Categorizing Cyber-Attacks: Analysis of Industry, Event-Type, Economic and Criminological Factors

SR - Behavioral and Social Sciences

Aaron D’Souza, Grade 9, The Cambridge School

Abstract

My research problem is to determine which industries are most affected by cyber-attacks, which types of attacks are most prevalent, and understanding potential causes for such industries being more vulnerable and attacked. Then, comparing with other industries with lower cyber events to find unique causes.

My hypothesis is that healthcare would be the victim with the greatest number of cyber-attacks with the primary type being exploitative due to a large concentration of personal data, recurrent access to the data by numerous individuals, and potentially outdated systems.

First, I obtained data from the publicly available Center for International and Security Studies at Maryland (CISSM) website. I then performed data analysis and data visualization in Python programming using the Google Colab environment to arrive at insights and classify cyber-attacks based on underlying economic and criminological factors.

The analysis indicates that the public administration industry has the highest number of cyber-attacks, followed by health care and social assistance. This finding is close but not exactly consistent with the research hypothesis. The work then proceeded with a detailed analysis of cyber-attacks and their classification by attack type, motive and actors, and to explain the discrepancy versus the hypothesis.

From an applications perspective, such classification of attacks is expected to assist industries and governments set policies and implement procedures to help safeguard against cyber-attacks. Further, such a study illustrates the efficacy of data science and data visualization in understanding social science.

Introduction

- **Research problem**
  - My research problem is to determine which industries are most affected by cyber-attacks, which types of attacks are most prevalent, and understanding causes for certain industries being more vulnerable and attacked. Then, compare with other industries with lower cyber events to find unique causes.
  - I became interested in this area from my participation in cybersecurity competitions, and classes on data science and artificial intelligence. Upon reading research papers, I became interested in pursuing this topic since it combines my interests in cybersecurity and data science
  - This topic is very important since cyber attacks affect real people threatening their privacy, financial security, and every aspect of one's life tied to data and/or technology. Additionally, cyber attacks on government agencies can threaten national security affecting public safety, social infrastructure, and defense.
    - Although important, as far as I know, this topic has not been investigated in a prior GSDSEF

- **Prior work**
  - I read several papers listed in the reference section of this document. One relevant reference is "A study of cyber attacks: In the healthcare sector" by Karuna Bhosale. This reference highlights challenges including endpoint device management, security awareness issues, and inadequate protection of connected medical devices and hospital information systems, demonstrating a need for enhanced cybersecurity strategies and solutions.
  - Limitation of prior work is that it does not identify which industry sector is most vulnerable to cyber attacks and why
    - My research addresses these limitations through a detailed data science analysis of cyber attack records to identify which industries are most impacted and the underlying causes

- **Hypothesis**
  - My hypothesis is that healthcare would be the victim with the greatest number of cyber-attacks with the primary type being exploitative due to a large concentration of personal data, recurrent access to the data by numerous individuals, and potentially outdated systems.
  - Basis of hypothesis:
    - My experience and knowledge of cyber attacks from participating in Cyberpatriot, Southern California Cyber Cup and National Cyber League competitions, my personal visits to doctors' offices and reading current news events on cyber attacks. As I researched this topic further, I read literature which supported my hypothesis and helped me refine it, e.g., U.S. Center for Disease Control article which explains that in the U.S. a vast majority of physicians use electronic medical records; thus allowing more cyber attacks

Materials

- Dear Judges, Since this is a data science project, I did not use any chemicals, instruments, or other physical tools. May I skip this slide and include an additional results slide instead?

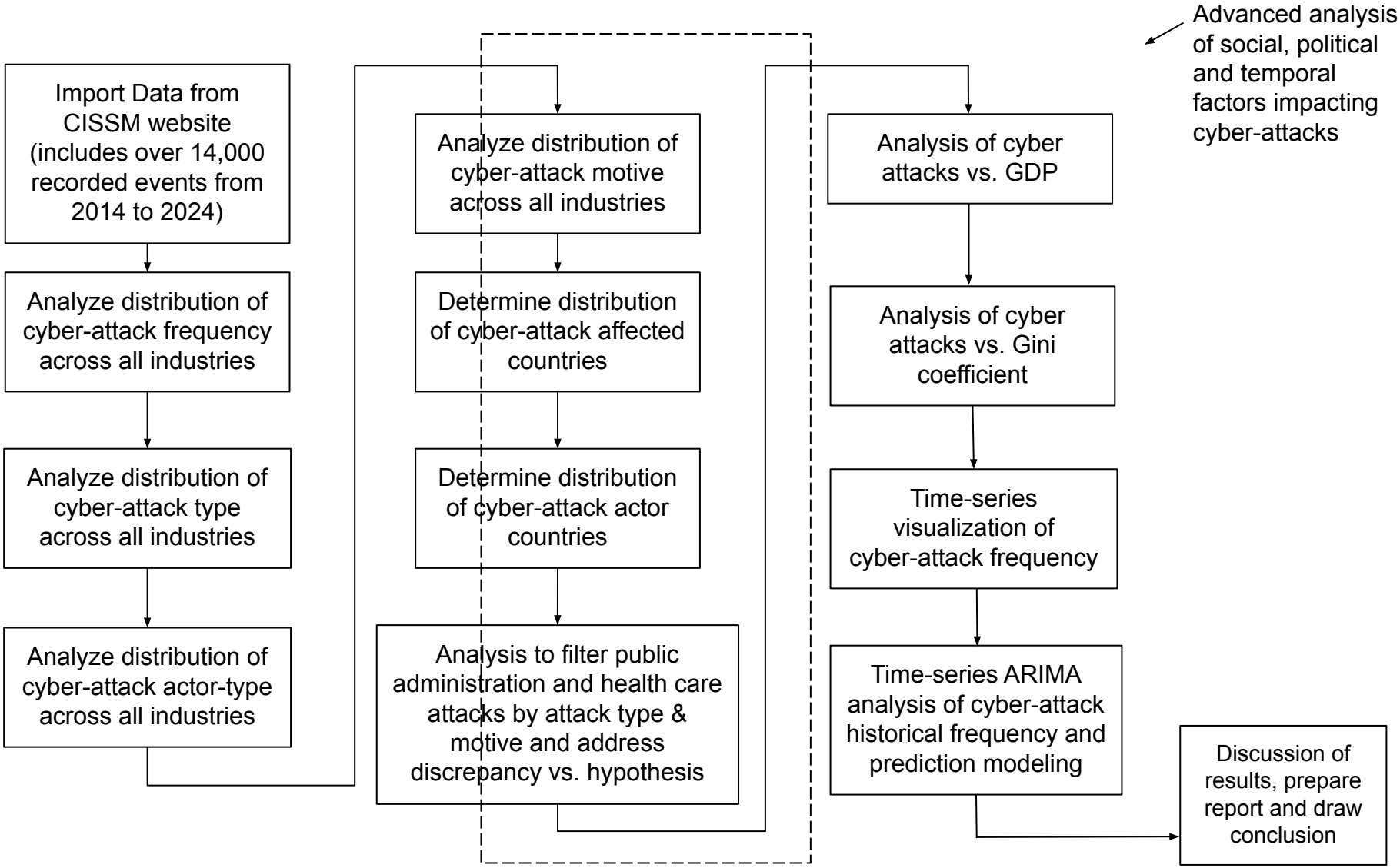
- This data science project uses data was sourced from Center for International and Security Studies at University of Maryland(CISSM).
  - The data, “Cyber Events Database” is a record of scraped publications of cyber-attack events occurring since 2014 to 2024 with over 14,000 recorded events.
  - I downloaded the cyber events dataset that I used from this link:
    - https://cissm.umd.edu/cyber-events-database

- I also used data from The World Bank’s Poverty & Inequality Indicators (PIP) which include various economic describing-variables per country including GDP, wealth gap, gini coefficient, etc.
  - I downloaded the PIP dataset that I used from this link:
    - https://pip.worldbank.org/poverty-calculator

- The cyber events are classified by date, actor type (criminal, hacktivist, nation-state, undetermined), actor, organization (which organization was impacted by the event), industry, motive (protest, sabotage, espionage, financial, undetermined), event type (disruptive, exploitative, mixed undetermined), event description, source url, target country, and actor country.
  - The data collects only publicly available information, using python to “scrape” data.
  - Then researchers from CISSM reviewed and categorized the data to ensure the source material information is reflected consistently in the data and the event identified fits their definition of a cyber event. The industries were categorized per the North American Industry Classification System.

Methods / Procedure

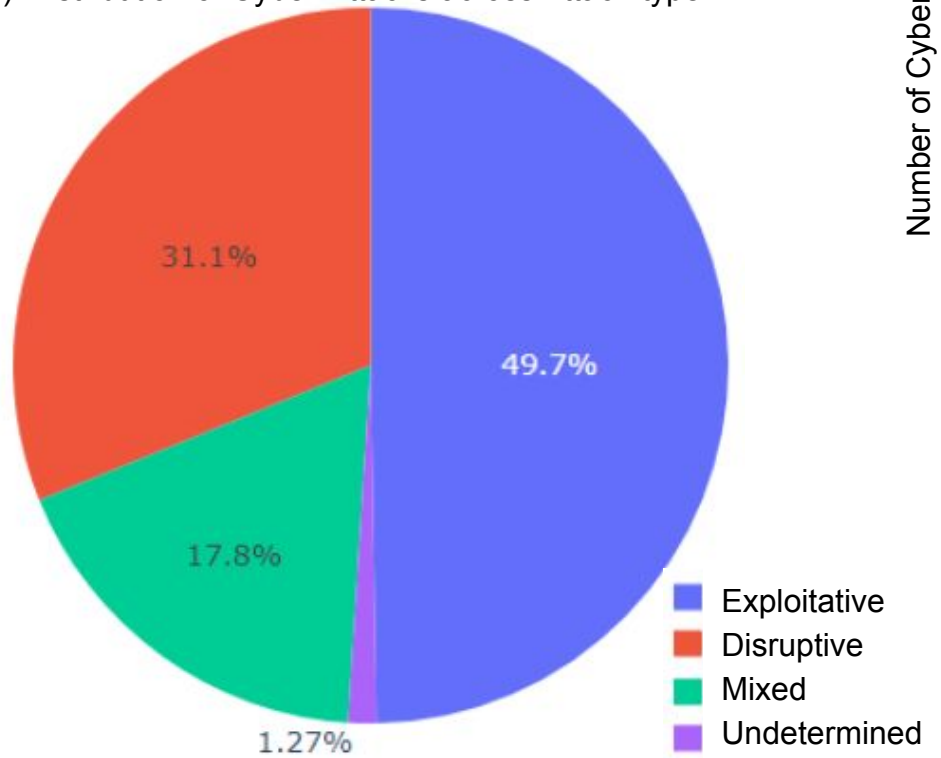
- I Investigated the research problem using the following procedural steps shown with a block diagram flow chart
- Implementation of these steps in my Python program in the Google colab environment is at the link: https://colab.research.google.com/drive/1ThXrjvK0lxD_CQ6YjgvV7VEjMUwJWke7?usp=sharing
  - Full set of results in Google colab, only selected results are on next slides due to restriction on number of slides

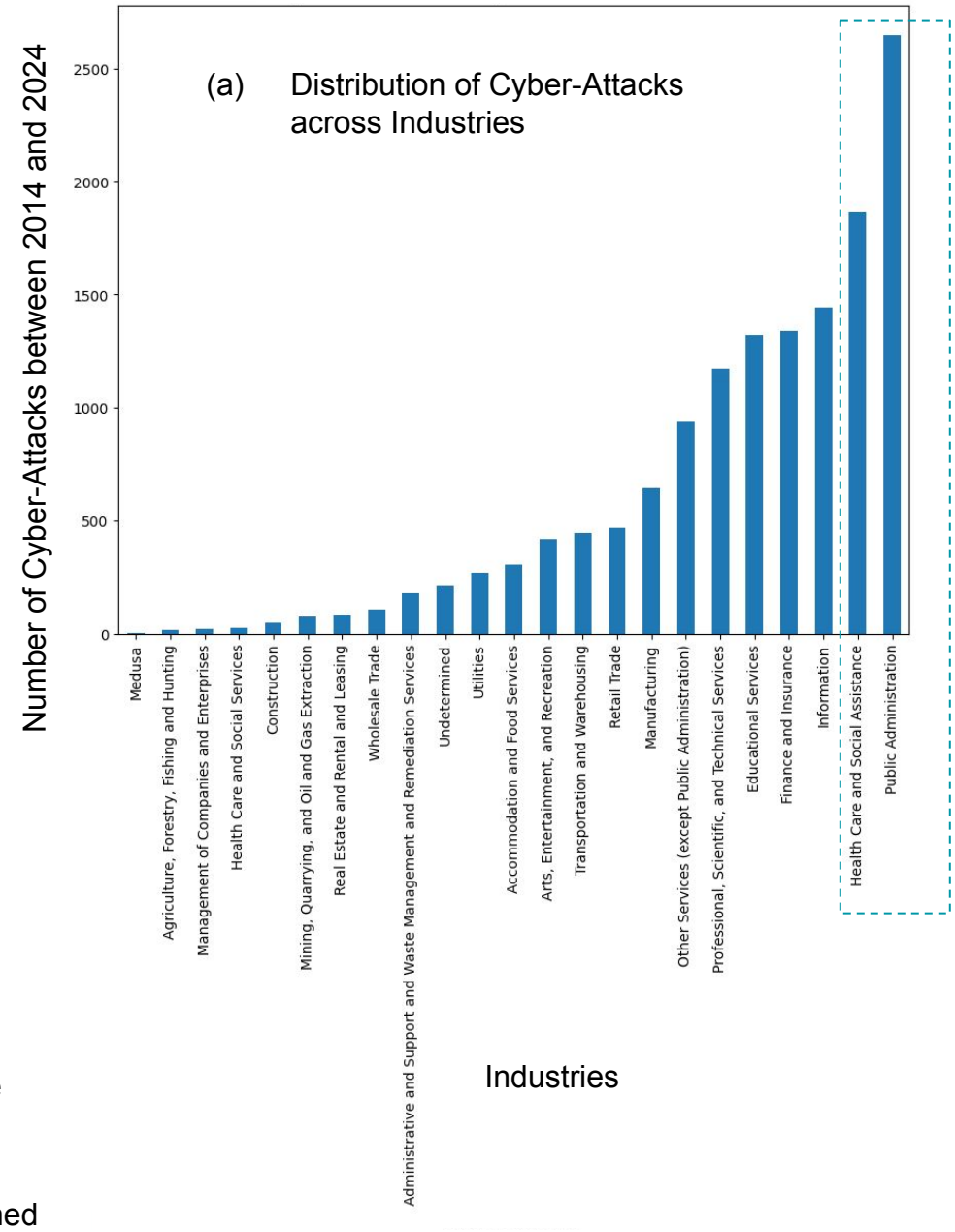


Results (1): Distribution of Cyber Attacks (a) Across Industries & (b) Attack-type

- Findings:
  - (a) Public Administration has the highest number of cyber attacks, followed by Healthcare and Social Assistance
  - (b) Almost 50% of the total number of cyber-attacks between 2014-2024 are exploitative, followed by ~31% disruptive
- My hypothesis that health care would have the highest is close, but not exactly accurate
  - Motivated my further analysis on the causes of the distribution of the attacks on the following slides considering attack-type, motive and other factors

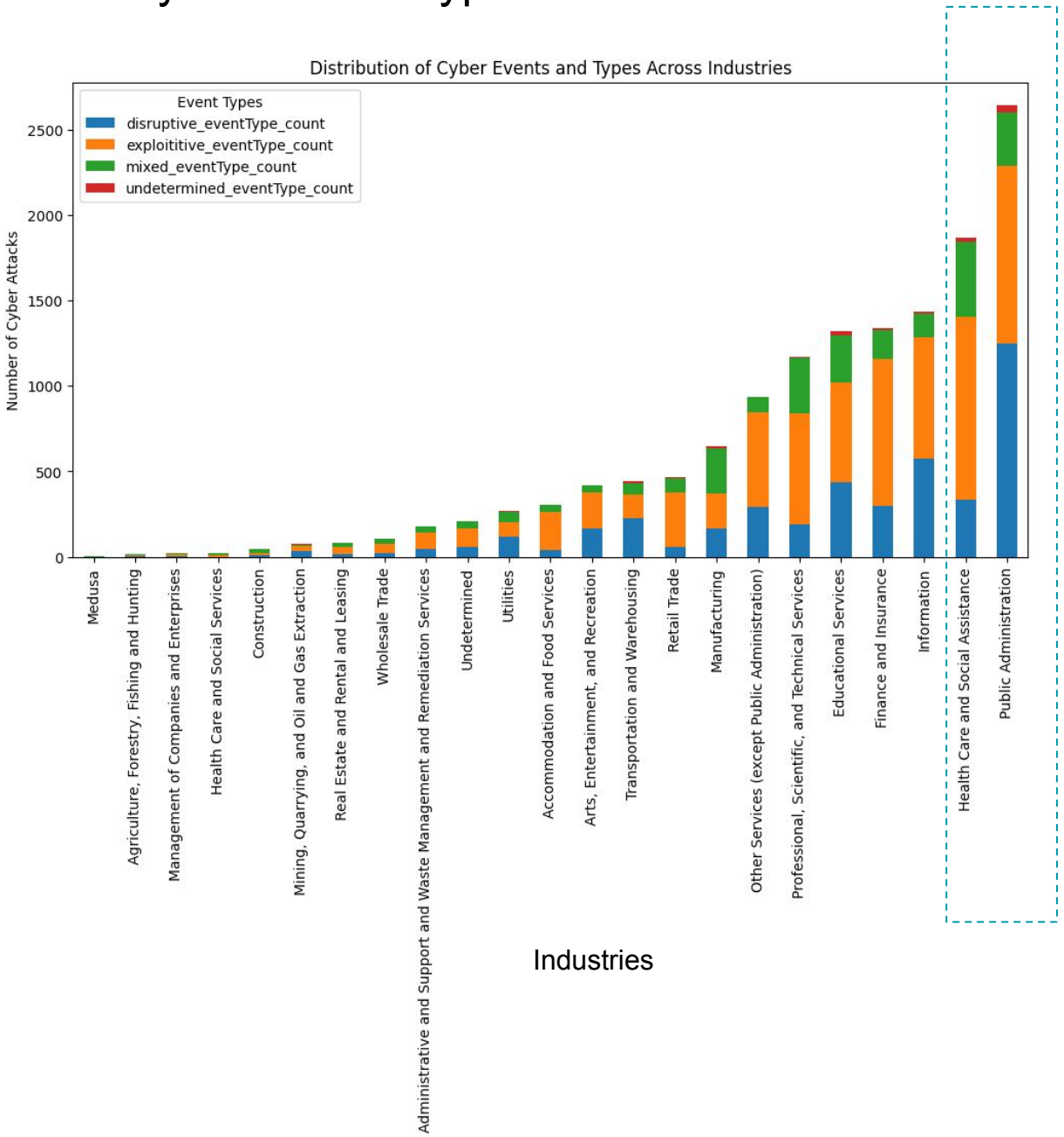
(b) Distribution of Cyber-Attacks across Attack-type





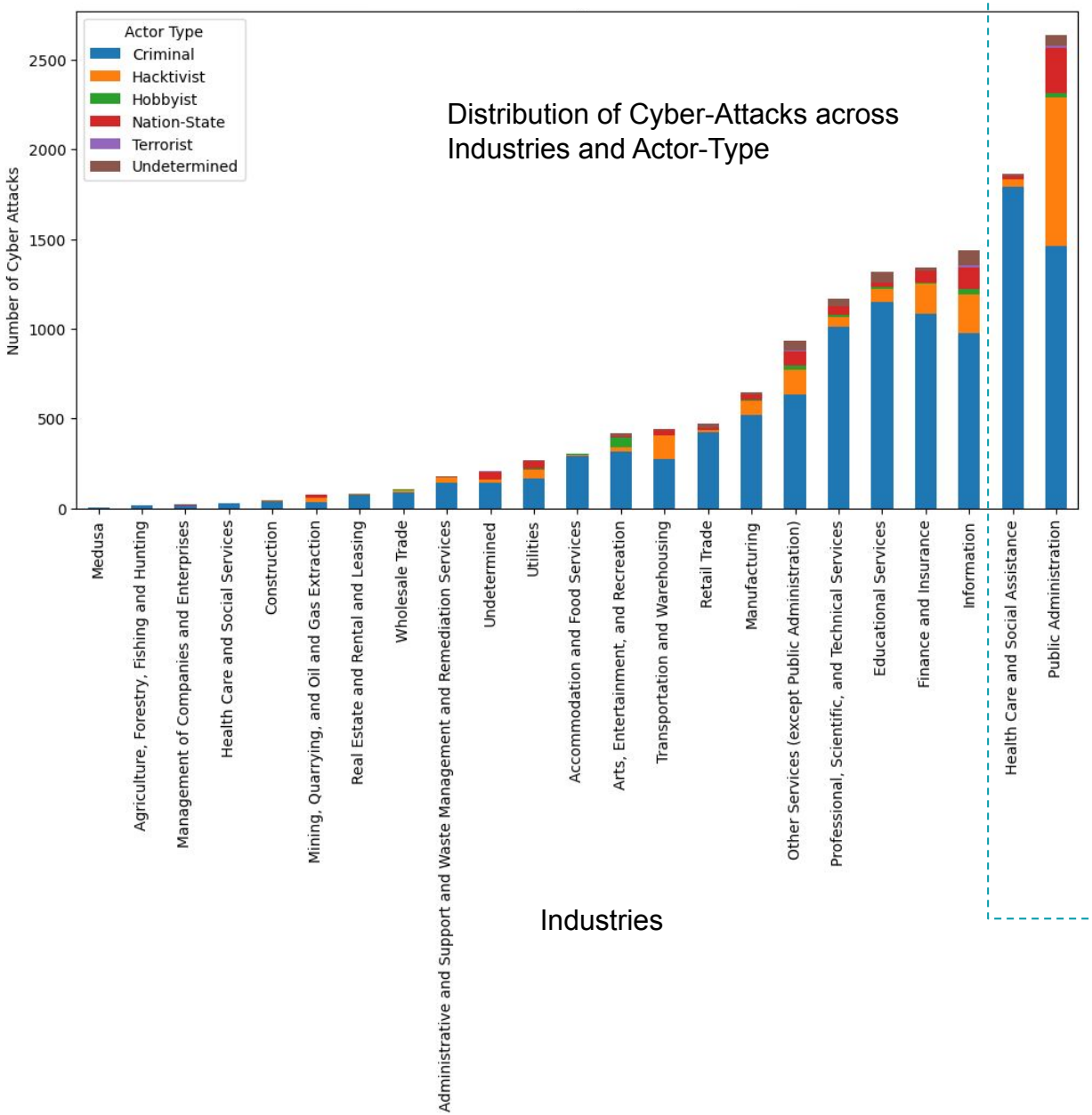
Results (2): Cyber Attacks by industry and Attack-type

- Findings:
  - (a) Public Administration has the highest number of cyber attacks, with the most number of attacks being disruptive attacks followed by exploitative attacks
  - (b) Healthcare and Social Assistance is the second number of cyber attacks, with the most number of attacks being exploitative attacks followed by mixed attacks, and then disruptive attacks



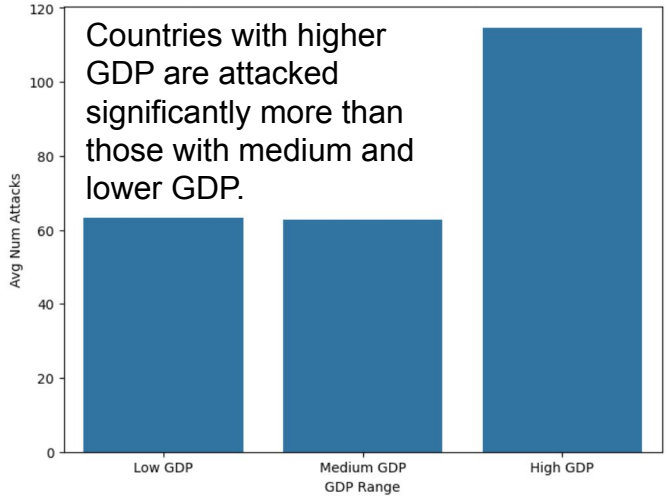
Results (3): Cyber Attacks by industry and actor-type

- Findings:
  a. Attacks on healthcare and social assistance are almost all done by criminals
  b. Attacks on public administration are approximately two-third done by criminals, followed by hactivists and nation-states.
  c. As can be seen from the bar graphs, *after removing hactivist, nation-states and terrorists*, health care and social assistance has higher number of cyber attacks as compared to public administration

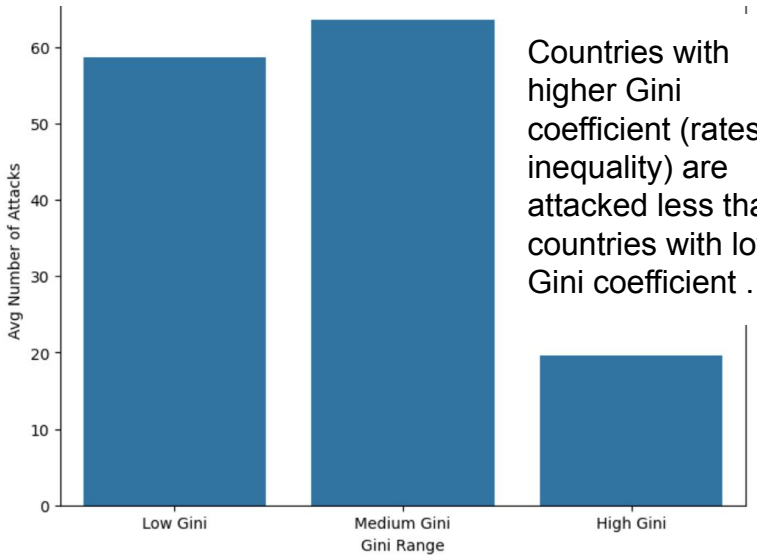


Results (4): Cyber-attack vs (a) Country GDP, (b) Country Gini-Coefficient, (c) Time-series autocorrelation and (d) ARIMA model predictions

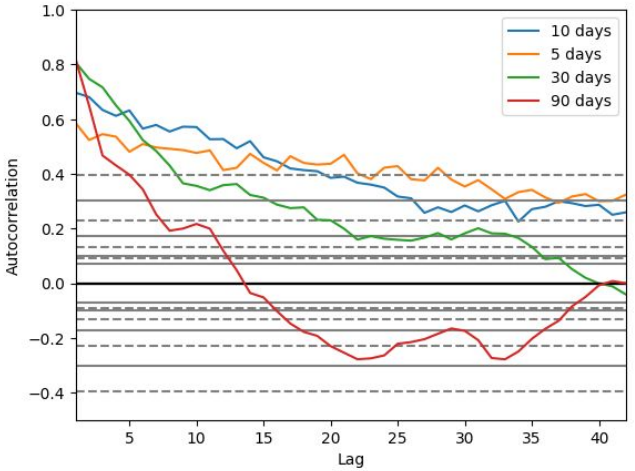
(a) Number of attacks by Country GDP



(b) Number of attacks by Country Gini Coefficient

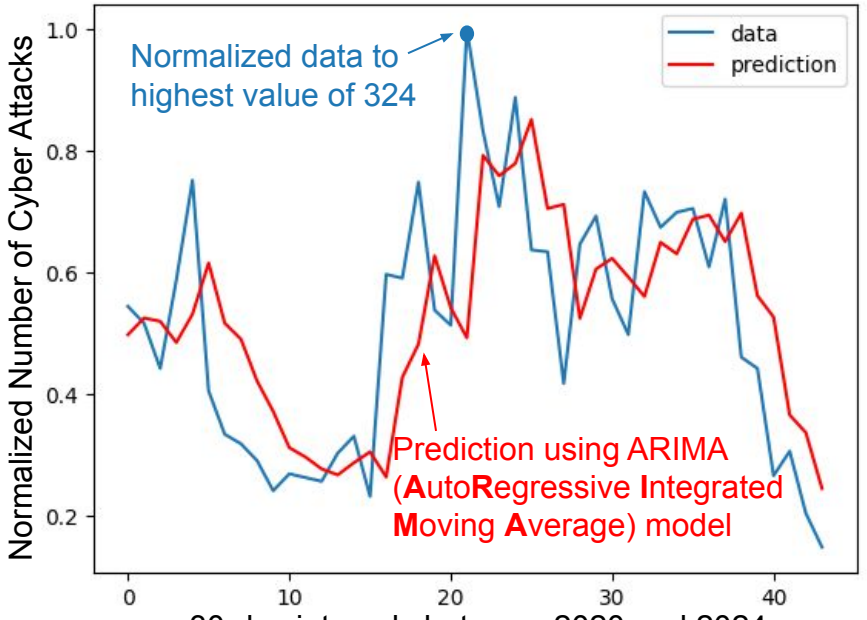


(c) Autocorrelation of cyber-attack time-series vs. lag



Autocorrelation lag used to calculate ARIMA coefficients and identify how many lags are statistically significant for each time period.

(d) ARIMA time-series model prediction vs. data



Discussion

- The public administration industry has the highest number of cyber attacks, followed by health care and social assistance. This finding is close but not exactly consistent with the research hypothesis that health care would have the highest number of attacks.

- The work then proceeded with a detailed analysis of cyber attacks and their classification by attack type, motive and actors.
  - The findings indicate that attacks on public administration, and health care & social assistance differ in terms of these underlying factors: While public administration is found to have an approximately 20% higher number of disruptive attacks than exploitative attacks, the health care and social assistance industry is dominated by exploitative attacks which are approximately 3 times higher than disruptive attacks.
  - Additionally, in terms of actor type, public administration is attacked by a moderately higher (56%) criminals than politically or ideologically motivated nation-states and hacktivists (42%), while attacks on health care and social assistance industry are dominated by criminals (96%).
  - Then, from a motive perspective, 51% of attacks on public administration industry had a non-financial motive, whereas approximately 97% of attacks on the health care and social assistance industry had a financial motive.
  - Furthermore, the motive/actor profiles differ between the exploitative and disruptive attack types.
    - The motive of exploitative attacks is mostly financial (78%), while disruptive attacks have a slight majority of non-financial motives (52%).
    - Similarly, with actor profile, exploitative attacks have an overwhelming majority of non-political criminal actors (85%) while disruptive attacks also have a large prevalence of criminal actors, though less at 60% with a larger number of political actors.

- Considering cyber attacks only by non-political actors, i.e., not by nation-states or hacktivists, health care and social assistance is found to have the highest number of cyber attacks which supports the hypothesis of this research.

- Additionally, countries with higher GDP and countries with greater equality (using Gini coefficient) are attacked more frequently
  - This is because such countries have more economic goods to capture and are more likely to be targets of political actors due to their greater influence across the world.

- Based on my ARIMA time-series model of the cyber attack data, it is clear that trends in cyber attacks are predictable.
  - This finding could be used to enhance security during periods of increased vulnerability.

Please note: Because of space restrictions, not all of my results are in this slide deck. Please see my Google colab which includes all results https://colab.research.google.com/drive/1ThXrjvK0lxD_CQ6YjgvV7VEjMUwJWke7?usp=sharing

Conclusion

- My hypothesis that healthcare would be the victim with the greatest number of cyber-attacks was not entirely accurate.
  - However, when considering only either exploitative attacks or those with financial motive, healthcare has the greatest number of cyber-attacks.
    - This is due to the large concentration of personal data, recurrent access to the data by numerous individuals, and potentially outdated systems.
  - The only industry that surpassed health care in number of cyber was Public administration due to a large number of politically motivated disruptive attacks.

- This research identifies the industry with greatest number of cyber attacks and criminological and economic factors unlike prior studies found in my literature review:
  - Politically motivated disruptive attacks resulted in Public administration having the greatest number of cyber attacks
  - Financially motivated attacks resulted in a large number of attacks on the healthcare industry
  - Using GDP, countries with higher economic resources are attacked more frequently as attackers see a greater reward and that they are more influential on the global scene resulting in more enemies and politically-motivated attacks.
  - Using Gini coefficient, countries with less inequality are attacked more frequently as those countries generally have greater overall prosperity to exploit.

- My time series analysis finds that there are predictable trends in cyber attacks that can help to enhance security at high-risk moments.

- Application-wise, my research can help to identify the where and why of cyber-attacks such that government policy and private cybersecurity work can focus on solving the problems where they exist.

- Also, my work demonstrates the efficacy of using data science to recognize patterns in records of social science events.

Reference list

Anderson, R. and Moore, T. (2006). The Economics of Information Security. *Science*, [online] 314(5799), pp.610–613. Available at: https://www.jstor.org/stable/20031627 [Accessed 23 Jul. 2024].

Bhosale, K.S., Nenova, M. and Iliev, G. (2021). A study of cyber attacks: In the healthcare sector. *IEEE Xplore*, [online] pp.1–6. doi:https://doi.org/10.1109/Lighting49406.2021.9598947.

Center for International and Security Studies at Maryland. (2024). *Cyber Events Database*. [data set] Available at: https://cissm.umd.edu/cyber-events-database.

Coventry, L. and Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas*, [online] 113(113), pp.48–52. doi:https://doi.org/10.1016/j.maturitas.2018.04.008.

CyberPeace Institute (n.d.). *Cyberattacks Impact and Harm on the Public administration sector | CyberPeace Institute*. [online] cyberconflicts.cyberpeaceinstitute.org. Available at: https://cyberconflicts.cyberpeaceinstitute.org/impact/sectors/public-administration [Accessed 21 Sep. 2024].

Electronic Privacy Information Center (2024). *Medical Records*. [online] Epic.org. Available at: https://archive.epic.org/privacy/consumer/med_record.html.

Myrick, K.L., McNeal, M. and DeFrances, C. (2022). QuickStats: Percentage of Office-Based Physicians Using Telemedicine Technology, by Specialty — United States, 2019 and 2021. *Morbidity and Mortality Weekly Report (MMWR)*, [online] 71. doi:https://doi.org/10.15585/mmwr.mm7149a6.

Office of the Director of National Intelligence (2024). *Recent Cyber Attacks on US Infrastructure Underscore Vulnerability of Critical US Systems*. [online] Available at: https://www.dni.gov/files/CTIIC/documents/products/Recent_Cyber_Attacks_on_US_Infrastructure_Underscore_Vulnerability_of_Critical_US_Systems-June2024.pdf.

Sophos (2024). *Cybersecurity as a Service Delivered | Sophos*. [online] Available at: https://www.sophos.com/en-us/press/press-releases/2024/09/two-thirds-healthcare-organizations-hit-ransomware-four-year-high.

World Bank (2025), Poverty and Inequality Platform (version 20240627_2017_01_02_PROD) [data set]. pip.worldbank.org. Available at: https://pip.worldbank.org/poverty-calculator