SPATIAL AND TEMPORAL VARIABILITY OF CYBER ATTACKS IN KENYA

By

LORNA MAKENA

Project report submitted to the department of Geomatics Engineering and geospatial Information Systems degree of Bachelor of Science in Geospatial Information Science.

Department of Geomatic Engineering and Geospatial Information Systems

DECLARATION

I declare that this project is my own work and has not been submitted by anybody else in any other university for the award of any degree to the best of myknowledge.

Sign…… Date……..

Lorna Makena

ENC222-0128/2018

Department of Geomatics Engineering and Geospatial Information Systems (GEGIS)

Jomo Kenyatta University of Agriculture and Technology

CERTIFICATION

This project has been submitted for examination with my approval as the candidate's supervisor.

Sign……. Date ……

Dr. Eunice Nduati.

Lecturer

Department of Geomatics Engineering and Geospatial Information Systems (GEGIS)

©GEGIS 2022

Acknowledgements

The results in this document were not fully accomplished by the student author only. External help from lecturers and experts in the field of cyber security mentored and provided logistic support during the process.

First and foremost, I'd want to express my sincere gratitude to my major advisor and supervisor, Dr. E. Nduati, PhD from the Department of Geomatic Engineering and Geospatial Information Systems-GEGIS. Your support and mentorship during the whole exercise pushed me in to ensuring provision of excellent results. You have been an amazing role model for me. You believed in my ideas and gave me the freedom to pursue them. Working under your guidance hasprovided me with the resources and confidence I need to accomplish my career achievements. God bless you.

I'd also like to offer appreciation to Mrs. Mugambi, lecturer at the department of Computer Science and IT. By guiding me through field of cyber security which was a new field to me helped me be able to achieve the results below. By also connecting me to other sources where I could learn more about cyber attacks and cyber security as a whole which helped in broadening information I acquired. Thank you.

I sincerely thank the Directorate of Criminal investigation and Communication Authority of Kenya for the provision of victims of cyber- attack incidents and cyber-attack IP addresses data required for the research of this project.

Sincere thanks to Mr. Matthew, cyber security consultant from Serianu Cyber Security Company on offering advice on how I can implement GIS in a cyber space environment. To Mr. Hassan Boneya, GIS expert working at International Center for Humanitarian Agency (ICHA), who directly impacted me with knowledge in various aspects and skills that have played key roles in my project. To you all, may you be blessed abundantly.

Abstract

Geospatial data is often perceived as only being related to maps, compasses and locations. However, the application areas of geospatial data are far wider and extend to the field of cybersecurity. Geospatial data can feed into intelligence systems, help with analysis, information sharing, and help create situational awareness. Geospatial data is very powerful and can help to prioritize cyber threats and identify critical areas of concern. This study's main aim is to investigate the integration of geospatial data in to cyber attacks monitoring and tracking.

The findings in the hotspot analysis and autocorrelation suggest that cyber attacks victims are not randomly distributed but are instead concentrated in certain areas with higher population densities and greater levels of internet connectivity as well as access to electricity. Specifically, the results indicate that central Kenya, including regions such as Kiambu, Nakuru and Nairobi, have a high cluster, which suggests a higher concentration of cyber attack victim in this regions. This could be due to a variety of factors, such as the presence of a higher density of businesses or financial institutions that are targeted by cybercriminals or a higher level of internet usage or access to technology in this region, which could make it more vulnerable to cyber threats.

Through the Ordinary least squares analysis, a range of factors was considered that may influence the incidence of cyber attacks, including total population, income, access to ICT, and access to electricity. When examining the relationship between Cyber victimization and the various socio-economic factors, it was found out that total population, access to electricity and ICT access were strongly positively correlated with the cyber-attacks incidents.

Regions with higher populations and greater economic activity, as well as better access to electricity, had higher rates of cyber attacks. Additionally, access to ICT, such as internet connectivity and access to technology, was also positively correlated with the incidence of cyber-attacks. Regions with better access to ICT tended to have higher rates of cyber attacks, possibly due to a greater number of internet users and greater use of online platforms for financial transactions and other activities targeted by cybercriminals.

As part of the analysis, spatial autocorrelation techniques were used to identify clusters and outliers of Cyber victimization across the country.The results indicated that Cyber victimization is not randomly distributed in Kenya, but instead, there are distinct patterns of clustering and outliers. These findings suggest that cyber threats are concentrated in specific regions of the country, and

targeted interventions may be necessary to address these challenges.

The identification of clusters and outliers of Cyber victimization can help policymakers and cybersecurity experts to develop more targeted interventions to address the root causes of cyber threats in specific regions. By identifying regions with high clusters of cyber attacks, policymakers can prioritize interventions to improve internet infrastructure and security measures in those areas, as well as provide cybersecurity training and resources to businesses and financial institutions operating in those regions. Similarly, addressing the unique challenges faced by outlier regions can help to prevent the spread of Cyber victimization and enhance the overall cybersecurity posture of the country.

The web application was created to serve five purposes: to create a platform in which cyber-attack incidents can be reported this solves the problem of data collection, platform to host a database to store the source IP addresses which can be used by cyber security experts in digital forensic investigations, real time tracking and monitoring in the web application, increases cyber security awareness through offering maps for visualization and data mining and lastly coordination and collaboration between cyber security agencies in Kenya.

Table of Contents

Acknowledgement.....-III-

Abstract.....-IV-

Table of Contents.....-VI-

List of Figures.....-VIII-

List of Tables.....-XI-

Acryonmyms and Abbreviations.....-X-

1 Introduction.....-13-

1.1 Background.....-13-

1.2 Motivation and problem statement.....-16-

1.3 Justification.....-17-

1.4 Research Identification.....-18-

1.4.1 Research Objectives.....-18-

1.4.2 Research Questions.....-18-

1.5 Study Outline.....-19-

1.6 Structure of the Report.....-19-

2.0 Literature Review.....-20-

2.1 Types of cyber attacks.....-24-

2.2 Gaps.....-26-

3. Study Area and Methods.....-27-

3.1 Study Area.....-27-

3.2 Data.....-28-

3.2.1 Cyber-attack IP addresses.....-29-

3.2.2 Correlation data.....-29-

3.3 Methodology……-30-

3.3.1 Processes……-31-

3.3.1.2 Normalization of IP addresses……-31-

3.3.1.3 Aggregation of the IP addresses……-31-

3.3.1.4 Density Computation and Hotspot Analysis……..-32-

3.3.1.5 Regression……32-

3.3.1.6 Web Application……-32-

3.3.1.6.1 Streamlit web framework……-32-

3.3.1.6.2 Django web framework……..-33-

4. Results……-335-

4.1 Spatial characterization of cyber attacks……-35-

4.2 Cyber attacks patterns analysis……..-37-

4.2.1 Hotspot Analysis……..-37-

4.2.2. Spatial Autocorrelation……-38-

4.2.3 Correlation Analysis……-41-

4.3 Regression……-43-

4.4 Web application……-45-

5. Discussions……-53-

6. Conclusion……-54-

7. Recommendation……..-55-

References……-56-

Appendix……-60-

List of Figures

Figure 1: WannaCry cyber-attack hacking……....17
Figure 2: Stages involved in cyber-attack……..20
Figure 3: Flowchart of framework of how GIS can be applied in cyber security…….....21
Figure 4: Study Area map……...27
Figure 5: Methodology Flowchart……..30
Figure 6: Flowchart of steps involved in Streamlit web development……...33
Figure 7: Configuration of Leaflet in Django…….33
Figure 8: Flowchart of steps in Django web framework……34
Figure 9: Cyber victim hotspots for 2020 and 2021……...35
Figure 10: Heat map on cyber attacks Source IP addresses……35
Figure 11: Pie chart showing countries with the highest cyber attacks source IP count…….37
Figure 12: Hotspots for cyber attacks victims…….38
Figure 13: Spatial Autoccorrelation report……..39
Figure 14: Moran's I cyber attack output……40
Figure 15: Cyber victims and ICT access correlation……..41
Figure 16: Cyber attack victims and electricity access correlation……..41
Figure 17: Cyber attack victims and income correlation……..42
Figure 18: Cyber attacks victims and population……..42
Figure 19: OLS analysis between independent and dependent variables……..43
Figure 20: Regression analysis of standard residuals…….44
Figure 21: Homepage of the frontend web application……..46
Figure 22: Analysis page……....46
Figure 23: Contact Page……47
Figure 24: Report page for the frontend web application……..47
Figure 25: Report pages……48
Figure 26: User login of backend web application……49
Figure 27: Cyber Attack data on the database……....49
Figure 28: Example of details of cyber attack incident……..50
Figure 29: Cyber attack victims on the database……....50
Figure 30: Add cyber attack IP form……..51

Figure 31: Add cyber attack victims form…….51
Figure 32: Cyber attack victims in Kenya…….52
Figure 33: Cyber attacks source IPs targeting Kenya……52

List of Tables

Table 1: Dataset used in the study……28
Table 2: Cyber-attack IP addresses data……..29

Acronyms and Abbreviations

1. OLS Ordinary Linear Regression
2. GIS Geospatial Information Science
3. GWR Geographical Weighted Regression
4. GEGIS Geomatics Engineering and Geospatial Information Systems
5. IDS Intrusion Detection System
6. KE-CIRT Kenya Computer Incidence Response Team
7. ISP Internet Service Provider
8. IP Internet Protocol
9. GIS Geospatial Information Science
10. CAK Communication Authority of Kenya
11. DCI Directorate of Criminal Investigation
12. SOC Security Operation Analyst
13. KRA Kenya Revenue Authority
14. IEBC Independent Electoral and Boundaries Commission

1 Introduction

1.1 Background

With the increased use and reliance on technology, negative factors such as cyber attacks are on the rise. Globally, over 4.66 billion individuals have access to internet services and connections (Cisco, 2020). According to a survey by Jumia Business Intelligence and GSMA Mobile in 2020, Kenya has the greatest internet penetration in Africa, with over 31 million individuals having access to the internet.

Kenya has become a profitable target for hackers losing around Sh21.2 billion to cyber attacks in 2020 (Standard media newspaper, 2020). This may be due to a lack of comprehensive legislation and regulatory frameworks for cyber attacks as well as limited and inefficient measures and methods used in cyber security sectors to prevent cyber attacks from occurring (Serianu annual report, 2020). In addition, the lack of public awareness on how cyber attacks occur and how easily information can be hacked into online by third parties add to the targeting of local citizens and increase in cyber attacks (Communication Authority of Kenya annual cybercrime report, 2020).

To address the rising cases of cybercrime in Kenya, the government enacted the Computer Misuse and Cybercrime Act in 2018. The Act seeks to provide a legal framework for the prevention, investigation, and prosecution of cybercrime. The Act defines cybercrime as any offense such as unauthorized access to a computer system, unauthorized interception and interference with one's private information, computer forgery and child pornography committed through a computer system or any other electronic communication device (Cybercrime Act, 2018).

The Act also provides for the establishment of a National Computer and Cybercrimes Coordination Committee, which is tasked with coordinating the fight against cybercrime in Kenya. According to the Computer Misuse and Cybercrime Act, penalties for cybercrime offenses, are fines of up to 10 million Kenyan shillings or imprisonment for up to 10 years according to the offense committed.

Example, for phishing, the penalty for the commission of this offence is a fine not exceeding three hundred thousand shillings, imprisonment for a term not exceeding three years or both according to the Computer Misuse and Cybercrime Act (2018). The Act also provides for the forfeiture of assets, money or property used in the commission of cybercrime offenses, this means that what the hacker gains such as money is recovered and the victims are compensated.

According to the Communication Authority of Kenya, the most prevalent types of cyber attacks now confronting the country are: Malware, spam, botnets, defacements, phishing and financial fraud (Communication Authority of Kenya annual cybercrime report, 2020).

Between July and September of 2020, the Kenya Computer Incident Response Team - CoordinationCentre (National KE-CIRT/CC) detected 35.2 million cyber attacks occurrences. According to Kenya's Communications Authority statistics, this represents a 52.9 percent increase over the previous quarter's 13.9 million cyber attacks incidents. The increasing detection of cyber attackswas related to the shift to remote working and greater use of e-commerce in reaction to the Covid-19 outbreak.

Most cyber attacks go unreported, and many more go undetected. As a result, the scope of cybercrime is significantly broader than was previously thought (Detica, 2013a). Through the worldwide networked communications, the cyberspace decentralizes criminal conduct around the world while simultaneously, generates new social places in which this type of activity can hide (Morely & Robbins, 1995). Traditional modes of surveillance, such as CCTV cameras or intelligence analysts, are incapable of monitoring and filtering huge amounts of data, which might reach up to 900 million occurrences per day on a single network (Detica,2013a). Thus comes the need of geospatial layer data to be able to visualize large data.

Research studies on the integration between geospatial science and cyber attacks are limited and have not been explored fully. Merien et al. proposed an entropy-based model for representing cyber attacks as a pathway from source to target (Merien, Bellekens, & Brossets, 2018). Their findings show a pattern among cyber attacks origins that might be utilized to categorize internet cyber attacks.

Chen at el. proposed predictability measures using a probability matrix to characterize cyber attack patterns (Chen, & Huang, 2015). They demonstrated a correlation between large- scale cyber attacks and their predictabilities. Hu et al. proposed a Geographical Internet Protocol (GEO IP)-based analysis (Hu, Baynard, & Hu. (2015). GEO-IP was applied in their study to determine the location of cyber-attack origin and advanced spatial statistical analysis was used toinvestigate cyber-attack patterns.

Many corporations and institutions that constitute the large part that is targeted by hackers have yet to take into consideration using geospatial data as one solution to monitor and track cyber

attacks. Just as businesses are beginning to incorporate geo-analytics into their operations to gain better business insight and speed decision-making,their security IT departments must geo-enrich their threat assessment data by integrating location analytics to improve the analysis and better deal with outside threats.

Geospatial data can be used to display locations of interest and emerging network traffic patterns, predict cyber attack growth trends, and identify affected areas and the growth of specific types of cyber risks. In terms of spatial analysis, there is evident need for a deeper knowledge about cyber attacks. This can provide not just a unique view on the origins and victims of cyber attacks, but also possible concerns for regions that are the target of cyber attacks.

1.2 Motivation and problem statement

Cyber attacks are a continuing issue globally and Kenya. Thus, varies methods and solutions should be applied to help monitor and track the cyber attacks. Cyber attacks on firms throughout the world have increased by 29% by the year 2020. This growing threat in Africa is being driven by an increase in internet users, which creates a wider pool of targets for criminals.

The surge has seen the need to find solutions and countermeasures using statistical methods and geospatial science. As a result, geospatial science is a required need for cyberattack monitoring and tracking (Koike, Ohno, & Koizumi, 2005). Cyber attack patterns and growth over time are not apparent until data is represented visually in maps and statistical models are evaluated (Du, & Yang, 2013). Geospatial science can be used in real time event monitoring for cyber attacks. This gives the security experts a visual on where cyber- attacks are concentrated.

Moreover**,** previous researches have revealed little about cyber attacks, especially in its spatial analysis, in relation to the physical geographic location of the attacks (Lin, Chiu, & Lee, 2013)**.** Most studies in the literature use the cyber attacks geocoded IP addresses for analysis, making the analysis difficult when adding covariates to find the relationships between attacks and other variables rather than geolocations (Du, & Yang, 2013)**.** In addition**,** previous studies have dealt with only trivial source-target relationships in cyber attacks.

Many organizations and institutions, example, universities in Kenya have not yet embraced the use of geospatial science in cyber security monitoring. Cyber security experts, in Kenya, are yet to see or acknowledge the use of geospatial layer data in their daily use in monitoring cyber attacks. Moreover, the research studies on cyber attacks monitoring and GIS are limited, scarce and its aspects have not been fully explored.

1.3 Justification

High-profile incidents like the Equifax data breach and WannaCry ransomware attack have ensured that cybersecurity is a top priority for organizations of all kinds (GIS Lounge, 2021). Retailers that process payments, hospitals that are entrusted with confidential patient information and the local government conducting elections are all highly conscious of the threats posed by phishing scams, denial-of-service attacks and malware (Zhuang, Zamir & Liang, 2021). Systems detect more cyber attacks with every passing second around the world. GIS helps us to understand the scale of this problem and detect meaningful trends. Mapping cyber attacks in real time reveals just how common such incidents are and how important it is for organizations to have updated countermeasures in place.

Figure 1: WannaCry cyber-attack hacking

1.4 Research Identification

The main aim of this research is to investigate the integration of geospatial data in to cyber-attacks monitoring and tracking.

This is achievable through the following specific objectives:

1.4.1 Research Objectives

1) To identify and visualize where cyber attacks originate and model its growth and show its patterns.
2) To identify the dependency factors that correlate to cyber attacks such as demographic factors, social and economic factors.
3) To create a web application of real time monitoring and tracking for cyber attack reporting that includes a spatial database.

1.4.2 Research Questions

The following questions are articulated with respect to above-mentioned objectives:

1. Where are cyber attacks occurring in the research area?
2. What are the intended target locations?
3. Where do the attacks originate (source)?
4. What are the types of cyber attacks taking place?
5. What are the trends (patterns) of the cyber attacks?
6. How has the pattern of cyber attacks changed over time?
7. What are the dependencies that are associated with cyber attacks?
8. What types of cyberattack are most prevalent, and where do they arise from (location)?
9. What time of day do they occur most frequently?
10. What are methods are used for the cyber attacks?

1.5 Study Outline

Whilst the advancing in technology has immediate negative problems such as cyber attacks, its impact and monitoring research expansion is still limited as more and more attacks occur. An example in Kenya, on average, businesses deal with almost 540 weekly cyber attacks than theirpeers across the globe, a study by Check Point Software Technologies cyber security firm (Cisco, 2020)

The aim of this study is to investigate how Geospatial Information Science can be integrated and used as a counter measure against cyber attacks by mapping clusters of where the attacks originate, show their patterns and dependency factors e.g. economic or political factors. The scope of this study will include cyber attacks victims' data from the Directorate of Criminal Investigations. The data collected and analyzed will be for a period of 2 years that is 2020 to 2022. The study area will encompass the whole country.

1.6 Structure of the Report

The report is outlined as follows:

Chapter 1 presents the background information, a problem statement, study objectives and justification of the study. The chapter underlines why it is becoming increasingly important to identify and examine the link between geospatial analysis and cyber-attack monitoring and tracking. Chapter 2 introduces an analytical summary of the current level of knowledge regarding geospatial analysis and cyber-attack tracking. In examining the context of cyber attacksincidences, it emphasizes the significance of these variables and associated arguments. Chapter 3describes how data was acquired, what observation units were utilized, and how the data was processed It explains how the theoretical framework informs multi-contextual crime analysis Chapter 4 describes the study results. Chapter 5 interprets the study results and how it connects with the study objectives. It also relates present work to previous studies. Chapter 6 presents the conclusion to the study from the findings any recommendations for future research in this field.

2.0 Literature Review

In a society driven increasingly by big data, social networks, online transactions, information saved via the internet, and automated operations done using IT systems, information security anddata privacy are constantly at risk. Cybercrime is expanding in terms of the number of attacksand the degree of damage inflicted to its victims as new tools and approaches by hackers are developed (Lin, Chiu, & Lee, 2013).

Attackers aim to jeopardize the confidentiality, availability, and integrity of information by creating new ways of gaining unauthorized access to networks, programs, and data (Andrea, 2016). Their targets range from individuals to small or medium-sized businesses, and even commercial giants.Lockheed Martin (2011) explained the cycle of cyber attacks, the attack chain starts at the recon stage (reconnaissance stage). This is where the hacker screens the target computer for any vulnerabilities. The second stage is the weaponization stage where the hacker produces malware to exploit the vulnerability found in the first stage.

During the third stage, that is the delivery stage, the intruder sends the malware via fraudulent emails or any other medium. The malware then starts executing on the target system during the exploit stage. The malware installs a backdoor accessible to the attacker during the installation stage. The hacker gains persistent direct access to the victim's network during the call back stage. The hacker initiates overall goal activities, such as data corruption, data destruction, or data theft, during the last stage that is the persist stage (Lockheed Martin, 2011).

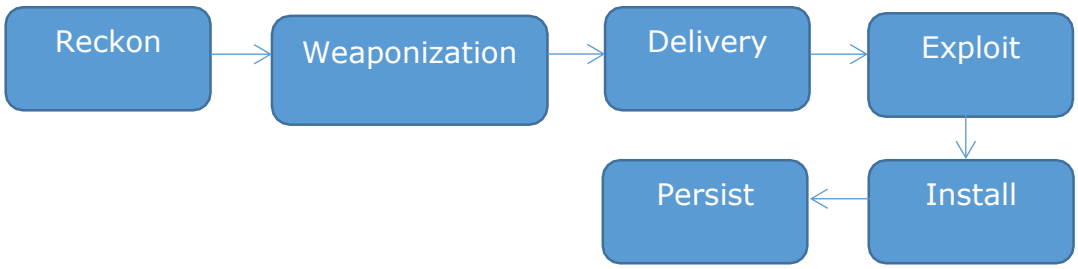


Figure 2: Stages involved in cyber-attack by Lockeheed Martin (2011)

By using an Intrusion Detection System (IDS) which is a monitoring system that detects suspicious activities and generates alerts when they are detected, a security operations center (SOC) analyst or incident responder can investigate the issue and take the appropriate actions (Serianu Cyber Security Company annual report, 2020). The alerts are the cyber attacks IP addresses of the hackers who are at the weaponization stage of the cyber attacks cycle. The

system allows immediate and efficient way to stop the hackers from accessing the unauthorized information (Cisco, 2011).

Jiang and Orm (2000) introduced a framework which shows how GIS can be applied and integrated into a cyber security environment.

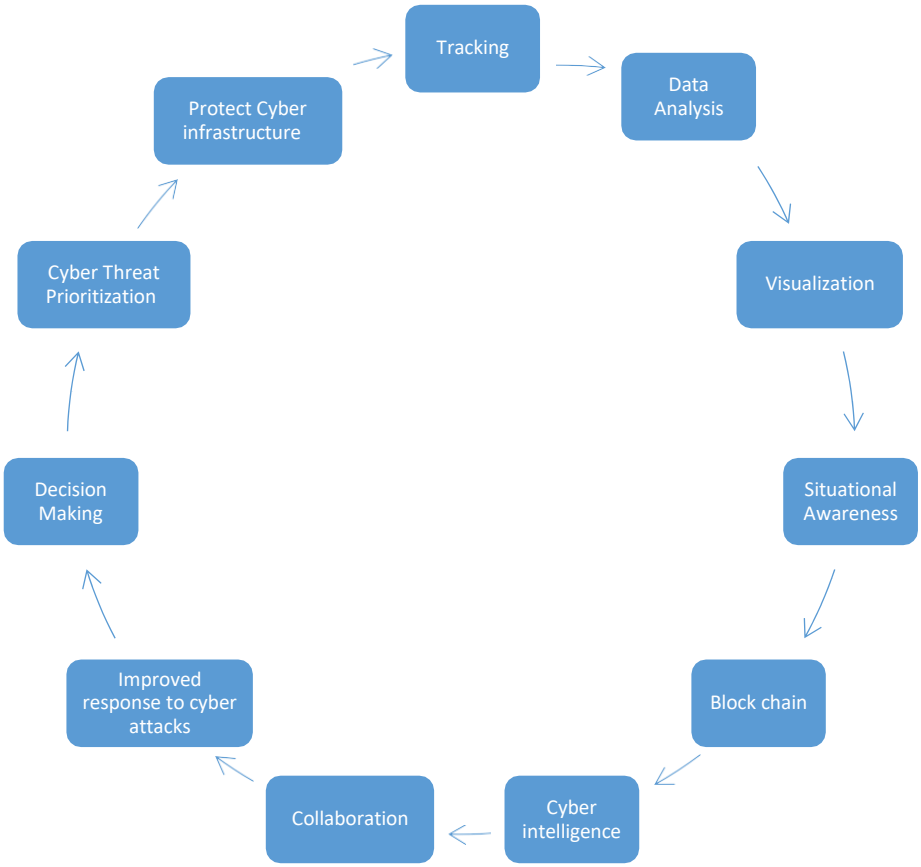


Figure 3: Flowchart of framework of how GIS can be applied in cyber security

- Tracking: Using Location intelligence, connections can be shown and links found. Geospatial data can help visualize trends, patterns, hotspot areas and conduct further analysis such as regression to understand the data.
- Data analysis: Establish correlations between objects and events. An example, data can be presented in tables showing various locations of cyberattack. When this data is plotted into a map or grouped, aggregated and processed by geographic region, then the prevalence of a specific cyberattack in certain location can be seen.

- Visualization: graphical representation e.g. maps graphs e.t.c
- Situational awareness: perceives the relevant information by being able to access it, recognize it, and comprehend the information and prediction of possible outcomes. Situational awareness is critical for cyber attacks to gain an understanding of the environment, to support decision making.
- Cyber intelligence: Raw data is processed to form information which can be processed and analyzed together with other data sources to create intelligence. An example, time- lapse of attacks may show decline or increase in a cyber attack overtime and may relate to a certain event.
- Collaboration: reveal patterns to determine dependencies and correlation between various factors e.g. economic factors, political factors and time factors
- Decision making: GIS acts as a decision making support as with the analysis and visualization provided the relevant authorities can make informed decisions.
- Cyber threat prioritization: through collaboration of different agencies using geospatial Science cyber attacks and ways to mitigate them becomes a priority.

Merien et al. proposed an entropy-based model for representing cyber attacks as a pathway from source to target (Merien, Bellekens, & Brossets, 2018). Their findings show a pattern among cyber attacks origins that might be utilized to categorize internet cyber attacks. Chen et al. proposed predictability measures using a probability matrix to characterize cyberattack patterns (Chen, & Huang, 2015). They demonstrated a correlation between large- scale cyber attacks and their predictabilities. Hu et al. proposed a geographical Internet Protocol (GEO IP)-based analysis (Hu, Baynard, & Hu, 2015). GEO-IP was applied in their study to determine the location of cyber-attack origin and advanced spatial statistical analysis was used toinvestigate cyber-attack patterns.

Lin et al. demonstrated a method for detecting malicious URLs (Uniform ResourceLocators) using lexical analysis (Lin, Chiu, & Lee, 2013). According to the findings in big data, employing a two-step filtering on the data, 90 percent of malicious URLs could be identified. Feroz and Mengel proposed yet another URL based classification technique (Feroz, & Mengel, 2015). They used URL clusters to classify them onto various categories, which was used to predict cyber attacks by the classification model.

Du and Yang took a slightly different strategy to identifying and classifying cyber attacks, they grouped sources into coordinated teams (Du, & Yang, 2013). When compared to trivial statistics methodologies, their approach provides an additional approach on high impact attacks. Koike et al. focused on the visualization of cyber attacks using a 2D IP address matrix (Koike, Ohno, & Koizumi, 2005). Their hypothesis was that the given visualization framework could boost detection algorithms.

Xiu and Li (2014) proposed the design of cybercrime spatial analysis system which has 3 aspects(Xiu & Li, 2014): Building a spatial database for cybercrime with IP addresses using web searches and IP tracking to achieve automatic data loading and real-time updating, second by creating a network resource spatial database and a cybercrime spatial database, it is possible to archive fast tracking and spatial orientation for complex IPs like proxies, which addresses the problem of the authenticity and credibility of the IPs that will be tracked and lastly by establishing a spatial analysis model that can be used to predict cyber attacks.

Research by Octavian Bordeanu in “How’s the weather like in cyberspace” (Octavian, 2018) proposed the use of Self Organizing Map (SOM) which is a neural network used to identify clusters and hotspots within the topological structure of cyber-attack data. He hypothesized that physical mapping of cybercrime is visually informative, based on the idea that visual representation of cybercrime data can provide valuable insights but have little to offer in terms of understanding the underlying structures and patterns in the data they are built on. By using an artificial neural network with unsupervised learning preserves, mimics and superimposes the topological structure of the data on a two dimensional grid.

There are some successful applications of GIS for effective monitoring, analysis, quick orientation and making cybercrime maps and applications in developed countries. An example ofa popular real-time cyber map is Kaspersky map. In Britain, Germany and United States, the IP spatial database has been applied in practice to seek clues and for tracking hackers. Experts say that by 2030 most transactions and interactions will be done in the cyber environment, that is online, thus measures should be identified and used to countermeasure cyber attacks growth that will be witnessed then.

2.1 Types of cyber attacks.

1. Phishing Attacks:

Phishing attacks are a common type of cyber attack in Kenya. Phishing attacks involve tricking individuals into providing their sensitive information, such as passwords and credit card details, by posing as legitimate entities (Communication Authority of Kenya annual report, 2020). According to a study conducted by Serwaa-Boateng et al. (2019), phishing attacks were the most prevalent cybercrime in Kenya, accounting for 48% of all reported cases. The study also revealed that phishing attacks were the most prevalent type of cybercrime against financial institutions in Kenya, with 60% of all reported cases.

Phishing attacks in Kenya have targeted both individuals and businesses. For instance, in 2018, the Kenya Revenue Authority (KRA) issued a warning to the public regarding phishing attacks that impersonated the tax authority (Oduor, 2018). In the same year, Safaricom, one of Kenya's leading mobile network operators, warned its customers of a phishing scam targeting M-PESA users (Mumbi, 2018). The scam involved sending fraudulent messages to customers and requesting them to provide their M-PESA PIN.

2. Malware Attacks:

Malware attacks are another common type of cyber attack in Kenya. These attacks involve the use of malicious software to gain unauthorized access to a victim's computer system (Communication Authority of Kenya annual report, 2020). The malware can be used to steal sensitive information, damage the victim's system, or spy on the victim. According to a report by Kaspersky Lab (2020), Kenya ranked sixth in Africa for malware attacks in 2019.

Malware attacks in Kenya have targeted both individuals and businesses. For instance, in 2019, a malware attack targeted the National Hospital Insurance Fund (NHIF), one of Kenya's largest state-owned corporations (Kinyanjui, 2019). The attack led to the loss of sensitive data belonging to the corporation's members.

3. Botnets Attacks:

Botnets attacks involve overwhelming a victim's computer system with traffic to the point where it becomes inaccessible to legitimate users (Communication Authority of Kenya annual report, 2020). These attacks can be used to disrupt the operations of businesses or websites. According to

a study by Serwaa-Boateng et al. (2019), botnets attacks accounted for 17% of all reported cybercrimes in Kenya.

Botnets attacks in Kenya have targeted both individuals and businesses. For instance, in 2018, a botnets attack targeted the Independent Electoral and Boundaries Commission (IEBC) during the repeat presidential elections (Standard media newspaper, 2018). The attack led to the disruption of the election process and the subsequent cancellation of the elections.

4. Spams

Spams, which are unsolicited emails or messages, are used by fraudsters to lure unsuspecting victims into parting with their money or personal information (Communication Authority of Kenya annual report, 2020). In Kenya, spam messages often promise huge financial gains or job opportunities, and they ask the recipient to provide their personal information or to pay a small fee to access the supposed reward. Once the fraudster gets hold of the victim's information, they can use it to commit identity theft or other financial crimes.

5. Financial fraud

Financial fraud in Kenya takes many forms, including phishing, fake investment schemes, pyramid schemes, and mobile money fraud. One of the most common forms of financial fraud is phishing, which involves tricking victims into divulging their personal information by posing as a legitimate organization such as a bank or a government agency (Communication Authority of Kenya annual report, 2020). A notable case in Kenya involved a pyramid scheme called Ekeza Sacco, which saw the owners of the scheme disappear with over 1 billion Kenyan shillings in investors' money in the year 2020 (Standard media newspaper, 2020).

2.2 Gaps

Research studies on how GIS can be integrated in to cyber attacks monitoring and tracking are limited, scarce and its aspects have not yet been fully explored. During the analysis and visualization of the data provided, some challenges and gaps arose. These included:

a. Lack of standardized data: Data used in monitoring cyber attacks may come from various sources, and may be presented in different formats. This can make it challenging to combine the data from different sources and extract useful information.

b. Complexity of cyber attacks: Cyber attacks can be complex and involve multiple stages, making it difficult to identify the precise location of the attacker or the target. Additionally, hackers may use proxy addresses which make tracking them down hard. This can make it challenging to accurately map the attack using geospatial analysis.

c. Limited availability of data: This is due to the low rate in which the cyber attacks are reported as most are ignored. This can limit the effectiveness of using geospatial analysis to monitor cyber attacks and further analysis such as adding time factor example when do cyber attacks mostly occur during the day or night.

d. Legal and privacy concerns: The use of geospatial analysis in monitoring cyber attacks raises legal and privacy concerns. For instance, the collection and use of location data may be subject to legal restrictions, and the use of geospatial data may raise privacy concerns for individuals.

The building of the web application addresses two of the challenges. That is, building of web application that includes a spatial database solves the challenges on standardization of data where there is a common reporting page that victims use to report any incident, after which it is stored in a spatial database. This also solves the challenge of limited availability of data as with a web application it can reach many people as it removes the aspect of having to commute to a police station. The analysis focused on the data provided with future improvement to be done with the use of data in the web application built.

3. Study Area and Methods

3.1 Study Area

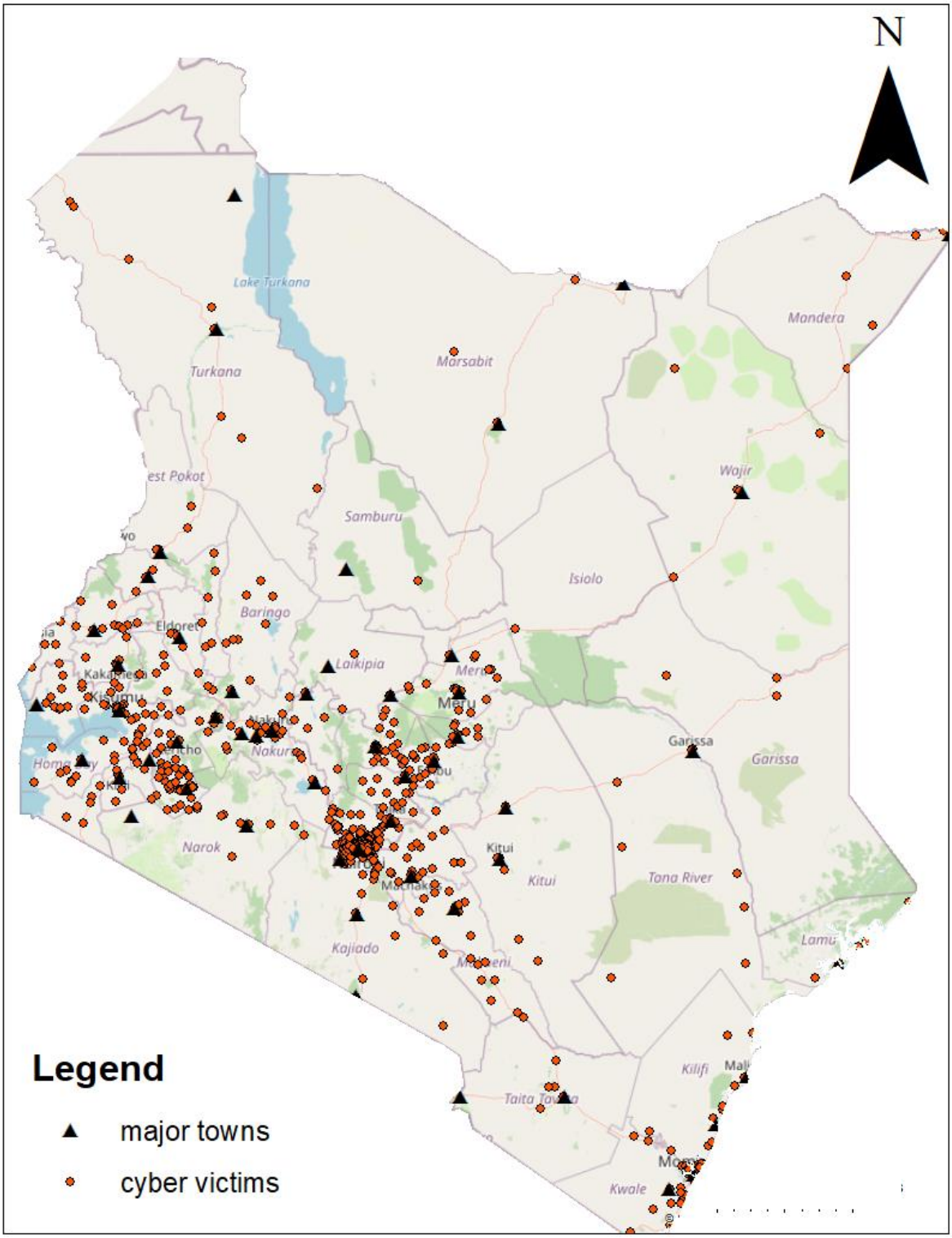


Figure 4: Study area map

Kenya is a country located in East Africa, bordered by Tanzania to the south, Uganda to the west, South Sudan to the northwest, Ethiopia to the north, Somalia to the east, and the Indian Ocean to the southeast. Kenya's area is approximately 580,367 square kilometers. The country is divided into 47 counties with Nairobi being the capital city of Kenya. Its spatial extent are coordinates from 33° 54' 33.12" E to 41° 53' 57.48" E and from 4° 37' 20.64" S to 4° 40' 42.52" S. Kenya's total area is 98% land and 2% water. The highest point in Kenya is Mount Kenya, which reaches an elevation of 5,199 meters (17,057 feet) above sea level. According to the 2019 census, the population of Kenya is around 51,629,122 citizens. Cyber attack victims' data was a total of 2000 points for both 2020 and 2021. In Figure 4 above, it shows the distribution of the cyber attack victims and major concentration can be seen around major towns such as Nairobi and Kiambu.

3.2 Data

Table 1: Dataset used in the study

Data	Source	Description
Source IP addresses	Communication Authority of Kenya	Data from CAK includes source IPs from various organization that hackers have attacked those institutions.
Cyber-attack victims	DCI	Reported incidents of cyber attacks 2020 and 2021
Population, income, access to ICT, internet access per county	Kenya National Bureau of Statistics	Rate of income, population access to ICT and internet access per county
Kenya shapefile	Kenya National Bureau of Standards	Kenya administration boundaries shapefile
World shapefile	ESRI	Global Countries shapefile.

3.2.1 Cyber-attack IP addresses

Cyber-attack IP addresses data showed source IPs which give the hackers location once geocoded and time of attack that an intrusion detection system was able to capture the attack. This data was used in the creation in spatial database. The IP addresses were obtained from an Intrusion Detection System in which once stored in a database more forensic investigation by a cyber security expert can be done. Below shows an example of the data collected:

Table 2: cyber-attack IP addresses data

IP	Country	City	Region Name	Time Zone	Latitude	Longitude	Type of attack
85.215.117.241	Germany	Berlin	Berlin	Europe/Berlin	51.2993	9.491	malware
74.82.195.39	United States	Los Angeles	California	America/ LA	34.0549	-118.2578	malware
77.40.0.17	Russia	Yoshkar-Ola	Mariy-El Republ	Europe/Moscow	56.6388	47.8908	malware
61.222.202.235	Taiwan	Taipei	Taipei City	Asia/Taipei	23.5	121.000	malware
196.188.106.242	Ethiopia	Addis Abbaba	Anhui	Africa/AddisAbbaba	8.000	38.0000	phishing
179.61.219.218	United States	Chicago	Illinois	America/Chicago	41.8483	-87.6517	phishing
122.117.51.33	Taiwan	Chang-hua	Changhua	Asia/Taipei	24.0709	120.5419	phishing
116.10.141.70	China	Shanghai	Guangxi	Asia/Shanghai	22.8137	108.3146	phishing
92.63.97.240	Russia	Moscow	Moscow	Europe/Mosco	55.7386	37.6068	financial fraud
192.241.211.143	United States	San Francisco	California	America/LA	37.7353	-122.3732	Financial fraud
37.32.10.93	Iran	Tehran	Tehran	Asia/Tehran	35.698	51.4115	Financial fraud
23.99.229.218	United States	Des Moines	Iowa	America/Chicago	41.6006	-93.6112	Financial fraud
167.94.145.21	United States	Chicago	Iowa	America/Chicago	37.751	-97.822	spam
159.65.150.25	India	Bengaluru	Karnataka	Asia/Kolkata	12.9721	77.5933	spam
195.19.4.22	Russia	Moscow	Moscow	Europe/Mosco	55.7527	37.6172	spam
192.241.200.47	United States	San Francisco	California	America/LA	37.7353	-122.3732	spam
189.78.95.214	Brazil	Taboao da Serra	Sao Paulo	America/California	-23.6116	-46.7827	botnets
128.199.129.68	Singapore	Singapore	Singapore	Asia/Singapore	1.314	103.6839	botnets
197.5.145.30	Tunisia	Tunis	Tunis	Africa/Tunis	34.000	9.00000	botnets

3.2.2 Correlation data

In order to identify the dependency factors that correlate to cyber attacks such as demographic factors, social and economic factors; population, ICT access, income, electricity access, and internet users are taken into consideration. Larger populations tend to have more potential victims for cyber attacks due to the sheer number of people who use technology in their daily lives (Shaw & Yu, 2010). Greater access to ICT infrastructure, such as high-speed internet and cyber cafes, make people are more susceptible to cyber attacks due to the increased use of technology in their societies. People with higher incomes tend to have more valuable assets and are more likely to use technology to manage their finances and personal information, making them more attractive targets for cyber criminals (Luoli, 2020) The widespread access to electricity is more likely to have higher level of technology adoption therefore, may be more vulnerable to cyber attacks. Internet users provide a measure of the overall level of technological literacy and adoption within a country plus the number of people who are able to access the internet, in which has potential for cyber attack victims .

3.3 Methodology

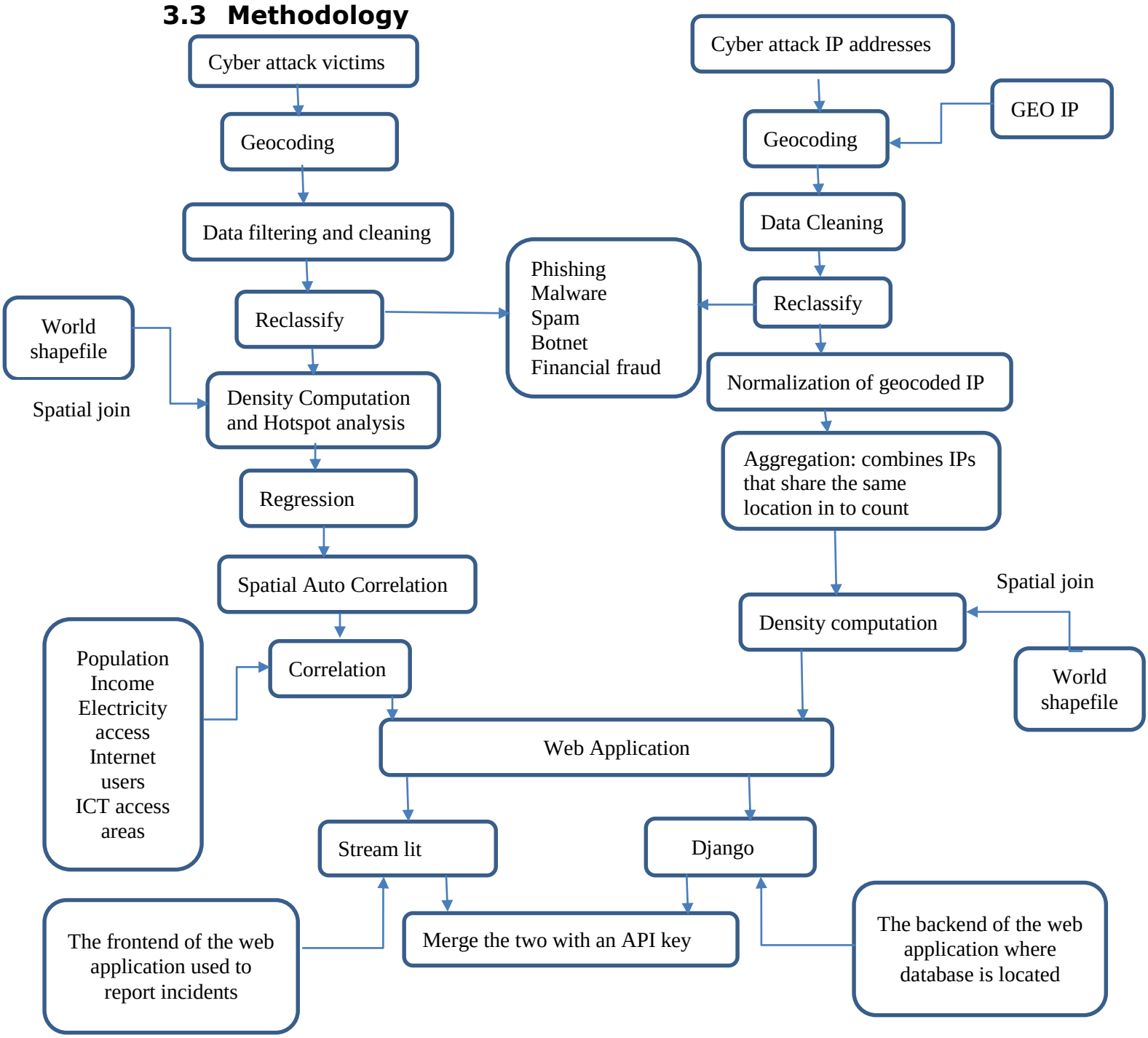


Figure 5: Methodology Flowchart

3.3.1 Processes

3.3.1.1 Geocoding

Once the data was collected, in order to determine the coordinates of both the cyber-attack victims and cyber-attack source IPs addresses, they were geocoded. The cyber attacks victims were geocoded using a python script which ran a google geocoding API in a Jupyter notebook and resulted in over 95% match for each of the incidents locations. Geocoding IP addresses using GEOIP is the process of determining the geographic location of an IP address (Hu & Baynard, 2015). By geographically mapping the IP address, it provides you with location information such as the country, state, city,zip code, latitude/longitude, ISP, area code, and other information using a python script and an IP geocoding API key. Example of these results can be seen in Table 2.

3.3.1.2 Normalization of IP addresses

Normalization of IP addresses are done for compact representation in a database. This is done for two main reasons (Hu & Baynard, 2015):

a) Normalizing IP addresses ensures that they are stored and compared consistently in a database. For example, without normalization, the same IP address could be stored as "192.168.1.1" in one record and "192.168.001.001" in another record, which could cause issues when querying or comparing records.

b) When querying a database for IP addresses, it is often faster to search for a normalized representation of the address than to search for the original string. This is because normalized representations can be indexed and searched more efficiently than strings.

Using the GEOIP software, normalization of the IP addresses can be achieved.

3.3.1.3 Aggregation of the IP addresses

Aggregation combines alerts that share the same root causes, e.g., originated from same source IP or attack the same target. Aggregation is done using similarity based clustering. Due to attacks occurring many times a day e.g. you can find the health sector such as a hospital has been attacked 700 times in a day. Thus using data from two years generates more than million cyber attacks. Aggregation narrows down the data for easy mapping and storage in a database (Buczak & Guven, 2016). Simplification of the process included, identifying the cluster centroids, reassigning the points based on minimum distance and identifying the new cluster centroids.

3.3.1.4 Density Computation and Hotspot Analysis

This explains the spatial patterns of cyber attacks. It is also performed to see whether the observations are random, uniform, or clustered (Janelle& Hodge, 2000). Density values were calculated using the algorithm from the Kernel Density tool in ArcGIS Pro. If the input features are points, the distances will be measured using the geodesic method. Ifthe input features are lines, the distances will be measured using the planar method (Kellerman, 2002). Hotspot analysis determines the confidence of hotspots and cold spots of the data obtained example 99% hotspot confidence.

3.3.1.5 Regression

Multiple linear regression models are widely known for their application in numerous research fields, including economics, demography, transportation, health, environmental sciences, etc. They enable analysts to establish the association between or among the tested variables along a straight line. It uses two or more explanatory variables and determines which of them are statistically important predictors of a given dependent variable. However, the data in independent variables should not show multicollinearity, and the Variance Inflation Factor (VIF) method should be applied and checked when values are lower than 7.5. Ordinary Least Squares (OLS) regression method was used in this study.

3.3.1.6 Web Application

Django and Streamlit web frameworks were used to build the web application due to various advantages both hold. The Django framework supports database connection such as PostgreSQL and has a great set of libraries such as scikit-learn which is designed to manage website content. Streamlit has the advantage of being free and open source with the provision of an API to build a web application.

3.3.1.6.1 Streamlit web framework

Streamlit is an open source python based framework for developing and deploying interactive data science dashboards, web apps and machine learning models. For this research Streamlit acts as the frontend of the web application. Frontend web development is the development of

the graphical user interface of a website. Here, the public can see the analysis and report any cyber attack incident. The report page is linked to a database in which the back end users can be able to see the report made. Coding was done in a VSCoduim notebook in a Linux environment Its processes include:

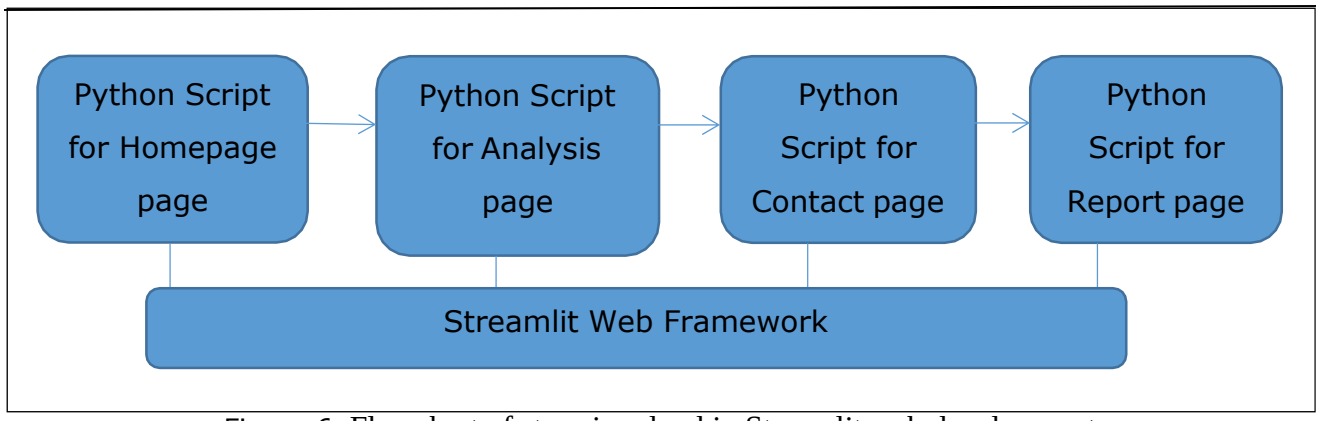


Figure 6: Flowchart of steps involved in Streamlit web development

3.3.1.6.2 Django web framework

Django is a free and open-source, Python-based web framework that follows the model– template–views architectural pattern. For this research Django acts as the backend of the web application. Backend development means working on server-side software, which focuses on everything you can't see on a website such as database administration (Zhuang & Zamir, 2021). Coding was done in a VSCoduim notebook in a Linux environment.

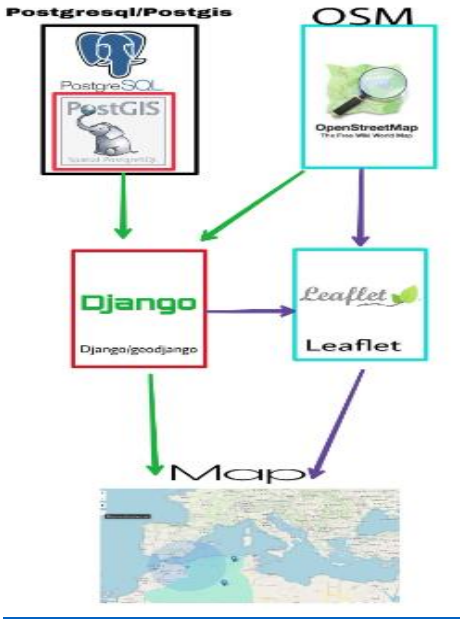


Figure 7: Configuration of Leaflet in Django

Figure 8: Flowchart of steps in Django web framework

Flowchart shows the basics steps in the building of the backend framework in Django (Andreea, 2016)

4. Results

4.1 Spatial characterization of cyber attacks

Cyber attack victims recorded for the years 2020 and 2021 were 922 and 1061 cases respectively for the period of January to December. The Kernel density tool in ESRI ArcGIS was used to show crime hotspots for varying years.

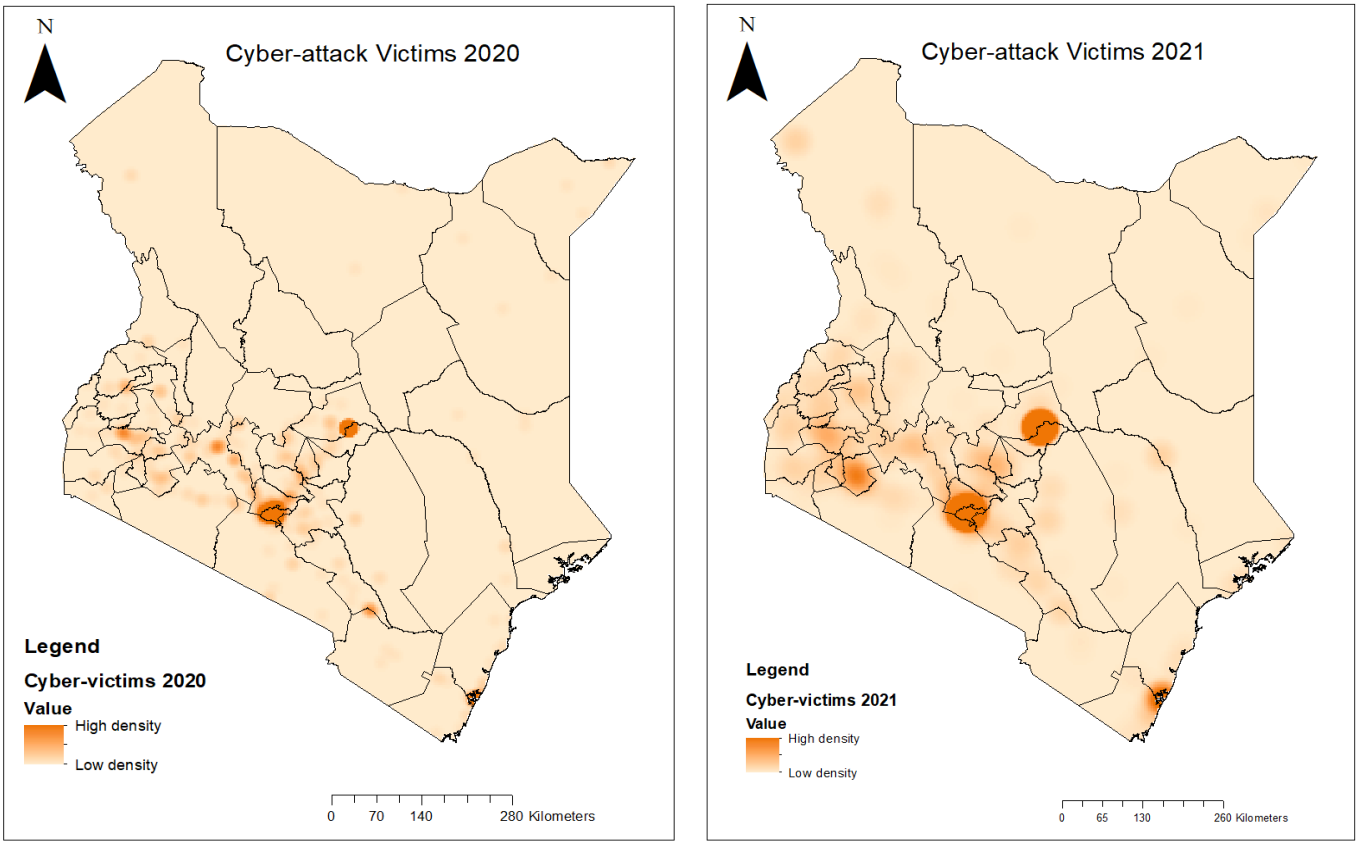


Figure 9: Cyber victim hotspots for 2020 and 2021

A 15.9% increase in cyber attacks victims was recorded in 2021 from 2020. Increased recording of cyber attacks in 2021 can be attributed to the Covid-19 pandemic that resulted in the shutting down of many businesses and also the growth of online activities such as online transactions. This shows that a dependency factor such as the Covid 19 pandemic which may not be a direct factor but it created an environment where cyber attacks became more prevalent in 2021 (Luoli, 2021).

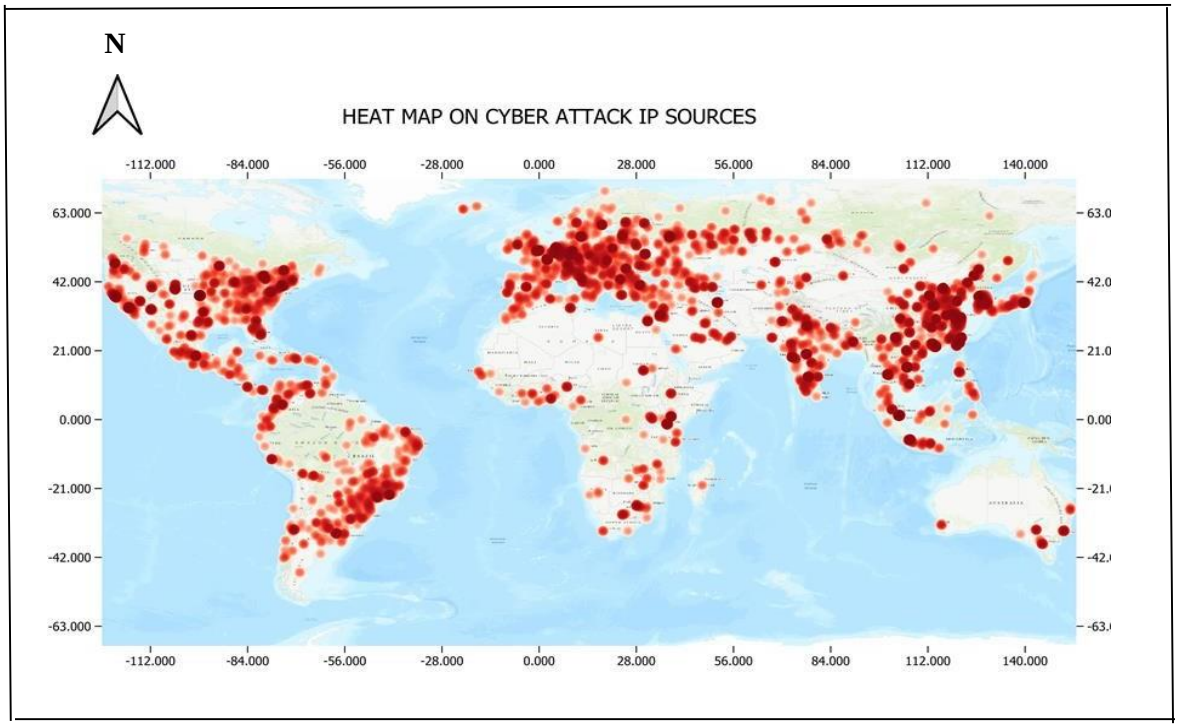


Figure 10: Heat map on cyber attacks Source IP addresses

The National KE-CIRT/CC detected 35.2 million cyber attacks in 2020 and the source IPs addresses of many of them were international. This number increased by 20% in 2021 due to increase in use and reliance on technology. In Figure 10, its shows that the European continent recorded the highest number of counts of cyber attacks directed to Kenya followed by Asian continent and North and South America continent. Russia had the largest number of cyber attacks source IPs targeting the Kenyan population of around 8000 cases followed by China with 5479 cases and the USA with 3990 cases of cyber attack source IPs according to the data obtained.

The Communication Authority of Kenya usually receives reports on cyber attacks source IPs that target the Kenyan market especially if they are international IP addresses. They then report to the respective countries authority about the IP addresses in their countries which are targeting Kenya and necessitate them to take action on the perpetrators. The heat map can provide a visualization of the countries that often target Kenya and support decision making and collaboration among countries and cyber security agencies in the world.

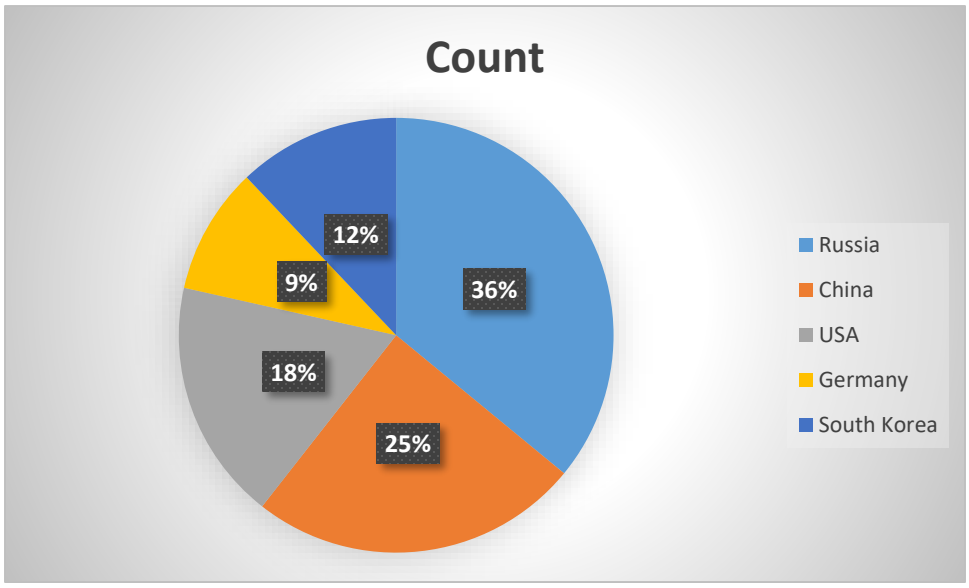


Figure 11: Pie chart showing countries with the highest cyber attacks source IP count

4.2 Cyber attacks patterns analysis

4.2.1 Hotspot Analysis

Figure 12 below displays a map of cyber attack hotspots in Kenya. From the map we can conclude that there are several hotspots where cybercrime incidents are more prevalent than in other areas. These hotspots may be influenced by a variety of factors, such as population, income, electricity and Technology access amongst others.

For example, counties in the central and western part of Kenya such Nairobi, Kiambu, Kisumu, Kakamega among others are hotspot areas where there are high levels of cyber attacks victims. This hotspot may be influenced by the high concentration of financial institutions and businesses that operate in this area, as well as the large number of tourists and visitors who may be more vulnerable to cyber attacks, large number of young people who reside in these areas, as well as the widespread use of social media and other online platforms.

Furthermore, the study provides valuable insights into the spatial distribution of cyber attacks, which may be useful for informing the development of more effective and targeted prevention and intervention efforts in the future.

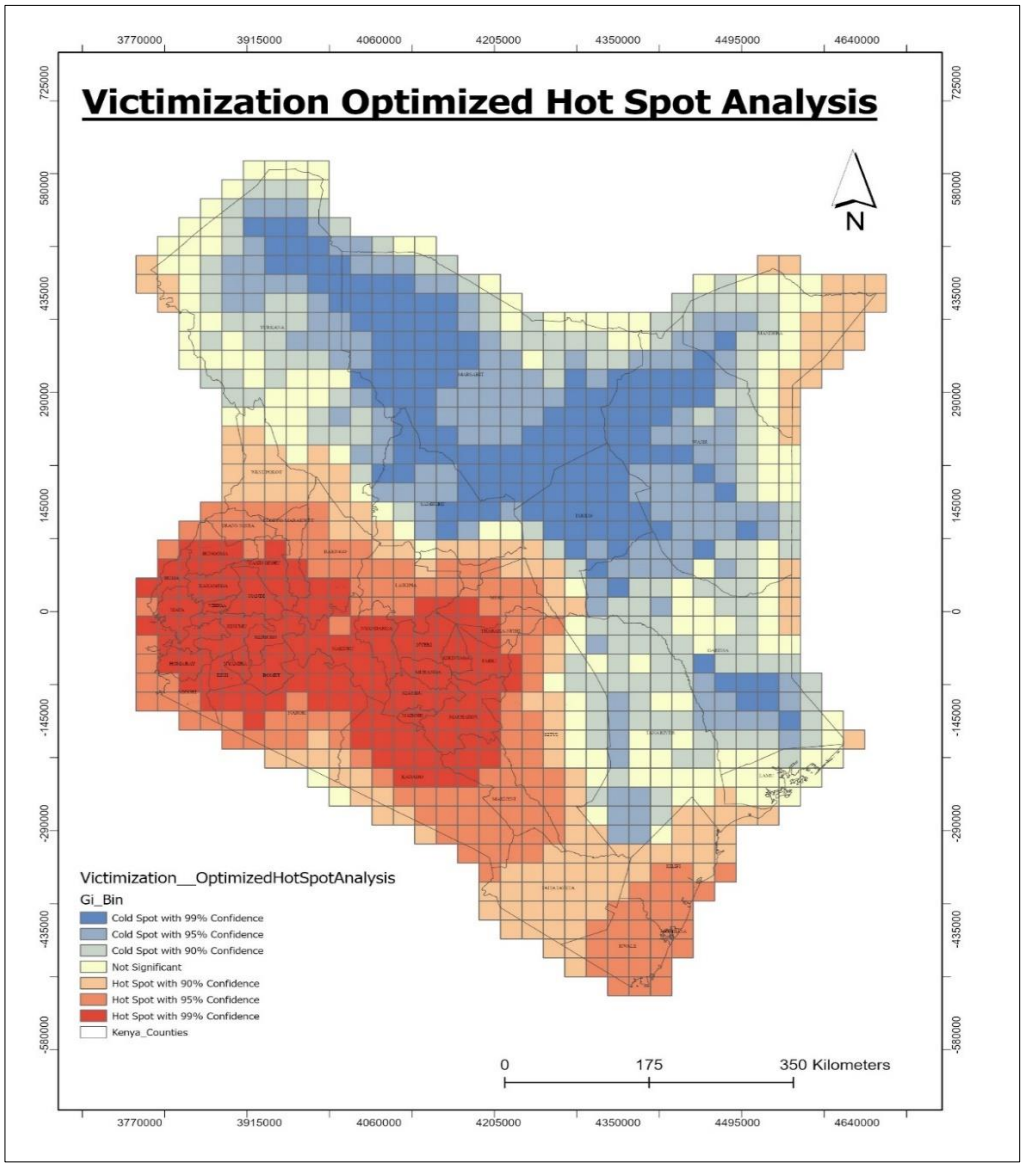


Figure 12: Hotspots for cyber attacks victims

4.2.2. Spatial Autocorrelation

Found significant spatial autocorrelation among the cyber activities and specific geographic regions, indicating that the spatial patterns and trends of cyber attack activities are not random. The Local Morans I statistic identified several significant spatial clusters and outliers of cyber attacks providing further evidence of the non-random distribution of Cyber victimization in Kenya. The results of the spatial autocorrelation are shown below.

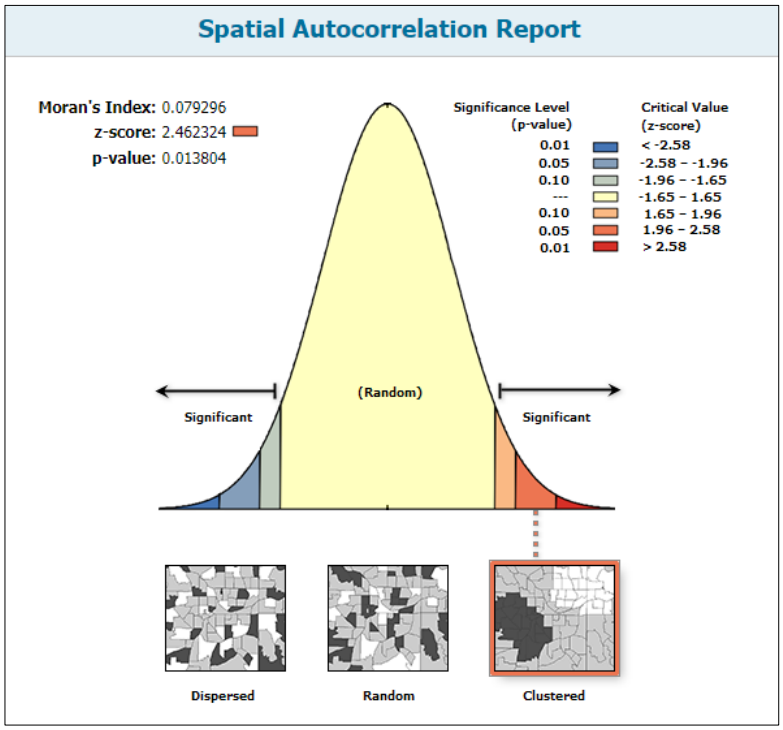


Figure 13: Spatial Autoccorrelation report

Given the z-score of 2.462324, there is a less than 5% likelihood that this clustered pattern could be the result of random chance.

Figure 14 below displays the clusters and outliers of total cybercrime activities per counties . The purpose of this figure is to examine the spatial autocorrelation between counties and cybercrime activities. The figure shows that certain counties within specific geographic region have same clusters, while other are assumed as outliers or insignificant. This suggests that counties with high cybercrime activities are geographically associated with their neighbors who are too at risk of cyber crime victimization.

Moran's I was used to quantify spatial autocorrelation. The tool measures a spatial autocorrelation that ranges from -1 to 1. A value of -1 indicates perfect dispersion, a value of 0 indicates no spatial autocorrelation, and a value of 1 indicates perfect clustering. It considers the value of each observation, the location of each observation, and the values and locations of neighbouring observations. Specifically, it compares the observed spatial pattern of a variable to what would be expected under the null hypothesis of random spatial distribution.

For example, The Clusters in the Map indicates clustering (i.e., similar values are close to each other) while the Outliers indicates dispersion (i.e., dissimilar values are close to each other). So, a Moran I value of 0 and above indicates that there is spatial autocorrelation in the variable.

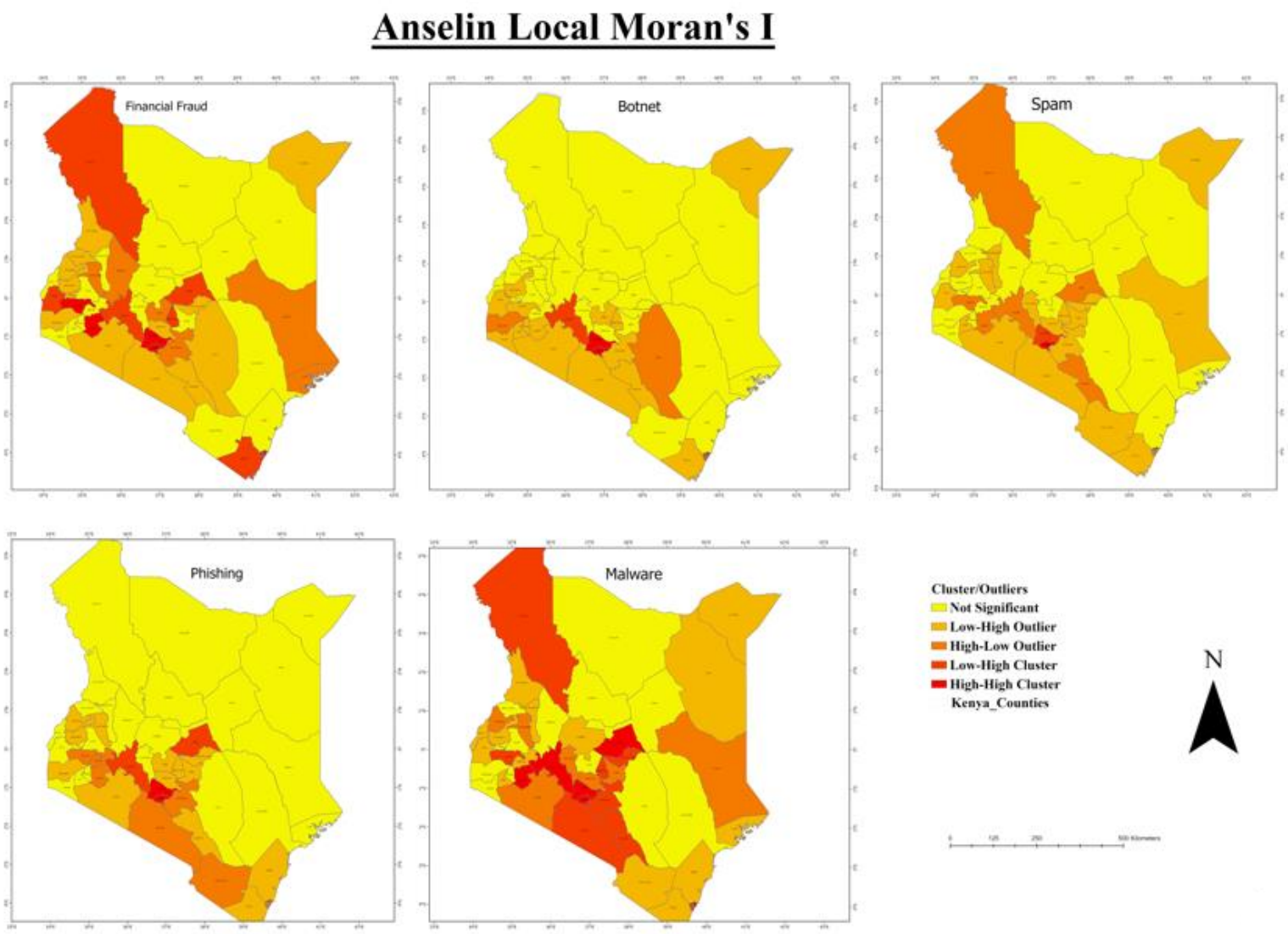


Figure 14: Moran's I cyber attack output

The Values are quantified based on spatial autocorrelation which ranges from not significant values, outliers, and Clusters. No significant ranges indicate counties that are not significantly correlated in cybercrime attack, Outliers indicates partial or no spatial autocorrelation, and Clusters indicates perfect spatial autocorrelation. It considers the value of each observation, the location of each observation, and the values and locations of neighboring observations. Specifically, it compares the observed spatial pattern of a variable to what would be expected under the null hypothesis of random spatial distribution. Example High-High cluster means that the cyber victims of that location have a high spatial autocorrelation to the geographical region.

4.2.3 Correlation Analysis.

The Pearson correlation coefficient (r) is the most common way of measuring a linear correlation. It is a number between –1 and 1 that measures the strength and direction of the relationship between two variables. When one variable changes, the other variable changes in the same direction. This analysis revealed a strong positive correlation between the number of cyber victims and the level of electricity connectivity, ICT Access, income and Population.

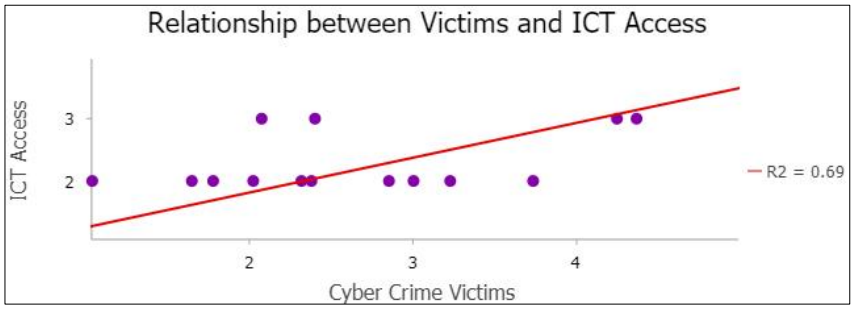


Figure 15: Cyber victims and ICT access correlation

The figure shows a scatter plot for the relationship between Total Cybercrime cases and ICT Access. From the figure, we can conclude that as access to ICT increases so as the victims of cybercrime. It has a strong positive correlation.

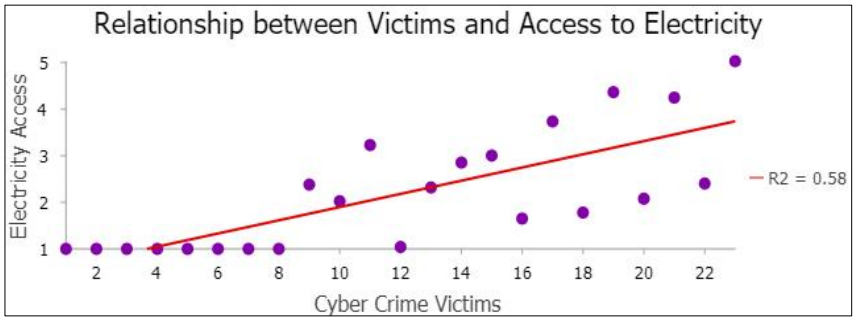


Figure 16: Cyber attack victims and electricity access correlation

The figure shows a scatter plot for a relationship between access to electricity and total cybercrime victims. High access to electricity is shown to be significant to the total cases of cybercrime victims. This shows a strong positive correlation.

Figure 17 below shows the relationship between total cybercrime attacks and total income. It has a Pearson's correlation coefficient of 0.2. This means that it has a very low positive correlation relationship. It means that cyber attack victims may not have been targeted due to their income levels. The hacker may have had another motive such as online identity theft or hackers do not care for the amount of money you have as long as they can extort you of something for them to gain it.

What's interesting to see is that according to the scatter plot there are more cyber victims in the population with low income than higher income. This may be a result of maybe those who earn more securely save their money in institutions not easily accessed by hackers while the ones with low income save their money in places which can be easily accessed by hackers. It can also mean that the higher earners are more aware of cyber attacks and crimes more than the low earners who may lack awareness.

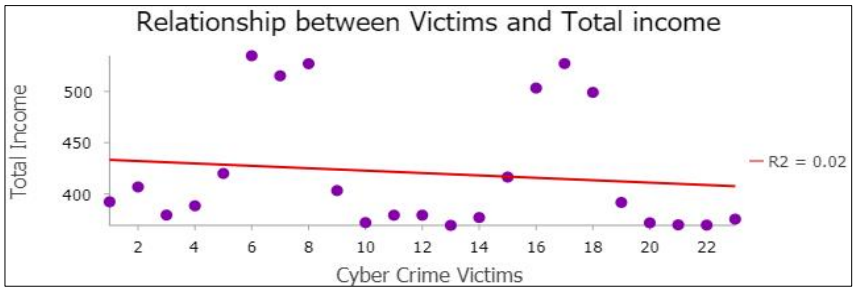


Figure 17: Cyber attack victims and income correlation

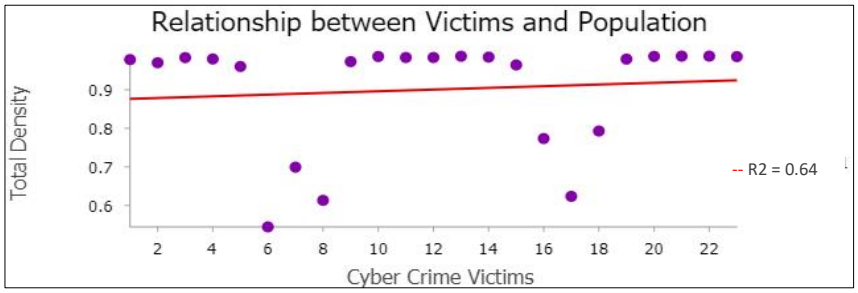


Figure 18: Cyber attacks victims and population

The figure shows the relationship between total cybercrime attacks and the total Population. cybercrime tends to occur where the population is high. It has a strong positive correlation. The more the population grows, the more the number of cyber attacks increase.

4.3 Regression

Ordinary Least Square regression is applied in order to examine the impact of explanatory variables and dependent variable. It is basically a way to predict unknown quantities from a batch of existing data, by applying OLS, we'll get an equation that takes the independent variable as an input, and dependent variable as output. Regression result shows that the Adjusted R-Squared value is 0.882, or 88%. This indicates that using the exploratory variables with the dependent variable, the model is explaining 88% of the variation in cyber attacks incidents.

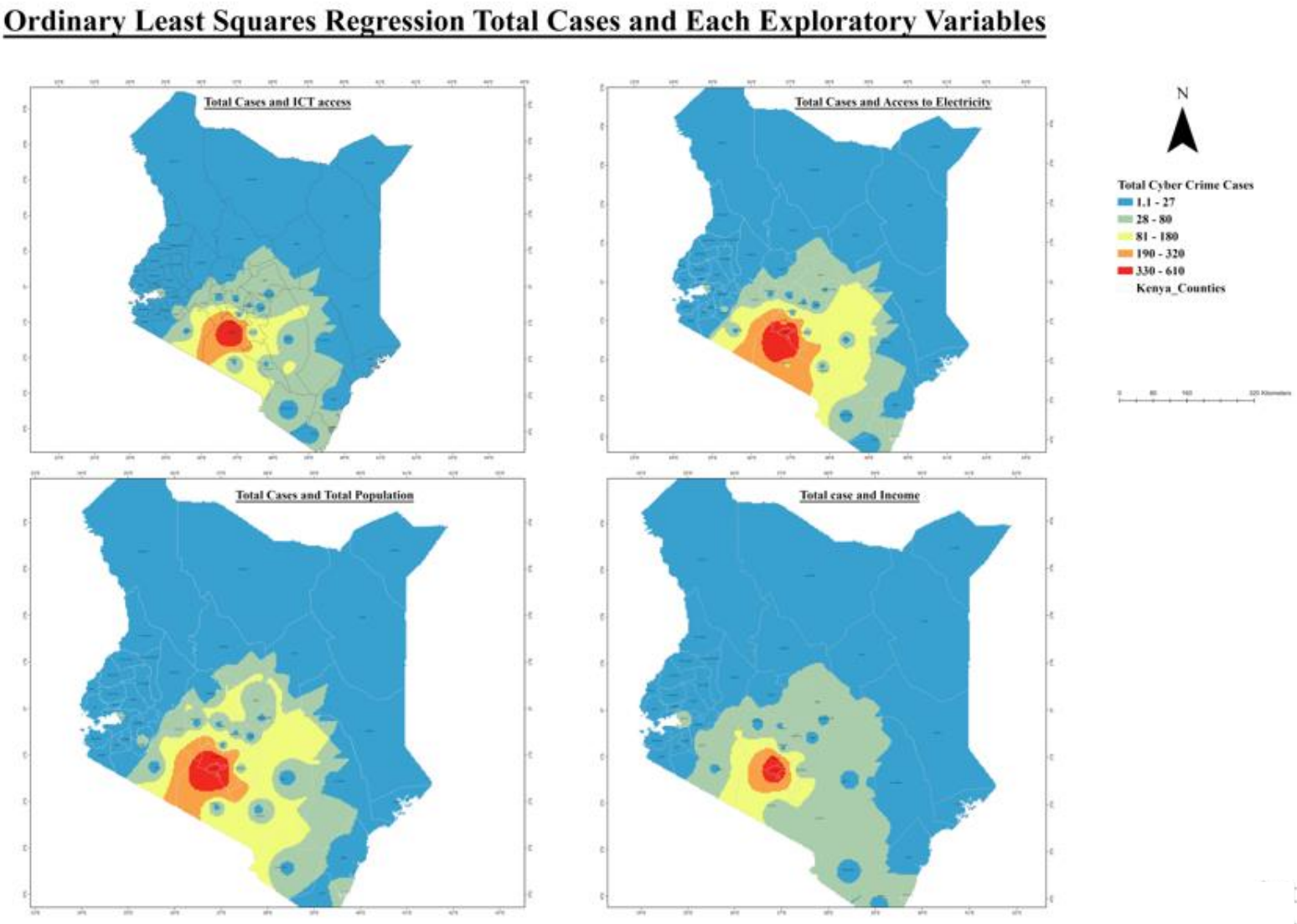


Figure 19: OLS analysis between independent and dependent variables.

It displays the proportion of victims of cybercrime and access to each exploratory variables.

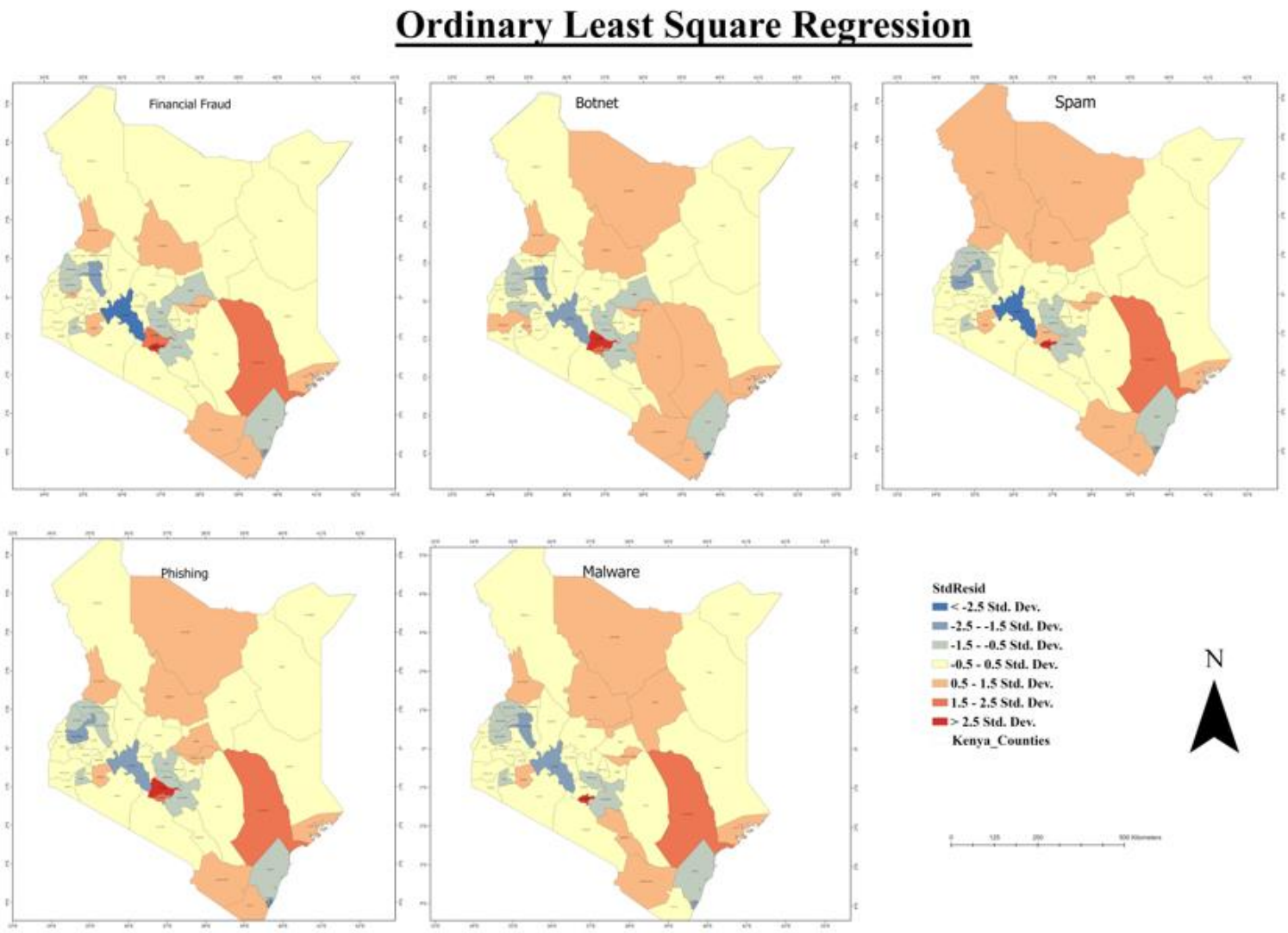


Figure 20: Regression analysis of standard residuals

Figure 20 displays the results of ordinary Least squares for cybercrime attack type with all other exploratory variables. The purpose of this figure is to determine whether the explanatory variables in the model have a consistent relationship to the individual attack type as a dependent variable, especially in geographic space. The independent variables are also spatially lagged to consider the potential spatial dependencies in the data.

The spatial lag operator calculates the average value of a variable among its neighboring observations and includes this spatially lagged value as an additional independent variable in the regression model. This is used for the identification and quantification of spatial dependencies and spatial autocorrelation in the data as shown in figure 20.

Standard deviation (SD) values generated show the variation from the slag means. It indicates how much the individual Cybercrime Attack varies from the exploratory variables. A Higher deviation means a higher cyber attack in the specific county while a negative deviation equates to a dispersion in which the counties do not correlate with other exploratory variables.

In the context of the figure 20, a standard deviation of over 2.5 means the relationship between specific attacks and other variables is ideal and therefore high cases of cyber attacks. A standard deviation of -2.5 is found to indicate that there is a high dispersion in the data and that means the counties involved are not ideal for cyber attacks or few are reported.

4.4 Web application

One of the challenges faced during the analysis of the project was standardization of the data obtained. A large amount of time was spent in data cleaning and filtering as some of the areas of the cyber attacks reported required to be written in a correct manner. Moreover, the low reporting rate in the country of victims of cyber attacks shows that there is low awareness on cybercrimes. To solve this a web application was created for cyber attacks incident reporting and visualization.

The first page is the Homepage which gives statistics on the current cyber-attack scenario in Kenya from the Communication of Authority of Kenya. Analysis page displays maps in which analysis has been done on data obtained. Contact page gives information on the relevant authority and their toll number to report a cyber attack. Lastly, report page allows victims to input data to report any cyber-attack incident.

The report page is connected to a database in the backend framework in which when the victim submits the data recorded, it is stored in the database.The backend web application is only available to administrators. This is gotten from a scenario where assuming that one works at the DCI cybercrime department, by logging in one’s credentials, access is given.

The administrator is able to access the data stored in the database and view them on a map with all the detail recorded. The administrator may also edit or create any new data example a call comes from a victim reporting an incident, using the add cyber attack or add cyber attack victims forms, the administrator records the incident.

The following are the screenshots of the web application both the frontend and backend of the application.

Frontend web application

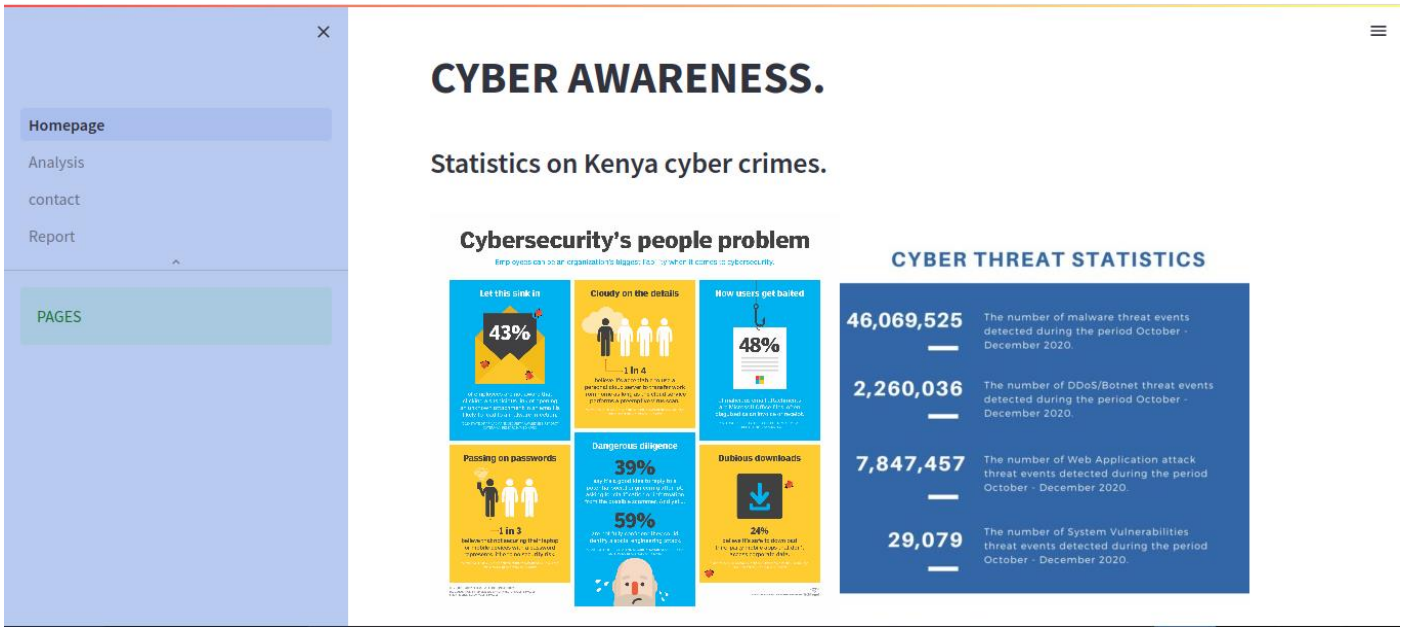


Figure 21: Homepage of the frontend web application

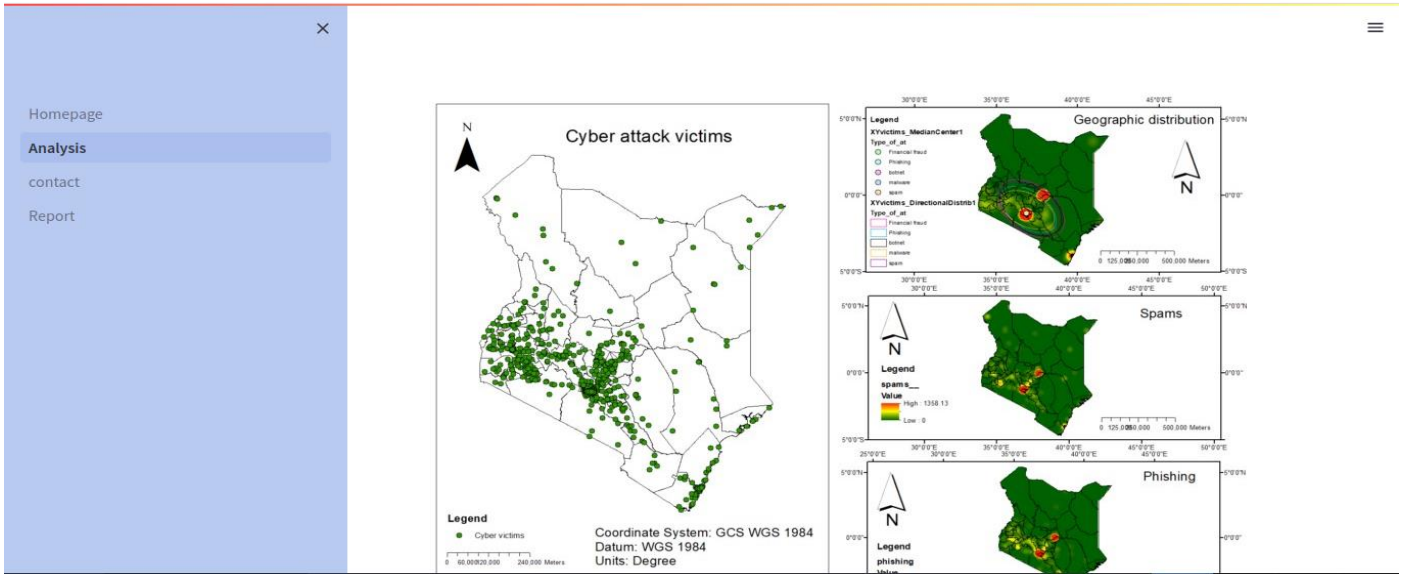


Figure 22: Analysis page

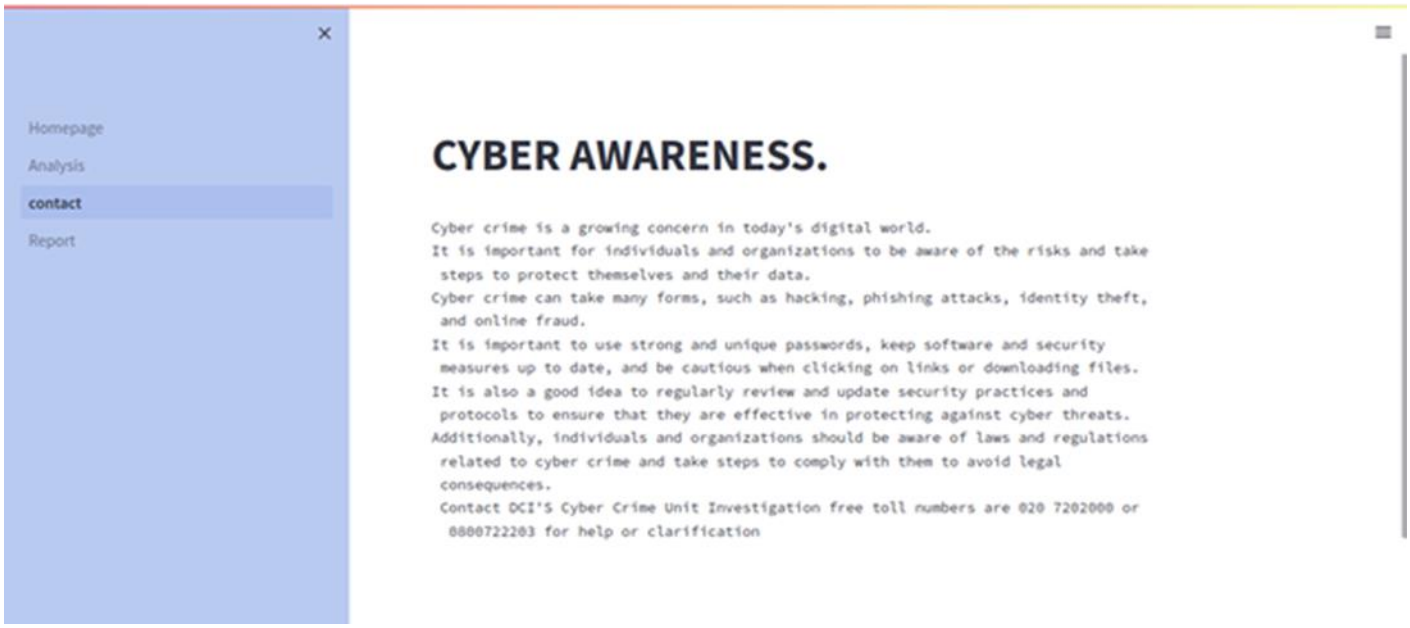
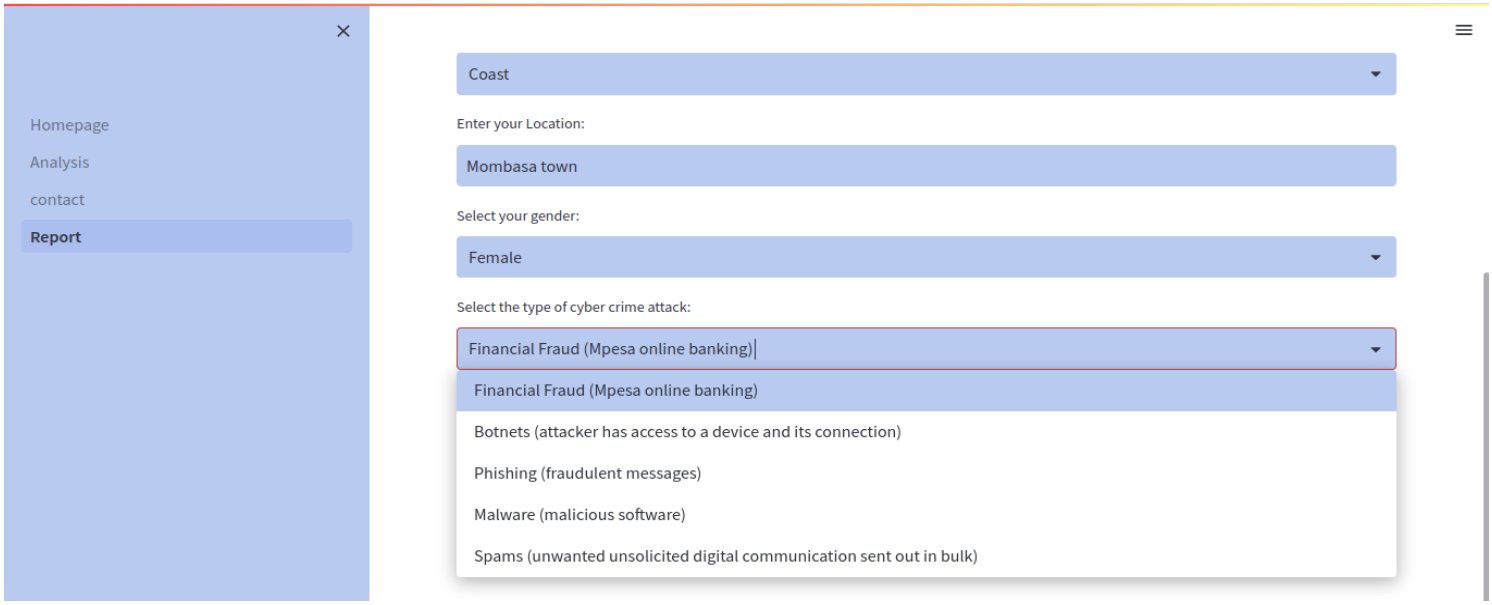


Figure 23: Contact Page

Figure 24: Report page for the frontend web application



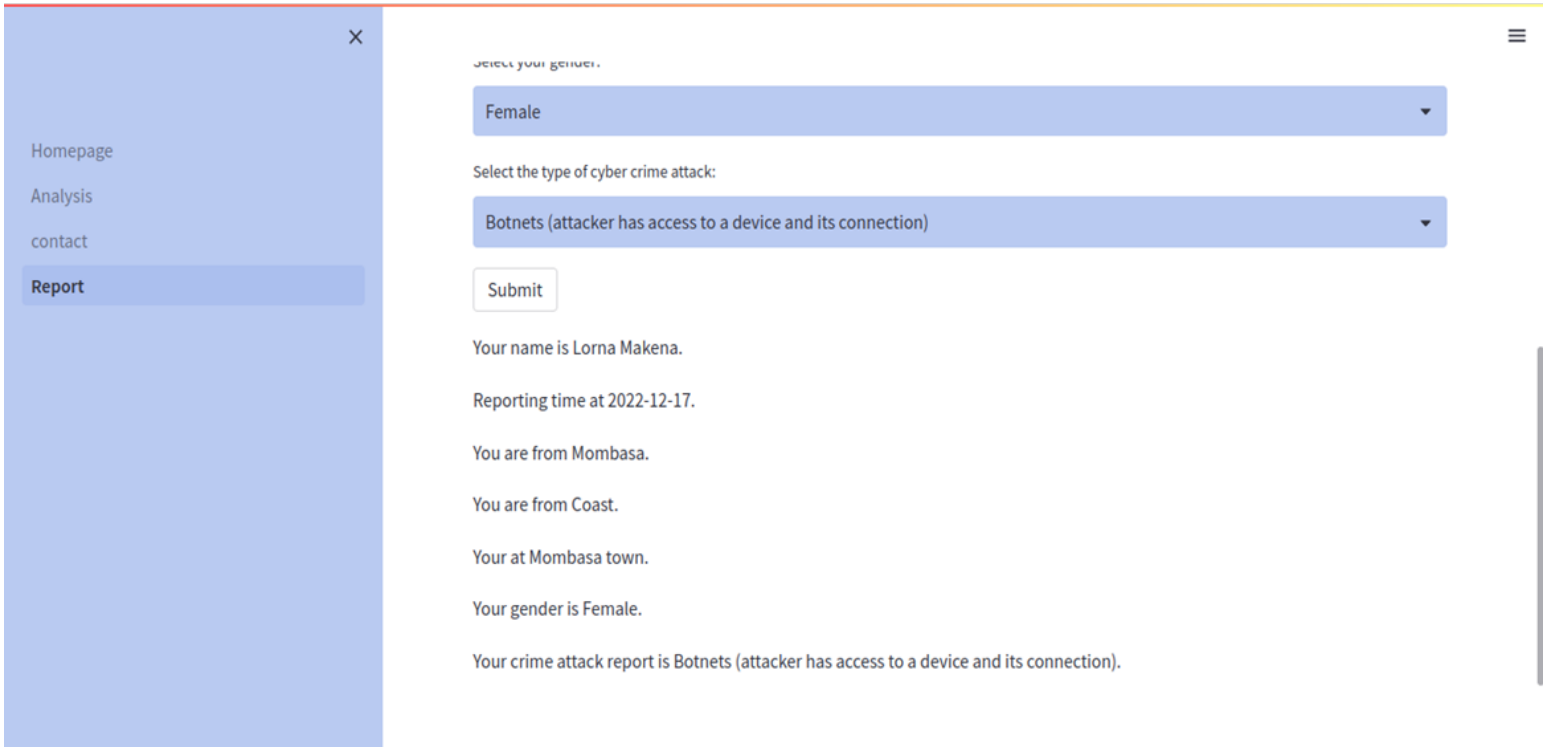


Figure 25: Report pages

Backend web application.

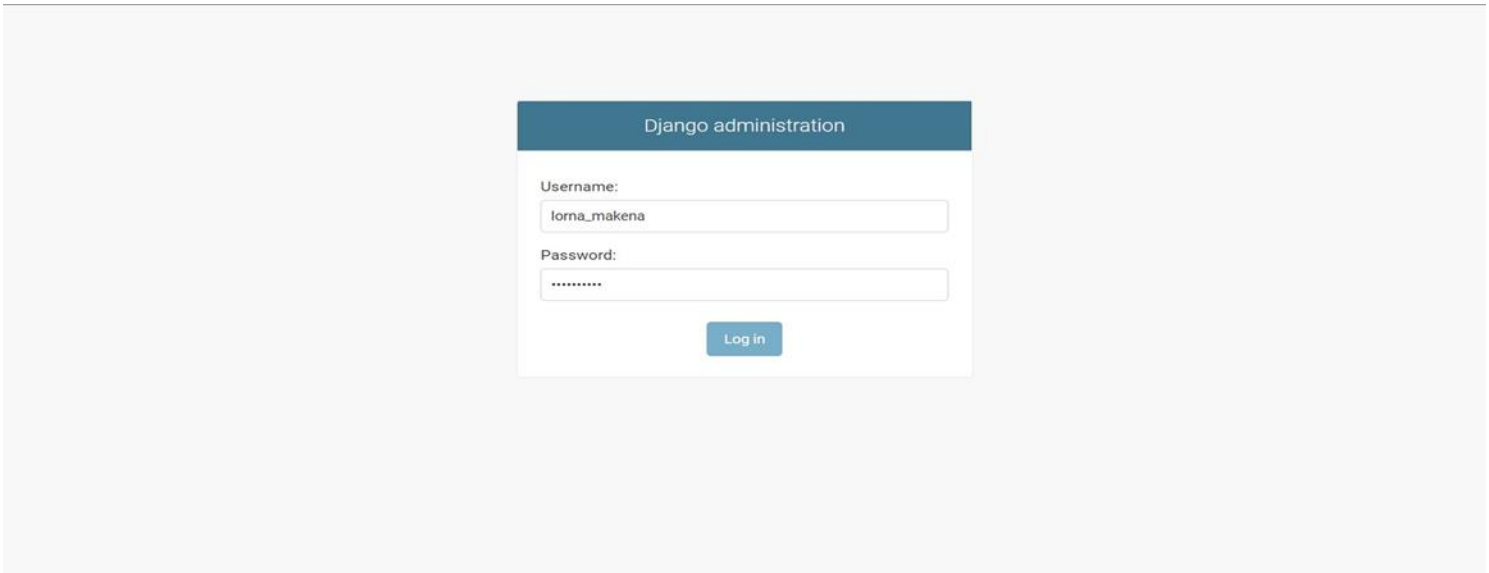


Figure 26: User login of backend web application

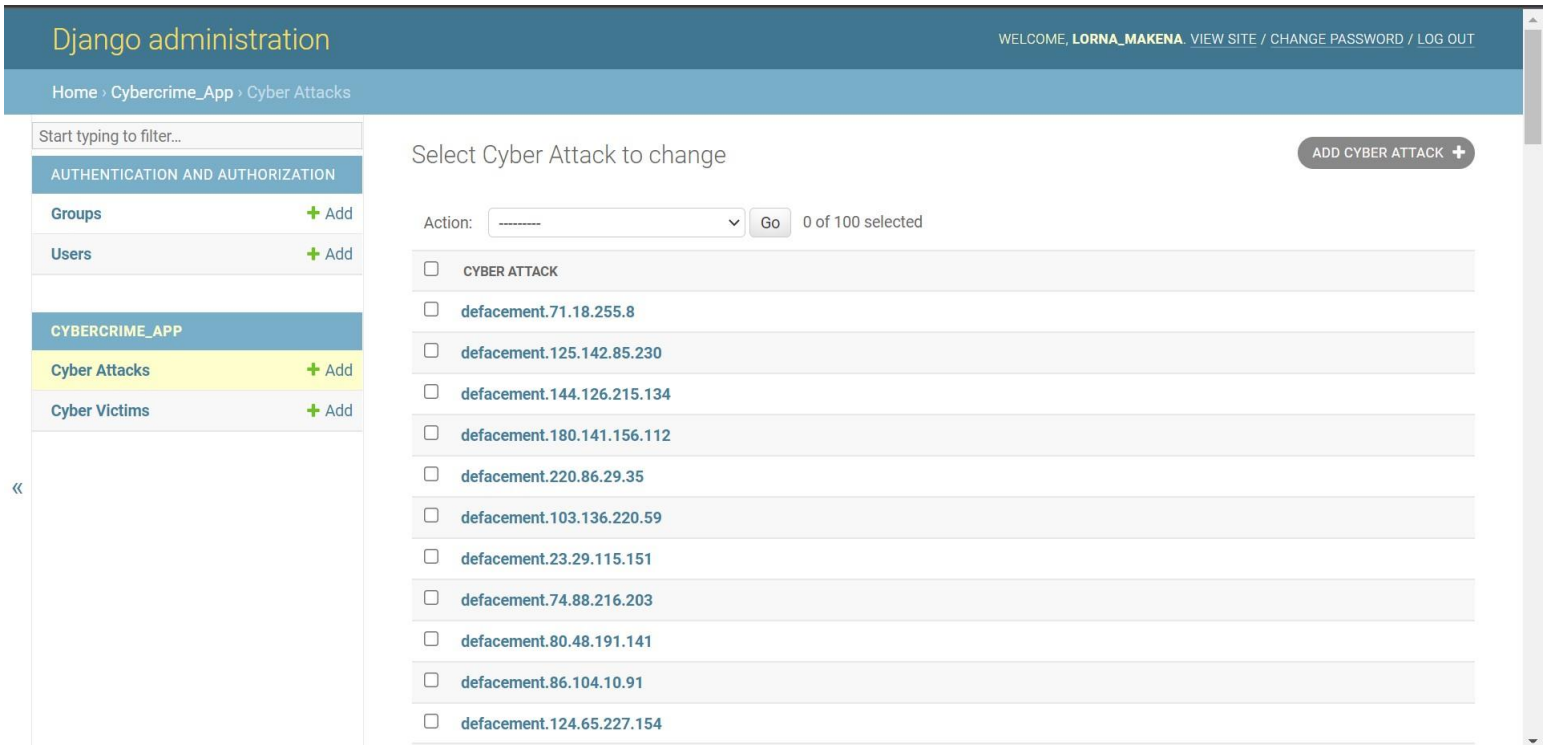


Figure 27:Cyber Attack data on the database

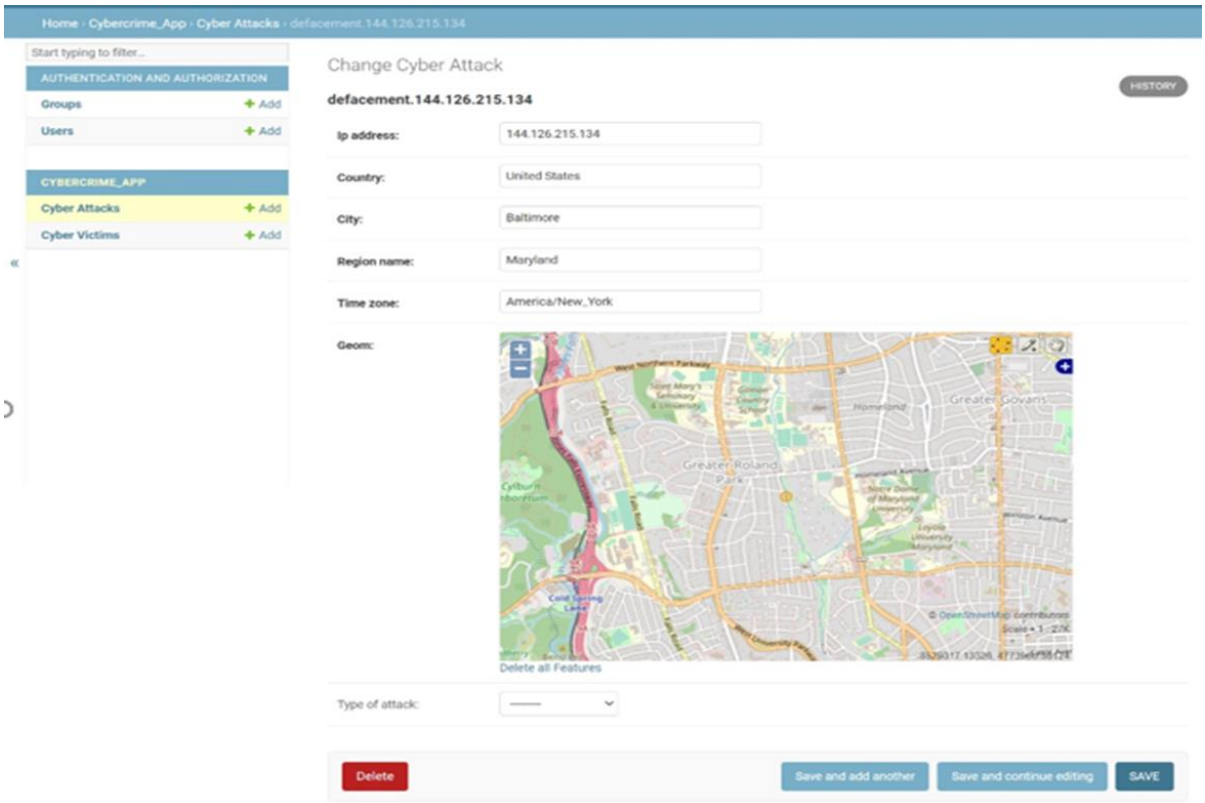


Figure 28: Example of details of cyber attack incident

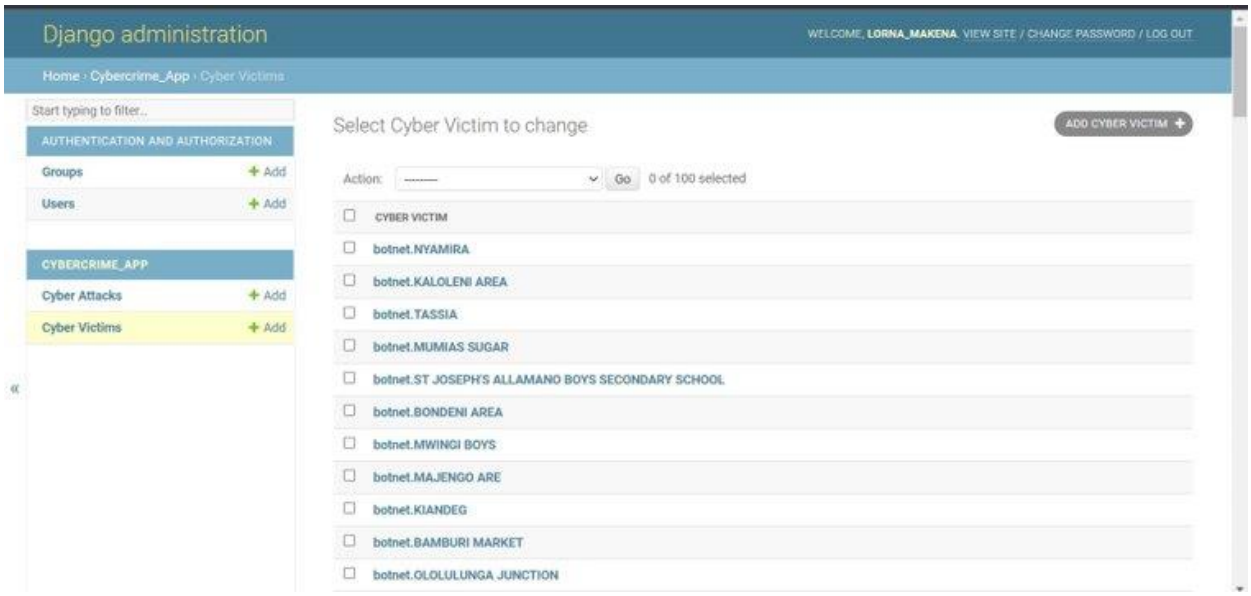


Figure 29: Cyber attack victims on the database

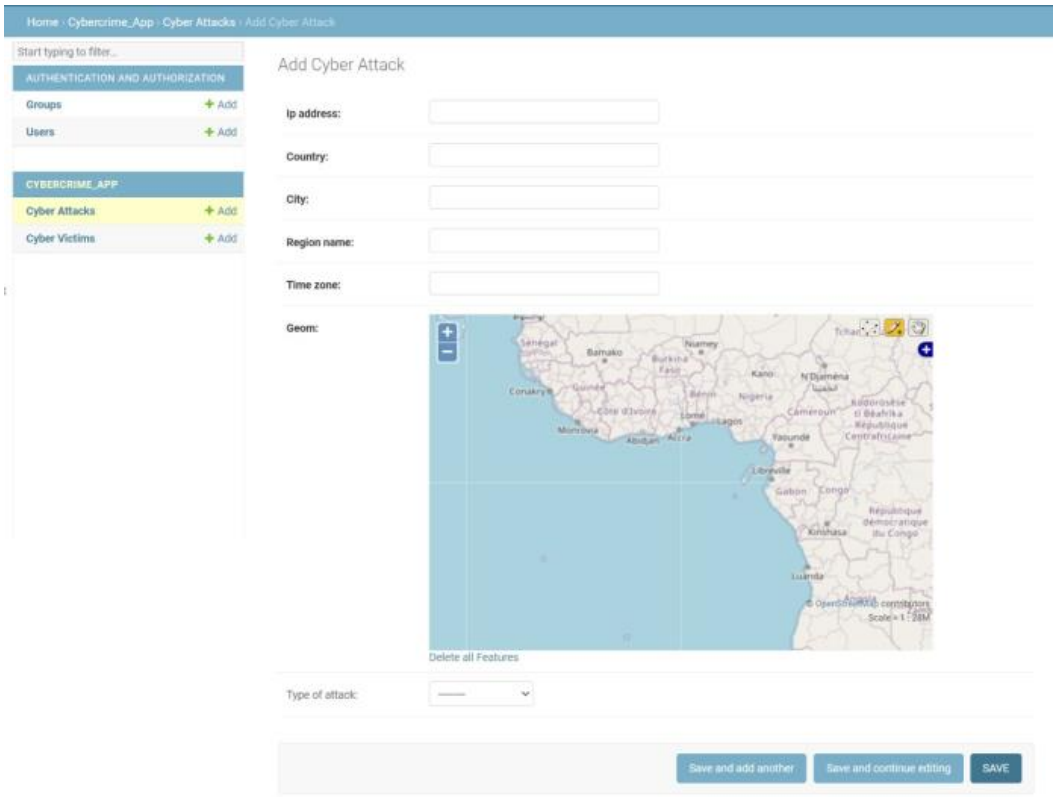


Figure 30: Add cyber attack IP form

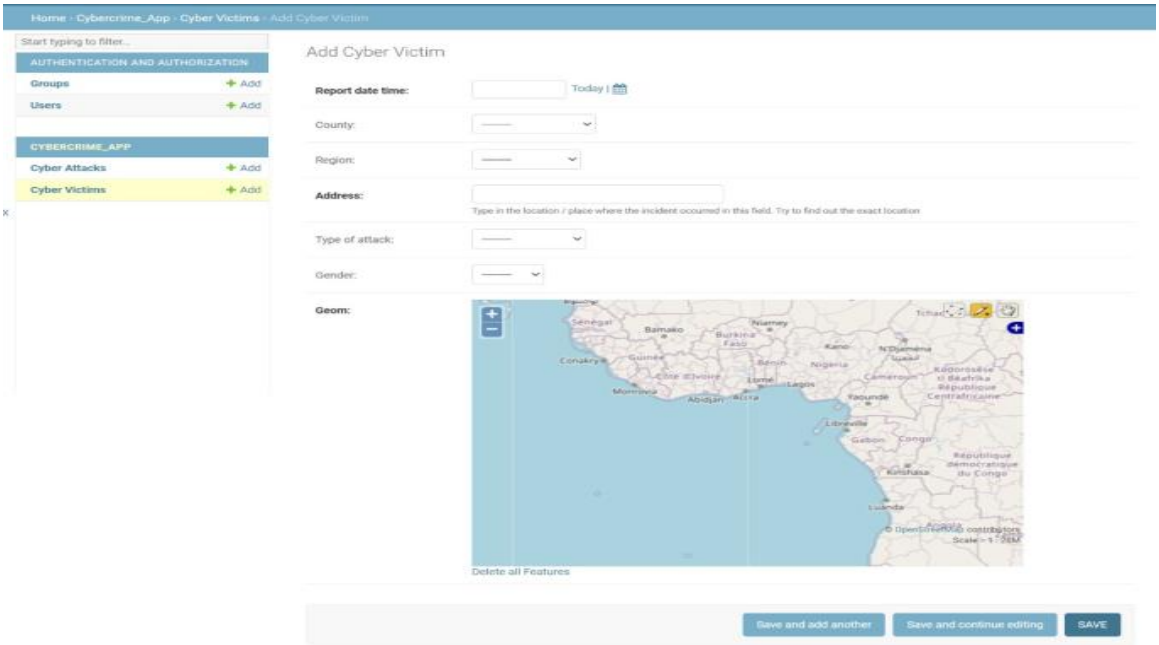


Figure 31:Add cyber attack victims form

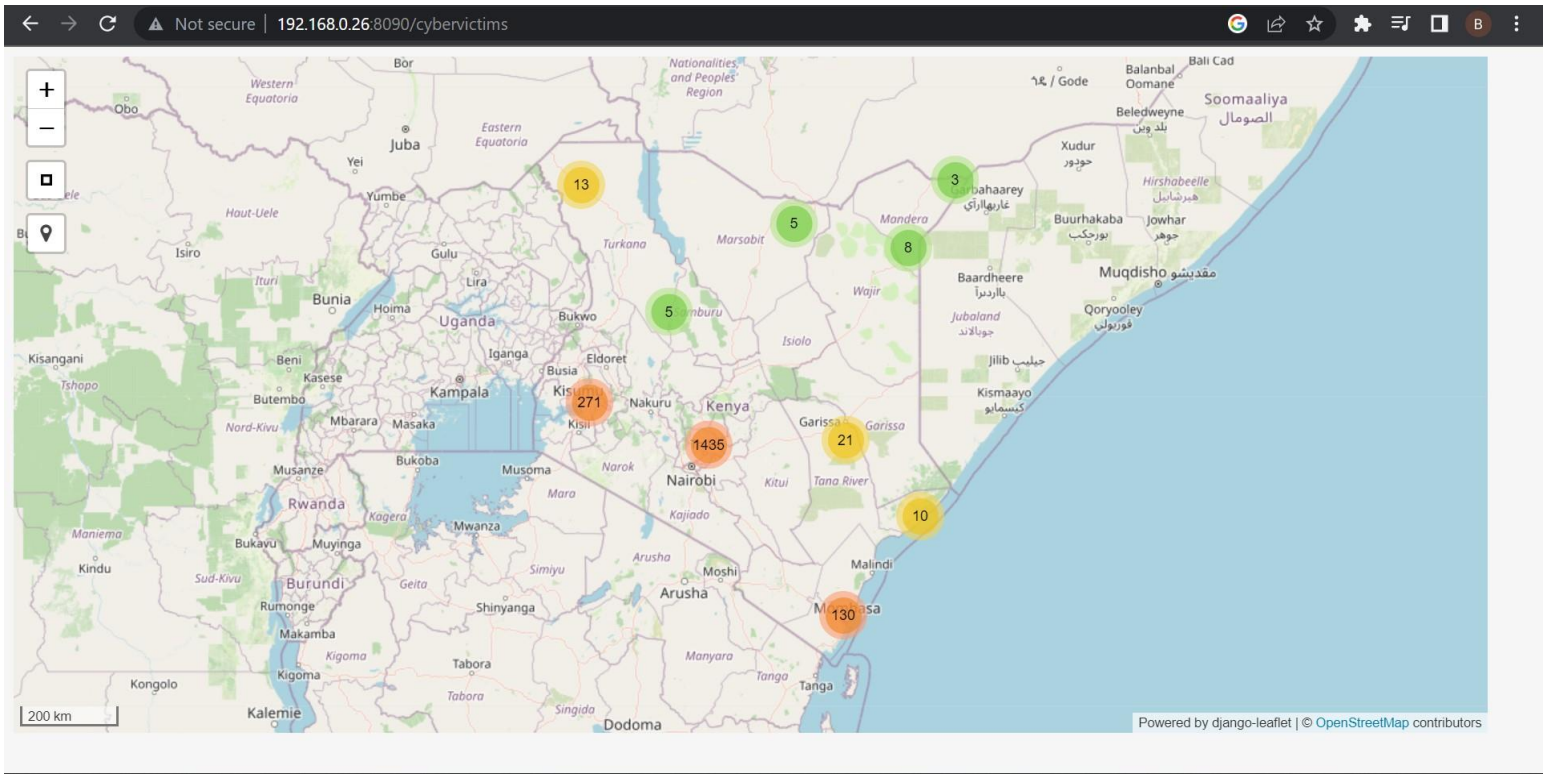


Figure 32: Cyber attack victims in Kenya

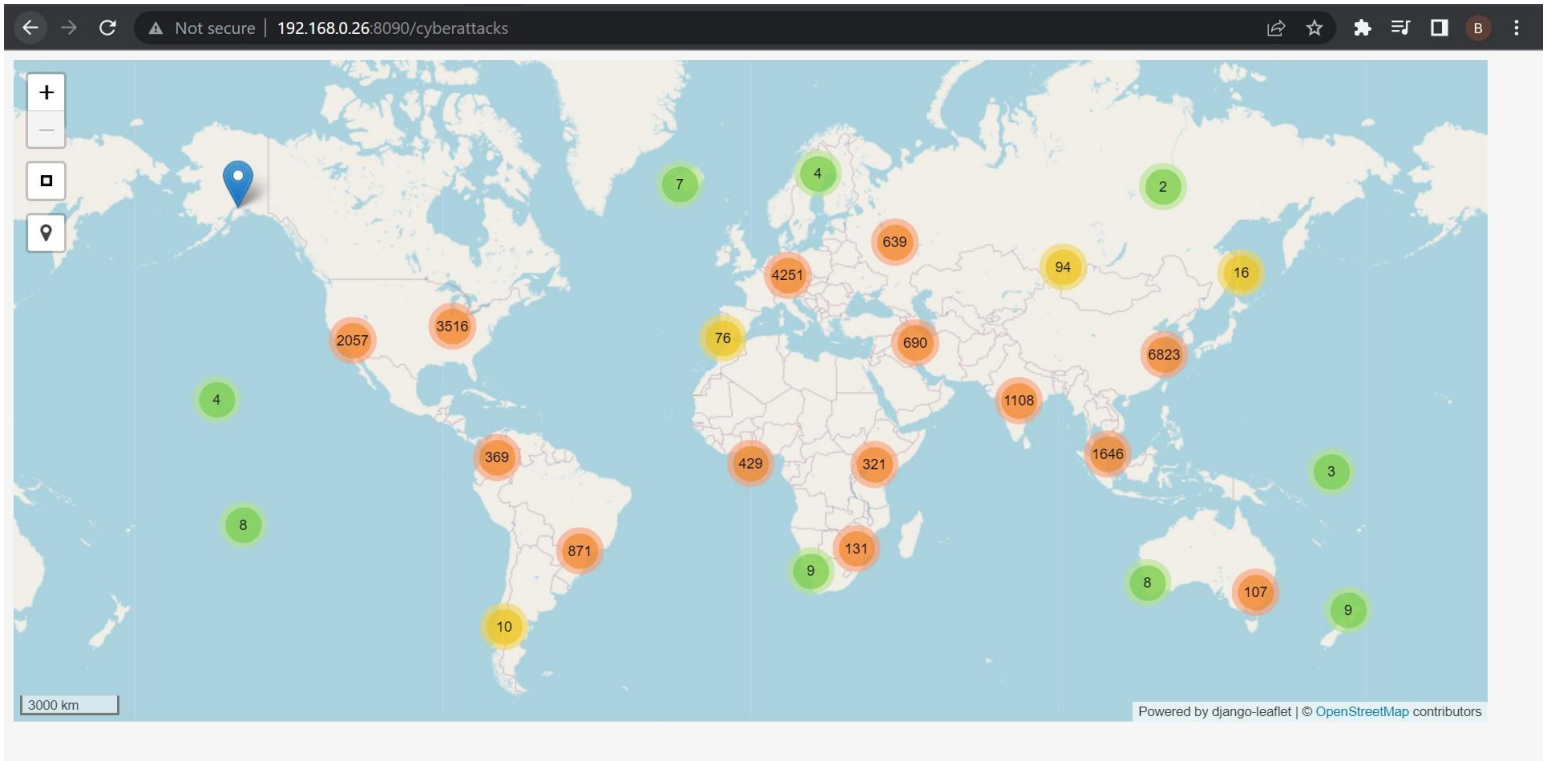


Figure 33: Cyber attacks source IPs targeting Kenya

5. Discussions

The findings suggest that Cyber victimization is not randomly distributed but are instead concentrated in certain areas with higher population densities and greater levels of internet connectivity as well as access to electricity.

Specifically, the results indicate that central Kenya, including regions such as Kiambu and Nakuru, Nairobi, and Kisumu have a high cluster, which suggests a higher concentration of these Cyber victimization in this region. This could be due to a variety of factors, such as the presence of a higher density of businesses or financial institutions that are targeted by cybercriminals or a higher level of internet usage or access to technology in this region, which could make it more vulnerable to cyber threats.

This study focused on analyzing the spatial patterns of cyber attack incidents, in Kenya. Through the OLS analysis, considered a range of factors that may influence the incidence of cyber attacks, including total population, income, access to ICT, and access to electricity.When examining the relationship between Cyber victimization and the various socio-economic factors, it was found out that total population, income, and access to electricity were positively correlated with the incidence of cyber-attacks.

Regions with higher populations and greater economic activity, as well as better access to electricity, had higher rates of cyber attacks. Additionally, access to ICT, such as internet connectivity and access to technology, was also positively correlated with the incidence of cyber-attacks. Regions with better access to ICT tended to have higher rates of cyber attacks, possibly due to a greater number of internet users and greater use of online platforms for financial transactions and other activities targeted by cybercriminals

As part of the analysis, spatial autocorrelation techniques was used to identify clusters and outliers of Cyber victimization across the country. The results indicated that Cyber victimization is not randomly distributed in Kenya, but instead, there are distinct patterns of clustering and outliers. There was high clusters of incidents in the central, rift valley, and western regions of Kenya.

Conversely, Nakuru and Uasin Gishu had negative standard deviation residuals, indicating a lower incidence of Cyber victimization. These findings suggest that cyber threats are concentrated in specific regions of the country, and targeted interventions may be necessary to address these challenges.

To further explore these patterns, Anselin Local Moran's I was used to analyze spatial autocorrelation and identify clusters and outliers of Cyber victimization in Kenya.

6. Conclusion

From the spatial characterization of cyber attacks, there is a noticeable increase in cyber attacks from 2020 to 2021 in the hotspots maps (Figure 9). It is shown by the map that there is a centralization of cyber attack victims in the central part of Kenya, that is mostly Nairobi and Kiambu counties. There is a huge gap between cyber attacks victims from Nairobi to the rest of the towns such as Mombasa, Turkana or Mandera. This affirms that Nairobi is a major hotspot for cyber attacks given it is the capital of Kenya and where most companies and industries are located in Kenya. It also boasts as one of the counties with the highest population in Kenya which proves the positive correlation between large number of population and cyber attacks.

Most cyber-attacks source IPs (Figure 10) originate from outside of Kenya with 88% of hackers are located globally and 12% of hackers are located internally (Communication of Authority of Kenya 2021 annual report, 2021). Cyber attack source IP addresses were not included in the analysis due to lack of credibility of the data but were used in the creation of a database in which security experts can perform digital forensic investigation on the IPs to determine their credibility. Hackers do not use their own IP addresses but a proxy i.e. a hacker might be from Kenya however his IP address reads it's from Afghanistan. However visually displaying the data can show huge clustering of areas or locations which target Kenya.

Financial Fraud, phishing and spams are the most types of cyber attacks that were reported by Kenyans in the 2020 and 2021. Nairobi and Kiambu counties had the highest number of all five types of cyber attacks. This gives a picture to how also because of proximity in location between the two there is high number of cyber incidents.

Marsabit, Wajir, Isiolo and Garissa had little to none cyber-attack incidents. This displayed a problem or error on the data obtained. The is because the percentage of internet users in Kenya by 2020 was 42% of the whole population, according to the Communicaton Authority of Kenya, this value should have risen in 2021 due to the Covid pandemic which shifted money transactions and any mode of communication online. Thus, displaying a county with little to none number of cyber attack victims against the internet users in the same area proves to show the low reporting rate of cyber attacks.

This can be solved through the creation of a web application that connects to a database. The web application serves five purposes: creating a platform in which cyber-attack incidents can be

reported this solves the problem of data collection, platform to host a database to store the source IP addresses which can be used by cyber security experts in digital forensic investigations, real time tracking and monitoring in the web application, increases cyber security awareness through offering maps for visualization and data mining and lastly coordination and collaboration between cyber security agencies in Kenya.

Ordinary Least Squares (OLS) is a regression method that allows one to see the relationship between a dependent variable and exploratory variables. By using regression, dependency factors such as population, ICT access, electricity access that had a significant correlation with the dependent factor, cyber attacks, were able to be deduced and this were population, internet users per county and ICT access per county.

The identification of clusters and outliers of Cyber victimization can help policymakers and cybersecurity experts to develop more targeted interventions to address the root causes of cyber threats in specific regions. By identifying regions with high clusters of cyber attacks, policymakers can prioritize interventions to improve internet infrastructure and security measures in those areas, as well as provide cybersecurity training and resources to businesses and financial institutions operating in those regions. Similarly, addressing the unique challenges faced by outlier regions can help to prevent the spread of Cyber victimization and enhance the overall cybersecurity posture of the country

In conclusion, geospatial data can be successfully integrated in a cyber space environment to monitor and track cyber attacks in an area of study

7. Recommendation

Design of Cybercrime Spatial Analysis System which has 2 aspects to this:

- Archiving fast tracking and spatial orienting for complex IP such as proxy, this is through building a Network resource spatial database and cybercrime spatial database this solves the issue on the credibility of the IPs that will be tracked.
- Lastly establishing a spatial analysis model that can be used to predict cyber attacks.

References

1. A. L. Buczak & E. Guven (2016). A Survey of Data Mining and Machine Learning
2. Altaweel, M. (2021). Cyber Security and GIS. GIS Lounge. https://www.gislounge.com/cybersecurity-and-gis/
3. Amir, B. et al., (2014). Geographic Profiling for serial cybercrime investigation. Police Support Agency of Bosnia and Herzegovina. https//dx.doi.org/10.1016/j.aebj.2014.10.-12
4. Andreea Bendovschi. (2016). Cyber attacks – trends, patterns and security countermeasures. doi: 10.1016/S2212-5671(15)01077-1
5. Back, S., LaPrade, J. & Sadhika, S. (2018). Spatial and Temporal patterns of Cyber attacks: Effective cybercrime prevention strategies around the globe. International Journal of protection, security and investigation. Publisher J-Institute, Japan
6. Bakis, H. (2001). Understanding the geo-cyberspace: A major task for geographers and planners in the next decade. Netcom, 2001, 15(1/2):9-16.
7. Barney, W. (2007). Geographies of the tropical Internet: An overview. Singapore Journal ofTropical Geography, 2007, 28(2):219-238.
8. Batten, P. & Micheal, J. (2005). Some principles for landscape mapping on the Internet. Journal of Map Geography Libraries, 2005, 2(1):99-110.
9. Batty, M. & MILLER, H., (2000). Representing and Visualizing Physical, Virtual and Hybrid Information Spaces. In Information, Place and Cyberspace Issues in Accessibility (edited by D. Janelle and D. Hodge, Berlin: Springer) 2000, 133-146.
10. Bordeanu, O. (2018). How's the weather like in cyberspace? Pretty hot in south-west and south-east according to the cybercrime heatmaps. University of London thesis, 2018.
11. Charvi Vij & Shruti Keshari (2022), "Study on Lexical Analysis of Malicious URLs using Machine Learning", 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), pp.120-127, 2022.
12. Chen, & Huang. (2015). Spatio-Temporal patterns and predictability of cyber attacks.
13. Cisco Annual Report (2020). Internet Connectivity. https://www.cisco.com/c/dam/en_us/about/annual-report/cisco-annual-report-2020.pdf
14. Communication Authority of Kenya annual cybercrime report (2020). https://ke-cirt.go.ke
15. Communication Authority of Kenya annual report (2021). Annual Report and Financial statement 2021. https://www.ca.go.ke/downloads/publications/annual-reports/

16. Computer Misuse and Cybercrime Act (2018) https://nc4.go.ke/the-computer-misuse-act
17. Detica, P., & Parkes, R (2013). Rise of Cyber Attacks and its negative implications. University of Carlifonia. https://doi.org/10.1093/oso/9780198863007.001.0001
18. Dodge, M. & Kitchin,R. (2001). Mapping Cyberspace, Routledge, London. DOI:10.4324/9780203279731
19. Du, & Yang. (2013). Temporal and spatial analyses for large-scale cyber attacks. *In* Environment and Planning (B), 2001, 28(1):3-19.
20. Feroz, & Mengel. (2015). Phishing URL detection using URL ranking. Published in: 2015 IEEE International Congress on Big Data. DOI: 10.1109/BigDataCongress.2015.97 Handbook of Computational Approaches to Counterterrorism; Springer.
21. Geo-statistical Dengue Risk Model Case Study of Lahore Dengue Outbreaks 2011. 10.1109/GEOINFORMATICS.2015.7378714
22. Hu, Baynard, & Hu. (2015). GIS mapping and spatial analysis of cybersecurity attacks on a Florida University. Publication Title International Conference on Geo-informatics
23. ISBuzz Staff. The Role of Geospatial Data in Cybersecurity. https://informationsecuritybuzz.com/the-role-of-geospatial-data-in-cybersecurity/
24. Janelle, D. & Hodge, D. (2000), Information, Place and Cyberspace Issues in Accessibility (Berlin: Springer), 2000.
25. Jiang, B. & Orm E. (2000). Mapping cyberspace: Visualizing, analyzing and exploring virtual worlds. The Cartographic Journal, 2000, 37-2:117-122.
26. Kellerman, A. (2002). The Internet on Earth: Geography of Information. WS: John Wiley and Sons Ltd, 2002:1-282.
27. Kinyanjui, P. (2019) Cyber attack on the National Hospital Insurance Fund. Article from Standard media newsapaper, 2019.
28. Koike, Ohno, & Koizumi. (2005). Visualizing cyber attacks using IP matrix.
29. Lin, Chiu, & Lee. (2013). Malicious URL filtering—A big data application. 2015 23rd International Conference on Geoinformatics, 1–5. Conference Location: New York, NY, USA.
30. LinLiu, R. (2008), Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems, Information Science Reference Imprint of IGI Publishing Hershey, PA 2008.

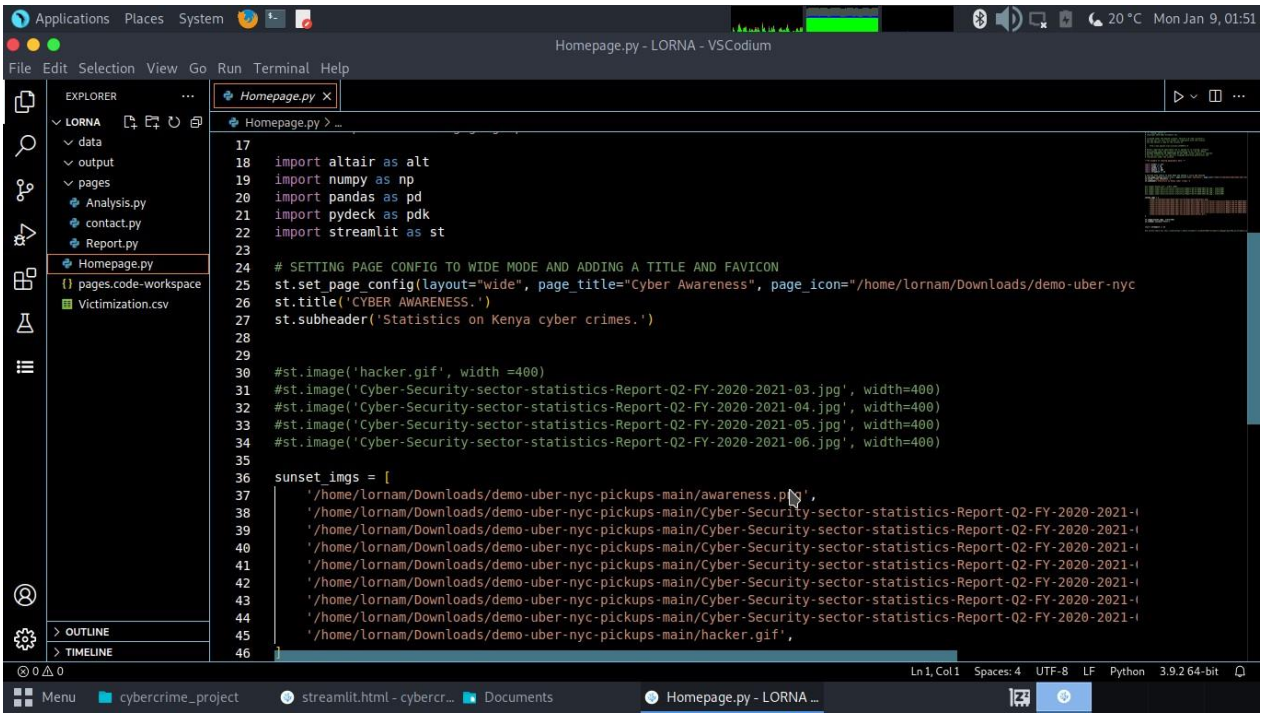
31. Luoli Wang (2021) J. Phys, A Spatial-Temporal correlation based method for advanced persistent threat detection. Conf. Ser. 2113 012037
32. Malecik, E. & Bennet, C. (2002). The Economic geography of the Internet's infrastructure. EconomicGeography, 2002, 78(4):399-424.
33. Martin, L. (2011) Cycle of Cyber Attacks from hacker to the computer access. Publication on Conference on Geospatial Science and Cyber Security Journal. 2011, 37-2:117-122
34. Merien, Bellekens, & Brossets. (2018). A Spatio-Temporal entropy-based approach for the analysis of cyber attacks. In Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.249450
35. Mohammedreza, P. & Vahid, B.P. (2018). Spatial Analysis of the crime of financial fraudof citizens in the virtual space of Tehran. Amin University of Police Sciences, Tehran.
36. Mumbi, R. & Mungai, L. (2018). Safaricom Scams in Kenya. Article from standard media newspaper, 2020.
37. Oduor, O. (2018) Phishing instances and cases in Kenya and their impacts
38. P. Gao et al., (2021) An Intelligent Threat-Detection Method for Power Monitoring System Based on Attack Chain Knowledge in 2021 IEEE 4th International Electrical and Energy Conference (CIEEC), May 2021, pp. 1–6. doi: 10.1109/CIEEC50170.2021.9510852.
39. Serianu Cyber Security Company Annual report 2020. https://www.serianu.com/annual-reports.html
40. Serwaa-Boateng et al. (2019) The most prevalent cybercrimes in Kenya. Springer, 2019. https//dx.doi.org/10.1016/j.aebj.2014.10.-12
41. Shaw, S. & Yu. H. (2010) A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space, Urban Planning International, 2010,6, 27-35
42. WarfB, G. (2001). Segue ways into cyberspace: Multiple geographies of the digital divide.
43. Wenqun Xiu & Xiaming Li (2010). The Design of Cybercrime Spatial Analysis System.
44. Yu, H. & Shaw, S-L (2008). Exploring potential human activities in physical and virtual spaces: A Spatial-Temporal GIS approach. International Journal of Geographic Information Science, 2008, 22(4), 409-430.

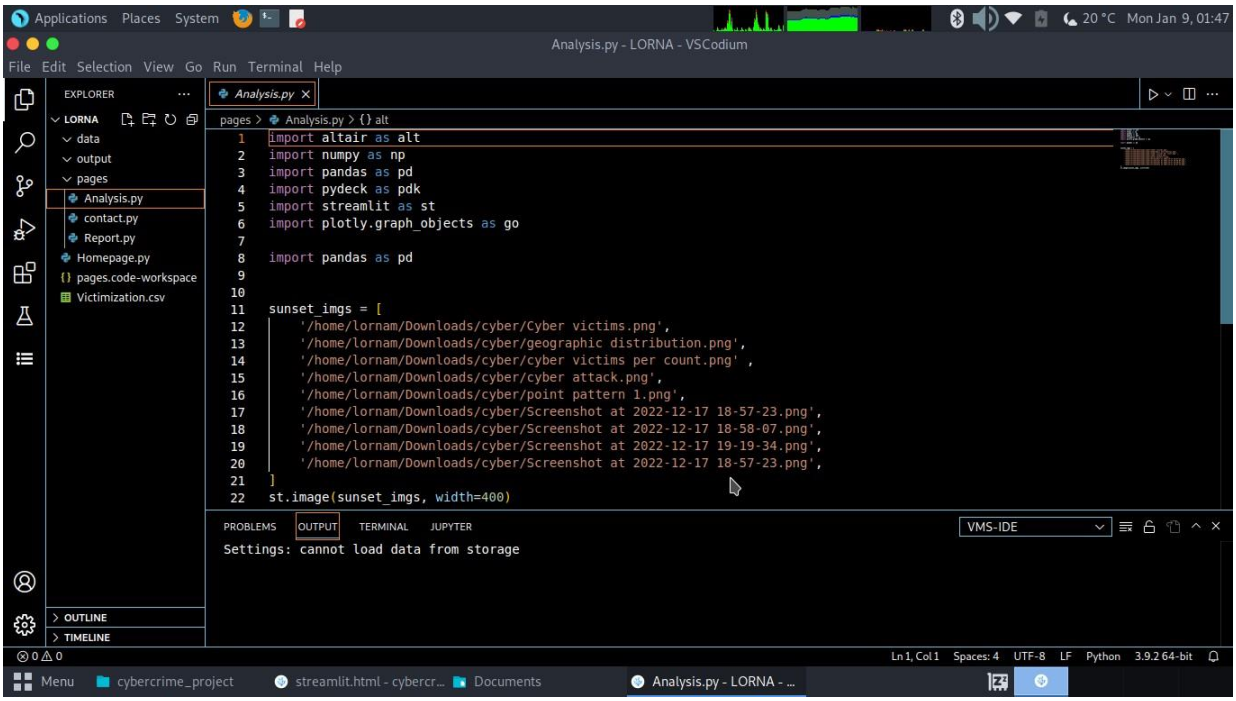
45. Zhuang P, Zamir T, & Liang H (2021). Blockchain for Cybersecurity in Smart Grid: A Comprehensive Survey. IEEE Trans Ind Inf. 2021 Jan;17(1):3–19.

Appendix

Below are the extra materials that were relevant to my study but could not fit in the main text of my write up. This consist of codes that were used to build both the frontend and backend of the web application in VS codium in a Linux environment.

Code for Streamlit pages:





F

Report.py - LORNA - VSCodium

pages > Report.py > {} st

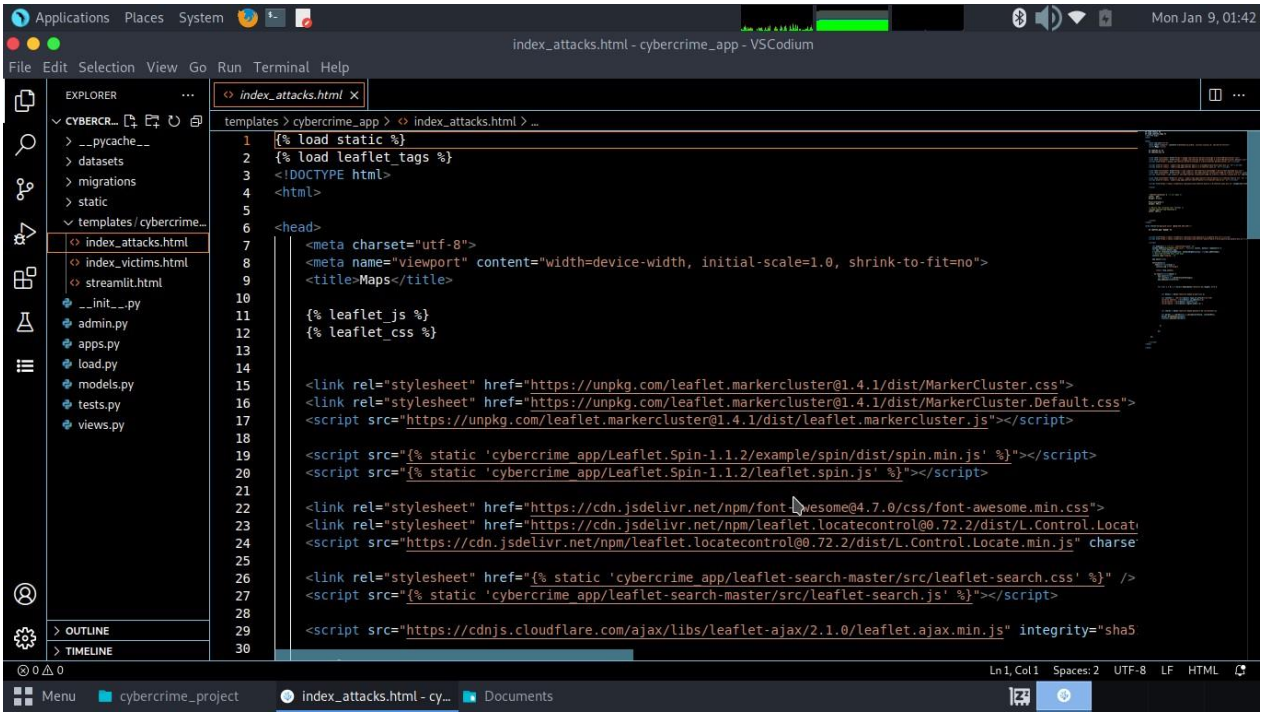
```python
import streamlit as st

st.title('REPORT INCIDENT.')

# Text input
name = st.text_input("Enter your name:")

Report_date_time =st.date_input("Enter the date")

# Number input
county = st.selectbox("Enter your County:", [
        "Baringo",
        "Bomet",
        "Bungoma",
        "Busia",
        "Elgeyo Marakwet",
        "Embu",
        "Garissa",
        "Homa Bay",
        "Isiolo",
        "Kajiado",
        "Kakamega",
        "Kericho",
        "Kiambu",
        "Kilifi",
        "Kirinyaga",
        "Kisii",
        "Kisumu",
        "Kitui",
        "Kwale",
```

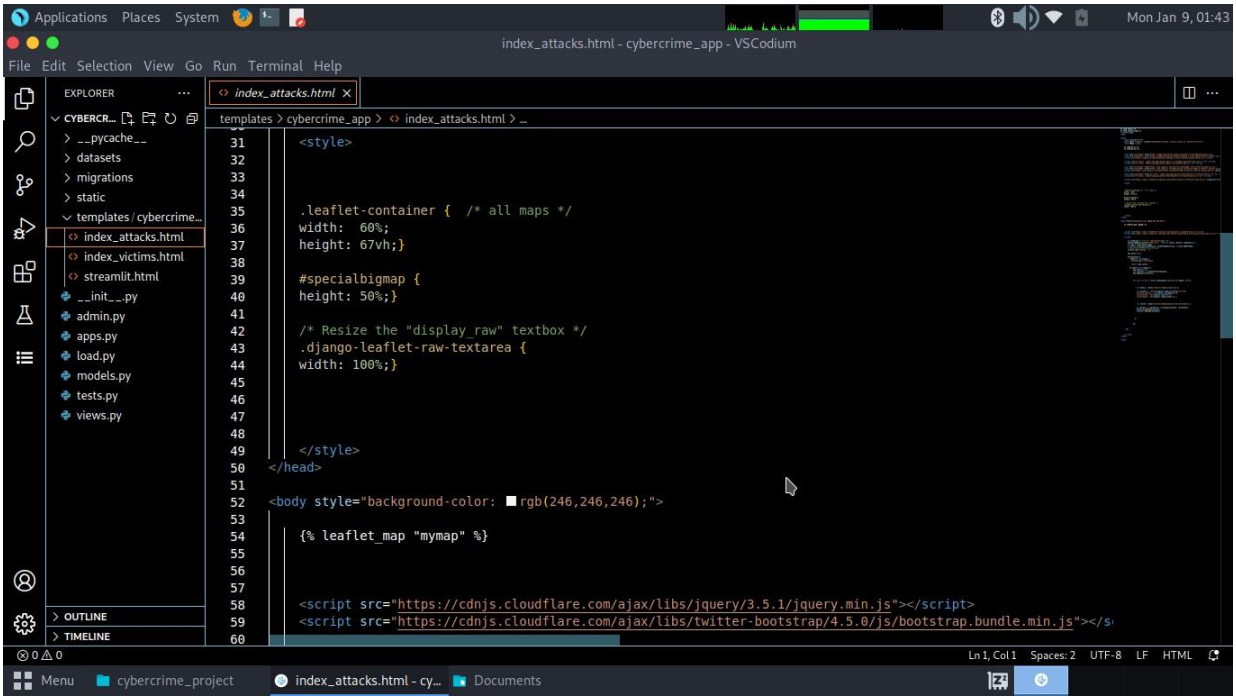
Report.py - LORNA - VSCodium

pages > Report.py > {} st

```python
region = st.selectbox("Enter your region:", [
    'Coast',
    'Central',
    'Lower Eastern',
    'North Eastern',
    'North Rift',
    'South Rift',
    'Upper Eastern',
    'West Kenya',
    'Nairobi',
] )

Address_location = st.text_input("Enter your Location:")

# Dropdown menu
gender = st.selectbox("Select your gender:", ["Male", "Female", "Other"])

# Checkbox

Type_of_attack = st.selectbox("Select the type of cyber crime attack:", [
    "Financial Fraud (Mpesa online banking)",
    "Botnets (attacker has access to a device and its connection)",
    "Phishing (fraudulent messages)",
    "Malware (malicious software)",
    "Spams (unwanted unsolicited digital communication sent out in bulk)"
])


# Submit button
if st.button("Submit"):
```

Codes for back-end Framework





models.py

models.py > COUNTY_CHOICES

```python
from pyexpat import model
from django.contrib.gis.db import models

# Create your models here.


REGIONS_OPTS = [

    ('Coast','Coast'),
    ('Central','Central'),
    ('Lower Eastern','Lower Eastern'),
    ('North Eastern','North Eastern'),
    ('North Rift','North Rift'),
    ('South Rift','South Rift'),
    ('Upper Eastern','Upper Eastern'),
    ('West Kenya','West Kenya'),
    ('Nairobi','Nairobi')
]

COUNTY_CHOICES = [

        ("Baringo", "Baringo"),
        ("Bomet", "Bomet"),
        ("Bungoma","Bungoma"),
        ("Busia", "Busia"),
        ("Elgeyo Marakwet", "Elgeyo Marakwet"),
        ("Embu",    "Embu"),
        ("Garissa", "Garissa"),
        ("Homa Bay", "Homa Bay"),
        ("Isiolo", "Isiolo"),
```

Applications Places System Mon Jan 9, 01:38

models.py - cybercrime_app - VSCodium

File Edit Selection View Go Run Terminal Help

EXPLORER

CYBERCR...
- __pycache__
- datasets
- migrations
- static
- templates / cybercrime...
  - index_attacks.html
  - index_victims.html
  - streamlit.html
- __init__.py
- admin.py
- apps.py
- load.py
- models.py
- tests.py
- views.py

OUTLINE

TIMELINE

models.py > COUNTY_CHOICES

```python
        ("Kericho", "Kericho"),
        ("Kiambu", "Kiambu"),
        ("Kilifi", "Kilifi"),
        ("Kirinyaga", "Kirinyaga"),
        ("Kisii", "Kisii"),
        ("Kisumu", "Kisumu"),
        ("Kitui", "Kitui"),
        ("Kwale", "Kwale"),
        ("Laikipia", "Laikipia"),
        ("Lamu", "Lamu"),
        ("Machakos", "Machakos"),
        ("Makueni", "Makueni"),
        ("Mandera", "Mandera"),
        ("Marsabit", "Marsabit"),
        ("Meru", "Meru"),
        ("Migori", "Migori"),
        ("Mombasa", "Mombasa"),
        ("Murang'a", "Murang'a"),
        ("Nairobi", "Nairobi"),
        ("Nakuru", "Nakuru"),
        ("Nandi", "Nandi"),
        ("Narok", "Narok"),
        ("Nyamira", "Nyamira"),
        ("Nyandarua", "Nyandarua"),
        ("Nyeri", "Nyeri"),
        ("Samburu", "Samburu"),
        ("Siaya", "Siaya"),
        ("Taita Taveta", "Taita Taveta"),
        ("Tana River", "Tana River"),
        ("Tharaka Nithi", "Tharaka Nithi"),
```

Ln 69, Col 6 Spaces: 4 UTF-8 LF Python 3.9.2 64-bit

Menu cybercrime_project models.py - cybercrime... Documents

models.py - cybercrime_app - VSCodium

```python
ATTACK_OPTS =[

    ('botnet','Botnet'),
    ('Financial fraud','Financial Fraud'),
    ('malware','Malware'),
    ('Phishing','Phishing'),
    ('spam','Spam'),
]

GENDER_OPTS = [

    ('Male','Male'),
    ('Female','Female'),
    ('Other','Other'),
]

class CyberVictims (models.Model):


    report_date_time = models.DateField(null=True)



    county= models.CharField(null=True,  blank= True, max_length=45, choices= COUNTY_CHOICES)

    region = models.CharField(null=True, blank= True, max_length=45, choices= REGIONS_OPTS)

    address= models.CharField(null=True, max_length=50, help_text="Type in the location / place where the \
        incident occurred in this field. Try to find out the exact location")
```

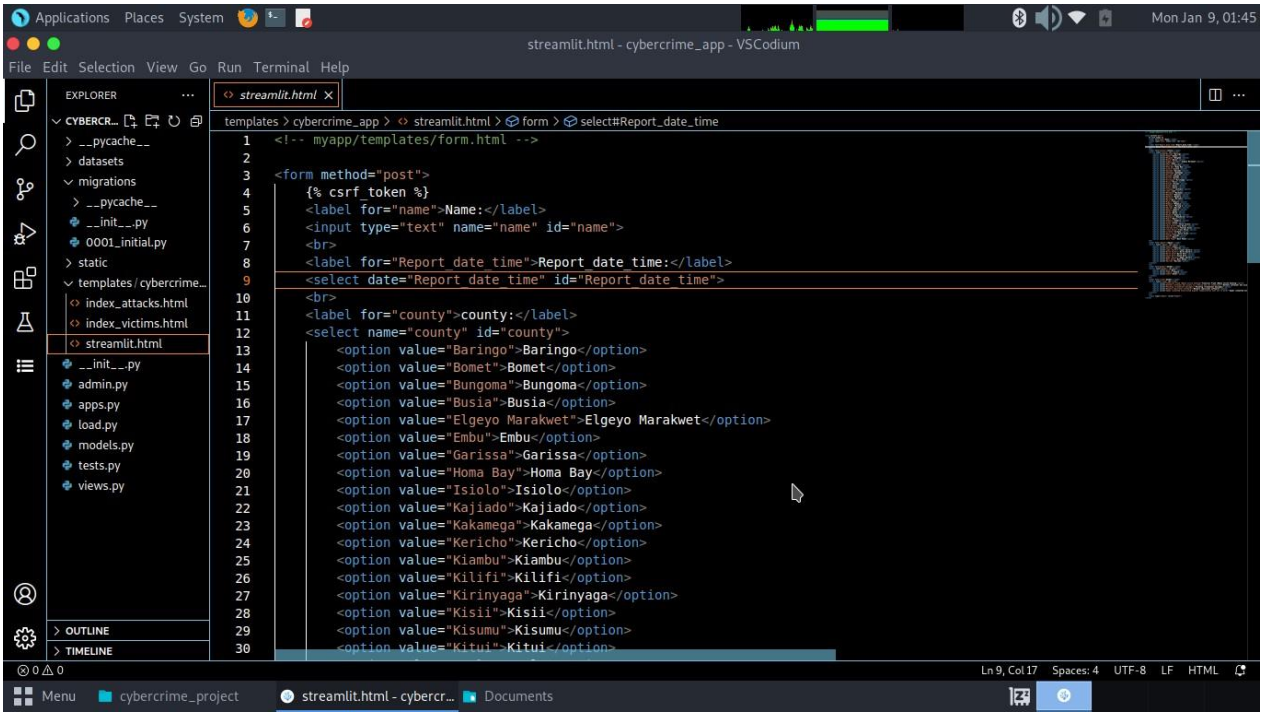
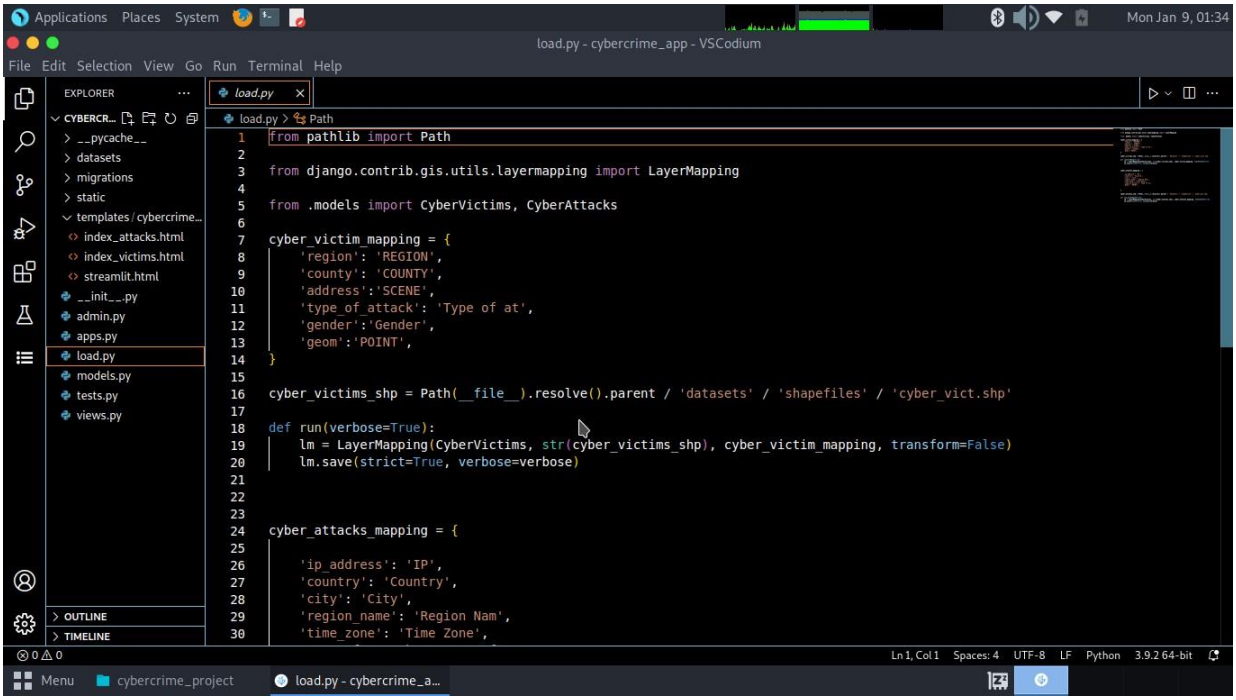
models.py - cybercrime_app - VSCodium

```python
    type_of_attack = models.CharField(null=True, blank= True, max_length=15, choices= ATTACK_OPTS)

    gender = models.CharField(null=True, blank= True, max_length=10, choices= GENDER_OPTS)

    geom = models.PointField(srid=4326, null=True)



    def __str__(self):

        return self.type_of_attack

    class Meta:

        verbose_name = "Cyber Victim"
        verbose_name_plural = "Cyber Victims"



class CyberAttacks (models.Model):

    ip_address = models.CharField(max_length=55, null= True)

    country = models.CharField(max_length=70, null=True,  blank=False)

    city = models.CharField(max_length=70, null=True,blank=False)

    region_name = models.CharField(max_length=70, null=True)
```

models.py - cybercrime_app - VSCodium

```python
class CyberAttacks (models.Model):

    ip_address = models.CharField(max_length=55, null= True)

    country = models.CharField(max_length=70, null=True,  blank=False)

    city = models.CharField(max_length=70, null=True,blank=False)

    region_name = models.CharField(max_length=70, null=True)

    time_zone = models.CharField(max_length=70, null=True, blank=False)

    geom = models.PointField(srid=4326)

    type_of_attack = models.CharField(null=True, blank= True, max_length=35, choices= ATTACK_OPTS)

    class Meta:

        verbose_name = "Cyber Attack"
        verbose_name_plural = "Cyber Attacks"
```

views.py - cybercrime_app - VSCodium

```python
from django.core.serializers import serialize
from django.http import HttpResponse
from django.views.generic import TemplateView

from .models import CyberAttacks, CyberVictims

# Create your views here.

def cyberattack_dataset(request):

    cyber_attacks = serialize  ('geojson',CyberAttacks.objects.all() )
    return HttpResponse (cyber_attacks, content_type='json')

def cybervictims_dataset(request):

    cyber_victims = serialize  ('geojson',CyberVictims.objects.all() )
    return HttpResponse (cyber_victims, content_type='json')


class CyberAttackView (TemplateView):

    template_name = 'cybercrime_app/index_attacks.html'

class CyberVictimsView (TemplateView):

    template_name = 'cybercrime_app/index_victims.html'

# myapp/views.py

from django.shortcuts import render
```

Connecting streamlit to Django