Beyond the Crime Rates: Integrating Deprivation Data in Urban Crime Analysis

Dimitris Nikandrou
School of Computer Science, City University of London
Email: Dimitris.Nikandrou@city.ac.uk Student ID: 200013603

ABSTRACT

This study examines the correlation between socio-economic deprivation and crime types and frequencies in London. Utilizing June 2019 crime data from the Metropolitan Police and the Index of Multiple Deprivation (IMD), we applied statistical analyses to identify patterns in crime occurrence across various socio-economic landscapes. Initial findings suggest a complex relationship, with higher deprivation often coinciding with increased crime. However, exceptions and limitations within the data highlight the need for further, more nuanced research. Insights from this work emphasize the significance of context-driven policies to mitigate crime and deprivation effectively.

I. Introduction

The complex relationships between crime-rates and socio-economic factors were the basis of this analytical study. In a period where urban safety has been a hot topic, understanding the dynamics of crime in the context of social and economic deprivation is crucial. This project's goal is to break up and understand these complexities by analyzing data provided by the Metropolitan Police, alongside key deprivation indicators.

There are several reasons that are motivating enough to choose this topic as an area of study. Firstly, such an analysis extends beyond the traditional crime statistics, offering a more variated view of the conditions supporting criminal activities. Secondly, from a practical perspective, this project seeks to inform policy making and law enforcing strategies by indicating areas of deprivation that are more susceptible to criminal activities.

By integrating crime data with deprivation metrics, this project attempts to uncover patterns and correlations that are not immediately apparent. The analysis aims to understand how different socio-economic conditions may influence the nature and frequency of criminal activity in various London localities. This analysis also wants to contribute to the ongoing conversations regarding public safety and social equity and guide targeted interventions to enhance the safety of London's population and wellbeing by addressing the underlying socio-economic challenges that the city faces.

The project's ultimate goal is to contribute to the creation of safer and more equal urban communities in London by providing data-driven solutions to these problems.

II. Analytical Questions

1. How do crime types and frequencies vary across different levels of socio-economic deprivation?

This question aims to explore the relationship between the variety and incidence of crimes and the socio-economic status of different areas. It's crucial for understanding if certain crimes are more common in areas with higher deprivation scores.

2. Does the severity of crimes correlate with different levels of socio-economic deprivation?

This question shifts the focus to the severity or seriousness of crimes in relation to socio-economic factors. The analysis of crime data in conjunction with deprivation indicators aims to determine if areas with higher socio-economic challenges experience more severe types of crime.

3. Does the concentration of crimes differ in areas with distinct socio-economic profiles?

The aim is to analyze whether the density or concentration of crimes (number of crimes per unit area or population) varies significantly in areas with different socio-economic characteristics.

4. What are the common characteristics of areas with high crime rates?

This question aims to identify common socio-economic traits (like high unemployment, low education levels, etc.) in areas with elevated crime rates, offering insights into potential underlying causes.

III. DATA (MATERIALS)

o Dataset Overview

Two datasets were used for the analysis part of this project. The crime dataset was obtained from the Metropolitan Police website for crimes in London for 2019 and more specifically for June 2019. The deprivation dataset was obtained from ArcGIS Online and the dataset was called Index of Multiple Deprivation (2019).

o Key Characteristics

The crime dataset included variables such as Crime ID, Latitude, Longitude, Location, LSOA Code, LSOA Name, Crime Type and Outcome. On the other hand, the deprivation dataset included variables such as LSOA Code, IMD Score, Inc Score, Emp Score, Edu Score, ASS Score, Total Population etc., for all of England and hence it was aggregated on the LSOA Codes of the crime data so it could be merged with it.

o Data Pre-processing

The crime dataset was aggregated to the LSOA level and transformed (pivot wide) to allow for more efficient data manipulation. The new pivoted dataset included the various types of crimes along with their frequencies as well as a new Total Crime column, which was a sum for all crimes in a particular LSOA Code. The deprivation dataset was aggregated with the pivoted crime data, on the LSOA Code and some deprivation metrics were removed from the dataset. The new merged dataset included the crime types and their frequency, total crimes in each LSOA Code and Borough, Total Population, and the deprivation scores.

o Assumptions & Limitations

A few assumptions that were made when using these two datasets, were, but not limited to: Geographical Consistency, meaning that it was assumed that the LSOA Codes that were aggregated were consistent in both datasets. Moreover, it's worth stating that the analysis of these two datasets was correlational and does not necessarily imply causation for the relationship between crime types and deprivation indices.

IV. ANALYSIS

As every Data Science project, this study applies the standard pipeline procedure.

A. Data Preparation

1. Data Transformations and Feature Selection

The raw data that was obtained, required some data treatment methods, such as feature selection and transformations. For the crime dataset, the first thing required was to transform the dataset to allow for easier manipulation of the crime types. The dataset was pivoted wide, meaning that the all the crime types along with their frequency were listed next to each LSOA code and an additional column was created, Total Crimes, that summed the total crimes in each LSOA code. Meaning that each locality had its own metrics for both crime types and deprivation, that was integrated in the next step. The deprivation dataset obtained, had some columns removed from it, which were deprivation scores that were not considered as important as other deprivation metrics. For example, the HDD score which is a measure of qualitative food consumption and the BHS Score, which is a measure of the negative attitude for the future. Moreover, the IDC and IDO scores, which are measures of deprivation in children and old people. The deprivation scores that remained in the dataset, were the Income, Employment, Education and Adult Skills, which is a measure of how skilled each adult is considered meaning that he could be flexible in making money. It is worth mentioning that the deprivation dataset, included Crime scores as well which were removed.

2. Merging the Datasets

The two datasets, were then merged with each other according to their LSOA codes, meaning that now each locality had metrics for crime types and deprivation indicators, that allowed for efficient manipulation of the new merged dataset.

B. Data Derivation

1. Computed Data

When transforming the crime dataset, a Total Crimes column was computed by the summation of each crime type in each London locality. Moreover, the Total Population of each locality was aggregated with each unique LSOA code.

2. Insights from Data Combination

By manipulating the two datasets, some insights were obtained for the relationships between the crime types and the deprivation indices.

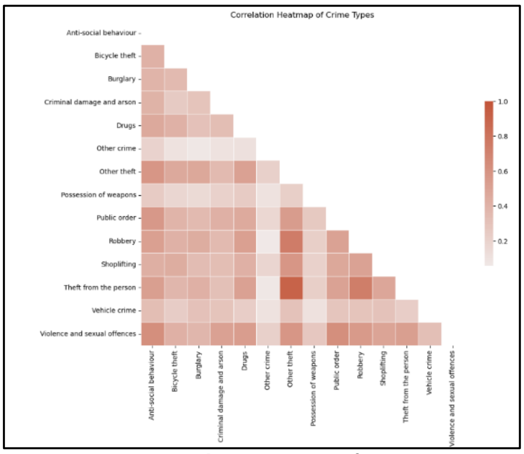

Figure 1: Correlation Heatmap of Crime Types

In the figure above, it can be observed that there is a strong correlation between robbery and theft, and anti-social behaviour and violence and sexual offences, which was expected.

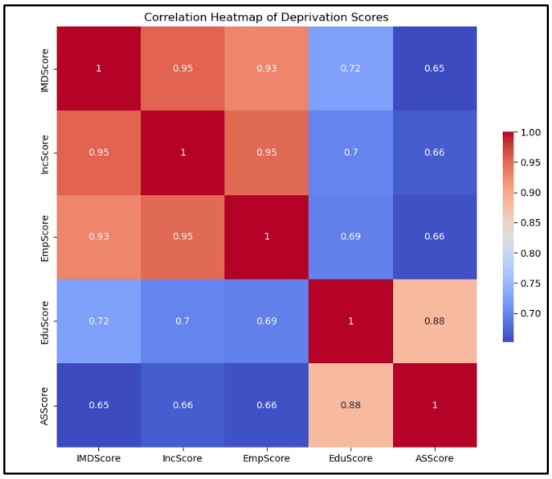

Figure 2: Correlation Heatmap of Deprivation Scores

In the figure above, it can be understood that all deprivation scores are relatively strongly correlated between them. Some, however, stand out for example, employment and income, education, and skill.

C. *Construction of Models*

1. Correlation between Crimes and Deprivation

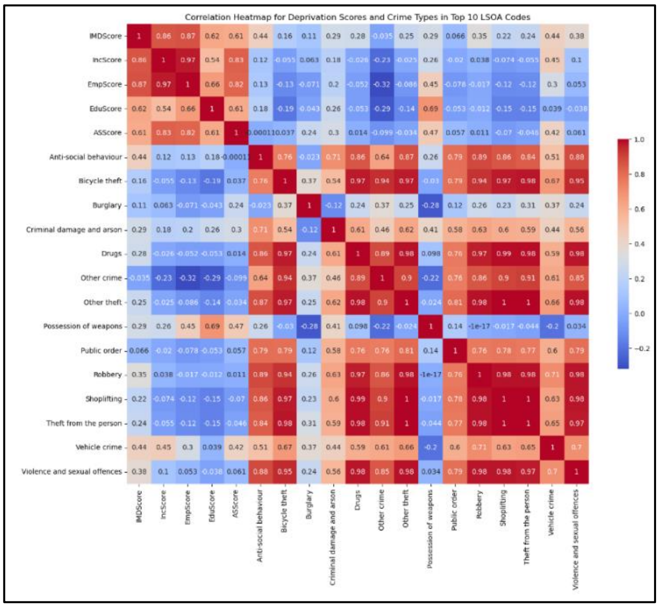

Figure 3: Correlation heatmap between Deprivation Scores and Crime Types in top 10 LSOA codes by population

Having a look at this correlation heatmap, a few interesting insights can be found. For example, there is a strong positive correlation between weapon possession and educational deprivation, moreover it is observed that the overall deprivation score (IMD) is relatively strongly correlated with anti-social behaviour. Another interesting relationship is found between drug possession and various crimes involving theft, robbery, and criminal damage. This indicates that a certain part of the population usually resorts to theft and other similar types of crimes to fund their drug addictions.

2. Population and IMD Score

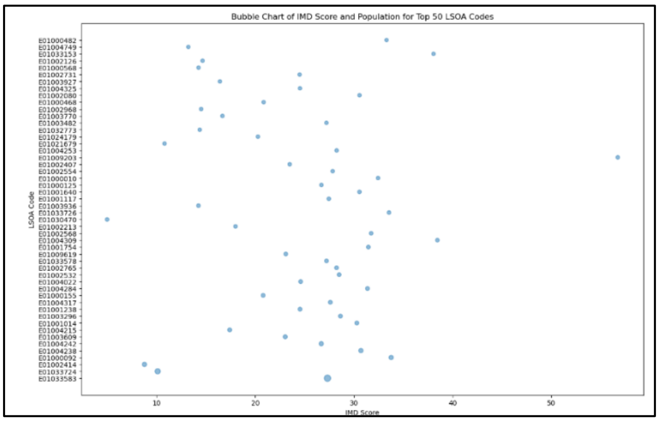

Figure 4: Bubble chart of Population and IMD Score

This bubble chart was created to visualize the overall deprivation of the top 50 most populated LSOA codes. LSOA areas with higher IMD scores, which are further to the right on the x-axis, are those with greater levels of deprivation. The LSOA codes on the y-axis are ranked from most populated to least populated. Even though the hypothesis is that the most populated areas would also be more susceptible to higher IMD Scores, this is not the case since the LSOA code with the higher IMD score is found at around the middle of the population range.

3. Hierarchical Clustering Dendrogram

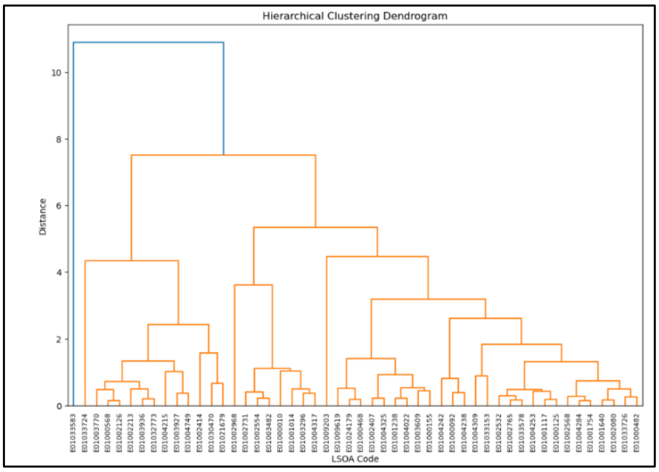

Figure 5: A hierarchical clustering dendrogram

This dendrogram was created by incorporating IMD score, Total Crimes and Total Population. The reason for creating this dendrogram was to determine the number of clusters by cutting the dendrogram at a specific distance. For example, cutting the dendrogram halfway down the y-axis might result in a certain number of clusters based on where the cut intersects the vertical lines. It is often used to understand the data's underlying structure or to decide on the number of clusters to use in further analysis.

D. *Validation of Results*

The reasons behind using these types of visualization and analytical techniques were multiple. As seen in the previous sections of the analysis, correlation heatmaps and calculations were extremely important in trying to answer the analytical questions. This is because, they are one of the most important techniques and methods if someone wants to get a deeper understanding of the complex relationships between crime rates and deprivations. It provides insights not only on the correlation measurements between variables that someone with common sense would assume high correlations, but also on variables that someone could have not guessed that were correlated or showed very little correlation. Moreover, the potential for the application of machine learning algorithms and classifies is emphasized by the creation of the hierarchical dendrogram for deprivation scores and crime rates, that could make the basis of a classifier being applied to the data to make predictions on topics such as potential LSOA codes where deprivation and crime rate might increase, hence giving a head start to law enforcement and policy making.

V. FINDINGS, REFLECTIONS & FURTHER WORK

It was found that there is a pattern where areas with higher socio-economic deprivation also experience different types or frequencies of crimes. This can be seen from the bubble chart where LSOA codes that are more to the right of the chart experience higher levels of deprivation. This can be validated using the figure below:

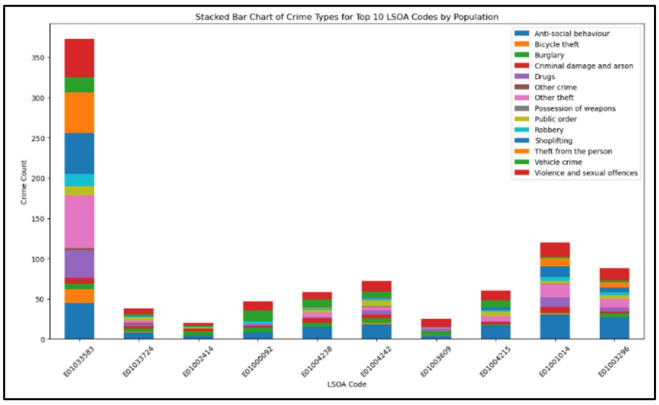

Figure 6: Stacked bar chart of crime types

It can be observed that in these areas, the most predominant crime types are shoplifting, violence and sexual offences and antisocial behavior, which would indicate that deprivation scores such as income and education might be higher than the average.

It can be said that the seriousness of the crimes is not directly correlated with higher deprivation scores, since it was found that the most prominent crimes in London were antisocial behavior and violence and sexual offences (which the proportions between the 2 are unknown). Even though these crimes are serious, they are not considered as severe as weapon possession, burglaries, and drugs.

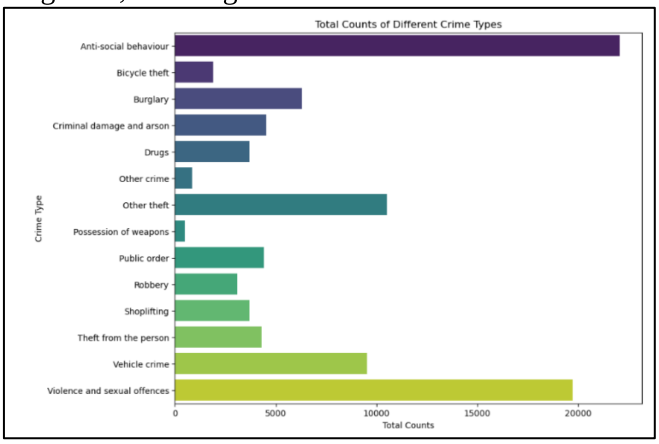

Figure 7: Total Number of different Crime Types

It can be said with confidence that areas that experience high crime rates are very likely to experience increased deprivation scores in the aspects of skill level, employment status, income status and education level. Even though the are strong correlations that support these statements, there are some localities which experience high levels of some crime types, for example burglaries and theft. These areas can be considered as outliers when thinking about the relationship between crime rate and deprivation, since certain wealthier areas can be susceptible to these kinds of crimes from individuals do not reside in them, but instead commute there to commit these offenses.

Reflecting on this study, the link between deprivation and crime is complicated. The analysis showed that areas with higher deprivation often saw more crime, but there were exceptions that need more looking into. The project's limitations, like not having data that would allow for time-series analysis, point out what should be studied next. This work has made it clear how important detailed analysis is for making good policies. It has also shown how much the environment and deprivation impact crime. The findings from this project can help to the addressing of both crime and deprivation, with the goal of making communities safer and fairer for the residents of London.

VI. REFERENCES

[1] Curtis, S.E., 1990. Use of survey data and small area statistics to assess the link between individual morbidity and neighbourhood deprivation. *Journal of Epidemiology and Community Health*, *44*(1), p.62.

[2] Malleson, N. and Andresen, M.A., 2016. Exploring the impact of ambient population measures on London crime hotspots. *Journal of Criminal Justice*, *46*, pp.52-63.

[3] Wang, H., Kifer, D., Graif, C. and Li, Z., 2016, August. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 635-644).

[4] Bradford, B., 2011. Police numbers and crime rates–a rapid evidence review. *HM Inspectorate of Constabulary*, pp.1-10.

[5] Mason, P., Kearns, A. and Livingston, M., 2013. "Safe Going": the influence of crime rates and perceived crime and safety on walking in deprived neighbourhoods. *Social science & medicine*, *91*, pp.15-24.

[6] Tarling, R. and Dennis, R., 2016. Socio-Economic Determinants of Crime Rates: Modelling Local Area Police-Recorded Crime. *The Howard Journal of Crime and Justice*, *55*(1-2), pp.207-225.

[7] Sutherland, A., Brunton-Smith, I. and Jackson, J., 2013. Collective efficacy, deprivation and violence in London. *British Journal of Criminology*, *53*(6), pp.1050-1074.

[8] Deprivation Data: https://www.arcgis.com/home/item.html?id=85c8b350c1a4442eacefc1bbe1b03205

[9] Crime Data: https://www.met.police.uk/sd/stats-and-data/

Section	Word Count
Abstract	90
Introduction	238
Analytical Questions	186
Data (Materials)	300
Analysis	917
Findings, Reflections and Further Work	387