Cross-Bank Fraud Detection Documentation

Step 1: Environment Setup & IP Enrichment

I began by installing the required Python packages and tools necessary for data handling, API calls, and profiling.

Then, using the `ipinfo.io` API, I enriched the original IP dataset by querying each IP address to extract additional metadata including:

- IP location
- country
- City
- Timezone
- Org

Profile 1: Identity & Bank Usage Analysis

The dataset was grouped by identity and bank usage to detect potential cross-bank fraud activity. Key findings:

- Multiple identities appeared in more than one bank.
- An anonymous identity (`"-"`) was linked to 11 banks, which is highly anomalous.

Sample output:

identity	bank_count	anomaly
2577384	1	1
13944790	2	1

15330202	2	1
-	11	-1

Profile 2: Shared Identities fingerprints and Device Fingerprint Clustering

This stage focused on identifying:

- Identities reused across multiple banks: 11 such identities were found.
- Overlap between specific banks (e.g., Bank8 & Bank9 shared the anonymous identity "-").
- Anonymous identity's fingerprints: A single anonymous identity used 17 distinct device fingerprints, further suggesting automation or spoofing.

These fingerprints were distributed across many banks:

Bank8: 398

Bank9: 247

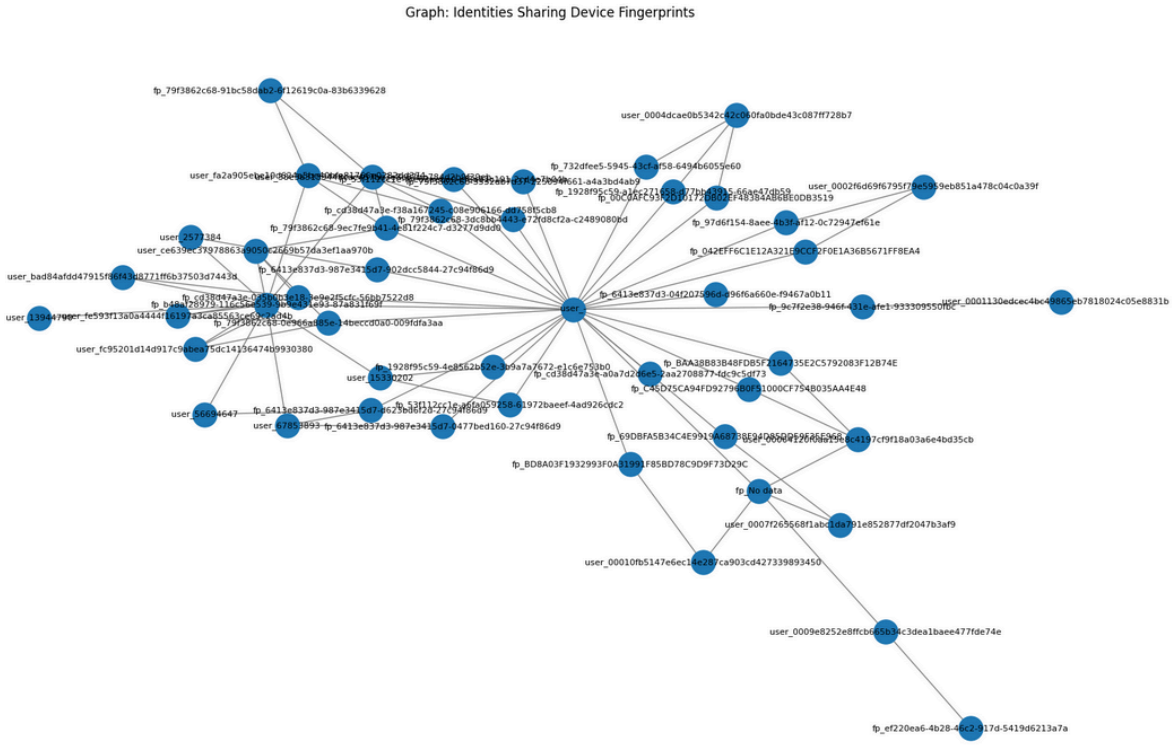
Bank2: 224

…

So :

- Clusters of users (identities) using the same devices, which is unlikely under normal behavior.

- Cluster sizes varied, with some as large as 168.
- A network graph was created to visualize these overlaps.



Additionally, the clustering helped uncover links like:

- Identity 2577384 and anonymous "-" sharing the same device.

- Identity 37294473 linked to a device also seen in other profiles.

Profile 3: Bot/Emulator Behaviour Detection

From the enriched metadata, 708 suspicious environments were flagged based on:

- Outdated Android WebView and iOS WKWebView versions.
- Unusual or fixed screen resolutions (e.g., `(720, 360, 24)`).
- Missing GPU renderer data.

Top patterns:

- `Chrome Mobile WebView 78.0–79.0` (Android)
- `Mobile Safari UI/WKWebView 13.x` (iOS)
- `PowerVR Rogue GE8300` and missing GPU renderers

This strongly suggests emulator farms or automated browser agents.

Fingerprints reused across 5+ identities:

	device_fingerprint	identity_count
1	1928f95c59-a1ec271658-d77bb43915-66ae47db59	116
3	C45D75CA94FD92796B0F51000CF754B035AA4E48	195

Profile 4: Fingerprint Reuse Across Multiple Identities

This step examined how many unique identities shared the same device fingerprint:

Some fingerprints were reused across 100+ identities, suggesting account farming or scripted identity spoofing.

Fingerprint: 1928f95c59-a1ec271658-d77bb43915-66ae47db59

Used by: 116 identities

Fingerprint: C45D75CA94FD92796B0F51000CF754B035AA4E48

Used by: 195 identities

🗒️ Profile 5: Fraud Ring Clusters via Shared IPs and Location Anomalies

All features (device, IP, fingerprint, cluster, and geo-data) were combined into a single dataset. Highlights:

- Dataset shape: (10,248 rows, 16 columns)
- Saved file: `profile5_many_identities_per_ip.xlsx`
- IPs used by 10+ identities: flagged and saved.
- Timezone vs OS mismatches: 4,109 detected — indicating user spoofing behavior.
- VPN-like ISPs or orgs: 87 cases identified.

Country mismatches:

flag: 9248 records (suspicious behaviour)

ok: 1000 records

This profile focuses on detecting coordinated fraud behaviors through analysis of shared IP addresses and enriched IP metadata. By merging the main user activity dataset with IP intelligence data from external enrichment, we investigated several red flags indicating suspicious behavior patterns.

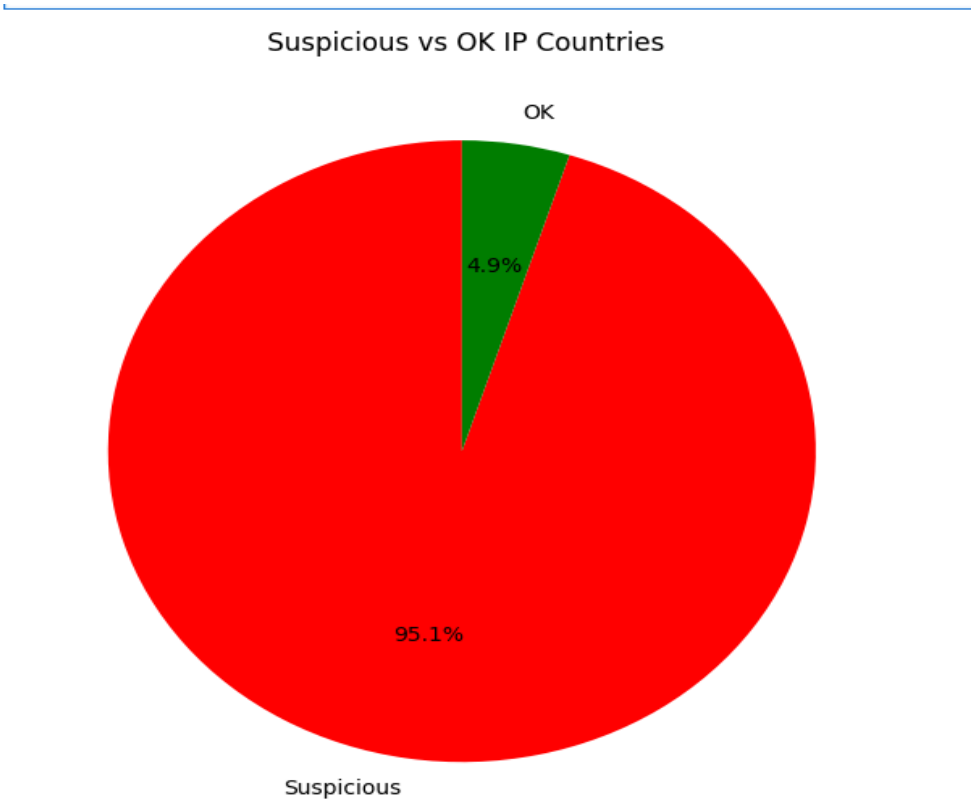
by exploding multi-IP fields and grouping by IP to identify addresses used by **10 or more unique identities**, flagging them as potentially shared infrastructure for account farms or fraud rings. These were saved separately for review.

To further enrich the risk signal, we detected **mismatches between timezones and operating systems** such as devices claiming `macOS` but appearing in `Russia` or `Asia` time zones which often indicate spoofed environments or VPNs.

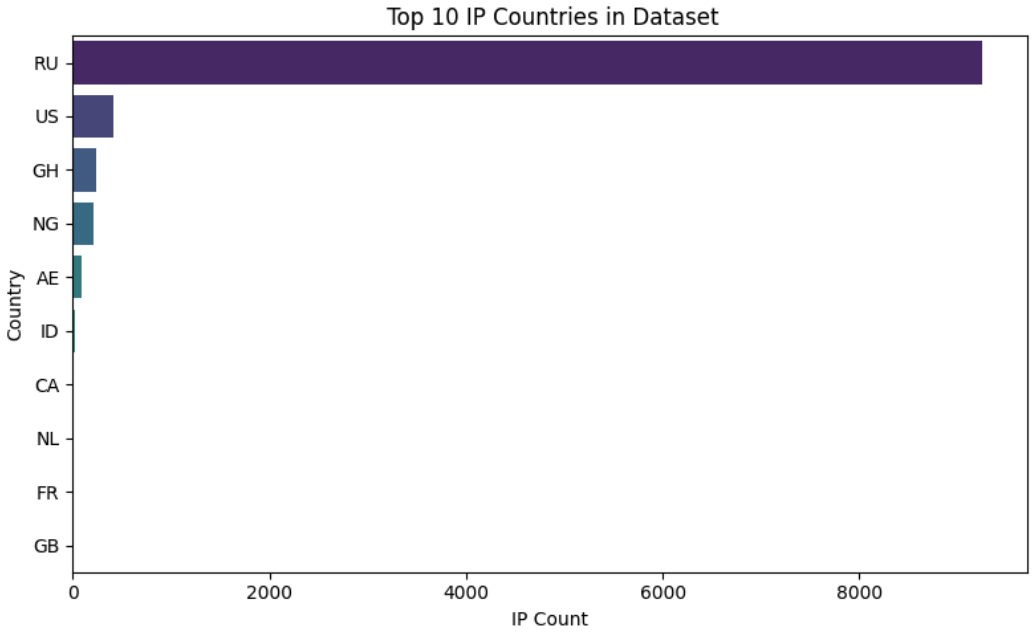
we flagged IPs associated with known **cloud infrastructure or VPN providers** like `Amazon`, `Google`, and `Contabo`, using organization keywords in the metadata. We also applied **country-level filtering**, marking IPs from high-risk or untrusted geolocations (e.g., `RU`, `CN`, `IR`) as suspicious.

Visual analysis was used to support these findings:

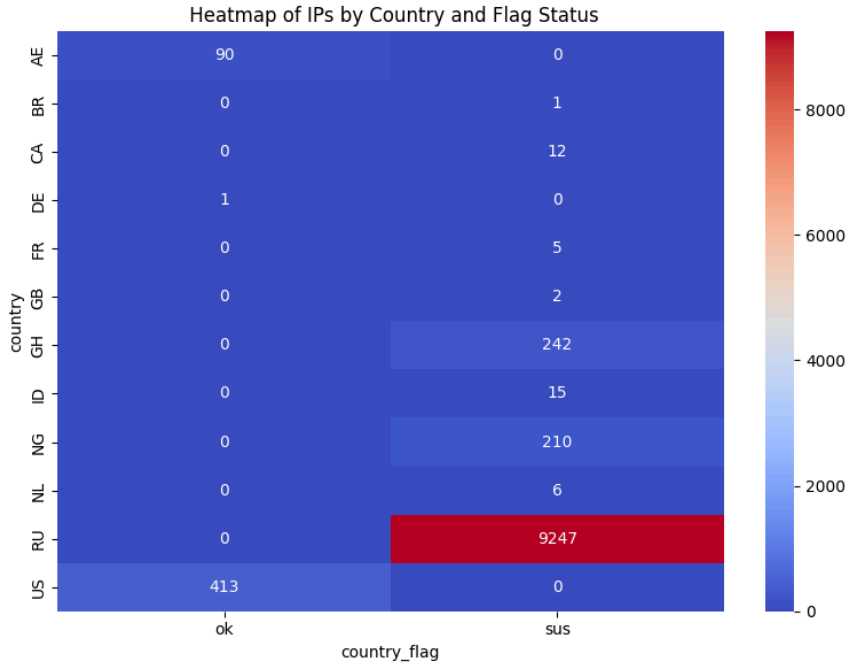
- A **pie chart** of suspicious vs OK IP geolocations.



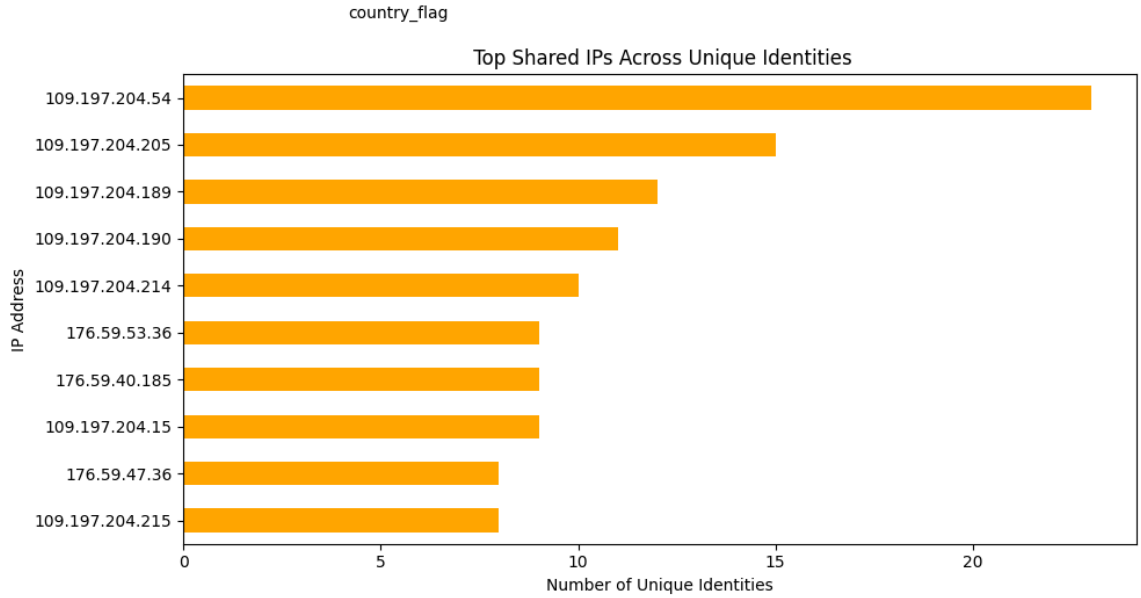
- A **bar chart** of the top IP-originating countries.



- A **heatmap** showing the distribution of flagged IPs by country.



- A **horizontal bar chart** identifying the top IPs used across the most unique identities.



This layered approach allowed us to uncover coordinated fraud networks, VPN usage, and location spoofing, which traditional rule-based fraud engines often miss.

Conclusion

The layered profiling approach uncovered several strong indicators of coordinated fraudulent activity:

- Reused device fingerprints across banks,
- Anonymous or shared identities,
- IPs and devices linked to automation or emulator signatures,
- Mismatches in timezone, country, and OS all pointing to behavior outside typical user interaction.