INTRODUCTION

In spite of the use of social media and internet as the massive way to advertise something, nowadays we keep seeing advertising in the middle of the cities and streets like posters or billboards. This means that promote a product or event through posters among the city is an old but still efficient way to publicize. So, the question that I will try to answer in this project is: where would you put an advertising in a city like Bogotá in order to impact the greatest number of people?

Even though the impact of a publicity depends on other factors like what product or event are your promoting, what are your target customers and what is the strategy used in the poster, to choose a correct place to show your advertising can be an advantage and also the definitive factor when we measure the success or fail of the publicity. It doesn't have sense to spend many dollars for a poster if at the end it will be display in an incorrect place and nobody will see it.

Hence, we should choose areas with a good visibility where you can impact the right audience. That's why it is always a good idea to show our advertising in public spaces, events, companies, restaurants, hotels, coffee shops, shopping centers, and trending places where there is a big affluent of people.

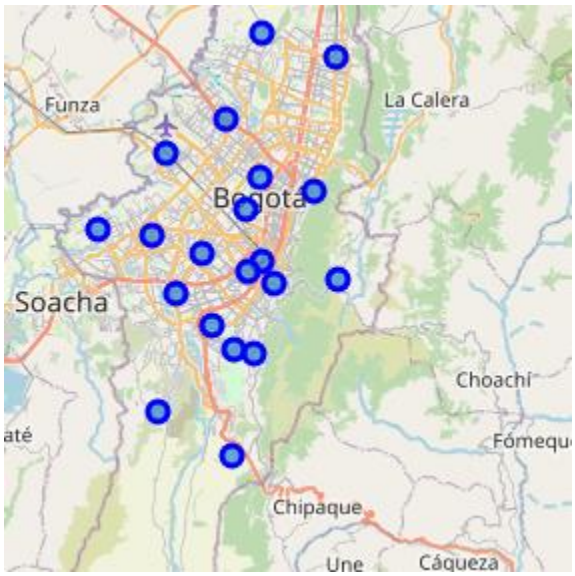
Therefore, the objective in this project will be found the busiest areas in the city of Bogotá where we can put our advertising and get positive results.

DATA

The data is based in the city of Bogotá, the capital city of Colombia. The city is divided into 20 different borough and we will use foursquare API to identify the venues for each borough and neighborhood.

I will use a table like the following with latitude and longitude of every location in Bogotá.

	LOCALIDAD	LONGITUD	LATITUD	CODIGO	gp
0	CHAPINERO	-74.0467	4.6569	2	-74.0467,4.6569
1	TUNJUELITO	-74.1407	4.5875	6	-74.1407,4.5875
2	ANTONIO NARIÑO	-74.1009	4.5486	15	-74.1009,4.5486
3	PUENTE ARANDA	-74.1227	4.6149	16	-74.1227,4.6149
4	USAQUÉN	-74.0312	4.7485	1	-74.0312,4.7485



Map of the localities of Bogota

Additionally, the table with the information about venues like name, ID, location and category from Foursquare API is the following:

```
(1351, 7)
```

	LOCALIDAD	LATITUD	LONGITUD	LUGAR	LUGAR LATITUD	LUGAR LONGITUS	CATEGORIA
0	CHAPINERO	4.6569	-74.0467	Bandido Bistro	4.661514	-74.050307	French Restaurant
1	CHAPINERO	4.6569	-74.0467	Quebrada La Vieja	4.650833	-74.049511	Scenic Lookout
2	CHAPINERO	4.6569	-74.0467	El Caracol Azul	4.656121	-74.053203	Peruvian Restaurant
3	CHAPINERO	4.6569	-74.0467	Harry Sasson	4.659021	-74.054525	Restaurant
4	CHAPINERO	4.6569	-74.0467	Brot Bakery & Cafe	4.663257	-74.050578	Bakery

So, we get 1351 venues from all the localities of Bogotá.

METHODOLOGY

Exploratory Data Analysis

First, I determine how many venues we have for each locality, so I decide to eliminate the localities with less than 15 venues. Therefore, we eliminate: CIUDAD BOLIVAR, SAN CRISTOBAL, SANTA FE, SUMAPAZ, and USME.

Machine Learning

The objective was to find the busiest areas of the city, so I use DBSCAN to find the areas with more places in the city.

After the fitting, I get 21 labels, so here is a small part of the table with its respective labels.

	LOCALIDAD	LATITUD	LONGITUD	LUGAR	LUGAR LATITUD	LUGAR LONGITUS	CATEGORIA	Labels
0	CHAPINERO	4.6569	-74.0467	Bandido Bistro	4.661514	-74.050307	French Restaurant	0
1	CHAPINERO	4.6569	-74.0467	Quebrada La Vieja	4.650833	-74.049511	Scenic Lookout	0
2	CHAPINERO	4.6569	-74.0467	El Caracol Azul	4.656121	-74.053203	Peruvian Restaurant	0
3	CHAPINERO	4.6569	-74.0467	Harry Sasson	4.659021	-74.054525	Restaurant	0
4	CHAPINERO	4.6569	-74.0467	Brot Bakery & Cafe	4.663257	-74.050578	Bakery	0

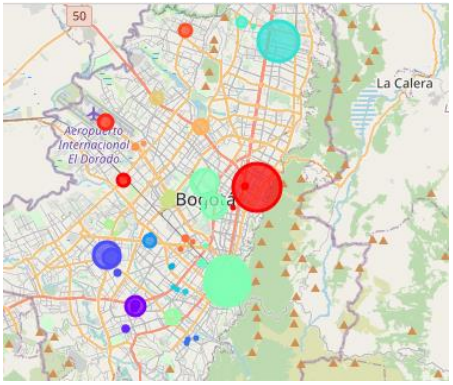
RESULTS

Therefore, we could identify that the localities with more closest venues were CHAPINERO, LA CANDELARIA, BOGOTÁ, USAQUÉN, and LOS MÁRTIRES.

	LOCALIDAD	Labels	LUGAR LATITUD	LUGAR LONGITUS	Count	sdt Count
7	CHAPINERO	0	4.659989	-74.053965	100	2.020202
19	LA CANDELARIA	11	4.604675	-74.070945	93	1.878788
6	BOGOTÁ	11	4.607775	-74.070968	90	1.818182
49	USAQUÉN	8	4.743890	-74.041702	82	1.656566
24	LOS MÁRTIRES	11	4.603669	-74.071974	74	1.494949

Where count represent the number of venues in that locality that belong to its respective Label.

Additionally, we can identify the proportion of all localities in the following map:



Map with clusters.

DISCUSSION

I can notice that our project was limit by the number of venues per call, so it will be interest to realize the same procedure dividing the localities into neighborhood and finding more venues of each one. Another aspect to improve is the change of the radius that we use to make the call in Foursquares API, another improvement can be to find new models with different radius.

For the DBSCAN method I use the parameter min_samples equal to 10. It is an opportunity to find new results if we decide to change this parameter. Finally, It would interesting if we could add more features to the clustering method because the successful of an advertising doesn't only depends of the population density.

CONCLUSION

In this project, I create a list with at most 100 venues per locality, then I cluster then using the method DBSCAN, so I cluster it based in density or looking for the areas with more public venues. I could group the venues and finally found that the better areas to advertise something using a poster are CHAPINEIRO, LA CANDELARIA, BOGOTÁ, USAQUÉN and LOS MARTIRES. There are still parameters to improve but I could find a good model.