# Coursera Capstone

# IBM Applied Data Science Capstone

# Opening a Brazilian Restaurant in Rio de Janeiro, Brazil

## 1. Introduction

Final Part of final IBM Capstone Project, this is my opportunity to work on real life data and understand its properties to make a business decision. The learning outcome was to define a business problem and to use data science methodologies to understand the data and eventually make a business conclusion.I´m also had to scrape data from the web and use Foresquare API to extraction location details.

In this project, we start with defining a problem, identifying the needed data to answer the problem, analyze data to study its properties and give a conclusion to answer the business question.

## 2. Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Rio de Janeiro, Brazil to open a new Brazilian Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: where would you recommend that they open it?

## 3. Target Audience

- Travellers and residents who want to have an experience to try Brazilian Food.

- Business personnel who are planning to invest in a restaurant

The analysis is going to give a view of demographics before they decide to invest

## 4. Data

- Name of the 167 Neighboors, from web scrapping

https://pt.wikipedia.org/wiki/Lista_de_bairros_da_cidade_do_Rio_de_Janeiro

- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data

- Use Foresquare Data to obtain info about restaurants, We will use this data to perform clustering on the neighbourhoods.

## 5. Methodology

I´m going to use the basic methodology as taught in Week 3 lab, I add the calculus to select optimal number of clusters in KMeans Algorithm (Silhouette Score)

### 5.1 Data Scraping

I will do web scraping using Python pandas to extract the list of neighbourhoods data. I organized the CSV file in Excel and upload the file to my notebook.this is just a list of names.

## 5.2 Get the geographical coordinates

I need to get the geographical coordinates in the form of latitude and longitude, using the package Geocode, in in order to be able to use Foursquare API. To do so,

I converted the neighbourhoods into their equivalent latitude and longitude values.

Put into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package.

## 5.3 Use the Foursquare API to explore the neighborhoods

I use the Foursquare API to explore neighborhoods, I need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical. Next, we will use Foursquare API to get the top 300. After that, explore function to get the Brazilian Restaurant in each neighborhood, the foursquare brings just 2 Brazilians Restaurants in Copacabana per example. We are going to use in out analysis, all the restaurants.

## 5.4 Analyze the restaurants fo Each Neighborhood

I group rows by neighborhood and by taking the mean of the frequency of occurrence of each type of restaurant then we discovered each neighborhood along with the top 5 most common venues

## 5.5 Predictive Modeling

I have to define ideal Number of clusters, and the doing the Kmeans Clustering

## 5.6 Clustering

## 5.6.1 Silhouette Score

This is a better measure to decide the number of clusters to be formulated from the data. It is calculated for each instance and the formula goes like this:

Silhouette Coefficient = (x-y)/ max(x,y)

```
K-Means Clustering - Confusion Matrix

 Predicted      0  1  2  3  4
Actuall
Abolição       1  0  0  0  0
Acari          0  0  1  0  0
Anchieta       1  0  0  0  0
Andaraí        0  0  1  0  0
Anil           0  0  1  0  0
...           .. .. .. .. ..
Vila Militar   0  0  0  0  1
Vila Valqueire 0  0  1  0  0
Vila da Penha  0  0  1  0  0
Vista Alegre   0  0  1  0  0
Zumbi          0  0  1  0  0

[123 rows x 5 columns]

Silhouette-score: 0.303
```

```
[41]: for k in range(5,15):
          k_means_fitk = KMeans(n_clusters=k,max_iter=500)
          k_means_fitk.fit(x)
          print ("For K value",k,",Silhouette-score: %0.3f" % silhouette_score(x, k_means_fitk.labels_, metric='euclidean'))

      For K value 5 ,Silhouette-score: 0.342
      For K value 6 ,Silhouette-score: 0.331
      For K value 7 ,Silhouette-score: 0.331
      For K value 8 ,Silhouette-score: 0.347
      For K value 9 ,Silhouette-score: 0.360
      For K value 10 ,Silhouette-score: 0.359
      For K value 11 ,Silhouette-score: 0.397
      For K value 12 ,Silhouette-score: 0.387
      For K value 13 ,Silhouette-score: 0.391
      For K value 14 ,Silhouette-score: 0.411
```

The silhouette coefficient values in the preceding results shows that K value 14 has better scores than all the other values. As a thumb rule, we need to take the next K value of the highest silhouette coefficient.

```
[44]: # elbow curve - Avg. within-cluster sum of squares
      fig = plt.figure()
      ax = fig.add_subplot(111)
      ax.plot(K, avgWithinSS, 'b*-')
      plt.grid(True)
      plt.xlabel('Number of clusters')
      plt.ylabel('Average within-cluster sum of squares')
```

```
[44]: Text(0, 0.5, 'Average within-cluster sum of squares')
```
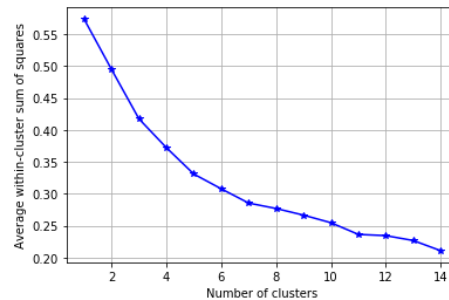


```
[45]: # elbow curve - percentage of variance explained
      fig = plt.figure()
      ax = fig.add_subplot(111)
      ax.plot(K, bss/tss*100, 'b*-')
      plt.grid(True)
      plt.xlabel('Number of clusters')
      plt.ylabel('Percentage of variance explained')
      plt.show()
```
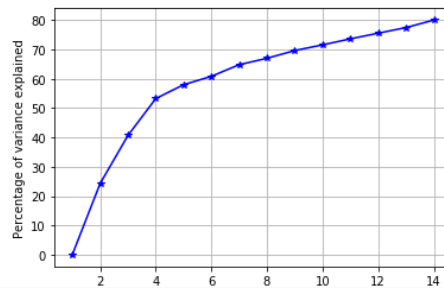


The total percentage of variance explained value should be greater than 80 percent to decide the optimal number of clusters. Even here, a k-value of fourteen seems to give a decent value of total variance explained. Hence, we can conclude from all the preceding metrics (silhouette, average within cluster variance, and total variance explained), that fourteen clusters are ideal.

We use the distortion score and elbow method for K means clustering to identify the correct value of K. From the graph it can be seen that the best value is K=14

### 5.6.2 K-Mean

K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem.

Using K-mean to clustering data neighborhoods with most common restaurant in each Neighborhood

The results will allow us to identify which neighbourhoods have higher concentration of each type of restaurant .Based on the occurrence of restaurant in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open what kind of restaurant.

6. **Results section**.



**Cluster 0**

| [78]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 0 | Saúde | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 12 | Paquetá | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 18 | Cosme Velho | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 21 | Urca | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 30 | São Conrado | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 34 | Maracanã | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 47 | Vigário Geral | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 53 | Inhaúma | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 57 | Rocha | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 87 | Oswaldo Cruz | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 100 | Tauá | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 102 | Portuguesa | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 105 | Guadalupe | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 134 | Vila Militar | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 137 | Magalhães Bastos | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 152 | Pedra de Guaratiba | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 154 | Jacarezinho | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |

There are 17 neighborhoods in Cluster 0. We can easily notice that most neighborhoods in Cluster 0 have Brazilian Restaurants, Vegan Restaurants in the most Commom Restaurants.

## Cluster 1

| [79]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 2 | Santo Cristo | Restaurant | Brazilian Restaurant | Italian Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| | 37 | Grajaú | Restaurant | Japanese Restaurant | Sushi Restaurant | Brazilian Restaurant | Empanada Restaurant |
| | 39 | Bonsucesso | Restaurant | Latin American Restaurant | Brazilian Restaurant | Japanese Restaurant | Jewish Restaurant |
| | 68 | Piedade | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 113 | Pavuna | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 114 | Jacarepaguá | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 118 | Curicica | Restaurant | Brazilian Restaurant | Italian Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| | 124 | Vila Valqueire | Restaurant | Japanese Restaurant | Sushi Restaurant | Brazilian Restaurant | Fast Food Restaurant |
| | 138 | Realengo | Seafood Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Empanada Restaurant | Italian Restaurant |
| | 139 | Padre Miguel | Restaurant | Japanese Restaurant | Fast Food Restaurant | Italian Restaurant | Vegetarian / Vegan Restaurant |
| | 158 | Parque Colúmbia | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |

There are 11 neighborhoods in Cluster 1. We can easily notice that most neighborhoods in Cluster 1 have "Restaurants and Brazilian Restaurants.

## Cluster 2

| [80]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 20 | Humaitá | Italian Restaurant | Vegetarian / Vegan Restaurant | Brazilian Restaurant | Restaurant | Argentinian Restaurant |
| | 24 | Ipanema | Brazilian Restaurant | Japanese Restaurant | Italian Restaurant | Mexican Restaurant | Portuguese Restaurant |
| | 27 | Jardim Botánico | Brazilian Restaurant | Comfort Food Restaurant | Italian Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant |
| | 32 | Tijuca | Italian Restaurant | Brazilian Restaurant | Middle Eastern Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| | 36 | Andaraí | Japanese Restaurant | Brazilian Restaurant | Italian Restaurant | Seafood Restaurant | Restaurant |
| | 63 | Todos os Santos | Italian Restaurant | Fast Food Restaurant | Southeastern Brazilian Restaurant | Seafood Restaurant | Brazilian Restaurant |
| | 64 | Cachambi | Japanese Restaurant | Fast Food Restaurant | Italian Restaurant | Brazilian Restaurant | Seafood Restaurant |
| | 82 | Madureira | Brazilian Restaurant | Fast Food Restaurant | Restaurant | Japanese Restaurant | Italian Restaurant |
| | 85 | Rocha Miranda | Italian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 90 | Ribeira | Japanese Restaurant | Seafood Restaurant | German Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| | 91 | Zumbi | Japanese Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 101 | Moneró | Brazilian Restaurant | Fast Food Restaurant | Italian Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant |
| | 115 | Anil | Japanese Restaurant | Italian Restaurant | Brazilian Restaurant | Seafood Restaurant | Fast Food Restaurant |
| | 121 | Taquara | Japanese Restaurant | Restaurant | Fast Food Restaurant | Italian Restaurant | Sushi Restaurant |
| | 127 | Barra da Tijuca | Seafood Restaurant | Vegetarian / Vegan Restaurant | French Restaurant | Italian Restaurant | Sushi Restaurant |
| | 133 | Deodoro | Comfort Food Restaurant | Italian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| | 143 | Campo Grande | Japanese Restaurant | Sushi Restaurant | Brazilian Restaurant | Restaurant | Fast Food Restaurant |
| | 155 | Complexo do Alemão | Chinese Restaurant | Fast Food Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant |

There are 18 neighborhoods in Cluster 2. We can easily notice that most neighborhoods in Cluster 2 have Japanese, Italian and Fast Food in their top 5 restaurants.

## Cluster 3

| [81]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 42 | Penha | Southeastern Brazilian Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 54 | Engenho da Rainha | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 84 | Turiaçu | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 94 | Praia da Bandeira | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 106 | Anchieta | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 116 | Gardénia Azul | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 131 | Recreio dos Bandeirantes | Fast Food Restaurant | Mexican Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 161 | Vila Kennedy | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |

There are 8 neighborhoods in Cluster 3. We can easily notice that most neighborhoods in Cluster 3 have Fast Food and Vegetarian Restaurants in their top 5 restaurants.

## Cluster 4

[82]:

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 6 | Rio Comprido | Brazilian Restaurant | Fast Food Restaurant | Italian Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant |
| 7 | Cidade Nova | Brazilian Restaurant | Italian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Middle Eastern Restaurant |
| 8 | Estácio | Brazilian Restaurant | Restaurant | Italian Restaurant | Middle Eastern Restaurant | Vegetarian / Vegan Restaurant |
| 14 | Flamengo | Brazilian Restaurant | Japanese Restaurant | Middle Eastern Restaurant | Restaurant | Italian Restaurant |
| 15 | Glória | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Peruvian Restaurant | Asian Restaurant |
| 17 | Catete | Brazilian Restaurant | Italian Restaurant | Fast Food Restaurant | Sushi Restaurant | Chinese Restaurant |
| 19 | Botafogo | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Middle Eastern Restaurant | Japanese Restaurant |
| 22 | Leme | Brazilian Restaurant | Restaurant | Japanese Restaurant | Italian Restaurant | Seafood Restaurant |
| 23 | Copacabana | Brazilian Restaurant | Japanese Restaurant | Molecular Gastronomy Restaurant | Tapas Restaurant | French Restaurant |
| 25 | Leblon | Brazilian Restaurant | Italian Restaurant | Restaurant | Japanese Restaurant | Thai Restaurant |
| 26 | Lagoa | Restaurant | Brazilian Restaurant | American Restaurant | Mediterranean Restaurant | French Restaurant |

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 25 | Leblon | Brazilian Restaurant | Italian Restaurant | Restaurant | Japanese Restaurant | Thai Restaurant |
| 26 | Lagoa | Restaurant | Brazilian Restaurant | American Restaurant | Mediterranean Restaurant | French Restaurant |
| 28 | Gávea | Brazilian Restaurant | Japanese Restaurant | Sushi Restaurant | Restaurant | Middle Eastern Restaurant |
| 29 | Vidigal | Brazilian Restaurant | Italian Restaurant | Empanada Restaurant | Mediterranean Restaurant | Sushi Restaurant |
| 31 | Praça da Bandeira | Brazilian Restaurant | Restaurant | Fast Food Restaurant | American Restaurant | Southeastern Brazilian Restaurant |
| 35 | Vila Isabel | Brazilian Restaurant | Southeastern Brazilian Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| 38 | Manguinhos | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| 41 | Olaria | Brazilian Restaurant | Asian Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant |
| 44 | Braz de Pina | Brazilian Restaurant | Japanese Restaurant | Fast Food Restaurant | Mexican Restaurant | Restaurant |
| 45 | Cordovil | Brazilian Restaurant | Comfort Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 58 | Riachuelo | Asian Restaurant | Brazilian Restaurant | Restaurant | Portuguese Restaurant | Vegetarian / Vegan Restaurant |
| 62 | Méier | Brazilian Restaurant | Japanese Restaurant | Middle Eastern Restaurant | Restaurant | Portuguese Restaurant |
| 65 | Engenho de Dentro | Japanese Restaurant | Mexican Restaurant | Brazilian Restaurant | Restaurant | Fast Food Restaurant |

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 72 | Vicente de Carvalho | Brazilian Restaurant | Chinese Restaurant | Portuguese Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant |
| 73 | Vila da Penha | Brazilian Restaurant | Fast Food Restaurant | Chinese Restaurant | American Restaurant | Mexican Restaurant |
| 74 | Vista Alegre | Japanese Restaurant | Brazilian Restaurant | Mexican Restaurant | Restaurant | Fast Food Restaurant |
| 97 | Freguesia | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| 98 | Jardim Guanabara | Brazilian Restaurant | Japanese Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 99 | Jardim Carioca | Brazilian Restaurant | Sushi Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| 119 | Freguesia | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| 120 | Pechincha | Brazilian Restaurant | German Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 130 | Vargem Grande | Brazilian Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant | Japanese Restaurant | Mediterranean Restaurant |
| 146 | Cosmos | Brazilian Restaurant | Mineiro Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 148 | Santa Cruz | Brazilian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| 153 | Rocinha | Brazilian Restaurant | Japanese Restaurant | Fast Food Restaurant | Sushi Restaurant | Asian Restaurant |
| 157 | Vasco da Gama | Brazilian Restaurant | Restaurant | Portuguese Restaurant | Vegetarian / Vegan Restaurant | Empanada Restaurant |
| 160 | Lapa | Brazilian Restaurant | Restaurant | German Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant |

There are 40 neighborhoods in Cluster 4. We can easily notice that most neighborhoods in Cluster 4 have Brazilian Restaurant.

## Cluster 5

[83]:

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 103 | Galeão | Seafood Restaurant | Brazilian Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| 132 | Grumari | Seafood Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| 144 | Senador Vasconcelos | Mexican Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant | Empanada Restaurant | Italian Restaurant |
| 151 | Barra de Guaratiba | Brazilian Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |

There are 4 neighborhoods in Cluster 5. We can easily notice that most neighborhoods in Cluster 5 have SeaFood Restaurant.

**Cluster 6**

| [84]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 4 | Centro | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 46 | Parada de Lucas | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 49 | Higienópolis | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 61 | Lins de Vasconcelos | Restaurant | Sushi Restaurant | Vegetarian / Vegan Restaurant | Empanada Restaurant | Italian Restaurant |
| | 92 | Cacuia | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 96 | Bancários | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 107 | Parque Anchieta | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 111 | Barros Filho | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 117 | Cidade de Deus | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 123 | Praça Seca | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 128 | Camorim | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |
| | 140 | Bangu | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |

There are 12 neighborhoods in Cluster 6. We can easily notice that most neighborhoods in Cluster 1 have Restaurant and Vegetarian Restaurant.

**Cluster 7**

| [60]: | | Bairro | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| | 110 | Acari | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant | Asian Restaurant |

There is 1 neighborhoods in Cluster 7. Vegetarian Restaurant

**Cluster 8**

| [85]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 1 | Gamboa | Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 59 | Sampaio | Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 69 | Abolição | Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 70 | Pilares | Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 77 | Campinho | Fast Food Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |
| | 145 | Inhoaíba | Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant | Argentinian Restaurant |

There are 6 neighborhoods in Cluster 8. Fast Food

**Cluster 9**

| : | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 9 | I. de São Cristóvão | American Restaurant | Restaurant | Portuguese Restaurant | Empanada Restaurant | Italian Restaurant |
| | 76 | Colégio | American Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant | Brazilian Restaurant |

There are 2 neighborhoods in Cluster 9, American and Thai Restaurant

**Cluster 10**

| [87]: | | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|
| | 126 | Itanhangá | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant | Brazilian Restaurant |

There is 1 neighborhoods in Cluster 10. Vegetarian and Thai Restaurant.

## Cluster11

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 40 | Ramos | Brazilian Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 52 | Del Castilho | Fast Food Restaurant | Brazilian Restaurant | Middle Eastern Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |
| 75 | Irajá | Brazilian Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 81 | Cascadura | Brazilian Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 129 | Vargem Pequena | Brazilian Restaurant | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant |
| 156 | Maré | Brazilian Restaurant | Fast Food Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Jewish Restaurant |

There are 6 neighborhoods in Cluster11. Brazilian Restaurant and Fast Food

## Cluster 12

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 11 | Benfica | Asian Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Brazilian Restaurant |

There is 1 neighborhood in Cluster 12. Asian Restaurant

## Cluster 13

| | Bairro | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 43 | Penha Circular | Dumpling Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Argentinian Restaurant | Asian Restaurant |

There is 1 neighborhood in Cluster 13. Dumpling Restaurant.

### *Cluster 14. It doesn´t apper any neighborhood*

Brazilian restaurants as the most common restaurants.

## 7. Discussion section.

We have some database problems. Per example: The Copacabana has just two Brazilians Restaurants, The classification of restaurants has problems as well, the most common type of restaurant is "restaurant". The vegetarian restaurant appears in a lot of neighborhoods, we don´t see in Brazil too much old people vegetarian, the young people has preference with this kind of restaurant, they have more affinity with technology, the people who update the app are the youngest, we have this bias in the database.

Based the database for each Neighborhood above, I believe that classification for each cluster can be done better with calculation of restaurants (most common) in each Neighborhood. Refering to each cluster, I can't achieve clearly what represent in each cluster by using Foursquare - Most Common Restaurant data.

Rio de Janeiro is a Big City with a high population density in a narrow área and high income difference. This project, I only consider one factor i.e. frequency of occurrence of Restaurants there are. other factors such as population and income of residents that could influence the location decision

8. **Conclusion**

I would like to conclude saying this project gave me the opportunity to use python and applying the data science methodologies in a real problem. For this Project, I used CRISP-DM, I learned how important is to know and validate the data, I have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 14 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders regarding the best locations to open a Brazilian restaurant.

**The answer proposed by this project is: The neighbourhoods in cluster 3,6,7,8 and 13 are the most preferred locations to open a Brazilian Restaurant**.