Range Data Science Project

Analyze a dataset of blockchain transactions to uncover insights, detect anomalies, and explore clustering/grouping of addresses based on activity patterns.

Dataset Provided:
Set of Solana payments containing addresses, transaction hashes, block heights, token types, and amounts.

Field Explanations:
- sender/recipient/amount: These fields represent the sender, recipient, and amount of each payment—self-explanatory.
- token: This represents the currency used in the transaction. Similar to traditional currencies like USD or EUR, the values in this field represent different cryptocurrencies.
- height: The dataset does not include timestamps for payments but instead provides block heights. Each new block is assigned the next sequential height, with blocks generated every 400ms. -> Consecutive heights are 400ms apart.
- tx_hash: Transactions sharing the same tx_hash were created by the same entity and committed to the blockchain simultaneously.

Here's the file (200MB):
https://drive.google.com/file/d/1i1DP27sFfy-xAlPm48JqsIXKyfVbCnlo/view?usp=sharing

Tasks (Choose one or more):
1. **Clustering & Address Profiling:**
   a. Group addresses based on transaction behavior (e.g., frequent transactors, large-value movers, inactive accounts, etc.).
   b. Suggest and use clustering algorithms for the task.

2. **Anomaly Detection:**
   a. Identify outliers (e.g., abnormally high-value transfers, sudden spikes in activity).
   b. Suggest and use statistical methods for the task.

3. **Network Analysis:**
   a. Construct a transaction graph and analyze network properties (e.g., centrality, connected components).
   b. Identify potential hubs or influential addresses.

Deliverables:
A short write-up explaining your findings, methodology, and assumptions.
Code snippets or visualizations to support your analysis.