

Battle of boroughs: where should one live in London?

Nerine Joewondo

March 1, 2021

Abstract

One of the biggest challenge of moving to London is to find a suitable accommodation. To begin their search, one must narrow down from the 32 boroughs of London, which span 1,572 km² in area, to a list of few boroughs they are interested to explore further. Generally, factors contributing to one's decision on accommodation are: rental price, safety, and types of venues nearby. In this study, we use K- mean clustering to group 33 boroughs of London into four distinct groups based on venues. Additionally, we overlay this information on top of (i) current average rental price of a 1-bedroom flat and (ii) crime population.

1 Introduction

London is the second most populous city in Europe. In 2018/2019 approximately 450,000 people migrated to London [1]. One of the biggest challenge of moving to London is to find a suitable accommodation. To begin their search, one must narrow down from the 33 boroughs of London, which span 1,572 km² in area [2], to a list of few boroughs they are interested to explore further.

As there could be varying reasons for one to move to London, there could also be varying factors affecting one's decision to choose a borough over another. In this study, we use K-means clustering to group London boroughs into K groups based on the most common venue type. Furthermore, this information is overlayed on top of a (i) heatmap plot of crime prevalence and (ii) chloropleth plot of average (private) rental by borough. The two resulting plots would allow one to quickly narrow down their search into a few boroughs based on their preferences on rental price, venues, and safety tolerance.

2 Data

The dataset required to gain insights on the problem described are:

- List of boroughs in London [2] is scraped to extract coordinate and population data for each borough. London consists of 33 boroughs.
- Top 500 venues in each borough will be obtained via the [Foursquare API](#). K-means clustering will be used to produce venue-based grouping of London boroughs.
- Crime statistics by borough in CSV format is downloaded from London Data Store [3]. The dataset contains records of crime based on its type and location for the past 24 months. This data is visualized on a heatmap plot along with the venue-based cluster plotted as a point on the map.
- Average private rent of a 1 bedroom flat by borough is obtained by scraping [rightmove website](#) using rightmove_webscraper API.

3 Methodology

3.1 Coordinates

Beautiful Soup package is used to scrape tables [2] listing boroughs of London. Along with the borough name, coordinates and population of each borough are extracted.

3.2 Venues from Foursquare

For each borough, 500 top venues are obtained using the Foursquare API. The resulting dataset contains a list of 1507 venues, which are categorized into:

- Arts & Entertainment

	Borough	Area	Population	Coordinates	Borough	Area	Population	Latitude	Longitude	
0	Barking and Dagenham	13.93	212906	.mw-parser-output .geo-default,.mw-parser-output...	0	Barking and Dagenham	13.93	212906	51.5607	0.1557
1	Barnet	33.49	395896	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W /...	1	Barnet	33.49	395896	51.6252	-0.1517
2	Bexley	23.38	248287	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E /...	2	Bexley	23.38	248287	51.4549	0.1505
3	Brent	16.70	329771	51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W /...	3	Brent	16.70	329771	51.5588	-0.2817
4	Bromley	57.97	332336	51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E /...	4	Bromley	57.97	332336	51.4039	0.0198
					5	Camden	8.40	270029	51.5290	-0.1255

(a)

(b)

Figure 1: (a) Data frame constructed by scrapping Wikipedia table. (b) Data frame after cleaning.

	Borough Total Crime									
	MajorText	MinorText	LookUp_BoroughName	201901	201902	201903	201904	201905	201906	
0	Arson and Criminal Damage	Arson	Barking and Dagenham	5	2	5	5	11	3	39407
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	97	127	138	130	140	113	58917
2	Burglary - Business and Community	Burglary -	Barking and Dagenham	45	24	29	27	21	27	33637
3	Burglary	Burglary - Residential	Barking and Dagenham	114	107	99	96	114	96	59123
4	Drug Offences	Drug Trafficking	Barking and Dagenham	6	2	6	5	9	6	47177
										69364

(a)

(b)

Figure 2: (a) Data frame constructed by reading the CSV file on recorded crime. (b) Data frame after cleaning.

- College & University
- Events
- Food
- Nightlife
- Outdoor & Recreation
- Professional
- Shop & Services
- Travel & Transport

For each borough, the top 5 venues are established. This dataset is binarized using one-hot encoding method before processing it using K-means clustering.

3.3 K-means clustering

The K-means is an unsupervised machine learning algorithm. It partitions a set of N samples X into K disjoint clusters C . Each cluster is described by the mean μ_j of the samples within it, or commonly called "centroids".

The objective of the K-means algorithm is to choose centroids that minimise the inertia, which is defined as:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2) \quad (1)$$

K-means is implemented to partition London boroughs into K clusters based on existing venues in each borough. An "elbow" curve is plotted to determine the optimal number of cluster.

3.4 Crime data

Crime statistics by borough in CSV format is downloaded from London Data Store [3]. The dataset contains records of crime based on its type and location for the past 24 months. A new data frame is constructed by calculating the

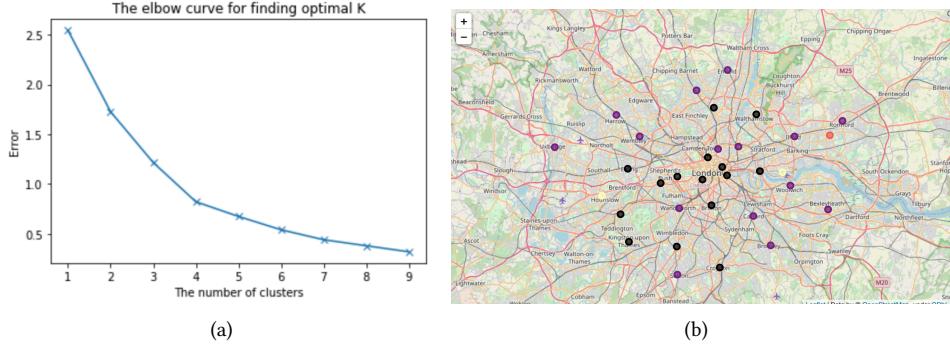


Figure 3: (a) Elbow method to determine the optimal number of clusters. (b) Visualization of clusters (Cluster 0: black, Cluster 1: purple, Cluster 2: pink/orange, Cluster 3: yellow).

cumulative crime for each borough for 24 months. A normalized crime index, I is calculated by dividing total crime by population:

$$I = \frac{\text{crime}_{\text{total}}}{\text{population}_{\text{total}}} \quad (2)$$

3.5 Rent offerings

To get an idea of rent price, we scrape the [rightmove website](#) using rightmove_webscraper API, filtering for offers of '1 bedroom flat within the greater London area'. As of today (28 February 2021), there are 1050 offers resulting from this query. From the query, we extract the postcode and price for each property. There are many instances of 'NaN' for the postcode data, therefore the data frame is cleaned to remove 'NaN' value before processing.

4 Result

The elbow method indicates that the optimal number of cluster is 4. Folium package is used to visualize the clustering results on a map of London (figure fig:first). The resulting clusters are listed below. Cluster 0 consists of 3 boroughs, Cluster 1 of 15 boroughs, Cluster 2 of 14 boroughs, and Cluster 3 of 1 borough.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Camden	Food	Shop & Service	Travel & Transport	Nightlife Spot	Outdoors & Recreation
6	Croydon	Food	Nightlife Spot	Shop & Service	Outdoors & Recreation	Arts & Entertainment
7	Ealing	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Travel & Transport
11	Hammersmith and Fulham	Food	Shop & Service	Nightlife Spot	Travel & Transport	Outdoors & Recreation
12	Haringey	Food	Outdoors & Recreation	Shop & Service	Nightlife Spot	Travel & Transport
18	Kensington and Chelsea	Food	Shop & Service	Outdoors & Recreation	Travel & Transport	Arts & Entertainment
19	Kingston upon Thames	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Travel & Transport
20	Lambeth	Food	Nightlife Spot	Shop & Service	Outdoors & Recreation	Arts & Entertainment
22	Merton	Food	Shop & Service	Outdoors & Recreation	Travel & Transport	Professional & Other Places
25	Richmond upon Thames	Food	Shop & Service	Nightlife Spot	Travel & Transport	Outdoors & Recreation
26	Southwark	Food	Nightlife Spot	Outdoors & Recreation	Arts & Entertainment	Shop & Service
28	Tower Hamlets	Food	Travel & Transport	Shop & Service	Outdoors & Recreation	Arts & Entertainment
29	Waltham Forest	Food	Outdoors & Recreation	Nightlife Spot	Arts & Entertainment	Shop & Service
31	Westminster	Food	Travel & Transport	Shop & Service	Nightlife Spot	Arts & Entertainment
32	City of London	Food	Outdoors & Recreation	Arts & Entertainment	Shop & Service	Nightlife Spot

Figure 4: List of Cluster 0: shopper's paradise

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Barnet	Shop & Service	Travel & Transport	Food	Professional & Other Places	Outdoors & Recreation
2	Bexley	Shop & Service	Food	Nightlife Spot	Travel & Transport	Outdoors & Recreation
3	Brent	Food	Shop & Service	Travel & Transport	Nightlife Spot	Outdoors & Recreation
4	Bromley	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Travel & Transport
8	Enfield	Shop & Service	Food	Nightlife Spot	Travel & Transport	Arts & Entertainment
9	Greenwich	Food	Shop & Service	Travel & Transport	Nightlife Spot	Outdoors & Recreation
10	Hackney	Food	Shop & Service	Nightlife Spot	Travel & Transport	Outdoors & Recreation
13	Harrow	Food	Shop & Service	Travel & Transport	Arts & Entertainment	Professional & Other Places
14	Havering	Shop & Service	Food	Nightlife Spot	Travel & Transport	Outdoors & Recreation
15	Hillingdon	Shop & Service	Food	Nightlife Spot	Travel & Transport	Outdoors & Recreation
17	Islington	Food	Shop & Service	Nightlife Spot	Arts & Entertainment	Outdoors & Recreation
21	Lewisham	Food	Shop & Service	Travel & Transport	Nightlife Spot	Arts & Entertainment
24	Redbridge	Shop & Service	Food	Travel & Transport	Arts & Entertainment	Outdoors & Recreation
27	Sutton	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Travel & Transport
30	Wandsworth	Food	Shop & Service	Nightlife Spot	Travel & Transport	Outdoors & Recreation

Figure 5: List of Cluster 1: nightlife focused.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	Outdoors & Recreation	Travel & Transport	Shop & Service	Professional & Other Places	Nightlife Spot

Figure 6: List of Cluster 2: outdoorsy.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
16	Hounslow	Travel & Transport	Outdoors & Recreation	Food	Shop & Service	Professional & Other Places
23	Newham	Travel & Transport	Food	Shop & Service	Outdoors & Recreation	Professional & Other Places

Figure 7: List of Cluster 3: frequent traveler

5 Discussion

Upon examining the venues existing in London, one would notice that Food related venues are commonly found in most of the boroughs of London. In fact, the most common venue of both Cluster 0 and 1 is mostly Food related venues. Generally, Cluster 0 and 1 are composed of similar types of venues. Notably, the 2nd most common venue of Cluster 1 is mostly Nightlife spot while of Cluster 0 is mostly Shop & Service.

There are two airports in London: Heathrow Airport and London City Airport, which are located in Hounslow and Newham respectively. Rightfully, the algorithm includes these two boroughs into Cluster 3. Cluster 2 consists of only Barking and Dagenham, which has mostly outdoors and recreation venues.

To summarize, most London boroughs consist of a similar blend of venues, with the exception of Barnet, Hounslow, Newham and Barking & Dagenham. For the frequent travellers, living in one of the boroughs in Cluster 3 (Hounslow, Newham) might be beneficial due to short commute to the airport. An outdoorsy person might be suited to live in Barking and Dagenham. Otherwise, the one's nightlife preference may be the determining factor to choose between the boroughs in Cluster 0 and 1.

A chloropleth map is chosen to visualize the crime instances per population for the past 24 months (figure fig:second (a)). The highest crime instances is recorded in Westminster, which is plotted in blue. Notably, the trend of crime occurrence diminishes radially outward from Westminster, which may correspond well to the TfL travel zones. Higher crime prevalence is observed in highly touristy area, i.e Westminster, Kensington & Chelsea, City of London etc. This may be correlated to crimes related to tourism, such as theft, which may not be relevant to the experience of living in the area. Future works should group crime instances based on its relevance.

The average rent of 1 bedroom flat for each borough is plotted on a heat map. The boroughs with the highest average rent are: Westminster, Kensington & Chelsea, and City of London. Similarly, this may correspond well to

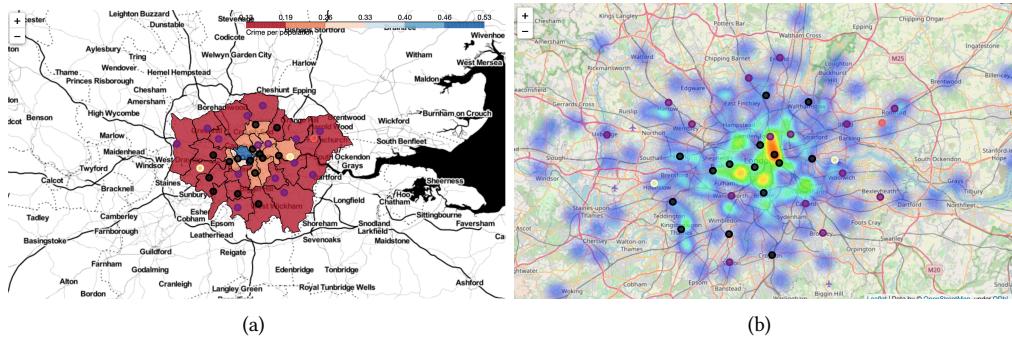


Figure 8: (a) Chloropleth map of crime instances per population for the past 24 months. (b) Heat map of average rental price of a 1 bedroom flat currently being offered on Rightmove.

the TfL travel zones.

6 Conclusion

K-means algorithm is implemented to partition boroughs of London into 4 distinct cluster based on venues. Most boroughs of London offer a similar blend of venues with the exception of Hounslow, Newham and Barking & Dagenham. Barnet, Hounslow, and Newham is distinctly different as the most common venue is travel related venue, while for Barking & Dagenham is outdoor and recreation venue. The rest of the boroughs of London fall into Cluster 0 and 1, which are similar in terms of the most common venue being food related venue. The second most common venue differentiates cluster 0 and 1 (shop & service and nightlife spot, respectively).

Boroughs in London are relatively similar in terms of venues. However, difference in crime prevalence and average rental price can be observed. Particularly, Westminster is observed to have the highest criminal prevalence and among the highest rental price. Both criminal prevalence and rental price seems to have similar trend to London's travel zone, and is observed to correlate to touristic areas of London. For this reason, future works are recommended to explore the correlation between crime and tourism.