

2. Data Description and Preparation

2.1 Data Description

The data used in this project has been acquired from three different sources. The first data source is that of London crime data, showing crime in each borough of London. The dataset consists of following columns:

- **Isao_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 second dataset has been scraped from the wikipedia page containing the list of London boroughs. It contains additional information about the boroughs with following columns:

- **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 controls 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

The third data source is the wikipedia page providing the list of Neighborhoods in the Royal Borough of Kingston upon Thames. It has following columns:

- **Neighborhood**: Name of the neighborhood in the borough
- **Borough**: Name of the borough
- **Latitude**: Latitude of the borough
- **Longitude**: Longitude of the borough

2.2 Data Preparation

Each of the three datasets have been cleaned separately. The crimes pertaining to the most recent year (2016) have been selected from London crime data. The major categories of crime are pivoted to get the total crimes for each borough as per the category. (see *fig 2.1*)

	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

Fig 2.1 London crime data after preprocessing

The python library **Beautiful Soup** has been used for scraping the second data from wikipedia page. We have extracted the data in a tabular format using this library. After web scraping, we need to manipulate the string in order to get the names of the boroughs in appropriate form (see *fig 2.2*). This step helps us afterwards in merging the two datasets together using borough names.

	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	North London Business Park, Oakleigh Road South	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

Fig 2.2 List of London Boroughs

We merge the two datasets using borough names resulting into a new dataset that combines the necessary information in one dataset (see *fig 2.3*). This dataset will be used to visualize the crime rates in each borough and identify the borough that recorded the least number of crimes during the year 2016.

	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	North London Business Park, Oakleigh Road South	33.49	369088	51°37'31"N 0°09'08"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

Fig 2.3 London Borough Crime

After visualizing the crime in each borough, we can identify the borough with least crime rate and mark that borough as the safest borough. The third dataset has been acquired from the list of neighborhoods in the safest borough on wikipedia. This dataset has been created from scratch. The pandas dataframe has been created with the neighborhood names and borough name with blank latitude and longitude columns (see fig 2.4).

	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		

Fig 2.4 Neighborhoods of the safest borough

The neighborhood coordinates can be obtained using **Google Maps API Geocoding** for the final dataset (see fig 2.5).

	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

Fig 2.5 Neighborhoods of the safest borough

The final dataset will be used to generate the venues for all neighborhoods using **Foursquare API**.