The Battle of Neighborhoods: Report
By Prabhat Ranjan

1. Introduction

1.1. Business Problem:

Due to crime rate in London, how do we determine which borough in London will be the safest for staying?

Our objective for this assignment is to find 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**.

1.2. Target Audience:

The target audience for the report will be the people who are looking to relocate to London or looking for safest borough in London. In order to finalize a neighborhood to hunt for an apartment, safety is considered as a top concern when moving to a new place. The crime statistics will provide an insight into this issue.

2. Data Acquisition & 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 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 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

London crime data after preprocessing:

	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. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form . This is important because we will be merging the two datasets together using the Borough names.

List of London Boroughs:

	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	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

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.

London Borough crime:

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Burglary	Criminal Damage	Drugs	Other Notifiable Offences
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
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	3402	2183	906	499
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

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.

	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		

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

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

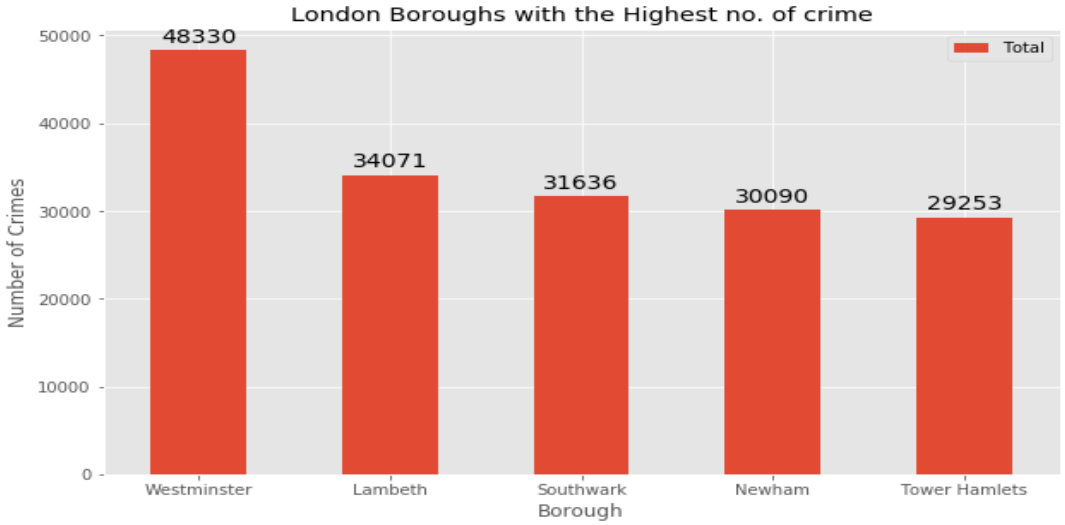
The describe function in python is used to get statistics of the London crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime

	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'.

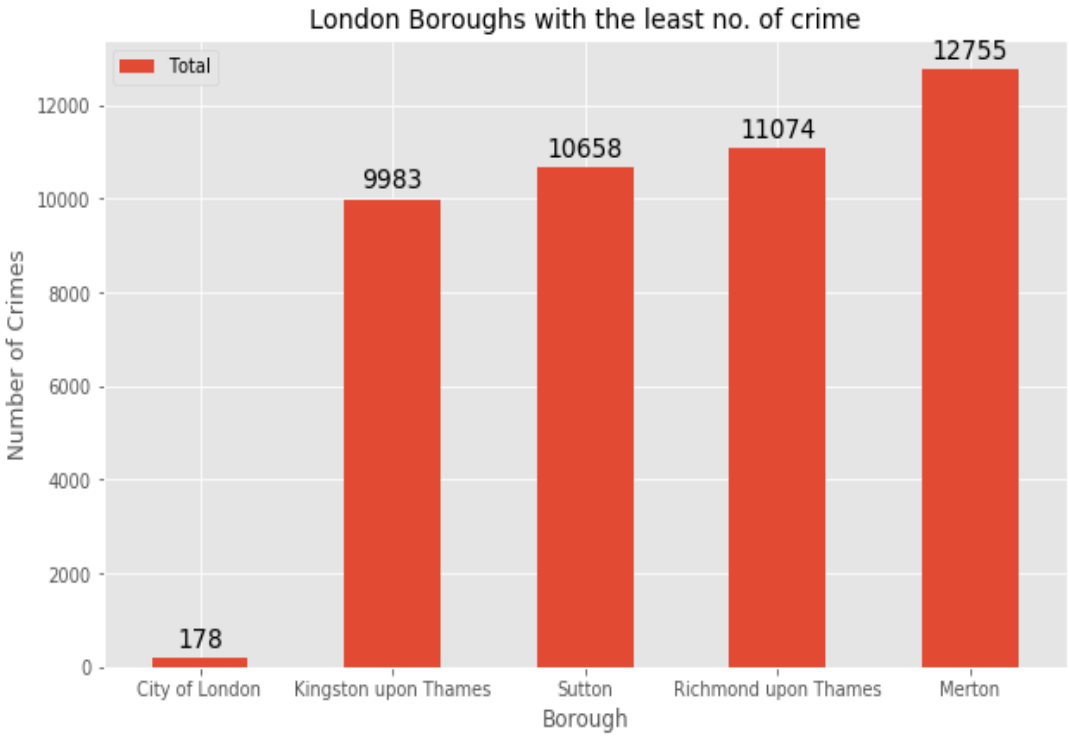
3.1.2. Boroughs with the highest crime rates

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



3.1.3. 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

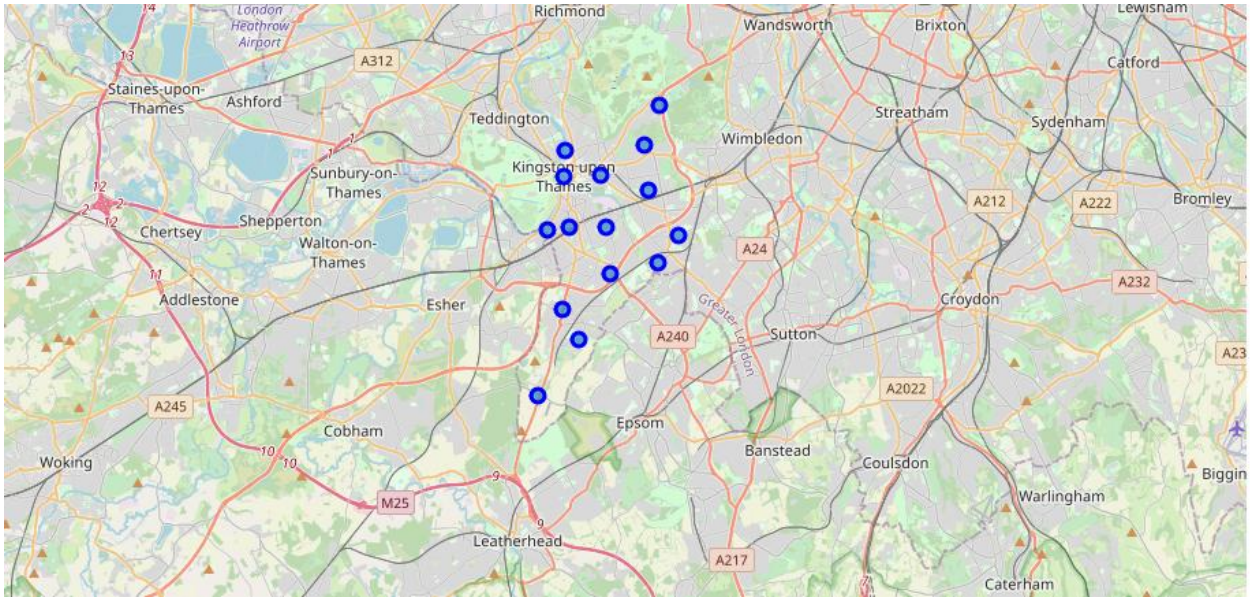


City of London has a significantly lower crime rate because it i is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area. Hence we will consider the next borough with 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

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



3.2. Modeling

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	Kamala Food and Wine	51.397810	-0.284045	Wine Shop
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

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.

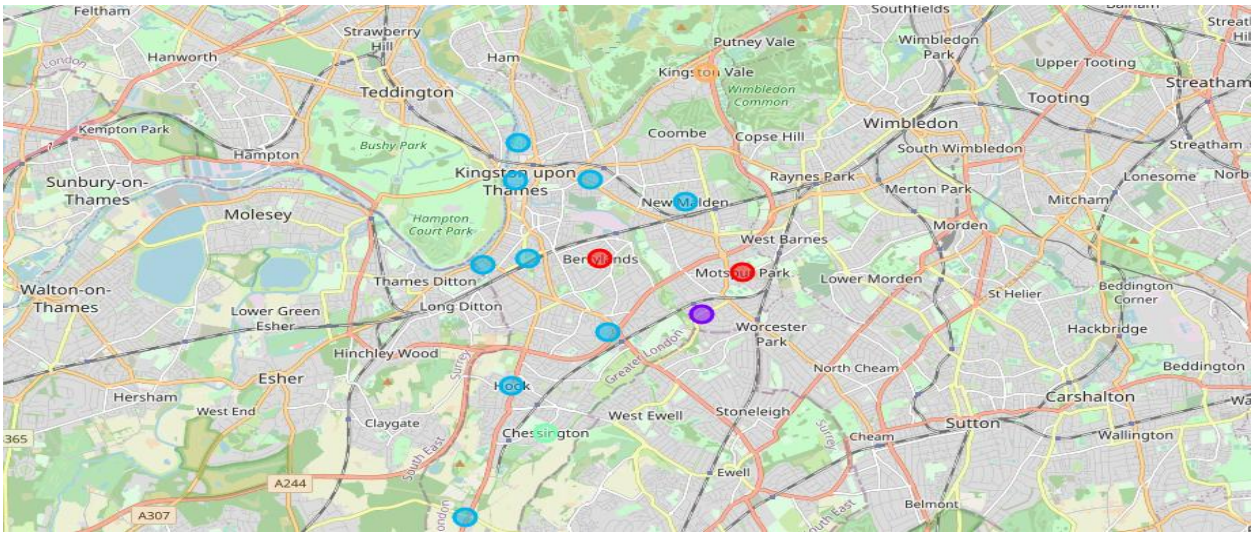
To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.

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 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.



Cluster 1: Looking into the neighborhoods. 1n the 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 Mo Commc Venu
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	0	Wine Shop	Park	Bus Stop	Gym / Fitness Center	Thai Restaurant	Golf Course	Gastropub	Garden Center
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	0	Gym	Bus Stop	Park	Restaurant	Soccer Field	Furniture / Home Store	Fried Chicken Joint	French Restaura

The cluster one has 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 wine shop, gym, park, bus stop etc.

Cluster 2: Looking into the neighborhoods in the 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 M Comm Ver
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	1	Food	Pub	Construction & Landscaping	Train Station	German Restaurant	Gastropub	Garden Center	Furnitu / Home Store

The second cluster has one neighborhood which consists of Venues such as food, pub, construction & landscaping etc.

Cluster 3: Looking into the neighborhoods in the 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 Mos Commo Venu
1	Canbury	Kingston upon Thames	51.417499	-0.305553	2	Pub	Gym / Fitness Center	Spa	Plaza	Café	Shop & Service	Hotel
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Bakery	Indian Restaurant	Supermarket	Fish & Chips Shop	Wine Shop	Food	Departmen Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	2	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Asian Restaurant	Furniture / Home Stor
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Fast Food Restaurant	Deli / Bodega
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Gastropub	Bar	Sushi Restaurant	Supermarket	Indian Restaurant	Chinese Restaurant	Korean Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	2	Food	Indian Restaurant	Italian Restaurant	Platform	Pub	Wine Shop	Rental Car Location

The third cluster has nine neighborhoods which consists of Venues such as Pub, grocery stores, coffee shop, food, Restaurants, bakery shops etc.

Cluster 4: Looking into the neighborhoods in the 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 M Com Ve
2	Chessington	Kingston upon Thames	51.358336	-0.298622	3	Construction & Landscaping	Food	Deli / Bodega	Department Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish Chips Shop

The fourth cluster has one neighborhood in it, this neighborhood have common venues such as construction/landscaping, food. Deli bodega, electronic stores, wine shop etc.

Cluster 5: Looking into the neighborhoods in the 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Soccer Field	Grocery Store	Bar	Sandwich Place	Food	Department Store	Electronics Store	Farmers Market

The fifth cluster has one neighborhood which consists of Venues such as soccer field, grocery store, bar, sandwich place, food, farmers market etc.

5. Discussions

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 1 and 2 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 third cluster is suitable. For a family I feel that the neighborhoods in Cluster 1, 4 & 5 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, Electronics Stores and Soccer fields which is ideal for a family.

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 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.