Exploring London Boroughs

Demis Thanopoulos

15 February 2021

1 Introduction

Finding a place to live in no small task. In this report we will explore London neighborhoods and use crime data along data science machine learning to come up with selections to live.

2 Business Problem

When you want to move in to London the choice can be overwhelming as there are 32 boroughs with hundreds of neighborhoods to choose from. Before you decide on the location, one criterion is to avoid areas with high crime. Once you have selected a low crime area then you could examine the venues and facilities availability by each area that complements your lifestyle. From then on you have reached a level of analysis that has narrowed down your options from a hundred to less than 20!

In this notebook we will use data science along with Fourquare venue details to help guide a daunting task as such.

3 Data

For this project we will scrape London Boroughs from wikipedia List_of_London_boroughs and relevant homicide crime data from Crime_in_London in order to sort neighborhoods via homicides (serious crime).

We will also download a comprehensive crime list per area from data.london.gov to check the mean crime rates per area and complement the serious crimes analysis.

We will then decide which area to focus on and get a list of London neighborhoods from List_of_areas_of_London and link their geographical coordinates from geohack.toolforge.org

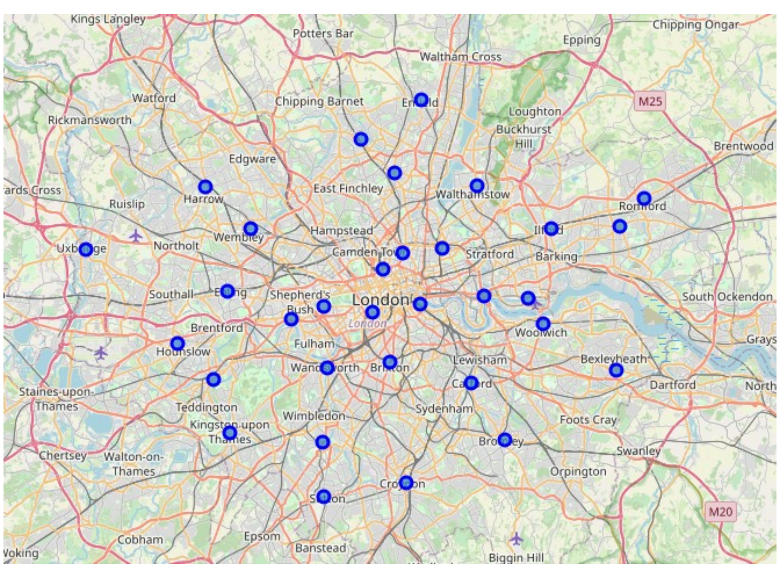
For the selected borough we will use the Fourquare API to get venues in each neighborhood and cluster them using the k means algorithm from the scikit learn python library. All visualizations will use folio.

4 Methodology

We scrape data from List_of_London_boroughs and get the data in a pandas dataframe.

	Borough	Latitude	Longitude
0	Barking and Dagenham	51.5607	0.1557
1	Barnet	51.6252	-0.1517
2	Bexley	51.4549	0.1505
3	Brent	51.5588	-0.2817
4	Bromley	51.4039	0.0198
5	Camden	51.5290	-0.1255
6	Croydon	51.3714	-0.0977
7	Ealing	51.5130	-0.3089
8	Enfield	51.6538	-0.0799
9	Greenwich	51.4892	0.0648
10	Hackney	51.5450	-0.0553
11	Hammersmith and Fulham	51.4927	-0.2339
12	Haringey	51.6000	-0.1119
13	Harrow	51.5898	-0.3346
14	Havering	51.5812	0.1837
15	Hillingdon	51.5441	-0.4760
16	Hounslow	51.4746	-0.3680
17	Islington	51.5416	-0.1022
18	Kensington and Chelsea	51.5020	-0.1947
19	Kingston upon Thames	51.4085	-0.3064
20	Lambeth	51.4607	-0.1163
21	Lewisham	51.4452	-0.0209
22	Merton	51.4014	-0.1958
23	Newham	51.5077	0.0469

Then we can visualize the boroughs on a map:



Then we get crime data from Crime_in_London and data.london.gov and we create 2 new sorted borough lists per crime. The first one has the number of homicides:

	Rank	Borough	Number of homicides 2000 to 2012
31	32	Richmond upon Thames	14
30	31	Kingston upon Thames	17
29	30	Kensington & Chelsea	23
28	29	Harrow	24
27	28	Sutton	25

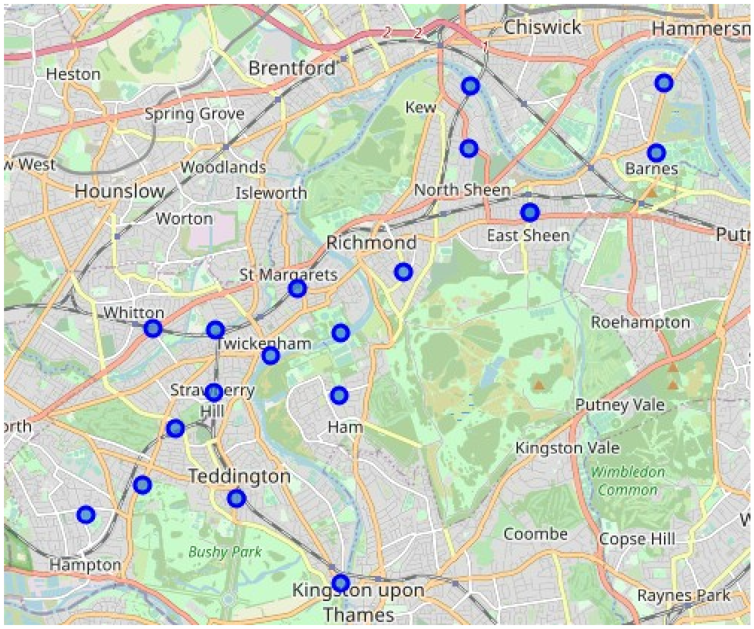
and the second has sum of the mean number of crimes per category:

	24month
Borough	
London Heathrow and London City Airports	214.7
Kingston upon Thames	1023.9
Richmond upon Thames	1049.0
Sutton	1123.7
Merton	1167.9

Richmond upon Thames looks like a good choice.

We next get the areas in Richmond from List_of_areas_of_London and create a visual on the map using folio.

	Borough	Neighborhood	Code	Latitude	Longitude
0	Richmond upon Thames	Barnes	TQ225765	51.474209	-0.237571
1	Richmond upon Thames	Castelnau	TQ226776	51.484074	-0.235750
2	Richmond upon Thames	East Sheen	TQ205755	51.465651	-0.266694
4	Richmond upon Thames	Mortlake	TQ205755	51.465651	-0.266694
6	Richmond upon Thames	Eel Pie Island	TQ164731	51.444938	-0.326478
7	Richmond upon Thames	Fulwell	TQ149719	51.434458	-0.348442
8	Richmond upon Thames	Ham	TQ175725	51.439318	-0.310856
9	Richmond upon Thames	Hampton	TQ135705	51.422157	-0.369021
10	Richmond upon Thames	Hampton Hill	TQ144710	51.426470	-0.355922
11	Richmond upon Thames	Hampton Wick	TQ176695	51.412335	-0.310413
12	Richmond upon Thames	Kew	TQ195775	51.483838	-0.280407
13	Richmond upon Thames	North Sheen	TQ195765	51.474850	-0.280745
14	Richmond upon Thames	Petersham	TQ175735	51.448306	-0.310525
15	Richmond upon Thames	Richmond	TQ185745	51.457085	-0.295807
16	Richmond upon Thames	St Margarets	TQ168742	51.454742	-0.320362
17	Richmond upon Thames	Strawberry Hill	TQ155725	51.439729	-0.339618
18	Richmond upon Thames	Teddington	TQ159708	51.424368	-0.334422
19	Richmond upon Thames	Twickenham	TQ155735	51.448717	-0.339292
20	Richmond upon Thames	Whitton	TQ145735	51.448920	-0.353676



Having decided on the borough and having the list of areas in it, we then use Foursquare API to locate the top venues per neighborhood.

We start looking at the venues by counting them per neighborhood:

	Neighborhood Latitude
Neighborhood	
Hampton Wick	74
Eel Pie Island	47
Teddington	30
North Sheen	28
Whitton	21
East Sheen	20
Mortlake	20
Barnes	18
Kew	16
St Margarets	15
Fulwell	14
Castelnau	11
Hampton Hill	9
Petersham	8
Strawberry Hill	8
Twickenham	6
Hampton	4
Ham	4
Richmond	4

We ctrate a table with mean values:

	Neighborhood	American Restaurant	Asian Restaurant	Australian Restaurant	Bakery	Bar	Beer Bar	Beer Garden	Beer Store	Boat or Ferry	...
0	Barnes	0.00	0.0	0.0	0.055556	0.000000	0.0	0.000000	0.00	0.0	...
1	Castelnau	0.00	0.0	0.0	0.000000	0.090909	0.0	0.000000	0.00	0.0	...
2	East Sheen	0.05	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.05	0.0	...
3	Eel Pie Island	0.00	0.0	0.0	0.021277	0.000000	0.0	0.021277	0.00	0.0	...
4	Fulwell	0.00	0.0	0.0	0.071429	0.000000	0.0	0.000000	0.00	0.0	...

and perform a one-hot analysis to prepare our data for clustering using the k-means algorithm:

	Neighborhood	American Restaurant	Asian Restaurant	Australian Restaurant	Bakery	Bar	Beer Bar	Beer Garden	Beer Store	Boat or Ferry	...
0	Barnes	0	0	0	0	0	0	0	0	0	...
1	Barnes	0	0	0	0	0	0	0	0	0	...
2	Barnes	0	0	0	0	0	0	0	0	0	...
3	Barnes	0	0	0	0	0	0	0	0	0	...
4	Barnes	0	0	0	0	0	0	0	0	0	...

Before clustering we create a table to list the top venues per neighborhood:

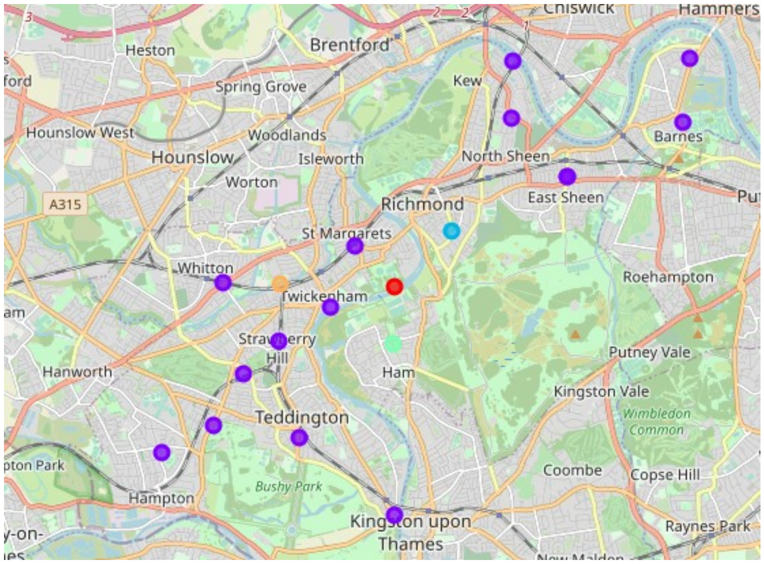
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barnes	Park	Sports Club	Movie Theater	Restaurant	Food & Drink Shop
1	Castelnau	Café	Indian Restaurant	Cosmetics Shop	French Restaurant	Chinese Restaurant
2	East Sheen	Coffee Shop	Pub	Grocery Store	Pizza Place	American Restaurant
3	Eel Pie Island	Coffee Shop	Pub	Italian Restaurant	Indian Restaurant	Pharmacy
4	Fulwell	Café	Chinese Restaurant	Golf Course	Pub	Gastropub

We can use the relevant amenities to help pick a neighborhood that better suits our lifestyle.

5 Results - Clustering using k-means

The final step in our data analysis comes by clustering the neighborhoods - venues table into 5 different clusters. The final clusters can be visualised on a map to help us locate the preferred cluster on the map.

	Borough	Neighborhood	Code	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Richmond upon Thames	Barnes	TQ225765	51.474209	-0.237571	1	Park	Sports Club	Movie Theater	Restaurant
1	Richmond upon Thames	Castelnau	TQ226776	51.484074	-0.235750	1	Café	Indian Restaurant	Cosmetics Shop	French Restaurant
2	Richmond upon Thames	East Sheen	TQ205755	51.465651	-0.266694	1	Coffee Shop	Pub	Grocery Store	Pizza Place
4	Richmond upon Thames	Mortlake	TQ205755	51.465651	-0.266694	1	Coffee Shop	Pub	Grocery Store	Pizza Place
6	Richmond upon Thames	Eel Pie Island	TQ164731	51.444938	-0.326478	1	Coffee Shop	Pub	Italian Restaurant	Indian Restaurant



6 Discussion

From the above analysis we ended up with 5 clusters of a London Borough with low rate crime we can select to live in. Our individual preferences and needs can be superimposed on the clusters's most common venues to make the final selection.

7 Conclusions

In this report we showcased how we can narrow down seemingly overwhelming tasks, like selecting a neighborhood in London, with the help of python. By getting relevant data from the internet such as areas of interest and crime data and then implementing data science methodology to clean, analyse and visualise series of data we can reach a point of insight to help informed based decisions.