Executive Summary

Web Traffic Anomaly Detection

In today's rapidly evolving digital landscape, cybersecurity remains a critical concern—especially for cloud-based infrastructures. This project focuses on analyzing web traffic logs collected from AWS CloudWatch to detect and interpret suspicious or potentially harmful activity targeting a production web server.

The dataset consisted entirely of traffic labeled as suspicious, offering a valuable foundation for pattern discovery and anomaly detection. Using a combination of exploratory data analysis and machine learning techniques, the investigation uncovered several key insights:

- **High-risk IPs Identified**: A small number of source IP addresses were repeatedly observed, with one IP (`155.91.45.242`) responsible for the majority of detected anomalies.

- **Traffic Anomalies Detected**: An Isolation Forest model revealed 15 traffic entries (~5.3%) as anomalous. These anomalies clustered in specific hours, suggesting patterns of automated or targeted behavior.

- **Temporal Trends**: Traffic analysis showed a sudden spike in volume after a period of zero activity—potentially indicating downtime, scheduled maintenance, or a coordinated attack.

- **Geographic Origins**: Most traffic originated from the United States and Canada, with remaining activity spread across a few other countries.

The insights suggest the presence of bot-like behavior, potential infiltration attempts, and suspicious spikes in network usage. These findings are significant for establishing early-warning mechanisms, informing firewall rules, and guiding future threat prevention strategies.

The model and framework developed here can be adapted for real-time anomaly detection using AWS services like Lambda or SageMaker, providing scalable security monitoring for cloud environments.