

# **Coursera Capstone IBM Applied Data Science Capstone**

## **The better sector to invest in Brasilia, Brazil.**

By: Simone Pereira

2021



## 1. Introduction

For many investors who wish to open a new business, many doubts arise in what to invest their capital and in which place the central location and consumers of the place offers a great starting point.

As a result, we will check within the city of Brasilia the best place to open a new business.

Of course, as with any business decision, opening a new location requires serious consideration and it is much more complicated than it looks.

Particularly, the location is one of the the most important decisions that will determine whether the site will be a success or a failure.

## 2. Business Problem

The objective of this project is to analyze and select the best locations in the city of Brasilia, Brazil to open a new business. Using data science methodology and machine learning techniques such as clustering, this project aims to provide solutions to meet the business.

Question: What is the best business in Brasília based on consumer tastes as measured by the booking of places that are most frequented by Forsquare and by income?

### Data:

The dataset used to solve the problem:

- Data from all administrative regions of Brasilia and the respective per capita income
- FourSquare to find the venus for each city and
- The coordinates of each city and each venues for analysis.

### 2. Materials and Methods

The data for the administrative regions of Brasília with the respective per capita incomes were extracted from Wikipedia, address: ([https://pt.wikipedia.org/wiki/Lista\\_de\\_regi%C3%B5es\\_administrativas\\_do\\_Distrito\\_Federal\\_p\\_or\\_renda\\_per\\_capita](https://pt.wikipedia.org/wiki/Lista_de_regi%C3%B5es_administrativas_do_Distrito_Federal_p_or_renda_per_capita)). The BeautifulSoup library was used for the extraction.

### Analysis methodolog

Initially, all 31 administrative regions (ARs) in Brasília were used for the analysis. There is no need to exclude any for the first analyzes. However, when extracting the most visited sites, the

RA Jardim Botânico returned and did not use it when grouping the clusters with the most visited sites. The data related to per capita income were transformed into a float to serve as an X variable. Clustering method: k-means.

### 3. Data Analysis - Development

The parameter for Kmeans was the dataframe composed of the following variables:

- Income,
- Latitude and longitude.

The number of k was determined using the silhouette coefficient closest to 1.

### 4. Discussion of Results and Conclusion

The results of k-means clustering showed that we can group customer profiles based on their income and their geographic location. Three clusters were found, which were later grouped to the most visited locations in each cluster group, thus showing the most visited locations by each group.

- Cluster 0: The most visited places by class D:

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Bakery)	9.0	7.0	14.0	7.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Longitude, Bakery)	9.0	7.0	14.0	7.0
(Venue Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0

- Cluster 1: The most visited places by class A:

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Longitude, Shopping Plaza)	1.0	8.0	0.0	0.0
(Venue Latitude, Shopping Plaza)	1.0	8.0	0.0	0.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Shopping Plaza)	1.0	8.0	0.0	0.0

- Cluster 2: The most visited places by class B.

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0

Cluster 3: The most visited places by class C.

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(City Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0

The analysis shows that if the investor who wants to invest in a city in Brasilia should invest in Restaurants, which is the most marked place on FourSquare by Brasilienses. And if you want high-income consumers, you can choose a Shopping Center as the location that will obtain a higher number of high-income customers. Given that the frequency of shopping centers is equal to the frequency of appointments in restaurants according to FourSquare.

## **5. References**

IBM Data Science Professional Certificate - <https://www.coursera.org/IBM Data Science Professional Certificate>.