A social-economic comparative between Barcelona and Madrid

Jose Marquez

December 20, 2020

1. Introduction

Madrid and Barcelona are the largest cities in Spain. Madrid is the administrative capital and Barcelona the industrial one. The Spanish government is interested on understanding how different or similar are these cities in order to study if they can apply the same actions on both cities or not.

For a better understanding of the problem, we first study each city individually. Then, we compere the results on both cities and try to find similarities between the cities or between districts of the different cities.

The annual net income per capita is a good indicator to distinguish between the rich and poor districts of the city. We can also use as an indicator the number of services that can generate a good quality of life. A neighborhood may not have a high income but be well equipped. We are going to consider the following services as indicators of a good quality of life: hospitals and other medical centers, pharmacies, police stations, supermarkets, parks, playgrounds, Athletics & Sports venues, post offices, parking, event spaces, Arts & Entertainment venues, music venues, libraries, cultural centers, music schools and language schools (as an indicator of the education since there are schools and high schools in almost all districts).

2. Data

In order to study this case, we need first to define the data that we need and find the data sources. Then, we work on them to create the dataset that we will use for our purpose.

First, we are going to study the income in each district for each city. So, we need to get the information about the income that are distributed geographically. Searching on internet, we found the income for each postal code in Spain in 2018 being the data source the site https://www.epdata.es/ which is a portal (in Spanish) that contains many studies and data sets on population, unemployment, income, crime and many other subjects. The data is stored on a csv file. To visualize this information on a map we need the coordinates of each postal codes. To do this we need to know the postal codes of each city, the information was found on the site http://es.postcode.info/ and then we get the coordinates using the geocoder python library for these postal codes.

Second, to study the venues in each district of the city. For this purpose, we will use the FourSquare API to get the information needed. Knowing the coordinates of each district we can explore the selected venues of any type around.

In our study we will plot data about the different districts on a map, so we will need two json files with the geographical districts limits for each city. We found some files on the internet but they need some modifications to fit our problem (this is why I'm not giving the source)

3. Methodology

To easily identify the quality of the life in each district, we will plot the data using a choropleth map for the income and for the total number of "quality" venues. We will discuss the insights for each city trying to understand where are a good standard of living.

Then, we can group each district into clusters with similar characteristics. Like this we can find which districts are similar inside each city, and how if there are similar districts between the cities.

With this information we will be able to decide if we can take common actions in both cities and which actions, we could take in order to improve the standard of living in the poor districts.

We have the net mean **income** for each district for both cities and the district coordinates, so we can plot a **choropleth** map to visualize the geographical income distribution having a better understanding on how the wealth is distributed inside each city.

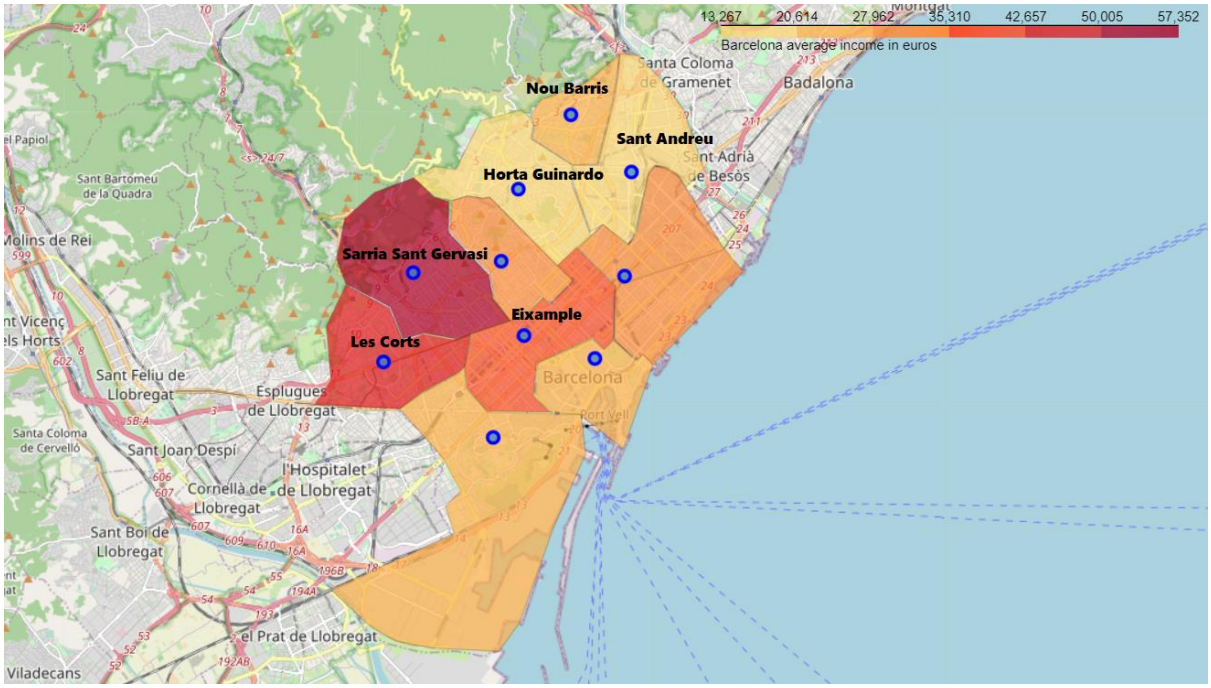

Fig 1. Barcelona net mean income distribution by district.

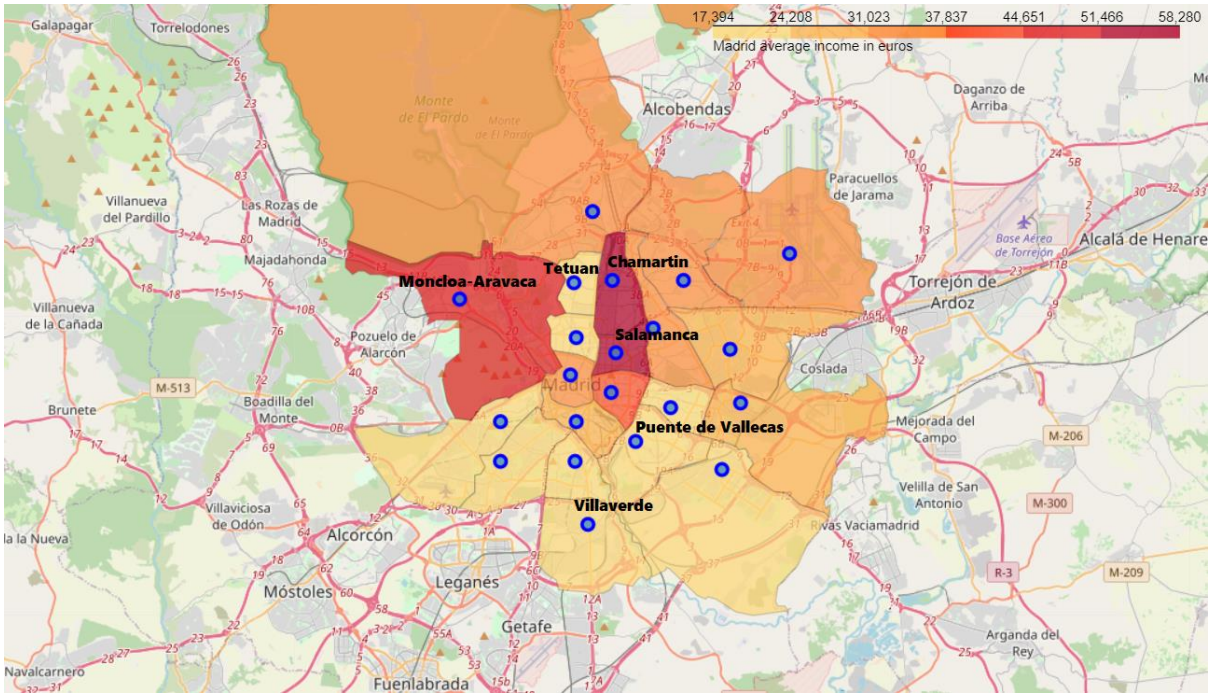

Fig 2. Madrid net mean income distribution by district.

We can identify the wealthy and poorest districts in both cities. In **figure 1** we observe that the wealthy districts of Barcelona are across the Diagonal Avenue, with the highest income being at the east side of the city, Sarria-San Gervasi (56.920 €), Les Corts(48.995,50 €) and the center the Eixample(38.479,70 €)) while the poorest are Sant Andreu(13699 Eur), Horta Guinardo(17.905 €) and Nou Barris(21.154,25 €) all of them in the north side of the city.

In **figure 2**, we have the choropleth map for the net mean income of Madrid. The wealthy districts in Madrid are Salamanca (57.879 €), Chamartin (54.030 €), Moncloa-Aravaca (51.101,33 €) and poorest are Puente de Vallecas (17.795 €), Tetuan(18.031 €) and Villaverde (19.709 €)

Using the data from **FourSquare** we can have a look on **venues related with the life quality and try to understand how them are related to the income distribution**. We consider that these venues should be related with health, education, cultural activities, places where people can feel away from the anxiety of the city like parks, security and other services. From all the venues that we can find on FourSquare we have selected the following: park, playground, athletic and sport, parking, post office, police station, pharmacy, medical center, hospital, supermarket, bookstore, library, music venue, cultural center, art and entertainment, event space, music school and language school (these two latest as an indicator for educational services of the district since almost all districts have schools and high schools)

From all these some are relevant, potting another **choropleth map** with the number of these venues per district we can visualize how they are related to the income distribution.

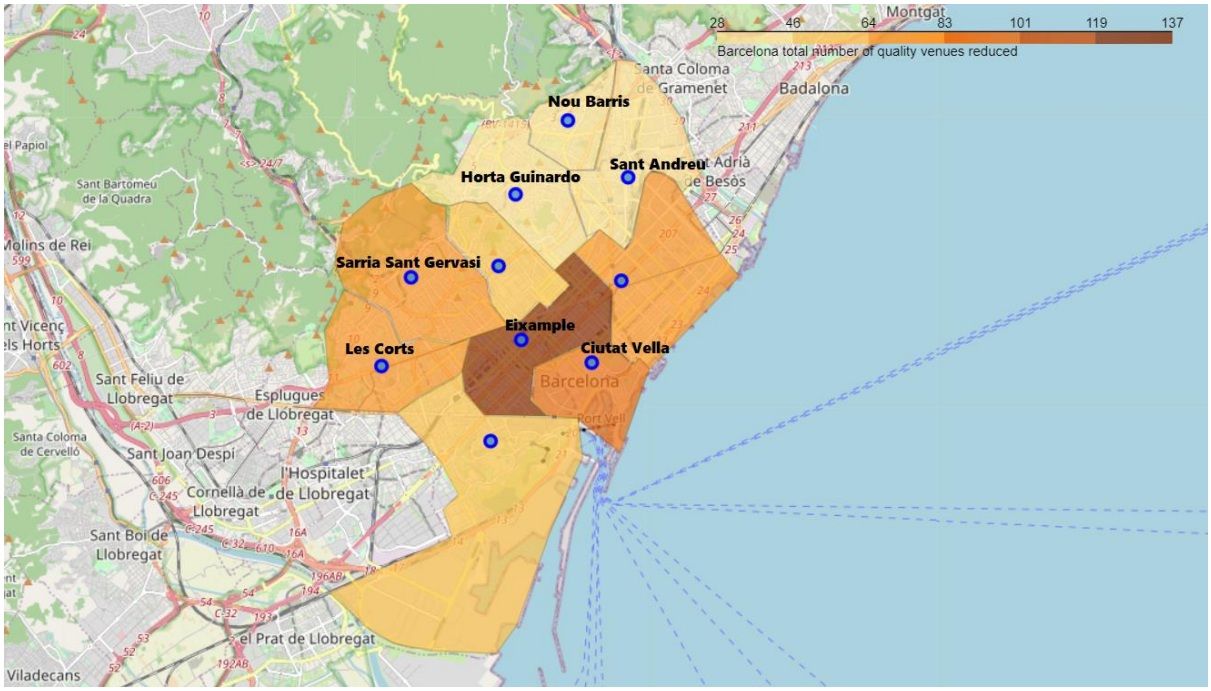

Fig 3. Barcelona venues distribution by district.

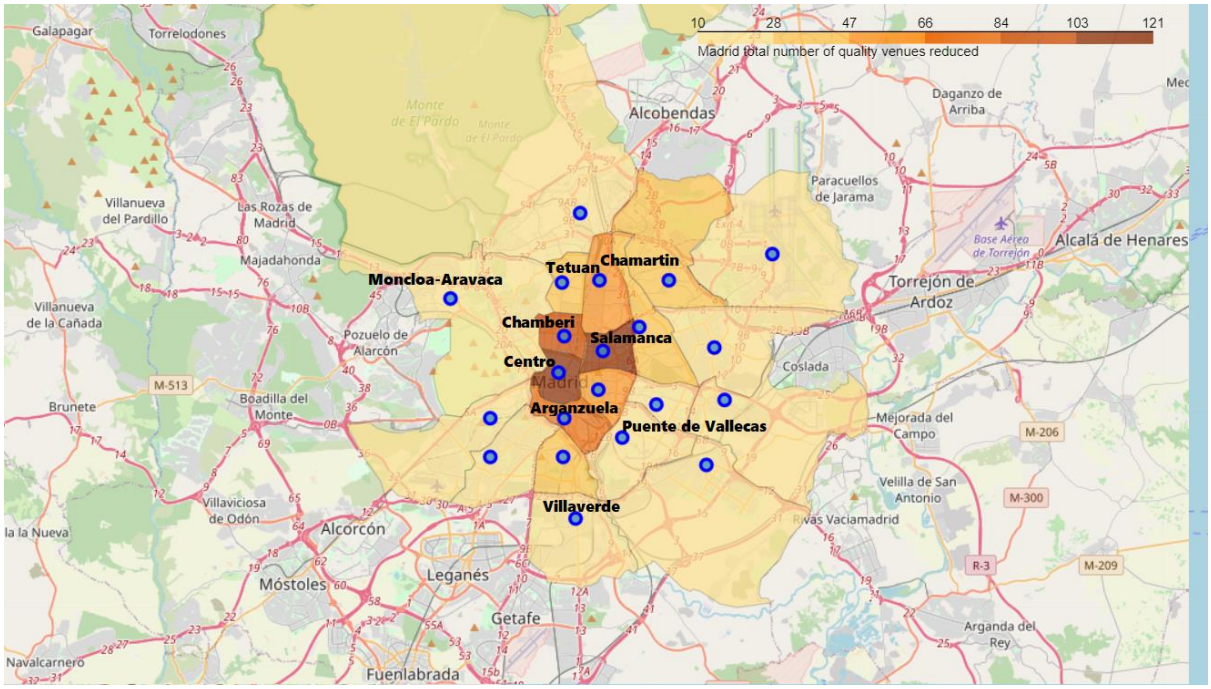

Fig 4. Madrid venues distribution by district.

In figure 3 and figure 4 we can observe how well correlated are these venues with the income, Especially for Barcelona districts. Madrid has some exception like Moncloa-Aravaca that seems to has less services than expected or for the contrary Tetuan or Chamberi have more than expected. The two latest maybe due to proximity to the city center where the vast majority of venues are located.

Applying **k-means** algorithm we can **group the different districts of each city in clusters** that would share **similar characteristics**. After fining tune the model and training it with the data set, we can find the following results:

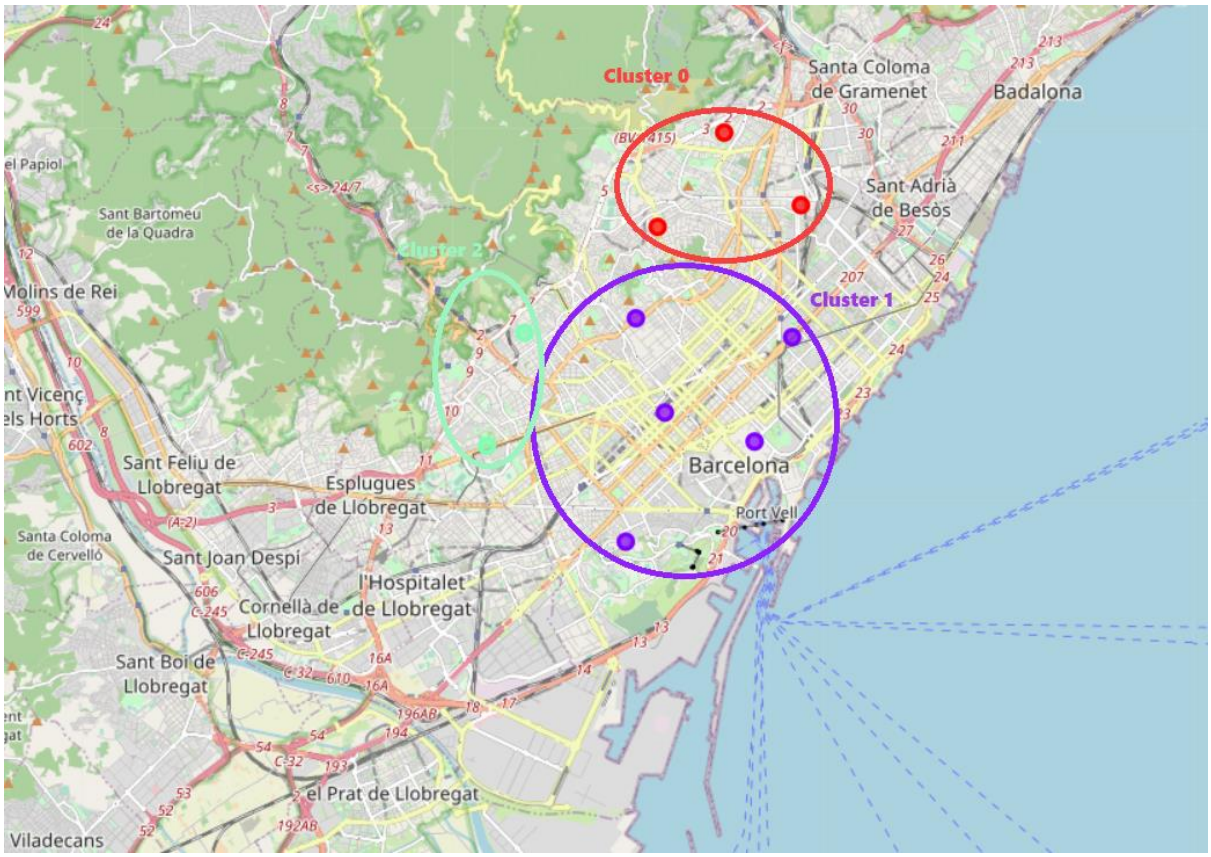

Fig 5. Barcelona districts grouped by similar characteristics.

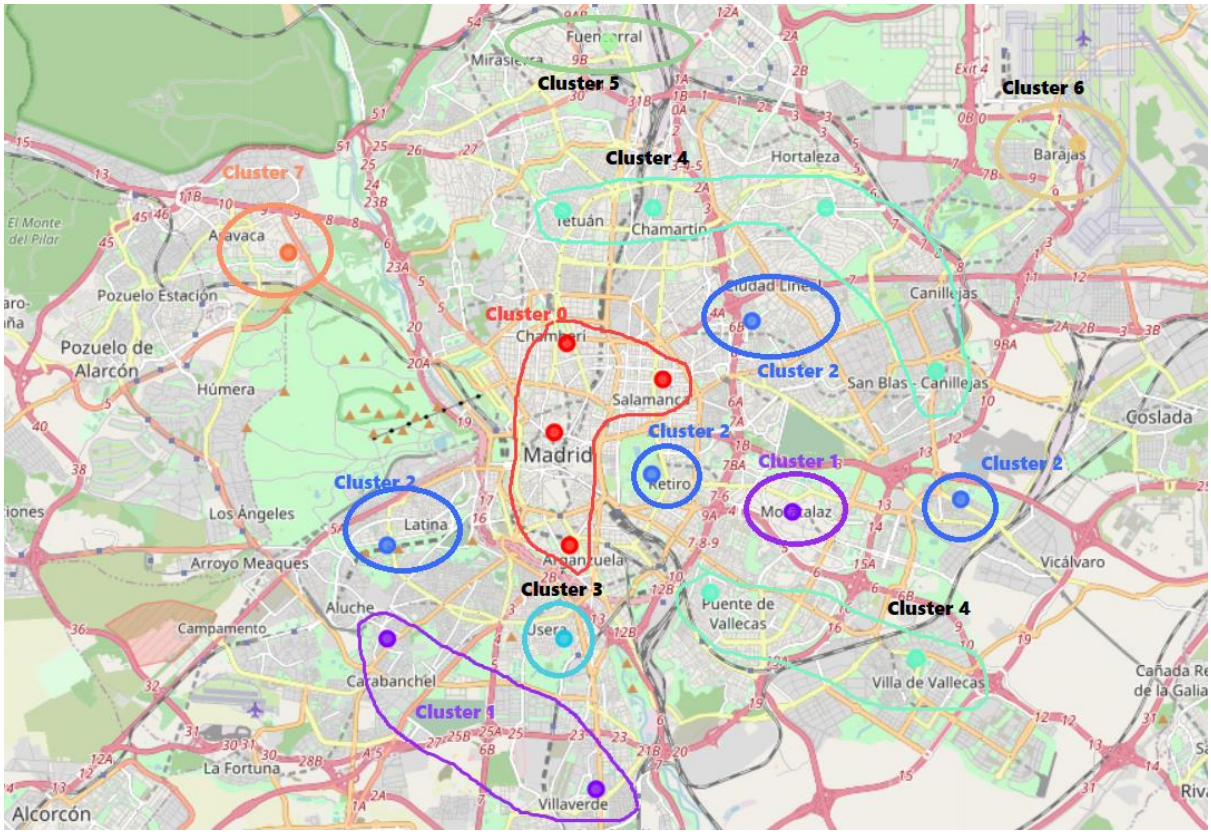

Fig 6. Madrid districts grouped by similar characteristics.

In **figure 5** we observe that the **poorest** districts of **Barcelona** are grouped in the **same cluster** so share similar characteristics. In **figure 6** shows that **Madrid** has more **heterogeneous poor districts**. They have been grouped in three different clusters.

To find similarities between districts of Madrid and Barcelona we merge the data sets into one a repeat the process with the k-means. In this case:

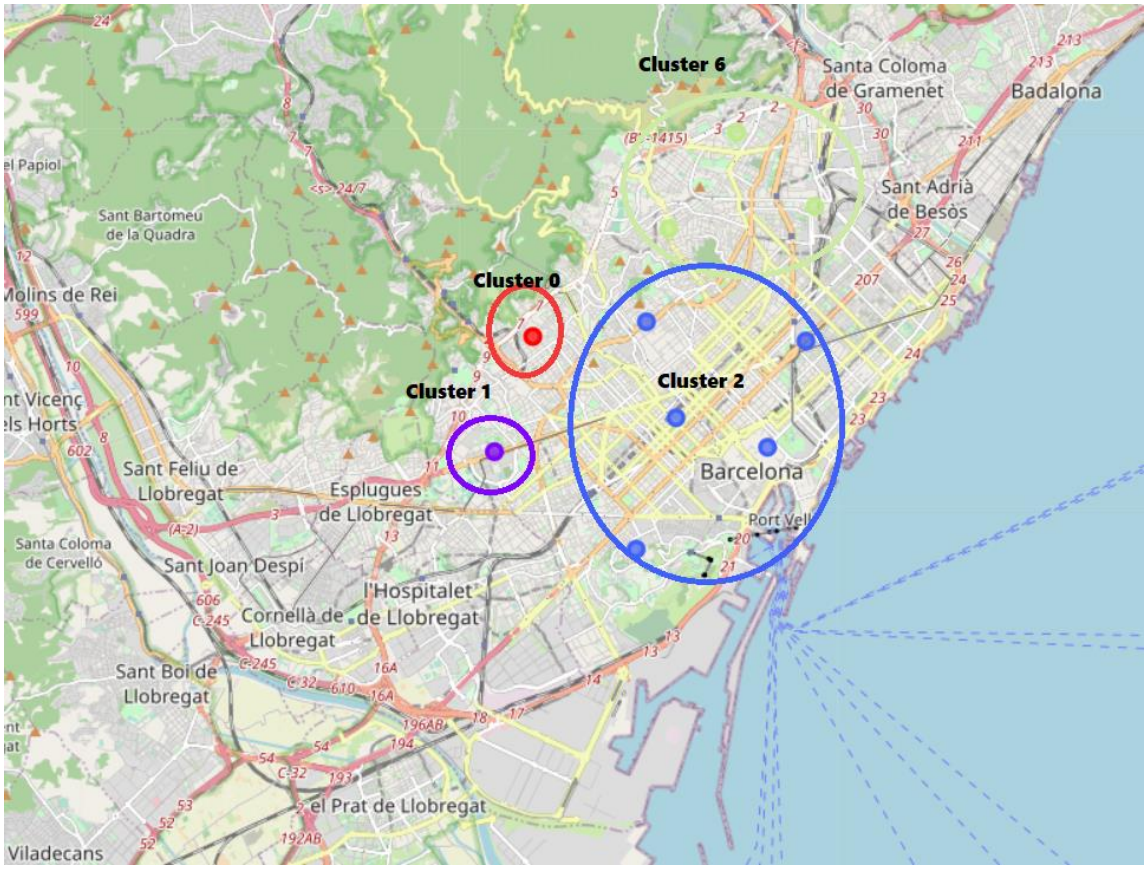

Fig 7. Barcelona clusters when grouping with the Madrid districts.

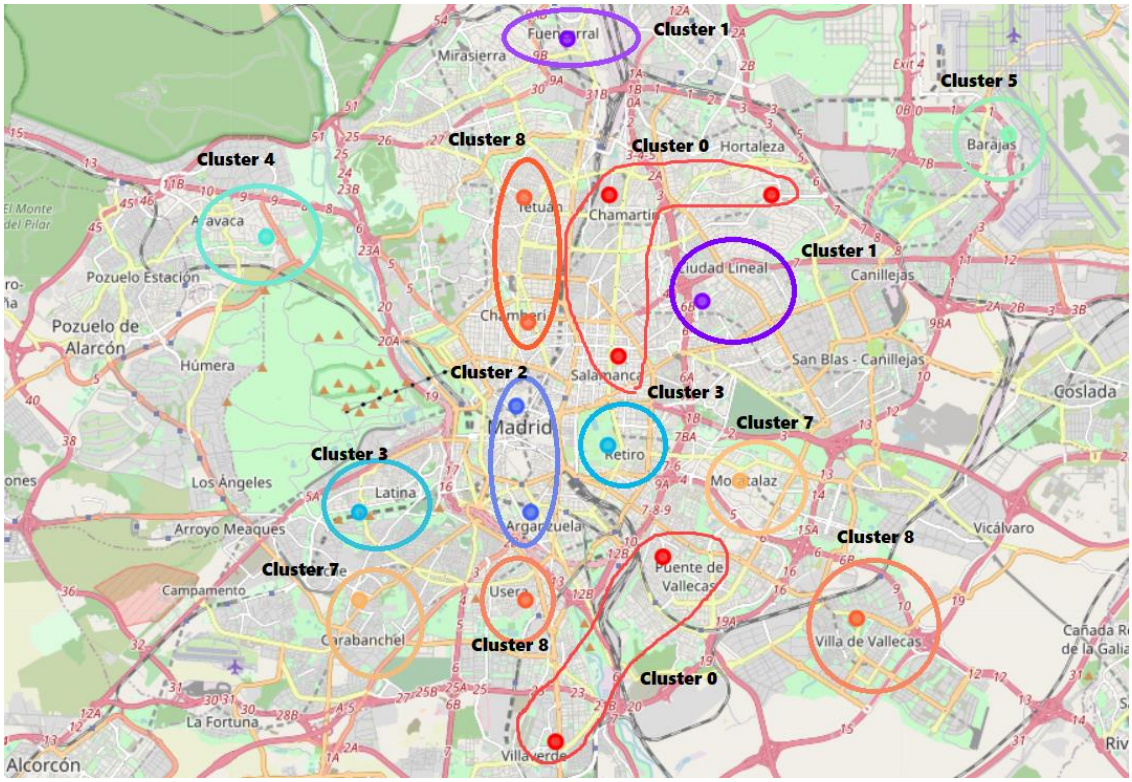

Fig 8. Madrid clusters when grouping with the Barcelona districts.

In **figure 7** and **figure 8** we can observe that the **poorest districts of Barcelona still** stick together on the **same cluster**, while now the **poorest districts of Madrid are grouped into two different** clusters. The cluster for the **poorest districts of Barcelona** is **different from** the two clusters for the **poorest district of Madrid**. Also, we can observe that the city **center of Madrid are quite similar to the city center of Barcelona** sharing both the **same cluster**. Also, two of the wealthy districts of Madrid, Salamanca and Chamartin, share the same cluster with one of the wealthy districts of Barcelona, Sarria-Sant Gervasi.

We can **identify which venues are lacking the poorest districts** comparing with the wealthy ones> In order to do this, we compare the top ten venues for each district.

For **Barcelona**:

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Eixample	Medical Center	Event Space	Art Gallery	Gym / Fitness Center	Gym	Park	Language School	Hospital	Yoga Studio	Theater
Sarria Sant Gervasi	Medical Center	Doctor's Office	Event Space	Park	Dentist's Office	Cultural Center	Music School	Gym / Fitness Center	Hospital	Language School

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Horta Guinardo	Park	Cultural Center	Language School	Event Space	Medical Center	Dentist's Office	Music School	Gym / Fitness Center	Athletics & Sports	Doctor's Office
Nou Barris	Event Space	Park	Medical Center	Gym	Cultural Center	Language School	Gym / Fitness Center	Gym Pool	Escape Room	Doctor's Office
Sant Andreu	Park	Event Space	Medical Center	Language School	Cultural Center	Gym	Soccer Field	Dentist's Office	Music School	Veterinarian

The poorest districts of Barcelona seem to lack of **health care** services.

For **Madrid**:

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Chamartin	Medical Center	Event Space	Doctor's Office	Park	Hospital	Gym / Fitness Center	Gym	Language School	Dentist's Office	Cultural Center
Salamanca	Medical Center	Gym / Fitness Center	Doctor's Office	Event Space	Gym	Hospital	Language School	Music School	Cultural Center	Dentist's Office

Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Puente de Vallecas	Medical Center	Gym	Event Space	Park	Athletics & Sports	Dentist's Office	Language School	Rock Club	Doctor's Office	Concert Hall
Tetuan	Gym	Medical Center	Dentist's Office	Gym / Fitness Center	Doctor's Office	Dance Studio	Park	Event Space	Veterinarian	Gymnastics Gym
Villaverde	Medical Center	Park	Dance Studio	Art Gallery	Event Space	Doctor's Office	Dentist's Office	Veterinarian	Basketball Court	Athletics & Sports

The poorest districts of Madrid seem to lack of cultural and educational venues.

4. Discussion

Regarding the distribution of the net mean income, we can observe that the wealthy districts in Madrid are located west and east of the city center, while the poorest are located in the south of Madrid and at the north of the city center. Barcelona has a different topography from Madrid. While Madrid is located in the center of the Iberian Peninsula, Barcelona is in the Mediterranean cost. That can condition how the income is distributed in each city. Barcelona net incomes seems to be distributed along the Diagonal, the main venue that cross the city from the north-west to the south-

east. The wealthy districts in Barcelona are over this avenue meanly in the north-west side and the neuralgic center of the city, while the poorest are all located in the north-east of the city.

When analyzing the venues distribution, we found that many of them are more correlated with their income. After dropping the less relevant venues we can see that for Barcelona the matching between the reduced venues map and the income are very close. Madrid shows a less clear image than Barcelona. It seems that Madrid concentrated the most of the venues close to the city center independently for the income level of the districts, while Barcelona seems to be more homogenously distributed. This information will be helpful when discussing the similarity between both cities or the possible actions to improve the poorest districts.

Regarding the clustering of the Madrid districts, the poorest districts are separated into three different clusters, being the cluster 1 for Villaverde the one most close to the income image, and for the wealthy one cluster 7 where the district Moncloa-Aravaca was located. In the case of Barcelona there are three clusters that meets prety well the income distribution of the city. On cluster 0 we have all the poorest districts Sant Andreu, Horta Guinardo and Nou Barris, on cluster 1 the Eixample and other medium income level districts and on cluster 2 the two most wealthy districts Sarria-San Gervasi and Les Corts. So, we can clearly identify that the poorest districts of Barcelona are very similar, while for Madrid are distributed into heterogenous gorups. When analyzing together all the districts for both cities, the image of Barcelona didn't change, while Madrid get closer to the income distribution. As we can see for Madrid, now the poorest district is only distributed into two groups, but still are mixed with other wealthier districts. Tetuan is inside the culster 8 and Villaverde and Puente de Vallecas are inside cluster 0. The only similarity between both cities is located in their city center. Both city center is located into the same cluster, number 2, which consists on Madrid-Centro/Arganzuela from Madrid and Eixample/Ciutat Vella/Gracia/Sant Marti/Sants-Montjouic. Any poor district is related between both cities.

Having a look to the top ten venues for each district, Barcelona poorest districts maybe have a lack of health care services comparing with the wealthy ones. While for Madrid poorest districts seem to have a different problem with a lack of venues related to the cultural and educational ones. Looking at the top ten for the poorest districts of Madrid, we can see that districts on different clusters were more similar that we think.

5. Conclusion

We can conclude that **Madrid and Barcelona districts are**, in general, quite **different especially the poorest ones**. Only the city **centers of both cities seem to share a strong similarity**. Thus, in principle, **a common strategy cannot be defined between the two cities**.

If we compare the poorest district inside each city, we conclude that the **poorest districts of Barcelona can be grouped together but the poorest districts of Madrid not**. Thus, commons actions for the poorest districts of Barcelona can be taken, while in principle for Madrid is not possible but after looking at the top ten venues for the **poorest districts of Madrid, we can observe a lack of similar kind of venues for the two clusters inside Madrid**.

A solution for **Barcelona poorest districts can be to increase the health care services**, while for **Madrid the solution could be to increase cultural and educational venues**.