Applied Data Science Capstone

(Coursera – IBM Course)

Final Assignment - The Battle of Neighborhoods

Clusters of boroughs in

Rio de Janeiro and São Paulo

Author:

Luciana Guerra

March, 2021

1. INTRODUCTION

São Paulo and Rio de Janeiro are two of the most important cities in Brazil and, perhaps, in the world.

São Paulo is the main financial and corporate center in South America and also the most populous city in the Southern Hemisphere, with more than 12 million inhabitants. Some cultural sites are widely known, such as "Memorial da América Latina", "Parque Ibirapuera" and "Avenida Paulista".

Rio de Janeiro is the main tourist destination in South America, mainly because of the most famous event: the carnival. It is the second largest city in Brazil and has almost 7 million inhabitants. Here are some famous postcards: "Maracanã", "Arcos da Lapa", "Corcovado", "Pão de Açúcar" and "Praia de Conpacabana".

It is very common for people to have to move from one of these cities to the other. Perhaps because of work, studies or even health care.

Although there are many similar characteristics, the lifestyle in Rio and São Paulo is very different. Thus, when it comes to moving from one to the other, it is good to choose a neighborhood with similar locations, so that adaptation is easier.

In this study, I will organize the neighborhoods of the two cities into clusters according to the headquarters of Foursquare.

2. DATA

For the present study, we will need the names of the boroughs of São Paulo and Rio de Janeiro, as well as the latitude and longitude of each one. In addition, we will use the Foursquare API to discover the most frequent category of locations in each neighborhood.

Below, I present where the data will be obtained.

- Boroughs of São Paulo: https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_S%C3%A3o_Paulo_por_popula%C3%A7%C3%A3o
- Boroughs of Rio de Janeiro: https://www.data.rio/datasets/limite-de-bairros/data?geometry=-44.752%2C-23.138%2C-42.139%2C-22.695&page=6

- Geographic coordinates: https://batchgeo.com/
- Most frequent category of locations in each neighborhood: Foursquare API

3. METHODOLOGY

Initially, I imported the names of the neighborhoods of the two cities from the mentioned sources and created a dataframe for each city. In São Paulo there are 96 neighborhoods and in Rio de Janeiro, 163.

For the neighborhoods in Rio, I was unable to import directly from the url mentioned, so I downloaded the csv, put it on GitHub and read the csv using pandas.

From the neighborhood names, I generated a kml file using the BatchGeo website and then created an xls file containing the neighborhoods and their geographic coordinates. I made each city file available on my GitHub and read it using pandas.

Using the folium, I generated two maps, one for each city, marking the respective neighborhoods.

Using a function, I accessed the Foursquare API to register the locations of the neighborhoods on a dataframe, up to a maximum number of 100 locations, considering a radius of 500 meters from the geographic coordinates of the neighborhood.

Analyzing the result, it was possible to know the 10 categories of most frequent locations in each neighborhood. See the result for 15 neighborhoods:

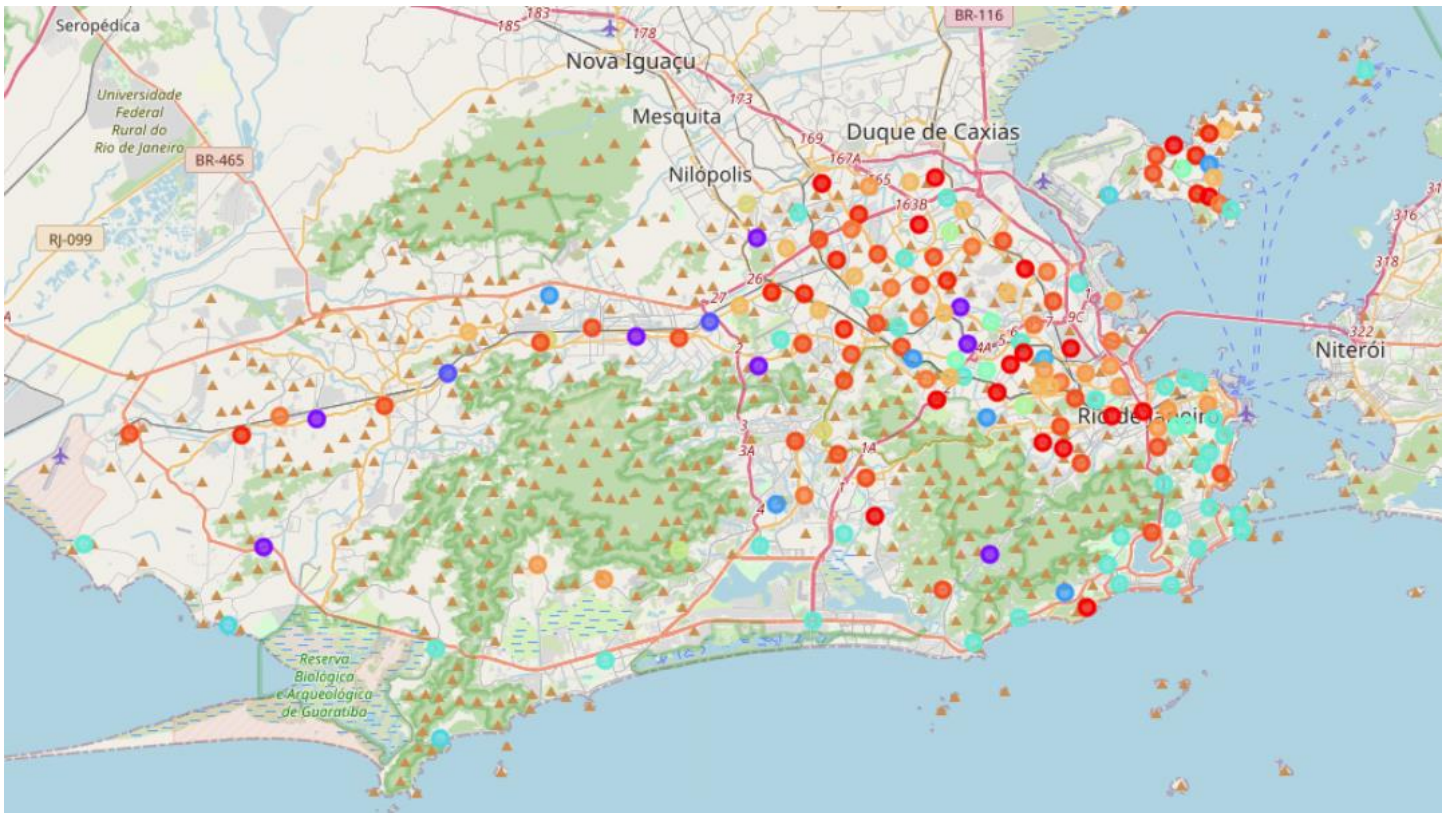
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abolição	BBQ Joint	Pizza Place	Deli / Bodega	Plaza	Dive Bar	Burger Joint	Food Truck	Soccer Field	Snack Place	Bar
1	Acari	Metro Station	Sushi Restaurant	Flea Market	Coffee Shop	Snack Place	Market	Pizza Place	Food Truck	Outdoor Sculpture	Other Nightlife
2	Alto da Boa Vista	Soccer Field	Big Box Store	Bakery	Fast Food Restaurant	Church	Nightclub	Other Great Outdoors	Pastelaria	Peruvian Restaurant	Optical Shop
3	Alto de Pinheiros	Convenience Store	Plaza	Bar	Supermarket	Flea Market	Tennis Court	Coffee Shop	Fast Food Restaurant	Trail	Market
4	Anchieta	Department Store	Food Truck	Breakfast Spot	Soccer Field	Park	Comfort Food Restaurant	Middle Eastern Restaurant	Diner	Other Repair Shop	Pastelaria
5	Andaraí	Bar	Pizza Place	Bakery	Sandwich Place	Dive Bar	Music Venue	Fruit & Vegetable Store	Hotel	Food Truck	Gym / Fitness Center
6	Anhanguera	Snack Place	Burger Joint	Grocery Store	Pizza Place	Bakery	ATM	Paella Restaurant	Pedestrian Plaza	Pastry Shop	Pastelaria
7	Anil	Bar	Brazilian Restaurant	BBQ Joint	Burger Joint	Northeastern Brazilian Restaurant	Pizza Place	Sandwich Place	Nightclub	Japanese Restaurant	Fried Chicken Joint
8	Aricanduva	Pizza Place	Grocery Store	Gym / Fitness Center	Bakery	ATM	Outlet Store	Pedestrian Plaza	Pastry Shop	Pastelaria	Park
9	Artur Alvim	Sporting Goods Shop	Coffee Shop	Hot Dog Joint	Gymnastics Gym	ATM	Paella Restaurant	Performing Arts Venue	Pedestrian Plaza	Pastry Shop	Pastelaria
10	Bancários	Food Truck	Plaza	Burger Joint	Pizza Place	Arts & Crafts Store	Pet Store	Pedestrian Plaza	Snack Place	BBQ Joint	Comfort Food Restaurant
11	Bangu	Fast Food Restaurant	Bar	Restaurant	Accessories Store	Pharmacy	Seafood Restaurant	Toy / Game Store	Pie Shop	Construction & Landscaping	BBQ Joint
12	Barra Funda	Bar	Restaurant	Nightclub	Plaza	Theater	Martial Arts School	Brazilian Restaurant	Sandwich Place	Fast Food Restaurant	Museum
13	Barra da Tijuca	Pet Store	Art Gallery	Farmers Market	Theater	Thai Restaurant	Park	Ice Cream Shop	Café	Music Venue	Other Repair Shop
14	Barra de Guaratiba	South American Restaurant	Health Food Store	Snack Place	Bus Station	ATM	Pedestrian Plaza	Pastry Shop	Pastelaria	Park	Paper / Office Supplies Store

Using the scikit-learn library, I organized the neighborhoods into 20 clusters. Let's see which group the first 15 neighborhoods are in:

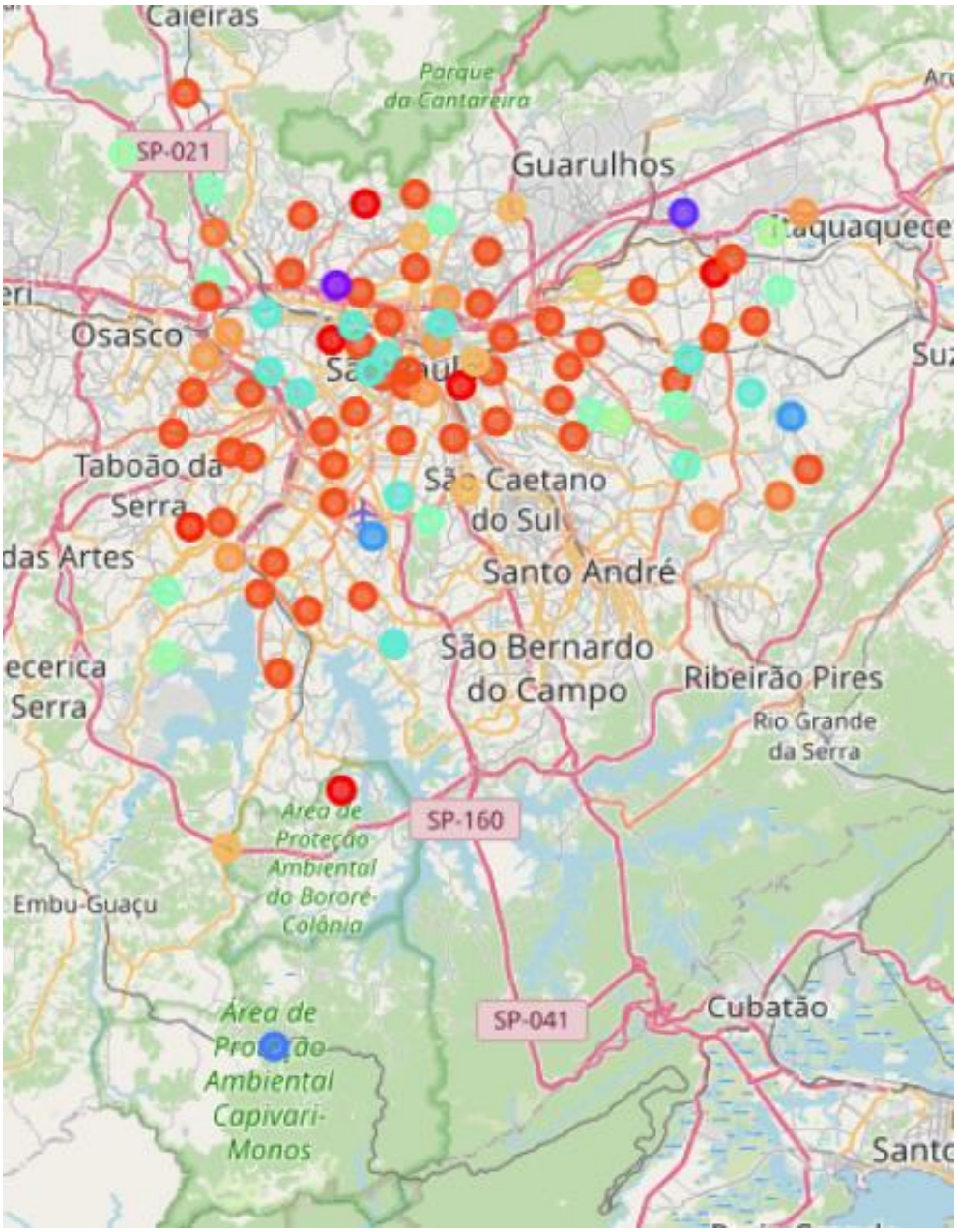
	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Paquetá	Rio de Janeiro	-22.763625	-43.108358	8.0	Beach	Brazilian Restaurant	Harbor / Marina	Park	Boarding House	Food & Drink Shop	Pier	Hotel	Market	Bar
1	Freguesia (Ilha)	Rio de Janeiro	-22.789645	-43.173205	15.0	Farmers Market	Bar	Burger Joint	Gym	Food Truck	Pizza Place	ATM	Paintball Field	Pedestrian Plaza	Pastry Shop
2	Bancários	Rio de Janeiro	-22.790984	-43.180925	19.0	Food Truck	Plaza	Burger Joint	Pizza Place	Arts & Crafts Store	Pet Store	Pedestrian Plaza	Snack Place	BBQ Joint	Comfort Food Restaurant
3	Galeão	Rio de Janeiro	-22.817114	-43.227823	7.0	Seafood Restaurant	Churrascaria	Steakhouse	Food Truck	Gym	Gym / Fitness Center	Comedy Club	Pizza Place	Paella Restaurant	Pedestrian Plaza
4	Tauá	Rio de Janeiro	-22.800505	-43.186953	19.0	Tanning Salon	Plaza	Bakery	Farmers Market	Pharmacy	Martial Arts School	Food Truck	Mountain	Shop & Service	Optical Shop
5	Portuguesa	Rio de Janeiro	-22.799995	-43.205779	18.0	Department Store	Clothing Store	Plaza	Fast Food Restaurant	Cosmetics Shop	Chocolate Shop	Brazilian Restaurant	Italian Restaurant	Bar	Mobile Phone Shop
6	Moneró	Rio de Janeiro	-22.795768	-43.197717	0.0	Park	Bar	Dive Bar	Creperie	Bakery	Food Truck	ATM	Paper / Office Supplies Store	Peruvian Restaurant	Perfume Shop
7	Vigário Geral	Rio de Janeiro	-22.809455	-43.308952	0.0	Samba School	Bar	Snack Place	Music Venue	Cave	Pizza Place	ATM	Pedestrian Plaza	Pastry Shop	Pastelaria
8	Cocotá	Rio de Janeiro	-22.803838	-43.180770	5.0	Gym / Fitness Center	Shopping Plaza	Fast Food Restaurant	Park	Paintball Field	Sandwich Place	Gym Pool	Tanning Salon	Skate Park	Burger Joint
9	Jardim América	Rio de Janeiro	-22.811487	-43.320718	15.0	Pharmacy	Plaza	Gym	Pizza Place	Steakhouse	Japanese Restaurant	Burger Joint	Snack Place	Sandwich Place	Video Store
10	Jardim Carioca	Rio de Janeiro	-22.805766	-43.193233	10.0	Pizza Place	Brazilian Restaurant	Plaza	Burger Joint	Boutique	Gym	Sushi Restaurant	Gym / Fitness Center	Tea Room	Pharmacy
11	Pavuna	Rio de Janeiro	-22.812226	-43.361972	0.0	Bakery	Bar	Gym / Fitness Center	Gym	Theater	Health & Beauty Service	Convenience Store	Department Store	Burger Joint	Other Repair Shop
12	Cordovil	Rio de Janeiro	-22.823441	-43.295936	15.0	Record Shop	Grocery Store	Gym	Food Stand	Outlet Store	Pedestrian Plaza	Pastry Shop	Pastelaria	Park	Paper / Office Supplies Store
13	Jardim Guanabara	Rio de Janeiro	-22.808059	-43.207250	18.0	Gym / Fitness Center	Burger Joint	Pharmacy	Japanese Restaurant	Brazilian Restaurant	Pizza Place	Sandwich Place	Health Food Store	Pet Store	Health & Beauty Service
14	Parada de Lucas	Rio de Janeiro	-22.818451	-43.303014	8.0	ATM	Grocery Store	Train Station	Tea Room	Fruit & Vegetable Store	Furniture / Home Store	Bar	Brazilian Restaurant	Organic Grocery	Other Great Outdoors

Using *folium*, I generated a map with the neighborhoods identified with the color corresponding to your cluster:

Rio de Janeiro:



São Paulo:



In addition, to get to know each cluster better, I visualized individual tables for each group (see in the next section).

4. RESULTS

Next, I will present the size and composition of the clusters. We can see that there are 10 clusters with less than 5 neighborhoods and 2 clusters with more than 49 neighborhoods.

Number of Neighborhoods	City		
nº Cluster	Rio de Janeiro	São Paulo	Total Geral
18	28	45	73
8	38	11	49
0	17	5	22
15	14	5	19
16	11	8	19
17	12	2	14
10	4	8	12
19	10	1	11
1	8	1	9
5	6	2	8
14	3	1	4
12	2	2	4
7	3		3
9	2	1	3
11		2	2
3	2		2
13	1		1
2		1	1
4		1	1
6	1		1
Total Geral	**162**	**96**	**258**

- **0.0**
  - **Rio de Janeiro**
    - Água Santa
    - Andaraí
    - Anil
    - Cachambi
    - Grajaú
    - Manguinhos
    - Maracanã
    - Maria da Graça
    - Méier
    - Moneró
    - Olaria
    - Pavuna
    - Pitangueiras
    - Praça da Bandeira
    - Vidigal
    - Vigário Geral
    - Vista Alegre
  - **São Paulo**
    - Cachoeirinha
    - Grajaú
    - Mooca
    - Perdizes
    - Vila Jacuí

- **1.0**
  - **Rio de Janeiro**
    - Alto da Boa Vista
    - Engenho da Rainha
    - Guaratiba
    - Inhoaíba
    - Jardim Sulacap
    - Padre Miguel
    - Pilares
    - Ricardo de Albuquerque
  - **São Paulo**
    - Limão

- **2.0**
  - **São Paulo**
    - Ermelino Matarazzo

- **3.0**
  - **Rio de Janeiro**
    - Magalhães Bastos
    - Senador Vasconcelos

- **4.0**
  - **São Paulo**
    - Marsilac

- **5.0**
  - **Rio de Janeiro**
    - Cocotá
    - Curicica
    - Lins de Vasconcelos
    - Quintino Bocaiúva
    - Rocinha
    - Vila Kennedy
  - **São Paulo**
    - Guaianases
    - Jabaquara

- **6.0**
  - **Rio de Janeiro**
    - Jacarezinho

- **7.0**
  - **Rio de Janeiro**
    - Galeão
    - Grumari
    - Pedra de Guaratiba

- **9.0**
  - **Rio de Janeiro**
    - Catumbi
    - Santa Teresa
  - **São Paulo**
    - São Mateus

- **10.0**
  - **Rio de Janeiro**
    - Abolição
    - Inhaúma
    - Jardim Carioca
    - Todos os Santos
  - **São Paulo**
    - Capão Redondo
    - Cursino
    - Jaraguá
    - Parque do Carmo
    - São Domingos
    - Sapopemba
    - Tucuruvi
    - Vila Curucá

- **11.0**
  - **São Paulo**
    - Anhanguera
    - Jardim Ângela

- **12.0**
  - **Rio de Janeiro**
    - Brás de Pina
    - Engenho Novo
  - **São Paulo**
    - Aricanduva
    - Jardim Helena

- **13.0**
  - **Rio de Janeiro**
    - Camorim

- **14.0**
  - **Rio de Janeiro**
    - Parque Anchieta
    - Senador Camará
    - Tanque
  - **São Paulo**
    - Cangaíba

- **15.0**
  - **Rio de Janeiro**
    - Bento Ribeiro
    - Complexo do Alemão
    - Cordovil
    - Encantado
    - Freguesia (Ilha)
    - Guadalupe
    - Jardim América
    - Praia da Bandeira
    - Riachuelo
    - Rocha Miranda
    - Sampaio
    - Santíssimo
    - Tomás Coelho
    - Vila Militar
  - **São Paulo**
    - Belém
    - Jaçanã
    - Mandaqui
    - Parelheiros
    - Sacomã

- **16.0**
  - **Rio de Janeiro**
    - Benfica
    - Centro
    - Cidade Nova
    - Cidade Universitária
    - Estácio
    - Jacaré
    - Parque Colúmbia
    - São Cristóvão
    - Vargem Grande
    - Vargem Pequena
    - Vasco da Gama
  - **São Paulo**
    - Brás
    - Cambuci
    - Itaim Paulista
    - Jaguaré
    - Jardim São Luís
    - São Rafael
    - Vila Guilherme
    - Vila Leopoldina

- **17.0**
  - **Rio de Janeiro**
    - Caju
    - Cavalcanti
    - Cidade de Deus
    - Coelho Neto
    - Cosmos
    - Higienópolis
    - Penha Circular
    - Piedade
    - Ramos
    - Rocha
    - Vaz Lobo
    - Zumbi
  - **São Paulo**
    - Iguatemi
    - Pirituba

- **19.0**
  - **Rio de Janeiro**
    - Acari
    - Bancários
    - Cacuia
    - Deodoro
    - Honório Gurgel
    - Marechal Hermes
    - Osvaldo Cruz
    - Paciência
    - Tauá
    - Vila Kosmos
  - **São Paulo**
    - Campo Limpo

8.0	
Rio de Janeiro	
Anchieta	
Barra da Tijuca	
Barra de Guaratiba	
Botafogo	
Campo dos Afonsos	
Catete	
Copacabana	
Cosme Velho	
Costa Barros	
Del Castilho	
Engenheiro Leal	
Engenho de Dentro	
Gamboa	
Gardênia Azul	
Gávea	
Glória	
Humaitá	
Ipanema	
Irajá	
Jacarepaguá	
Jardim Botânico	
Joá	
Lapa	
Laranjeiras	
Leblon	
Leme	
Mangueira	
Maré	
Paquetá	
Parada de Lucas	
Recreio dos Bandeirantes	
Ribeira	
Santo Cristo	
São Conrado	
Saúde	
Sepetiba	
Turiaçú	
Urca	
São Paulo	
Alto de Pinheiros	
Artur Alvim	
Barra Funda	
Consolação	
José Bonifácio	
Lapa	
Pari	
Pedreira	
Pinheiros	
República	
Saúde	

18.0	
Rio de Janeiro	
Bangu	
Barros Filho	
Bonsucesso	
Campinho	
Campo Grande	
Cascadura	
Colégio	
Flamengo	
Freguesia (Jacarepaguá)	
Itanhangá	
Jabour	
Jardim Guanabara	
Lagoa	
Madureira	
Pechincha	
Penha	
Portuguesa	
Praça Seca	
Realengo	
Rio Comprido	
Santa Cruz	
São Francisco Xavier	
Taquara	
Tijuca	
Vicente de Carvalho	
Vila da Penha	
Vila Isabel	
Vila Valqueire	
São Paulo	
Água Rasa	
Bela Vista	
Bom Retiro	
Brasilândia	
Butantã	
Campo Belo	
Campo Grande	
Carrão	
Casa Verde	
Cidade Ademar	
Cidade Dutra	
Cidade Líder	
Cidade Tiradentes	
Freguesia do Ó	
Ipiranga	
Itaim Bibi	
Itaquera	
Jaguara	
Jardim Paulista	
Lajeado	
Liberdade	
Moema	
Morumbi	
Penha	
Perus	
Ponte Rasa	
Raposo Tavares	
Rio Pequeno	
Santa Cecília	
Santana	
Santo Amaro	
São Lucas	
São Miguel Paulista	
Sé	
Socorro	
Tatuapé	
Tremembé	
Vila Andrade	
Vila Formosa	
Vila Maria	
Vila Mariana	
Vila Matilde	
Vila Medeiros	
Vila Prudente	
Vila Sônia	

5. DISCUSSION

We can see that half of the clusters are composed of few neighborhoods, while two clusters have many neighborhoods (49 and 79 boroughs). With the exception of these two large clusters, the other groupings, in fact, bring together similar neighborhoods (I can say that since I know both cities well).

In these two larger clusters, there are some neighborhoods that are notoriously very different, such as Irajá (RJ) and Pinheiros (SP), in cluster 8.0, and Madureira (RJ) and Moema (SP), in cluster 18.0. That is, it is not recommended that a person use this information to choose which neighborhood will live in when moving to another city, and it is necessary to carry out a lot of additional research.

As a refinement of this study, a new division of the two cities can be made, grouping small and close neighborhoods. Additionally, you can exclude neighborhoods that have little information on Foursquare (less than 10 venues, for example). Also, if large clusters remain, a new division into clusters can be made considering only these neighborhoods.

6. CONCLUSION

The aim of this study was to group similar neighborhoods in the cities of Rio de Janeiro and São Paulo, using the category of local venues as parameters.

To some extent, the result was satisfactory. However, is necessary to make some refinements to obtain a more reliable result.