Capstone Project - The Battle of Neighborhoods Report

Prepared By – Pranav Kumar

September 7, 2020

Table of Contents

1. Introduction

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 bring us satisfaction? 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 Boroughs of London from 2008 to 2016. The year 2016 being the latest we will be considering the data of that year which is actually 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 neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Interest

Expats who are considering relocating to London will be interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.

2. Data Acquisition and 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 the boroughs for each major category (see Figure 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

Figure 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 Figure 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 [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

Figure 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 Figure 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	18103	18888	9188	2819	6105	50999	43091	149447
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	36981	21024	9796	2953	7374	87285	46565	212191
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	14973	17244	7346	1999	2338	40071	30037	114136
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	28923	20569	25978	3711	12473	72523	63178	227551
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	27135	24039	8942	2637	4868	69742	46759	184349

Figure 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 left blank (see Figure 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		

Figure 4 - Neighborhood of the safest borough

The coordinates of the neighborhoods are be obtained using Google Maps API geocoding to get the final dataset (See Figure 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

Figure 5 - Neighborhoods of the safest borough

The new dataset is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

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 Figure 5).

	Burglary	Criminal Damage	Drugs	Fraud or Forgery	Other Notifiable Offences	Robbery	Sexual Offence	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	33.000000	33.000000
mean	22857.363636	19119.333333	14265.606061	161.363636	3222.696970	7844.636364	38.575758	80662.454545	47214.575758	195386.606061
std	7452.366846	5942.903618	7544.259564	81.603775	1362.107294	4677.643075	15.139002	45155.624776	17226.165191	79148.057551
min	15.000000	16.000000	33.000000	0.000000	17.000000	24.000000	0.000000	561.000000	114.000000	780.000000
25%	18103.000000	17244.000000	8942.000000	106.000000	2358.000000	4744.000000	27.000000	52609.000000	33968.000000	149447.000000
50%	24871.000000	20405.000000	14101.000000	157.000000	3293.000000	7688.000000	40.000000	77940.000000	50943.000000	203879.000000
75%	27980.000000	22755.000000	18389.000000	207.000000	3963.000000	10084.000000	47.000000	92523.000000	59993.000000	228613.000000
max	36981.000000	31218.000000	34031.000000	323.000000	6504.000000	18408.000000	71.000000	277617.000000	72726.000000	455028.000000

Figure 6 - 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 Figure 7).

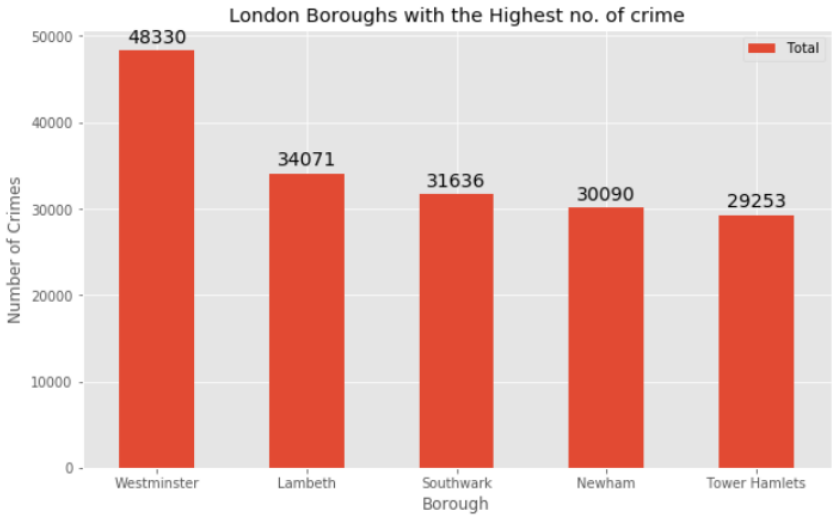

Figure 7 - Boroughs with the highest crime rates

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 Figure 8).

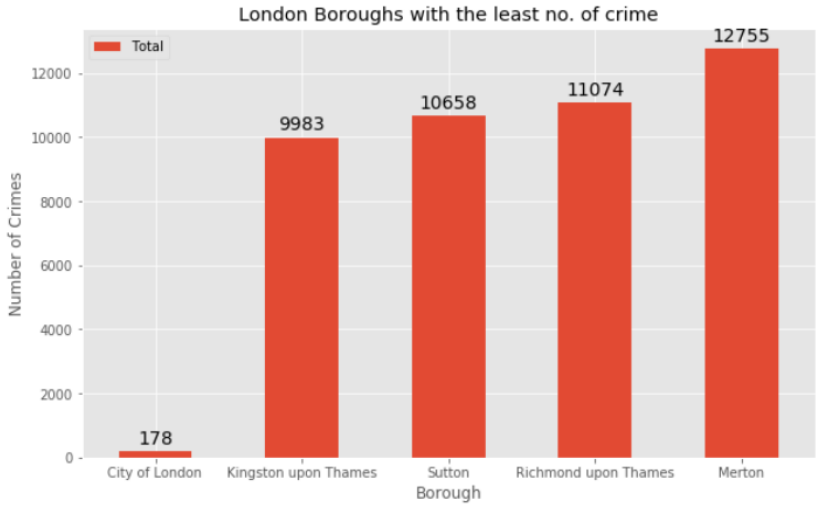

Figure 8 - Boroughs with the lowest crime rates

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 Figure 9). 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]
6	City of London	780	1.12	7000

Figure 9 - 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 Figure 10).

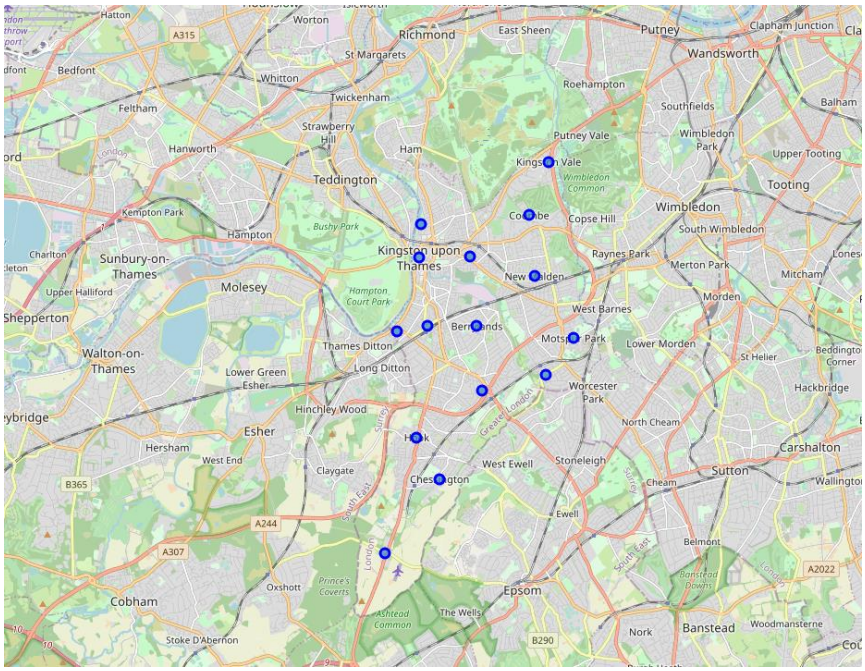

Figure 10 - Neighborhoods in Kingston upon Thames

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 data frame. This data frame contains all the venues along with their coordinates and category (see Figure 11).

	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

Figure 11 - 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 Figure 12)

	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	0	Pub	Park	Coffee Shop	Platform	Convenience Store	College Soccer Field	Bus Stop	Gym / Fitness Center	Train Station	Gift Shop
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Coffee Shop	Café	Clothing Store	Hotel	Park	Thai Restaurant	Italian Restaurant	Department Store	Sushi Restaurant
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Pub	Coffee Shop	Café	Clothing Store	Italian Restaurant	Thai Restaurant	Burger Joint	Sushi Restaurant	Park	Department Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	0	Park	Bus Stop	Furniture / Home Store	Pub	Grocery Store	Rugby Pitch	Restaurant	Japanese Restaurant	Soccer Field	Mediterranean Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Korean Restaurant	Supermarket	Fast Food Restaurant	Grocery Store	Coffee Shop	Indian Restaurant	Café	Bus Stop	Pub	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Pub	Gastropub	Bar	Indian Restaurant	Thai Restaurant	Gym / Fitness Center	Food	Italian Restaurant	Japanese Restaurant	Pizza Place
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Train Station	Park	Steakhouse	Gym / Fitness Center	Grocery Store	Bakery	English Restaurant	Japanese Restaurant	Falafel Restaurant	Fast Food Restaurant
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Pub	Coffee Shop	Indian Restaurant	Platform	Fish & Chips Shop	Italian Restaurant	Gastropub	Bakery	Pharmacy	Grocery Store
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Pub	Coffee Shop	Indian Restaurant	Grocery Store	Park	Italian Restaurant	Hotel	Gym / Fitness Center	Pharmacy	Platform
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Discount Store	Restaurant	Soccer Field	Bowling Alley	Coffee Shop	Climbing Gym	Sandwich Place	Pizza Place

Figure 12 - Cluster 1

The cluster one is the biggest cluster with 10 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, Train Station, and stores.

Looking into the neighborhoods in the second, fourth and fifth clusters, we can see these clusters have only one neighborhood in each. This is because of the unique venues in each of the neighborhoods, hence they couldn't be clustered into similar neighborhoods (see Figure 14,15 and 16).

	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
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Theme Park Ride / Attraction	Zoo Exhibit	Pub	Restaurant	Exhibit	Fast Food Restaurant	Fried Chicken Joint	Garden Center	Gift Shop	Grocery Store

Figure 13 - 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
3	Coombe	Kingston upon Thames	51.41945	-0.265398	3	Hotel	Bus Stop	Rest Area	Sports Club	Stables	Food	Diner	Discount Store	Donut Shop	Electronics Store

Figure 14 - 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Stables	Grocery Store	Bar	Coffee Shop	Sandwich Place	Soccer Field	Outdoors & Recreation	Zoo Exhibit	Discount Store	Donut Shop

Figure 15 - 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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	2	Train Station	Fast Food Restaurant	Supermarket	Convenience Store	Platform	Golf Course	Breakfast Spot	Zoo Exhibit	Falafel Restaurant	Farmers Market
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Convenience Store	Supermarket	Indian Restaurant	Café	Park	Breakfast Spot	Fish & Chips Shop	Fast Food Restaurant	Bakery	Platform

Figure 16 - Cluster 3

The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Train Station, Supermarket, Breakfast Spot etc. (see Figure 17).

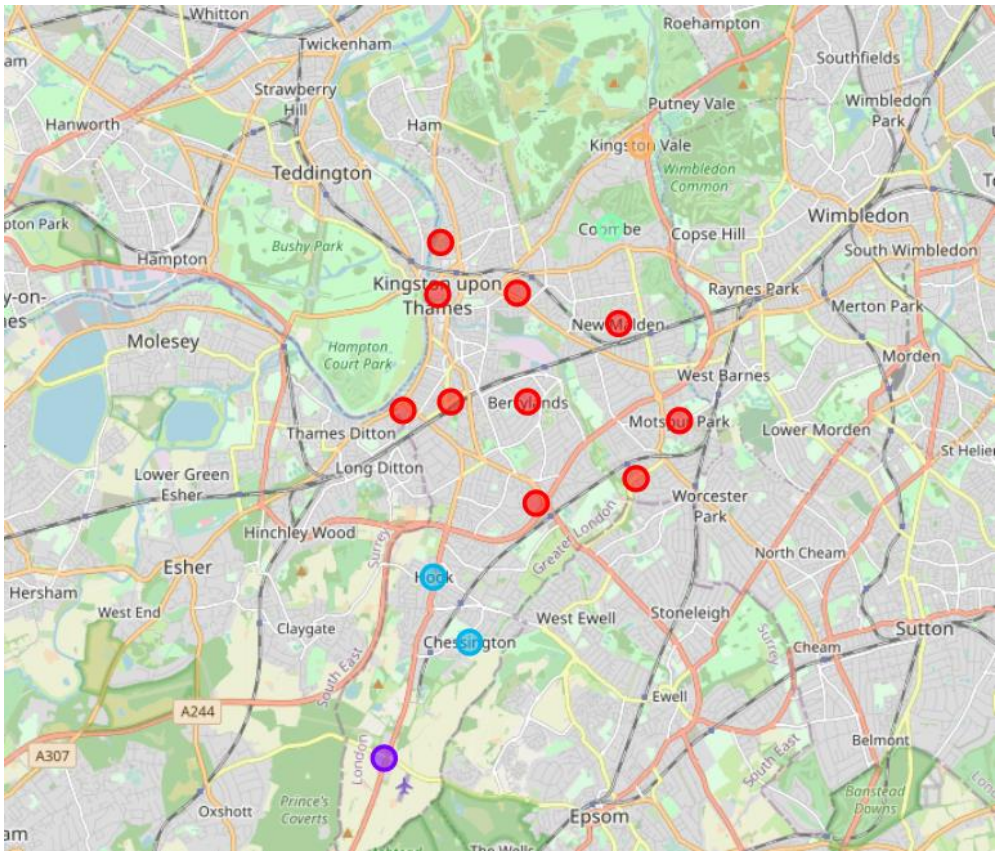

Figure 17 - Clustered neighborhoods in the Borough of Kingston upon Thames

5. Discussion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can 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 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 first cluster is suitable. For a family I feel that the neighborhoods in Cluster 3 and 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Bus Stops, Restaurants, Electronics Stores and Grocery Store which is ideal for a family. The choices of neighborhoods may vary from person to person.

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.