Applied Data Science

Capstone Project

The Battle of Neighborhoods

By: Tushar Supe

1. Introduction

1.1 Introduction / Business Problem

Change is the only constant in today's modern age. People move from place to place to find comfort in life whether it is due to safety reasons, transportation or better food availability or job opportunities. But it's hard to relocate to different places and experience their conditions by actually living there. If it can be done virtually or with the help of data analysis then a lot of money and time will be saved and a relocator will definitely find the best place to live which fulfills all his/her needs.

So here in this project best and safest places to relocate in London are found out with analysis. The London crime data is obtained from Kaggle. Different boroughs in London are studied based on the total crimes, and neighbourhoods are explored for safest borough and analysed using k-mean clustering.

1.2 Interest

This report is targeted to people who are relocating to London considering safety as top priority. The crime data is taken into consideration to select safest neighbourhoods in London. Furthermore, for the safest borough, the neighbourhoods are clustered based on most common venues which are obtained from Foursquare API. Based on clusters a person can find best neighbourhood to relocate.

2. Data Loading

2.1 Data Acquisition

The data acquired from following data sources:

- The dataset consisting of the crime statistics of each borough in London obtained from Kaggle
- Co-ordinate of neighborhood obtained using Google Maps API geocoding
- Most common venues obtained by Foursquare API

The London crime data is obtained from Kaggle. The dataset contains the following 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 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

For the safest borough the coordinate of neighborhood obtained using Google Maps API geocoding. For that the neighborhood list is obtained from Wikipedia.

For the different neighborhoods the most common venues are obtained from Foursquare API.

2.2 Data Cleaning

The London crime data is loaded from Kaggle and is shown in fig 2.1

	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

Fig 2.1 London crime data before preprocessing

The data of year 2016 is only selected. As shown in fig 2.2

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01004177	Sutton	Theft and Handling	Theft/Taking of Pedal Cycle	1	2016	8
1	E01000733	Bromley	Criminal Damage	Criminal Damage To Motor Vehicle	1	2016	4
2	E01003989	Southwark	Theft and Handling	Theft From Shops	4	2016	8
3	E01002276	Havering	Burglary	Burglary in a Dwelling	1	2016	8
4	E01003674	Redbridge	Drugs	Possession Of Drugs	2	2016	11

Fig 2.2 London crime data 2016

Later for year 2016 only Borough, Major crime category and Number of crimes is selected. As shown in fig 2.3

	Borough	Major_Category	No_of_Crimes
0	Sutton	Theft and Handling	1
1	Bromley	Criminal Damage	1
2	Southwark	Theft and Handling	4
3	Havering	Burglary	1
4	Redbridge	Drugs	2

Fig 2.3 London crime data after preprocessing

Further the data is pivoted across number of crimes of different crimes as shown in fig 2.4

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	33482
1	Barnet	3402	2183	906	499	464	9731	7499	49368
2	Bexley	1123	1673	646	294	209	4392	4503	25680
3	Brent	2631	2280	2096	536	919	9026	9205	53386
4	Bromley	2214	2202	728	417	369	7584	6650	40328

Fig 2.4 London crime data further processing

Later for the safest borough in London the neighborhoods are obtained from Wikipedia and the coordinates are obtained from Google API geocoder. Final data frame is shown in fig 2.5

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames	51.3938	-0.284802
1	Canbury	Kingston upon Thames	51.4175	-0.305553
2	Chessington	Kingston upon Thames	51.3583	-0.298622
3	Coombe	Kingston upon Thames	51.4194	-0.265398
4	Hook	Kingston upon Thames	51.3679	-0.307145

Fig 2.5 Neighborhoods of the safest borough along with coordinates

For the neighborhoods best 10 most common venues are obtained using the Foursquare API,

	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	Café	Bus Stop	Wine Shop	Food	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant
1	Canbury	Pub	Hotel	Shop & Service	Café	Plaza	Indian Restaurant	Fish & Chips Shop	Park	Spa	Supermarket
2	Hook	Bakery	Supermarket	Fish & Chips Shop	Indian Restaurant	Wine Shop	French Restaurant	Electronics Store	Farmers Market	Fast Food Restaurant	Food
3	Kingston Vale	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food
4	Kingston upon Thames	Coffee Shop	Pub	Sushi Restaurant	Café	Burger Joint	Asian Restaurant	Gift Shop	Furniture / Home Store	French Restaurant	Electronics Store

Fig 2.6 Neighborhoods with most common places

3. Methodology

3.1 Data Analysis

Statistical summary

The statistical summary is shown in fig 3.1

	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
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	44613.393939
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	17656.457498
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	356.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	33806.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	45460.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	54348.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	96660.000000

Fig 3.1. Statistical description of the London crimes

Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses.

Boroughs with the highest crime rates

Top five boroughs with the highest crime rate are Westminster, Lambeth, Southwark, Newham, Tower Hamlets. From fig 3.2

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
32	Westminster	3218	2179	2049	708	1822	27520	10834	96660
21	Lambeth	3087	2764	2738	635	1196	13155	10496	68142
27	Southwark	2946	2621	1838	494	1317	12946	9474	63272
24	Newham	2115	2496	1684	713	1472	11964	9646	60180
29	Tower Hamlets	2794	2357	1629	678	1234	10953	9608	58506

Fig 3.2 Boroughs with the highest crime rates

Boroughs with the lowest crime rates

Top low crime rate borough are Kingston upon Thames, Sutton, Richmond upon Thames and Merton from fig 3.3

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
25	Redbridge	1997	1650	1017	381	599	7447	6411	39004
15	Havering	1826	1804	718	389	311	5919	5936	33806
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	33482
14	Harrow	1994	1212	473	267	377	4537	4293	26306
2	Bexley	1123	1673	646	294	209	4392	4503	25680
23	Merton	1419	1418	466	249	283	4894	4026	25510
26	Richmond upon Thames	1359	1148	320	217	106	4769	3155	22148
28	Sutton	1233	1316	461	253	165	3516	3714	21316
20	Kingston upon Thames	879	1054	743	189	121	3803	3194	19966

Fig 3.3 Boroughs with the lowest crime rates

Neighborhoods in Kingston upon Thames

There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python in fig 3.4

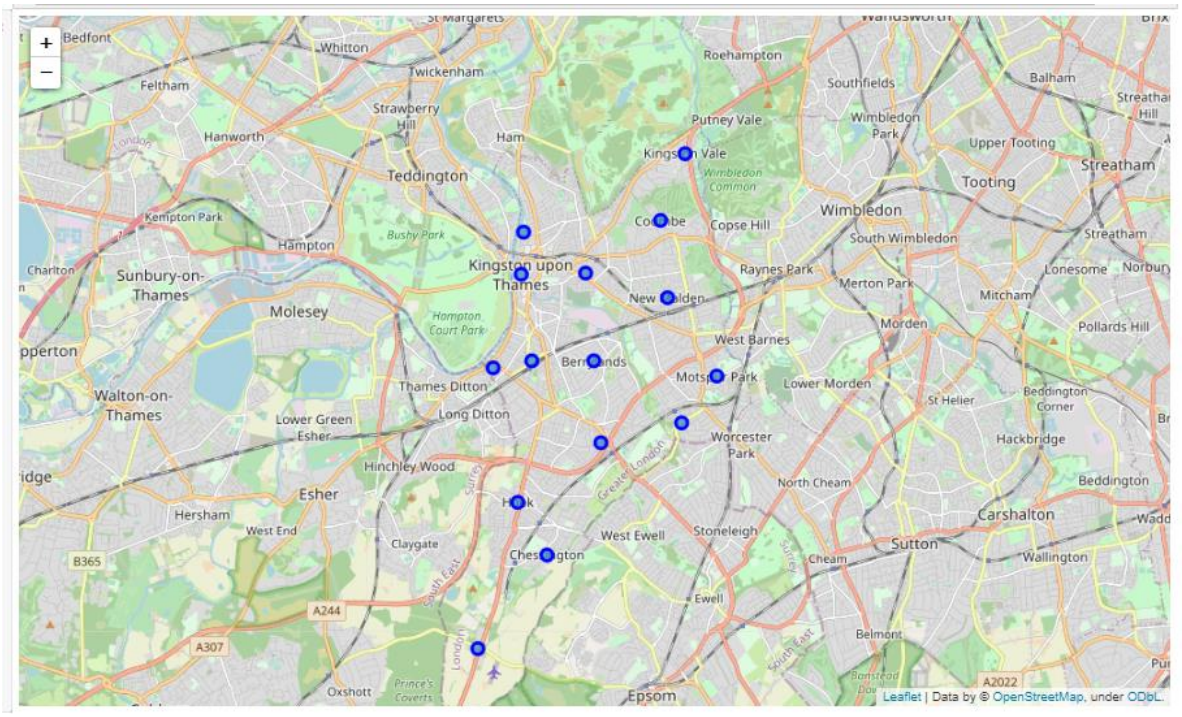

Fig 3.4 Neighborhoods in Kingston upon Thames

3.2 Modelling

For neighborhoods in Kingston upon Thames all the venues within a 500 m radius of are obtained from Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a pandas data frame. As shown in fig 3.5

	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	Canbury Gardens	51.417409	-0.305300	Park

Fig 3.5 Venue details of each Neighborhood

One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

The neighborhoods are clustered using K - means clustering and visualized data using color coding the clusters.

4. Results

After K-means clustering the neighborhoods are classified into 5 different clusters

	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	51.4318	-0.258138	0	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Electronics Store	Farmers Market	Fast Food Restaurant	Fish Chip Sho
7	Malden Rushett	Kingston upon Thames	51.3411	-0.319076	0	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Fish & Chips Shop	Department Store	Electronics Store	Farmer Marke
8	Motspur Park	Kingston upon Thames	51.391	-0.248898	0	Gym	Park	Restaurant	Soccer Field	Fish & Chips Shop	Department Store	Electronics Store	Farmers Market	Fast Foo Restaurar
14	Tolworth	Kingston upon Thames	51.3789	-0.28286	0	Grocery Store	Pharmacy	Bowling Alley	Coffee Shop	Italian Restaurant	Pizza Place	Hotel	Café	Bus Sto

Fig 4.1 Cluster 1

The 1st cluster has 4 of 15 neighborhoods in the borough Kingston upon Thames. The most common venues are Gym, sandwich place and grocery stores, pharmacy, park. .

	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
11	Old Malden	Kingston upon Thames	51.3825	-0.25909	1	Construction & Landscaping	Train Station	Food	Deli / Bodega	Department Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

Fig 4.2 Cluster 2

The second cluster has one neighborhood with Construction & Landscaping as most common venue.

	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
4	Hook	Kingston upon Thames	51.3679	-0.307145	2	Bakery	Supermarket	Fish & Chips Shop	Indian Restaurant	Wine Shop	French Restaurant	Electronics Store	Farmers Market	Fast Food Restaurant

Fig 4.3 Cluster 3

The third cluster has one neighborhood with Bakery as most common venue.

	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 M Comm Ve
0	Berrylands	Kingston upon Thames	51.3938	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Food	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Fre Restau

Fig 4.4 Cluster 4

The fourth cluster has one neighborhood with Gym / Fitness Center as most common venue.

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
Canbury	Kingston upon Thames	51.4175	-0.305553	4	Pub	Hotel	Shop & Service	Café	Plaza	Indian Restaurant	Fish & Chips Shop	Park	Spa
Kingston upon Thames	Kingston upon Thames	51.4096	-0.306262	4	Coffee Shop	Pub	Sushi Restaurant	Café	Burger Joint	Asian Restaurant	Gift Shop	Furniture / Home Store	French Restaurant
New Malden	Kingston upon Thames	51.4053	-0.263407	4	Gym	Gastropub	Sushi Restaurant	Supermarket	Bar	Indian Restaurant	Korean Restaurant	Food	Electronics Store
Norbiton	Kingston upon Thames	51.41	-0.287396	4	Pub	Italian Restaurant	Indian Restaurant	Food	Platform	Wine Shop	Coffee Shop	Hotel	Hardware Store
Seething Wells	Kingston upon Thames	51.3926	-0.314366	4	Indian Restaurant	Coffee Shop	Pub	Café	Gym	Restaurant	Park	Pet Café	Fast Food Restaurant
Surbiton	Kingston upon Thames	51.3938	-0.30331	4	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	French Restaurant	Train Station	Gym / Fitness Center	Hotel

Fig 4.5 Cluster 5

The fifth cluster is biggest cluster with 6/15 neighborhood which consists of Venues such as Pub, Coffee shop, gym, indian restaurant

Visualising the clustered neighborhoods on a map using the folium library fig 4.6.

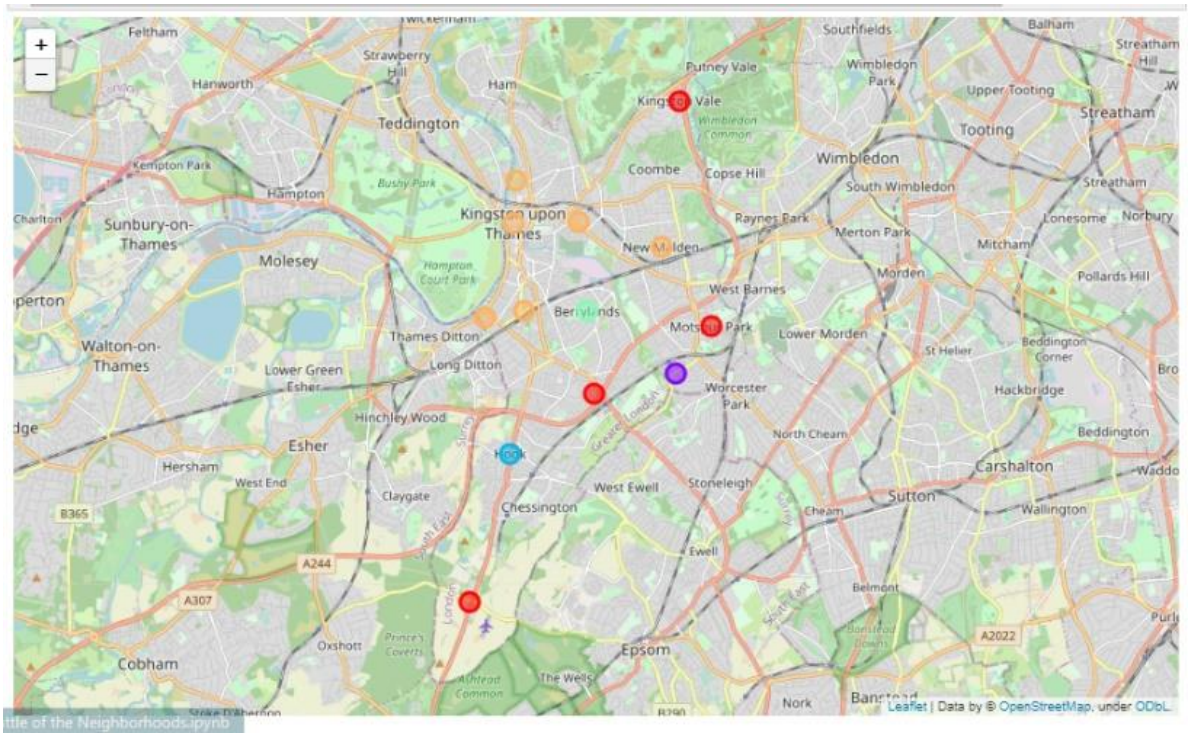

Fig 4.6 Clustered neighborhoods in the Borough of Kingston upon Thame

5. Discussion

The aim of this project is to find safest borough in London for people who are relocating to London. For the safest borough the neighbourhood need to be analyzed so that person will be able to enjoy most common venues and could enjoy life along with safety. Most important venues include pharmacy, gym, restaurants, ease of transportation and grocery stores.

From the analysis of neighborhoods we can see that 1st cluster includes venues like Gym, sandwich place and grocery stores, pharmacy, park. 2nd cluster includes Construction & Landscaping, Train Station, Food, Deli / Bodega, Department Store, Electronics Store, Farmers Market. 3rd cluster includes Bakery, Supermarket, Fish & Chips Shop, Indian Restaurant, Wine Shop, French Restaurant, Electronics Store. 4th cluster includes Gym / Fitness Center, Park, Café, Bus Stop, Wine Shop, Food, Farmers Market. 5th cluster includes Pub, Coffee shop, gym, Indian restaurant, pharmacy, mostly restaurants and eatery.

For a family 1st cluster is more suitable dues to the common venues in that cluster includes most of the essential services. For a foodie person I think 5th cluster is more suitable.

.

6. Conclusion

Relocation is most difficult and hectic process and it involves moving to new unknown location where we don't about food, transportation services and neighboring area and safety. so due to new techniques in data science we can analyze different neighborhoods as if we are virtually present there and we can choose best area to live which is safe for our family and also fulfills all necessary needs. Here in this project we analyzed different borough of London based on crime rates and shortlisted some neighborhood of safe borough and finally categorized different clusters based on most common venues. this can be helpful for person relocating to London and choosing best neighborhood to live.