Bad Bot Threat Analytics: Geographic, ISP & Use Case Intelligence Dashboard

---

1. OBJECTIVE

The provided text describes a dataset focused on analysing bot requests for various customers. It outlines four distinct datasets: "Use Case Stats", "Geo Stats", "ISP Stats", and "Customer Details", each offering different perspectives on bot activity. "Use Case Stats" categorises bad bot requests by their purpose, such as "Account Takeover" or "Price Scraping", while "Geo Stats" identifies the country of origin for these attacks. "ISP Stats" details the internet service provider linked to the bot requests, and "Customer Details" provides general information about the customer, including their industry and traffic plan. The ultimate goal is to visualise this data to understand and classify bad bot requests based on unique customer IDs.

2. Power BI Data Cleaning & Transformation

Step 1: Load Data into Power BI

- Imported data from Excel into Power BI Desktop.
- Multiple sheets/tables were loaded, including key ones such as Customer Details.

Step 2: Open Power Query Editor

- Clicked on “Transform Data” to open the Power Query Editor.
- Performed data cleaning and transformation operations table by table.

Step 3: Clean Each Table

Remove Unnecessary Columns/Rows

- Action: Scanned all tables for irrelevant blank rows or columns.
- Result: Removed them to ensure a clean data structure.

Rename Columns

- Action: Right-clicked on column headers and chose Rename.
- Goal:
    - Removed extra spaces, special characters, and inconsistent casing.
    - Ensured columns have clear, consistent, and meaningful names.

Set Correct Data Types

- Reviewed and assigned appropriate data types for all columns:
    - Date fields → set to Date type.
- Numerical fields (e.g., counts, amounts) → set to Whole Number or Decimal Number.

Step 4: Enhance 'Customer Details' Table

Add a Custom Numeric Column for “Traffic Slab / Month”

- Field Used: Traffic Slab / Month (e.g., values like 50 MILL, 10 BILL)
- Objective: Convert this text-based metric into a pure numeric field for analysis.

Steps Followed:

1. Clicked on Add Column → Custom Column
2. Gave the new column the name: TrafficSlabNumeric
3. Used the following Power Query M code as the formula:

M Language

```
let
    txt = [Traffic Slab / Month],
    num = Number.FromText(Text.BeforeDelimiter(txt," ")),
    unit = Text.Upper(Text.AfterDelimiter(txt," ")),
    factor = if Text.StartsWith(unit,"BILL") then 1000000000
             else if Text.StartsWith(unit,"MILL") then 1000000
             else 1
in num * factor
```

Final Checks

- Verified the structure and cleanliness of all transformed tables.
- Confirmed all data types are correct and consistent.
- Ensured the newly added column (TrafficSlabNumeric) can be used in future reports, KPIs, and visuals.

Establish Relationships Between Tables

Relationships Created:

1. Customer Details[Customer ID] → Use Case Stats[Customer ID]
2. Customer Details[Customer ID] → Geo Stats[Customer ID]
3. Customer Details[Customer ID] → ISP Stats[Customer ID]

🔧 How Relationships Were Created:

- Drag and drop the Customer ID column from Customer Details onto the corresponding Customer ID column in the related tables.
- Repeated this for all three target tables:
    - Use Case Stats
    - Geo Stats
    - ISP Stats

3. Data Analysis

Largest geographic presence based on Bad Bot Requests?

Insights:

1. Customer A has the largest geographic presence (194 countries)
    - This suggests Customer A's platform or services are exposed globally, or are targeted by bots from the widest range of locations.
2. High Risk ≠ Wide Geo Spread
    - In the previous chart, Customer G had the highest risk score, but here it has the smallest geographic footprint (46 countries).
    - Insight: Concentrated bot attacks from fewer countries can still be high risk (based on intensity/IP density).
3. Customers E, D, F, and C are in the middle cluster
    - These customers are targeted across 150–170 countries, showing moderate-high geographic exposure.
    - If they belong to a high-risk vertical (e.g., financial or e-commerce), they may require global bot mitigation policies.
4. Customers B & G have limited reach
    - Customer B: 82 countries
    - Customer G: 46 countries
    - This might imply a regional focus or a smaller customer base.
5. Geo distribution can influence mitigation strategy
    - For Customer A, geo-blocking from specific regions may be infeasible — instead, adaptive filtering or behavioral detection is more effective.
    - For Customer G, geo-specific blocking or WAF rules could quickly reduce exposure due to the narrow footprint.

A bad bot scoring system using the following parameters:

- **Bad Bot Requests**
- **Unique Bot IPs**

- **Use Cases**
- **Industry Vertical.**

Key Insights

1. Risk is concentrated.
    - Top-3 customers (G, B, D) = 49 % of total risk.
    - A focused mitigation plan for those three could halve your exposure.
2. G is the critical pain-point.
    - Its score is ~10 pts above the baseline and ~2 pts above the next-highest peer, signalling very high volumes of bad-bot requests and/or a large pool of unique malicious IPs for its use-case / vertical.
3. Middle cluster (C–F) is tightly packed.
    - Only 2 pts separate C, D, F. These customers share similar risk drivers; a single policy (e.g., stricter WAF rules or rate-limiting tuned to their use-cases) could address all three.
4. "Low-risk" still isn't low.
    - Even the "best" performer (A) is only ~4 pts below the benchmark. Nobody is truly in the safe zone, implying the current baseline (15.8) might be conservative or the overall environment is hostile.
5. Industry vertical weightings appear to matter.
    - If G/B/D belong to the same vertical, that vertical weighting may be too light; conversely, if A/E's vertical is different, it could be lowering their scores even though raw bot activity is sizable.
6. Model validation tip.
    - Check multicollinearity between Bad-Bot Requests and Unique IPs so one is not drowning out the other. Consider normalising scores by traffic volume to avoid penalising large-traffic customers purely for scale.

Top 5 ISPs for each customer based on Bad Bot Percentage

Key Insights

1. More requests occurred from "Not Exceeded" customers (111M)
    - But these customers stayed within their permitted slab, which indicates good traffic control and effective monitoring.
2. Four customers exceeded their slab:
    - Customer D is the major offender (45M requests) → Likely either underestimated their traffic needs or had a spike in activity.
    - Customers C, E, and F each had moderate overages (10M, 10M, and 5M respectively).
    - Customer D needs immediate attention
3. Responsible for 64% of the Exceeded group's traffic.
    - ➢ May require:
    - Traffic slab reassessment
    - Optimization
    - Billing update or contract renegotiation
    - Customer G has very low traffic (6M) and is well within limits
4. Possibly a smaller or less active client. Minimal risk for overage penalties or scalability issues.
    - Customers A & B manage large volumes (52M, 53M) but remain within slab
5. They are high-traffic clients but well-managed.
    - Consider providing volume incentives, or an early warning system to keep them from crossing thresholds in future.

Customers that exceeded their monthly Traffic Slab

Key Insights

1. Customer F is most impacted by high-bot ISPs
    - All 5 ISPs for Customer F have Bad Bot % > 97.79%
    - Top ISP: Pars Parva System Co. Ltd. – 97.86%
    - Insight: These ISPs are nearly entirely malicious for Customer F.

- Customer F should blacklist or throttle traffic from these ISPs immediately.

2. Customer A is also heavily hit by high-bot ISPs
    - Top ISPs (e.g., IP Express Ibadan – 94.87%, Hetzner – 94.36%) show very high bot rates.
    - Overall average for A's ISPs ~93–94%
    - Suggests targeted and sophisticated bot activity.
    - Consider implementing ISP-based blocking, rate-limiting, or behavioral anomaly detection.
3. Customer D and Customer E show mid-high bot percentages
    - Customer D: 97.62% (for at least one ISP)
    - Customer E: 97.42%
    - These values are close to F's range, meaning bot attacks are very ISP-specific for these customers as well.
    - Recommendation: Treat these as critical even if their overall traffic volume is lower.
4. Customer C and Customer B show lower Bad Bot %
    - Customer C: 93.25%
    - Customer B: 81.14%
    - While still high, they are below the total average (91.93%)
    - Suggests some mixed or semi-legitimate traffic is also present.
    - Requires deeper segmentation (bad bot vs. suspicious human traffic).
5. Total Line Insight
    - Overall Avg Bad Bot % = 91.93%
    - Most ISPs in this table are highly suspicious.
    - If you use a threshold like 80%, nearly all ISPs here are dangerous.

Total Use-Case Distribution

Decomposition Tree:

- Insights – Customer D focus
- Two-client concentration: D + C generate 86 % of global malicious traffic.

- Customer D is almost single-vector: 95 % = content-scraping bots.
- Price-scraping is emerging: small but growing niche for D.

Recommendations

Priority	Action	Expected gain
🔴	Deploy anti-scraping counter-measures (JS challenges, rotating CSS, honeypot fields) on D's content endpoints.	Potentially cut 45 % of global bot load.
🟠	Throttle or CAPTCHA price API calls for D.	Stops second-tier abuse.
🟢	Repeat tree drill-down for Customer C to isolate its top use-case, then replicate controls.	Systematic reduction of remaining 39 %.

Bot Score Over Date Month

Stacked Area by Vertical:

- Vertical-specific risk: E-commerce remains the prime target even after first mitigation wave.
- Adaptive attackers: Plateau after Day 10 implies bots adjusted to initial blocks.
- Silent sectors could be next: Low, steady lines often precede sudden campaign pivots.

Recommendations

Horizon	Action	Why
Now	Add behaviour-based detection (device fingerprinting, session-velocity rules) to e-commerce sites.	Captures bots rotating IPs/UA strings.

Horizon	Action	Why
30 days	Perform synthetic-bot testing on low-volume verticals to validate defences.	Avoid surprise spikes later.
Quarterly	Share threat-intel signatures across verticals when a new botnet is blocked.	Prevent attacker lateral movement.

Total Bad-Bot Requests by Date Month

Waterfall:

What it shows

- 25 Jul: +3 M surge driven mainly by Customer C (and F).
- 06 Jul / 02 Jul: Sharp decreases after fixes on Customer E and D.
- Post-mid-month: oscillations trend downward, but periodic upticks from Customer F & "Other".

Insights

1. Spike analysis is crucial: One-day surges are tied to individual customers.
2. Remediation works: Drops on 06 Jul & 02 Jul validate rapid-response effectiveness.
3. Residual volatility: Customers F and "Other" still trigger intermittent rises—monitoring gaps remain.

Recommendations

Phase	Action	KPI / Metric
Detection	Implement 24-hr alert: flag any customer adding >250 k bad requests in a day.	Mean time-to-detect < 2 hrs.

Phase	Action	KPI / Metric
Root-cause	For each green bar, run WAF log deep-dive & tag bot family → update blocklists.	Recurrence of same bot family < 10 %.
Reporting	Automate weekly waterfall KPI in Power BI; colour-code customers by SLA breach risk.	Stakeholders receive Friday digest with week-on-week deltas.

1. Visuals & DAX artefacts

Area	Visual type	Key fields / measures used
Geographic Presence by Countries	Filled Map (Bing / Azure)	Country (Location)Customer Name (Legend, colour saturated)Bad Bot Requests (Size/Colour saturation)
Bad Bot Scoring System	Donut chart (Ring 100 %)	Risk Score measure CustomerName (Detail / Legend)
Top 5 ISPs per Customer	Matrix (hierarchy)	Rows: CustomerName ▸ ISPNameValues: Bad Bot % measure (Conditional formatting → bar & font colour)Page

Area	Visual type	Key fields / measures used
		filter: Is Top 5 ISP = 1
Geo Presence by Customers	Horizontal bar	CustomerName (Axis)Distinct Countries (Values)
Customers Exceeded Monthly Limit	Stacked column	Traffic Slab Status (Exceeded / Not) on X-axisTotal Requests stacked by CustomerName (Values)DAX flag: Exceeded = IF(TotalRequests > MonthlySlab, "Exceeded", "Not Exceeded")
Hidden helper measures (power interactivity)	• Bad Bot %• Risk Score Avg• Is Top 5 ISP (1/0)• Geo Footprint (DISTINCTCOUNT Country)	

2. Interactive controls (Slicers / Filters)

Control	Where & how it's implemented
Customer tabs (A … G) at top	A horizontal slicer styled as "Tile" and synced across pages. Single-select mode ensures only one customer context at a time.
Dropdown filter (right-hand pop-out pane)	Page-level CustomerName filter for power-users (hidden in reading view but editable when Filter Pane shown).
Top-5 ISP page filter	In the Filter Pane ➜ "Filters on this visual" → Is Top 5 ISP = 1.
Edit Interactions	Each slicer is configured via ribbon ► Format ► Edit Interactions so that all six visuals respond in sync to the customer selection.

3. Navigation & Bookmark logic

Item	Purpose	How it's wired
Reset ↩ button (top-right arrow)	Clears all slicers/filters back to default dashboard state.	Button ► Action = Bookmark ➜ Reset_Page. Bookmark captures:• Default slicer state (no customer selected)• Default visual sorts• Filter-pane settings
Customer buttons	Provide quick "tabbed" navigation feeling.	Each tile triggers Sync Slicer change (no extra bookmark needed).

Item	Purpose	How it’s wired
Selection Pane groups	Visuals are grouped (e.g., “Header”, “Cards”, “Charts”) so the Reset bookmark can easily Show/Hide entire groups if you ever switch to drill-through view.	

4. Styling, layout & UX polish

Feature	Notes
Custom theme (.json)	Green gradient palette for consistency across donut, bars and matrix conditional formatting.
Rounded-corner containers	Each visual sits inside a Shape → Rounded Rectangle (thicker border) to create card-style separation.
Iconography	Robot logo (top-left) imported as Image visual; drives brand recall.
Tool-tips	Default tooltip shows Bad Bot %, Risk Score, Geo Count (pulled via multi-row card).
Responsive design	Mobile-portrait layout copy exists (hidden page), sharing same slicers via Sync Slicer.

4. Key Insights & Conclusion

Executive Summary

The analysis reveals a critical cybersecurity crisis with over 91.93% of total traffic being malicious bot activity, representing 41.5M bad bot requests across all customers. This constitutes a severe, coordinated attack targeting multiple industry verticals with sophisticated data harvesting operations.

Key Threat Indicators

Geographic & Scale Analysis

- Eastern European concentration: Primary attack origin in Ukraine region
- Customer A: Largest geographic footprint (194 requests, 23.7%)
- Customer D: Highest volume threat (19.7M requests, 45M traffic limit exceeded)
- 4 out of 7 customers exceeded monthly traffic limits

Attack Sophistication

- Content scraping dominates: 18.7M requests targeting intellectual property
- Price scraping campaigns: 943K requests for competitive intelligence
- Multi-industry targeting: Travel, Real Estate, E-Commerce, and Financial sectors
- Persistent threat patterns: Sustained 5M+ bot scores across all verticals

ISP Compromise Analysis

- Critical ISP infiltration: 97.82-97.86% bot traffic through specific providers

- Systematic network compromise: Multiple ISPs showing 90%+ malicious traffic
- Coordinated infrastructure: Suggests organized cybercriminal operations

Strategic Implications

Immediate Risks

- Data theft at massive scale - Content and pricing intelligence being harvested
- Infrastructure overload - Traffic limits exceeded by 70M+ requests
- Revenue impact - Legitimate traffic being crowded out by bots
- Competitive disadvantage - Proprietary data being systematically extracted

Long-term Consequences

- Erosion of data integrity across multiple industry verticals
- Potential regulatory compliance issues due to security breaches
- Operational cost escalation from bot-driven infrastructure demands
- Customer trust degradation if service quality deteriorates

Critical Recommendations

Immediate Actions (0-30 days)

- Emergency bot mitigation deployment across all customers
- ISP blacklisting for providers showing >95% bot traffic
- Geographic blocking for Eastern European attack vectors
- Traffic throttling implementation for customers exceeding limits

Medium-term Strategy (1-6 months)

- Advanced bot detection systems with behavioral analysis

- Customer-specific traffic profiling and anomaly detection
- ISP partnership program for threat intelligence sharing
- Industry vertical security frameworks tailored to specific attack patterns

Long-term Security Posture (6+ months)

- Predictive threat modeling based on temporal attack patterns
- Zero-trust architecture implementation across all customer touchpoints
- Automated response systems for real-time threat neutralization
- Continuous security monitoring with machine learning capabilities

Conclusion

This analysis reveals a sophisticated, large-scale cybercriminal operation targeting multiple customers across various industries. The 91.93% bot traffic rate represents an existential threat requiring immediate, comprehensive response. The coordinated nature of attacks, geographic concentration, and systematic ISP compromise suggests this is not isolated incidents but rather organized cybercrime infrastructure requiring enterprise-level security transformation.