Spatiotemporal Analysis of Crime in London

Sajeel Nadeem Alam

Abstract—An extensive spatiotemporal analysis of crime incidents in London can help police forces make better policies and allocate their resources more effectively. This study conducts a temporal analysis to investigate how crime incidents have varied over the years as well as spatial analysis to study the spatial distribution of crime across the city. The study uses LSOA level crime data updated by the Metropolitan Police Service (MPS) and sourced from London Datastore. It ranges from 2010 to 2021, and has information on various categories of crime in each LSOA. This was merged with the 2011 London census data to compare correlations between crime rates and demographic/socioeconomic factors. The distribution of crime across MPS Business Command Units (BCUs) was also analysed. For temporal analysis, line graphs and heatmaps were used to analyse crimes counts across the years and months. To discover spatial patterns, K-Means clustering was performed to group together areas with similar crime rates which were visualised on choropleth maps of London. Results show that the number of crime incidents increased till 2019 and then dropped suddenly in 2020. Areas in Central London have much higher crime rates as compared to areas further away from the centre of the city.

1 Problem Statement

Crime is defined as "An act or omission constituting an offence (usually a grave one) against an individual or the state and punishable by law" (OED, 2024). To ensure that people committing crimes are held accountable, police forces must detain them first. The force responsible for handling crime in London boroughs is the Metropolitan Police Service (MPS) which is divided into 12 Business Command Units (BCUs). It is essential for the MPS to understand spatiotemporal patterns of crime and identify crime hotspots within London to allocate resources appropriately. This is the purpose of the study and in order to help achieve it, the research questions it aims to answer are:

1. Over the years, what time of the year experiences the highest crime rates across London?
2. What is the spatial distribution of crime across London and which areas have the highest crimes rates?
3. Are MPS Business Command Units (BCUs) distributed appropriately according to the crime rate?
4. Is there a correlation between crime rates and demographic/socioeconomic factors of London areas?

This study uses a crime dataset provided by the MPS (MPS, 2017). It contains monthly counts of various categories of crime in each Lower layer Super Output Area (LSOA) of every borough from 04/2010 to 07/2021 which will help analyse the temporal and spatial patterns of crime. BCU assignments were added to analyse their distribution (Wikipedia, 2025). This was merged with 2011 census data to investigate correlations between crime rates of London boroughs and their demographic/socioeconomic factors (GLA).

2 State of the Art

(Maciejewski et al., 2010) built a system to use visual analytics to understand spatiotemporal hotspots when dealing with multiple multivariate datasets by linking them. They worked with West Lafayette Police Department's crime data. For temporal analysis, line charts displaying count of crime incidents were made, where users could display multiple series on a single plot. In geospatial analysis, they displayed incidents using coordinates on grids and aggregated them for privacy when zoomed in. KDE with a variable kernel helped in identifying crime hotspots on choropleth maps which worked for both sparse and dense data. Their system allowed integrating choropleth maps and corresponding line graphs with the functionality to zoom into hotspots to find new peaks within them. Inspired from this work, line graphs can be used in this study to visually analyse crime incidents across the years. Aggregating incidents across all crime categories and months while using choropleth maps for visualisation will help with understanding the spatial pattern of crimes. However, KDE cannot be implemented due to the unavailability of geocoordinates in the dataset.

(Garcia et al., 2021) focuses on the visual analysis of crime in urban areas such as Sao Paulo which have a high crime rate and variability in the patterns of crime. They built a dataset with the Sao Paolo police department holding information about crime incidents across 7 years. Their aim was to identify crime hotspots using crime counts along with recurrence rate and variability. Their system 'CrimAnalyzer' allows both spatial and temporal crime analysis in particular regions of the city and compare them which is relevant for this study. They state that using KDE is problematic since poor parameter choices can lead to inaccurate results. Hence, they identified crime hotspots using non-negative matrix factorization and Getis-Ord G* for comparison while displaying them on choropleth maps. London is an urban area hence this work is comparable. Similar to this work, this study will also explore the identification of crime hotspots using choropleth maps. Moreover, they try to understand seasonal changes in crime which is relevant to this study.

(Joshi et al., 2017) visually analysed crime by performing K-Means clustering. They used a dataset from the Bureau of Crime Statistics and Research of New South Wales which contained 9000 records of crime that took place in the New South Wales region of Australia. The dataset contained attributes like location, sublocation, crime type, crime subtype and year which is similar to the attributes of the dataset in this study. The goal of their work was to cluster regions according to the count of crime incidents and find the crime rate of each crime type and cities with high crime rates. Optimal number of clusters were selected using silhouette analysis and the region and type of crime with the highest crime rate in each cluster were identified. Since the dataset used in this work is similar to the MPS dataset, clustering can be implemented (and validated using silhouette analysis) to visually analyze the spatial distribution of crime across London.

3 Properties of the Data

This study uses the MPS crime dataset which contains the count of each minor category of crime in each major category across all the LSOAs in each London borough from 04/2010 to 07/2021. It contains 148358 rows and 141 columns. Each row represents one minor category of crime in a specific LSOA. Columns 1-5 represent the LSOA code, LSOA name, borough, major category and minor category and each subsequent column after this contains the count of crime as integers for a specific month. This was merged with the 2011 London census data where each row represents one borough and contains information about various demographic and socioeconomic factors. Borough names from both datasets were used to link them. The MPS has divided London into 12 business command units (BCUs) each containing a certain number of boroughs. BCUs were added to the dataset by mapping each borough to the BCU it belongs to (Wikipedia, 2025). Precision in terms of spatial data is down to the Lower layer Super Output Area (LSOA) level. Each borough has multiple LSOAs and the dataset contains crime data on all of them. In terms of temporal precision, the dataset contains crime counts on a monthly interval.

Using pandas, the dataset was checked for missing values and duplicate rows, none were found. The distribution of the monthly crime counts was evaluated using the mean and standard deviations and no outliers were detected. There is no gap in the temporal coverage from 04/2010 to 07/2021. The dataset contains crime data of all 32 boroughs of London and within each borough, on each LSOA. There are 4829 LSOAs across all the boroughs. There are 10 major categories of crime and each category has its own minor categories. Fig 1 shows the crime count of each major category across all the boroughs and all the months.

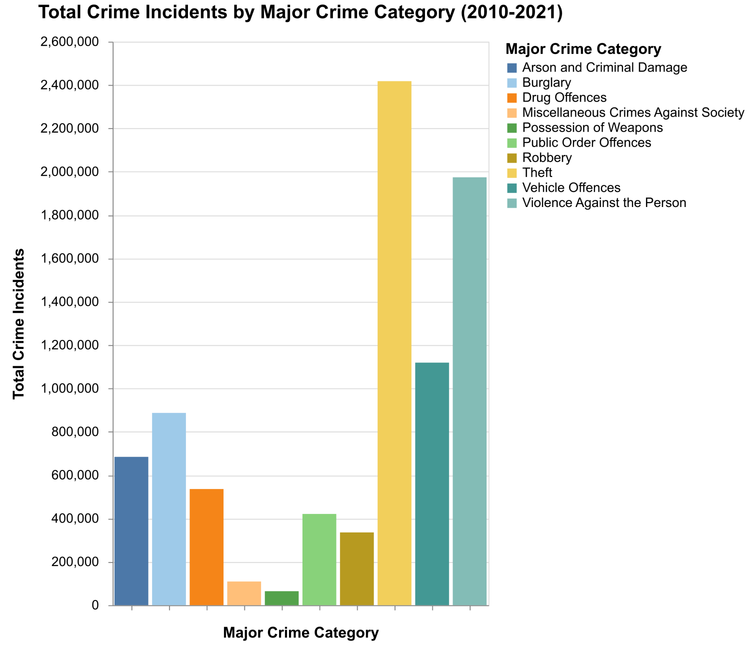

Fig. 1. Total crime incidents of each major crime category across all boroughs (2010-2021)

Fig. 1 shows that the crime incidents are not evenly distributed across the categories. 'Theft' has the highest count while 'Possession of Weapons' has the least. This information will be essential while comparing the spatial distribution of different crime categories across the city. In order to check how the MPS BCUs are spatially distributed across the city, they are plotted on a choropleth map of London with borough and LSOAs boundaries as shown in Fig 2.

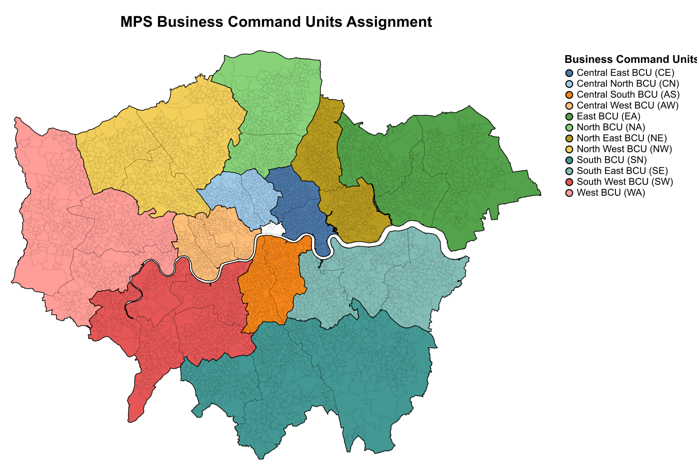

Fig. 2. MPS BCU assignments

It can be seen that not all BCUs contain the same number of boroughs; they either consist of 2, 3 or 4. The BCU sizes in central London are small and they increase in size as we move further away. This is because the spatial size of the boroughs also increases as we move further away from central London. Hence when comparing boroughs, the crime counts will have to be normalised by the borough population for a fair comparison. The map shows that the City of London is not included in any BCU since it is not a borough and hence is not policed by the MPS, thus it will not be a part of this study.

4 ANALYSIS

4.1 Approach

The analytical approach consists of a series of sequential steps, each serving an essential purpose as mentioned in the workflow diagram.

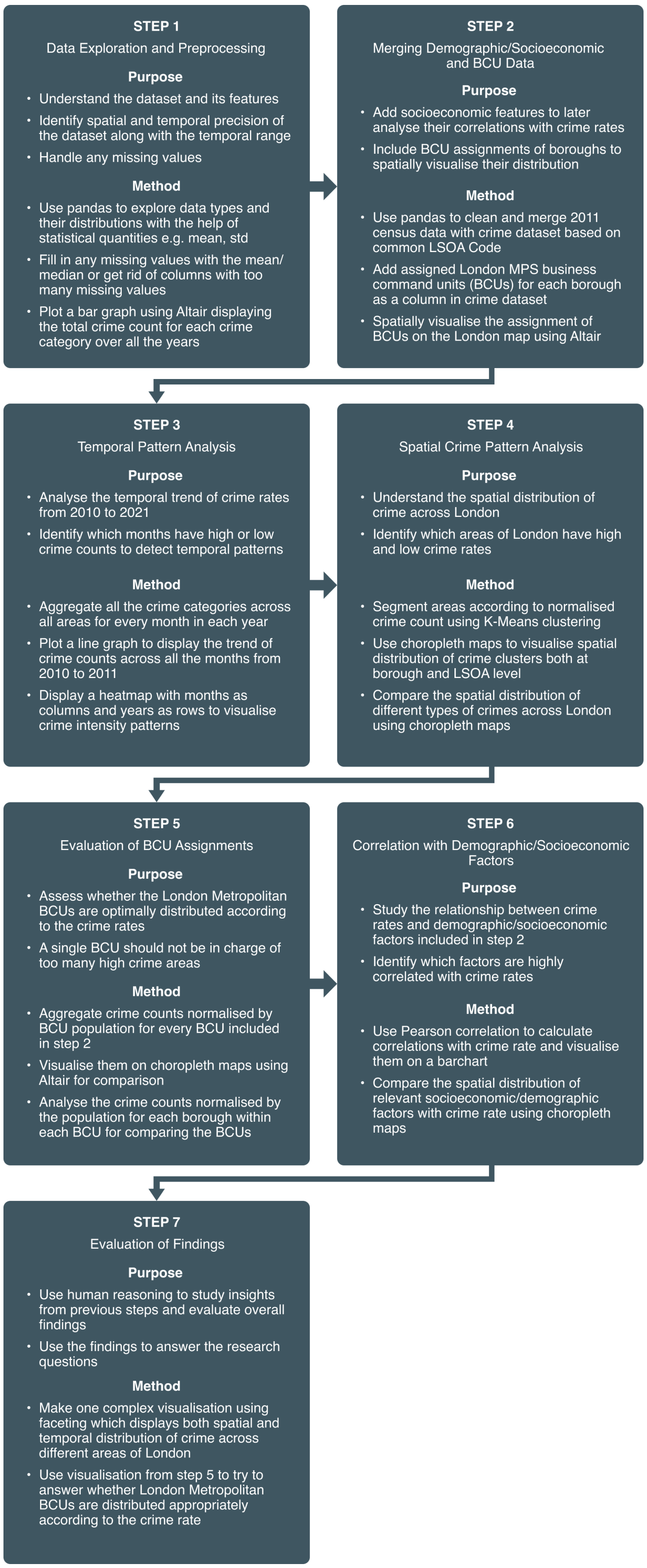

Fig. 3. Analytical approach workflow diagram

Initially data exploration is carried out where the features of the dataset are understood along with their distributions using human reasoning to analyse the mean and standard deviations. This also includes handling any missing values and noting the spatial and temporal precisions. This is followed by merging the crime dataset with the 2011 London census data using borough names to get access to demographic/socioeconomic factors. Before doing so, the census data was cleaned by converting relevant features to numerical. MPS BCUs are mapped to their respective boroughs and their spatial distribution across the city is analysed using a choropleth map mentioned earlier for which human reasoning is essential. In the next step, temporal analysis is carried out using the cleaned dataset from step 2. Crime counts across all categories and boroughs are aggregated for each month and displayed using a line graph to check for temporal trends across the years. They are also displayed using a heatmap to analyse any seasonal crime intensity patterns. Human reasoning is used to analyse both graphs and extract existing temporal trends.

In step 4, spatial analysis is done using the dataset from step 2 to analyse the spatial pattern of crime in London. For this purpose, K-Means clustering is implemented using the crime count normalised by the population to group together boroughs with similar crime rates. These clusters are displayed on a choropleth map where they are coloured according to the average crime rate of the cluster. Human reasoning is used to understand the spatial pattern of crime and identify the boroughs with low and high crime rates with the help of the colour intensity of the clusters. Clustering is also done at the LSOA level to identify crime hotspots, and also for different crime categories to analyse the difference in spatial patterns for each category, for which human reasoning is essential.

Step 5 uses the BCU assignment from step 2 for comparing the crime rates in each BCU. Crime counts of each BCU normalised by its population are displayed on a choropleth map using a sequential colour scheme. Human reasoning is required to comprehend the link between the colour intensity and the crime rate for comparing the BCUs. Moreover, the crime rates of each borough within each BCU are displayed using a choropleth map to understand the spatial distribution of crime within each BCU. Step 6 calculates the correlations between the demographic/socioeconomic factors from the census data to the crime rates in each borough. Correlations are displayed on a barchart using a diverging colour scheme. Boroughs are clustered using the feature with the highest correlation and the spatial distribution is compared with that of the crime rate.

Results are presented using a faceted graph displaying the how the spatial pattern of crime evolves over the years. The choropleth map of BCU crime rates from step 5 is used to evaluate their assignments.

4.2 Process

In order to check what time of the year over all the years experiences the highest number of crimes across London, the crime counts across all major categories and boroughs were aggregated for each month and plotted on a line graph. Individual line graphs for each major category were also plotted for comparison.

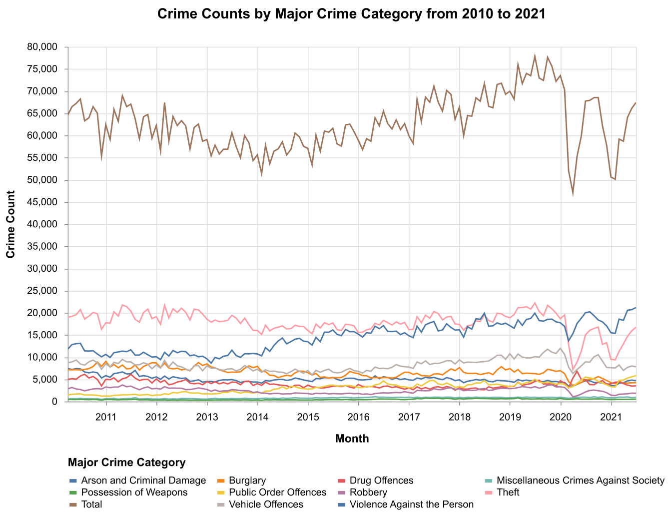

Fig. 4. Line chart of crime counts across all major categories

Fig. 4 shows the total crime counts across all major categories. Using human reasoning shows that there is a slight downwards trend from 2010 to 2014 after which there is an upwards trend till the end of 2019. At the beginning of 2020 there is a significant dip which may be due to the Covid-19 lockdown restrictions resulting in lesser crimes as people stayed in their own homes. Theft has a similar trend whereas Arson and Criminal Damage has a slight upwards trend after 2014. Since seasonal patterns are not apparent in this visualisation, a heatmap is made to visualise crime intensities across all the months from 2010-2021.

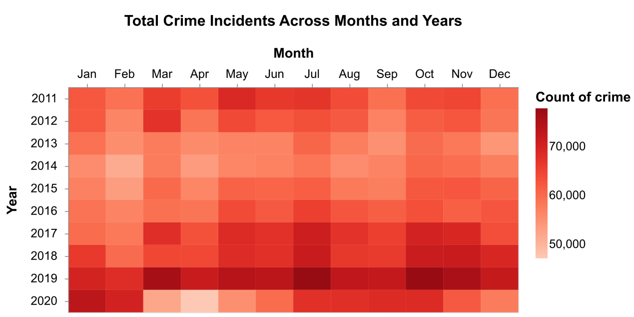

Fig. 5. Heatmap of monthly crime counts from 2010-2021

Human reasoning is used to check for seasonal patterns using the intensity of the colour. In 2011, May had the highest crime rate followed by March in 2012. From 2017 to 2019, July, October and November had high crime rates accompanied by March in 2019. In 2020, January had the highest count of crime. Hence it can be seen that there is no apparent seasonal pattern. Across all the years, no specific time of the year has a higher crime rate however, from 2010-2021 the year 2019 had the greatest number of crimes followed by a dip probably due to the Covid-19 lockdown.

To visually analyse the spatial distribution of crime, instead of directly comparing the aggregated count of crimes of different areas with each other, the aggregated count of crimes of an area is divided by its population divided by 1000. This gives the crime rate per 1000 people of that area resulting in a fair comparison since each area has a different population. Mathematically this is represented as:

$$Crime\ rate\ per\ 1000\ people = \frac{AggregatedCrimeCount}{AreaPopulation/1000}$$

When comparing boroughs, AggregatedCrimeCount is the total number of crimes in that borough and AreaPopulation is the population of that borough. Hence the crime rate per 1000 people is calculated for all the boroughs and this is used for K-Means clustering to group together boroughs with similar crime rates in order to get the spatial pattern of crime. The parameter k i.e., the optimal number of clusters is validated using silhouette scores and silhouette plots for which human reasoning is essential in order to determine what number of clusters gives the highest silhouette score and balanced thickness of silhouette plots. The spatial pattern of crime across all months and categories is first visualised at the borough level followed by at the LSOA level where the clusters are coloured according to the average crime rate of the cluster.

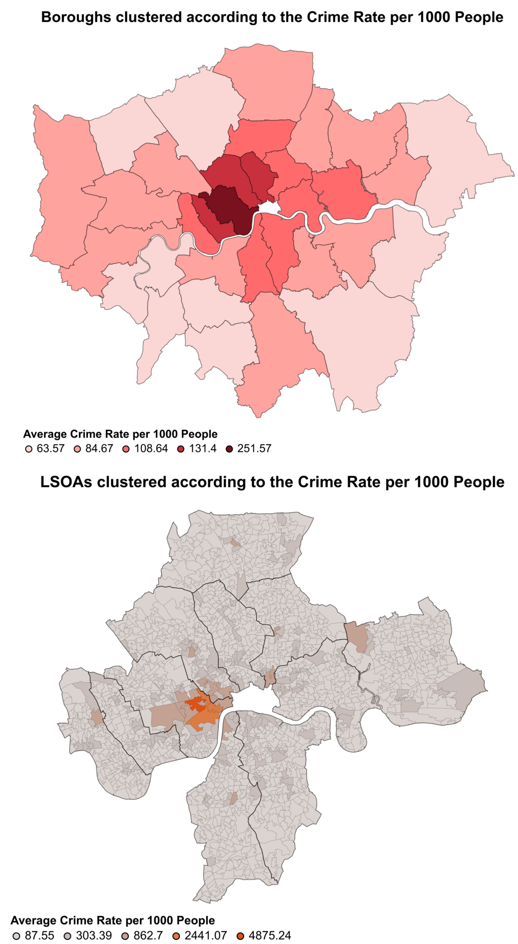

Fig. 6. Spatial pattern of crime at borough and LSOA levels

It can be seen that in the case of boroughs, the average crime rate of the clusters increases as we get closer to central London with Westminster having the highest crime rate of 251.57 crimes per 1000 people, whereas the peripheries have lower crime rates. This is inferred by human reasoning by looking at how the colour varies. The reason for this may be that areas in central London are more affluent hence being a more attractive target for criminals. Repeating K-Means clustering at the LSOA level (and validating the number of clusters using silhouette plots) for only the top three clusters with the higher crime rates reveals a hotspot with a crime rate of 4875.24. Using the LSOA code, this area is traced back to a section of Oxford Street, Regent Street and New Bond Street. Human reasoning was required in order to make this identification and link it to the fact that these streets are filled with tourists and shoppers posing more opportunities for criminals. K-Means clustering is then performed for each category of crime across all the months to analyse how their individual spatial patterns vary as they may differ from the combined count.

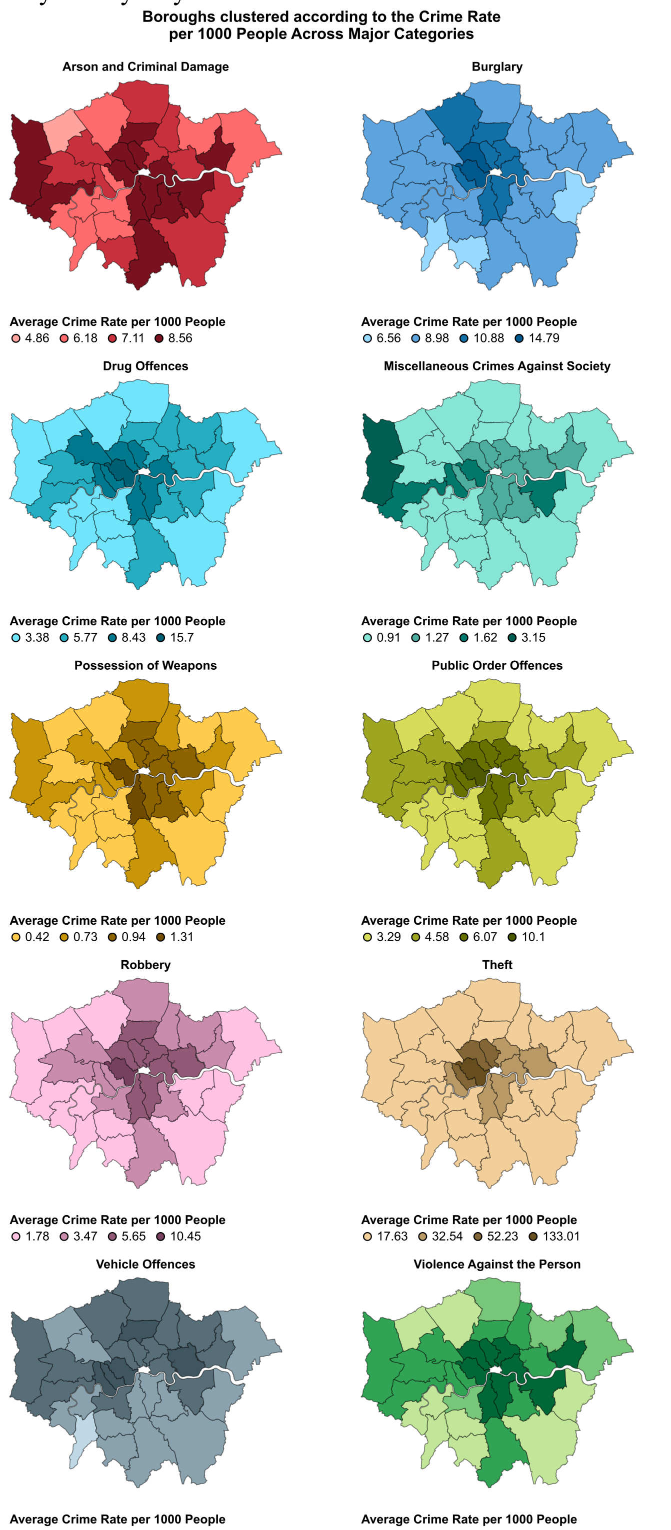

Fig. 7. Spatial pattern of crime across major categories

The parameter k i.e., number of clusters is validated using silhouette scores and plots. For most of the major categories such as Burglary, Drug Offences, Possession of Weapons, Public Order Offences, Robbery, Theft and Violence Against the Person the crime rate increases as we get closer to central London. However, for Arson and Criminal Damage along with some parts of central London, boroughs in the west and south also have a high crime rate. Miscellaneous Crimes Against Society has the highest crime rate in Hillingdon whereas for Vehicle Offences, the crime rate is higher in north and west London. Human reasoning is essential in order identify these spatial patterns.

In order to assess the assignment of the MPS BCUs that were added in step 2, and check if a single BCU is overburdened by being in charge of multiple high crime areas, the crime rate per 1000 people is calculated for all the boroughs and displayed on a choropleth map. The colour of the borough corresponds to the crime rate. The map contains BCU boundaries to help identify which group of boroughs make up one BCU and hence use human reasoning to judge which BCUs have high crime areas.

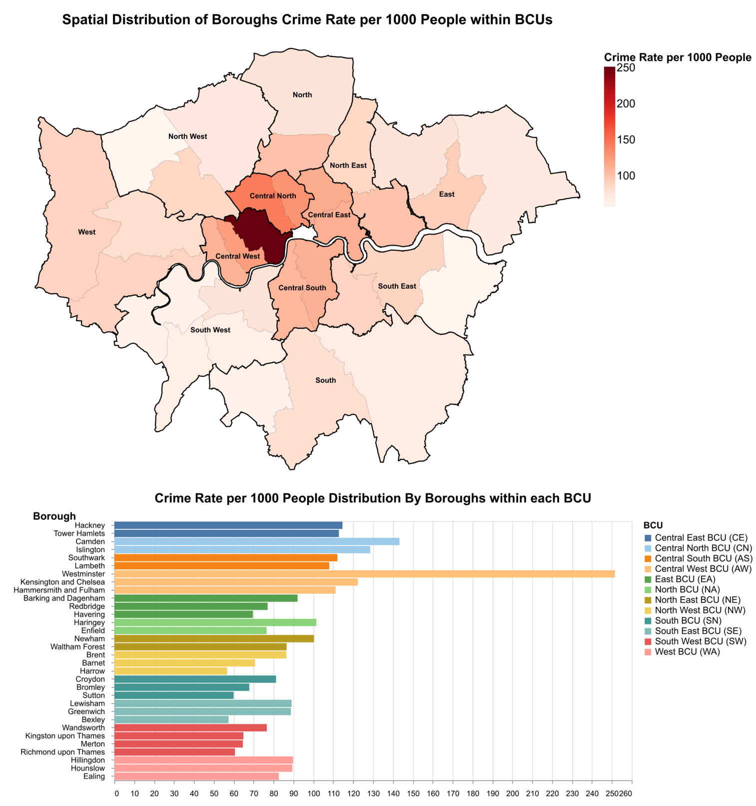

Fig. 8. Borough crime rate distribution within BCU assignments

The map shows the BCUs in central London have areas with higher crime rates as compared to the BCUs on the outskirts. However, with the help of human reasoning it can be seen that spatially the BCUs in central London are smaller in size which means that the area they have to cover is lesser than the BCUs further away from central London but having less area does not necessarily mean lesser population to account for as boroughs in central London are population dense. The accompanying barchart helps to analyse the crime rates of the boroughs within each BCU in more detail. It shows that the Central West BCU has the borough with the highest crime rate i.e., Westminster, as can also be seen on the map. This is followed by Camden and Islington which make up the Central North BCU. Hence central London which has areas with higher crime rates is also divided into more BCUs to account for the greater population density whereas as we move further away from central, the number of BCUs reduce and they become larger in size as the areas that they are in charge of have lower crime rates.

The correlations between the borough crime rates and various demographic/socioeconomic features added in step 2 via the census data are calculated to explore any existing relationships.

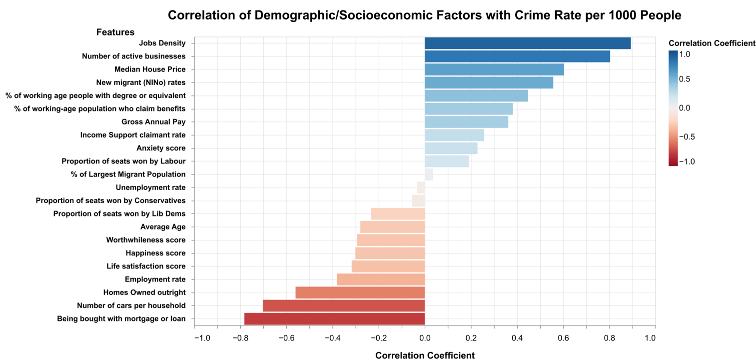

Fig. 9. Correlation analysis between demographic/socioeconomic factors with crime rate

The figure shows that Jobs Density has the highest positive correlation followed by Number of active businesses. This does make sense as boroughs where more people commute to work to and have a greater number of offices, present a greater opportunity for criminals to take advantage of people. Residences being bought with mortgage or loans has the highest negative correlation with crime rate followed by number of cars per household and houses owned outright. This may be because houses in central London where the crime rates are higher, are more expensive as compared to the peripheries hence people prefer to rent as opposed to buying them (Benham & Reeves, 2022). Moreover, the public transport network in central London is well-connected and people live closer to their workplaces which is why they own fewer cars as opposed to those living further away thus resulting in a negative correlation with crime rate as central London has higher crime rate areas. In order to compare the spatial pattern of Jobs Density with crime rate, K-Means clustering is performed on Jobs Density and the clusters are displayed on a choropleth map where they are coloured according to the average Jobs Density of the cluster.

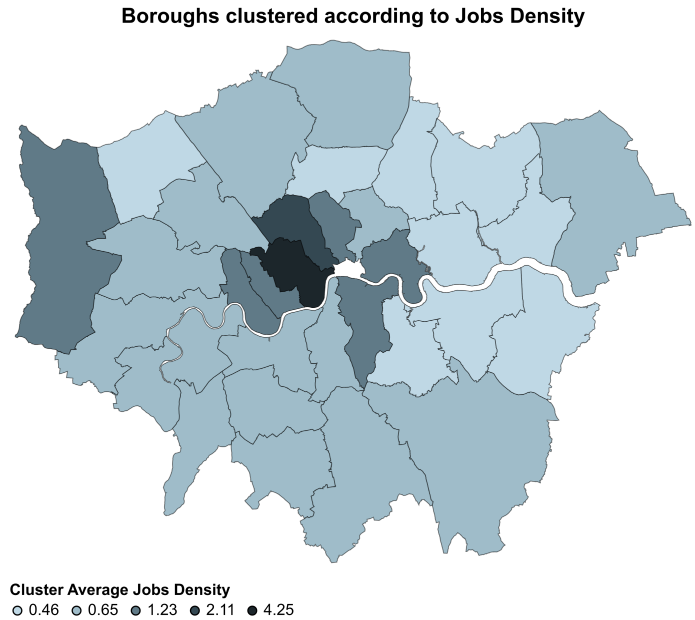

Fig. 10. Boroughs clustered according to their Jobs Density

Comparing the spatial pattern of Jobs Density with that of crime rates in Fig. 6 shows certain similarities. The values of both features are higher in central London and decrease as the boroughs get further away from the city centre. Using human reasoning, this can be explained by the fact that central London has many office buildings both in the private and public sector where majority of the people commute to work to. It also has parliament offices for government officials. Supporting infrastructure such as high-end restaurants and shopping outlets for luxury brands makes the area more affluent as compared to the peripheries. This results in a greater number of well-earning individuals spending most of their day in the city centre, and becoming a target for criminals leading to a higher crime rate.

4.3 Results

This figure examines the spatiotemporal distribution of crime in different times of the year as shown below.

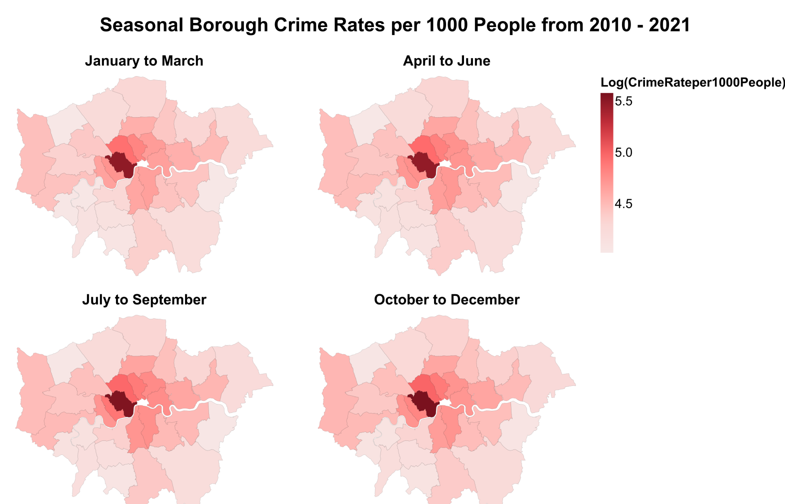

Fig. 11. Crime rate per 1000 People faceted over the yearly quarters

Since there is a huge difference between the highest and lowest crime rate, its logarithm is used to colour boroughs to enhance interpretability. For the first research question, it can be seen that crime rates for all boroughs stay somewhat constant across the four quarters thus there is no specific time of the year which experiences higher crime rates. For the second question, the spatial pattern of crime stays consistent with crime rates increasing as we move towards central London. Westminster has the highest crime rate.

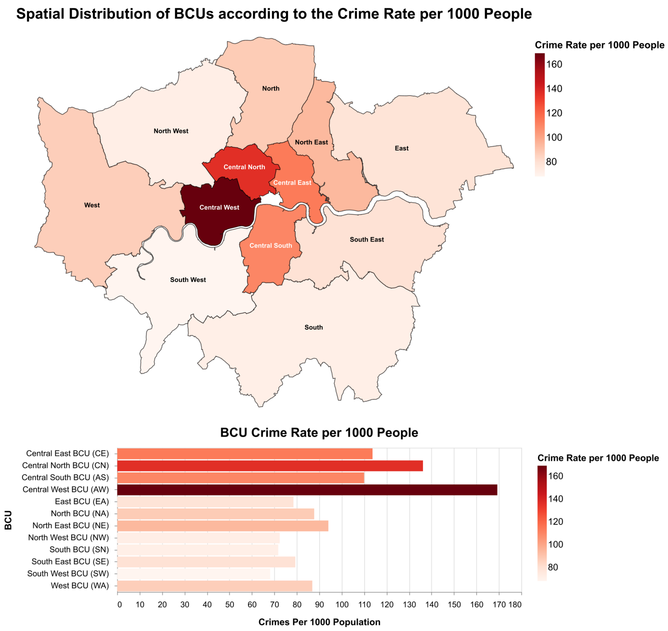

Fig. 12. Spatial distribution of BCU crime rates.

For the third question, Fig.12 shows BCUs in central London deal with higher crimes rates but their areas are smaller allowing them to focus their resources on high crime areas. It can be said that up till a certain extent, BCUs are distributed appropriately according to the crime rate but more information on their resources is needed for a complete answer. Lastly, Fig.9 answers the fourth question as it shows jobs density and number of active businesses have strong positive correlations with crime rate, whereas houses being bought with mortgage and number of cars have strong negative correlations with crime rate.

5 Critical reflection

The goal of this study was to carry out a robust spatiotemporal analysis of crime in London to better inform police forces for an effective allocation of their resources. However, there were some limitations which acted as learning curves. In temporal analysis, a line graph was made to visually analyse the monthly crime counts across the years however, seasonal trends were difficult to identify in this visualisation since the individual months were not labelled. A suggestion to other analysts: a heatmap of the monthly counts with months and years on the axes made it easier to identify times of the year with higher crime rates. The temporal precision of the data was only till the month level. Having the exact date and timestamp of the crime incident would have allowed for a more thorough analysis by applying kernel density estimation (KDE) to identify hours during the day or night when the crime rate is highest for different parts of the city.

For exploring the spatial pattern of crime in London, K-Means clustering was performed using crime rates to group together areas with similar crime rates so that the spatial pattern is more apparent when visualised on a choropleth map. Clustering at the borough level presented a clear spatial pattern however, clustering at the LSOA level resulted in a confusing choropleth map with no apparent spatial pattern due to the immense number of datapoints which served as a lesson learnt. Hence clustering at the LSOA level was done only on the LSOAs belonging to high crime rate boroughs identified earlier, to visualise hotspots; a suggestion to analysts whose datasets do not have exact geocoordinates. Precision of the spatial data was limited to the LSOA level hence hotspots could only be identified to the nearest LSOA. It would have been preferable if exact geocoordinates of crime incidents were available as it would allow for the identification of hotspots down to the street level using KDE. This implementation along with a software with the ability to zoom in to the street level would make the visual analysis more precise by allowing police officials to identify exact street locations with high crime rates.

The distribution of BCUs according to their crime rates was justified partially however, for a complete answer it is essential to have data regarding the resources allocated to each BCU which was not available in this study. Having access to demographic/socioeconomic factors at the LSOA level would have resulted in more precise and robust correlation calculations. It is also important to understand that this study explores correlation and correlation does not equate to causation.

Similar research questions along with the visualisation and computational methods used in this study can be generalised to other spatiotemporal datasets that lack precision. The data used and methods applied are appropriate for a partially detailed analysis of spatiotemporal data however for exact, precise answers it is necessary for the dataset to be more precise with its datapoints so that an in-depth analysis can be carried out.

Table of word counts

Problem statement	250/250
State of the art	500/500
Properties of the data	498/500
Analysis: Approach	500/500
Analysis: Process	1424/1500
Analysis: Results	204/200
Critical reflection	500/500

References

[1] *Crime, N.* (2024) *Oxford English Dictionary*. Available at: https://doi.org/10.1093/OED/5425155625

[2] *MPs recorded crime: Geographic breakdown* (2017) *London Datastore News*. Metropolitan Police Service. Available at: https://data.london.gov.uk/dataset/recorded_crime_summary

[3] *Organisation and structure of the metropolitan police* (2025) *Wikipedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Organisation_and_structure_of_the_Metropolitan_Police

[4] *London Borough Profiles*. GLA. Available at: https://londondatastore-upload.s3.amazonaws.com/instant-atlas/borough-profiles/atlas.html

[5] Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., Cleveland, W. S., Grannis, S. J. and Ebert, D. S. (2010) 'A visual analytics approach to understanding spatiotemporal hotspots', *IEEE Transactions on Visualization and Computer Graphics*, 16(2), pp. 205–220. doi: 10.1109/tvcg.2009.100.

[6] Garcia, G., Silveira, J., Poco, J., Paiva, A., Nery, M. B., Silva, C. T., Adorno, S. and Nonato, L. G. (2021) 'Crimanalyzer: Understanding crime patterns in São Paulo', *IEEE Transactions on Visualization and Computer Graphics*, 27(4), pp. 2313–2328. doi: 10.1109/tvcg.2019.2947515.

[7] Joshi, A., Sabitha, A. S. and Choudhury, T. (2017) 'Crime analysis using K-means clustering', *2017 3rd International Conference on Computational Intelligence and Networks (CINE)*, pp. 33–39. doi: 10.1109/cine.2017.23.

[8] Benham & Reeves (2022) *London's most expensive homes are 30 minutes away from the city centre*, *Benham and Reeves*. Benham and Reeves. Available at: https://www.benhams.com/press-release/buying/londons-most-expensive-homes-are-30mins-away-from-the-city-centre/