THE BATTLE OF NEIGHBORHOODS

Applied Data Science Capstone - IBM

Nishit Shetty

Table of Contents

Introduction

With a rising cases of crimes every year, it is very important to understand the crime rates in the nearby locality before buying a house. So, considering that I want to move to London, I will be looking for a safe locality to stay where the crime rate is low and safe environment for me to stay.

It can also be used by expats who are looking for safe localities to shift in London.

Also, it is important to consider the trending venues and the host spots of the city to understand places that are most important for you to consider when you want to move. I will also be looking at the Top 10 Common Venues each and around the neighbourhood using k-means clustering.

Data

Data for this project will be from three references over the internet.

1. Crime Data for London
   The crime data for London will be picked from Kaggle: https://www.kaggle.com/jboysen/london-crime. The data set contains the following fields: lsoa_code, borough, major_category, minor_category, value, year, month

2. List of Boroughs in London
   The data for the same is being picked from Wikipedia: https://en.wikipedia.org/wiki/List_of_London_boroughs. The data set contains the following fields: Borough, Inner, Status, Local authority, Political control, Headquarters, Area (sq mi), Population (2013 est)[1], Co-ordinates & Nr. in map

3. List of Localities in Royal Borough of Kingston upon Thames
   Post analysis, the data of neighbourhoods is being picked from Wikipedia: https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames. The data set contains Neighbourhood, borough, latitude & longitude data.

4. Foursquare API
   The Foursquare API will be used for the project as well.

Methodology

Exploratory Data Analysis

Statistical Analysis of Crime Data

The crime data for London was analysed using descriptive statistics using .describe() function within Python to understand the data set in a better manner. The descriptive analysis includes mean, count, 1st quartile, 2nd quartile & 3rd quartile for each of the categories of crimes in London.

```
In [43]: London_crime.describe()
```

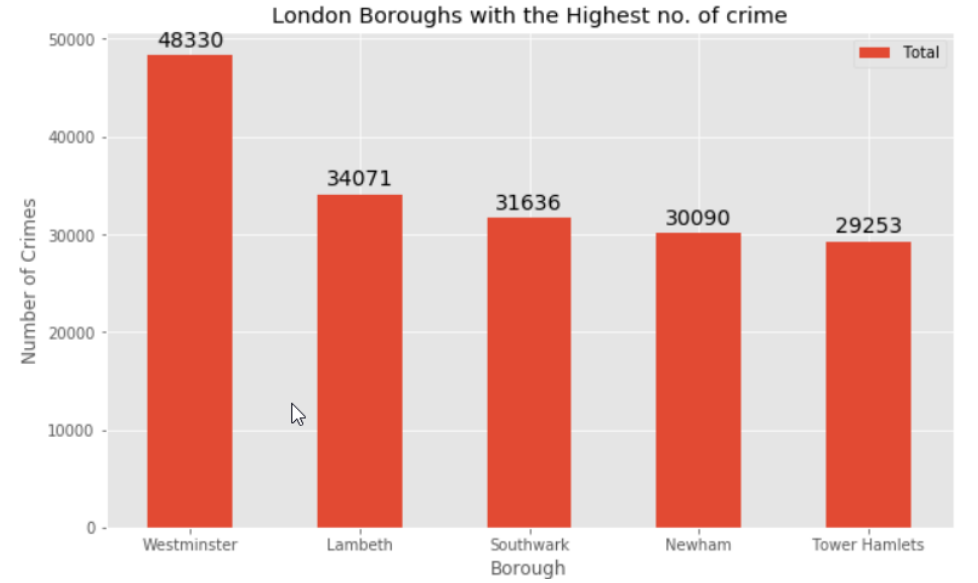
Out[43]:

	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

Crime Rate Analysis – Highest Crime Rate Borough-wise Data

top five boroughs with highest crime rates were analysed and displayed using a column graph. Based on the data, Westminister has the highest crime rate in London, followed by Lambeth, Southwark, Newham and Tower Hamlets.

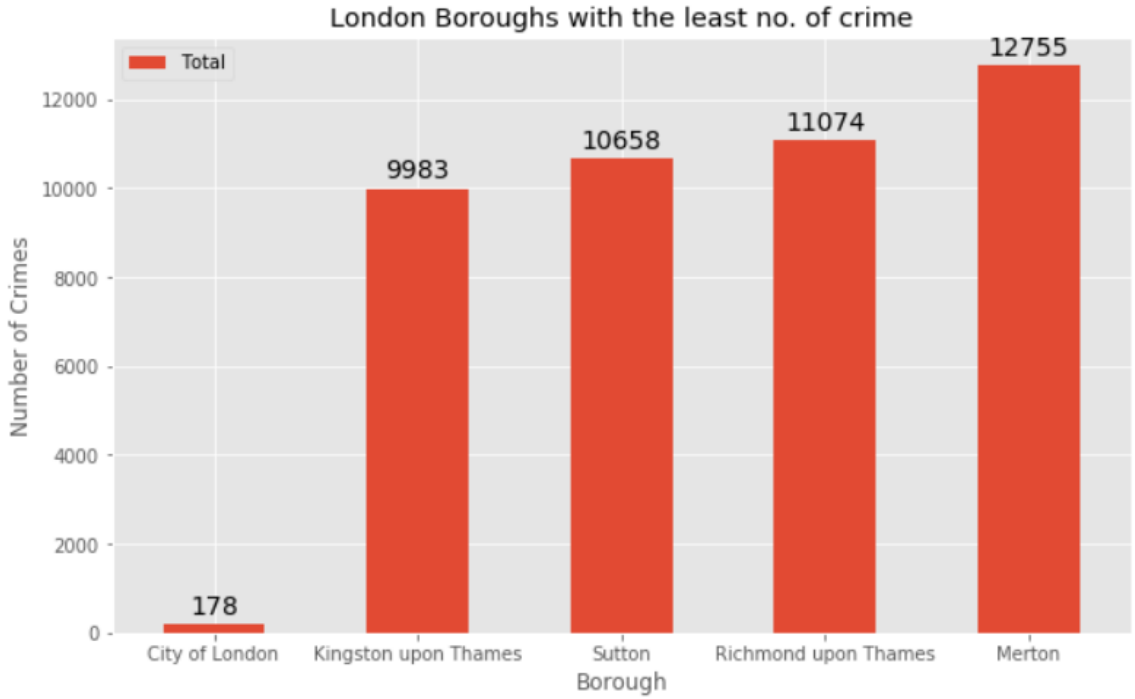
```
plt.show()
```



Crime Rate Analysis – Lowest Crime Rate Borough-wise Data

The lowest crime rate recorded for boroughs in London is the City of London. It is followed by Kingdom upon Thames, Sutton Borough, Richmond upon Thames & Merton. The graph of the same is as attached below:

```
plt.show()
```



For the sake of further analysis, the City of London will not be considered as it is the 33rd principal division of Greater London but nor a borough specifically. Additionally, the area of the City of London is just 1.12 sq. miles which is a very small area as compared to the other boroughs.

In [50]:
```
df_col = df_bot5[df_bot5['Borough'] == 'City of London']
df_col = df_col[['Borough','Total','Area (sq mi)','Population (2013 est)[1]']]
df_col
```

Out[50]:

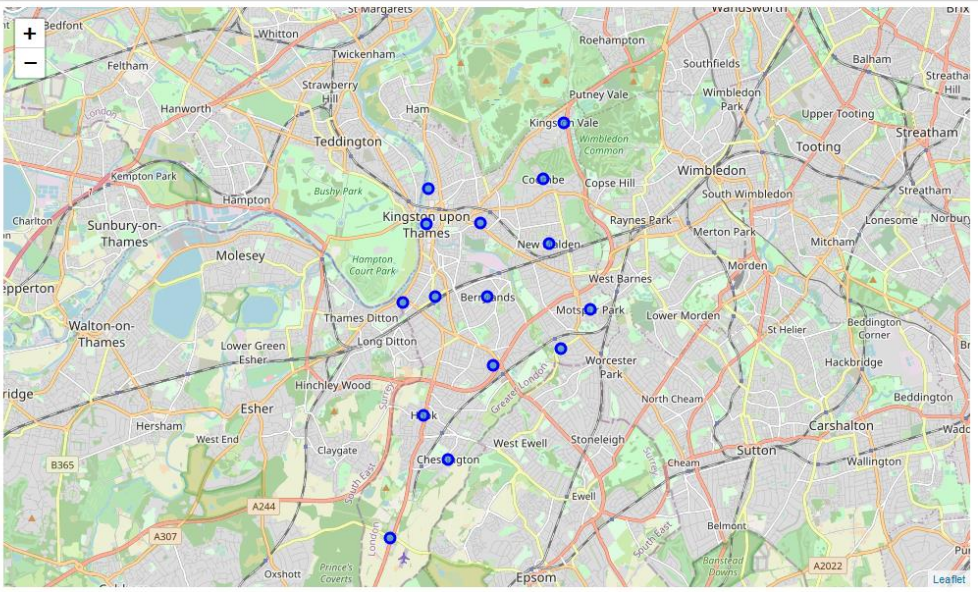
	Borough	Total	Area (sq mi)	Population (2013 est)[1]
6	City of London	178	1.12	7000

Neighbourhoods in Kingston upon Thames

15 neighbourhoods have been identified within Kinston upon Thames which have been visualised as follows using Folium on Python:

```
map_lon
```

Out[57]:



Neighbourhood Analysis

On finalising Kingston upon Thames for further analysis, the neighbourhoods within the borough were stated and their corresponding latitude and longitude was appended.

Out[55]:

	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

Post which, within a 500 m radius of each of the neighbourhood, venues were picked up from Foursquare using the API which was realised from the fetched JSON file. Data of the same is as follows:

Out[60]:

	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	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

One hot encoding helped to convert the data for better interpretation of the ML algorithm. Venues in the safest borough of London was then analysed using k-means algorithm. The cluster size determined was 5 which will categorise 14 neighbourhoods into 5 clusters finally. The clusters would encapsulate the venues of similar types which will help me identify the most appropriate neighbourhood for me to shift in London.

Results

Based on the k-means clustering, we found out the most common venues for each and every neighbourhood in Kingdom upon Thames. Below-mentioned is the image of the output:

```
neighborhoods_venues_sorted.head()
```

Out[68]:

	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	Bus Stop	Turkish Restaurant	Fish & Chips Shop	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Food
1	Canbury	Pub	Park	Café	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Fish & Chips Shop	Plaza	Hotel
2	Chessington	Construction & Landscaping	Food	Deli / Bodega	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant	Grocery Store
3	Hook	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Turkish Restaurant	Food	Department Store	Discount Store	Farmers Market	Fast Food Restaurant
4	Kingston Vale	Grocery Store	Sandwich Place	Bar	Soccer Field	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food

Based on the k-means clustering, the following clusters were finally devised:

Cluster 1

Cluster 1 includes the following neighbourhoods: Kingston Vale, Malden Rushett, Old Malden & Tolworth

Upon further inspection, we can understand that most common venues in these neighbourhoods are: Grocery Store, Pub & Bar.

```
In [77]: kut_merged[kut_merged['Cluster Labels'] == 0]
```

Out[77]:

	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 Mos Commo Venu
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	0	Grocery Store	Sandwich Place	Bar	Soccer Field	Department Store	Discount Store	Farmers Market	Fast Foo Restaura
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Grocery Store	Pub	Garden Center	Restaurant	Fish & Chips Shop	Deli / Bodega	Department Store	Discour Stor
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Construction & Landscaping	Grocery Store	Food	Train Station	Bagel Shop	Department Store	Golf Course	Gift Sho
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Restaurant	Pharmacy	Pizza Place	Train Station	Hotel	Indian Restaurant	Italia Restaura

Cluster 2

Cluster 2 includes the following neighbourhoods: Chessington

Upon further inspection, we can understand that most common venues in these neighbourhoods are: Construction & Landscaping and Food & Deli / Bodega.

```
In [78]: kut_merged[kut_merged['Cluster Labels'] == 1]
```

Out[78]:

	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 Con V
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Construction & Landscaping	Food	Deli / Bodega	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	F Resta

Cluster 3

Cluster 3 includes the following neighbourhoods: Hook & New Malden

Upon further inspection, we can understand that most common venues in these neighbourhoods are: Gym, Bakery, Supermarket & Gastropub

```
In [79]: kut_merged[kut_merged['Cluster Labels'] == 2]
```

Out[79]:

	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
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Turkish Restaurant	Food	Department Store	Discount Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Gym	Gastropub	Office	Sushi Restaurant	Supermarket	Bar	Korean Restaurant	Indian Restaurant

Cluster 4

Cluster 4 includes the following neighbourhoods: Berrylands & Motspur Park

Upon further inspection, we can understand that most common venues in these neighbourhoods are: Gym / Fitness Center , Park, Bus Stop and Fish & Chips Shop.

```
In [80]: kut_merged[kut_merged['Cluster Labels'] == 3]
```

Out[80]:

	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 Mo Comm Ven
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Bus Stop	Turkish Restaurant	Fish & Chips Shop	Department Store	Discount Store	Farmers Market	Fast Fo Restaura
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Fish & Chips Shop	Department Store	Discount Store	Farmers Market	Fast Fo Restaura

Cluster 5

Cluster 5 includes the following neighbourhoods: Canbury, Kingston upon Thames, Norbiton, Seething Wells & Surbiton

Upon further inspection, we can understand that most common venues in these neighbourhoods are: Pub, Café, Indian Restaurant & Coffee Shop.

```
In [81]: kut_merged[kut_merged['Cluster Labels'] == 4]
```

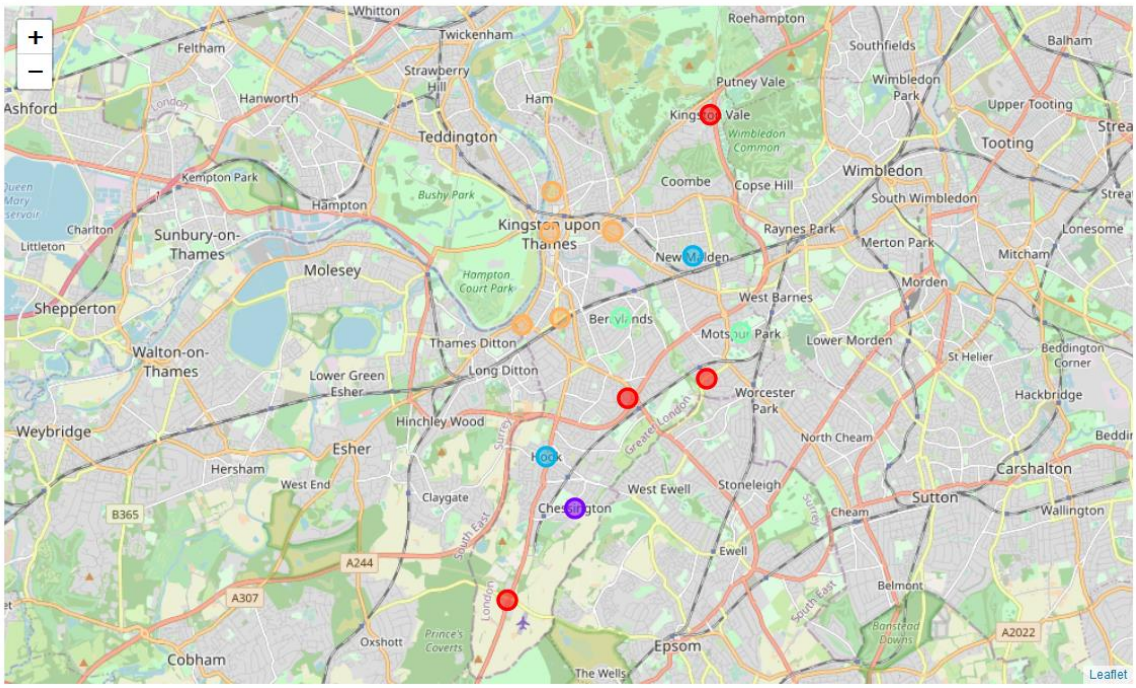
Out[81]:

	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	4	Pub	Park	Café	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Fish & Chips Shop
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	4	Café	Pub	Burger Joint	Sushi Restaurant	Coffee Shop	Turkish Restaurant	French Restaurant	Gym / Fitness Center
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	4	Indian Restaurant	Italian Restaurant	Food	Platform	Pub	Japanese Restaurant	Hotel	Coffee Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	4	Indian Restaurant	Coffee Shop	Pub	Café	Gym	Chinese Restaurant	Fast Food Restaurant	Fish & Chips Shop
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	4	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Breakfast Spot	French Restaurant	Train Station

Based on the above clusters, it has been visualised on the map mentioned below:

```
map_clusters
```

Out[76]:



Discussion

This project has helped me choose my preferred locality in London which has a safe neighbourhood along with venues around the neighbourhood which I would prefer. This can also be used by expats who would want to shift to London in a safe neighbourhood.

For a family relocating to London, I would recommend them shifting to Tolworth (Within Cluster 1) which has a grocery store nearby along with restaurant, pharmacy, pizza place and train station.

For a bachelor who is a health freak and likes to explore the city on weekends and chill, I would prefer New Malden (Within Cluster 2) which has a gym, gastro pub, office and super market nearby.

Conclusion

This project can help people relocating to London find the most suitable safe neighbourhood for them which will lead to easy accessibility and safety. Technology & data has contributed to make this decision easier for people which is why would recommend this project.