أكاديمية سدايا
SDAIA Academy

Project Proposal:

Cybersecurity Attacks Data Analysis

Exploratory Data Analysis (EDA) Of Cybersecurity Attacks Dataset

SDAIA ACADEMY (T5)

By Shaikha Bin Ateeq

أكاديمية سدايا
SDAIA Academy

Introduction :

Organizations are becoming more vulnerable to cyber threats because of the growing global reliance on computers, networks, programs, social media, and data Data breaches, a typical cyber attack, have a huge negative impact on businesses and are frequently caused by inadequately safeguarded data.For information security, it's no longer enough to rely on traditional IT personnel and security protocols. Threat intelligence tools and security programs are clearly needed to lower your organization's cyber risk and identify potential attack surfaces. I intend to work with data on cybersecurity attacks that I obtained from Kaggle, which recorded the events of various cybersecurity attacks of varying durations and patterns.[1]

In this project, I will be using Exploratory Data Analysis approach as known as EDA, which aims to analyze and investigate data sets and summarize their main characteristics mostly by employing data visualization methods. [2] However I am planning to understand the the patterns of Cybersecurity attacks and the Most Logical Ports attacked and the common type of attacks in additions the Different time of the day , (hours,night,morning) and so on The goal of these analyzes is to know the origin and type of attacks and to try to raise the level of protection in those vulnerabilities

Data Overview :

The data will help us understand the nature of Cybersecurity attacks and raising the level of protection in the port where cyber-attacks often occur

The features provided in the dataset are:

- ◊ Attack category : Type of registered cybersecurity attack
- ◊ Attack subcategory: A subcategory of the type of cybersecurity attack registered
- ◊ Attack Name: The technical name for the cybersecurity attack
- ◊ Time: Start and end date of the attack in timestamp format
- ◊ Protocol: The protocol used for the attack.
- ◊ Source IP: IPv4 address where the attack came from.
- ◊ Source Port: The logical port where the attack came from.
- ◊ Destination IP: Destination IPv4 address.
- ◊ Destination Port: Logical destination port

Tools:

To explore and analyze the data, I will be using Jupyter notebook to use python language and Python libraries, such as:

- Matplotlib and Seaborn for data visualization.
- Numpy and Panda for data read and write operations

References :

[1] What is Cybersecurity Risk? A Thorough Definition | UpGuard. (2021). Retrieved 13 November 2021, from https://www.upguard.com/blog/cybersecurity-risk
[2] (2021). Retrieved 13 November 2021, from http://web.mta.info/developers/resources/nyct/turnstile/ts_Field_Description.txt