Power BI, Practical Activity

Presented by: Siervo Ivan López Carvajal

2025

Contents

[Data Exploration and Cleaning: Summary Report 2](#_Toc197272860)

[Data Loading and Inspection 2](#_Toc197272861)

[Data Cleaning and Transformation 2](#_Toc197272862)

[Exploratory Data Analysis and Visualization 3](#_Toc197272863)

[Insights and Recommendations 5](#_Toc197272864)

[Conclusion 5](#_Toc197272865)

Table of Figures

[Figure 1 Alerted Threats by Attack Type Over Tim 3](#_Toc197277961)

[Figure 2 Distribution of Attack Types and Severity Levels 4](#_Toc197277962)

[Figure 3 Firewall Logs by Year 4](#_Toc197277963)

[Figure 4 Data Transformation Evidence Through Advanced Editor - Power Bi 5](#_Toc197277964)

[Figure 5 Power Bi Dashboard 5](#_Toc197277965)

# Data Exploration and Cleaning: Summary Report

This report includes the cleaning, transformation and analyzing a raw cybersecurity dataset using Power Bi. The main scope of this task was to prepare the data for effective visualization, extract useful information and highlight the value of proper data handling in business and cybersecurity environments.

# Data Loading and Inspection

The dataset that I used for this work has been obtained through www.keggle.com from the dataset section, I initially looked for many different datasets, but since I wanted to deep dive into the cybersecurity field, I looked for different options. I finally found an interesting option about a raw data of a cybersecurity report. The dataset was first imported from a .csv file containing network traffic logs, source and destination Ips, protocols, anomaly scores, device information and attack types. After importing the headers were adjusted and the column data typed were also modified to match their correct formats. (Timestamp as datetime, Packet Length as integer, etc.) Data Cleaning and Transformation

# Data Cleaning and Transformation

A major part of this project was focused on data cleaning. The Payload Data column was removed as per considering that it contained unnecessary information. Duplicated were removed and null/missing values in critical fields like malware indicators, Alerts/Warning, Proxy information, Firewall Logs and IDS/IPS Alerts were replaced with the placeholder “unknown” to maintain the dataset integrity without deletion.   
  
Considering that Geo-Location data was originally stored as a single column including city and state, there was necessary to separate city and state columns using a delimiter. Each new column was trimmed for whitespace. In addition, new custom columns were created for data integrity and better visibility:  
  
*1. Severity Rank* (numeric level based on *Severity Level*)  
*2. Detected* and *Alerted* flags for cleaner filtering  
*3. Browser*, *Operating System* and *Device Type* extracted from user-agent data in the Device *Information* field

The previous transformations were made on the dataset. The dataset was more structured and allowed easier analysis and data handling when filtering.

# Exploratory Data Analysis and Visualization

Once the clean-up of the data was performed, the visualization has been built up to identify trends and patterns. This included:

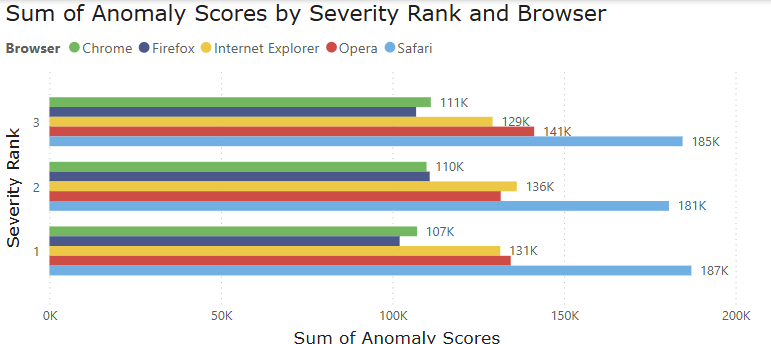


Figure 1 Total Anomaly Scores by Browser and Severity Level

A graph showing the number of data

AI-generated content may be incorrect.

Figure Alerted Threats by Attack Type Over Tim

A colorful pie chart with text

AI-generated content may be incorrect.

Figure Distribution of Attack Types and Severity Levels

A graph of firewall logs

AI-generated content may be incorrect.

Figure Firewall Logs by Year

The visuals allowed me to interpret that browser like Chrome and Safari had the highest anomaly scores, and that DDOS alerts showed steady numbers until a sharp drop in 2023. Despite the fewer total alerts in recent years, the average anomaly score remained high, revealing more advanced or target threats.

A screenshot of a computer program

AI-generated content may be incorrect.

Figure Data Transformation Evidence Through Advanced Editor - Power Bi

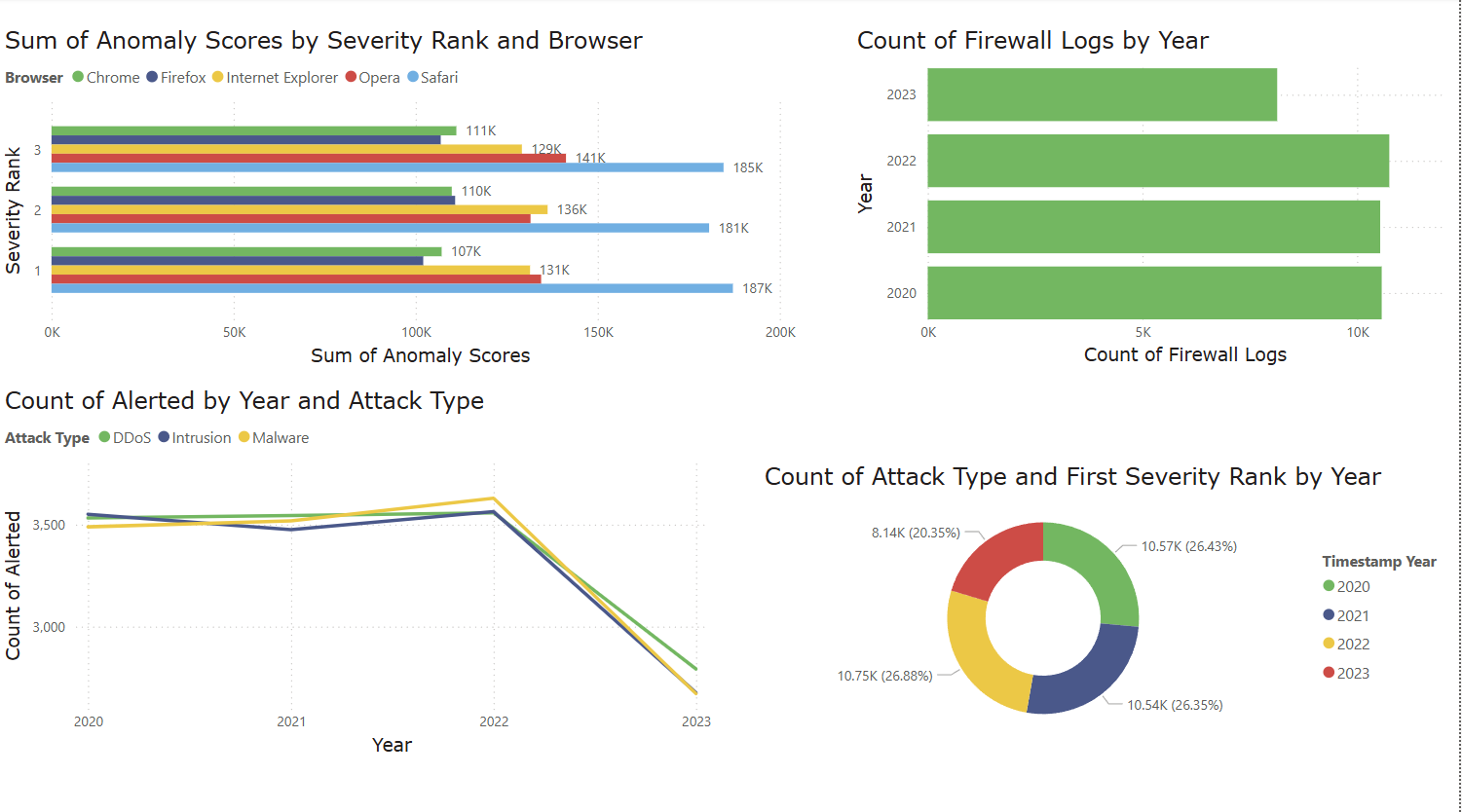


Figure Power Bi Dashboard

# Insights and Recommendations

# The data shows that high severity alerts are often linked to desktop uses running outdated browsers, which is important to highlight the risk of unpatched systems. While the analysis shows that volumes of alert have dropped in recent records, anomaly scores remain high, suggesting more complex or targeted threats. Organizations should focus on updating software, monitoring high-risk systems and tracking changes in threat behavior, not just volume. Anomaly trends offer valuable early warnings when alert counts alone may seem low.

# Conclusion

This project showed how proper data cleaning improves the quality of analysis. By removing errors, structuring data, and creating new insights through custom columns, I turned raw logs into meaningful patterns. Power Bi supported both transformation and visualization data, helping uncover critical trends. All data was treated to maintain data integrity, ensuring ethical use. Overall, this process reflects the real value of data preparation in cybersecurity analytics.