

The Battle of Neighborhoods - MADRID

Sergio García García

April 2020

1. Introduction

1.1 Background

The capital and economic center of Spain, Madrid is a very diverse city with a great culture. Full of charming corners with great restaurants and places of great interest such as the Prado museum or the Santiago Bernabeu stadium.

For new great chefs it is crucial to correctly select the area where to open a new restaurant. This analysis tries to segment and classify the best areas for the opening of a new restaurant.

1.2 Problem

The data produced after this analysis will allow us to segment the different neighborhoods of Madrid by classifying them. After carrying out this study, we will see which areas are best to open a Spanish food restaurant and which are the areas to open a tapas restaurant or hotel.

1.3 Interest

Those interested will be new entrepreneurs willing to open restaurants. Madrid is a great gastronomic city, therefore the location becomes crucial.

2. Data acquisition and cleaning

2.1 Data sources

To begin with the analysis, information has been obtained from the Madrid municipality. This information is the postal codes and the neighborhoods with their different sections.

The links used are as follows

Page:

<http://www.madrid.org/iestadis/fijas/clasificaciones/barrios.htm>

Excel:

<http://www.madrid.org/iestadis/fijas/clasificaciones/descarga/cobar18.xls>

The information obtained and put into table format is as follows:

	munic	distr	ldistr	barrio	descrip	secci
0	796	1	Centro	1	Palacio	1
1	796	1	Centro	1	Palacio	2
2	796	1	Centro	1	Palacio	3
3	796	1	Centro	1	Palacio	4
4	796	1	Centro	1	Palacio	6

2.2 Data cleaning

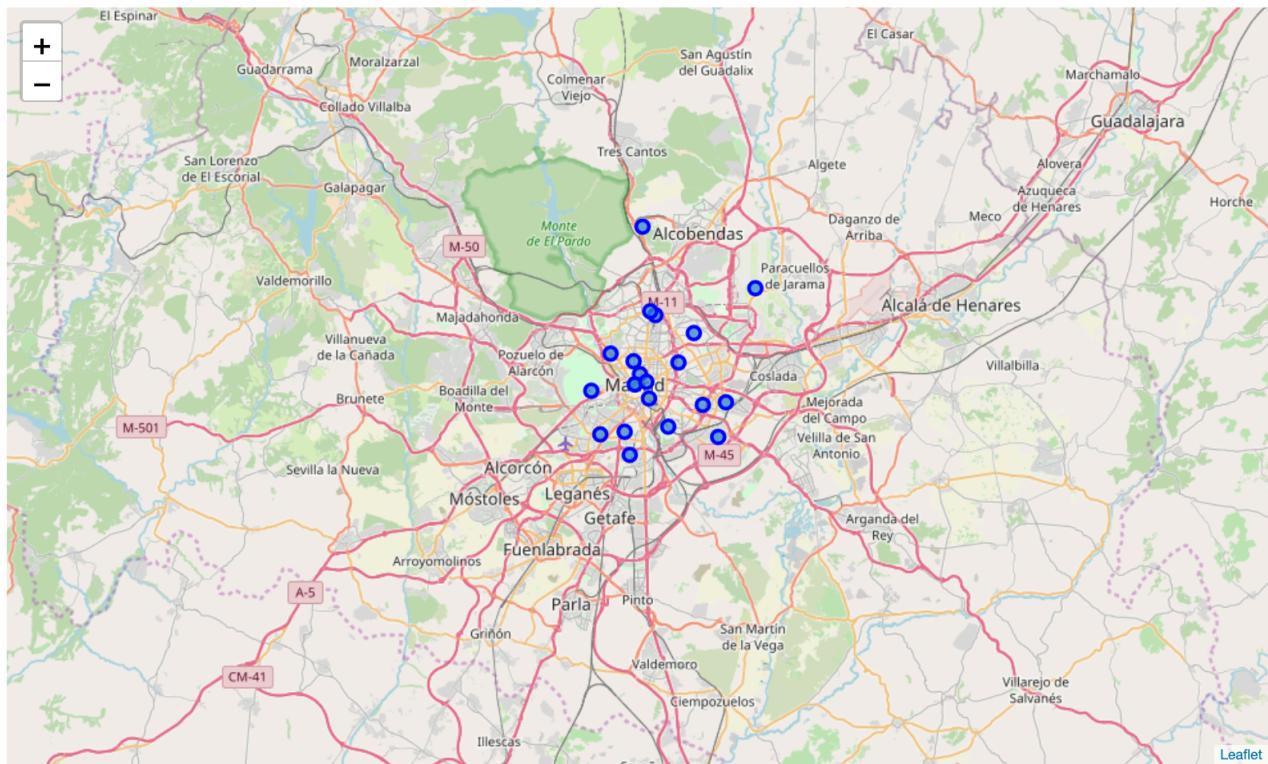
This information is cleaned and put into a suitable format for its exploitation. The first thing we do is concatenate several fields to obtain the postal code. Later we were left with the neighborhoods and borough. We group neighborhoods by borough in the same line, separate with ','.

	PostalCode	Borough	Neighborhood
0	7961	Centro	Palacio, Embajadores, Cortes, Justicia, Univer...
1	79610	Latina	Los Cármenes, Puerta del Angel, Lucero, Aluche...
2	79611	Carabanchel	Comillas, Opañel, San Isidro, Vista Alegre, Pu...
3	79612	Usera	Orcasitas, Orcasur, San Fermín, Almendrales, M...
4	79613	Puente de Vallecas	Entrevías, San Diego, Palomeras Bajas, Palomer...

Finally we obtain the latitude and longitude for each of them. This last step is performed in order to obtain information from the Foursquare API. Where we will get the most popular venues of each one

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	7961	Centro	Palacio, Embajadores, Cortes, Justicia, Univer...	40.425356	-3.698190
1	79610	Latina	Los Cármenes, Puerta del Angel, Lucero, Aluche...	40.411603	-3.749912
2	79611	Carabanchel	Comillas, Opañel, San Isidro, Vista Alegre, Pu...	40.375855	-3.740910
3	79612	Usera	Orcasitas, Orcasur, San Fermín, Almendrales, M...	40.377540	-3.715229
4	79613	Puente de Vallecas	Entrevías, San Diego, Palomeras Bajas, Palomer...	40.381633	-3.668024
5	79614	Moratalaz	Pavones, Horcajo, Marroquina, Media Legua, Fon...	40.400081	-3.631538
6	79615	Ciudad Lineal	Ventas, Pueblo Nuevo, Quintana, Concepción, Sa...	40.433980	-3.657251
7	79616	Hortaleza	Palomas, Piovera, Canillas, Pinar del Rey, Apo...	40.458139	-3.641003
8	79617	Villaverde	Villaverde alto, Casco Histórico de Villaverde...	40.358858	-3.708645

If we put it on a map it would be as follows:



3. Exploratory Data Analysis

Using the Foursquare API using longitude and latitude, we look at the venues. We have selected a total of 710 for the analysis.

Once obtained, we keep the top 5 for each area using percentages. This would look like this:

```
----Bellas Vistas, Cuatro Caminos, Castillejos, Almenara, Valdeacederas, Berruguete----  
venue freq  
0 Spanish Restaurant 0.17  
1 Sandwich Place 0.10  
2 Hotel 0.10  
3 Restaurant 0.07  
4 Italian Restaurant 0.07
```

Finally we select the 10 most common and establish them in a table:

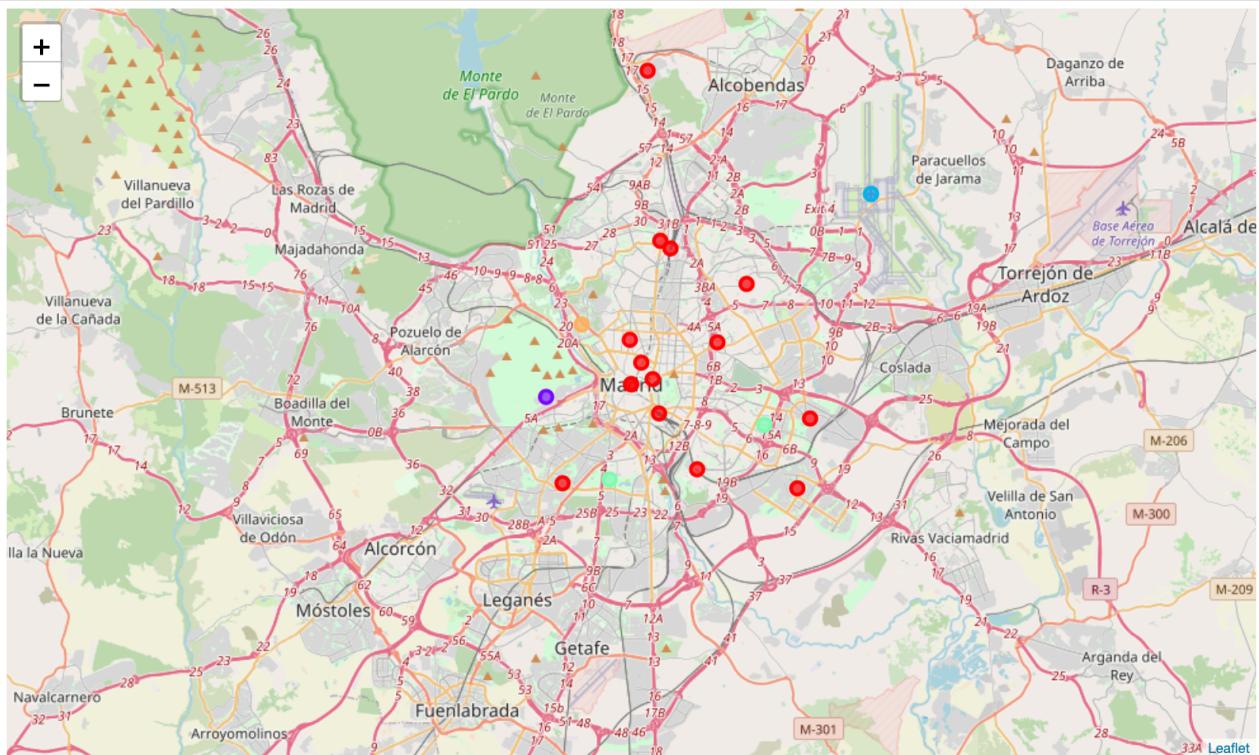
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alameda de Osuna, Aeropuerto, Casco Histórico ...	Apres Ski Bar	Wine Shop	Department Store	Flower Shop	Fast Food Restaurant	Falafel Restaurant	Event Space	Electronics Store	Diner	Dessert Shop
1	Bellas Vistas, Cuatro Caminos, Castillejos, Al...	Spanish Restaurant	Hotel	Sandwich Place	Restaurant	Breakfast Spot	Italian Restaurant	Business Service	Fast Food Restaurant	Building	Burger Joint
2	Casa de Campo, Argüelles, Ciudad Universitaria...	Restaurant	Museum	Metro Station	Bus Station	Flower Shop	Falafel Restaurant	Event Space	Electronics Store	Diner	Dessert Shop
3	Casco Histórico de Vallecas, Santa Eugenia, En...	Tapas Restaurant	Pet Store	Soccer Field	Grocery Store	Restaurant	Asian Restaurant	Supermarket	Wine Shop	Electronics Store	Diner
	Casco Histórico										

4. Predictive Modeling

A cluster technique has been applied to perform the segmentation. For this, the number of the cluster has been established as 5.

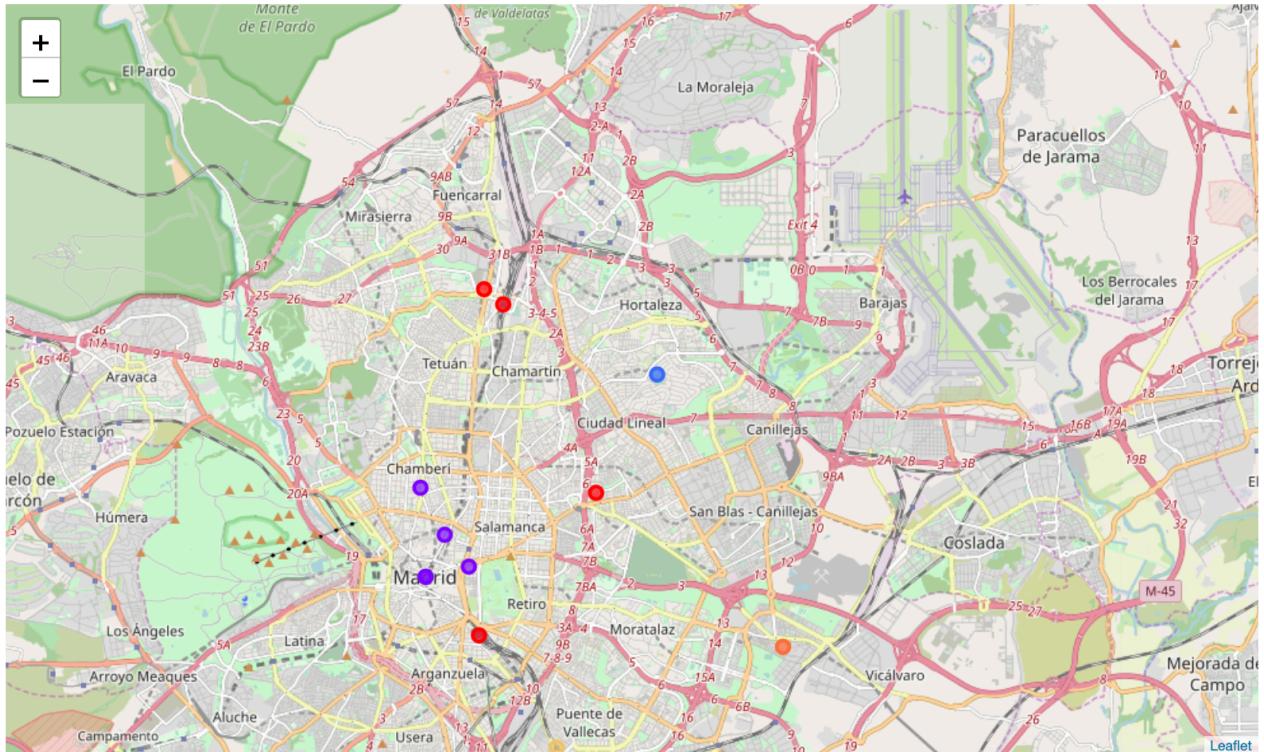
With this we obtain the most prone areas to establish a restaurant.

The red dots mark these areas. As we can see Madrid is a very gastronomic area, so this analysis does not give us much information



But our analysis wants to go further. Therefore, within cluster number 1, the most prone cluster, a re-segmentation has been carried out to further adjust the most appropriate areas for each type of restaurant.

For this reason, a clustering technique has been applied to better understand the areas. The cluster number has been set to 8.



With this new analysis we are able to find the most prone areas for each type of restaurant.

In the case of red dots as we can see in the following table. They are the most suitable areas for a Spanish food restaurant:

	Borough	Cluster Labels	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Ciudad Lineal	0	Spanish Restaurant	Grocery Store	Restaurant	Chinese Restaurant	Mediterranean Restaurant	Tapas Restaurant	Hotel	Gourmet Shop	Park	Butcher
11	Arganzuela	0	Spanish Restaurant	Hotel	Restaurant	Grocery Store	Sandwich Place	Train Station	Museum	Gym	Brewery	Latin American Restaurant
16	Chamartin	0	Spanish Restaurant	Café	Platform	Hotel	Gym / Fitness Center	Restaurant	Bar	Sandwich Place	Train Station	Skating Rink
17	Tetuan	0	Spanish Restaurant	Hotel	Sandwich Place	Restaurant	Breakfast Spot	Italian Restaurant	Burger Joint	Building	Business Service	Fast Food Restaurant

In the case of purple spots as we can see they are more prone to hotels or informal restaurants:

	Borough	Cluster Labels	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Centro	1	Restaurant	Spanish Restaurant	Bakery	Hotel	Cocktail Bar	Cosmetics Shop	Café	Vegetarian / Vegan Restaurant	Bar	Italian Restaurant
12	San Blas	1	Hotel	Plaza	Tapas Restaurant	Spanish Restaurant	Hostel	Wine Bar	Gourmet Shop	Clothing Store	Restaurant	Dessert Shop
14	Retiro	1	Hotel	Plaza	Tapas Restaurant	Spanish Restaurant	Hostel	Wine Bar	Gourmet Shop	Clothing Store	Restaurant	Dessert Shop
15	Salamanca	1	Hotel	Spanish Restaurant	Plaza	Restaurant	Theater	Park	Café	Japanese Restaurant	Art Gallery	BBQ Joint
18	Chamberí	1	Restaurant	Tapas Restaurant	Theater	Bar	Spanish Restaurant	Café	Bakery	Multiplex	Plaza	Mexican Restaurant

5. Conclusions

This analysis has allowed us to know which sites are more suitable for the opening of a new restaurant with only the use of data from the community of Madrid and the help of the Foursquare API. This allows us to make better decisions based on real data.

This will make our new open business more likely to succeed. What improves the conditions for people who have and use this analysis.

6. Future directions

As future steps we should try to expand the analysis including new variables and new sources of information. Since in this case we have based ourselves on the advantages of Foursquare. But important data such as the average income in the area or the saturation level of restaurants could be very rich analyzes for decision-making when opening a new restauran