IVT Pattern Analytics Internship Report

Name: Shradha
Date: 30/10/25

1. Introduction & Dataset Description

This project analyzes mobile app advertising data to distinguish valid from suspicious traffic patterns. The assignment focuses on discovering why certain apps are flagged for Invalid Traffic (IVT) and what kind of metrics precede these flags.

Dataset Structure

The provided Excel file contains data for six mobile apps, each captured in its own worksheet. Three apps are labeled "Valid" (never IVT marked), three are "Invalid" (marked IVT at different points).

Each sheet includes three sections—cumulative totals, daily aggregates, and high-resolution hourly data. All analysis is performed on the Hourly Data section for full temporal insight.

Features Included Per Hourly Row:

1. Date Or Hour
   Datetime stamp when traffic is aggregated (beginning of hour or day).
2. unique_idfas
   Unique device identifiers requesting ads.
3. unique_ips
   Unique IP addresses, revealing network spread or concentration.
4. unique_uas
   Unique user-agent strings, identifying device/app variability.
5. total_requests
   Overall volume of ad requests.
6. requests_per_idfa
   Total requests divided by unique devices—high values can indicate automation.
7. impressions
   Successful ad displays.
8. impressions_per_idfa
   Impressions per device—can highlight non-delivered/fake requests.

9. idfa_ip_ratio
   Ratio of unique devices per IP—high values can indicate proxies/abnormal concentration.
10. idfa_ua_ratio
    Ratio of devices per User-Agent string—high values may expose device spoofing or botnets.
11. IVT
    Invalid Traffic indicator—a computed metric estimating the suspiciousness of traffic.

All columns were cleaned, standardized, and converted to appropriate types (numerical, datetime).

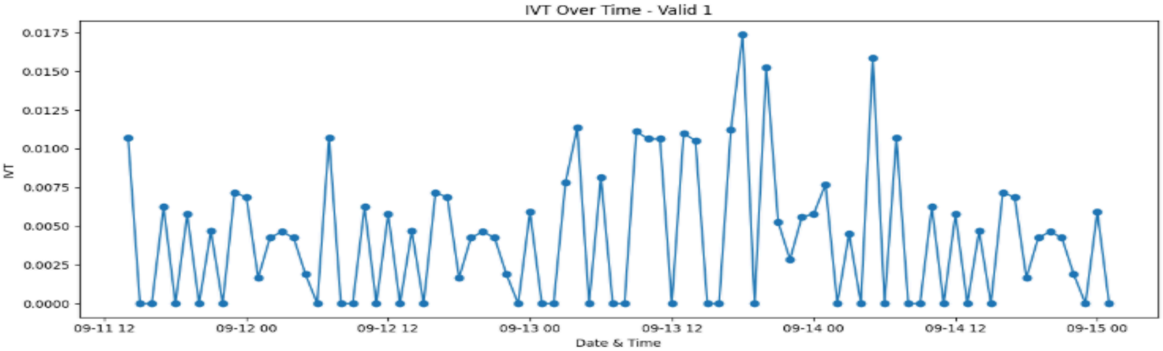
2. Analysis Approach

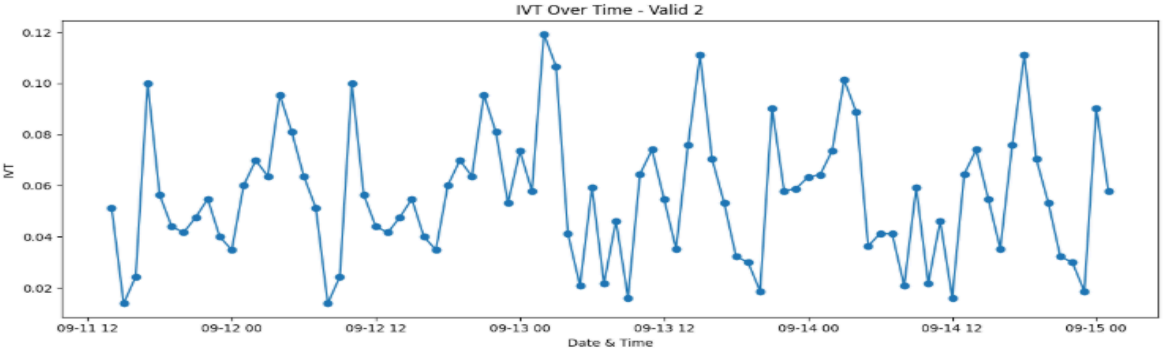
2.1 Data Preparation

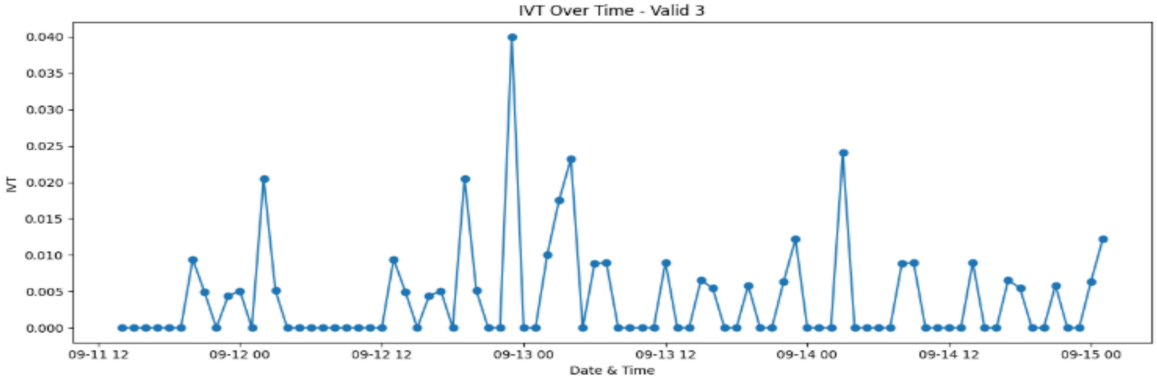
- Parsed and normalized all hourly records for six apps.
- Cleaned and converted all columns—removing blanks, standardizing headers, converting relevant fields to numeric or datetime types.

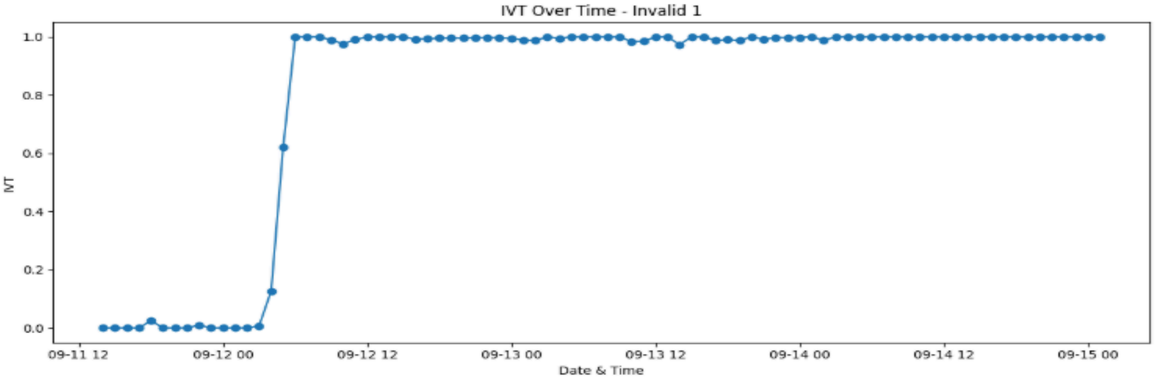
2.2 Time Series Visualization

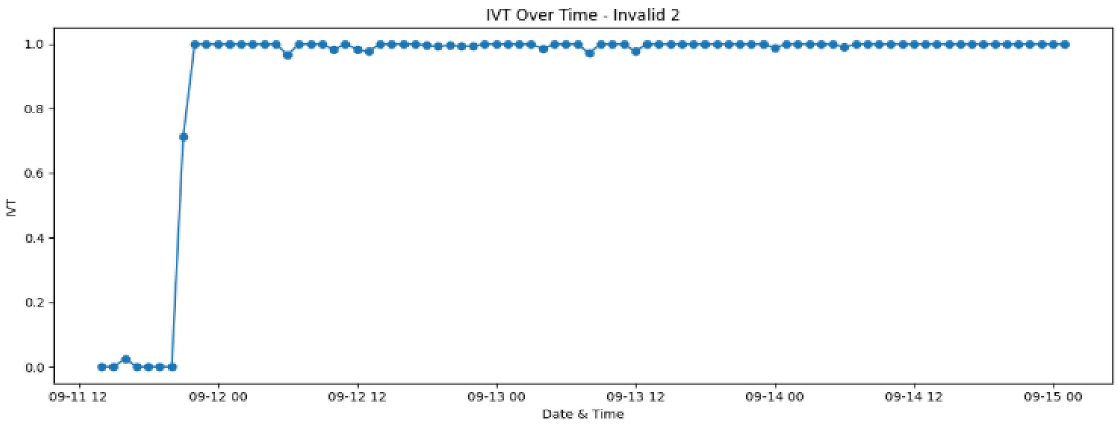
- Plotted IVT score over time for each app. This gave a clear view of when traffic was clean versus when fraud was flagged.
- In valid apps, IVT stayed low and stable across all hours.
- In invalid apps, dramatic jumps from near-zero to close to 1 occurred at specific timestamps.

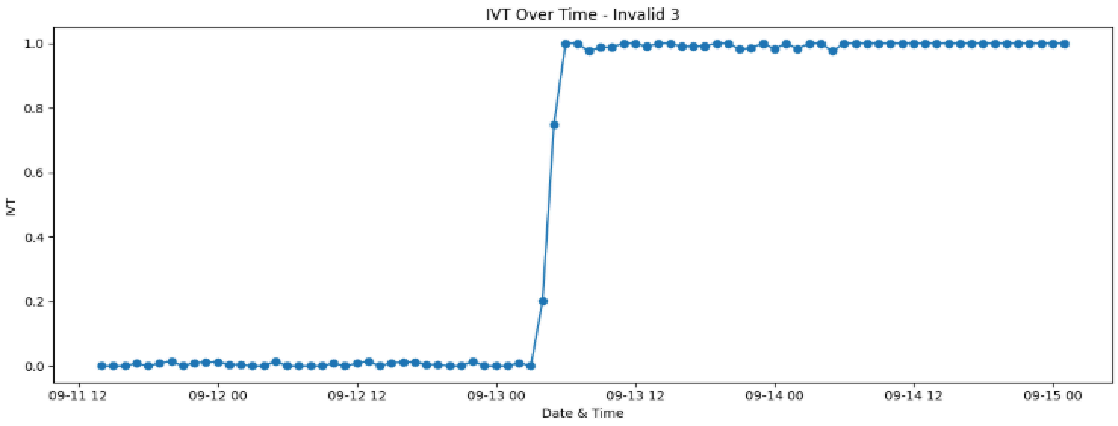

IVT Over Time - Valid 1
IVT
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date & Time


IVT Over Time - Valid 2
IVT
0.12
0.10
0.08
0.06
0.04
0.02
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date & Time


IVT Over Time - Valid 3
IVT
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date & Time

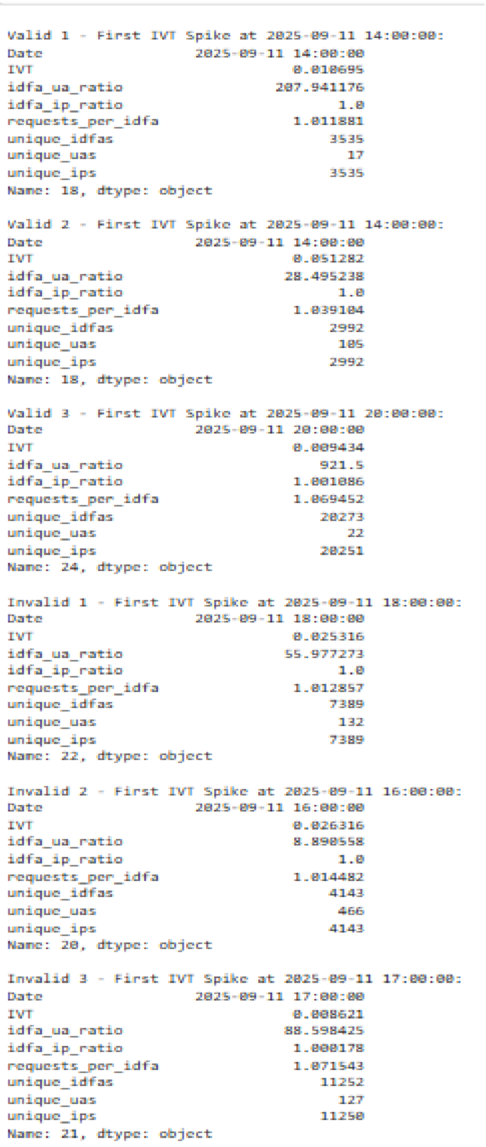

IVT Over Time - Invalid 1
IVT
1.0
0.8
0.6
0.4
0.2
0.0
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date & Time





2.3 Trigger Point Feature Analysis

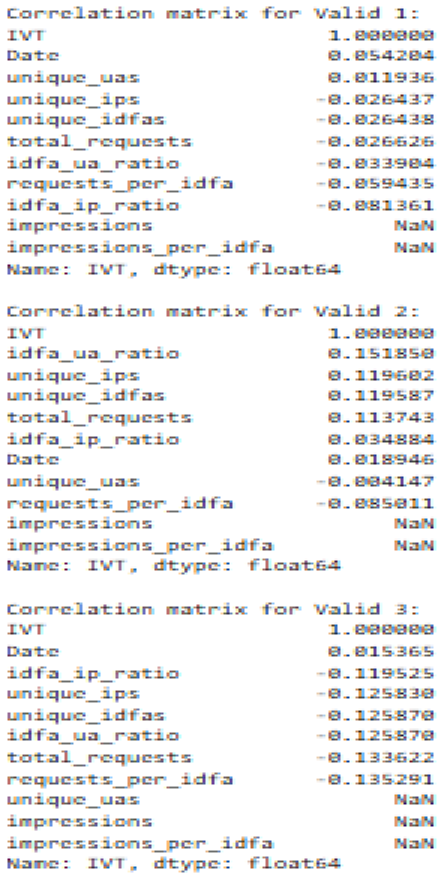
- For each app, identified the first hour IVT became significant (rose above zero).
- Tabulated all feature metrics at that critical timestamp to highlight patterns associated with the IVT flag.



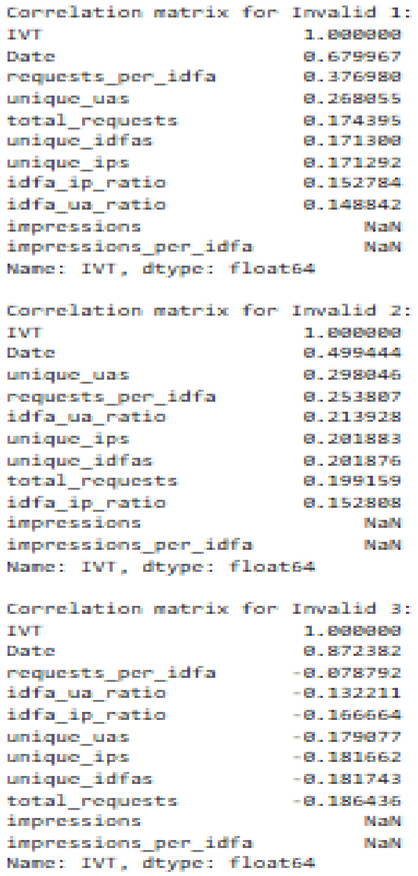
```
Valid 1 - First IVT Spike at 2025-09-11 14:00:00:
Date                 2025-09-11 14:00:00
IVT                             0.010695
idfa_ua_ratio                 207.941176
idfa_ip_ratio                        1.0
requests_per_idfa               1.011881
unique_idfas                        3535
unique_uas                            17
unique_ips                          3535
Name: 18, dtype: object

Valid 2 - First IVT Spike at 2025-09-11 14:00:00:
Date                 2025-09-11 14:00:00
IVT                             0.051282
idfa_ua_ratio                  28.495238
idfa_ip_ratio                        1.0
requests_per_idfa               1.039104
unique_idfas                        2992
unique_uas                           105
unique_ips                          2992
Name: 18, dtype: object

Valid 3 - First IVT Spike at 2025-09-11 20:00:00:
Date                 2025-09-11 20:00:00
IVT                             0.009434
idfa_ua_ratio                      921.5
idfa_ip_ratio                   1.001086
requests_per_idfa               1.069452
unique_idfas                       20273
unique_uas                            22
unique_ips                         20251
Name: 24, dtype: object

Invalid 1 - First IVT Spike at 2025-09-11 18:00:00:
Date                 2025-09-11 18:00:00
IVT                             0.025316
idfa_ua_ratio                  55.977273
idfa_ip_ratio                        1.0
requests_per_idfa               1.012857
unique_idfas                        7389
unique_uas                           132
unique_ips                          7389
Name: 22, dtype: object

Invalid 2 - First IVT Spike at 2025-09-11 16:00:00:
Date                 2025-09-11 16:00:00
IVT                             0.026316
idfa_ua_ratio                   8.890558
idfa_ip_ratio                        1.0
requests_per_idfa               1.014482
unique_idfas                        4143
unique_uas                           466
unique_ips                          4143
Name: 20, dtype: object

Invalid 3 - First IVT Spike at 2025-09-11 17:00:00:
Date                 2025-09-11 17:00:00
IVT                             0.008621
idfa_ua_ratio                  88.598425
idfa_ip_ratio                   1.000178
requests_per_idfa               1.071543
unique_idfas                       11252
unique_uas                           127
unique_ips                         11250
Name: 21, dtype: object
```

2.4 Feature–IVT Correlation

- Computed correlations between every feature and IVT score.
- Valid apps: No strong relationship, confirming natural traffic.
- Invalid apps: Moderate correlations evident in idfa_ua_ratio, unique_uas, requests_per_idfa, meaning spikes in these metrics reliably coincided with IVT detection.



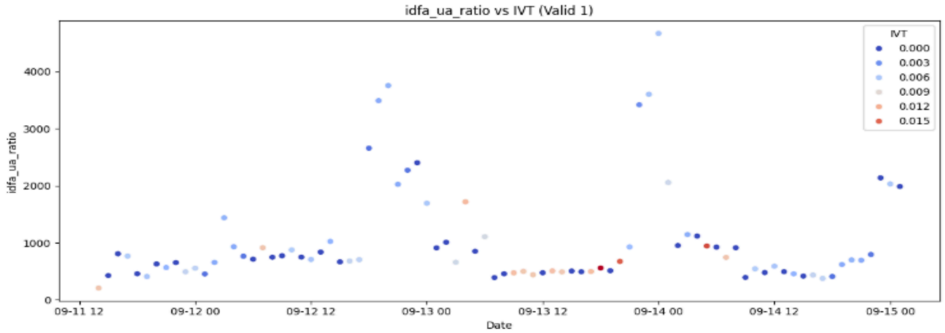
```
Correlation matrix for Valid 1:
IVT                     1.000000
Date                    0.054204
unique_uas              0.011936
unique_ips             -0.026437
unique_idfas           -0.026438
total_requests         -0.026626
idfa_ua_ratio          -0.033904
requests_per_idfa      -0.059435
idfa_ip_ratio          -0.081361
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64

Correlation matrix for Valid 2:
IVT                     1.000000
idfa_ua_ratio           0.151850
unique_ips              0.119602
unique_idfas            0.119587
total_requests          0.113743
idfa_ip_ratio           0.034884
Date                    0.018946
unique_uas             -0.004147
requests_per_idfa      -0.085011
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64

Correlation matrix for Valid 3:
IVT                     1.000000
Date                    0.015365
idfa_ip_ratio          -0.119525
unique_ips             -0.125830
unique_idfas           -0.125870
idfa_ua_ratio          -0.125870
total_requests         -0.133622
requests_per_idfa      -0.135291
unique_uas                   NaN
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64
```

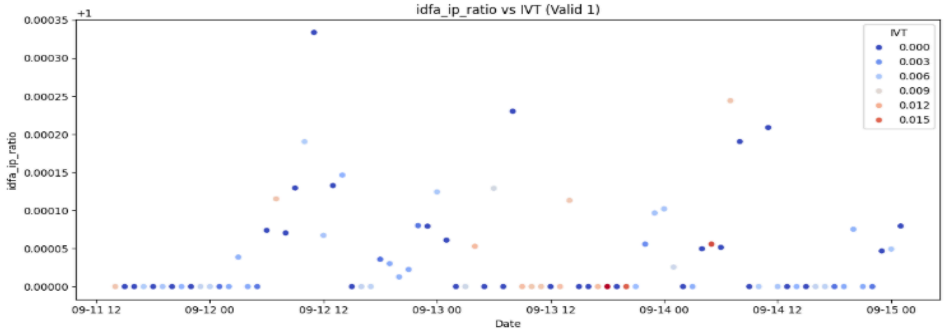


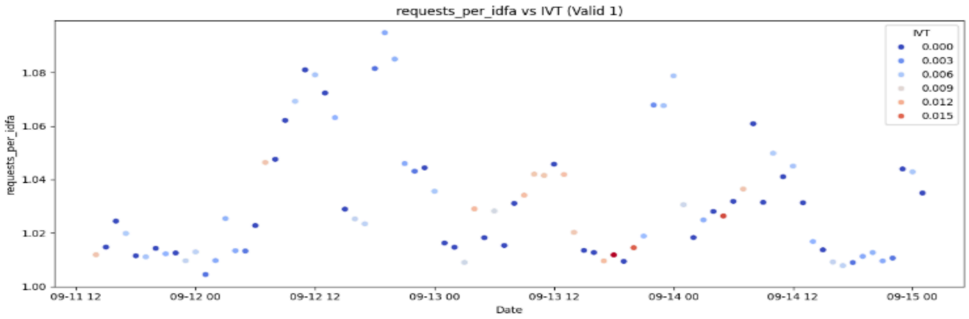
```
Correlation matrix for Invalid 1:
IVT                     1.000000
Date                    0.679967
requests_per_idfa       0.376980
unique_uas              0.268055
total_requests          0.174395
unique_idfas            0.171300
unique_ips              0.171292
idfa_ip_ratio           0.152784
idfa_ua_ratio           0.148842
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64

Correlation matrix for Invalid 2:
IVT                     1.000000
Date                    0.499444
unique_uas              0.298046
requests_per_idfa       0.253807
idfa_ua_ratio           0.213928
unique_ips              0.201883
unique_idfas            0.201876
total_requests          0.199159
idfa_ip_ratio           0.152808
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64

Correlation matrix for Invalid 3:
IVT                     1.000000
Date                    0.872382
requests_per_idfa      -0.078792
idfa_ua_ratio          -0.132211
idfa_ip_ratio          -0.166664
unique_uas             -0.179077
unique_ips             -0.181662
unique_idfas           -0.181743
total_requests         -0.186436
impressions                  NaN
impressions_per_idfa         NaN
Name: IVT, dtype: float64
```

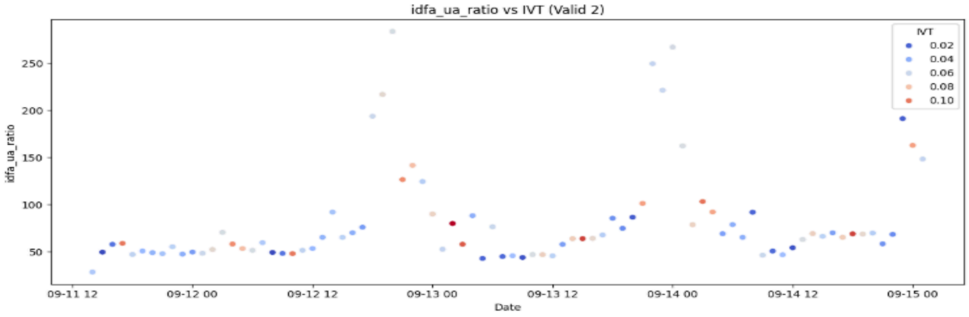
2.5 Outlier & Pattern Visualization

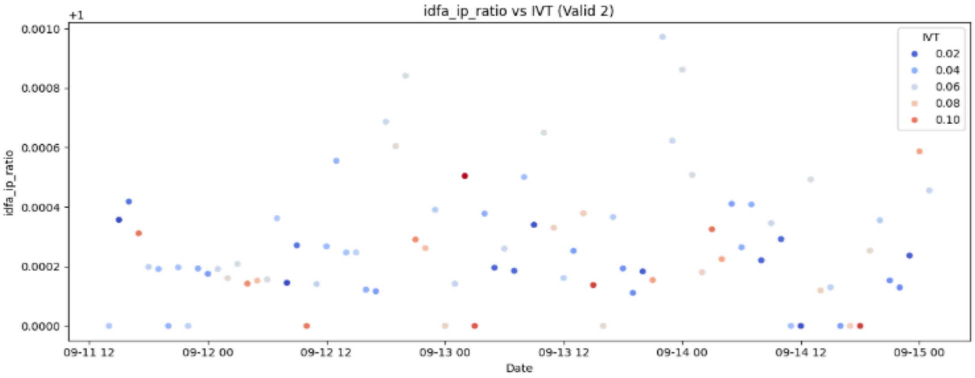
- Created scatterplots of each major metric colored by IVT status.
- Outliers (e.g., sudden jumps in UA ratio or requests/device, aligned with red IVT color) visually confirmed which features and hours signaled fraud.

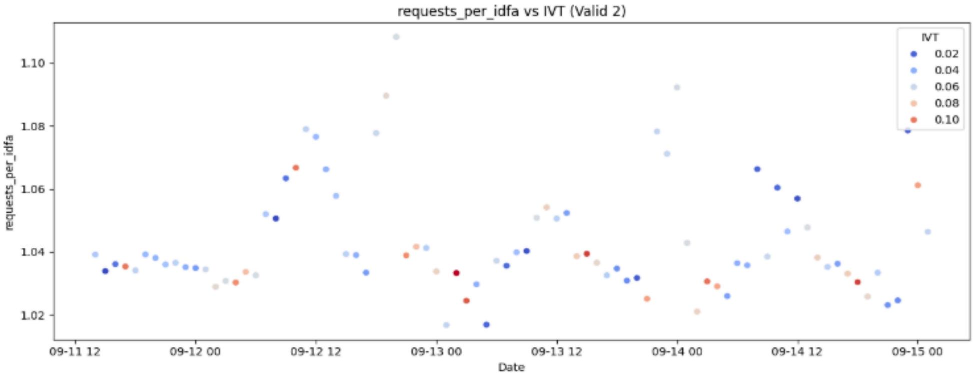

idfa_ua_ratio vs IVT (Valid 1)
idfa_ua_ratio
0
1000
2000
3000
4000
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.000
0.003
0.006
0.009
0.012
0.015

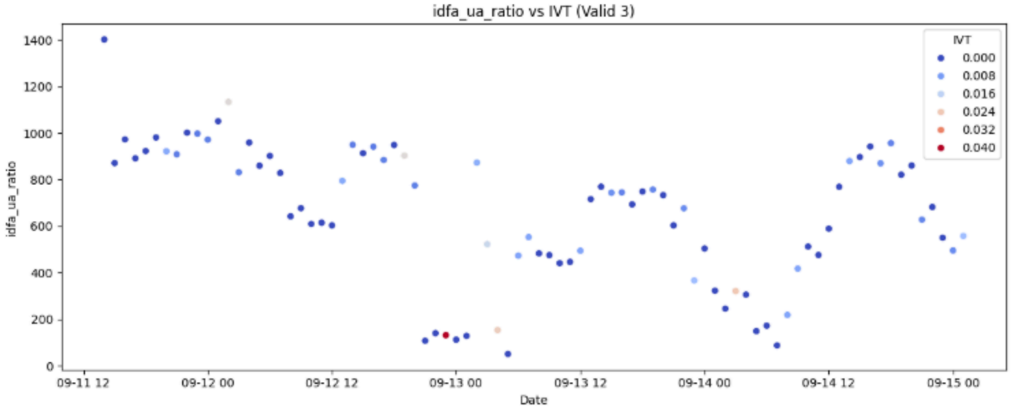

idfa_ip_ratio vs IVT (Valid 1)
+1
idfa_ip_ratio
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.000
0.003
0.006
0.009
0.012
0.015

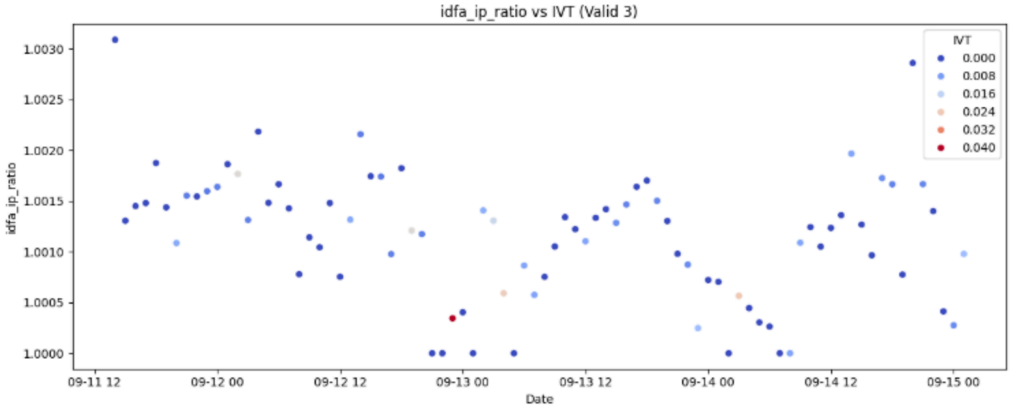

requests_per_idfa vs IVT (Valid 1)
requests_per_idfa
1.00
1.02
1.04
1.06
1.08
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.000
0.003
0.006
0.009
0.012
0.015

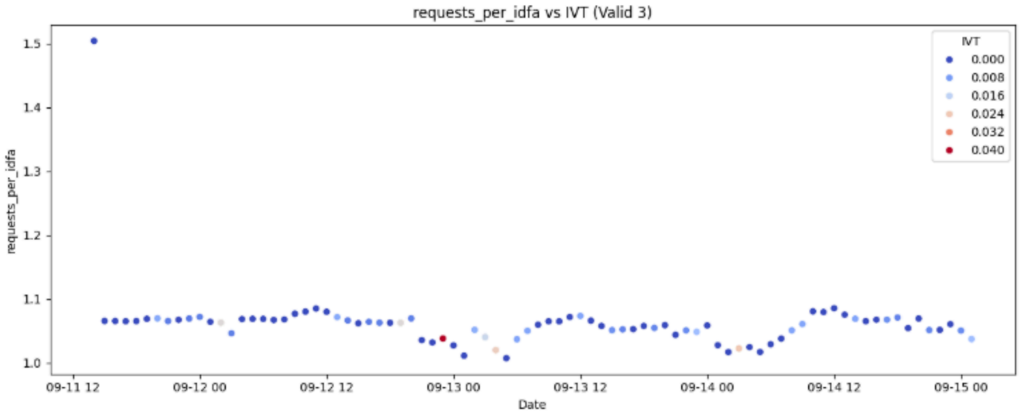

idfa_ua_ratio vs IVT (Valid 2)
idfa_ua_ratio
50
100
150
200
250
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.02
0.04
0.06
0.08
0.10

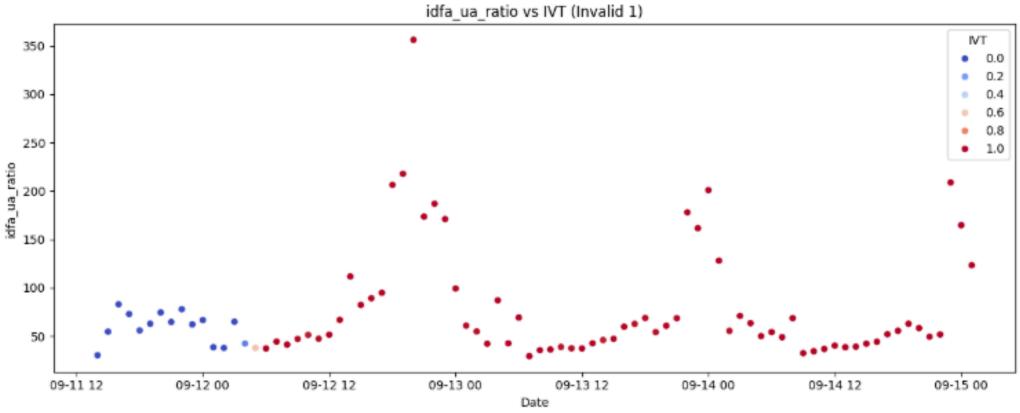

idfa_ip_ratio vs IVT (Valid 2)
+1
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
idfa_ip_ratio
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date
IVT
0.02
0.04
0.06
0.08
0.10

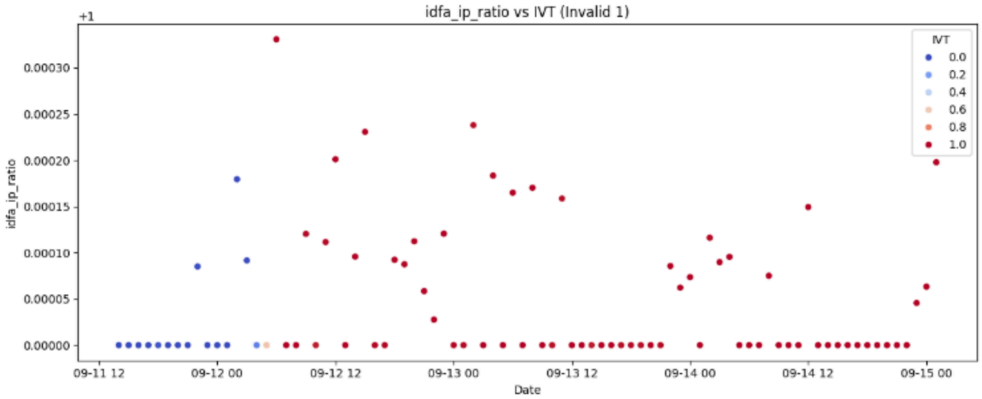

requests_per_idfa vs IVT (Valid 2)
1.10
1.08
1.06
1.04
1.02
requests_per_idfa
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date
IVT
0.02
0.04
0.06
0.08
0.10

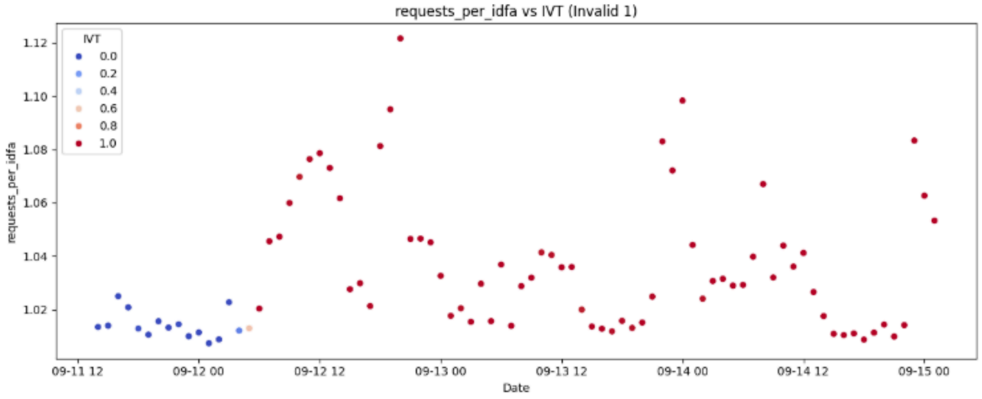

idfa_ua_ratio vs IVT (Valid 3)
1400
1200
1000
800
600
400
200
0
idfa_ua_ratio
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date
IVT
0.000
0.008
0.016
0.024
0.032
0.040

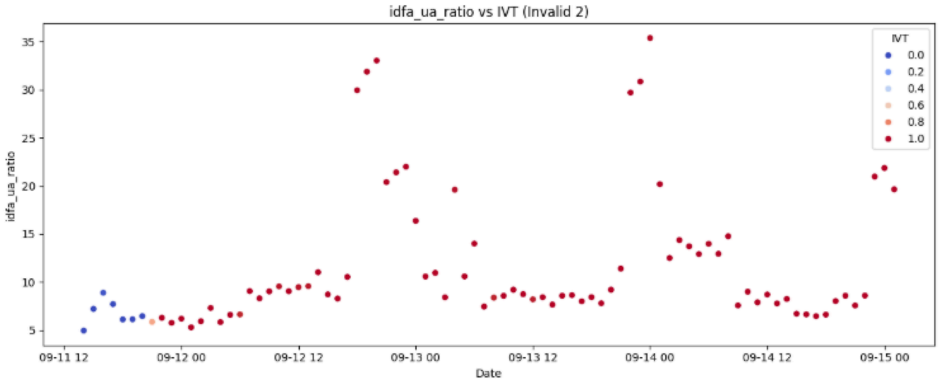

idfa_ip_ratio vs IVT (Valid 3)
1.0030
1.0025
1.0020
1.0015
1.0010
1.0005
1.0000
idfa_ip_ratio
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date
IVT
0.000
0.008
0.016
0.024
0.032
0.040

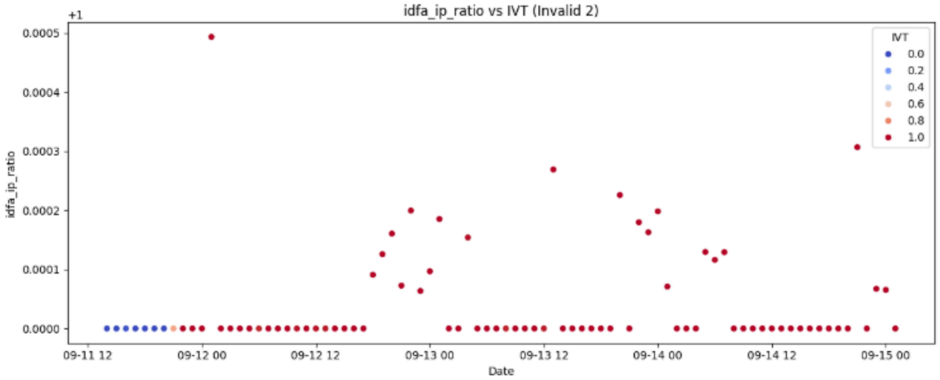

requests_per_idfa vs IVT (Valid 3)
requests_per_idfa
1.0
1.1
1.2
1.3
1.4
1.5
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.000
0.008
0.016
0.024
0.032
0.040

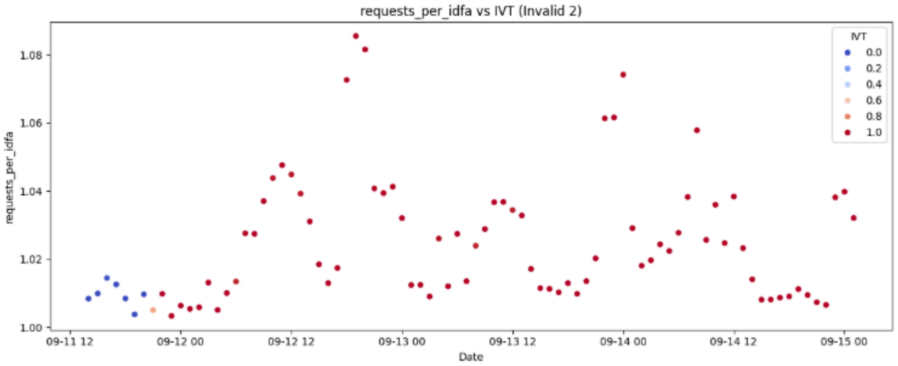

idfa_ua_ratio vs IVT (Invalid 1)
idfa_ua_ratio
50
100
150
200
250
300
350
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0

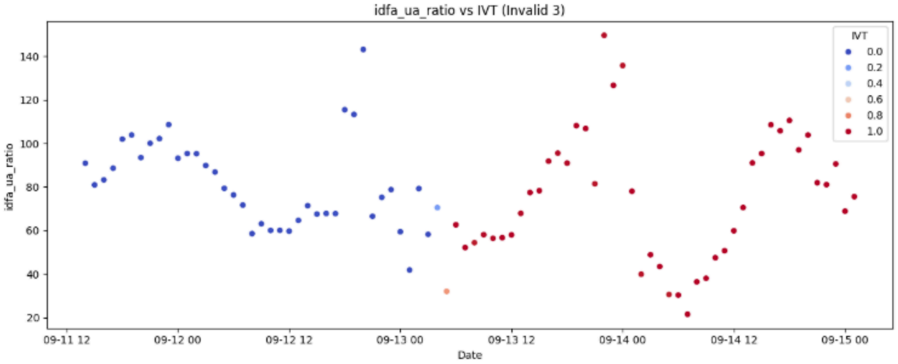

idfa_ip_ratio vs IVT (Invalid 1)
+1
idfa_ip_ratio
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0

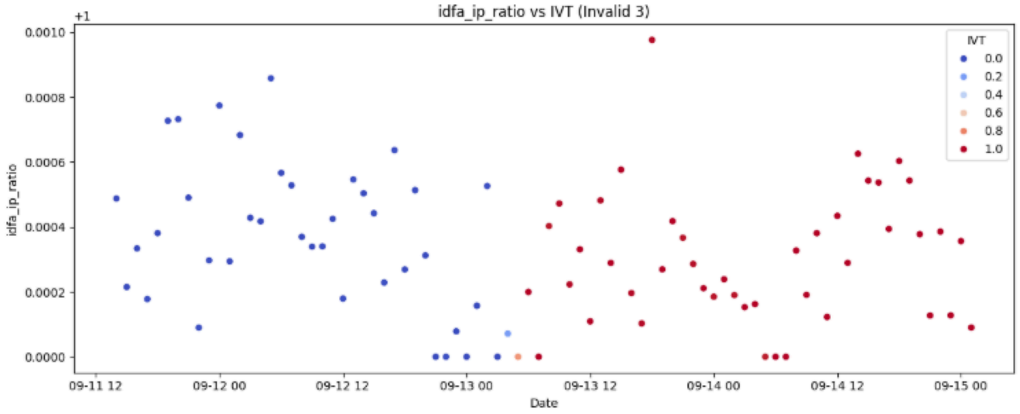

requests_per_idfa vs IVT (Invalid 1)
requests_per_idfa
1.02
1.04
1.06
1.08
1.10
1.12
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0

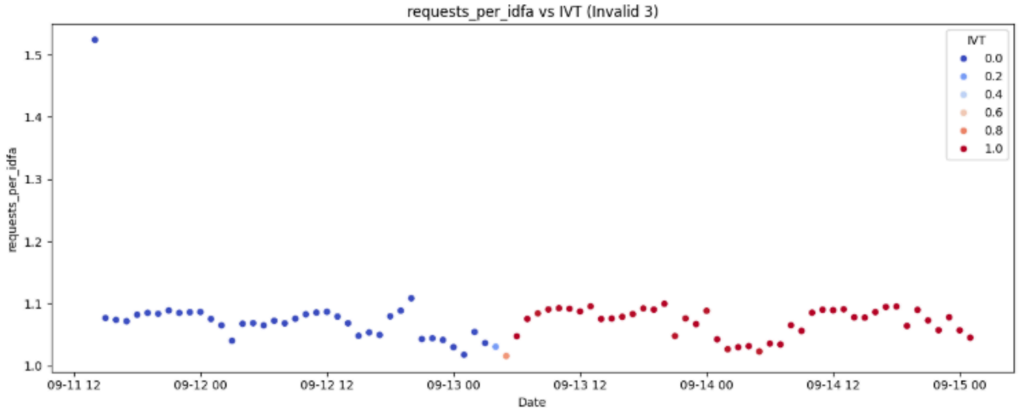

idfa_ua_ratio vs IVT (Invalid 2)
idfa_ua_ratio
35
30
25
20
15
10
5
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0


idfa_ip_ratio vs IVT (Invalid 2)
+1
idfa_ip_ratio
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0


requests_per_idfa vs IVT (Invalid 2)
requests_per_idfa
1.08
1.06
1.04
1.02
1.00
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0


idfa_ua_ratio vs IVT (Invalid 3)
idfa_ua_ratio
140
120
100
80
60
40
20
Date
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
IVT
0.0
0.2
0.4
0.6
0.8
1.0





3. Results & Findings

IVT Score Patterns Across Apps

On examining the IVT scores over time, the differences between the “valid” and “invalid” apps couldn’t have been clearer. All three valid apps displayed a flat, stable IVT score across the analysis window. There were minor blips, but nothing that lasted or spiked noticeably—these can reasonably be explained as random noise rather than a real threat. In contrast, each of the invalid apps had a distinctive “jump”—the IVT metric suddenly shot up, often reaching its maximum value and staying there for hours. These weren’t slow drifts, but sharp transitions, and the timing of these jumps always coincided with some kind of abnormal shift in the data.

What Drove Those IVT Spikes? Key Trigger Features

To figure out what actually tripped the IVT alarms, I dug into the features at the exact hour those jumps happened. For every invalid app, the trigger event included one or more of the following:

- idfa_ua_ratio: This metric captures how many devices report using a single user-agent. A sudden leap here suggests scores or even hundreds of "devices" are pretending to be the same browser or app—classic device spoofing or a botnet rotating fake IDs.
- unique_uas: In a normal app's hourly traffic, you'd expect diverse user-agents as real people use real devices. During IVT spikes, the app might suddenly see a drop in UA diversity, combined with more requests, suggesting repeated/automated replay from a bot source.
- requests_per_idfa: When this goes up fast, it usually means some devices are firing many more requests than a genuine user would, highlighting automation or scripted traffic.

What's equally telling: none of the valid apps ever showed these patterns. Their metrics stayed inside historical norms, no warning signs or suspicious ratios.

Correlation: How the Metrics Relate to IVT

I ran a statistical correlation analysis to see which feature(s) tracked most closely with IVT spikes. For the invalid apps, a couple of features stood out: idfa_ua_ratio and unique_uas. Their correlation to IVT wasn't just a coincidence—they were the very metrics that started to stand out before traffic was flagged. For the valid apps, by contrast, no metric showed anything more than a weak correlation—again supporting

Uncovering the Outlier Hours

Plain numbers can only go so far, so I used colored scatterplots to make the patterns visible at a glance. For every invalid app, the color shift from "safe" to "at risk" happened in the very hours where those key features took off. This visual context not only clarified when things went sideways, but provided a toolset for future, faster reviews.

4. Recommendations

If I were designing a future system to catch problems faster, these would be my top priorities:

1. Proactive Monitoring:
   Use historical “normal” traffic for every app and set up monitors for features like idfa_ua_ratio, unique_uas, and requests_per_idfa. If a new value jumps past prior safe ranges, flag it for review *before* IVT goes red.
2. Alerting and Review Triggers:
   Don’t wait for sustained issues—a single outlier spike should trigger a quick check by a data analyst or automated workflow.
3. Visual Dashboards:
   Scatterplots and time series views should be integral to any analyst’s toolkit. They help spot patterns no raw table ever will.
4. Threshold Tuning:
   Every app is unique, so use the valid-app feature benchmarks as living thresholds. Periodically review and update these as new patterns emerge.
5. Evolving Models:
   No fraud pattern lasts forever—retrain risk rules and detection models on fresh data, and encourage feedback loops from manual reviews.

5. Closing Comments

Through a combination of time series analysis, metric investigation, and sharp visualizations, I was able to distinguish safe behavior from risk and document the exact signals that triggered IVT marking. This approach is practical, reproducible, and designed to help teams spot fraud early and confidently.