

Living in Milan – a practical guide for students

1. Introduction

Nowadays moving to Milan for studying purposes is a fairly common practice. A lot of students move (especially from Middle and Southern Italy) to Milan to attend university (bachelor, master degrees and PhDs). Moving to a complex and dynamic city such as Milan is not always easy, especially for all those students who come from smaller cities (or even small villages) spread all over Italy. Being transplanted in such a caothic city may lead students to feel lost and disorienteted in the middle of all possibilities that the city has to offer.

2. Business Problem

The aim of the project is to help students to have a synthetic and global view of all the possibilities that Milan has to offer and how this opportunities are spread all over the territor. Where useful facilities are located? Which NILs (which are divisions of the territory that can be interpreted exactly as neighbourhoods) offer the best amusement possibilities? Where and which kind of restaurants, bar, pubs are available in the city? The project tries to answer to all this kind of questions and, moreover, takes into consideration even the average age of each NIL and the distribution of all the universities in the city to best fit the necessities of new students.

3. Data

A huge amount of useful datasets about the city of Milan can be found at the following link: <https://dati.comune.milano.it/dataset>. Here the municipal administration, stores a lot of different datasets (more than 1000) that contain informations about very different topics. The datasets we are going to use in the project are mostly in the “.json” and “.geojson” format.

3.1 First dataset

The first dataset is about the different NILs (or most commonly neighbourhoods) of Milan. The fields in the dataset must be interpreted as follows:

1. ID_NIL: stands for "Nuclei di identità locali", which represents neighbourhoods of Milan
2. NIL: name description of the neighbourhood
3. Valido_dal: indicates the saring validity date of the data
4. Valido_al: indicates if the neighbourhood is still valid nowadays
5. Fonte: indicates if the neighbourhoods data are approved
6. Shape_Length: linear extension of the neighbourhood
7. Shape_Area: area extension of the neighbourhood
8. OBJECTID: unique id
9. geometry: spatial coordinates of the boundaries of the neighbourhood

3.2 Second dataset

The second dataset is necessary to associate to each NIL its litude and longitude coordinates, which will be necessary for the application of the Foursquare API. The fields in the dataset must be interpreted as follows:

1. id_nil: stands for "Nuclei di identità locali", which represents neighbourhoods of Milan
2. nil_name: name description of the nil (neighbourhood)
3. lng: calculated longitude for the nil (neighbourhood)
4. lat: calculated latitude for the nil (neighbourhood)

3.3 Third dataset

The third is huge and contains a lot of different data about population. Different indicators about population are stored in the dataset, but in the project we will be interested in the average age for each NIL of the city. The fields in the dataset must be interpreted as follows:

1. Anno: year of census (the data are available for 1991, 2001 and 2011)
2. Territorio: division of the territory (for example division by NIL)
3. Indicators: indicator analyzed (different types of indicators are stored all together in the dataset)
4. Definizione: description of the indicator taken into consideration
5. Valore indicatore: numeric value for the indicator taken into consideration

3.4 Fourth dataset

The last dataset is about universities in Milan. The fields in the dataset must be interpreted as follows:

1. T_LAUREA: course of study available inside the university
2. DENOMINAZ: denomination of the university
3. LONG_X_4326: longitude coordinate of the university
4. LANG_Y_4326: latitude coordinate of the university

3.5 Foursquare API

In the project we will make use of Foursquare API. We will need data about different venues in the city. In order to get that kind of informations we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

4. Methodology

4.1 Data about venues in Milan

To explore the most relevant venues in the city of Milan, we decided to preliminarily explore NIL s of the city (i.e. neighborhoods) using Foursquare API, then we segment them, and finally we grouped them into clusters to find similar NILs. The clusterization of all the NIL gathered using Foursquare API is carried out using a form of unsupervised machine learning algorithm from Scikit-Learn package: k-means clustering algorithm. In *Figure 1* and *Figure 2* is shown part of the code we wrote to implement the clustering algorithm.

id_nil	nil_name	lng	lat	Cluster Labels	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	
0	48	RONCHETTO SUL NAVIGLIO - Q. RE LODOVICO IL MORO	9.137260	45.438460	3.0	Theater	Supermarket	Italian Restaurant	Pizza Place	Fast Food Restaurant	Fish & Chips Shop	Flea Market	Flower Shop	Food
1	64	TRENNO	9.101675	45.492822	0.0	Pizza Place	Adult Education Center	Pub	Sandwich Place	Soccer Field	Ice Cream Shop	Plaza	Park	Bakery
2	67	PORTELLO	9.153947	45.484490	3.0	Italian Restaurant	Café	Japanese Restaurant	Hotel	Bar	Pub	Pizza Place	Restaurant	Pastry Shop
3	81	BOVISASCA	9.156731	45.517433	3.0	Soccer Field	Shopping Plaza	Park	Supermarket	Restaurant	Art Gallery	Gym / Fitness Center	Bus Stop	Italian Restaurant
4	84	PARCO NORD	9.184235	45.523514	2.0	Football Stadium	Playground	Wine Shop	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	Fountain	Food Truck	Food Court

Figure 1

4.2 Data about average age of Milan

To obtain data about average of each NIL of Milan we used a dataframe where are stored a lot of useful informations about different kind of indicators concerning age of the population. We ended up with the average age for every NIL in 3 different decades: 1991, 2001, and 2011 (*Figure 3*). Since no data are available for the year 2021, we implemented a simple linear regression model which predict the trend of age of the population from data about previous years censi (*Figure 4*).

	NIL	Età_media_1991	Età_media_2001	Età_media_2011	Età_media_2021
0	1	42.8	44.0	45.3	0.0
1	2	43.2	44.9	45.9	0.0
2	4	42.9	44.8	46.5	0.0
3	5	43.5	45.2	46.3	0.0
4	6	42.5	44.1	45.3	0.0
...
64	79	41.8	43.4	42.9	0.0
65	80	41.3	43.8	43.7	0.0
66	81	39.3	44.6	46.9	0.0
67	82	43.2	45.4	43.6	0.0
68	83	38.1	42.9	45.3	0.0

Figure 2

```

# build linear regression model
Y_pred = []
for i in range(len(age_df)):
    X = [1991, 2001, 2011]
    Y = [age_df["Età_media_1991"].loc[i], age_df["Età_media_2001"].loc[i], age_df["Età_media_2011"].loc[i]]

    X = np.array(X).reshape(-1, 1)
    Y = np.array(Y)

    model = LinearRegression()
    model.fit(X, Y)

    Y_pred.append(float(model.predict(np.array([2021]).reshape(-1, 1))))

# round values obtained from linear regression
Y_pred_rounded = []
for num in Y_pred:
    num = round(num, 1)
    Y_pred_rounded.append(num)

# show dataframe with prediction for 2021 population by NIL
age_df["Età_media_2021"] = Y_pred_rounded
age_df

```

Figure 4

4.3 Data about universities locations in Milan

To get informations about all the different universities located in Milan, we explored a dataframe where are stored all the universities with their latitude and longitude coordinates; together with the denomination of the university are made available all the different degrees that the university has to offer (Figure 5).

	T_LAUREA	DENOMINAZ	LONG_X_4326	LAT_Y_4326
0	ACCADEMIA	ACCADEMIA DI BELLE ARTI DI BRERA	9.187669	45.472245
1	ARTI E ANTROPOLOGIA DEL SACRO	ACCADEMIA DI BELLE ARTI DI BRERA	9.187669	45.472245
2	SCUOLA DI ARTI VISIVE	ACCADEMIA DI BELLE ARTI DI BRERA	9.187669	45.472245
3	SCUOLA DI COMUNICAZIONE DIDATTICA DELL'ARTE	ACCADEMIA DI BELLE ARTI DI BRERA	9.187669	45.472245
4	SCUOLA DI COMUNICAZIONE E ORGANIZZAZIONE PER L...	ACCADEMIA DI BELLE ARTI DI BRERA	9.187669	45.472245
...
587	MEDICINA E CHIRURGIA	VITA-SALUTE SAN RAFFAELE	9.267605	45.506910
588	NEUROSCIENZE COGNITIVE	VITA-SALUTE SAN RAFFAELE	9.267605	45.506910
589	PSICOLOGIA CLINICA	VITA-SALUTE SAN RAFFAELE	9.228690	45.499750
590	SCIENZE DELLA COMUNICAZIONE	VITA-SALUTE SAN RAFFAELE	9.267605	45.506910
591	SCIENZE PSICOLOGICHE	VITA-SALUTE SAN RAFFAELE	9.267605	45.506910

Figure 5

All the informations about datasets are explained in the previous section and also available in the Notebook of the project. In the Notebook is also possible to follow all the steps that we followed to clean all the datasets and to arrange data in a clear and useful way.

5. Results

5.1 Results about venues section

The clustering algorithm lead to division of venues in Milan in 5 different clusters (Figure 6). Each cluster is clearly analyzed in the Notebook of the project.

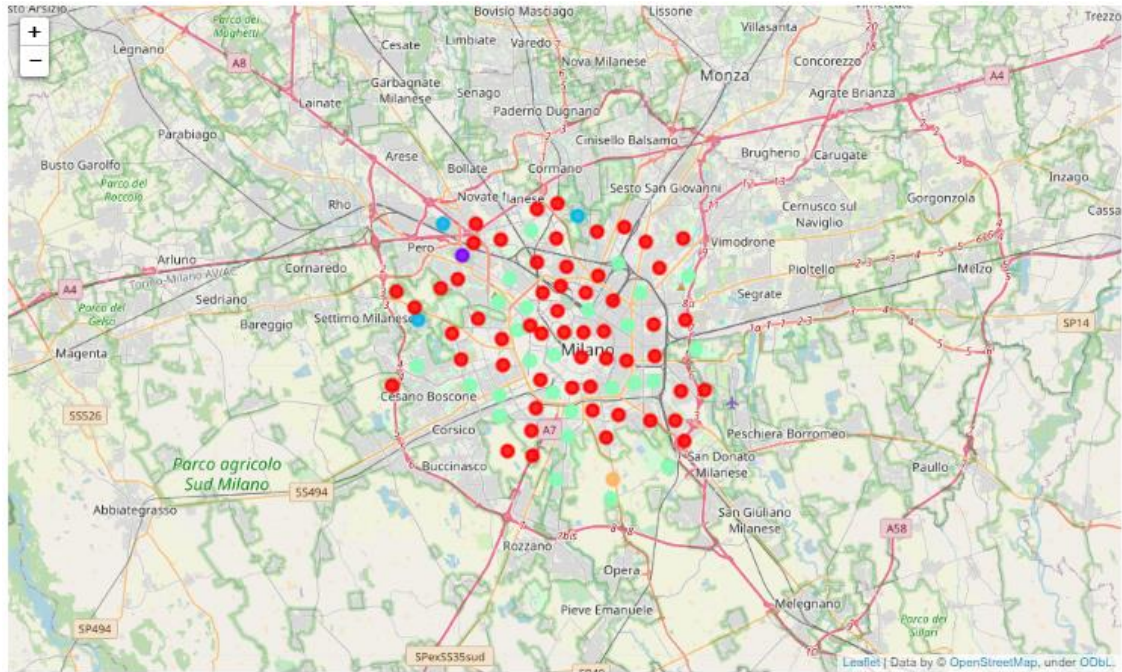


Figure 6

5.2 Results about average age section

Analysis on the average age of the city is summed up in Figure 7. For each NIL is indicated the average age of the population. Take a look to the Notebook of the project for the full analysis on the average age of NILs of Milan.

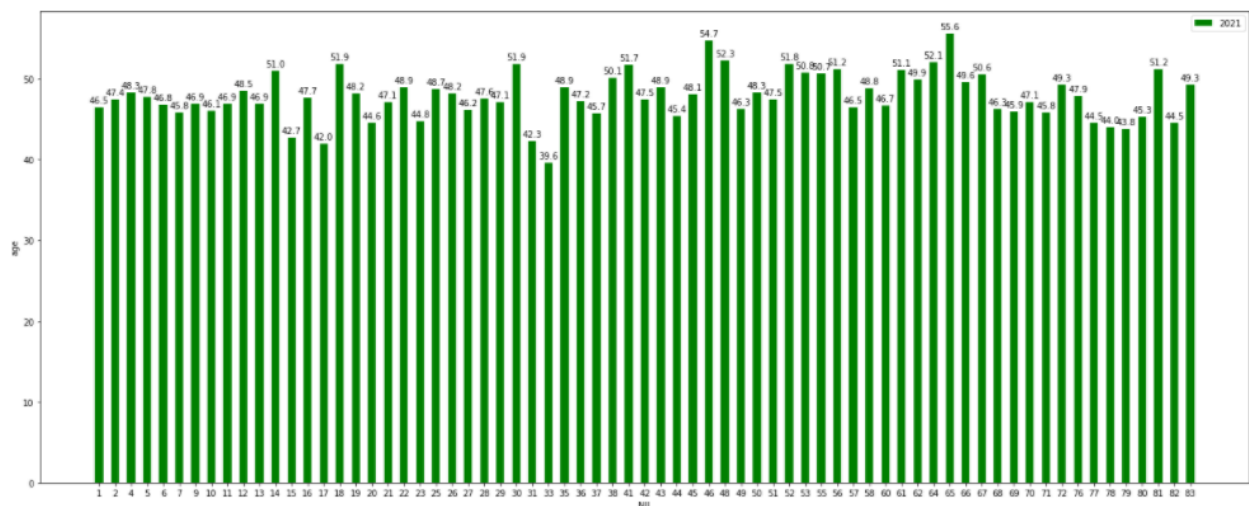


Figure 7

5.3 Results about universities section

The location of all the universities (and the available degrees) is depicted in *Figure 8*. Take a look to the Notebook of the project for the full analysis on the universities of Milan.

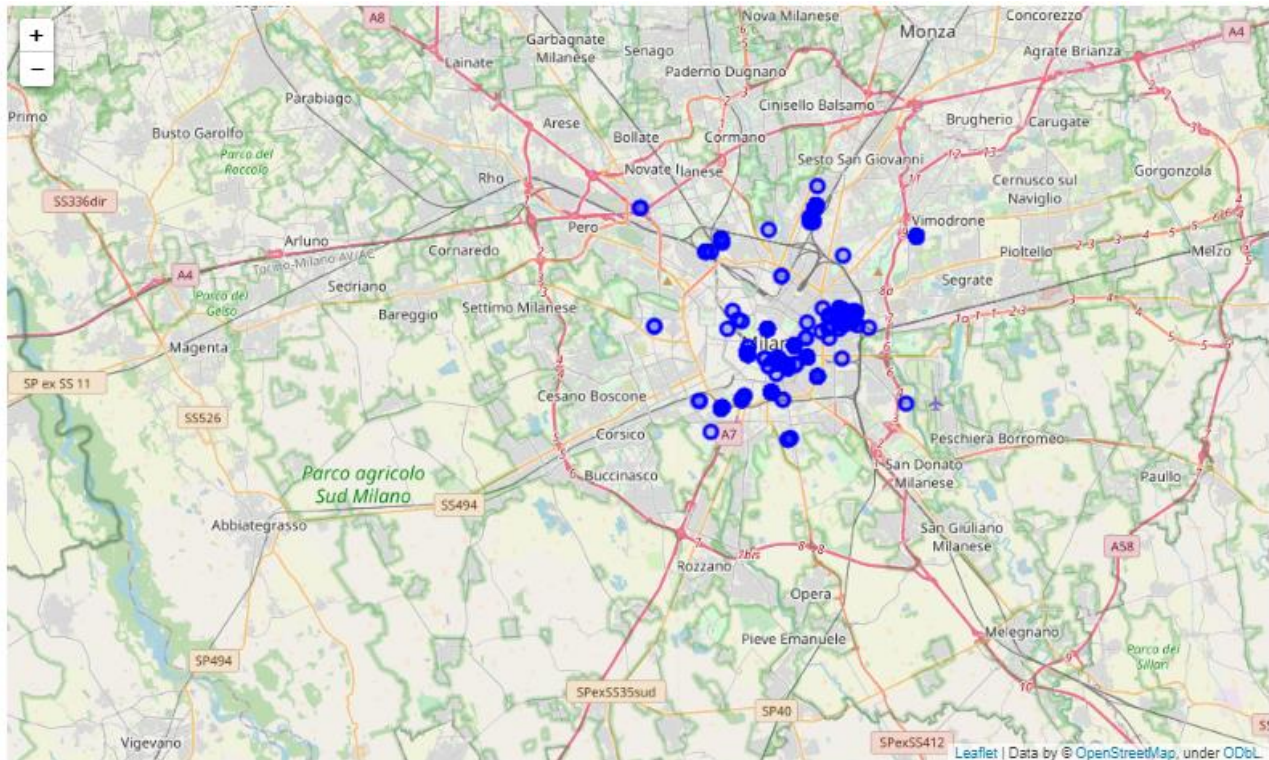


Figure 8

6. Discussion

Generally speaking, the city of Milan has many opportunities to offer to a student moving there for the first time. A lot of different venues have been found using Foursquare API. Different facilities, entertainment places are well spread all over the city (*Figure 9*).

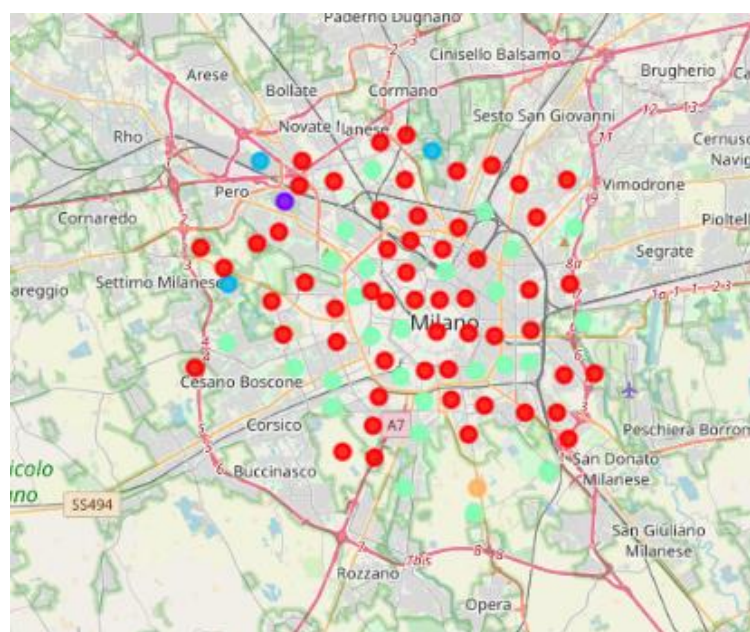


Figure 9

The average age of the city is estimated at 47.8 years, anyway the age distribution between NILs may vary a lot, ranging from the minimum 39.6 years up to 55.6 years as a maximum (*Figure 10*).

	NIL	Età_media_1991	Età_media_2001	Età_media_2011	Età_media_2021
28	33	41.9	41.7	40.2	39.6
14	17	38.3	40.1	40.6	42.0
27	31	35.3	38.2	39.8	42.3
12	15	39.6	40.6	41.7	42.7
64	79	41.8	43.4	42.9	43.8
...
26	30	45.8	49.8	49.4	51.9
52	64	36.5	41.7	46.9	52.1
39	48	39.6	44.9	47.8	52.3
38	46	39.7	45.5	49.5	54.7
53	65	42.0	47.9	50.7	55.6

Figure 10

Universities are abundant in the city and still remain one of the main attractive the city has to offer for students coming from all over Italy. In the pool of universities of the city, there are excellences recognized all over Europe and the World, such as Politecnico di Milano. Anyway the distribution of universities is not as fair as the distribution of venues in the city: most of universities (with thei respective offered degrees) are located in the South-est part of the city (*Figure 11*).



Figure 11

7. Conclusion

The city of Milan has so many things to offer to incoming students who can benefit from the big variety of universities on the territory and, at the same time, enjoy life thanks to all the facilities and events spread all over the city.

This project can be seen as a starting point to build a more wide analysis on all the relevant aspects a student has to face when moving to a big city such as Milan. An interesting analysis could be made on the average houses prices, particularly the average cost for renting a room or an house depending on the different NILs of Milan could be very useful for a new student moving to the city.

In the project the following Python libraries were used:

Pandas: library to create and manipulate dataframes.

Folium: library visualize interactive leaflet maps.

Numpy: scientific library.

Request: http library.

Folium: map rendering library

Geocoder: library to retrieve location data.

Geopandas: library to work with spatial data.

Scikit Learn: machine learning algorithms library.

Shapely: library to work with geometric figures.

Matplotlib: Python Plotting Module.