Finding the best area to open a new leisure centre in London

Luca Montalto Giampaoli

10 April 2021

Introduction and description of the problem

London is one of the world's most important global cities. It exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation. London is also the most populous region, urban zone and metropolitan area in the United Kingdom.

For these reasons entrepreneurs have been investing in various business activities spanning from the hospitality to wellbeing sectors, from the entertainment to cultural venues. One of the big concerns for entrepreneurs when deciding to open a new business would be finding the right area.

Therefore in this project, **I intend to explore different Inner boroughs of London and identify the best area to find new business opportunities. This research is expected to benefit investors who want to open a new leisure centre in a profitable location.**

Data Acquisition and Data processing

The following datasets were used:

a. Greenspace Information for Greater London

- Percentage and count of residential households with access to at least one open space
- source: London Datastore
- url: https://data.london.gov.uk/dataset/access-public-open-space-and-nature-ward

b. List of London Boroughs

- List of the boroughs in greater London
- source: Wikipedia
- url: https://en.wikipedia.org/wiki/List_of_London_boroughs
- Method: The list of the Boroughs will be taken from Wikipedia through web scraping by BeautifulSoup
- The complete dataset includes both the "Outer boroughs" and the "Inner boroughs" of London, but only the "Inner" ones was considered

c. FourSquare

- A FourSquare Developer token was generated and applied to identify the most popular venues in each individual area

Import required libraries

A number of libraries have been imported to allow to handle various functions such as data vectorisation, chart visualisation and clustering.

2.2 Download and explore the dataset that describes the accessibility to parks in London

The analysis was based on the idea to combine the information available on the accessibility to parks and green areas across London and compare this to the cluster analysis in order to identify the boroughs and highlight the most popular type of venues.

The dataset containing the accessibility to the parks is available on the London Datastore (data.london.gov.uk). This dataset includes the boroughs across the Greater London including both Inner and Outer Boroughs. After a cleaning process, the final dataset includes the unique boroughs and the mean of the access to green areas as a percentage.

We are interested in analysing only the Inner Boroughs, therefore the Outers ones will be removed along the process.

We ended up with a data frame of 34 rows and 2 columns in total.

	Borough	Access to nature %
0	Barking and Dagenham	71.705882
1	Barnet	82.333333
2	Bexley	83.190476
3	Brent	82.095238
4	Bromley	82.454545
5	Camden	70.555556
6	City of London	0.000000
7	Croydon	70.333333
8	Ealing	74.826087
9	Enfield	54.047619
10	Greenwich	87.117647
11	Hackney	74.105263
12	Hammersmith and Fulham	64.125000
13	Haringey	82.210526
14	Harrow	61.428571
15	Havering	58.944444
16	Hillingdon	84.454545
17	Hounslow	77.250000
18	Islington	67.562500
19	Kensington and Chelsea	90.444444
20	Kingston upon Thames	67.875000
21	Lambeth	63.333333
22	Lewisham	86.166667
23	London	74.000000
24	Merton	86.450000
25	Newham	60.600000
26	Redbridge	82.000000
27	Richmond upon Thames	88.777778
28	Southwark	70.428571
29	Sutton	70.500000
30	Tower Hamlets	64.941176
31	Waltham Forest	57.700000
32	Wandsworth	90.550000
33	Westminster	73.150000

2.3 Download and explore the dataset with the list of boroughs in Greater London

The data set containing the list of boroughs across the Greater London was downloaded from Wikipedia using the Web scraping technique by BeautifulSoup library.
The data set was cleaned to end up with only the Inner boroughs. The coordinates (latitude and longitude) of each borough were also available in the same dataset therefore it wasn't necessary to use geographical API to convert locations into coordinates.

The final data frame has 12 unique rows and 3 columns.

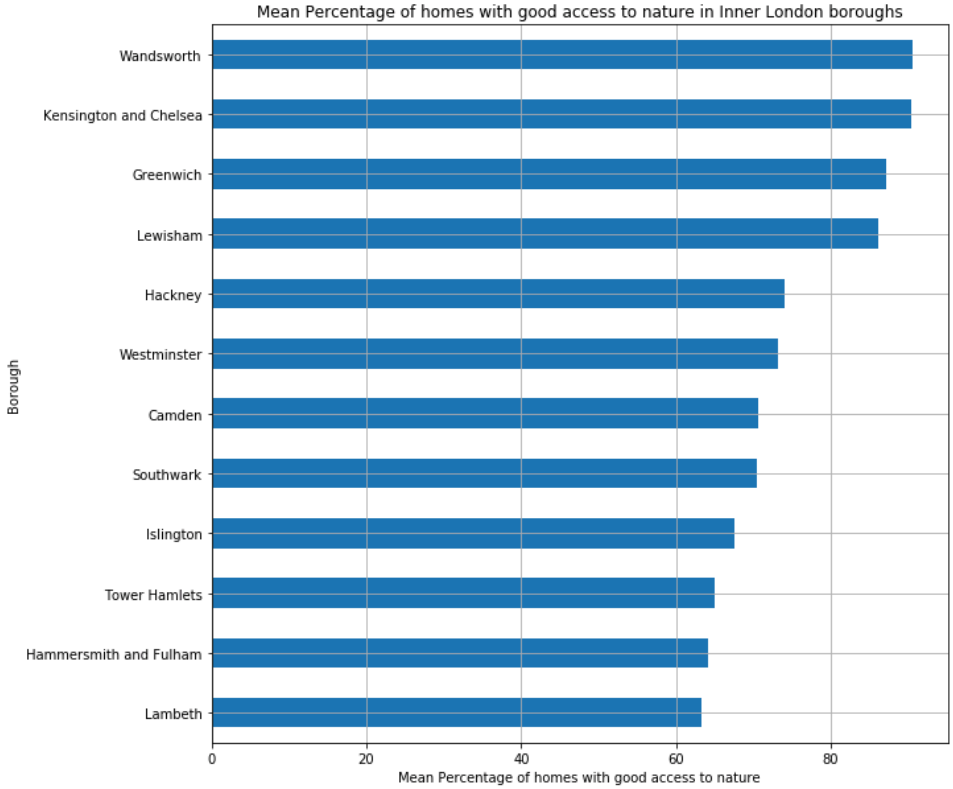
	Borough	Latitude	Longitude
0	Camden	51.5290	-0.1255
1	Greenwich	51.4892	0.0648
2	Hackney	51.5450	-0.0553
3	Hammersmith and Fulham	51.4927	-0.2339
4	Islington	51.5416	-0.1022
5	Kensington and Chelsea	51.5020	-0.1947
6	Lambeth	51.4607	-0.1163
7	Lewisham	51.4452	-0.0209
8	Southwark	51.5035	-0.0804
9	Tower Hamlets	51.5099	-0.0059
10	Wandsworth	51.4567	-0.1910
11	Westminster	51.4973	-0.1372

2.4 Combine Access to parks and London boroughs dataset¶

The two data frames previously cleaned have been combined to provide for each borough its coordinates along with the percentage of access to the green areas.

	Borough	Access to nature %	Latitude	Longitude
5	Camden	70.555556	51.5290	-0.1255
10	Greenwich	87.117647	51.4892	0.0648
11	Hackney	74.105263	51.5450	-0.0553
12	Hammersmith and Fulham	64.125000	51.4927	-0.2339
18	Islington	67.562500	51.5416	-0.1022
19	Kensington and Chelsea	90.444444	51.5020	-0.1947
21	Lambeth	63.333333	51.4607	-0.1163
22	Lewisham	86.166667	51.4452	-0.0209
28	Southwark	70.428571	51.5035	-0.0804
30	Tower Hamlets	64.941176	51.5099	-0.0059
32	Wandsworth	90.550000	51.4567	-0.1910
33	Westminster	73.150000	51.4973	-0.1372

This allowed to come up with a chart that shows the Mean Percentage of homes with good access to nature in Inner London boroughs.



The chart suggests that Lambeth has the smallest percentage of homes with good access to nature, followed by Hammersmith and Fulham (H&F), Tower Hamlets (TH), and Islington. Therefore our analysis will focus over those four boroughs.

2.5 Exploration of the boroughs and their segmentation

The Foursquare API was used to carry out the segmentation of the venues across the different boroughs identified by the coordinates.

To optimise the analysis the number of popular venues was limited up to ten.

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	10th M Comm Ve
Camden	51.5290	-0.1255	2	Coffee Shop	Café	Hotel	Pub	Italian Restaurant	Pizza Place	Breakfast Spot	Train Station	Burger Joint	Fa Restau
Greenwich	51.4892	0.0648	1	Coffee Shop	Clothing Store	Supermarket	Pub	Pharmacy	Hotel	Grocery Store	Fast Food Restaurant	Plaza	Sand F
Hackney	51.5450	-0.0553	2	Pub	Coffee Shop	Cocktail Bar	Bakery	Café	Brewery	Park	Clothing Store	Organic Grocery	H
Hammersmith and Fulham	51.4927	-0.2339	2	Pub	Café	Indian Restaurant	Italian Restaurant	Gastropub	Hotel	Coffee Shop	Clothing Store	Chinese Restaurant	Bus S
Islington	51.5416	-0.1022	0	Pub	Bakery	Ice Cream Shop	Park	Boutique	Burger Joint	Cocktail Bar	Turkish Restaurant	Music Venue	The
Kensington and Chelsea	51.5020	-0.1947	2	Bakery	French Restaurant	Juice Bar	Café	Hotel	Burger Joint	Italian Restaurant	Spa	Restaurant	
Lambeth	51.4607	-0.1163	2	Caribbean Restaurant	Market	Pub	Coffee Shop	Gym / Fitness Center	Beer Bar	Pizza Place	Nightclub	Grocery Store	Sand F
Lewisham	51.4452	-0.0209	4	Supermarket	Grocery Store	Coffee Shop	Platform	Train Station	Italian Restaurant	Pizza Place	Bakery	Shopping Mall	Bus S
Southwark	51.5035	-0.0804	2	Coffee Shop	Pub	Bar	Scenic Lookout	Cocktail Bar	Restaurant	French Restaurant	Hotel	English Restaurant	It Restau
Tower Hamlets	51.5099	-0.0059	3	Hotel	Sandwich Place	Coffee Shop	Italian Restaurant	Fried Chicken Joint	Light Rail Station	Café	Pub	Convenience Store	Gro S
Wandsworth	51.4567	-0.1910	1	Coffee Shop	Pub	Clothing Store	Breakfast Spot	Supermarket	Gym / Fitness Center	Indian Restaurant	Chaat Place	Gift Shop	Bus S
Westminster	51.4973	-0.1372	3	Hotel	Coffee Shop	Sandwich Place	Italian Restaurant	Theater	Sushi Restaurant	Pub	Gym / Fitness Center	Hotel Bar	Juice

2.6 Cluster Analysis

Clustering is a set of techniques used to partition data into groups, or clusters. Clusters are loosely defined as groups of data objects that are more similar to other objects in their cluster than they are to data objects in other clusters. Conventional k-means requires only a few steps. The first step is to randomly select k centroids, where k is equal to the number of clusters you choose. Centroids are data points representing the centre of a cluster.

The main element of the algorithm works by a two-step process called expectation-maximization. The expectation step assigns each data point to its nearest centroid. Then, the maximization step computes the mean of all the points for each cluster and sets the new centroid.

The quality of the cluster assignments is determined by computing the sum of the squared error (SSE) after the centroids converge, or match the previous iteration's assignment. The SSE is defined as the sum of the squared Euclidean distances of each point to its closest centroid. Since this is a measure of error, the objective of k-means is to try to minimize this value.

The quality of the cluster assignments is determined by computing the sum of the squared error (SSE) after the centroids converge, or match the previous iteration's assignment. The SSE is defined as the sum of the squared Euclidean distances of each point to its closest centroid. Since this is a measure of error, the objective of k-means is to try to minimize this value.

We identified five clusters hence k-means equal to five.

Cluster 0

```
# Cluster 0: Lively area
london_merged.loc[london_merged['Cluster Labels'] == 0, london_merged.columns[[0] + list(range(3, london_merged.shape[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
4	Islington	0	Pub	Bakery	Ice Cream Shop	Park	Boutique	Burger Joint	Cocktail Bar	Turkish Restaurant	Music Venue	Theater

By looking at the results, the cluster 0 identifies the lively area in London made of mainly pubs, bakery, ice cream shops, etc

Cluster 1

```
# Cluster 1 Quiet area
london_merged.loc[london_merged['Cluster Labels'] == 1, london_merged.columns[[0] + list(range(3, london_merged.shape[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	Greenwich	1	Coffee Shop	Clothing Store	Supermarket	Pub	Pharmacy	Hotel	Grocery Store	Fast Food Restaurant	Plaza	Sandwich Place
10	Wandsworth	1	Coffee Shop	Pub	Clothing Store	Breakfast Spot	Supermarket	Gym / Fitness Center	Indian Restaurant	Chaat Place	Gift Shop	Bus Stop

By the type of venues, Cluster 1 identifies quiet areas with coffee shops and stores.

Cluster 2

```
# Cluster 2 busy area
london_merged.loc[london_merged['Cluster Labels'] == 2, london_merged.columns[[0] + list(range(3, london_merged.shape[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
0	Camden	2	Coffee Shop	Café	Hotel	Pub	Italian Restaurant	Pizza Place	Breakfast Spot	Train Station	Burger Joint	Falafel Restaurant
2	Hackney	2	Pub	Coffee Shop	Cocktail Bar	Bakery	Café	Brewery	Park	Clothing Store	Organic Grocery	Hotel
3	Hammersmith and Fulham	2	Pub	Café	Indian Restaurant	Italian Restaurant	Gastropub	Hotel	Coffee Shop	Clothing Store	Chinese Restaurant	Bus Stop
5	Kensington and Chelsea	2	Bakery	French Restaurant	Juice Bar	Café	Hotel	Burger Joint	Italian Restaurant	Spa	Restaurant	Pub
6	Lambeth	2	Caribbean Restaurant	Market	Pub	Coffee Shop	Gym / Fitness Center	Beer Bar	Pizza Place	Nightclub	Grocery Store	Sandwich Place
8	Southwark	2	Coffee Shop	Pub	Bar	Scenic Lookout	Cocktail Bar	Restaurant	French Restaurant	Hotel	English Restaurant	Italian Restaurant

By the type of venues, Cluster 2 identifies busy areas whit pubs, market and restaurants.

Cluster 3

```
# Cluster 3 traveller area

london_merged.loc[london_merged['Cluster Labels'] == 3, london_merged.columns[[0] + list(range(3, london_merged.shape[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
9	Tower Hamlets	3	Hotel	Sandwich Place	Coffee Shop	Italian Restaurant	Fried Chicken Joint	Light Rail Station	Café	Pub	Convenience Store	Grocery Store
11	Westminster	3	Hotel	Coffee Shop	Sandwich Place	Italian Restaurant	Theater	Sushi Restaurant	Pub	Gym / Fitness Center	Hotel Bar	Juice Bar

By the type of venues, Cluster 3 identifies traveller area with hotels, restaurant and public transport hubs.

Cluster 4

```
# Cluster 4 family area

london_merged.loc[london_merged['Cluster Labels'] == 4, london_merged.columns[[0] + list(range(3, london_merged.shape[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
7	Lewisham	4	Supermarket	Grocery Store	Coffee Shop	Platform	Train Station	Italian Restaurant	Pizza Place	Bakery	Shopping Mall	Bus Stop

By the type of venues, Cluster 4 identifies family area with supermarkets, stores, shopping mall and public transports.

3. Results and Recommendations

Now we will review all the analysis made in this project before making a conclusion on which area in London is more indicated to open a leisure centre.

Like mentioned in the beginning, our key criteria of location decision will be based on current data of accessibility to green areas and parks.

Accessibility to nature

The boroughs in London with little access to green areas have been identified as more attractive areas with more opportunities for the opening of a leisure centre, as more likely habitants will need a space with exercise facilities.

	Borough	Access to nature %	Latitude	Longitude
21	Lambeth	63.333333	51.4607	-0.1163
12	Hammersmith and Fulham	64.125000	51.4927	-0.2339
30	Tower Hamlets	64.941176	51.5099	-0.0059
18	Islington	67.562500	51.5416	-0.1022
28	Southwark	70.428571	51.5035	-0.0804
5	Camden	70.555556	51.5290	-0.1255
33	Westminster	73.150000	51.4973	-0.1372
11	Hackney	74.105263	51.5450	-0.0553
22	Lewisham	86.166667	51.4452	-0.0209
10	Greenwich	87.117647	51.4892	0.0648
19	Kensington and Chelsea	90.444444	51.5020	-0.1947
32	Wandsworth	90.550000	51.4567	-0.1910

By setting a threshold at 70%, the boroughs to analise are Lambeth, H&F, TH and Islington in order.

Lambeth

Lambeth is within Cluster 2 and according to my interpretation Cluster 2 identifies a busy area. Also, by looking at the most common venues, Gyms and Fitness centres have been identified as 5th most common venue.

Hammersmith and Fulham

H&F is also within Cluster 2, however looks like gyms and leisure centres are not popular venues.

Tower Hamlets

TH is within Cluster 3 and according to my interpretation Cluster 2 identifies a traveller area with Hotels as top venues. Also in this case, gyms and leisure centres are not popular venues.

Islington

Islington is within Cluster 0 and according to my interpretation Cluster 0 identifies a lively area with Hotels as top venues. Also in this case, gyms and leisure centres are not popular venues however parks have been included as one of the most common venues (4th in the rank).

4. Conclusion

From our analysis, we can suggest that the four boroughs above are potentially the best places to build a leisure centre, based on accessibility to green areas across the boroughs of London. However, the clustering analysis helps to determine further which of the above boroughs may have more potential.

Hammersmith and Fulham, and Tower Hamlets seems to be good options where to open new leisure centres as they might fill a gap in the market and compensate the lack of such venues.

5. Limitations and opportunities¶

To have more accurate results, we will need further data such as the actual number of gyms, fitness and leisure centres across London as well as the average age of the residents. However, this research was still meaningful as we were able to carry out a deep analysis of the boroughs.