Capstone Project- The Battle of Neighborhoods Report -

Raghavendra Kumar Punugu

1. Introduction

1.1 Background

A lot of people keep switching to different places. These relocations could be based on the several factors including Job Relocation, Better climatic conditions or any other unmentioned reasons. When someone wants to move to different location, he/she looks for several factors including climatic conditions, cost of living, safety etc. Safety can be of paramount importance among the others. He/she would want to know the neighborhoods, past crime records, occurrence of crime based on the degree etc. If one does not feel safe staying in one's home, it is not worth living there and enjoying the neighborhood.

1.2 Problem

The Crime dataset used in this project is obtained from Kaggle, the dataset contains information about Crime statistics in London. We will be exploring 2016 crime data, as 2016 is the latest year data points we have. The aim of this is project to find the safest neighborhoods in and around London based on the crimes committed. We will find the safest neighborhoods and explore the neighborhood using K-means clustering.

1.3 Interest

People who are moving to different cities for any reason including short term to long term stay etc. will benefit from this study in making a well-informed decision.

2. Data Cleaning

2.1 Data Preparation

The dataset 1 (London Crime) has different columns:

- **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
- **value:** monthly reported count of categorical crime in given borough
- **year**: Year of reported counts, 2008-2016
- **month**: Month of reported counts, 1-12

Dataset 2 (List of London borough), contains the list of London boroughs which is obtained from wikipedia. It contains the following columns:

- **Borough:** Names of Borough in London
- **Inner:** Categorizing the borough as Inner London Borough or an Outer London Borough
- **Status:** Categorizing the borough as Royal, City or other

- **Local Authority**: Local Authority assigned to that borough
- **Political Control**: Political party that controls the borough
- **Headquarters:** HQ of the borough
- **Area (sq mi):** Area of the borough
- **Population**: Population in the borough
- **Nr. In map**: The number assigned to each borough to represent on a map

2.2 Data Cleaning

The data preparation is done individually for each of the data set. The below dataset in table 1 represents the London Data Crime (head = 5) for year 2016,

	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
2	E01000677	Bromley	Violence Against the Person	Other violence	0	2015	5
3	E01003774	Redbridge	Burglary	Burglary in Other Buildings	0	2016	3
4	E01004563	Wandsworth	Robbery	Personal Property	0	2008	6

Table 1: London Crime Data

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	Barking and Dagenham [note 1]	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

Table 2: Longitude and Latitude obtained from Wikipedia using Beautiful soap

Table 2 represents the borough in London with longitudes and latitudes. The two datasets are merged based on Borough which is represented in Table 3.

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
Westminster	Westminster City Council	Conservative	Westminster City Hall, 64 Victoria Street	8.29	226841	51°29′50″N 0°08′14″W / 51.4973°N 0.1372°W	NaN	2179	2049	708	1822	NaN	NaN	48330
Lambeth	Lambeth London Borough Council	Labour	Lambeth Town Hall, Brixton Hill	10.36	314242	51°27′39″N 0°06′59″W / 51.4607°N 0.1163°W	NaN	2764	2738	635	1196	NaN	NaN	34071
Southwark	Southwark London Borough Council	Labour	160 Tooley Street	11.14	298464	51°30′13″N 0°04′49″W / 51.5035°N 0.0804°W	NaN	2621	1838	494	1317	NaN	NaN	31636
Newham	Newham London Borough Council	Labour	Newham Dockside, 1000 Dockside Road	13.98	318227	51°30′28″N 0°02′49″E / 51.5077°N 0.0469°E	NaN	2496	1684	713	1472	NaN	NaN	30090
Tower Hamlets	Tower Hamlets London Borough Council	Labour	Town Hall, Mulberry Place, 5 Clove Crescent	7.63	272890	51°30′36″N 0°00′21″W / 51.5099°N 0.0059°W	NaN	2357	1629	678	1234	NaN	NaN	29253

Table 3: The merged dataset based on Borough

3.Methodology

3.1 Exploratory Data Analysis

3.1.1 Statistical Summary of Crimes

Using the Describe () function in pandas, we get the descriptive statistics including mean, count, standard deviation, min, max and so on.

	Bulgary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and handling	Violence against people	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

Table 4: Descriptive Statistics of Crime Data

From the table 4 we can infer that, Theft and handling is the highest reported crime in year 2016 followed by Violence against person, Burglary, Criminal Damage and so on. Other Notifiable offences is lowest in reported crime for year 2016.

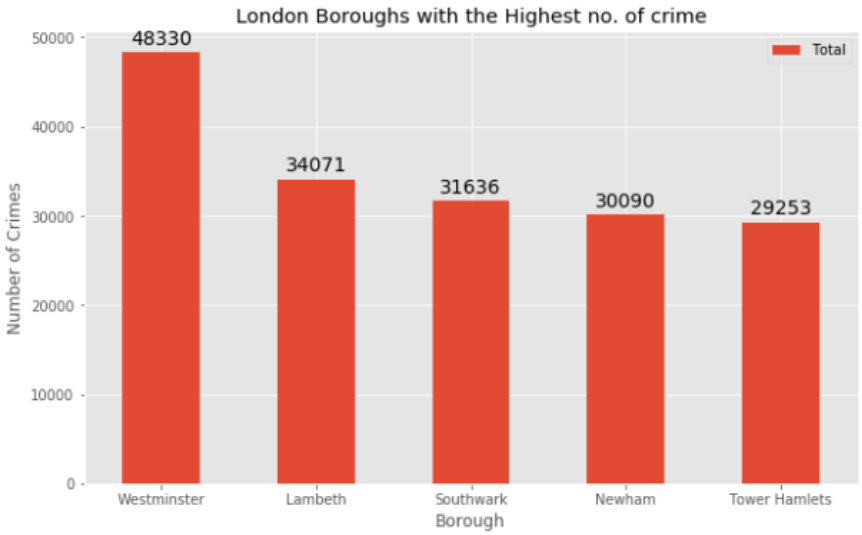

Figure 1: London Borough with the Highest number of Crime

Figure 1 represents the London Borough with Highest number of crimes reported in descending order. Westminster has the highest number of reported crimes which is 48330 which is followed by Lamberth (34071), Southwark (31636), Newham(30090) and Tower Hamlets (29253).

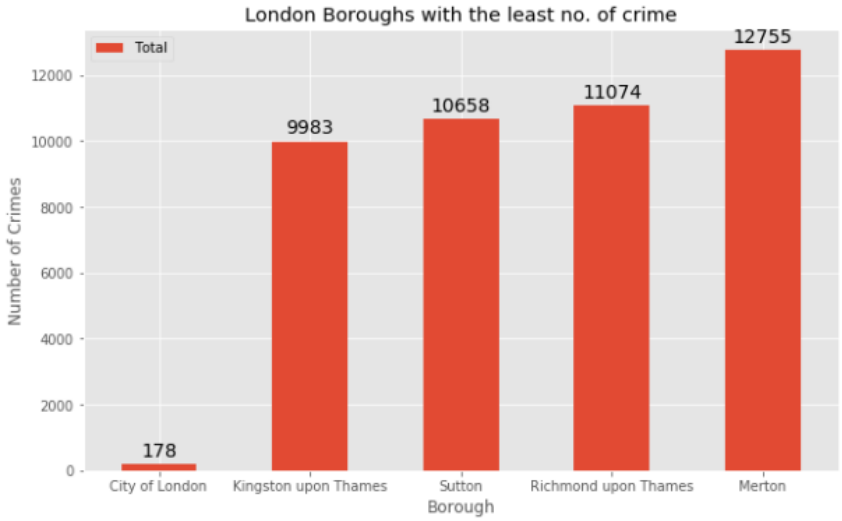

Figure 2: London Borough with the least number of Crimes

Figure 2 represents the London borough with least number of Crimes reported for year 2016. City of London has the least number of reported crime of 178 instances among all the Boroughs in London, followed by Kingston upon Thames (9983), Sutton Borough (10658), Richmond upon Thames(11074) and Merton (12755).

Figure 3: Neighborhoods in Kingston upon Thames

Figure 3 represents the 15 Neighborhoods in Kingston upon Thames, they are represented in table 7 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

Table 7: 15 Neighborhoods in Kingston upon Thames

3.2 Modeling

Using the Final Dataset containing the 15 neighborhoods in Kingston upon Thames along with longitudes and latitudes, we explore venues within 500 meters of each neighborhood using Foursqaure API. This returns a json file containing all the venues in each neighborhood which is converted to Pandas DataFrame.

	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	Cafe Rosa	51.390175	-0.282490	Café
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

Table 8: Some neighborhoods around Kingston upon Thames

One hot encoding is done to convert categorical variables to numerical form that can be used by ML algorithm for quicker processing. We finally select 10 common venues near each of the neighborhood. Using k-means clustering we are going to cluster the neighborhood which have similar profile and group them. People can short list their venues/neighborhoods based on this clustering.

	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 Co
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	0.0	Convenience Store	Bus Stop	Gym / Fitness Center	Colombian Restaurant	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4.0	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Plaza	Rest
2	Chessington	Kingston upon Thames	51.358336	-0.298622	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Coombe	Kingston upon Thames	51.419450	-0.265398	3.0	Health & Beauty Service	Wine Shop	Food	Department Store	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	
4	Hook	Kingston upon Thames	51.367898	-0.307145	4.0	Bakery	Supermarket	Fish & Chips Shop	Indian Restaurant	Wine Shop	Food	Department Store	Discount Store	Elect

Table 9: Top 5 neighborhoods around Kingston upon Thames.

4. Results

When have obtained the neighborhoods based on the five clusters and we see that cluster 5 has 7 Neighborhoods, where as cluster 2 has 3 neighborhoods and cluster 3 has 2, while cluster 1 and 4 have one each.

Cluster 1

Based on the cluster 1, we see that this cluster has just neighborhood which is Berrylands in Kingston upon Thames, and the most common venue is Convenience store followed by Bus Stop and Gym/Fitness center.

	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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	0	Convenience Store	Bus Stop	Gym / Fitness Center	Colombian Restaurant	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

Table 10: Neighborhood based on Cluster 1

Cluster 2

Based on the below table, we see that the borough in cluster 2 has 3 neighborhoods, and among the three neighborhood, the most common venue is Grocery store for Kingston Vale and Tolworth, whereas for Mostspur Park, its most common venue is Bus Stop.

	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 Mos Commo Venu
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	1	Grocery Store	Italian Restaurant	Bar	Soccer Field	Department Store	Discount Store	Electronics Store	Farmers Market	Fast Foo Restaura
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Bus Stop	Gym	Park	Soccer Field	Wine Shop	Department Store	Discount Store	Electronics Store	Farme Mark
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Sandwich Place	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Discount Store	Coffe Sho

Table 11: Neighborhood based on cluster 2

Cluster 3

Based on the below table, we see that this cluster has 2 neighborhoods, the most common venue for Maiden Rushett is Garden center where as for Old Maiden it is Convenience Store. Followed by Pub which is the second most common venue for both the neighborhood.

	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 M Comm Ven
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Garden Center	Pub	Restaurant	Convenience Store	Wine Shop	Farmers Market	Deli / Bodega	Department Store	Disco St
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	2	Convenience Store	Pub	Food	Train Station	Bakery	Bar	Gift Shop	German Restaurant	Gastrop

Table 12: Neighborhood based on cluster 3

Cluster 4

In cluster 2 we have just one neighborhood which is Coombe in Kingston upon James borough, and the most common venue for this neighborhood is Health and Beauty Service followed by Wine Shop.

	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	10 Co
3	Coombe	Kingston upon Thames	51.41945	-0.265398	3	Health & Beauty Service	Wine Shop	Food	Department Store	Discount Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Res

Table 13: Neighborhood based on cluster 4

Cluster 5

Cluster 5 has the most neighborhoods among the five cluster we have modeled. Indian Restaurant is popular in two neighborhoods of Norbiton and Seething Wells, while Coffee Shop is the most common venue for Kingston upon Thames and Surbiton.

	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	9 C
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Plaza	Re
4	Hook	Kingston upon Thames	51.367898	-0.307145	4	Bakery	Supermarket	Fish & Chips Shop	Indian Restaurant	Wine Shop	Food	Department Store	Discount Store	El
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	4	Coffee Shop	Café	Sushi Restaurant	Pub	Burger Joint	Asian Restaurant	Department Store	Electronics Store	A Re
9	New Malden	Kingston upon Thames	51.405335	-0.263407	4	Gastropub	Supermarket	Korean Restaurant	Chinese Restaurant	Indian Restaurant	Bar	Office	Sushi Restaurant	
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	4	Indian Restaurant	Pub	Italian Restaurant	Food	Platform	Wine Shop	Hardware Store	Pharmacy	J Re
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	4	Indian Restaurant	Coffee Shop	Café	Italian Restaurant	Pub	Golf Course	Gym	Gym / Fitness Center	
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	4	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Gastropub	French Restaurant	Train Station	Gym / Fitness Center	

Table 14: Neighborhood based on cluster 5

5. Discussions

The aim of this project is to help people migrating to different London Borough to make well informed decisions based on their preferences. For example, if some one wants to move to a place with Indian restaurant around in his neighborhood he can prefer either Norbiton or Seething Wells, however, if their immediate next preference is a pub they may most likely prefer Norbiton. A family that prefers to stay in the vicinity of grocery store can choose Kingston vale or Tolworth. The above results would be a significant help for people/ families trying to move to Kingston Upon Thames borough and select the neighborhood which suits their preference. Also, the clustering was done for 500 m, which in my opinion is not a deal breaker for any person/persons, Some venues may be a little far than the desired distance but then people tend to get a better holistic picture of the neighborhoods and their most common venues.

6. Conclusion

These results will help people make well-informed decisions ahead of their time and is hassle free process. In the future work, we will work on considering several additional factors including cost of living, schooling district etc.