THE BATTLE OF NEIGHBORHOODS

PROJECT REPORT

Capstone Project provided by IBM on COURSERA

BY:

SREERAG RAJEEVAN

CONTENTS

1. INTRODUCTION:

Human migration is the movement of people form one place to another with the intention of settling, permanently or temporarily at a new location. The movement is often over long distances and from one country to another, but internal migration is also possible. People may migrate as individuals, in family units or in large groups. A person who moves from their home due to forced displacement (such as natural disaster or civil disturbance) may be described as a displaced person or, if remaining in the home country, an internally displaced person. A person who is seeking refuge in another country can, if the reason for leaving the home country is political, religious, or another form of persecution, make a formal application to that country where refuge is sought and is then usually described as an asylum seeker. If this application is successful this person's legal status becomes that of a refugee.

Having talked about what migration is, who is considered a refuge, etc., the main focus of this project is to help people shifting to London from other countries or are planning on shifting neighborhoods within London. Refugees moving into a country and to help them would be another project all-together.

London is the capital and the largest city in England and the United Kingdom. It's considered to be the world's most important global cities and has been called the world's most powerful, desirable, influential, visited, expensive, sustainable, investment-friendly, and popular-for-work city.

Taking all these factors into consideration, moving into the city is rather a herculean task. Along with all the wonderful things listed above, the city of London also tops in crime rates in the UK. We will analyze different neighborhoods in the city based on based on their safety. Along with that, we will also be looking at the accessibility of the neighborhood to bus stations, parks, restaurants, supermarkets, pharmacies, theatre, hospital, airport, etc.

This project will work as a good guide for expats planning on shifting to London, as well as people shifting neighborhoods within London, to know about the neighborhood they are shifting to.

2. EXTRACTING AND CLEANING DATA:

Multiple sources are available on the internet showing crime rates in the city of London. We will be analyzing crime rates with the help of data available on Kaggle.
https://www.kaggle.com/jboysen/london-crime
This dataset provides us with column-wise information on type of crime, when was the crime committed, how frequently crimes are committed in a particular area, year and month of the crime committed and most importantly categorization of crime based on its impact. We will get to cleaning the dataset in a while.

Wikipedia offers information regarding anything and anyone. We will be extracting some additional information regarding London city from:
https://en.wikipedia.org/wiki/List_of_London_boroughs

This dataset provides us with in-depth details regarding the co-ordinates of the boroughs, the area(in square miles), population, political party that controls the borough, etc.

We will also extract the latitudes and longitudes of certain neighborhoods within boroughs from multiple Wikipedia pages. This dataset will be built from scratch within the Jupyter Notebook using python codes.

In-built Python Libraries will be used to clean the dataset that were extracted. Data extracted from the internet will have many unwanted values and characters. Cleaning the dataset is as important as plotting a graph. A good clean dataset gives excellent results. Python libraries such as Pandas, Beautiful-Soup, Requests, LXML, Folium, Scikit-Learn, Geocoder and Matplotlib makes cleaning, plotting, and analyzing easy. Apart from these amazing libraries available on python, we will be using Foursquare API as the primary data gathering source for venues around a given borough. By providing latitude and longitude values into the Foursquare API any person can obtain details regarding most common venues around the given location.

As mentioned earlier, extracting information from data obtained from the internet is a vital process after which the most important step is cleaning of the dataset. Data obtained from three of the sources mentioned above is cleaned separately. From the dataset obtained from Kaggle, we will only take the latest available values, i.e., values of the year 2016. This is shown in Fig 1.

	Borough							No_of_Crimes	Total
Major_Category		**Burglary**	**Criminal Damage**	**Drugs**	**Other Notifiable Offences**	**Robbery**	**Theft and Handling**	**Violence Against the Person**	
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
5	Camden	2652	1935	1493	490	899	14088	7626	29183
6	City of London	2	2	10	6	4	129	25	178
7	Croydon	2738	3219	1367	718	1139	9229	10302	28712
8	Ealing	2492	2562	1355	613	669	10040	9396	27127
9	Enfield	2541	2136	1063	492	807	8037	7409	22485

Fig 1: London Crime data after initial cleaning

Our next source is a Wikipedia page. This data is extracted using Beautiful-Soup library available on Python. Using this library, the tabular data is obtained from the website, using its URL. After extraction, a couple of cleaning processes must be done in order to use the dataset further, mainly because we have to merge both this table and the previously created table. A glimpse of this dataset is shown below, Fig 2.

	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	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: Details of each Borough in London

Next step, we merge the two datasets. This gives us a better understanding of the Boroughs, also, this will help us in identifying the Boroughs with the greatest number of crimes committed and the ones with least number of crimes committed. Shown in Fig 3 below.

	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	North London Business Park, Oakleigh Road South	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 3: Merged dataset

We continue with our analysis. After initial analysis we identify Kingston upon Thames to be the safest Borough. How we arrived at this be explained in detail in the methodology section. Here, we will be creating another dataset which contains the latitude and longitude values of neighborhoods within Kingston upon Thames. For this dataset, we will be visiting a Wikipedia page. Unlike the previous Wikipedia page – where we had a table available, and we just had to extract it – here there are many other values available which is of no significance to our project. Therefore, we will be creating a table using python codes, and later we will extract the required latitude and longitude values, which will be added into respective columns of the table. The initial table that we created will look as shown below in Fig 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		
5	Kingston upon Thames	Kingston upon Thames		
6	Kingston Vale	Kingston upon Thames		
7	Malden Rushett	Kingston upon Thames		
8	Motspur Park	Kingston upon Thames		
9	New Malden	Kingston upon Thames		
10	Norbiton	Kingston upon Thames		
11	Old Malden	Kingston upon Thames		
12	Seething Wells	Kingston upon Thames		
13	Surbiton	Kingston upon Thames		
14	Tolworth	Kingston upon Thames		

Fig 4: Neighborhoods in Kingston upon Thames – Initial

We extract the latitude and longitude values into python using Geocoder library in python. After obtaining the values, we add them to the respective columns to complete our table. The final table is shown in Fig 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
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 5: Neighborhoods in Kingston upon Thames – Final

In the methodolgy section we will be exploring the initial dataset in detail and we will be arriving at the conclusion as to which Borough is the safest in London. Using the table in Fig 5, we will explore each of the areas in detail to identify most common venues around the place using Foursquare API. Then we will do our finial analysis using K-Means clustering algorithm, which will help us in identifying similar neighborhoods.

3.<u>METHODOLOGY:</u>

After we extract and clean our data, the next step is to start with various analysis methods, to obtain the desired end result. The target audience for this project is anyone (that could be a person, a family, group of friends shiftting in together, etc.) planning to move in to London from elsewhere or shift neighborhoods with-in London. Now, these people would not understand heavy coding, so, the main objective will be to keep this as simple as possible.

We have the table that depicts the type of crime committed in various boroughs in London. The table in Fig 3 contains all the necessary details. But it would be hectic process if we were to go through each row one-by-one (imagine the time it would take!!). In python, we have an in-built code - Describe – which shows us statistical description of each column. This is shown in Fig 6.

	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

Fig 6: Statistical Distribution of various crimes

As we can see, the table gives us a complete idea as to which crime has been committed the most number of times, least number of times, the standard deviation, etc. Also from the table is clear that Theft and Handling is the crime that has been committed most number of times, across all boroughs.

Our next step is segregate the table an find out the top five areas with most crimes committed. The areas are, Westminster, Lambeth, Southwark, Newham and Tower Hamlets. We plot these values onto a graph. Fig 7 depicts the graph we created.

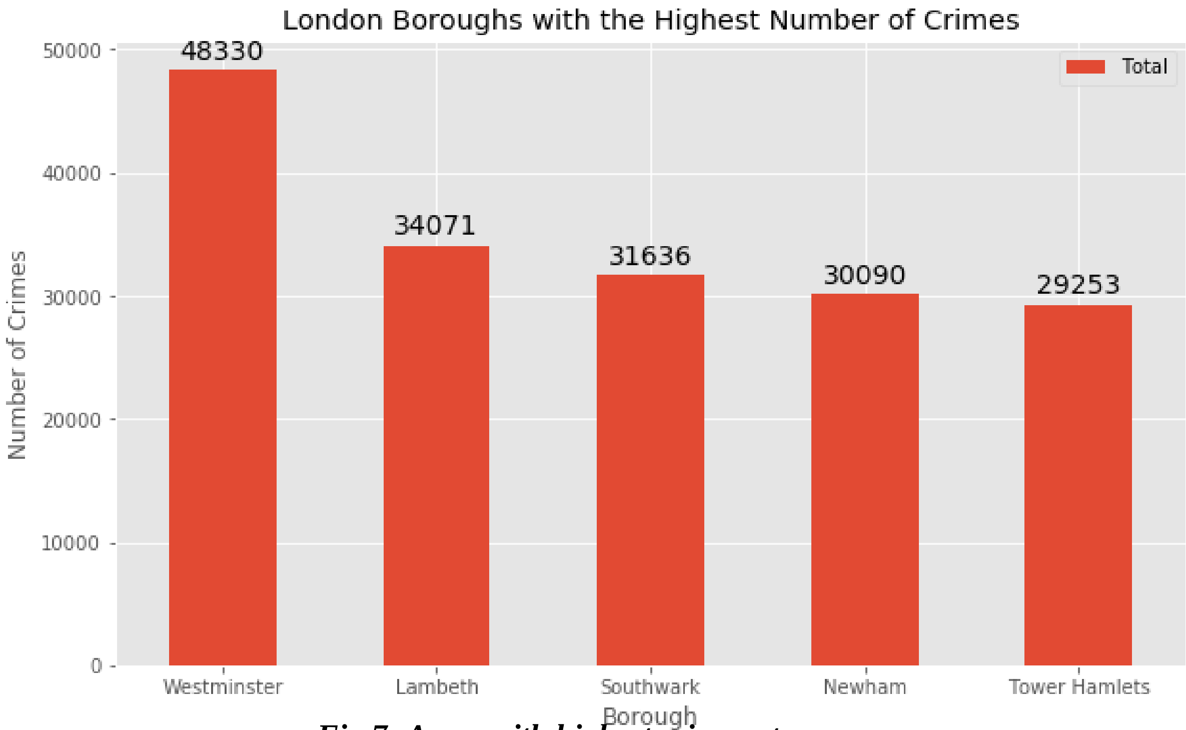

Fig 7: Areas with highest crime rates

From the graph we can have a basic idea as to which are the places that can be avoided.

Since our aim is to find the safest borough, we have to analyze the top five boroughs with lowest number of crimes committed. Kingston upon Thames, Sutton, Richmond upon Thames, Merton and Bexley are the areas that have recorded lowest crime rates. We put these values onto a graph to have better understanding. Fig 8 depicts the graph.

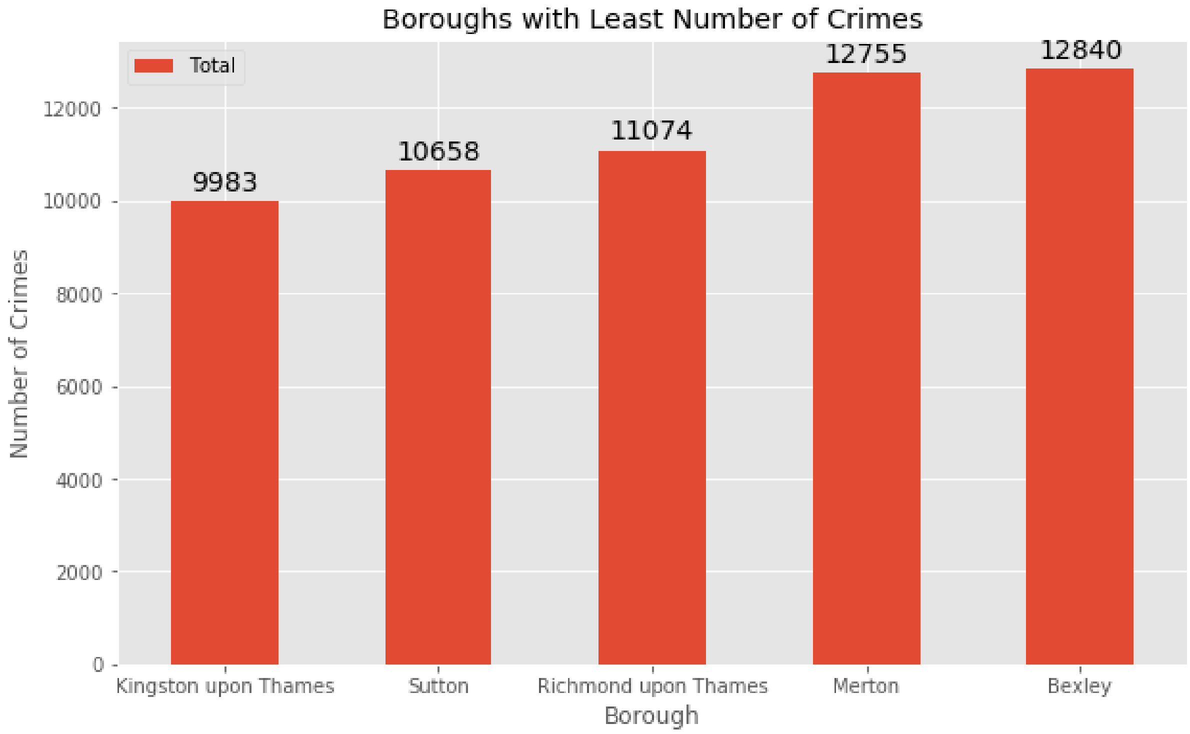

Fig 8: Areas with lowest crime rates

We can see that Kingston upon Thames is the borough with lowest crime rates. This is the borough on which we have to conduct detailed analysis. The figure below (Fig 9), shows the graph depicting the type of crime committed in Kingston upon Thames.

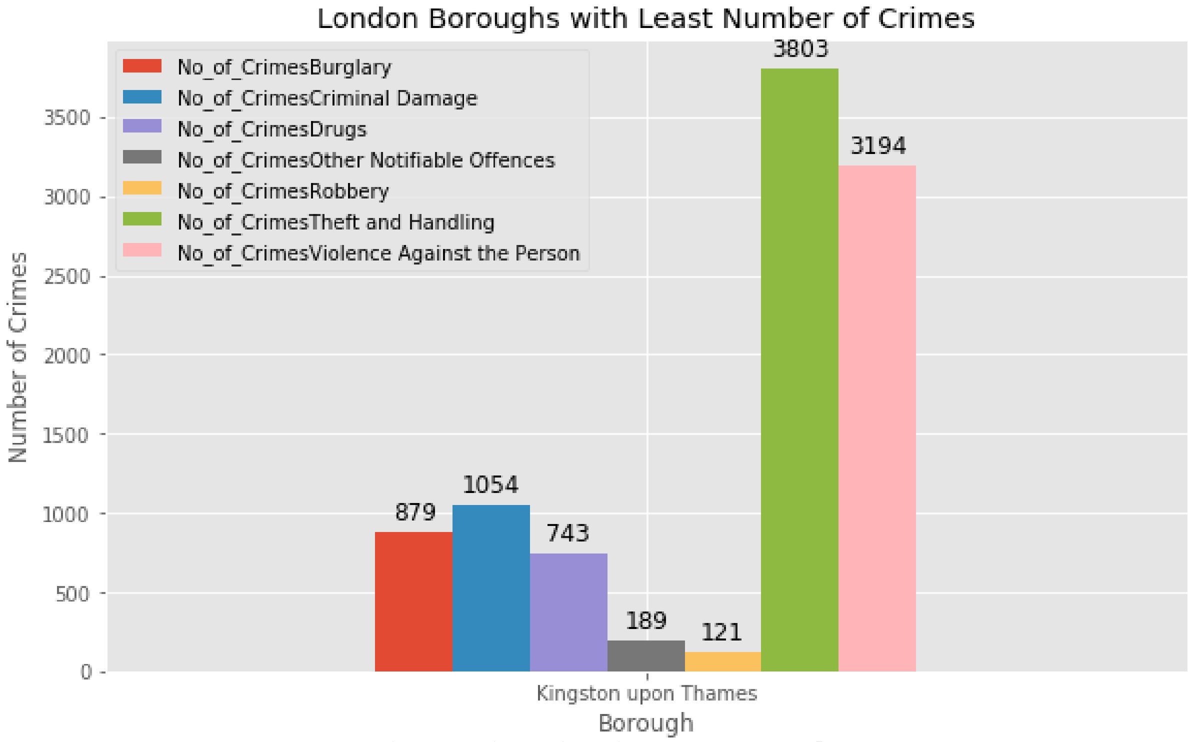

Fig 9: Crimes in Kingston upon Thames

From the table that we created showing neighborhoods within Kingston upon Thames, we see that there is total of fifteen neighborhoods in the area. We plot the map of the area to have a better understanding. Fig 10 show the map.

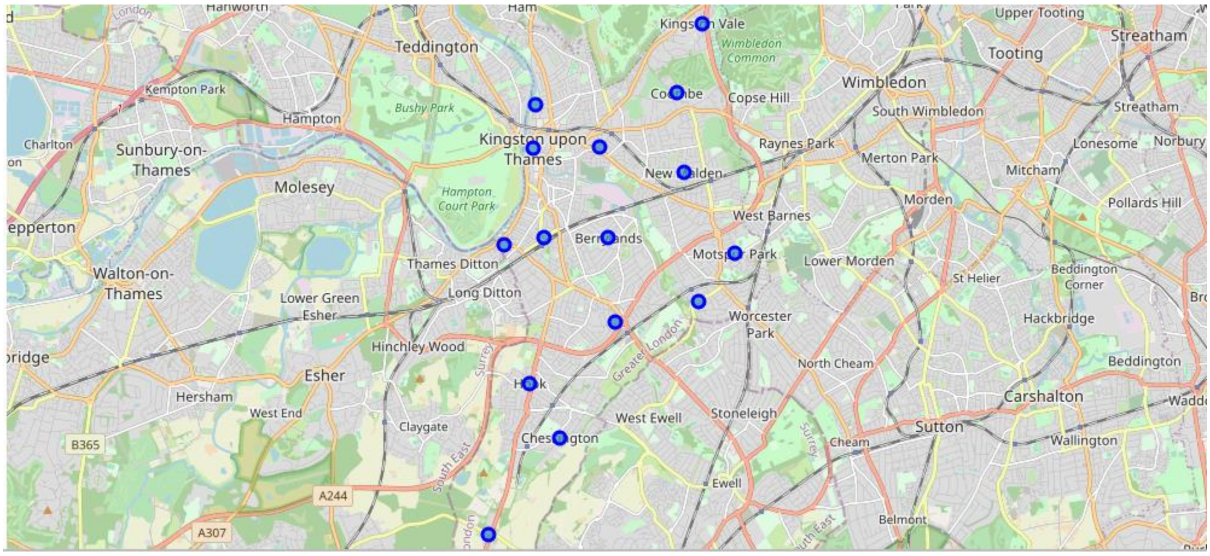

Fig 10: Neighborhoods in Kingston upon Thames

With the help of Foursquare API and our dataset containing the co-ordinates of the neighborhoods in Kingston upon Thames, we can find out most common venues within five hundred meters of each neighborhood. This will help us in identifying the best neighborhood in the borough. Foursquare API will return a JSON file containing all the information regarding the venues along with their co-ordinates. Once again, we will have to extract the data and create a dataset out of it for easier analysis. A glimpse of this dataset is shown below in Fig 11.

	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	Berrylands	51.393781	-0.284802	SK Superstores	51.389901	-0.283278	Convenience Store
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

Fig 11: Venues near each neighborhood (first five readings)

Machine Learning algorithms are the best suited for predicting the outcomes of an event, showing us what would be the best place to invest, etc. Here, in this case, we will use these algorithms to predict which neighborhood/neighborhoods would be the best suited to move in for a new person or family. But, before we start with Machine Learning algorithms, we will to Onehot encoding to our data. Onehot encoding is a process in which any categorical data is converted into a form which would help the Machine Learning algorithm to come up with better predictions. The dataset will be grouped by the neighborhood, and mean of the venues will be calculated, finally we will get the ten most common venues in each of the neighborhoods.

K-Means is a type of unsupervised Machine Learning algorithm. It is a type of partitioning clustering, which divides the data into ‘K’ non-overlapping subsets, or clusters, without any cluster internal structure or labels. The reason to use this method in our analysis is because it divides the dataset into groups depending upon similarities among them. This makes our job easy, to provide a detailed explanation as to why a particular neighborhood was selected and certain neighborhood excluded. We have fifteen neighborhoods. Therefore, we will use five clusters (K = 5). This will help use in obtaining better results.

4.RESULTS:

K-Means clustering groups neighborhoods based on the similarities. Once K-Means has been executed, we can view each cluster to explore the results. Fig 12 shows us cluster ONE.

	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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

Fig 12: Cluster ONE

This cluster of neighborhoods contain Train Stations, Restaurants, Supermarkets, Pharmacies, Cafes, Gyms, etc. Looks like an ideal place for a family to move into. Fig 13 shows cluster TWO.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
3	Coombe	Kingston upon Thames	51.41945	-0.265398	1	Health & Beauty Service	Wine Shop	Food	Department Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

Fig 13: Cluster TWO

Fig 14 shows cluster THREE.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Grocery Store	Food	Construction & Landscaping	Train Station	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint

Fig 14: Cluster THREE

Fig 15 shows cluster FIVE.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant

Fig 15: Cluster FIVE

Figures 13, 14 and 15 shows cluster two, three and five. Form the images we can see that even though we have some essential items in each cluster, these neighborhoods would not be advisable for families to move into. Fig 16 shows cluster FOUR.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1 Most Common Venue	2 Most Common Venue	3 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Food	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Restaurant	Fast Food Restaurant	Department Store	Dry Cleaner	Electronics Store

Fig 16: Cluster FOUR

This cluster, compared to cluster two, three and five, shows better results. Considering better connectivity, supermarkets, and other venues it is clear that cluster ONE is the ideal cluster. Fig 16 shows the final cluster map of Kingston upon Thames.

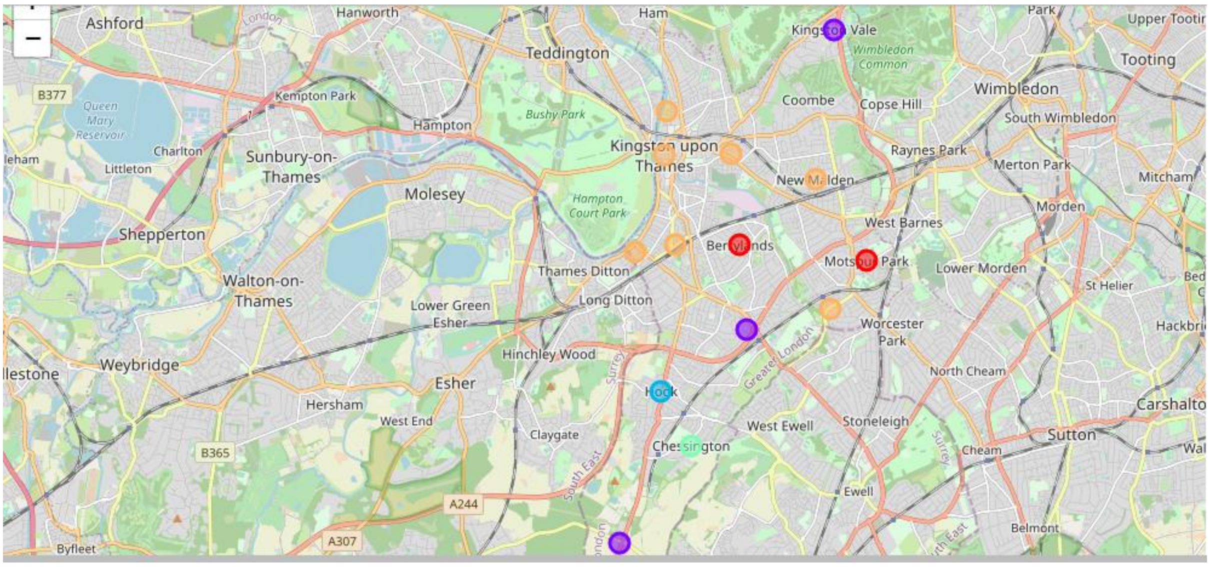

Fig 17: Kingston upon Thames – Final Cluster

From the figure above, we can see that K-Means clustering has clustered the neighborhoods in the borough by looking into their similarities. They are color coded to show similar clusters. Orange color depicts the cluster we selected, i.e., cluster ONE.

5. DISCUSSION:

The aim of this project is to help people to shifting to London. They would have a clear understanding about the place they will be calling their home. Taking safety as the main aspect, we have done analysis on various boroughs in London to finally identify one best option. After obtaining the best result, we analyzed that particular borough in further detail. Common venues near each neighborhood was looked into, so that our analysis becomes easier, and moreover, any person wanting to know the safest borough could do so with ease. For a family the best suited option would be cluster ONE, which contains neighborhoods like Canbury, Hooke, Kingston upon Thames, Malden Rushett, New Malden, Norbiton, Seething Wells, Surbiton, and Tolworth.

6. CONCLUSION:

The project helps people shifting to London, to identify different features of the area they will shift to. Technological advancements are for us to use it to stay one step ahead at all times. Using various tools and API's we analyzed different boroughs in London, showing people the safety of each borough. This was taken a step further and the most ideal borough was explored in detail. The basis of the project has been safety of people moving in. Further advancements can be done on the project by taking other factors into consideration. For example, a person moving in would definitely want to know the cost of living in the area, the rent he/she would have to pay, etc. This could be ideal next step in our exploration.