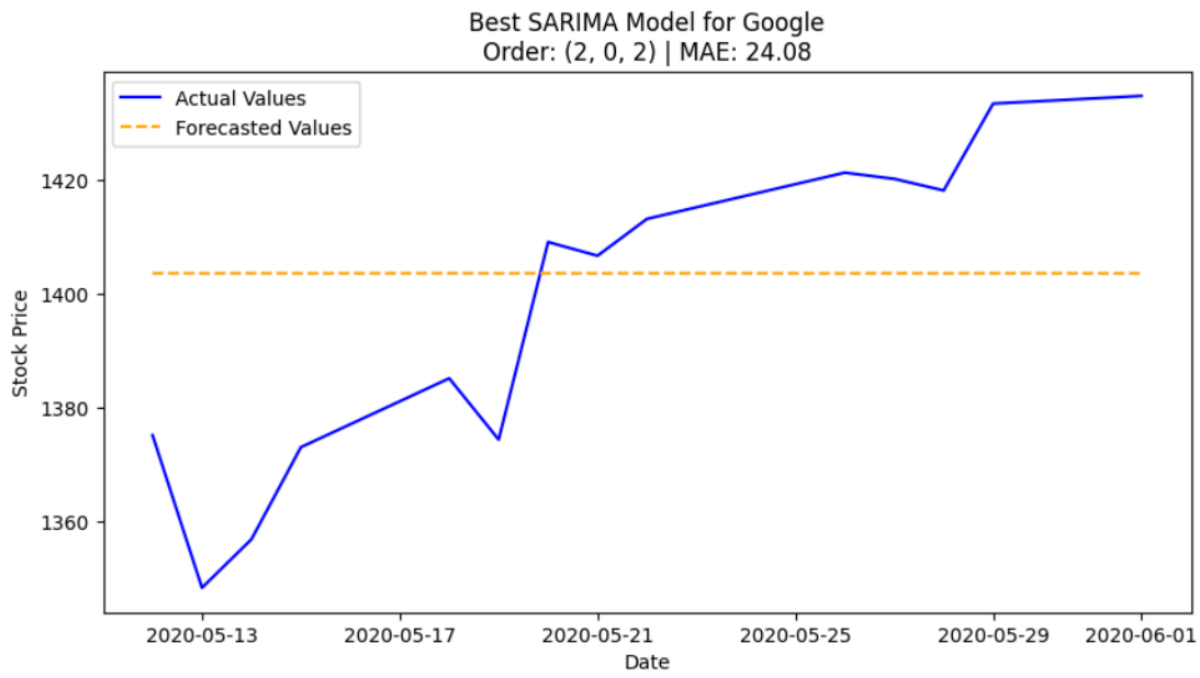
where $\varepsilon_t$ is a white noise error term.

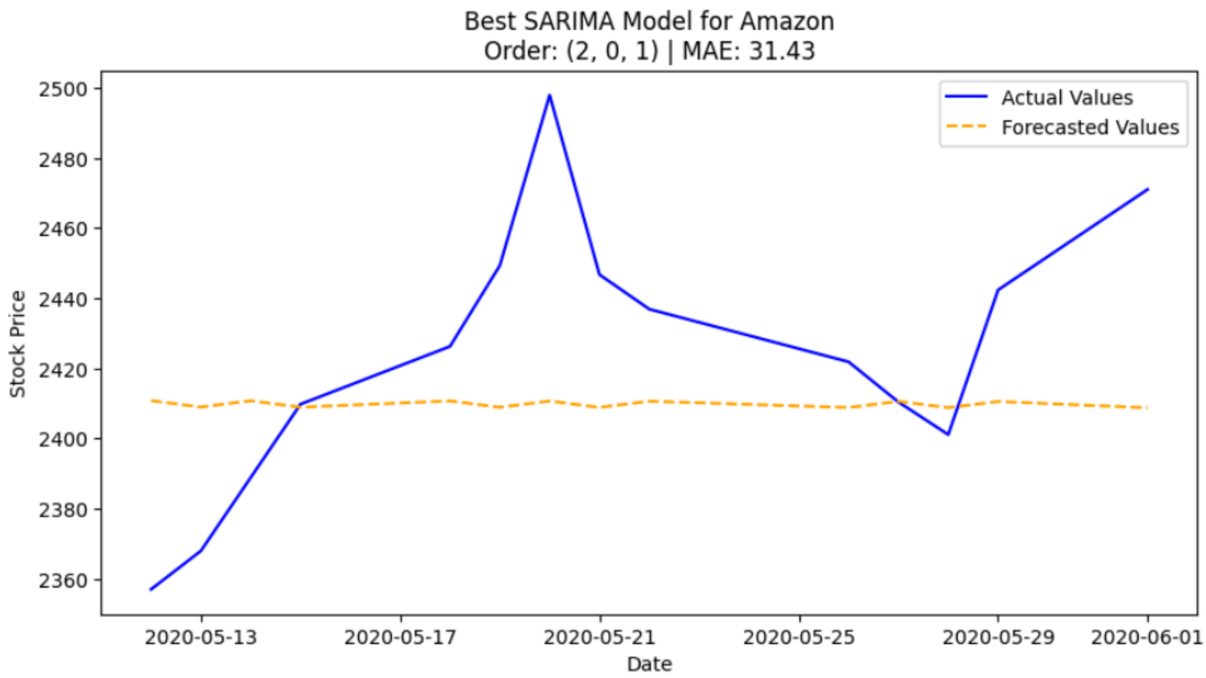
RESULTS:

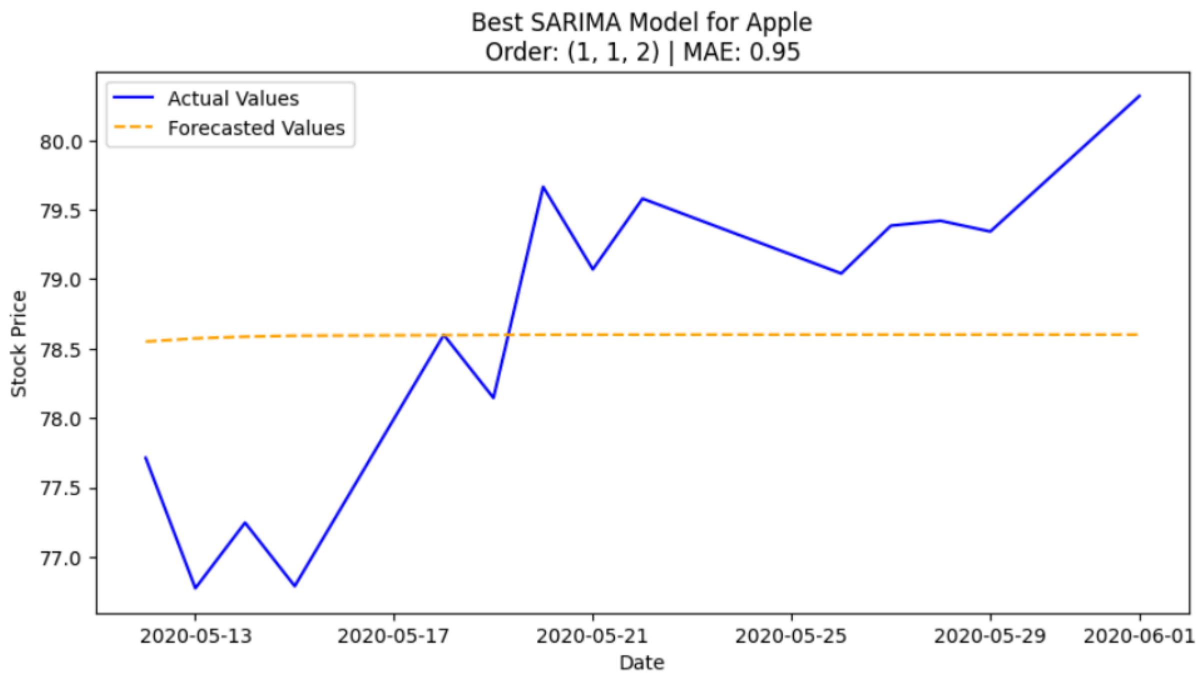
After fitting SARIMA models over a grid of parameters, the best model for each stock series was selected based on the lowest MAE. The Mean Absolute Error (MAE) for each stock series is presented below, alongside actual vs. forecasted values.


Best SARIMA Model for Netflix
Order: (1, 0, 0) | MAE: 12.64
Actual Values
Forecasted Values
Stock Price
455
450
445
440
435
430
425
420
415
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
Date


Best SARIMA Model for Facebook
Order: (2, 1, 1) | MAE: 11.26
Actual Values
Forecasted Values
Stock Price
235
230
225
220
215
210
205
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
Date


Best SARIMA Model for Google
Order: (2, 0, 2) | MAE: 24.08
Actual Values
Forecasted Values
Stock Price
1360
1380
1400
1420
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
Date

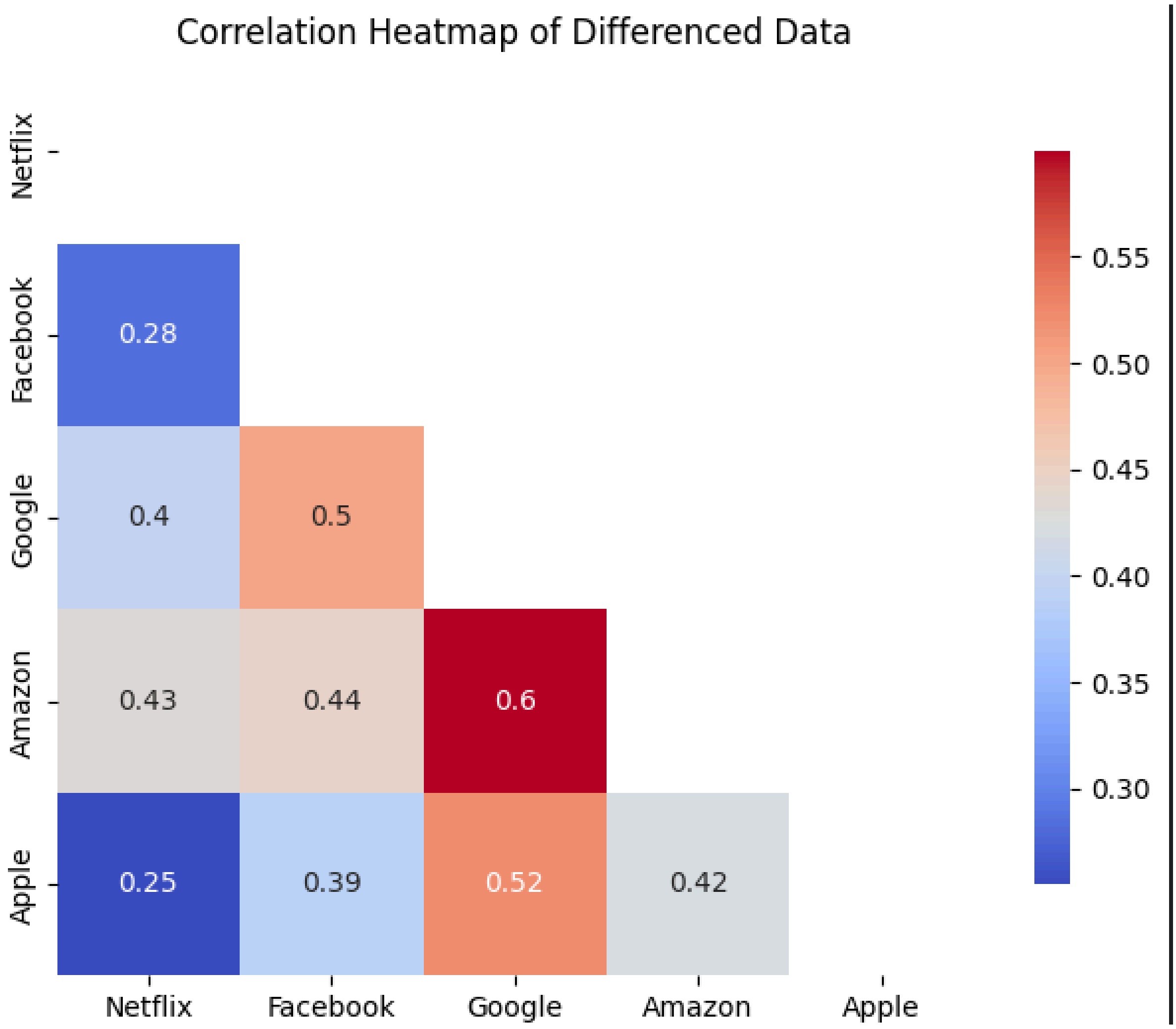

Best SARIMA Model for Amazon
Order: (2, 0, 1) | MAE: 31.43
Actual Values
Forecasted Values
Stock Price
2360
2380
2400
2420
2440
2460
2480
2500
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
Date



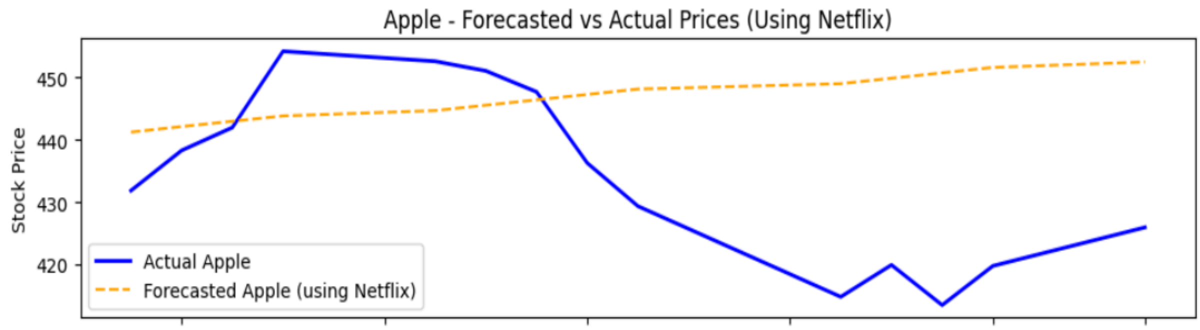

# Vector Time Series Modeling:

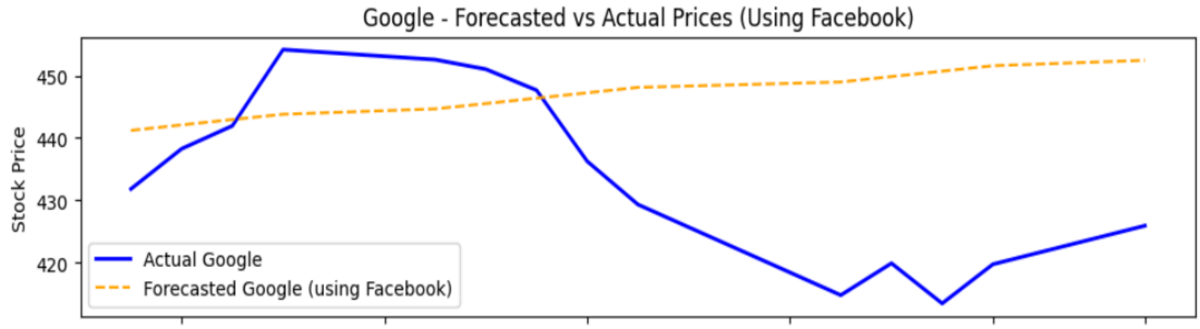
To investigate potential interactions among the different stock series, we conducted a Granger causality test, which evaluates whether past values of one time series help predict future values of another. The test revealed statistically significant causal relationships among several stock pairs, indicating that certain series influenced others over specific lags. These causative links suggested that a model capable of capturing these multivariate dependencies would be advantageous for more accurate forecasting. The VAR model was thus chosen as it is well-suited to capturing the dynamics between multiple time series that may influence each other.

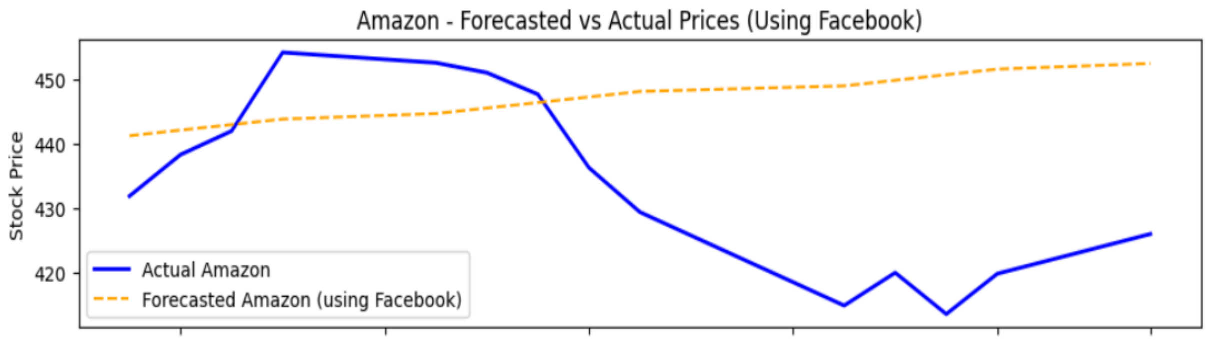
```
Significant causation pairs with p-value below 0.05:
Column Netflix causes Column Apple at lag 1 with p-value 0.0204
Column Facebook causes Column Google at lag 1 with p-value 0.0311
Column Facebook causes Column Amazon at lag 1 with p-value 0.0233
Column Facebook causes Column Apple at lag 1 with p-value 0.0009
Column Google causes Column Apple at lag 1 with p-value 0.0016
Column Amazon causes Column Facebook at lag 1 with p-value 0.0013
Column Amazon causes Column Google at lag 3 with p-value 0.0352
Column Amazon causes Column Apple at lag 1 with p-value 0.0040
Column Apple causes Column Netflix at lag 1 with p-value 0.0052
```

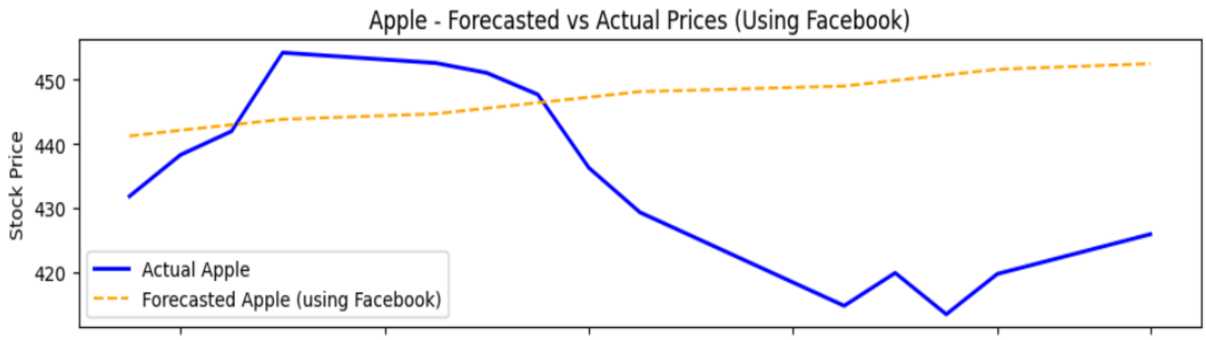


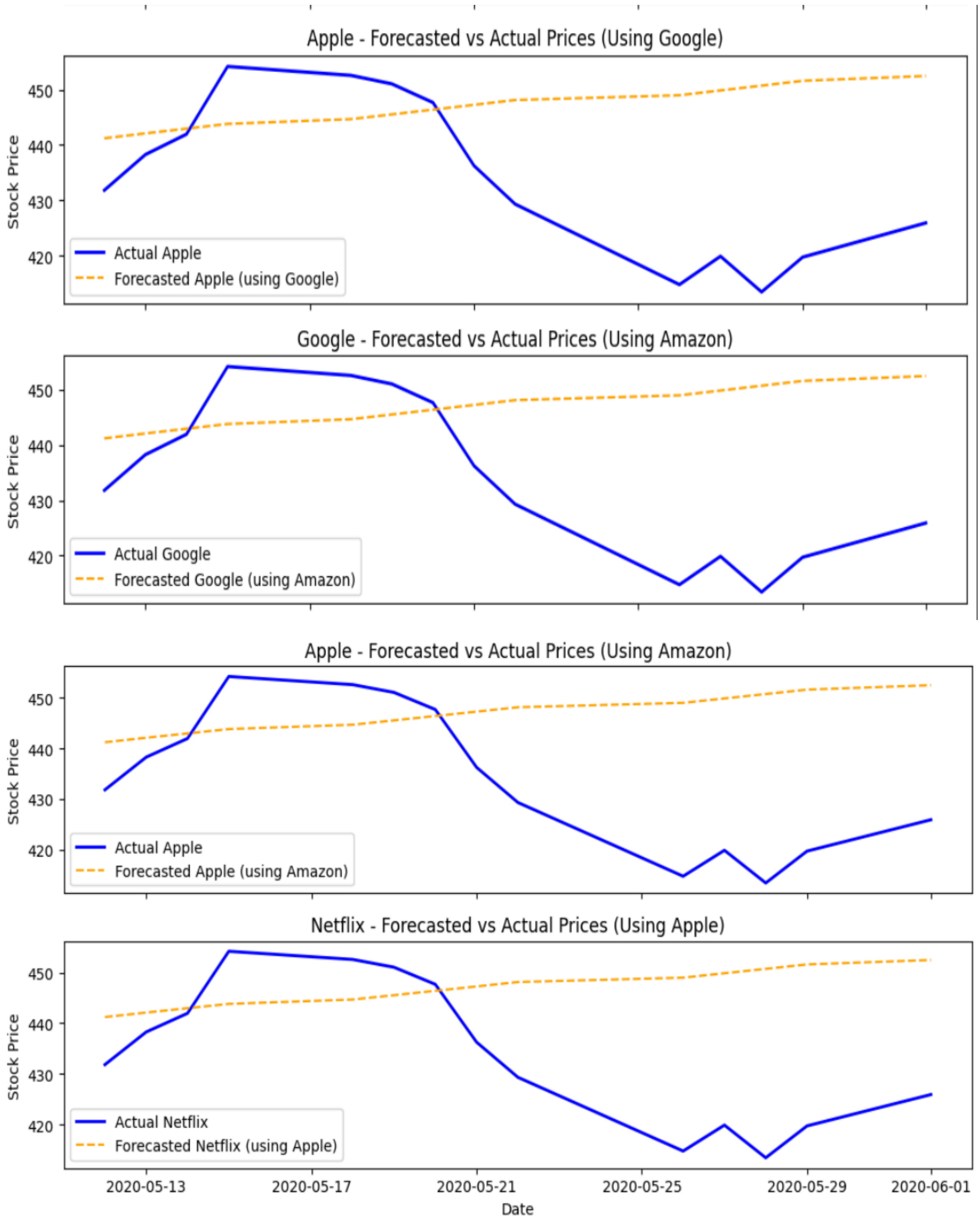

Next, we applied a VAR model to a set of five stock prices, each representing a different company. We aimed to capture the dependencies among these stock prices, assuming that the movement in one could potentially influence or be influenced by the others. For each pair of stocks identified as significantly causative in our Granger causality tests, we fitted a separate VAR model, using Mean Absolute Error (MAE) as the primary metric for evaluating forecast accuracy. After fitting VAR models to the stock pairs with significant Granger causation, we compared forecasted versus actual values for each pair over the 14-day test period.

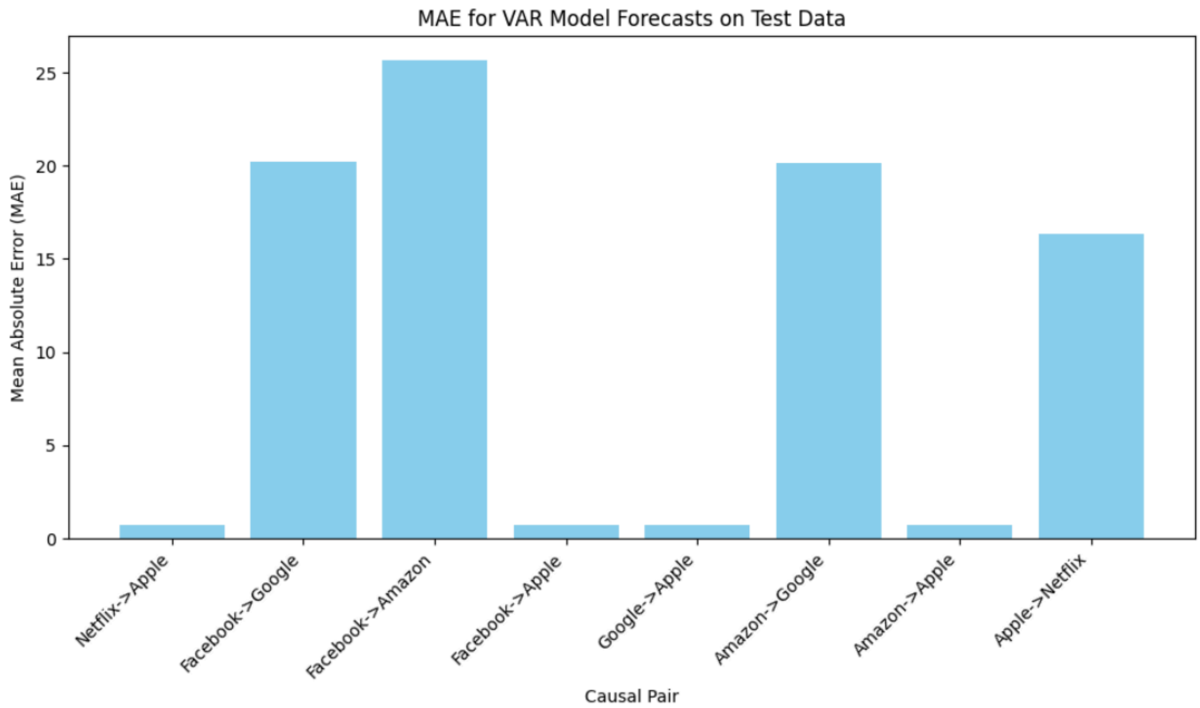



Apple - Forecasted vs Actual Prices (Using Google)
Stock Price
450
440
430
420
Actual Apple
Forecasted Apple (using Google)
Google - Forecasted vs Actual Prices (Using Amazon)
Stock Price
450
440
430
420
Actual Google
Forecasted Google (using Amazon)
Apple - Forecasted vs Actual Prices (Using Amazon)
Stock Price
450
440
430
420
Actual Apple
Forecasted Apple (using Amazon)
Netflix - Forecasted vs Actual Prices (Using Apple)
Stock Price
450
440
430
420
Actual Netflix
Forecasted Netflix (using Apple)
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
2020-06-01
Date




# Results

After analysing and comparing the results from the Random Walk model, SARIMA, and Vector Time Series (VAR) models, the following conclusions were drawn:

1. **Random Walk Model**:
   - The Random Walk model consistently achieved lower Mean Absolute Error (MAE) compared to SARIMA and VAR models.
   - This indicates that the Random Walk model, despite its simplicity, is highly effective in capturing the underlying dynamics of the FAANG stock prices.
   - The Random Walk model assumes stock prices follow a stochastic process with no predictable mean reversion, making it a robust choice in scenarios where prices are highly volatile and exhibit no clear trend or seasonality.
2. **SARIMA Models**:
   - While SARIMA models performed well in identifying and modeling autoregressive and seasonal components, they struggled to achieve the same level of forecast accuracy as the Random Walk model.
   - The reliance on parameter tuning and the assumption of stationarity may have limited their adaptability to the highly dynamic and stochastic nature of stock prices.
3. **Vector Time Series Models (VAR)**:

- VAR models provided insights into the interdependencies between stocks, particularly for pairs with significant Granger causality.
- However, the complexity of VAR models and their sensitivity to parameter tuning led to higher forecasting errors compared to the Random Walk model.
- VAR's focus on capturing relationships between variables may have overshadowed the individual price movements, which the Random Walk model handles effectively.

The findings of this project highlight that the simplicity of the Random Walk model often surpasses more complex models in stock price forecasting. While SARIMA and VAR offer valuable insights into patterns and relationships, they may not be as practical for prediction in volatile markets like stocks. Further exploration of hybrid models or machine learning-based approaches could potentially combine the strengths of these methods.