Rio de Janeiro's Neighborhoods Clustering and HDI Analysis

Enzo Casemiro Zuccaro

February 8, 2020

IBM Data Science Professional Certificate

Introduction

- Rio de Janeiro is one of the biggest and probably the most famous brazilian city

- Rio's social inequality is unknown for the majority of visitors

- It can be exposed by making a Human Development Index (HDI) analysis of each neighborhood

- The neighborhoods' HDI ranges from 0.970 to 0.700

Objectives

- Verify the relationship between the HDI of Rio de Janeiro's neighborhoods and the number of venues of each one

- Make a cluster analysis using the k-means algorithm to observe similarities between some neighborhoods based on the kind of the venues present in each one

Data Acquisition

- All data related to Rio's neighborhoods was acquired in the following Wikipedia article:
  - https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_por_IDH

- The venues information were acquired using the Foursquare API

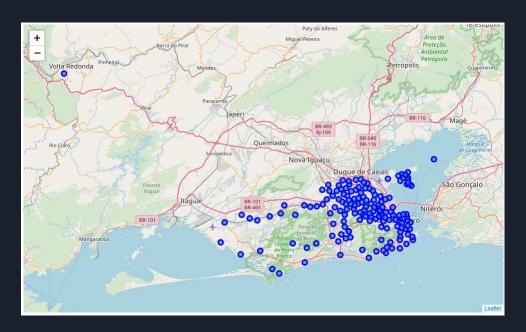- The coordinates of each neighborhoods were acquired using GeoPy library

Using GeoPy

	Neighborhood	HDI
0	Gávea	0.970
1	Leblon	0.967
2	Jardim Guanabara	0.963
3	Ipanema	0.962
4	Humaitá	0.959

	Neighborhood	HDI	Latitude	Longitude
0	Gávea	0.970	-22.981424	-43.238324
1	Leblon	0.967	-22.983556	-43.224938
2	Jardim Guanabara	0.963	-22.812836	-43.200779
3	Ipanema	0.962	-22.983956	-43.202216
4	Humaitá	0.959	-22.954641	-43.200480

Neighborhoods Visualization with Folium

- Mangueira was misplaced by GeoPy, so it was removed

Using the Foursquare API

- Radius of 1 km
- A limit of 100 venues per call

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Gávea	-22.981424	-43.238324	Instituto Moreira Salles (IMS)	-22.981627	-43.239637	Cultural Center
1	Gávea	-22.981424	-43.238324	Empório Jardim	-22.981632	-43.239682	Breakfast Spot
2	Gávea	-22.981424	-43.238324	Sociedade Germania	-22.980419	-43.239628	Athletics & Sports
3	Gávea	-22.981424	-43.238324	Bosque da PUC	-22.980186	-43.234375	College Quad
4	Gávea	-22.981424	-43.238324	Parque da Cidade	-22.981210	-43.242434	Park

- A total of 4670 venues were found for 159 neighborhoods
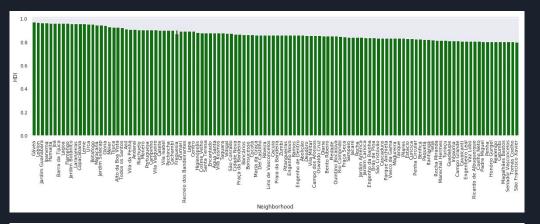
One Hot Encoding

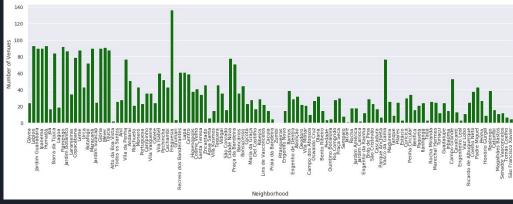- It was necessary to apply a one hot encoding algorithm to apply the clustering method

	Neighborhood	ATM	Acai House	Accessories Store	Adult Boutique	Airport Lounge	American Restaurant	Amphitheater	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Enter
0	Abolição	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	Acari	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	Alto da Boa Vista	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	Anchieta	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	Andaraí	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Ten Most Common Venues

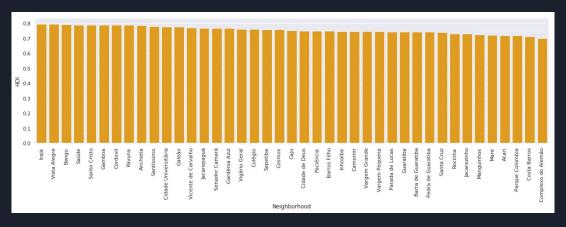|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abolição	BBQ Joint	Food Truck	Soccer Field	Bar	Deli / Bodega	Gym / Fitness Center	Gym	Music Venue	General Entertainment	Burger Joint
1	Acari	Market	Plaza	Burger Joint	Bus Station	Buffet	Supermarket	Brazilian Restaurant	Gym / Fitness Center	Soccer Field	Snack Place
2	Alto da Boa Vista	Mountain	Scenic Lookout	Zoo	French Restaurant	Food & Drink Shop	Food Court	Food Service	Food Stand	Food Truck	Football Stadium
3	Anchieta	Plaza	Gym / Fitness Center	Pizza Place	Fast Food Restaurant	Nightclub	Fruit & Vegetable Store	Gym	Pet Store	Church	General Entertainment
4	Andaraí	Bar	Bakery	Gym / Fitness Center	Food Truck	Japanese Restaurant	Music Venue	Gym	Pizza Place	Farmers Market	Plaza

HDI Analysis - Neighborhoods with a HDI equal or higher than 0.8
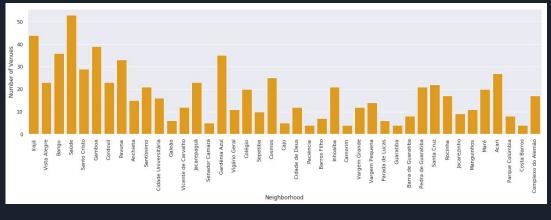

- For these neighborhoods the average number of venues is 35.40

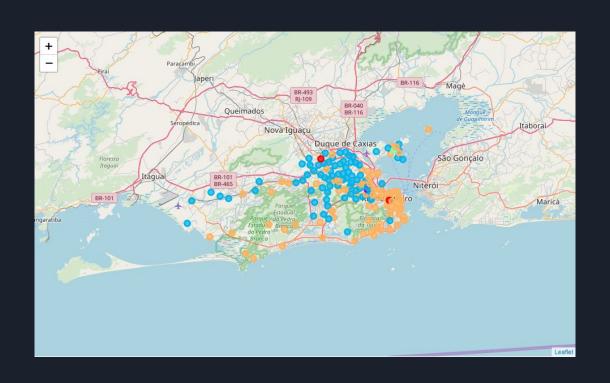# HDI Analysis - Neighborhoods with a HDI lower than 0.8

- For these neighborhoods the average number of venues is 17.85

Clustering

- Five clusters were used to divide the neighborhoods using the k-means algorithm

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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 Venu
0	Gávea	0.970	-22.981424	-43.238324	24.0	4.0	Sushi Restaurant	Park	Cultural Center	Trail	Theater	Gym	Brazilian Restaurant	Fruit Vegetab Stor
1	Leblon	0.967	-22.983556	-43.224938	93.0	4.0	Brazilian Restaurant	Steakhouse	Bar	Restaurant	Pizza Place	Italian Restaurant	Juice Bar	Frenc Restaura
2	Jardim Guanabara	0.963	-22.812836	-43.200779	90.0	4.0	Brazilian Restaurant	Gym	Pizza Place	Burger Joint	Beach Bar	Japanese Restaurant	Bakery	Ba
3	Ipanema	0.962	-22.983956	-43.202216	90.0	4.0	Food Stand	Brazilian Restaurant	Pizza Place	Bar	Italian Restaurant	Japanese Restaurant	Hostel	Hot
4	Humaitá	0.959	-22.954641	-43.200480	93.0	4.0	Bar	Brazilian Restaurant	Café	Vegetarian / Vegan Restaurant	Nightclub	Plaza	Pie Shop	Restaura

Cluster Visualization with Folium

Cluster 0

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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	9t Co
90	Estácio	0.829	-22.916626	-43.203630	3.0	0	Plaza	Tunnel	Zoo	Frame Store	Food & Drink Shop	Food Court	Food Service	Food Stand	
157	Costa Barros	0.713	-22.824700	-43.369839	4.0	0	Plaza	Border Crossing	Mountain	Liquor Store	Zoo	French Restaurant	Food Court	Food Service	

- The average HDI for this cluster is 0.771

Cluster 1

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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 M Comm Ve
76	Jacaré	0.839	-22.893117	-43.2572	1.0	1	Market	Zoo	Flower Shop	Food & Drink Shop	Food Court	Food Service	Food Stand	Food Truck	Foo Stac

- The average HDI for this cluster is 0.839

Cluster 2

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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
7	Lagoa	0.959	-22.962466	-43.202488	19.0	2	Scenic Lookout	Gym / Fitness Center	Japanese Restaurant	Trail	Dive Bar	Pool	Plaza	Playgroun
16	Jardim Sulacap	0.944	-22.895680	-43.392997	25.0	2	Gym	Plaza	Restaurant	Food Truck	Toll Booth	Tunnel	Snack Place	B:
21	Todos os Santos	0.922	-22.894563	-43.285060	26.0	2	Soccer Stadium	Bar	Furniture / Home Store	Pizza Place	Clothing Store	Tattoo Parlor	History Museum	Shoppir Ma
23	Vila da Penha	0.909	-22.843507	-43.310058	77.0	2	Bakery	Pizza Place	Brazilian Restaurant	Bar	Plaza	Fast Food Restaurant	Japanese Restaurant	Gym Fitnes Cent
25	Riachuelo	0.905	-22.902695	-43.255175	21.0	2	Gym	Restaurant	Brazilian Restaurant	Train Station	Farmers Market	Organic Grocery	Supermarket	Gym Fitnes Cent

- The average HDI for this cluster is 0.824

Cluster 3

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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 Co...
80	Jardim Carioca	0.836	-22.805249	-43.193195	2.0	3	Samba School	Asian Restaurant	Zoo	French Restaurant	Food & Drink Shop	Food Court	Food Service	Food Stand	

- The average HDI for this cluster is 0.836

Cluster 4

	Neighborhood	HDI	Latitude	Longitude	Number of Venues	Cluster Label	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
0	Gávea	0.970	-22.981424	-43.238324	24.0	4	Sushi Restaurant	Park	Cultural Center	Trail	Theater	Gym	Brazilian Restaurant	Fruit Vegetab Stor
1	Leblon	0.967	-22.983556	-43.224938	93.0	4	Brazilian Restaurant	Steakhouse	Bar	Restaurant	Pizza Place	Italian Restaurant	Juice Bar	Frenc Restaura
2	Jardim Guanabara	0.963	-22.812836	-43.200779	90.0	4	Brazilian Restaurant	Gym	Pizza Place	Burger Joint	Beach Bar	Japanese Restaurant	Bakery	Ba
3	Ipanema	0.962	-22.983956	-43.202216	90.0	4	Food Stand	Brazilian Restaurant	Pizza Place	Bar	Italian Restaurant	Japanese Restaurant	Hostel	Hot
4	Humaitá	0.959	-22.954641	-43.200480	93.0	4	Bar	Brazilian Restaurant	Café	Vegetarian / Vegan Restaurant	Nightclub	Plaza	Pie Shop	Restaura

- The average HDI for this cluster is 0.858

Conclusion

- The results shows that there is some relationship between the HDI and the number of venues present in a neighborhood in Rio de Janeiro

- For neighborhoods with a HDI equal or higher than 0.8, which is considered a high HDI, the average number of venues is 35, twice the average number of venues with a HDI lower than 0.8 and higher than 0.5

- The k-means method divided the neighborhood into 2 large clusters (2 and 4), and 3 smaller ones (0, 1 and 3). The 2 largest clusters seems to be divided into groups of higher HDI neighborhoods (cluster 4) and lower HDI neighborhoods (cluster 2)