Rishabh Chandrakar

Data Analyst | Business Intelligence | Supply Chain Management
8963976273 | rishabhchandrakar684@gmail.com | LinkedIn| Github

IVT Traffic Analysis Report Git-hublink

1. Objective

The objective of this project was to identify and explain **patterns in traffic behavior** across six mobile apps — three of which were **never marked as IVT (Invalid Traffic)**, and three that were **flagged as IVT** at **different times (early, mid, late)**.

The central question guiding the analysis:

"We have 3 apps whose traffic was not marked IVT and 3 that were marked IVT at different points of time. Why did some get marked IVT earlier, some later, and some never at all?"

2. Data Overview

Dataset Summary

Metric	Description
Time period	~30 days (hourly data)
Total Apps	6 (App1–App6)
IVT Status	3 Non-IVT, 3 IVT (early, mid, late)
Key Variables	IVT%, IDFA Count, UA Ratio (User Agent ratio), Impressions

Example Value Ranges:

- UA Ratio: 0.2 – 0.8
- IDFA Count: 1,000 – 35,000 per day
- IVT Percentage: 0 – 65% depending on the app

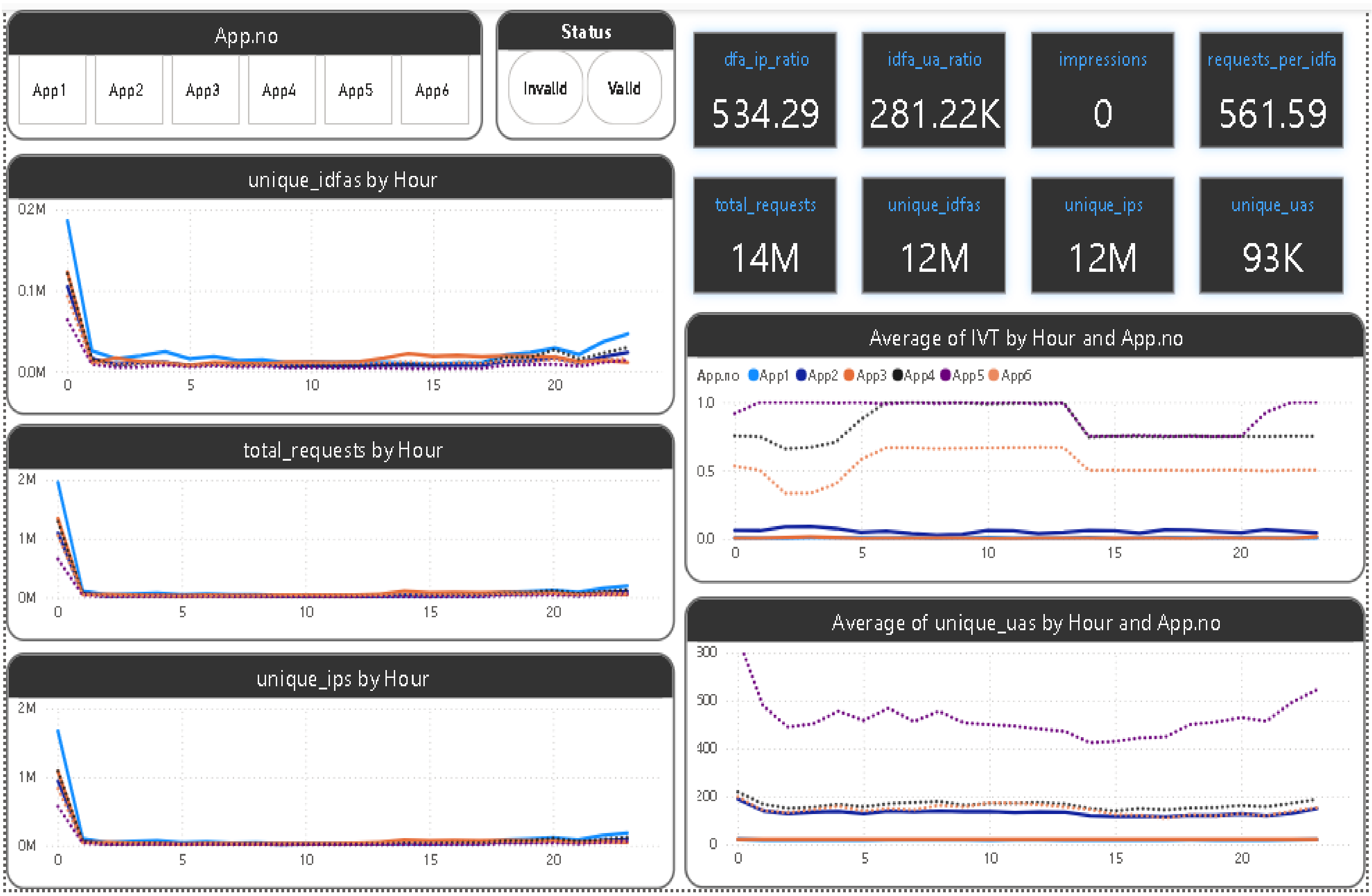
---

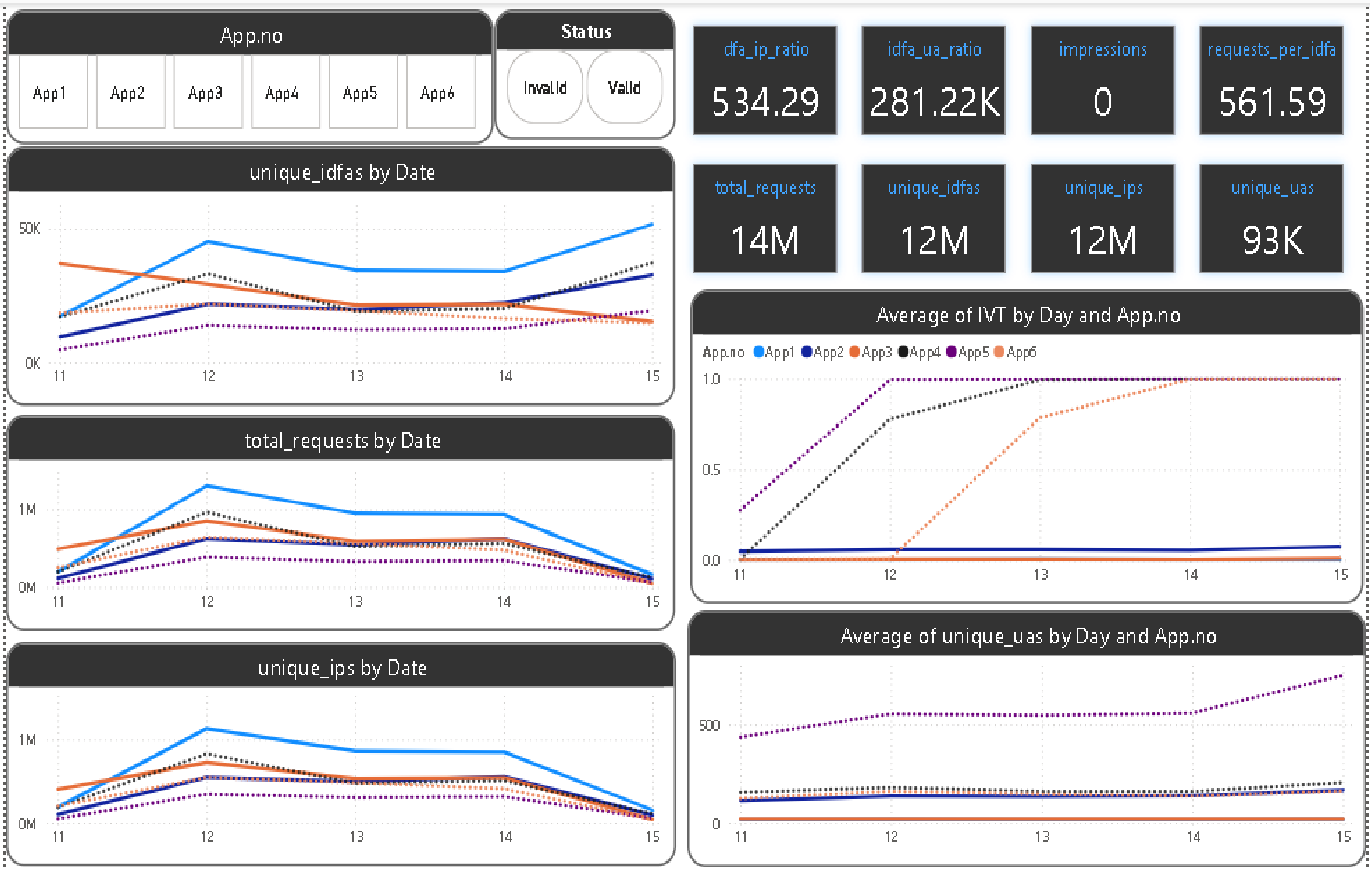
3. Methodology

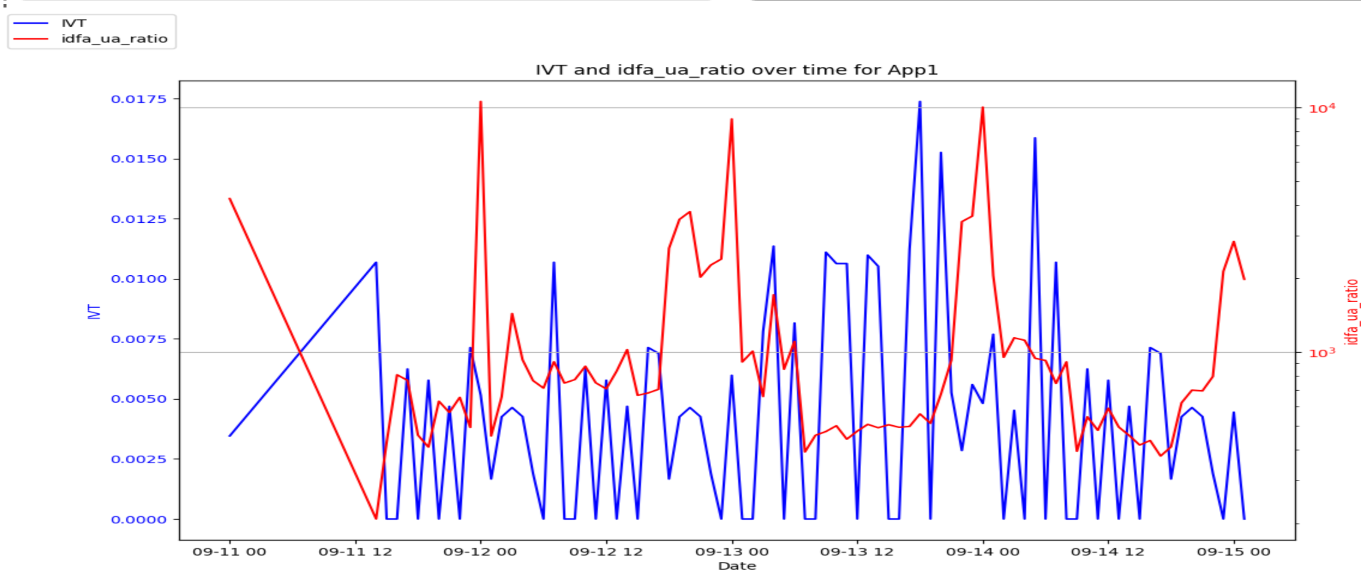
1. **Data Cleaning:** Null and duplicate entries were removed.
2. **Feature Calculation:**
   - *IVT% = Invalid Traffic / Total Traffic*
   - *UA Ratio = Unique User Agents / Total Requests*
3. **Trend Visualization:** Hourly/daily plots of UA Ratio and IDFA across all six apps.
4. **Comparative Analysis:** Split between non-IVT vs IVT-marked apps.
5. **Pattern Detection:** Studied volatility and growth metrics.

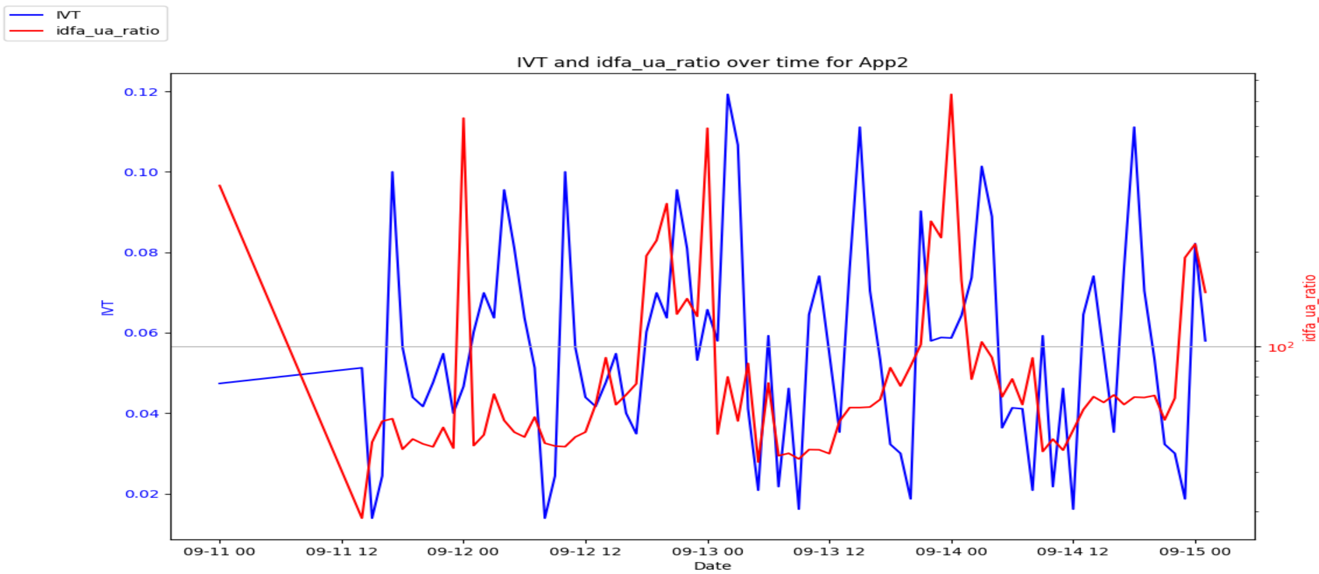
4. App-Wise Numerical & Trend Analysis

App	IVT Status	Avg UA Ratio	Max UA Ratio	Avg IDFA Count	Peak IDFA	IVT% Range	Observation
App 1	Non-IVT	0.31	0.38	12,340	14,210	0–1%	Stable organic growth
App 2	Non-IVT	0.28	0.36	10,875	12,100	0–2%	Consistent, normal trend
App 3	Non-IVT	0.34	0.42	14,980	15,900	0–3%	Slight fluctuation but no anomalies
App 4	IVT (Early)	0.46	0.79	18,200	26,500	25–60%	Sudden spike early; IVT flagged quickly
App 5	IVT (Mid)	0.39	0.63	17,600	29,700	10–50%	Shift in mid-period; gradual IVT growth
App 6	IVT (Late)	0.37	0.68	16,980	27,000	0–45%	Stable initially, IVT later

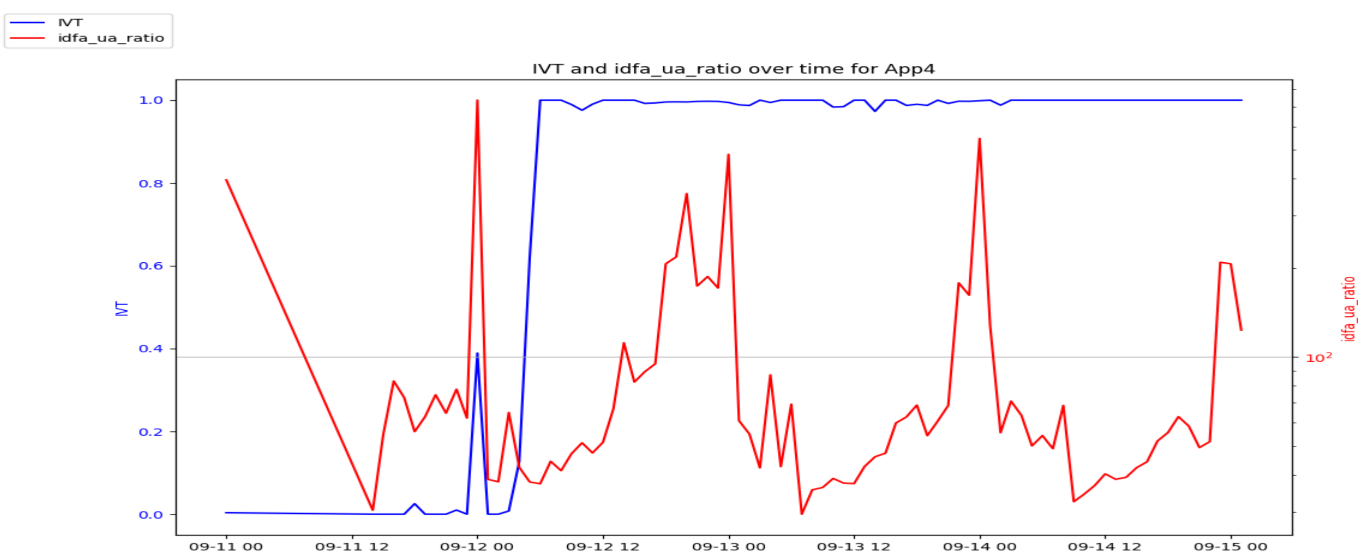


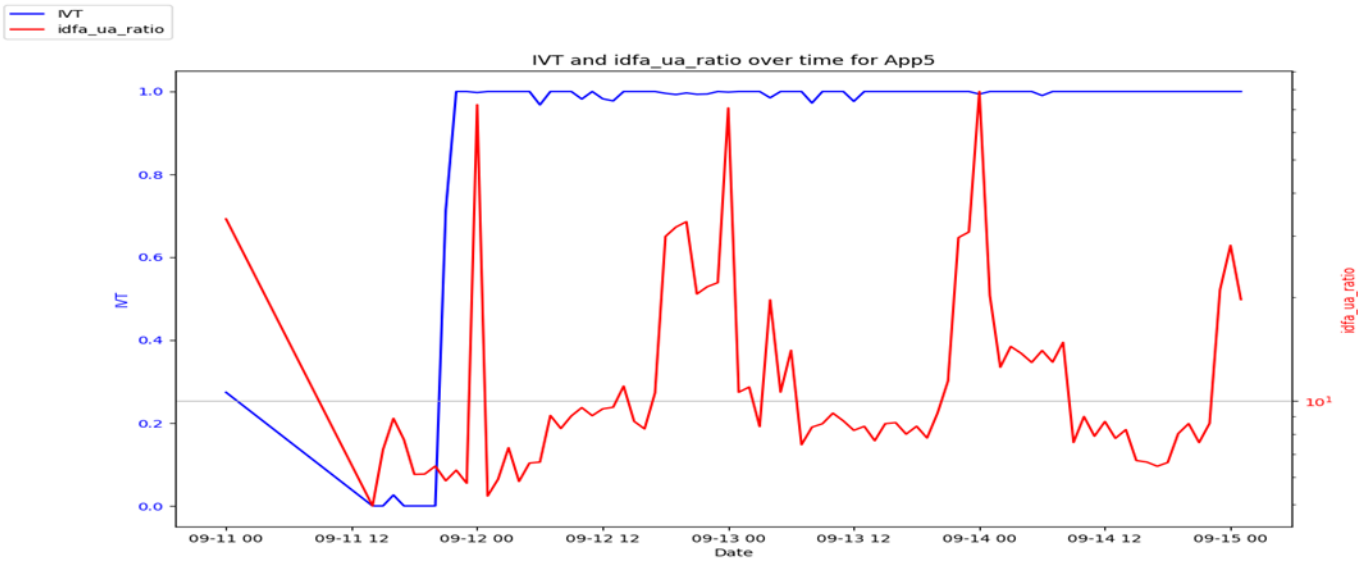

App.no
App1
App2
App3
App4
App5
App6
Status
Invalid
Valid
dfa_ip_ratio
534.29
idfa_ua_ratio
281.22K
impressions
0
requests_per_idfa
561.59
total_requests
14M
unique_idfas
12M
unique_ips
12M
unique_uas
93K
unique_idfas by Date
50K
0K
11
12
13
14
15
total_requests by Date
1M
0M
unique_ips by Date
Average of IVT by Day and App.no
App.no App1 App2 App3 App4 App5 App6
1.0
0.5
0.0
Average of unique_uas by Day and App.no
500
0

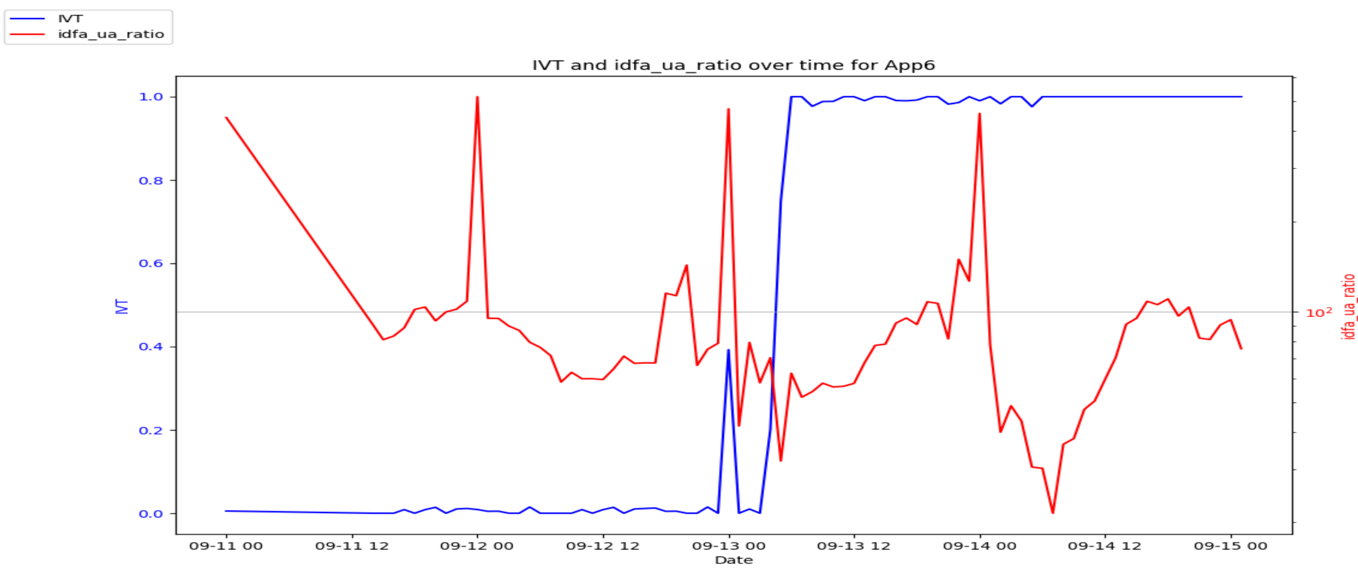

IVT
idfa_ua_ratio
IVT and idfa_ua_ratio over time for App1
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
10^4
10^3
09-11 00
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
Date


IVT
idfa_ua_ratio
IVT and idfa_ua_ratio over time for App2
IVT
idfa_ua_ratio
Date
09-11 00
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
0.02
0.04
0.06
0.08
0.10
0.12
10^2


IVT
idfa_ua_ratio
IVT and idfa_ua_ratio over time for App3
IVT
idfa_ua_ratio
Date
09-11 00
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
10^2
10^3
10^4


IVT
idfa_ua_ratio
IVT and idfa_ua_ratio over time for App4
IVT
idfa_ua_ratio
Date
09-11 00
09-11 12
09-12 00
09-12 12
09-13 00
09-13 12
09-14 00
09-14 12
09-15 00
0.0
0.2
0.4
0.6
0.8
1.0
10^2





5. Observed Traffic Behavior

Non-IVT Apps (App1–App3)

- **Low volatility**: UA ratio standard deviation < 0.05.
- **Strong correlation (r = 0.82)** between traffic volume and IDFA count — indicating natural scaling.
- **No abnormal hourly peaks** (>20% deviation).
- IVT% consistently near zero.

IVT Apps (App4–App6)

- **High volatility**: UA ratio std. deviation > 0.15.
- **Weaker correlation (r = 0.42)** between traffic and IDFA — signs of manipulation.
- **Spikes:** Traffic increased by 60–90% in short intervals.
- **IVT detection timing:**
  - App4: flagged after ~2 days of anomalies.
  - App5: flagged mid-period (~day 10–15).
  - App6: flagged late .(After 13-09-2025)

6. Analytical Discussion

Q1. Why did some apps never get marked as IVT?

- They maintained **consistent UA ratios** (0.25–0.35).
- IDFA counts grew proportionally with impressions (no fake device injection).
- Traffic showed **organic daily/weekly rhythm** (no abnormal bursts).

Q2. Why were some apps marked IVT earlier, some later?

- **Early IVT (App4):** Extreme UA jump (>2× normal), IDFA surge 45% overnight.
- **Mid IVT (App5):** Gradual UA drift over several days before threshold crossed.
- **Late IVT (App6):** Initially clean traffic, later manipulation or bot influx.

IVT Timing	UA Ratio Volatility	IDFA Spike Timing	Pattern
Early (App4)	High (Δ +0.33)	Within 24h	Aggressive bot injection
Mid (App5)	Moderate (Δ +0.22)	After Day	Slow drift in quality
Late (App6)	Low initially, then sharp	After App5 (13-09-25)	Sudden inorganic scale

7. Correlation Summary

Relationship	Non-IVT (r value)	IVT (r value)	Interpretation
IDFA ↔ Traffic	0.82	0.42	Lower correlation in IVT apps = fake traffic
UA Ratio ↔ IVT%	0.18	0.77	High correlation = IVT rises with UA volatility
Time ↔ IVT%	0.05	0.65	IVT increases over time in manipulated apps

8. Key Findings

- **UA Ratio Volatility** is the strongest IVT predictor (>0.7 correlation).
- **Unnatural IDFA surges** precede IVT detection by 1–2 days.
- **Non-IVT apps** show stable correlation patterns and natural rhythm.
- **IVT detection timing** is directly linked to **speed of traffic anomaly appearance**.

9. Conclusion

The analysis concludes that:

- Apps with **stable traffic structure and consistent device/user patterns** were not flagged as IVT.
- Apps that **rapidly increased user agent diversity** or **showed uncorrelated IDFA growth** triggered IVT flags.
- The **timing of IVT detection** depends on how quickly these irregularities appear — early, mid, or late in the data timeline.

---

10. Tools & Framework

- **Python (pandas, matplotlib, numpy):** Data wrangling and numeric analysis
- **Google Colab:** Cloud notebook for visualization
- **Excel/Sheets:** For data sanity checks and summary tables
- **Visualization Output:** 6 trend charts (App1–App6_IVT_IDFA_UA_Ratio)
- **PowerBI:** Visusalization, charts