Geoanalysis Of The Restaurant Categories In The Metropolitan Lima Area, Peru

Jair Flores

April 20, 2020

1. Introduction

1.1. Background

In recent years, the restaurant sector experienced a significant growth according to INEI figures, reaching 4.87% in 2019[1], becoming in one of the most thriving businesses. For example until 2018 the number of formal restaurants in Peru amounted to 60 thousand, where Lima (Metropolitan Lima mainly) concentrated 41% with 25 thousand restaurants; meanwhile, in the rest of Peru the figure exceeded 35 thousand restaurants formally constituted according to the Lima Chamber of Commerce (LCC)[2].

1.2. Business Problem

The thriving restaurant sector is growing at a great rate throughout Metropolitan Lima area. In this sector, there are different categories or types of restaurants due to the diversification of people's tastes and specific foods.
But how much do we know about these categories and their distribution in our metropolitan area? Which are the mains? Is it possible to classify the metropolitan districts of Lima according to these categories? In this project we will use data science to answer these questions.

1.3. Target Audience

This project is aimed both the public interested on finding these answers and the Lima's people who want to start their medium or small business and want to find out what type of restaurant business is right in their district.

2. Methodology

For this project, we will direct our efforts to answer the questions of the commercial problem, we will find the distribution of the types of restaurants by the districts and subregions (groups of districts) which make up the Lima metropolitan area and obtain the main categories of restaurants, then we will cluster the types/categories of restaurants to generate a new classification of Lima Metropolitan area.The project methodology is shown below:

- Data Preprocessing
    - Data Reduction: Feature selection of the primary data to obtain districts of Lima Metropolitan
    - Data Collecting: Collecting metadata of a sample of restaurants
    - Data Cleaning: Resolving conflicts between data and drop duplicates

- Data Analysis
    - Normalization of the restaurant categories
    - Classification map of the Lima Metropolitan area by districts and subregions
    - Obtaining the main restaurant categories by districts and subregions
    - Heat map of the main restaurant categories
- Clustering Model
    - Linear Dimensional Reduction by PCA using SVD
    - Internal validation metrics to choose the optimal k
    - K-Means Clustering Algorithm
    - Mapping K-Means clusters

3. Data Preprocessing

3.1. Primary Data

Based on definition of our problem, we will need a shapefile file of the Peruvian districts in where are the Lima Metropolitan districts, this will provide us with a relevant geographic information that is available at the following link provided by the National Geographic Institute of Peru.

3.2. Data Reduction

Data downloaded is a spatial data file where is found the Peruvian districts geometry, therefore, we only select the Lima Metropolitan districts of our primary source (IGN) through ArcGIS site packages in Python like arcgis and arcpy.

3.3. Data Collecting

Once selected the districts, we need a sample of restaurants to analyze, for this reason we will create uniform fishnet points over Lima Metropolitan area, in total were 4,964 points according to Rate Limits in Foursquare.
These points will feed the Foursquare API that will recommend up to 50 restaurants for each fishnet point in a assigned radius of 1250 meters to completely cover the Lima Metropolitan area.

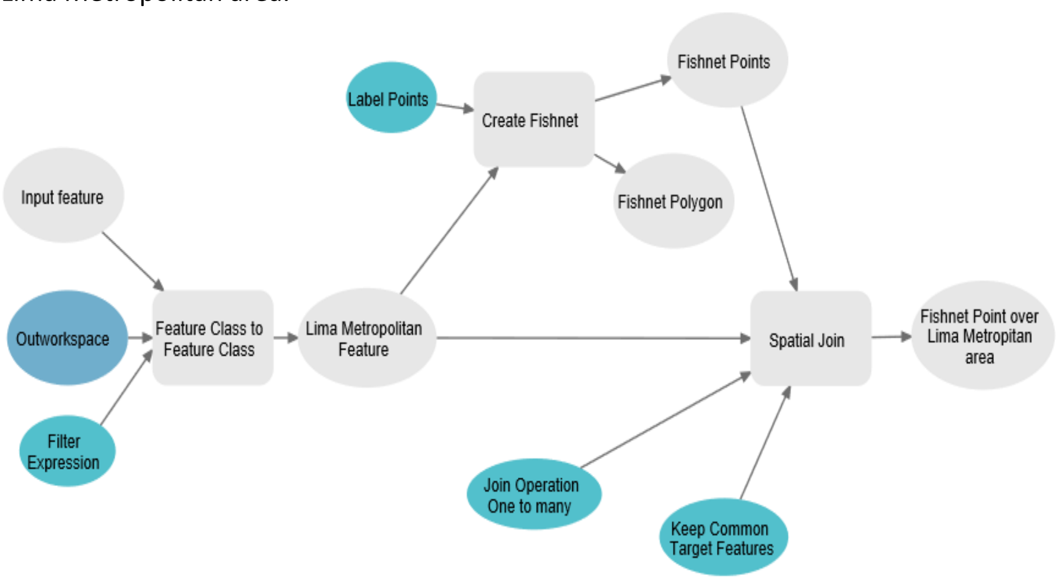


Figure 1 Workflow of data preprocessing made with Model Builder in ArcGIS Pro 2.5

3.4. Data Cleaning

Consists in cleaning the metadata (category, name, latitude and longitude) of the sample of restaurants saved in a dataframe and obtained through the Foursquare API feeded with our fishnet points.
For it, we verify if restaurant locations are contained in respective district and then drop the duplicate records to obtain a definitive dataframe with 5189 restaurants and 85 unique categories.

4. Data Analysis

We will analyze both spatial data of Lima Metropolitan like the sample of restaurants data to answer the first two questions of the project.
In this part you will find how types of restaurants are distributed throughout the metropolitan area by districts and subregions (groups of districts) and which are the main types/ categories of restaurants. Let's get started

4.1. Normalization of the restaurant categories

We normalize the number of restaurants categories by **Min Max** Method.

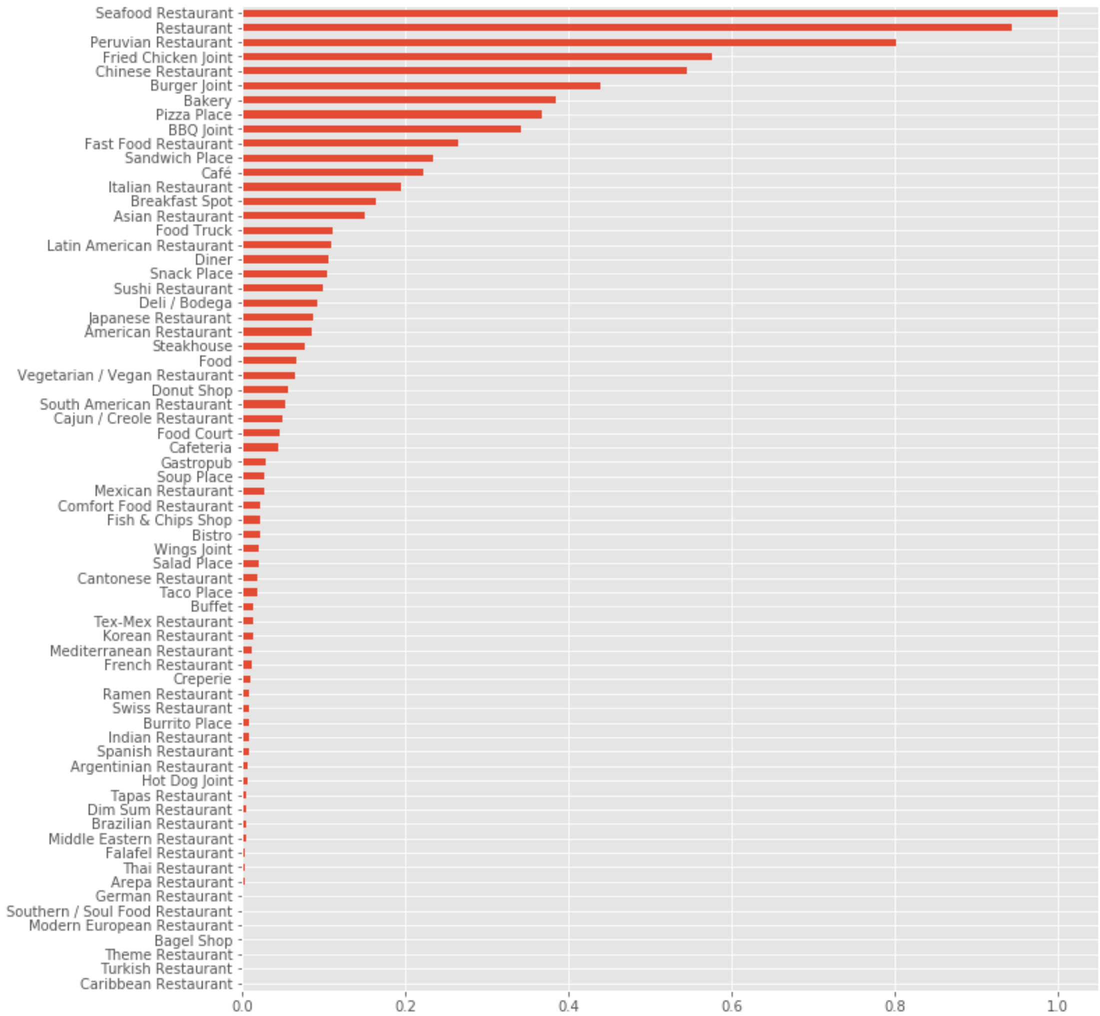


Figure 2 Min Max Normalization of the restaurant categories in a bar plot

From this graph we can comment Seafood Restaurant is the category of restaurant with the most presence in Metropolitan Lima according to our sample of 5189 restaurants and 68 categories of 85 (excluding the minimums that when normalized obtained 0).

Doing a statistical summarize it was found that seven types of restaurants are above the 90th percentile (0.373), this mean that are greater than 90% of the data, and fifteen categories are above the mean (with 0.121), the data have a median of 0.024 and there is a positive bias since the standard deviation (0.212) is greater than the mean, which corroborates the aforementioned, we have few "high" values and many "low" values.

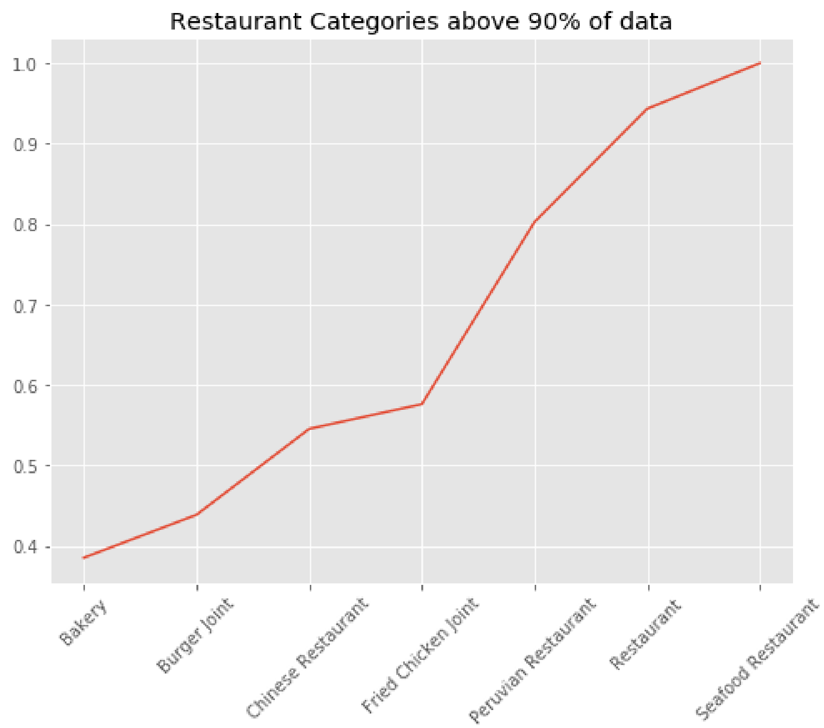


Figure 3 Restaurant Categories above 90% of data

4.2. Classification map of the Lima Metropolitan area by districts and subregions

The classification of the Lima metropolitan area is one of the most important steps in this project, since this will allow us to know how are distributed the types of restaurants.

We will classify them by districts (territories divided politically) and subregions (groups of districts divided by their geography, population and economic-social condition), to know how different they can be gastronomically talking

4.2.1. By districts

Lima Metropolitan is made up of 50 districts of which 6 belong to the Constitutional Province of Callao.

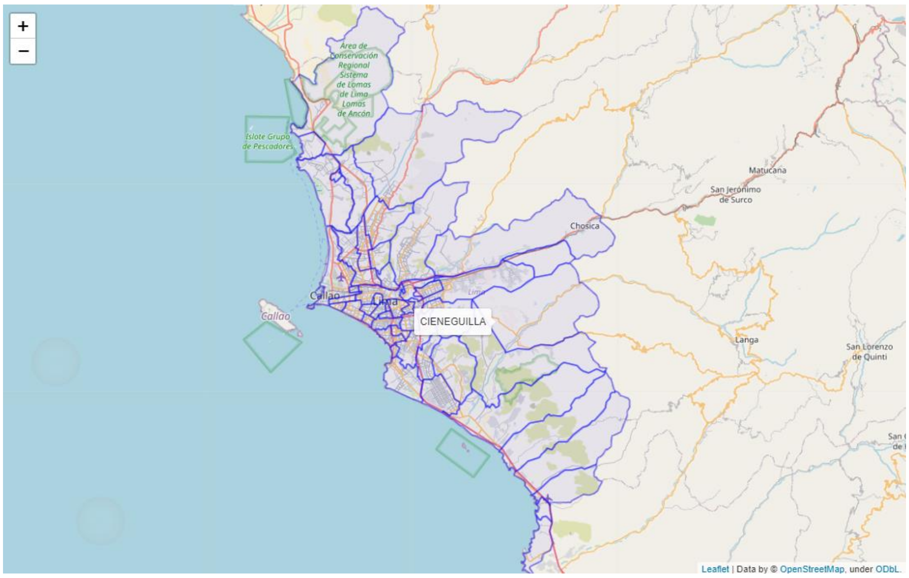


Figure 4 Lima Metropolitan districts

4.2.2. By subregions

Lima's population is geographically divided in six subregions: Lima Norte (turquoise), Lima Sur (red), Lima Este (yellow), Callao (orange), Central Lima (blue) and Residential Lima (purple).

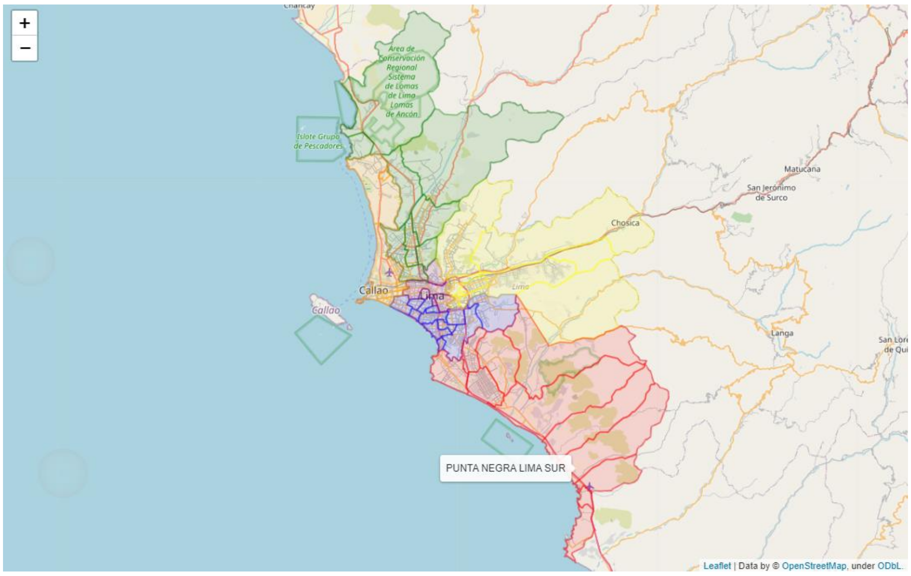


Figure 5 Lima Metropolitan subregions

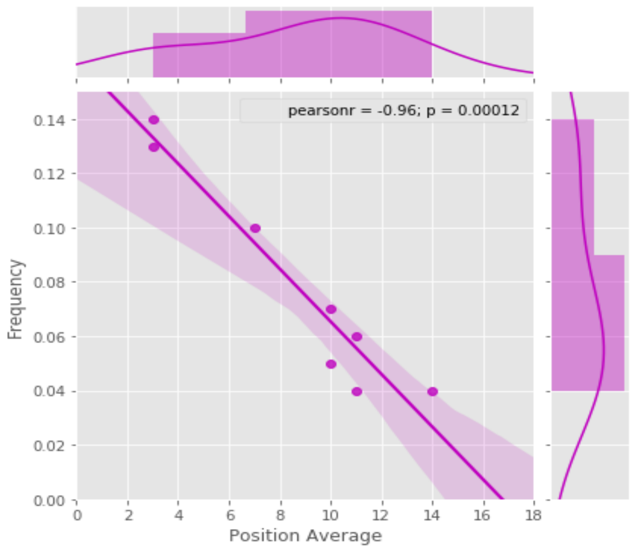
4.3. Obtaining the main restaurant categories by districts and subregions

We will start to manipulate the data of the dataframe to obtain the frequency of the types/categories of restaurants in the subregions and regions.
To finding the main categories, we will keep mind that the position that have a category in each district or subregion and its frequency play an important role to determining a main category.

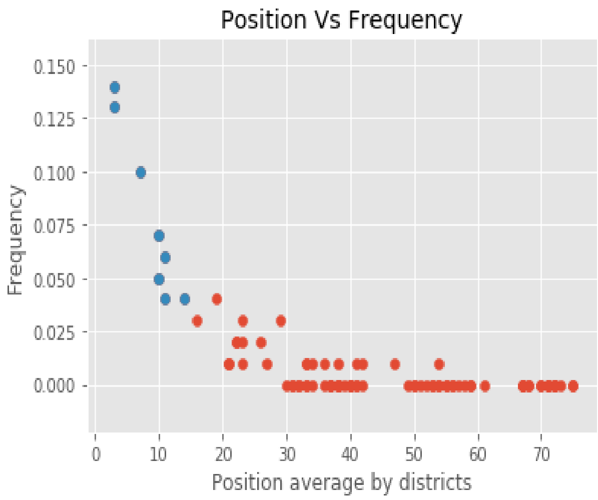
For it, we will do a scatterplot of Position Vs Frequency with the total average of their Position and Frequency, and will take only 10% of the data that are above 90% of the frequency data (90th percentile) and that are below 10% of the average positions (10th percentile).

A lower position (first positions) must indicate a higher frequency, therefore, the categories must present a strongest negative correlation, a min p- value would indicate that our hypothesis is right and we will see how dispersed are its frequency values between the districts and subregions with the Coefficient of Variation

In the case of the **districts**, we have a Position Vs Frequency scatterplot, in where the blue points are the 10% of the data that are above 90% of the frequency data (0.036) and that are below 10% of the average of positions (17). Of the Jointplot we know that these categories have a strongest negative correlation and a minimum p – value expected.
We have a greater dispersion with respect to the frequency average due to positive bias that becomes more evident in the categories and also because there are districts with a much smaller sample from restaurants, especially in the case of small districts, a problem that por example, at the sub-regional level it would be hidden, despite this, the CV in the categories does not exceed 80% with exception of the BBQ Joint category, so that can be still considered as homogeneous categories but with an frequency average with a low representation due to the high values that have.



	Category	Frequency	Position Average
0	Restaurant	0.13	3
1	Seafood Restaurant	0.14	3
2	Peruvian Restaurant	0.10	7
3	Chinese Restaurant	0.05	10
4	Fried Chicken Joint	0.07	10
5	Bakery	0.04	11
6	Burger Joint	0.06	11
7	BBQ Joint	0.04	14



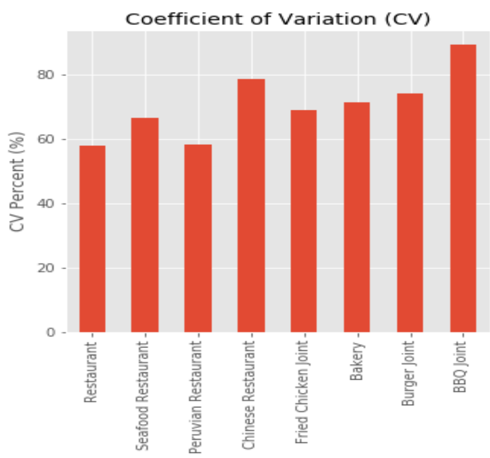
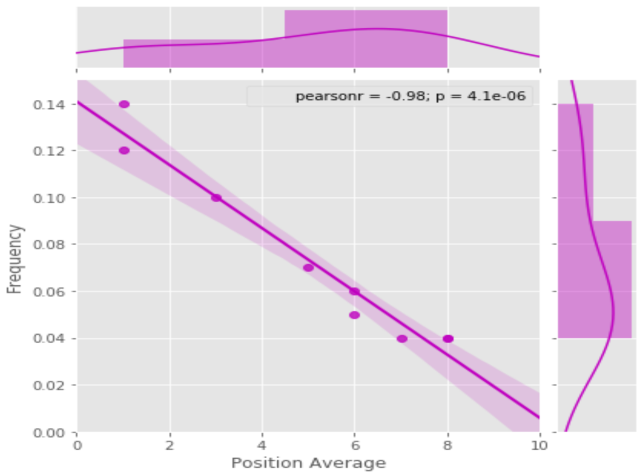


Figure 6 Graph Set that describe the Frequency vs Position, dispersion, correlation and p-value of the main categories

While in the **subregions** we have a Position Vs Frequency scatterplot, in where the blue points are the 10% of the data that are above 90% of the frequency data (0.036) and that are below 10% of the average of positions (9). Of the Jointplot we know that these categories have a stronger negative Pearson's correlation (-0.98) and a lower p-value (0.0000041), being better than the main categories by districts, regarding their dispersion they are considered homogeneous, being their maximum of CV the 40% in the Chinese Restaurant category, what means the restaurant categories have a good distribution in the subregions and a representative frequency average due to the high values that have decrease with at subregion level.



	Category	Frequency	Position Average
0	Seafood Restaurant	0.14	1
1	Restaurant	0.12	1
2	Peruvian Restaurant	0.10	3
3	Fried Chicken Joint	0.07	5
4	Chinese Restaurant	0.06	6
5	Burger Joint	0.05	6
6	Bakery	0.04	7
7	BBQ Joint	0.04	8
8	Pizza Place	0.04	8

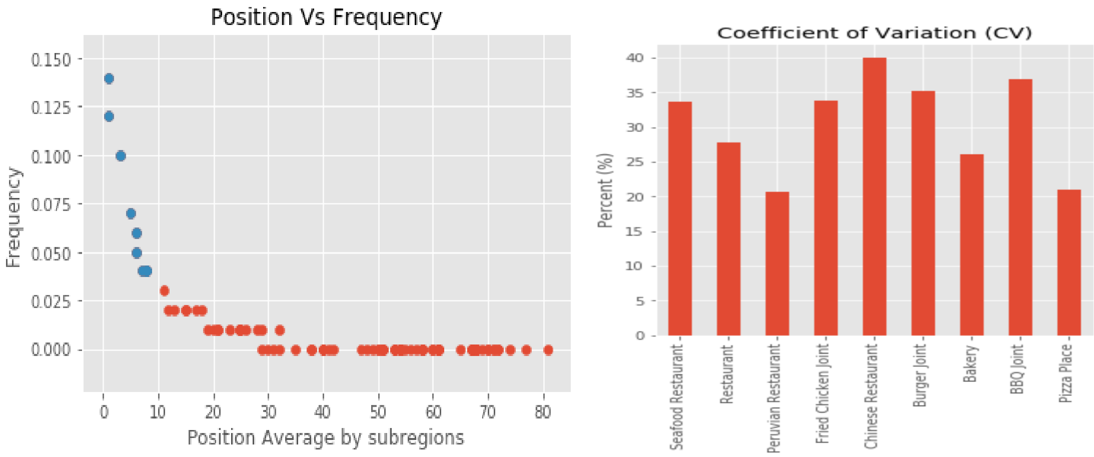


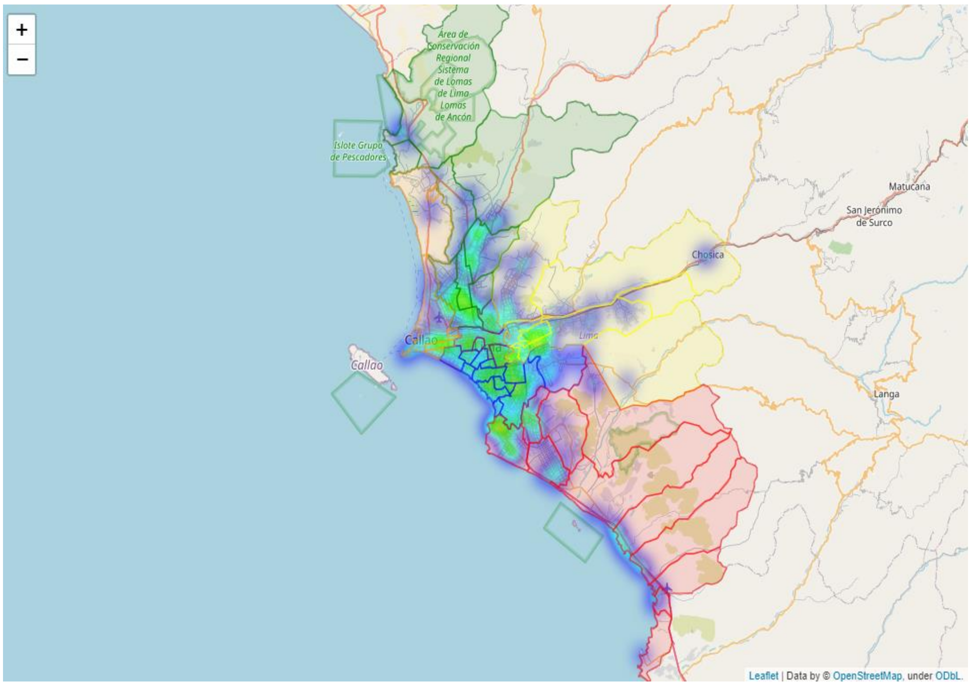
Figure 8 Position vs Frequency of restaurant categories by subregions

4.4. Heat Map of the main restaurant categories

To visualize the distribution and concentration of the main categories of restaurants on a map, we will make a heat map of each main restaurant category into Lima Metropolitan.

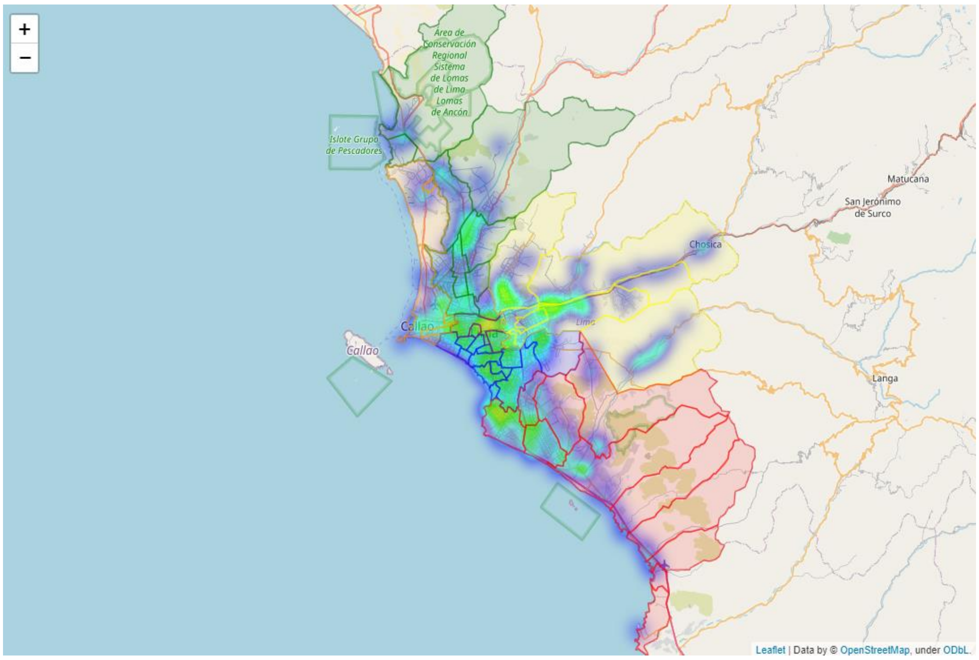
4.4.1. Seafood Restaurant Category

To visualize completely you can access in this [link]



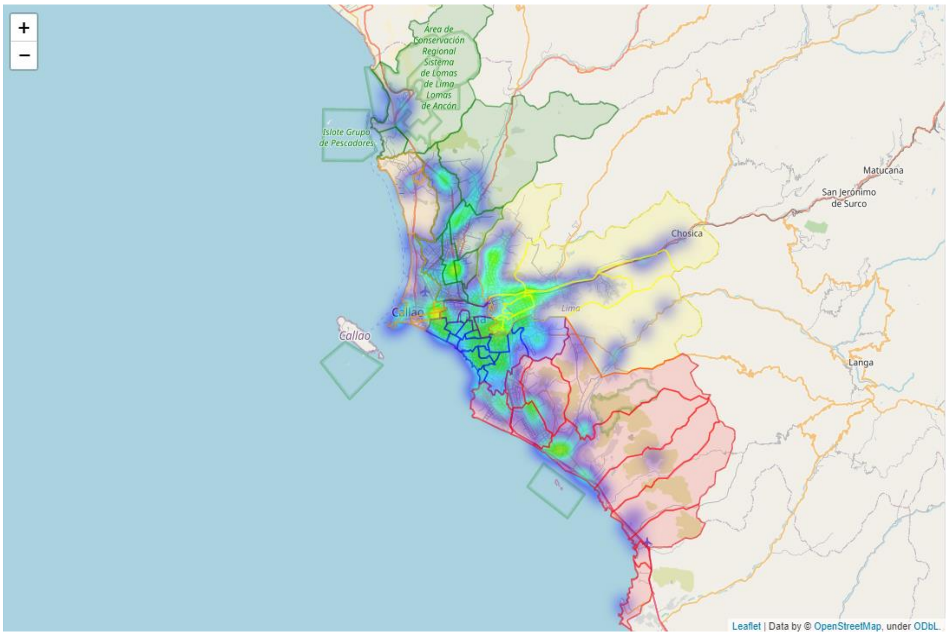
4.4.2. Restaurant Category

To visualize completely you can access in this [link]



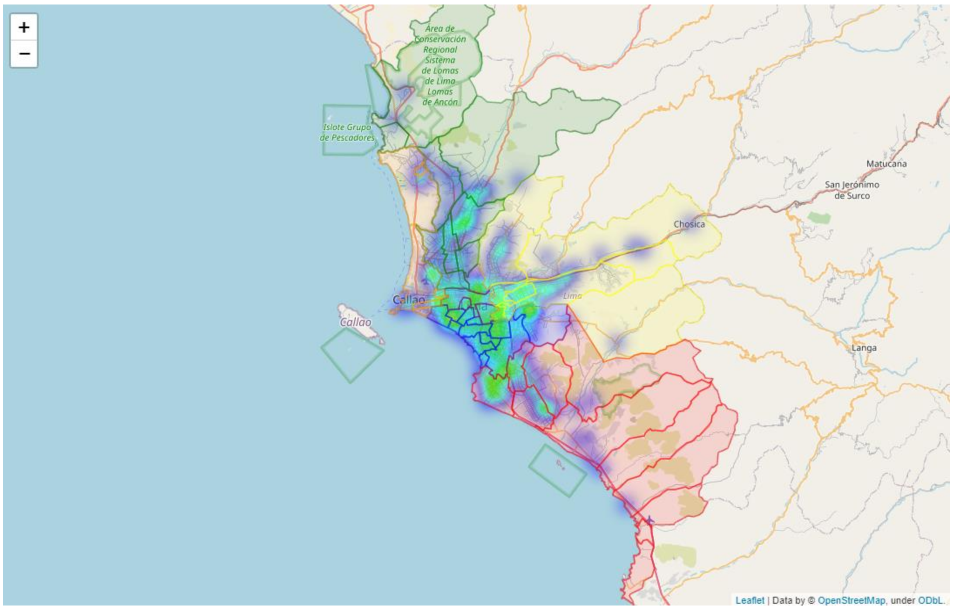
4.4.3. Peruvian Restaurant Category

To visualize completely you can access in this link



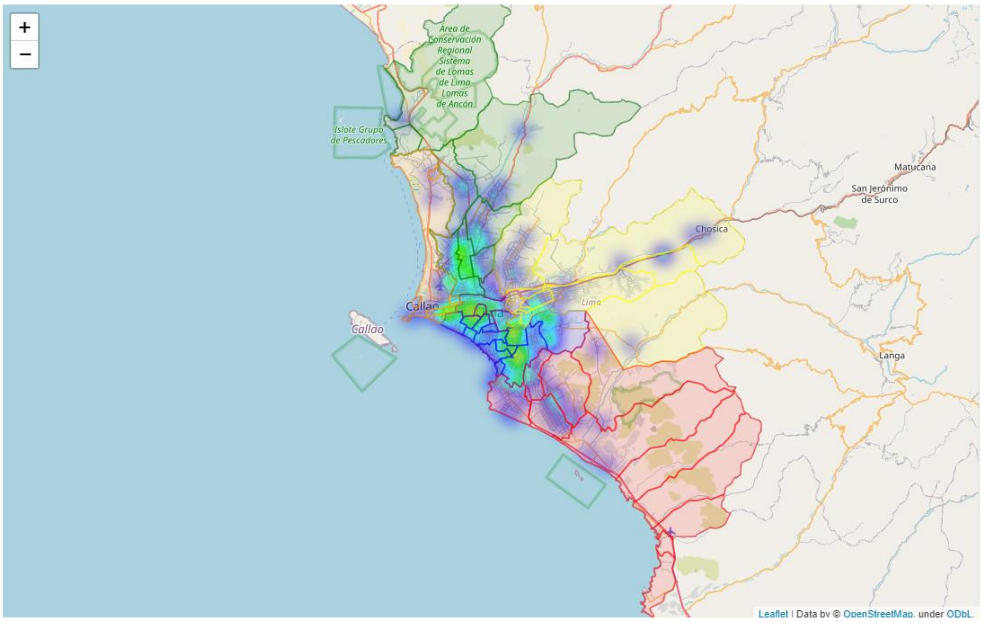
4.4.4. Fried Chicken Joint Category

To visualize completely you can access in this link



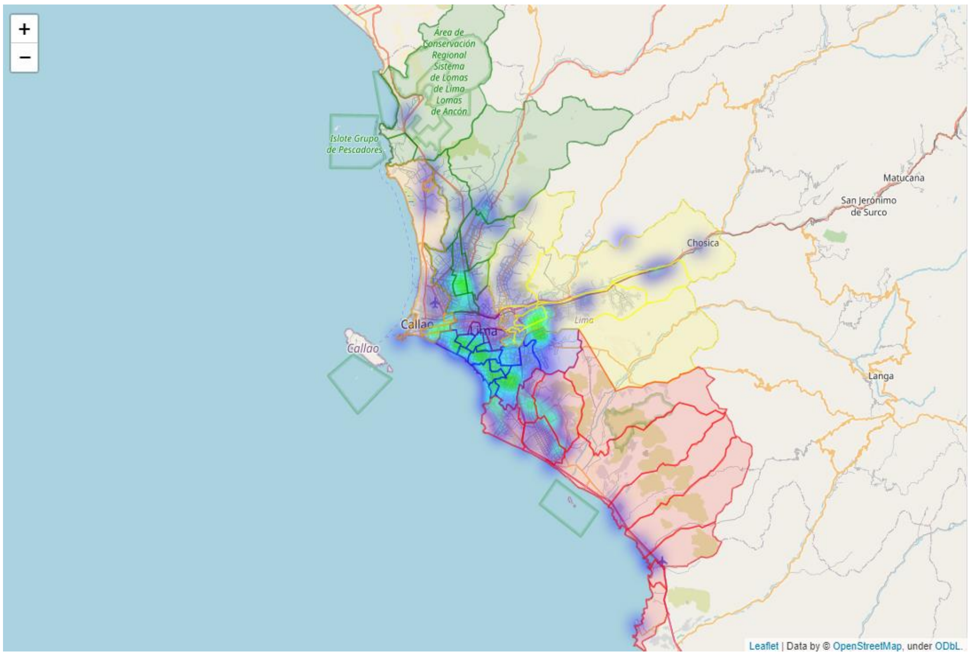
4.4.5. Chinese Restaurant Category

To visualize completely you can access in this link



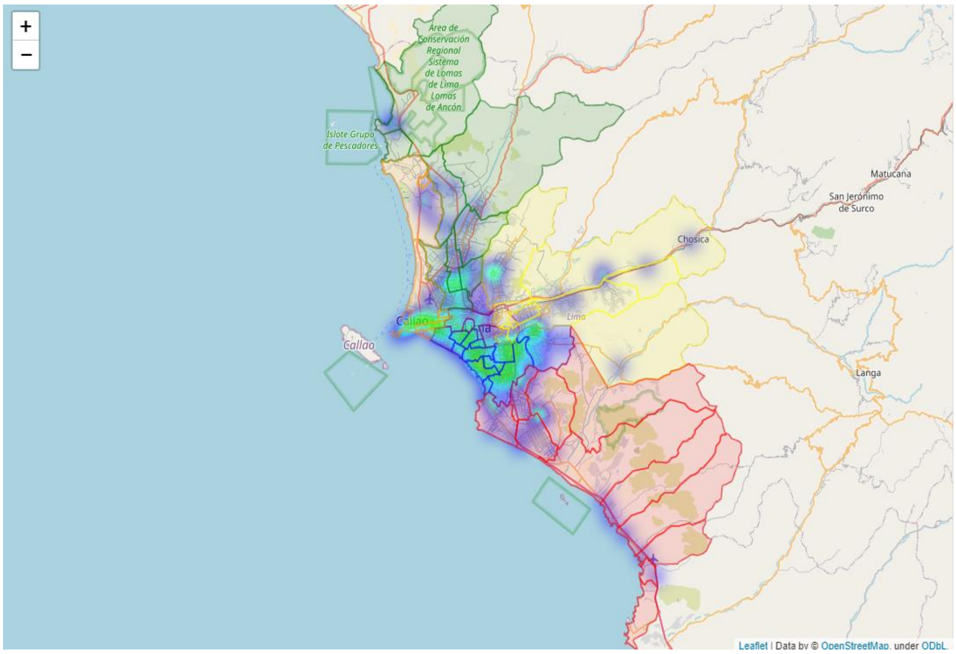
4.4.6. Burger Joint Category

To visualize completely you can access in this link



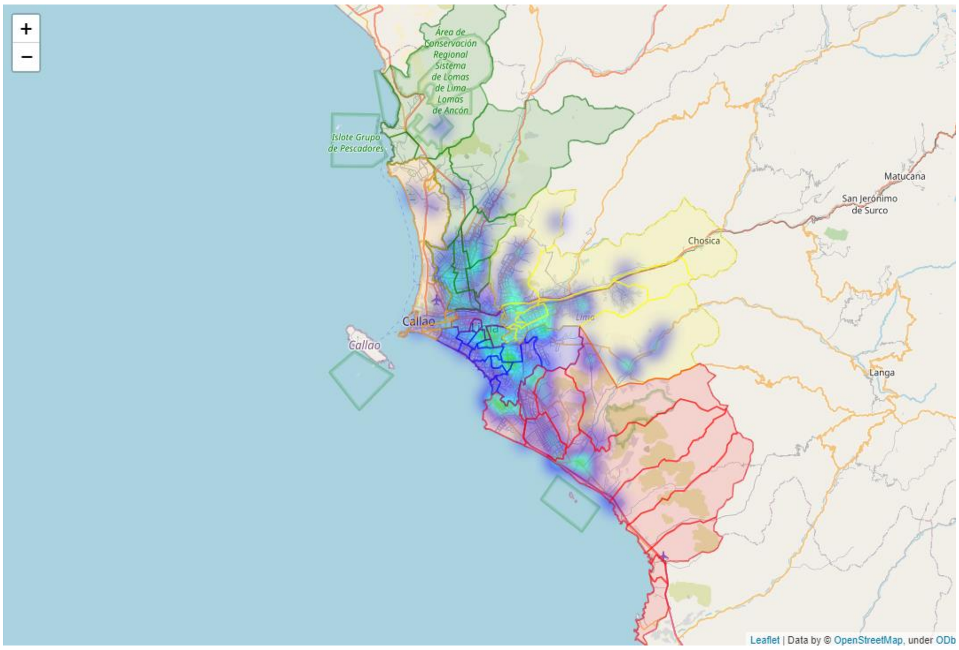
4.4.7. Bakery Category

To visualize completely you can access in this link



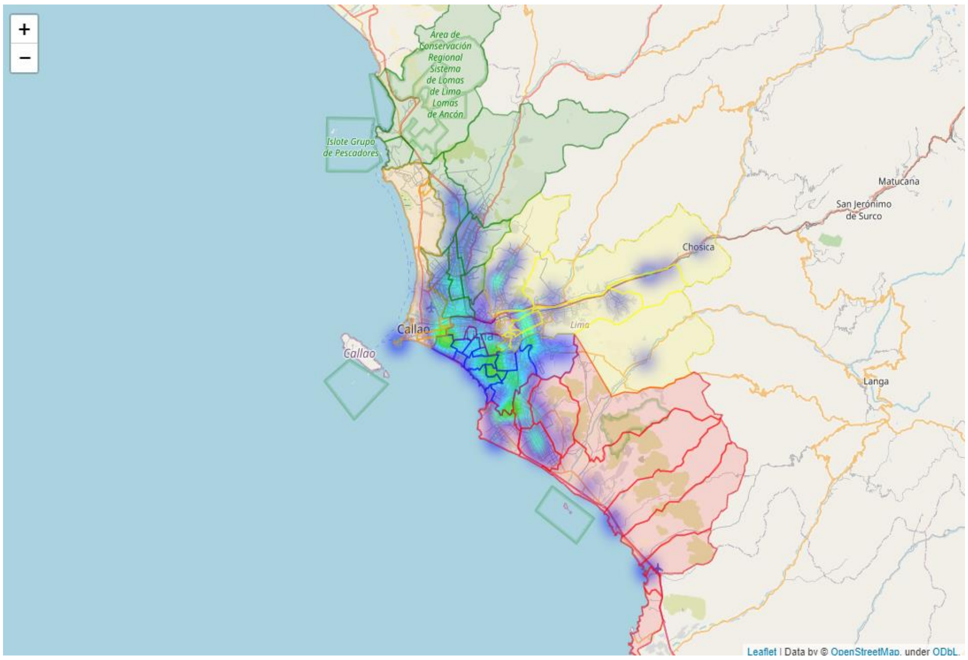
4.4.8. BBQ Joint Category

To visualize completely you can access in this link



4.4.9. Pizza Place Category

To visualize completely you can access in this link



In the heat maps of the main categories in the Lima Metropolitan area, is observed a large concentration of restaurant categories throughout the Residential and Central Lima though also in the large urban and commercial centers of Lima Norte, Sur, Este and Callao. Later we will see how much this influence

5. Clustering Model

In this section, we'll talk a little bit about the machine learning algorithm to answer this question: Is it possible to group the metropolitan districts of Lima according to these categories?

To start, we will use the clustering algorithm called K-Means that make the districts is grouping according into separable spherical groups so that the mean value converges towards the center of the group and each object is in the group where closest to mean value (of frequency of their categories), this method is one of the most common in clustering.

5.1. Linear Dimensionality Reduction

Doing an overview, our data frame contains 85 categories and 50 records from districts, if we want to group the districts with respect to all the categories, we would be facing a great problem since in spaces of very high dimensions (more than eighty in this case), Euclidean distances tend to inflate (this is an instance of the so-called "dimensionality curse"). For it, a good solution is to run a dimensionality reduction algorithm with the Principal Component Analysis (PCA) to obtain a simplest form (in 2-D) and better work with the clustering algorithm.

5.1.1. Principal Component Analysis (PCA)

The PCA with Singular Value Decomposition is a linear dimensionality reduction in where the axis are the Principal Components that are adjusted and oriented towards where the variance is maximum, through Sum of Square Distances (SSD), and are used the two components, with more variance accounted, as axes in the algorithm normalized by the Singular Value Decomposition.
We use the two components with the maximum variance.

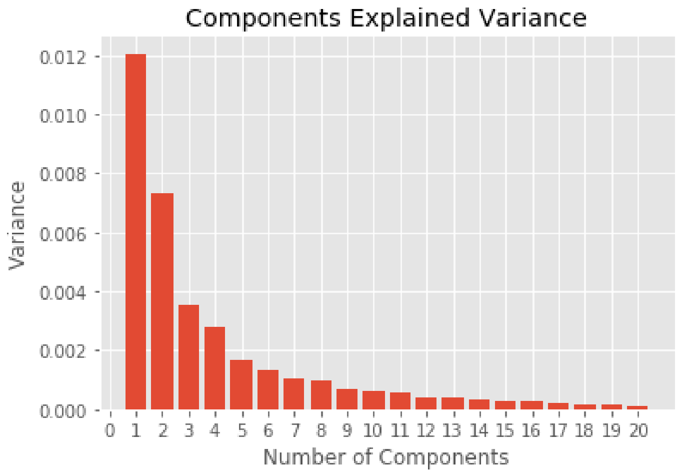


Figure 6 Principal Component Variance

5.2. Internal validation metrics to choose of the optimal k

Next to reduce the dimension of our data at two dimensions, the k means clustering is easy, but how k we choose it? We have answer, for this case we do not know the outcomes of cluster labels, then we only must evaluate the clusters internally through the internal validation metrics. Let's get started, there are two types of metrics in any clustering: Cohesion, separation and Internal Indexes [3]

5.2.1. Cohesion by Within – Cluster – Sum of Squares (WSS)

Cohesion means that objects of the same cluster are the most similar possible (closest to center value) this can be measured with Within-Cluster-Sum of Squared (WSS) that is the sum of the square distances of the point from its predicted cluster center and the optimal k is choose according to the elbow method.
In this case, the optimal k according to elbow method would be four or eight

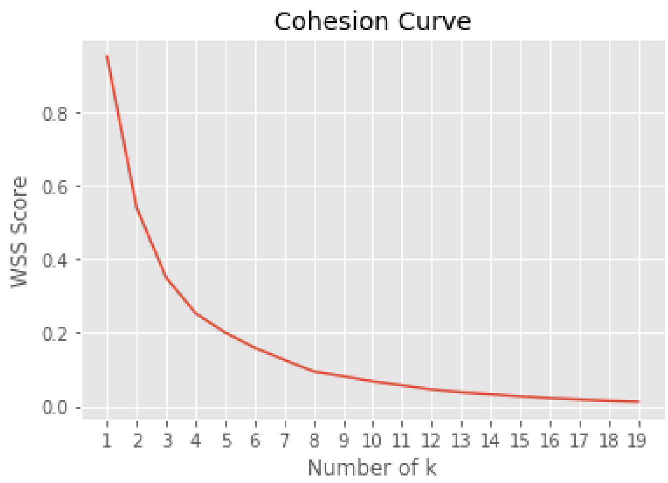


Figure 7 Cohesion Curve with WSS

5.2.2. Separation by Between – Clusters – Sum of Squares (BSS)

Separation means that clusters must be widely separated between them, A measurement of separation used to evaluate the intercluster distance is Between-Cluster-Sum of Squared (WSS) that is the sum of the square distances of the cluster centers with the mean of objects multiply by its object number, while longer distance, the clustering will be better.

In this case, the optimal k would be when BSS is high for example six or eight

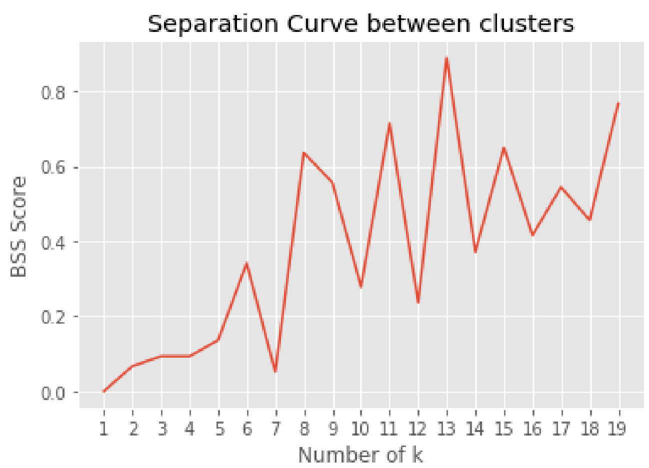


Figure 8 Separation Curve between clusters

5.2.3. Internal Indexes

Davies Bouldin Index and Silhouette Coefficient are the two main internal metrics

Davies Bouldin Index is calculated by the summation divided by the cluster numbers of maximum distances of the average of the distances between the points of the cluster "i" with the cluster "j" divided by the intercluster distance

$$DB = \frac{1}{k}\sum_{i=1, i\neq j}^{k} \max\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$$

From this graph, four and eleven are the number of clusters that minimizes the DB index, therefore, are taken as the optimal.

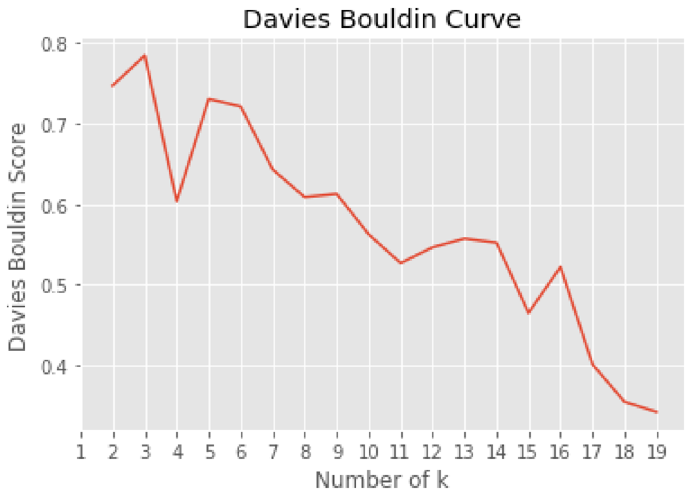


Figure 9 Davies - Bouldin Curve

Silhouette Coefficient is defined as:

$$SC = \frac{1}{N}\sum_{i=1}^{N} s(x)$$

Where s(x) is: $s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}}$

And b(x) is the average distance from “x” point to all other points in the same cluster and a(x) is the average distance from “x” point to all other points in the closest clusters.

From this curve, the optimal k without doubt is four since the Silhouette coefficient is between -1 and 1 in where the greatest value close to one is taken as optimal

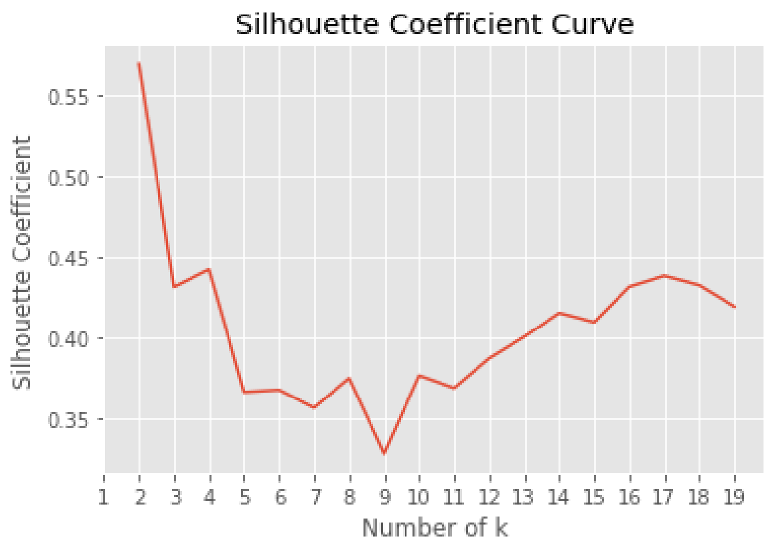


Figure 10 Silhouette Coefficient Curve

We conclude that, therefore, the number of k clusters is equal to **four** since with this number will be obtained a correct clustering.

5.3. K – Means Clustering Algorithm

We carry out the K – Means clustering with four number of clusters to obtain the classification of Lima Metropolitan districts by restaurant categories. We will visualize how the districts are clustered in our reduced dimension (2- D) in where the two dimensions are the two principal components.

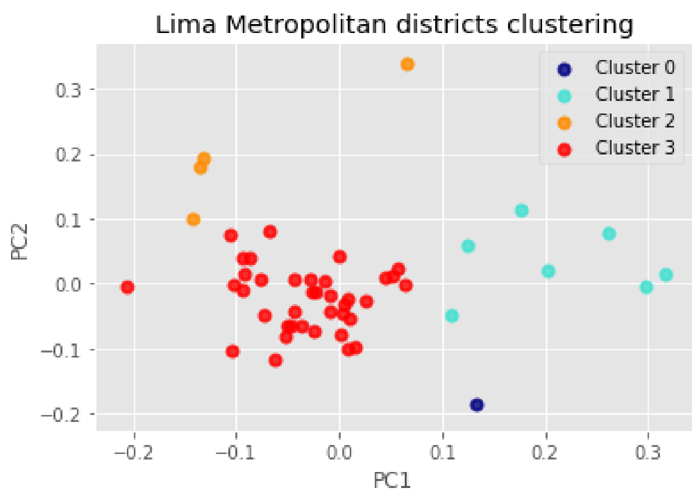


Figure 11 Lima Metropolitan districts clustering

We found that exists one cluster with only districts, to know which district is exactly. We can map K – Means Clusters to have a good visualization.

5.4. Mapping K-Means clusters

Here is the result! we show the Lima Metropolitan clustering by restaurant categories on the map. We have four clusters: 0(turquois), 1(red), 2(gray), 3(blue). In the results and discussion section we will examine these clusters in more detail.

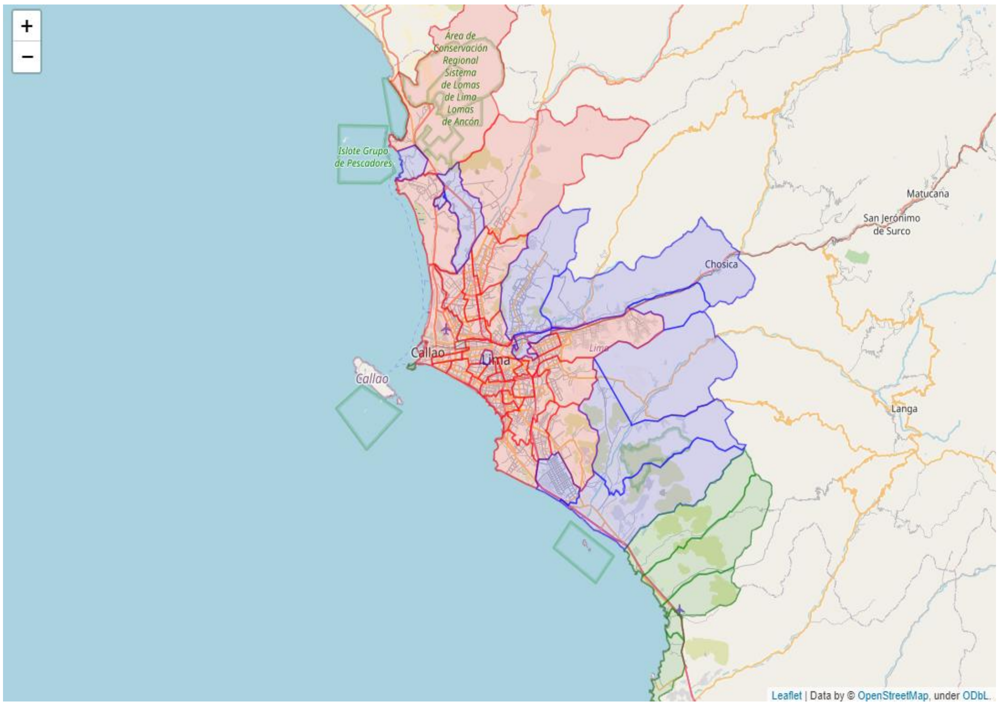


Figure 12 Lima Metropolitan clustering by restaurant categories

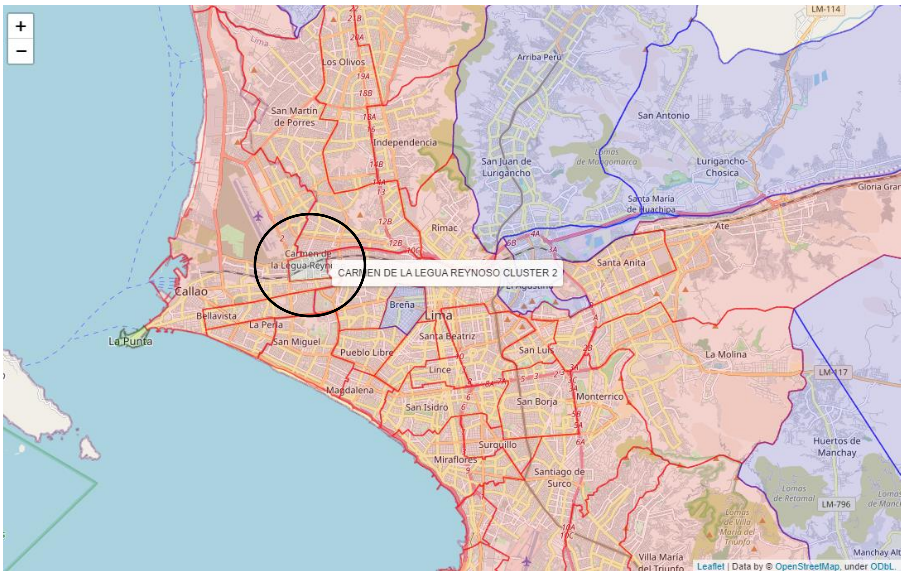


Figure 13 Cluster 2 with only Carmen de la Legua Reynoso district

6. Results and Discussion

Once selected the districts, we need a sample of restaurants to analyze, for this reason we create uniform fishnet points with 108 rows and 109 columns over Lima Metropolitan area, giving a total of 4,964 points according to Rate Limits in Foursquare. The radius of 1250 meters assigned to completely cover the Lima Metropolitan area would be deficient to take a sample of restaurants of small districts.

About the sample, knowing from the heat maps that Metropolitan Lima is also centralized, creating a network of points in the main centers of

Metropolitan Lima to obtain a much larger sample would be a good method, although it is far from the main focus of the project.
Our random sample of 5,188 restaurants representing approximately **5%** of restaurants in Metropolitan Lima, were identified 85 types or restaurant categories.
We find that certain categories have high frequency values and many categories with low values, this creating a positive bias in the distribution of restaurant categories.
In the case of districts we have a greater dispersion with respect to the frequency average due to positive bias that becomes more evident in the categories and also because there are districts with a much smaller sample from restaurants, especially in the case of small districts, a problem that por example, at the sub-regional level it would be hidden, despite this, the CV in the categories does not exceed 80% with exception of the BBQ Joint category, so that can be still considered as homogeneous categories but with an frequency average with a low representation.
The main categories are almost the same between districts and subregions with exception to Pizza Place Category are present in the main categories by subregions, the heat maps made for each one indicate the distribution of the main categories in the Lima Metropolitan area, where is observed a large concentration of restaurant categories throughout the Residential and Central Lima though also in the large urban and commercial centers of Lima Norte, Sur, Este and Callao

6.1. In the Clustering Model

After dimensionally reducing our data set and evaluating its optimal k, the result gave us a classification with four clusters. But what does it mean?

The cluster 0 (turquoise in the map) is made up by south Lima Metropolitan districts with extensive coastal areas and known for their holiday resorts and beaches and La Punta, have like main categories the Seafood Restaurant (all in the first position), Bakery and Pizza Place.

DISTRITO	1ST MOST COMMON CATEGORY	2ND MOST COMMON CATEGORY	3RD MOST COMMON CATEGORY	4TH MOST COMMON CATEGORY	5TH MOST COMMON CATEGORY
LA PUNTA	Seafood Restaurant	Pizza Place	Restaurant	Bakery	Italian Restaurant
PUCUSANA	Seafood Restaurant	Latin American Restaurant	Italian Restaurant	Restaurant	Cajun / Creole Restaurant
SANTA MARIA DEL MAR	Seafood Restaurant	Pizza Place	Restaurant	Deli / Bodega	Middle Eastern Restaurant
PUNTA NEGRA	Seafood Restaurant	Restaurant	Peruvian Restaurant	Bakery	Burger Joint
PUNTA HERMOSA	Seafood Restaurant	Restaurant	Sandwich Place	Bakery	Pizza Place
SAN BARTOLO	Seafood Restaurant	Pizza Place	Restaurant	Burger Joint	Italian Restaurant

Table 1 Districts within the Cluster 0

The second cluster 1 (red in the map) is the most extent cluster with almost the 60% of districts and is made up by central Lima Metropolitan and the larger and commercial urban centers.
In this cluster, we see the rise of fast food categories like Burger and Fried Chicken Joint as well as BBQ Joint and Chinese Restaurant, but with a clear predominance of Seafood and Peruvian Restaurant

DISTRITO	1ST MOST COMMON CATEGORY	2ND MOST COMMON CATEGORY	3RD MOST COMMON CATEGORY	4TH MOST COMMON CATEGORY	5TH MOST COMMON CATEGORY
ANCON	Peruvian Restaurant	Seafood Restaurant	Restaurant	Fried Chicken Joint	Fish & Chips Shop
SAN ISIDRO	Peruvian Restaurant	Restaurant	Seafood Restaurant	Café	Chinese Restaurant
MAGDALENA DEL MAR	Seafood Restaurant	Bakery	Fried Chicken Joint	Italian Restaurant	Burger Joint
SAN BORJA	Chinese Restaurant	Seafood Restaurant	Restaurant	Bakery	Sushi Restaurant
LINCE	Seafood Restaurant	Peruvian Restaurant	Restaurant	Italian Restaurant	Food Truck
SANTIAGO DE SURCO	Chinese Restaurant	Restaurant	Seafood Restaurant	Burger Joint	Bakery
PUEBLO LIBRE	Chinese Restaurant	Burger Joint	Restaurant	Pizza Place	Café
CARABAYLLO	Restaurant	Fried Chicken Joint	Chinese Restaurant	Peruvian Restaurant	Burger Joint
JESUS MARIA	Peruvian Restaurant	Chinese Restaurant	Restaurant	Seafood Restaurant	Bakery
CHORRILLOS	Seafood Restaurant	Restaurant	Fried Chicken Joint	Peruvian Restaurant	Pizza Place
BARRANCO	Seafood Restaurant	Restaurant	Peruvian Restaurant	Burger Joint	Café
VILLA MARIA DEL TRIUNFO	Seafood Restaurant	Restaurant	Burger Joint	Pizza Place	Fast Food Restaurant
SAN JUAN DE MIRAFLORES	Fried Chicken Joint	Burger Joint	Restaurant	Pizza Place	Seafood Restaurant
MIRAFLORES	Seafood Restaurant	Peruvian Restaurant	Café	Restaurant	Italian Restaurant
SURQUILLO	Seafood Restaurant	Peruvian Restaurant	Restaurant	Burger Joint	Sandwich Place
LA PERLA	Seafood Restaurant	Burger Joint	Chinese Restaurant	Restaurant	BBQ Joint
SAN LUIS	Seafood Restaurant	Restaurant	Pizza Place	Burger Joint	Fried Chicken Joint
SAN MIGUEL	Seafood Restaurant	Chinese Restaurant	Sandwich Place	Burger Joint	Pizza Place

LA VICTORIA	Seafood Restaurant	Restaurant	Peruvian Restaurant	Fried Chicken Joint	Chinese Restaurant
LA MOLINA	Restaurant	Seafood Restaurant	Chinese Restaurant	Peruvian Restaurant	Bakery
BELLAVISTA	Seafood Restaurant	Chinese Restaurant	Peruvian Restaurant	Fried Chicken Joint	Burger Joint
SANTA ANITA	Fried Chicken Joint	Seafood Restaurant	Peruvian Restaurant	Burger Joint	Chinese Restaurant
LIMA	Restaurant	Seafood Restaurant	Chinese Restaurant	Fried Chicken Joint	Peruvian Restaurant
RIMAC	Peruvian Restaurant	Fried Chicken Joint	Seafood Restaurant	Sandwich Place	Bakery
ATE	Restaurant	Seafood Restaurant	Peruvian Restaurant	BBQ Joint	Bakery
INDEPENDENCIA	Fried Chicken Joint	Fast Food Restaurant	Peruvian Restaurant	Restaurant	BBQ Joint
CALLAO	Seafood Restaurant	Peruvian Restaurant	Restaurant	Fried Chicken Joint	Bakery
SAN MARTIN DE PORRES	Fried Chicken Joint	Chinese Restaurant	Seafood Restaurant	BBQ Joint	Peruvian Restaurant
LOS OLIVOS	Seafood Restaurant	Fried Chicken Joint	Restaurant	Chinese Restaurant	Peruvian Restaurant
COMAS	Seafood Restaurant	Fried Chicken Joint	Restaurant	Peruvian Restaurant	BBQ Joint
VENTANILLA	Seafood Restaurant	Restaurant	Bakery	Fried Chicken Joint	Burger Joint

Table 2 Districts within Cluster 1

The third cluster 2 (gray in the map) only contains a district but looking the five first position of its categories we could integrate it into the first cluster. But keep in mind that is a small district, the sample taken from this district was lower than the others, making the value of the frequencies in these categories low and does not fitted in any other cluster.

DISTRICT	1ST MOST COMMON CATEGORY	2ND MOST COMMON CATEGORY	3RD MOST COMMON CATEGORY	4TH MOST COMMON CATEGORY	5TH MOST COMMON CATEGORY
CARMEN DE LA LEGUA REYNOSO	Pizza Place	Bakery	Seafood Restaurant	Café	Cafeteria

Table 3 Districts within Cluster 2

The last cluster 3(blue in the map) is made up by Southeast and two northern metropolitan districts that in which predominate the Peruvian and common restaurant categories with a great presence of the fast food categories.

DISTRICT	1ST MOST COMMON CATEGORY	2ND MOST COMMON CATEGORY	3RD MOST COMMON CATEGORY	4TH MOST COMMON CATEGORY	5TH MOST COMMON CATEGORY
LURIN	Peruvian Restaurant	BBQ Joint	Restaurant	Seafood Restaurant	Diner
SANTA ROSA	Restaurant	Seafood Restaurant	Italian Restaurant	Brazilian Restaurant	Peruvian Restaurant
VILLA EL SALVADOR	Restaurant	Peruvian Restaurant	Fried Chicken Joint	Seafood Restaurant	BBQ Joint
PACHACAMAC	Peruvian Restaurant	Restaurant	BBQ Joint	Seafood Restaurant	Food Truck
BREÑA	Restaurant	Fried Chicken Joint	Peruvian Restaurant	Chinese Restaurant	Fast Food Restaurant
CIENEGUILLA	Restaurant	Peruvian Restaurant	BBQ Joint	Italian Restaurant	Food Truck
EL AGUSTINO	Peruvian Restaurant	Fried Chicken Joint	Seafood Restaurant	Restaurant	Fast Food Restaurant
CHACLACAYO	Restaurant	Peruvian Restaurant	Pizza Place	Burger Joint	Italian Restaurant
LURIGANCHO	Peruvian Restaurant	Restaurant	Fried Chicken Joint	Seafood Restaurant	Pizza Place
SAN JUAN DE LURIGANCHO	Restaurant	Peruvian Restaurant	Fried Chicken Joint	Seafood Restaurant	BBQ Joint
PUENTE PIEDRA	Restaurant	Peruvian Restaurant	Seafood Restaurant	Pizza Place	Fried Chicken Joint
MI PERU	Restaurant	Burger Joint	Food Truck	Peruvian Restaurant	Chinese Restaurant

Table 4 Districts within Cluster 3

7. Conclusion

In conclusion we will answer the three questions of the project.

How much do we know about these categories and their distribution in our metropolitan area? Which are the mains?

From our analyzed sample (which represents 5% of the restaurants in Metropolitan Lima) we have noticed that there are few categories of restaurants that concentrate a large part of the sector: Seafood, Peruvian and Common Restaurant are undoubtedly the main ones and as a replacement for We have BBQ Joint, Bakery and Chinese Restaurant, but it should be noted that the fast food sector is well represented in the Metropolitan area with categories such as Burger, Fried Chicken Joint and Pizza Place. After these we only have unrepresentative and even sporadic categories.

Is it possible to classify the metropolitan districts of Lima according to these categories?

Sure, in this project we have made the classification of Metropolitan Lima through k-means clustering where the most optimal k validated internally was chosen. Which resulted in a satisfactory classification.

8. References

[1] INEI “INFORME TÉCNICO N°2 DE PRODUCCIÓN NACIONAL DICIEMBRE 2019– FEBRERO 2020”, 14 DE FEBRERO DE 2020

[2] PERU RETAIL. " EL CRECIENTE NEGOCIO DE LOS RESTAURANTES EN EL MERCADO PERUANO, 18 DE JULIO DE 2018, P. 1

[3] Elizabeth León Guzmán, Universidad Nacional de Colombia. MÉTRICAS PARA LA VALIDACIÓN CLUSTERING, Recovered from [https://disi.unal.edu.co/~eleonguz/cursos/mda/presentaciones/validacion_Clustering.pdf]