The Battle of the Neighborhoods

Where open a Pizza places in São Paulo - Brazil

Introduction: Business Problem

Currently, in the consumption of pizza, the city of São Paulo only loses to New York, where it competes fiercely for the title of world capital of pizza.

This project aims to find the best place to open a **Pizza Place** in the city of **São Paulo**.

Data - Acquisition

Based on definition of our problem, factors that will influence our decission are:

- number of existing pizza centers in the neighborhood
- number of inhabitants in the neighborhood
- pizzeria density in determined clusters

Following data sources will be needed to extract/generate the required information:

- will be necessary to create a database containing the neighborhoods, their geolocation and habitation data using sources of **IBGE institute**.
- number of pizza places and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of São Paulo center will be obtained using **Google Maps**

Data - Acquisition

| DISTRICT	LATITUDE	LONGITUDE	POPULATION
SE	-23.550	-46.634	23.651
TUCURUVI	-23.474	-46.611	98.438
PINHEIROS	-23.562	-46.695	65.364
MOOCA	-23.564	-46.597	75.724
ITAIM BIBI	-23.585	-46.672	92.570
VILA MARIANA	-23.595	-46.632	113.463
BRAS	-23.539	-46.602	29.265
AGUA BRANCA	-23.527	-46.673	84.963
CAMPO BELO	-23.622	-46.679	65.752
ALTO DA LAPA	-23.527	-46.708	65.739
INDIANOPOLIS	-23.610	-46.660	130.780
JARDIM PAULISTA	-23.572	-46.657	88.692
CERQUEIRA CESAR	-23.563	-46.666	57.365
ACLIMACAO	-23.566	-46.632	36.948
SANTA IFIGENIA	-23.536	-46.638	14.383
PERDIZES	-23.544	-46.676	111.161
CASA VERDE	-23.501	-46.650	85.624
SANTANA	-23.500	-46.624	118.797
VILA MARIA	-23.511	-46.593	130.484
PARI	-23.525	-46.615	17.299
VILA LEONOR	-23.500	-46.596	54.331
VILA BANDEIRANTE	-23.496	-46.662	85.624
IPIRANGA	-23.582	-46.601	106.865
BROOKLIN	-23.603	-46.690	69.460

About the Data:

- 24 Districts

- ~1,8Mi of people lives in this districts

- More then 25.000 venues colected

- 94 pizzerias

Neighborhoods and correlations

Correlation from Pizza places and population of neighborhoods:

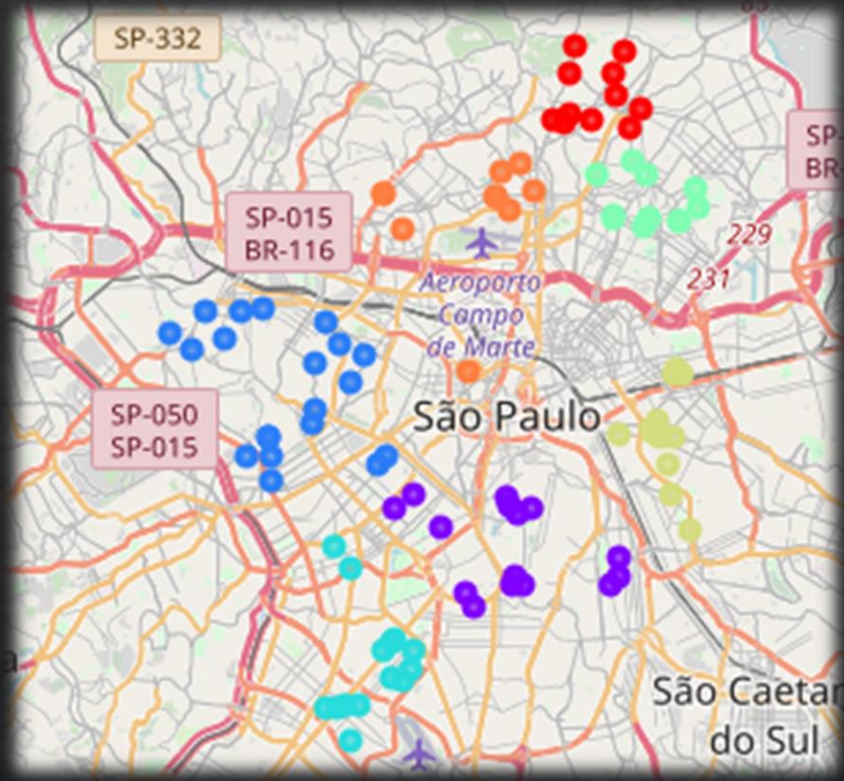Venues Colected from Foursquare:
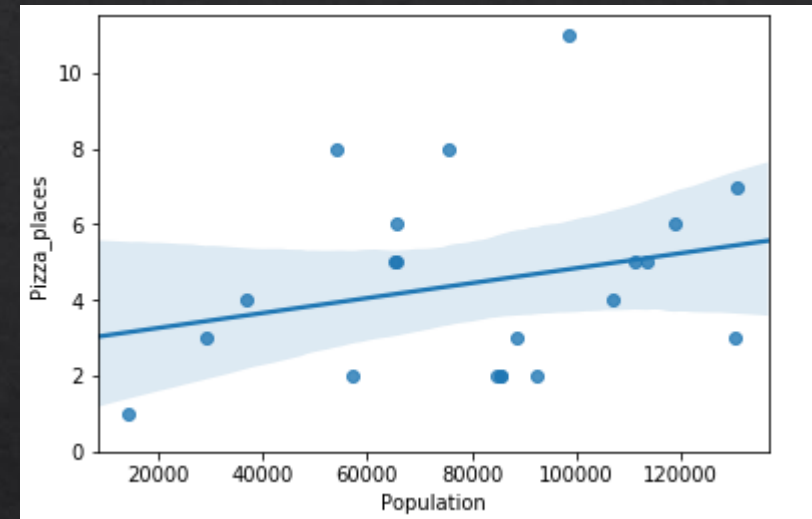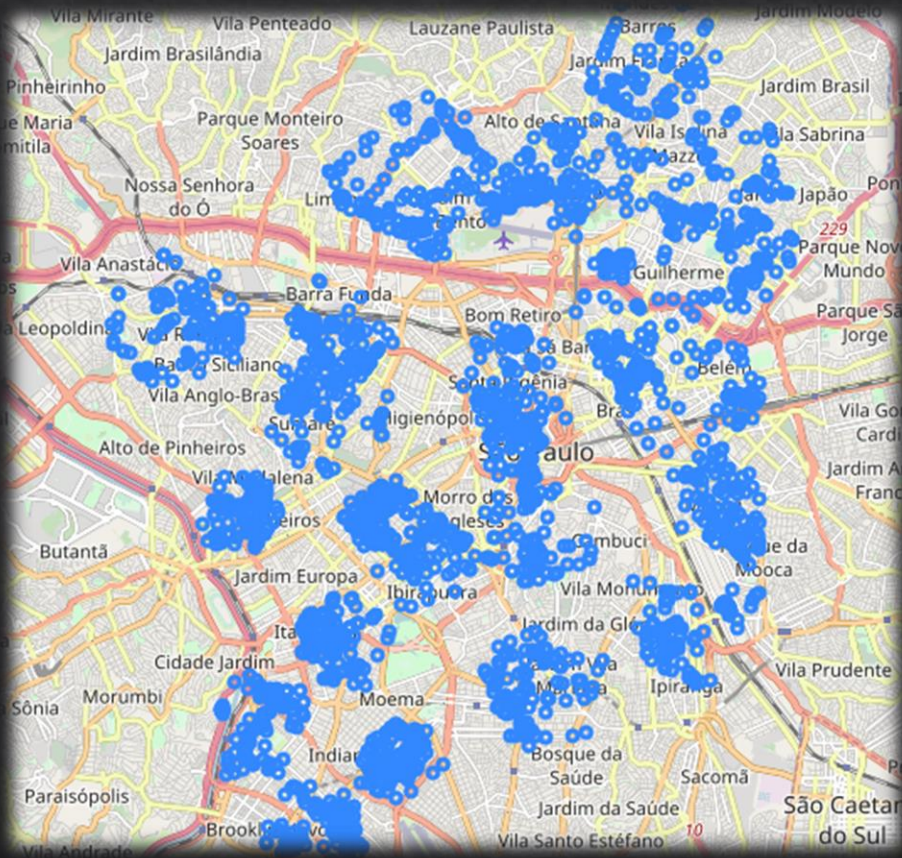
Small correlation between the variables, but we have oportunities when the population is large and number of pizzerias is low

Analysis – K-Means

Pizzerias Grouped whith K-means

Looking at the Clusters, we can see that the ratio of population vs. number of pizzerias is better in cluster 6.

Cluster	Pizza_Places	Population	Pop / Pizza
6	5	185.631	37.126
4	12	303.612	25.301
1	15	345.968	23.065
3	14	289.102	20.650
2	20	384.592	19.230
5	11	104.989	9.544
0	11	98.438	8.949

Analysis – Best K and Neighborhoods

Cluster	Pizza_Places	Population	Pop / Pizza
6	5	185.631	37.126
4	12	303.612	25.301
1	15	345.968	23.065
3	14	289.102	20.650
2	20	384.592	19.230
5	11	104.989	9.544
0	11	98.438	8.949

Santa Ifigênia
Casa Verde
Santana
Vila Bandeirantes

Analysis – Best zone
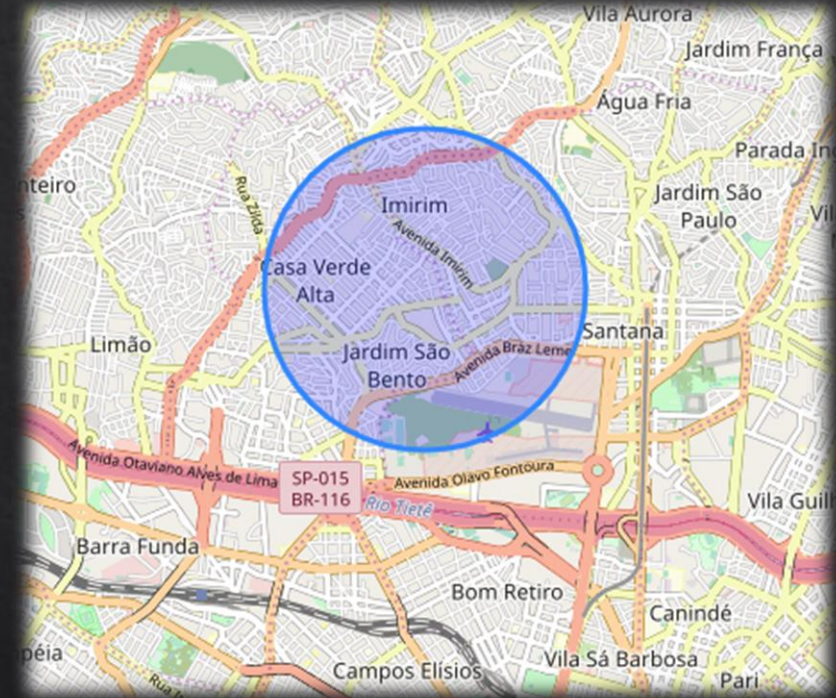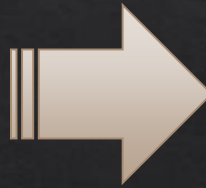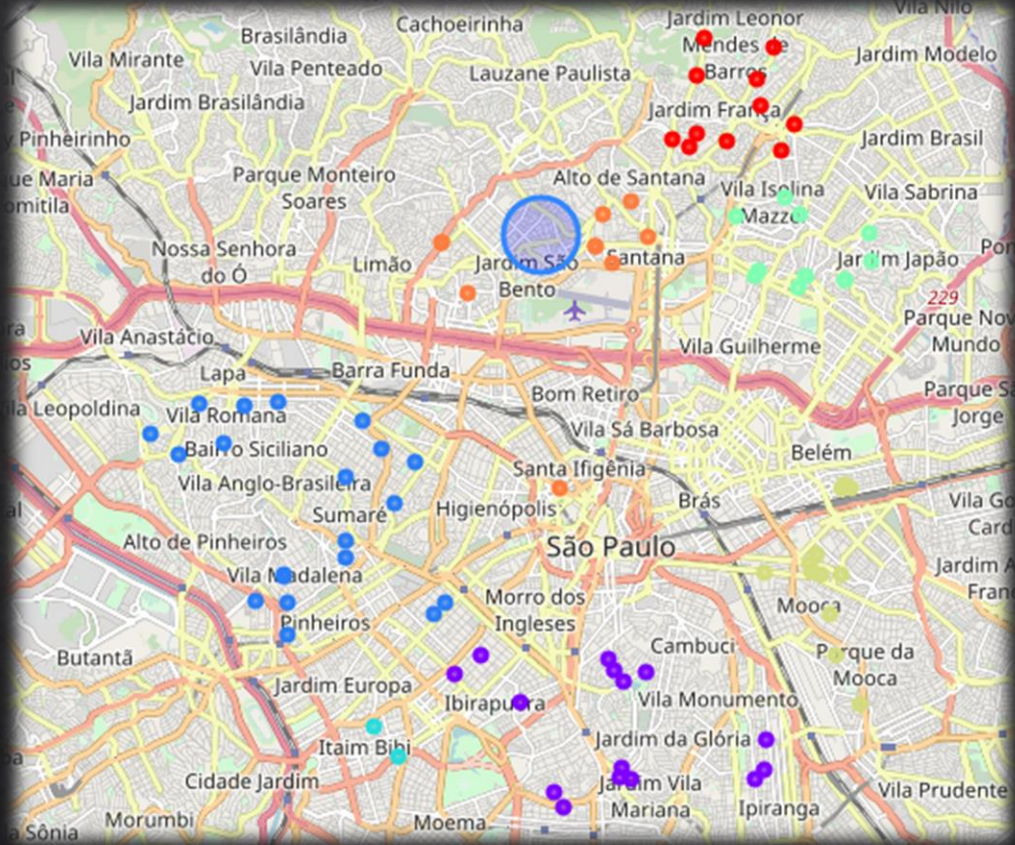
Conclusion

In our analysis we raised the pizzerias in the main neighborhood around the center of Sao Paulo. There are over **90 pizzerias in 20 neighborhoods**.

We can observe that we have some places with very few restaurants or pizzerias, but also those areas with low population, because they are commercial center.

The focus of this project was to find gaps between clusters where we can have opportunities to install a pizzeria and reach the largest possible

At the beginning of the analysis we can see that neighborhoods like Itaim Bibi would be an opportunity. But analyzing graphically on a map we can see that we have other regions with greater potential

From the moment the cluesters were created and then estimated the population inserted in each one, it was possible to create the necessary analyzes to then **suggest a zone**, which will be the starting point for the stakeholders to check more carefully in which exact location open the pizzeria.