

Comprehensive Exploratory Data Analysis (EDA) of Stock Price Time Series Data

Team Semicolon

Introduction

Exploratory Data Analysis (EDA) is a crucial step in understanding the underlying patterns, trends, and potential anomalies in a dataset. This report provides a comprehensive EDA on a financial time series dataset, specifically analyzing stock market data spanning from **March 17, 1980, to December 27, 2024**. The dataset consists of various price-related attributes, including **Open, High, Low, Close, Adjusted Close, and Volume**.

The objective of this EDA is to:

- Gain insights into the data distribution, trends, and seasonality.
- Identify missing values and inconsistencies in the dataset.
- Analyze correlations between different financial indicators.
- Detect outliers or anomalies in price movements and trading volume.
- Visualize key patterns that may aid in predictive modeling and decision-making.

The following sections present a detailed analysis, supported by statistical summaries and visualizations.

1. Handling Missing Values

Before proceeding with analysis, we examined missing values in the dataset:

Column	Missing Values
Date	110
Adj Close	93
Close	117
High	95
Low	127
Open	103
Volume	145

To address these:

- **Rows with missing dates (110) were removed** to maintain time series consistency.
- **Price-related columns (up to 127 missing values) were forward-filled (**ffill**) and backward-filled (**bfill**)** to ensure smooth data continuity.
- **Missing Volume values (145) were replaced with 0**, assuming no trading occurred on those days.

This preprocessing ensures a complete and consistent dataset, making it suitable for further exploration and modeling.

2. Descriptive Statistics

A summary of the dataset's key statistics provides insights into the distribution of stock prices and trading volume:

Stock Prices (Adj Close, Close, High, Low, Open)

Statistic	Adj Close	Close	High	Low	Open
Count	11,181	11,181	11,181	11,181	11,181
Mean	63.67	72.13	72.55	71.71	68.02
Std Dev	52.35	51.36	51.62	51.04	55.92
Min	2.26	3.24	3.24	3.24	0.00
25%	19.22	27.55	27.74	27.55	0.00
50% (Median)	50.62	66.06	66.66	65.47	66.00
75%	104.84	114.40	114.98	113.68	114.50
Max	254.77	254.77	255.23	253.59	255.00

- The **mean closing price is \$72.13**, with a **maximum of \$254.77** and a **minimum of \$3.24**.
- The **high standard deviation (51.36)** suggests significant price volatility.
- The **Open price has many zero values**, which may indicate missing or unrecorded trading sessions.

Trading Volume

Statistic	Volume
Count	11,181
Mean	211,905
Std Dev	387,354
Min	0
25%	12,600
50% (Median)	85,300
75%	289,400
Max	18,582,700

- **Trading volume varies significantly**, with a **mean of ~211K shares** but a **maximum of over 18.5M shares** in a single day.
- The **minimum value is 0**, aligning with the previous missing value treatment.
- The **wide range and high standard deviation** suggest periods of both high and low market activity.

3. Stationarity Test: Augmented Dickey-Fuller (ADF) Test

To check if the '**Close**' price series is stationary, we performed the **Augmented Dickey-Fuller (ADF) test**.

Results:

- **ADF Statistic:** -0.4874
- **p-value:** 0.8944

Interpretation:

- The **p-value (0.89)** is **much higher than 0.05**, meaning we fail to reject the **null hypothesis (H_0)** that the series has a **unit root** (i.e., it is **non-stationary**).
- This suggests the '**Close**' price series **has a trend or seasonality**, which must be addressed before applying time series models.

4. Stock Closing Prices Over Time

The time series plot of stock closing prices from **1980 to 2025** reveals key trends and patterns in price movement.



Observations:

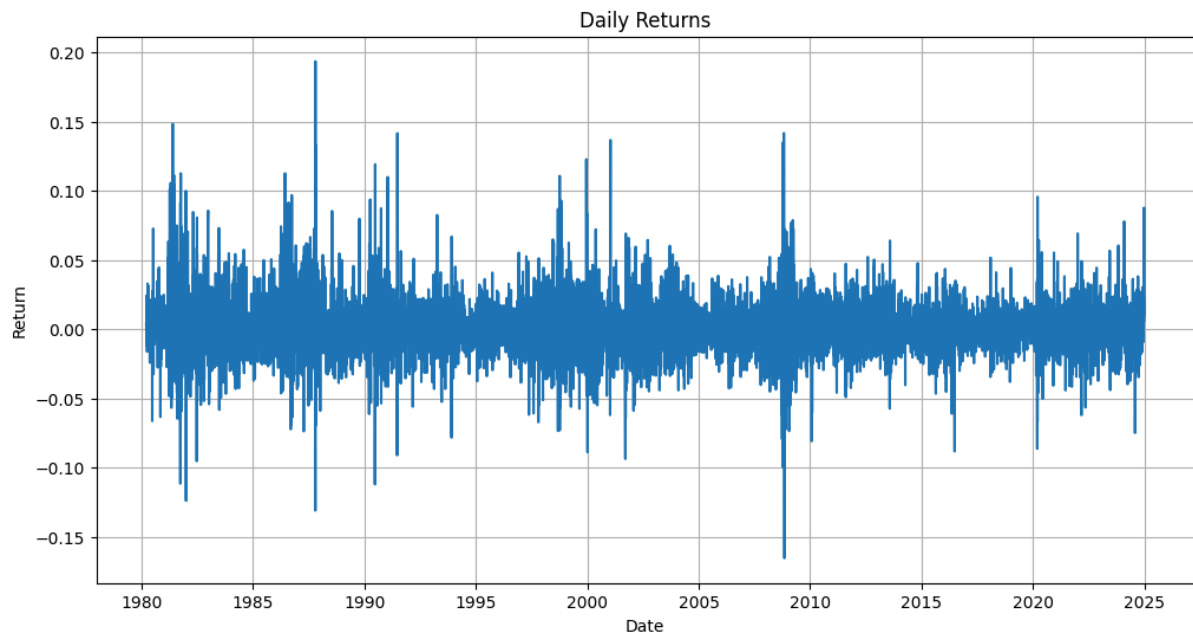
- The stock price has shown **significant long-term growth**, with multiple phases of rapid appreciation and corrections.
- Periods of high volatility are evident, particularly around **2000, 2008, and 2020**, likely due to major economic events or market shifts.
- A strong **upward trend** is observed, indicating long-term market appreciation despite short-term fluctuations.

Key Insights:

- The stock price is **non-stationary**, as confirmed by the **Augmented Dickey-Fuller test**, meaning it exhibits a trend over time.
- The presence of **market cycles and sharp fluctuations** suggests that external factors influence price movements.

5. Daily Returns Analysis

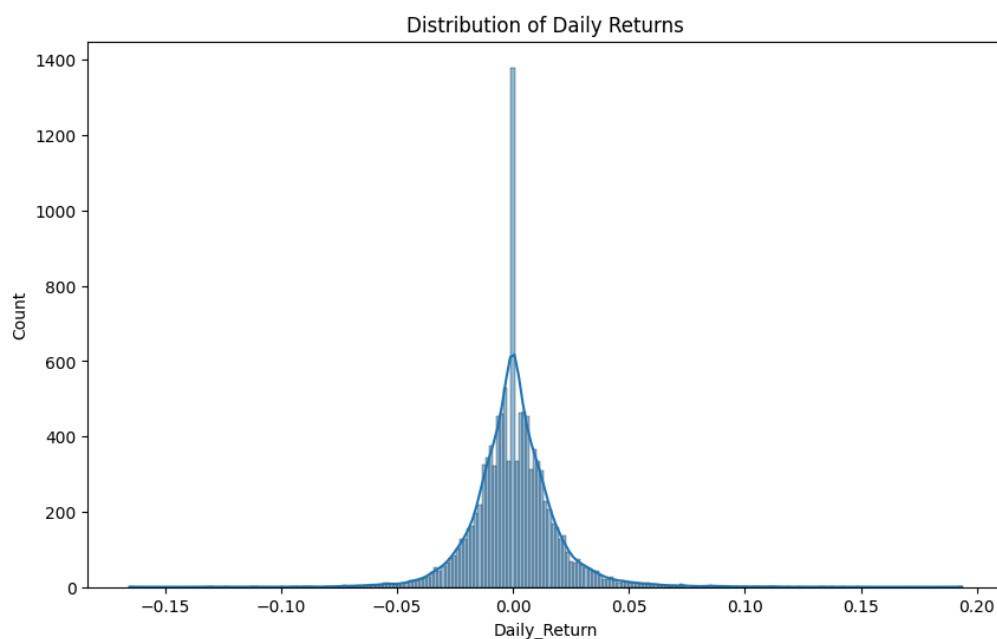
The plot above represents the **daily returns** of the stock over time, calculated as the **percentage change** in closing prices.



Observations:

- Daily returns exhibit **high volatility**, with **frequent spikes and dips**, indicating significant price fluctuations.
- There are **periods of increased volatility**, especially during market crashes or economic crises (e.g., around 2000, 2008, and 2020).
- While the returns fluctuate around **zero**, extreme positive and negative movements suggest **potential market shocks**.

6. Distribution of Daily Returns



The histogram of daily returns shows a **near-normal distribution**, centered around **zero**, indicating that most price changes are small. However, a **sharp peak at zero** suggests many days with little to no movement. The **fat tails** reveal occasional large price swings, highlighting periods of high volatility.

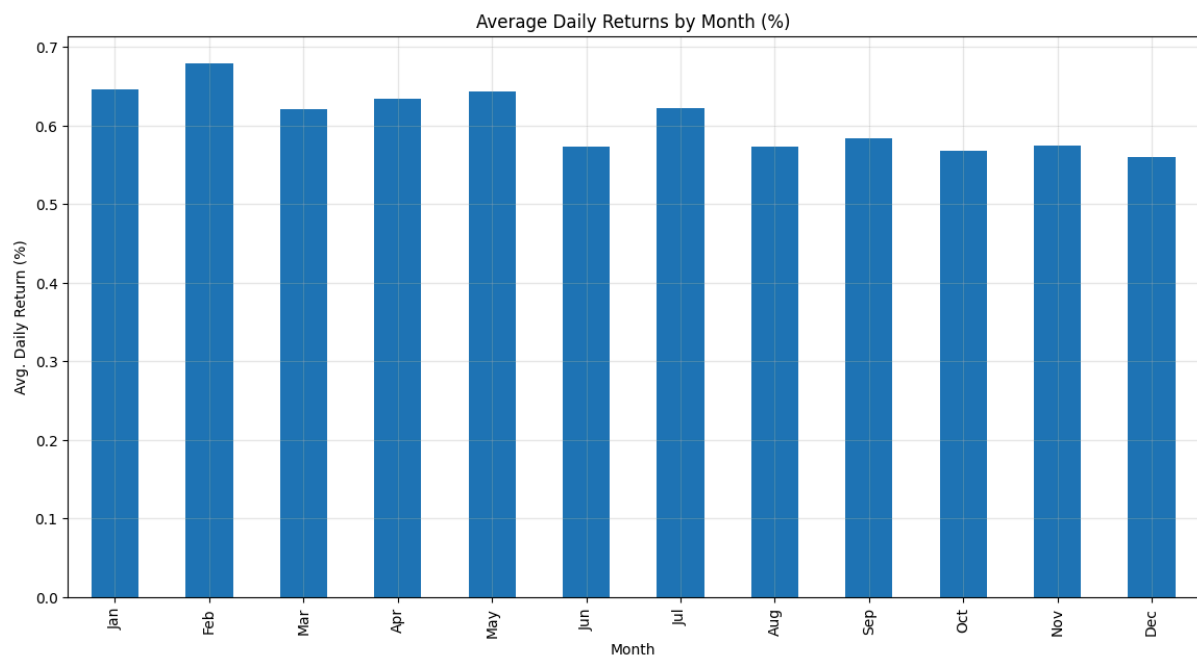
Key Takeaways:

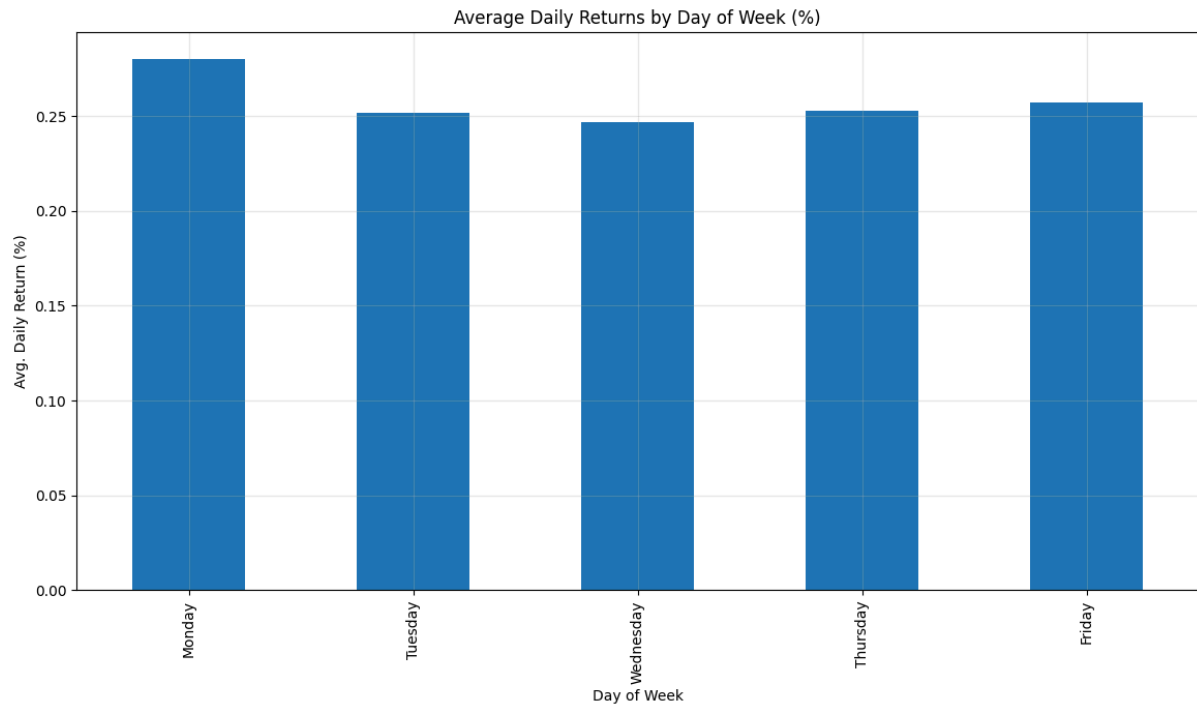
- **Frequent small returns** with occasional extreme values.
- **Fat tails** indicate higher-than-normal risk.
- **High peak at zero** suggests low activity on certain days.

This insight is crucial for understanding volatility and risk in the asset's price movements.

7. Seasonality in Returns

To analyze potential seasonal patterns, we examined **average daily returns by month and day of the week**.





Key Observations:

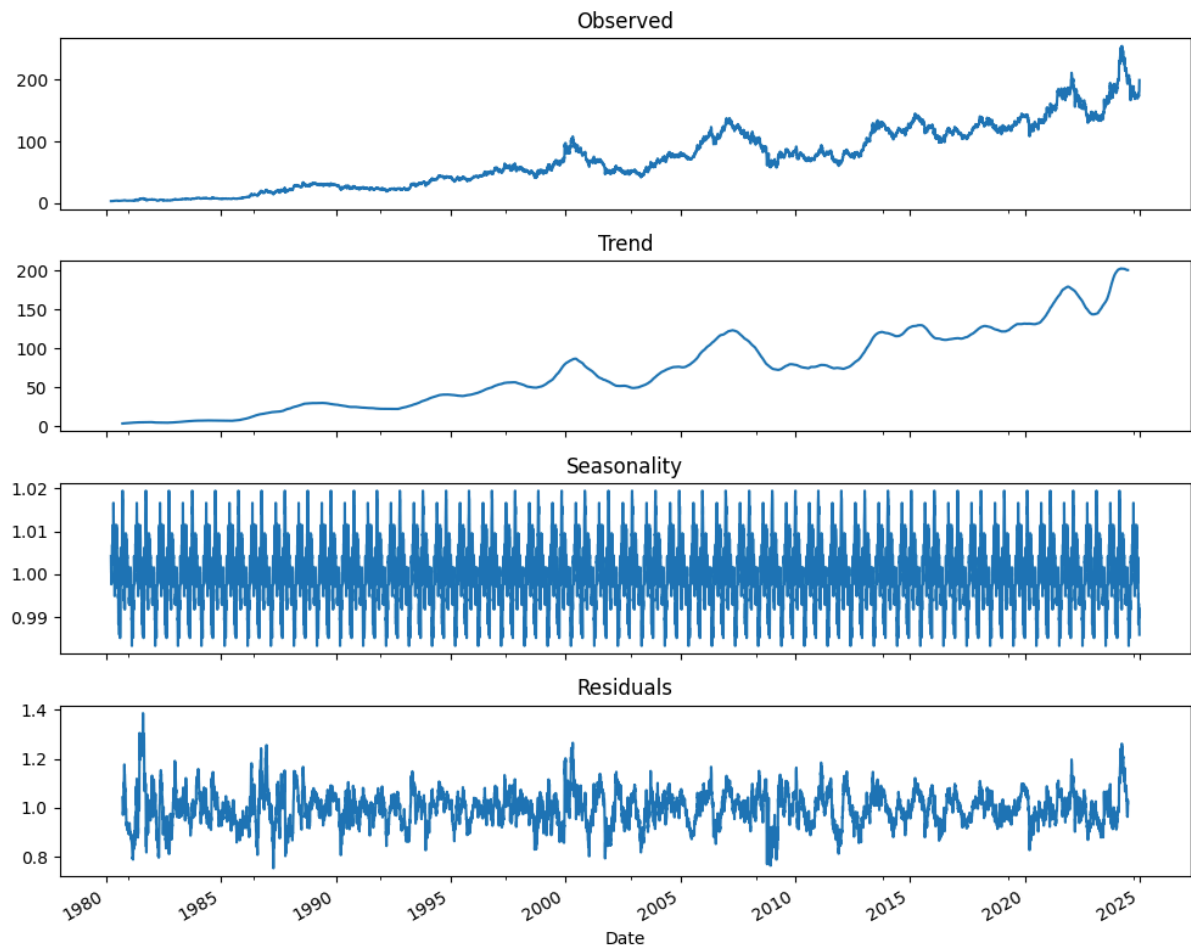
- **Monthly Returns:** There is no significant variation across months, suggesting **no strong seasonal trend** in returns.
- **Day-of-Week Returns:** Similarly, returns across weekdays are fairly uniform, indicating **no clear weekday effect**.

Implications:

- No strong seasonal bias means that **returns are not heavily influenced by calendar effects**.
- Other factors like macroeconomic events or market conditions may play a larger role in price movements.

8. Seasonal Decomposition of Closing Prices

This seasonal decomposition breaks the **closing price** into four components: **Observed**, **Trend**, **Seasonality**, and **Residuals**.



Key Observations:

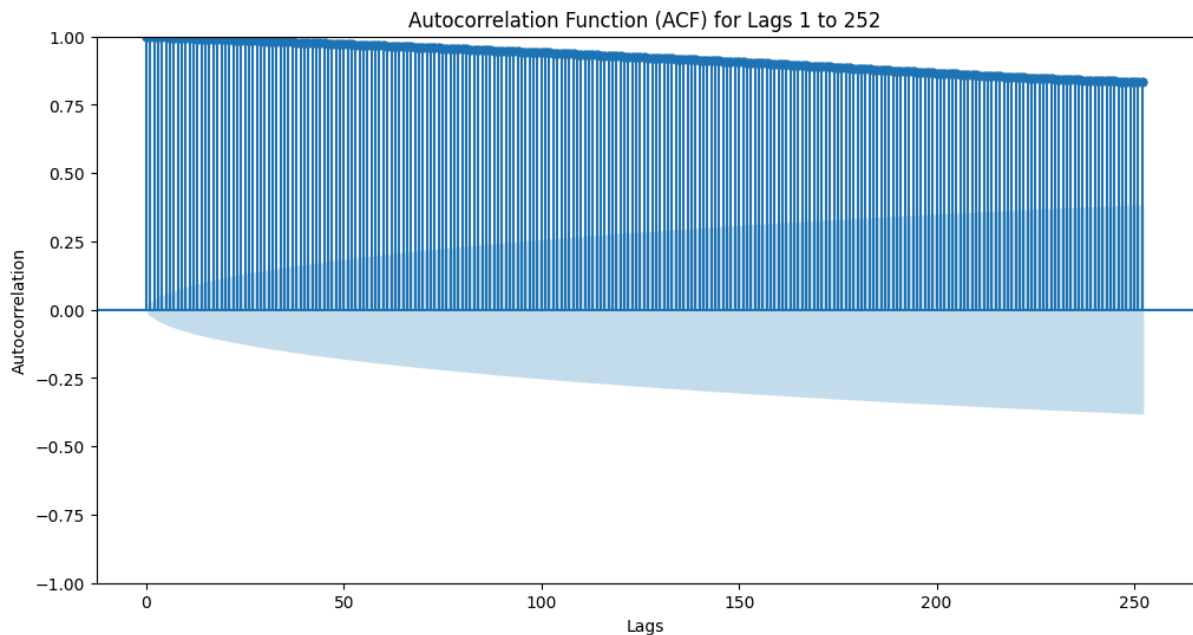
- **Trend:** A clear upward trend is visible, showing long-term price growth with periodic dips.
- **Seasonality:** Small but consistent seasonal fluctuations, likely reflecting recurring market patterns.
- **Residuals:** The noise component shows higher volatility in some periods, indicating possible external influences like economic events.

Implications:

- The **strong trend** suggests long-term growth, making this asset potentially attractive for long-term investors.
- The **seasonality effect is minor**, meaning short-term trading strategies based purely on seasonality may be less effective.
- The **residuals indicate varying volatility**, which could impact risk management and forecasting accuracy.

9. Autocorrelation Analysis (ACF) of Closing Prices

The **Autocorrelation Function (ACF)** plot helps analyze how past closing prices influence future prices over different lags (up to 252 trading days).



Key Observations:

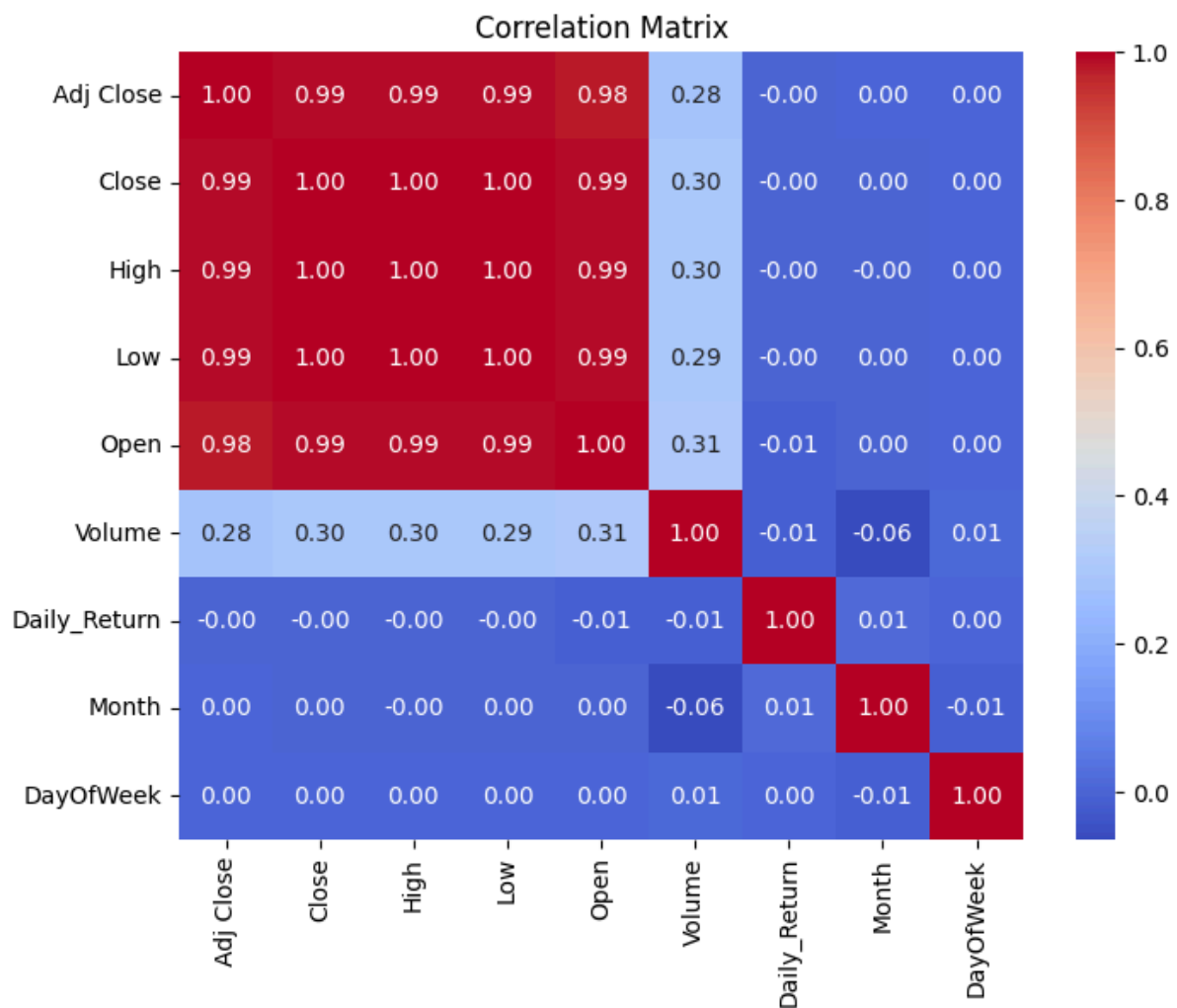
- **High autocorrelation at all lags**, gradually declining, indicates **strong persistence in price movements**.
- This suggests the stock exhibits a **long-term memory effect**, meaning past prices heavily influence future prices.
- The values remain **well above the confidence band**, confirming statistical significance.

Implications:

- The **strong autocorrelation** suggests that price movements are not purely random, supporting the use of trend-following strategies.

10. Correlation Analysis

A heatmap visualizing feature correlations reveals:



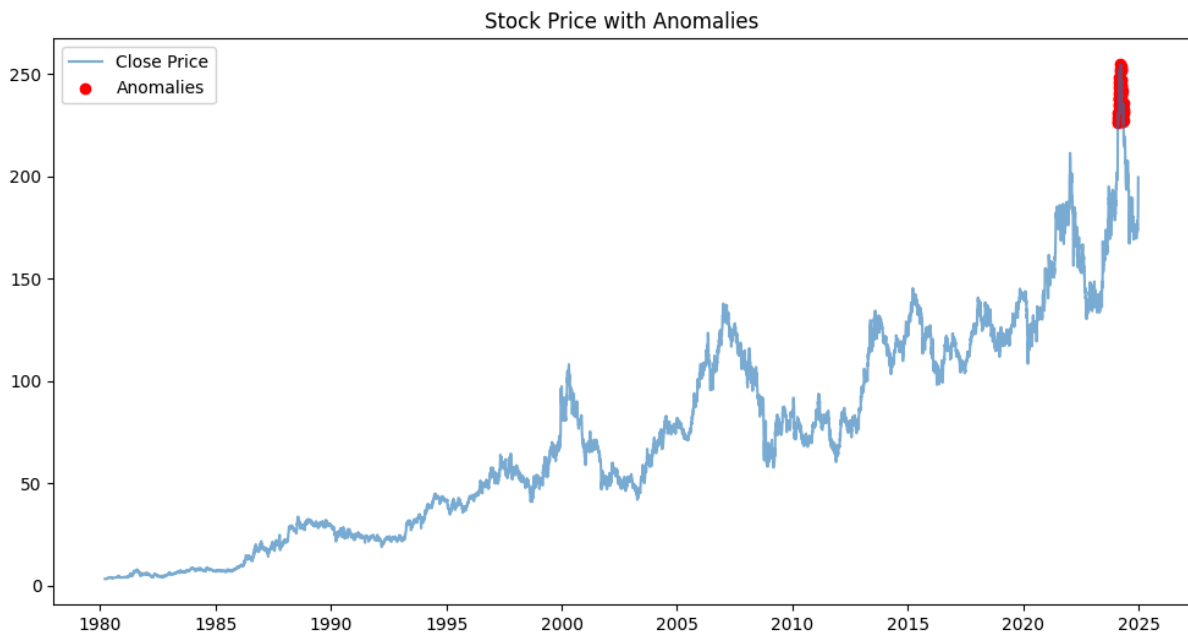
- **Strong Positive Correlations:** Prices (**Adj Close**, **Close**, **High**, **Low**, **Open**) are nearly identical, indicating redundant information.
- **Weak Price-Volume Link:** **Volume** shows a weak positive correlation (~0.28–0.31) with prices.
- **Minimal Impact of Time & Returns:** **Daily_Return**, **Month**, and **DayOfWeek** have negligible correlations with prices and volume.

11. Anomaly Detection with Z-Score

To identify unusual price movements, we used the **Z-score method**:

- **Z-Score Calculation:** Measures how many standard deviations a value is from the mean.

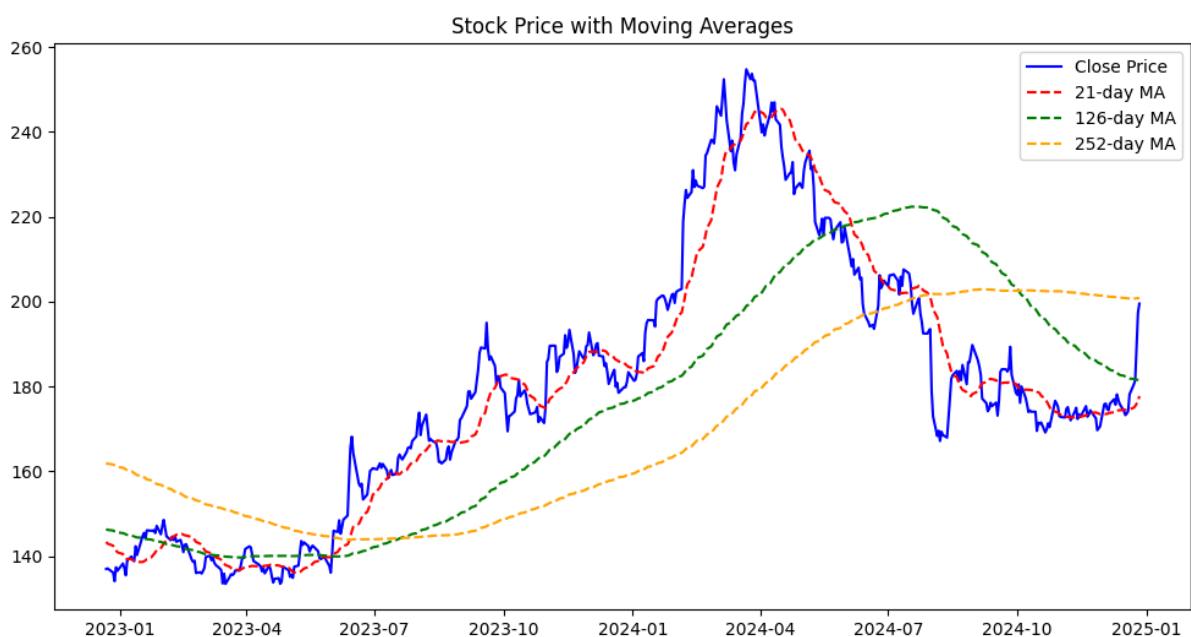
- **Threshold for Anomalies:** Points with Z-scores beyond ± 3 are flagged as anomalies.



The plot shows:

- **Close Price Trend:** The stock price generally trends upward with periods of volatility.
- **Detected Anomalies:** Red points highlight significant deviations, mostly concentrated in recent years

12. Stock Price with Moving Averages



The plot shows the stock's closing price over time, alongside three key moving averages:

- **21-day Moving Average (MA)** (short-term trend)
- **126-day Moving Average (MA)** (medium-term trend)
- **252-day Moving Average (MA)** (long-term trend)

These indicators help smooth out price fluctuations and reveal underlying trends.

Key Observations:

Trend Analysis:

- The closing price exhibits significant upward momentum through most of 2023, peaking around early 2024, followed by a steep decline.
- After the drop, the price consolidates and starts to stabilize toward the end of 2024, with signs of a slight upward bounce.

Moving Averages Behavior:

- The **21-day MA** closely tracks the stock price, reacting quickly to short-term price changes.
- The **126-day MA** lags more, smoothing out medium-term fluctuations and acting as a dynamic support/resistance level.
- The **252-day MA** is the slowest to react, showing the long-term trend. It peaks later than the price and flattens as the stock stabilizes.

Crossovers:

- Several **bullish crossovers** occur when the shorter moving average (e.g., 21-day) crosses above the longer ones — signaling potential buying opportunities.
- Conversely, **bearish crossovers** signal potential downtrends, such as when the 21-day MA falls below the longer moving averages.

Volatility and Price Action:

- The stock price shows significant volatility, with sharp spikes and steep drops.
- Periods of large divergence between the price and long-term MAs (e.g., early 2024) suggest potential overbought/oversold conditions.