Analyzing Business Opportunities in Lima

Alex Felipe Miguel Peláez

February 19, 2020

1. Introduction

1.1. Background

Lima, capital city of Peru, is one of the largest cities in Latam with outstanding growing in the recent years. Peru has constantly maintained its macroeconomics indicators healthy. Its major indicators such as GDP, employment and inflation have shown good performance and have been stable. Due to that stability, there is an expecting increase of entrepreneurs who are looking for business opportunities. The whole market, mainly in Lima, is providing hopeful signs (growing demand for goods and services, exchange trade controlled, and lower barriers for entrance) which should be consider to invest in a new entrepreneurship.

1.2. Problem

Considering that Peru is a developing country in which source of information in good quality is not developed and is not available for everyone, we have considered to work with the scarce data presented on websites. The following report not only would provide input information about what are businesses' clustering segmentation in Lima but also would provide an important tool to identify if it is appropriate to start a new business in a determined geographical area based on competitor's composition. Finally, as an output of our analysis we can suggest a location to a new start-up company mainly related to its sector competitors.

1.3. Interest

Small entrepreneurs would be very interested in accurate prediction of the best location to start a new business, because with that information they will make more thoughtful decisions. Other stakeholders, such as retailers, banks and real estate companies will find this information useful. They can fit their market forecast to the cluster segmentation analysis provided in this report.

2. Data Preparation

The following report will consider two sources of information. The first one is related to a website in which we can find demographic composition of the whole of Lima districts. To get the main input information required for the analysis, we just have dropped some data in order to have cleaned our new dataframe. Then we have to convert some data sources into the proper data type to start our analysis. Then we just going to plot some bar charts identifying demographic composition of each district in order to identify the spot that might best fit to our desired demand.

This dataset exists for free on the web. Link to the dataset is:
https://es.wikipedia.org/wiki/Anexo:Distritos_de_Lima#Distritos_de_Lima

Distritos ⬍	Ubigeo ⬍	Área (km²) ⬍	Población ⬍	Densidad (/km²) ⬍	Fundado en ⬍	Código Postal ⬍	Ubicación
Ancón	150102	299.22	29419	98.3	29 de octubre de 1874	02	
Ate Vitarte	150103	77.72	419663	5399.7	2 de enero de 1857	03	
Barranco	150104	3.33	45922	13790.4	26 de octubre de 1874	04	
Breña	150105	3.22	94808	29443.5	15 de julio de 1949	05	

Illustration 1: Lima Districts source

The second data source was elaborate by us because of the scarcity of resources on public websites and official sites. We have developed manually all the Lima districts´ coordinates and have made a dataframe. We have converted the coordinates from DMS (degrees, minutes and seconds) to decimal degrees. Then we are going to plot each district in Lima map to show the composition of each cluster.

This dataset exists for free on the web. Link to the dataset is:

https://www.antipodas.net/coordenadaspais/peru/lima.php

3. Methodology

The methodology that we are going to implement is divided in two sections. The first one is based on graphic analysis. The second is related to clustering data in which we are going to use Foursquare API.

Indeed, in the first part we are going to plot two bar charts in which we can identify the density and area in each district. To do that, we first need to modify the data and to sort in descending order because is more straightforward to view.

Moreover, in the second section, we going to load coordinates from each district and create a dataframe. Then is important to plot our city (Lima) and identify the localization of each district. Then we just can apply clustering techniques with K-means in order to get venues per district, the most common venues and other important outcome.

At the end, our goal is to define in which districts is more attractive to invest and what kind of business is more common around the city, assuming that this kind of company is the most demanded in this Peruvian city.

3.1. Methodology part 1:

- Extracting data from website: I'm going to use Beautiful Soup library to get my data extracted.

Illustration 2: Extracting data from Lima Districts

- Scrapping data: I have to drop some data from my new dataframe in order to get the necessary information.

	Distritos	drop1	Ubigeo	drop2	Área	drop3	Población	Densidad	drop4	drop5	Fundado	en
0	Ancón		150102		299.22		29419	98.3			18741029 29 de octubre de 1874	
1	Ate Vitarte		150103		77.72		419663	5399.7			18570102 2 de enero de 1857	
2	Barranco		150104		3.33		45922	13790.4			18741026 26 de octubre de 1874	
3	Breña		150105		3.22		94808	29443.5			19490715 15 de julio de 1949	
4	Carabayllo		150106		346.88		188764	544.2			18210804 4 de agosto de 1821	
5	Chaclacayo		150107		39.5		39686	1004.7			19400424 24 de abril de 1940	
6	Chorrillos		150108		38.94		262595	6743.6			18570102 2 de enero de 1857	
7	Cieneguilla		150109		240.33		15784	65.7			19700303 3 de marzo de 1970	
8	Comas		150110		48.75		464745	9533.2			19611212 12 de diciembre de 1961	
9	El Agustino		150111		12.54		165425	13 191.8			19650106 6 de enero de 1965	

Illustration 3: Scrapping data from my new dataframe

- Cleaning data: I have to review the data type of each column and make changes if it is required. For example, It is needed to have area and population column with float data type.

	Distritos	Área	Población	Densidad
0	Ancón	299.22	29419	98.3
1	Ate Vitarte	77.72	419663	5399.7
2	Barranco	3.33	45922	13790.4
3	Breña	3.22	94808	29443.5
4	Carabayllo	346.88	188764	544.2
5	Chaclacayo	39.5	39686	1004.7
6	Chorrillos	38.94	262595	6743.6
7	Cieneguilla	240.33	15784	65.7
8	Comas	48.75	464745	9533.2
9	El Agustino	12.54	165425	13 191.8
10	Independencia	14.56	197000	13 551.4

Illustration 4: Dataframe Cleaned

- Plotting chart 1: In this bar chart I'm going to show Lima districts sorted by their areas values in ascending order.
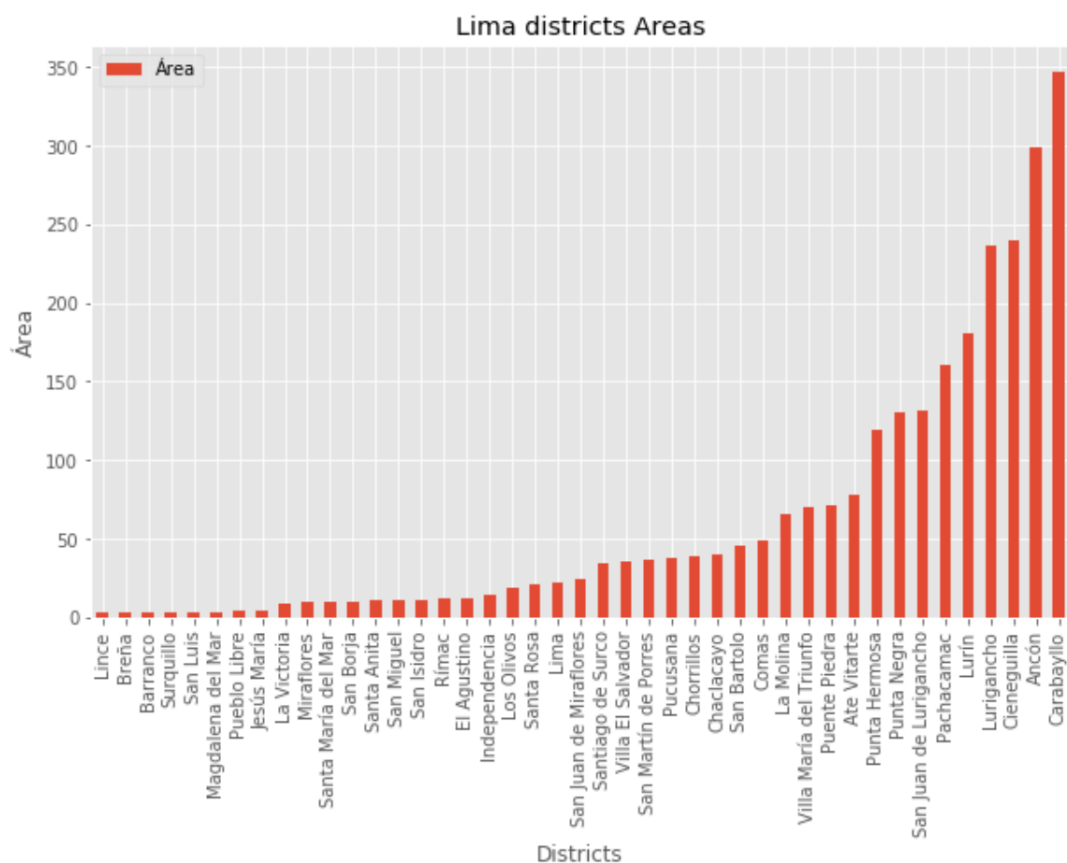
Illustration 5:Lima districts sorted by areas

- According to the graphic above, we can identify that Lurigancho, Cieneguilla, Ancon, and Carabayllo are the districts with the greatest area in Lima.

- Plotting chart 2: In this bar chart I'm going to show Lima districts sorted by their population values in ascending order.
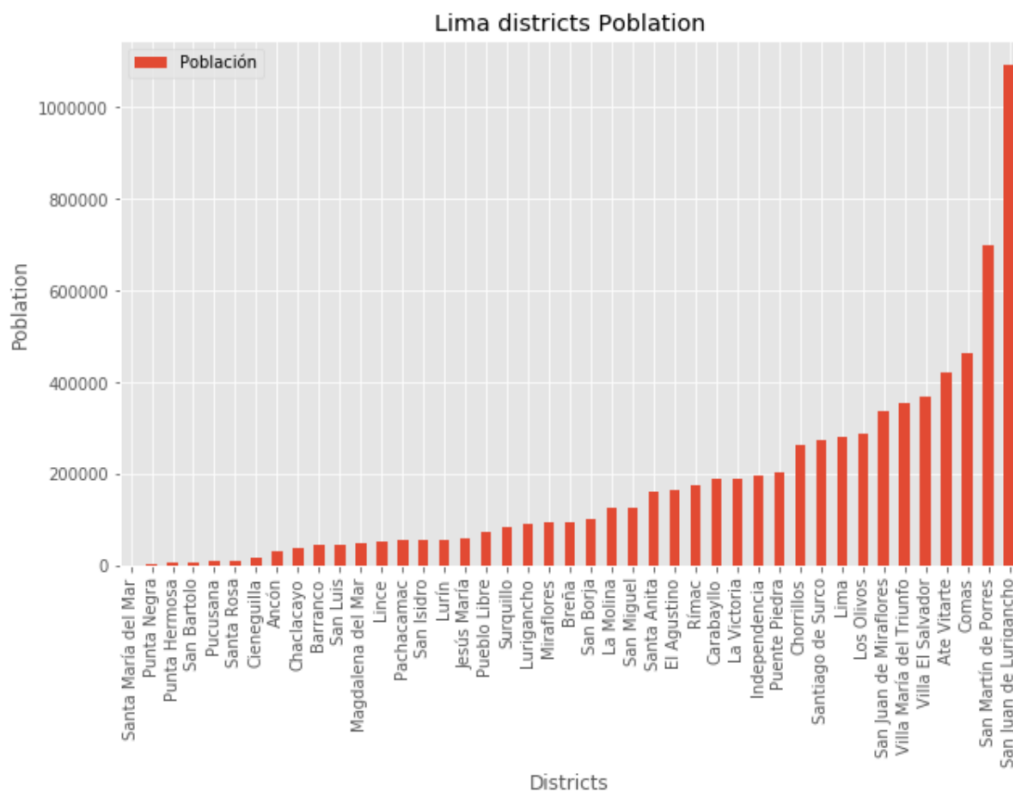
Illustration 6: Lima districts sored by population

- According to the graphic above, we can identify that Comas, San Martin de Porres and San Juan de Lurigancho are the districts with the greatest population in Lima.

- By both charts presented above we can infer that there is no correlation between area and population. This result is helpful because we can define later if the spots found are the ones that will have the greatest business concentration.

3.2. Methodology part 2:

- Creating new dataframe: due to the scarce of information in good quality, I decided to create my own database.

```
data = [['Lima01', 'Lima', -14.046071,-75.704294], ['Lima02', 'Ancon', -11.696553,-77.111654],['Lima03', 'Ate Vitarte', -14.046071,-75.704294],
       ['Lima04', 'Barranco', -12.143959,-77.0202268],['Lima05', 'Breña', -12.059700,-77.050118],['Lima06', 'Carabayllo', -11.794993,-76.989292],
       ['Lima07', 'Comas', -11.932861,-77.040674],['Lima08', 'Chaclacayo', -11.992479,-76.776176],['Lima09', 'Chorrillos', -12.192349,-77.008962],
       ['Lima10', 'El Agustino', -12.042052,-76.995714],['Lima11', 'Jesús María', -12.078186,-77.046411],['Lima12', 'La Molina', -12.090176,-76.922337],
       ['Lima13', 'La Victoria', -12.073357,-77.016417],['Lima14', 'Lince', -12.086567,-77.036647],['Lima15', 'Lurigancho', -11.948832,-76.762701],
       ['Lima16', 'Lurin', -12.238049,-76.783862],['Lima17', 'Magdalena del Mar', -12.491734,-75.911147],['Lima18', 'Miraflores', -12.121498,-77.025906],
       ['Lima19', 'Pachacamac', -12.251096,-76.906592],['Lima20', 'Pucusana', -12.482091,-76.797452],['Lima21', 'Pueblo Libre', -12.076638,-77.076638],
       ['Lima22', 'Puente Piedra', -11.876827,-77.074482],['Lima23', 'Punta Negra', -12.365557,-76.795190],['Lima24', 'Punta Hermosa', -12.332678,-76.825698],
       ['Lima25', 'Rimac', -12.020304,-77.035462],['Lima26', 'San Bartolo', -12.387071,-76.777945],['Lima27', 'San Isidro', -12.097902,-77.035366],
       ['Lima28', 'Independencia', -11.989307,-77.047330],['Lima29', 'San Juan de Miraflores', -12.159910,-76.969140],['Lima30', 'San Luis', -12.072355,-76.995890],
       ['Lima31', 'San Martín de Porres', -11.986759,-77.097655],['Lima32', 'San Miguel', -12.078655,-77.095283],['Lima33', 'Santiago de Surco', -12.125104,-76.981919],
       ['Lima34', 'Surquillo', -12.114197,-77.010474],['Lima35', 'Villa  Maria del triunfo', -12.176643,-76.918967],['Lima36', 'San Juan de Lurigancho', -11.948832,-76.762701],
       ['Lima37', 'Santa Maria del Mar', -12.401402,-76.775465],['Lima38', 'Santa Rosa', -12.035851,-77.086616],['Lima39', 'Los Olivos', -11.965985,-77.073071],
       ['Lima40', 'Cieneguilla', -12.073166,-76.777071],['Lima41', 'San Borja', -12.096451,-76.995689],['Lima42', 'Villa El Salvador', -12.213503,-76.937026],
       ['Lima43', 'Santa Anita', -12.223382,-76.847707]]
newdf = pd.DataFrame(data, columns = ['PostCode', 'District', 'Latitude', 'Longitude'])
newdf
```

Illustration 7: creating a new dataframe by setting decimal degrees coordinates of each district

- Building the new dataframe in columns showing latitude and longitude of each Lima districts.

	PostCode	District	Latitude	Longitude
0	Lima01	Lima	-14.046071	-75.704294
1	Lima02	Ancon	-11.696553	-77.111654
2	Lima03	Ate Vitarte	-14.046071	-75.704294
3	Lima04	Barranco	-12.143959	-77.020227
4	Lima05	Breña	-12.059700	-77.050118
5	Lima06	Carabayllo	-11.794993	-76.989292
6	Lima07	Comas	-11.932861	-77.040674
7	Lima08	Chaclacayo	-11.992479	-76.776176
8	Lima09	Chorrillos	-12.192349	-77.008962
9	Lima10	El Agustino	-12.042052	-76.995714
10	Lima11	Jesús María	-12.078186	-77.046411
11	Lima12	La Molina	-12.090176	-76.922337
12	Lima13	La Victoria	-12.073357	-77.016417
13	Lima14	Lince	-12.086567	-77.036647
14	Lima15	Lurigancho	-11.948832	-76.762701

Illustration 8: Coordinates Lima Districts

- Plotting Lima map: I'm going to use folium map to plot Lima map. Moreover, I'm going to show the districts location in the Lima map.

Illustration 9: Lima Districts locations

- Getting venues: I'm going to use Foursquare API. Then I'm going to run a code to get the closer venues in each neighborhood.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Ate Vitarte	2	2	2	2	2	2
Barranco	100	100	100	100	100	100
Breña	9	9	9	9	9	9
Carabayllo	1	1	1	1	1	1
Chorrillos	5	5	5	5	5	5
Comas	4	4	4	4	4	4
El Agustino	2	2	2	2	2	2
Jesús María	52	52	52	52	52	52
La Victoria	8	8	8	8	8	8
Lima	2	2	2	2	2	2
Lince	70	70	70	70	70	70
Los Olivos	4	4	4	4	4	4
Miraflores	75	75	75	75	75	75
Pachacamac	3	3	3	3	3	3
Pucusana	7	7	7	7	7	7
Pueblo Libre	34	34	34	34	34	34
Punta Hermosa	20	20	20	20	20	20
Punta Negra	8	8	8	8	8	8
Rimac	4	4	4	4	4	4
San Bartolo	20	20	20	20	20	20
San Borja	30	30	30	30	30	30

Illustration 10: Nearby venues per districts

- Sorting dataframe: I´m going to sort the above dataframe by number of venues.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Jesús María	53	53	53	53	53	53
Lince	71	71	71	71	71	71
Miraflores	76	76	76	76	76	76
San Isidro	94	94	94	94	94	94
Barranco	100	100	100	100	100	100

Illustration 11: Nearby venues dataframe sorted by number of venues in ascending order

- We can see in the above table that the districts with major venues are Miraflores, San Isidro and Barranco. None have a relationship with the neighborhoods found them in the beginning analysis (relationship between area and population).

- Clustering districts: Firstly, we have to use kMeans library to standardize area and population data.

```
array([[-0.70109223],
       [ 1.20415246],
       [-0.62052148],
       [-0.38185082],
       [ 0.07686006],
       [-0.6509668 ],
       [ 0.43731691],
       [-0.76766087],
       [ 1.42425127],
       [-0.03708533],
       [ 0.11857348],
       [-0.55868368],
       [-0.23704512],
       [ 0.08395877],
       [ 0.5164523 ],
```

Illustration 12 Data standardized

- Clustering districts: Secondly, we have to create labels for each cluster. To do that we define number of cluster =5.

```
[0 2 0 0 3 0 3 0 2 3 3 0 0 3 3 0 3 0 0 0 0 0 0 0 3 0 0 3 0 0 0 1 2 0 4 0 3
 0 0 3 0 2 2]
```

Illustration 13 Clusters Labels obtained

- Clustering districts: Then, we have to merge labels with our dataframe related to venues.

	PostCode	District	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Lima01	Lima	-14.046071	-75.704294	0	Pizza Place	Fast Food Restaurant	Event Space	Food & Drink Shop	Food
1	Lima02	Ancon	-11.696553	-77.111654	2	NaN	NaN	NaN	NaN	NaN
2	Lima03	Ate Vitarte	-14.046071	-75.704294	0	Pizza Place	Fast Food Restaurant	Event Space	Food & Drink Shop	Food
3	Lima04	Barranco	-12.143959	-77.020227	0	Bar	Seafood Restaurant	Burger Joint	Café	Pizza Place
4	Lima05	Breña	-12.059700	-77.050118	3	Fast Food Restaurant	South American Restaurant	Restaurant	Peruvian Restaurant	Pharmacy

Illustration 14: Labels merged in a new dataframe

- Clustering districts: Then, we are going to plot our new clusters (5) in Lima map. Each cluster will have different color.
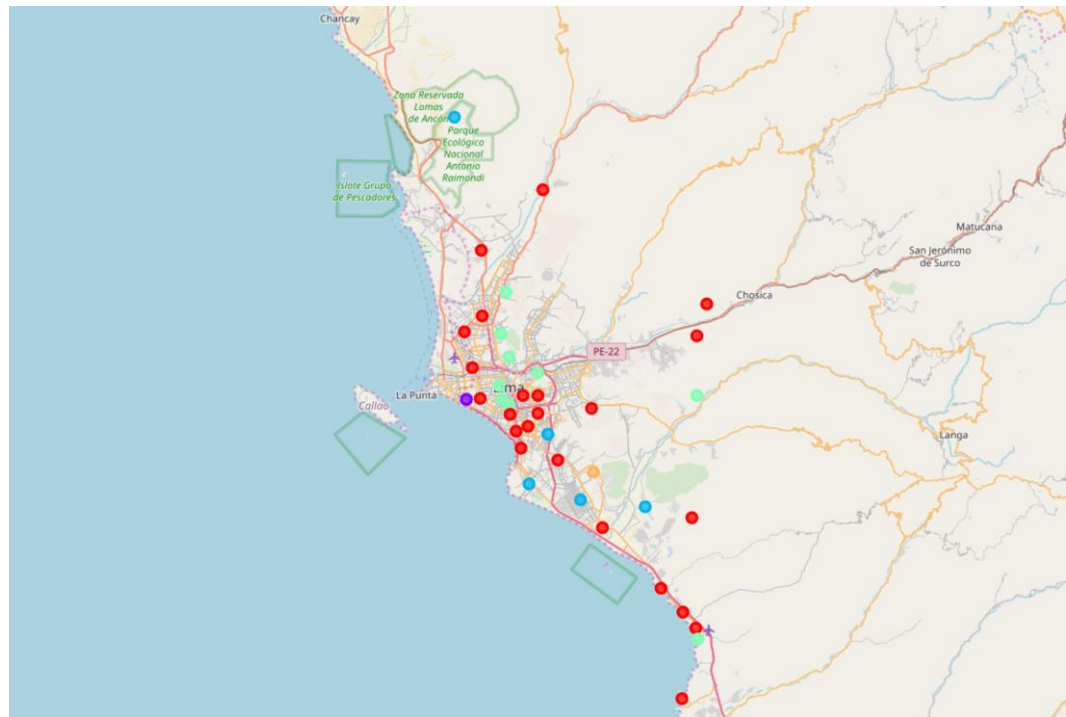

Illustration 15: Plotting Clusters in Lima Districts

- From the graphic above, we can say that clusters N° 0 and 2 are close to the center of the city and have an important concentration.

Cluster N°0: **Most common venues: Hotel, Seafood Restaurant, Peruvian Restaurant and Fast Food Restaurant**

	PostCode	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	Lima01	Pizza Place	Fast Food Restaurant	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Farmers Market	Farm	Falafel Restaurant	Electronics Store
2	Lima03	Pizza Place	Fast Food Restaurant	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Farmers Market	Farm	Falafel Restaurant	Electronics Store
3	Lima04	Bar	Seafood Restaurant	Burger Joint	Café	Pizza Place	Restaurant	Breakfast Spot	Sandwich Place	Peruvian Restaurant	Bakery
5	Lima06	Chinese Restaurant	Yoga Studio	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant
7	Lima08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	Lima12	Park	Yoga Studio	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant
12	Lima13	BBQ Joint	Bus Stop	Furniture / Home Store	Clothing Store	Yoga Studio	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
15	Lima16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	Lima18	Hotel	Seafood Restaurant	Coffee Shop	Café	Bar	Sandwich Place	Cocktail Bar	Peruvian Restaurant	Ice Cream Shop	Bookstore
18	Lima19	Farm	History Museum	Yoga Studio	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant
19	Lima20	Beach	Harbor / Marina	Park	Italian Restaurant	Cajun / Creole Restaurant	Bed & Breakfast	Yoga Studio	Farm	Food	Fish & Chips Shop
20	Lima21	Park	Sandwich Place	Seafood Restaurant	Burger Joint	Gym / Fitness Center	Fried Chicken Joint	Korean Restaurant	Soccer Field	Shopping Mall	Bookstore
21	Lima22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	Lima23	Nightclub	Plaza	Other Great Outdoors	Fried Chicken Joint	Restaurant	Beach	Electronics Store	Fast Food Restaurant	Farmers Market	Farm
23	Lima24	Seafood Restaurant	Beach	Sandwich Place	Bakery	Italian Restaurant	New American Restaurant	Restaurant	Cafeteria	Pizza Place	Farmers Market
25	Lima26	Seafood Restaurant	Nightclub	Pizza Place	Italian Restaurant	Peruvian Restaurant	Restaurant	Burger Joint	Scandinavian Restaurant	Park	Playground
26	Lima27	Peruvian Restaurant	Hotel	Italian Restaurant	Café	Sushi Restaurant	Restaurant	Coffee Shop	Seafood Restaurant	Latin American Restaurant	Salad Place
28	Lima29	Seafood Restaurant	Restaurant	Health & Beauty Service	Light Rail Station	Chinese Restaurant	Farmers Market	Soccer Field	Event Space	Fish & Chips Shop	Fast Food Restaurant
29	Lima30	Restaurant	Seafood Restaurant	BBQ Joint	Peruvian Restaurant	Auto Garage	Breakfast Spot	Park	Farmers Market	Intersection	Bakery
30	Lima31	BBQ Joint	Seafood Restaurant	Yoga Studio	Food Stand	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
33	Lima34	Café	Soccer Field	Fast Food Restaurant	Shopping Mall	Supermarket	Park	Food Court	Coffee Shop	Sporting Goods Shop	Dessert Shop
35	Lima36	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
37	Lima38	Skate Park	Café	Yoga Studio	Falafel Restaurant	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Event Space

Illustration 16: Cluster N° 0

- Miraflores, San Isidro and Barranco belong to this cluster and we can see that the majority of businesses in this segment are related to restaurants and hotels.

Cluster N°1: **Most common venues: Pharmacy and Event Space**

PostCode	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
34 Lima35	Pharmacy	Event Space	Food & Drink Shop	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Electronics Store

Illustration 17: Cluster N° 1

- Lima 35 is Villa Maria del Triunfo and is located far away from the center of the city.

Cluster N°2: **Most common venues: Fast Food Restaurant, Peruvian Restaurant and Department Store**

PostCode	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
4 Lima05	Fast Food Restaurant	South American Restaurant	Restaurant	Peruvian Restaurant	Pharmacy	Grocery Store	Fried Chicken Joint	Electronics Store	Farmers Market	Farm
6 Lima07	Department Store	Park	Shopping Mall	Movie Theater	Yoga Studio	Falafel Restaurant	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
9 Lima10	Fried Chicken Joint	Café	Yoga Studio	Falafel Restaurant	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Event Space
10 Lima11	Peruvian Restaurant	Seafood Restaurant	Park	BBQ Joint	Burger Joint	Restaurant	Boutique	Sandwich Place	Chinese Restaurant	Café
13 Lima14	Seafood Restaurant	Chinese Restaurant	Nightclub	Fried Chicken Joint	Restaurant	Bar	Peruvian Restaurant	Vegetarian / Vegan Restaurant	Burger Joint	Spa
14 Lima15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16 Lima17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24 Lima25	American Restaurant	Seafood Restaurant	Soccer Field	Pizza Place	Dance Studio	Falafel Restaurant	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
27 Lima28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
36 Lima37	Beach	Campground	Pool	Italian Restaurant	Yoga Studio	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
39 Lima40	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

illustration 18: Cluster N° 2

- Lima 05, 07, 10 and 11 are represented by Breña, Comas, El Agustino and Lince respectively, and those zones are closer to the center of the city.

Cluster N°3: **Most common venues: Seafood Restaurant and Chinese Restaurant**

PostCode	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
31 Lima32	Seafood Restaurant	Chinese Restaurant	Dance Studio	Pizza Place	Playground	Sandwich Place	Restaurant	Martial Arts Dojo	Soccer Field	Soccer Stadium

illustration 19: Cluster N° 3

- Lima 32 is San Miguel and is located in the coast and closer to the airport and seaport. That´s could be a factor to develop seafood restaurants in the zone.

Cluster N°4: **Most common venues: Hostel, Bar and Chinese Restaurant**

PostCode	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
1 Lima02	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8 Lima09	Hostel	Beer Garden	Clothing Store	Restaurant	Gas Station	Falafel Restaurant	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
32 Lima33	Park	Chinese Restaurant	Bakery	BBQ Joint	Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Italian Restaurant	Fried Chicken Joint	Ice Cream Shop
41 Lima42	Bar	Pizza Place	Dive Bar	Food	Market	Gym	Electronics Store	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
42 Lima43	Park	Moving Target	Yoga Studio	Event Space	Food	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant

Illustration 20: Cluster N° 4

- Lima 09 and 33 are "Chorrillos" and "Santiago de Surco", respectively. In those zones the most important businesses are Hostels and Chinese Restaurants.

4. Results

Firstly, I can conclude that the most common business in Lima is Restaurants. It seems that there is an important proportion in income spend related to feeding.

Moreover, from clustering segmentation we discover that Clusters N° 0 and 2 are the ones that have the most concentrated offer.

Finally, the predominance of restaurants and hotels might be related to tourism attractions. Is well known that Lima is continuously increasing its transit of tourists each year.

5. Recommendations

In order to increase the accuracy in our analysis, I strongly recommend to incorporate correlation among demographic factors such as income level, age, gender and education level. By analyzing those factors, we can get better conclusions between demographic pattern and business relationships.

Also is important to apply a demand analysis because our model is only based on supply (competitors). To have a better understanding about the market behavior I recommend to implement a regression analysis of the demand in each district.

6. Conclusions

According to the analysis made in this report I can conclude that in Lima is common to invest in restaurants, especially those related to typical food, bars and hotels. We can infer that those investments are strongly related to the increasing number of tourists in that city.

Moreover, according to the data, there is an important cluster, N°0, mainly represented by the districts of San Isidro, Miraflores and Barranco, that agglomerate the majority of the venues in the city. Because of that, if I have to suggest a business recommendation, we are going to bet for those spots.

Finally, I consider that there is an opportunity for other kind of businesses such as Gyms, department stores and Yoga studios that are currently emerging in other districts related to Cluster N°0. Investors can take advantage of this trending and replicate those businesses in other districts that belong to Cluster N°2 and N°4.