Introduction/Business Problem

1 Introduction

1.1 Background

Job plays a central role in our life, we spend about 14% of our lifetime at work. But because we spent such huge amount of time working, find a fitting job is not an easy task. Finding a dream job in your hometown, if you are living in a small city, is mission impossible. So often we have to move to other cities to chasing our dream job.
With this propose in mind find a well placed house, not too far from office or other areas where we can practice our hobbies, in a new city, is a critical task.

1.2 Business Problem

The objective of this capstone project is to find the most suitable location for living in a new city. Within the project we will explore the neighborhood of Milano and search for a better place where to live depending of the hobbies and interests of a person.

1.3 Target

A person who have to move to another city chasing his dream job.

2 Data

2.1 Data sourcers

To solve this problem, I will need below data :
- geographical coordinate of Milano: the postal codes of Milano go from 20121 to 201612 so we will retrieve the geographical coordinates of these postal codes
- Forsquare API to get venues data related to Milano areas. The main focus will be on gyms, swimming pools and fitness centers because these are my favorite hobbies.

2.2 Extracting data

- getting Latitude and Longitude data of the Milano neighborhoods via Nominatim package
- use Foursquare API to get venue data related to Milano neighborhoods

3 Methodology

3.1 Identify areas coordinates

Using Nominatim package we can get the geographical coordinates for all areas in Milano

```python
#get coordinates
for i in range(20121, 20162+1):
    location = geolocator.geocode(str(i), address)
    latitudeMi = location.latitude
    longitudeMi = location.longitude
    latitude.append(latitudeMi)
    longitude.append(longitudeMi)
    print('The geograpical coordinate of {} are {}, {}.'.format(str(i) + address, latitudeMi, longitudeMi))
```

```
The geograpical coordinate of 20121Milan Italy are 45.47209965286037, 9.188083637357634.
The geograpical coordinate of 20122Milan Italy are 45.461913126654856, 9.196374983587853.
The geograpical coordinate of 20123Milan Italy are 45.4632179225897, 9.177475393185716.
The geograpical coordinate of 20124Milan Italy are 45.4831028, 9.1994731.
The geograpical coordinate of 20125Milan Italy are 45.4996703215885, 9.204921034636818.
The geograpical coordinate of 20126Milan Italy are 21.9308311, -102.2843864.
The geograpical coordinate of 20127Milan Italy are 45.496602297442486, 9.220526978547303.
The geograpical coordinate of 20128Milan Italy are 45.51493449289599, 9.225577792516473.
The geograpical coordinate of 20129Milan Italy are 45.47140199137647, 9.213718798757128.
The geograpical coordinate of 20130Milan Italy are 43.24453390258106, -1.990582383566761.
The geograpical coordinate of 20131Milan Italy are 45.48376029871447, 9.222420693236819.
The geograpical coordinate of 20132Milan Italy are 41.9338015, 9.145888411588466.
The geograpical coordinate of 20133Milan Italy are 45.4675063, 9.2268744.
The geograpical coordinate of 20134Milan Italy are 45.477234481757336, 9.244661731071343.
The geograpical coordinate of 20135Milan Italy are 45.45471703248223, 9.2111613960378.
The geograpical coordinate of 20136Milan Italy are 38.678395, -77.4946056.
The geograpical coordinate of 20137Milan Italy are 45.455728877159025, 9.223286168104481.
The geograpical coordinate of 20138Milan Italy are 45.4542043, 9.246882395444766.
The geograpical coordinate of 20139Milan Italy are 45.439910917576, 9.218291845094784.
The geograpical coordinate of 20140Milan Italy are 43.217506979853425, -2.0211183862154223.
The geograpical coordinate of 20141Milan Italy are 45.4376494, 9.2000118.
The geograpical coordinate of 20142Milan Italy are 41.9046375, 9.005198021341464.
The geograpical coordinate of 20143Milan Italy are 41.715047, 8.992941095594375.
The geograpical coordinate of 20144Milan Italy are 53.57447692138757, 9.977849102726172.
The geograpical coordinate of 20145Milan Italy are 45.4649158, 9.1505807.
The geograpical coordinate of 20146Milan Italy are 45.45815364737022, 9.144515216447875.
The geograpical coordinate of 20147Milan Italy are 45.46058598852061, 9.127363478257072.
The geograpical coordinate of 20148Milan Italy are 45.4833723, 9.1350923.
The geograpical coordinate of 20149Milan Italy are 45.47963701722943, 9.152823260786677.
```

3.2 Identify office coordinates

My new office is located in "Piazza Gae Aulenti, 20124 Milano" so using Nominatim package we can get geographical coordinates.

```python
office = 'Piazza Gae Aulenti, 20124 Milano' # IBM Client Center address

geolocator = Nominatim(user_agent="DreamJobCity")
location = geolocator.geocode(office)
latitudeOffice = location.latitude
longitudeOffice = location.longitude

print('The geographical coordinate of {} are {}, {}.'.format(office, latitudeOffice, longitudeOffice))
```

```
The geographical coordinate of Piazza Gae Aulenti, 20124 Milano are 45.483456000000004, 9.190440363642619.
```

3.3 Explore the venues in each area in order to find the most enjoyable area

For me an enjoyable area means and area with a large number of swimming pools and gyms. This mean that i will use this criteria for querying Forsquare API.

```
In [14]: #Fetch the top 100 venues that are in Milan
         import requests

         query = 'Swimming Pool'
         resultsSwimming = requests.get( VENUES_URI.format(CLIENT_ID, CLIENT_SECRET, latitudeOffice, longitudeOffice, VERSION, query, RADIUS, LIMIT) ).js
         resultsSwimming
                    'state': 'Lombardia',
                    'country': 'Italia',
                    'formattedAddress': ['Corso como', 'Milano Lombardia', 'Italia']},
                   'categories': [{'id': '4bf58dd8d48988d105941735',
                     'name': 'Gym Pool',
                     'pluralName': 'Gym Pools',
                     'shortName': 'Pool',
                     'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/pool_',
                      'suffix': '.png'},
                     'primary': True}],
                   'photos': {'count': 0, 'groups': []}},
                  'referralId': 'e-0-4c1b498f63750f474498b467-4'},
                 {'reasons': {'count': 0,
                   'items': [{'summary': 'This spot is popular',
                     'type': 'general',
                     'reasonName': 'globalInteractionReason'}]},
                  'venue': {'id': '4bc76f5f6501c9b6a86d3e29',
                   'name': 'Piscina Bacone',
                   'location': {'address': 'Via Piccinni 10',
                    'lat': 45.482629906470486,
```

```
In [21]: nearby_venues = get_result_data(resultsSwimming)
         nearby_venues.head()
```

Out[21]:

	id	name	categories	address	lat	lng
0	5244274e11d2a351e0178739	Ceresio 7 Pools & Restaurant	Italian Restaurant	Via Ceresio, 7	45.484025	9.179849
1	4bd198dbb221c9b66ebfd5d0	Piscina Cozzi	Pool	Viale Tunisia 35	45.478397	9.201517
2	4ba4f699f964a52053c938e3	Miele	Pool	Via Lambertenghi 12	45.489098	9.186258
3	4c8b62663dc2a1cd1fa7b532	Centro Sportivo Murat	Pool	Via Dino Villani, 2	45.499327	9.190580

3.4 Apply K-Means Clustering Algorithm

3.4.1 Select the features for clustering purpose and scaled them:

```
In [57]: milan_onehot.columns

Out[57]: Index(['Art Gallery', 'Art Museum', 'Boxing Gym', 'Building', 'Campground',
                'Castle', 'Climbing Gym', 'College Gym', 'Dance Studio', 'Garden',
                'General Entertainment', 'Gym', 'Gym / Fitness Center', 'Gym Pool',
                'Hotel', 'Italian Restaurant', 'Lake', 'Martial Arts Dojo',
                'Monument / Landmark', 'Office', 'Park', 'Parking', 'Playground',
                'Plaza', 'Pool', 'Resort', 'Road', 'Skate Park', 'Spa', 'Sports Club',
                'Stadium', 'Track', 'Water Park', 'Yoga Studio'],
               dtype='object')
```

3.4.2 K-Means CLustering

With k=5, i use k-means clustering to agglomerate all data based on the transformed features.

Examine Clusters

```
In [111]: #Cluster 1
          milan_merged.loc[milan_merged['Cluster Labels'] == 0, milan_merged.columns[[0] + list(range(4, milan_merged.shape[1]))]]
```
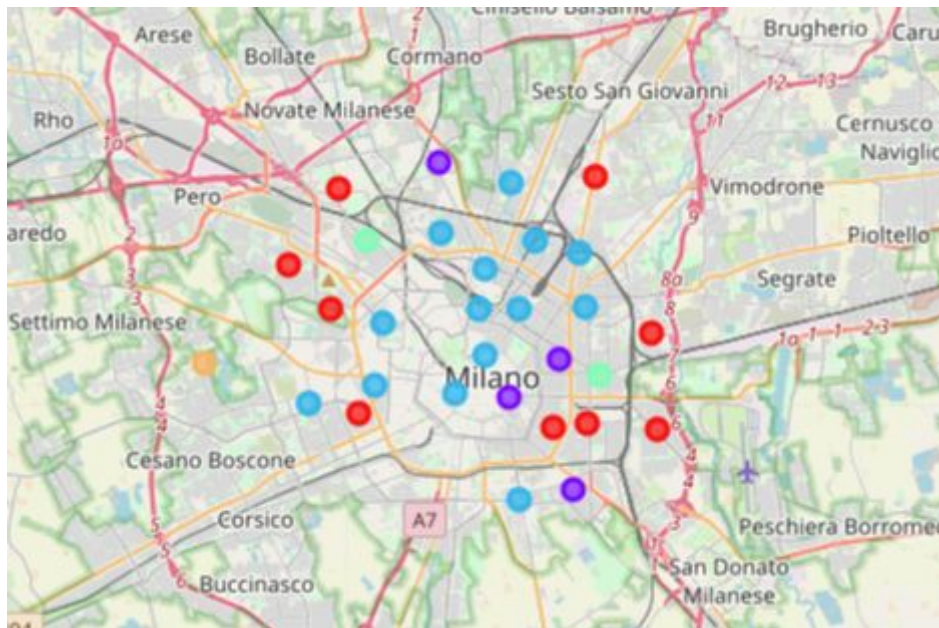
Out[111]:

	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
7	20128	Pool	Gym	Park	Martial Arts Dojo	Climbing Gym	Dance Studio	Gym / Fitness Center	General Entertainment	Garden	Yoga Studio
13	20134	Park	Gym	Gym / Fitness Center	Yoga Studio	Dance Studio	Gym Pool	General Entertainment	Garden	College Gym	Italian Restaurant
14	20135	Park	Gym	Pool	Martial Arts Dojo	Monument / Landmark	Yoga Studio	Sports Club	Boxing Gym	Campground	Castle
16	20137	Park	Gym	Pool	Boxing Gym	Martial Arts Dojo	Monument / Landmark	College Gym	Garden	Gym / Fitness Center	General Entertainment
17	20138	Pool	Gym	Park	Climbing Gym	Dance Studio	Gym Pool	Gym / Fitness Center	General Entertainment	Garden	Yoga Studio
25	20146	Park	Gym / Fitness Center	Gym	Martial Arts Dojo	Yoga Studio	Italian Restaurant	Gym Pool	General Entertainment	Garden	Dance Studio
27	20148	Park	Pool	Playground	Track	Gym	Hotel	Gym / Fitness Center	General Entertainment	Garden	Dance Studio
30	20151	Park	Gym	Pool	Gym / Fitness Center	Sports Club	Office	College Gym	General Entertainment	Garden	Dance Studio
36	20157	Park	Pool	Italian Restaurant	Gym Pool	Gym / Fitness Center	Gym	General Entertainment	Garden	Dance Studio	Yoga Studio

```
In [112]: #Cluster 2
          milan_merged.loc[milan_merged['Cluster Labels'] == 1, milan_merged.columns[[0] + list(range(4, milan_merged.shape[1]))]]
```

Out[112]:

	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
1	20122	Gym	Gym / Fitness Center	Park	Plaza	Martial Arts Dojo	Monument / Landmark	General Entertainment	Art Gallery	Spa	College Gym
5	20126	Gym / Fitness Center	Gym	Park	Yoga Studio	Dance Studio	Gym Pool	General Entertainment	Garden	College Gym	Italian Restaurant
8	20129	Gym / Fitness Center	Gym	Park	Yoga Studio	College Gym	Playground	Plaza	Pool	Building	Gym Pool
18	20139	Gym / Fitness Center	Gym Pool	Gym	Park	Pool	Italian Restaurant	General Entertainment	Garden	Dance Studio	Yoga Studio

4 Results

A large number of Swimming Pools, Gyms and Parks are present in Milan, however hobbies relevance is different for everyone. For me the most important hobby is swimming followed by gym and park walks. So the search for my new home will be focused in the areas(postal code) where the most common venues will allow me to enjoy my hobbies and be as close as possible to the office.

4.1 Clusters Analysis

4.1.1 Cluster #1

In this cluster the most common venues are parks, but park activities are my last favorite hobby, so the areas in this cluster aren't the most suited place in town

```
#Cluster 1
milan_merged.loc[milan_merged['Cluster Labels'] == 0, milan_merged.columns[[0] + list(range(4, milan_merged.shape[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
7	20128	Pool	Gym	Park	Martial Arts Dojo	Climbing Gym	Dance Studio	Gym / Fitness Center	General Entertainment		Yoga Studio
13	20134	Park	Gym	Gym / Fitness Center	Yoga Studio	Dance Studio	Gym Pool	General Entertainment	Garden	College Gym	Italian Restaurant
14	20135	Park	Gym	Pool	Martial Arts Dojo	Monument / Landmark	Yoga Studio	Sports Club	Boxing Gym	Campground	Castle
16	20137	Park	Gym	Pool	Boxing Gym	Martial Arts Dojo	Monument / Landmark	College Gym	Garden	Gym / Fitness Center	General Entertainment
17	20138	Pool	Gym	Park	Climbing Gym	Dance Studio	Gym Pool	Gym / Fitness Center	General Entertainment	Garden	Yoga Studio
25	20146	Park	Gym / Fitness Center	Gym	Martial Arts Dojo	Yoga Studio	Italian Restaurant	Gym Pool	General Entertainment	Garden	Dance Studio
27	20148	Park	Pool	Playground	Track	Gym	Hotel	Gym / Fitness Center	General Entertainment	Garden	Dance Studio
30	20151	Park	Gym	Pool	Gym / Fitness Center	Sports Club	Office	College Gym	General Entertainment	Garden	Dance Studio
36	20157	Park	Pool	Italian Restaurant	Gym Pool	Gym / Fitness Center	Gym	General Entertainment	Garden	Dance Studio	Yoga Studio

4.1.2 Cluster #2

Is on of the biggest clusters but is not a good choice because the most common venues are Gyms and Parks. I know that is 2 of 3 but my office address is not in this cluster.

```
#Cluster 2
milan_merged.loc[milan_merged['Cluster Labels'] == 1, milan_merged.columns[[0] + list(range(4, milan_merged.shape[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
1	20122	Gym	Gym / Fitness Center	Park	Plaza	Martial Arts Dojo	Monument / Landmark	General Entertainment	Art Gallery	Spa	College Gym
5	20126	Gym / Fitness Center	Gym	Park	Yoga Studio	Dance Studio	Gym Pool	General Entertainment	Garden	College Gym	Italian Restaurant
8	20129	Gym / Fitness Center	Gym	Park	Yoga Studio	College Gym	Playground	Plaza	Pool	Building	Gym Pool
18	20139	Gym / Fitness Center	Gym Pool	Gym	Park	Pool	Italian Restaurant	General Entertainment	Garden	Dance Studio	Yoga Studio
19	20140	Gym / Fitness Center	Park	Yoga Studio	Dance Studio	Gym Pool	Gym	General Entertainment	Garden	College Gym	Italian Restaurant
23	20144	Gym / Fitness Center	Park	Yoga Studio	Pool	Spa	Gym	Martial Arts Dojo	Building	Campground	Castle
29	20150	Gym / Fitness Center	Hotel	Gym	Yoga Studio	Resort	Pool	Spa	Campground	Castle	Building
40	20161	Gym / Fitness Center	Park	Pool	Gym	Italian Restaurant	Gym Pool	General Entertainment	Garden	Dance Studio	Yoga Studio

4.1.3 Cluster #3

The most relevant cluster because contains my office Postal Code **20124**. I can see that the most common venues are in reverse order of preference, however all my hobbies are the most common venues in many areas of this cluster, for example in *20149* pools are the most common venue, in *20162* are the second one.

```
#Cluster 3
milan_merged.loc[milan_merged['Cluster Labels'] == 2, milan_merged.columns[[0] + list(range(4, milan_merged.shape[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
0	20121	Gym	Park	Gym / Fitness Center	Plaza	Yoga Studio	Castle	Gym Pool	Hotel	Monument / Landmark	Lake
2	20123	Gym	Park	Gym / Fitness Center	Yoga Studio	Plaza	Pool	Spa	Road	Castle	Gym Pool
3	20124	Park	Gym / Fitness Center	Gym	Gym Pool	Martial Arts Dojo	Yoga Studio	Pool	Castle	Climbing Gym	College Gym
4	20125	Park	Track	Gym / Fitness Center	Gym	Playground	Yoga Studio	Dance Studio	General Entertainment	Garden	College Gym
6	20127	Park	Gym	Gym / Fitness Center	Gym Pool	Pool	Martial Arts Dojo	Playground	Plaza	Track	Climbing Gym
10	20131	Gym	Gym / Fitness Center	Park	Pool	Yoga Studio	Martial Arts Dojo	College Gym	Castle	Climbing Gym
20	20141	Gym	Park	Pool	Hotel	Martial Arts Dojo	Gym / Fitness Center	General Entertainment	Garden	Dance Studio	Yoga Studio
24	20145	Park	Gym / Fitness Center	Gym	Yoga Studio	Martial Arts Dojo	Gym Pool	Italian Restaurant	General Entertainment	Garden	Dance Studio
26	20147	Park	Gym	Dance Studio	Martial Arts Dojo	Gym / Fitness Center	Italian Restaurant	Gym Pool	General Entertainment	Garden	Yoga Studio
28	20149	Pool	Gym	Park	Gym / Fitness Center	Yoga Studio	Hotel	Stadium	Campground	Castle	Climbing Gym
33	20154	Gym	Park	Gym Pool	Gym / Fitness Center	Yoga Studio	Italian Restaurant	Pool	Climbing Gym	College Gym	Castle
37	20158	Gym	Park	Gym / Fitness Center	College Gym	Gym Pool	Skate Park	Dance Studio	General Entertainment	Garden	Yoga Studio
38	20159	Gym	Park	Pool	Gym / Fitness Center	Parking	Office	Yoga Studio	Gym Pool	Climbing Gym	College Gym
41	20162	Gym	Pool	Gym Pool	Gym / Fitness Center	Park	Italian Restaurant	General Entertainment	Garden	Dance Studio	Yoga Studio

4.1.4 Cluster #4

Is a small cluster where gyms are the most common venues but there few swimming pools are present.

```
#Cluster 4
milan_merged.loc[milan_merged['Cluster Labels'] == 3, milan_merged.columns[[0] + list(range(4, milan_merged.shape[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
12	20133	Gym	College Gym	Park	Garden	Dance Studio	Gym Pool	Gym / Fitness Center	General Entertainment	Yoga Studio	Italian Restaurant
35	20156	Gym	Park	Yoga Studio	Dance Studio	Gym Pool	Gym / Fitness Center	General Entertainment	Garden	College Gym	Italian Restaurant

4.1.5 Cluster #5

The smallest cluster with only one postal code where my favorite activity is in 6th position.

```
#Cluster 5
milan_merged.loc[milan_merged['Cluster Labels'] == 4, milan_merged.columns[[0] + list(range(4, milan_merged.shape[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
32	20153	Water Park	Campground	Park	Garden	Hotel	Gym Pool	Gym / Fitness Center	Gym	General Entertainment	Yoga Studio

5 Discussion

The aim of this project is to find the most suitable location for living in a new city. I found that the most suited area for me is the **20124**, but this area should be considered only as starting point because all the nearby areas (Cluster #3) of this postal code are suited to my lifestyle.

6 Conclusion

The main purpose of this project was to identify most suitable location for living in Milan in order to aid me have my Dream Job without limit my way of life.
Final decision will be made based on additional factors like real estate availability and prices.