Capstone project – Final report

Jatin Ajmera for Coursera capstone project

Introduction

- A customer, Marco, wants to open a new burger joint in Bogota, Colombia
- Due to Bogota's high diversity and very large size, he asked me for help in order to find the best spot to open a burger joint
- Bogota has 20 different localities-Districts and we aim to find the best one
- We need to find a location with good amount of customers and low amount of competition

Data

- To help Marco in his search we need to access the following data
- The localities of Bogota, Colombia from Wikipedia

  https://es.wikipedia.org/wiki/Anexo:Localidades_de_Bogot%C3%A1

- The coordinates (Latitude, Longitude) of these localities of Bogota from open street map APIs
- From foursquare we will need following venues data
    - The burger joint ventures of the localities
    - The offices venues of the localities
    - The high school 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

Methodology

- For each locality, all office, school, universities and burger joints venues data have been collected from Foursquare.
- Then for each locality, the sums of the office, school, universities and burger joints were calculated
- For each of this 4 categories, a weight has been defined according to what Marco considers most important
    - Burger joints have been weighted as -1, since we want to avoid concurrence
    - Schools have been weighted as 1, since students are good customers
    - Universities have been weighted as 1.5, since students are good customers
    - Offices have been weighted as 2, since employees are even better customers
- Note that weights can be modified according to 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, Offices, Burger joints)

Localities of Bogota, Colombia
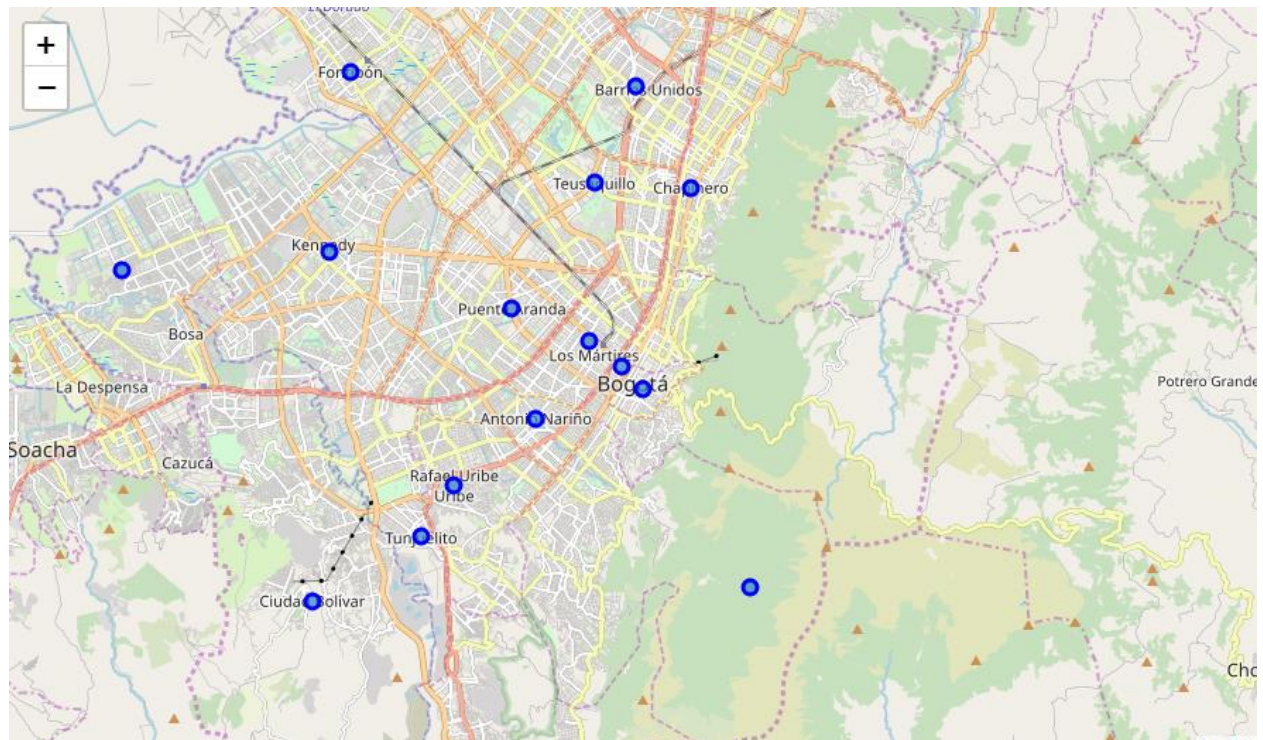
	Localidades	Latitude	Longitude
0	Usaquén	4.694969	-74.031093
1	Chapinero	4.645377	-74.061943
2	Santa Fe (Bogotá)	4.602204	-74.078837
3	San Cristóbal (Bogotá)	4.548658	-74.047473
4	Usme	4.411136	-74.129108
5	Tunjuelito	4.561049	-74.127523
6	Bosa (Bogotá)	4.625492	-74.200280
7	Kennedy (Bogotá)	4.629682	-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)	4.545440	-74.154000
19	Sumapaz (Bogotá)	4.097379	-74.342307

Burger joints in Bogota localities

High school in Bogota localities

Universities in Bogota localities
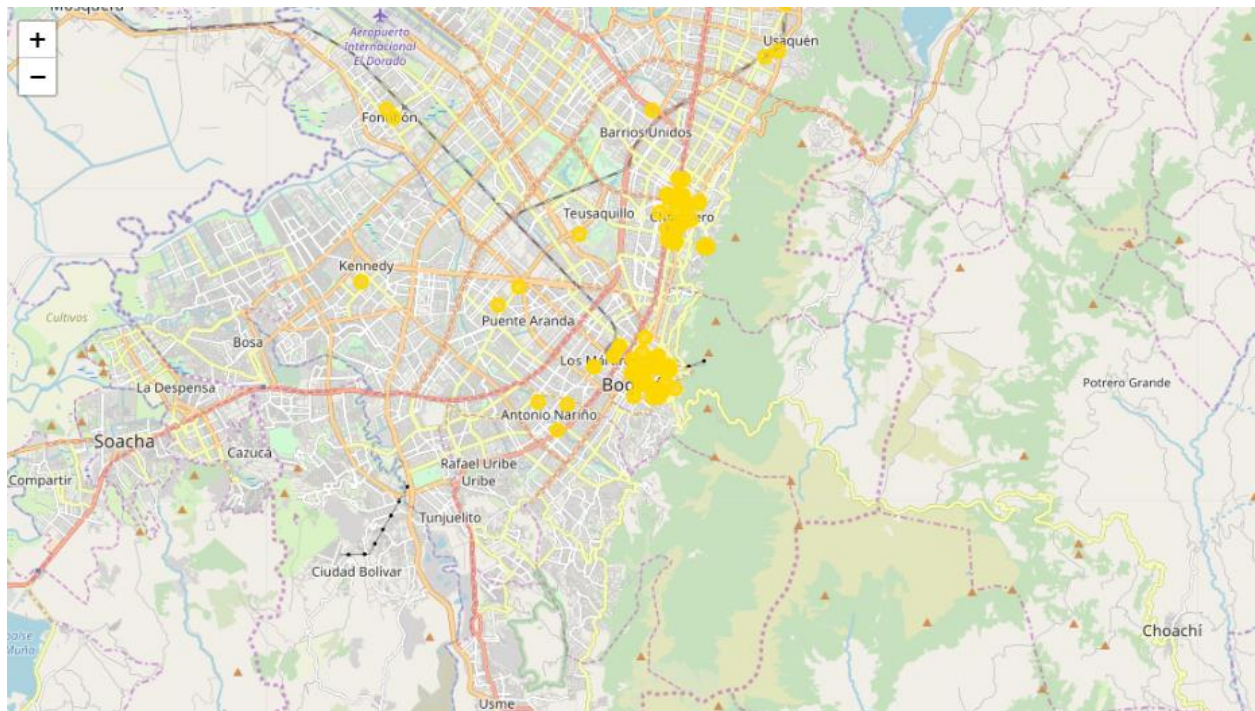
Offices in Bogota localities
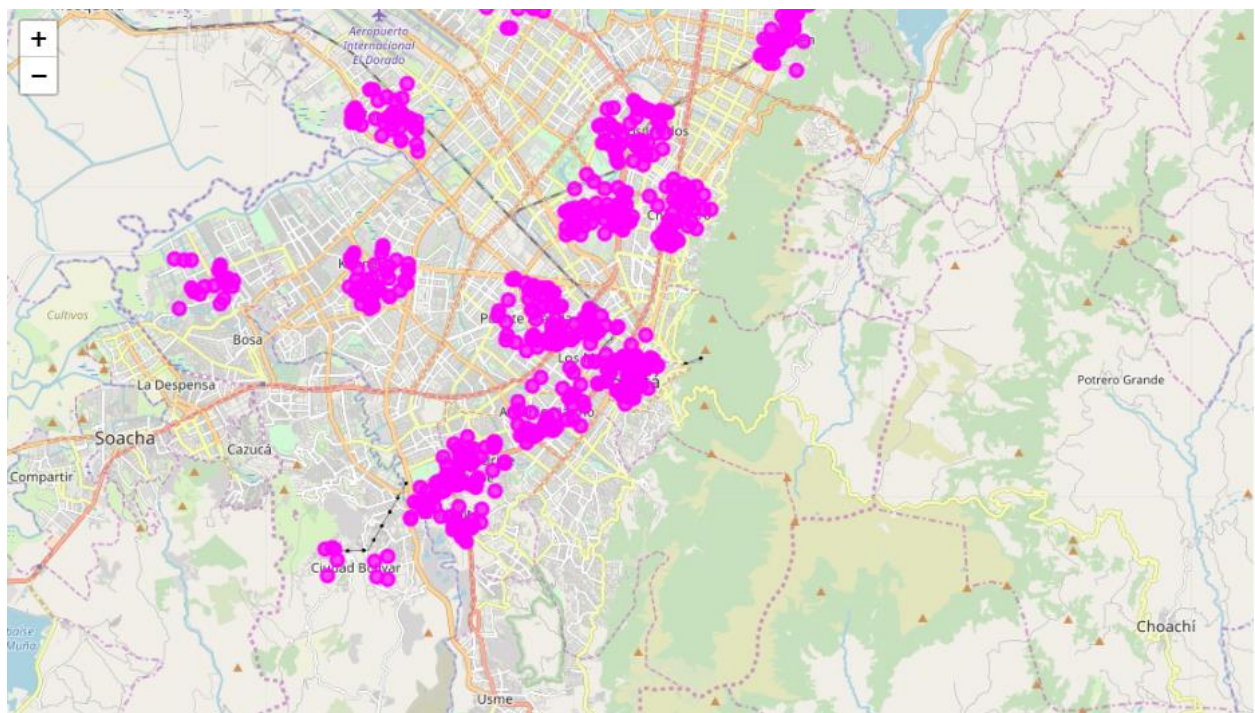
Results

	Localidad	Score
16	La Candelaria	131.5
1	Chapinero	116.5
2	Santa Fe (Bogotá)	112.0
17	Rafael Uribe Uribe (Bogotá)	95.0
15	Puente Aranda	95.0
14	Antonio Nariño (Bogotá)	94.5
12	Teusaquillo	94.5
8	Fontibón	92.0
11	Barrios Unidos (Bogotá)	91.5
13	Los Mártires	87.5
9	Engativá	86.5
5	Tunjuelito	85.0
0	Usaquén	84.5
7	Kennedy (Bogotá)	73.5
10	Suba	52.0
6	Bosa (Bogotá)	45.0
18	Ciudad Bolívar (Bogotá)	18.0
4	Usme	0.0
3	San Cristóbal (Bogotá)	0.0
19	Sumapaz (Bogotá)	0.0

- The locality with best score is "La Candelaria" with 131.5, being the best option
- Follows closely "Chapinero" with 116.5

These options maximizes the number of potential customers from offices and universities and at the same time have not too large competence.

Best place for Burger joint in Bogota is "La Candelaria"
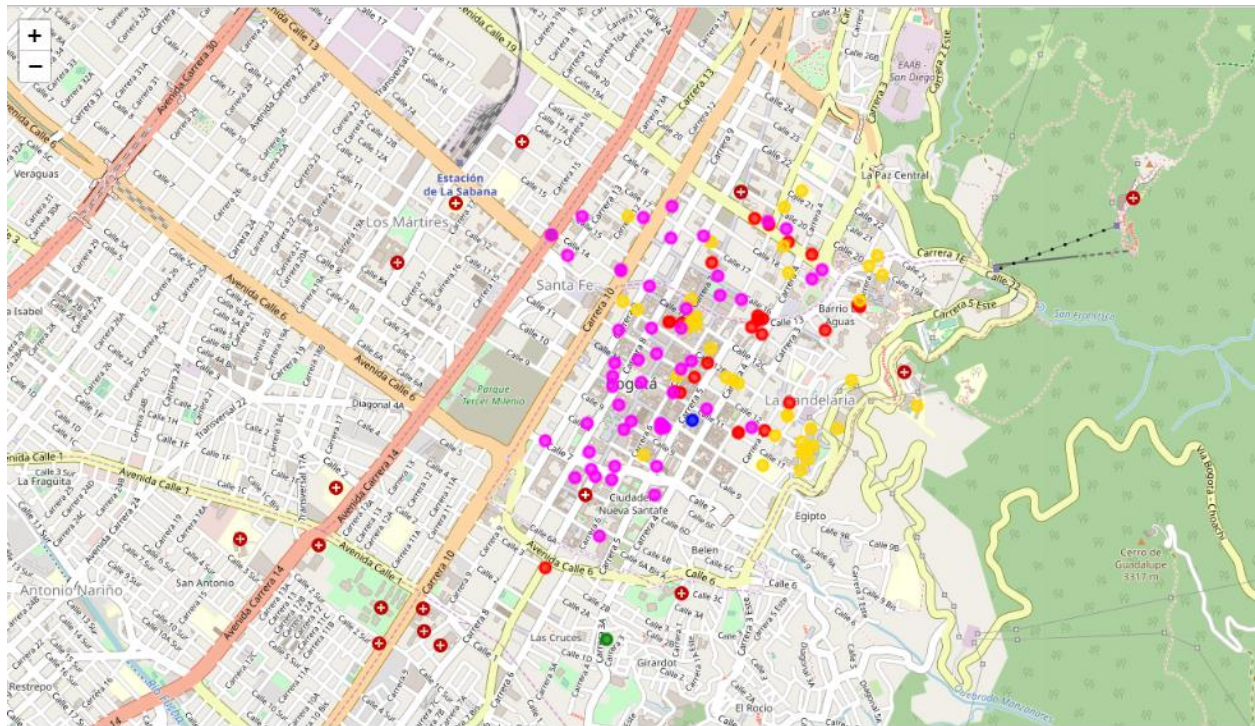
Recommendation

The following analysis can be improved with following extensions:

- Consider more categories, For example like "Night life" which is good source for customers. But also like "Restaurants", which even if not burger joints maybe 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 most number of potential customers.
- Using smaller geographical areas like neighbourhoods could improve the accuracy for the scores.