Safest and most suitable neighborhood in London

Krishnanunni B

11 June, 2020

1. Introduction.

1.1 Background:-

Lots of students, workers, bussiness men, etc are migrating to London and don't know where in London to settle. This leads to many problems within the mind of an indiviual as well as in the society in which he is living. There is very high risk for these people to fall into the traps of robbers, thugs, smugglers, etc. They may get involved in crimes or may become a victim. Several incidents have been reported previously and those people are not able to overcome these problem. Another problem which needs to be addressed is the interests of an individual about a region before migrating to that region. Without knowing what kind of place or what facilities are there in that place may cause so many misunderstanding and mental illness. These migrated people may become less productive than before or may not become productive as expected.
So if there is a system that can guide them to the correct place, it would be great and advantageous. With the help of modern technology it is possible to solve this problem.

1.2 Problem:-

Data that might contribute to determining the safest Borough in London improvement might include crime data from the police department in London. This includes crime type, date, year, may be sub category of crime, location of the crime, etc.Then location data of the safest Borough and its neighborhood, location data of all the venues in each neighborhood.
This project aims to help people to select the safest and best neighborhood in London according each individuals preference based on these data.

1.3 Interest:-

Obviously, the people who are planning to migrate in London will be very interested to know these informations about the places, venues, areas etc. This project is very helpful for those people who has not travelled to London before, who are planning to migrate and want to know these informations.

2. Data acquisition and cleaning.

2.1 Data sources:-

Data needed to tackle this problem and find a solution can be divided into three types;
a. London crime data of recent years with crime location, type of crime, date etc. It can be downloaded from kaggle. Using this data we can find the Borough in London with least crimes.
b. Location data of neighborhoods. It can be obtained using geopy client of python. It is used to find the Location of each neighborhood in their respective Borough.
c. Foursquare venues data of each neighborhood. It can be obtained using Foursquare API. It is used find the most visited venues which is later used for applying Machine Learning algorithms.

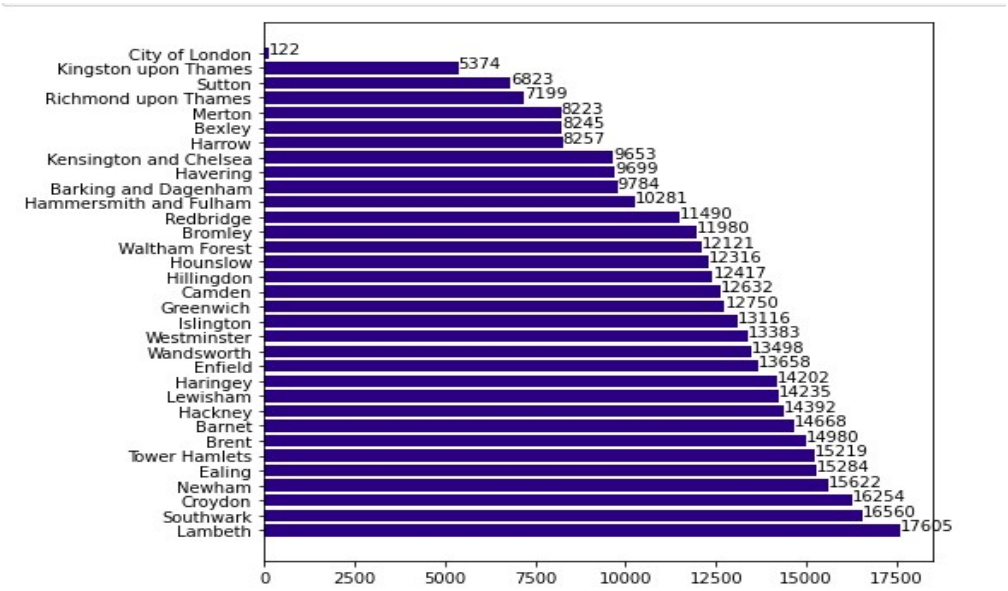
2.2 Data cleaning:-

The data downloaded from kaggle ie Crime data of London data was cleaned beforehand itself by the person who has uploaded it in the kaggle platform. So, the job of cleaning that huge data set was saved. I took only the recent year’s data of crimes ie 2016 which had number of cases per month greater than zero.
Location data of neighborhoods can be obtained using geopy client of python. It is used to find the Location of each neighborhood in their respective Borough. The raw data was appended to lists and then converted to pandas data frame. Foursquare venues data was retrieved in json format and was cleaned using one hot encoding technique, etc.

****Feature Selection is avoided here because this project uses an unsupervised machine learning algorithm.**

3. Exploratory Data Analysis

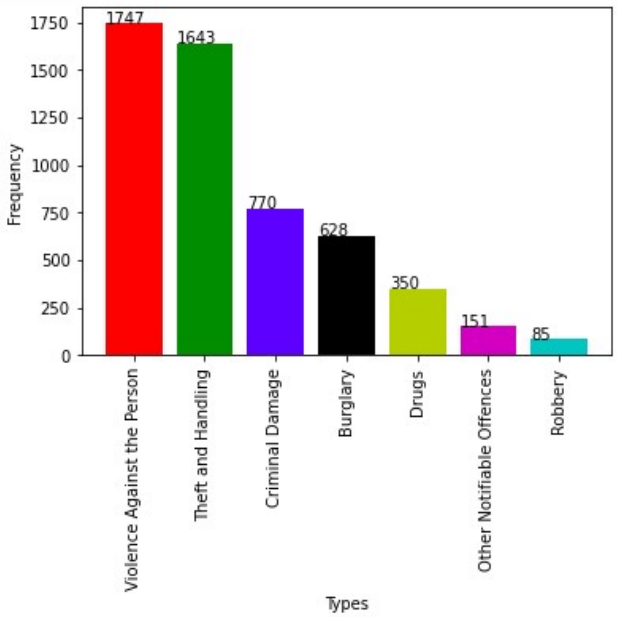
Different types of exploratory data analysis was performed on the crime dataset.
First the top Borough in London which has highest number of crimes was found. Then summary statistics was applied on the data. Then found the most occuring crime in London.
Then took the most recent year's data ie 2016. Discarded the rows that has numer of crimes in a month value 0. Then summary statistics was applied to the new data.
Arranged the Boroughs with respect to their total number of crimes and plotted a graph,



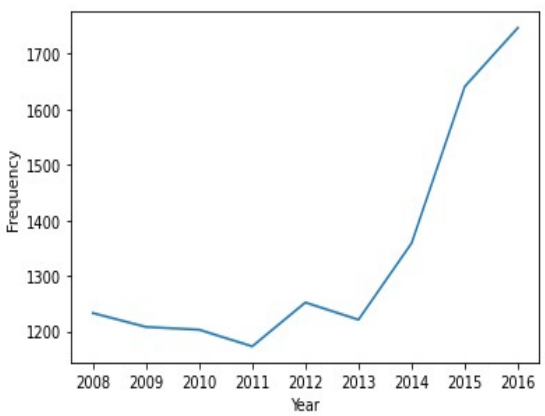
From the above figure, it is clearly understood that the *City of London* is the safest Borough in London. *Kingston upon Thames* can also be considered as safe.The chart also shows that places like *Lambeth*, *Southwark* and *Croydon* are unsafe comparing to other Boroughs.

City of London is the safest place in London according to the crime data of the year 2016.But according to wikipedia data, City of London is not a London Borough(for reference follow this).So the next safest Borough which we can explore is *Kingston upon Thames*

Next another graph was plotted to find out which category of crime is most prominent in Borough which we have found through analysis.



From the above chart it is clearly understood that v*iolence against the person* is the most prominent crime occuring in Kingston upon Thames, followed by *theft and handling*.



After this analysis we plotted the trend of the most prominent category of crime from the year 2008 to 2016.

From the above plot it was understood that unntil 2011 the graph tend to go downwards. But in 2011 there is an upward movement in the graph. According several sources in the internet and my knowledge it is because of the riot in London which took place in august 2011.
But from 2013 there is a rapid increae in the number of crimes. My be it is due to the increase in number of gangs in London.

So, we found that Kingston upon Thames is the safest Borough.
The **first phase** of the project was understanding, preparing and cleaning of the huge datasets. In this project we only considered the most recent and relevant data from the crime dataset.Then we applied small statistics upon the data to find the overall behaviour and structure of data.
The **second phase** of the project was to find the safest Borough in London. In the analysis we found out that City of London is the safest one followed by Kingston upon Thames. We also found out that Westmister and Lambeth are some of the places which has very high crime numbers. We couldn't explore City of London as it is not considered as a Borough according to wikipedia data. So, we explored the Borough, Kingston upon Thames. There we found out that theft and handling is the most prominent crime occuring in Kingston upon Thames, followed by violence against the person. We clarified it through a bar chart.

Next, our task was to find the best neighborhood in **Kingston upon Thames** by comparing the amenities and facilities in these neighorhoods.

Method
Following are the steps to find the neighorhoods and cluster them into different clusters:
1. Retrieving location data of the respective neighborhoods.

2. Collecting details of all the venue in these neighborhoods.

3. Finding the top 10 venues of these neighborhoods.

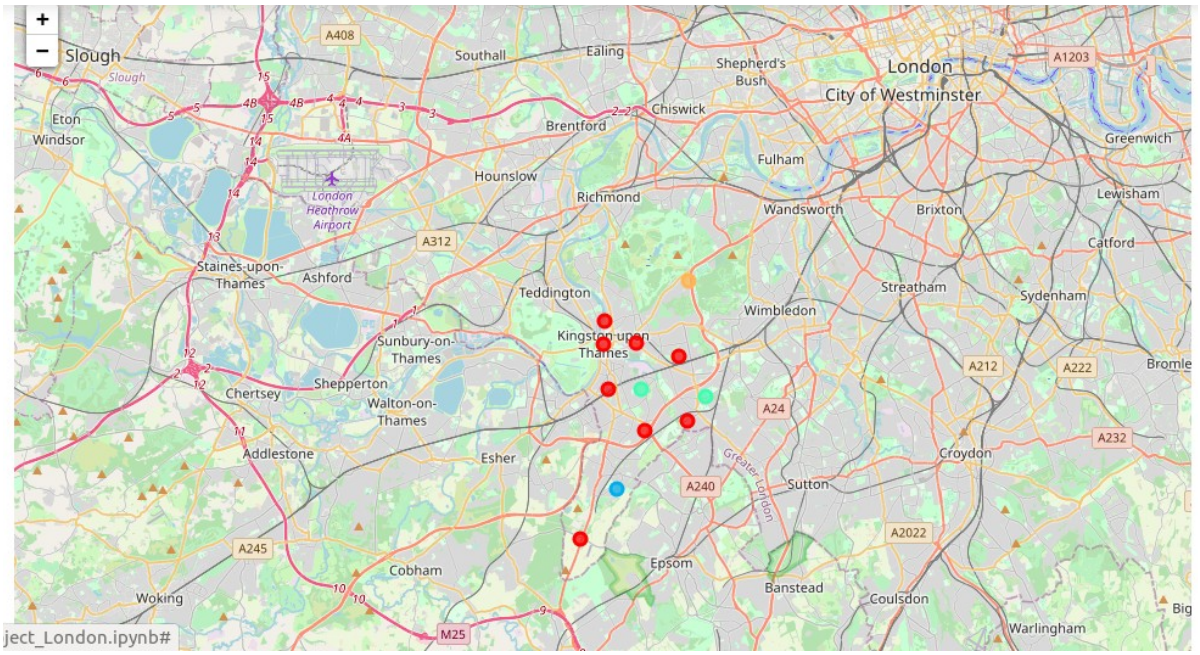
4. Clustering them using KMeans algorithm.

5. Selecting the best neighborhood according to individual’s interests.

4. ML -model

The **third phase** of the project was to find the different neighborhoods in the borough, Kingston upon Thames, retrieve their respective venues, find the top 10 venues of each of them and to cluster them according to similar properties.We retrieved the neighborhood data using geopy package, a geocoding library. Venue data was retrieved by using Foursquare API. And clustering ML algorithm that is used here is KMeans algorithm.

For this project an unsupervised machine learning model is used. The reason behind it is that there are no labels in the data and we want to cluster similar neighborhoods . So Kmeans algorithhm is used.
After aplying that algorithm, we get a map as shown below,



5. Results and Disussons

The result is that the neighborhoods are clustered considerng their top 10 venues. There are fve clusters in total and each have different charateristics. Now we an look and discuss the results and outomes of this project. Let’s see how it found a solution to the problem and how the solution is structured.

Solution to the problem.

Thus the problem addrressed well and an optimal solution has been found out.

Following are the 5 clusters which would help people select neighborhood according to their interests and way of living.

1. The first cluster consist of neighborhoods where there are many varieties of restaurants, pubs, clubs, etc and has all the amenities. This cluster also has the most number of neighborhoods with 8 neighborhoods.
2. The second cluster consist of a single neighborhood and it is the only neighborhood with beach as frequently visisted place.
3. The third cluster consist of construction & Landscaping, Wine Shops etc. It consist of a single neighborhood.
4. The fourth cluster consist of 2 neighborhoods with parks, gym, wine shops, etc being frequently visited venues.
5. The fifth and final cluster consist of a single neighborhood with grocery store, bar, soccer field, etc being the most visited venues.
   And, I personally prefer the final cluster (fifth) ie Kingston Vale .

These are the 5 clusters or groups of similar neighborhood and an individual can select easily among these neighborhood according to their own interests.

6. Conclusion

The project was about finding the best neighorhood in London with least crime number and with different types of amenities and facilities. It was able to achieve its goal as it has found the safest Borough in London and clustered the best neighborhood with similar characteristics together. However, the project only deals with the above mentioned criterias and does not consider the value of the property, taxes and laws, employability, etc in that particular neighorhood . For that another project these criterias and parameters can be considered in future.
