Capstone Project - The Battle of Neighbourhoods

Report

Nitesh Ranjan Singh

University of Petroleum and Energy Studies, Dehradun-248001

November 2020

1. <u>Introduction</u>

1.1 Background

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will satisfy us? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort?

To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move, you'll end up regretting. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

1.2 Problem

The crime statistics dataset of London found on Kaggle has crimes in each borough of London from 2008 to 2016. The year 2016 is the latest we will be considering the data of that year which is old information as of now. The crime rates in each borough may have changed over time.

This project aims to select the safest borough in London based on the total crimes, explore the neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering.

1.3 Interest

Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around each neighbourhood.

2. <u>Data Acquisition and Cleaning</u>

2.1 Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a London crime data that shows the crime per borough in London.

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 a major category.
- ❖ **value**: monthly reported count of categorical crime in a given borough.
- ❖ **year**: Year of reported counts, 2008-2016
- ❖ **month**: Month of reported counts, 1-12

The second source of data is scraped from a Wikipedia page that contains the list of London boroughs.

This page contains additional information about the boroughs, the following are the 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 boroughs.
- ❖ **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 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 neighbourhoods available on the site, the following are columns:

- ❖ **Neighbourhoods:** Name of the neighbourhood in the Borough.
- ❖ **Borough**: Name of the Borough.
- ❖ **Latitude:** Latitude of the Borough.
- ❖ **Longitude:** Longitude of the Borough.

2.2 Data Cleaning

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 the boroughs** for each major 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	93	130	51	28	33	336	412	1083
1	Barnet	232	154	57	37	24	628	514	1646
2	Bexley	70	129	58	17	9	319	312	914
3	Brent	198	154	136	37	48	591	596	1760
4	Bromley	147	134	56	30	19	487	428	1301

Fig 2.1 London crime data after pre-processing

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 on the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (see fig 2.2). This is important because we will be merging the two datasets using the 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 [note 1]	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

Fig 2.2 List of London Boroughs

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset (see fig 2.3). 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.

	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	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	232	154	57	37	24	628	514	1646
1	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	70	129	58	17	9	319	312	914
2	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	198	154	136	37	48	591	596	1760
3	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	147	134	56	30	19	487	428	1301
4	Camden	Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	229719	51°31′44″N 0°07′32″W / 51.5290°N 0.1255°W	177	163	166	31	72	804	541	1954

Fig 2.3 London Borough Crime

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 left blank (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		
5	Kingston upon Thames	Kingston upon Thames		
6	Kingston Vale	Kingston upon Thames		
7	Malden Rushett	Kingston upon Thames		
8	Motspur Park	Kingston upon Thames		
9	New Malden	Kingston upon Thames		
10	Norbiton	Kingston upon Thames		

Fig 2.4 Neighbourhoods of the safest borough

The coordinates of the neighbourhoods are be obtained using Google Maps API geocoding to get 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
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

Fig 2.5 Neighbourhoods of the safest borough

3. Methodology

3.1 Exploratory Data Analysis (EDA)

3.1.1 Statistical Summary 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 (See fig 3.1.1).

	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	137.242424	132.969697	85.333333	30.969697	44.181818	606.121212	484.696970	1521.515152
std	49.477918	44.057125	50.658579	13.375549	27.881955	309.912262	178.707597	608.764020
min	0.000000	0.000000	0.000000	0.000000	2.000000	5.000000	1.000000	8.000000
25%	99.000000	100.000000	57.000000	21.000000	22.000000	392.000000	397.000000	1158.000000
50%	142.000000	134.000000	78.000000	31.000000	43.000000	591.000000	514.000000	1554.000000
75%	176.000000	154.000000	129.000000	41.000000	63.000000	735.000000	618.000000	1909.000000
max	232.000000	211.000000	235.000000	60.000000	115.000000	1790.000000	759.000000	3249.000000

Fig 3.1.1 Statistical description of the London crimes

3.1.2 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 boroughs (see fig 3.1.2).
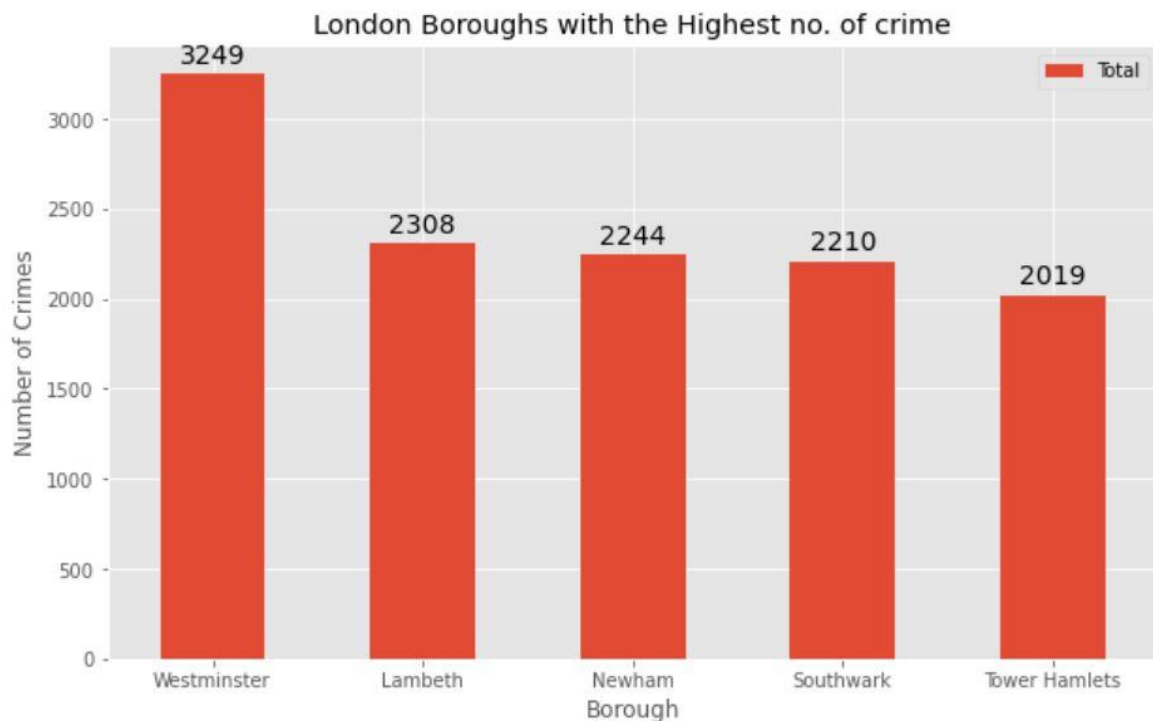

Fig 3.1.2 Boroughs with the highest crime rates

3.1.3 Boroughs with the lowest crime rates

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 3.1.3).
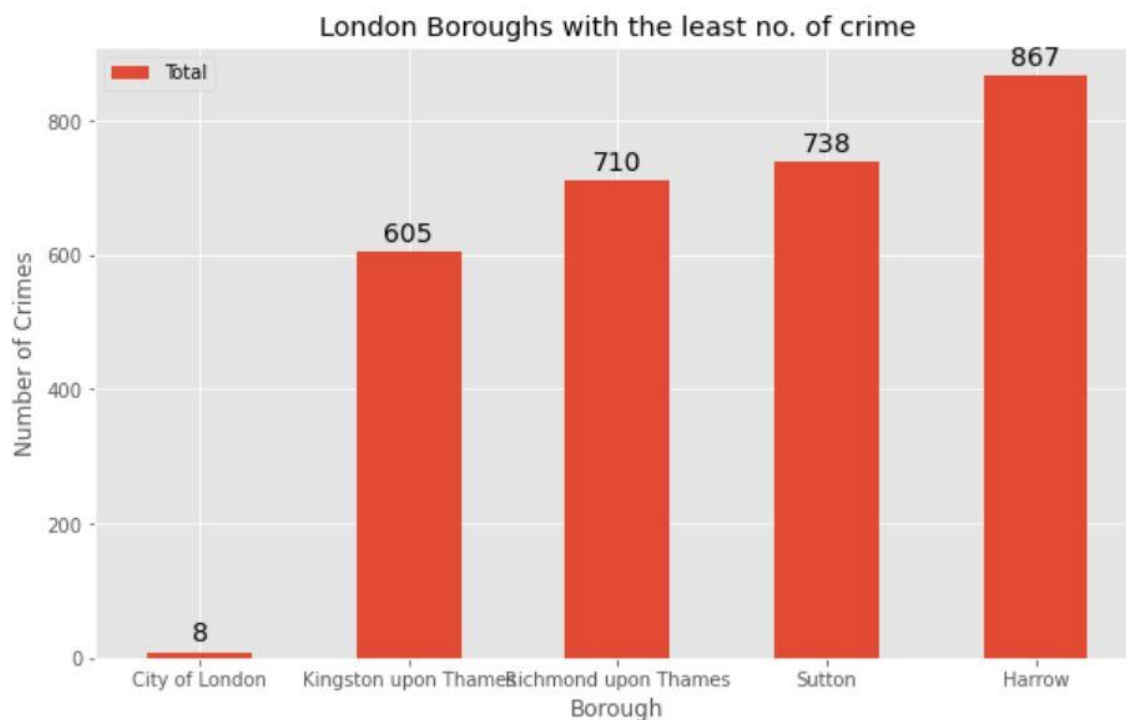

Fig 3.1.3 Boroughs with the lowest crime rates

The city of London has a significantly lower crime rate because it 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 fig 3.1.3.1). Hence, we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

	Borough	Total	Area (sq mi)	Population (2013 est)[1]
5	City of London	8	1.12	7000

Fig 3.1.3.1 City of London

3.1.4 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 3.1.4).
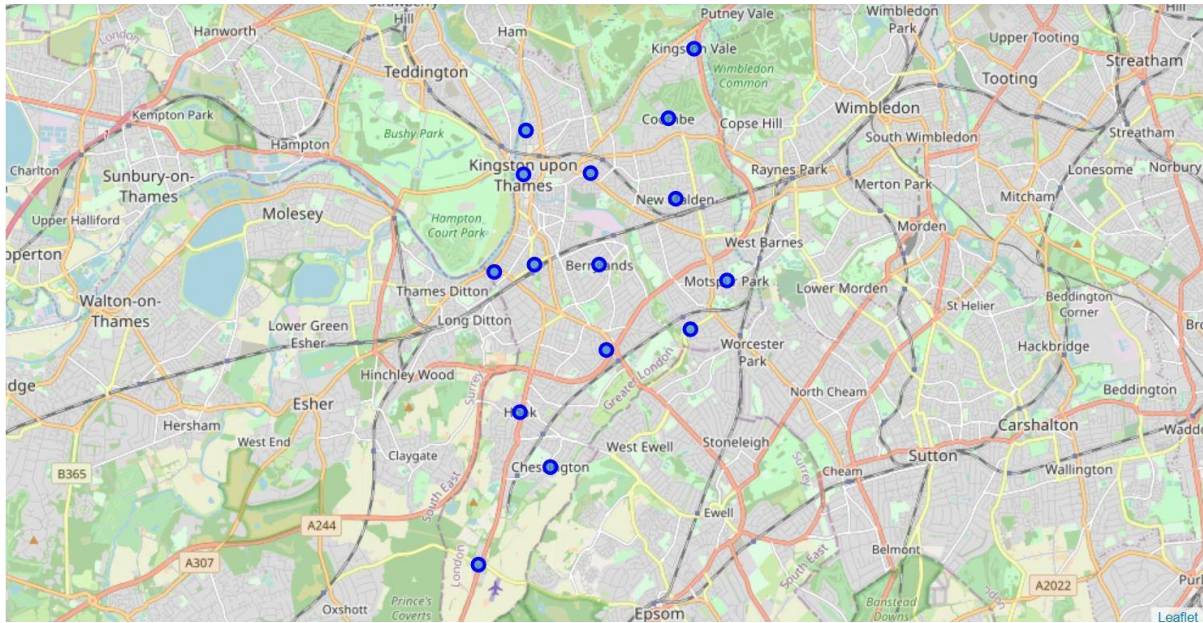
Fig 3.1.4 Neighbourhoods in Kingston upon Thames

3.2 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 (see fig 3.2.1).

	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	Cafe Rosa	51.390175	-0.282490	Café
4	Berrylands	51.393781	-0.284802	Kamala Food and Wine	51.397810	-0.284045	Wine Shop

Fig 3.2.1 Venue details of each neighbourhood

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 to do a better job in prediction). The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally, the 10 common venues are calculated for each of the neighbourhoods.

To 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.

4. <u>Results</u>

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 in the 1st cluster (see fig 4.1)

	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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	0	Park	Wine Shop	Fast Food Restaurant	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Farmers Market	Fish & Chips Shop

Fig 4.1 Cluster 1

. Looking into the neighbourhoods in the 2nd cluster (see fig 4.2)

	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 Con V
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Café	Shop & Service	Gym / Fitness Center	
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Fish & Chips Shop	Pub	Bakery	Supermarket	Indian Restaurant	Fast Food Restaurant	Deli / Bodega	Department Store	Cl
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Pub	Sushi Restaurant	Café	Burger Joint	Asian Restaurant	Department Store	Gym / Fitness Center	M
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Korean Restaurant	Gastropub	Sushi Restaurant	Supermarket	Bar	Indian Restaurant	Wine Shop	Farmers Market	B
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Food	Italian Restaurant	Pub	Wine Shop	Japanese Restaurant	Hotel	Hardware Store	Ch
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Restaurant	Pet Café	Chinese Restaurant	Café	Italian Restaurant	
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Thai Restaurant	Tea Room	Gastropub	S

Fig 4.2 Cluster 2

. Looking into the neighbourhoods in the 3rd cluster (see fig 4.3)

	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 Mo Comm Ven
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Grocery Store	Sandwich Place	Bar	Soccer Field	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner	Electron St
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Garden Center	Pub	Restaurant	Electronics Store	Cosmetics Shop	Deli / Bodega	Department Store	Clean
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Restaurant	Pharmacy	Café	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Cof Sh

Fig 4.3 Cluster 3

. Looking into the neighbourhoods in the 4th cluster (see fig 4.4)

	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	1 C
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Wine Shop	Gym / Fitness Center	Park	Café	Bus Stop	Farmers Market	Department Store	Dry Cleaner	Electronics Store	F R
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Wine Shop	Farmers Market	Deli / Bodega	Department Store	Dry Cleaner	E

Fig 4.4 Cluster 4

. Looking into the neighbourhoods in the 5th cluster (see fig 4.5)

	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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	4	Train Station	Pub	Food	Construction & Landscaping	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner

Fig 4.5 Cluster 5

Visualising the clustered neighbourhoods on a map using the folium library (see fig 4.6)
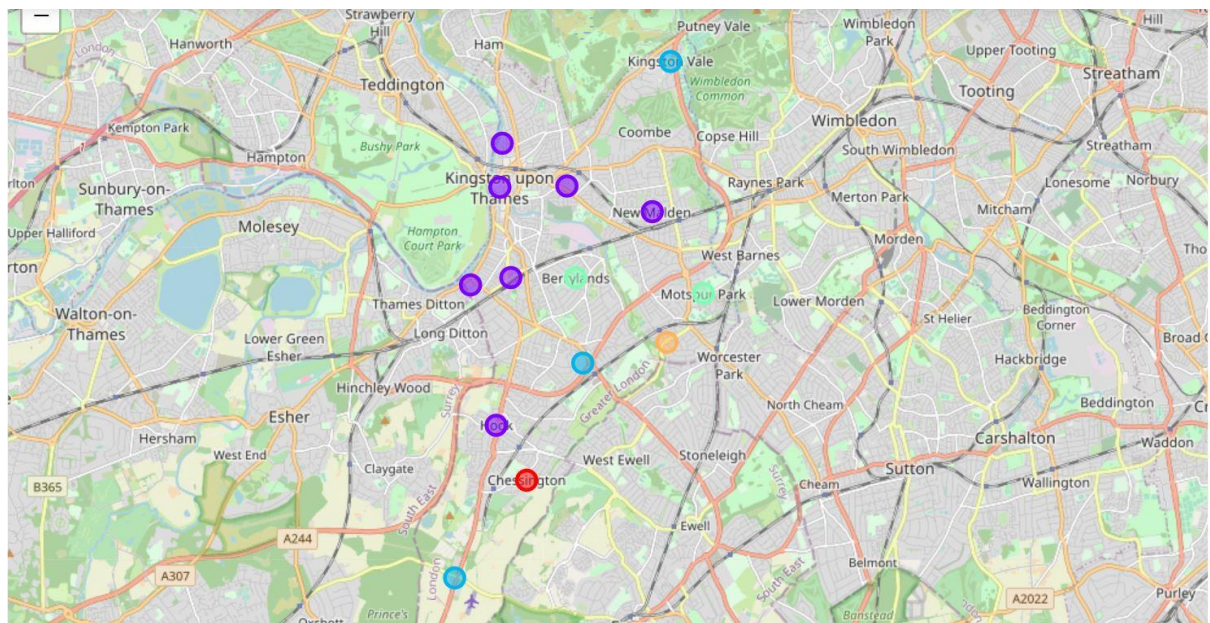

Fig 4.6 Clustered neighbourhoods in the Borough of Kingston upon Thames

Each cluster is colour coded for the ease of presentation; we can see that majority of the neighbourhood falls in the purple cluster which is the 2nd cluster. Two neighbourhoods have their cluster (Red and Yellow), these are clusters 1st and 5th. The green cluster consists of two neighbourhoods which is the 4th cluster, the blue cluster consists of three neighbourhoods which is the 3rd cluster.

5. Discussion

This project aims to help people who want to relocate to the safest borough in London, ex-pats can choose 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 neighbourhood with good connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues. If a person is looking for a neighbourhood with stores and restaurants in proximity then the neighbourhoods in the first cluster are suitable. For a family I feel that the neighbourhoods in Cluster 4 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. The choices of neighbourhoods may vary from person to person.

6. Conclusion

This project helps a person get a better understanding of the neighbourhoods concerning the most common venues in that neighbourhood. 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 neighbourhood. We have just taken safety as a primary concern to shortlist the safest borough of London. The future of this project includes taking other factors such as the cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.