

Capstone Project - The Battle of Neighbourhoods

1. BUSINESS PROBLEM

One of the client/owner wants to open a branch of his restaurant in a new venue Madrid. Madrid is the capital and largest city of Spain, one of the busiest cities in Europe. More than 3 million people live in the city.

Statistic shows that annual number of international tourists visiting the Community of Madrid are increasing every year, on an average 800,000 visiting each month.

Client has some branches in European countries, after having successfully opened a branch in Valencia city in El Carme neighbourhood, he is planning to restaurant in Madrid.

Problem Statement :

Analyzing data by applying tools and methodologies using data science need to find out which is the neighbourhood from Madrid will be having similar characteristics like El Carme in Valencia city to open a new restaurant.

2. DATA

The data to be used for this project comes from different sites / locations:

- **Foursquare** is a social location service that allows users to explore the world around them, which provides information on different types of entertainment, drinking and dining venues.

The Foursquare API allows application developers to interact with the Foursquare platform, The API itself is a RESTful set of addresses to which you can send requests and find information related to the venues, such as location, overall category, reviews and tips.

- Madrid Neighborhood Names and geographic coordinates. Available on <https://datos.madrid.es/>, this is used to obtain the neighborhood location information from the city.

- Valencia City Neighborhood Names and geographic coordinates. Data available on <http://mapas.valencia.es/lanzadera/opendata/Barrios/SHAPE>
- Madrid census data, where we can get the population and income statistics, available in <http://www-2.munimadrid.es/CSE6/jsps/menuBancoDatos.jsp>

Below the details of how we will use each data source during this project.

2.1. Foursquare API data

For this project we will use the Foursquare Places API. One of the features of this API is to provide a list of venues within a specific location, based on the Lat/Lon coordinates and a radius.

In order to obtain a list of venues within a specified area, we use the “explore” endpoint from the API. By passing the proper parameters via an HTTP request to the *explore* endpoint, we get a JSON object.

The *location* object contains the coordinates of each venue, which will be used to associate it with its respective neighborhood.

The *categories* array will be used to categorize the neighborhood. Basically, we will count how many venues from all available categories are found on each neighborhood, and then use that information to compare neighborhoods from Madrid with El Carme in Valencia.

2.2. Madrid Neighborhoods

The Madrid city government has made available to the public a series of datasets with information of interest. We will be using the “Divisiones administrativas: distritos, barrios y divisiones históricas” dataset, available in the following URL: <https://datos.madrid.es/egob/catalogo/200078-10-distritos-barrios.zip>.

The data inside the .zip file is in ESRI format. To convert this to a dataframe that we can use, *geopandas* python library.

2.3. Valencia City Neighborhoods

Valencia City Neighborhood Names and geographic coordinates. Data available on <http://mapas.valencia.es/lanzadera/opendata/Barrios/SHAPE>.

This data is also available in ESRI format.

2.4. Madrid Census data

To complement our analysis we will be using the statistics of the population and average income per neighborhood in Madrid. This data is available in the municipality data bank, <http://www-2.munimadrid.es/CSE6/jsps/menuBancoDatos.jsp>