

CLUSTERING NEIGHBORHOODS IN MADRID

LUIS MUÑOZ DEL POZO
DECEMBER, 2019

INTRODUCTION

- 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.
- Madrid is administratively divided in 21 districts (“distritos”) and 131 wards (“barrios”).
- 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.

DATA

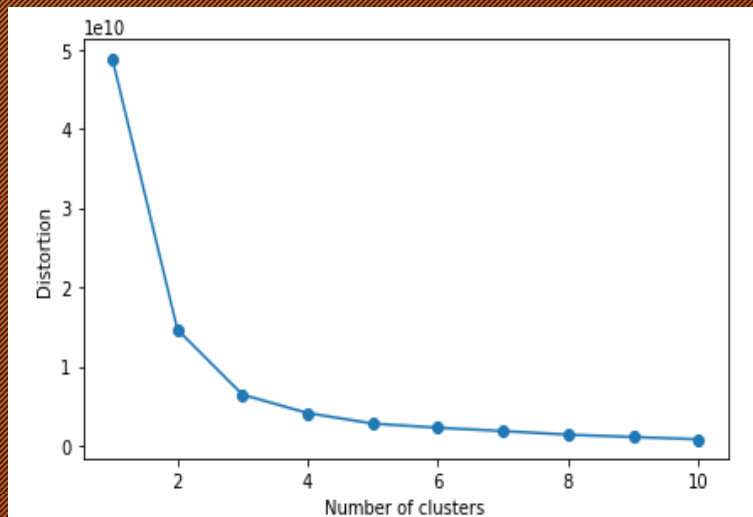
- Information of the administrative distribution in district and wards. This information will be obtained from the Madrid City Hall official website
- Information of all of the district such as population, size (area), average income, number of crimes, or number of hotel beds.
- Information of the coordinates of every ward by geocoding
- Information about the type and location of restaurants from Foursquare API

METHODOLOGY

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 restaurant to invest in, considering the clusters created in phase 2.

Phases 1 and 2 will be developed by clustering the “distritos” and “barrios” respectively, using the k-mean method, optimizing the number of clusters using the elbow method.

CLUSTERING “DISTRITOS” BY DEMOGRAPHIC DATA



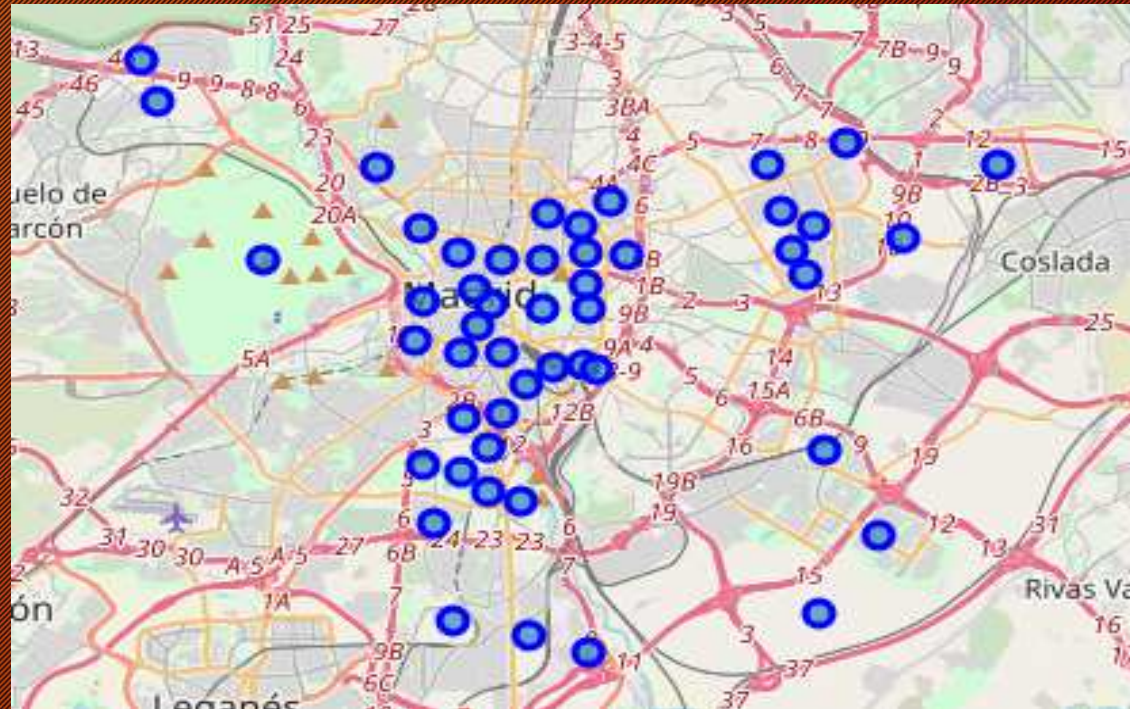
Optimizing number of clusters using elbow method (k=3)

“Distritos” Clusters Summary

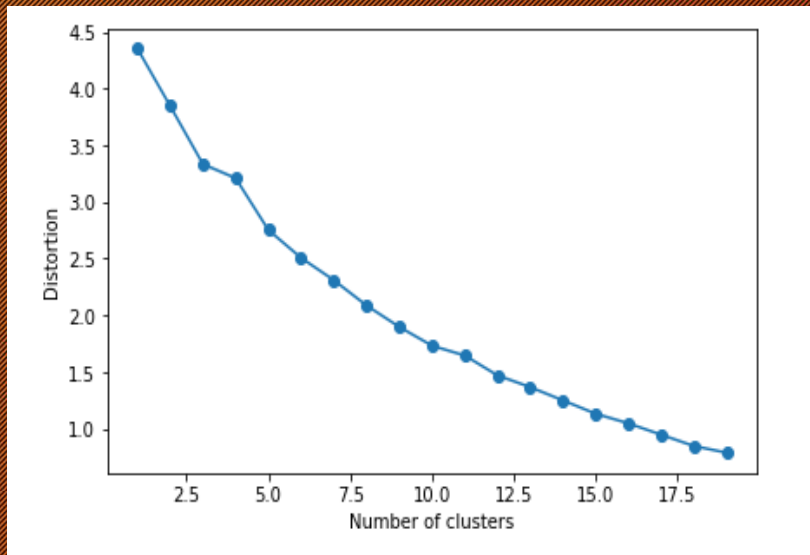
	Renta (Euro/hab)	Pob(16_79)	Superf. (Ha)	Plazas Hotel	Delitos
Distritos_clust					
0	14426.666667	61976.000000	2776.423333	1552.666667	215.333333
1	14091.166667	197749.333333	5518.800000	1335.000000	660.000000
2	17455.000000	120235.083333	1583.620000	5658.583333	796.083333

CLUSTERING “DISTRITOS” BY DEMOGRAPHIC DATA (cont.)

“Distritos”
Cluster 2 Map

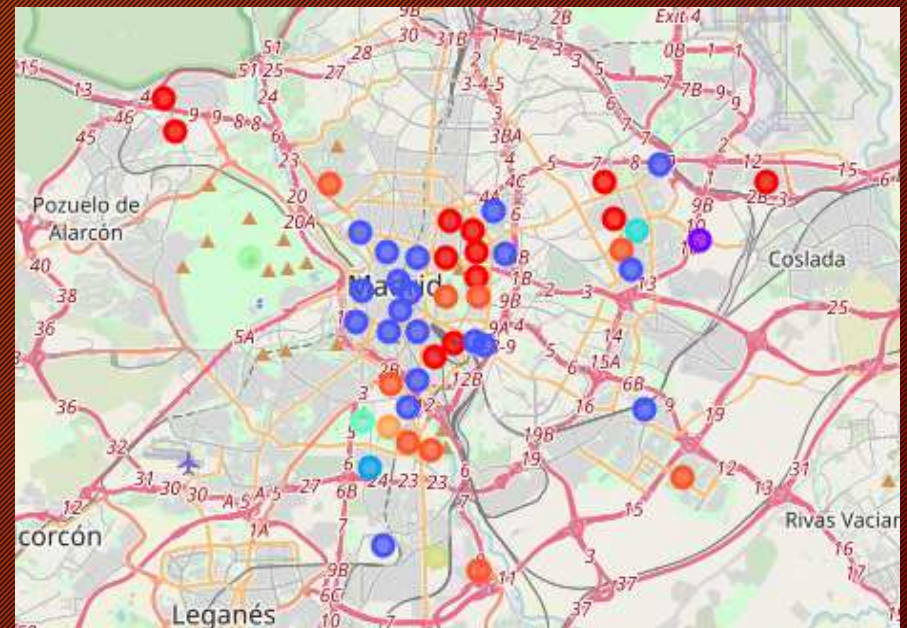


CLUSTERING “BARRIOS” BY VENUES



Optimizing number of clusters using elbow method (k=10)

“Barrios” in Madrid



DISCUSSION

10 CLUSTERS SELECTED:

- CLUSTERS 2, 4, 5, 6, 7, 8, and 9 with only one ward, and no many restaurants. Residential areas.
- CLUSTER 10 with 9 wards, but not many hotels and restaurants.
- CLUSTER 1 (13 wards) and CLUSTER 3 (20 wards) with a lot of restaurants and all kind of venues. Most of the venues are Spanish Food restaurants

CONCLUSION

- Clusters 1 and 3 were selected as the ones that will be the best to invest in a restaurant
- As most of the venues are Spanish Food restaurants, it is suggested to invest in other kind of restaurant in the selected wards