Capstone Project - The Battle of Neighborhoods

Applied Data Science Capstone IBM Coursera
Mina Jamshidian
2021

Introduction

Background

People worldwide change several homes in their lives to find the best place to live for their entire lives. The most important parameter for a good place for living in the safety which is helping people to stay in the area for a long time instead of consequently changing their area for finding his project aims to find a good place a soon as possible they can save their money and the makes a question for us that the safe place is one of the important parameters for choosing the home for a stay in a place or people like to change their area by chance?

Problem

The Dataset that was used for this project is "London Crime Data" from Kaggle. This dataset covers the number of criminal reports by month, LSOA borough, and major/minor category from Jan 2008-Dec 2016. This project aims to find the safest borough for finding the five common venues in every neighbourhood. Then clustering the area by using K-mean clustering.

Interest

This project is for someone who wants to immigrate to London. Therefore the most important concern for those who are looking to relocate to a new place is safety. Because if people do not feel safe on their property they would not enjoy their life.

Data

Data Description

The decision would be infected by some factors based on the problem definition:

- The total crimes number per borough in the 2016 year.
- The most common venues per neighbourhood in the safest borough that was selected.

Three datasets would be used for this project:

Following data sources will be needed to extract/generate the required information:

- **First Data**: Preprocessing a real-world data set from Kaggle showing the London Crimes from 2008 to 2016: A dataset consisting of the crime statistics of each borough in London obtained from Kaggle
- **Second Data**: Scraping additional information of the different Boroughs in London from a Wikipedia page.: More information regarding the boroughs of London is scraped using the Beautifulsoup library
- **Third Data**: Creating a new dataset of the Neighborhoods of the safest borough in London and generating their coordinates.: Co-ordinate of the neighbourhood will be obtained using "Google Maps API geocoding".

First Dataset

https://www.kaggle.com/jboysen/london-crime
"London crime data" shows the crime per borough in London. This dataset contains the following columns:

- lsoa_code: code for Lower Super Output Area in Greater London.
- lsoa_code: code for Lower Super Output Area in Greater London.
- rough: Common name for London borough.
- major_category: High-level categorization of crime
- minor_category: Low-level categorization of crime within the major category.
- value: monthly reported count of categorical crime in the given borough.
- year: Year of reported counts, 2008-2016.
- month: Month of reported counts, 1-12.

Second Dataset

https://en.wikipedia.org/wiki/List_of_London_boroughs
"Wikipedia " contains the list of London boroughs and the boroughs information. This dataset contains the 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 another 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): The population in the borough was 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.

Third Dataset

https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames
Creating a new dataset of the Neighborhoods of the safest borough in London and generating their coordinates. This dataset was created from scratch using the list of neighbourhoods available on the Wikipedia site. This dataset contains the following columns:

- Neighbourhood: Name of the neighbourhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

Data Cleaning

First Dataset

From the London crime dataset, the crimes during 2016 which is the last recent year in this dataset were selected. The major categories of crime are pivoted to get the total crimes per borough as per the 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

Figure 1

Second Dataset

This dataset comes from a Wikipedia page that the "Beautiful Soup" library in python was used. The "Beautiful Soup" library can extract the data in the tabular format. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form. This is important for merging two datasets by using the Borough names.
After the two datasets are merged on the Borough names to a new dataset. The purpose of this new dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded in the year 2016.

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]	Population (2019 est) [1]	Co-ordinates	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling
6	City of London	Corporation of London;Inner Temple;Middle Temple	?	Guildhall	1.12	9721.0	NaN	51°30′56″N 0°05′32″W / 51.5155°N 0.0922°W	2	2	10	6	4	129
20	Kingston upon Thames	Kingston upon Thames London Borough Council	Liberal Democrat	Guildhall, High Street	14.38	NaN	177507.0	51°24′31″N 0°18′23″W / 51.4085°N 0.3064°W	879	1054	743	189	121	3803
28	Sutton	Sutton London Borough Council	Liberal Democrat	Civic Offices, St Nicholas Way	16.93	NaN	206349.0	51°21′42″N 0°11′40″W / 51.3618°N 0.1945°W	1233	1316	461	253	165	3516
26	Richmond upon Thames	Richmond upon Thames London Borough Council	Liberal Democrat	Civic Centre, 44 York Street	22.17	NaN	198019.0	51°26′52″N 0°19′34″W / 51.4479°N 0.3260°W	1359	1148	320	217	106	4769
23	Merton	Merton London Borough Council	Labour	Civic Centre, London Road	14.52	NaN	206548.0	51°24′05″N 0°11′45″W / 51.4014°N 0.1958°W	1419	1418	466	249	283	4894

Figure 2

Third Dataset

After visualizing the crime in each borough it would be shown the borough with the lowest crime rate and hence tag that borough as the safest borough. The third dataset is acquired from the list of neighbourhoods in the safest borough on Wikipedia. This dataset is created from scratch, the

panda’s data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude are left blank.
The coordinates of the neighbourhoods are be obtained using "Google Maps API geocoding" to get the final dataset
This dataset would be used to generate the venues per neighbourhood by using the Foursquare API.

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames	51.399008	-0.280911
1	Canbury	Kingston upon Thames	51.423977	-0.275941
2	Chessington	Kingston upon Thames	51.349515	-0.317273
3	Coombe	Kingston upon Thames	51.359819	-0.060080
4	Hook	Kingston upon Thames	51.642031	-0.170341
5	Kingston upon Thames	Kingston upon Thames	51.412928	-0.301858
6	Kingston Vale	Kingston upon Thames	51.403074	-0.303220
7	Malden Rushett	Kingston upon Thames	51.337256	-0.320270
8	Motspur Park	Kingston upon Thames	51.394875	-0.239608
9	New Malden	Kingston upon Thames	51.404433	-0.253936
10	Norbiton	Kingston upon Thames	51.413697	-0.289023
11	Old Malden	Kingston upon Thames	51.384725	-0.261245
12	Seething Wells	Kingston upon Thames	51.393460	-0.315256
13	Surbiton	Kingston upon Thames	51.392411	-0.303999
14	Tolworth	Kingston upon Thames	51.376875	-0.279442

Figure 3

Methodology

Exploratory Data Analysis

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.
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 offences’.

	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

Figure 4

Boroughs with the lowest crime rates

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

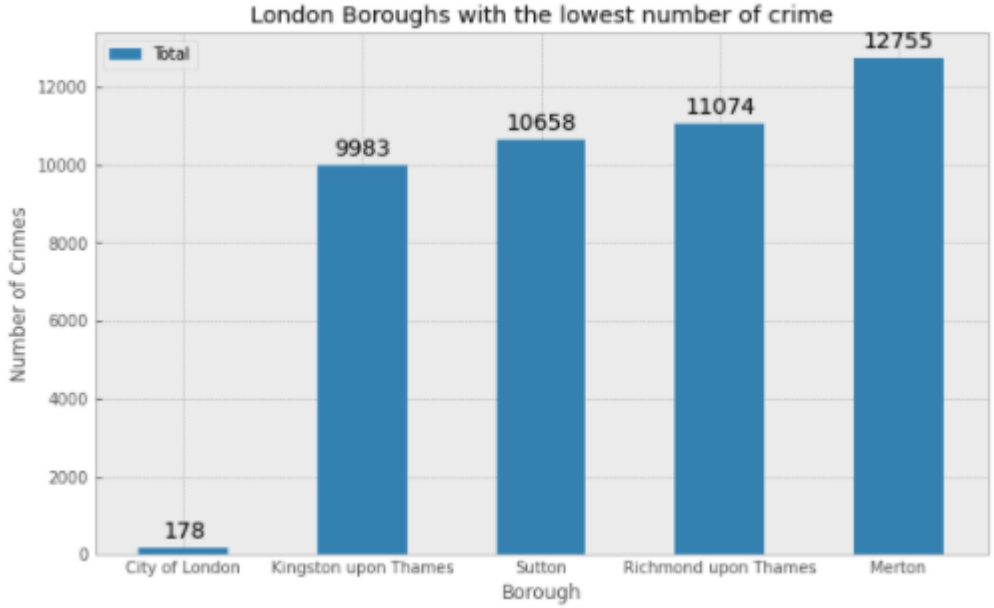

Figure 5

The 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.

Neighbourhoods in Kingston upon Thames

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

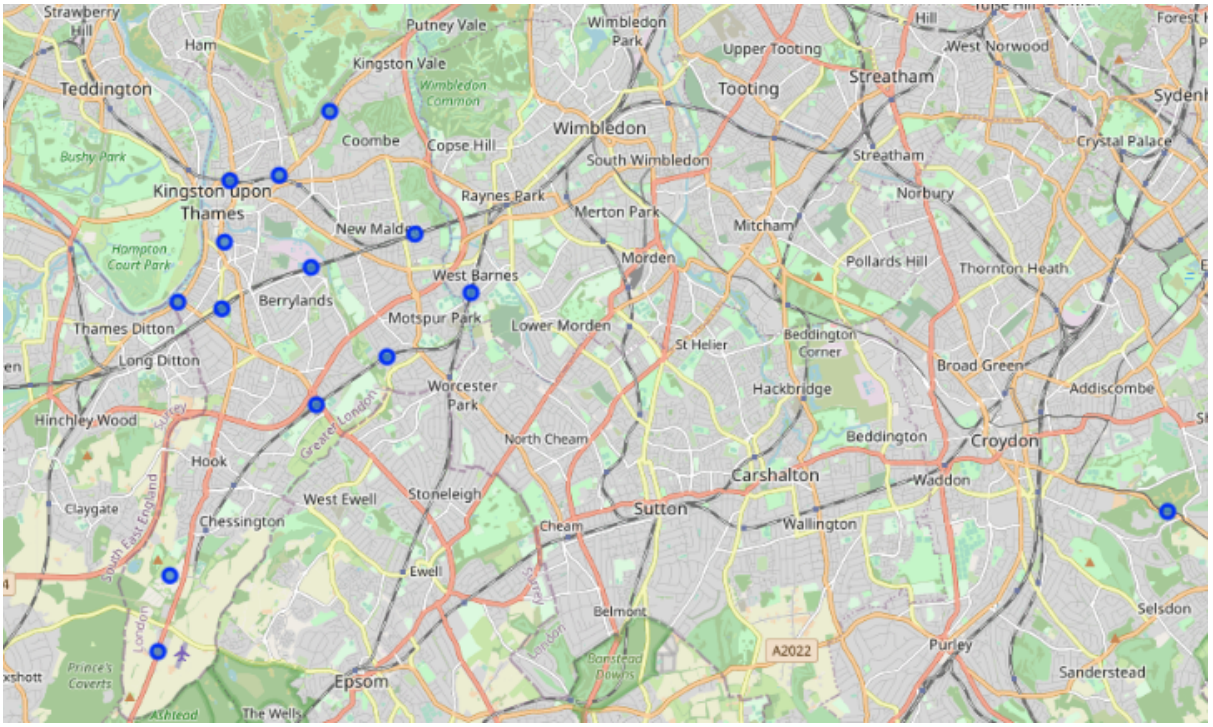

Figure 6

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.

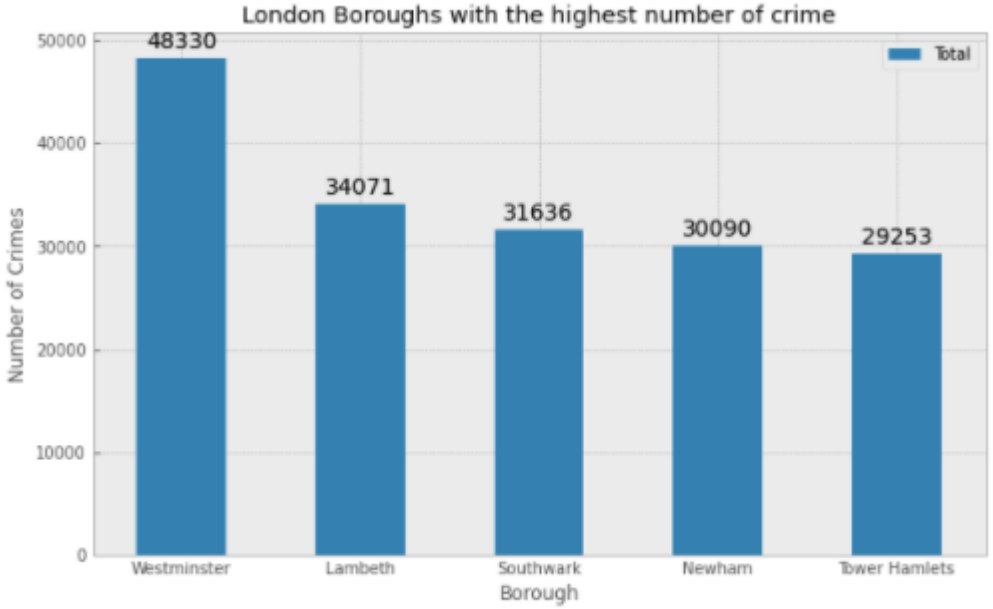

Figure 7

Modelling

Using the final dataset containing the neighbourhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500-meter radius of each neighbourhood by connecting to the Foursquare API. This returns a JSON file containing all the venues in each neighbourhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category.

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 neighbourhoods. To help people find similar neighbourhoods in the safest borough we will be clustering similar neighbourhoods using Kmeans 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 neighbourhoods into 5 clusters. The reason to conduct a Kmeans clustering is to cluster neighbourhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighbourhood.

Results

After running the Kmeans clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five clusters. Looking into the neighbourhoods in the first cluster.

Cluster one is the biggest cluster with 9 of the 15 neighbourhoods in the borough Kingston upon Thames. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Pubs, Cafes, Supermarkets, and stores.

Looking into the neighbourhoods in the second, third and fifth clusters, we can see these clusters have only one neighbourhood in each. This is because of the unique venues in each of the neighbourhoods, hence they couldn't be clustered into similar neighbourhoods.
The second cluster has seven neighbourhood which consists of Venues such as Restaurants, Pub, Hotel, Zoo and park.

The most common venues in these neighbourhoods are Restaurants, Pubs, Cafes, Supermarkets, and stores.

```
1 kut_merged[kut_merged['Cluster Labels'] == 0]
```

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	10tl Mos Commo Venu
Hook	Kingston upon Thames	51.642031	-0.170341	0	Pub	Fish & Chips Shop	Grocery Store	Coffee Shop	Garden	Fast Food Restaurant	Food	Forest	French Restaurant	Frie Chicke Join
Malden Rushett	Kingston upon Thames	51.337256	-0.320270	0	Pub	Grocery Store	Garden Center	Fast Food Restaurant	Fish & Chips Shop	Food	Forest	French Restaurant	Fried Chicken Joint	Furnitu / Hom Ston

Figure 8

The second cluster has seven neighbourhood which consists of Venues such as Restaurants, Pub, Hotel, Zoo and park.

```
kut_merged[kut_merged['Cluster Labels'] == 1]
```

	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 M Comn Ve
2	Chessington	Kingston upon Thames	51.349515	-0.317273	1	Theme Park Ride / Attraction	Zoo Exhibit	Playground	Hotel	Aquarium	Forest	Theme Park	Theater	BBQ J
5	Kingston upon Thames	Kingston upon Thames	51.412928	-0.301858	1	Coffee Shop	Sushi Restaurant	Hotel	Clothing Store	Record Shop	Public Art	Pub	Department Store	Electror St
9	New Malden	Kingston upon Thames	51.404433	-0.253936	1	Korean Restaurant	Coffee Shop	Supermarket	Café	Gym / Fitness Center	Japanese Restaurant	Fast Food Restaurant	Office	Newsag
10	Norbiton	Kingston upon Thames	51.413697	-0.289023	1	Pub	Italian Restaurant	Indian Restaurant	Food	Hardware Store	Pizza Place	Gastropub	Grocery Store	Japan Restau
12	Seething Wells	Kingston upon Thames	51.393460	-0.315256	1	Pub	Indian Restaurant	Café	Harbor / Marina	Fast Food Restaurant	Restaurant	Park	Coffee Shop	Chin Restau
13	Surbiton	Kingston upon Thames	51.392411	-0.303999	1	Coffee Shop	Pub	Grocery Store	Café	Indian Restaurant	Italian Restaurant	Gastropub	French Restaurant	Fis Ch St
14	Tolworth	Kingston upon Thames	51.376875	-0.279442	1	Hotel	Bowling Alley	Coffee Shop	Sandwich Place	Furniture / Home Store	Soccer Field	Pizza Place	Garden Center	Bus S

Figure 9

The third cluster has three neighbourhoods their entire lives to find worldwide is to find the neighbourhood that consists of Venues such as Train stations, Restaurants, zoos, Park.

```
kut_merged[kut_merged['Cluster Labels'] == 2]
```

borhood	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	10t Mo Commo Venu
errylands	Kingston upon Thames	51.399008	-0.280911	2	Platform	Train Station	Coffee Shop	Park	Pub	Zoo Exhibit	Fried Chicken Joint	Fast Food Restaurant	Fish & Chips Shop	Foo
spur Park	Kingston upon Thames	51.394875	-0.239608	2	Park	Pub	Indian Restaurant	Cosmetics Shop	Rugby Pitch	Soccer Field	Mediterranean Restaurant	Train Station	Golf Course	Groce Sto
d Malden	Kingston upon Thames	51.384725	-0.261245	2	Food	Train Station	Construction & Landscaping	Park	Zoo Exhibit	Garden	Fish & Chips Shop	Forest	French Restaurant	Frie Chicke Joi

Figure 10

The fourth cluster has one neighbourhood in it, this neighbourhood has common venues such as Tram Station, Performing Arts Venue, Scenic Lookout, Chinese Restaurant, Zoo Exhibit and Garden.

```
kut_merged[kut_merged['Cluster Labels'] == 3]
```

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 Mos Commo Venu
Coombe	Kingston upon Thames	51.359819	-0.06008	3	Tram Station	Performing Arts Venue	Scenic Lookout	Chinese Restaurant	Zoo Exhibit	Garden	Fish & Chips Shop	Food	Forest	Frenc Restaura

Figure 11

The fifth cluster has two neighbourhoods which consist of Venues such as a Hotel, park, Garden, Gym and Pub.

```
1 kut_merged[kut_merged['Cluster Labels'] == 4]
```

	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 Mo Commc Venu
1	Canbury	Kingston upon Thames	51.423977	-0.275941	4	Hotel	Park	Garden	Bus Stop	Gym Pool	Gym / Fitness Center	Fish & Chips Shop	Food	Forest	Frenc Restaura
6	Kingston Vale	Kingston upon Thames	51.403074	-0.303220	4	Park	Grocery Store	Pub	Hotel	Breakfast Spot	Café	Burger Joint	Hookah Bar	Bowling Alley	Bookst

Figure 12

Visualising the clustered neighbourhoods on a map using the folium library. Each cluster is colour coded for ease of presentation, we can see that majority of the neighbourhood falls in the red cluster which is the first cluster. Three neighbourhoods have their cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two neighbourhoods which is the 4th cluster.

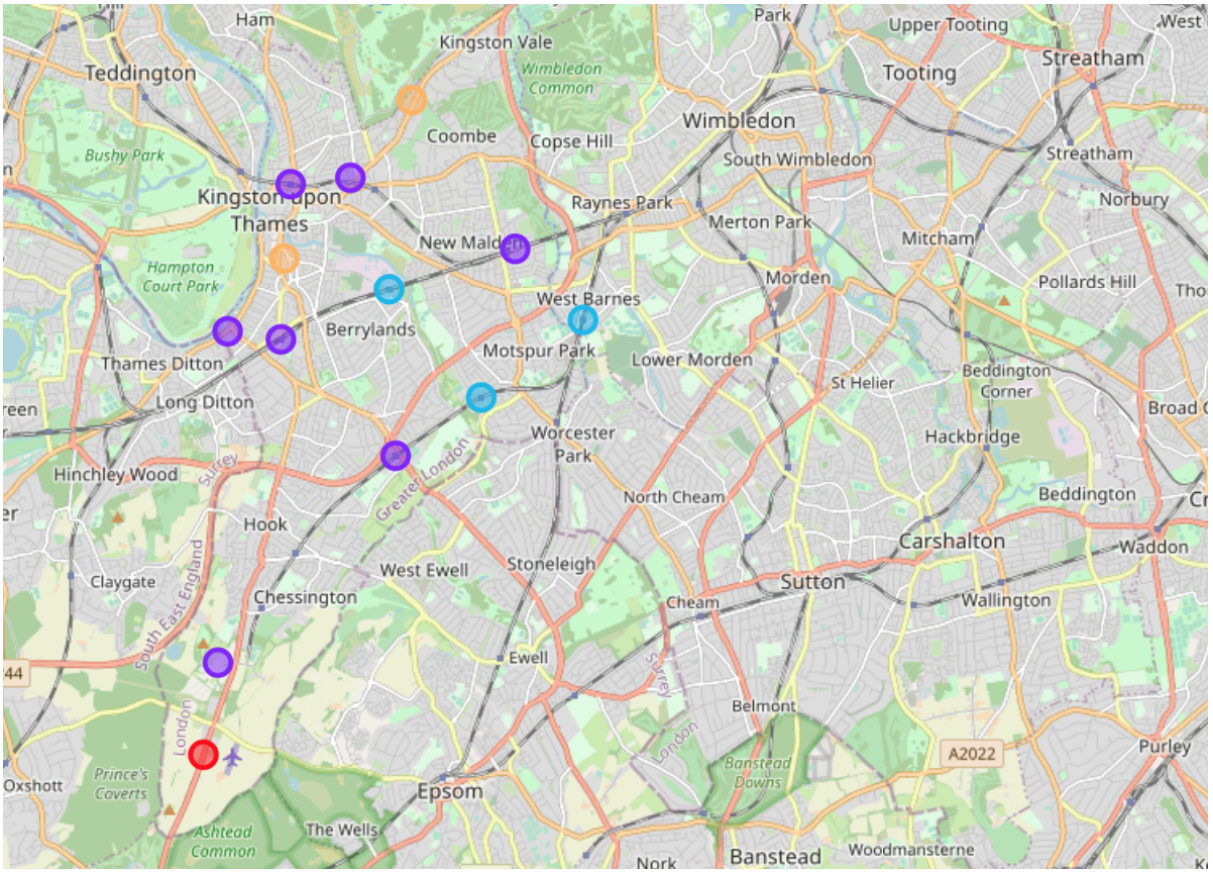

Figure 13

Discussion

This project aims to help people who want to relocate to the safest borough in London, ex-pats can choose the neighbourhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighbourhood with good connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues. If a person is looking for a neighbourhood with stores and restaurants nearby then the neighbourhoods in the first cluster are suitable. For a family I feel that the neighbourhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family. The choices of neighbourhoods may vary from person to person.

Conclusion

This project helps a person get a better understanding of the neighbourhoods concerning the most common venues in that neighbourhood. 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 neighbourhood. 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 the cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.