# How can the public administration guide Barcelona city towards a more equal and pleasant city for it´s citizens?

**Fernando Paulino Nogueira**

**February 12, 2020**

## Introduction

We know that cities are not all the same. Their configuration is unequal, the number and types of venues, the public spaces for leisure activities, the criminality rates, the rental prices, between others, all of these aspects are different and in some part that represents a better or worse quality of live for the individuals and families as well as for the comunity as a whole.

If the public authority have up-to-date information about how the neighborhoods are configured and the metrics that are relevant in order to orient the public policies it can make the best decisions aiming to provide a better distributed city to all.

A more equally distributed city would improve the general quality of life. People wouldn´t need to go far away from their homes to work, to entertainment, to shop, to have family time, in short, to live. And that would reflect on smaller traffic jams, better air quality, healthier ways of transportation, growth of the local neighborhood economy, less time on the way and more time doing the things you like/need.

## Objectives

The main objective of this work is to provide the stakeholders with relevant information to decision making.

## Possible stakeholders

***Public Power***

- Explore the final result in order to see what are the neighborhoods with fewer number of public squares (green area) in order to plan the construction of new ones.
- See what are the predominant kind of venues by neighborhood in order to create orient which categpories should be subsidized and where.

***Civil Society Groups***

- Know what areas of the city are under real state speculation proccess in order to plan actions against it.

***Individuals***

- Analyse different aspects of the neighborhoods in order to choose the most interesting one for buying an appartment or invest in a business.

## Data

I´ll use 3 data sources:

***The foursquare API:***

The foursquare API will provide us with the neighborhoods´ venues´ names, their category and location.

Description:

The foursquare API will provide us with the neighborhoods´ venues, their category and location.

***Prices per square meter per neighborhood:***

https://www.bcn.cat/estadistica/castella/dades/timm/ipreus/habllo/a2020/tbarris2.htm

Description:

This table, provided by the statstics´ city hall website, have the monthly average prices per square meter. This will be our source in order to get the rental prices for each neighborhood in Barcelona city.

***The Annual Statistics Report for the city of Barcelona:***

http://www.bcn.cat/estadistica/catala/dades/anuari/Anuari2020_AAFF.pdf

Description:

This pdf is the Annual Statistics Report for the city of Barcelona. It encompasses many different information about the city. The information I´m after is on page 213, where we have the criminality rates for the boroughs of the city. The numbers I´ll be using are the "whole" numbers of criminality rates. That means that we won´t explode the numbers into the specific crimes, because, in my opinion, it would prejudice visualization and comprehensiveness of the work.

***GeoJson file containing the Borough´s boundaries:***

https://raw.githubusercontent.com/martgnz/bcn-geodata/master/districtes/districtes.geojson

Description:

We found this file in a repository in GitHub. It cointains all the necessary lat and long point marking all the boundaries of Barcelona´s boroughs.

## Methodology

The final work consists in 6 parts.

***Part 1:***

Creation of the base dataframe containing the following information: Borough, Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venues Names, Venues Latitudes, Venues Longitudes and Venues Categories.

Here I´ll use Pandas, bs4 and Request libraries.

***Part 2:***

Exploration and visualization of the information gathered so far. I´ll show the neighborhoods and the venues categories and their number of occurrences. I´ll also plot different graphics showing the amount of venues gathered by neighborhood, the prices per quare meter per neighborhood and also the amount of criminal complaints per neighborhood.

I´ll use Pandas, Seaborn and Matplotlib libraries.

***Part 3:***

Create new dataframe and preprare data for clustering, using one-hot encoding.

I´ll use Pandas library

***Part 4:***

Cluster neighborhoods using their venues categories similarity as the criterion. After some tests I have decided to use 4 clusters. That´s why because 5 presented too similar neighborhoods and 3 seemed to excessively condense them.

I´ll use Sklearn and Pandas modules.

***Part 5:***

Analysis after clustering. This is where i´ll try to describe each cluster based on the predominance of their venues.

I´ll make use of Pandas library.

***Part 6:***

Create final dataframes, create Barcelona´s map and plot all the relevant information on it.

I´ll make use of Pandas, Colormap and Folium libraries.

## Results

We have discovered what are the most expensive areas of the city, the most targeted for criminality, we also could learn what are the predominant kinds of venues in each neighborhood in order to understand a little better what are the dynamics of each one of them.

Regarding the clustering of he neighborhoods we came to this final result:

*Cluster 0* - "Diverse Neighborhoods" - Neighborhoods with more diverse kinds of venues. More oriented to locals than tourists.

*Cluster 1* - "Restauration Neighborhoods" - Strong presence of restaurants and bars/cafés.

*Cluster 2* - "Touristic Hosting Neighborhoods" - Very small area concentrating a lot of Hotels.

*Cluster 3* - "Amusement Neighborhoods" - Concentration of restaurants, bars and squares.

## Discussion

It´s important to notice that this work has limitations, amongst them we have:

- We don´t know the Foursquare API criterion to find and present the venues to us. In a number of neighborhoods it has reached the limit of 100 delimited by me. In order to deeply study the neighborhods the ideal would be to know every significant venue inside each neighborhood.

- The information we have on the criminal complaints are by borough and not nieghborhoods, it is from 2019 and it´s condensated for simplification reasons. It would be important to get it by neighborhood and to explore it in a expanded way.
- I raise a doubt if it wouldn´t be more accurate for the clustering algorithm if we separated the venues categories by "macro-categories" grouping some similar categories together.

## Conclusion

I believe the work achieved the result it aimed for. We could condensate all the information on just one map, showing in a clear and easy way crucial aspects of Barcelona city.

The information gathered could help policy makers, civil society, individuals and companies better understand the city and guide their decisions with metrics to compare and well grounded data.