Analysis for new Restaurant in Lima-Perú

Carlos Gutiérrez

July 05, 2020

1. Introduction

Perú is a country located in western South America and well known by its ancient history as well as for its tourist places. Also the Peruvian food is considered as one of most important in the world due to its diversity and special taste, however there are also many restaurants from different countries which in some way compete in a business that is growing year by year and has become very competitive. Lima is the capital of Perú and a very crowded city with almost 10 million people living there and grouping almost the 40% of restaurants of the country. So in order to enter in the business of restaurants is important to analyze some aspects before jump into it, for example the location to define in which neighborhood would be good to place a new restaurant, also what kind of food are going to prepare?, would be a good idea to place a Peruvian restaurants or a foreign one?. We will try to answer these questions in order to define what kind and where a new restaurant could be placed.

2. Data

To solve the problem defined in section 1, we will only analyze most popular neighborhoods, these places correspond to Lima Center and is composed by 14 neighborhoods whose information (names, latitude, longitude, etc.) will be gathered from Wikipedia. The data downloaded contain 43 neighborhoods so we will clear the data to only have the 14 neighborhoods to analyze. From our data the most popular neighborhoods are “Miraflores” and “San Isidro”, the first one because of their venues like bars and restaurants and for the presence of many tourist, the second one is better known as a business neighborhood but also with many well-known restaurants.

We will use Geopy as a geocoding tool to get latitude and longitude for each selected neighborhood. One problem with this service was that “Miraflores” neighborhood was located far from the center causing the data collected not cover places where many restaurants are, to fix this we will move the location of “Miraflores” closer to center. The venues of each neighborhoods will be gathered using Foursquare API with a limit of 100 venues by neighborhood and a radius of 2 km. Since the distance between each neighborhood are not constant is inevitable that many venues will be duplicated, so after the clean of duplicated venues we have 903 venues in total and 170 unique categories.

Table 1: Neighborhoods venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	San Miguel	-12.078656	-77.095283	Do it!	-12.081012	-77.082683	Accessories Store
1	Lima	-12.062107	-77.036526	Tony Roma's	-12.057406	-77.030081	American Restaurant
2	Lima	-12.062107	-77.036526	Chili's	-12.066808	-77.047318	American Restaurant
3	San Miguel	-12.078656	-77.095283	T.G.I. Fridays	-12.076010	-77.084234	American Restaurant
4	San Miguel	-12.078656	-77.095283	T.G.I. Friday's	-12.078503	-77.087624	American Restaurant

Table 2: Neighborhoods information

	Neighborhood	Latitude	Longitude
0	Lima	-12.062107	-77.036526
1	Barranco	-12.143959	-77.020268
2	Breña	-12.059700	-77.050119
3	Jesús María	-12.078186	-77.046412
4	La Victoria	-12.073358	-77.016417
5	Lince	-12.086568	-77.036647
6	Magdalena del Mar	-12.095654	-77.068213
7	Pueblo Libre	-12.076639	-77.067858
8	Miraflores	-12.118814	-77.040097
9	San Borja	-12.096452	-76.995690
10	San Isidro	-12.097902	-77.035367
11	San Miguel	-12.078656	-77.095283
12	Santiago de Surco	-12.125105	-76.981919
13	Surquillo	-12.114198	-77.010475

3. Methodology

The firsts analysis we performed is the cluster of all neighborhoods in order to get information about how similar are neighborhoods, later we will use it as part of our analysis to select in which neighborhood we could place a new restaurant. We modified the data frame ordered by the category of each venue there and taking the mean in order to generate values and used in clustering process, we define 5 clusters.

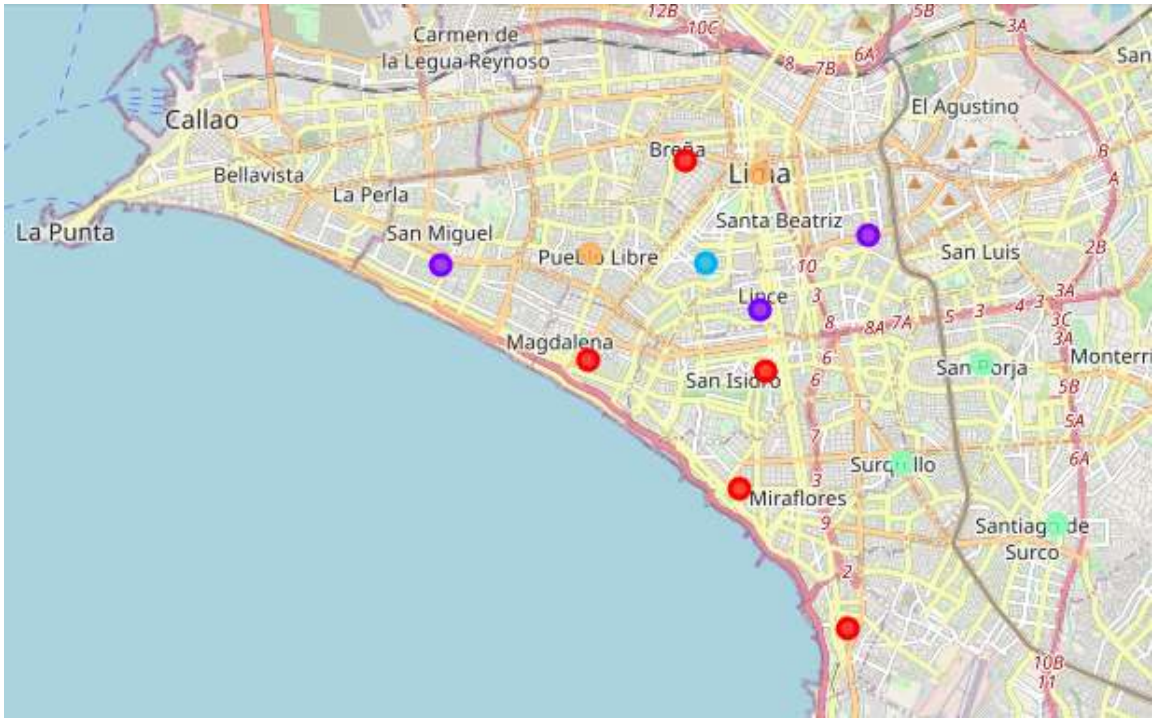

Figure 1: Neighborhoods Clustered

The cluster that have more neighborhoods is cluster 0 (red circles) which have 5 of 14 neighborhoods, so this places are similar in terms of venues, on the other hand the cluster 2 (blue circle) have only one neighborhood which could indicate that the most of venues there not necessary are the same of others neighborhoods.

We grouped the venues by category resulting the three main categories are:

- Park
- Seafood restaurant and
- Peruvian restaurant

Also the fourth grouped venue is labeled as “Restaurant”, this first result just confirm that the most popular venues in Lima are restaurants.

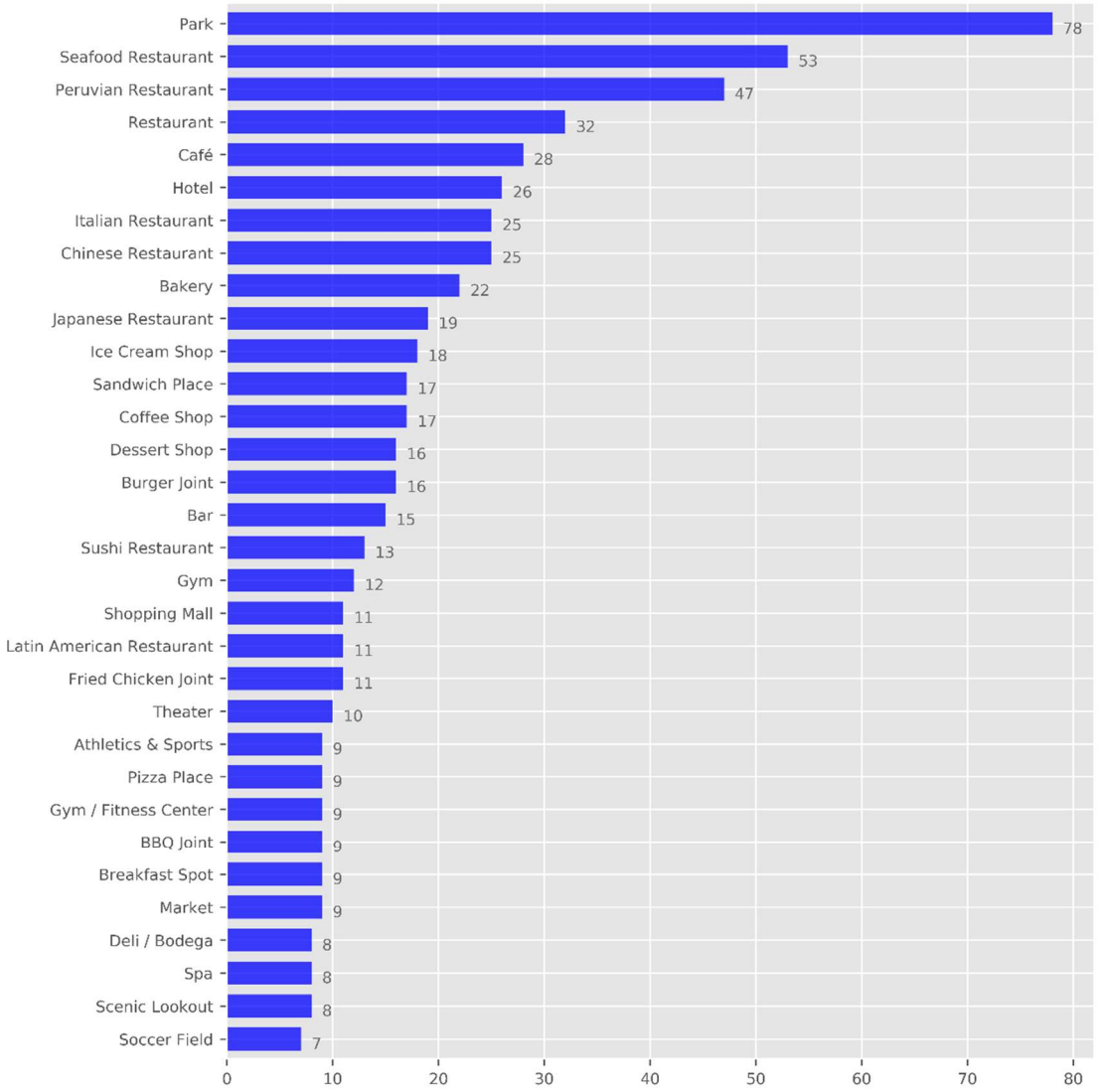

Figure 2: Lima Venues by Category

As we are looking for only restaurants then will drop all information not related to this category like, cafes gyms, etc. We will filter the data using “Restaurant” word.

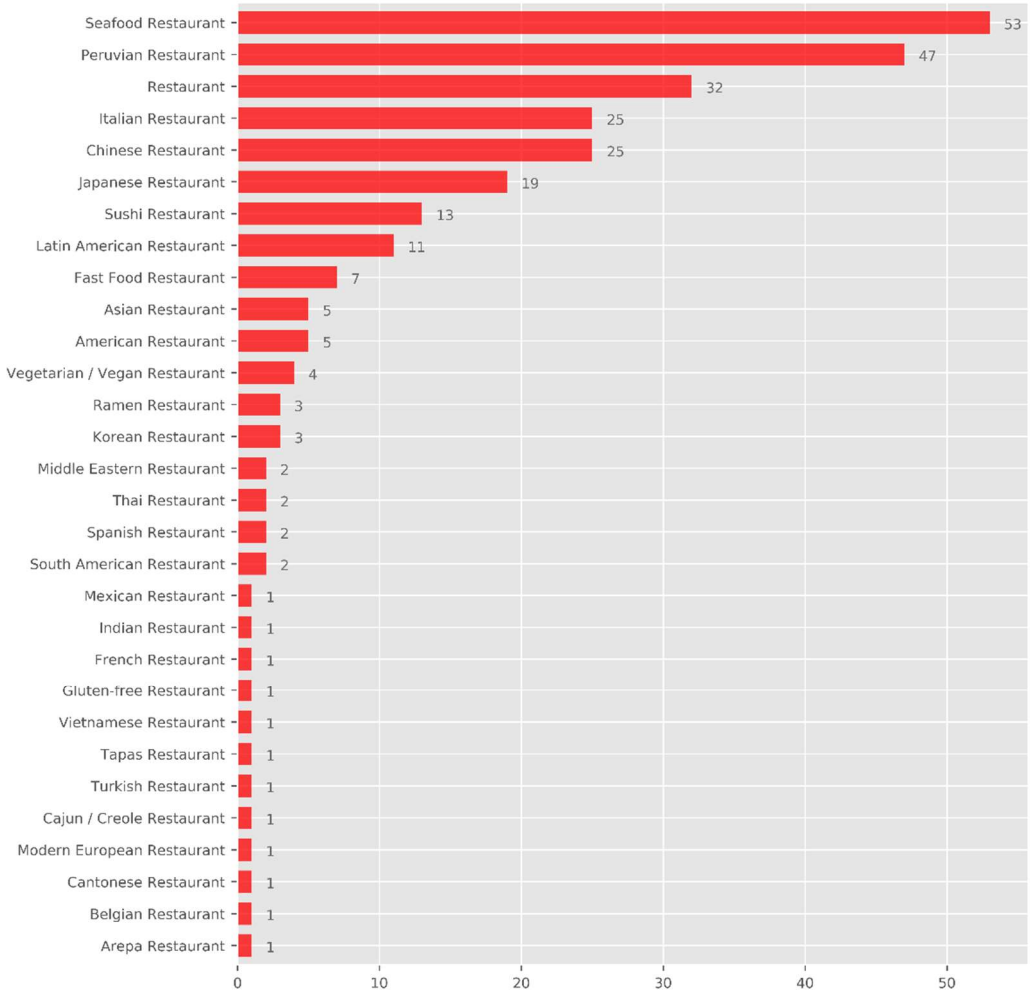

Figure 3: Restaurants in Lima

Now we are going to group the data by categories:

- 1° Category: This will group all "Peruvian Restaurant" also "Seafood Restaurant" since correspond also to Peruvian food. We will also include the "Restaurant" labeled data because from a review of the dishes correspond to Peruvian. Also after review the data of "South American Restaurant" and "Latin American Restaurant" we realize that belongs to Peruvian food.
- 2° Category: Italian
- 3° Category: Chinese
- 4° Category: Japanese, which will include the data labeled as "Japanese Restaurant", "Sushi Restaurant" and "Ramen Restaurants".
- 5° Category: Foreign what correspond to all the foreign restaurants, except Italian, Japanese and Chinese.
- 6° Category: Vegan, all vegan restaurants including "Gluten free Restaurant".
- 7° Category: Fast food.

We will drop the "Tapas Restaurant" since in this case is not really a restaurant is more a bar which offers snacks, also will drop the "Confort Food Restaurant" since is a combination of café

and different kinds of food and due to the dishes. Let's label "Cajun / Creole Restaurant" as "Peruvian Restaurants". From the Asian restaurants we will rename two of them as Chinese since correspond to this category. Finally we rename all the rest of restaurants as “Foreign Restaurants”.

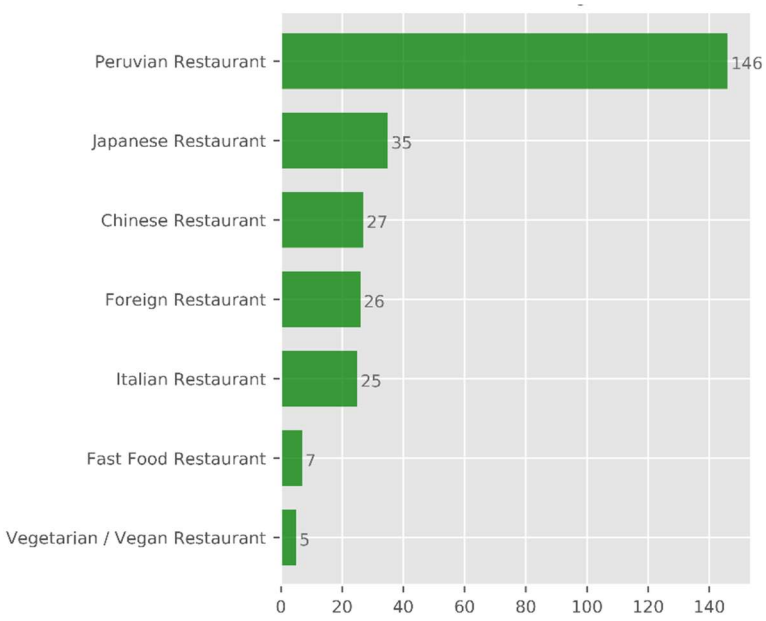

Figure 4: Restaurants by Category

From the chart we can see that there is no doubt that “Peruvian Food” is by far the most popular category having even more restaurants than the other categories together. But also is important to note that there are 4 foreign categories which have similar quantities of restaurants. We will draw this information in a Lima map by category in order to visualize the location of restaurants in each neighborhood.

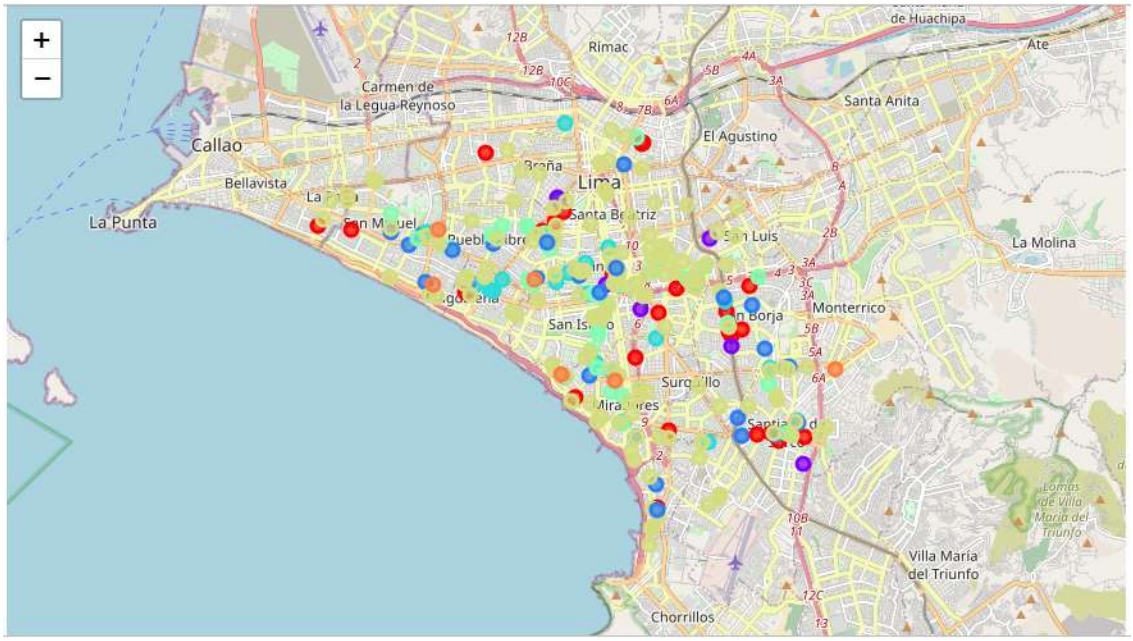

Figure 5: Restaurants in Lima by Category

We will focus only in the four main categories and modify the data frame to show how many restaurants we have by neighborhood.

Table 3: Number of Restaurants by Category

Venue Category	Neighborhood	Venue Latitude	Venue Longitude	Venue
Chinese Restaurant	27	27	27	27
Italian Restaurant	25	25	25	25
Japanese Restaurant	35	35	35	35
Peruvian Restaurant	146	146	146	146

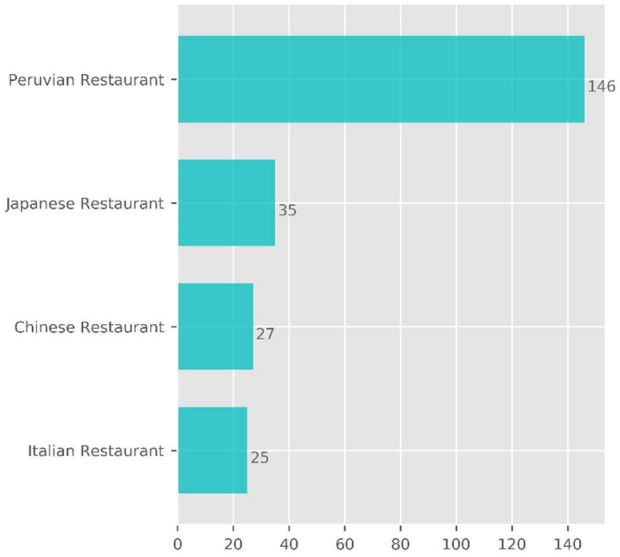

Figure 6: Restaurants in Lima (main categories)

We modify again the data in order to have the numbers of restaurants by category for each neighborhood.

Table 2: Number of Restaurants by Category for each neighborhood

Neighborhood	Peruvian Restaurant	Japanese Restaurant	Chinese Restaurant	Italian Restaurant
La Victoria	28	6	1	4
Barranco	14	0	2	0
San Miguel	13	2	3	0
San Isidro	12	5	2	2
Lince	11	1	1	1

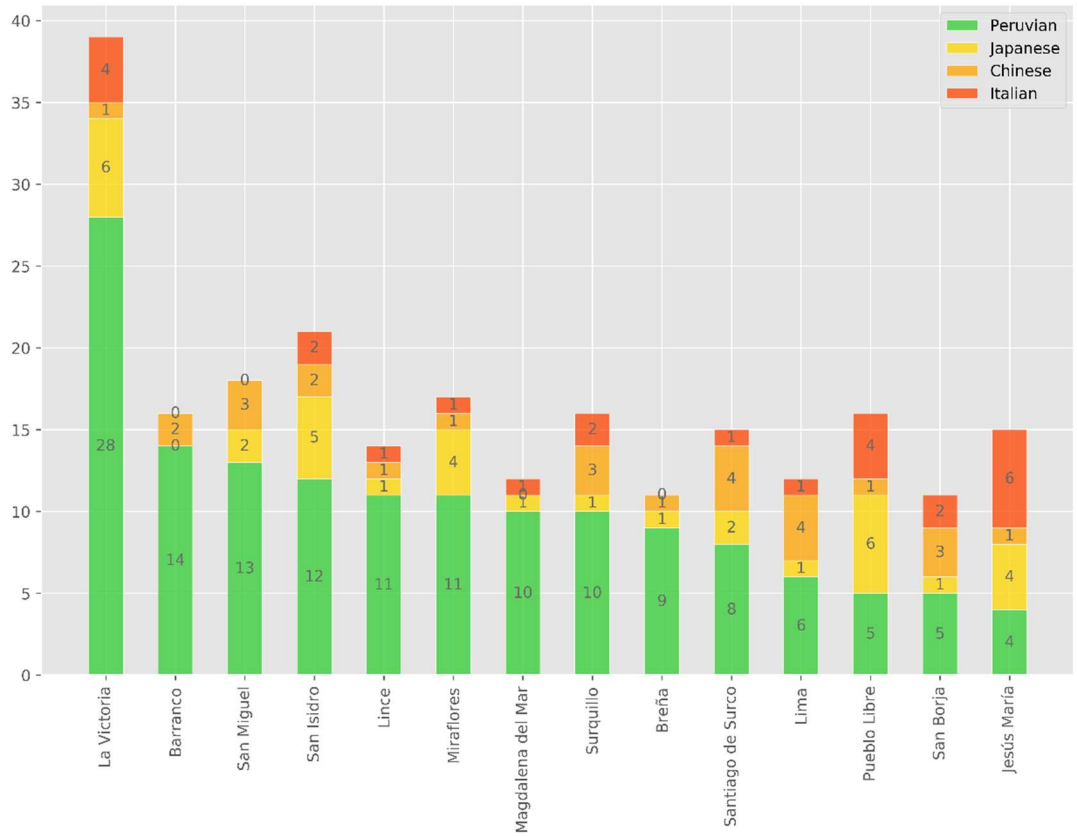

Figure 7: Restaurants by Category for each neighborhood

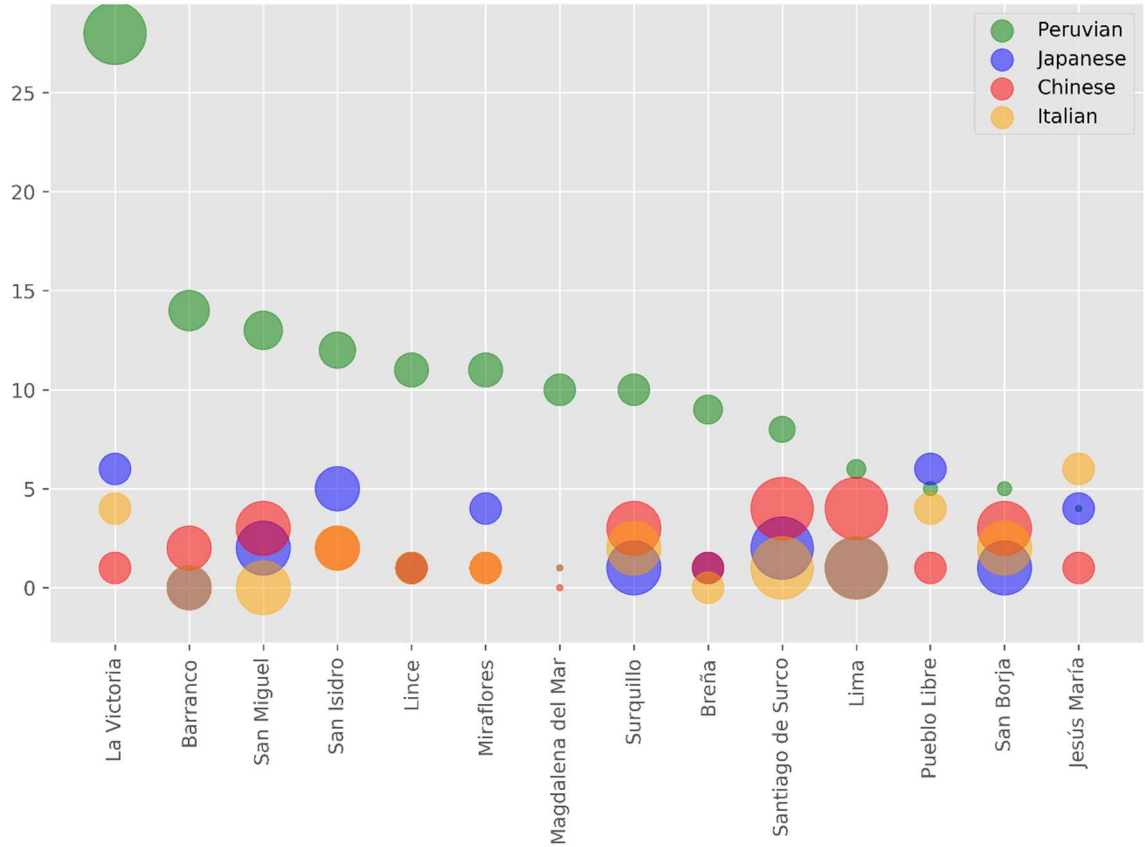

Figure 8: Restaurants by Category for each neighborhood

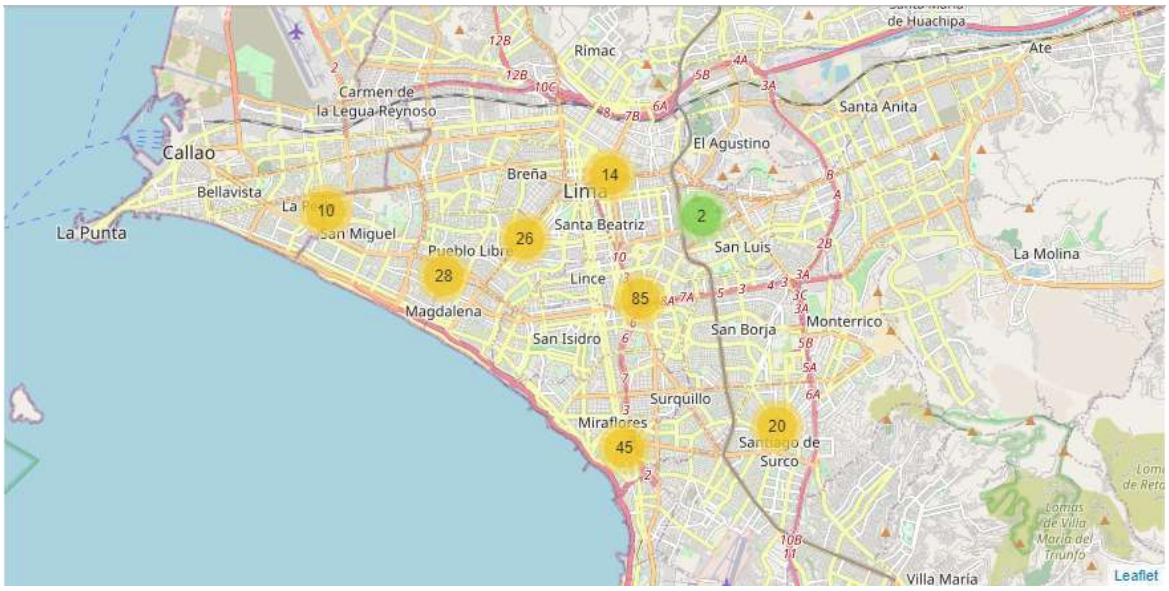

Figure 9: Restaurants by Category for each neighborhood

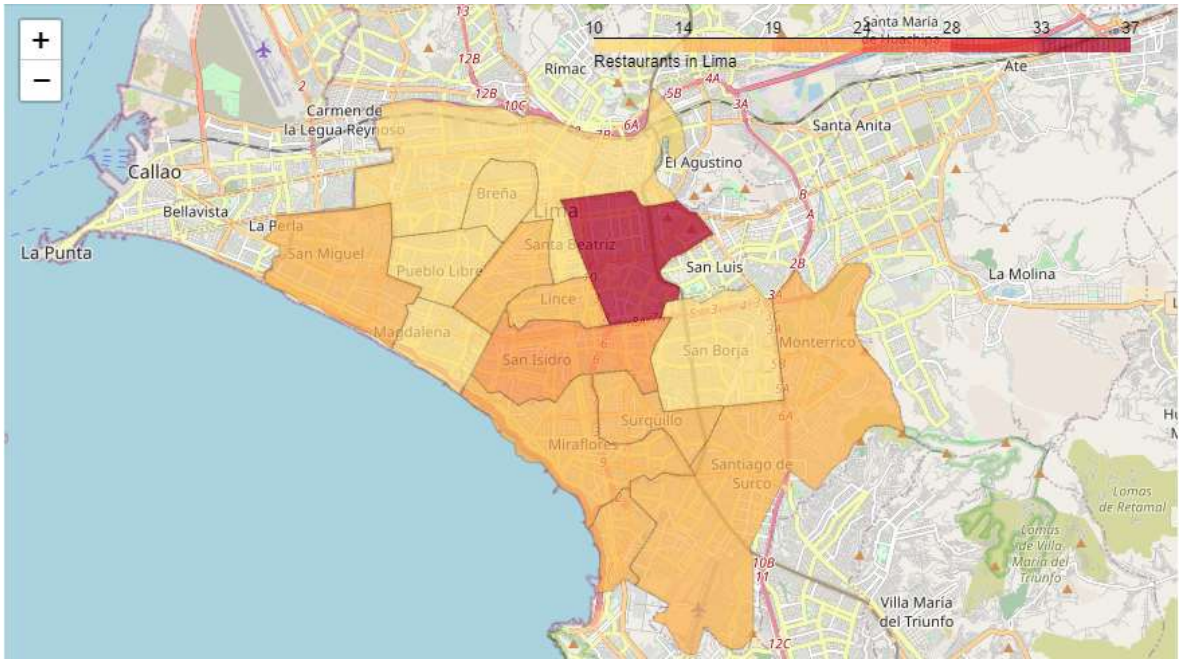

Figure 10: Restaurants by Category for each neighborhood

4. Results section where you discuss the results

From the previous charts we see that “La Victoria” is the neighborhood with most restaurants having the four categories but by far the “Peruvian Food” is most common, so maybe it wouldn't be a good idea to open a new restaurant in this neighborhood, at least not a Peruvian one. The others neighborhoods have a similar quantities of restaurants and most of them have the same tendency, we mean, the Peruvian Restaurant are the most common. Only in two neighborhoods (Pueblo Libre and Jesús María) the Peruvian restaurants are no the most common, in this two neighborhood prevails the Japanese and Italian food respectively, so it looks that there is room to new Peruvian restaurants.

Barranco does not have Japanese and Italian restaurants, so could be a good candidates for these restaurants. In Lince, Miraflores, Magdalena del Mar and Breña the Chines and Italian restaurants are minimum or there is not, so could be good places to this kind of food.

5. Discussion section where you discuss any observations

In this analysis we do not have in account the prices for the rent of restaurants room, which change a lot depending the neighborhood you decide to open a new restaurant. For example the most expensive rent correspond to two neighborhoods (Miraflores and San Isidro), so if you open a business here provably the prices of dishes will have to be above the average.

Regarding to data gathering from Foursquare we note that the data changed very slow depending of the day you gather, this also impact slowly in the results, so if you run the Jupiter Lab could have a small difference from the results presented here . Another issue with data gathering is that since the limit of venues was 100 some restaurants where not included, to tackle this issue we can run the gathering of data for each neighborhood using different location in the same neighborhood, of course many venues will be repeated but filtering we will have most restaurants for our analysis.

6. Conclusion

In this study, we analyzed where could be a good place to open a new restaurant in Lima Perú, identifying the quantities and categories in each neighborhoods we can define the possible place as well as the category. Through clustering of neighborhoods we can analyze their similarity from the point of view of venues which gave as insight to select the neighborhood. Also we verified that the Peruvian restaurants are the most popular by far and that one neighborhood has almost the double of restaurants than the average of the others neighborhoods.