

Capstone Project

The Battle of Neighborhoods

Report

Mylavarabhatla Pratyush

May, 2020

1. Introduction

1.1 Background

It is found that people in the UK shift houses a lot. But what causes them to move their houses? Shifting is a very tedious and painstaking process, there must be a strong reason behind their movement. According to a survey which included 1,000 people, slightly more than half said that the chief motivation was for upping sticks. Other popular reasons for moving were access to shops and amenities (35 per cent), being close to family/friends (32 per cent), reducing running costs (31 per cent) and needing a bigger home (30 per cent).

Each year, approximately 4-5% of the population will move to a different city or county in the UK.

London being the city of choice, is quickly increasing in population density. On a yearly basis, nearly 200,000 people move from elsewhere in the UK to live in London, while only 25,000 move the opposite way.

1.2 Problem

A survey of 2,00 people by “Which?” has shown that 18 per cent of the people have trouble finding a home in an ideal area which ticks all the right boxes.

Neighborhood can play a very important role in choosing a home and so this is the problem that I wish to solve.

1.3 Aim

Aim of this project is to find clusters of neighborhoods which are suitable for people to move in.

2. Data Requisites

2.1 Acquiring Data

The data required for this project is compiled from various data sources. The first data source comprises of the London Crime Data which 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 categorizations of crime within 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 Borough.
- **Local authority:** The local authority assigned to the borough.
- **Political control:** The political party that controls the borough.
- **Headquarters:** Headquarters of the borough.

- **Area (sq. mi):** Area of the borough in square miles.
- **Population (2013 Est):** 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** found on the Wikipedia page. This dataset is created from scratch using the list of neighborhoods available on the site. Columns are as follows:

- **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

Preparation of the data for all the 3 sources of data are done individually. From the London Crimes dataset, we have only considered the most recent year i.e. 2016.

	Borough	Burglary	Criminal_damage	Drugs	Other_notifiable_offense	Robbery	Theft_and_handling	Violence_against_the_person	Total_crime
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 data preprocessing

The List of London Boroughs dataset is scraped from the Wikipedia page using Beautiful Soup library. After the dataset is scraped from the website, the dataset is processed to get the names of the Boroughs in correct form (see *fig 2.2*).

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]	Co-ordinates	Nr. in map	Population(2011 est)	Nr. inmap
0	Barking and Dagenham	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352.0	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E	25.0	NaN	NaN
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088.0	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W	31.0	NaN	NaN
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687.0	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E	23.0	NaN	NaN
3	Brent	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers	16.70	317264.0	51°33'32"N 0°16'54"W / 51.5588°N	12.0	NaN	NaN

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 single dataset (see *fig 2.3*). The purpose of this dataset is to visualize the crime rates in each Borough with the least crimes recorded during the year 2016.

Borough	Burglary	Criminal_damage	Drugs	Other_notifiable_offense	Robbery	Theft_and_handling	Violence_against_the_person	Total_crime	Inner	Status	
Barking and Dagenham	1287	1949	919		378	534	5607	6067	16741	NaN	NaN
Barnet	3402	2183	906		499	464	9731	7499	24684	NaN	NaN
Bexley	1123	1673	646		294	209	4392	4503	12840	NaN	NaN
Brent	2631	2280	2096		536	919	9026	9205	26693	NaN	NaN
Bromley	2214	2202	728		417	369	7584	6650	20164	NaN	NaN

Fig 2.3 Crimes in London Borough

3. Exploratory Data Analysis

3.1 Boroughs with Highest crime records

After merging the two datasets we can find the list of Boroughs which have the highest crime records during the year 2016. It is clear 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*).

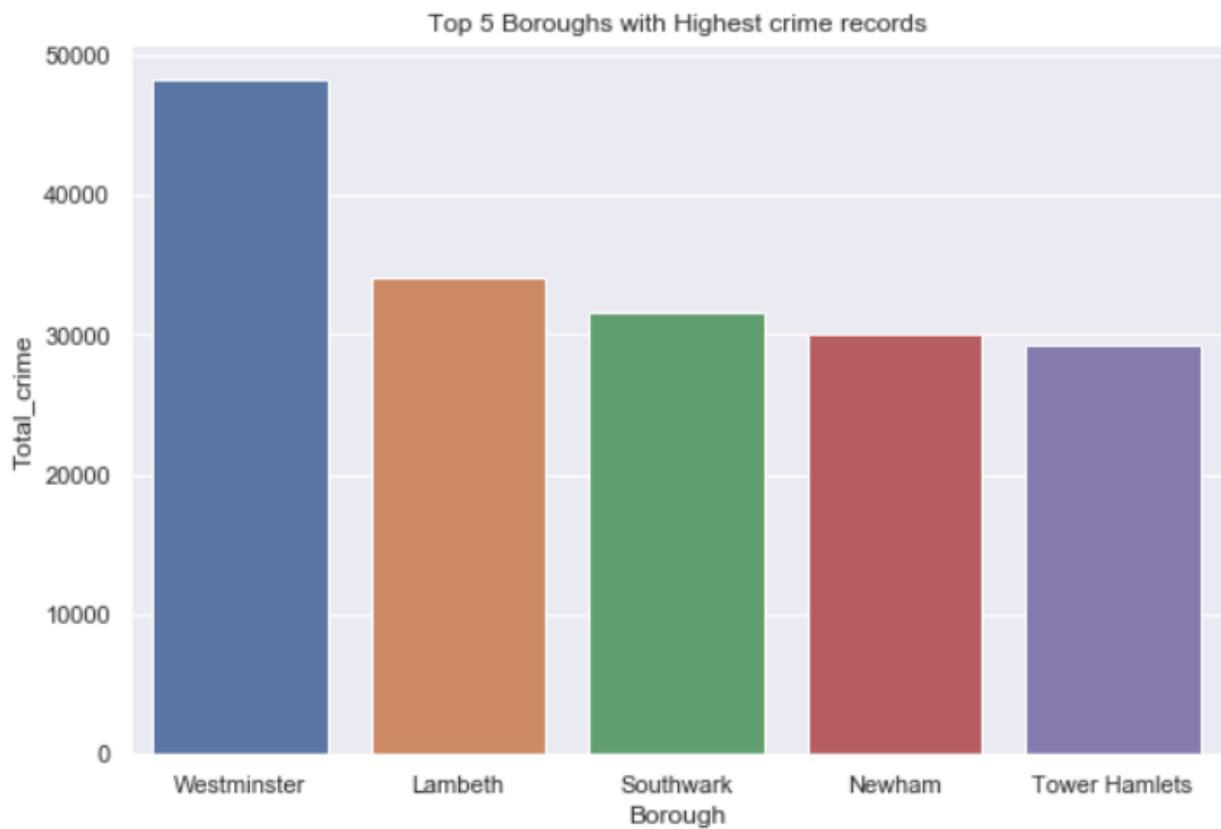


Fig 3.1 Boroughs with highest crime records

3.2 Boroughs with the Lowest crime records

Comparing the five Boroughs with the lowest crime records 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 3.2.1*).

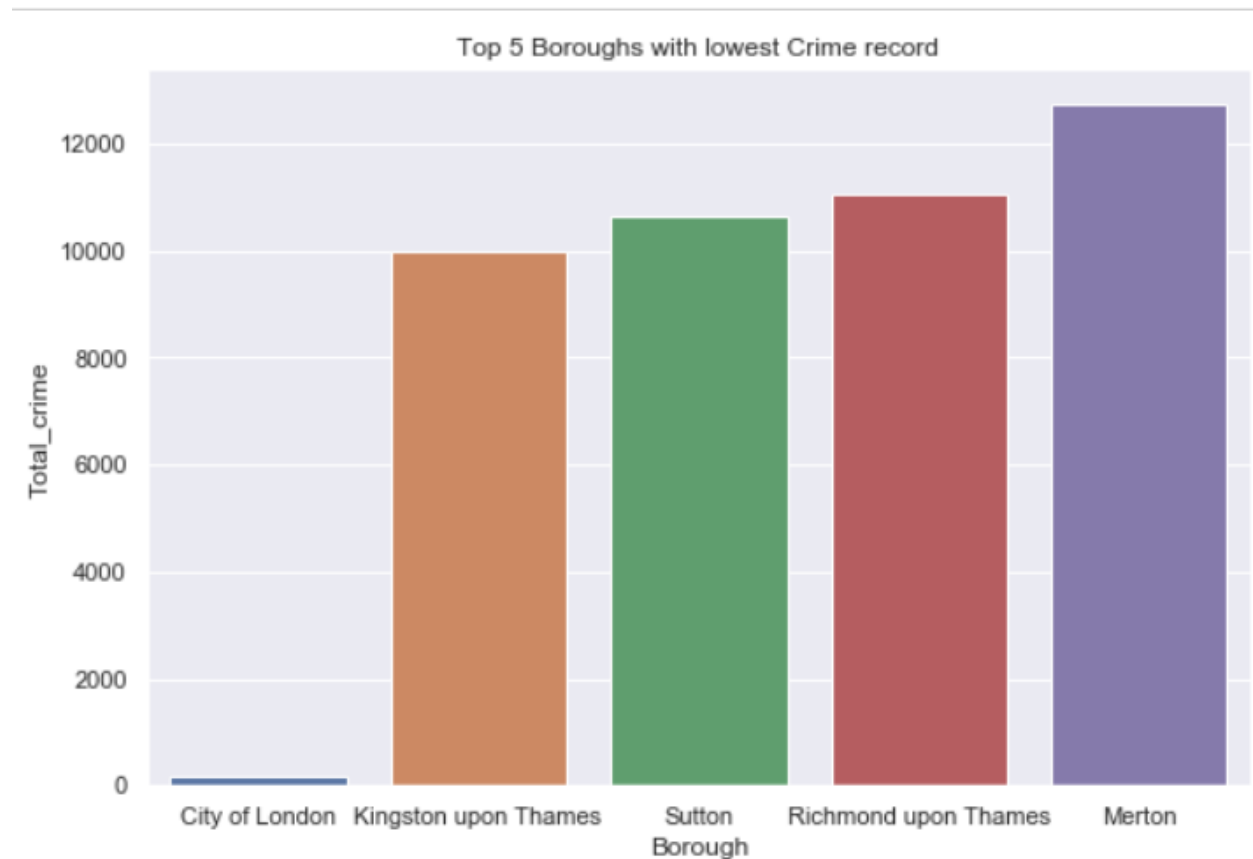


Fig 3.2.1 Boroughs with lowest crime records

City of London has 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 2011 which suggests that it is a small area (see *fig 3.2.2*). Hence, we will consider the next Borough with the lowest crime record as the safest Borough in London which is Kingston upon Thames.

	Borough	Area (sq mi)	Population(2011 est)
0	City of London	1.12	7000

Fig 3.2.2 City of London

3.3 Neighborhoods in Kingston upon Thames

There are 15 neighborhoods in the Royal Borough of Kingston upon Thames, they are visualized on the map using the folium library in python (see fig 3.3).

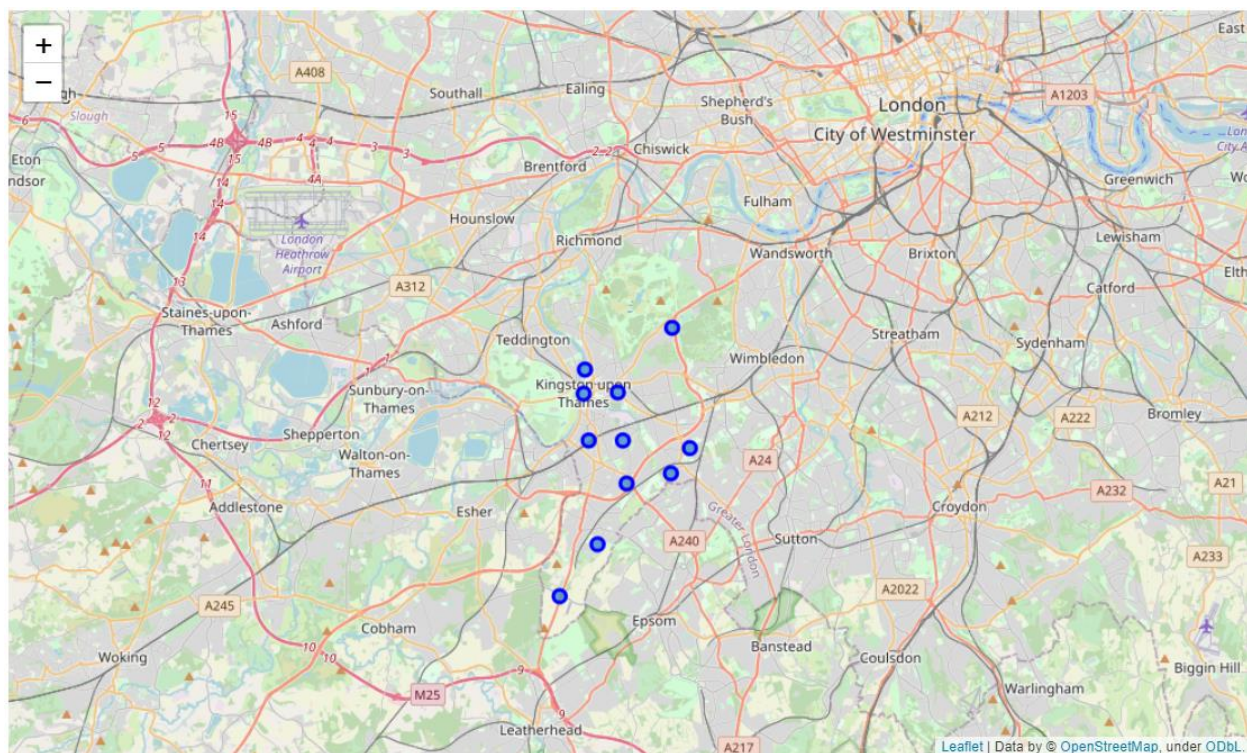


Fig 3.3 Neighborhoods in Kingston upon Thames

3.4 Data 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 data frame. This data frame contains all the venues along with their coordinates and category (see *fig 3.4*).

	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	SK Superstores	51.389901	-0.283278	Convenience Store
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
...
224	Tolworth	51.378876	-0.282860	Mantex UK	51.377009	-0.285152	Furniture / Home Store
225	Tolworth	51.378876	-0.282860	LloydsPharmacy	51.381273	-0.283112	Pharmacy
226	Tolworth	51.378876	-0.282860	Tolworth Bus Stop B	51.377780	-0.279041	Bus Stop
227	Tolworth	51.378876	-0.282860	Viana Restaurant	51.382067	-0.284743	Restaurant
228	Tolworth	51.378876	-0.282860	Tolworth Railway Station (TOL)	51.377385	-0.279454	Train Station

Fig 3.4 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 pre-defined size. We will use a cluster size of 5 for this project that will cluster the neighborhoods into 5 clusters. The reason to conduct a K – means clustering is to cluster neighborhoods with similar venues together so people can shortlist the area if their interest based on the venues/amenities around each neighborhood.

4. Results

After running the K – means clustering we can access each cluster 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	8th Most Common Venue	9th Most Common Venue
10	Old Malden	Kingston upon Thames	51.382484	-0.25909	0	Train Station	Indian Restaurant	Food	Pub	Women's Store	Falafel Restaurant	Deli / Bodega	Department Store	Dessert Shop

Fig 4.1 Cluster 1

Cluster 1 is one of the shortest clusters with only 1 out of the 15 neighborhoods. Upon closely examining these neighborhoods we can see that the most common venues are Train Station, Restaurants, Pubs 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 Most Common Venue	9th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	1	Gym / Fitness Center	Convenience Store	Park	Bus Stop	Deli / Bodega	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner
7	Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Gym	Restaurant	Park	Bus Stop	Soccer Field	Falafel Restaurant	Dessert Shop	Discount Store	Donut Shop

Fig 4.2 Cluster 2

Cluster 2 contains two neighborhoods with Gym, Convenience Store, Restaurants, Parks and Bus stop being the most common venue (see *fig 4.2*).

Cluster 3 is by far the largest cluster with 7 neighborhoods. The most common venues in these neighborhoods are Pubs, Café's, Restaurants, Supermarket, Clothing store and Parks (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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	2	Pub	Gym / Fitness Center	Shop & Service	Plaza	Park	Hotel	Indian Restaurant	Supermarket
4	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	2	Coffee Shop	Café	Pub	Clothing Store	Italian Restaurant	Bakery	Department Store	Ice Cream Shop
6	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Restaurant	Garden Center	Pub	Falafel Restaurant	Deli / Bodega	Department Store	Dessert Shop
8	New Maldon	Kingston upon Thames	51.731198	0.679185	2	Pub	Italian Restaurant	Coffee Shop	Supermarket	Mexican Restaurant	Department Store	Sandwich Place	Brewery
9	Norbiton	Kingston upon Thames	51.409999	-0.287396	2	Indian Restaurant	Food	Platform	Pub	Italian Restaurant	Breakfast Spot	Coffee Shop	Pharmacy
11	Surbiton	Kingston upon Thames	51.393756	-0.303310	2	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Platform	Breakfast Spot	French Restaurant
12	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Pharmacy	Restaurant	Bowling Alley	Sandwich Place	Bus Stop	Pizza Place	Coffee Shop

Fig 4.3 Cluster 3

Cluster 4 and 5 contain 1 neighborhood each. This is because the most common venues for these neighborhoods are unique in nature.

Most common venues for Cluster 4 include Beach, Women's store, Dessert shop, Electronics Store and Restaurants (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
3	Coombe	Kingston upon Thames	50.878156	-4.547739	3	Beach	Women's Store	Fish & Chips Shop	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner	Electronics Store	English Restaurant

Fig 4.4 Cluster 4

For Cluster 5 most common venues include Grocery store, Bar, Soccer Field, Departmental Store and Donut shops (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
5	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Sandwich Place	Soccer Field	Farmers Market	Department Store	Dessert Shop	Discount Store	Donut Shop

Fig 4.5 Cluster 5

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

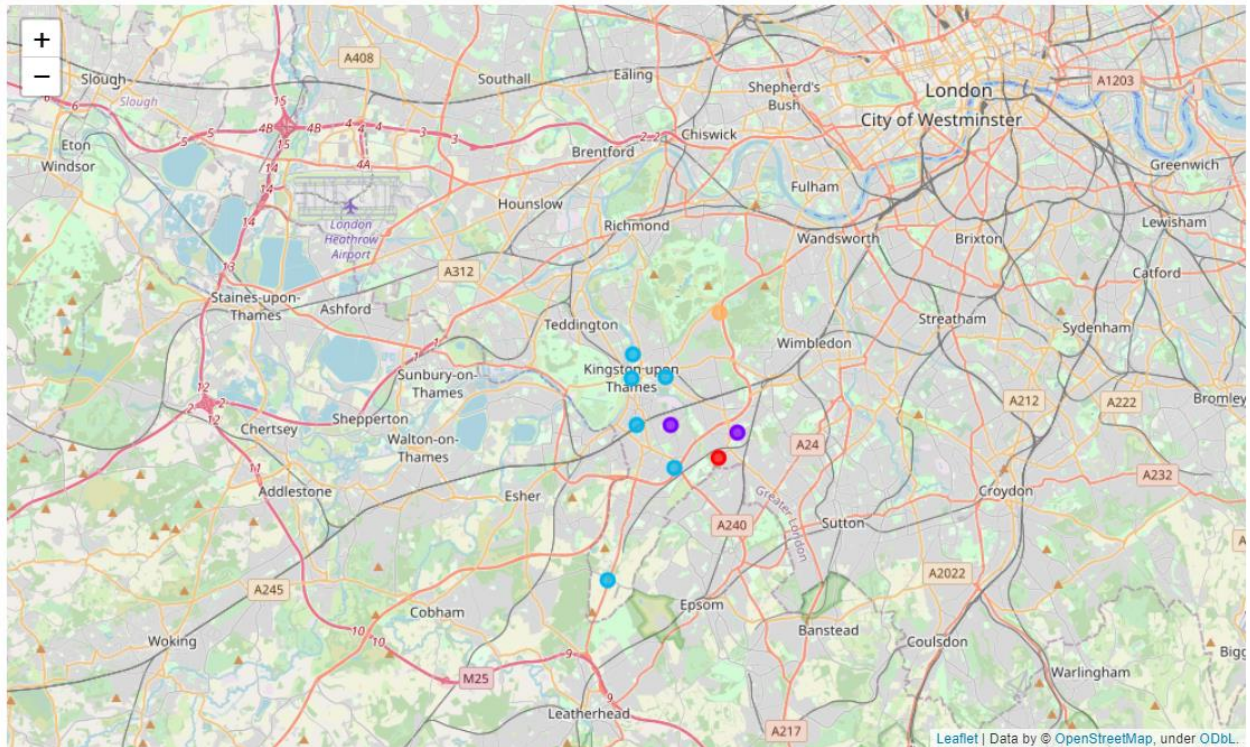


Fig 4.6 Clustered Neighborhoods in the Borough of Kingston upon Thames

Each cluster is color coded for the ease of presentation; we can see that majority of the neighborhood falls in the Blue cluster which is the third cluster. Two neighborhoods have their own cluster (Red and Orange), these are clusters four and five. The Purple cluster consists of two neighborhoods which is the second 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, people who are looking for a neighborhood with good daily commute options then clusters 1 and 2 would be

recommended. If people are looking for neighborhoods with good eating joints, then clusters 4 and 5 would be advisable. For people who are looking to shift with family, cluster 3 would be recommended because it has the most amenities/facilities in close vicinity.

6. Conclusion

This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhoods. The neighborhoods selected in this project were done based on crime records in various Boroughs. Improvements can be made to this project by taking into consideration more factors such as Cost of living, Budget etc. Projects like these can really help people out and prevent them from moving out later due to neighborhood issues.