Exploratory Data Analysis Report

1. Data Overview

- The dataset contains **historical stock price data from 1980 onward**.

- **Key attributes:**
  - `Date`: The trading day.
  - `Open`: Opening price of the stock.
  - `High`: Highest price reached during the trading day.
  - `Low`: Lowest price reached during the trading day.
  - `Close`: Closing price of the stock.
  - `Adj Close`: Adjusted closing price (accounts for stock splits & dividends).
  - `Volume`: Number of shares traded.
  -
- **Initial Observations:**

  - The dataset has missing values in key columns, which need to be addressed.
  - There are periods of high and low volatility, indicating fluctuating market conditions.

---

2. Handling Missing Data

- **Identified Issues:**
  - Missing values in `Close`, `Open`, `High`, `Low`, and `Volume`.
  - Some `Open` values are recorded as `0`, which might indicate missing data.
- **Solution Applied:**
  - Forward-fill method was used to propagate the last valid price value to maintain continuity.
  - Dropped rows where `Date` was missing to preserve time consistency.
  - Ensured all processed data is consistent for feature engineering.

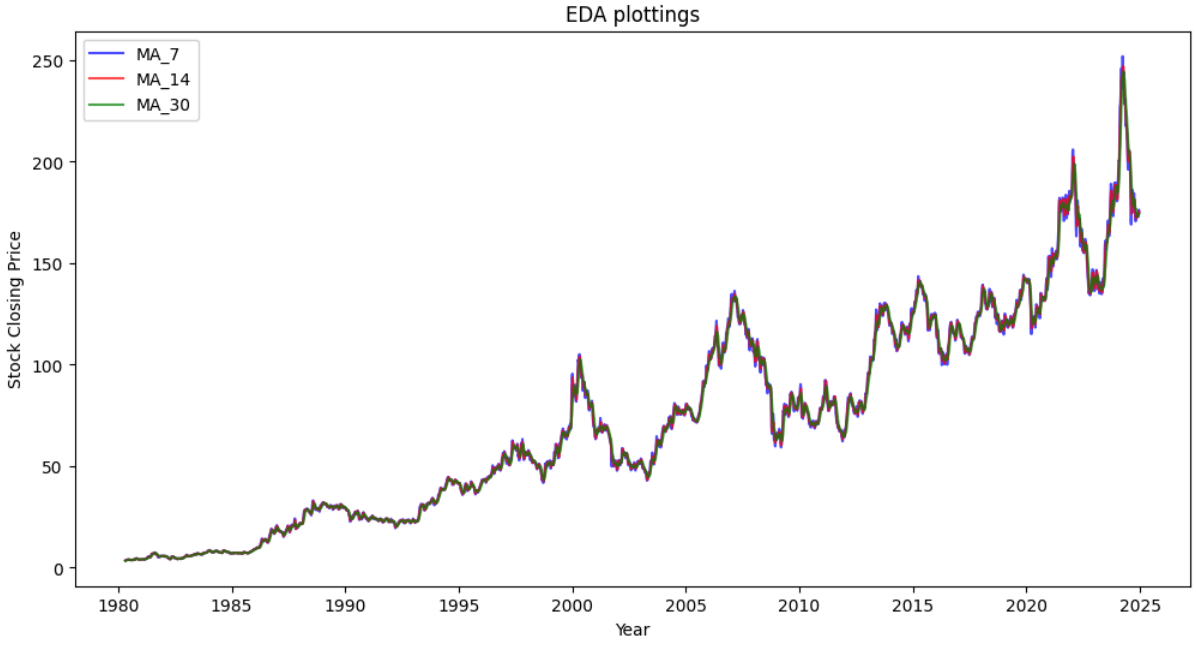
---

3. Key Visualizations & Relationships

3.1 Stock Price Trends Over Time

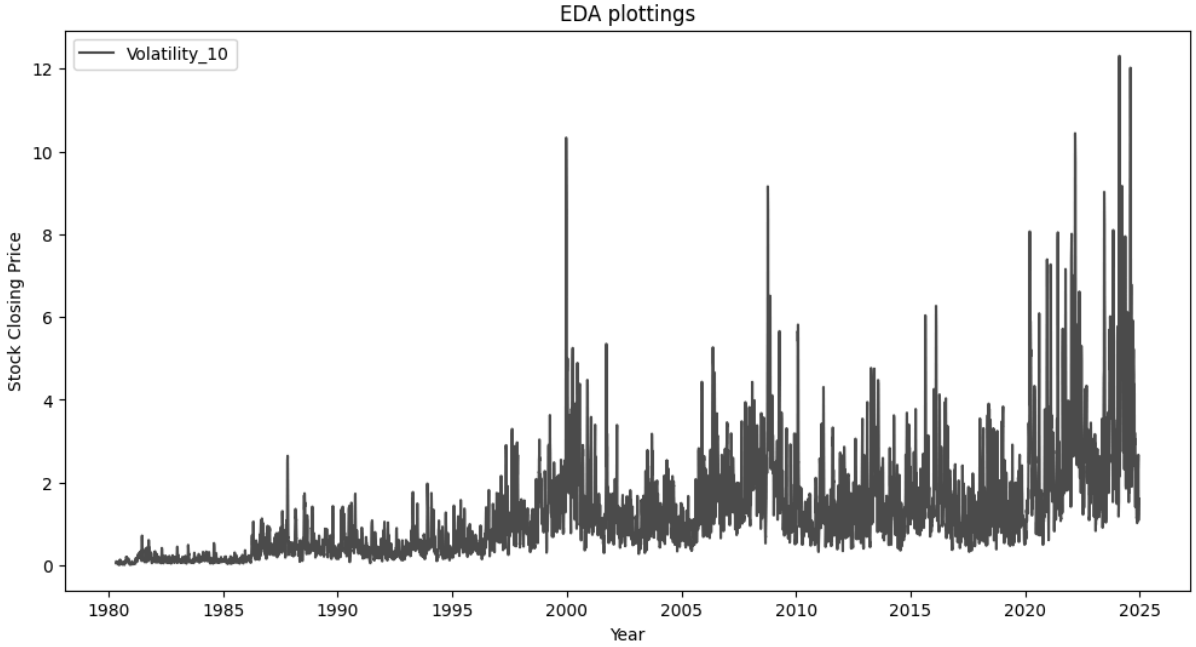
- **Time series plot** of closing prices shows a long-term **upward trend** with periodic fluctuations.
- Certain periods exhibit **sharp increases/decreases**, possibly linked to market events.

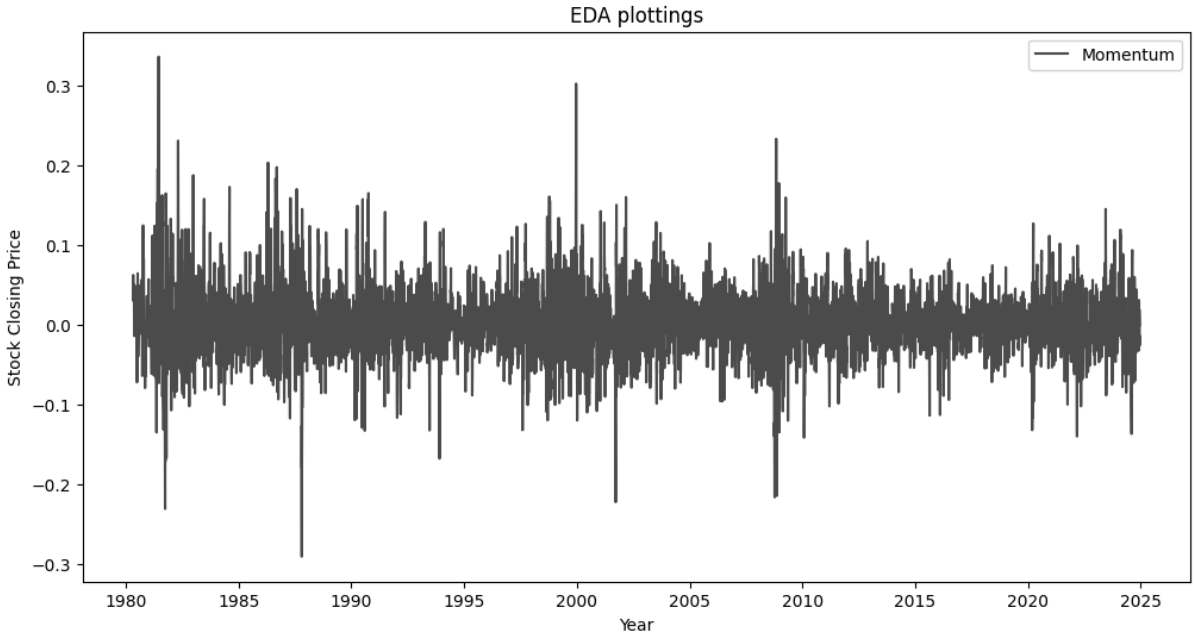
3.2 Moving Averages for Trend Identification



- Applied **7-day, 14-day, and 30-day Moving Averages** to smooth price fluctuations.
- Moving averages help in identifying short-term and long-term trends.

3.3 Seasonality & Volatility Analysis





- **Stock prices are cyclical**, with periods of **high volatility** followed by stabilization.
- Rolling **10-day standard deviation** used to track volatility over time.

---

4. Trends, Seasonality, and Anomalies

4.1 Long-Term Trend

- Overall, the stock price has shown a strong **upward trend**, driven by growth in market capitalization and investor interest.

4.2 Seasonality

- The stock market **follows cycles**, where prices tend to be **higher during certain periods (e.g., Q4 earnings seasons, holiday spikes)**.
- **Yearly seasonality patterns observed**: Stronger gains at the start of the year, potential dips mid-year.

4.3 Volatility & Anomalies

- Certain **sharp price drops/spikes** are **not explained by normal trends** (potential market crashes, economic shifts).
- Periods of **high volatility** coincide with significant market events (e.g., financial crises, policy changes).

---

5. Feature Selection: Justification & Choices

The following features were selected based on **historical stock market trends and financial indicators**:

5.1 Moving Averages (MA_7, MA_14, MA_30)

- **Purpose**: Smoothens price fluctuations and identifies trends.
- **Justification**: Short-term traders use 7-day MA, while long-term investors track 30-day MA.

5.2 Volatility Indicator (Rolling 10-day Standard Deviation)

- **Purpose**: Measures price fluctuation intensity.
- **Justification**: High volatility often signals market uncertainty or upcoming trend reversals.

5.3 Momentum Indicator (5-day Rate of Change)

- **Purpose**: Measures price movement speed.
- **Justification**: If momentum is high, the stock is experiencing strong buying/selling pressure.

5.4 Relative Strength Index (RSI_14)

- **Purpose**: Determines if a stock is **overbought (>70) or oversold (<30)**.
- **Justification**: Helps detect trend reversals by identifying extreme conditions.

---

6. Data Preprocessing & Normalization

6.1 Normalization Approach

- Used **StandardScaler** to normalize numerical features (`Close`, `MA_7`, `MA_14`, `Momentum`, etc.).
- **Why?**
  - Ensures all features are on a **similar scale** to prevent model bias.
  - Helps in better model convergence for regression-based approaches.

6.2 Final Processed Dataset

- No missing values after preprocessing.
- All numerical features standardized.
- The dataset is now **clean and ready** for model training.

Final Takeaways

- **Stock price follows trends with clear seasonal cycles**.
- **Volatility & anomalies** indicate significant market events.
- **Feature selection is based on financial indicators** commonly used by traders.
- **Processed dataset is well-structured for predictive modeling**.