1. **Project Overview**

**Title:**

**Global Cybersecurity Threat Analysis (2015-2024)**

**Introduction/Summary:**

Technology has improved greatly over the last two decades, and it is amazing what we can now achieve with it. Sadly, these advancements also come with new risks. As the world has become more digital, almost all our important information is now stored or accessed online.

Since much of this information is valuable, cybercriminals go to great lengths to steal it for their own benefit. When they succeed, the impact can be devastating — causing major financial losses, reputational damage, and other serious consequences.

For this project, I analyzed a global cybersecurity threat dataset covering the years 2015 to 2024. The goal was to uncover trends, understand the growing risks, and highlight ways organizations and individuals can better protect themselves in an increasingly connected world. The dataset contains records of cybersecurity incidents from ten (10) countries across seven (7) industries, detailing the attack source, financial loss, number of affected users, security vulnerabilities, defense mechanisms, and incident resolution time.

The primary objectives of this analysis were to:

* Understand the global impact of cybersecurity threats across countries and industries from 2015 to 2024.
* Identify the most common attack types and their effects on financial loss and user exposure.
* Analyze key threat sources, exploited vulnerabilities, and the resolution time of incidents.
* Evaluate the effectiveness of different defense mechanisms in minimizing financial losses.
* Reveal yearly patterns and trends in threat occurrence and damage severity.

A dashboard was created to visualize key metrics and indicators, including:

* **Total Threats, Financial Loss, Users Affected, and Resolution Time** – Assessing the overall scale and urgency of cybersecurity incidents.
* **Financial Loss and Users Affected by Country and Industry** – Identifying high-risk regions and sectors.
* **Threat Impact by Attack Source, Vulnerability, and Defense Strategy** – Pinpointing systemic weaknesses and response effectiveness.
* **Yearly Trends** – Tracking how different attack types have evolved over the years in terms of scale and impact.

**Key Findings:**

* **Most Financially Damaging Attack Source:** Nation-state attacks caused the highest financial loss, totaling $40.50B, followed closely by unknown sources (**$38.81B**).
* **Top Security Vulnerability:** Zero-day vulnerabilities led to the greatest financial losses ($39.55B), emphasizing the need for proactive threat detection.
* **Most Affected Industry:** The IT industry suffered the highest financial loss ($24.81B) and most users affected (250.09M).
* **Country with Highest Loss:** The UK recorded the highest financial loss among all countries, totaling $16.50B.
* **Country with Most Users Affected:** Brazil topped the list with 168.81M users impacted.
* **Year with Highest Financial Loss:** 2017 ($16.26B) had the highest financial loss with Phishing ($3.32B) being the most impactful attack type followed closely by the year 2023 with a total financial loss of $15.96B.
* **Most Effective Defense Mechanism:** Firewall had the lowest financial loss ($29.09B), indicating better effectiveness compared to other methods.

This analysis provides critical insights into the evolving nature of cybersecurity threats globally. The findings can help organizations prioritize investments in security infrastructure, understand where and how they’re most vulnerable, and make informed decisions to strengthen their digital defenses — therefore reducing financial loss and minimizing user impact.

1. **Data Description**

The dataset used for this analysis was obtained from Kaggle. It contains 3,000 records of cybersecurity incidents reported globally across 10 countries from 7 target industries between 2015 and 2024. It records key details for each incident, including the year of occurrence, financial losses, number of affected users, attack source, security vulnerabilities exploited, defense mechanisms applied, and the time taken to resolve each incident.

**Source**

This dataset was sourced from Kaggle and contains structured information on cybersecurity incidents across 10 countries and 7 target industries.

**Content**

This dataset includes a mix of categorical, numerical, and time-based data, allowing for comprehensive analysis of global cybersecurity threats.

* **Number of Rows:** 3000
* **Number of Columns:** 10
* **Data Types:**
* **Categorical Data:** Country, Attack Type, Target Industry, Attack Source, Security Vulnerability Type, Defense Mechanism Used.
* **Numerical Data:** Financial Loss, Number of Affected Users, Incident Resolution Time.
* **Date-Based Data:** Year.

**Column Descriptions**

* **Country:** Country where the attack occurred.
* **Year:** Year of the incident.
* **Attack Type:** Type of cybersecurity threat (Malware, DDoS).
* **Target Industry:** Industry targeted (e.g. Finance, Healthcare).
* **Financial Loss:** Estimated financial loss in millions.
* **Number of Affected Users:** Number of users impacted by the attack.
* **Attack Source:** Origin of the attack.
* **Security Vulnerability Type:** Type of vulnerability exploited (e.g. Zero-Day, SQL Injection).
* **Defense Mechanism Used:** Cyber-defense strategy applied (e.g. Firewall, VPN).
* **Incident Resolution Time:** Time taken to fully resolve the incident (in hours).

**Key Features**

* **Global Coverage:** The dataset spans cybersecurity incidents reported across 10 different countries, providing a worldwide perspective on cyber threats.
* **Multi-Year Range:** Covers incidents from 2015 to 2024, enabling trend analysis over a 10-year period.
* **Industry Focus:** Includes data on 7 target industries, allowing comparison of sector-specific vulnerabilities and impacts.
* **Threat Specifics:** Captures critical threat dimensions such as attack types, sources, and exploited vulnerabilities.
* **Impact Metrics:** Quantifies the consequences of attacks through financial loss (in millions), number of affected users, and resolution time (in hours).
* **Defense Mechanism Tracking:** Highlights the types of defenses used and their outcomes in mitigating threats.

This dataset gives a clear view of how cyber threats have affected different countries and industries over time. It provides a solid foundation for identifying patterns, assessing vulnerabilities, and evaluating the effectiveness of defense strategies.

1. **Data Preprocessing**

Effective data analysis begins with proper data preparation. This step involved importing, exploring, cleaning, formatting, and engineering features to ensure the dataset was analysis-ready.

**Initial Exploration**

The dataset was first imported into Microsoft Excel for exploration. A duplicate of the original dataset was created to ensure that the cleaning process could be tracked and reversed if necessary. As a standard practice for improving readability and organization, the cells were formatted with appropriate column widths and row heights using the shortcuts ALT + H + O + I (for AutoFit Column Width) and ALT + H + O + A (for AutoFit Row Height).

To assess data quality, preliminary checks were performed to identify duplicate and blank entries:

* **Duplicate Check**: Using the **Remove Duplicates** function located in the *Data Tools* section of the **Data** tab, the dataset was scanned for repeated rows. No duplicates were found, confirming the uniqueness of all entries.
* **Blank Check**: The **Go To Special** option, accessible from *Find & Select* under the **Home** tab, was used to highlight any blank cells. This inspection revealed no missing or null values across the dataset

This initial phase confirmed that the dataset was complete and contained no immediately visible structural errors.

**Data Cleaning and Formatting**

Some columns required adjustments to align with standard data types and ensure accuracy in analysis. The following transformations were made to prepare the data for analysis:

* **Date Formatting:**

The original 'Year' column contained only 4-digit year values (e.g., 2015) and was stored in ‘General’ format, which isn’t compatible with time-based tools in Excel.  
To resolve this:

* A new column was created using the formula =DATE(C2, 1, 1), which transforms each year into a complete date format representing January 1st of the corresponding year.
* This created a proper **Date** column that allowed for time-based visualizations.
* **Year Extraction and Cleaning:**

After converting the year into date format, a separate Year column was generated using the formula =TEXT(C2, "yyyy") to isolate the year as a string.

* This string-based year was then pasted as values using the **Paste Special > Values** option to eliminate formula dependencies.
* The final column was renamed YEAR and retained for filtering, grouping, and year-based trend analysis.
* **Column Cleanup:**

The intermediate formula-based columns used for date manipulation were deleted after extracting the necessary values. Only the finalized DATE and YEAR columns were kept.

**New Features**

Additional columns were introduced to improve insights during analysis:

* **Distinct Country:**

To calculate the number of unique countries, a new column named DISTINCT COUNTRY was created using:  
=IF(COUNTIF($A$2:A2, A2) = 1, 1, 0)  
This logic returns “1” for the first occurrence of each country and “0” for repetitions. It enables quick aggregation of distinct countries for metrics like country-level comparisons.

* **Distinct Target Industry:**

Using a similar method, the DISTINCT TARGET INDUSTRY column was generated with the formula:  
=IF(COUNTIF($F$2:F2, F2) = 1, 1, 0)  
This helps identify how many unique target industries are represented in the dataset.

* **Actual Financial Loss:**

The existing column titled Financial Loss (in Millions) stored financial loss figures in shorthand format (e.g., 80.53 = $80.53M).  
To calculate the actual values in full currency terms, a new column named Financial Loss (Actual) was created using the formula:  
=H2\*1000000  
This conversion ensures accurate aggregation and visualization of financial loss data in raw numerical format (e.g., $80,530,000).

* **Average Incident Resolution Time:**

Although not stored as a column, the **average** time it took to resolve incidents was calculated using the formula:  
=AVERAGE(N2:N3001)  
This helped to provide an overall picture of how long it typically takes organizations to resolve cybersecurity threats.

Other columns, including Attack Type, Attack Source, Security Vulnerability Type, Defense Mechanism Used, and Number of Affected Users, were already clean and correctly formatted, and required no changes.

**Summary of New Features**

1. **Date:** Full date created from the 4-digit year column using =DATE().
2. **Year:** Clean standalone year extracted using =TEXT().
3. **Distinct Country:** Identifies the first occurrence of each country using COUNTIF().
4. **Distinct Target Industry:** Identifies the first occurrence of each industry using COUNTIF().
5. **Actual Financial Loss:** Converts shorthand financial loss to actual amount using =H2\*1000000.

After cleaning and organizing the data, and creating some helpful new columns, the dataset was ready for analysis.

1. **Exploratory Data Analysis**

This section summarizes key patterns and trends in global cybersecurity threats reported from 2015 to 2024. The analysis focuses on threat types, financial losses, affected users, regional and industry impact, as well as year-over-year trends. It also evaluates vulnerabilities and the effectiveness of different defense mechanisms.

**Key Performance Indicators (KPIs)**

To summarize the overall impact of cybersecurity threats over the years, the following KPIs were calculated:

* **Total Threats:** 3000
* **Countries Affected:** 10
* **Total Financial Loss:** $151.48B
* **Total Users Affected:** 1.51B
* **Average Incident Resolution Time:** 36hrs

****

These KPIs highlight the severity of cybersecurity breaches globally, with billions in financial damages and hundreds of millions of users compromised across multiple countries and industries**.**

**Financial Loss Analysis**

**Financial Loss by Country**

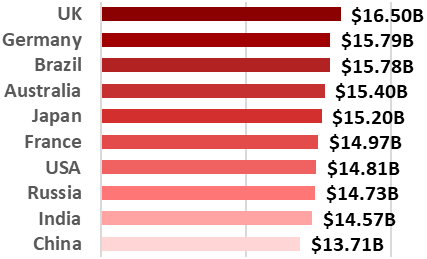
* UK recorded the highest financial loss, losing a total of $16.50B.
* Germany and Brazil followed closely with a loss of $15.79B and $15.78B respectively.
* China experienced the least loss losing a total of $13.71B.

**Users Affected by Country**

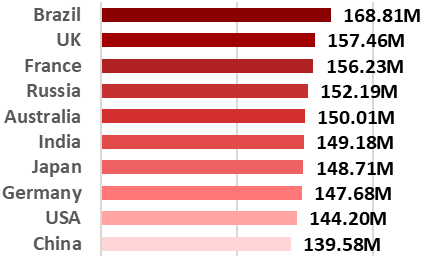
* Brazil had the highest number of affected users (168.81M), followed by UK (157.46M), and France (156.23M).
* China again had the least with a total of 139.58M affected users.

**Insights**

* The most financially impacted and most affected countries are not the same, suggesting different attack intensities or defense capabilities.
* The variation in loss amounts highlights the need for each country to assess its own threat landscape and prepare accordingly.



**Financial Loss by Country**



**Users Affected by Country**

**Financial Loss by Industry**

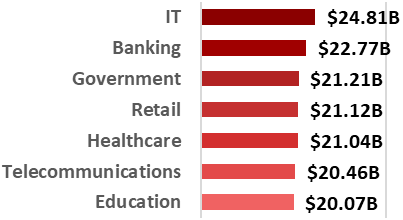
* The IT industry was affected the most recording a financial loss of $24.81B, followed by Banking ($22.77B).
* Education industry was the least affected with a loss of $20.07B.

**Users Affected by Industry**

* IT was also the most affected industry in terms of users (250.09M), followed by Banking (225.10M).

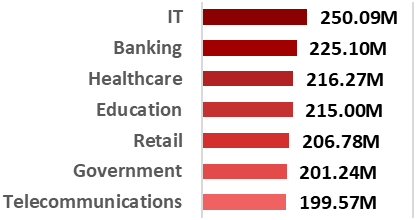
**Insights**

* IT and Banking industries were the biggest targets for both financial loss and user exposure, indicating a high vulnerability to cyberattacks.



**Financial Loss by Industry**

**Users Affected by Industry**



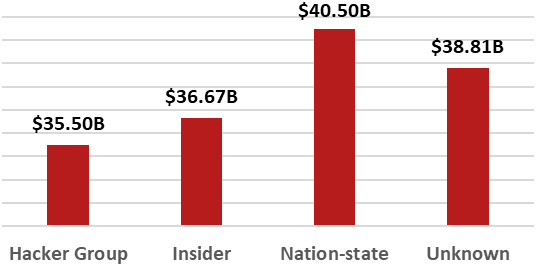
**Financial Loss by Attack Source**

* The Nation-state attack source caused the highest financial loss and was responsible for the loss of $40.50B, followed by an unknown source responsible for the loss of $38.81B.
* Insider threats contributed to $36.67B in losses, highlighting risks from compromised or malicious employees.
* Hacker Group had the least impact with a financial loss of $35.50B.

**Insights**

* External actors (nation-states and unknown sources) accounted for the majority of financial losses.
* However, insider threats still pose a significant risk and must not be overlooked, given their substantial contribution to total losses.
* The relatively lower figure from hacker groups does not diminish their threat level, as their methods could evolve or escalate over time.

**Financial Loss by Attack Source**



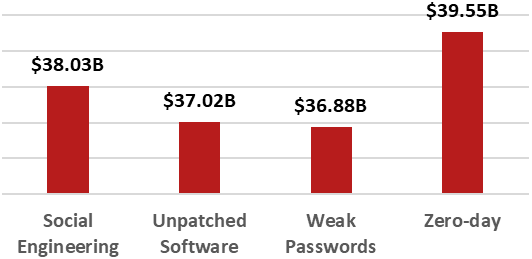
**Financial Loss by Security Vulnerability**

* The most exploited vulnerability was Zero-day attacks as it caused the loss of $39.55B.
* Social Engineering followed closely, responsible for $38.03B in losses.
* Weak Passwords is also a major risk as it was responsible for the loss of $36.88B.

**Insights**

* Human behavior remains the weakest link in cybersecurity—social engineering and weak passwords together contributed to nearly $75B in losses.
* Password management practices and multi-factor authentication should be prioritized to reduce the impact of easily exploitable weaknesses.
* The high loss from zero-day vulnerabilities highlights the critical need for proactive threat detection and faster patch deployment.

**Financial Loss by Security Vulnerability**



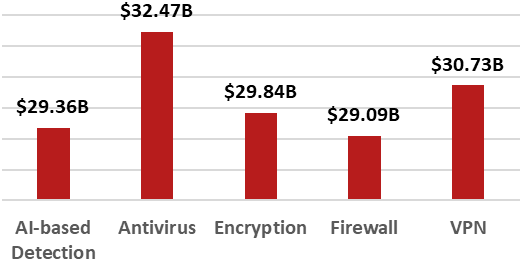
**Financial Loss by Defense Mechanism Used**

* Firewall ($29.09B) had the lowest financial loss, indicating better effectiveness.
* Encryption ($29.84B) and AI-based Detection ($29.36B) followed closely, showing relatively good performance but not as effective as Firewalls.
* Antivirus ($32.47B) and VPN ($30.73B) had a high effect on financial loss meaning they were the most compromised.

**Insights**

* The relatively lower losses with Firewalls suggest they are still the most effective defense mechanism.
* Technologies like AI-based Detection and Encryption show promise, performing better than traditional tools but still needing refinement.
* Antivirus and VPNs, though widely used, appear increasingly vulnerable to advanced attack methods, highlighting the need for modern, adaptive security layers.
* Regular updates, threat detection systems, and zero-trust frameworks should complement existing mechanisms to reduce vulnerabilities.

**Financial Loss by Defense Mechanism Used**



**Threat Type Distribution**

**Distribution of Users Affected by Attack Types**

Among the six major cybersecurity threats analyzed from 2015 to 2024**:**

* **DDoS** attacks affected the highest number of users, accounting for 17.5% of the total affected population (approximately 265.20 million users).
* **SQL Injection** and **Phishing attacks** followed closely, each impacting around 17.0% of users (257.77M and 257.72M respectively).
* **Ransomware** affected 16.4% of users (247.89M), while **Malware** and **Man-in-the-Middle (MitM)** attacks accounted for 16.3% (246.76M) and 15.8% (238.71M) of the total respectively.

****

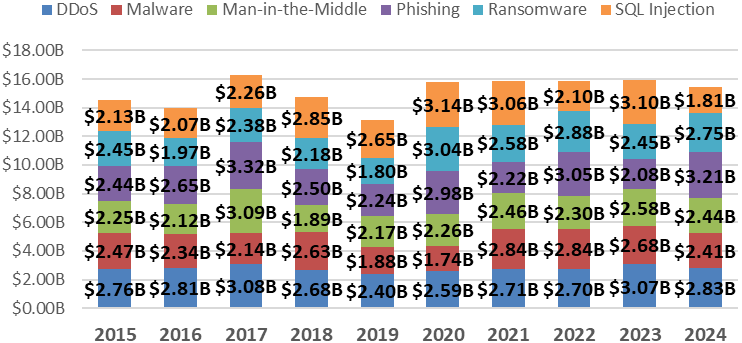
**Insights**

* These figures suggest that although user impact is fairly evenly distributed across these attack types, DDoS attacks have had the most widespread reach over the decade.

**Yearly Trend of Attack Types**

**Yearly Trend of Attack Types by Financial Loss**

* The years 2017 ($16.26B) and 2023 ($15.96B) witnessed the highest financial losses, with Phishing ($3.32B) and SQL Injection ($3.10B) being the major contributors, respectively.
* A noticeable dip occurred in 2019, followed by a steady increase in losses from 2020 to 2023.

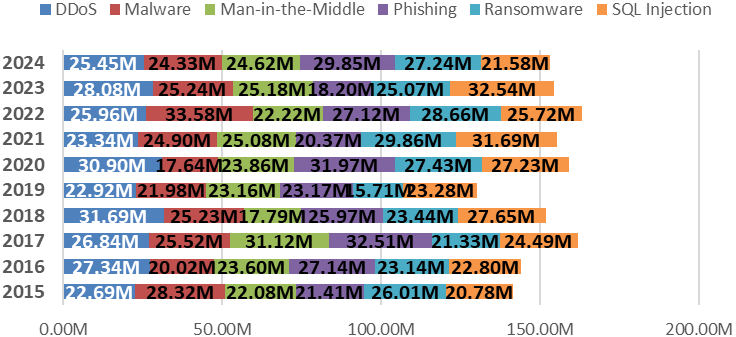


**Yearly Trend of Attack Types by Financial Loss**

**Yearly Trend of Attack Types by Users Affected**

* More users were affected in 2022 (163.26M), followed closely by 2017 (161.81M).
* 2019 recorded the lowest number of affected users, marking a notable dip in user impact that year.

**Yearly Trend of Attack Types by Users Affected**



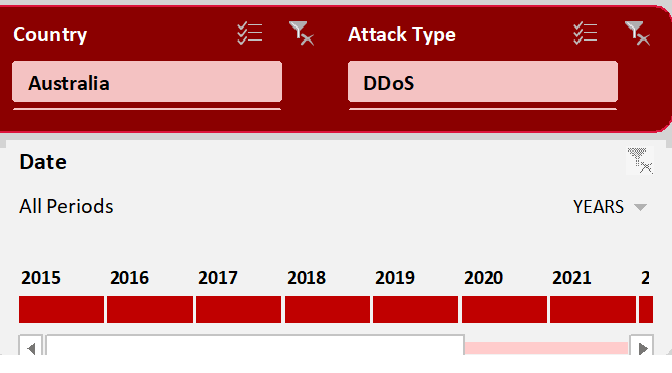
**Insights**

* Cyberattacks are becoming increasingly financially damaging, with millions of users affected annually.
* The consistent rise in both losses and user impact emphasizes the urgent need for stronger cybersecurity measures to reduce long-term consequences.

**Filters for Interactive Analysis**

To support dynamic analysis, the following filters were added:

* **Country** – To examine threat distributions across different regions.
* **Attack Type** – To compare the frequency and impact of various cyberattack methods.
* Year – To explore trends and changes in threat over time.



**Key Takeaways**

* UK recorded the highest financial loss, while Brazil had the highest number of affected users.
* The IT industry was the most impacted, threat-wise and financial-loss wise.
* The Nation-state attack source caused the highest financial loss, while Hacker Group had the least impact on financial loss.
* Zero-Day was the most exploited vulnerability.
* Firewall was the most effective defense mechanism.
* The year 2017 recorded the highest financial loss, while more users were affected in 2022.
* DDoS attacks affected the highest number of users.

**5. Conclusion and Recommendations**

This analysis, spanning data from 2015 to 2024, provides a comprehensive overview of how different attack types, sources, and defense mechanisms have influenced both financial loss and user impact worldwide. The findings offer actionable insights into the most critical areas of vulnerability and highlight opportunities for stronger, more proactive cybersecurity strategies.

**Key Findings**

* Countries such as the **United Kingdom**, **Germany**, and **Brazil** bore the heaviest financial burdens, while **Brazil**, **the UK**, and **France** saw the highest number of users affected. This suggests that these countries are frequent targets, likely due to their scale, data presence, and vulnerability to disruption.
* **The IT industry** was the most impacted sector, both in terms of financial losses and number of users affected, suggesting it's a primary target for attackers.
* **Nation-state** attacks were the most financially damaging, closely followed by breaches from **unknown sources** and **insider threats**. This highlights that both well-organized external actors and internal organizational weaknesses present serious risks.
* **Zero-Day** vulnerabilities were the most exploited type of security flaw, underscoring the importance of proactive patching and vulnerability detection.
* Among defense mechanisms analyzed, **firewalls** corresponded with the least financial loss, suggesting they remain highly effective when properly configured and maintained. In contrast, **VPNs** and **Antivirus software** were associated with higher losses, potentially reflecting outdated configurations or being common points of failure.
* **DDoS** emerged as the most common and damaging threat type, accounting for the highest percentage of attacks and users affected.
* **DDoS**, **SQL Injection**, and **Phishing** attacks also contributed significantly to financial losses, pointing to persistent technical vulnerabilities and insufficient infrastructure defenses across organizations.
* There has been a rising trend in both financial losses and user impact, especially from 2020 onward. Years like **2017** and **2023** recorded the highest cumulative losses, while **2022** affected the most users, showing that cyber incidents are becoming increasingly severe and large-scale.

**Recommendations**

Based on the insights derived from the data, the following actions are recommended to prevent future cyber threats and reduce their financial and human impact:

1. **Strengthen Employee Awareness and Training:**

* Implement regular cybersecurity training programs to educate employees on identifying phishing attempts, social engineering, and suspicious behavior.
* Enforce strict protocols for handling sensitive data and login credentials.
* Simulate phishing attacks internally to test and improve employee response.

1. **Improve Cybersecurity Readiness in High-Risk Sectors:**

* Target investments and resources towards industries like IT, which were the most affected.
* Set industry-specific standards and regulations to guide compliance and preparedness.
* Share threat intelligence within and across sectors to anticipate and prevent widespread attacks.

1. **Improve Patch Management and Zero-Day Response:**

* Establish a rapid response protocol for patching vulnerabilities once they’re discovered.
* Subscribe to global vulnerability databases and security advisories to stay informed.
* Encourage vendors to release timely updates and test patches before full deployment.

1. **Adopt Stronger Data Backup and Recovery Policies:**

* Maintain secure, encrypted backups that are regularly tested for integrity and availability.
* Store backups both on-site and off-site or in the cloud to prevent ransomware lockdowns.
* Create clear business continuity and disaster recovery plans.

1. **Increase Government and International Collaboration:**

* Promote public-private partnerships for sharing cyber threat data and best practices.
* Encourage the establishment of cross-border response units to address global cybercrime.
* Support global efforts to track, identify, and prosecute nation-state-sponsored or transnational attackers.

**Limitations of the Analysis**

Despite the insights gained from this cybersecurity threat analysis, certain limitations should be noted:

1. **Limited Time Frame of Data (2015–2024):**

* The dataset spans only ten years, which may not capture longer-term trends or the full evolution of cyberattack techniques.

**Recommendation:**

* Extend future analyses to include historical data beyond 2015 and monitor trends continuously in real-time to build a more comprehensive outlook**.**

1. **Lack of Granular Industry Breakdown:**

* While the IT sector was identified as the most affected, the dataset did not provide detailed sub-industry classifications (e.g., fintech, software, hardware).

**Recommendation:**

* Collect or request data with deeper segmentation to better understand vulnerabilities within specific industry niches and tailor defenses accordingly.

1. **Incomplete Attribution of Attack Sources:**

* A large number of cyberattacks were attributed to "unknown" sources, limiting insight into attacker profiles and motivations.

**Recommendation:**

* Strengthen attribution efforts through collaboration with cybersecurity intelligence agencies and by using forensic tools that trace digital footprints more accurately.

1. **Financial Loss Estimates May Be Conservative or Incomplete:**

* The reported financial losses may not capture hidden costs such as reputational damage, long-term operational disruptions, or legal penalties.

**Recommendation:**

* Incorporate broader economic impact assessments, including indirect and intangible costs, in future evaluations for a more realistic financial impact scope.

1. **User Impact Was Measured in Volume, Not Severity:**

* The dataset quantifies users affected but doesn’t account for severity, such as data stolen, identity theft, or duration of exposure.

**Recommendation:**

* Future studies should include severity metrics, perhaps through user surveys or breach detail reports, to better understand the actual harm caused to individuals.

**Summary**

This analysis highlights the scale and complexity of global cybersecurity threats over the last decade. Financial and user impact trends emphasize the urgent need for improved security measures across industries and countries.

By implementing the recommendations in this report:

* Organizations will be better positioned to prevent cyberattacks, respond to them swiftly, and reduce their overall impact.
* Enhanced training, upgraded security technologies, and regular system assessments will collectively lead to fewer breaches, lower financial losses, and improved data protection across sectors.

As these actions are put into practice, continued monitoring and future assessments will be essential to track progress, adjust strategies, and stay ahead of evolving threats.