Applied Data Science - Capstone project

Opening a climbing gym in London

May 1, 2020

1 Introduction

Climbing is a rising activity, gaining popularity as the sport becomes more and more mainstream. Following its introduction in the next olympics in Tokyo, the demand for new climbing gyms is now higher than ever. This project aims at finding out the best borough in London to open a new climbing gym and could be of interest for anyone looking for a new business opportunity in a young and thriving market.

The Foursquare location data of the existing clymbing gyms in London was used to figure out the best borough to open a new gym, based on the proximity of the existing gyms and the size of potential customer base.

2 Data

As already mentioned, this project used the Foursquare location data of the existing clymbing gyms in London to figure out the best borough to open a new gym, based on the proximity of the existing gyms and the size of potential customer base. This was initially assesed from this Wikipedia table, which lists all Greater London's boroughs, their geographical coordinates and population. Unfortunately this dataset was found to be unreliable and new coordinates had to be fetched from GitHub.

	Borough	Area (sqmi)	Population	Latitude	Longitude
0	Barking and Dagenham	13.93	194352	51.5607	0.1557
1	Barnet	33.49	369088	51.6252	0.1517
2	Bexley	23.38	236687	51.4549	0.1505
3	Brent	16.70	317264	51.5588	0.2817
4	Bromley	57.97	317899	51.4039	0.0198

Figure 1: First dataframe, containing the data imported from Wikipedia.

3 Methodology

To make things simple, the best borough was considered to be the one where no gym yet exists and whose five closest boroughs also lack a climbing gym. This is of course an approximation which doesn't take into account many important factors, such as the potential customer base in each borough, the other existing businesses in the area and the average income of people living and working in the borough, just to name a few.

Upon importing the data from Wikipedia, these were organised into a dataframe (Fig. 1) and the central coordinates of each borough were displayed on a Folium map, as shown in Fig. 2

It is clear that the coordinates of the boroughs are wrong. A different dataset was then downloaded from Github as a .xlsx file and imported into python as a new dataframe, shown in Fig. 3.

Upon cleaning and reorganising the data, these were plotted again in Fig. 4, whre we can see how this new database yielded indeed better results.

Next, the location data of the existing climbing gyms in London were retrieved from Foursquare using the "search" key and the word "Climbing" as a query. Specifically, for each borough, identified by its coordinates in the Github dataset, the search was limited within a radius $r = \sqrt{\frac{A}{\pi}}$ from the borough's coordinates, with A being the area of each borough taken from the Wikipedia database and assumed correct. Due to this imperfect research method, gyms close to the border of one borough could be assigned to multiple boroughs. When this was the case, the gym was assigned to the closest borough. The gym data was then compiled into a dataframe (Fig. 5) and displayed on a Folium map in Fig. 6, together with the boruoghs.

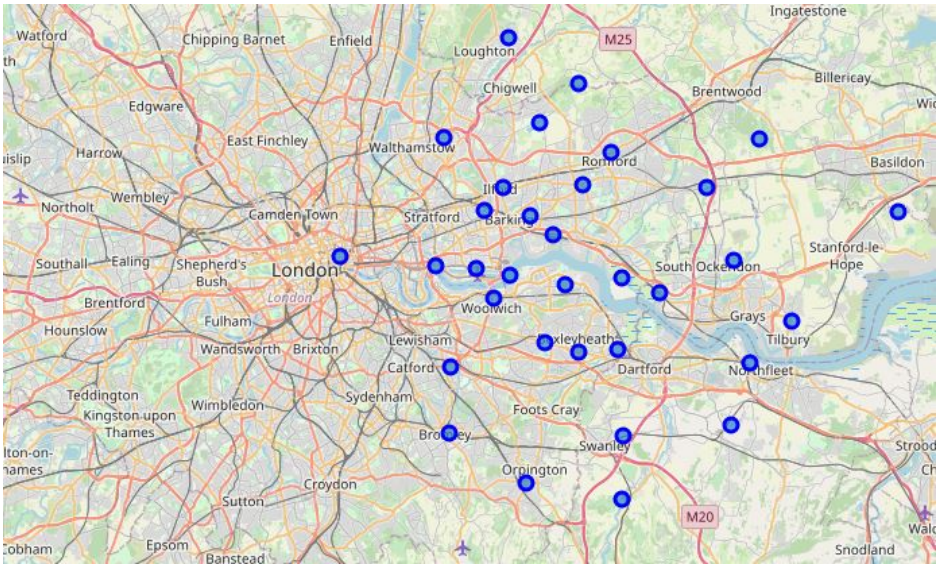

Figure 2: Folium map of London, where the blue dots mark the different boroughs as obtained from Wikipedia.

It was now possible to select, according to the criterion explained above, the best borough to open a new climbing gym.

4 Results

First, a new dataframe was created which contained the boroughs without climbing gym. For each of these boroughs, the distance from all the other boroughs was calculated and the five closest boroughs were then selected, as shown in Fig. 7.

For each row in this dataframe, it was simply checked whether any of these five closest boroughs hosted a climbing gym and, in this case, the row was discarded. What was left was the only borough who, in addition to not having a gym, has got its five closest boroughs which also do not have a climbing facility. This was the borough of Croydon, as you can see in Fig. 8 and Fig. 9.

	Borough	Population density (people/km2)	Latitude	Longitude
0	Barking and Dagenham	5792.891651	51.545277	0.133528
1	Barnet	4491.630649	51.560700	-0.210017
2	Bexley	4034.402238	51.457389	0.138861
3	Brent	7678.080133	51.454900	-0.267821
4	Bromley	2183.928180	51.371999	0.051524

Figure 3: Second dataframe, containing the data imported from Github. The Population density column is obtained using the data from Fig. 1.

5 Discussion

As naive as this study can be, we can see that it leads to a reasonable result: the southern outskirt of London, where Croydon is located, is far from any existing climbing facility, and therefore looks like a promising area to start a business. However, it needs to be stressed that this analysis overlooks many other variables which might indedd play a big role when choosing the best borough, as already outlined earlier.

Moreover, this study highlights the importance of working with trusted and reliable data, and is a reminder to always double-check the origin and trustworthiness of the data.

6 Conclusion

Despite its simplicity, this project was a useful and entertaining way to learn some basic Python skills and apply them to data science.

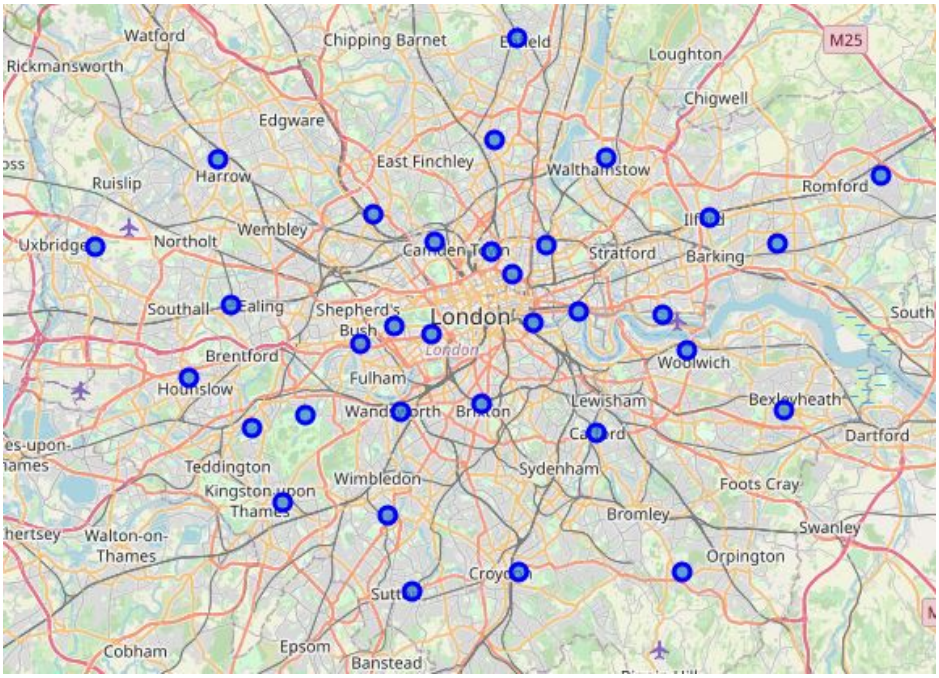

Figure 4: Folium map of London, where the blue dots mark the different boroughs as obtained from Github.

	Borough	Borough Latitude	Borough Longitude	Gym	Gym Latitude	Gym Longitude
0	Barnet	51.560700	-0.210017	Westway Climbing Wall	51.515455	-0.220350
1	Camden	51.546394	-0.157424	Swiss Cottage Climbing Centre	51.542497	-0.173436
2	Camden	51.546394	-0.157424	Better Swiss Cottage Leisure Centre	51.542550	-0.172956
3	Ealing	51.513000	-0.331026	The Arch Acton	51.506720	-0.265233
4	Greenwich	51.489200	0.056244	The Reach	51.494364	0.042757

Figure 5: Third dataframe, containing the location data of the gyms as found on Foursquare, togehter with the location of their boroughs.

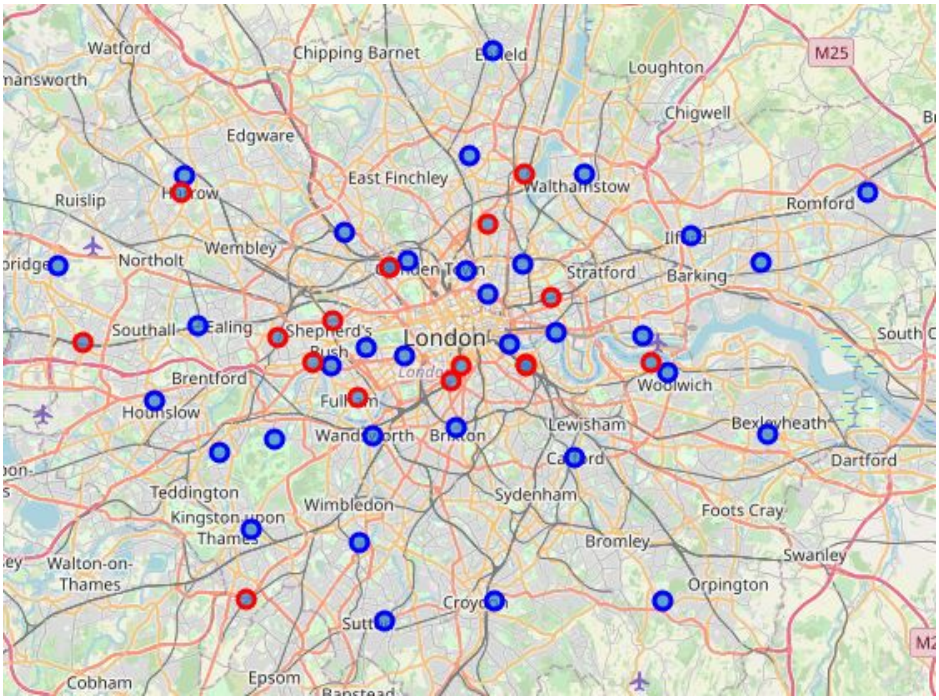

Figure 6: Folium map of London, where the blue dots mark the different boroughs as obtained from Github and the red ones mark the different gyms, as obtained from Fromsquare.

	Borough	Population density (people/km2)	Latitude	Longitude	Neighbours
0	Barking and Dagenham	5792.891651	51.545277	0.133528	Redbridge Havering Newham Greenwich Bexley
2	Bexley	4034.402238	51.457389	0.138861	Greenwich Newham Barking and Dagenham Lewisham...
3	Brent	7678.080133	51.454900	-0.267821	Richmond upon Thames Hammersmith and Fulham Ki...
4	Bromley	2183.928180	51.371999	0.051524	Lewisham Croydon Bexley Greenwich Newham
6	City of London	3033.645891	51.529000	-0.092171	Islington Hackney Southwark Tower Hamlets Camden

Figure 7: Dataframe containing the five closest boroughs for each borough without a gym.

	Borough	Population density (people/km2)	Latitude	Longitude	Neighbours
7	Croydon	4466.560887	51.3714	-0.087157	Sutton Merton Lewisham Bromley Lambeth

Figure 8: Remaining dataframe after filtering for neighbouring boroughs containing a gym.

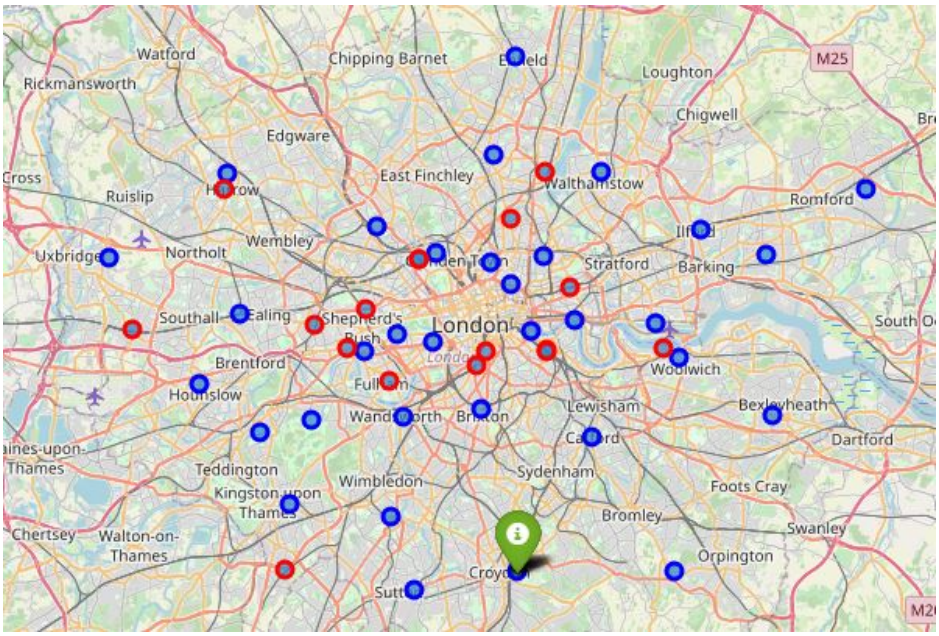

Figure 9: Folium map of London, where the blue dots mark the different boroughs as obtained from Github and the red ones mark the different gyms, as obtained from Fromsquare. The green label shows the borough of Croydon which, according to our analysis, is the most convenient one to open a new climbing gym.