Dataset

World Cyber Incidents 2005 to 2020
(Source: https://www.kaggle.com/fireballbyedimyrnmom/2020-cyber-incidents)

Description

This dataset consisted records of major cyber-attack incidents around the world from 2005 to 2020.

Each record holds the types & descriptions of the attack, the name of the attackers & the country that they are from, the victims and the economy sectors that been targeted.

Total 481 rows of incidents had been recorded and 12 columns of attributes.

7 out of all 12 columns have been identified important from this dataset:

Col. No.	Attributes	Descriptions
1	Date	The date of the cyber attack recorded *e.g. "7/20/2018"*
2	Affiliations	The potential attacker (group/organization) that was responsible for the attack *e.g. "Believed to be the work of Whitefly"*
3	Description	A summary of the impact of the cyber-attack *e.g. "Personal data of over 1.5 million patients was compromised, including the pharmaceutical prescriptions of 160,000 people."*
4	Victims	The victim of the attack (country, firm, organization etc) *e.g. "Singhealth"*
5	Sponsor	The potential country that host the attacker
6	Type	The objective type of the cyber-attack *e.g. "Espionage", "Financial Theft", "Sabotage"*
7	Category	The economy sector that been targeted *e.g. "Government", "Private", Civil Society"*

5 columns that identified less important:

Col. No.	Attributes	Descriptions
8	Title	The title of the attacks *e.g. "Compromise of SingHealth, a large health-care provider in Singapore"*
9	Response	The victim's feedback to the attack. Most of them are empty

		e.g. “Criminal charges”, “Denouncement”
11, 12, 13	Source 1,2 & 3	The reference link/sources to the record

First 5 records from the dataset:

	Title	Date	Affiliations	Description	Response	Victims	Sponsor	Type	Category
0	Attack on Austrian foreign ministry	2/13/2020	Turla	The suspected Russian hackers conducted a week...	Confirmation https://www.theregister.co.uk/2...	Austrian Foreign Ministry	Russian Federation	Espionage	Government
1	Spear-phishing campaign against unnamed U.S. g...	1/23/2020	Konni Group	The suspected North Korean threat actor Konni ...	NaN	Employees of the U.S. government	Korea (Democratic People's Republic of)	Espionage	Government
2	Australian Signals Directorate	4/6/2020	NaN	Responsible for attacking infrastructure that ...	NaN	NaN	Australia	Data destruction	Private sector
3	Catfishing of Israeli soldiers	2/16/2020	APT-C-23	The Hamas-associated threat actor APT-C-23 tar...	Hack Back https://www.bleepingcomputer.com/n...	Israeli Defense Forces (IDF) soldiers	Palestine, State of	Espionage	Military
4	Targeting of U.S. companies and government age...	8/10/2020	Fox Kitten	Iranian hackers attacked high-end networking e...	NaN	U.S. government agencies, U.S. companies	Iran (Islamic Republic of)	Espionage	Government, Private sector

Challenges anticipated:

- To fill up empty records. Might be able to extract it out from “Title” or “Description” column
- Extracting useful insight from “Description” column that might not be found in other columns. Using Word Cloud.
- To extract information (e.g. country) from the “victims” column, as not all record was in country name.
- Some records in the “Category” columns consist of multiple sectors in an entry. How can I extract it out individually?

Goals

Problem Statement:

What is the most affected area to cyber attack in the world for the past 15years?

Questions:

1. The overall trend of the cyber-attacks around the world from 2005 to 2020?
2. The country and the sector where the most hit of cyber-attacks?
3. The trend/size of each cyber-attack type?
4. Relationship of the cyber-attack types to each targeted sectors?
5. The attackers/sponsors that responsible to most of the cyber-attack incidents around the world?
6. The frequency of the top 5 attackers/sponsor from 2005 to 2020?
7. Which period in a year have the highest rate of attack?
8. The relationship of the attackers and their attack type?
9. The forecast of future cyber-attack direction?
10. The recommendation of any vulnerable area to look into?