

IBM Data Science

Capstone Project

***Find myself the best neighborhood
to live in Barcelona***

Yu Deng

2020.3

1. Introduction

This report is for the capstone course of IBM Data Science, a professional certification series provided by Coursera.com. In this section, I am going to utilize many data science tools taught in all of 9 courses and produce a data analysis report by myself.

I am a business school student in China. This autumn I will exchange in one of the best business school in Europe, ESADE, Barcelona, and stay there for about 3-4 months. The school doesn't offer on-campus accommodation, so I have to rent an apartment together with my schoolmates. It's a tough task for me to decide where to live. However, the capstone course brings us Foursquare API and the methods to cluster neighborhoods based on their similarity. I decide to deploy methods alike to choose the best place to live.

My analysis goes through following processes:

- a) Pick out neighborhoods near my school, ESADE, as candidates.
- b) Use data from Foursquare to explore famous venues in selected neighborhoods.
- c) Cluster these neighborhoods base on their similarity with the help

of machine learning.

- d) Sort out the most common venues in each neighborhood, calculate their proportions.
- e) Transform my needs and interests into a vector which consists of scores on different venues. Multiply the neighborhoods dataframe by the vector to choose the best place to live.

2. Data

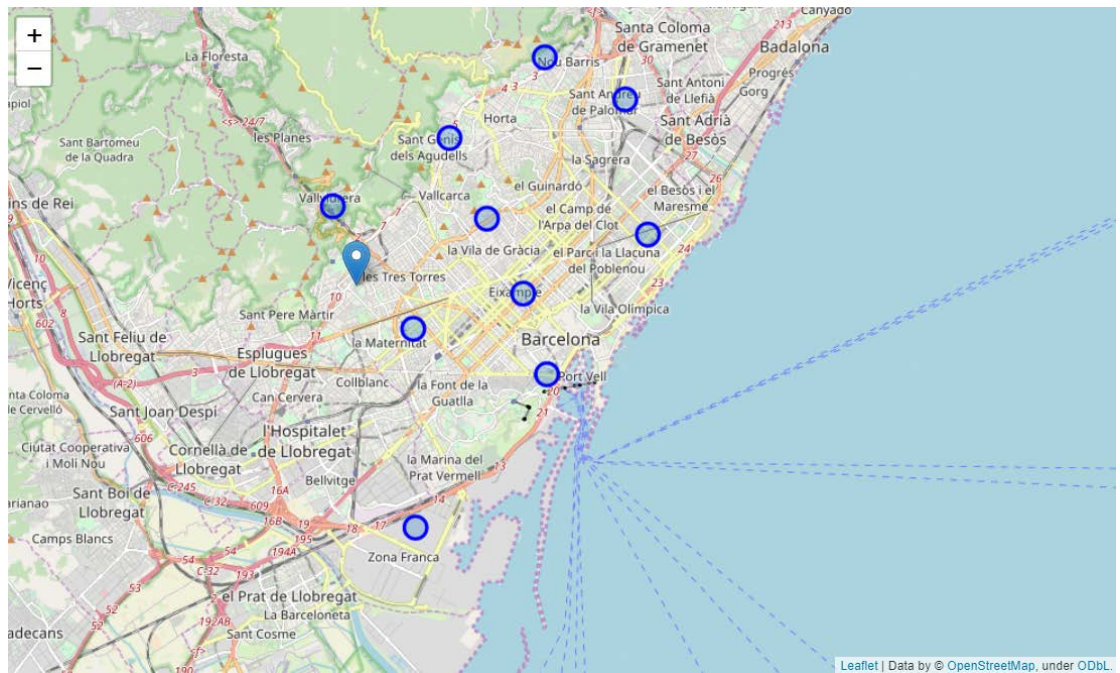
The data of my report consists of three part:

- a) The geospatial data of ESADE Business School.

The geographical coordinate of ESADE Business School are 41.3947771, 2.1150794468981124.

- b) The list of neighborhoods in Barcelona and their geospatial data.

	District	Density	Neighbourhoods	Latitude	Longitude
0	Ciutat Vella	24786	La Barceloneta, El Gòtic, El Raval, Sant Pere,...	41.374962	2.173265
1	Eixample	35586	L'Antiga Esquerra de l'Eixample, La Nova Esque...	41.393394	2.166085
2	Sants-Montjuïc	8321	La Bordeta, la Font de la Guatlla, Hostafrancs...	41.340234	2.133347
3	Les Corts	13584	les Corts, la Maternitat i Sant Ramon, Pedralbes	41.385244	2.132863
4	Sarrià-Sant Gervasi	6992	El Putget i Farró, Sarrià, Sant Gervasi - la B...	41.413043	2.108356
5	Gràcia	28660	Vila de Gràcia, el Camp d'en Grassot i Gràcia ...	41.410171	2.155136
6	Horta-Guinardó	14217	El Baix Guinardó, El Guinardó, Can Baró, El Ca...	41.428540	2.143597
7	Nou Barris	20520	Can Peguera, Canyelles, Ciutat Meridiana, La G...	41.446727	2.172565
8	Sant Andreu	21737	Baró de Viver, Bon Pastor, El Congrés i els In...	41.437439	2.196859
9	Sant Martí	20466	El Besòs i el Maresme, el Clot, El Camp de l'A...	41.406782	2.203655



- c) The geographical coordinates of the top 5 closest districts to ESADE Business School, which are **Les Corts**, **Sarrià-Sant Gervasi**, **Eixample**, **Gràcia** and **Horta-Guinardó**.

	District	Neighborhood	Latitude	Longitude
0	Eixample	L'Antiga Esquerra de l'Eixample	41.388765	2.156597
1	Eixample	La Nova Esquerra de l'Eixample	41.382816	2.149966
2	Eixample	Dreta de l'Eixample	41.394124	2.166471
3	Eixample	Fort Pienc	41.395925	2.182325
4	Eixample	Sagrada Família	41.403479	2.174410
5	Eixample	Sant Antoni	41.380578	2.160102
6	Les Corts	les Corts	41.385244	2.132863
7	Les Corts	la Maternitat i Sant Ramon	41.381975	2.115474
8	Les Corts	Pedralbes	41.390140	2.112218
9	Sarrià-Sant Gervasi	El Putget i Farró	41.407476	2.143283
10	Sarrià-Sant Gervasi	Sarrià	41.399373	2.121513
11	Sarrià-Sant Gervasi	Sant Gervasi - la Bonanova	41.409637	2.129637

.....

d) Information of venues in specific neighborhoods provided by Foursquare.com.

	District	Neighborhood	Venue	Venue Category
0	Eixample	L'Antiga Esquerra de l'Eixample	Roast Club Café	Coffee Shop
1	Eixample	L'Antiga Esquerra de l'Eixample	Foc i Oli	Sandwich Place
2	Eixample	L'Antiga Esquerra de l'Eixample	Tandem	Cocktail Bar
3	Eixample	L'Antiga Esquerra de l'Eixample	The Corner Hotel Barcelona	Hotel
4	Eixample	L'Antiga Esquerra de l'Eixample	Mikkeller Bar Barcelona	Beer Bar

.....

3. Methodology

In this part, I calculate the means of occurrence of each venue, to find out which venues are most common in these neighborhoods.

Neighborhood	Accessories Store	African Restaurant	American Restaurant	Animal Shelter	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery
0 Can Baró	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.01	0.0	0.00	0.01	0.00	0.050000
1 Dreta de l'Eixample	0.0	0.0	0.0	0.0	0.0	0.01	0.0	0.00	0.0	0.00	0.00	0.00	0.030000
2 El Baix Guinardó	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	0.0	0.00	0.01	0.00	0.050000
3 El Carmel	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.00	0.0	0.00	0.00	0.00	0.036585
4 El Guinardó	0.0	0.0	0.0	0.0	0.0	0.01	0.0	0.00	0.0	0.01	0.01	0.01	0.060000

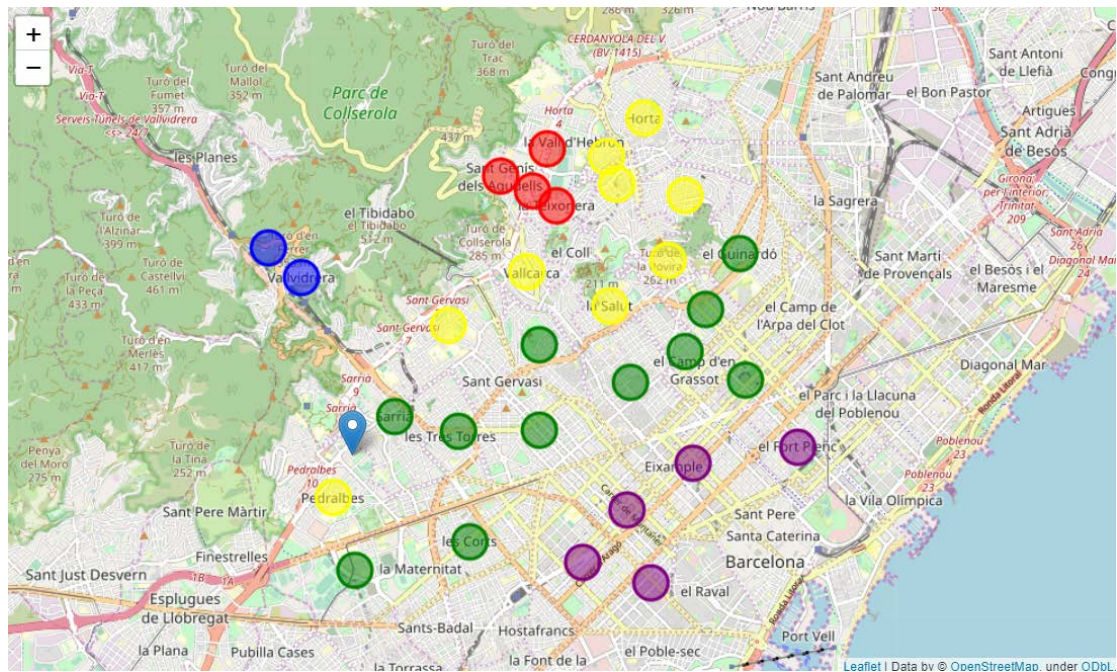
.....

K-means clustering in machine learning can help me easily classify the neighborhoods into designated number of categories base on their similarity. I set the number of clusters as 5 and come up with the results below. The map can visualize the results accordingly.

Neighborhood Cluster

0	Can Baró	4
1	Dreta de l'Eixample	3
2	El Baix Guinardó	1
3	El Carmel	4
4	El Guinardó	1

.....



Then, I sort out the top 10 most common venues in each neighborhood to figure out their patterns and characteristics.

Cluster 1

	Neighborhood	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
10	La Vall d'Hebron.	Park	Spanish Restaurant	Grocery Store	Bakery	Tennis Court	Metro Station	Tapas Restaurant	Plaza	Argentinian Restaurant	Chinese Restaurant
11	Montbau	Spanish Restaurant	Grocery Store	Restaurant	Plaza	Park	Soccer Field	Soccer Stadium	Hotel	Tennis Court	Chinese Restaurant
15	Sant Genís dels Agudells	Plaza	Grocery Store	Spanish Restaurant	Park	Soccer Stadium	Tapas Restaurant	Tennis Court	Argentinian Restaurant	Outdoor Sculpture	Café
28	la Teixonera	Park	Grocery Store	Spanish Restaurant	Plaza	Scenic Lookout	Tapas Restaurant	Tennis Court	Hotel	Chinese Restaurant	Mountain

Cluster 2

	Neighborhood	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
2	El Baix Guinardó	Restaurant	Tapas Restaurant	Bakery	Plaza	Bar	Italian Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Brewery	Hotel
4	El Guinardó	Restaurant	Tapas Restaurant	Bakery	Spanish Restaurant	Italian Restaurant	Café	Bar	Brewery	Hotel	Gym
5	El Putget i Farró	Bakery	Pizza Place	Plaza	Italian Restaurant	Wine Bar	Mediterranean Restaurant	Park	Japanese Restaurant	Café	Deli / Bodega
13	Sagrada Família	Italian Restaurant	Burger Joint	Hotel	Restaurant	Bar	Plaza	Pizza Place	Mexican Restaurant	Steakhouse	Sandwich Place
16	Sant Gervasi - Galvany	Restaurant	Spanish Restaurant	Mediterranean Restaurant	Pizza Place	Tapas Restaurant	Japanese Restaurant	Italian Restaurant	Deli / Bodega	French Restaurant	Wine Bar
18	Sarrià	Hotel	Mediterranean Restaurant	Bakery	Grocery Store	Café	Burger Joint	Japanese Restaurant	Park	Gym / Fitness Center	Restaurant
22	Vila de Gràcia	Tapas Restaurant	Plaza	Restaurant	Café	Cocktail Bar	Mediterranean Restaurant	Bar	Pizza Place	Wine Bar	Indie Movie Theater
23	el Camp d'en Grassot i Gràcia Nova	Tapas Restaurant	Plaza	Mediterranean Restaurant	Indie Movie Theater	Vegetarian / Vegan Restaurant	Restaurant	Bar	Bakery	Hotel	Japanese Restaurant
26	la Maternitat i Sant Ramon	Hotel	Pizza Place	Café	Spanish Restaurant	Bar	Bakery	Nightclub	Park	Coffee Shop	Mediterranean Restaurant
29	Iles Corts	Restaurant	Pizza Place	Sandwich Place	Café	Spanish Restaurant	Hotel	Burger Joint	Supermarket	Cocktail Bar	Asian Restaurant
30	Iles Tres Torres	Spanish Restaurant	Japanese Restaurant	Italian Restaurant	Mediterranean Restaurant	Café	Hotel	Restaurant	Bakery	Pizza Place	Cocktail Bar

Cluster 3

	Neighborhood	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
19	Tibidabo i les Planes	Building	BBQ Joint	Soccer Field	Fountain	National Park	General Entertainment	Scenic Lookout	Light Rail Station	Farmers Market	Tourist Information Center
21	Valldirera	Light Rail Station	Building	BBQ Joint	Soccer Field	Fountain	National Park	Scenic Lookout	Lake	Farmers Market	Restaurant

Cluster 4

	Neighborhood	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
1	Dreta de l'Eixample	Hotel	Tapas Restaurant	Boutique	Mediterranean Restaurant	Restaurant	Spanish Restaurant	Hostel	Bookstore	Clothing Store	Bakery
6	Fort Pienc	Hotel	Coffee Shop	Mediterranean Restaurant	Bistro	Hostel	Burger Joint	Spanish Restaurant	Restaurant	Tapas Restaurant	Breakfast Spot
8	L'Antiga Esquerra de l'Eixample	Hotel	Spanish Restaurant	Mediterranean Restaurant	Cocktail Bar	Tapas Restaurant	Japanese Restaurant	Beer Bar	Wine Bar	Pizza Place	Restaurant
9	La Nova Esquerra de l'Eixample	Hotel	Spanish Restaurant	Coffee Shop	Tapas Restaurant	Mediterranean Restaurant	Café	Japanese Restaurant	Korean Restaurant	Gym	Restaurant
14	Sant Antoni	Cocktail Bar	Tapas Restaurant	Coffee Shop	Mediterranean Restaurant	Hotel	Bar	Beer Bar	Japanese Restaurant	Spanish Restaurant	Wine Bar

Cluster 5

	Neighborhood	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
0	Can Baró	Bar	Spanish Restaurant	Park	Gym	Bakery	Grocery Store	Café	Plaza	Tapas Restaurant	Restaurant
3	El Carmel	Spanish Restaurant	Tapas Restaurant	Restaurant	Plaza	Bakery	Park	Mediterranean Restaurant	Metro Station	Pizza Place	Café
7	Horta	Spanish Restaurant	Tapas Restaurant	Grocery Store	Restaurant	Plaza	Mediterranean Restaurant	Park	Hostel	Hotel	Food & Drink Shop
12	Pedralbes	Hotel	Garden	Gym / Fitness Center	Mediterranean Restaurant	Park	Restaurant	Café	Bar	College Cafeteria	Spanish Restaurant
17	Sant Gervasi - la Bonanova	Bakery	Mediterranean Restaurant	Spanish Restaurant	Seafood Restaurant	Plaza	Restaurant	Café	Grocery Store	Bar	Park
20	Vallcarca i els Penitents	Park	Plaza	Café	Bakery	Hotel	Grocery Store	Supermarket	Spanish Restaurant	Breakfast Spot	Mediterranean Restaurant
24	la Clota	Spanish Restaurant	Plaza	Supermarket	Park	Restaurant	Metro Station	Soccer Field	Café	Sandwich Place	Food & Drink Shop
25	la Font d'en Fargues	Tapas Restaurant	Spanish Restaurant	Restaurant	Bar	Café	Grocery Store	Plaza	Supermarket	Italian Restaurant	Pizza Place
27	la Salut	Plaza	Park	Restaurant	Bar	Hotel	Tapas Restaurant	Café	Mediterranean Restaurant	Italian Restaurant	Scenic Lookout

According to the characteristics of each cluster, it is easy to figure out cluster 1 and 3 are holiday resorts. Besides, cluster 4 is too far away from ESADE Business School. Cluster 2 and 5 can be perfect final candidates, they include restaurants, cafés, bars, plaza, groceries and parks. All make them suitable neighborhoods for me to live.

The final candidate list includes 20 neighborhoods in cluster 2 and 5, and I sort out the top 20 common venue in these neighborhoods combined.

0	Can Baró	Restaurant	0.051725
2	El Baix Guinardó	Spanish Restaurant	0.046304
3	El Carmel	Tapas Restaurant	0.043235
4	El Guinardó	Plaza	0.041651
5	El Putget i Farró	Mediterranean Restaurant	0.038035
7	Horta	Café	0.037032
12	Pedralbes	Bakery	0.036278
13	Sagrada Família	Hotel	0.033610
16	Sant Gervasi - Galvany	Park	0.033396
17	Sant Gervasi - la Bonanova	Pizza Place	0.031454
18	Sarrià	Bar	0.030328
		Italian Restaurant	0.025904
		Grocery Store	0.020429
		Burger Joint	0.020082
		Sandwich Place	0.019351
		Supermarket	0.018693
		Japanese Restaurant	0.016000
		Gym	0.015786
		Coffee Shop	0.014592
		Gym / Fitness Center	0.014237

.....

Finally, I create the matrix of occurrence of venues, and also a vector which represents my preference to some specific venues by assigning a score range from 0 to 10 to each of them.

	Accessories Store	African Restaurant	American Restaurant	Animal Shelter	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	ARTS & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Basketball Stadium
0	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.010000	0.00	0.000000	0.010000	0.00	0.050000	0.070000	0.00
2	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.010000	0.00	0.050000	0.040000	0.00
3	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.036585	0.012195	0.00
4	0.00	0.00	0.000000	0.00	0.0	0.010000	0.00	0.000000	0.00	0.010000	0.010000	0.01	0.060000	0.030000	0.00
5	0.00	0.00	0.000000	0.00	0.0	0.000000	0.01	0.000000	0.01	0.000000	0.000000	0.00	0.090000	0.020000	0.00
7	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.025316	0.000000	0.00
12	0.00	0.00	0.018182	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.018182	0.00	0.000000	0.036364	0.00
13	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.030000	0.040000	0.00
16	0.01	0.01	0.000000	0.00	0.0	0.000000	0.02	0.000000	0.00	0.000000	0.000000	0.00	0.030000	0.000000	0.00
17	0.00	0.00	0.000000	0.01	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.070000	0.030000	0.00
18	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.070000	0.010000	0.00
20	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.010000	0.00	0.000000	0.000000	0.00	0.050000	0.040000	0.00
22	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.02	0.010000	0.000000	0.00	0.020000	0.050000	0.00
23	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.030000	0.040000	0.00
24	0.00	0.00	0.000000	0.00	0.0	0.018868	0.00	0.000000	0.00	0.000000	0.000000	0.00	0.018868	0.018868	0.00
25	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.011628	0.000000	0.00	0.011628	0.046512	0.00
26	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.000000	0.030000	0.00	0.030000	0.040000	0.01
27	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.013158	0.00	0.000000	0.000000	0.00	0.013158	0.052632	0.00
29	0.00	0.00	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.00	0.030000	0.000000	0.00	0.010000	0.010000	0.00
30	0.00	0.00	0.000000	0.00	0.0	0.000000	0.01	0.000000	0.00	0.020000	0.000000	0.00	0.030000	0.020000	0.00

```
pre_dict = {
    'Art Museum': 5,
    'Bakery': 6,
    'Bar': 3,
    'Cafeteria': 3,
    'Café': 3,
    'Chinese Restaurant': 8,
    'Convenience Store': 6,
    'Fast Food Restaurant': 6,
    'Food & Drink Shop': 6,
    'Football Stadium': 3,
    'Fruit & Vegetable Store': 8,
    'Grocery Store': 8,
    'Japanese Restaurant': 4,
    'Market': 5,
    'Mediterranean Restaurant': 3,
    'Museum': 5,
    'Restaurant': 4,
    'Soccer Field': 3,
    'Soccer Stadium': 3,
    'Spanish Restaurant': 4,
    'Supermarket': 5,
    'Tapas Restaurant': 4
}
```

	Accessories Store	African Restaurant	American Restaurant	Animal Shelter	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	ARTS & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Basketball Stadium
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	6.0	3.0	0.0

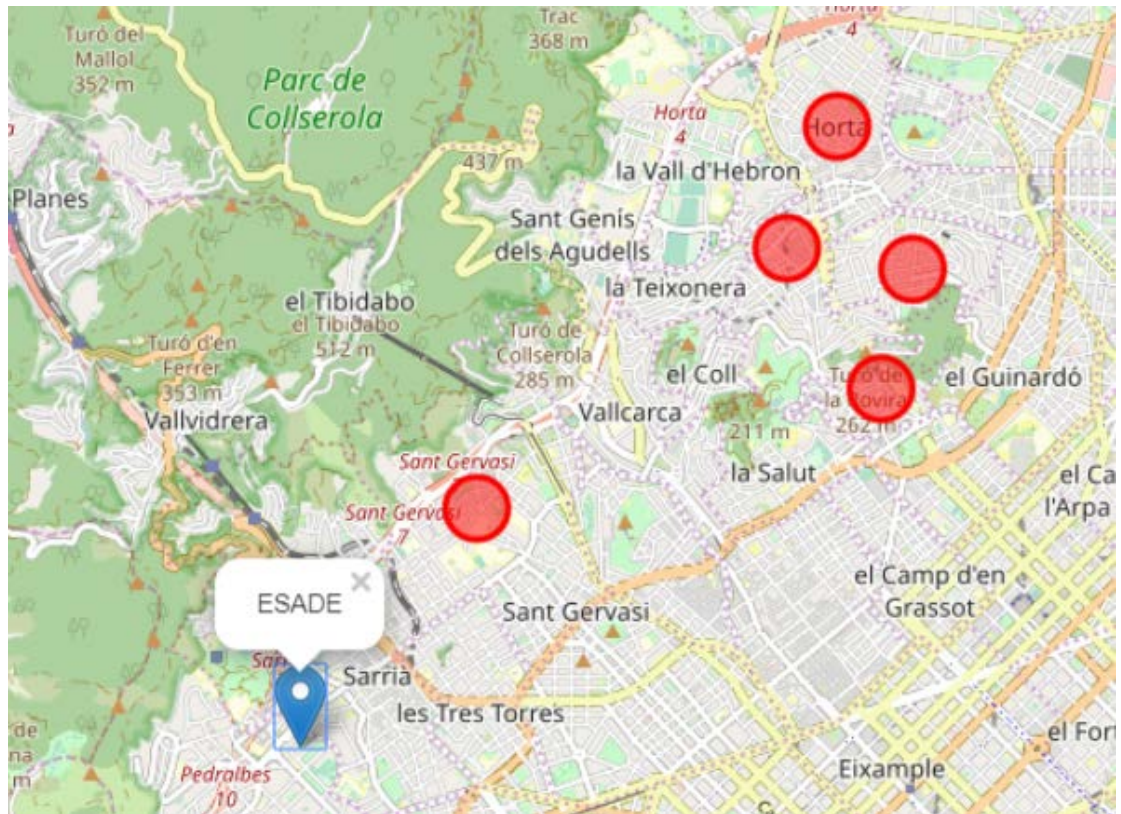
4. Results

By multiplying the neighborhoods matrix by score vector, I get an array of final preference scores. The idea comes from recommendation system.

```
array([[2.15      ],
       [1.36      ],
       [2.01219512],
       [1.8       ],
       [1.51      ],
       [2.29113924],
       [1.09090909],
       [1.05      ],
       [1.56      ],
       [2.13      ],
       [1.98      ],
       [1.72      ],
       [1.37      ],
       [1.34      ],
       [1.79245283],
       [2.18604651],
       [1.63      ],
       [1.30263158],
       [1.22      ],
       [1.37      ]])
```

In the end, it's easy to choose top 5 suitable neighborhoods for me to live in Barcelona based on the venues there and my preference towards those venues. They are ***Horta, la Font d'en Fargues, Can Baró, Sant Gervasi - la Bonanova, and El Carmel.***

	Neighborhood	Score	Latitude	Longitude
0	Horta	2.291139	41.433021	2.159088
1	la Font d'en Fargues	2.186047	41.424364	2.165320
2	Can Baró	2.150000	41.416966	2.162693
3	Sant Gervasi - la Bonanova	2.130000	41.409637	2.129637
4	El Carmel	2.012195	41.425591	2.154959



5. Discussion

For anyone who wants to find a best neighborhood to live, this recommendation system based on geographical data and venues information can provide a reasonable scenario.

Since two schoolmates are together with me towards Barcelona, maybe it is a great idea to ask them for their preference scores. By calculating the most suitable neighborhoods for each of us, we can select a best place to live.

6. Conclusion

In this report, I gather data from Wikipedia and Foursquare, then use data tools including NumPy and Pandas. Machine learning algorithm and data visualization methods help me do further analysis and demonstration. In the end, a recommendation system based on matrix multiplication helps me reach a final decision.

The most suitable neighborhoods for me to live in Barcelona are ***Horta, la Font d'en Fargues, Can Baró, Sant Gervasi - la Bonanova, and El Carmel.***