ANALYSIS OF RESIDENTIAL

NEIGHBORHOODS IN

MADRID, SPAIN

*Capstone Project: The battle of Neighborhoods*

****

*Madrid Downtown – Getty images*

Author: Raúl Castellanos

1. **INTRODUCTION**

As a part of Coursera IBM Data Science Professional Certificate Capstone Project, we were asked to elaborate an analysis of a city and its neighborhoods.

A common problem nowdays is the uprising prices of houses and cost of living. This is more problematic in the big cities as are the most crowded places so I find pretty interesenting for every family that wants to move to a new city to be able to know which are the best neighborhoods in their new destiny. As I am from Madrid, Spain I decided to elaborate this kind of research in this city and retrieve a list of its neighborhoods, find their geographical coordinates and use the coordinates as input of the Foursquare API, which we have used previously in the course, to obtain the top venue categories in each neighborhood. Using the frequency of venue categories we can use the k-means algorithm to cluster neighborhoods of similar venue categories and identify which ones are residential neighborhood, so a family can decide where to move to live in this city.

In addition to venue categories, I introduced data from the public schools of the city in each neighborhood in order to check if the information obtained from the cluster analysis is correct, and there is a great number of schools in those neighborhoods as it would be expected from a residential neighborhood.

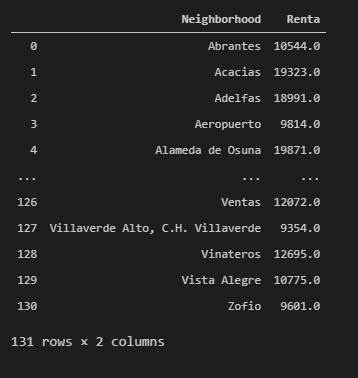
As a result we will be able to identify which are the most suitable parts of the city to live avoiding those neighborhoods with less facilities for a new family.

1. **DATA**

In this section I introduce the datasets that will be used and their sources.

***Neighborhoods information***

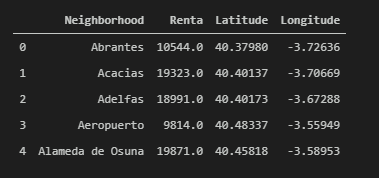
My initial dataset is a csv downloaded from the Madrid’s public wepage where we can find the name of every neighborhood and its Per Capita Income.



*Figure 1: Per Capita Income and neighborhoods name in Madrid.*

***Geographical coordinates***

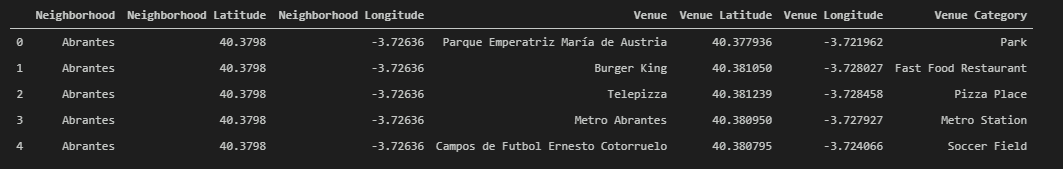
The information of every neighborhood will be enriched with geographical coordinates using the Geopy library from which I will add its latitude and longitude.



*Figure 2: Neighborhoods in Madrid with latitude and longitude information.*

***Venue categories***

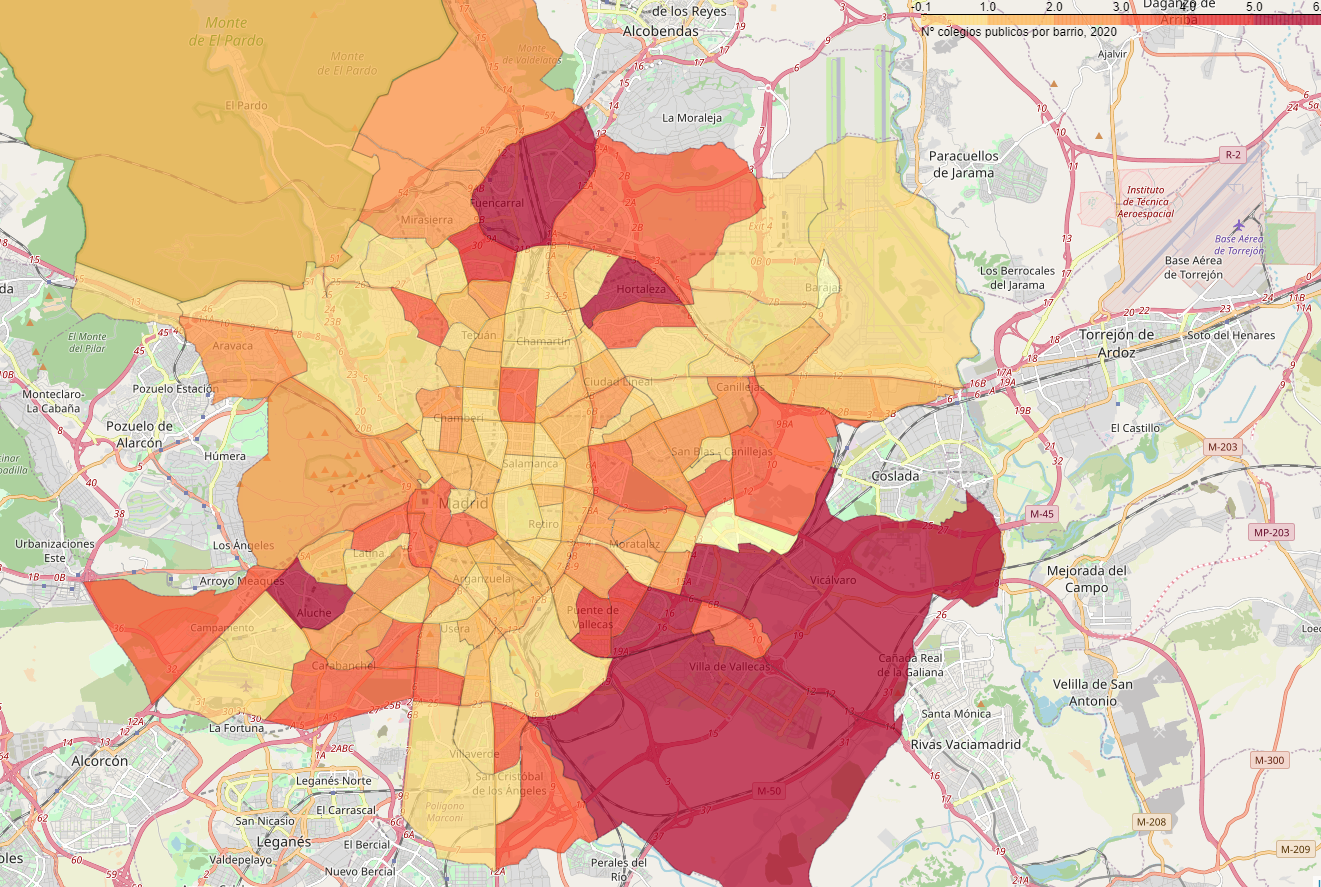
Next, I will use the Foursquare API, using the latitude and logitude included from the Geopy for every neighborhood to retrieve venues in a given radius around each location.



*Figure 3: Sample Venue Categories returne by Foursquare API per neighborhood.*

***Public schools***

Finally, information of amount of public schools in every neighborhood is added from the public dataset from https://datos.madrid.es/portal/site/egob/

**

*Figure 4: Number of schools per neighborhoods in Madrid.*