



Capstone Project – The Battle of the Neighborhoods

Coursera

Paulo Calabria, *PhD*
December 23, 2019

Week 1 - required to submit the following (23/Dec):

1. A description of the problem and a discussion of the background. (15 marks)
Problem Statement
2. A description of the data and how it will be used to solve the problem. (15 marks)

1 – Problem Statement

In the Americas there are many important cities that perform a leading role in the world economy. In particular we have the cities of New York, in the USA and the city of Sao Paulo in Brazil. Both cities have many things in common, like the fact that both are a melting pot of different cultures and both were built by many immigrants from various countries in Europe, Asia, Africa. In modern days, it would be helpful to have a way to compare the neighborhood of those cities to help the expatriates to choose places to go, to go for short or long assignment knowing in advance what they can encounter.

Research Question: What are similar neighborhoods I can find in Sao Paulo and New York City?

2 – Data description and how it will be used

We will need geo-location information about that specific borough and the neighborhoods for the two cities. In Sao Paulo I found that in a WiKi and built up a data base. For New York, the data was acquired based on Module 3.

We will need data about different venues in different neighborhoods of each specific borough. In order to obtain that information we will use "Foursquare" locational information. Foursquare data will be used for segmentation and clustering.

Using segmentation and clustering, the goal is to determine:

1. Similarity or dissimilarity of both cities.
2. Classification of area located inside each city and find similar clusters.
3. Generate data that can be used in the future for apps for the travelers.

Week2 – The final deliverables of the project will be (27/Dec):

1. A link to your Notebook on your Github repository, showing your code. (15 marks)
2. A full report consisting of all of the following components (15 marks) (incorporated in the Notebook):
 - Introduction where you discuss the business problem and who would be interested in this project.
 - Data where you describe the data that will be used to solve the problem and the source of the data.
 - Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
 - Results section where you discuss the results.
 - Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
 - Conclusion section where you conclude the report.
3. Your choice of a presentation or blogpost. (10 marks)

VI-Methodology

- Analyze Sao Paulo
- K-mean Cluster Sao Paulo
- Analyze New York
- K-mean Cluster New York

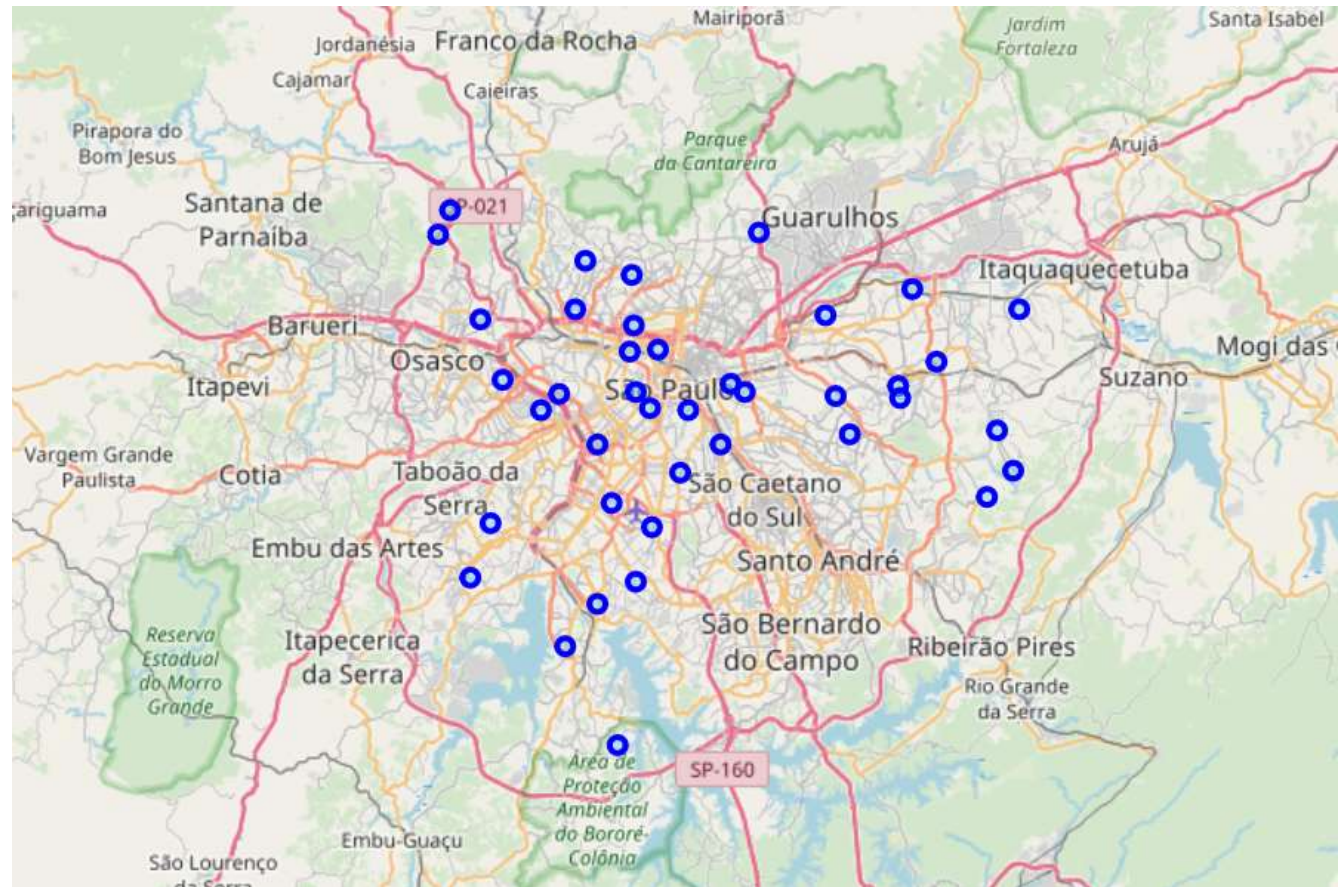
We found 89 “Bairros” in Sao Paulo with their Latitude and Longitude. For this work we used 40

For those 40 Boroughs we got all the information about the venues from Foursquare to perform a K-mean Clustering

```
Saopaulo_neighborhoods.head()
```

	Borough	Neighbourhood	Latitude	Longitude
0	Água Rasa	Água Rasa	-23.5532	-46.5819
1	Alto de Pinheiros	Alto de Pinheiros	-23.5537	-46.7088
2	Anhanguera	Anhanguera	-23.4544	-46.7902
3	Aricanduva	Aricanduva	-23.5795	-46.5110
4	Artur Alvim	Artur Alvim	-23.5488	-46.4772

“Bairros” plotted in Sao Paulo Map

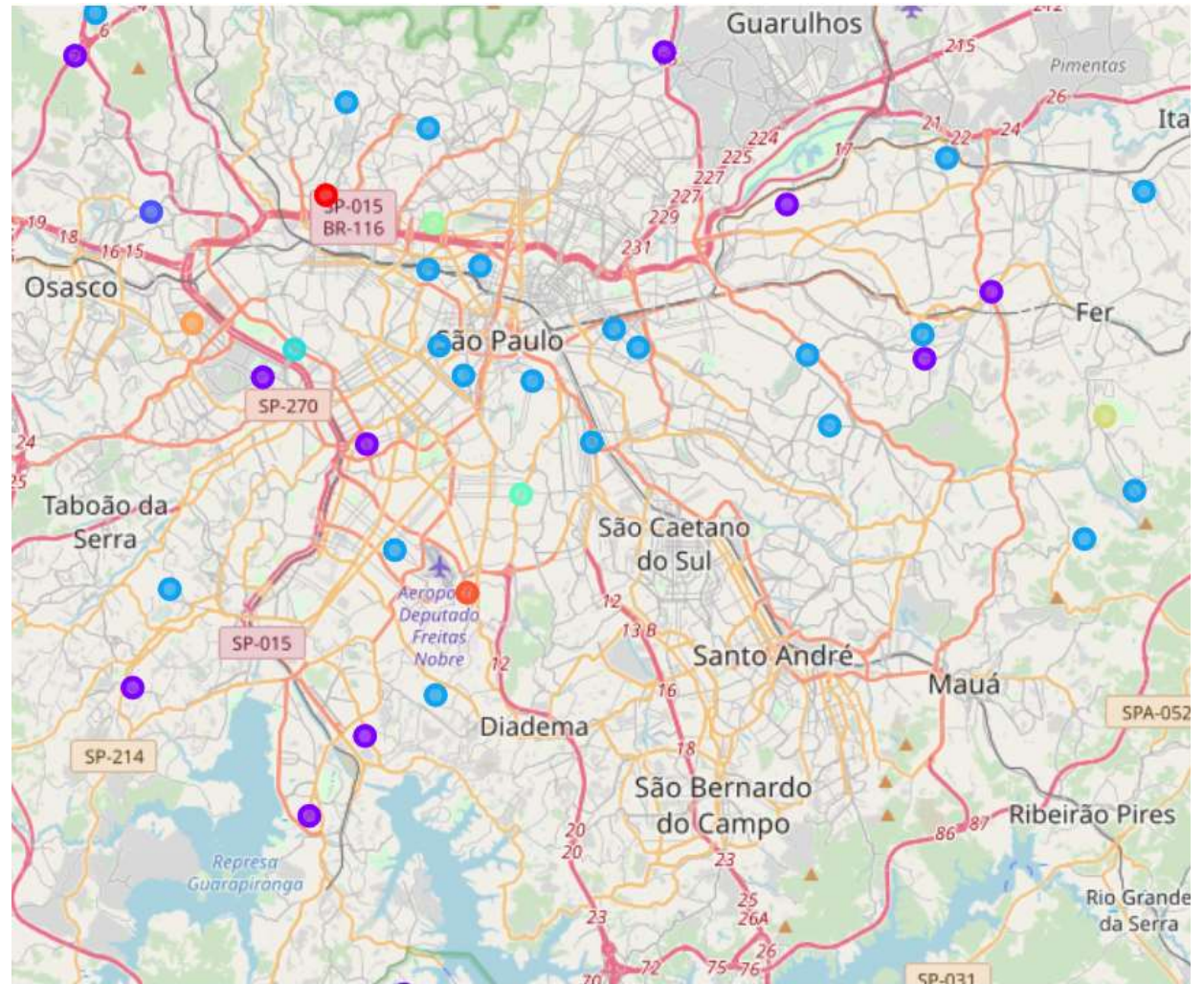


10 Most common venues per neighborhood (Sao Paulo)

```
spneighborhoods_venues_sorted.sample(5)
```

	Neighbourhood	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
22	Cidade Tiradentes	Snack Place	BBQ Joint	Flea Market	Fast Food Restaurant	Bus Station	Bakery	Pizza Place	Theater	Gym / Fitness Center	Dessert Shop
21	Cidade Líder	Bakery	Dessert Shop	Gym / Fitness Center	Convenience Store	Comfort Food Restaurant	Restaurant	Brazilian Restaurant	Market	Food Truck	Liquor Store
24	Cursino	Bakery	Acai House	Fruit & Vegetable Store	Bar	Pizza Place	Burger Joint	Middle Eastern Restaurant	Furniture / Home Store	Flower Shop	Medical Center
39	Água Rasa	Farmers Market	Brazilian Restaurant	Bakery	Gym	Sushi Restaurant	Pharmacy	Pastelaria	Convenience Store	Deli / Bodega	Diner
3	Artur Alvim	Soccer Stadium	Boxing Gym	Grocery Store	Stadium	Tour Provider	Burger Joint	Clothing Store	Farmers Market	Soccer Field	Event Space

Sao Paulo Clusters visualization



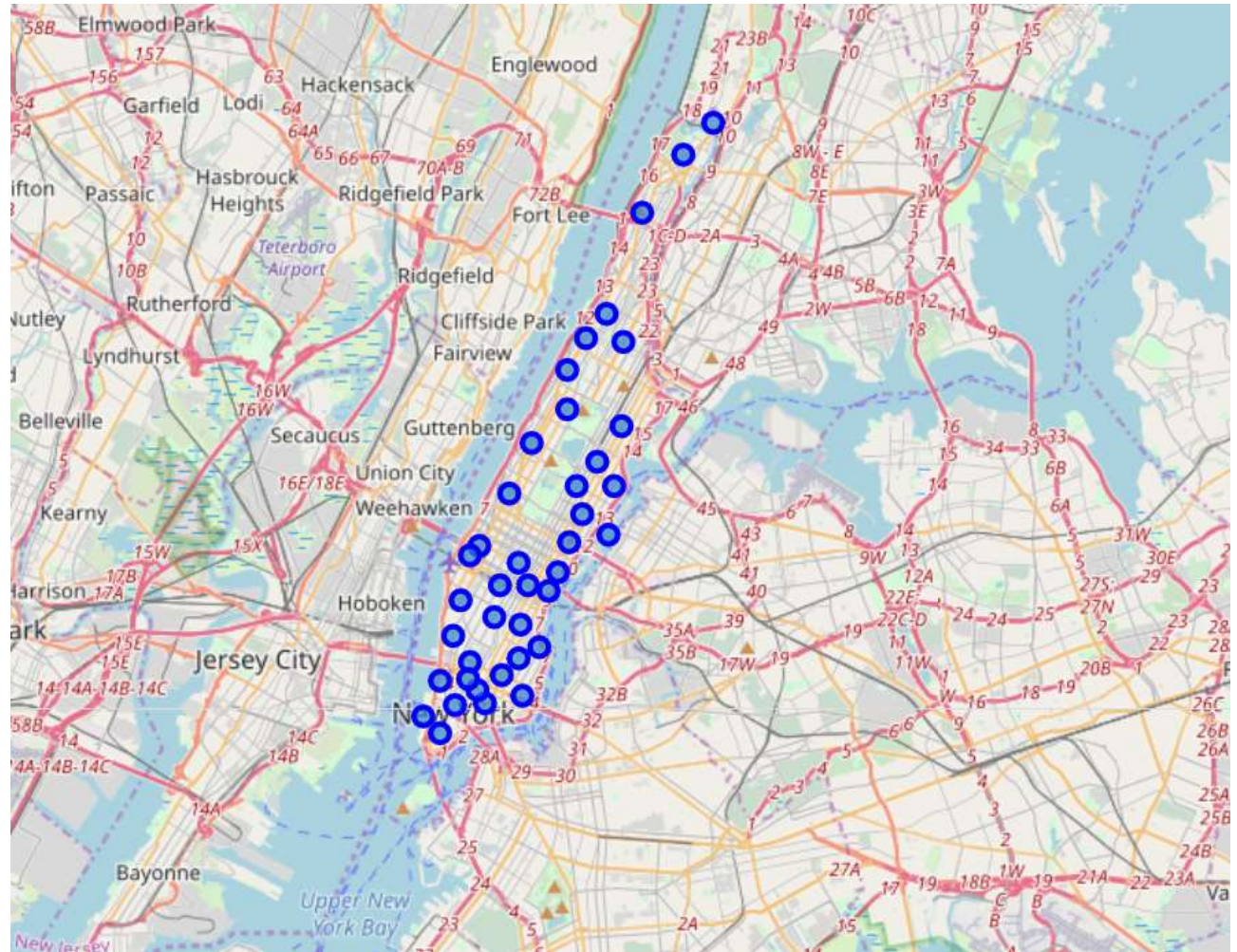
Sao Paulo Cluster 2 List

	Neighbourhood	Cluster Labels	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
3	Aricanduva	1	Pizza Place	Bakery	Gym / Fitness Center	Grocery Store	Pharmacy	Pastelaria	Fish Market	Fast Food Restaurant	Farmers Market	Exhibit
11	Cachoeirinha	1	Pizza Place	Food Truck	Bakery	Soccer Field	Shoe Store	Food Stand	Steakhouse	Supermarket	Dumpling Restaurant	Plaza
15	Campo Limpo	1	Food Truck	Plaza	Dumpling Restaurant	Gym / Fitness Center	Pizza Place	Construction & Landscaping	Dessert Shop	Park	Coffee Shop	Clothing Store
17	Capão Redondo	1	Bakery	Gym	Pizza Place	Plaza	Fast Food Restaurant	Convenience Store	Sandwich Place	Brazilian Restaurant	Soup Place	Health & Beauty Service
18	Carrão	1	Gym / Fitness Center	Bus Station	Juice Bar	Dessert Shop	Comfort Food Restaurant	Chocolate Shop	Sandwich Place	Fast Food Restaurant	Smoke Shop	Flower Shop
22	Cidade Líder	1	Bakery	Dessert Shop	Gym / Fitness Center	Convenience Store	Comfort Food Restaurant	Restaurant	Brazilian Restaurant	Market	Food Truck	Liquor Store
23	Cidade Tiradentes	1	Snack Place	BBQ Joint	Flea Market	Fast Food Restaurant	Bus Station	Bakery	Pizza Place	Theater	Gym / Fitness Center	Dessert Shop
29	Guaianases	1	Gym / Fitness Center	Food Truck	Pharmacy	Department Store	Shoe Store	Zoo Exhibit	Electronics Store	Flea Market	Fish Market	Fast Food Restaurant
33	Itaim Paulista	1	Dessert Shop	Bowling Alley	Japanese Restaurant	Pizza Place	Fast Food Restaurant	Shipping Store	Brewery	Brazilian Restaurant	Pharmacy	Food Truck
35	Jabaquara	1	Brazilian Restaurant	Japanese Restaurant	Park	Café	History Museum	Sushi Restaurant	Gym	Gym / Fitness Center	Hobby Shop	Massage Studio

Sao Paulo Cluster 4 List

	Neighbourhood	Cluster Labels	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	Água Rasa	3	Farmers Market	Brazilian Restaurant	Bakery	Gym	Sushi Restaurant	Pharmacy	Pastelaria	Convenience Store	Deli / Bodega	Diner
1	Alto de Pinheiros	3	Plaza	Convenience Store	Bar	Restaurant	Coffee Shop	Café	Business Service	Fast Food Restaurant	Flea Market	Food Truck
4	Artur Alvim	3	Soccer Stadium	Boxing Gym	Grocery Store	Stadium	Tour Provider	Burger Joint	Clothing Store	Farmers Market	Soccer Field	Event Space
5	Barra Funda	3	Nightclub	Theater	Pub	Museum	Korean Restaurant	Martial Arts Dojo	Lounge	Café	Farmers Market	Exhibit
6	Bela Vista	3	Bar	Antique Shop	Brazilian Restaurant	Gymnastics Gym	Italian Restaurant	Pizza Place	Plaza	Café	Coffee Shop	Jazz Club
7	Belém	3	Bar	Burger Joint	Restaurant	Gym	Ice Cream Shop	Market	Brazilian Restaurant	Chocolate Shop	Coffee Shop	Italian Restaurant
8	Bom Retiro	3	Women's Store	Café	Korean Restaurant	Diner	Pizza Place	Pastelaria	Chocolate Shop	Restaurant	Coffee Shop	Pharmacy
9	Brasilândia	3	Gym	Food Truck	Department Store	Dessert Shop	Convenience Store	Pizza Place	Diner	Comfort Food Restaurant	Recording Studio	Restaurant
10	Butantã	3	Science Museum	Track	Bus Station	Hot Dog Joint	Zoo Exhibit	Bakery	College Gym	College Stadium	Food Truck	Gym / Fitness Center
12	Cambuci	3	Brazilian Restaurant	Bakery	Restaurant	Market	Burger Joint	Sandwich Place	Chocolate Shop	Clothing Store	Food & Drink Shop	Spa
13	Campo Belo	3	Spa	Dessert Shop	Ice Cream Shop	Chocolate Shop	Beer Bar	Spanish Restaurant	Food & Drink Shop	Brazilian Restaurant	Burger Joint	Pie Shop
14	Campo Grande	3	Pizza Place	Bakery	Market	Snack Place	Bar	Juice Bar	Photography Studio	Gym	Playground	Restaurant
16	Cangaíba	3	Plaza	Bakery	Music Venue	Chocolate Shop	Café	Soccer Field	Bookstore	Steakhouse	Pharmacy	Gay Bar
19	Casa Verde	3	Gym / Fitness Center	Restaurant	Bar	Pizza Place	Dessert Shop	Burger Joint	Shoe Store	Seafood Restaurant	Plaza	Pet Store
21	Cidade Dutra	3	Dessert Shop	Bar	Brazilian Restaurant	Pharmacy	Department Store	Burger Joint	Plaza	Pet Store	Pizza Place	Steakhouse
24	Consolação	3	Bar	Dance Studio	Burger Joint	Food Truck	Gym / Fitness Center	Comedy Club	Miscellaneous Shop	Movie Theater	Baiano Restaurant	Fish Market
25	Cursino	3	Bakery	Acai House	Fruit & Vegetable Store	Bar	Pizza Place	Burger Joint	Middle Eastern Restaurant	Furniture / Home Store	Flower Shop	Medical Center
27	Freguesia do Ó	3	Burger Joint	Pizza Place	Bar	Acai House	Tapiocaria	Chocolate Shop	Martial Arts Dojo	Church	Brewery	Snack Place
31	Ipiranga	3	Warehouse Store	Gym / Fitness Center	Gastropub	Skate Park	Buffet	Burger Joint	Spa	Italian Restaurant	Bar	Supermarket
32	Itaim Bibi	3	Restaurant	Japanese Restaurant	Bar	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Burger Joint	Buffet	Lounge	Salon / Barbershop
34	Itaquera	3	Convenience Store	Pizza Place	Bakery	Martial Arts Dojo	Farmers Market	Electronics Store	Nightclub	Pharmacy	Rock Club	Café

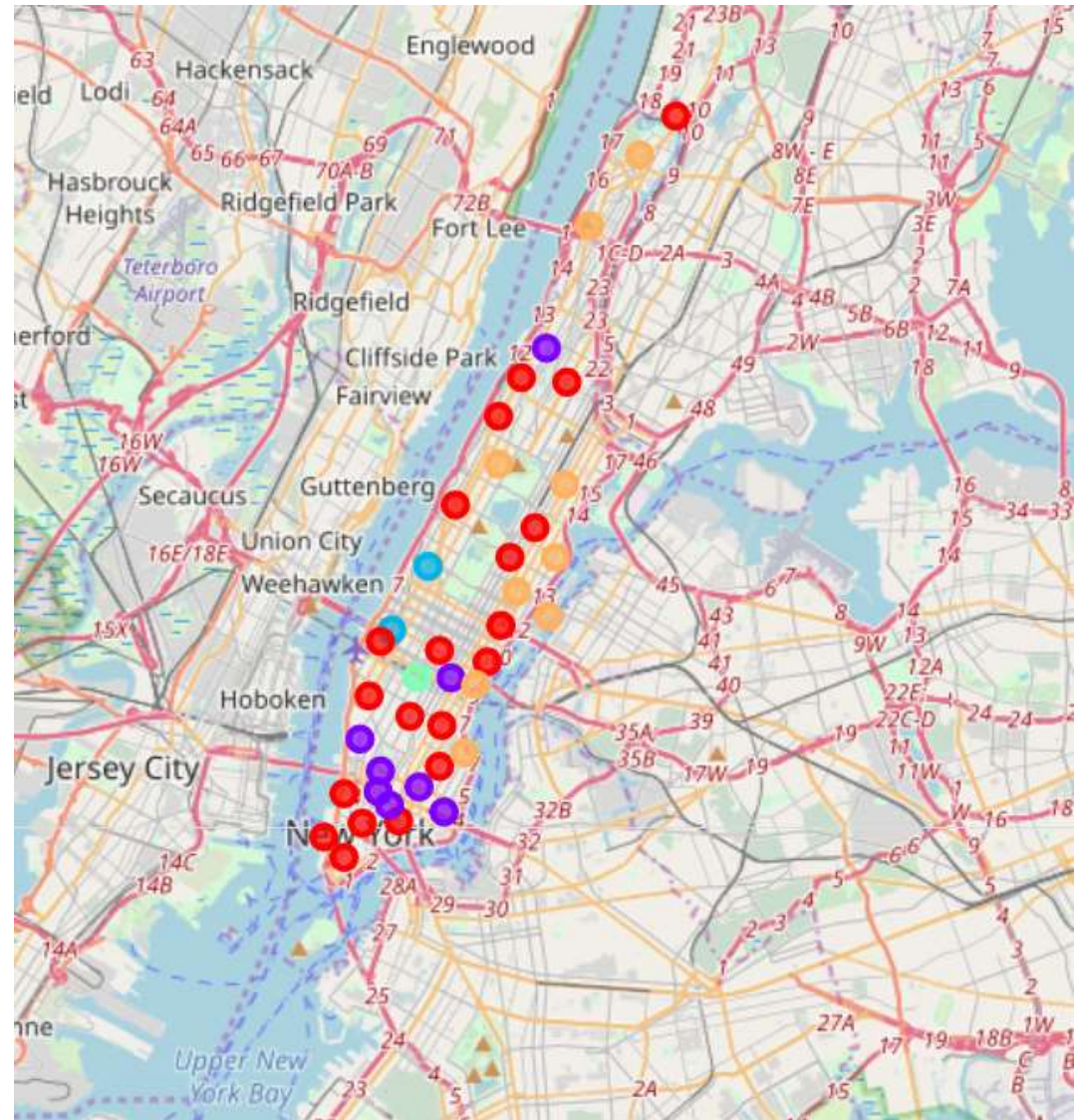
Nearby Neighborhoods plotted in New York (Manhattan) Map



10 Most common venues per neighborhood (Manhattan)

	Neighbourhood	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	Battery Park City	Park	Memorial Site	Coffee Shop	Sandwich Place	Food Court	Building	Shopping Mall	Sushi Restaurant	Burrito Place	Smoke Shop
1	Carnegie Hill	Gym	Bookstore	Coffee Shop	Pizza Place	Italian Restaurant	Gym / Fitness Center	Community Center	Café	French Restaurant	Spa
2	Central Harlem	French Restaurant	African Restaurant	American Restaurant	Bar	Cosmetics Shop	Bookstore	Beer Bar	Gym / Fitness Center	Library	Music Venue
3	Chelsea	Hotel	Italian Restaurant	Nightclub	Theater	Indian Restaurant	Chinese Restaurant	Beer Bar	Speakeasy	Café	Seafood Restaurant
4	Chinatown	Chinese Restaurant	Spa	Sandwich Place	Noodle House	Spanish Restaurant	Bubble Tea Shop	Salon / Barbershop	Roof Deck	Cocktail Bar	New American Restaurant

Manhattan Clusters Visualization



Manhattan Cluster 2 List

	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
4	Hamilton Heights	Coffee Shop	Yoga Studio	Café	Caribbean Restaurant	Cocktail Bar	Mexican Restaurant	Japanese Restaurant	School	Pizza Place	Park
16	Murray Hill	Japanese Restaurant	Coffee Shop	Hotel	Burger Joint	Italian Restaurant	Tea Room	Restaurant	Ramen Restaurant	Lounge	Speakeasy
18	Greenwich Village	Italian Restaurant	Café	French Restaurant	Sushi Restaurant	Clothing Store	Cosmetics Shop	Coffee Shop	Cocktail Bar	Salon / Barbershop	Beer Bar
20	Lower East Side	Café	Japanese Restaurant	Ramen Restaurant	Coffee Shop	Cocktail Bar	Art Gallery	Yoga Studio	Speakeasy	Pet Café	Mediterranean Restaurant
22	Little Italy	Café	Sandwich Place	Optical Shop	Ice Cream Shop	Wine Bar	Chinese Restaurant	French Restaurant	Spanish Restaurant	Martial Arts Dojo	Karaoke Bar
23	Soho	Women's Store	Clothing Store	Men's Store	Shoe Store	Boutique	Salon / Barbershop	Optical Shop	Ski Shop	French Restaurant	Miscellaneous Shop
24	West Village	Italian Restaurant	Cosmetics Shop	Cocktail Bar	Bakery	Chinese Restaurant	Gastropub	Gourmet Shop	Coffee Shop	Accessories Store	French Restaurant
31	Noho	French Restaurant	Coffee Shop	Wine Shop	Ice Cream Shop	Boutique	Cocktail Bar	Italian Restaurant	Rock Club	Hotel	Grocery Store

Manhattan Cluster 5 List

	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
2	Washington Heights	Café	Wine Shop	Bakery	Park	Indian Restaurant	Pool	Caribbean Restaurant	Scenic Lookout	Cocktail Bar	Coffee Shop
3	Inwood	Mexican Restaurant	Frozen Yogurt Shop	Park	Café	Bakery	American Restaurant	Deli / Bodega	Wine Bar	Diner	Dog Run
7	East Harlem	Mexican Restaurant	Bakery	Latin American Restaurant	Thai Restaurant	Pizza Place	Taco Place	Street Art	Steakhouse	Spanish Restaurant	French Restaurant
9	Yorkville	Deli / Bodega	Wine Shop	Park	Coffee Shop	Italian Restaurant	Liquor Store	Asian Restaurant	Monument / Landmark	Sandwich Place	Beer Store
10	Lenox Hill	Gym	Wine Shop	Burger Joint	Gym / Fitness Center	Thai Restaurant	Japanese Restaurant	Smoke Shop	French Restaurant	Salad Place	Liquor Store
11	Roosevelt Island	Coffee Shop	Park	Sandwich Place	Japanese Restaurant	Bus Line	Metro Station	Soccer Field	School	Scenic Lookout	Liquor Store
25	Manhattan Valley	Pizza Place	Bar	Coffee Shop	Yoga Studio	Hawaiian Restaurant	Park	Noodle House	Mexican Restaurant	Korean Restaurant	Italian Restaurant
36	Tudor City	Park	Pizza Place	Mexican Restaurant	Deli / Bodega	Thai Restaurant	Restaurant	Spanish Restaurant	Bridge	Burger Joint	Café
37	Stuyvesant Town	Boat or Ferry	Bar	Park	Baseball Field	Harbor / Marina	Gas Station	Farmers Market	German Restaurant	Cocktail Bar	Coffee Shop

VI – Discussion

With the data on the clusters for each cities above, we can have an idea of each one and the type of venues categories (most common) would you find in each cities. Our original idea is to use this data to help the expatriates to chose places to go, to go for short of long assignment, based on the similarities of the type of Neighborhood they will find.

We can certainly say that Sao Paulo has many "Bairros" (Neighborhoods) that are very similar to what you would find in Manhattan and the data helps to identify them

In general, with the initial data obtained for this project we can make the following assumptions for each cluster: If you are looking for places to stay and have many alternatives nearby:

- Cluster 2 and 5 in Manhattan
- Cluster 2 and 4 in Sao Paulo

One idea to enhance this initial project would be to go deeper in the correlation of the Neighborhood by finding a quantitative way to identify and distinguish their difference based on most common venues founded via Foursquare. Another idea, is to design an App where the user could give her/his most important venue that she or he would like to find in the target Neighborhood at the city and run this analysis and come back with possible suggestion. I think this would be very helpful and can be done in a future project.

I believe that this project is a first step towards a quantitative and systematic comparison of the different cities, and specially that is bring information about the city of Sao Paulo, which I could not find any similar work done before.

VII – Conclusion

Although I think the goals of this project were met (Applied Data Science Capstone Project) there is definitely room for further improvement and development as discussed above.

My initial experience as a Data Scientist trying to do this project and others, is that the time you spent in getting the right data to answer your initial "Research Question" represents the majority of the time spent. The right definition of the problem to be solved is key to make sure you go after the right data needed. In this aspects I may say from my experience that the Data Science work and methodology is very similar to scientific methodology, and the risks associated with finding out that your objective is too aggressive and you will need more analysis / data to accomplish are present.

Thank You !

Contact Info:

Paulo Cesar Calabria
pcalabria@fei.edu.br