

Cluster of types of restaurants in the city of Salvador using foursquare api.

Elton Moreno Costa Lima

June 08/2021

Report

The final assignment of this course is the so-called "Capstone Project" in which many of the tools and methods learned over the past few months are applied in a challenge of the general idea p of a "Battle of Neighborhoods", for example, the analysis and useful preparation of data on locations in urban environments. I chose the city of Salvador for the case. To pass the final task, you need to create a Jupyter notebook with Python as the programming language and all code, as well as a report and final presentation.

Business problem

Salvador is a city that attracts many tourists, is a historic city, was the first capital of Brazil, has a full cultural diversity. For tourists, finding the right place to eat can be a challenge, however. This is just one reason to give tourists a good overview of what to eat where.

So, the goal I want to achieve with this exercise is to give a simple recommendation to tourists in Salvador: in which district of the city will you find a large number or even concentration of what types of restaurants? Where to eat Bahian food, where to get fast food? The target audience is foreign tourists.

Data

Foursquare is a U.S. technology company from New York with a focus on location data. Its technology and data power up apps like Apple's Maps, Uber, Twitter and many other household names. I will use foursquare data, such as restaurant name, ID, location and food category (vegetarian, Italian, etc.).

In addition, I'll use the districts/parts of Salvador city overview from Wikipedia:
https://pt.wikipedia.org/wiki/Categoria:Bairros_de_Salvador

Methodology

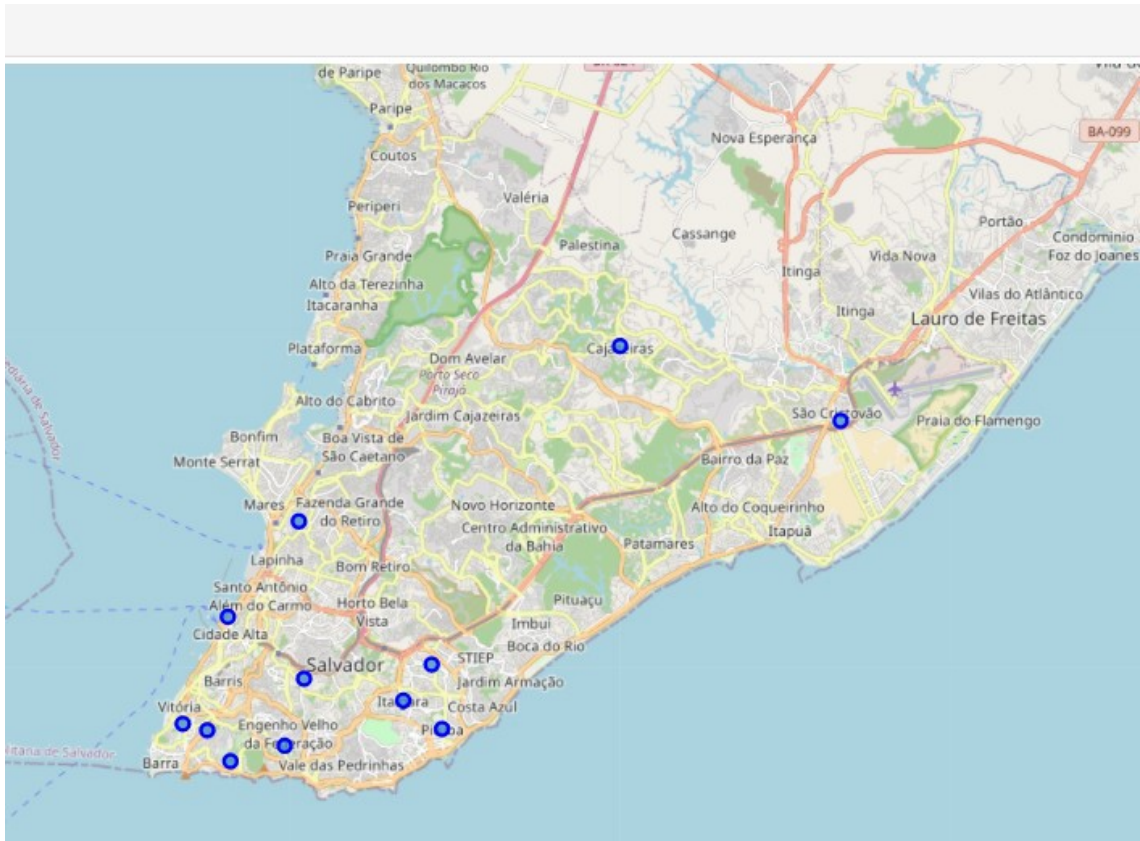
Extracting data from Wikipedia to create a dataframe with the districts of the city of Salvador. For this, I used the panda reading function. I had to clear the resulting data frame in terms of unnecessary information or data that could not be handled in a data frame.

We use geopy functions by installing the conda-forge geopy package. I used the nominatim function to add geospatial data to the data frame, which is latitude and longitude.

[6]:

	Bairros	Latitude	Longitude
0	Barra, Salvador	-12.999437	-38.518218
1	Brotas, Salvador	-12.986140	-38.492384
2	Cajazeiras, Salvador	-12.899938	-38.408157
3	Caminho das arvores, Salvador	-12.982453	-38.458303
4	Campo grande, Salvador	-11.744337	-38.447595
5	Comercio, Salvador	-12.970068	-38.512473
6	Graça, Salvador	-12.997703	-38.524663
7	Itaigara, Salvador	-12.991887	-38.465944
8	Itapoa, Salvador	-12.919344	-38.349614
9	Liberdade, Salvador	-12.945203	-38.493752
10	Ondina, Salvador	-13.007408	-38.511721
11	Pituba, Salvador	-12.999260	-38.455498
12	Rio vermelho, Salvador	-13.003685	-38.497519

Using the folium package to create map

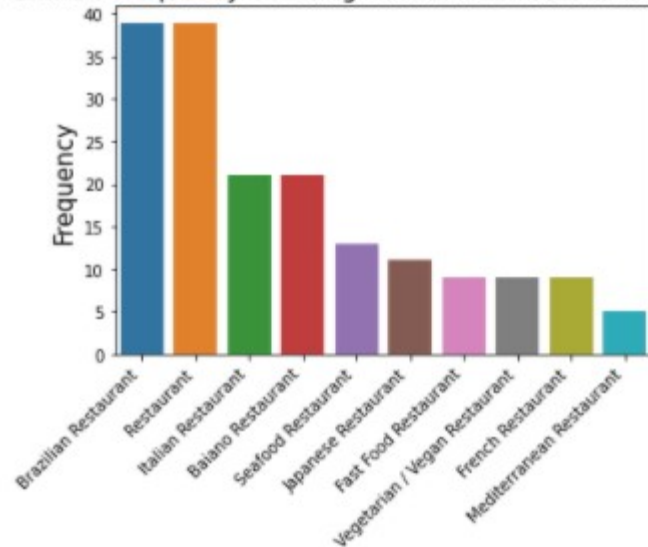


We checked which 10 categories are the most frequented.

	Venue_Category	Frequency
0	Brazilian Restaurant	39
1	Restaurant	39
2	Italian Restaurant	21
3	Baiano Restaurant	21
4	Seafood Restaurant	13
5	Japanese Restaurant	11
6	Fast Food Restaurant	9
7	Vegetarian / Vegan Restaurant	9
8	French Restaurant	9
9	Mediterranean Restaurant	5

We plotted a bar chart with the frequency of the 10 most frequently occurring restaurants throughout the city, using seaborn/matplotlib packages. We can see that Brazilian restaurants, Bahians are the most frequent restaurants in Salvador, which seems quite reasonable, and these cuisines are excellent and highly appreciated by a large part of the population.

10 mais Frequently Occuring Venues in 10 bairros of Salvador



To find clusters of restaurant types in the different neighborhoods of the city, I first transformed the data board with the restaurant locations, associated with the city's neighborhoods, by single coding (0/1), as seen in the image below.

[34]:

Neighborhood	African Restaurant	Argentinian Restaurant	Asian Restaurant	Baiano Restaurant	Brazilian Restaurant	Chinese Restaurant	Dumpling Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Latin American Restaurant	Mediterranean Restaurant
1 Barra, Salvador	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2 Barra, Salvador	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Barra, Salvador	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
4 Barra, Salvador	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 Barra, Salvador	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

I then used the grouping to show the frequency of each restaurant category in each district of the city.

[35]:

Neighborhood	African Restaurant	Argentinian Restaurant	Asian Restaurant	Baiano Restaurant	Brazilian Restaurant	Chinese Restaurant	Dumpling Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Latin American Restaurant	Mediterranean Restaurant
0 Barra, Salvador	0.000000	0.000000	0.000000	0.000000	0.100000	0.0000	0.050000	0.000000	0.050000	0.050000	0.000000	0.150000	0.100000	0.000000	0.050000
1 Brotas, Salvador	0.000000	0.055556	0.000000	0.111111	0.166667	0.0000	0.000000	0.000000	0.111111	0.000000	0.055556	0.055556	0.055556	0.000000	0.055556
2 Cajazeiras, Salvador	0.000000	0.000000	0.000000	0.000000	0.142857	0.0000	0.000000	0.571429	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3 Caminho das Amélias, Salvador	0.000000	0.000000	0.000000	0.200000	0.150000	0.0000	0.050000	0.050000	0.000000	0.000000	0.000000	0.150000	0.000000	0.050000	0.000000
4 Comércio, Salvador	0.000000	0.000000	0.000000	0.166667	0.166667	0.0000	0.000000	0.000000	0.111111	0.000000	0.000000	0.000000	0.055556	0.000000	0.055556
5 Graça, Salvador	0.000000	0.000000	0.000000	0.040000	0.080000	0.0000	0.040000	0.000000	0.080000	0.040000	0.000000	0.080000	0.120000	0.000000	0.040000
6 Itaipava, Salvador	0.000000	0.000000	0.055556	0.166667	0.166667	0.0000	0.000000	0.055556	0.000000	0.000000	0.000000	0.166667	0.000000	0.055556	0.000000
7 Itapoa, Salvador	0.000000	0.000000	0.125000	0.000000	0.312500	0.0625	0.000000	0.125000	0.000000	0.000000	0.000000	0.062500	0.062500	0.000000	0.062500
8 Liberdade, Salvador	0.111111	0.000000	0.000000	0.111111	0.222222	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9 Ondina, Salvador	0.000000	0.047619	0.000000	0.047619	0.190476	0.0000	0.047619	0.000000	0.047619	0.047619	0.047619	0.047619	0.095238	0.000000	0.000000
10 Pituba, Salvador	0.000000	0.000000	0.000000	0.173913	0.217391	0.0000	0.000000	0.043478	0.000000	0.000000	0.000000	0.130435	0.000000	0.043478	0.000000
11 Rio Vermelho, Salvador	0.000000	0.040000	0.040000	0.080000	0.240000	0.0400	0.000000	0.000000	0.040000	0.000000	0.040000	0.160000	0.040000	0.000000	0.000000

I used this information to create a data frame in which you can see the most common restaurant location types for each district in the city.

[38]:

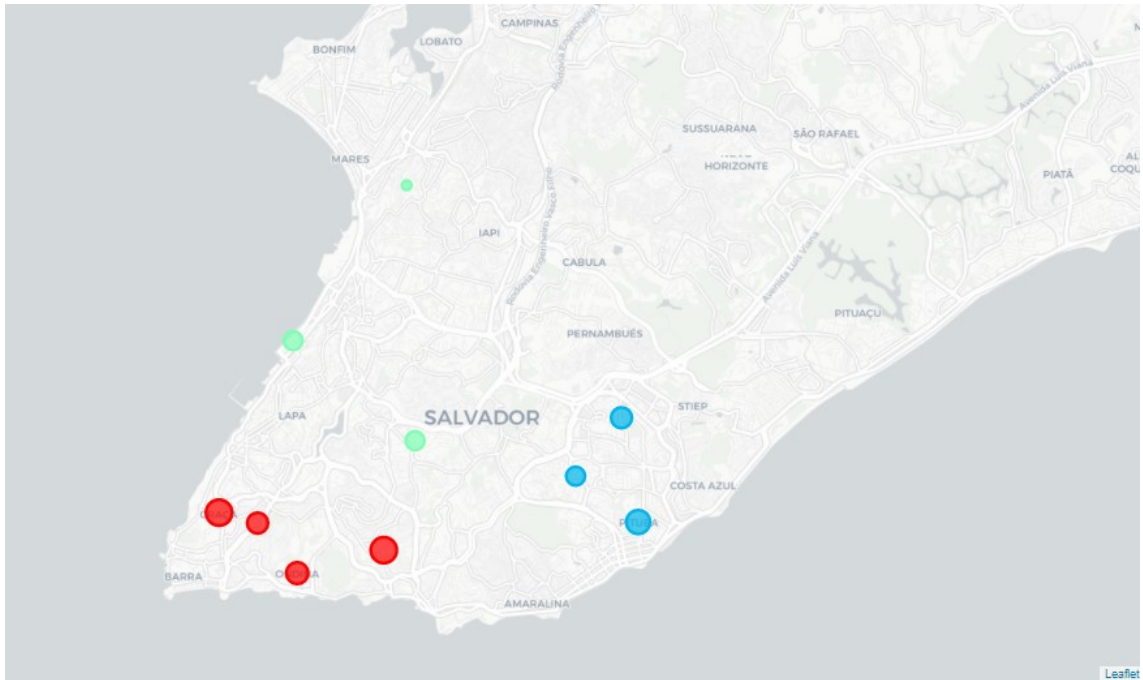
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barra, Salvador	Italian Restaurant	Restaurant	Brazilian Restaurant	Japanese Restaurant	Portuguese Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	French Restaurant
1	Brotas, Salvador	Restaurant	Brazilian Restaurant	French Restaurant	Balano Restaurant	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Indian Restaurant	Tapas Restaurant
2	Cajazeiras, Salvador	Fast Food Restaurant	Brazilian Restaurant	Seafood Restaurant	Restaurant	Vietnamese Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Balano Restaurant	Chinese Restaurant
3	Caminho das arvores, Salvador	Balano Restaurant	Restaurant	Italian Restaurant	Brazilian Restaurant	Seafood Restaurant	Latin American Restaurant	Dumpling Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
4	Comercio, Salvador	Restaurant	Balano Restaurant	Brazilian Restaurant	French Restaurant	Vietnamese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Indian Restaurant
5	Graça, Salvador	Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Brazilian Restaurant	French Restaurant	Italian Restaurant	Balano Restaurant	German Restaurant	Swiss Restaurant	Sushi Restaurant
6	Itaigara, Salvador	Italian Restaurant	Balano Restaurant	Brazilian Restaurant	Restaurant	Seafood Restaurant	Latin American Restaurant	Asian Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
7	Itapoa, Salvador	Brazilian Restaurant	Asian Restaurant	Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant
8	Liberdade, Salvador	Restaurant	Brazilian Restaurant	African Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Dumpling Restaurant
9	Ondina, Salvador	Brazilian Restaurant	Restaurant	Japanese Restaurant	Portuguese Restaurant	Balano Restaurant	German Restaurant	Indian Restaurant	Italian Restaurant	Dumpling Restaurant	French Restaurant
10	Pituba, Salvador	Restaurant	Brazilian Restaurant	Balano Restaurant	Italian Restaurant	Seafood Restaurant	Latin American Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant	Asian Restaurant
11	Rio vermelho, Salvador	Brazilian Restaurant	Italian Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Portuguese Restaurant	Thai Restaurant	Tapas Restaurant	Argentinian Restaurant	Asian Restaurant	Seafood Restaurant

Now, with all this data, I could finally run an unsupervised machine learning algorithm, more specifically, a k-means clustering algorithm from the scikit-learn package. One can use the elbow method to systematically set the k value, but I simply chose k to be 5, having been inspired by one of the course courses to do so.

[45]:

	Cluster Labels	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	0	Barra, Salvador	Italian Restaurant	Restaurant	Brazilian Restaurant	Japanese Restaurant	Portuguese Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	French Restaurant
1	3	Brotas, Salvador	Restaurant	Brazilian Restaurant	French Restaurant	Balano Restaurant	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Indian Restaurant	Tapas Restaurant
2	1	Cajazeiras, Salvador	Fast Food Restaurant	Brazilian Restaurant	Seafood Restaurant	Restaurant	Vietnamese Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Balano Restaurant	Chinese Restaurant
3	2	Caminho das arvores, Salvador	Balano Restaurant	Restaurant	Italian Restaurant	Brazilian Restaurant	Seafood Restaurant	Latin American Restaurant	Dumpling Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
4	3	Comercio, Salvador	Restaurant	Balano Restaurant	Brazilian Restaurant	French Restaurant	Vietnamese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Indian Restaurant
5	0	Graça, Salvador	Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Brazilian Restaurant	French Restaurant	Italian Restaurant	Balano Restaurant	German Restaurant	Swiss Restaurant	Sushi Restaurant
6	2	Itaigara, Salvador	Italian Restaurant	Balano Restaurant	Brazilian Restaurant	Restaurant	Seafood Restaurant	Latin American Restaurant	Asian Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
7	4	Itapoa, Salvador	Brazilian Restaurant	Asian Restaurant	Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant
8	3	Liberdade, Salvador	Restaurant	Brazilian Restaurant	African Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Dumpling Restaurant
9	0	Ondina, Salvador	Brazilian Restaurant	Restaurant	Japanese Restaurant	Portuguese Restaurant	Balano Restaurant	German Restaurant	Indian Restaurant	Italian Restaurant	Dumpling Restaurant	French Restaurant
10	2	Pituba, Salvador	Restaurant	Brazilian Restaurant	Balano Restaurant	Italian Restaurant	Seafood Restaurant	Latin American Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant	Asian Restaurant
11	0	Rio vermelho, Salvador	Brazilian Restaurant	Italian Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Portuguese Restaurant	Thai Restaurant	Tapas Restaurant	Argentinian Restaurant	Asian Restaurant	Seafood Restaurant

Now we can use cluster labels to show the city marked with a specific cluster color on a map, again using folium



Cluster 1 - The Italian Food Cluster

```
[46]: salvador_merged.loc[salvador_merged['Cluster Labels'] == 0, salvador_merged.columns[[1] + list(range(5, salvador_merged.shape[1]))]]
```

```
[46]:
```

	Latitude	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	-12.999437	Restaurant	Brazilian Restaurant	Japanese Restaurant	Portuguese Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	French Restaurant
5	-12.997703	Vegetarian / Vegan Restaurant	Japanese Restaurant	Brazilian Restaurant	French Restaurant	Italian Restaurant	Balano Restaurant	German Restaurant	Swiss Restaurant	Sushi Restaurant
9	-13.007408	Restaurant	Japanese Restaurant	Portuguese Restaurant	Balano Restaurant	German Restaurant	Indian Restaurant	Italian Restaurant	Dumpling Restaurant	French Restaurant
11	-13.003685	Italian Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Portuguese Restaurant	Thai Restaurant	Tapas Restaurant	Argentinian Restaurant	Asian Restaurant	Seafood Restaurant

Cluster 2 - the Sea Food Cluster

```
[47]: salvador_merged.loc[salvador_merged['Cluster Labels'] == 1, salvador_merged.columns[[1] + list(range(5, salvador_merged.shape[1]))]]
```

```
[47]:
```

	Latitude	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
2	-12.999938	Brazilian Restaurant	Seafood Restaurant	Restaurant	Vietnamese Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Balano Restaurant	Chinese Restaurant

Cluster 3 - the Bahian Food Cluster

[48]: `salvador_merged.loc[salvador_merged['Cluster Labels'] == 2, salvador_merged.columns[[1] + list(range(5, salvador_merged.shape[1]))]]`

	Latitude	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
3	-12.982453	Restaurant	Italian Restaurant	Brazilian Restaurant	Seafood Restaurant	Latin American Restaurant	Dumpling Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
6	-12.991887	Balano Restaurant	Brazilian Restaurant	Restaurant	Seafood Restaurant	Latin American Restaurant	Asian Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant
10	-12.999260	Brazilian Restaurant	Balano Restaurant	Italian Restaurant	Seafood Restaurant	Latin American Restaurant	Mineiro Restaurant	Fast Food Restaurant	Argentinian Restaurant	Asian Restaurant

Cluster 4 - the Brazilian Food Cluster

[49]: `salvador_merged.loc[salvador_merged['Cluster Labels'] == 3, salvador_merged.columns[[1] + list(range(5, salvador_merged.shape[1]))]]`

	Latitude	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
1	-12.986140	Brazilian Restaurant	French Restaurant	Balano Restaurant	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Indian Restaurant	Tapas Restaurant
4	-12.970068	Balano Restaurant	Brazilian Restaurant	French Restaurant	Vietnamese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Indian Restaurant
8	-12.945203	Brazilian Restaurant	African Restaurant	Balano Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Dumpling Restaurant

Cluster 5 - the Asian Food Cluster

[50]: `salvador_merged.loc[salvador_merged['Cluster Labels'] == 4, salvador_merged.columns[[1] + list(range(5, salvador_merged.shape[1]))]]`

	Latitude	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
7	-12.919344	Asian Restaurant	Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant

It is really possible to define clusters of certain cuisines in the city of Salvador.

Sources

The primary source that we can not fail to mention is the courses of the ibm/coursera Professional Certificate of Data Science, it was of infinite value to me.

Examples of excellent solutions to the capstone project mentioned in coursera: [\(52\)](#)
[Applied Data Science Capstone Project — The Restaurant Battle of Neighborhoods in Cologne | LinkedIn, "LinkedIn"](#). <https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/>,
<https://medium.com/@radialee/capstone-project-the-battle-of-neighborhoods-in-tokyo-restaurants-45a503e65ff>, and various others.

<https://stackoverflow.com>

The notebook is available at [Coursera_Capstone_DataScienceIBM/Capstone Week 5.ipynb at main · eltonmoreno1/Coursera_Capstone_DataScienceIBM \(github.com\)](#)

Conclusion

We reached the goal presented at the beginning so tourists can see in the results which neighborhoods of the city best match their food cravings. This is just one example of fantastic data science uses cases that one can perceive applying technology that is available for free.

I thank the global data science community for sharing so many important things.

IBM/Coursera Data Science Professional Certificate opened the doors of knowledge from the world of data science, and made me more prepared and inspired to continue expanding the search for knowledge in data science my special thank you Coursera/IBM!