

# Capstone Project - The Battle of Neighbourhoods Report

## 1. Introduction

### 1.1 Background

The average UK citizen moves about four times in their lifetime. The common question we asked is whether people move until they find a place to settle down where they truly feel happy? Or the opposite happen, that we move to a new area without knowing exactly what we are moving into.

In order to minimize the chances of moving to the places that we do not like, we should always conduct our due diligence like doing research when planning our next move in to another location or even country.

Safety should be the critical evaluation criteria when moving to a new location or country - no one will enjoy living when he or she does not feel safe at their home.

### 1.2 Problem

The crime statistics dataset of London (from Kaggle) has crimes in each Boroughs of London. The data range is from 2008 to 2016.

This project has 3 main steps:

- to select the borough in London which is safest, based on the total crimes each borough;
- explore the neighbourhoods of that borough to find the 10 most common location/venues in each neighbourhood;
- and perform clustering of the neighbourhoods using k-mean clustering.

### 1.3 Interest

This project finding can be also used by expats who are considering to relocate to London because they will also be interested to identify the safest borough in London and explore its neighbourhoods.

## 2. Data Acquisition and Cleaning

### 2.1 Data Acquisition

The data acquired for this project come from three sources:

- London crime data (Kaggle)
- list of London boroughs (Wikipedia)

## Capstone Project - The Battle of Neighbourhoods Report

- list of Neighbourhoods in the Royal Borough of Kingston upon Thames (Wikipedia)

The first data source of the project uses a London crime data that shows the crime per borough in London. The columns of the data as below:

- **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
- **month**: Month of reported counts

The second source of data is scraped from a Wikipedia that contains the list of London boroughs. The page is useful because it contains additional information about the boroughs, the column of the data as below:

- **Borough**: The names of the 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**: 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. This dataset is created from scratch using the list of neighbourhood found online and coordinates from Google Maps API geocoding the column of the data as below:

- **Neighbourhood**: Name of the neighbourhood in the Borough
- **Borough**: Name of the Borough
- **Latitude**: Latitude of the Borough
- **Longitude**: Longitude of the Borough

# Capstone Project - The Battle of Neighbourhoods Report

## 2.2 Data Cleaning

The data preparation for each of the three data source is done one by one. 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.

	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 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 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.

## Capstone Project - The Battle of Neighbourhoods Report

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	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	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	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	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	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 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		

*Fig 2.4 Neighbourhoods of the safest borough*

The coordinates of the neighbourhoods is be obtained using Google Maps API geocoding to get the final dataset (See *Fig 2.5*) .

## Capstone Project - The Battle of Neighbourhoods Report

	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 Neighbourhoods of the safest borough*

The new dataset is used to generate the 10 most common venues for each neighbourhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighbourhoods 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 *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	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

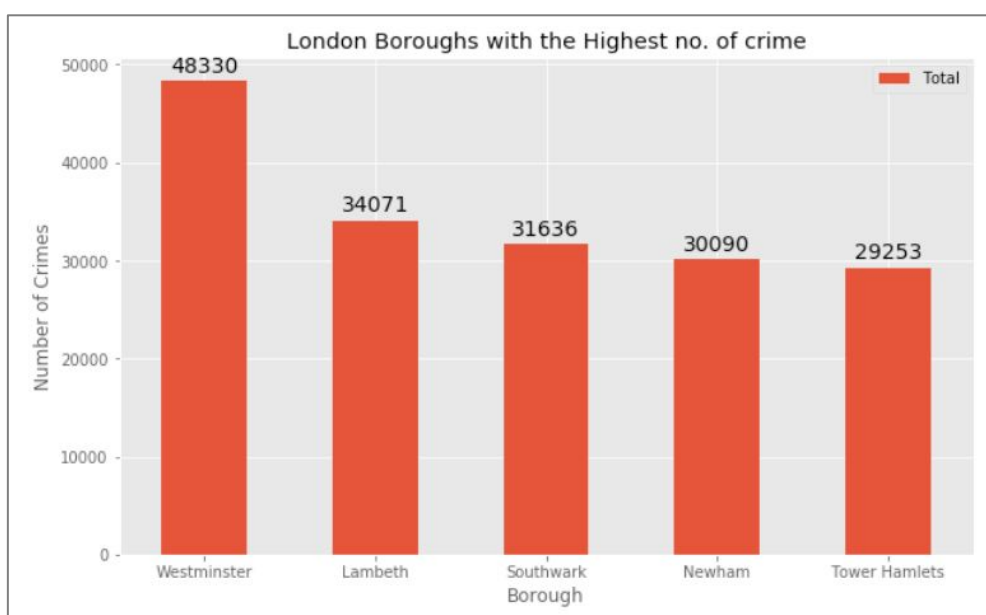
*Fig 3.1.1 Statistical description of the London crimes*

## Capstone Project - The Battle of Neighbourhoods Report

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).

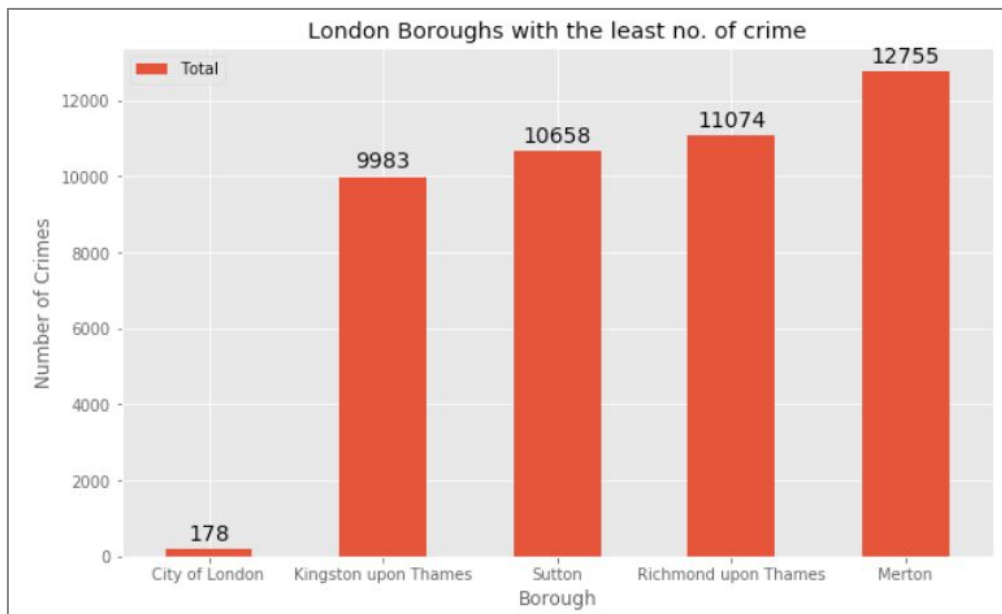


*Fig 3.1.2 Boroughs with the highest crime rates*

### 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).

## Capstone Project - The Battle of Neighbourhoods Report



*Fig 3.1.3 Boroughs with the lowest crime rates*

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]
6	City of London	178	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*).

# Capstone Project - The Battle of Neighbourhoods Report

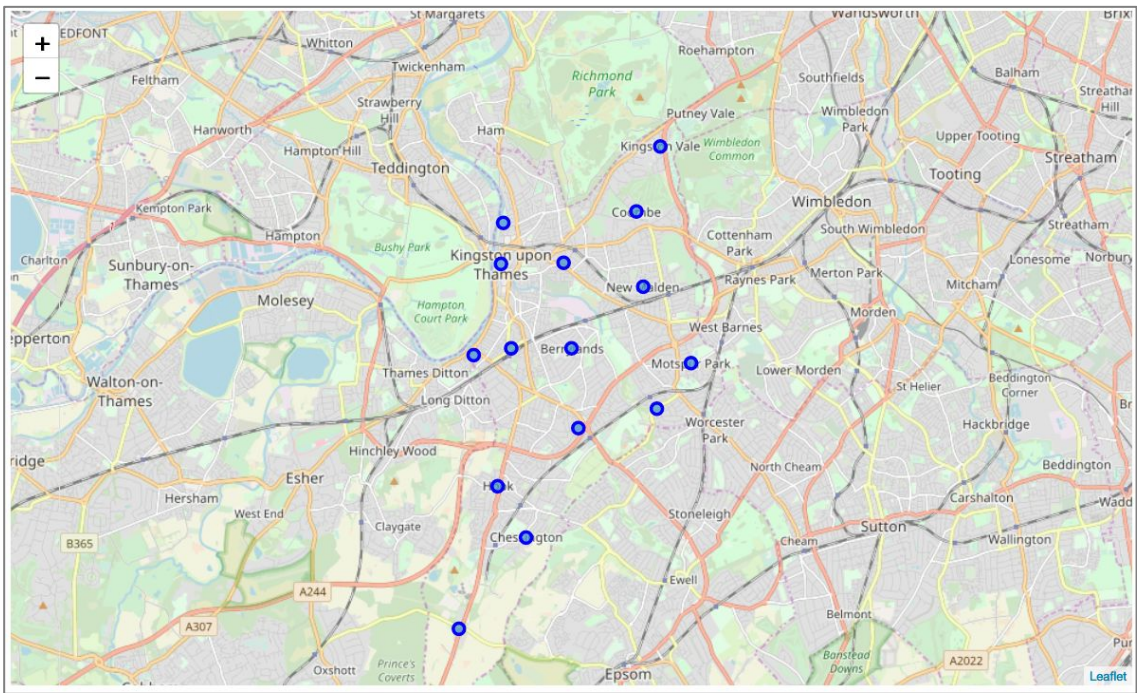


Fig 3.1.4 Neighbourhoods in Kingston upon Thames

## 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 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	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 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 Neighbourhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighbourhoods.



## Capstone Project - The Battle of Neighbourhoods Report

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. Results

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 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
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*

The cluster one is the biggest cluster with 9 of the 15 neighbourhoods in the borough Kingston upon Thames. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

Looking into the neighbourhoods in the second, third and fifth clusters, we can see these clusters have only one neighbourhood in each. This is because of the unique venues in each of the neighbourhoods, hence they couldn't be clustered into similar neighbourhoods (see *figures 4.2, 4.3 and 4.4*).

## Capstone Project - The Battle of Neighbourhoods Report

	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*

The second cluster has one neighbourhood which consists of Venues such as Restaurants, Golf courses, and wine shops.

	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*

The third cluster has one neighbourhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

	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*

The fifth cluster has one neighbourhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores. We will look into the neighbourhoods in the fourth cluster (*see fig 4.5*).

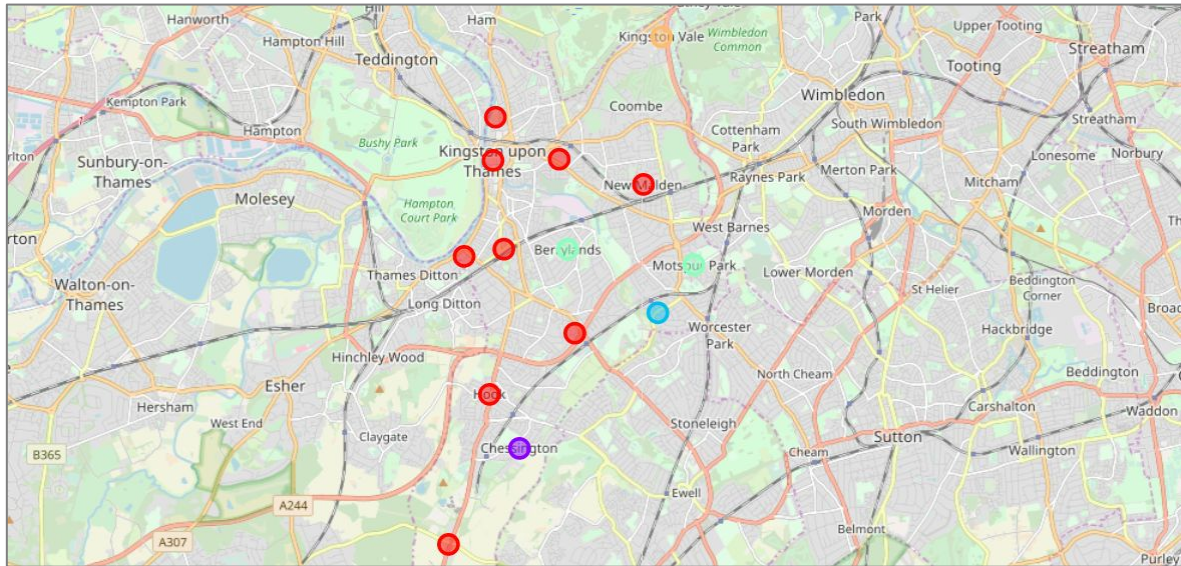
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	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*

## Capstone Project - The Battle of Neighbourhoods Report

The fourth cluster has two neighbourhoods in it, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

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 red cluster which is the first cluster. Three neighbourhoods have their own cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two neighbourhoods 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 chose 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 a close proximity then the neighbourhoods in the first cluster is suitable. The choices of neighbourhoods depends on the person.

## **Capstone Project - The Battle of Neighbourhoods Report**

### **6.Conclusion**

This project helps a person get a better understanding of the neighbourhoods with respect to 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.