**Capstone Project - The Battle of the Neighborhoods**

Price comparison for hotels in Northeast Brazil

Gael T.Yota

June 9, 2020

## **Introduction: Business Problem**

Two Brazilian cities were ranked in a ranking that honored the destinations that are highest among tourists in 2013. Fortaleza, in Ceará, and João Pessoa, in Paraíba, were among the 54 places in the world that had more positive opinions and increased in interest throughout the year according to TripAdvisor users.Both cities are located in the Northeast region of Brazil.

Fortaleza has 313,140 km ² of area and 2,643,247 inhabitants estimated in 2018, in addition to the greater population density among the capitals of the country, and it is also the largest city of Ceará in population and the fifth in Brazil.

João Pessoa is a Brazilian municipality, capital and main financial and economic center of the state of Paraíba. With a population of 800,323 inhabitants estimated in 2018, the capital of Paraíba is the eighth most populous city in the Northeast Region and the 23rd in Brazil, being in its state the most populous municipality.

As shown in the TripAdvisor ranking these two cities have a very strong tourist potential. Thinking as a tourist, in general, we want to stay in a place close to sights and we also want to pay a fair price for the accommodation.

When we consider this problem, we can create a graphical map informing the values of the hotels and group them near the best places to visit in the cities. In addition, we can also compare the two Brazilian capitals by analyzing the cost of lodging and the tourist sites according to the city guide of Foursquare.

## **Data**

As both capitals have several neighborhoods, I decided to pick up the major neighborhoods to compare hotel prices. For Fortaleza city I selected the following neighborhoods: Praia de Iracema, Meireles, Mucuripe, Praia do Futuro and Downtown.For city of João Pessoa I selected: Manaíra, Tambaú, Cabo Branco, Bessa and Downtown.

I'll also use OpenStreetMap with Nominatim, google maps data to manually collect hotel values and the foursquare API for finding venues in both capitals.

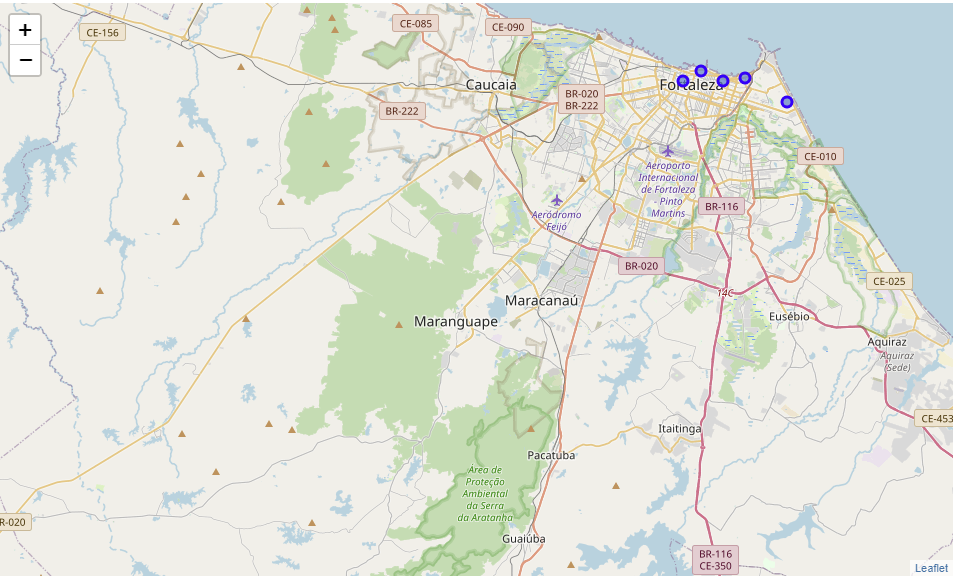


Fig 1. Visualization of the Neighborhoods

Now that we have our location candidates, let's use Foursquare API to get info on venues in each neighborhood. We will check the number of bars, restaurants, theaters, steakhouse, coffee shop and malls in each neighborhood.

## **Methodology**

In this project, we will direct our efforts to find the best place to stay in the cities of João Pessoa and Fortaleza. In the first step, we collect the required hotel location and price data using google map data manually. We also identified the quantity and bars, restaurants, theaters, steakhouse, malls and coffees.

The second step will be the creation of two new features with the average price of hotels and a score of neighborhoods considering the number of locations found through the foursquare API.

The third and last step is to use the K-Means algorithm to perform automatic grouping by similarity, and we will create 3 clusters that will be presented and analyzed next.

## **Results**

Let us now group the neighborhoods by value of lodging and good location based on the places found in foursquare.

These clusters are the end result of our analysis.

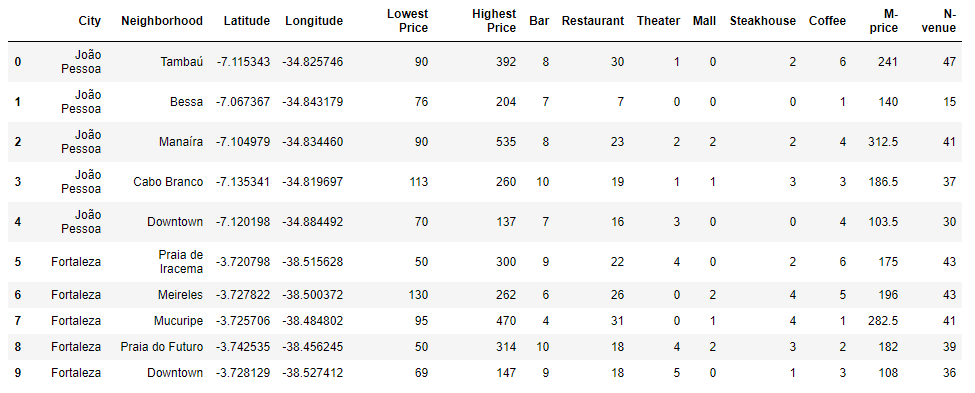


Table 1. Data table result using Foursquare data

Now, using K-Means let’s group these data by clusters. We get clusters as seen on the last column of this table;

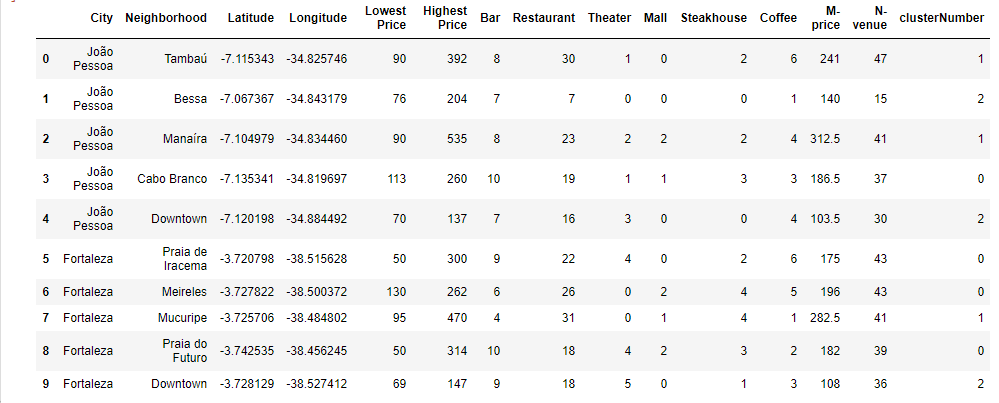


Table 2. Data table result after clustering with K-Means

Now let us see the scatter plot of the classified groups.

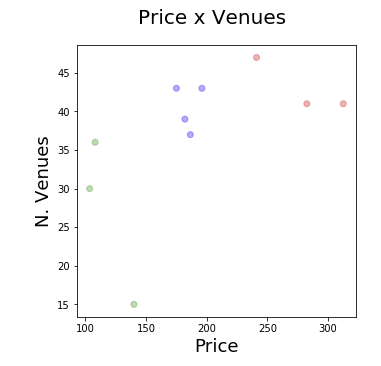


Fig. 2 Data visualization of the clusters

## **Observations and Discussion**

Our analysis shows which local neighborhoods are the best neighborhoods of the two Brazilian Northeast cities. In addition, after the classification we noticed that both cities downtowns had been embedded in the same cluster, demonstrating that the algorithm did a good classification of the data.

## **Conclusion**

This project had as objective to analyze two capitals of the northeast that already were in the ranking of the positive opinions in the tripadvador. For this we analyzed the tourist sites that are available on foursquare and also the values of the hotels shown in google maps.To group the data we used a K-Means classification algorithm that grouped the muds of these cities plotting a graph in which we can visually choose the best destination based on this classification.The strategy used was very simple considering only the number of bars, restaurants, theaters, coofe shop, shopping and steakhouse, in addition to the price of hotels. Future analyzes could be made based on the police occurrences found in the neighborhoods, suggesting a safer location for tourists.