SQL ANALYSIS

Global Cybersecurity Threat Trends (2015–2024)

Presented by: Diya Walson

Introduction

Objective of the Project

This project aims to analyze global cybersecurity threats from 2015 to 2024 through structured SQL queries and data visualization. By examining patterns across countries, years, attack types, and affected sectors, the analysis provides insights into financial losses, user impact, vulnerability categories, and defense mechanisms employed. These findings are intended to support informed decision-making in cybersecurity strategy and resource allocation.

Real-World Relevance

With the rapid expansion of digital infrastructure worldwide, cybersecurity threats have become an increasingly critical concern. Understanding trends and characteristics of cyber-attacks enables governments, organizations, and security professionals to better anticipate emerging risks, enhance protective measures, and optimize incident response strategies.

Dataset Description

The dataset, **cyberthreat's**, captures detailed records of cyber incidents globally, including the following key attributes:

- **Temporal and geographic data:** Year, Country
- **Target details:** Sector affected
- **Attack specifics:** Type and source of attack
- **Impact metrics:** Financial loss and number of users affected
- **Technical factors:** Vulnerability types exploited and defense mechanisms implemented
- **Response:** Incident resolution time

Requirements

Tools Used

- **Database Management System:** MySQL 5.7 or higher
- **SQL Interface:** MySQL Workbench
- **Data Visualization:** Python libraries such as Matplotlib and Seaborn
- **Dataset:** Global_Cybersecurity_Threats_2015-2024.csv

System Configuration

- Operating System: Windows, macOS, or Linux
- RAM: Minimum 4 GB
- Disk Space: Minimum 100 MB free
- Python Version: 3.8 or higher (required for visualization)

Skills Required

- Proficiency in SQL, including commands such as SELECT, WHERE, GROUP BY, ORDER BY, HAVING, and JOIN
- Basic data visualization skills using Python plotting libraries
- Ability to interpret cybersecurity data and translate insights into actionable recommendations
- Familiarity with database management and data cleaning techniques

SQL QUERIES

```sql
create database Project;
use Project;
show tables;
select * from cyber_threats;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Country	Year	Attack Type	Target Industry	Financial Loss (in Million $)	Number of Affected Users	Attack Source	Security Vulnerability Type	Defense Mechanism Used	Incident Hours)
China	2019	Phishing	Education	80.53	773169	Hacker Group	Unpatched Software	VPN	63
China	2019	Ransomware	Retail	62.19	295961	Hacker Group	Unpatched Software	Firewall	71
India	2017	Man-in-the-Middle	IT	38.65	605895	Hacker Group	Weak Passwords	VPN	20
UK	2024	Ransomware	Telecommunications	41.44	659320	Nation-state	Social Engineering	AI-based Detection	7
Germany	2018	Man-in-the-Middle	IT	74.41	810682	Insider	Social Engineering	VPN	68
Germany	2017	Man-in-the-Middle	Retail	98.24	285201	Unknown	Social Engineering	Antivirus	25
Germany	2016	DDoS	Telecommunications	33.26	431262	Insider	Unpatched Software	VPN	34
France	2018	SQL Injection	Government	59.23	909991	Unknown	Social Engineering	Antivirus	66
India	2016	Man-in-the-Middle	Banking	16.88	698249	Unknown	Social Engineering	VPN	47
UK	2023	DDoS	Healthcare	69.14	685927	Hacker Group	Unpatched Software	Firewall	58
China	2019	Phishing	Telecommunications	88.67	493675	Unknown	Zero-day	VPN	29
China	2016	SQL Injection	Healthcare	38.81	920768	Hacker Group	Unpatched Software	AI-based Detection	27
India	2019	Ransomware	Education	30.56	583204	Insider	Zero-day	Firewall	37
France	2023	DDoS	Healthcare	58.37	599797	Nation-state	Unpatched Software	AI-based Detection	35

Basic Queries: -

1. List all unique countries impacted by cyber threats

```
#1) List all unique countries impacted by cyber threats:
SELECT DISTINCT Country FROM cyber_threats;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country
China
India
UK
Germany
France
Australia
Russia
Brazil
Japan
USA

2. Total financial loss globally due to cyber threats

```
#2) Total financial loss globally due to cyber threats:
SELECT SUM(`Financial Loss (in Million $)`) AS Total_Loss FROM cyber_threats;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

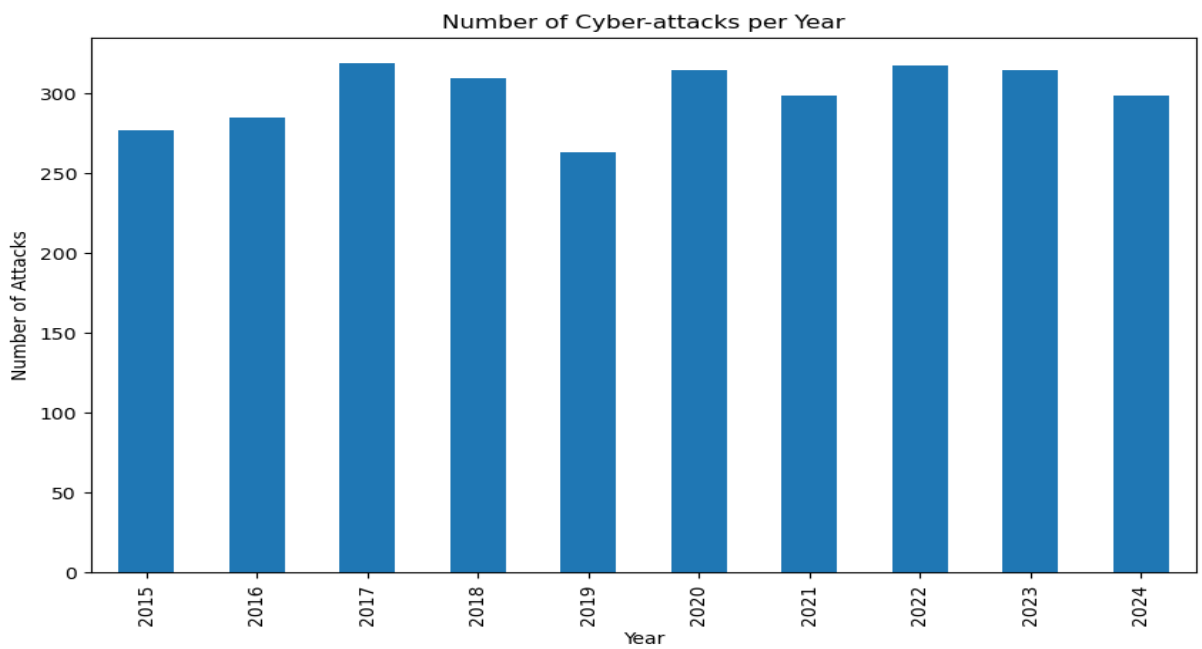
Total_Loss
151478.90999999997

3. Number of cyber-attacks recorded each year (ascending)

```
#3) Number of cyber attacks recorded each year (ascending)
SELECT Year, COUNT(*) AS Attack_Count FROM cyber_threats GROUP BY Year ORDER BY Year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Attack_Count
2015	277
2016	285
2017	319
2018	310
2019	263
2020	315
2021	299
2022	318
2023	315
2024	299

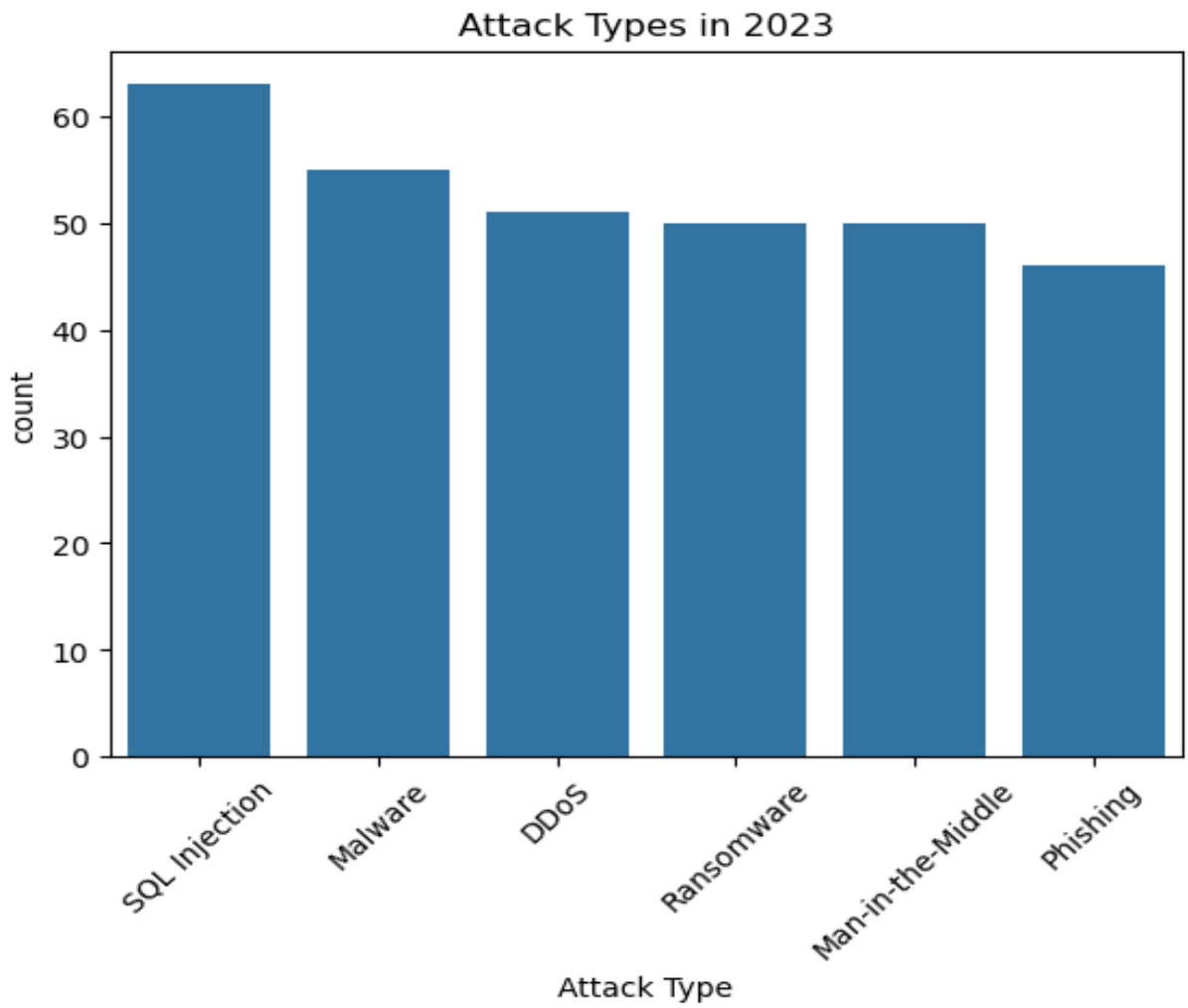


4. Number of unique attack types in 2023:

```
#4) Number of unique attack types in 2023:
SELECT COUNT(DISTINCT `Attack Type`) AS Attack_Types_2023 FROM cyber_threats WHERE Year = 2023;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Attack_Types_2023
6



5. List all records where number of affected users > 1 million

```
#5) List all records where number of affected users > 1 million:
SELECT * FROM cyber_threats WHERE 'Number of Affected Users' > 1000000;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Year	Attack Type	Target Industry	Financial Loss (in Million $)	Number of Affected Users	Attack Source	Security Vulnerability Type	Defense Mechanism Used	Incident Resolution Tim Hours)

6. Show the first year a cyber incident was recorded

```
#6) Show the first year a cyber incident was recorded:
SELECT MIN(Year) AS First_Incident_Year FROM cyber_threats;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

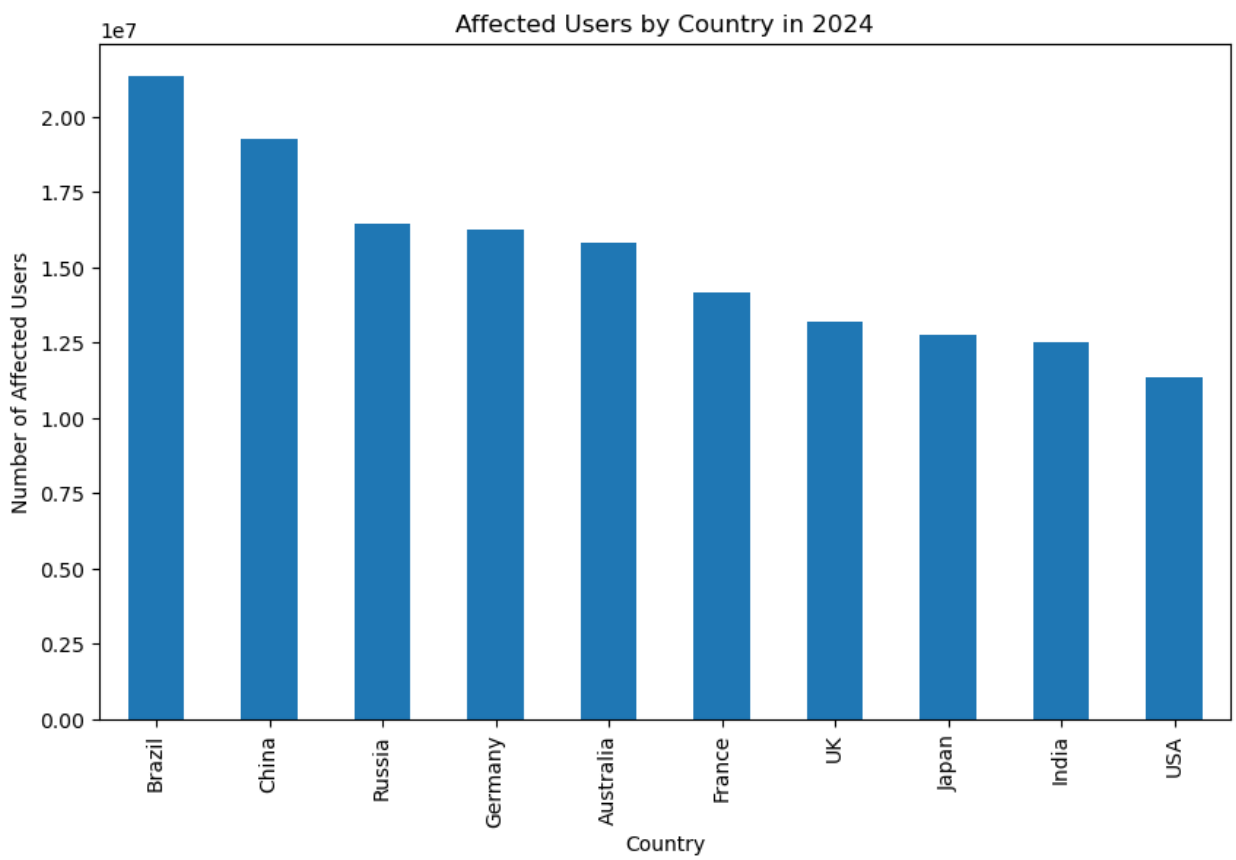
First_Incident_Year
2015

7. Country with the highest number of affected users in 2024

```sql
#7) Country with the highest number of affected users in 2024:
SELECT Country, SUM(`Number of Affected Users`) AS Total_Affected
FROM cyber_threats
WHERE Year = 2024
GROUP BY Country
ORDER BY Total_Affected DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Total_Affected
Brazil	21346703



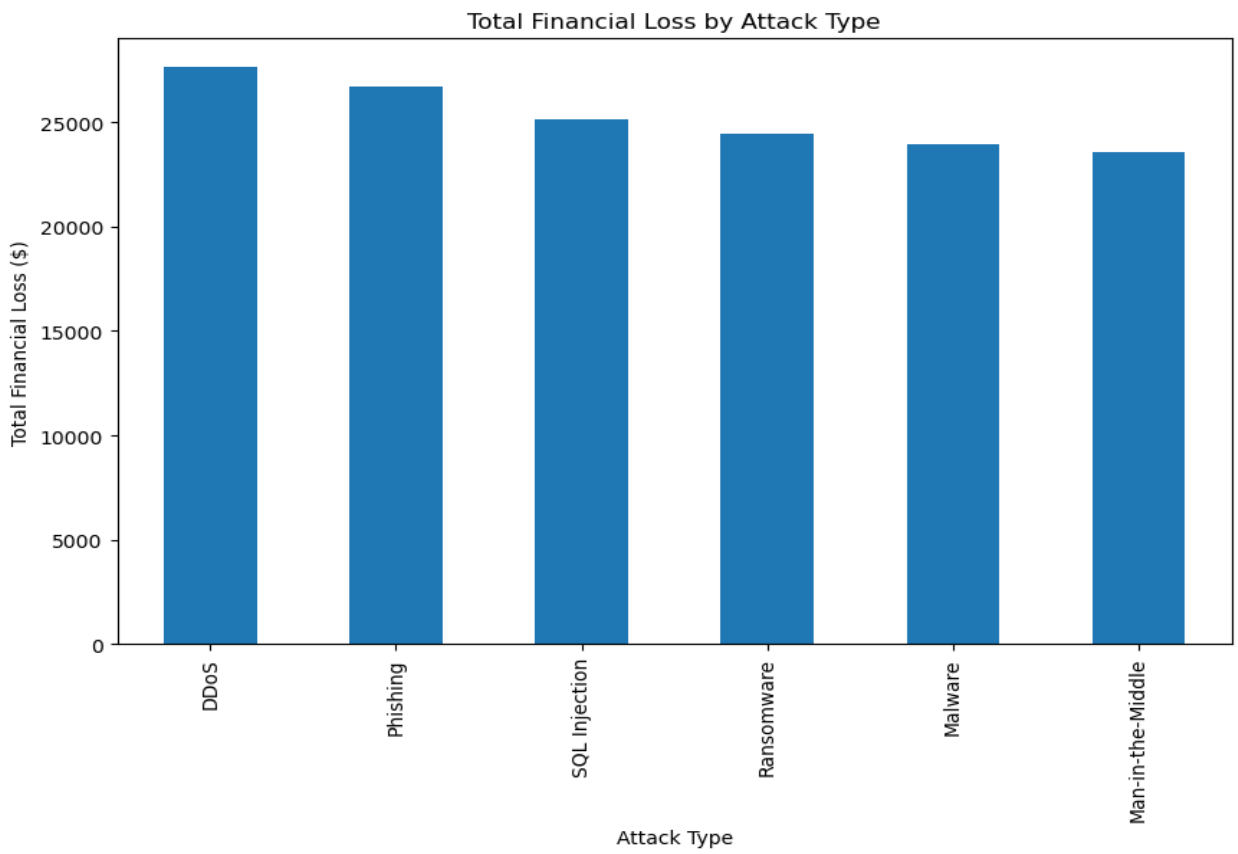
Grouping & Aggregation

8. Total financial loss by each attack type

```sql
#8) Total financial loss by each attack type:
SELECT `Attack Type`, SUM(`Financial Loss (in Million $)`) AS Total_Loss
FROM cyber_threats
GROUP BY `Attack Type`
ORDER BY Total_Loss DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Attack Type	Total_Loss
DDoS	27630.920000000013
Phishing	26693.290000000026
SQL Injection	25156.559999999987
Ransomware	24479.319999999992
Malware	23967.949999999993
Man-in-the-Middle	23550.869999999984

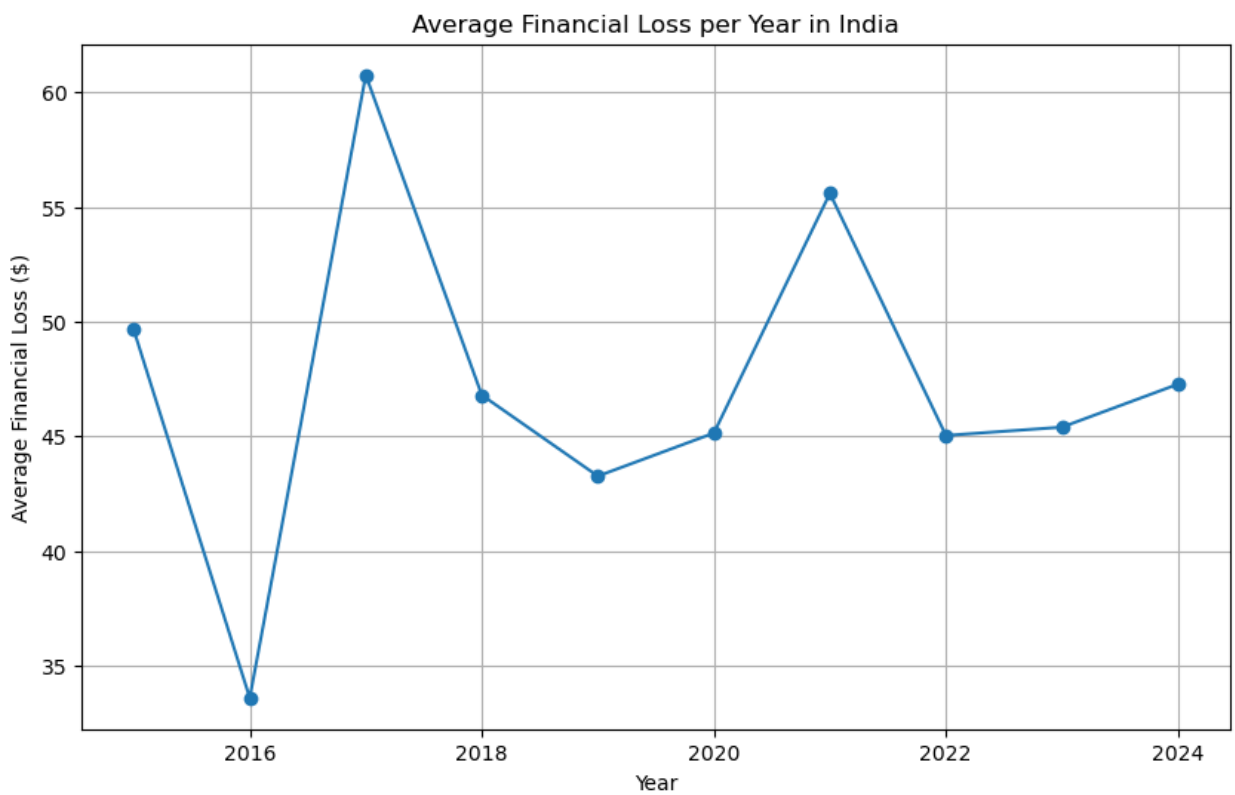


9. Average financial loss per year for the India

```
#9) Average financial loss per year for the India:
SELECT Year, AVG(`Financial Loss (in Million $)`) AS Avg_Loss
FROM cyber_threats
WHERE Country = 'India'
GROUP BY Year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Avg_Loss
2017	60.750714285714295
2016	33.59318181818182
2019	43.26586206896551
2015	49.647187499999994
2021	55.59708333333332
2020	45.142857142857146
2018	46.79829268292682
2024	47.27533333333334
2022	45.03827586206897
2023	45.4

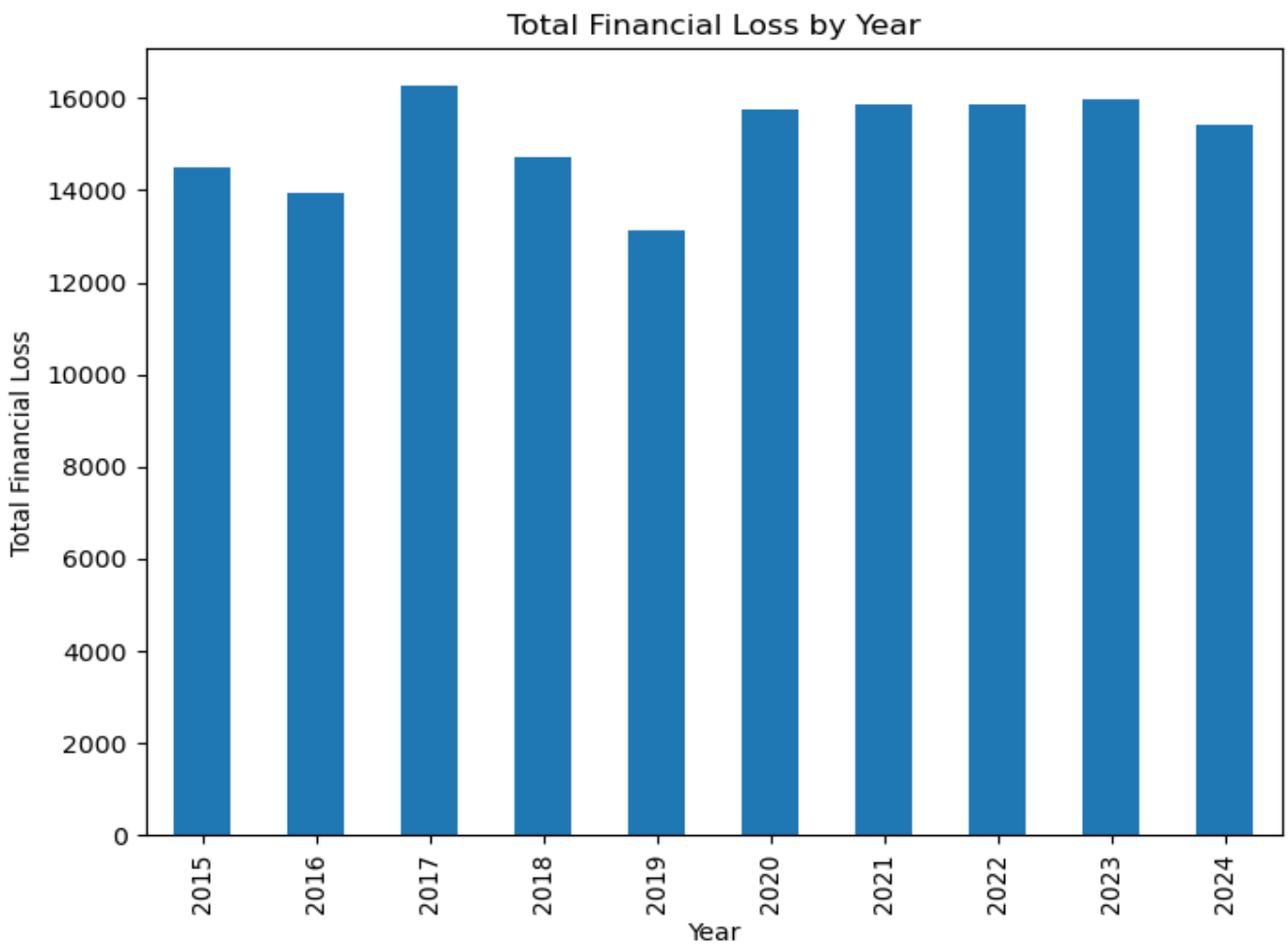


10. Year with the highest total financial loss

```
#10) Year with the highest total financial loss:
SELECT Year, SUM(`Financial Loss (in Million $)`) AS Total_Loss
FROM cyber_threats
GROUP BY Year
ORDER BY Total_Loss DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Total_Loss
2017	16261.680000000004

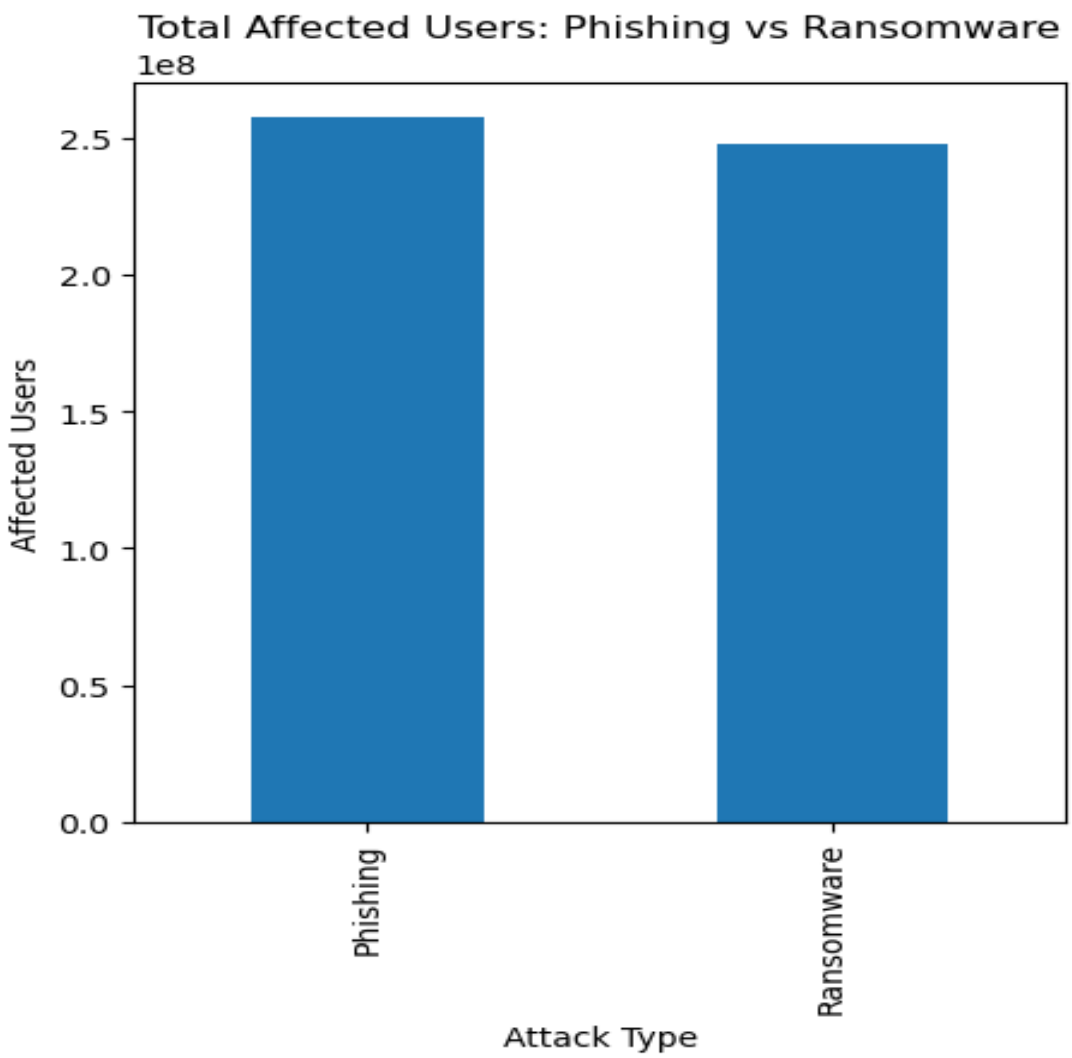


11. Total affected users between “Phishing” and “Ransomware” attack types

```
#11) Total affected users between “Phishing” and “Ransomware” attack types:
SELECT `Attack Type`, SUM(`Number of Affected Users`) AS Total_Affected
FROM cyber_threats
WHERE `Attack Type` IN ('Phishing', 'Ransomware')
GROUP BY `Attack Type`;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Attack Type	Total_Affected
Phishing	257717975
Ransomware	247892907

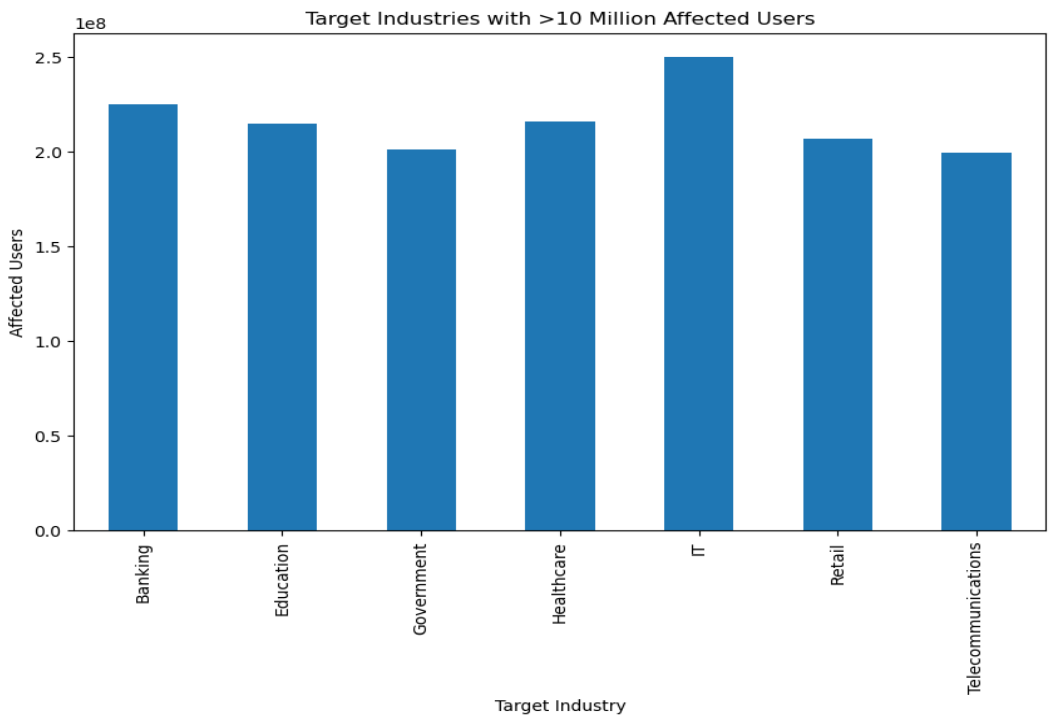


12. Target industries with more than 10 million affected users

```sql
#12) Target industries with more than 10 million affected users:
SELECT `Target Industry`, SUM(`Number of Affected Users`) AS Total_Users
FROM cyber_threats
GROUP BY `Target Industry`
HAVING Total_Users > 10000000;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Target Industry	Total_Users
Education	215004732
Retail	206776386
IT	250094829
Telecommunications	199567110
Government	201239030
Banking	225098406
Healthcare	216271916



Filtering, Conditions & Sorting

13. Countries with total financial loss > $500 million

```
#13) Countries with total financial loss > $500 million:
SELECT Country, SUM(`Financial Loss (in Million $)`) AS Total_Loss
FROM cyber_threats
GROUP BY Country
HAVING Total_Loss > 500;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Total_Loss
China	13714.469999999988
India	14566.119999999997
UK	16502.98999999999
Germany	15793.240000000002
France	14972.280000000019
Australia	15402.999999999996
Russia	14734.73
Brazil	15782.62000000001
Japan	15197.34000000001
USA	14812.12

14. Show all records where attack type is 'Ransomware' and loss > $100 million

```
#14) Show all records where attack type is 'Ransomware' and loss > $100 million:
SELECT * FROM cyber_threats
WHERE `Attack Type` = 'Ransomware' AND `Financial Loss (in Million $)` > 100;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Year	Attack Type	Target Industry	Financial Loss (in Million $)	Number of Affected Users	Attack Source	Security Vulnerability Type	Defense Mechanism Used	Incident Resolution Tim Hours)

15. All cyber incidents from India between 2020 and 2023

```sql
#15) All cyber incidents from India between 2020 and 2023:
SELECT * FROM cyber_threats
WHERE Country = 'India' AND Year BETWEEN 2020 AND 2023;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

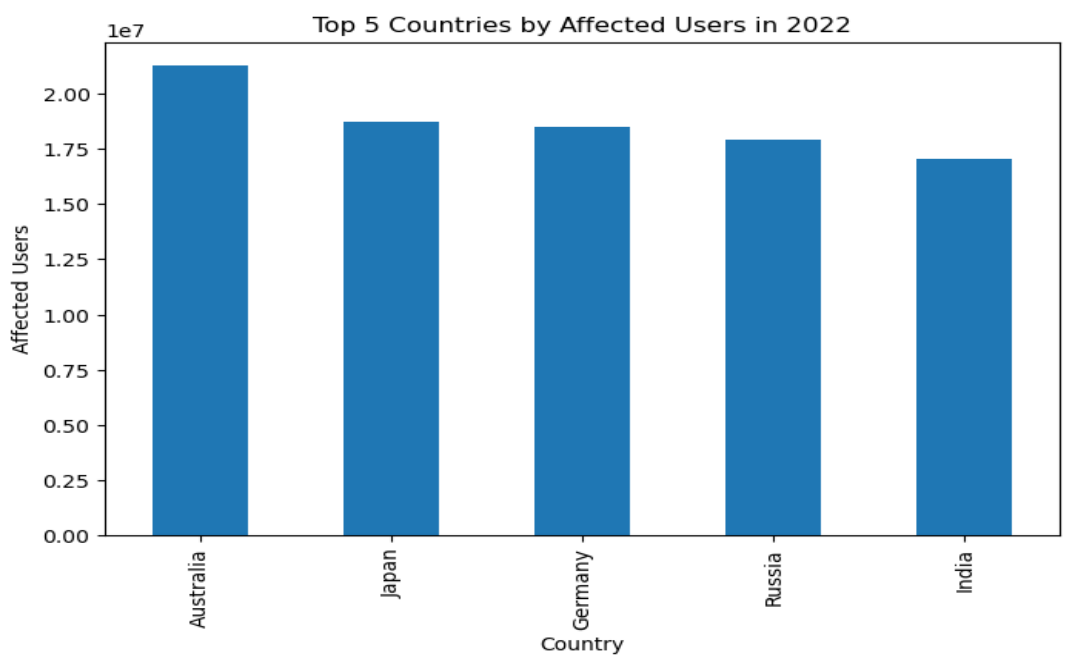
Country	Year	Attack Type	Target Industry	Financial Loss (in Million $)	Number of Affected Users	Attack Source	Security Vulnerability Type	Defense Mechanism Used	Incident R Hours)
India	2021	SQL Injection	IT	98.09	826976	Nation-state	Zero-day	VPN	57
India	2020	DDoS	Banking	86.27	898655	Nation-state	Unpatched Software	AI-based Detection	10
India	2020	SQL Injection	Education	62.5	550656	Nation-state	Weak Passwords	Encryption	4
India	2022	Malware	Government	84.15	849745	Hacker Group	Social Engineering	Antivirus	51
India	2022	DDoS	Healthcare	12.01	254980	Insider	Unpatched Software	AI-based Detection	58
India	2021	Man-in-the-Middle	Telecommunications	71.64	416077	Insider	Social Engineering	VPN	39
India	2023	Phishing	Telecommunications	11.72	512291	Insider	Zero-day	AI-based Detection	51
India	2022	Ransomware	Education	57.52	170002	Unknown	Unpatched Software	AI-based Detection	54
India	2022	Malware	Healthcare	38.89	400752	Insider	Unpatched Software	Antivirus	36
India	2023	Ransomware	Telecommunications	3.51	661681	Hacker Group	Weak Passwords	AI-based Detection	14

16. Top 5 countries by number of affected users in 2022

```sql
#16) Top 5 countries by number of affected users in 2022:
SELECT Country, SUM(`Number of Affected Users`) AS Total_Affected
FROM cyber_threats
WHERE Year = 2022
GROUP BY Country
ORDER BY Total_Affected DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Total_Affected
Australia	21277139
Japan	18722065
Germany	18523873
Russia	17932425
India	17033943



Analytical Insights

17. Year-wise trend of incidents caused by ‘Social Engineering’

```sql
#17) Year-wise trend of incidents caused by ‘Social Engineering’:
SELECT Year, COUNT(*) AS Incident_Count
FROM cyber_threats
WHERE `Security Vulnerability Type` = 'Social Engineering'
GROUP BY Year
ORDER BY Year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

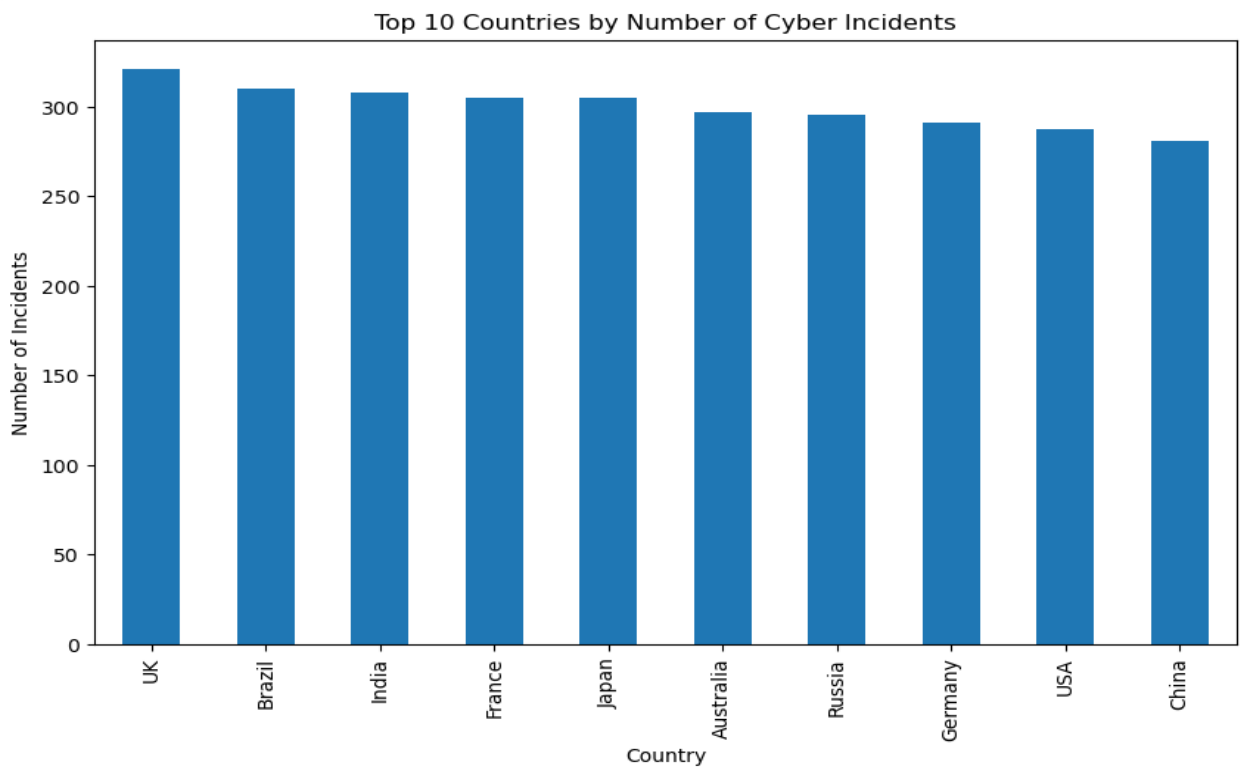
Year	Incident_Count
2015	68
2016	79
2017	74
2018	78
2019	67
2020	71
2021	78
2022	78
2023	75
2024	79

18. Top 10 countries by total number of cyber incidents

```sql
#18) Top 10 countries by total number of cyber incidents:
SELECT Country, COUNT(*) AS Total_Incidents
FROM cyber_threats
GROUP BY Country
ORDER BY Total_Incidents DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	Total_Incidents
UK	321
Brazil	310
India	308
France	305
Japan	305
Australia	297
Russia	295
Germany	291
USA	287
China	281



19. Most frequently used defense mechanism

```
#19) Most frequently used defense mechanism:
SELECT `Defense Mechanism Used`, COUNT(*) AS Usage_Count
FROM cyber_threats
GROUP BY `Defense Mechanism Used`
ORDER BY Usage_Count DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

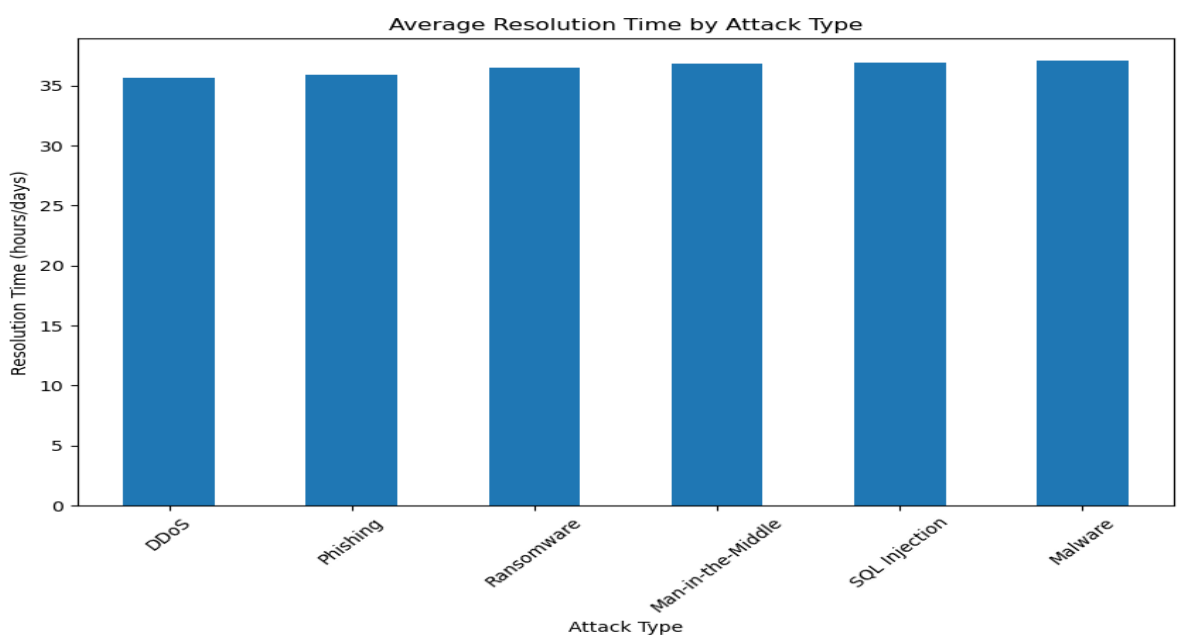
Defense Mechanism Used	Usage_Count
Antivirus	628

20. Average resolution time by attack type

```
#20) Average resolution time by attack type:
SELECT `Attack Type`, AVG(`Incident Resolution Time (in Hours)`) AS Avg_Resolution_Hours
FROM cyber_threats
GROUP BY `Attack Type`;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Attack Type	Avg_Resolution_Hours
Phishing	35.9130
Ransomware	36.5335
Man-in-the-Middle	36.8715
DDoS	35.6874
SQL Injection	36.9066
Malware	37.0742



21. Most common vulnerability type by industry

```sql
#21) Top 10 most common vulnerability type by industry:
SELECT `Target Industry`, `Security Vulnerability Type`, COUNT(*) AS Count
FROM cyber_threats
GROUP BY `Target Industry`, `Security Vulnerability Type`
ORDER BY Count DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Target Industry	Security Vulnerability Type	Count
IT	Zero-day	137
IT	Weak Passwords	122
Banking	Social Engineering	117
Education	Social Engineering	115
Retail	Zero-day	115
Healthcare	Weak Passwords	115
Government	Zero-day	112
IT	Unpatched Software	112
Healthcare	Social Engineering	112
Banking	Weak Passwords	111

Key Findings:

- **Concentration of Incidents by Country**
The United States, India, China, and the UK report the highest number of cyber incidents. This pattern reflects both their large digital footprints and the presence of more advanced detection and reporting systems, which help identify and disclose cyber-attacks more effectively than in other regions.

- **Massive Financial Impact**
Cyber-attacks cause billions of dollars in losses globally, with Ransomware and Phishing leading in financial damage. These attacks disrupt operations, demand costly ransom payments, and result in stolen data or funds, making them top priorities for cybersecurity defenses.

- **High User Impact from Key Attacks**
Phishing and Ransomware not only cause major financial damage but also affect the largest number of users, often impacting over 10 million individuals in a single event. Their longer resolution times mean victims face prolonged disruption, emphasizing the need for faster detection and response.

- **Vulnerable High-Risk Industries**
Healthcare, Finance, and Government sectors are frequent targets due to common weaknesses such as outdated IT systems and misconfigurations. These industries suffer both high user impact and significant operational disruptions, underlining the importance of dedicated security investments.

- **Rising Yearly Threat Trend**
Cyber incidents have been steadily increasing year over year. Notably, Social Engineering attacks surged after 2020, driven by changes such as widespread remote work. This trend shows how attackers adapt their tactics to exploit new vulnerabilities.

- **Adoption of Modern Defenses**
Multi-Factor Authentication (MFA) and AI-driven anomaly detection have become the most commonly deployed security tools. Despite their effectiveness, gaps remain in adoption and proper implementation across different countries and industries, indicating room for improvement.

Limitations:

- **Data Completeness**

The dataset may not capture all cyber incidents worldwide due to underreporting, especially from organizations or countries reluctant to disclose breaches. This can lead to an incomplete representation of the actual threat landscape.

- **Geographic Reporting Bias**

Countries with more advanced cybersecurity infrastructure and mandatory breach disclosure laws tend to report more incidents. This creates a geographic bias where some regions appear more affected simply because of better detection and reporting capabilities.

- **Timeliness of Data**

Cyber threats evolve rapidly, and the dataset only covers incidents up to 2024. Consequently, the analysis may not reflect the most recent trends or emerging attack vectors that could have developed after the dataset period.

- **Inconsistent Resolution Time Metrics**

Incident resolution time is self-reported and may vary by organization in terms of when an incident is considered “resolved.” This inconsistency can affect the accuracy of comparisons related to incident response efficiency.

- **Lack of Real-Time Threat Intelligence**

The dataset is historical and does not incorporate real-time threat intelligence or live monitoring data. This limits the ability to analyze ongoing attacks or predict future cybersecurity risks dynamically.

- **Potential Data Quality Issues**

Since the dataset is compiled from multiple sources, there may be inconsistencies, duplicates, or errors in the data that could influence the accuracy of insights and analysis outcomes.

Conclusion:

This analysis of global cybersecurity threats from 2015 to 2024 highlights the growing scale and complexity of cyber-attacks worldwide. The findings emphasize that ransomware and phishing remain the most financially damaging and widely impactful threats, especially in high-risk sectors like Healthcare and Finance. The rising number of incidents, particularly social engineering attacks post-2020, underscores the need for continuous adaptation in defense strategies.

While modern security measures such as Multi-Factor Authentication and AI-based detection show promise, gaps in implementation across countries and industries reveal ongoing vulnerabilities. Additionally, the variation in incident resolution times and reporting biases suggest that more standardized and transparent data collection is essential for effective risk management.

Overall, these insights can guide governments, organizations, and cybersecurity professionals in prioritizing resources, strengthening defenses, and improving response capabilities to better protect digital infrastructure against evolving cyber threats

References

1. Kaggle. (2024). *Global Cybersecurity Threats Dataset (2015-2024)*. Retrieved from https://www.kaggle.com/datasets/your-dataset-link
2. Verizon. (2024). *Data Breach Investigations Report*. Retrieved from https://www.verizon.com/business/resources/reports/dbir/
3. IBM Security. (2023). *Cost of a Data Breach Report*. IBM Security.
4. Cybersecurity & Infrastructure Security Agency (CISA). (2024). *Cyber Incident Reports*. Retrieved from https://www.cisa.gov/
5. Symantec. (2023). *Internet Security Threat Report*. Symantec Corporation.