

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 homogeneous 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.

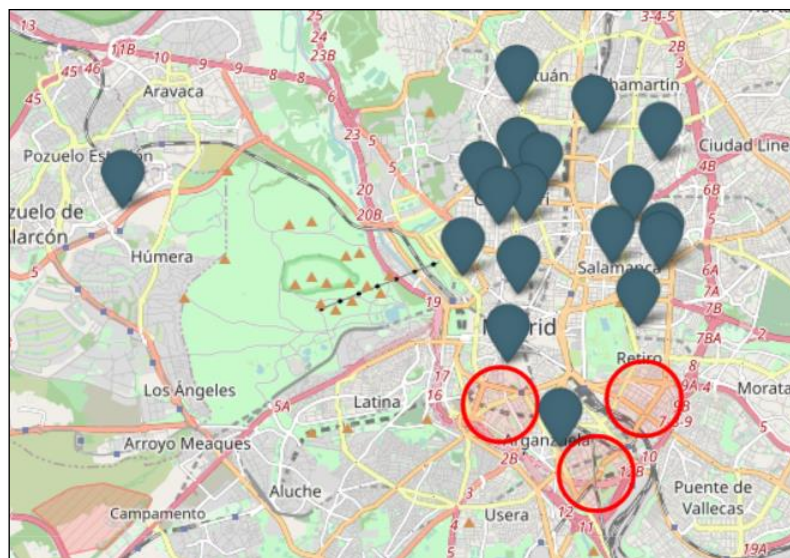
- **Third dataset** will contain information about the bakeries around the possible locations: This dataset will be used to get the second criteria and help identify which should be the best location of next Levaduramadre.

Data assumption

As said before, we assumed that Levaduramadre owner gives us 3 locations to choose 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) three zones in the south-center of Madrid without Levaduramadre.



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

3. 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.

	NAME	LAT	LON
0	Option_01	40.403677	-3.703524
1	Option_02	40.405408	-3.676730
2	Option_03	40.392024	-3.688217

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.

	NAME	LAT	LON
0	Levaduramadre_0	40.440712	-3.698919
1	Levaduramadre_1	40.426730	-3.671556
2	Levaduramadre_2	40.410534	-3.706628
3	Levaduramadre_3	40.422757	-3.704190
4	Levaduramadre_4	40.424831	-3.701068
5	Levaduramadre_5	40.425808	-3.716907
6	Levaduramadre_6	40.434561	-3.708538
7	Levaduramadre_7	40.435278	-3.702351
8	Levaduramadre_8	40.417014	-3.676648
9	Levaduramadre_9	40.427962	-3.682627
10	Levaduramadre_10	40.395948	-3.694386
11	Levaduramadre_11	40.428400	-3.671153
12	Levaduramadre_12	40.439179	-3.712758
13	Levaduramadre_13	40.434825	-3.678034
14	Levaduramadre_14	40.443327	-3.704027
15	Levaduramadre_15	40.455721	-3.704375
16	Levaduramadre_16	40.450629	-3.687048
17	Levaduramadre_17	40.445819	-3.671484
18	Option_01	40.403677	-3.703524
19	Option_02	40.405408	-3.676730
20	Option_03	40.392024	-3.688217

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 obtain 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.

	NAME	LAT	LON	REF	CATEGO
0	Levaduramadre	40.440712	-3.698919	Levaduramadre_0	Bakery
1	El Secreto de Ponzano	40.440700	-3.699052	Levaduramadre_0	Tapas Restaurant
2	El Escudo	40.440719	-3.699143	Levaduramadre_0	Spanish Restaurant
3	Arima-Basque Gastronomy	40.440947	-3.699293	Levaduramadre_0	Cocktail Bar
4	Kemuri 49	40.440665	-3.699301	Levaduramadre_0	Japanese Restaurant
...
54	Calle Bronce 10	40.391248	-3.687983	Option_03	Residential Building (Apartment / Condo)
55	bababebe.es HQ	40.392478	-3.690823	Option_03	Toy / Game Store
56	Exclusive Lounge Bar	40.390910	-3.688034	Option_03	Cocktail Bar
57	Rocódromo Planetario	40.393651	-3.687054	Option_03	Rock Climbing Spot
58	New Park	40.390036	-3.687717	Option_03	Restaurant

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).

	NAME	DIST	OPT
0	Granier	162.0	Option_01
1	Panaderia Rovier	44.0	Option_01
2	La Rinconada Pasteleria	154.0	Option_01
3	Ytalia Bakery & Coffee	261.0	Option_01
4	PANISHOP	127.0	Option_01
0	La Petite Sara	128.0	Option_02
1	VillaGarcia Pasteleria	49.0	Option_02
2	Panaderia San Miguel	146.0	Option_02
3	Horno San Miguel	124.0	Option_02
4	L'atelier Del Pan	151.0	Option_02
5	The Old Bakery	155.0	Option_02
6	Panaderia Pasteleria Los Canasteros	209.0	Option_02
7	Taberna De Panaderos	153.0	Option_02
8	Panaderia Trigal	278.0	Option_02

Picture 5.- Example of third dataset