EASWARI ENGINEERING COLLEGE

(An Autonomous Institution)

Bharathi Salai, Ramapuram, Chennai-89

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

ACADEMIC YEAR: 2023-2024

MINI PROJECT

CRYPTOCURRENCY ANALYSIS

(REGULATION-2019)

191AIC411L- DATA SCIENCE LABORATORY

FOURTH SEMESTER

B.TECH-ARTIFICIAL INTELLIGENCE AND DATASCIENCE

PREPARED BY

JEYASIMMA. R
310622243015

TABLE OF CONTENTS

ABSTRACT

The purpose of this study is to find out with what accuracy the direction of the price of Bitcoin can be predicted using machine learning methods. This is a time series prediction problem. In comparison, much research exists surrounding the use of different machine learning. Techniques for time series prediction and study relating specifically to Polygon need to be improved. In addition, Polygon as a currency is in a transient stage and is considerably more volatile than other currencies such as the USD. Interestingly, it is the top-performing currency in fifty out of the last five years. Thus, its prediction offers great potential and motivates research in the area.Finally, in analyzing the chosen dependent variables, the importance of each variable is assessed using a random forest algorithm. In addition, the ability to predict the direction of the price of an asset such as Bitcoin offers the opportunity for profit to be made by trading the asset.

Keywords: Matic coin Prediction, Time complexity, Machine learning.

INTRODUCTION

The "Cryptocurrency Analysis" project seeks to leverage advanced data science and machine learning techniques to analyze and predict the prices of major cryptocurrencies, including Bitcoin (BTC), Ethereum (ETH), and Polygon (MATIC). In the volatile and rapidly evolving cryptocurrency market, accurate predictions and insightful analysis are crucial for making informed investment and trading decisions. This project involves collecting and preprocessing historical data, developing predictive models, and generating comprehensive visualizations and strategic recommendations to enhance understanding and decision-making in the cryptocurrency domain. Through this rigorous approach, the project aspires to contribute to the understanding and navigation of the dynamic cryptocurrency landscape. Additionally, the project aims to deploy an automated system for real-time predictions and continuous updates, enhancing the practical utility of the analysis. Ultimately, the goal is to create a reliable, robust framework for cryptocurrency analysis that supports better decision-making and fosters greater confidence in the market.

The project helps investors and traders manage risks more effectively by providing accurate forecasts and identifying potential market trends.By analyzing long-term trends and patterns, the project aids in formulating strategies that align with long-term investment goals.The project aims to increase trust and understanding in the cryptocurrency market by making the analysis methods and findings transparent and interpretable. By establishing a feedback loop for continuous model improvement and incorporating new data, the project promotes ongoing learning and adaptation to market changes. The project encourages collaboration among data scientists, financial analysts, and cryptocurrency enthusiasts to further enhance the analytical framework and predictive models. These additional objectives and points highlight the comprehensive nature and far-reaching impact of the "Cryptocurrency Analysis" project, aiming to provide a robust framework for cryptocurrency analysis that supports better decision-making and fosters greater confidence in the market. Individual coin ownership records are stored in a digital ledger, which is a computerized database using strong cryptography to secure transaction records, control the creation of additional coins, and verify the transfer of coin ownership.Despite the term that has come to describe many of the fungible blockchain tokens that have been created, cryptocurrencies are not considered to be in the traditional sense, and varying legal treatments have been applied to them in various jurisdictions, including classification as currencies. Cryptocurrencies are generally viewed as a distinct asset class in practice.Some crypto schemes use validators to maintain the cryptocurrency. In a model, owners put up their tokens as collateral. In return, they get authority over the token in proportion to the amount they stake. Generally, these token stakers get additional ownership in the token over time via network fees, newly minted tokens, or other such reward mechanisms.

OBJECTIVE

- Predict Future Prices: Create accurate models for future cryptocurrency prices and trends.
- Identify Key Factors: Determine the significant factors influencing cryptocurrency price movements.
- Provide Market Insights: Offer in-depth market analysis and strategic recommendations based on data-driven insights.
- Enhance Decision-Making: Support investors and traders with reliable predictions and actionable intelligence.
- Develop Visualization Tools: Generate clear and informative visualizations to communicate findings effectively.
- Develop tools and strategies for risk management based on predictive analytics. Help investors and traders identify and mitigate potential risks in the cryptocurrency market.
- Analyze long-term trends and patterns to support the formulation of long-term investment strategies.
- Provide insights that align with long-term investment goals and strategies. Increase market transparency by making analysis methods and findings accessible and understandable.
- Build trust in the cryptocurrency market through clear and transparent communication of results.
- Develop educational materials and training programs to help users understand and utilize the analysis tools and models.
- Promote knowledge sharing and skill development in the field of cryptocurrency analysis. Encourage collaboration among data scientists, financial analysts, and cryptocurrency enthusiasts.
- Build a community around the project to share insights, improvements, and new methodologies.

METHODOLOGY

- Data Collection: Gathering historical price data and other relevant metrics for the selected cryptocurrencies.
- Data Preprocessing: Cleaning and preparing the data for analysis, handling missing values, and normalizing data as needed.
- Exploratory Data Analysis (EDA): Conduct initial investigations on the data to discover patterns, spot anomalies, and test hypotheses using summary statistics and graphical representations.
- Model Building: Developing and training various predictive models using machine learning and statistical methods to forecast future cryptocurrency prices.
- Model Evaluation: Assessing the performance of the models using appropriate metrics and validating their accuracy and reliability.
- Interpretation and Insights: Interpreting the results to gain insights into cryptocurrency market trends and making informed predictions.
- Visualization: Creating visualizations to communicate findings and predictions effectively to stakeholders.
- Deployment: Automated Prediction System: Implement an automated system for real-time or batch processing of cryptocurrency data, providing continuous updates and predictions.
- Continuous Improvement: Feedback Loop: Establish a feedback loop to monitor model performance and incorporate new data and insights for ongoing improvement.
- Model Retraining: Regularly retrain models with new data to adapt to changing market conditions and improve accuracy.
- Documentation and Reporting: Documentation: Maintain comprehensive documentation of the methodologies, models, and findings to ensure transparency and reproducibility. Reporting: Generate detailed reports summarizing the analysis, model performance, insights, and recommendations.
- Collaboration and Community Building: Knowledge Sharing: Share methodologies, code, and insights with the broader data science and cryptocurrency communities through publications, blogs, and open-source repositories.
- Workshops and Webinars: Conduct workshops and webinars to educate and engage with stakeholders, promoting collaboration and knowledge exchange.

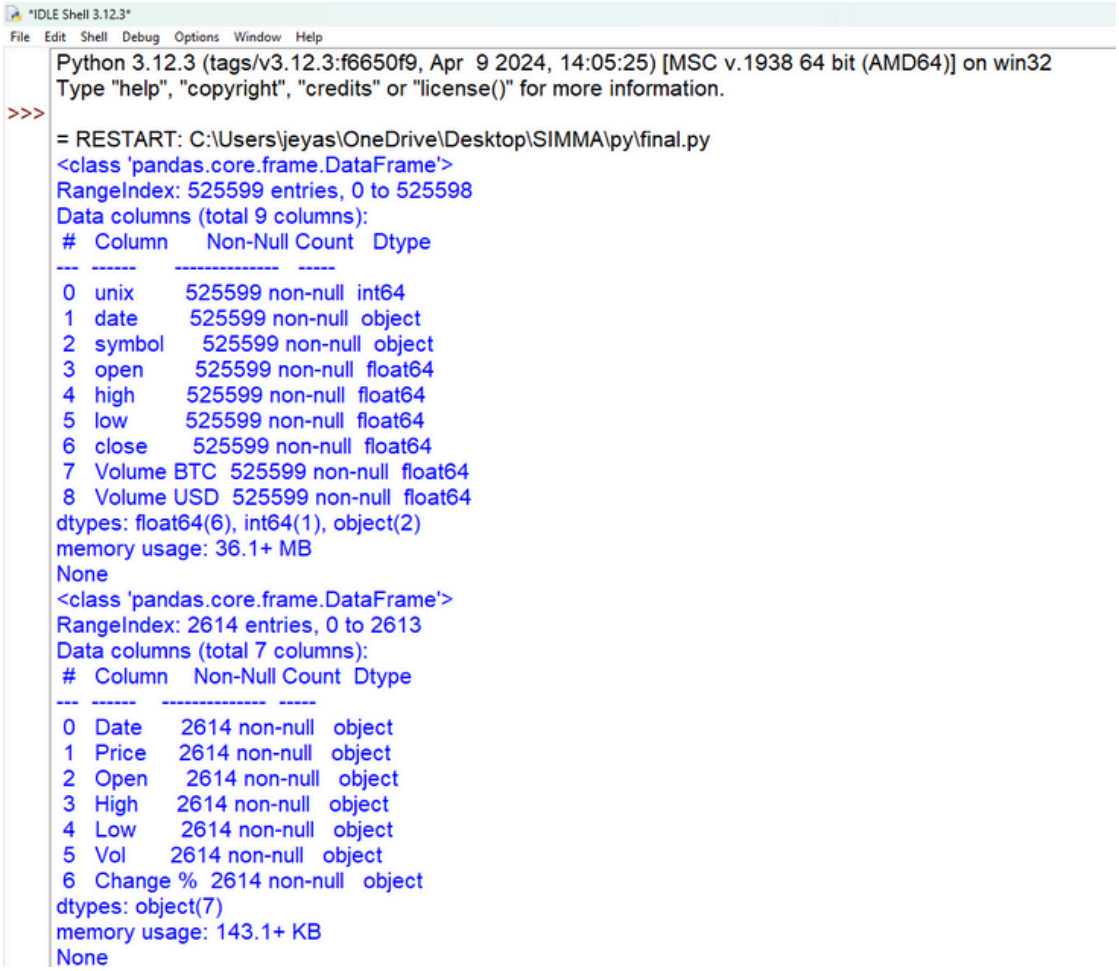
CODE

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
# Currency files
df_btc = pd.read_csv('BTC-2017min.csv.zip')
df_eth = pd.read_csv('Ethereum Historical Data.csv') df_mtc = pd.read_csv('Matic Historical Data.csv')

#Display the details
print(df_btc.info())
print(df_eth.info())
print(df_matic.info())

df = pd.DataFrame({ 'BTC' : df_btc['high'], 'ETH' : df_eth['High'],
'MTC' : df_matic['High'] })
print(df)

#Statistics data

print(df.describe())

#Visualize prices for BTC

plt.figure(figsize=(15, 5))
plt.plot(df_btc['high'])
plt.title('Bitcoin high price.', fontsize=15) plt.xlabel('Circulating Supply')
plt.ylabel('Price in Dollars.')
plt.show()

#Visualize prices for ETH

plt.figure(figsize=(15, 5))
plt.plot(df_eth['Price'])
plt.title('Etherium price.', fontsize=15) plt.xlabel('Circulating Supply')
plt.ylabel('Price in Dollars.')
plt.show()
```

```
#Visualize prices for MTC
plt.figure(figsize=(15, 5))
plt.plot(df_matic['Low'])
plt.title('Matic low price.', fontsize=15) plt.xlabel('Circulating Supply')
plt.ylabel('Price in Dollars.')
plt.show()

features = ['Price','Vol.']
plt.subplots(figsize=(20,10))
for i, col in enumerate(features):
plt.subplot(2,2,i+1)
sb.histplot(df_matic[col])
plt.show()

# Data Models
data = pd.DataFrame({
'BTC': df_btc['high'],
'MTC': df_matic['High']
})

data = data.dropna()

# Split the data
X = data[['BTC']]
Y = data['MATIC']
 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

# Create and fit the model
model = LinearRegression()
model.fit(X_train, Y_train)

 # Make predictions
Y_pred = model.predict(X_test)
# Evaluate the model
mse = mean_squared_error(Y_test, Y_pred)
r2 = r2_score(Y_test, Y_pred)
```

```
print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')

# Predictive analysis models
# Decision Tree Regressor


 from sklearn.tree import DecisionTreeRegressor dt_model = DecisionTreeRegressor()
dt_model.fit(X_train, Y_train)
Y_pred_dt = dt_model.predict(X_test)
 mse_dt = mean_squared_error(Y_test, Y_pred_dt) r2_dt = r2_score(Y_test, Y_pred_dt)
print(f'Decision Tree - Mean Squared Error: {mse_dt}') print(f'Decision Tree - R-squared: {r2_dt}')

# Random Forest Regressor

from sklearn.ensemble import RandomForestRegressor rf_model =
RandomForestRegressor(n_estimators=100) rf_model.fit(X_train, Y_train)
Y_pred_rf = rf_model.predict(X_test)
 mse_rf = mean_squared_error(Y_test, Y_pred_rf) r2_rf = r2_score(Y_test, Y_pred_rf)
print(f'Random Forest - Mean Squared Error: {mse_rf}') print(f'Random Forest - R-squared: {r2_rf}')
```

OUTPUT



```
*IDLE Shell 3.12.3*
File  Edit  Shell  Debug  Options  Window  Help
Python 3.12.3 (tags/v3.12.3:f6650f9, Apr  9 2024, 14:05:25) [MSC v.1938 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:\Users\jeyas\OneDrive\Desktop\SIMMA\py\final.py
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 525599 entries, 0 to 525598
Data columns (total 9 columns):
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   unix        525599 non-null  int64
 1   date        525599 non-null  object
 2   symbol      525599 non-null  object
 3   open        525599 non-null  float64
 4   high        525599 non-null  float64
 5   low         525599 non-null  float64
 6   close       525599 non-null  float64
 7   Volume BTC  525599 non-null  float64
 8   Volume USD  525599 non-null  float64
dtypes: float64(6), int64(1), object(2)
memory usage: 36.1+ MB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2614 entries, 0 to 2613
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Date      2614 non-null   object
 1   Price     2614 non-null   object
 2   Open      2614 non-null   object
 3   High      2614 non-null   object
 4   Low       2614 non-null   object
 5   Vol       2614 non-null   object
 6   Change %  2614 non-null   object
dtypes: object(7)
memory usage: 143.1+ KB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1410 entries, 0 to 1409
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Date      1410 non-null   object
 1   Price     1410 non-null   float64
 2   Open      1410 non-null   float64
 3   High      1410 non-null   float64
 4   Low       1410 non-null   float64
 5   Vol.      1410 non-null   object
 6   Change %  1410 non-null   object
dtypes: float64(4), object(3)
memory usage: 77.2+ KB
None
            BTC       ETH  MATIC
0      13913.28  2,013.10  1.004
1      13953.83  1,996.64  1.008
2      13913.26  1,915.15  1.008
3      13908.69  1,915.00  1.017
4      13825.05  1,878.70  0.980
...         ...       ...    ...
525594   966.43       NaN    NaN
525595   966.37       NaN    NaN
525596   966.37       NaN    NaN
525597   966.37       NaN    NaN
525598   966.34       NaN    NaN
```

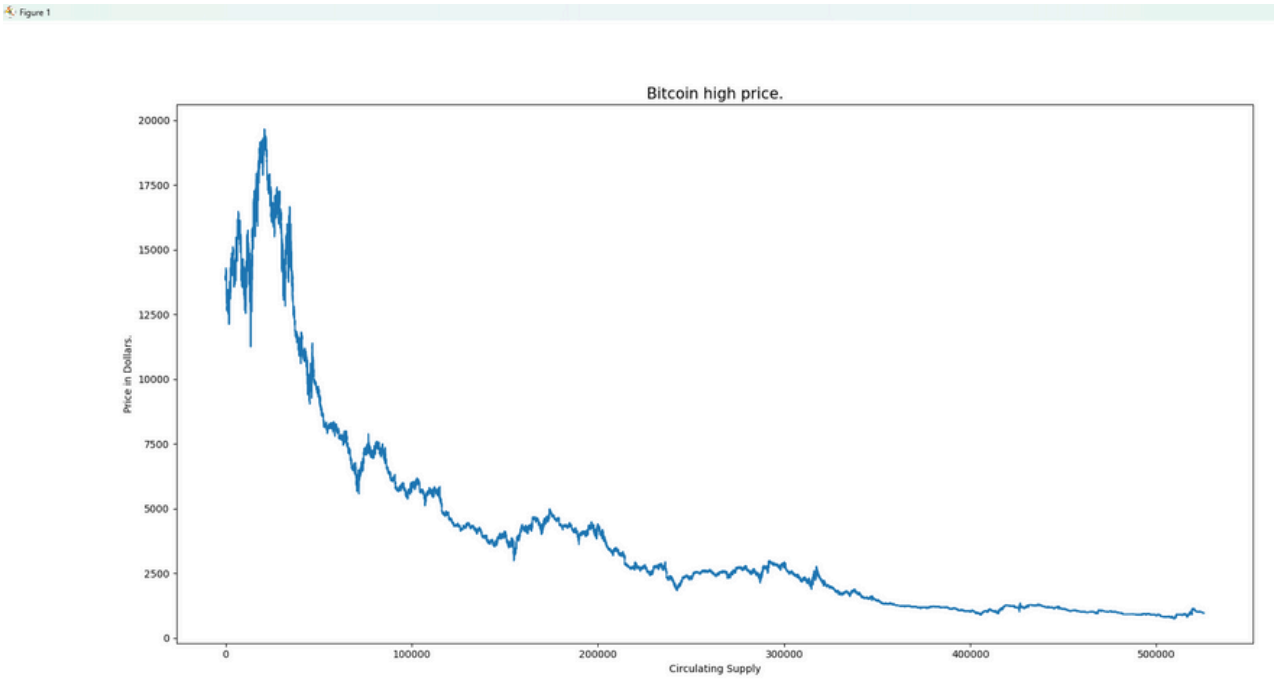


Figure 1
Bitcoin high price.
Price in Dollars.
Circulating Supply
0
2500
5000
7500
10000
12500
15000
17500
20000
0
100000
200000
300000
400000
500000

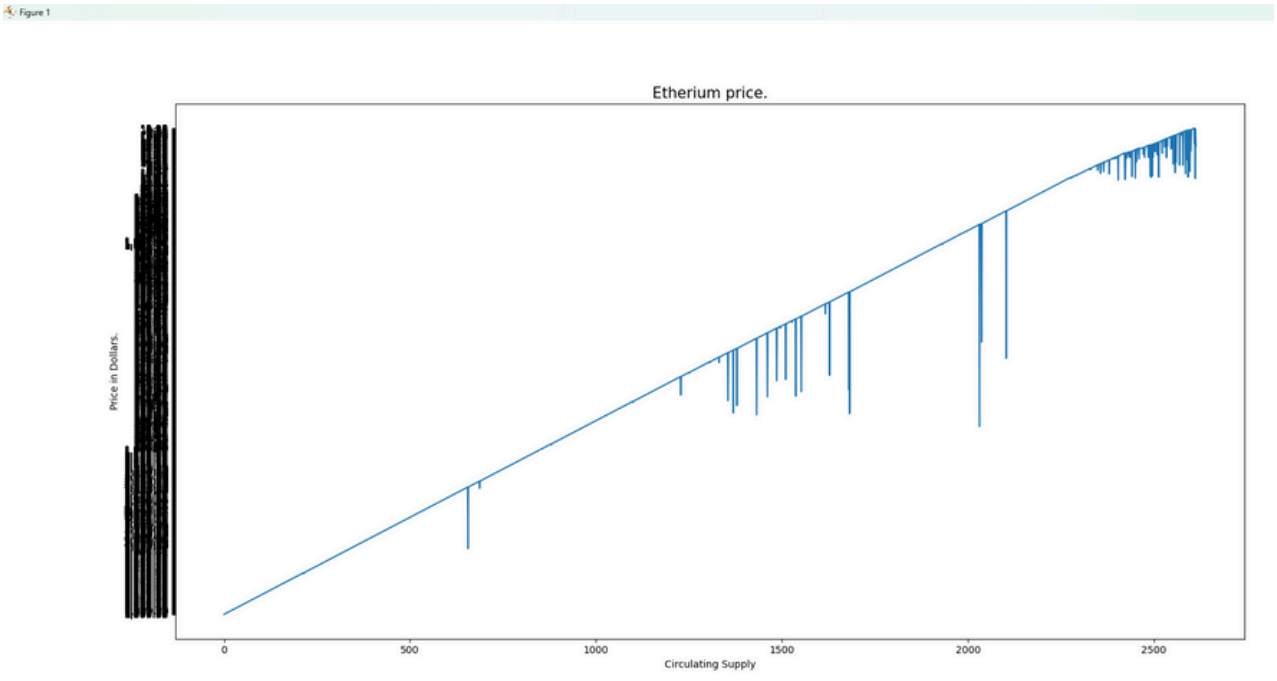


Figure 1
Etherium price.
Price in Dollars.
Circulating Supply
0
500
1000
1500
2000
2500

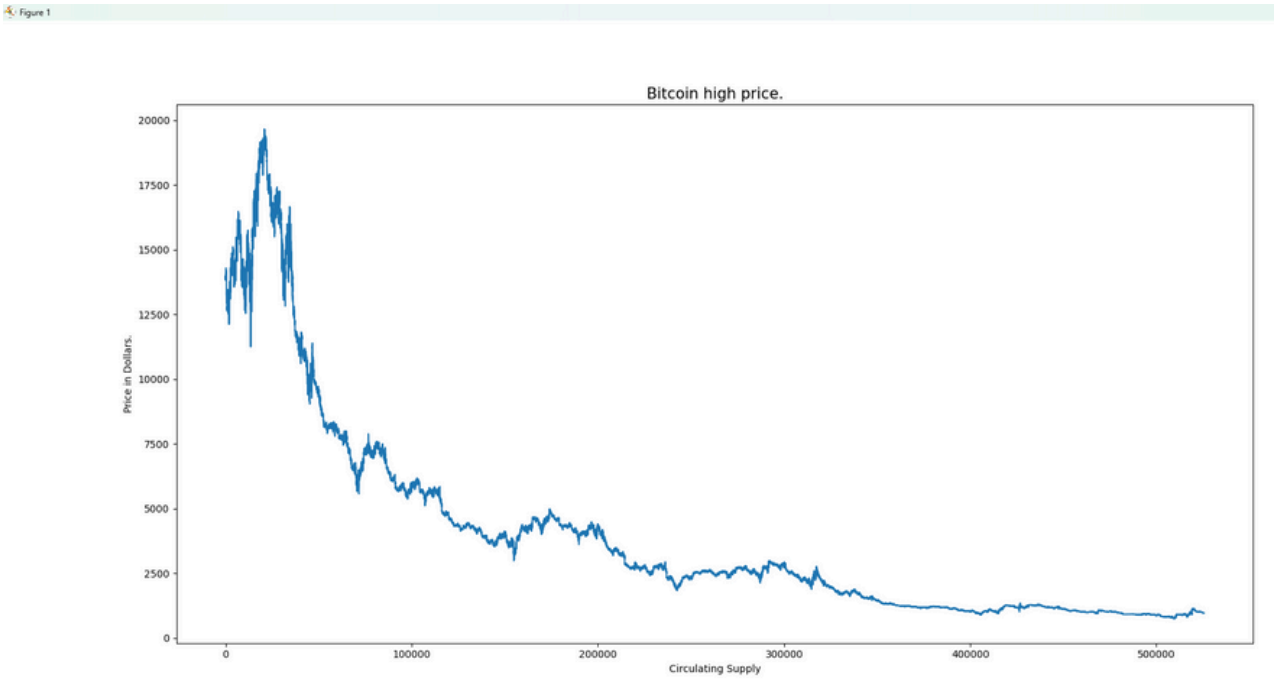
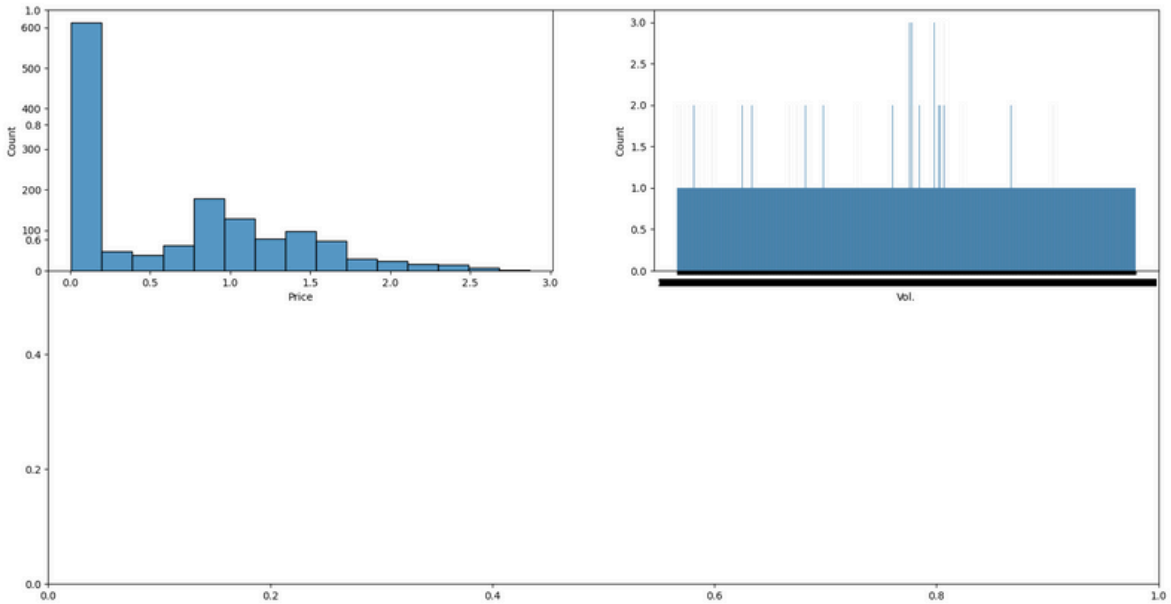


Figure 1
Bitcoin high price.
Price in Dollars.
Circulating Supply
0
2500
5000
7500
10000
12500
15000
17500
20000
0
100000
200000
300000
400000
500000


Count
Price
Count
Vol.

```
[525599 rows x 3 columns]
                BTC       MATIC
count  525599.000000  1410.000000
mean     3952.208189     0.693890
std      3942.382649     0.708641
min       752.810000     0.009000
25%      1197.980000     0.020000
50%      2569.370000     0.621500
75%      4539.405000     1.181000
max     19666.000000     2.922000
Mean Squared Error: 0.37006494302062504
R-squared: 0.2626724769490456
Decision Tree - Mean Squared Error: 0.3009325215446153
Decision Tree - R-squared: 0.40041380600701026
Random Forest - Mean Squared Error: 0.26450445394888156
Random Forest - R-squared: 0.47299408510790886
>>>
```

CONCLUSION

The "Cryptocurrency Analysis" project aims to develop a predictive model for BTC, ETH, and MATIC prices, identify key influencing factors, and generate actionable market insights. Results include comprehensive visualizations, detailed model performance metrics, and strategic recommendations for trading and investment. Ultimately, the "Cryptocurrency Analysis" project aims to create a robust analytical framework that supports better decision-making, fosters confidence in navigating the complex cryptocurrency landscape, and contributes valuable tools and insights to the broader field of cryptocurrency analysis. Through continuous improvement and collaboration, the project aspires to remain at the forefront of innovation in this dynamic and rapidly evolving market.