Methodology Document: Decentralized Wallet Credit Scoring on Compound V2

- ## Objective

The aim of this project is to build a decentralized credit scoring system for wallets that interact with the Compound V2 protocol. Each wallet receives a score from 0 to 100 based solely on its historical transaction behavior. The score reflects financial responsibility, liquidity contribution, and risk profile.

- ## Data Source

Three largest JSON files from the Compound V2 dataset were selected for maximum protocol coverage.

Each file includes transaction-level logs for:

Deposits: Wallets supplying liquidity

Borrows: Wallets taking loans

Repays: Wallets repaying loans

Withdraws: Wallets withdrawing funds

Liquidations: Wallets forcibly closed due to undercollateralization

Each record contains the wallet address, transaction type, timestamp, asset amount, value in USD, and token used.

- ## Feature Engineering

To evaluate wallet behavior, the following features were engineered from raw transaction logs:

Repay-to-Borrow Ratio: Indicates how much a wallet repays compared to how much it borrows. Higher is better.

Liquidation Ratio: Shows how frequently a wallet is liquidated relative to its borrow actions. Lower values are preferred.

Net Flow (USD): Total deposits minus total withdrawals. Positive flow suggests long-term commitment and liquidity support.

Token Diversity: Number of unique tokens interacted with. Suggests strategic participation and broader engagement.

Active Days: Total number of days between the wallet's first and last transaction. Reflects consistency and maturity.

Each of these features was normalized and used in the final scoring logic.

Scoring Logic

Each feature was scaled to a 0–1 range using MinMaxScaler to ensure comparability across different scales.

A weighted sum of these features was then computed to calculate a raw score for each wallet. The following weights were assigned based on domain intuition and behavioral significance:

Repay-to-Borrow Ratio – 30%

This is the most critical indicator of responsible borrowing behavior. Wallets that repay what they borrow are considered the most trustworthy, which is why this feature carries the highest weight.

Liquidation Ratio – -20%

Frequent liquidation suggests poor risk management or exploitative use. This feature is negatively weighted to penalize wallets that are repeatedly liquidated.

Net Flow in USD – 25%

A wallet that deposits more than it withdraws is adding value to the protocol. This feature reflects liquidity contribution and is highly rewarded.

Token Diversity – 15%

Wallets interacting with a wide range of tokens are more likely to be real, engaged users rather than bots or flash-loan exploiters. A moderate weight is given to this behavioral signal.

Active Days – 10%

Longer participation suggests reliability and protocol familiarity. While important, it plays a supporting role compared to financial behavior.

These weights were selected using a combination of:

DeFi domain knowledge

Observed behavior patterns

Manual inspection of top- and bottom-ranked wallets

After computing the weighted sum, the score was normalized to a 0–100 range. Wallets with higher scores exhibit stronger repayment behavior, contribute net liquidity, maintain consistent activity, and demonstrate responsible protocol use.

- ## Why Use Credit Tiers

Credit scores are helpful but not always easily interpretable.

To make results actionable and segment wallet behavior more clearly, we used **KMeans clustering** with k=4 to categorize wallets into behavior-based **credit tiers**.

The tiers were labeled after sorting clusters by their average credit score:

Tier 1 (Excellent): Strong contributors with responsible and consistent behavior

Tier 2 (Good): Generally healthy behavior, minor risk signs

Tier 3 (Average): Mixed activity and engagement

Tier 4 (Risky): Poor repayment, high liquidations, or exploit-like usage

Tiers make it easier for protocols to act on credit data (e.g., by adjusting lending terms or filtering wallets for incentives).

- ## Final Output

All wallets were scored and categorized into tiers.

A **CSV file** named wallet_credit_scores.csv contains wallet addresses, their credit scores, tiers, and underlying features.

Another CSV file named top_1000_wallets.csv contains the 1,000 highest-scoring wallets, sorted in descending order.

These outputs are suitable for downstream analysis, visual dashboards, or integration with DeFi risk engines.

- ## Top vs. Bottom 5 Wallet Feature Comparison

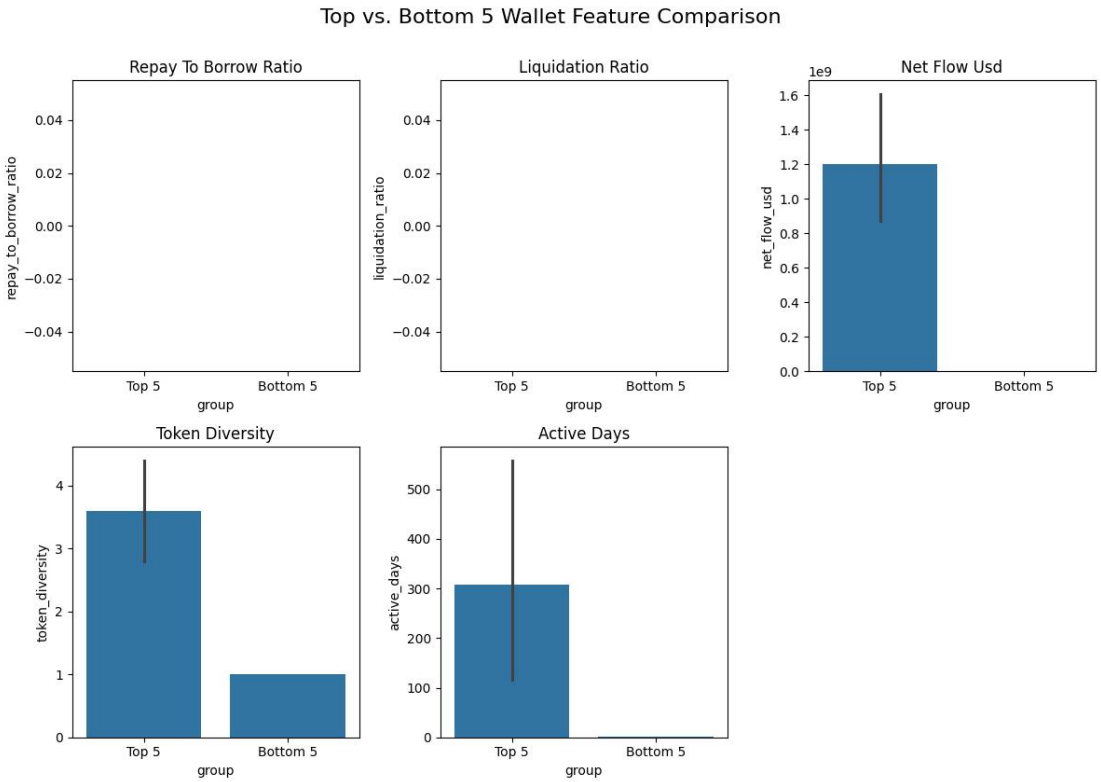
This chart highlights key differences in behavior between the top 5 and bottom 5 wallets:

Repay-to-Borrow Ratio and **Liquidation Ratio**: The bottom 5 wallets have nearly zero values, confirming they rarely repay and often get liquidated.

Net Flow USD: Top wallets consistently contribute liquidity, while bottom wallets show negligible or negative net flow.

Token Diversity: High-scoring wallets interact with a broader range of assets, reflecting more complex and stable usage.

Active Days: The top wallets engage over significantly longer periods compared to bottom wallets, which often act in a short, bursty, or exploitative manner.



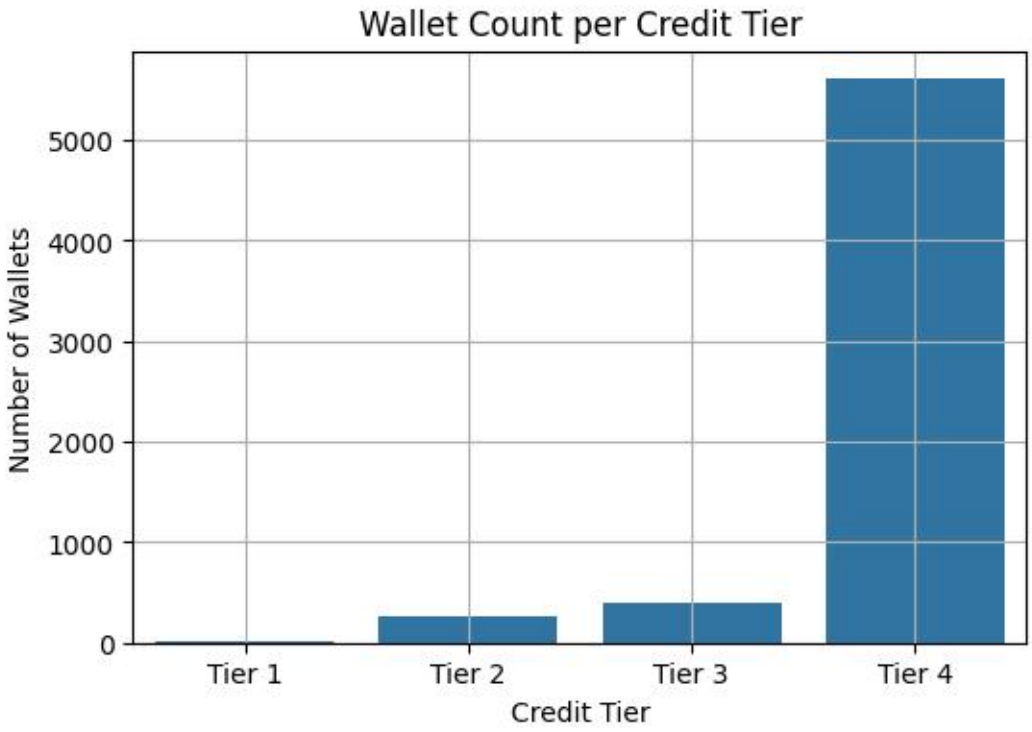
- **Wallet Distribution Across Credit Tiers**

This bar chart displays how wallets are distributed among the four credit tiers:

Tier 1 (Excellent) contains very few wallets — indicating that high-performing behavior is rare and exclusive.

Tier 2 and Tier 3 contain moderate numbers of wallets with fair-to-average behavior.

Tier 4 (Risky) holds the majority of wallets, signaling a large number of users with high-risk or poor protocol behavior — such as defaulting on loans or acting like bots.



- **Conclusion**

This system delivers an explainable, fully on-chain credit scoring model using no external labels or pretrained models. It transforms transaction data into normalized behavioral scores and tiers that reflect user trustworthiness in a decentralized setting. It can be scaled to other protocols or adapted as an open standard for DeFi user risk profiling.