Establishing a restaurant in an district of Madrid City

Final report of the “Applied Data Science Capstone” course taught by IBM on Coursera

Marius Claudiu Spalatelu

June 31, 2021

Glossary

1. Introduction

If you have ever traveled to Madrid, you will have seen that there are many restaurants, bars, coffee shops, shopping centers and more. In our case, we have the need to open a restaurant in Madrid city and we do not know exactly where to establish ourselves and we want to know a little more what could be the right place where we can have profitability with the business. For this we are going to work on knowing a little more information about each district of Madrid and have an idea where it is convenient for us to establish the restaurant.

We are going to study the density of the population and the prices of the houses for each district, to know which district is more expensive and which is cheaper and thus we get an idea of the prices.

The main objective is to decide where we are going to locate the restaurant and it would also help us with this information if we can afford to rent the premises or later buy the place where we will install ourselves.

Using Data Science we are going to study the information that exists between different districts of Madrid, which will help us to finish the site where we can establish the restaurant.

Next we will see the processes that will be followed in the process.

1. How many venues in each district? By answering the question, this will help us better understand the level for each district.
2. How many categories in each district? By answering the question, this will help us better understand the diversity of businesses for each district.
3. How many venues in each category? In answering the question, this shows us the magnitude of the categories for each district.
4. What are the most popular categories in each district? In this section we will see relevant information that will help investors understand which businesses are more profitable compared to opening a restaurant.
5. How many clusters we can use to categorize the districts based on the popularity of restaurants? Here we will visualize the districts for each cluster on the map.

6. Choose the place we consider and view it on the map. We will see the site where we are going to locate our restaurant after having studied the previous aspects and we determine the site under our criteria.

2. Data sources

In the research, different data sources have been searched, which will be explained below.

1. List of Madrid City districts from Wikipedia. List of Madrid City districts from Wikipedia. This provides us with information about Population (1 Jan 2020) and Area (ha) and also calculates the population density for each district of Madrid.

2. List of housing prices of Madrid city by district from Madrid City Council website. With the information provided by the website of the Madrid city council, we can know the list of prices of the houses per € m / 2 and it will help us to understand if a district is more expensive or cheaper compared to other districts. In this way it helps us to evaluate the establishment of the restaurant.

3. Methodology

1. Extract the information by scraping table of Madrid city from wikipedia webpage and the average price for each housing district. This provides us with relevant information on the quality of life in each district. I will use the BeautifulSoup library to extract the information as it is very useful.

2. The Density column is calculated to see the number of people in each district based on the extracted columns Population (1 Jan 2020) and Area (ha).

3. The Nominatim functionality provided by the geopy.geocoders library is used to obtain the coordinates of each district and manipulate the information to assign it in a dataframe with Pandas.

4. I will use the Foursquare API to explore the venues in each district.

5. I will use the Folium library represent each restaurant in their respective locations and also the districts.

6. Matplotlib library will also be used to display the charts.

4. Data Analysis

4.1 **Venues per district**

In this step we are going to compare the number of venues between the districts.

Looking at the results of the table shown below (Figure 1), it can be seen that there are certain districts that have a large population and more density per m2. We can see districts such as Retiro, Salamanca, Chamartin, Centro, Arganzuela that have a higher population density compared to Hortaleza, Villaverde, Usera, Ciudad Lineal for example.
In Figure 2 we can see the number of venues for each district, where the difference between some districts with respect to others is clearly observed.

	District	Population (1 Jan 2020)	Area (ha)	Density (pop/m2)
0	Centro	140,991	522.82	2.697
1	Arganzuela	156,176	646.22	2.417
2	Retiro	120,873	546.62	2.211
3	Salamanca	148,405	539.24	2.752
4	Chamartín	148,039	917.55	1.613
5	Tetuán	161,991	537.47	3.014
6	Chamberí	141,397	467.92	3.022
7	Fuencarral-El Pardo	250,636	23,783.84	0.105
8	Moncloa-Aravaca	122,164	4,653.11	0.263
9	Latina	242,923	2,542.72	0.955
10	Carabanchel	261,118	1,404.83	1.859
11	Usera	143,365	777.77	1.843
12	Puente de Vallecas	241,666	1,496.86	1.614
13	Moratalaz	95,907	610.32	1.571
14	Ciudad Lineal	220,598	1,142.57	1.931
15	Hortaleza	193,833	2,741.98	0.707
16	Villaverde	154,915	2,018.76	0.767
17	Villa de Vallecas	114,832	5,146.72	0.223
18	Vicálvaro	74,235	3,526.67	0.210
19	San Blas-Canillejas	161,672	2,229.24	0.725
20	Barajas	50,158	4,192.28	0.120

Figure 1: The main dataframe.

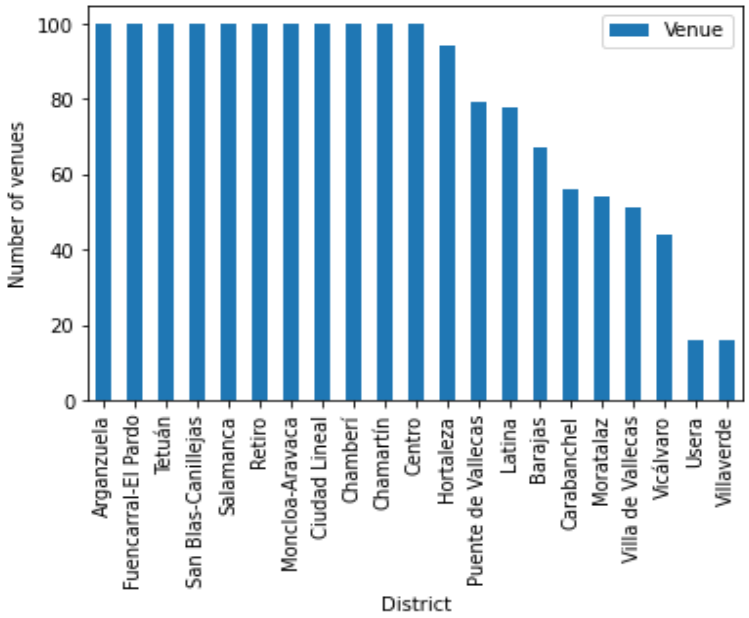


Figure 2: The number venues in each district.

4.2 **Housing prices**

Below we will see the average price of homes in the districts of Madrid, provided by the Madrid city council.

This information is very useful, since it tells us which districts are more expensive or cheaper depending on the location of each one of them.

This reveals to us where we can locate the restaurant based on the economic situation we have and we can sustain to open the business.

	District	Average Housing Price (€/m2)
0	Centro	4802
1	Arganzuela	4000
2	Retiro	4692
3	Salamanca	5949
4	Chamartín	5115
5	Tetuán	3637
6	Chamberí	5248
7	Fuencarral-El Pardo	3518
8	Moncloa-Aravaca	4004
9	Latina	2265
10	Carabanchel	2155
11	Usera	1975
12	Puente Vallecas	1909
13	Moratalaz	2544
14	Ciudad Lineal	2945
15	Hortaleza	3754
16	Villaverde	1721
17	Villa Vallecas	2313
18	Vicálvaro	2438
19	San Blas-Canillejas	2520
20	Barajas	3131

4.3 **Categories per district**

In figure 3, we can see the magnitude of business categories in each district of Madrid.

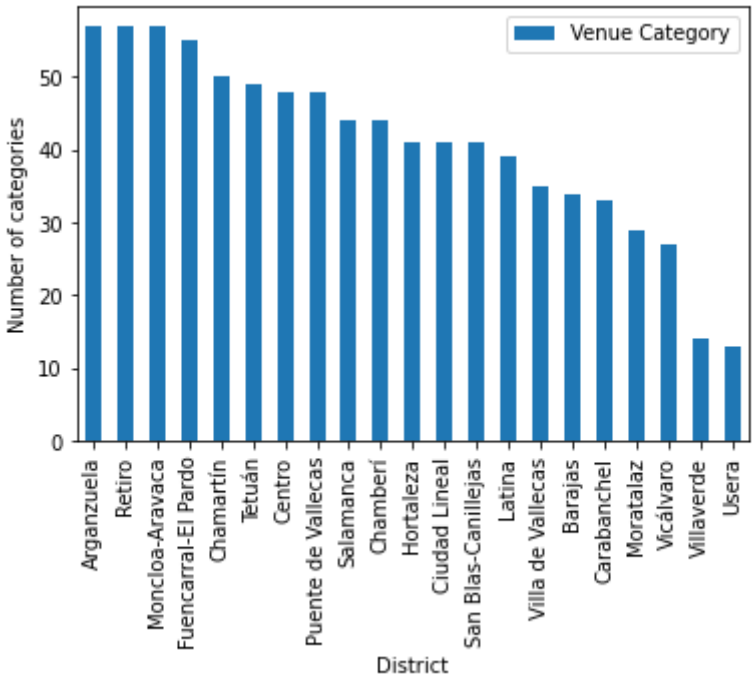


Figure 3: Number of categories in each district.

As we have previously seen the districts with the highest population density, in this we continue to see that there are certain districts such as Arganzuela, Retiro, Moncloa-Aravaca, Chamarin, Tetuán and Centro that have a high number of different types of businesses compared to other districts.

4.4 **District per category**

As can be seen in the following figure, we see the different categories for each district. In this case, the first 5 districts are displayed. We see what each type of business brings us for each district.

	District	Accessories Store	Airport	Airport Lounge	Airport Service	American Restaurant	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	...	Trail	Train Station	Travel Lounge	Turkish Restaurant	Vegetarian / Vegan Restaurant	Wine Bar	Wine Shop	Winery	Yoga Studio	Zoo
0	Arganzuela	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.0	0.0	0.020000	...	0.00	0.00	0.00	0.0	0.0	0.010000	0.00	0.00	0.0	0.0
1	Barajas	0.014925	0.014925	0.044776	0.074627	0.00	0.00	0.0	0.0	0.029851	...	0.00	0.00	0.00	0.0	0.0	0.014925	0.00	0.00	0.0	0.0
2	Carabanchel	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.0	0.0	0.000000	...	0.00	0.00	0.00	0.0	0.0	0.000000	0.00	0.00	0.0	0.0
3	Centro	0.000000	0.000000	0.000000	0.000000	0.00	0.01	0.0	0.0	0.000000	...	0.01	0.00	0.00	0.0	0.0	0.030000	0.01	0.01	0.0	0.0
4	Chamartín	0.000000	0.000000	0.000000	0.000000	0.01	0.00	0.0	0.0	0.000000	...	0.00	0.02	0.01	0.0	0.0	0.010000	0.00	0.00	0.0	0.0

4.5 **Top most common venue categories in each district**

In the following figure we can see the 10 most popular categories for each district. This gives us an idea of which districts have more variety of businesses and if we have a lot of competition against other restaurants.

	District	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	Arganzuela	Spanish Restaurant	Restaurant	Park	Bakery	Art Gallery	Grocery Store	Gym	Burger Joint	Chinese Restaurant	Indie Theater
1	Barajas	Hotel	Spanish Restaurant	Airport Service	Duty-free Shop	Coffee Shop	Airport Lounge	Rental Car Location	Restaurant	Argentinian Restaurant	Snack Place
2	Carabanchel	Fast Food Restaurant	Restaurant	Tapas Restaurant	Grocery Store	Concert Hall	Spanish Restaurant	Bar	Pizza Place	Café	Plaza
3	Centro	Spanish Restaurant	Bar	Restaurant	Plaza	Bookstore	Mexican Restaurant	Brewery	Tapas Restaurant	Wine Bar	Hotel
4	Chamartín	Spanish Restaurant	Restaurant	Hotel	Bar	Gym / Fitness Center	Supermarket	Fast Food Restaurant	Café	Italian Restaurant	Breakfast Spot
5	Chamberí	Restaurant	Spanish Restaurant	Plaza	Tapas Restaurant	Japanese Restaurant	Italian Restaurant	Bar	Hotel	Coffee Shop	Bookstore
6	Ciudad Lineal	Spanish Restaurant	Restaurant	Bar	Supermarket	Tapas Restaurant	Café	Park	Hotel	Asian Restaurant	Coffee Shop
7	Fuencarral-El Pardo	Spanish Restaurant	Restaurant	Burger Joint	Coffee Shop	Pizza Place	Café	Italian Restaurant	Gym	Pharmacy	Japanese Restaurant
8	Hortaleza	Spanish Restaurant	Restaurant	Bar	Chinese Restaurant	Pizza Place	Gastropub	Supermarket	Bakery	Park	Hotel
9	Latina	Bar	Tapas Restaurant	Park	Restaurant	Italian Restaurant	Train Station	Pizza Place	Shopping Mall	Spanish Restaurant	Chinese Restaurant

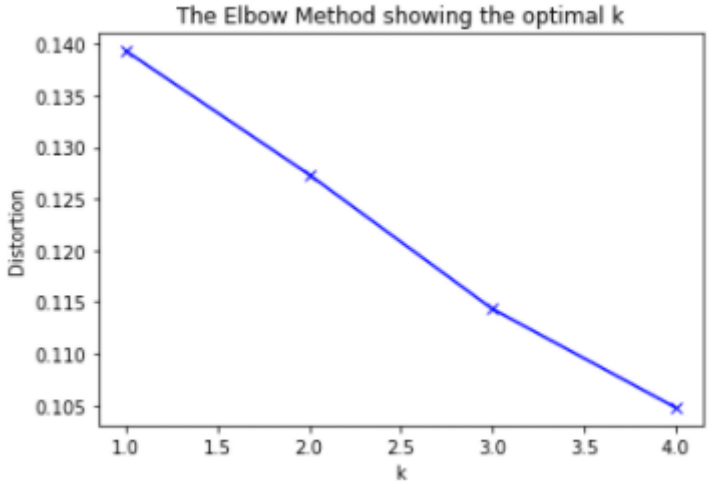
4.6 **How many clusters should we use?**

Considering the information of the restaurant category, we want to group this information of the restaurants into different groups. For this we need to determine the number of groups (K for K-means method).

We are going to use the elbow method to determine the number of groups we are going to need. In the research that has been done, it has been determined that the number 3 is the most suitable.

	District	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
10	Moncloa-Aravaca	Spanish Restaurant	Café	Tapas Restaurant	Hotel	Supermarket	Bar	Coffee Shop	Park	Restaurant	Ice Cream Shop
1	Barajas	Hotel	Spanish Restaurant	Airport Service	Duty-free Shop	Coffee Shop	Airport Lounge	Rental Car Location	Restaurant	Argentinian Restaurant	Snack Place
3	Centro	Spanish Restaurant	Bar	Restaurant	Plaza	Bookstore	Mexican Restaurant	Brewery	Tapas Restaurant	Wine Bar	Hotel
4	Chamartín	Spanish Restaurant	Restaurant	Hotel	Bar	Gym / Fitness Center	Supermarket	Fast Food Restaurant	Café	Italian Restaurant	Breakfast Spot
5	Chamberí	Restaurant	Spanish Restaurant	Plaza	Tapas Restaurant	Japanese Restaurant	Italian Restaurant	Bar	Hotel	Coffee Shop	Bookstore
6	Ciudad Lineal	Spanish Restaurant	Restaurant	Bar	Supermarket	Tapas Restaurant	Café	Park	Hotel	Asian Restaurant	Coffee Shop
7	Fuencarral-El Pardo	Spanish Restaurant	Restaurant	Burger Joint	Coffee Shop	Pizza Place	Café	Italian Restaurant	Gym	Pharmacy	Japanese Restaurant
8	Hortaleza	Spanish Restaurant	Restaurant	Bar	Chinese Restaurant	Pizza Place	Gastropub	Supermarket	Bakery	Park	Hotel
16	Tetuán	Spanish Restaurant	Restaurant	Seafood Restaurant	Bar	Hotel	Japanese Restaurant	Pub	Soccer Stadium	Tapas Restaurant	Ice Cream Shop
15	San Blas-Canillejas	Spanish Restaurant	Restaurant	Bar	Supermarket	Tapas Restaurant	Café	Park	Hotel	Asian Restaurant	Coffee Shop
14	Salamanca	Spanish Restaurant	Restaurant	Tapas Restaurant	Hotel	Boutique	Japanese Restaurant	Italian Restaurant	Jewelry Store	Clothing Store	Furniture / Home Store
18	Vicálvaro	Exhibit	Pizza Place	Grocery Store	Spanish Restaurant	Café	Tapas Restaurant	Park	Fast Food Restaurant	Gym / Fitness Center	Restaurant
13	Retiro	Spanish Restaurant	Hotel	Restaurant	Museum	Garden	Bakery	Park	Grocery Store	Gym	Italian Restaurant
0	Arganzuela	Spanish Restaurant	Restaurant	Park	Bakery	Art Gallery	Grocery Store	Gym	Burger Joint	Chinese Restaurant	Indie Theater
11	Moratalaz	Pizza Place	Spanish Restaurant	Tapas Restaurant	Park	Grocery Store	Café	Plaza	Bar	Bakery	Gym / Fitness Center
19	Villa de Vallecas	Grocery Store	Gym	Spanish Restaurant	Fast Food Restaurant	Restaurant	Supermarket	Bar	Bakery	Tapas Restaurant	Soccer Field
9	Latina	Bar	Tapas Restaurant	Park	Restaurant	Italian Restaurant	Train Station	Pizza Place	Shopping Mall	Spanish Restaurant	Chinese Restaurant
2	Carabanchel	Fast Food Restaurant	Restaurant	Tapas Restaurant	Grocery Store	Concert Hall	Spanish Restaurant	Bar	Pizza Place	Café	Plaza
12	Puente de Vallecas	Spanish Restaurant	Restaurant	Pizza Place	Supermarket	Bar	Italian Restaurant	Grocery Store	Gym	Pub	Tapas Restaurant
17	Usera	Park	Fast Food Restaurant	Beer Garden	Accessories Store	Athletics & Sports	Pet Store	Café	Diner	Museum	Sporting Goods Shop
20	Villaverde	Spanish Restaurant	Pizza Place	Deli / Bodega	Mediterranean Restaurant	Brewery	Bar	Furniture / Home Store	Diner	Gastropub	Park

Using the elbow method, it has been decided that the most suitable number for K is the number 3, so we see the graph below.



For each cluster, the information of the categories is grouped by different districts and we can see the information below:

Cluster 1:

	Population (1 Jan 2020)	Longitude	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	140,991	-5.664080	0	Spanish Restaurant	Bar	Restaurant	Plaza	Bookstore	Mexican Restaurant	Brewery	Tapas Restaurant	Wine Bar	Hotel
3	148,405	-3.683587	0	Spanish Restaurant	Restaurant	Tapas Restaurant	Hotel	Boutique	Japanese Restaurant	Italian Restaurant	Jewelry Store	Clothing Store	Furniture / Home Store
4	148,039	-3.675252	0	Spanish Restaurant	Restaurant	Hotel	Bar	Gym / Fitness Center	Supermarket	Fast Food Restaurant	Café	Italian Restaurant	Breakfast Spot
5	161,991	-3.700422	0	Spanish Restaurant	Restaurant	Seafood Restaurant	Bar	Hotel	Japanese Restaurant	Pub	Soccer Stadium	Tapas Restaurant	Ice Cream Shop
6	141,397	-3.697201	0	Restaurant	Spanish Restaurant	Plaza	Tapas Restaurant	Japanese Restaurant	Italian Restaurant	Bar	Hotel	Coffee Shop	Bookstore
7	250,636	-3.663639	0	Spanish Restaurant	Restaurant	Burger Joint	Coffee Shop	Pizza Place	Café	Italian Restaurant	Gym	Pharmacy	Japanese Restaurant
8	122,164	-3.718717	0	Spanish Restaurant	Café	Tapas Restaurant	Hotel	Supermarket	Bar	Coffee Shop	Park	Restaurant	Ice Cream Shop
14	220,598	-3.637071	0	Spanish Restaurant	Restaurant	Bar	Supermarket	Tapas Restaurant	Café	Park	Hotel	Asian Restaurant	Coffee Shop
15	193,833	-3.656963	0	Spanish Restaurant	Restaurant	Bar	Chinese Restaurant	Pizza Place	Gastropub	Supermarket	Bakery	Park	Hotel
19	161,672	-3.637071	0	Spanish Restaurant	Restaurant	Bar	Supermarket	Tapas Restaurant	Café	Park	Hotel	Asian Restaurant	Coffee Shop
20	50,158	-3.577779	0	Hotel	Spanish Restaurant	Airport Service	Duty-free Shop	Coffee Shop	Airport Lounge	Rental Car Location	Restaurant	Argentinian Restaurant	Snack Place

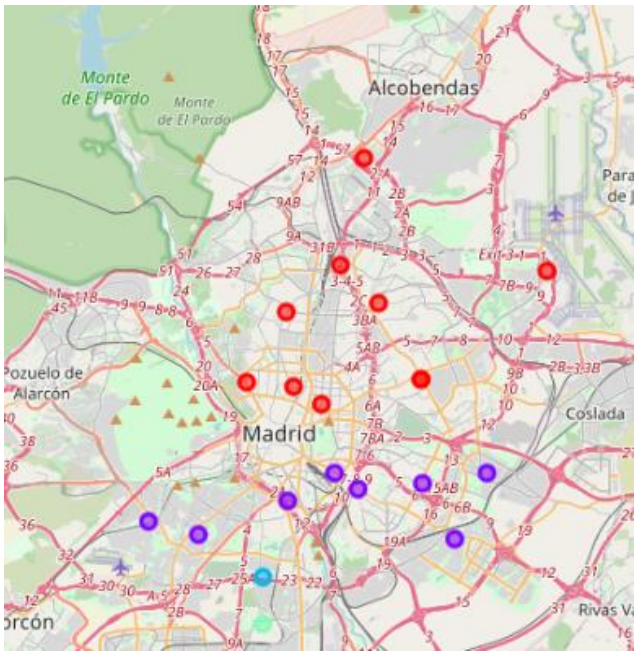
Cluster 2:

	Population (1 Jan 2020)	Longitude	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
1	156,176	-3.699313	1	Spanish Restaurant	Restaurant	Park	Bakery	Art Gallery	Grocery Store	Gym	Burger Joint	Chinese Restaurant	Indie Theater
2	120,873	-3.677427	1	Spanish Restaurant	Hotel	Restaurant	Museum	Garden	Bakery	Park	Grocery Store	Gym	Italian Restaurant
9	242,923	-3.764642	1	Bar	Tapas Restaurant	Park	Restaurant	Italian Restaurant	Train Station	Pizza Place	Shopping Mall	Spanish Restaurant	Chinese Restaurant
10	261,118	-3.741459	1	Fast Food Restaurant	Restaurant	Tapas Restaurant	Grocery Store	Concert Hall	Spanish Restaurant	Bar	Pizza Place	Café	Plaza
12	241,666	-3.666470	1	Spanish Restaurant	Restaurant	Pizza Place	Supermarket	Bar	Italian Restaurant	Grocery Store	Gym	Pub	Tapas Restaurant
13	95,907	-3.636776	1	Pizza Place	Spanish Restaurant	Tapas Restaurant	Park	Grocery Store	Café	Plaza	Bar	Bakery	Gym / Fitness Center
17	114,832	-3.621516	1	Grocery Store	Gym	Spanish Restaurant	Fast Food Restaurant	Restaurant	Supermarket	Bar	Bakery	Tapas Restaurant	Soccer Field
18	74,235	-3.606052	1	Exhibit	Pizza Place	Grocery Store	Spanish Restaurant	Café	Tapas Restaurant	Park	Fast Food Restaurant	Gym / Fitness Center	Restaurant

Cluster 3:

	Population (1 Jan 2020)	Longitude	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
11	143,365	-3.711469	2	Park	Fast Food Restaurant	Beer Garden	Accessories Store	Athletics & Sports	Pet Store	Café	Diner	Museum	Sporting Goods Shop

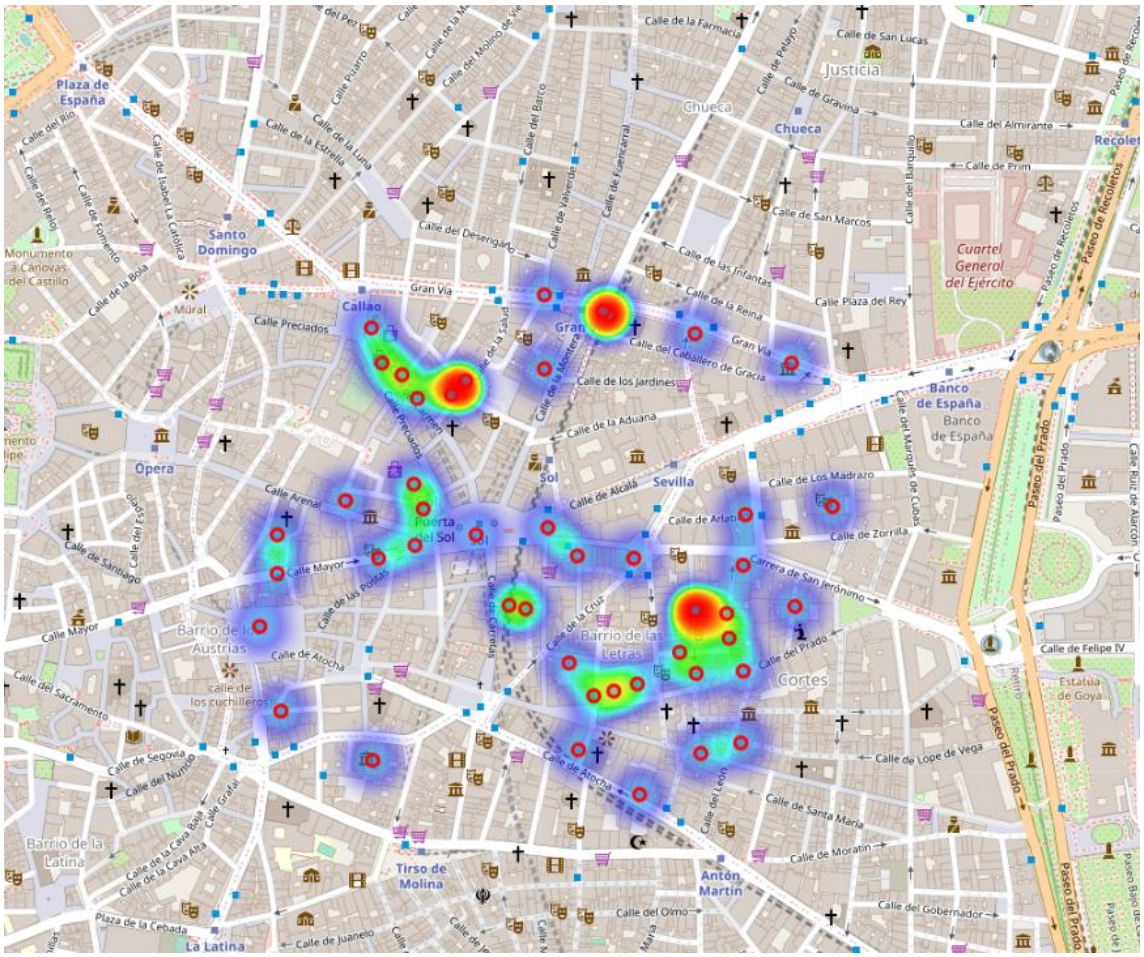
See map:



4.7 **What are the restaurants in the center of Madrid?**

The restaurants between the different districts of Madrid have been analyzed, and it has been decided to establish the business in the center of Madrid.
For this, the central restaurants have been searched and grouped to show them on the map in their corresponding location.

We can see it below:



5. Conclusion

After the complete analysis of the districts of Madrid, seeing all the relevant information that has helped us make the decision to establish our restaurant, it has been concluded that our establishment will be located in the central district of Madrid.

This has been taken based on different factors:

1. We believe that the center is an area where many people pass every day and since we are going to open a seafood restaurant, there are not many places in Madrid where we can eat a good seafood platter, so we believe that it is a good choice.

2. We know that the center of Madrid is an expensive area that will surely take a lot of work and effort at the beginning but we are hopeful that it will turn out well and we hope that it will be profitable.

Below we can see where the restaurant will be located.