Capstone Project – Report

Introduction and Business Problem

Nowadays people want to star their own businesses and dream for become succesfull in their trip into the business world, also there are a lot of options and opportunities for doing that. We have a lot information from different sources that helps us to analyse and decide where we can open a new business and when we are going to get profit. In such way, Abigail wants to open a fast food restaurant in some Locality (District) of Lima Metropolitana (Lima and Callao), Peru and she is asking our help for finding the best Locality to start her business.

In order to open Abigail's restaurant, we will have to use some information about the Localities in Lima Metropolitana.

We must be sure that the place where the restaurant is going to be open will have enough customers and also that there aren't a large ammount of restaurants near to that place. In addition we have to consider the kind of customers that are near to the place Abigail choose in order to decide if the place will be profitable.

Data

To help Abigail in her search we will need to access following data:

- The districts of Lima Metropolitana, Peru from Wikipedia: https://es.wikipedia.org/wiki/Anexo:Distritos_de_Lima

- The coordinates (latitude, longitude) ot these districts of Lima from Open Street Map APIs

- From Foursquare we will need the following venues data:
  - the restaurants 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 new fast food restaurant.

Methodology

First, we collected the data of the districts in Lima Metropolitana and got the latitude and longitude for each one.

Table 1. Localities in Lima Metropolitana

	District	LocationCode	Postcode
0	Ancón	150102	02
1	Ate	150103	03
2	Barranco	150104	04
3	Breña	150105	05
4	Carabayllo	150106	06
5	Chaclacayo	150107	08
6	Chorrillos	150108	09
7	Cieneguilla	150109	40
8	Comas	150110	07
9	El Agustino	150111	10
10	Independencia	150112	28
11	Jesús María	150113	11
12	La Molina	150114	12
13	La Victoria	150115	13
14	Lima	150101	01
15	Lince	150116	14
16	Los Olivos	150117	39
17	Lurigancho	150118	15
18	Lurín	150119	16
19	Magdalena del Mar	150120	17
20	Miraflores	150122	18
21	Pachacamac	150123	19
22	Pucusana	150124	20
23	Pueblo Libre	150121	21
24	Puente Piedra	150125	22
25	Punta Hermosa	150126	24
26	Punta Negra	150127	23
27	Rímac	150128	25
28	San Bartolo	150129	26
29	San Borja	150130	41
30	San Isidro	150131	27
31	San Juan de Lurigancho	150132	36
32	San Juan de Miraflores	150133	29
33	San Luis	150134	30
34	San Martín de Porres	150135	31
35	San Miguel	150136	32
36	Santa Anita	150137	43
37	Santa María del Mar	150138	37
38	Santa Rosa	150139	38
39	Santiago de Surco	150140	33
40	Surquillo	150141	34
41	Villa El Salvador	150142	42
42	Villa María del Triunfo	150143	35
43	Bellavista	070102	Callao 2
44	Callao	070101	Callao 1
45	Carmen de La Legua-Reynoso	070103	Callao 3
46	La Perla	070104	Callao 4
47	La Punta	070105	Callao 5
48	Ventanilla	070106	Callao 6
49	Mi Perú	070107	Callao 7

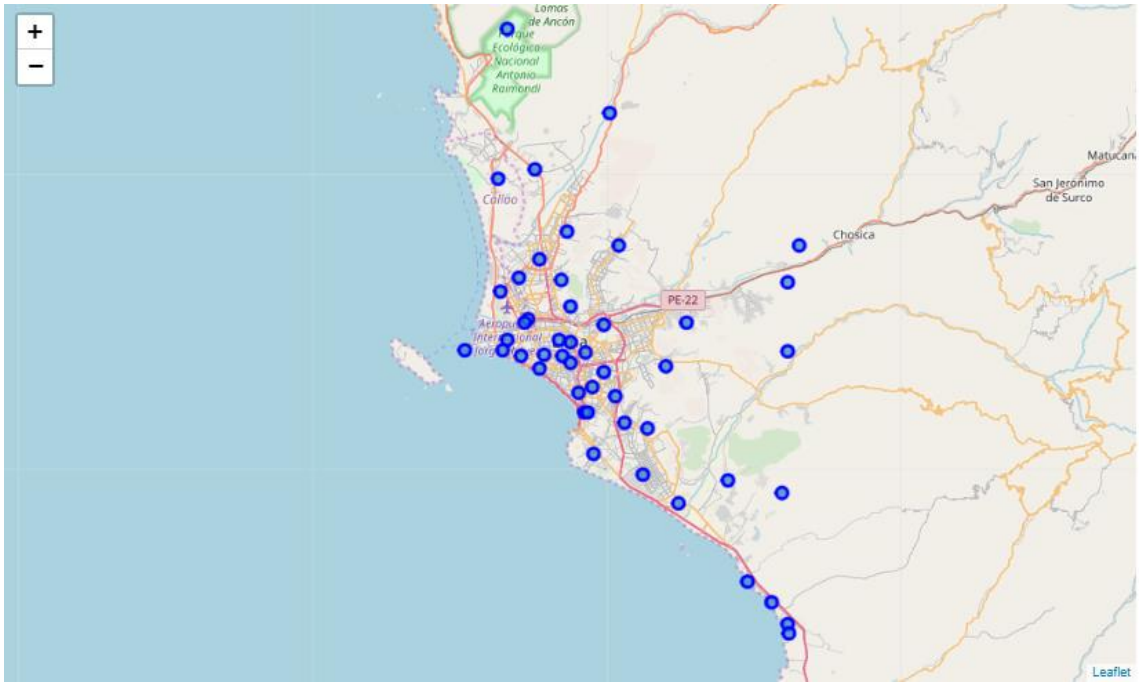

Figure 1. Localities in Lima Metropolitana

After that, we got the data for each locality, all offices, schools, universities and fast food restaurants venues that have been collected from Foursquare.

Then for each locality, we computed the sum of offices, schools, universities and fast food restaurants.

For get the better to place to set the new fast food restaurant we had to consider some variables and put a weight for each one that represents the impact of these variables in our choise. Thus, we had the following:

1. **weight_fast_food = -1** negative value, because we want a place that does not have enough fast food restaurants.
2. **weight_high = 1** positive value, because students of high school are often a good customer.
3. **weight_uni = 1.5** a more positive value, because university students are regularly a good customer.
4. **weight_office = 2** a very positive value, because employees are even better customer.

It is important to notice that each weights could be modified according to the importance or relevance for each one.

Finally, we computed a score for each locality as the weighted sum of the number of venues in each of the four categories: schools, universities, offices and fast food restaurants.

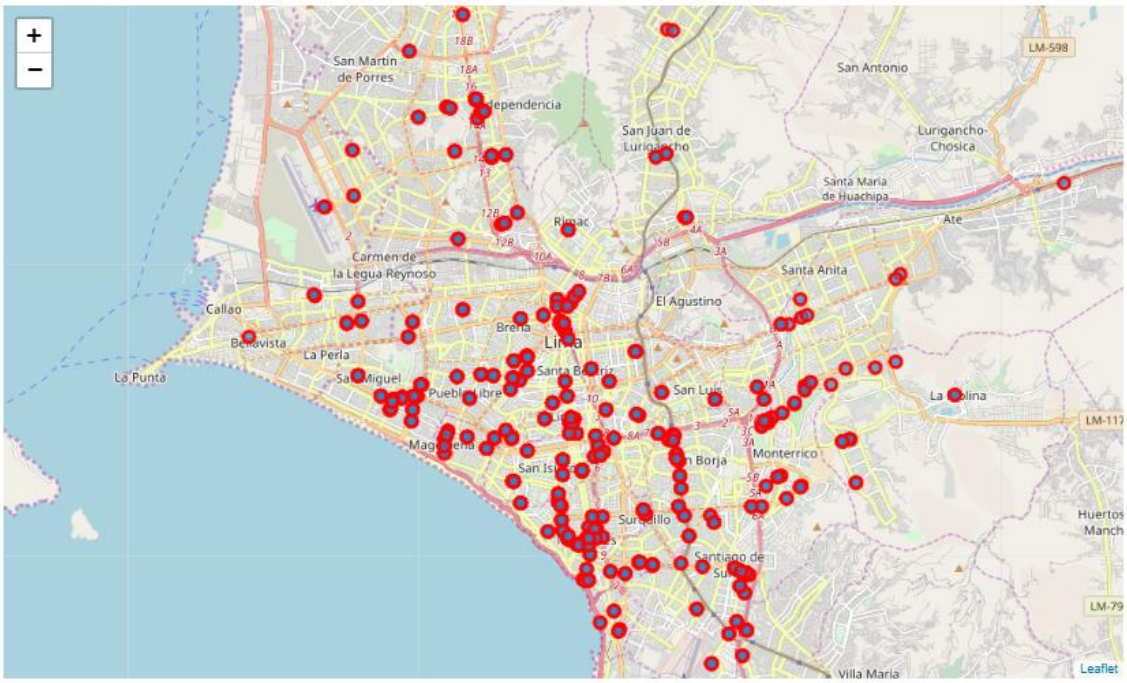

Figure 2. Fast Food Restaurants in Lima Metropolitana

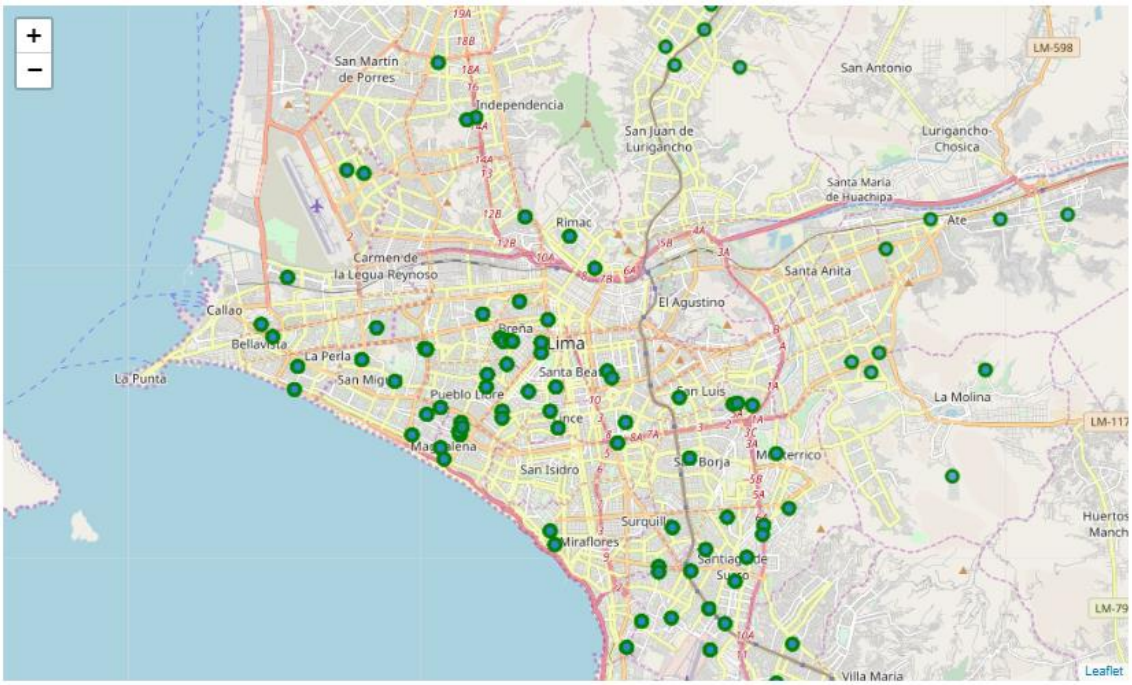

Figure 3. High Schools in Lima Metropolitana

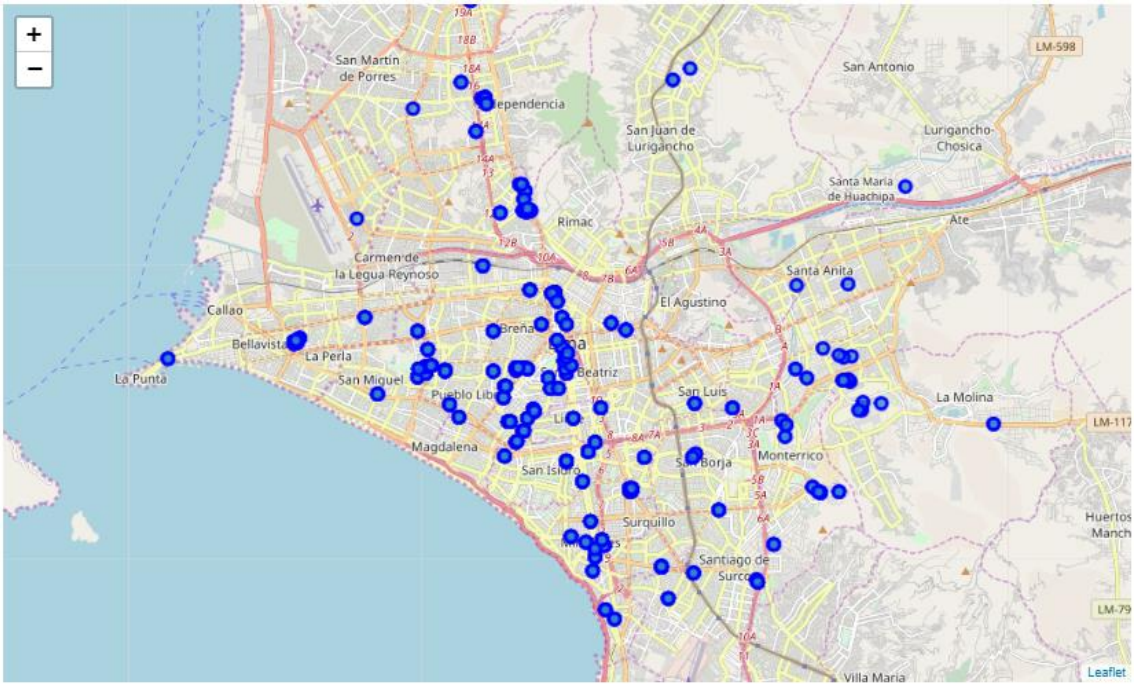

Figure 4. Universities in Lima Metropolitana

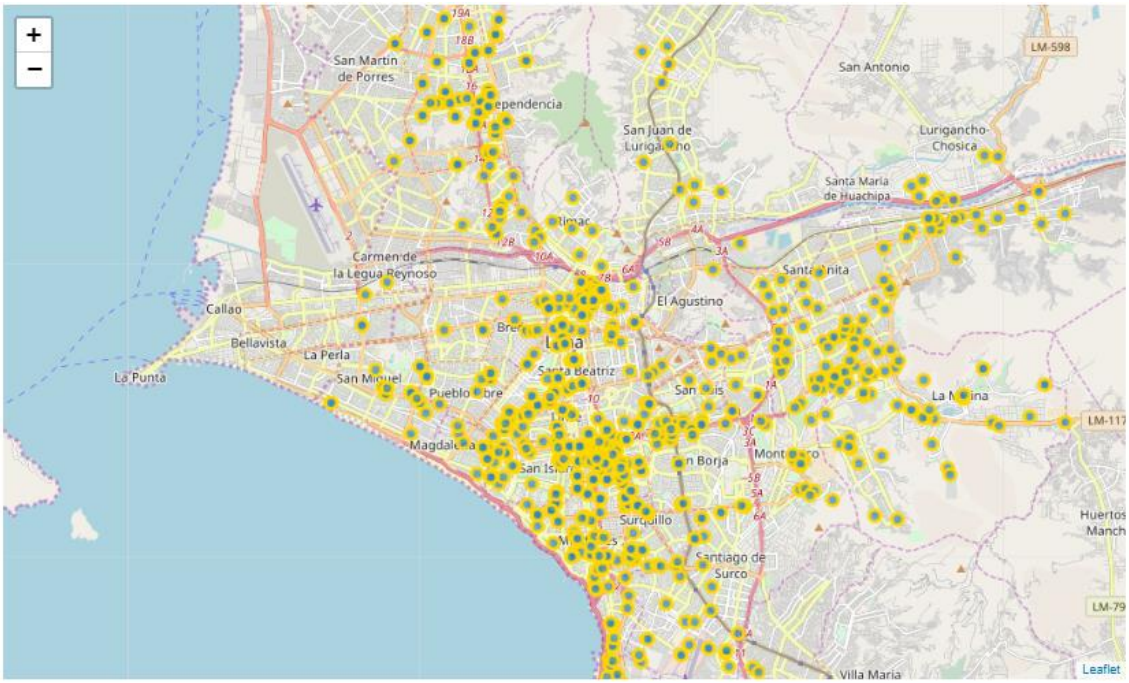

Figure 5. Offices in Lima Metropolitana

Table 2. Sum of venues for each locality in Lima Metropolitana

	District	LocationCode	Postcode	Latitude	Longitude	Fast Food Restaurants	High Schools	Universities	Offices
0	Ancón	150102	02	-11.696554	-77.111655	0.0	0.0	0.0	0.0
1	Ate	150103	03	-12.038728	-76.896873	4.0	5.0	1.0	36.0
2	Barranco	150104	04	-12.143959	-77.020268	81.0	18.0	29.0	100.0
3	Breña	150105	05	-12.059700	-77.050119	100.0	38.0	76.0	100.0
4	Carabayllo	150106	06	-11.794993	-76.989292	0.0	0.0	0.0	7.0
5	Chaclacayo	150107	08	-11.992479	-76.776176	0.0	3.0	0.0	9.0
6	Chorrillos	150108	09	-12.192350	-77.008962	16.0	7.0	8.0	86.0
7	Cieneguilla	150109	40	-12.073167	-76.777071	0.0	0.0	0.0	5.0
8	Comas	150110	07	-11.932861	-77.040674	6.0	5.0	4.0	29.0
9	El Agustino	150111	10	-12.042052	-76.995714	40.0	8.0	14.0	100.0
10	Independencia	150112	28	-11.989308	-77.047331	27.0	6.0	21.0	83.0
11	Jesús María	150113	11	-12.078186	-77.046412	100.0	38.0	77.0	100.0
12	La Molina	150114	12	-12.090177	-76.922338	17.0	5.0	22.0	84.0
13	La Victoria	150115	13	-12.073994	-77.018197	100.0	28.0	66.0	100.0
14	Lima	150101	01	-12.062107	-77.036526	92.0	33.0	76.0	100.0
15	Lince	150116	14	-12.086568	-77.036647	100.0	35.0	81.0	100.0
16	Los Olivos	150117	39	-11.965985	-77.073071	24.0	5.0	11.0	60.0
17	Lurigancho	150118	15	-11.948832	-76.762701	0.0	4.0	0.0	11.0
18	Lurín	150119	16	-12.238050	-76.783863	0.0	0.0	0.0	5.0
19	Magdalena del Mar	150120	17	-12.092369	-77.073309	75.0	32.0	54.0	100.0
20	Miraflores	150122	18	-12.121498	-77.025906	100.0	22.0	42.0	100.0
21	Pachacamac	150123	19	-12.250682	-76.906646	3.0	4.0	4.0	21.0
22	Pucusana	150124	20	-12.482092	-76.797453	0.0	0.0	0.0	5.0
23	Pueblo Libre	150121	21	-12.076639	-77.067858	100.0	34.0	67.0	100.0
24	Puente Piedra	150125	22	-11.860845	-77.078938	2.0	3.0	0.0	15.0
25	Punta Hermosa	150126	24	-12.341667	-76.825278	1.0	2.0	0.0	7.0
26	Punta Negra	150127	23	-12.365558	-76.795191	1.0	3.0	0.0	8.0
27	Rímac	150128	25	-12.020304	-77.035463	44.0	11.0	31.0	100.0
28	San Bartolo	150129	26	-12.391579	-76.776192	1.0	2.0	0.0	6.0
29	San Borja	150130	41	-12.096452	-76.995690	100.0	23.0	49.0	100.0
30	San Isidro	150131	27	-11.468296	-77.212225	2.0	4.0	0.0	4.0
31	San Juan de Lurigancho	150132	36	-11.948751	-76.977911	2.0	6.0	2.0	15.0
32	San Juan de Miraflores	150133	29	-12.155852	-76.972129	31.0	17.0	9.0	100.0
33	San Luis	150134	30	-11.031128	-77.605457	0.0	0.0	1.0	5.0
34	San Martín de Porres	150135	31	-11.986759	-77.097655	24.0	6.0	10.0	99.0
35	San Miguel	150136	32	-12.078656	-77.095283	51.0	28.0	37.0	100.0
36	Santa Anita	150137	43	-12.223383	-76.847707	0.0	3.0	5.0	12.0
37	Santa María del Mar	150138	37	-12.401403	-76.775465	1.0	2.0	0.0	6.0
38	Santa Rosa	150139	38	-12.035851	-77.086616	41.0	19.0	35.0	100.0
39	Santiago de Surco	150140	33	-12.125105	-76.981919	71.0	21.0	34.0	100.0
40	Surquillo	150141	34	-12.114198	-77.010475	100.0	30.0	40.0	100.0
41	Villa El Salvador	150142	42	-12.216894	-76.949150	8.0	9.0	4.0	44.0
42	Villa María del Triunfo	150143	35	-12.162520	-76.944275	6.0	3.0	1.0	55.0
43	Bellavista	070102	Callao 2	-12.059736	-77.111161	31.0	13.0	22.0	100.0
44	Callao	070101	Callao 1	-12.003654	-77.119244	7.0	3.0	2.0	65.0
45	Carmen de La Legua-Reynoso	070103	Callao 3	-12.039379	-77.090561	39.0	20.0	31.0	100.0
46	La Perla	070104	Callao 4	-12.071343	-77.117564	26.0	13.0	21.0	100.0
47	La Punta	070105	Callao 5	-12.070896	-77.162264	5.0	4.0	7.0	48.0
48	Ventanilla	070106	Callao 6	-11.871983	-77.122204	4.0	4.0	1.0	19.0
49	Mi Perú	070107	Callao 7	-12.143956	-77.015792	81.0	18.0	29.0	100.0

Result

We computed a score for each locality in Lima Metropolitana based on the following equation:

```
df_score['Score'] = df_data['Fast Food Restaurants'] * weight_fast_food + df_data['High Schools'] * weight_high + df_data['Universities'] * weight_uni + df_data['Offices'] * weight_office
```

Considering these values:

```
weight_fast_food = -1
weight_high = 1
weight_uni = 1.5
weight_office = 2
```

We noticed that the localities of the lowest score were:

- Ancon with 0
- Cieneguilla with 10
- Lurín with 10
- Pucusana with 10
- San Isidro with 10

We also noticed tha the localities of the highest score were:

- Lince with 256.5
- Lima with 255.0
- Jesús María with 253.5
- Breña with 252.0
- Magdalena del Mar with 238.0

Table 3. Score for each locality in Lima Metropolitana

	District	Score
0	Ancón	0.0
1	Ate	74.5
2	Barranco	180.5
3	Breña	252.0
4	Carabayllo	14.0
5	Chaclacayo	21.0
6	Chorrillos	175.0
7	Cieneguilla	10.0
8	Comas	63.0
9	El Agustino	189.0
10	Independencia	176.5
11	Jesús María	253.5
12	La Molina	189.0
13	La Victoria	227.0
14	Lima	255.0
15	Lince	256.5
16	Los Olivos	117.5
17	Lurigancho	26.0
18	Lurín	10.0
19	Magdalena del Mar	238.0
20	Miraflores	185.0
21	Pachacamac	49.0
22	Pucusana	10.0
23	Pueblo Libre	234.5
24	Puente Piedra	31.0
25	Punta Hermosa	15.0
26	Punta Negra	18.0
27	Rímac	213.5
28	San Bartolo	13.0
29	San Borja	196.5
30	San Isidro	10.0
31	San Juan de Lurigancho	37.0
32	San Juan de Miraflores	199.5
33	San Luis	11.5
34	San Martín de Porres	195.0
35	San Miguel	232.5
36	Santa Anita	34.5
37	Santa María del Mar	13.0
38	Santa Rosa	230.5
39	Santiago de Surco	201.0
40	Surquillo	190.0
41	Villa El Salvador	95.0
42	Villa María del Triunfo	108.5
43	Bellavista	215.0
44	Callao	129.0
45	Carmen de La Legua-Reynoso	227.5
46	La Perla	218.5
47	La Punta	105.5
48	Ventanilla	39.5
49	Mi Perú	180.5

Discussion

The above analysis can be improved with following extensions:

- Consider more categories or variables. For example: "Night life" which is also a good source for customers. Another instance is "Restaurants" in general, which even if they are not Fast Food Restaurants maybe some concurrence are higher.

- 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 Neighborhoods could improve the accuracy of scores.

Conclusion

In order to find the best place to open Abigail's Fast Food Restaurant, we concluded that the most appropriate locality is Lince with a score of 256.5.

We also concluded that our accuracy could be better by considering another variables or categories like “Night life” or other types of restaurants and by computing the analysis in neighborhoods.