



Presentation Capton Project

The Battle of Neighborhoods in SAO PAULO-BR

Big diversity of the São Paulo city

- The city of Sao Paulo, Sao Paulo state in Brazil, is one of the bigger world city with more of twenty millions of habitants.
- The city has multiple territory division, some of which are not compatible. This incompatibility causes doubts about spatial references, for both the population and the state representatives.
- For this work, it will have use of three territory divisions:
 - First level: The 32 sub prefectures.
 - Second level: The 96 districts (each district belongs to one sub prefecture).]
 - Third level: The hundreds of neighborhoods.

V · D · E		Subdivisões da cidade de São Paulo	[Esconder]
Zonas político-administrativas		Centro e Centro histórico · Centro-Sul · Leste 1 e 2 · Nordeste · Noroeste · Norte · Oeste · Sudeste · Sudoeste · Sul	
Subprefeituras		Aricanduva/Formosa/Carrão · Butantã · Campo Limpo · Capela do Socorro · Casa Verde · Cidade Ademar · Cidade Tiradentes · Ermelino Matarazzo · Freguesia do Ó/Brasilândia · Guaianases · Ipiranga · Itaim Paulista · Itaquera · Jabaquara · Jaconã/Tremembé · Lapa · M'Boi Mirim · Mooca · Parelheiros · Penha · Perus · Pinheiros · Pirituba/Jaraguá · Santana/Tucuruvi · Santo Amaro · São Mateus · São Miguel Paulista · Sapopemba · Sé · Vila Maria/Vila Guilherme · Vila Mariana · Vila Prudente	
Distritos		Água Rasa · Alto de Pinheiros · Anhangüera · Aricanduva · Artur Alvim · Barra Funda · Bela Vista · Belém · Bom Retiro · Brasilândia · Brás · Butantã · Cachoeirinha · Cambuci · Campo Belo · Campo Grande · Campo Limpo · Cangaíba · Capão Redondo · Carrão · Casa Verde · Cidade Ademar · Cidade Dutra · Cidade Líder · Cidade Tiradentes · Consolação · Cursino · Ermelino Matarazzo · Freguesia do Ó · Grajaú · Guaianases · Iguatemi · Ipiranga · Itaim Bibi · Itaim Paulista · Itaquera · Jabaquara · Jaconã · Jaguará · Jaguaré · Jaraguá · Jardim Helena · Jardim Paulista · Jardim São Luís · Jardim Ângela · José Bonifácio · Lajeado · Lapa · Liberdade · Limão · Mandaqui · Marsilac · Moema · Mooca · Morumbi · Parelheiros · Pari · Parque do Carmo · Pedreira · Penha · Perdizes · Perus · Pinheiros · Pirituba · Ponte Rasa · Raposo Tavares · República · Rio Pequeno · Sacomã · Santa Cecília · Santana · Santo Amaro · São Domingos · São Mateus · São Miguel Paulista · São Lucas · São Rafael · Saúde · Sapopemba · Sé · Socorro · Tatuapé · Tremembé · Tucuruvi · Vila Andrade · Vila Curuçá · Vila Formosa · Vila Guilherme · Vila Jacuí · Vila Leopoldina · Vila Maria · Vila Mariana · Vila Matilde · Vila Medeiros · Vila Prudente · Vila Sônia	

The clustering of these neighborhoods, in function of the more represented venues categories and in function of socio-economic indicators, should give interesting insights to help deciding where open a new business.

Data acquisition

- The list of the neighborhood by the scrap of one table in the Web Site page:

<https://www.agenteimovel.com.br/imoveis/a-venda/sp/sao-paulo/>

<https://www.tudoaquisaopaulo.com.br/bairros/sao-paulo-sp>

- The latitude and longitude coordinates of the center of each neighborhood calling the API of the provider ARCGIS with the Python library Geocoder.
- The limits of the Sao Paulo city sub prefectures (32) and districts (96), downloading shape files from the official Sao Paulo web site

http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx

- The IDH (Human Development Index) by district of Sao Paulo City by scrap of the Brazilian Wiki page

https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_S%C3%A3o_Paulo_por_%C3%8Dndice_de_Desenvolvimento_Humano

- The venues of each neighborhood has been recovered calling a API of the Foursquare site.

Data transformation/cleaning

- List of the neighborhoods

Like it has not be possible to find information about the complete list of the neighborhood (more of 1700), it has been used the limited list of 525 neighborhood, scraped with some information about the real state.

- Conversion of the district/sub prefecture coordinates limits from unit UTM Cartesian (EPSG:29193) to spherical system (latitude/longitude degrees EPSG:4326.

To can use the library folium, to mount the map of the Sao Paulo city with his division.

- Sub prefecture and district of each neighborhood

Like it hasn't been found a web site with the indications, easy to recover, of the sub prefecture and district of each Sao Paulo city neighborhood, it has been used a geometric approach to define these: for each neighborhood loop on all the sub prefectures and all districts to find these whose the polygon contains the neighborhood central point.

Features selection

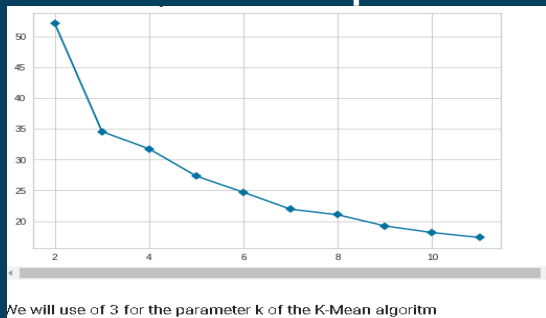
- The sub prefecture and district limits are used to show these administrative limits in a Sao Paulo map.
- The center neighborhood coordinates are showed in the same map to have an idea of the selected neighborhoods distribution between the different sub prefectures and districts of Sao Paulo city.
- The first cluster of these neighborhoods has been mounted with the top ten venues categories of each neighborhood.
- The second cluster of these neighborhoods has been mounted with the normalized data about the IDH and price/m2 of each neighborhood.

Vila Uberabinha
MOEMA/VILA MARIANA
-23.60078-
-46.67094

MOEMA

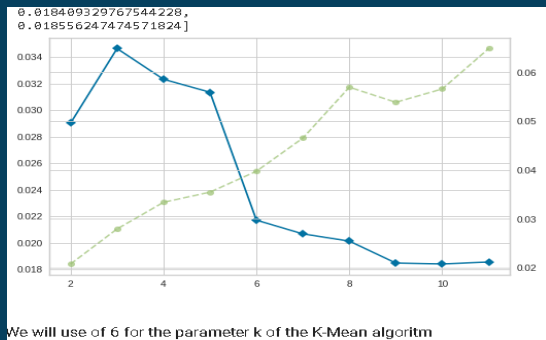
Clustering of the neighborhood using the K-Means ML algorithm

- By the TOP 10 venue categories:
 - Verification of the optimum k hyper parameter by the elbow method of the library yellowbrick.cluster



➤ Repartition in 3 clusters.

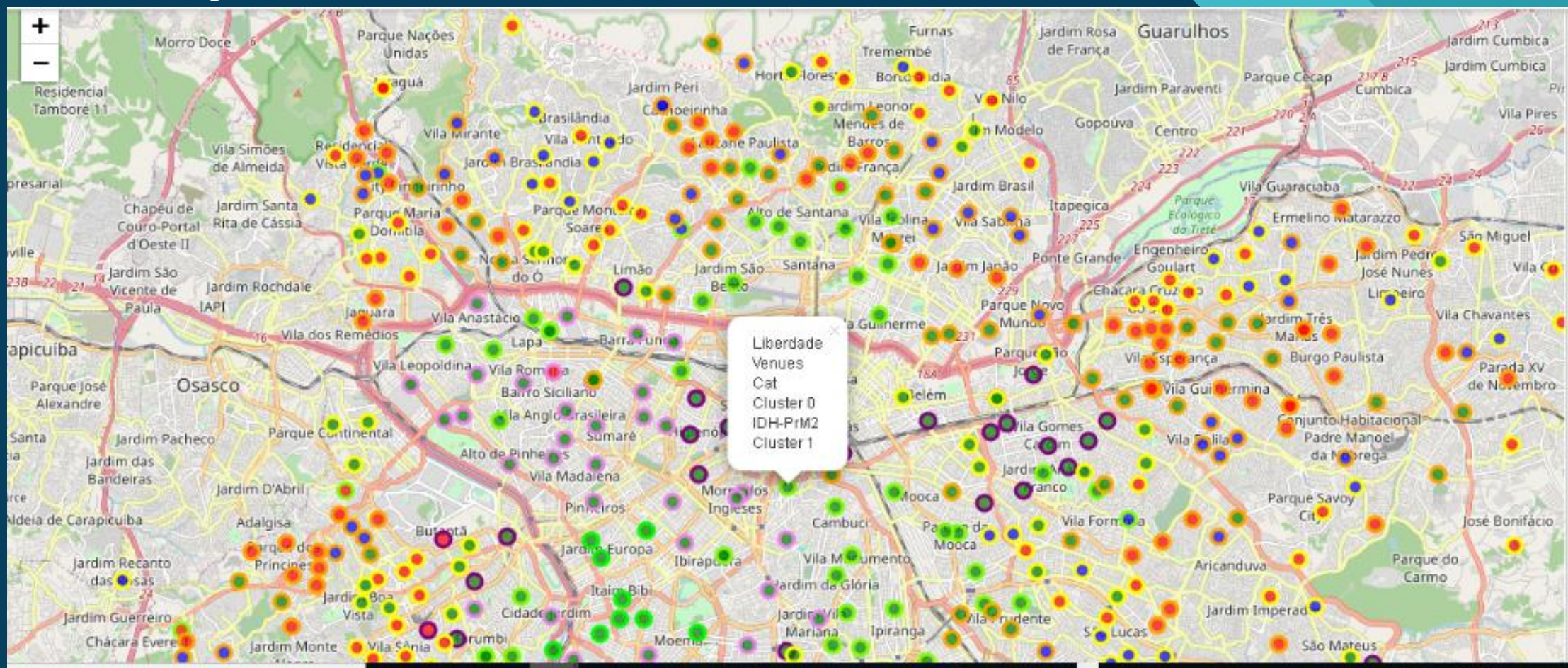
- By the socio-economic indicators (Human Development Indicator and house price by m2)
 - Verification of the optimum k hyper parameter by the elbow method of the library yellowbrick.cluster



➤ Repartition in 6 clusters.

Representation on a Sao Paulo city map of the two clusters of each studied neighborhood

Using a circle color different by label of the IDH/priceM2 cluster and a filling color different by label of the TOP 10 venue categories cluster:



Descriptions inputted manually for each Cluster groups:

- The Cluster Groups are the resulting of the cartesian product of the labels of the two clusters)

Cluster By TOP Venues Categ.	Cluster By IDH-M2Price	DescLocal	DescSocEco	DescVenCat
0	0	Internal Perifery and Center	Without IDH and low price/m2	Restaurant (principally Brazilian)
	1	Internal Periphery	Medium IDH and price/m2	Restaurant (principally Brazilian) an
	2	External Periphery	Low IDH and price/m2	Bakery and Pizzeria
	3	Center West	Very High IDH and price/m2	Restaurant (principally Italian)
	4	West North-Center-South	Very High IDH and high price/m2	Restaurant (principally Brazilian) an
	5	Center, East and South	Without IDH and medium price/m2	Restaurant and Bar
1	0	External Periphery	Without IDH and very low price/m2	Bakery
	1		Low IDH and medium price/m2	Pizzeria and Bakery
	2	External Periphery	Low IDH and price/M2	Pizzeria
	5		Without IDH and medium price/m2	Pizzeria
2	0		Without IDH and very low price/m2	Bakery
	2	External Periphery	Very low IDH and low price/m2	Bakery

Results, Discussion and Conclusion

- As was to be expected, given the size and population of the city, Sao Paulo presents a wide variety of neighborhood.
- At most of the administrative divisions of these neighborhoods, it has been shown:

- Three clusters of these neighborhoods considering their principal venue categories.

The principals venue categories are restaurant, pizzeria and bakery.

- Six clusters of these neighborhoods considering there IDH and house price by m2.

The neighborhood cluster group with the better socio-economics indicators is formed of 9 neighborhoods and is located in the south east of the extended center of Sao Paulo (label 0 for the TOP 10 venue categories cluster and 3 for the IDH-price/m2 cluster).

- The study could have been more precise, if we has been able to access data from all of the neighborhood of Sao Paulo (more than 1700).
- It would have been interesting to use more socio-economics indicators of these neighborhoods, if the Internet site with the las referencing had been on-line
- This study allows having a better vision of the disparities between the thousands of neighborhood of the megalopolis Sao Paulo.
- It should give insights for the businessman to identify the best local to open a new business.



Thank you