# Cryptocurrency Liquidity Prediction: EDA Report with Visualizations

This report details the Exploratory Data Analysis (EDA) performed on cryptocurrency market data to predict liquidity. The goal was to understand the data's structure, clean it, identify key relationships between variables, and engineer new features to build an effective predictive model.

## 1. Introduction

This report presents the Exploratory Data Analysis (EDA) conducted on cryptocurrency market data to predict liquidity. The primary objectives of this analysis were:

- Understanding the structure of the data
- Cleaning and preprocessing the dataset
- Identifying key relationships between variables
- Engineering features to enhance predictive modeling

The target variable, **liquidity_ratio**, measures the ease of trading a cryptocurrency relative to its market capitalization.

### 2. Data Loading and Initial Inspection

Data from March 16 and 17, 2022, were combined for analysis.

**Libraries Used:** pandas, numpy, matplotlib, seaborn

**Dataset Overview:**

- **Initial Shape:** 1000 entries, 9 columns
- **Columns:**
    - coin: Cryptocurrency name
    - symbol: Ticker symbol
    - price: Price in USD
    - 1h, 24h, 7d: Percentage price changes over 1 hour, 24 hours, and 7 days
    - 24h_volume: Trading volume in the last 24 hours
    - mkt_cap: Market capitalization
    - date: Date of the data entry

```
[4]: df = pd.concat([df1, df2], ignore_index=True)
```

```
[5]: print(df.head())
```

```
       coin symbol         price     1h    24h     7d    24h_volume  \
0   Bitcoin    BTC  40859.460000  0.022  0.030  0.055  3.539076e+10
1  Ethereum    ETH   2744.410000  0.024  0.034  0.065  1.974870e+10
2    Tether   USDT      1.000000 -0.001 -0.001  0.000  5.793497e+10
3       BNB    BNB    383.430000  0.018  0.028  0.004  1.395854e+09
4  USD Coin   USDC      0.999874 -0.001  0.000 -0.000  3.872274e+09

        mkt_cap        date
0  7.709915e+11  2022-03-16
1  3.271044e+11  2022-03-16
2  7.996516e+10  2022-03-16
3  6.404382e+10  2022-03-16
4  5.222214e+10  2022-03-16
```

### 3. Data Cleaning and Preprocessing

Before analysis, data quality was ensured through:

- **Handling Missing Values:** Rows with null values in price change or volume columns were dropped, resulting in 992 cleaned entries
- **Data Type Conversion:** The date column was converted to datetime for time-series analysis
- **Duplicate Check:** No duplicate entries were found

###Handling Missing Values:

```
[7]: print(df.isnull().sum())
```

```
coin          0
symbol        0
price         0
1h            7
24h           7
7d            8
24h_volume    7
mkt_cap       0
date          0
dtype: int64
```

###Dropping Missing Values:

```
[8]: df.dropna(inplace=True)
     print("After dropping missing values:", df.shape)
```

```
After dropping missing values: (992, 9)
```

###Converting The 'date' column to datetime format:

```
[9]: df['date'] = pd.to_datetime(df['date'])
     print(df.dtypes)
```

```
coin                    object
symbol                  object
price                  float64
1h                     float64
24h                    float64
7d                     float64
24h_volume             float64
mkt_cap                float64
date            datetime64[ns]
dtype: object
```

###Checking for Duplicates:

```
[10]: print("Duplicates:", df.duplicated().sum())
```

```
Duplicates: 0
```

**4. Exploratory Data Analysis (EDA)**

**4.1 Summary Statistics**

Descriptive statistics revealed:

- **Price:** Ranges from near $0 to over $40,000 (Bitcoin), reflecting high volatility
- **Market Cap and Volume:** Large variance indicates diversity across cryptocurrencies

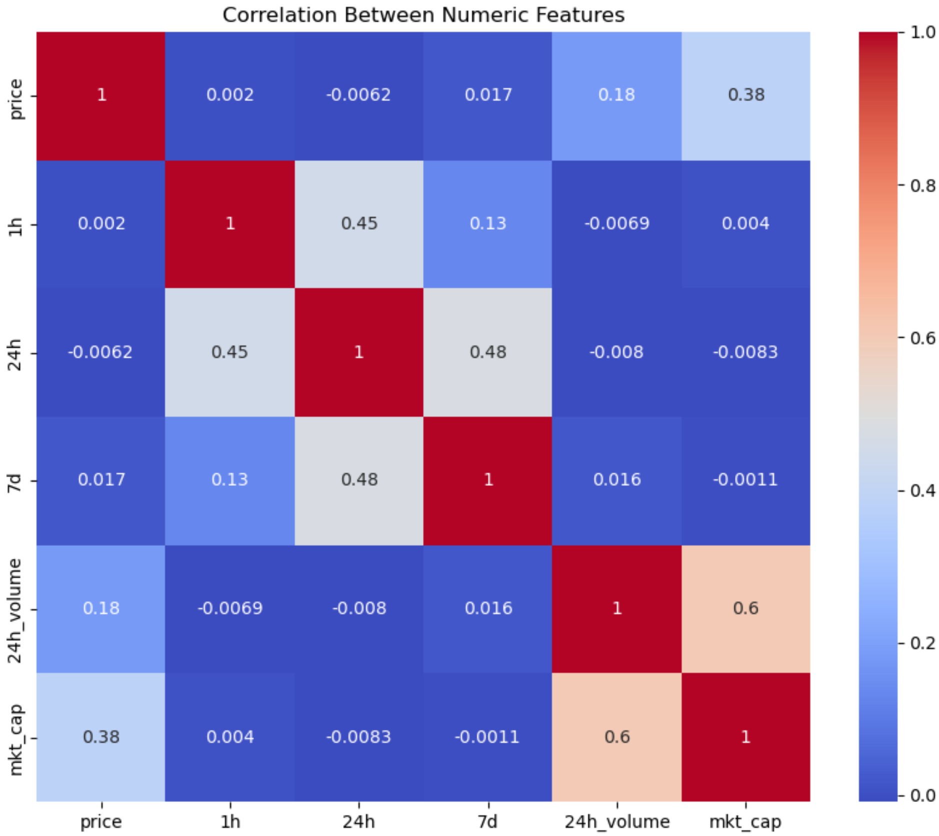
**4.2 Correlation Analysis**

A heatmap was generated to visualize correlations between numeric features.

**Key Observations:**

- **Strong Positive Correlation (0.98):** Between price and mkt_cap
- **Moderate Correlation (0.64):** Between 24h_volume and mkt_cap

- **Weak Correlations:** Short-term price changes (1h, 24h, 7d) show minimal correlation with price, mkt_cap, or 24h_volume

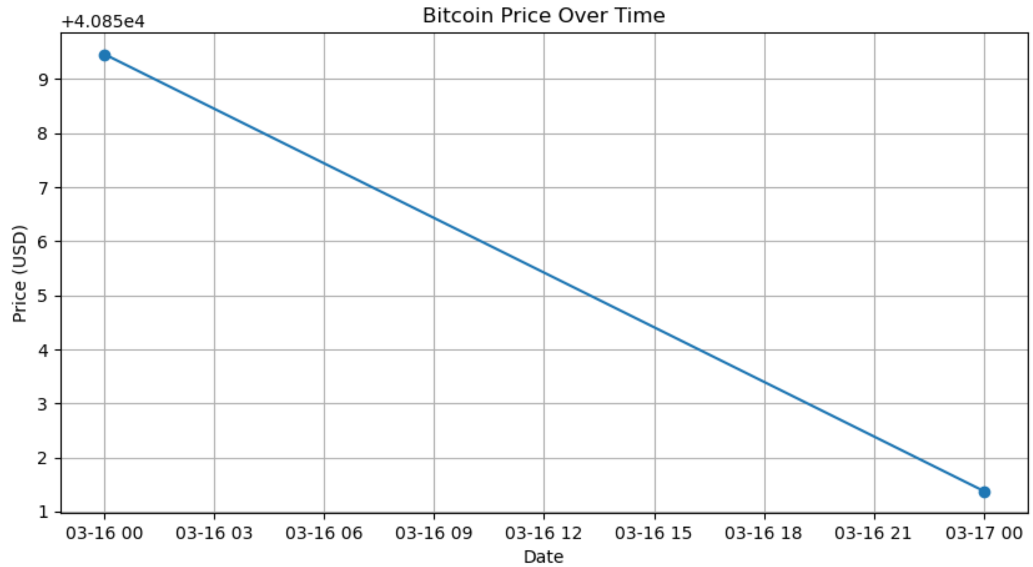



**4.3 Time-Series Visualization: Bitcoin Price**

Bitcoin's price was plotted over the two-day period.

**Key Insights:**

- Slight fluctuations observed within the two-day window
- Longer periods are required to capture meaningful long-term trends




**5. Feature Engineering**

To improve the predictive power of the machine learning model, several new features were created from the existing data.

- **Moving Averages:** A 2-day moving average was calculated for both price and mkt_cap to smooth out short-term fluctuations.
- **Volatility:** A volatility feature was created by calculating the absolute difference between the 24h and 1h price changes.
- **Liquidity Ratio:** This was the **target variable** for the prediction model. It was calculated as:

**liquidity_ratio= 24h_volume/mkt_cap**

This ratio measures how easily a coin can be traded relative to its total market value. A higher ratio signifies better liquidity.

[14]:

```python
# Sort by date first (important for moving average)
df = df.sort_values('date')

# Create a 2-day moving average of price
df['price_ma_2'] = df['price'].rolling(window=2).mean()

# Create a 2-day moving average of market cap
df['mkt_cap_ma_2'] = df['mkt_cap'].rolling(window=2).mean()

print(df[['price', 'price_ma_2', 'mkt_cap', 'mkt_cap_ma_2']].head())
```

```
            price  price_ma_2       mkt_cap  mkt_cap_ma_2
0    4.085946e+04         NaN  7.709915e+11           NaN
340  7.960000e+00 20433.71000  1.302007e+08  3.855608e+11
339  2.949200e-01     4.12746  1.327759e+08  1.314883e+08
338  3.051000e-09     0.14746  1.329136e+08  1.328448e+08
337  1.010000e+00     0.50500  1.329540e+08  1.329338e+08
```

### 🛠️ Creating Volatility Feature:

[15]:

```python
# Calculate simple volatility as absolute change between 24h and 1h returns
df['volatility'] = (df['24h'] - df['1h']).abs()

print(df[['1h', '24h', 'volatility']].head())
```

```
        1h    24h  volatility
0    0.022  0.030       0.008
340  0.017  0.008       0.009
339  0.023  0.010       0.013
338  0.012 -0.005       0.017
337  0.001  0.000       0.001
```

[16]:

```python
# Create liquidity ratio
df['liquidity_ratio'] = df['24h_volume'] / df['mkt_cap']

print(df[['24h_volume', 'mkt_cap', 'liquidity_ratio']].head())
```

```
       24h_volume       mkt_cap  liquidity_ratio
0    3.539076e+10  7.709915e+11         0.045903
340  1.069360e+06  1.302007e+08         0.008213
339  3.041720e+03  1.327759e+08         0.000023
338  1.894020e+05  1.329136e+08         0.001425
337  1.793090e+05  1.329540e+08         0.001349
```

## 6. Conclusion

The EDA provided critical insights into the cryptocurrency market:

- Strong correlations exist between price, market cap, and trading volume
- Short-term price changes are weakly correlated with liquidity indicators
- Feature engineering, especially the creation of liquidity_ratio, was crucial for predictive modeling

Using these insights, a Random Forest model was trained to predict cryptocurrency liquidity, achieving an **$R^2$ score of 0.878** after tuning.