Applied Data Science - Coursera Capstone Project

The Battle of the Neighbourhoods

Identifying the safest borough in London and exploring its neighbourhoods.

Report By,

Prajwal Premkumar Khairnar

1. Introduction

1.1 Background

We can often observe that people move to different cities for the purpose of education, jobs, business, etc. The UK's migrant population is mostly concentrated in London. Being one of the prime cities in the UK, there are significant number of people who move to London from either other cities or other countries.

While moving to a new city, we must thoroughly research about various factors that may affect our stay and life there. To ensure a comfortable life, a sensible decision must be made to not end up wasting time or money or risking any other factor. Safety in the locality is one of the top most priority, without which one cannot be comfortable with living there.

1.2 Problem

The statistical data for Crimes in London which is available on kaggle.com. It contains data regarding various crimes in different boroughs of London. The dataset information may not be up-to-date and the current crime rates may have changed over time.

The main objective of this project is to exploring the data for London to find the safest borough in London which is based on the total crimes. The neighbourhoods in that borough are explored to find the most common venues in each neighbourhood which may include restaurants, parks, gymnasiums, etc. which can help the user to know the neighbourhoods better and be aware of all the available facilities/resources. Ultimately the neighbourhoods are clustered using k-means clustering algorithm to identify similar neighbourhoods.

1.3 Target Audience:

The people who are considering or are planning to relocate to London will be interested to understand and identify the safest borough in London. When relocating to a new city, especially prominent cities which have vast areas, population, it is very difficult to identify and understand the areas regarding their safety. It is important to make sure that one is fairly informed regarding this factor before relocating to avoid any waste or money, time or unsafe incidents. People who are facing this problem, can use this analysis to easily identify a suitable and rather safest borough in London. Additionally, they can explore the neighbourhoods in that borough and the most common venues in those neighbourhoods. For instance, the food joints, parks, etc. that can be found in every neighbourhood.

2. Data Acquisition and Data Cleaning

2.1. Data Acquisition

This project is implemented using a combination of multiple data sources. Three datasets are used of which some are directly available while some are scraped from Wikipedia.

2.1.1. London Crime Data

This data covers the number of criminal reports by month, LSOA borough, and major/minor category from Jan 2008-Dec 2016. The crimes per borough in London are shown in the dataset. The columns in the dataset are as follows:

- **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

This is the main dataset which consists a total of 13M rows. Each entry/row describes a crime in such a way that we can understand the name of the London borough in which the crime has taken place along with the LSOA code for the area. The further attributes mention the major and minor category in which the crime can be classified. The value which is the monthly count of reported crime for that category of crime in the mentioned borough. The year of the reported entry (2008-2016) along with the month (1-12) are the last two attributes.

2.1.2. London Boroughs

The specific information about the boroughs present in London. This data is scraped from a Wikipedia page. The columns in this dataset are as follows:

- **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)**: 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.

This data consists of detailed information regarding the 33 boroughs in London with a row for each borough with 10 attributes/features for each, which we did not have in the previous dataset. The borough name will be a common column in this and the London Crime Data datasets. The Inner/Outer category of the borough is mentioned along with the status of the borough as Royal, City or Other Borough. The other attributes include the local authority and political control for each borough along with the headquarter location and the total area of the borough. The co-ordinates on the map is an important attribute for each borough. Population for the borough is also an attribute. Nr. In map is a unique number given to each borough just for the purpose of representation on the map.

2.1.3. Neighbourhoods in the Royal Borough of Kingston upon Thames

The list of neighbourhoods data is scraped from Wikipedia page for the Royal Borough of Kingston upon Thames. The columns in this dataset are as follows:

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

The neighbourhoods in the borough which is identified safe are explored using this scraped data. The dataset consists or multiple entries, each associated with a neighbourhood in the Royal Borough of Kingston upon Thames. The neighbourhood name, borough name are the first two attributes. The other attributes mention the latitude and longitude values for that neighbourhood entry.

2.2. Data Cleaning

2.2.1. London Crime Data

From the entire dataset, the data associated with the latest year (2016) is selected for further processing. The major categories of crimes are used and pivoted to get the total crimes in each borough based on 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

Fig. London Crime Data (after Pre-Processing)

Considering the first entry in the above data, we can see that 'Barking and Dagenham' borough is the borough name. The following attributes give us the detailed total summary for the count of crimes based on the specific categories. For instance, total crimes under the burglary category are recorded as 1287, while under Criminal Damage are 1949. Similarly the values for Drugs, Other Notifiable Offences, Robbery, 'Theft and Handling', 'Violence Against the Person' have recorded values as 919, 378, 534, 5607 and 6067 respectively. The last column named as 'Total' gives us the count of total crimes for that borough. Understanding such attribute values for each borough, we can identify the borough with maximum number of crimes and the total counts for each category as well. We can understand the overall characteristics of each borough as far as the safety factor is concerned. The conclusion which can be drawn from this data is valuable, considering the objective of this project.

2.2.2. London Boroughs

The data from Wikipedia page is scraped using the **Beautiful Soup** python library to extract it in a tabular format. String manipulation is required to ensure that the names of Boroughs match with that in the previously collected data. We need the names to match exactly as we will be merging the datasets further.

	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	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

Fig. List of London Boroughs

The scraped data gives us detailed description about various attributes of the boroughs in London. Consider any such entry, for example 'Barking and Dagenham' borough. The various attribute values give us the description that includes information such as local authority which is 'Barking and Dagenham London Borough Council'. The political control is in the category 'Labour'. The headquarter address is also available for each borough along with the geographical co-ordinates for the borough. The area for this borough is 13.93 (sq. mi) and the population is 194,352. The unique number associated is 25 (used to represent on the map). The additional features that this dataset can be used to briefly study any of the boroughs while deriving the conclusions.

2.2.3. Merged Dataset

Visualizing the most crimes in each borough using the dataset created by merging the two previous datasets.

	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	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	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

Fig. London Crimes per Borough

The previous datasets are merged using the common column 'Borough' to combine all the available data together to collectively analyse the data for every borough. The example of 'Barking and Dagenham' here now has all the attributes/features combined which we separately observed in the previous datasets.

2.2.4. Neighbourhoods in the Royal Borough of Kingston upon Thames

The neighbourhoods in the borough which is specifically identified as safe, the Royal Borough of Kingston upon Thames, are explored. The data (name of borough and neighbourhoods) is scraped from the Wikipedia page and the latitude and longitude values are obtained using the Google Maps API geocoding to form the final dataset. Further, Foursquare API will be used to generate venues for each neighbourhood.

	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. Neighbourhoods of the safest boroughs

Focusing on the identified borough, each entry in this dataset now refers to a unique neighbourhood. Each neighbourhood record has a latitude and longitude value associated with it. We can use this location data to use FourSquare API to explore the various venues in these neighbourhoods.

3. Methodology

3.1. Exploratory Data Analysis

3.1.1. Statistical Summary

The `describe()` function is used to obtain the data summary for the London Crime Rate data. This function returns the count, mean, standard deviation, minimum value, maximum value for all the numeric columns. 1st Quartile (25%), 2nd Quartile (50%) and 3rd Quartile (75%) is also returned for the columns. The values returned after the function call on the data is shown below.

	Area (sq mi)	Population (2013 est) [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	33.000000	33.000000
mean	18.408485	255026.878788	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	12.645367	71891.280393	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	1.120000	7000.000000	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	10.360000	215667.000000	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	14.520000	263386.000000	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	21.780000	310516.000000	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	57.970000	372752.000000	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

Fig. Statistical Description of London Crime Data

The count for all the Crimes is 33 as there are total 33 boroughs. The highest reported crime is 'Theft and Handling' followed by 'Violence Against the Person' and 'Criminal Damage'. The categories with least reported crimes are 'Drugs', 'Robbery' and 'Other Notifiable Offences'.

3.1.2. Visualizing total crimes for each Borough

The dataset values for Borough and Total crimes for each borough are visualized to understand the distribution of values.

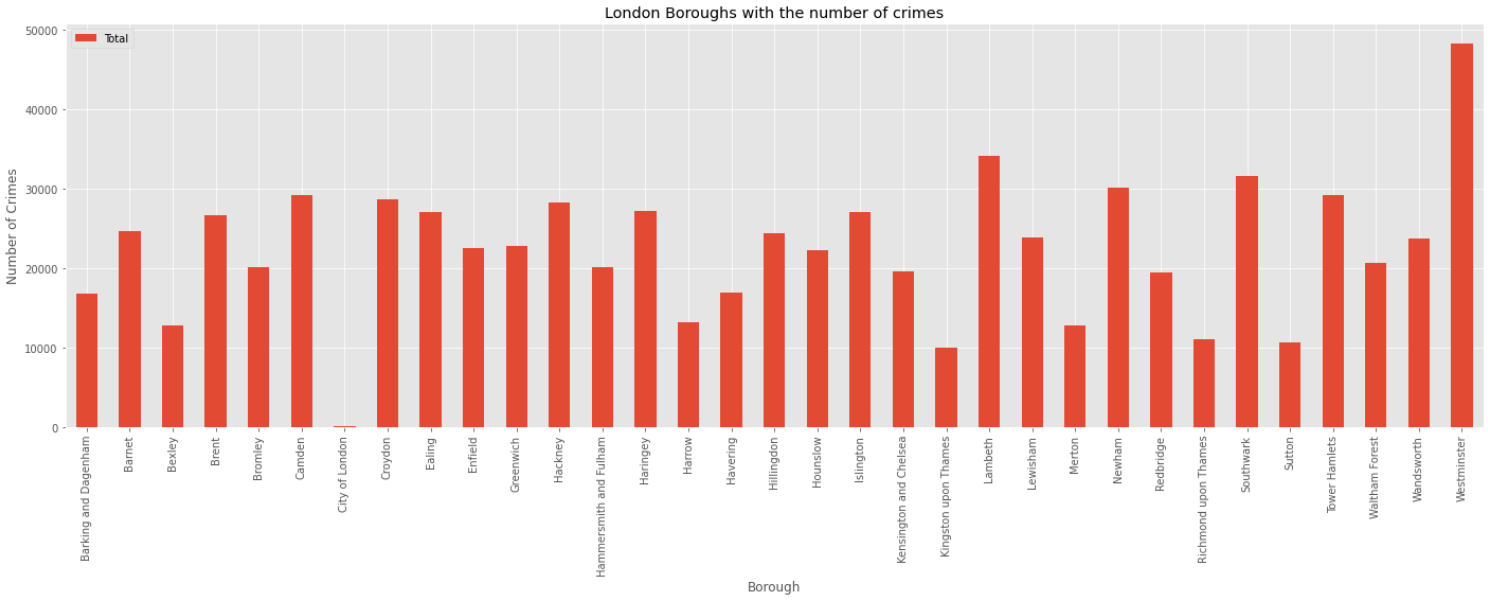

Fig. Total Crimes per Borough in London

It is evident that Westminster Borough has the maximum number of reported crimes whereas other boroughs show relatively lower crimes.

3.1.3. Boroughs with Lowest Crime Rate

The dataset is sorted in ascending order to obtain the boroughs with lower number of crimes first. The top 5 boroughs are separated and visualized using a bar graph.

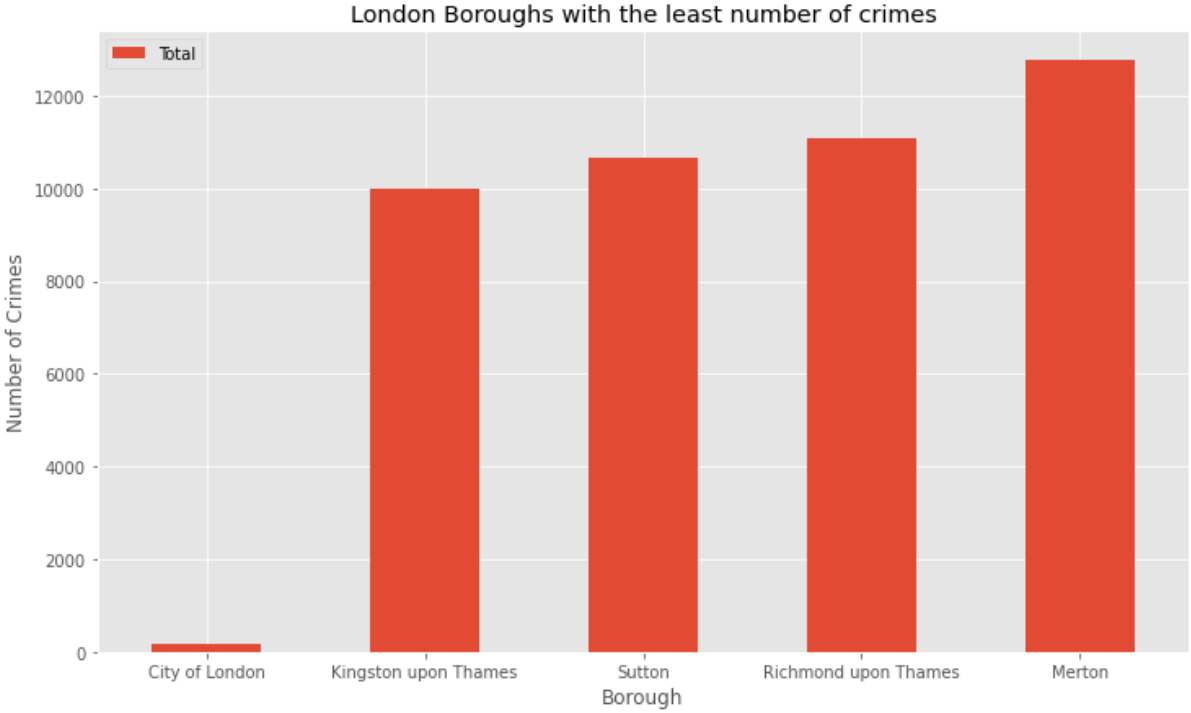

Fig. Boroughs with Least Crimes

From the graph, we can observe that the 'City of London' has extremely low number of crimes as compared to other 4 boroughs. Below is the detailed record for the 'City of London'.

	Borough	Total	Area (sq mi)	Population (2013 est)[1]
6	City of London	178	1.12	7000

Fig. City of London

The City of London has an area of 1.12 sq. miles (relatively small area), Population of 7000 people and 178 reported crimes, which is a significantly low number. As per the information available on wikipedia.org, City of London administratively forms one of the 33 local authority districts of London; however, the City of London is not a London borough.

Since, City of London is not a London Borough, we will select the next borough in the order; **Kingston upon Thames**.

3.1.4. Neighbourhoods in Kingston upon Thames

There are 15 neighbourhoods in the borough 'Kingston upon Thames'. The data is available on this Wikipedia page.

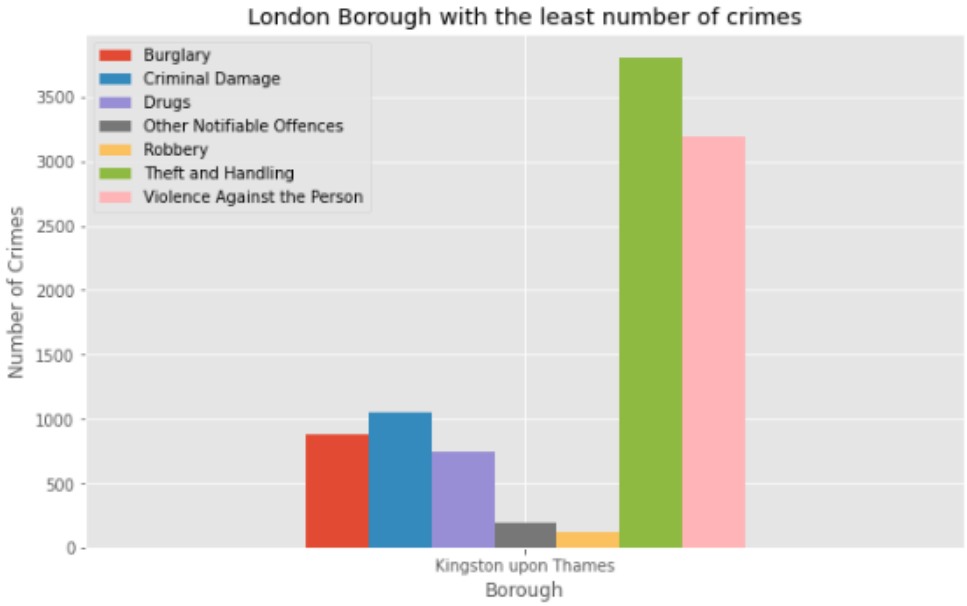

Fig. Crimes in Kingston upon Thames

The above bar graph shows us the total number of crimes for each category of crime in 'Kingston upon Thames'. 'Theft and Handling' has the largest number of reported crimes, followed by 'Violence

Against the Person'. 'Robbery' and 'Other Notifiable Offences' are the categories with least number of reported crimes.

The Google API Geocoder is used to get the latitude and longitude values for the neighbourhoods (15) in Kingston upon Thames. The data is combined to form the dataframe below.

	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
11	Old Malden	Kingston upon Thames	51.382484	-0.259090
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366
13	Surbiton	Kingston upon Thames	51.393756	-0.303310
14	Tolworth	Kingston upon Thames	51.378876	-0.282860

Fig. Neighbourhoods in Kingston upon Thames

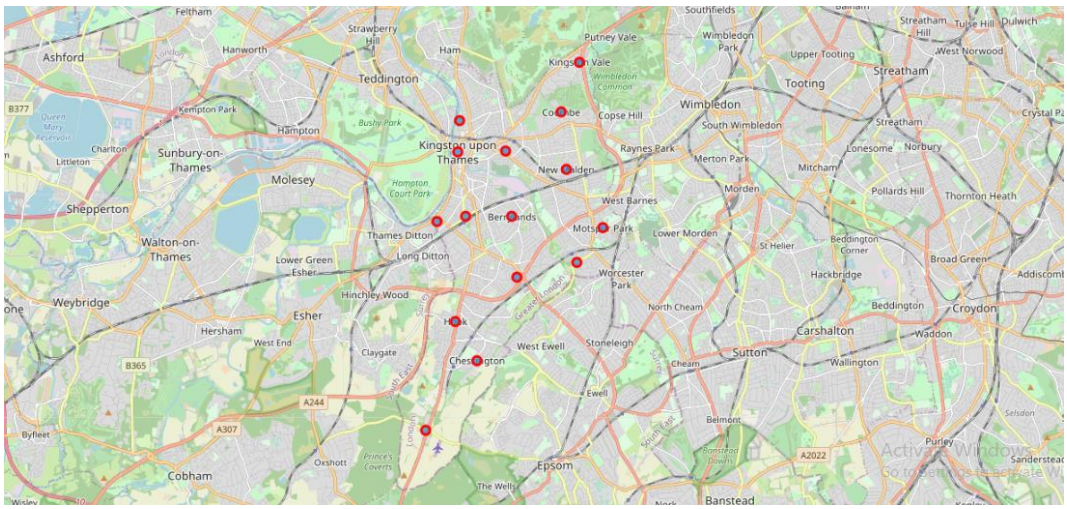

Fig. Visualizing the neighbourhoods using folium python library

3.2. Modelling

As we have the neighbourhoods along with their geographical location, we now use FourSquare API to obtain the venues for these neighbourhoods in the radius of 500 metres. The API call returns a JSON file which we convert to a pandas dataframe for processing. The venues along with their category, borough and venue's latitude and longitude data for each venue is present.

	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. Venues Data for Neighbourhoods in Kingston upon Thames

The 'Venue Category' is a categorical field. We need to convert the data in this field to numeric data so that we can process it easily later while clustering the neighbourhoods. We use a technique called One Hot Encoding for this purpose.

One Hot Encoding involves conversion of categorical variables to numeric variable values so that we our Machine Learning algorithm can process this information. The venues data is grouped by Neighbourhood and venue count for each neighbourhood is obtained. From this data, 10 most common venues are calculated for each neighbourhood.

	Neighborhood	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	Gym / Fitness Center	Park	Bus Stop	Turkish Restaurant	Fish & Chips Shop	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Food
1	Canbury	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Plaza	Café	Hotel
2	Chessington	Construction & Landscaping	Turkish Restaurant	Deli / Bodega	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food
3	Coombe	Health & Beauty Service	Turkish Restaurant	Food	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant
4	Hook	Bakery	Supermarket	Fish & Chips Shop	Indian Restaurant	Turkish Restaurant	Food	Discount Store	Dry Cleaner	Electronics Store	Farmers Market

Fig. Ten Most Common Venues for each Neighbourhood

4. Results

K – Means Clustering algorithm is used to cluster the neighbourhoods into 5 clusters (K=5). We can list the neighbourhoods in any particular cluster using the cluster label assigned to each neighbourhood.

Cluster 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	10th Most Common Venue
3	Coombe	Kingston upon Thames	51.41945	-0.265398	0	Health & Beauty Service	Wine Shop	Furniture / Home Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food	French Restaurant	Fried Chicken Joint	Garden Center

Fig. Neighbourhoods in Cluster 0

Cluster 0 has only one neighbourhood. The most common venues in this cluster are **Health & Beauty Services, Wine Shops, Furniture / Home Stores and Farmers Markets.**

Cluster 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 Most Common Venue	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Hotel	Spa	Plaza	Fish & Chips Shop	Café	Indian Restaurant	Shop & Service	Park	Supermarket
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Bakery	Fish & Chips Shop	Indian Restaurant	Supermarket	Wine Shop	Fried Chicken Joint	Farmers Market	Fast Food Restaurant	Food	French Restaurant
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Pub	Sushi Restaurant	Coffee Shop	Burger Joint	Asian Restaurant	Furniture / Home Store	Turkish Restaurant	French Restaurant	Electronics Store
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Grocery Store	Pub	Restaurant	Garden Center	Food	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gym	Supermarket	Grocery Store	Chinese Restaurant	Korean Restaurant	Gastropub	Indian Restaurant	Sushi Restaurant	Bar	German Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Food	Indian Restaurant	Italian Restaurant	Platform	Pub	Wine Shop	Rental Car Location	Fried Chicken Joint	Hardware Store	Hotel
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Pub	Coffee Shop	Harbor / Marina	Chinese Restaurant	Gym	Gym / Fitness Center	Turkish Restaurant	Hotel	Italian Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	Gym / Fitness Center	French Restaurant	Hotel	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Coffee Shop	Italian Restaurant	Train Station	Thai Restaurant	Furniture / Home Store	Pizza Place	Soccer Field	Bowling Alley

Fig. Neighbourhoods in Cluster 1

Cluster 1 has 9 neighbourhoods among the total 15 neighbourhoods. The most common venues in the neighbourhoods in this cluster are **Pubs, Cafés, Stores and Gyms.**

Cluster 2

	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	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food	French Restaurant	Wine Shop
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Soccer Field	Bus Stop	Park	Gym	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	Electronics Store

Fig. Neighbourhoods in Cluster 2

Cluster 2 has two neighbourhoods. The most common venues in this cluster are **Gym & Fitness Centers, Parks, Bus Stops and Soccer Fields.**

Cluster 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 Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	3	Construction & Landscaping	Turkish Restaurant	Deli / Bodega	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	3	Train Station	Pub	Food	Wine Shop	French Restaurant	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Fried Chicken Joint

Fig. Neighbourhoods in Cluster 3

Cluster 3 has two neighbourhoods. The most common venues in this cluster are **Construction & Landscaping, Restaurants, Train Stations and Pubs.**

Cluster 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 Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Sandwich Place	Soccer Field	Gift Shop	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Department Store

Fig. Neighbourhoods in Cluster 4

Cluster 4 has only one neighbourhood. The most common venues in this cluster are **Grocery Stores, Sandwich Places, Bar and Soccer Fields.**

Visualizing the clustered neighbourhoods on a map using folium python library.

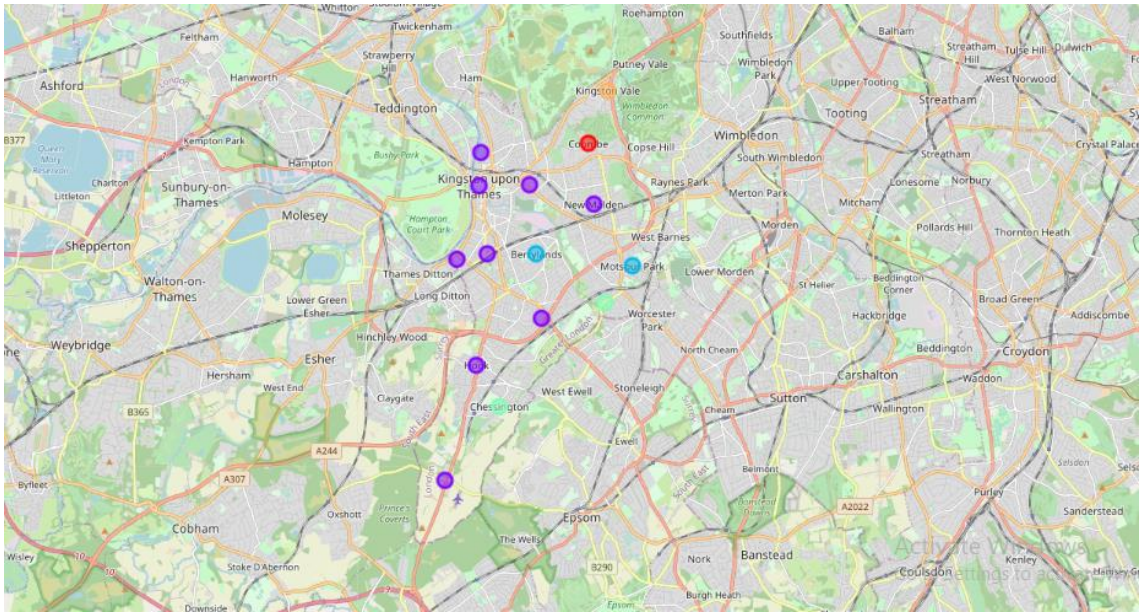

Fig. Clustered Neighbourhoods

The clusters are represented using different colours on the map. We can see one neighbourhood in Cluster 0 (Red), Clusters 1 (Purple) has maximum neighbourhoods, Cluster 2 (Blue) have two neighbourhood, Cluster 3 (Green) has two neighbourhoods while Cluster 4 (Yellow) has one neighbourhood.

5. Discussion

Any person considering to relocate to London can use the results of this project to identify the safest borough in London, Kinston upon Thames. The further clustering of the neighbourhoods is done based on the most common venues in the neighbourhoods. The person who is relocating, can select a neighbourhood from one of these clusters based on his/her preference considering the characteristics of the clusters.

For instance, a person who prefers a social lifestyle would likely relate more to a place with more Pubs, Cafés, and Restaurants. The neighbourhoods in Cluster 1 have these characteristics and therefore he/she can choose a neighbourhood from this cluster for a suitable experience.

6. Conclusion

The aim of the project is achieved by identifying the safest borough in London. The neighbourhoods in this borough are clustered into 5 total clusters. Clustering is implemented on the basis of most common venues in each neighbourhood. The Google API Geocoder and FourSquare API are used to generate geographical coordinates and venues for each neighbourhood. The characteristics of neighbourhoods in each cluster are discussed for the user to choose a neighbourhood that is best suitable to his/her needs and personality.