

CAPSTONE PROJECT

The Battle of Neighborhoods

Best Neighborhoods in Sao Paulo to Open an Italian Restaurant

Author: Gustavo Ribeiro Testa

email: gustavo.r@globo.com

April 11, 2020

1. Introduction

To open an Italian restaurant in a big city like Sao Paulo could be a risky business - open it in the wrong neighborhood and you are ruined. Sao Paulo is the biggest city of Brazil with more than 10 millions inhabitants, with lots of different places and neighborhoods to search for.

Data Science offer tools that can help to solve this problem, and identify the best places in town to open an Italian restaurant. But before we start collecting and analyzing data, and in order to achieve success in this task, it is **very important to know the business first**. The more knowledge you have regarding the business you are dealing with, better results and insights will come out of this.

The Business Problem

To open a successful Italian restaurant in the city, it is very important to define who is the target. Who will eat in the restaurant? The restaurant prices are going to be cheap or expensive? In this research, I am going to search for the best neighborhoods in town to open a Italian restaurant focused on people who does not bother to pay a little bit more than average prices.

So, this research is useful for anyone who wants to open an Italian restaurant in the city of Sao Paulo, targeting customers with medium to high income.

2. Data Section

2.1 Sources of Data

Data used in this research comes from 2 sources:

- a) Google maps
- b) FourSquare

Google maps provided geolocation data to all 91 neighborhoods of this research. One you access google maps, each Neighborhood Latitude and Longitude can be obtained by clicking on a location, and then click "What's Here?"

As a result google maps provides you the latitude and longitude of that location. I did this for all 91 neighborhoods of this research, and stored the values on an excel file.

As per FourSquare, I used an API that enable me to retrieve a venues list (up to 100 venues) for an specific geolocation. For this research, I retrieved a venue list for each of the 91 neighborhoods in the city, which result at the end 4,647 venues.

2.2 Data Structure

In order to process and analyze all these information, I will work mainly with **2 dataframes**:

a) Neighborhoods **dataframe**

b) Venues **dataframe**

2.2.1 Neighborhoods **dataframe**

it will be named **spdf**, and structured with the following columns:

- Neighborhood: the name of the Neighborhood
- Latitude....: the latitude of the Neighborhood
- Longitude....: the longitude of the Neighborhood
- Size.....: This column will be used to define the relative "Size" of the Neighborhood. Size could be, for example, the number of expensive restaurants in the area. For now, it will be set to 10.

index	Neighborhood	Latitude	Longitude	Size
0	Pinheiros (Paes Leme)	-23.566997	-46.696343	10
1	Pinheiros (Fradique)	-23.563412	-46.684643	10
2	Vila Madalena	-23.553547	-46.688649	10
3	Alto de Pinheiros	-23.553213	-46.708974	10
4	Boacava	-23.541327	-46.720548	10
5	Clinicas	-23.556694	-46.675425	10
6	Pacaembu	-23.546327	-46.668858	10
7	Jardim Europa	-23.574492	-46.681769	10
8	Itaim	-23.584443	-46.679146	10
9	Vila Olímpia	-23.594352	-46.677099	10

Latitude and Longitude values, as explained on item 2.1, was obtained using *google maps*.

2.2.2 Venues dataframe

It will be named **sp_venues**, and structured with the following columns:

- Neighborhood.....: the name of the Neighborhood
- Neighborhood Latitude : the latitude of the Neighborhood
- Neighborhood Longitude: the longitude of the Neighborhood
- Venue.....: Venue name
- Venue Latitude.....: Venue latitude
- Venue Longitude.....: Venue longitude
- Venue Category.....: Venue Category
- Price.....: Client evaluation price,
from 1 (cheap) to 4 (expensive).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Price
0	Pinheiros (Paes Leme)	-23.566997	-46.696343	Pitico	-23.566305	-46.696066	Beer Garden	1
1	Pinheiros (Paes Leme)	-23.566997	-46.696343	Comedoria	-23.567765	-46.697848	Buffet	1
2	Pinheiros (Paes Leme)	-23.566997	-46.696343	Bráz Eletrica	-23.566013	-46.695895	Pizza Place	1
3	Pinheiros (Paes Leme)	-23.566997	-46.696343	St. Chico Padaria	-23.565137	-46.695565	Bakery	1
4	Pinheiros (Paes Leme)	-23.566997	-46.696343	Um Coffee Co.	-23.568214	-46.698818	Coffee Shop	1
5	Pinheiros (Paes Leme)	-23.566997	-46.696343	Remelexo Brasil	-23.567562	-46.698365	Nightclub	1

I will use FourSquare API to get venues data for each neighborhood, and store it into the **venues dataframe**, see details in item 2.2.

3. Methodology

As explained before, I will use 2 sources of data: Google Maps and FourSquare. Once collected, data is going to be processed and analyzed. I will use **K-means, a machine learning unsupervised method** to analyze all venues in the city. As a result, neighborhoods are going to be grouped into clusters, based on its similarities and dissimilarities.

Since we are looking for neighborhoods that were visited by people with medium to high income (future clients of our Italian restaurant!), the clusters identified by the machine learning process will be overlapped with the customer perception of price for each neighborhood. This method will produce powerful insights and reveal the best neighborhoods to open a restaurant in Sao Paulo.

3.1 Define Neighborhood Database

Since Neighborhoods data was not available for download, I develop a way get the geolocation (Latitude and Longitude) of 91 neighborhoods of Sao Paulo. The data was retrieved manually, as follows:

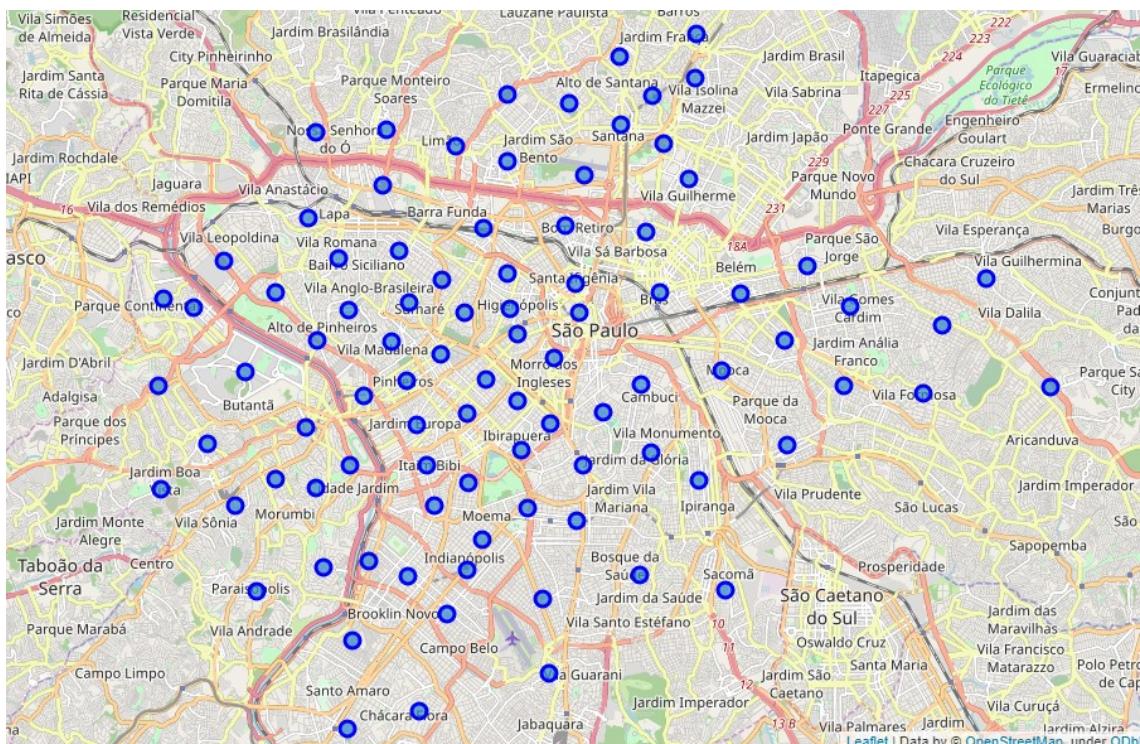
- a) An excel file named 'sp-neighborhoods.xls', was created with the following columns:

 - **Neighborhood**: Store the name of the Neighborhood
 - **Latitude**: Store the latitude of the Neighborhood
 - **Longitude**: Store the longitude of the Neighborhood
 - **Size**: This column will be used to define the relative "Size" of the Neighborhood. Size will receive the total value of restaurants in the neighborhood, filtered by price.

b) Then I went to maps.google.com/maps and took note of the latitude and longitude of each neighborhood I selected

c) Set Column 'Size' was set to 10 as default value (it will be modified later)

At the end, I had built a database with 91 neighborhoods for the city of São Paulo, with latitude and longitudes. I will use Folium to visualize all 91 neighborhoods:



Calculate the average distance between neighborhoods

In order to evaluate all venues in an specific area (defined by its latitude and longitude), Foursquare API requires a radius, so it will return every venues that is found within it and that satisfies the criteria specified.

I will set this radius as half of the average distance between neighborhoods.

Distance between Neighborhoods	
count	91.000000
mean	1.459429
std	0.447308
min	0.741260
25%	1.137664
50%	1.387757
75%	1.720193
max	3.452532

From that we conclude: **the radius required by Foursquare API** should be 1.459km / 2, which is approximately: $R = 750\text{m}$

3.2 Get Venues from Foursquare

To build our database **sp_venues** described in item 2. of this report, I need to make a foursquare request, using the GET method and the EXPLORE endpoint to retrieve the data.

The Foursquare request will have the following information:

- CLIENT_ID: Personal info for 'login'
- CLIENT_SECRET: Personal info for 'login'
- VERSION: Foursquare API version
- price_filter : Price filter from 1 to 4
- LIMIT: Maximum number of venues to return in the request
- RADIUS: All venues found within this radius will be returned in the request

I am also going to use a criteria to filter venues based on the price evaluation on the costumer. Costumers can classify a Venue using 4 different price, as follows:

- \$ (cheap)
- \$\$
- \$\$\$
- \$\$\$\$ (expensive)

Since there are 4 price, four requests will have to be made to FourSquare API, each of them asking for venues with a specific price. All of them are then concatenated into one *dataframe*, which contains a list off all venues in the city for the neighborhoods and price requested.

3.3 One Hot Encoding

Now, lets process the venues database and apply K-Means, a machine learning unsupervised method, to identify neighborhoods by its similarities and dissimilarities.

The venues database has 4647 Venues and 153 unique Venues Categories. The machine Learning method will identify similarities in Venues Categories and group them together into clusters. Here is an example of 10 different Venues Categories:

- Wine Bar
- Kebab Restaurant
- Tea Room
- Irish Pub
- Yoga Studio
- Greek Restaurant
- Fish Market
- Sake Bar
- Spanish Restaurant

But before we can apply K-Means, we need to prepare our data. K-Means function in Python only accept numerical data. So, we can do a “One Hot Encoding” to our database. This will create a *dataframe* with 4647 rows and 153 columns, each column being an Unique Venues Category.

	Neighborhood	Price	Acai House	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	...	Theater	Theme Park
0	Pinheiros (Paes Leme)	1	0	0	0	0	0	0	0	0	...	0	0
1	Pinheiros (Paes Leme)	1	0	0	0	0	0	0	0	0	...	0	0
2	Pinheiros (Paes Leme)	1	0	0	0	0	0	0	0	0	...	0	0
3	Pinheiros (Paes Leme)	1	0	0	0	0	0	0	0	0	...	0	0
4	Pinheiros (Paes Leme)	1	0	0	0	0	0	0	0	0	...	0	0

Data in each row and column is either one or zero. One means true, that is, the neighborhood in that row has a venue that belongs to the venues category specified by the column name. In other words, when data is true, neighborhoods and venues categories are related. And when false, no relationship.

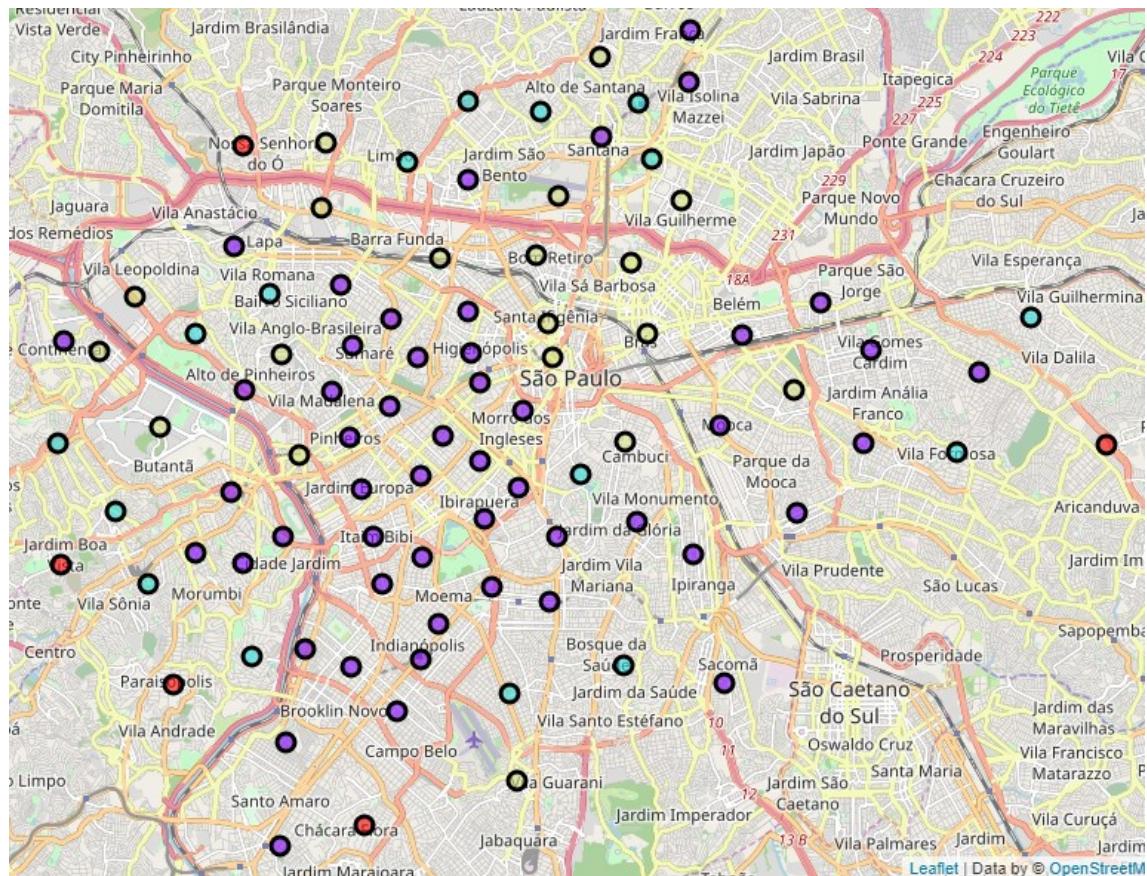
3.4 K-Means

One algorithms that can be used for neighborhood segmentation is K-Means clustering. K-Means can group neighborhoods in an unsupervised way, based on the similarity and dissimilarity of neighborhoods from each other.

K-Means will analyze the One Hot Encoded *dataframe* and group similar neighborhoods, assigning each of them to a cluster, based on whether they have similar attributes. All similar items should fall into a cluster, and dissimilar items should fall into different clusters.

After applied, K-Means assign to each neighborhoods an specific cluster, and groups them.

Here is an example for number of clusters = 4:



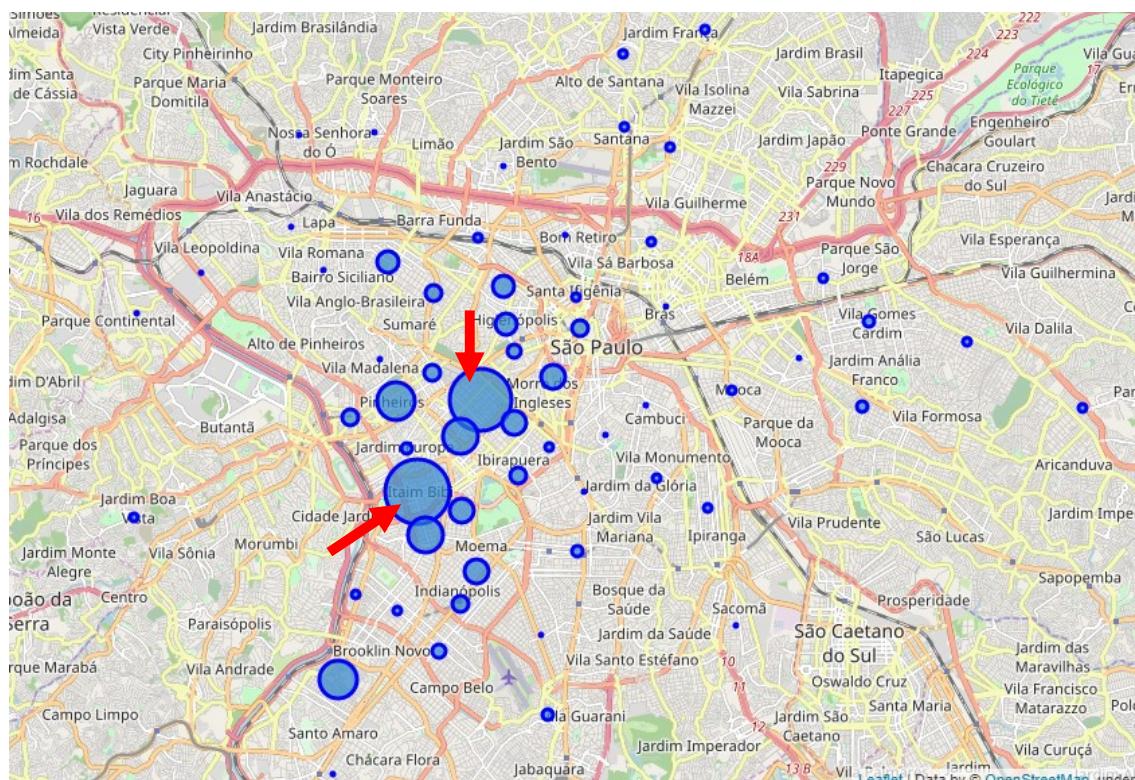
4. Results

Considering all 91 Neighborhoods and 4 different price ranges, Foursquare returned a total of 4647 venues in the city. In this section we will explore the results and take a closer look on them.

4.1 Top Neighborhoods with Italian restaurants

Total Italian Restaurants	
Neighborhood	
Itaim	19
Cerqueira Cesar	18
Pinheiros (Fradique)	11
Chacara Santo Antonio	11
Jardins	10
Vila Olimpia	10
Moema	7
Bixiga	7
Vila Nova Conceicao	7
Pamplona	7

TOTAL ITALIAN RESTAURANTS PER NEIGHBORHOOD



(Bigger radius means more Italian restaurants in that neighborhood)

4.2 Price: Venues and Restaurants per Neighborhoods

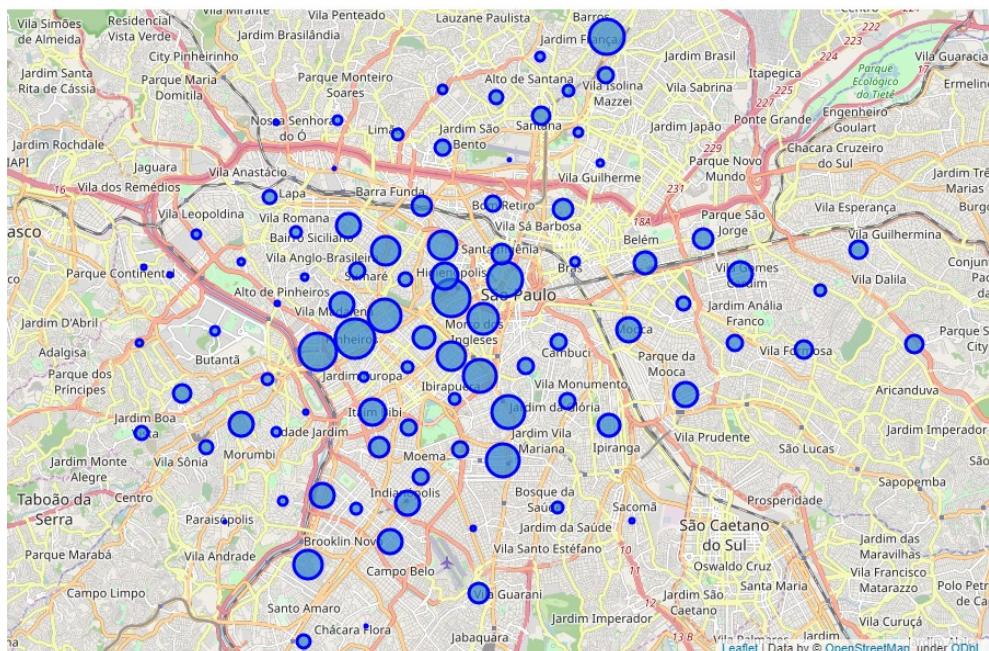
FourSquare returned a list of venues for each neighborhood and price filter specified. Since there were 4 different price (from 1 to 4), this section of the report will show the total venues for each neighborhood and price requested, and their location on the city. Lets start with price equals 1:

--- PRICE = 1 (\$) ---

TOTAL VENUES AND RESTAURANTS PER NEIGHBORHOOD

	Neighborhood	Venues		Neighborhood	Restaurants
0	Pinheiros (Fradique)	41	0	Clinicas	7
1	Consolacao	39	1	Tucuruvi	7
2	Pinheiros (Paes Leme)	39	2	Barra Funda	6
3	Tucuruvi	37	3	Santa Ifigenia	6
4	Ahangabau	36	4	Paraiso	5
5	Vila Clementino	35	5	Santa Cecilia	5
6	Vila Mariana	34	6	Higienopolis	5
7	Paraiso	34	7	Pinheiros (Fradique)	4
8	Clinicas	34	8	Mooca	4
9	Bixiga	32	9	Vila Clementino	4

TOTAL VENUES PER NEIGHBORHOOD



(Bigger radius means more venues in that neighborhood)

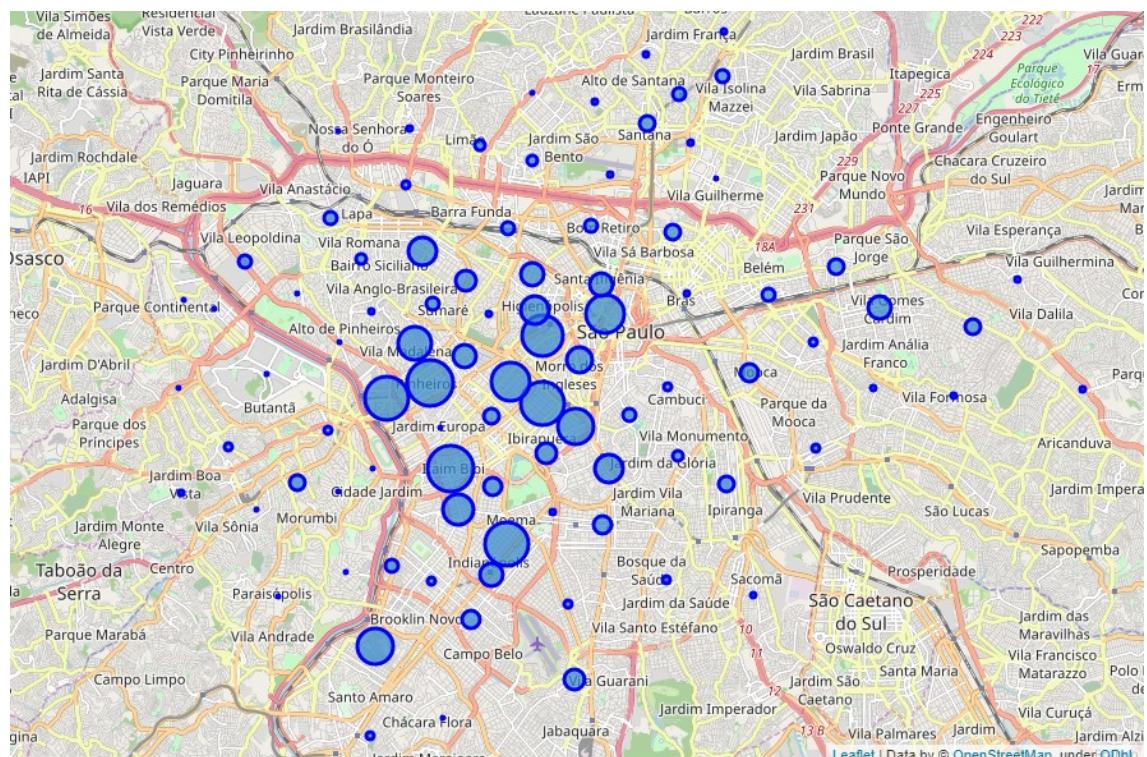
The total venues per neighborhood (radius size on map) has a better distribution through the city, although a little bit more venues are concentrated in the central area, as expected.

--- PRICE = 2 (\$\$) ---

TOTAL VENUES AND RESTAURANTS PER NEIGHBORHOOD

	Neighborhood	Venues		Neighborhood	Restaurants
0	Itaim	83	0	Itaim	59
1	Pinheiros (Fradique)	81	1	Pamplona	52
2	Pamplona	80	2	Pinheiros (Paes Leme)	52
3	Pinheiros (Paes Leme)	78	3	Pinheiros (Fradique)	46
4	Moema	77	4	Ahangabau	45
5	Consolacao	73	5	Moema	45
6	Ahangabau	71	6	Chacara Santo Antonio	38
7	Cerqueira Cesar	69	7	Paraiso	37
8	Chacara Santo Antonio	64	8	Cerqueira Cesar	36
9	Paraiso	64	9	Consolacao	33

TOTAL VENUES PER NEIGHBORHOOD



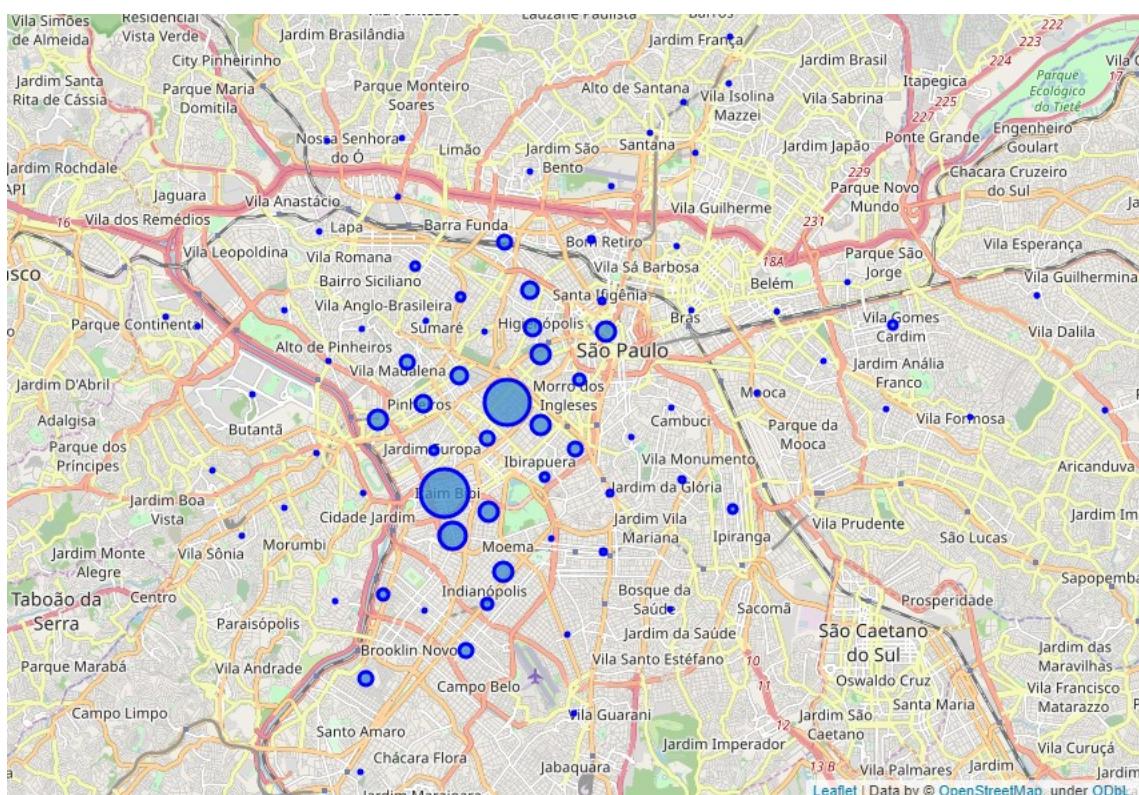
(Bigger radius means more venues in that neighborhood)

--- PRICE = 3 (\$\$\$) ---

TOTAL VENUES AND RESTAURANTS PER NEIGHBORHOOD

	Neighborhood	Venues		Neighborhood	Restaurants
0	Itaim	53	0	Itaim	34
1	Cerqueira Cesar	49	1	Cerqueira Cesar	28
2	Vila Olímpia	30	2	Vila Olímpia	14
3	Ahangabau	22	3	Vila Nova Conceição	11
4	Pamplona	22	4	Jardins	11
5	Pinheiros (Paes Leme)	22	5	Pinheiros (Fradique)	10
6	Consolação	21	6	Moema	9
7	Vila Nova Conceição	20	7	Chácara Santo Antônio	8
8	Moema	20	8	Higienópolis	8
9	Clinicas	18	9	Santa Cecília	8

TOTAL VENUES PER NEIGHBORHOOD



(Bigger radius means more venues in that neighborhood)

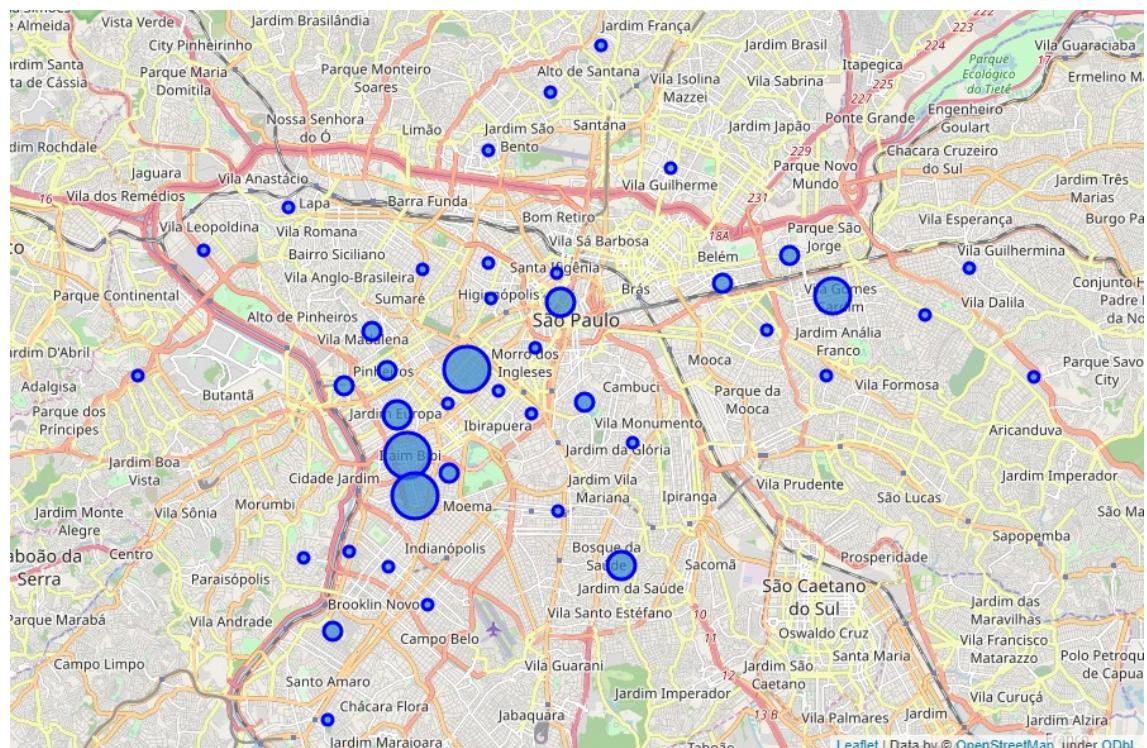
Please note that, as the price range goes up, venues and restaurants tends to concentrate in specific neighborhoods of São Paulo: *Itaim* and *Cerqueira Cesar* in this case

--- PRICE = 4 (\$\$\$\$) ---

TOTAL VENUES AND RESTAURANTS PER NEIGHBORHOOD

	Neighborhood	Venues		Neighborhood	Restaurants
0	Itaim	5	0	Cerqueira Cesar	3
1	Vila Olimpia	5	1	Pinheiros (Fradique)	2
2	Cerqueira Cesar	5	2	Vila Madalena	2
3	Vila Gomes Cardim	4	3	Ahangabau	1
4	Ahangabau	3	4	Bosque da Saude	1
5	Bosque da Saude	3	5	Itaim	1
6	Jardim Europa	3	6	Jardim Europa	1
7	Aclimacao	2	7	Jardins	1
8	Pinheiros (Fradique)	2	8	Pamplona	1
9	Tatuape	2	9	Perdizes	1

TOTAL VENUES PER NEIGHBORHOOD



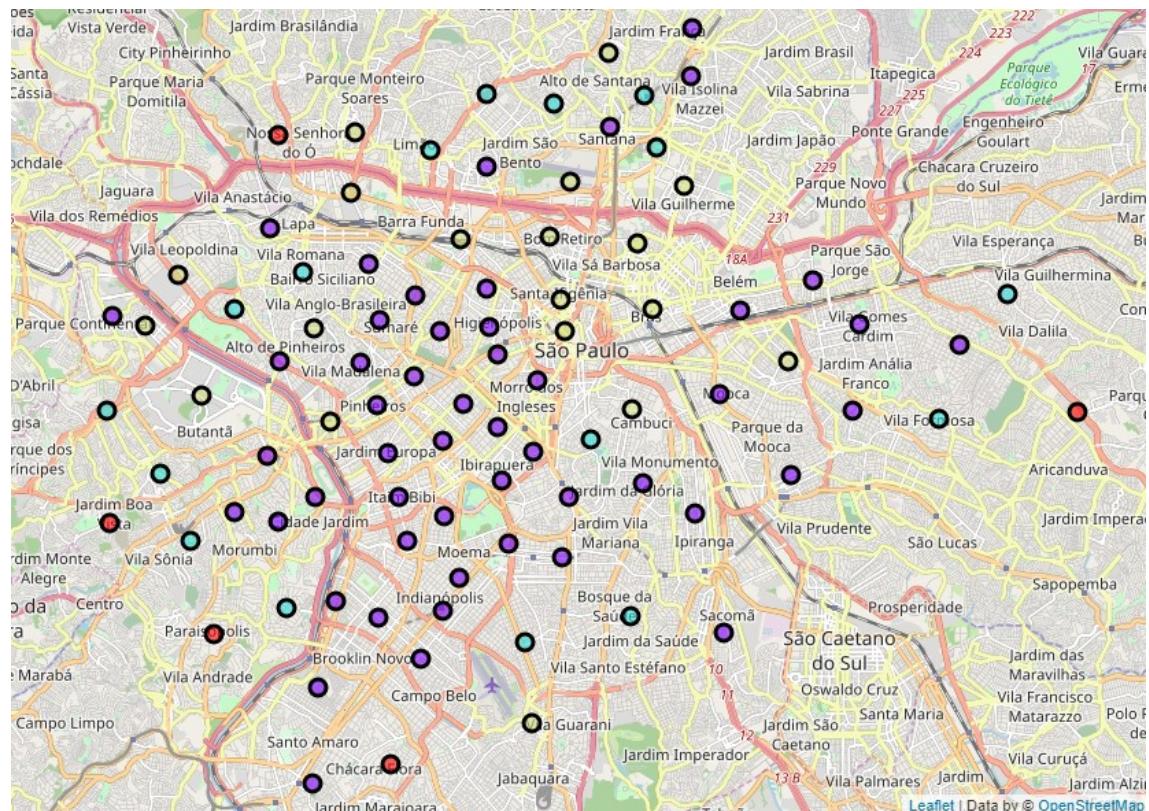
As we can see, the most expensive venues are concentrated mainly in 4 neighborhoods of São Paulo: *Itaim*, *Vila Olimpia*, *Cerqueira Cesar* and *Vila Gomes Cardim*.

4.3 Neighborhoods Clusters

The neighborhoods clusters were generated by K-Means, a machine learning unsupervised method. K-Means yielded the following results for number of clusters = 4:

- Cluster 1: 51 neighborhoods
- Cluster 3: 19 neighborhoods
- Cluster 2: 16 neighborhoods
- Cluster 0: 5 neighborhoods

SAO PAULO NEIGHBORHOOD CLUSTERS



Cluster 1: 51 neighborhoods

4.3.1 – Cluster 1: Top 10 Neighborhoods, sorted by price

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	Price
Cerqueira Cesar	Italian Restaurant	Restaurant	Coffee Shop	Brazilian Restaurant	Bar	Bakery	French Restaurant	Ice Cream Shop	Burger Joint	Pizza Place	2.255172
Itaim	Italian Restaurant	Japanese Restaurant	Bar	Restaurant	Burger Joint	Brazilian Restaurant	Coffee Shop	Middle Eastern Restaurant	Sushi Restaurant	Pizza Place	2.207101
Jardim Europa	Italian Restaurant	Brazilian Restaurant	Middle Eastern Restaurant	Café	Bakery	Japanese Restaurant	Noodle House	Museum	Coffee Shop	Movie Theater	2.200000
Vila Olímpia	Italian Restaurant	Bar	Steakhouse	Restaurant	Middle Eastern Restaurant	Café	Coffee Shop	Nightclub	Pizza Place	Brazilian Restaurant	2.172727
Jardins	Italian Restaurant	Brazilian Restaurant	Bakery	Burger Joint	Middle Eastern Restaurant	Spanish Restaurant	Bar	Café	Portuguese Restaurant	Tapas Restaurant	2.111111
Vila Nova Conceição	Restaurant	Italian Restaurant	Burger Joint	Pizza Place	Coffee Shop	Brazilian Restaurant	Gastropub	Japanese Restaurant	Café	Middle Eastern Restaurant	2.098592
Moema	Brazilian Restaurant	Japanese Restaurant	Italian Restaurant	Bar	Bakery	Sushi Restaurant	Dessert Shop	Restaurant	Pizza Place	Burger Joint	2.035398
Pamplona	Coffee Shop	Middle Eastern Restaurant	Japanese Restaurant	Dessert Shop	Brazilian Restaurant	Italian Restaurant	Burger Joint	Café	Restaurant	Ice Cream Shop	1.962121
Pq Ibirapuera	Bar	Brazilian Restaurant	Italian Restaurant	Café	Burger Joint	Restaurant	Sushi Restaurant	Bakery	Middle Eastern Restaurant	Pizza Place	1.928571
Vila Madalena	Bar	Pizza Place	Restaurant	Ice Cream Shop	Bakery	Brazilian Restaurant	Burger Joint	Café	Cocktail Bar	Vegetarian / Vegan Restaurant	1.923077

Cluster 1: Top 10 Venues for
1st to 3rd most Common Venue:

Bar	23
Brazilian Restaurant	21
Pizza Place	18
Restaurant	16
Italian Restaurant	14
Bakery	12
Dessert Shop	8
Burger Joint	8
Coffee Shop	8
Café	7

As shown above, all top 10 venues from 1st to 3rd most common venues are food related.

From all clusters ordered by price evaluation of the customer, Cluster 1 is the only one having Italian Restaurants on top 5 positions of 1st most common venues. According to customer evaluation, Italian Restaurants are the most expensive common venue in this cluster.

Cluster 3: 19 neighborhoods

4.3.2 – Cluster 3: Top 10 Neighborhoods, sorted by price

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	Price
Parque Anhembi	Brazilian Restaurant	Bar	Nightclub	Churrascaria	Snack Place	Café	Brewery	Food Truck	Beer Garden	Hotel Bar	2.000000
Aqua Branca	Brazilian Restaurant	Restaurant	Bar	Fast Food Restaurant	Pizza Place	Food Truck	Coffee Shop	BBQ Joint	Furniture / Home Store	Churrascaria	2.000000
Ahangabau	Brazilian Restaurant	Café	Coffee Shop	Bakery	Bar	Restaurant	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Italian Restaurant	Dessert Shop	1.939394
Pinheiros (Paes Leme)	Brazilian Restaurant	Burger Joint	Restaurant	Coffee Shop	Bar	Japanese Restaurant	Café	Italian Restaurant	Nightclub	Vegetarian / Vegan Restaurant	1.907801
Barra Funda	Nightclub	Restaurant	Brazilian Restaurant	Fast Food Restaurant	Café	Sandwich Place	Italian Restaurant	Bar	Bakery	Japanese Restaurant	1.896552
Favela Sao Remo (Jaguaré)	Brazilian Restaurant	Japanese Restaurant	Café	Fast Food Restaurant	Lounge	Snack Place	Churrascaria	Restaurant	Nightclub	Pizza Place	1.846154
Santa Ifigenia	Brazilian Restaurant	Bar	Middle Eastern Restaurant	Peruvian Restaurant	Bakery	Café	Coffee Shop	Vegetarian / Vegan Restaurant	Sandwich Place	Dessert Shop	1.842857
Vila Leopoldina	Brazilian Restaurant	Food Truck	Dessert Shop	Pizza Place	Japanese Restaurant	BBQ Joint	Sushi Restaurant	Burger Joint	Restaurant	Buffet	1.794118
Bom Retiro	Brazilian Restaurant	Korean Restaurant	Café	Restaurant	Diner	Dessert Shop	Bar	Coffee Shop	Pizza Place	Greek Restaurant	1.744681
Vila Beatriz	Brazilian Restaurant	Bakery	Pizza Place	Restaurant	Fast Food Restaurant	Japanese Restaurant	Farmers Market	Bar	Cupcake Shop	Gluten-free Restaurant	1.722222

**Cluster 3: Top 10 Venues for
1st to 3rd most Common Venue:**

Brazilian Restaurant	18
Restaurant	7
Café	6
Bar	4
Bakery	4
Pizza Place	2
Nightclub	2
Food Truck	2
Middle Eastern Restaurant	2
Burger Joint	2

Similar to Cluster 1, most of the top 10 venues are also food related venues. Restaurants also appear 2 times on top positions.

Cluster 2: 16 neighborhoods

4.3.3 – Cluster 2: Top 10 Neighborhoods, sorted by price

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	Price
Bosque da Saude	Bar	Bakery	Pizza Place	Asian Restaurant	Dessert Shop	Steakhouse	Deli / Bodega	Fast Food Restaurant	Chinese Restaurant	Molecular Gastrology Restaurant	1.903226
Aclimacao	Pizza Place	Bar	Bakery	BBQ Joint	Korean Restaurant	Sushi Restaurant	Dessert Shop	Chinese Restaurant	Steakhouse	Burger Joint	1.804878
Planalto Paulista	Pizza Place	Restaurant	Bakery	Bar	Brazilian Restaurant	Sandwich Place	Steakhouse	Italian Restaurant	BBQ Joint	Tea Room	1.777778
Jardim São Paulo	Bar	Pizza Place	Japanese Restaurant	Bakery	Brazilian Restaurant	Food Truck	Nightclub	Restaurant	Churrascaria	Fast Food Restaurant	1.769231
Real Parque	Bakery	Café	Pizza Place	Hotel Bar	Japanese Restaurant	Middle Eastern Restaurant	Nightclub	Restaurant	Chinese Restaurant	Burger Joint	1.750000
Carandiru	Pizza Place	Bakery	Bar	Brazilian Restaurant	Italian Restaurant	Café	Restaurant	Snack Place	Seafood Restaurant	Japanese Restaurant	1.750000
Boacava	Snack Place	Sushi Restaurant	Bakery	Speakeasy	Japanese Restaurant	Burger Joint	Juice Bar	Brazilian Restaurant	Pizza Place	Falafel Restaurant	1.666667
Vila Romana	Pizza Place	Bar	Bakery	Café	Brazilian Restaurant	Ice Cream Shop	Bagel Shop	Deli / Bodega	Coffee Shop	Restaurant	1.620690
Rio Pequeno	Bakery	Pizza Place	Restaurant	Brazilian Restaurant	Burger Joint	Food Truck	Japanese Restaurant	Bar	Fish Market	Fish & Chips Shop	1.615385
Limao	Restaurant	Bakery	Bar	Burger Joint	Food Truck	Japanese Restaurant	Diner	Pizza Place	Brazilian Restaurant	Sushi Restaurant	1.600000

Cluster 2: Top 10 Venues for
1st to 3rd most Common Venue:

Bakery	15
Pizza Place	12
Bar	10
Restaurant	4
Café	1
Sushi Restaurant	1
Snack Place	1
BBQ Joint	1
Burger Joint	1
Japanese Restaurant	1

This cluster shows a lower presence of restaurants from 1st to 3rd most common venues.

Cluster 0: 5 neighborhoods

4.3.4 – Cluster 0: Top 10 Neighborhoods, sorted by price

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	Price
Favela Paraisópolis	Pizza Place	Burger Joint	Fast Food Restaurant	Bakery	Brazilian Restaurant	Falafel Restaurant	Flower Shop	Fish Taverna	Fish Market	Fish & Chips Shop	1.500000
Favela Princesinha (Shop.Raposo)	Café	Pizza Place	Brazilian Restaurant	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Juice Bar	Restaurant	Sandwich Place	Food Truck	1.434783
Jardim Aricanduva	Fast Food Restaurant	Dessert Shop	Bakery	Café	Burger Joint	Italian Restaurant	Fried Chicken Joint	Pizza Place	Sandwich Place	Brewery	1.428571
Piqueri	Pizza Place	BBQ Joint	Bakery	Dessert Shop	Fast Food Restaurant	Italian Restaurant	Sandwich Place	Juice Bar	Fish & Chips Shop	Farmers Market	1.727273
Chacara Flora	Hookah Bar	Fast Food Restaurant	Restaurant	Pizza Place	Gym / Fitness Center	Doner Restaurant	Farmers Market	Falafel Restaurant	English Restaurant	Empanada Restaurant	1.250000

Cluster 0: Top 10 Venues for
1st to 3rd most Common Venue:

Pizza Place	3
Fast Food Restaurant	3
Bakery	2
Dessert Shop	1
BBQ Joint	1
Restaurant	1
Burger Joint	1
Brazilian Restaurant	1
Café	1
Hookah Bar	1

This is the cluster with the lowest presence of restaurants from 1st to 3rd most common venues.

This cluster also show much less venues when compared with other clusters in São Paulo.

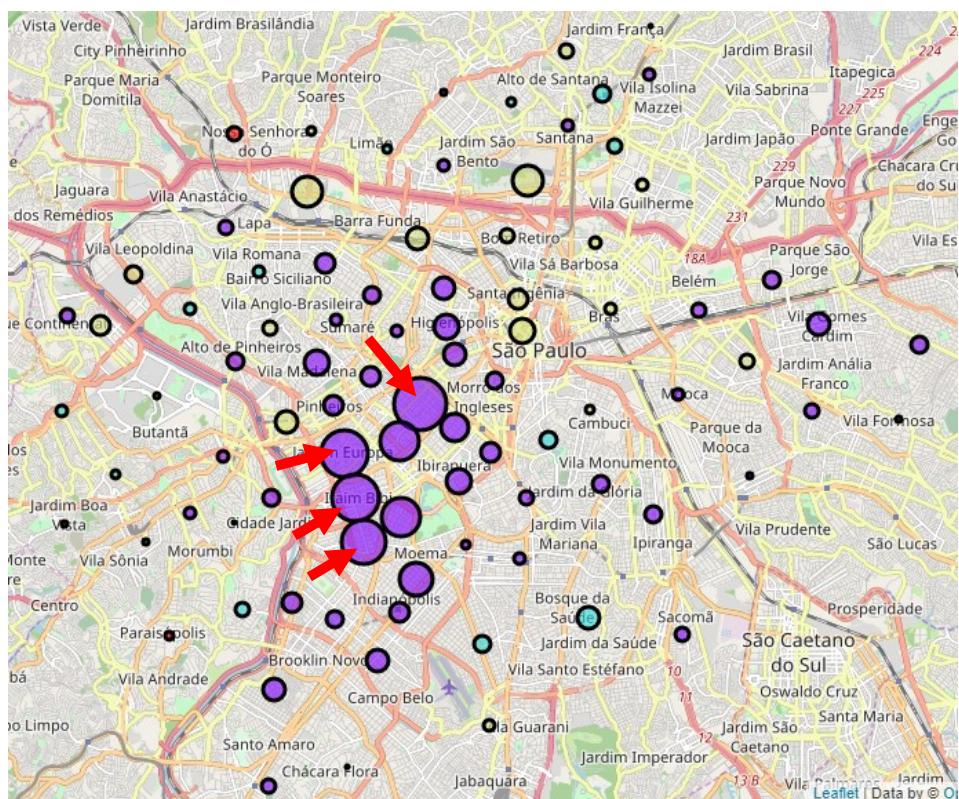
4.4 Neighborhoods Clusters x Venue Prices

The goal of this work is to find the best places in town do open a Italian restaurant, for customers with medium to high income. So, let's explore the data we collected and see if there is a relationship between venues price evaluation and clusters identified by K-Means.

To do this, let's plot 3 different folium maps of neighborhoods clusters, and set the radius of each neighborhood on the map equals to:

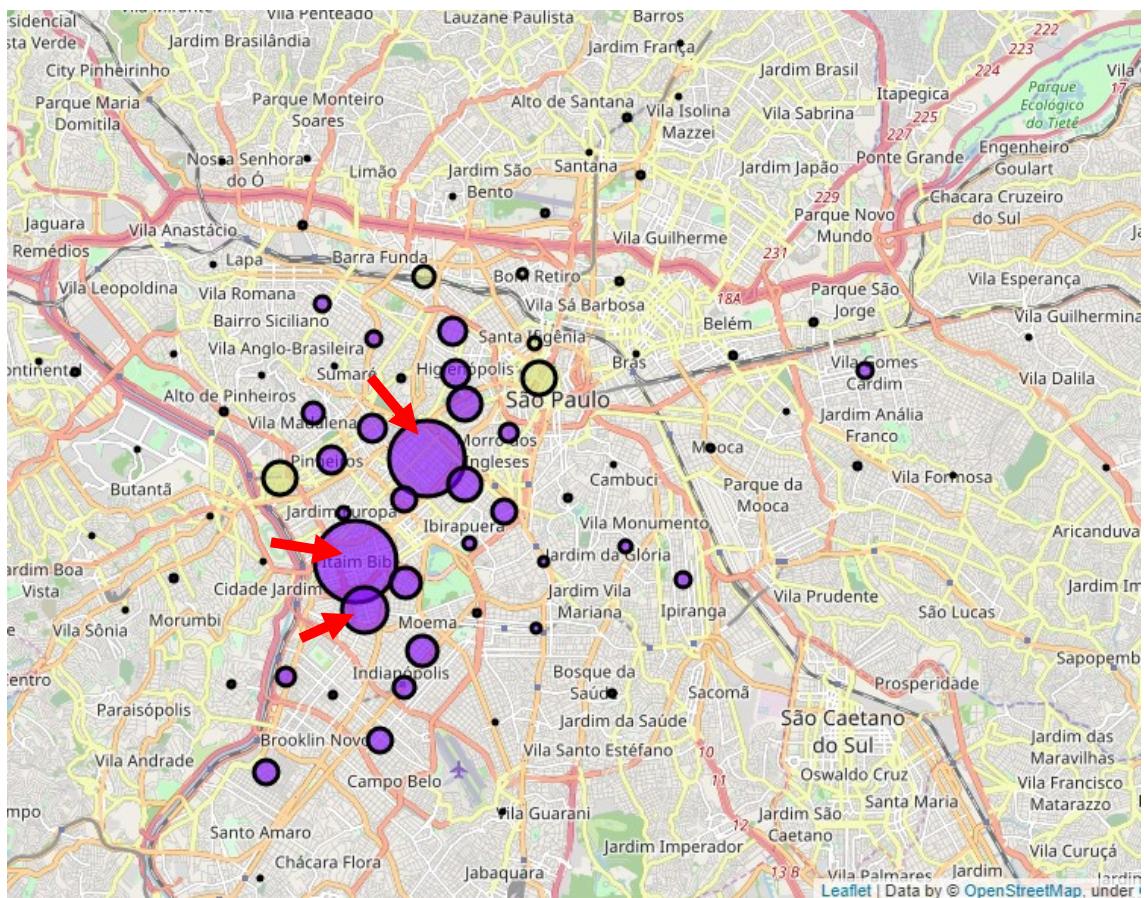
- A) Average venue price for each neighborhood
- B) Total neighborhood venues with price = 3
- C) Total neighborhood venues with price = 4

4.4.1 Clusters x Average Venue Price for each Neighborhood



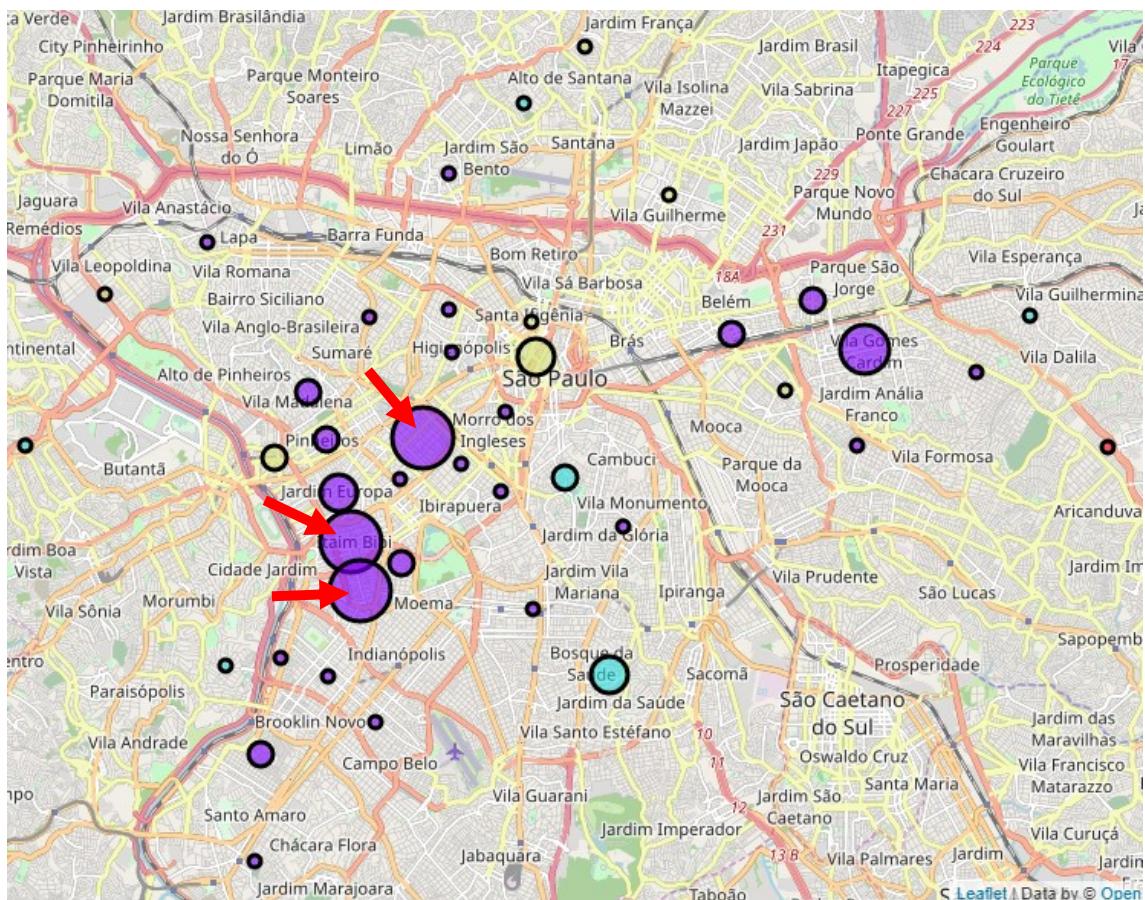
Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Price
Cerqueira Cesar	1	Italian Restaurant	Restaurant	Coffee Shop	2.255172
Itaim	1	Italian Restaurant	Japanese Restaurant	Bar	2.207101
Jardim Europa	1	Italian Restaurant	Brazilian Restaurant	Middle Eastern Restaurant	2.200000
Vila Olímpia	1	Italian Restaurant	Bar	Steakhouse	2.172727
Jardins	1	Italian Restaurant	Brazilian Restaurant	Bakery	2.111111
Vila Nova Conceicao	1	Restaurant	Italian Restaurant	Burger Joint	2.098592
Moema	1	Brazilian Restaurant	Japanese Restaurant	Italian Restaurant	2.035398
Parque Anhembi	3	Brazilian Restaurant	Bar	Nightclub	2.000000
Agua Branca	3	Brazilian Restaurant	Restaurant	Bar	2.000000
Pamplona	1	Coffee Shop	Middle Eastern Restaurant	Japanese Restaurant	1.962121

4.4.2 Clusters x Total Neighborhood Venues with Price = 3



Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Venues
Itaim	1	Italian Restaurant	Japanese Restaurant	Bar	53
Cerqueira Cesar	1	Italian Restaurant	Restaurant	Coffee Shop	49
Vila Olímpia	1	Italian Restaurant	Bar	Steakhouse	30
Pinheiros (Paes Leme)	3	Brazilian Restaurant	Burger Joint	Restaurant	22
Ahangabau	3	Brazilian Restaurant	Café	Coffee Shop	22
Pampulha	1	Coffee Shop	Middle Eastern Restaurant	Japanese Restaurant	22
Consolação	1	Bar	Café	Brazilian Restaurant	21
Vila Nova Conceição	1	Restaurant	Italian Restaurant	Burger Joint	20
Moema	1	Brazilian Restaurant	Japanese Restaurant	Italian Restaurant	20
Clinicas	1	Restaurant	Bar	Pizza Place	18

4.4.3 Clusters x Total Neighborhood Venues with Price = 4



Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Venues
Cerqueira Cesar	1	Italian Restaurant	Restaurant	Coffee Shop	5
Itaim	1	Italian Restaurant	Japanese Restaurant	Bar	5
Vila Olímpia	1	Italian Restaurant	Bar	Steakhouse	5
Vila Gomes Cardim	1	Pizza Place	Bar	Dessert Shop	4
Bosque da Saúde	2	Bar	Bakery	Pizza Place	3
Ahangabau	3	Brazilian Restaurant	Café	Coffee Shop	3
Jardim Europa	1	Italian Restaurant	Brazilian Restaurant	Middle Eastern Restaurant	3
Tatuape	1	Restaurant	Pizza Place	Coffee Shop	2
Pinheiros (Fradique)	1	Italian Restaurant	Brazilian Restaurant	Restaurant	2
Aclimacão	2	Pizza Place	Bar	Bakery	2

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Venues
Cerqueira Cesar	1	Italian Restaurant	Restaurant	Coffee Shop	5
Itaim	1	Italian Restaurant	Japanese Restaurant	Bar	5
Vila Olímpia	1	Italian Restaurant	Bar	Steakhouse	5
Vila Gomes Cardim	1	Pizza Place	Bar	Dessert Shop	4
Bosque da Saúde	2	Bar	Bakery	Pizza Place	3
Ahangabau	3	Brazilian Restaurant	Café	Coffee Shop	3
Jardim Europa	1	Italian Restaurant	Brazilian Restaurant	Middle Eastern Restaurant	3
Tatuape	1	Restaurant	Pizza Place	Coffee Shop	2
Pinheiros (Fradique)	1	Italian Restaurant	Brazilian Restaurant	Restaurant	2
Aclimacão	2	Pizza Place	Bar	Bakery	2

5. Discussion

Italian restaurants are concentrated on the central area of São Paulo, mostly on *Itaim* and *Cerqueira Cesar* neighborhoods, followed by *Pinheiros*, *Chacara Santo Antonio*, *Jardins* and *Vila Olímpia*, as shown on item 4.1 of this report.

But knowing the top neighborhoods with Italian Restaurants is not enough to choose the best Neighborhood to start a new one. We also need to know the neighborhoods a little better, and see if there are groups of neighborhoods in the city with similar characteristics, for example, groups of neighborhoods that people tend to come to eat, have fun, etc.

K-Means helped to identify these groups, clustering neighborhoods with similar characteristics. Let's explore the similarities and dissimilarities between them:

- Cluster 1:
 - 1st cluster in size, with 51 neighborhoods.
 - All top 10 venues from 1st to 3rd most common venues of this cluster are food related.
 - Italian Restaurants only appear on this cluster from 1st to 3rd most common venues.
- Cluster 3:
 - 2nd biggest cluster with 19 neighborhoods.
 - Top 10 venues are also food related venues.
 - Restaurants frequency is lower than Cluster 1.
- Cluster 2:
 - 3rd cluster in size, with 16 neighborhoods.
 - Top 10 venues are also food related venues.
 - Restaurants are not the main activities on these neighborhoods, which are bakery and Pizza places.
- Cluster 0:
 - Smallest cluster in town, with 5 neighborhoods.
 - Lowest number of restaurants from 1st to 3rd most common venues.
 - Also have much less venues when compared to other clusters.

Neighborhoods in cluster 0 has the lowest number of venues in the city. When this cluster was sorted by customer price evaluation, please note the first 2 neighborhoods are *favelas* (brazilian name for shantytown), extremely poor areas in the city, densely populated but with poor services available.

Since the low number of venues is a characteristic of this cluster, maybe that's why K-Means grouped shantytowns into this cluster. Other clusters also with low number of venues, such as *Chacara Flora*, but this time a very rich area in the city, as also grouped in this cluster for the same reason – it's a residential area, with low number of commerce and venues available.

Since people who live in *favelas* are poor, they tend to see prices as more expensive than other inhabitants of the same cluster – that's the reason why the top 2 neighborhoods of this cluster (sorted by price) are shantytowns. *Favela Paraisópolis* is the 2nd biggest *favela* in São Paulo, with more than 100,000 inhabitants. In the other extreme is *Chacara Flora*, rated with very cheap prices inside cluster 0 – probably because money is not a problem for people who live there.

Folium maps presented on item 4.2 show cheapest venues (price = 1) more evenly distributed over the city, when compared to venues evaluated as more expensive.

As the price evaluation increase, venues tend to agglomerate on richest neighborhoods. This is very clear for price equals 3 and 4 - please see results on item 4.2.

From all clusters ordered by price evaluation of the customer, Cluster 1 is the only one having Italian Restaurants on top 5 positions of 1st most common venues – please see item 4.3.1. According to customer evaluation, Italian Restaurants are the most expensive common venue in this cluster.

As an example, *Itaim*, *Cerqueira César* and *Vila Olímpia* are the top 3 neighborhoods with the most number of expensive venues and restaurants on the city. For price equals 3, *Vila Gomes Cardim* occupy the 4th place, with a considerable number of venues evaluated as expensive. These neighborhoods could be a good location to start our Italian Restaurant.

So, what are best places in town to open an Italian restaurant for customers with medium to high income?

Results from K-Means have shown that cluster neighborhoods that have restaurants in the top common venues are good candidates, that is, cluster 1.

After filtering data obtained from FourSquare, we saw that neighborhoods with the most number of expensive venues (in customer evaluation) are also good candidates.

After cross referencing both of these results, we could have a very clear view where to open our Italian restaurant in São Paulo. This is clearly shown on Folium maps 4.1.1, 4.1.2 and 4.1.3.

They all point to the same conclusion, the top 3 neighborhoods to open an Italian restaurant in São Paulo are:

1. *Cerqueira César*
2. *Itaim*
3. *Vila Olímpia*

6. Conclusion

The aim of this study was to identify the best places in São Paulo to open an Italian Restaurant, for people from medium to high income.. To achieve this, I analyzed the characteristics of 91 neighborhoods in town.

Data acquired from FourSquare was stored in a *dataframe* with 4,647 venues, 153 unique venues categories, 4 different price tags and 91 neighborhoods.

The first characteristic to be analyzed was venue price: I want to identify which neighborhoods had the most number of expensive venues in town. This was easily achieved by counting how many venues of a specific price tag were present in each neighborhood.

The second characteristic to be analyzed was neighborhood clusters, that should be grouped by similarities and separated by dissimilarities. To do this, I applied a machine learning unsupervised method (K-Means). One of the 4 clusters generated revealed a characteristic I was looking for: neighborhoods that offer good places to eat.

But the best insights were revealed when I cross reference venues prices with neighborhood clusters. The solution to the initial problem became clear, and 3 neighborhoods out of initial 91 were revealed as the best places in town to open an Italian Restaurant, considering as customers people with medium to high income:

1. *Cerqueira Cesar*
2. *Itaim*
3. *Vila Olímpia*

7. References

1. [*FourSquare API*](#)
2. [*Google Maps*](#)