Capstone Project - The Battle of Neighborhoods
Report

IDDI ABDUL AZIZ

MARCH, 2020.

1. Introduction

1.1 Background

As the capital of the United Kingdom, London is a large and bustling city with an incredible number of things to see and do. Visitors and residents alike are able to experience a modern, developed city that also holds a long royal history.

What are the best and safest places to live in London? This is a question that will be asked by many people as they prepare for their next big move. However, what may be the best place in London for someone else, might not be the best place for you. It is a big decision that requires a lot of research. But we all know that; serene and safety environs is what everyone would think of when locating to a new place. Families, ambitious young professionals, and those seeking an adventure all yearns for a safety, comfortable, and peaceful environments. Safety they say brings first aid to the uninjured.

In our hunt for an apartment, Edward Coke's dictum "Precaution is better than cure" reminds that safety is deemed prudent and a top concern when relocating to a new place. There is no delight when you don't feel safe at home. The crime statistics will provide an insight into this issue.

In this project we will try to solve this problem by using location data and machine learning techniques to identify family friendly neighborhoods in London, UK. To determine which neighborhoods are deemed safe; we will focus on neighborhoods with low crime rates.

1.2 Problem

The London crime dataset obtained from Kaggle datasets, in it are crimes in each Boroughs of London ranging from 2008 to 2016. Though its been four (4) years now and crimes in each Borough might have changed over time. But in this project, we assume 2016 being the latest year.

This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of that borough to find the ten (10) most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Stakeholders interest

This project is would be of interest to families looking for new homes, expats, stakeholders who offer housing to families or city planners who are looking to make neighborhoods safer. Additionally, Police forces and safety officials would benefit from knowing which places have higher crimes so they can implement more resources in those locations to lower crime.

2. Data

2.1 Data Utilized

For this project we need data about:

- Ø London Crime Rates obtained from Kaggle;
- Ø List of London boroughs; obtained and scraped from a Wikipedia page.
- Ø List of neighborhoods in the Royal Borough of Kingston upon Thames also from Wikipedia page.

2.2 Data Cleaning

ü *London Crime Data:*

Data cleaning of the three (3) data sets were done separately; in the London crime data, the latest year (2016), is only selected. The major categories of crime are swiveled to get the total crimes per the boroughs for each major category. *(See the below fig 2.1, After Preprocessing the data)*

[14]:

	Borough	No_of_Crimes							Total
Major_Category		Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	
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
5	Camden	2652	1935	1493	490	899	14088	7626	29183
6	City of London	2	2	10	6	4	129	25	178
7	Croydon	2738	3219	1367	718	1139	9229	10302	28712
8	Ealing	2492	2562	1355	613	669	10040	9396	27127
9	Enfield	2541	2136	1063	492	807	8037	7409	22485
10	Greenwich	1780	2476	867	521	486	8010	8590	22730

ü *List of London Boroughs (Wikipedia):*
The London boroughs data set taken from Wikipedia is scraped using BeautifulSoup library in python. This library is used in extracting the data in tabular format as in the Wikipedia website. String manipulation is obvious after scraping, to get names of the boroughs in the exact format. Its very essential to as we will be merging the two (2) datasets together using Borough names. *(See the below fig. 2.2; List of London Boroughs)*

[22]:

	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

ü *Datasets Merged (London Crime Data & List of London Boroughs):*
In other to gain the necessary information we need; the two datasets are merged into one dataset *(Ld_crime)*. In other to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016. (*See the below fig. 2.3; London Borough Crime*)

[38]:

	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

ü *Safest Borough:*
The borough with the lowest crime rate, automatically is the safest borough; after we visualized the crime rate in each borough. The third set of data is created from scratch, with pandas data frame the names of the neighborhoods and the name of the borough with the latitude left blank. Coordinates of the neighborhood is to be gotten from Google Maps API geocoding to get the final dataset. *(See the below fig. 2.4-5; Safest borough)*

[47]:

	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		

[49]:

	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
10	Norbiton	Kingston upon Thames	51.409999	-0.287396

The Foursquare API will be used to generate and obtain the 10 most common venues for each neighborhood, K-Mean clustering algorithm will be used finally to cluster similar neighborhoods together.

3. Methodology

The methodology in this project consists of two parts:

- 3.1 *Exploratory Data Analysis:* Visualise the crime rates in the London boroughs to identify the safest borough and extract the neighborhoods in that borough to find the 10 most common venues in each neighborhood.

- 3.2 *Modelling:* 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(ML) 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.

3.1 *Exploratory Data Analysis:*

ü *Statistical Summary of Crimes:*

To get the statistical information of London crime data, we use the describe function in python. 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 the below fig 3.1.; London crimes statistical description)*

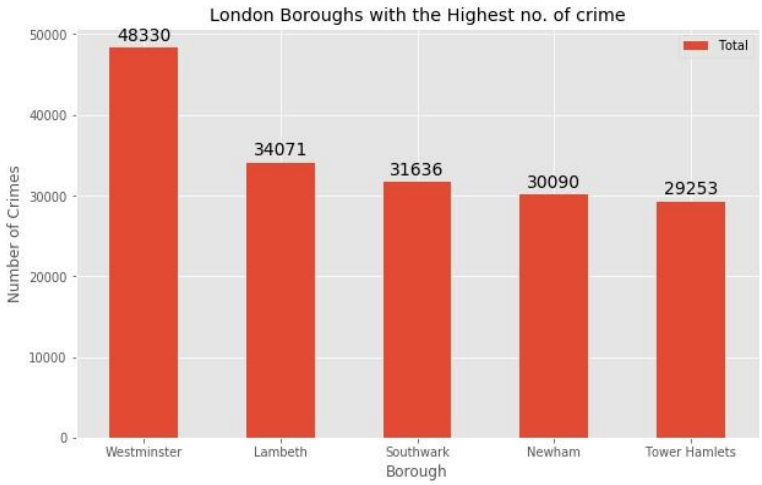
[19]:

	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

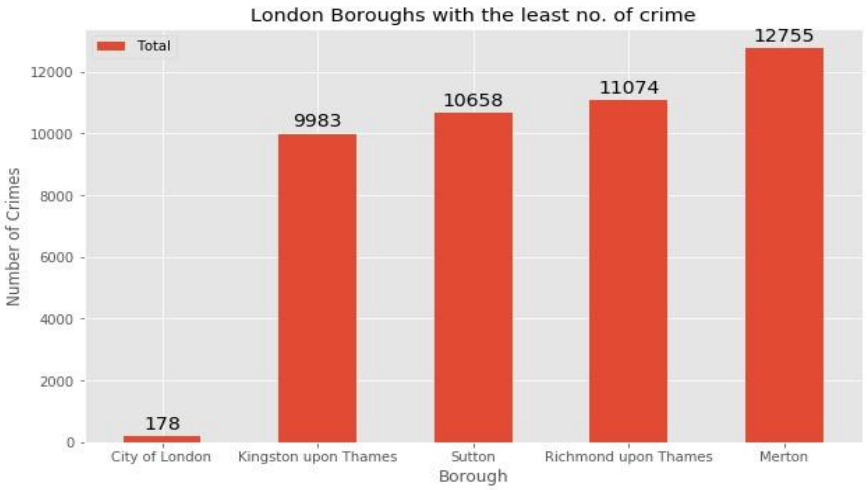
From the above table, the count for each major categories of crime returns the value of 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.

ü *Highest Crime Rate in Borough:*

In comparing the five (5) boroughs with highest crime rate during the year (2016) is quite evident that the Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has significantly higher crimes rate than the other 4 boroughs. *(See the below fig. 3.2; Borough with the highest crime rate)*

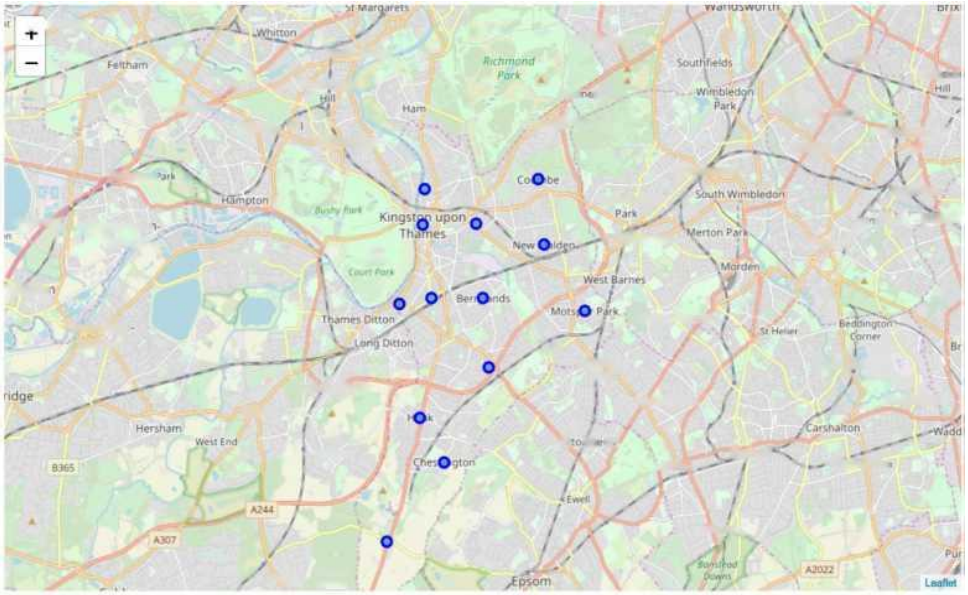


ü *Lowest Crime Rate in Borough:*
It is evident that in comparing the five (5) boroughs with lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton *(See the below fig. 3.3; Borough with the lowest crime rate)*



Obviously from the above diagram (graph), City of London recorded the lowest crime rate. But according to Wikipedia, City of London is the 33rd principal division of Greater London, but it is not a London borough. Hence we consider the next borough (Kingston upon Thames) as the safest borough.

ü *Kingston upon Thames Neighborhoods:*
Fifteen (15) neighborhoods are in the royal borough of Kingston upon Thames. Below is a visualized map using folium on python *(See the below fig. 3.4; Neighborhoods in Kingston upon Thames)*



3.2 *Modelling:*

We will be using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, with that 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 data frame. This data frame contains all the venues along with their coordinates and categories *(See the below fig. 3.2.1; Details of venues of each neighborhood)*

[54]:

	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[1] 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.

In helping interested folks find similar neighborhoods in the safest borough we will be clustering similar neighbourhoods using K-means clustering which is form of unsupervised ML 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 five (5) clusters. Reason is to conduct a K-means clustering to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the amenities (parks, playgrounds, Libraries, coffee, etc) around each neighborhood.

4. Results

After running the K-means clustering we can access each cluster created to see which neighborhood were assigned to each of the five clusters. Looking into the neighborhoods in the first cluster *(See the below fig. 4.1, 4.2, 4.3, 4.4, and 4.5; cluster 1, 2, 3, 5, and 4 respectively)*

[86]:

	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	0	Pub	Café	Spa	Indian Restaurant	Plaza	Hotel	Shop & Service	Gym / Fitness Center	Supermarket	Fish & Chips Shop
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Fish & Chips Shop	Indian Restaurant	Bakery	Supermarket	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Department Store	German Restaurant	Furniture / Home Store	French Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Chinese Restaurant	Korean Restaurant	Sushi Restaurant	Supermarket	Bar	Gym	Indian Restaurant	Electronics Store	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Food	Pub	Italian Restaurant	Fried Chicken Joint	Dry Cleaner	Grocery Store	Hardware Store	Hotel	Japanese Restaurant
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Café	Coffee Shop	Pub	Pet Café	Fish & Chips Shop	Fast Food Restaurant	Golf Course	Chinese Restaurant	Gym
13	Surbiton	Kingston upon Thames	51.399756	-0.303310	0	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Train Station	French Restaurant	Pizza Place	Breakfast Spot	Deli / Bodega

[1] A process by which categorical variables are converted into a form that could be provided to Machine Learning algorithms to do a better job in prediction.

The cluster one (1) is the biggest cluster with seven (7) of the fifteen (15) neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Café, Supermarkets, and stores.

Further looking into other neighborhoods, the second(2nd), third(3rd) and fifth(5th) clusters, have only one neighborhood in each. This is because of the unique venues in each neighborhoods, hence they couldn't be clustered into similar neighborhoods.

[87]:

	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	1	Park	Gym / Fitness Center	Bus Stop	Wine Shop	Farmers Market	Department Store	Discount Store	Dry Cleaner	Electronics Store	Fast Food Restaurant

The above second (2nd) cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wine shops.

[89]:

	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
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Park	Gym	Bus Stop	Soccer Field	Restaurant	Electronics Store	Cosmetics Shop	Deli / Bodega	Department Store	Discount Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Restaurant	Pharmacy	Italian Restaurant	Pizza Place	Café	Bus Stop	Hotel	Discount Store	Bowling Alley

The third (3rd) cluster above has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

[75]:

	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
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	Gastropub

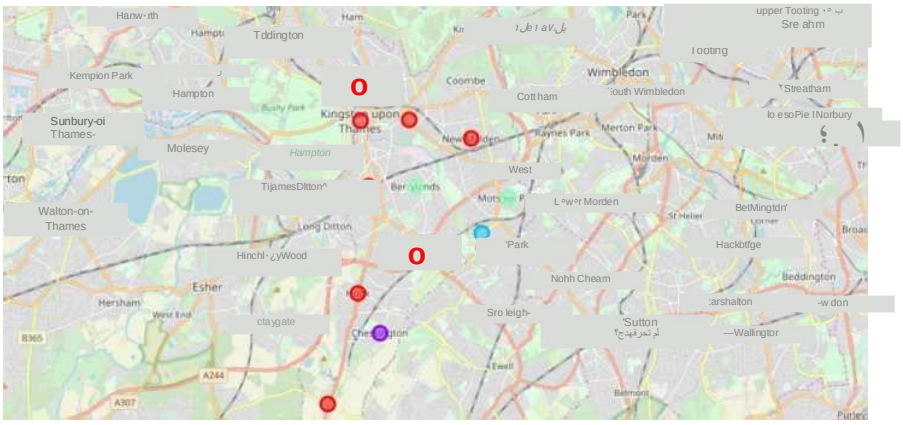
The fifth (5th) cluster above has one neighborhood which consist of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores. Now we will look in the neighborhoods in the fourth (4th) cluster.

[74]:

	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	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store	Farmers Market	Fast Food Restaurant	Food
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	Farmers Market

The fourth (4th) cluster above has two neighborhoods, these neighborhoods have common venues such Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics stores and Soccer fields etc.

Now we are visualizing the clustered neighborhoods on a map using folium library (See the below fig. 4.6; Clustering neighborhoods in the Borough of Kings upon Thames)



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 (1st) cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two (2), three (3) and five (5). The green cluster consists of two neighborhoods which is the fourth(4th) cluster.

5. Discussion

With more neighborhoods to choose from the clusters, it should be relatively easy for and of interest to families looking for new homes, expats, stakeholders who offer housing to families or city planners who are looking to make neighborhoods safer based on their preferences.

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 third (3rd) and fourth (4th) 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(1st) cluster is suitable. For a family I feel that the neighborhoods in Cluster fourth (4th) 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.

In areas of high crime rates should be avoided by families. But those in charge of police forces (Police Commissioners) and safety would benefit from implementing more resources into these areas to bring the crime down.

6. Conclusion

We conclude that there are many safest and family-friendly neighborhoods to choose from. Which is great news as many people believe it is impossible to find safer and family-friendly neighborhoods in Big Cities. Now that you have read about the neighborhoods of London and several of the areas the city has to offer, I hope some of your stress is relieved. Remember to think deeply about what is right for you and your situation, and not get caught up in generalizations. Without a doubt, London is a great place to live, work, and raise a family.
Before making the big move, 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. Do your research and ask people who have lived there. But, above all remember:
It's not the place. It's the people.
Good luck with your choice!