BITCOIN PRICE AND TRADE STRATEGY PREDICTION

BAN 5600 - 01 Advanced Big Data Computing and Programming
Dr. Jin Fang

Karan Chaudhary
Sayara Malla Thakuri
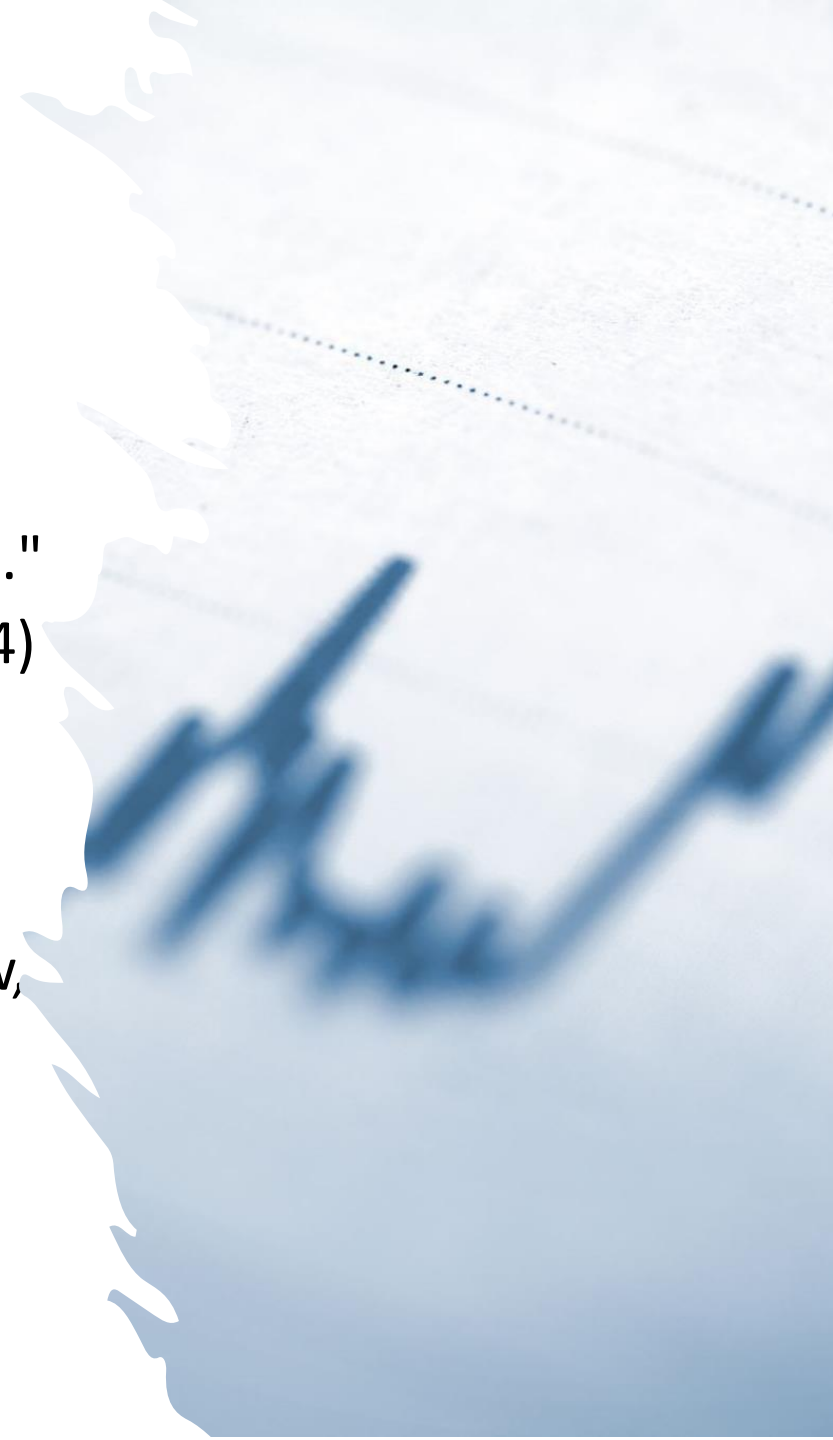**Sri Navya Guduru**
Swetha Pothu

REASEARCH PROBLEM

- **Predictive Complexity**:
  - "Challenges in forecasting Bitcoin prices due to extreme volatility and market unpredictability."

- **Data Analysis Challenges**:
  - "Limited historical data and the inadequacy of traditional financial models in predicting cryptocurrency trends."

- **Technological and Regulatory Dynamics**:
  - "Influences of evolving blockchain technology and diverse global regulations on Bitcoin's market behavior."

- **Strategy Formulation Difficulties**:
  - "Developing adaptable and resilient trading strategies to navigate the highly fluctuating Bitcoin market."

- **Real-Time Decision Making**:
  - "Need for advanced analytical tools to facilitate real-time decision making in a rapidly changing market environment."
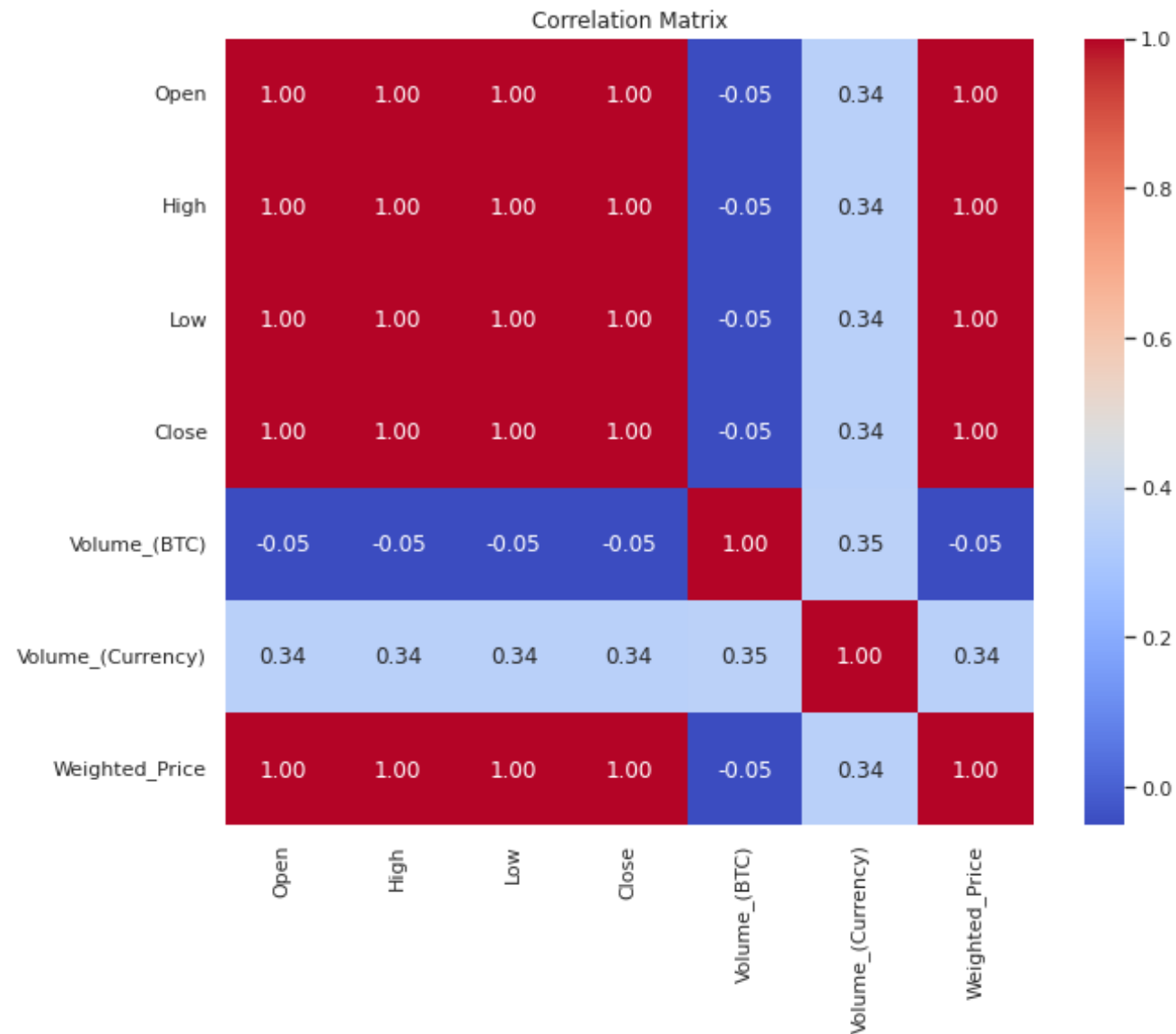
DATA DESCRIPTION

- **Data Source and Structure**:
  - "Sourced from Kaggle: CSV file capturing real-time Bitcoin trading activity."
  - "Covers January 2012 to March 2021."

- **Data Features**:
  - **Timestamps**: "Unix time at the start of each 60-second window."
  - **OHLC Data (Open, High, Low, Close)**: "Prices at the start, highest, lowest, and end of each time window."
  - **Volume Data**:
    - Volume (BTC): "Bitcoin volume within the time window."
    - Volume Currency: "Corresponding currency volume transacted."
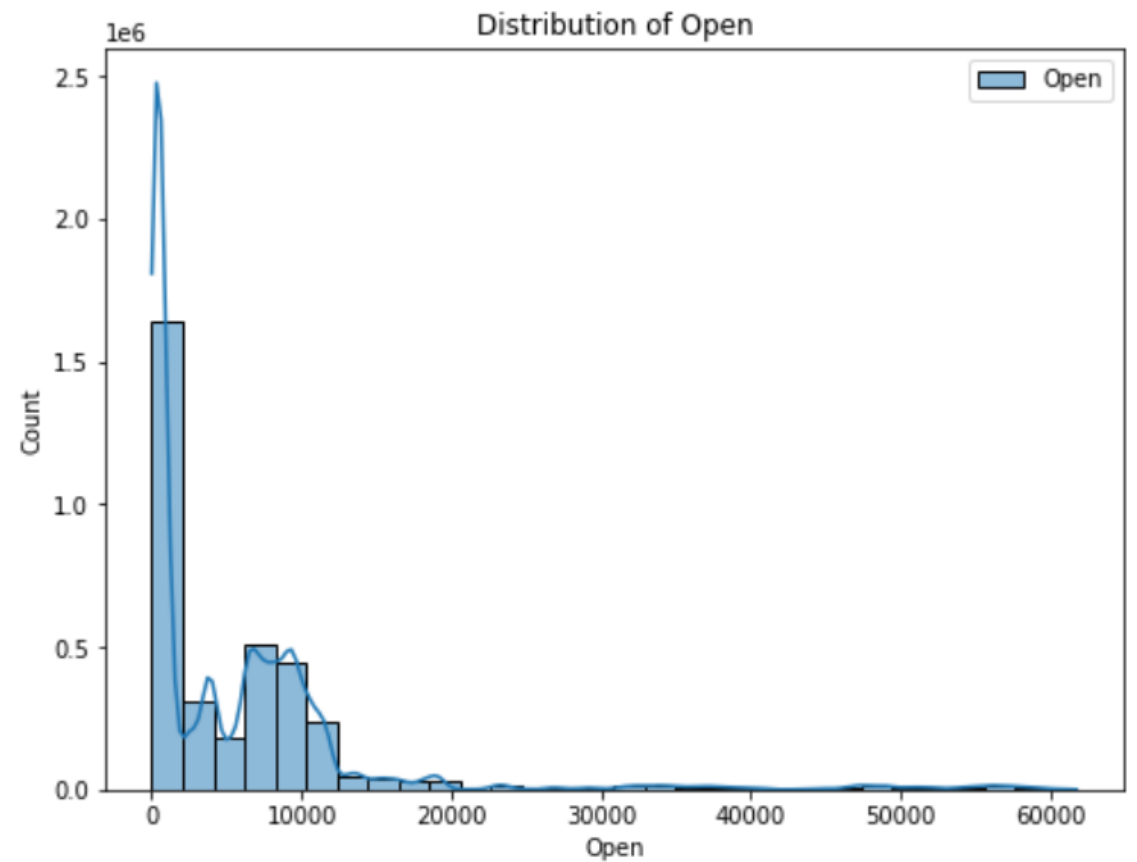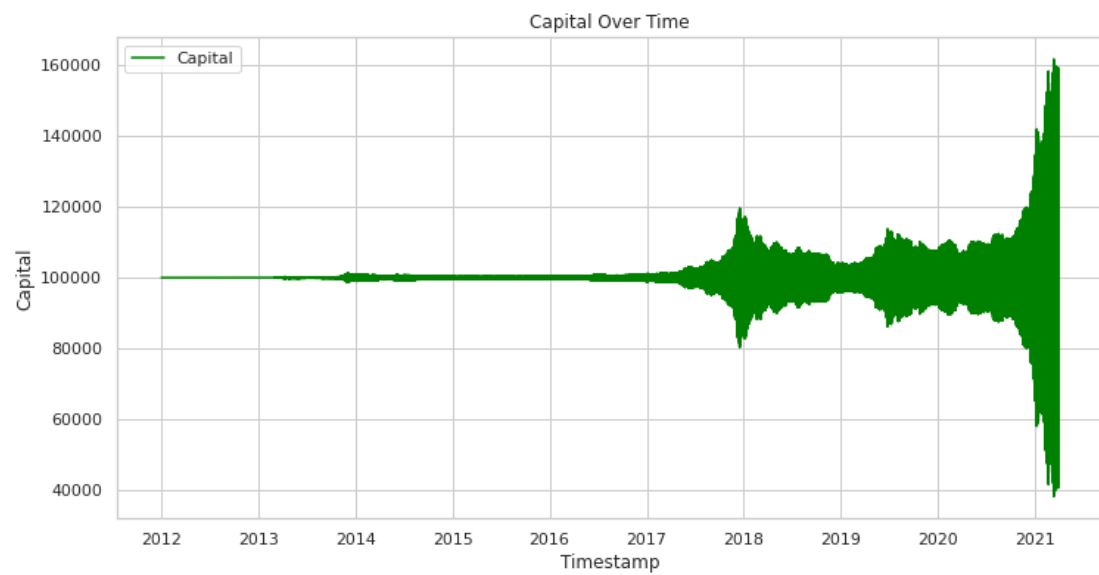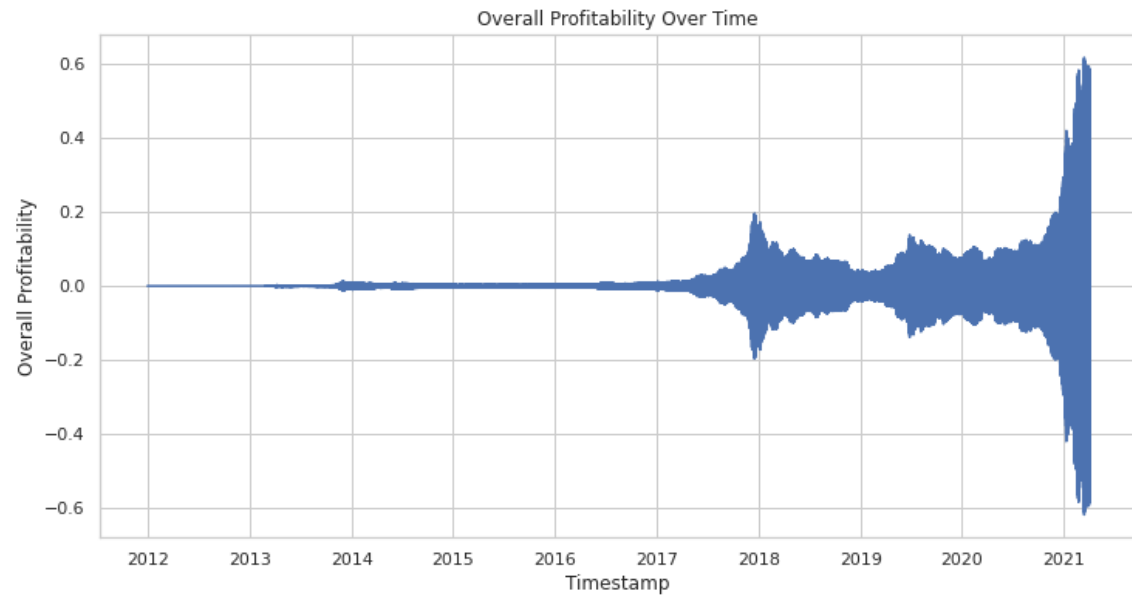 - **Weighted Bitcoin Price**: "Volume Weighted Average Price (VWAP)."

- **Data Integrity**:
- "Gaps filled with NaNs indicating no trading activity."
- "Deduplication efforts ensure data reliability, even in case of errors."
- **Data Analysis**:
- **DataFrame Details**: "Comprises 4,857,377 rows and 8 columns."
- **Data Types**: "Seven columns of floating-point numbers (float64) and one column of integers (int64)."
- **Memory Usage**: "Approximately 296.5 megabytes."
- **Statistical Summary**:
- **Mean** and **Median** values for each key metric (Open, High, Low, Close, Volume, etc.).
- "Standard Deviation to understand price volatility."
- "Minimum and Maximum values to gauge market range."
- "25th, 50th (Median), and 75th Percentiles for a detailed distribution insight."
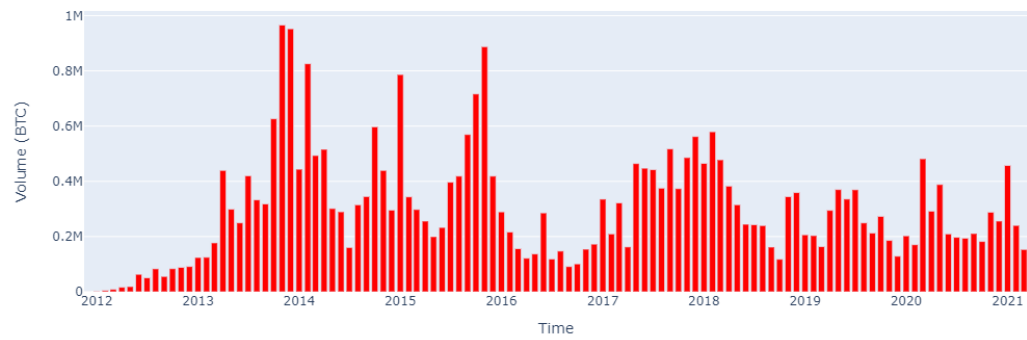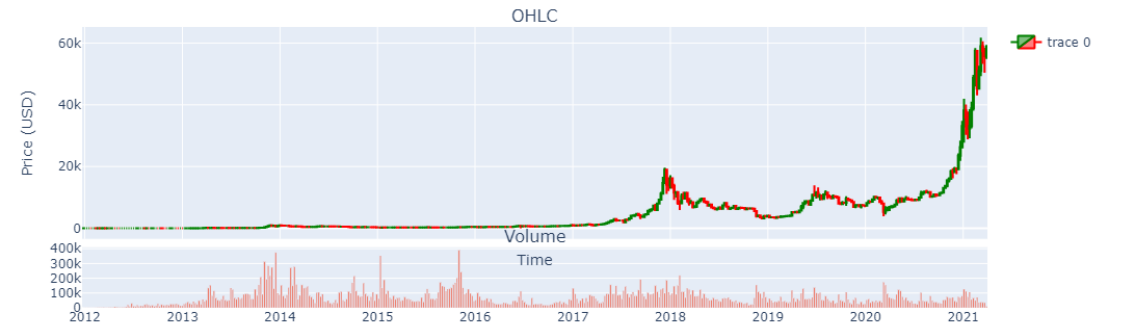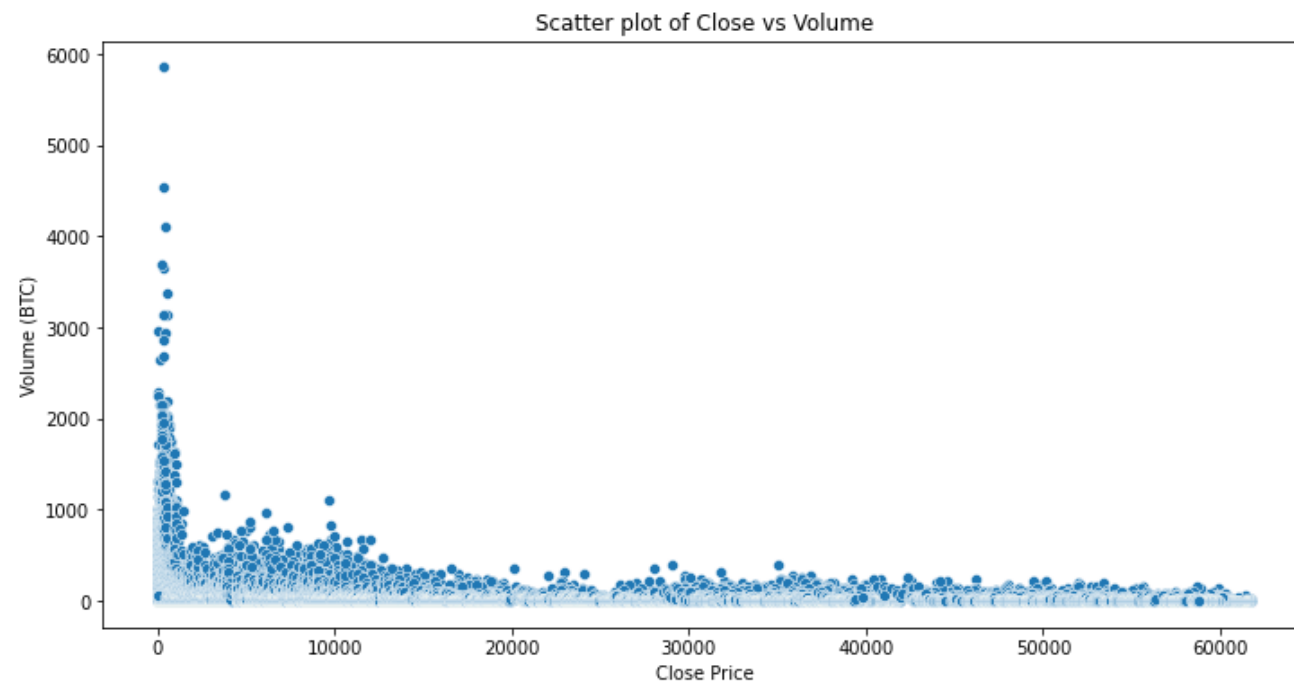
Descriptive Analysis

Overall Profitability Over Time

Capital Over Time

Distribution of Open

Bitcoin Monthly Trading Volume

Bitcoin Candlestick Chart with Volume

Scatter plot of Close vs Volume

Bitcoin Price and 5-Minute SMA

Time Series Plot of Close Prices

DATA MODELLING

- Close as Target Value

- Performed data scaling and split data into 80% & 20% for train and test data respectively

- Random Forest RMSE: 1529.9224460927794

- Gradient-Boosted Trees RMSE: 1177.5499589090648

- Linear Regression RMSE: 7.093107252725778

- Decision Tree Classifier Area Under ROC: 0.4598783833384746

- Best Model: RandomForestRegressionModel: uid=RandomForestRegressor_36eb5daa5027, numTrees=20, numFeatures=6
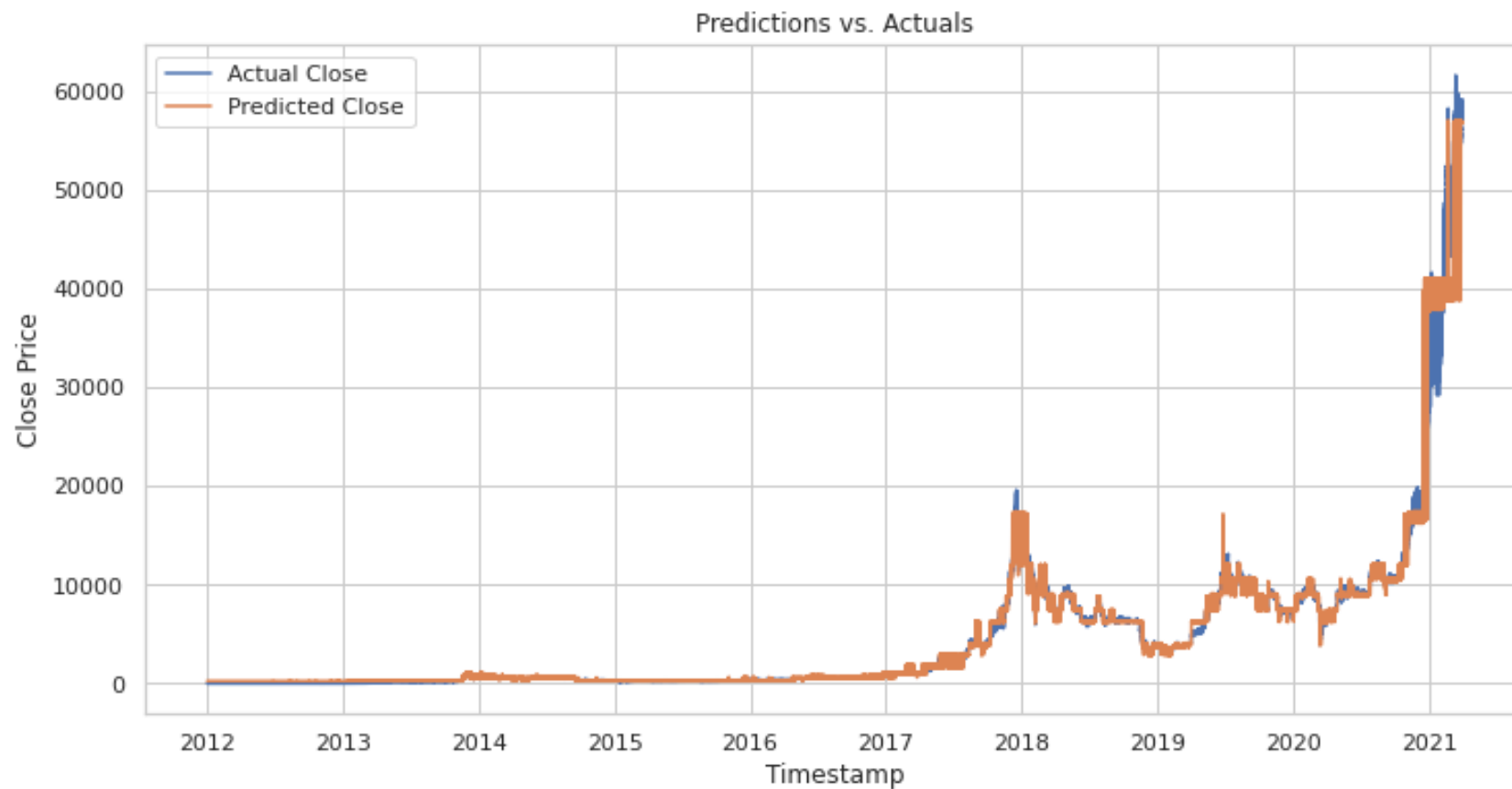
Predictive Analysis

SUMMARY

- Bitcoin's price has seen significant fluctuations, notably in 2018 and late 2020 notably the price saw considerable changes.

- Time Series Analysis reveals an overall rising trend in prices interspersed with irregular abnormalities.

- Apart from trade volumes in Bitcoin and currencies, most measured factors exhibit strong correlations.

- Based on moving averages, the short- and long-term price trends are comparable.

- RANDOM FOREST and linear regression models are being used to forecast future Bitcoin prices.

- Overlapping predictors and test data is a positive outcome, indicating that the RANDOM FOREST model is well fit to the time series data.

- The project's significance rests in its ability to use past data insights to support strategy creation, risk assessment, and decision-making in the unpredictable cryptocurrency market.

Thank You!!!