THE BATTLE OF NEIGHBORHOODS

IBM Data Science Program — Capstone Project

DECEMBER 29, 2019

JAIRO IVAN ORDOÑEZ

INTRODUCTION

In the last module of the Data Science Program from IBM, we need to demonstrate the learned in the last 8 courses. This document represents the final result of the capstone Project. The business problem chooses for this project is the recommendation to a freelancer to live in an Neighborhood in Cartagena or Bogota (Colombia) across his requirements. Below the business problem description:

A Colombian freelancer is looking for a Neighborhood to live in Cartagena or Bogotá, however, he needs these special facilities:

1. Vegetarian or Vegan Restaurant, because He cannot eat animal protein (meat, fish, chicken, etc.)
2. A gym or Public Park to make physic exercise.
3. Farmer market to buy the vegetables and fruit for his diet.
4. Pharmacy to buy medicament to his allergy

The Colombian ask to help for us to, with cluster models applied in the Data Science Program, we can recommend him the Neighborhoods than accommodate to his requests.

Note: We will not evaluate the security or living price in the Neighborhoods, only the closer venues.

DATA

The data to use is the postal codes to Colombia. This file contains the postal code to every neighborhood from the Cities of Colombia. We need to search these postal codes in the Foursquare API to get the venues in the neighborhoods searched. In this file we have:
1. noid: Is only an id to the row
2. codigo_departamento: state's id (In Colombia the state is called "departamento")
3. nombre_departamento: state's name
4. codigo_municipio: city's id (When the state is the Capital, in this column we have the neighborhood group)
5. nombre_municipio: city's name
6. zona_postal: postal zone
7. codigo_postal: postal Code
8. limite_norte: what is the north limit of the zone. (Urban limit, rural limit, another city, etc.)
9. limite_sur: what is the south limit of the zone.
10. limite_este: what is the east limit of the zone.
11. limit_oeste: what is the west limit of the zone.
12. tipo: kind of city (Urban or Rural)
13. barrios_contenidos_en_el_codigo_postal: Neighborhoods in the postal code (separated by minus "-")

METHODOLOGY

First, we wrangled the data to use only the relevant columns to solve the problem, then we found the neighborhoods for both cities (Bogota and Cartagena) and with Foursquare API we found the nearby venues of the Neighborhoods to cluster by KMeans method.

Based on the examples with Toronto and New York, we found that the cities are very different, and possibly one of both was to qualify better than the other, in part, because Bogota is bigger than Cartagena and it's more diversity only for be the Capital.

Finally we explain the conclusions to the stakeholder and the final recommendations to his requirement.

RESULTS

1. Bogota is the Capital of Colombia, so It is the bigger city from this country.
2. Bogota has many different venues to classify the Neighborhoods, but the most common venue is the Fast Food Restaurants.
3. Cartagena is a coast city, so it is most probably to find sea restaurants.
4. There is not much information to the venues in Cartagena, so the analysis is more basically than Bogota.
5. In Bogota, the five clusters designed is not sufficient to demonstrate an useful analysis in the distribution of the Neighborhood, however is very useful to solve the Business Problem because one of the Clusters accomplish the requirements to the Stakeholder.

DISCUSSION

1. If the customer wants, we can to compare another couple of cities, like Barranquilla – Cartagena and Bogota – Medellin. This allow have less differences between the type of the city and the population.
2. Foursquare API, in Colombia has not very much information about the venues, so is harder to make a complete analysis around the neighborhoods.

CONCLUSION

So, the conclusion to our freelancer Colombian is, if he decides to live in Bogotá:

1. We recommend the "LOCALIDAD SUBA" Neighborhood because its has 3 of the 4 requirements in 2 distinct places.
2. In "LOCALIDAD BOSA" there are two places with Restaurants Vegetarian or Vegan, but is most common the bars, so if he likes the drink and the dance, we recommend this Neighborhood.
3. And the most common venue in most of the Neighborhoods in Bogota is the Fast Food Restaurants.

If he decides to live in Cartagena:

1. In Cartagena there are not Neighborhoods with the special facilities, because it is a coast city and its first product is the fish, so the major of the venues are restaurants but no vegan or vegetarian.
2. In Cartagena there are not many parks so, if he needs do physic exercise, he would have to swim in the beach.