SOCIO-ECONOMICAL ANALYSIS OF MADRID'S NEIGHBORHOODS

Mar Campillos Serrano

May 2, 2020

INTRODUCTION

As the financial capital in Southern Europe, Madrid is a safe and stable environment for companies to grow, and the chosen headquarter location of 2000 companies. Driving the young talent, Madrid offers extensive opportunities for further education, with 17 universities and over 30 research centers. With over 75 million tourists visiting Spain every year, the country occupies a significant worldwide economic position. It is the fourth metropolis in the EU by Gross Domestic Product thanks to leading employers such as Telefónica, Iberia, and BBVA. Sharing frontiers with 8 countries by land and sea, Spain is a logical destination for international trade, facilitated by the country's high-speed rail system, the second-longest network in the world.

The city is divided into 21 districts and 131 neighborhoods. At first, we planned to use the Foursquare API, but we realized that there isn’t enough data for some neighborhoods. Fortunately, Madrid's City Council has an excellent Open Data Website, that provides all kinds of socioeconomic data for the city. Madrid also has an Statistics Portal with so many useful data that is hard to choose just a few indicators. We will use datasets provided for both websites to perform a socio-economic analysis of Madrid and classify their neighborhoods based in their most common venues, the size and age of their population and their average income.

We consider that this is information can be useful for companies that want to invest in the city. They will be able to determine the best part of the city to locate their businesses depending on the kind of activity they perform (commercial, financial, industrial,...) and how their customer base is.

DATA

As I said earlier, the City Council Open Data website has lots of information, for this project we are going to use:

- Registry of venues and the activity that is performed in them. This is a huge dataset with more than 163000 rows and 46 columns. The most important columns are: venue id, district, neighborhood, venue situation ('open', 'closed'…) and description of the activity ('hairdresser service', 'bar', 'restaurant'...). It needs a lot of data cleaning and preprocessing to extract the relevant information, all the open venues in each neighborhood, and their activities. This information will be used to cluster the neighborhoods.

- Demographic indicators by neighborhood in 2016. It shows the total population, the average age, the percentage of the population that it's 65 years old and older... We consider that population and average age are relevant data that can help us to classify the neighborhoods.
- The average income per person and neighborhood. We will use it as well to classify the neighborhoods. For example, a company that sells luxury products needs a customer base with a high average income.
- Table with the coordinates of each neighborhood. This information was obtained using geopy and Nominatim and will be used to create a map of the neighborhoods and its clusters.
- Geojson file of Madrid that will be used in choropleth maps.

After cleaning and preprocessing all the datasets, we will combine them in a pandas data frame to perform cluster analysis of the neighborhoods. We will represent the clusters in a map, alone and combined with choropleth maps.

Data Cleaning and Preprocessing

To get the venues data ready for analysis we have to drop most of the columns because contain incomplete or irrelevant information. These are the columns we keep:

- 'id_local', that contains a unique number for each venue.
- 'desc_barrio_local', that contains the name of the neighborhood where the venue is located.
- 'desc_situacion_local', that indicates if the venue is open, closed, used as a home...
- 'desc_epigrafe', that contains the description of the activities developed in each venue based on the Economic Activities Index.

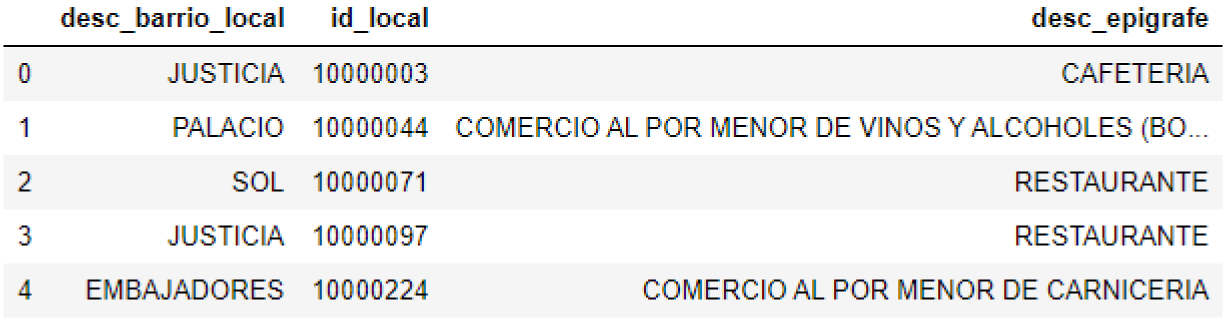
We should have a unique value for each 'id_local', but we realize that we have duplicate values. That's because some venues perform two or more activities. For instance, most schools, high schools and childcare centers have canteens, and many of them have assigned 'SERVICIOS DE COMEDOR EN CENTROS EDUCATIVOS Y CENTROS DE CUIDADO INFANTIL' as primary activity. This can affect the results of the analysis, so we drop this activity and the duplicate 'id_local' values.

The next step is select only those venues classified as 'Abierto' (Open). When we check the column 'desc_epigrafe' of the open venues, we see some problems:

- It has null values.
- It contains the values 'SIN ACTIVIDAD' (no activity) and 'LOCAL SIN ACTIVIDAD' (venue with no activity)
- Many descriptions are too long.

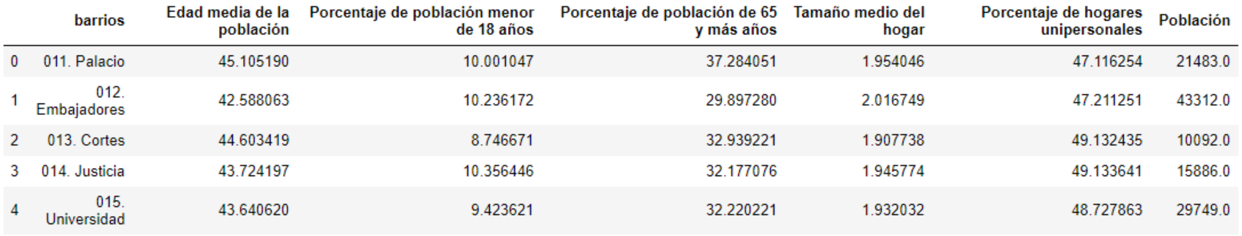
To fix all these problems we drop 'SIN ACTIVIDAD', 'LOCAL SIN ACTIVIDAD' and the null values and translate the most common descriptions.

After dropping the column 'desc_situacion_local' and changing the order, the data frame looks like this.



	desc_barrio_local	id_local	desc_epigrafe
0	JUSTICIA	10000003	CAFETERIA
1	PALACIO	10000044	COMERCIO AL POR MENOR DE VINOS Y ALCOHOLES (BO...
2	SOL	10000071	RESTAURANTE
3	JUSTICIA	10000097	RESTAURANTE
4	EMBAJADORES	10000224	COMERCIO AL POR MENOR DE CARNICERIA

Next, we process the excel file with the population data, it looks like this:



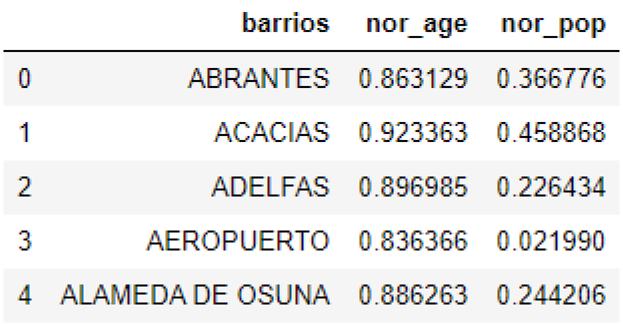
	barrios	Edad media de la población	Porcentaje de población menor de 18 años	Porcentaje de población de 65 y más años	Tamaño medio del hogar	Porcentaje de hogares unipersonales	Población
0	011. Palacio	45.105190	10.001047	37.284051	1.954046	47.116254	21483.0
1	012. Embajadores	42.588063	10.236172	29.897280	2.016749	47.211251	43312.0
2	013. Cortes	44.603419	8.746671	32.939221	1.907738	49.132435	10092.0
3	014. Justicia	43.724197	10.356446	32.177076	1.945774	49.133641	15886.0
4	015. Universidad	43.640620	9.423621	32.220221	1.932032	48.727863	29749.0

We only need the average age ('Edad media de la población') and the total population of each neighborhood. If we want to combine both data frames, the names of the neighborhoods in the population data frame must match exactly the names of the neighborhoods in the abiertos data frame. Therefore, we have to analyze which names are contained in the abiertos data frame and modify the names in the population data frame accordingly. To do this:

- we create a list called br with the name of the neighborhoods in the population data frame and
- a list called nombres with the names of the neighborhoods in the abiertos data frame.
- Then we modify the nombres list so it matches exactly the br list. We called barrios_norm to this new list.
- Now we use a dictionary to combine both lists and replace the values of barrios in the population data frame.

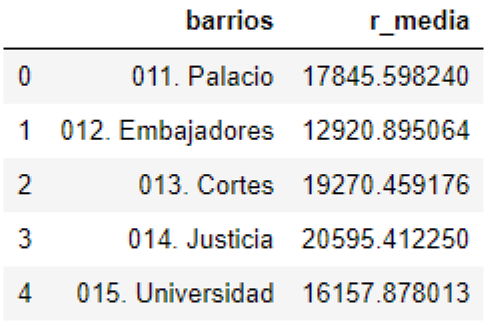
Because the file is from 2016, we don't have information for some neighborhoods, so we decide to fill the null values using the mean.

Next, we sort the values of the population data frame, reset the index, and normalize the data. The new data frame looks like this:



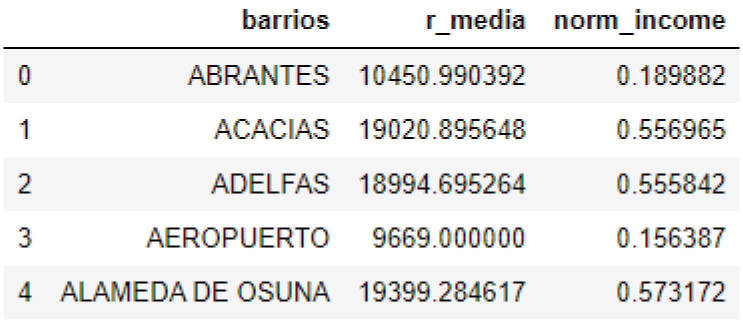
	barrios	nor_age	nor_pop
0	ABRANTES	0.863129	0.366776
1	ACACIAS	0.923363	0.458868
2	ADELFAS	0.896985	0.226434
3	AEROPUERTO	0.836366	0.021990
4	ALAMEDA DE OSUNA	0.886263	0.244206

We apply the same process to the average income (‘renta media’) data. We go from this:



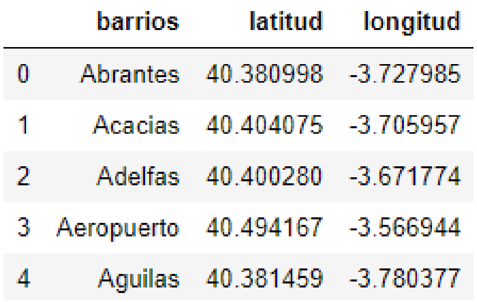
	barrios	r_media
0	011. Palacio	17845.598240
1	012. Embajadores	12920.895064
2	013. Cortes	19270.459176
3	014. Justicia	20595.412250
4	015. Universidad	16157.878013

To this:



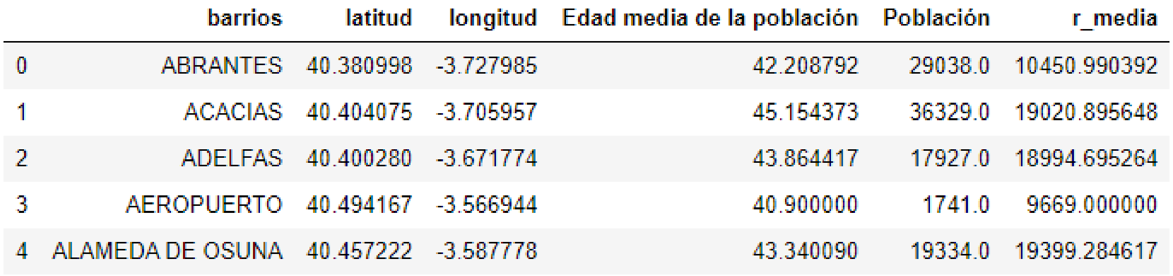
	barrios	r_media	norm_income
0	ABRANTES	10450.990392	0.189882
1	ACACIAS	19020.895648	0.556965
2	ADELFAS	18994.695264	0.555842
3	AEROPUERTO	9669.000000	0.156387
4	ALAMEDA DE OSUNA	19399.284617	0.573172

The next dataset that we need to process is the table with the geographic coordinates of each neighborhood. It looks like this:



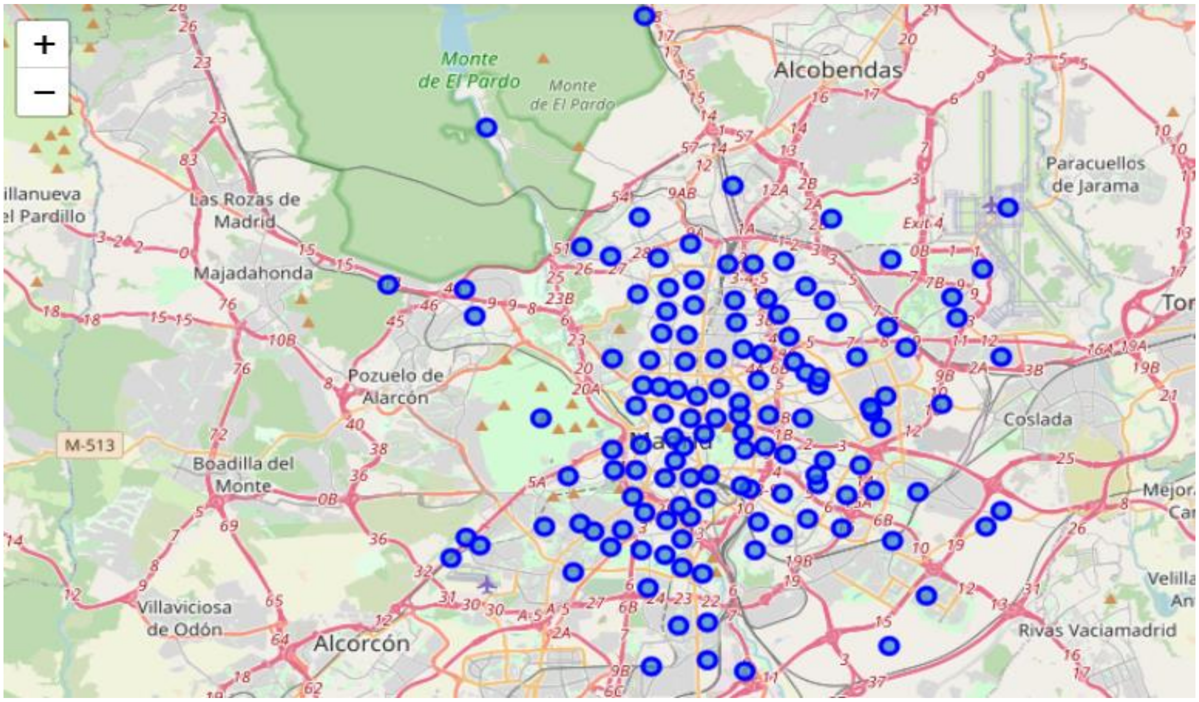
	barrios	latitud	longitud
0	Abrantes	40.380998	-3.727985
1	Acacias	40.404075	-3.705957
2	Adelfas	40.400280	-3.671774
3	Aeropuerto	40.494167	-3.566944
4	Aguilas	40.381459	-3.780377

After changing the values in the barrios column, we combine the coordinates data with the population and income data.



	barrios	latitud	longitud	Edad media de la población	Población	r_media
0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392
1	ACACIAS	40.404075	-3.705957	45.154373	36329.0	19020.895648
2	ADELFAS	40.400280	-3.671774	43.864417	17927.0	18994.695264
3	AEROPUERTO	40.494167	-3.566944	40.900000	1741.0	9669.000000
4	ALAMEDA DE OSUNA	40.457222	-3.587778	43.340090	19334.0	19399.284617

Now we can represent the neighborhoods on a map of Madrid.



METHODOLOGY

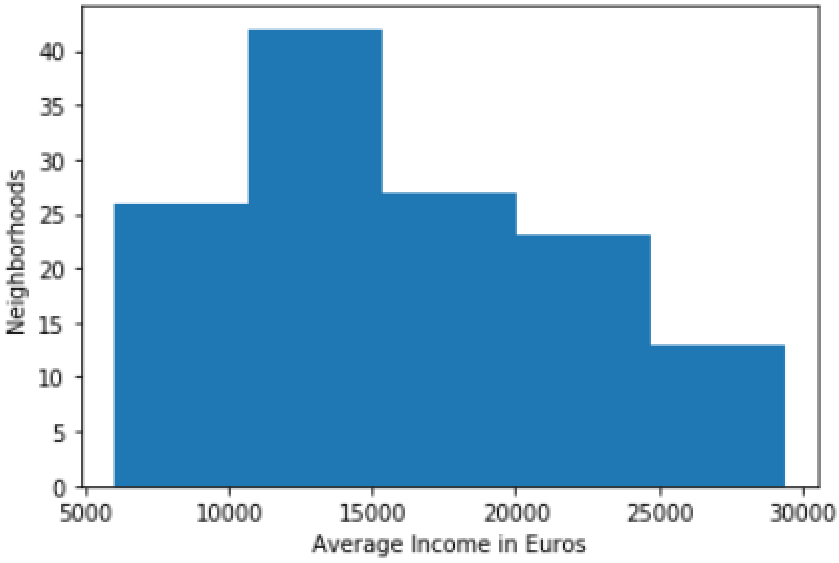
Now that we have all the data ready, we can start the analysis. We will:

- Analyze the age, population and income data to see how are distributed, and add level labels to the data: low, medium-low, medium, medium-high and high.
- Combine all this data with the venues data and apply KMeans algorithm to cluster the neighborhoods.
- Compare the income, age and population data of all the clusters.
- Compare the most common venues of each cluster.
- Represent all this information in maps.

ANALYSIS

Income Data

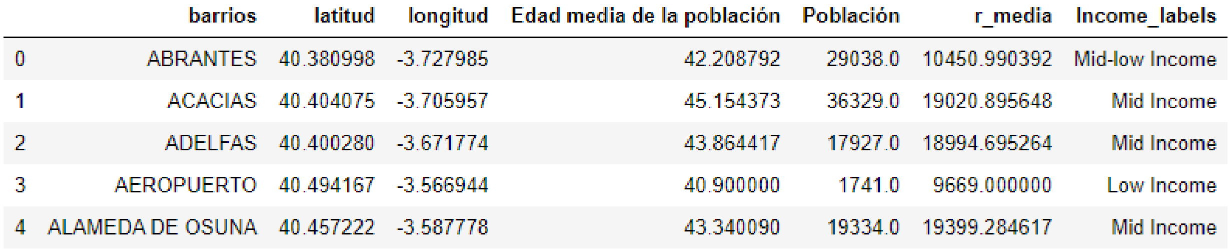
Let's plot a histogram of the income data grouped in 5 bins.



Based on this, the average income levels are:

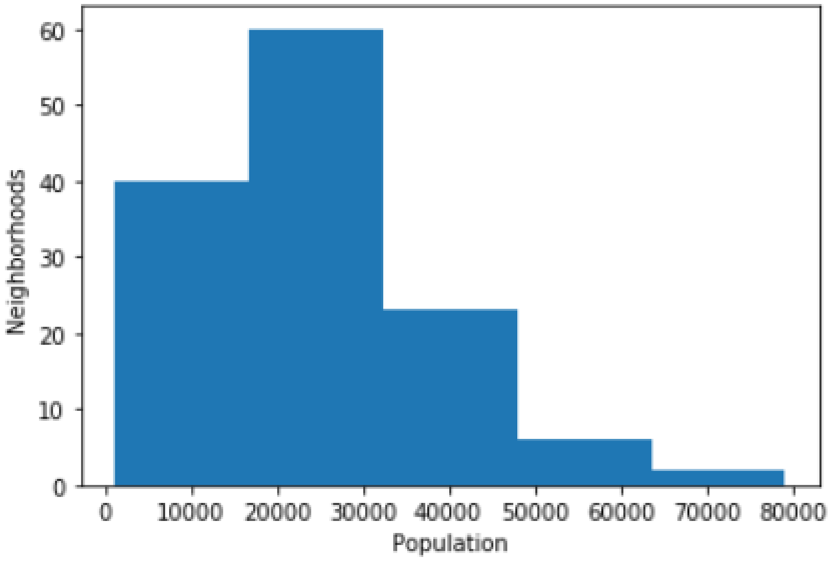
- Low Income: 10000 euros
- Mid-low Income: 15000-10000 euros
- Mid Income: 20000-15000 euros
- Mid-high Income: 25000-20000 euros
- High Income: 30000-25000 euros

We can create the labels with these levels and add them to the barrios data frame:



	barrios	latitud	longitud	Edad media de la población	Población	r_media	Income_labels
0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income
1	ACACIAS	40.404075	-3.705957	45.154373	36329.0	19020.895648	Mid Income
2	ADELFAS	40.400280	-3.671774	43.864417	17927.0	18994.695264	Mid Income
3	AEROPUERTO	40.494167	-3.566944	40.900000	1741.0	9669.000000	Low Income
4	ALAMEDA DE OSUNA	40.457222	-3.587778	43.340090	19334.0	19399.284617	Mid Income

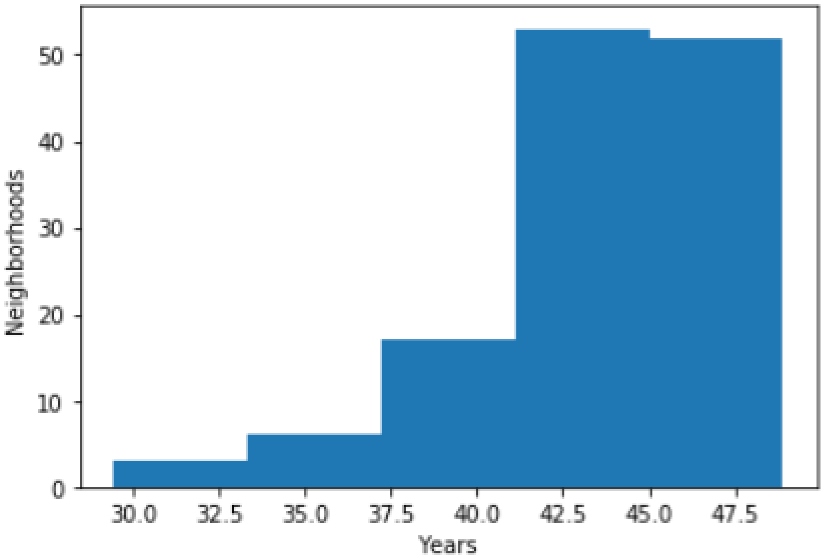
Population Data



Based on this, the population levels are:

- Low pop: 15000 people
- Mid-low pop: 30000-15000 people
- Mid pop: 45000-30000 people
- Mid-high pop: 65000-45000 people
- High pop: 80000-65000 people

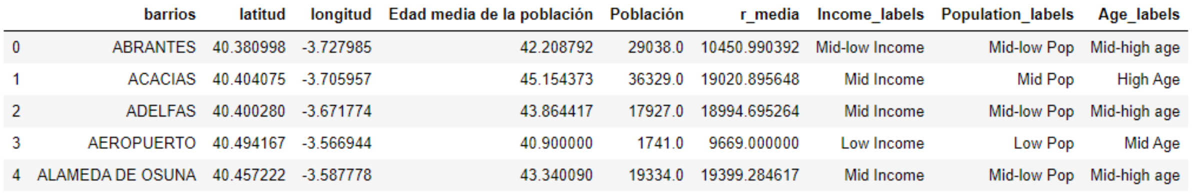
Average Age Data



Based on this, we can clasify the neighborhoods in:

- Low age: 33 years
- Mid-low age: 33-37.5 years
- Mid age: 41-37.5 years
- Mid-high age: 45-41 years
- High age: 49-45 years

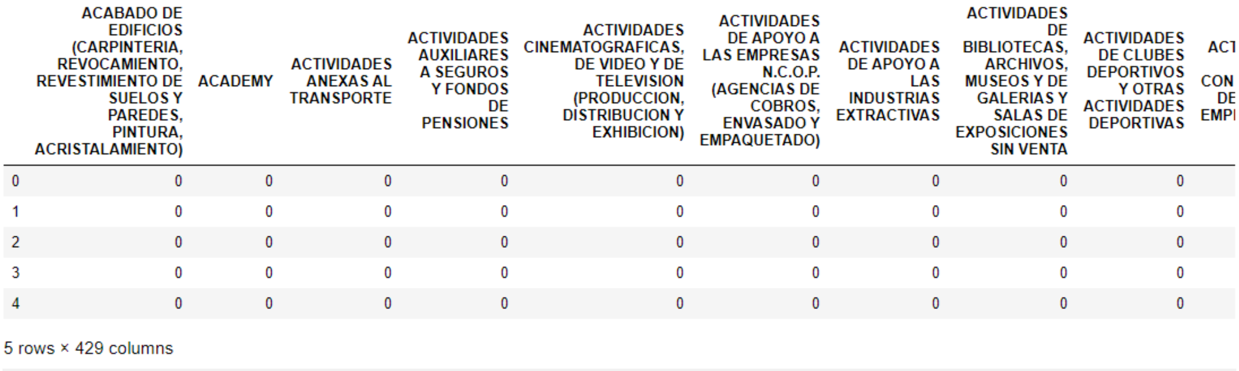
Once we create the labels for population and average data and add them to **barrios**, the data frame looks like this:



	barrios	latitud	longitud	Edad media de la población	Población	r_media	Income_labels	Population_labels	Age_labels
0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age
1	ACACIAS	40.404075	-3.705957	45.154373	36329.0	19020.895648	Mid Income	Mid Pop	High Age
2	ADELFAS	40.400280	-3.671774	43.864417	17927.0	18994.695264	Mid Income	Mid-low Pop	Mid-high age
3	AEROPUERTO	40.494167	-3.566944	40.900000	1741.0	9669.000000	Low Income	Low Pop	Mid Age
4	ALAMEDA DE OSUNA	40.457222	-3.587778	43.340090	19334.0	19399.284617	Mid Income	Mid-low Pop	Mid-high age

Cluster Analysis

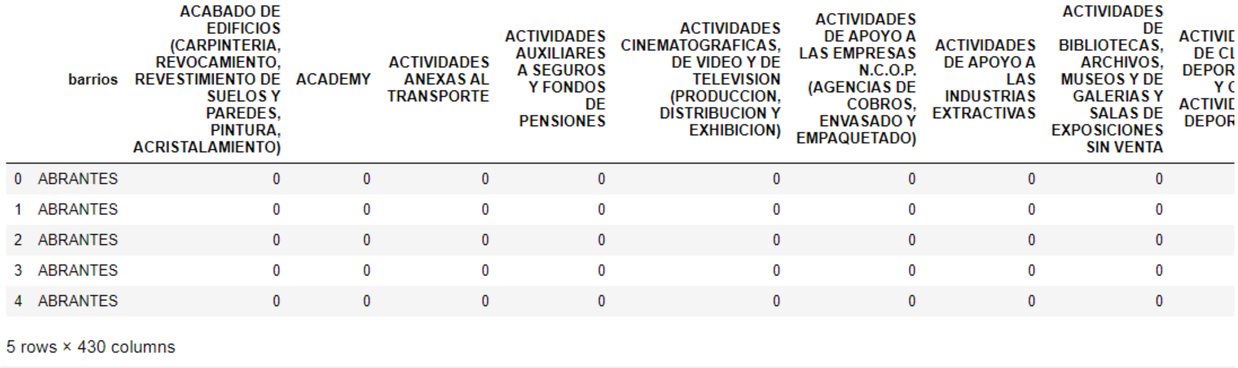
Using the abiertos data frame, let's find out how many unique activities can be curated from all venues. There are 429. Now we use one-hot encoding to transform all the unique activities into numbers.



	ACABADO DE EDIFICIOS (CARPINTERIA, REVOCAMIENTO, REVESTIMIENTO DE SUELOS Y PAREDES, PINTURA, ACRISTALAMIENTO)	ACADEMY	ACTIVIDADES ANEXAS AL TRANSPORTE	ACTIVIDADES AUXILIARES A SEGUROS Y FONDOS DE PENSIONES	ACTIVIDADES CINEMATOGRAFICAS, DE VIDEO Y DE TELEVISION (PRODUCCION, DISTRIBUCION Y EXHIBICION)	ACTIVIDADES DE APOYO A LAS EMPRESAS N.C.O.P. (AGENCIAS DE COBROS, ENVASADO Y EMPAQUETADO)	ACTIVIDADES DE APOYO A LAS INDUSTRIAS EXTRACTIVAS	ACTIVIDADES DE BIBLIOTECAS, ARCHIVOS, MUSEOS Y DE GALERIAS Y SALAS DE EXPOSICIONES SIN VENTA	ACTIVIDADES DE CLUBES DEPORTIVOS Y OTRAS ACTIVIDADES DEPORTIVAS	ACT CON DE EMP
0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	

5 rows × 429 columns

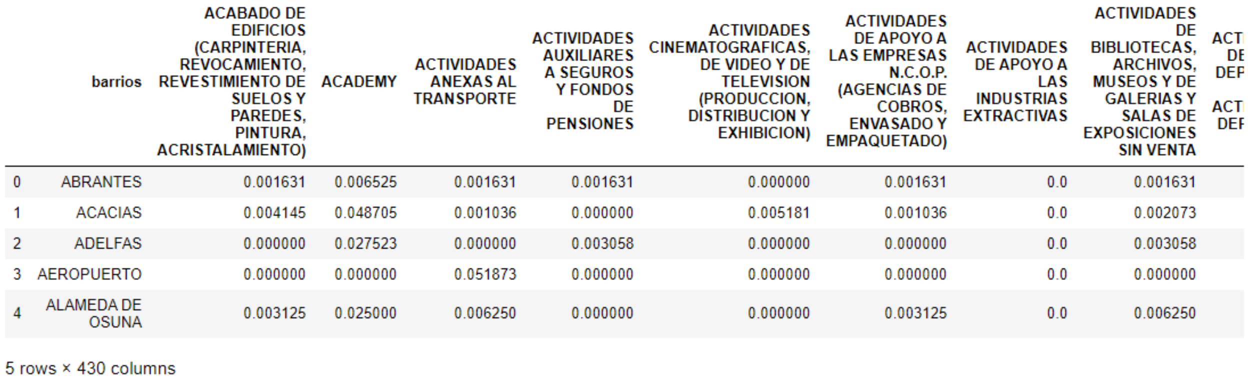
Then we merge abiertos_onehot to the 'barrios' column of abiertos.



	barrios	ACABADO DE EDIFICIOS (CARPINTERIA, REVOCAMIENTO, REVESTIMIENTO DE SUELOS Y PAREDES, PINTURA, ACRISTALAMIENTO)	ACADEMY	ACTIVIDADES ANEXAS AL TRANSPORTE	ACTIVIDADES AUXILIARES A SEGUROS Y FONDOS DE PENSIONES	ACTIVIDADES CINEMATOGRAFICAS, DE VIDEO Y DE TELEVISION (PRODUCCION, DISTRIBUCION Y EXHIBICION)	ACTIVIDADES DE APOYO A LAS EMPRESAS N.C.O.P. (AGENCIAS DE COBROS, ENVASADO Y EMPAQUETADO)	ACTIVIDADES DE APOYO A LAS INDUSTRIAS EXTRACTIVAS	ACTIVIDADES DE BIBLIOTECAS, ARCHIVOS, MUSEOS Y DE GALERIAS Y SALAS DE EXPOSICIONES SIN VENTA	ACTIVID DE CL DEPOR Y C ACTIVID DEPOR
0	ABRANTES	0	0	0	0	0	0	0	0	
1	ABRANTES	0	0	0	0	0	0	0	0	
2	ABRANTES	0	0	0	0	0	0	0	0	
3	ABRANTES	0	0	0	0	0	0	0	0	
4	ABRANTES	0	0	0	0	0	0	0	0	

5 rows × 430 columns

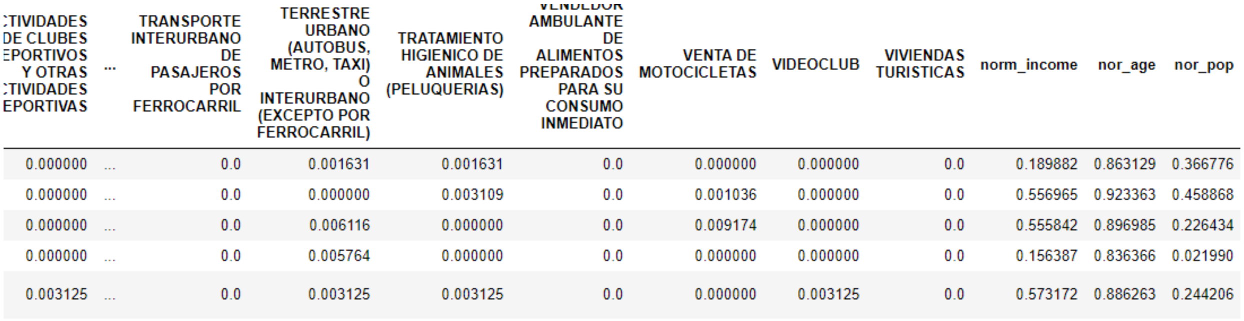
Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each activity.



	barrios	ACABADO DE EDIFICIOS (CARPINTERIA, REVOCAMIENTO, REVESTIMIENTO DE SUELOS Y PAREDES, PINTURA, ACRISTALAMIENTO)	ACADEMY	ACTIVIDADES ANEXAS AL TRANSPORTE	ACTIVIDADES AUXILIARES A SEGUROS Y FONDOS DE PENSIONES	ACTIVIDADES CINEMATOGRAFICAS, DE VIDEO Y DE TELEVISION (PRODUCCION, DISTRIBUCION Y EXHIBICION)	ACTIVIDADES DE APOYO A LAS EMPRESAS N.C.O.P. (AGENCIAS DE COBROS, ENVASADO Y EMPAQUETADO)	ACTIVIDADES DE APOYO A LAS INDUSTRIAS EXTRACTIVAS	ACTIVIDADES DE BIBLIOTECAS, ARCHIVOS, MUSEOS Y DE GALERIAS Y SALAS DE EXPOSICIONES SIN VENTA	ACT DE DEF ACT DEF
0	ABRANTES	0.001631	0.006525	0.001631	0.001631	0.000000	0.001631	0.0	0.001631	
1	ACACIAS	0.004145	0.048705	0.001036	0.000000	0.005181	0.001036	0.0	0.002073	
2	ADELFAS	0.000000	0.027523	0.000000	0.003058	0.000000	0.000000	0.0	0.003058	
3	AEROPUERTO	0.000000	0.000000	0.051873	0.000000	0.000000	0.000000	0.0	0.000000	
4	ALAMEDA DE OSUNA	0.003125	0.025000	0.006250	0.000000	0.000000	0.003125	0.0	0.006250	

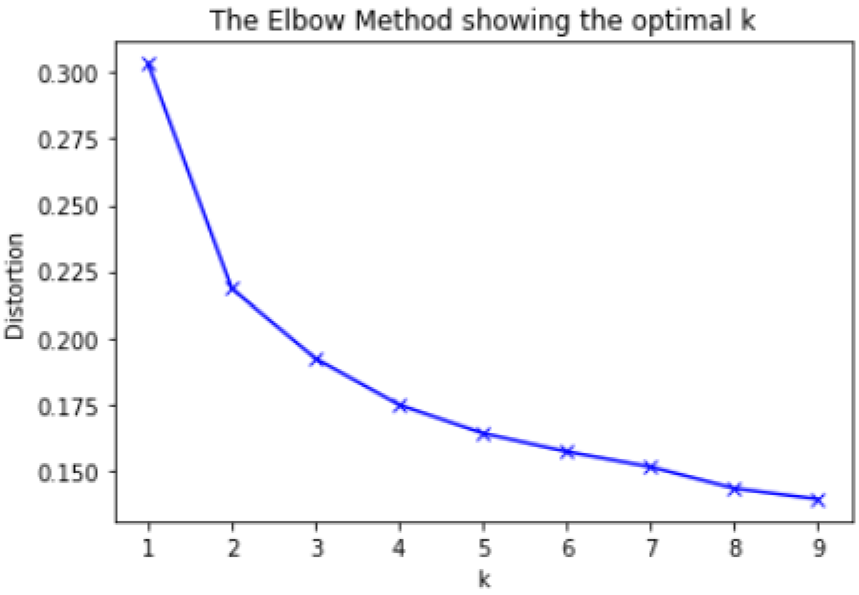
5 rows × 430 columns

Now we combine the venue's data with the normalized income, population, and age data.

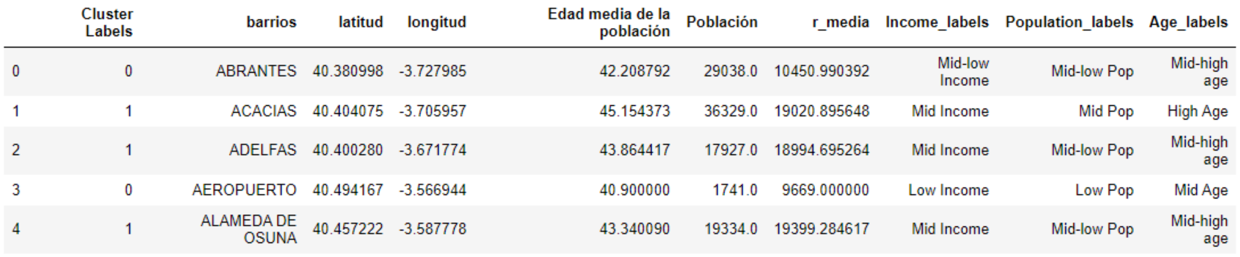


CTIVIDADES DE CLUBES EPORTIVOS Y OTRAS CTIVIDADES EPORTIVAS	...	TRANSPORTE INTERURBANO DE PASAJEROS POR FERROCARRIL	TERRESTRE URBANO (AUTOBUS, METRO, TAXI) O INTERURBANO (EXCEPTO POR FERROCARRIL)	TRATAMIENTO HIGIENICO DE ANIMALES (PELUQUERIAS)	VENDEDOR AMBULANTE DE ALIMENTOS PREPARADOS PARA SU CONSUMO INMEDIATO	VENTA DE MOTOCICLETAS	VIDEOCLUB	VIVIENDAS TURISTICAS	norm_income	nor_age	nor_pop
0.000000	...	0.0	0.001631	0.001631	0.0	0.000000	0.000000	0.0	0.189882	0.863129	0.366776
0.000000	...	0.0	0.000000	0.003109	0.0	0.001036	0.000000	0.0	0.556965	0.923363	0.458868
0.000000	...	0.0	0.006116	0.000000	0.0	0.009174	0.000000	0.0	0.555842	0.896985	0.226434
0.000000	...	0.0	0.005764	0.000000	0.0	0.000000	0.000000	0.0	0.156387	0.836366	0.021990
0.003125	...	0.0	0.003125	0.003125	0.0	0.000000	0.003125	0.0	0.573172	0.886263	0.244206

Now we can perform cluster analysis applying the KMeans algorithm. We use the elbow method to determine the optimal number of clusters.

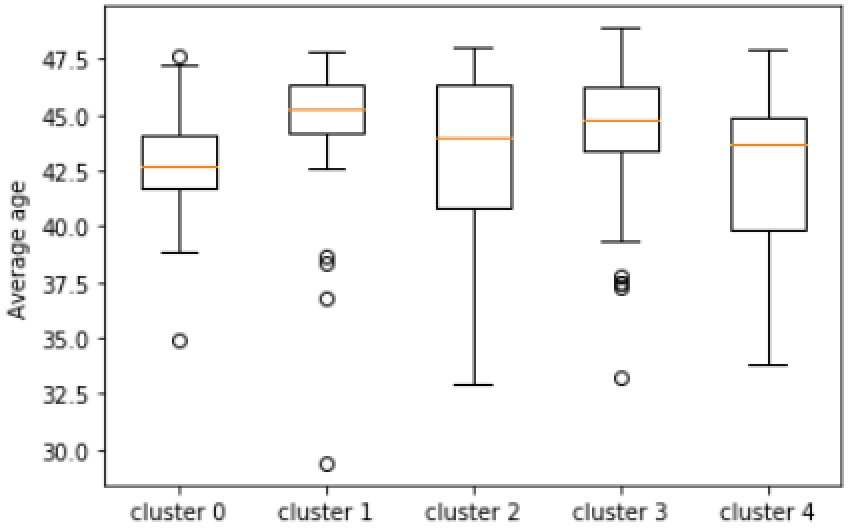


We run the algorithm with 5 clusters and add the cluster labels to barrios

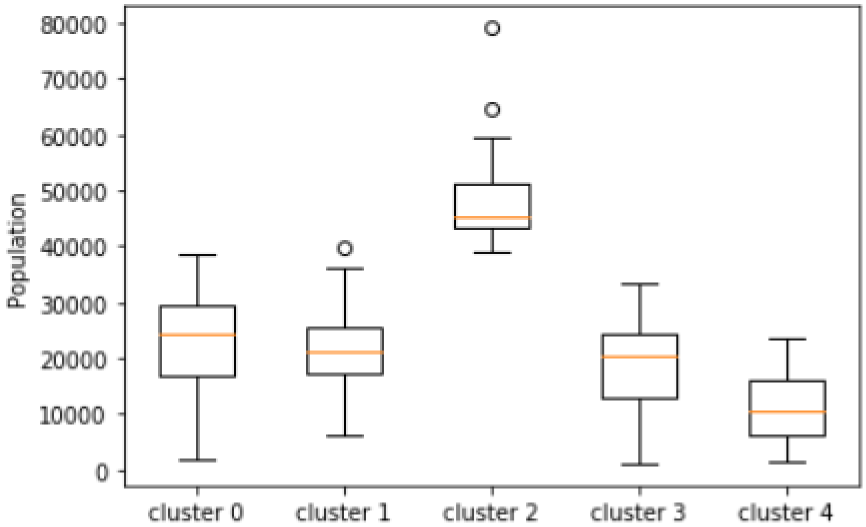


	Cluster Labels	barrios	latitud	longitud	Edad media de la población	Población	r_media	Income_labels	Population_labels	Age_labels
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age
1	1	ACACIAS	40.404075	-3.705957	45.154373	36329.0	19020.895648	Mid Income	Mid Pop	High Age
2	1	ADELFAS	40.400280	-3.671774	43.864417	17927.0	18994.695264	Mid Income	Mid-low Pop	Mid-high age
3	0	AEROPUERTO	40.494167	-3.566944	40.900000	1741.0	9669.000000	Low Income	Low Pop	Mid Age
4	1	ALAMEDA DE OSUNA	40.457222	-3.587778	43.340090	19334.0	19399.284617	Mid Income	Mid-low Pop	Mid-high age

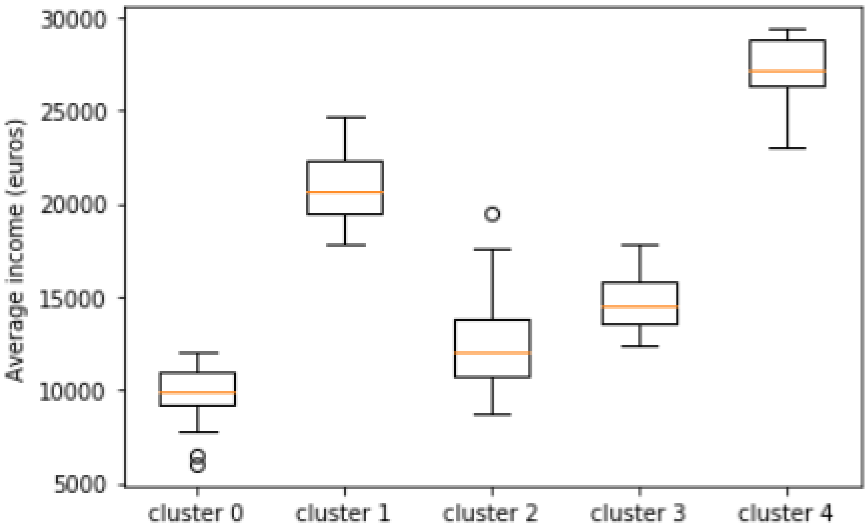
Let's plot boxplots of the average age, average income, and population of the neighborhoods of each cluster to see if we can find any patterns.



We have seen already that Madrid is an 'old' city and most of the neighborhoods have an average age of 45 years or more. The few neighborhoods with 'low_age' label appear as outliers in cluster 1 and give a negative skewness to cluster 2. 'Mid-low age' neighborhoods appear as outliers in clusters 0 and 3 and give a negative skewness to cluster 4.

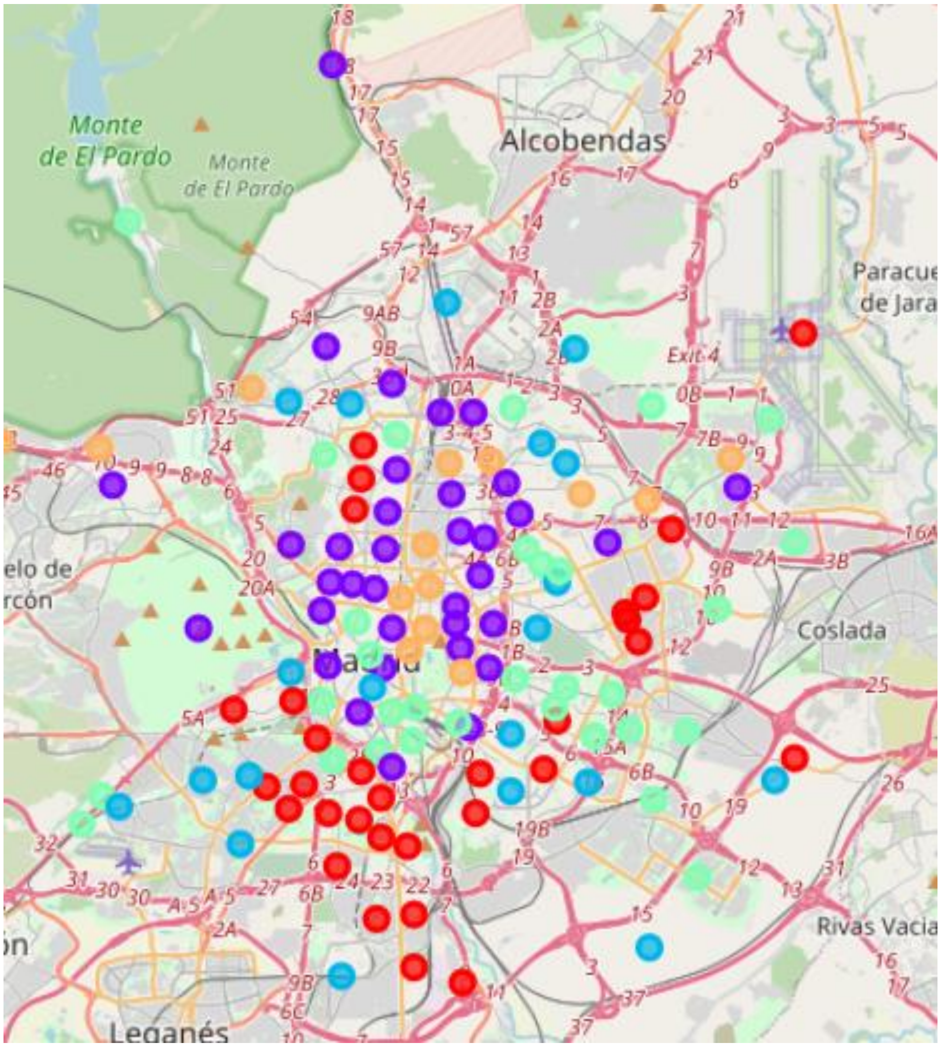


It's quite clear that cluster 2 groups the most populated neighborhoods and cluster 4 the least populated.

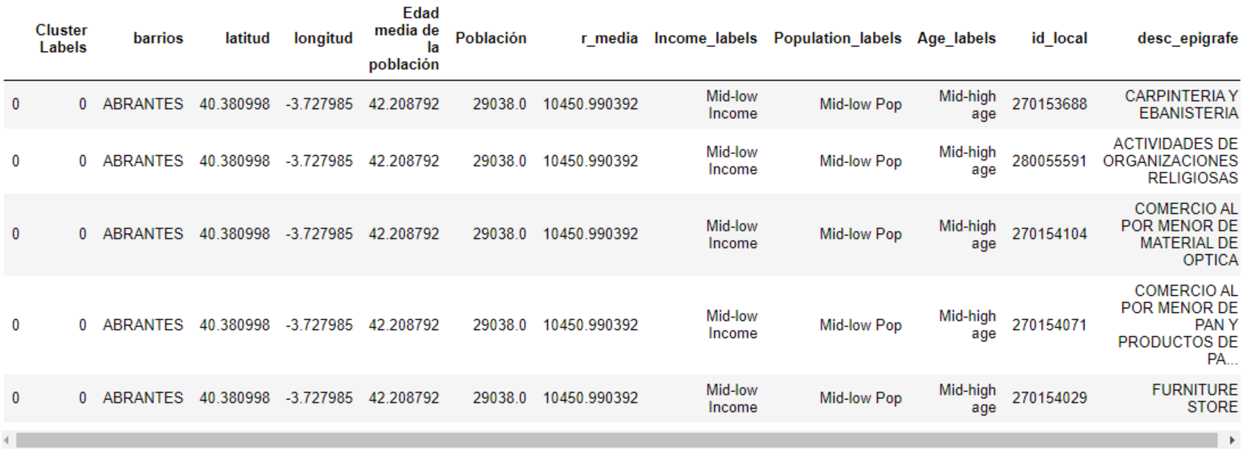


The order of the clusters from the richest to poorest is cluster 4, cluster 1, cluster 3, cluster 2, and cluster 0.

Let's see the clusters in the map:



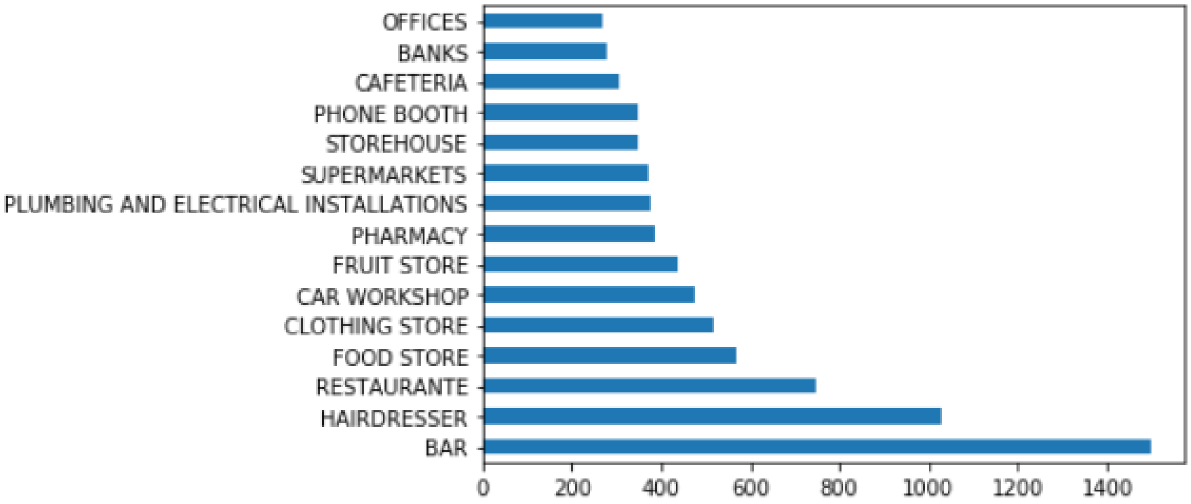
Let's check which are the 15 most common venues for each cluster. First we merge barrios and abiertos.



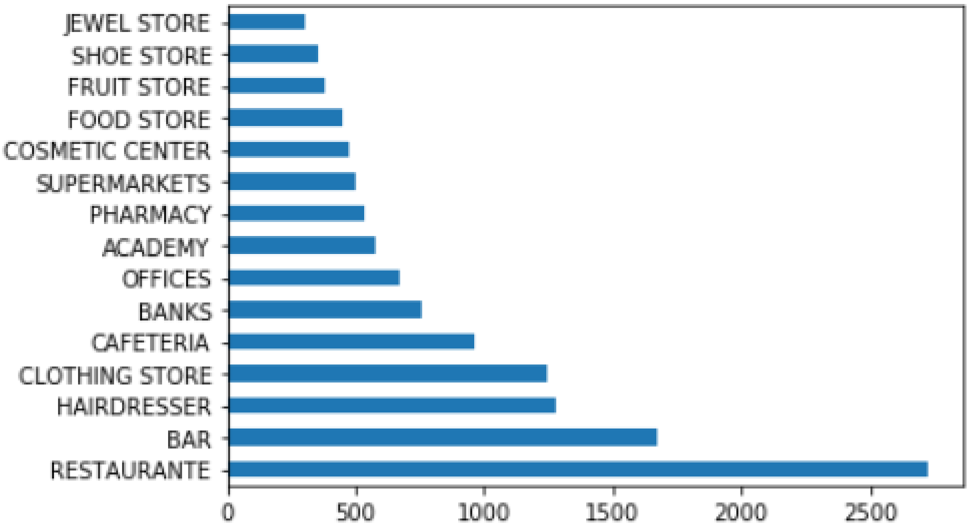
	Cluster Labels	barrios	latitud	longitud	Edad media de la población	Población	r_media	Income_labels	Population_labels	Age_labels	id_local	desc_epigrafe
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age	270153688	CARPINTERIA Y EBANISTERIA
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age	280055591	ACTIVIDADES DE ORGANIZACIONES RELIGIOSAS
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age	270154104	COMERCIO AL POR MENOR DE MATERIAL DE OPTICA
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age	270154071	COMERCIO AL POR MENOR DE PAN Y PRODUCTOS DE PA...
0	0	ABRANTES	40.380998	-3.727985	42.208792	29038.0	10450.990392	Mid-low Income	Mid-low Pop	Mid-high age	270154029	FURNITURE STORE

Now we extract the data for each cluster label and plot the 15 most common venues for each of them.

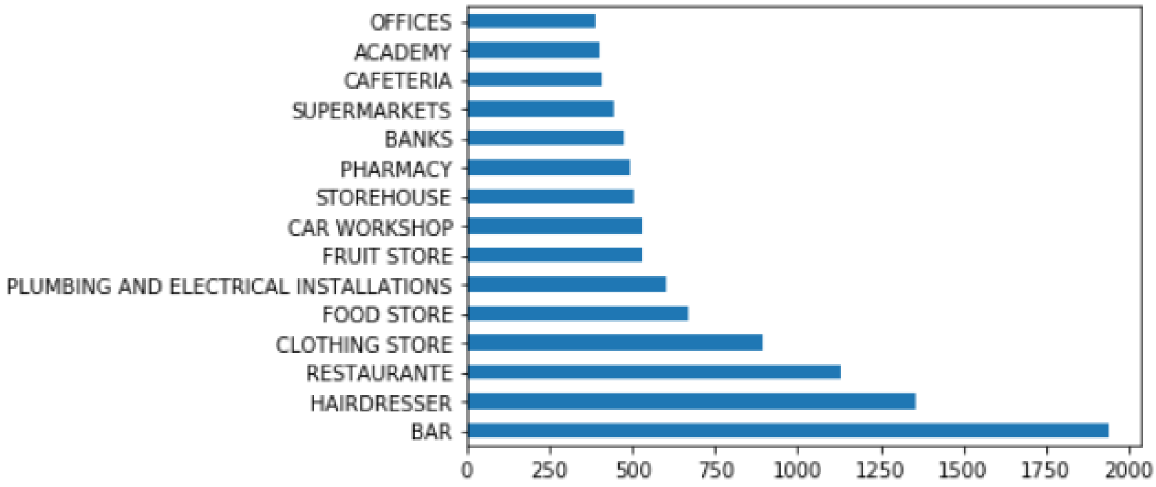
Cluster 0



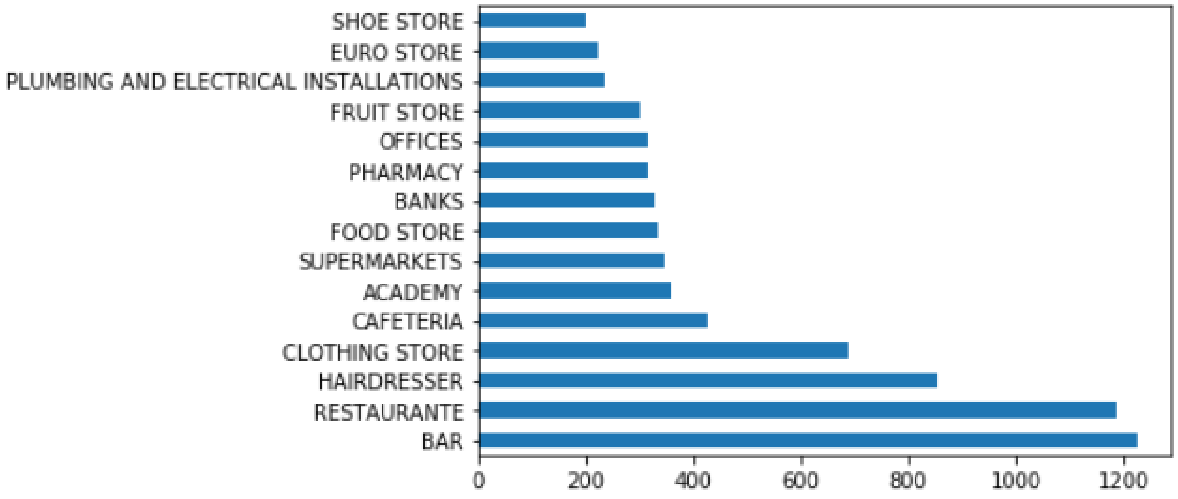
Cluster 1



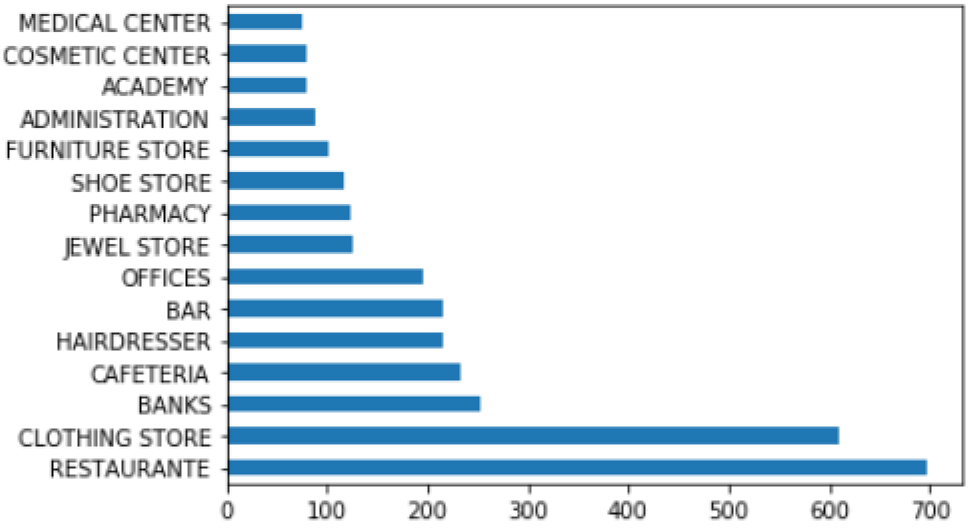
Cluster 2



Cluster 3



Cluster 4

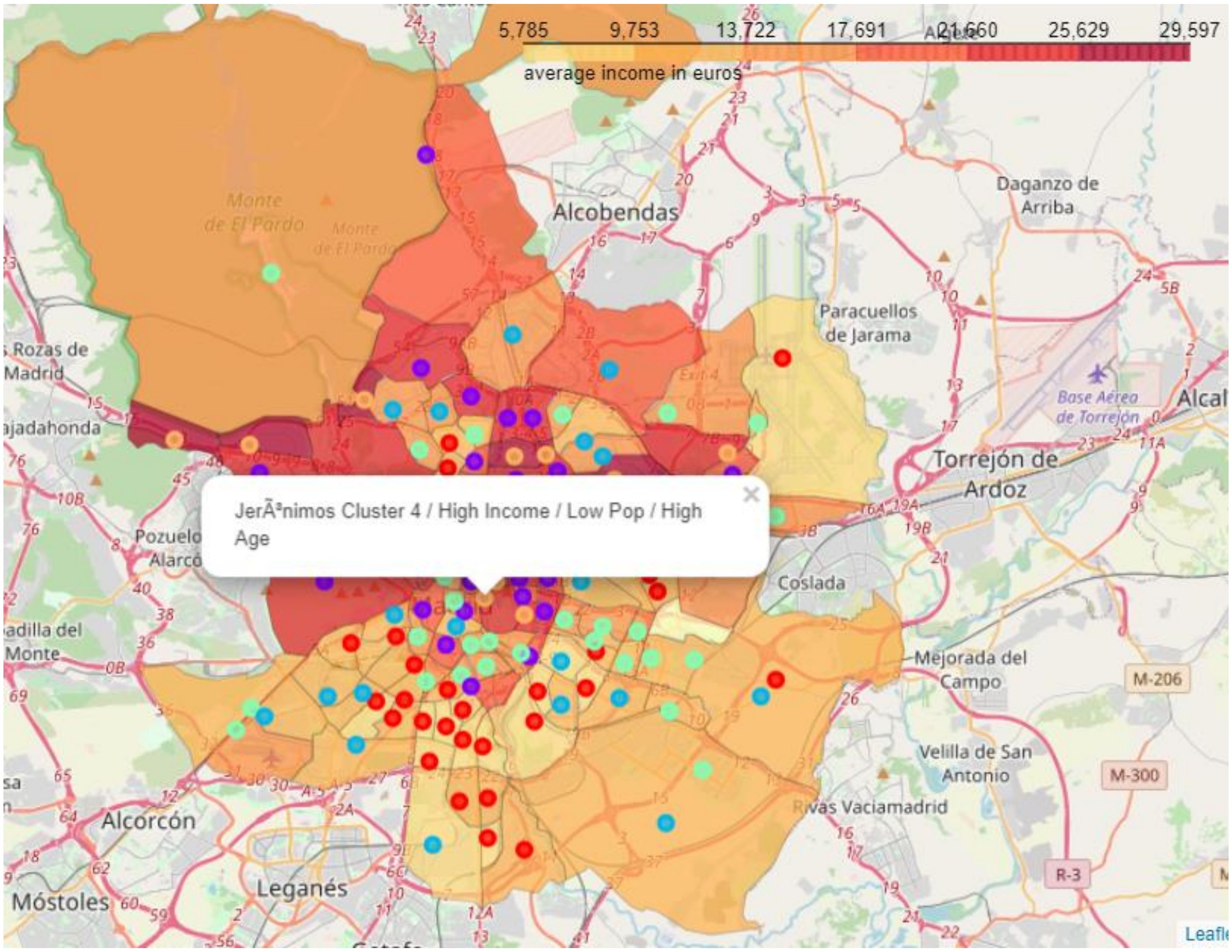


Now we are going to represent the clusters and the other data in choropleth maps. To do that we need a Geojson file with the geographical data of Madrid's neighborhoods. Sadly the file doesn't include the newest neighborhoods: Ensanche de Vallecas, Valdebernardo, Valderribas, and El Cañaveral.

To bind barrios and the geojson file successfully, the name of the neighborhood in barrios must match exactly the name of the neighborhood in the geojson file. Therefore, we have to analyze which names are contained in the geojson file and modify the names in barrios accordingly.

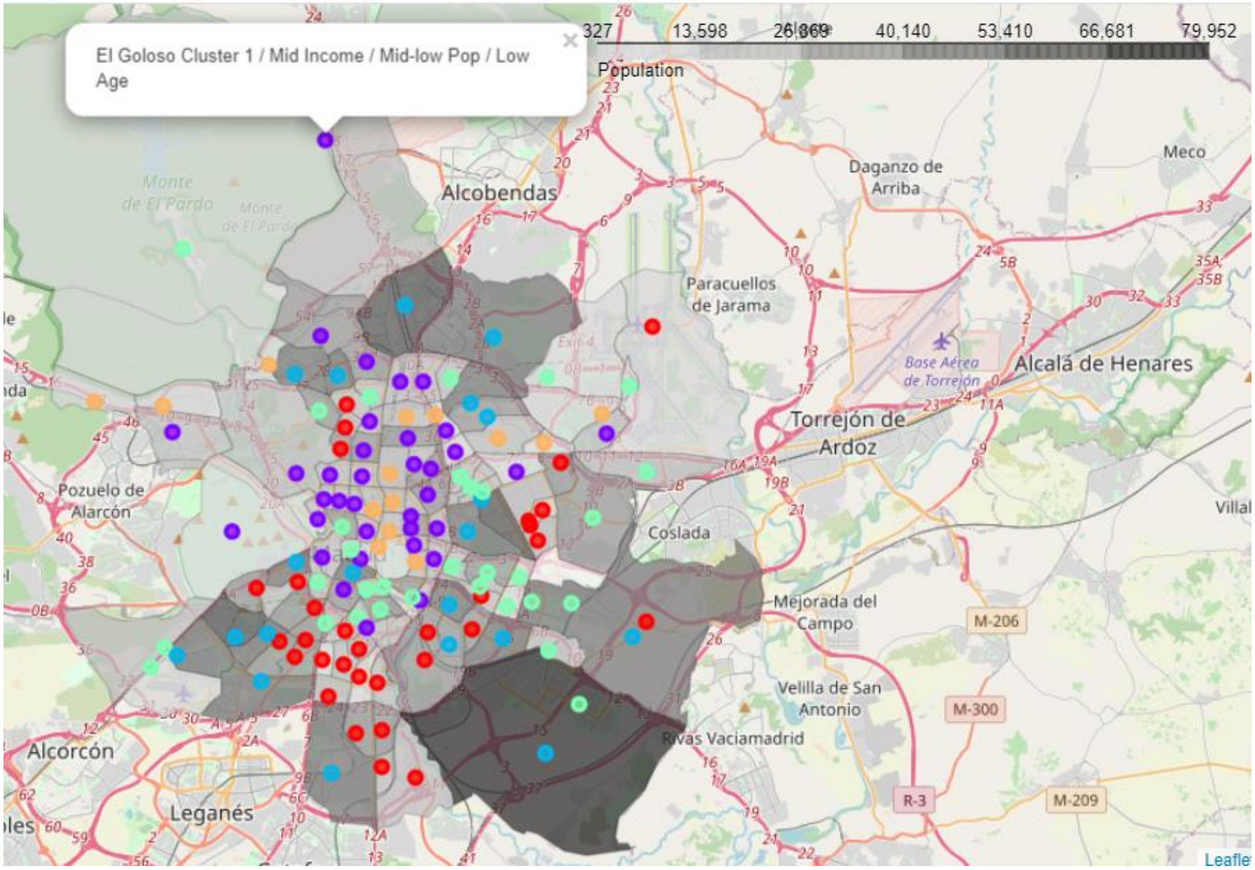
As shown above, the json.load() method returns a Python dictionary. Then, we loop through the dictionary to obtain the names of the neighborhoods. We have created a list with geojson file names in the same order as they appear in barrios data frame.

Let's see the average income map, we have added all the labels to the pop-up.

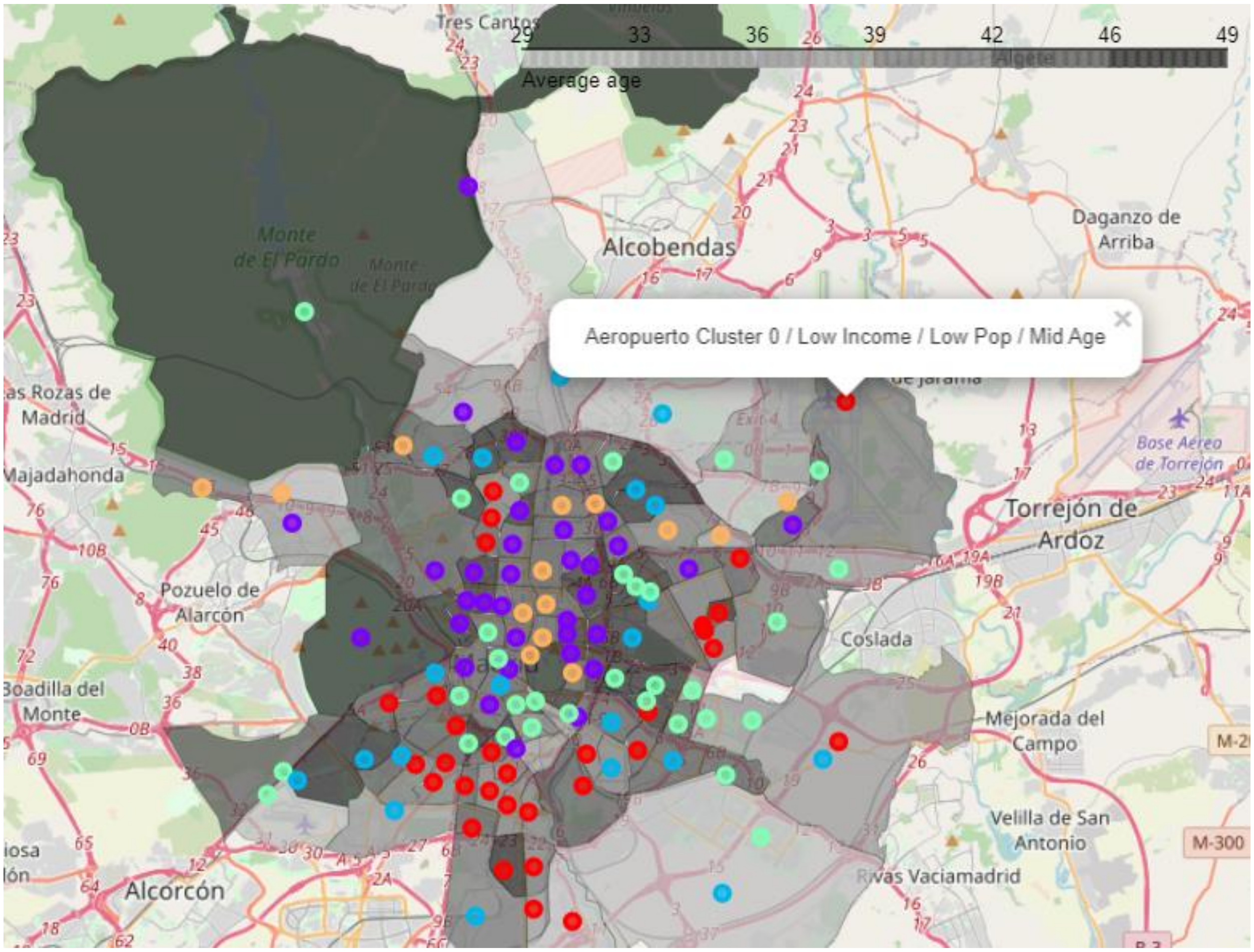


JerÃ³nimos Cluster 4 / High Income / Low Pop / High Age

Now the population map:



And the average age map



Results and Discussion

Based on all the analysis performed in the data, these are all the observations that we have made:

- Cluster 0 (Red) groups neighborhoods with low and medium-low average income, mid-high average age, and medium-low population. They are mostly located in the periphery. Some of the most common venues are car workshops, plumbing and electrical installations, and phone booths.
- Cluster 1 (Purple) groups neighborhoods with medium and medium-high average income, high and medium-high average age, and medium-low population. They are located mostly in the center of the city. There are some neighborhoods in the periphery but none of them are in the south. Some of their most common activities are clothing stores, banks, offices, cosmetic centers, and jewel stores.
- Cluster 2 (Blue) groups neighborhoods with mostly medium-low average income, high and medium-high average age, and medium and medium-high population. They are all in the periphery with the exception of Embajadores. Some of their most common activities are plumbing and electrical installations, car workshops, and storehouses.

- Cluster 3 (Cyan) groups neighborhoods with medium and medium-low average income, high and medium-high average age, and medium-low population. They are located all around the city. It has more academies and offices than plumbing and electrical installations. It also has 'euro shops'.
- Cluster 4 (Orange) groups neighborhoods with high average income, high and medium-high average age, and low and medium-low population. They are in the center and north of the city. The most common activities are banks, offices, stores, administration, and medical centers.

We can see that there's a clear divide between north and south and between the center and the periphery of the city in terms of income, population, and the activities performed. The richest and least populated neighborhoods (clusters 1 and 4) are in the center and north of the city and concentrate commercial, administrative and financial activities.

Some of the neighborhoods with average income and low population (cluster 3) are in the center of the city (Sol, Universidad) but the rest are in the periphery. That explains that the activities performed are quite diverse. Sol is one of the most commercial neighborhoods and attracts a lot of tourists, but other neighborhoods have a more industrial or administrative profile.

The poorest and most populated neighborhoods (clusters 0 and 2) are all in the periphery, especially in the south. They have an industrial profile because some of the most common activities are car workshops, plumbing and electrical installations and storehouses. Many of these neighborhoods have a high immigrant population, which explains the more than 400 phone booths in cluster 0.

I haven't mention bars, restaurants, and hairdressers until now because they are all around the city, the only difference is how many there are. Restaurants are the most common activity in clusters 1 and 4, but in clusters 0 and 2 are the third most common. Bars are the most common activity in clusters 0, 2, and 3 and the second most common activity in cluster 1. Something similar happens with hairdressers, they are one of the five most common activities in all the clusters except cluster 4.

Conclusion

The purpose of this project was to classify Madrid's neighborhoods and help companies to identify the best parts of the city to locate their businesses. To do that we have combined venue data with income and population data for each neighborhood to cluster them. Thanks to this analysis we have identified the parts of the city with a financial and commercial profile, and the ones with a more industrial profile. Logically, the financial neighborhoods are also the richest, but also the least populated. We have also identified the 'youngest' neighborhoods and we have classified most of them as medium-high average income, which gives them a potential for growth.

Unfortunately, most of the data is from 2016, and Madrid has been badly affected by the Covid-19 pandemic. If we could perform the same analysis with updated data we are sure that the results would be quite different.