

Capstone Project

The Battle of Neighborhoods

By Eliana Salvemini

Problem explanation

- The problem I would like to analyze arises from a personal need: where should I choose to live? In the next month I have to move for job in another city and I will deal with the problem of where taking home.
- This is the typical problem that a person who has to be relocated for job has to face.
- A person could say to choose a place not too far from the office, but this is not enough. I want to find an area where I can practice my hobbies.
- First of all since I am a swimmer I want an area full of swimming pools.
- Then since I love to practice sport in general I want to live in an area with gyms and parks where to run and do jogging.
- Another person who loves arts and exhibitions could prefer an area full of theaters or cinemas for example. Other people could like going and chilling out in the evening so an area with pubs and clubs would be perfect.
- So my problem is in general: find the better place where to live depending on the hobbies and interests of a person.

Data

- The data I want to use are the geographical coordinates of Milan city, retrieved from the geolocator API.
- In particular I know that the postal codes of Milan city go from 20121 to 20162, so I will get the geographical coordinates of these postal codes.
- Next, I will explore with the Foursquare API the different venues for each postal code paying particular attention to the gyms, swimming pools, parks and at the end I will use K-means to cluster the data and choose the cluster that better fits to my hobbies.
- I would like to use more data like for example the prices of the houses in the different area but I was not able to retrieve no datasets regarding this. But this could be an idea for future work.

Methodology

- First of all I will get the geographical coordinates of the postal codes of Milan City: from 20121 to 20162.

	Postal Code	Latitude	Longitude
0	20121	45.472178	9.188044
1	20122	45.461812	9.196310
2	20123	45.462639	9.188515
3	20124	45.484604	9.200817
4	20125	45.499771	9.204911
5	20126	45.513290	9.217624
6	20127	45.496563	9.220420
7	20128	45.515129	9.224564
8	20129	45.470966	9.213798
9	20130	43.244552	-1.990588

Methodology

- I know that my future office is located in 'Via Castellanza 11, 20151 Milan' so with geolocator I get its coordinates.

```
myaddress = 'Via Castellanza 11, 20151 Milan'
geolocator = Nominatim()
location = geolocator.geocode(myaddress)
latitudeMy = location.latitude
longitudeMy = location.longitude
print('The geographical coordinate of {} are {},
{}.'.format(myaddress, latitudeMy, longitudeMy))
The geographical coordinate of Via Castellanza 11,
20151 Milan are 45.5009523, 9.109888.
```

Methodology

- Then for each postal code with Foursquare API I get the top 100 venues with appear in a radius of 1000 meters using as key words: *swimming pool, gym* and *park*.

search_query = Swimming Pool or Gym or Park

radius = 1000

LIMIT = 100

CLIENT_ID = # my Foursquare ID

CLIENT_SECRET = # my Foursquare Secret

VERSION = 20180605

- and compute the URL:

url =

`https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)`

Foursquare API Results for Pool

	id	name	categories	address	lat	lng
0	4c308fa366e40f4734aac38b	Piscina Comunale Lampugnano	['id': '4bf58dd8d48988d15e941735', 'name': 'P...']	Via Adolfo Omodeo	45.491386	9.115659
1	4f0fd180e4b0d3f8a3c90de1	Virgin Swimming Pool	['id': '4bf58dd8d48988d15e941735', 'name': 'P...']	Via Alcide de Gasperi 1	45.496361	9.128447
2	51dbaf19498ea602283851a2	The Hub's pool	['id': '4bf58dd8d48988d15e941735', 'name': 'P...']	Via Privata Polonia 10	45.513616	9.119248
3	51699a22e4b04259cd600bf	Superspa by Angelo Caroli @ B4	['id': '4bf58dd8d48988d105941735', 'name': 'G...']	NaN	45.511416	9.124666

Foursquare API Results for Gym

	id	name	categories	address	lat	lng
0	4cac3b1744a8224bda8e3140	Virgin Active	['id': '4bf58dd8d48988d175941735', 'name': 'G...']	Viale Alcide De Gasperi, 2	45.496443	9.128483
1	4d42a020b6e73704fbda8609	Way Out	['id': '4bf58dd8d48988d176941735', 'name': 'G...']	Via Sapri, 64	45.503863	9.130028
2	4defe40fd16486e86e4d90a3	Capelli e Sforza	['id': '4bf58dd8d48988d176941735', 'name': 'G...']	Via Lampugnano 80	45.490882	9.113795
3	52ed2d6a498e23a6618ffdae	CrossFit San Siro	['id': '4bf58dd8d48988d176941735', 'name': 'G...']	Via Gallarate 207	45.499128	9.124770
4	51c3530c498e5911e791b1fd	Rockspot	['id': '503289d391d4c4b30a586d6a', 'name': 'C...']	NaN	45.511383	9.091387

Foursquare API Results for Park

	id	name	categories	address	lat	lng
0	4bf026f1d4e4d13a97ea15a7	Parco di Trenno	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Cascina Bellaria	45.484267	9.107022
1	55af50bb498e3c69cda54dfb	Cascina Merlata	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Gallarate	45.506671	9.103671
2	4f07fdcbe4b0e624d6a9eb3e	Parco Sandro Pertini	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Francesco Cilea	45.497012	9.110018
3	4c963c94f7cfa1cd734fc415	Circolo Ricreativo RCS	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	NaN	45.500119	9.117218
4	4f89ac89e4b09efba1b8554b	Spazio Gorlini	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Parco Di Trenno	45.489359	9.102798
5	4cbd58244495721ec7215f7a	Giardini dei Caduti di Nassiriya	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Benedetto Croce	45.494014	9.128630
6	4d8a291399c2a1cdde908ad7	Parco Naturale Pero Atahotel	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Giovanni Keplero, 21	45.503373	9.086651
7	4e144d4414951daa08a8dccf	Parchetto Trenno	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	Via Trenno	45.488790	9.126559

Milan_venue dataframe

- Each Foursquare API run produces a venue dataframe. I concat the 3 dataframes for swimming pool, gym and park and obtain milan_venues.

	Postal Code	Postal Code Latitude	Postal Code Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Address	Venue Category
0	20121	45.472178	9.188044	Boscolo Milano, Autograph Collection	4ba68fd0f964a5208 b5e39e3	45.466514	9.193668	Corso Matteotti 4/6	Park
1	20121	45.472178	9.188044	acqua go	54fefc01498ea7e04 9439d53	45.477745	9.184486	NaN	Gym Pool
2	20122	45.461812	9.196310	Boscolo Milano, Autograph Collection	4ba68fd0f964a5208 b5e39e3	45.466514	9.193668	Corso Matteotti 4/6	Park
3	20122	45.461812	9.196310	Physioclinic	4c9daae40e9bb1f74 4c1df5f	45.461423	9.205257	Via Fontana 18	Gym
4	20123	45.462639	9.188515	Boscolo Milano, Autograph Collection	4ba68fd0f964a5208 b5e39e3	45.466514	9.193668	Corso Matteotti 4/6	Park

One hot encoding

- Then I applied the *one hot encoding* using the `pd.get_dummies()` method and produce the `milan_onehot` dataframe.

	Art Gallery	Art Museum	Athletics & Sports	Boxing Gym	Campground	Castle	Climbing Gym	College Gym	Event Space	Garden	...	Public Art	Resort	Road
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0
9	0	0	0	0	0	0	0	0	0	0	...	0	0	0

One hot encoding

- Here are the different venues categories that the one hot encoding produces:
- *Art Gallery, Art Museum, Athletics & Sports, Boxing Gym, Campground, Castle, Climbing Gym, College Gym, Event Space, Garden, General Entertainment, Gym, Gym / Fitness Center, Gym Pool, Harbor / Marina, Hotel, Italian Restaurant, Lake, Martial Arts Dojo, Monument / Landmark, Neighborhood, Park, Playground, Plaza, Pool, Public Art, Resort, Road, Spa, Sports Club, Stadium, Supermarket, Track, Vegetarian / Vegan Restaurant, Yoga Studio*

One hot encoding mean

- Then I group this dataframe for postal code and compute the mean of each category frequency.

	Postal Code	Art Gallery	Art Museum	Athletics & Sports	Boxing Gym	Campground	Castle	Climbing Gym	College Gym	Event Space	...
0	20121	0.020000	0.020000	0.0	0.0	0.0	0.020000	0.0	0.000000	0.0000	...
1	20122	0.016667	0.000000	0.0	0.0	0.0	0.016667	0.0	0.016667	0.0000	...
2	20123	0.017857	0.017857	0.0	0.0	0.0	0.017857	0.0	0.017857	0.0000	...
3	20124	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.0	0.024390	0.0000	...
4	20125	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0625	...

5 top common venues

- For each postal code I compute the 5 top most common venues along with their frequency of occurring.

- ----20121----

- venue freq
- 0 Park 0.38
- 1 Gym 0.16
- 2 Gym / Fitness Center 0.08
- 3 Plaza 0.06
- 4 Hotel 0.06

- ----20122----

- venue freq
- 0 Park 0.45
- 1 Gym 0.15
- 2 Gym / Fitness Center 0.12
- 3 Plaza 0.03
- 4 Monument / Landmark 0.03

- ----20123----

- venue freq
- 0 Park 0.43
- 1 Gym / Fitness Center 0.14
- 2 Gym 0.12
- 3 Plaza 0.04
- 4 Monument / Landmark 0.04

- ----20124----

- venue freq
- 0 Park 0.34
- 1 Hotel 0.17
- 2 Plaza 0.10
- 3 Gym 0.10
- 4 Gym / Fitness Center 0.10
-

10 top common venues

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	20121	Park	Gym	Gym / Fitness Center	Hotel	Plaza	Monument / Landmark	Yoga Studio	Playground	Art Museum	Castle
1	20122	Park	Gym	Gym / Fitness Center	Hotel	Monument / Landmark	Plaza	Lake	Castle	College Gym	Garden
2	20123	Park	Gym / Fitness Center	Gym	Hotel	Monument / Landmark	Plaza	Lake	Art Museum	Castle	College Gym
3	20124	Park	Hotel	Gym	Gym / Fitness Center	Plaza	Playground	Garden	Martial Arts Dojo	College Gym	Pool
4	20125	Park	Plaza	Event Space	Track	Gym / Fitness Center	Gym	Playground	Yoga Studio	General Entertainment	Garden

Cluster Neighborhoods analysis

- After removing the postal code column, I decided to run k-means to cluster the postal code into 5 different clusters.

Cluster #1

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	20138	Park	Gym / Fitness Center	Pool	Yoga Studio	Hotel	Gym Pool	Gym	General Entertainment	Garden	Event Space
22	20143	Park	Gym	Gym / Fitness Center	Harbor / Marina	Athletics & Sports	College Gym	Hotel	Pool	Garden	Gym Pool
23	20144	Park	Gym / Fitness Center	Yoga Studio	Pool	Public Art	Plaza	Martial Arts Dojo	Boxing Gym	Campground	Castle
25	20146	Park	Gym / Fitness Center	Gym	Martial Arts Dojo	Yoga Studio	Hotel	Gym Pool	General Entertainment	Garden	Event Space
26	20147	Park	Gym	Pool	Martial Arts Dojo	Gym / Fitness Center	Yoga Studio	Gym Pool	General Entertainment	Garden	Event Space
32	20153	Park	Campground	Gym	Yoga Studio	Garden	Harbor / Marina	Gym Pool	Gym / Fitness Center	General Entertainment	Event Space
35	20156	Park	Gym Pool	Playground	Yoga Studio	Hotel	Gym / Fitness Center	Gym	General Entertainment	Garden	Event Space
36	20157	Park	Gym / Fitness Center	Pool	Yoga Studio	Hotel	Gym Pool	Gym	General Entertainment	Garden	Event Space

Cluster #2

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	20121	Park	Gym	Gym / Fitness Center	Hotel	Plaza	Monument / Landmark	Yoga Studio	Playground	Art Museum	Castle
1	20122	Park	Gym	Gym / Fitness Center	Hotel	Monument / Landmark	Plaza	Lake	Castle	College Gym	Garden
2	20123	Park	Gym / Fitness Center	Gym	Hotel	Monument / Landmark	Plaza	Lake	Art Museum	Castle	College Gym
3	20124	Park	Hotel	Gym	Gym / Fitness Center	Plaza	Playground	Garden	Martial Arts Dojo	College Gym	Pool
6	20127	Park	Gym	Plaza	Gym / Fitness Center	College Gym	Martial Arts Dojo	Event Space	Playground	Pool	Track
8	20129	Park	Gym / Fitness Center	Plaza	Gym	Garden	Monument / Landmark	Yoga Studio	College Gym	Gym Pool	Playground
10	20131	Park	Gym	Gym / Fitness Center	Plaza	Pool	Yoga Studio	Playground	Martial Arts Dojo	College Gym	Supermarket
11	20132	Park	Gym / Fitness Center	Martial Arts Dojo	Event Space	Gym	Neighborhood	College Gym	Plaza	Garden	General Entertainment
12	20133	Park	Gym / Fitness Center	Gym	Plaza	College Gym	Supermarket	Monument / Landmark	Garden	Event Space	General Entertainment
29	20150	Hotel	Resort	Gym / Fitness Center	Gym	Pool	Sports Club	Park	Climbing Gym	Event Space	Campground
40	20161	Park	Gym / Fitness Center	Hotel	Gym	Pool	Yoga Studio	Gym Pool	General Entertainment	Garden	Event Space

Cluster 3

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	20130	Gym	Yoga Studio	Event Space	Harbor / Marina	Gym Pool	Gym / Fitness Center	General Entertainment	Garden	College Gym	Italian Restaurant

Cluster 4

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19	20140	Gym / Fitness Center	Park	Yoga Studio	Italian Restaurant	Harbor / Marina	Gym Pool	Gym	General Entertainment	Garden	Event Space
21	20142	Gym / Fitness Center	Park	Yoga Studio	Italian Restaurant	Harbor / Marina	Gym Pool	Gym	General Entertainment	Garden	Event Space

Cluster 5

	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	20125	Park	Plaza	Event Space	Track	Gym / Fitness Center	Gym	Playground	Yoga Studio	General Entertainment	Garden
5	20126	Park	Pool	Gym	Event Space	Martial Arts Dojo	Climbing Gym	Playground	Gym / Fitness Center	General Entertainment	Garden
7	20128	Park	Pool	Gym	Event Space	Martial Arts Dojo	Climbing Gym	Gym Pool	Gym / Fitness Center	General Entertainment	Garden
13	20134	Park	Gym	Garden	Gym / Fitness Center	Neighborhood	Plaza	Yoga Studio	Gym Pool	General Entertainment	Event Space
14	20135	Park	Gym	Pool	Yoga Studio	Boxing Gym	Garden	Martial Arts Dojo	Monument / Landmark	Art Gallery	Sports Club
15	20136	Park	Gym	Gym / Fitness Center	Pool	Yoga Studio	Gym Pool	Vegetarian / Vegan Restaurant	Martial Arts Dojo	Monument / Landmark	Plaza
18	20139	Park	Gym / Fitness Center	Gym	Gym Pool	Pool	Yoga Studio	Hotel	General Entertainment	Garden	Event Space
20	20141	Park	Hotel	Gym / Fitness Center	Gym	Pool	Martial Arts Dojo	Yoga Studio	Gym Pool	General Entertainment	Garden
24	20145	Park	Gym / Fitness Center	Gym	Yoga Studio	Art Museum	Martial Arts Dojo	Gym Pool	Garden	General Entertainment	Event Space
27	20148	Park	Pool	Gym Pool	Gym	Gym / Fitness Center	Playground	Hotel	Yoga Studio	General Entertainment	Garden
28	20149	Park	Gym	Pool	Gym / Fitness Center	Gym Pool	Yoga Studio	Hotel	Monument / Landmark	Playground	Lake
30	20151	Park	Pool	Gym	Gym / Fitness Center	Playground	Yoga Studio	Gym Pool	General Entertainment	Garden	Event Space
31	20152	Park	Pool	Gym	College Gym	Yoga Studio	Gym Pool	Gym / Fitness Center	General Entertainment	Garden	Event Space
33	20154	Park	Gym	Gym / Fitness Center	Lake	Art Museum	Italian Restaurant	Castle	Road	Monument / Landmark	Plaza
34	20155	Park	Gym	Gym / Fitness Center	Monument / Landmark	Playground	College Gym	Yoga Studio	Garden	Gym Pool	General Entertainment
37	20158	Park	Gym Pool	Gym / Fitness Center	Gym	College Gym	Yoga Studio	Garden	Harbor / Marina	General Entertainment	Event Space
38	20159	Park	Gym	Pool	Gym / Fitness Center	Yoga Studio	Gym Pool	Playground	Hotel	General Entertainment	Garden
39	20160	Pool	Park	College Gym	Yoga Studio	Gym Pool	Gym / Fitness Center	Gym	General Entertainment	Garden	Event Space
41	20162	Park	Pool	Gym	Gym / Fitness Center	Playground	Yoga Studio	Gym Pool	General Entertainment	Garden	Event Space

Discussion

- In order of importance, the venues I am mostly interested in are:
 1. swimming pool
 2. gym
 3. Park
- so I would like to find a postal code where this order is more or less respected.
- I will not consider **cluster 1** because the first most common venues are parks which are the one I am less interested in. Then at the second position there are gyms and swimming pools are only at the third position.
- For **cluster 2**: parks are the most common venues, in the second position there are
- For **cluster 3** we have only postal code 20130, but swimming pool are only at 5th position in term of importance. the gyms and pools appear only at 5th position.
- In **cluster 4** there are only 2 postal codes 20140 and 20142 where swimming pools are only in six position.

Discussion

- **Cluster 5** is the biggest one and this is the cluster I want to concentrate on. This cluster contains the postal code of my office which is *20151*. I can see how for this postal code the first most common venues are Park but Pools are the second most common venues. Moreover swimming pools appears even in the 7th position. For postal code *20152* which is close to my office as well, swimming pools appear in the second position and in the 6th one. I can also see how in this cluster swimming pools appear in different postal codes. For example for *20128*, *20151* and *20162* pools appear in the second and in the 7th position. For *20136* pools appear in 4th and 6th position. For *20139* pools appear in the 4th and 5th position. For *20141* pools appear in the 5th and 8th position. For *20148* pools appear in the second and third position. For *20149* pools appear in third and 5th position. For *20159* pools appear in the third and 6th position. But there is only one postal code in which pools appear in the first position and in the 5th one: the **20160**.

Conclusion

- In postal code 20160 Park are in the second position and there are also many gyms at third, 4th, 6th and 7th positions. In this area there also gardens and place where event can take place. I think this is the postal code where I want to live. Now I can go more in details in the different streets and search for some apartments or rooms to rent.