Applied Data Science - Coursera Capstone Project
The Battle of the Neighbourhoods

Identifying the safest borough in London and exploring its neighbourhoods.

By,

Prajwal Premkumar Khairnar

Introduction

- 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.

- 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.

Introduction

- 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.

Data Acquisition

- 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.

- London Boroughs

The specific information about the boroughs present in London. This data is scraped from a Wikipedia page

- 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.

Data Cleaning

- 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.

- 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.

- 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.

Methodology

- Exploratory Data Analysis
  - Statistic Description of London Crime Data

	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

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’.

Methodology

- Exploratory Data Analysis
  - Crimes in each Borough

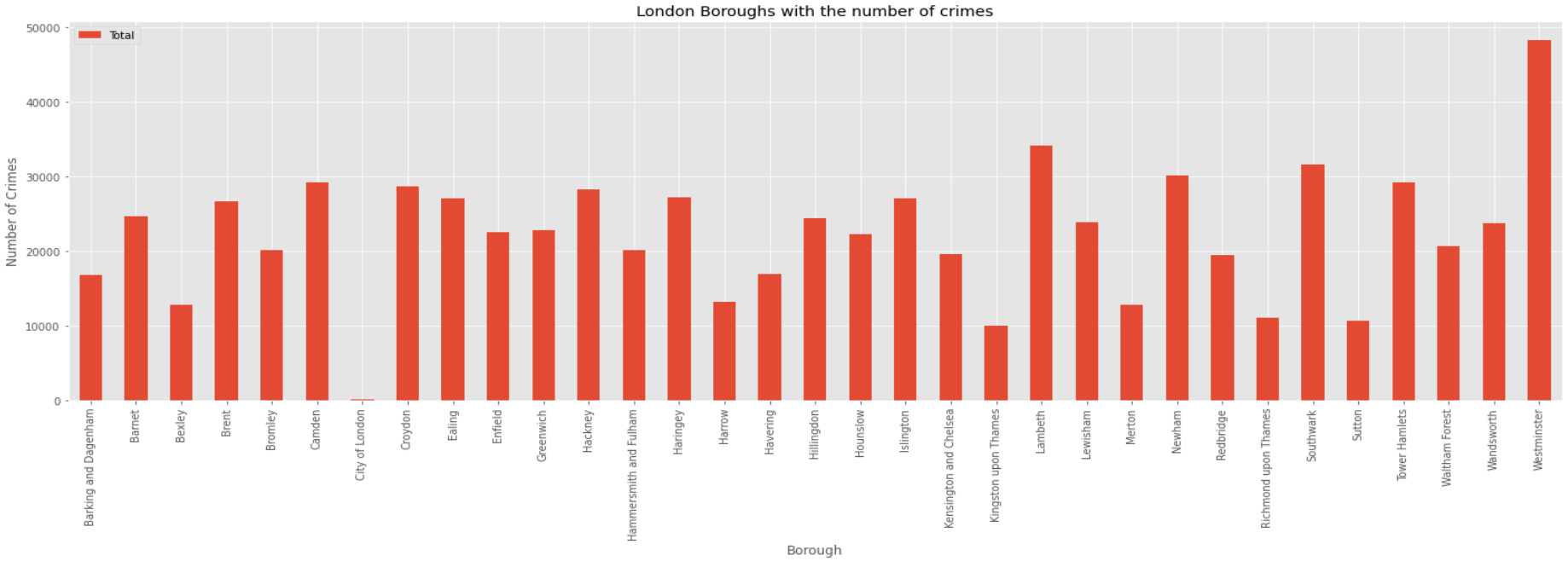

Fig. Total Crimes per Borough in London

Methodology

- Exploratory Data Analysis
  - 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.

    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'.

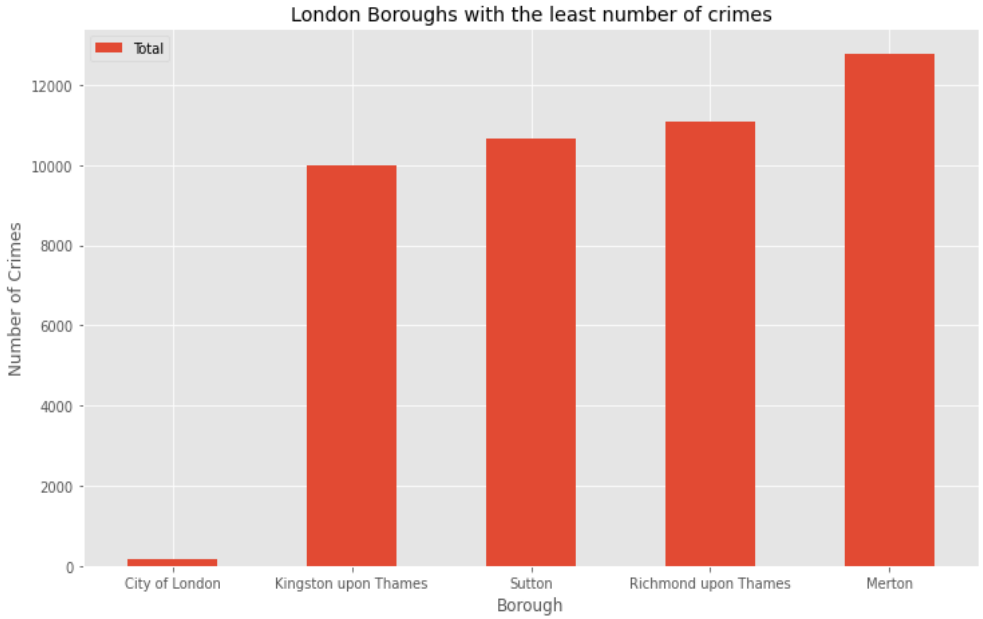

Fig. Boroughs with Least Crimes

Methodology

- Exploratory Data Analysis
  - Boroughs with Lowest Crime Rate

	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**.

Methodology

- Exploratory Data Analysis
  - Neighbourhoods in Kingston upon Thames

    There are 15 neighbourhoods in Kingston upon Thames

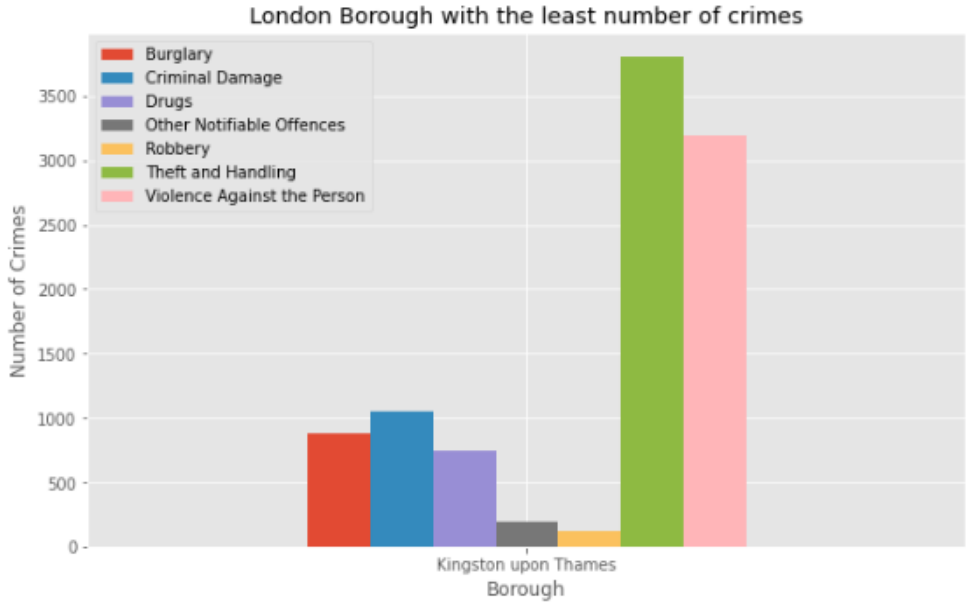

Fig. Crimes in Kingston upon Thames

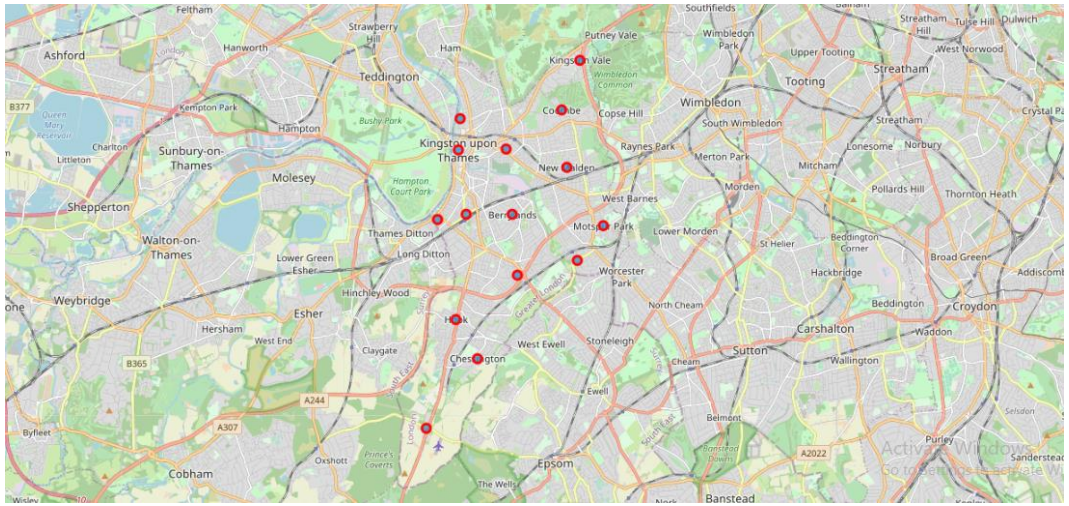

Fig. Visualizing the neighbourhoods using folium python library

Methodology

- Exploratory Data Analysis
  - Neighbourhoods in Kingston upon Thames

  The Latitude and Longitude values for each neighbourhood are obtained using Google API Geocoder.

	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

Methodology

- 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

Methodology

- Modelling

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

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.**

Results

- **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.**

Results

- **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.**

Results

- **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.**

Results

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

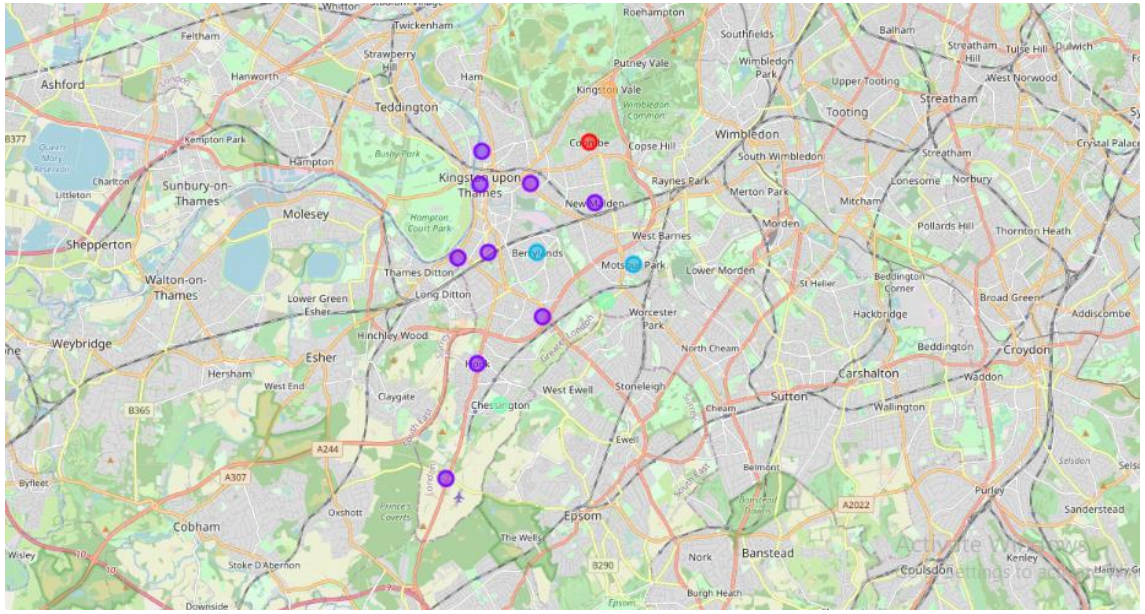

Fig. Clustered Neighbourhoods

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.

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.

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.

- Considering safety to be one of the prime concerns, this project is extremely useful to ensure that this factor is well researched while taking the decision to relocate to London. The clustering of neighbourhoods provides a personally suitable choice for all the users.