Project Report on

Cyber Security Attacks

Submitted by:

Aanchal Kumari

&

Chanchal

In partial fulfillment of completion of the course

Advanced Diploma in IT, Networking and Cloud Computing.

Under Guidance of:

Year 2022-2023

Abstract

The primary objectives of this research include the collection and analysis of diverse datasets encompassing various cyber-attacks types, such as malware infections, phishing attempts, ransom ware incidents, and other forms of malicious activities. Advanced data analytics and machine learning techniques will be applied to identify patterns and trends within these datasets, allowing for a nuanced understanding of attack vectors and evolving threat landscapes.

The project aims to categorize cyber threats based on their characteristics, methodologies, and impacts. By examining historical attack data, we intend to identify recurring patterns and emerging trends, providing insights into the evolving tactics employed by cyber adversaries.

Acknowledgement

At this juncture of our journey, we wish to express our heartfelt gratitude to all those who have contributed to the creation and success of **"Cyber Security Attacks".** This project has been a labor of passion and dedication, and it would not have been possible without the unwavering support and guidance we have received.

First and foremost, we offer our thanks to the boundless creativity and inspiration that flows from the universe. We are grateful for the opportunity to embark on this venture.

We extend our sincerest appreciation to our mentors, **Mrs. Mala Mishra & Ms. Ankita Shukla**, whose wisdom and guidance have been instrumental in shaping the vision of **"Cyber Security Attacks".** Your support at every crucial turn has illuminated our path and fueled our determination to create a meaningful platform.

To our dedicated team of developers, designers, and content creators, we extend our deepest gratitude. Your tireless efforts, innovation, and creativity have breathed life into **"Cyber Security Attacks".** It is your collective dedication that has made this project a reality.

Our appreciation also goes to our colleagues and friends who provided invaluable insights and feedback during the development process. Your input has been instrumental in refining our ideas and enhancing the user experience.

We acknowledge the contributions of the broader IT community, whose open-source ethos has been a wellspring of knowledge and inspiration. The collaborative spirit of this community has been a guiding light.

Last but not least, we owe a debt of gratitude to our families and friends who have stood by us throughout this journey. Your unwavering support, encouragement, and belief in our vision have been our constant motivation.

ADVANCE DIPLOMA IN IT NETWORKING & CLOUD COMPUTING

The Advanced Diploma in IT Networking and Cloud Computing program offered by NSTI (W) Noida in collaboration with Edunet Foundation is a comprehensive course designed to equip students with advanced skills in information technology and cloud computing. This program covers a wide range of topics, including Computer Networking, Database Management, Virtualization, Cloud Technologies, and Cybersecurity. Students will gain hands-on experience through practical labs, workshops, and real-world projects, enabling them to excel in the rapidly evolving IT industry. Upon completion of the program, Graduates will have a strong foundation in both IT Fundamentals and Cloud Computing, making them highly sought-after professionals in the field.

Project Requirements

Project Name	**Cyber Security Attacks**
Languages Used	**Python**
Editor	**Jupyter Notebook, Google Colab**
Web Browser	**Google Chrome, Microsoft Edge**

Team Composition and Workload Division

Aanchal Kumari	Data Analysis, Synopsis
Chanchal	Data Analysis, Synopsis

Tables of Content

1. Introduction to Problem

The increasing frequency, complexity, and severity of cybersecurity attacks present a critical challenge to individuals, organizations, and nations worldwide.

Cyber threats, including malware infections, phishing attacks, ransomware incidents, and other malicious activities, exploit vulnerabilities in digital systems, leading to data breaches, financial losses, and compromised critical infrastructure. The evolving nature of cyber threats poses a significant problem, necessitating a comprehensive understanding and effective mitigation strategies.

2. Requirements

3.1 Technology Stack

Python: High-level programming language used for server-side scripting.

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text, providing an interactive and collaborative environment for data science and analysis.

3.2 Hardware

Laptop/ Computer

3.3 Software

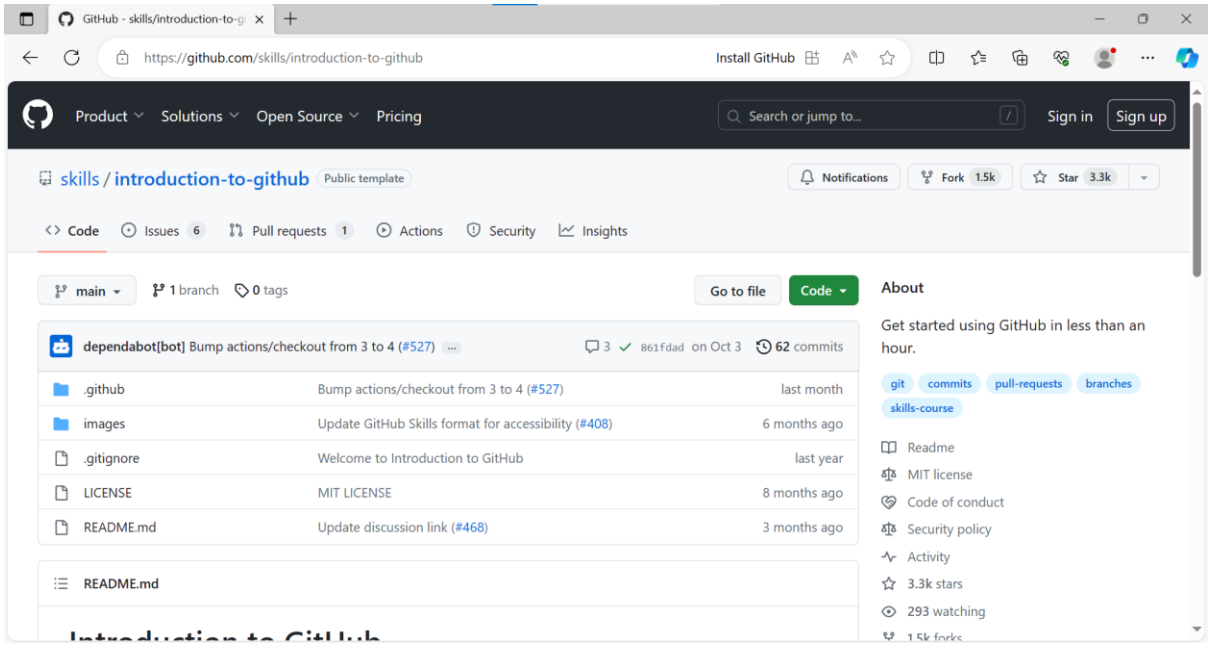
Operating System (OS)

Version Control System

Text Editors and Integrated Development Environments (IDEs)

3.4 Deployment Environment

Github



3. Overview

The data analysis project aims to investigate and derive meaningful insights from a specific dataset. It involves collecting, cleaning, and processing raw data to uncover patterns, trends, and correlations. Using statistical methods and visualization tools, the project seeks to provide a comprehensive understanding of the data, enabling informed decision-making. The analysis may involve exploring relationships between variables, identifying outliers, and creating predictive models. Throughout the project, a systematic approach is followed, including hypothesis testing and validation of results. The ultimate goal is to offer actionable recommendations or conclusions based on the data findings. The project typically employs programming languages such as Python or R, along with tools like Jupyter Notebooks, to facilitate a transparent and reproducible analytical workflow. Overall, the data analysis project serves to extract valuable insights, enhance understanding, and support evidence-based decision-making in a given domain.

4. Project Module

1. Import the required libraries.
2. Load/ Read the Dataset
3. Do Visualizations
4. Effect of different gases on different states
5. Prepare Heatmap/ Confusion Matrix

6. Prepare Profile Report

6. Sample Screenshots

Cyber Attack Data Exploration

```
[196] #importing all the necessary Python libraries and the dataset
      import warnings
      warnings.filterwarnings('ignore')
      import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
      %matplotlib inline
      import seaborn as sns

      data=pd.read_csv("cybersecurity_attacks.csv")
```

```
import warnings
warnings.filterwarnings('ignore')
```

```
data
```

	Timestamp	Source IP Address	Destination IP Address	Source Port	Destination Port	Protocol	Packet Length	Packet Type	Traffic Type	Payload Data	...	Action Taken	Severity Level	User Information	Device Information	Network Segment	Geo-location Data
0	2023-05-30 06:33:58	103.216.15.12	84.9.164.252	31225	17616	ICMP	503	Data	HTTP	Qui natus odio asperiores nam. Optio nobis ius...	...	Logged	Low	Reyansh Dugal	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT ...	Segment A	Jamshedpur, Sikkim
1	2020-08-26 07:08:30	78.199.217.198	66.191.137.154	17245	48166	ICMP	1174	Data	HTTP	Aperiam quos modi officiis veritatis rem. Omni...	...	Blocked	Low	Sumer Rana	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT ...	Segment B	Bilaspur, Nagaland
2	2022-11-13 08:23:25	63.79.210.48	198.219.82.17	16811	53600	UDP	306	Control	HTTP	Perferendis sapiente vitae soluta. Hic delectu...	...	Ignored	Low	Himmat Karpe	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Segment C	Bokaro, Rajasthan
3	2023-07-02 10:38:46	163.42.196.10	101.228.192.255	20018	32534	UDP	385	Data	HTTP	Totam maxime beatae expedita explicabo porro l...	...	Blocked	Medium	Fateh Kibe	Mozilla/5.0 (Macintosh; PPC Mac OS X 10_11_5; ...	Segment B	Jaunpur, Rajasthan
4	2023-07-16 13:11:07	71.166.185.76	189.243.174.238	6131	26646	TCP	1462	Data	DNS	Odit nesciunt dolorem nisi iste iusto. Animi	...	Blocked	Low	Dhanush Chad	Mozilla/5.0 (compatible; MSIE 5.0; Windows NT	Segment C	Anantapur, Tripura

```
# general information
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 138 entries, 2020-01-31 to 2023-10-31
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   Attack Type  138 non-null    object
 1   value        138 non-null    int64 
dtypes: int64(1), object(1)
memory usage: 3.2+ KB
```

Data Cleaning

```
# check for null values
data.isna().sum()
```

```
Timestamp                  0
Source IP Address          0
Destination IP Address     0
Source Port                0
Destination Port           0
Protocol                   0
Packet Length              0
Packet Type                0
Traffic Type               0
Payload Data               0
Malware Indicators      1176
Anomaly Scores             0
Alerts/Warnings         1187
Attack Type                0
Attack Signature           0
Action Taken               0
Severity Level             0
User Information           0
Device Information         0
Network Segment            1
Geo-location Data          1
Proxy Information       1148
Firewall Logs           1197
IDS/IPS Alerts          1163
Log Source                 1
dtype: int64
```

```
[67] # calculate the mean , std, min, max and count of every attributes
     data.describe()
```

	Source Port	Destination Port	Packet Length	Anomaly Scores
count	2348.000000	2348.000000	2348.000000	2348.000000
mean	32314.992760	32838.227428	788.548552	50.104647
std	18781.241745	18571.544069	413.974742	28.932539
min	1031.000000	1030.000000	64.000000	0.060000
25%	15965.500000	17096.250000	428.750000	24.657500
50%	31733.000000	32502.000000	786.000000	50.520000
75%	48357.000000	49076.500000	1143.250000	75.160000
max	65521.000000	65535.000000	1500.000000	99.990000

```
[68] # Check for duplicate values
     data.duplicated().sum()
```

```
0
```

```
# Checking Skewness from 'Source Port' to 'Anomaly Scores'

df=data.loc[:,'Source Port':'Anomaly Scores']
df=df.select_dtypes([np.int, np.float])
for i, col in enumerate(df.columns):
    print("\nSkewness of "+col +" is", df[col].skew()) #measures skewness
```

```
Skewness of Source Port is 0.05946941104053445

Skewness of Destination Port is 0.03858148110055104

Skewness of Packet Length is -0.017914267379731216

Skewness of Anomaly Scores is -0.02083080381904235
```

```
[70] # Check unique values
     data["Traffic Type"].unique()
```

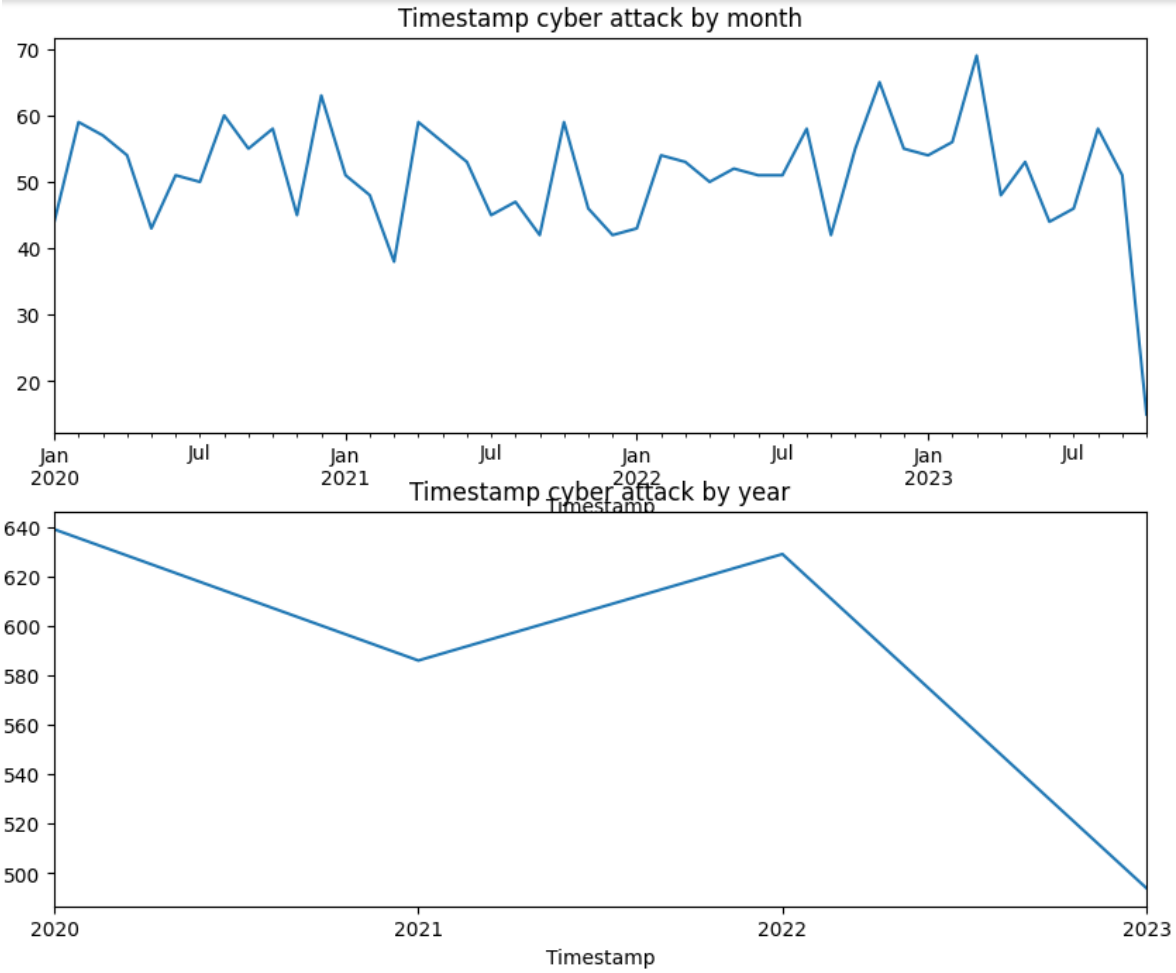
```
array(['HTTP', 'DNS', 'FTP'], dtype=object)
```

```
[71] data["Attack Type"].unique()
```

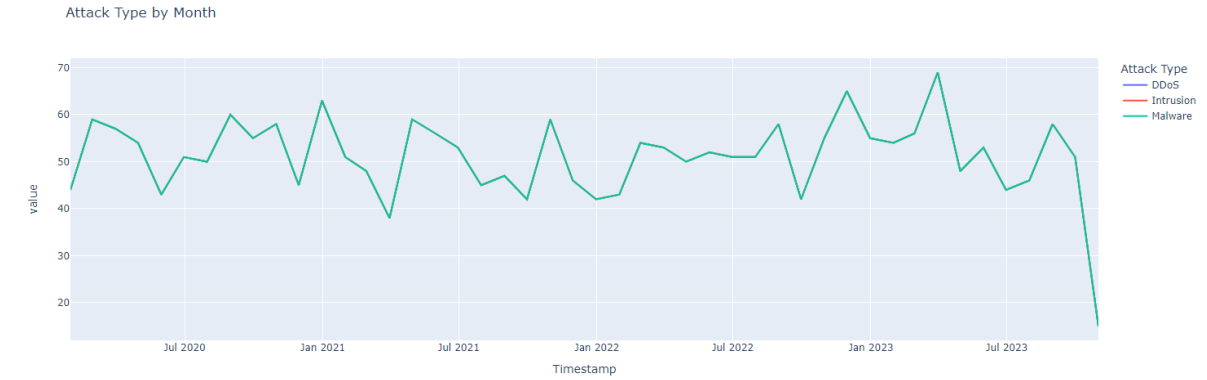
```
array(['Malware', 'DDoS', 'Intrusion'], dtype=object)
```

```
# Assuming 'data' is your DataFrame and 'Timestamp' is the column containing timestamps
data['Timestamp'] = pd.to_datetime(data['Timestamp'])

fig = plt.figure(figsize=(10, 8))
fig.add_subplot(211)
data.resample('M', on='Timestamp')['Attack Type'].count().plot(title='Timestamp cyber attack by month')


fig.add_subplot(212)
data.resample('Y', on='Timestamp')['Attack Type'].count().plot(title='Timestamp cyber attack by year')
```

```
<Axes: title={'center': 'Timestamp cyber attack by year'}, xlabel='Timestamp'>
```

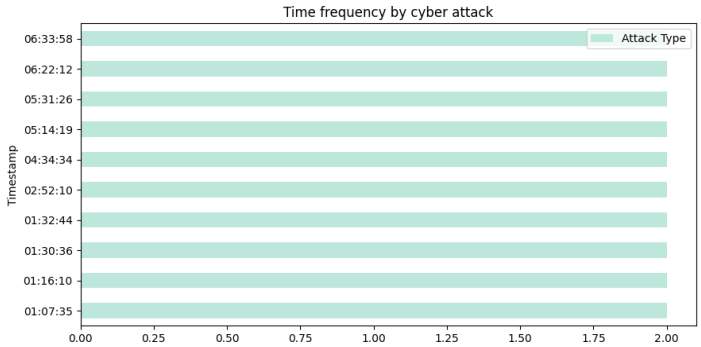


```
import plotly.express as px

data = pd.crosstab(data['Timestamp'], data['Attack Type']).resample('M').count().melt(ignore_index=False)
px.line(data, x=data.index, y='value', color='Attack Type', title='Attack Type by Month').show()
```



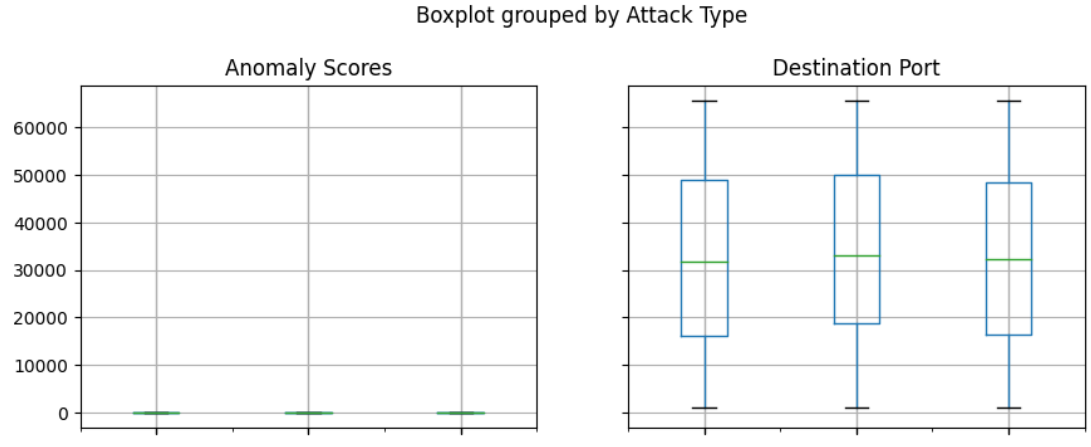

```
data['Timestamp'] = pd.to_datetime(data['Timestamp'])
data.groupby(data['Timestamp'].dt.time).agg({'Attack Type': 'count'}).nlargest(10, 'Attack Type').plot(kind='barh', figsize=(10, 5), colormap='icefire', title='Time frequency by cyber attack')
```

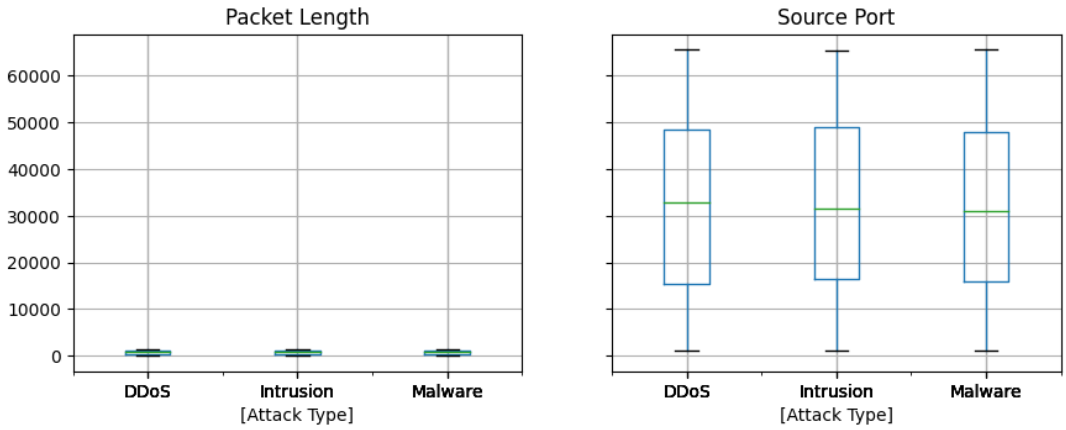
```
<Axes: title={'center': 'Time frequency by cyber attack'}, ylabel='Timestamp'>
```



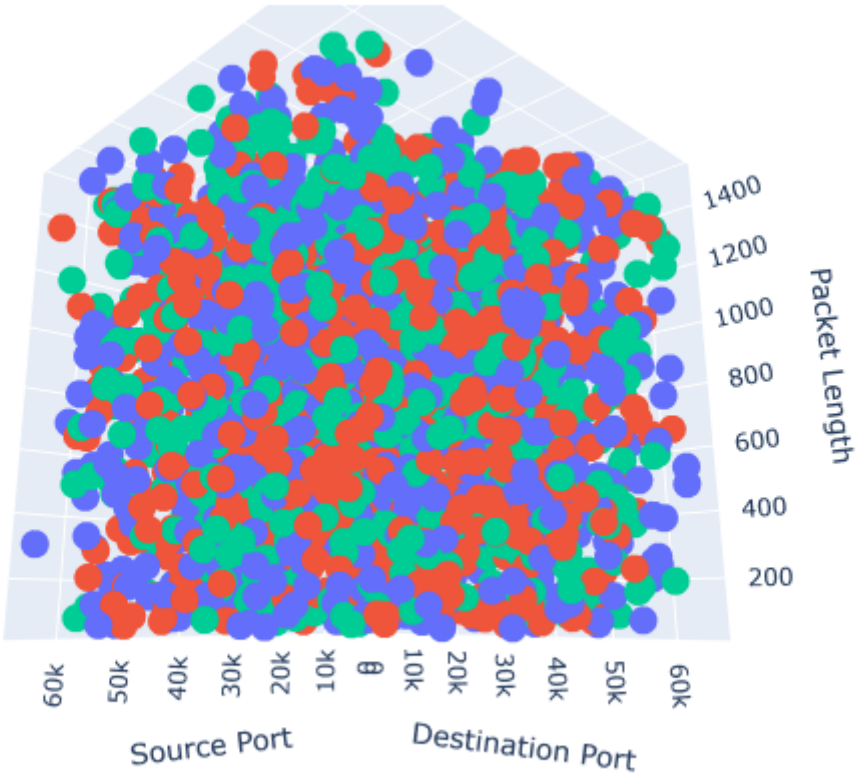

```
data.boxplot(figsize=(10,8), by='Attack Type')
```

```
array([[<Axes: title={'center': 'Anomaly Scores'}, xlabel='[Attack Type]'>,
        <Axes: title={'center': 'Destination Port'}, xlabel='[Attack Type]'>],
       [<Axes: title={'center': 'Packet Length'}, xlabel='[Attack Type]'>,
        <Axes: title={'center': 'Source Port'}, xlabel='[Attack Type]'>]],
      dtype=object)
```

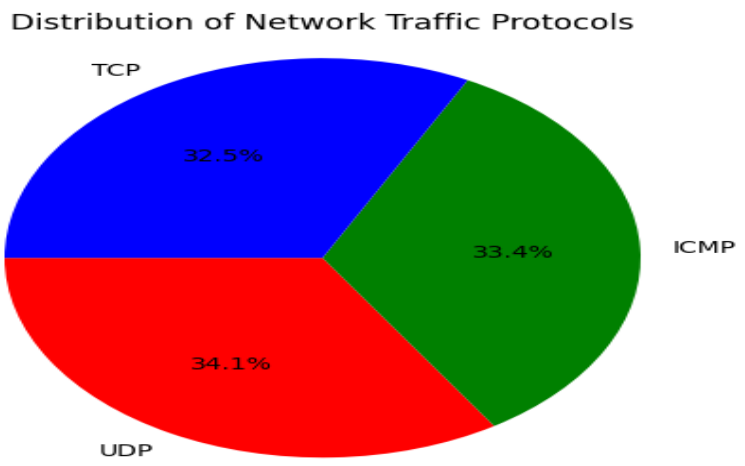




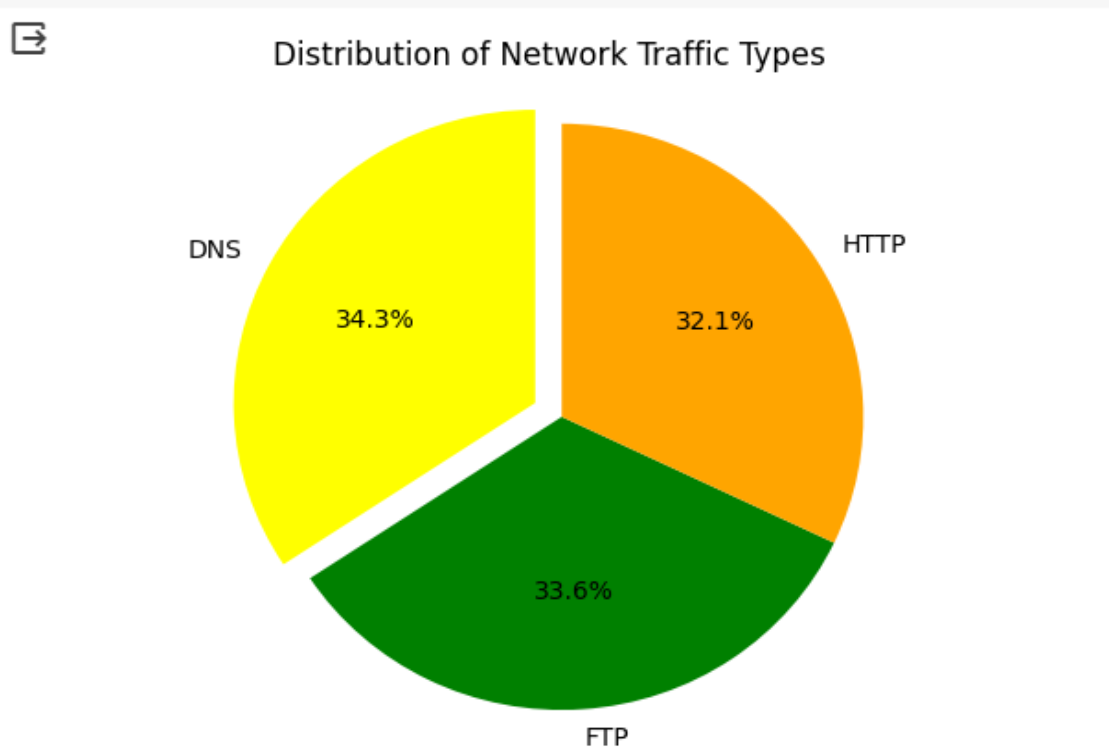

```
[215] px.scatter_3d(data, x='Source Port', y='Destination Port', z='Packet Length', color='Protocol').show()
```



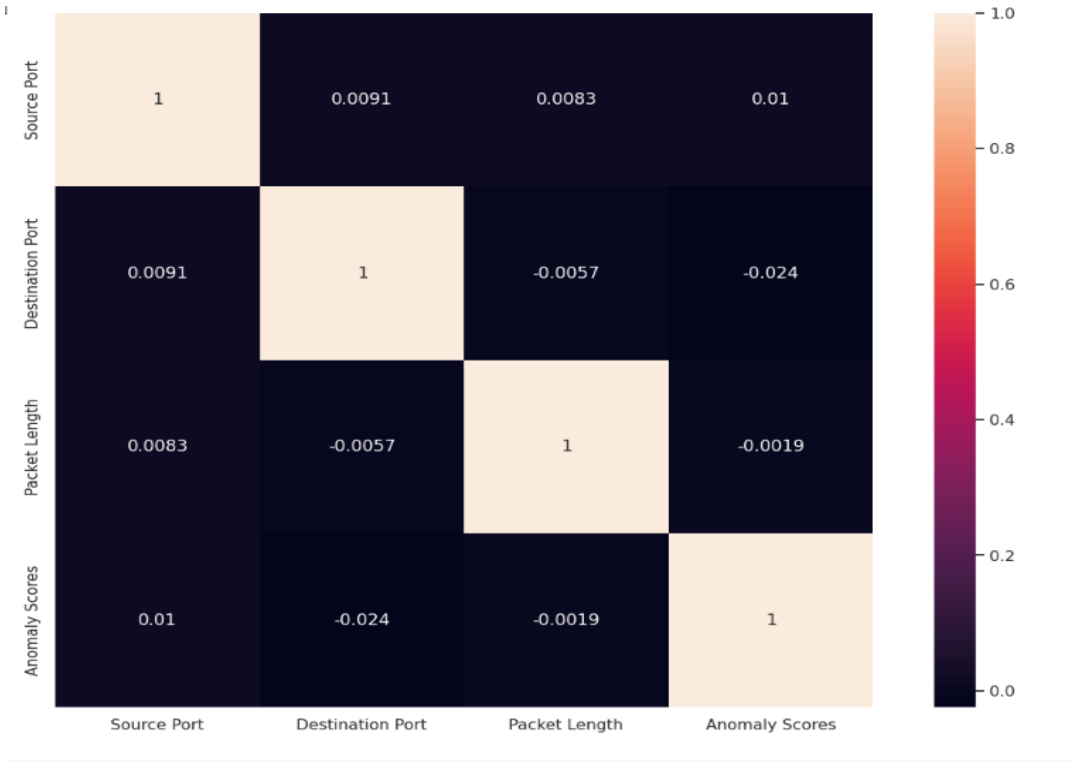

```
labels = ['UDP', 'ICMP', 'TCP']
sizes = data['Protocol'].value_counts()  # Proportional sizes of each category
colors = ['red', 'green', 'blue']  # Color for each category segment
explode = (0.1, 0, 0)  # Explode a slice if needed (0 means no explosion)


# Create a pie chart
plt.pie(sizes, labels=labels, colors=colors,  autopct='%1.1f%%', startangle=180)

plt.axis('equal')
plt.title('Distribution of Network Traffic Protocols')
plt.show()
```



```
[102] # Data for the pie chart

      labels =['DNS','FTP','HTTP']
      sizes = data['Traffic Type'].value_counts()
      colors = ['yellow', 'green', 'orange']
      explode = (0.1, 0, 0)

      # Create a pie chart
      plt.pie(sizes, labels=labels, colors=colors, explode=explode, autopct='%2.1f%%', startangle=90)

      plt.axis('equal')
      plt.title('Distribution of Network Traffic Types')

      plt.show()
```




```
[233] import warnings
      warnings.filterwarnings('ignore')

      #Correlation matrix
      corrmat = data.corr()
      f, ax = plt.subplots(figsize = (15, 10))
      sns.heatmap(corrmat, vmax = 1, square = True, annot = True)
```



8 Future Scope

The future scope for a cyber security attacks data analysis project is vast, given the evolving nature of cyber threats and the increasing reliance on digital technologies. Here are several areas where such a project could have significant impact:

Implementing advanced machine learning algorithms and artificial intelligence to enhance the detection of patterns and anomalies in cyber attack data.

Analyzing user and network behavior to identify abnormal activities that may indicate a cyber attack.

9 Conclusion

In conclusion, this cyber security project aims to provide a comprehensive understanding of the current threat landscape, empowering organizations and individuals with the knowledge needed to safeguard their digital assets. By combining data analysis, threat intelligence, and proactive cybersecurity measures, the project seeks to contribute to the ongoing efforts to create a secure and resilient cyberspace.

10 References

https://www.kaggle.com/datasets