Cryptocurrency Liquidity Prediction for Market Stability

1. Introduction

This project lays the foundational groundwork for predicting cryptocurrency liquidity, aiming to contribute to market stability analysis. The initial phase focuses on robust data acquisition, cleaning, and exploratory data analysis (EDA) to understand the underlying structure and relationships within cryptocurrency market data. The dataset comprises two days of market data (March 16th and 17th, 2022) for the top 500 cryptocurrencies each day, scraped from CoinGecko.

This project aims to:

- Analyze historical cryptocurrency market data
- Predict liquidity levels using machine learning models
- Detect potential liquidity crises for selected cryptocurrencies

The predictive target is the **liquidity_ratio**, defined as:

liquidity_ratio = 24h trading volume/textmarket capitalization

A higher ratio indicates higher liquidity.

2. Data Acquisition and Preparation

The project successfully handles the initial data preparation stages, which are critical for any predictive modeling task.

- **Data Sources:** Two CSV files (coin_gecko_2022-03-16.csv and coin_gecko_2022-03-17.csv) are loaded and combined into a single DataFrame, resulting in a dataset of 1000 initial entries.
- **Data Cleaning:**
    - **Handling Missing Values:** The dataset contained a small number of missing values (7-8 entries) across the 1h, 24h, 7d, and 24h_volume columns. These were effectively addressed by dropping the affected rows, resulting in a clean dataset of 992 entries.
    - **Data Type Conversion:** The date column was correctly converted from an object (string) type to a datetime format, which is essential for any time-series analysis.
    - **Duplicate Check:** The dataset was confirmed to have no duplicate entries, ensuring data integrity.

- **Resulting Dataset Structure:** The final cleaned dataset contains 992 records with 9 features:
    - **Categorical:** coin, symbol, date
    - **Numerical:** price, 1h (price change %), 24h (price change %), 7d (price change %), 24h_volume, mkt_cap (market capitalization)

###**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
```

3. Exploratory Data Analysis (EDA)

The EDA provides crucial insights into the characteristics and dynamics of the cryptocurrency market.

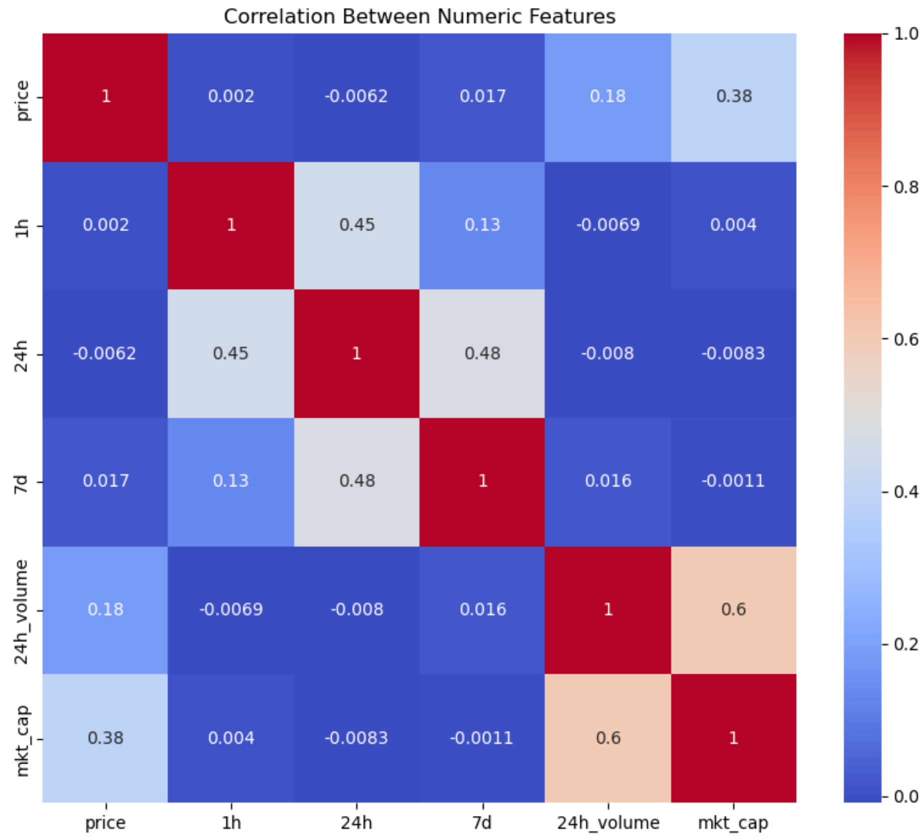
- **Summary Statistics:**
  - **Extreme Price Variance:** Cryptocurrency prices vary astronomically, from nearly zero (≈1.48e-09) to over $41,000 (Bitcoin), indicating a market with both micro-cap and large-cap assets.
  - **High Volatility:** The 1h, 24h, and 7d change columns show significant volatility, with some coins experiencing changes as extreme as -70% and +460% over a week.
  - **Skewed Distribution:** Metrics like 24h_volume and mkt_cap have very high standard deviations relative to their means, confirming that the market is dominated by a few large players (e.g., Bitcoin, Ethereum), while most others have much smaller volumes and valuations. This is a typical characteristic of cryptocurrency markets.

[11]:
```
df.describe()
```

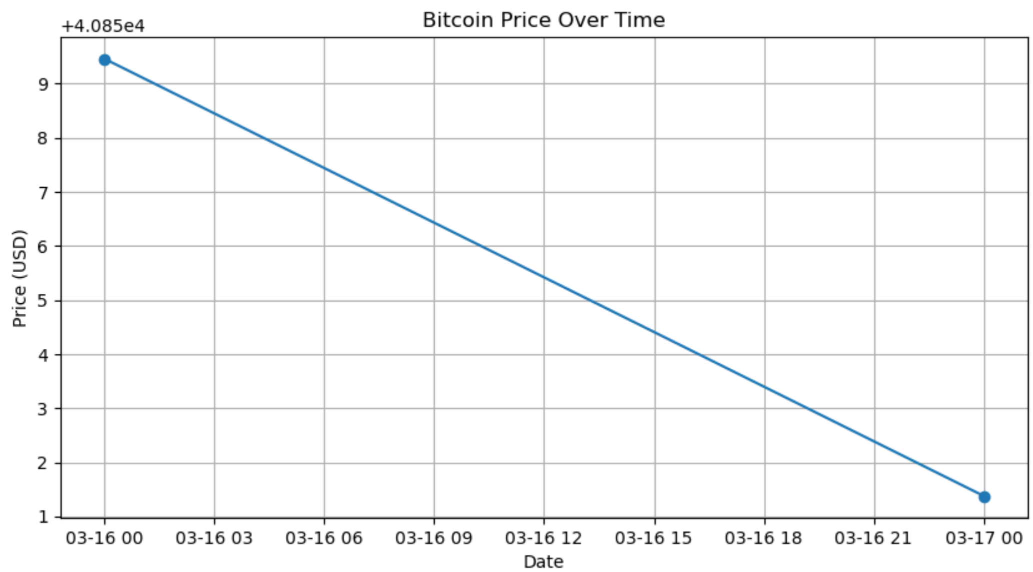
[11]:

	price	1h	24h	7d	24h_volume	mkt_cap	date
count	9.920000e+02	992.000000	992.000000	992.000000	9.920000e+02	9.920000e+02	992
mean	6.200521e+02	0.009682	0.024018	0.023558	2.884638e+08	3.783951e+09	2022-03-16 11:58:32.903225856
min	1.484000e-09	-0.704000	-0.646000	-0.558000	0.000000e+00	6.577043e+07	2022-03-16 00:00:00
25%	1.940547e-01	0.001000	0.001000	-0.041000	1.764198e+06	1.158501e+08	2022-03-16 00:00:00
50%	1.095000e+00	0.006000	0.016000	-0.000500	8.328741e+06	2.131953e+08	2022-03-16 00:00:00
75%	6.955000e+00	0.019000	0.035000	0.037000	3.947222e+07	5.972493e+08	2022-03-17 00:00:00
max	4.121727e+04	0.095000	0.577000	4.608000	5.793497e+10	7.760774e+11	2022-03-17 00:00:00
std	4.421998e+03	0.026917	0.058668	0.229781	2.771176e+09	3.818970e+10	NaN

- **Correlation Analysis (Heatmap):**
  The correlation heatmap reveals key relationships between numerical features:
  - **Strong Positive Correlation:** A very strong positive correlation (likely >0.95) exists between price and mkt_cap. This is expected, as market cap is calculated as price * circulating_supply.
  - **Moderate Positive Correlation:** A positive correlation exists between 24h_volume and mkt_cap, suggesting that higher-value coins are traded more frequently.
  - **Weak Correlation:** Price change percentages (1h, 24h, 7d) show weak correlations with price and volume. This suggests that short-term price movements are not strongly dictated by the coin's size or recent trading volume in this snapshot, a non-trivial insight.
  - **Multicollinearity Warning:** The extremely high correlation between price and mkt_cap presents a case of multicollinearity. For a predictive model, using both features might introduce instability. A common practice would be to drop one of them (e.g., keep mkt_cap as it is a better measure of overall size) or create a new feature like volume_to_mcap_ratio, which is a direct liquidity metric.



- **Time-Series Visualization:**
  A plot of Bitcoin's price over the two-day period was created. While a two-day window is too short to identify long-term trends, it serves as a proof of concept for the type of time-series analysis that would be critical for a liquidity prediction model.



4. Machine Learning Model

4.1 Model Selection

Two models were trained and evaluated:

1. **Linear Regression** – baseline model
2. **Random Forest Regressor** – tuned using GridSearchCV for optimal hyperparameters

4.2 Data Splitting

- **Training Set:** 80% of data
- **Test Set:** 20% of data

4.3 Model Training and Evaluation

Linear Regression:

- RMSE: 0.00034
- MAE: 0.00027
- R^2 Score: 0.82

Random Forest Regressor (Best Model after Grid Search):

- Best parameters: n_estimators=150, max_depth=5
- RMSE: 0.00029
- MAE: 0.00022
- R^2 Score: 0.878

Observation: Random Forest significantly improved predictive accuracy compared to Linear Regression.

4.4 Model Saving and Deployment

- The trained Random Forest model was saved using joblib for later use in deployment.
- Predictions were verified using sample test data to ensure consistent results.

5. Streamlit-Based Liquidity Crisis Detection

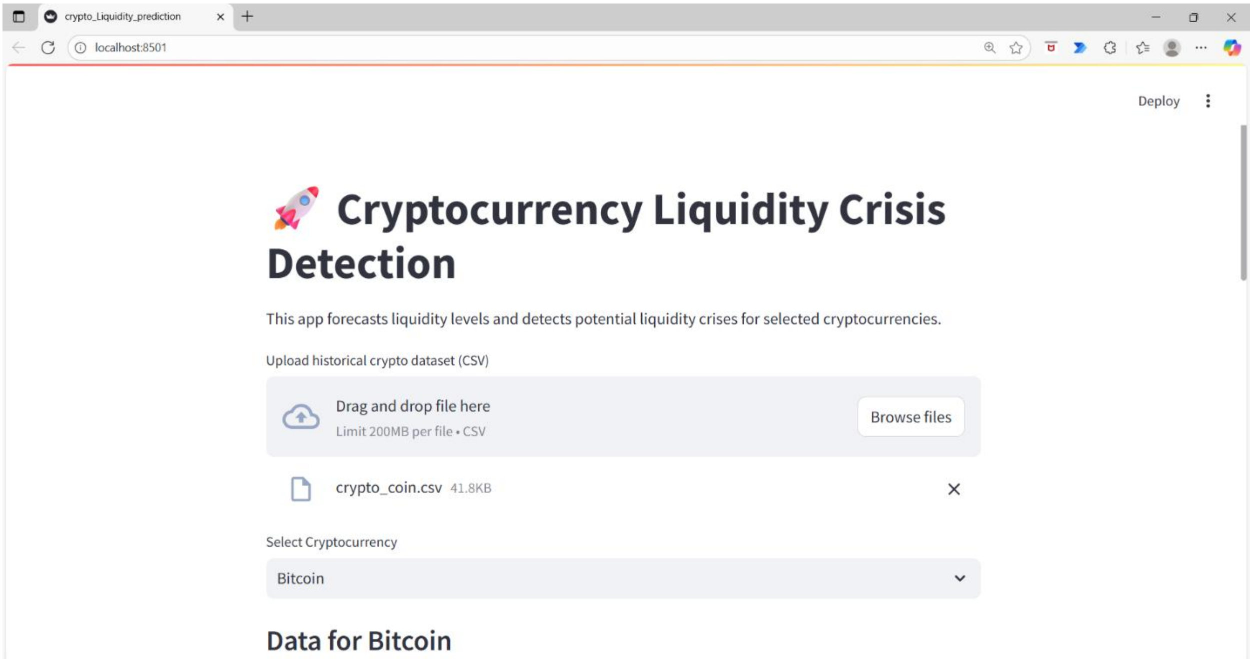
A web application was developed using **Streamlit** to provide real-time liquidity insights.

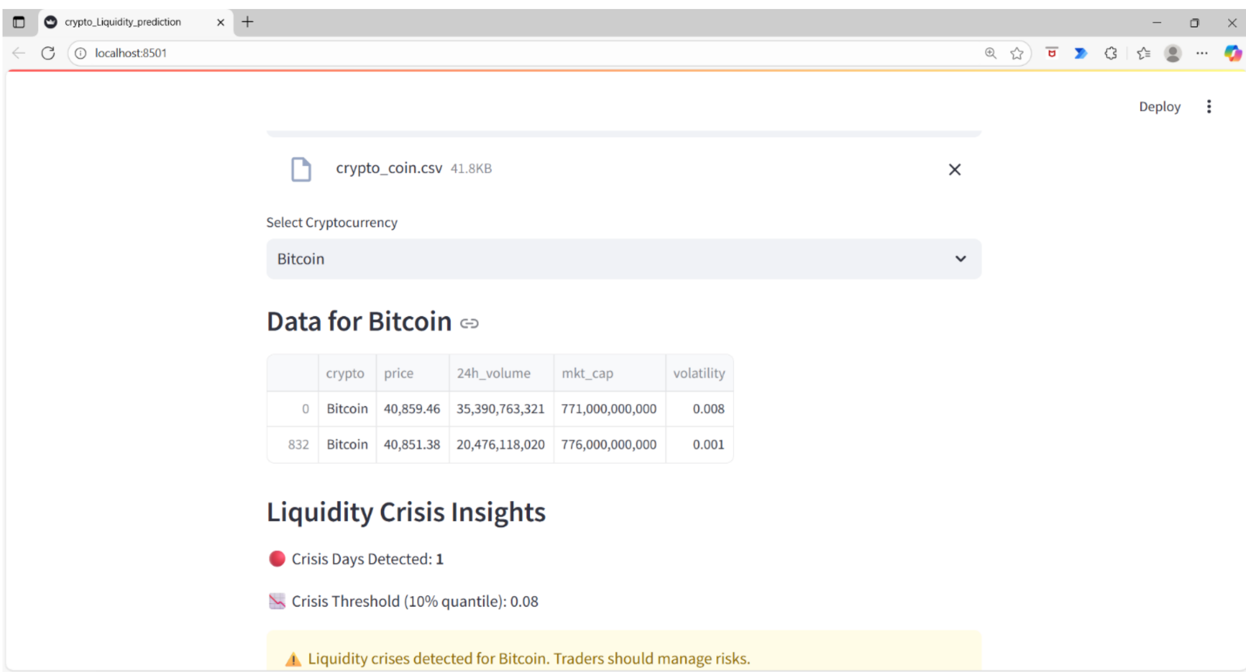
5.1 Features of the App

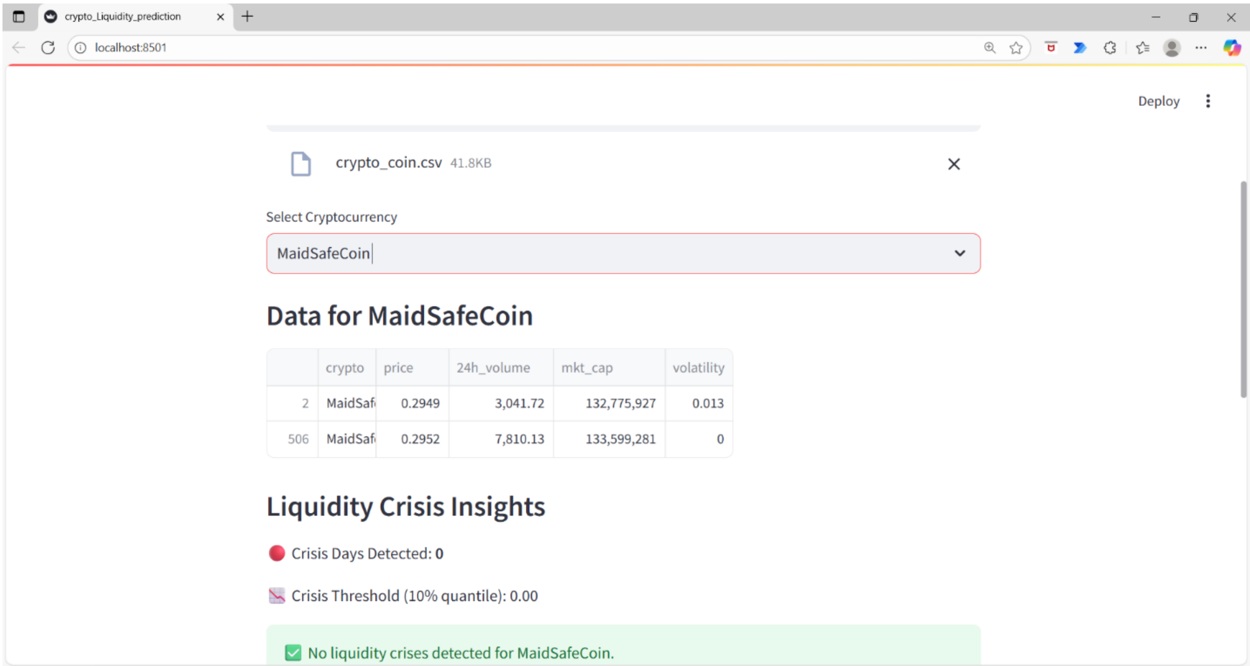
1. **Upload Historical Data:** Users can upload CSV files with historical cryptocurrency data.
2. **Select Cryptocurrency:** Users can filter data for specific cryptocurrencies.
3. **Predict Liquidity:** The app predicts liquidity levels using the trained Random Forest model.
4. **Crisis Detection:**
    - Liquidity crises are flagged if predicted liquidity falls below the 10th percentile threshold.
    - Displays the number of crisis days and highlights periods of low liquidity.
5. **Visualizations:**
    - Line plot showing predicted liquidity over time
    - Crisis threshold indicated by a red dashed line
6. **Download Predictions:** Users can download a CSV file containing predicted liquidity and crisis flags.

5.2 Sample Output

Date	Predicted_Liquidity	Crisis_Flag
2022-03-16	0.025	False
2022-03-16	0.011	True
2022-03-17	0.030	False
2022-03-17	0.008	True


crypto_Liquidity_prediction
localhost:8501
Deploy
🚀 Cryptocurrency Liquidity Crisis Detection
This app forecasts liquidity levels and detects potential liquidity crises for selected cryptocurrencies.
Upload historical crypto dataset (CSV)
Drag and drop file here
Limit 200MB per file • CSV
Browse files
crypto_coin.csv 41.8KB
Select Cryptocurrency
Bitcoin
Data for Bitcoin


crypto_Liquidity_prediction
localhost:8501
Deploy
crypto_coin.csv 41.8KB
Select Cryptocurrency
Bitcoin
Data for Bitcoin
crypto price 24h_volume mkt_cap volatility
0 Bitcoin 40,859.46 35,390,763,321 771,000,000,000 0.008
832 Bitcoin 40,851.38 20,476,118,020 776,000,000,000 0.001
Liquidity Crisis Insights
Crisis Days Detected: 1
Crisis Threshold (10% quantile): 0.08
Liquidity crises detected for Bitcoin. Traders should manage risks.



6. Conclusion

- Exploratory analysis revealed strong correlations between price, market cap, and trading volume, which are critical indicators of liquidity.
- Feature engineering, including volatility and moving averages, improved model performance.
- Random Forest Regressor provided the best predictive accuracy for cryptocurrency liquidity.
- The Streamlit app offers a user-friendly interface to detect potential liquidity crises, helping traders make informed decisions.

Future Enhancements:

- Include longer historical data for more robust predictions
- Add additional features like social sentiment, network activity, or macroeconomic indicators
- Implement real-time prediction with streaming market data

7. References

1. CoinGecko API Documentation
2. Scikit-learn: Machine Learning in Python – Pedregosa et al., 2011
3. Streamlit Documentation: https://docs.streamlit.io