

# Exploratory Data Analysis (EDA) Report

## 1. Introduction

This Exploratory Data Analysis (EDA) report is part of the **Cryptocurrency Volatility Prediction** project. The goal of EDA is to understand historical cryptocurrency market data, identify patterns and trends, detect anomalies, and analyze relationships between variables that influence market volatility.

The insights gained from EDA guide feature engineering and model selection for predicting cryptocurrency volatility.

---

## 2. Dataset Overview

The dataset contains **daily historical records** for more than **50 cryptocurrencies**.

### 2.1 Features Description

Feature	Description
date	Trading date
symbol	Cryptocurrency symbol (e.g., BTC, ETH)
open	Opening price of the day
high	Highest price of the day
low	Lowest price of the day
close	Closing price of the day
volume	Trading volume
market_cap	Market capitalization

### 2.2 Dataset Size

- Records: Daily entries across multiple cryptocurrencies
  - Time span: Multiple years of historical data
- 

## 3. Data Quality Checks

### 3.1 Missing Values

- Some cryptocurrencies contained missing values in **volume** and **market\_cap** columns.

- Forward-fill method was used to maintain time-series continuity.
- Rows with remaining missing values after preprocessing were removed.

### 3.2 Data Consistency

- Data was sorted by **symbol** and **date**.
  - Duplicate records were checked and removed if found.
- 

## 4. Statistical Summary

Key numerical statistics were analyzed for price and volume-related features.

### 4.1 Observations

- Prices show a **wide range**, indicating high volatility.
  - Volume and market capitalization are **right-skewed**, with few cryptocurrencies dominating trading activity.
  - High standard deviation in prices suggests frequent market fluctuations.
- 

## 5. Feature Engineering for EDA

To better analyze volatility patterns, the following features were engineered:

- **Log Returns**: Measures percentage price change
  - **14-Day Rolling Volatility**: Standard deviation of log returns (target variable)
  - **Simple Moving Averages (7 & 14 days)**: Trend detection
  - **Price Range (High – Low)**: Intraday volatility
  - **Volume to Market Cap Ratio**: Liquidity indicator
- 

## 6. Univariate Analysis

### 6.1 Volatility Distribution

- The volatility distribution is **positively skewed**.
- Most days show moderate volatility, with **occasional spikes** representing market shocks.
- Extreme volatility values often align with major crypto market events.

### 6.2 Price Distribution

- Closing prices vary significantly between cryptocurrencies.
  - Large-cap cryptocurrencies show more stable price distributions compared to smaller-cap assets.
-

## 7. Bivariate Analysis

### 7.1 Volume vs Volatility

- Higher trading volume often corresponds to **increased volatility**.
- Sudden volume spikes usually precede or accompany volatile price movements.

### 7.2 Market Cap vs Volatility

- Large market-cap cryptocurrencies tend to have **lower relative volatility**.
  - Smaller market-cap coins exhibit **frequent high-volatility periods**.
- 

## 8. Correlation Analysis

A correlation heatmap was used to analyze relationships between features.

### 8.1 Key Insights

- **High, Low, Open, Close prices** are strongly correlated.
  - **Volatility** has moderate correlation with:
    - Price range
    - Trading volume
  - Liquidity ratio shows a meaningful relationship with volatility, making it a useful predictive feature.
- 

## 9. Trend Analysis

- Rolling volatility reveals **cyclical volatility patterns**.
  - Periods of market uncertainty show sustained high volatility clusters.
  - Bull and bear market phases can be visually identified through volatility trends.
- 

## 10. Outlier Detection

- Outliers are present in volume and volatility features.
  - These outliers represent real-world events (market crashes, sudden rallies) and were **not removed** to preserve market behavior.
- 

## 11. Key EDA Insights

- Cryptocurrency markets are **highly volatile and non-stationary**.
- Volatility is influenced by:
  - Trading volume

- Liquidity
  - Intraday price range
  - Feature engineering is essential to capture market dynamics.
  - Tree-based models are suitable due to non-linear relationships.
- 

## 12. Conclusion

The EDA highlights strong volatility patterns and complex relationships between market variables. These insights justify the use of engineered volatility features and machine learning models such as Random Forest for accurate volatility prediction.

The findings from EDA form the foundation for feature selection, model training, and risk-aware cryptocurrency market analysis.

---

**End of EDA Report**