# A Japanese Restaurant in Madrid

Margarida Santos

December 19, 2021

## 1 Introduction/Business Problem

In this project we will try to find possible locations for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Japanese restaurant in Madrid, Spain.

Since there are lots of restaurants in Madrid we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Japanese restaurants in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## 2 Data

Based on definition of our problem, the factors that will influence our decision are:

- number of existing restaurants in the neighboorhood (any type of restaurant)
- number of and distance to Japanese restaurants in the neighboorhood, if any
- distance of neighboorhood to city center

The following data sources will be needed to extract/generate the required information:

- coordinates of Madrid center will be obtained using Geopy of well known Madrid location (Kilómetro Cero - Puerta del Sol, 7, 28013 Madrid, Spain)
- location and limits of each district will be obtained using geopy and an open-source json file
- number of restaurants and their type and location in every neighboorhood will be obtained using Foursquare API

Table 1: Districts of Madrid and their average coordinates

| District | Latitude | Longitude |
|---|---|---|
| Centro | 40.418309 | -3.704902 |
| Arganzuela | 40.401842 | -3.698664 |
| Retiro | 40.410488 | -3.677238 |
| Salamanca | 40.433693 | -3.671852 |
| Chamartin | 40.467394 | -3.675737 |
| Tetuan | 40.468377 | -3.701191 |
| Chamberi | 40.439211 | -3.706746 |
| Fuencarral-El Pardo | 40.557248 | -3.745399 |
| Moncloa-Aravaca | 40.460135 | -3.766995 |
| Latina | 40.392334 | -3.777401 |
| Carabanchel | 40.388535 | -3.731706 |
| Usera | 40.375043 | -3.694493 |
| Puente de Vallecas | 40.379683 | -3.670539 |
| Moratalaz | 40.408704 | -3.641184 |
| Ciudad Lineal | 40.447280 | -3.649623 |
| Hortaleza | 40.494441 | -3.629036 |
| Villaverde | 40.337174 | -3.675403 |
| Villa de Vallecas | 40.345142 | -3.633995 |
| Vicalvaro | 40.395885 | -3.554856 |
| San Blas | 40.437099 | -3.587188 |
| Barajas | 40.481393 | -3.568474 |

The first thing we should do is create a map of the region of Madrid where to place all the information needed.

We can start by determining the coordinates of the well known Madrid location, Kilómetro Cero at Puerta del Sol, 7, 28013 Madrid, Spain using Geopy library. These coordinates, [40.4163, -3.7034], will represent the center of Madrid, which will be marked on the future map.

Next step, is finding the existent districts of Madrid and their respective limits. Fortunately, there are some open source resources which we can use, namely a json file at `https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/madrid-districts.geojson`. Table 1 show the list of the districts in Madrid and their "central" coordinates[1].

For finding the optimal location, first we need to create area candidates. As such, we will create a grid of area candidates, equally spaced, centered around city center and within $\approx 3\ km$ from Kilometro Cero. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

[1]It is just the averages of the coordinates latitude and longitude, finding a location inside the district to place a marker when the map is done

This process results in 90 area candidates, which will also be shown in our map.

Figure 1 shows the map of Madrid centered on Kilometro Cero, with all the districts limited in blue and marked, the area candidates grid in yellow and radius of 1km, 2km and 3km around Kilometro Cero in red.

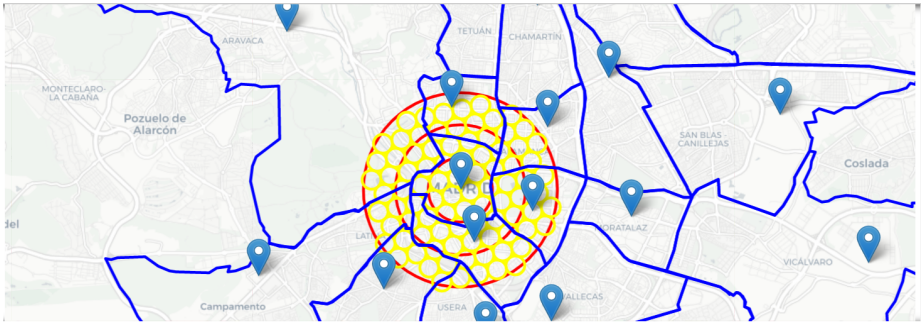


Figure 1: Map of Madrid with districts identified and first area candidates.

Still in the phase of gathering data, we need to use the Foursquare API to get info on restaurants in each district. We're interested in venues in "food" category, but only those that are proper restaurants - coffe shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have "restaurant" in category name, and we'll make sure to detect and include all the subcategories of specific "Japanese restaurant" category, as we need info on Japanese restaurants in the district.

All the information will be organise on two dataframes, one concerning all the restaurants and one with only the Japanese restaurants. However, given the length of these dataframes they are not shown here.

The locations of all the restaurants are shown on Figure 2 in green and the Japanese restaurants in red.

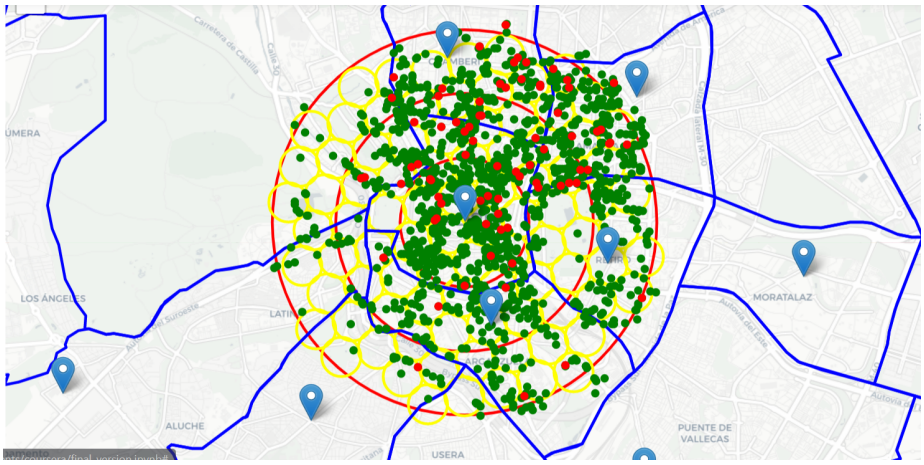


Figure 2: Map of Madrid with restaurants found in green and Japanese restaurants in red.

This concludes the data phase, which includes its gathering and treatment.

# 3 Methodology

We intend to select the best location or locations for a Japanese restaurant in Madrid.

For this, we want locations which fulfill as best as possible the following conditions:

- have few restaurants around the candidate location;
- have few Japanese restaurants around the candidate location;
- are close to the center of the city,

In this project we will direct our efforts on detecting areas of Madrid that have low restaurant density, particularly those with low number of Japanese restaurants. We will limit our analysis to area $\approx$ 2 $km$ around city center.

In first step we have collected the required data - location and type (category) of every restaurant within 3km from Madrid center (Kilometro Cero). We have also identified Japanese restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Madrid - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Japanese restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Japanese restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## 3.1 Analysis

With the data gathered we can observe that in the region of interest, exist 1475 restaurants, of which 77 are Japanese restaurants, i.e, 5.22 %. Also, each neighborhood have an average of 11.33 restaurants.

Also, measuring the distance of the Japanese restaurants found using Foursquare API, we can see that the average distance of the closest Japanese restaurant to each area center is 481.18 meters.

To better understand the best candidates, we need to analyse the density of restaurants and Japanese restaurants at different locations. For that, we can use the heatmaps shown in Figures 3 and 4.

In Figure 3 we can see that within 2km to the center of the city, that there are some pockets in Centro, Arganzuela, Retiro, Latina and Moncloa-Aravaca.

In Figure 4, we can conclude that there are not many Japanese restaurants in Madrid, which is great news for our stakeholders. The districts observed in the last heatmap are still good options.

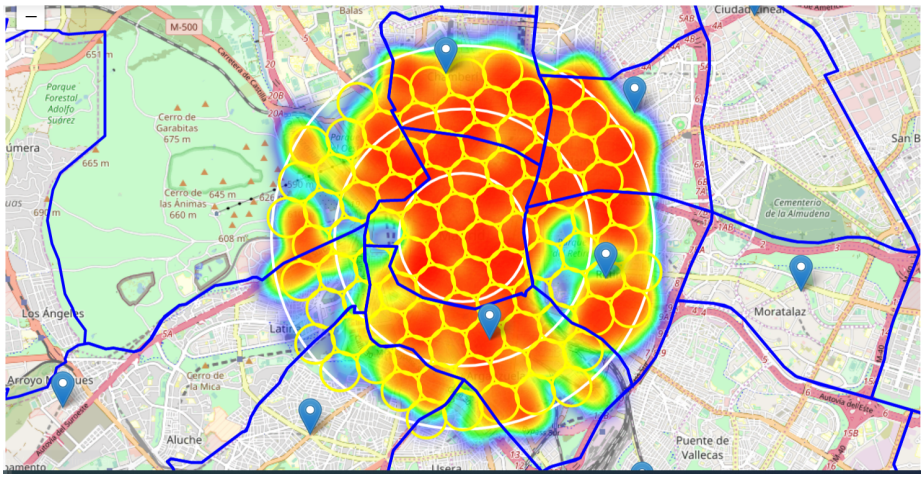


Figure 3: Heatmap showing the density of restaurants in Madrid.


Figure 4: Heatmap showing the density of Japanese restaurants in Madrid.

As such, we have four districts which can be eligible candidates, Centro, Arganzuela, Retiro, Latina and Moncloa-Aravaca. Given that the Centro, Arganzuela and Retiro have closest pockets to the center, let's consider only these ones.

## 3.2 Centro, Arganzuela and Retiro

Given that these seem to be best districts, we can reduce the analysis to within 2km to the center of city. Also, create smaller candidate areas, only 100 meters apart. This creates 1453 candidate areas.

Getting back to our objective, we can filter location with no more than two restaurants in radius of 250 meters, and no Japanese restaurants in radius of 400 meters. This results in:

- 121 locations with no more than two restaurants within 250 meters
- 343 locations with no Japanese restaurants within 400 meters
- 108 locations with no more than two restaurants within 250 meters and no Japanese restaurants within 400 meters

The 108 locations that meet both conditions are represented as green circles on Figure 5, as expected in "'cold" regions of the heatmap.

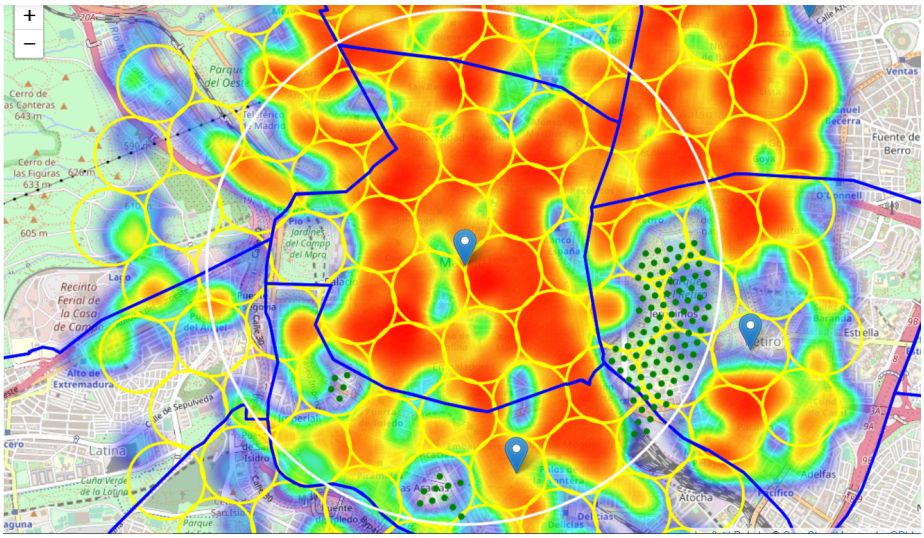


Figure 5: Good locations for Japanese restaurants as green circles in a heatmap of Madrid.

We have now a clear indication of zones with low number of restaurants in vicinity, and no Japanese restaurants at all nearby. We can see that this are in Arganzuela and Retiro.

Finally, we must cluster those locations and create centers of zones containing good locations. Given that, observing Figure 5, it seems that we do not have many good regions, we will only try to find 5 clusters.

The clusters found are shown in Figure 6 in black. Using Geopy, we can find the addresses of these locations. The following are the addresses found and respective distances to Kilometro Cero:

- Povets, Albocàsser, l'Alt Maestrat, Castelló / Castellón, Comunitat Valenciana - 1.7km from Kilometro Cero
- Albocàsser, l'Alt Maestrat, Castelló / Castellón, Comunitat Valenciana - 1.8km from Kilometro Cero
- Barranc d'en Cabrera, Albocàsser, l'Alt Maestrat, Castelló / Castellón, Comunitat Valenciana - 1.7km from Kilometro Cero
- Barranc d'en Cabrera, Albocàsser, l'Alt Maestrat, Castelló / Castellón, Comunitat Valenciana - 1.4km from Kilometro Cero
- Barranc d'en Cabrera, Albocàsser, l'Alt Maestrat, Castelló / Castellón, Comunitat Valenciana - 1.8km from Kilometro Cero

As we can see, three of them are the same address. This is expected since in Figure 6, three clusters in Retiro overlap. Meaning that, we have found three different addresses to propose to our stakeholders.

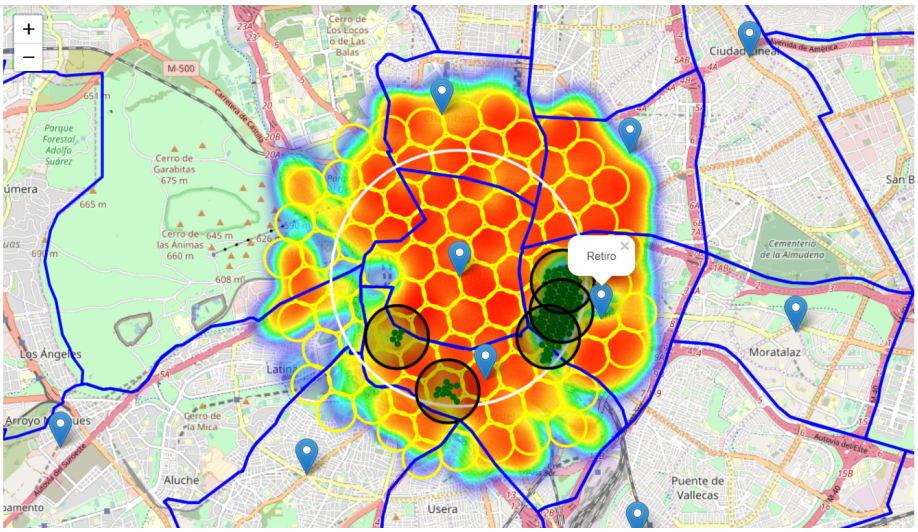


Figure 6: Clusters of good locations for Japanese restaurants in black in a heatmap of Madrid.

This concludes our analysis. We have created 3 addresses representing centers of zones containing locations with low number of restaurants and no Japanese restaurants nearby, all zones being fairly close to city center (all less than 2km from Kilometro Zero). Although zones are shown on the map as circles, their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. The zones found are located in Arganzuela and Retiro districts, which are very close to city center.

# 4 Results and Discussion

Our analysis shows that although there are a great number of restaurants in Madrid ($\approx$ 1500 within 3km from Kilometro Cero), there are pockets of low restaurant density fairly close to city center. From the analysis of heatmaps, we decided to focus on the districts Centro, Retiro and Arganzuela, since these had a lesser concentration of restaurants.

After directing our attention to this more narrow area of interest (within 2km from Kilometro Cero) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Japanese restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 3 zones in Arganzuela and Retiro (we found that Centro was not a good option) containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants

in general and Japanese restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Madrid center but not crowded with existing restaurants (particularly Japanese) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

# 5 Conclusion

Purpose of this project was to identify Madrid areas close to center with low number of restaurants (particularly Japanese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general districts that justify further analysis (Centro, Retiro and Arganzuela), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders. All of these points belong to Retiro and Arganzuela.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of districts and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.