APPLIED DATA SCIENCE CAPSTONE PROJECT:
Statistical solutions on finding optimal places to start a local business in Barcelona

1. Description of the problem and discussion of the background

Barcelona is one of the most densely populated Mediterranean cities which does not stop growing, both in tourism and in new business models. That is why it has become an excellent location to start a new commerce.

However, land expansion is a Barcelona tough problem, given by its topography situation and being surrounded by other big cities like Hospitalet de Llobregat. This circumstances increase the risk of opening a new store due the very competitive market and the high investment required.

For this reason, now it is becoming more important than ever knowing how Barcelona territorial distribution works, the differences between each district and neighborhood, how its population is distributed, and understanding the virtues that enhance every zone. Using all this information will make easier to find bigger opportunities and accomplish the greatest success when opening a new store.

This study expects to answer decisive questions in the private sector, as for example:

- Is my business model too crowned in a specific district? Which alternative districts would be a better option to open a particular store?
- In order to open a second store of the same kind, which would be the most similar districts to the main store district?
- Which neighborhood has the least number of restaurants per square meter? Which of these options has a bigger population density?

But there also solutions for the public sector:

- Analyzing the number of students per square meter, which district is more in need of a library?
- If we want to build a new nursing home, which neighborhoods have the largest amount of old man without one near? ¿And if we want to build a new children's park?

In conclusion, this study main objective is to support the decision of where to start a specific business, using all the information from the city and showing graphically the best options to have in mind.

2. Description of the data and how it will be used to solve the problem

- Barcelona population Dataset

Every statistical data from Barcelona city and its neighborhoods is extracted from the statistical department section of Barcelona local government webpage, as:

  - Population density (https://www.bcn.cat/estadistica/castella/dades/barris/terri/sup/sup417.htm)
  - Age distribution (https://www.bcn.cat/estadistica/castella/dades/barris/tpob/pad/ine/a2017/ine08.htm)

I decided to use 2017 datasets because it is the most actual year which was fully loaded in every section.

- Longitude and latitude from Barcelona neighborhoods and districts:

Unfortunately, I have been unable to find a current dataset which contains the longitude and latitude of all districts. I even contacted the statistical department of Barcelona local government to request this information, but I was told that they did not have this dataset.

So, understanding how important this data for this project is, I decided to collect manually this information for every single district using Google Maps application and save it in a CSV file that will be allocated at my GitHub repository.

- Venues from every Barcelona district

Additionally, through Foursquare API (https://developer.foursquare.com/) I will get every relevant venue from each district, processing and grouping them in order to record how many locals of each kind exist in each neighborhood.