# Business Problem:

*Barcelona* is a city on the coast of northeastern Spain. It is the capital and largest city of the autonomous community of Catalonia, as well as the second most populous municipality of Spain. With a population of 1.6 million within city limits, its urban area extends to numerous neighboring municipalities within the Province of Barcelona and is home to around 4.8 million people, making it the fifth most populous urban area in the European Union. It is one of the largest metropolises on the Mediterranean Sea, located on the coast between the mouths of the rivers Llobregat and Besòs, and bounded to the west by the Serra de Collserola mountain range, the tallest peak of which is 512 metres (1,680 feet) high.

Barcelona is a major cultural, economic, and financial centre in southwestern Europe, as well as the main biotech hub in Spain. As a leading world city, Barcelona's influence in global socio-economic affairs qualifies it for global city status (Beta +).

Barcelona is a transport hub, with the Port of Barcelona being one of Europe's principal seaports and busiest European passenger port, an international airport, Barcelona–El Prat Airport, which handles over 50 million passengers per year, an extensive motorway network, and a high-speed rail line with a link to France and the rest of Europe.

# Data Collection

To consider the above problem, the data is collected as following:

- I found the List of areas of Barcelona with its boroughs and houses rent prices.

https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-mercat-immobiliari-lloguer-mitja-mensual

- I used ForsquareAPI to get the most common venues of given Borough of Barcelona.

https://developer.foursquare.com/

- For choropleth maps I used. geojson file of Barcelona Districts / Barris.

https://github.com/martgnz/bcn-geodata

https://github.com/martgnz/bcn-geodata/blob/master/barris/barris.geojson

https://github.com/martgnz/bcn-geodata/blob/master/districtes/districtes.geojson

# Data Preprocessing

First of all, I removed all null values and then get rid of unwanted columns and only kept ‘Area’ and ‘Price’ columns.

Then, by using geocoder library I find the Longitudes and Latitudes of the Location and add a column of each in my DataFrame. I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 1000 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

Finally, by using the Foursquare API in conjunction with the created datasets, a table of most common visited venues in Barcelona neighborhoods is generated.

# Machine Learning

We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs.

K-Means algorithm is one of the most common cluster method of unsupervised learning. First, I run K-Means to cluster the boroughs into 6 clusters because when I analyze the K-Means with elbow method it ensured me the 6 degree for optimum k of the K-Means.

Then, I merged table with cluster labels for each borough. After examining each cluster, I labeled each cluster as follows:

1. Mixed Social Venues
2. Hotels and Social Venues
3. Stores and seafood restaurants
4. Pubs and Historic places
5. Sports and Athletics
6. Restaurants and Bars

After examining Average Prices, I label each price as follows:

- Less than 738: “Low level”
- 738–1009: “Average Level”
- 1009–1280: “Above Average level”
- 1280–1551: “High Level”
- More than 1551: “Very High level”

# Result

I came to the result that the house prices in the downtown and with Hotels and Social venues nearby are very high you can clearly visualize in the map above while in the suburbs and the neighborhoods away from the city center have low prices but the facilities are also good. Almost all low price neighborhoods are close to restaurants, pubs, sports facilities etc.

Some Boroughs such as Les Corts, Sarrià-Sant Gervasi and Sant Martí have very high house prices.

Sants-Montjuïc, Horta-Guinardó and Nou Barris Boroughs have very low house prices but have good venues to visit nearby.

# Conclusion

As people are turning to big cities to start a business or work. For this reason, people can easily interpret where to live with all facilities and cheaply. Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

# References


- https://en.wikipedia.org/wiki/Barcelona
- https://opendata-ajuntament.barcelona.cat/data/en/dataset/est-mercat-immobiliari-lloguer-mitja-mensual
- https://developer.foursquare.com/
- https://github.com/martgnz/bcn-geodata
- https://github.com/martgnz/bcn-geodata/blob/master/barris/barris.geojson
- https://github.com/martgnz/bcn-geodata/blob/master/districtes/districtes.geojson