

Healthy Food Restaurant - Report

Brayan Camilo Riveros Castellanos

March 31, 2021

Introduction



1.1 Background

A chain restaurant wants to start a new restaurant of healthy food because in Bogotá city there are many fast restaurants or common restaurants, also the citizens over time people have improved in their diet.

1.2 Problem

In Bogotá, Colombia in the last few years several restaurants of all kinds have been opened, in this case, there is a minimal group that prepares healthy food and the number of people interested in healthy food is also growing, for this reason, the stakeholders will want to know which it's the best place for starting their business.

1.3 Interest

The parts interest are restaurant chains which it will want to open a new restaurant for fit people, people that workout or simply people interested in healthy food.

Data acquisition and cleaning



2.1 Data sources

It will be need to use two databases, the first data contain latitude, longitude for each neighborhood, and the second data contain population and rural territorial extension so on. These data will use looking for venues in Foursquare

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table, then I will delete row columns irrelevant to the objective. There are problems with names or other data, so I should validity it for the neighborhood code and select a specific year (2015) for data amount of population.

2.3 Feature selection

After data cleaning, I will find related places with restaurants, GYM and shopping mall since these places are the most visited for people interesting in healthy food.

Exploratory data analysis



From the data downloaded in bogota-laburbano.opendatasoft.com and library python folium was graphed the neighborhoods (Localidades) of Bogotá with a radius of 3000 meters.

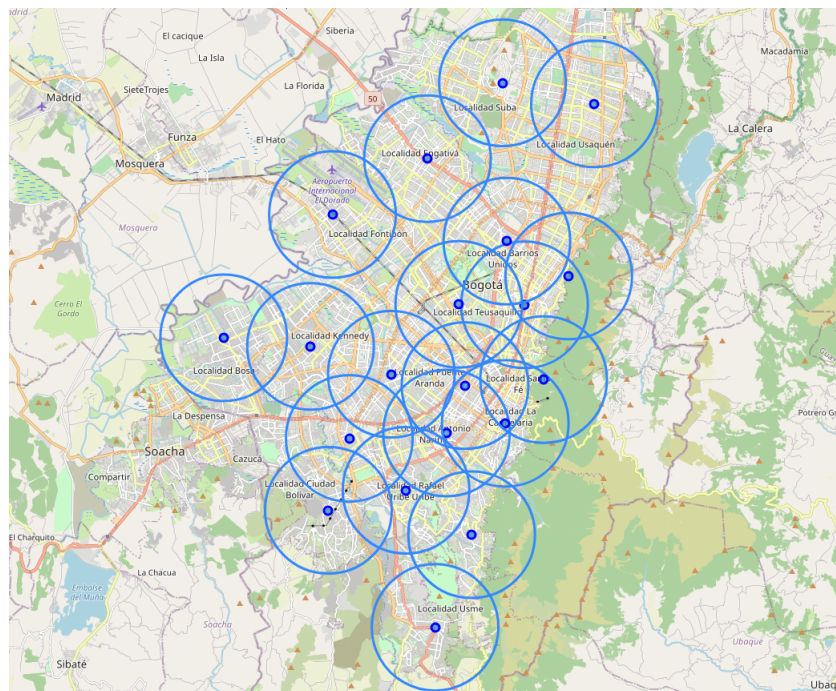


Figure 3.1: Clustering

3.1 Getting venues

The Foursquare API was used to find locations in each neighborhood, but, in this case, we worked with two data sets, first venues such as Restaurants (all types), GYM and Shopping Mall were analyzed because those places are important to take a decision, the other hand, the data all venues were used to analyze the similarity of the neighborhoods.

Neighborhood	Venue	Venue Category
CHAPINERO	Bandido Bistro	French Restaurant
CHAPINERO	Quebrada La Vieja	Scenic Lookout
CHAPINERO	El Caracol Azul	Peruvian Restaurant
CHAPINERO	Harry Sasson	Restaurant
CHAPINERO	Brot Bakery & Cafe	Bakery

In this way, I found the number of restaurants, GYM, Shopping Center and amount population

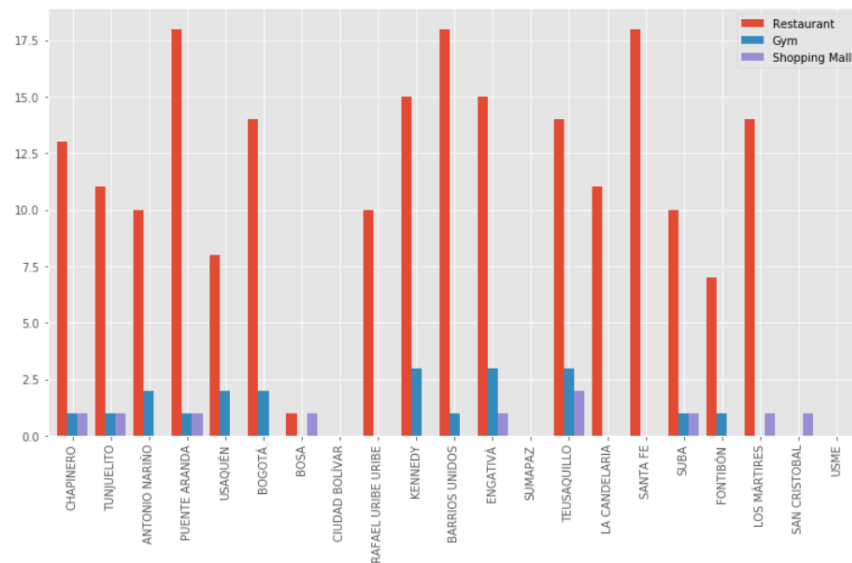


Figure 3.2: Restaurants, GYM and Shopping Center

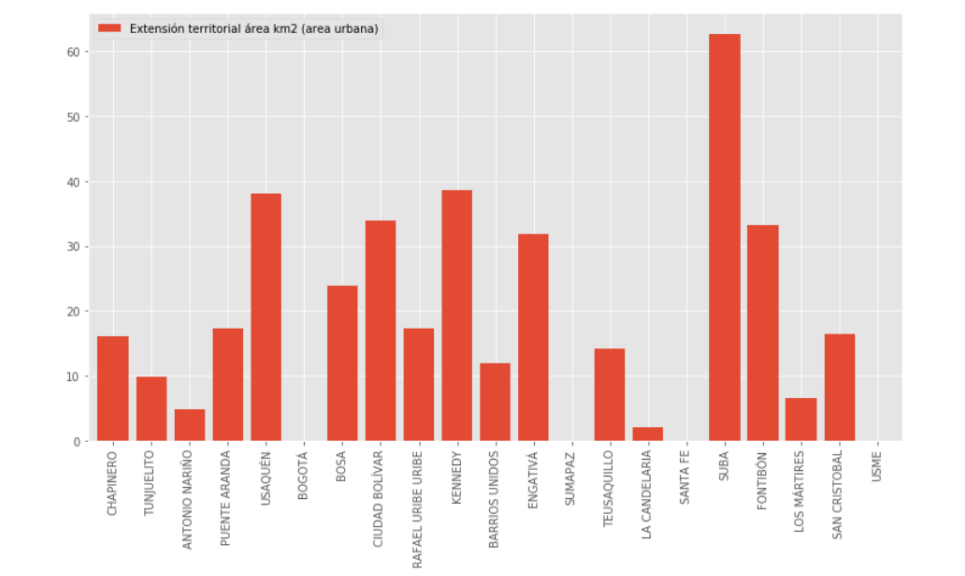


Figure 3.3: Population

And finally, I found the most frequent venues for each neighborhood as shown in the table

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
ANTONIO NARIÑO	Pizza Place	Park	Café	Seafood Restaurant	Fried Chicken Joint	Restaurant	Art Museum
BARRIOS UNIDOS	Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	French Restaurant	Asian Restaurant	BBQ Joint
BOGOTÁ	Coffee Shop	Hotel	Bookstore	Bakery	Italian Restaurant	Vegetarian / Vegan Restaurant	Pizza Place
BOSCA	Grocery Store	Movie Theater	Department Store	Pharmacy	Restaurant	Ice Cream Shop	Park
CHAPINERO	Hotel	Bakery	Coffee Shop	Restaurant	Italian Restaurant	Breakfast Spot	Burger Joint

Figure 3.4: The most frequent venues

Model

The model will be classification, specifically it will be used the algorithm K-means, I make seven clustering with all venues to be able to determine similar neighborhood in Bogotá.

4.1 K-means

With the data collected from Foursquare (all venues) it has been made seven clusters as shown in the image

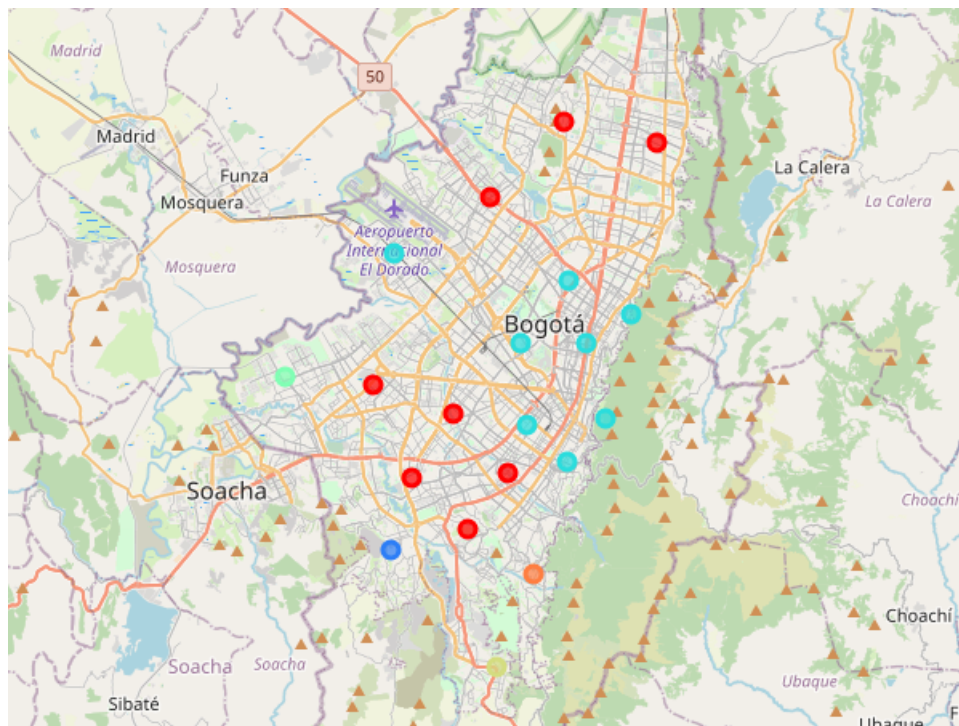


Figure 4.1: Clustering

With this model together with data on population, the number of restaurants (all types), GYM, and Shopping malls, you may be able to choose the best place (neighborhood) to start your healthy food restaurant. So the group that contains some GYM restaurants and the mall was group 1 as shown in the table

Neighborhood	Population	Restaurant	Gym	Shopping Mall
TUNJUELITO	200048.0	11.0	1.0	1.0
ANTONIO NARIÑO	108941.0	10.0	2.0	0.0
PUENTE ARANDA	258414.0	18.0	1.0	1.0
USAQUÉN	494066.0	8.0	2.0	0.0
RAFAEL URIBE URIBE	375107.0	10.0	0.0	0.0
KENNEDY	1069469.0	15.0	3.0	0.0
ENGATIVÁ	874755.0	15.0	3.0	1.0
SUBA	1174736.0	10.0	1.0	1.0

In theory any neighborhood of this set is recommended.

4.2 Conclusion

With previous model and data of population, the number of restaurants (all types), GYM, and Shopping malls seems that the best place for healthy food restaurant is PUENTE ARANDA, because there are a population moderately high and restaurants with high frequency as shown in the list

1. Seafood Restaurant
2. Mexican Restaurant
3. Restaurant
4. Burger Joint
5. BBQ Joint
6. French Restaurant
7. Park
8. Sandwich Place
9. Creperie
10. South American Restaurant

Besides, more 50 percent the most frequent venues are related to food or restaurants and at least there is a GYM a shopping mall in this neighborhood.

4.3 Future directions

It is possible to improve the model with data with more qualitative and quantitative variables of the population and updating of places in Bogota.