Fraud Analysis

The purpose of this analysis is to discover user IDs that are related to IDs suspected of being involved in fraudulent operations.

The dataset contains information about devices and personal user accounts potentially involved in cross-bank fraud. The two known compromised entities are:

- Device ID: 91b12379-8098-457f-a2ad-a94d767797c2 (Bank4)
- User Account: 0007f265568f1abc1da791e852877df2047b3af9 (Bank8)

Dataset Columns

The dataset includes the following fields:

- device_id - Unique device identifier
- identity - Unique user identifier
- bank - Identifier of the bank linked to the user
- device_fingerprint - Unique digital signature of the user's device
- gpu_vendor - Manufacturer of the device's GPU
- screen - Screen resolution
- os - Operating system
- browser - Browser used by the user
- ips - IP addresses used by the device or user

Project Steps

The main steps in this project include:

1. Data Preparation:

- Load and clean the dataset.

- Normalize fields and handle missing values.

2. Extract Information Based on IP:

- Parse and split IPs if necessary.

- Geolocate IPs to find their origin.

- Detect IP reuse across different accounts/devices.

3. Extract Related Accounts:

- Map all accounts linked to the compromised device and user.

- Use graphs to show connections through shared devices, IPs, and fingerprints.

4. Visualization:

- Generate graphs using networkx.

- Show distribution of IP usage, account connections, and clustering.

5. Reporting:

- Summarize suspicious clusters.

- Export graphs and findings into the final report.