## **Objective**

Develop a credit scoring system (0–100) for Compound V2 wallets based on transaction behavior, where higher scores indicate reliable usage and lower scores reflect risky or exploitative behavior.

## **Data Selection**

Select the three largest JSON files from the Compound V2 dataset to capture significant protocol activity.

Assume files contain transaction logs with wallet addresses, transaction types, amounts, and timestamps.

## **Defining Good vs. Bad Behavior**

**Good Behavior**:

* + High deposit frequency and volume.
  + Timely repayments (high repayment-to-borrow ratio).
  + Low or no liquidations.
  + Stable borrow-to-deposit ratio (<1).
  + Consistent transaction patterns over time.

**Bad Behavior**:

* + High borrow-to-deposit ratios (>1).
  + Frequent liquidations.
  + Low repayment rates.
  + Bot-like patterns (e.g., rapid deposit-withdraw cycles).
  + Sporadic or exploitative activity (e.g., borrowing without repaying).

## **Feature Engineering**

Aggregate transactions by wallet to create features:

* **Volume-Based**: Total deposits, borrows, repayments, withdrawals, liquidation amounts.
* **Ratio-Based**: Borrow-to-deposit ratio, repayment-to-borrow ratio.
* **Frequency-Based**: Number of transactions per type, average time between transactions.
* **Temporal Patterns**: Entropy of transaction timestamps (to detect bot-like regularity).
* **Risk Indicators**: Liquidation count, liquidation-to-borrow ratio.
* **Activity Span**: Time between first and last transaction.

## **Modeling Approach**

* **Unsupervised Learning**: Use K-means clustering to group wallets by behavior (k=5 clusters, determined via elbow method).
* **Feature Scaling**: Standardize features to ensure equal weighting.
* **Cluster Scoring**:
  + Analyze cluster centroids to identify "good" (low liquidation, high repayment) vs. "bad" (high liquidation, low repayment) clusters.
  + Assign scores (0–100) based on cluster characteristics, normalized to a 0–100 scale.
* **Validation**: Cross-check scores against domain rules (e.g., wallets with zero liquidations score >50).

## **Scoring System**

* Scores reflect reliability and protocol health:
  + High scores (80–100): High deposits, consistent repayments, no liquidations.
  + Medium scores (50–80): Moderate borrowing, some repayments, rare liquidations.
  + Low scores (0–50): High borrowing, frequent liquidations, low repayments.
* Normalize scores to ensure a smooth distribution.

## **Wallet Analysis**

* Select five high-scoring and five low-scoring wallets.
* Analyze transaction patterns (e.g., frequency, ratios, liquidations) to justify scores.
* Highlight protocol health implications (e.g., low-scoring wallets may destabilize the protocol).