A firm location decision process for a new opening

**Abstract**

This project consists in comparing and analyzing the metropolitan areas of the two biggest cities of Spain (Madrid and Barcelona) by exploring the places around the subway stations. The aim of this project is to find the best areas to extend a very successful gym brand from Madrid to Barcelona based on the neighborhood similarity. To perform the analysis, the unsupervised machine learning K-means method will be used for clustering the stations. The differences between clusters are based on the venue’s categories around the station area. The result of this study will help the gym company to decide the places that are suitable for the next opening.

# Introduction/Business problem

## Background

Barcelona is the capital and the largest city of Catalonia, which is in the northeastern of Spain in the Mediterranean Sea coast. Barcelona has a population of 1.6 million within the city limits and 5.2 million considering the metropolitan area, being the sixth most populous area in the European Union. The population density is 15.9 k inhabitants per square kilometer, but the city center has a population density of 40.3 k inhabitants per km2. Inhabitants between 25 and 44 years old, which are the main target, represent the 30.6 percent of the population. This makes the city a perfect place for business expansion.

The historical attractions, the sports, the good climate and the night entertainment are keeping the tourist income growing. It is the 4th preferred place to visit among tourists in Europe. Last year Barcelona hosted 9 million of tourists. Nevertheless, this is not the unique motor of the city, the best universities of Spain, industries and many international businesses are located here, making Barcelona attractive for investments.

On the other hand, Madrid is the capital of Spain and the biggest city situated in the strategic center of the country. The capital has a population of 3.3 million and 6.5 million considering metropolitan area. It is considered the major financial center and the leading economic hub of the Iberian Peninsula. The population density is 5 k inhabitants per square kilometer and the city center has a population density 26 k inhabitants per km2. And the middle age structure is 38% of the population. Apart from tourist impact, similar features could be found between two cities.

## Problem definition

A successful gym club chain from Madrid is planning to open a new gym in Barcelona. The locations of the gyms are usually placed in residential areas, near to downtown and close to a subway or train station. They are looking for locations in Barcelona with similar features to place the new gym. This study will cluster the subway and train stations of both cities, taking in account the kind of venues around. This will give us a base for similarities between the city’s locations. Finally, to concrete the best places, we will make a comparison the population density in the neighborhoods.

# Data

## Data requirements

* List of subway and train stations and coordinates within Madrid metropolitan area
* List of subway and train stations and coordinates within Barcelona metropolitan area
* List neighborhoods and its demographic data of Madrid and Barcelona.
* Information of the venues around the stations obtained from Foursquare. This will be the key to identify similar areas.
* List of the gyms near to the stations, obtained from Foursquare data

## Data sources

The list of all stations , names, type and coordinates within the metropolitan area of Barcelona con be easily download from the Open Data service of the city: <https://opendata-ajuntament.barcelona.cat/data/es/dataset/transports/resource/e07dec0d-4aeb-40f3-b987-e1f35e088ce2>

For stations data of Madrid metropolitan area there is also an open data service, but in this case the data is divided. The subway (metro) stations coordinates can be downloaded from here: [https://datos.madrid.es/egob/catalogo/200073-12-puntos-transporte-navegadores.gpx](https://datos.madrid.es/egob/catalogo/200073-12-puntos-transporte-navegadores.gpx'), and the train (cercanias) stations can be found here: [https://datos.madrid.es/egob/catalogo/200073-1-puntos-transporte-navegadores.gpx](https://datos.madrid.es/egob/catalogo/200073-1-puntos-transporte-navegadores.gpx'). The links may be temporally, so here there is the general link for all kind of transports: <https://datos.madrid.es/sites/v/index.jsp?vgnextoid=08055cde99be2410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

The administrative division of the Madrid neighborhoods can be downloaded as shape file from: <https://datos.madrid.es/egob/catalogo/200078-10-distritos-barrios.zip>

The administrative division of the Barcelona neighbourhoods can be downloaded as json here: <https://github.com/martgnz/bcn-geodata/tree/master/barris>

The demographic data of Barcelona neighborhoods can be scrapped from here: <https://www.bcn.cat/estadistica/castella/dades/inf/lecpadro/lec19/t13.htm>

The demographic data of Madrid neighborhoods can be downloaded with a tool in this page: <http://www-2.munimadrid.es/TSE6/control/seleccionDatosBarrio>