Bike -Sharing Stations in Bogota

Lucia Perez Ramirez

January 27, 2019

Abstract

In this project I propose suitable locations for bike-sharing stations in Bogota (if the system was to be implemented) based on the most visited places withing the city. For this, I use several tools including the Foursquare API, the Folium visualization library and the DBSCAN clustering package.

1 Introduction

Bogota is ranked as the fifth most traffic congested city of the world. It takes the first place among South American capitals. According to the Global Traffic Scorecard of INRIX (2018), a Bogota citizen loses 75 hours a year in traffic jams. In fact, most Latin American cities have to deal with a rapid and unplanned urban growth which represents a major challenge in mobility and traffic dynamics.

Given this panorama, the promotion of the bicycle as a daily and safe mode of transport has become a common objective within the policies of sustainability and equity in large cities. The use of the bicycle not only reduces carbon emissions within big cities, it also helps to alleviate traffic congestion, decreases travel times, and favors people's health and wellbeing.

Still, Bogota is famous across the world for being a bike friendly city. It has a population of around eight million people and cycle paths covering more than 360km (220 miles) of the citys surface. Almost 84,000 people use Bogotas cycle route network every day, which only stands for around 1% of the total population. This has made local government to ask themselves 'How to make the bicycle a daily and safe mean of transport for most people?'.

I think that a bike-borrowing system would be appropriate for a city like Bogota in order to answer this question. This solution also deals with other concerns among citizens which include vandalism, parking or storage, and maintenance.

But then again another question arises and this is **which would be the ideal locations to put bike-sharing points within the city?**

2 Data

Bogota is divided in 20 localities or districts, which are also subdivided in Zonal Planning Units (UPZ in Spanish) and Rural Planning Units (UPR). There is a total of 112 UPZs and 4 UPRs in the city.

To segment the UPZs and explore them, we essentially need a dataset that contains the 20 districts with their corresponding UPZs as well as the the latitude and longitude coordinates of each UPZ.

I was able to download a dataset with all the UPZs and their location from this website as a csv file (Mapas Bogota). I uploaded it to my GitHub repository here. However, they don't have their associated district. Luckily, I found a table that included each UPZs number and to which district it belongs in this Wikipedia page. I used the **BeautifulSoup** package in Python to scrape the wiki page and get the dataframe, then used the merge command to join both dataframes to get a final one. I tried to illustrate the process here:

Dataset 1

	UPZ	Name UPZ	Long	Lat
0	100	Galerias	-74,072	4,643
1	83	Las Margaritas	-74,178	4,638
2	107	Quinta Paredes	-74,090	4,631
3	101	Teusaquillo	-74,075	4,626
4	91	Sagrado Corazon	-74,064	4,619
	...	...	...	...

join on 'UPZ' ⟷

Dataset 2

	UPZ	DisID	District
0	1	1	Usaquen
1	9	1	Usaquen
2	10	1	Usaquen
3	11	1	Usaquen
4	12	1	Usaquen
	...	...	...

Resulting dataframe

	UPZ	Name UPZ	Longitude	Latitude	Dis_ID	District
0	100	Galerias	-74.072083	4.642904	13	Teusaquillo
1	83	Las Margaritas	-74.178175	4.637752	8	Kennedy
2	107	Quinta Paredes	-74.090252	4.631541	13	Teusaquillo
3	101	Teusaquillo	-74.075133	4.626522	13	Teusaquillo
4	91	Sagrado Corazon	-74.063995	4.619246	3	Santa Fe
	...	...	...	...	...	...

Figure 1: Data exploration and wrangling.

I also found some useful GeoJSON files for visualization purposes. One of them delimits each district within the city and can be downloaded from this GitHub repository. The other one demarcates the existing bike routes in the city, I downloaded from this webpage (Movilidad Bogotá) and can as well be found in my GitHub repository here.

3 Methodology

To find the ideal location for bike-sharing stations I started from the assumption that the stations should be near to the most visited places in the area (e.g. restaurants, bars, stores, etc.). For this prupose I used the Foursquare API to get the top venues of each UPZ and cluster them using a Density-based spatial (DBSCAN) machine learning algorithm. The center of the clusters would be defined as the location of the bike-sharing stations.

To reduce a bit the scope of the project and the amount of information, I only worked with the **Teusaquillo** district which is a locality where most of the cultural life and commercial activity of the city takes place. Teusaquillo has 6 UPZs, as shown in this table:

Table 1: Zonal Planning Units of Teusaquillo district.

	UPZ	**Name UPZ**	**Longitude**	**Latitude**	**Dis_ID**	**District**
0	100	Galerias	-74.072083	4.642904	13	Teusaquillo
1	107	Quinta Paredes	-74.090252	4.631541	13	Teusaquillo
2	101	Teusaquillo	-74.075133	4.626522	13	Teusaquillo
3	106	La Esmeralda	-74.086950	4.647636	13	Teusaquillo
4	109	Ciudad Salitre Oriental	-74.101445	4.644015	13	Teusaquillo
5	104	Parque Simon Bolivar - CAN	-74.091185	4.648992	13	Teusaquillo

Finally, visualization was aided by the use of the Folium package for interactive maps.

3.1 Foursquare API

In order to use the Foursquare API it is necessary to create a Foursquare developer account and have your credentials handy (CLIENT_ID and CLIENT_SECRET). After this, it is possible to construct a URL to send a request to the API. In this case, I used the *'explore'* call, which returns a list of recommended venues near the current location. To define the URL, some other parameters are necessary (LATITUDE, LONGITUDE, VERSION, RADIUS and LIMIT). The URL will be defined as follows:

https://api.foursquare.com/v2/venues/explore?client_id=**CLIENT_ID**
&client_secret=**CLIENT_SECRET**&ll=**LAT**,**LNG**&v=**VERSION**&radius=**RADIUS**
&limit=**LIMIT**

I used a function to loop through each Teusaquillo's UPZ to get their top 100 venues (this is the limit in Foursquare API) using a radius of 1600 m.

3.2 DBSCAN

DBSCAN is a density based data clustering algorithm (its long name is *Density-based spatial clustering of applications with noise*). Given a set of points in some space, it groups

together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away). DBSCAN library can be imported from the sklearn.cluster package in Python.

The two most important parameters to define when using DBSCAN are *eps*, which is the maximum distance between two samples for them to be considered as in the same neighborhood, and *min_samples*, which is the number of samples in a neighborhood for a point to be considered as a core point, including the point itself. In this case I chose $eps = 0.24$ and $min\_samples = 12$.

4 Results

The DBSCAN algorithm returned a total of 11 clusters, which accounts for 11 bike-sharing stations (excluding the 'outliers' cluster). I plotted the clusters along with their centroids using the Folium visualization library for interactive maps. Note that I also put a GeoJSON layer which shows the bike routes around the city and are plotted in blue. The results are shown in Figure 1.

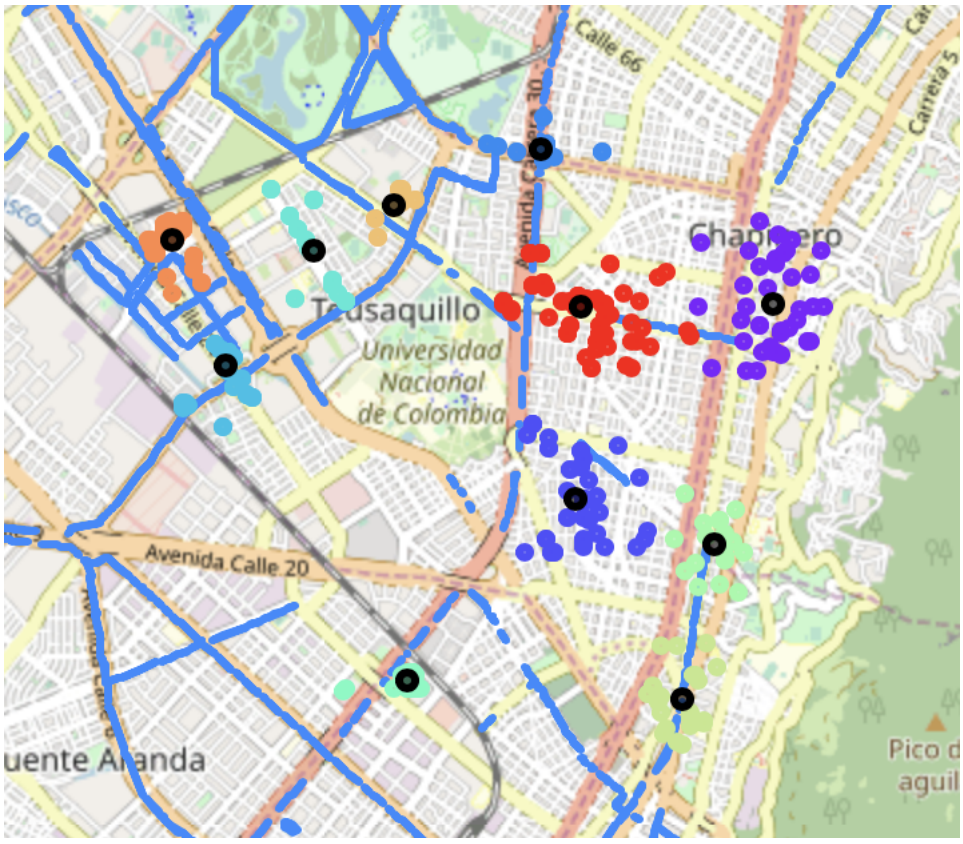

Figure 2: Clusters of top venues in Teusaquillo

The centroids (plotted in black) were obtained by calculating the mean of the latitude and longitude values of each cluster. There was a total of 138 outliers (from 600 total venues).

5 Discussion

If we take a closer look at the centroids of the clusters (thus, the location of our bike stations), we see most of them are located in known spots of the city, like shopping malls, hotels or parks. The list of the spots would be as follows.

Bike stations in shopping malls:

- Calima
- Galeras Centro Comercial
- Gran Estacin
- San Marcos

Bike stations near hotels:

- Hotel Tryp

Bike stations in parks:

- Parkway
- Paque Central Bavaria
- Parque Nacional
- Parque La Esmeralda

Others:

- Carulla supermarket (cra 54 cll 57)
- Movistar Arena

This is an interesting finding, as it shows that DBSCAN is a convenient algorhitm for the purpose required in this project.

6 Conclusions

It was possible to find the location for possible bike-sharing stations in Bogota using the Foursquare API and DBSCAN clustering algorithm.

If there was to be a bike sharing system in the city, in the district of Teusaquillo, it would be convenient to have 11 bike-sharing points, according to the density-based algorithm.

Most of the points were found to be near shopping malls or parks in the city, which makes it convenient in terms of infrastructure, if the stations were to be implemented.