

## **Description of the data and how it will be used to solve the problem**

### **Data Acquisition:**

In this project three combination of data used from different 3 sources. The first data set The crime statistics dataset of London is used from Kaggle. This data shows the crime per borough in London. The data descriptions are following:

- 1. First Data source:** This data counts the number of crimes at two different geographic levels of London (LSOA and borough) by year, according to crime type. The dataset contains the following columns.
  - lsoa\_code: code for [Lower Super Output Area](#) in Greater London.
  - borough: Common name for London borough.
  - major\_category: High level categorization of crime
  - minor\_category: Low level categorization of crime within major category.
  - value: monthly reported count of categorical crime in given borough
  - year: Year of reported counts, 2008-2016
  - month: Month of reported counts, 1-12

The list of the crime types covered by major Category and minor Category

- 2. Second Data source:** The second source of data is scraped from a Wikipedia page that contains [the list of London boroughs](#). The following columns are used for analysis.
  - Borough: The names of the 33 London boroughs
  - Inner: Categorizing the borough as an Inner London borough or an Outer London Borough
  - Status: Categorizing the borough as Royal, City or other borough
  - Local authority: The local authority assigned to the borough
  - Political control: The political party that control the borough
  - Headquarters: Headquarters of the Boroughs
  - Area (sq. mi): Area of the borough in square miles
  - Population (2013 Est)[1]: The population in the borough recorded during the year 2013.
  - Co-ordinates: The latitude and longitude of the boroughs
  - Nr. in map: The number assigned to each borough to represent visually on a map
- 3. Third Data source:** The third data source is the [list of Neighbourhoods in the Royal Borough of Kingston upon Thames](#) as found on a Wikipedia page. This dataset is created from scratch using the list of neighbourhood available on the site, the following are columns:
  - Neighbourhood: Name of the neighbourhood in the Borough.
  - Borough: Name of the Borough.
  - Latitude: Latitude of the Borough.
  - Longitude: Longitude of the Borough

## Data Preprocessing:

Table-1 shows the London crime data, most recent year (2016).

**Table1: London crime data of 2016**

	Isa_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

The data preparation for each of the three sources of data is done separately. From the London crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per borough as per the category(see Table-2).

**Table2: London crime data after pre-processing**

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741
1	Barnet	3402	2183	906	499	464	9731	7499	24684
2	Bexley	1123	1673	646	294	209	4392	4503	12840
3	Brent	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	2214	2202	728	417	369	7584	6650	20164

The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (see Table-3). This is important because we will be merging the two datasets together using the Borough names.

**Table-3: List of London Boroughs**

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	Barking and Dagenham	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E	25
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W	31
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E	23
3	Brent	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264	51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W	12
4	Bromley	NaN	NaN	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E	20

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset (see **Table-4**). The purpose of this dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

**Table-4: London Borough Crime**

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	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	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E	1287	1949	919	378	534	5607	6067	16741
1	Barnet	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W	3402	2183	906	499	464	9731	7499	24684
2	Bexley	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E	1123	1673	646	294	209	4392	4503	12840
3	Brent	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264	51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E	2214	2202	728	417	369	7584	6650	20164

After visualizing the crime in each borough we can find the borough with the lowest crime rate and hence tag that borough as the safest borough. The 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 data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude. The coordinates of the neighbourhoods is be obtained using Google Maps API geocoding to get the final dataset. The new dataset is used to generate the venues for each neighbourhoods using the Foursquare API(see **Table-5**).

**Table-5: Neighbourhoods of the safest borough**

	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
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898
9	New Malden	Kingston upon Thames	51.405335	-0.263407
10	Norbiton	Kingston upon Thames	51.409999	-0.287396