Research Methodology

Philosophy and Analytical Strategy

To understand the approach for this project, we need to identify what kind of data is being collected, used and filtered. Every data set has a unique structure which makes the data analysis mandatory. Following a typical data analysis structure, we store the data and try to read the data for a better understanding of the data set. This chapter will delve into the strategies used to analyse the data and explain the different algorithms which can be implemented.

The data being used in this project can be labelled as quantitative data and categorical data. Numerical data such as location’s longitude and latitude fall under the category of quantitative data which offers a precise pinpoint of the crime. Since the coordinates represent real life values on a map it can be used for quantitative analysis and geo-mapping. Data such as crime types, context of the crime, date of the crime, name and code of LSOA (Lower Layer Super Output Areas) as well as the outcome of the crime constitute a diverse set of information which we can label as categorical data. ADD REFERENCE CITATION HERE

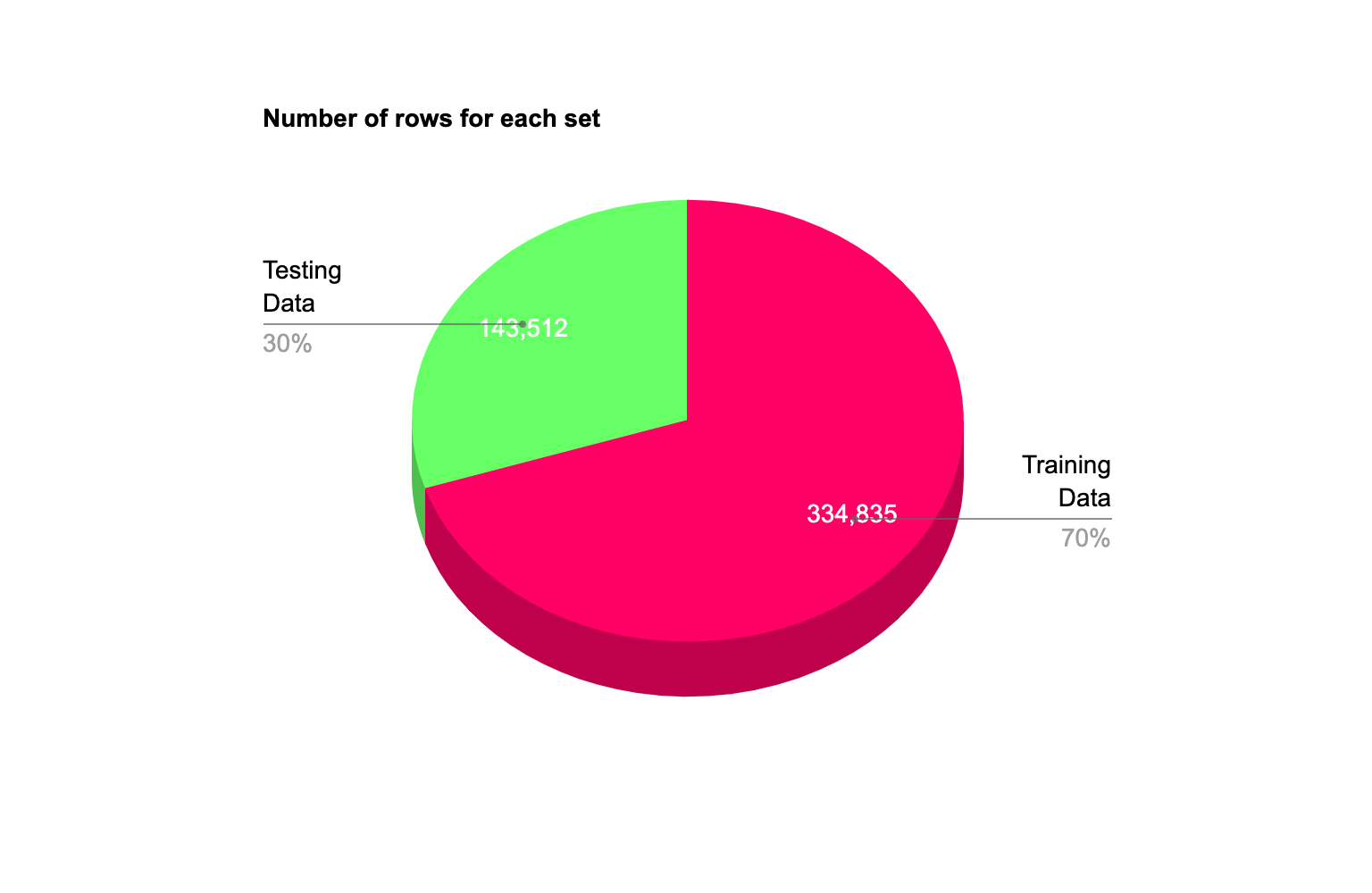
Data Collection and Analysis

The research is based on a comprehensive open-source database called 'ASB Incidents, Crimes and Consequences', a resource documenting individual incidents of crime and behaviour in England, Wales and Northern Ireland. The document, published by People at Home in the Digital World under the Open Government License v3.0, provides detailed information, including street-location, crime type, and court outcomes for each case. The geographic scope and specificity of the data make it a powerful resource for understanding the nature of crime and abuse in the context of AI-driven crime problem solving.

To begin the research, we performed qualitative data collection on the data set in CSV format. This process is initiated using a Python script designed to simplify the data preparation phase, with a specific focus on the main steps of classifying data into training and testing. The script, implemented in the Jupyter notebook or Python environment, not only collects multiple CSV files into a compatible DataFrame, but also uses an efficient way to sort the data by type.

Segmentation profiling is an important part of the research process and demonstrates optimization and reproducibility to ensure that training and testing results in a representative distribution of “crime type”. The script uses a function called split\_data\_by\_type that iterates over the dataset and efficiently splits it into training and testing data for each crime type. This stratified approach is important to minimize bias and ensure that the predictive models developed in the next phase of the study are robust and applicable to a wide range of circumstances for offending.

The script easily allows changes to the data distribution for training and testing, with the default being 70% training and 30% testing. This change is important and provides us with the flexibility to adjust the balance based on specific information or characteristics of the data. The resulting data is called "training\_data.csv" and "testing\_data.csv" and form the basis for subsequent analysis and testing.



Additionally, the script includes a random seed for data sampling to ensure repeatability of the next run. These features are important to me since I am looking for consistency in testing and analysis, facilitating comparison of results, and the ability to analyse results from multiple iterations.

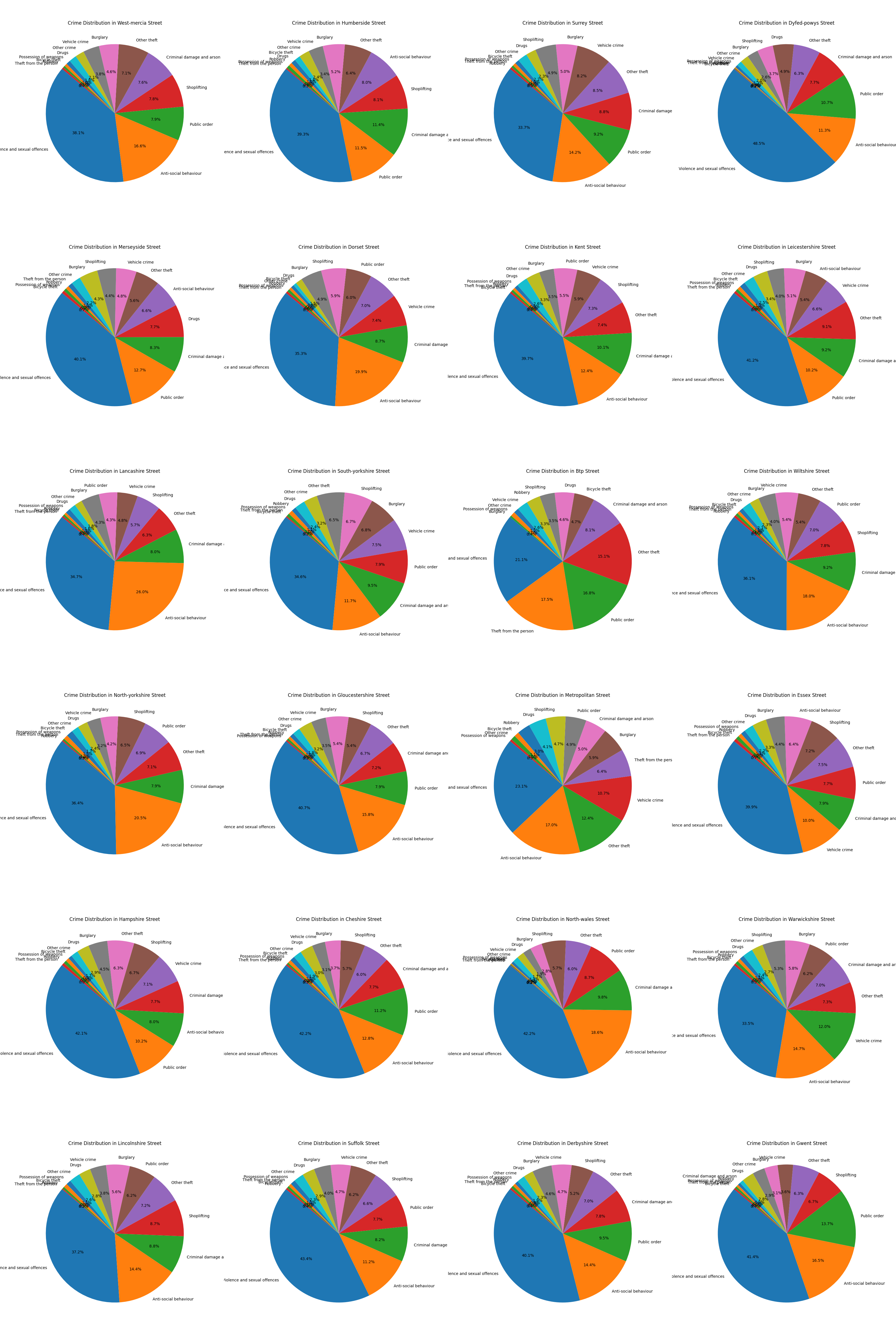
Although the main purpose of AI solutions is to investigate serious crimes, it is also important to understand the distribution of crimes in a particular location. The Metropolitan police has the most crimes for each crime type, which makes sense since it is the most densely populated area which this dataset covers.

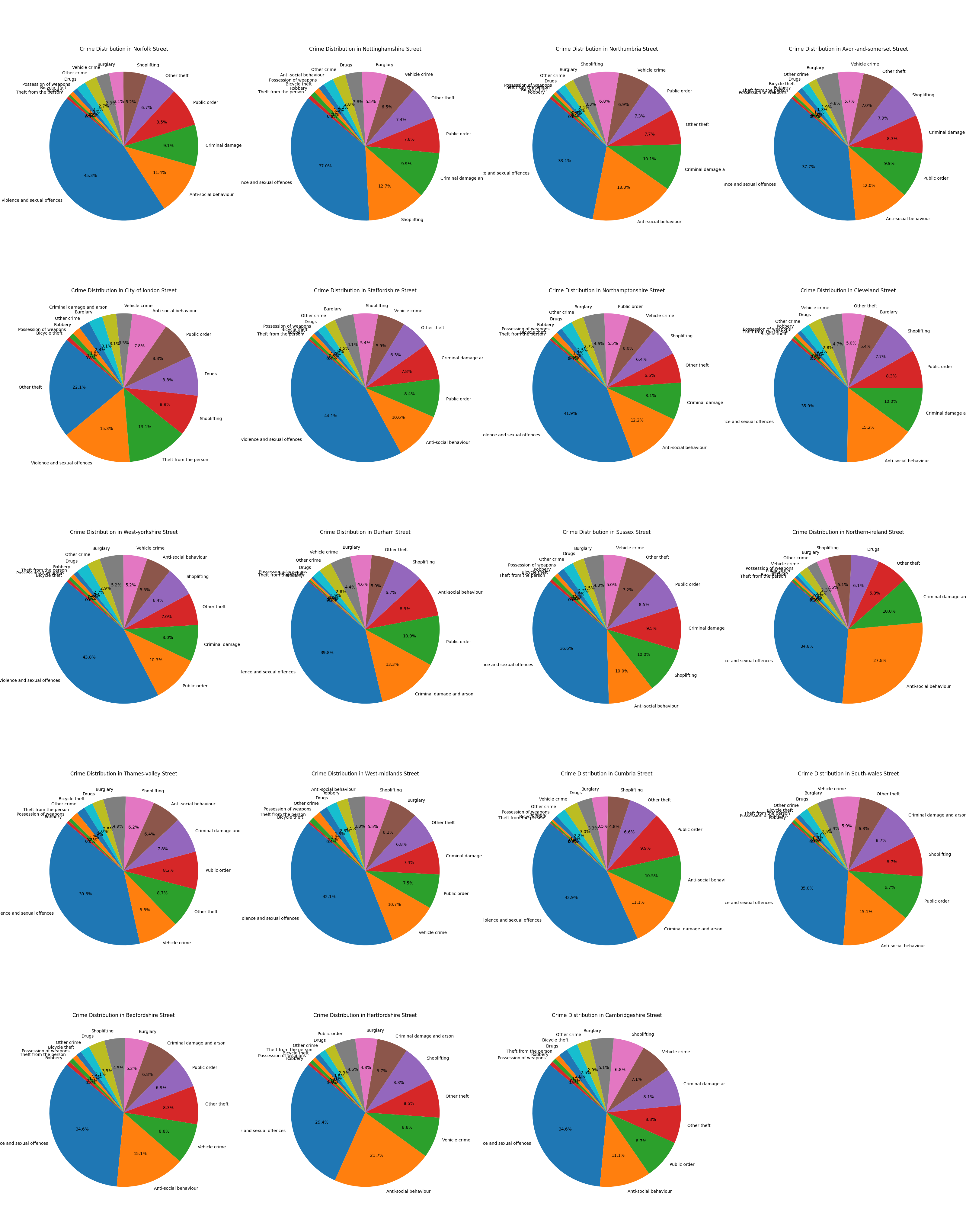
|  |  |  |
| --- | --- | --- |
| **County** | **Crime type** | **Count** |
| **2023-01-metropolitan-street** | Anti-social behaviour | 14822 |
| **2023-01-metropolitan-street** | Bicycle theft | 987 |
| **2023-01-metropolitan-street** | Burglary | 5143 |
| **2023-01-metropolitan-street** | Criminal damage and arson | 4362 |
| **2023-01-metropolitan-street** | Drugs | 3596 |
| **2023-01-metropolitan-street** | Other crime | 934 |
| **2023-01-metropolitan-street** | Other theft | 10851 |
| **2023-01-west-midlands-street** | Possession of weapons | 570 |
| **2023-01-metropolitan-street** | Public order | 4256 |
| **2023-01-metropolitan-street** | Robbery | 2658 |
| **2023-01-metropolitan-street** | Shoplifting | 4063 |
| **2023-01-metropolitan-street** | Theft from the person | 5570 |
| **2023-01-metropolitan-street** | Vehicle crime | 9350 |
| **2023-01-metropolitan-street** | Violence and sexual offences | 20178 |

Anti-social behaviour tops the list with more than 14,800 cases that specifically point to potential problems with public decision-making or social cohesion in this area. Vehicle theft (more than 9,300 incidents) and other thefts (more than 10,800 incidents) are also concerning. The table also shows the presence of other crimes such as bicycle theft, petty theft and robberies.

This data helps us understand the distribution of crime types in different counties. Even though the metropolitan police took over all the crime types we can get a general idea of the dataset.

To investigate the data in more detail we must research new methods which can showcase the most prevalent crime type. Results reveal concerns, in particular, about the amount of violence and sexual offences, which have become the most common crimes.





A closer look at the data shows that violent crime and sex offences together lead the highest crime rates in all regions. In Dyfed-Powys in particular, these crimes accounted for 48.5% of all reported incidents, demonstrating their regional concern. This is followed by Norfolk, which has the second highest number of such crimes, accounting for 45.3% of reported cases.

For example, in this case, the rate of violence and sexual offences are low in some regions. It is worth noting that the incidence of these offenses in the British Transport Police (BTP) area is low, accounting for only 21.1% of reported cases. This observation is based on expectations, given the particular nature of policing in the transportation system, where crimes of this nature would occur infrequently compared to other jurisdictions.

These findings highlight the need for response plans and resource allocation to address violent and sexual violence, especially in areas where it is a major problem. This data also shows small differences in the distribution of crimes; There should be a regional approach to suit the specific problems faced by each county.