The Battle of Neighborhoods

Best Neighborhoods in London Borough

Author: Mohammed Abdulla T

Date: February 2020

Contents

1 Introduction

1.1 Background

Moving when you have a family can be daunting, especially when you're moving to a city as diverse as London is. Consider to set up new home in London, even though it was exciting, it was certainly not an easy task. Upon some research, I could find that the slump in London's housing market may be about to end. And this is the right time to invest in a property. Before making an investment, it is necessary to have a study about London neighbourhood.

1.2 Problem

Data that might contribute to determining the best neighbourhood include how safe the neighbourhood is, affordability, accessibility and facilities like parks, restaurants etc. This project aims to predict the best neighbourhood based on the above factors.

1.3 Interest

Obviously, expats who are considering to relocate to London as part of career change could utilise this report for handing a safe borough and district in London for buying/renting house. People who are currently in London, but has not invested in a house yet due to the current political/economical scenario, could also use this report.

2 Data acquisition and cleaning

2.1 Data sources

- The first step in the analysis was to find all the boroughs in London. This data could be obtained from a wiki page, List of London Boroughs.
- The latest crime data records could not be obtained for the boroughs. However, the data pertaining to year 2016, is there as a kaggle dataset, Kaggle dataset for London Crime data .
- Next, we can check the affordability of the houses using a dataset from Office for National Statistics. ONS Dataset for Houseprice.
- During the research, I also came across a survey by itv, which records the happiness index of the people in uk. The link for the report is Happiest Boroughs revealed by ITV. The data from this web page can be scrapped and best boroughs can be found out.

- Finally find the district in Sitton which is the best borough buying a house, we scrap the wiki page.

2.2 Data Cleaning

To find the London borough names, we use BeautifulSoup to scrap the wikipage. After that, string manipulation is done using regular expression and the exact name of the boroughs are extracted. Extracting the exact names is very important, as through out the project, we will be using the Borough names as the index for the data frames.

Out[4]:

	London_Borough
0	Barking and Dagenham
1	Barnet
2	Bexley
3	Brent
4	Bromley
5	Camden

Figure 1: List of London Borough

For the crime rate analysis, the data set from Kaggle is downloaded. The dataset consist of:

- 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

The csv file is read using pandas. Only the data from the year 2016 is extracted. If the 'value' column is 0, then is filtered out. Finally the dataset can be grouped together on borough name to get the count/crime rate.

Out[9]:

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01001116	Croydon	Burglary	Burglary in Other Buildings	0	2016	11
1	E01001646	Greenwich	Violence Against the Person	Other violence	0	2016	11
3	E01003774	Redbridge	Burglary	Burglary in Other Buildings	0	2016	3
9	E01004177	Sutton	Theft and Handling	Theft/Taking of Pedal Cycle	1	2016	8
14	E01002398	Hillingdon	Theft and Handling	Theft/Taking Of Motor Vehicle	0	2016	2
18	E01002945	Kingston upon Thames	Theft and Handling	Theft From Shops	0	2016	11
29	E01003325	Lewisham	Violence Against the Person	Common Assault	0	2016	2
32	E01000733	Bromley	Criminal Damage	Criminal Damage To Motor Vehicle	1	2016	4
36	E01002006	Haringey	Criminal Damage	Criminal Damage To Other Building	0	2016	12
64	E01003989	Southwark	Theft and Handling	Theft From Shops	4	2016	8

Figure 2: London Crime Data for the year 2016

For the analysis of price per sq.m of houses in the boroughs, the dataset from Office of National Statistics is extracted. This contains:

- Local authority code
- Local authority name
- Year
- Price per m2

Only local authority name which are there in the Borough list extracted from the wiki page is filtered and taken into a dataset. Next, the local authority code and year can be dropped. This will give the dataset with the borough name and price/m2. The dataset when sorted in the ascending order will give the most affordable areas.

Out[11]:

	local authority code	local authority name	year	price per m2
0	E06000001	Hartlepool	2016	987
1	E06000002	Middlesbrough	2016	1120
2	E06000003	Redcar and Cleveland	2016	1182
3	E06000004	Stockton-on-Tees	2016	1254
4	E06000005	Darlington	2016	1260
5	E06000006	Halton	2016	1339
6	E06000007	Warrington	2016	1750
7	E06000008	Blackburn with Darwen	2016	1053
8	E06000009	Blackpool	2016	1107
9	E06000010	Kingston upon Hull, City of	2016	1127
10	E06000011	East Riding of Yorkshire	2016	1588

Figure 3: Comparison of House price per m2 in London(2016)

From the analysis of the above three data source, we conclude the best borough for real estate in London to be Sutton. Next we need to build a dataset with the neighborhoods in Sutton. For that, the wiki page London Borough of Sutton #Districts is scrapped and the district names are obtained. Borough name will be Sutton. The geographical cordinated are obtained using geopy client.

Out[29]:

	District	Borough	Latitude	Longitude
0	Bandon Hill	Sutton	51.364777	-0.134833
1	Beddington	Sutton	51.371988	-0.132393
2	Beddington Corner	Sutton	51.386942	-0.149532
3	Belmont	Sutton	51.343785	-0.201152
4	Benhilton	Sutton	51.371642	-0.191571
5	Carshalton	Sutton	51.365788	-0.161086

Figure 4: Geographical Coordinates of Sutton Disticts

Finally Using Foursquare Location Data, the 100 most popular venues in a radius of 500m for each district is Sutton is obtained. The data obtained is a JSON _le, and we need to turn that into a data-frame. This final dataset will contain:

- District
- District Latitude
- District Longitude
- Venue
- Venue Latitude
- Venue Longitude
- Venue Category

Out[35]:

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bandon Hill	51.364777	-0.134833	Asif Balti House	51.367795	-0.132356	Indian Restaurant
1	Bandon Hill	51.364777	-0.134833	Sharon's Perfect Pantry	51.367678	-0.131585	Bakery
2	Bandon Hill	51.364777	-0.134833	Mellows Park	51.360692	-0.134101	Park
3	Beddington	51.371988	-0.132393	Carew Manor	51.370983	-0.136604	Park
4	Beddington	51.371988	-0.132393	Wickes	51.375519	-0.130732	Hardware Store
5	Beddington	51.371988	-0.132393	Asif Balti House	51.367795	-0.132356	Indian Restaurant
6	Beddington	51.371988	-0.132393	The Plough	51.367633	-0.132089	Pub
7	Beddington Corner	51.386942	-0.149532	Traq	51.389091	-0.144979	Racetrack
8	Beddington Corner	51.386942	-0.149532	TNT Croydon Depot	51.389079	-0.144074	Business Service
9	Belmont	51.343785	-0.201152	Belmont Railway Station (BLM)	51.343857	-0.198677	Train Station

Figure 5: Top 100 Venues in each Neighborhoods of Sutton

3 Methodology

3.1 Exploratory Data Analysis

In this section, we will analysis the data frames built in the above section. And come up with the analysis on the best neighbourhood in Sutton. It consists of the below steps:

- Visualise the crime rates in the London boroughs to identify the safest borough and extract the neighbourhoods in that borough to _nd the 15most common venues in each neighbourhood.
- Visualise the price per m2 in each boroughs of London and extract the 15values with the least values.
- Analyse the happiness index and get the one with top 15 values.

3.1.1 Analysis of crime data

Using the describe function, we could get the statistical values of the columns in the dataset.

Out[17]:

	lsoa_code	borough	major_category	minor_category	value	year	month
count	392042	392042	392042	392042	392042.000000	392042.0	392042.000000
unique	4835	33	7	28	NaN	NaN	NaN
top	E01004763	Lambeth	Theft and Handling	Harassment	NaN	NaN	NaN
freq	256	17605	129159	36213	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	1.877659	2016.0	6.543077
std	NaN	NaN	NaN	NaN	2.650033	0.0	3.423461
min	NaN	NaN	NaN	NaN	1.000000	2016.0	1.000000
25%	NaN	NaN	NaN	NaN	1.000000	2016.0	4.000000
50%	NaN	NaN	NaN	NaN	1.000000	2016.0	7.000000
75%	NaN	NaN	NaN	NaN	2.000000	2016.0	10.000000
max	NaN	NaN	NaN	NaN	149.000000	2016.0	12.000000

Figure 6: Crime data Analysis

From the above analysis, we could _nd that among the 33 boroughs of London, Lambeth has the highest crime rate. Out of the 392042 crimes were reported in the year 2016, Theft and Handling were most of them.

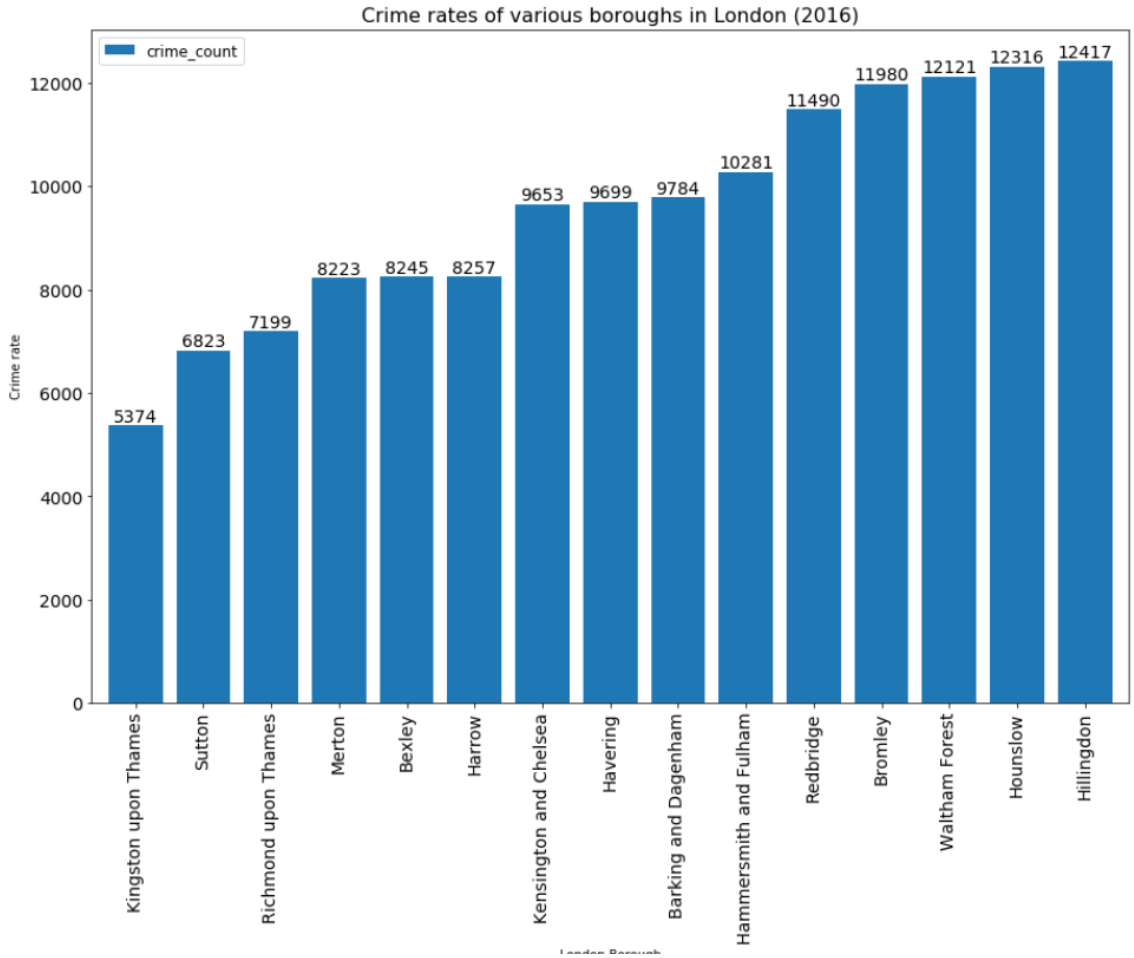


Figure 7: Boroughs with lowest crime rates

From the plot, we could see that Kingston Upon Thames has the lowest crime rate, followed by Sutton, Richmond upon Thames and Merton.

3.1.2 Analysis of Housing price

We need to analyse the boroughs which are affordable by comparing the general trend in the price of the houses in the neighbourhood. For that we have built a dataframe with housing price published by ONS.

Out[22]:

	price per m2
count	32.000000
mean	7473.187500
std	3423.874446
min	3994.000000
25%	5262.500000
50%	6510.000000
75%	8549.750000
max	19439.000000

Figure 8: Analysis of Housing price

From the above analysis, the mean value of the housing price per m2 in London is 7473.2, with a minimum value of 3994 and a maximum value of 19439. Now we can check the boroughs which has the lowest rates:

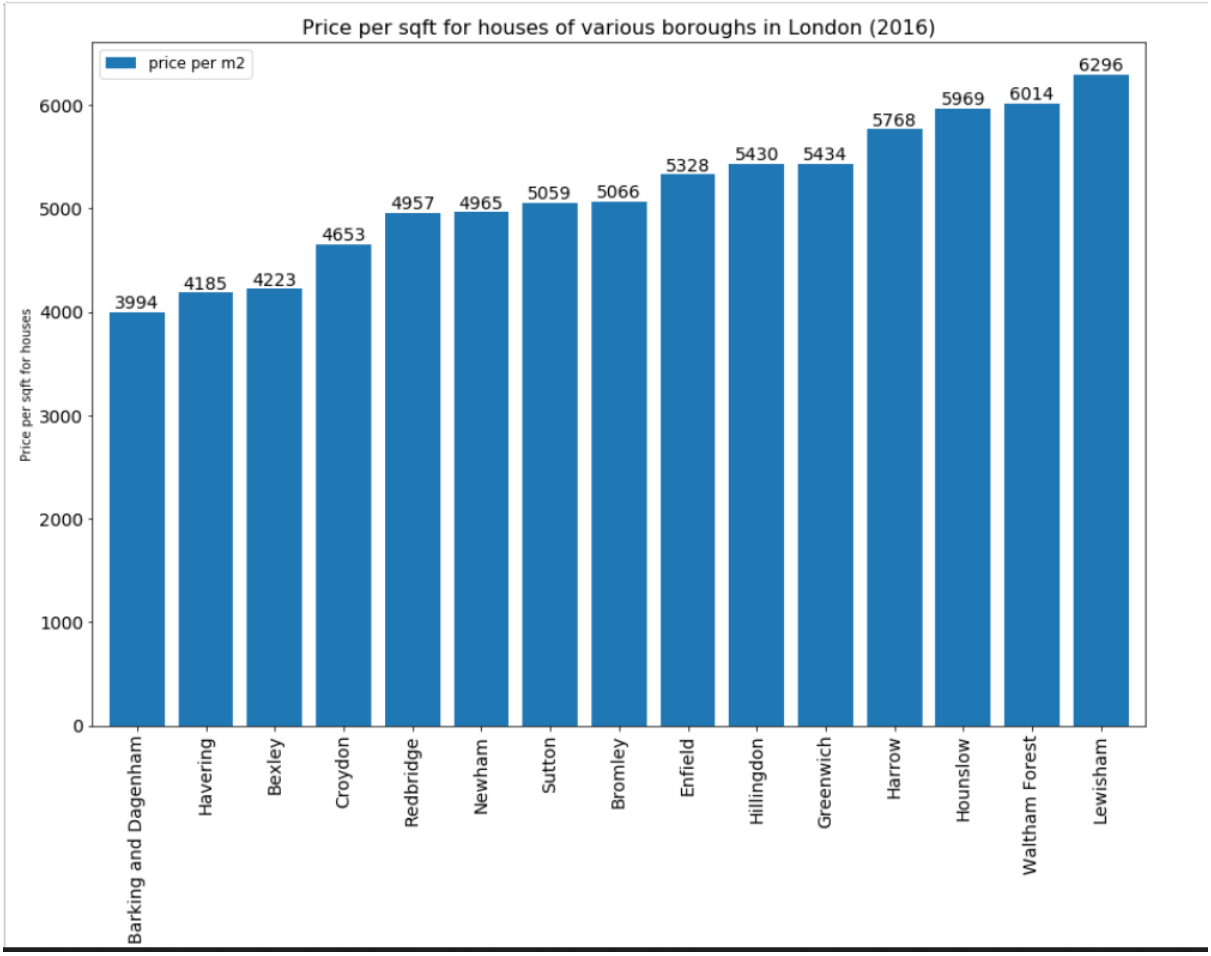


Figure 9: Boroughs with lowest Housing price

From the plot above, some of the most affordable regions in London are in Barking and Dagenham, Havering, Bexley and Croydon.

3.1.3 Analysis of happiness index

The boroughs with the best happiness index can be shown as:

	London_Borough
0	Richmond Upon Thames
1	Kingston upon Thames
2	Bromley
3	Sutton
4	Wandsworth
5	Camden
6	Barnet
7	Ealing
8	Greenwich
9	Havering
10	Hackney
11	Waltham Forest
12	Merton
13	Kensington & Chelsea
14	Hammersmith & Fulham

Figure 10: Boroughs with best Happiness index

Considering the crime rates, affordability in terms of price per sq feet and the happiness index, I have decided that these boroughs in London are the best ones for buying a house: Sutton, Bromley, Havering, Waltham Forest, with Sutton being the best choice. It is placed second in terms of least crime rate, seventh in affordability, which is much less than the mean value of houses in London (7473.1875) and fourth in happiness index.

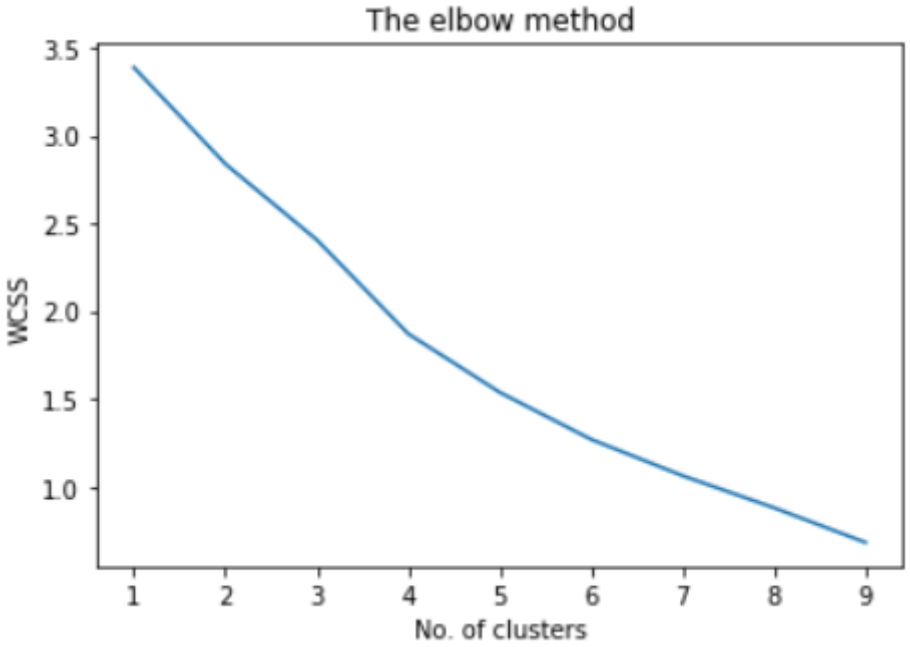
3.2 Modelling

Now, using the data set containing the geographical coordinates of Sutton Districts and FourSquare API, we can _nd the 100 top most venues in each district. There are 60 unique categories of venues in Sutton. One hot encoding is then performed on these venues to convert the categorical values to a form understandable by the ML algorithm called binarization. Now by grouping these venues on the basis of neighbourhood, we calculate the mean of the venues. Based on the mean values, we can _nd out the top 10 venues in each venue.

	District	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	Bandon Hill	Bakery	Park	Indian Restaurant	Veterinarian	Cosmetics Shop	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub
1	Beddington	Indian Restaurant	Hardware Store	Park	Pub	Hotel	Gym / Fitness Center	Gym	Grocery Store	Gastropub	Garden Center
2	Beddington Corner	Business Service	Racetrack	Veterinarian	Convenience Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub	Garden Center
3	Belmont	Train Station	Asian Restaurant	Bus Stop	Pub	Veterinarian	Convenience Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub
4	Benhilton	Indian Restaurant	Gym / Fitness Center	Grocery Store	Supermarket	Park	Pizza Place	Clothing Store	Coffee Shop	Cosmetics Shop	Gym

Figure 11: Top 10 venue categories the neighbourhood

Finally we can utilize K-Means clustering algorithm, to group the similar neighbourhoods. The neighbourhoods are classified into 6 categories using elbow method.



4 Results

Clusters are as shown below

	District	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
6	Carshalton Beeches	Sutton	51.357196	-0.169351	0	Train Station	Italian Restaurant	Bakery	Grocery Store	Veterinarian	Convenience Store	Gym / Fitness Center	Gym	Gastropub
15	Sutton	Sutton	51.357511	-0.173640	0	Train Station	Italian Restaurant	Bakery	Grocery Store	Park	Veterinarian	Convenience Store	Gym / Fitness Center	Gym

Figure 12: First cluster

	District	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
0	Bandon Hill	Sutton	51.364777	-0.134833	1	Bakery	Park	Indian Restaurant	Veterinarian	Cosmetics Shop	Hardware Store	Gym / Fitness Center	Gym	Grocery Store
1	Beddington	Sutton	51.371988	-0.132393	1	Indian Restaurant	Hardware Store	Park	Pub	Hotel	Gym / Fitness Center	Gym	Grocery Store	Gastropub
14	South Beddington	Sutton	51.371988	-0.132393	1	Indian Restaurant	Hardware Store	Park	Pub	Hotel	Gym / Fitness Center	Gym	Grocery Store	Gastropub

Figure 13: Second cluster

	District	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
3	Belmont	Sutton	51.343785	-0.201152	2	Train Station	Asian Restaurant	Bus Stop	Pub	Veterinarian	Convenience Store	Gym / Fitness Center	Gym
4	Benhilton	Sutton	51.371642	-0.191571	2	Indian Restaurant	Gym / Fitness Center	Grocery Store	Supermarket	Park	Pizza Place	Clothing Store	Coffee Shop
5	Carshalton	Sutton	51.365788	-0.161086	2	Grocery Store	Park	Pub	Coffee Shop	Gym / Fitness Center	Platform	Hotel	Train Station
8	Cheam	Sutton	51.357616	-0.216241	2	Pub	Italian Restaurant	Grocery Store	Thai Restaurant	Restaurant	Gym / Fitness Center	Indian Restaurant	Creperie
9	Hackbridge	Sutton	51.379613	-0.156754	2	Train Station	Supermarket	Park	River	Veterinarian	Construction & Landscaping	Gym	Grocery Store
11	North Cheam	Sutton	51.371578	-0.220225	2	Coffee Shop	Soccer Field	Pub	Gym / Fitness Center	Fast Food Restaurant	Seafood Restaurant	Social Club	Supermarket
12	Rosehill	Sutton	51.012505	-0.140639	2	Restaurant	Grocery Store	Business Service	Pub	Veterinarian	Gym / Fitness Center	Gym	Gastropub
16	Sutton Common	Sutton	51.375373	-0.196032	2	Indoor Play Area	Gym / Fitness Center	Athletics & Sports	Tennis Court	Grocery Store	Park	Convenience Store	Gym
17	Sutton High Street	Sutton	51.359765	-0.190991	2	Pub	Coffee Shop	Pizza Place	Café	Bar	Italian Restaurant	Portuguese Restaurant	Gym
18	The Wrythe	Sutton	51.367059	-0.162956	2	Pub	Café	Grocery Store	Park	Veterinarian	Platform	Gym / Fitness Center	Coffee Shop
19	Wallington	Sutton	51.357945	-0.149562	2	Grocery Store	Supermarket	Pharmacy	Bookstore	Fast Food Restaurant	Coffee Shop	Pizza Place	Portuguese Restaurant
21	Worcester Park	Sutton	51.378400	-0.241602	2	Grocery Store	Pub	Pharmacy	Cosmetics Shop	Pizza Place	Coffee Shop	Pet Store	Pet Service

Figure 14: Third cluster

	District	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
10	Little Woodcote	Sutton	51.346076	-0.145932	3	Construction & Landscaping	Park	Garden Center	Coffee Shop	Convenience Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store
20	Woodcote Green	Sutton	51.347991	-0.146830	3	Construction & Landscaping	Park	Garden Center	Coffee Shop	Convenience Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store

Figure 15: Fourth cluster

	District	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
2	Beddington Corner	Sutton	51.386942	-0.149532	4	Business Service	Racetrack	Veterinarian	Convenience Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub

Figure 16: Fifth cluster

	District	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
13	St. Helier	Sutton	51.386695	-0.180057	5	Breakfast Spot	Coffee Shop	Veterinarian	Convenience Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub

Figure 17: Sixth cluster

5 Discussion

After clustering neighbourhood into 6 clusters we could see that third cluster is the biggest one with all amenities, like Train station, Restaurant, Grocery store, Pub etc. If people prefer to live in all amenities location then third cluster is the best. These are the areas which attract lot of floating crowd. First cluster contains two neighbourhood where most common place is train station. People who use travel can use this area. Second cluster containing mostly bakery and restaurants which can opt preferably for foodies. Fifth cluster contain only one place where most common place is business service and racetrack etc. Last cluster also contain only one place where mainly for those who prefer to eat outside. Amenities contain breakfast spot, coffee shops etc.

6 Conclusion

This project helps me get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. We have just taken safety, affordability and accessibility as the primary concerns to shortlist the borough of London. We also have issues in getting the latest dataset for analysis. The future of this project includes taking factors such as cost of living in the areas, employment rate, transportation etc into consideration to shortlist the borough.