**Introduction**

Crime, no matter its shape, size or form is detrimental to the peace, progress and growth of any society. By creating a culture of fear and distrust, deterring investment, and reducing the quality of life, the incidences of crime have wider implications both internally and externally. It can lead to a breakdown of social order, economic loss, and a total sense of insecurity among citizens depending on its intensity.

To properly understand crime, you have to look at it from the context of a framework, the usual definition of crime which often focuses on the illegal act itself fails to capture the full extent of the harm inflicted on the victims and the far reaching consequences of criminal activities in the individual in particular and the society in general. The Office of National Statistics (ONS) evaluated crime harm – the negative impacts of crime by looking at the possible victims or people experiencing the impact of crime. Accordingly, 4 categorises of harm levels were developed: Individual, community, institutional and societal levels. These categories were further classified into 5 domains which provides insights into the type of crime harm experienced by these victims. These domains are: physical, emotional or psychological, financial or economic, community safety, and privacy (ONS, 2002).

The crime rate in Bradford and by extension West Yorkshire is a major source of concern. According to CrimeRate, Bradford is the 3rd most dangerous city in the whole of England, Wales and Northern Ireland and evidently the most dangerous city in West Yorkshire. The overall crime rate in Bradford in 2022 was 159 crimes per 1,000 people which was 26% higher that the West Yorkshire rate of 126 per 1,000 people (CrimeRate, 2023).

The Office of National Statistics puts the crime rate in Bradford at 13,875 incidences per 100,000 people with a total of 75,893 incidences for the years ending July 2022 which was 12.1% higher than previous year. According to the Police, for the first 3 months of this year 2023, Bradford City alone have recorded a total of 1,618 crimes which is an average of over 539 crimes per month. Of the crimes recorded so far this year violence and sexual offences ranked highest with 594 (36.7% of total) reported cases, followed by shoplifting (231 cases, 14.3% of total) and public order (220, 13.6% of total) (Police.UK, 2023).

You cannot isolate a city from the county where it is located. The crime rate in West Yorkshire is on the increase. According to the Office of National Statistics for the year ending July 2022, West Yorkshire crime rate was 12,961 per 100,000 people arising from 304,585 reported incidences which is 42,841, or 16.4 percent higher than reported crimes in 2021. Violence related crime has been a serious cause of worry in West Yorkshire, for example, 24 people lost their lives to violent deaths in the period between October 2020 and September 2021, during this same period, 2325 people were victims of serious violence. In fact, when put in monetary terms, violence cost West Yorkshire almost a billion pounds in 2021 (West Yorkshire Violence Reduction Unit (VRU), 2022).

**Motivation**

With the figures above, there is a need to urgently address the issues of crime in West Yorkshire and the best way to start would be to use machine learning models to evaluate the underlying hidden patterns in the incidences of crime in the county. It is evident that the 5,680 police officers in West Yorkshire are overwhelmed and their efforts to combat crime though very commendable is not yielding enough. This research work is intended to provide the West Yorkshire Police and the West Yorkshire Combined Authority with valuable insights into crime patterns, enabling them more develop effective strategies to combat the menace of crime in the county.

On a personal level, this project serves as my way of giving back to the city of Bradford and West Yorkshire. By applying the knowledge that I have acquired from the University of Bradford as a student of Data Analytics and Artificial Intelligence, to analyse crime patterns and contribute to evidence-based solutions, I aspire to make a meaningful difference in the lives of the people by promoting safer communities for livelihood and for business to thrive.

**Research Question**

How can the clustering models be applied to analyse the underlying crime patterns in West Yorkshire, and identify and predict distinct hotspots or clusters of criminal activities to aid in targeted interventions and efficient resource allocation to combat criminality?

**Objectives of Study**

1. Evaluate the effectiveness of the clustering algorithms in identifying crime hotspots and clusters in West Yorkshire
2. Analyse the spatial characteristics of the identified clusters and their relationship to socio-demographic factors or environmental features
3. Access the stability and robustness of the clustering results by varying the algorithm parameters or employing alternative clustering techniques
4. Explore the temporal dynamics of the crime clusters to understand if there are any recurring patterns or changes over time
5. Examine the crime types and modus operandi associated with each cluster to gain insights into the underlying factors contributing to criminal activities in specific areas.
6. Provide actionable recommendations for targeted interventions, resource allocation, and crime prevention strategies based on the identified crime clusters
7. Promote knowledge-sharing and collaboration with local law enforcement agencies and community organizations to ensure the practical application of the clustering results in crime prevention initiatives

**Literature Review**

**Crime Causation Theories**

A growing body of research have considered the importance of neurobiological factor in criminal or antisocial behaviours. These studies have identified several brain regions and neural system that may be particularly relevant to criminality. To this end, a number of ways have been identified that could potentially help to understand the importance of neurobiology in criminality. First, is the research on criminal behaviour, second is the early discovery and possible prediction of behaviours considered to be criminal or antisocial in nature, third it can help with the determination and direction of criminal proceedings and also it can also help with prescribing the treatment of criminality (Focquaert, 2018)

Ling et al 2019, evaluated the relationship between 3 biological factors: psychophysiology, brain, and genetics and criminality or antisocial behaviour. According to them despite these growing body of research these factors should be seen as complementary to current research and theories and possibly another way to investigate the treatment option for criminality.

Several research studies have identified a significant relationship between psychopathy – neuropsychiatric disorder marked by lack of empathy, resulting into frequent antisocial behavioural patterns and criminality and physiological processes. Macdougall 2016, investigated the possible relationship between the levels of heart rate (HR), Respiratory Sinus Arrhythmia (RSA), and skin conductance (SC) and adolescence psychopathic behaviours. She found positive and significant correlation between skin conductance level and Psychopathy Checklist-Youth Version (PCL:TV)

psychopath tendencies amongst youths. Another study shows that poor skin conductance conditioning at a very early age is associated with aggression at 8 years old and predisposed people to criminality during adolescent ages (Gao et al 2010).

Psychoanalytic theories of crime causation suggest that unconscious psychological processes may contribute to criminal behaviour which according to them is a representation of “psychological conflict”. These theories are based on the work of Sigmund Freud and other psychoanalytic theorists, who proposed that early childhood experiences, particularly related to the parent-child relationship, can shape an individual's personality and behaviour. According to them, to understand criminality, we must as a matter of fact gain insight into interaction between the unconscious motives of individuals and their behaviours (James Byrne, 2010).

**Crime Patterns and Factors in West Yorkshire**

The nature of crime in West Yorkshire is diverse and multifaceted, the countyis located in Northern England and comprises of 5 metropolitan boroughs: Leeds, Wakefield, Bradford, Kirklees and Calderdale. Crime prevention is at the front burner of the West Yorkshire Combined Authority, Tracy Brabin the first mayor of West Yorkshire elected in May 2021 centred her election manifesto on effective policing and crime prevention with specific emphasis on tackling drugs related crimes, preventing anti-social behaviours of all forms, tackling modern slavery, and building a system that lays emphasis on preventing crimes through early intervention and a strategy that is victim-centred in its responses to crime in the county (West Yorkshire Combined Authority, 2022).

Incidents of violence with injury in West Yorkshire have risen by 100.6% between 2011 and 2021 from 12,111 to 24,298 incidents per year. During that same period however, murders or homicides decreased by 10.7% from 28 to 25 per year. The table below shows the types of crimes in West Yorkshire ranked by total incidents between 2020 to 2022.

**Table 1: Types of Crimes in West Yorkshire Ranked by Incidence 2021 to 2022**

|  |  |  |  |
| --- | --- | --- | --- |
| TYPE | INCIDENTS | CRIME RATE | PERCENTAGE |
| Stalking and harassment | 53,834 | 2,291 | 18% |
| Violence without injury | 46,173 | 1,965 | 15% |
| Public order offences | 43,285 | 1,842 | 14% |
| Violence with injury | 29,538 | 1,257 | 10% |
| Criminal damage | 29,207 | 1,243 | 10% |
| All other theft offences | 21,298 | 906 | 7% |
| Shoplifting | 16,399 | 698 | 5% |
| Vehicle offences | 16,136 | 687 | 5% |
| Residential burglary | 9,667 | 411 | 3% |
| Sexual offences | 9,414 | 401 | 3% |
| Drug offences | 8,299 | 353 | 3% |
| Miscellaneous crimes | 7,003 | 298 | 2% |
| Non-residential burglary | 3,959 | 168 | 1% |
| Theft from the person | 3,095 | 132 | 1% |
| Robbery | 2,771 | 118 | 1% |
| Possession of weapons | 2,425 | 103 | 1% |
| Bicycle theft | 1,985 | 84 | 1% |
| Death by dangerous driving | 68 | 3 | 0% |
| Homicide | 29 | 1 | 0% |

Source: Varbes.com

The table shows that stalking and harassment, violence without injury, public order offences and violence with injury ranked highest in incidences while death by dangerous driving and homicide had zero percent increase during the years in review.

**Table 2: West Yorkshire ranked by Safety**

|  |  |  |  |
| --- | --- | --- | --- |
| RANK | AREA | INCIDENTS | CRIME RATE |
| 1 | Kirklees | 46445 | 10718 |
| 2 | Calderdale | 25493 | 12326 |
| 3 | Wakefield | 45931 | 12982 |
| 4 | Leeds | 110823 | 13698 |
| 5 | Bradford | 75893 | 13875 |

Source: Varbes.com

Table 2 shows that Kirklees is safest region while Bradford is by far the most dangerous region in West Yorkshire by crime rate.

Organised crime is also on the increase in West Yorkshire, between 27th February and 5th March 2023 the West Yorkshire police arrested 37 people with hard drugs and modern slavery. £11,000 worth of cocaine, 1,200 wraps of crack cocaine, 500 wraps of heroin, 2,400 cannabis plants, 100 illegal vapes and £186,000 cash, 3 knives, 2 axes, a Taser and a firearm were confiscated. The operation saw 158 vulnerable people including 38 children safeguarded (BBC News, March, 2023).

As at 2022, the county has about 146 organised crime groups (OCGs) which prior to that has increased significantly over a period of time. The OCGs are classified into groups based on their primary form of criminality. This classification is shown in the table below

**Table 3: Organised Crime Groups in West Yorkshire**

|  |  |
| --- | --- |
| **Group** | **Total** |
| Drug Supply | 69 |
| Organized Acquisitive Crime | 34 |
| Modern Day Slavery | 12 |
| Firearm | 3 |
| Criminal Financing | 13 |
| Child Sexual Exploitation and Abuse | 3 |
| County Line | 12 |

***Source: Programme Precision, 2022***

The activities of these OCGs are known to be widespread, 46 West Yorkshire OCGs operate across several counties affecting 24 other policing units in the United Kingdom and internationally in 11 different countries and on the flip side West Yorkshire is affected by 293 OCGs from 21 other policing regions (Programme Precision, 2022).

What are the factors responsible for criminality in West Yorkshire? The need assessment conducted by the West Yorkshire Violence Reduction Unit (VRU), which was aimed at understanding the nature and ramifications of violent crimes in West Yorkshire, highlighted 9 key problems leading to violent crimes in the county. These are: derivation, housing, unemployment, substance use, mental health, neurodiversity, disability, gender education – the lack of it, technology and social media influences. All these factors combined together have in one way or the other contributed to the escalation of violent crime in West Yorkshire (West Yorkshire Violence Reduction Unit (VRU), 2022).

**Crime Analysis Method and Techniques**

The New York Police Department has traced the history of crime mapping to at least 1900, when police departments began using pins or dots on maps to represent crime incidents. The limitation of this process however is that it was difficult to trace patterns when you update crime locations on the map except you photograph the previous location on the maps. Not only was this process cumbersome it was also static because physical maps cannot be queried for more information about crime modes, types, perpetrators, victims etc. To this end, historical analysis of crime is lost; a pin on a map doesn’t say much (Keith Harries, 1999). However, the advent of Geographic Information Systems (GIS) technology in the 1960s and 1970s revolutionized crime mapping, allowing for more sophisticated spatial analysis and visualization of crime data.

Other techniques of crime analysis that have been used in times past include: spatial analysis, temporal analysis, hot spot analysis, predictive analysis, and social network analysis all these and more shall be reviewed in details in chapter 2 of the project. However, we are interested in using machine learning model(s) in this project.

**Methodology**

First, we start with the brief review of literature providing empirical evidences supporting the use of machine learning models in general and clustering model in particular in crime analysis.

Kshatri et al, 2022 used Naïve Bayes and random tree classification models to forecast violent crimes using the National Crime Records Bureau (NCRB) for all states in India and discovered that the machine leaning models outperformed other models on the same dataset Top of Form

Llaha, 2020 evaluated the effectiveness and performances of machine learning methods in conjunction with other models in analysing historical crime data and she discovered that the Decision Tree achieved better results with respect to correctly classified instances and accuracy in prediction

Bharati and Sarvanaguruin 2018 using machine learning models on the Chicago crime dataset were able to predict the type of crime that may likely happen given other variables with 79% accuracy on the test dataset

Asor, et al 2022 used some machine learning algorithms to develop models in crime reports in Laguna Province in the Philippines. They discovered that ensembles like random forest and gradient boasting performed better in the classification of crime when compared to decision tree. In fact, gradient boasting was more efficient for both classification and regression compared to random forest when the number of trees were less than 100.

Bokde et al, 2018 using KMeans clustering model found it very easy to identify underlying crime patterns and trends in historical crime data to this end, they concluded that with machine learning crime data mining has a promising future

Jain et al 2017 also using the KMean model concluded that the model was efficient in determining common crime trends and identifying specific crime prone areas. According to them the model was very useful with complex crime scenarios and a valuable analytical tool for crime detection and prevention.

**Data Sources and Collection**

The United Kingdom has very rich crime data collection and gathering sources therefore the project relies entirely on secondary data. We have comprehensive datasets from West Yorkshire Police and also data from Office of National Statistics which is the official data gathering and collection agency for the United Kingdom government. The relevant datasets shall be downloaded from the official websites of the data sources mentioned above.

**Data Cleaning and Preparation**

Using the Python programming language and specifically depending on the Pandas library we shall clean and prepare the datasets for the modelling stages. During the course of cleaning, we shall look out for missing data, data duplications, incorrectly signed data, irrelevant columns, rows or data points, irregular or inconsistent data etc. at the end of the day we shall select the relevant features which should include but not limited to: geographical coordinates, crime types, number of incidences, data and time and possible demographic characteristics. After this, we shall perform feature scaling using normalization or standardization to ensure that all different features are on the same scale.

**Model Selection**

The project intends to use the K-Means clustering algorithm. Clustering provides insights into hidden underlying patterns in data by grouping similar data points together based on certain features, characteristics and proximity. It will help us identify distinct crime clusters or hotspots of criminal activities in West Yorkshire and consequently we can predict the crime patterns associated with each cluster. K-Means follows the Expectation-Minimization approach leading us to 2 computational steps, details to be provided in chapter 3 of the project.

**Test The Validity of the Identified Clusters**

There are 3 popular metrics in literature for testing the validity of clusters aimed at measuring the compactness and separation of each cluster.

1. Dunn Index
2. Silhouette score
3. Davis Bouldin score

We shall also test the validity of the clusters empirically by comparing them with known crime patterns or expert knowledge. Note that at this stage of the analysis we shall run and rerun the model using different parameters and hyper parameters to ensure that the clusters created satisfy the standard clustering requirements: ethically, mathematically and intuitively.

**Cluster Exploratory Data Analysis**

We shall perform a detailed exploratory data analysis on each cluster identified by the model as follows:

1. Visualize the clusters on a map to understand the spatial distributions of crime patterns and identify hotspots
2. Profile each cluster with respect to type of crime, temporal patterns or demographic profiles, socio-economic variables or environmental factors and victim profiles,
3. Provide visual representations of the identified crime patterns, hotspots, and cluster characteristics

**Analytical Tools**

As stated earlier, the Python Programming language shall be used extensively and the following Python libraries shall be deployed.

1. Scikit-Learn – for deploying the models
2. Pandas and Numpy– for data wrangling
3. Matplotlib, Seaborn and Plotly Express – for data visualisations
4. Pydeck or Folium for geospatial data visualisation

Finally, I hope to be able to create a dashboard to be hosted online detailing my findings to be used as presentation to the West Yorkshire Combined Authority and the West Yorkshire Police.