DS102 DS104 :
Final Project Proposal

World Cyber Incidents (2005 - 2020)

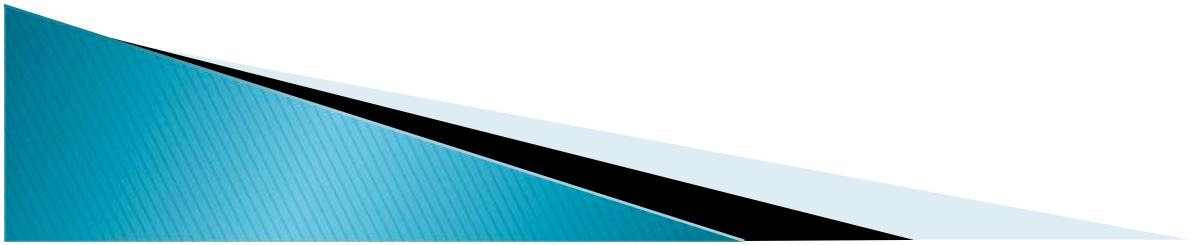
(Source: https://www.kaggle.com/fireballbyedimyrnmom/2020-cyber-incidents)

GitHub link to project: https://github.com/jiayang243/CyberAttacks

By: Jia Yang

Problem Statement

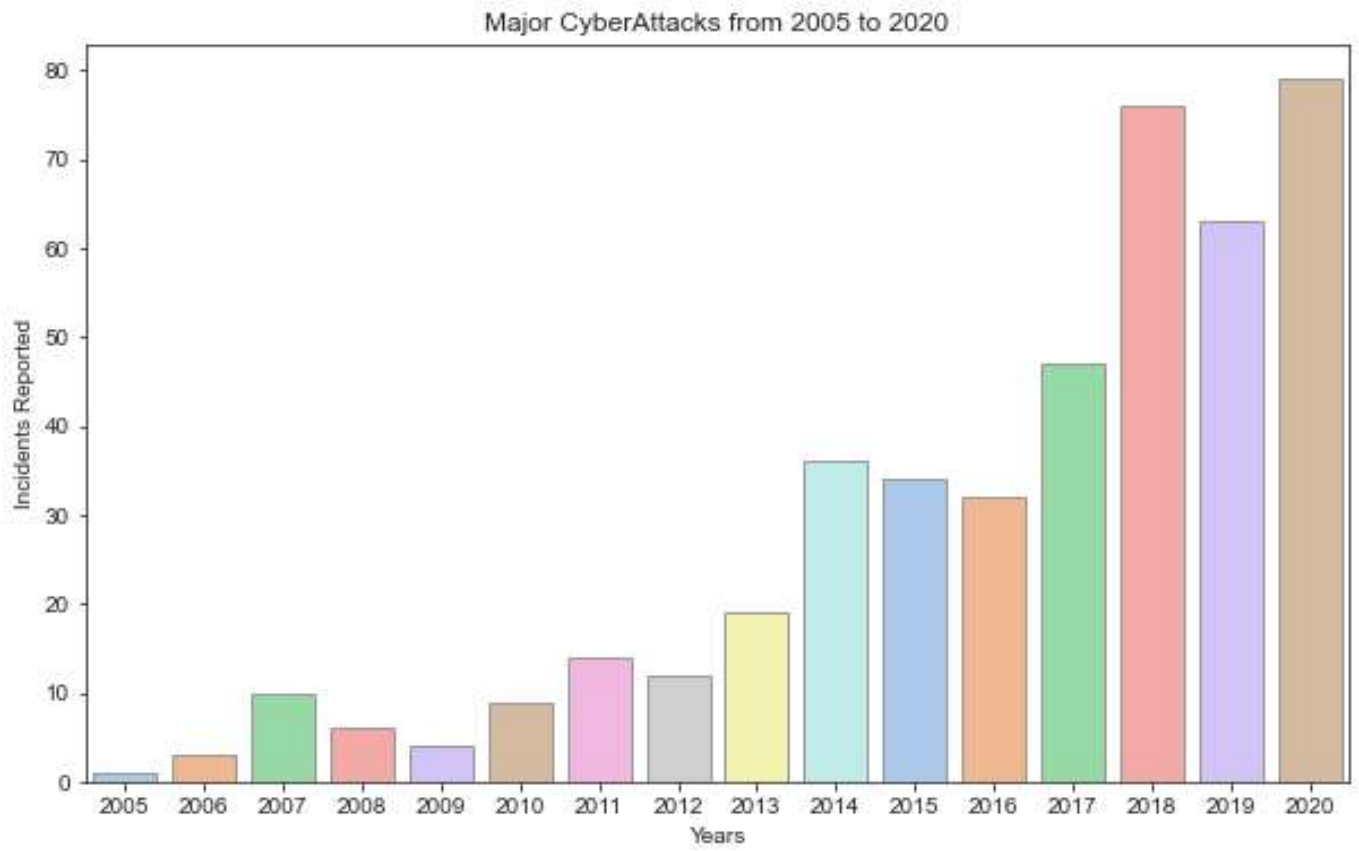
What is the most affected area by cyber-attack in the world for the past 15years?



Major Cyber-Attack Trend Around the World (2005 to 2020)

Total 445 incidents reported for the past 15 years.

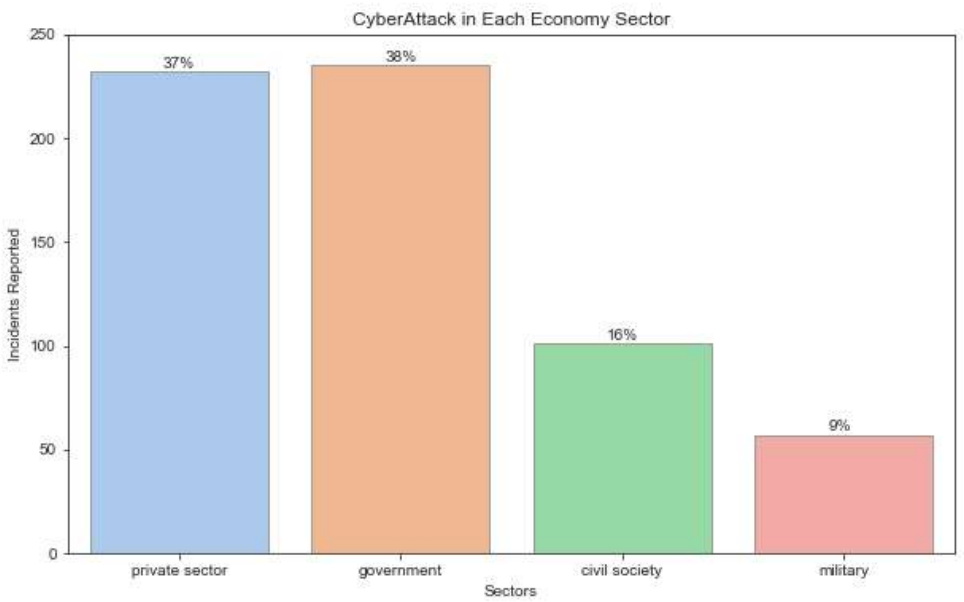
There was a increase of 80% in Cyber-Attacks since 2005.

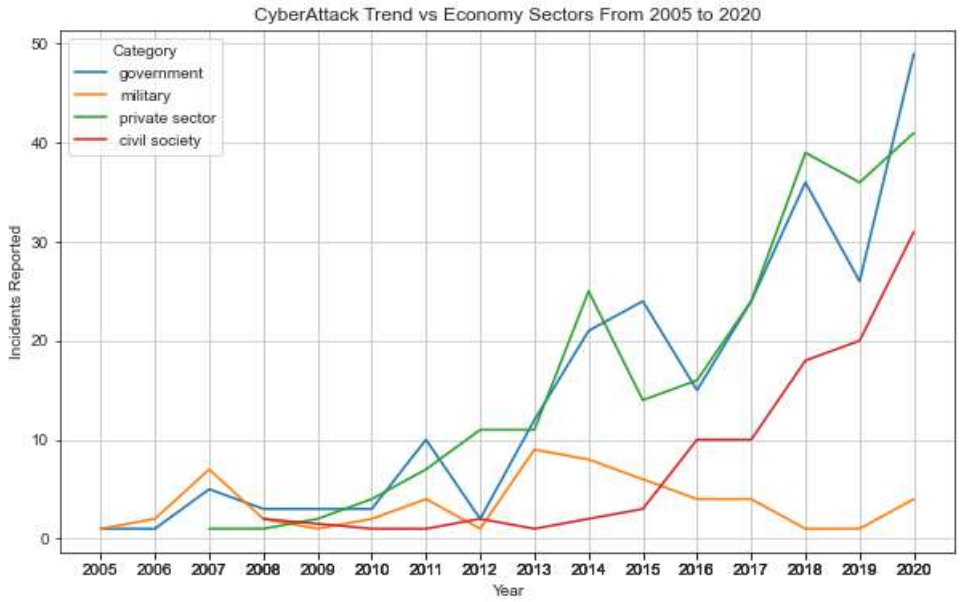


Cyber-Attack in Economy Sector

Majority attacks(75%) targeted on Private & Government sectors.

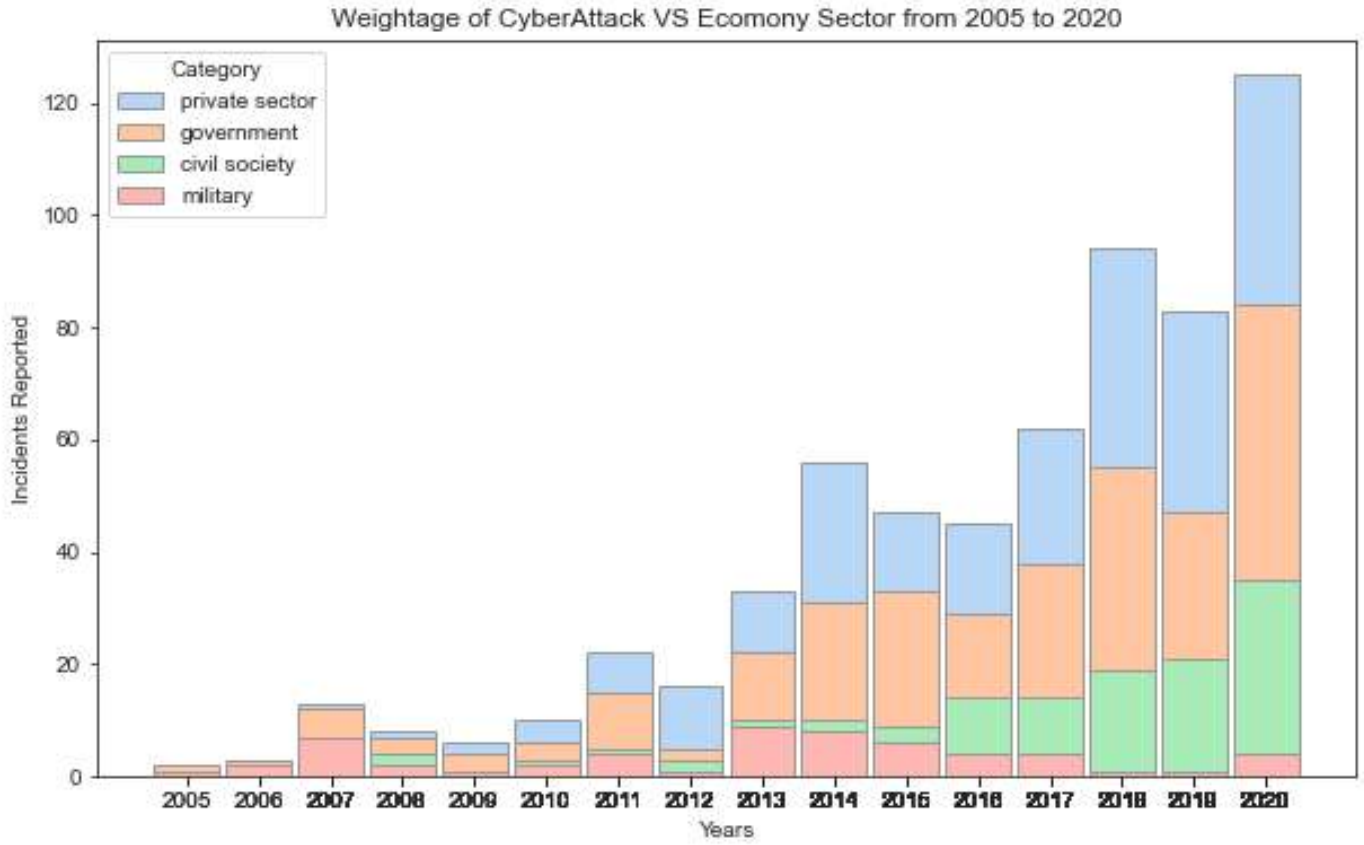
Can see that global growth of Cyber-Attack was also due to attacks to Private & Government Sectors





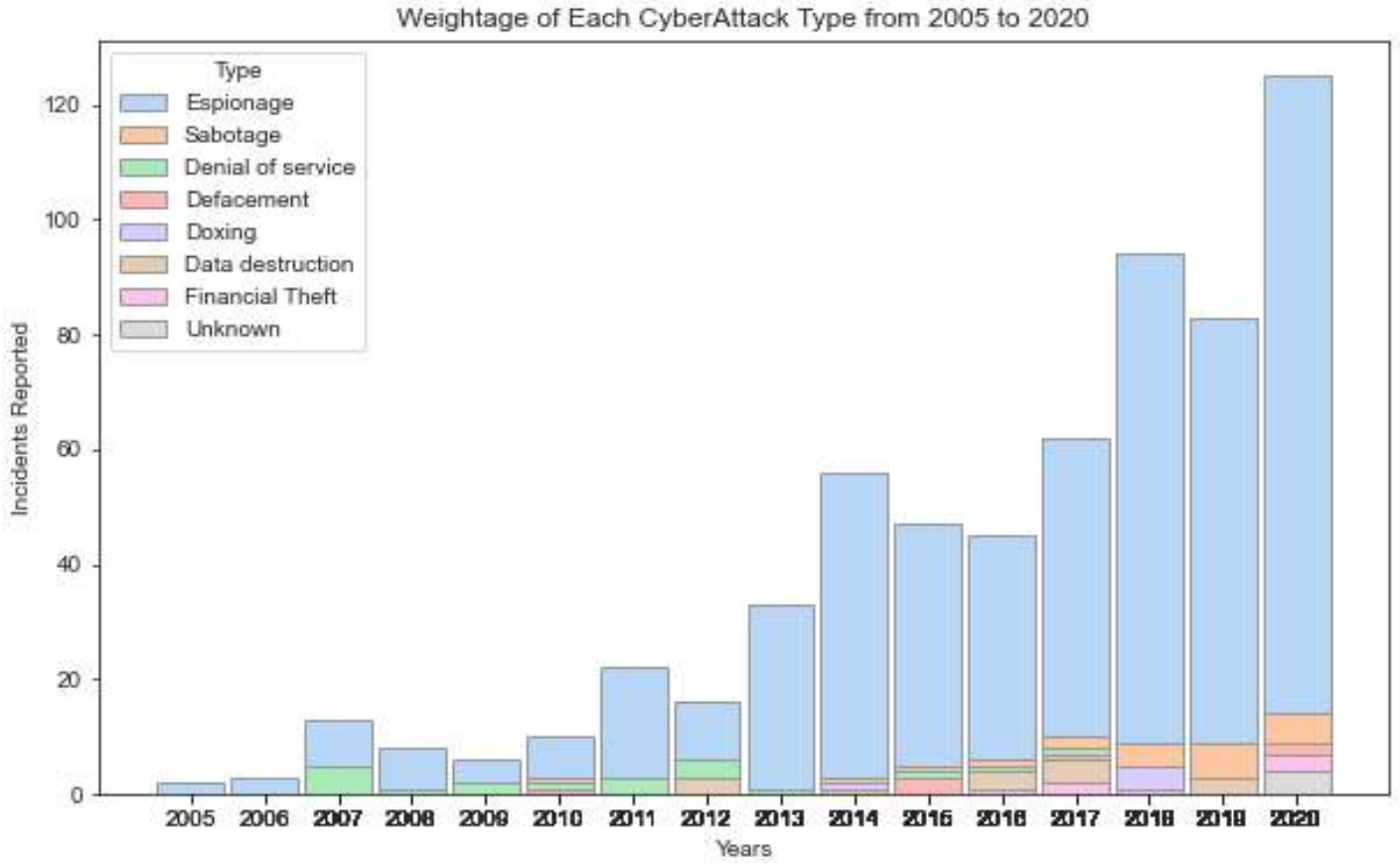
Cyber-Attack in Economy Sector

Attacks to the private & government sectors appeared in the past 15years seems to be increasing.



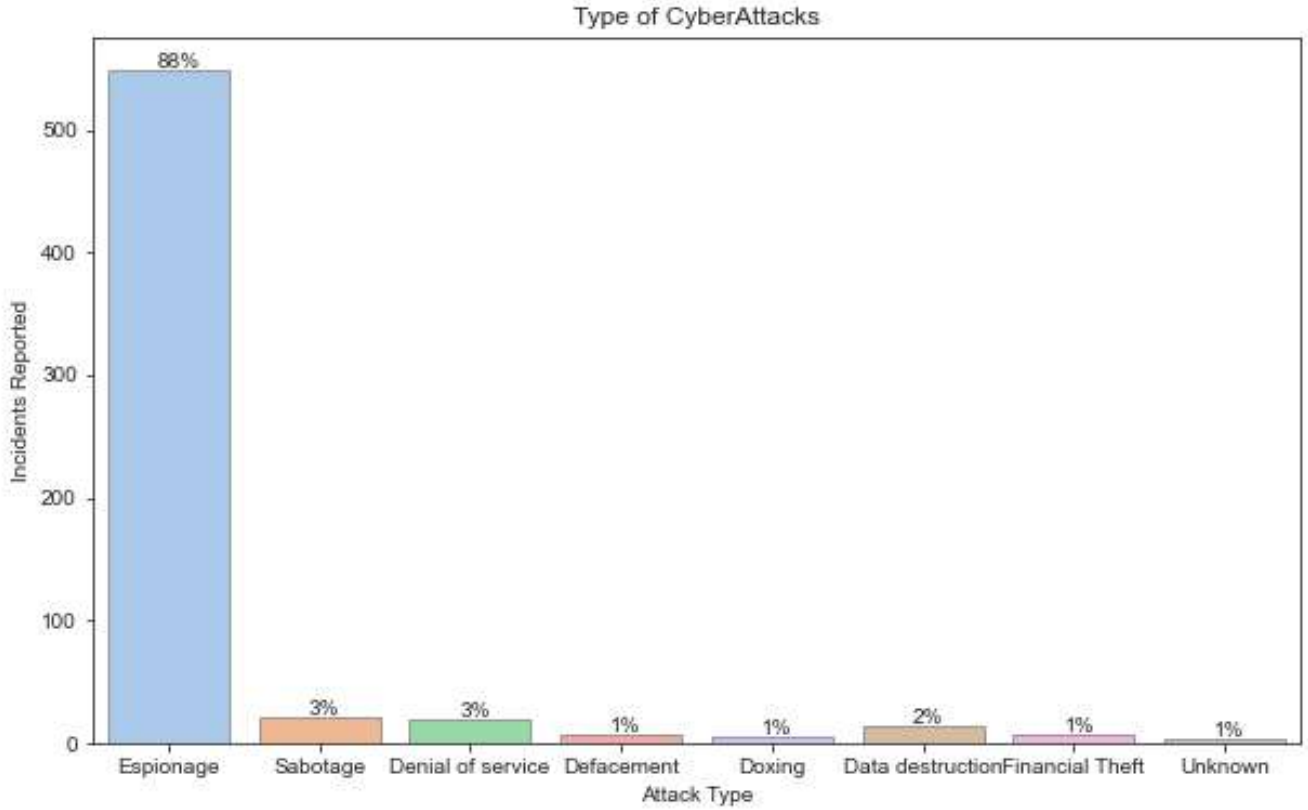
Types of Cyber-Attacks

Espionage form majority of the cyber-attack from 2005 to 2020.



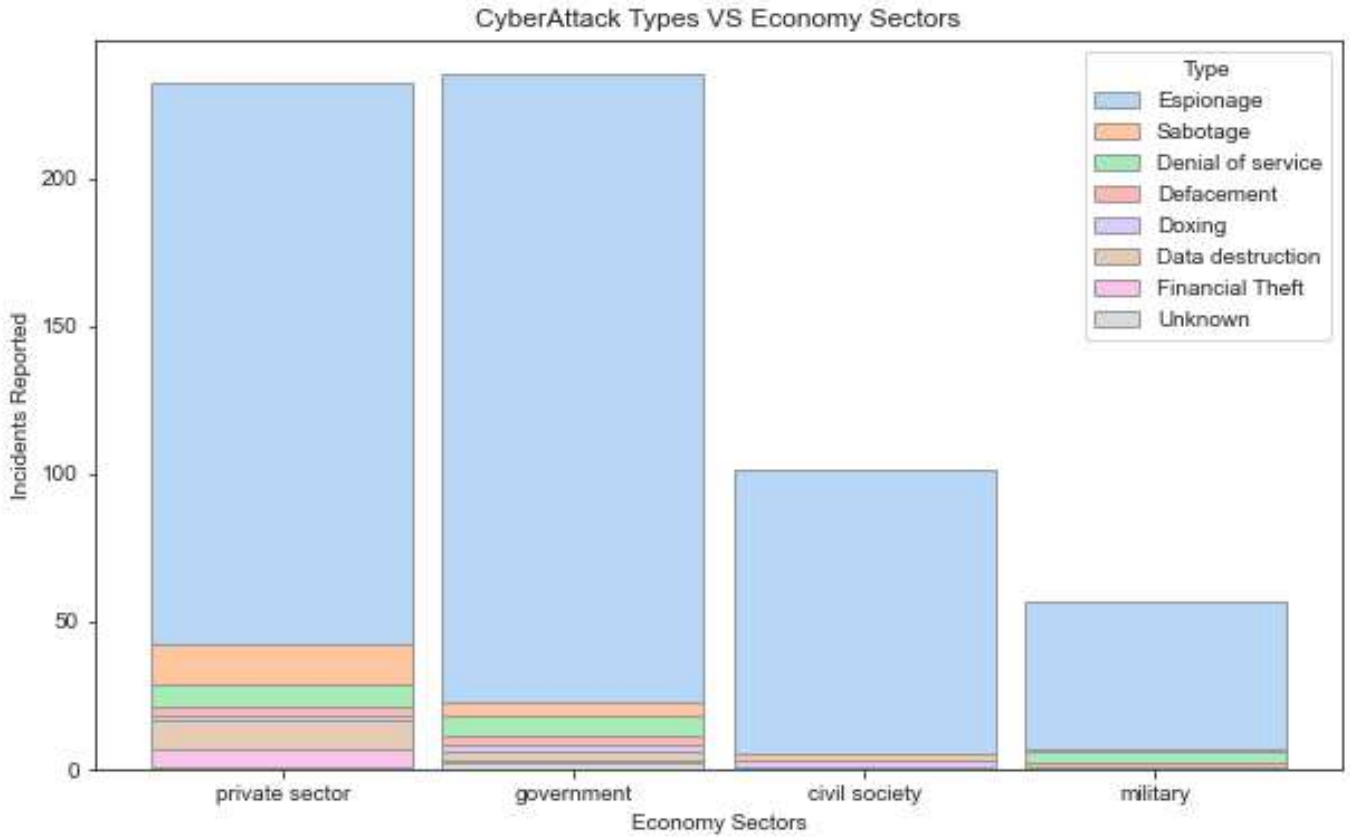
Types of Cyber-Attacks

88% of the Cyber-Attacks involved Espionage



Types of Cyber-Attacks

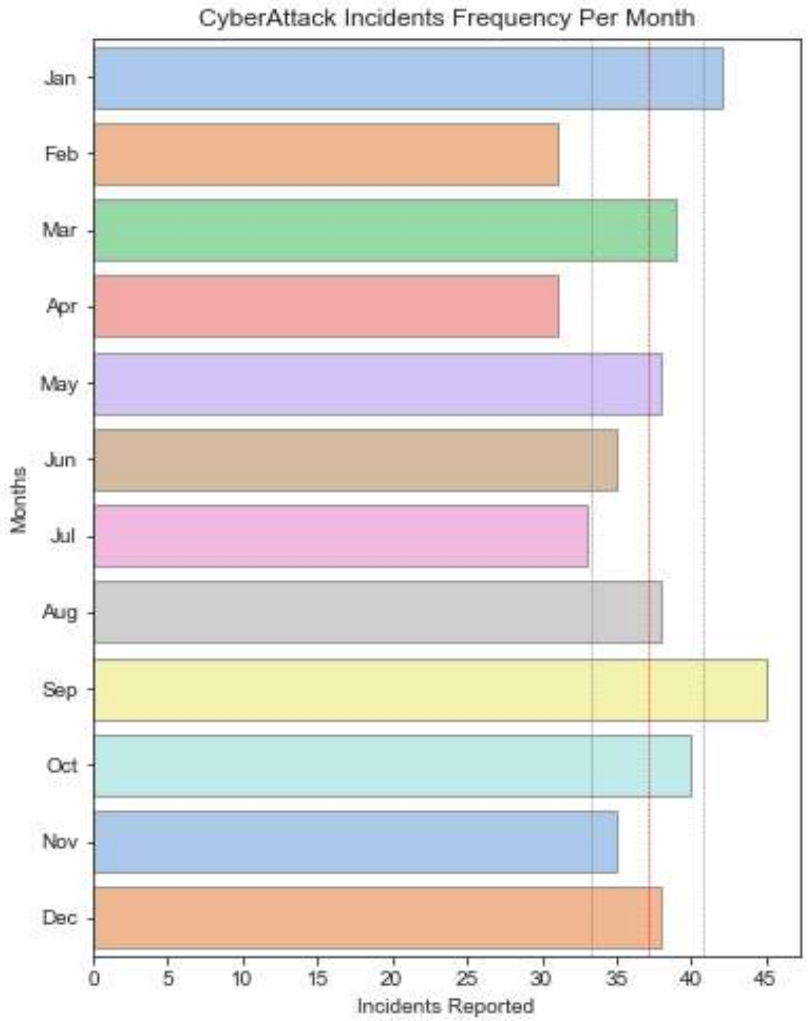
- Majority of the cyber-attack type found to be Espionage in all Economy Sectors.



Cyber-Attack Frequency Per Month

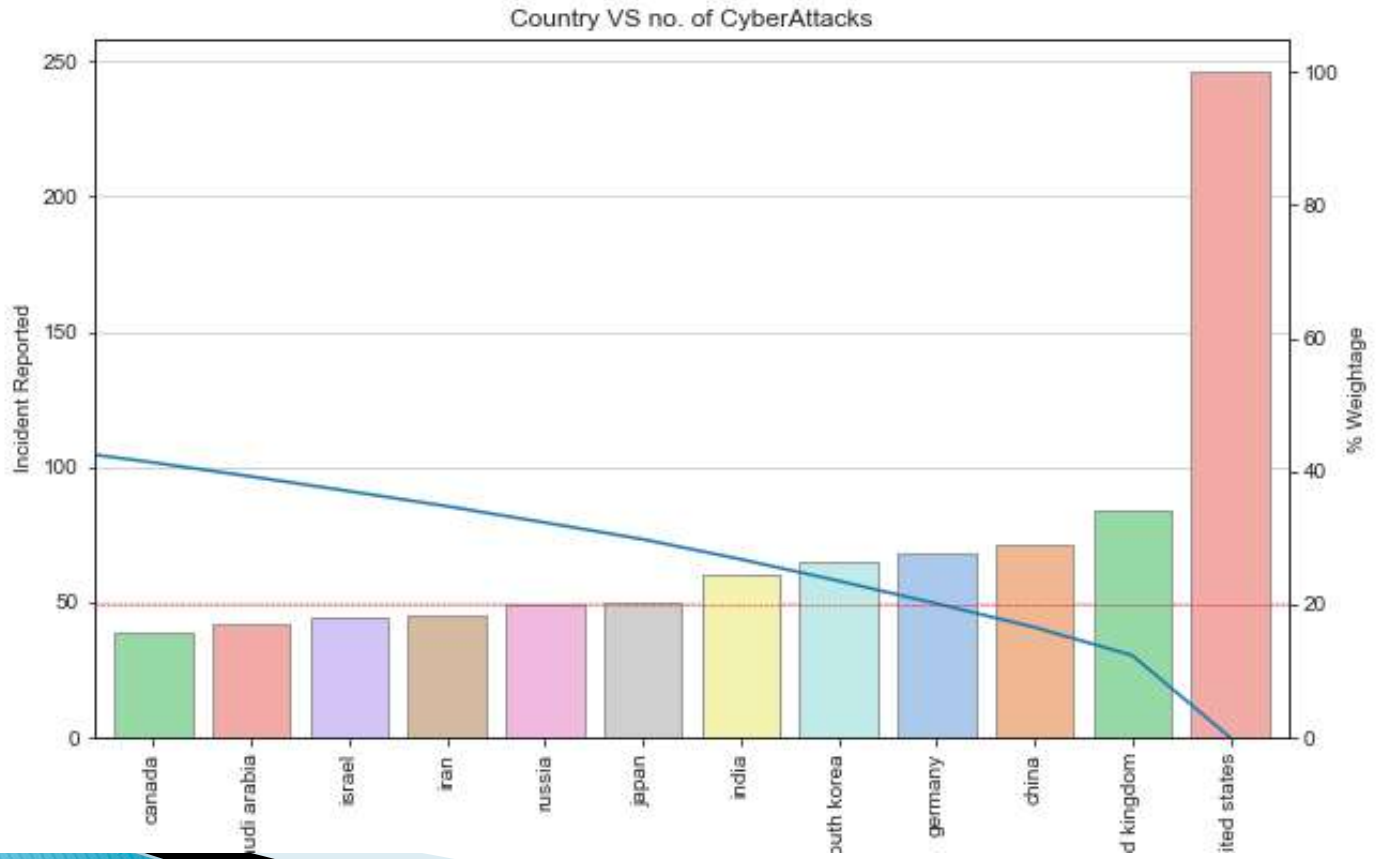
.

Average 38 cyber-attacks a months throughout a year.



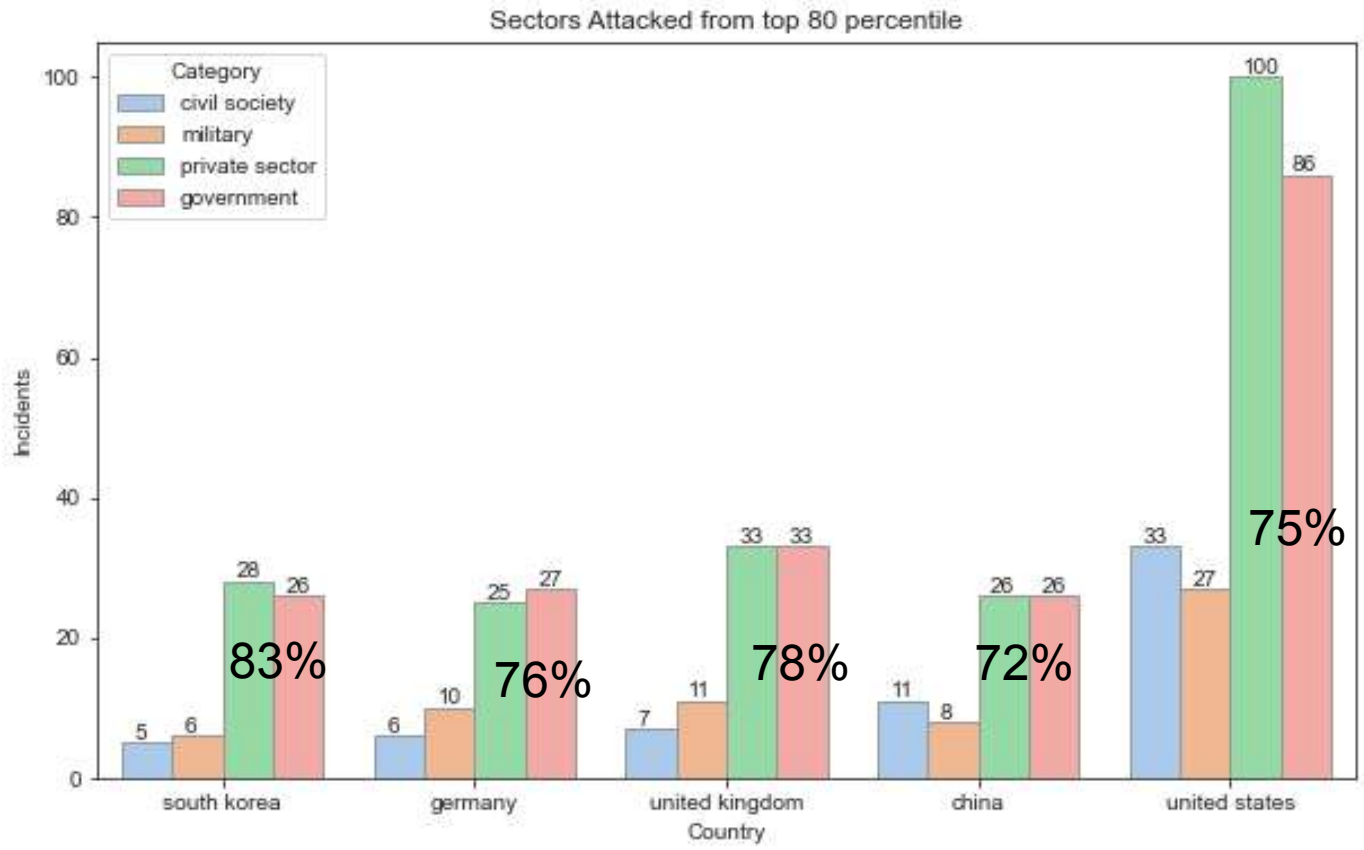
Country VS no. of CyberAttacks

80%˙of the Cyber-attacks targeted United States, United Kingdom & China.



Country VS no. of CyberAttacks

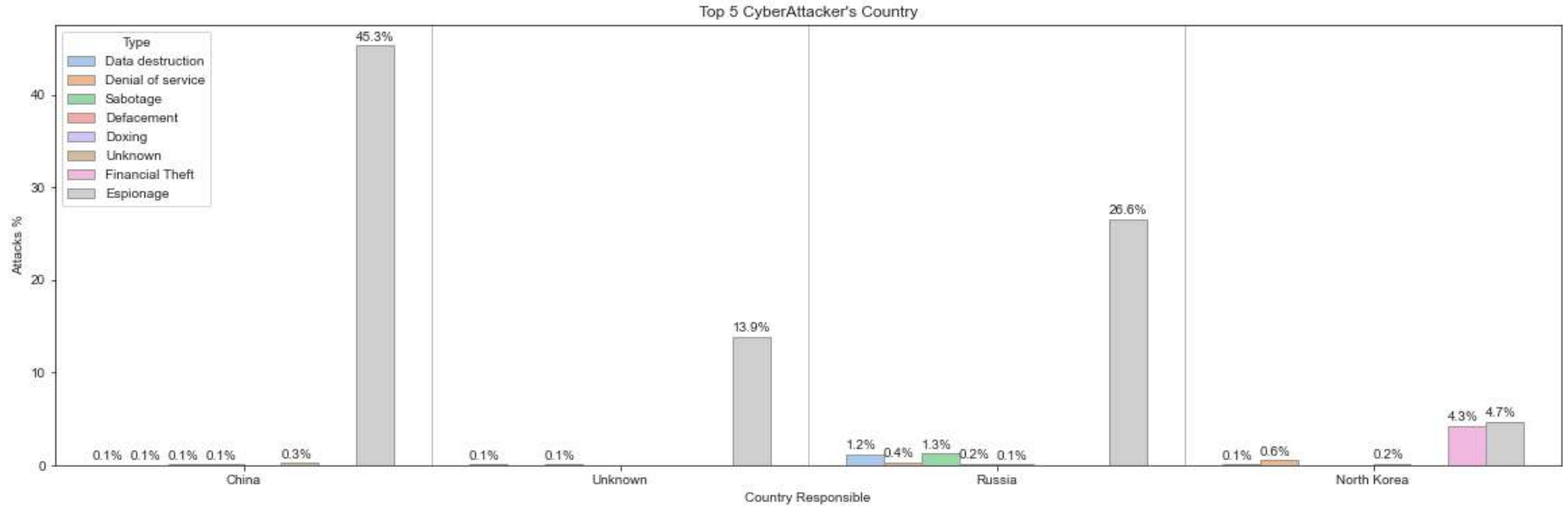
Graph show that consistency of around ~75% attacks targeted on Private and Government Sector to the top 5 countries that contribute to 80% of the global cyber-attack incidents.



Top 5 Attackers (Country)

.

Graph shown that majority attacks were Espionage type from all 5 countries.



Word Cloud On China Attacker Description

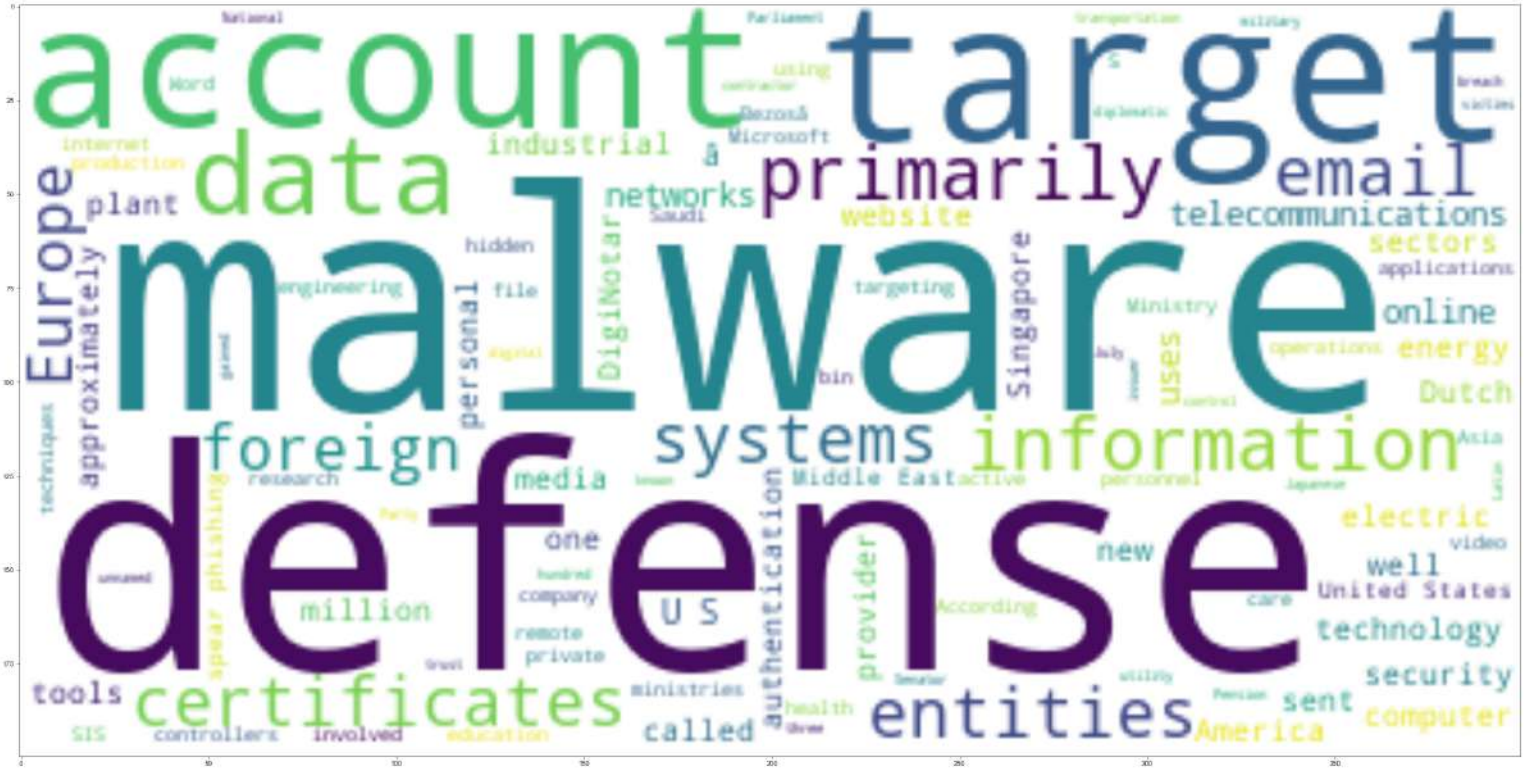
.

From Word cloud that attacks most related to **stealing of data/information on defense**, technology by deploying **malware**, **phishing email** and attacking through **network to company/firm or military**

Word Cloud On Unknown Attacker Description

.

From word cloud that attacks from unknown sources related to **stealing of information** on **defense, technology, personal** by deploying **malware, phishing email or attacking through network**



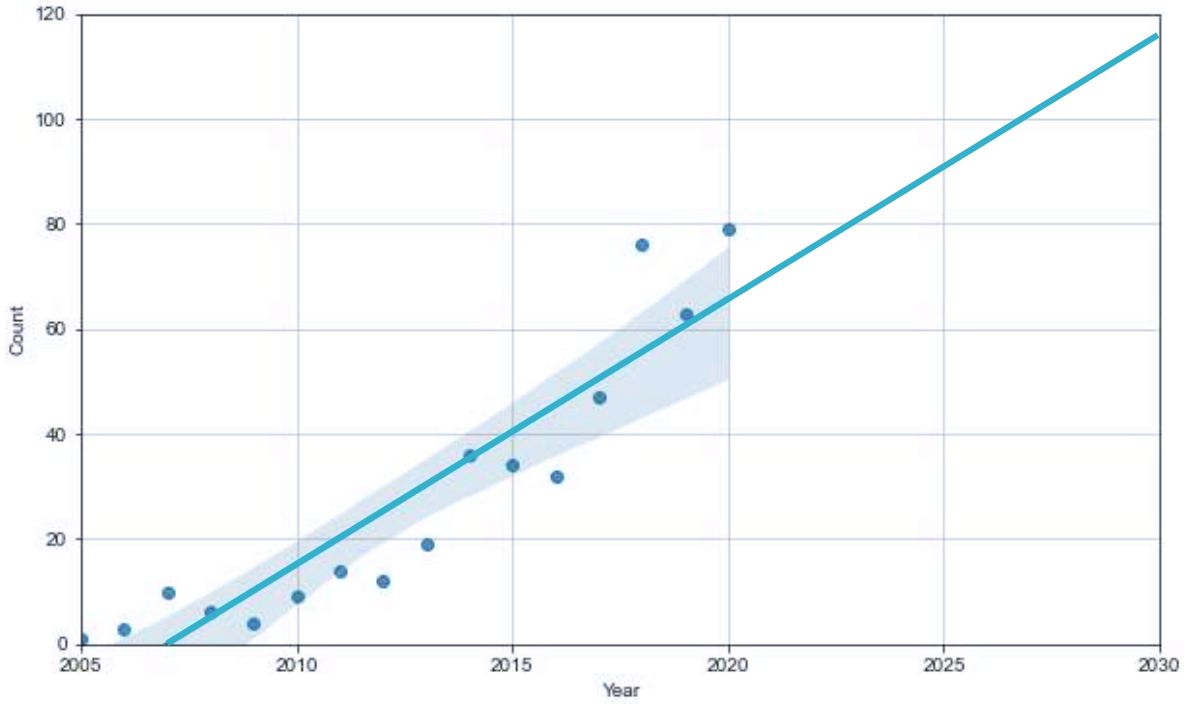
Cyber-Attack Forecast

Based on regression plot, Cyber-attack forecast to increase by 43% by 2030.

Attacks to Private & Government sectors will also increase from 235 to 336 as it proportional to the growth of total cyber-attacks

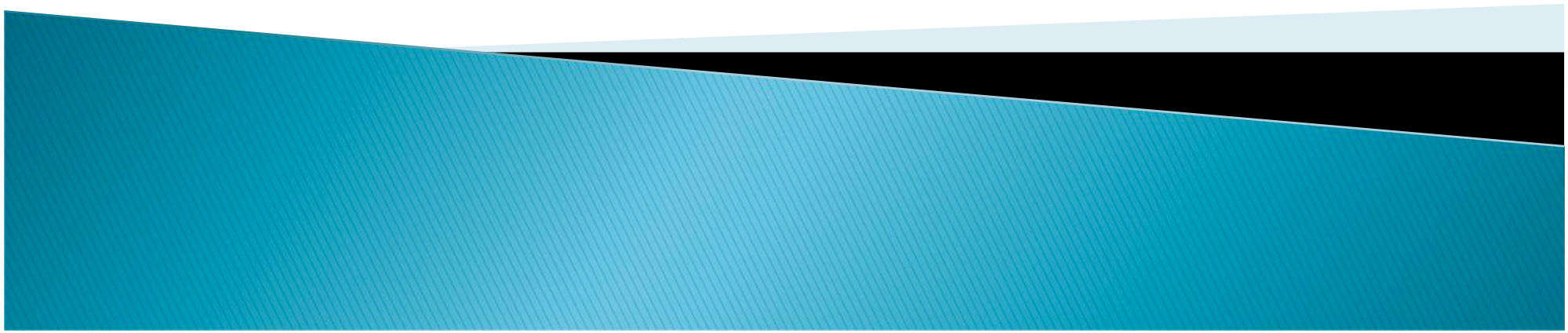
88% of the attacks will be Espionage

Average attacks per month will increase to 54 from 38

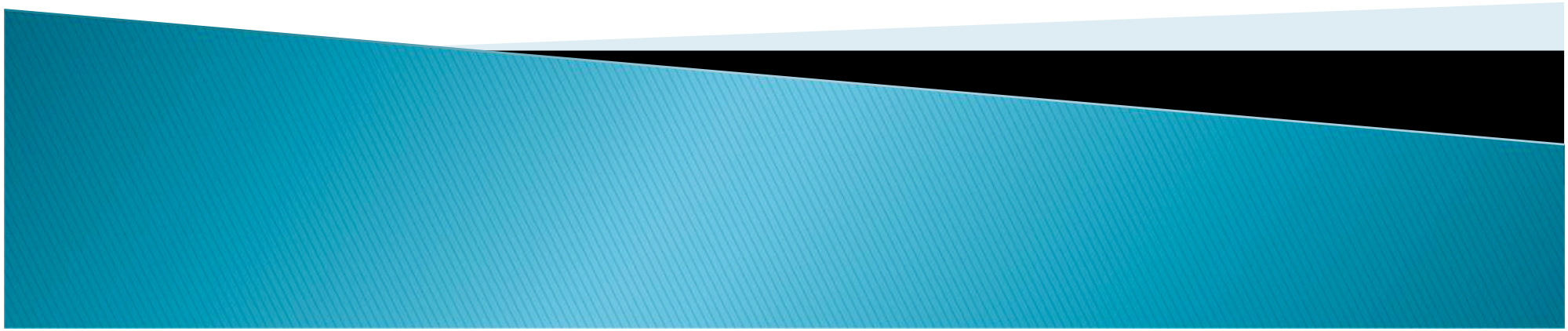


Due to the growth of cyber attacks and based on the dataset, companies & militaries in government & private sector should look into improving their protection in espionage attacks from other countries (other

Thank You!



Appendix

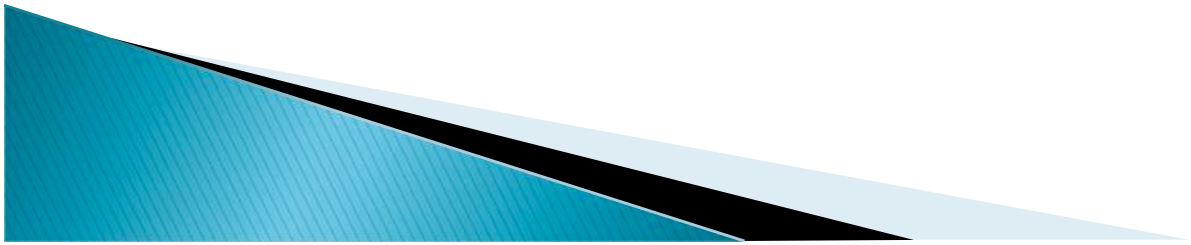


Dataset Overview

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.



Records From 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

Records From Dataset

Title – The title of the attacks

e.g. "Compromise of SingHealth, a large health-care provider in Singapore"

***Date** – The date of the cyber-attack recorded

e.g. "7/20/2018"

***Affiliations** – The potential attacker (group/organization) that was responsible for the attack

e.g. "Believed to be the work of Whitefly"

***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."

* - Important Column

Records From Dataset

Response – The victim's feedback to the attack. Most of them are empty
e.g. "Criminal charges", "Denouncement"

*Victims – The victim of the attack (country, firm, organization etc)
e.g. "Singhealth"

*Sponsor – The potential country that host the attacker

*Type – The objective type of the cyber-attack
e.g. "Espionage", "Financial Theft", "Sabotage"

*Category - The economic sector that been targeted
e.g. "Government", "Private", Civil Society"

Source 1, 2 & 3 – The reference link/sources to the record

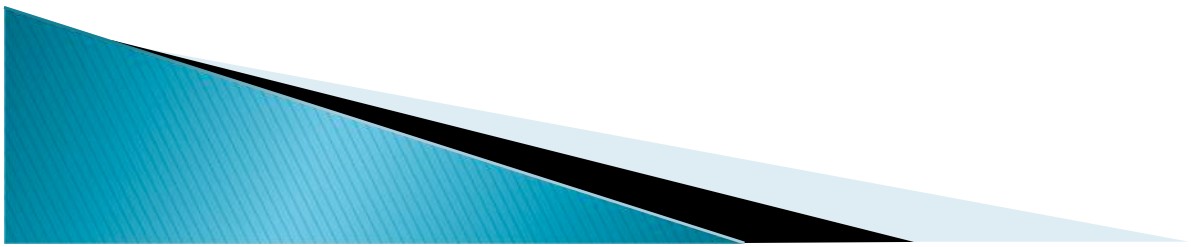
* - Important Column

Questions to Find Insights From Dataset

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 type to the country & sector
5. The attackers/sponsor that responsible to most of the cyber-attack around the world.
6. The frequency of the top 5 attackers/sponsor.
7. Which period in a year have the highest rate of attack?
8. The relationship of the attackers and their attack type.
9. The relationship of the attackers and the targeted country & sector
10. The forecast of future cyber-attack direction?
11. Recommendation of any vulnerable area to look into.

Challenges Foresee With Dataset

- To fill up empty records. Might be able to extract it out from “Title” or “Description” columns.
- Extracting useful insight from “Description” column that might not be found in other columns. Using of 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?



Attack Frequency from the top 5 Cyber-Attackers (2005 to 2020)

No proper frequency pattern. Attacks from China found throughout the years.