

Moving to London which is the best London Borough to live in?

Ian Dixon - July 2020

Introduction

London is a large and vibrant city, with a population of around 9 million people. It is popular not only as the capital of the United Kingdom and the work opportunities it provides, but also because it provides great services and amenities.

London is a city with many coffee shops, great restaurants and place to visit, it is also a green city with parks and open spaces which make it attractive to people looking to live and work there. Of course, the downside of any large city is Crime, in 2018/19 there were a total of 912,000 crimes across London.

Problem Description

London is comprised of 32 boroughs and the City of London (not classed as a Borough), if someone was moving to London and didn't know much about the city, which Borough should they choose to search for a place to live?

Obviously, they would like to have the advantages which London brings such as great places to eat and drink and be able to get out and enjoy parks for recreation and exercise, but it is also to live in an area with low crime rate.

For this exercise the focus will be on identifying boroughs with three key amenities Restaurants, Pubs and Parks

Not everyone has unlimited funds to buy a house or apartment either, so knowing how much it costs to buy a place to live is also an important consideration.

Interest

Anyone moving to London would be interested to know the areas which have the amenities they are looking for, and what crime is like in the areas they are looking to move too.

Data Collection and Cleaning

Data

To undertake the analysis the following datasets were used

London Crime Data

London Crime data is available to download in a number of formats at https://data.london.gov.uk/dataset/recorded_crime_summary, multiple files are available which provide a breakdown of crimes at different geographical levels within London.

For the analysis, as we are looking at London Boroughs the MPS Borough Level Crime (most recent 24 months) data set was used.

The above file doesn't include the City of London (as it is not a Borough) but we can scrape this data from the City of London police website https://www.ukcrimestats.com/Police_Force/City_of_London_Police

London Borough Information

Wikipedia resource provides a list of all the London boroughs, which can be scraped at https://en.wikipedia.org/wiki/List_of_London_boroughs, this dataset includes the latitude and longitude of each borough, along with population estimates and the area in square miles of the borough.

House Price Information

House price information can be scraped from the UK Government website <https://www.gov.uk/government/publications/uk-house-price-index-england-march-2020/uk-house-price-index-england-march-2020>

Venue Data

Information on types and numbers of venues such as restaurants, pubs and parks withing a geographical area is available from <https://foursquare.com/>. The foursquare API provides functions to search for different categories of venues though a variety of methods such as venues within a set radius of a given latitude and longitude.

The Foursquare API is used to generate a list of venues for each London borough.

Geographical Data

A .json file providing the coordinates of the boundaries of the London boroughs is available at https://skgrange.github.io/www/data/london_boroughs.json this was used in the generation of map graphics

Data Preparation

Crime data was downloaded and cleaned, the total number of crimes for the year 2019 calculated for each London borough

London borough data was scraped from Wikipedia, and cleaned (removal of white spaces, reformatting of Latitude and longitude to a format supported by the Foursquare API)

House price data was scraped, cleaned and filtered

The three sources of data were then combined into one data frame for further analysis.

Foursquare venue data was downloaded, through the Foursquare API using the latitude and longitude co-ordinates provided by Wikipedia, with a radius of 2000 meters to give a list of venues by category for each borough.

Foursquare provides multiple identifications for types of restaurants e.g. Indian restaurant, Chinese restaurant etc, so the Venue data was recategorized to a higher level e.g. Restaurant

The venue data was then one-hot encoded and the mean frequency of each venue category for each borough calculated, the top ten venue categories per borough were calculated

Data Analysis

Folium (<https://python-visualization.github.io/folium/>) was used to generate a Choropleth map of the crime rate per 1000 population (fig. 1) for each of the London Boroughs.

Light Areas are low crime, while dark areas represent higher crime rates As expected the centre of London has the boroughs with the highest crime rates, while the suburbs particularly in the south west, and some boroughs in the east of London show lower crime rates.

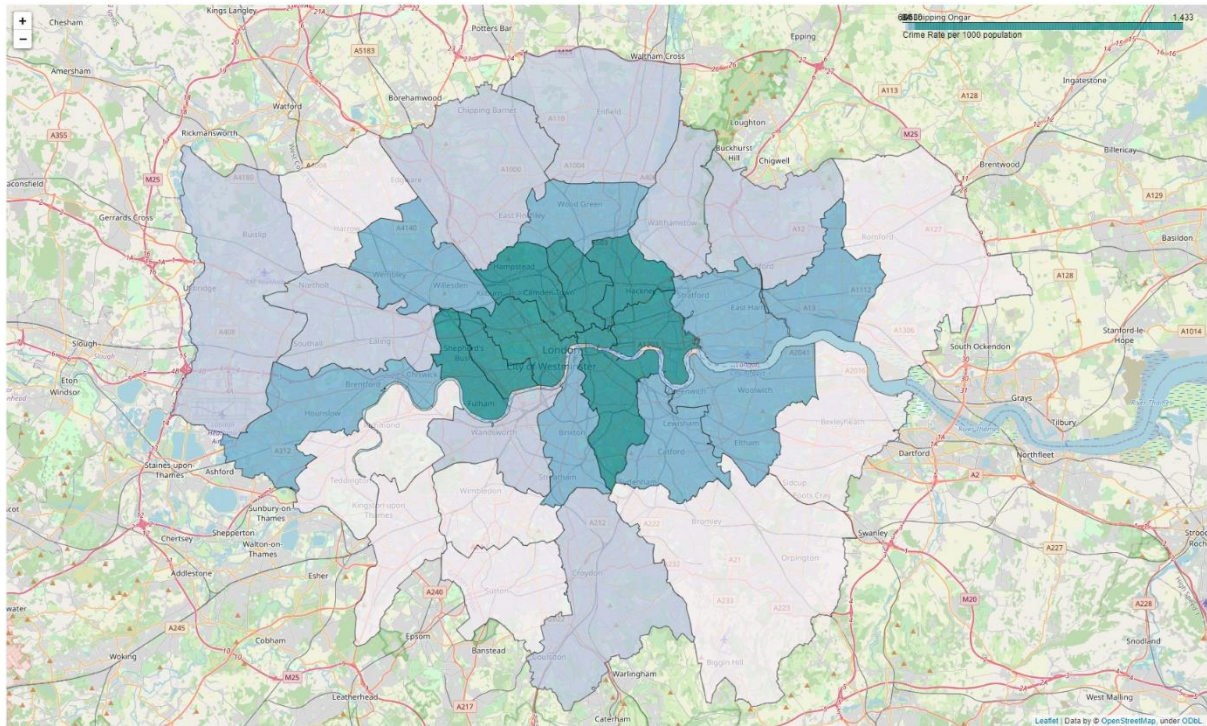


Fig. 1 Crime rate per 1000 population for London boroughs

Clustering Analysis

K Means clustering was utilized on the venues data to identify boroughs with similarities in venues, with the aim to identify a cluster with the target characteristics of restaurants pubs and parks. The elbow method was used to determine the optimal value of K.

Initial plot of the squared error cost (fig.2) produced a curve with no defined elbow, therefore additional visualization tools were used to determine the optimal value of K.

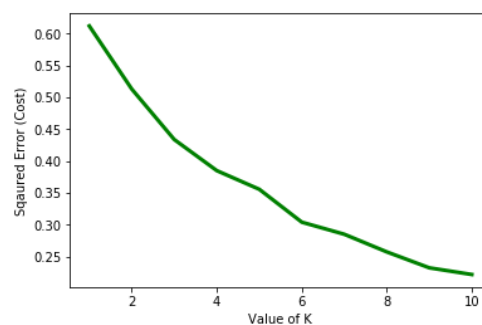


fig.2 elbow plot

Yellowbrick (<https://www.scikit-yb.org/en/latest/>) provides additional tools for visualisation for machine learning, their elbow method plot was utilized to determine the value of K (fig.3)

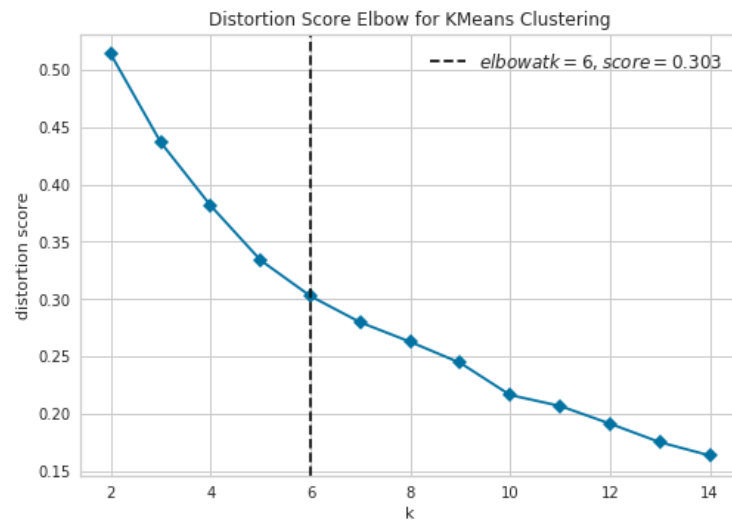


Fig. 3 Yellowbrick elbow method to determine value of k

K Means clustering was run on the venues data with a K of 6, the resulting clusters were then added to the dataset and plotted over the map of London Crime rates (fig. 4).

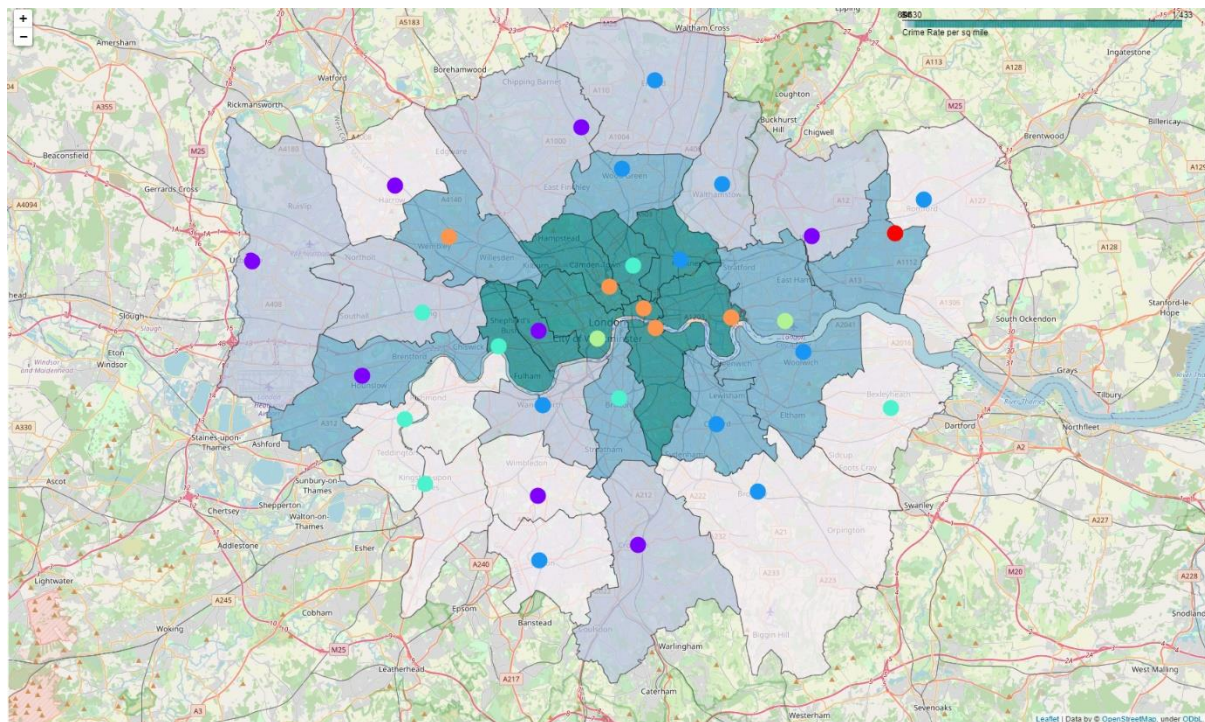


Fig. 4 Clusters plotted over London crime rate per 1000 population choropleth

Cluster data was reviewed to determine the main features of each cluster

Cluster 0:

	Borough	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 Most Common Venue
0	Barking and Dagenham	0	Grocery Store	Gas Station	Park	Supermarket	Gym / Fitness Center	Restaurant	Platform	Pool	Coffee Shop	Soccer Field

Cluster 1:

	Borough	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 Most Common Venue
1	Barnet	1	Restaurant	Coffee Shop	Grocery Store	Pub	Park	Café	Pharmacy	Hotel	Supermarket	Pizza Place
6	Croydon	1	Restaurant	Pub	Coffee Shop	Hotel	Clothing Store	Supermarket	Park	Café	Grocery Store	Bookstore
13	Harrow	1	Restaurant	Coffee Shop	Grocery Store	Sandwich Place	Café	Supermarket	Park	Bar	Pub	Gym
15	Hillingdon	1	Restaurant	Coffee Shop	Pub	Gym	Supermarket	Hotel	Gym / Fitness Center	Burger Joint	Sandwich Place	Pharmacy
16	Hounslow	1	Restaurant	Coffee Shop	Bus Stop	Clothing Store	Hotel	Supermarket	Sandwich Place	Convenience Store	Discount Store	Pharmacy
18	Kensington and Chelsea	1	Restaurant	Pub	Bar	Café	Bakery	Garden	Park	Supermarket	Science Museum	Burger Joint
22	Merton	1	Restaurant	Coffee Shop	Supermarket	Grocery Store	Park	Café	Clothing Store	Hotel	Pizza Place	Bar
24	Redbridge	1	Restaurant	Grocery Store	Supermarket	Coffee Shop	Clothing Store	Park	Bakery	Irish Pub	Department Store	Gym / Fitness Center

Cluster 2:

	Borough	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 Most Common Venue
4	Bromley	2	Restaurant	Pub	Clothing Store	Pizza Place	Coffee Shop	Gym / Fitness Center	Park	Sandwich Place	Gastropub	Department Store
8	Enfield	2	Restaurant	Pub	Coffee Shop	Supermarket	Pizza Place	Clothing Store	Grocery Store	Pharmacy	Train Station	Gym / Fitness Center
9	Greenwich	2	Restaurant	Grocery Store	Pub	Coffee Shop	Park	Bakery	Gym / Fitness Center	Hotel	Sandwich Place	Café
10	Hackney	2	Restaurant	Pub	Coffee Shop	Bakery	Café	Brewery	Park	Wine Shop	Bookstore	Pizza Place
12	Haringey	2	Restaurant	Pub	Café	Park	Grocery Store	Bar	Bakery	Coffee Shop	Food & Drink Shop	Pizza Place
14	Havering	2	Restaurant	Coffee Shop	Grocery Store	Pub	Clothing Store	Café	Shopping Mall	Supermarket	Furniture / Home Store	Multiplex
21	Lewisham	2	Coffee Shop	Restaurant	Pub	Café	Bar	Park	Supermarket	Theater	Fish & Chips Shop	Pizza Place
27	Sutton	2	Restaurant	Pub	Coffee Shop	Grocery Store	Café	Clothing Store	Supermarket	Park	Pizza Place	Sandwich Place
29	Waltham Forest	2	Restaurant	Pub	Café	Grocery Store	Coffee Shop	Supermarket	Pizza Place	Gym / Fitness Center	Hotel	Sandwich Place
30	Wandsworth	2	Restaurant	Pub	Park	Coffee Shop	Pizza Place	Bar	Café	Supermarket	Burger Joint	Grocery Store

Cluster 3:

	Borough	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 Most Common Venue
2	Bexley	3	Restaurant	Pub	Clothing Store	Coffee Shop	Supermarket	Grocery Store	Hotel	Pharmacy	Rugby Pitch	Garden
7	Ealing	3	Restaurant	Pub	Park	Coffee Shop	Hotel	Pizza Place	Café	Burger Joint	Sandwich Place	Gym / Fitness Center
11	Hammersmith and Fulham	3	Restaurant	Pub	Café	Coffee Shop	Bar	Park	Pizza Place	Gastropub	Plaza	Hotel
17	Islington	3	Restaurant	Pub	Park	Bar	Coffee Shop	Theater	Gastropub	Fish Market	Bakery	Pizza Place
19	Kingston upon Thames	3	Restaurant	Pub	Coffee Shop	Café	Burger Joint	Department Store	Park	Gastropub	Bookstore	Bar
20	Lambeth	3	Restaurant	Pub	Coffee Shop	Bar	Pizza Place	Market	Park	Brewery	Café	BBQ Joint
25	Richmond upon Thames	3	Restaurant	Pub	Café	Park	Coffee Shop	Rugby Stadium	Garden	Hotel	Historic Site	Gastropub

Cluster 4:

	Borough	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 Most Common Venue
23	Newham	4	Hotel	Restaurant	Coffee Shop	Gym / Fitness Center	Café	Park	Grocery Store	Light Rail Station	Sandwich Place	Bar
31	Westminster	4	Hotel	Restaurant	Plaza	Café	Clothing Store	Park	Boutique	Bar	Art Museum	Ice Cream Shop

Cluster 5:

	Borough	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 Most Common Venue
3	Brent	5	Restaurant	Coffee Shop	Clothing Store	Hotel	Sandwich Place	Bar	Grocery Store	Pizza Place	Sporting Goods Shop	Gym / Fitness Center
5	Camden	5	Restaurant	Coffee Shop	Bar	Hotel	Bookstore	Breakfast Spot	Pizza Place	Bakery	Plaza	Canal
26	Southwark	5	Restaurant	Coffee Shop	Bar	Hotel	Pub	Bakery	Garden	Brewery	Street Food Gathering	Gym / Fitness Center
28	Tower Hamlets	5	Restaurant	Coffee Shop	Bar	Hotel	Burger Joint	Park	Pub	Plaza	Gym / Fitness Center	Lounge
32	City of London	5	Restaurant	Bar	Coffee Shop	Hotel	Scenic Lookout	Gym / Fitness Center	Grocery Store	Burger Joint	Street Food Gathering	Bike Shop

Looking at the top 3 venues for each cluster, clusters 1,2,3,5 all have restaurants as their top venue, clusters 2 and 3 have pubs as their second top venues, cluster 3 has parks in the third top venues.

So, it appears that cluster 3 is meets our criteria of boroughs with restaurants, pubs and parks.

Conclusion

Based on the clustering analysis cluster 3 meets the criteria for areas with restaurants, pubs and parks.

The following table (fig.5) shows the Boroughs in cluster 3 and the associated Average House Price and Crime per 1000 population.

	Borough	Average House Price	Crime per 1000 pop
0	Richmond upon Thames	£ 633,481	68
1	Bexley	£ 331,679	76
2	Kingston upon Thames	£ 506,000	79
3	Ealing	£ 454,027	93
4	Lambeth	£ 525,487	115
5	Hammersmith and Fulham	£ 747,174	134
6	Islington	£ 632,270	141

Richmond upon Thames, Bexley, Kingston upon Thames and Ealing meet the criteria of low crime rate (less than 100 crimes per 1000 population in 2019) and Restaurants, Pubs and Parks.

Future

The analysis was done at a high-level e.g. total crime, Borough as the geographical area, and average house price. There is scope dig deeper into the data at a more local level below the Borough such as the electoral ward.

Crime data - the data sources provide different categories of crimes recorded and a breakdown by further geographical areas such as electoral ward. Therefore, more detailed analysis could be done to identify different types of crime and the frequency of those e.g. Violent crime etc.

House price data – this was just the average, but further data maybe available to segment this into the average price depending on the number of rooms in the property and if it is a house, or apartment etc.