Exploring venues in Madrid and Barcelona

using Foursquare location data

1. INTRODUCTION

Whenever a company or a person decides to start a new journey in a new city or a new country, there are a lot of research to do ahead before taking this big step on moving to a new city. It would be wise to start collecting and analysing a wide variety of information available about the new future home city. Depending on the type and area of business of the company, or personal lifestyle of the person, age, and many other personal factors, different information will be useful to know beforehand, and which will most probably help in deciding which city would be best to move in. From municipal services and cable providers, to the best local coffee shops and healthcare practices, as well as cultural offer, these are only some of the thing everyone wants to know about the host city. Proximity to stores, schools, coffee shops, parks and many other venues are some of the key elements to analyse when searching for the right neighbourhood in the new city, as well as information about local internet providers, best restaurants and businesses in the area.

The purpose of this project is to explore the venues in 2 cities from Spain: Madrid and Barcelona. Madrid is the capital city of Spain. After Madrid, the second major city in Spain, in terms of population as well as notoriety, is without doubt Barcelona. In both cities there are a lot of venues to explore and which can provide valuable information to anyone who is interested in moving to Madrid or Barcelona. To explore the information, during this project the Foursquare API, together with external datasets (csv file, json files, etc.), are the main tools used to provide the expected analyses of different neighbourhoods from Madrid and Barcelona. Different maps will be plotted in order to better visualize the neighbourhoods and different venues from both cities, highlighting the different venue categories in each city, and maybe what would be very interesting is to find the similarity or dissimilarity around venues categories in each city. The comparison between venues categories that you can find in each city could be a valuable element which could make the difference when choosing a new city where to move in.

The target audience is any person interested in living in Madrid or Barcelona, and wants to access information from both cities, as a comparison tool. On the other hand,

based on the information provided, firstly on neighbourhood level, secondly on venues level, business decisions can be taken with respect to the right area and city where to expand or start a new business.

2. DATA

External csv and json file will be used in order to get information about neighbourhoods in Madrid and Barcelona, as well as their location, etc. Main information used from external files in this project:

- Neighbourhoods name
- Latitude
- Longitude
- Population
- Area of each neighbourhood, among others

On the other hand, in order to start analysing the venues in each neighbourhood, additional information regarding venues will be used throughout this project, by using the Foursquare API as main tool. It entails to extract complete information of various venues (name, location, etc.) in each city, plot them on maps, and highlight them based on their category, frequency, or other information available. This enables any person to take a quick glance and decide which neighbourhood better adapt to his/her lifestyle, based on venue location and other parameters.

All previous data will be used to cluster neighbourhoods and venues based on different characteristics, in order to identify similar neighbourhoods and venues, as well as the relationship amongst them. Neighbourhoods and venues will be grouped in different groups/clusters, plotted and coloured for a better and easy visualization of components from different group which are similar, and easy identify the points of interest for any person (type of venues in neighbourhood, etc.) who wants to explore Madrid and Barcelona.

More about data used for analysis:

Madrid has 21 neighbourhoods and more than 3.3 million inhabitants, and it´s located in the centre of Spain. Barcelona, on the other hand, which is the second largest city of Spain, with a population of 1.6 million within city limits, is located on the coast of northeaster Spain, and has only 10 neighbourhoods.

Four different external datasets were used to make possible this analysis. Two of them are csv files, one for each city, with information around the name of the neighbourhoods, the area (ha) of each neighbourhoods, total inhabitants in each neighbourhood, the latitude and longitude coordinates for the centre of the neighbourhood. Other two external files were used, json files, containing the geographical coordinates of all the neighbourhoods in Madrid and Barcelona. We can take a first look over how the population is spread across neighbourhoods in both cities:

Fig.1. Madrid´s population per neighbourhoods

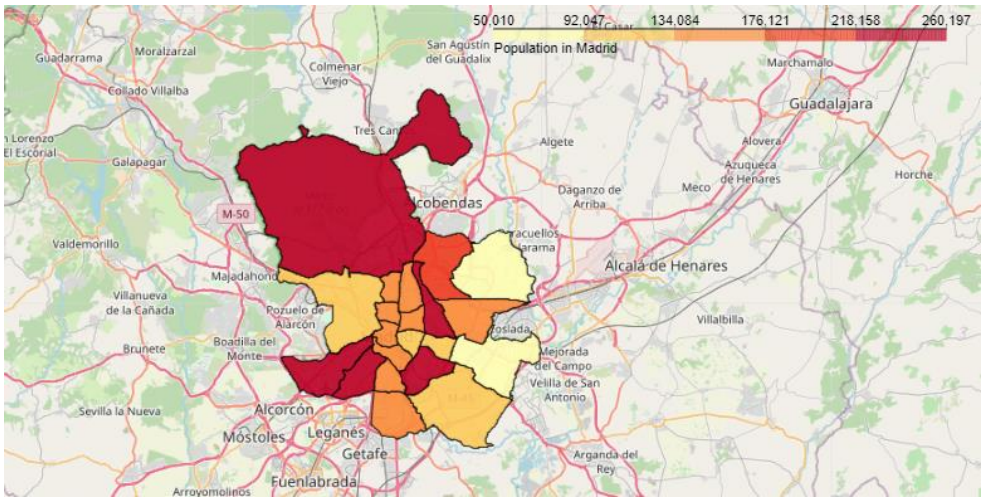
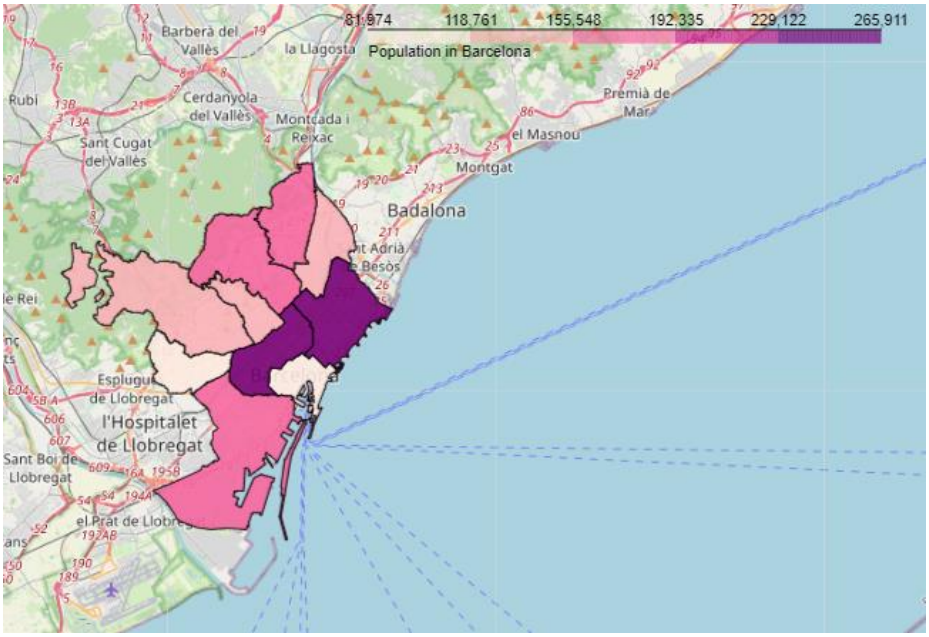


Fig.2. Barcelona´s population per neighbourhoods



From previous 2 figure we can appreciate that in Madrid there any much more inhabitants living in not necessarily centric neighbourhood, while in Barcelona more population in living in centric neighbourhood, mainly in 2 neighbourhood.

3. METHODOLOGY AND EXPLORATORY DATA ANALYSIS

As a first step, venues in Madrid and Barcelona were retrieved using the Foursquare API in order to explore the neighbourhoods and segment them. The location data was extracted from the Foursquare API for all venues up to a distance of 1.5 kilometre from the centre of each neighbourhood. As a result, 2 datasets were created, one for each city, containing the following information for each neighbourhood, in addition to the previous ones:

- **Venues** – name of the venue
- **Venue Latitude** - the latitude value of the venue.
- **Venue Longitude** - the longitude value of the venue
- **Venue Category** – the category in which the venue belongs to

For Madrid 194 unique venues categories were returned, and for Barcelona 184 unique categories were retrieved, with a limit of 100 venues per neighbourhood considered for analysis.

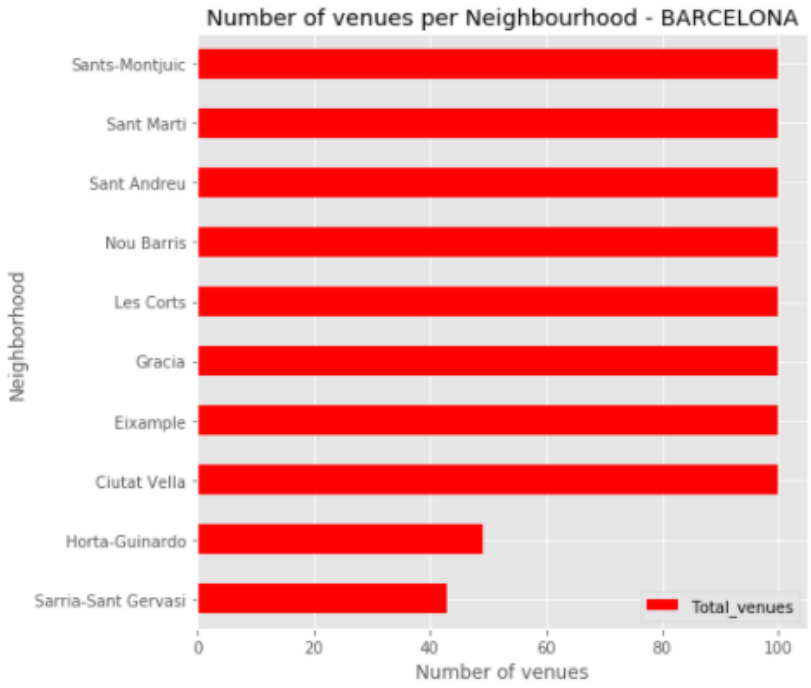


Fig.3. Barcelona

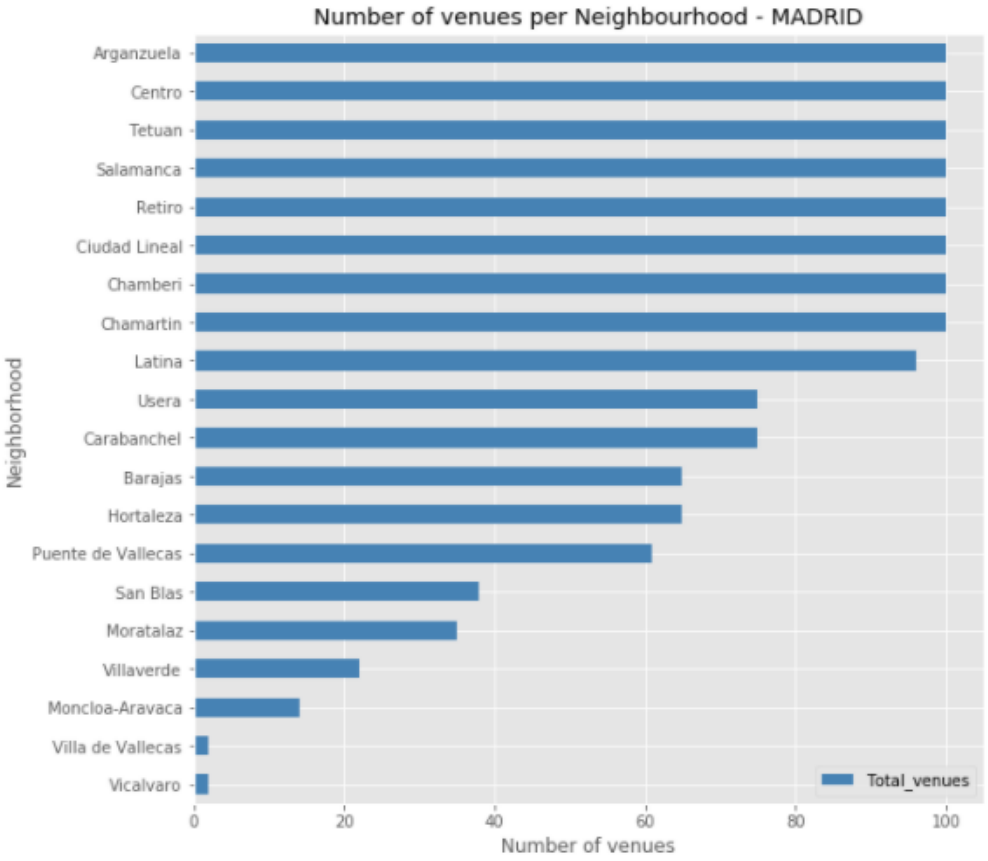


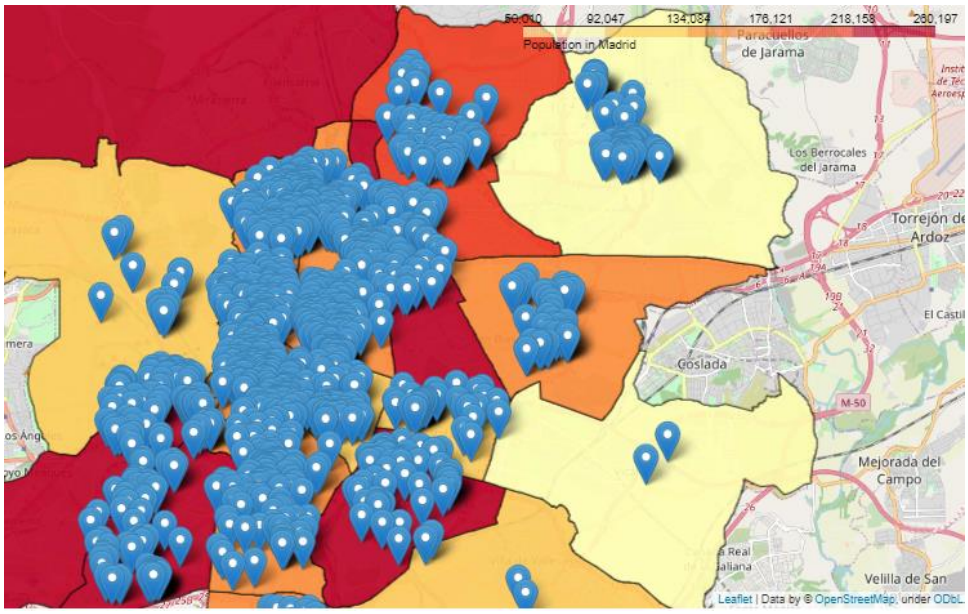
Fig.4. Madrid

From figure 3 and 4, we can see the number of venues returned for each neighbourhood, with a limit of 100 venues information retrieved. Therefore, they are much more venues in each city that could be explored, and in this project, we have considered a limit of 100 venues per neighbourhood, on a radius of 1.5 kilometre.

Explore neighbourhoods in Madrid and Barcelona

The analysis begins by taking a look at the various categories of venues that exist in each city. As there are many categories and due to the high number of venues and neighbourhood, we decided to focus the analysis on the most common venue categories in each neighbourhood, first 10^{th} categories, and then use this feature to group the neighbourhoods into clusters. In order to provide clustering of neighbourhood based on venues categories, the k-means clustering algorithm will be used to complete this task.

Fig.5. Venues retrieved by neighbourhood - MADRID



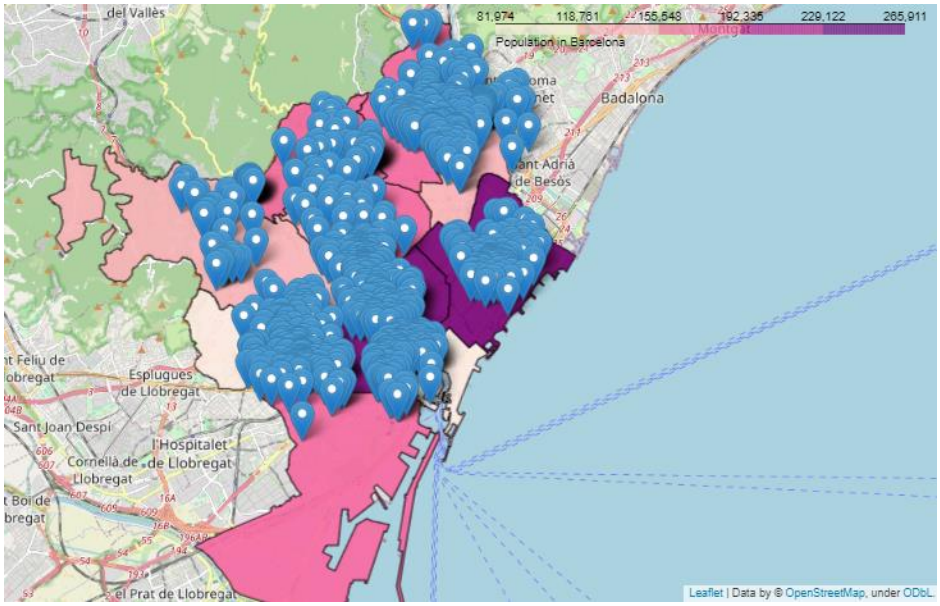


Fig.6. Venues retrieved by neighbourhood - BARCELONA

In this step, the venues by neighbourhoods were retrieved for Madrid and Barcelona, and next step is to analyse each neighbourhood, in each city. Since data in our datasets is categorical, categories of venues which cannot be used to make sum or any mathematical operations, using one-hot-encoding we could somehow transform categorical feature into an array whose size is the number of possible choices, in this case number of categories of venues, with value 0 those categories that are not in a specific neighbourhood, and value 1 for those categories that exists in a specific

neighbourhood. Neighbourhoods were group by taking the mean of the frequency of occurrence of each category.

MADRID: example of first 10th most common venue categories in each neighbourhood

	Neighborhood	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
0	Arganzuela	Art Gallery	Plaza	Bar	Spanish Restaurant	Park	Restaurant	Coffee Shop	Art Museum	Tapas Restaurant	Argentinian Restaurant
1	Barajas	Hotel	Airport Service	Spanish Restaurant	Coffee Shop	Airport Lounge	Restaurant	Tapas Restaurant	Duty-free Shop	Airport	Accessories Store
2	Carabanchel	Bar	Clothing Store	Spanish Restaurant	Tapas Restaurant	Gym	Supermarket	Coffee Shop	Music School	Metro Station	Fast Food Restaurant
3	Centro	Plaza	Hotel	Tapas Restaurant	Spanish Restaurant	Café	Restaurant	Hostel	Bookstore	Coffee Shop	Wine Bar
4	Chamartin	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Hotel	Tapas Restaurant	Japanese Restaurant	Steakhouse	Pizza Place	Plaza	Athletics & Sports

BARCELONA: example of first 10th most common venue categories in each neighbourhood

	Neighborhood	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
0	Ciutat Vella	Tapas Restaurant	Hotel	Cocktail Bar	Coffee Shop	Spanish Restaurant	Bar	Italian Restaurant	Plaza	Ice Cream Shop	Pizza Place
1	Eixample	Hotel	Tapas Restaurant	Mediterranean Restaurant	Boutique	Coffee Shop	Restaurant	Spanish Restaurant	Bookstore	Sandwich Place	Wine Bar
2	Gracia	Plaza	Mediterranean Restaurant	Wine Bar	Tapas Restaurant	Bakery	Cocktail Bar	Restaurant	Indie Movie Theater	Italian Restaurant	Pizza Place
3	Horta-Guinardo	Spanish Restaurant	Tapas Restaurant	Park	Café	Gym	Grocery Store	Plaza	Soccer Field	Mediterranean Restaurant	Sandwich Place
4	Les Corts	Restaurant	Pizza Place	Tapas Restaurant	Spanish Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Hostel	Middle Eastern Restaurant	Italian Restaurant	Japanese Restaurant

Cluster Neighbourhoods:

Finally, neighbourhoods were clustered based on the category of venues that were retrieved for each neighbourhood, using K-Means clustering, and decide to cluster the neighbourhoods into seven separate groups.

In the following maps we can see the clusters created. We can see also that there are neighbourhoods with no venue information retrieved, therefore, those were not clustered in the clusters created.

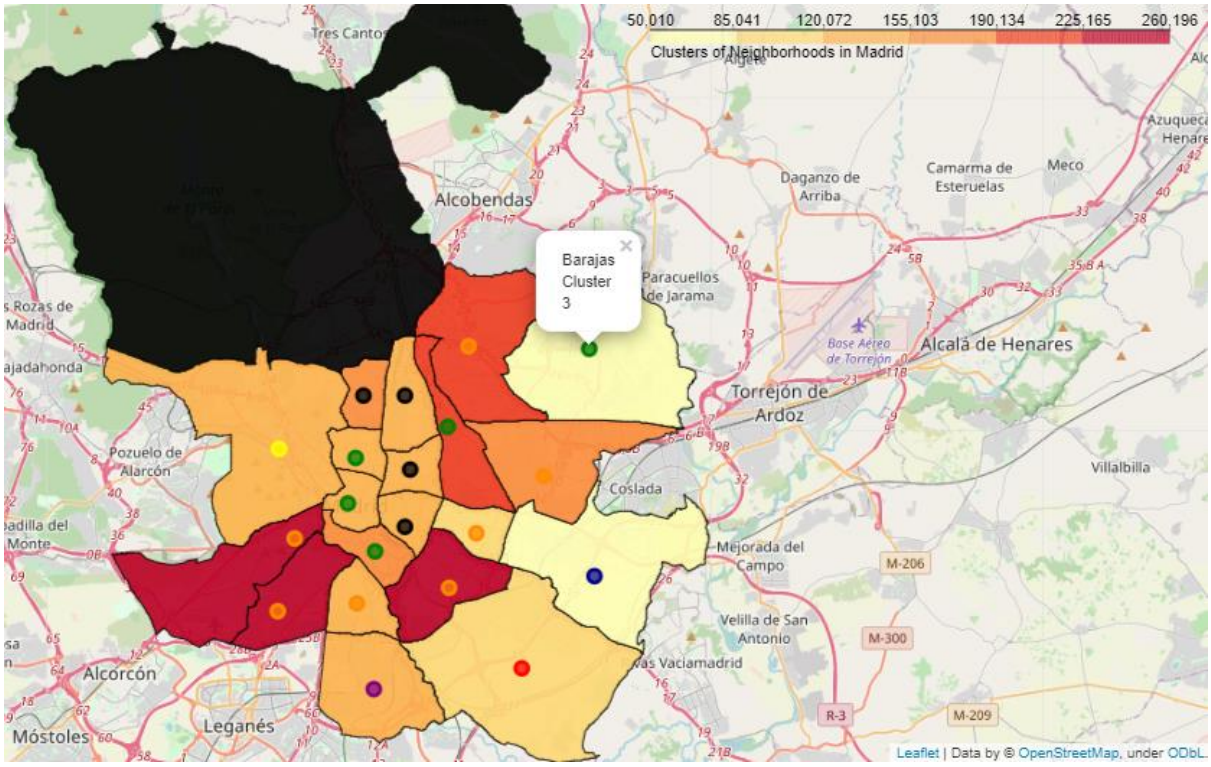


Fig.7. Neighbourhood Clusters in Madrid

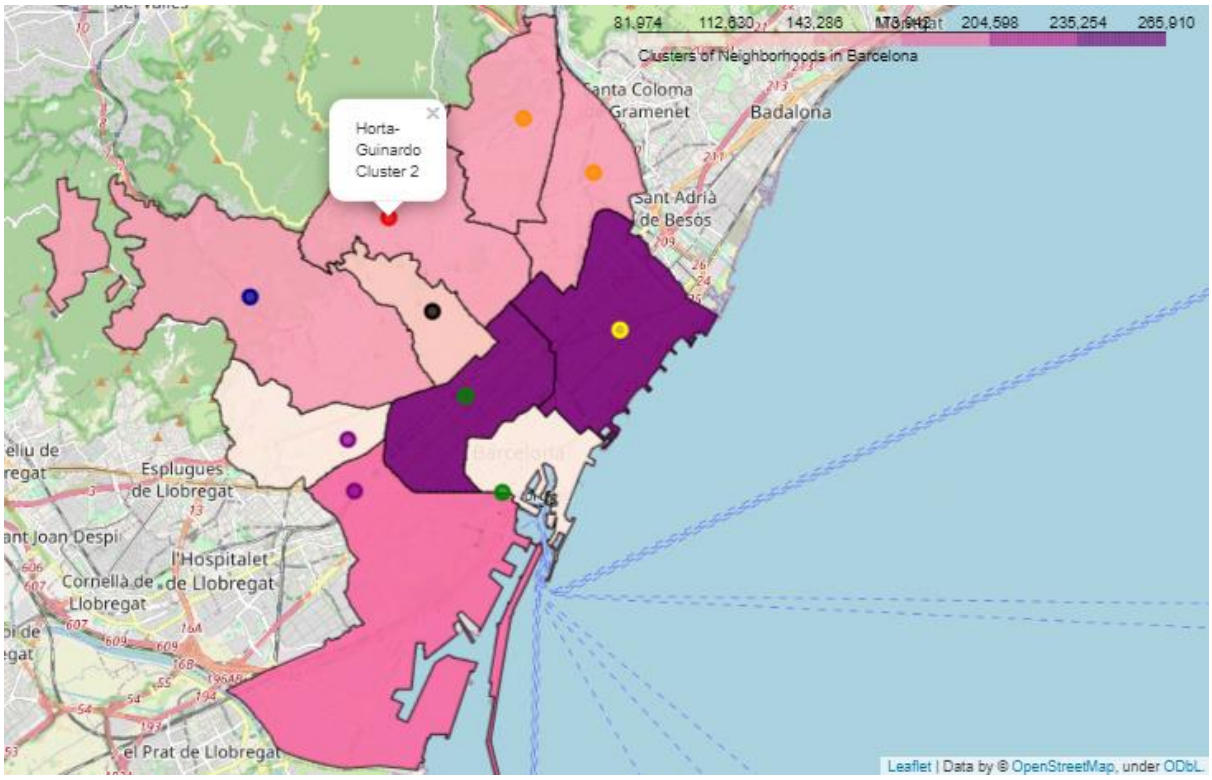


Fig.8. Neighbourhood Clusters in Barcelona

In **Madrid**, the below clusters were created, with the following more common venue categories:

 ✓ Cluster0 : *Spanish Restaurants* and *Restaurants* in general

 ✓ Cluster1: *Toll Booth*, *Mediterranean Restaurant*

 ✓Cluster2: *Shoe Store*, *Scenic Lookout*

 ✓Cluster3*: Art Gallery*, *Plaza*, *Bar*

 ✓ Cluster4*: Grocery Store, Theme Park Ride / Attraction, Park, Clothing Store*

 ✓ Cluster5*: Train Station, Brewery, Electronics Store*

 ✓ Cluster6: *Pool, Salad Place, Museum, Pub*

In **Barcelona**, the following clusters were created:

 ✓ Cluster0: *Plaza, Mediterranean Restaurant, Wine Bar*

 ✓ Cluster1: *Theme Park*, *Scenic Lookout*

 ✓Cluster2: *Gym, Spanish Restaurant, Tapas Restaurant,*

 ✓Cluster3*: Tapas Restaurant, Hotel, Coffee Shop*

 ✓ Cluster4*: Spanish Restaurant, Grocery Store, Supermarket*

 ✓ Cluster5*: Pizza Place, Japanese Restaurant*

 ✓ Cluster6: *Italian Restaurant, Mediterranean Restaurant, Park*

Results:

We notice that Madrid and Barcelona have slightly similar neighbourhood clusters. However, due to the higher number of neighbourhood, and of course which mean

more inhabitants and venues, we can notice that in Madrid there is a higher variety of venues, and more equally spread across its neighbourhoods. On the other hand, the central part of Madrid has more restaurants and art galleries and bars, same situation as in Barcelona.

Al least 4 neighbourhoods from Madrid have been included in the same cluster 1, where the most common categories of venues are *Spanish Restaurant*, mostly in the centre of the city. Another details we can detect is that the *Theme Park Ride / Attraction, Parks* in both cities, Madrid and Barcelona are located in the peripheric neighbourhoods. On the other hand, Barcelona even if it´s smaller than Madrid, in terms of area and number of inhabitants, it has very similar distribution of venues across its territory. Anyone who is looking on moving in Madrid or Barcelona, can use this information and choose what neighbourhood to visit or live in, in a nutshell, have the big picture on what´s the grade of similarity of dissimilarity between these 2 great cities.

4. CONCLUSION

In general, both cities offer a wide variety of *Restaurants* and places to visit. Madrid and Barcelona have similar categories of venues to visit and to discover. The purpose of this project was to highlight the most common and useful places in both cities, where they are located and how are they distributed across their territory. Both cities offer a wide variety of restaurants and grocery markets, they both have park and attraction parks, where anyone can spend real good time with friend, family or maybe a good place where to start a new business.