# Capstone Project - The Battle of the Neighborhoods

Soubam Kohinoor(Oct 2019)

# 1. Introduction

▶ **Background:** Safety is top concern when moving to new area. If you don't feet safety in your own home , you are not going to be able to enjoy living there.

▶ **Problem:** The project aim to select the safest borough in London based on the total crimes , explore the neighborhoods of the borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhood using k-mean clustering

▶ **Interest:** Expats who are considering to relocate to London will be interest to identify the safest borough in London and explore its neighborhood and common venues around each neighborhood.

# 2. Data Acquisition and Cleaning

▶ **Data Acquisition:** The data acquired for this project is a combination of data from three sources.

  ▶ The first data source of the projects uses London crimes data that shows the crime per borough in London

  ▶ The second source of data is scraped from a Wikipedia page that contain the list of London boroughs. This page contains additional information about the boroughs.

  ▶ The third data source is the list of Neighborhood in the Royal Borough of the Kingston upon Thames as found on the Wikipedia page.

# 2. Data Acquisition and Cleaning

▶ **Data Cleaning:** The data cleaning process for each of the three sources of data are 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.

▶ 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.

▶ The two data sets are merged on the Borough names to form a new data set. The purpose of this data set is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

▶ After visualizing the crime in each borough we can find the borough with the lowest crime rate. The third data set is created, with the names of the neighborhoods and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.

▶ The new data set 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

▶ **Exploring Data Analysis**

## Statistical summary of crimes

| | 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 |

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. Methodology

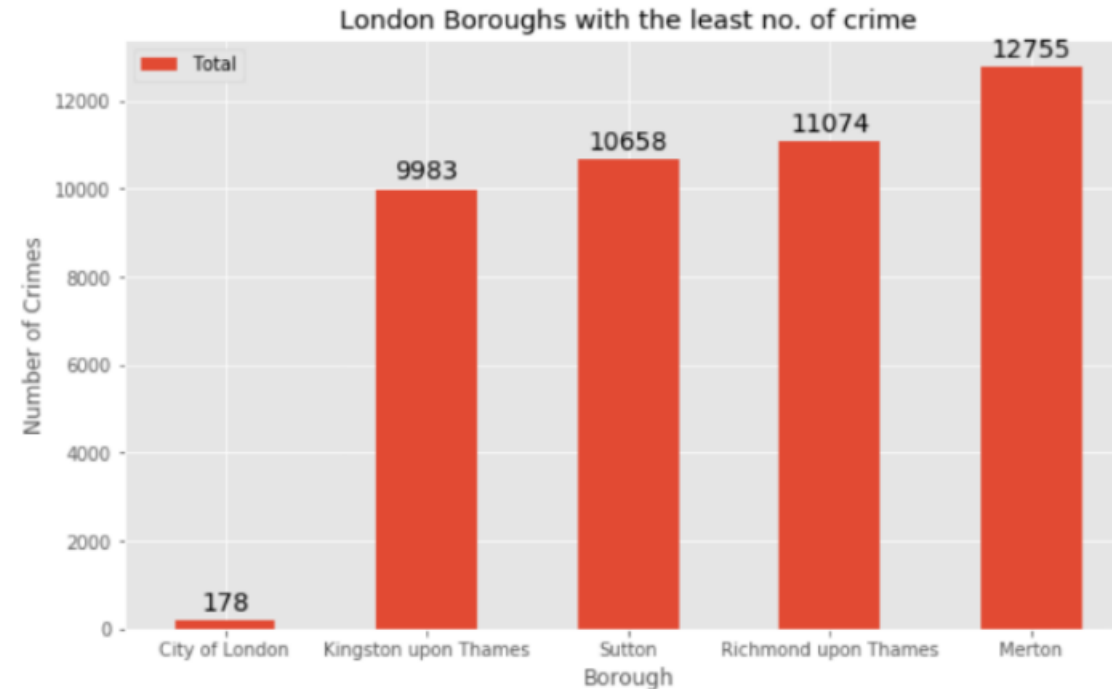## Boroughs with highest crimes rate
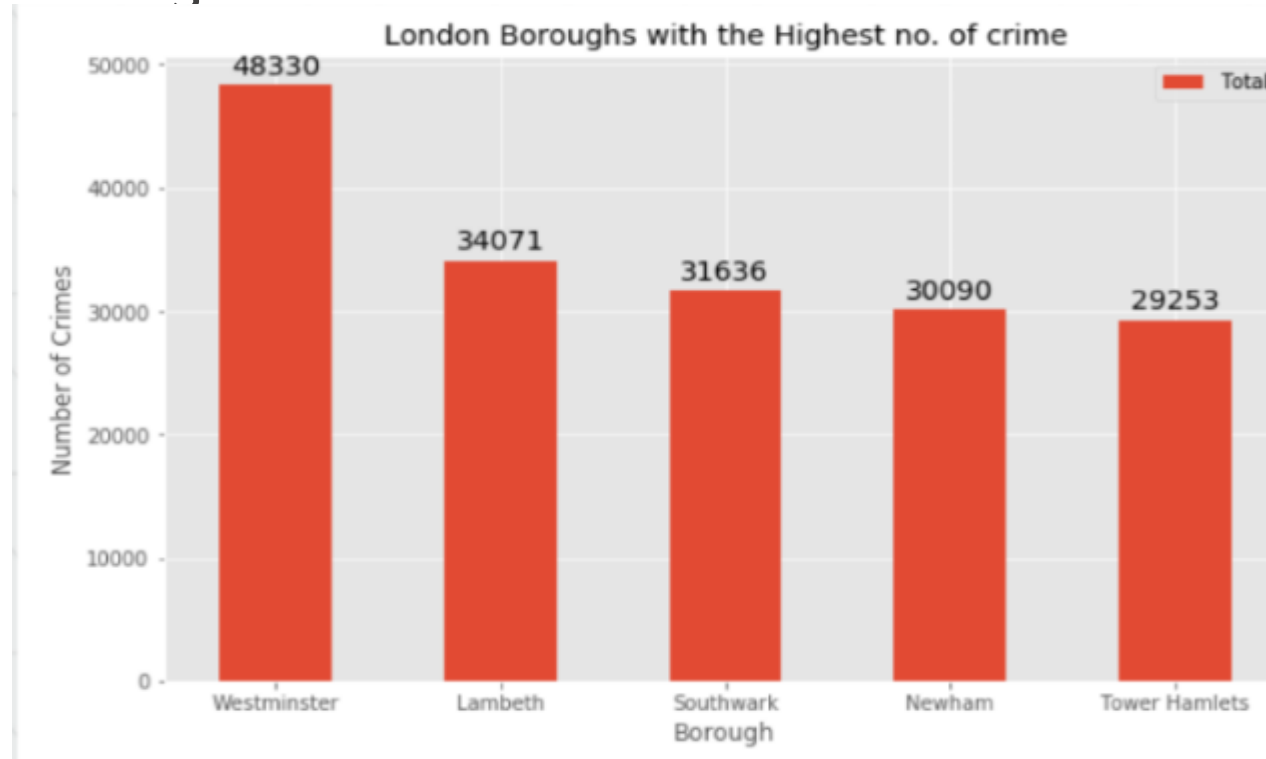


Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the h...
crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly hi...
crime rate than the other 4 boroughs.

# 3. Methodology

Boroughs with lowest crimes rat



London Boroughs with the least no. of crime

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 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.

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

# 3. Methodology

Neighborhood in Kingston upon Thames

There are 15 neighborhoods in the royal boroughs of Kingston upon Thames, They are visualized on a map using folium on python
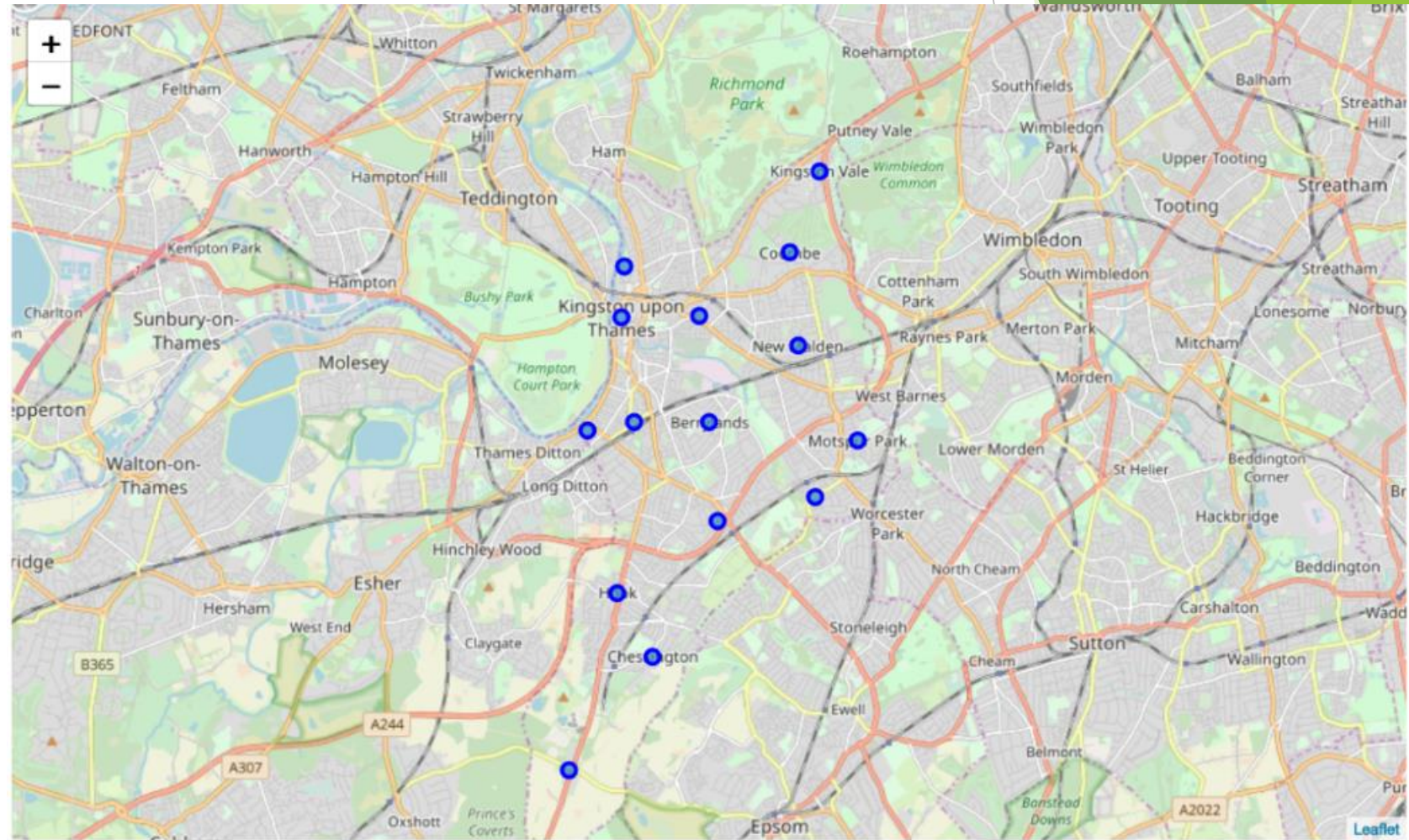
# 3. Methodology

**Modelling :**

► Using the final data set 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.

| | 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 |

• One hot encoding is done on the venues data. 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. Visualizing the clustered neighborhoods on a map using the folium library.
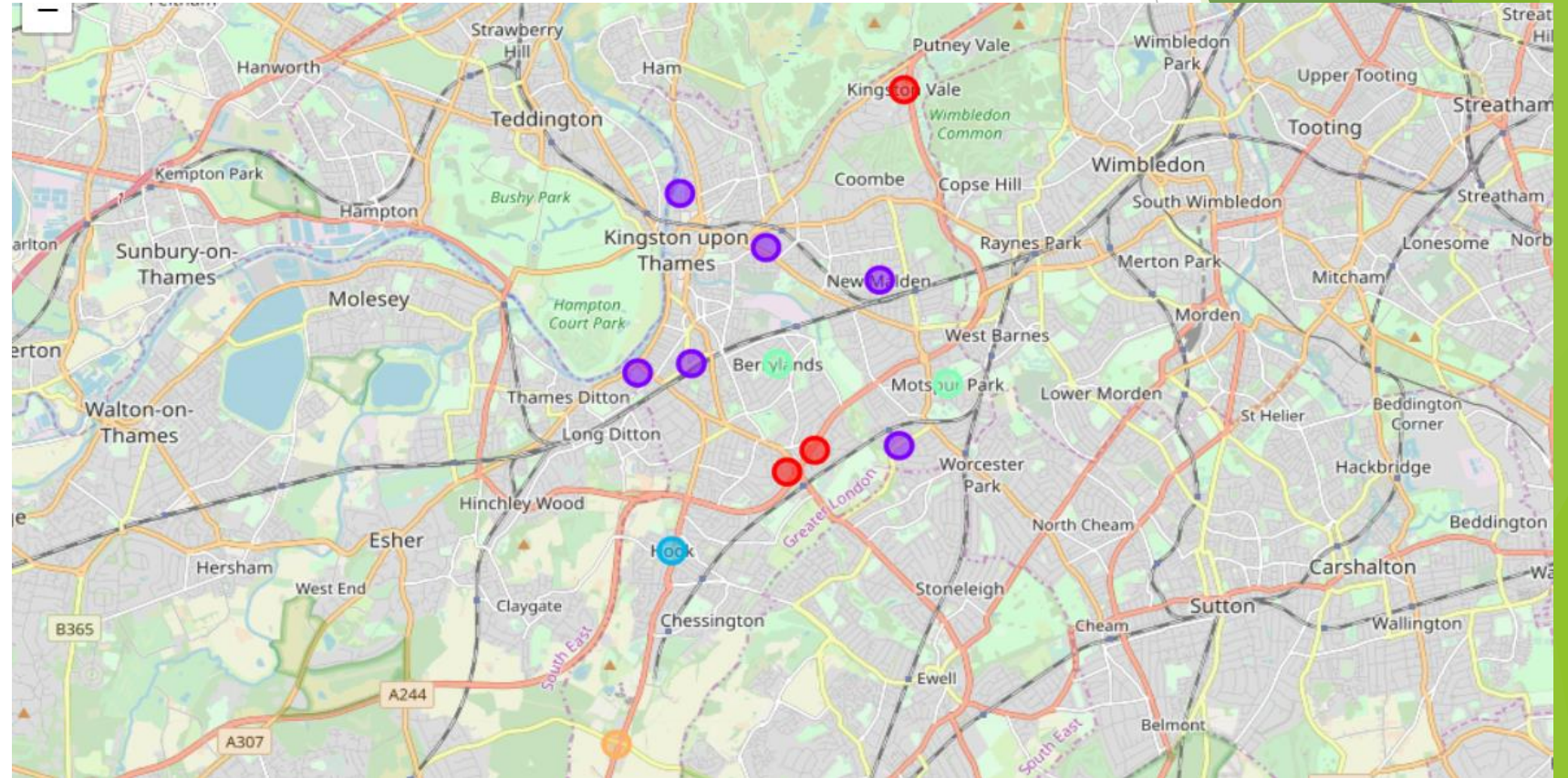


Each cluster is Color coded for the ease of presentation, we can see that majority of the neighbourhood falls in the purple cluster which is the Second cluster. Three neighbourhoods have their own cluster (Blue, Red and Yellow), these are clusters one three and one. The green cluster consists of two neighbourhoods which is the 4th cluster.

# 4.Results

Cluster1 : Looking into the neighborhoods in first cluster

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | Kingston upon Thames | Kingston upon Thames | 51.381776 | -0.276987 | 0 | Grocery Store | Pharmacy | Sandwich Place | Pizza Place | Discount Store | Bowling Alley | Bus Stop | Coffee Shop | Restaurant | Fish & Chips Shop |
| 6 | Kingston Vale | Kingston upon Thames | 51.431850 | -0.258138 | 0 | Grocery Store | Bar | Italian Restaurant | Soccer Field | Farmers Market | Golf Course | Gastropub | Garden Center | Furniture / Home Store | Fried Chicken Joint |
| 14 | Tolworth | Kingston upon Thames | 51.378876 | -0.282860 | 0 | Grocery Store | Pharmacy | Sandwich Place | Italian Restaurant | Furniture / Home Store | Discount Store | Bowling Alley | Train Station | Soccer Field | Bus Stop |

The first cluster having common venues in these neighbourhoods are Grocery Stores and Pharmacy and Restaurants

# 4.Results

Cluster2 : Looking into the neighborhoods in Second cluster

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Canbury | Kingston upon Thames | 51.417499 | -0.305553 | 1 | Pub | Park | Gym / Fitness Center | Supermarket | Spa | Fish & Chips Shop | Shop & Service | Plaza | Hotel | Café |
| 9 | New Malden | Kingston upon Thames | 51.405335 | -0.263407 | 1 | Gym | Bar | Sushi Restaurant | Supermarket | Gastropub | Indian Restaurant | Chinese Restaurant | Korean Restaurant | Fast Food Restaurant | Golf Course |
| 10 | Norbiton | Kingston upon Thames | 51.409999 | -0.287396 | 1 | Platform | Food | Indian Restaurant | Italian Restaurant | Pub | Bagel Shop | Grocery Store | Farmers Market | Hardware Store | Hotel |
| 11 | Old Malden | Kingston upon Thames | 51.382484 | -0.259090 | 1 | Train Station | Indian Restaurant | Food | Pub | Wine Shop | Farmers Market | Golf Course | Gastropub | Garden Center | Furniture / Home Store |
| 12 | Seething Wells | Kingston upon Thames | 51.392642 | -0.314366 | 1 | Indian Restaurant | Coffee Shop | Pub | Café | Harbor / Marina | Fast Food Restaurant | Golf Course | Italian Restaurant | Restaurant | Chinese Restaurant |
| 13 | Surbiton | Kingston upon Thames | 51.393756 | -0.303310 | 1 | Coffee Shop | Pub | Grocery Store | Hotel | Farmers Market | Lounge | Deli / Bodega | Pharmacy | Train Station | Gym / Fitness Center |

The second cluster is the largest cluster which having venues such as Platform , Restaurant , Park, Pubs etc

# 4.Results

Cluster3 : Looking into the neighborhoods in Third cluster

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | Hook | Kingston upon Thames | 51.367898 | -0.307145 | 2 | Bakery | Indian Restaurant | Convenience Store | Fish & Chips Shop | Wine Shop | Fast Food Restaurant | Grocery Store | Golf Course | Gastropub | Garden Center |

The third cluster has one neighbourhood which are having Bakery, Restaurant, Fish & Chips shop , wine shop

# 4.Results

Cluster4 : Looking into the neighborhoods in Fourth cluster

| | 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 | 3 | Wine Shop | Bus Stop | Park | Gym / Fitness Center | Fast Food Restaurant | Grocery Store | Golf Course | Gastropub | Garden Center | Furniture / Home Store |
| 8 | Motspur Park | Kingston upon Thames | 51.390985 | -0.248898 | 3 | Soccer Field | Bus Stop | Park | Gym | Wine Shop | Farmers Market | Golf Course | Gastropub | Garden Center | Furniture / Home Store |

The Fourth cluster has two neighbourhood which are having Fields, Bus stop , Park, Garden etc

# 4.Results

Cluster5 : Looking into the neighborhoods in Fifth cluster

| | 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 | 4 | Pub | Garden Center | Restaurant | Convenience Store | Discount Store | Golf Course | Gastropub | Furniture / Home Store | Fried Chicken Joint | French Restaurant |

The Fifth cluster has one neighbourhood which are having Pubs, Garden, Restaurant and stores.

# 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 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 2 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 close proximity then the neighborhoods in the first cluster is suitable.

- For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Garden which is ideal for a family.

- The preference of venues may vary from person to person, they can select a neighborhood based on ones priorities.

# 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.