Capstone project

Coursera IBM Data Science Certification

Seyed Jamal Haddadi

Report Contents

1- Introduction:

> ➤ The "business problem" to be solved by this project and who may be interested.

2- Data:

> ➤ Describe Data requirements and Sources needed to solve the problem.

3- Methodology:

> ➤ Execute data processing, describe/discuss any exploratory data analysis and/or inferential statistical testing performed, and/or machine learnings used.

4- Execution and Results:

> ➤ Discussion of the results and finding of answer.

5- Recommendation:

> ➤ Finding the best option based on comparison.

1- Introduction

The goal of this project is to use Data Analysis from web sources to select the best spot in Bogota to open a new burger joint. Due to high population diversity and enormous area of Bogota, all the aspects must be taken into consideration to provide a list of best spots as a suggestion to the entrepreneur.

There are 2 different localities and the aim is to find the place which has not many burger joint around to avoid competition. Also, the selected place must be a rich neighborhood and has an acceptable number of customers for best profits.

2- Data

To perform this task, we will need to access the following data:

- The Localities of Bogota in Colombia from Wikipedia: *https://es.wikipedia.org/wiki/Anexo:Localidades_de_Bogot%C3%A1*

- The coordinates (latitude, longitude) of these Localities in Bogota from Open Street Map APIs

  Also, from Foursquare, we will need the following venues data:

- The burger joint venues of the Localities
- The offices venues of the Localities
- The high schools' venues of the Localities
- The universities venues of the Localities
- We will then leverage the data in order to determine which locality is the most appropriate in order to locate the burger joint.

3- Methodology

The strategy to find the answer:

Our strategy is based on mapping the descried data in section 2, in order to facilitate the choice

For each locality, all office, school, university and burger joints venues data have been collected from Foursquare. Then, for each locality, the sums of the office, school, university and burger joints were

computed. For each of these 4 categories, a weight (or penalty) has been defined according to what Client (Soheil) considers the most important. Therefore:

- Burger Joints have been weighted with -1, since Paolo wants to avoid concurrence.
- Schools have been weighted with 1, since student are good customers.
- Universities have been weighted with 1.5, since students are good customers.
- Offices have been weighted with 2, since employees are even better customers.
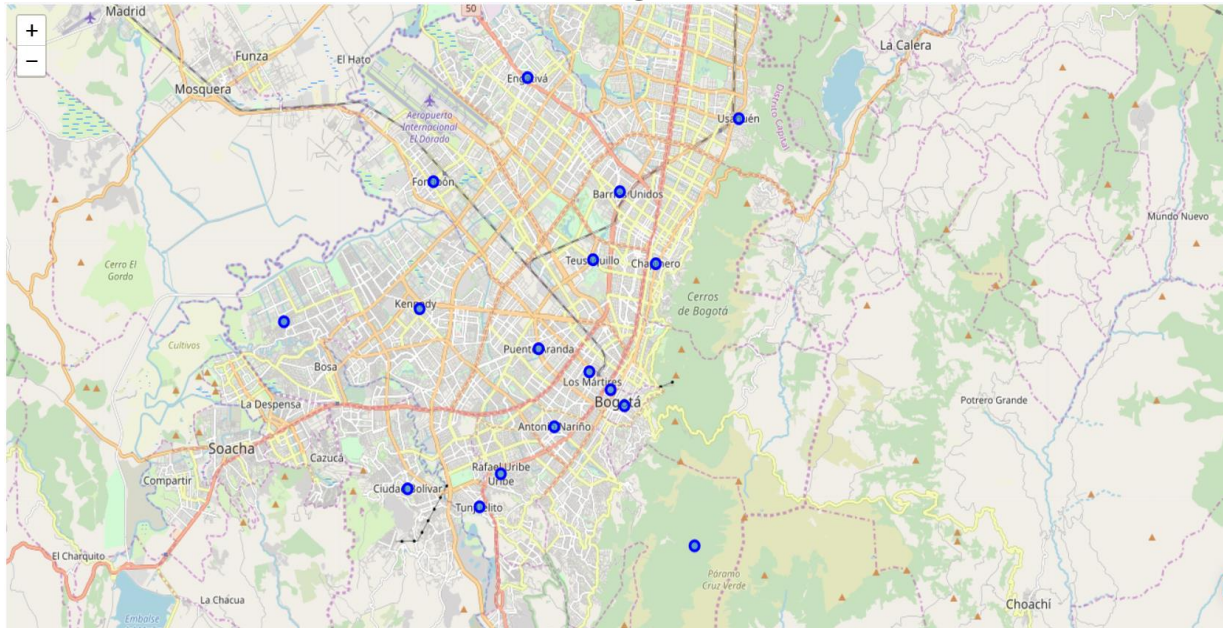
Note that the weights can be modified according to the importance of each category.

Lastly, a score was computed for each locality as the weighted sum of the number of venues in each of the 4 categories (school, university, office, burger joints).

4- Execution and Results

In this section, the results have been obtained via map acquired.

Localities of Bogotá, Colombia

	Localidades	Latitude	Longitude
0	Usaquén	4.694969	-74.031093
1	Chapinero	4.645377	-74.061943
2	Santa Fe (Bogotá)	4.602204	-74.078837
3	San Cristóbal (Bogotá)	4.548658	-74.047473
4	Usme	4.411136	-74.129108
5	Tunjuelito	4.562204	-74.127647
6	Bosa (Bogotá)	4.625492	-74.200280
7	Kennedy (Bogotá)	4.629682	-74.149935
8	Fontibón	4.673327	-74.144732
9	Engativá	4.708695	-74.109643
10	Suba	4.761197	-74.082518
11	Barrios Unidos (Bogotá)	4.669679	-74.075483
12	Teusaquillo	4.646410	-74.085441
13	Los Mártires	4.608375	-74.086538
14	Antonio Nariño (Bogotá)	4.589429	-74.099568
15	Puente Aranda	4.616071	-74.105433
16	La Candelaria	4.596515	-74.073492
17	Rafael Uribe Uribe (Bogotá)	4.573319	-74.119473
18	Ciudad Bolívar (Bogotá)	4.568300	-74.154185
19	Sumapaz (Bogotá)	4.097379	-74.342307

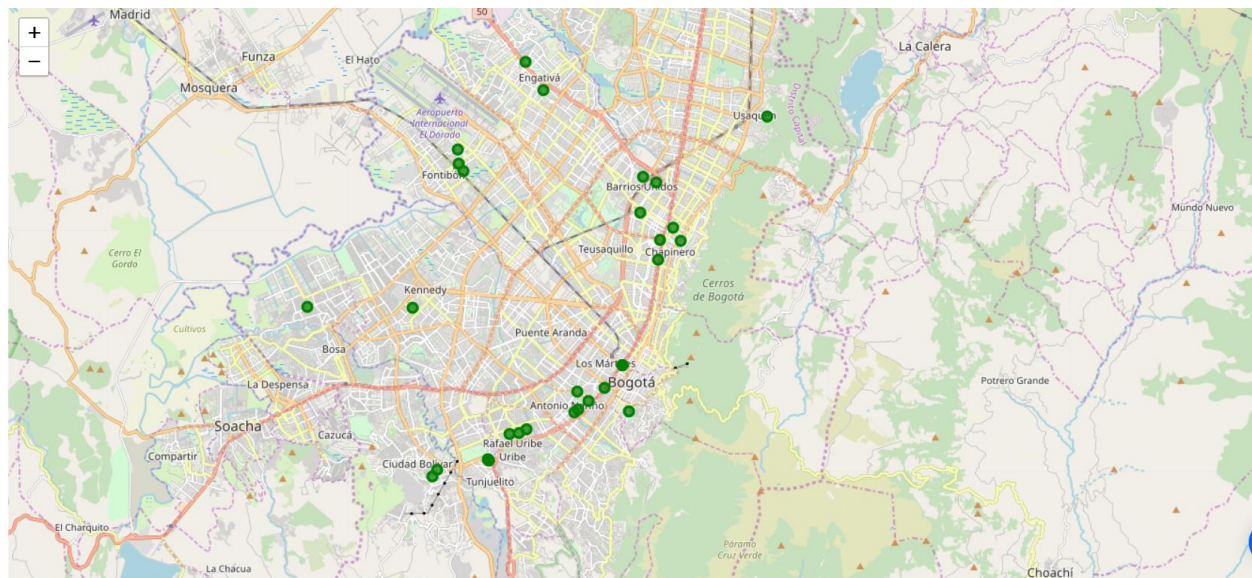# Burger Joints in Bogotá Localities
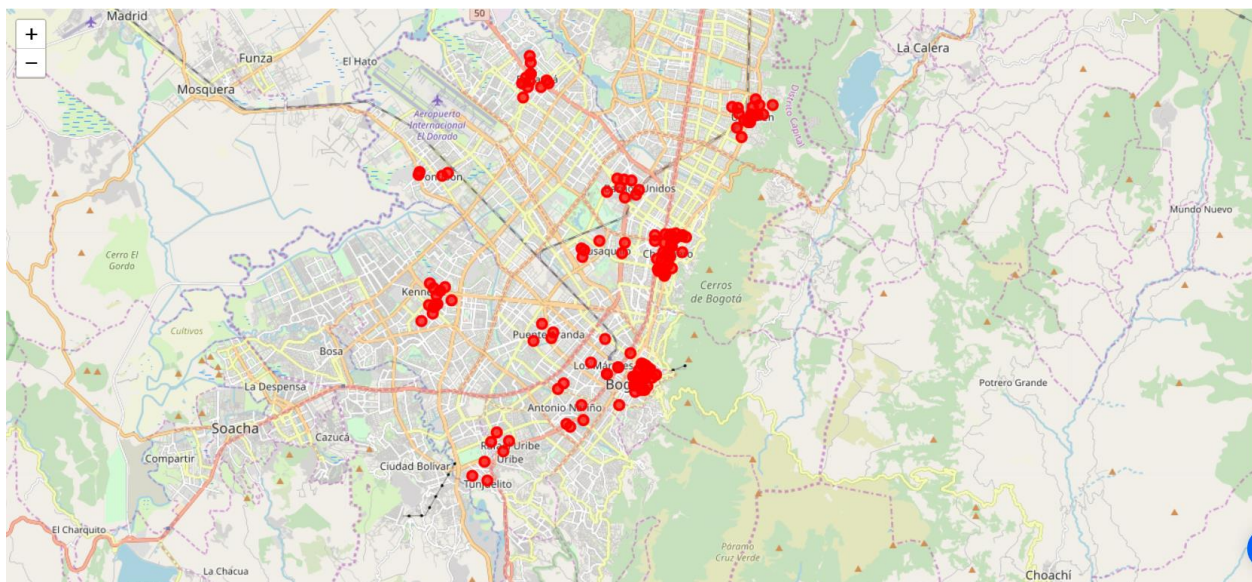
Universities in Bogotá Localities
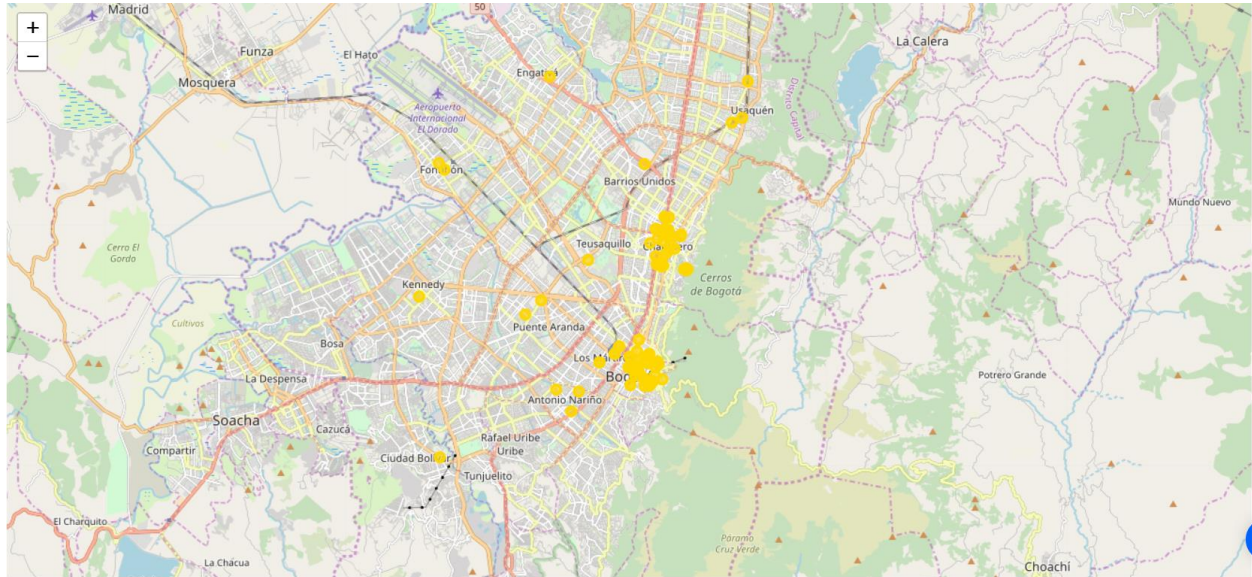
High Schools in Bogotá Localities

Out[55]:

	Localidad	Score
16	La Candelaria	131.5
1	Chapinero	115.0
2	Santa Fe (Bogotá)	112.0
14	Antonio Nariño (Bogotá)	98.5
8	Fontibón	98.0
13	Los Mártires	97.5
15	Puente Aranda	97.0
17	Rafael Uribe Uribe (Bogotá)	95.0
11	Barrios Unidos (Bogotá)	94.5
12	Teusaquillo	89.5
5	Tunjuelito	89.0
9	Engativá	87.5
0	Usaquén	82.5
7	Kennedy (Bogotá)	77.5
10	Suba	52.0
18	Ciudad Bolívar (Bogotá)	47.5
6	Bosa (Bogotá)	45.0
4	Usme	0.0
3	San Cristóbal (Bogotá)	0.0
19	Sumapaz (Bogotá)	0.0

Best Place for the Burger Joint in Bogotá is "La Candelaria"
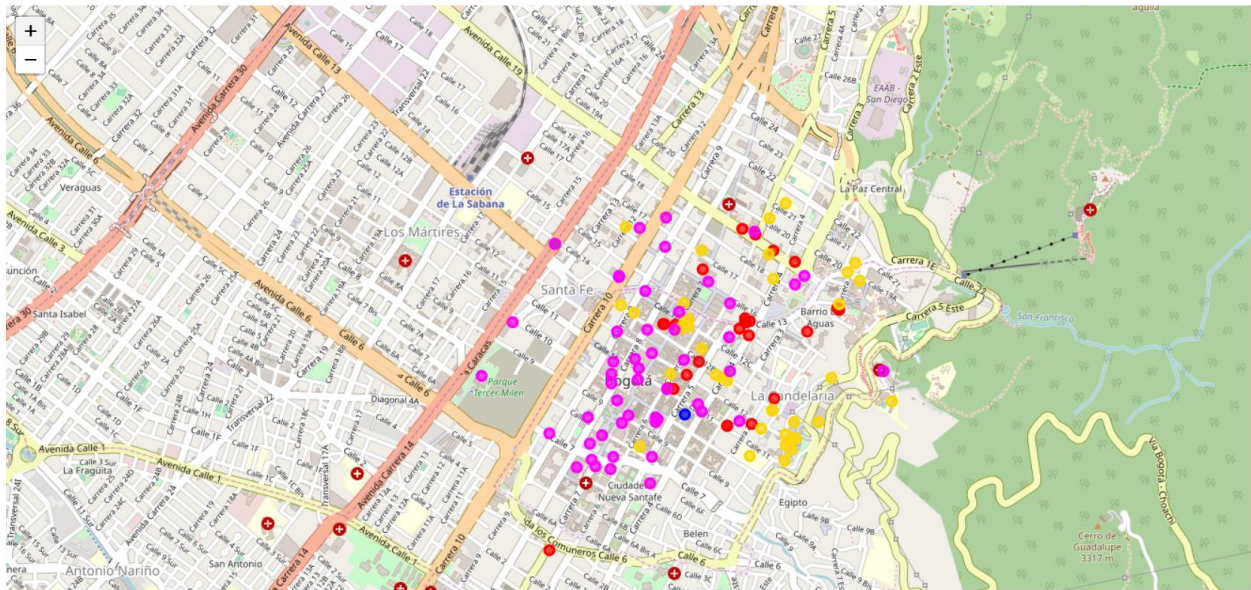
As a result, "La Candelaria" with the score od 109.5 is the most appropriate spot for this job. "Antonio Nariño" and "Fontibon" with 101 and 99 are the second and third options that could be suitable, respectively. In fact, these options maximize the number of potential customers from offices and universities and at the same time have not too large competence.

5- Recommendation

The following analysis can be improved with following extensions:

➤ Consider more categories. For example, like "Night life" which is also a good source for customers. But also like "Restaurants", which even if not burger joints may be some concurrence if too many.

➤ In the Locality itself, it can also be computed the distance between all the venues in order to find a place with the greatest number of potential customers.

➤ Using smaller geographical areas like Neighborhoods could improve the accuracy for the scores.

6- Acknowledgment

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course. I feel this Capstone project presented me a great opportunity to practice and apply the Data

Science tools and methodologies learned. Indeed, I have created a good project that I can present as an example to show my potential. I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.