CAPSTONE PROJECT:

THE BATTLE OF NEIGHBORHOODS
MADRID (SPAIN)

ALICIA MARTINEZ-ZAPORTA GIMENO

05-05-2021

TABLE OF CONTENTS:

1.- Introduction - Business Problem

2.- Data

2.1.- Data cleaning

3.- Methodology and analysis

3.1.- K-means: Clusters

3.2.- Distribution of venues

3.3.- Location of Japanese Restaurants

3.4.- Distribution of restaurants per neighborhood

3.5.- Distribution of other types of venues per neighborhood

4.- Results

5.- Discussion

6.- Conclusion

Capstone Project - The Battle of Neighborhoods - Madrid (Spain)

1.- Introduction - Business Problem

In this project we will try to find the best location for a Japanese restaurant in the city center of Madrid, Spain.

Madrid is not only a touristic city but also the center of many businesses with their employees having lunch every day near their offices. The city center is the crowdest area but also has many restaurants established there already.

The locations that we will analyze are the ones located near "Puerta del sol" which is considered the center point of the city. Those locations will have as potential customers not only the tourists, but also the people working nearby.

We will be interested in the location that doesn't have any Japanese restaurant (if any) and also those with less restaurants in general. We will also study other venues located in those areas, as it can be as well a decisive point to choose one area over another. If there are a good variety of venues, it might attract more people to the area rather than an area with a lack of them.

Taking all this information into account, advantages of each area will be explained so the stakeholders can choose the most suitable area for their restaurant.

2.- Data

We will use Foursquare to get the data necessary to analyze the areas.

- Are there other Japanese restaurants in the area?
- What kind of restaurants apart from Japanese are around the area?
- Are there leisure activities to attract people in that area?

We will display the information into a map to visually understand the data. In order to do that, we will extract from Wikipedia the list of boroughs and neighborhoods of Madrid. We will also download from github the coordinates of the boundaries of the boroughs and neighborhoods in Madrid.

2.1.- Data cleaning

From Wikipedia we extracted a list of boroughs and neighborhoods of Madrid. However, other columns as surface, density of population or images of the area were also included. We discarded these columns as were redundant for the purpose of the study.

As the study is trying to find the best location near Puerta del sol, we decided to continue only with the borough that contains that location, and study the neighborhoods within.

We also used geopy to find and add the latitude and longitude of the neighborhoods. We combined everything in the following table:

	Borough	Neighborhoods	Latitude	Longitude
0	Centro	Palacio	40.415129	-3.715618
1	Centro	Embajadores	40.409681	-3.701644
2	Centro	Cortes	40.414348	-3.698525
3	Centro	Justicia	40.423957	-3.695747
4	Centro	Universidad	40.425426	-3.706023
5	Centro	Sol	40.416947	-3.703489

From Foursquare we extracted the venues located per neighborhood. As we want to know the number of restaurants per area, we divided the data in 2 different data frames to continue the study: One containing the restaurants and the other one with the remaining types of venues.

After a further analysis of the remaining venues, we got a table of the top 10 most common venues of each area:

Neighborhood	1st	2nd	3rd	4th	5th	6th	7th	8th e	9th	10th
Cortes	Plaza	Bar	Hotel	Café	Theater	Book store	Pizza Place	Deli / Bodega	Coffee Shop	Cocktail Bar
Embajadores	Bar	Café	Plaza	Hostel	Art Gallery	Book store	Coffee Shop	Pizza Place	Hotel	Theater
Justicia	Bakery	Book store	Hotel	Deli / Bodega	Flower Shop	Plaza	Bar	Bistro	Café	Lounge
Palacio	Plaza	Bar	Historic Site	Music Venue	Beer Bar	Church	Pizza Place	Ice Cream Shop	Performing Arts Venue	Garden
Sol	Plaza	Hotel	Gourmet Shop	Hostel	Book store	Cosmetic Shop	Coffee Shop	Cocktail Bar	Clothing Store	Pastry Shop
Universidad	Café	Bar	Book store	Coffee Shop	Plaza	Cocktail Bar	Theater	Hotel	Miscellan. Shop	Bakery

From the data we also identified the number of Japanese restaurants depending of the type of food they serve: Japanese restaurant (general Japanese food), sushi restaurant and ramen restaurant:

Neighborhood	Japanese Restaurant	Sushi Restaurant	Ramen Restaurant
Cortes	0	2	0
Embajadores	0	2	0
Justicia	1	1	0
Palacio	1	0	1
Sol	0	0	1
Universidad	1	0	0

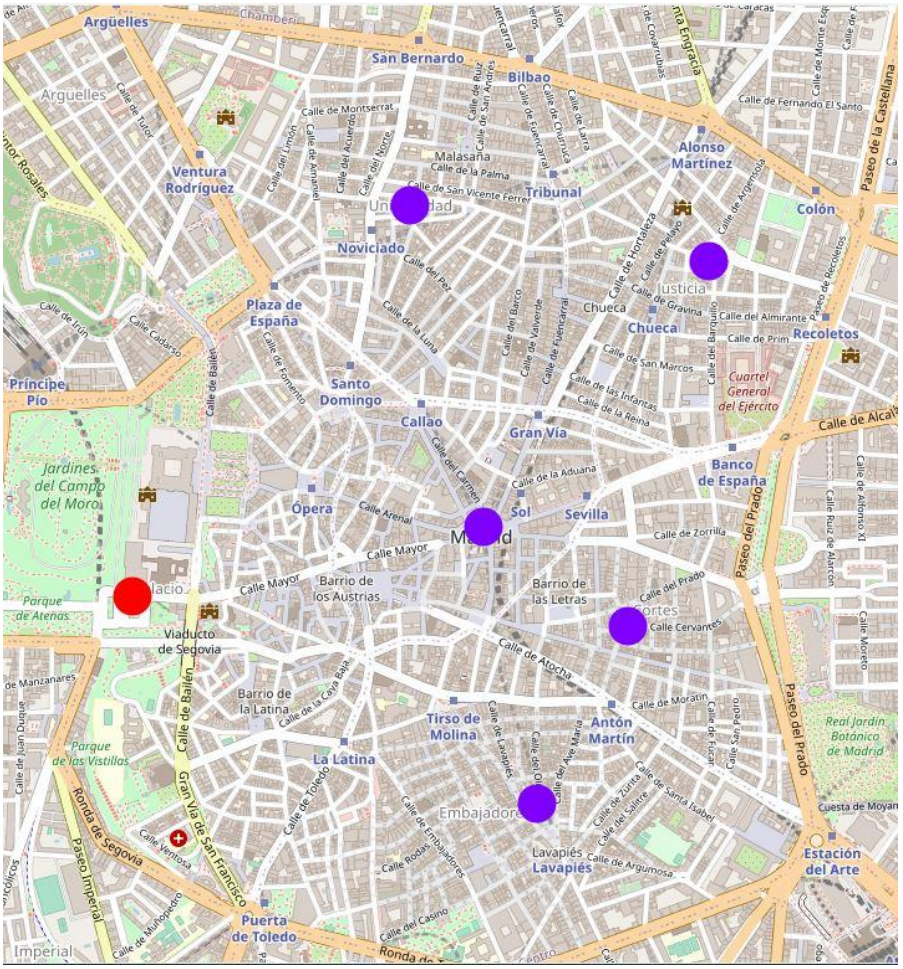
3.- Methodology and analysis

In the first step we have collected and prepared all necessary data for our project. Type and location of restaurants and other venues in the areas of study, the borough of Centre and the neighborhoods in it. We have also identified among the restaurants the ones serving Japanese food.

In the second step we have studied the different neighborhoods. We have used the K-means algorithm as it's one of the most used clustering algorithms due to its simplicity.

3.1.- K-means: Clusters

In order to get an idea of the structure of the data, we created subgroups according to its similarity. The data analyzed has been the venues without restaurants and using the elbow method, we have clustered the data into 2 subgroups.

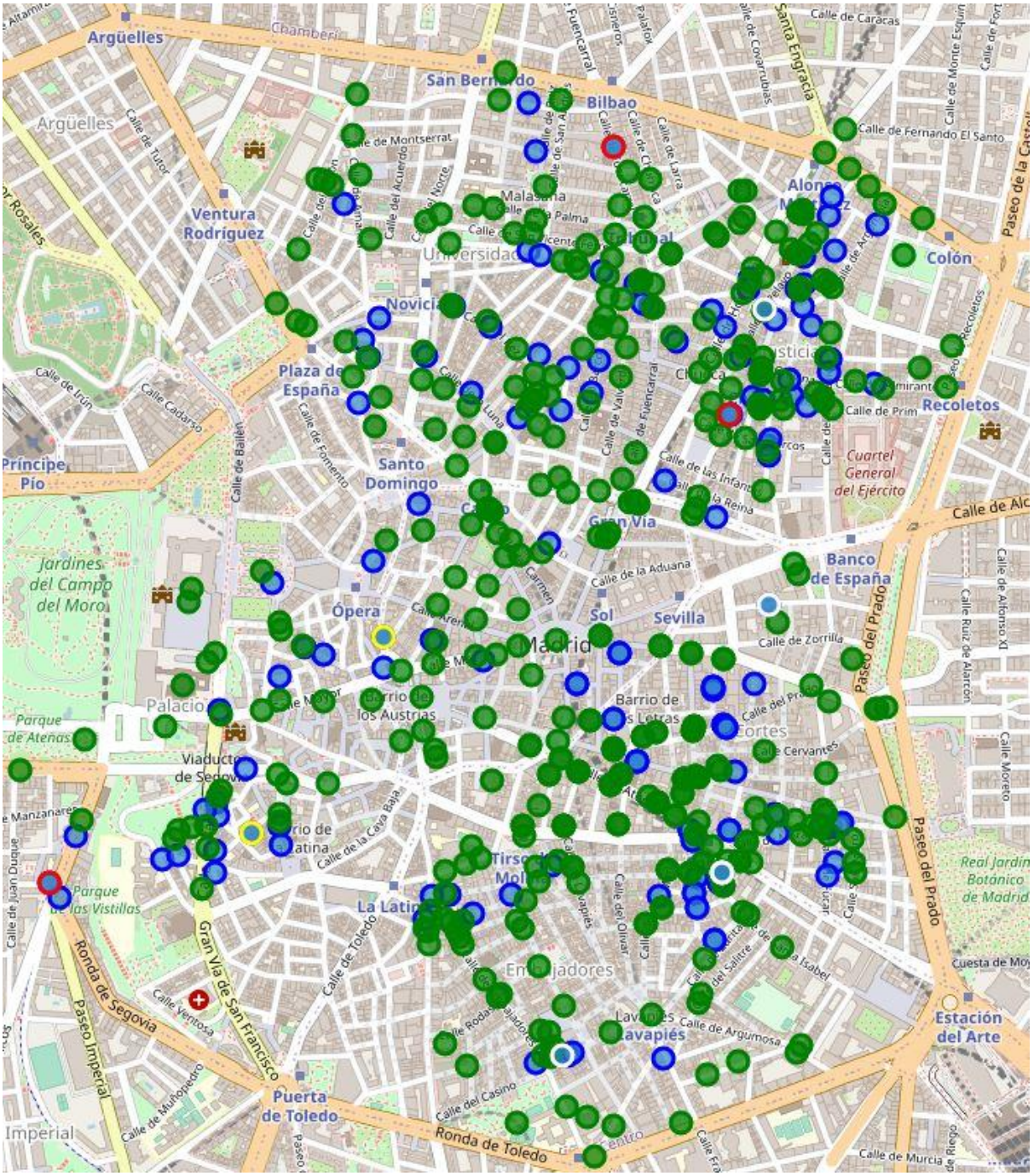


As we can see, five neighborhoods have similar data and Palacio is the dissimilar one. This helps us to discard Palacio as a good location for the restaurant.

3.2.- Distribution of venues

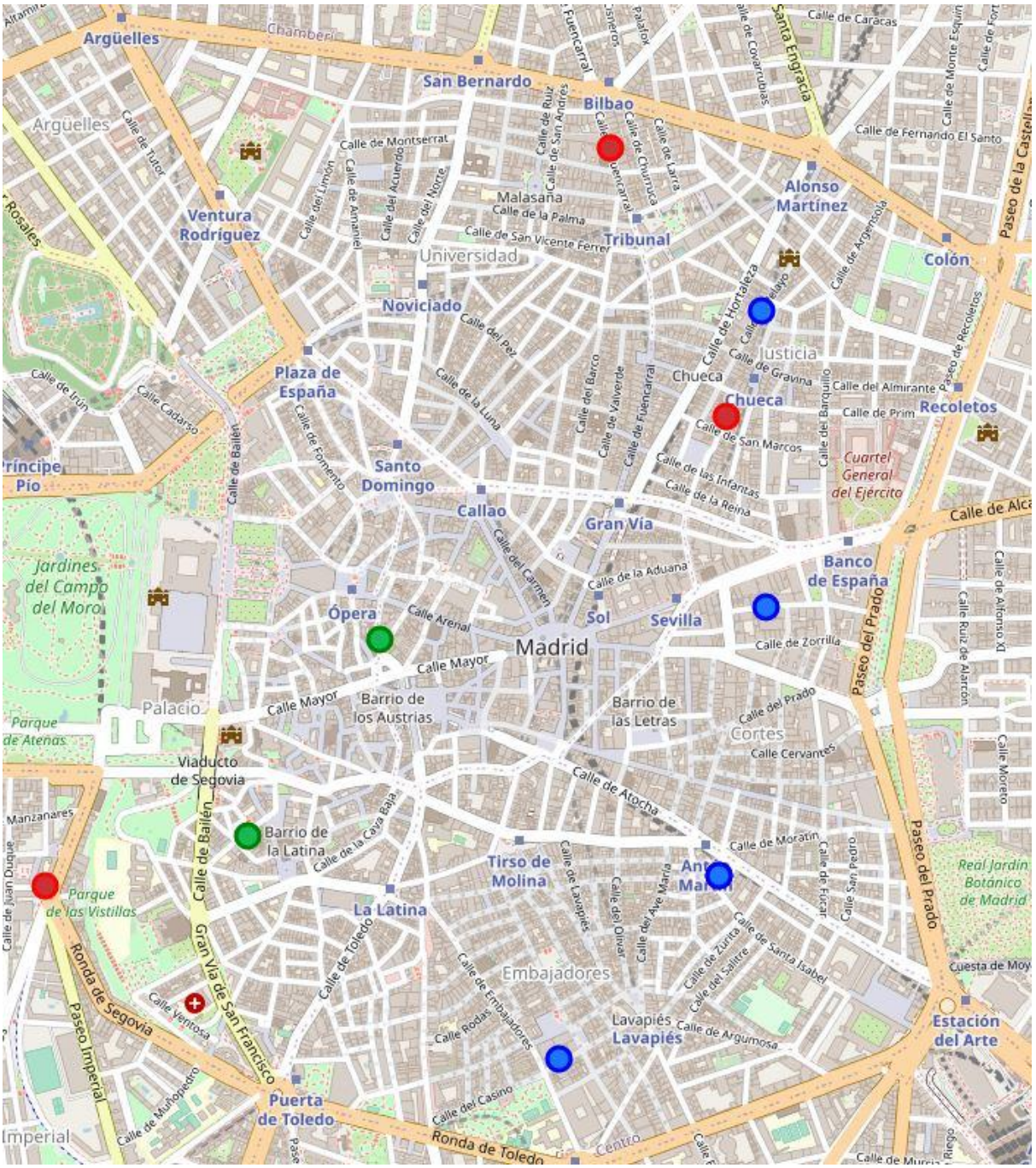
We show now a map with the restaurants in blue, the different venues in green and the Japanese restaurants in different colours to be able to differentiate them (Japanese: red, Sushi: white, Ramen: yellow).

With this map we can also have an idea about the amount of restaurants per neighborhood.



3.3.- Location of Japanese Restaurants

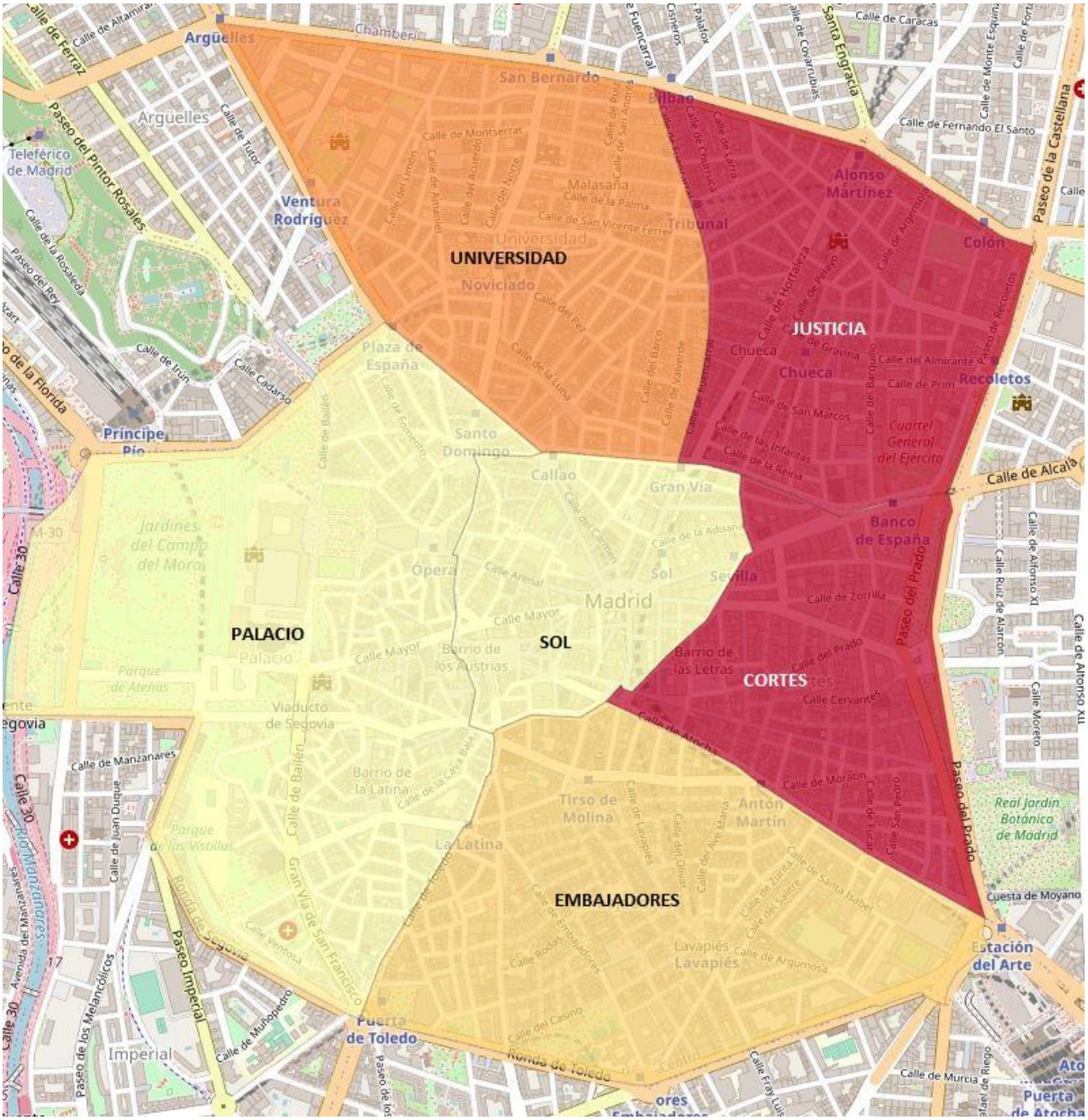
In order to facilitate the location of the Japanese restaurants, we plot them in a different map using different colours to represent the different types of restaurants: Japanese: red, sushi: blue, ramen: green.



We can see from this map that sushi restaurants are the most popular ones in this borough.

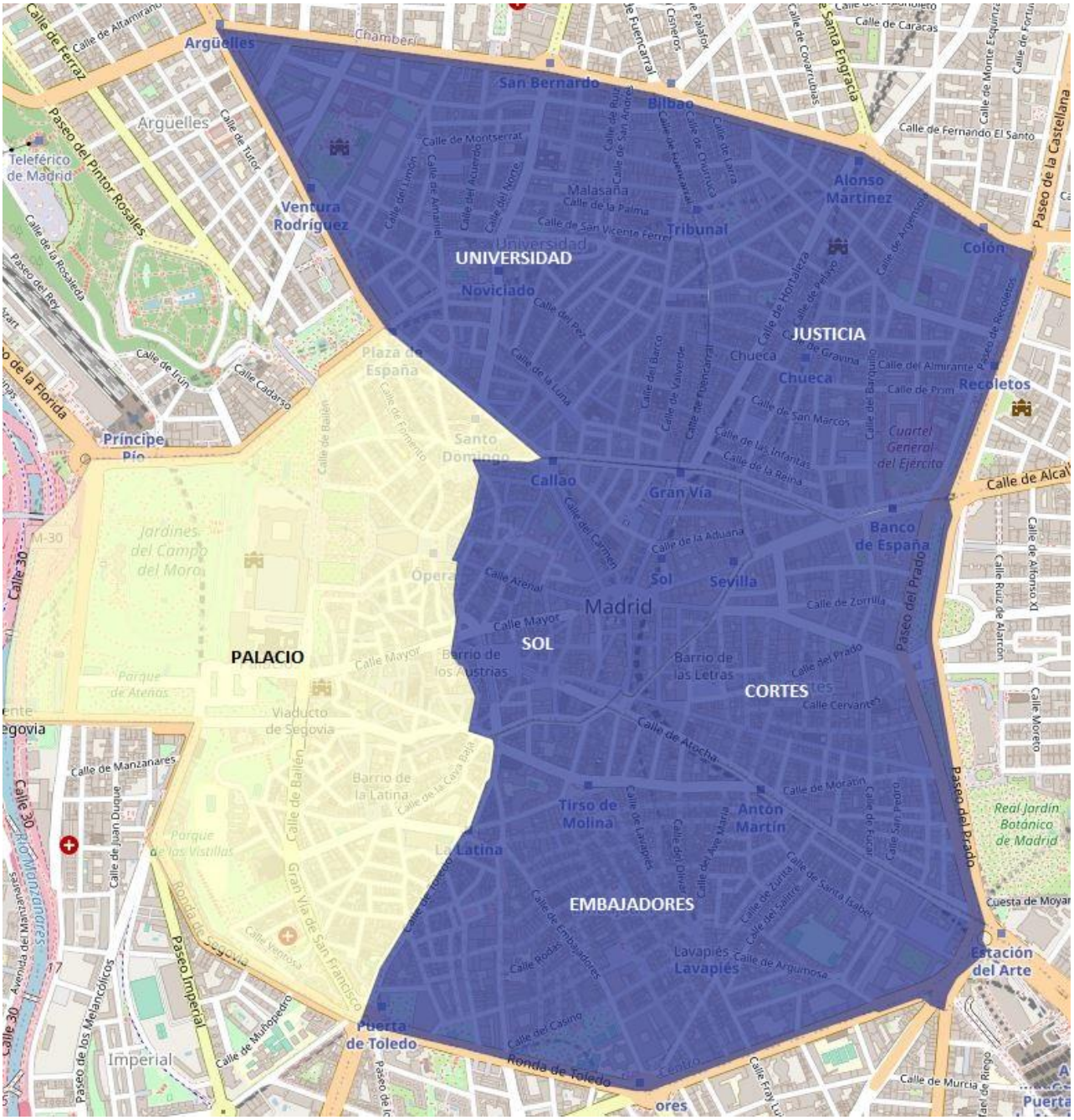
3.4.- Distribution of restaurants per neighborhood

In this choropleth map, we colorize the neighborhoods according to their number of restaurants. The darker is the colour, the more restaurants we find in the area.



3.5.- Distribution of other types of venues per neighborhood

In the same way as before, we plot a choropleth map in which we colorize the neighborhoods according to their number of venues. The darker is the colour, the more venues we find in the area.



4.- Results

- From the K-means clustering we see that the neighborhood of "Palacio" is the dissimilar one in terms of venues we can find in the area. This is also supported for both of the choropleth maps, which show that there are few restaurants and barely venues in the area compared to the other ones. Apart from this, we realise that there are already 2 restaurants located there, a Japanese restaurant and a ramen restaurant.
- "Sol" is the neighborhood with a lower amount of restaurants but it has a high amount of venues. It also has 1 Ramen restaurant in the area.
- "Embajadores" has a medium amount of restaurants and the amount of venues is high. This neighborhood has 2 sushi restaurants.
- "Universidad" has a medium amount of restaurants and the amount of venues is high. This neighborhood has 1 Japanese restaurant.
- "Justicia" has a high amount of both restaurants and venues. It also has 1 Japanese restaurant and 1 sushi restaurant.
- "Cortes" has a high amount of both restaurants and venues. It also has 2 sushi restaurants in the area.

5.- Discussion

Observing the results, we can confirm that "Palacio" is the less appealing area to open the new restaurant. There aren't enough venues that might attract people to the area and also already has 2 Japanese food restaurants.

The other 5 districts have a high amount of venues, so from this point of view any of them could be a good option. However, taking into account the number of restaurants in the area, "Justicia" and "Cortes" wouldn't be a good choice as the areas are already full of restaurants. On top of this, both neighborhoods have already 2 Japanese restaurants located there.

The final options are "Sol" and "Embajadores". Although it's true that "Sol" has fewer restaurants in the area, the difference is not that big to make a difference and discard "Embajadores" (17 vs 21). Regarding the Japanese restaurants located already in the area, "Embajadores" has 2 sushi restaurants, which could be enough to attract people to those restaurants instead of the new one, but a Japanese restaurant can offer a big variety of food, including sushi, so this reason could make a difference among the potential customers to be the chosen one. The neighborhood of "Sol" has only 1 ramen restaurant in the area, so a Japanese restaurant could get the potential customers as the variety will be wider than only one type of meal that it's in this case the ramen.

As a final consideration, "Sol" is the neighborhood in which "Puerta del sol" is located, and being around this area was one of the main points of this study.

6.- Conclusion

The purpose of this project was to identify potential neighborhoods to open a Japanese restaurant near Puerta del sol in Madrid, Spain, taking also into account other restaurants in the area, including Japanese restaurants, and other types of venues located in those areas. Using Foursquare data we could identify the amount of restaurants and other venues per neighborhood and also clustering the areas to verify the similarity between them.

Although this study has identified as potential neighborhoods "Sol" and "Embajadores", this is only a starting point for the stakeholders. They might take into account other factors and considerations which can make these options not optimal, like prices for renting the local or other social and economic factors.