

High-Level Design (HLD)

Objective

Build a data pipeline to analyze and potentially model cryptocurrency market data using historical data scraped from CoinGecko.

Scope

- Preprocessing of cryptocurrency datasets
- Feature engineering and transformation
- Visualization and correlation analysis
- Model training and evaluation (assumed from full review)

Technology Stack

- **Languages:** Python
- **Libraries:** NumPy, Pandas, Seaborn, Matplotlib, scikit-learn
- **Data Source:** CSV files exported from CoinGecko

Module Breakdown

1. Data Ingestion

- Read two CSV datasets from consecutive dates
- Merge into a unified DataFrame

2. Data Cleaning & Transformation

- Drop unnecessary columns
- Convert all feature values to float type

3. Exploratory Data Analysis

- Statistical summaries
- Correlation heatmaps
- Distribution plots

4. Data Preparation

- Feature scaling
- Train-test split

5. **Modeling (likely)**

- Model training and testing
 - Performance evaluation metrics
-