The Battle of Neighbourhoods on London Crime Data

1 Introduction

I. Background:

From 2010 to 2016, an estimated annual average of 495,000 non-British citizens moved to the UK with the intention of staying for 12 months or longer. An estimated annual average of 190,000 left the country to live abroad during the same period. Now questions comes

Does peoples move until they find a place to settle down where they truly happy?

Is the place is safe for them?

Is there other facility available like school garden, etc.?

So without knowing about the area what they are getting into, forcing them to turn tail and run at the first sign of discomfort?

To minimize the chances of this happening we should always do proper research where planning our next move in life. Consider factors when picking a new place for living so you don't end up wasting your valuable time and money making a move you will end you regretting. Safety is the top concern when moving to a new area.

II. Problem statement:

London is one of the most ethnically diverse cities in the world. London had a population of 8,173,941 according to 2011 census,. Of this number 44.9% were White British. 37% of the population were born outside the UK, including 24.5% born outside of Europe. From the above statistic it was seen that people are move to London for better life. But the crime rate is increased in London . In 2015/16 the crime rate in London was 82.26 per 1000 peoples. In 2019/20 the crime rate in London stood at 101.48 crimes committed per thousand people.

This is the overall statistic in London but the crime rates in each brough have not same. It may be changed over time and brough.

This projects aims to find out the least crime rate brough on basis of total crimes. Explore the neighbourhoods of that brough to find out the 10 most common venues in each neighbourhood. Finally clustering was done by K Means clustering. The crime statistics dataset of London is used for this project on Kaggle. The year 2016 is consider for analysis.

III. Interest:

Peoples who are want to relocate to London will be interested to identify the safest borough in London and explores its neighbourhoods and common venues around each neighbourhood.

2 Data Acquisition and Prepossessing

I. 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:

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

Second Data source: The second source pf 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 year2013.
- Co-ordinates: The latitude and longitude of the boroughs
- Nr. in map: The number assigned to each borough to represent visually on a map

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

II. Data Prepossessing:

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

Table1: London crime data of 2016

	lsoa_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

3 Methodology

I. Exploratory Data Analysis:

1) Descriptive statistics of crimes:

The describe function in python is used to get statistics of the London crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile(50%), and the 3rd quartile (75%) for each of the major categories of crime. The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.

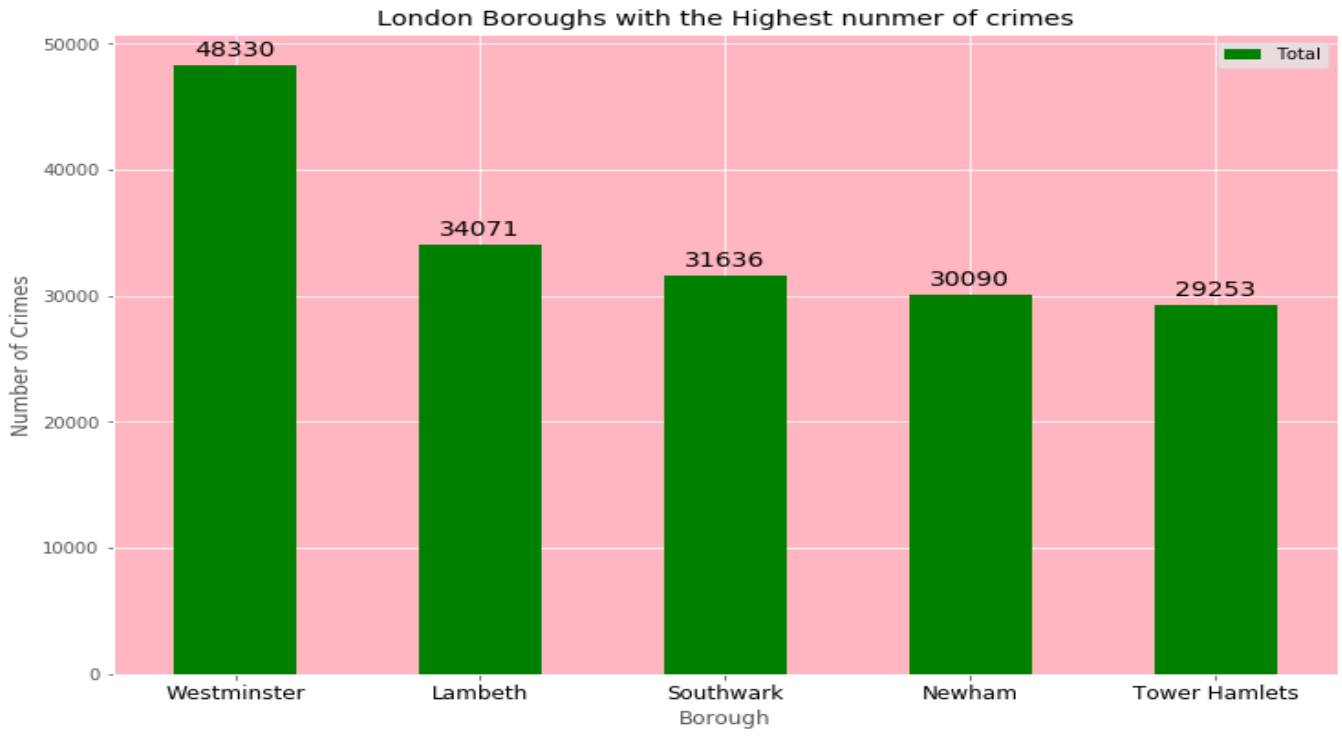
Table-6: Descriptive statistics of crimes

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

2) Boroughs with the highest crime rates:

Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 borough.

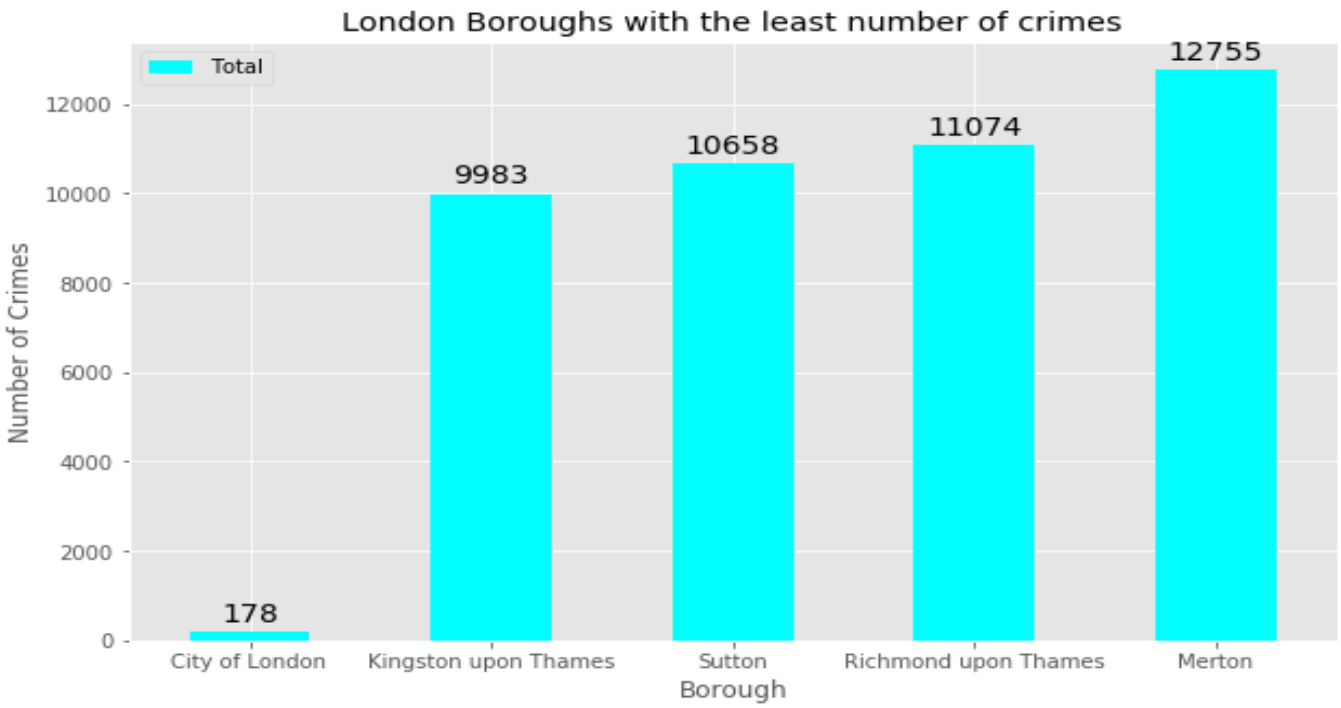
Fig-1: Boroughs with the highest crime rates



3) Boroughs with the lowest crime rate:

Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton (see **Fig-2**)

Fig-2: Boroughs with the lowest crime rate



City of London has a significantly lower crime rate because it i is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area (see **Table-7**). Hence we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames

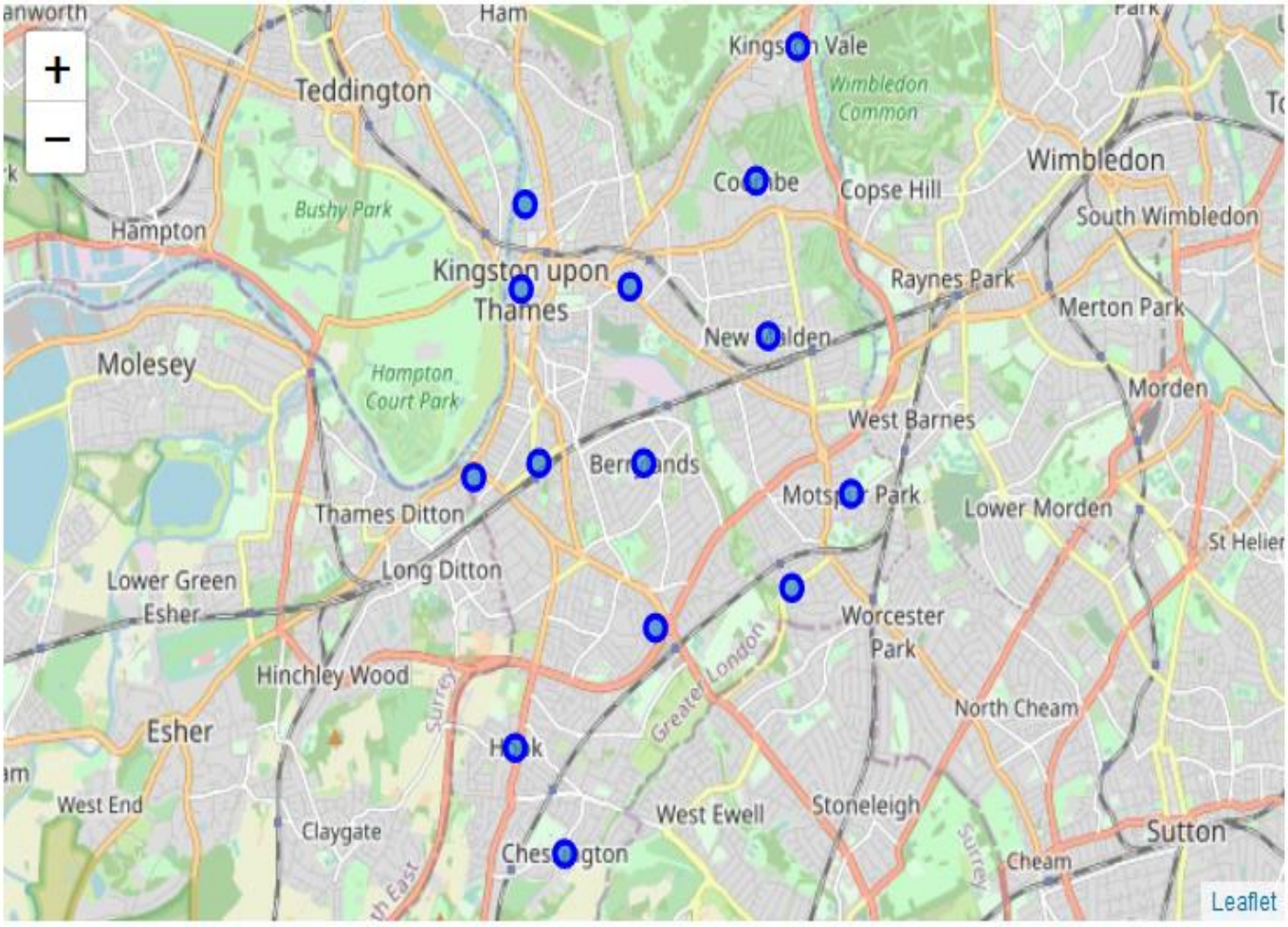
Table-7: City of London

	Borough	Total	Area (sq mi)	Population (2013 est)[1]
6	City of London	178	1.12	7000

4) Neighbourhoods in Kingston upon Thames:

There are 15 neighbourhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python (see **Fig-3**)

Fig-3: Neighbourhoods in Kingston upon Thames



II. Modelling

Using the final dataset containing the neighbourhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighbourhood by connecting to the Foursquare API. This returns a json file containing all the

venues in each neighbourhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category.

One hot encoding is done on the venues data(One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms is a better job in prediction). The Venues data is then grouped by the Neighbourhood and the mean of the venues are calculated, finally the 10 common venues are calculated foreach of the neighbourhoods' help people find similar neighbourhoods in the safest borough we will be clustering similar neighbourhoods 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 neighbourhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighbourhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighbourhood.

Table-8: Venue details of each Neighbourhood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	Alexandra Park	51.394230	-0.281206	Park
2	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
3	Berrylands	51.393781	-0.284802	Kamala Food and Wine	51.397810	-0.284045	Wine Shop
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

4 Results

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five clusters. Looking into the neighbourhoods the first cluster.

Table-8: Cluster 1

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	0	Pub	Food	Train Station	Department Store	Dessert Shop	Discount Store	Donut Shop	Electronics Store	English Restaurant	Falafel Restaurant

The second cluster is the biggest cluster with 9 of the 15 neighbourhoods in the borough Kingston upon Thames. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Pubs, Gym, Cafe, Supermarkets, and Grocery Store.

Table-9: Cluster 2

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Hotel	Supermarket	Indian Restaurant	Fish & Chips Shop	Park	Plaza	Café	Shop & Service	Spa
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Women's Store	Food	Dessert Shop	Discount Store	Donut Shop	Electronics Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Café	Clothing Store	Italian Restaurant	Pub	Bakery	Department Store	Hotel	Ice Cream Shop	Lebanese Restaurant
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Pub	Grocery Store	Restaurant	Garden Center	Fast Food Restaurant	Department Store	Dessert Shop	Discount Store	Donut Shop	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gym	Indian Restaurant	Bar	Gastropub	Sushi Restaurant	Supermarket	Korean Restaurant	Fish & Chips Shop	Dessert Shop	Discount Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Food	Italian Restaurant	Pub	Hotel	Breakfast Spot	Fried Chicken Joint	Hardware Store	Wine Shop	Japanese Restaurant
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Pub	Coffee Shop	Gym	Fast Food Restaurant	Playground	Pet Café	Restaurant	Chinese Restaurant	Café
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	Gym / Fitness Center	Gastropub	French Restaurant	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Bowling Alley	Pizza Place	Coffee Shop	Discount Store	Café	Restaurant	Sandwich Place	Bus Stop

The third cluster has one neighbourhood which consists of Venues such as Grocery shops, Bars, Electronics Store, Women's Store, and Fish & Chips Shop.

Table-10: Cluster 3

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	2	Grocery Store	Bar	Sandwich Place	Soccer Field	Women's Store	Fish & Chips Shop	Dessert Shop	Discount Store	Donut Shop	Electronics Store

The fourth cluster has one neighbourhood which consists of Venues such as Dessert Shop, Tea Room , Electronics Store, Women's Store, and Discount Store.

Table-11: Cluster 4

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Coombe	Kingston upon Thames	51.41945	-0.265398	3	Tea Room	Women's Store	French Restaurant	Dessert Shop	Discount Store	Donut Shop	Electronics Store	English Restaurant	Falafel Restaurant	Farmers Market

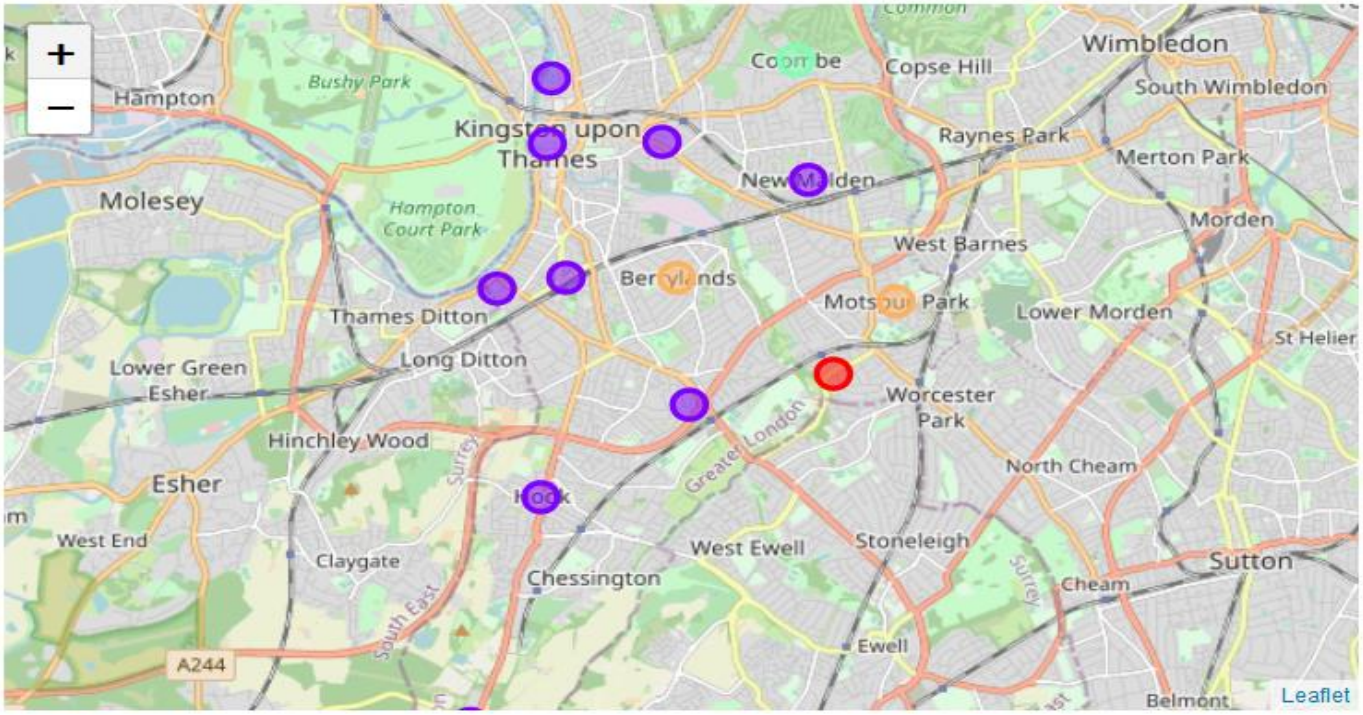
The fifth cluster has two neighbourhoods in it, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

Table-12: Cluster 5

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	4	Wine Shop	Gym / Fitness Center	Park	Bus Stop	Food	Discount Store	Donut Shop	Electronics Store	English Restaurant	Falafel Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Gym	Park	Bus Stop	Soccer Field	Women's Store	Fast Food Restaurant	Dessert Shop	Discount Store	Donut Shop	Electronics Store

Visualising the clustered neighbourhoods on a map using the folium library.

Fig-4: Clustered neighbourhoods in the Borough of Kingston upon Thames



5 Discussion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighbourhoods to which they want to relocate based on the most common venues in it. For example if a person is looking for a neighbourhoods with good connectivity and public transportation we can see that Clusters 1,2 and 5 have Train stations and Bus stops. If a person is looking for a neighbourhood with stores and restaurants in a close proximity then the neighbourhoods in the third cluster is suitable. For a family I feel that the neighbourhoods in Cluster 5 are more suitable dues to the common venues in that cluster, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.

6 Conclusion

This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhoods. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighbourhoods. We have just taken safety as a primary concern to shortlist the borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough based on safety and a predefined budget.