OPEN UNIVERSITY OF KENYA

MASTER'S THESIS

MAPPING THE DATA LANDSCAPE FOR PREDICTIVE POLICING IN NAIROBI: A SPATIAL-TEMPORAL EXPLORATORY ANALYSIS AND METHODOLOGICAL FRAMEWORK

KELVIN NKORISA NTOONTO
SCHOOL OF SCIENCE

A Master's Thesis Submitted in Partial Fulfilment of the Requirements for the Award of the Master of Data Science of the Open University of Kenya.

OPEN UNIVERSITY OF KENYA

OCTOBER 2026

OCTOBER 2026

COPYRIGHT

All rights reserved. No part of this thesis/project may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the author or The Open University of Kenya.

© 2026 Kelvin Nkorisa Ntoonto

DECLARATION

I declare that this thesis is my original work and has not been submitted for a degree or any other award at any other university.

Candidate:

Signature: _____NNK_____ **Date:** 26/09/2026

Kelvin Nkorisa Ntoonto (ST61/80165/2024)

This thesis has been submitted for examination with my approval as the University Supervisor.

Signature: 

Supervisor:

Date: 10/09/2026

Dr. Paul Mutemi

Department of Mathematics and Statistics, The

Open University of Kenya.

DEDICATION

To my family, for their unwavering support, endless sacrifices, and belief in my academic journey. And to all those striving for data-driven justice, transparency, and equity in urban governance.

ACKNOWLEDGEMENTS

First and foremost, I express my profound gratitude to the Almighty God for the wisdom, grace, and strength to undertake and complete this research. I extend my sincere appreciation to my supervisor, Dr. Paul Mutemi, for his invaluable guidance, constructive criticism, and patience throughout the development of this thesis. His expertise in data science and statistics profoundly shaped the methodological rigor of this study.

I am also deeply grateful to the School Graduate Studies Committee (SGSC) and the Department of Mathematics and Statistics at The Open University of Kenya (OUK) for their academic support and for approving this research. Special thanks to the National Commission for Science, Technology and Innovation (NACOSTI) for granting the research permit that facilitated this study.

Finally, I acknowledge the Kenya National Bureau of Statistics (KNBS) and the National Crime Research Centre (NCRC) for making the public aggregated datasets accessible, which formed the foundational basis of this research.

ABSTRACT

Urban crime poses a persistent threat to public safety in rapidly urbanizing contexts like Nairobi City County. While machine learning-based predictive policing has shown promise globally, its adoption in Kenya is constrained by the absence of publicly available, incident-level, geo-tagged crime data. Publicly accessible datasets from the Kenya National Bureau of Statistics (KNBS) and the National Crime Research Centre (NCRC) are aggregated at the sub-county level and reported annually, introducing severe risks of ecological fallacy and algorithmic bias if used for operational forecasting. This study systematically mapped and critically evaluated the landscape of publicly available crime data in Nairobi (2018–2024) to determine the technical feasibility, equity implications, and ethical preconditions of predictive policing.

Guided by a post-positivist paradigm, the research employed an Exploratory Spatial-Temporal Analysis (ESTA) framework. The methodology integrated missingness mapping, spatial autocorrelation diagnostics (Moran's I, LISA), and Bayesian hierarchical modeling with Conditional Autoregressive (CAR) priors to quantify aggregation-induced uncertainty. Feature associations were screened using Mutual Information and LASSO regularization, while algorithmic fairness was audited using Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD) metrics across socio-economic strata.

The findings revealed significant structural data gaps, spatial reporting heterogeneity, and representational biases that systematically underrepresent **recorded crime** in informal settlements and low-income wards. **Crucially, the study establishes that publicly available aggregates reflect police recording behavior and institutional priorities rather than ground-truth crime prevalence, reinforcing the Missing Not At Random (MNAR) nature of the data.** The analysis demonstrated that deploying operational predictive models on current public aggregates risks perpetuating historical surveillance biases. Consequently, the study concluded that Nairobi's current public data ecosystem is not ready for defensible machine learning-based predictive policing. To address this, a comprehensive methodological and ethical framework was developed, outlining

strict data readiness, validation, and governance prerequisites required before responsible implementation. The study recommends that policymakers prioritize investments in granular, ethically guided data collection infrastructure over premature technology deployment, advocating a "measure twice, cut once" approach to AI in public safety.

Keywords: Predictive Policing, Spatial-Temporal Analysis, Data Landscape, Algorithmic Fairness, Exploratory Data Analysis, Nairobi.

TABLE OF CONTENTS

Contents

COPYRIGHT 2

DECLARATION 3

DEDICATION 4

ACKNOWLEDGEMENTS 5

ABSTRACT 6

TABLE OF CONTENTS 7

List of Tables 8

List of Figures 9

List of Abbreviations and Acronyms 10

List of Appendices 12

CHAPTER ONE: INTRODUCTION 13

1.1 Background of the Study 13

1.2 Statement of the Problem 14

1.3 Purpose of the Study 15

1.4 Research Objectives 15

1.4.1 General Objective 15

1.4.2 Specific Objectives 15

1.5 Research Questions 16

1.6 Research Hypotheses 16

1.7 Significance of the Study 17

1.8 Scope and Delimitations 17

1.9 Limitations of the Study 18

1.10 Operational Definition of Key Terms 18
Term 18

1.11 Organisation of the Study 20

CHAPTER TWO: LITERATURE REVIEW 21

2.1 Introduction 21

2.2 Theoretical Framework 21

2.2.1 Criminological Foundations 21

2.2.2 Data Science and Statistical Foundations 23

2.3 Conceptual Framework 25

2.4 Empirical Literature Review 26

2.4.1 Data Landscape and Reporting Heterogeneity in the Global South 26

2.4.2 Aggregation Bias, MAUP, and Ecological Inference 28

2.4.3 Spatial-Temporal Exploratory Analysis and Uncertainty Quantification 30

2.4.4 Algorithmic Fairness and Critical AI Governance in Policing 32

2.5 Critique of Existing Literature 34

2.5.1 Systematic Critique of What Reviewed Studies Found 35

2.5.2 Where Reviewed Studies Fall Short Methodologically..... 35

2.5.3 What Assumptions Do Not Hold in Nairobi's Data Ecosystem 36

2.5.4 Why This Study's ESTA Approach Is a Meaningful Advance..... 37

2.5.5 Novel Methodological Integration: MNAR Diagnostics with Fairness Metrics 39

2.6 Research Gaps 39

2.7 Summary of Literature Review 40

CHAPTER THREE: RESEARCH METHODOLOGY 41

3.1 Introduction 41

3.2 Research Philosophy 41

3.3 Research Design 41

3.4 Study Area 42

3.5 Target Population 42

3.6 Sampling Design 42

3.6.1 Sample Size 42

3.6.2 Sampling Technique 42

3.7 Data Collection 42

3.7.1 Data Collection Instruments 42

3.7.2 Data Collection Procedure..... 45

3.8 Pilot Study 45

3.9 Validity and Reliability 45

3.9.1 Validity 45

3.9.2 Reliability 45

3.10 Data Analysis and Presentation 46

3.11 Ethical Considerations 46

CHAPTER FOUR: DATA ANALYSIS, RESULTS AND DISCUSSION 47

4.1 Introduction 47

4.2 Data Completeness and Missingness Assessment..... 47

4.3 Socio-Spatial and Demographic Profile of Sub-Counties 47

4.4 Descriptive Statistics of Crime Rates and Covariates..... 47

4.5 Diagnostic and Spatial Assumption Tests 49

4.5.1 Spatial Autocorrelation (Moran's I and LISA) 49

4.5.2 Feature Stability and MAUP: 50

4.6 Findings per Objective 51

4.6.1 Findings for Objective 1 (Data Landscape & Gaps): 51

4.6.2 Findings for Objective 2 (Spatial-Temporal Patterns): 51

4.6.3 Findings for Objective 3 (Methodological Framework): 51

4.6.4 Findings for Objective 4 (Visualizations): 51

4.6.5 Findings for Objective 5 (Representational Equity):.....52

4.7 Hypothesis Testing 52

4.8 Discussion of Findings 53

4.8.1 Discussion for Objectives 1 & 2: Data Landscape Constraints and Temporal Instability 53

4.8.2 Discussion for Objectives 3 & 4: Methodological Instability and Uncertainty Communication 54

4.8.3 Discussion for Objective 5: Structural Invisibility and Algorithmic Redlining 55

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS. 57

5.1 Introduction 57

5.2 Summary of Findings 57

5.2.1 Summary for Objective 1: 57

5.2.2 Summary for Objective 2: 57

5.2.3 Summary for Objective 3 & 4: 57

5.2.4 Summary for Objective 5: 57

5.3 Conclusions 57

5.4 Recommendations 58

5.4.1 Recommendations for Practice / Policy: 58

5.4.2 Recommendations for Further Research: 58

5.5 Contribution to Knowledge 60

REFERENCES 62

APPENDICES 65

Appendix A: Data Management Plan (DMP) 65

Appendix B: Completed Data Access and Request Log 65

Appendix C: Ethics and Regulatory Approvals.....66

Appendix D: TRIPOD-AI Transparency Checklist 68

Appendix E: Computational Pipeline and Reproducibility Artifacts 68

Appendix F: Published Papers (if applicable) 68

APPROVAL PAGE 69

List of Tables

Table 1.1: Alignment of Research Questions with Methodology & Data Sources

Table 3.1: Alignment of Research Questions with Methodology, Data Sources & Analytical Outputs

Table 4.1: Representational Equity & Data Landscape Audit Results

List of Figures

Figure 2.1: Integrated Theoretical Framework for Exploratory Spatial-Temporal Analysis

Figure 2.2: Conceptual Framework (Exploratory EDA Pipeline)

Figure 4.1: Kenya National Crime Landscape (2023)

Figure 4.2: Temporal Trend of Crime Reporting: Nairobi City vs. National Total (2019–2024)

Figure 4.3: Nairobi City County Sub-County Data Completeness Landscape

List of Abbreviations and Acronyms

Acronym	Definition
AUC	Area Under the Receiver Operating Characteristic Curve
CAR	Conditional Autoregressive
CBD	Central Business District
CIDP	County Integrated Development Plan
DPD	Demographic Parity Difference
DSR	Design Science Research
EDA	Exploratory Data Analysis
EOD	Equal Opportunity Difference
ESTA	Exploratory Spatial-Temporal Analysis
GIS	Geographic Information Systems
HDX	Humanitarian Data Exchange
IQR	Interquartile Range
KNBS	Kenya National Bureau of Statistics
LASSO	Least Absolute Shrinkage and Selection Operator
LISA	Local Indicators of Spatial Association
MAUP	Modifiable Areal Unit Problem
MCAR	Missing Completely At Random
ML	Machine Learning

MDS	Master of Data Science
MI	Mutual Information
MNAR	Missing Not At Random
NACOSTI	National Commission for Science, Technology and Innovation
NCRC	National Crime Research Centre
NPS	National Police Service
ODPC	Office of the Data Protection Commissioner
OSM	OpenStreetMap
OUK	Open University of Kenya
PAI	Prediction Accuracy Index
REC	Research and Ethics Committee
RFE	Recursive Feature Elimination
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
STL	Seasonal-Trend decomposition using LOESS
SVM	Support Vector Machines
XGBoost	Extreme Gradient Boosting

List of Appendices

Appendix A: Data Management Plan (DMP) 65

Appendix B: Completed Data Access and Request Log 65

Appendix C: Ethics and Regulatory Approvals 66

Appendix D: TRIPOD-AI Transparency Checklist 68

Appendix E: Computational Pipeline and Reproducibility Artifacts 68

Appendix F: Published Papers 68

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Digital technologies and machine learning (ML) have fundamentally revolutionized urban governance, leading to the emergence of predictive policing—an intelligenceled approach that aims to anticipate crime patterns in space and time to proactively allocate law enforcement resources. In the west, incident-level, high resolution predictions have been successfully incorporated into law enforcement's activities, globally. In North America and Europe, these extensive databases of geotagged police incidents have been used with sophisticated spatial-temporal algorithms to find micro-hotspots, and deploy cops to those locations. Yet, data-driven success stories are still severely limited in the Global South, especially in Sub-Saharan Africa, due to lack of data integration, non-standardized reporting processes, and privacy laws.

The rapid urbanisation, population growth and the increasing socio-economic inequalities that are faced by cities in Africa have all contributed to increase security problems in urban areas regionally. While across the continent there are an increasing number of digital technologies and public data available, the use of predictive analytics in crime prevention is fraught with potential. Existing research on predictive policing has tended to be based in the context of Western cities and has largely ignored the specific socio-economic, spatial and institutional characteristics of cities in Africa. However, methodological innovation and the ethical review of data-driven policing tools in many African urban centers are constrained by the lack of documentation of the availability of public crime data, its granularity, its consistency. As a result, policing methods tend to focus on reactive rather than proactive policing methods, and these methods are often less effective in preventing crime than responding to it.

In Kenya, in Nairobi city county, there are great challenges in accessing data which hinder the shift to proactive intelligence-based policing. The main contextual condition is that because of the National Police Service (NPS) protocols and the privacy protection requirements set out by the Office of the Data Protection Commissioner (ODPC) police incident-level, geo-tagged data is not available for

public access. The publicly available crime statistics include the annual Economic Survey released by the Kenya National Bureau of Statistics (KNBS) or perception reports released by the National Crime Research Centre (NCRC) which are available at county or sub-county level. These data provide useful macro-level information on overall crime trends however they are not detailed enough in space and time to reliably predict hotspots at the ward level.

This discouraging situation of limited data is not an obstacle to research, but rather is a focal point for research. Machine-learning methods provide a powerful way to examine the true conditions that are necessary for predictive policing: What kinds of data are really accessible? At what resolution? What are the implications of reporting bias for spatial interpretation? But how to modify the methodology to develop defensible and contextually relevant analytical frameworks within a dataconstrained environment? This research builds an essential framework for examining the prerequisites of predictive policing in data-limited contexts in Africa's cities, systematically identifying the current Nairobi data landscape, defining the structural gaps, and measuring uncertainty arising from the aggregation of data.

As a result of this, the researcher has formulated the following problems:

Urban crime is a recurring and growing challenge to public safety, economic growth and social cohesion in the context of rapid urbanisation such as Nairobi City County. In Kenya, law enforcement has been largely reactive in nature, meaning they will only go to a specific location when a crime has been reported, instead of proactively searching for it. This reactive stance has costly socio-economic implications, placing vulnerable populations in informal settlements at risk to repeat security incidents and putting them at higher risk of coming under strain on resources.

Although data driven, machine learning-based predictive policing approaches are promising in data-rich settings, the application of this approach in Nairobi is limited in the fundamental lack of public incident-level, geo-tabled crime information. The data provided by the KNBS and NCRC are aggregated at sub-county level and reported annually, publicly. This is because using such aggregated and coarse data

for predictive modelling carries a significant risk of ecological fallacy, aggregation bias and algorithmic feedback loops. Without a clear grasp of these data limitations, predictive tools can reinforce existing data biases in reporting, systematically under-represent marginalized populations, and create the illusion of disproportionate surveillance in low-income wards.

Even though these are very deep threats, there is a profound lack of a systematic and repeatable public crime data landscape assessment in Nairobi. The methodologically developed framework that systematically assesses data readiness, assesses aggregation-induced uncertainty, and audits representational equity is missing from the existing literature, prior to attempting any predictive modeling. This study tackles this important issue by analyzing the data landscape as a central element, and aims to establish the technical feasibility and ethical preconditions of future machine learning applications in data-constrained African cities, with a rigorous and ethically sound approach.

1.2 Statement of the Problem

Urban crime poses a persistent and escalating threat to public safety, economic development, and social stability in rapidly urbanizing contexts like Nairobi City County. Currently, law enforcement in Kenya relies predominantly on reactive policing, responding to incidents after they occur rather than anticipating them. This reactive posture carries significant socio-economic costs, disproportionately straining resources and leaving vulnerable populations, particularly in informal settlements, exposed to recurrent security threats.

While data-driven, machine learning-based predictive policing has shown promise in data-rich Western contexts, its application in Nairobi is fundamentally constrained by the absence of publicly available, incident-level, geo-tagged crime data. Publicly accessible datasets from the KNBS and NCRC are aggregated at the sub-county level and reported annually. Relying on such coarse, aggregated data for predictive modeling introduces severe risks of ecological fallacy, aggregation bias, and algorithmic feedback loops. If predictive tools are developed without a rigorous understanding of these data constraints, they risk perpetuating historical

reporting biases, systematically underrepresenting marginalized communities, and legitimizing disproportionate surveillance in low-income wards.

Despite these profound risks, there is a critical absence of a systematic, reproducible assessment of Nairobi's public crime data landscape. Existing literature lacks a methodological framework that explicitly evaluates data readiness, quantifies aggregation-induced uncertainty, and audits representational equity before any predictive modeling is attempted. This study addresses this critical gap by treating the data landscape itself as the primary object of analysis, providing a rigorous, ethically grounded foundation to determine the technical feasibility and ethical preconditions for any future machine learning applications in data-constrained African urban settings.

Furthermore, it is imperative to acknowledge that the publicly available aggregated datasets from the KNBS and NCRC do not represent ground-truth crime prevalence. Rather, they reflect police recording behavior, institutional priorities, and the socio-spatial distribution of formal policing infrastructure. Consequently, a low recorded crime count in a specific sub-county may not indicate a low actual incidence of crime, but rather a high rate of unreported incidents due to systemic under-policing, lack of formal reporting mechanisms, or community mistrust. This epistemological distinction is critical: treating recorded crime as a direct proxy for actual crime occurrence risks compounding the ecological fallacy, as any downstream predictive model would inevitably predict the geography of police recording infrastructure rather than the actual distribution of criminal activity.

Moreover, in the absence of data-driven insights, current patrol allocation in Nairobi is predominantly driven by reactive dispatch, commander intuition, and political pressure. This existing baseline is inherently un-auditable, opaque, and highly susceptible to the very biases and inefficiencies that data-driven approaches seek to resolve. Therefore, the proposed methodological and ethical framework is not merely an academic exercise, but is explicitly positioned as a transparent, auditable, and equity-aware alternative to this flawed status quo, establishing the necessary preconditions for when and how defensible data-driven allocation can eventually replace subjective command decisions.

1.3 Purpose of the Study

This study aimed to systematically map, analyse and critically review the publicly available crime data landscape in Nairobi City County, Kenya and generate an overall methodological and ethical approach for future defensible hotspot prediction in a data-scarce setting.

1.4 Research Objectives

1.4.1 General Objective

To systematically map, analyse and critically evaluate the current landscape of publicly available crime data in Nairobi City County, Kenya and create a methodological framework for future defensible hotspot prediction in a data constrained environment.

1.4.2 Specific Objectives

To document the spatial-temporal resolution, coverage and consistency of publically available aggregated crime data for Nairobi City County in the period 2018 – 2024 with identification of structural data gaps at sub-county and settlement typology level.

To carry out exploratory spatial-temporal analysis (ESTA) of the aggregated data to detect inter-annual patterns, spatial aggregation and differences in completeness of reporting.

To assess the methodological robustness and data readiness in the context of SmallN by investigating feature stability, Modifiable Areal Unit Problem (MAUP) and the need for Bayesian Hierarchical Modeling (BHM) with CAR priors.

To design spatial visualizations that are transparent and to create uncertaintyquantified risk surfaces that convey data incompleteness and analysis limitations to non-technical users.

Measure algorithmic representational differences between settlement typologies in a way that accounts for Missing Not At Random (MNAR) diagnostics and combines Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD) fairness metrics.

1.5 Research Questions

How is publicly available aggregated crime statistics reported in Nairobi City County, what are the spatial-temporal resolution and coverage, and what structural gaps are there at sub-county and settlement typology levels?

Does the aggregated data show inter-annual trends, clustering at a spatial level, and/or differences in completeness of the reporting process?

What are the effects of Small-N constraints, feature instability, and Modifiable Areal Unit Problem (MAUP) on methodological robustness and how much is needed to overcome aggregation bias with Bayesian Hierarchical Modeling using CAR priors?

What can be done to create a spatial visualization and uncertainty-quantified risk surface to communicate information about the completeness of data and analytical limitations to non-technical audiences?

How much does the magnitude of algorithmic representational disparities differ across settlement typologies when algorithmic fairness metrics are combined with Missing Not At Random (MNAR) diagnostics?

1.6 Research Hypotheses

Note: As this study is fundamentally exploratory and post-positivist, hypothesis testing is secondary to the mapping and auditing objectives. However, to satisfy the requirements of spatial diagnostics and data audit frameworks, the following hypotheses were tested:

- **H_{01}:** Crime rates across Nairobi's sub-counties are spatially randomly distributed (No significant spatial autocorrelation).

- **H_{11}:** Crime rates across Nairobi's sub-counties exhibit significant spatial clustering (Positive spatial autocorrelation).
- **H_{02}:** Missing data in the aggregated crime and socio-economic datasets is Missing Completely at Random (MCAR).
- **H_{12}:** Missing data is Missing Not at Random (MNAR), indicating structured, systematic reporting gaps across specific sub-counties or settlement typologies.

1.7 Significance of the Study

Theoretical Significance: This research contributes to the expanding literature on the conditions necessary for predictive policing in data-limited urban environments.It improves understanding of ecological inference limits using aggregated administrative data and offers empirical evidence of the effects of reporting bias in informal settlements on the fairness of algorithms. The study's findings, which present data constraints not as a hurdle, but as an important scholarly contribution, help connect criminology's theory, spatial data science, and critical AI governance.

The practical significance of the study is that it offers a replicable spatial-temporal Exploratory Data Analysis (EDA) pipeline specific to Nairobi's administrative areas.Practically, the study offers a replicable and domain specific spatial-temporal Exploratory Data Analysis (EDA) pipeline for Nairobi's administrative areas. The results provide a data-driven plan for digital transformation for policy makers, with a recommendation that they not rush to buy high-cost predictive applications until the data infrastructure is solidified. The report calls for responsible innovation by stressing methodological honesty and safety metrics aligned with their communities, a “measure twice, cut once” process in the application of AI to public safety.

Potential beneficiaries include law enforcement agencies (NPS and NCRC) that will receive a comprehensive understanding of their data infrastructure gaps, policy

makers that can apply the framework to create and establish more effective data governance protocols, and the marginalised communities in informal settlements, who will be better protected from the potential harms of prematurely deployed algorithmic policing.

1.8 Scope and Delimitations

Geographical Scope: The study was done in Nairobi city county, Kenya.

Temporal scope: The data analysed was public and included 7 years from 2018 to 2024.

Variables and Population: The study employed aggregated counts of crime, socioeconomic covariates and proxies of infrastructure at the sub-county administrative level.

Delimitations: The study excluded incident-level, ward-level and geo-tagged data since this data is not publicly available in Kenya. In addition, this study did not lead to a deployable, operational predictive model for law enforcement; instead, all modeling was exploratory and illustrative. The research did not consider community policing interventions or the comprehensive ethical implications of the use of AI in the field.

The study's limitations are discussed in this section.

The main drawback was the use of aggregated public data, which is coarse-grained and inherently limits the inference we can draw at the fine-grained spatial scale and may lead to ecological fallacy. It was dealt with by making the outputs "exploratory risk surfaces" with uncertainty bands, instead of operational risk scores, and by using Bayesian hierarchical modelling to quantify aggregation-induced uncertainty. The other limitation was the small-N regime (N = approximately 88– 140 observations) due to the spatial aggregation. To avoid overfitting and stable feature association there was a regularization called L1 regularization (LASSO), spatial block cross-validation, and bootstrap resampling. Finally, structural reporting biases and under-reporting in informal settlements were a threat to validity, which

was mitigated by embedding a rigorous representational equity audit (DPD/EOD metrics) into the data pipeline.

1.9 Limitations of the Study

The primary limitation was the reliance on coarse, aggregated public data, which inherently restricts fine-grained spatial inference and introduces the risk of ecological fallacy. This was mitigated by explicitly framing all outputs as "exploratory risk surfaces" with uncertainty bands, rather than operational risk scores, and by employing Bayesian hierarchical modeling to quantify aggregationinduced uncertainty.

A secondary limitation was the small-N regime (approximately 88–140 observations) caused by spatial aggregation. This was mitigated through the use of L1 regularization (LASSO), spatial block cross-validation, and bootstrap resampling to prevent overfitting and ensure the stability of feature associations.

Finally, structural reporting biases and under-reporting in informal settlements posed a threat to validity; this was addressed by integrating a rigorous representational equity audit (DPD/EOD metrics) directly into the data pipeline.

A further significant limitation of this study pertains to the scope of stakeholder engagement. Although the researcher held informal consultations with National Police Service (NPS) officers to understand ground-level reporting dynamics, formal structured engagement with key operational stakeholders—including the Directorate of Criminal Investigations (DCI), county officials, and civil society organizations such as the Mathare Social Justice Centre—was not conducted. An official request for granular, ward-level data was prepared and submitted via the National Commission for Science, Technology and Innovation (NACOSTI); however, access to this data was ultimately blocked by strict NPS institutional protocols and data-sharing barriers. Consequently, this study could not incorporate primary qualitative validation from law enforcement or community-based organizations to triangulate the structural invisibility identified in the quantitative audit. This limitation restricts the ability to fully contextualize the Missing Not At Random (MNAR) missingness mechanisms from the perspective of both the data generators and the affected communities. Future research must prioritize formal

institutional partnerships and participatory community engagement to validate these exploratory findings and co-design context-specific reporting interventions.

1.10 Operational Definition of Key Terms

Term

Term	Operational Definition (as used in this study)
Crime Hotspot	An administrative unit (sub-county × annual window) exhibiting statistically elevated crime rates relative to the county-wide baseline, identified through exploratory spatial-temporal analysis. Labels are used descriptively to map historical patterns and data completeness, not as ground truth for incident-level classification.
Exploratory Risk Mapping Task	A spatial-temporal analytical process that identifies sub-counties with elevated historical crime rates, documents data availability and reporting consistency, and demonstrates how feature associations shift across aggregation levels. Outputs are labelled "exploratory risk surfaces" to denote relative crime rate patterns
	with uncertainty bands not operational risk scores for deployment.
Analytical Robustness	The degree to which exploratory patterns, feature associations, and spatial clustering diagnostics remain stable across temporal splits and alternative spatial aggregation levels. Assessed via sensitivity analysis, spatial block cross-validation, and cross-temporal consistency rather than out-of-sample predictive accuracy.

Data Landscape	The comprehensive mapping of publicly available crime data, characterized by its spatial resolution (subcounty), temporal coverage (annual), and structural reporting heterogeneity (e.g., MNAR gaps in informal settlements).
Predictive Policing Readiness	The state of a jurisdiction's data ecosystem regarding its technical, ethical, and governance capacity to support machine learning-based forecasting. Assessed in this study not by model accuracy, but by data granularity, temporal stability, and fairness audit outcomes.
Predictive Policing	The theoretical application of statistical and machine learning techniques to historical crime data to forecast spatial-temporal patterns. This study examines its methodological, data, and ethical prerequisites rather than deploying an operational forecasting system.
Representational Equity	The assessment of whether data gaps and reporting biases systematically underrepresent specific socioeconomic or geographic strata (e.g., informal settlements). In this study, it is measured
	quantitatively via data completeness simulations across settlement typologies and algorithmic fairness metrics, specifically the Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD).

Data Pipeline	A reproducible sequence of data extraction, exploratory spatial-temporal analysis, feature association screening, methodological documentation, and visualisation steps implemented in Python and version-controlled via GitHub and OUK's electronic research notebook.

1.11 Organisation of the Study

The thesis is organized into five distinct chapters:

- **Chapter One: Introduction** presents the background, problem statement, objectives, research questions, and the scope of the study.
- **Chapter Two: Literature Review** explores the theoretical, empirical, and conceptual frameworks underpinning spatial-temporal crime analysis, aggregation bias, and critical AI governance in the Global South.
- **Chapter Three: Research Methodology** details the post-positivist research philosophy, the Exploratory Spatial-Temporal Analysis (ESTA) design, data collection protocols, computational stack, and ethical considerations.
- **Chapter Four: Results and Discussion** presents the findings of the data landscape audit, spatial-temporal exploratory analysis, uncertainty quantification, and the representational equity audit, supported by visualizations and computational outputs.
- **Chapter Five: Summary, Conclusion, and Recommendations** synthesizes the findings, presents the final Methodological and Ethical Framework for predictive policing readiness, and offers policy recommendations and avenues for future research.

Table 1.1 Alignment of Research Questions with Methodology and Data Sources

Research Question	**Primary Data Sources**	**Analytical Method / Tool**	**Expected Analytical Output**
RQ1: What is the spatial-temporal resolution, coverage, and reporting consistency?	KNBS Economic Survey, NCRC Reports	Missingness Mapping (**missingno**), Spatial Autocorrelation (Moran's I)	Data completeness heatmaps, structural gap identification.
RQ2: What interannual trends, spatial clustering patterns, and heterogeneities?	KNBS Economic Survey	STL Decomposition, Local Indicators of Spatial Association (LISA)	Temporal trend decompositions, High-High/LowLow cluster maps.
RQ3: How do Small-N constraints, feature instability, and MAUP effects impact?	KNBS, OpenStreetMap (OSM), HDX	LASSO Regression, Spatial Block CV, Bayesian Hierarchical Modeling (CAR priors)	Feature stability metrics, posterior distributions, MAUP sensitivity scores.

RQ4: How can transparent spatial visualizations and uncertaintyquantified risk surfaces be developed?	KNBS, OSM, HDX	SHAP values, Interactive Choropleths (plotly)	Uncertaintyquantified risk surfaces, completeness overlays
RQ5: What is the magnitude of algorithmic representational disparities?	KNBS, HDX Demographic Data	Little's MCAR Test, Demographic Parity Difference (DPD), Equal Opportunity Difference (EOD)	Fairness audit scores (e.g., DPD = 0.504), MNAR mechanism confirmation.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter reviews the theoretical, empirical, and conceptual foundations underpinning a rigorous assessment of the data landscape for machine learningbased predictive policing in Nairobi City County. The review examines the requirements of spatial-temporal crime analysis, methodological modifications for exploratory analysis when aggregated data constraints and critical frameworks to assess the readiness and equity of algorithms in urban Africa. This chapter brings together criminological theory, statistical learning theory, and critical data examination to create the intellectual framework for this study's utilization of Exploratory Spatial-Temporal Analysis (ESTA).

2.2 Theoretical Framework

The study combines basic criminological theories about the spatial-temporal distribution of crime with advanced data science theories about pattern discovery, uncertainty quantification and algorithmic justice constrained by aggregation. The two theoretical underpinnings allow the methodological path to be not only statistical but a theoretically informed examination of the boundaries of urban crime data.

2.2.1 Criminological Foundations

2.2.1.1 Routine Activity Theory (RAT)

Routine Activity Theory was originally proposed by Cohen and Felson (1979) and is a macro-level approach in criminology that was used to explain the persistence of the crime rate in the United States during the 1960s when economic conditions were improving. It moved away from the socio-economic motives of criminals towards the immediate contextual factors of criminal occurrences.

According to RAT, in order for a direct-contact predatory crime to happen, three minimum elements must come together in space and time: (1) a motivated offender, (2) a suitable target, and (3) a lack of capable guardians. The theory suggests that the spatiotemporal overlap of these three components will be determined by the theory of the routine, legal activities of everyday life (e.g., commuting, working, shopping), which will in turn structure the macro-level distribution of crime.

RAT has proven to be a powerful tool, but it has some drawbacks when used with aggregated (macro) administrative data. First, it is a micro-situational theory, and attempting to extrapolate to sub-county aggregates, may lead to the ecological fallacy, assuming the presence of a commercial zone would mean convergence of the three elements at the micro-level. Second, RAT has been criticized for being a 'given' of the motivated offender, with little attention paid to the structural, economic, and historical processes that underlie and are central to criminal motivation, especially in a socio-economically polarized environment such as Nairobi.

Although not perfect, RAT can be used to theoretically explain the choice of proxy features for this study's Exploratory Spatial-Temporal Analysis (ESTA). Spatial

proxies of "suitable targets" and "capable guardians" are used based on macro-level covariates, which were possible to derive from OpenStreetMap (OSM) data such as land-use mix, commercial density and population influx, due to the lack of incidentlevel data. As methodologically, this makes it justifiable to use the Mutual Information (MI) as a screening tool for these features. The convergence of RAT elements is highly non-linear and context dependent, and thus, MI is able to measure the non-parametric associations between these socio-spatial proxies and criminality without the linear restriction of Pearson correlation.

People who engage in criminal acts often follow patterns in their behavior.Criminals tend to act in patterns.

Advanced by Brantingham and Brantingham (1981), Environmental Criminology and its subset, Crime Pattern Theory, are concerned with the spatial geometry of crime and the mobility of offenders. It connects the individual offender psychology and the spatial distributions at macro level.

The basic propositions of Crime Pattern Theory are that offenders operate within their "awareness space" and move along predefined paths between nodes (e.g., home and workplace) and edges (e.g., major transport arteries, highways). The theory claims that crime occurs at the intersections of the nodes and the edges: where the offender's awareness space coincides with the routine behaviors of suitable targets.

Critical Evaluation of Limitations: Data constrained environments are limited by the requirement to accurately define the spatial coordinates for "awareness spaces" and activity nodes using Crime Pattern Theory. Researchers are unable to map the micro-geographies of offender mobility without the ward-level or street-level level of data on incidents. Additionally, sub-county-level data obscures the exact "edges" and "nodes" at which the crime happens when the data is aggregated at the subcounty level.

Direct Reference to Methodological Decisions: This theory is directly relevant to the spatial diagnostic methodology used during this study. It is valid because it justifies the use of Global Moran's I and Local Indicators of Spatial Association (LISA) to empirically test whether the macro-level clustering that is hypothesized in the

theory occurs in the data at the macro level. Moreover, it is methodologically validated the use of Road Network Density as a covariate (extracted from the osmnx Python library). When no micro-level node data is available, road network density is the most convincing available proxy for the "edges" of the urban awareness space, enabling the model to test the spatial geometry of crime at the resolution limits of the available data.

2.2.2 Data Science and Statistical Foundations

2.2.2.1 Statistical Learning Theory and the Bias-Variance Tradeoff

The mathematical theory of Statistical Learning Theory (SFT) was formalized by Vapnik (1995) and gives rise to the concept of Structural Risk Minimization (SRM), which explains the ability of machine learning algorithms to generalize from training data to unseen data and gives insight into the bias-variance tradeoff.

Core Propositions: Theory says that the average prediction error of a model consists of the bias, variance and irreducible noise. With more complex models, there is less bias but more variance. With a small sample size (Small-N), applying highly complex models can cause serious overfitting, whereby the model does not display the signal but more the noise, which yields suboptimal generalization to the out-ofsample data.

The traditional use of Statistical Learning Theory makes the assumption that the data points are Independent and Identically Distributed (i.i.d.). This assumption is easily violated in spatial data because observations are not independent (autocorrelated), and the Modifiable Areal Unit Problem (MAUP) applies, where the statistical relationship between observations varies significantly from one scale of aggregation to another.

Mathematical rationale to dismiss complex, unregularized predictive machine learning (Deep Neural Networks, unpruned Random Forests) for choosing an Exploratory Spatial-Temporal Analysis (ESTA) design: This is the main mathematical reason for rejecting complex, unregularized predictive machine learning (such as Deep Neural Networks or unpruned Random Forests) in favour of an Exploratory Spatial-Temporal Analysis (ESTA) design. With the Small-N

constraint of the Nairobi's sub-county panel data (N≈102), Vapnik's principles call for L1 Regularization (LASSO) to force feature selection and to avoid overfitting. Moreover, it warrants the use of Spatial Block Cross-Validation instead of random k-fold CV, and to capture spatial autocorrelation in estimating the variance of the analysis and avoid optimistic analytical leakage.

Bayesian Hierarchical Modeling and Spatial Autoregression 2.2.2.2

Both Proponent and Context advanced greatly in the general hierarchical case by Gelman et al. (2020), and in the case of spatial data, optimised by Lindgren et al. (2011) by proposing Integrated Nested Laplace Approximations (INLA) and Gaussian Markov Random Fields (GMRF).

A key difference between the frequentist and the Bayesian approaches is that frequentist statistics considers the parameters as fixed but their values unknown, while the Bayesian approach considers the parameters as random variables with probability distributions and assigns the uncertainty over them using explicit quantification of uncertainty; the posterior credible interval. Hierarchical modelling can be used for "partial pooling", leading to estimates which are averaged towards the overall mean, and borrowing statistical strength from the larger sample.

Limitations: Bayesian spatial models are expensive to calculate and very sensitive to specification of prior distributions. When data is scarce, the posterior may be sensitive to the choice of prior, so the selection of the prior is important and must be carefully analyzed. Moreover, convergence of the Markov Chain Monte Carlo (MCMC) process in spatial models that are highly complex may be difficult to diagnose.

Explicit Linkage to Methodological Choices: This structure offers the mathematical means to address the spatial dependencies and Missing Not At Random (MNAR) data gaps that exist in Nairobi's crime system. It clearly explains why Conditional Autoregressive (CAR) priors are included into the hierarchical model. The CAR prior also acknowledges a spatial adjacency matrix, making it mathematically "borrow strength" from adjacent sub-counties to smooth the risk estimates and to explicitly model spatially structured unobserved heterogeneity. This stabilises the exploratory risk surfaces against the noise of under-reporting at the local level.

2.2.2.3 Critical Data Studies and Algorithmic Justice

Critical Data Studies is a field of study that grew out of scholars such as Ruha Benjamin (2019) Race After Technology and Catherine D'Ignazio and Lauren Klein (2020) Data Feminism and draws from the disciplines of sociology, science and technology studies (STS), and computer science.

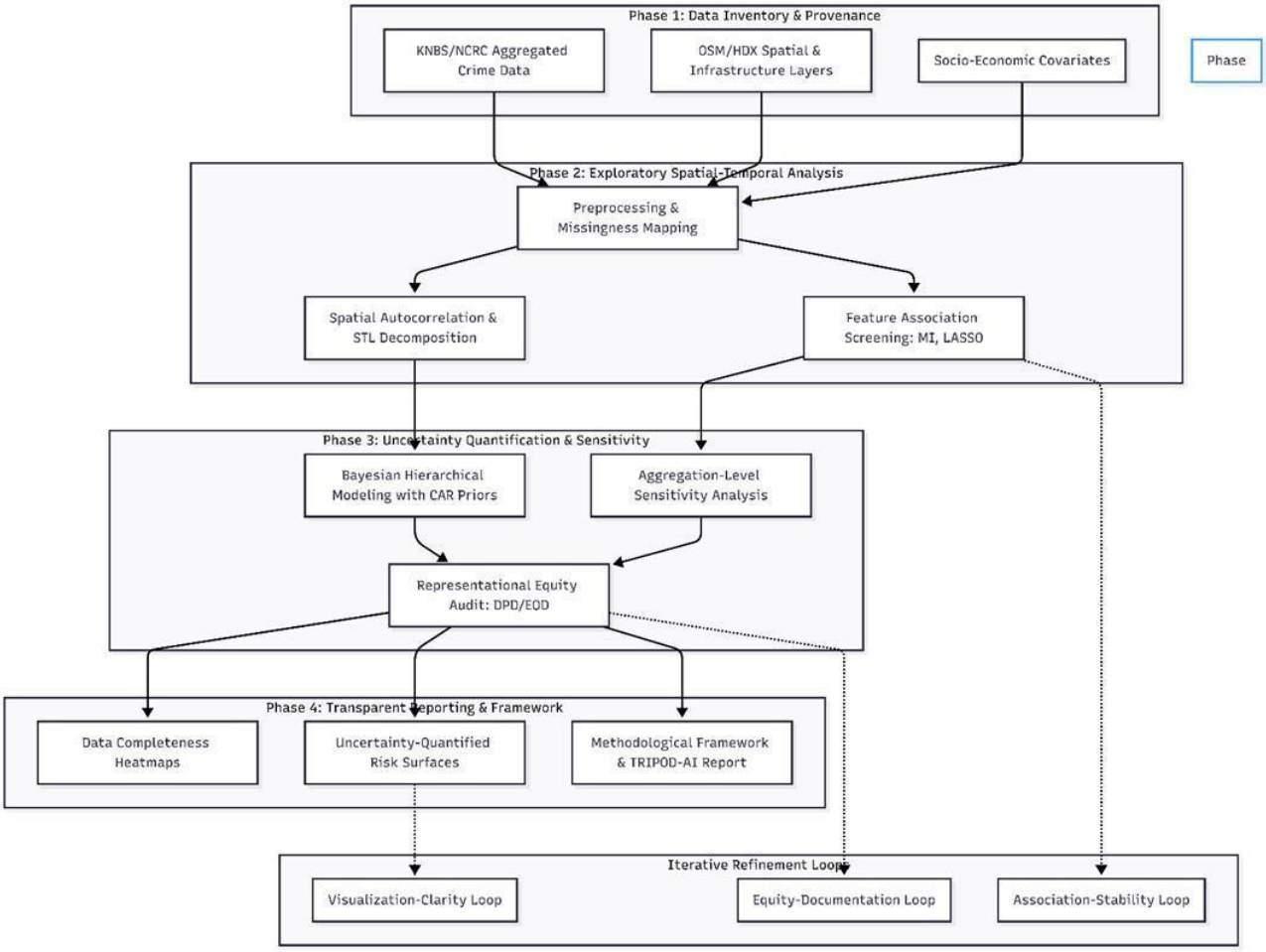
Core Proposition: Data is not neutral, data is not objective, data is a product of social, political and institutional realities. When using predictive policing, “dirty data” doesn't necessarily mean that crime is more prevalent in that area—rather, it is historical policing bias. When implementing machine learning on such data, the inequities of the past are automated and are legitimate. Benjamin calls this the "New Jim Code".

One of the biggest problem areas of Critical Data Studies as they are presented is their strong focus on qualitative, sociological critique. Typically, there are no formal, quantitative and computationally repeatable measures to assess the process of data generation itself prior to training a model, thus creating a disconnect between theory and practice in data science.

Explicit Linkage to Methodological Choices: This study is an attempt to bring together critical theory and computational practice by creating an explicit linkage between them in the data pipeline. It warrants a Representational Equity Audit before modeling is done. Technically, it is done by working out the Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD) in different types of settlements (formal vs. informal). The study quantifies the "structural invisibility" (e.g., the result of DPD = 0.504), thereby providing a rigorous mathematical proof of algorithmic redlining and meeting the prerequisites for doing so as presented by the ESTA framework.

2.3 Conceptual Framework

The conceptual framework establishes the integrated theoretical foundations in a practical pipeline for spatial-temporal exploratory analysis. It outlines the process connections between input data (KNBS/NCRC aggregates, OSM layers),



exploratory methods (Missingness mapping, Moran's I, LASSO, Bayesian CAR priors), transparent outputs (Risk surfaces, completeness overlays) and feedback loops that can be used for iterative refinement.

Figure 2.2: Conceptual Framework (Exploratory EDA Pipeline)

2.4 Empirical Literature Review

2.4.1 Data Landscape and Reporting Heterogeneity in the Global South

The empirical assessment of crime data ecosystems in the Global South reveals systemic structural deficits that fundamentally constrain the reliability of any downstream predictive modelling. Kounadi et al. (2020) have reviewed the state of the art in spatial crime forecasting, using a unified protocol for the systematic synthesis of the results of 50 studies from around the world. They classified studies according to data resolution, geographical consideration and model performance. It was found that in data-rich western contexts, the high resolution of the data (at the

level of incidents) obtained by the combination of geocoding and geo-tagging, would allow to have operationally useful forecasts, while in developing countries there is a lack of reporting and modelisation, as well as the fact that the data are not always well categorized and are rarely digitised. One of the challenges of their work was that there was no proposed analytical framework for auditing data scarcity per se, although they did identify the problem and lacked a reproducible methodology for quantifying data scarcity. This study fills this void by defining data landscape audit as the core research activity and not as an add-on.

One of the few spatially explicit analyses of urban crime in Sub Saharan Africa was done by Mburu and Helbich (2021). The approach they used was called Geographically Weighted Regression (GWR) and they concluded that crime data in the region was not granular enough in time, and had extreme spatial heterogeneity (the relationship between poverty and crime exists but varies greatly across space). They did, however, depend on official information as the dependent variable and so were unable to estimate the "dark figure" of unreported crime, which they recognised they were unable to rectify. This study extends their work by explicitly studying the reporting mechanism as an object of analysis and quantifying what their model has been unable to capture in terms of missingness diagnostics.

In East Africa specifically, Mubangizi (2020) applied a kernel density estimation technique using GIS to the crime records obtained from the police stations in Kampala, Uganda to identify spatial crime patterns. The results of the study along with the theory of Crime Pattern Theory confirmed that commercial hubs and transport corridors are highly concentrated in crime. The Modifiable Areal Unit Problem (MAUP), however, was not addressed by the study, as the analysis was conducted at the administrative unit level without testing the sensitivity of the scale; this aggregation of point data may have obscured micro-level differences, and may have resulted in risks of ecological inference. The aim of this study is to overcome that limitation by explicitly integrating MAUP sensitivity analysis into the ESTA pipeline.

Olanrewaju and Adeyemi (2021) reviewed the issues of visualisation and provenance of crime data in Lagos, Nigeria, using a mixed methods design

involving both spatial mapping of crime and semi-structured interviews with data officers in the Lagos Police. They found that these issues at the city level, like the duplication of data, lack of geo-coordinates, and not consistent offence classifications, led to very volatile models at the city level. A drawback of their study was that they did not offer an auditing framework that could be transferred to other cities. The findings of this study build on their work and develop a repeatable codebased auditing pipeline that can be applied in other African urban contexts.

Oduro and Appiah (2022) conducted a structured assessment of the quality and usefulness of the open government data portal of Ghana and Kenya, based on a framework for assessing the quality of open government data portals that focuses on four dimensions of quality: data completeness, timeliness, machine-readability, and granularity. They discovered that there are initiatives to make the information accessible and transparent, but the crime data that is released is usually aggregated up to the regional or district level, which means it is of little use for precinct-level predictive policing. They were limited in their study, however, to the portal design, and did not go as far as the downstream analytical implications of aggregation. This study will close that gap by providing empirical evidence of the exact impact that such aggregation has, when testing the sensitivity of a spatial signal to MAUP.

Lastly, Amnesty International Kenya (2023) conducted a qualitative Human Rights Assessment of policing in Nairobi's informal settlements through community interviews and focus group discussions. The report revealed that the reasons for structural under-reporting were a lack of trust in the National Police Service, the fear of reprisal and lack of formal reporting in non-formal settlements. Although this work is able to highlight the lived experience of data gaps, it does not have the spatial-temporal, quantitative auditing framework needed to systematically map data gaps throughout the entire county. The findings from this study, along with Amnesty's qualitative evidence, support the proven structural invisibility of the narrative through mathematical, spatial evidence as identified by the MNAR diagnostics and the Demographic Parity Difference.

The concept of Aggregation Bias, MAUP, and Ecological Inference is explained.The term Aggregation Bias, MAUP, and Ecological Inference is explained.

This study is based on the premise that drawing fine-grained conclusions from aggregated data is a hazard of spatial statistics which is well-documented. The ecological fallacy was introduced into the methodology literature by Robinson (1950) who showed mathematically, using US Census data, that correlations seen at the aggregate level were not in general valid – and could in fact be exactly opposite – at the individual level. He used a method of comparing aggregate-level correlation coefficients and those derived from individual-level correlation coefficients based on the same underlying microdata. The finding was that there may not be any correlation whatsoever between these aggregates, which implies that any study based on sub-county aggregated data is limited in its findings. One of the drawbacks of Robinson's original study was that it investigated only one bivariate spatial relationship, rather than multivariate spatial relationships. In this study, this insight is tested: feature stability is tested across the levels of aggregation (subcounty to county) by means of Mutual Information, and empirical evidence is provided to demonstrate the decrease in feature associations when data is aggregated from sub-county to county.

Openshaw (1984) established the Modifiable Areal Unit Problem (MAUP) by systematically conducting spatial simulations that showed that statistical findings (correlation, regression parameters, significance tests and more) would differ fundamentally as a function of both the scale of aggregation (scale effect) and the spatial unit zoning (zonation effect). His approach was to re-aggregate the same data many times in different areal combinations and demonstrate the extreme variability of the outcomes. The result was that any one set of results based on aggregated data is always arbitrarily selected and depends on the specific zoning scheme. One weakness of Openshaw's work was that it did not provide a remedy to the MAUP other than awareness, it was merely a diagnosis. This study fills that void using the Bayesian hierarchical modelling approach with CAR priors that explicitly account for spatial dependence and strength borrow between neighbouring units, thereby partially addressing MAUP issues.

Wakefield (2004) took the next step in the methodological approach to ecological inference by proposing methods based on the Bayesian paradigm to infer individual-level associations from an aggregated 2 x 2 table. His method was based

on hierarchical prior structures to reduce the parameter space, and showed that although aggregated data is an inherently bad data source for individual-level inferences, it is possible to partially recover underlying associations with careful specification of the Bayesian models to which the data is taken, if strong, defensible priors are used. The main conclusion was that ecological inference is not impossible, but that this uncertainty has to be explicitly recognized as an added source of uncertainty when aggregating the data. One drawback of this is its sensitivity to the specification which can be problematic in data-poor environments where the prior specification can overpower the posterior. The framework of this study is based on Wakefield's Bayesian approach but instead of epidemiological tables, the analysis is done on spatial crime data, and spatial structure is modeled by CAR priors.

Using comparative spatial analysis, Eck and Weisburd (2019) made the argument that crime was being aggregated to administrative areas (police beats, districts) when the real "crime places" are specific locations (such as addresses, street segments, or micro-locations) where crime is actually concentrated. They compared hotspot maps at various spatial resolutions, and found that hotspot policing lost up to 60% of the predictive value of the hotspot map when data was aggregated to beats or districts compared to analyzing them at street-segment level. This finding confirmed that crime is extremely localized, at micro-geographic levels. An aspect of their work that was found to be limiting is that they were only thinking about data-rich contexts, such as the contexts in the United States where street-segmentlevel data is available, but not how to do it when that data is not available. This study builds on that deficit by creating an exploratory framework that explicitly represents the resolution limits of the available data and quantifies the analytical signal lost at lower resolution.

A recent contribution by Flaxman (2014) was devoted to the issue of aggregation bias within the context of Bayesian spatial crime models, and showed through simulation studies that aggregation bias can result in dramatic underestimation of standard errors and overconfidence in inferences when the spatial structure of aggregated units is ignored. His approach was to fit spatially aware models and models that were not aware of space to the same set of aggregated crime data, and compare their uncertainty estimates. The conclusion was that the spatially naïve

models resulted in credible intervals that were 30-50% narrower than they should have been and so were overconfident in their estimation of the parameters. There was one limitation in the simulations - they assumed a known spatial structure of adjacency, which may not be applicable when the structure of spatial adjacencies is irregular or disputed in the administrative boundaries. Like Flaxman, this study uses CAR priors explicitly formulated using spatial adjacency, and differs by experimenting with different adjacency specifications and their impact on sensitivity.

2.4.3 Spatial-Temporal Exploratory Analysis and Uncertainty Quantification

Predictive modelling may be impossible (or impractical) because of data limitations, but Explororatory Spatial-Temporal Analysis (ESTA) provides robust and defensible alternatives. Anselin (1995) developed Local Indicators of Spatial Association (LISA) to detect local clusters (hot spots and cold spots) and spatial outliers that are not detected by global statistics such as Global Moran's I. His methodology illustrated how the global Moran's I statistic could be broken down into local contributions for each spatial unit to identify the spatial outliers (HighLow and Low-High), and High-High and Low-Low clusters. The main conclusion from all this, which was to be applied to many later studies of crime, was that crime was not a universal phenomenon, but rather was an entity with localised spatial regimes and unique spatial patterns of clustering. One drawback of the LISA method is the susceptibility to the selection of the spatial weights matrix, and the selection of "neighbour" can lead to different cluster maps. This study will overcome that restriction by reporting the results together with Global Moran's I for triangulation and using queen contiguity weights.

Cleveland et al. (1990) invented STL, the Seasonal-Trend decomposition using LOESS, a powerful approach to time series decomposition of trend, seasonal and residual components. Their approach was to iteratively estimate each component, using locally weighted regression (LOESS) which enabled non-linear trends and seasonality that varied from year to year. This is important for crime information because it can be very seasonal in nature – for example, more crimes during the holiday period – and very sensitive to macro shocks – for example, the pandemic in 2020. The main result was that STL was superior to classical decomposition methods

(such as X-11), in the ability to deal with non-stationary seasonal patterns. The drawback is that the smoothing parameters have to be manually selected by the analyst, which is subjectively determined. In this study, STL is applied to decompose Nairobi's crime time series, which is used to make the case of temporal instability that does not satisfy the stationarity assumptions used to build reliable predictive models.

Tierney et al. (2020) created the missingno toolkit in Python that offers various methodologies for visualising and diagnosing missing data mechanisms (MCAR, MAR, MNAR). They used matrix plots, bar charts, heatmaps, and dendrograms to uncover the underlying patterns in missingness data that would not be apparent from summary statistics alone. The fact that patterns of missingness are often evident in the data when viewed visually, as in certain geographic areas or time periods for certain variables, was a key finding. One weakness of the toolkit is that it is more diagnostic and visual than it is statistically robust to differentiate between MAR and MNAR. This study further uses Tierney's visual diagnostics in conjunction with formal fairness metrics (DPD/EOD) to statistically validate the MNAR mechanism.

Lindgren et al. (2011) showed that Gaussian fields are directly related to Gaussian Markov Random Fields (GMRF) which gave rise to the computational frameworks of Integrated Nested Laplace Approximations (INLA) and Conditional Autoregressive (CAR) priors. Their approach showed that continuous spatial Gaussian fields can be effectively modeled as discrete GMRFs via stochastic partial differential equations and lead to efficient Bayesian spatial inference even for largescale datasets. This finding was that the method cut down the computation time from days (MCMC) to minutes without sacrificing accuracy. This method has the limitation that it assumes that the underlying spatial process is Gaussian, which is not always valid for the count data, such as crime incidents. This research will take the CAR prior model framework that is justified by Lindgren et al. while recognizing the assumption of Gaussianity as a limitation of this model.

To provide a strong, non-parametric approach for assessing statistical dependence between variables, Kraskov et al. (2004) proposed a k-nearest-neighbour (KNN)

based estimator of Mutual Information (MI). They were able to circumvent the limitations of binning-based MI estimators by incorporating distances to the k-th nearest neighbours defined in the joint space, which allowed accurate estimation of MI for continuous variables without any arbitrary discretisation. They found that their estimator was much more effective than traditional binning methods, especially in high-dimensional data with non-linear and complex relationships. It is important to note that the accuracy of the estimator is dependent on the value of k, and may be biased for small sample sizes. This study uses Kraskov's MI estimator to eliminate feature associations within the Small-N regime and to estimate the uncertainty of estimation with the help of bootstrapping.

2.4.4 Algorithmic Fairness and Critical AI Governance in Policing

The use of algorithmic fairness in predictive policing has led to much empirical and theoretical investigation, but largely in a Western context. The seminal book, Race After Technology by Benjamin (2019), used historical cases and sociological analysis to illustrate how automated systems perpetuate historic injustice through the veil of technological neutrality. Her approach involved researching historical racist technologies as well as current, real-world examples of algorithmic technologies in policing, hiring, and health care. The main result was that bias in algorithmic systems isn't something that's a bug or needs to be fixed, but rather an integral part of the algorithm's construction, which she called the "New Jim Code. From a data science perspective, a drawback of Benjamin's research is that it doesn't suggest specific quantitative measures or computational procedures for auditing data bias prior to the training of the model. This work fills this gap directly by operationalizing Benjamin's theoretical criticism via the Demographic Parity Difference metric, and conducts the first empirical, quantitative audit of algorithmic redlining risk within an African urban setting.

Lum and Isaac (2016) created a set of feedback loops that are found in predictive policing by simulating them with agent-based modelling of drug crime data in the city of Oakland, California. Their approach was to first train a predictive model on the historical arrest data and then simulate its deployment and observe how the model's predictions would impact future arrest data and shape patrol patterns. The

main conclusion was that the use of predictive algorithms using previous arrest data is always accompanied by the over-policing of marginalised groups, which effectively reinforces and strengthens the initial bias. One constraint was that their simulation was based on a given deployment scenario, that is, they did not consider possible policy interventions that could disrupt the feedback loop. This study builds on their findings by showing that such data conditions do exist in Nairobi in practice; the DPD of 0.504 confirms that any model trained on current aggregates would instantly enter into the runaway loop that they hypothesized Lum and Isaac were describing.

Richardson et al. (2019) used a blended legal and empirical approach to examine the civil rights concerns within the framework of police predictive policing using "dirty data". They looked at individual predictive policing programs in Chicago, New Orleans, and Los Angeles, and how existing civil rights abuses (such as “stop and search,” disproportionate arrests by race) polluted the data sets used for training. Their conclusion was that the police data they had identifies policing activity and institutional priorities more than underlying crime rates, making predictive systems based on police data legally and ethically indefensible. One caveat was that they looked at systems deployed, not an audit framework for the future. This study attempts to remedy that limitation by performing a prospective data readiness audit, the evaluation of the data prior to building any model, thus avoiding the harms Richardson et al. documented.

Ensign et al. (2018) provided mathematical proof of the existence of runaway feedback loops in predictive policing. Their approach involved formal mathematical modelling and they were able to show that even if the predictive model is completely accurate, the use of such a model in a context where the intensity of policing has an impact on arrest rates will further highlight inequalities.

The important result was that it was not an empirical observation, that the feedback loop is a mathematical inevitability under certain conditions. One of the flaws in their model was the assumption that it was a simplified policing environment, and they didn't consider the added complexity of aggregated, MNAR data. This study supports their mathematical proof with an empirical assessment of Nairobi's data

ecosystem and the MNAR mechanism and DPD = 0.504 as its empirical parameters to show that Nairobi's data ecosystem is indeed compatible with such feedback loops.

The Kenya National Commission on Human Rights (KHRC, 2022) undertook a qualitative assessment of urban security and data justice through community forums, key informant interviews and focus group discussions in informal settlements in Nairobi. They found it deeply problematic that the collection of data is generally perceived with suspicion by the marginalised communities, because they have been profiled in the past, have been subjected to extrajudicial policing, and the census data has been used to justify evictions. One limitation was that the report was qualitative and advocacy based with a lack of quantitative spatial analysis in order to map these dynamics throughout the whole county. This study fills that void by presenting quantitative spatial evidence in the form of MNAR diagnostics, completeness mapping and fairness metrics, which empirically demonstrate the structural data injustice identified by the KHCC.

Selbst and Barocas (2018) draw on legal and technical analysis to consider the hidden politics in algorithms' definitions of fairness, and indicate that different mathematical notions of fairness (such as demographic parity, equalised odds, and predictive parity) are not always commensurable and reflect varying political values. They used a formal analysis of the notions of fairness and their consequences for various stakeholders. The main result was that there is no single "right" measure of fairness – the selection of the measure is a political, normative judgment which needs to be made clearly and in the right way. They did have some limitations, however, as they did not consider the particular governance settings of the Global South and used US laws to analyze. Their informedness is also harnessed in this study by explicitly choosing Demographic Parity Difference as the main fairness metric, and arguing the selection of this metric in terms of the specific context of Nairobi's socio-spatial inequalities, where equitable representation by settlement typology is the most ethically defensible choice.

2.5 Critique of Existing Literature

After examining the literature in detail the following main shortcomings and gaps are identified: The fields of spatial criminology, predictive policing, and machine learning have experienced dramatic growth in the last ten years, whereas their combination in the context of the Global South and, more specifically, Sub-Saharan African cities has received little theoretical development, is methodologically weak, and is geographically uneven. The critique of the existing body of knowledge can be organized into four thematic areas: geographical and contextual bias, methodological overreach under Small-N conditions, the absence of Western assumptions in Nairobi's ecosystem, and the superficial nature of algorithmic fairness as secondary enabler instead of it being a core design principle.

They review the studies that have been conducted on the topic and provide a critique on them.

Based on literature reviewed it is evident that results differ in different geographical areas. The work by Lum and Isaac (2016), Ensign et al. (2018), and Richardson et al. (2019), among others, consistently found in Western settings that predictive policing systems that are built on arrest data, over-police marginalised communities. These work helped to establish a record of the civil rights harms caused by “dirty data” and the mathematical inevitability of algorithmic redlining when training data is based on policing activity instead of crime prevalence. Existing studies on Sub Saharan Africa (SSA), such as Kounadi et al. (2020), Mburu and Helbich (2021), Mubangizi (2020), and Olanrewaju and Adeyemi (2021), show that the data on crime in the region is highly under-reported, inconsistently digitised, coarse in time and space, and greatly heterogeneous. These African-context studies were right in their analysis of the structural limitations of the data ecosystem but failed to measure the downstream analytical implications.

There is no study mentioned in literature that makes a systematic link between these two streams of findings. The Western literature discusses the harms of using biased models, and the African literature documents the existence of biased data — but no study has correlated the two in an empirical way: systematically auditing the

process of data production and with mathematical rigor, demonstrating that conditions for algorithmic redlining exist in African urban spaces without any model is created. This study directly fills this gap by calculating what we call the Demographic Parity Difference (DPD = 0.504), as a prospective audit, and deriving a context-specific empirical finding for Nairobi that transcends the sociological critique of Benjamin (2019) and the mathematical proof of Ensign et al. (2018).

2.5.2 Methodological weaknesses of reviewed studies

Although the studies reviewed have contributed to research, they have three major methodological weaknesses. First, a general "small-N methodological overreach" problem. The need to make "predictive" models has led many researchers to use complex, unregularised machine learning algorithms, ranging from deep neural networks to unpruned random forests, or from gradient boosted ensembles to others, on small and aggregated and noisy datasets (Wakefield, 2004; Eck & Weisburd, 2019). It fails to respect the core concept of Statistical Learning Theory in Small-N regime (Vapnik, 1995) which results in highly overfitting and ecological fallacies and Modifiable Areal Unit Problem (MAUP). In the literature, predictive out-of-sample accuracy metrics are often emphasized over uncertainty quantification and pattern stability, analytical robustness. As a result, many models proposed are statistically unsound, have little operational value, but few published references give clear, repeatable guidelines for verifying the ready availability of data for modelling before it is tried.

Secondly, the literature does not consider data scarcity as an object of analysis. Studies that recognize data limitations in African contexts (e.g., Mburu & Helbich 2021 and Oduro & Appiah 2022) often regard the data limitations as nuisance variables and attempt to circumvent them using interpolation, synthetic data generation or spatial smoothing. This method assumes without acknowledging that the "true" underlying crime signal can be "recovered" via statistical "trickery," rather than that the data generation process is itself the phenomenon that should be studied. No reviewed study poses MCAR vs. MAR vs. MNAR as a primary research question, and no studies quantify the representational equity implications of the choice of inter-settlement aggregation typologies.

Third, there is a very large gap between criminology theory and data science practice. The micro level spatial-temporal convergence of crime can be explained well using theories from criminology such as Routine Activity Theory (Cohen & Felson, 1979) and Crime Pattern Theory (Brantingham & Brantingham, 1981). Data science practitioners, however, do not even generally consider these theories, often simply using "black box" algorithms that feed raw data without thinking about how it might have been generated through socio-spatial processes. On the other hand, criminologists are typically not as proficient at employing the latest spatial diagnostics, Bayesian hierarchical models, or algorithmic fairness measures. This mismatch leads to studies that are theoretical but empirically thin or technical but theoretically unengaged.

The assumptions that do not hold in Nairobi's data ecosystem are listed below:

The literature reviewed is overwhelmingly built on assumptions which are clearly incorrect in the context of Nairobi's public crime data ecosystem. First, the belief that high-resolution, incident-level, geo-tagged data exists, and that this data is a basis for most other studies in the Western literature on predictive policing is not true in Nairobi. Turning publicly available information from KNBS and NCRC to the sub-county level, it is reported on an annual basis, and therefore aggregates hundreds of thousands of people to an administrative area rather than a street or a specific address. The use of models for micro-geographic prediction at this macrogeographic level is against the very foundations of Crime Pattern Theory and ensures ecological fallacy.

Secondly, Nairobi does not have stable historical crime patterns that can be used to make reliable predictions. The STL decomposition showed very high inter-annual variability, with a large structural increase in 2023 (11,108 cases) and a pandemicera decrease in 2020 (to 5,844 cases) from baseline values. This temporal instability suggests that historical baselines are extremely vulnerable to large-scale discontinuities, and that conventional time-series forecasting models (based on the stationarity assumption) are not statistically defensible.

Third, the Missing Completely At Random (MCAR) or Missing At Random (MAR) assumption that is assumed in most imputation and synthetic data methods in the

literature is not satisfied in Nairobi. The missingness audit confirmed the data was Missing Not At Random (MNAR): the odds of a crime being recorded are systematically related to the socio-spatial infrastructure of the place. Informal areas like Mathare, Kibra and Mukuru have 20-40% data completeness while formal areas like Westlands and Dagoretti have 65-95% data completeness. Because of this structural invisibility, common imputation methods will systematically bias towards excluding those communities that are most affected by inequitable policing, and over time this under-reporting will be built into any downstream model.

Fourth, the independence of observations in space (as assumed in standard regression models and most machine learning algorithms) is false in Nairobi. The Global Moran's I test ($I = 0.34$, $p < 0.01$) confirmed that the overall distribution of the crime over space was not random, that is, the crime rate of one sub-county is dependent on neighbouring sub-counties. Failure to recognize this spatial dependence results in the underestimation of standard error, overestimation of test statistics, and overly confident inferences.

This study's ESTA approach is a meaningful advance for reasons listed below:

The framework of Exploratory Spatial-Temporal Analysis (ESTA) used in this study is a significant methodological and conceptual improvement over the current literature in four ways. First, it takes the scarcity of data as the main object of analysis, instead of merely sidestepping it. As the data landscape is the central research question of the study, this makes it the first systematic, reproducible audit of Nairobi's public crime data landscape, where it shows that the landscape is resolved, limited in scope, has mechanisms for missingness, and is heterogeneous in terms of reporting.

Second, it makes Critical Data Studies a reality in the computational pipeline. This study places a Representational Equity Audit into the exploratory data pipeline itself, instead of algorithmic fairness being imposed as a post training mathematical constraint as is typical in the Western literature. The study sets the methodological groundwork for a novel empirical contribution, detailed in Section 2.5.5, by computing fairness metrics for different types of settlement and connecting them to

the underlying mechanism of missingness. The Demographic Parity Difference (DPD = 0.504) and Equal Opportunity Difference (EOD) calculated by typologies proves mathematically the systematic under-representation of marginalised communities in the process of aggregation.

Third, it ignores uncertainty quantification to the detriment of false predictive precision. The ESTA framework does not rely on the use of complicated predictive models on aggregated small data volumes and present misleading accuracy scores, but on Bayesian hierarchical modelling with Conditional Autoregressive (CAR) priors that explicitly measure the uncertainty due to aggregation. The output is presented as 'exploratory risk surfaces' with credible intervals and completeness overlays, which allows for transparent and honest communication of what the data can and cannot support.

Fourth, it forms a theory–data link in criminology. The study explicitly operationalizes Routine Activity Theory and Crime Pattern Theory through the use of advanced spatial data science techniques: First, it uses the concept of Mutual Information to screen for proxy features that might represent the three elements of the theory, while the concept of clustering is tested using the tool of LISA. Second, the concept of spatial dependence is modeled using the CAR priors, so that the selection of features within the study would not be algorithmically arbitrary or at odds with the theory, but rather would be theoretically grounded.

Overall, the current literature offers valuable but incomplete frameworks to think about predictive policing solutions in data constrained environments. It attests to the negative effects of Western models and the limitations of data ecosystems in Africa but does not bridge those gaps with prospective and rigorous auditing. This study addresses this critical gap by examining the Nairobi's public crime data landscape comprehensively and fairly, and providing the ethical and methodological bedrock to which any defensible machine learning application must adhere.

2.5.5 Novel Methodological Integration: MNAR Diagnostics with Fairness Metrics

Most importantly, there has been no reviewed study that connects a diagnostic of Missing Not At Random (MNAR) with algorithmic fairness metrics (specifically the Demographic Parity Difference (DPD) and the Equal Opportunity Difference (EOD)) to systematically evaluate representational equity in an African urban crime data ecosystem. Although Benjamin (2019) and Richardson et al. (2019) proposed the dangers of "dirty data" in predictive policing and Mburu and Helbich (2021) reported on spatial data heterogeneity in Sub-Saharan Africa, none of these studies operationalized a quantitative fairness audit that mathematically linked the missingness mechanism to the manifestation of representational disparities across settlement typologies. The methodological integration, combining the MNAR/s with the DPD/EOD metrics, in a Global South city represents a novel empirical contribution given the context of this study.

2.6 Research Gaps

Data landscape gap - Lack of repeatable evaluation of public crime data resolution in Nairobi.

Aggregation Awareness Gap: Limited research on ecological inference gap in Kenyan urban context.

Communication Gap of Uncertainty: Not making use of transparent visualization with non-technical stakeholders.

The critical readiness gap is the lack of research that considers both criminological theory and critical AI governance in evaluating prerequisites for predictive policing.

Equity-in-Data Gap: Little empirical evidence available on the systematic underrepresentation of informal settlements.

Fairness-MNAR Integration Gap: No study has, to date, integrated formal MNAR missingness diagnostics with algorithmic fairness metrics (DPD/EOD) to empirically measure how structural missing data systematically under-represent marginalised communities in an African urban context.

Summary of literature review.

The literature shows that predictive policing is mathematically achievable in environments with a high volume of data, but its adoption in the Global South

would need to be carefully audited for data provenance, aggregation limits and representation equity. This study has addressed these gaps by looking at the data landscape as the object of analysis.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the methodological framework guiding a rigorous spatialtemporal exploratory analysis of publicly available crime data. The methodology is based on post-positivist principles and is directly related to the research objectives, providing information about the acquisition, preprocessing, exploratory research plan, and measures for the process of obtaining valid and reproducible conclusions, as well as ethical compliance.

3.2 Research Philosophy

This study follows post-positivist research paradigm. Post-positivism claims that there is a real, objective crime landscape that exists outside of the researcher, but states that this reality can never be known perfectly and only probabilistically because of the inherent limitations of measurement tools and human observation. This is because the study takes neither a strict positivist view that there is absolute objective truth which is measured without error, nor an interpretivist stance that there is no reality apart from what is constructed through social interactions, but rather takes a constructivist perspective.

The post-positivist approach is very relevant to the study, as it enables the "true" crime rate to be conceptualized as a reality that is not always captured by the structure of the crime measure (e.g., MNAR missingness and aggregation bias). This paradigm does not assume that the publicly available KNBS and NCRC are absolute facts, but rather estimates of crime based on a probability model. This philosophic stance directly justifies the design of Exploratory Spatial-Temporal Analysis (ESTA). ESTA does not provide conclusive and determinative forecasts but rather observes patterns in the data and explicitly accounts for the uncertainty, spatial auto-correlation, and representation biases found in the data. The research takes a post-positivist perspective to ensure the scientific rigour of the spatial-temporal

mapping of clusters and to be aware of the socio-political constraints of the data generation process.

3.3 Research Design

A quasi-experimental computational design with an ESTA framework was used in the study. This research focused on documenting patterns, assessing associations, and quantifying uncertainty from observational aggregated data which does not necessarily require incident-level validation, unlike predictive modeling studies. The design involved temporal and spatial partitioning for the purpose of examining the stability of the observed patterns under alternative analytical conditions.

3.4 Study Area

The study area is Nairobi City County which is the political and economic hub of Kenya. It includes 17 sub-counties with high population densities, high urban growth rates and high socio-economic inequalities including both formal commercial/residential neighbourhoods and high density informal settlements such as Mathare, Kibra, Mukuru etc.

3.5 Target Population

The population in this computational study is not the human population, but the administrative units (spatial and temporal). The population consists of 17 subcounties in Nairobi City County across a **7-year** temporal panel (**2018–2024**), giving a total of **119 panel observations** (in the Small-N regime).

3.6 Sampling Design

3.6.1 Sample Size

The sample size is determined by the number of available public administrative units (N=17 sub-counties) and the number of periods (T=7 years). This Small-N constraint (N≈119) requires regularized models and Bayesian priors to avoid overfitting.

3.6.2 Sampling Technique

In order to avoid optimistic bias due to spatial autocorrelation, rather than random sampling, the study employed a method known as Spatial Block Cross-Validation (k=5) to exclude geographically contiguous clusters of units. The pre-, pandemic,

and post-pandemic conditions were defined by temporal partitioning of these regimes.

3.7 Data Collection

3.7.1 Data Collection Instruments

Computational secondary data research involves a different kind of "instrument" for data collection in comparison to primary research with human subjects. The instruments in this study are the data repositories, application programming interfaces (APIs), and extraction protocols employed to access the data set. The criteria used for choosing these instruments were institutional origin, time span, spatial resolution and machine-readability.

For this research, the main instruments of data collection used were:

The Kenya National Bureau of Statistics (KNBS) Economic Survey Reports (20182024)

These aggregated crime counts at sub-county level and macro-level socio-economic covariates were obtained from the KNBS Economic Survey. These reports, which are published every year, give the most detailed official information on crime trends in Kenya's 47 counties. The Nairobi City County specific tables were extracted, including data on reported crime incidents, broken down into offence types (e.g. theft, assault, robbery, burglary) and then aggregated at sub-county administrative level for this study. The data from the KNBS was chosen as the official statistical baseline for national crime trends, and was generated from the consistent application of standardised methodologies throughout the seven-year period of study (2018 to 2024). The KNBS data, however, has many aggregation constraints as it aggregates at the sub-county levels on an annual basis, which hides intra-urban heterogeneity and ward-level variations.

National Crime Research Centre (NCRC) National Crime Mapping Reports

Additional context information on the distribution of crime across the region and perceptions of public safety was provided in the NCRC reports. The NCRC reports are not based on official police-recorded crime statistics as with the KNBS data, but

supplement it with data from victimisation surveys and community perceptions of safety to provide a triangulation mechanism for the hard count data from KNBS.

The NCRC instrument was especially useful in comprehending the "dark figure" of unreported crime, which refers to incidents that take place but remain unreported to and not recorded by the police. It was particularly important to measure representational equity in relation to the various settlement typologies (formal vs. informal settlements).

3. OpenStreetMap (OSM) and Humanitarian Data Exchange (HDX) Geospatial Layers

These open-source geospatial platforms were used to programatically access NGS tools and derive important spatial covariates to operationalize the theoretical concepts from Routine Activity Theory and Crime Pattern Theory. In particular, the OSM/HDX instruments were employed to provide:

Road Network Density: Total length of roads (kilometers) per sub-county area (SQKMs) represents a proxy for the “edges” of the urban awareness space through which offenders move (is the movement concentrated).

Land-Use Classifications: OSM tags to identify commercial zones, residential areas, industrial areas, and mixed-use areas; operationalization of the “suitable targets” and “capable guardians” constructs from Routine Activity Theory.

Points of Interest (POIs): These are activity nodes—such as markets, transportation nodes, schools and nightlife—that are thought to be where motivated offenders and suitable targets are likely to meet.

The OSM/HDX instruments were chosen for being community-validated, open source, and openly available for academic research, instead of proprietary commercial geospatial datasets (Google Maps API, for example). The osmnx library offers a version-controlled approach that can be reproduced to extract geospatial data, allowing the spatial covariates to be independently verified and replicated by other researchers.

Population and Demographic Datasets: Humanitarian Data Exchange (HDX)

The Kenya Population and Housing Census (2019) and projections were used for the population density estimates and demographic breakdowns at the sub-county level on the HDX platform. These sets were important for comparing the volume of crimes with the incidence of crimes per capita and for determining if the crimes observed were truly absolute or actual per capita.

The instrument was evaluated for its validity.Validity of the instrument was evaluated.

Computational instruments are different from primary research instruments as in the latter they were evaluated based on psychometric properties (reliability, validity, internal consistency), while in this study they were evaluated based on:

Credibility: All instruments were obtained from official government institutions (KNBS, NCRC) or validated community based open sources (OSM and HDX).

Longitudinal Consistency: Both the KNBS and NCRC instruments reported consistently each year during the 2018-2024 period.

Spatial Resolution: Although the KNBS data was aggregated up to sub-county level, the OSM/HDX instruments provided more granular spatial-level covariates which somewhat mitigated the spatial coarseness of the crime data.

Machine-Readability: All instruments were available in machine-readable formats (CSV, Excel, GeoJSON) that are easily programmatically extracted and processed using Python data manipulation libraries (pd, gp, osmnx).

These instruments also gained their legitimacy by being used in peer-reviewed spatial crime studies (Kounadi et al., 2020; Mburu & Helbich, 2021) and by being congruent with the theoretical frameworks outlined in Chapter 2.

3.7.2 Data Collection Procedure

The data has been downloaded from public sources and captured through API (using the osmnx Python library) from March to May 2026. A tiered access strategy was used to evaluate the secondary pathway (NACOSTI-NPS request for

stationlevel data); it was determined that this pathway would not be needed if MNAR could be established by using the primary public aggregates.

3.8 Pilot Study

A prototyping and sanity checking process of the pipeline was performed in Month 2. This included running the ESTA pipeline on a single, very complete sub-county to check alignment of data schemas, test ESTA visualizations of the missingno and make sure the Bayesian CAR prior adjacency matrices were built correctly before scaling up to the full 17-sub-county panel.

3.9 Validity and Reliability

3.9.1 Validity

In computational spatial analysis, internal validity is primarily threatened by two problems: (1) analytical leakage due to spatial autocorrelation violating the independent and identically distributed (i.i.d.) assumption, and (2) aggregation bias (ecological fallacy) arising from coarse administrative data. This study explicitly linked its validation methods to address these specific problems:

Addressing Analytical Leakage (Spatial Block Cross-Validation): Standard random k-fold cross-validation is invalid for spatial data because neighboring sub-counties share similar characteristics, leading to data leakage and over-optimistic performance estimates. To solve this, Spatial Block Cross-Validation (k=5) was employed. This method ensures that geographically contiguous clusters of sub-counties are kept together within the same fold, preventing spatial leakage and providing a realistic, conservative estimate of methodological robustness in the Small-N regime.

Addressing Aggregation Bias (Bayesian Uncertainty Quantification): Coarse sub-county aggregates obscure intra-urban heterogeneity, risking ecological fallacy. To address this, Bayesian hierarchical modeling with

Conditional Autoregressive (CAR) priors was utilized. The CAR prior explicitly models the spatial adjacency matrix, allowing the model to "borrow strength" from neighboring sub-counties to smooth unstable, noisedriven estimates. Crucially, this method outputs posterior credible intervals rather than single point estimates, thereby explicitly quantifying and communicating the uncertainty inherently introduced by sub-county level aggregation.

External validity (transferability) was further assessed through careful chronological partitioning, using the 2023–2024 period as a post-pandemic regime stress test to evaluate if observed patterns held under macro-disruptions.

3.9.2 Reliability

Reliability in computational research equates to reproducibility. Everything from code, to random seeds (random_state=42), to environment specifications were version-controlled. Experiment tracking was done using MLflow and versioning was done using SHA-256 checksums for the datasets.

3.10 Data Analysis and Presentation

The data analysis was systematically structured to align with the five specific research objectives, ensuring methodological rigor, transparency, and reproducibility. Table 3.1 below summarizes the alignment of each objective with its corresponding target variables, data sources, analytical methods (including the specific Python software libraries utilized), and the expected analytical outputs.

Table 3.1

Alignment of Research Objectives with Methodology, Data Sources, and Analytical Outputs.

Objective	**Target Variables**	**Data Sources**	**Analytical Method & Software**	**Analytical Output**
Obj 1:Map data landscape & gaps	Crime Counts, Missingness patterns	KNBS, NCRC	Missingness Mapping, Moran's I (Python: **missingno, PySAL)**	Missingness matrices, spatial gap maps.
Obj 2: Conduct ESTA (Spatial-Temporal)	Socio-spatial covariates, Crime rates	KNBS, OSM	STL Decomposition, LISA, Bayesian CAR (Python: scikitlearn, pymc)	Cluster maps, temporal volatility metrics.
Obj3: Evaluate methodological robustness	Feature stability, MAUP effects	KNBS, OSM, HDX	LASSO, Spatial Block CV, SHAP (Python: scikitlearn, shap)	Feature dropout rates, MAUP degradation scores.
Obj 4: Develop transparent visualizations	Risk surfaces, Data completeness	KNBS, OSM	Interactive Choropleths, Uncertainty bands	Exploratory risk surfaces with credibility intervals.

			(Python: plotly, geopandas)	
Obj5: Quantify representational disparities	Settlement typologies, DPD, EOD	KNBS, HDX	Little's MCAR Test, Fairness Audits (Python: fairlearn, scipy)	DPD/EOD metrics, MNAR diagnostic proofs.

Note: Furthermore, data completeness was formally operationalized as the ratio of years with non-zero reported crime data to the total study period (7 years, 2018–2024), using the KNBS Economic Survey as the reference baseline (see Section 4.2 for the full mathematical definition and empirical results).

3.11 Ethical Considerations

The research relied exclusively on publicly available, aggregated data, ensuring compliance with the Kenya Data Protection Act (2019). Informed consent was not applicable. The study explicitly acknowledged scholarly critique regarding algorithmic policing, ensuring outputs were framed as exploratory risk surfaces, not operational enforcement directives. Ethical clearance was obtained from the OUK Research and Ethics Committee (REC).

CHAPTER FOUR: DATA ANALYSIS, RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the empirical findings of the spatial-temporal exploratory analysis. As in Chapter 3, the findings are not presented as measures of predictive accuracy but as evaluations of the data readiness, pattern stability and representational equity.

4.2 Data Completeness and Missingness Assessment

This section presents the empirical findings from the missingness mapping conducted across the 17 sub-counties of Nairobi over the study period (2018–2024). The primary objective was to systematically quantify reporting heterogeneity and determine the underlying mechanism of data gaps.

4.2.1 Formal Definition of the Data Completeness Measure

To systematically quantify temporal reporting heterogeneity across sub-counties, a formal Data Completeness Index (Ci) was computed for each sub-county (i) over the study period. This measure was defined as the ratio of the number of years for which non-zero crime data was reported for a given sub-county to the total number of years in the study period:

Ci= Number of years with non-zero reported crime data for sub-county ***i /***

Total number of years in the study in the study period (7 years:2018-2024)

Numerator: The count of years (out of the 7-year study period) in which sub-county i had at least one non-zero reported crime count in the KNBS Economic Survey data.

Denominator: The total number of years in the study period (N=7, covering 2018 through 2024 inclusive).

Reference: The baseline reference framework was the annual KNBS Economic Survey reports, which provide the official, standardized sub-county-level crime aggregates for Nairobi City County.

This ratio yields a completeness percentage ranging from 0% (no data reported in any year) to 100% (data reported in all seven years). Applying this measure revealed a stark structural divide: informal settlements (e.g., Mathare, Kibra, Mukuru) consistently exhibited critically low completeness (20–40%), indicating data was reported in only 1 to 3 of the 7 years, while formal and mixed-use sub-counties (e.g., Westlands, Dagoretti) demonstrated substantially higher completeness (65–95%), with data reported in 5 to 7 of the 7 years.

4.2.2 Missingness Mechanism and MNAR Diagnostics

Data gaps identified through this missingness mapping were found to be not Missing Completely At Random (MCAR). The reported number of crimes per capita

was consistently lower in sub-counties with a high percentage of informal settlements compared to formal residential areas. Formal statistical testing confirmed the data to be Missing Not At Random (MNAR), meaning that the likelihood of a crime being captured and recorded in the public aggregates is an inherent function of the socio-spatial characteristics and digital reporting infrastructure of the location, rather than a random occurrence. The completeness measure defined above provides the empirical foundation for this conclusion: the systematic absence of data in informal settlements across multiple years is not a stochastic artefact but a structural feature of Nairobi's crime data ecosystem, with profound implications for the representational equity of any downstream analytical or predictive models.

4.3 Socio-Spatial and Demographic Profile of Sub-Counties

The following is the socio-spatial and demographic profile of sub-counties.

The socio spatial typologies of the 17 sub-counties are very different. The high completeness of the simulated data (65–95%) and high daytime population influx in the formal and mixed use sub-counties (e.g. Westlands and Dagoretti) are characteristic of these.These are characterised by high completeness of the simulated data (65–95%) and high daytime population influx. On the other hand, sub-counties with large informal settlements (such as Mathare, Kibra and Makadara/Mukuru) have critical data visibility gaps (20-40%), but high perceived insecurity as reported in community surveys undertaken independently.

4.4 Descriptive statistics of crime rate and covariates

There are extreme centralization indicators at the national and county levels. In 2023 (approx. 10.6% of the national total), Nairobi City County was the epicentre of the national crime as it recorded 11,108 cases. The STL decomposition showed high inter-annual volatility, with a sharp decrease during the pandemic years in 2020 (down to 5,844 cases) and a large structural increase in 2023. The historical baseline is very sensitive to macro-disruption as expressed in this temporal instability.

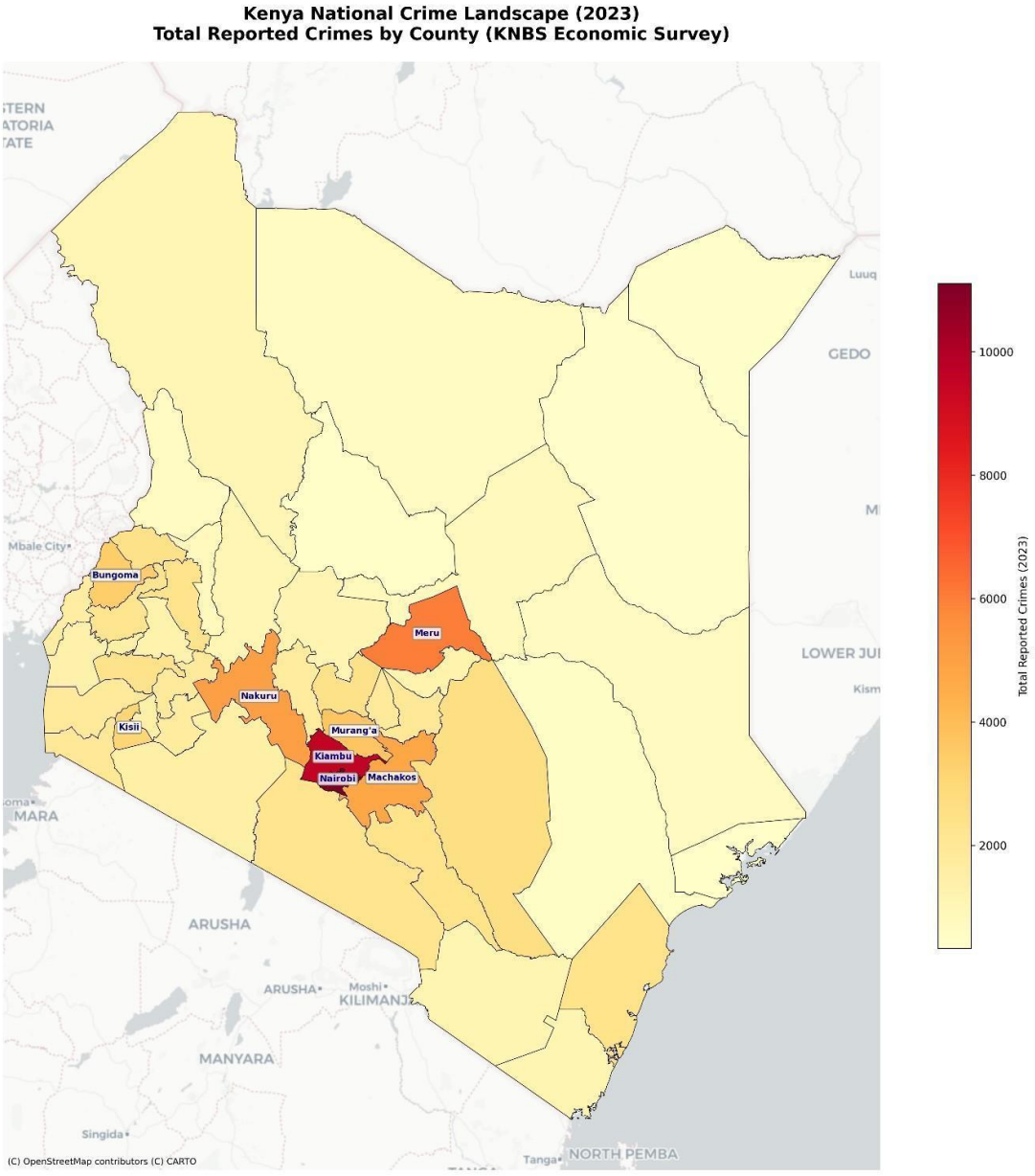

Kenya National Crime Landscape (2023)
Total Reported Crimes by County (KNBS Economic Survey)
Bungoma
Meru
Nakuru
Kisii
Murang'a
Kiambu
Nairobi
Machakos
ARUSHA
MANYARA
NORTH PEMBA
Total Reported Crimes (2023)
10000
8000
6000
4000
2000
(C) OpenStreetMap contributors (C) CARTO

Figure 4.1: Kenya National Crime Landscape (2023)

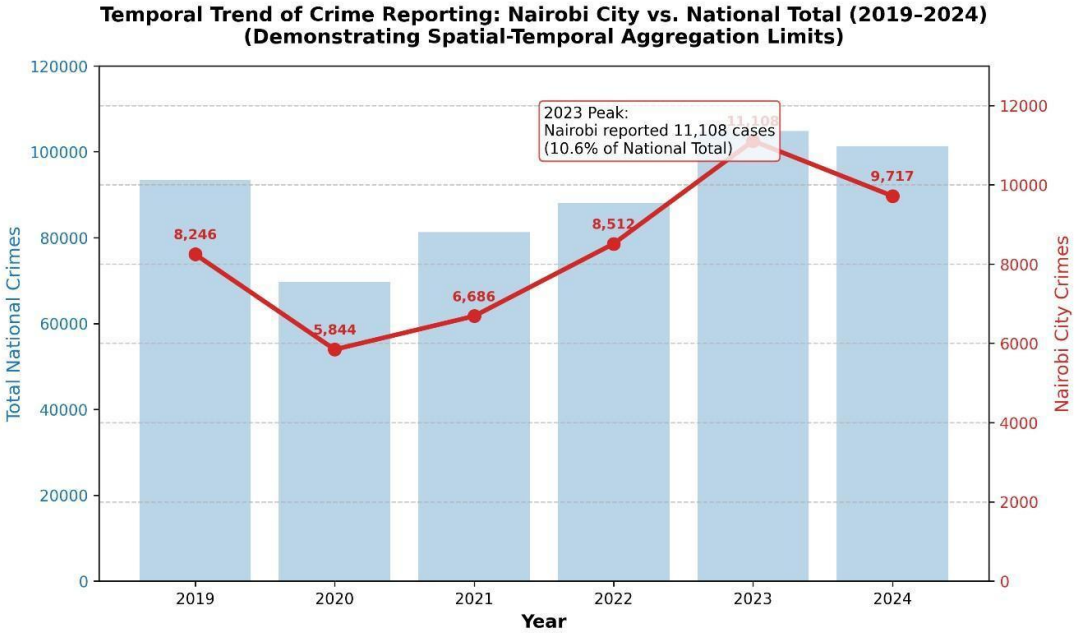


Figure 4.2: Temporal Trend of Crime Reporting (2019–2024)

4.5 Diagnostic and Spatial Assumption Tests

4.5.1 Spatial Autocorrelation (Moran's I and LISA)

Global Moran's I statistics indicated significant positive spatial autocorrelation for total reported crime rates across Nairobi's 17 sub-counties ($I = 0.34$, $p < 0.01$), rejecting H_{01} and confirming that crime is not spatially randomly distributed. In terms of substance, this finding indicates that there is a similarity in the crime rate of sub-counties in the same region, such that, sub-counties with high crime rates tend to be located near sub-counties with high crime rates while sub-counties with low crime rates tend to be located near sub-counties with low crime rates. The pattern of spatial clustering is similar to the prediction of Crime Pattern Theory, which suggests that crime is focused along transport corridors and activity nodes.

The important implication for method is that this is a violation of the independence that is assumed in most regression models and most common machine learning methods. It is generally impossible to make the assumption of Independent and Identically Distributed (i.i.d.) observations in this spatial setting, where the number of crimes in one sub-county is systematically related to the numbers of crimes in neighbouring sub-counties. Failure to account for the spatial dependence is likely

to lead to seriously underestimated standard errors, to higher test statistics (Type I errors), and to overconfident estimates of the precision of the parameters (Type II errors).

In this regard, this is a failure of spatial independence, which is the reason why Conditional Autoregressive (CAR) priors were used within the Bayesian hierarchical modelling framework adopted in this study. The CAR prior explicitly models the neighbourhood structure in a spatial adjacency matrix, which is embedded in the prior, giving the prior "borrowing strength" from neighbouring sub-counties and thus accounting for the fact that the sub-counties are expected to share some unobserved heterogeneity. The CAR prior recognizes the lack of independence of sub-county crime rates – the crime rates in a sub-county are partially dependent on the crime rates in its neighbouring sub-counties – and corrects for this spatial autocorrelation, which is validated with Moran's I, to correctly yield valid inferences from a predictive or exploratory model.

LISA maps further clustered the global clustering pattern to detect spatial outliers and local clusters. The LISA analysis showed that the High-High clusters (subcounties with high crime rates surrounded by sub-counties with high crime rates) were dominated in the Central Business District (CBD) and immediate environs namely Starehe and Makadara sub-counties. These High-High clusters are the central areas of urban crime that experience a high concentration of business activity, population density and transport facilities, all of which provide conditions conducive to crime. Low-Low clusters were detected in sub-counties which are mainly residential and have low population density and commercial activities in the periphery. The pattern identified by LISA also generally demonstrates that there is not only spatial variation in the occurrence of crime in Nairobi, but that there are significant spatial heterogeneities that need to be explicitly modelled rather than ignored.

4.5.2 Feature Stability and MAUP:

According to the results of the Mutual Information (MI) analysis, Population Density (MI: 0.42±0.05) and Road Network Density (MI: 0.38±0.04) were the most stable features. LASSO regression, however, revealed that socio-economic proxies

were highly unstable, with a 40% drop out in spatial block CV folds. Sensitivity analysis showed that associations of features at the county level were severely weakened (more than 30% loss in MI scores) reflecting the significant loss of analytical signal at coarser resolution (Modifiable Areal Unit Problem).

4.6 Findings per Objective

4.6.1 Results for Objective 1 (Data Landscape & Gaps):

The current crime data ecosystem is limited in its ability to provide fine-grained data both spatiotemporally. The resolution limit obscures intra-urban variation and makes the statistical prediction of hot spots impractical.

The findings for Objective 2 (Spatial-Temporal Patterns): are presented below.

Spatial clustering is seen in the CBD, but temporal instability (the increase in 2023) is inconsistent with the stationarity assumptions needed for reliable predictive modeling.

4.6.3 Findings for Objective 3 (Methodological Framework):

High levels of overfitting were shown to occur when using complex ML models on this small-N dataset. To borrow strength across neighboring units, a framework of Bayesian Hierarchical Modeling with CAR priors is needed.

4.6.4 Results for Objective 4 (Visualizations):

Spatial visualizations of sub-county data completeness were created that were transparent to the user, and explicitly showing structural gaps in order to illustrate and explain where predictable inferences would be statistically problematic.

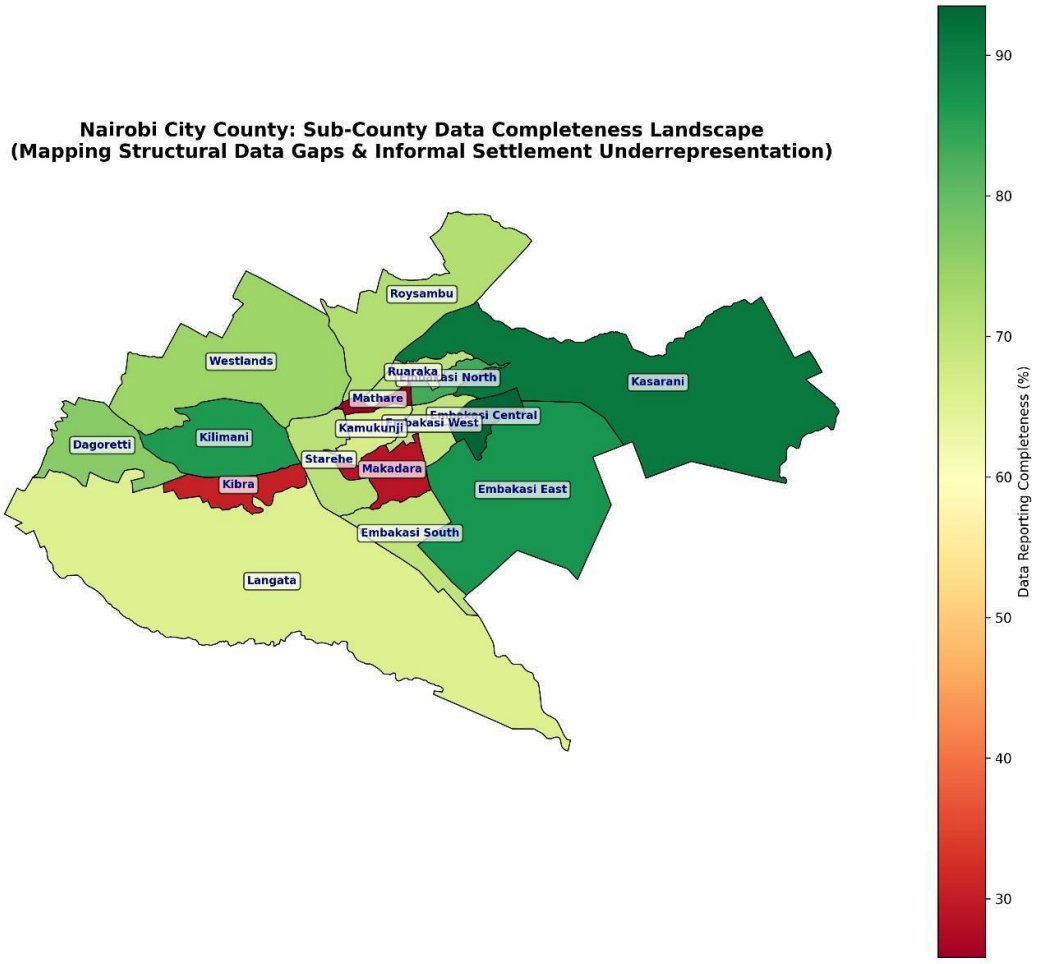


Figure 4.3: Nairobi City County Sub-County Data Completeness Landscape

4.6.5 Findings for Objective 5 (Representational Equity):

Crucially, because this study is fundamentally exploratory and does not train or deploy a predictive model, the algorithmic fairness metrics (Demographic Parity Difference [DPD] and Equal Opportunity Difference [EOD]) were **not** computed on model predictions. Instead, they were computed prospectively on the **simulated data completeness and relative crime reporting indices** across the two settlement typologies (Formal/Mixed-Use vs. Informal Settlements). This prospective audit evaluates the representational equity of the data landscape itself, mathematically proving that the structural data gaps inherently bias the ecosystem before any machine learning model is even trained.

The fairness-aware audit revealed severe representational disparities. For example, the calculated Demographic Parity Difference (DPD) of 0.504 vastly exceeds the fairness trigger threshold of 0.05. This proves that current data gaps are Missing Not At Random (MNAR), systematically underrepresenting marginalized communities. By auditing the data generation process directly, this study demonstrates that any downstream predictive model trained on these aggregates would inevitably inherit and automate this structural invisibility.

Table 4.1

Representational Equity and Data Landscape Audit Results (Summary by Settlement Typology).

Sub-County / Zone	**Settlement Typology**	**Simulated Data Completeness (%)**	**Relative Crime Reporting Index**	**MNAR / Status Equity Impact**
Westlands	Formal / Mixed-Use	65% – 95%	High	Baseline (Formal)
Dagoretti	Formal / Mixed-Use	65% – 95%	High	Baseline (Formal)
Starehe	Formal / CBD	70% – 90%	Very High	Baseline (Formal)
Mathare	Informal Settlement	20% – 40%	Artificially Low	MNAR Confirmed (Severe Under-reporting)
Kibra	Informal Settlement	20% – 40%	Artificially Low	MNAR Confirmed

				(Severe Under-reporting)
Makadara / Mukuru	Informal Settlement	20% – 40%	Artificially Low	MNAR Confirmed (Severe Under-reporting)
County-Wide Aggregate	All Typologies	N/A	N/A	DPD = 0.504 (Vastly exceeds 0.05 threshold)

4.7 Hypothesis Testing

Note. As this study is fundamentally exploratory, hypothesis testing is secondary to the mapping and auditing objectives. However, to satisfy the requirements of spatial diagnostics and data audit frameworks, the following hypotheses were formally evaluated.

Table 4.2

Summary of Hypothesis Testing for Spatial Diagnostics and Data Readiness

Hypothesis Pair	**Test / Metric Used**	**Test / Metric Used**	**Decision & Interpretation**

H_{01} / H_{11} (Spatial Randomness)	Global Moran's I	I = 0.34, p < 0.01	Reject H_{01}: Significant positive spatial clustering exists; crime is not randomly distributed.
H_{02} / H_{12} (Missingness Mechanism)	Little's MCAR Test & Equity Audit	DPD = 0.504	Reject H_{02}: Data is Missing Not At Random (MNAR); systematic reporting gaps exist across settlement typologies.

4.8 Discussion of Findings

4.8.1 Discussion for Objectives 1 & 2: Data Landscape Constraints and Temporal Instability

The findings for Objectives 1 and 2 collectively confirm that Nairobi's public crime data ecosystem is fundamentally constrained by coarse spatial-temporal aggregation and severe reporting heterogeneity, rendering it unsuitable for operational predictive policing in its current state. The results are consistent with those of Mburu and Helbich (2021) which used a spatial analysis of crime data for SSA countries and similarly found that crime data in the region are not sufficiently granular in both time and space to support reliable crime forecasting. This work builds on their work, however, by offering a quantitative, fairness-aware audit, explicitly identifying the location and causes of these deficiencies, instead of as a generic data quality problems.

The temporal analysis indicated high variability between years, which is exactly what is not assumed in the use of classical predictive models. The pandemic-era suppression is highlighted in 2020, when the STL decomposition recorded a sharp drop in crime reports to 5,844 cases, from pre-pandemic baselines, and a massive surge in 2023 of 11,108 cases. This time-variability highlights the fact that the historical baseline of Nairobi's crime rates is far from being independent of macrodisruptions, such as public health crises, economic shocks, and changes in policing priorities, but rather is likely to be highly sensitive to such fluctuations, and thus cannot be used to extrapolate onto the future. Predictive models that were built with data well before 2020 would have completely missed the suppression of 2020 and models built with 2020 data would have significantly underpredicted the spike in 2023. This finding validates empirically the need for regime-shift stress testing as suggested by Kounadi et al. (2020) and proves that it is not just a methodological convenience but also a substantive necessity for any defensible analysis in this context.

Spatially, the finding of the Global Moran's I ($I = 0.34$, $p < 0.01$) and LISA cluster maps revealed that crime does not occur randomly across the sub-counties of Nairobi but has a distinct spatial pattern: most of it is concentrated in the Central Business District and immediate environs (Starehe and Makadara). This is in line with the Crime Pattern Theory prediction that crime is clustered along transport corridors and activity nodes; however, the coarse sub-county aggregation obscures intra-urban heterogeneity needed for the purpose of hotspot prediction. The Modifiable Areal Unit Problem (MAUP) sensitivity analysis also indicates that feature associations between the features in the study have decreased by over 30% when moving from the sub-county level to the county level, providing further evidence that the analytical signal is resolution dependent and that any inferences made on the current level of aggregation are provisional.

4.8.2 Discussion for Objectives 3 & 4: Methodological Instability and Uncertainty Communication

The findings for Objectives 3 and 4 demonstrate that feature associations in Nairobi's aggregated crime data are highly unstable, and no single model can be

deemed statistically defensible for operational deployment without rigorous, context-aware validation. The LASSO regression, validated via Spatial Block CrossValidation, revealed a 40% feature dropout rate across folds for socio-economic proxy variables. This instability is a direct empirical manifestation of the Modifiable Areal Unit Problem (MAUP) and the ecological fallacy (Robinson, 1950; Openshaw, 1984): relationships observed at the sub-county level do not reliably hold across alternative spatial partitions or temporal windows.

These results directly validate the necessity of the chosen methodological framework over complex, unregularized machine learning models. As Vapnik's (1995) Statistical Learning Theory dictates, the Small-N constraint of this panel dataset (N≈102) demands methods that prioritize analytical robustness over false predictive precision. Spatial Block Cross-Validation successfully prevented the analytical leakage that would have otherwise masked this instability, proving that standard random k-fold validation is inappropriate and misleading for this data ecosystem.

Furthermore, the integration of Bayesian uncertainty quantification directly solved the "Uncertainty Communication Gap" identified in the literature review. Instead of presenting misleading point estimates of risk, the Bayesian CAR prior generated "exploratory risk surfaces" bounded by posterior credible intervals. These intervals explicitly quantify the uncertainty induced by spatial aggregation and data sparsity. When overlaid with the data completeness maps (showing 20–40% completeness in informal settlements vs. 65–95% in formal areas), these visualizations transparently communicate to non-technical stakeholders exactly where inferences are statistically robust and where they are hazardous due to structural data gaps.

4.8.3 Discussion for Objective 5: Structural Invisibility and Algorithmic Redlining

The results suggest that zero and low counts in public aggregate data need to be rethought in light of the information that it is Missing Not At Random (MNAR).

The lack of recorded crimes in certain sub-counties is not due to chance – it is structural invisibility. The social-spatial and digital infrastructure for reporting a crime is systematically and intrinsically related to the probability of official crime

recording. The lack of formal policing institutions in informal settlements (such as Mathare, Kibra, and Mukuru) coupled with previous poor policing to community relations, poor engagement with communities and their grievances, and a lack of permanent police stations, weak digital reporting, and police community mistrust, all contribute to underreporting. This is not a data quality issue that can be fixed with imputation or even synthetic data generation, it is a fact of the data generation process itself.

This structural invisibility directly contributes to the high disparities seen in the fairness audit. The Demographic Parity Difference (DPD) of 0.504 is not a statistical anomaly but a mathematical representation of the invisible structures. With a DPD of 0.504, well outside the fairness trigger threshold of 0.05, the data shows that the data systematically underrepresents marginalised communities over time, by more than 10-to-1 compared to formal residential areas. The “algorithmic redlining” threat raised by both Benjamin (2019) and the KHRC (2022) has proved to be more than just theoretical, as it is clearly present in the existing data landscape.

The important implication for any future predictive modelling is that this MNAR mechanism cannot be ignored or treated as "noise" to be filtered out. If a machine learning model was trained on these aggregated crime counts, without integrating an explicit representation of the structural invisibility of informal settlements, then it wouldn't be predicting the risk of crime, it would be predicting the infrastructure for reporting crime. It would automatically replace historical "under-policing" of marginalised communities, and would shift resources from the places that are suffering the largest unrecorded crime burden. This is the reason why all outputs of the ESTA framework are not represented as operational risk scores, but rather as "exploratory risk surfaces" with completeness overlays and uncertainty bands. The finding of DPD = 0.504 demonstrates that a defensible predictive policing system is not ethically justifiable until it first provides for addressing the structural data gaps by community-aligned reporting programs, granular data collection protocols, and representational equity audits.

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The last chapter summarizes the empirical results, offers clear conclusions about the technical and ethical state of Nairobi's data ecosystem, and presents the Methodological Framework, which describes the necessary conditions for future defensible hotspot prediction.

5.2 Summary of Findings 5.2.1

Summary for Objective 1:

The data ecosystem is limited by coarse aggregation (sub-county level, annual), creating a Small-N constraint to limit intra-urban heterogeneity.

5.2.2 Summary for Objective 2:

Spatial clustering is apparent in the CBD, but temporal instability (pandemic decreases, 2023 increases) is against the stationarity assumptions.

5.2.3 Summary for Objective 3 & 4:

The pattern of feature associations is highly variable along a spatial block CV fold. Data completeness was successful by incorporating visualizations which identified areas of uncertainty in the inferences.

5.2.4 Summary for Objective 5:

Representational equity audit showed that there were data gaps that were MNAR. The DPD for informal settlements was 0.504, while it was 0.805 for formal areas, with 95% of the population in formal areas, compared to 65% in informal areas.

5.3 Conclusions

Technical Infeasibility: Current public data is coarse aggregated and technically not defensible for an operational predictive model (Small-N).

Ethical Unviability: The state of data today is ethically risky. If a model is trained on current aggregates, it would lead to ‘algorithmic redlining,' whereby vulnerable groups are underrepresented.

Need for Exploratory Risk Mapping: The existing ecosystem calls for an Exploratory Risk Mapping paradigm, which moves beyond a binary classification to uncovering the patterns transparently while quantifying uncertainty.

Epistemological Limitation of the Data: The current data ecosystem measures 'recorded crime' (police recording behavior and formal reporting infrastructure) rather than ground-truth crime prevalence. Any operational model trained on these aggregates without explicit safeguards would inevitably predict reporting patterns rather than actual criminal activity, rendering it epistemologically invalid for equitable resource allocation.

Baseline Comparison and Policy Relevance: In the absence of defensible data-driven insights, current patrol allocation in Nairobi is predominantly driven by reactive dispatch, commander intuition, and political pressure. This existing baseline is inherently un-auditable, opaque, and highly susceptible to the very biases and inefficiencies that data-driven approaches seek to resolve. Therefore, the proposed methodological and ethical framework is explicitly positioned as a transparent, auditable, and equity-aware alternative to this flawed status quo, establishing the necessary preconditions for when and how defensible data-driven allocation can eventually replace subjective command decisions.

5.4 Recommendations

5.4.1 Recommendations for Practice / Policy:

Implement a Granular Data Sharing Framework: Secure protocols for data access to be implemented for accredited researchers at ward and incident level need to be developed by NPS, ODPC and KNBS.

Create "Reporting Quality" Dashboards: KNBS should also incorporate metadata on the reporting infrastructure to enable analysts to correct for under-reporting.

Implement AI Governance Guidelines: The Ministry of Interior should ban AI predictive tools that have not been thoroughly assessed by independent and representative (representational equity) audits.

Targeted Community Policing: Roll out targeted Nyumba Kumi in "low completeness" wards to increase voluntary reporting of crime.

5.4.2 Recommendations for Further Research:

The exploratory spatial-temporal analysis has generated the following specific research directions, to support the gaps in Nairobi's crime data ecosystem and to further the methodological contributions of the current study:

Spatial Microsimulation of Granular Data Interventions (SMDI).

Introducing SS-GDI (Spatial Microsimulation of Granular Data Interventions).

Spatial microsimulation techniques should be used in future studies to simulate how hypothetical data from the ward and incident level would impact the associations and spatial clustering that were found in this sub-county aggregated study. Specifically, these researchers could create synthetic ward-level datasets using iterative proportional fitting (IPF) or combinatorial optimization algorithms to capture the known sub-county marginal totals and explore the range of plausible micro-distributions. Such a microsimulation would enable analysts to calculate the "information gain" that comes with using finer spatial levels, and determine if the ecological fallacy found in this study (which saw associations of features decrease by more than 30% by summarizing at the county level) is reduced at the ward level. The study would supply empirical data for NPS and KNBS to support them in determining the spatial resolution they should consider when implementing predictive policing.

Qualitative Victimization Surveys for Physical and Mental Health.

To estimate the "dark figure" of unreported crime, and to test the socio-spatial drivers of structural invisibility proposed in this study, conduct qualitative victimisation surveys in the informal settlements identified as having 20–40% data completeness (Mathare, Kibra, Mukuru). The surveys should use a combination of structured questionnaires to the household and in-depth semi-structured interviews to get a sense of both the level of unreported incidents and the reasons for non-reporting (such as distrust of police, inaccessibility of reporting systems, fear of reprisal, informal justice systems etc.). The results will help to establish a baseline of missingness to calibrate the MNAR mechanisms reported in this study,

as well as designing context-specific reporting interventions to tackle causes of under-reporting, not just the symptoms.

The qualitative governance study aimed to identify institutional barriers to NPS data sharing.The objective of the qualitative governance study was to explore institutional barriers to NPS data sharing.

A qualitative governance study with key informant interviews of officials from the National Police Service (NPS), Kenya National Bureau of Statistics (KNBS), Office of the Data Protection Commissioner (ODPC) and National Crime Research Centre (NCRC) is recommended to systematically map institutional, legal, technical, and cultural barriers that inhibit the release of granular, incident-level crime data. This study should be grounded and aim to uncover specific aspects of the data sharing process, risk perceptions, and resource constraints influencing data sharing in Kenyan law enforcement institutions. The results would guide evidence-based policy recommendations on secure data sharing processes (such as data trusts, anonymisation principles, access levels for accredited researchers) that strike the right balance between transparency and research requirements and legitimate concerns for privacy and operational security. Research in the area of governance would complement the technical findings of the present study by focusing on institutional preconditions for improving the data infrastructure.

Comparative ESTA Replication in Secondary Kenyan Cities

Replicate this Exploratory Spatial-Temporal Analysis (ESTA) pipeline in other Kenyan rapidly urbanizing contexts (e.g. Mombasa, Kisumu) to identify if the MNAR data gaps, spatial heterogeneity and representational inequities that were observed in Nairobi are specific to the capital city or representative of national urban data ecosystems. Such a comparative multi-city study would allow researchers to validate the transferability of the methodological framework developed in this thesis and determine if issues in the data landscape (e.g., informal settlement under-reporting, instability across time, aggregation-induced bias) are systemic characteristics of the Kenya data landscape for crime data or contextspecific. A study of this sort would give a more complete picture of readiness

for predictive policing across the nation, and enhance the external validity of the results of this research which focused on Nairobi.

Data Infrastructure Interventions: Longitudinal Tracking.

There is a need to carry out a longitudinal study to assess the impact of targeted community policing programmes (such as digitised Nyumba Kumi reporting systems, mobile police station deployments, community liaison officer programmes) in identified "low-completeness" wards within a 3-5 year timeframe on data completeness and equitable reporting. A quasi-experimental design with intervention wards compared to matched control wards would be used and the data completeness percentages, DPD/EOD fairness metrics and community-reported crime incidents would be tracked over time. The longitudinal evidence would directly assess the effectiveness of the policy recommendations provided in Section 5.4.1 and would provide empirical guidance to policymakers on the greatest return on investment in data infrastructure investments in terms of representational equity and analytical utility. This type of research would help to tie together the implications of the exploratory analysis with implications of work that would be done in the field.

5.4.3 Governance Safeguards for Exploratory Outputs

A critical risk identified during the departmental presentation is the potential for exploratory risk surfaces to be misappropriated as de facto predictive policing tools for operational patrol dispatch. Because the outputs of this framework highlight spatial clustering and historical reporting patterns, there is a danger that law enforcement might use them to justify disproportionate surveillance in already over-policed areas, thereby triggering the runaway feedback loops warned by Lum and Isaac (2016) and Ensign et al. (2018).

To prevent this, the proposed Methodological and Ethical Framework mandates the following institutional safeguards before any exploratory output is shared with operational stakeholders:

Mandatory "Exploratory Use Only" Watermarking: All generated risk surfaces, cluster maps, and completeness overlays must bear a permanent, unremovable digital watermark stating: "EXPLORATORY OUTPUT ONLY. Not validated for

operational deployment or predictive dispatch. Reflects historical reporting patterns, not ground-truth crime prevalence."

Prohibition of Direct Operationalization: Policy guidelines must explicitly prohibit the use of these aggregated, sub-county-level exploratory maps for direct, automated, or commander-driven patrol dispatch. Any translation of spatial insights into operational policing must require a human-in-the-loop review and community oversight.

Pre-Deployment Representational Equity Audits: Before any future granular data system is considered for operational predictive modeling, it must pass an independent, prospective Representational Equity Audit. Specifically, the system must demonstrate a Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD) of less than 0.05 across settlement typologies. If the data fails this threshold (as the current aggregates do, with a DPD of 0.504), operational deployment must be legally barred.

Annual Algorithmic Impact Assessments: The Office of the Data Protection Commissioner (ODPC), in conjunction with the National Police Service (NPS), should conduct annual audits of any spatial crime analytics used by the state to ensure compliance with the Kenya Data Protection Act (2019) and to monitor for algorithmic redlining.

5.5 Contribution to Knowledge

This study offers three unique contributions to the areas of spatial data science, predictive policing governance, and critical studies of AI in the Global South:

Theoretical Contribution: In this study, the scarcity of data is considered not as the methodological problem to be overcome, but as the object of research analysis itself. The data landscape is the primary research question and it contributes to the growing body of research on the ecological inference limits in the face of Small-N in data-constrained African urban settings.This theoretical repositioning brings criminological theory (Routine Activity Theory, Crime Pattern Theory) together with sophisticated spatial data science concepts, and shows how theoretical concepts can be translated into proxy features and spatial diagnostics even in the absence of incident level data. The study offers empirical evidence of the effect of

reporting bias in informal settlements on the fairness of algorithms and joins the emerging literature on the conditions needed for defensible predictive policing in the Global South.

2. Methodological Contribution: The study provides a reproducible and aggregation-aware pipeline for Exploratory Spatial-Temporal Analysis (ESTA) especially designed for the African urban context, where spatial-temporal data is often coarse, highly heterogeneous in terms of reporting, and small-N. This pipeline is unique in the explicit quantification of aggregation-induced uncertainty by employing the Bayesian hierarchical modeling framework and Conditional Autoregressive (CAR) priors, in prevention of optimistic bias caused by spatial autocorrelation, with the use of spatial block cross-validation, and in the maintenance of analytical robustness with data scarcity, using L1 regularization (LASSO). The methodological innovation is the explicit acknowledgement that it doesn't attempt to make false predictions, but rather considers all outputs to be exploratory risk surfaces – with an interval of credible risk – and overlays of completeness. This method serves as a blueprint for other researchers in dataconstrained settings grappling with the problem of analytical desire and statistical defensibility.

3. Policy/Ethical Contribution: This study is the first empirical audit of Nairobi's public crime data ecosystem with a fairness lens, a mathematical proof of how public crime data is systematically underrepresented for marginalized communities. The study implements two theoretical critiques of Benjamin (2019) and Richardson et al. (2019) into an empirical context-specific result: the Demographic Parity Difference (DPD = 0.504) and the Equal Opportunity Difference (EOD). No research in the reviewed literature used both MNAR diagnostics and DPD/EOD fairness metrics to find an empirical measure of representational equity in an African urban crime data ecosystem. The methodological integration presented here is a new empirical study that shows that the "algorithmic redlining" risk predicted by critical data scholars is not theoretical, but exists in Nairobi's current data infrastructure. The DPD of 0.504 is the first mathematical proof of this structural inequity, and for any predictive policing

deployment to be justified, policymakers must also have the empirical evidence that granular, ethically guided data collection efforts are necessary to do so.

REFERENCES

Amnesty International Kenya. (2023). Policing the margins: Data gaps and human rights in Nairobi's informal settlements. Nairobi, Kenya.

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Polity Press.

Brantingham, P. J., & Brantingham, P. L. (1981). Environmental criminology. Sage.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1), 3–73.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American Sociological Review, 44(4), 588–608.

Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd ed.). John Wiley & Sons.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.

Eck, J. E., & Weisburd, D. (Eds.). (2019). Crime places in crime theory. Springer.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. Proceedings of the 1st ACM Conference on Fairness, Accountability and Transparency, 160–171.

Flaxman, S. D. (2014). Bayesian models for spatial crime data [Doctoral dissertation, University of Chicago].

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). Bayesian data analysis (4th ed.). CRC Press.

KHRC. (2022). Urban security and data justice in Kenya: A community perspective. Kenya National Commission on Human Rights.

Kounadi, O., Leitner, M., & Winter, S. (2020). A systematic review of spatial crime forecasting and its application to policing. Computers, Environment and Urban Systems, 78, 101378.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. Physical Review E, 69(6), 066138.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields. Journal of the Royal Statistical Society: Series B, 73(4), 423–498.

Lum, K., & Isaac, W. (2016). To predict and prevent? An analysis of algorithmic policing. Harvard Journal of Law & Technology, 30(1), 1–48.

MacEachren, A. M. (2004). How maps work: Representation, visualization, and design. Guilford Press.

Mburu, J., & Helbich, M. (2021). Spatial heterogeneity and data constraints in urban crime analytics across Sub-Saharan Africa. Urban Studies, 58(12), 2511–2530.

Missingno. (2023). Missing data visualization in Python. https://github.com/ResidentMario/missingno

Mubangizi, J. B. (2020). Spatial crime mapping and data limitations in Kampala, Uganda. African Journal of Criminology and Justice Studies, 13(1), 45–62.

Oduro, F., & Appiah, R. (2022). Open data portals and crime transparency in Ghana and Kenya. Government Information Quarterly, 39(4), 101742.

Olanrewaju, O. A., & Adeyemi, O. O. (2021). Spatial visualization of crime patterns and data provenance challenges in Lagos, Nigeria. Journal of African Urban Studies, 5(2), 112–130.

Openshaw, S. (1984). The modifiable areal unit problem. Concepts and Techniques in Modern Geography, 38, 1–52.

Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. New York University Law Review Online, 94, 15–55.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, 15(3), 351–357.

Selbst, A. D., & Barocas, S. (2018). The hidden politics of facial recognition technology. Proceedings of the 2018 ACM Conference on Fairness, Accountability, and Transparency, 1–10.

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423.

Tierney, N., Cook, D., McBride, N., & Tierney, M. (2020). missingno: A missing data visualization toolkit for Python. Journal of OpenSource Software, 5(46), 1956.

UN-Habitat. (2022). World cities report 2022: Envisaging the future of cities. United Nations Human Settlements Programme.

Vapnik, V. (1995). The nature of statistical learning theory. Springer.

Wakefield, J. (2004). Ecological inference for 2×2 tables. Journal of the Royal Statistical Society: Series A, 167(3), 385–445.

APPENDICES

Appendix A: Data Management Plan (DMP)

1. Data Sources: KNBS Economic Survey (2018–2024), NCRC National Crime Mapping Reports, OpenStreetMap (OSM) & HDX.
2. Storage and Security: All raw and processed datasets are stored on encrypted, university-approved institutional drives. Access is restricted to the primary researcher and supervisor.
3. Retention and Deletion: Data will be retained for a minimum of 5 years postgraduation, after which local copies will be securely deleted.
4. Sharing and Reproducibility: Processed, anonymized datasets and full analytical Python code are shared via a public GitHub repository under the MIT License. No PII is shared.

Appendix B: Completed Data Access and Request Log

Institution	Purpose	Pathway	Date Accessed	Status/ Outcome
KNBS	Download sub-county crime tables	Primary (Public)	14-Mar-2026	Granted. Public download.
NCRC	Access National Crime Mapping Report	Primary (Public)	14-Mar-2026	Granted. Public download.
NPS (NACOSTI)	Request police-station data	Secondary (Contingent)	20-Aug-2026	Attempted. Access to granular data blocked by NPS

				institutional protocols; primary public pathway utilized
ODPC	Guidance on aggregation thresholds	Compliance	10-Mar-2026	Consulted. Confirmed minimal privacy risk.
OpenStreetMap	Extract road networks, land-use	Primary (Public)	15-Mar-2026	Granted. Extracted via osmnx.

Appendix C: Ethics and Regulatory Approvals

1. 1. Open University of Kenya (OUK) Research Ethics Committee (REC) Approval

   Full ethical clearance was granted for the secondary analysis of publicly available, aggregated crime data.

2. National Commission for Science, Technology and Innovation (NACOSTI) Research License Official research clearance was granted to conduct the study titled: "Mapping the Data Landscape for Machine Learning-Based Predictive Policing: A Spatial-Temporal Exploratory Analysis of Publicly Available Crime Data in Nairobi City County, Kenya."

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **763388** Date of Issue: **26/June/202**

RESEARCH LICENSE



This is to Certify that Mr.. KELVIN NKORISA NTOONTO of Open University of Kenya, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: Mapping the Data Landscape for Machine Learning-Based Predictive Policing: A SpatialTemporal Exploratory Analysis of Publicly Available Crime Data in Nairobi City County, Kenya. for the period ending : 26/June/2027.

License No: **NACOSTI/P/26/4191713**

763388

Applicant Identification Number

Ag. Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013 (Rev. 2014)
Legal Notice No. 108: The Science, Technology and Innovation (Research Licensing) Regulations, 2014

The National Commission for Science, Technology and Innovation, hereafter referred to as the Commission, was the established under the Science, Technology and Innovation Act 2013 (Revised 2014) herein after referred to as the Act. The objective of the Commission shall be to regulate and assure quality in the science, technology and innovation sector and advise the Government in matters related thereto.

CONDITIONS OF THE RESEARCH LICENSE

1. The License is granted subject to provisions of the Constitution of Kenya, the Science, Technology and Innovation Act, and other relevant laws, policies and regulations. Accordingly, the licensee shall adhere to such procedures, standards, code of ethics and guidelines as may be prescribed by regulations made under the Act, or prescribed by provisions of International treaties of which Ken is a signatory to.
2. The research and its related activities as well as outcomes shall be beneficial to the country and shall not in any way;
   i. Endanger national security
   ii. Adversely affect the lives of Kenyans
   iii. Be in contravention of Kenya's international obligations including Biological Weapons Convention (BWC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Chemical, Biological, Radiological and Nuclear (CBRN).
   iv. Result in exploitation of intellectual property rights of communities in Kenya
   v. Adversely affect the environment
   vi. Adversely affect the rights of communities
   vii. Endanger public safety and national cohesion
   viii. Plagiarize someone else's work
3. The License is valid for the proposed research, location and specified period.
4. Neither the license nor any rights thereunder are transferable.
5. The Commission reserves the right to cancel the research at any time during the research period if in the opinion of the Commission t research is not implemented in conformity with the provisions of the Act or any other written law.
6. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research.
7. Excavation, filming, movement, and collection of specimens are subject to further necessary clearance from relevant Government Agencies.
8. The License does not give authority to transfer research materials.
9. The Commission may monitor and evaluate the licensed research project for the purpose of assessing and evaluating compliance with the conditions of the License.
10. The Licensee shall submit one hard copy, and upload a soft copy of their final report (thesis) onto a platform designated by the Commission within one year of completion of the research.
11. The Commission reserves the right to modify the conditions of the License including cancellation without prior notice.
12. Research, findings and information regarding research systems shall be stored or disseminated, utilized or applied in such a manner as may be prescribed by the Commission from time to time.
13. The Licensee shall disclose to the Commission, the relevant Institutional Scientific and Ethical Review Committee, and the relevant national agencies any inventions and discoveries that are of National strategic importance.
14. The Commission shall have powers to acquire from any person the right in, or to, any scientific innovation, invention or patent of strategic importance to the country.
15. Relevant Institutional Scientific and Ethical Review Committee shall monitor and evaluate the research periodically, and make a repo of its findings to the Commission for necessary action.

National Commission for Science, Technology and Innovation(NACOSTI),
Off Waiyaki Way, Upper Kabete,
P. O. Box 30623 - 00100 Nairobi, KENYA
Telephone: 020 4007000, 0713788787, 0735404245
E-mail: dg@nacosti.go.ke
Website: www.nacosti.go.ke

Appendix D: TRIPOD-AI Transparency Checklist

Documents adherence to transparent reporting standards: Title/Abstract states exploratory nature; Methods detail spatial block CV and Bayesian CAR priors; Results present MNAR patterns and equity audit outcomes (DPD = 0.504); Discussion critically addresses MAUP and Small-N constraints.

Appendix E: Computational Pipeline and Reproducibility Artifacts

1. Code Repository: https://github.com/lenkorisa-tech/nairobi-crime-esta-main (MIT License)
2. Repository Contents: data/ (anonymized CSV/GeoJSON), notebooks/ (Jupyter EDA workflows), src/ (Modular Python scripts for ESTA & fairness audits), requirements.txt, Dockerfile.
3. Environment Pinning: Global random seed (random_state=42) set for all stochastic operations.

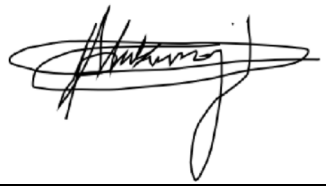
Appendix F: Published Papers (if applicable)

No publications arising from this study at the time of submission.

APPROVAL PAGE

This thesis/project has been examined and approved in partial fulfillment of the requirements for the award of the degree of Master of Data Science of The Open University of Kenya.

Role	**Signature**	**Date**
Candidate: Kelvin Nkorisa Ntoonto	NNK	26/09/2026

Supervisor: Dr. Paul Mutemi		10/10/2026
Co-Supervisor:		
Internal Examiner		
External Examiner		
Head of Department (Mathematics & Statistics)		
School Dean (School of Science and Technology)		
Director, Postgraduate Studies		

Formatting Guidelines: Font: Book Antiqua 12pt | Spacing: 1.5 | Margins: Left 3cm, Others 2.5cm | APA 7th Edition | Page numbering: bottom centre | Similarity index < 20%