

1. Exploratory Data Analysis (EDA)

Dataset Overview

- **Period:** January 1, 2010, to December 31, 2024
- **Features:** Open, High, Low, Close, Adjusted Close, Volume

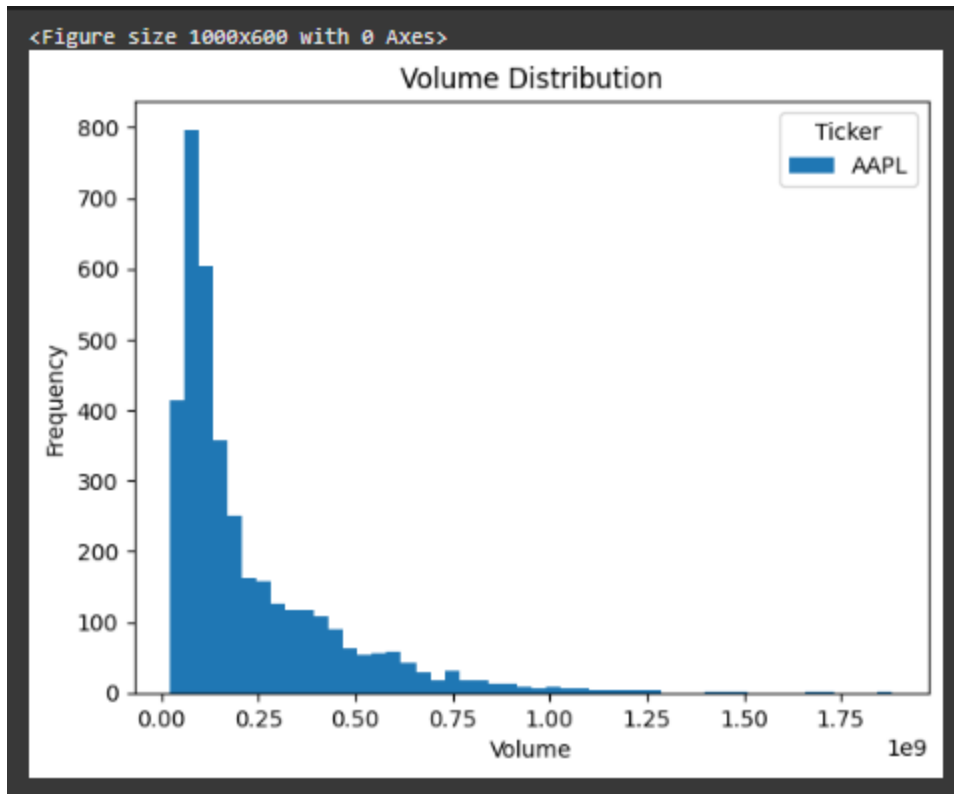
Price	Close	High	Low	Open	Volume
Ticker	AAPL	AAPL	AAPL	AAPL	AAPL
Date					
2010-01-04	6.440332	6.455078	6.391279	6.422878	493729600
2010-01-05	6.451466	6.487879	6.417460	6.458087	601904800
2010-01-06	6.348845	6.477044	6.342225	6.451465	552160000
2010-01-07	6.337110	6.379843	6.291067	6.372319	477131200
2010-01-08	6.379242	6.379844	6.291369	6.328685	447610800
...
2024-12-23	254.989655	255.369227	253.171646	254.490204	40858800
2024-12-24	257.916443	257.926411	255.009620	255.209412	23234700
2024-12-26	258.735504	259.814335	257.347047	257.906429	27237100
2024-12-27	255.309296	258.415896	252.782075	257.546826	42355300
2024-12-30	251.923019	253.221595	250.474615	251.952985	35557500

3773 rows x 5 columns

- **Data Integrity:**
 - Missing values were imputed using forward fill.
 - No duplicate records were identified.

Visualizations:

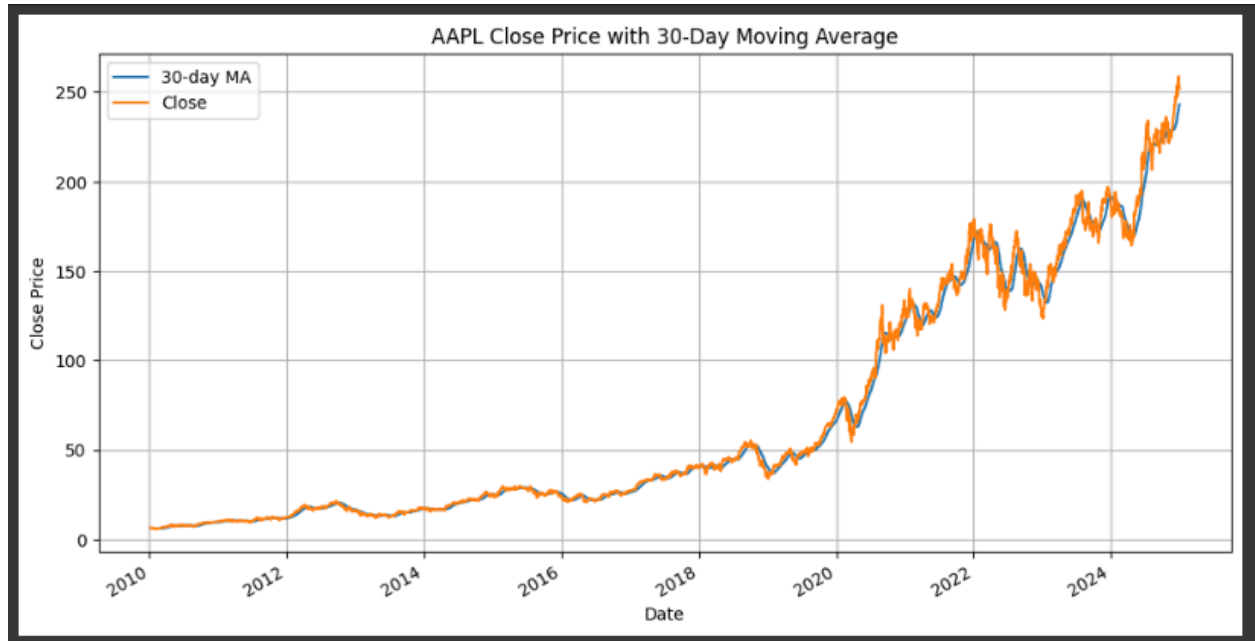
- **Volume Distribution:** A histogram indicated skewness and potential outliers in trading volume.



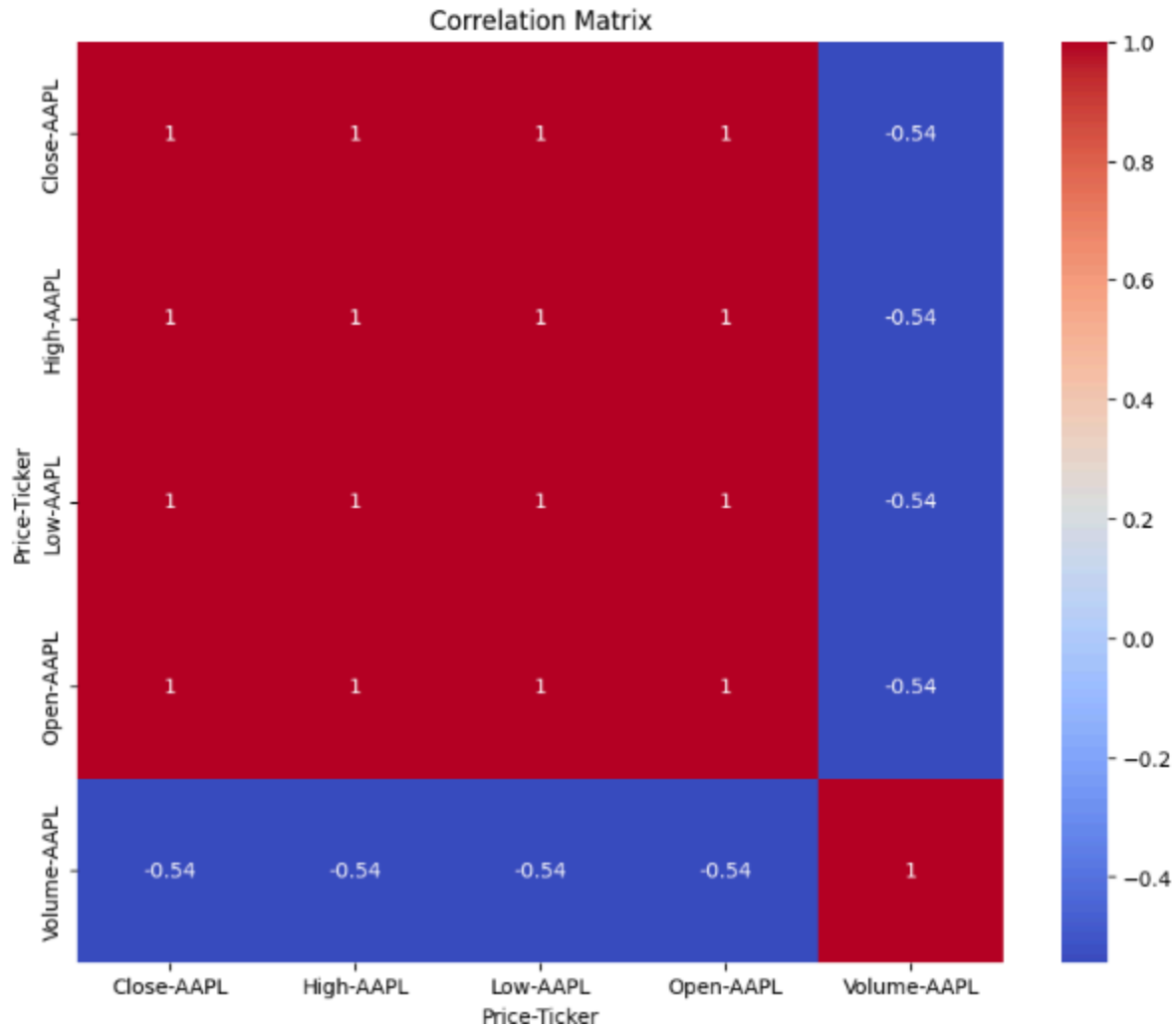
- **Price Trends:** Line plots demonstrated a consistent upward trend in prices over the period.



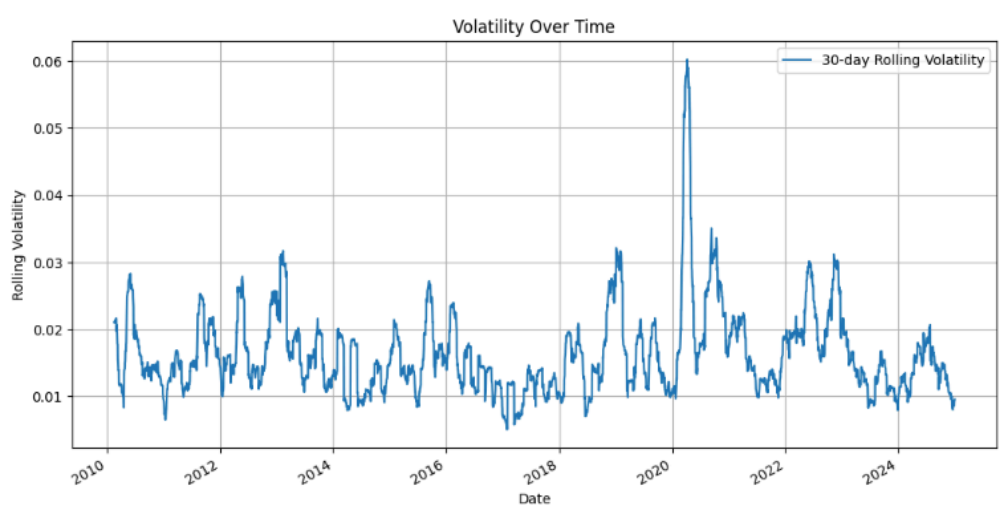
- **30-Day Moving Average:** Applied to the Close price, this smoothed short-term fluctuations, highlighting long-term trends.



- **Boxplots:** Illustrate the variance and spread of price-related features.
- **Correlation Matrix:** Strong positive correlations among price features were revealed.



- **Volatility Analysis:** Calculated as the rolling standard deviation of daily returns, indicating periods of fluctuating market activity.



2. Modeling Results

Preprocessing

- **Feature Scaling:** Standardized the features (Open, High, Low, Close, Volume) using StandardScaler to ensure consistent model performance.

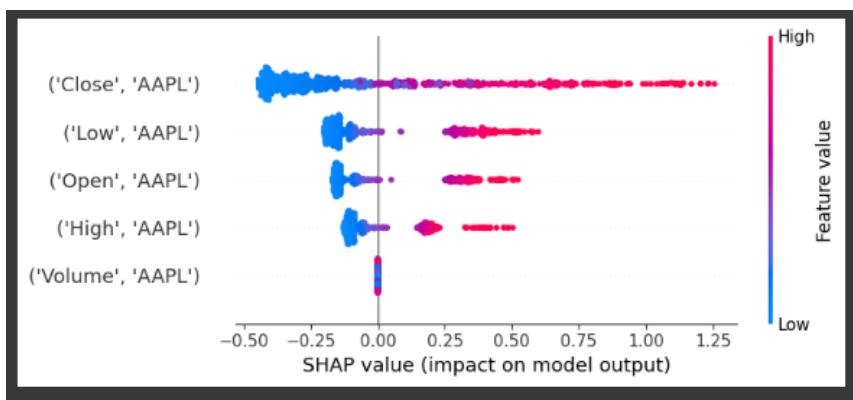
Model: Random Forest Regressor

- **Train/Test Split:** 80% training, 20% testing.
- **Evaluation Metrics:**
 - Mean Squared Error (MSE): Computed during model evaluation.
 - Root Mean Squared Error (RMSE): Derived from MSE.
 - R-squared (R^2): Measured model fit.
- **Hyperparameter Tuning:**
 - Performed using GridSearchCV with 5-fold cross-validation.
 - Optimal Parameters:
 - `n_estimators`: 200
 - `max_depth`: 20
 - `min_samples_split`: 5

3. Interpretability Techniques

SHAP (SHapley Additive Explanations)

- Quantified the contribution of each feature to model predictions.
- Key Insight: 'Open' and 'High' prices exhibited the most significant positive impact on predictions.



LIME (Local Interpretable Model-agnostic Explanations)

- Provided explanations for individual predictions, enhancing transparency.
- Visualized the localized effect of features on specific instances.



Partial Dependence Plots

- Depicted the average model response to variations in 'Open' and 'Volume'.
- Key Observation: 'Open' displayed a strong positive linear relationship with the target variable.

