



The Battle of Neighborhoods

Applied Data Science Capstone - IBM

Tavishi Garg

TABLE OF CONTENTS

Problem Statement

Methodology.....	3
Exploratory Data Analysis.....	3
Statistical Analysis of Crime Data	3
Crime Rate Analysis – Highest Crime Rate Borough-wise Data.....	3
Crime Rate Analysis – Lowest Crime Rate Borough-wise Data	4
Neighbourhoods in Kingston upon Thames	5
Neighbourhood Analysis	6
Results.....	7
Cluster 1.....	7
Cluster 2.....	7
Cluster 3.....	8
Cluster 4.....	8
Cluster 5.....	8
Discussion	9
Conclusion.....	10

WHAT IS THE PROBLEM STATEMENT?

I want to move to London, but considering the rising cases of crime in the city. I am looking for safe localities before making any decision. I will shortlist the locality where crime rate is low and provides a safe environment for my stay.

The analysis will help other expats who are looking to relocate themselves. I will do the analysis by finding out the trending venues and the city host stops which are important to me while making the decision.

I am planning to use k-means clustering to shortlist top 10 common venues around the neighbourhood

DATA

I have found 3 references from internet to collect data for this project:

1. Data of crime in London
2. List of Boroughs in London
3. List of Localities in Royal Borough of Kingston upon Thames

Data of crime in London

The following link from Kaggle will be used to extract data: <https://www.kaggle.com/jboysen/london-crime>.

The following fields will be extracted:

- Lsoa_code
- Borough
- Major_category
- Minor_category,
- Value,
- Year
- Month

List of Boroughs in London

The following link from Wikipedia will be used to extract data:

https://en.wikipedia.org/wiki/List_of_London_boroughs.

The following fields will be extracted:

- Borough
- Inner
- Status
- Local authority
- Political control
- Headquarters
- Area (sq mi)
- Population (2013 est)
- Co-ordinates & Nr. in map

List of Localities in Royal Borough of Kingston upon Thames

After the Analysis is completed, the following link from Wikipedia will be used to extract data:

https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames

The following fields will be extracted:

- Neighbourhood
- Borough
- Latitude & longitude data.

Foursquare API

I will also use Foursquare API for the project

METHODOLOGY

EXPLORATORY DATA ANALYSIS

STATISTICAL ANALYSIS OF CRIME DATA

I was descriptive analysis using **Python**. I used describe() function to understand the data set. The data was extracted from Crime Data for London.

Through Descriptive Analysis I could extract mean, count, 1st quartile, 2nd quartile & 3rd quartile for each of the categories of crimes in London.

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

Out [43]:

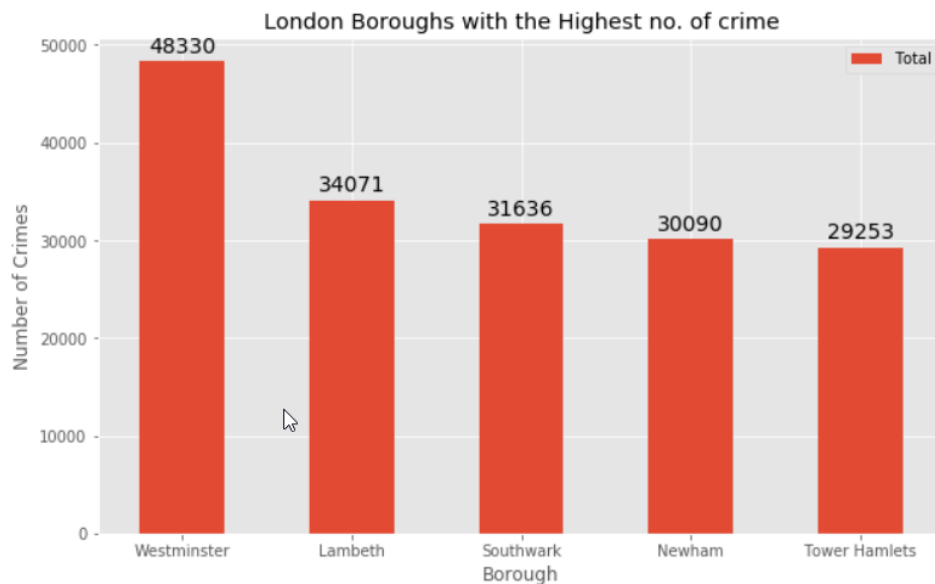
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
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

I have used column graph to display top 5 boroughs with highest crime rates.

I could conclude that Westminster has the highest crime rate in London, followed by Lambeth, Southwark, Newham and Tower Hamlets.

```
plt.show()
```

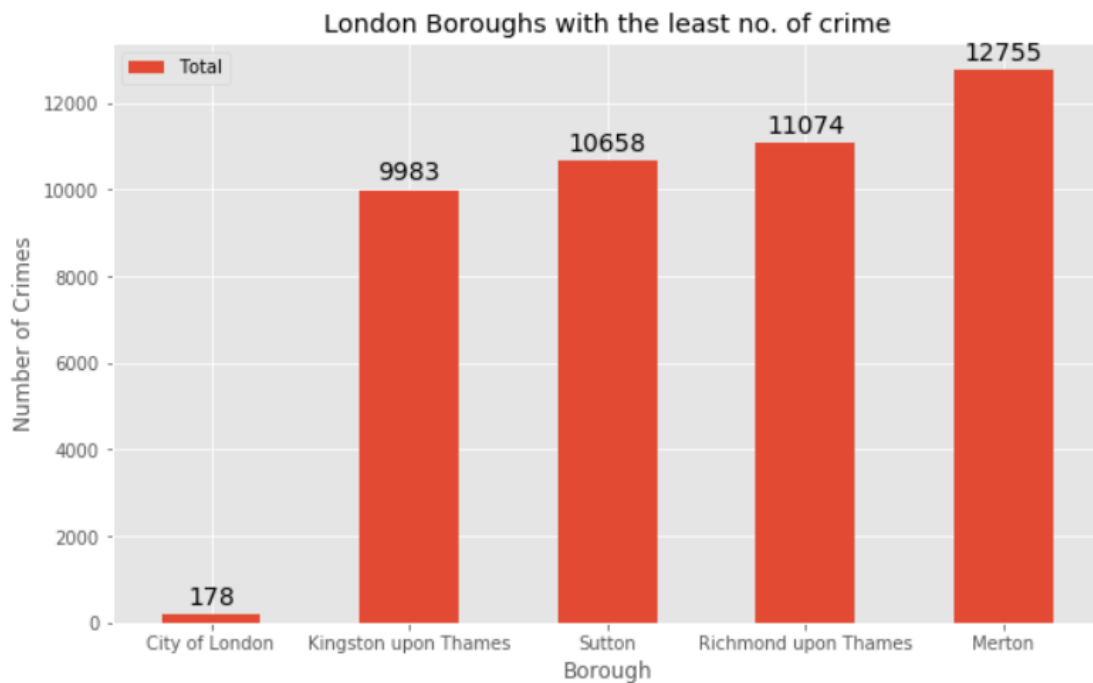


CRIME RATE ANALYSIS – LOWEST CRIME RATE BOROUGH-WISE DATA

To find the lowest crime rate I have used the column graph.

It was evident that city of London has lowest crime rate followed by by Kingdom upon Thames, Sutton Borough, Richmond upon Thames & Merton.

```
plt.show()
```



To bring uniformity to the analysis, I have decided not to consider City of London.

This is because, the city is less than 2 miles thus it is a very small area as compared to 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
```

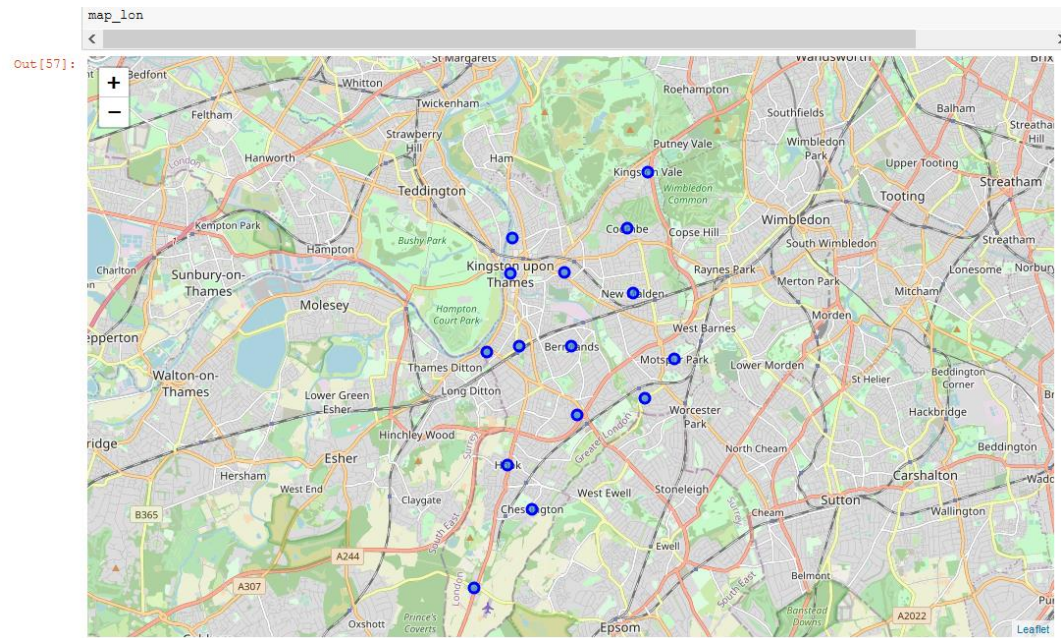
Out[50]:

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

NEIGHBOURHOODS IN KINGSTON UPON THAMES

The below map shows 15 neighborhoods within Kinston upon thames

Using Folium on Python, the map is prepared.



NEIGHBOURHOOD ANALYSIS

The below table shows further analysis of Kingston Upon Thames

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

The table shows neighbourhoods within 500m radius, using Foursquare API

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

RESULTS

Using k-means clustering, I could shortlist most common venues in each neighbourhood in Kingdom upon Thames

```
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

Post shortlisting, following clusters were made:

CLUSTER 1

Following neighbourhoods are studied together:

- Kingston Vale
- Malden Rushett,
- Old Malden
- Tolworth

Grocery Store, Pub & Ba are the most visited venues

```
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 Most Common Venue
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 Food Restaurant
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Grocery Store	Pub	Garden Center	Restaurant	Fish & Chips Shop	Deli / Bodega	Department Store	Discount Store
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 Shop
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Restaurant	Pharmacy	Pizza Place	Train Station	Hotel	Indian Restaurant	Italian Restaurant

CLUSTER 2

Following neighbourhoods are studied together:

- Chessington

Construction & Landscaping and Food & Deli / Bodega are the most visited venues


```
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 Most Common Venue
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	Fish & Chips Shop

CLUSTER 3

Following neighbourhoods are studied together:

- Hook
- New Malden

Gym, Bakery, Supermarket & Gastropub are the most visited venues

```
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

Following neighbourhoods are studied together:

- Berrylands
- Motspur Park

Fitness Centre, Park, Bus Stop and Fish & Chips Shop are the most visited venues

```
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 Most Common Venue
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 Food Restaurant
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 Food Restaurant

CLUSTER 5

Following neighbourhoods are studied together:

- Canbury
- Kingston upon Thames
- Norbiton,
- Seething Wells
- Surbiton

Pub, Café, Indian Restaurant & Coffee Shop are the most visited venues

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

Out[81]:

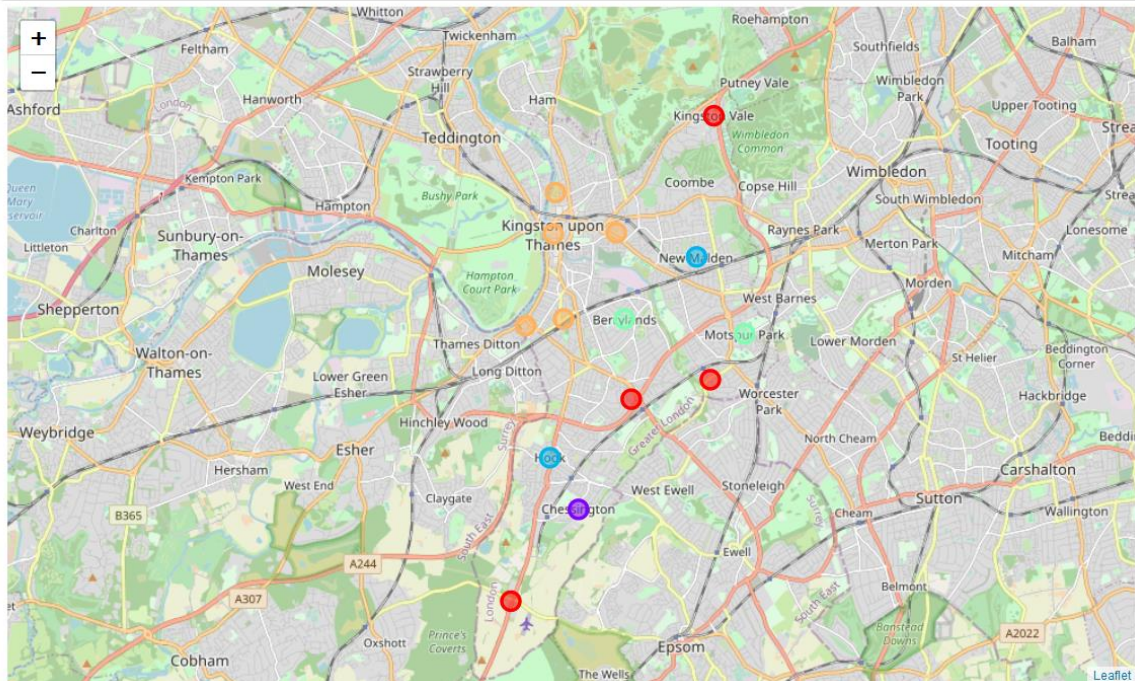
	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4	Pub	Park	Café	Supermarket	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

The following map is prepared:

map_clusters

<

Out[76]:



DISCUSSION

- With the help of this project, I was able to choose a safe neighbourhood in London
- The analysis can also be used by expats to make a similar discussion
- My Recommendations:
 - For Families
 - They can relocate themselves to Tolworth (within cluster-1)
 - This is because the neighborhood has a grocery store along with restaurants, pharmacy and a train station as well

- For Bachelors
 - They can relocate themselves to New Malden (within cluster-2)
 - This is because they are health freak and they can explore the city and visit gym, gastro pub, offices etc

CONCLUSION

With the help of this project, I was able to find a safe and suitable environment in London. This made the relocating process much easier. I would recommend this project to everyone.