Cryptocurrency Volatility Analysis (Phase 1 – Phase 3)

University: COMSATS University Islamabad, Sahiwal Campus
Course: Introduction to Data Science
Instructor: Dr. Fakhar Mustafa

Students:
Ahmer Bilal Mustafa (SP23-BCS-009)
Ali Hassan (SP23-BCS-012)
Salman (FA22-BCS-109)

1. Introduction and Project Overview

The rapid growth of cryptocurrency markets has introduced new challenges due to their highly volatile nature. Unlike traditional financial assets, cryptocurrencies experience frequent and significant price fluctuations within short time periods. This project aims to analyze the volatility behavior and price dynamics of major cryptocurrencies using historical market data. The analysis follows a complete data science workflow consisting of data importing, cleaning, exploratory analysis, and predictive modeling. The structured phase-wise approach ensures clarity, reproducibility, and analytical depth.

2. Dataset Description and Sources

The datasets used in this project consist of daily OHLC (Open, High, Low, Close) records for three major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB). The data was collected from reputable public sources including Yahoo Finance and CoinMarketCap. Each dataset contains time-stamped price values, daily trading volume, and market-related attributes such as market capitalization. All datasets were downloaded in CSV format and stored as raw data without any modification to preserve originality.

3. Phase 1: Data Importing and Cleaning

Phase 1 focused on preparing clean and reliable datasets suitable for further analysis. The raw CSV files were imported into R Studio using libraries such as readr and data.table. During the importing process, several challenges were encountered including incorrect delimiter usage and improper column separation, which resulted in all data being read into a single column. These issues were resolved by explicitly specifying delimiters and restructuring the datasets. Data cleaning involved removing zero-volume records, converting character values into numeric formats, handling missing values, standardizing date formats, and retaining only relevant OHLC attributes.

4. Phase 1: Data Transformation

After completing the cleaning process, multiple data transformation techniques were applied to enhance analytical usefulness. Logarithmic returns were calculated to capture relative price changes while stabilizing variance in the time series data. Daily intraday volatility was computed using the difference between high and low prices normalized by the opening price, providing a percentage-based measure of price fluctuation. In addition, close prices were normalized to a base value to enable meaningful comparison across cryptocurrencies with different price scales. The transformed datasets were stored as cleaned CSV files.

5. Phase 2: Exploratory Data Analysis (EDA)

Phase 2 involved exploratory data analysis to understand patterns and relationships within the cleaned datasets. Time-series visualizations were generated to examine price trends, log return behavior, and volatility movement over time. Correlation analysis was conducted to measure interdependencies between cryptocurrencies at both price and return levels. Candlestick charts were used to visually represent OHLC movements. All visual outputs were automatically exported into a multi-page PDF file to ensure reproducibility and proper documentation of results.

6. Phase 3: Predictive Modeling

In Phase 3, predictive modeling techniques were applied to analyze future price behavior. A linear regression model was developed to predict next-day log returns using lagged return features, volatility measures, and moving averages. Additionally, a logistic regression model was implemented to classify future price direction as upward or downward. Model performance was evaluated using metrics such as RMSE, MAE, and classification accuracy, providing insights into the predictive capability of the models.

7. Conclusion

This project demonstrates a complete and systematic data science pipeline applied to cryptocurrency market data. Careful data importing, cleaning, and transformation in Phase 1 enabled meaningful exploratory analysis and predictive modeling in later phases. The project highlights the importance of clean data, reproducible workflows, and structured analysis, providing a strong foundation for future research and advanced machine learning applications.