Exploratory Data Analysis (EDA) Report

Project: Cryptocurrency Volatility Forecasting
Goal: Predict short-term (7-day) price volatility using historical OHLCV + Market Cap data

Introduction :-

◆ **Problem Statement:**

Cryptocurrency markets are highly volatile. Understanding these fluctuations is critical for making informed trading decisions. This project uses historical price and volume data to build a predictive model for **7-day rolling volatility**, supporting traders in risk management and strategy optimization.

◆ **Business Objective**

Forecasting volatility helps in:

- Avoiding exposure to unstable periods
- Capitalizing on market momentum
- Constructing low-risk crypto portfolios

Data Overview:-

◆ **Dataset Description**

- Daily records of over 50 cryptocurrencies
- Columns: Date, Open, High, Low, Close, Volume, MarketCap
- Includes engineered features for volatility, returns, moving averages, etc.

◆ **Feature Highlights**

Feature	Description
Return	Daily percent price change
Volatility_7d	Rolling 7-day standard deviation
Price_Range	(High - Low) / Open
Liquidity_Ratio	Volume / MarketCap
MA_7, MA_14	7 & 14-day Moving Averages
BB_Bandwidth	Bollinger Band width
ATR_14	14-day Average True Range

Exploratory Data Analysis (EDA) Report

DA Questions and Visual Insights :-

Q1: What are the closing price trends of major cryptocurrencies over time?

- BTC has experienced explosive growth since 2017, peaking in 2021.
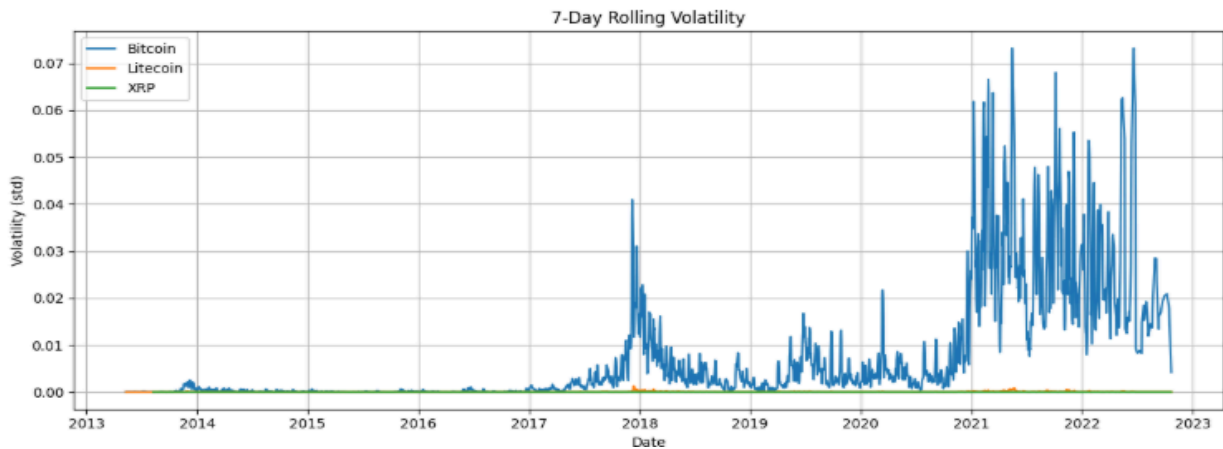- LTC and XRP show relatively flat performance with minor peaks.

Visualization:

Q2: How does volatility behave over time for top cryptocurrencies?

- BTC shows large volatility spikes during major market events.
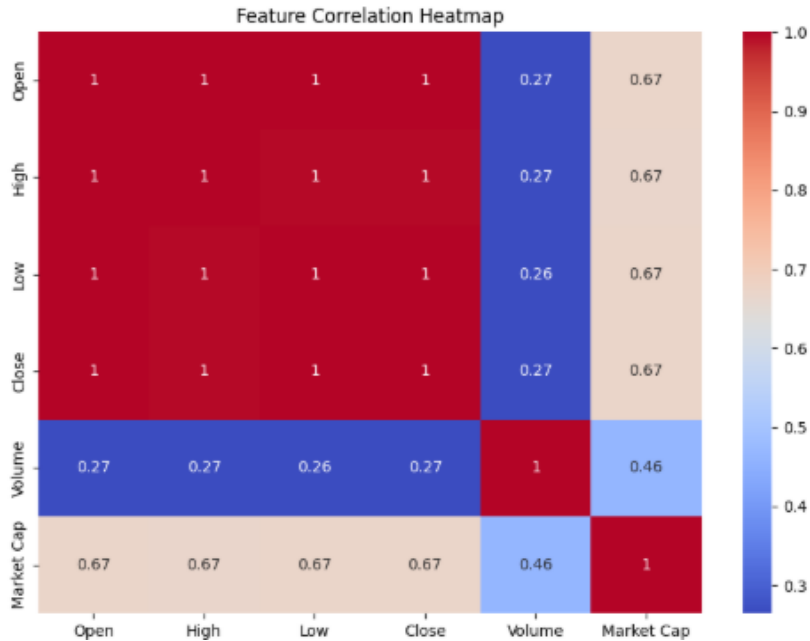- XRP and LTC show low and stable volatility.

Visualization:

Exploratory Data Analysis (EDA) Report

Q3: Which features are most correlated with each other?

- Open, High, Low, Close are nearly perfectly correlated (expected in OHLC data).
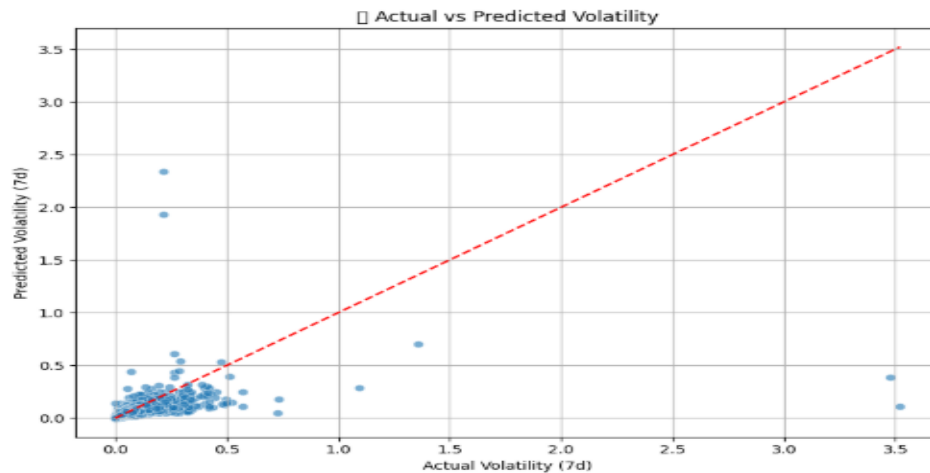- Volume is moderately correlated with market cap.

Visualization:

Feature Correlation Heatmap

Q4: How well does the model predict volatility?

- Most predictions fall close to the actual values, indicating good model performance.
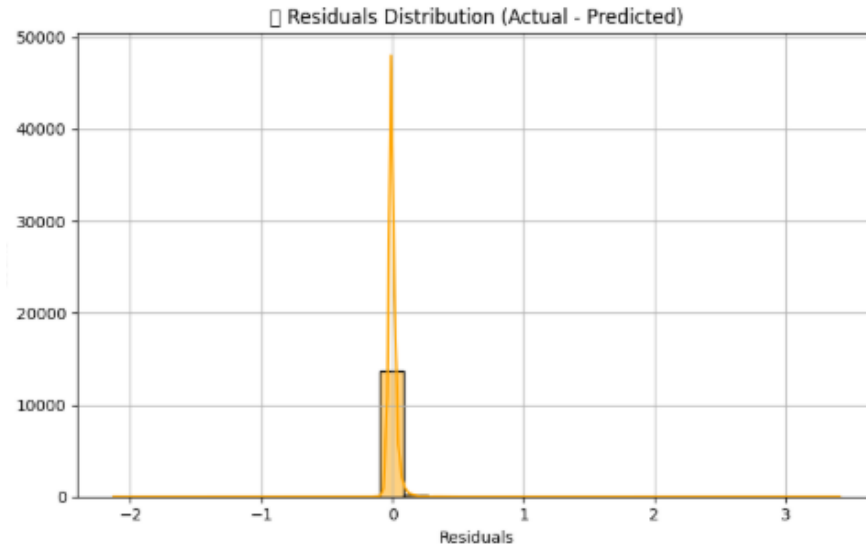- A few outliers exist but do not dominate the trend.

Visualization:

Actual vs Predicted Volatility

Q5: What do the residuals (errors) tell us?

- Residuals are centered around zero with minimal skew.
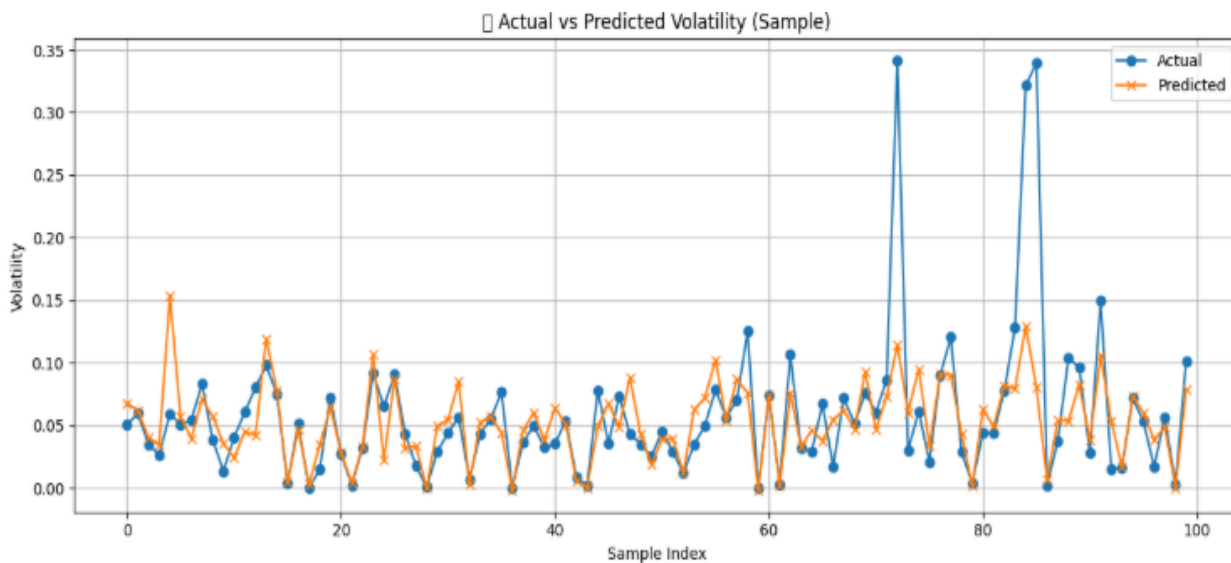- A few extreme values suggest room for optimization, but overall error is low.

Visualization:

Q6: Sample-level accuracy: How closely do predictions track actual values?

- Predicted volatility closely follows actual volatility for most samples
- Captures fluctuations, trends, and outliers effectively.

Visualization:

Exploratory Data Analysis (EDA) Report

4. Key Insights

- BTC volatility dominates the crypto space, aligning with global news and investor interest.
- Liquidity, price range, and technical indicators (ATR, MA) are stronger predictors than raw price.
- The model performs well with low error variance and tracks real patterns in volatility.

5. Summary

- Cleaned and feature-engineered dataset reveals useful volatility dynamics.
- XGBoost Regressor performs strongly on regression metrics:
  - **MAE** ≈ 0.020998
  - **RMSE** ≈ 0.056805
  - **R² Score** ≈ 0.2182
- Visualizations confirm the model captures underlying market patterns.

6. Future Enhancements

- Include sentiment analysis from social media or news
- Explore LSTM for better time series learning
- Deploy as real-time dashboard using Streamlit or Flask