Crime Prediction in Nairobi using Time Series Analysis

GROUP MEMBERS

- **CINDY GLORIA OTIENO**
- **MAXWELL SIMIYU – I63/4274/2019**
- **CLAUDIA CHEPTOO**
- **FRANCIS MUTHEE NJARAMBA**
- **PAUL SAYIANKA**

GROUP SUPERVISOR

- **MS. ANNE WANG'OMBE**

UNIVERSITY OF NAIROBI

A RESEARCH PROJECT REPORT SUBMITTED TO THE FACULTY OD SCIENCE AND TECHNOLOGY IN FULFILMENT FOR THE REQUIREMENTS OF THE AWARD OF THE DEGREE IN BACHELOR OF SCIENCE IN PURE STATISTICS AT THE UNIVERSITY OF NAIROBI.

Date: June, 2023

DECLARATION

This Research project is the original work for the group members listed above and has not been presented for a degree in any other University.

Signature ____________ Date

Group Leader: Cindy Gloria Otieno

I63//2019

SUPERVISOR:

This project has been submitted with approval as the University Supervisor.

Signature ____________ Date

Ms. Anne Wang'ombe

Department of Mathematics

University of Nairobi

DEDICATION

We dedicate this project to our families, friends and to the Department of Mathematics at The Faculty of Science and Technology in the University of Nairobi. We thank them for their continuous support.

ACKNOWLEDGEMENT:

We would like to express our sincere gratitude to all those who contributed to the successful completion of this research. From the group members dedication and the group leader in that case. We are grateful to the Ministry of Health and Security in the County of Nairobi for granting us letters and permission to access sub county levels to get this data. We also extend our heartfelt appreciation to the Kasarani Police Station and Health Center for granting us access to their data and for their cooperation throughout the research process. Their willingness to share information and insights has been instrumental in our analysis. We also like thank the participants who willingly provided their time and cooperation during the data collection phase. Each person's contribution has been invaluable in enhancing the quality and reliability of our findings. Additionally, we extend our thanks to our research team supervisor Ms. Anne Wang'ombe for her dedication, support, and collaboration throughout the project even during uncertain terms.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

This paper presents a comprehensive analysis of crime trends in Nairobi, with a specific focus on data collected from Kasarani Police Station and Health Center. The objective of this research is to gain an understanding of the patterns and dynamics of crime in Nairobi. In particular examining the influence of socioeconomic factors on criminal activities. The study involved the collection of crime data over a specified period from 2018-2022 data recorded at the specific center. The goal was to collect as much information from all sub counties but it was limited by the deputy IG of police due to privacy of information. This projects utilized statistical techniques and data visualization tools to analyze and interpret the findings. From our analysis, it is revealed that there is a wide range of criminal incidents occurring in Nairobi such as; theft, assault, robbery, and other related offenses. By examining this temporal distributions of these crimes, we are then tasked with identifying peak crime periods and thus shedding light on potential underlying factors driving the various criminal behavior. Furthermore, the research explored the correlation between crime rates and various socioeconomic indicators, such as poverty levels, unemployment rates, and educational attainment in the population.

The outcomes of this research contribute to the growing body of knowledge on crime in urban settings and offer practical implications for policymakers, law enforcement agencies, and other stakeholders involved in crime prevention efforts. The findings underscore the importance of addressing socioeconomic disparities and implementing targeted interventions to address the root causes of crime in Nairobi. Ultimately, this research aims to contribute to creating safer and more secure communities for all residents of the city.

CHAPTER 1: INTRODUCTION

Resolving crime has been a challenging topic towards maintaining order in many countries. Eons have passed on and we are still taken back to the systems ran by Romans and Napoleon. The Mesopotamia Sistema Judicial was a significant milestone known to man housing judicial codes of King Ur-Nammu (2032 B.C) and legal systems such as Hammurabi (1792-1750 B.C). Most crimes back then were considered private faults against property interests of the rich and wealthy in the society. These systems have been reforming as years rolled on. Different societies sought independence and we had law systems based from religion to administration and much more. Crime solving has as seen has come a long way from the early days of policing to the current digital age. In the past, solving crimes was a laborious and time-consuming process that relied on witness testimony and physical evidence. Law enforcers would have to canvas neighborhoods, interview witnesses and analyze physical evidence to try and piece together the details of a crime. In the new age, advancements in forensics science in late 20th century revolutionized the field of crime-solving. The introduction of DNA and bio-analysis of evidence made it easier to start identifying suspects and linking them to crime scenes.

The digital technology has once again revolutionized how crimes are addressed. Law enforcement organizations may now use a significant amount of information to find intuitions in criminal activity. Thanks to developments in data science and technology. Law enforcement officers can actively prevent crimes from happening by training machine learning algorithms to recognize and categorize illegal conduct. There is a great deal of information and tools at our disposal which make a huge difference in solving crime at this age.

This study seeks to explore how statistical analysis and data science can be leveraged to reduce crime rates. We ask the question how the new age technology can be implemented in Nairobi to help reduce crime. The study will involve examining earlier crime patterns and trends in different known crime hot spots and identify the key factors contributing to crime in the city. Moreover, we seek to understand how effective existing crime prevention strategies are and propose new strategies for reducing crime.

Background of the study

Crime is a persistent challenge in many urban areas around the world and has become a constant menace to the authorities trying to fight crime. Nairobi, the capital city of Kenya is no in this case and has had its fair share on fighting lawbreakers. Despite various efforts by law enforcement agencies, crime rates in Nairobi remain high. The reports are stunning how crime rate since 2015 has scaled upwards even with huge spending on police force by the government. According to a report by The Kenya National Bureau of Statistics published on *Daily Nation*, the overall crime rate in Nairobi increased by 5.9% between 2019 and 2020. Topping this list we have crimes such as theft, robbery and assault. Another report by Nairobi Crime Research Center stated that at least 2 in 10 individuals in Nairobi have had an experience of the listed crimes with assault being reported the highest. As of November 2022, *The Star* reported in an interview with Dr. Lillian Munyua where she states that Kenyatta National Hospital receives a daily average cases of seven stabbings (Emmanuel, 2022). She also specifically states the specific zones where most of these incidents takes places and areas like Thika Road and Globe Roundabout are mentioned. In addition to the high rates of crime in Nairobi, the city also faces a shortage of medical professionals and healthcare facilities. Apart from the national hospital, private hospices and a few varsity medical centers, most of the clinics cannot take more than 50 inpatients a day. In a survey conducted by World Bank and World Health Organization (WHO), Kenya has only 0.2 physicians and 1.4 nurses per 1000 people. Which is well below the WHO's recommended minimum of 2.3 health workers per 1,000 people (Bank, 2022). Nairobi's hospitals and clinics frequently lack the necessary staff and resources to adequately care for patients, especially those with critical and severe injuries. This shortage is hugely significant on how authorities respond to the overwhelming crime activities around the city. The unavailability of paramedic ambulances to act on response on scenes of crime are also scarce. This situation of hit and miss medical attention leaves medical centers in Nairobi being often inundated with patients who have been injured in robberies, assaults, and other forms of criminal activity. This puts a substantial strain on the resources and capacity of medical facilities, which are often ill-equipped to handle such high volumes of patients.

We assume to correlate the high rate of rampant crime with the population of the area. While this may hold true, we are yet to prove for other factors such as unemployment and substance abuse as other contributors to the proliferating crime. It is worth noting Nairobi is one of the fastest growing cities in Africa with a population of over four million people. The city is constantly growing diverse-wise and in terms of demography. This growth poses a challenge on how Nairobi is managed by the municipal authority. Numerous challenges such as traffic congestion, pollution and fast growing crime rate come along with this growth in population. High population density can create an environment that is conducive to criminal activity especially in areas where poverty and inequality are prevalent. In a county where the Gross Domestic Product (GDP) per Capita of about $27,798 (CEIC, 2021), we have nearly 30% of the population living below the poverty line. This shows how the city's residents struggle to meet basic need such as food and accessing healthcare. Despite evidence of increased economic power, there is also an increased poverty rate. The high poverty rate in Nairobi can contribute to the city's crime problem. Individuals living in poverty are often more susceptible to engaging in criminal activity as a means of survival. The feeling of hopelessness among the poor contributes to a sense of frustration leading to violent behavior. Suppose we place some of the sub counties into light, areas like Embakasi have the highest population in the city and a high prevalence crime rate. But correlation doesn't mean causation therefore no conclusion is made with no statistical proof. Other areas include Kasarani having more than half a million population and a disturbing high rate of crime activities.

The absence of data and evidence-based methods to crime prevention is one of the biggest obstacles to reducing crime in Nairobi. In Nairobi, a lot of crime prevention measures are not based on reliable statistical and data-driven analysis, but rather on anecdotal evidence. As a result, the efficacy of these tactics is often unclear, necessitating the development of evidence-based crime preventive methods. Traditional policing strategies are reactive, which means they react to crime after it has already happened. This strategy has drawbacks because it doesn't address the underlying causes of crime and might not be successful in deterring it. Therefore, there is a need for proactive approaches to crime prevention that use data and technology to identify crime patterns and trends.

Looking around the world, new advanced technological methods have dawned ushering a new era of solving problems technologically. Decisions aren't made solely by people but by help of insights from data. Urban areas around the world are changing their management methods such as garbage collection, road maintenance and crime prevention strategies based on data fueled decisions. This cities are referred to as smart cities because of the way they utilize this technology (Yin, et al., 2015). Smart cities ensure there is complete transformation of a city turning it into a productive and manageable center where citizens find it habitable. Smart cities around the world have been experimenting with data-driven approaches to crime prevention, including the use of predictive analytics and machine learning algorithms. For example, in India, The Ministry of Housing and Urban Affairs in 2015 launched a comprehensive program to use data science and machine learning to improve infrastructure management, service providence and public safety (Kumar, Sonali, & Pradeep, 2018).The program is yet to completion but India is showing determination of completing these projects after being halted momentarily by COVID-19. This program involves the installation of thousands of CCTV cameras across the city, which are all connected to a central hub. Robust video analytical methods are used by metropolitans to solve problems facing the city. Similarly, in Chicago, the police department has implemented a predictive analytics program called the Strategic Subject List (SSL) (ChicagoPD, 2017). The SSL analyzes information from a variety of sources, including arrest records, social media, and criminal histories. They then pinpoint people who are most likely to be associated with violent crimes. The police department makes use of this data to concentrate its efforts on these high-risk individuals, including by offering social assistance to deter them from committing crimes.

In recent years, there has been increasing interest in the potential of data-driven approaches to address crime and improve public safety in Kenya. We have heard mentions of Konza City and its implementation by the *Vision 2030*. The framework behind Konza is integrating urban information and data science algorithms to come up with a network that supports efficient service delivery. The same can also be made out of Nairobi given that the progress in developing Konza has slacked down the years. There is great need and urgency that the city will be part of the technological revolution around the world and avoid losing status once other smart cities come of age. The field of statistics and data science offers a range of techniques that could be implemented around hot spots to track down people's activities. There are tools that can help to identify crime patterns, forecast future trends, and determine the effectiveness of crime prevention strategies. One of the promising applications of data science in crime prevention is the use of predictive analytics to identify areas that are at high risk of crime. This approach involves analyzing historical crime data from local law enforcement, demographic data, information from health centers and other relevant crime research centers. This information can be used to allocate resources and deploy law enforcement personnel to prevent crime in areas with a high crime rate. Other resources such as rescue ambulances can also be deployed to these areas in order to counter up with crime victims. Another promising application of data science in crime prevention is the use of machine learning algorithms to analyze CCTV footage and detect criminal activity in real-time. With this strategy, machine learning models are trained to spot particular illegal activity, such as suspicious conduct, drug usage or the presence of weapons. Alerts can be made to law enforcement agencies when these behaviors are seen so they can take action and stop the crime from happening.

Statement of the problem

Crime does not only affect the individual, it affects the community, society, or even the state as it affects its economic growth and productivity. A criminal act may have several consequences; physical, emotional, or even psychological. A 2011 Gallop poll that targeted the American population showed that there is public concern about crime and more than a third of the target population said they would be "afraid to walk alone at night". This is about 86 million adults. About 114 million adults are troubled that their homes could be robbed. Another percentage of adults fear getting murdered during a crime. This is an indication that public concern about crime may be higher than documented. The media also plays a role in the cause of unrest among the public by giving more coverage to violent crimes giving the impression that most crimes are violent and also profiling offenders.

Petty offenses like stealing are the most reported cases in Kenya. There has also been a rise in reported banditry, ethnic violence, carjacking, and corruption. With a crime index of 60.14, Kenya is ranked 23rd globally on the organized crime index. According to an UN-HABITAT survey, which gathered responses from over 10,500 residents in Nairobi, about 37% of the respondents were have been targets of robbery, and 22% victims of theft. These incidents occurred at least once during the previous year. The analysis shows that if this rate persists to the following year, 1 in 5 residents will be a victim of physical assault, 1in 5 will likely fall prey to theft and snatching and 2 in 5 could be victims of a robbery. The high crime rate has led to increased levels of poverty with little to no access of basic services

Most affected households are those from low income areas. This is so due to limited access to resources and general poverty prevailing therein. Understanding this relationship and knowing how it affects the social economic growth forms the basis to develop ideas and strategies on crime reduction in Nairobi City County

Objectives of the study

General objective

The general objective of the study was to highlight the rate of insecurity in Nairobi and show how statistical analysis and data science algorithms can be used to help reduce crime.

Specific objectives

- To analyze the patterns and trends of crime in Nairobi between 2018 and 2022.
- To determine the key factors contributing to crime in Nairobi.
- To evaluate the effectiveness of existing crime prevention strategies in Nairobi and come up with new strategies to reduce crime.
- To explore the potential of statistical models and machine learning algorithms in predicting and preventing crime in Nairobi.
- To understand how other smart cities like Abu Dhabi implement ML in reducing crime rates.
- Create a dashboard displaying the demographic information and areas that are susceptible to crime based on the real-time population information of the area.

Research questions

- What are the patterns and trend of crime in Nairobi between 2018 and 2022?
- What are the factors contributing to crime in Nairobi?
- What is the effectiveness of the existing crime prevention strategies and can we come up with new ones?
- How can statistical models and machine learning algorithms be used to predict and prevent crime in Nairobi?

- How do smart cities like Abu Dhabi implement ML in reducing crime?

Justification of the study

Crime has historically been known as a disruption to the socioeconomic development of a society. Despite concerted efforts by the government, civil society, and the international community to address run-away crime, police statistics point to a growing spike in crime commissions across the country. It was envisaged that the outcome of crime analysis would help facilitate sharing of actionable information on specific locations where crime is likely to occur and general crime trends and typologies across the Counties. The city fosters and strengthens the economic prospects of the whole country. A state of lawlessness could have negative impacts not only on the city itself but also on the whole country.

Years ago, the acts of being robbed or assaulted in broad daylight were rare and unthinkable, nowadays these petty crimes happen on daily basis in the streets of Nairobi.

Nairobi County has had an annual growth rate of 4% over the decades according to UN Urbanization Prospects. This large growing population poses a challenge to enforcement agencies to suppress criminal activities. Increased population is also associated with increased poverty rates as resources become limited. In an effort to reduce the associated crimes such as robbery, burglary, home breaking, possession of illicit brews and narcotic drugs, and physical assault among other criminal activities, we have to incorporate new advances in technology. We have seen it work in the fields of banking, marketing, and airlines.

The study is premised on the conviction that conducting research on the key elements of security will culminate in facilitating the dissemination of evidence-based information that will assist policy formulation and ultimately improve human security in the country. Crime analysis will make use of the knowledge of data mining and machine learning tools such as image recognition to obtain meaningful vast amounts of data and disseminate them to officers in charge to assist in their effort to apprehend criminals and suppress criminal activity.In developed countries, CompStat, a data-driven management technique which reduces and prevents crime, creates predictive models to analyze crime and arrest data which drives the police to areas that need stringent supervision. It has reduced the rate of crime by a rate of 75% in a span of 20 years. Our Security forces and Nairobi Crime Research Center have the databases populated with details of incidence reports, criminal descriptions, fingerprints, and individual mobile phone records. This valuable information can help to visualize and discover association rules of prevalent crimes happening around the city for better policy planning. The review will help future academicians and analysts who would like to partake in an analysis of the same.

Significance of the study

The findings provide valuable insights in different ways. Society has an increased expectation for police to be data-driven and use evidence-based practices. Crime analysis will provide better data quality and automation of data access for police will enable them to act swiftly thus safeguarding the hotspot areas. This will also help in budget formation as the crime statistics can show where more resources are needed, as well as where fewer resources are needed as a community grows safer. Crime analysis also allows police departments to review trends and adjust their staffing accordingly in order to counter act potential criminal activities.

The realization of social, economic, and environmental objectives in order to ease the transition toward more sustainable development will be evident. People will be engaging in their daily businesses without fear or anxiety of being attacked. This will also moralize foreign investors to set up firms in our city which will generate revenue in country. The efficiency of this model will push governments to create real-time crime centers that facilitate crime analysis for immediate incidents. This will be a source of employment.

The community would benefit from community-oriented policing approach which will create awareness about criminal activities taking place in the area so they can be avoided. By making crime data public increases transparency and opens scrutiny to the criminal justice professionals. The community may decide to improve public areas like parks and schools as they may be the targets due to poor designs which need to be changed. This builds the relationships between the departments, government entities and the community.

Scope of the study

With new technological innovations being developed, they can be beneficial in prevention and fighting of crime. In view of this context, the study explores how we can implement statistical analysis and data science techniques to aid in crime reduction. The study involves an analysis of crime committed over the time period ranging from 2015-2022 and the factors contributing to those crimes. The geographical location this research will be one sub county in Nairobi County, Kasarani Sub County.

Limitations of the study

The study was subjected to certain limitations. The apparently narrow scope of the study was a result of insufficient funds and inadequate time, which made it hard to acquire data for Nairobi County as a whole, hence restricting the research to one sub counties. The researcher had trouble collecting data from the police station and the hospital. The obstacle was alleviated by offering a letter that clarified the academic purpose of using the data and assured the maintenance of confidentiality.

Organization of the study

Chapter one contains the introduction and background of the study. It also highlights the statement of the problem, objectives of the study, the research questions, justification, significance and limitations of the study.

Chapter two presents the literature on the trends of crime in Nairobi, factors contributing to the crimes, how statistical model can be used in crime management and how smart cities use machine learning in crime management. At the end of the chapter there is a summary and a conceptual framework of the study.

Chapter three contains the methodology used for the research. It highlights the research design, the target population, the sample size determination, the sampling techniques and procedures, data collection tools, validity and reliability of the data, data collection methods and procedures, a summary of the data analysis methods and procedures, and the ethical consideration.

Chapter four highlights the findings, the data analysis and interpretation of the findings. Tables, pie charts, histograms, and a statistical model was used to present the findings. The chapter also highlights the demographic characteristics of the area, the residents' age group and age distribution. At the end of the chapter there are key informant findings gathered from the target population.

Chapter five presents the summary of the study, conclusion, recommendations and suggestions for future research. The objectives of the study was to observe the patterns and trends of crime in Nairobi, analyze the factors contributing to crime in Nairobi and create a statistical model that would predict crime for the purpose of crime control. The overview is based on the stated objectives.

CHAPTER 2: LITERATURE REVIEW

Empirical literature Review

Patterns and trends of crime in Nairobi County.

Crimes do not occur randomly or uniformly across neighborhoods, or social groups, or during an individual's daily activities or during an individual's lifetime. There are hot spots and cold spots; there are high repeat offenders and high repeat victims. In fact a very small proportion of people commit most of the known crimes and also account for a large proportion of victimization. Thus the argument of uniformity and complete randomness of targets and victims seems indefensible and no longer plausible. Bar fights occur with a greater frequency on Weekend nights than on weekday afternoons. Analyzing crime requires concepts and models that can be used to account for the patterned non-uniformity and non-randomness that characterizes real criminal events.

Crime activities are divided into three different categories. Property crime, crime against Person, and hate crime. Crimes against persons include Murder, aggravated assault, rape, and robbery are commonly referred to as personal crimes. Property crimes do not entail bodily injury, and they include burglary, robbery, vehicle theft, and arson, to mention a few. While hate Crimes are crimes that occur when someone is motivated by politics or personal feelings based on tribe culture or belief (Braga 2016).

Criminal activities have been a threat to national, regional, and international governments and only a few researchers in sociology and criminology have ever identified the trends of criminal activities on the safety of persons and property. The trends and patterns can only be handled individually due to difference in jurisprudence, crime classification and reporting. In most cases national crime statistics only concentrate on selected crime types, different data gathering method and different survey methods.

The Nairobi Annual Crime Observatory report established more criminal activities hot spots in urban slums than in residential estates in Nairobi City County, whereby muggings, robberies and assaults have been on the rise. Some of the common places within Nairobi with increased crime being; Kasarani, Kibra, Kamkunji, Uhuru Highway, Nyamakima, Gikomba, National Archives, and Kayole among others. This report was supported by Kenya National Police Service reports which are released annually. They indicate that inhabitants of the city live in continuous fear of insecurity. This is because the crime rate has been on the rise for the last ten years, with a 6.2 percent rise in crime from 2011 to 2020.The changing patterns of criminal activities are associated with mushrooming of unplanned settlement such as Kibra, Mathare and Mukuru Kwa Jenga. This slums have presented inhabitants with limited opportunity for legal employment which create helplessness and hopelessness which drive individuals to theft and robbery in the most affluent neighborhood.

The target population was drawn from Nairobi County which has a population of 4.397 million. The target sample for the study was drawn from one sub-county ie Kasarani. The researcher used secondary data that were collected through a review of police reports records, National Crime Research Center (NCRC) studies and Police Annual Reports on crime. The researcher visited the police stations to check on documents such as Occurrence books and Cell registers to gain insight into the types of criminal activities common in the Nairobi City County. A crime report was generated and the results are as follows;

In **2017**, Nairobi County recorded 7,434 cases which was the highest in the country in terms of county crime analysis. In **2018**, Nairobi County recorded the highest number of cases reported to police at 7,128. The largest number of cases reported in Nairobi involved Offenses against Persons at 1,513 cases (Assault 1,243, Creating Disturbance 216 and Affray 54). The highest number of Assault cases reported is an indication that most offenses are between persons known to each other. In **2019**, Nairobi County recorded the highest number of cases reported to police at 8,246. In **2020**, Nairobi County recorded the highest number of cases reported to police at 5,844. The largest number of cases reported in Nairobi involved Offenses against Persons at 1,080 cases. The highest number of Assault cases reported (902 cases) is an indication that most offenses are between persons known to each other. In **2021,** Nairobi County recorded the highest number of cases reported to police at 6,686. The largest number of cases reported in Nairobi involved Offenses against Persons at 1,373 cases. The highest number being of Assault cases reported (1,129 cases) is an indication that most offenses are between persons known to each other. According to the Nairobi County Crime Outlook 2020 report, the most prevalent crimes are; Stealing with 53.3%, possession of narcotic drugs and illicit brews had 53.3% and 49.3% respectively, burglary and house breaking at 47.6% and robbery violence and gender based violence stood at 44.6%.

Factors contributing to crime in Nairobi County

Kenya has struggled with many challenges to its development since attaining independence, the toughest among them being Crime, having an upward trend. Nairobi being the most criminalized urban area in the country due to high population and social inequality has seen a high increase in crime rate. The Economic survey 2021 by Kenya National Bureau of Statistics (KNBS) reported that the number of crime in Nairobi rose by 16.6 percent from 2020. The rise was attributed in part to the difficult economic period that was brought about by the restrictions imposed to curb the spread of corona virus, which led to job losses, limited economic opportunities and increase in the cost of living. Crime prevention and reduction is a major challenge for the government and private sectors, hence the importance of identifying the causes of crime. Crime is a complex problem that arises from a variety of sources, but unemployment, poverty, drug and substance abuse are the main causes. Child abuse and neglect and peer pressure also play a role in criminal activities. Each of these views provides insight into the cause of crime, but it appears that none can be isolated.

In Kenya, the unemployment rate measures the number of people looking for a job as a proportion of the work force. According to macro trends, unemployment rate for 2021 was 5.74 percent; a 1.49 percent increase from 2018. 5.7 percent of Kenya's labor force was out of work in 2021 (World bank) , United Nations Development Program (UNDP) classified Kenya as the country with the highest unemployment rate in the Eastern African region having the percentage of unemployed twice the 2.7 percent of East Africa average. Ineffective education system and low level of education has been the primary cause of unemployment. The system teaches students to be good in theoretical knowledge instead of providing knowledge that enables them to be job creators and put so much emphasis on white- collar jobs during conventional training, undermining the mentality of young people who now think they can't make it with without employment opportunity. Also a great deal of Kenyans lack certificates in higher education excluding them in job consideration due low level of education.

Despite Nairobi having strong economic growth over the years and generating job opportunities, they have mostly been low paying, informal and the job opportunities coming at a low rate that is unable to absorb the rapid population growth. The youths are hit the hardest by the lack of employment (regardless of them having university degrees and other qualification) compared to their counterpart above 35 years recording unemployment rates of 2.8 percent with 20-34 years recording unemployment rate of 10 percent (KNBS, 2019). Regardless of the type of unemployment whether involuntary, disguised or frictional unemployment the reality remains: People in Nairobi especially in the poor areas are resorting to crime due to unemployment.

Poverty has a great influence on the crime rate in Nairobi. A substantial number of people in Nairobi experience relative poverty, in which they lack the minimum amount of income necessary to maintain the average standard of living in the society. In 2022, 17 percent of Kenya's population lived below the international poverty line (1.90 us. Dollar per day), this indicated that over 8.9 million Kenyans were living in extreme poverty (World Bank). Poverty in our country still remains relatively high compared to other lower income economy countries. Nairobi is dealing with rapid urbanization, yet the gap between the rich and the poor is growing wider with 60 percent of population now residing in slums and occupying just 6 percent of the land NCSS reports. Crime is widely spread in urban slums such as Kibera, crime surveys done by the Security Research and Information Centre(SRIC) with the Government of Kenya and United Nation Development Programme (UNDP) Kenya reported, 98.8 per of the respondent that undertook the survey witnessed crime in the last three month period of the study (Borgen project, 2022). The determinant of crime especially in the urban slums of Nairobi is poverty, urban poverty occurs due to inequality, corruption, lack of opportunity. There is extreme in equality in Nairobi despite the economic growth, the wealthy and the middle class individuals are gaining from the county's economic performance, but the economic prosperity is failing to benefit the poorest. The Inadequate resource of allocation affects the urban poor; the lack of infrastructure, poor provision of sanitation and good health care foster a deprived quality of life. Corruption also doesn't help the case as often it accompanies centralization of power; leaders fail to be accountable to those they serve, this inhibits development, when the leader takes in the money allocated for the public to benefit them individually.

Equally there is a correlation between drug and substance abuse and crime activities. The spike in drug use rate has been witnessed in almost every part of the country, particularly in urban areas because of its easy access. NACADA report stated that, 18.2 per cent of Kenyans aged 15-65 years are currently using at least one drug or substance abuse(DSA); 12.2 per cent are using alcohol; 8.3 per cent using tobacco; 4.1 per cent using miraa/khat and 1.0 per cent using bhang (NACADA, 2017). The rate of drug use, for the most part illegal drugs has threatened to sabotage societal advancement. There has been overwhelming evidence that drug use count-ermines security in many countries in the world, leading to major threats to political and social stability and economic developments. Drug abuse is often brought about by involvement in negative peer group, family problems, influence of culture and society, being out of school also increases the risk of alcohol and drug abuse. Although drug and substance abuse (DSA) affect the general population, Kenyans youth are at risk the most, with male leading; based on survey done in Nairobi urban slum adolescent male are more likely to consume alcohol and use illegal drugs 20 times and 5 times respectively that the girls (Science direct). The social and criminal justice problems that are frequently linked to drug and substance abuse; drug trafficking, gangs, prostitution, property crime, robbery and the growing number of youth homicide.

Child abuse and neglect is not only a cause of crime but a crime itself. Child abuse and neglect are commonly reported as frequent concern in cities (25 per cent) than in rural areas (20 per cent) and the matter of children not attending school the same 5- percentage point gap (Afro barometer survey). Parenting practices have been known to be important in determining the involvement of a minor in delinquent or criminal activities. It is importance to note that not all children who are neglected will be involved in criminal activities, but this is a risk factor that should be taken into consideration.

Although peer pressure is not a major factor contributing to crime, it can and it does lead to criminal activities. Peer relation is a recurring phenomenon in criminal behavior in most cases pertaining adolescents. Though teens experience peer pressure, only a number of them give in to the negative peer pressure, and are strongly influenced by their delinquent peers. Teens affected by negative peer pressure acquire immoral habits such as Drug and substance abuse, drug dealing, prostitution; the peer influence also causes school dropout which all in all leads to criminal activities.

Easy access to weapons, broken families, politics and gender supremacy being tertiary causes also evokes crime occurrence, but are not the root cause of the problem. If all this facets are allowed to worsen and grow in a community they create an environment that intercedes with the growth and productivity of the people in the society hence generating criminal behaviors.

Statistical models and machine learning algorithms in predicting and preventing crime.

Machine learning (ML) is a subdivision of artificial intelligence and computer science. In machine learning, data is analyzed using statistical methods to create an algorithm which is used to make classifications and predictions. ML has been an important tool to researchers when it comes to studying crime patterns and predicting crime before it happens. This is crucial as it protects lives of those who would have fallen victim to violent crimes and property damage. Extensive research has been conducted on the topic and there has been enormous steps into crime prediction using machine learning. Although the primary goal is to predict crime, the datasets and methods the researchers used that are applied are multiple. Crime is a rising pandemic that affects the country's social and economic development (Matekere, 2021). Social and economic issues are some of the factors contributing to crime and these patterns can also be used as data in crime prediction which will reduce crime rates in our communities (Hajela, 2020).

There are also other technologies of AI that enable researchers give accurate predictions. The divisions include natural language processing (NLP), speech vision, expert system, and robotics. ML creates algorithms which are used to analyze past data while NLP is a division of AI that studies the interaction between computers and the human language. There are five steps in NLP to ensure the computer comprehends text and spoken word the way humans can. Computer vision is another subfield of AI that allows computers collect data from videos and photos and act or make founded recommendations from the data (Aggarwal, 2021). Robotics is the use of robots to do errands without involving humans, it is used in AI to create systems that mimic human’s decision making and learning ability. A system that uses artificial intelligence to solve problems is called an expert system. Researchers need a combination of these technologies to give clear and reliable predictions, to increase efficiency of their research materials and save on time.

The nature of crime is an important aspect to consider in order to have an accurate crime prediction (Elluri, 2019). These are aspects related to the crime and they include gender, assailant(s) age, victim(s) age, location, education status etc. These aspects assist in giving a clear picture of the crimes and gives a more accurate prediction. Data is an important tool in ML. Local authorities have tons of data on reported crime that can be used be for analysis (Yuki, 2019). Some governments have invested in the research and given researchers access to this reports while in other areas access to this information is restricted. Several crime predicting models have been proposed by many researchers (Pratibha, 2020). There have been theories from crime pattern that suggest that offenders are more likely to commit crime on familiar streets. Streets they have struck before (Jalil, 2017). Other theories suggest that mapping hot-spots will enable researchers explain the regular crimes in that neighborhood (Kadar, 2019). A hot spot area is an area more prone to crime and is generally considered unsafe.

There has been a rise in reported crime in Nairobi which prompted researchers to study the patterns and look into the factors promoting crime. There have also been some attempts to analyze the data using statistical methods and to predict future crime occurrences. In a study that used Bayesian theory and a spatial-temporal Bayesian model to analyze the crime patterns in Nairobi, particularly robberies, showed that poverty and unemployment rates were significant factors to the rise in robbery incidents. There have been extensive advances in Bayesian methods when it comes to research, this has been key to Kenyan researchers in crime prediction and management. The study mapped hot-spots in Nairobi and gave a spatial distribution of crime in Nairobi and some studies explore the data to predict spatial patterns. The government has made an effort by increasing security expenditure from 7.5% in 2015 to 7.9% in 2018. This money goes to increasing the number of security personnel and research to better crime prevention strategies. While there have been evident steps in crime prediction in Nairobi, further studies should be made in the distribution of crime in the city.

One of the challenges encountered is data collection. Some governments and local authority have not made these records available to the public, some have not been properly documented while others not properly stored. The rise of interest in this field has been attributed to the availability of data which prompted research on the topic. Another challenge is the availability of the software needed in making evaluation. Such software include; PreCobs, PredPol, Crime anticipation system, and Hunchlab. These are recently designed systems and they have their limitations. However, the systems have been used to predict crime and provide crime precluding strategies (Carvalho and Pedrosa 2021). One of the most important factor about research in any field is the validity of the data. Efforts have been made by researchers to ensure the validity of the data used in this field, efforts which include automated searches instead of manual reading of printed journals, using a query design on the data before executing it on databases, conducting thorough data collection methods by using effective research questions and updating records during the review. Protocols have been put in place by a team of researchers called the SLR protocol to ensure validity of the data collected.

The ML models are subjected to certain limitations which include; huge reliance on previous data which may not always be accurate, the lack of linearity between the urban metrics and the population size (Alves, 2018), the availability of resources and data to be analyzed, and some technical issues like data storage and system performance.

Implementation of data science techniques to curb crime across different cities in the world.

Since the turn of the century, technology has improved immensely and has affected day to day human lives. From analog devices performing simple functions to complex digital devices that power up our homes. We simply can’t ignore the impact of these devices on how we perform tasks. Jobs have been much simpler and get done by time. This led to a need for automation, where repetitive tasks would be programmed to be performed by machines. Automation has been a major part of the technological revolution since the onset of 20^{th} century. There were various simple applications in automation's initial phase such as manufacturing and assembly-line production (Frohhm, et al., 2006). This led to development of software and hardware that automate complex tasks. The internet and the proliferation of smart devices in the 21st century have further expanded the potential applications of automation, making it possible to automate tasks ranging from simple household chores to complex industrial processes.

One notable field that has benefited from the impact of these technological advancements is the field of crime prevention. The development of data science and Artificial Intelligence, AI has made it possible to analyze vast amounts of data from various sources such as CCTV video footages and social media data. Image-subject identification in video analysis and sentimental analysis in social media data make it possible to pull strings and make crime identification models to carry out conclusions on criminal activities. We can look at how Nairobi can be transformed into a data driven city, where every single piece of demographic data is valued and the goal is to maximize every insight to curb crime. Several measures can be carried out in order to make Nairobi a smart city. Creating an extensive strategy for gathering and managing data from numerous sources, such as CCTV video and social media data. In order to do this, a large investment would need to be made in the infrastructure for data collection as well as in the creation of complex algorithms for data analysis. Large security cooperation with law enforcement organizations and other stakeholders would also be essential in such transformation.

Several cities across the world have already implemented data science techniques to curb crime. For instance, in New York, the New York City Police Department (NYPD) deploys analysis on CCTV footage and social media data to determine and predict crime (Nitsure, et al., 2020). Similarly, in London, the Metropolitan Police has implemented a system known as PredPol which utilizes data science techniques to predict crime locations based on past crime patterns (Utset, 2016). Other countries such as South Africa have embraced the PredPol in fighting crime and has proved beneficial to the police force. One notable city behind all this is Abu Dhabi. In Abu Dhabi, data science has been a backbone to tackle crime through the deployment of a program known as "Hassantuk".

Abu Dhabi, the capital of the United Arab Emirates, is renowned for the economic stability and the cultural diversity. Abu Dhabi has been able to maintain its position as the safest city in the world for the fourth consecutive year through the utilization of new technologies in the city's security measures. The introduction of artificial intelligence, cameras, sensors, drones, and smart traffic systems have contributed immensely to enhancing Abu Dhabi's security. The city's police force uses monitoring and control devices in public and private facilities to prevent crime and maintain public security. In addition, the police use smart glasses, which are augmented virtual reality glasses equipped with a micro-camera that uses AI to automatically scan hundreds of faces and notify authorities when it detects a wanted person (Noori, et al., 2020). The police also use drones to monitor traffic flow, detect congestions, and direct patrols at hot spots.

As reported by Khaleej Times in an interview with Major Eng. Ahmed Surour Al Shamsi – head of Safe City Department at Abu Dhabi Police, he reveals on the significant step the technology has helped reduce crime in the city. Since the introduction of AI, crime rate regressed by 57% while traffic fatalities decreased by 4%. The department ensured sensors and cameras have covered 85% of the city’s infrastructure. This wide coverage ensured there is a reduction in response time in getting to accidents spots or emergencies. Al Shamsi further stated that the system recommends appropriate locations for deploying patrol vehicles based on vulnerability and history of crime occurrence. If there’s a dangerous driver on the road, the operation rooms immediately sends out a patrol vehicle to corner the driver. Another strategy employed by the police is the development of smart gates. There are five smart gates which communicate with each other providing alerts to drivers on fog, rain, accidents or traffic congestion. The Khaleej Times report stated that the system can detect errant drivers over-speeding, those using mobile phones in traffic and other violations. The Abu Dhabi Police's current focus is on the utilization of advanced drone robotics to enhance the capabilities of response to Chemical, Biological, Radiological, and Nuclear (CBRN) incidents (Vodák, et al., 2021). They are working with international partners and universities on how to improve on the current research and development and contribute to this knowledge field.

The use of smart technologies, such as ML, in law enforcement is not a new thing. But its integration with cameras and sensors is innovative. Abu Dhabi has shown that its milestone ahead and other cities should implement from it or move a notch higher. For Nairobi to reduce crime rates adversely, we should borrow a leaf oversees and try to implement it. These model to run a metropolis allows police to gather information faster and more efficiently in turn leading to quicker response times to crime scenes. However, there are concerns over AI use and data privacy. Citizens are subjected to constant surveillance which at some point makes them feel uncomfortable with their gadgets. Another risk is possibility of biasedness, for instance, there is a possibility of biased decision-making based on the information fed in the model. If the parameters in the model are fed to target a certain ethnicity or a group of people, then the target will be mostly the ones focused by the model. The system's accuracy and reliability are also dependent on the quality of the data fed into it. All in all, it is essential to ensure it the system is deployed then all ethical and legal frameworks are in place to regulate the use of AI in law enforcement.

Theoretical Framework

The study will be informed by rational theory by Jeremy Bentham (1789) and learning theory by Clinard, Quinney & Wilderman (2014).

Rational Theory.

This theory suggests that individuals engage in criminal behaviour after making a rational decision based on the cost-benefit analysis most importantly, the goal is to maximize the benefits while minimizing costs. Based on this analysis. This theory suggests that people in areas experiencing high crime rate are more likely to engage in criminal behaviour if they perceive that the benefits would outweigh the costs. This brings the aspect of rationality which assumes that individuals are rational actors who make decisions based on logical considerations. Cornish & Clarke (1985) came up with this theory to assist in thinking of how to prevent crime using this knowledge.

According to Cornish &Clarke (1985) a criminal makes two major types of decisions; the crime involvement and crime event decision. A crime involvement is where an offender has decided that crime is the only way to make it in life. Factors that might influence this decision are the social characteristics such as social class, particular background and temperament of the offender. The other decision is the crime event decision where an offender thinks and makes a decision on where to offend, what to offend and how to go about it. There are specific factors that influence the decision to commit a criminal offense e.g what types of costs matter & what factors affect perception of those costs (Cornish& Clarke, 1986)

Jeremy Bentham (1789) in his Hedonistic calculus states that people will weigh the possible pleasures from committing the crime against possible pain from punishment and act accordingly. According to the deterrence theory, potential costs of committing a crime such as fine, imprisonment can deter individuals from engaging in crime. If the potential costs are not perceived as severe enough, the individual may still choose to commit crime. Bentham's idea is that motivations for action whether criminal or otherwise are based on the desire to maximize pleasure and minimize pain and thus punishment has to be applied rationally if they are to influence people's perception of pain and pleasure associated with their choices.

In the deterrence approach Beccaria (1764) proposes that effective punishment need to be swift and certain. However, punishment effects are conditional in that the less the crime the less the punishment is effected. For example petty crimes like disturbance of peace in public areas might just be meted by a fine while hard-core crimes like robbery with violence can earn an individual a long jail term sentence. Thus an individual would assess the benefits and costs outcome and make a decision. The benefits that an individual might gain from crime include money, emotional satisfaction, and the excitement of the crime and other times it can be the status or respect gained from being a hard-core criminal feared by many including the state. The costs include but not limited to punishment, rejection by significant other, stigma or even sanction and loss of legitimate income.

According to Lawrence E. Cohen in his book "The Illegal City: A study of the Urban Informal Economy" (1979), argues that criminals engage in cost-benefit analyses when deciding whether to commit a crime or not. The availability of criminal opportunities is a key factor in explaining high crime rate in some areas. Access to opportunities for committing crime such as unsecured buildings or vulnerable individuals may likely lead an individual to engage in crime.

Learning Theory

According to this theory, individuals learn criminal behaviour through observation and interaction with others in their social environment. In effect, criminal behaviour is premised on the notion that it is a learned behaviour (Clinard, Quinney & Wildeman, 2014). The theory assumes that at birth no one has the motivation to commit crime. Thus an individual commits crime through interactions between thoughts, behaviour and the particular environment.

Learning theory has several concepts that shed light on high crime rate. Differential association theory (Sutherland 1947) suggests that criminal behaviour is learned through interactions with others who hold favourable attitude towards crime. Through interactions with others, individuals learn the values, attitudes, methods and motives for criminal behaviour. The interactions according to Sutherland happens through communications either verbally or through gestures. This means that an individual is more likely to engage in crime if the people they associate with are crime-oriented. The main part of this concept is that it occurs within intimate personal groups. Sometimes these persons are close friends or even relatives and as such the bond is easily created and would greatly influence the behaviour of an individual.

Another concept of learning theory is the cognitive learning theory. Here learning is a mental process in which an individual is involved in problem-solving reasoning and decision-making. It looks at how individuals inclined to commit a crime vie the moral constraints and fear of punishment. Ronald Akers proposes the principle of operant conditioning of how criminal behaviour is learned. According to this principle, an individual behaviour is modeIled through observing the actions of others. For instance if a respected person engages in criminal behaviour like stealing, those who look up to him might be influenced. According to Sigmund Feud, an individual turns to crime in response to unconscious mental process resulting from early childhood experiences.

Summary of the literature review and research gaps

Author & Year	The focus of the study	Findings	Research gaps
Sila and Dr Masiga, 2023	The study sought to provide the framework for understanding the frequency of crimes on annual basis.	The study established that stealing, burglary and housebreaking were the most prevalent crimes.	The study did not cater for unreported cases.
Sabastian (2020)	The aspects that contribute to criminal activities in Nairobi from time period of 2018 to 2022.	Criminal activities occur due to unemployment, poverty, drug and substance abuse, child abuse and negligence and peer pressure.	Only the major cause of crime were intensively reviewed.
Noori Negar, 2020	Implementation of data science techniques to curb crime across different cities in the world.	The study showed that Abu Dhabi is the safest city around the world. The city has invested on having city's utilities ran by	The system implemented is more focused on traffic control around the city. Even though law enforcement

		Artificial Intelligence models. There is a 5 gate system in the city that enables effective communication between different essentials such as traffic, waste management and water supply.	officers are involved, they are tasked to find drivers breaking or slowing the movement within the city
He J, Zheng H (2021)	How statistical models and machine learning algorithms in predicting and preventing crime.	The study gave a background on machine learning algorithms then showed how the algorithms can be used in predicting and preventing crime. The study also mentions the challenges and limitations of ML in crime management.	Further studies should be made in the distribution of crime in the city.

Table 1: Summary of Literature Review and Research Gaps.

Conceptual Framework

The conceptual framework shows the relationship between the independent, intervening and dependent variables. This is important as it shows how your variables come together and helps in illustrating coherent findings.

CHAPTER 3: RESEARCH METHODOLOGY

Research Design

Research design is the specific procedure involved in research process: data collection, data analysis and report writing (Creswell, 2014). In this study, the researchers chose a descriptive qualitative and quantitative research design. This type of research was chosen in accordance with study objectives that tend to describe factors contributing to crime in Nairobi and getting to know the effectiveness of existing crime prevention strategies in Nairobi and come up with new strategies to reduce crime. Descriptive research is useful for a researcher to describe and interpret things and trends that are prevalent in a study environment. Also the in depth knowledge gained from a descriptive research can be useful to inform future policies in the area of study. This also is in line with the researchers 5th objective, "How smart cities like Abu Dhabi implement machine learning in reducing crime rate.

Study site

Kasarani is a suburb located in the northeastern part of Nairobi, Kenya at 1.2254° S and 36.8976° E. The population of Kasarani is estimated to be around 200,000 people, with a majority of residents being of the Kikuyu tribe. The sub county houses several government facilities including sub county police headquarters and a health center. Kasarani has been known to have higher crime rates compared to other parts of Nairobi, with theft and robbery being the most common types of crimes reported. The area is characterized by high levels of poverty, unemployment, and informal settlements, which are factors that have been associated with higher crime rates. Kasarani was chosen as a study site due to its high crime rates and availability of data sources. The health center provided cover information of what we and received from the police report. This study provided information on the relationship between crime and health outcomes in a crime prone area.

Target population

The target population was Kasarani Sub County in Nairobi County. According to (Zhao, et al.), the group of people or things that the research or study is meant to reflect or draw conclusions about is known as the target population. The target population forms the group of interest that the researcher wants to generalize their findings to. This is the group that the researcher wants to draw conclusions on the basis of the data collected.

Sample size calculation

The sample was drawn using Cochran's formula due to the large population.

That is;

$$n = z^2pq/e^2$$

Where;

$$n = desired\ sample\ size$$

Z = the z-value of 1.96 which corresponds to confidence level

P = estimated proportion of the population

$$Q = 1 - p(0.5)$$

e= desired margin error

Therefore;

$$n = (1.96^2 * 0.5 * 0.5)/(0.05)^2 = 385$$

Sampling Techniques

For facilitation of the required information from the respondents and informants, we deployed two sampling techniques i.e. simple random sampling and purposive sampling. We administered questionnaires randomly among the residents of Kasarani Sub-county. This allowed equal chances of selection. The key informants were the Kenya Police Service and

Ministry of Health, whereby we accessed Kasarani Police station and Kasarani Health Centre. The choice of purposive sampling was thus premised on the fact that both are reliable sources of information on crimes.

Data collection Instruments

The study used secondary data. Secondary data collected from Kasarani police station and Kasarani Health Center. Data collected was from the year 2018 – 2022. At Kasarani Police Station, the researchers got a comprehensive data month by month on crimes such as murder, manslaughter etc. These were the preferred data collection instruments since our study aims at finding a pattern and using statistical model to predict future crime. The study also used some data from the Kenya National Bureau of Statistics (KNBS), news reports and other relevant articles.

Validity and reliability of the Data Collection Instruments

Validity

Validity in research methodology has been described as the extent to which the research accurately measures or represents the objectives under study (Kothari, 2013). It is a critical part of research quality and it ensures the research findings obtained from a study are valid, thus conveying the soundness of that tool. The researcher utilized the knowledge and guidance of the supervisor and made sure that the data obtained from the police station was beneficial in accomplishing the study objectives. Also external validity was used to check whether the findings of the research study can be generalized to a larger population setting, this was accomplished by using a representative sample that closely resembles the target population.

Reliability of Research Instruments

Creswell (2014) defines reliability as the consistency, stability and dependability of the data collected in a study. It refers to the extent to which a research instrument or procedure yields consistent results from repeated trials under similar conditions, hence ensuring that the measurements are free from random error that could affect the study validity.

Data collection Procedure

The researchers obtained an introduction letter from the University of Nairobi-department of Mathematics and addressed to the Ministry of Interior and National Administration-Ethics Board, Nyayo House. Here the researchers were requesting for authorization to collect data from Kasarani Health Centre. After obtaining the authorization letter, the researchers went to the doctor in charge of the facility-Kasarani Health Centre. He requested to go through the research project to get a clear understanding of our work before agreeing to give the data requested. He also agreed to be interviewed while on record.

The researchers also obtained an introduction letter from the department of Mathematics addressed to the deputy inspector general, Vigilance House. The researchers were seeking authorization to obtain data from Kasarani police station. We then proceeded to the police station ICT office but were redirected to the OCPD office to grant permission to access data we needed. The OCPD was expecting pour visit as he had been informed prior from Vigilance House. Permit was granted and we obtained the data we needed from the station's records office. We then proceeded to the Deputy OCPD who agreed to conduct an interview.

Time Series Analysis

Cryer Jonathan, 1986, describes time series analysis as a statistical technique used to analyze and predict patterns in a data sequence collected over time. Examining historical patterns in a past data enables the modelling and making informed forecasts to track future behavior. Our aim in utilizing the time series model on Nairobi's historical data is to make it easy on understanding the underlying structure and characteristics of historical crimes in the capital. Time series involves identifying trends, seasonality, cyclic patterns and irregular fluctuations (noise). We have had multinational organizations utilizing the power of time series in segments such as finance at forex market levels, weather forecasting and in our case, crime prediction.

A common approach in time series analysis is the Autoregressive Integrated Moving Average (ARIMA) model. The model combines Autoregressive (AR), differencing (I) and the Moving Average (MA) components to capture sequential dependencies and stationarity properties in a sample set (Gilbert 2005). The ARIMA model is specified as ARIMA (p, d, q), where p, d, and q are the orders of the autoregressive, differencing, and moving average components, respectively. The relationship between our observed values and their lagged values is accounted in the AR model. The AR model assumes the current value of the variable is influenced by it historical values. The model is denoted as p which is a representation of the number of lagged observations utilized in the model. On the other hand, the MA component models the error term as a combination of past values. This model accounts for random fluctuations in the time series that aren't explained by the autoregressive and differencing components. The MA model is denoted by 'q' which represents the number of lagged errors used in the model. The I component deals with the differencing of the data to transform a non-stationary time series into a stationary one. Stationarity is an assumption under time series analysis as it ensures that the statistical properties remain constant over time (Priestley and Subba 1969). The differencing component involves taking the difference between consecutive observations to remove the trend component and seasonality. This component is denoted by'd' which is a representation of the number of times differencing is applied.

The ARIMA equation represents the mathematical formula used to describe the ARIMA model.

$$\boldsymbol{Y_t = c + \phi_1 Y_t + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t}$$

Where: $\boldsymbol{Y_t}$ represents the value of the time series at time t

$\boldsymbol{C}$ is the intercept

$\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \dots \boldsymbol{\phi_p}$ Are the autoregressive (AR) parameters that represent the weights of the lagged values of the time series.

$\boldsymbol{Y_t}, \boldsymbol{Y_{t-2}}, \dots \boldsymbol{Y_{t-p}}$; lagged values of the time series.

$\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \dots, \boldsymbol{\theta_q}$; are the moving average (MA) parameters that represent the weights of the lagged error terms.

$\boldsymbol{\varepsilon_{t-1}}, \boldsymbol{\epsilon_{t-2}}, \boldsymbol{\epsilon_{t-q}}$**;** are the lagged error terms.

$\boldsymbol{\epsilon_t}$; The error term at time t, normally distributed with mean zero and constant variance.

Data Analysis and Presentation

After data collection, the researchers drafted a data profiling report which contained the basis statistics, raw counts, data structure, missing data profile, the Univariate distribution, the Correlation analysis and the Principal component analysis. The quantitative data was analyzed using the statistical package R to obtain the patterns and trends of crime in Kasarani. Line graphs, bar charts, line plots, and histograms were used to present findings of the quantitative data. The researchers used time series analysis to present the qualitative data. The researchers performed a hypothesis testing on the impact of the COVID pandemic on crime rates.

Ethical Consideration

Ethical rules and regulation were a foremost consideration in this project. Great effort was made to ensure that all the rules and standards associated with data collection were fulfilled. The researchers got an introductory letter from their department at The University of Nairobi then headed to both Nyayo house and Vigilance house. At Nyayo house the researchers were granted approval by the county commissioner in the office of the Ministry of Interior and National Administration to collect data from Kasarani Health Centre. At Vigilance house, the reseachers' request to collect data from Kasarani Police Station was approved by the Deputy Inspector General of the Kenya Police Service. Data collected was treated with outmost integrity, there was no fabrication or duplication done on the data.

CHAPTER 4: PRESENTATION OF FINDINGS AND DISCUSSION

Demographic Characteristics

Gender.

Table 4.2 shows that females are the majority in this sub-county by 51%. The margin is not large as males have a 49% representation. The minority in the population are the intersex with a population of 37 individuals from a total population of 780656 individuals.

Gender	Population
Female	399385
Male	381234
Intersex	37
Total	**780656**

Table2: Gender of the residents, *source:* Kenya National Bureau of Statistics (2019)

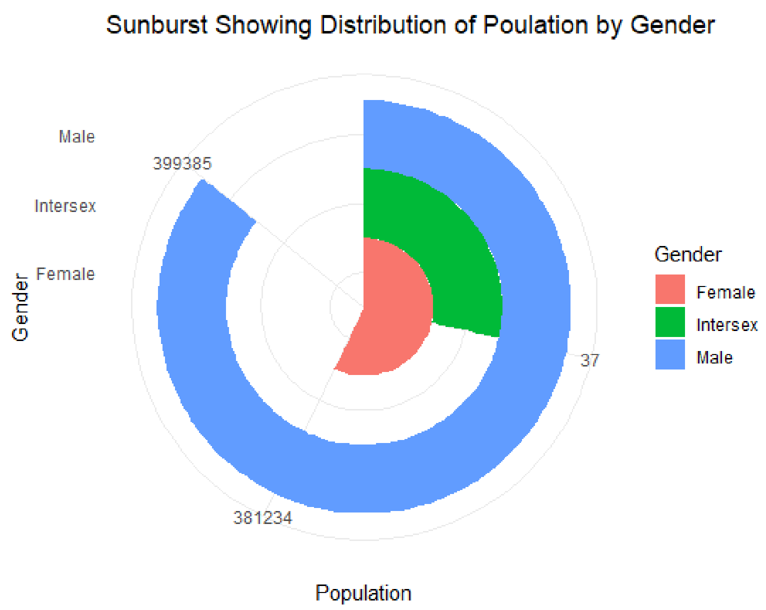

Figure 1: Gender

Age.

0-9 years	10-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60-69 years	70-79 years	80-89 years	90+ years
175303	128466	212331	152127	69015	29757	9852	2785	728	229

Table3: Residents age distribution

Table 3 shows that 23% of the population are of the age between 0 to 9 years, between the ages of 10 to 19 years had a 17% representation of the population with 128466 persons. Further, 212331 (27%) were between 20 to 29 years, 152127 (20%) were between 30 to 39 years, 69015 (9%) were between 40 to 49 years, 29757 (4%) were between 50 to 59 years, 9852 (1.26%) were between the age 60 to 69 years, 2785 (0.36%) were between the age 70 to 79 years, 728 (0.1%) were between the age 80 to 89 years while 229 (0.03%) were the age 90 years and above. The information is critical as to understand the distribution of this population. Figure 4.2 shows that the age group 15 to 64 years is the majority in this population. This is the age group with individuals at their prime and productive age.

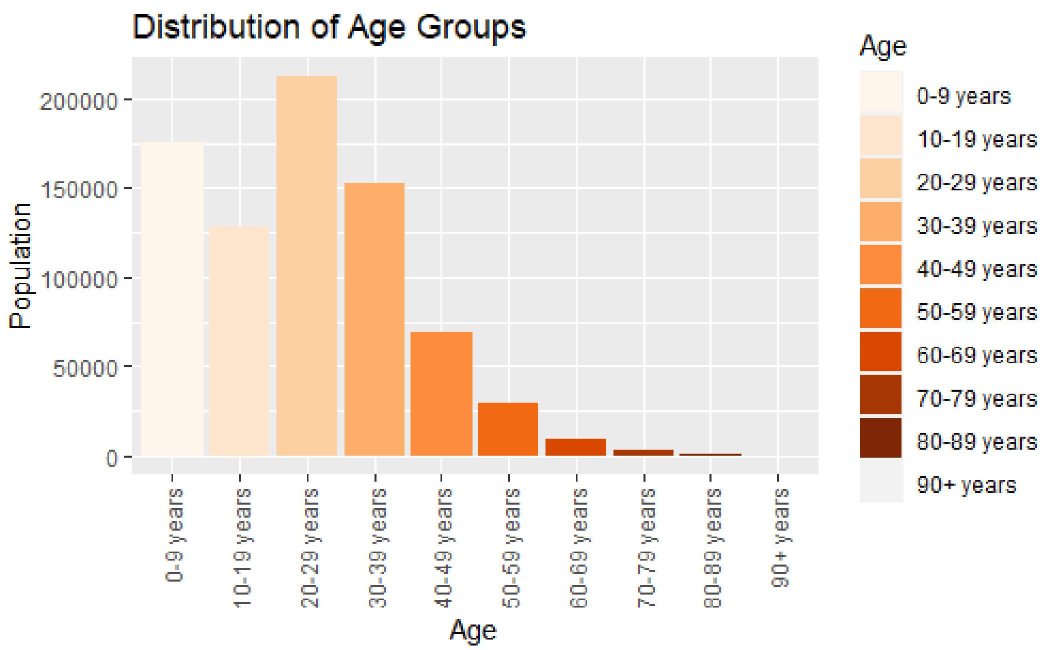

Figure 2: Distribution of Age

Descriptive findings.

Patterns and trends of crime in Kasarani sub-county Between 2018-2022.

OFFENCE	2018	2019	2020	2021	2022	TOTAL
GENDER-BASED VIOLENCE	99	123	113	121	202	658
THEFT AND STEALING	128	163	109	161	223	784
MURDER (ATTEMPTS)	44	14	11	22	31	122
DRUG AND SUBSTANCE ABUSE	23	33	14	15	27	112
BREAKINGS/BURGLARY	49	40	53	51	76	269
ALL OTHER PENAL CODE OFFENCES	37	120	91	146	139	533
TOTAL	**380**	**493**	**393**	**523**	**698**	**2487**

Table 4: Reported crimes in Kasarani sub county

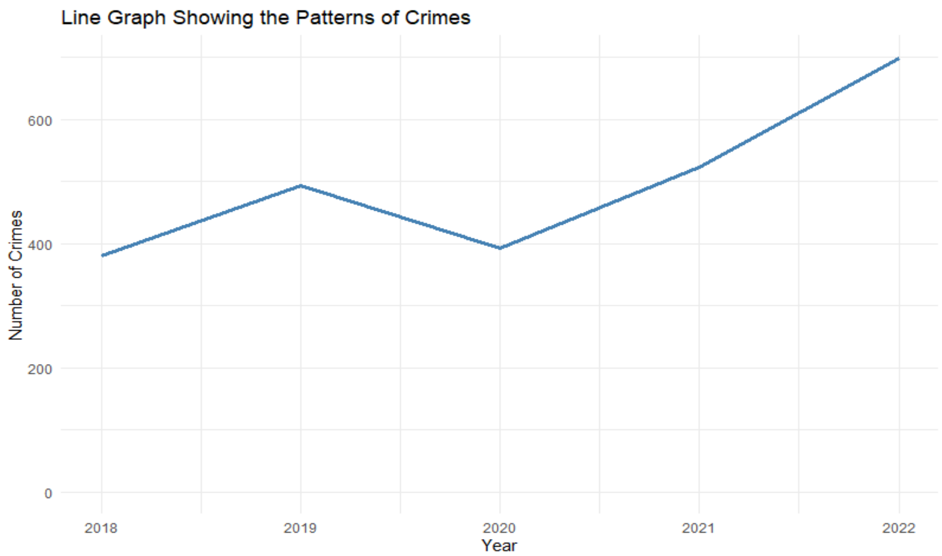

Figure 3: Crime pattern

The line graph above indicates an upward trend in the total number of crimes committed from 2020-2022. The decreased rate in 2020 can be explained by restrictions reinforced by the government during the COVID 19 pandemic. The sharp upward trend can be explained by inflation and unemployment factors which rose due to the graph and trend.

Offences	Mean	Std. Dev	Min	Max	Crime rate/100,000 People.
GENDER-BASED VIOLENCE	132	40.47	99	202	84.29
THEFT AND STEALING	157	43.45	109	223	100.43
MURDER (ATTEMPTS)	24	13.43	11	44	15.63
DRUG AND SUBSTANCE ABUSE	22	8.05	14	33	14.35
BREAKINGS/BURGLARY	59	13.36	40	76	34.46
ALL OTHER PENAL CODE OFFENCES	107	44.35	37	146	68.28

Table 5: Descriptive Statistics on Forms of crime

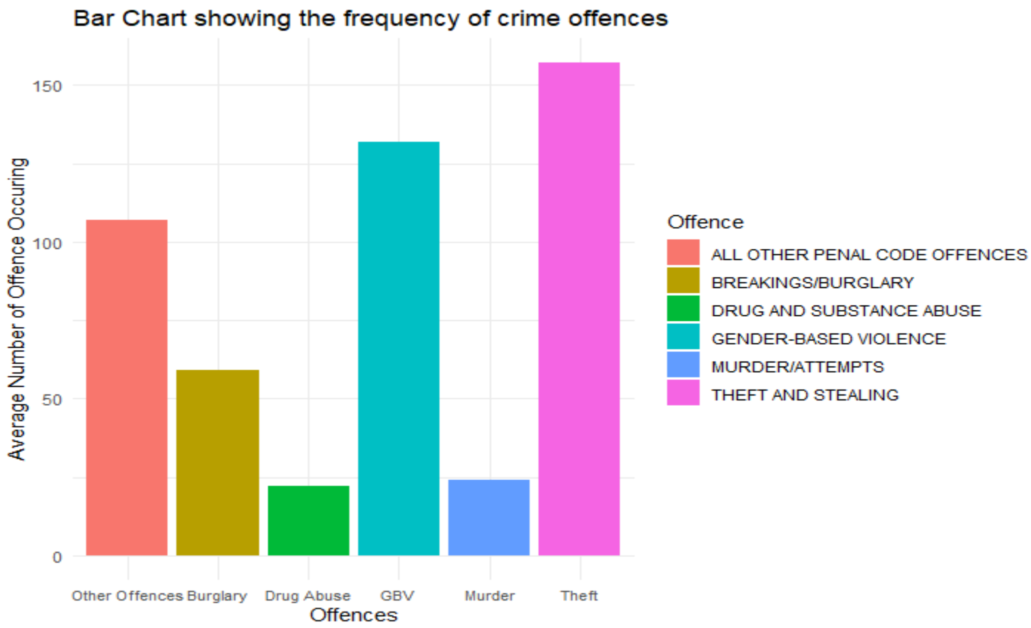

Figure 4: Frequency of crime offences.

The most prevalent form of crime, as indicated in the chart above, is Theft and stealing, and gender-based violence. All other penal-code offences were petty crimes that happen in the streets regularly. Theft and stealing has an average of 157 cases per year ranging 109-223 cases. It is followed by Gender-Based Violence with an average of 132 cases and ranges between 99-202 cases in the year between 2018-2022. The crime rate, calculated per 100,000 people, for theft and stealing and GBV is approximately 100 and 83, which is relatively high during a span of 5 years.

Factors contributing to crime in Nairobi.

Factors contributing to crime	**Gender**	
	Male	Female
Conflict over natural resources	62	37
Retrogressive cultural, religious beliefs and practices	67	39
Alcohol, drug and substance abuse	326	249

Proliferation of illicit arms and weapons	8	12
Greed for quick wealth	70	52
High cost of living	51	31
High rate of school drop out	20	21
Trafficking of illicit alcohol and drugs	39	24
Illiteracy and ignorance	229	162
Weak law enforcement	87	52
Marginalization and unequal distribution of resources	28	20
Peer pressure/influence	132	88
Political incitement	32	14
Poor parenting/upbringing and moral decadence	54	65
Poverty	493	392
Crime of passion factors (e.g., infidelity)	43	23
Stress and depression	6	17
Technological exposure	3	0
Negative ethnicity and nepotism	47	28
Unemployment	795	572
Hereditary factors	6	8
Revenge after crime factors	13	15

Radicalization	13	5
Corruption	27	38
Family/domestic disputes	20	20
Migration	4	4
Poor transport and communication infrastructure	10	12
Total	**1176**	**872**

Table 6: Factors contributing to crime.

Source: Crime mapping (2016)

According to data from crime mapping (2016), unemployment topped the list among the factors contributing to crime with responses from 795 men and 572 women. Poverty is close with 493 men and 392 women citing. Alcohol, drug and substance abuse with 326 men and 249 women responding. The other top factors are illiteracy and ignorance (229 men, 162 women) and peer pressure/influence (132 men and 88 women).

As at 2016, Kenya had a population of 47.89 million with the youth amounting to approximately 75% of the population. This is data according to the World Bank. Most of the youths are unemployed and majority come from low income areas (slums) where poverty is rampant. This makes them feel marginalized in terms of access to opportunities, representation and participation. Thus most resort to crime as the means to get out of poverty. Others due to idleness and peer influence abuse drugs which lead them to commit crime as they have already been intoxicated.

Time Series Analysis

Line Plot

The line plot it plotted for total criminal offences against years to have a general view of how the variable crime progresses over the years. The data is collected from 2018 to 2022 at a monthly frequency from January to December. No preprocessing has been done to the original data if not having the null value being filled with zero, which is still an implication there was no crime activity taking place.

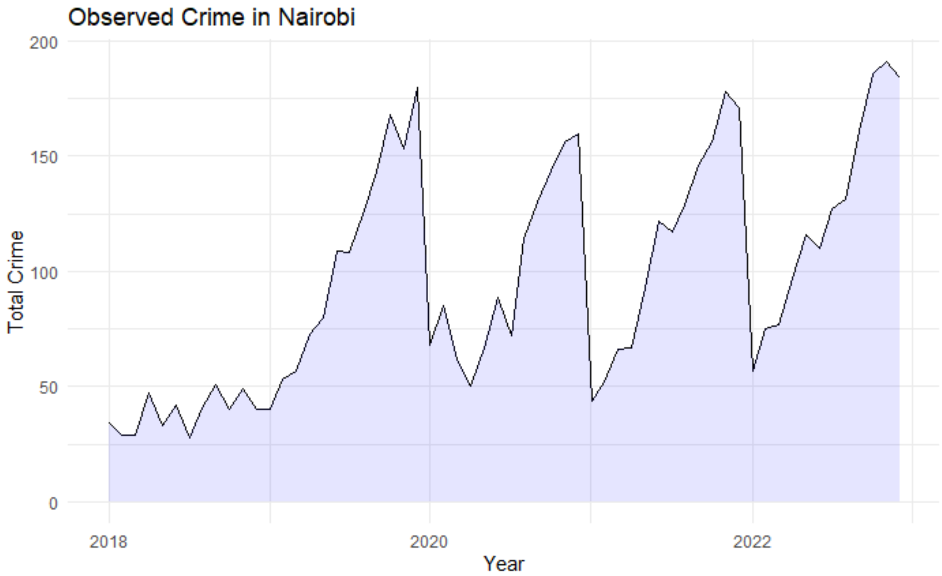

Figure 5: Observed crime in Nairobi.

The line plot starts gradually from 2018 and peaks towards 2019 and trough through onset of 2020, the pattern goes on towards the end of a year and peaks again on the start of the New Year. These irregular patterns at around 2018-2019 can be described as outliers, they are the points that have deviated from our general time series pattern and may alter the general performance and evaluation of our model. There is presence of an inconsistent trend from start of 2021 towards the end of 2022. (Bovenga, Guido and Alberto 2021) Define trend as the long-term behavior or direction of a time series. Guido particularly describes trend as an indicator for if a variable is increasing, decreasing or remaining stable over time. From the line plot we don't get a clear depiction of seasonality as there is no repetitive patterns occurring at fixed stages in the time series. The time series is peaking at around July-September and forming troughs on December to May. The former presence of outliers dully affects the seasonality component. The time series therefore follows an additive model. The log transformed of the time series can help us pick on what model is appropriate for the time series.

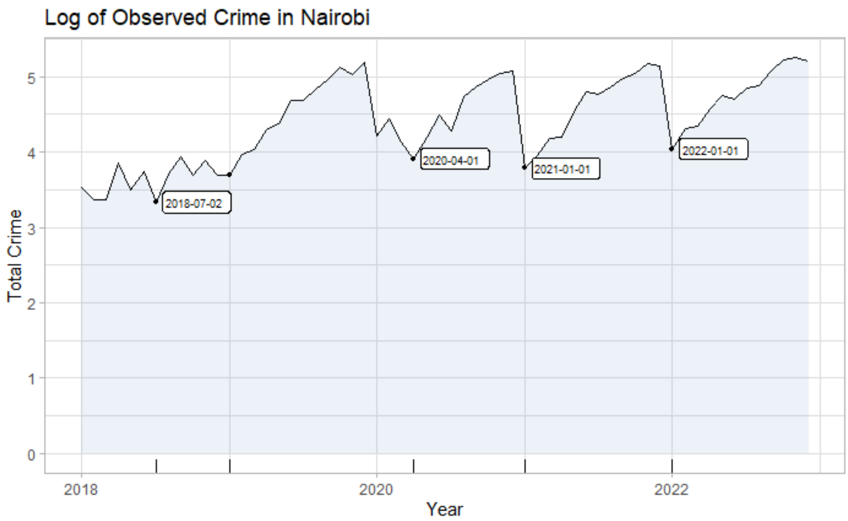

Figure 6: Log of observed crime in Nairobi.

From the log-transformed time series, the seasonal fluctuations and the random fluctuations seem to be constant over time. We can therefore describe the time series using an additive model instead of using the multiplicative model.

Time Series Decomposition.

This refers to splitting the time series into its underlying components; trend, seasonality and random fluctuations. An additive model follows the sum of it individual components such that.

$$\boldsymbol{Time\ Series\ =\ Trend\ +\ Seasonality\ +\ Residual}$$

This process involves us estimating the trend and the seasonality component and then calculating the residual component. The appropriate procedure for the decomposition of the additive model will be the simple moving average[1]. This method will help in the smoothing process of the time series.

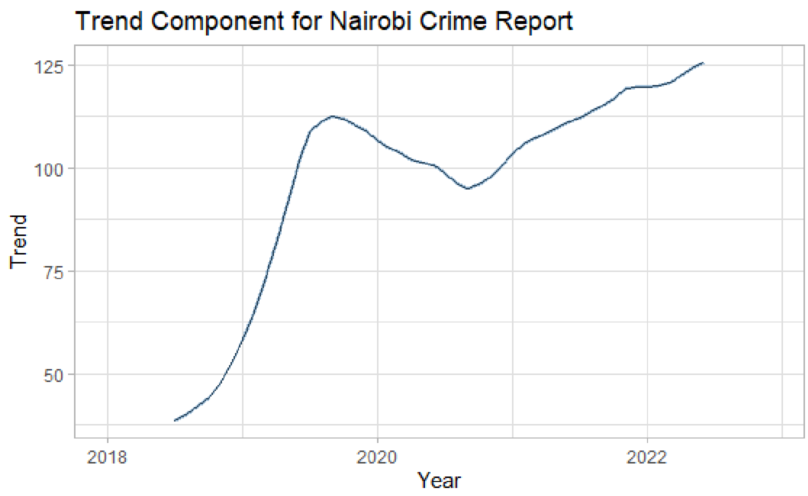

Figure 7: Trend Component of Nairobi crime Report.

A smoothing using simple average method of order 5 is applied on the time series and provides us with the trend component and its progression. One can simply notice how the crime rate peaked at the end of 2022 unlike any other year recorded in the data set. Removing the trend component leaves us with the following plot.

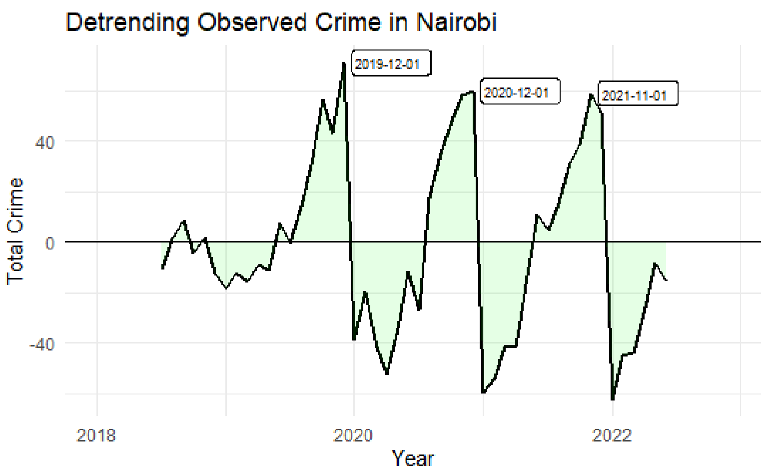

Figure 8: Detrending Observed Crime in Nairobi.

[1] SMA is a technique in time series analysis to smooth out fluctuations and identify underlying trends in the data. The procedure involves calculating the average of a specified number of data points over a moving window to provide smoothed time series representation

Decomposing the whole time series gets us the following using the decompose function in ggplot gets us the single attributes of the time series separated.

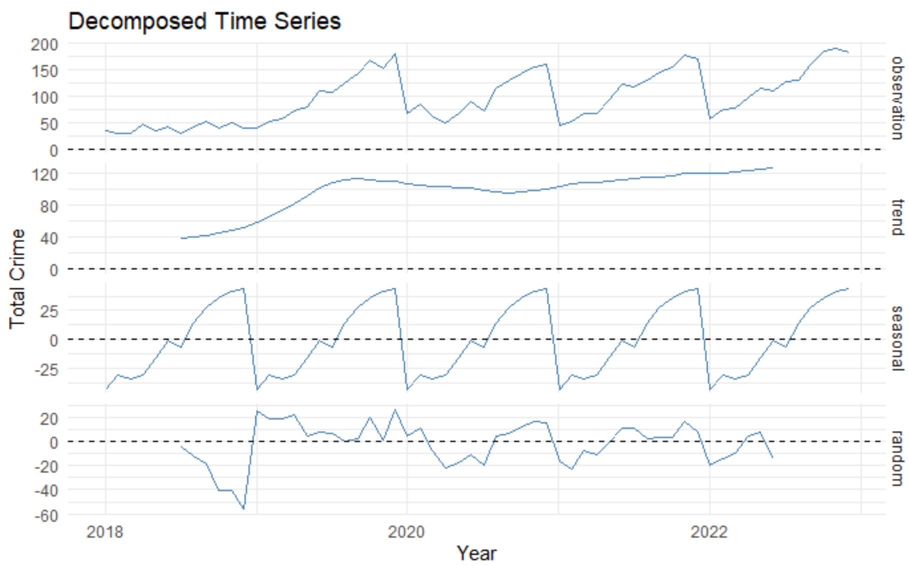

Figure 9: Decomposed Time Series.

We can deduce that all assumptions for ARIMA modelling are met according to the stationarity and seasonality. We remove the non-stationary part for the sake of creating the ARIMA model.

Augmented Dickey-Fuller Test for Stationarity

ADF test is a statistical test that determines whether a time series has a unit root or is stationary (Yaya, et al. 2021). This is a commonly used unit root test for a time series which addresses autocorrelation and lagged differences issue in the data. This method uses the hypothesis testing methods to compare the critical values and determine significance of the test. For example;

$$H_0: The\ time\ series\ contains\ a\ unit\ root, indicating\ non\ stationarity$$

$$H_1:\ The\ time\ series\ has\ no\ unit\ root, the\ time\ series\ is\ stationary.$$

Calling for the test function, we get the following tabulated results

Augmented Dickey-Fuller Test		
Dickey-Fuller	Lag order	p-value
-3.8447	3	0.002249
Alternative Hypothesis	Stationary	

Table 7: Augmented Dickey-Fuller Test.

Based on this results, The ADF test statistic is -3.8447 performed on lag order 3. The p-value is 0.02 which is less than 0.05. We therefore reject the null hypothesis that states that the time series contains a unit root at 0.05 significance level. We then accept the null hypothesis that the time series is stationary. The ADF test therefore implies that the crime report time series does not possess a unit root so we consider the time series for ARIMA modelling.

Parameter Estimation

After testing for stationarity, we analyze the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. These determines the order of the AR and MA components respectively. From the cutoff points of this plots, we acquire the p value from the ACF and the q value from the PACF plot. This values will be our initial estimates for the model as we test for the optimal order that balances model fit and complexity. The general process is iterative and revolves around refining order estimation. The plot below is an ACF plot and cuts off after lag 3.

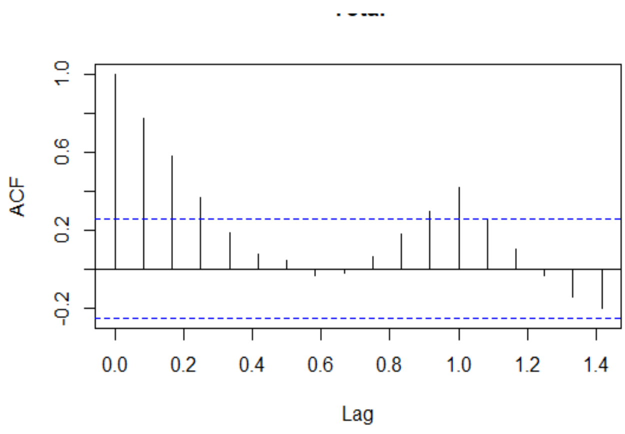

Figure 10: ACF plot

The initial lags up to lag 3 display higher correlation values indicating a significant correlation between the current observation and the previous 4 observations. After the 3rd lag the correlation between the observations of greater lags diminishes. Sporadic correlations reappear at lag 12 and 13 but they would not affect the determination of p since we use the lags before the first cutoff. We can't rely on the ACF solely to get informed and near-accurate values of p. For effectiveness, utilizing model diagnostics procedures such as Bayesian Information Criterion (BIC) model selection criterion will ensure that our estimation of p will lead to an optimal measure.

The PACF plot also helps us in the determination of the p value for the ARIMA/ARMA model. Unlike the ACF, the PACF is focused on the direct correlation between observations at different lags removing indirect effects (Yakubu and Moch 2022).

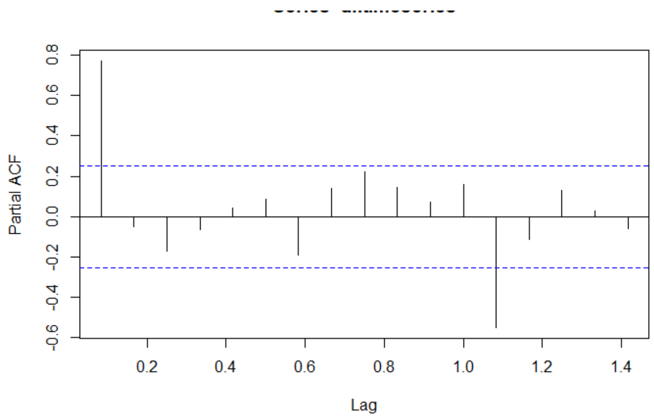

Figure 11: PACF plot.

From the plot, there are no significant partial autocorrelations beyond lag 1. The plot depicts most of the points falling below the CI level. At lag 1, the partial autocorrelation means that there is a strong direct relationship between our current observation and what is immediately preceding our observation (Mestre, et al. 2021). It's easy to note that all other partial autocorrelations fall within the CI. We deduce that the smaller partial autocolerrations are due to random variables in the data leading to no significant direct relationship between current observation and observations at those lags. However, there is one exception at lag 12, which shows a significant partial autocorrelation.

Using forecasts, we generate the p, d, q values as follow: $ARIMA\ (1,1,0)(0,1,0)[12]$. The model includes an autoregressive component of order 1, a differencing component of order 1 and a moving average component of order 0. The seasonal part of the model includes a seasonal differencing component of order 1 and a seasonal moving average component of order 0 with a 12-seasonal period. This model shows a lesser standard deviation from most of values used by other models. Also, the auto ARIMA model showed lesser for the p and q values drawn from the plot

$\boldsymbol{Model\ (p, d, q)}$	$\mathbf{0, 0, 3}$	$\mathbf{1, 0, 0}$	$\mathbf{1, 1, 3}$	$\mathbf{0, 0, 0}$	$\mathbf{(1, 1, 0)(0, 1, 0)[12]}$
$\boldsymbol{\sigma^2}$	854.2	846.6	829.9	2413	547.7
AIC	586.53	581.88	574.5	641.58	432.82

Table 8: The auto ARIMA model.

Variance and AIC aid in model selection, the lesser these two values are suggests that the models predictions are closer to the actual values. Lesser variance means fewer spread of residuals around the mean. On the other hand, the AIC determines the relative quality of the statistical model. The lower the value the better the fit which accounts for the model's complexity and goodness of fit (Cavanaugh and Andrew 2019).

The model selected is the last model with seasonal element *[1, 0, and 0] [0, 1, and 0]*. The table below represents the model summary.

Coefficients	AR(1)
	-0.2697
S.E.	0.1395

Table 9: The model summary

The coefficient AR (1) corresponds to the autoregressive term. The AR term is the coefficient of the lagged value of the time series variable. The AR (1) coefficient in this case is -.2697 with .1395 standard error. The coefficient being negative means there is a negative relationship with the lagged observation (Zemoul and Youcef 2022). As the lagged value decrease, our current value tends to increase and vice versa. Our standard error is slightly lower indicating a more precise p estimate. Our model selected belongs to SARIMA models. Its stands for Seasonal Autoregressive Integrated Moving Average and is simply an extension of ARIMA. This is a suitable selection since our data patterns exhibits both seasonal and non-seasonal properties along 2018-2019.

The equation for the model is as given below after estimation of our parameters:

$$\boldsymbol{Y_t = (1 - 0.2697 * B)(1 - B^{12}) * X_t}$$

- Y_t Represents the value of the time series at time t
- B is the backshift operator that represents the lag operator
- X_t Is the differenced series at time t

The non-seasonal autoregressive (AR) component is AR (1), indicated by the coefficient of -0.2697. The other non-seasonal differencing (I) component is of order 1, indicated by the (1, 1, 0) part. The seasonal differencing (D) component is of order 1, indicated by the (0, 1, 0) [12] part while the seasonal period is 12 indicated by [12] from Auto ARIMA.

Model Diagnostics

This is an essential step in evaluation of our time series model. This process identifies model adequacy and reliability. Some common procedures include residual analysis which is the difference between the observed values and the predicted values (Khan and Rajiv 2021). A residual plot will help us visualize the patterns and the deviations from the characteristics of the residuals such as; zero mean, independence and constant variance. Other measures in diagnostics include; ACF and PACF for residuals, Ljung-Box Test and normality test. First we look at how the model's predicted crime offences match with the observed crime offences from forecast model using our AR(1) component

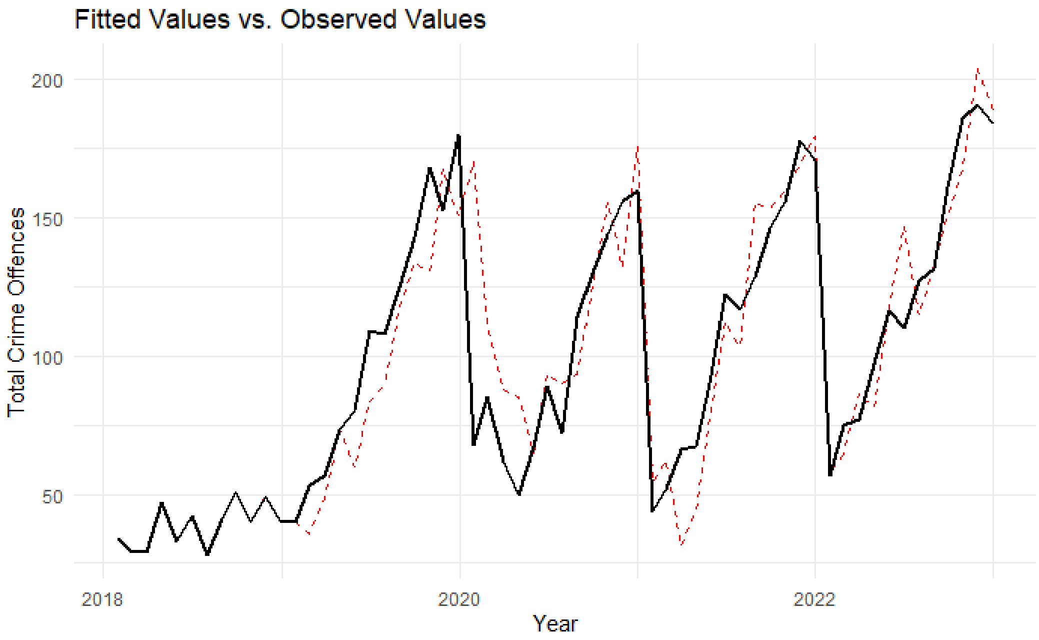

Figure 12: Fitted values vs Observed values

The plot shows the observed crime numbers plotted with the predicted crime numbers from our model. It's safe to say the predicted values sit tight on our observed values with less variance. The purpose of examining this fit is to see ow well our model captures the underlying patterns and variability in the data. This is how the descriptive statistics between the two groups sit.

	Min	**Median**	**Mean**	**StdDev**	**Max**
Observed	28.00	87.00	96.70	49.53	191.00
Fitted	28.00	88.76	96.57	50.55	203.95

Table 10: Descriptive statistics.

From this summary statistics, we notice that the minimum value of both groups is 28. This is an indication that our model captures the lower range of the data accurately. The median value is quite close between our groups cementing further that the model captures the central tendency parameters effectively. The same can be said about the mean. Considering the standard deviation, the observed group has a standard deviation of 49.53 while the fitted group has a standard deviation of 50.55. This shows some variability introduced by the model in its prediction. The maximum values is where we have greater differences between the two groups, we might say this is due to overestimation of the extreme values by the model.

Box-Ljung test

This test is used to assess the overall autocorrelation in the residuals. A low p-value suggests that there is significant autocorrelation remaining in the residuals suggesting a lack of fit in the model.

Box-Ljung test results:	
Test statistic	0.1464485
p-value	0.7019527

Table 11: Box-Ljung test results.

The null hypothesis tests for no autocorrelation while the alternative hypothesis checks for presence of correlation (Hassani and Mohhammad 2020). From the results, the test statistic is reported as 0.1464485, and the associated p-value is 0.7019527. The test results provide a relatively high p-value of suggesting that there is no sufficient evidence to reject the null hypothesis of no autocorrelation. Our conclusion is led to that the model displays significant correlation at different lags and there is no autocorrelation present in the residuals model. The model therefore captures the temporal dependencies in the data and residuals are relatively independent, ultimately passing the independent test of residuals

Normal Probability Plots for Residuals

They are also referred to Quantile-Quantile (QQ) Plots, They assess the distributional similarity between the observed values and the fitted values of the model (Yusn, Hongyun and Yuting 2021). Data points are plotted along a diagonal line to check for good fit between the observed and expected distribution. Having deviation from this line suggest differences in the distributional characteristics. From our plot, we have all but one point lining along the diagonal line. This one point is affected by extreme outliers in the observed data set from the year 2018 where criminal activities were low. This outlier in the QQ plot also suggest possibility of skewness in the distribution of the residuals as we are going to test for normality in the next sub section

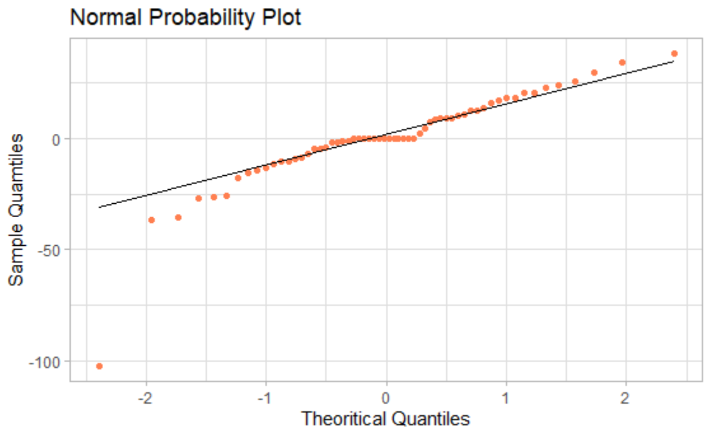

Figure 13: Normal Probability Plot

Distribution of Residuals

Histograms provides us with the visual check of how the distribution of the data is by dividing it into bins along the axis. To check for normality, we expect a bell shaped frequency polygon in the histogram. From our QQ plots, we see a deviation along and the diagonal and we expect to have it replicated along the histogram.

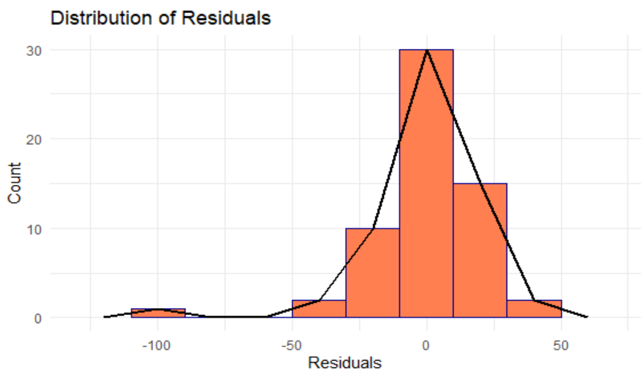

Figure 14: Distribution of Residuals

The histogram displays a distribution heavily skewed to the left and no sign of a normal distribution. Our study involves outliers in the data form criminal activity in 2018 and this is quite rightly brought into the models and its residuals.

Forecasting

The forecasting process includes fitting the model to the historical data, and then using the model to generate forecasts for future periods (Torres, et al. 2021). This step will involve accuracy assessment as we will see how our model performs along with the test set. By forecasting, we can modify the time period ahead to see the number of crimes in the months to come. Our model constitutes of autoregressive (AR) component of order 1, a non-seasonal differencing (I) component of order 1, and a seasonal differencing (D) component of order 1 with a seasonal period of 12. These component enables the model to capture the temporary correlation within the crime information making it possible to capture the trends and patterns in the data.

The coefficient estimate for the AR components is -0.2697, this values suggests that the previous observed value has a negative influence on the current value at time t.

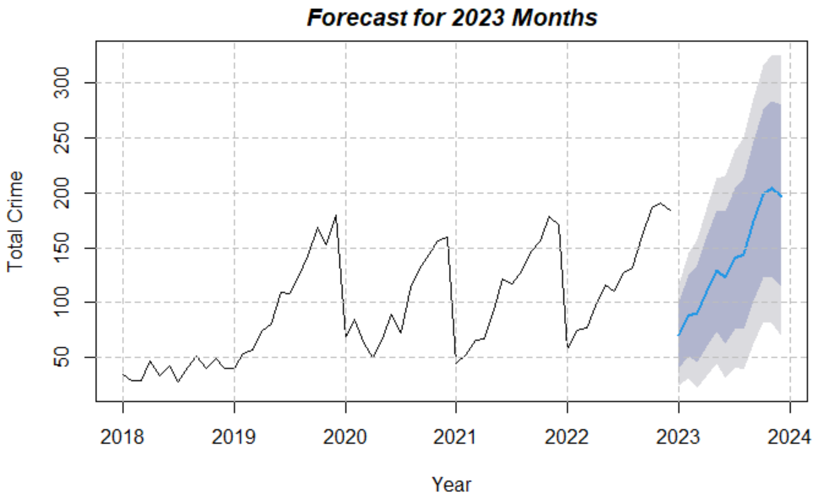

Figure 15: Forecast for 2023 months.

The plot shows the forecast values for 12 months of 2023. The shaded regions are the 95% and 80% confidence intervals for the mean of the forecast months. The plot looks to be following the same trend as observed from 2019 onwards.

Summary statistics Comparison for the forecast values of 2023 Vs 2022 are as follows

	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Forecast_2023	70.0	105.0	134.5	139.0	179.8	204.0
Observed_2022	57.0	92.0	121.0	126.0	166.8	191.0

Table 12: Summary statistics.

This gives us a comparison of how the forecasted values for 2023 perform against the previous year 2022. The 1st quartile shows the approximate values for March 2023 while the 3rd quartile shows the expected values for September 2023. The mean monthly number of reported is expected to rise to 139 from the 2022 reported mean of 126 cases. The year 2023 also expects a record high of 204 crimes to be reported from the previous high value in 191.

Forecasting Rape cases for 2023

We can't use the total crimes model to single out each type of offence. Instead we define new coefficients for the new variable that captures the variability and patterns within the sample. The values in the data set for the rape category are minimal and almost no recording happen in certain months.

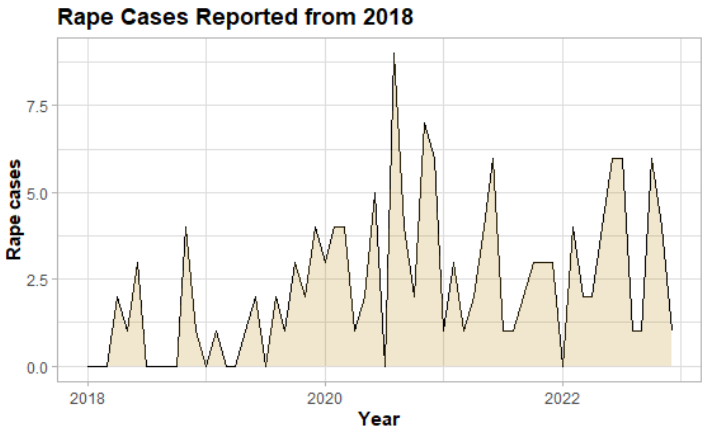

Figure 16: Rape cases reported from 2018

This plot shows the line plot for rape cases. Highest cases were reported in 2020 while the least cases were reported in 2018. It’s also easy to note that 2020 had most criminal activities reported if we keeping in mind the general report collected from the area. One could devise a test to find comparisons between the numbers of crime before March 2020 (onset of COVID) and after that using a two-samples-Independent test. More information about this discovery can be established from the trend.

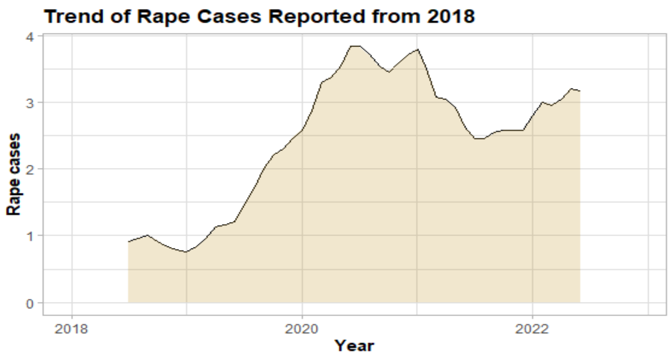

Figure 17: Trend of Rape cases reported from 2018

From this we can discover the trend of how the rape cases have been progressing from 2018. There is a concerning peaking from 2018 to 2020 and then dips from 2021(probably after end of lockdown). The variable display characteristics of a time series and would favor us working on the prediction and selection of the model.

Using auto ARIMA, we can find the best components for our model and carry out our forecast. It would be better if the model could only predict for he numbers up to June to increase the accuracy of the prediction and model validity.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**
2023	0	3	1	1	3	5

Table 13: Predicted rape cases in 2023

This are the numbers predicted for rape cases in 2023. The actual number of rape cases in 2023 may vary due to various factors such as social, economic, and demographic changes, as well as the effectiveness of crime prevention measures and law enforcement efforts. To keep the modelling more effective, regular monitoring and updating of the data must be implemented to train the model effectively.

Research Findings

Two-Sample Independent Test on Crime Trends before and after onset of COVID-19

It's quite obvious on how the pattern of crime in Nairobi has been before COVID and after COVID declaration in Kenya. The novel virus was declared a pandemic in Kenya in March 2020 and a national lockdown was reinstated. The country was adversely affected including economic wise and social interactions limited. Curfews from 7 pm to 4 am were put in place to minimize gatherings and interactions to prevent the spread of the virus. During this period, Nairobi was on total lockdown and no one was allowed to get in or out of the city. As reported by (Mathew & Dan, 2021), before the onset of the pandemic, an estimated 85% of residents in Nairobi in informal areas were food insecure. This is largely based on the fact that there was a high poverty index and high unemployment rate. This factors are large contributors to crime and many youths opted to crime activities to achieve ends meet. From our line plots before, criminal activities before the pandemic were low figures and slowly rose as the global scare of COVID hit the international media. This section reports on paired sample t test done to find the effect of the pandemic on crime numbers in Nairobi.

A Two-sample t-test is a statistical test that is used to compare means of two related or paired groups (Keselman, et al., n.d.). The test is used in a test involving same subjects observed under two different conditions or at two different time points. Given our study, we have cases of crime reported *before* COVID-19 onset and *after* declaration of COVID-19. Our sample is going to work under assumption that the observations are independent of each other (Derrick, et al., 2018). Also, we assume that the differences between the paired observations are approximately normally distributed (Rochon, et al., 2012). Lastly, homogeneity of variances is also assumed in the sample, (Moser & Garry, 1992) state that the variance of the differences between the paired observations are equal. To perform the test, we first design our null and alternative hypothesis. The null hypothesis will form comparison with the p-value to determine if we reject or keep the null hypothesis. Here are the hypothesis statements;

- H_0: There is no significant difference in the mean rate of crime after COVID-19 onset
- H_1: There is a significant difference in the mean rate of crime after onset of COVID-19.

We are working with an underlying assumption that there is a significant difference in rate of crime during COVID and after the COVID pandemic. This is the summary table after mutation and transformation. The variable `COVID.Period` refers to the time before and after.

COVID Period	Count	Mean	SD
After	26	103.5	42.01167
Before	26	73.26923	46.42289

Table 14; Crime rate before and after the onset of COVID-19

The table provides with summary statistics for the two groups. From the table we can get a better understanding of the variability and distribution of the crime rate before and after the onset of COVID-19. The "After" group consists of 26 observations, with a mean crime rate of 103.5 and a standard deviation of 42.01. This means that on average, the crime rate after peak of COVID was higher and there was moderate variability in the recorded crime. On the other hand, the “Before” group consisted of 26 observations with mean crime rate of 73.27, the standard deviation was 46.42. This group had a lower mean crime rate. The before group also showed a higher amount of variability in the crime rates recorded. For visualization, a boxplot will help us understand the distribution of the groups.

Figure 18: Box plot for before and after onset of COVID-19

We get the impression that the median crime rate after COVID-19 hit was much higher than the median crime rate before the start of the pandemic. The boxplot helps assess the notable differences between these two groups.

Next step is checking for normality of the two groups. A density plot will help us assess the normality of the two distributions. Normality is usually displayed by bell shaped curves or symmetric shapes.

Figure 19: Density plots for before and after.

The after group displays a symmetric distribution whereas the before group is skewed towards the right. We can conclude that only one group displays normality while the other distribution is skewed.

After the normality checks, the hypothesis test was conducted and results are recorded in the table below.

T-statistic	Df	p-value
3.0365	25	0.002765

Table 15: The Hypothesis testing results

The paired t-test compared the mean crime reported between the “After” and “Before” onset of COVID-19. The test had a t-value of 3.0365 with 25 degrees of freedom. The p-value associated with the test is 0.002765 with the alternative hypothesis set to determine if the true mean difference is greater than zero. What we can capture from the t statistic and p value is that there is sufficient evidence to reject to reject the null hypothesis. We conclude that there is a significant difference in the mean crime counts between the two periods before and after onset of COVID-19.

Key informant findings

Perception of the security situation in the study area

There are various articles about the rise of insecurity in Kasarani sub-county which gives the impression of a `hotspot sub-county'. According to the analyzed data, there has been an increase in reported crime from the period 2018-2022. However, the Deputy OCS noted that there has been an increase in crime rates in the other sub-counties across the country. During the COVID pandemic and its aftermath has the highest recorded crime incidences. The researchers did not get the opportunity to conduct interviews with the residents to get their perception of the security situation in their area.

Community policing in Kasarani Sub County

The concept of community policing is valued in the sub-county. The staff at Kasarani Health Center knew what community policing is and they appreciate its impact. At the health center, community policing is controlled by the Ministry of Interior and National Administration through its representatives. The ministry has made initiative s towards community policing including having a hotline for rape victims and having candid discussions with the community at every chance. The community outreach is not limited to cases related to crime, rising issues like TB and the COVID pandemic are also in the discussion.

Kasarani Police Station has made notable steps in engaging with the community. The area OCPD and his deputy are both integrated into the community. They are both in the `nyumba kumi' initiative, an initiative started by the former President of the Republic of Kenya to ensure that everyone knows their next door neighbor. The station has also had initiatives for boda boda operators. Initiatives that are still in their developing stage but are already yielding results. The station also has a hotline and a trained officer on the receiving end to anyone who is in distress and cannot come to the station.

Impacts of community participation on crime prevention.

The community has a sense of responsibility when it comes to their safety. They have active conversations with the police where they share their concerns, ideas and appreciation. The community once in a while have their moment of strength and they come together and make a great impact. An example is the area popularly known as `Hunters' in Kasarani sub-county. The area was a well-known hotspot and residents lived in constant fear of someone breaking into their homes or being mugged in the streets. The community had a series of discussions and decided to partition their plots into estates. They hired watchmen, security guards and even installed CCTV in their neighborhood. This reduced crime in the area by a considerable percentage and the area is no longer a `no-go zone'. This is one of the examples to show that the residents are a part of this discussion and they are, in their own way, taking responsibility for their safety.

Suggested improvement measures for better community policing.

The Deputy OCS at Kasarani Police Station noted with concern that the community sometimes harbor criminals. This happens with offenders they are familiar with. However, they report cases where the offender is a stranger to them. In this case, they are very willing to share who the culprits are. The Deputy OCS suggested an active partnership between the police and the residents and for the residents to have more confidence in the police. There was also a mention of initiating projects for the youths to keep them engaged. At Kasarani Health Center, improvement measures for better community policing can only be initiated by the Ministry of Interior and National Administration and the researchers did not get a chance to have that conversation with them.

Factors contributing to crime in Kasarani Sub County.

The main factor mentioned by the Deputy OSC at Kasarani Police Station was unemployment. There are varied causes of unemployment which include lack of education, corruption, declining economic growth among others. Another factor is the recent COVID pandemic where there was a 16.6 percent rise in crime incidences in Nairobi County according to the Kenya National Bureau of Statistics. Other factors include poverty, alcohol and substance abuse, peer pressure, political incitement among others.

Has there been an attempt to use Statistical models to predict crime in the area?

At Kasarani Police Station, effort has been made to predict crime occurrences. The system was put in place in 2021 and has been a massive contribution towards reducing crime incidences. When a crime is reported, it is recorded in the physical occurrence book then transferred to a digital occurrence book (Digital OB). They use a system generated model to predict crime using reported crime as their reference data. From this, they can generate a crime clock that is month based. This is critical since there are festive months and crime rates tend to increase during these months. It gives police officers a number to expect so that they can prepare adequately and allocate resources appropriately.

Challenges faced by the police officers when dealing with criminal activities in the area.

The main challenge discussed was lack of resources. The police officers need their gears in order to face whatever situation comes their way. Some of the gears they need include handgun, baton or nightstick, pepper spray, flashlight, radio, first aid kit, body armor or a bullet proof vest among others. Another challenge is the lack of confidence between the police and the residents. There are times the residents perceive the police as the threat and they fail to be cooperative. In recent times there has also been the rising issue about the mental health of the police officers.

CHAPTER 5: CONCLUSION

Summary of findings

Our study has demonstrated the effectiveness of time series modeling in understanding and predicting crime patterns in Nairobi. The developed model highlights the patterns into the temporal dynamics of crime which allows for accurate prediction and forecasting of future crime trends. The findings of this study have several implications for crime prevention and law enforcement strategies. Firstly, the time series model enables authorities to anticipate and proactively address potential crime hotspots. By identifying high-risk periods and areas with a higher likelihood of criminal activity, law enforcement agencies can allocate resources effectively and implement targeted interventions. This proactive approach can lead to a more efficient use of resources and a reduction in crime rates.

Furthermore, the time series model can support the development of early warning systems for real-time crime detection and prevention. By monitoring deviations from the expected patterns, authorities identify the emerging crime trends or unusual activity and take immediate action. This proactive approach can help prevent criminal incidents and improve public safety.

The predictive capabilities of the time series model also have implications for policy evaluation. By comparing the actual crime rates with the predicted values, policymakers assess the impact of crime prevention policies and interventions put in place. This evaluation provides ensures there is feedback on the effectiveness of a specific initiative and thus helps guide evidence-based decision-making in crime prevention strategies. Also, it is quite noticeable that the success of the time series model relies on the availability of accurate and up-to-date data. Continuous data collection, recording and monitoring will be crucial to the maintaining the model's accuracy and relevance. Additionally, regular model updates and refinements based on new data and evolving crime patterns are necessary to ensure its effectiveness over time.

In conclusion, the application of time series modeling in crime prediction offers a significant potential for improving crime prevention strategies in Nairobi. By harnessing the power of data and predictive analytics, authorities can make informed decisions, allocate resources efficiently, and implement proactive measures to reduce crime rates and enhance public safety. These findings from this study will provide a foundation for evidence-based approaches to crime prevention and contribute to the broader goal of creating safer communities and metropolitan areas around the country.

Recommendations

Measures to be placed by the government include:

- Crime Prediction and Forecasting: Utilize the time series model or existing statistical models to forecast future crime trends in Nairobi. The models could provide insights into the expected patterns and fluctuations in crime rates in the county. This will enable law enforcement agencies to allocate resources effectively and address potential crime hotspots.
- Resource Allocation: Utilize the time series model to optimize resource allocation for crime prevention and response. By understanding the temporal patterns of crime, authorities can allocate law enforcement personnel, patrol routes, and surveillance resources in a way that maximizes coverage during high-risk periods or in areas with a higher likelihood of criminal activity.
- Early Warning Systems: Develop early warning systems based on the time series model that can detect and prevent crime in real-time. This can be done by monitoring deviations from the expected patterns. Authorities then can identify emerging crime trends or unusual activity and take immediate action to prevent this criminal incidents.
- Policy Evaluation: Assess the impact of crime prevention policies and interventions by comparing the actual crime rates with the predicted values from the time series model. This evaluation can will determine the effectiveness of specific initiatives which will therefore enable policymakers to make data-driven decisions and adjust strategies as necessary.

The government should also implement the model to various counties in the country to effectively lower the crime rate in the country. The instrumentation of the security sector by investing in tools such as CCTV cameras, motion sensors and better computer hardware to run more than one models at a go in real time.

Recommendations on Future research.

1. The study should cover other sub counties in Nairobi and also other counties in the Republic of Kenya.
2. A study should be conducted that mainly focuses on the community policing and how to increase its effectiveness.
3. Further study should be conducted on the use of statistical models in predicting crime and there should be more proposed models that predict crime.

CHAPTER 6: REFERENCES

Apley, Daniel W, and Zhu Jingyu . 2016. *Visualizing the effects of predictor variables in black box supervised learning models.*

Bovenga, Fabio, Pasquariello Guido, and Refice Alberto. 2021. "Statistically-based trend analysis of MTInSAR displacement time series." In *Remote Sensing 13.12*, 2302.

Cavanaugh, Joseph E, and A Neath Andrew. 2019. "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements." In *Wiley Interdisciplinary Reviews: Computational Statistics 11.3*, e1460.

Chukhrova, Nataliya, and Johannssen Arne . 2019. "Fuzzy regression analysis: systematic review and bibliography." In *Applied Soft Computing 84*, 105708.

Gilbert, Kenneth. 2005. "An ARIMA supply chain model." In *Management Science 51.2*, 305-310.

Hassani, Hossein, and Reza Yeganegi Mohhammad. 2020. "Selecting optimal lag order in Ljung–Box test." In *Physica A: Statistical Mechanics and its Applications 541* , 123700.

Khan, Farhan Mohammad, and Gupta Rajiv. 2021. "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India." In *Journal of Safety Science and Resilience 1.1*, 12-18.

Derrick, B., Annalise, R., Deirdre, T. & Paul, W., 2018. Tests for equality of variances between two samples which contain both paired observations and independent observations. In: *Journal of Applied Quantitative Methods 13, no 2.* s.l.:s.n., pp. 36-47.

Keselman, H. J., Othman, A. R., Wilcox, r. r. & Fradette, K., n.d. The new and improved two-sample t test.. In: *Psychological Science, 15(1).* s.l.:s.n., pp. 47-51.

Mathew, S. & Dan, P., 2021. *Kenya's COVID-19 lockdown is forcing people to make difficult food and household energy decisions.* [Online]
[Accessed 31 May 2023].

Moser, B. K. & Garry, R. S., 1992. Homogeneity of variance in the two-sample means test.. In: *The American Statistician 46.1.* s.l.:s.n., pp. 19-21.

Rochon, J., Mathias, G. & Meinhard , K., 2012. To test or not to test: Preliminary assessment of normality when comparing two independent samples.. In: *BMC medical research methodology 12.1.* s.l.:s.n., pp. 1-11.

Mestre, Guillermo, Portela José , Rice Gregory , Muñoz San Roque Antonio, and Alonso Estrella. 2021. "Functional time series model identification and diagnosis by means of auto-and partial autocorrelation analysis." In *Computational Statistics & Data Analysis 155*, 107108.

Priestley, M B, and Rao T Subba. 1969. "A test for non-stationarity of time-series." In *Journal of the Royal Statistical Society: Series B (Methodological) 31.1*, 140-149.

Torres, José F, Hadjout Dalil , Sebaa Abderrazak , Martínez-Álvarez Francisco , and Troncoso Alicia . 2021. "Deep learning for time series forecasting: a survey." In *Big Data 9, no. 1*, 3-21.

Yakubu, Usman Abbas, and Panji Agung Moch. 2022. "Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for Ewallet Transactions during a Pandemic." In *nternational Journal of Global Operations Research 3.3*, 80-85.

Yaya, OlaOluwa S, E Ogbonna Ahamuefula, Furuoka Fumitaka, and A Gil-Alana Luis. 2021. "A new unit root test for unemployment hysteresis based on the autoregressive neural network." In *Oxford Bulletin of Economics and Statistics 83, no. 4*, 960-981.

Yusn, Ke-Hai, Liu Hongyun, and Han Yuting. 2021. "Differential item functioning analysis without a priori information on anchor items: QQ plots and graphical test." In *psychometrika 86*, 345-377.

Zemoul, Sara-Imane, and Berkoun Youcef . 2022. "Asymptotic behavior of LSE estimator of an AR (1) coefficient with associated innovations." In *Communications in Statistics-Theory and Methods*, 1-8.

CHAPTER 7: APPENDICES

Appendix I: Authorization letter from the Police Headquarters

KENYA POLICE SERVICE

Telegraphic address: "VIGILANCE", Nairobi
Telephone: Nairobi 341411-6
Fax: 330495
When replying please quote

POLICE HEADQUARTERS
P.O. Box 30083-00100
NAIROBI

Ref. No. **KPS/DIG/ADM/21/VOL.II (23)** and date

28th April, 2023

Prof Stephen W. Luketero, PLD
University of Nairobi
Faculty of Science of Technology
Department of Mathematics
P.O Box 30197
Nairobi, Kenya.

RE: REQUEST FOR DATA COLLECTION FOR UNDERGRADUATE RESEARCH PROJECT.

Your letter Ref. UON/FST/ 1/2 dated 26TH April 2023 refers.

Please be informed that your request has been granted. The students named below to report to Sub County Police Commander Kasarani for further guidance on the issue

STUDENT'S NAME	REG NO
1.Cindy Gioria Oluoch	163/2194/2019
2.Maxwell Simiyu	163/4274/2019
3.Claudie Cheptoo	163/4243/2019
4.Francis Njaramba	163/4239/2019
5.Paul Sayianca	163/4270/2019

SUB-COUNTY POLICE COMMANDER
KASARANI
P. O. Box 65020 - 00618,
RUARAKA

Kindly assist them as per my letter A/EST/6/1/VOL.III/193 dated 06/05/2023

Thank you.

FREDRICK M.LAI
For: DEPUTY INSPECTOR GENERAL
KENYA POLICE SERVICE

Appendix II: Authorization letter from the Ministry of Interior and National Administration

OFFICE OF THE PRESIDENT
MINISTRY OF INTERIOR AND NATIONAL ADMINISTRATION
STATE DEPARTMENT OF INTERNAL SECURITY AND NATIONAL ADMINISTRATION

Telegrams..............................
Telephone: Nairobi 316845, 341666
When replying please quote

COUNTY COMMISSIONER
NAIROBI COUNTY
P.O. BOX 30124-00100
NAIROBI

REF NO. ED 10/6 VOL. XXVI (167)

26th April, 2023

Cindy Gloria Otieno
Maxwell Simiyu
Claudia Cheptoo
Francis Njaramba
Paul Sayianka
University of Nairobi

RE: RESEARCH AUTHORIZATION

Your letter dated 25th April, 2023 refers.

This office has no objection and authority is hereby granted to conduct a research on "**Crime Incidences relating to burglary, physical assault, sex related crimes, manslaughter, drug trafficking, carjacking and pick pocketing"** at Kasarani Health Centre within Kasarani Sub County.

J. K. KIAMBI
FOR: COUNTY COMMISSIONER

CC: Deputy County Commissioner
Kasarani Sub County