Finding the best place to open a new hostel

Andrés Herrera

February 9, 2020

Introduction

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 tourists came for holiday, recreation and 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 and maybe even impossible 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.

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?

Data description (City analyzed: Bogotá)

Bogotá is divided in 20 districts and 112 subdivisions that are given an UPZ number associated to one district. There are two main webpages from which this information can be acquired:

1. https://es.wikipedia.org/wiki/Unidades_de_Planeamiento_Zonal
2. https://mapas.bogota.gov.co/

Once the longitude and latitude information is obtained, Foursquare can be used to explore the venues and cluster the districts according to the main venues. To do this, the Foursquare API that will be primordially used is:

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

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. This was mysterious to find because it can be concluded that all that vast area has similar characteristics. It doesn't give much information about what place could be the perfect one to open a new hostel which was a problem to tackle.
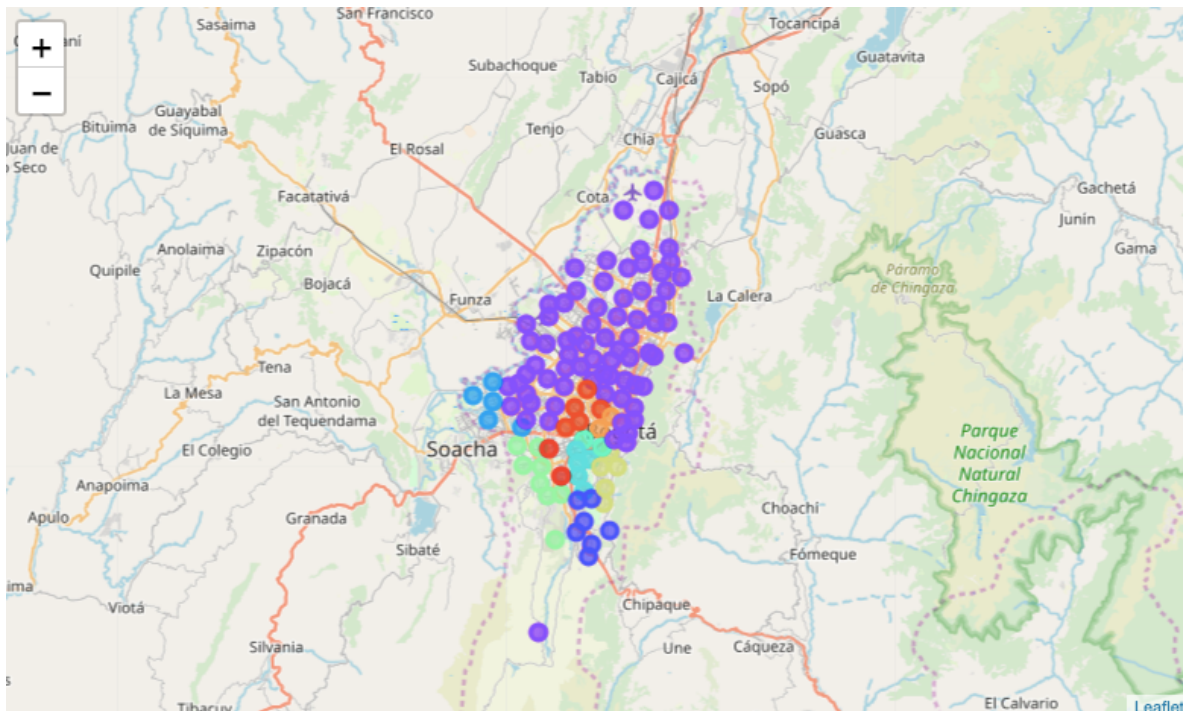
Looking at the clusters and the main avenues, all clusters except the purple one (cluster 1) didn't have main avenues that a tourist would particularly consider to go during their leisure vacations. Therefore, a proper visual analysis of the districts and their main avenues gave as result that there is a District which 1st most common avenue is a Hostel avenue. That District is called La Candelaria and its most common avenues are depicted in the following image:

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

Certainly, if the 1st most common venue is a Hostels venue it gives valuable information because one can infer that other investors chose this District to open a hostel for a particular reason. Making a further analysis, the most common venues are 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.
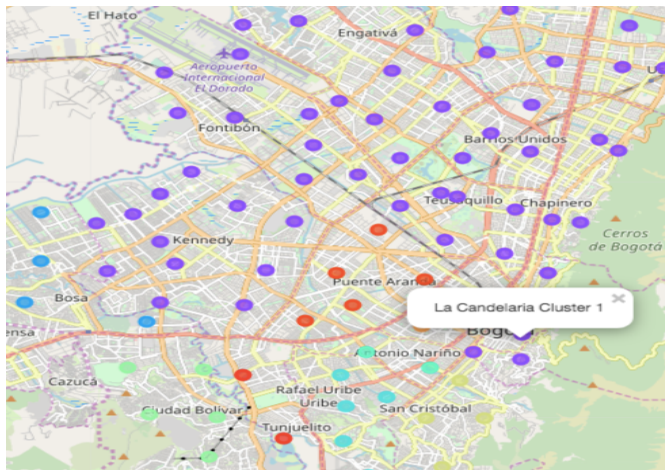
Discussion

The vastness of the purple cluster (cluster 1) can not be unnoticed. With a visual analysis of the information gathered, it was possible to make a distinction between the districts and notice one that *luckily* gave information about where numerous hostels can be found. However, a proper analysis of the cluster 1 can be made to look for alternative options to open a new hostel if La Candelaria presents competitive problems for the investors.

Conclusion

In conclusion, Cluster 1 was found to be the one as a potential area to open a new hostel in terms of its most common avenues. In particular, a district called La Candelaria has the biggest potential for opening a new hostel because of common avenues in the district. Certainly, a more advanced analysis should be given by the investors in terms of costs and potential revenues.

Furthermore, in case that La Candelaria has numerous hostels and presents competitive problems for the investor, efforts should be given towards clustering the cluster 1 and make the same that was done in this project but focusing in clustering the purple section in the map. Special attention should be given to Teusaquillo and Chapinero because by a simple visual analysis, the most common venues could be of interest for tourists.

Bibliography

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