Analyzing Security Incident Logs to Detect Potential Cyber Threats

Data Analytics Capstone
By: Lani Nguyen

EXECUTIVE SUMMARY

This project analyzes Microsoft security incident logs to identify high-risk accounts, organizational vulnerabilities, abnormal authentication patterns, and MITRE ATT&CK techniques associated with severe alerts. Using SQL, Excel, and Tableau dashboards, the analysis reveals that a single user account (673934) is responsible for an overwhelming majority of high-severity incidents, exhibits near-constant alert activity, and triggers multiple high-risk attack categories.

Several organizations (738, 73, and 116) show chronic unresolved medium-to-high severity alerts, especially involving ransomware, phishing, and script-execution techniques mapped to MITRE ATT&CK. Off-hours activity is abnormally high across multiple entities, suggesting automated attempts, beaconing malware, or misconfigured processes.

This analysis provides actionable recommendations for identity investigation, endpoint hardening, improved triage workflows, and enhanced monitoring.

PROBLEM STATEMENT

As an IT auditor, my objective is to analyze security logs to identify suspicious activity that may indicate security breaches, non-compliance with controls, or policy violations. The goal is to use data analytics to detect anomalies, trends, and behavioral indicators that align with elevated cybersecurity risk.

DATASET

Source: Microsoft Security Incident Prediction
https://www.kaggle.com/datasets/Microsoft/microsoft-security-incident-prediction

The dataset includes fields such as account identifiers, timestamps, organizational units, alert categories, severity levels, verdicts, MITRE techniques, and entity types.

GOAL

Use data analytics to identify patterns in security incident logs that:

- Indicate potential cyber threats

- Highlight weaknesses in incident response
- Identify abnormal user behavior
- Surface unresolved high-risk alerts
- Map activity to MITRE ATT&CK for threat classification

SCOPE OF WORK

This project analyzes a subset of logs covering recent security incidents.

The analysis includes:

- User accounts with frequent high-severity incidents
- Monthly and weekly incident patterns
- Business units with unresolved alerts
- Off-hours access attempts
- Alert frequency and time-interval anomalies
- MITRE ATT&CK techniques associated with severe activity

Deliverables include:

- Defined research questions
- Data dictionary and MITRE mappings
- Cleaned SQL datasets and GitHub repository containing datasets and queries
- Excel analysis with pivot tables and correlations
- Tableau dashboards
- Summary of results and recommendations
- Lessons learned

RESEARCH QUESTIONS

1. Which user accounts have triggered more than 3 high-severity incidents in a single week?
    - Useful for identifying potential insider threats or compromised credentials.

2. What is the average time between an alert being generated and the next alert for the same user account, and does that interval decrease during known attack periods?
    - Alert velocity indicates automated attacks or persistent compromise.

3. Which departments or business units generate the highest volume of unresolved medium-to-high alerts, and what controls can be strengthened there?
    - Pinpoints control weaknesses and policy enforcement gaps.

4. How many incidents involve access attempts outside normal working hours (6PM-6AM), and how often do those lead to escalation?
    - Anomalies outside typical hours may indicate suspicious or unauthorized activity.

5. What was the trend in failed login attempts over the last 30 days, and were there any spikes on specific dates?

- Identifies brute-force or credential-stuffing indicators.

DASHBOARD OVERVIEW

To support the analysis, I developed a series of Tableau dashboards that summarize key findings and highlight high-risk security behavior in a clear and interpretable format.

Dashboard 1 – High-Risk Account Security Overview
Focuses on:

- Top user accounts with high-severity alerts
- Unresolved alerts by organization
- Off-hours access attempts
- Average time between alerts

This dashboard helps identify compromised accounts, alert velocity anomalies, and organizational exposure.

Dashboard 2 – MITRE ATT&CK Technique Analysis
Focuses on:

- Most frequent and most severe attack categories
- Highest-risk MITRE techniques triggered by organization
- Accounts most associated with specific techniques

This dashboard visualizes threat behavior patterns mapped to the MITRE framework, helping illustrate attacker tactics, techniques, and procedures (TTPs).

Together, these visualizations support IT auditing by surfacing the most critical threats, highlighting trends, and validating the quantitative findings presented in the analysis section.

ANALYSIS

High Severity Accounts

Account 673934 triggered 217,265 high-severity incidents, with 77% tied to:

- SuspiciousActivity (52%)
- CommandAndControl (15%)
- Exfiltration (9%)

By comparison Account 30 had 1,916 incidents and Account 41 had 944 incidents.

Account 673934 should be considered a critical-risk identity.

Alert Velocity

Account 673934 generated alerts at extremely short intervals (average in seconds):

- Impact — 0.0067 seconds
- InitialAccess — 0.0482 seconds
- Discovery — 0.0660 seconds

This account also generated the most frequent severe alerts:

- Exfiltration — 0.1399 seconds
- Ransomware — 4.5506 seconds

Other accounts average 6–22 seconds between alerts.

These alert intervals indicate potential automated scripts, malware beaconing, and/or compromised accounts generating repeated events.

Organizational Analysis

Organization 738

- 6,253 unresolved medium-high alerts
- 6,238 tied to Ransomware

Organization 73

- 853 unresolved alerts
- 100% tied to Initial Access (phishing)

Organization 116

- 832 unresolved alerts
- 97% tied to Execution (PowerShell)

These organizations show critical triage backlogs and require immediate response hardening.

Off-Hours Activity

Account 673934 generated over 164 million off-hours activities.

- 986,071 were labeled as malicious
- 99% remaining alerts remain unresolved

Organization 93 has 95+ million access attempts.

- Nearly all escalated
- 100% left unresolved

This strongly suggests beaconing malware, brute force attempts, or misconfigured automation.

Attack Categories

Most frequent:

- Impact (77% severe, 0% resolved)
- Initial Access (23% severe, 6% resolved)
- Discovery (16% severe, 0% resolved)

Most Severe:

- Exfiltration (86% severe, 0% resolved)
- Ransomware (97% severe, 53% resolved)

Rare (but dangerous):

- Credential Stealing (0% severe, 0% resolved)
- Web Exploit (95% severe, 0% resolved)
- Weaponization (100% severe, 0% resolved)

Rare alerts being never resolved is a major audit concern.

MITRE ATT&CK Technique Findings

The most frequent MITRE Techniques associated with certain attacks.

Discovery
T1087 + T1087.002 account for 97% of discovery alerts.

Exfiltration
T1041 accounts for 70% of events.

Phishing
T1566 is responsible for 100% of credential-stealing attempts.

Organization Techniques

- Org 738 → Ransomware → T1486
- Org 73 → Initial Access → T1566.002
- Org 116 → Execution → T1059.001 (PowerShell)

These techniques correspond directly to real-world ransomware, phishing, and automation attack paths.

RECOMMENDATIONS

1. Investigate Account 673934 immediately.

This identity shows signs of compromise across:

- Alert volume
- Frequency
- Severity
- Off-hours anomalies
- MITRE mappings

2. Improve incident responses at Organizations 738, 73, and 116.

Focus on:

- Ransomware containment
- Phishing mitigation
- Endpoint execution controls
- Implementing SLAs and automated triage

3. Strengthen off-hours monitoring.

Cross-reference user access with HR data, VPN logs, and device inventories.

4. Address low volume but severe categories.

Credential theft, weaponization, and web exploits should never have a 0% resolution rate.

LESSONS LEARNED

- Understanding dataset structure early would have prevented misaligned questions.
- Saving multiple dataset versions is critical for recovery from cleaning mistakes.
- Exporting cleaned datasets progressively into Excel avoids re-running SQL.
- SQL mastery requires continued practice, especially with joins, timestamps, and analytic functions.

CONCLUSION

This analysis revealed multiple areas of high security risk, including a severely compromised account, ransomware and phishing vulnerabilities in specific business units, and evidence of automated or scripted attack behavior. Using SQL, Excel, and Tableau dashboards, the project demonstrates how data analytics can support IT auditing by uncovering hidden patterns, detecting anomalies, and informing targeted security improvements.