Cyber Security Attack Analysis

Created by: Nikhil Mohan Tatfale — February 2026. Analysis of 40,000 recorded cyber events to surface attack patterns, high-risk times and places, malware and alert trends, and severity distributions to inform defenses.

Project Overview

Scope

40,000 events across multiple attack types and network segments.

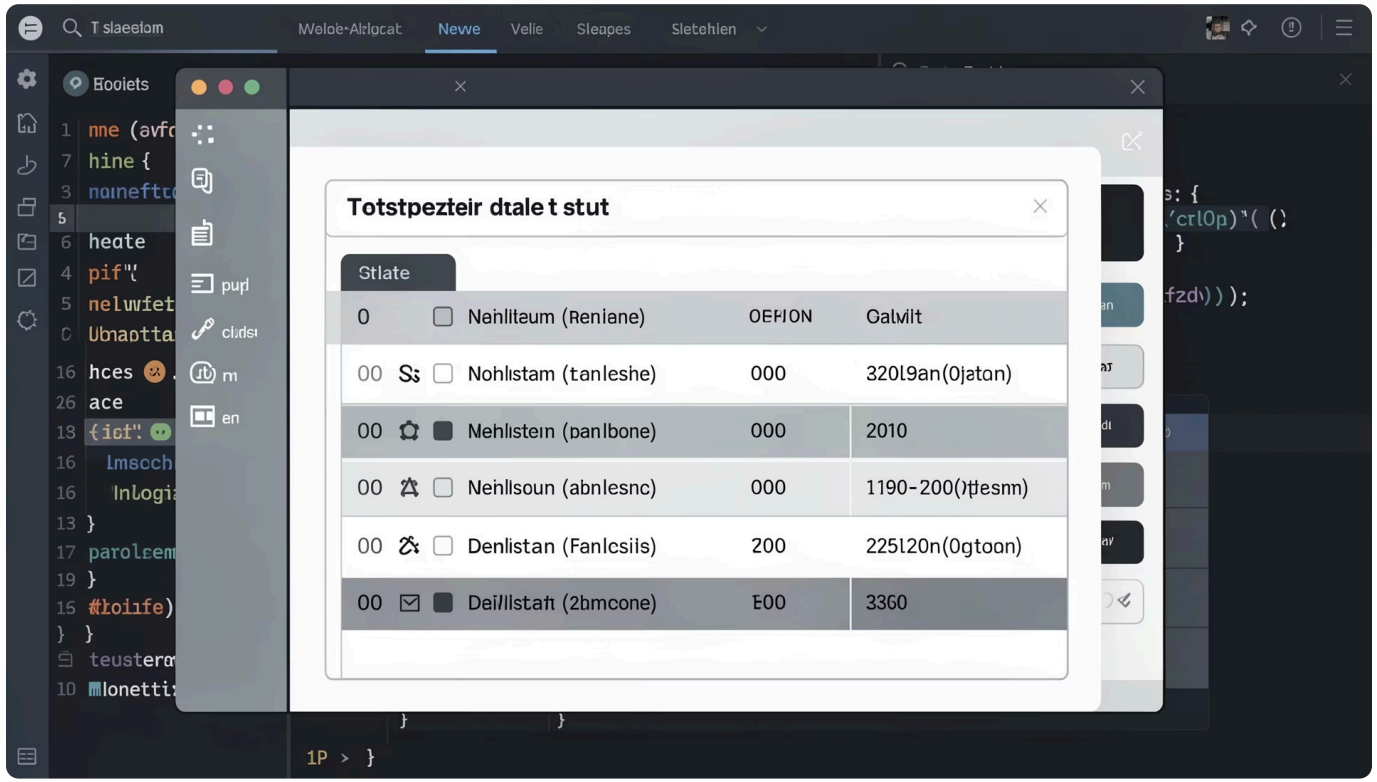
Goals

Discover attack patterns, high-risk periods/regions, malware/alert trends, and severity distributions.

Workflow

Data from Kaggle → EDA in Python → Cleaning & feature extraction → Power BI dashboard.

Dataset Summary

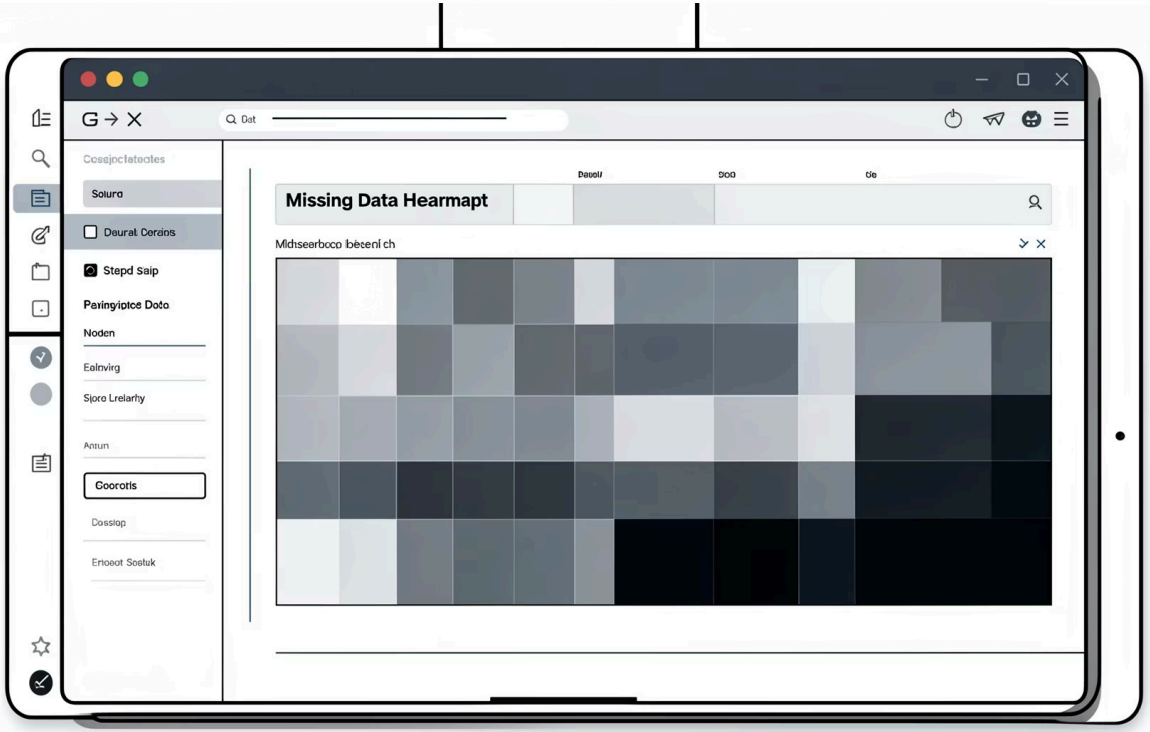


Key facts

Rows: 40,000 • Columns: 25

- Network IDs, Traffic Details, Security Events
- Device & Location, Proxy/Firewall/IDS logs
- Missing data concentrated in several security signal columns

Missing Data



Malware Indicators

20,000 nulls

Alerts/Warnings

20,067 nulls

Proxy / Firewall / IDS

~19,800–20,050 nulls across columns

Descriptive placeholders used to preserve records while flagging missing signals.

Data Preparation Highlights

Loading & EDA

Pandas in Google Colab — df.info() and df.describe() to profile types and counts.

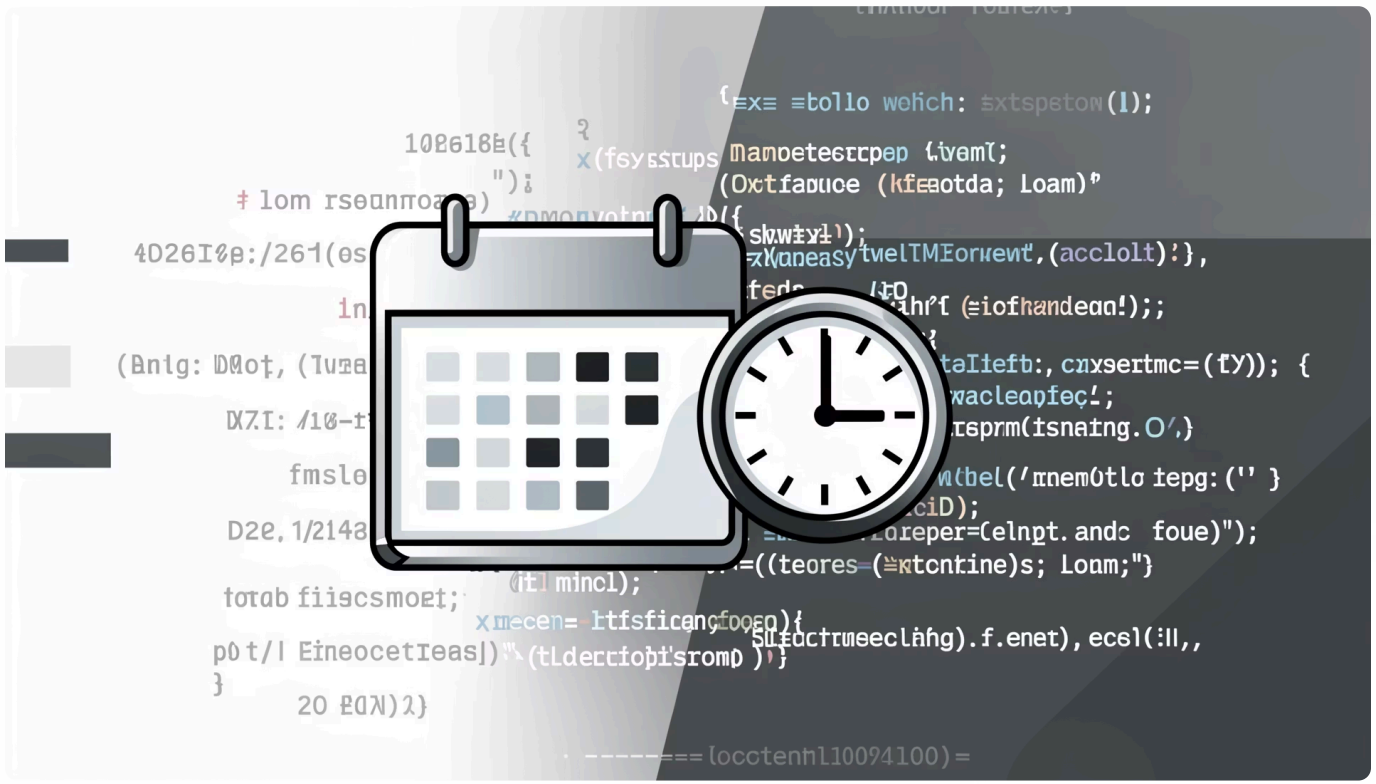
Missing-value strategy

Fill nulls with contextual defaults (e.g., 'No Malware Detected', 'Not Triggered', 'Unknown').

Column standardization

Converted to snake_case and renamed problematic headers for consistency.

Feature Engineering



Date & Time

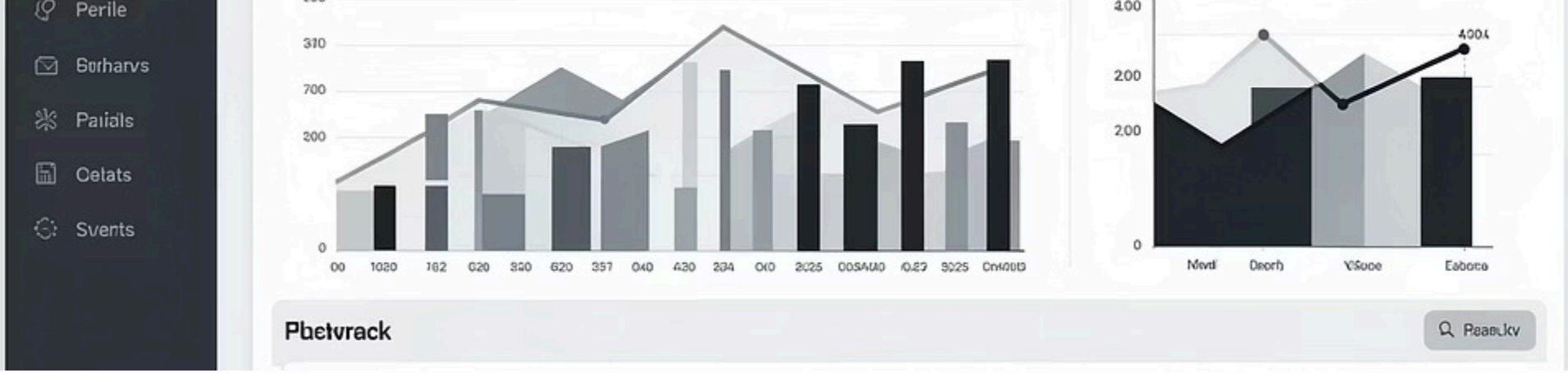
Parsed timestamp into separate date and time columns.

City & State

Split geo-location into city and state columns.

Operating System

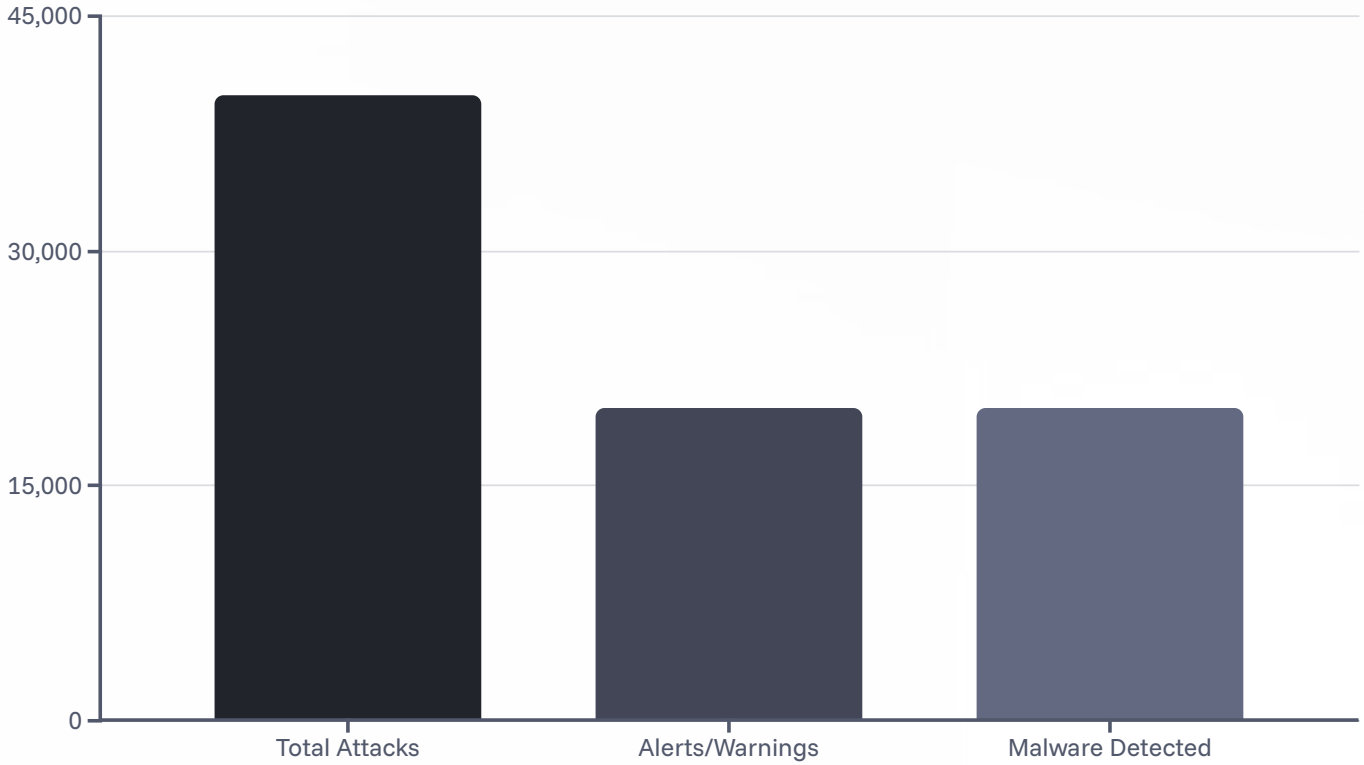
Extracted OS from device_information (Windows, Linux, Mac OS, iOS, Android, Windows CE, Other).



Power BI Dashboard

Interactive overview enabling drill-down by Attack Type, Traffic Type, Network Segment, and Year.

Dashboard Metrics & Visuals



Additional visuals: Attacks per Month (Mar peak 3.7K → Dec 2.7K), Network Type distribution (DNS 13,376; HTTP 13,360; FTP 13,264), Severity distribution (OK 10K; Medium 13,435; High 13,382; Low 13,183), OS breakdown (Windows 44.88%, Linux 28.97%, Mac OS 22.1%).

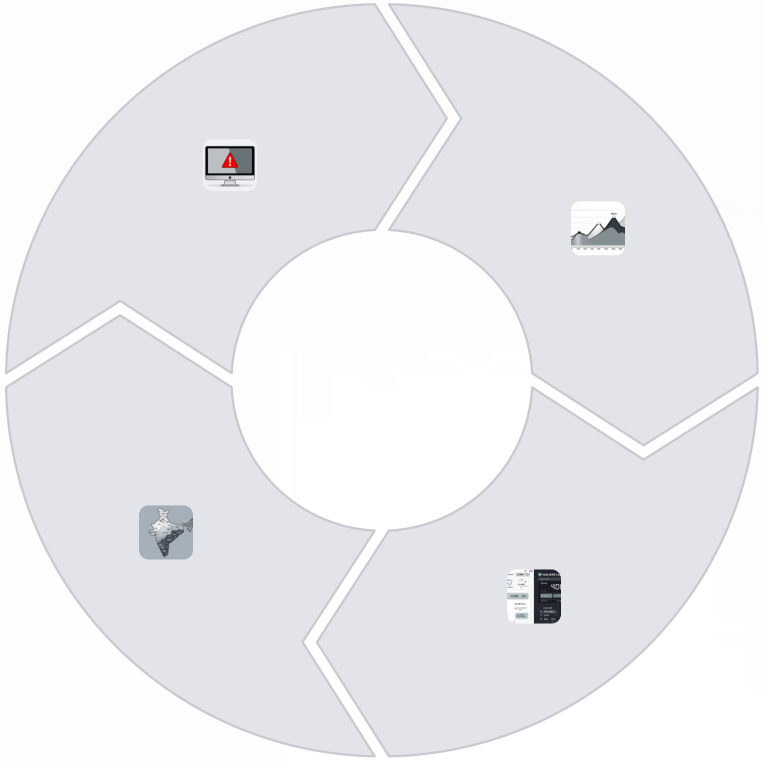
Key Findings

Windows Most Targeted

Windows accounts for nearly 45% of events; Linux and Mac OS follow.

Geographic Concentration

Top states: Manipur, Uttar Pradesh, Gujarat. Top cities: Ghaziabad, Aurangabad, Rourkela.



Peak Periods

March highest (3.7K); July/August high (3.6K); November/December lowest (2.7K).

Signal Coverage ~50%

~20K records triggered alerts and ~20K detected malware — half of events lacked signals.

Conclusions & Next Steps

- Prioritize Windows defenses given highest exposure.
- Investigate visibility gap where ~50% events lack alerts/malware traces.
- Focus monitoring in peak months (Mar, Jul, Aug) and high-risk regions.
- Maintain protocol-agnostic defenses — attackers exploit DNS, HTTP, FTP similarly.
- Refine dashboard with additional filters and anomaly detection measures.