

Where to open an Italian Restaurant in Barcelona

Introduction

Barcelona is a global city in the coast of Mediterranean sea. With an incredible coast, a warm climate, historical monuments and a growing economy it has become the 5th most visited city in Europe and 20th in the World.

It has 519 hotels (March 2016) including 35 five stars hotels and a developed tourist infrastructure also with a high density of international restaurants, bars, pubs, bakeries or coffee shops.

One of the critical aspects for the success of a restaurant in a city like Barcelona is where to place it. Specifically, this project aims to predict whether a new Italian restaurant in Barcelona should be placed in order to have more probabilities to success.

Obviously, any person trying to start a business like that or any person or company interested on investing on this kind of business would be interested on that.

Data

Data about the Neighborhoods of Barcelona, their districts, population, territory, density, urban territory and net density can be found in the next link:

<https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-densitat/resource/08c88565-5b50-47d1-8234-6d3dd2da4c58>

For the purpose of this project, urban territory and net density will be ignored as we are interested in the total population, territory and density of the city and its neighborhoods.

After importing the data, we've used *geopy* function to get the coordinates for each neighborhood in the city of Barcelona. In order to do that we've renamed the neighborhood of *Vallvidrera, el Tibidabo i les Planes* by only *Vallvidrera*, otherwise it was not found. In addition, we've added the name of the district and the name of the city to get the right coordinates instead of the ones from some neighborhoods out from Barcelona city. This happened because Barcelona is not only the name of the city but also from the province and some neighborhoods share names with other towns in the province.

As an example, for the neighborhood of Sant Martí de Provençals we are searching Sant Martí de Provençals, el Clot, Barcelona, which are the name of the neighborhood, the name of the district and the name of the city, respectively.

Moreover, for those neighborhoods that have the same name than the district we are only using the name of the neighborhood without specifying the district name. One example of this is Sant Andreu neighborhood which district is also named Sant Andreu, therefore we've just done the search as Sant Andreu, Barcelona, skipping the district name.

With that, we've been able to find the latitude and longitude of each Neighborhood.

Once we had that, we've used Foursquare API to find 100 venues in a radius of 1000 meters from the coordinates of each neighborhood.

Methodology

First, we’ve explored how is the city of Barcelona shaping in terms of total population, territory, density, neighborhoods and districts. Density has been calculated as we cannot add density of different neighborhoods to get the total density of the city.

```
Total population of Barcelona is: 1636762
Total Area of Barcelona is: 10135.7
Density of Barcelona city is: 161.48485057766112
Total number of Neighborhoods are: 73
Total number of Districts are: 10
```

Then, we’ve explored the population and density of the different neighborhoods considering the fact that it would not be a good idea to open a restaurant in a very low populated or very low dense neighborhood and it would be more difficult to get clients.

	District	Neighborhood	Population	Area (ha)	Density
41	Horta-Guinardó	la Clota	683	17.8	38
11	Sants-Montjuïc	la Marina del Prat Vermell	1196	1403.4	1
55	Nou Barris	Vallbona	1405	60.8	23
46	Nou Barris	Can Peguera	2288	11.9	193
57	Sant Andreu	Baró de Viver	2597	23.0	113

	District	Neighborhood	Population	Area (ha)	Density
11	Sants-Montjuïc	la Marina del Prat Vermell	1196	1403.4	1
21	Sarrià-Sant Gervasi	Vallvidrera, el Tibidabo i les Planes	4698	1132.9	4
53	Nou Barris	Torre Baró	2914	174.2	17
55	Nou Barris	Vallbona	1405	60.8	23
39	Horta-Guinardó	Montbau	5109	205.5	25

Our approach is then to identify which are the 5 most populated and 5 most dense neighborhoods in the city of Barcelona as these would be potentially the ideal areas to open a restaurant.

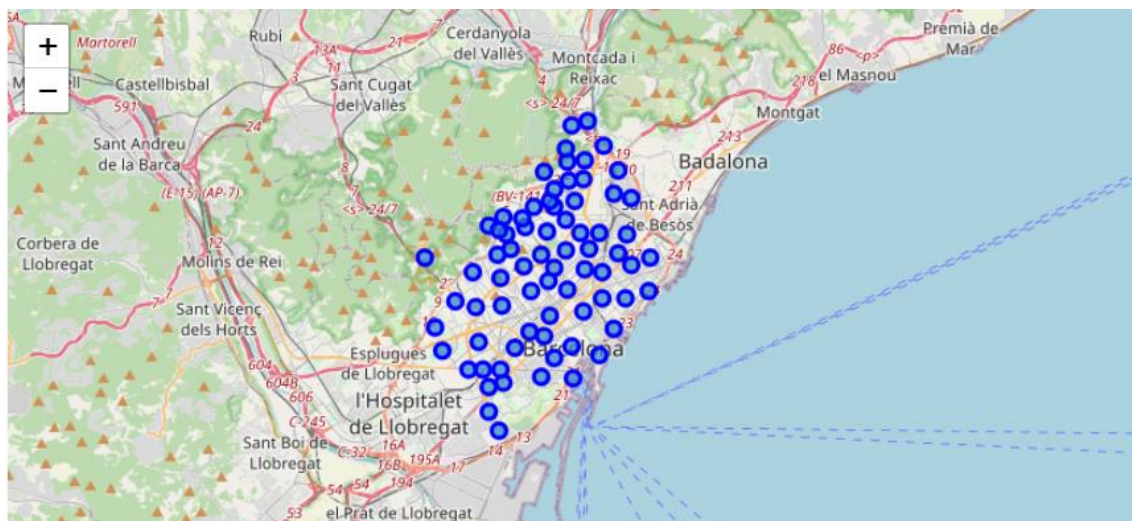
	District	Neighborhood	Population	Area (ha)	Density
8	Eixample	la Nova Esquerra de l'Eixample	58032	134.1	433
59	Sant Andreu	Sant Andreu	57843	186.7	310
5	Eixample	la Sagrada Família	51385	104.2	493
30	Gràcia	la Vila de Gràcia	50102	132.1	379
25	Sarrià-Sant Gervasi	Sant Gervasi - Galvany	47588	166.9	285

	District	Neighborhood	Population	Area (ha)	Density
16	Sants-Montjuïc	Sants - Badal	24312	41.5	586
31	Gràcia	el Camp d'en Grassot i Gràcia Nova	34911	65.1	537
50	Nou Barris	Verdun	12531	23.7	528
62	Sant Andreu	Navas	22176	42.4	523
63	Sant Martí	el Camp de l'Arpa del Clot	38663	74.1	522

However, we would not just stick to this information as in reality it may happen that these neighborhoods have already a long list of Italian Restaurant so that we should avoid to place our restaurant there.

For that reason, the idea is to group the different neighborhoods by clusters made from similarities in terms of venue categories. Once we get that, we'll choose to place the new Italian restaurant in the cluster that already has more Italian restaurants as for some reason looks like the place where this kind of restaurant succeed. However, from this cluster we'll pick the neighborhood that has the lowest density of Italian restaurant by population as it would be the place were there are less competitors.

In order to get this data we will first find the coordinates of each Barcelona neighborhood using geopy and paint them in a map:



Once we've located the neighborhoods in the map , we've searched for each of them the 100 venues in a radius of 1000 m from the neighborhood center location and we've found a list of 5851 different venues distributed in 288 unique categories.

We've exported to excel the list of categories with the number of occurrences and group some of them by one single category, i.e. Italian Restaurant and Pizza Place, the Tapas Restaurant and Spanish Restaurant or Gym and Fitness Centers. This would allow us to have a more accurate clustering of neighborhoods and relevancy of data.

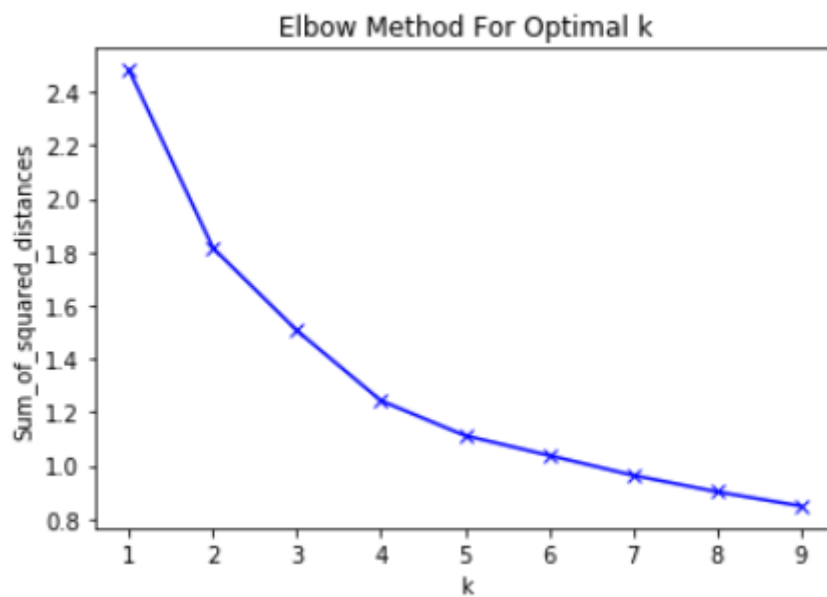
After doing that, we've reduced the 288 unique categories to **26** categories.

Then we've explored which are the most frequent categories for each neighborhood. As an example, these are Sant Andreu neighborhood top 5 categories:

----Sant Andreu----

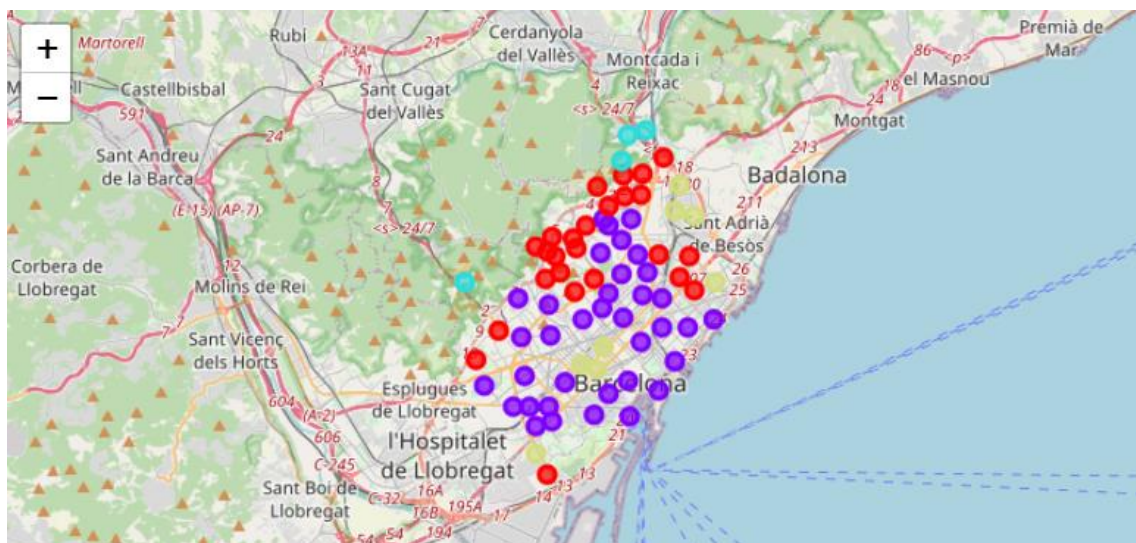
	venue	freq
0	Non-food stores	0.24
1	Spanish restaurants	0.16
2	Markets and food stores	0.14
3	Bakeries and coffee places	0.11
4	Fast food restaurants	0.09

Now it is time to cluster the neighborhoods. To do that we've selected the top 10 grouped venue categories for each neighborhood and performed an Elbow Method analysis to find the optimal number of clusters:



In this case we've selected 4 number of clusters.

Let's finally plot the different clusters in a map of Barcelona to see how the different neighborhoods have been grouped:



Then, we've found that Cluster 1 is the one with more Italian Restaurants:

Italian restaurants	
Cluster Labels	
0	54
1	207
2	0
3	23

and from there the Neighborhood with lowest density of Italian Restaurant is *La Nova Esquerra de l'Eixample*

	District	Neighborhood	Population	Area (ha)	Density	latitude	longitude	Italian restaurants	Cluster Labels	Italian_density
6	Eixample	la Nova Esquerra de l'Eixample	58032	134.1	433	41.382816	2.149966	3	1	0.000052
0	Ciutat Vella	el Raval	47353	110.0	430	41.379518	2.168368	3	1	0.000063
19	Gràcia	la Vila de Gràcia	50102	132.1	379	41.403178	2.157166	4	1	0.000080
4	Eixample	el Fort Pienc	32349	92.9	348	41.395925	2.182325	4	1	0.000124
13	Les Corts	les Corts	46274	141.0	328	41.385244	2.132863	6	1	0.000130

Conclusion

In this study we've explored the city of Barcelona with the objective to find the best neighborhood to open a new Italian Restaurant.

We've clustered the neighborhoods based on their venue categories and identified in which cluster we can find more Italian Restaurants. This means that for some reason this should be the cluster where an Italian restaurant has more options to succeed.

Finally, we've analyzed which Neighborhood from this cluster has the lowest density of Italian Restaurants by population so we make sure not to open the restaurant in an already overloaded neighborhood.

Discussion

We've noted that foursquare API is very useful to get different venues close to a given coordinate. However, the way it categorizes some of them is not very accurate. Having 288 different categories for 5821 venues is not really useful for clustering as for example there were 44 categories with one unique venue. For that reason we think that it is needed to perform an analysis of the obtained categories and group some of them to get more relevant data.

Also, we've seen that most of the clusters follow a geographical aggregation which is a good sign as neighborhoods in Barcelona are close to each other and usually there are not much differences between close ones.