Clustering Analysis of Crime Patterns in the City of London: A Data-Driven Approach to Identifying Safe and Risky Areas

Amartey Pius

Faculty of Computer Science and Engineering

University of Wolverhampton

Wolverhampton United Kingdom

p.q.amartey@wlv.ac.uk

***Abstract*— This report analyses the City of London dataset from January 2022 to January 2023 to determine the City of London neighbourhood with the lowest crime rate. The study employs descriptive statistics, data visualisation techniques, and clustering analysis to identify areas with high crime rates and recurrence patterns. The report identifies neighbourhoods with high crime rates and provides recommendations for secure neighbourhoods based on the research. This work can assist those seeking safer areas in the City of London and also contribute to the broader community of crime analysis**

Keywords— Clustering, crime, analysis, crime data, visualisation, crime prediction, prediction method

Introduction

The City of London is a major financial centre with a population of about 9,000 people, which grows to around 450,000 during the working week. Crime is a major issue in the area, with a wide range of offences reported to the police daily. Many people are concerned about their safety and are looking for ways to reduce the risks associated with living in London's financial district.

The City of London Crime Dataset contains 50,714 rows and 12 columns. The columns in the dataset are as follows:

* Crime ID: A unique identifier for each crime reported.
* Month: The month in which the crime was reported.
* Reported by: The authority that said the crime.
* Falls within: The geographic area where the crime occurred.
* Longitude: The longitude of the location where the crime occurred.
* Latitude: The latitude of the location where the crime occurred.
* Location: A description of the location where the crime occurred.
* LSOA code: Lower Layer Super Output Area code.
* LSOA name: Lower Layer Super Output Area name.
* Crime type: The type of crime that was committed.
* Last outcome category: The outcome of the investigation.
* Context: Additional contextual information about the crime.

This study aims to provide a comprehensive overview of crime in the City of London using the available dataset from January 2022 to January 2023 to understand the problem better. The study will use a variety of methodologies, including descriptive statistics, data visualisation, and spatial analysis, to identify high-crime areas and make recommendations for safer neighbourhoods to live in.

This study will be divided into sections that will provide an overview of crime in the City of London, analyse crime data, perform a spatial analysis of the crime data using GIS software, and make recommendations for safer neighbourhoods based on the study's findings. The first section will give an overview of crime in the City of London, including the types of offences and the frequency with which they occur. The second section will use descriptive statistics and data visualisation techniques to analyse the crime data. The third section will employ GIS software to conduct a spatial analysis of crime data and identify crime hotspots in the City of London. The final section will recommend safer neighbourhoods based on the study's findings.

This study will use the following methodology to achieve its goal:

## Descriptive statistics

The study will use summary statistics to describe the overall trend in crime over the study period.

## Data visualisation

The study will employ charts, graphs, and maps to provide a visual representation of the data.

## Spatial analysis:

The study will use GIS software to conduct a spatial data analysis to identify crime hotspots in the City of London. The study intends to identify the areas with the highest crime rates and recommend safer neighbourhoods by analysing the data. This report will be helpful for people looking for a safer neighbourhood in the City of London and the larger research community in crime analysis and prevention.

The methodologies mentioned above will be analysed in detail in Chapter 3

Contribution:

In this report we contribute a study of data from January 2022 to January 2023 for the City of London that sheds important light on the trends and patterns of local crime. Making educated choices about where to live, work, or visit within the City of London can be made easier if the communities with the lowest crime rates are identified.

We introduce many methods for data visualization, which will contribute to make it simpler for policymakers and law enforcement organizations to identify locations with high crime rates and devise focused responses if charts, graphs, and maps are used to visually depict the data. This method may be used in different datasets and places to offer insightful information and help identify communities with high crime rates.

Next, we develop GIS software for spatial data analysis to give a fresh perspective on crime analysis. The location of crime hotspots in the City of London can help guide the distribution of resources to those communities who most require them. The recommendations made in our study for safer communities based on the analysis may be helpful to locals, tourists, legislators, and law enforcement organizations. The results of this study can help crime analysis and prevention by offering a case study of how spatial analysis can be used to pinpoint crime hotspots and suggest safer communities.

Overall, the contributions of this study offer insightful information about how to employ clustering analysis, descriptive statistics, and data visualization approaches to pinpoint problematic and safe locations in the City of London. These techniques may be used on different datasets and places to offer insightful information on crime patterns and trends, eventually helping to achieve the objective of lowering crime rates and enhancing community safety.

# literature review

## Crime Trend Analysis

A study on the effectiveness of stop and search as a crime-prevention strategy in London was done by M. Tiratelli et al. (2018). According to Tiratelli et al. (2018), London has experienced a decline in crime rates over the past decade. Nonetheless, some crimes have become more prevalent recently, including burglaries and knife crimes. Their data analysis spanning 2006 to 2016 revealed that while stop and search operations rose throughout this time, they had little impact in reducing crime.

Graphical user interface

Description automatically generated

*Figure 2.0 Crime Trend Analysis Sample data*

Map

Description automatically generated

*Figure 2.1 Crime Trend Analysis Sample data*

Graphical user interface, chart

Description automatically generated

*Figure 2.2 Crime Trend Analysis Sample data*

Figure 2.0 is an example of crime trend analysis using line graphs to demonstrate how crime rates have changed. Time is shown on the x-axis, while the number of offences is shown on the y-axis. Long-term patterns in crime, such as an overall rise or fall over several years, maybe found using this kind of graph. Other methods include bar graphs(figure2.0), heatmaps(figure2.1) and scatterplots(figure 2.2).

Tiratelli et al. (2018) examined data from the London Metropolitan Police Department and discovered that stop and search had minimal impact on lowering crime. According to their findings, stop-and-search policies may cause crime to be displaced to other locations. This study suggests that targeted interventions and community policing should be considered as alternative crime-prevention measures by legislators and law enforcement organisations.

## Gap in the Research

Only data from the London Metropolitan Police Department is analysed. The results might not, therefore, apply to other areas or nations. The study also failed to consider additional elements, such as modifications in social and economic situations, that would have contributed to the drop-in crime rates in London.

The study by Tiratelli et al. (2018) offers insightful information on London's crime patterns and the efficiency of stop and search as a deterrent. Nonetheless, their findings imply that governments and law enforcement organisations must concentrate on developing new ways to combat crime in high-risk locations.

## Crime Trends

Analyzing crime trends may give important insights into patterns and guide judgements on allocating resources for the best efficient enforcement techniques. T. Cheng and D. Williams (2018) used space-time analysis to examine crime trends in the heart of London.

Cheng and Williams (2018) examined crime trends in central London using a space-time analysis methodology. They analysed data from the London Metropolitan Police Department and discovered that crime hotspots congregated in specific locations, including shopping malls, entertainment districts, and transportation hubs. The theft survey also found that some crimes, like theft, happen more frequently at times of the day and on days of the week.

## Finding Safe Neighborhoods

Finding safer neighborhoods in London is a topic Cheng and Williams (2018) did not specifically examine. Yet, their examination of crime trends might offer insightful information. The study discovered that crime hotspots frequently tended to gather places, suggesting that these regions could need greater enforcement resources. Policymakers and law enforcement organisations may use this data to identify high-risk locations and spend resources appropriately.

## Space-Time Analysis Method

Cheng and Williams (2018) applied a space-time analysis method to better understand crime trends. This method can pinpoint crime hotspots and assist in making predictions about the appearance of new crimes by analysing crime trends in both the spatial and time dimensions. With the use of this data, resources may be assigned to efficient policing tactics like focused patrols and preventative actions.

These findings imply that crime hotspots frequently congregate places, highlighting the necessity of tailored policing approaches. A useful method for determining high-risk regions and directing resources for efficient enforcement tactics is space-time analysis.

Text

Description automatically generated

*Figure 2.3 Crime and theft data plotted totime of day in hours*

Figure 2.3 above depicts the variations of thefts was comparable to all other crime categories, with the exception of just one peak in the thefts, at about 3 pm, and a lengthier

dip or recline from 2 am until 10 am.

Text

Description automatically generated

*Figure 2.4 Bird’s eye view of clusters in Space-Time Cube*

The outcomes of using this visualisation approach are crystal obvious and simple to understand. The method follows the analysis and is used to pinpoint crime hotspots and assist in making predictions.

Text

Description automatically generated

*Figure 2.5 The SOM method*

In Figure 2.5.2 The SOM Method is a high-dimensional dataset often represented as a two-dimensional discretised pattern using the self-organising map (SOM). This would be an effective method in this case as is extremely effective and very useful particularly for the visualisation and analysis of the attribute space. The clusters in figure 2.5 are clear and distinct and with the linked views of the PCP, reorderable matrix and map matrix, the trends can be clearly interpreted.

## Hot Street Analysis

Wu and Li (2022) conducted a study on crime patterns in London using a "hot street" analysis approach to identify areas with high crime rates.

Wu and Li (2022) looked at how lockdowns affected London's crime trends. They discovered that while crime rates in residential areas increased during the COVID-19 lockdowns, they reduced in business areas. These findings imply that lockdowns may alter crime patterns unintentionally, and policymakers should take this into account when putting interventions into place.

The research offers useful information for locating safer areas of London. Their "hot street" analytical method may be used to locate high-risk regions that need specialised enforcement measures. Moreover, their research on how lockdowns affect crime trends might help with resource allocation decisions for efficient enforcement tactics.

Graphical user interface, scatter chart

Description automatically generated

*Figure 2.6 LISA Cluster Map*

In figure 2.6 Hot Street analysis The LISA Cluster Map shows the significant locations color coded by type of spatial autocorrelation. They show that in figure 2.6 had highlighted the "hot street" and "cold street" crime areas and discussed the connection between the target place and its surroundings. The "LH" streets (legend in light blue) are those with low crime rates but are bordered by streets with high crime rates. The "HH" streets (red segments on the left map) are streets with high crime rates and are flanked by streets with high crime rates. Streets shown in grey (with the value "ns") had been assumed to have offences that were spread geographically randomly.

ns") had been assumed to have offences that were spread geographically randomly.

Graphical user interface, application, Word

Description automatically generated

*Figure 2.7 Hot Street maps of crime in Haringey*

Hot spot analysis is a useful technique for finding problem areas, more efficiently allocating resources, enhancing decision-making, directing planning and policy formulation, and tracking change through time.

## Environment Population Measures

To pinpoint crime hotspots in London, Malleson and Andresen (2016) employed ambient population metrics including foot traffic, traffic volume, and public transportation use. They discovered that these variables had a positive correlation with crime, indicating that places with heavy pedestrian or vehicular traffic were more likely to encounter crime. Also, their study showed that the influence of ambient population measurements on different types of crime varied, with theft and robbery having a stronger association with foot traffic than other crimes.

## Spatial and Temporal Analysis

Malleson and Andresen (2016) conducted a geographical and temporal examination of crime patterns to identify crime hotspots in London. Using a kernel density analysis to map crime rates across the city, they found that crime hotspots were concentrated in specific areas, such as industrial and residential districts with high population densities. Also, their research revealed that crime is more common on weekends and at night.

A picture containing graphical user interface

Description automatically generated

*Figure 2.8 Hot Street maps of crime in Haringey*

In figure 2.8 shows the spatial distribution of the actual and predicted values for community crime rate in August. These results show that the predictive distributions can reflect the overall distribution function of the actual crime rate.

## Safer Neighborhoods in London

Malleson’s and Andresen's (2016) study provides valuable insights into identifying safer neighborhoods in London. The use of population measures helped to identify areas more likely to encounter crime with high footfall or traffic volume. Their geographical and temporal analysis can also help decision-makers decide how to best allocate resources for efficient police operations.

Gaps in the research include that it only looked at crime trends in London (which we are focusing on) but would be helpful, which may restrict the applicability of its results to other areas or nations. In addition, the study did not take into consideration additional elements, such as alterations and changes in social and economic situations, that may have contributed to crime trends. In conclusion safer neighborhoods in the city of London and analysing crime trends is a crucial area of research. Research has given a thorough summary of the available research on this subject, highlighting the many approaches taken to pinpoint safer neighborhoods and study crime patterns. The assessment also highlighted these approaches' drawbacks and difficulties, such as the trustworthiness and availability of the data. According to the study, finding safer neighborhoods in the City of London requires a multidimensional approach that considers a number of factors such as socioeconomic traits, the physical environment, and community involvement. On the other hand, crime trend analysis requires the use of trustworthy and legitimate data, reducing analytical methods, and a thorough understanding of the underlying factors influencing crime trends. Overall, the study has underlined the need to understand how crime is distributed geographically in the City of London and the requirement for efficient crime prevention measures that may address the underlying causes of crime. Future studies in this field should continue to investigate novel approaches for determining safer neighborhoods and examining crime patterns, as well as consider the effects of cutting-edge technology like artificial intelligence and machine learning on crime prevention and public safety.

# Methododlogy

This study's methodology included several phases, each of which contributed to the overall analysis of the City of London crime dataset between January 2022 and January 2023. The methodology is divided into the following sections:

* Data Collection
* Data Cleaning
* Data Analysis
* Clustering Analysis
* Geographical Analysis
* Recommendations
* Future work

Each block will be discussed in detail below.

## Data Collection

The first step in the methodology was to collect the City of London crime dataset from January 2022 to January 2023. The dataset was obtained from the UK police website and included information on reported crimes in the City of London, such as the date and time of the crime, its location, the type of crime committed, and the outcome of the investigation. Figure 1 below shows the first five data in the dataset.

Graphical user interface, text

Description automatically generated

Figure 3 Sample data

## Data Cleaning

Before performing clustering analysis on the City of London crime dataset, the data was cleaned using various Python functions and tools to ensure that it was accurate and reliable. Inconsistencies in the data were found and addressed whenever it was practicable to do so. In total, 3,738 rows of data were removed due to missing information. One of the cleaning steps involved dropping the 'LSOA code' column from the 'clustering data' Data Frame.

In addition to dropping the 'LSOA code' column, the 'context' and 'ID' columns were also dropped from the original City of London crime dataset.

The 'context' column contains additional information about the crime irrelevant to our analysis. In contrast, the 'ID' column is simply an identifier for each crime that is unnecessary for our analysis. Therefore, dropping these columns helps to simplify the dataset and make it easier to work with.

## Data Analysis

Several Python modules and functions were utilised in order to get insights into the different types of crimes committed, the locations of those crimes, and the trends that have developed over time. The dataset was initially evaluated to identify which types of crimes were committed the most frequently and which locations had the highest incidence of those crimes. To find patterns over the course of time, the number of crimes committed each month was also studied. Figure 2 below shows the different crimes and the number of times they occur.

Text

Description automatically generated

Figure 3.1 Various crimes and the frequency of each

Figure 3 below shows the number of crimes which occur each month.

Table

Description automatically generated

Figure .2 Number of crimes in each month

## 3.4 Clustering Analysis

Clustering analysis was carried out so that recurring patterns and associations within the dataset could be discovered. First, the dataset was filtered so that it only contained records originating from the City of London. After that, the data were cleaned up by first transforming the 'Crime type' column into binary indicators, and then by scaling the data. After then, hierarchical clustering was carried out in order to discover connections hidden within the data. Folium was used to create a visual representation of the generated clusters on a map. Figure 4 below shows the clusters on a dendrogram.

Chart, box and whisker chart

Description automatically generated

Figure 3.3. determining number of clusters using dendrograms

# Results

Geographical analysis of the City of London crime dataset revealed that certain areas have a higher incidence of criminal activity than others. The locations marked with red dots on the map have a high incidence of criminal activity, indicating that residents and visitors to these areas may be at a higher risk of crime. These high-risk areas are often associated with nightlife, tourist attractions, and other busy areas with a high volume of people, making them more attractive to criminals.

On the other hand, locations marked with blue dots on the map have a low incidence of criminal activity, indicating that these areas are relatively safe. These low-risk areas are often residential neighbourhoods or commercial areas with less foot traffic, making them less attractive to criminals.

Areas marked with black spots on the map have a moderate level of crime, indicating that residents and visitors to these areas should exercise caution. These moderate-risk areas are often business districts or neighbourhoods with a mix of residential and commercial activity.

Lastly, the regions marked with green dots have a mild level of crime, indicating that these areas are relatively safe. These mild-risk areas are often residential neighbourhoods or commercial areas with less foot traffic and lower crime rates.

Based on this analysis, we would advise anyone looking for a safer place to live in the City of London to consider neighbourhoods with low or mild levels of crime, marked with blue or green dots on the map. We also recommend that visitors to the City of London exercise caution in high-risk areas marked with red dots on the map.

The purpose of this analysis was to provide a visual representation of the data and to help identify high-crime areas in the City of London. This analysis can be useful for a variety of purposes, such as helping residents and visitors make more informed decisions about where to live, work, or travel in the City of London. It can also be used by law enforcement agencies to identify areas that require more attention and resources for crime prevention and control. Figure 5 below shows thee geographical analysis and implementation on a map.

**Map

Description automatically generated with medium confidence**

Figure 3.4 geographical analysis of crimes on map

Figure 6 further explains the geographical analysis on a graph.

Chart, scatter chart

Description automatically generated

Figure 4 Crime analysis on graph

# Limitations

This report suffers from several shortcomings, any one of which may compromise the reliability of the findings. To begin, the span of time covered by the data utilised in the analysis is only one year, which may not be sufficient to uncover long-term tendencies or patterns in criminal activity. Second, the investigation only considers offences that were recorded, which may not be representative of the true rate of crime in the City of London. Finally, the quality of the data that was obtained as well as the methods that were utilised for the cleaning and analysis of the data may have an impact on the precision of the results.

# Recommendations

Considering the findings of the investigation, the research suggests that residents of the City of London who are looking for safer neighbourhoods should consider locations that have lower overall crime rates. Those who are looking for a safer place to live should probably avoid living in areas that have a high crime rate because these areas are likely to have a higher risk of crime. In addition, the report suggests that the City of London Police Department should concentrate on regions that have high crime rates and devise ways to minimise crime in those areas.

# Future Work

To better understand the long-term patterns of criminal activity in the City of London, it may be beneficial for subsequent research to broaden the scope of the analysis so that it incorporates data from additional years. It may also be beneficial to investigate the connection between criminal activity and other factors, such as population characteristics, socioeconomic conditions, and urban planning. Finally, doing research into the efficacy of various crime prevention measures and coming up with ideas for improving the effectiveness of crime prevention in the City of London may prove to be valuable.

# Conclusion

In conclusion, the City of London crime dataset covering the time period January 2022 through January 2023 was evaluated with a variety of libraries and functions available in Python. The data had to be cleaned before it could be considered trustworthy and accurate. The tools of data analysis were then utilised in order to get insights into the different sorts of crimes committed, the locations of those crimes, and patterns over time. At long last, a clustering analysis was carried out in order to discover recurring themes and connections hidden within the information. The generated clusters were represented on a map using Folium, which provided useful insights into the linkages between the different forms of criminal activity and places within the City of London.

##### References

1. Monika and A. Bhat, "An analysis of Crime data under Apache Pig on Big Data," (2019) Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, (2019).
2. Tiratelli, M. et al. (2018) Does Stop and Search Deter Crime? Evidence From Ten Years of London-wide Data. British journal of criminology. [Online] 58 (5), 1212–1231.
3. T. Cheng & D. Williams (2018) SPACE-TIME ANALYSIS OF CRIME PATTERNS IN CENTRAL LONDON. International archives of the photogrammetry, remote sensing and spatial information sciences. [Online] XXXIX-B247–52.
4. Wu, Y. & Li, Y. (2022) ‘Hot street’ of crime detection in London borough and lockdown impacts. Geo-spatial information science. [Online] ahead-of-print (ahead-of-print), 1–17.
5. Malleson, N. & Andresen, M. A. (2016) Exploring the impact of ambient population measures on London crime hotspots. Journal of criminal justice. [Online] 4652–63.
6. M. Sathiyanarayanan, A. K. Junejo and O. Fadahunsi, "Visual Analysis of Predictive Policing to Improve Crime Investigation," (2019) International Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019,
7. M. Hassan and M. Z. Rahman, "Crime news analysis: Location and story detection," (2017) 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, (2017)
8. Hu, T., Zhu, X., Duan, L. and Guo, W., 2018. Urban crime prediction based on spatio-temporal Bayesian model. *PloS one*, *13*(10), p.e0206215.
9. Biron, K., Mansoor, W., Miniaoui, S., Atalla, S., Mukhtar, H., & Bin Hashim, K. F. (2019). Data Science Tools for Crime Investigation, Archival, and Analysis.
10. Carvalho, A., & Pedrosa, I. (2021). Data Analysis in Predicting Crime: Predictive Policing. In 2021 16th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). Chaves, Portugal. doi: 10.23919/CISTI52073.2021.9476393.
11. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. Computer, 37(4), 50-56. doi: 10.1109/MC.2004.1297301.
12. Almuhanna, A. A., Alrehili, M. M., Alsubhi, S. H., & Syed, L. (2021). Prediction of Crime in Neighbourhoods of New York City using Spatial Data Analysis.
13. Alqahtani, A., Garima, A., & Alaiad, A. (2019). Crime Analysis in Chicago City. In 2019 10th International Conference on Information and Communication Systems (ICICS) (pp. 166-172).
14. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017). An overview on crime prediction methods. In 2017 6th ICT International Student Project Conference (ICT-ISPC) (pp. 1-5). Johor, Malaysia. doi: 10.1109/ICT-ISPC.2017.8075335.
15. Al-Rummana, G. A., Al-Ahdal, A. H. A., & Shinde, G. N. (2021). The Role of Big Data Analysis in Increasing the Crime Prediction and Prevention Rates. In Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques, and Applications (pp. 209-220). Wiley. doi: 10.1002/9781119711629.ch10.
16. Ibrahim, R., & Shafiq, M. O. (2017). On the Measurement and Analysis of Safety in a Large City. In 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech) (pp. 1068-1075). Orlando, FL, USA. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.174.
17. Vassilakos, Andreas, Matthias Gottlieb and Maurice E. Dawson. “Towards Crime Prevention Using Big Data Analytics: A Literature Review with an Explorative Case Study.” (2019).
18. Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016). Crime prediction and forecasting in Tamilnadu using clustering approaches. In 2016 International Conference on Emerging Technological Trends (ICETT) (pp. 1-6). IEEE. doi: 10.1109/ICETT.2016.7873764.
19. Wang, J., Hu, J., Shen, S., Zhuang, J., & Ni, S. (2020). Crime risk analysis through big data algorithm with urban metrics. Physica A: Statistical Mechanics and its Applications, 545, 123627.
20. Taylor, D. & Abarno, D. (2022). Using big data from probabilistic genotyping to solve crime. Forensic Science International: Genetics, 57, 102631.