Report on Time Series Analysis with Stock Price Data

Introduction

The Time Series Data Analysis project focuses on analyzing and forecasting stock prices over time to understand their behavior and trends. This report details the steps taken to analyze historical stock prices for the company AABA. The dataset includes features such as Open, High, Low, Close, Volume, and Name, which are examined using various time series analysis techniques.

Data Preparation

The dataset was imported from a CSV file named 'stock_data.csv'. The following steps were performed to prepare the data for analysis:

1. **Data Cleaning:**
   - **Dropped Unnecessary Columns:** Removed the 'Unnamed: 0' column, which was not relevant to the analysis.
   - **Set Index:** Set the 'Date' column as the index to facilitate time series analysis.

```
df = pd.read_csv('stock_data.csv', parse_dates=True, index_col='Date')
df.drop('Unnamed: 0', axis=1, inplace= True)
```

2. **Initial Data Inspection:**
   - Displayed the first 10 rows of the dataset to understand its structure and contents.

Date	Open	High	Low	Close	Volume	Name
2006-01-03	39.69	41.22	38.79	40.91	24232729	AABA
2006-01-04	41.22	41.90	40.77	40.97	20553479	AABA
2006-01-05	40.93	41.73	40.85	41.53	12829610	AABA
2006-01-06	42.88	43.57	42.80	43.21	29422828	AABA
2006-01-09	43.10	43.66	42.82	43.42	16268338	AABA
2006-01-10	42.96	43.34	42.34	42.98	16288580	AABA
2006-01-11	42.19	42.31	41.72	41.87	26192772	AABA
2006-01-12	41.92	41.99	40.76	40.89	18921686	AABA
2006-01-13	41.00	41.08	39.62	39.90	30966185	AABA
2006-01-17	39.09	40.39	38.96	40.11	42429911	AABA

Exploratory Data Analysis (EDA)

1. **Plotting High Prices Over Time:**
   - Used Seaborn to plot the 'High' prices over time, providing a visualization of the stock's highest prices on a daily basis.

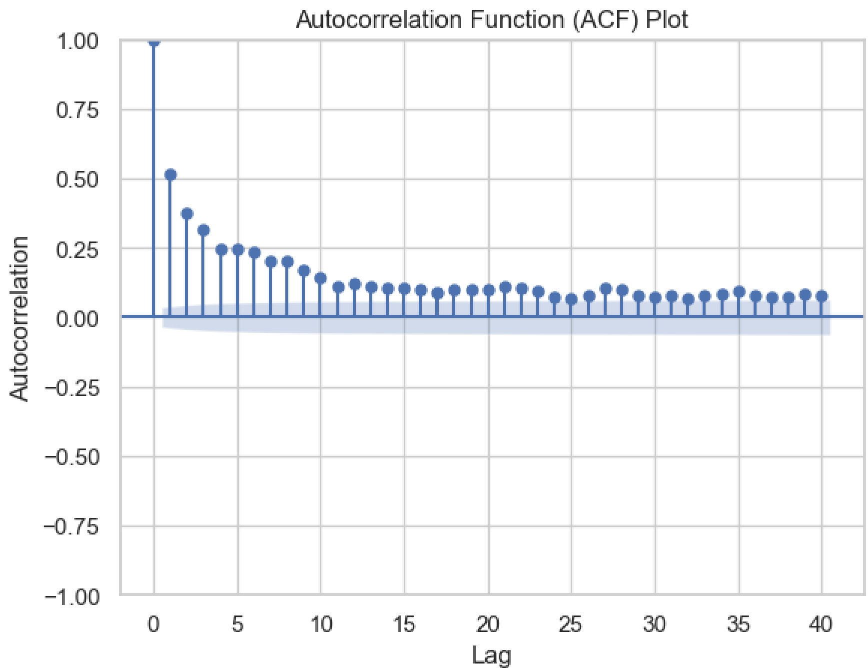
- The plot clearly shows the daily variations in the highest stock prices, indicating periods of volatility and stability in the stock market.

2. **Monthly Resampling:**
   - Resampled the data to a monthly frequency using the mean aggregation function to observe long-term trends.

The monthly resampling plot smooths out short-term fluctuations and highlights the overall trend in the stock prices over a longer period. We have observed an upward trend in the resampled monthly volume data. An upward trend indicates that, over the monthly intervals, the "high" column tends to increase over time.

Seasonality and Stationarity

1. **Autocorrelation Function (ACF):**
   - Plotted the ACF of the 'Volume' to detect seasonality in the data.



The ACF plot helps identify any repeating patterns or seasonality in the stock's trading volume. Significant spikes at specific lags would indicate a strong seasonality component.

2. **Stationarity Test:**
    - Conducted the Augmented Dickey-Fuller (ADF) test on the 'High' prices to check for stationarity.

```
ADF Statistic: 0.7671404880535945
p-value: 0.9910868050318213
Critical Values: {'1%': np.float64(-3.4325316347197403), '5%': np.float64(-2.862503905260741), '10%': np.float64(-2.5672831121111113)}
```

The initial ADF test resulted in an ADF Statistic of 0.767 and a p-value of 0.991, indicating that the 'High' prices are non-stationary. This is evident as the p-value is significantly higher than the typical threshold of 0.05.

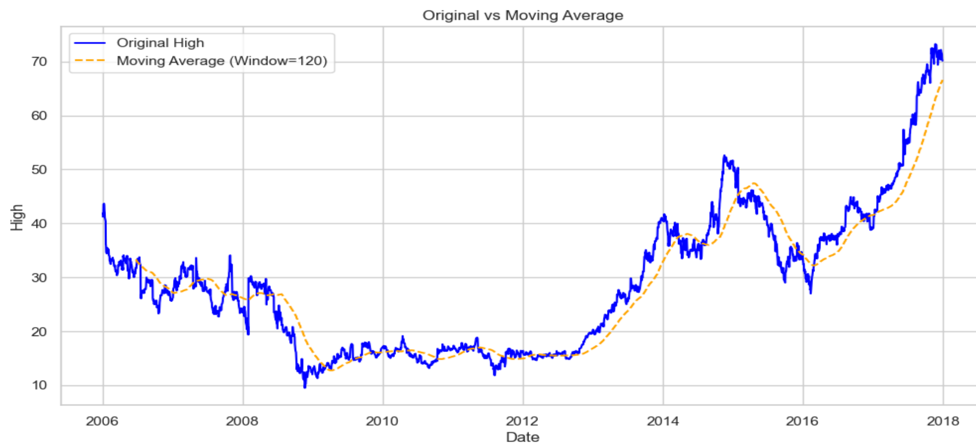
3. **Data Transformation**
   Differencing:
    - Applied differencing to the 'High' prices to remove trends and achieve stationarity.

```
            High  high_diff
Date
2006-01-03  41.22        NaN
2006-01-04  41.90       0.68
2006-01-05  41.73      -0.17
2006-01-06  43.57       1.84
2006-01-09  43.66       0.09
```

Differencing helps to stabilize the mean of a time series by removing changes in the level of a time series, thus eliminating trend and seasonality.

Moving Average:

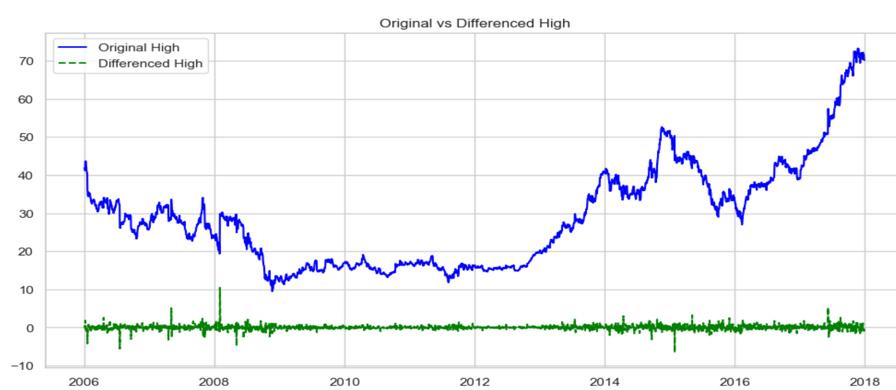
- Used a moving average with a window size of 120 to smooth the data.



Smoothing the data using a moving average helps to highlight the underlying trends by reducing noise. This is particularly useful for visualizing long-term trends in stock prices.

Comparison and Final Stationarity Test

1. **Visual Comparison:**
   - Compared the original 'High' prices with the differenced and smoothed versions to visualize the transformations.



The comparison plot shows how differencing and smoothing affect the data. The differenced series fluctuates around zero, indicating that trends have been removed. The smoothed series reduces short-term fluctuations, highlighting the overall trend.

2. **Final ADF Test:**
   - Conducted a second ADF test on the differenced data, which confirmed stationarity with a significantly low p-value.

```
ADF Statistic: -12.14836747834325
p-value: 1.5912766134148351e-22
Critical Values: {'1%': np.float64(-3.4325316347197403), '5%': np.float64(-2.862503905260741), '10%': np.float64(-2.5672831121111113)}
+ Code  + Markdown
```

The second ADF test resulted in an ADF Statistic of -12.148 and a p-value of 1.591e-22, indicating that the differenced series is stationary. This suggests that the transformations successfully removed trends and seasonality from the data.

Conclusion

- This project demonstrated the application of time series analysis techniques on stock price data. Key steps included data cleaning, exploratory data analysis, seasonality detection, and stationarity testing. Transformations like differencing and moving averages were crucial in preparing the data for further analysis and forecasting.
- The insights gained from this analysis provide a foundation for more advanced time series modeling and prediction, enabling better investment decisions based on historical stock price trends. Future work could involve implementing forecasting models such as ARIMA or LSTM to predict future stock prices based on the transformed, stationary data.