

PREDICTING THE APPROPRIATE LOCATIONS FOR A STATIONERY BUSINESS



PREDICTING PLACES FOR THE IMPLEMENTATION OF A STATIONERY

In a business model, it is necessary to evaluate some areas that interfere with the implementation:

- Competition - check the area to see if there are many businesses in the same industry there
- Opportunity - who will be like people and institutions to buy with your future business
- Locality - best place to invest in a new establishment
- Good movement of people around the place

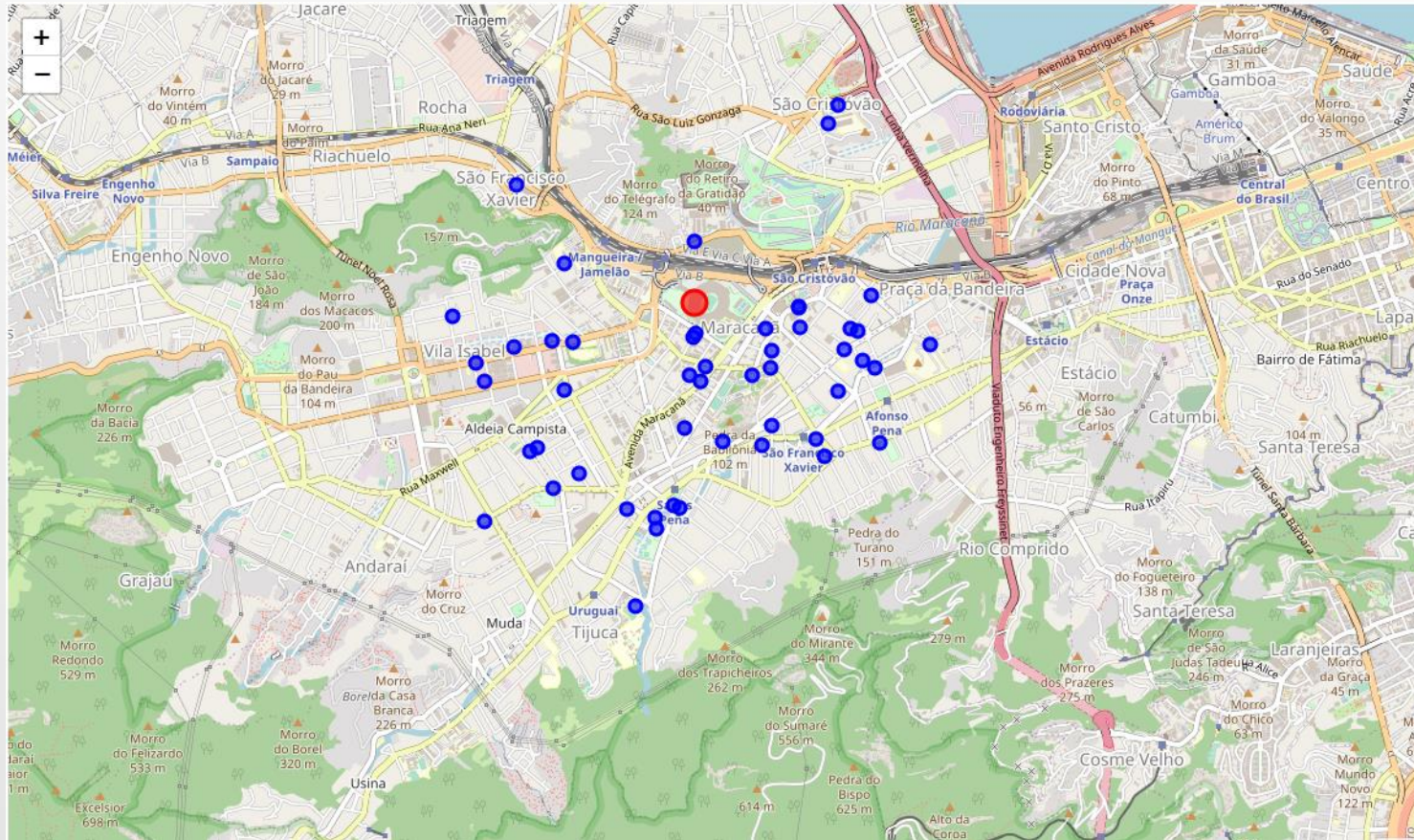
DATA ACQUISITION AND CLEANING

- Number of schools in the neighborhood
- Number and distance of stationers in the neighborhood, if any
- Distance from the neighborhood to the central point located in Maracanã
- The number of schools, names, their type and location in each neighborhood will be obtained using the **Foursquare API**
- The number of stationers, names, their type and location in each neighborhood will also be obtained using the **Foursquare API**

DATA FRAME OF SCHOOLS CLOSE TO THE CENTRAL POINT

	name	categories	address	lat	lng	labeledLatLngs	distance	formattedAddress	id
0	Colégio Militar do Rio de Janeiro (CMRJ)	School	R. S. Francisco Xavier, 267	-22.916797	-43.227150	[{'label': 'display', 'lat': -22.9167972204570...	660	[R. S. Francisco Xavier, 267, Rio de Janeiro, ...	4dc1528422713750ba79c2ad
1	Colégio Batista	None	Rua Visconde de Itamarati, 75	-22.916780	-43.231521	[{'label': 'display', 'lat': -22.9167803740600...	515	[Rua Visconde de Itamarati, 75 (Maracanã), Rio...	4dbfe92c4b2221ec2d640bac
2	Colégio Ressurreição	School	Rua Oto de Alencar, 23	-22.915196	-43.225821	[{'label': 'display', 'lat': -22.9151963527452...	644	[Rua Oto de Alencar, 23, Rio de Janeiro, RJ, B...	512b53c1e4b0fd3b1159b204
3	Colégio Pedro II	School	R. S. Francisco Xavier, 204/208	-22.916347	-43.225898	[{'label': 'display', 'lat': -22.9163471631107...	714	[R. S. Francisco Xavier, 204/208 (Campus Tijuca...	5071a03ce4b0e1b65ea15489
4	Colégio Colégio Nossa Senhora de Lourdes (NSL)	Private School	R. Oito de Dezembro, 328	-22.909630	-43.240261	[{'label': 'display', 'lat': -22.9096297441251...	973	[R. Oito de Dezembro, 328, Rio de Janeiro, RJ,...	4ea5556577c8d0ce5f233bbb

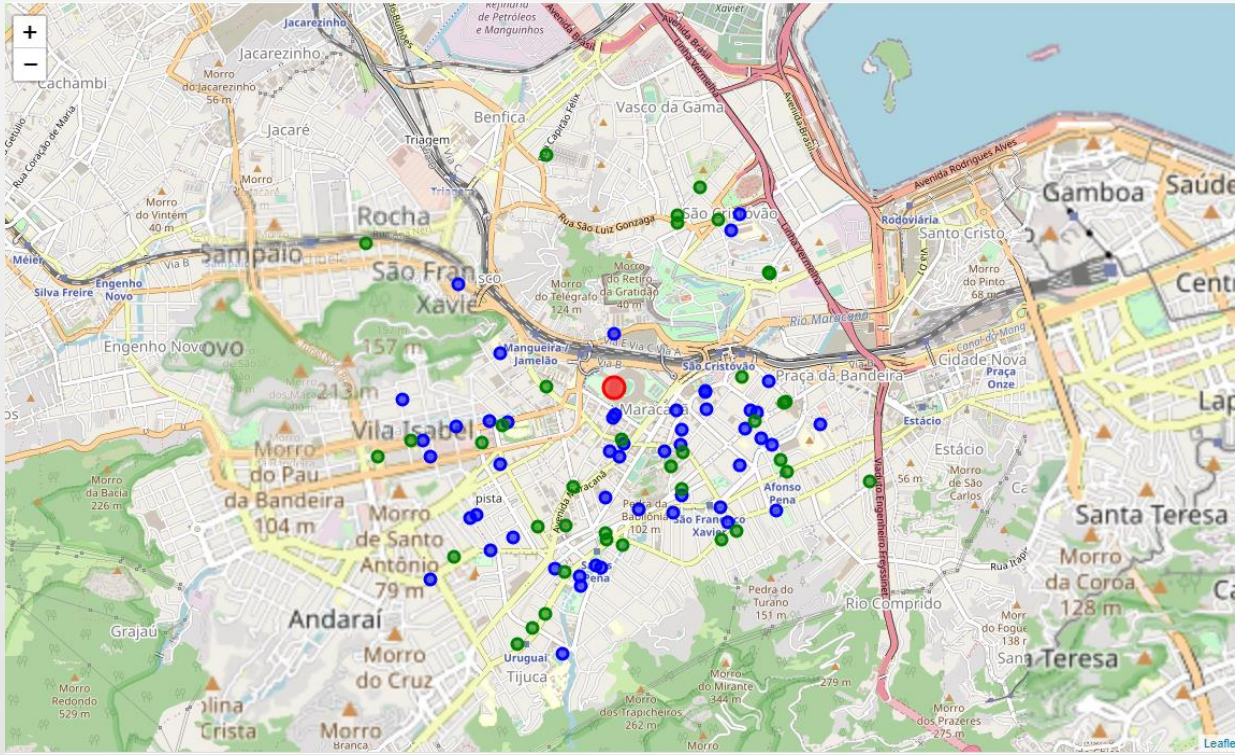
POINTS WHERE COLLEGES ARE LOCATED ON THE MAP



DATA FRAME OF STATIONERS NEAR SCHOOLS IN THE REGION

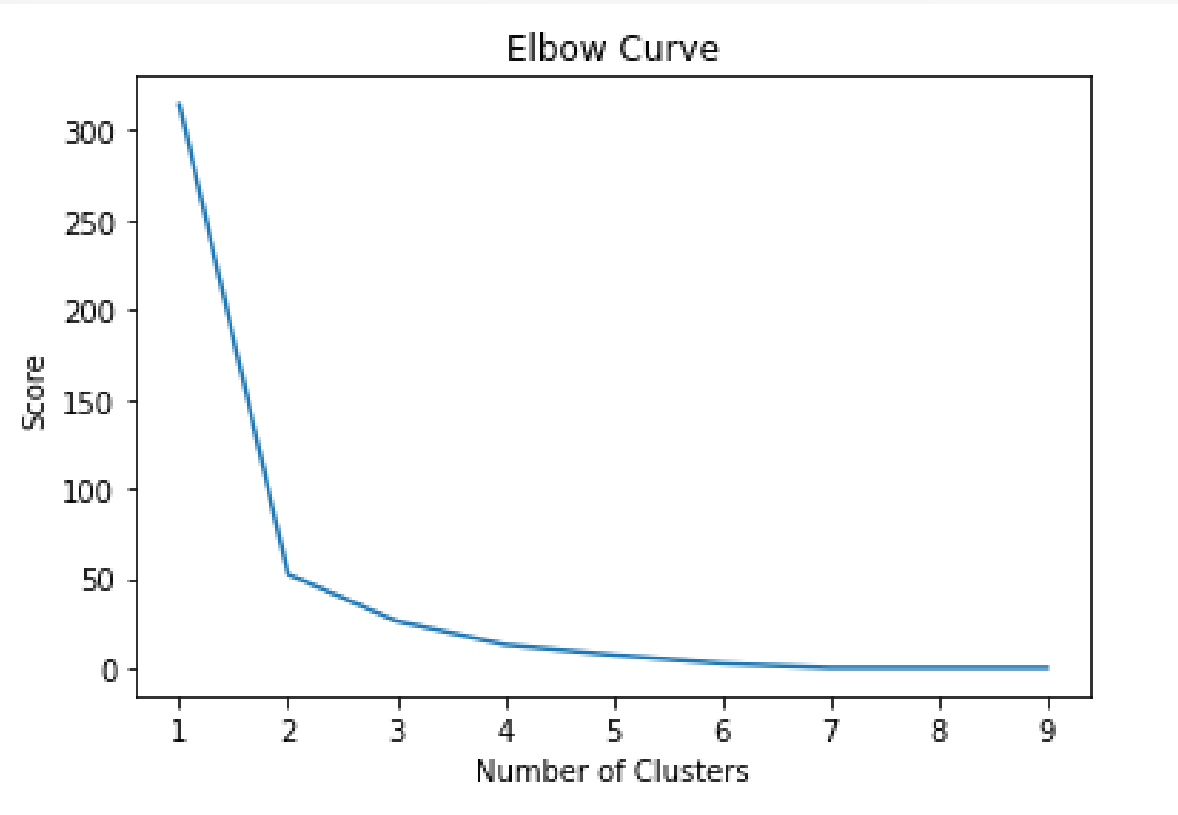
	name	categories	address	lat	lng	labeledLatLngs	distance	formattedAddress	id
0	Papelaria Porto Seguro	Paper / Office Supplies Store	Rua São Francisco Xavier	-22.915969	-43.230628	[{'label': 'display', 'lat': -22.9159694730426...	427	[Rua São Francisco Xavier, Rio de Janeiro, RJ,...	4e946c658b81ef41a94b7377
1	Papelaria Yunes	Bookstore	NaN	-22.912098	-43.236651	[{'label': 'display', 'lat': -22.9120980467721...	561	[Brasil]	4f6258a3e4b0832f00d7d3cd
2	papelaria raposo	Locksmith	NaN	-22.919445	-43.234497	[{'label': 'display', 'lat': -22.9194454163645...	879	[Brasil]	53c547c1498ef1a0284198aa
3	Papelaria Papel Moderno	Stationery Store	R. Sto. Afonso, 101	-22.922829	-43.231915	[{'label': 'display', 'lat': -22.9228290454082...	1189	[R. Sto. Afonso, 101, Rio de Janeiro, RJ, Brasil]	549da21b498e520a4e30b6b6
4	Papelaria Risco e Rabisco	Paper / Office Supplies Store	R. Mariz e Barros, 653 A	-22.914617	-43.219935	[{'label': 'display', 'lat': -22.9146166683652...	1184	[R. Mariz e Barros, 653 A, Rio de Janeiro, RJ,...	4ea69f80cc219396a47722c6

DISTRIBUTION OF SCHOOLS AND STATIONERY STORES AFTER THE ADDITION OF TWO DATA FRAMES



- First step was to join the two data frames for the analysis;
- The second step is to evaluate the data samples and their proportions on the graph;
- Green points - Stationery in the area;
- Blue points – Schools.

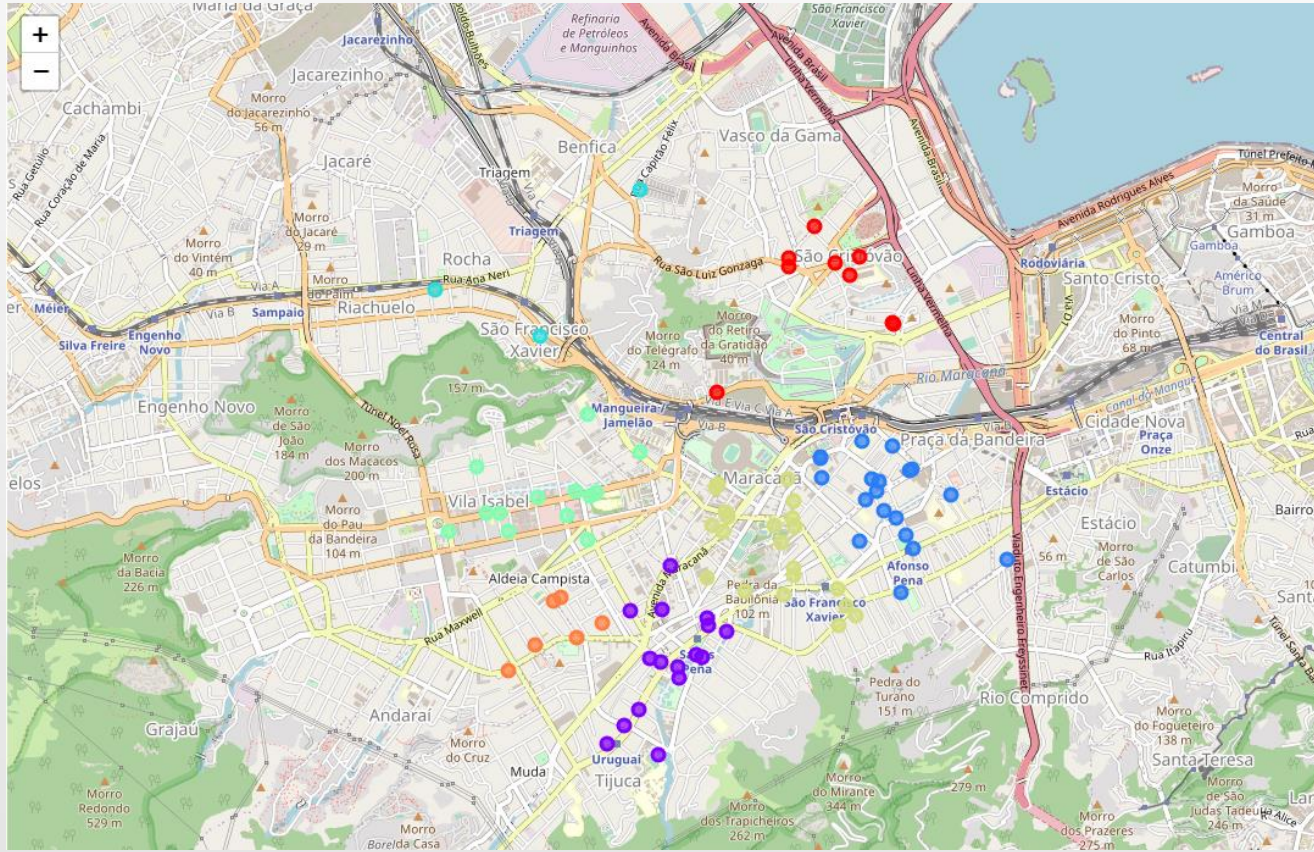
ANALYSIS OF K-MEANS CLUSTERS



The elbow method tries to solve the problem of the best choice of the 'k' argument.

So, in the problem we identified a "k = 3" being the best performance of the algorithm, so from it we can start the case study with K-Means

GRAPHICAL VISUALIZATION OF K-MEANS CLUSTERS



- With $k = 7$ our graphical distribution occurred satisfactorily with precision over the grouped points.
- We can see that the points away from the other clusters were left in new clusters.

NUMERICAL DISTRIBUTION OF SCHOOL CLUSTERS BY STATIONERY STORES

Cluster	Labels	id_global
0	0	9
	1	11
1	0	4
	1	17
2	0	5
	1	7
3	0	6
	1	3
4	0	10
	1	6
5	0	1
	1	5
6	0	2
	1	1

The model used has large numbers of stationery stores near some schools, as well as distractions with few stationery stores and many schools.

CONCLUSION AND FUTURE DIRECTIONS

- The number of schools is higher in the regions of Maracanã, São Cristovão and Vila Isabel.
- In these three areas there are few stationery stores for many schools
- In other locations on the map, many stationers are grouped
- Ideas include:
 1. Include in the analysis data from other establishments that use the services of the proposed business.