

# Data Science Project | Where to live in Bogota, Colombia

Mario A. Serrano | July 2020

## 1. INTRODUCTION

### 1.1 THE BACKGROUND



Bogotá is the largest city, and also the capital of Colombia, and according to [worldpopulationreview.com](http://worldpopulationreview.com) has about 11 million people living in the metropolitan area.

Bogotá is located in the center of Colombia, on a high plateau known as the Bogotá savanna, part of the Altiplano Cundiboyacense located in the Eastern Cordillera of the Andes. It is the third-highest capital in South America and in the world after Quito and La Paz, at an average of 2,640 meters (8,660 ft) above sea level. Subdivided into 20 localities, Bogotá has an area of 1,587 square kilometers (613 square miles) and a relatively cool climate that is constant through the year.

Bogotá stands out for its economic strength and associated financial maturity, its attractiveness to global companies and the quality of human capital. It is the financial and commercial heart of Colombia, with the most business activity of any city in the country. The capital hosts the main financial market in Colombia and the Andean natural region and is the leading destination for new foreign direct investment projects coming into Latin America and Colombia. It has the highest nominal GDP in the country, responsible for almost a quarter of the nation's total (24.7%).

The city's airport, El Dorado International Airport, named after the mythical El Dorado, handles the largest cargo volume in Latin America, and is third in number of people. Bogotá is home to the largest number of universities and research centers in the country, and is an important cultural center, with many theaters, libraries and museums.

Colombia is located at the north side of South America, and due to its economic stability compared to the rest of the Latin America, has become the hub for the rest of the region.

## **1.2 THE PROBLEM**

A lot of companies have located their Latin America headquarters in Bogota, which makes Bogota a place for a lot of expats.

Usually with a vast city like Bogota, expats don't know which will be the best place to live, sometimes they receive advice from local people, but don't have a data driven decision.

## **1.3 TARGET AUDIENCE**

This project has as target audience, people that want to live in Bogota, and take a data driven decision before moving to certain neighborhood.

## **2. DATA COLLECTION**

Before starting data mining, is important to have clarity of what kind of data will be needed to take the decision.

According to Twin Cities habitat for humanity web site, people first look for neighborhood safety, especially if you have a child, as will give greater peace of mind.

Food, medications, and toiletries are necessities for daily life. If convenience is important, you may want to consider choosing a location that's close to where you will be running a majority of your errands.

It is still important to check out the outdoor amenities a neighborhood has to offer, as sidewalks, parks, walking trails and dog parks.

Where you work in relation to where you live is an important factor to consider, especially if your goal is to have a short commute. If driving yourself isn't an option, or one you prefer to avoid, look into transportation options.

## 2.1 BOGOTA ZIP CODES

To find the zip codes of neighborhood in Bogota, the postal code website was used, as it contains a csv file with all the required info:

[http://visor.codigopostal.gov.co/472/visor/Codigos\\_Postales\\_Nacionales.csv](http://visor.codigopostal.gov.co/472/visor/Codigos_Postales_Nacionales.csv).

As the name of the link indicates, the file contains all info for Colombia, and need to be filtered to keep only Bogota data.

## 2.2 NEIGHBORHOODS SAFETY

A report from a local government web site was found, where information need to be scrapped, as it was published as a pdf file.

[https://scj.gov.co/sites/default/files/documentos\\_oaiee/Boletin\\_2020\\_05\\_Reporte\\_bogota\\_2020\\_05.pdf](https://scj.gov.co/sites/default/files/documentos_oaiee/Boletin_2020_05_Reporte_bogota_2020_05.pdf)

## 2.3 VENUES

As source of venues data, foursquare will be used, taking advantage of the zip codes information.

For work information I find that is really dependable of where the office will be located at, for that reason, the scope of the project will be to show where will be the best places to live, using the classification named above.

## 3. METHODOLOGY

In order to manipulate the info, pandas was used as the main source. All the info was taken and converted to **PANDAS** data frames, where columns names where changed for better understanding and manipulation.

Once the information was ready for usage, **MATPLOTLIB.PYPILOT** was used to visualize top 5 boroughs, depending the data that wanted to show. Horizontal bars, histograms and box plot were mainly used.

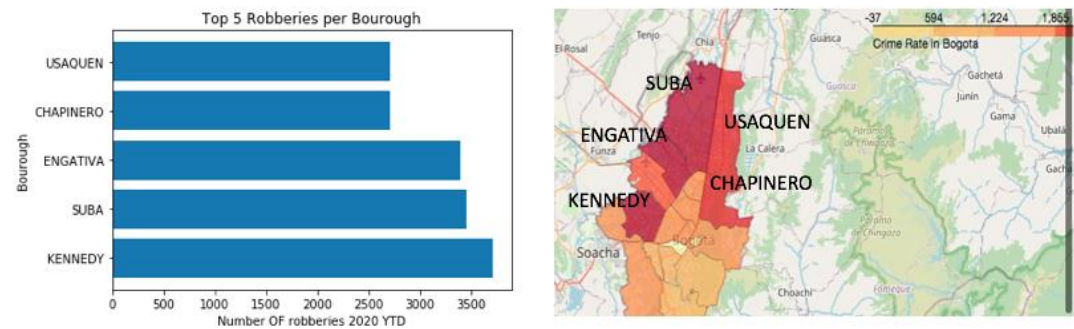
To visualize the distribution of robberies in the cities was used **CLOROPETH**, as we used it on san Francisco crime data.

To respond the question where to live in Bogota, **KMEANS** methodology was used to cluster all Bogota, and make the conclusions, this methodology was

used, as for many people that arrives to a new city, it is important to know what will be near by their living place.

4. RESULTS

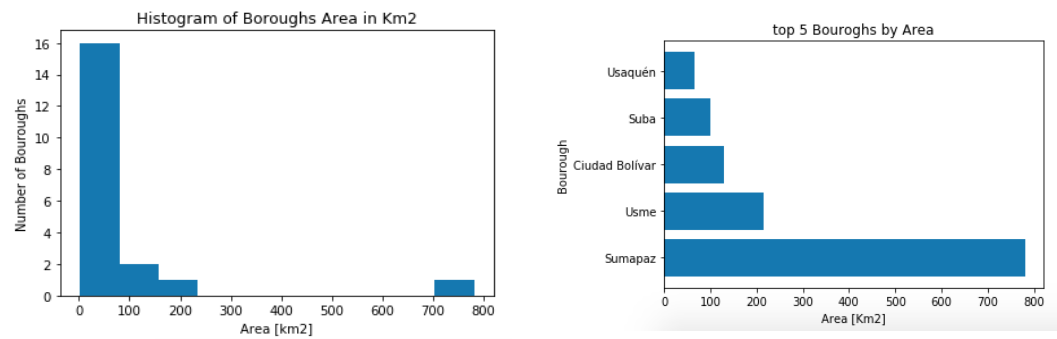
4.1 BOGOTA SAFETY INSIGHTS



Bogota is a city with 20 Boroughs. For a person that doesn’t know where to live in the city, the first thing to know will be how is the safety of each borough. Taking into account that Bogota is a city with 10M habitants including metropolitan area, number of robberies are vast.

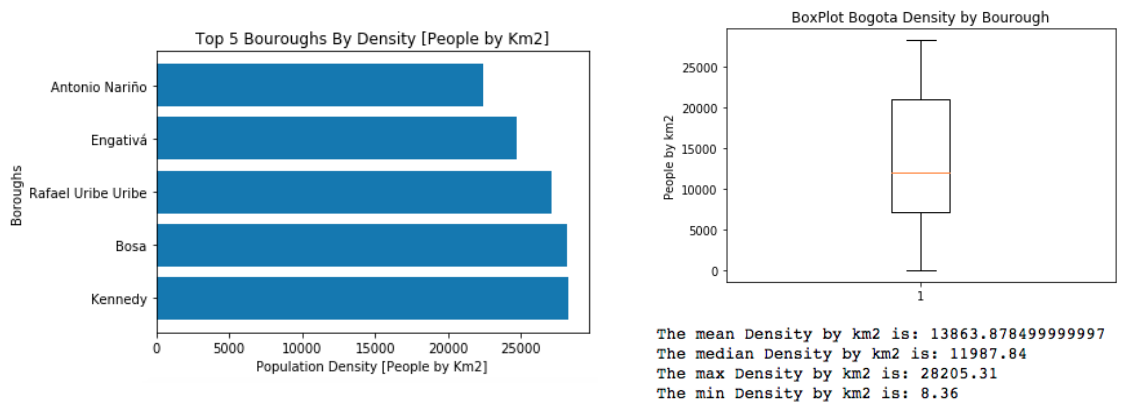
The top 5 boroughs from robberies data are the one shown on the horizontal bar above, their geographical location is shown at the right, where it can be seen that this top 5 boroughs are close each other, and are located at the north side of the city.

4.2 BOGOTA BOROUGHS AREAS IN KM2



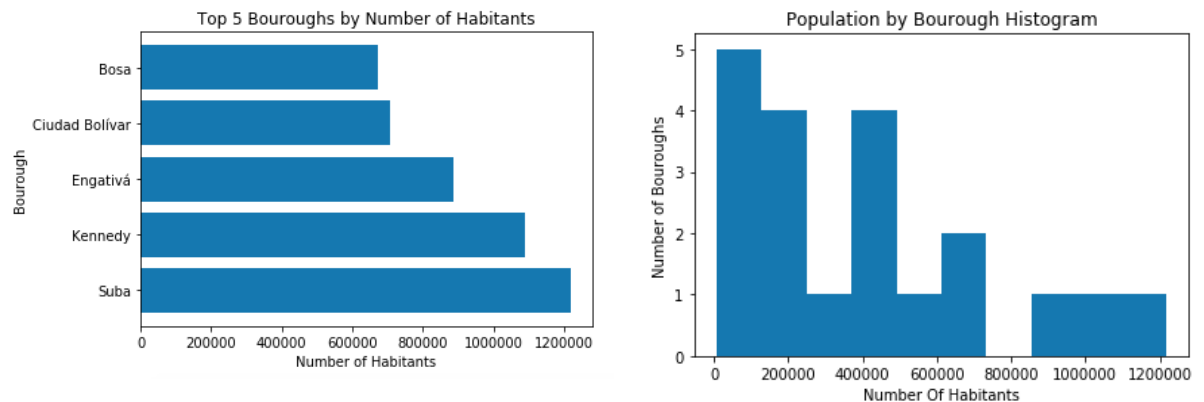
Boroughs are mainly below 100km2 as shown on histogram above. The major anomaly is Sumapaz, which is above 700km2, and includes a national park.

4.3 BOGOTA BOROUGHS DENSITY (HABITANTS X KM2)



The major density in Bogotá is located in Kennedy, also this borough was part of the top 5 by robberies. Sumapaz represents an anomaly as its vast area, doesn't have too much people living on it. The mean density for the entire city is 13.8k habitants x km2 and its median is 11.9k habitants x km2. To have a benchmark, the webpage worldpopulation.com says that NY has a population density of 27.3k per km2.

#### 4.4 BOGOTA BOUROUGH'S POPULATION



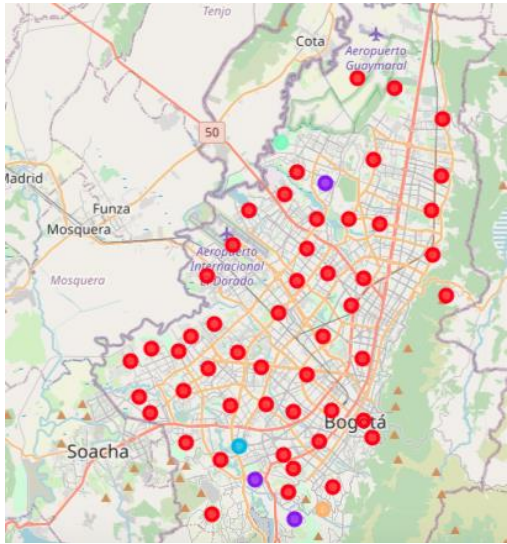
Population numbers shows 50% of Boroughs have a mean population below 250k, but the Kennedy and Suba account for 25% of the entire population of the city. It must be noticed that for density and total population, Usaquén and Chapinero are not in the top 5, but they are part in the top 5 of robberies, this is because the GDP per capita increase in Usaquén and Chapinero boroughs.

#### 4.5 CLUSTERING BOGOTA BOUROUGH'S AND NEIGHBORHOODS

Top 5 venues of largest cluster for all Bogotá Boroughs.

Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0 111321	0.0	Restaurant	Pizza Place	Latin American Restaurant	Burger Joint	Sandwich Place
2 110931	0.0	Bar	Bakery	Theme Park	Furniture / Home Store	Cupcake Shop
4 110841	0.0	Pizza Place	Grocery Store	American Restaurant	Athletics & Sports	Cultural Center

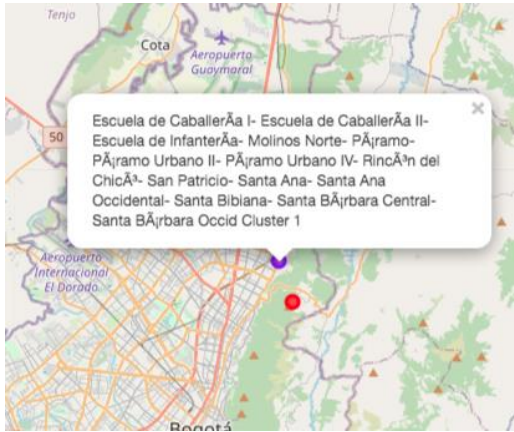
Bogota Map | Clustered



Kmeans was used trying to cluster Bogota neighborhoods. 5 clusters were used, where is noticed that most of the city is red dotted, which is cluster 0. This cluster corresponds to places with restaurants and parks as shown in table above.

In order to narrow clustering, the bouroughs of chapinero and usaquen were used only.

4.6 USAQUEN AND CHAPINERO CLUSTERING



	Postal Code	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
I	110111	1.0	Asian Restaurant	Steakhouse	Hotel	French Restaurant	Coffee Shop

After applying 5 clusters to Usaquen and Chapinero, was found that the zone where most expats will be like to live is cluster number 1. This cluster top 5 is surrounded by international restaurants and hotels, which give an insight of the office density that has the area.

## **5. DISCUSSION SECTION**

The main objective was to find which will be the best place to live in Bogota, based on available data. Before doing the clustering as important to have insights of the total city, to be able to understand overall city dynamics.

After applying clustering to the entire city, the obtained results didn't have much conclusion, as most of the neighbors were one cluster. Was necessary to reduce the boroughs and see the results.

Data science is a subject that it is not well organized yet in the country, and data still very difficult to find, a lot of thing can be found in pdf format, which difficult the fast mining.

## **6. CONCLUSION**

Bogota is a city with 8M habitants and 10M with metropolitan area. Safety data it is necessary to review before living on any city to take the most data driven decision. Also, it is important to segment data when a large city is used as it happened with Bogota.