**ML-Based Credit Score Analysis Report**

**This document provides an analysis of credit scores generated using an unsupervised machine learning model (K-Means clustering) applied to Aave V2 wallet behaviour data.**

**Score Distribution**

**Wallets were grouped into 5 clusters using K-Means and mapped to score ranges:**

| **Cluster ID** | **Score** | **Number of Wallets** | **Behaviour Summary** |
| --- | --- | --- | --- |
| **0** | **200** | **TBD** | **Low interaction, often liquidated, single asset usage** |
| **1** | **400** | **TBD** | **Low repayment ratio, moderate borrowing, few redeems** |
| **2** | **600** | **TBD** | **Average behaviour, some deposits, repayments, mixed risk** |
| **3** | **800** | **TBD** | **Reliable with consistent deposits, active usage, diversified assets** |
| **4** | **1000** | **TBD** | **Top-tier users, high deposits, full repayments, active and diverse** |

**Replace TBD with actual counts using features\_df['cluster'].value\_counts() if analyzing live.**

**Behavior Insights by Score Range**

**0–200 (Cluster 0)**

* **Minimal deposit or repay activity**
* **Often appear in liquidation events**
* **Use only 1 or 2 assets**
* **Long inactivity gaps between transactions**

**400–600 (Clusters 1 & 2)**

* **Modest interaction volume**
* **Some borrow/repay, not fully covered**
* **Mixed levels of activity and asset diversity**
* **Few liquidations or redemptions**

**🟢 800–1000 (Clusters 3 & 4)**

* **Large deposit and repay volumes**
* **High asset diversity (5+ tokens)**
* **Daily to weekly activity**
* **Consistently avoid liquidation**
* **High repay-to-borrow ratios**

**Visualization Suggestion**

**import matplotlib.pyplot as plt**

**plt.hist(features\_df['credit\_score'], bins=[200,400,600,800,1000], edgecolor='black')**

**plt.title("Credit Score Distribution from KMeans Clustering")**

**plt.xlabel("Credit Score Range")**

**plt.ylabel("Number of Wallets")**

**plt.grid(True)**

**plt.show()**

**Conclusion**

**Using clustering to assign scores based on behavior lets us:**

* **Identify risk-prone users vs. responsible actors**
* **Quantify on-chain behavior in a score without needing labels**
* **Create scalable heuristics for lending, whitelisting, or fraud prevention**

**This method is modular and extensible to more DeFi protocols and time series analysis.**