Exploratory Data Analysis (EDA) Report

Objective

The objective of the EDA is to understand cryptocurrency price behaviour, volatility patterns, liquidity effects, and relationships between engineered features before building the machine learning model.

Dataset Overview

The dataset used in this project consists of **daily historical records for multiple cryptocurrencies**, covering a broad time horizon. Each record represents the market activity of a specific cryptocurrency on a given date, enabling both **cross-sectional** and **time-series analysis**.

The key features included in the dataset are:

- **Open Price**: The price at which the cryptocurrency started trading on a given day.
- **High Price**: The highest price reached during the trading day.
- **Low Price**: The lowest price recorded during the trading day.
- **Close Price**: The final trading price at the end of the day, commonly used as a reference price in financial analysis.
- **Trading Volume**: The total quantity of the cryptocurrency traded during the day, reflecting market activity and liquidity.
- **Market Capitalization**: The total market value of the cryptocurrency, calculated as price multiplied by circulating supply.
- **Date**: The calendar date of the observation, allowing chronological ordering and time-based computations.

The dataset spans **multiple years**, covering different market phases such as bullish trends, bearish corrections, and high-volatility events. This long time span makes the dataset suitable for analyzing **long-term trends, volatility clustering, and regime changes** in cryptocurrency markets.

Because the dataset includes **multiple cryptocurrencies**, it captures a wide range of behaviours—from highly liquid, large-cap assets to smaller, more volatile assets—making it ideal for studying how volatility differs across market conditions and asset types.

Summary Statistics

Volatility Characteristics

The summary statistics reveal that **most volatility values are relatively low**, indicating that cryptocurrencies spend a large portion of time in comparatively stable conditions. However, the presence of **extreme maximum values** shows that volatility occasionally spikes sharply.

These spikes correspond to **market shocks**, such as sudden price crashes, rallies, regulatory news, or macroeconomic events. This behaviour confirms that cryptocurrency volatility is

right-skewed and **non-normally distributed**, which is a well-known characteristic of financial time-series data.

Return Distribution

The distribution of log returns shows **heavy tails**, meaning that extreme positive and negative returns occur more frequently than would be expected under a normal distribution. This indicates the presence of **fat-tail risk**, where large price movements are more common.

Such behaviour violates the assumptions of simple linear or Gaussian models and highlights the need for:

- Robust feature engineering (rolling windows, volatility measures)
- Non-linear machine learning models that can handle extreme values

---

Trading Volume and Liquidity

Trading volume and liquidity metrics exhibit **high variability across cryptocurrencies and time periods**. Some assets consistently show high volume, reflecting strong market participation and liquidity, while others experience intermittent or low trading activity.

This variability implies that:

- Low-liquidity assets are more prone to sudden price swings
- High-liquidity assets tend to absorb shocks more effectively

As a result, volume-based and liquidity-based features play a crucial role in explaining and predicting volatility.

Overall Statistical Insight

The summary statistics collectively indicate that cryptocurrency markets are:

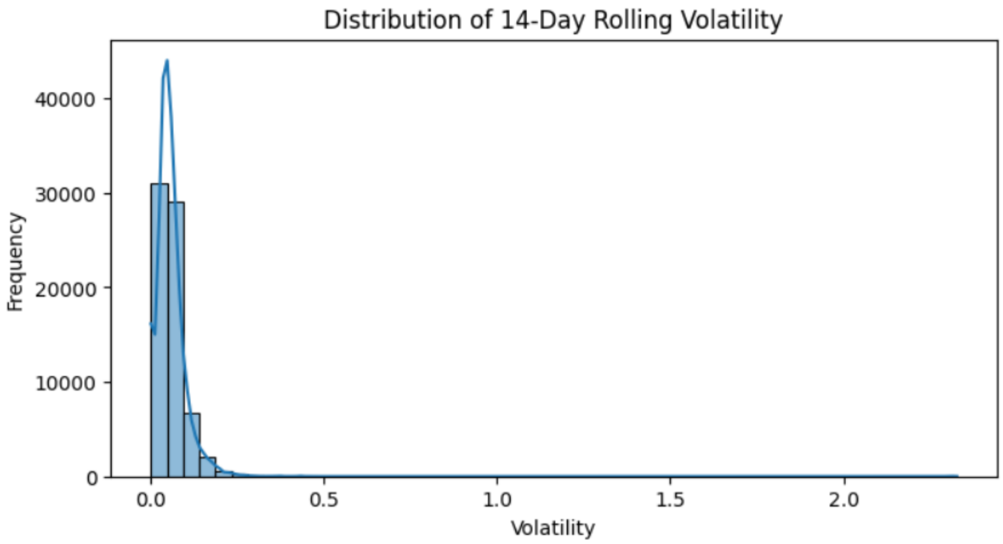
- **Highly volatile**
- **Non-stationary**
- **Influenced by liquidity conditions**
- **Prone to extreme events**

These observations justify the use of **rolling volatility measures**, **liquidity indicators**, and **advanced machine learning models** rather than simple statistical approaches.

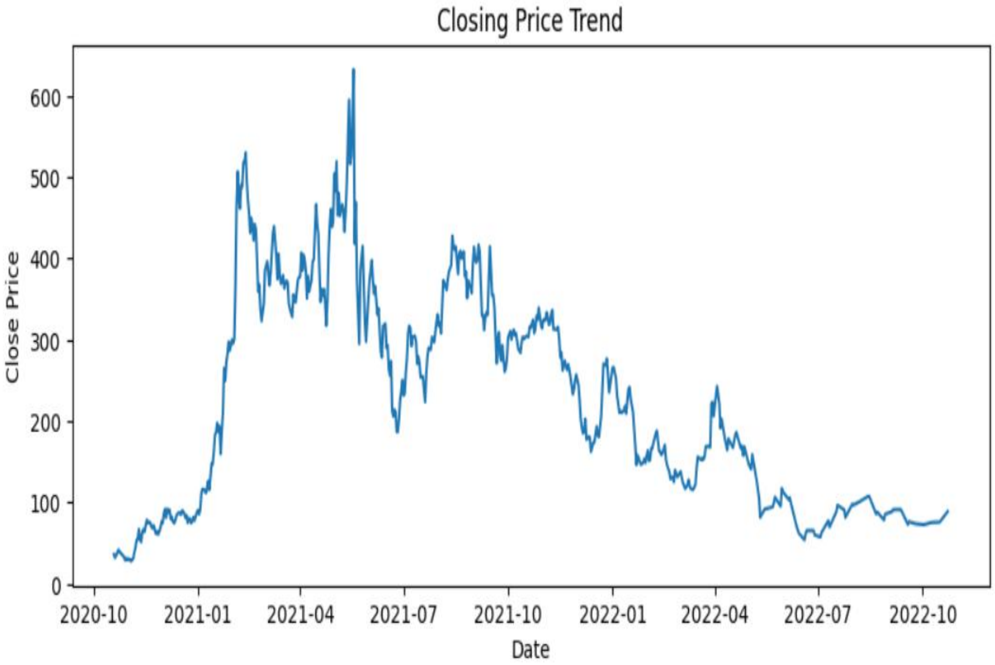
Visual Analysis & Insights

1. **Volatility Distribution**
   - The volatility distribution is right-skewed.
   - Most observations fall under low volatility.
   - Rare but extreme spikes indicate market shocks.

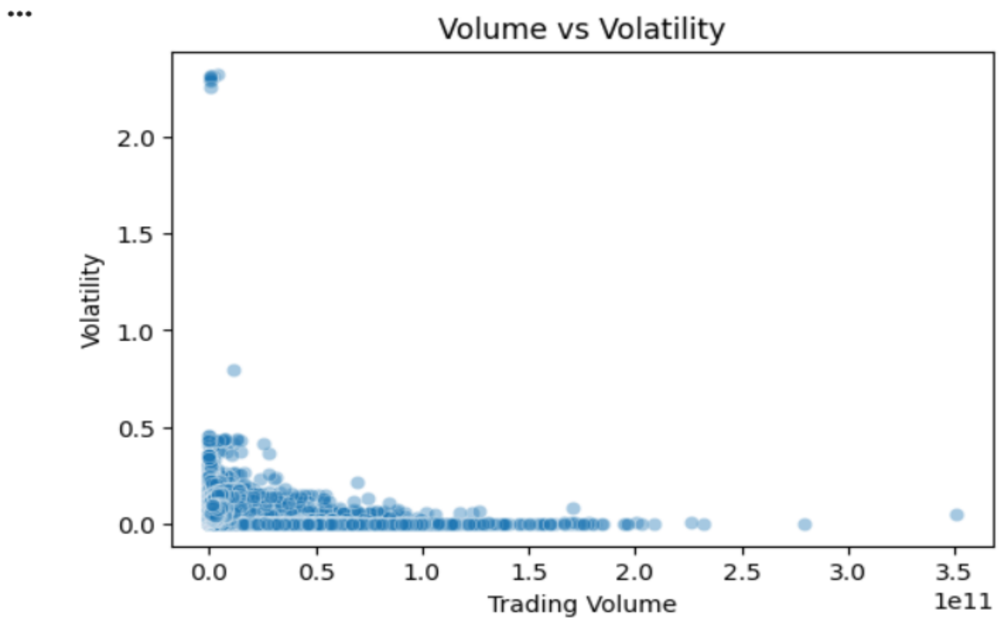
...



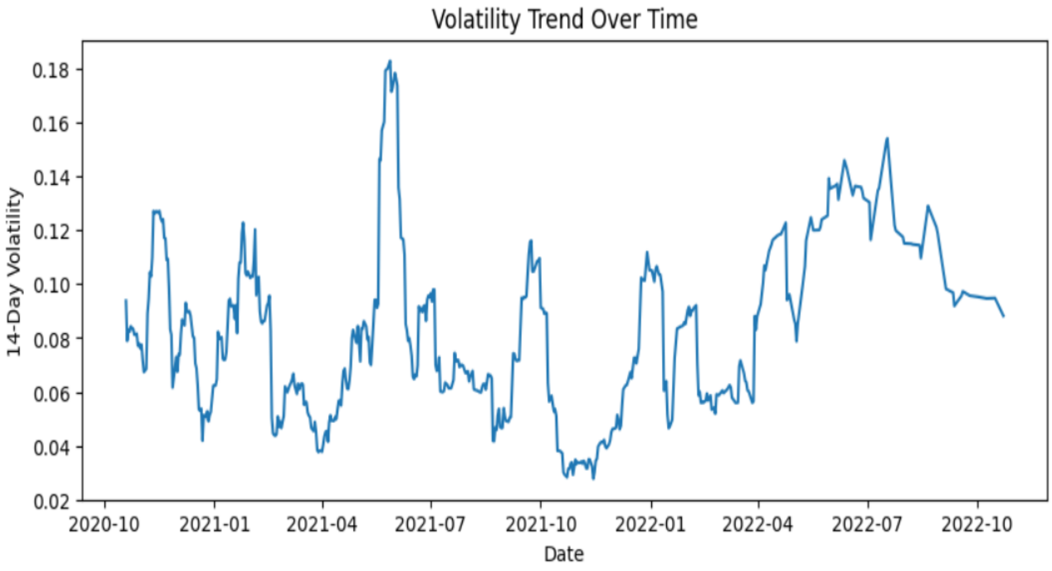
2. **Closing Price Trend**
   - Prices show sharp uptrends and downtrends.
   - Sudden jumps and crashes reflect speculative market behaviour.
   - Justifies the use of rolling and trend-based indicators.



3. **Volume vs Volatility**
    - Lower trading volume is associated with higher volatility.
    - Higher liquidity stabilizes price movements.
    - Confirms the importance of liquidity-based features.



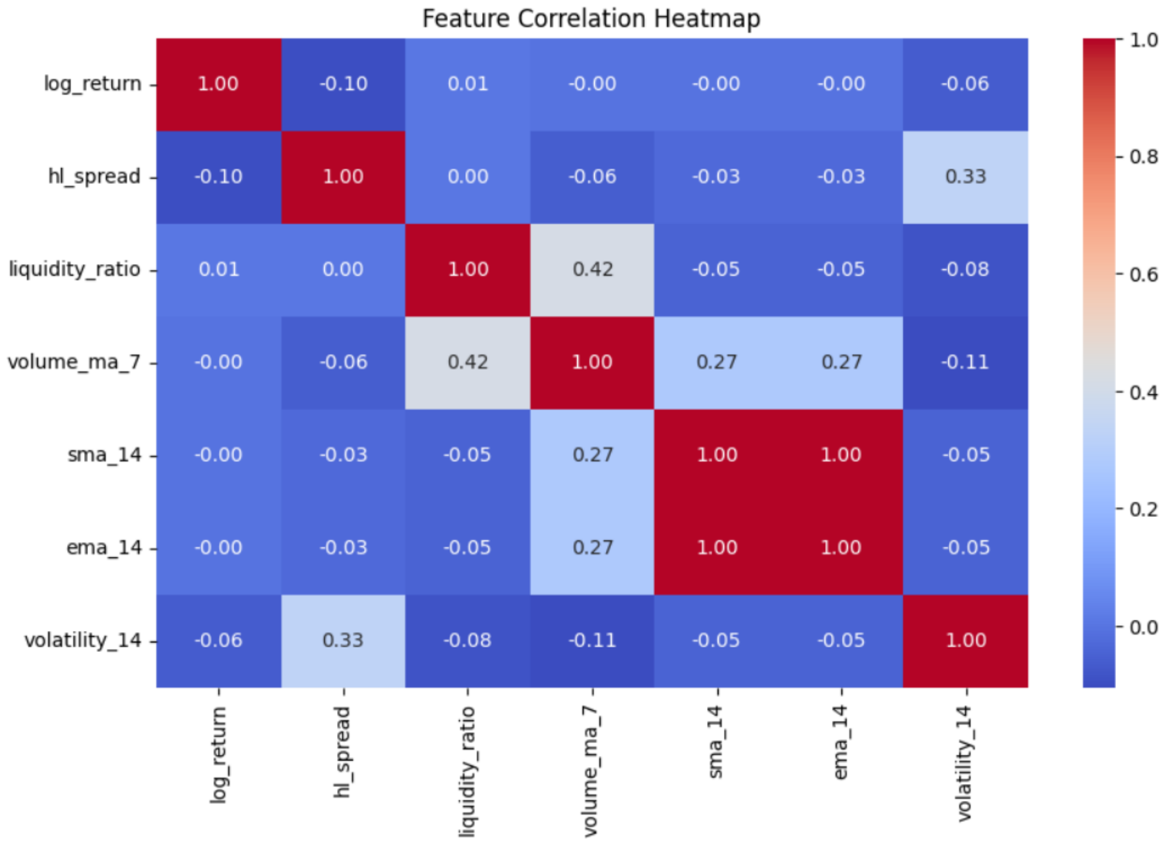
4. **Volatility Over Time**
    - Volatility occurs in clusters.
    - Periods of high volatility persist over time.
    - Confirms volatility persistence and predictability



.

Correlation Heat map

- High–Low spread shows positive correlation with volatility.
- Liquidity features influence volatility more than raw returns.
- No severe multicollinearity observed among selected features.



EDA Conclusion

The Exploratory Data Analysis (EDA) provided critical insights into the structural behaviour of cryptocurrency markets and directly informed the feature engineering and modelling strategy adopted in this project.

The analysis revealed that **cryptocurrency volatility is highly right-skewed**, with the majority of observations concentrated at low volatility levels and a small number of extreme values forming a long tail. This indicates that cryptocurrencies typically experience relatively stable periods, punctuated by sudden and intense volatility spikes caused by market shocks, news events, or speculative trading behaviour. Such a distribution confirms that volatility does not follow a normal pattern, making traditional linear assumptions insufficient.

Time-series visualizations of closing prices demonstrated **strong non-stationarity**, characterized by sharp upward rallies followed by abrupt corrections. These rapid price movements highlight the speculative nature of crypto assets and justify the use of rolling indicators and trend-based features, such as moving averages, to smooth short-term noise and capture underlying price direction.

The relationship between **trading volume and volatility** showed that lower-liquidity assets tend to exhibit higher and more unstable volatility. In contrast, cryptocurrencies with higher trading volumes displayed comparatively more stable volatility patterns. This finding emphasizes the importance of liquidity as a stabilizing factor and validates the inclusion of liquidity-based features such as volume moving averages and volume-to-market-cap ratios in the model.

Correlation analysis further revealed that **price range indicators**, particularly the High–Low spread, have a stronger relationship with volatility than raw returns. Liquidity features also showed meaningful correlations, while moving averages were more closely related to trend detection rather than direct volatility estimation. Importantly, the correlation structure indicated **no severe multicollinearity**, suggesting that the engineered features provide complementary information rather than redundant signals.

The volatility-over-time analysis clearly demonstrated **volatility clustering**, where periods of high volatility are followed by continued high volatility, and calmer periods persist similarly. This persistence implies that past volatility contains predictive information about future volatility, supporting the feasibility of forecasting volatility using historical patterns.

Overall, the EDA confirms that cryptocurrency markets are **highly volatile, liquidity-sensitive, and non-stationary**, with strong temporal dependencies. These insights justified the use of rolling statistical features, liquidity metrics, and non-linear machine learning models such as XGBoost. The findings from EDA laid a solid foundation for feature engineering, model selection, and deployment, ensuring that the predictive system is both data-driven and contextually aligned with real-world crypto market behaviour.

.