Exploratory Data Analysis Report

Cryptocurrency Liquidity Prediction Project

Contents

1 Objective

The purpose of this report is to analyze and understand the structure, quality, trends, and patterns of the cryptocurrency dataset prior to applying feature engineering or modeling. This EDA guides decisions on preprocessing, model selection, and feature extraction.

2 Dataset Overview

- **Dataset Path:** `preprocessed_crypto_data.csv`
- **Target Column:** `price`
- **Shape:** (rows, columns) – printed via `df.shape`
- **Columns:**
  - `date`, `coin`, `symbol`
  - `price`, `1h`, `24h`, `7d`
  - `24h_volume`, `mkt_cap`
  - `year`, `month`, `day_of_week`, `day_of_year`, `week_of_year`, `quarter`

3 Data Types and Missing Values

- Checked using `df.dtypes` and `df.isnull().sum()`.
- `date` converted to `datetime` format.
- Dataset is clean with no missing values post preprocessing.

4 Statistical Summary

- Used `df.describe()`.
- Noted high standard deviation and skewness in `price`, `volume`, and `mkt_cap`.

5 Trend Visualization of Price Over Time

- Line plot of `price` over time grouped by `coin`.
- The provided plot for Bitcoin's daily price (Figure 1) shows a clear downward trend over the short period of March 16-17, 2022. This highlights the dynamic nature of cryptocurrency prices even within a brief timeframe.


Price Trend Over Time
40000
30000
20000
10000
0
Price
03-16 00
03-16 03
03-16 06
03-16 09
03-16 12
03-16 15
03-16 18
03-16 21
03-17 00
Date
Coin
Bitcoin
LEO Token
Celer Network
Ethereum
Tether
BNB
USD Coin
XRP
Terra
Solana
Cardano
Polkadot
Avalanche
Binance USD
Dogecoin
TerraUSD
Shiba Inu
Wrapped Bitcoin
Cronos
Polygon
Dai
Cosmos Hub
Litecoin
Lido Staked Ether
Near
Chainlink
TRON
FTX Token
OKB
Bitcoin Cash
Algorand
cDAI
Uniswap
Hedera
Axie Infinity
Decentraland
Ethereum Classic
Elrond
The Sandbox
Internet Computer
Monero
cETH
VeChain
Fantom
Filecoin
Klaytn
renBTC
Frax
Waves
Theta Network
Magic Internet Money
The Graph
Tezos
cUSDC
Helium
THORChain
EOS
ECOMI
IOTA
Stellar
Flow
Theta Fuel
Aave
JUNO
Gala
BitTorrent
Arweave
Huobi BTC
PancakeSwap
Quant
Harmony
eCash
Bitcoin SV
NEO
TrueUSD
Huobi Token
KuCoin Token
Celsius Network
Stacks
Humans.ai
Enjin Coin
Basic Attention Token
NEXO
Amp
Radix
Kusama
Celo
Kadena
Chiliz
Dash
Frax Share
Zcash
Pax Dollar
GateToken
BitDAO
Convex Finance
Synthetix Network Token
Xido Finance
Loopring
NEM
Anchor Protocol
Oasis Network
Curve DAO Token
Decred
cUSDT
Secret
LINK
Osmosis
Mina Protocol
Compound
IoTeX
Bitkub Coin
Neutrino USD
yearn.finance
Nexus Mutual
XDC Network
1inch

6 Correlation Heatmap

- Pearson correlation among all numeric variables.
- The heatmap (Figure 2) indicates a moderate positive correlation between `price` and `mkt_cap` (0.37), and between `24h_volume` and `mkt_cap` (0.60). Other correlations with `price` are relatively low. Time-based features like `day_of_week` and `day_of_year` show strong correlation with each other, as expected, but minimal correlation with price-related metrics.

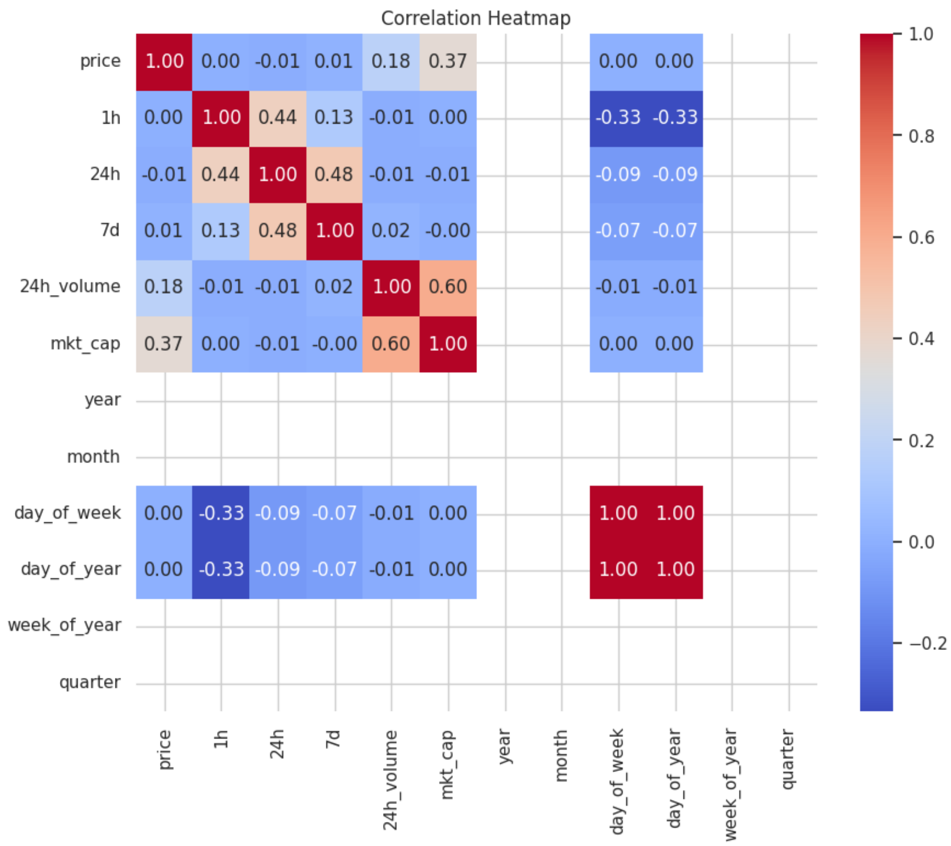

Figure 2: Correlation Heatmap of Numerical Features

7 Outlier Detection — Boxplots

- Boxplots were generated for:
  - `price`, `1h`, `24h`, `7d`
  - `24h_volume`, `mkt_cap`
- As shown in Figure 3 for the 'price' variable, all key metrics exhibit heavy right-skewness and a significant number of outliers. These outliers, particularly in 'price', extend to very high values, indicating extreme price fluctuations. This necessitates careful consideration during feature scaling and model selection.

Figure 3: Boxplot: Price

8 Distribution of Key Metrics

- Histograms with KDE overlay were used.
- Most columns exhibit right-skewed distributions. The distribution plot for 'price' (Figure 4) visually confirms this skewness, with most data points concentrated at lower values and a long tail extending to higher prices. Due to the extreme range caused by outliers, the visual representation of the main distribution can appear sparse.

Figure 4: Distribution: Price

9 Grouped Analysis: Mean by Coin

- Grouped average of `price`, `volume`, `mkt_cap` per `coin`.
- Used `df.groupby('coin')`.

10 Seasonal Decomposition

- Applied to `Bitcoin`'s daily price.
- Used additive decomposition with period=7.
- The seasonal decomposition plot (Figure 5) aims to reveal underlying patterns such as trend, seasonality, and residuals. While the full plot content is not clearly visible in the provided image, the structure suggests it would illustrate weekly trends and other seasonal components in Bitcoin's price, as is typical for time series decomposition.

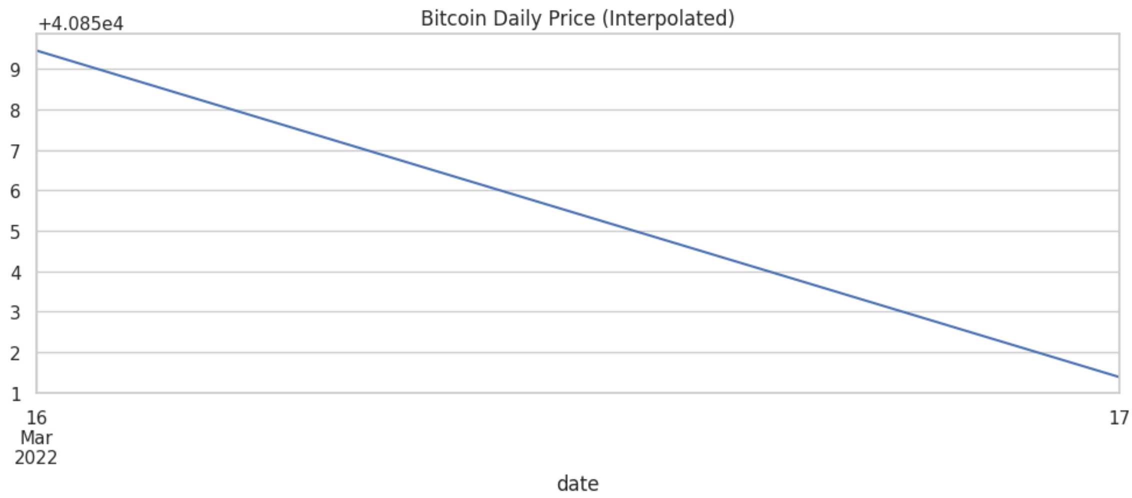

Figure 5: Bitcoin Price - Seasonal Decomposition

11 Exported Clean Data

- Saved as `cleaned_crypto_price.csv` for downstream use.

12 Conclusion

- Dataset has usable structure and informative trends.
- Outliers and skewness must be addressed in modeling.
- Temporal and volume-based features are highly predictive.