Methodology Document

Compound V2 Wallet Scoring

---

1. Introduction

The Compound protocol enables decentralized money market operations such as lending, borrowing, and liquidation. While raw transactional data offers a granular view of wallet activity, it does not directly capture the *quality*, *intent*, or *sustainability* of user behavior.

To address this gap, we propose a **behavioral scoring framework** that quantifies wallet interaction patterns on Compound V2, focusing on long-term value-added participation. The methodology blends descriptive statistics with dimensionality reduction via **Principal Component Analysis (PCA)** to distill complex wallet activities into a single, interpretable score.

---

2. Feature Engineering

We begin by constructing a robust set of features designed to capture diverse wallet behaviors across the protocol. These features fall into three broad categories:

A. Activity and Engagement

- `tx_count`: Total number of Compound-specific transactions.
- `active_days`: Number of distinct days on which the wallet interacted with the protocol.
- `asset_diversity`: Number of unique tokens/assets the wallet has interacted with (deposited, borrowed, repaid, or withdrawn).

These indicators reflect *consistency*, *breadth*, and *loyalty* of usage over time.

B. Financial Behavior

- `net_deposit`: Total value of deposits minus withdrawals.

- `net_borrow`: Total value of borrows minus repayments.
- `total_liquidated`: Total USD value liquidated (i.e., value lost due to protocol-enforced liquidation).
- `num_liquidated`: Number of times the wallet was itself liquidated.

These represent the wallet's capital positioning and exposure to risk. High net deposits and managed borrowing are considered healthier than frequent liquidation or zero participation.

C. Liquidator Activity

- `liquidator_profit`: USD value gained from liquidating other users.
- `num_liquidator`: Number of times the wallet acted as a liquidator.

Liquidator behavior is a double-edged sword — useful for protocol health but also frequently exploited by flash loaners who game the system for short-term arbitrage.

---

3. Motivation for PCA

Our scoring methodology must account for:

- Multiple, possibly correlated features
- Non-obvious behavioral clusters
- The need for **relative** comparison (i.e., how a wallet performs in relation to others)

To solve this, we employ **Principal Component Analysis (PCA)** — a widely used linear technique that reduces dimensionality while retaining the maximal variance. PCA enables us to uncover latent factors that differentiate user behavior across multiple axes.

---

4. PCA Implementation Details

A. Preprocessing

All features were:

- Log-transformed (where applicable) to reduce skewness (e.g., tx_count, net_deposit)
- Standardized to have zero mean and unit variance

This ensures that all features contribute fairly to the PCA computation and aren't dominated by large numerical scales.

B. Variance Explained

PCA yielded the following variance distribution:

- **PC1**: ~52% of variance
- **PC2**: ~20% of variance
- Subsequent components had sharply diminishing returns

Hence, **PC1 alone was used as the primary score**, as it encapsulated the majority of behavioral variance while maintaining interpretability.

---

5. Score Construction

The first principal component (PC1) had the following notable feature loadings:

Feature	Loading (PC1)	Interpretation
tx_count	+	More transactions = higher score
net_deposit	+	More deposits = higher score
active_days	+	Longevity = higher score

asset_diversity	+	Broader usage = higher score
liquidator_profit	−	Flash liquidators = lower score
num_liquidator	−	Aggressive liquidation = penalized
total_liquidated	+	Complex signal: risky, but active
num_liquidated	−	High liquidation count = penalized

Scoring Formula

We rescaled the PC1 values into a 0–100 range using:

Score=100×PC1−min(PC1)max(PC1)−min(PC1)\text{Score} = 100 \times \frac{\text{PC1} - \min(\text{PC1})}{\max(\text{PC1}) - \min(\text{PC1})}Score=100×max(PC1)−min(PC1)PC1−min(PC1)

This normalization ensures interpretability and ranking consistency. A score of 100 represents the most engaged, well-behaved wallets, while a score of 0 denotes either inactive or exploitative wallets.

---

6. Behavioral Patterns Observed

High Scoring Wallets (Score > 98)

These wallets exhibited:

- Substantial `net_deposit` and moderate borrowing
- Multiple days of activity (`active_days` > 2000)
- Low or zero `liquidator_profit` — not profiting from others' loss

- No liquidation history
- Diversified across multiple assets

They represent long-term users who likely provide protocol liquidity, test various markets, and borrow responsibly — all healthy behaviors.

Low Scoring Wallets (Score < 1)

These wallets:

- Had either **0 deposits** or **negative net deposits** (indicating net withdrawals or flash loans)
- Zero borrowing or lending
- Earned significant `liquidator_profit` but had minimal or zero `active_days`
- Often had <10 transactions in total

These are likely **sybil actors**, flash loan bots, or extractive liquidators — behaviors not conducive to the health of the protocol.

---

7. Rationale & Justification

Why Not Just Use Raw Metrics?

Raw metrics can be misleading:

- High `liquidator_profit` may suggest skill, but often masks short-term bots.
- A wallet with 10,000 transactions could be spam or automation unless contextualized by `active_days`.
- Net deposit or borrow alone doesn't reflect intent or loyalty.

PCA acts as a *filter* to weigh these signals meaningfully, balancing the trade-offs.

Why Penalize Liquidators?

The intention isn't to discredit necessary protocol functions but to discourage actors whose **sole purpose is extraction**, not participation. Liquidators with no history of borrowing or supplying are likely capitalizing on system vulnerabilities.

---

8. Limitations & Future Work

- PCA assumes linear relationships. Non-linear techniques (e.g., t-SNE, UMAP) might uncover hidden clusters.
- Time-series analysis (e.g., volatility of deposits/borrows) could enrich the behavioral fingerprint.
- Flash loan identification could be made explicit and added as a feature.

---

9. Conclusion

The scoring system presented here offers a principled, interpretable, and behaviorally grounded way to distinguish between **healthy long-term protocol participants** and **short-term extractive agents**. Through feature engineering and PCA-based scoring, we highlight patterns that align with **protocol sustainability**, rewarding those who meaningfully contribute.