

Capstone Project  
The battle of Neighborhoods  
Report

# 1.Introduction

## a. Background

People move at least 5 times in their timelines. We are looking for a better life, a better place with more opportunity to grow up our career. Moving into a big city also is a big problem.

## b. Problem

We found many questions here:

- Do we simply move to a big city from town without any investigation?
- Have we ever considered the safety of a borough is a requirement when we looking for a place to live?
- Living in a borough nearby the school, or a café shop. It is not a bad idea?

Here we found a simple project on looking a borough to live in one of the biggest cities in Europe – London. This project can lead us to answer lots of questions by investigate the crime's ratio in every neighborhood in London and help us to explore the top venues nearby in every borough.

## c. Target

This project is aimed to help anyone who considers relocating their life in London. They will find out all the necessary information before the move.

# 2.Data Acquisition and Data Cleaning

## a. Data Acquisition

The data used in this project is a combination of multiple sources.

The most important source is the London Crime CSV which presents all the crime during 2008 – 2016. (Provide by Kaggle Data Link ). The dataset contains of 7 columns:

Isoa\_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

Next, a list of boroughs is found on the website Wikipedia ([Borough Data](#) )

- 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 a list of districts in the Royal Borough of Kingston upon Thames As found also on [Wikipedia](#). This dataset is created from scratch using the list of neighborhoods available on the site.

- Neighborhood: Name of the neighborhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

#### b. Data cleaning

The preparation of these data source is done separately. From the London Crime data, we select only the crime during the year 2016. The major categories of the crime are pivoted to get a better calculation for the crime ratio after.

	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

Next, we merge these data to mark a new data set that contains all the necessary information's. Now we have a complete dataset to visualize the crime rate in each borough to find out the safest borough in London.

	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

Here we can easily find the borough with the lowest crime ratio.

It is the time to use the third data source. 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 empty (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 can be filled by using the Google Map API geocoding. (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

### 3. Methodology

#### a. Exploratory Data analysis

##### i. Statistical summary of crimes

We use the describe function in pandas to get the statistic of the London crime data, the function returns mean, standard deviation, min, max , ...

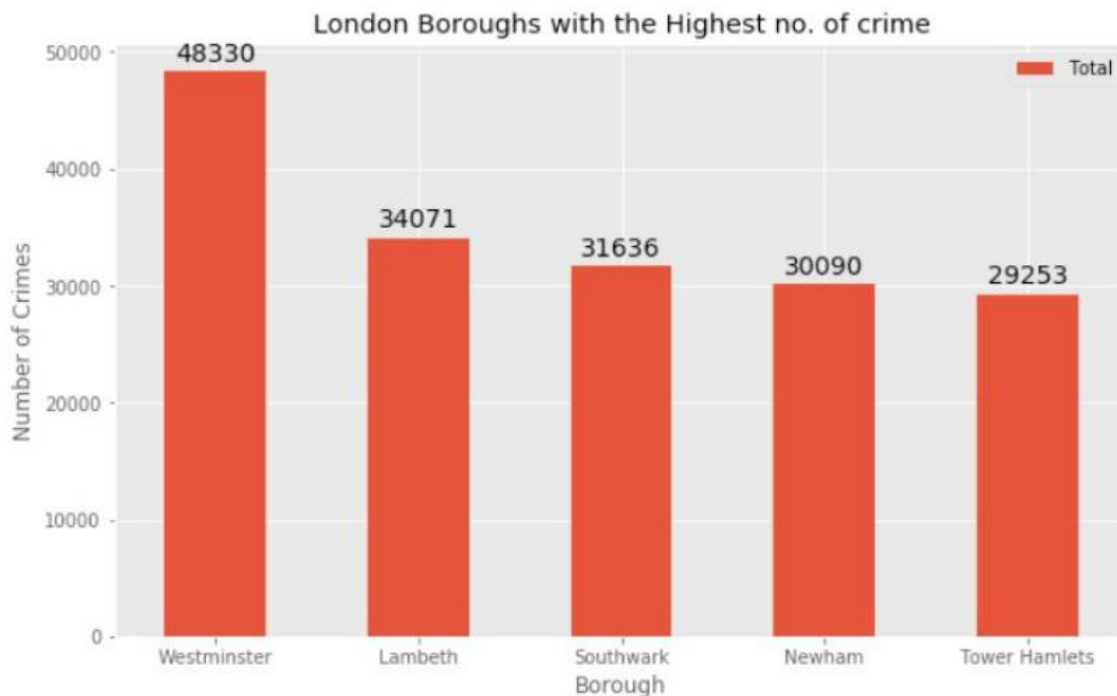
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
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

Look at the fig, here we find the 'Theft and Handling' is the highest reported crime category. It followed by 'Violence against the person' and 'Criminal damage'. The lowest reported crime is 'Drugs', 'Robbery' and 'Other'.

ii. Boroughs with highest crime ratio:

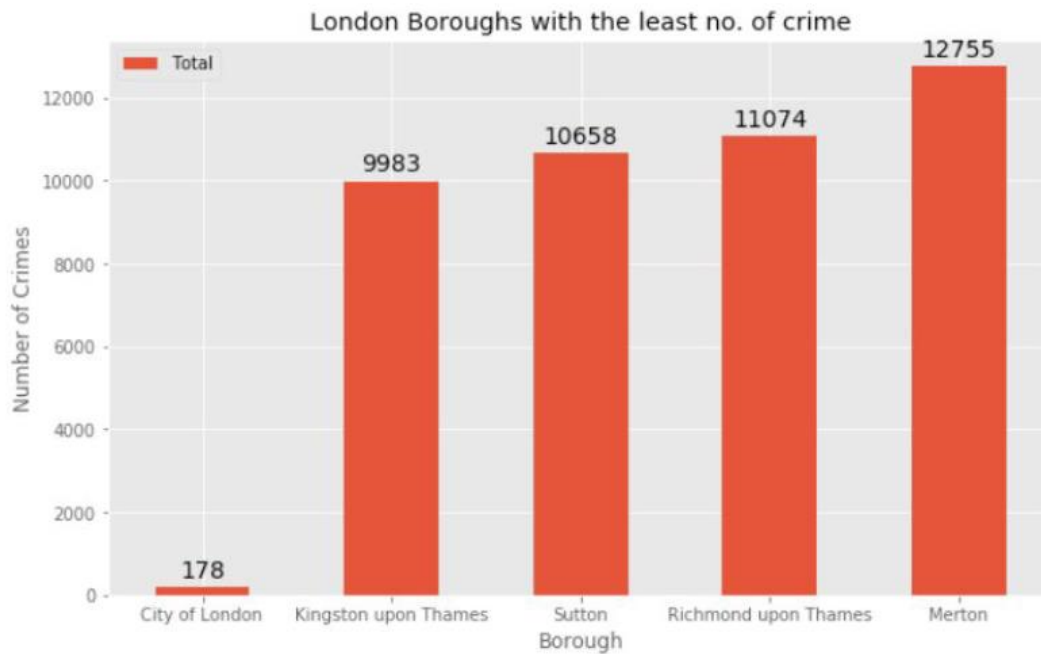
Comparing five top boroughs with the highest crime rate during the year 2016. Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs (Fig 3.1.2)



*Fig 3.1.2 Boroughs with the highest crime rates*

iii. Boroughs with lower crime rate

According to the figure 3.1.3, City of London is the safest borough in London. It followed by Kingston upon Thames, Sutton Borough, Richmond upon Thames, Merton.



*Fig 3.1.3 Boroughs with the lowest crime rates*

City of London has a best crime rate here because it is the 33<sup>rd</sup> principal division of Greater London and it is not considered like a London borough. It has only 1.12 square miles and a population about 7000 at year 2013. So that we consider the safest borough here is Kingston upon Thames.

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

#### iv. Neighborhoods in Kingston upon Thames

After modeling, there are 15 neighborhoods in the royal borough of Kingston upon Thames. The visualization below is created by folium python.



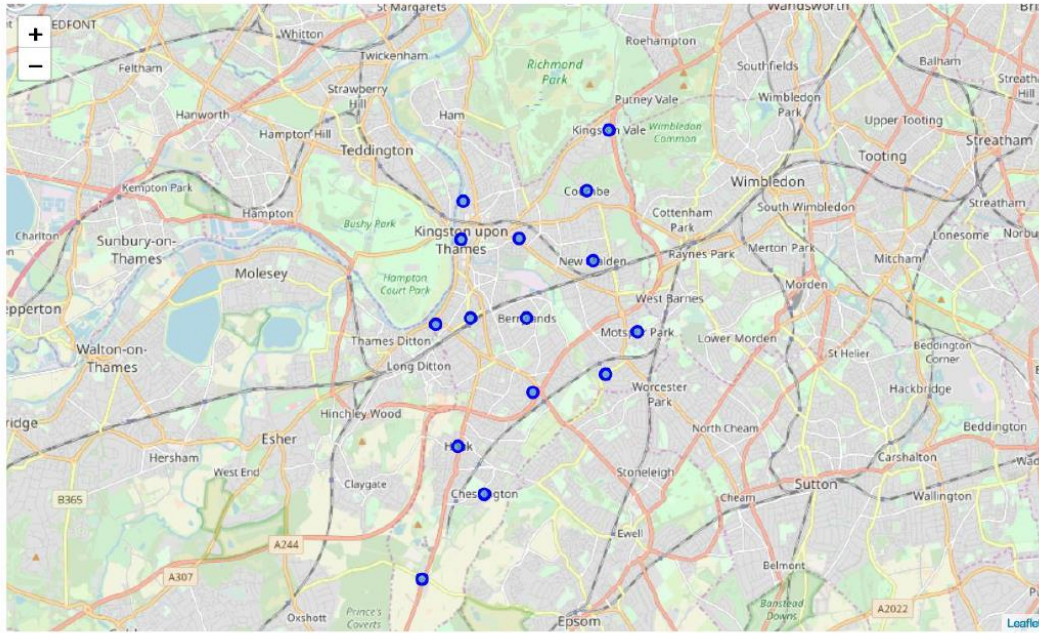


Fig 3.1.4 Neighborhoods in Kingston upon Thames

## 4. Modeling

Using the final dataset that contains the neighborhood in Kingston upon Thames along with the latitude and longitude. We start to find all the venues with 500 meters radius of each neighborhood by using the Foursquare API. This will return a json file that contains all the venues in each neighborhood as well. Now our mission is converting this into the data frame. (Fig 3.2.1)

The Venues data is then grouped by the Neighborhood and

	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	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Fig 3.2.1 Venue details of each Neighborhood

the mean of the venues is calculated, finally the 10 common venues are calculated for each of the neighborhoods.

In order to help people easily to identify the neighborhoods in the borough we are applying the clustering method using the K-Means clustering method. However, we divide our 15 neighborhoods into 5 different clusters.



## a. Result

After running the method, we can access to these clusters to see which neighborhoods were assigned to each of the five cluster.

Here is the first cluster 1. This is also the biggest cluster in our model with 9 of 15 neighborhoods. We can explore the most common venues in these neighborhoods are

- Restaurant
- Pubs
- Café
- Supermarket
- 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 Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

Fig 4.1 Cluster 1

Now, we move to the other clusters. These clusters contain only one neighborhood in each. (except 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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant

Fig 4.2 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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Deli / Bodega

Fig 4.3 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Park	Gym	Restaurant	Soccer Field	Bus Stop	Wine Shop	Fast Food Restaurant	Dry Cleaner	Electronics Store

Fig 4.5 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Italian Restaurant	Soccer Field	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Department Store

Fig 4.4 Cluster 5

Here we found that each cluster contain their hot venues. By selecting our best favorite categories venues. We can now decide where to live.

Let us visualize the geography of these cluster by the fig 4.6

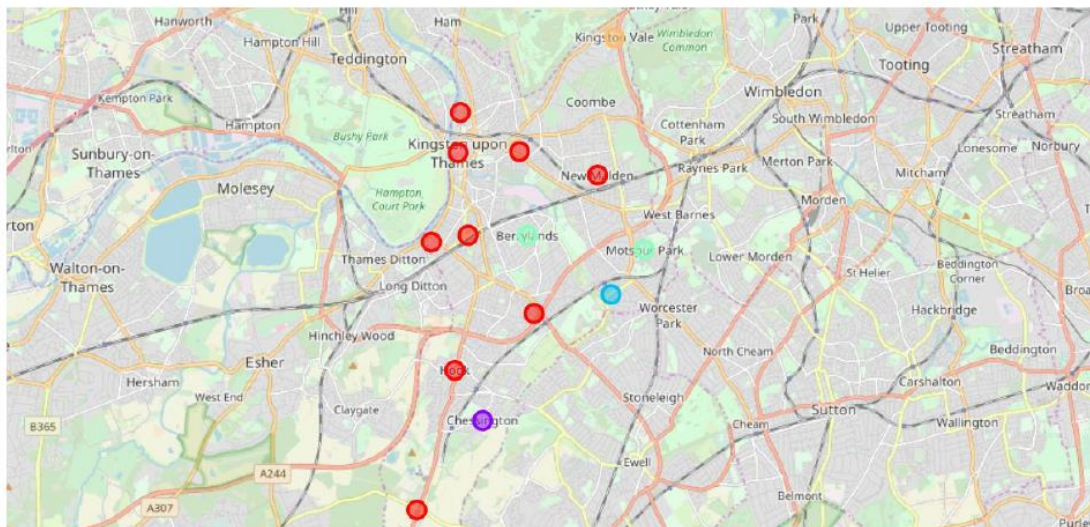


Fig 4.6 Clustered neighborhoods in the Borough of Kingston upon Thames

Each cluster is colored, we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two neighborhoods which is the 4th 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 3 and 4 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 4 are more suitable due 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. 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.

**Thank you for reading**