Capstone Project - The Battle of Neighborhoods - Analysis of Safest Boroughs in London

- Report

1. Introduction

1.1 Background

With increased globalization, job opportunities and a liberal world, humans are open to relocate to places other than their native ones. Also, it is human desire to move until they find a place to settle down where they truly feel happy. Sometimes our wants and needs change over time, making us leave a town we once called home and move to an area where we feel satisfied. We may even move to a new area too frequently without knowing exactly what we are getting into, forcing us to leave at the first sign of discomfort.

To reduce the chances of such a thing happening, we should do proper research before planning to move to a new location. Consider a few factors before choosing a new place, so that you don't end up wasting your valuable money and time, relocating to a place that you regret afterwards. Safety is one of the top concerns while relocating, as you would never be able to enjoy living in a place where you don't feel safe.

1.2 Problem

We will be using the crime statistics dataset of London from Kaggle, which consists of the crimes in each borough of London from the year 2008 to 2016. The year 2016 being the latest, we will consider the data of that year, which is actually old information as of now. Although, the crime rates in each borough may have changed over time.

In this analysis we will select the safest borough in London based on total crimes. We will explore the neighborhoods of that borough, finding the 10 most common venues in each neighborhood. We will cluster the neighborhoods using k-means clustering.

1.3 Interest

This project would interest expats who are looking forward to relocating to London, identifying its safest borough. They would like to explore its neighborhoods and common venues around each neighborhood.

2. Data Description and Preparation

2.1 Data Description

The data used in this project has been acquired from three different sources. The first data source is that of London crime data, showing crime in each borough of London. The dataset consists of 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 dataset has been scraped from the wikipedia page containing the list of London boroughs. It contains additional information about the boroughs with following 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 controls 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 wikipedia page providing the list of Neighborhoods in the Royal Borough of Kingston upon Thames. It has following 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

Each of the three datasets have been cleaned separately. The crimes pertaining to the most recent year (2016) have been selected from London crime data. The major categories of crime are pivoted to get the total crimes for each borough as per the category. (see *fig 2.1*)

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	823	1203	549	234	311	3528	3735	10383
1	Barnet	2082	1355	564	291	278	5789	4600	14959
2	Bexley	662	1057	412	176	111	2803	2676	7897
3	Brent	1571	1411	1247	305	513	5620	5563	16230
4	Bromley	1392	1359	440	260	227	4530	3908	12116

Fig 2.1 London crime data after preprocessing

The python library **Beautiful Soup** has been used for scraping the second data from wikipedia page. We have extracted the data in a tabular format using this library. After web scraping, we need to manipulate the string in order to get the names of the boroughs in appropriate form (see *fig 2.2*). This step helps us afterwards in merging the two datasets together using 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	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

Fig 2.2 List of London Boroughs

We merge the two datasets using borough names resulting into a new dataset that combines the necessary information in one dataset (see *fig 2.3*). This dataset will be used to visualize the crime rates in each borough and identify the borough that recorded the least number of crimes during the year 2016.

	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	823	1203	549	234	311	3528	3735	10383
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	2082	1355	564	291	278	5789	4600	14959
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	662	1057	412	176	111	2803	2676	7897
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	1571	1411	1247	305	513	5620	5563	16230
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	1392	1359	440	260	227	4530	3908	12116

Fig 2.3 London Borough Crime

After visualizing the crime in each borough, we can identify the borough with least crime rate and mark that borough as the safest borough. The third dataset has been acquired from the list of neighborhoods in the safest borough on wikipedia. This dataset has been created from scratch. The pandas dataframe has been created with the neighborhood names and borough name with blank latitude and longitude columns (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 Neighborhoods of the safest borough

The neighborhood coordinates can be obtained using **Google Maps API Geocoding** for the final dataset (see *fig 2.5*).

	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 Neighborhoods of the safest borough

The final dataset will be used to generate the venues for all neighborhoods using **Foursquare API**.

3. Methodology

3.1 Exploratory Data Analysis

3.1.1 Statistical Summary of Crimes

We use the describe function to get the statistics of London crime data. The mean, standard deviation, minimum, maximum, 1st quartile (25%), second quartile (50%) and third quartile (75%) for all 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	1256.515152	1195.121212	719.151515	285.151515	411.696970	5429.636364	4272.969697	13570.242424
std	451.837991	380.840334	362.618526	117.046497	264.500294	2850.484858	1515.739743	5377.517945
min	1.000000	1.000000	6.000000	1.000000	3.000000	70.000000	15.000000	97.000000
25%	904.000000	990.000000	475.000000	222.000000	230.000000	3629.000000	3661.000000	10398.000000
50%	1255.000000	1238.000000	667.000000	291.000000	362.000000	5423.000000	4578.000000	13854.000000
75%	1571.000000	1448.000000	984.000000	331.000000	558.000000	6662.000000	5340.000000	16601.000000
max	2082.000000	1926.000000	1741.000000	619.000000	1090.000000	17073.000000	6417.000000	29500.000000

Fig 3.1.1 Statistical description of London crimes

There are 33 boroughs in London, which can be seen in the count for each of the major categories of crime. We can see that 'Theft and Handling' is the highest reported crime in the

year 2016, which is followed by 'Violence against the person'. The least reported crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.

3.1.2 Boroughs with the lowest crime rates

After comparing the five boroughs with the lowest crime rate during the year 2016, we found that the City of London has the least recorded crimes, followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton. (see fig 3.1.2)
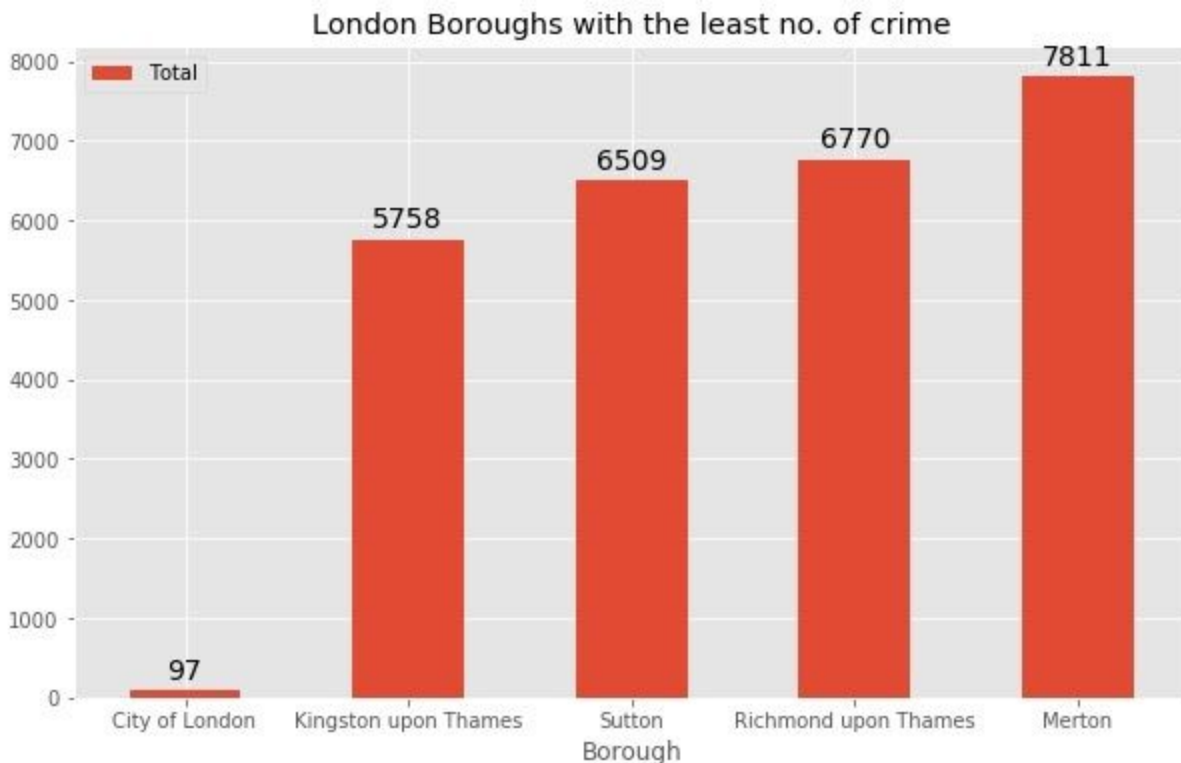

Fig 3.1.2 Boroughs with lowest crime rates

The City of London is 1.12 square miles in area and has a population of 7000 as of 2013, indicating that it is small. Since, City of London is not a London borough but the 33rd principal division of Greater London, we will consider the next borough with lowest crime rate as the safest, which is Kingston upon Thames.

	Borough	Total	Area (sq mi)	Population (2013 est)[1]
6	City of London	97	1.12	7000

Fig 3.1.3 City of London

3.1.3 Neighborhoods in Kingston upon Thames

There are 15 neighborhoods in royal borough of Kingston upon Thames as visualised on a map using folium in python (see fig 3.1.4)
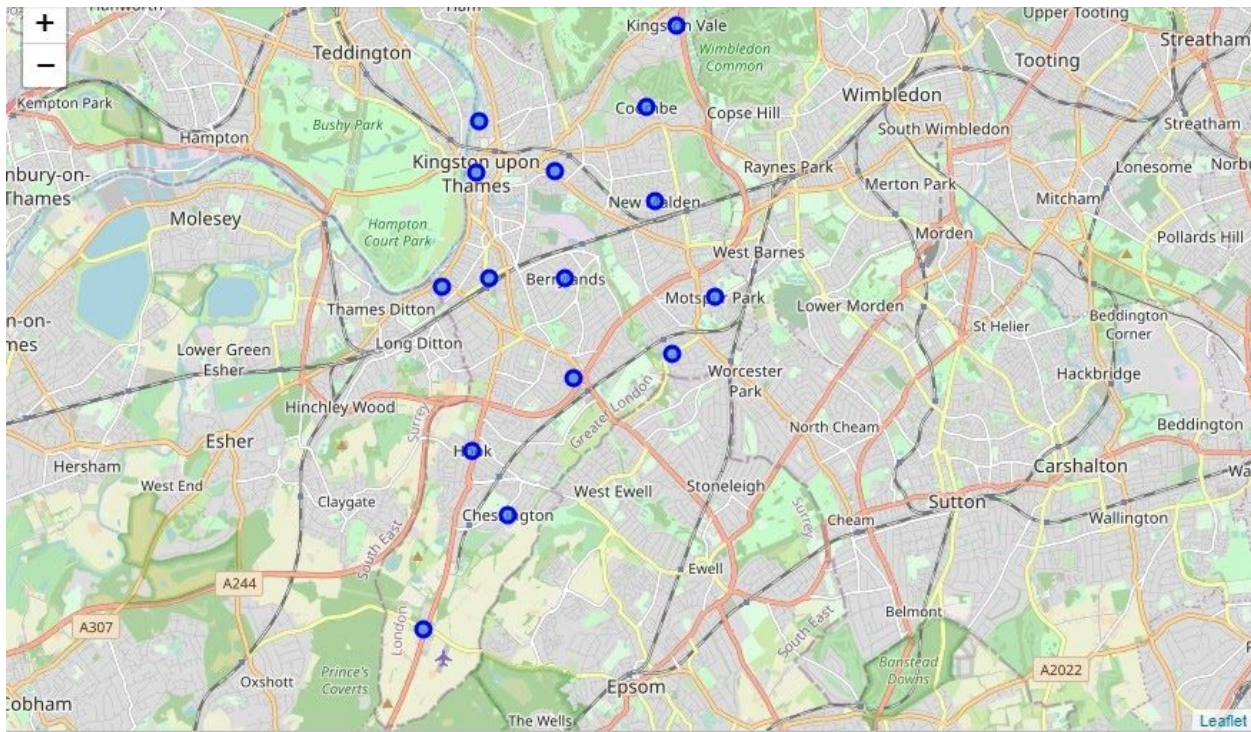

Fig 3.1.4 Neighborhoods in Kingston upon Thames

3.2 Modelling

We can find all the venues within 500m radius of each neighborhood of Kingston upon Thames using the final dataset with the neighborhoods along with latitude and longitude by connecting to Foursquare API. It returns the data in json format with all the venues in each neighborhood, which we convert to pandas dataframe. The dataframe contains all the venues with the 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	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Fig 3.2.1 Venue details of each neighborhood

After this we apply one hot encoding to the venues data, which is then grouped by the neighborhood and the mean of venues are calculated. Finally, the 10 most common venues are calculated for each of the neighborhoods.

Then we use k-means clustering to cluster similar neighborhoods. This will help people to find similar neighborhoods in the safest borough. We will use a cluster size of 5 in this project for clustering 15 neighborhoods into 5 clusters. This will cluster neighborhoods with similar venues together, so that the people would be able to shortlist an area of their interest on the basis of venues/amenities near each neighborhood.

4. Results

After applying k-means clustering, we will be able to access each cluster created to see which neighborhoods were assigned to each of the 5 clusters. Looking into the first cluster, we can see that it has one neighborhood with venues such as train station, restaurant and grocery store.

	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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	0	Construction & Landscaping	Pub	Food	Train Station	Bagel Shop	Department Store	Golf Course	Gift Shop	German Restaurant	Gastropub

Fig 4.1 Cluster 1

The second cluster has three neighborhoods consisting of venues such as wine shops, bars, restaurants, electronics stores, Bus stop, pharmacy, soccer field, farmers market and grocery stores.

	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	1	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Fast Food Restaurant	Department Store	Discount Store	Electronics Store	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Restaurant	Discount Store	Hotel	Bus Stop	Italian Restaurant	Sandwich Place	Bowling Alley	Soccer Field

Fig 4.2 Cluster 2

The third cluster has two neighborhoods with common venues as Gym, park, bus stop, restaurants, electronic stores and 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	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	2	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Food	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Gym	Park	Bus Stop	Restaurant	Soccer Field	Fast Food Restaurant	Department Store	Discount Store	Electronics Store	Farmers Market

Fig 4.3 Custer 3

The fourth cluster is the biggest of all with 6 of the 15 neighborhoods in the Kingston upon Thames borough. We can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

	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	3	Pub	River	Café	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Fish & Chips Shop	Plaza	Park
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	3	Café	Pub	Sushi Restaurant	Coffee Shop	Burger Joint	Asian Restaurant	German Restaurant	Gift Shop	French Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	3	Gym	Sushi Restaurant	Korean Restaurant	Bar	Supermarket	Indian Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	3	Food	Indian Restaurant	Italian Restaurant	Pub	Wine Shop	Rental Car Location	Hardware Store	Hotel	Japanese Restaurant	Coffee Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	3	Indian Restaurant	Coffee Shop	Pub	Gym	Pet Café	Playground	Chinese Restaurant	Café	Restaurant	Fast Food Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	3	Coffee Shop	Pub	Pharmacy	Italian Restaurant	Grocery Store	Gym / Fitness Center	Deli / Bodega	Farmers Market	Gastropub	Hotel

Fig 4.4 Cluster 4

The fifth cluster has one neighborhood which has venues such as wine shops, supermarkets, Restaurants and electronics stores.

	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	4	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Wine Shop	French Restaurant	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant

Fig 4.5 Cluster 5

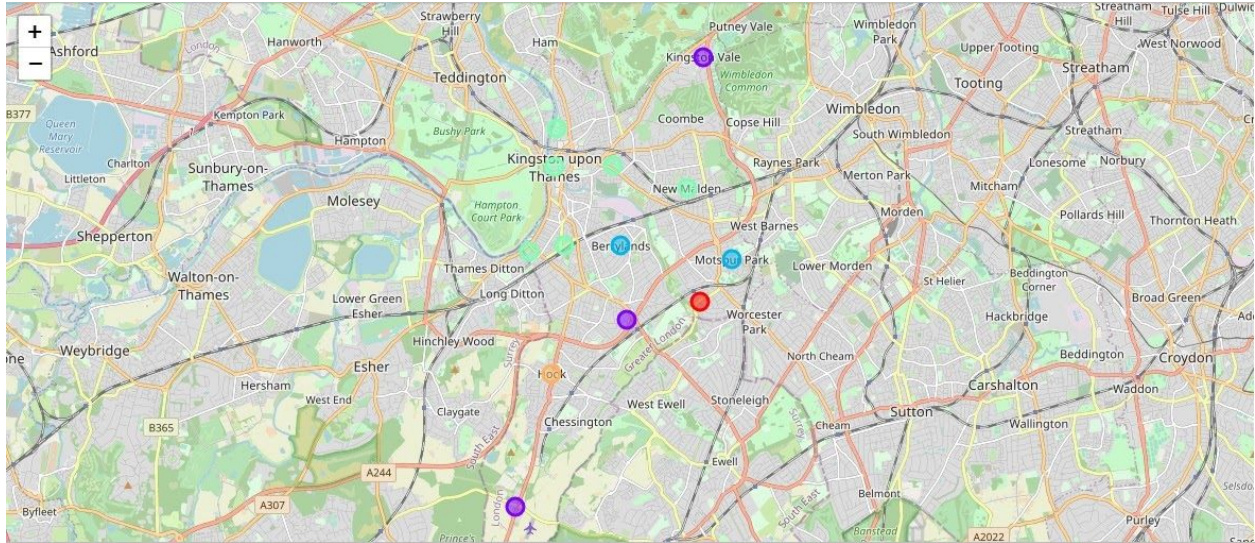We can visualize the clusters on a map using the folium library (see fig 4.6).

Fig 4.6 Clustered neighborhoods in Kingston upon Thames borough

Each cluster is represented by a different color. A majority of the neighborhoods fall in the green cluster, which is the fourth cluster. The Purple cluster has three neighborhoods, which is the second cluster and blue cluster has two neighborhoods, which is the third cluster. Two neighborhoods have their own cluster (Red and yellow), these are cluster 1 and cluster 5.

5. Discussion

This project helps people wanting to relocate to the safest borough in London. Expats can select a neighborhood of their choice based on its most common venues. So, if someone wants to live in a neighborhood with good connectivity and ease of transportation, they can choose from cluster 1, cluster 2 and cluster 3, which have train stations and bus stops as common venues. If the need is to live in a neighborhood with stores and restaurants in close proximity, then the neighborhoods in cluster 2 and 4 are suitable. Cluster 3 is most suitable for families because of the presence of Parks, Gym/Fitness centers, bus stops, restaurants, electronics stores and soccer fields.

6. Conclusion

In this project we have taken safety as a primary concern to choose from a number of boroughs in London. Also, it helps people get a better understanding of the neighborhoods in London with respect to the most common venues in those neighborhoods. Using technology to find out more about places before moving into a neighborhood, helps us to stay a step ahead. This project can further be extended to finding out other factors such as cost of living in selected areas, in order to move to a place that is safe and also well within budget.