1.- Introduction and Business Problem

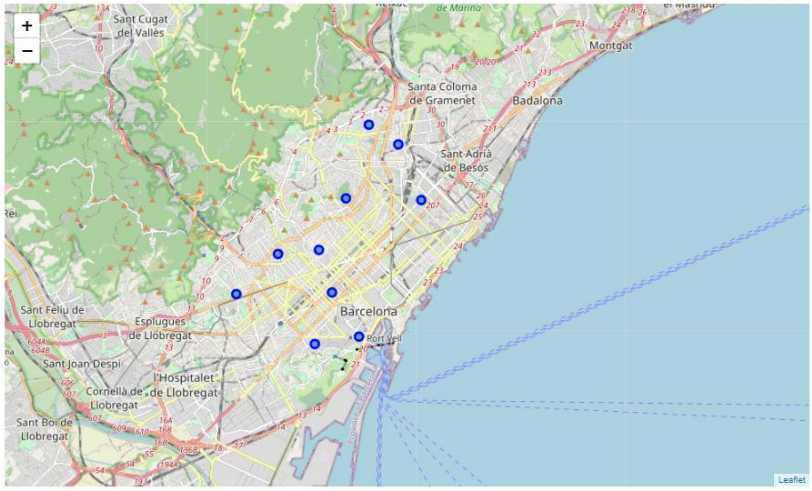
Barcelona is one of the most important nodes of tourism in the Mediterranean Sea, therefore the offer of accommodation facilities is incredibly competitive, and business related to it rise and fall very quickly.
My client is planning to open a Bed & Breakfast with a tiny bar to attract young tourists into a nicer and closer ambient than the normal hotels can offer. The location of both hospitality and restaurant facilities near the potential ubication are key factors for success. This work will determine the idoneal position of this establishment.

2.- Data

The data required to this project is essentially:
1.- Coordinates of Hotels and Hostals
2.- Names and Coordinates of Neighbourhoods
3.- Cultural venues and traffic in some tourist areas
The Names and Coordinates of Neighbourhoods can easily be scraped from Wikipedia or the Townhall page. They will be obtained by *selenium* and *selenium.webdriver.*
The information of both the hotels and the traffic will be obtained by the Foursquare API and both datasets will me merged in one.
A K-means clustering will be performed to classify the characteristics and get the similarities of the different districts and zones. Once that is done, we will see wich one is better suited to have this kind of establishment, essentially because of the absence of other like it.

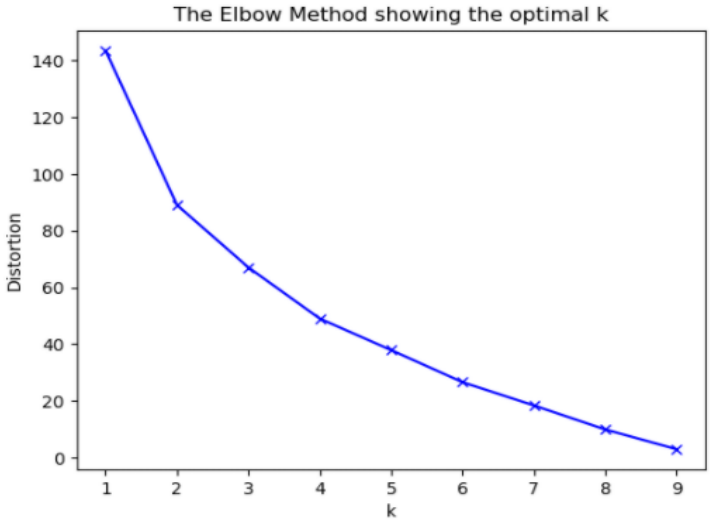
3.-Methodology

In the first place, an extractive process of the data has been carried out using selenium, first accessing Wikipedia to obtain data from the city's neighborhoods and then Foursquare to fill in these places with useful information. As we seek to locate a hostel with a bar in the ideal place, we are going to try to locate the ideal neighborhood within the cluster of neighborhoods that most interests us.



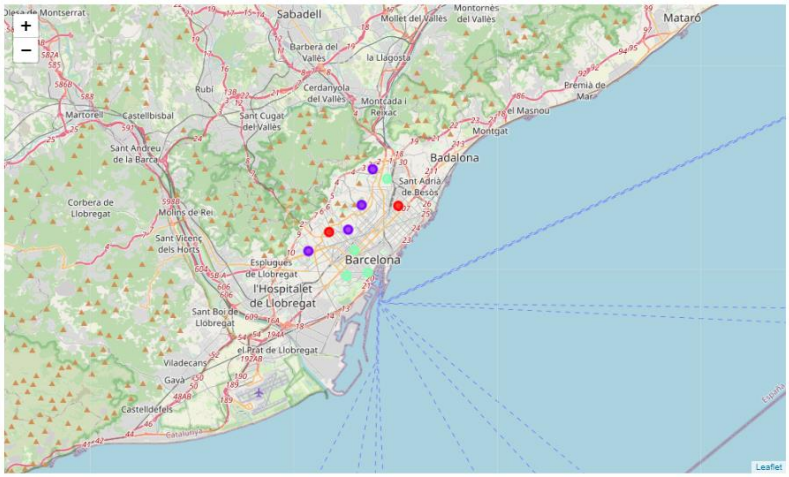
1. Centroids of Neighbourhoods

Second, once all the data has been obtained, we have carried out a normalization and clustering by the K-means method, with the aim of ruling out neighborhoods that contain characteristics that we are not looking for, that is, those with many hotels and many restaurants. that suppose a clear excess of competition to the place that is trying to open. Next we can see how we have decided to use a K = 3 for the K-means.



2.Optimal selection via Elbow Method

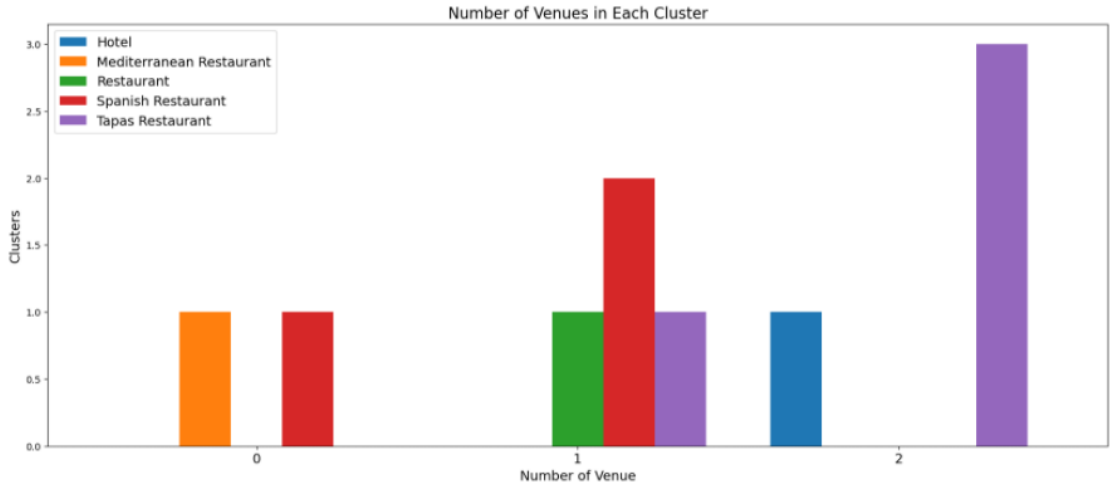
Below we can see the graphical result of the clustering in the neighborhoods.



3.K-means result

The groups have more or less precisely coincided with certain geographical positions, leaving a clear cluster in the interior, another in the southern part of the coast and the rest further inland, more mixed. This is really totally accidental and not causal, since the information that has been used to carry out the clusters is the type of business that exists in the zones.

We have worked the densities of the businesses in these clusters to decide to eliminate certain neighborhoods. The graph is as follows:



4.-Results

As we can see on the previous graphs, the ideal cluster to place a Hostal with a bar will be the cluster 0, that corresponds with Sant martí and Saint Gervasi, the two red dots.

We can analize them closely to figure out wich one is better suited to hold the Hostal, Based on the rest of venues:

```
----Sant Martí----
                      venue  freq
0         Spanish Restaurant  0.08
1                Supermarket  0.06
2   Mediterranean Restaurant  0.05
3                      Diner  0.05
4                     Bakery  0.05
5           Tapas Restaurant  0.05
6      Performing Arts Venue  0.03
7                Pizza Place  0.03
8                 Restaurant  0.03
9               Burger Joint  0.03
```

```
----Sarrià-Sant Gervasi----
                      venue  freq
0   Mediterranean Restaurant  0.08
1         Spanish Restaurant  0.08
2                 Restaurant  0.07
3         Italian Restaurant  0.06
4                       Park  0.05
5        Japanese Restaurant  0.04
6                Pizza Place  0.04
7                      Hotel  0.03
8         Seafood Restaurant  0.03
9       Gym / Fitness Center  0.03
```

We will choose Sant Martí to place the Hostal with bar because of this information, it has less hotels and hostals and also the offer of night activity is bigger, so the environment is perfect for the task.

5.-Discussion and Conclusion

As we can see from the results, the techniques that were applied resulted satisfactory. This clustering shows lots of information that would help to tons of business, not only on this particular area but on most of open-public relied on businesses.

For future work, I will suggest to make bigger divisions or even try other cities and compare even bigger clusters of neighbourhoods.