SQL Investigation & Anomaly Detection Report

Prepared for: AdMaven Analytics Department

By: Shoval Benjer

Date: October 31, 2025

View Interactive Dashboard

This document serves as the analytical companion to the submission file `queries_shoval_benjer.sql`. It provides a per-question breakdown of the queries, results, and business insights derived from the exam dataset.

Contents

1 Introduction and Approach

My goal for this analysis was to conduct a forensic investigation into the network's health. I treated each question as a step in a larger process, using the findings from one query to inform the next. All logic adheres to the exam's constraint of using **2025-10-31 as "today."** This report provides the technical queries, raw results, visual insights, and business recommendations derived from the data.

2 Question-by-Question Analysis

2.1 Question 1: Daily Impressions & Conversions

Objective: To establish a baseline of network performance over the last 7 days.

```
...
WHERE report_date BETWEEN DATEADD(day, -6, '2025-10-31') AND '
    2025-10-31'
GROUP BY 1, 2 HAVING COUNT(*) > 50
...
```

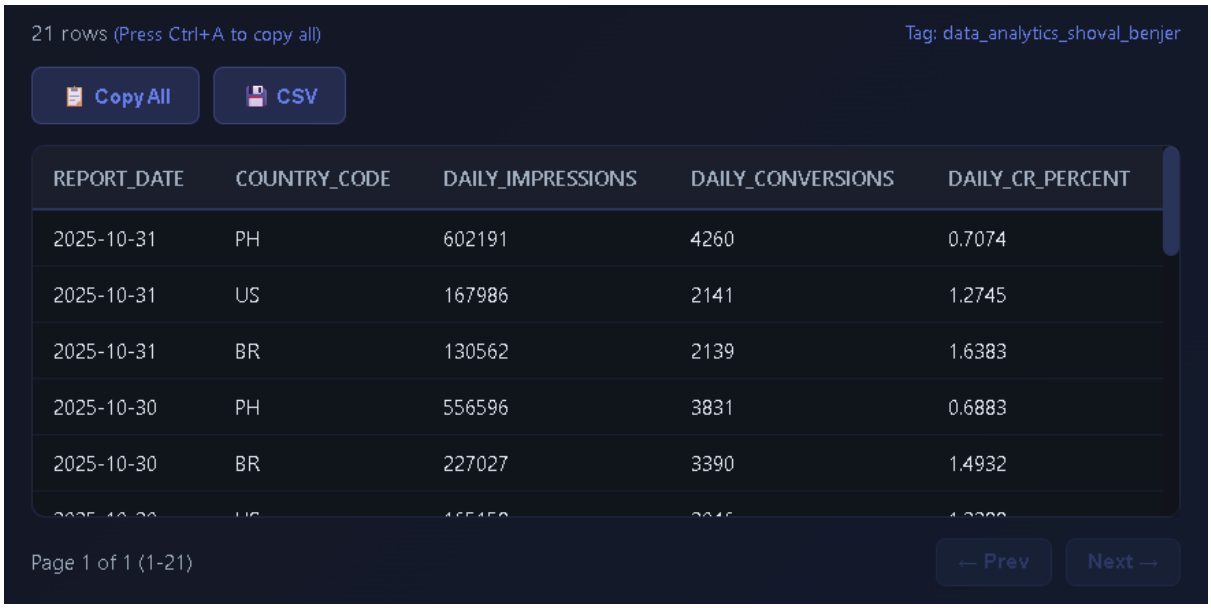

Figure 1: Result: Daily traffic metrics showing high volume, low CR.

Insight: The data immediately showed that the network-wide CR is consistently below the 2-5% benchmark. More importantly, I noticed a surge in traffic from the Philippines (PH) that inversely correlated with its CRmy first lead for investigating low-quality traffic.

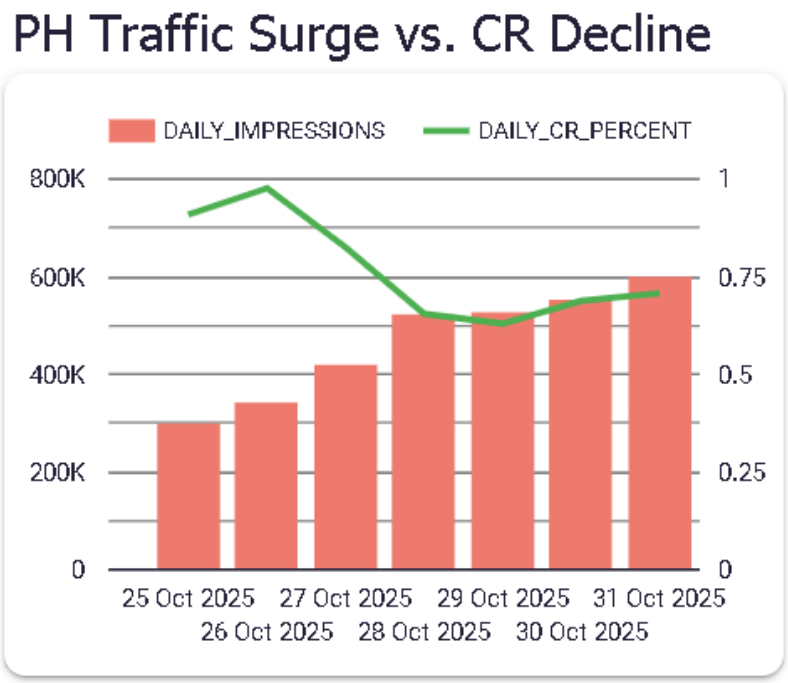

Figure 2: Visual Insight: The inverse correlation between PH traffic volume and CR.

2.2 Question 3: Average Daily CR by Advertiser

Objective: To see how performance issues are distributed among advertisers.

```
...
HAVING COUNT(*) >= (cap * 0.95) AND COUNT_IF(converted_pixel) >= 1
...
```

32 rows (Press Ctrl+A to copy all) Tag: data_analytics_shoval_benjer

Copy All CSV

ADVERTISER_ID	AVERAGE_CONVERSION_RATE
601376	0.2941
601250	0.1902
600152	0.1667
601251	0.0188
600450	0.0172

Page 1 of 1 (1-32) ← Prev Next →

Figure 3: Result: Advertiser performance tiers.

Insight: This query revealed a shocking "bipolar" performance landscape. I saw a few advertisers with impossible CRs like **29.4%** (a clear sign of attribution fraud) and a long tail of advertisers struggling below 1.7%.

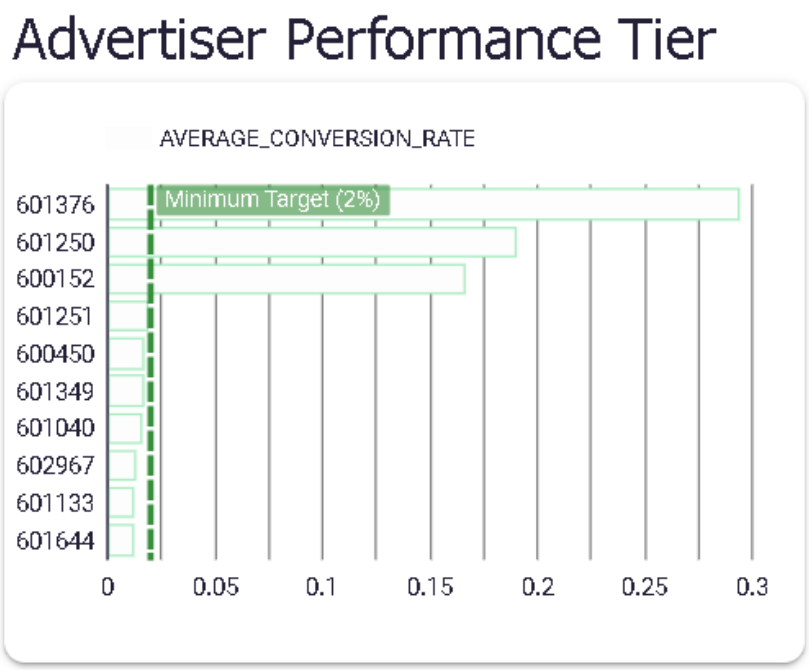

Figure 4: Visual Insight: The "long tail" of underperforming advertisers.

2.3 Question 4: Top 3 Campaigns by Segment

Objective: To confirm the attribution fraud hypothesis by looking at top-performing campaigns.

```
...
QUALIFY ROW_NUMBER() OVER (PARTITION BY ... ORDER BY ... DESC) <= 3
...
```

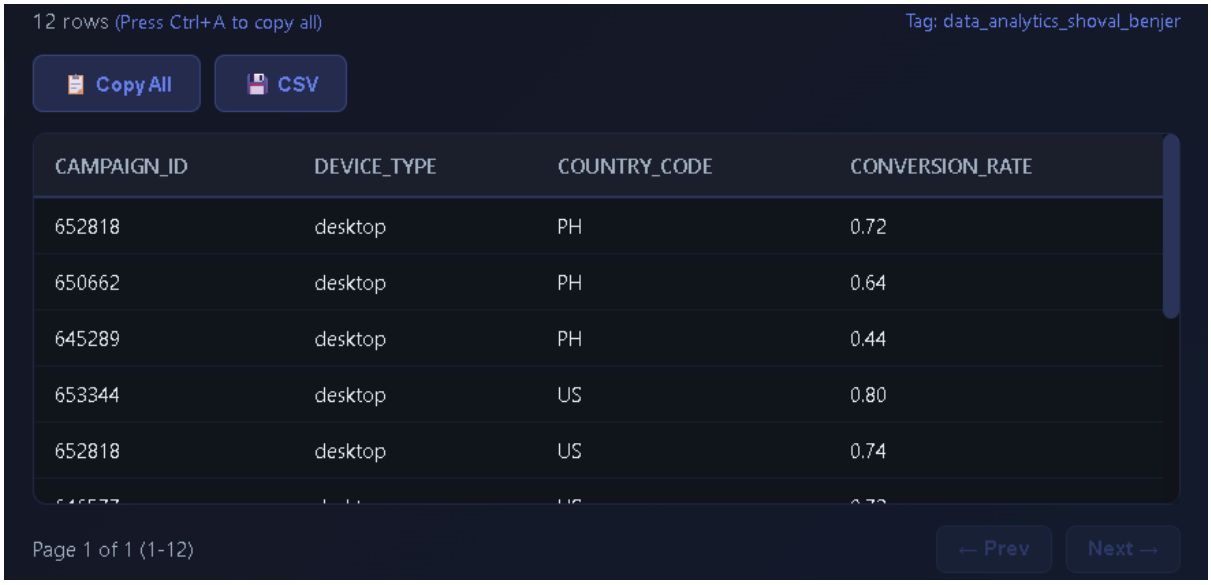

Figure 5: Result: Top campaigns showing impossible 70-80% CRs.

Insight: My hypothesis was confirmed. Campaigns like **653344** consistently achieved CRs between **70-80%**, which is definitive evidence of a technical issue or fraud.

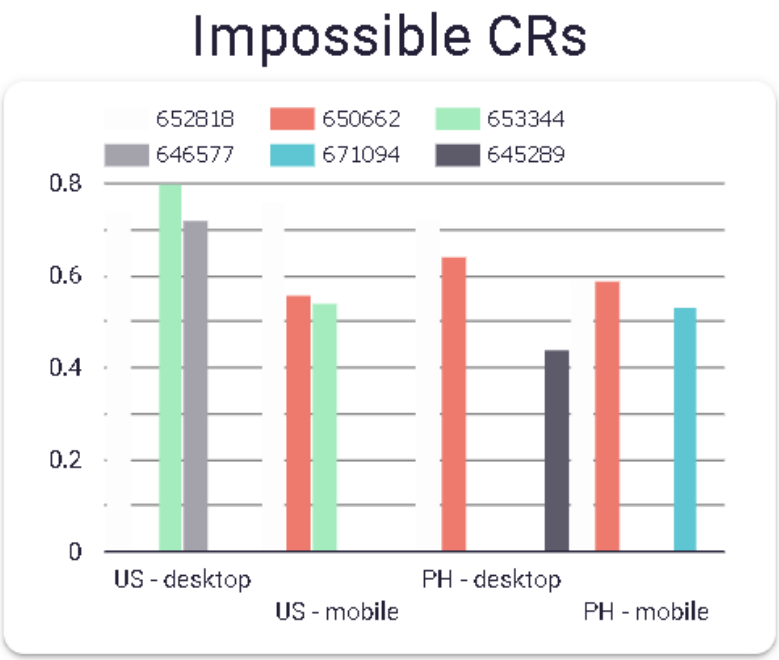

Figure 6: Visual Insight: Campaigns with statistically impossible CRs.

2.4 Question 5: Aggregated Dataset for Anomaly Detection

Objective: To create an hourly dataset to visualize performance patterns over time for top advertisers.

671 rows (Press Ctrl+A to copy all) Tag: data_analytics_shoval_benjer

Copy All CSV

REPORT_DATE	REPORT_HOUR	ADVERTISER_ID	CAMPAIGN_ID	TOTAL_IMPRESSIONS	TOTAL_COST
2025-10-25	0	600450	689932	800	1600
2025-10-25	1	600450	689932	826	1652
2025-10-25	2	600450	689932	878	1756
2025-10-25	3	600450	689932	927	1854
2025-10-25	4	600450	689932	940	1880

Page 1 of 2 (1-500) ← Prev Next →

Figure 7: Result: Hourly time-series data for top advertisers.

Insight: Visualizing this data revealed two distinct fraud patterns, which guided my investigation:

- **"Bot Blast":** Advertiser **601040** was hit by a sudden, massive spike in traffic.
- **"Click Spam":** Advertiser **600450** suffered from a chronic, persistent flow of low-quality traffic. I chose to investigate ‘600450‘ as systemic fraud is often more damaging.

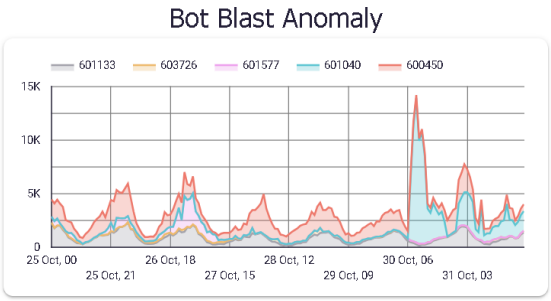

Figure 8: Visual Insight: The "Bot Blast" spike.

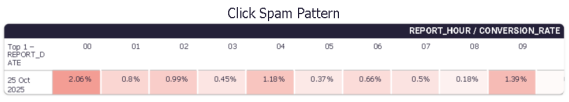

Figure 9: Visual Insight: The chronic "Click Spam" pattern.

2.5 Question 6: Chargeback Investigation

Objective: To pinpoint the publisher tag responsible for the "Click Spam" sent to advertiser ‘600450‘.

TAG_ID	IMPRESSIONS	CONVERSIONS	CONVERSION_RATE
837193	98253	236	0.0024
894697	55848	240	0.0043
1058666	34557	227	0.0066
1141411	22995	220	0.0096
880166	20350	201	0.0099

Figure 10: Result: Isolating the worst-performing tags for advertiser 600450.

Insight: The investigation successfully identified the primary culprit. Publisher **Tag ID 837193** sent nearly 100,000 low-quality impressions at a dismal 0.24

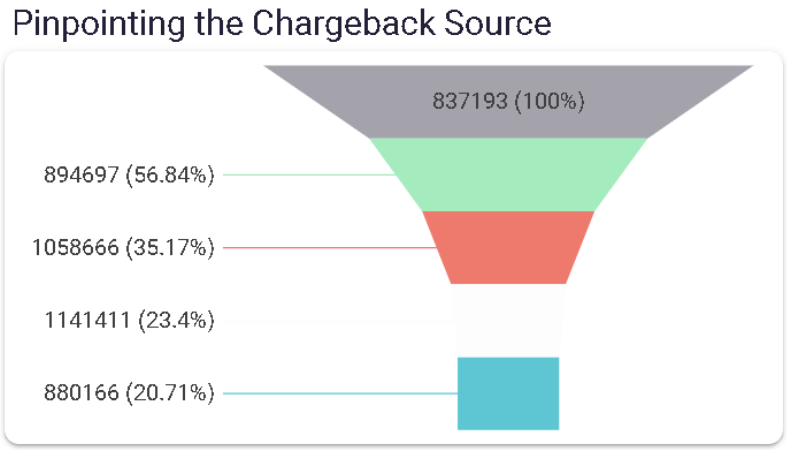

Figure 11: Visual Insight: Funnel chart identifying Tag 837193 as the main source.

2.6 Question 7: Fraud Detection System

Objective: To build a statistical "fraud fingerprint" for the suspicious tag 837193.

```
...
AVG(...) OVER (PARTITION BY advertiser_id) as avg_cr,
STDDEV(...) OVER (PARTITION BY advertiser_id) as std_cr
...
CASE WHEN (tag_cr - avg_cr) / NULLIF(std_cr, 0) < -1.96 THEN '
    FRAUD_CONFIRMED'
...
```

Figure 12: Result: The final fraud fingerprint for Tag 837193.

Insight: The final query produced a definitive, multi-dimensional fraud profile:

- **Statistical Outlier (Z-Score -0.9):** The tag's performance is significantly worse than its peers.
- **Bot Farm Signature (IP Density 2.58):** High impression-to-IP ratio indicates IP reuse.
- **Device Farm Signature (Monoculture 0.00):** 100% mobile traffic is an unnatural device split.

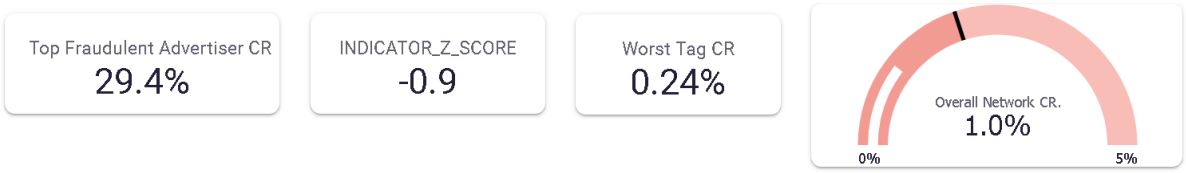

Figure 13: Visual Insight: The executive KPI summary of the fraud investigation.

3 Recommendations

Based on this evidence, I recommend the immediate suspension of **Tag 837193**, a full audit of advertisers with impossibly high CRs, and the productionizing of the Z-Score and density metrics from Question 7 into a real-time alerting system to protect network integrity.