THE NEXT BAKERY

[CAPSTONE PROJECT]

## CARLOS GONZALEZ ROMERO

# INTRODUCTION / BUSINESS PROBLEM

### Background

The Levaduramadre brand is a gourmet bakery that is in full expansion in the city of Madrid, Spain. Although it was created in 2007, it is in 2017 that this expansion has been made. This expansion is happened mainly in the center and north-center of Madrid.

### Problem

To make the expansion in a homogeneus way, we will try to find an optimal location for a new bakery of the brand Levaduramadre in the south-center of Madrid. To do this, we will assume that the bakery owner gives us a choice of 3 specific locations, and we will have to justify our choice between those 3 options based on two criteria:

* How similar are the neighborhoods that currently have a Levaduramadre compared to the neighborhood of future Levaduramadre.
* How many bakeries there are around the future Levaduramadre and what its the distance between them.

### Interest

This report has been requested by the owner of other Levaduramadre interested in opening a new bakery in the south-center of Madrid.

# DATA

### Data sources

For this work we will use the data obtained from Foursquare API. Although it will be the same API, we will obtain three datasets, one for each aim:

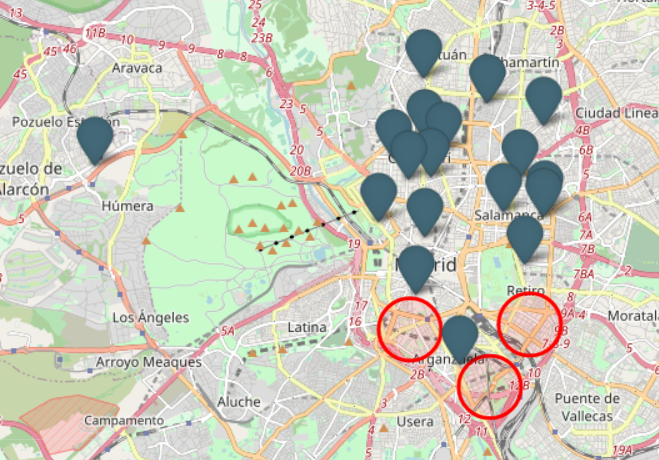
* **First dataset** will contain information about all Levaduramadre (existings and possibles future): This dataset will be used to obtain the second dataset.
* **Second dataset** will contain information about the venues around all Levaduramadre: This dataset will be used to create a cluster (using k-means clustering) of Levaduramadre’s locations to get the first criteria and help identify which should be the best location of next Levaduramadre.
* **Third dataset** will contain information about the bakeries around the possible locations: This dataset will be used to get the second criteria and help identify which should be the best location of next Levaduramadre.

### Data assumption

As said before, we assumed that Levaduramadre owner gives us 3 locations to choice the best option to open a new Levaduramadre.

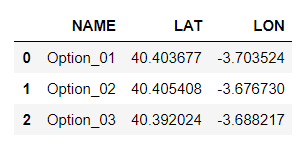
The locations of the possibles future Levaduramadre has been made by the following steps:

1. Search and paint on a map all existing Levaduramadre in Madrid. The search will be done with a call to Foursquare API, using the keyword “Levaduramadre” in the call, to obtain all venues named Levaduramadre in 20 km around Madrid center.
2. Select ([approximately](https://www.spanishdict.com/translate/approximately?langFrom=en)) three zones in the south-center of Madrid without Levaduramadre.



Picture 1.- Location of existings Levaduramadre and zones of possibles Levaduramadre.

1. Choose a random address in those zones, with the only criterion that it will be located on a street. Of the random address we get the latitude and longitude from google map.



Picture 2.- Dataframe with latitude and longitude of possibles future Levaduramadre

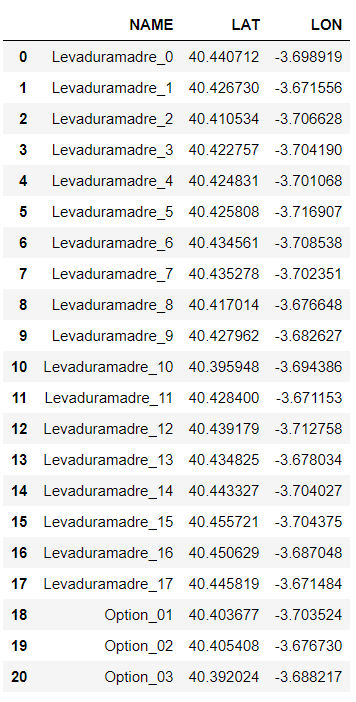
### Data cleaning

**First dataset**

To obtain the second dataset, I need clean the datas of the first dataset, because I obtain all venues named Levaduramadre, include the factory called “Obrador”, and other Levaduramadre that is not located in Madrid, so I deleted the venue with “Obrador” in its name, and filtered the rest by city equal to Madrid.

For easier identification of each venue, I will change the “name” each venue with the concatenated “Levaduramadre\_” plus a numerical value.

The last step for clean this first dataset will be get only the next information: name, latitude and longitude (also I will change the column name to match with the dataframe of possibles Levaduramadre). And then join with the dataframe of possibles future Levaduramadre.



Picture 3.- Example of first dataset

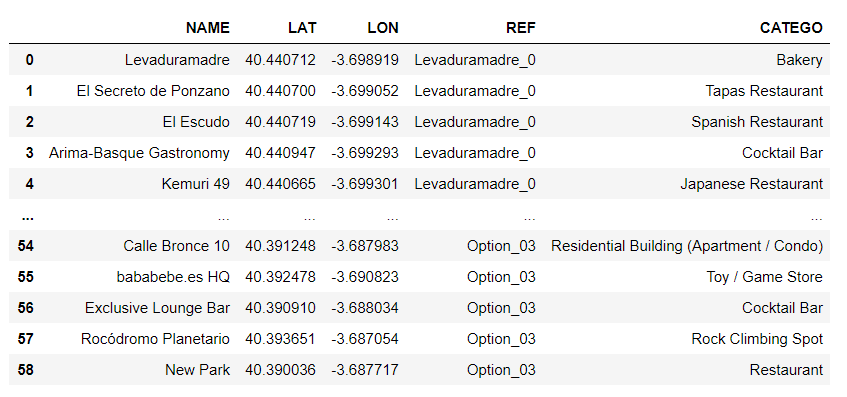
**Second dataset**

With the first dataset we can develop a search (with Foursquare API) and obtain the datas to make the second dataset. Obtaining those data will be made in the next 3 steps:

* Create an axiliar list.
* Make calls to Foursquare API and save the responses in auxiliar list generated: In this call we don´t use keyword, so we will obtein all venues in 200 meters around a point, in this case, each of venue in first dataset.
* Save the auxiliar list in a dataframe (named “df\_foursquare”).

**From the dataframe df\_foursquare** we get the useful information and create the second dataset. The useful information will be: name, latitude, longitude and the venue wich we use of reference in the search (in this case all Levaduramadre).

Then we add the category name to second dataset from the dataframe df\_foursquare.



Picture 4.- Example of second dataset

**Third dataset**

The process to create the third dataset is very similar to the second dataset, but we have to change the parameters of the call to Foursquare API, beacuse we will use a category ID (corresponding to Bakery) to make the calls. The response will be all bakeries around a point, in this case, the three possibles future Levaduramadre.

Once the call is made we proceed to obtain only the useful data: name, distance and the venue wich we use of reference in the search (in this case the possibles future Levaduramadre).



Picture 5.- Example of third dataset