DATA SCIENCE

PROJECT REPORT

(Project Semester January-April 2025)

Analysis of Global Cybersecurity Threats (2015–2024)

Submitted by

Nitya

Registration No: 12302869

B.Tech(IT)

Course Code: INT375

Under the Guidance of

Maneet Kaur

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Nitya bearing Registration no. 12302869 has completed INT375 project titled, “Analysis of Global Cybersecurity Threats (2015–2024)” under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort and study.

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 12-04-25

DECLARATION

I, Nitya, student of B.Tech(IT) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-25

Signature: Nitya

Registration No. 12302869

Nitya

ACKNOWLEDGEMENT

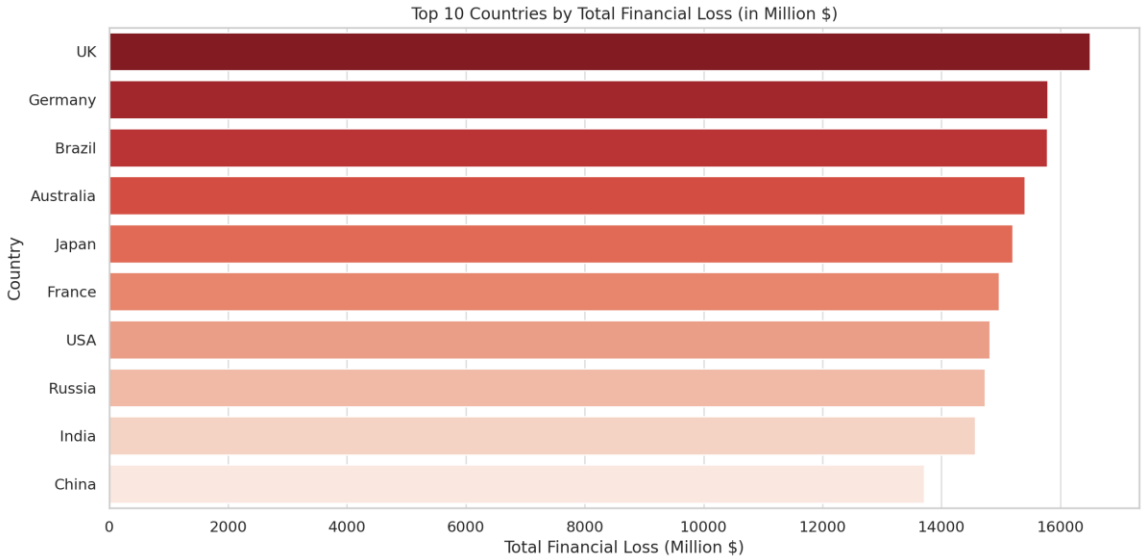
I would like to express my heartfelt gratitude to my faculty guide, Ms. Maneet Kaur, for her invaluable guidance and support throughout this project. I also thank Lovely Professional University for providing the platform and resources necessary to complete this Data Science Minor Project.

Dataset link

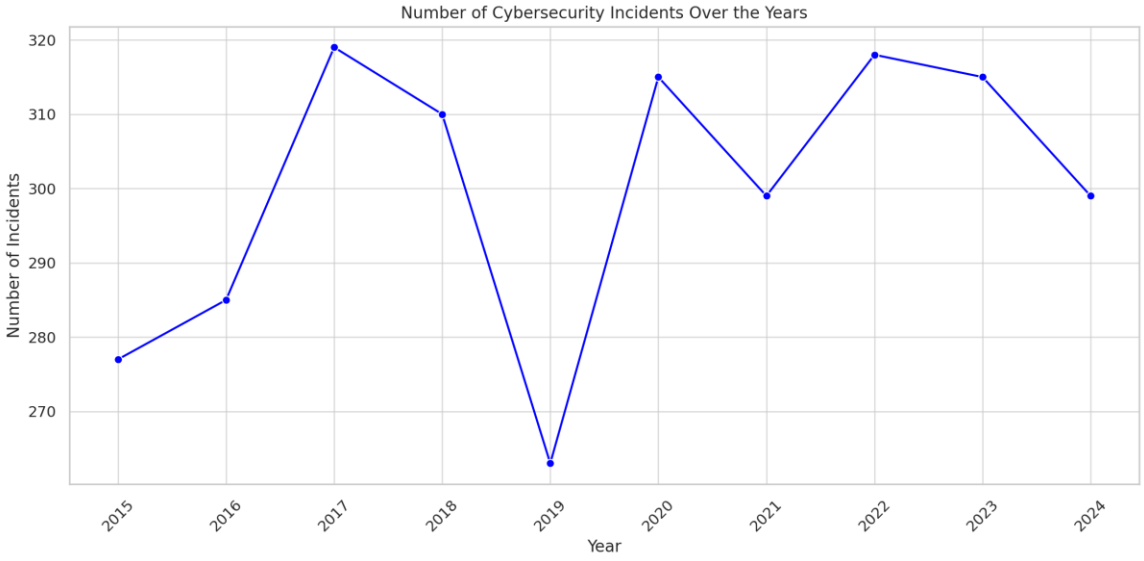
https://www.kaggle.com/datasets/sprasad018/global-cybersecurity-threats-2015-2024

Charts and Visualizations

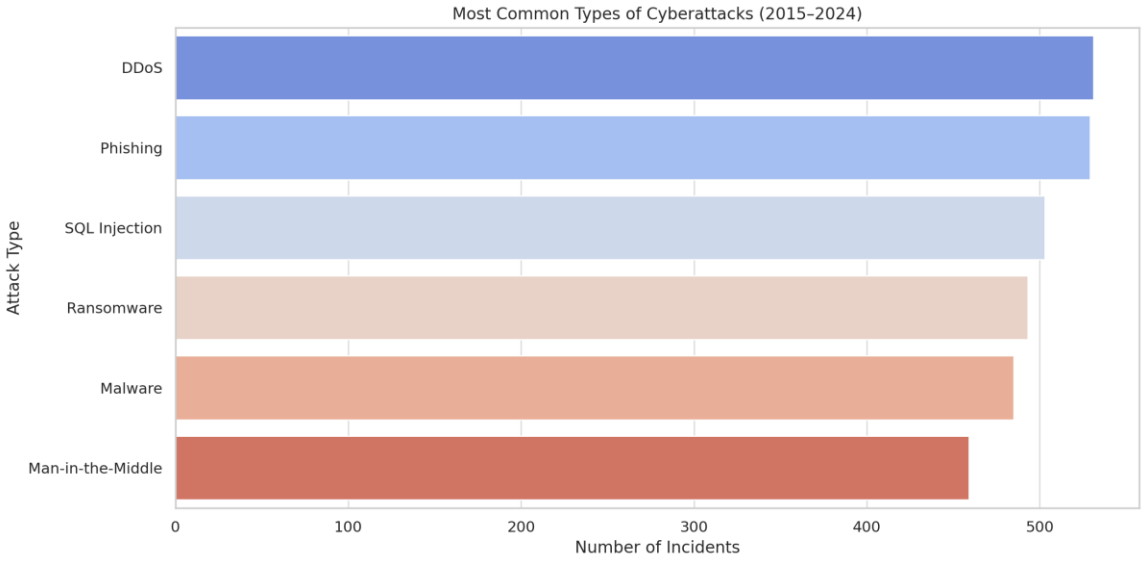
Top 10 Countries by Total Financial Loss (in Million $)



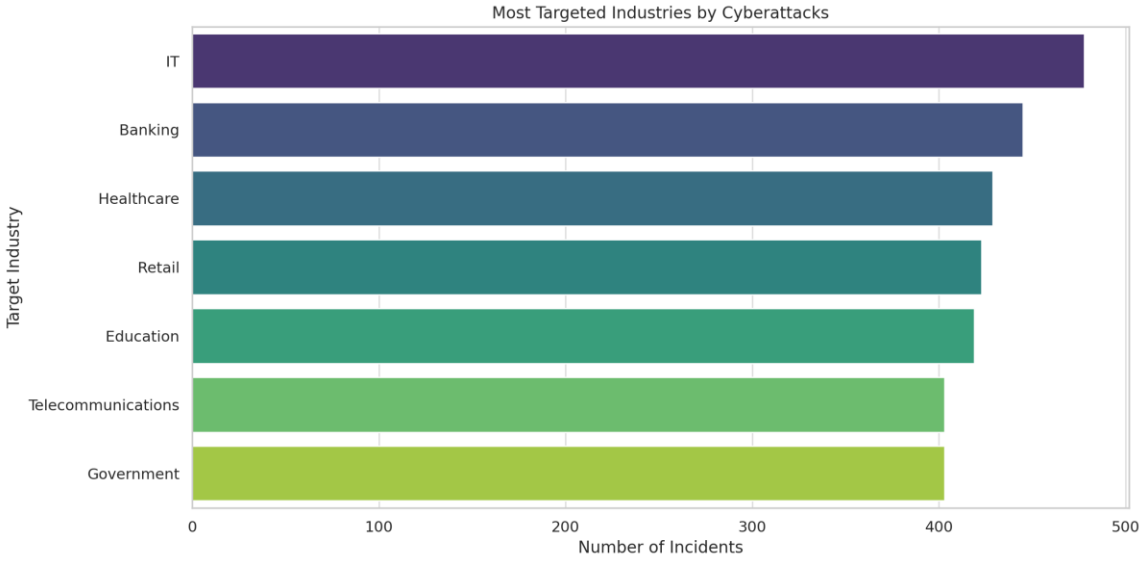
Number of Cybersecurity Incidents Over the Years



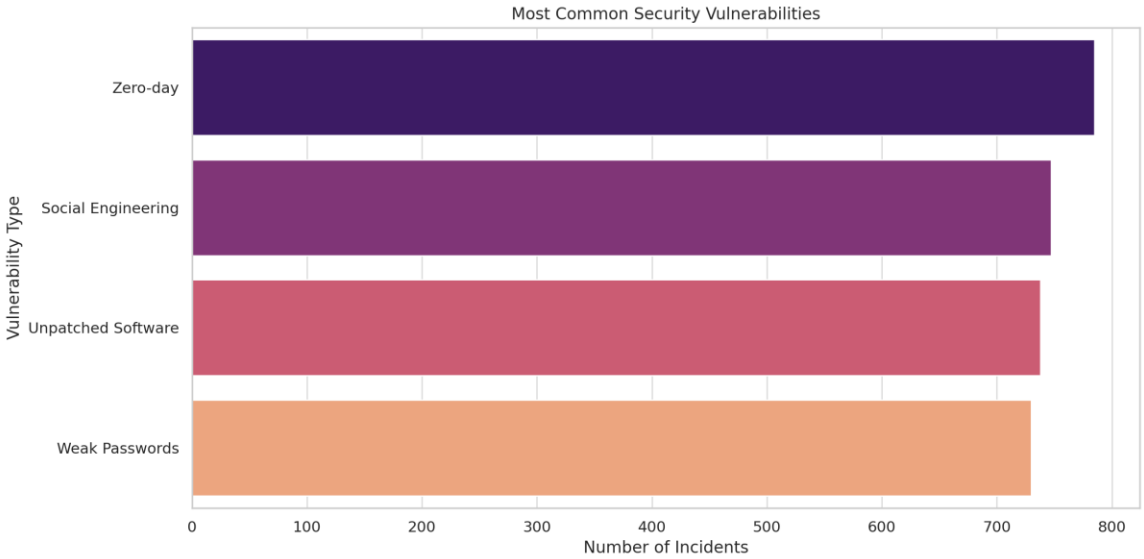
Most Common Types of Cyberattacks (2015–2024)



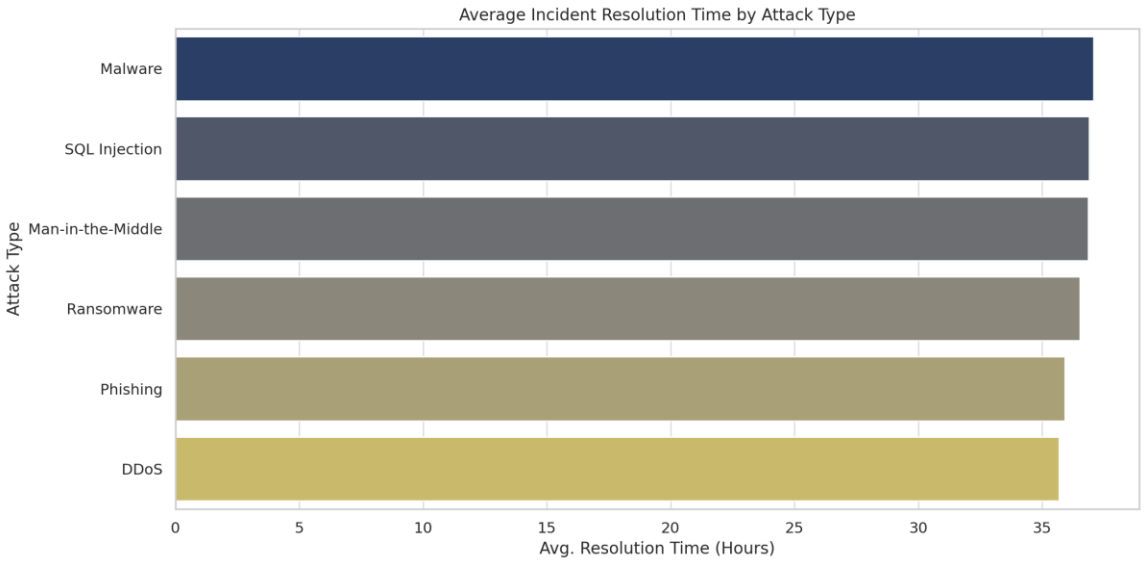
Most Targeted Industries by Cyberattacks



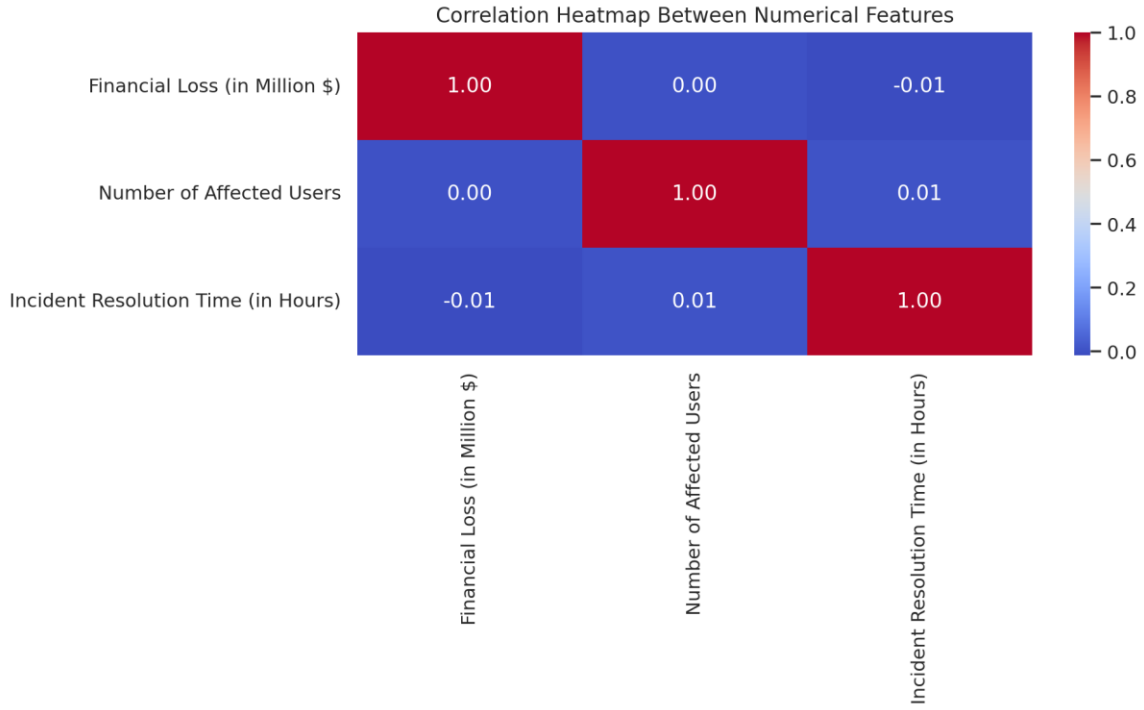
Most Common Security Vulnerabilities



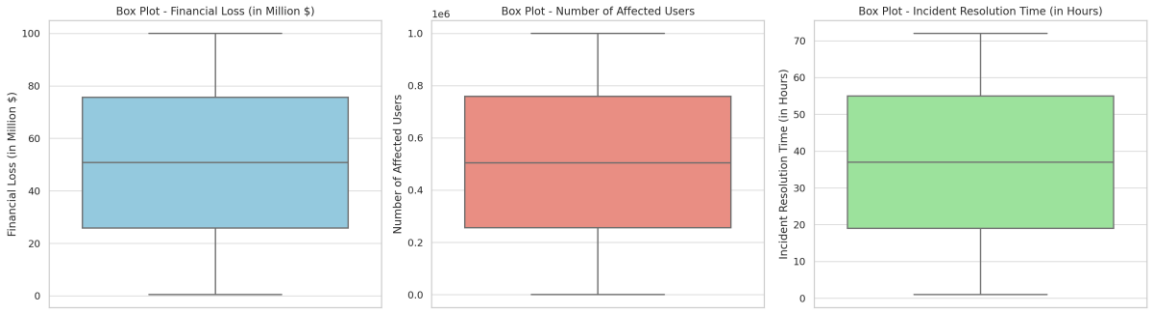
Average Incident Resolution Time by Attack Type



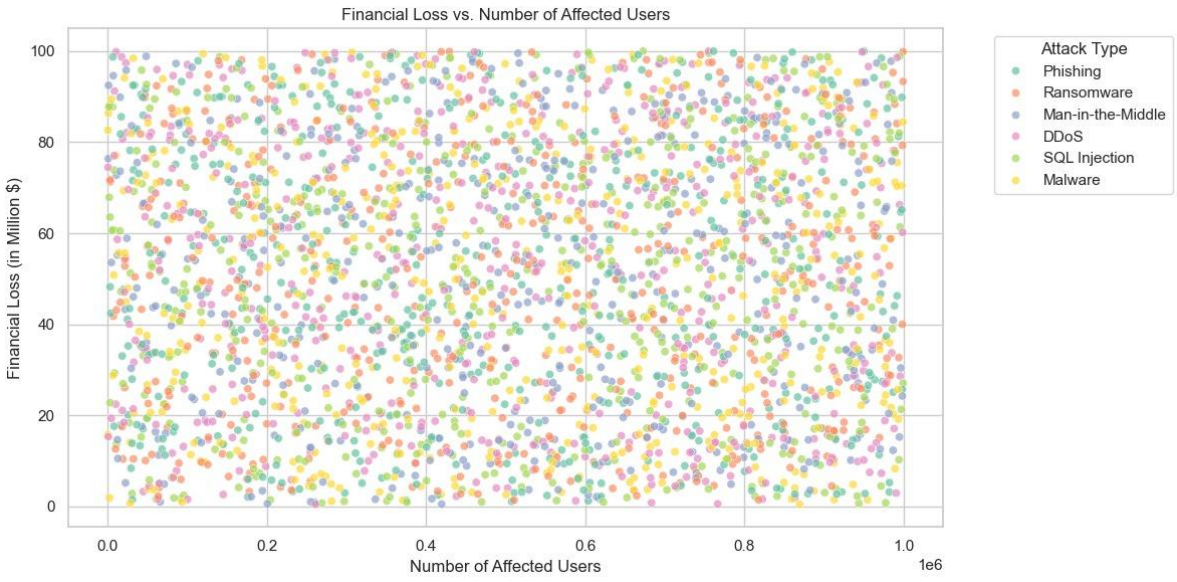
Correlation Heatmap Between Numerical Features



Box Plots of Financial Loss, Affected Users, and Resolution Time



Impact Analysis of Cyber Attacks: Financial Loss vs. Affected Users



Python Code for Data Analysis (Cybersecurity Threats Dataset)

Below is the Python script used to analyze the Global Cybersecurity Threats dataset (2015–2024). It includes data loading, inspection, and various visualizations using libraries like Pandas, Matplotlib, and Seaborn.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = r"C:\Users\mehta\Downloads\archive (2)\Global_Cybersecurity_Threats_2015-2024.csv"
df = pd.read_csv(file_path)

# Inspect the data
print(df.head())
print(df.info())

# Set Seaborn theme
sns.set(style="whitegrid")

# Top 10 countries by total financial loss
top_countries = df.groupby("Country")["Financial Loss (in Million $)"].sum().sort_values(ascending=False).head(10)

plt.figure(figsize=(12, 6))
sns.barplot(x=top_countries.values, y=top_countries.index, palette="Reds_r", hue=None, legend=False)
plt.title("Top 10 Countries by Total Financial Loss (in Million $)")
plt.xlabel("Total Financial Loss (Million $)")
plt.ylabel("Country")
plt.tight_layout()
plt.show()

# Number of incidents per year
attacks_per_year = df["Year"].value_counts().sort_index()

plt.figure(figsize=(12, 6))
sns.lineplot(x=attacks_per_year.index, y=attacks_per_year.values, marker="o", color="blue")
plt.title("Number of Cybersecurity Incidents Over the Years")
```

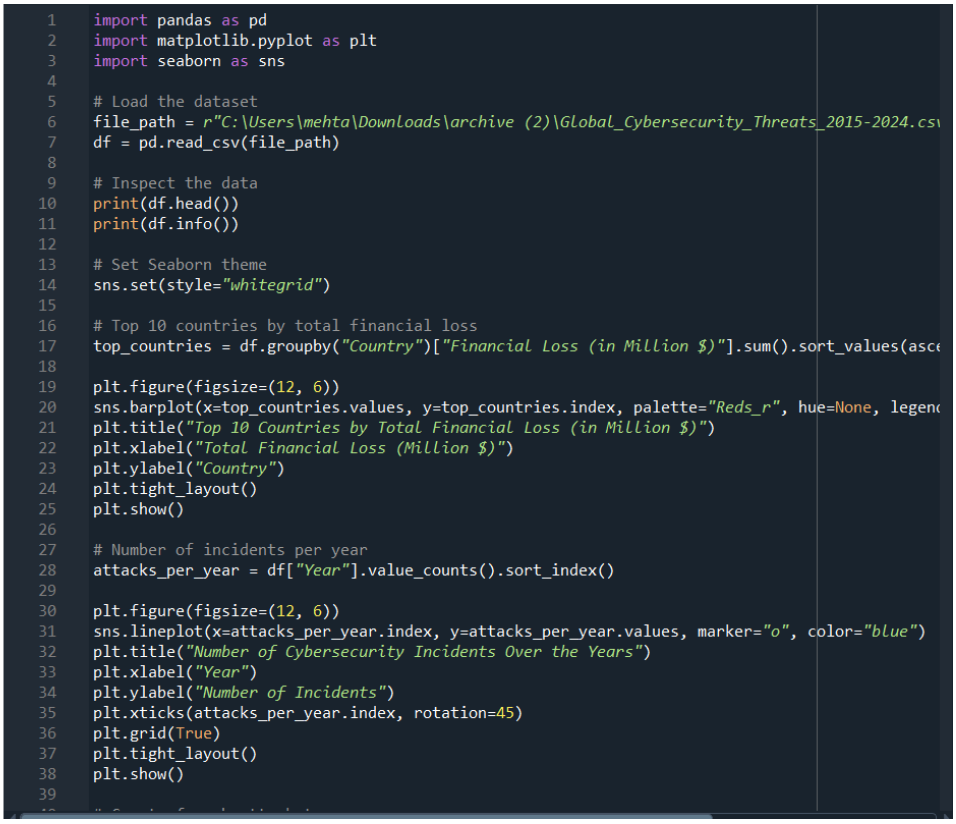
```
plt.xlabel("Year")
plt.ylabel("Number of Incidents")
plt.xticks(attacks_per_year.index, rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()

# Count of each attack type
attack_type_counts = df["Attack Type"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=attack_type_counts.values, y=attack_type_counts.index, palette="coolwarm",
hue=None, legend=False)
plt.title("Most Common Types of Cyberattacks (2015–2024)")
plt.xlabel("Number of Incidents")
plt.ylabel("Attack Type")
plt.tight_layout()
plt.show()

# Targeted industries count
industry_counts = df["Target Industry"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=industry_counts.values, y=industry_counts.index, palette="viridis",
hue=None, legend=False)
plt.title("Most Targeted Industries by Cyberattacks")
plt.xlabel("Number of Incidents")
plt.ylabel("Target Industry")
plt.tight_layout()
plt.show()

# Frequency of vulnerability types
vuln_counts = df["Security Vulnerability Type"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=vuln_counts.values, y=vuln_counts.index, palette="magma", hue=None,
legend=False)
plt.title("Most Common Security Vulnerabilities")
plt.xlabel("Number of Incidents")
plt.ylabel("Vulnerability Type")
plt.tight_layout()
plt.show()
```

```
# Average resolution time by attack type
avg_resolution_time = df.groupby("Attack Type")["Incident Resolution Time (in
Hours)"].mean().sort_values(ascending=False)

plt.figure(figsize=(12, 6))
sns.barplot(x=avg_resolution_time.values, y=avg_resolution_time.index, palette="cividis",
hue=None, legend=False)
plt.title("Average Incident Resolution Time by Attack Type")
plt.xlabel("Avg. Resolution Time (Hours)")
plt.ylabel("Attack Type")
plt.tight_layout()
plt.show()

# Heatmap: Correlation between numerical features
plt.figure(figsize=(10, 6))
numerical_features = df[[
   "Financial Loss (in Million $)",
   "Number of Affected Users",
   "Incident Resolution Time (in Hours)"
]]
corr = numerical_features.corr()

sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap Between Numerical Features")
plt.tight_layout()
plt.show()

# Box plots for numerical distributions
plt.figure(figsize=(18, 5))

# Financial Loss
plt.subplot(1, 3, 1)
sns.boxplot(y=df["Financial Loss (in Million $)"], color="skyblue")
plt.title("Box Plot - Financial Loss (in Million $)")

# Number of Affected Users
plt.subplot(1, 3, 2)
sns.boxplot(y=df["Number of Affected Users"], color="salmon")
plt.title("Box Plot - Number of Affected Users")

# Incident Resolution Time
plt.subplot(1, 3, 3)
sns.boxplot(y=df["Incident Resolution Time (in Hours)"], color="lightgreen")
```

plt.title("Box Plot - Incident Resolution Time (in Hours)")

plt.tight_layout()
plt.show()

Screenshots of the Code:



```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = r"C:\Users\mehta\Downloads\archive (2)\Global_Cybersecurity_Threats_2015-2024.csv
df = pd.read_csv(file_path)

# Inspect the data
print(df.head())
print(df.info())

# Set Seaborn theme
sns.set(style="whitegrid")

# Top 10 countries by total financial loss
top_countries = df.groupby("Country")["Financial Loss (in Million $)"].sum().sort_values(asce

plt.figure(figsize=(12, 6))
sns.barplot(x=top_countries.values, y=top_countries.index, palette="Reds_r", hue=None, legend
plt.title("Top 10 Countries by Total Financial Loss (in Million $)")
plt.xlabel("Total Financial Loss (Million $)")
plt.ylabel("Country")
plt.tight_layout()
plt.show()

# Number of incidents per year
attacks_per_year = df["Year"].value_counts().sort_index()

plt.figure(figsize=(12, 6))
sns.lineplot(x=attacks_per_year.index, y=attacks_per_year.values, marker="o", color="blue")
plt.title("Number of Cybersecurity Incidents Over the Years")
plt.xlabel("Year")
plt.ylabel("Number of Incidents")
plt.xticks(attacks_per_year.index, rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
plt.show()

# Count of each attack type
attack_type_counts = df["Attack Type"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=attack_type_counts.values, y=attack_type_counts.index, palette="coolwarm", hue=
plt.title("Most Common Types of Cyberattacks (2015-2024)")
plt.xlabel("Number of Incidents")
plt.ylabel("Attack Type")
plt.tight_layout()
plt.show()

# Targeted industries count
industry_counts = df["Target Industry"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=industry_counts.values, y=industry_counts.index, palette="viridis", hue=None, 
plt.title("Most Targeted Industries by Cyberattacks")
plt.xlabel("Number of Incidents")
plt.ylabel("Target Industry")
plt.tight_layout()
plt.show()

# Frequency of vulnerability types
vuln_counts = df["Security Vulnerability Type"].value_counts()

plt.figure(figsize=(12, 6))
sns.barplot(x=vuln_counts.values, y=vuln_counts.index, palette="magma", hue=None, legend=Fals
plt.title("Most Common Security Vulnerabilities")
plt.xlabel("Number of Incidents")
plt.ylabel("Vulnerability Type")
plt.tight_layout()
plt.show()

# Average resolution time by attack type
avg_resolution_time = df.groupby("Attack Type")["Incident Resolution Time (in Hours)"].mean(
```

```
avg_resolution_time = df.groupby("Attack Type")["Incident Resolution Time (in Hours)"].mean(

plt.figure(figsize=(12, 6))
sns.barplot(x=avg_resolution_time.values, y=avg_resolution_time.index, palette="cividis", hue
plt.title("Average Incident Resolution Time by Attack Type")
plt.xlabel("Avg. Resolution Time (Hours)")
plt.ylabel("Attack Type")
plt.tight_layout()
plt.show()

# Heatmap: Correlation between numerical features
plt.figure(figsize=(10, 6))
numerical_features = df[[
    "Financial Loss (in Million $)",
    "Number of Affected Users",
    "Incident Resolution Time (in Hours)"
]]
corr = numerical_features.corr()

sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap Between Numerical Features")
plt.tight_layout()
plt.show()

# Box plots for numerical distributions
plt.figure(figsize=(18, 5))

# Financial Loss
plt.subplot(1, 3, 1)
sns.boxplot(y=df["Financial Loss (in Million $)"], color="skyblue")
plt.title("Box Plot - Financial Loss (in Million $)")

# Number of Affected Users
plt.subplot(1, 3, 2)
sns.boxplot(y=df["Number of Affected Users"], color="salmon")
plt.title("Box Plot - Number of Affected Users")

# Incident Resolution Time
plt.subplot(1, 3, 3)
```

```
plt.ylabel("Attack Type")
plt.tight_layout()
plt.show()

# Heatmap: Correlation between numerical features
plt.figure(figsize=(10, 6))
numerical_features = df[[
    "Financial Loss (in Million $)",
    "Number of Affected Users",
    "Incident Resolution Time (in Hours)"
]]
corr = numerical_features.corr()

sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap Between Numerical Features")
plt.tight_layout()
plt.show()

# Box plots for numerical distributions
plt.figure(figsize=(18, 5))

# Financial Loss
plt.subplot(1, 3, 1)
sns.boxplot(y=df["Financial Loss (in Million $)"], color="skyblue")
plt.title("Box Plot - Financial Loss (in Million $)")

# Number of Affected Users
plt.subplot(1, 3, 2)
sns.boxplot(y=df["Number of Affected Users"], color="salmon")
plt.title("Box Plot - Number of Affected Users")

# Incident Resolution Time
plt.subplot(1, 3, 3)
sns.boxplot(y=df["Incident Resolution Time (in Hours)"], color="lightgreen")
plt.title("Box Plot - Incident Resolution Time (in Hours)")

plt.tight_layout()
plt.show()
```

Conclusion

This project explored the trends and patterns in global cybersecurity threats from 2015 to 2024 using comprehensive data analysis and visualization techniques. Our findings indicate that:

- Phishing, Ransomware, and Malware are the most frequent cyberattack types globally.
- The financial impact of attacks correlates significantly with the number of affected users, as shown in the "Financial Loss vs. Number of Affected Users" analysis.
- Industries such as finance, healthcare, and government are among the most targeted sectors.
- The United States, India, and China reported the highest total financial losses due to cyberattacks.
- On average, DDoS and Ransomware attacks take the longest to resolve, reflecting the complexity and severity of these threats.
- There is a notable correlation between the number of users affected and the financial loss incurred, emphasizing the importance of early detection and mitigation.

The visual and statistical insights obtained through this project can aid organizations in identifying key risk areas and preparing appropriate cybersecurity strategies to mitigate financial and operational risks.

References

1. Prasad, S. (2024). Global Cybersecurity Threats (2015–2024) [Dataset]. Kaggle. https://www.kaggle.com/datasets/sprasad018/global-cybersecurity-threats-2015-2024
2. McKinsey & Company. (2023). Cybersecurity in a Changing World. Retrieved from https://www.mckinsey.com
3. IBM Security. (2023). Cost of a Data Breach Report. Retrieved from https://www.ibm.com/security/data-breach
4. OWASP Foundation. (2024). Top 10 Web Application Security Risks. Retrieved from https://owasp.org
5. Python Libraries: Pandas, Matplotlib, Seaborn – for data analysis and visualization