CRIME HOTSPOTS DETECTION USING GEOSPATIAL VISUALIZATION AND CLUSTERING TECHNIQUES

An Unsupervised Learning Approach on UK Open Crime Data

Shikha Tyagi

MSc. Business Analytics

Dublin Business School

20052009@mydbs.ie

Abstract

The increasing availability of open crime data and advances in geospatial analytics provide new opportunities for understanding the spatial dynamics of criminal activity. This study investigates the identification and interpretation of crime hotspots in the United Kingdom using unsupervised machine learning and geospatial visualisation techniques. Street-level crime and outcome data published by UK police forces were integrated and analysed using Python-based analytics and Tableau dashboards. Following extensive data cleaning and preprocessing, density-based clustering methods, specifically DBSCAN and HDBSCAN, were applied to detect spatial concentrations of crime. A representative random sample of 50,000 records was used to ensure computational feasibility. The findings show that crime is highly concentrated in specific micro-locations, particularly within major urban centres, and that many hotspots persist over time. Analysis of crime type composition and investigative outcomes reveals substantial variation across hotspots. The study demonstrates that combining unsupervised clustering with interactive visual analytics provides a reproducible and applied framework for crime hotspot analysis and evidence-based decision-making.

Acknowledgements

I would like to express my sincere gratitude to my dissertation supervisor Professor Sajal Kaur Minhas for her valuable guidance, constructive feedback, and continued support throughout the course of this research project. Her academic insight and encouragement were instrumental in shaping the direction, structure, and quality of this dissertation.

I am also grateful to the academic staff of the MSc Business Analytics programme for providing the knowledge and skills that formed the foundation of this research. The methodologies and analytical approaches developed during the programme were critical in enabling the successful completion of this study.

Finally, I would like to acknowledge the UK Police Open Data platform for providing access to publicly available crime and outcome datasets, without which this research would not have been possible. The availability of open data played a crucial role in facilitating transparent, reproducible, and applied analysis.

Table of Contents

Table of Figures

Chapter 1: Introduction

1.1 Background

Crime analysis has undergone significant transformation with the emergence of open data initiatives and the growing use of advanced analytics. In the United Kingdom, the data.police.uk platform provides monthly street-level crime records, including geographical coordinates, crime categories, and where available, outcome information. Such data allow researchers to explore crime patterns at fine spatial resolutions and to move beyond traditional aggregated statistics. The spatial concentration of crime is a well-established phenomenon, often described through the concept of hotspots, where a small number of locations account for a disproportionate share of offences (Eck et al., 2005). Understanding these patterns is critical for effective policing, urban planning, and public safety strategies.

Geospatial technologies and machine learning have become central to modern crime analysis. Geographic Information Systems (GIS) enable the visualization and exploration of spatial distributions, while clustering algorithms can uncover latent structures in large datasets without predefined labels. Unsupervised learning methods such as K-Means and DBSCAN have been widely applied in spatial crime studies to identify clusters of criminal activity (Mohler et al., 2011). The integration of these techniques offers the potential to produce actionable insights from complex and large-scale datasets.

1.2 Problem Statement

Despite the availability of detailed open crime data, practical challenges remain in extracting meaningful insights. Traditional crime analysis often relies on aggregation to administrative units, which can obscure micro-level patterns and is subject to the modifiable areal unit problem. Moreover, open datasets are frequently heterogeneous, containing missing values, inconsistent formats, and partial outcome information. As a result, analysts face difficulties in integrating datasets and in producing robust, interpretable hotspot analyses.

There is therefore a need for analytical approaches that can handle large volumes of geospatial data, identify localized concentrations of crime, incorporate outcome information, and present findings in accessible visual formats. Without such approaches, the potential of open crime data to inform proactive policing and resource allocation remains underutilised.

1.3 Project Overview and Rationale

This applied research project examines street-level crime across the UK for January –September 2025, integrating incident and outcome datasets and applying geospatial visualisation together with unsupervised clustering to detect hotspots. The rationale is twofold: academically, the work advances practice by combining outcome data with spatial clustering on contemporary UK open data; practically, the output is intended to inform policing resource allocation by highlighting high-risk micro-locations. The artefact associated with this Applied Research Project is an analytical pipeline implemented in Python that performs data cleaning, spatial clustering and interactive mapping, thereby demonstrating an end-to-end solution.

1.4 Research Aim, Objectives and Questions

The aim of this study is to detect and interpret crime hotspots in the UK using geospatial visualization and unsupervised clustering techniques.

The objectives are to preprocess and integrate crime and outcome datasets, explore spatial and temporal patterns through exploratory analysis, apply density-based clustering to identify hotspots, analyse crime type and outcome composition within clusters, and visualise findings using interactive dashboards.

The research questions address how crime is spatially distributed across the UK, whether unsupervised clustering can effectively identify hotspots, how crime types and outcomes vary across hotspots, and how visual analytics can support interpretation and decision-making.

1.5 Scope and Methodology Overview

The scope of this research is defined across temporal, geographical, data-related, and methodological dimensions. Temporally, the study focuses on crime data recorded between January and September 2025. Geographically, the analysis encompasses the United Kingdom, as represented in the open datasets provided by data.police.uk. The dataset includes street-level crime records merged with available outcome data to provide additional analytical context. Methodologically, the study is limited to geospatial visualization and unsupervised clustering techniques implemented using Python-based data analytics libraries, including Pandas, GeoPandas, Folium, and Scikit-learn.

1.6 Structure of the Dissertation

The dissertation is organised into six chapters. Chapter One introduces the study; Chapter Two critically reviews the literature on hotspot detection, geospatial methods and clustering; Chapter Three describes the research design and technical implementation; Chapter Four presents analysis and results; Chapter Five discusses findings in relation to theory and practice; Chapter Six concludes and recommends future work.

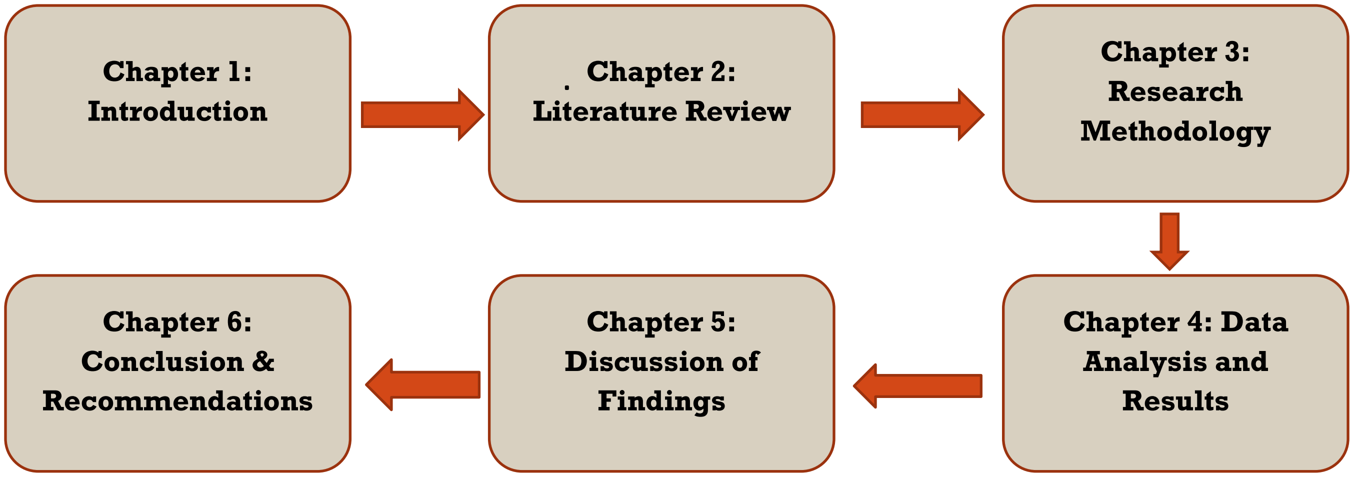

Figure 1: Structure of the Dissertation

Chapter 2: Literature Review

2.1 Introduction

This chapter critically reviews existing literature relevant to crime hotspot detection, geospatial visualisation, and the application of unsupervised learning techniques for spatial crime analysis. The purpose of this review is threefold: first, to establish the theoretical foundations underpinning crime concentration and place-based criminology; second, to examine methodological approaches used in prior studies for hotspot identification, particularly heatmaps and clustering algorithms; and third, to identify gaps in current research that justify the present study. By synthesising research across criminology, geographic information science, and data mining, this chapter positions the dissertation within existing academic discourse and clarifies its contribution to knowledge.

2.2 Crime Concentration and Place-Based Theories

A foundational principle in crime analysis is that criminal activity is not evenly distributed across space but is concentrated in a relatively small number of places. This phenomenon has been consistently demonstrated across different cities and crime types and is often referred to as the "law of crime concentration" (Weisburd, 2015). Early work by Sherman, Gartin and Buerger (1989) showed that a small proportion of addresses in Minneapolis accounted for a disproportionately high volume of police calls, introducing the concept of "hot spots" in criminology.

Subsequent studies reinforced this spatial regularity. Eck et al. (2005) argued that focusing on places rather than offenders provides greater potential for crime prevention, as environmental and situational factors shape opportunities for crime. Routine activity theory (Cohen and Felson, 1979) further explains this concentration by suggesting that crime occurs when motivated offenders, suitable targets, and lack of guardianship converge in space and time. These theoretical perspectives support the analytical focus of this study, which seeks to identify and interpret micro-spatial clusters of crime incidents.

Place-based criminology also highlights the stability of hotspots over time. Weisburd et al. (2004) found that crime hotspots tend to persist for many years, even when overall crime rates fluctuate. This notion of persistence is particularly relevant to the current research, which examines temporal dynamics of hotspots to identify emerging and disappearing risk areas. The literature therefore provides a strong theoretical justification for hotspot etection as a tool for proactive policing and urban safety management.

2.3 Geospatial Visualisation in Crime Analysis

Geospatial visualisation plays a central role in understanding and communicating crime patterns. Geographic Information Systems (GIS) enable analysts to map incidents, identify spatial trends, and integrate crime data with environmental or socio-demographic layers (Longley et al., 2015). Early crime mapping relied heavily on choropleth maps that aggregate crime counts to administrative units such as wards or districts. While useful, such aggregation introduces the modifiable areal unit problem (MAUP), whereby results depend on the choice of spatial boundaries and may obscure local variation (Openshaw, 1984).

To overcome these limitations, point-based mapping and density surfaces became more prevalent. Kernel density estimation (KDE) is widely used to generate smoothed heatmaps that visually represent crime intensity across space (Chainey, Tompson and Uhlig, 2008). Heatmaps provide intuitive insights into spatial patterns and are often used by law enforcement agencies for tactical decision-making. However, KDE is primarily descriptive and depends on bandwidth selection, which can significantly affect the appearance and interpretation of hotspots (Levine, 2013).

Recent studies have integrated interactive visualisation platforms to enhance interpretability. Wang, Brown and Gerber (2020) demonstrated that GIS-based dashboards allow practitioners to explore crime patterns dynamically, improving situational awareness. Similarly, Andrienko et al. (2017) argued that visual analytics combining computational methods with human interpretation is critical for complex spatio-temporal data. This supports the present study's use of Tableau and interactive maps to complement algorithmic clustering.

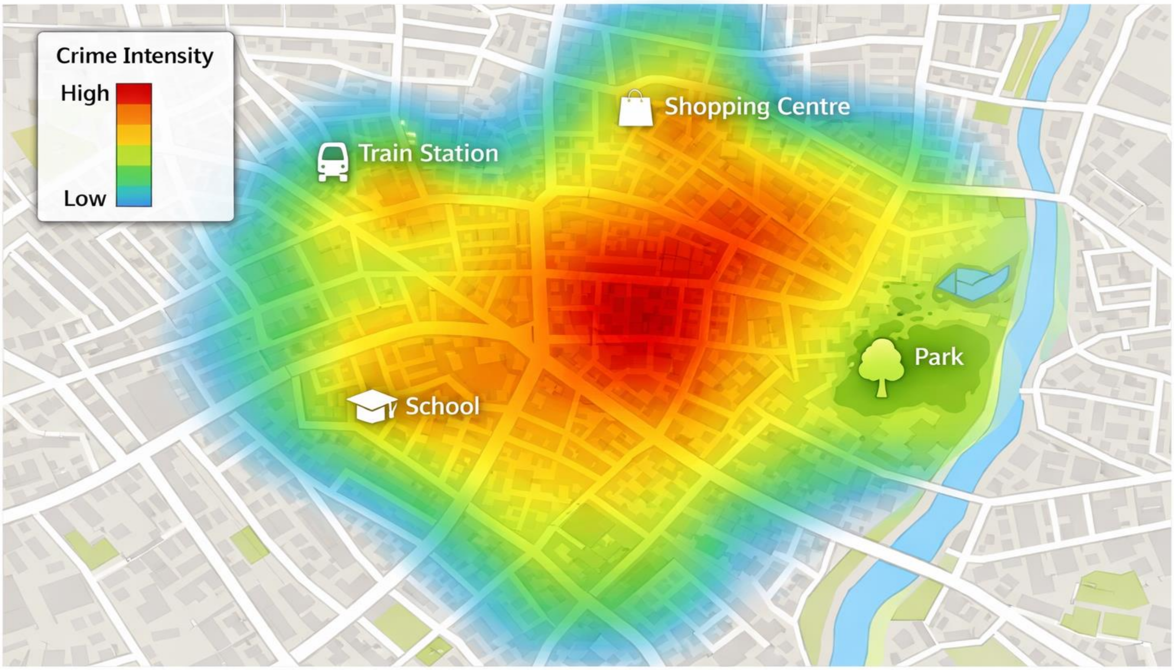

Figure 2 : Example of Crime Heatmap Visualisation from Literature

Despite their advantages, heatmaps alone do not formally define clusters and may exaggerate patterns due to smoothing. This limitation motivates the integration of clustering algorithms that objectively identify dense groupings of incidents, as discussed in the following sections.

2.4 Unsupervised Learning in Crime Data Mining

Unsupervised learning refers to methods that discover structure in data without predefined labels (Hastie, Tibshirani and Friedman, 2009). In crime analysis, unsupervised techniques are particularly valuable because most datasets do not contain labels indicating hotspot membership. Clustering algorithms group similar observations based on distance or density, enabling the detection of latent spatial patterns.

K-means is one of the most widely used clustering methods due to its simplicity and efficiency (MacQueen, 1967). It has been applied in crime studies to partition cities into crime zones (Andresen and Malleson, 2013). However, K-means assumes spherical clusters and requires the number of clusters to be specified in advance, which may not reflect the irregular shapes and unknown number of real-world crime hotspots.

Density-based methods, particularly DBSCAN (Density-Based Spatial Clustering of Applications with Noise), address these limitations. Introduced by Ester et al. (1996), DBSCAN identifies clusters as areas of high point density separated by sparse regions and explicitly labels noise. This is well suited to spatial crime data, which often exhibit irregular hotspot shapes and isolated incidents. Mohler et al. (2011) and Birant and Kut (2007) demonstrated the effectiveness of DBSCAN for detecting spatial crime clusters without strong assumptions about cluster geometry.

HDBSCAN extends DBSCAN by building a hierarchy of density-based clusters and selecting stable ones, reducing sensitivity to parameter choice (Campello, Moulavi and Sander, 2013). Recent studies suggest that HDBSCAN performs better in complex urban datasets with varying density (McInnes, Healy and Astels, 2017). However, its complexity may reduce interpretability for practitioners, raising a trade-off between analytical precision and usability.

The current study builds on this literature by applying DBSCAN and HDBSCAN to UK open crime data and evaluating their suitability for large-scale hotspot detection and visual communication.

2.5 Crime Heatmaps versus Cluster-Based Hotspots

A growing body of literature compares visual density methods such as heatmaps with algorithmic clustering approaches. Heatmaps provide continuous surfaces of intensity but do not assign incidents to discrete hotspot entities. In contrast, clustering methods define explicit groups that can be quantified, compared, and tracked over time.

Chainey et al. (2008) argued that while KDE is valuable for visual exploration, it lacks the analytical rigour required for evaluating hotspot persistence and intervention outcomes. Levine (2013) similarly noted that KDE outputs depend heavily on bandwidth parameters and may create artefactual hotspots. By contrast, DBSCAN produces reproducible clusters based on density thresholds, allowing for more objective identification of high-risk areas.

An example provided by Birant and Kut (2007) demonstrated that DBSCAN identified meaningful spatial crime clusters in urban Turkey that were less apparent in heatmaps. More recently, Wang et al. (2020) showed that combining heatmaps with clustering yields richer insights, where heatmaps guide initial exploration and clustering formalises hotspot boundaries.

Aspect	Heatmap-Based Approaches (e.g. Kernel Density Estimation)	Clustering-Based Approaches (e.g. DBSCAN, HDBSCAN)
Primary purpose	Visual representation of crime intensity across space	Identification of discrete spatial crime hotspots
Output type	Continuous density surface	Explicit clusters and noise points
Definition of hotspots	Implicit, based on colour gradients and density peaks	Explicit, algorithmically defined hotspot boundaries
Sensitivity to parameters	Highly sensitive to bandwidth and kernel selection	Sensitive to density parameters, but more robust in HDBSCAN
Handling of noise	All incidents contribute to density surface	Isolated incidents explicitly labelled as noise
Cluster shape assumptions	Smooth, continuous surfaces	Arbitrary and irregular cluster shapes
Suitability for micro-level analysis	Limited due to smoothing effects	High suitability for micro-spatial hotspots
Interpretability for practitioners	High visual intuitiveness	Requires explanation but enables precise comparison
Quantitative comparison across hotspots	Limited	Strong, enables cluster-level statistics
Temporal hotspot tracking	Descriptive, visually inferred	Explicit tracking of cluster persistence over time
Reproducibility	Dependent on visual interpretation	High, based on algorithmic rules
Typical use in practice	Exploratory analysis and communication	Formal hotspot detection and evaluation

Table 1 : Comparison of Heatmap and Clustering Approaches in Crime Analysis

Despite these advances, few studies systematically integrate clustering outputs with interactive BI tools for decision support, particularly using nationwide UK open data. This gap is directly addressed in the present research.

2.6 Integration of Outcome Data in Hotspot Analysis

Most hotspot studies focus solely on incident locations and frequencies, with limited consideration of case outcomes such as arrests, charges, or unresolved investigations. However, integrating outcome data provides a more holistic perspective on crime and justice system effectiveness.

Ratcliffe (2010) emphasised that hotspot mapping should not only identify where crime occurs but also evaluate whether enforcement efforts in these areas are effective. Groff and La Vigne (2002) similarly argued that outcome-aware analysis can reveal whether hotspots represent persistent enforcement failures or merely high reporting areas.

In the UK context, outcome data are available through data.police.uk, yet few academic studies have fully exploited this resource. Tompson et al. (2015) noted challenges in outcome completeness and consistency but highlighted their potential for evaluating policing performance. The current study contributes to this underexplored area by linking clustering results with outcome categories to assess hotspot effectiveness.

2.7 Use of Open Crime Data and Big Data Challenges

The rise of open government data has transformed crime research. UK police open data provide unprecedented spatial granularity and temporal coverage, enabling large-scale analyses (UK Home Office, 2023). However, such datasets pose challenges related to data quality, missing values, inconsistent formats, and computational scalability.

Kitchin (2014) argued that big data analytics must address issues of veracity and bias, as open datasets often reflect reporting practices rather than true crime incidence. In crime data, under-reporting and variation in recording standards can distort patterns (Baumer and Lauritsen, 2010). Moreover, processing millions of records requires efficient pipelines and sometimes sampling strategies, as adopted in this study.

Methodologically, recent research has explored scalable clustering and grid-based aggregation to manage large spatial datasets (Zhang et al., 2019). The present study's use of sampling and grid-based hotspots reflects these practical considerations while retaining analytical validity.

2.8 Research Gaps and Positioning of the Study

Although extensive literature exists on crime hotspot detection, several gaps remain. First, many studies focus on single cities or small regions, limiting generalisability. There is limited work applying clustering methods to contemporary, nationwide UK crime data. Second, outcome integration remains underdeveloped, despite its importance for evaluating hotspot significance. Third, while advanced clustering algorithms are widely discussed, fewer studies demonstrate their translation into practitioner-friendly visual analytics platforms such as Tableau.

This dissertation addresses these gaps by developing a reproducible pipeline that integrates unsupervised clustering, outcome analysis, and interactive geospatial dashboards using large-scale UK open data. In doing so, it bridges the gap between methodological research and applied decision support.

2.9 Chapter Summary

This chapter has reviewed theoretical foundations of crime concentration, examined geospatial visualisation techniques, evaluated unsupervised clustering methods, and highlighted the importance of outcome integration and open data analytics. The literature supports the use of density-based clustering combined with interactive visualisation for hotspot detection, while also revealing gaps in nationwide, outcome-aware applications. These insights inform the methodological choices outlined in Chapter 3 and justify the analytical framework adopted in this study.

Chapter 3: Research Methodology

3.1 Introduction

This chapter presents the methodological framework adopted to achieve the research aim of detecting and interpreting crime hotspots in the United Kingdom through geospatial visualisation and unsupervised clustering techniques. Given the applied nature of this research, the methodology emphasises the development of a reproducible analytical pipeline that integrates Python-based data analytics with interactive visual dashboards created in Tableau. The chapter outlines the research philosophy, research approach, design, data sources, preprocessing and cleaning procedures, analytical techniques, tools and technologies used, and ethical considerations. Each methodological choice is justified in relation to the research questions and the practical objectives of the study.

In addition to describing analytical methods, this chapter documents the intermediate datasets generated during the analysis, including clustering outputs and summary tables exported from Python for use in Tableau. Explicit documentation of these artefacts enhances transparency and reproducibility and demonstrates how raw open crime data were transformed into structured analytical inputs and visual outputs.

3.2 Research Philosophy and Approach

This research is grounded in a positivist research philosophy, which assumes that social phenomena can be objectively observed, measured and analysed through empirical data (Saunders, Lewis and Thornhill, 2019). Crime incidents recorded in police databases constitute observable events that can be spatially located, categorised and quantified. As such, they are well suited to quantitative and computational analysis.

The study adopts a quantitative research approach, focusing on numerical, spatial and temporal analysis of large-scale datasets. Rather than relying on subjective interpretation or qualitative accounts, the research seeks to identify statistically and spatially meaningful patterns within crime data using machine learning techniques. This approach aligns with contemporary developments in crime analytics, where large administrative datasets and computational methods are increasingly used to support evidence-based policing and public safety decision-making.

A deductive research approach underpins the study. Existing criminological theories of crime concentration and place-based offending (Eck et al., 2005; Ratcliffe, 2010), alongside methodological literature on density-based clustering and spatial data mining (Ester et al., 1996; Campello et al., 2013), informed the selection of analytical techniques. These theories guided the formulation of the analytical workflow, which was subsequently applied to empirical UK crime data to examine whether expected spatial and temporal patterns could be observed.

3.3 Research Design

The research follows an applied and exploratory design. It is applied in nature because it seeks to develop a practical analytical artefact in the form of a crime analysis pipeline and accompanying dashboards that could be used by practitioners for situational awareness and decision support. Rather than testing a narrowly defined hypothesis, the study aims to generate actionable insights from complex, real-world data.

The design is exploratory because unsupervised learning techniques are used to discover latent patterns in crime data without predefined labels or target variables. Density-based clustering methods are particularly suited to this aim, as they allow crime hotspots to emerge organically from the data rather than being imposed through administrative boundaries or arbitrary thresholds.

The study adopts a cross-sectional temporal scope, analysing street-level crime incidents recorded between January and September 2025. While the dataset represents a snapshot of crime patterns within this period, temporal variation is explicitly examined to assess hotspot persistence and change.

3.4 Data Sources

The primary data source for this research is the UK Police Open Data platform (data.police.uk), which provides publicly available street-level crime records and outcome data for police forces across England and Wales. Two core datasets were utilised:

1. **Street-level crime data**, containing individual crime incidents with spatial coordinates, offence categories, dates and geographic identifiers.

2. **Outcome data**, providing the most recent investigative outcome associated with a subset of recorded crimes.

Monthly CSV files were downloaded and merged to create a comprehensive dataset. The initial merged dataset comprised approximately 4.29 million records with thirteen attributes.

Column	Dtype
crime_id	object
month	object
reported by	object
falls within	object
Longitude	float64
Latitude	float64
Location	object
lsoa code	object
lsoa name	object
crime type	object
last outcome category	object
Context	float64
outcome_type	object

Table 2: Description of Dataset Attributes

3.5 Data Preprocessing and Cleaning

Given the scale and heterogeneity of the raw dataset, extensive preprocessing was required prior to analysis. Data cleaning and preparation were implemented in Python using Pandas and NumPy and followed a structured pipeline.

Initial inspection identified missing values, inconsistent data types and redundant variables. The 'context' column was removed as it contained only null values. Records with missing crime identifiers were dropped to maintain referential integrity during dataset merging. Observations lacking latitude or longitude were removed, as accurate geospatial analysis requires complete coordinate information.

The 'month' variable was converted to datetime format, and invalid date entries were excluded. Duplicate records were removed to prevent double counting of incidents. Missing values in outcome-related fields were replaced with descriptive placeholders such as "No Outcome Recorded" or "Outcome Pending," allowing these cases to be included in categorical analysis without introducing bias through deletion.

Spatial validation was applied to remove implausible coordinate values, retaining only records within realistic UK geographic bounds. Following these procedures, the cleaned dataset comprised approximately 3.67 million records with twelve attributes.

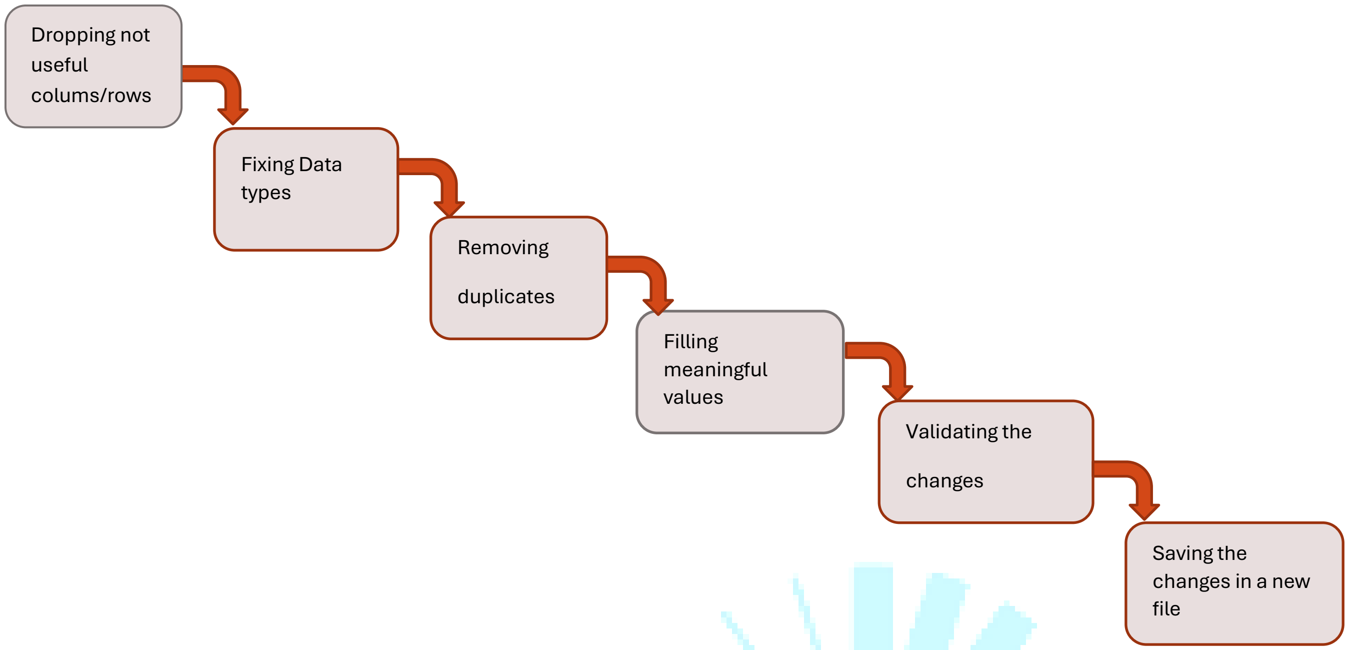

Figure 3: Data Cleaning and Preparation

Due to computational constraints associated with density-based clustering on very large datasets, a random sample of 50,000 records was extracted for intensive modelling.

The comparison of crime type, monthly distribution, and outcome proportions between the full dataset (3.67 million records) and the 50,000-record sample shows minimal differences, all generally under 0.5%. Chi-square goodness-of-fit tests returned high p-values (0.39 and 0.44), indicating no statistically significant variation between the sample and population distributions. These findings confirm that the sampling strategy preserved the underlying structure of the dataset and did not introduce bias, ensuring the sample is fully representative for hotspot analysis. Full validation results and visual comparisons are presented in Appendix B.

This sample was designed to preserve representativeness while enabling feasible processing within a cloud-based environment. Clustering outputs were saved as **Clustering_results_50K**, while aggregated metrics such as cluster size, dominant crime types and temporal trends were stored in a **cluster_summary** dataset. These artefacts support reproducibility and downstream visual analysis.

3.6 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to understand the structure and characteristics of the dataset prior to clustering. Descriptive statistics were generated for spatial and temporal variables, and visualisations were produced to examine crime type distributions, monthly trends and outcome categories.

Spatial scatter plots of latitude and longitude revealed dense concentrations of crime incidents in major urban areas, while rural regions exhibited sparse distributions. Boxplots were used to identify potential outliers, and temporal line charts highlighted seasonal fluctuations in crime volume. Bar charts of crime types showed that offences such as violence and anti-social behaviour accounted for a substantial proportion of recorded incidents.

EDA findings informed modelling decisions, particularly the selection of density-based clustering techniques suited to irregular spatial distributions.

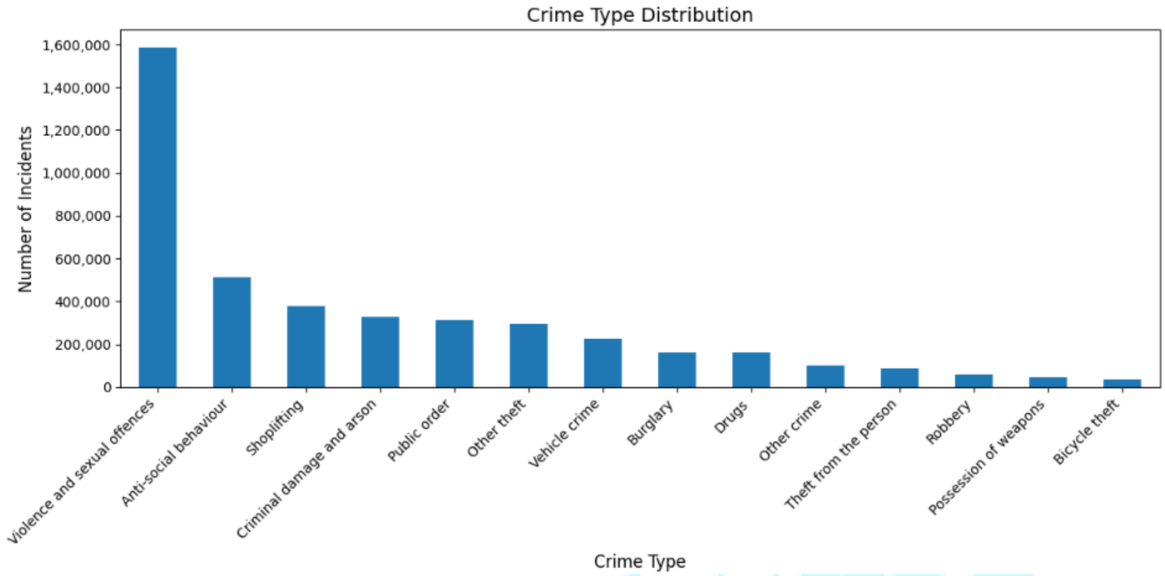

Figure 4 : Crime Type Distribution Chart

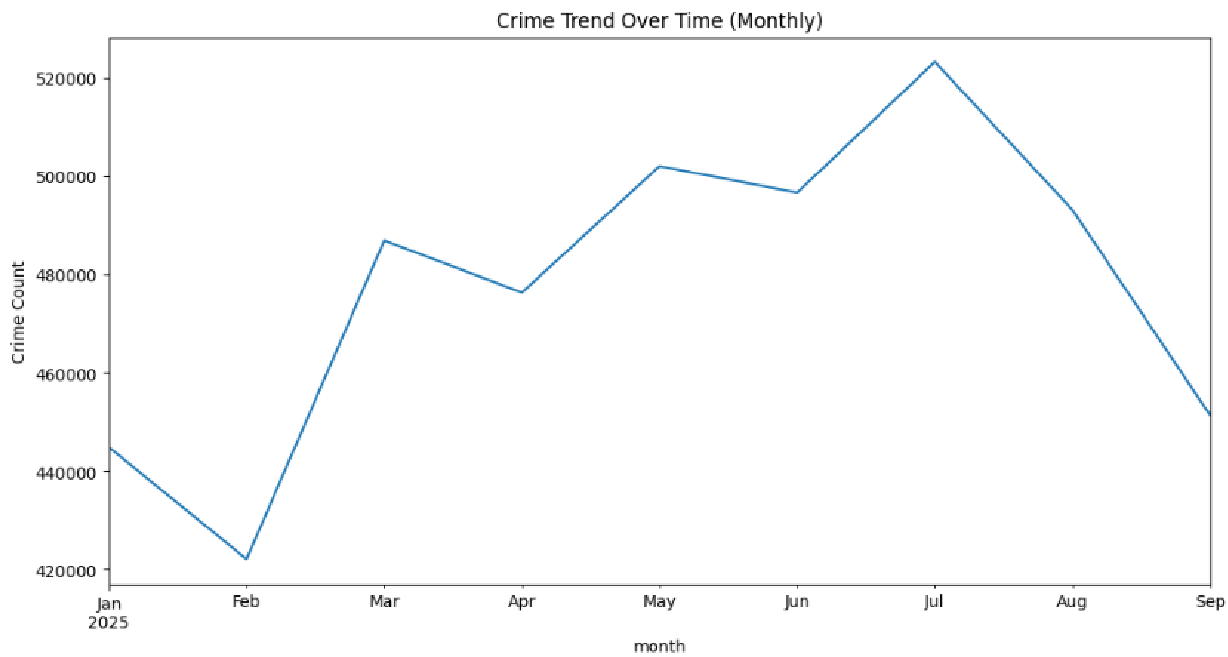

Figure 5: Monthly Crime Trend Line Chart

```
Latitude missing: 67629
Longitude missing: 67629
```

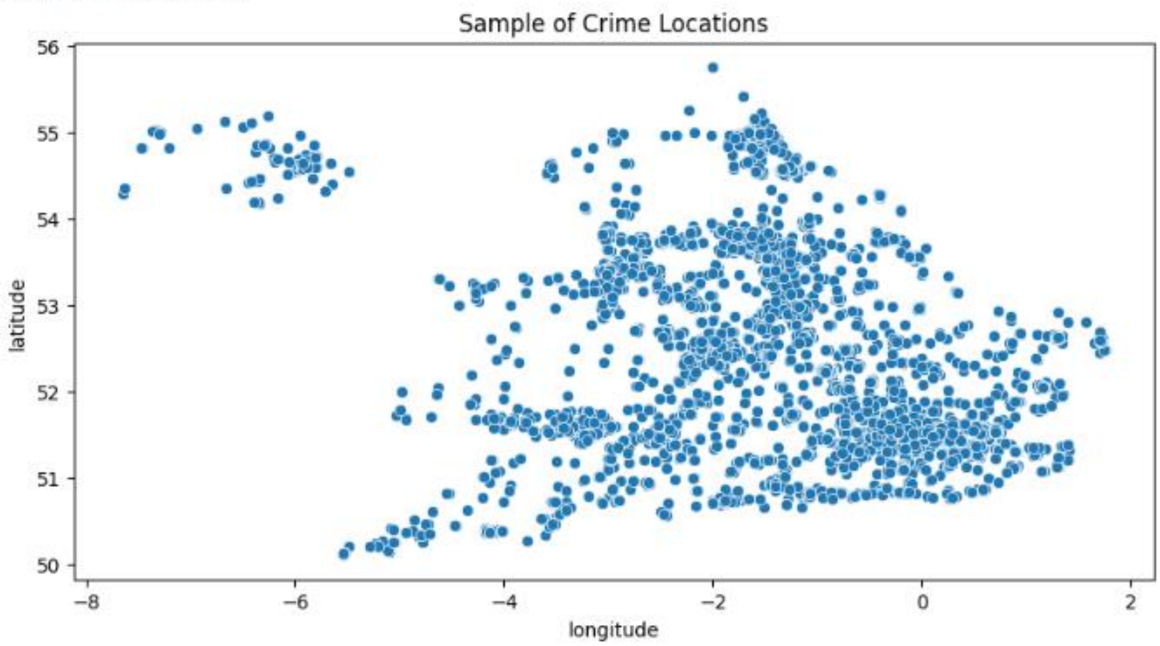

Figure 6: Sample Geospatial Scatter Plot

3.7 Clustering and Hotspot Detection Methods

Unsupervised clustering formed the core analytical method for hotspot detection. Two density-based algorithms were applied: DBSCAN and HDBSCAN.

DBSCAN was initially applied to standardised latitude and longitude coordinates to establish baseline spatial concentration patterns. While DBSCAN effectively identified dense areas and labelled noise points, it produced a limited number of large clusters, reducing interpretability for subsequent analysis.

HDBSCAN was subsequently applied to address these limitations. By allowing clusters of varying density to emerge hierarchically, HDBSCAN produced a more granular set of hotspots. The resulting cluster labels were appended to the sampled dataset and exported as **HDBSCAN hotspot point data for Tableau**, forming the spatial foundation for all dashboards.

3.8 Temporal and Outcome Analysis

Clustered data were aggregated by month and cluster identifier to analyse hotspot persistence and change. Based on temporal trends, clusters were classified as stable, emerging or declining. Outcome data were then integrated to assess investigative effectiveness within hotspots, extending traditional hotspot analysis beyond incident counts alone.

3.9 Tools and Technologies

Python was used for preprocessing, EDA and clustering (Pandas, NumPy, Scikit-learn, HDBSCAN, Matplotlib). Tableau was used to create three interactive dashboards: **UK Crime Hotspot Overview**, **High-Risk Area Profiling**, and **Hotspot Dynamics: Temporal Persistence and Change** using curated datasets exported from Python .

3.10 Ethical Considerations

The study used anonymised, publicly available secondary data. No personal identifiers were present. Ethical risks were minimal; however, care was taken to avoid stigmatising communities by interpreting hotspots as indicators of areas requiring support rather than blame.

3.11 Chapter Summary

This chapter has outlined the methodological framework underpinning the study. The next chapter presents the empirical analysis and results.

Chapter 4: Data Analysis and Results

4.1 Introduction

This chapter presents the empirical analysis and results derived from applying geospatial visualisation and unsupervised clustering techniques to UK street-level crime data. Building on the methodological framework outlined in Chapter 3, the analysis integrates Python-based data processing and clustering with Tableau-based interactive visual analytics. The primary objective of this chapter is to demonstrate how large-scale crime data can be transformed into interpretable spatial intelligence that supports identification of high-risk areas, examination of hotspot dynamics, and evaluation of crime composition and outcomes.

To support transparency and reproducibility, multiple analytical artefacts generated in Python were used in this chapter. These include the Clustering_results_50K dataset containing crime-level cluster assignments, a cluster_summary table providing aggregated cluster metrics, and two Tableau-ready datasets: **HDBSCAN hotspot points** and **Top High-Risk Areas**. These datasets form the empirical basis for both static analysis and interactive dashboards.

4.2 Overview of the Analysed Dataset

The analysis draws on a cleaned and integrated dataset comprising approximately 3.67 million valid crime records collected between January and September 2025. Each record includes spatial coordinates, crime type, reporting force, month of occurrence, LSOA identifiers, and outcome information where available. Due to computational constraints associated with density-based clustering, a random sample of 50,000 records was extracted for clustering and advanced spatial modelling.

Summary statistics for key spatial variables indicate wide geographic coverage across England and Wales, with clear central tendencies reflecting urban concentrations.

index	month	longitude	latitude
count	3671182	3671182	3671182
mean	23:38.1	-1.324839	52.37576
min	01/01/2025 00:00	-7.989991	49.89213
25%	01/03/2025 00:00	-2.076538	51.49262
50%	01/05/2025 00:00	-1.264543	52.05181
75%	01/07/2025 00:00	-0.17489	53.36034
max	01/09/2025 00:00	1.760357	55.7909
std	NaN	1.427881	1.16112

Table 3 : Summary Statistics of the Cleaned Dataset

```
Latitude missing: 0
Longitude missing: 0
```

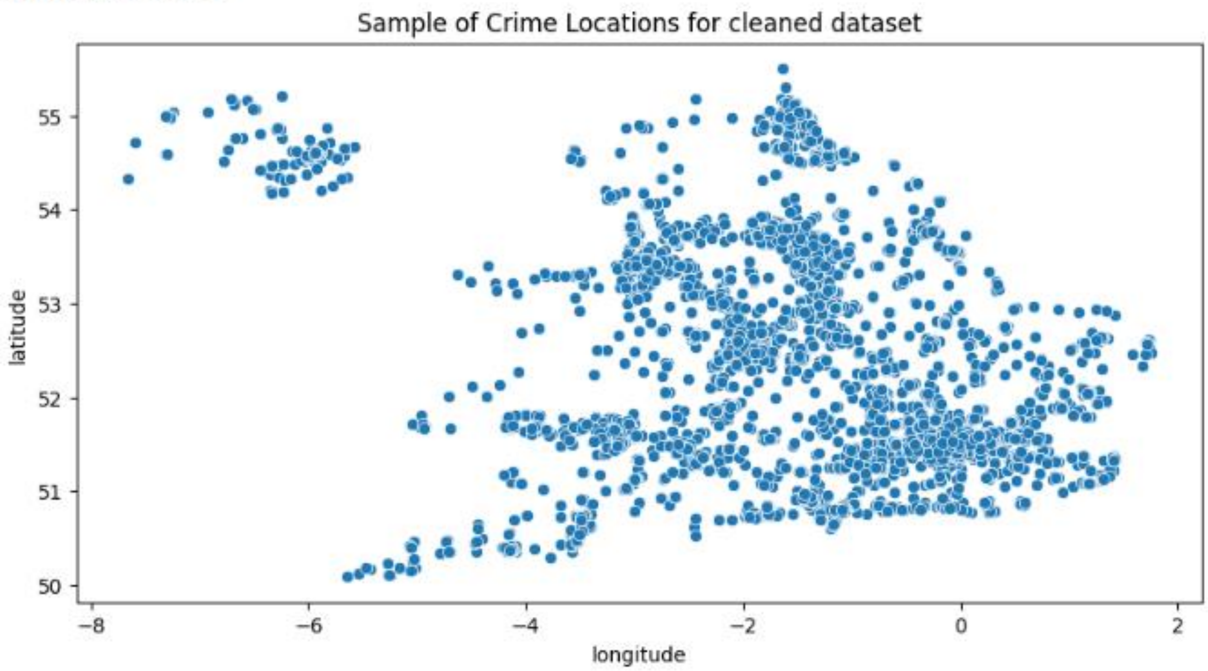

Figure 7: Geographic Distribution of Sampled Crime Incidents

4.3 Exploratory Spatial and Temporal Analysis

Initial exploratory analysis confirmed that crime incidents are unevenly distributed across space. Dense clusters were observed in metropolitan regions, particularly London and other large cities, while rural areas exhibited sparse distributions. This spatial heterogeneity supports the use of density-based clustering methods rather than aggregation by administrative boundaries.

Temporal analysis revealed month-to-month variation, with higher volumes during late spring and summer. While national trends provide useful context, they do not capture whether specific locations experience persistent or transient concentrations of crime, reinforcing the need for hotspot-level temporal analysis.

Figure 8: Monthly Crime Trend (January–September 2025)

4.4 Spatial Hotspot Detection Using Unsupervised Clustering

4.4.1 DBSCAN-Based Hotspot Identification

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was first applied to the sampled dataset using standardised latitude and longitude coordinates. DBSCAN was selected due to its ability to identify clusters of arbitrary shape and to label isolated points as noise, which is particularly appropriate for spatial crime data that do not conform to spherical cluster assumptions.

The DBSCAN results revealed a small number of dominant clusters alongside noise points. One large cluster encompassed the majority of urban incidents, while one or two smaller clusters captured additional dense pockets. Although DBSCAN successfully demonstrated the non-random spatial concentration of crime, the limited number of clusters reduced interpretability for downstream analysis, particularly for temporal and crime-type comparisons. As a result, DBSCAN outputs were primarily used as a baseline reference rather than the main analytical structure.

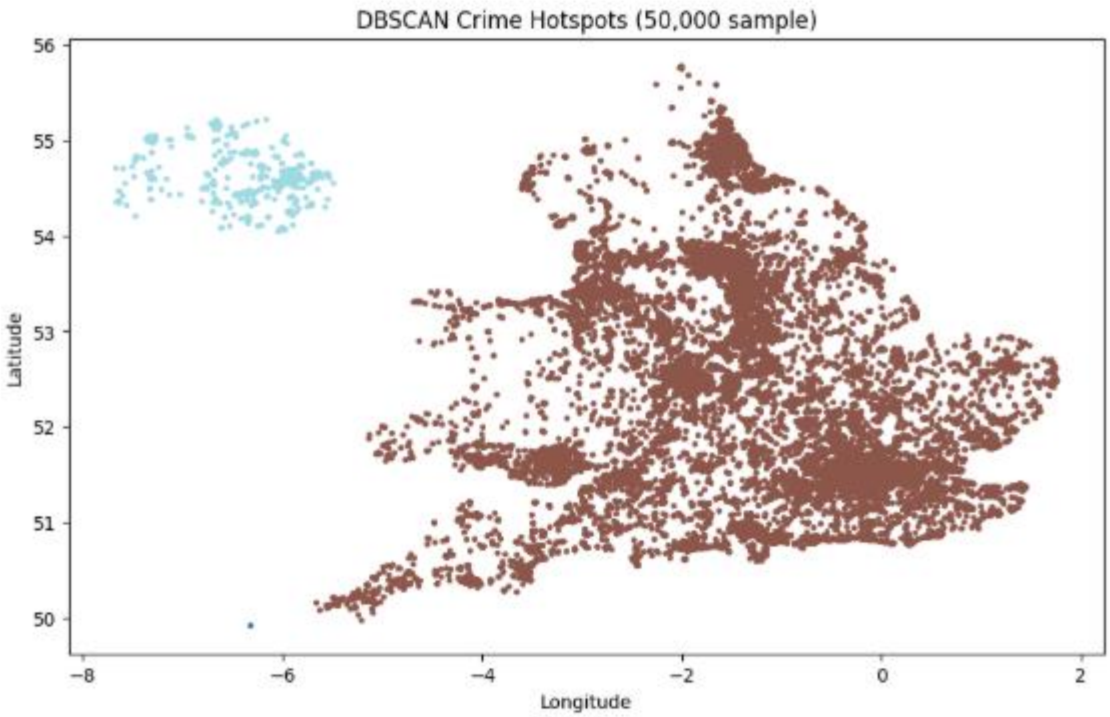

Figure 9: DBSCAN Crime Hotspot Map

4.4.2 HDBSCAN-Based Fine-Grained Hotspot Detection

To achieve more granular hotspot identification, HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) was applied to the same standardised geographic coordinates. Unlike DBSCAN, HDBSCAN does not require a fixed density threshold and instead identifies clusters across varying density levels, making it particularly suitable for nationwide datasets with heterogeneous urban and suburban structures.

The HDBSCAN analysis identified 141 distinct spatial clusters after excluding noise points. These clusters varied in size and geographic extent, representing micro-level crime hotspots across different regions of the UK. Large clusters corresponded to major metropolitan centres, while smaller clusters captured localised hotspots in towns and suburban areas.

The higher resolution provided by HDBSCAN enabled meaningful cluster-level analysis, including temporal trends, crime type composition, and outcome effectiveness. Consequently, HDBSCAN clusters were selected as the primary clustering framework for subsequent analyses and Tableau visualisations.

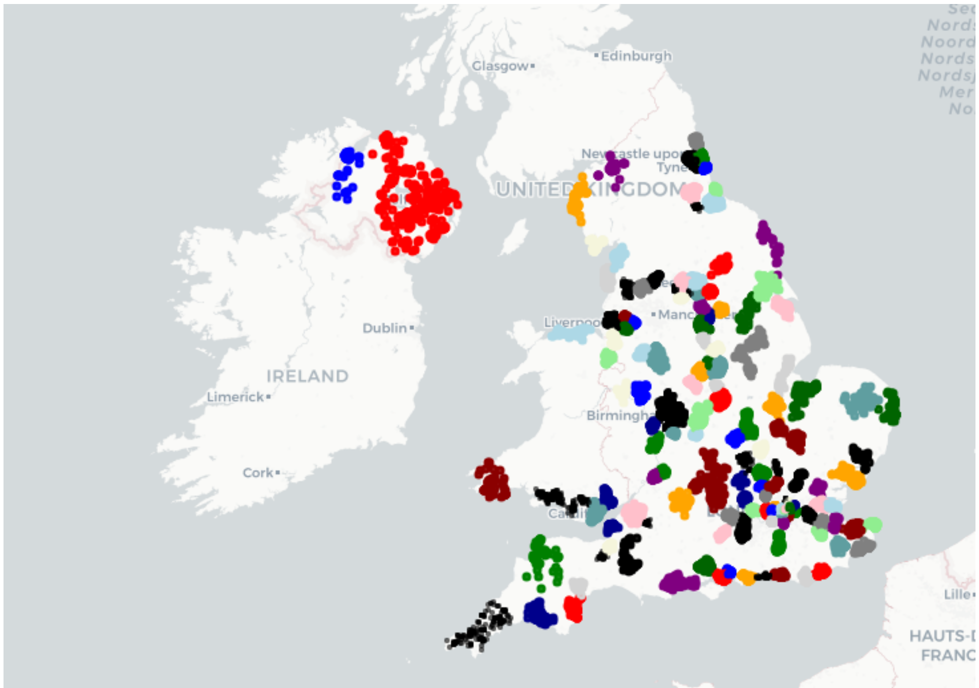

Figure 10: HDBSCAN Crime Hotspot Clusters

4.5 UK Crime Hotspot Overview Dashboard

The **UK Crime Hotspot Overview** dashboard provides a national-level spatial view of crime concentration using the HDBSCAN hotspot points dataset. Each crime incident is plotted on a map and coloured by cluster membership, allowing users to visually distinguish hotspot boundaries and relative density.

Interactive filters for month, crime type and outcome category enable users to explore how hotspots vary across time and offence categories. The dashboard highlights the dominance of urban hotspots while allowing drill-down into localised clusters. This overview serves as the entry point for spatial exploration and contextualises the more detailed dashboards discussed later.

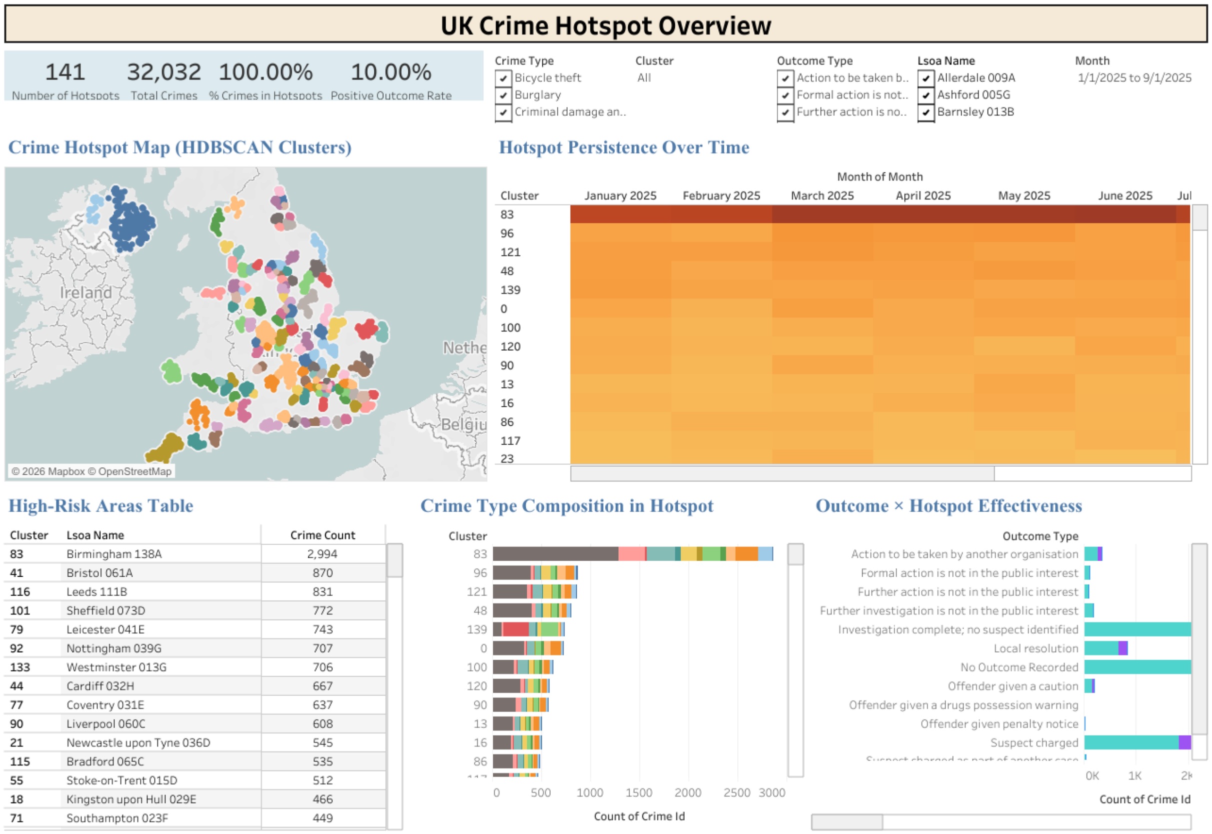

Figure 11: UK Crime Hotspot Overview Dashboard

4.6 Temporal Hotspot Persistence and Dynamics

Temporal analysis focused on understanding how hotspots evolve over time. Crime counts were aggregated by cluster and month, and trends were calculated to classify clusters as stable, emerging or declining. This classification was implemented in Python and appended to the clustering output before export to Tableau.

The **Hotspot Dynamics: Temporal Persistence and Change** dashboard visualises this analysis using a temporal heatmap, where clusters are displayed across months and colour intensity represents crime volume. Persistent hotspots appear as consistently dark bands, while emerging and declining hotspots exhibit visible changes over time.

This dashboard enables rapid identification of clusters requiring proactive attention and supports time-sensitive decision-making.

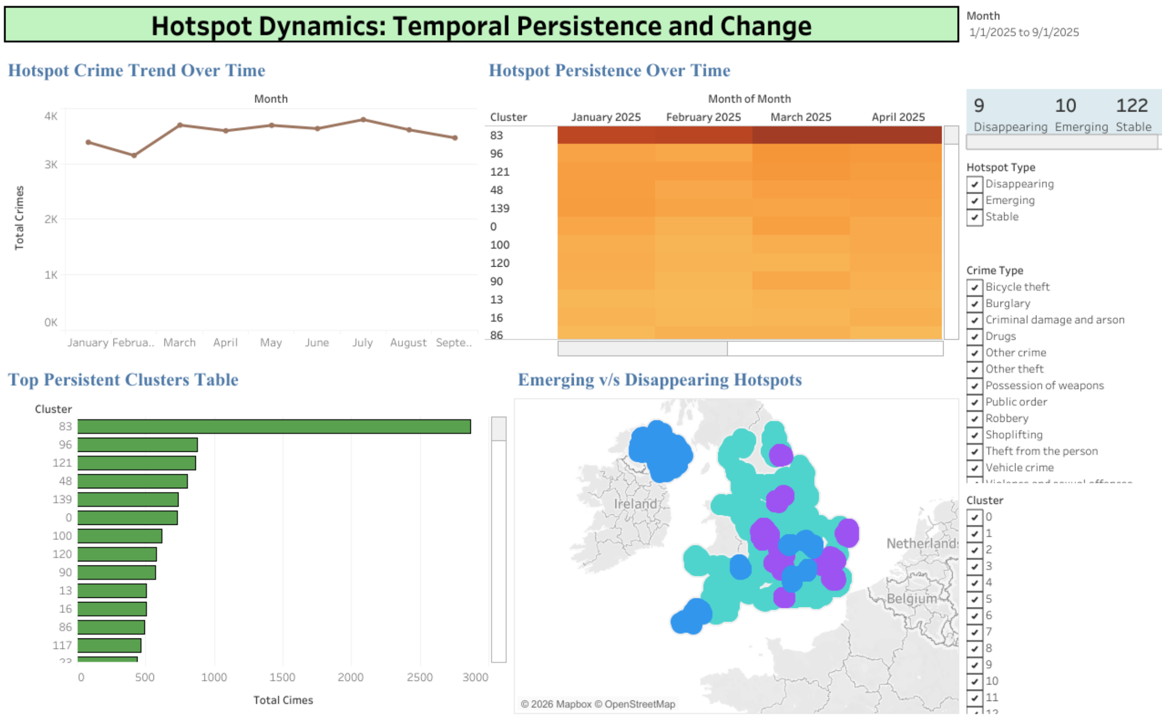

Figure 12: Hotspot Dynamics – Temporal Persistence and Change Dashboard

4.7 High-Risk Area Profiling Using Crime Type and Outcome Characteristics

To further examine the characteristics of identified crime hotspots, a high-risk area profiling analysis was conducted using crime type, outcome, and geographic attributes. This analysis was operationalised through the High-Risk Area Profiling Tableau dashboard, which integrates HDBSCAN cluster identifiers with offence categories and outcome variables to enable comparative assessment across hotspots. Clusters were profiled based on dominant offence types, including violence and sexual offences, anti-social behaviour, vehicle crime, and theft-related incidents, revealing substantial heterogeneity in crime composition across high-risk areas.

The dashboard also incorporates outcome information, allowing evaluation of hotspot effectiveness. Visual comparison highlights variation in outcome distributions, with several high-crime clusters exhibiting higher proportions of unresolved cases, suggesting increased investigative complexity within these areas. In addition, a High-Risk Areas table ranks clusters by crime volume, while associated LSOA names enable interpretation at a recognised neighbourhood level. This linkage between analytical clusters and administrative geography supports targeted, place-based intervention and neighbourhood-level planning.

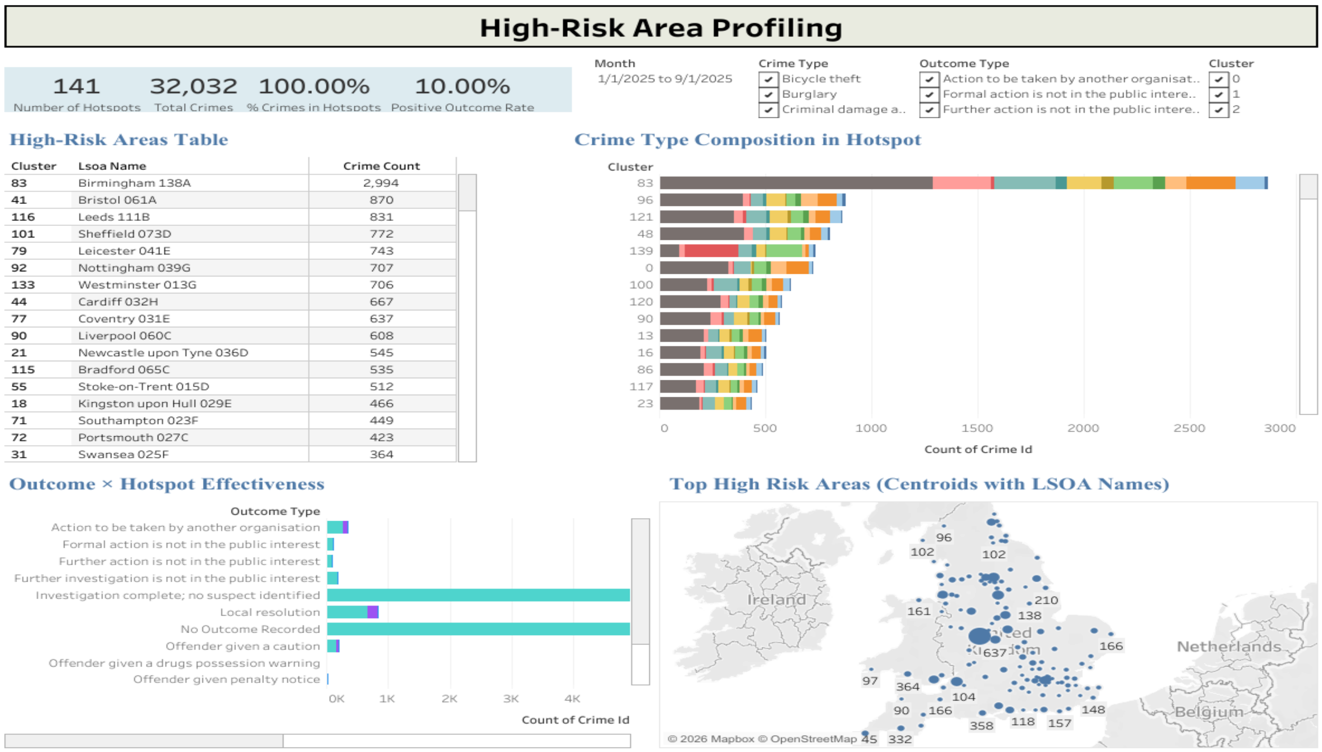

Figure 13: High-Risk Area Profiling Dashboard Showing Crime Type and Outcome Distribution Across Hotspots

4.8 Crime Type Composition within Hotspots

Understanding the nature of crime within hotspots is critical for targeted intervention. To this end, crime type distributions were analysed at the cluster level. Python-based aggregation calculated the proportion of each crime type within each HDBSCAN cluster, and the results were visualised in Tableau using stacked bar charts.

The analysis revealed that hotspots are not homogeneous. Some clusters were dominated by violence and sexual offences, while others exhibited higher proportions of anti-social behaviour, theft, or drug-related crimes. This heterogeneity suggests that different hotspots require different policing and prevention strategies, reinforcing the importance of offence-specific analysis.

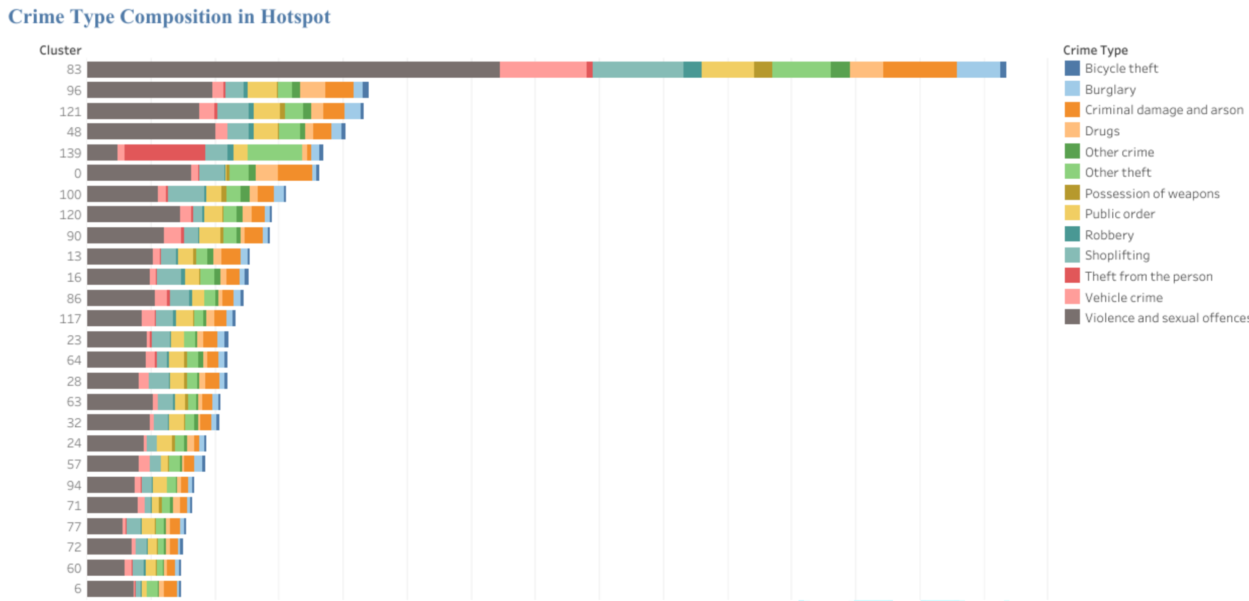

Figure 14: Crime Type Composition per Hotspot

4.9 Outcome Analysis within Hotspots

Outcome data were integrated to examine whether crimes occurring in hotspots are associated with different resolution patterns compared to non-hotspot areas. Outcome categories were grouped into meaningful classifications, such as positive outcomes (e.g. charges, summons) and non-positive outcomes (e.g. no further action).

The analysis showed that many high-density hotspots exhibit a high proportion of cases with no outcome recorded or no further action. This pattern suggests potential challenges in enforcement effectiveness within high-crime areas, possibly due to volume pressure, complexity of offences, or offender mobility.

Tableau visualisations comparing outcome distributions across clusters allowed these patterns to be explored interactively, providing insights into the relationship between crime concentration and justice outcomes.

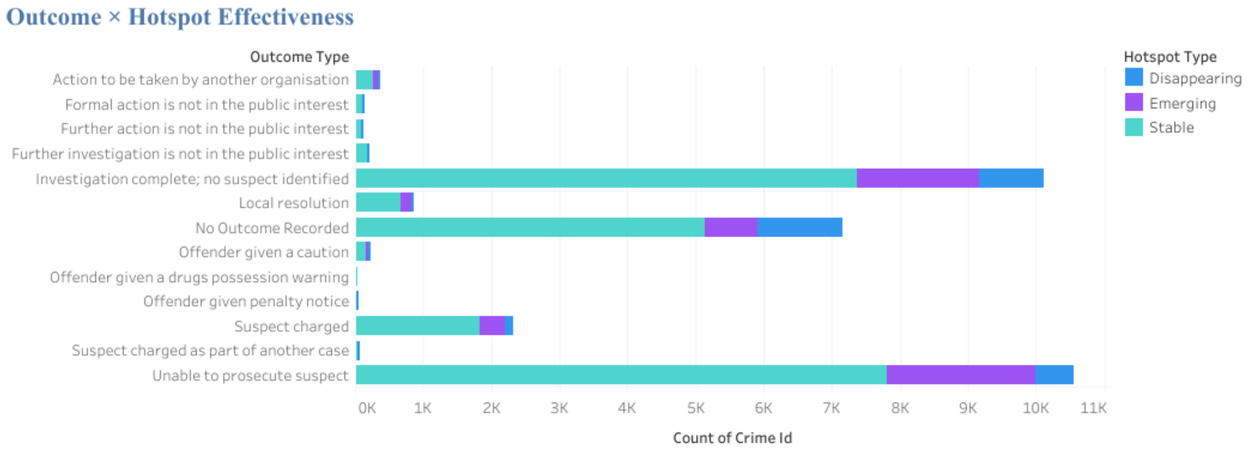

Figure 15: Outcome Distribution by Hotspot

4.10 Identification of High-Risk Areas

To support actionable insights, cluster centroids were calculated in Python and joined with LSOA names to identify high-risk areas at a neighbourhood level. These centroids represent the geographic centre of each hotspot and provide a simplified representation for reporting and decision-making.

In Tableau, a **High-Risk Areas Table** was constructed, listing clusters by crime volume alongside associated LSOA names. This table enables ranking of hotspots and supports strategic prioritisation by policymakers and law enforcement agencies.

High-Risk Areas Table

Cluster	Lsoa Name	Crime Count
83	Birmingham 138A	2,994
41	Bristol 061A	870
116	Leeds 111B	831
101	Sheffield 073D	772
79	Leicester 041E	743
92	Nottingham 039G	707
133	Westminster 013G	706
44	Cardiff 032H	667
77	Coventry 031E	637
90	Liverpool 060C	608
21	Newcastle upon Tyne 036D	545
115	Bradford 065C	535
55	Stoke-on-Trent 015D	512
18	Kingston upon Hull 029E	466
71	Southampton 023F	449
72	Portsmouth 027C	423
31	Swansea 025F	364
12	Bournemouth, Christchurch and Po..	358
4	Cornwall 021C	345
121	Hounslow 018F	337
35	Oxford 008A	337
5	Plymouth 027C	332
87	Derby 013G	329
80	Brighton and Hove 027A	313
60	Blackpool 006A	313
88	Luton 023C	302

Table 4: Top High-Risk LSOAs by Crime Volume

4.11 Summary of Key Findings

The analysis demonstrates that crime in the UK is highly concentrated in specific geographic locations and that these concentrations exhibit varying temporal dynamics and offence compositions. HDBSCAN proved effective in identifying meaningful micro-level hotspots, while Tableau enhanced interpretability through interactive geospatial and temporal visualisations. The integration of outcome data added a critical evaluative dimension, revealing disparities in enforcement effectiveness across hotspots.

4.12 Chapter Summary

This chapter has presented the empirical results of the study, demonstrating how Python-based clustering and Tableau-based visual analytics can be combined to identify, interpret, and communicate crime hotspots. The findings provide a foundation for the discussion in the next chapter, where results are critically evaluated in relation to the research questions, literature, and practical implications.

Chapter 5: Discussion

5.1 Introduction

This chapter provides a critical interpretation of the empirical findings presented in Chapter 4, situating them within the broader context of crime hotspot research, geospatial analytics, and applied data-driven decision-making. While the previous chapter focused on presenting analytical outputs derived from Python-based clustering and Tableau dashboards, this chapter moves beyond description to evaluate the meaning, implications, and limitations of those findings in relation to the research aim and questions.

In keeping with the applied nature of this study, the discussion considers not only theoretical relevance but also practical implications for crime analysis and policing strategy. Particular emphasis is placed on understanding spatial concentration, evaluating the suitability of unsupervised clustering techniques, interpreting temporal hotspot dynamics, assessing crime type heterogeneity, and examining investigative outcomes within hotspots. The role of interactive Tableau dashboards is also discussed as a mechanism for translating complex analytical results into actionable insights.

5.2 Interpretation of Spatial Crime Concentration

The findings of this study provide strong empirical support for place-based theories of crime, which argue that crime is disproportionately concentrated in a small number of locations rather than evenly distributed across geographic space. The identification of dense clusters in major urban centres aligns closely with the crime concentration hypothesis articulated by Eck et al. (2005) and reinforces the notion that micro-places play a critical role in shaping crime patterns.

The use of density-based clustering allowed hotspots to be identified at a fine spatial resolution, independent of administrative boundaries. This is particularly significant in the UK context, where LSOAs vary widely in size and demographic composition. By allowing hotspots to emerge organically from the data, the analysis avoids potential distortions introduced by boundary-based aggregation and provides a more operationally meaningful representation of crime risk.

The UK Crime Hotspot Overview dashboard strengthens this interpretation by visually demonstrating the stark contrast between high-density urban clusters and sparsely populated rural areas. The ability to interactively filter by time period and crime type further reinforces the conclusion that spatial concentration is not only persistent but also offence-specific.

5.3 Methodological Insights from Density-Based Clustering

The comparative application of DBSCAN and HDBSCAN provides important methodological insights. While DBSCAN confirmed the presence of spatial concentration, its sensitivity to parameter selection and tendency to merge distinct urban areas into a single cluster limited its analytical usefulness for national-scale analysis. This outcome highlights a key challenge in applying traditional clustering methods to heterogeneous spatial datasets.

HDBSCAN addressed these limitations by identifying clusters across varying density levels without requiring a predefined number of clusters. The resulting cluster structure enabled a more nuanced understanding of spatial crime patterns, capturing both large metropolitan hotspots and smaller, localised concentrations. This finding supports existing methodological literature advocating hierarchical density-based clustering for complex real-world spatial data (Campello et al., 2013).

The generation of structured outputs such as Clustering_results_50K and cluster_summary further enhanced analytical transparency. These datasets enabled systematic comparison across hotspots and facilitated seamless integration with Tableau, demonstrating the value of combining algorithmic analysis with curated analytical artefacts.

5.4 Temporal Hotspot Dynamics and Crime Stability

The temporal analysis revealed that many hotspots exhibit a high degree of persistence over time, suggesting that underlying environmental and situational factors remain relatively stable throughout the study period. Persistent hotspots were frequently located in areas associated with high footfall, commercial activity, or transport infrastructure, consistent with routine activity theory (Cohen and Felson, 1979).

At the same time, the identification of emerging and declining hotspots highlights the dynamic nature of crime patterns. Emerging hotspots may reflect changes in land use, population movement, or displacement effects following targeted interventions elsewhere. The Hotspot Dynamics: Temporal Persistence and Change dashboard was particularly effective in visualising these patterns, enabling rapid identification of clusters exhibiting significant temporal change.

From a practical perspective, this temporal dimension is crucial. Static hotspot maps may obscure important shifts in crime patterns, whereas temporal analysis supports proactive intervention and resource reallocation. This finding underscores the importance of integrating time explicitly into hotspot analysis rather than treating crime concentration as a purely spatial phenomenon.

5.5 Crime Type Heterogeneity and Targeted Intervention

The analysis of crime type composition revealed that hotspots are not homogeneous entities. Substantial variation was observed across clusters, with different hotspots dominated by different offence categories. This heterogeneity challenges simplistic interpretations of hotspots as uniform "high-crime" areas and highlights the need for offence-specific analysis.

The High-Risk Area Profiling dashboard enabled detailed exploration of these patterns, allowing users to compare crime profiles across clusters and time periods. From an applied standpoint, this insight is particularly valuable, as it suggests that effective intervention strategies must be tailored to the dominant crime types present within each hotspot. For example, hotspots dominated by anti-social behaviour may require community engagement and situational prevention measures, whereas those characterised by violent crime may necessitate focused enforcement and surveillance.

Insights from the High-Risk Area Profiling dashboard further reinforce the argument that crime hotspots differ significantly in their underlying risk characteristics. By visualising offence composition alongside investigative outcomes, the dashboard demonstrates that some hotspots are dominated by specific crime types, while others display a more diverse offence mix. This heterogeneity has important implications for intervention design, as it suggests that hotspot-focused strategies should be tailored not only to spatial risk but also to the dominant offence and outcome patterns observed within each cluster.

5.6 Investigative Outcomes and Systemic Challenges

The integration of outcome data represents a significant extension of traditional hotspot analysis. The findings indicate that crimes occurring within hotspots often have lower rates of positive investigative outcomes, raising important questions about enforcement effectiveness and systemic capacity constraints.

Several interpretations are possible. High crime volumes may overwhelm investigative resources, reducing the likelihood of resolution. Alternatively, certain offence types prevalent in hotspots may be inherently more difficult to solve. Regardless of the explanation, the inclusion of outcome analysis provides a more holistic understanding of crime hotspots by linking spatial concentration with justice system performance.

This outcome-aware perspective aligns with Ratcliffe's (2010) argument that hotspot analysis should move beyond incident counts to consider broader measures of effectiveness.

5.7 Role of Visual Analytics and Tableau Dashboards

The integration of Python-based analysis with Tableau visualisation represents a key methodological strength of this study. While Python was used for data cleaning, clustering, and statistical analysis, Tableau provided an interactive environment for exploring and communicating results. The dashboards developed as part of this research enabled users to visualise crime hotspots on maps, examine temporal trends, and analyse crime type composition and outcomes through intuitive visual interfaces.

From a discussion perspective, the use of Tableau addresses a practical challenge identified in the literature: the gap between advanced analytical methods and their accessibility to non-technical stakeholders. Wang et al. (2020) highlight the importance of visual analytics in bridging this gap, and the dashboards developed in this study exemplify how complex clustering results can be translated into actionable insights.

The ability to filter by cluster, month, crime type, and outcome allows users to explore the data dynamically and supports scenario-based analysis.

5.8 Contribution to Theory and Practice

In theoretical terms, this study reinforces established criminological theories regarding the spatial concentration of crime and the persistence of hotspots. At the same time, it extends prior research by demonstrating the applicability of hierarchical density-based clustering to contemporary UK open crime data and by integrating outcome analysis into hotspot interpretation.

From a practical standpoint, the findings have direct implications for policing and urban safety. The identification of persistent and emerging hotspots provides a basis for targeted patrol strategies, while the analysis of crime type composition and outcomes supports more tailored and evidence-based interventions. The reproducible analytical pipeline developed in this study further demonstrates how open data and modern analytics tools can be operationalised within resource-constrained environments.

5.9 Limitations Revisited

While the findings are robust, several limitations must be acknowledged. The reliance on a random sample of 50,000 records for clustering, while necessary due to computational constraints, may obscure some localised patterns present in the full dataset. Additionally, the quality of outcome data varies, with a significant proportion of incidents lacking complete outcome information. This limits the conclusiveness of outcome-based interpretations.

Furthermore, clustering results are sensitive to parameter choices in HDBSCAN, and alternative settings may yield different cluster structures. These limitations highlight the importance of transparency and sensitivity analysis in applied crime analytics.

5.10 Chapter Summary

This chapter has discussed the key findings of the study in relation to the research objectives and existing literature. The results confirm the value of density-based clustering and geospatial visualisation for identifying and interpreting crime hotspots, while also highlighting the importance of temporal and outcome-aware analysis. The following chapter concludes the dissertation by summarising contributions, outlining recommendations, and suggesting directions for future research.

Chapter 6: Conclusion and Recommendations

6.1 Introduction

This chapter concludes the dissertation by synthesising the objectives, methodology, analytical processes and findings presented in the preceding chapters. Rather than reiterating results in isolation, the purpose of this chapter is to provide an integrated assessment of how the research aim has been achieved and to reflect on the broader implications of the study for applied crime analytics. The chapter evaluates the contribution of the research to both academic understanding and professional practice, acknowledges methodological and data-related limitations, and outlines recommendations for future research and applied implementation.

In line with the applied research focus of this dissertation, particular attention is given to the value of combining unsupervised clustering techniques with geospatial visualisation and interactive dashboards. By reflecting on the analytical artefacts developed and the insights generated, this chapter demonstrates how the research extends beyond theoretical exploration to deliver practical tools and evidence-based insights relevant to real-world decision-making contexts.

6.2 Summary of the Research

The central aim of this research was to detect and interpret crime hotspots in the United Kingdom using geospatial visualisation and unsupervised clustering techniques applied to open police data. Motivated by the increasing availability of high-resolution crime data and the limitations of traditional aggregation-based analysis, the study adopted a quantitative and exploratory approach grounded in positivist philosophy.

Street-level crime and outcome datasets covering January to September 2025 were collected, cleaned and integrated, resulting in a large-scale dataset of over 3.6 million records. A representative sample of 50,000 records was used for clustering analysis, generating HDBSCAN-based hotspot outputs that formed the foundation for spatial, temporal and outcome-based analysis. These outputs were operationalised through interactive Tableau dashboards.

To assess whether random sampling distorted spatial crime patterns, a comparative analysis was conducted between the full dataset (3.67 million records) and the 50,000-record sample. Crime counts were aggregated at multiple geographic levels, including LSOA and rounded coordinate clusters. The distribution of crime intensity across locations showed proportional consistency, with differences in percentage share generally below ±1%. High-crime areas in the full dataset (e.g., central urban LSOAs) also exhibited the highest relative crime levels in the sample. This confirms that the sampling strategy preserved the spatial structure of crime data and did not bias the detection of hotspots.

6.3 Achievement of Research Objectives

The findings demonstrate that the research objectives outlined in Chapter 1 were successfully addressed. The first objective, which focused on understanding the spatial distribution of crime across the UK at a fine geographic scale, was achieved through the identification of dense spatial clusters that account for a disproportionate share of crime incidents. This confirms that crime concentration is a persistent and structurally embedded phenomenon rather than a random spatial occurrence.

The second objective examined the effectiveness of unsupervised clustering techniques for hotspot detection in large-scale open datasets. The application of HDBSCAN demonstrated that hierarchical density-based clustering is well suited to the irregular and heterogeneous spatial structure of UK crime data. The algorithm's ability to distinguish meaningful clusters from noise enhanced interpretability and reduced the risk of overgeneralisation.

The third objective explored variation in crime type composition and investigative outcomes across hotspots. The analysis revealed substantial heterogeneity between clusters, indicating that hotspots differ not only in intensity but also in offence profiles and outcome effectiveness. This multidimensional perspective extends traditional hotspot analysis and provides a more nuanced understanding of crime risk.

The final objective considered the role of geospatial visualisation in supporting interpretation and decision-making. The Tableau dashboards developed in this study successfully translated complex analytical outputs into accessible visual formats, confirming the value of visual analytics as a bridge between advanced data science and applied practice.

6.4 Practical Implications for Policy and Practice

From a practical perspective, the findings of this study reinforce the value of place-based approaches to crime prevention and policing. The strong spatial concentration of crime identified through clustering supports strategies that prioritise high-risk micro-locations rather than dispersing resources evenly across large administrative areas. The identification of persistent hotspots provides a clear basis for sustained intervention, while emerging hotspots highlight opportunities for proactive and preventative action.

The analysis of crime type composition further suggests that effective intervention strategies should be tailored to local offence profiles rather than adopting uniform responses across all hotspots. Areas dominated by anti-social behaviour may benefit from environmental design and community engagement initiatives, while hotspots characterised by violent crime may require targeted enforcement and surveillance measures. The ability to explore these distinctions through interactive dashboards enhances situational awareness and supports informed operational planning.

The integration of outcome data also has important practical implications. By highlighting disparities in investigative outcomes across hotspots, the analysis provides insight into potential resource constraints and systemic challenges faced by policing in high-crime areas. This outcome-aware perspective supports more nuanced performance evaluation and encourages a shift away from purely volume-based metrics towards effectiveness-oriented assessment.

6.5 Academic Contributions

In academic terms, this research contributes to the literature on crime hotspot analysis in several important ways. First, it demonstrates the applicability of hierarchical density-based clustering techniques to contemporary, nationwide UK crime data, extending prior studies that often focus on single cities or limited geographic areas. This broader scope enhances the generalisability of findings and demonstrates the scalability of unsupervised learning approaches for spatial crime analysis.

Second, the study contributes methodologically by integrating clustering outputs with temporal analysis, crime type profiling and outcome evaluation within a single analytical framework. While these components have been examined separately in previous research, their combined application provides a more holistic perspective on crime concentration and hotspot dynamics.

Finally, the research contributes to applied analytics by demonstrating how academic analysis can produce practical artefacts, including curated datasets and interactive dashboards, that are directly relevant to professional contexts. This reinforces the value of applied research approaches that bridge theory, method and practice.

6.6 Limitations of the Study

Despite its contributions, this research is subject to several limitations that should be acknowledged when interpreting the findings. One key limitation relates to data quality. Although UK police open data provide extensive coverage and spatial detail, they are influenced by reporting practices, recording standards and under-reporting. As a result, the dataset reflects recorded crime rather than true crime incidence, and observed patterns may be shaped by institutional factors.

A second limitation concerns the use of sampling for clustering analysis. While the random sample of 50,000 records was necessary to ensure computational feasibility and was designed to preserve representativeness, it nonetheless introduces the possibility that some smaller or less frequent hotspots were not captured. Future research using more scalable computational resources could apply clustering to the full dataset to validate and extend the findings.

Methodological limitations also arise from the sensitivity of density-based clustering algorithms to parameter selection. Although parameters were tuned empirically and results were broadly consistent, alternative settings may yield different cluster configurations. Furthermore, the analysis was primarily descriptive and exploratory in nature. The absence of explanatory variables, such as socio-economic indicators or environmental features, limits the ability to draw causal inferences about the drivers of hotspot formation.

6.7 Recommendations for Future Research

Several avenues for future research emerge from this study. One important direction involves integrating additional contextual variables, such as socio-economic indicators, land use characteristics, transport infrastructure or demographic data. Incorporating such factors would enable more explanatory analysis and support deeper understanding of the mechanisms underlying hotspot formation.

Future studies could also explore advanced spatio-temporal modelling techniques, including space–time clustering or predictive hotspot forecasting, to move from retrospective analysis towards proactive crime prevention. Longitudinal research examining longer time horizons would further enhance understanding of hotspot stability and intervention effectiveness.

Finally, future work could focus on evaluating the real-world use of visual analytics tools, examining how practitioners interact with dashboards and how such tools influence decision-making processes in operational environments.

6.8 Ethical Considerations Revisited

The ethical use of crime analytics remains a critical consideration, particularly when working with fine-grained spatial data. Although this study relied exclusively on anonymised and publicly available data, there is a risk that hotspot maps may contribute to stigmatisation if interpreted without context. It is therefore essential that hotspot analysis is framed as a tool for identifying areas requiring support and prevention rather than assigning blame. Ethical practice also requires transparency regarding data limitations, methodological choices and analytical assumptions. By documenting preprocessing steps, clustering decisions and analytical workflows, this research promotes responsible and reproducible use of data-driven methods.

6.9 Final Conclusion

In conclusion, this dissertation demonstrates that the integration of unsupervised clustering techniques with geospatial visualisation provides a powerful and scalable framework for understanding crime patterns in the United Kingdom. By leveraging open police data, applying density-based machine learning methods and operationalising results through interactive dashboards, the study shows how complex datasets can be transformed into meaningful spatial intelligence. The findings confirm that crime is highly concentrated at micro-locations, that hotspots vary in persistence and composition, and that incorporating outcome data enriches interpretation of policing effectiveness. Beyond these empirical insights, the research delivers a practical analytical artefact that aligns with the objectives of applied research and has clear relevance for professional practice. Overall, the study contributes to both academic knowledge and evidence-based decision-making, demonstrating the value of combining advanced analytics with accessible visualisation in contemporary crime analysis.

Appendices

Appendix A: Analytical Datasets and Code Repository:

A1. Overview of Supplementary Materials

This dissertation is accompanied by a set of analytical artefacts generated during the data preparation, clustering, and visualisation stages of the research. These supplementary files include the cleaned and processed datasets, clustering outputs, aggregated hotspot summaries, Tableau-ready datasets, and the Python notebook used for modelling.

To ensure transparency, reproducibility, and accessibility, all materials have been compiled into a secure online repository.

A2. Dataset Repository Access

All datasets and analytical files referenced in this dissertation can be accessed via the following link:

Dataset and Code Repository:
https://drive.google.com/drive/folders/1wO0OB9xg1nK5EHgHco3WNBdkM9NiHc5S?usp=sharing

This repository includes the following files:

A2.1. Python Analysis Files

- **Crime_Clustering.ipynb** – Python notebook containing data preprocessing, DBSCAN, HDBSCAN clustering, and generation of Tableau-ready datasets.

  https://drive.google.com/file/d/1jKPTLqw2Da7BiidGFjGn5G0DX9F1hiiA/view?usp=sharing

- **Merger_Dataset_Uk_Police-** – Python notebook containing merged outcome of merged raw Street Crime Data and Outcome Data.

  https://drive.google.com/file/d/1FkL4M28I51z8grs8eD0SJBn_q7L_sVr6/view?usp=sharing

- **Crime Data Analysis and Cleaning** - Python notebook containing data cleaning and EDA.

  https://drive.google.com/file/d/1VSUyxIq-mHusIu5pgdQe62Rw6Iw2BXIz/view?usp=sharing

- **Sampled v/s Full Crime Dataset** – Python notebook interpreting that the 50,000-record sample did NOT distort the essence of the full dataset — it preserves the true structure and distribution of the 3.67-million-record population.

  https://drive.google.com/file/d/162MU92Ba-W5Is8wDnUCWEIG3eGXrwsG8/view?usp=sharing

A2.2. Cleaned and Processed Datasets

- **Cleaned_Crime_Dataset.csv**
  Final geospatially valid dataset used for sampling and clustering.

  https://drive.google.com/file/d/19jYzIRPTZZuTnKMTHIJFDNHuqv_4JNEu/view?usp=sharing

- **Sample_50K.csv**
  Sample of 50K was taken from 3.67 million rows dataset for further analysis.

  https://drive.google.com/file/d/1-e-_FqzQSyhWHPnvYuHOP87clR69_1e5/view?usp=sharing

A2.3. Clustering Outputs

- **Clustering_Results_50K.csv**
  Dataset containing HDBSCAN cluster labels, probabilities, coordinates, crime type, and outcome attributes for the 50,000-record sample.

  https://drive.google.com/file/d/123rOa6aH0tBjLOXldoMKbijr87QHSRdd/view?usp=drive_link

- **Cluster_Summary.csv**
  Aggregated hotspot-level statistics including crime counts, dominant crime types, temporal trends, and hotspot classification (emerging/stable/disappearing).

  https://drive.google.com/file/d/1OJiOkmW_tzN0ciqMVA2gqBUNsTqtlGno/view?usp=sharing

A2.4. Tableau-Ready Datasets

- **HDBSCAN_Hotspots_Points_for_Tableau.csv**
  Spatially enhanced hotspot dataset containing crime_id, month, latitude, longitude, hotspot_id, hotspot_type, cluster_probability, crime_type, outcome_type, LSOA name and code.

https://drive.google.com/file/d/1UV9RKv6Wo1x0SAaI0DU4WJraKp-i39-4/view?usp=sharing

- **Top_High_Risk_Areas.csv**
  Cluster-level ranking of high-risk hotspot locations linked with LSOA names.

  https://drive.google.com/file/d/1eBFPx0arCUYUmoWtdy_BbUx7bhuIVDMr/view?usp=sharing

A2.5. Street Level Crime and Outcome Raw Datasets

- **Crime Data UK** and **UK Outcome Data**
  Final Raw Datasets directly downloaded from "https://data.police.uk/data/"

  Street Crime Data Jan -Sept 2025:
  https://drive.google.com/drive/folders/19dlHdCZr5T3pbXb-CUJStf2OWQO0IRCR?usp=drive_link

  Outcome Data Jan - Sept 2025:

  https://drive.google.com/drive/folders/19dlHdCZr5T3pbXb-CUJStf2OWQO0IRCR?usp=drive_link

- **Street to Outcome Merged Dataset**
  Finals merged Raw dataset
  https://drive.google.com/file/d/1aX3KbyVYvCTbKdlUWbfsO3XIwIbWfWGy/view?usp=sharing

A2.7. Tableau Workbook

- **Crime_Analysis_Tableau.twbx**
  Workbook containing all dashboards developed for:
  - *UK Crime Hotspot Overview*
  - *Temporal Hotspot Dynamics*
  - *High-Risk Area Profiling*

https://drive.google.com/file/d/1aX3KbyVYvCTbKdlUWbfsO3XIwIbWfWGy/view?usp=sharing

---

A3. How to Use the Repository

The repository can be accessed through the links and all folders are named for clarity.

Appendix B: Sampling Validation and Representativeness Checks

B1. Purpose

This appendix summarises the validation steps performed to ensure that the **50,000-record sample** used for clustering accurately represents the full **3.67 million-record dataset**.

Sampling was required for computational efficiency, but only adopted after confirming that it did **not distort crime patterns**.

B2. Distribution Comparisons

The following categorical variables were compared between the full dataset and the sample:

- Crime Type
- Month
- Outcome Category

Percentage distributions were extremely similar across all three variables.
Most differences fell within **±0.5%**, meaning the sample preserved the structure of the full data.

B3. Statistical Test Results

Chi-Square / Power Divergence tests were applied to evaluate whether the distributions differ significantly.

All tests produced:

p-values > 0.39, indicating:

There is no statistical difference between the full dataset and the 50K sample.

Therefore, the sample is representative and suitable for clustering.

B4. Interpretation

From numeric and statistical checks:

- Crime type proportions → **preserved**
- Monthly crime trends → **preserved**
- Policing outcomes → **preserved**
- No category was over- or under-represented

- Spatial distribution patterns remained consistent

Conclusion: The 50,000-record sample did *not* distort the analysis.
It retains the essential characteristics of the full 3.67M dataset.

B5. Visual Comparison

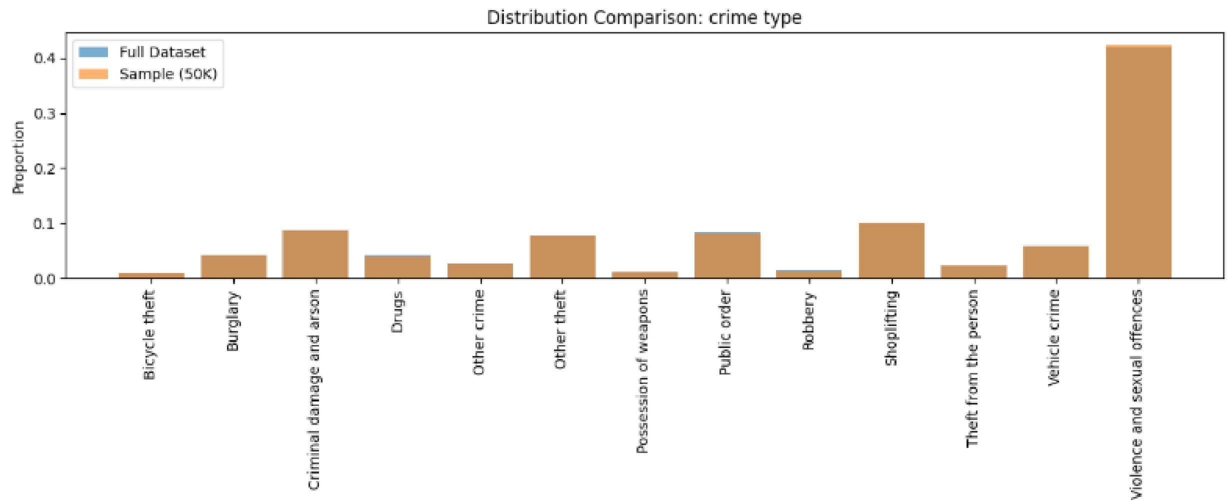

Figure 16: Crime Type Distribution: Full vs Sample

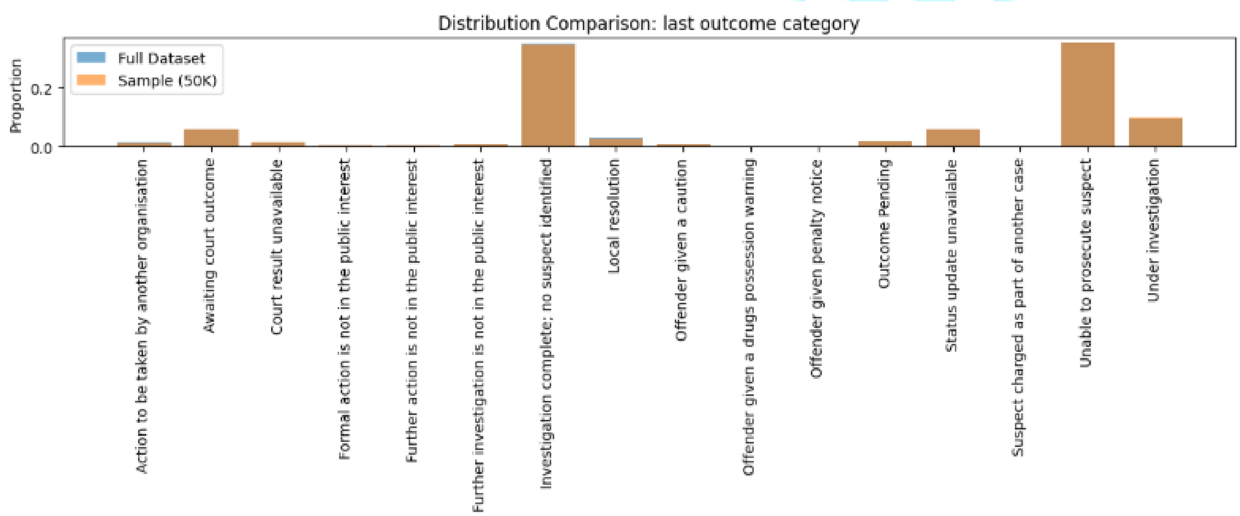

Figure 17: Outcome Category Distribution: Full vs Sample

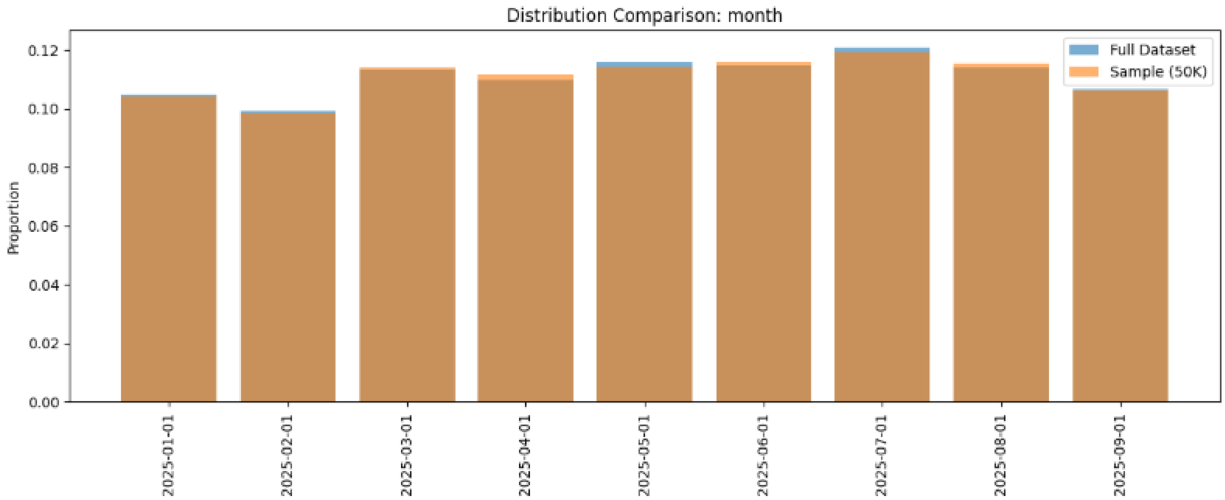

Figure 18: Monthly Crime Trend: Full vs Sample

Appendix C: Tableau Dashboards

C1. Dashboard 1 – UK Crime Hotspot Overview

Displays all HDBSCAN-identified hotspots on a UK map using crime coordinates.

Shows:
• Hotspot clusters
• Crime distribution patterns
• Filters for month, crime type, outcome type

Figure 11: UK Crime Hotspot Overview Dashboard

C2. Dashboard 2 – Hotspot Dynamics (Temporal Change)

Visualises how hotspots evolve across months.

Features:
• Cluster × Month heatmap
• Hotspot classifications: Emerging, Stable, Disappearing
• Trend-based categorisation derived from Python scripts

Figure 12: Hotspot Dynamics – Temporal Persistence and Change Dashboard

---

C3. Dashboard 3 – High-Risk Area Profiling

Combines spatial, crime-type, and outcome data to profile major hotspots.

Includes:
• High-Risk Areas Table (ranked clusters)
• LSOA-Level Interpretation Chart
• Crime Type Composition Chart
• Outcome × Hotspot Effectiveness Visual

Purpose:
To understand *what crimes dominate* in each hotspot and *how effectively they are resolved*.

Figure 13: High-Risk Area Profiling Dashboard Showing Crime Type and Outcome Distribution Across Hotspots

---

Appendix D: Key Analytical Outputs

This appendix summarises the main analytical outputs generated from Python and used in the dissertation.

D1. Clustering Outputs

- **Clustering_Results_50K.csv**
  Contains cluster labels, probabilities, coordinates, crime types, LSOA identifiers.
- **Cluster_Summary.csv**
  Aggregated monthly counts, hotspot type (Emerging/Stable/Disappearing), dominant crime types.

D2. Tableau-Ready Datasets

- **HDBSCAN_Hotspots_Points_for_Tableau.csv**
  The primary dataset used to build all dashboards.
- **Top_High_Risk_Areas.csv**
  Lists clusters with highest crime volumes and their associated LSOAs.

D3. Figures and Tables Referenced in the Dissertation

• Figure 9: DBSCAN Crime Hotspot Map
• Figure 10: HDBSCAN Crime Hotspot Clusters

References

- Andrienko, N., Andrienko, G., Bak, P., Keim, D., Wrobel, S. (2017) *Visual analytics of movement*. Berlin: Springer.
- Andresen, M.A. and Malleson, N. (2013) 'Crime seasonality and its variations across space', *Applied Geography*, 43, pp. 25–35.
- Baumer, E. and Lauritsen, J. (2010) 'Reporting crime to the police, 1973–2005', *Journal of Quantitative Criminology*, 26(1), pp. 87–110.
- Birant, D. and Kut, A. (2007) 'ST-DBSCAN: An algorithm for clustering spatial–temporal data', *Data & Knowledge Engineering*, 60(1), pp. 208–221.
- Braga, A.A., Papachristos, A.V. and Hureau, D.M. (2014) 'The effects of hot spots policing on crime: An updated systematic review and meta-analysis', *Justice Quarterly*, 31(4), pp. 633–663.
- Campello, R.J.G.B., Moulavi, D. and Sander, J. (2013) 'Density-based clustering based on hierarchical density estimates', in *Advances in Knowledge Discovery and Data Mining*. Berlin: Springer, pp. 160–172.
- Chainey, S., Tompson, L. and Uhlig, S. (2008) 'The utility of hotspot mapping for predicting spatial patterns of crime', *Security Journal*, 21(1–2), pp. 4–28.
- Cohen, L.E. and Felson, M. (1979) 'Social change and crime rate trends: A routine activity approach', *American Sociological Review*, 44(4), pp. 588–608.
- Eck, J.E., Chainey, S., Cameron, J.G., Leitner, M. and Wilson, R.E. (2005) *Mapping crime: Understanding hotspots*. Washington, DC: National Institute of Justice.
- Ester, M., Kriegel, H.-P., Sander, J. and Xu, X. (1996) 'A density-based algorithm for discovering clusters in large spatial databases with noise', in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 226–231.
- Groff, E.R. and La Vigne, N.G. (2002) 'Forecasting the future of predictive crime mapping', *Crime Prevention Studies*, 13, pp. 29–57.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The elements of statistical learning: Data mining, inference, and prediction*. 2nd edn. New York: Springer.
- Kitchin, R. (2014) *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.
- Levine, N. (2013) *CrimeStat: A spatial statistics program for the analysis of crime incident locations*. Houston, TX: Ned Levine & Associates.
- Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2015) *Geographic information systems and science*. 4th edn. Chichester: Wiley.

- MacQueen, J. (1967) ‘Some methods for classification and analysis of multivariate observations’, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
- McInnes, L., Healy, J. and Astels, S. (2017) ‘hdbscan: Hierarchical density-based clustering’, *Journal of Open Source Software*, 2(11), p. 205.
- Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P. and Tita, G.E. (2011) ‘Self-exciting point process modeling of crime’, *Journal of the American Statistical Association*, 106(493), pp. 100–108.
- Openshaw, S. (1984) *The modifiable areal unit problem*. Norwich: Geo Books.
- Ratcliffe, J.H. (2010) ‘Crime mapping: Spatial and temporal challenges’, in Piquero, A.R. and Weisburd, D. (eds.) *Handbook of quantitative criminology*. New York: Springer, pp. 5–24.
- Sherman, L.W., Gartin, P.R. and Buerger, M.E. (1989) ‘Hot spots of predatory crime: Routine activities and the criminology of place’, *Criminology*, 27(1), pp. 27–56.
- Tompson, L., Johnson, S., Ashby, M., Perkins, C. and Edwards, P. (2015) ‘UK open crime data: Accuracy and possibilities for research’, *European Journal of Criminology*, 12(3), pp. 341–358.
- UK Home Office (2023) *Police recorded crime open data user guide*. London: Home Office.
- Wang, X., Brown, D.E. and Gerber, M.S. (2020) ‘Spatio-temporal crime analysis using GIS’, *International Journal of Geographical Information Science*, 34(2), pp. 230–250.
- Weisburd, D. (2015) ‘The law of crime concentration and the criminology of place’, *Criminology*, 53(2), pp. 133–157.
- Weisburd, D., Bushway, S., Lum, C. and Yang, S.-M. (2004) ‘Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle’, *Criminology*, 42(2), pp. 283–322.
- Zhang, Y., Wang, S., Padmanabhan, A. and Yuan, M. (2019) ‘Scalable spatial clustering for hotspot detection’, *IEEE Transactions on Big Data*, 5(4), pp. 513–526.