Capstone Project: The Battle of Neighbourhoods, GyMania

IBM Data Science Professional Certificate

Muhsin B. Caglar, 16.02.2021

Introduction: Business Problem

In this project, we will try to outline the best locations for a gym. Specifically, the project will focus on stakeholders interested in opening a commercial gym in London, United Kingdom.

From the beginning of the 21st century, with increase in available leisure time, less people working in physically demanding jobs due to automation, more in-depth knowledge of dietary and health knowledge and the obesity epidemic, people are more inclined to start a physical activity routine such as getting a gym membership. Local and national gym chains have seen a very large surge in their number of members. Moreover, it is clear that once the lockdown restrictions in place due to COVID-19 are lifted, a considerable amount of people will be craving to get back out there and improve their health, wellbeing, and body image.

In this project, we will try to focus on factors that affect the success of a gym business and try to optimise certain variables to identify the areas which have the most chance of success if a new gym was opened. Specifically, we will look at possible areas of opportunity in London, United Kingdom.

There will be several factors that we will focus on. Obviously, the number one factor will be the availability of a gym nearby as this will increase the number of competitors. Population density will be another factor that will be considered. Also, we would look at possible opportunities near area centres as most people will be going to the gym prior to or after work, hence being closer to work hubs and offices will be a bonus. Also, we will consider square meter price of properties in the areas.

Data analysis and data science tools will be used in order to analyse the available data based on the abovementioned criteria. Advantages will be outlined and the areas with best potentials will be listed.

Data

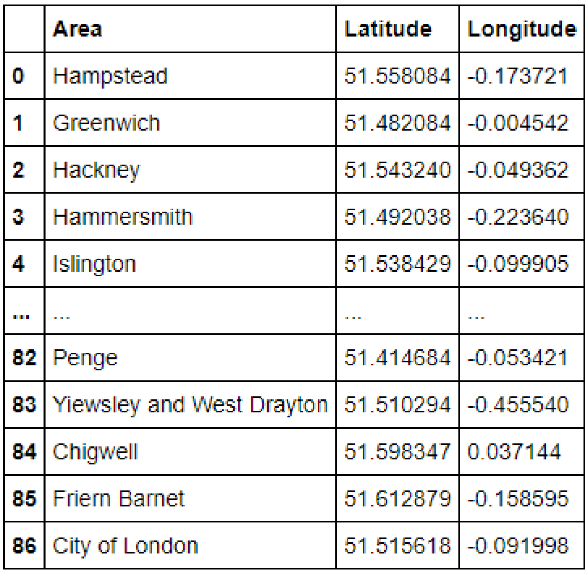
As previously outlined, we will require the following data:

- number of existing gyms in the given area
- population density of each area. (Based on Borough area and population)
- distance of neighbourhood from borough centre

The boroughs within greater London will be divided up into the areas within it and the required data will be collected.

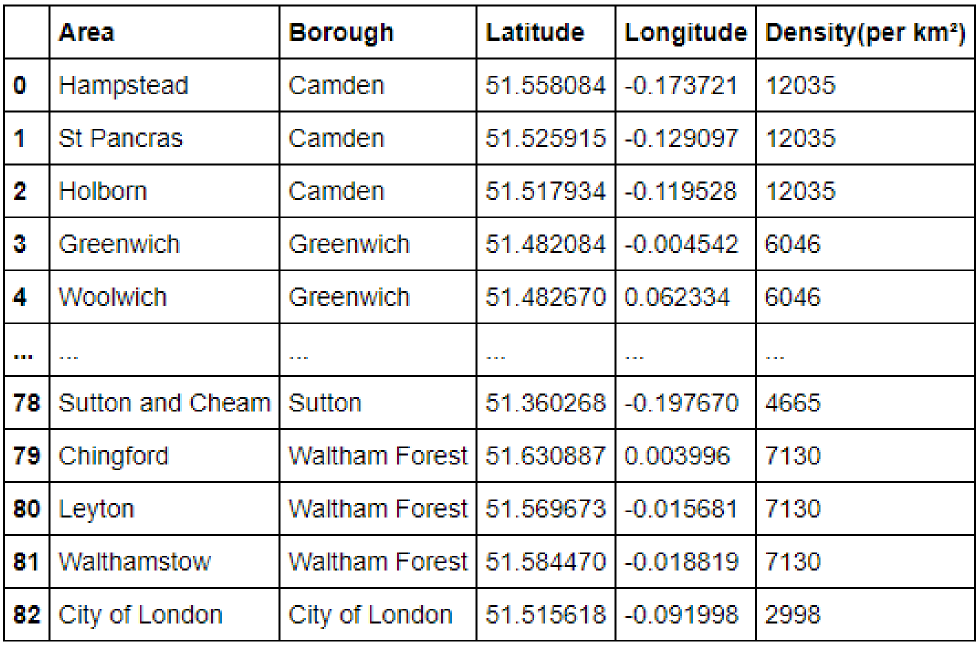
Following data sources will be needed to extract/generate the required information:

- population density and square meter price of properties and area centre names will be scraped from datasets on the internet.
- Approximate coordinates of centres of areas will be obtained using **Geopy libraries Neomatim tool**
- number of gyms and locations in every neighbourhood will be obtained using **Foursquare API**



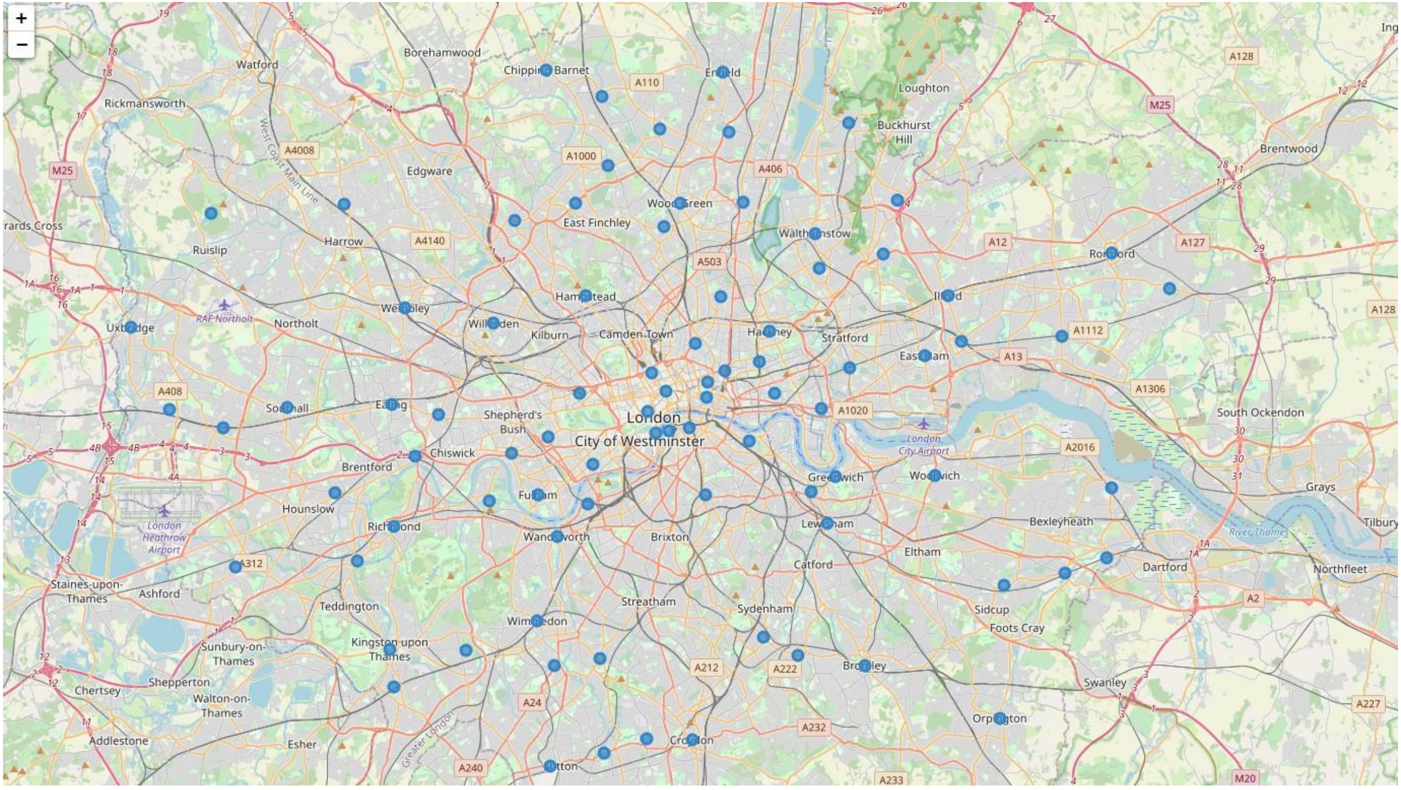
	Area	Latitude	Longitude
0	Hampstead	51.558084	-0.173721
1	Greenwich	51.482084	-0.004542
2	Hackney	51.543240	-0.049362
3	Hammersmith	51.492038	-0.223640
4	Islington	51.538429	-0.099905
...	...	...	...
82	Penge	51.414684	-0.053421
83	Yiewsley and West Drayton	51.510294	-0.455540
84	Chigwell	51.598347	0.037144
85	Friern Barnet	51.612879	-0.158595
86	City of London	51.515618	-0.091998

Location info Table

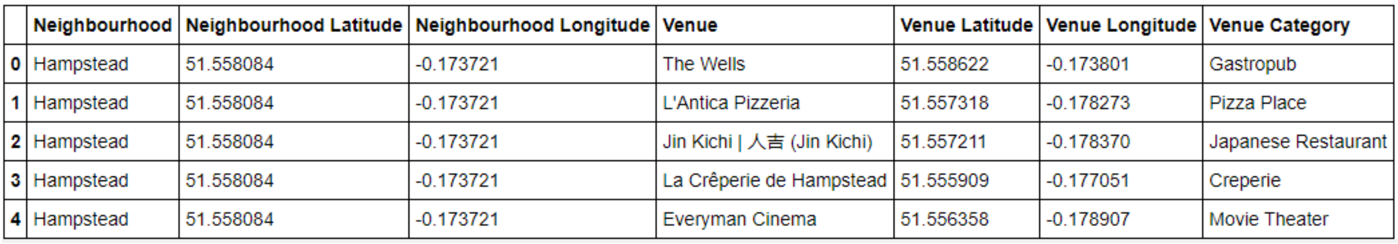


	Area	Borough	Latitude	Longitude	Density(per km^2)
0	Hampstead	Camden	51.558084	-0.173721	12035
1	St Pancras	Camden	51.525915	-0.129097	12035
2	Holborn	Camden	51.517934	-0.119528	12035
3	Greenwich	Greenwich	51.482084	-0.004542	6046
4	Woolwich	Greenwich	51.482670	0.062334	6046
...	...	...	...	...	...
78	Sutton and Cheam	Sutton	51.360268	-0.197670	4665
79	Chingford	Waltham Forest	51.630887	0.003996	7130
80	Leyton	Waltham Forest	51.569673	-0.015681	7130
81	Walthamstow	Waltham Forest	51.584470	-0.018819	7130
82	City of London	City of London	51.515618	-0.091998	2998

Location info with Population Density



Neighbourhood Centres Identified

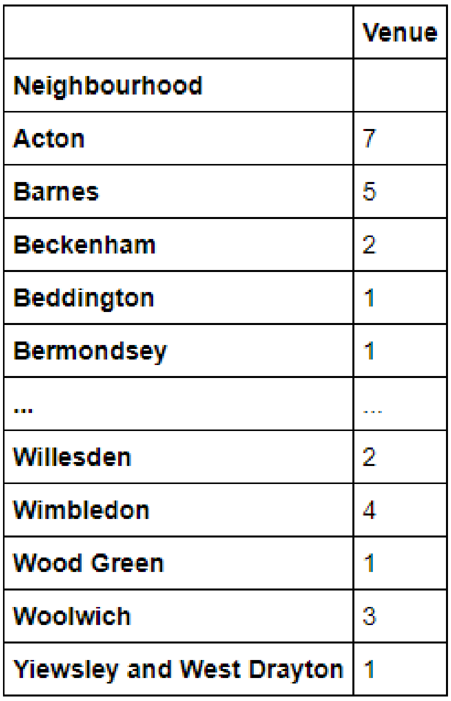


	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hampstead	51.558084	-0.173721	The Wells	51.558622	-0.173801	Gastropub
1	Hampstead	51.558084	-0.173721	L'Antica Pizzeria	51.557318	-0.178273	Pizza Place
2	Hampstead	51.558084	-0.173721	Jin Kichi | 人吉 (Jin Kichi)	51.557211	-0.178370	Japanese Restaurant
3	Hampstead	51.558084	-0.173721	La Crêperie de Hampstead	51.555909	-0.177051	Creperie
4	Hampstead	51.558084	-0.173721	Everyman Cinema	51.556358	-0.178907	Movie Theater

Initial Unprocessed Venues Found

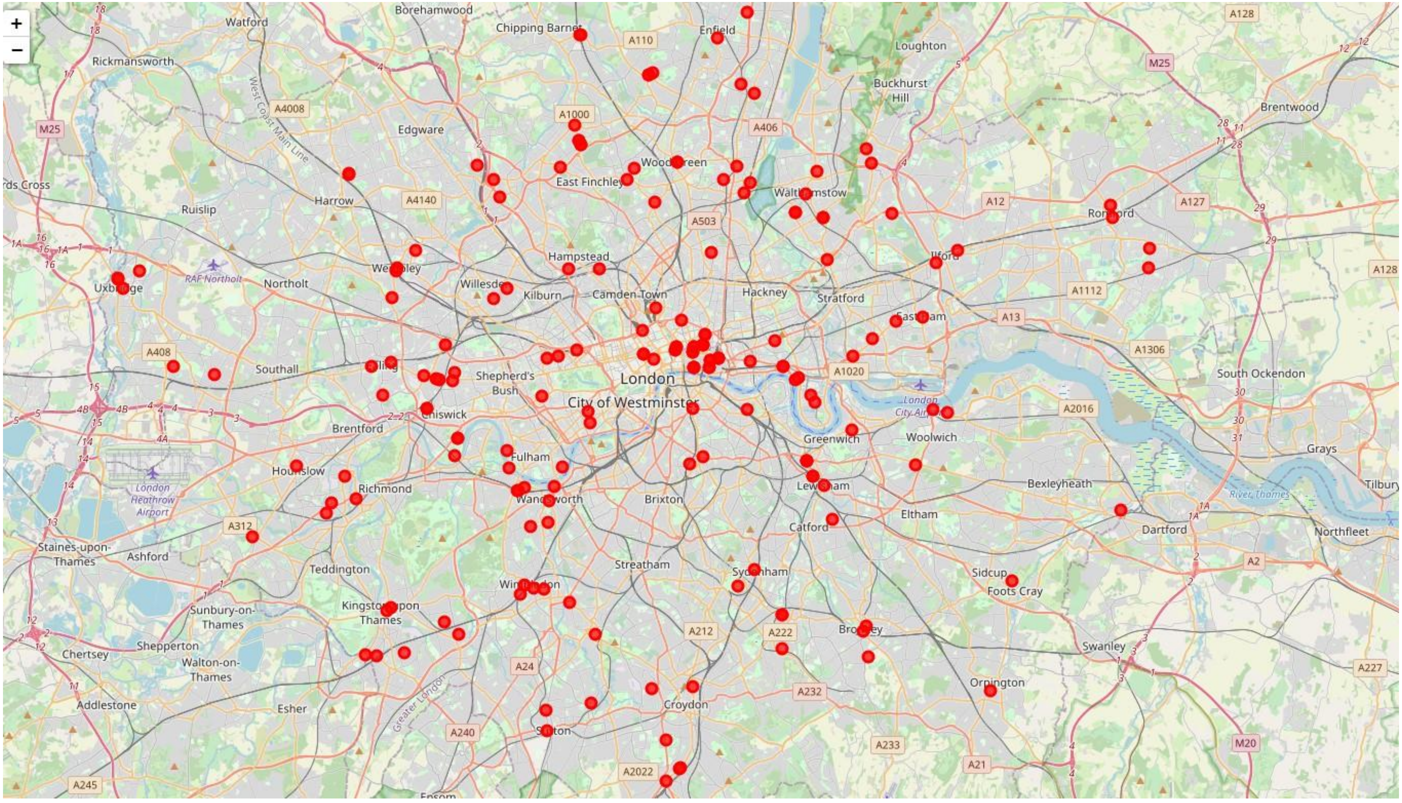
asdfds

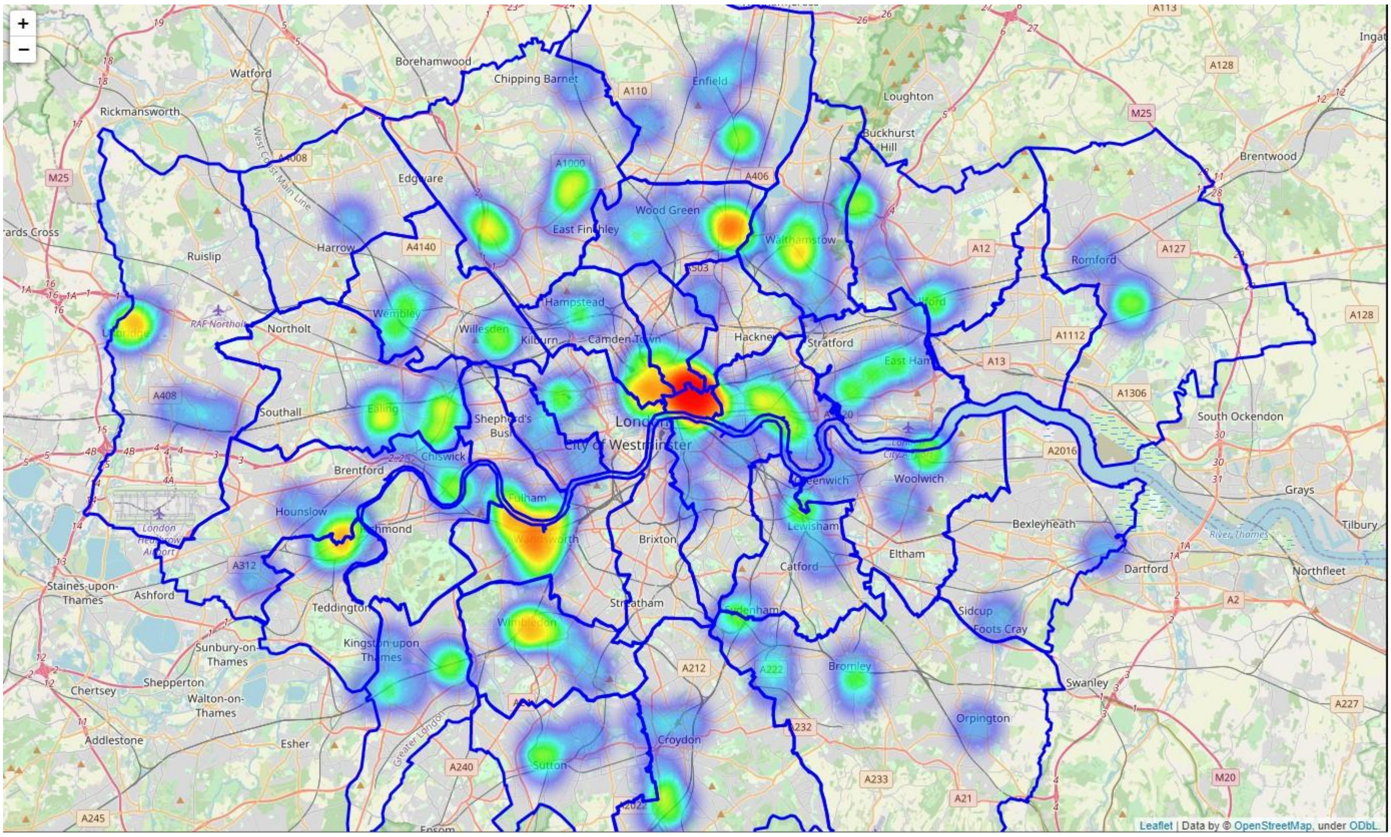
Venue Data Processed for Venues only Containing Gym in Category



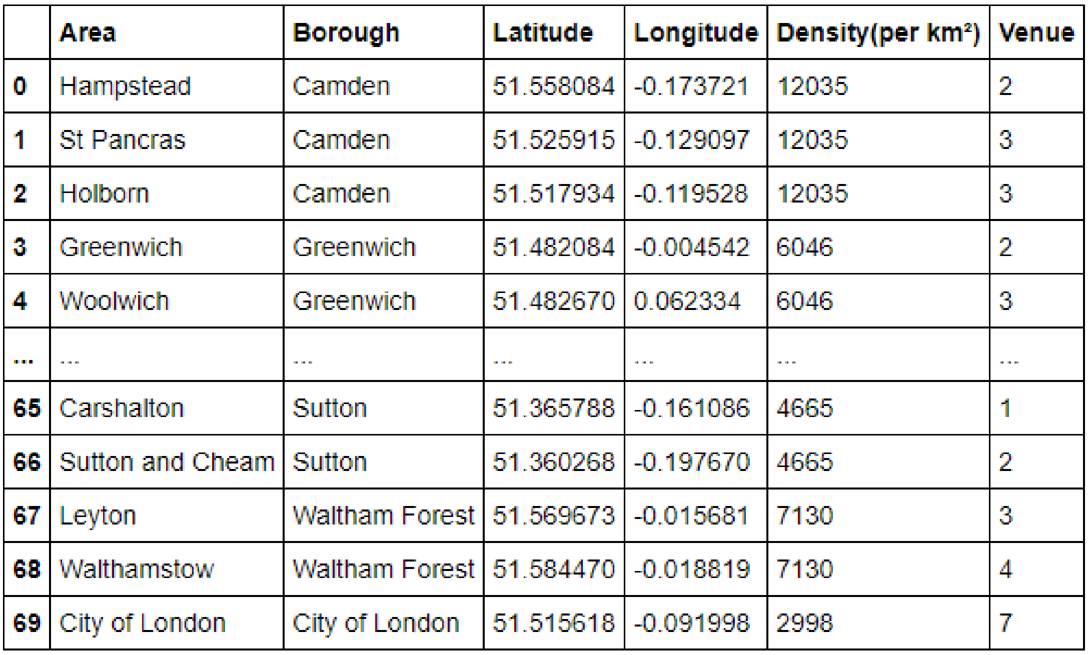
	Venue
Neighbourhood	
Acton	7
Barnes	5
Beckenham	2
Beddington	1
Bermondsey	1
...	...
Willesden	2
Wimbledon	4
Wood Green	1
Woolwich	3
Yiewsley and West Drayton	1

Number of Gyms in Each Neighbourhood





Maps of Distribution of Gyms



	Area	Borough	Latitude	Longitude	Density(per km²)	Venue
0	Hampstead	Camden	51.558084	-0.173721	12035	2
1	St Pancras	Camden	51.525915	-0.129097	12035	3
2	Holborn	Camden	51.517934	-0.119528	12035	3
3	Greenwich	Greenwich	51.482084	-0.004542	6046	2
4	Woolwich	Greenwich	51.482670	0.062334	6046	3
...	...	...	...	...	...	...
65	Carshalton	Sutton	51.365788	-0.161086	4665	1
66	Sutton and Cheam	Sutton	51.360268	-0.197670	4665	2
67	Leyton	Waltham Forest	51.569673	-0.015681	7130	3
68	Walthamstow	Waltham Forest	51.584470	-0.018819	7130	4
69	City of London	City of London	51.515618	-0.091998	2998	7

Number of Gyms Added to the Main Data frame (Venue Column)

Methodology ¶

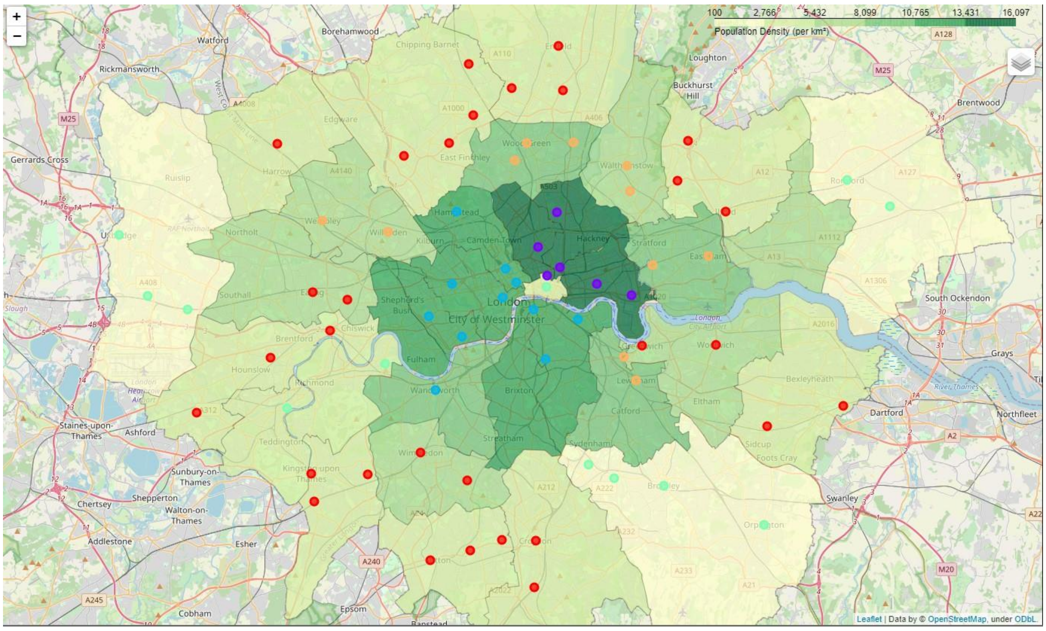
In this project, we will try to identify significant trends between data available and where current gyms in London are located. We will mainly focus on population density and gyms that are 1 mile away from the neighbourhood centre. First step was to collate the data required which was scraped of the internet and Foursquare was used to get data for the venues available in London. The data collected consisted of the population density of each borough, the areas within the borough, latitude, and longitude of the centre of each area.

Second step was to calculate gym density across different areas of London. Heatmaps were produced in order to identify promising areas close to neighbourhood centres where few or no gyms operated.

Third step was to pre-process the available data to allow for kNN algorithm to be utilised. The algorithm was used to cluster the available data of number gyms based on population density and the number of venues available in each area. This will allow for further analysis on the aspects that determine where gyms usually operate and identify possible areas that can be exploited for a successful gym business.

Analysis ¶

kNN method was used to cluster the data with population density and number of gyms taken into account. The map produced is below



Clusters and Population Density Map

Results and Discussion ¶

The results of the kNN algorithm clearly show that their is a positive correlation between population density and number of gyms operating in a given area. The heatmap produced also prove this point. The maps produced further showed that areas such as Brixton in the South of London and Hackney in the North of London have very few gyms for there given areas and population densities. These areas may be of intrest to stakeholders looking to open or expand their gym business. These areas may need further analysis and street level investigation to solidify the findings of this data analysis project.

The kNN algorithm is very robust and produced very good results in this case however, the scope of the project can be extended to include other machine learning algorithms to further solidify the findings. Further elbow method analysis on the kNN algorithm can be used to improve and optimise the k value that will be used.

Conclusion

The findings above are based on pretty concrete evedince. I believe with timely updates to this kind of data science projects, new lucaritive business oppurtunities can be found before they are realised by the general public or the common businessperson. Data science allows for huge amounts of knowledge to be acquired in a short time which is a huge leverage when making decisions of any kind. I believe data science will play a huge role in both finance and innovation in the near future and I hope to be a part of it. The project has immensely furthered my understanding of data science. I hope to use and improve the skills I have gained from this course in the future.

As this project will not be trusted by your computer I have included screenshot of all the maps and figures produced.