# Risk Scoring System: Compound Wallets on Ethereum

## 1. Data Collection Method

We used the Covalent API to gather real-time on-chain financial activity for Ethereum wallets. Specifically, we queried each wallet's activity in the Compound V2 protocol, which provides data on:  
- Assets supplied (lending)  
- Assets borrowed (debt)  
- The underlying balance in USD for each  
- Total assets interacted with  
  
Each request returned structured JSON with balances, which we parsed into quantitative features.

## 2. Feature Selection Rationale

We focused on financial behavior indicators that reflect how risky a wallet's position is:

|  |  |  |
| --- | --- | --- |
| Feature | Description | Rationale |
| total\_supply | Total USD value of tokens supplied (lent) | Indicates liquidity and potential stability |
| total\_borrow | Total USD value of tokens borrowed | Directly reflects leverage or debt |
| net\_position | Difference between supply and borrow | Negative value suggests higher risk |
| num\_assets | Number of unique assets used in lending/borrowing | Higher diversity implies complexity or risk spreading |

## 3. Scoring Method

We developed a custom risk score scaled from 0 to 1000 using normalized financial features:  
  
1. All four features were Min-Max normalized.  
2. We assigned weights based on their contribution to risk:  
  
 score = (total\_borrow\_norm \* 0.4) + (1 - total\_supply\_norm) \* 0.2 + (1 - net\_position\_norm) \* 0.3 + (1 - num\_assets\_norm) \* 0.1  
  
3. We inverted supply, net position, and asset count since higher values here lower risk, while higher borrow increases risk.  
4. Finally, we scaled the score:  
  
 df['score'] = score \* 1000  
  
This ensures that more leveraged and less diversified wallets score higher, i.e., are riskier.

## 4. Justification of Risk Indicators

- Borrowed Amount (`total\_borrow`) is a direct risk — more borrowed means more owed.  
- Net Position (supply - borrow) reflects safety buffer. A negative net means high exposure and vulnerability to liquidation.  
- Total Supply shows financial health; more supplied assets reduce risk of default.  
- Number of Assets indicates portfolio spread — overexposure to few assets increases volatility risk.  
  
These indicators balance accuracy and simplicity, making the model easy to scale across thousands of wallets with minimal data dependency.