

**Evaluation of the Optimal Location for Erecting a Mexican
Restaurant in the City of Madrid, Spain.**

by

Molokwu Reginald Chukwuka

**In partial fulfillment for the award of Specialization certificate &
Professional badge in Data science from IBM**

Date: 24th June 24, 2019

Type: Peer Review

Table of Content

1.1. Problem Description.....	3
1.2. Data Presentation.....	3
1.3. Target Audience.....	3
3.1. Methodology.....	4
3.2. Results.....	9
3.3. Discussion.....	9
4. Conclusions.....	10

1.1. Problem Description

In this project, the problem attempted to solve will be to find the best possible location or the most optimal, for a Mexican restaurant in the city of Madrid, Spain. To achieve this task, an analytical approach will be used, based on advanced machine learning techniques and data analysis, concretely clustering and perhaps some data visualization techniques.

During the process of analysis, several data transformations will be performed, in order to find the best possible data format for the machine learning model to ingest. Once the data is set up and prepared, a modeling process will be carried out, and this statistical analysis will provide the best possible places to locate the Mexican restaurant.

1.2. Data Presentation

The data that will be used to develop this project is based on two sites:

1. The Foursquare API: This data will be accessed via Python, and used to obtain the most common venues per neighborhood in the city of Madrid. This way, it is possible to have a taste of how the city's venues are distributed, what are the most common places for leisure, and in general, it will provide an idea of what people's likes are.

2. The Madrid City Hall's Web Portal: This site provides several data sources of great utility to solve this problem. The files are provided in Excel format, and they are built over a statistical exploitation and use basis. The data contains updated information about the immigrant population per country and per nationality. This data will be analyzed in such a way that one could determine the best location of a new venue/restaurant/other based on people's nationalities. For the sake of simplicity, it will be assumed for this exercise that people's likes varies according to their nationality, and that people from one specific country will be more attracted to place that matches the environment and culture of their own countries, rather than the ones from foreign countries.

You can access the data by clicking [this link](#)

1.3. Target Audience

The target audience of this project could be any business owner that is planning to open a new business local, restaurant, real state agency, shops, etc... Since this approach could be applicable not only to Mexican food restaurant but to other kind of businesses, anybody who is considering to place a new business local or even relocate it, could benefit of this project's approach.

2.1. Methodology

The methodology used to approach this problem includes some statistical exploration of the data and some visualizations. The main machine learning technique involved in the development of this project is clustering, in concrete the K-Means algorithm was used, implemented with Python. At a first moment, the main problem was how to obtain the necessary data to build a constructive approach to the problem to be tackled. Usually, to solve these kinds of optimal business location problems, a lot of consumer's data are needed, but for this example and for the sake of simplicity, the focus was put mainly on the population's nationality. A study was carried out over the inhabitants of Madrid, and it was assumed for this example that the national population from a certain country would prefer restaurants based on their national country and food, rather than restaurants from other countries or that have nothing to do with the culture of their countries, specially when it comes to immigrant populations, that are not in their countries, and certainly would like to usually have a taste of their food and original culture. Because in the end, it is not only about the food, it is also about having a piece of the country in question. When a someone enters in an Italian restaurant, or American, or Peruvian restaurant, they are not only consuming the food and culinary specialties of the country in question, but also the culture, the people, the music, the decoration. All of this must make people feel like they were there on the country. With all this being considered, it was decided that the main goal to efficiently solve this problem, was firstly to define what our target population is, and secondly, find the areas where this population is living, and finally, examine the venues and restaurants in this area to see if our product could work.

Here is an example of the data used:

Country of Pri	Total Ciudad	Centro	Arganzuela	Retiro	Salamanca	Chamartin	Tetuán
Rumanía	450360	8150	7540	4800	7530	6800	14680
China	372760	15080	13560	5640	7550	6520	19880
Ecuador	239530	6470	7410	2650	6190	3800	13950
Venezuela	233590	15630	9130	6380	15640	9330	13100
Colombia	226180	9980	7170	4830	8030	5510	8220
Marruecos	219090	11010	3900	1840	3220	2800	13930
Italia	203080	30300	12190	8400	18170	10600	11940
Perú	188290	5630	5210	2530	6120	4190	9650
Paraguay	186820	3640	4740	2370	5210	6570	33110
República Dor	175110	3650	6540	2040	3440	3220	22720
Honduras	159810	1490	2280	2320	3320	3370	7550

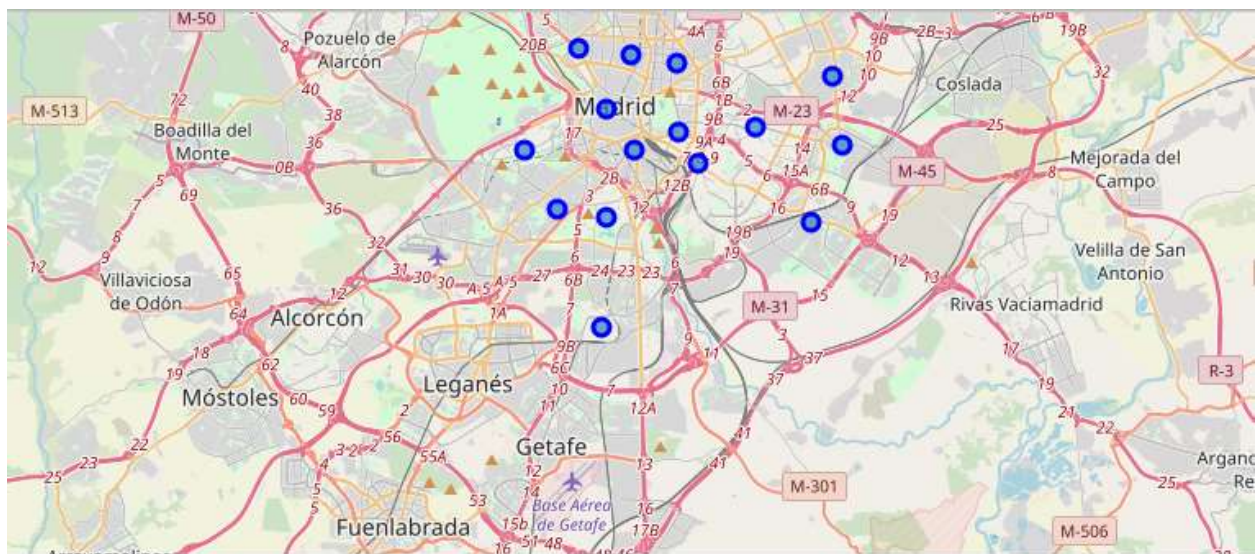
This data contains information about the quantities of immigrant populations in Madrid inside each Neighborhood. The main features are the country of precedence, which P á g i n a 4 | 10 indicates where the people of that lives in those neighborhoods come from. It contains also the quantities of people by country living in each neighborhood. So, with this, it is already possible to have an idea of where is our target population located. In this project, the idea is to open a Mexican restaurant in the city. With further analysis, this question will be answered. Nevertheless, this task could not be achieved only working with this raw data. It was also needed to obtain information about the most common venues in these neighborhoods, besides of the population kind that was inhabiting on the different neighborhoods. It was

also needed to determine somehow in what measure these neighborhoods were different or similar between them. To continue this line, The Foursquare API was used to obtain the needed data about the venues in each neighborhood, but to use the Foursquare API, it was first necessary to transform the raw data to something the Foursquare API was capable to handle. Basically, the coordinates of each neighborhood were needed.

This is an example of the transformed data:

Neighborhood	Latitude	Longitude
Centro	40415347	-3707371
Arganzuela	40402733	-3695403
Retiro	40408072	-3676729
Salamanca	4043	-3677778
Chamartin	40453333	-36775
Tetuán	40460556	-37
Chamberí	40432792	-3697186
Fuencarral-El P	40478611	-3709722
Moncloa-Aravaca	40435151	-3718765
Latina	40402461	-3741294
Carabanchel	40383669	-3727989
Usera	40381336	-3706856
Puente de Val	40398204	-3669059
Moratalaz	40409869	-3644436
Ciudad Lineal	4045	-365

Once the data was transformed into a format ingestible by the Foursquare API, the information about the venues could be obtained. The neighborhoods were then plotted into a map of Madrid, so it was possible to have an idea of their geographical situation:



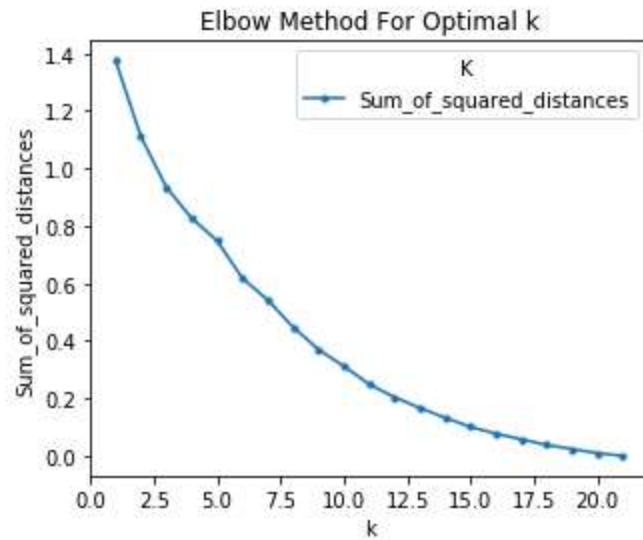
The next step was to obtain the nearby venues by neighborhood, together with their respective coordinates:

Neighborhood	Neighborhood	Neighborhood	Venue	Venue Latitud	Venue Longitud	Venue Category
Centro	40415347	-3707371	Plaza Mayor	4,0415E+16	-3,7076E+16	Plaza
Centro	40415347	-3707371	Mercado de S	4,0415E+15	-3,709E+16	Market
Centro	40415347	-3707371	La Taberna de	4,0415E+16	-3,7081E+15	Other Nightlife
Centro	40415347	-3707371	The Hat Madr	4,0414E+16	-3,7071E+14	Hotel
Centro	40415347	-3707371	Amorino	4,0416E+15	-3,7084E+16	Ice Cream Sho
Centro	40415347	-3707371	Bot��n	4,0414E+15	-3,7081E+15	Spanish Restai
Centro	40415347	-3707371	Bar El Cogollo	4,0414E+15	-3,7067E+15	Spanish Restai
Centro	40415347	-3707371	Chocolater��a	4,0417E+16	-3,7068E+16	Chocolate Sho
Centro	40415347	-3707371	Pinkleton & W	4,0415E+15	-3,7091E+16	Wine Bar

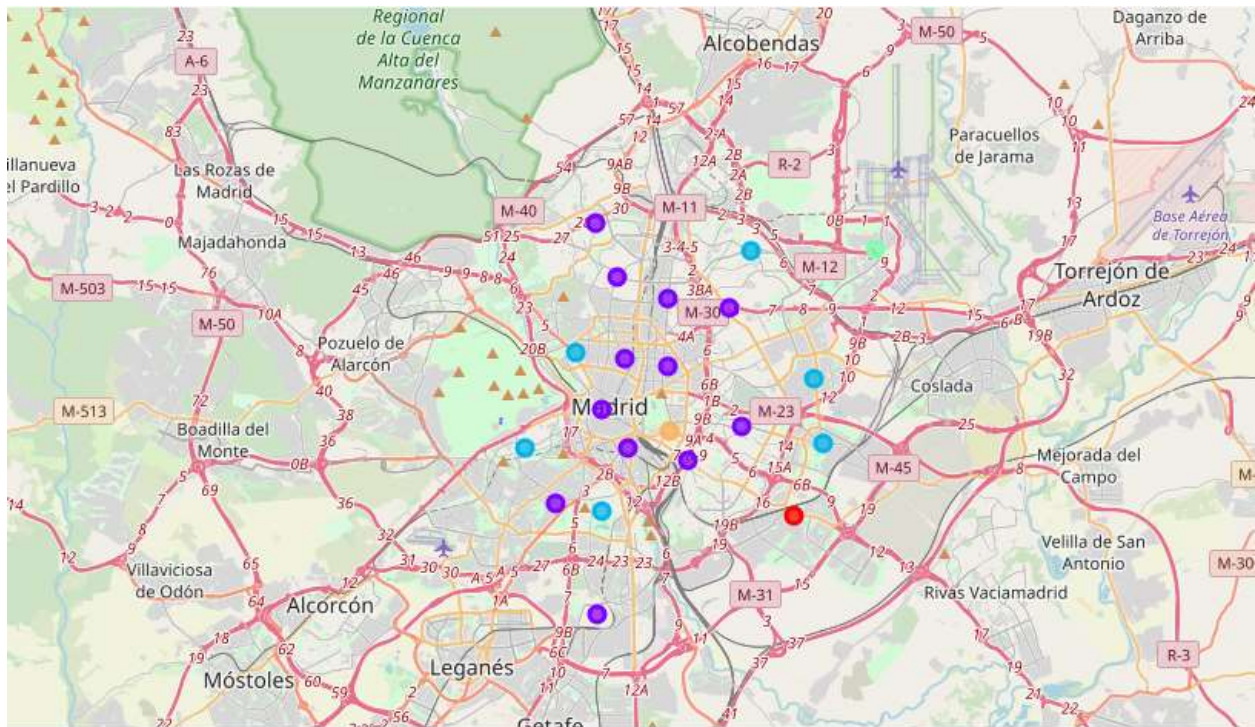
Looking at this sample, it is possible to see the names of the venues, their coordinates, and the category of each venue. The results are ordered by neighborhood. This is a vital step in the segmentation process, since all the important data about the venues is obtained from here. Once the venues per neighborhood were obtained, it was then needed to look at the mean occurrence of each venue by neighborhood:

----Arganzuela----			----Carabanchel----		
	venue	freq		venue	freq
0	Spanish Restaurant	0.12	0	Plaza	0.14
1	Restaurant	0.10	1	Bakery	0.14
2	Grocery Store	0.05	2	Soccer Field	0.14
3	Tapas Restaurant	0.05	3	Nightclub	0.14
4	Bakery	0.05	4	Fast Food Restaurant	0.14
----Barajas----			----Centro----		
	venue	freq		venue	freq
0	Hotel	0.23	0	Spanish Restaurant	0.13
1	Spanish Restaurant	0.10	1	Tapas Restaurant	0.08
2	Restaurant	0.10	2	Plaza	0.05
3	Coffee Shop	0.06	3	Restaurant	0.04
4	Supermarket	0.06	4	Hotel	0.04
----Carabanchel----			----Chamart��n----		

This process is progressive, once a piece of information is obtained, it is possible to go for the next one. With this data in hands, now the segmentation can be made, and the clusters created. But first it is necessary to determine somehow, what the appropriate number of clusters is. To perform this task, the elbow method was used. This method consists in plotting a hypothetical and usually large number of clusters in our data, and draw a curve representing the squared distances between each cluster. At some point, the distances will descend to a point where there is no need to keep increasing them. This means that creating more divisions in the data (clusters) is pointless as the difference between groups starts being highly difficult to appreciate:



This is our curve above. The distances start reducing importantly from cluster 5 on. So, it was determined that the optimal number of clusters for this problem was 5. With this being done, it is possible to build the clusters now and have a look at them:



This are the 5 clusters on the map of Madrid, it is possible to see how many neighborhoods belong to each cluster, which is also important information. Now it is possible to examine the data of each cluster:

Cluster 1:

Neighborhood	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	Latitude	Longitude
Usara	0	Seafood Restaurant	Bubble Tea Shop	Spanish Restaurant	Asian Restaurant	Noodle House	Café	Theater	Fast Food Restaurant	Chinese Restaurant	Falafel Restaurant	40.3796	-3.62135

Cluster 2:

Neighborhood	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	Latitude	Longitude
Centro	1	Spanish Restaurant	Tapas Restaurant	Plaza	Mexican Restaurant	Hotel	Restaurant	Ice Cream Shop	Gift Shop	Wine Bar	Bistro	40.430000	-3.677778
Arganzuela	1	Spanish Restaurant	Restaurant	Bakery	Tapas Restaurant	Grocery Store	Gym / Fitness Center	Breakfast Spot	Pizza Place	Hotel	Falafel Restaurant	40.415347	-3.707371
Retiro	1	Spanish Restaurant	Bar	Tapas Restaurant	Grocery Store	Museum	Gym	Mexican Restaurant	Mediterranean Restaurant	Pizza Place	Plaza	40.409859	-3.644436
Fuencarral-El Pardo	1	Italian Restaurant	Clothing Store	Fast Food Restaurant	American Restaurant	Burger Joint	Tapas Restaurant	Cosmetics Shop	Sandwich Place	Restaurant	Pizza Place	40.478611	-3.709722
Chamberí	1	Spanish Restaurant	Brewery	Bar	Restaurant	Japanese Restaurant	Tapas Restaurant	Hotel	Plaza	Café	Burger Joint	40.460556	-3.700000
Ciudad Lineal	1	Spanish Restaurant	Gastropub	Supermarket	Restaurant	Burger Joint	Argentinian Restaurant	Pharmacy	Diner	Bakery	German Restaurant	40.432792	-3.697186
Moncloa-Aravaca	1	Bar	Tapas Restaurant	Italian Restaurant	Bakery	Japanese Restaurant	Spanish Restaurant	Pizza Place	Coffee Shop	Mediterranean Restaurant	Restaurant	40.383659	-3.727990
Salamanca	1	Spanish Restaurant	Seafood Restaurant	Tapas Restaurant	Coffee Shop	Restaurant	Mediterranean Restaurant	Clothing Store	Mexican Restaurant	Gym	Bakery	40.450000	-3.650000
Fuente de Vallecas	1	Fast Food Restaurant	Gym	Grocery Store	Bakery	Bar	Hotel	Tapas Restaurant	Supermarket	Coffee Shop	Camera Store	40.398204	-3.669059

Cluster 3:

Neighborhood	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	Latitude	Longitude
San Blas-Canillejas	2	Metro Station	Supermarket	Snack Place	Pizza Place	Gym	Asian Restaurant	Grocery Store	Shopping Mall	Flea Market	Fish Market	40.469457	-3.640482
Latina	2	Pizza Place	Park	Train Station	Grocery Store	Fast Food Restaurant	Bakery	Metro Station	Falafel Restaurant	Women's Store	Electronics Store	40.402461	-3.741294
Villa de Vallecas	2	Soccer Field	Grocery Store	Spanish Restaurant	Sandwich Place	Plaza	Church	Train Station	Diner	Fabric Shop	Fish Market	40.426001	-3.612764
Hortaleza	2	Pizza Place	Breakfast Spot	Spanish Restaurant	Pharmacy	Supermarket	Park	Clothing Store	Donut Shop	Plaza	Pub	40.435151	-3.718765
Vicálvaro	2	Pizza Place	Spanish Restaurant	Breakfast Spot	Plaza	Café	Fast Food Restaurant	Sandwich Place	Beer Bar	Grocery Store	Dog Run	40.404200	-3.608060
Moratalaz	2	Brewery	Bar	Park	Soccer Field	Pizza Place	Food Truck	Plaza	Ice Cream Shop	Bakery	Nightclub	40.381336	-3.706856

Cluster 4:

Neighborhood	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	Latitude	Longitude
Villaverde	3	Spanish Restaurant	Pizza Place	Diner	Electronics Store	Flea Market	Fish Market	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fabric Shop	40.470196	-3.58489

Cluster 5:

Neighborhood	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	Latitude	Longitude
Carabanchel	4	Soccer Field	Burger Joint	Pizza Place	Plaza	Fast Food Restaurant	Bakery	Nightclub	Electronics Store	Fish Market	Farmers Market	40.408072	-3.676729

So, this kind of approach, allow us to perform an analysis of an entire city by looking at its venues and population. With this information, observations and conclusions can be made now.

3.2. Results

The results obtained were five clusters of very different population and venues distribution. The following is a description of the clusters:

- Cluster One: Occupied by Bulgarians and the most common venue is the seafood restaurant.
- Cluster Two: Mostly inhabited by south Americans, Europeans, and North Americans. The most common venues are tapas restaurants, Argentinian restaurants, pizza places, supermarkets and Spanish restaurants, among many others.
- Cluster Three: This cluster is composed only by 3 different population kinds: Americans Ukrainian people and Dominican Republic people. The most common venues are Pizza place, gym, shopping mall, church and bakeries etc.
- Cluster Four: This cluster is only composed by Bangladeshi people. The most common places are Spanish restaurant, falafel restaurants, fish markets, fast food restaurants and electronic stores.
- Cluster Five: Again, only people from Ecuador seems to live in this cluster. The most common venues are soccer fields, burger joint, plaza, fast food restaurants.

3.3. Discussions

It is interesting how the venues and people from different countries varies to one cluster to another. The main differentiation is located on these two variables. Each cluster has its own characteristics, but also common spots with other clusters. If we examine with more detail these results, some conclusions can be made. As a recommendation, it must be said in a study of this size, to make good predictions about where to open a certain business or shop, more data is needed. For example, socio-demographic data about the population, like their income level, if they have children or not, the education level, what kind of job do they make a living from, etc.... Also, one of the most important data to examine carefully are the data related to the people's likes and tastes about how they prefer to spend their leisure time, what kinds of food do they like, or what are their hobbies. With all these data gathered, a more indepth analysis could be performed, and the segmentations would be more accurate. For this project, these data weren't available, and was also out of the project's scope.

4. Conclusion

As far as we are able to see with this data, there are no mexican populations registered in Madrid. However, in Cluster 2, it is possible to notice that there's a mexican restaurant located in the "Centro" neighborhood, which is the town center.

If a deeper exam is performed into this cluster, it is noticeable that its the living population are mostly latinos, mixed with some other europeans, but mainly, the people living in this cluster come from south american countries. Apart of this fact, other kinds of latin restaurants can be found, like argetinian restaurants, tapas restaurants, and italian restaurants. So it is possible to tell that the inhabitants of this area likes these kinds of food.

By following this logic, if we would like to open a new mexican restaurant in the city or any kind of restaurant in fact, it would only be necessary to find a where are the restaurants similar the the one we want to open, study the population in that area, and find similar clusters of population in the city that don't have yet or have very few resturants like the one we would like to open.

In this example, clusters 4 and 5 could make a good match for our target population. Looking at the venues in these clusters, it is possible to find one mexican restaurant, and a good bunch of fast food, argentinian, and south american restaurants. So, in these clusters, it is possible to state that the existing restaurants matches the population's nationalities and tastes.

In conclussion, and taking into consideration the explanations given above as well as the data, it is highly possible that clusters 4 and five could be a good place to open our mexican restaurants. As explained above, the same logic could apply to oopen other kind of restaurant or business in any other area of the city. It is only necessary to to examine the the existing businesses in our target area, and study the population, then compare these 2 factors with the same ones in areas where there are existing business like the one we want to open, and then verify if the matching is correct