1.- Introduction and Background:

Perú is a country known, among other things, for its cuisine, its biodiversity and for its varyiety of landscapes. These are some reasons why, despite being considered a third-world country, the number of tourist is growing year by year.

Lima, the Capital City is well-known for many reasons. Firstly, is the arrival place for the mayority of tourist who visit Peru even for work or pleasure time. Added to this, and because of its proximity to the sea, Lima has the honour of being the peruvian region which have one of the best *Ceviche*. *Ceviche* is probably the most recognized peruvian dish made mainly by fresh fish.

However, this is not the only iconic dish in Lima, because being the capital, is a mandatory checkpoint for the people travelling inside the country, causing that in this city we can find a massive variety of restaurants with food from every corner of Perú

This is an exploratory analysis approach which can be interesting for enterpreneurs seeking for open a restaurant (including peruvian seafood, andean food, coast food and so on) in one of the gastronomic capitals of the world.

2.- Data acquisition and use of it

This project will use two types of data:

1. **Localization of different hotels, hostels, motels and all tourist related info and restaurants in lima.**
2. **Localization of different lima neighborhoods in a grid.**

The first type of data will be obtained using the **FOURSQUARE API**, this API bring us the possibility of extract all the venues in a certain radius from a specific location (latitud and longitud), this feature is interesting for us because we well disgreggate our research area in a grid, so we can obtain specific information about venues for every cell on that grid.

The second type of data will by obtained using the **Google Maps API,** making a 1x1km grid, which extends for the most populated parts of the city. This data is important to determine the cells with the high density of lodges by km2. Using this information we could have an idea about the capacity of reception of tourist in every specific point in the área.

3.- Methodology

First, we obtained the information of the location of the **'Plaza de Armas de Lima'**, one of the iconical places located in the center of the city. After that we made a grid cell of 16x20km, including in this grid the mayority of Lima districts known for having more tourist affluence than the others. Every cell had a 500m radius, as shown in the Figure 1

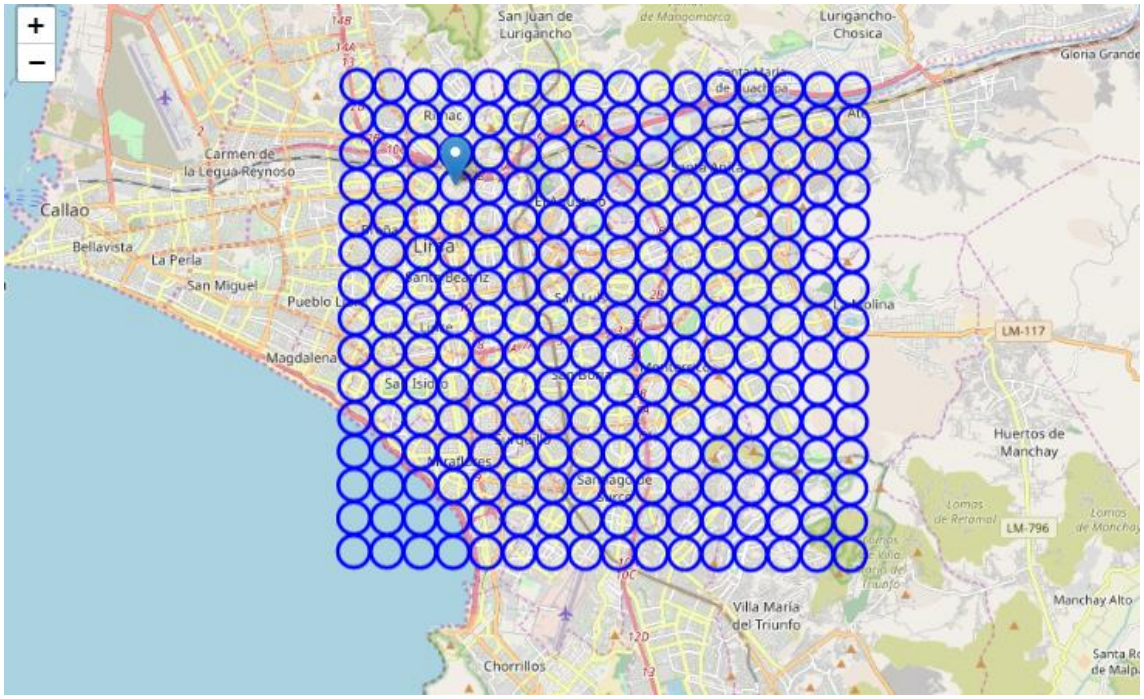


Figure 1

After making the grid, we acquire information about the venues of every grid. For this task, we use the "Foursquare API".

Once the data was collected, we select only the grids which had lodges (including hotels, hostels, motels) or places for information for tourists, as a measure of the capacity of receiving a large amount of tourists at the same time (tourists that come to Perú to have a taste of its cuisine, among other things).

The resulting table is shown below in the Table 1.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
36	6	6	6	6	6	6
51	6	6	6	6	6	6
50	5	5	5	5	5	5
35	4	4	4	4	4	4
80	3	3	3	3	3	3
100	3	3	3	3	3	3
81	3	3	3	3	3	3
67	2	2	2	2	2	2
85	2	2	2	2	2	2
4	2	2	2	2	2	2
20	2	2	2	2	2	2
54	1	1	1	1	1	1
97	1	1	1	1	1	1
178	1	1	1	1	1	1
162	1	1	1	1	1	1
146	1	1	1	1	1	1
138	1	1	1	1	1	1
99	1	1	1	1	1	1
90	1	1	1	1	1	1
66	1	1	1	1	1	1
86	1	1	1	1	1	1
27	1	1	1	1	1	1
82	1	1	1	1	1	1
37	1	1	1	1	1	1
19	1	1	1	1	1	1
52	1	1	1	1	1	1

Table 1

Determine the most common restaurants for grid cells with lodges.

For cells with lodges, we found the 10 most common restaurants in order to make a later clusterization and determine the best choice for a restaurant in that place. It is important to remind that this analysis is looking for a location of a ‘Peruvian cuisine’ restaurant, but this denomination is as broad as the variety of dishes and this is as broad as the variety of regions in Peru.

Due to this, we selected just some categories of restaurants that could be related to ‘Peruvian cuisine’, such as: *'Seafood Restaurant', 'Coffee Shop', 'Cafeteria','Café', 'Peruvian Restaurant','Restaurant','Food', 'Chinese Restaurant', 'Bakery','Fish & Chips Shop' , 'Buffet', 'Vegetarian / Vegan Restaurant', 'Dessert Shop', 'Ice Cream Shop'.*

The resulting table is shown below in the Table 2.

	Neighborhood	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	4	Seafood Restaurant	Restaurant	Café	Peruvian Restaurant	Coffee Shop	Bakery	Ice Cream Shop	Chinese Restaurant	Vegetarian / Vegan Restaurant	Food
1	5	Seafood Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Peruvian Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
2	6	Restaurant	Seafood Restaurant	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
3	7	Ice Cream Shop	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Peruvian Restaurant	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
4	8	Coffee Shop	Chinese Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Peruvian Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop

Tabla 2

Clustering the grid cells by the most common restaurants

Later, a clusterization will be done in order to determine similarities among different cells of the grid, based on the most common restaurants in every grid. This clusterization was carried out with a *KMeans* algorithm.

To determine the best number of clusters (best “k”), we used the *elbow method,* and the visual result is that X is the best “k” (Table 3).

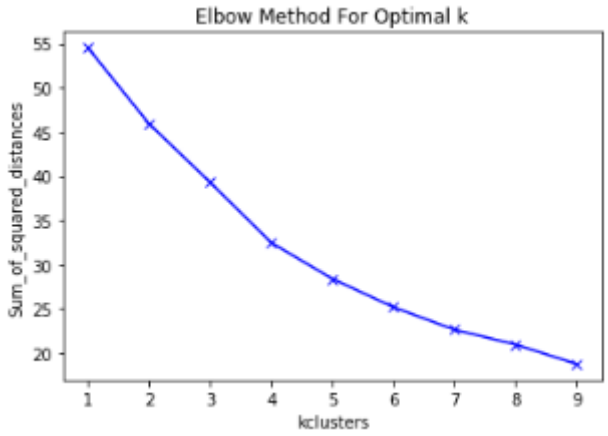


Table 3

In such table, is not so clear where is the “elbow”, but we can assume that is between 4 and 5, so we are going to choose 5, in order to have more possibilities for clustering among the restaurants.

Results and Discussion

The first results of the clusterization are shown in the Table 4

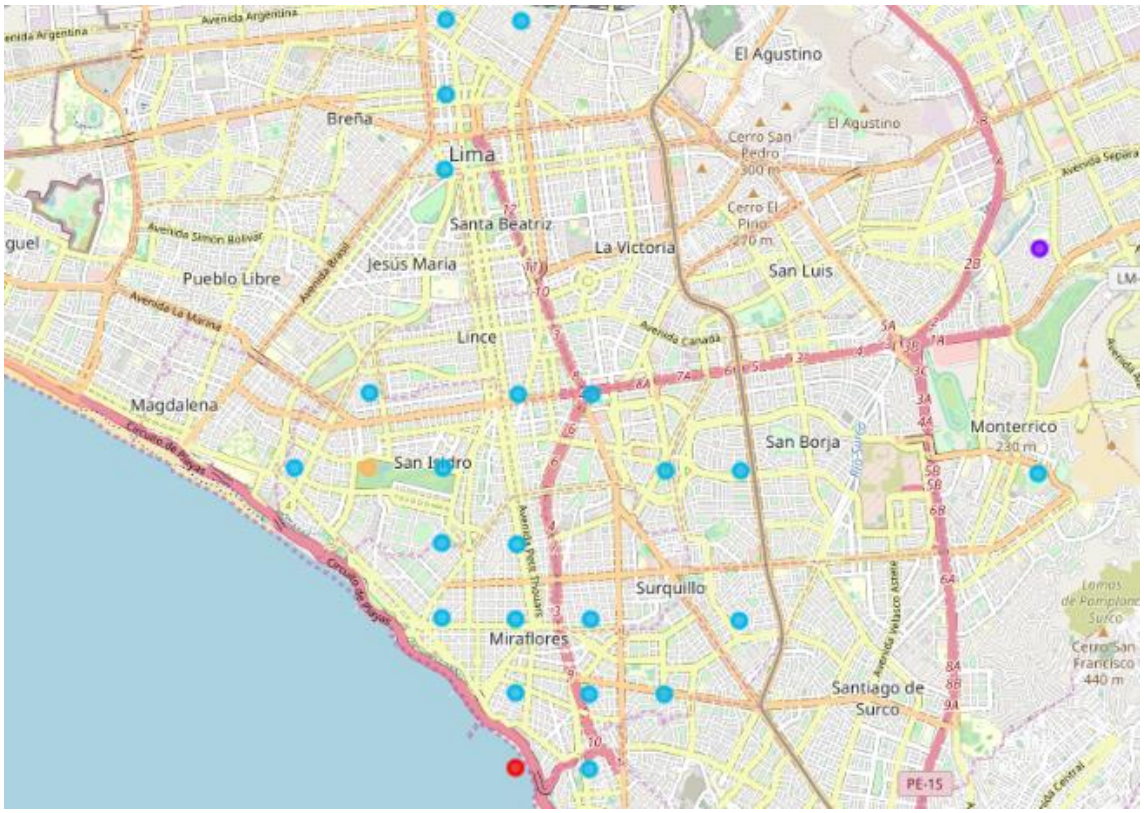
```
In [146]: neighborhoods_venues_sorted.sort_values(by='Cluster Labels')
```

Out[146]:

	Neighborhood	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
147	237	0	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
132	210	0	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
104	157	0	Peruvian Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop
126	196	0	Restaurant	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant
125	187	0	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant	Ice Cream Shop	Food	Fish & Chips Shop	Dessert Shop	Coffee Shop	Chinese Restaurant

Table 4

Finally, a map was made to have a view of the districts with more lodges and its clusters.



It is curious that cell grids with lodges, used to have the be on the same cluster, representing similarities between their most common restaurants. Being the cluster 2 the most common, with only 3 exceptions

As we can see, this is a first glimpse of where to see if someone is looking for open a new restaurant targeting to a tourist audience in Peru.

It is important to note that this project is just an example of a methodic approach that someone can use to have an business exploratory analysis. There are some parameters that we can optimize with a better understanding of the business (such as de number and the types of lodges for every grid, the categories of restaurants, etc).

That being said, this project can be a used as a guide of how to decide about what steps to make in order to start the research about the location of any kind of business

Conclusion

In this study, we evaluated the locations in Lima with the largest number of lodges, assuming this as a measure of tourist affluence. In those locations we made a clusterization process using de “KMeans” algorithm, using the most common restaurants on each one. We also made a map to locate this points and have a visual reference which, adding some experience about a place, serves as a guide for an exploratory analysis. This kind of simple but practical algorith research could be very useful to businessmen and women, for making the first decisions on where should they open a new business, in this case specially for new restaurants.