Project Title: Don't try to bug me

Introduction & Objective: This data analysis project aims to investigate and visualize the landscape of cyberattacks targeting major global economies and critical industries.The primary objective is to move beyond generic statistics and uncover specific patterns, correlations, and vulnerabilities unique to the most targeted countries. By analyzing the interplay between attack vectors, target industries, and geographical location, this project seeks to provide actionable intelligence for cybersecurity professionals, policymakers, and corporate leaders to prioritize defense strategies and resource allocation.

Data Source & Overview: "Kaggle dataset" The analysis will be conducted on a large synthesized dataset of cybersecurity incidents,a sample of which is provided. This dataset contains rich, multi-faceted records of cyberattacks with the following key attributes:

· Attack Characteristics: Type of attack (e.g., Phishing, DDoS, SQL Injection, Zero-Day Exploit), outcome (Success/Failure), and duration. · Target Information: The system targeted (e.g., Cloud Service, IoT Device, Database), the industry sector (e.g., Healthcare, Finance, Retail), and the data component affected. · Contextual Factors: The security controls in place at the time (e.g., Firewall, MFA, WAF), the role of the targeted user, and the response time. · Geographical & Mitigation Data: The location (country) of the attack and the mitigation action taken (e.g., Patch, Containment, Block IP).

Key Analysis Questions: The project will seek to answer critical questions such as:

· Which 10 countries experience the highest volume of cyberattacks, and does the primary attack type differ by region (e.g., Russia vs. USA)? · Which industries within these top-targeted nations are most vulnerable to specific types of attacks (e.g., is Healthcare in the UK disproportionately targeted by Ransomware)? · What is the correlation between the security controls deployed and the outcome (success/failure rate) of attacks in these countries? · Are certain attack types consistently more successful against specific target systems (e.g., Cloud Services vs. IoT Devices) across different nations? · How does the average response time to an incident vary by country and industry, and what impact does it have on the outcome?

Our Methodology:

· Data Cleaning & Preprocessing: Handle missing values, standardize country and industry names, and ensure consistent formatting for timestamps and numerical fields. · Exploratory Data Analysis (EDA): Calculate descriptive statistics to understand the distribution of attacks by country, industry, and type. Identify the initial list of top 10 countries by attack frequency. · Geospatial Analysis: Use libraries like geopandas or plotly to create choropleth maps visualizing attack density and heatmaps for specific attack types across the globe, focusing on the top 10 nations. · Trend Analysis: Analyze temporal patterns to identify if certain attacks are increasing in frequency within specific countries or industries over time. · Correlation & Comparative Analysis: Employ cross-tabulation and statistical tests to explore relationships between variables (e.g., is a faster response time correlated with a higher failure rate of attacks?).

Expected Outcome: The final deliverable will be a comprehensive dashboard(built using tools like Tableau, Power BI, or Python Dash) and a report that provides:

· A ranked list of the top 10 most targeted systems. · Interactive visualizations breaking down the threat profile for each country by industry and attack type. · Insights into the most effective security measures and mitigation strategies for different regions and sectors. · Data-driven recommendations for enhancing national and organizational cybersecurity posture based on the identified patterns