Introduction:

Barcelona is one of the most important cities in Europe. With an expanding economy for the growth in Tourism. Holding a population of **1.6 million** and more than **12 million** tourists that visit this city annually. Being in the top 20 most visited in the world it is a great city to place a lucrative business.

With more than 7400 commerce distributed around Barcelona, in **10 Districts and 73 neighborhoods**, you come with two main questions as an investor. **Where and what should I invest?**

In this study a map will be created showing for each neighborhood or as called locally "Barrio", which are the **most likely best** venue missing in each borough. By giving it a score from 20 – 0, being 20 a venue that should be well received by the neighborhood and 0 a place that is already filled out by the zone.

The information used (total values: 369.663):

- Barcelona neighborhood Population (2018)
- Average rent price for m^2 (2019)
- Venues in Barcelona (2020)
- Location and coordinates for each neighborhood and limits

Data Description:

The population for each neighborhood was taken from the latest population study (2018) from the official Barcelona website. It was done by doing a web scrapping from the official site. (https://www.bcn.cat/)

The average m2 price of rent by neighborhood was taken from the latest update (2019) by the "Ajuntament de Barcelona" at the official web site. (https://www.bcn.cat/). This information is considered as a direct relation of an economy data.

The venue information was taken making an API call to FourSquare, obtaining the list of venues through the neighborhoods.

The location (Latitude and Longitude) for each neighborhood was retrieved using the GeoPy tool API.

The mapping coordinates for creating each neighborhood map was done using a GeoJson file found at (https://opendata-ajuntament.barcelona.cat/es)

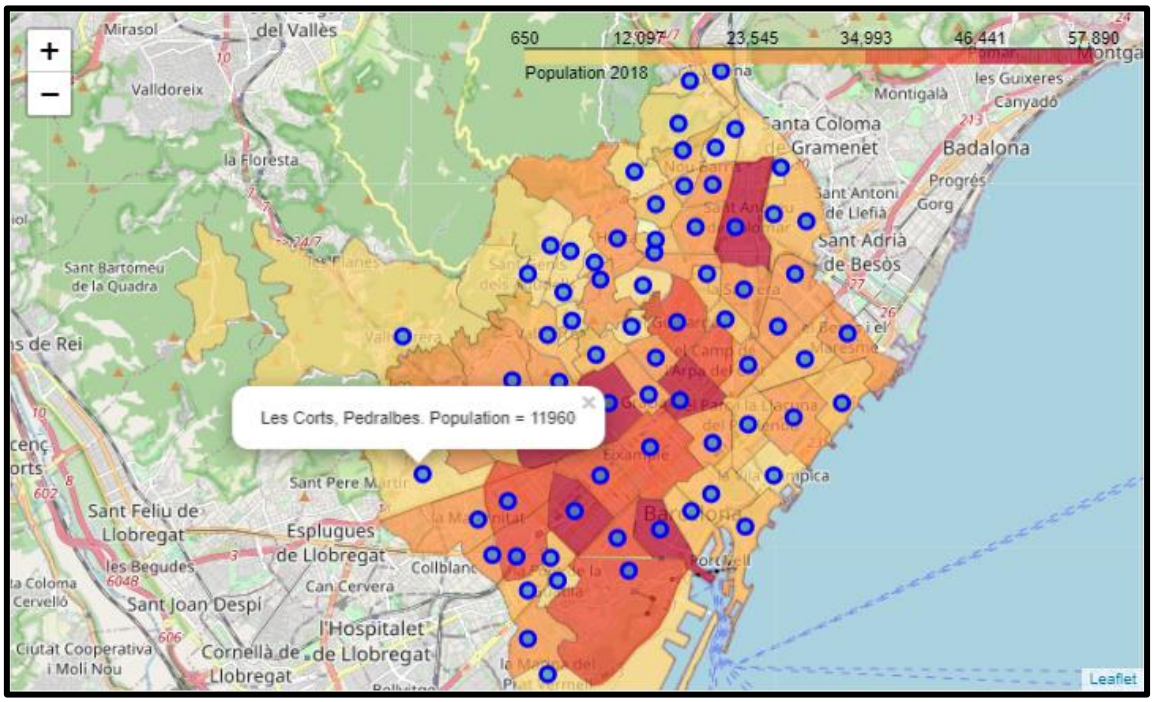
The result was 5 different datasets with a total of (369.663) values that were used for the study.

Methodology:

First the datasets of price by m2 and population needs to be cleaned and adjusted, then find the location (latitude and longitude) value for each neighborhood and join all of this into a combined Dataset:

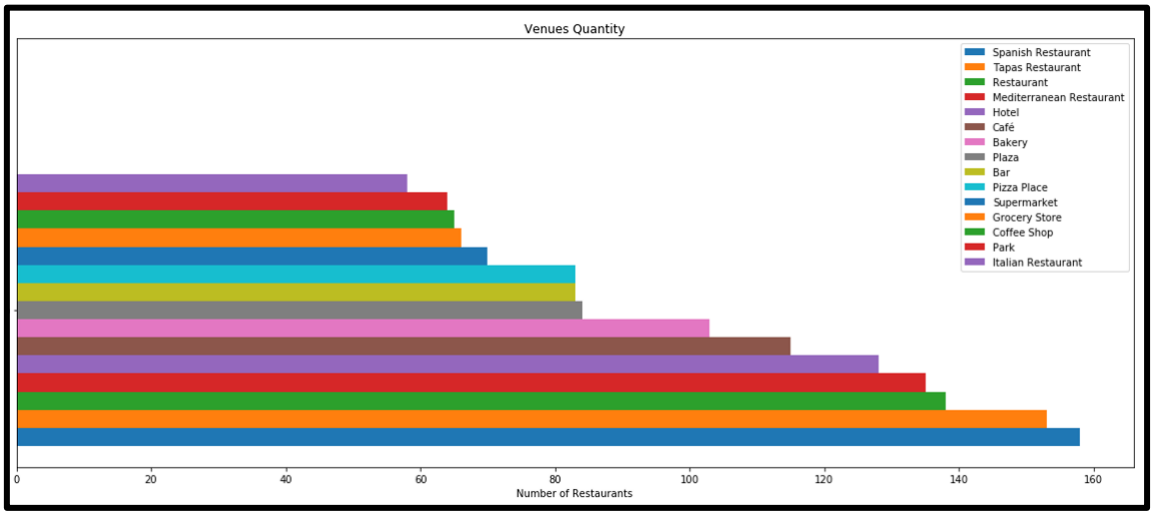
	Districts	Barrios	Population 2018	2019 Average m^2 price €	Latitude	Longitude
0	Ciutat Vella	el Raval	46948	14.004	41.380072	2.169613
1	Ciutat Vella	el Barri Gòtic	16731	14.003	41.383395	2.176912
2	Ciutat Vella	la Barceloneta	14739	17.005	41.380653	2.189927
3	Ciutat Vella	Sant Pere, Santa Caterina i la Ribera	22296	15.003	41.386697	2.181687
4	Eixample	el Fort Pienc	31922	13.009	41.395925	2.182325

Using the Folium and the Choropleth library, a geographical map and a superimpose choropleth or “heatmap” of the population was made. For each neighborhood, a popup icon was made with its information.



With the FourSquare API, and a range of a 600 meters radius for each zone center, the search for the venues was done with a total of **4589** venues found. With such a large range it was obvious that some venues were duplicates in the search. Working with the result and cleaning the repeated data, it ended with a total of **3197** different venues found in the city and 296 unique categories.

Manipulating the set, it can be found which are the most common venues in Barcelona:



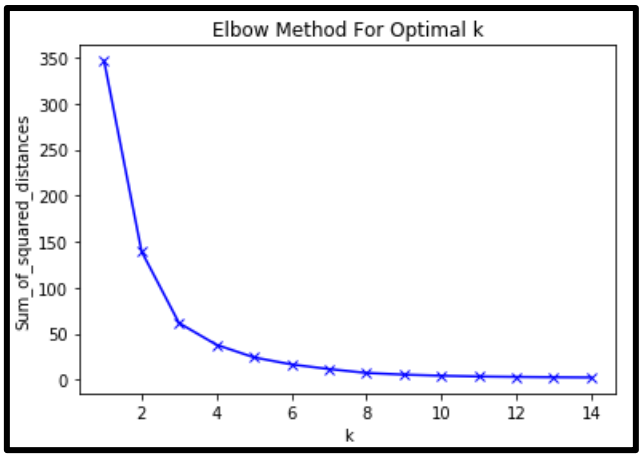
And the most common for each of the 73 neighborhoods:

Barrios	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
Baró de Viver	Breakfast Spot	Kids Store	Clothing Store	Perfume Shop	Asian Restaurant	Shoe Store
Can Baró	Spanish Restaurant	Soccer Field	Grocery Store	Pool	Bar	Bakery
Diagonal Mar i el Front Marítim del Poblenou	Beach Bar	Restaurant	Mediterranean Restaurant	Athletics & Sports	Hotel	Italian Restaurant

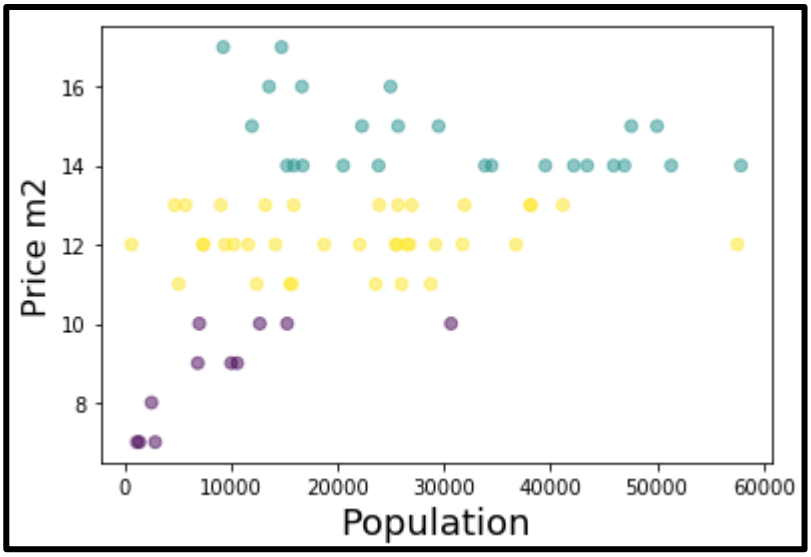
Having the venues, population and average rent price, the information can be handled to group the neighborhoods with a **K-meaning algorithm,** this is one of the most common method to handle unsupervised machine learning.

Doing some test mixing the datasets, it was found that the best clustering was made only using the venues and the avg m2 price.

It is done by first finding the best number of clusters for the Data, using the Elbow Method, the best k can be found.



The value 3 is the best k to use in this scenario, and the results can be checked by using a scatter plot with each cluster in a different color.



It can be said that the clusters are **High, medium, and low avg price for m2**, and with the following common venues:

	Groups	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	0	Restaurant	Café	Park	Spanish Restaurant	Grocery Store
1	1	Mediterranean Restaurant	Hotel	Spanish Restaurant	Restaurant	Tapas Restaurant
2	2	Tapas Restaurant	Spanish Restaurant	Restaurant	Café	Bakery

Procedure:

Now the data is gathered, its only missing the districts common venues, that is found using the same procedure.

With the Districts, Cluster, and Neighborhood top Venues it its possible to calculate the difference between the common venues by Districts and Cluster and common the existing venues.

Each data its defined by the 20 most common venues, giving a score between 20 to 1, being 20 to the most common venue and 1 to the 20th common venue.

Due to concentricity the District set is going to have most of the percentage score, and the cluster a smaller percentage. It will be done by 70% District and 30% cluster. This will be called "**Theorical score".**

The existing common venues are going to have a 100% of the score. Is going to be the **"Real score".**

The calculation would be done by:

$$\boldsymbol{Theorical\ Score - Real\ Score = Final\ Score}$$

For a better understanding here are two examples:

1) The 1st most common venue by district and 2nd in cluster are Italian Restaurants, and its 7th most common in the neighborhood.

$$\boldsymbol{Italian\ x\ 20\ x\ 70\% + Italian\ x\ 19\ x\ 30\% - Italian\ 14 = Italian\ 5.7}$$

The result it is a 5.7/20 this mean that an Italian Restaurant is roughly likely to be a good investment in the neighborhood.

2) The 3rd most common venue by district and 4th in cluster are Japanese Restaurants, and its 13th most common in the neighborhood.

$$\boldsymbol{Japanese\ \ x\ 18\ x\ 70\% + Japanese\ x\ 17\ x\ 30\% - Japanese\ 7 = Japanese\ 12}$$

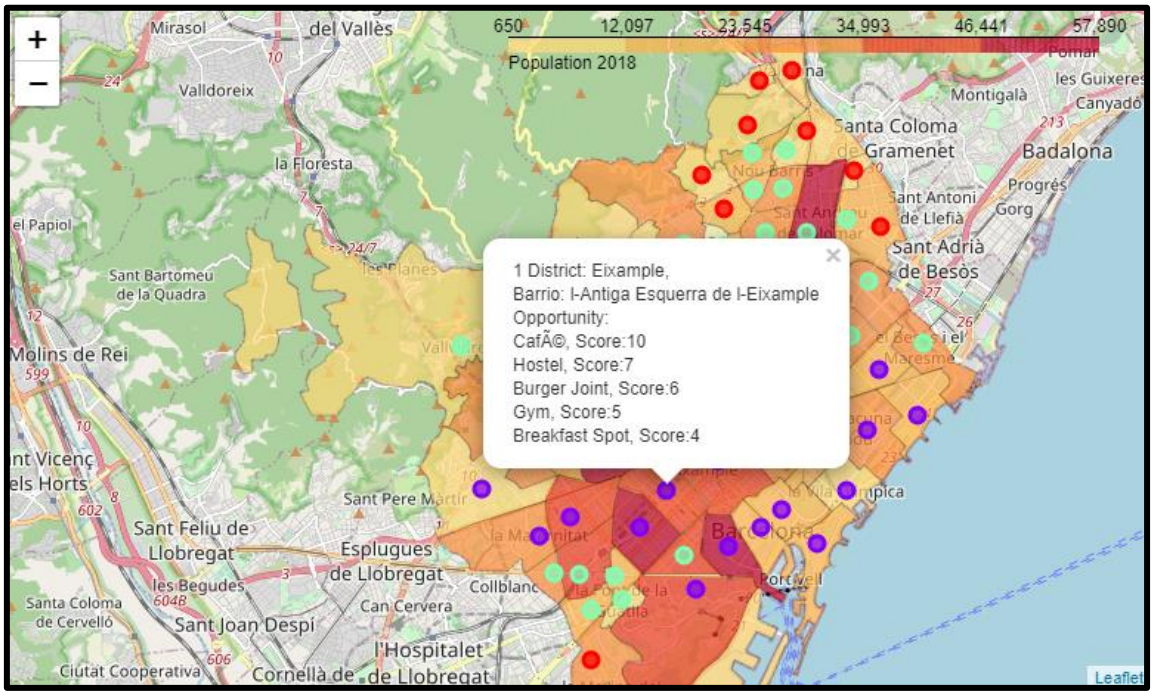
The result it is a 12/20 this mean that a Japanese Restaurant is most likely to be a good investment in the neighborhood.

Results:

At the end it returned a final Data Set, ordered as index (Districts, Groups, Barrios) and the rows with the 1st to the 20th position. In each value it contains the Venue and the final score.

Districts	Groups	Barrios	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Ciutat Vella	1	Sant Pere, Santa Caterina i la Ribera	Bar, Score:10	Ice Cream Shop, Score:9	Coffee Shop, Score:9	Italian Restaurant, Score:8	Burger Joint, Score:5
		el Barri Gòtic	Restaurant, Score:7	Breakfast Spot, Score:6	Beach, Score:4	Paella Restaurant, Score:3	Pizza Place, Score:2
		el Raval	Hotel, Score:12	Italian Restaurant, Score:8	Restaurant, Score:7	Breakfast Spot, Score:6	Wine Bar, Score:5

For a better use, this information was added to the Barcelona map, to the popup label from each neighborhood. If you click each dot, it will bring a label with the District, Barrio, and the first five venues with it scores.



Conclusion:

For the example in the last image shown the District “Eixample” in the neighborhood “L’Antiga Esquerra de L’Eixample”, the venue with most opportunity it is a Coffee Shop, Hostel and Burger Joint. Meaning these venues are general common along the district and the

segmentation done before, but not in this neighborhood. Being them the most likely venues to invest in this “Barrio”.

Although its an analytical study done with accurate and reliable information, its only a guideline to support any decision. Further analysis should be done before investing or preparing to make any economical action.

For better results more intensive study should be done, more information should be taken into consideration. Remember that only the foursquare database was used with **3197** venues, recalling there are more that 7400 restaurants, bars or coffees.

Important Districts Results:

Eixample:

Groups	Barrios	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
1	l'Antiga Esquerra de l'Eixample	Café, Score:10	Hostel, Score:7	Burger Joint, Score:6	Gym, Score:5	Breakfast Spot, Score:4	Mexican Restaurant, Score:2	Bar, Score:2	Dog Run, Score:0
	la Dreta de l'Eixample	Restaurant, Score:12	Coffee Shop, Score:10	Pizza Place, Score:8	Café, Score:7	Japanese Restaurant, Score:7	Gym, Score:5	Burger Joint, Score:5	Breakfast Spot, Score:4
	la Nova Esquerra de l'Eixample	Coffee Shop, Score:10	Pizza Place, Score:8	Hostel, Score:7	Burger Joint, Score:6	Restaurant, Score:5	Breakfast Spot, Score:4	Cocktail Bar, Score:4	Spanish Restaurant, Score:3
	la Sagrada Família	Mediterranean Restaurant, Score:11	Coffee Shop, Score:10	Hostel, Score:7	Japanese Restaurant, Score:5	Tapas Restaurant, Score:5	Gym, Score:5	Cocktail Bar, Score:4	Café, Score:2
2	Sant Antoni	Bakery, Score:9	Hostel, Score:7	Burger Joint, Score:5	Gym, Score:5	Cocktail Bar, Score:4	Sandwich Place, Score:3	Spanish Restaurant, Score:3	Flea Market, Score:0
	el Fort Pienc	Tapas Restaurant, Score:13	Pizza Place, Score:8	Japanese Restaurant, Score:7	Cocktail Bar, Score:4	Sandwich Place, Score:3	Mexican Restaurant, Score:2	Bar, Score:2	Beer Bar, Score:1

Les Corts:

Groups	Barrios	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Pedralbes	Restaurant, Score:11	Supermarket, Score:10	Hotel, Score:9	Pizza Place, Score:7	Bakery, Score:6
	la Maternitat i Sant Ramon	Café, Score:13	Garden, Score:11	Gym / Fitness Center, Score:7	Bar, Score:7	Restaurant, Score:6
	les Corts	Garden, Score:11	Soccer Stadium, Score:8	Gym / Fitness Center, Score:7	Coffee Shop, Score:5	Mediterranean Restaurant, Score:5

Ciutat Vella:

Groups	Barrios	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Sant Pere, Santa Caterina i la Ribera	Bar, Score:10	Ice Cream Shop, Score:9	Coffee Shop, Score:9	Italian Restaurant, Score:8	Burger Joint, Score:5
	el Barri Gòtic	Restaurant, Score:7	Breakfast Spot, Score:6	Beach, Score:4	Paella Restaurant, Score:3	Pizza Place, Score:2
	el Raval	Hotel, Score:12	Italian Restaurant, Score:8	Restaurant, Score:7	Breakfast Spot, Score:6	Wine Bar, Score:5
	la Barceloneta	Hotel, Score:12	Cocktail Bar, Score:11	Coffee Shop, Score:9	Italian Restaurant, Score:8	Breakfast Spot, Score:6

You can see the map, and all the code in this link:

https://nbviewer.jupyter.org/github/jorgequintanap/JQ-Files/blob/master/BarcelonaFinal.ipynb