

COURSERA CAPSTONE:

OPENING A GYM IN BARCELONA

1. INTRODUCTION

The city of Barcelona, Spain is one of the most visited cities in the world. Its beautiful beach landscapes, important monuments and culture are, without a shadow of doubt relevant factors affecting its tourism. Its economy is based upon this area, and many families and individuals depend on it.

It is a well acknowledged truth that during the current situation (where the whole Spanish population is in quarantine due to the coronavirus crisis), many people have lost their job, as hotels/restaurants/museums have had to close due to the disappearance of clientele.

This situation has left many people unemployed, seeking for new opportunities to sustain their life in Barcelona.

Generally, it is expected that these people are going to try to use the current situation to start their own business, and the area of fitness and health, obviously related to gyms is going to be an emergent opportunity in the near future.

Moreover, this is a priori one of the most interesting areas to start a new business, as when the quarantine is over, it is expected that many people will be returning to their previous lifestyle of going to the gym, and as well, many new people will be entering gyms, as they will be fed up of their inactivity, caused by the isolation.

2. BUSSINESS PROBLEM

The aim of this project is to analyse whether opening a gym in the city of Barcelona, Spain is a good opportunity or not.

Throughout the use of Data Science, paired with Machine Learning algorithms (such as K-means clustering), the different suburbs in the city will be classified according to the amount of gyms they have or do not have. By doing so, it will be determined if opening a gym is a good opportunity as a whole, and if so, what are the best suburbs to do so in Barcelona.

3. DATA

For solving this problem, various data sources will be used. Firstly, the following Wikipedia page, https://en.wikipedia.org/wiki/Districts_of_Barcelona, will be used to obtain the name of all the neighborhoods of the city. We will do so by performing web-scraping techniques, more specifically Python requests and the beautiful soup package.

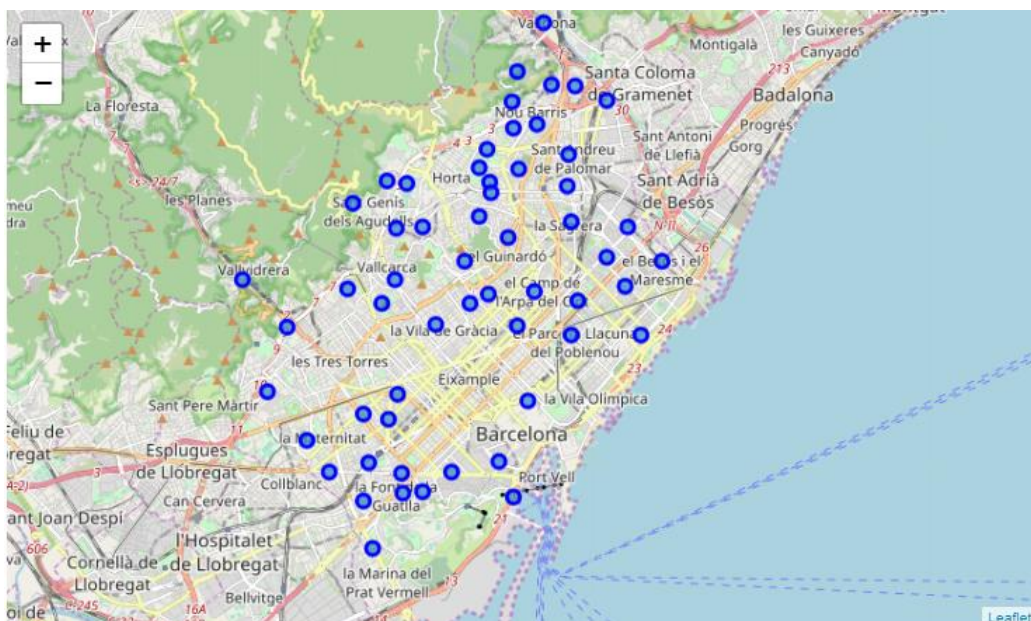
By doing this we define the scope of the project, that is no other than the city of Barcelona.

We will also need information about venues in every neighborhood, specially gyms, that we will obtain through the Foursquare API. To do so, it is necessary to previously obtain the data regarding latitude and longitude for every neighborhood in the city, for what we will use the python geocoder package.

4. METHODOLOGY

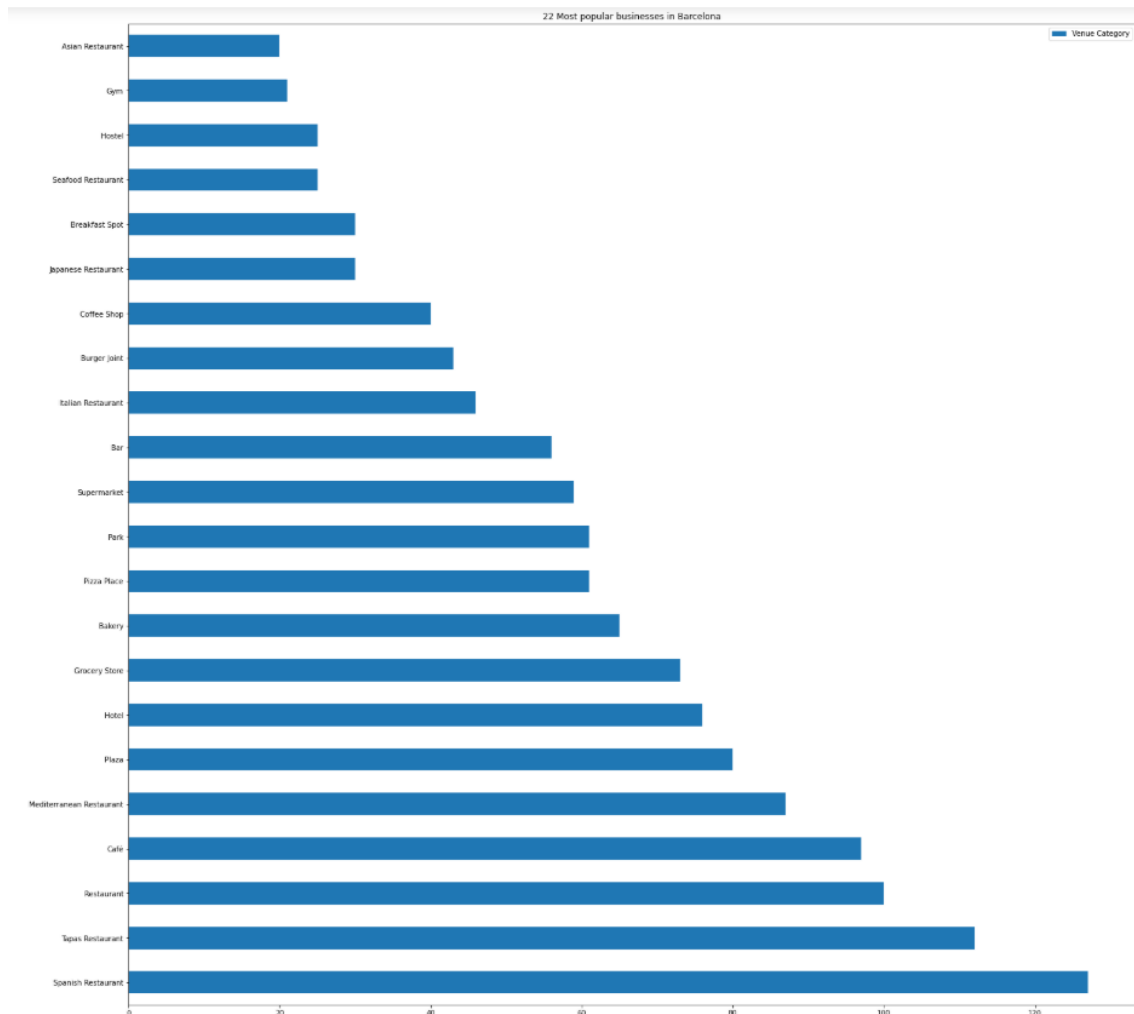
To develop any kind of Data Science project, one of the initial states is to obtain the necessary information to then obtain our conclusions and take decisions based upon those conclusions. As mentioned previously this will be done using various sources like the Wikipedia page, the API and the geocoder package, which is of extreme utility to utilize the Foursquare API.

With this information we will start off by creating a map of the city, and positioning each neighborhood, in order to develop a previous understanding of the distribution of these in the city (each different neighborhood is represented by a blue spot in the following map).



We then introduce the geospatial information of each neighborhood into the Foursquare API. By doing so we are going to obtain the existing venues in every neighborhood, with a determined limit and radius of exploration (in this case, 100 venues were obtained in a radius of 500 from the neighborhood's centre). Two hundred and forty-three different kind of venues were obtained overall.

For the initiation of any kind of businesses one of the most relevant factors to consider is clearly the possible competence against which we will be fighting.



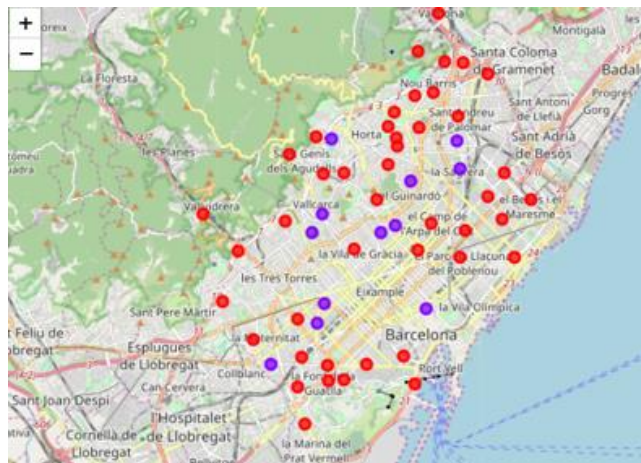
To have a clear vision of the distribution of gyms in the city we proceed to cluster the neighborhoods into 3 clusters, depending on the number of gyms present. The machine learning algorithm used for this purpose has been K-Means clustering, as we want to partition our observations into k ($k=3$) clusters, where every observation belongs to the cluster with the nearest mean of gym presence.

As we can observe from the previous horizontal bar chart, where we have plotted the 22 most popular business types in Barcelona, the Gym industry places twenty-first, which is not very high at all.

5. RESULTS

The results were surprisingly positive towards our hypothesis, as only of the neighborhoods were included into the group with many gyms, whereas the rest were divided between having little number of gyms or being in between.

This means that as a whole, the industry of gyms in Barcelona is yet to be exploited, which is great for our people of interest. (It can be seen by the graph, where 15 sample neighborhoods were introduced into a table with their relative gym frequency. It is clear that they are really small).



Neighborhood	Gym
Can Baró	0.000000
Can Peguera	0.016129
Canyelles (neighbourhood)	0.000000
Ciutat Meridiana	0.035714
Dreta de l'Eixample	0.000000
El Baix Guinardó	0.000000
El Camp d'en Grassot i Gràcia Nova	0.000000
El Carmel	0.090909
El Coll	0.032258
El Guinardó	0.000000
El Gòtic	0.023810
El Poble-sec, Barcelona	0.000000
El Putget i Farró	0.037037
El Raval	0.030000
El Turó de la Peira	0.000000
Fort Pienc	0.011905

Knowing this, it is obvious that anyone interested in opening a new business would rather do it somewhere with less competence.

This group would correspond to our red cluster, which is indeed the most populated one.

6. DISCUSSION

Neighborhood	Neighborhood	Neighborhood	Neighborhood
7 El Carmel	1 Can Peguera	0 Can Baró	55 Vila de Gràcia
	23 La Font de la Guatlla	56 Vilapicina i la Torre Llobeta	27 La Maternitat i Sant Ramon
	3 Ciutat Meridiana	29 La Prosperitat	26 La Marina del Prat Vermell
	49 Sarrià, Barcelona	30 La Salut	28 La Nova Esquerra de l'Eixample
	10 El Gòtic	31 La Teixonera	24 La Guineueta, Barcelona
	46 Sant Pere, Santa Caterina i la Ribera	32 La Trinitat Nova	2 Canyelles (neighbourhood)
	45 Sant Gervasi - la Bonanova	33 La Vall d'Hebron	4 Dreta de l'Eixample
	8 El Coll	34 Les Corts (neighbourhood)	5 El Baix Guinardó
	12 El Putget i Farró	35 Les Roquetes, Barcelona	6 El Camp d'en Grassot i Gràcia Nova
	39 Pedralbes	36 Les Tres Torres	9 El Guinardó
	13 El Raval	37 Montbau	25 La Marina de Port
	48 Sants-Badal	38 Montjuïc	11 El Poble-sec, Barcelona
		40 Porta, Barcelona	14 El Turó de la Peira
		41 Sagrada Família (Eixample)	57 Zona Franca - Port
		42 Sant Antoni, Barcelona	16 Horta, Barcelona
		43 Sant Genís dels Agudells	17 Hostafrancs
		44 Sant Gervasi - Galvany	18 L'Antiga Esquerra de l'Eixample
		47 Sants	19 La Barceloneta, Barcelona
		50 Torre Baró	22 La Font d'en Fargues
		51 Vallbona, Barcelona	21 La Clota
		52 Valcarlos i els Penitents	20 La Bordeta, Barcelona
		53 Vallvidrera, Tibidabo i les Planes	15 Fort Pienc
		54 Verdu	

The previous 3 tables correspond from left to right to: green cluster, purple cluster and red cluster.

Having acknowledged that the neighborhoods from the third cluster have less relative frequency of gyms, it would be a wise decision for the possible gym owner/creator to start his business in any of the 44 neighborhoods belonging to that cluster.

However, having such a wide range of options to choose from, it would possibly be convenient to follow with this research in the future, taking into account other variables like population per neighborhood, mean age, sex... which can all influence the probability of success of opening a gym and making it work properly.

7. CONCLUSION

During the development of this project, a business problem has been identified, and several procedures have been done to find a solution to it.

Starting of by web scraping and the obtention of geospatial data from neighborhoods in the place of interest, a clustering approach has been introduced in order to find the best possibilities for the people interested in this area.

As a final answer to the question proposed in the introduction of this report “Is it a good opportunity to start a gym business in Barcelona?”, YES, it is a fantastic opportunity, considering both the results obtained as well as the current situation in which the country is submerged.

The conclusions obtained with this research are going to help the relevant stakeholders to take advantage of the potential locations mentioned beforehand.

One of the main advantages of this project is that it can easily be extrapolated to other regions/countries and business areas, providing information of interest to a wide range of differentiated people.