Where to Install a New Hostel In Rio de Janeiro?

A Data Analysis Approach

Capstone Project
IBM Data Scientist Professional Certificate

Felipe N. Ruperti

July 2019

Introduction

Some facts about Rio de Janeiro:

- The city had 18,204,831 arrivals in the year of 2014, according to the Secretary of Tourism of the State of Rio de Janeiro.
- Rio only had 75 hostels and, when considering hotel rooms, Rio was placed on the 77th position, according to the Price of Travel website.

Conclusion: there are unexplored opportunities in the low-budget tourism sector.

Objective: this study provides a data analysis approach in order to spot the best neighborhoods of where to install a new hostel in the city of Rio de Janeiro.

Methodology - Data Preparation

The dataset was created following two steps:

1. The initial data consists of 139 geospatial points/spots, which had their latitude and longitude gathered manually from the GPS website.

2. The initial data latitude/longitude of each point was used in order to obtain two variables: main/primary and specific/secondary venue types from the Foursquare developer api. The radius from each point was set in 350 meters and the limit of venues retrived per neighborhood was set in 500[1]. As there are many venues per neighborhood, the dataset has a total of 4,737 observations and 8 columns/variables. Table 1 shows the first 3 observations of the dataset.

Methodology - Data Preparation

	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Main Venue Category	Venue Category
1	Leme 1	-22.962373	-43.168171	Bar do David	-22.960691	-43.168878	nightlife	Bar
2	Leme 1	-22.962373	-43.168171	Ponta do Leme	-22.962557	-43.166152	parks_outdoors	Beach
3	Leme 1	-22.962373	-43.168171	Calçadão do Leme	-22.962679	-43.167250	parks_outdoors	Pedestrian Plaza

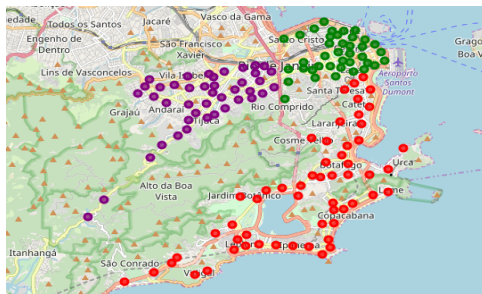Table: First 3 Observations of the Dataset

Figure: Rio's Districts - Neighborhoods

Methodology - Descriptive Analysis and K-Means

- **Descripive Analysis**: was used in order to have a better understanding of the venues and its spatial distribution across neighborhoods. Statistics as mean, median, mode and graphs as heatmaps, bar charts and histograms were employed to enhance insight.

- **K-Means**: was the algorithm herein applied for data clustering, because it is simple to understand and deploy and provides further insight into the data, by grouping the neighborhoods according to their secondary venue types.
  - Metric: euclidian distance.
  - Initialization: k-means++
  - Optimal k: iteration from $k = 3$ to 20, using the total within sum of squares in the silhouette/elbow method.

Methodology - Ranking
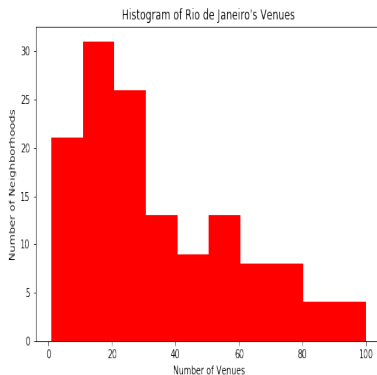
- **Variables**:

    - Number of Venues for Tourism
    - Number of Venues for Tourism/Per Hostel + Newcomer
    - Counts of Scenic Lookouts and Beaches

- **Data Preparation**: min-max normalization of the variables.

- **Ranking Composition**:

    60%N of Venues for Tourism +
    30%N of Venues for Tourism/Per Hostel & Newcomer +
    10%N of Scenic Lookouts & Beaches

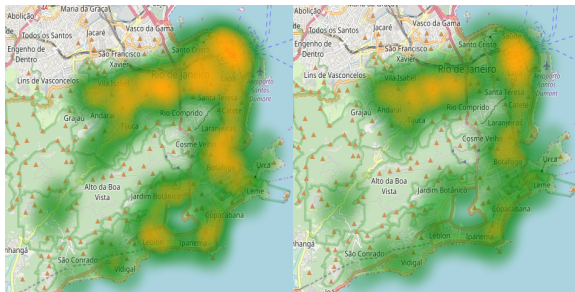# Results - Descriptive Analysis

Figure

Mean = 35 & Median = 27

Results - Descriptive Analysis - Heatmap 1

The number of distinct specific/secondary venues for Rio totals 296. On the other hand, there are 8 main/primary venue categories found on Foursquare: arts and entertainment, building, education, events, food, nightlife, parks & outdoors and shops.
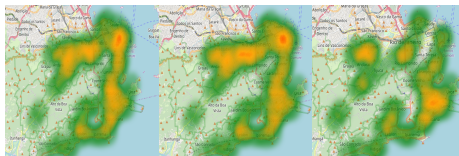

(a) Parks and Outdoors   (b) Shops
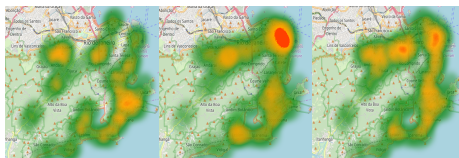
Figure: Primary Venue Heatmaps - Part I

Results - Descriptive Analysis - Heatmap 2

(a) Arts & Entertainment   (b) Building   (c) Education
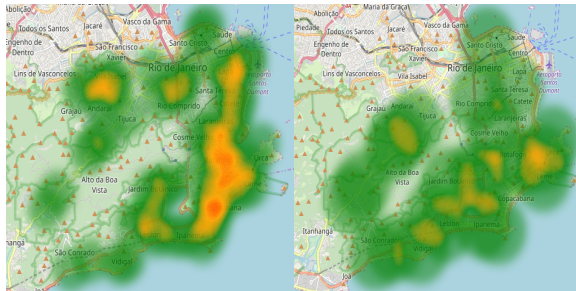

(d) Events   (e) Food   (f) Nighlife

Figure: Primary Venue Heatmaps - Part II

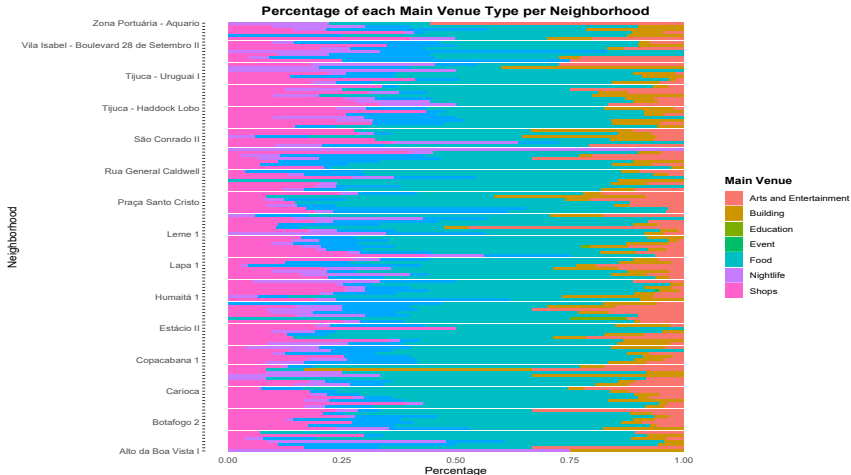# Results - Descriptive Analysis - Heatmap 3

(a) Hostels

(b) Scenic Lookouts and Beaches

Figure: Hostels, Scenic Lookouts and Beaches Heatmap

Results - Descriptive Analysis - Number of Hostels

	Neighborhood	N° Hostels
1	Copacabana 8	7
2	Lapa 1	6
3	Glória - Santa Teresa	6
4	Copacabana - Bairro Peixoto	4
5	Botafogo 3	4
6	Botafogo 4	3
7	Santa Teresa - Largo dos Guimarães	3
8	Leblon 3	3
9	Botafogo 5	3
10	Copacabana 2	2

Results - Descriptive Analysis - Bar Chart 1

Results - Descriptive Analysis - Bar Chart 2

Counts of Main Venue Type per Neighborhood

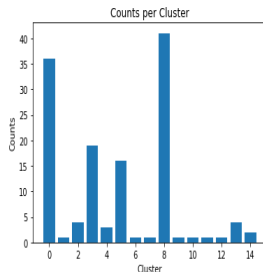**Main Venue**

- Arts and Entertainment
- Building
- Education
- Event
- Food
- Nightlife
- Shops
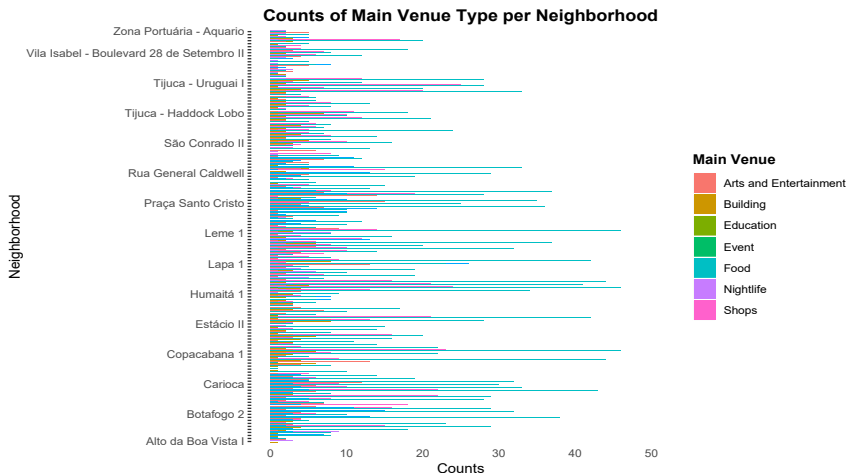
Results - K-Means
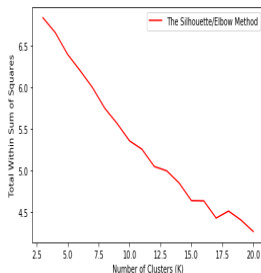
(a) Silhouette Graph   (b) Counts per Cluster

Figure: Silhouette Graph and Counts per Cluster Label Bar Chart

K = 15 is chosen.

Results - Ranking

	index	Cluster Label	Cluster Description	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Ranking
0	1	8	Bars, Restaurants, Pizza and Coffe	Leblon 4	Zona Sul	-22.985782	-43.227286	92.50
1	2	0	Restaurants, Bars and Culture	Praça Passeio Público	Centro	-22.912930	-43.177173	79.48
2	3	0	Restaurants, Bars and Culture	Cinelândia	Centro	-22.910366	-43.175907	78.17
3	4	8	Bars, Restaurants, Pizza and Coffe	Leme 2	Zona Sul	-22.963840	-43.173703	78.17
4	5	8	Bars, Restaurants, Pizza and Coffe	Gávea 1	Zona Sul	-22.975750	-43.228035	76.86
5	6	8	Bars, Restaurants, Pizza and Coffe	Ipanema 1	Zona Sul	-22.984576	-43.198638	70.57
6	7	0	Restaurants, Bars and Culture	Camelôdromo	Centro	-22.903379	-43.182878	70.28
7	8	0	Restaurants, Bars and Culture	Botafogo 1	Zona Sul	-22.956530	-43.193978	68.97
8	9	0	Restaurants, Bars and Culture	Praça Tiradentes	Centro	-22.906478	-43.182956	68.97
9	10	8	Bars, Restaurants, Pizza and Coffe	Lapa 2	Centro	-22.912367	-43.186272	67.65
10	11	0	Restaurants, Bars and Culture	Rua do Ouvidor	Centro	-22.903613	-43.178112	67.65
11	12	0	Restaurants, Bars and Culture	Praça XV	Centro	-22.902859	-43.173220	66.34
12	13	0	Restaurants, Bars and Culture	Praça Mauá	Centro	-22.897001	-43.180842	65.03
13	14	0	Restaurants, Bars and Culture	Castelo	Centro	-22.911183	-43.173301	61.08
14	15	8	Bars, Restaurants, Pizza and Coffe	Humaitá 2	Zona Sul	-22.955930	-43.197163	61.08
15	16	8	Bars, Restaurants, Pizza and Coffe	Botafogo 5	Zona Sul	-22.947694	-43.185878	60.06
16	17	0	Restaurants, Bars and Culture	Rua do Senado	Centro	-22.909510	-43.185593	59.77
17	18	8	Bars, Restaurants, Pizza and Coffe	Copacabana 2	Zona Sul	-22.969604	-43.185210	59.28
18	19	8	Bars, Restaurants, Pizza and Coffe	Ipanema 2	Zona Sul	-22.984335	-43.204910	58.49
19	20	3	Bars, Bakeries and Restaurants	Lapa 1	Centro	-22.913835	-43.181443	56.15

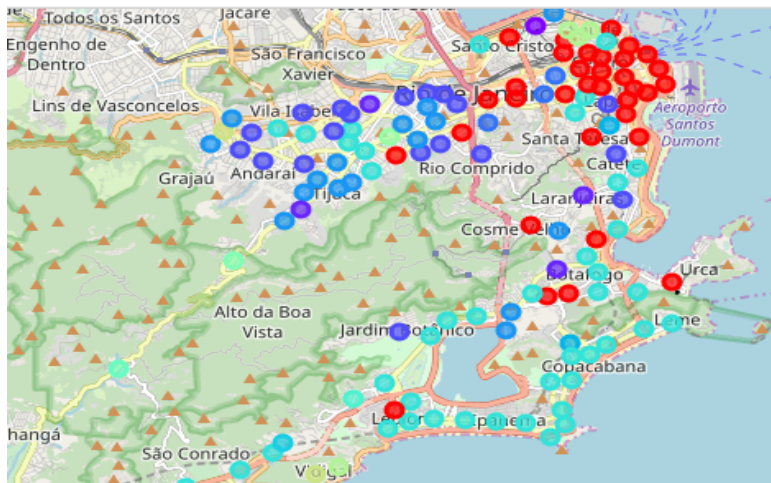Table: Cluster Description

Results - K-Means - Clusters

Figure: Spatial Distribution of Clusters

Results - K-Means - Cluster Description

	index	Cluster Label	Cluster Description	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Ranking
0	1	8	Bars, Restaurants, Pizza and Coffe	Leblon 4	Zona Sul	-22.985782	-43.227286	92.50
1	2	0	Restaurants, Bars and Culture	Praça Passeio Público	Centro	-22.912930	-43.177173	79.48
2	3	0	Restaurants, Bars and Culture	Cinelândia	Centro	-22.910366	-43.175907	78.17
3	4	8	Bars, Restaurants, Pizza and Coffe	Leme 2	Zona Sul	-22.963840	-43.173703	78.17
4	5	8	Bars, Restaurants, Pizza and Coffe	Gávea 1	Zona Sul	-22.975750	-43.228035	76.86
5	6	8	Bars, Restaurants, Pizza and Coffe	Ipanema 1	Zona Sul	-22.984576	-43.198638	70.57
6	7	0	Restaurants, Bars and Culture	Camelôdromo	Centro	-22.903336	-43.182878	70.28
7	8	0	Restaurants, Bars and Culture	Botafogo 1	Zona Sul	-22.956530	-43.193978	68.97
8	9	0	Restaurants, Bars and Culture	Praça Tiradentes	Centro	-22.906478	-43.182956	68.97
9	10	8	Bars, Restaurants, Pizza and Coffe	Lapa 2	Centro	-22.912367	-43.186272	67.65
10	11	0	Restaurants, Bars and Culture	Rua do Ouvidor	Centro	-22.903613	-43.178112	67.65
11	12	0	Restaurants, Bars and Culture	Praça XV	Centro	-22.902859	-43.173220	66.34
12	13	0	Restaurants, Bars and Culture	Praça Mauá	Centro	-22.897001	-43.180842	65.03
13	14	0	Restaurants, Bars and Culture	Castelo	Centro	-22.911183	-43.173301	61.08
14	15	8	Bars, Restaurants, Pizza and Coffe	Humaitá 2	Zona Sul	-22.955930	-43.197163	61.08
15	16	8	Bars, Restaurants, Pizza and Coffe	Botafogo 5	Zona Sul	-22.947694	-43.185878	60.06
16	17	0	Restaurants, Bars and Culture	Rua do Senado	Centro	-22.909510	-43.185593	59.77
17	18	8	Bars, Restaurants, Pizza and Coffe	Copacabana 2	Zona Sul	-22.969604	-43.185210	59.28
18	19	8	Bars, Restaurants, Pizza and Coffe	Ipanema 2	Zona Sul	-22.984335	-43.204910	58.49
19	20	3	Bars, Bakeries and Restaurants	Lapa 1	Centro	-22.913835	-43.181443	56.15

Table: Ranking - Top 20 Neighborhoods for Hostels

Conclusion

The results of the top 20 best neighborhoods for the installment of a new hostel based on the number of venues, the competition from existing hostels and nearby scenic lookouts and beaches is consistent to what would be expected when considering the characteristics of the neighborhoods and touristic opportunities.

There will be a trade-off regarding the neighborhoods in the city Rio de Janeiro between rent price vs safety at night and between beaches vs number of cultural/historic venues, that is reflected in the differences between the South Zone and Downtown district.

This creates specific strategies, notably for marketing, that the entrepreneur should take into account, because the neighborhood/district/location will determine not only the price per stay to be charged, but, specially, the hostel's touristic profile.

References

📄 Foursquare.
https://developer.foursquare.com/.
Accessed on: 2019-07-04.

📄 GPS Coordinates.
https://www.gps-coordinates.net/.
Accessed on: 2019-07-04.

📄 Price of Travel.
Cities with the Most Hotels (including Hostels).
https://www.priceoftravel.com/1588/
cities-with-the-most-hotels-around-the-world/.
Accessed on: 2019-07-04.