

Main Pipeline

Step 1 – Import Data

- Load **BTC dataset**.
 - Load **ETH dataset**.
 - Columns (raw): Unix, Date, Symbol, Open, High, Low, Close, Volume, tradecount.
 - Convert Date → datetime, sort chronologically.
-

Step 2 – Handle Missing Values

Step 3 – Contributions (Apply separately to BTC & ETH)

1. **Encoding (nethm)**
 - Add column Coin (BTC or ETH).
 - Or One-Hot encode if needed later.
2. **Outlier Removal (amil)**
 - Detect extreme spikes in Volume / tradecount using IQR.
 - Remove/cap anomalies.
3. **Scaling (ber)**
 - Normalize Open, High, Low, Close, Volume, tradecount.
 - Use Min-Max or StandardScaler.
4. **Feature Selection (nor)**
 - Use correlation heatmaps to drop redundant OHLC columns.
 - Keep Close as primary price feature.
5. **Dimensionality Reduction (ran)**
 - Apply PCA on scaled numeric features.
 - Reduce to 2–3 components for visualization/analysis.
6. **Feature Creation – Indicators (mn)**

- From Close & Volume: Moving Average, RSI, MACD, Bollinger Bands.
 - Add new engineered columns.
-

Step 4 – Merge Datasets

- After preprocessing, **concatenate BTC and ETH datasets**.
 - The Coin column differentiates them.
 - Now you have **one combined dataset** with:
 - Coin, Date, Close, Volume, tradecount,
 - Moving_Average, RSI, MACD, Bollinger_Upper, Bollinger_Lower
-

Step 5 – Final Output

- One **integrated, cleaned, feature-rich dataset**.
- Ready for **EDA visualizations** (BTC vs ETH comparison).