Crime in London: Searching for Strategies

Benedict Collins
Department of Computer Science
City, University of London
London, United Kingdom
benny.collins@city.ac.uk

Abstract—**It is suggested by the results that the number of Metropolitan Police does not have much of an effect on crime levels, whereas the percentage of the population that are homeless, population density, the percentage of population with qualification level NVQ1 or less, as well as the percentage of households that are workless all display higher levels of influence on crime levels in a borough. This identifies several socioeconomic factors that should be prioritised for government spending, as opposed to increasing Metropolitan Police numbers.**

There are implications that crime prevention techniques should generally be split between Outer and Inner London, because of tourism and homelessness in Inner London. Crime in these areas could be reduced through public information campaigns, helping visitors stay aware of pickpockets.
In the years 2019-2021, the proportion of crimes categorised as 'Violence Against the Person' increased significantly, along with the rates of violent crimes in the city.
. (*Abstract*)

Keywords—component, formatting, style, styling, insert (*key words*)

I. Introduction (*Heading 1*)

In a world of increasingly polarised politics and high levels of access to independent sources of information across the internet, the argument surrounding the key causes of crime is more prevalent than ever. In 2019, London saw its highest number of homicides in 10 years [1], with figures showing a “surge in knife and gang-related killings since 2014” [1].
The austerity measures that have been implemented in the UK, since the 2008 recession, are often spoken about as having a significant effect on crime levels in the UK. This is due to the association of austerity with depreciating quality of life for many of the disadvantaged members of society. In 2020, the UK’s corporate income tax was one of the lowest in Europe (19%), and many argue that policies such as these are reducing government funding, which in turn reduces the amount of funding available for public services such as community centres, schools, affordable housing, and healthcare. After a decade of austerity, “spending on welfare benefits for the UK’s poorest families will have shrunk by nearly a quarter” [2], resulting in a reduction of the quality of life for many of these families.
In response to the increasing number of homicides and violent crimes, the Serious Violence Strategy, published by the Home Office in 2018, asserts that “tackling serious violence is not a law enforcement issue alone” [3], and the strategy set out in the document, will focus on “a number of sectors such as education, health, social services, housing, youth services, and victim services” [3]. This displays a shift from the focus on “character traits in individuals” [4] seen in the Home Offices’ 2016 Modern Crime Prevention Strategy, to a more generalised, area and demographic-based analysis of the causes of crime, where public services are considered essential to the strategy.

II. Analytical Questions

This project will attempt to provide an overview of crime in London by searching for answers to the following questions:

1. Does the number of people working for the Metropolitan Police have the largest influence on fluctuations in crime levels?
2. Which measurements of quality of life have the largest impact on crime levels?
3. Which London boroughs are seen as being priority areas for crime intervention?
4. What key events and policy decisions may have affected crime levels in the city?
5. What key events and socio-economic factors influence the proportions of types of crime?
6. What aspects of society may be important to prioritise in order to help lower London’s crime rates?

These questions aim to collate ideas for targeted, pre-emptive crime prevention, as well as identifying which areas require greater attention and funding from the relevant government authorities.
The selected measurements of quality of life for this project are income levels for employees, household working status, homelessness, housing affordability, education level and housing density.

III. Data

The Metropolitan Police’s crime datasets for 2008-2018 and 2012-2021 contain the number of monthly crimes in each borough for the different major and minor category crime classifications. This provides us with the opportunity to look at crime category classifications, as well as the total number of crimes for each borough.
The ONS’ population estimates dataset for 2001-2020 contains the yearly population estimates for each borough. This dataset can be combined with our other datasets to obtain per capita, or percentage of population, measurements.
Datasets containing the number of people working for the Metropolitan Police each year are not available by borough, reducing how much insight we can gain on Police numbers’ effects on specified boroughs. However, we can still use it as an estimate of police presence in London boroughs.
Employee income will be analysed using a dataset containing the yearly percentages of people earning below the London Living Wage (LLW) in each borough.
The household working status dataset contains percentages of working households, households with mixed employment status, and workless households. The percentage of workless households has been selected as an attribute, as unemployment in a city as expensive as London has the potential to be detrimental to one's lifestyle.
Homelessness will be analysed using a dataset containing the total number of people sleeping rough, by borough for the years 2010-2020.

Housing affordability will be analysed via a dataset containing the amount of affordable housing made available each financial year for each borough.
Education level is measured using the dataset listing qualification levels of people of working age for each borough in the years 2004-2020. The qualification levels of interest are 'no qualifications' and level ‘NVQ1 qualifications’, described as ‘1-4 GCSEs or equivalent’.
The final dataset is a population density dataset, containing yearly density measurements for each individual borough.

IV. Analysis

Data preparation largely consisted of joining datasets measuring the same things over different periods of time, such as the crime datasets for 2008-2018 and 2012-2021, and the 2007-2009, 2010-2012 and 2013-2021 datasets for the number of people working for the Metropolitan Police. The time periods of the datasets were all different, so the time period had to be aligned, leaving us with the years 2010-2018. Crime datasets containing measurements outside of these years were maintained and left for use in analysis of crime category proportions, as these have been heavily affected by the coronavirus pandemic, as well as time series analysis.
The final datasets used for the rest of the analysis were created by collecting borough DataFrames, containing the measurements for each variable, independent and dependent, over the years 2010-2018. These separate borough DataFrames were then concatenated, resulting in a DataFrame where the columns list the variables and the rows list of all the yearly values of said variables, over all the boroughs.
The heatmaps and k-means clustering algorithms required a DataFrame containing single, summary values for each variable in each of the boroughs, which act as row indices. These values were obtained for each borough by summing the values of all raw number variables (e.g. number of crimes), and taking the mean of any proportion-based variables (i.e. percentages, per capita and density measurements).
The final DataFrame that was defined lists the variables for London as a whole, over the specified years. This DataFrame is for use in the time series analysis.

Data derivation for this project predominantly involved using our population estimates dataset to calculate per capita variable equivalents for crime, Metropolitan Police workforce and affordable housing made available yearly. The same population data was used to acquire the percentage of population variable equivalents for the number of people sleeping rough and the number of people of working age with qualification level NVQ1 or less. Datasets that displayed monthly measurements, such as our crime datasets, had to be adjusted to yearly measurements, due to all the independent variable datasets showing yearly measurements, apart from Metropolitan Police workforce. This was done by summing values across the months of a year.
The DataFrame showing the percentage of people of working age with qualification level NVQ1 or less was derived by creating two separate DataFrames, from the original dataset, via filtering. One DataFrame for the number of people with ‘no qualifications’, and one for the number of people with ‘NVQ1 or equivalent’ were both obtained and then added together in order to obtain the DataFrame for the number of people with qualification level NVQ1 or less. This was then combined with our population DataFrame to obtain the percentage of population equivalent of this count.
The London DataFrame for population density had to be calculated by summing the population over all London boroughs, then dividing by the total square km covered by the city. This was done for each year in order to ascertain yearly London population density measurements.

K-means was picked as a suitable model to group together crime prevention priority boroughs, as well as identifying unifying features of boroughs that could contribute to contrasts in crime rates between different clusters. Three clustering algorithms were run. One split was based on borough values for independent variables, and another was split on crime category proportions. The remaining algorithm was split on crime per capita values, however this final split was applied to a dataset excluding Westminster due to its dominance of this particular measure. The elbow method was applied to each set of variables, in order to estimate the optimal value of k for each model. The results for these elbow plots were not certain, so the silhouette method was employed, with the silhouette score acting as a goodness-of-fit metric that compares cohesion and separation in the model. Each model was assigned an optimal k value of 2. Annual measurements for each borough in each cluster were concatenated into separate cluster DataFrames for further analysis. The validation of the different clusters was done by comparing borough maps, coloured according to their clusters, with borough heatmaps for each different variable, in order to search for any meaningful patterns.
Hyperparameter tuning for the random forest models involved using cross-validation testing over 500 random combinations of different specified input values for the hyperparameters maximum tree depth, maximum number of features to consider at each split, minimum number of samples to consider at each split, minimum number of samples that can be contained in each leaf, and the number of decision trees to build. Once the optimised hyperparameters have been acquired, we use them to construct a function that runs a train/test split a DataFrame, then trains and tests a random forest regressor on the respective DataFrames. The function then prints the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and the Mean Squared Error (MSE), which act as validation metrics. The main metric used to compare the models' performances is MAPE, as it allows for comparison of models for predicting values on different scales, in this case crime number and crimes per capita. The permuted predictor importance estimates for the attributes in each model should provide an understanding of the extent to which the different attributes affect crime levels in the area covered by the model. Although it won't tell us whether the relationship between the attribute and target variable is positive or negative, it will tell us the extent to which the attribute can be used to predict crime level.

Further analysis of crime categories and total crime levels for each borough, and London as a whole, will be done using pie charts and time series analysis. This should provide insight into how crime levels, and types of crime, have changed over time.

V. Findings, Reflections and Further Work

Our results suggest that the number of Metropolitan Police per capita doesn't have much of an effect on crime levels. This can be seen in the low Pearson correlation coefficients when comparing this variable to total crimes and crimes per capita. This assertion is further supported by the permuted predictor importance estimates for our different random forest models. The importance estimate for police per capita is one of the lowest importance estimates across our 14 random forest models.

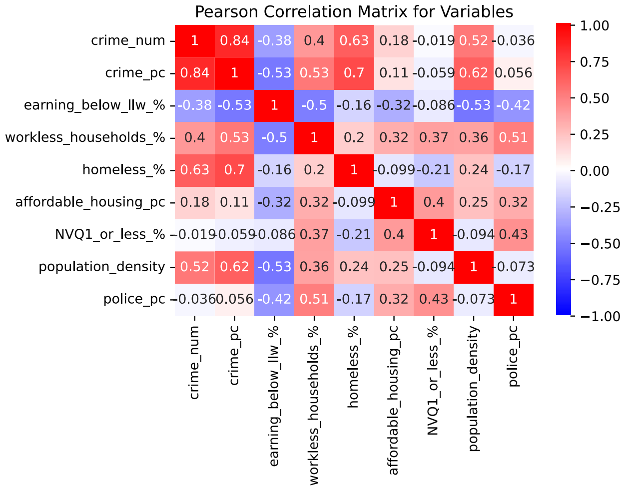

Fig. 1. Pearson correlation coefficient matrix.

The variables that provided the most predictive power to the random forests, suggesting higher importance in affecting crime levels, were the percentage of the population that are homeless, and population density. We also see reasonable levels of predictor importance for the percentage of population with qualification level NVQ1 or less, as well as the percentage of households that are workless. This implies that education, homelessness and employment are areas for governments to prioritise spending.

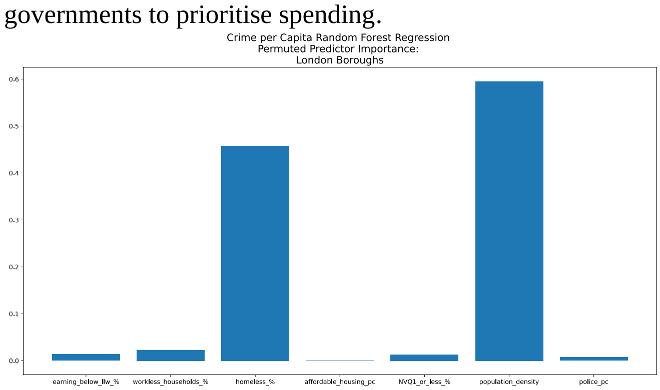

Fig. 2. Permuted predictor importance estimates for random forest applied to all borough measurements.

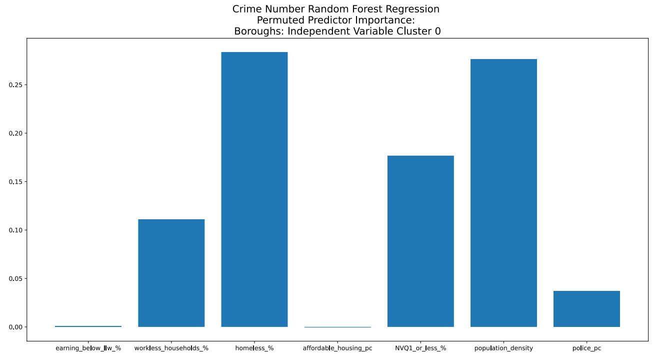

Fig. 3. Permuted predictor importance estimates for random forest applied to all boroughs in the independent variable cluster 0.

The k-means clusters illustrate a clear pattern. Clustering based on independent variables has grouped the Inner London boroughs together, as well as Haringey. These boroughs are considered to have similar characteristics based on the independent variables selected for this project, a proposition confirmed by our heatmaps, which show a rough split between measurements for Inner and Outer London boroughs. This implies that crime prevention techniques should generally be split between Outer and Inner London.

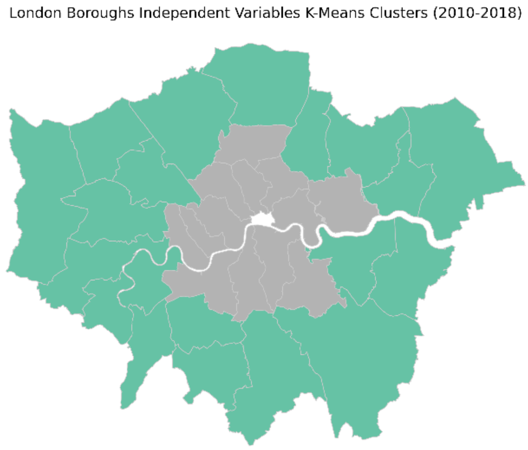

Fig. 4. Independent variable k-means clusters.

Clustering based on crime per capita has grouped together all the Inner London boroughs, except Wandsworth and Lewisham. Similarly to the last set of clusters, Haringey has been included in the cluster with the Inner London boroughs. These boroughs are considered to have similarly high annual crimes per capita, suggesting that these are the boroughs to be prioritised for crime prevention.

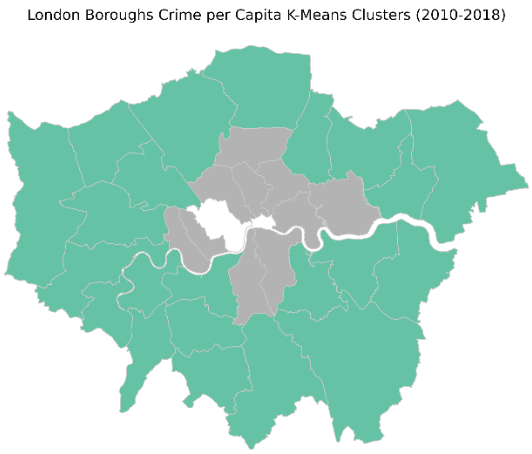

Fig. 5. Crime per capita (excluding Westminster) k-means clusters.

Clustering based on crime category proportions grouped the boroughs Hammersmith & Fulham, Kensington & Chelsea, Westminster, Camden and Islington; boroughs are often associated with high levels of tourism, which often attracts pickpockets. The pie charts show that, for the years 2008-2018, the proportion of crimes classed as ‘Theft and Handling’ in each of these boroughs was over 45%. This implies crime in these areas could be reduced through public information campaigns, helping visitors stay aware and keep their property safe.

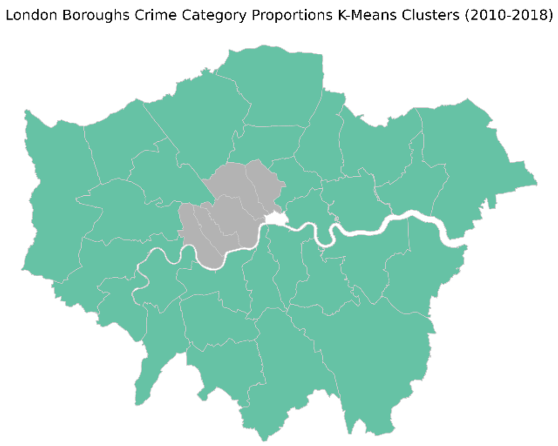

Fig. 6. Crime category proportions k-means clusters.

In the years 2019-2021, the proportion of crimes categorised as 'Violence Against the Person' increased significantly, in line with 2019 headlines. This increase could also possibly be partially explained by the rise in reports of domestic violence in many areas around the world. In the year ending March 2020, the UK and Wales (excluding Manchester) saw an increase of 9% in domestic abuse-related crimes from the previous year [5].

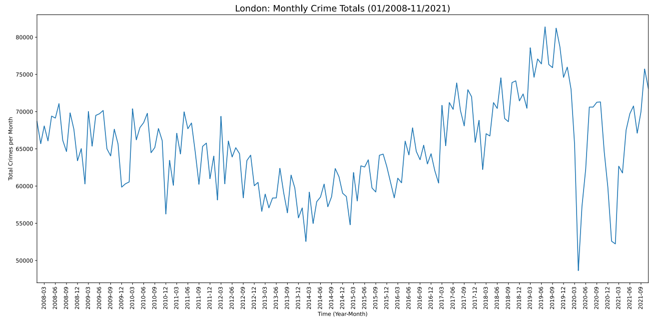

Fig. 7. Timeseries for total number of crimes in London (2008-2021).

Fig. 7. shows huge dips in crime numbers at the start of 2020 and 2021, due to lockdown measures posed by the pandemic. There has been a rise in the number of domestic abuse crimes, committed against women and children, during the pandemic however the increase in these types of crimes being reported will not necessarily yield the same increase in Metropolitan Police reports on such crimes, due to the difficulties posed to victims attempting to receive support. This identifies domestic abuse as a crime category that requires better detection and reporting methods.

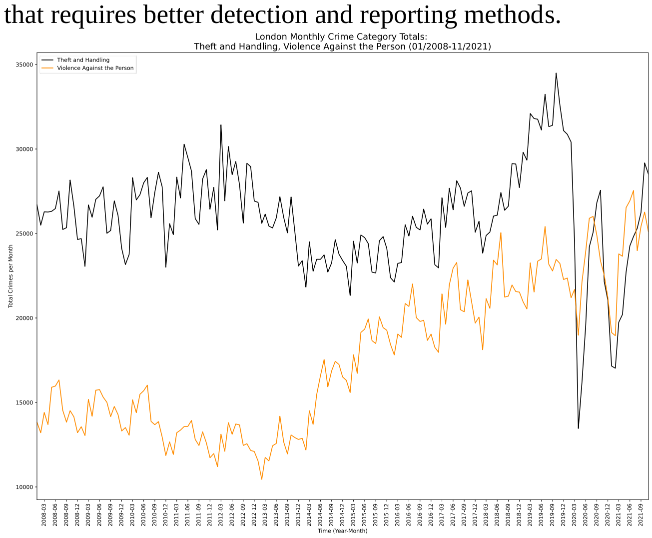

Fig. 8. Timeseries for total number of ‘Theft and Handling’ and ‘Violence Against the Person’ crimes in London (2008-2021).

The efficacy of the results would be greatly improved by data collected more consistently, over a wider timeframe, as low sample sizes didn’t help with modelling. The findings also could have been improved by looking at cuts in benefits and the availability of after school activities for underprivileged students. Further analysis into boroughs individually would also be likely to produce a more in-depth understanding of the importance of the selected features.

References

[1] E. Graham-Harrison *et al.*, “Lockdowns around the world bring rise in domestic violence,” *The guardian*, The Guardian, 28-Mar-2020.K. Elissa, “Title of paper if known,” unpublished.

[2] P. Butler, “Welfare spending for UK’s poorest shrinks by £37bn,” *The guardian*, The Guardian, 23-Sep-2018.

[3] “Serious Violence Strategy,” *Gov.uk*, 09-Apr-2018. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/698009/serious-violence-strategy.pdf. [Accessed: 23-Dec-2021].

[4] "Modern crime prevention strategy," *Gov.uk*, 23-Mar-2016. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/509831/6.1770_Modern_Crime_Prevention_Strategy_final_WEB_version.pdf. [Accessed: 23-Dec-2021].

[5] N. Stripe, "Domestic abuse in England and Wales overview - Office for National Statistics," *Gov.uk*, 25-Nov-2020. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2020. [Accessed: 23-Dec-2021].