



DISCOVERING VENUES IN SAO PAULO FOR NEWCOMERS: A CLUSTERING APPROACH

IVAN E. M. KÜHNE

JUNE 20, 2020



DISCOVERING VENUES IN SAO PAULO - BRAZIL

- Our first goal here is to use the Foursquare API to retrieve data about the venues that exist around some well know subway stations in the city of São Paulo;
- Then we will use the Python programming language to clean and analyse the retrieved data, grouping the stations according to the venues that are more common around each one;
- Our target audience are people that are coming to São Paulo, either as tourists or newcomers, and are interested in discovering the surroundings that are more suited to their interests;

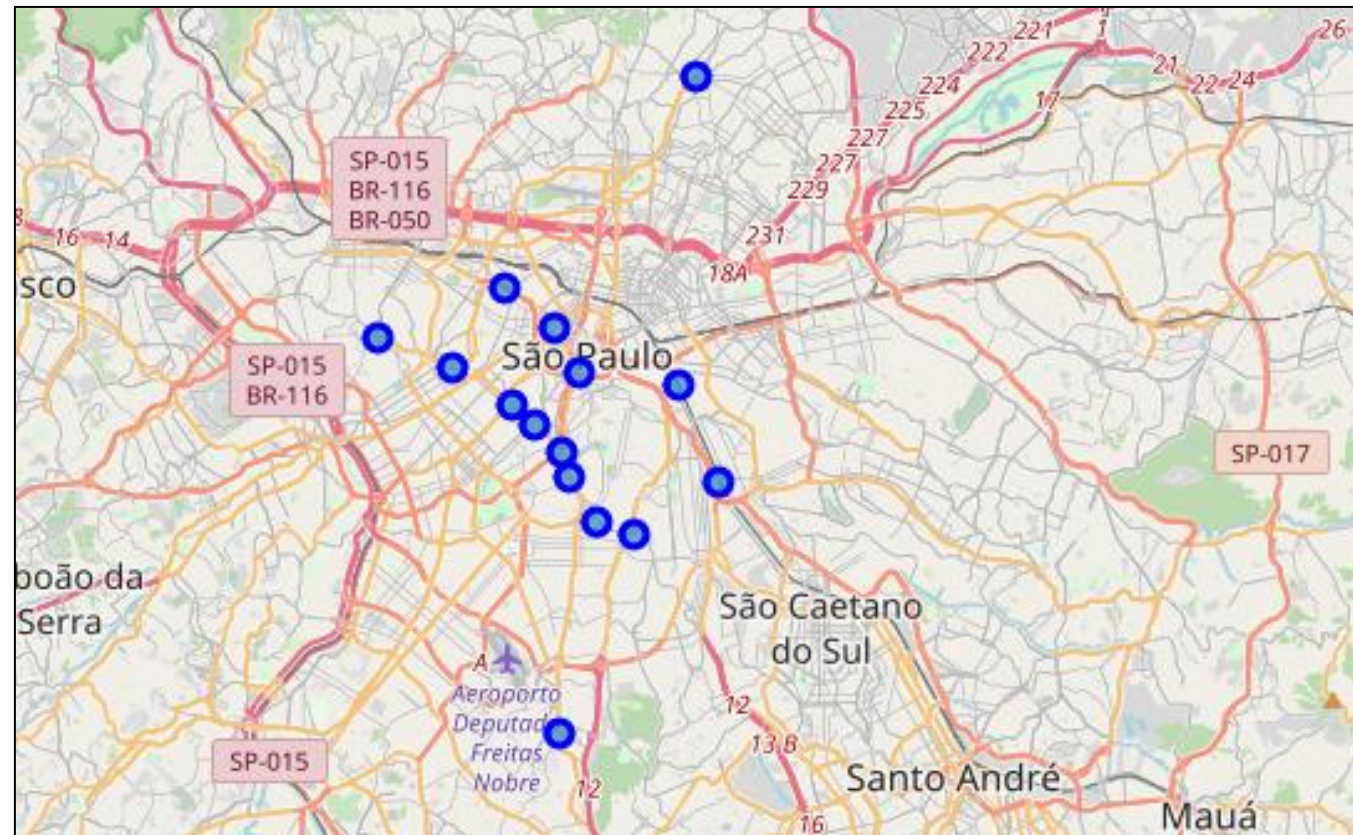
SAMPLE OF A MAP OF THE SAO PAULO SUBWAY STATIONS



SELECTED SUBWAY STATIONS AND NUMBER OF VENUES

Subway Station	Number of Venues
Ana Rosa	59
Brigadeiro	60
Chacara Klabin	51
Clinicas	29
Ipiranga	56
Jabaquara	42
Japao Liberdade	55
Juventus Mooca	24
Marechal Deodoro	33
Paraiso	47
Republica	100
Santos Imigrantes	34
Trianon MASP	88
Tucuruvi	99
Vila Madalena	49

LOCATION OF THE SELECTED STATIONS



EXAMPLE OF VENUES FREQUENCY: JAPAO LIBERDADE

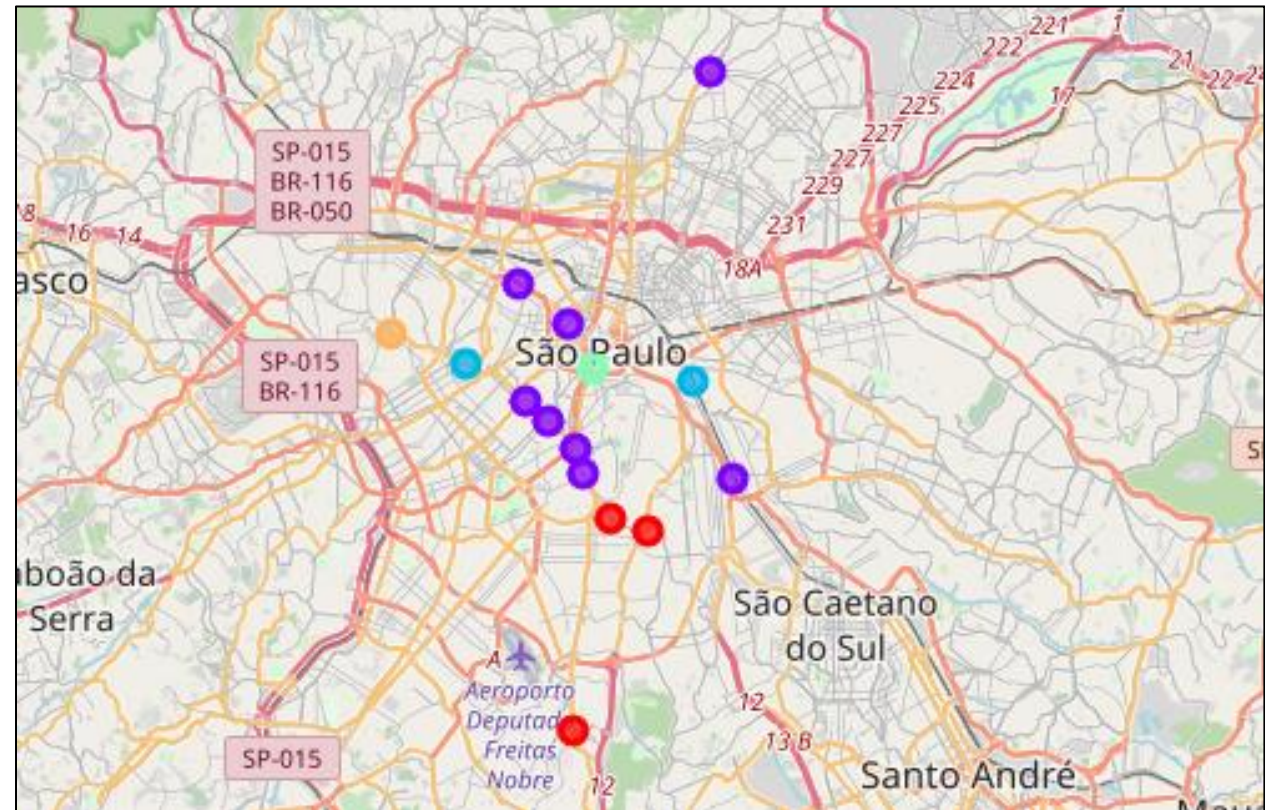
Order	Venue Category	Frequency
01	Japanese Restaurant	0.16
02	Sake Bar	0.09
03	Cosmetics Shop	0.05
04	Grocery Store	0.05
05	Bakery	0.04

EXAMPLE OF VENUES FREQUENCY: VILA MADALENA

Order	Venue Category	Frequency
01	Hostel	0.16
02	Burger Joint	0.08
03	Pharmacy	0.06
04	Gym	0.06
05	Thrift / Vintage Store	0.04

CLUSTERED SUBSTATIONS

Subway Station	Cluster
Ana Rosa	1
Brigadeiro	1
Chacara Klabin	0
Clinicas	2
Ipiranga	1
Jabaquara	0
Japao Liberdade	3
Juventus Mooca	2
Marechal Deodoro	1
Paraiso	1
Republica	1
Santos Imigrantes	0
Trianon MASP	1
Tucuruvi	1
Vila Madalena	4





CLUSTER 0

The Cluster 0 includes the subway stations Chacara Klabin, Jabaquara and Santos Imigrantes. Looking for similarities between these surroundings, we see that, for people interested in food, there a lot of pizza places, food trucks, snack places. But, if you are feeling sick, there are a lot of pharmacies in this cluster. But, if you are interested in keeping yourself healthy before feeling sick, there are a lot of gyms and fitness centers.



CLUSTER 1

The Cluster 1 includes the subway stations Ana Rosa, Brigadeiro, Ipiranga, Marechal Deodoro, Paraiso, Republica, Trianon MASP and Tucuruvi. Looking for similarities between these surroundings, we see that many of them have Brazilian restaurants, restaurants, coffe shops and chocolate shops. At first glance, none of these are novelty for Brazilian people, being more attractive to people from outside the country that want to discover the local recipes.



CLUSTER 1

But, looking a little deeper, we find some exotic places that may please some specific people, even when born in Brazil. Around the Station Marechal Deodoro, we can find restaurants specialized in kosher and Korean food that we do not see around the other stations. For people interested in these cuisines, it is the perfect place to explore. For people interested in Middle Eastern cuisine, the recommendation is to explore the surroundings of the stations Ana Rosa, Brigadeiro and Paraiso.



CLUSTER 2

The Cluster 2 includes the subway stations Clinicas and Juventus Mooca. Looking for similarities between these surroundings, we can find a prevalence of Brazilian restaurants, cafés, pizza places and restaurants in general. For people interested in another kinds of fun, we can see that around the station Juventus Mooca there are music venues and gaming cafes, venues that we don't find around the other subway stations.



CLUSTER 3

The Cluster 3 includes only the subway station Japao Liberdade. The top ten venue categories in this neighborhood are japanese restaurant, sake bar, grocery store, cosmetics shop, sushi restaurant, theater, bakery, bookstore, nightclub and asian restaurant. As presented before, this neighborhood congregates a lot of people with Asian ancestry, specially from Japan and China. For newcomers, this is a great place to try the Japanese and Asian food.



CLUSTER 4

The Cluster 4 includes only the subway station Vila Madalena. The top ten venue categories in this neighborhood are hostel, burger joint, pharmacy, gym, farmers market, bakery, thrift/vintage store, ice cream shop, health food store and dessert shop. As presented before, this neighborhood is a gathering point for the bohemian youth. For newcomers that are adept of this lifestyle, it is a great place to meet people with similar interests in the hostels, burger joints, bakeries and dessert shops.



DISCUSSION

- Each cluster has a different atmosphere, being capable of attracting people with different tastes, lifestyles and interests;
- Cluster 3 and 4 have only one station each, that have venues around them that are very different from the ones that are found in the other clusters;
- Cluster 2 have two stations, and one of them, Juventus Mooca, has some venues that are not found anywhere else: gaming cafes and music venues;
- Cluster 0 and 1 have more stations, that share a lot of similar venues arounding the stations that are grouped in each one.



CONCLUSION

As future directions, the analysis techniques can be deepened to deal with some ambiguities in the naming of some categories of venues that lead us to think that some of them describe the same type of venues, despite having different names, like "Café" and "Coffe Shop". The same occurs with "Churrascaria", the Portuguese word for the restaurant that serves the Brazilian BBQ, and perhaps can be merged with the "BBQ Joint" category.