The Battle of Neighbourhoods

A. Data:

A.1. Description of Data:

A. 1.a) Dataset 1: "MPS Ward Level Crime (most recent 24 months)"

- The dataset, "MPS Ward Level Crime Data (most recent 24 months)", has been extracted from the Metropolitan Police Service's "Recorded Crime: Geographic Breakdown" Data available on the London Datastore
- The extracted data is the most recent data available updated till June 2021
- This data counts the number of crimes per month according to crime type at the geographic level of London's Wards for the period July 2019 to June 2021
- In March 2019, the Metropolitan Police Service started to provide offences grouped as per the updated Home Office crime classifications
- Below is a list of the crime types covered under the new Home Office categories:

No.	Major Category (Labeled as "Crime Head")	Minor Category (Labeled as "Crime Sub-Head")
1.	• Arson and Criminal Damage	• Arson • Criminal Damage
2.	• Burglary	• Burglary - Business and Community • Burglary - Residential
3.	• Drug Offences	• Drug Trafficking • Possession of Drugs
4.	• Miscellaneous Crimes Against Society	• Absconding from Lawful Custody • Bail Offences • Bigamy • Concealing an Infant Death Close to Birth • Dangerous Driving • Disclosure, Obstruction, False or Misleading State • Exploitation of Prostitution • Forgery or Use of Drug Prescription • Fraud or Forgery Associated with Driver Records • Going Equipped for Stealing • Handling Stolen Goods • Making, Supplying or Possessing Articles for use • Obscene Publications • Offender Management Act

		• Other Forgery • Other Notifiable Offences • Perjury • Perverting Course of Justice • Possession of False Documents • Profiting from or Concealing Proceeds of Crime • Soliciting for Prostitution • Threat or Possession with Intent to Commit Crime • Wildlife Crime
5.	• Possession of Weapons	• Other Firearm Offences • Possession of Firearm with Intent • Possession of Firearms Offences • Possession of Other Weapon • Possession of Article with Blade or Point
6.	• Public Order Offences	• Other Offences Against the State, or Public Order • Public Fear Alarm or Distress • Racially or Religiously Aggravated Public Fear • Violent Disorder
7.	• Robbery	• Robbery of Business Property • Robbery of Personal Property
8.	• Sexual Offences	• Rape • Other Sexual Offences
9.	• Theft	• Bicycle Theft • Other Theft • Shoplifting • Theft from Person
10.	• Vehicle Offences	• Aggravated Vehicle Taking • Interfering with a Motor Vehicle • Theft from a Motor Vehicle • Theft or Taking of a Motor Vehicle
11.	• Violence Against the Person	• Homicide • Violence with Injury • Violence without Injury

- Before cleaning the data, the dataset contained 29 columns
- Post cleaning and processing the data, 5 columns have been renamed and 1 has been added
- The dataset now contains the following 30 columns:
  - **Ward Code:** Code of the Ward in the London Borough

- **Ward:** Name of the Ward in the London Borough
- **Borough:** Name of the London Borough
- **Crime Head:** High level categorisation of crime
- **Crime Sub-Head:** Low level categorisation of crime within Crime Head
- **201907 … 202106:** 24 separate columns for the Year and Month of the Reported Crime starting from 201907 to 202106. These columns show the number of reported crimes in the month for a particular Ward.
- **Total:** Total Crimes Reported for a particular Ward from July 2019 to June 2021

- While the exploring the dataset, it was found that there were two Wards by the name of "Belmont" in Harrow as well as in Sutton. Hence, in order to segregate them so as not to cause any confusion during analysis, their names were changed to "Belmont Harrow" and "Belmont Sutton".
- Further, in order to maintain consistency, the names of these two Wards were also changed in the fourth dataset, which had the London Postcodes
- The original dataset contained a total of 22,403 records
- Once the dataset was processed to include only the Top 5 safest Boroughs of London, the number of records reduced to 3,007 from 22,403 records
- After the dataset was processed further, to include only the Top 50 safest Wards of London, the number of records reduced to 1,549 from 3,007 records

A. 1.b) Dataset 2: "List of London Boroughs"

- The dataset, "List of London Boroughs", has been extracted from Wikipedia.org
- Post cleaning and processing the data, 2 columns have been dropped and 5 columns have been renamed
- The dataset now contains the following 8 columns:
  - **Borough:** Name of the London Borough
  - **Local Authority:** Name of the Local Authority
  - **Political Control:** Name of the Political Party controlling the Local Authority
  - **Head Quarters:** Address of the Local Authority
  - **Area (sq mi):** Area of the Borough in square miles
  - **Population (2019 estimate):** 2019 estimate of the Population of the Borough
  - **Co-ordinates:** Latitude and Longitude of the Borough
  - **Borough No. on Map:** Designated Number of the Borough on the Map of London
- The dataset contains a total of 32 records, which is the total number of London Boroughs, excluding the City of London

A. 1.c) Dataset 3: Merged Dataset of "MPS Ward Level Crime (most recent 24 months)" and "List of London Boroughs"

- The third dataset has been created by merging the first two datasets, i.e., by merging the datasets, "MPS Ward Level Crime (most recent 24 months)" and "List of London Boroughs"
- The two datasets have been merged on the common column present in both the datasets, i.e., the "Borough" column
- After merging and reindexing the columns, the dataset contains the following 37 columns:
  - **Ward Code:** Code of the Ward in the London Borough
  - **Ward:** Name of the Ward in the London Borough
  - **Borough:** Name of the London Borough
  - **Local Authority:** Name of the Local Authority
  - **Political Control:** Name of the Political Party controlling the Local Authority
  - **Head Quarters:** Address of the Local Authority
  - **Area (sq mi):** Area of the Borough in square miles
  - **Population (2019 estimate):** 2019 estimate of the Population of the Borough
  - **Co-ordinates:** Latitude and Longitude of the Borough
  - **Borough No. on Map:** Designated Number of the Borough on the Map of London
  - **Crime Head:** High level categorisation of crime
  - **Crime Sub-Head:** Low level categorisation of crime within Crime Head
  - **201907 … 202106:** 24 separate columns for the Year and Month of the Reported Crime starting from 201907 to 202106. These columns show the number of reported crimes in the month for a particular Ward.
  - **Total:** Total Crimes Reported for a particular Ward for the period July 2019 to June 2021
- The dataset contains a total of 22,403 records

A. 1.d) Dataset 4: "London Postcodes"

- The fourth dataset has been created to find Venues in the Neighbourhood of London using the Foursquare API
- The dataset, "London Postcodes", has been extracted from Doogal.co.uk
- The dataset has a complete list of London postcode districts
- The original dataset had 49 columns, but after the cleaning process, 28 columns were dropped as the same were not required for analysis

- Post cleaning and processing the data, the dataset contains the 21 columns, like Postcode Data, Nearest Station, Distance to Station, Ward, District, Constituency, etc.
- Even though this dataset already had the Latitude and the Longitude data available, I have used the ArcGIS API to re-fetch the coordinates of the preferred locations
- Before cleaning the data, the dataset contained a total of 3,24,634 records
- The number of records were reduced to 1,79,704 from 3,24,634 after removing the Postcodes that were not in use
- Once the dataset was processed to include only the Top 5 safest Boroughs of London, the number of records reduced to 20,249 from 1,79,704
- After the dataset was processed further, to include only the Top 50 safest Wards of London, the number of records reduced to 10,083 from 20,249
- Now, we could have used these 10,083 Postcodes of the Top 50 Wards of London to find their coordinates, but the process of fetching the coordinates for so many postcodes would have taken a lot of time. Hence, it was necessary to reduce the number of records further.
- Therefore, in order to reduce the dataset further, I selected the location that was nearest to the Station
- After processing the dataset, it was found that there were a total of 81 locations in the Top 50 safest Wards of London that were nearest to the Stations
- Thus, the number of Postcodes were reduced from 10,083 to 81

A.2. Use of Data to Solve the Problem:

A.2.a) Dataset 1: "MPS Ward Level Crime (most recent 24 months)"

- This dataset has been used to find Boroughs that have the highest and the lowest crime rate
- After having found the boroughs with the lowest crime rate, the data was sorted, and the 5 safest Boroughs in London were identified
- Though the 5 Boroughs identified could have served the purpose, as these 5 Boroughs were the safest as compared to the other Boroughs of London; I wanted to further eliminate the areas with crime so as to find the most secure venues for my target audience
- Also, had I taken all the 92 Wards from these 5 safe Boroughs, there was still a possibility that some of the Venues could fall in the 'unsafe' Ward of that particular safe Borough
- Therefore, in order to avoid such a scenario and to ensure that the Venues found were from the most secure areas of London, another layer of safety was added to identify the 10 Most Safest Wards within each of the 5 Most Safest Boroughs

- Thus, out of a total of 615 Wards in the whole of London, I have shortlisted only the 50 Most Safest Wards

A.2.b) Dataset 2: "List of London Boroughs"

- This dataset was used to fetch more information on the respective Boroughs of London

A.2.c) Dataset 3: Merged Dataset of "MPS Ward Level Crime (most recent 24 months)" and "List of London Boroughs"

- The merged dataset provides more information on the different Boroughs of London, like the local authority of the borough, the political party controlling the local authority, the address of the local authority, the area of the Borough, its population, its coordinates and its designated number on the map of London
- Thus, with this information we can get more insight in to the various Boroughs of London

A.2.d) Dataset 4: "London Postcodes"

- As the name suggests, this dataset has been used to fetch the Postcodes of the different neighbourhoods in London
- The dataset was processed to identify the postcodes of the Top 50 safest Wards of London
- This data has been used to fetch the geographical coordinates, i.e., the Latitude and Longitude, of the different neighbourhoods within the Top 50 safest Wards of London
- We have used the **ArcGIS API** to gather the Latitude and Longitude coordinates of the neighbourhoods based on their postcodes
- In order to reduce the number of neighbourhoods, the neighbourhoods nearest to the station have been selected
- This will greatly assist the target audience, as finding venues that are nearer to the stations will not only reduce their travel time but will also be more convenient to them
- These location coordinates were then used to highlight the locations on the Map of London
- The Latitude and Longitude coordinates were also linked with the **Foursquare API** to identify the different venues near these neighbourhoods
- Later, this data has been used to build a model to cluster the neighbourhood by using Machine Learning (k-means clustering)

A.3. Data Source:

A.3.a) Dataset 1: "MPS Ward Level Crime (most recent 24 months)"

- The dataset, "MPS Ward Level Crime Data (most recent 24 months)", has been extracted from the Metropolitan Police Service's "Recorded Crime: Geographic Breakdown" Data available on the London Datastore: https://data.london.gov.uk/dataset/recorded_crime_summary

A.3.b) Dataset 2: "List of London Boroughs"

- The dataset, "List of London Boroughs", has been extracted from the Wikipedia.org page: https://en.wikipedia.org/wiki/List_of_London_boroughs

A.3.c) Dataset 3: Merged Dataset of "MPS Ward Level Crime (most recent 24 months)" and "List of London Boroughs"

- Created by merging Datasets 1 & 2

A.3.d) Dataset 4: "London Postcodes"

- The dataset, "London Postcodes", has been extracted from Doogal.co.uk: **https://www.doogal.co.uk/london_postcodes.php**

A.3.e) ArcGIS API Data:

- ArcGIS is an online API that enables us to connect people, locations, and data using interactive maps
- We use the ArcGIS API to get the geographical coordinates (Latitude and Longitude) of the neighbourhoods of London by providing the Postcodes of the desired locations
- The following information is obtained for each Postcode,
  - **Postcode:** Postcode
  - **Latitude:** Latitude of the Postcode
  - **Longitude:** Longitude of the Postcode

A.3.f) Foursquare API Data:

- Foursquare is a location data provider with information about different venues and events within an area of interest
- The information obtained from the Foursquare API includes venue names, locations, menus, reviews, photos, etc.

- The Foursquare location platform is, thus, used by us as a data source since all the required information about the different venues in various neighbourhoods of the desired Borough or Ward can be obtained through their API
- In order to get the required information, we provide the Foursquare API with the Latitude and Longitude coordinates of the preferred neighbourhood
- Based on the Latitude and Longitude coordinates, the Foursquare API acquires information about different venues within each neighbourhood
- The data retrieved from the Foursquare API contains information of venues, which are within the specified distance of the latitude and longitude of said postcode.
- The venues identified are in the radius of 500 meters from the preferred neighbourhood
- The following information is obtained for each venue,
    - **Neighbourhood:** Name of the Neighborhood
    - **Neighbourhood Latitude:** Latitude of the Neighborhood
    - **Neighbourhood Longitude:** Longitude of the Neighborhood
    - **Venue:** Name of the Venue
    - **Venue Latitude:** Latitude of the Venue
    - **Venue Longitude:** Longitude of the Venue
    - **Venue Category:** Category of the Venue
- We then use the information collected to build a Machine Learning model using the K-Means Clustering algorithm
- The model clusters the neighbourhoods together based on similar venue categories
- This data is then used to present our observations and findings and assist us in taking the necessary decisions