

## *2. Description of the Data and how it used to solve the problem*

### **Data Acquisition and cleaning**

#### **2.1 Data Sources and consists of the following:**

The data required for the project will be taken from the following sources

- i) Section 1: London crime data
- ii) Section 2: List of London boroughs
- iii) Section 3: Wikipedia, list of neighbourhoods in the Royal Borough of Kingston.

The cleansed data will then be used alongside Foursquare data, which is readily available. Foursquare location data will be leveraged to explore or compare boroughs around London

#### **Data purposes:**

- The London crime data will find the following: lsoa\_code, borough, major\_category, minor\_category, value, year, month
- The list of London Borough will provide the following: Borough, Inner, Status, Local authority, Political control, Headquarters, Area, Population, Co ordinates, Nr in map.
- The list of neighbourhoods will provide the following: Neighbourhood, Borough, Latitude, Longitude

#### **2.2 Data cleaning**

Data downloaded or scraped from multiple sources were combined into one table.

This is done in the following sequences

The London crime data is accessed from Kaggle and this is then uploaded to Amazon S3 bucket.

This then downloaded into the notebook and into a dataframe.

The data was then processed to take off the null entries and the most recent data since 2016 kept.

##### *2.2.1 Pre processing data for recent crimes*

Table 1: London Crime after pre-processing

In three sections, data will be handled separately. From the London crime data, recent crimes are in selected. The major categories of crime are pivoted to get the total crimes per Borough

```
Out[12]:
```

|   | isoa_code | borough   | major_category     | minor_category                   | value | year | month |
|---|-----------|-----------|--------------------|----------------------------------|-------|------|-------|
| 0 | E01004177 | Sutton    | Theft and Handling | Theft/Taking of Pedal Cycle      | 1     | 2016 | 8     |
| 1 | E01000733 | Bromley   | Criminal Damage    | Criminal Damage To Motor Vehicle | 1     | 2016 | 4     |
| 2 | E01003989 | Southwark | Theft and Handling | Theft From Shops                 | 4     | 2016 | 8     |
| 3 | E01002276 | Havering  | Burglary           | Burglary in a Dwelling           | 1     | 2016 | 8     |
| 4 | E01003674 | Redbridge | Drugs              | Possession Of Drugs              | 2     | 2016 | 11    |

Change the column names

## 2.2.2 Scraping Data for List of Borough using beautiful soup

Table 2: List of London Boroughs

The second data scraped a Wikipedia page using the Beautiful soup library in Python. This helps us to extract the data in the tabular form as shown on the website. After web scraping, string manipulation is required the names of the boroughs in the correct form. This is important because of the merging of the datasets using borough names.

```
Out[61]:
```

|    | Borough                         | Inner | Status | Local authority                               | Political control          | Headquarters                             | Area (sq mi) | Population (2013 est)[1] | Co-ordinates                              | Loc. in map |
|----|---------------------------------|-------|--------|---|----------------------------|--|--------------|--------------------------|---|-------------|
| 0  | Barking and Dagenham (note 1)   | Nah   | Nah    | Barking and Dagenham London Borough Council   | Labour                     | Town Hall, 1 Town Square                 | 13.93        | 194352                   | 51°33'39"N 0°9'21"E / 51.5607°N 0.1557°E  | 25          |
| 1  | Barnet                          | Nah   | Nah    | Barnet London Borough Council                 | Conservative               | Barnet House, 2 Bristol Avenue, Coledale | 33.49        | 369088                   | 51°37'31"N 0°9'06"W / 51.6252°N 0.1517°W  | 31          |
| 2  | Berley                          | Nah   | Nah    | Berley London Borough Council                 | Conservative               | Civic Offices, 2 Watling Street          | 23.38        | 236687                   | 51°27'18"N 0°9'02"E / 51.4549°N 0.1505°E  | 23          |
| 3  | Brent                           | Nah   | Nah    | Brent London Borough Council                  | Labour                     | Brent Civic Centre, Engineers Way        | 16.70        | 317264                   | 51°33'32"N 0°18'54"W / 51.5588°N 0.2817°W | 12          |
| 4  | Bromley                         | Nah   | Nah    | Bromley London Borough Council                | Conservative               | Civic Centre, Stockwell Close            | 57.97        | 317899                   | 51°24'14"N 0°9'11"E / 51.4039°N 0.0198°E  | 20          |
| 5  | Camden                          | Nah   | Nah    | Camden London Borough Council                 | Labour                     | Camden Town Hall, Judd Street            | 8.40         | 229719                   | 51°31'44"N 0°7'32"W / 51.5290°N 0.1255°W  | 11          |
| 6  | Croydon                         | Nah   | Nah    | Croydon London Borough Council                | Labour                     | Bernard Weatherill House, Met Walk       | 33.41        | 372752                   | 51°22'17"N 0°9'52"W / 51.3714°N 0.0977°W  | 19          |
| 7  | Ealing                          | Nah   | Nah    | Ealing London Borough Council                 | Labour                     | Perceval House, 14-16 Uxbridge Road      | 21.44        | 342494                   | 51°30'47"N 0°18'32"W / 51.5130°N 0.3089°W | 13          |
| 8  | Enfield                         | Nah   | Nah    | Enfield London Borough Council                | Labour                     | Civic Centre, Silver Street              | 31.74        | 320524                   | 51°39'14"N 0°04'48"W / 51.6538°N 0.0799°W | 30          |
| 9  | Greenwich (note 2) (note 3)     | Royal | Royal  | Greenwich London Borough Council              | Labour                     | Woolwich Town Hall, Wellington Street    | 18.28        | 264008                   | 51°29'21"N 0°03'53"E / 51.4892°N 0.0648°E | 22          |
| 10 | Hackney                         | Nah   | Nah    | Hackney London Borough Council                | Labour                     | Hackney Town Hall, Mare Street           | 7.36         | 257379                   | 51°32'42"N 0°03'19"W / 51.5450°N 0.0553°W | 9           |
| 11 | Hammersmith and Fulham (note 4) | Nah   | Nah    | Hammersmith and Fulham London Borough Council | Labour                     | Town Hall, King Street                   | 6.33         | 178685                   | 51°29'24"N 0°14'02"W / 51.4927°N 0.2339°W | 4           |
| 12 | Haringey (note 3)               | Nah   | Nah    | Haringey London Borough Council               | Labour                     | Civic Centre, High Road                  | 11.42        | 263386                   | 51°36'00"N 0°06'43"W / 51.6000°N 0.1119°W | 29          |
| 13 | Harrow                          | Nah   | Nah    | Harrow London Borough Council                 | Labour                     | Civic Centre, Station Road               | 19.49        | 243372                   | 51°35'23"N 0°20'05"W / 51.5888°N 0.3345°W | 32          |
| 14 | Havering                        | Nah   | Nah    | Havering London Borough Council               | Conservative (council NOC) | Town Hall, Main Road                     | 43.35        | 242980                   | 51°34'52"N 0°11'01"E / 51.5812°N 0.1837°E | 24          |
| 15 | Hillingdon                      | Nah   | Nah    | Hillingdon London Borough Council             | Conservative               | Civic Centre, High Street                | 44.67        | 286805                   | 51°32'39"N 0°28'34"W / 51.5441°N 0.4760°W | 33          |
| 16 | Hounslow                        | Nah   | Nah    | Hounslow London Borough Council               | Labour                     | Hounslow House, 7 Bath Road              | 21.51        | 262407                   | 51°28'29"N 0°22'05"W / 51.4746°N 0.3680°W | 14          |
| 17 | Islington                       | Nah   | Nah    | Islington London Borough Council              | Labour                     | Customer Centre, 222 Upper Street        | 5.74         | 215687                   | 51°32'50"N 0°09'06"W / 51.5416°N 0.1522°W | 10          |
| 18 | Kensington and Chelsea          | Nah   | Royal  | Kensington and Chelsea London Borough Council | Conservative               | The Town Hall, Hornton Street            | 4.68         | 155594                   | 51°30'07"N 0°11'41"W / 51.5020°N 0.1947°W | 3           |
| 19 | Kingston upon Thames            | Nah   | Royal  | Kingston upon Thames London Borough Council   | Liberal Democrat           | Guildhall, High Street                   | 14.38        | 166793                   | 51°24'31"N 0°18'23"W / 51.4085°N 0.3064°W | 16          |
| 20 | Lambeth                         | Nah   | Nah    | Lambeth London Borough Council                | Labour                     | Lambeth Town Hall, Bidon Hill            | 10.36        | 314242                   | 51°27'39"N 0°06'59"W / 51.4607°N 0.1163°W | 6           |
| 21 | Lewisham                        | Nah   | Nah    | Lewisham London Borough Council               | Labour                     | Town Hall, 1 Catford Road                | 13.57        | 286180                   | 51°26'43"N 0°11'15"W / 51.4452°N 0.0209°W | 21          |
| 22 | Merton                          | Nah   | Nah    | Merton London Borough Council                 | Labour                     | Civic Centre, London Road                | 14.52        | 203223                   | 51°24'05"N 0°11'45"W / 51.4014°N 0.1958°W | 17          |

## 2.2.3 Merging dataframes to get a single dataframe.

Table 3: London Borough Crime

The two datasets are merged on the borough names form a new dataset that combines the necessary information. Purpose of this dataset is to visualise the crime rates in each borough and identify the borough with the least crimes recorded during 2016.

| Borough | Local authority      | Political control | Headquarters                                    | Area (sq m) | Population (2013 est)[1] | Co-ordinates   | Burglary | Criminal Damage | Drugs | Other Notifiable Offences | Robbery | Theft and Handling | Violence Against the Person | Total |
|---------|----------------------|-------------------|---|-------------|--------------------------|--|----------|-----------------|-------|---------------------------|---------|--------------------|-----------------------------|-------|
| 0       | Barking and Dagenham | Labour            | Town Hall, 1 Town Square                        | 13.93       | 194352                   | 51°33'39"N<br>0°09'21"E /<br>51.5607°N 0.1557°E<br>f...    | 1287     | 1949            | 919   | 378                       | 534     | 5607               | 6067                        | 16741 |
| 1       | Barnet               | Conservative      | North London Business Park, Oakleigh Road South | 33.49       | 369088                   | 51°37'31"N<br>0°09'06"W<br>f<br>51.6252°N 0.1517°W<br>f... | 3402     | 2183            | 906   | 499                       | 464     | 9731               | 7499                        | 24684 |
| 2       | Bexley               | Conservative      | Civic Offices, 2 Watling Street                 | 23.38       | 236687                   | 51°27'18"N<br>0°09'02"E /<br>51.4549°N 0.1505°E<br>f...    | 1123     | 1673            | 646   | 294                       | 209     | 4392               | 4503                        | 12840 |
| 3       | Brent                | Labour            | Brent Civic Centre, Engineers Way               | 16.70       | 317264                   | 51°33'32"N<br>0°16'54"W<br>f<br>51.5588°N 0.2817°W<br>f... | 2631     | 2280            | 2096  | 536                       | 919     | 9026               | 9205                        | 26693 |
| 4       | Bromley              | Conservative      | Civic Centre, Stockwell Close                   | 57.97       | 317899                   | 51°24'14"N<br>0°01'11"E /<br>51.4039°N 0.0198°E<br>f...    | 2214     | 2202            | 728   | 417                       | 369     | 7584               | 6650                        | 20164 |

#### 2.2.4 Finding the neighbourhood with the safest borough.

Table 4: Neighbourhoods of the safest borough

After visualising the crime in each borough, we will find the lowest crime rate and hence rate that borough as the safest borough. Third source of data is acquired from the list of neighbourhoods in the safest borough on Wikipedia. This dataset is created from scratch. The pandas dataframe is created with the names of the neighbourhoods the name of the borough with the latitude and longitude are left blank.

|   | Neighborhood | Borough              | Latitude | Longitude |
|---|--------------|----------------------|----------|-----------|
| 0 | Berrylands   | Kingston upon Thames |          |           |
| 1 | Canbury      | Kingston upon Thames |          |           |
| 2 | Chessington  | Kingston upon Thames |          |           |
| 3 | Coombe       | Kingston upon Thames |          |           |
| 4 | Hook         | Kingston upon Thames |          |           |

Using Google Maps API geocoding for the final dataset

#### 2.2.5 Merging dataframes to get a single dataframe.

Table 5: Neighbourhoods with safest borough

The coordinates of the neighbourhoods are obtained using Google Maps API geocoding get the final dataset

|   | Neighborhood | Borough              | Latitude  | Longitude |
|---|--------------|----------------------|-----------|-----------|
| 0 | Berrylands   | Kingston upon Thames | 51.393781 | -0.284802 |
| 1 | Canbury      | Kingston upon Thames | 51.417499 | -0.305553 |
| 2 | Chessington  | Kingston upon Thames | 51.358336 | -0.298622 |
| 3 | Coombe       | Kingston upon Thames | 51.419450 | -0.265398 |
| 4 | Hook         | Kingston upon Thames | 51.367898 | -0.307145 |

This will be used to get the venues for each neighbourhood using the Foursquare API.

Foursquare location data will be leveraged to explore or compare boroughs around London.

Data manipulation and analysis to derive subsets of the initial data.

## 2.3 Feature Selection

After data cleaning, the following features will be kept and the others dropped.

Feature selection during data cleaning

| Kept features                       | Dropped features                  | Reason for dropping features                         |
|-------------------------------------|-----------------------------------|--|
| Categories of Crime (major , minor) | Isoa code                         | This did not add value to the prediction             |
| Borough                             |                                   |  |
| Number of crimes                    |                                   |  |
| Year (recent)                       | Month, years before 2016 rejected | This is to give a recent set of figures to work with |

## Methodology

The methodology in this project consists of two parts:

1. Exploratory Data Analysis: Visualise the crime rates in the London boroughs for chain restaurant owners to identify the safest borough and extract the neighborhoods in that borough to find the 10 most common venues in each neighborhood
2. Modelling: To help chain restaurant owners to find other amenities in the safest boroughs. We will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar amenities. This is so that chain restaurant owners can look at other amenities that go hand in hand whilst running a food business, such as transport and other businesses in the area.

### 3 Exploratory Data Analysis

Using descriptive statistics, the borough with the highest number of crimes is obtained. The advice then is to avoid this particular borough,

- **Visualization:**
  - Analysis and plotting visualizations.
  - Data visualization using various mapping libraries.

K-means Clustering is used to find neighbourhood details attached with each clusters..

- **Discussion and Conclusions:**
  - Recommendations and results based on the data analysis.
  - Discussion of any limitations and how the results can be used, and any conclusions that can be drawn.