Madrid Airbnb Market

IBM Data Science Capstone Project

20/07/2021

—

Juan Manuel Valero

Introduction

Airbnb Is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. This has been increasing rented apartments significantly, and it is expected to continue increasing in the coming years.

Within the world of real estate, the option of using your apartment as Airbnb is increasing thanks to the profitability it provides. So in this project I will dedicate myself to analyze the Airbnb market in Madrid.

Business Problem

When it comes to investing in real estate for Airbnb, you have to mainly take into account the location of the flat, the venues around it and the type of real estate. Eventually when looking for neighborhoods to invest in, you come across neighborhoods with too much Airbnb supply, neighborhoods with no interest for tenants, or low rental prices.

That is why through the analysis of the Madrid Airbnb supply data, and the characteristic venues of each neighbourhood, I will find some neighbourhoods with investment potential.

Data

I. Madrid Airbnb Data

This set of data is made up of all the apartments advertised in Madrid, with their respective prices per night, location, etc. It comes from the Kaggle database website.

II. Madrid Neighbourhoods Coordinates

It is composed of the latitude and longitude of each neighbourhood, extracted from the python Geopy API.

III. Neighbourhoods Infrastructure

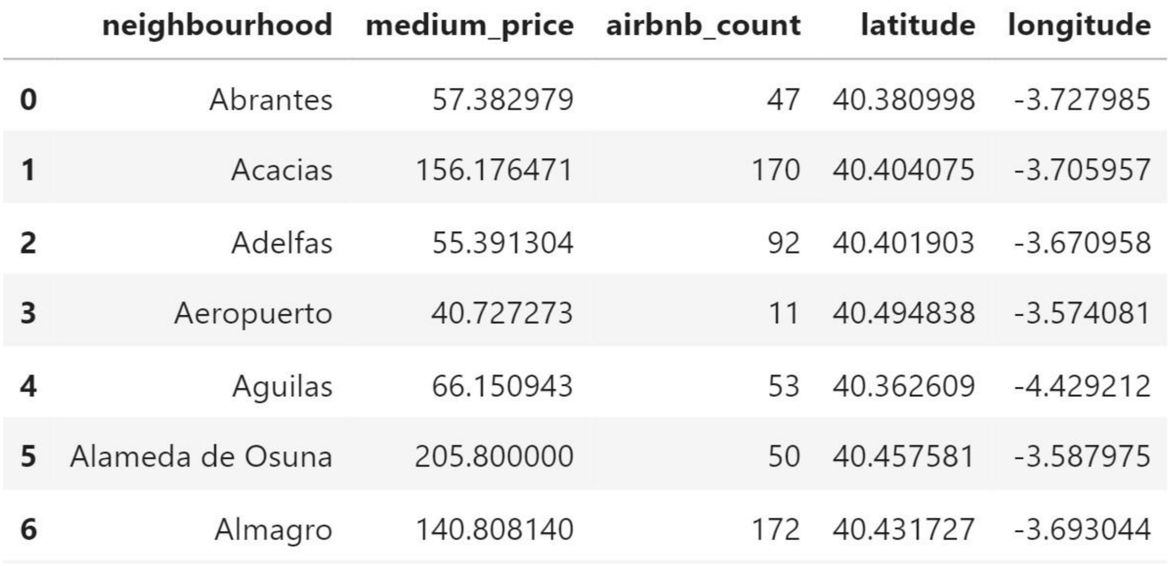
It is made up of the different venues in each neighbourhood, specifying the category, the name and the coordinates of the venue. This database was taken from the Foursquare API.

Methodology

In this project I will start obtaining the data from reliable resources, cleaning and modeling it. Then I will explore and analyse the data once the necessary data is selected. Decide which model fits better, using it, and finally extract the results with a respective conclusion.

Data Exploration

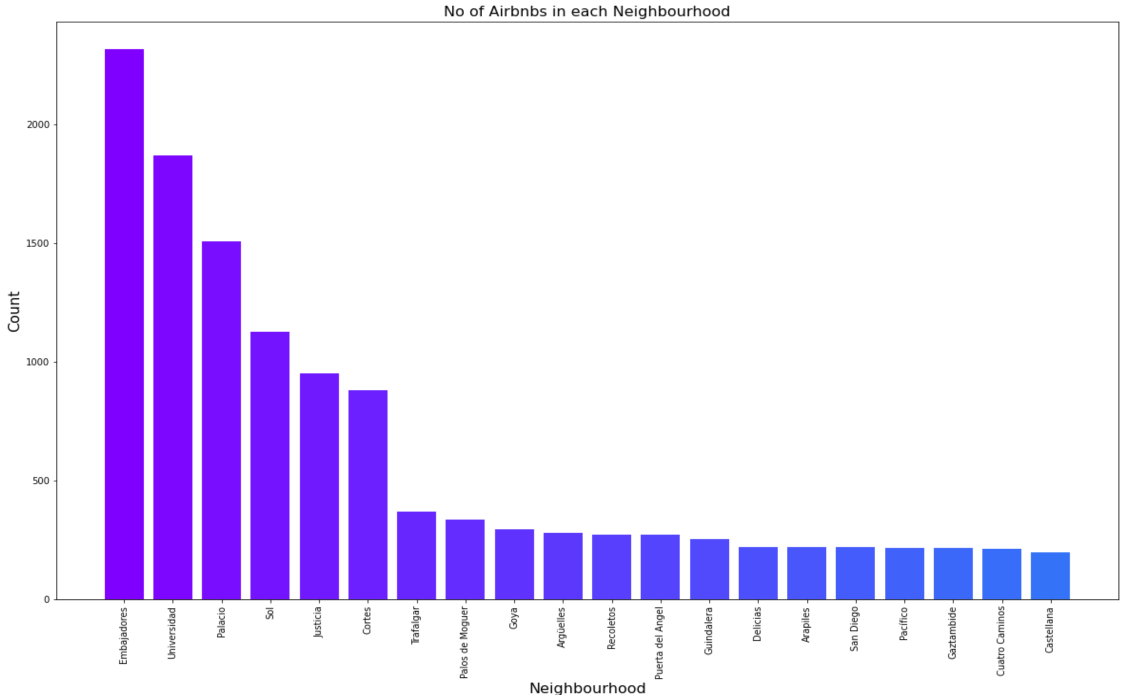
Firstly, I imported the Madrid Airbnb data and cleaned the irrelevant data. Then we get the latitude and longitude of each neighbourhood with the Geopy API, and merge both into an only dataset.



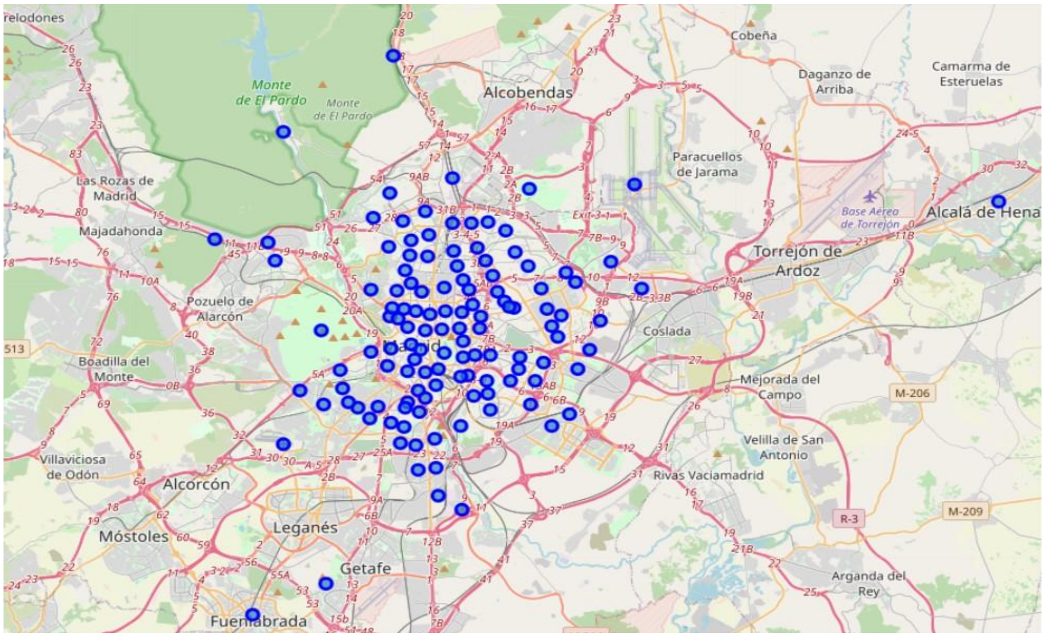
	neighbourhood	medium_price	airbnb_count	latitude	longitude
0	Abrantes	57.382979	47	40.380998	-3.727985
1	Acacias	156.176471	170	40.404075	-3.705957
2	Adelfas	55.391304	92	40.401903	-3.670958
3	Aeropuerto	40.727273	11	40.494838	-3.574081
4	Aguilas	66.150943	53	40.362609	-4.429212
5	Alameda de Osuna	205.800000	50	40.457581	-3.587975
6	Almagro	140.808140	172	40.431727	-3.693044

Data Visualization

We can see this previous data clearly, sorting the neighbourhoods by their number of Airbnbs, which means they are points of interest for the tenants. Having at the top Embajadores, Universidad, Palacio, Sol and Justicia.

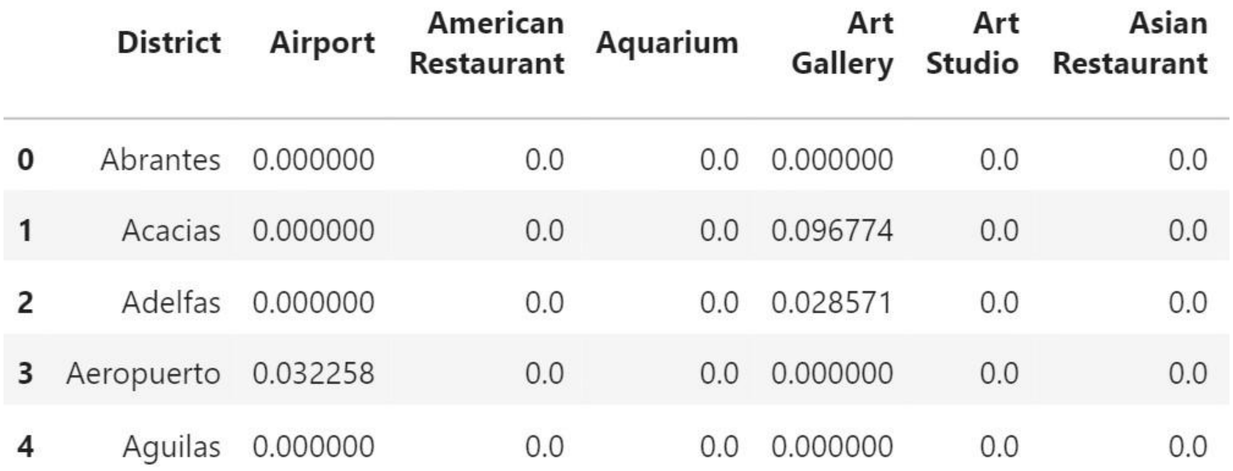


Then we plot all the neighbourhoods in a map with the folium library.



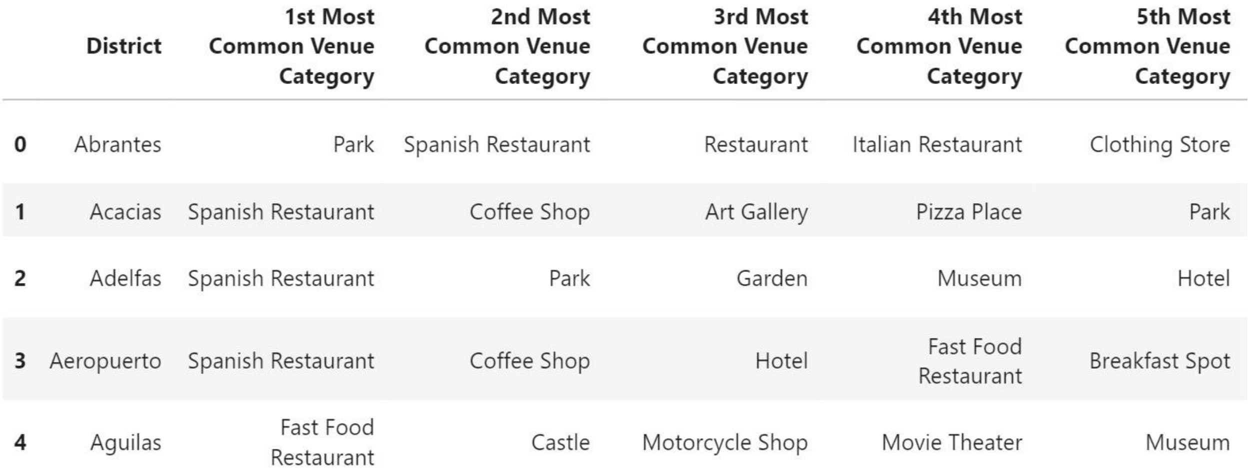
Neighbourhood Infrastructure Analysis

Using the Foursquare API, we get the venues of each neighbourhood with a distance of 25km, and only the venues with the required categories. We organize the unique venues categories obtained and create a one-hot encoding to analyse each neighbourhood. Then we group the rows by venue category and take the mean of the frequency of occurrence of each category.



	District	Airport	American Restaurant	Aquarium	Art Gallery	Art Studio	Asian Restaurant
0	Abrantes	0.000000	0.0	0.0	0.000000	0.0	0.0
1	Acacias	0.000000	0.0	0.0	0.096774	0.0	0.0
2	Adelfas	0.000000	0.0	0.0	0.028571	0.0	0.0
3	Aeropuerto	0.032258	0.0	0.0	0.000000	0.0	0.0
4	Aguilas	0.000000	0.0	0.0	0.000000	0.0	0.0

Finally we organize each neighbourhood according to its category and frequency.

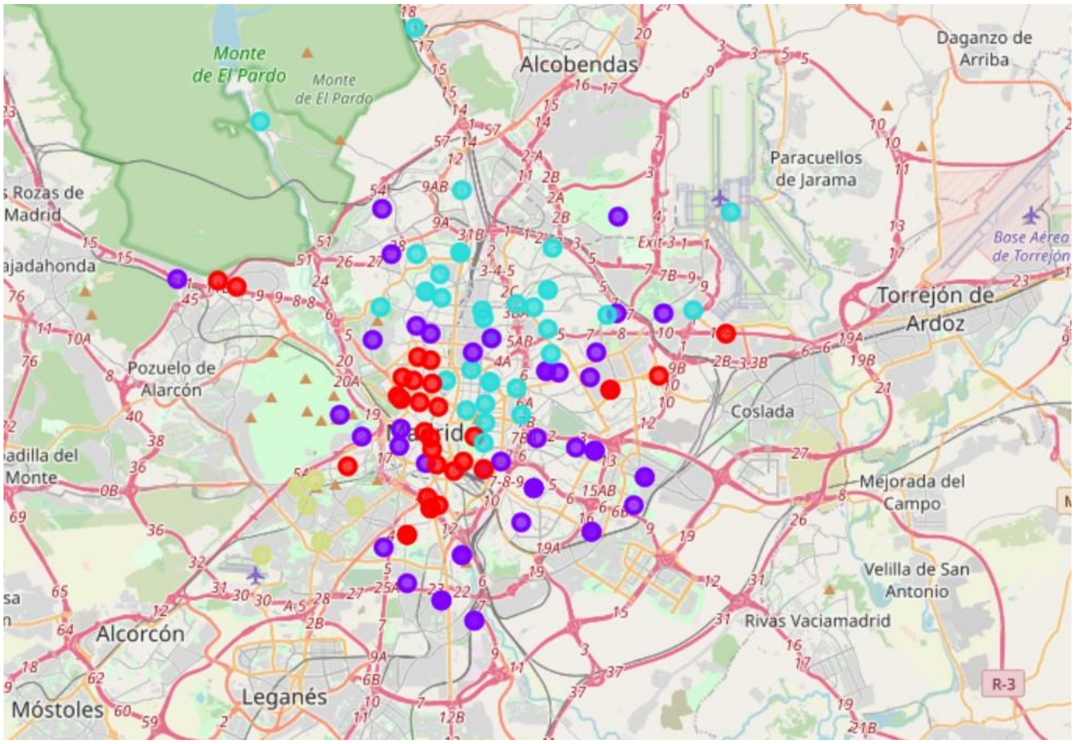


	District	1st Most Common Venue Category	2nd Most Common Venue Category	3rd Most Common Venue Category	4th Most Common Venue Category	5th Most Common Venue Category
0	Abrantes	Park	Spanish Restaurant	Restaurant	Italian Restaurant	Clothing Store
1	Acacias	Spanish Restaurant	Coffee Shop	Art Gallery	Pizza Place	Park
2	Adelfas	Spanish Restaurant	Park	Garden	Museum	Hotel
3	Aeropuerto	Spanish Restaurant	Coffee Shop	Hotel	Fast Food Restaurant	Breakfast Spot
4	Aguilas	Fast Food Restaurant	Castle	Motorcycle Shop	Movie Theater	Museum

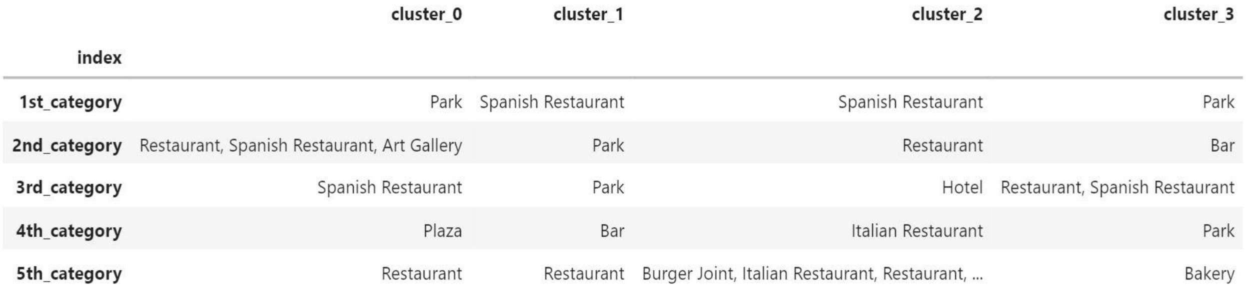
K-Means Clustering Analysis

We then use the k-means clustering algorithm to group the neighbourhoods into clusters based on its main venues, and we decide the number of clusters to use with the elbow method, concluding 4 clusters are the optimal.

The different neighbourhoods with their respective cluster number, are plotted into a map once again, with different colors to differentiate their cluster.
(Cluster 0 = Red, Cluster 1 = Purple, Cluster 2 = Blue, Cluster 3 = Yellow)

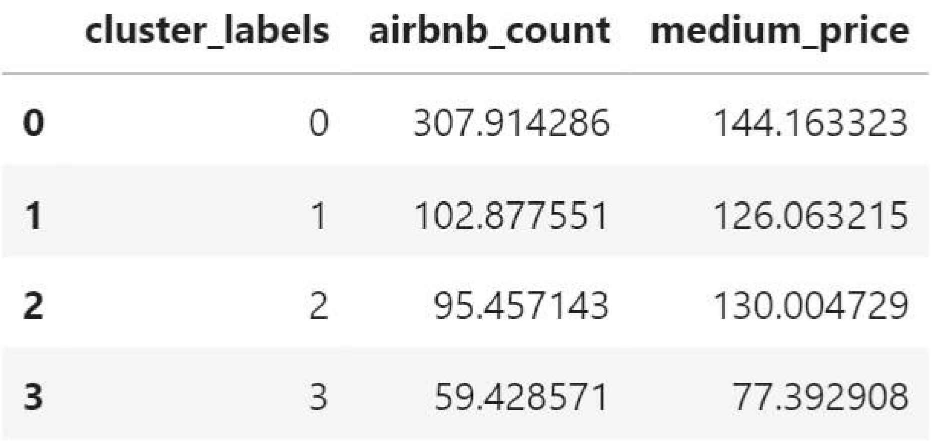


Having the clusters defined, we proceed to analyse each cluster. First by the most common venues in each cluster, although to find the significant differences you have to see all the venues in each cluster.



	cluster_0	cluster_1	cluster_2	cluster_3
index				
1st_category	Park	Spanish Restaurant	Spanish Restaurant	Park
2nd_category	Restaurant, Spanish Restaurant, Art Gallery	Park	Restaurant	Bar
3rd_category	Spanish Restaurant	Park	Hotel	Restaurant, Spanish Restaurant
4th_category	Plaza	Bar	Italian Restaurant	Park
5th_category	Restaurant	Restaurant	Burger Joint, Italian Restaurant, Restaurant, ...	Bakery

And also analyzing the clusters by the medium number of airbnbs and the medium price.



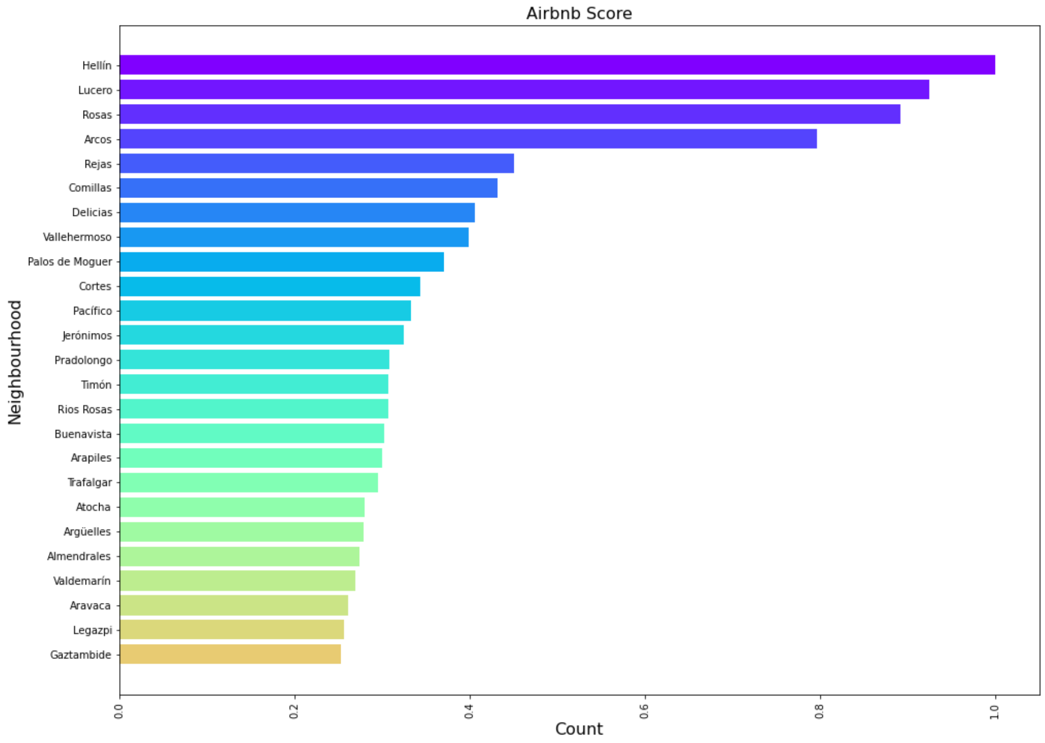
	cluster_labels	airbnb_count	medium_price
0	0	307.914286	144.163323
1	1	102.877551	126.063215
2	2	95.457143	130.004729
3	3	59.428571	77.392908

Results

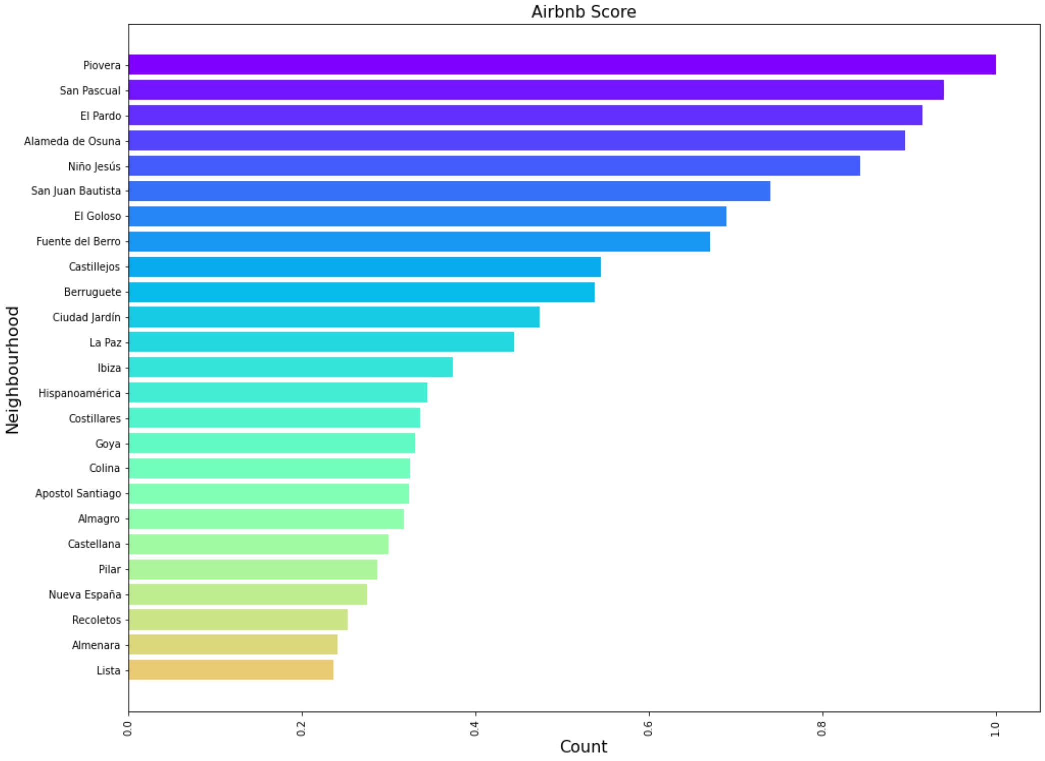
Analyzing the different clusters we can conclude that:

- Cluster 0, has the greater number of Airbnbs as well as the higher rent prices. Also coinciding is the part of the city center, which has lots of restaurants and shops.
- Cluster 1, has the second higher number of Airbnbs but with a medium price not too high, and as for venues has less business and more parks, that's because is the surrounding part of the center, which is mainly residencial.
- Cluster 2, being the north part of the center, is similar to Cluster 0 in the venues nearby, with lots of restaurants and business. But this one has less number of Airbnbs and a quite high rent price.
- And finally Cluster 3, has the lower number of Airbnbs and rent price. Also with less business and less interest points. There are residential zones.

With this analysis we conclude that Cluster 0 has the interest points more valued by the tenants, so we then analyse the different neighbourhoods in Cluster 0. Searching for those with higher rent prices and less Airbnb supply. We standardized the number of airbnbs with negative value, and summed the medium rent prices standardized, to get the final score. Then we sorted the neighbours by their score and plotted them. Having at the top Hellín, Lucero, Rosas, Arcos and Rejas.



It is also interesant to make the same analysis to Cluster 2, because it has similar venues characteristics as Cluster 0, but with less number of Airbnbs. So after the same procedure the top of the chart is for Piovera, San Pascual, El Pardo, Alameda de Osuna y Niño Jesús.



Discussion

The neighbourhoods at the top of the last charts, are some of the neighbourhoods in Madrid with more potential to invest in an Airbnb, taking into account their nearby business, medium rent price and the total number of Airbnbs. Before finally investing in some apartment you will have to investigate the zone inside the neighbourhood, the type of apartment and obviously the price of the floor. Having calculated the profitability that this investment can give you, you are ready to decide your investment. Although I leave this last part for further analysis.

Conclusion

In this project an attempt has been made to analyse the Madrid Airbnb market, making use of the Foursquare API and the Airbnb supply data. Using a K-mean clustering algorithm to classify the different neighbourhoods in similar groups based on the frequency of business in each neighbourhood. Finally merging the data we find out some neighbourhoods that can have investment potential.

Future possible research could help the system to make a more accurate analysis, including factors such as the price per m2, the zone of the neighbourhood, or the type of apartment, to conclude your investment.