The best neighborhood to your restaurant in São Paulo

CONTEXT

Main topics about São Paulo

São Paulo is a metropolis of many faces

15 thousand restaurants

52 countries culinary

12 million inhabitants


AUDIENCE
Stakeholders and interested parts

- ✓ Business owners who want to expand, defend or maintain position

- ✓ Entrepreneurs who want to make a market entry

- ✓ Employees in order to select the best places to work

The most interested parts

Explore, monitor and understand the market is a concern that suits very well many situations.

To do so, is essencial to identify the best neighborhoods to start a new restaurant and help this public.

BUSINESS PROBLEM

The chalenges in the restaurant market

What makes a restaurant successful

- Strategic locations: comfortable surrounding, next to correlated businesses, easy accessibility are primor factors to open the restaurant's to the
- Good suppliers: fresh and high quality foods in time
- Well elaborated menu: it is essencial that the menu fits the locals
- High service level attendance

High level competition

- In this São Paulos's cenario, we see a very consolidated and diverse restaurant's market that attracks a large public. Thus, to start a new restaurant in the capital it is essencial a well planned market entry

Solution

- One factor that we can trace and analyse with robust machine learning resources is the geolocation.
- We can analyse the current market with machine learning techniques and deliver the best spots and types of cusine to start a new business.

SUCCESS CRITERIA

Data modeling success criteria

- Identify the neighborhood with the highest lack of restaurants inside the cluster of restaurants(the one that restaurant is the predominant business and consequently it has the best correlation and sinergy with betewen restaurants and other businesses)

- Select the best number of cluster with elbow tests.
- Highest accuracy model: select the best training sample by selecting the right sets of cluster for each linear regression tested
- Select the neighborhood with the highest lack of restaurants

DATA EXTRATCION

1)The neighborhood

2)The geospacial coordinates

3)The business located in each venues of these neighborhoods

4)The frequency of occurrences of each business in each neighborhood

The data

We will extract, clean and prepare the data to be analysed. The main features or informations are

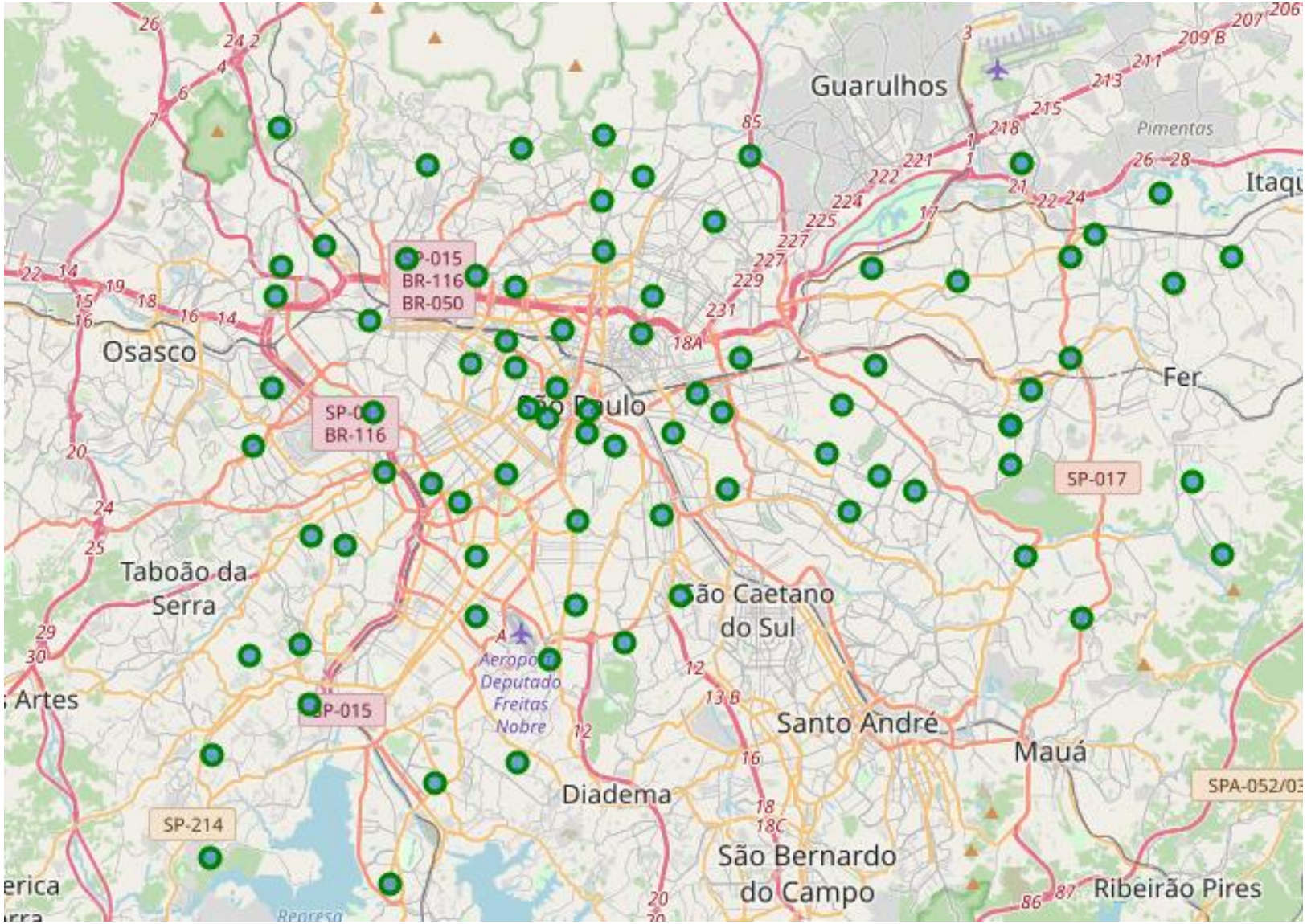
1)Collect the neighborhoods from São Paulo and the geospatial coordinates. Source: https://www.estadosecapitaisdobrasil.com/listas/lista-dos-bairros-de-sao-paulo/

For this, I manually downloaded and organized in a excel file.

The total number of official neighborhoods is 90.

2)The googlemaps api allow us to extract the geospacial coordinates from the neighborhood name. In order to extract, I created a imported the googlemaps library and I run a loop:

	FullAddress	long	lat
0	Água Rasa, São Paulo	-46.5819	-23.5532
1	Alto de Pinheiros, São Paulo	-46.7096	-23.5533
2	Anhanguera, São Paulo	-46.7279	-23.4976
3	Aricanduva, São Paulo	-46.511	-23.5795
4	Artur Alvim, São Paulo	-46.469	-23.546



Covered area

All the point are the neighborhood centers

Places arround - Foursquare API

For each neighbordhood center, in a 1000 meters radius, we have tha proeminets venues

Água Rasa, São Paulo	-23.553209	-46.581890	Padaria Santa Branca	-23.553953	-46.583706	Bakery
Água Rasa, São Paulo	-23.553209	-46.581890	Bona's Carnes	-23.552434	-46.583091	Restaurant
Água Rasa, São Paulo	-23.553209	-46.581890	Temaki Station	-23.553987	-46.583660	Restaurant
Água Rasa, São Paulo	-23.553209	-46.581890	Padaria Carillo	-23.553214	-46.578554	Bakery

It was collected **5504** places

Ocurrances ranks

Places ocurrances rank

Restaurant	1399
Bakery	296
Bar	188
Gym / Fitness Center	175
Burger Joint	151
Gym	136
Pharmacy	124
Dessert Shop	119

Restaurants ocurrances rank by neighborhood

Itaim Bibi, São Paulo	Restaurant	47
Jardim Paulista, São Paulo	Restaurant	41
Santa Cecília, São Paulo	Restaurant	39
Iguatemi, São Paulo	Restaurant	36
Liberdade, São Paulo	Restaurant	33
Pari, São Paulo	Restaurant	32
Santana, São Paulo	Restaurant	32
República, São Paulo	Restaurant	32
Sé, São Paulo	Restaurant	30
Saúde, São Paulo	Restaurant	30
Campo Belo, São Paulo	Restaurant	30
Consolação, São Paulo	Restaurant	30

MODELING

1)Run a clustering model

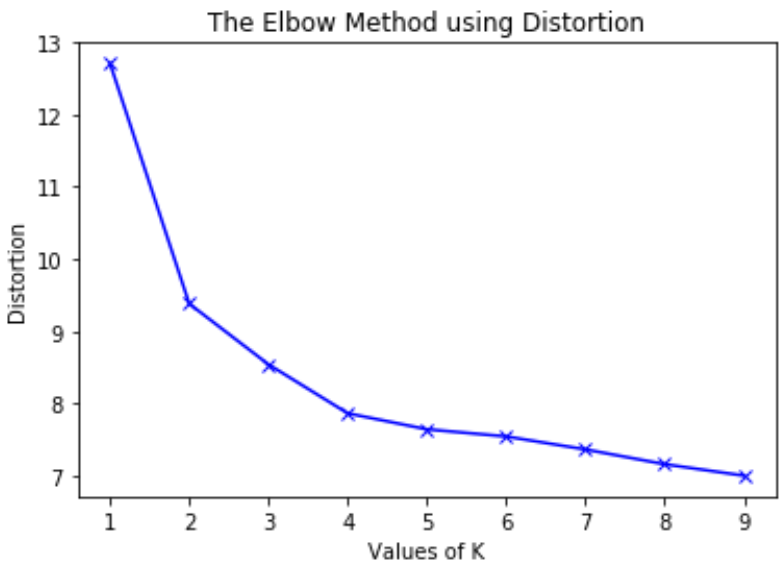
2)Select the clusters with more restaurants

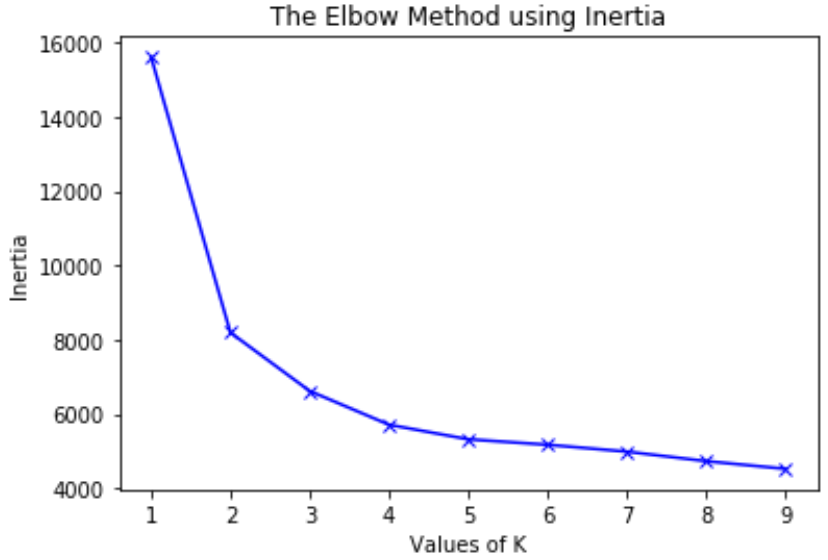
3)Train and test a linear regression for restaurant's ocurrances x other places for each cluster alone and together. Select the best method by mean error.

4)With the winner method, run a train the linear regression without a target neighborhood and project the number of restaurants of that neighborhood. The difference between the projection and the real number represents the lack of restaurants in that specific neighborhood.

The modeling process

In São Paulo we have common venues. To understand and perform better correlations, I created the following method.





Best K=4

Best number of cluster

In order to cluster the model, I used the K-means method for 10 Ks. The best number os clusters was 4, according to the Elbow and Inertia Method:

Method A

Cluster	Mean Error
0	2.46%
1	1.94%
2	2.56%
3	4.28%
Total mean	2.81%

Method B

Cluster	Mean Error
0,1,2,3	5.26%

The best method is **A**

04/04/2020

Linear regression

Method A

Train and test the restaurant projection for each cluster alone

Method B

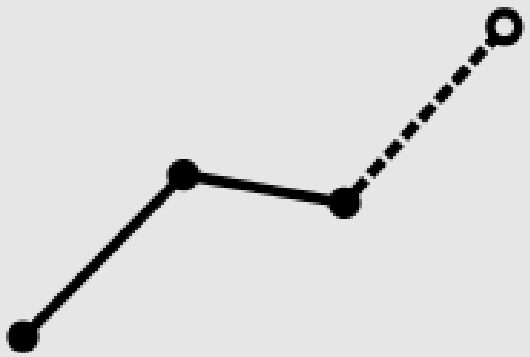
Train and test the restaurant projection for all clusters together

Independent variables:

Ocurrances of each category, except restaurants: bars, gyms, bakerys, etc…

Dependent variable:

Ocurrance of restaurant



Restaurants projections Model

1)Remove the target neighborhood from the cluster sample

2)Train the linear regression with restaurants(dependent variable) and other places(independent variable)

3)Predict the number of restaurants for the target cluster with its depend variable

4)The difference between the predicton and the real number of restaurants is the lack of restaurants in the target neighborhood.

5)This results is stored and the algorithm moves on to the next neighborhood.

Neighborhood	Actual of Number Restaurants	Predicted Number of Restaurants	Lack of Restaurants
Alto de Pinheiros, São Paulo	3.0	14.483317	11.483317
Moema, São Paulo	17.0	24.306804	7.306804
Consolação, São Paulo	30.0	37.252818	7.252818
Casa Verde, São Paulo	18.0	23.175517	5.175517
Brasilândia, São Paulo	4.0	8.894292	4.894292
Saúde, São Paulo	30.0	34.626458	4.626458
Perdizes, São Paulo	20.0	24.390057	4.390057
Cidade Ademar, São Paulo	9.0	13.053998	4.053998
República, São Paulo	32.0	35.483911	3.483911
Capão Redondo, São Paulo	10.0	13.033215	3.033215

The neighborhood with highest lack of restaurants is **Alto de Pinheiros** with a lack of **11.48**

04/04/2020

The result

The loop output for each regression is a lack of restaurants rank.

The diference betwen the projection and the real number is the lack of restaurants.

The tendency

We can see there is tendecy of lack of restauraunt **towards the west from south and north**, where Alto dos Pinheiros is located

CONCLUSION

- The recomendation is that the best place to start or maintain a restaurant is Alto de Pinheiros in termos of competition.
- This neigborhood has others businesses correlated to restaurants that indicates by linear regression a lack of 11.48 restaurants.