

# Japanese Restaurant in Madrid. Where and why?

## Business Problem

In this project we will try to find an optimal location 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 Spain 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.

## Data

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

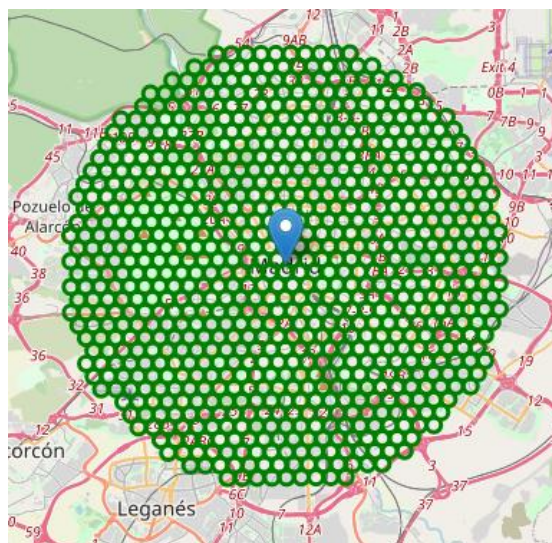
- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Italian restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of Madrid center will be obtained using Google Maps API geocoding of well known Madrid location (Puerta del Sol)

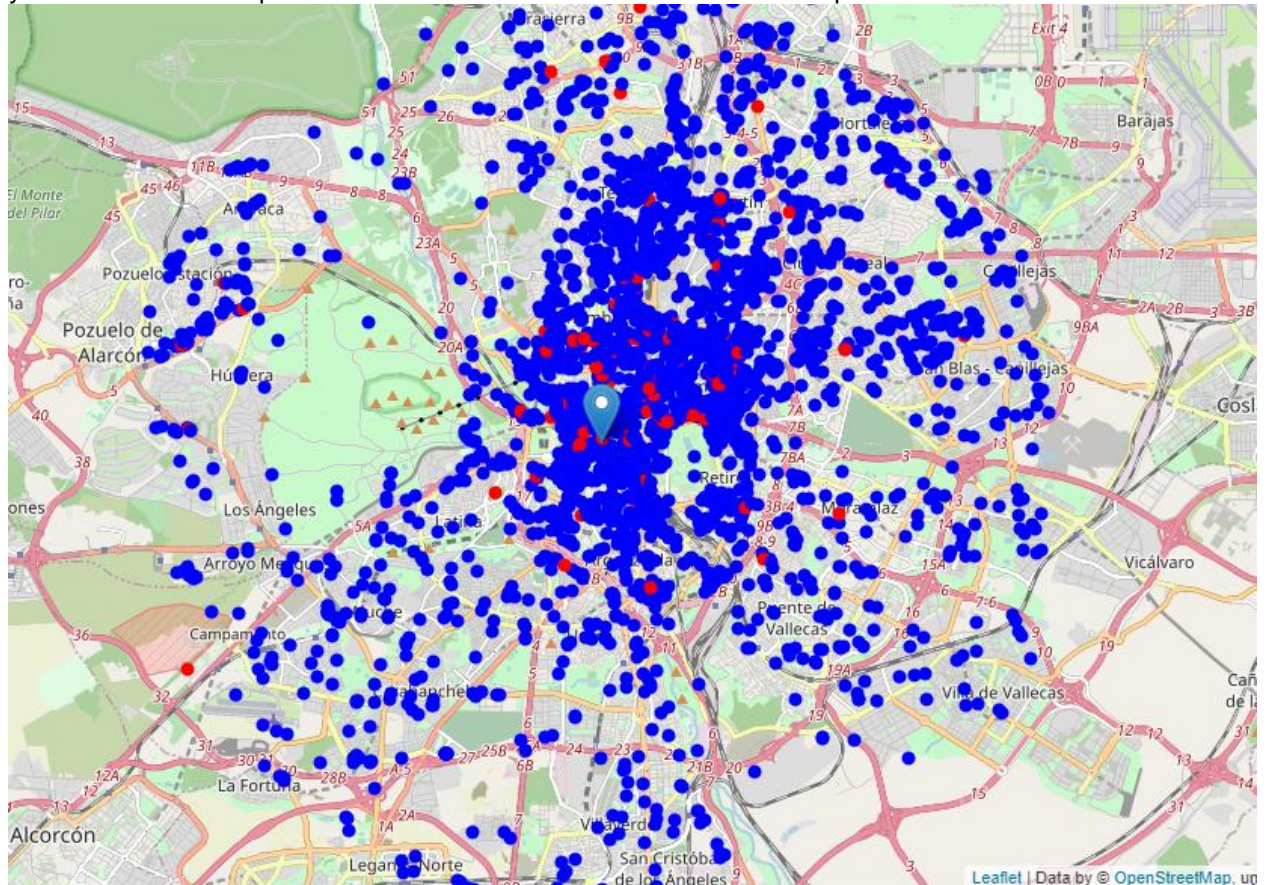
We create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approx. 18x18 kilometers centered around Madrid city center. We will get approximately a full city area coverage. We first find the latitude & longitude of Madrid city center, using specific, well known address and Google Maps geocoding API.



**Fig.1** Cell grid for further analysis. Marker is located at Madrid center (Puerta del Sol)

After we have our location candidates, we use Foursquare API to get info on restaurants in each cell

We're only interested in restaurants, so we filter it from all other venues like bakeries, etc. In the next figure you can see location plot of all Madrid restaurants as blue dots and Japanese restaurants as red dots.



**Fig.2** Restaurants on Madrid map (blue circles – all restaurants, red circles - Japanese restaurants) (9 km radius from center)

## Methodology

In first step we have collected the required **data: location and type (category) of every restaurant within 9km from Madrid center** (Puerta del Sol). 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.

## Analysis

We perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of restaurants in every area candidate**:

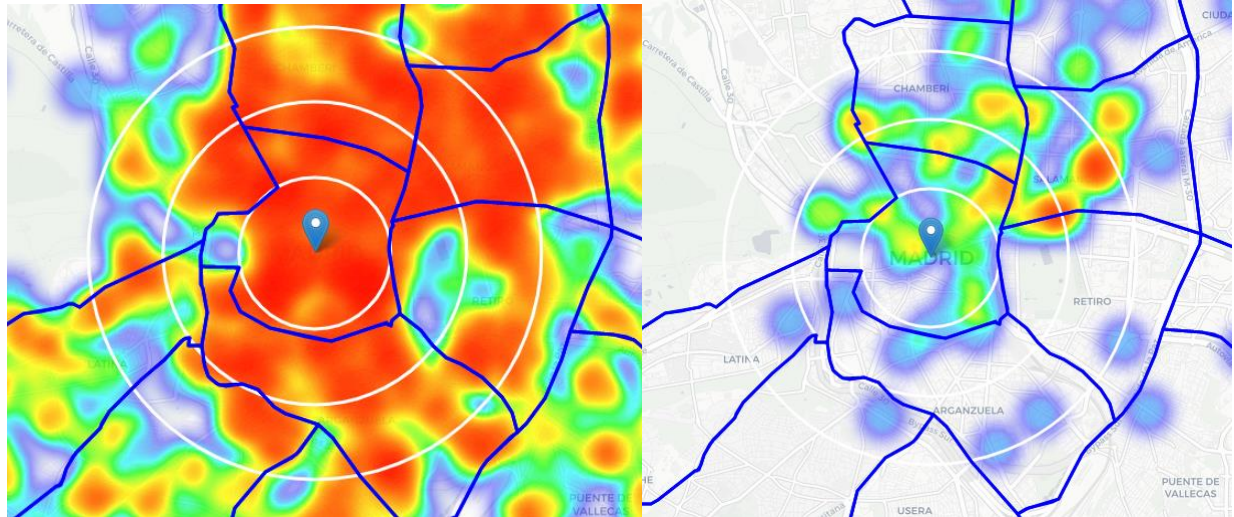
- Average number of restaurants in every area with radius=300m: **3.5**

Now let's calculate the **distance to nearest Japanese restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).

- Average distance to closest Japanese restaurant from each area center: **1745 meters**



So on average Japanese restaurant can be found within ~2km from every area center candidate.



**Fig. 3.** All restaurants heatmap (left) and only Japanese restaurants heatmap (right)

We can indicate higher density of existing Japanese restaurants directly north and east from Puerta del Sol, and absence of the such type of restaurants at south and south-west.

Based on this we will now focus our analysis on areas *south-west, south from Madrid center* - we will move the center of our area of interest and reduce its size to have a radius of **1.5km**. This places our location candidates mostly in boroughs **Arganzuela, Carabanchel and Usera**

### **Arganzuela, Carabanchel and Usera**

These boroughs are located not so far from city center. Some words about these districts:

**Arganzuela:** 'Located just south of the centre on the banks of Madrid's Manzanares River, Arganzuela is the best of both worlds: an easy walk into central Madrid and far enough away to have a local atmosphere and none of the crowds. It is also home to some great attractions including Madrid Río park, a huge renovation of the river banks that was completed in 2011. It includes play parks, kiosks and terraces, football pitches and lots of space to walk, cycle or rollerblade. Another key sight is the Matadero, Madrid's former slaughterhouse that is now a thriving cultural space with regular exhibitions, markets and its own cinema.'

**Usera:** 'Known as Madrid's Chinatown, Usera is home to much of the city's Chinese community and is – unsurprisingly – where the best and most authentic Chinese restaurants can be found. It is also becoming one of Madrid's most trendy areas for its green spaces and reasonable rents. Located just south of the River Manzanares, a new riverside shopping centre, Plaza Río 2 has also helped attract more attention to the area. In 2017, Airbnb named the area one of the “17 neighbourhoods to watch in 2017”. Usera is also home to the Manzanares Linear Park, a riverside park with a manmade hill topped by the impressive sculpture La Dama del Manzanares, by Valencian artist Manolo Valdés.'

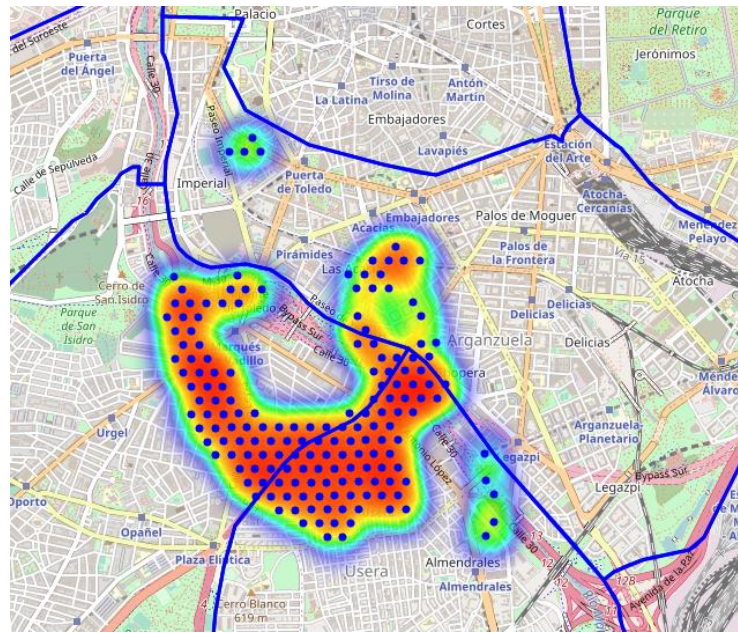
**Carabanchel:** 'Another neighbourhood just to the south of the River Manzanares, Carabanchel was first mentioned in historical documents in the 12th century. The area is known for its green spaces, especially San Isidro Park, the epicentre of the week-long Festival of San Isidro in May, dedicated to Madrid's patron saint.'

Material from site <https://theculturetrip.com/europe/spain/articles/a-guide-to-madrids-most-up-and-coming-neighbourhoods/>

Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Japanese restaurant**.

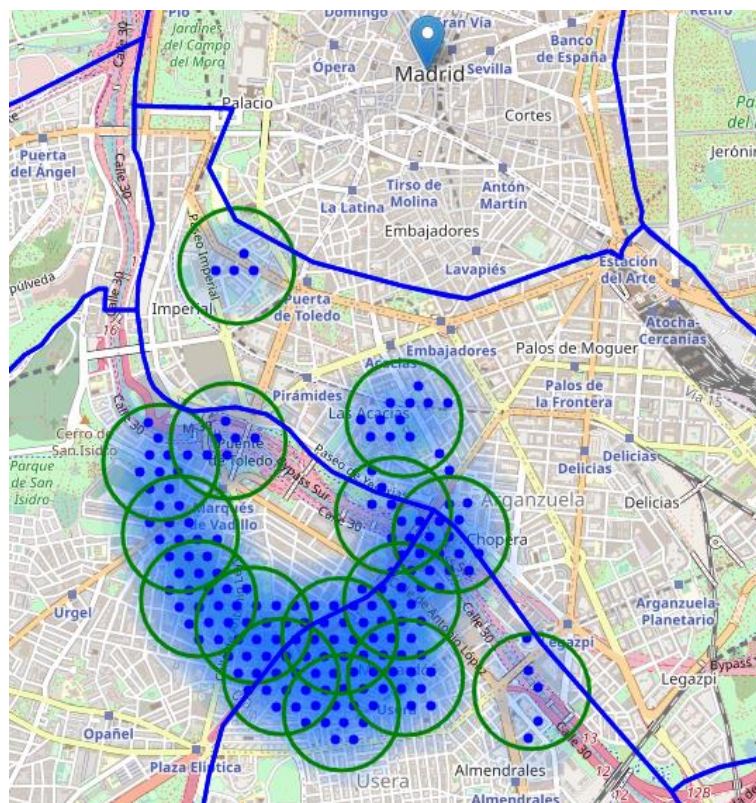
- Locations with no more than two restaurants nearby: 187
- Locations with no Japanese restaurants within 400m: 419
- Locations with both conditions met: 171
-

We now have a bunch of locations fairly close to Puerta del Sol, and we know that each of those locations has no more than two restaurants in radius of 250m, and no Japanese restaurant closer than 400m.



**Fig.4.** Heatmap of possible locations for Japanese restaurants

Let us now **cluster** those locations to create **centers of zones containing good locations** using k-means machine learning algorithm. Those zones, their centers and addresses will be the final result of our analysis.



**Fig.5.** K-Means output as 15 clusters.

### **Addresses of centers of areas recommended for further analysis**

- Calle de la Verdad, 20, 28019 Madrid      => 2.9km from Puerta del Sol
- Calle Torero, 6, 28026 Madrid      => 2.7km from Puerta del Sol
- Paseo de la Esperanza, 21, 28005 Madrid      => 1.8km from Puerta del Sol
- Calle del Dr. Carmena Ruiz, 1, 28026 Madrid      => 2.9km from Puerta del Sol
- Autopista de Circunvalación M-30, 28019 Madrid      => 2.1km from Puerta del Sol
- Calle de Gil Imón, 1, 28005 Madrid      => 1.4km from Puerta del Sol
- Pasarela de la Princesa, Unnamed Road, 28026 Madrid      => 3.2km from Puerta del Sol
- Calle San Nicomedes, 4, 28026 Madrid      => 3.2km from Puerta del Sol
- Paseo del Quince de Mayo, 28, 28019 Madrid      => 2.4km from Puerta del Sol
- Paseo de la Chopera, 2, 28045 Madrid      => 2.4km from Puerta del Sol
- Calle Mirasierra, 32, 28026 Madrid      => 3.1km from Puerta del Sol
- Calle de Jacinto Verdaguer, 36, 28019 Madrid      => 2.7km from Puerta del Sol
- Calle Quince de Agosto, 2, 28026 Madrid      => 3.3km from Puerta del Sol
- Los Pino, 28005 Madrid      => 2.3km from Puerta del Sol
- Calle Antonio Corpas, 6, 28019 Madrid      => 3.0km from Puerta del Sol
- 

This concludes our analysis. We have created 15 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 ~3km from Puerta del Sol, and about half of those less than 2km from Alexanderplatz). Most of zones are located near borders of three boroughs (Arganzuela, Carabanchel and Usera).

### **Results and Discussion**

Our analysis shows that although there is a great number of restaurants in Madrid (~3000 in our initial area of interest which was 18x18km around Puerta del Sol), there are absolutely no pockets of low restaurant density fairly close to city center. Our attention was brought by near absence of Japanese restaurant south to the center.

After directing our attention to this more narrow area of interest (covering approx. 3x3km south from Puerta del Sol) we first created a dense grid of location candidates (spaced 100m apart); 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 15 zones 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! Some of addresses are near the highways - I'm sure this is not the best place for a restaurant.

### **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 boroughs that justify further analysis (Arganzuela, Usera and Carabanchel), 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.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods 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.