

## Forecasting the Closing Price of Amazon Stock

### Week 2: Data Visualisation

Github Repository: [AMZN](#)

The goals of this week's task are as follows:

- Understand the data
- Visualise the variables and extract the trends
- Make possible distribution plots for different periods
- Perform trend decomposition and make your assumptions about seasonality existence in the data

#### Data Description

The Stock Price for Amazon (AMZN) for 3019 business days has been given in the provided dataset, starting from 3rd January 2006 to 29th December 2017. The dataset includes the following variables:

- ❖ Date : The date of the trading period.
- ❖ Open Price: The price of the stock at the beginning of the trading period.
- ❖ High Price: The highest price reached during the trading period.
- ❖ Low Price: The lowest price reached during the trading period.
- ❖ Close Price: The price of the stock at the end of the trading period.
- ❖ Volume: The total number of shares traded during the trading period.

There are no missing values in the dataset, and it is our view that all the data is required.

#### Data Visualisation and Trend Extraction

It is evident from Figure 1, that there is an obvious upward trend. The upward trend observed in the Opening, Closing, Highest, and Lowest Prices of Amazon stock suggests positive momentum and overall growth in its value over the given time period. Identifying such trends in time series data is crucial for choosing the right modelling approach enabling accurate forecasting and trend analysis.



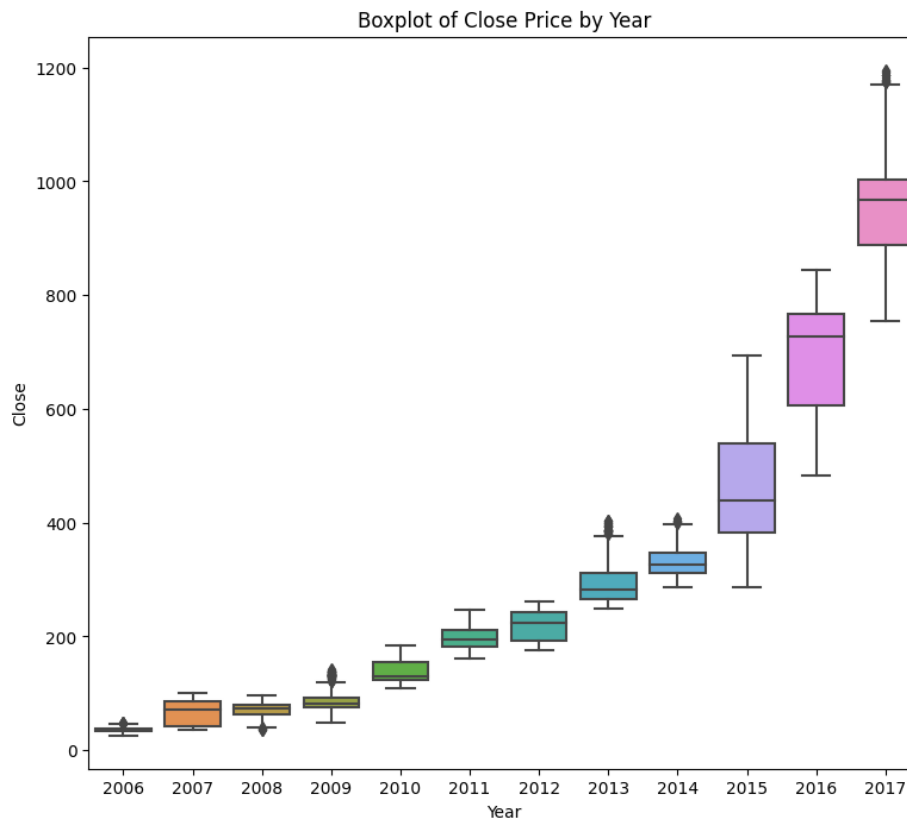
**Figure 1:** Stock Price Analysis of Amazon (AMZN): Opening, Closing, Highest, and Lowest Prices, along with Trading Volume Over Time

### Distribution Plots for Different Periods

Given that the Opening, Closing, Highest, and Lowest Prices are identical, we will consider one of them in this case.

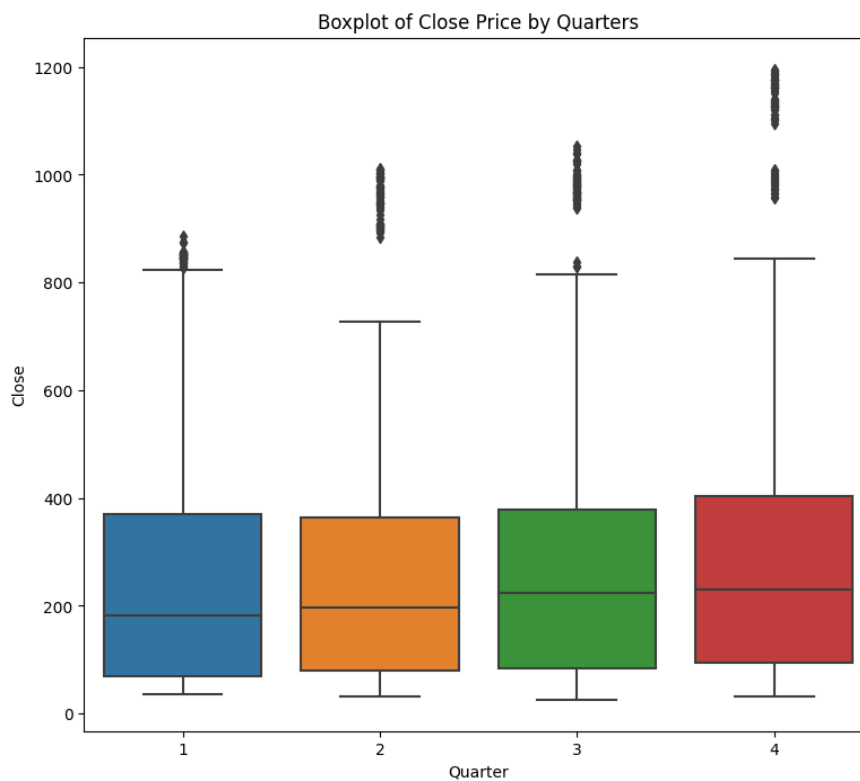
#### Yearly Distribution Plots

Figure 2 suggests a consistent upward trend in the central tendency of the data over the period between 2006 and 2017. For most of the boxplots, the distribution is less skewed towards lower values, implying a tendency for higher values. The increasing median/mean in the yearly boxplot aligns with the previously mentioned upward trend in Close, Open, High, and Low prices. Both trends suggest a positive trajectory for the stock, reinforcing the notion of overall growth in its value.



**Figure 2:** *Yearly Distribution of Close Prices*

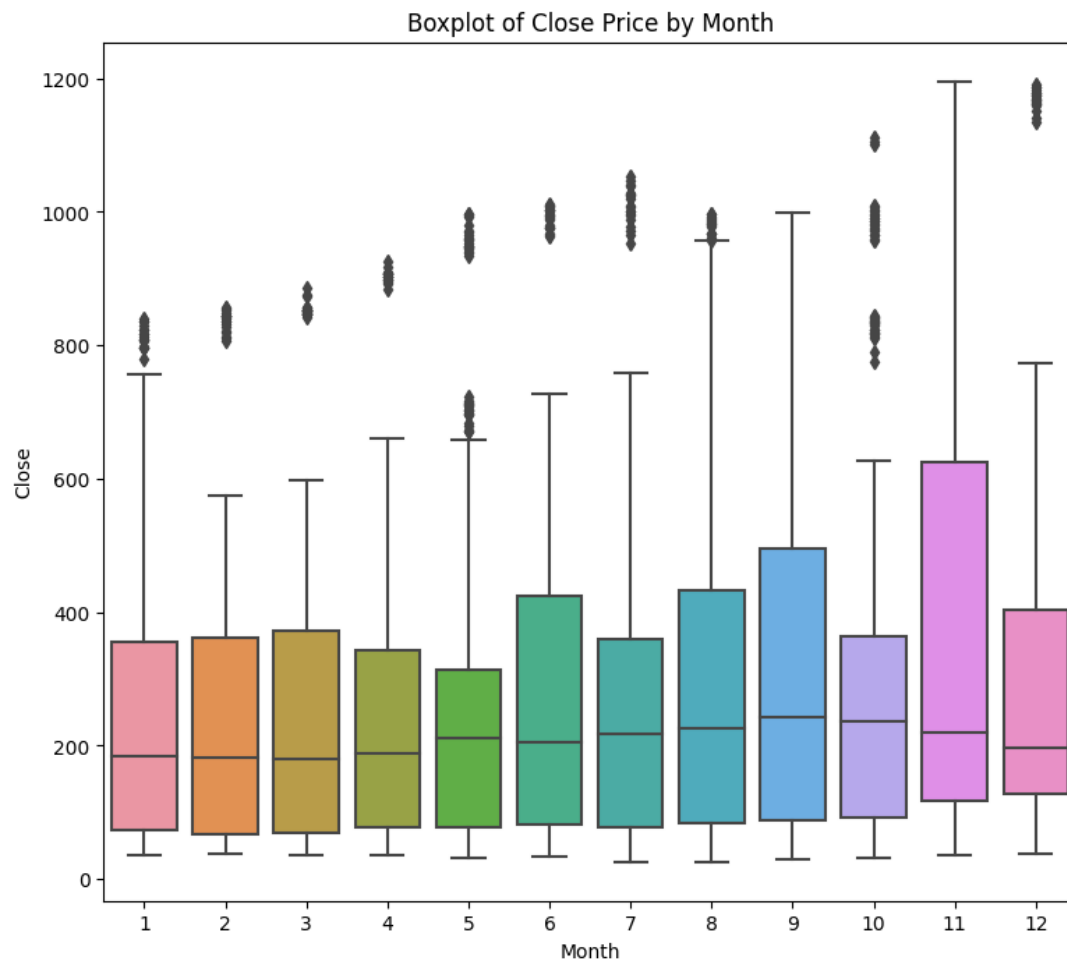
### Quarterly Distribution Plots



**Figure 3:** *Quarterly Distribution of Close Prices*

The rising trend is clearly observable in the quarterly distribution plots presented in Figure 3. This observation aligns with common intuition, as the overall upward trend dictates that the values in the later quarters are expected to be higher compared to the earlier quarters. In simpler terms, as time progresses, the data shows an increasing pattern, and this is reflected in each quarter's distribution, where later quarters tend to have higher values due to the overall upward trajectory.

### Monthly Distribution Plots

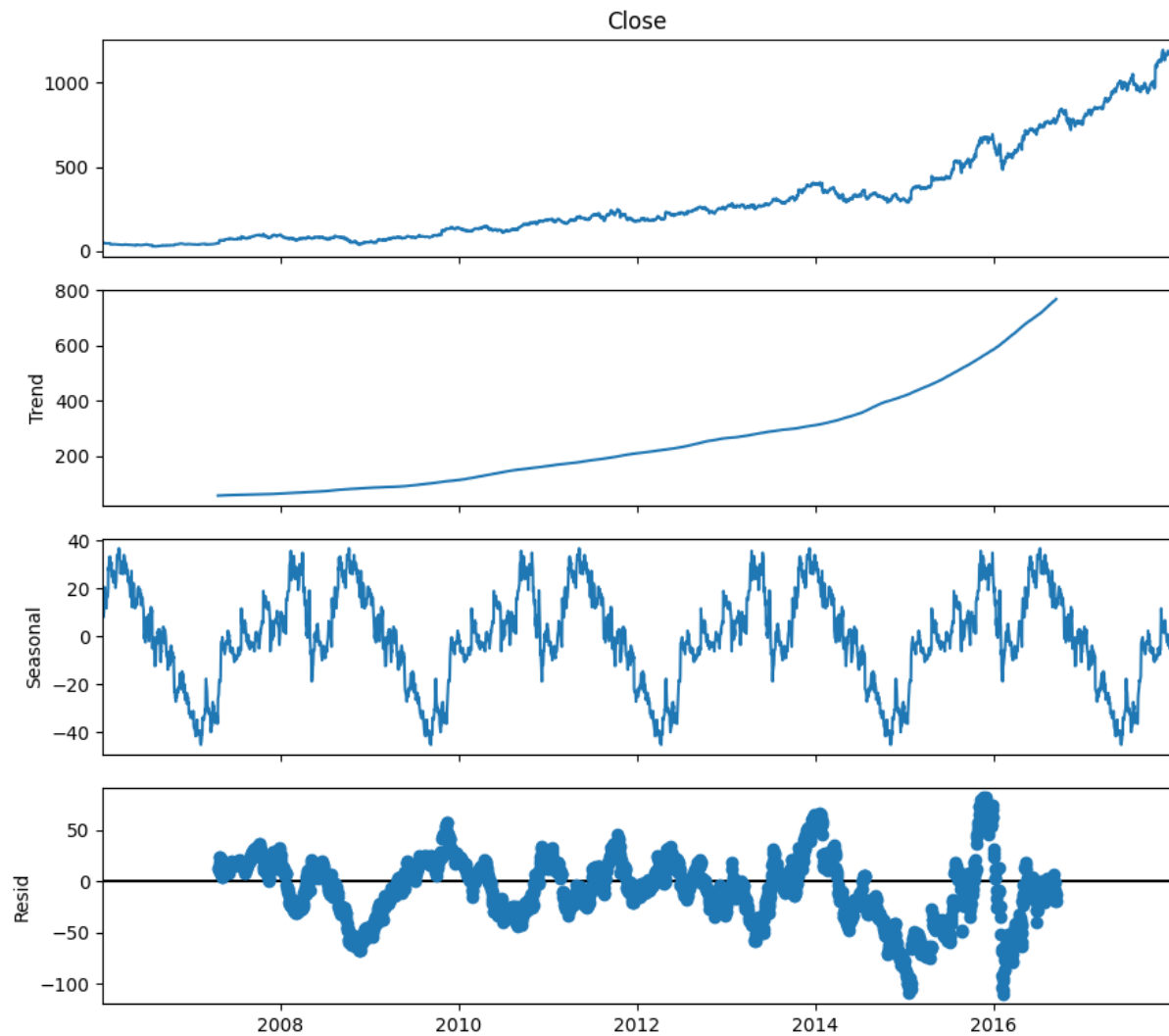


**Figure 4:** *Monthly Distribution of Close Prices*

Figure 4 shows a much varied picture with no clear monotonic trend. Regarding, based on the central tendency line, it is evident that overall the months of December, January, February, March and April have relatively lower closing prices than the rest of the months of the year. Moreover, disregarding the outliers, these months have on average less variance than their counterparts.

### Trend Decomposition

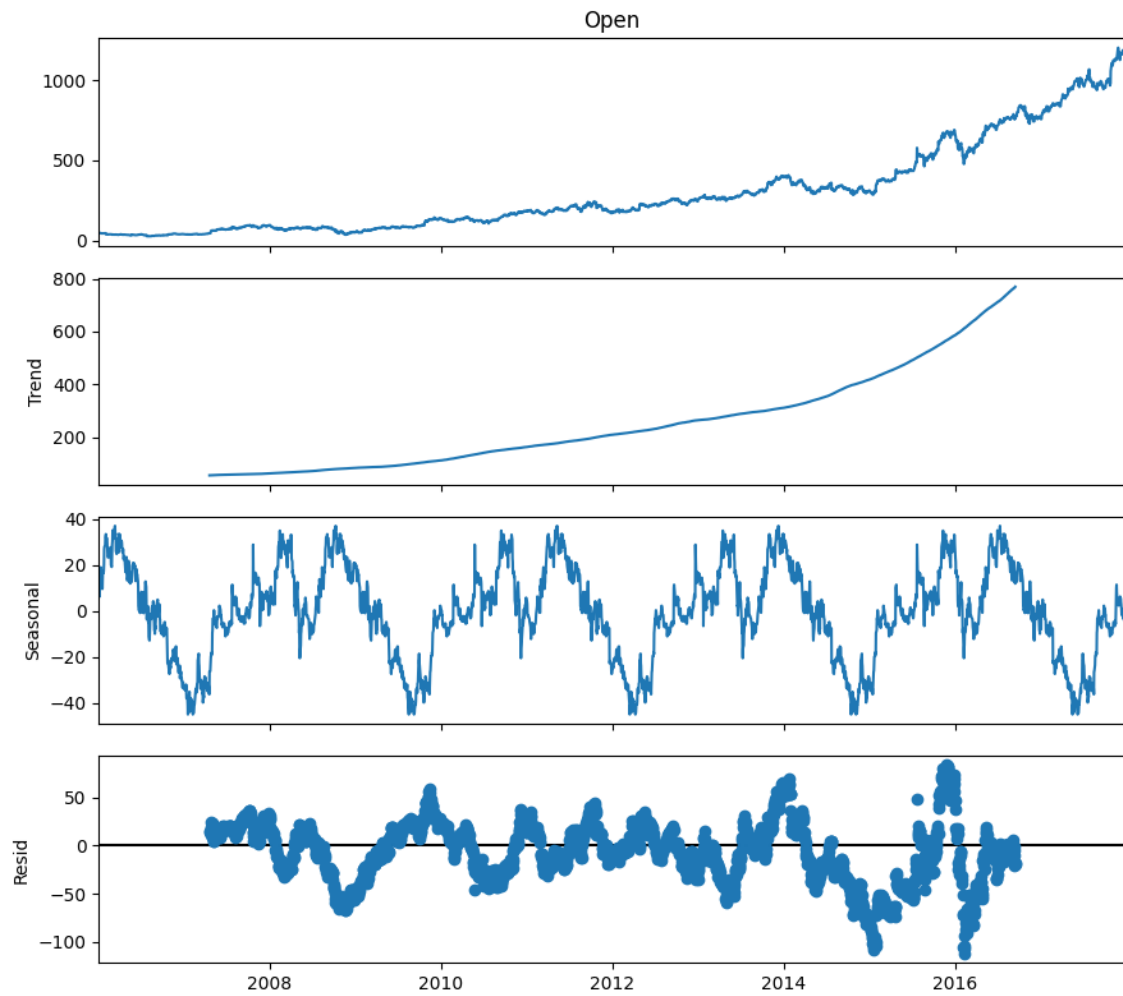
After experimenting with different seasonality values, we settled on cycles of 2.5 years. The trend decomposition of the closing price with 2.5 years cycles is represented in Figure 5.



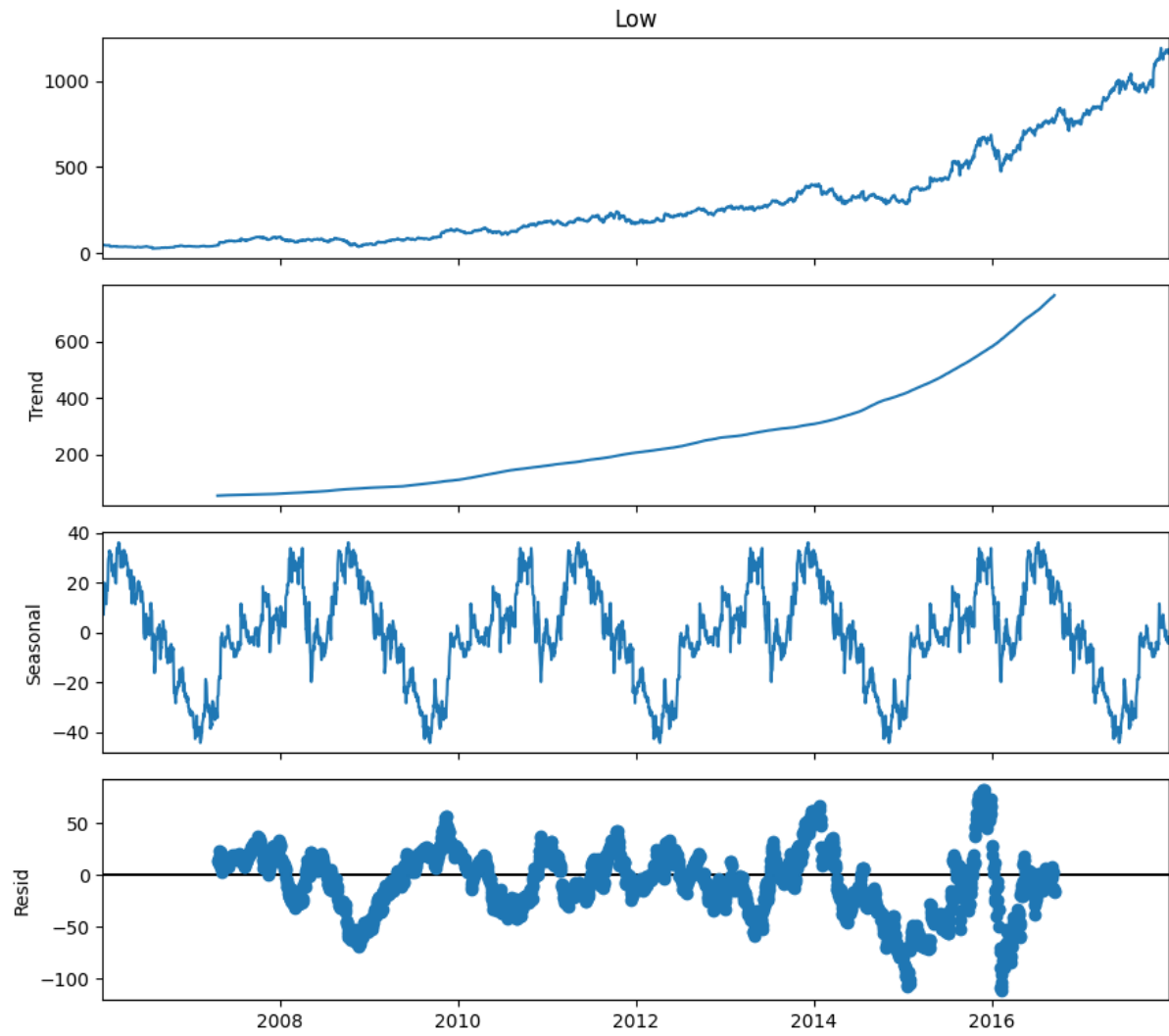
**Figure 5:** Time Series Decomposition of Closing Price Using Additive Model: Seasonal, Trend, and Residual Components

Examining the time series decomposition depicted in Figure 5, a roughly two-and-a-half-year seasonality pattern becomes evident, indicating an extended cyclical trend or multi-year influence. This longer cycle is likely shaped by larger-scale economic or environmental factors, contributing to the observed multi-year trends within the data. For the trend, an upward trend is evident for the period of observation and is in line with our earlier observations. However, looking at the residuals, it is obvious that there are still some patterns, which implies that there might still be some non-random structure in the data that was not captured by the trend and seasonality.

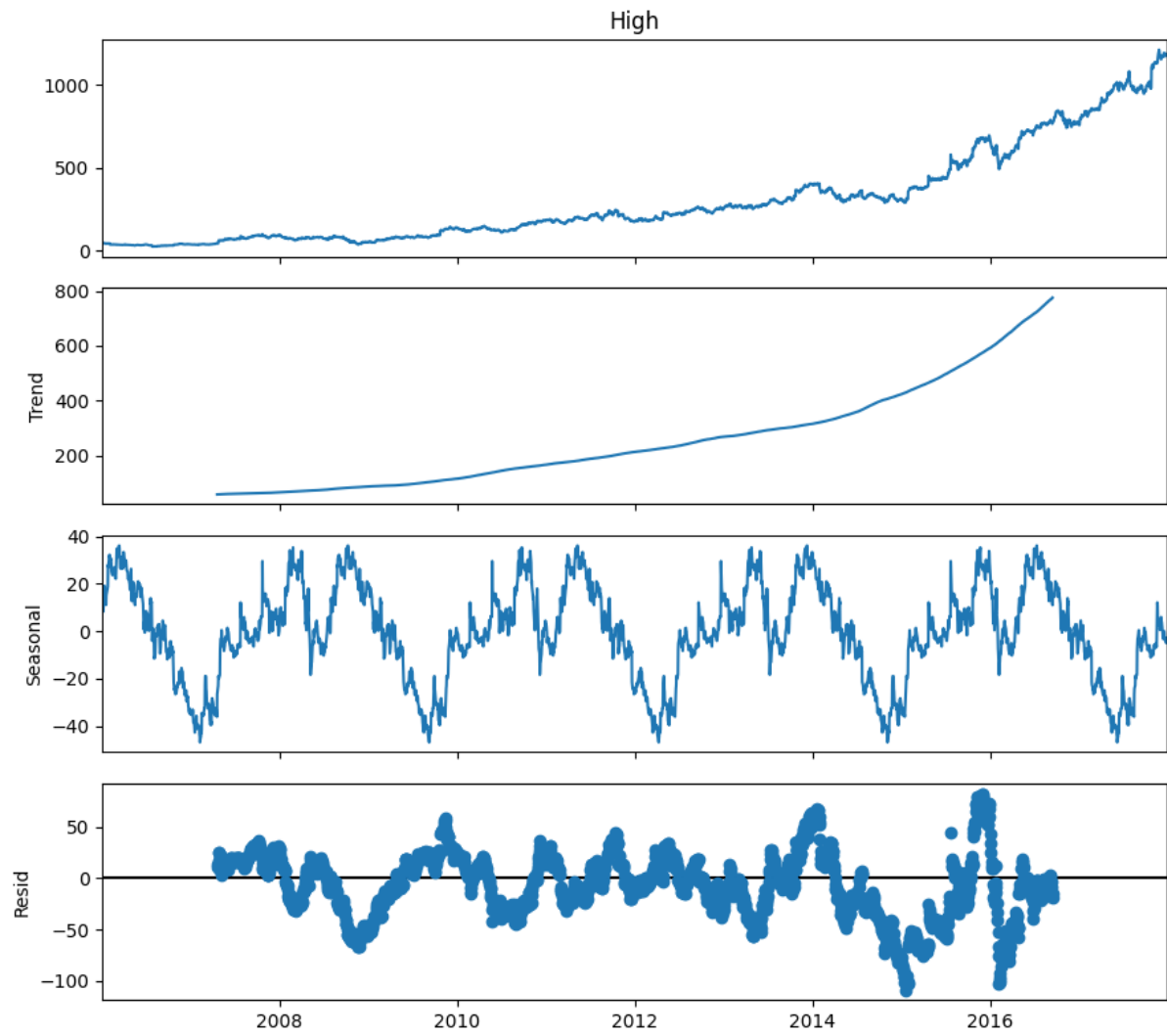
For the time series decomposition of the opening, lowest and highest prices, see Figures 6, 7 and 8. As earlier mentioned, the datasets for these three (including closing price) variables are identical and the patterns evident in Figure 5 are also obvious in Figures 6, 7 and 8.



**Figure 6:** Time Series Decomposition of Opening Price Using Additive Model: Seasonal, Trend, and Residual Components



**Figure 7:** Time Series Decomposition of Lowest Price Using Additive Model: Seasonal, Trend, and Residual Components



**Figure 8:** Time Series Decomposition of Highest Price Using Additive Model: Seasonal, Trend, and Residual Components