The Battle Of Neighborhoods

-

Madrid / Spain

Table of Contents

1 Introduction

This project is aimed at analyzing the best areas to buy a house in Madrid. The analysis will be based on the number of restaurants, schools, shops or metro lines available in the different neighborhood as well as on the potential renting profit on the area. Where the annual renting profit is computed as the renting price per month multiplied by the months in a year and divided by the sale price.

This problem could be of interest for anyone who is considering buying a house in a big city, either for living or investing.

2 Data

In order to find the best neighborhoods in Madrid to buy a house, the following data will be used:

- List of venues provided by foursquare. The following categories will be considered:
    - Food (Restaurants)
    - Food & Drink Shop
    - Clothing Store
    - Bus Stop
    - Metro Station
    - Arts & Entertainment
    - Park
    - Athletics & Sports
    - Hospital
    - School
    - University
    - Bank
    - Pharmacy
- Sale price per square meter in the different districts and Neighborhoods of Madrid. This information will be obtained from https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Estadistica/Areas-de-informacion-estadistica/Edificacion-y-vivienda/Mercado-de-la-vivienda/Precios-de-la-vivienda/?vgnextfmt=default&vgnextoid=bf281b47a277b210VgnVCM1000000b205a0aRCRD&vgnextchannel=22613c7ea422a210VgnVCM1000000b205a0aRCRD
- Rent price per month and per square meter in the different districts of Madrid. Note that the rent price is not available per neighborhood and therefore, it will be assumed the same rent price for all the neighborhoods within each district.
- List of Neighborhoods and their coordinates obtained from the following geojson: https://github.com/codeforamerica/click_that_hood/blob/master/public/data/madrid.geojson

3 Methodology

The methodology followed in this project is detailed hereafter:

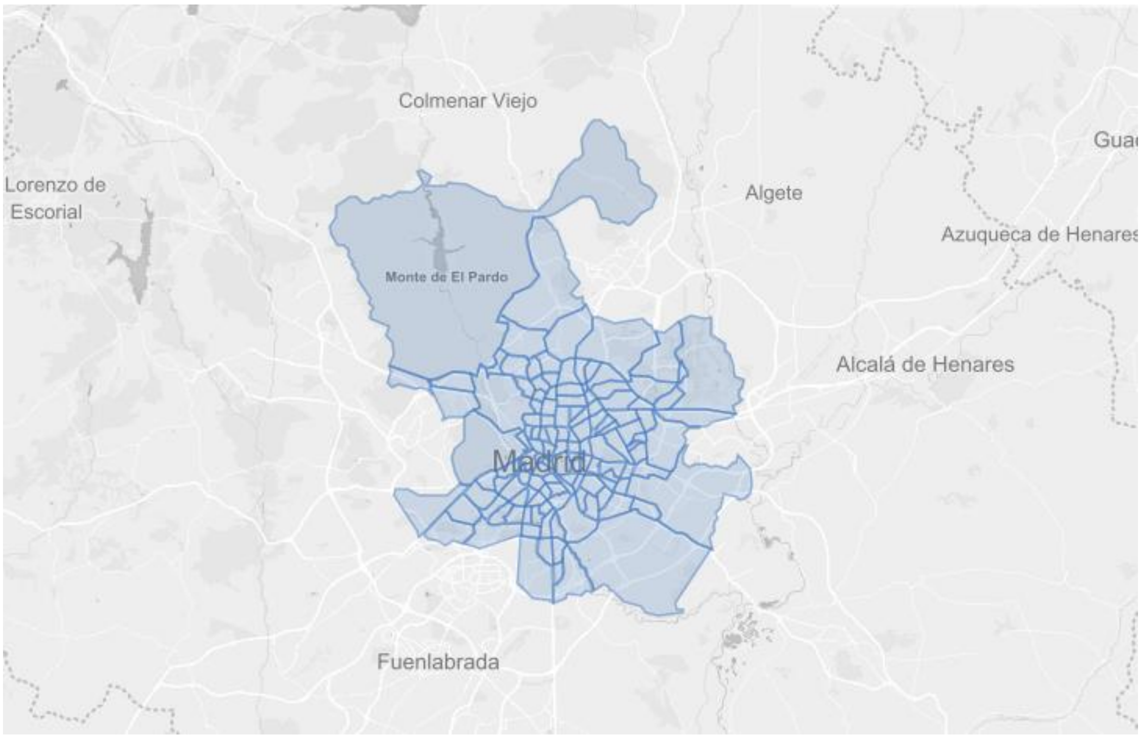
- Process madrid.geojson to obtain the list of Neighborhoods in Madrid
- Compute the reference point of each Neighborhood as the centroid of each polygon

- Using foursquare, look for the list of hospitals, schools, parks, restaurants, metro and bus stations, etc. within 5 minutes-walk (500 meters distance) from the reference position (refer to section 2 for the list of categories to be considered for the analysis).
- Store the data provided by foursquare in a pandas dataframe
- Process the resulting dataframe to obtain the number of venues of each type available in each neighborhood
- Using Kmeans and based on the number of venues available, classify the neighborhoods in 5 different groups. Display them in a folium map.
- Check the obtained groups to identify the most suitable groups (i.e. those with good communications, several schools and parks, at least one hospital, and with a relevant number of restaurants and shops in the area).
- Process the rent and sales database, merge them with the previous generated dataframe (containing the neighborhoods location, venues per category and cluster labels) and display the rent and sale prices in a folium map.
- Compute the renting profit as the ratio between the rent (multiplied by 12) and the sale price.
- Filter out the non-suitable group(s) and identify the most profitable neighborhoods
- Recommend one or two neighborhoods to buy a house based on the neighborhood features and the potential renting profit.

4 Results

4.1 List of Neighborhoods in Madrid

The list of neighborhoods in Madrid and their coordinates are obtained from the madrid.geojson obtained from:
https://github.com/codeforamerica/click_that_hood/blob/master/public/data/madrid.geojson



The json is processed to save in a dataframe the name in each neighboord and its coordinates, computed as the centroid of the coordinates defining the polygon of each area.

An excerpt of the resulting dataframe is showed below:

	Neighborhood	latitude	longitude
123	Alameda de Osuna	40.4566	-3.59166
124	Aeropuerto	40.4769	-3.55006
125	Casco Histórico de Barajas	40.4755	-3.57775
126	Timón	40.4895	-3.59815
127	Corralejos	40.4676	-3.60347

4.2 Available venues in each neighborhood

We use foursquare to look for the list of hospitals, schools, parks, restaurants, metro and bus stations, etc. within 5 minutes-walk (500 meters distance) from the reference position. The categories are searched in foursquare.

- 'Food': '4d4b7105d754a06374d81259',
- 'Food & Drink Shop': '4bf58dd8d48988d1f9941735',
- 'Clothing Store':'4bf58dd8d48988d103951735',
- 'Bus Stop': '52f2ab2ebcbc57f1066b8b4f',
- 'Metro Station': '4bf58dd8d48988d1fd931735',
- 'Arts & Entertainment': '4d4b7104d754a06370d81259',
- 'Park':'4bf58dd8d48988d163941735',
- 'Athletics & Sports':'4f4528bc4b90abdf24c9de85',
- 'Hospital':'4bf58dd8d48988d196941735',
- 'School':'4bf58dd8d48988d13b941735',
- 'University':'4bf58dd8d48988d1ae941735',
- 'Bank':'4bf58dd8d48988d10a951735',
- 'Pharmacy':'4bf58dd8d48988d10f951735'

The results provided by foursquare are saved in a dataframe. An example of the data stored in this dataframe for a given neighborhood is shown below:

	Unnamed: 0	Address	Category	Latitude	Longitude	Name	Neighborhood
0	0	Pl. Oriente, 2	Food	40.418081	-3.711867	Café de Oriente	Palacio
1	1	Plaza De Oriente 6	Food	40.418598	-3.711352	La Lonja del Mar	Palacio
2	2	27 C. de Preciados	Food	40.419668	-3.706135	Starbucks Preciados 27	Palacio
3	3	C. Pasadizo de San Ginés, 5	Food	40.416754	-3.707079	Chocolatería San Ginés	Palacio
4	4	Calle Mayor	Food	40.416095	-3.708119	Torrons Vicens: Artesa D' Agramunt	Palacio
5	5	C. Mayor, 7	Food	40.416343	-3.705682	Museo del Jamón	Palacio
6	6	NaN	Food	40.415045	-3.708306	Mesón De San Miguel	Palacio
7	7	C. Mayor, 2	Food	40.416651	-3.704662	La Mallorquina	Palacio
8	8	Plaza Mayor	Food	40.415446	-3.707157	Una Copa de Balon	Palacio
9	9	Mercado de San Miguel	Food	40.415313	-3.711930	Horno La Santiaguesa	Palacio
10	10	C. Toledo, 28	Food	40.413888	-3.707478	CUBIERTOS DE GLORIA	Palacio
11	11	Preciados 42	Food	40.419819	-3.707394	Steak Burguer Bar	Palacio
12	12	C. Mayor, 22	Food	40.416359	-3.707021	100 Montaditos	Palacio
13	13	C. Ciudad Rodrigo, 5	Food	40.415733	-3.708718	Starbucks San Miguel	Palacio
14	14	C. Preciados, 3	Food	40.417497	-3.704686	Club del Gourmet Corte Ingles	Palacio
15	15	Calle de los Cuchilleros, 17, LOC;DUP 17-19;	Food	40.414019	-3.708136	Botín	Palacio

In order to properly process this information and to be able to classify the neighborhoods based on the number of facilities of each type available in the area, we use the pivot_table method of pandas to convert the venues dataframe in a table with the number of banks, hospitals, metro and bus stops, etc. in each neighborhood.

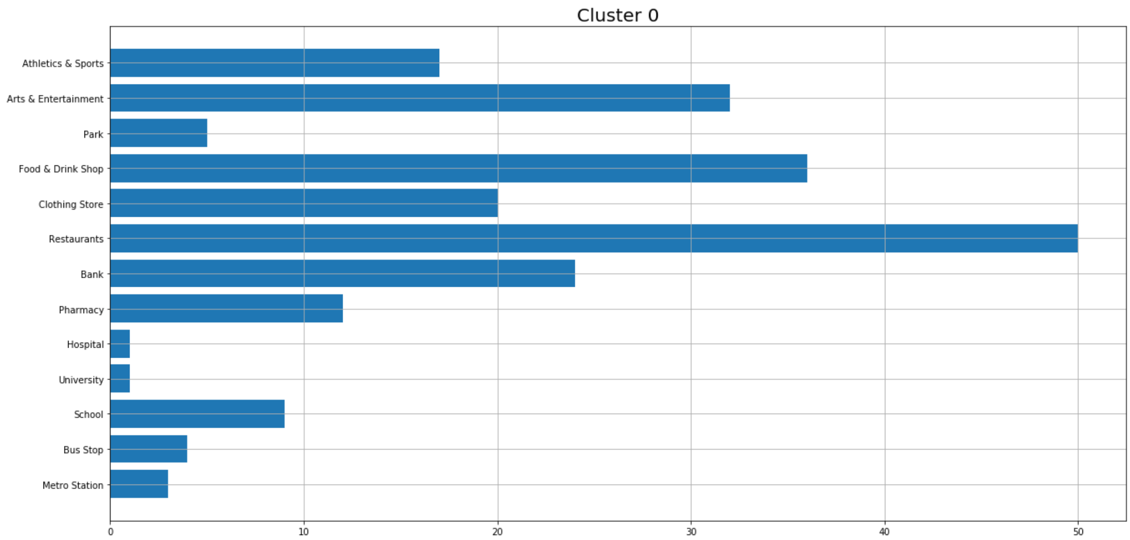
See below the first rows of the resulting table:

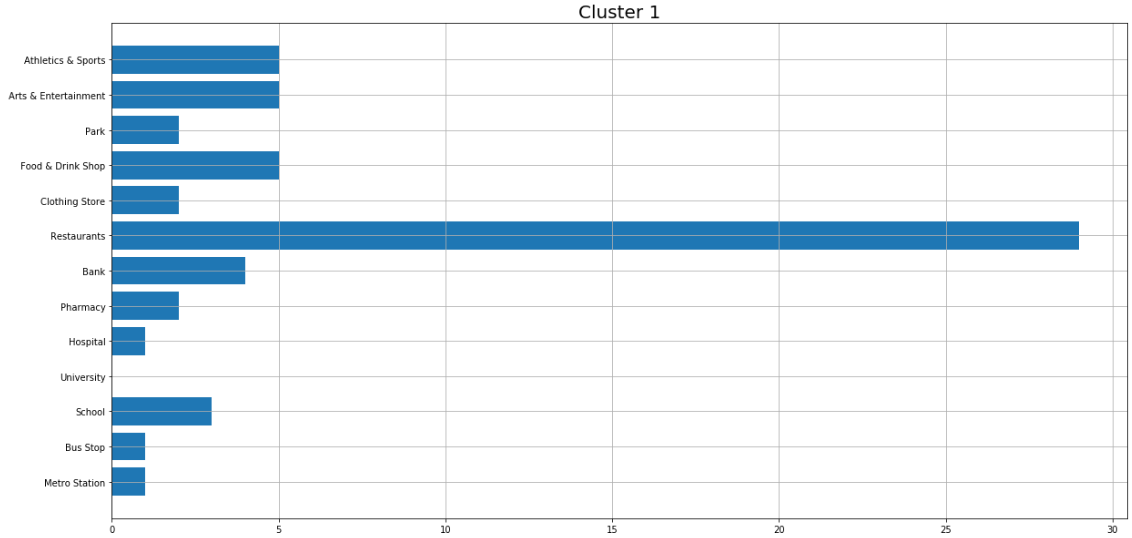
	Neighborhood	Metro Station	Bus Stop	School	University	Hospital	Pharmacy	Bank	Restaurants	Clothing Store	Food & Drink Shop	Park	Arts & Entertainment	Athletics & Sports
0	Palacio	3.0	3.0	9.0	3.0	0.0	22.0	23.0	50.0	48.0	48.0	7.0	50.0	18.0
1	Embajadores	5.0	2.0	12.0	1.0	0.0	20.0	26.0	50.0	46.0	48.0	3.0	50.0	26.0
2	Cortes	6.0	8.0	9.0	1.0	0.0	20.0	49.0	50.0	50.0	48.0	1.0	50.0	34.0
3	Justicia	9.0	3.0	11.0	7.0	0.0	30.0	47.0	50.0	50.0	48.0	7.0	50.0	45.0
4	Universidad	8.0	5.0	21.0	7.0	1.0	43.0	45.0	50.0	50.0	47.0	6.0	50.0	49.0

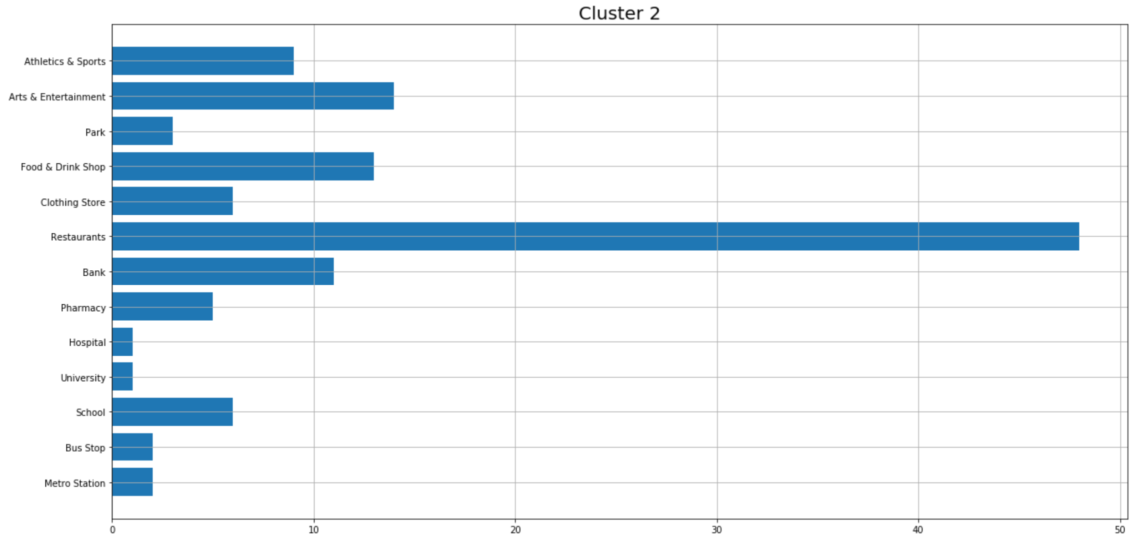
4.3 Kmeans classification

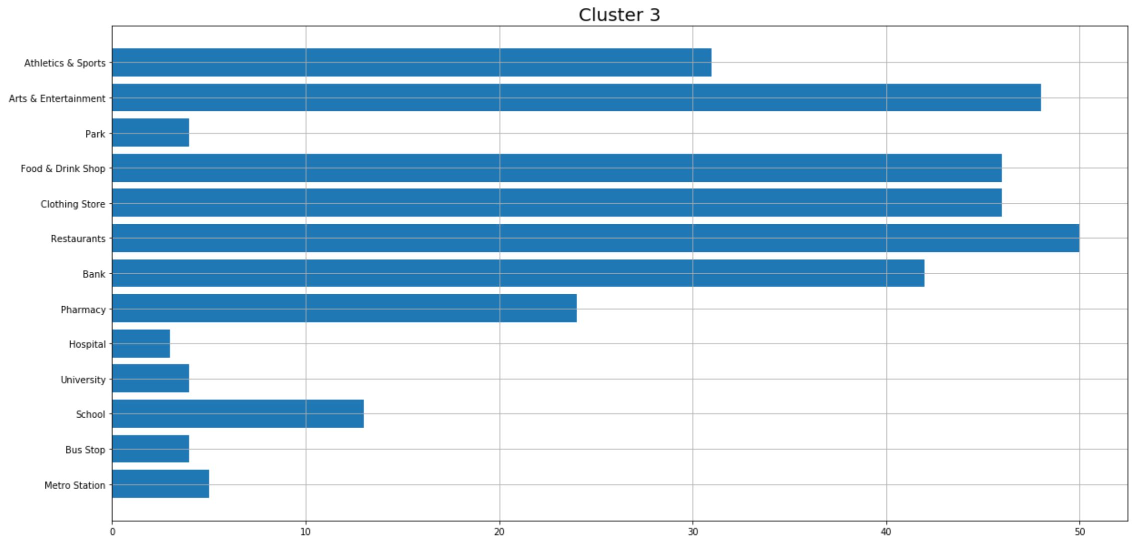
Using sklearn library, we classify the neighborhood based on the communication lines, facilities, parks or schools in each area.

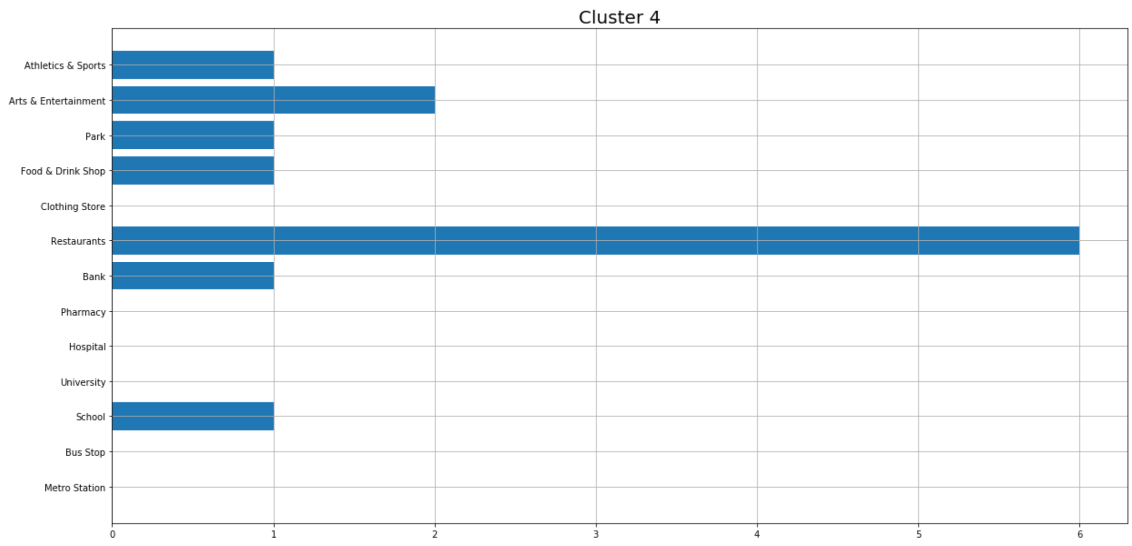
To have a better understanding of the features of each group, we compute the mean value of each category among all the neighborhoods in each group. The results obtained are depicted in the following bar charts:




Cluster 1
Athletics & Sports
Arts & Entertainment
Park
Food & Drink Shop
Clothing Store
Restaurants
Bank
Pharmacy
Hospital
University
School
Bus Stop
Metro Station
0
5
10
15
20
25
30


Cluster 2
Athletics & Sports
Arts & Entertainment
Park
Food & Drink Shop
Clothing Store
Restaurants
Bank
Pharmacy
Hospital
University
School
Bus Stop
Metro Station
0
10
20
30
40
50

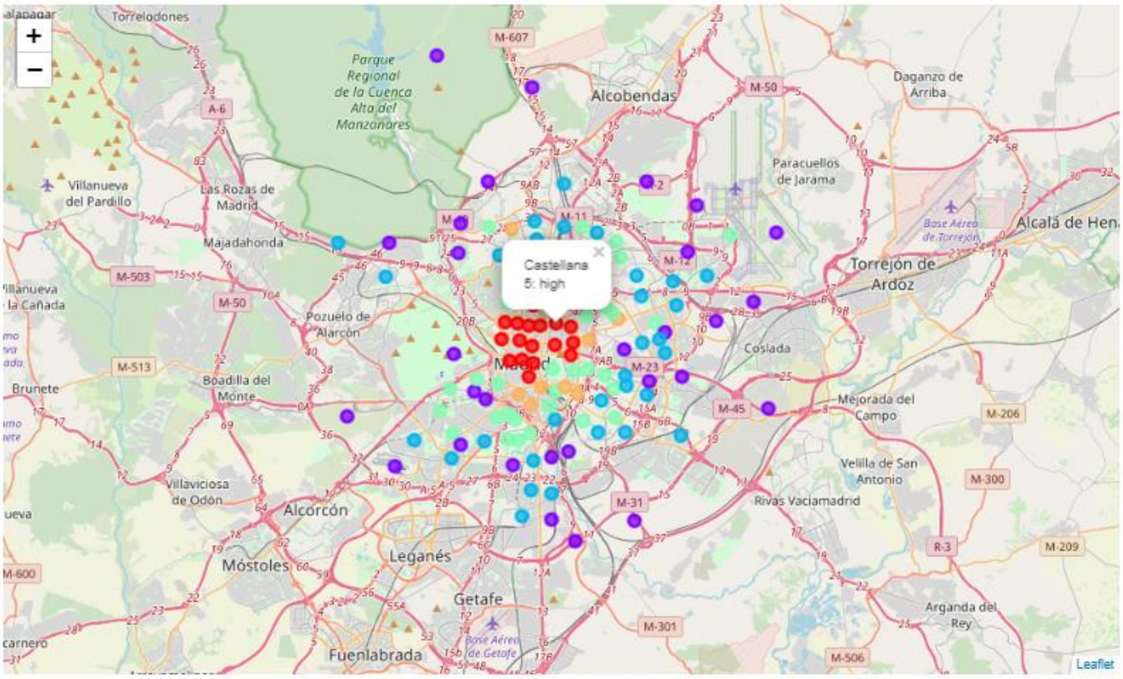

Cluster 3
Athletics & Sports
Arts & Entertainment
Park
Food & Drink Shop
Clothing Store
Restaurants
Bank
Pharmacy
Hospital
University
School
Bus Stop
Metro Station
0
10
20
30
40
50



The analysis of the figures above shows that:

- **Cluster 0**: Has a reasonable number of metro and bus stations. There are several food, clothes stores, pharmacy, banks and restaurants. There is also a hospital and several schools. We therefore classified this cluster as **medium-high**
- **Cluster 1**: In general terms this cluster has worse communications (metro and bus stops), no universities and fewer facilities (shops, restaurants, arts & entertainment) than cluster 0. Compared to cluster 2, this group is also less equipped. Therefore, it is classified as **medium-low**
- **Cluster 2**: In terms of facilities, this cluster is between cluster 0 and cluster 1. Therefore, it is considered as **medium**
- **Cluster 3**: This is the cluster with more clothes, restaurants, banks, pharmacies, schools, etc., therefore it classified as **high**
- **Cluster 4**: Neighborhoods in this group have poor or non-communications, few facilities, no hospitals... so it is considered as l**ow**

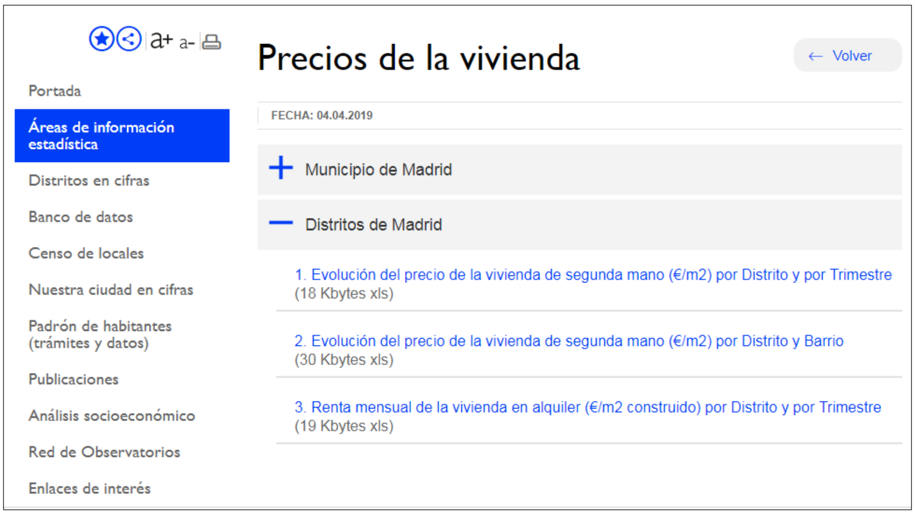
Finally, the neighborhoods grouped in clusters are depicted in a folium map.



As shown above, and as expected, the neighborhoods with more facilities are located in the center of Madrid (in red color). However, there are also several also in the borders of the city with good facilities, especially those depicted in orange (medium-high areas) and light green (medium cluster)

4.4 Sale and Rent prices analysis

Sale price per square meter in the different districts and Neighborhoods of Madrid is obtained from the https://www.madrid.es website:



Rent price per month and per square meter in the different districts of Madrid is obtained as well from the same site.

However, given that the rent price is not available per neighborhood but only per district, it is assumed the same rent price for all the neighborhoods within each district. It is also important to consider that:

- Not all the neighborhoods in the madrid.json (used to get the list of neighborhoods) in Madrid are available in the sale price excel file. There are a couple of them missing that will be filled with zeros
- There are some discrepancies in the names used by both lists. For instance, there is no consistency in the use of accents. For that reason, before merging both lists, we will remove the accents. Then, we will add them again since we need to plot the map (neighborhoods names should be aligned with the names on the json)

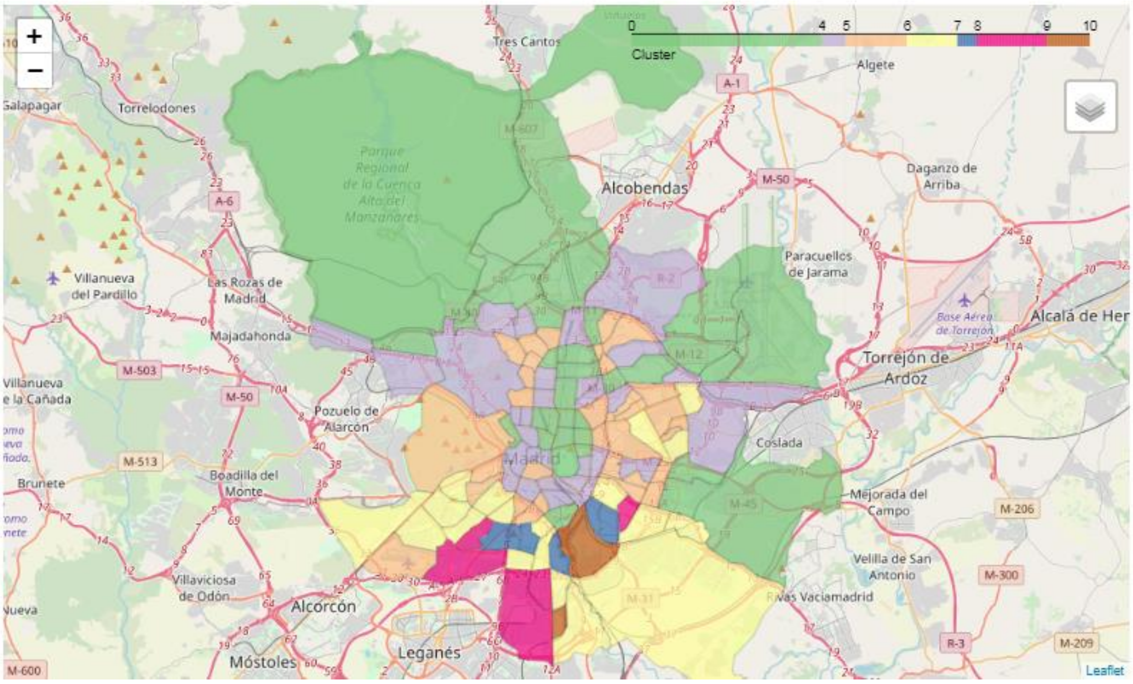
After processing the sale and rent prices database, the following dataframe is obtained, including the sale price and the monthly rent price per square meter:

	District ID	District	Neighborhood	Sale Price	Rent Price
0	011	Centro	Palacio	4873	19.0
1	012	Centro	Embajadores	4239	19.0
2	013	Centro	Cortes	5106	19.0
3	014	Centro	Justicia	5989	19.0
4	015	Centro	Universidad	5068	19.0
5	016	Centro	Sol	5263	19.0
6	021	Arganzuela	Imperial	3896	15.1
7	022	Arganzuela	Acacias	4046	15.1
8	023	Arganzuela	Chopera	3727	15.1
9	024	Arganzuela	Legazpi	4263	15.1

With this information, the potential renting profit is computed as the ratio between the annual rent price and the sale price. The following values are obtained:

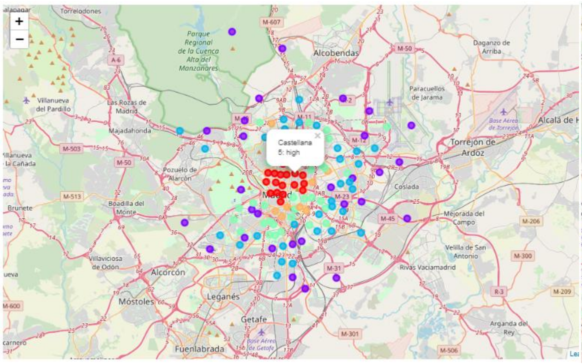
	District ID	District	Neighborhood	Sale Price	Rent Price	Profit
0	011	Centro	Palacio	4873	19.0	4.67884
1	012	Centro	Embajadores	4239	19.0	5.37863
2	013	Centro	Cortes	5106	19.0	4.46533
3	014	Centro	Justicia	5989	19.0	3.80698
4	015	Centro	Universidad	5068	19.0	4.49882
5	016	Centro	Sol	5263	19.0	4.33213
6	021	Arganzuela	Imperial	3896	15.1	4.65092
7	022	Arganzuela	Acacias	4046	15.1	4.4785
8	023	Arganzuela	Chopera	3727	15.1	4.86182
9	024	Arganzuela	Legazpi	4263	15.1	4.25053
10	025	Arganzuela	Delicias	3777	15.1	4.79746

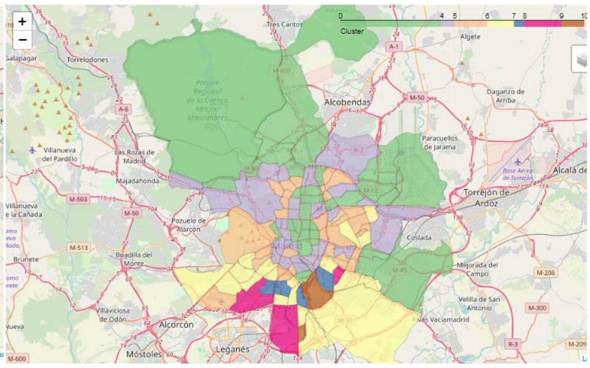
This information is finally depicted in a Choropleth map:



5 Discussion

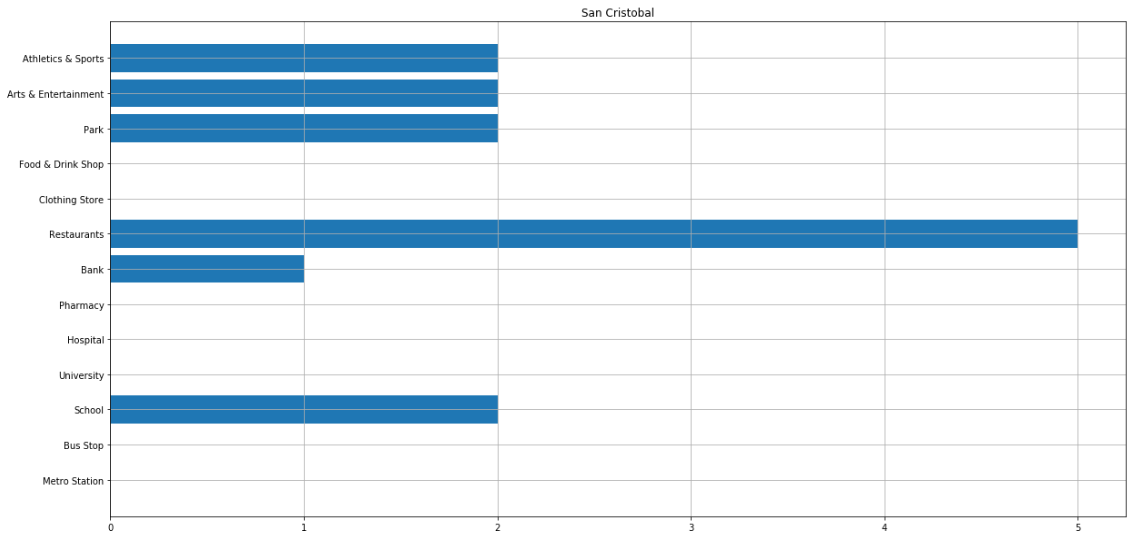
The comparison of the cluster and profit maps included in the previous section shows that while the areas with more facilities are in the central part of the city, the neighborhoods with better renting profit are in the outskirts.





For instance, focusing on the profit column only we could conclude that the best neighborhood for buying a hose in Madrid is San Cristobal. However, if we check the cluster label we observe that this area is classified as “low” facilities:

	Neighborhood	Cluster Labels	District	Sale Price	Rent Price	Profit
107	San Cristobal	1: low	Villaverde	1252	10.8	10.3514



Therefore, a trade off shall be made between the profit and the neighborhood quality. Considering as well that the house is not bought for investing, a good solution would be looking for those neighborhoods with a good enough renting profit (e.g. 5% or above) and select among them, the area(s) with the best classification.

6 Conclusion

As detailed in the previous paragraph, it is proposed to make the neighborhood selection by looking for those neighborhoods located in “high” or “medium-high” areas and with a renting profit above the 5%.

For instance, the most profitable neighborhoods in the “high” areas are:

	Neighborhood	Cluster Labels	District	Sale Price	Rent Price	Profit
1	Embajadores	5: high	Centro	4239	19.0	5.37863
37	Gaztambide	5: high	Chamberi	4664	18.4	4.73413
0	Palacio	5: high	Centro	4873	19.0	4.67884
38	Arapiles	5: high	Chamberi	4855	18.4	4.54789
4	Universidad	5: high	Centro	5068	19.0	4.49882

Doing the same for the “medium-high” areas we obtain:

	Neighborhood	Cluster Labels	District	Sale Price	Rent Price	Profit
31	Bellas Vistas	4: medium-high	Tetuan	3354	14.8	5.29517
93	Quintana	4: medium-high	Ciudad Lineal	2933	12.9	5.27787
21	Fuente del Berro	4: medium-high	Salamanca	4275	18.6	5.22105
8	Chopera	4: medium-high	Arganzuela	3727	15.1	4.86182
13	Pacífico	4: medium-high	Retiro	4059	16.2	4.78936

So based on the results above, the neighborhood recommended to buy a house in Madrid is **Embajadores.** The renting profit in this area is 5.37% and as shown in the following bar chart, there are good communications, parks, many facilities (restaurants, shops, banks, pharmacies, etc.) in the area: