5th July 2020

The Battle of London Neighborhoods

-The Capstone Project

Sunil M. Daware

1. INTRODUCTION

1.1 Background

With globalization more and more people are migrating from one country to another for green pasture. Due recent change in immigration policy by USA, there is spurt in immigration from India to London. Even though, new country offers lots of economic, social benefits along with great carrier opportunities, settling into new place is challenging decision, as one do not know, what that place is going to be, how will be locality from safety point of view, whether all amenities will be there or not etc.

To minimize the chances of this happening, one 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 do not end up wasting your valuable time and money making a move one will end up regretting. Safety is a top concern when moving to a new area. If one does not feel safe in your own home, you are not going to be able to enjoy living there.

1.2 Problem

A local Immigration consulting firm, which assist Students and Immigrants relocate to various countries, seeking my expertise to assist their client to identify the best Borough/Neighborhood to settle in London.

This report will be targeted to people who are looking to relocate to London. To finalize a neighborhood to hunt for an apartment, safety is considered as a top concern when moving to a new place. If one does not feel safe in own home, he/she is not going to be able to enjoy living there. The crime statistics will provide an insight into this issue.

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

We will focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood suited to an individual's needs can be selected.

The crime statistics dataset of London was found on Kaggle has crimes in each Boroughs of London from 2008 to 2016. The year 2016 being 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, then explore the neighborhoods of that borough, using FOURSQUARE API, to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

2. DATA ACQUISITION & CLEANING

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 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 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 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.

The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a Wikipedia page. This dataset is created from scratch using the list of neighborhoods available on the site, the following are 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 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 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 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 *fig 2.2*). This is important because we will be merging the two datasets together 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	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

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	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′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

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 neighborhoods in the safest borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighborhoods and the name of the borough with the latitude and longitude are 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		

Fig 2.4 Neighborhoods of the safest borough

The coordinates of the neighborhoods is 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

Fig 2.5 Neighborhoods of the safest borough

The new dataset is used to generate the venues for each neighborhood using the Foursquare API.

3. Methodology

3.1 Exploratory Data Analysis

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*).

```
London_crime.describe()
```

	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

Fig 3.1.1 Statistical description of the London crimes

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.

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*).

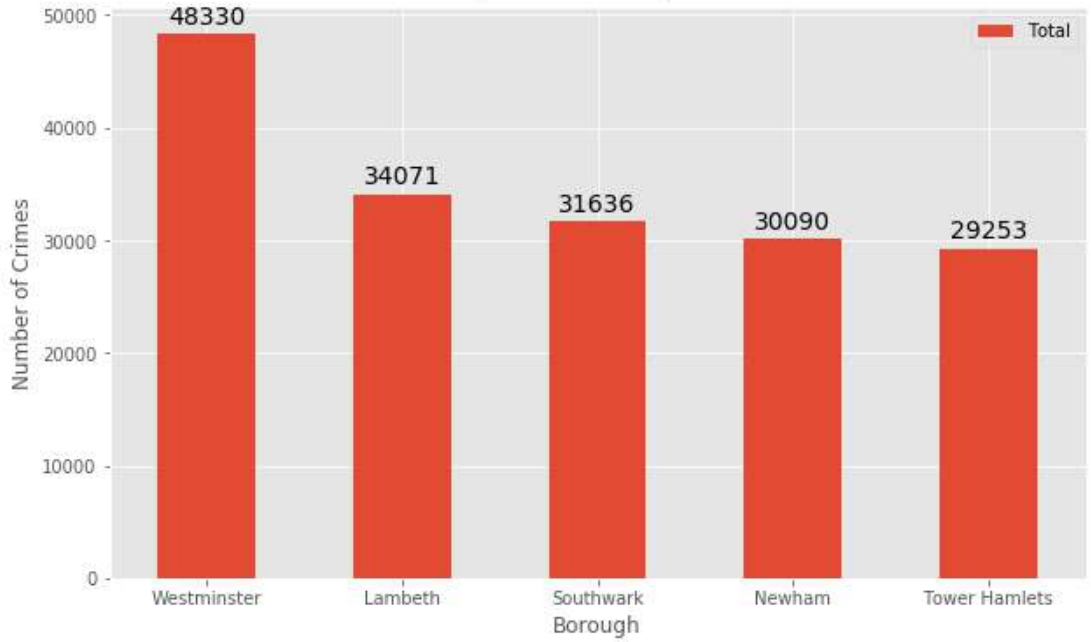

Fig 3.1.2 Boroughs with the highest crime rates

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*).

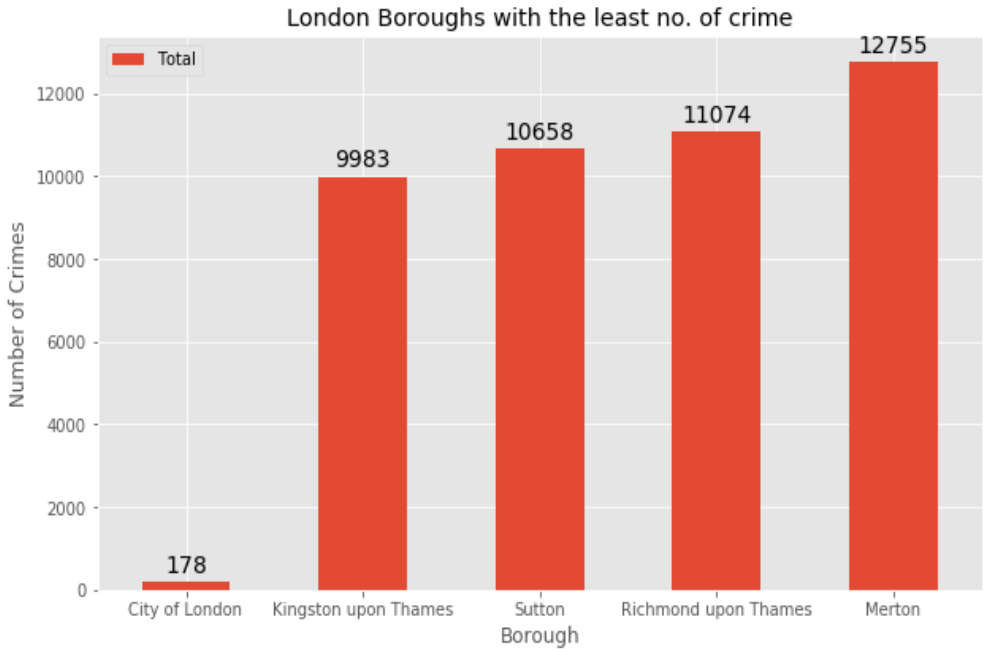

Fig 3.1.3 Boroughs with the lowest crime rates

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 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	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	City of London	([note 5]	Sui generis;City;Ceremonial county	Corporation of London;Inner Temple;Middle Temple	?	Guildhall	1.12	7000	51°30′56″N 0°05′32″W / 51.5155°N 0.0922°W	1

Fig 3.1.3.1 City of London

3.1.4 Neighborhoods in Kingston upon Thames

There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualized on a map using folium on python (*see fig 3.1.4*).

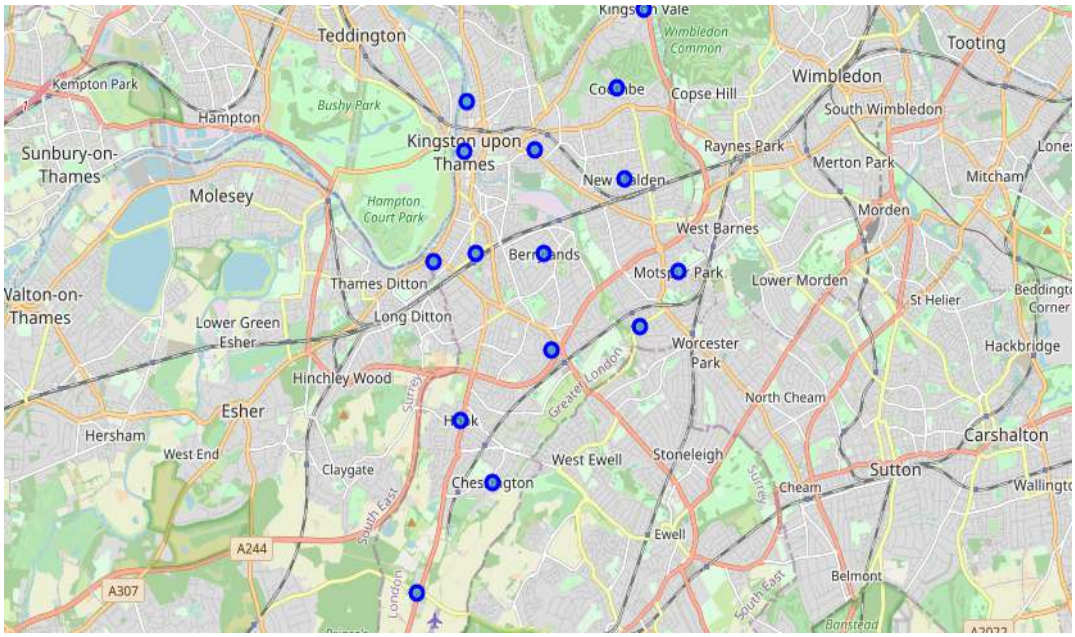

Fig 3.1.4 Neighborhoods in Kingston upon Thames

3.2 Modelling

Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a pandas dataframe. 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	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Fig 3.2.1 Venue details of each Neighborhood

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 neighborhoods.

To help people find similar neighborhoods in the safest borough 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 venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

4. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Looking into the neighborhoods in the first 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	
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	0	Construction & Landscaping	Pub	Food	Train Station	Bakery	Department Store	German Restaurant	C

Fig 4.1 Cluster 1

The First cluster has one neighborhood which consists of Venues such as Construction & Landscaping, Pub, Food add Train Station.

	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 M Com Ve
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Fish & Chips Shop	Supermarket	Spa	Plaza	Shop & Service	Café
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Pub	Burger Joint	Sushi Restaurant	Coffee Shop	Turkish Restaurant	Electro Store
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	1	Grocery Store	Sandwich Place	Bar	Soccer Field	Department Store	Discount Store	Dry Cleane
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gastropub	Chinese Restaurant	Sushi Restaurant	Supermarket	Bar	Korean Restaurant	Indian Restau
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Italian Restaurant	Food	Platform	Pub	Pharmacy	Hotel
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Café
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Train Station	Thai Restau
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Restaurant	Pharmacy	Train Station	Hotel	Indian Restaurant	Italian Restau

Fig 4.2 Cluster

The Second cluster is the biggest cluster with 8 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

	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 Mc Comm Ven
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Garden Center	Pub	Restaurant	Farmers Market	Department Store	Discount Store	Dry Cleane

Fig 4.3 Cluster 3

The third cluster has one neighborhood which consists of Venues such as Grocery Store, Garden Center, Pub, Restaurants, Farmers Market etc.

	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 Com V
4	Hook	Kingston upon Thames	51.367898	-0.307145	3	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Food	Discount Store	Dry Cleaner	Electr Store

Fig 4.4 Cluster 4

The fourth cluster has one neighborhood in it, this neighborhoods is famous for Indian Restaurant, bakery, Supermarket, Discount Store, Dry Cleaner and Farmers Market etc.

	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 Com Ve
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	4	Gym / Fitness Center	Park	Bus Stop	Turkish Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electro Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Park	Gym	Bus Stop	Soccer Field	Restaurant	Turkish Restaurant	Farmers Market	Discou Store

Fig 4.5 Cluster 5

The fifth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc. (*see fig 4.5*).

Visualizing the clustered neighborhoods on a map using the folium library (see fig 4.6).

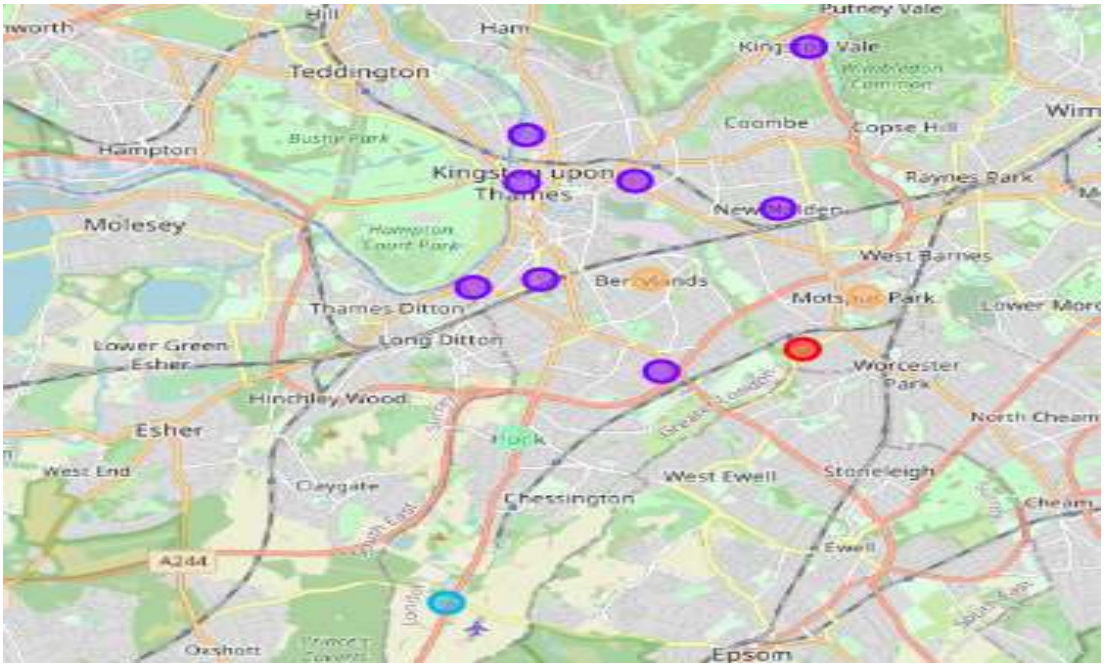

Fig 4.6 Clustered neighborhoods in the Borough of Kingston upon Thame

Each cluster is color coded for the ease of presentation; we can see that majority of the neighborhood falls in the Purple cluster which is the first cluster. Three neighborhoods have their own cluster (Green, Blue, Red and Yellow), these are clusters Four, Three, One and Five. The Yellow cluster consists of two neighborhoods which is the 5th cluster.

5. Discussion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighborhood with good connectivity and public transportation, we can see that Clusters 1, 5 and few neighborhoods in 2 have Train stations and Bus stops as the most common venues. If a person is looking for a neighborhood with stores and restaurants in a proximity, then the neighborhoods in the cluster3 is suitable. For a family I feel that the neighborhoods in Cluster 5 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, 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 neighborhoods with respect to the most common venues in that neighborhood. 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 neighborhood. 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 cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.