

CLUSTERS OF BOROUGHES IN RIO DE JANEIRO AND SÃO PAULO

APPLIED DATA SCIENCE CAPSTONE (COURSERA - IBM COURSE)

Luciana Guerra
March, 2021

WHERE TO LIVE IN RIO DE JANEIRO AND SÃO PAULO

- 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.
- 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.



DATA AND SOURCES

- 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

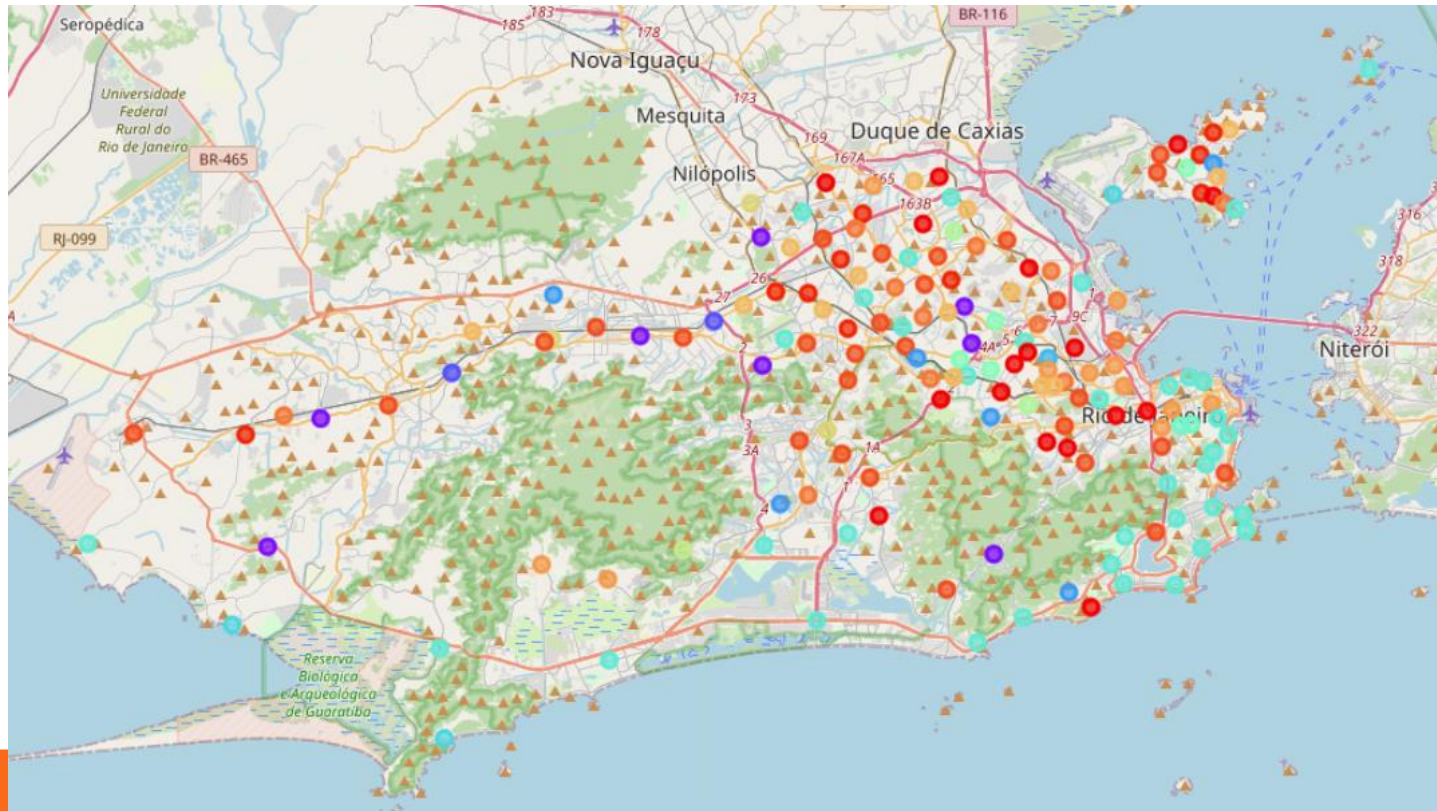


METHODOLOGY

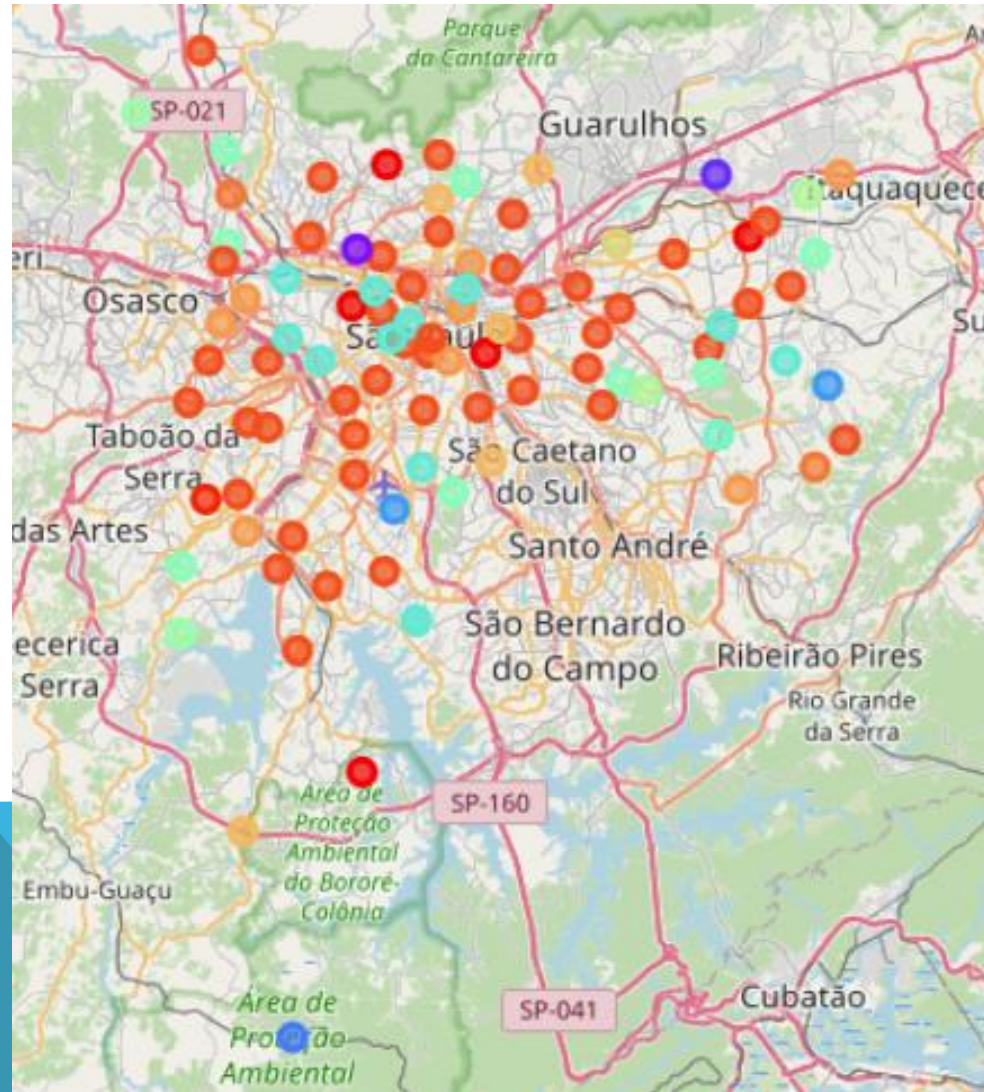
- Get the names of the boroughs
- Get latitude and longitude of the boroughs
- Get the most frequent category venues of each borough
- Create clusters



RESULTS - CLUSTERS - BOROUGHES IN RIO



RESULTS – CLUSTERS - BOROUGHS IN SÃO PAULO



RESULTS – CLUSTERS

Number of Neighborhoods	City		
nº Cluster	Rio de Janeiro	São Paulo	Total
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	162	96	258

RESULTS – CLUSTERS

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

5.0

Rio de Janeiro

Cocotá
Curicica
Lins de Vasconcelos
Quintino Bocaiúva
Rocinha
Vila Kennedy

São Paulo

Guaianases
Jabaquara

2.0

São Paulo

Ermelino Matarazzo

6.0

Rio de Janeiro

Jacarezinho

3.0

Rio de Janeiro

Magalhães Bastos
Senador Vasconcelos

7.0

Rio de Janeiro

Galeão
Grumari
Pedra de Guaratiba

4.0

São Paulo

Marsilac

RESULTS – CLUSTERS

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

São Paulo

Alto de Pinheiros
Artur Alvim
Barra Funda
Consolação
José Bonifácio
Lapa
Pari
Pedreira
Pinheiros
República
Saúde

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
Turiaçu
Urca

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

RESULTS – CLUSTERS

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
Jaguareé
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

RESULTS – CLUSTERS

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

19.0

☐ Rio de Janeiro

Acari
Bancários
Cacuia
Deodoro
Honório Gurgel
Marechal Hermes
Osvaldo Cruz
Paciência
Tauá
Vila Kosmos

☐ São Paulo

Campo Limpo

DISCUSSION AND CONCLUSION

- 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.
- 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.



DISCUSSION AND CONCLUSION

- As a **refinement** of this study, I suggest:
 - A different division of the two cities, grouping small and close neighborhoods.
 - Exclude neighborhoods that have little information on Foursquare (less than 10 venues, for example).
 - If large clusters remain, make a new division into clusters considering only these neighborhoods.
- To some extent, the result was satisfactory. However, is necessary to make some refinements to obtain a more reliable result.

