

# **Coursera Capstone**

## **IBM Applied Data Science Capstone**

### **Business Geolocation in Bogota Colombia.**

**By: Kevin Gamarra**

#### **1. Introduction: Business Problem**

Bogotá is the capital of the Republic of Colombia. In 2018 it had an estimated population of 7.4 million people. It is one of the most important cities in Latin America both in number of inhabitants and in GDP. The city has a dynamic economy, according to data from the Chamber of Commerce of Bogotá, 35855 new businesses were created in 2018, although it is estimated that only 30% survived a year after its creation.

Given this scenario, it is very important for people interested in creating new businesses to have as much information as possible about structures and the behavior of the markets in which they want to enter.

#### **Business Problem**

This analysis aims to expand knowledge through statistical tools about the geolocation of the different types of businesses throughout the 20 locations in which the city of Bogota is divided. In this way and depending on the particular analysis of each investor you can decide in which location you can establish your business, this will depend on many factors because some people prefer to take advantage of the agglomeration of companies, while others prefer to establish their businesses in places where There is not much competition among other important factors to consider.

#### **2. Data**

In order to meet the above objectives, we will need the following data:

List of "Localidades" of the Bogota, city.

Latitude and Longitude coordinates of those territorial subdivisions. This information is required to plot the map.

List of all VenueCategories in the "Localidades" in Bogota. This information is required in order to clustering the localidades.

### **Sources of data.**

The list of "Localidades" is in the page <https://es.wikipedia.org/wiki/Bogot%C3%A1>, the table it will be extracted with Pandas' method read\_html.

The Latitude and Longitude coordinates of localidades will be obtained with Python Geocoder Package.

The List of VenueCategories will be obtained from Foursquare API.

### **Methodology**

To achieve the objective of this project, we will now list the VenueCategories generated by Foursquare API within a limit of 100 meters around the specified coordinates with a limit of 100. We will generate a DataFrame in which the name of each Venue is specified, its category, its Latitude and Longitude, then we will count how many were obtained by each of the locations, in addition to finding out how many different VenueCategories returned the query to the API.

The above information will be useful to construct a DataFrame in which it is possible to identify if a type of Venue Category exists in a locality or not. Then a classification of the 5 VenueCategories or types of businesses more frequent in each of the localities will be made and finally, the K-Means algorithm will be used to generate 5 clusters in which the territorial units of Bogota are segmented according to the most common type of business in each of them. To obtain a better understanding of the results obtained, we will generate a map through the Folium package, in which the different clusters are clearly identified.

## 4. Results and Discussion

### Map's "Localidades" of Bogotá

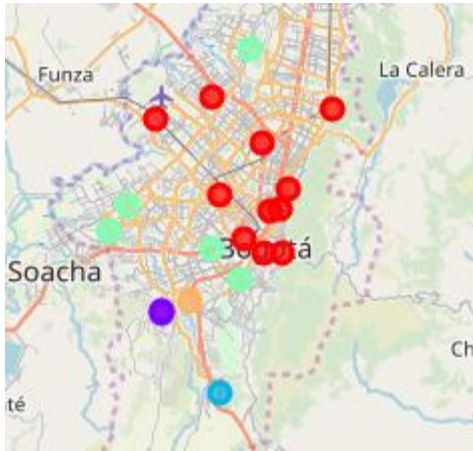
With the list of towns in Bogotá with their respective coordinates, we can draw the map of the city and how it is subdivided territorially.



- Cluster number 1 is made up of Usaquen, Chapinero, Santa Fe, Fontibon, Barrios Unidos, Teusaquillo, Los Martires, Puente Aranda, La Candelaria and Rafael Uribe. The VenuesCategories that characterize this group are Restaurants, Hotels, Cafeterias, Coffe Shops, Bakeries and Hamburgers.
- Cluster number 2 is formed by the towns of Sumapaz and Ciudad Bolivar. The types of places Categories that characterize this group are Parks, Gyms, Burguer Joints and Breakfast Spot.
- Cluster 3 is formed only by the locality of Usme. In this cluster a bus station was obtained as the most representative VenueCategory, then by CoctailBar, Pool, Mountain and DogRun
- Cluster 4 groups the towns of San Cristóbal, Bosa, Kennedy, Suba and Antonio Nariño. The main venues Categories in restaurants, gyms, supermarkets, pizzerias, cinemas and Grocery stores.
- Cluster 5 is formed only by the town of Tunjuelito and the Top 5 VenuesCategories are Park, Construction, Department Store, Fast Food Restaurant and Cajun Creole Restaurant.

- Cluster 2 and 3 do not have much variety of VenuesCategories, it should be noted that the town of Sumapaz located in Cluster 2 is above all a rural area with the lowest population density rate in the city.

#### **Map's Clusters Of Localidades Acoording The Venues Categories of Forsquare**



### **5. Conclusions**

- When deciding in which location to establish a business and depending on whether you want to take advantage of agglomeration economies or if you want to avoid a lot of competition. We have that Restaurants, Hotels and Cafes, are located mainly in cluster 1 and especially in the towns of Usaquén, Chapinero, Teusaquillo and Los Martires.
- Gyms, supermarkets, pizzerias and vegetable stores stand out in cluster 4, mainly in the towns of Bosa, San Cristóbal and Suba.