

# Finding the best place to open a new hostel

Andrés Herrera

February 9, 2020

## 1) Introduction

### a) *Background*

The objective of this project is to find the optimal location to open a new hostel in Bogotá, the capital city of Colombia. Over the last years, Colombia has been highly esteemed by tourists as one of their travel destinations. According to town hall, 12 million people came to Bogotá in 2018 and 54% of tourism was leisure tourism [1]. The number of tourists has been increasing throughout the years, turning tourism into a market niche to take into consideration for potential investors.

Data science is a highly effective tool to study big amounts of data and take out of it useful information that can be used as an advantage. Analyzing a city, its districts, venues and making divisions according to its homogenous and heterogeneous characteristics can be a hard task to do with only the naked eye. Therefore, using computational science techniques such as k-means clustering can be useful to study a big city such as Bogotá and make distinctions between its districts.

Bogotá is divided in 20 districts which have a lot of homogenous characteristics. In this project, k-mean clustering and Foursquare data will be used as a way to divide the city districts into homogenous groups that will provide information about which area could be more appreciated by tourists.

## **b) Business problem**

In this project, I focus on the problem of where to open a new hostel in Bogotá city. Imagine company Z wants to open a new hostel in the city and wants to find the best area that will provide their clients with recreation and leisure tourism which includes guided tours and experiencing local tourist attractions such as local restaurants, museums, local shops and night life. Therefore, I have been given the task to provide them with the information needed to take a final decision. The business problem can be summed up in these questions:

- How can the city be divided into clusters with homogenous characteristics?
- Which cluster or area can provide tourists with most of local tourist attractions?
- What are the main venues categories that tourist could appreciate more?

## **2) The code**

### **a) Data description**

Bogotá is divided in 20 Boroughs and more than 3,000 neighborhoods. There are two webpages from which this information was acquired:

1. [https://es.wikipedia.org/wiki/Unidades\\_de\\_Planeamiento\\_Zonal](https://es.wikipedia.org/wiki/Unidades_de_Planeamiento_Zonal)
2. <https://mapas.bogota.gov.co/>

After data cleaning, a table with Bogota's boroughs, neighborhoods and their respective latitude and longitude was obtained. The table with its first 5 columns after data cleaning:

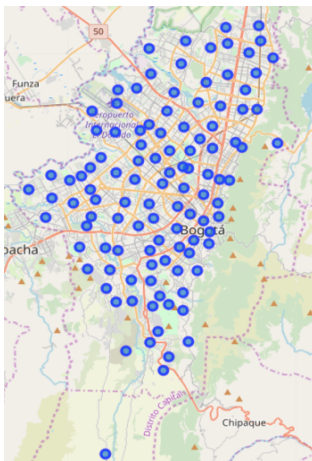
|   | Borough     | Neighborhood    | Latitude | Longitude  |
|---|-------------|-----------------|----------|------------|
| 0 | Teusaquillo | GALERIAS        | 4.642904 | -74.072083 |
| 1 | Kennedy     | LAS MARGARITAS  | 4.637752 | -74.178175 |
| 2 | Teusaquillo | QUINTA PAREDES  | 4.631541 | -74.090252 |
| 3 | Teusaquillo | TEUSAQUILLO     | 4.626522 | -74.075133 |
| 4 | Santa Fe    | SAGRADO CORAZON | 4.619246 | -74.063995 |

The above table was used to explore the main venues via the Foursquare API:

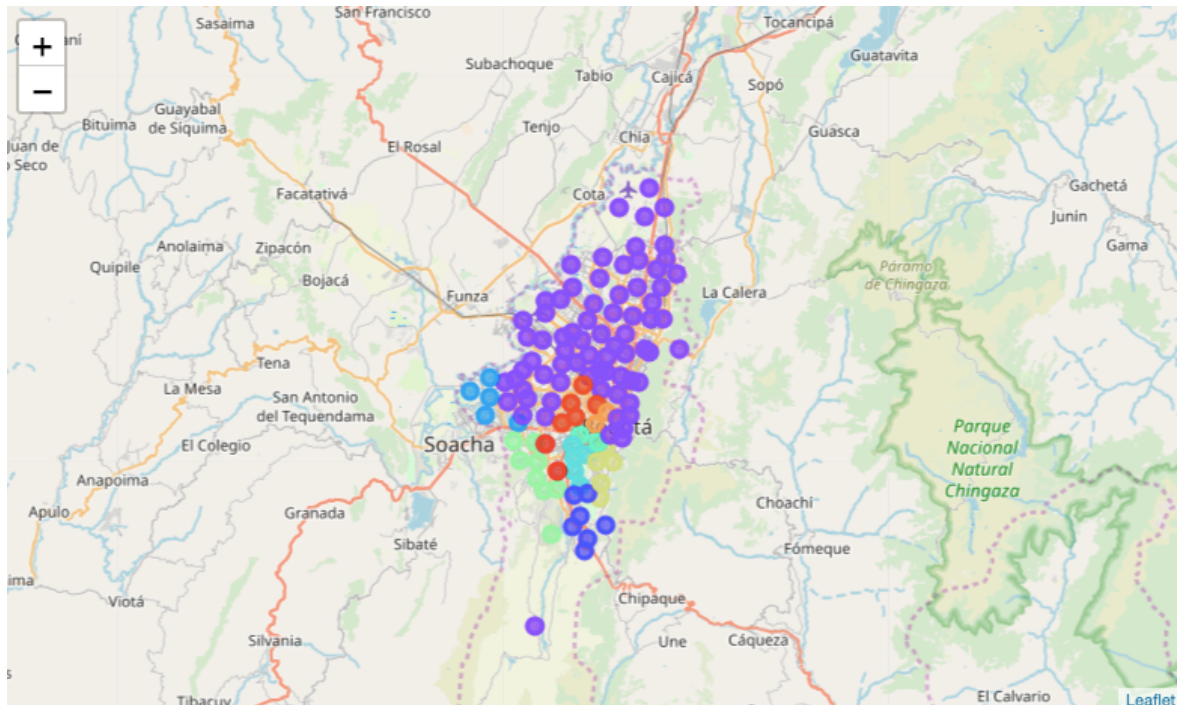
<https://api.foursquare.com/v2/venues/explore?>

## b) Map Visualization and clustering

With help of Folium it was possible to visualize Bogota and show the neighborhoods of the city.



With help of the Foursquare API and the **explore** function, the most common venue categories in each neighborhood were obtained. Afterwards, this feature was used to group the neighborhoods into clusters. This was achieved by using the k-means clustering algorithm and the emerging clusters were visualized.



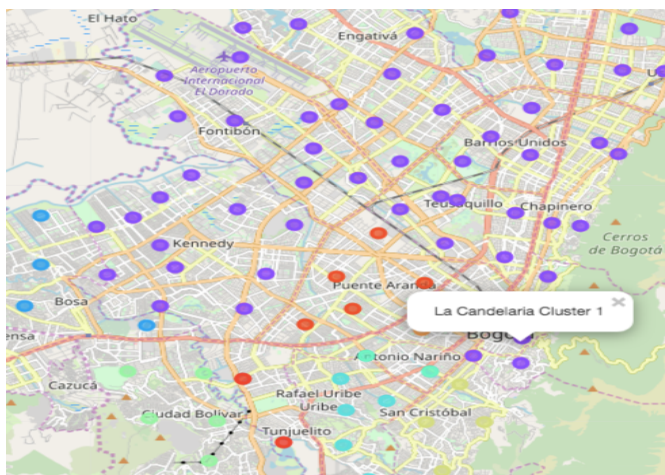
### 3) Results

Via K-means clustering, Bogotá city was divided into groups that gave information about the heterogeneous characteristics of the city. There was a particular cluster (purple) that encompassed a considerable part of the city. However, the other clusters main venues were related with industrial areas or didn't have a considerable amount of venues related with hotels, restaurants, museums or parks, which could be upsetting areas for leisure tourists.

Therefore, a further analysis was given to the purple cluster and compelling information was found. A deeper search was given and it was found that the neighborhood La Candelaria 1<sup>st</sup> most common avenue is a Hostel avenue. Likewise, Chapinero and Usaquen were found to have Hotels in their most common venues.

| District      | Cluster Labels | 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 | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------------|-----------------------|-----------------------|-----------------------|------------------------|
| La Candelaria | 1              | Hostel                | Breakfast Spot        | Restaurant            | Italian Restaurant    | Coffee Shop           | History Museum            | Café                  | Theater               | Art Museum            | Pub                    |
| Chapinero     | 1              | Bar                   | Restaurant            | Coffee Shop           | Fast Food Restaurant  | Bakery                | Latin American Restaurant | Hotel                 | Supermarket           | Japanese Restaurant   | Italian Restaurant     |
| Usaquén       | 1              | Burger Joint          | Bakery                | Coffee Shop           | Italian Restaurant    | Restaurant            | Hotel                     | Café                  | Steakhouse            | Asian Restaurant      | Mexican Restaurant     |

Clearly, La Candelaria is an optimal neighborhood to open a new hostel because it counts with restaurants, Coffee shops (Colombian Coffee is known worldwide), Museums and Pubs. It seems that La Candelaria is a district with *social and touristic movement* that makes it an optimal place to open a new hostel. Furthermore, Chapinero and Usaquen have appealing venues for leisure tourism. Therefore, special consideration should be given to these districts.



#### 4) Conclusion

In particular, a district called La Candelaria has the biggest potential for opening a new hostel because of the most common venues in the district. Likewise, Chapinero and Usaquen are potential areas to open a new hostel because there is presence of hotels and common venues that could be appreciated by tourists. Certainly, a more advanced analysis should be given by the investors in terms of costs and potential revenues in order to make a more accurate decision in terms of their objectives and expectations.

## **Bibliography**

[1] <https://www.eltiempo.com/bogota/cuantos-turistas-visitaron-bogota-en-2018-cifras-en-indicadores-de-turismo-347100>