



# BATTLE OF NEIGHBORHOODS

## APPLIED DATA SCIENCE CAPSTONE

Starting up a business in Fitness Industry in  
Central London

Ashish Nagpal

# INTRODUCTION

## Background

The fitness industry has been growing continuously over the past two decades. People are getting more and more health conscious and aware. Along with this, an ongoing trend of joining gyms has also boosted rapidly among the people and has taken shape of its own community. Hence, to cater to this community, a potential business in this industry is expected to be both profitable and reliable. The area of this business is to cater to fitness needs of the people – fitness equipment (personal and commercial), clothing, supplements and other accessories.

## Problem

One key aspect of starting up a business is location. It can play a determining factor in the success of the business. Hence, the main problem solved in this report is the potential locations in Central London to start up this business in the fitness industry.

## Interest

This report seeks the interest of entrepreneurs or businesspeople looking to enter the fitness industry, or a community of people who want to grow in this industry such as competitions, fitness trainers etc.

# DATA

## Data Source

The acquired data for this project is a list of Neighborhoods in London, England. This data is scraped from the Wikipedia Page: [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London) and pre-processed into the relevant data frame, and Foursquare API is used for analysis.

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge[9]	ILFORD	IG2	020	TQ455895
Aldgate	City[10]	LONDON	EC3	020	TQ334813
Aldwych	Westminster[10]	LONDON	WC2	020	TQ307810
Alperton	Brent[11]	WEMBLEY	HA0	020	TQ185835
Anerley	Bromley[11]	LONDON	SE20	020	TQ345695
Angel	Islington[8]	LONDON	EC1, N1	020	TQ345665
Aperfield	Bromley[11]	WESTERHAM	TN16	01959	TQ425585
Archway	Islington[12]	LONDON	N19	020	TQ285875
Ardleigh Green	Havering[12]	HORNCHURCH	RM11	01708	TQ535895
Arkley	Barnet[12]	BARNET, LONDON	EN5, NW7	020	TQ225955

## Data Pre-processing

1. This data is extracted into a data-frame with column names of “Location, London borough, Post town, Postcode district, Dial code and OS grid ref changed” to “Neighborhood, Borough, Town, PostCode, DialCode and OS grid ref” respectively.
2. The values in the data-frame are cleaned. As seen in the figure, the values contain numbered citations in square brackets, and ‘also known as’ for different locations which are all removed.
3. Any rows containing null OS grid ref values are dropped. This OS grid ref is another type of reference system of the geographical location which is based on grids. These are converted into geographical latitude and longitude values and added to the data-frame.
4. Finally, irrelevant columns – OS grid ref and DialCode are dropped.

Below is the data-frame head for all the neighborhoods in London:

	Borough	Neighborhood	Town	PostCode	Latitude	Longitude
0	Bexley, Greenwich	Abbey Wood	LONDON	SE2	51.486484	0.109318
1	Ealing, Hammersmith and Fulham	Acton	LONDON	W3, W4	51.510591	-0.264585
2	Croydon	Addington	CROYDON	CR0	51.362934	-0.025780
3	Croydon	Addiscombe	CROYDON	CR0	51.381625	-0.068126
4	Bexley	Albany Park	BEXLEY, SIDCUP	DA5, DA14	51.434929	0.125663

Now, this data frame is sliced to extract out all the Neighborhoods which come under Central London. This is done by extracting the rows with Boroughs listed under Central London, which can be found here: [https://en.wikipedia.org/wiki/Central\\_London](https://en.wikipedia.org/wiki/Central_London)

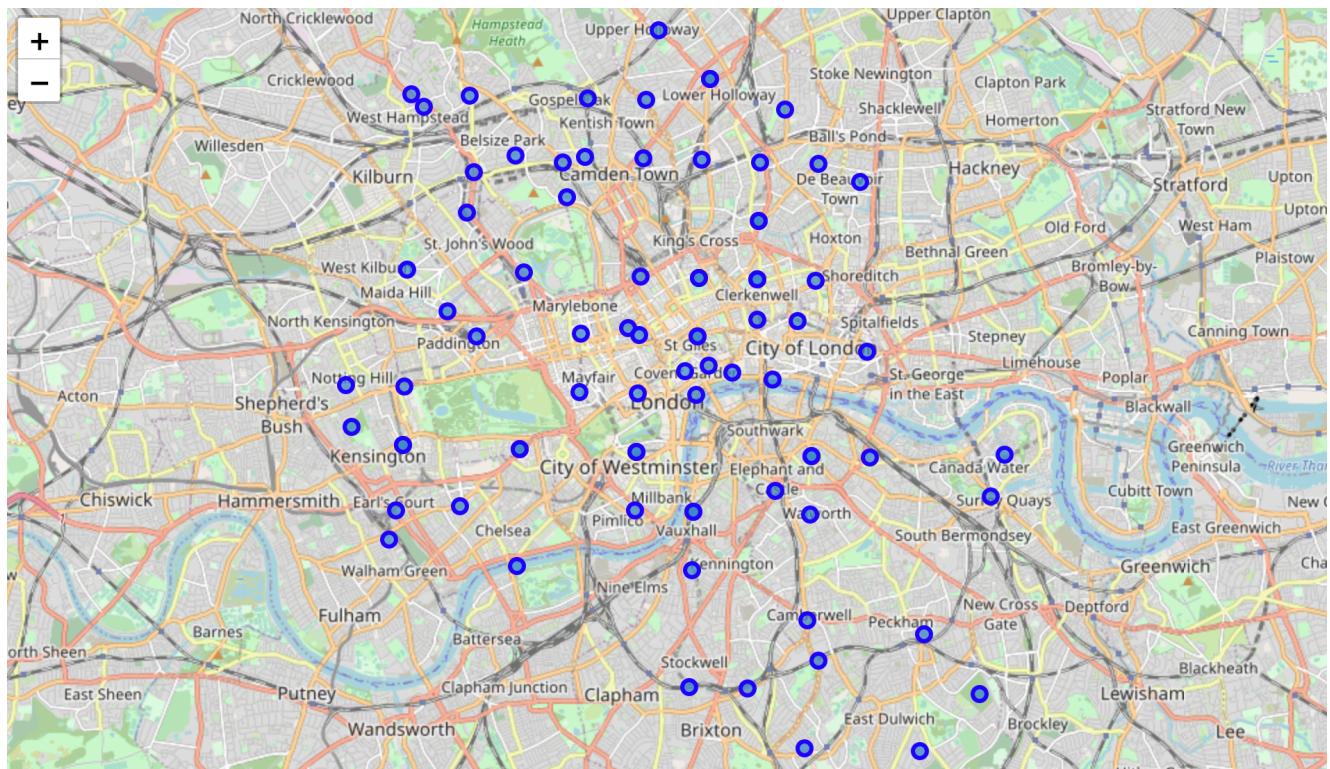
Below is the final data-frame head for Neighborhoods in Central London.

	Borough	Neighborhood	Town	PostCode	Latitude	Longitude
0	Camden	Belsize Park	LONDON	NW3	51.545049	-0.165131
1	Camden	Bloomsbury	LONDON	WC1	51.526345	-0.119715
2	Camden	Camden Town	LONDON	NW1	51.544548	-0.133398
3	Camden	Chalk Farm	LONDON	NW1	51.543969	-0.153628
4	Camden	Fitzrovia	LONDON	W1	51.518533	-0.137347

This data consists of 89 neighborhoods in Central London in 12 different Boroughs.

# METHODOLOGY

Firstly, the coordinates of London are extracted from geolocator and saved. Using these, a Folium Map is created plotting all the neighborhoods in Central London.



Next, one of these neighborhoods – Soho, Westminster is explored for top 100 venues. For this, Foursquare credentials are defined, and using the latitude and longitude values of Soho neighbourhood, top 100 venues are extracted from Foursquare.

	name	categories	lat	lng
0	The Punch Room	Cocktail Bar	51.516905	-0.136151
1	The London Edition (The London EDITION)	Hotel	51.516762	-0.136049
2	Kaffeine	Coffee Shop	51.516785	-0.137080
3	Roka Japanese Restaurant	Japanese Restaurant	51.518992	-0.135308
4	Vagabond	Wine Bar	51.518695	-0.135003

Now, the above process of exploring a neighborhood and extracting top 100 venues of the neighborhood using Foursquare is defined into a function and the process is repeated for all neighborhoods in Central London, using their respective location coordinates.

A data-frame containing all the venues near all the Central London neighborhoods is shown below. This data-frame consists of 4329 venues with 318 unique categories.

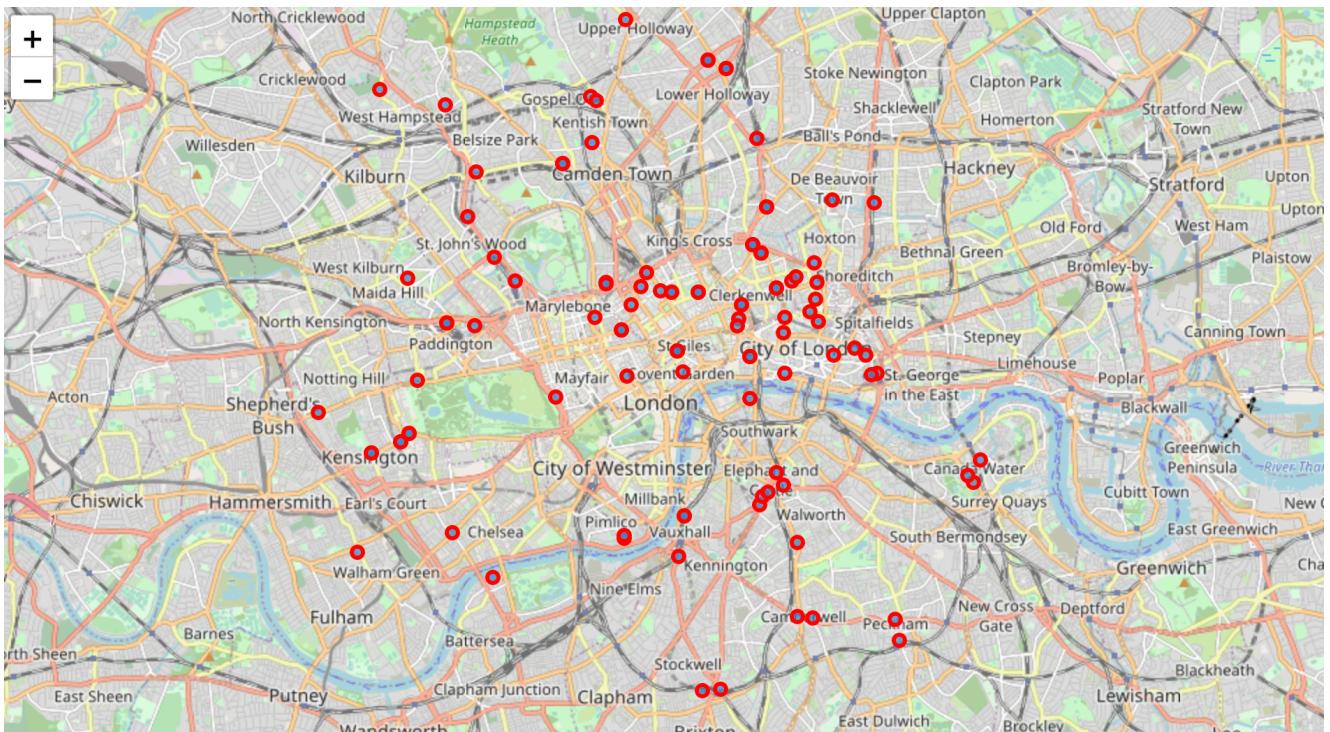
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Belsize Park	51.545049	-0.165131	Chamomile	51.545729	-0.162398	Café
1	Belsize Park	51.545049	-0.165131	Sable D'or	51.545990	-0.162048	Café
2	Belsize Park	51.545049	-0.165131	Black Truffle	51.545977	-0.162530	Deli / Bodega
3	Belsize Park	51.545049	-0.165131	Starbucks	51.545459	-0.162607	Coffee Shop
4	Belsize Park	51.545049	-0.165131	The Washington	51.545467	-0.162768	Pub
5	Belsize Park	51.545049	-0.165131	Primrose Hill Market	51.541779	-0.162012	Market
6	Belsize Park	51.545049	-0.165131	Grain Artisan Sourdough	51.545910	-0.162218	Café
7	Belsize Park	51.545049	-0.165131	M Lounge	51.542057	-0.170372	Hotel Bar
8	Belsize Park	51.545049	-0.165131	Co-op Food	51.543196	-0.166608	Grocery Store
9	Belsize Park	51.545049	-0.165131	Haverstock Hill	51.548460	-0.165340	Hill
10	Belsize Park	51.545049	-0.165131	Bonjour Brioche	51.546747	-0.161015	Café
11	Belsize Park	51.545049	-0.165131	The Sir Richard Steele	51.547000	-0.158916	Pub
12	Belsize Park	51.545049	-0.165131	London Marriott Hotel Regents Park	51.542148	-0.170034	Hotel
13	Belsize Park	51.545049	-0.165131	Oliver's Fish and Chips	51.546892	-0.158567	Fish & Chips Shop
14	Belsize Park	51.545049	-0.165131	Executive Lounge	51.542079	-0.170306	Hotel Bar

Now, the unique Venue Category values for this data-frame are analysed. For this project, the aim is to extract all venues with category 'Gym' or related to 'Gym', 'Fitness', etc. It is found that categories 'Gym', 'Gym / Fitness Center', 'Gym Pool', 'Gymnastics Gym' all relate to the target category I am aiming for, hence all the venues with these categories are extracted into another data-frame for further analysis.

The data-frame shows that there are 103 gyms listed in Foursquare in Central London.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bloomsbury	51.526345	-0.119715	The Gym	51.524175	-0.125031	Gym
1	Bloomsbury	51.526345	-0.119715	Nuffield Health Fitness & Wellbeing Gym	51.524178	-0.118479	Gym
2	Holborn	51.517359	-0.120085	Good Vibes	51.515064	-0.123741	Gym
3	Somerstown	51.526575	-0.134133	Bannatyne Health Club	51.524415	-0.127815	Gym
4	St Giles	51.517359	-0.120085	Good Vibes	51.515064	-0.123741	Gym
5	St Pancras	51.526345	-0.119715	The Gym	51.524175	-0.125031	Gym
6	St Pancras	51.526345	-0.119715	Nuffield Health Fitness & Wellbeing Gym	51.524178	-0.118479	Gym
7	De Beauvoir Town	51.540992	-0.080142	Crossfit Hackney	51.537875	-0.075192	Gym
8	St Luke's	51.525880	-0.090878	London Fight Factory	51.528507	-0.089985	Gym
9	Kensington	51.500517	-0.192876	Equinox Kensington	51.501369	-0.191634	Gym
10	North Kensington	51.500517	-0.192876	Equinox Kensington	51.501369	-0.191634	Gym
11	West Brompton	51.486182	-0.196326	The Gym	51.484301	-0.202171	Gym
12	Elephant and Castle	51.493669	-0.100876	Gymbox	51.494664	-0.097604	Gym
13	Elephant and Castle	51.493669	-0.100876	Snap Fitness	51.491668	-0.103291	Gym
14	Rotherhithe	51.499037	-0.044414	Maple Quays Gym	51.498429	-0.049103	Gym
15	Walworth	51.489934	-0.092381	The Gym	51.485818	-0.094187	Gym
16	Covent Garden	51.512013	-0.123190	Good Vibes	51.515064	-0.123741	Gym

For visual representation, all these gyms are plotted on the Folium Map to get a rough idea of the target location of the business.



It is somewhat clear from the map about the location of gyms in Central London. For starting up the suggested business in Central London, it is best to start with Neighborhoods with most gyms than Neighborhoods with less/no gyms.

Now, One-hot coding is performed on the above data-frame and all the neighborhoods with different category of gyms are listed. Finally, these are grouped, and each category is summed up for respective neighborhood.

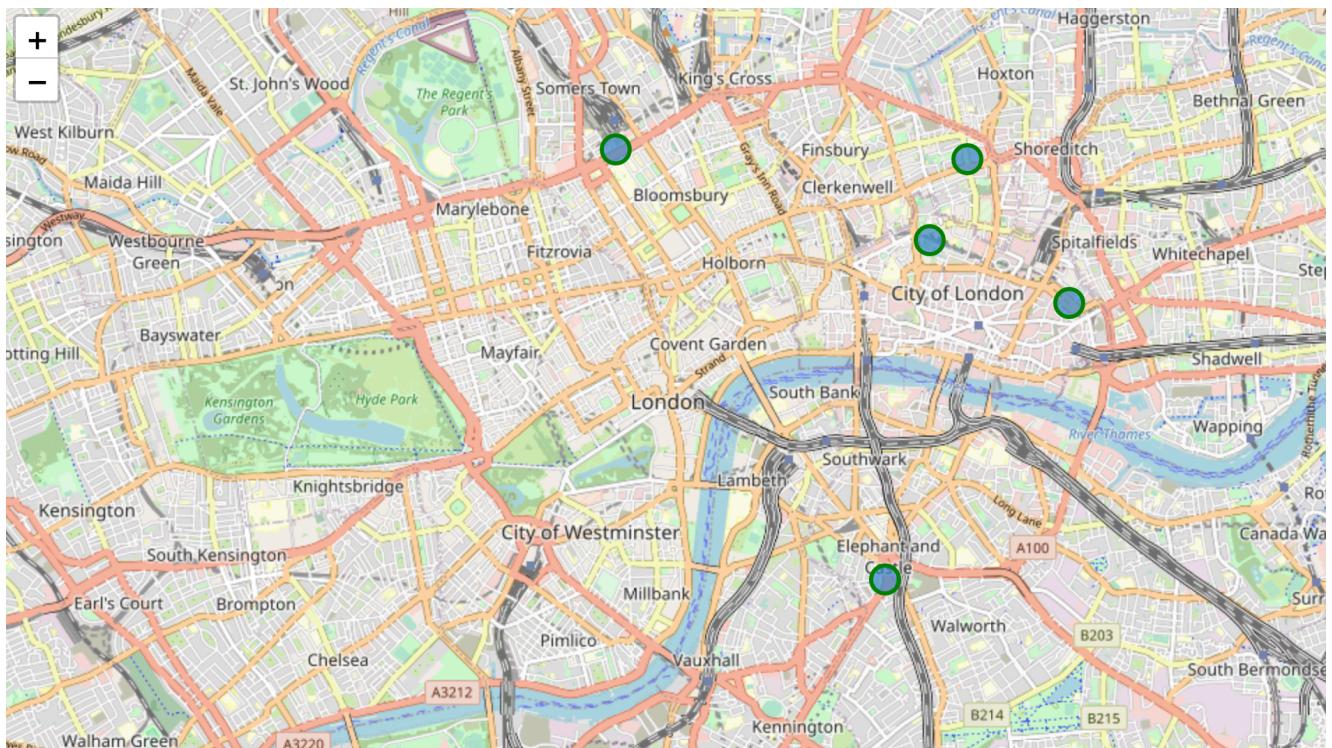
	<b>Neighborhood</b>	<b>Gym</b>	<b>Gym / Fitness Center</b>	<b>Gym Pool</b>	<b>Gymnastics Gym</b>
0	Aldgate	0	5	0	0
1	Aldwych	0	1	0	0
2	Barbican	1	3	0	0
3	Bayswater	0	1	0	0
4	Blackfriars	0	3	0	0
5	Bloomsbury	2	0	0	0
6	Brixton	0	1	0	0
7	Camberwell	0	2	0	0
8	Chalk Farm	0	0	0	1

Finally, these gyms of different categories are totalled, and the data is arranged in descending order, the neighborhood with the most gyms on the top.

	<b>Neighborhood</b>	<b>Gym</b>	<b>Gym / Fitness Center</b>	<b>Gym Pool</b>	<b>Gymnastics Gym</b>	<b>Total</b>
40	Somerstown	1	5	0	0	6
13	Elephant and Castle	2	3	0	0	5
45	St Luke's	1	3	1	0	5
0	Aldgate	0	5	0	0	5
2	Barbican	1	3	0	0	4

A new data-frame is extracted from the Central London Neighborhoods data-frame with the top 5 neighborhoods with the most amount of gyms and their respective coordinates and plotted on the Folium Map.

	Borough	Town	PostCode	Latitude	Longitude
Neighborhood					
<b>Somerstown</b>	Camden	London	NW1	51.526575	-0.134133
<b>Elephant and Castle</b>	Southwark	London	SE1, SE11, SE17	51.493669	-0.100876
<b>St Luke's</b>	Islington	London	EC1	51.525880	-0.090878
<b>Aldgate</b>	City	London	EC3	51.514885	-0.078356
<b>Barbican</b>	City	London	EC1	51.519660	-0.095466



## RESULT AND DISCUSSION

It is seen from the analysis, the Neighborhoods - "Somerstown", "Elephant and Castle", "St Luke's", "Aldgate", "Barbican" are the ones with the greatest number of gyms. It is safe to assume that these neighborhoods would house more a larger community of fitness related people than any other neighborhood. Hence, in order to start-up a business in the fitness industry, these locations are the suggested top 5.

However, it is not taken into account that nowadays, universities, offices and even housing complexes have their private gyms. This can be an influencing factor in the above data and analysis. This is very specific, and thus a topic for future detailed analysis.

## CONCLUSION

It can be concluded from the report that the ideal locations for starting up a business in the fitness industry can be determined using data analysis. Not limited to this, the same process can be followed for any potential business in order to determine its location. The key aspect in this is determining the target customer. In this case, it is the fitness community. It is assumed that a large community will be present in areas with the greatest number of gyms. Hence, it can be concluded that business location can be determined using data analysis and compared, once the target customers are known.