Looking for the optimal place for a new restaurant in Barcelona

Lluís Puigdemont

June 2019

1. Introduction: Discussion of the problem and background

Barcelona is one of the European capitals with the greatest tourist attraction in the world. The combination of a good climate with the Spanish culture and its people make Barcelona an ideal destination for all kinds of people: friends, families, business trips...

New leisure establishments are constantly opening and competing fiercely with each other in order to survive the hard beginnings and make a name for themselves in the city. In the world of gastronomy, there is a wide variety of establishments: from bars that simply serve beer and tapas to luxurious restaurants that prefer to deal with a more selective and demanding clientele. Let's suppose we wanted to build a new restaurant in Barcelona. The strategy to be implemented has already been defined: the food to be served has been decided, as well as its prices. We have also determined the opening hours and the workers needed to correctly develop the different services that will be offered. The golden question here is: where should I locate my premises and why?

The location of the premises is one of the key aspects for a new establishment to be able to compete with the rest of the market taking advantage of the attributes of the surrounding area. There are many variables to consider: price of the area, size of the premises, population density…

To perform this exercise, we will basically focus on two of those attributes:

- The existing population density in each one of the neighborhoods of Barcelona.

- The similarity of the neighborhoods comparing them according to the main places of interest.

With these two variables we can place our premises in a neighborhood where the population density is high, making sure that as many people as possible enter our restaurant and, at the same time, avoiding those neighborhoods where the density of restaurants is too high so we can have less competition.

2. Data description

In order to carry out the analysis described above, the first step is to determine the data required to carry it out. For this purpose, the following repositories have been consulted:

- Data on the different neighborhoods in each district of Barcelona[1]: the city of Barcelona is divided into 10 districts, which in turn are divided into a total of 73 neighborhoods. In order to be able to add the postal code to each of the districts I had to do a manual work on the file as I could not find such information already added.

- Population density in Barcelona segregated by neighbourhood[2]: the repository contains information on the surface area of each neighborhood in Barcelona, as well as its population density (both gross and net). I have selected only the attribute corresponding to the number of habitants per unit area of the city of Barcelona (gross population density) and linked it to the previous repository, because it considers the effect of the tourism, and not only the residential density.

- Finally, I have used Foursquare API for looking for the main places of interest of each neighborhood, as well as Arcgis to determine the latitude and longitude of each neighborhood.

Thus, our main repository will contain the following attributes:

- Postal code referred to every neighborhood in Barcelona.
- Name of every neighborhood of Barcelona.
- Poblation: it will always be Barcelona.
- Population density: number of habitants per unit of surface in Barcelona.
- Latitude of each neighborhood.
- Longitude of each neighborhood.

Finally, we will add the venue data of each neighborhood using the Foursquare API.

Using every neighborhood of Barcelona and its common venues, we will perform a K-Mean algorithm and look for those neighborhoods whose most common venues aren't also restaurants, so we can set our new restaurant in a neighborhood without too much competence. At the same time, we will look to set our restaurant location in a neighborhood that, considering the previous condition, has the highest population density.

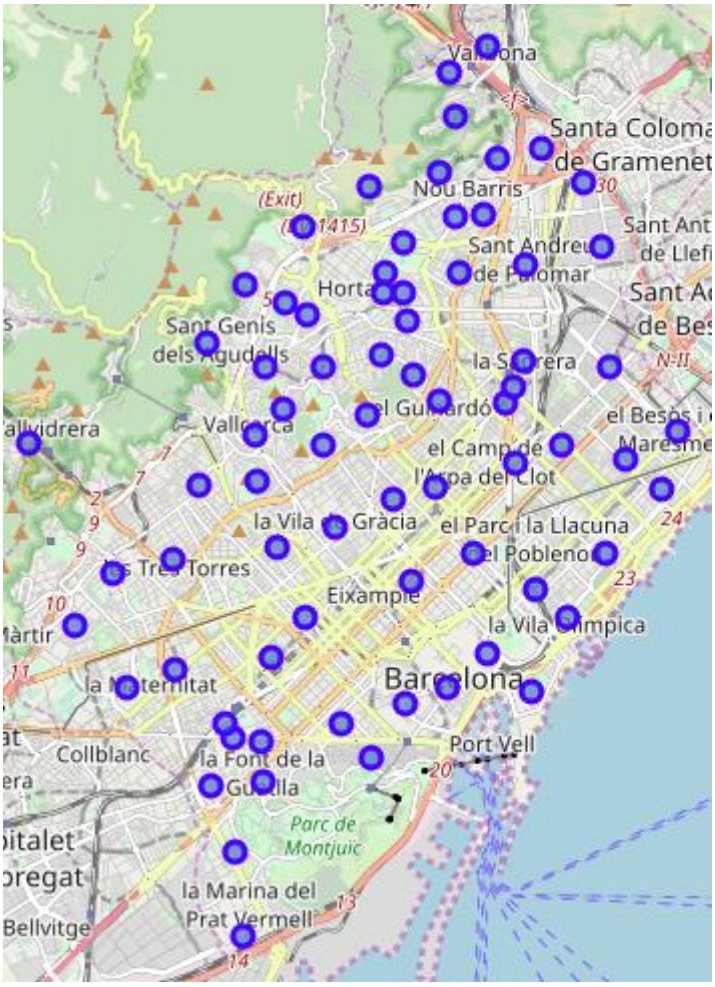
3. Methodology

The first step of the project consists of loading the information that contains the neighborhoods of Barcelona, the neighborhood code and the population density. To do this we use the Jupyter Notebook option to load information and convert it into a dataframe. In order to add the postal code belonging to each neighborhood, I had to manually search for the associated postal codes, as this information did not appear in a suitable format for the automatic loading of information. To do this, I have added in a new column of the dataframe the associated postal code of each neighborhood.

Then, to add the coordinates of longitude and latitude of each of the neighborhoods I used the functions of locating coordinates of *arcgis* within the *geocoder* library. With all this, I got the dataframe from which to start the analysis:

	Neig_Code	Neighborhood	Poblation_Density	Postal_Code	Latitude	Longitude
0	1	el Raval	433	08001	41.379673	2.169027
1	2	el Barri Gòtic	191	08002	41.381812	2.176201
2	3	la Barceloneta	114	08003	41.381132	2.190669
3	4	Sant Pere, Santa Caterina i la Ribera	204	08003	41.386169	2.183101
4	5	el Fort Pienc	345	08013	41.398863	2.180921

The next step was to build a map using the functions of the folium library, adding the coordinates of the city of Barcelona and superimposing the different neighborhoods that make up the city:



I then used Foursquare API to explore the main places of interest in each neighborhood and to be able to segment them. To do this, I first used a search function of the nearest venues of each neighborhood, setting as a search radius 750 meters and as a limit of 100 venues. In the following table you can see how each venue relates to the neighborhood to which it belongs:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	el Raval	41.379673	2.169027	Filmoteca de Catalunya	41.378540	2.171101	Movie Theater
1	el Raval	41.379673	2.169027	La Monroe	41.378795	2.170692	Spanish Restaurant
2	el Raval	41.379673	2.169027	A Tu Bola	41.380096	2.169054	Tapas Restaurant
3	el Raval	41.379673	2.169027	La Robadora	41.379500	2.170463	Gastropub
4	el Raval	41.379673	2.169027	33/45	41.381059	2.167399	Cocktail Bar

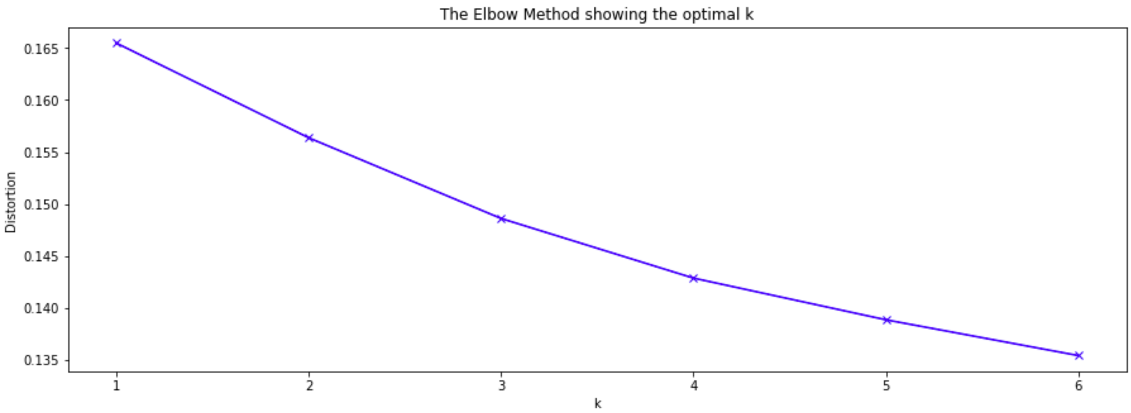
The total of different unique categories that appear is 291. By grouping the different venues by neighborhood and obtaining the average of venues for each neighborhood, I have obtained a table of 73 neighborhoods with the proportion of the 291 venue typologies that appear in it. An example of this table is as follows:

	Neighborhood	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant
0	Baró de Viver	0.000000	0.000000	0.00	0.032787	0.00	0.000000	0.00	0.000000
1	Can Baró	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.00	0.000000
2	Can Peguera	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.00	0.000000
3	Canyelles	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.00	0.000000
4	Ciutat Meridiana	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.00	0.000000

The next step was to analyze the most common venues in each neighborhood. For doing this we have used a function that returns the 10 most popular venues of each neighborhood and we have added it related to each of the neighborhoods in a new dataframe:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Baró de Viver	Clothing Store	Burger Joint	Italian Restaurant	Sandwich Place	Supermarket
1	Can Baró	Grocery Store	Plaza	Spanish Restaurant	Café	Soccer Field
2	Can Peguera	Tapas Restaurant	Grocery Store	Supermarket	Pizza Place	Spanish Restaurant
3	Canyelles	Soccer Field	Market	Mediterranean Restaurant	Skate Park	Café
4	Ciutat Meridiana	Metro Station	Park	Café	Mediterranean Restaurant	Supermarket

From here, we have been able to see with a visual analysis in the form of tables the distribution of venues in each neighborhood. We have detected different aspects: the difference in the amount of venues found and the difference in the type of venues for each one. The first aspect is due to the power of Foursquare API with the introduced parameters latitude, longitude, radius, limit... In different iterations different results can arise. The second aspect will be analysed below using the K-Means clustering method to analyse the similarity and difference of the different neighborhoods of Barcelona. One of the key parameters is the initial value of K, so before using the algorithm, I have tried to visually replicate the optimal choice of K using the Elbow method:

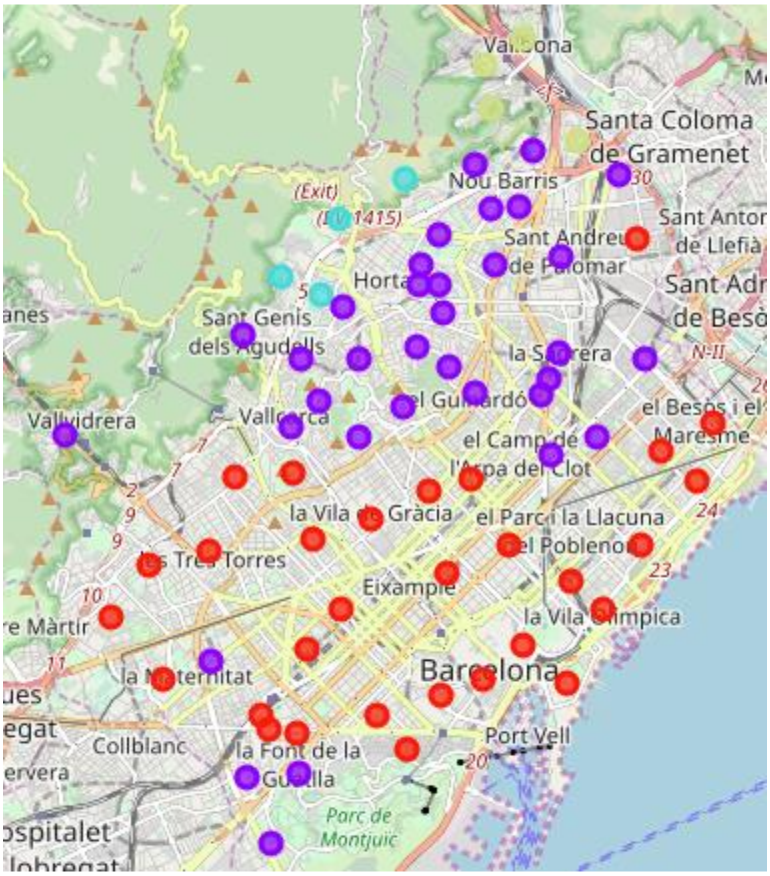


Although it is not very marked the elbow, we will assume that using the value of 4 for K we will have a good clustering of the neighborhoods.

So, I proceed to feed our model with the data obtained so far and we use the K means algorithm to cluster our neighborhoods according to the type of venues they have. The result is added to the dataframe and we obtain the following table:

	Neig_Code	Neighborhood	Poblation_Density	Postal_Code	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	1	el Raval	433	08001	41.379673	2.169027	0	Tapas Restaurant	Bar	Spanish Restaurant	Cocktail Bar	Restaurant
1	2	el Barri Gòtic	191	08002	41.381812	2.176201	0	Tapas Restaurant	Plaza	Spanish Restaurant	Cocktail Bar	Bar
2	3	la Barceloneta	114	08003	41.381132	2.190669	0	Tapas Restaurant	Mediterranean Restaurant	Paella Restaurant	Bar	Ice Cream Shop
3	4	Sant Pere, Santa Caterina i la Ribera	204	08003	41.386169	2.183101	0	Tapas Restaurant	Bar	Cocktail Bar	Wine Bar	Hotel
4	5	el Fort Pienc	345	08013	41.398863	2.180921	0	Restaurant	Spanish Restaurant	Coffee Shop	Hotel	Theater

The result of the clustering of the different neighborhoods of Barcelona according to their venues can be represented on a map in the following way:



In order to be able to detail the category that we want to associate to each one of the cluster labels, I have grouped the diagram by cluster label and type of venue, making a count of each venue to see which are the cluster labels with fewer restaurants as the most common areas. As an example:

	Cluster Labels	1st Most Common Venue	Neig_Code	Neighborhood	Poblation_Density	Postal_Code
10	0	Tapas Restaurant	9	9	9	9
0	0	Bakery	4	4	4	4
4	0	Hotel	4	4	4	4
6	0	Mediterranean Restaurant	3	3	3	3
9	0	Spanish Restaurant	3	3	3	3
8	0	Restaurant	2	2	2	2
1	0	Bar	1	1	1	1
2	0	Clothing Store	1	1	1	1
3	0	Garden	1	1	1	1
5	0	Italian Restaurant	1	1	1	1
7	0	Pizza Place	1	1	1	1
22	1	Tapas Restaurant	8	8	8	8
20	1	Spanish Restaurant	7	7	7	7
14	1	Grocery Store	5	5	5	5
18	1	Restaurant	3	3	3	3
21	1	Supermarket	3	3	3	3

If we look at the whole table, we can conclude that the neighborhoods where we will have less competition are those where restaurants are not part of the most common venues. These would be neighborhoods with cluster labels 2 and 3.

So, the neighborhoods in which we could put our restaurant for competition reasons are those that are in blue and light green:

Finally, we select each of the neighborhoods whose cluster label is 2 or 3 and we choose the one with the highest population density to comply with the condition of containing the maximum possible population density as well. The result is:

	Neig_Code	Neighborhood	Poblation_Density	Postal_Code	Latitude	Longitude	Cluster Labels	1st Most Common Venue
54	55	Ciutat Meridiana	291	08033	41.460682	2.176591	3	Metro Station

4. Results and discussion

As we have observed, after applying the K Means algorithm to our dataframe with the venues of each neighborhood, those neighborhoods in which a restaurant appeared in less frequency as most common venue were those belonging to cluster labels 2 and 3. The neighborhoods with this cluster label are:

- Horta
- Canyelles
- La Vall d'Hebron
- Montbau
- Ciutat Meridiana
- La Trinitat Vella
- Vallbona
- Torre Baró

In all these neighborhoods the restaurants did not appear as the most frequent place of interest. As you can see on the map, these neighborhoods have in common that they are not in the center of Barcelona, so it is normal that there is not a high concentration of restaurants in these areas.

On the other hand, if we observe the population density in these neighborhoods, we can see that, except for Ciutat Meridiana, they do not stand out for their magnitude. It makes sense to be as we have seen that those ara neighborhoods outside the center of Barcelona.

However, in the case of Ciutat Meridiana, population density it is above average (291>249.87).

If we base ourselves on the two conditions mentioned in the introduction, we can conclude that Ciutat Meridiana would be a good candidate as the neighborhood where we can locate our restaurant.

5. Conclusion

Competition between restaurants is the toughest in the neighborhoods of Barcelona. Although the analysis described above has reached a valid conclusion, it is still a simple model that does not take into account many other variables such as the cost of the premises, the power of the brand, the average price compared to competitors ...

Ciutat Meridiana has turned out to be a reasonable neighborhood in which to locate our premises if we are looking for less competition and a good influx of possible customers based on the population density of the area. However, before embarking on an operation that may entail a high initial cost, we should consider other factors and work on a good marketing strategy.

Even so, for all those adventurers who are curious to start a stage in the world of culinary restoration, they can use this report as a starting point to embrace an exciting adventure.

Bon appétit!

6. References:

1. https://opendata-ajuntament.barcelona.cat/data/es/dataset/20170706-districtes-barris/resource/4cc59b76-a977-40ac-8748-61217c8ff367
2. https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-densitat/resource/c2377d82-774c-4d54-8e56-6c8978189df9