Exploratory Data Analysis Report Cryptocurrency Volatility Analysis

Prepared by: Muthuswamy Nadar

**Contents**

1. [Objective](#_bookmark0) 2
2. [Dataset Overview](#_bookmark1) 2
   1. [Dataset Statistics](#_bookmark2) 2
3. [Data Preprocessing](#_bookmark3) 2
   1. [Data Loading](#_bookmark4) 2
   2. [Data Cleaning](#_bookmark5) 2
   3. [Sorting](#_bookmark6) 2
   4. [Feature Engineering](#_bookmark7) 3
   5. [Handling Missing Values](#_bookmark8) 3
4. [Exploratory Data Analysis](#_bookmark9) 3
   1. [Visualization 1: Bitcoin Close Price Over Time](#_bookmark10) 3
   2. [Visualization 2: Bitcoin 7-Day Rolling Volatility](#_bookmark11) 3
   3. [Visualization 3: Bitcoin Liquidity Ratio](#_bookmark12) 3
5. [Key Insights](#_bookmark13) 4
6. [Limitations](#_bookmark14) 4
7. [Recommendations for Modeling](#_bookmark15) 4
8. [Conclusion](#_bookmark16) 4

# Objective

The purpose of this exploratory data analysis (EDA) is to examine historical market data of cryptocurrencies, focusing on Bitcoin, to understand price trends, volatility, and liq- uidity patterns. This analysis supports the development of a volatility prediction model by identifying key features and trends in the dataset.

# Dataset Overview

The dataset contains daily market data for multiple cryptocurrencies, sourced from a CSV file (dataset.csv). Key columns include:

* Date: Date of the record
* Crypto Name: Name of the cryptocurrency (e.g., Bitcoin)
* Open, High, Low, Close: OHLC prices
* Volume: Trading volume
* Market Cap: Market capitalization
* Unnamed: 0: Index column (dropped during cleaning)

## Dataset Statistics

* + - Number of Records: Approximately 72,945
    - Time Period: 2013–2022 (assumed based on context)
    - Cryptocurrencies: Includes Bitcoin, Litecoin, XRP, and others
    - Missing Values: Handled by dropping rows with NaN values after feature engineer- ing

# Data Preprocessing

The following steps were performed to clean and prepare the data:

## Data Loading

Loaded the dataset using Pandas’ read\_csv function.

## Data Cleaning

* + - Dropped the Unnamed: 0 column, which served as an index.
    - Converted the Date column to datetime format.

## Sorting

Sorted the data by Crypto Name and Date to ensure chronological order for time-series calculations.

## Feature Engineering

* + - Daily Return: Calculated as the percentage change in close price: *closet−closet−*1 .

*close −*1

*t*

* + - Rolling Volatility (7-day): Standard deviation of daily returns over a 7-day rolling window, grouped by Crypto Name.
    - Liquidity Ratio: Calculated as  *volume* , indicating trading activity relative to mar-

*marketCap*

ket size.

* + - 7-day Moving Average of Close Price: Mean of close prices over a 7-day rolling win- dow.

## Handling Missing Values

Dropped rows with NaN values resulting from rolling calculations to create a cleaned dataset.

# Exploratory Data Analysis

The EDA focused on Bitcoin, the most prominent cryptocurrency in the dataset. Three visualizations were generated to explore price trends, volatility, and liquidity.

## Visualization 1: Bitcoin Close Price Over Time

* + - Description: A line plot of Bitcoin’s closing price against date.
    - File: bitcoin\_close\_price.png
    - Observations:
      * Bitcoin’s price shows significant growth, with peaks likely around 2017 and 2021.
      * Sharp increases are followed by corrections, indicating high volatility.
      * The upward trend suggests long-term appreciation with short-term fluctua- tions.

## Visualization 2: Bitcoin 7-Day Rolling Volatility

* + - Description: A line plot of Bitcoin’s 7-day rolling volatility against date.
    - File: bitcoin\_volatility.png
    - Observations:
      * Volatility fluctuates, with spikes during major market events.
      * Low volatility periods often precede sharp increases, suggesting consolidation phases.
      * Volatility clusters, with high-volatility periods lasting weeks to months.

## Visualization 3: Bitcoin Liquidity Ratio

* + - Description: A line plot of Bitcoin’s liquidity ratio against date.
    - File: bitcoin\_liquidity.png
    - Observations:
      * Periodic spikes indicate surges in trading activity relative to market size.
      * High liquidity coincides with volatile periods, suggesting increased market participation.
      * Long-term trends show varying liquidity, with some periods of lower trading activity.

# Key Insights

* Price Trends: Bitcoin’s closing price has a long-term upward trajectory but is highly volatile, making it suitable for volatility modeling.
* Volatility Patterns: The 7-day rolling volatility captures market dynamics, with spikes aligning with known events, critical for prediction models.
* Liquidity Dynamics: The liquidity ratio highlights intense trading periods that cor- relate with volatility, emphasizing volume-based features.
* Feature Importance: Engineered features (daily return, rolling volatility, liquidity ratio, moving average) provide meaningful insights for modeling.

# Limitations

* Missing Values: Rolling calculations introduce NaN values for initial records, re- ducing dataset size after cleaning.
* Single Cryptocurrency Focus: Visualizations focus on Bitcoin, which may not rep- resent smaller cryptocurrencies.
* Data Scope: Lacks external factors (e.g., news sentiment, macroeconomic indica- tors) that could influence volatility.

# Recommendations for Modeling

* Use engineered features (daily return, rolling volatility, liquidity ratio, moving av- erage) as model inputs.
* Employ time-series models (e.g., RandomForestRegressor with TimeSeriesSplit) to respect temporal dependencies.
* Explore additional features like Bollinger Bands or ATR for enhanced market dy- namics.
* Validate findings on other cryptocurrencies for generalizability.

# Conclusion

The EDA reveals that Bitcoin’s market data exhibits complex patterns in price, volatility, and liquidity, well-captured by engineered features. The visualizations provide a clear understanding of temporal trends, laying a strong foundation for a volatility prediction model. The cleaned dataset and derived features are ready for subsequent modeling steps.