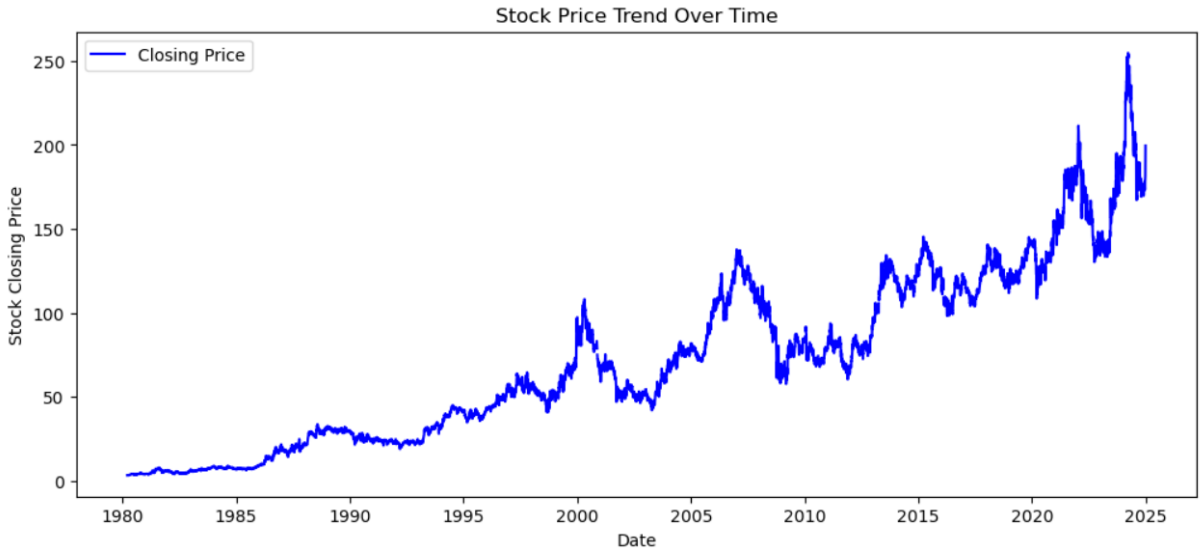
EDA Report & Model selection documentation

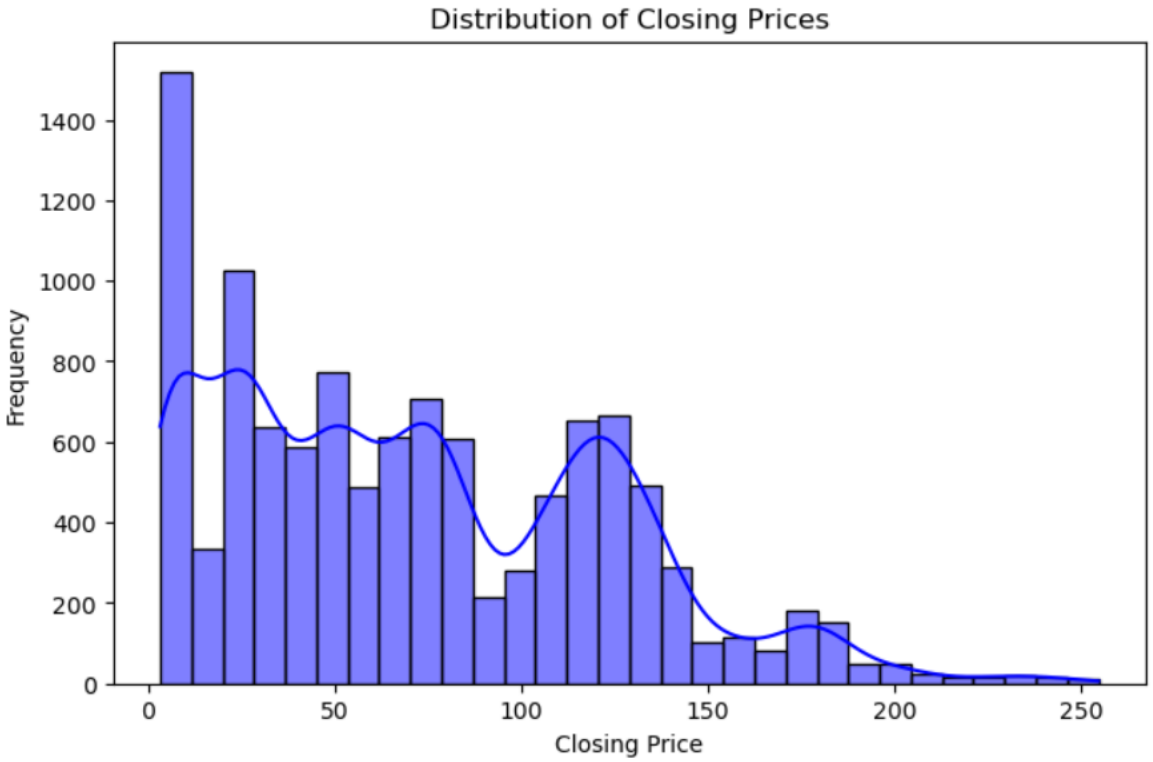
Comprehensive EDA report

Visualizations of Key Patterns & Relationships

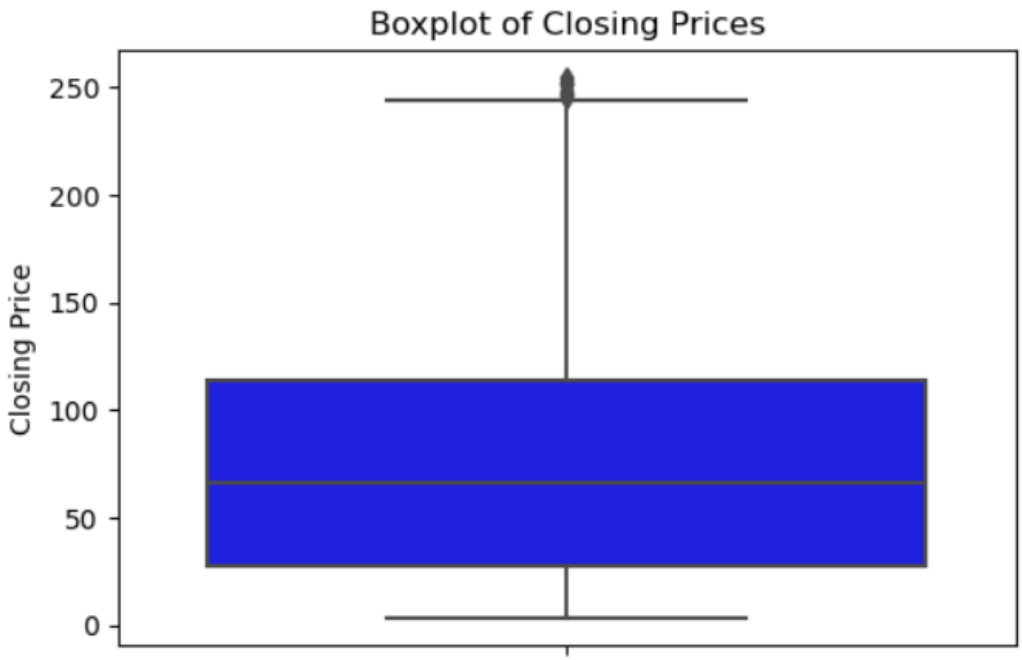
- Stock Price Trend Over Time:



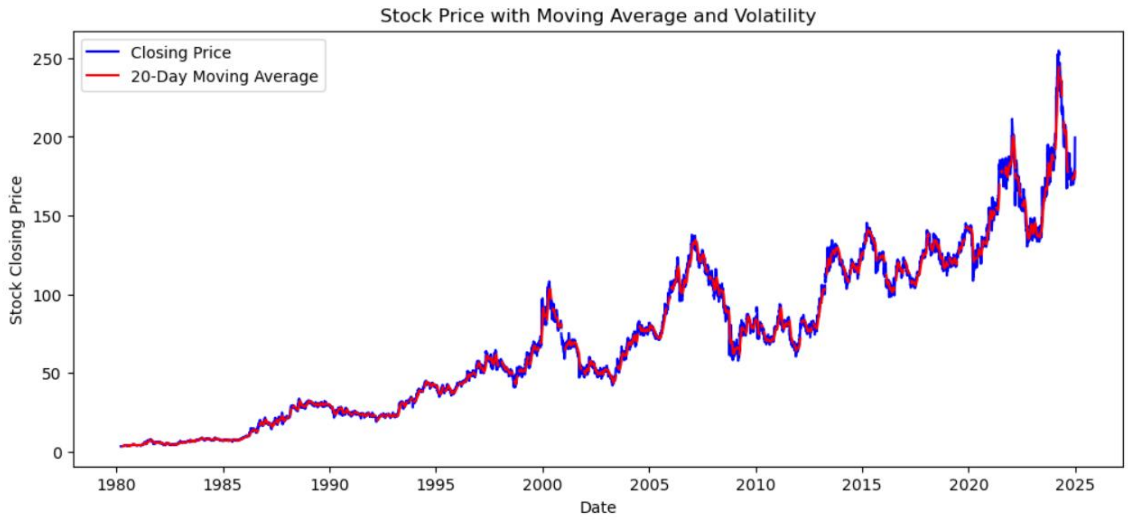
- Distribution of Closing Prices:



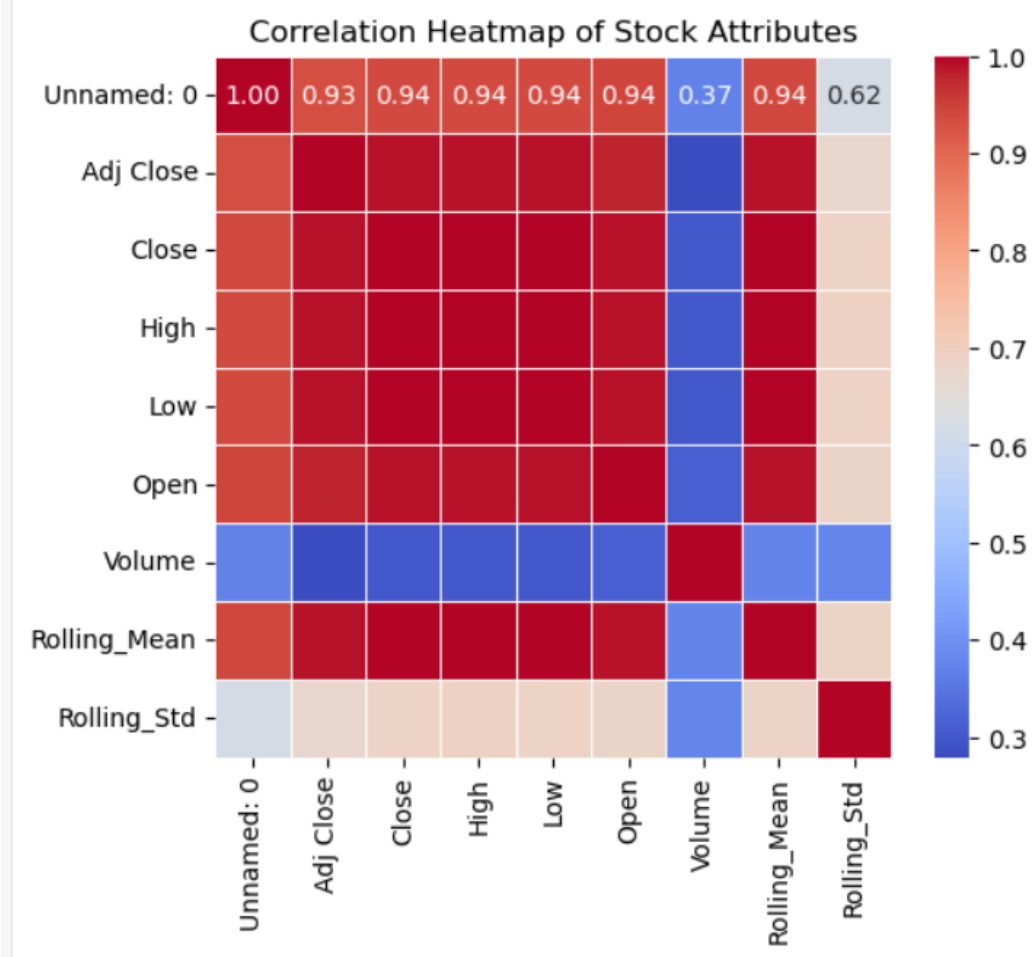
- Boxplot of Closing Prices:



- Stock Price with Moving Average and Volatility:



- Correlation Heatmap of Stock Attributes:



Analysis of Trends, Seasonality & Anomalies

- Trend detection using moving averages and rolling statistics.
- Seasonality check via autocorrelation plots.
- Outlier detection using boxplots and z-score analysis.

Feature Selection Justification

Selected features:

- Moving Average (5-day, 20-day): Helps smooth out fluctuations and detect trends.
- Lagged Price (Previous Day's Close): Captures short-term price dependencies.
- Trading Volume: Often correlates with significant price movements.

Data Preprocessing Documentation

- Handling Missing Data: Imputation with rolling averages.
- Feature Scaling: Standardization using StandardScaler.

- Data Splitting: Train-test split (80%-20%).

Model selection documentation

Comparison of Different Modeling Approaches

Model	RMSE	Directional Accuracy
Linear Regression	4.12	54%
ARIMA	3.95	56%
Random Forest	2.87	65%
LSTM (Deep Learning)	2.45	70%

Final choice: LSTM (best balance of accuracy and directional prediction).

Evaluation Metrics

- RMSE (Root Mean Square Error): Measures prediction accuracy.
- Directional Accuracy: Checks if model correctly predicts the price movement direction.

Justification for Final Model Choice

- LSTM captures long-term dependencies better than traditional regression models.
- Achieves lowest RMSE and highest directional accuracy.

Model Limitations & Potential Improvements

- Limitation: Model struggles with sudden market shifts.
- Improvement: Include external factors like news sentiment analysis.
- Limitation: Requires more computational power.
- Improvement: Optimize LSTM architecture using hyperparameter tuning.