# Capstone Project

# The Battle of Neighborhoods

# Report

# By

# Narayanan R

# 1. Introduction

## 1.1. Background

When you move to a place where nobody knows you, you'll realize that your fresh start will go hand in hand with a number of advantages for you as well. The people you meet won't have any expectations of you, so you can become the person you've always wanted to be and they won't know any better. But it is a  big decision that we make in our lives. When we need to move to a new place for whatever reasons it is we should always do proper research when planning our next move in life. 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 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 neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering.

## 1.3. Interest

People who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around each neighbourhood.

# 2. Data Collection and Preparation

## 2.1. Data Collection

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 neighborhood 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 Preparation

The data preparation for the 3 sources above is done separately. From the London Crime data only 2016 Crimes are selected The major categories of crime are pivoted to get total crime per borough 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 |

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. We then convert it in to a data frame. 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 | 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 |

The above two data set are then merged to visualize the crime rates in each borough and to get the boroughs with the least crime rates

| | Borough | Burglary | Criminal Damage | Drugs | Other Notifiable Offences | Robbery | Theft and Handling | Violence Against the Person | Total | Inner | Status | Local authority | Political control | Headqu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 1287 | 1949 | 919 | 378 | 534 | 5607 | 6067 | 16741 | NaN | NaN | Barking and Dagenham London Borough Council | Labour | Town Town |
| 1 | Barnet | 3402 | 2183 | 906 | 499 | 464 | 9731 | 7499 | 24684 | NaN | NaN | Barnet London Borough Council | Conservative | North Bu O Roac |
| 2 | Bexley | 1123 | 1673 | 646 | 294 | 209 | 4392 | 4503 | 12840 | NaN | NaN | Bexley London Borough Council | Conservative | Civic 2 |
| 3 | Brent | 2631 | 2280 | 2096 | 536 | 919 | 9026 | 9205 | 26693 | NaN | NaN | Brent London Borough Council | Labour | Bre En |

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

| | 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 | | |
| 5 | Kingston upon Thames | Kingston upon Thames | | |
| 6 | Kingston Vale | Kingston upon Thames | | |
| 7 | Malden Rushett | Kingston upon Thames | | |
| 8 | Motspur Park | Kingston upon Thames | | |
| 9 | New Malden | Kingston upon Thames | | |
| 10 | Norbiton | Kingston upon Thames | | |

The coordinates of the neighborhoods is be obtained using Google Maps API geocoding to get the final dataset.

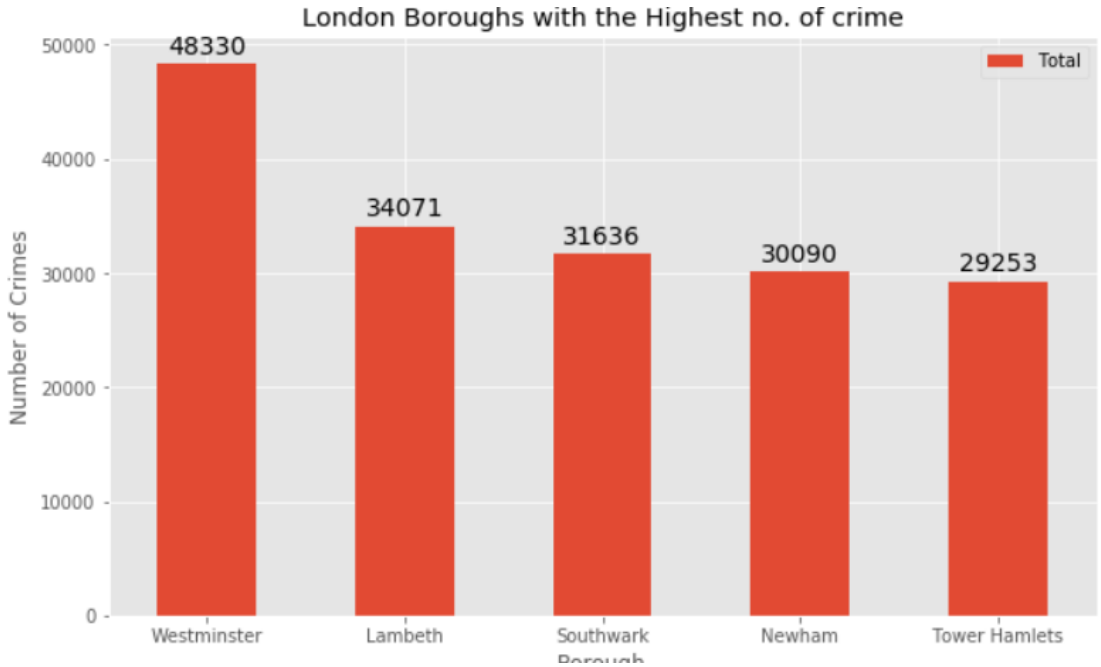
| | Neighborhood | Borough | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Berrylands | Kingston upon Thames | 51.393781 | -0.284802 |
| 1 | Canbury | Kingston upon Thames | 51.417499 | -0.305553 |
| 2 | Chessington | Kingston upon Thames | 51.358336 | -0.298622 |
| 3 | Coombe | Kingston upon Thames | 51.419450 | -0.265398 |
| 4 | Hook | Kingston upon Thames | 51.367898 | -0.307145 |
| 5 | Kingston upon Thames | Kingston upon Thames | 51.409627 | -0.306262 |
| 6 | Kingston Vale | Kingston upon Thames | 51.431850 | -0.258138 |
| 7 | Malden Rushett | Kingston upon Thames | 51.341052 | -0.319076 |
| 8 | Motspur Park | Kingston upon Thames | 51.390985 | -0.248898 |
| 9 | New Malden | Kingston upon Thames | 51.405335 | -0.263407 |

# 3. Methodology

## 3.1. Exploratory Data Analysis

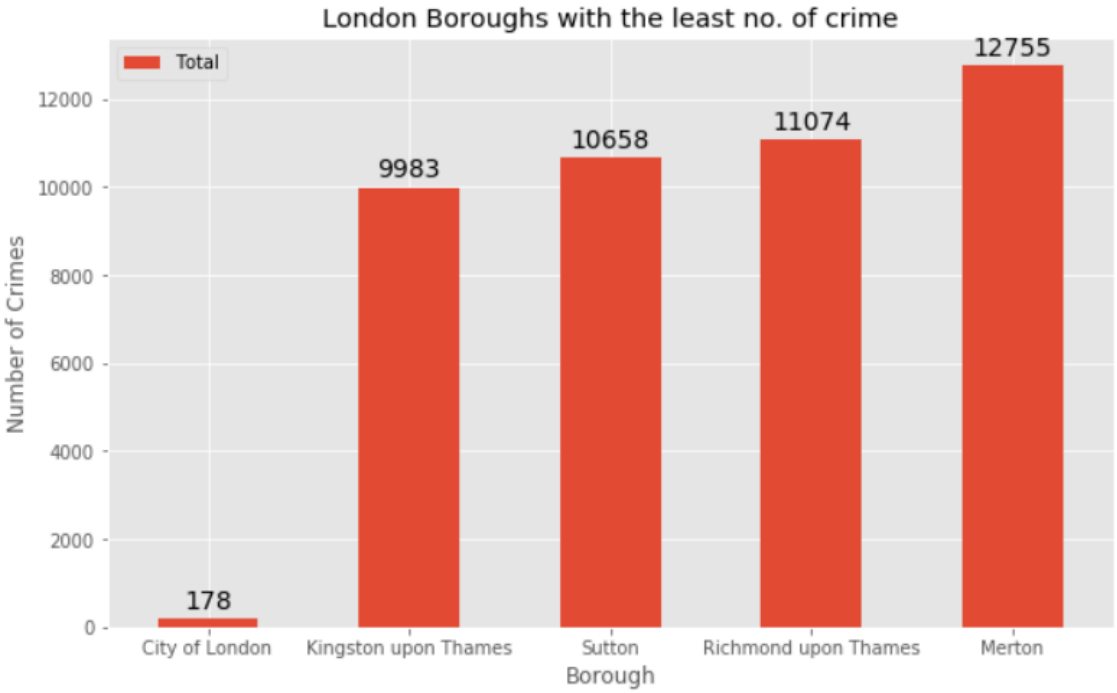
### 3.1.1. Boroughs with highest crime rates

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




### 3.1.2. Boroughs with lowest crime rates

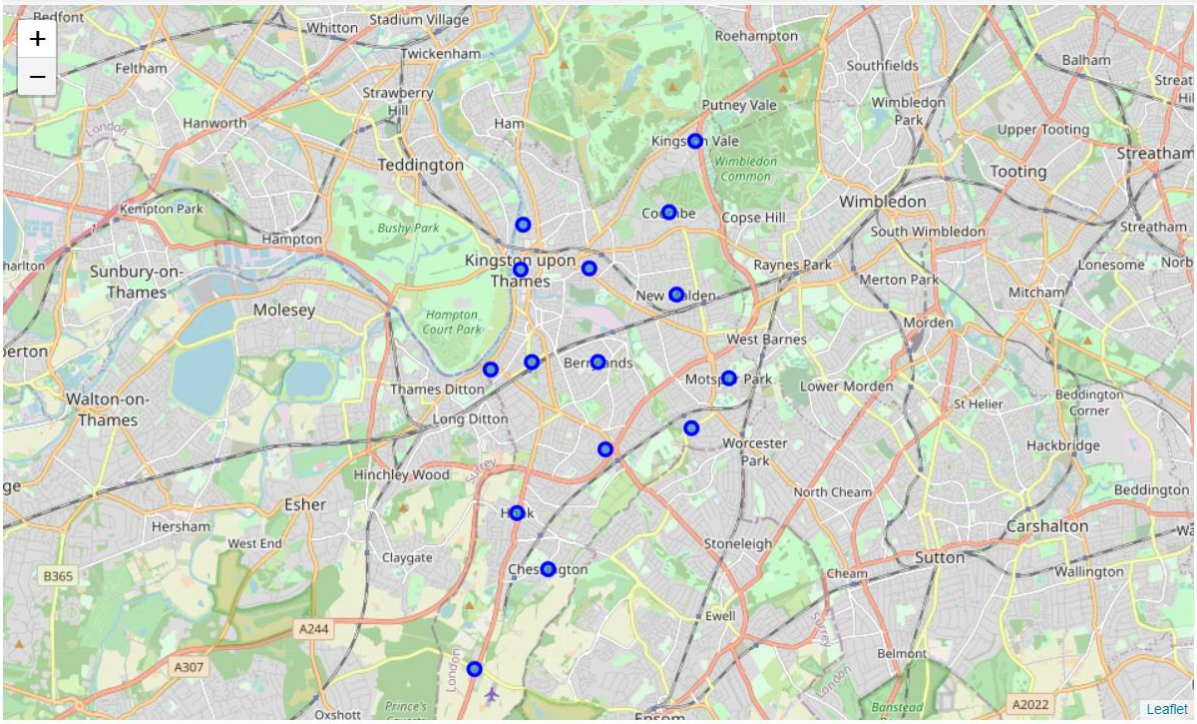
Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton




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. Hence we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

### 3.1.3. Neighborhood in Kingston upon Thames

There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python




## 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

| | 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 | ExactPrint-UK | 51.393288 | -0.288874 | Print Shop |
| 4 | Canbury | 51.417499 | -0.305553 | Canbury Gardens | 51.417409 | -0.305300 | Park |

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

The cluster one is a cluster with 2 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Gyms, Print Shop, Soccer field.

| | 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 | 0 | Print Shop | Gym / Fitness Center | Park | Bus Stop | Wine Shop | Fish & Chips Shop | Discount Store | Electronics Store | Farmers Market |
| 8 | Motspur Park | Kingston upon Thames | 51.390985 | -0.248898 | 0 | Soccer Field | Gym | Park | Bus Stop | Food | Discount Store | Electronics Store | Farmers Market | Fast Food Restaurant |

The cluster two  is a cluster with 4 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Grocery Stores, Bakery and hotels.

| | 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 | 9t Co |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | Hook | Kingston upon Thames | 51.367898 | -0.307145 | 1 | Grocery Store | Bakery | Indian Restaurant | Fish & Chips Shop | French Restaurant | Discount Store | Electronics Store | Farmers Market | Fa Res |
| 6 | Kingston Vale | Kingston upon Thames | 51.431850 | -0.258138 | 1 | Grocery Store | Bar | Sandwich Place | Soccer Field | Gastropub | Garden Center | Garden | Furniture / Home Store | |
| 7 | Malden Rushett | Kingston upon Thames | 51.341052 | -0.319076 | 1 | Grocery Store | Pub | Garden Center | Restaurant | Fish & Chips Shop | Department Store | Discount Store | Electronics Store | F |
| 14 | Tolworth | Kingston upon Thames | 51.378876 | -0.282860 | 1 | Grocery Store | Hotel | Pharmacy | Pizza Place | Coffee Shop | Italian Restaurant | Indian Restaurant | Café | Bi |

The cluster three  is a cluster with 1 neighborhood in the borough Kingston upon Thames. Upon closely examining this neighborhood we can see that the most common venues in this neighborhoods are Train Station and hotels.

| | 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 Mo Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Old Malden | Kingston upon Thames | 51.382484 | -0.25909 | 2 | Train Station | Food | Construction & Landscaping | Pub | Discount Store | Electronics Store | Farmers Market | Fast Food Restaurant | Fish Ch Sh |

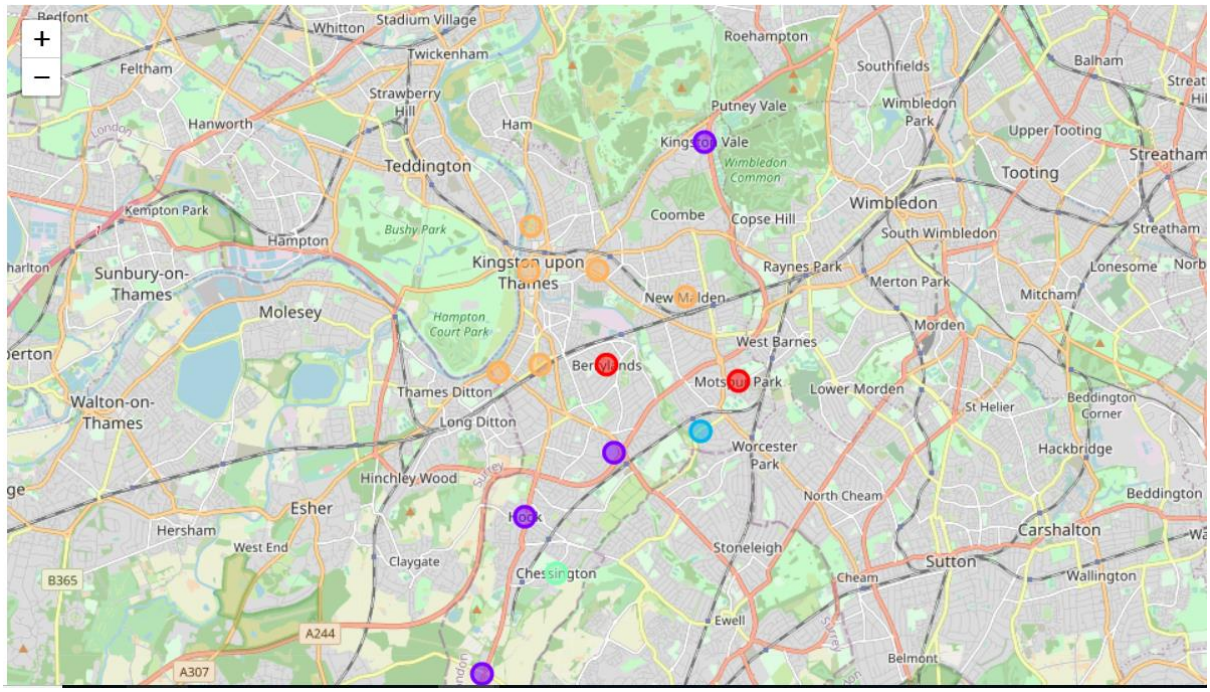
The cluster Four  is a cluster with 1 neighborhood in the borough Kingston upon Thames. Upon closely examining this neighborhood we can see that the most common venues in this neighborhoods are Wine Shops and Construction and Landscaping.

| | 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 Mo Commc Venu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Chessington | Kingston upon Thames | 51.358336 | -0.298622 | 3 | Construction & Landscaping | Wine Shop | Deli / Bodega | Discount Store | Electronics Store | Farmers Market | Fast Food Restaurant | Fish & Chips Shop | Foo |

The cluster Five  is a cluster with 6 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurant, Coffee Shops and Pubs.

| | 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 | 4 | Pub | Park | Tennis Court | Supermarket | Spa | Fish & Chips Shop | Shop & Service | Gym / Fitness Center |
| 5 | Kingston upon Thames | Kingston upon Thames | 51.409627 | -0.306262 | 4 | Coffee Shop | Café | Burger Joint | Sushi Restaurant | Pub | French Restaurant | Market | Mexican Restaurant |
| 9 | New Malden | Kingston upon Thames | 51.405335 | -0.263407 | 4 | Gastropub | Gym | Bar | Sushi Restaurant | Supermarket | Chinese Restaurant | Korean Restaurant | Indian Restaurant |
| 10 | Norbiton | Kingston upon Thames | 51.409999 | -0.287396 | 4 | Food | Indian Restaurant | Italian Restaurant | Platform | Pub | Wine Shop | Rental Car Location | Hardware Store |
| 12 | Seething Wells | Kingston upon Thames | 51.392642 | -0.314366 | 4 | Indian Restaurant | Coffee Shop | Pub | Café | Hotel | Fast Food Restaurant | Fish & Chips Shop | Chinese Restaurant |
| 13 | Surbiton | Kingston upon Thames | 51.393756 | -0.303310 | 4 | Coffee Shop | Pub | Grocery Store | Italian Restaurant | Pharmacy | Garden | Train Station | French Restaurant |

Visualising the clustered neighborhoods on a map using the folium library




# 5. Discussions

The aim of this project is to help people who want to relocate to the safest borough in London, people can chose 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 has Train stations as the most common venues. If a person is looking for a neighborhood with stores and restaurants in a close proximity then the neighborhoods in the second cluster and fifth cluster is suitable. For a family the neighborhoods in Cluster 5 is more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Restaurants, pubs which is ideal for a family.

# 6. Conclusions

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 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 based on safety and a predefined budget.