# Round 2 Assignment: Wallet Risk Scoring From Scratch - Documentation

Author: Shivam Singh
Date: 26/07/2025

## Overview

This project aims to evaluate the risk associated with various Ethereum wallets by analyzing their on-chain DeFi activity, specifically on the Compound protocol. Using transaction summaries and position data, a comprehensive risk profile is created for each wallet. The final deliverable is a CSV file with wallet addresses and corresponding risk scores ranging from 0 to 1000.

## 1. Data Collection Method

- Source: Moralis API
- Chain: Ethereum Mainnet (chain_id=1)
- Data Fetched:
  - Protocol Summary (/defi/summary)
  - Compound V2 Positions (/defi/compound/positions)
  - ERC20 Token Balances (/erc20)
- Tools Used: Python, requests, tqdm, pandas
- Fallback: Random fallback features if data is missing

## 2. Feature Selection Rationale

| Feature | Description | Risk Signal |
|---|---|---|
| Leverage Risk | Borrowed/Supplied assets | High leverage = Higher risk |
| Liquidation Risk | Past liquidations | More = Higher risk |
| Portfolio Size | Total position value (USD) | Larger portfolio = Higher risk |
| Diversification | Count of tokens held | Low diversification = Higher risk |
| Health Risk | Proxy for health factor | Lower health = Higher risk |

## 3. Scoring Method

- Base Score: Weighted sum of 5 features
- Weights: Leverage (30%), Liquidation (25%), Size (20%), Diversification (15%), Health (10%)
- KMeans clustering (k=5) added to adjust scores based on wallet groups
- Final scores normalized to 0–1000
- Random forest model used for refinement (depth=10, trees=100)

## 4. Justification of Risk Indicators

- Compound usage is central to DeFi lending
- Leverage & liquidation are direct risk indicators

- Large portfolio = larger liquidation impact
- Low diversification = concentrated risk
- Health factor simulates on-chain stability

## 5. Deliverables

- wallet_risk_scores.csv: wallet_id, score
- wallet_risk_scores_detailed.csv: extended feature data
- Visuals: Histogram + Pie Chart

## Status

- Code Cleaned ✅
- Scoring Model ✅
- Feature Engineering ✅
- API Handling ✅
- Deliverables ✅

## Risk Score Visualizations