Capstone project:

Finding the safest neighborhood in Sao Paulo, Brazil.

Finding the safest neighborhood in Sao Paulo, Brazil.

Problem Introduction

- The focus of this project is to find a location for opening a safe comercial store in the city of Sao Paulo, Brazil.
- The report aims to different stakeholders interested in opening any business type.
- The first step will be to choose a safe borough, analysing the crimes and listing the amount of it committed per borough in Sao Paulo.
- Next we will use K-Means to clustered the similar neighborhoods so as to group data based on existing venues which will help in the decision making process.

Finding the safest neighborhood in Sao Paulo, Brazil.

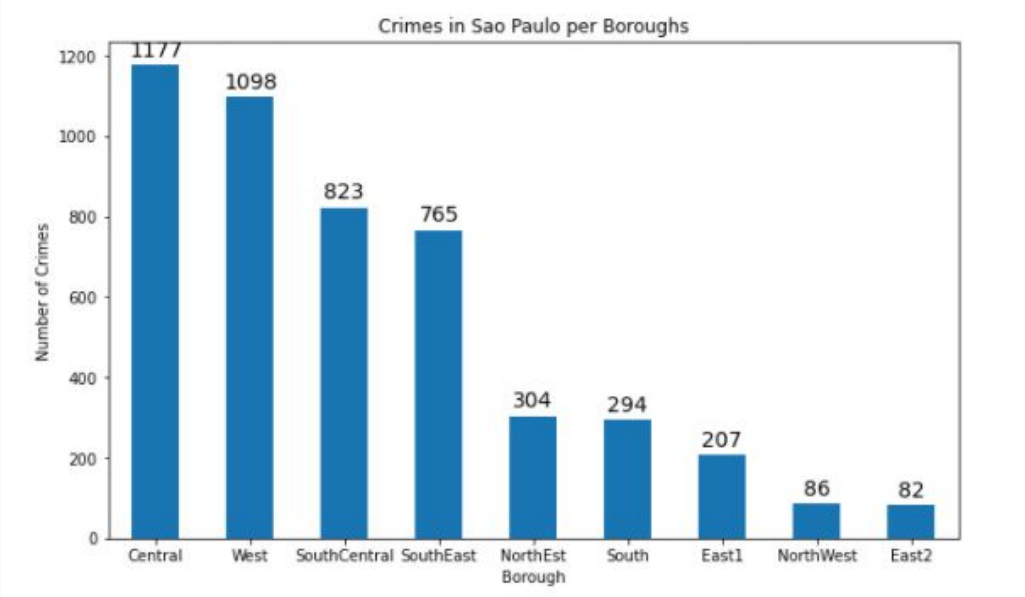
About the data

- Data set from Kaggle containing the Sao Paulo crimes from 2010 to 2018
  - In total 12899 rows and 36 columns
- Additional information of the list of officially categorized boroughs in Sao Paulo from Wikipedia
  - Used to prepare a Dataset with 9 Boroughs and 93 Neighborhoods
- New dataset of the neighborhoods, crime data and the respective neighbourhood cords. This data will be bring us using Geopy to find the safest borough and explore the neighbourhood by plotting it on maps using folium and perform exploratory data analysis

Finding the safest neighborhood in Sao Paulo, Brazil.

Crimes per Boroughs

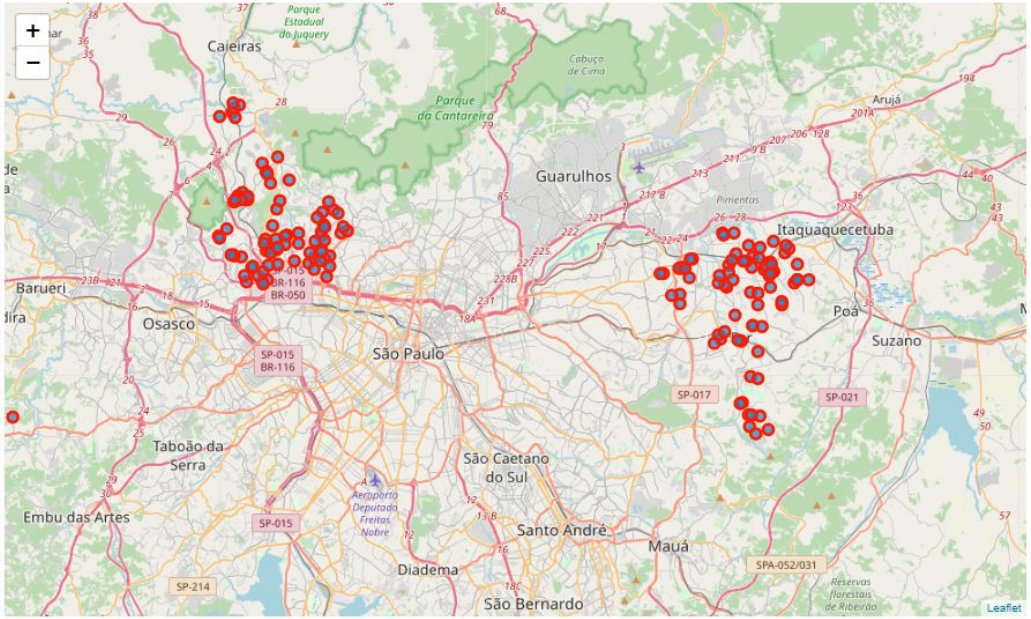
- Central and West have the large sum of crimes in Sao Paulo
- The best boroughs for any type of business in Sao Paulo are North West and East2.
- This two last boroughs will be use to model our K-model cluster



Finding the safest neighborhood in Sao Paulo, Brazil.

Folium to visualize this two boroughs

- We can see Sao Paulo in the center of these two boroughs: NorthWest and East2



Finding the safest neighborhood in Sao Paulo, Brazil.

Venues per neighborhood using FourSquare

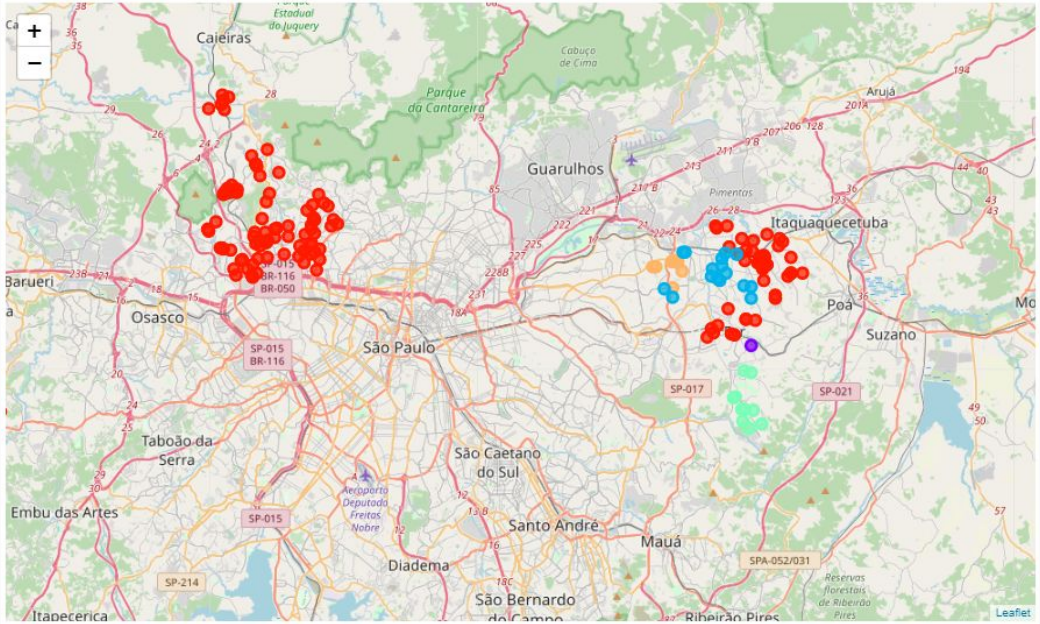
- This show us a head of the list Venues name, category, latitude and longitude informations from Foursquare API.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Brasilândia	Bakery	Pizza Place	Gym / Fitness Center	Convenience Store	Food Truck	Gym	Market	Snack Place	Pharmacy	Grocery Store
1	Cidade Tiradentes	Pizza Place	Bakery	Convenience Store	Gym	Snack Place	Gym / Fitness Center	Grocery Store	Pharmacy	Furniture / Home Store	Comfort Food Restaurant
2	Freguesia do Ó	Bakery	Pizza Place	Pharmacy	Gym / Fitness Center	Bar	Brazilian Restaurant	Pet Store	Athletics & Sports	BBQ Joint	Ice Cream Shop
3	Guaianases	Pizza Place	Farmers Market	Shoe Store	Drugstore	Women's Store	Dog Run	Fish Market	Fireworks Store	Fast Food Restaurant	Farm
4	Itaim Paulista	Pizza Place	Bakery	Japanese Restaurant	Dessert Shop	Fast Food Restaurant	Brazilian Restaurant	Food Truck	Flower Shop	Snack Place	Supermarket

Finding the safest neighborhood in Sao Paulo, Brazil.

Results: Using K-Means and clustering

- We use k = 5 to cluster this two boroughs
- In NorthWest all the neighborhoods were clustered in the first cluster
- In East2 we can see plenty the five clusters of our model



Finding the safest neighborhood in Sao Paulo, Brazil.

Results

- The problem has been achieved by first making use of Sao Paulo crime data to identify a safe borough with variety of neighborhoods for any type of business.
- After selecting the borough it was imperative to choose the right neighborhood where the commercial stores were not among venues in a close proximity to each other
- We made this by grouping the neighborhoods into clusters to assist the stakeholders by providing them with relevant data about venues and safety of a given neighborhood.

Finding the safest neighborhood in Sao Paulo, Brazil.

Conclusions

- We used the crime data to take a look of crimes in all neighborhoods of Sao Paulo, Brazil and later categorized them into different boroughs
- We achieve this picking the two safest boroughs in Sao Paulo and then checking the most common venues for both
- we use K-means to cluster this two boroughs and this way we can take a look for choosing the best neighborhood for any type of commercial business.
- Once we confirmed the borough the number of neighborhoods for consideration also comes down, we further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.