

The sweet decision

The best location for a new bakery in Madrid

Jose J. Ruano

August, 2020

1. Introduction

1.1 Background

Spanish people love bakeries, they enjoy eating good bread, cupcakes or pies. A bakery is a good business and it is possible to open a new one by yourself or supported by a franchise.

Madrid has more than 20 boroughs and 100 neighborhoods, but most of the new neighborhoods are in **the north of the city**. For that reason I am going to focus in that area where a lot of young people with children live and where large numbers of apartments are being built.

1.2 Problem

The idea is to open a new bakery in *Madrid*, the capital of Spain.

Some neighborhoods in the north of Madrid are quite new and **there are no venues enough**: supermarkets, grocery stores or bakeries. Collecting the proper data, it will be possible **to determine the best location for a new bakery** or bakeries.

2. Data acquisition and cleaning

2.2. Data sources

To start the analysis it has been necessary to collect the information above:

- Madrid neighborhoods
 - Name and latitude and longitude coordinates were scraped from wikipedia.
 - Population of every neighborhood was found on the web of Madrid City Hall.
- Venues of every neighborhood
 - Foursquare API was used.

Links list:

- https://es.wikipedia.org/wiki/Anexo:Barrios_administrativos_de_Madrid
- <http://www-2.munimadrid.es/TSE6/control/seleccionDatosBarrio>

2.2. Data cleaning and wrangling

The **first step** was to scrape Wikipedia to extract the name of every neighborhood. I got a dataframe and I deleted some not important columns and translated the name of headers from spanish to english.

Then, it was necessary to collect all the links from the same web of every neighborhood in order to obtain the coordinates (latitude and longitude). After a small manual manipulation of the list with these links, I collected the coordinates and did a new dataframe. Then, I merged previous dataframes.

Finally, I created a dataset (csv file) with the population of every neighborhood from a website and then I imported this set and performed a new dataframe. The final issue was to merge this dataframe with the previous one (main dataframe).

The **second step** was to use Foursquare API to obtain the venues. Previously, I filtered the main dataframe and I selected only the neighborhoods in the north of Madrid. These neighborhoods belong to the following boroughs: “Fuencarral - El Pardo”, “Hortaleza” and “Barajas”. There are 19 neighborhoods.

According to the way that I worked in Labs (previous weeks), I used Foursquare to get the venues, check which neighborhood has a bakery and which does not. I also performed a map using the “Folium” library. This map has marks with the selected neighborhoods and a prominent mark in the neighborhoods with a bakery.

3. Methodology

3.1. Exploratory data analysis

The analysis has been focused on to find out how many bakeries there are in every neighborhood in the north of Madrid. In order to do that, it has been necessary to used Foursquare API and explore the result:

Spanish Restaurant	14
Restaurant	9
Coffee Shop	7
Fast Food Restaurant	6
Breakfast Spot	5
Hotel	5
Tapas Restaurant	5
Asian Restaurant	5
Italian Restaurant	5
Park	4
Sandwich Place	4
Bakery	4
Soccer Field	4
Grocery Store	3
Plaza	3
Supermarket	3

Figure 1. Most common venues in the neighborhoods

Then it was useful to locate every bakery in the map and also the neighborhoods using the “Folium” library. The points in blue are neighborhoods and the red ones are bakeries:

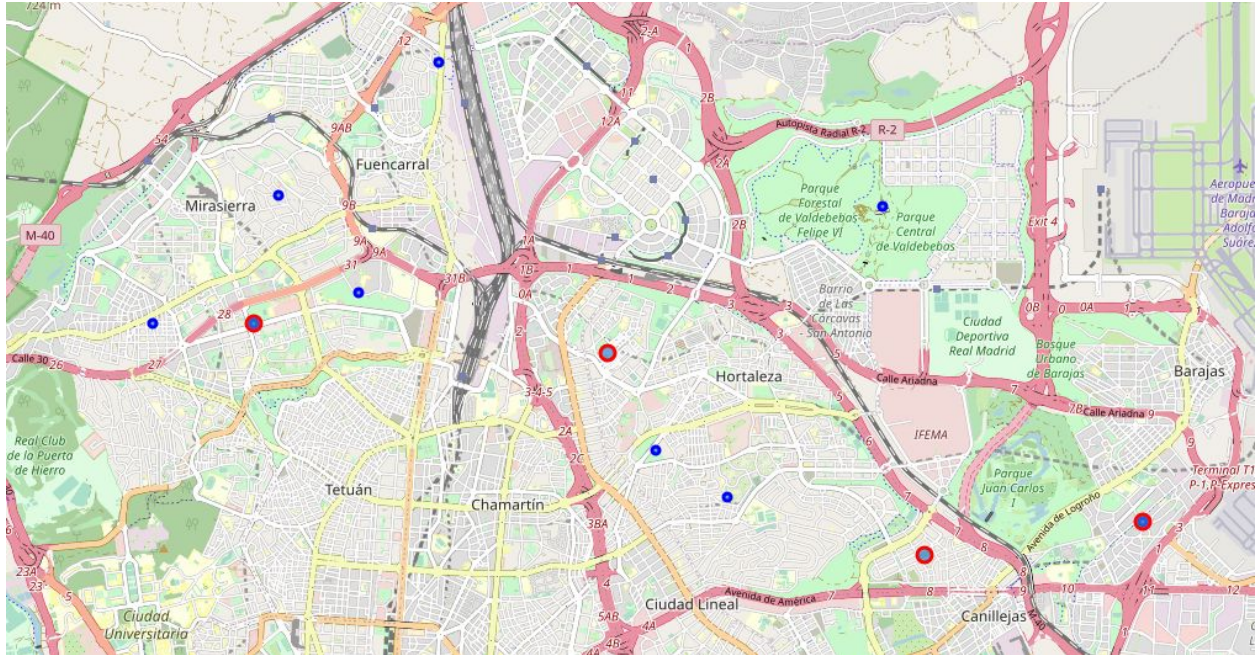


Figure 2. Map: bakeries in the north of Madrid

3.2. Machine Learning model: Clustering

In order to find some insights, Clustering model was used. The model involved 19 neighborhoods. The 4 neighborhoods which currently have a bakery were checked to find out a pattern:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	Cluster
1	Alameda de Osuna	Shop & Service	Smoke Shop	Metro Station	Italian Restaurant	Chinese Restaurant	Hotel	Pizza Place	0
2	Apóstol Santiago	Grocery Store	Metro Station	Bakery	Flea Market	Department Store	Diner	Donut Shop	0
7	El Pilar	Fast Food Restaurant	Tapas Restaurant	Burger Joint	Video Game Store	Pizza Place	Frozen Yogurt Shop	General Entertainment	0
11	Palomas	Coffee Shop	Restaurant	Café	Asian Restaurant	Spanish Restaurant	Sandwich Place	Food Truck	0

Figure 3. Neighborhoods with a bakery

The most seven common venues have been used to perform the clustering model.

4. Results

The result shows that **only 4 neighborhoods** in the north of Madrid **have a bakery**. These neighborhoods are far enough from each other.

About the clustering, **it has been difficult to find out some insights**. The model has created 3 different clusters but almost 75% of neighborhoods are in the same one. The main important pattern has been that the 4 neighborhoods which currently have a bakery belong to the same cluster.

5. Discussion

According to the data, the neighborhoods farther from the center of the city have less bakeries. This is a good opportunity to run a business like this, especially if it considers that the 3 most populated neighborhoods (“Valverde”, “Valdelasfuentes” and “Pinar del Rey”) do not have a bakery yet.

Another discussion could be the idea to open more than one new bakery. According to the result, it will be possible to create several new bakeries. From a conservative point of view, one to three new bakeries would be the recommendation.

6. Conclusion

This analysis clearly shows that a bakery could be an opportunity for someone who wants to run a food store in the north of Madrid. There are only a few bakeries in this area and it is known that bakeries are well valued stores for most people in Spain.

According to the data watched during this analysis, a similar one about supermarkets or other food stores could be an idea to take in account.