CLUSTERING NEIGHBORHOODS IN MADRID

LUIS MUÑOZ DEL POZO

TABLE OF CONTENTS

1. INTRODUCTION – BUSINESS PROBLEM

The objective of this project is to help investors to find the best area for opening a restaurant in Madrid, Spain.

Madrid is the largest Spanish city with more than 3 million inhabitants in a metropolitan area of approximately 600 square kilometers, and it is indeed a very good place to open a restaurant, thanks to its busy business centers, and lively streets either by day or night.

Madrid is administratively divided in 21 districts ("distritos" in Spanish), and 131 wards ("barrios" in Spanish).

Thanks to the data from several Madrid City Hall webpages and Foursquare, and an appropriate data science approach, at the end of the project stakeholders will have the proper information to select the best wards and the best kind of restaurant to invest in.

2. DATA

Based on the objective of the business problem, the following data will be retrieved:

- Information of the administrative distribution in district and wards. This information will be obtained from the Madrid City Hall official website

- Information of all of the district such as population, size (area), average income, number of crimes, or number of hotel beds.

- Information of the coordinates of every ward by geocoding

- Information about the type and location of restaurants from Foursquare API

References of the places to find the information is summarized in section 7.

3. METHODOLOGY

In order to accomplish this objective, three main phases will be considered:

1. Selecting the best "distritos" to open a restaurant. This selection will be done by clustering the "distritos based on its size, population, average income, number of crimes per year, and number of hotel beds (as a good indicator of tourism), .

2. Selecting the best "barrios" to open a restaurant within the best "distritos" selected in phase 1. This selection will be done by clustering the "barrios" based on the number of restaurants and type.

3. Selecting the best three types of restaurant to invest in, considering the clusters created in phase 2, based on the less common kinds of restaurants of each cluster.

Phases 1 and 2 will be developed by clustering the "distritos" and "barrios" respectively, using the k-mean method (as it is an unlabeled problem), optimizing the number of clusters using the elbow method.

4. RESULTS

In this section the main results are presented, following the steps described in section 3.

4.1. Clustering Districts by demographic data

As the information of income, population, hotel beds, etc. was found for districts, not wards, the first phase of the study was to cluster districts based on this info. K-mean method was chosen and the elbow method the way to pick the optimum k, as it can be shown in the following picture.
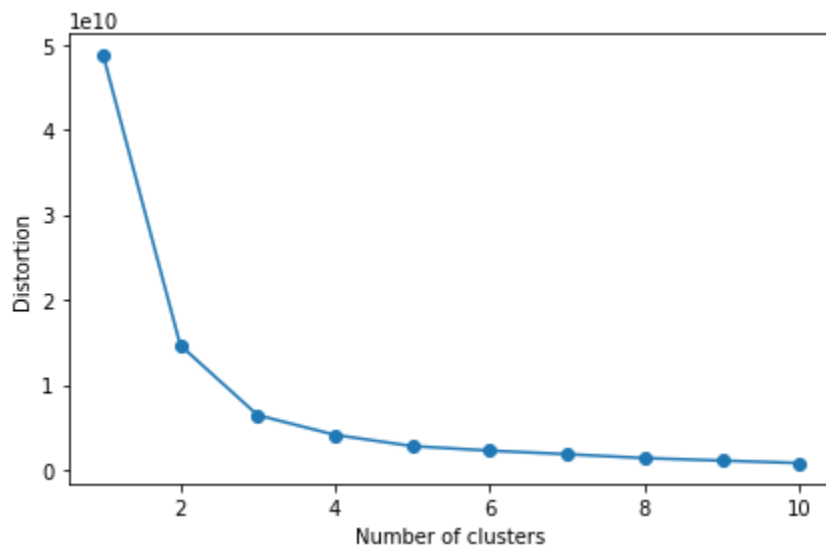

Figure 1. Elbow method in clustering Madrid districts.

Clustering with k = 3, the following results were obtained.

Distritos_clust	Renta (Euro/hab)	Pob(16_79)	Superf. (Ha)	Plazas Hotel	Delitos
0	14426.666667	61976.000000	2776.423333	1552.666667	215.333333
1	14091.166667	197749.333333	5518.800000	1335.000000	660.000000
2	17455.000000	120235.083333	1583.620000	5658.583333	796.083333

Table 1. Mean values for the district´s clusters chosen (income, population between 16 and 79 years, area in 0.01 km^2, hotel beds and number of crimes).

Cluster number 2 is chosen, because it is the highest income per inhabitant ("Renta" in Spanish), and it is the one with more hotel beds ("Plazas Hotel" in Spanish), so it is a good place to invest in a restaurant because the population has money and there are a lot of tourists.

Merging the information from Cluster number 2 and the name of the wards of the chosen districts, a list of wards was elaborated, considering 55 wards (of a total of 131).

Considering this information and "geocode" application, the location coordinates of each ward was retrieved, and the following map was created.

Map 1. Cluster 2 Madrid districts

4.2. Clustering Wards by venues

For clustering wards by venues, Foursquare is used to retrieve information in a radius of 500m of each ward (500 m is considered a good estimation since most of the Madrid wards are approximately 1 km^2).

One-hot method was used to accommodate the tables for using the k-mean method. The k optimum value was selected using the elbow method, as it can be shown in the graph below:
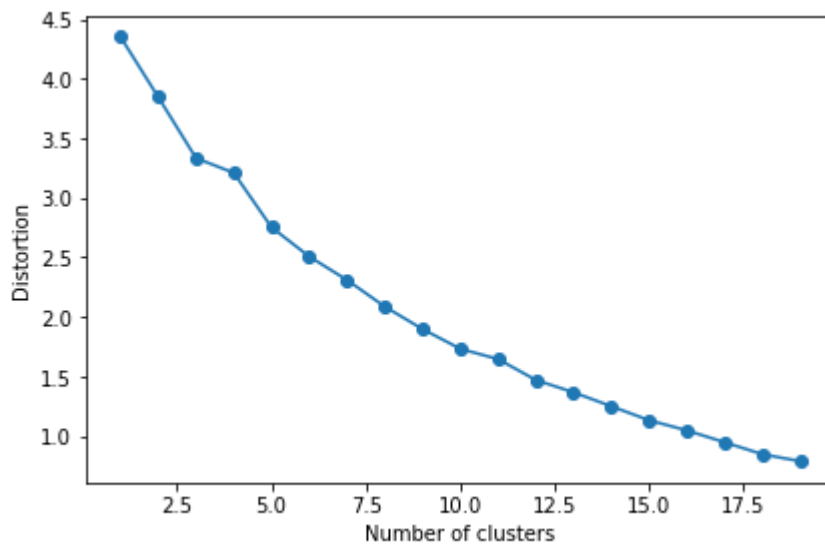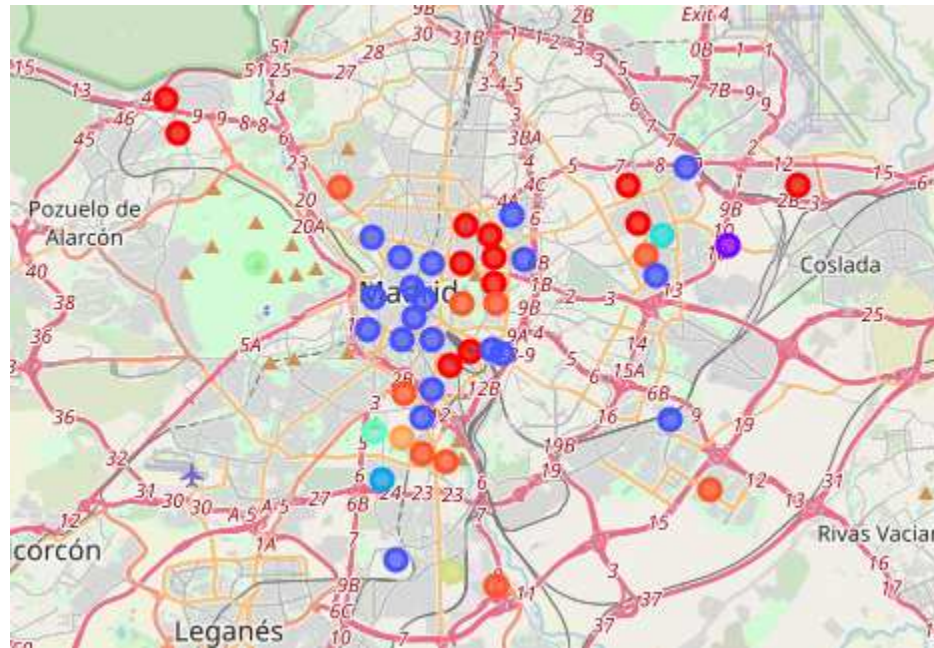

Figure 2. Elbow method in clustering Madrid wards.

In this case, it is not very clear which the best k value for clustering is. Based on the distortion values it was decided that k = 10 should be a good enough number of clusters.

Considering k = 10, the following map was created to show the final result.

Map 2. Clustering Madrid wards for districts in Cluster 2.

5. DISCUSSION

In this section all clusters are examined in order to conclude which one is the best for investing in a restaurant in Madrid.

5.1. Cluster 1

In the following table, it can be seen the wards included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Delicias	Coffee Shop	Chinese Restaurant	Restaurant	Mediterranean Restaurant	Bakery
12	Atocha	Restaurant	Bar	Spanish Restaurant	Italian Restaurant	Bakery
16	Ibiza	Spanish Restaurant	Tapas Restaurant	Italian Restaurant	Restaurant	Seafood Restaurant
19	Recoletos	Restaurant	Spanish Restaurant	Italian Restaurant	Japanese Restaurant	Hotel
20	Goya	Spanish Restaurant	Restaurant	Bakery	Tapas Restaurant	Seafood Restaurant
23	Lista	Spanish Restaurant	Restaurant	Seafood Restaurant	Hotel	Tapas Restaurant
24	Castellana	Spanish Restaurant	Restaurant	Coffee Shop	Boutique	Tapas Restaurant
29	Valdemarin	Restaurant	Spanish Restaurant	Asian Restaurant	Burger Joint	Breakfast Spot
31	Aravaca	Restaurant	Deli / Bodega	Soccer Field	Pizza Place	Pharmacy
47	Simancas	Spanish Restaurant	Restaurant	Café	Hotel	Mediterranean Restaurant
52	Rejas	Breakfast Spot	Café	Indie Movie Theater	Restaurant	Hostel
54	Salvador	Mediterranean Restaurant	Restaurant	Spanish Restaurant	Italian Restaurant	Asian Restaurant

Table 2. Madrid ward cluster 1 – 5th most common venues

It is clear that this cluster is full of restaurants (especially Spanish ones) and hotels, so it could be a good place to open a new restaurant different than Spanish.

5.2.   Cluster 2

In the following table, it can be seen the ward included and the 5th most common venues.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
51	Rosas	Gastropub	Electronics Store	Financial or Legal Service	Garden	Furniture / Home Store

Table 3. Madrid ward cluster 2 – 5th most common venues

This cluster seems to be a kind of suburb ward, so it does not seem to be a good place to open a new restaurant.

5.3. Cluster 3

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Palacio	Tapas Restaurant	Plaza	Spanish Restaurant	Historic Site	Bar
1	Embajadores	Bar	Café	Tapas Restaurant	Hotel	Art Gallery
2	Cortes	Spanish Restaurant	Hotel	Restaurant	Bar	Plaza
3	Justicia	Spanish Restaurant	Restaurant	Bakery	Gay Bar	Hotel
4	Universidad	Café	Cocktail Bar	Bookstore	Tapas Restaurant	Argentinian Restaurant
5	Sol	Spanish Restaurant	Hotel	Tapas Restaurant	Clothing Store	Restaurant
6	Imperial	Spanish Restaurant	Hotel	Coffee Shop	Café	Mediterranean Restaurant
7	Acacias	Bar	Pizza Place	Tapas Restaurant	Spanish Restaurant	Art Gallery
9	Legazpi	Spanish Restaurant	Bar	Plaza	Grocery Store	BBQ Joint
11	Palos de Moguer	Spanish Restaurant	Restaurant	Beer Garden	Breakfast Spot	Brewery
13	Pacífico	Spanish Restaurant	Bar	Food & Drink Shop	Bakery	Grocery Store
14	Adelfas	Spanish Restaurant	Bar	Tapas Restaurant	Fast Food Restaurant	Hotel
21	Fuente del Berro	Spanish Restaurant	Gym / Fitness Center	Bar	Bakery	Supermarket
22	La Guindalera	Hotel	Restaurant	Seafood Restaurant	Japanese Restaurant	Spanish Restaurant
26	Argüelles	Spanish Restaurant	Bar	Tapas Restaurant	Pub	Sandwich Place
35	Almendrales	Spanish Restaurant	Seafood Restaurant	BBQ Joint	Nightclub	Noodle House
39	Villaverde Alto	Train	Playground	Pizza Place	Spanish Restaurant	Train Station
45	Santa Eugenia	Spanish Restaurant	Farmers Market	Gym	Tapas Restaurant	Supermarket
50	Arcos	Burger Joint	Lottery Retailer	Soccer Field	Multiplex	Shopping Mall
53	Canillejas	Hotel	Plaza	Breakfast Spot	Café	Smoke Shop

Table 4. Madrid ward cluster 3 – 5th most common venues

This cluster is full of restaurants, as in cluster 1, but no so many hotels. It could be a good place to open a new restaurant, but it seems to be more a Spanish wards than a tourist wards. This distinction between cluster 1 and 3 is important in order to choose properly the kind of restaurant to be opened.

5.4. Cluster 4

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
32	Orcasitas	Plaza	Café	Grocery Store	Sporting Goods Shop	Yoga Studio

Table 5. Madrid ward cluster 2 – 5th most common venues

This cluster does not seem to be a good place to open a new restaurant, since it is more a residential area.

5.5. Cluster 5

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
48	Hellín	Music Venue	Snack Place	Pharmacy	Gym	Park

Table 6. Madrid ward cluster 2 – 5th most common venues

The same as for Cluster 4, this cluster does not seem to be a good place to open a new restaurant, since it is more a residential area.

5.6.   Cluster 6

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
37	Zofio	Spanish Restaurant	Beer Garden	Athletics & Sports	Park	Flea Market

Table 7. Madrid ward cluster 2 – 5th most common venues

As in the previous clusters, this one is more a suburb type of ward, so it is not recommended to consider it a good place to open a restaurant.

5.7.   Cluster 7

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
25	Casa de Campo	BBQ Joint	Trail	Fast Food Restaurant	Furniture / Home Store	Frozen Yogurt Shop

Table 8. Madrid ward cluster 2 – 5th most common venues

This cluster is definitely a residential area, so it is not going to be considered as a good place to open a new restaurant.

5.8. Cluster 8

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
40	San Cristóbal	Furniture / Home Store	Train Station	Athletics & Sports	Park	Yoga Studio

Table 9. Madrid ward cluster 2 – 5th most common venues

This cluster is a residential area, so it cannot be considered as a good place to invest in a new restaurant.

5.9. Cluster 9

In the following table, it can be seen the ward included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
38	Pradolongo	Park	Burger Joint	Beer Garden	Theater	Yoga Studio

Table 10. Madrid ward cluster 2 – 5th most common venues

This cluster is a residential area, so it will not be considered to open restaurant.

5.10.   Cluster 10

In the following table, it can be seen the wards included and the 5th most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	Jerónimos	Park	Fountain	Coffee Shop	Diner	Plaza
18	Niño Jesús	Italian Restaurant	Park	Spanish Restaurant	Coffee Shop	Plaza
27	Ciudad Universitaria	Music Venue	Museum	Restaurant	Park	Bus Station
33	Orcasur	Clothing Store	Pharmacy	Grocery Store	Farmers Market	Coffee Shop
34	San Fermín	Furniture / Home Store	Metro Station	Pharmacy	Restaurant	Park
36	Moscardó	Bakery	Restaurant	Coffee Shop	Clothing Store	Pool
41	Butarque	Plaza	Grocery Store	Park	Sports Bar	Yoga Studio
46	Ensanche de Vallecas	Clothing Store	Coffee Shop	Fast Food Restaurant	Park	Sandwich Place
49	Amposta	Metro Station	Spanish Restaurant	Supermarket	Café	Food

Table 11. Madrid ward cluster 1 – 5th most common venues

This cluster is a mix of suburb-type of ward and a more tourist ward, but it does not seem to be an excellent location for a new business.

6. CONCLUSION

This study tries to help investors to choose the best wards in Madrid to open a restaurant.

First of all, Madrid district data from different webpages (see section 7) has been retrieved to select the districts with highest income and more tourist by clustering the districts by k-mean method, and optimizing k by elbow method. Then, using Foursquare API service and geocode location, information about the venues in each ward of the selected districts was obtained. Using this info, wards were clustered using k-mean based on the venues. Ten different clusters were chosen as optimum in this case.

Finally clusters 1 and 3 were selected as the ones that will be the best to invest in a restaurant.

The final decision will be done depending on the type of restaurant the investors would like to open, but as a first advice, it can be said that all types of restaurants could match in the selected wards, except a Spanish restaurant, since it seems Madrid is packed of them.

7. REFERENCES

- Administrative distribution of Madrid:
  https://es.wikipedia.org/wiki/Anexo:Barrios_administrativos_de_Madrid

- Information of Madrid (income, population, etc. by "distritos"):
  - https://www.ine.es/experimental/experimental.htm
  - https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=46b55cde99be2410VgnVCM1000000b205a0aRCRD&

- Foursquare API

- Geocode API