

# Capstone Project

## The Battle of Neighbourhoods

Scrivi per inserire testo

Family-Friendly  
neighbourhoods  
in Milano

# Background

Milano, is the Business and Financial capital of Italy. Most important companies have their Italian Head-Quarter in Milano. That means a lot of people move in Milan to find the best working place, and to improve their career

Starting from Expo '15 Milano is Living a fast growing period in terms of business opportunities, and that is increasing tourism, the born of new companies, and obviously, people who decide to live here.

Milano, is well know also for its “Movida”, and is full of restaurants, pubs, discos, also its cultural offer is huge, with its museums, theatres and university.

# Business Problem

In a city with a business heart, that is living a fast-growing period, finding the best place to live with young children is becoming really hard, especially if you don't know the city, because you are relocating here for work reasons.

The question to answer, thus, is:

**“which is the most  
family-friendly  
neighbourhood in Milano?”**

# TARGET

The target is huge and include:

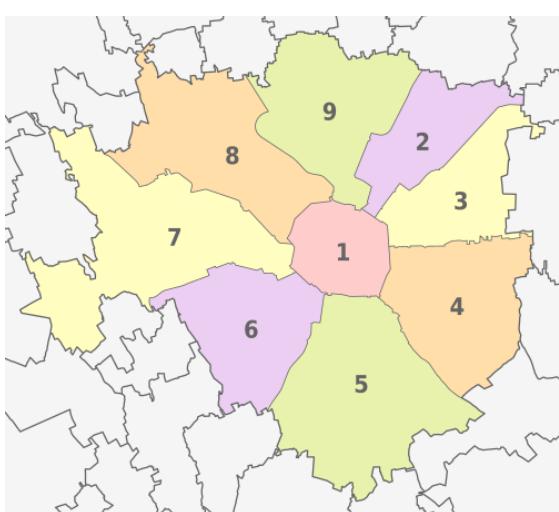
- Families living in Milan who are looking for a new location which fits better their needs
- Families moving to Milan
- Real Estate Agencies, in order to provide a better service to their clients

# SUCCESS CRITERIA

The success criteria of this project will be to provide to families with young children, the right information which allows them to choose the best area of Milano in which living.

# DATA DESCRIPTION

## Milano division



The municipality of Milan is divided into 9 municipalities. Each municipality contains a variable number of neighbourhoods.

This is the administrative division of the municipality.

From the point of view of the population and services census, Milan is divided into 88 NIL (local

identity nucleus), and all available data refer to this variable. Therefore we take NIL as Neighbourhoods, and municipality as Borough

## Family-Friendly concept

In order to find which is the neighbourhood most family friendly, following some surveys results found in the web, we took into the account following parameters:

- Number of schools:
  - Kindergartens
  - Primary Schools
  - Secondary Schools
- Number of Paediatricians

Once we have ranked neighbourhoods, we use Foursquare venues location in order to clustering that and to find the place which fits also parents expectations and needs.

## DATA SOURCES



The municipality of Milano, provides a web portal where it is possible to have access to several dataset: <http://dati.comune.milano.it/>

Data are available in different format:

- csv
- JSON
- Geojson

# LINK TO DATA

Kindergartens:

[http://dati.comune.milano.it/dataset/836d6a8c-c266-43bf-a3e4-16f382c4d9d8/resource/11e0d884-5032-4244-b79a-3567ce3efd2e/download/ds297\\_infanzia\\_quartiere-cittad.csv](http://dati.comune.milano.it/dataset/836d6a8c-c266-43bf-a3e4-16f382c4d9d8/resource/11e0d884-5032-4244-b79a-3567ce3efd2e/download/ds297_infanzia_quartiere-cittad.csv)

Primary Schools:

[http://dati.comune.milano.it/dataset/c3709838-21c8-4ddb-a910-dff6455894bb/resource/d9a5ed54-b8cd-4a04-b07b-914e10849c54/download/ds298\\_primaria\\_quartiere-cittad.csv](http://dati.comune.milano.it/dataset/c3709838-21c8-4ddb-a910-dff6455894bb/resource/d9a5ed54-b8cd-4a04-b07b-914e10849c54/download/ds298_primaria_quartiere-cittad.csv)

Secondary Schools:

[http://dati.comune.milano.it/dataset/28f7446d-7016-454d-9d27-781300a7ef38/resource/25c1faad-a14c-4fab-bc9c-010172beec9f/download/ds299\\_secondaria-i-grado\\_quartiere-cittad.csv](http://dati.comune.milano.it/dataset/28f7446d-7016-454d-9d27-781300a7ef38/resource/25c1faad-a14c-4fab-bc9c-010172beec9f/download/ds299_secondaria-i-grado_quartiere-cittad.csv)

Paediatricians:

[http://dati.comune.milano.it/dataset/b48fdaec-a868-47a8-8b59-fc9ed65c84a4/resource/22b05e1f-c5d2-4468-90e5-c098977856ef/download/ds235\\_sociale\\_pediatri\\_libera\\_scelta\\_2014\\_.csv](http://dati.comune.milano.it/dataset/b48fdaec-a868-47a8-8b59-fc9ed65c84a4/resource/22b05e1f-c5d2-4468-90e5-c098977856ef/download/ds235_sociale_pediatri_libera_scelta_2014_.csv)

# DATA DESCRIPTION

As we need to explore which is the most family-friendly neighbourhood in Milano, data with geolocation of Milano's neighbourhood is one of the key for this project.

Other key data are the distribution of schools and paediatricians.

Unfortunately data containing neighbourhoods coordinates are not available ready to use, so we need to process and clean what Milano data portal gives.

To carry on our analysis we need a data frame containing. At least:

- Neighbourhood ID
- Neighbourhood Name
- Latitude
- Longitude

From official Milano data portal we get a file with the following features:

FIELD - ITALIAN NAME	FIELD - ENGLISH TRANSLATION
CODICE_VIA	ID STREET
DUG	KIND OF STREET
DENOMINAZIONE VIA	STREET NAME
NUMERO CIVICO	STREET NUMBER
BARRA 2	SUB NUMBER
CAP	ZIP CODE
COMUNE	CITY
ID_Qri	NEIGHBOUR ID
MUNICIPIO	ID BOROUGH
ID_NIL	ID_NIL
NIL	NIL NAME

Starting from this data, and using another file with street coordinates, we derive the data frame we need, containing the following informations:

FIELD - ITALIAN NAME	FIELD - ENGLISH TRANSLATION
ID_NIL	NIL NAME
NIL	NIL NAME
LATITUDE	LATITUDE
LONGITUDE	LONGITUDE

To find NIL's coordinate we use centroid method starting from streets latitude and longitude.

		<b>NIL</b>	<b>ID_NIL</b>	<b>Latitude</b>	<b>Longitude</b>
<b>0</b>	ADRIANO	17	45.511567	9.244967	
<b>1</b>	AFFORI	80	45.514848	9.172868	
<b>2</b>	BAGGIO	55	45.461216	9.088377	
<b>3</b>	BANDE NERE	52	45.460334	9.138853	
<b>4</b>	BARONA	46	45.432494	9.154109	
...	...	...	...	...	...
<b>83</b>	VIGENTINA	5	45.451591	9.192724	
<b>84</b>	VILLAPIZZONE	71	45.496731	9.144505	
<b>85</b>	VLE MONZA	16	45.511586	9.227185	
<b>86</b>	WASHINGTON	51	45.460164	9.155163	
<b>87</b>	XXII MARZO	26	45.462359	9.214274	

In order to calculate the Family-Friendly Index, we need informations about schools and paediatricians.

Regarding files containing data on schools, these report the number of pupils for each school and for each school year, referring to the various NILs.

Since we are only interested in knowing whether or not a school is present in a given NIL, we processed the data until we obtained a table containing this information.

	<b>ID_NIL</b>	<b>Kindergarten</b>	<b>Primary School</b>	<b>Secondary School</b>
<b>0</b>	1	1.0	1.0	1.0
<b>1</b>	2	1.0	1.0	1.0
<b>2</b>	3	1.0	0.0	0.0
<b>3</b>	4	1.0	1.0	1.0
<b>4</b>	5	1.0	1.0	1.0

# METHODOLOGY

We adopt a standard working flow, which comprises four stages:

1. Collect Inspection Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling

After importing and analysing data, the most important step is to define to define a way to measure how a Neighbourhood is Family Friendly.

Surfing on the web, we found that a family with children thinks it is important to have close to its home schools and doctors.

Where high schools and universities are located is not important, because sons and daughters will be older enough to move with public transport, and also because the choice of those schools is based on several factors such as the school ranking.

We have created a **Family-Friendly Index (FFI)** which is built on the following assumptions (derived from our research):

$$\text{FFI} = \text{Presence of school} + \text{Presence of paediatricians}$$

- Presence of school: 20% each category --> Max score = 60
- Presence of paediatricians: Max score = 40

$$\text{Presence of paediatricians} = (\text{PNIL}/\text{mean(paediatricians)}) * n$$

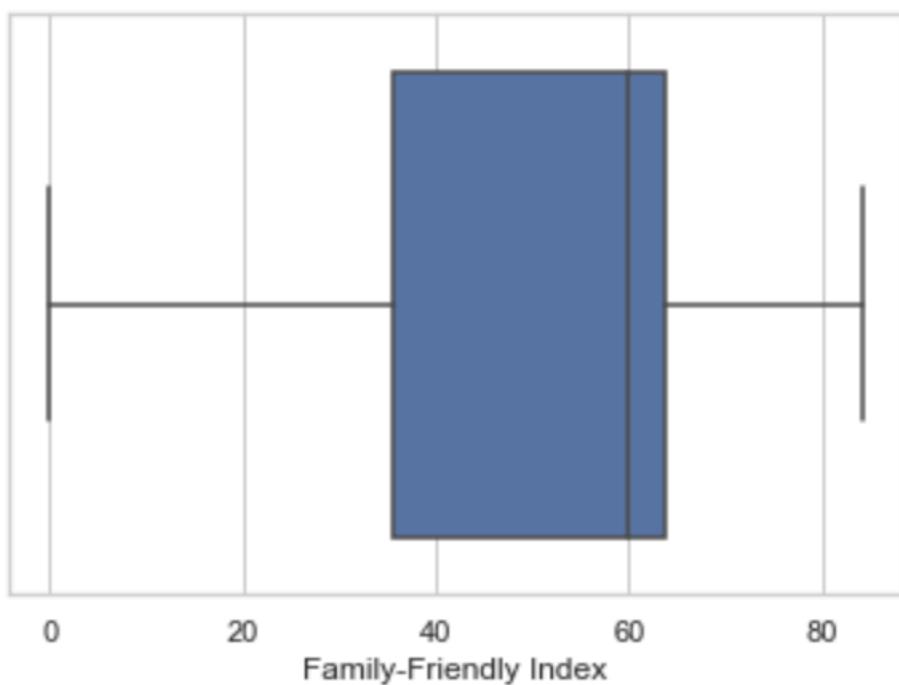
where:

$$n = \max(\text{paediatricians}) / (\text{mean}(\text{paediatricians}))$$

PNIL = number of paediatricians in a given NIL

ID NIL	Kindergarten	Primary School	Secondary School	Pediatrician	Family-Friendly Index	
0	1	1.0	1.0	1.0	2.0	63.718310
1	2	1.0	1.0	1.0	2.0	63.718310
2	3	1.0	0.0	0.0	0.0	20.000000
3	4	1.0	1.0	1.0	0.0	60.000000
4	5	1.0	1.0	1.0	3.0	65.577465

We can quickly analyse the index distribution using a simple boxplot:



From the graph we see that no NIL observes the maximum value of **Family-Friendly Index**, which means that ,in order to select the best neighbourhood, a family must also base its choice on other parameters. For this reason we will investigate through the **Foursquare API** which venues there are in neighbourhoods.

After fetching venues from Foursquare, we'll cluster results using K-Means algorithm.

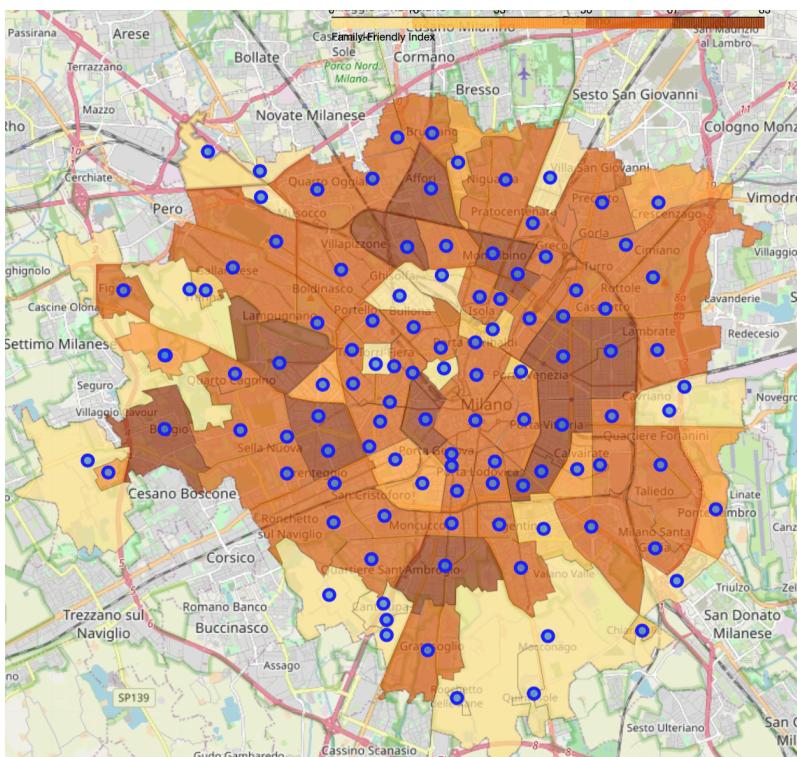
# Results and Discussion section

First of all, we can say that Milano could be not so bad for families with young children, despite its being a crowded and business oriented city.

Said this, we may discuss our results under two main perspectives:

- Family-Friendly Index
- Venues got from Foursquare

Analysing the Family-Friendly Index we have found the best neighbourhoods (in our dataset NIL) in terms of presence of schools and paediatricians.

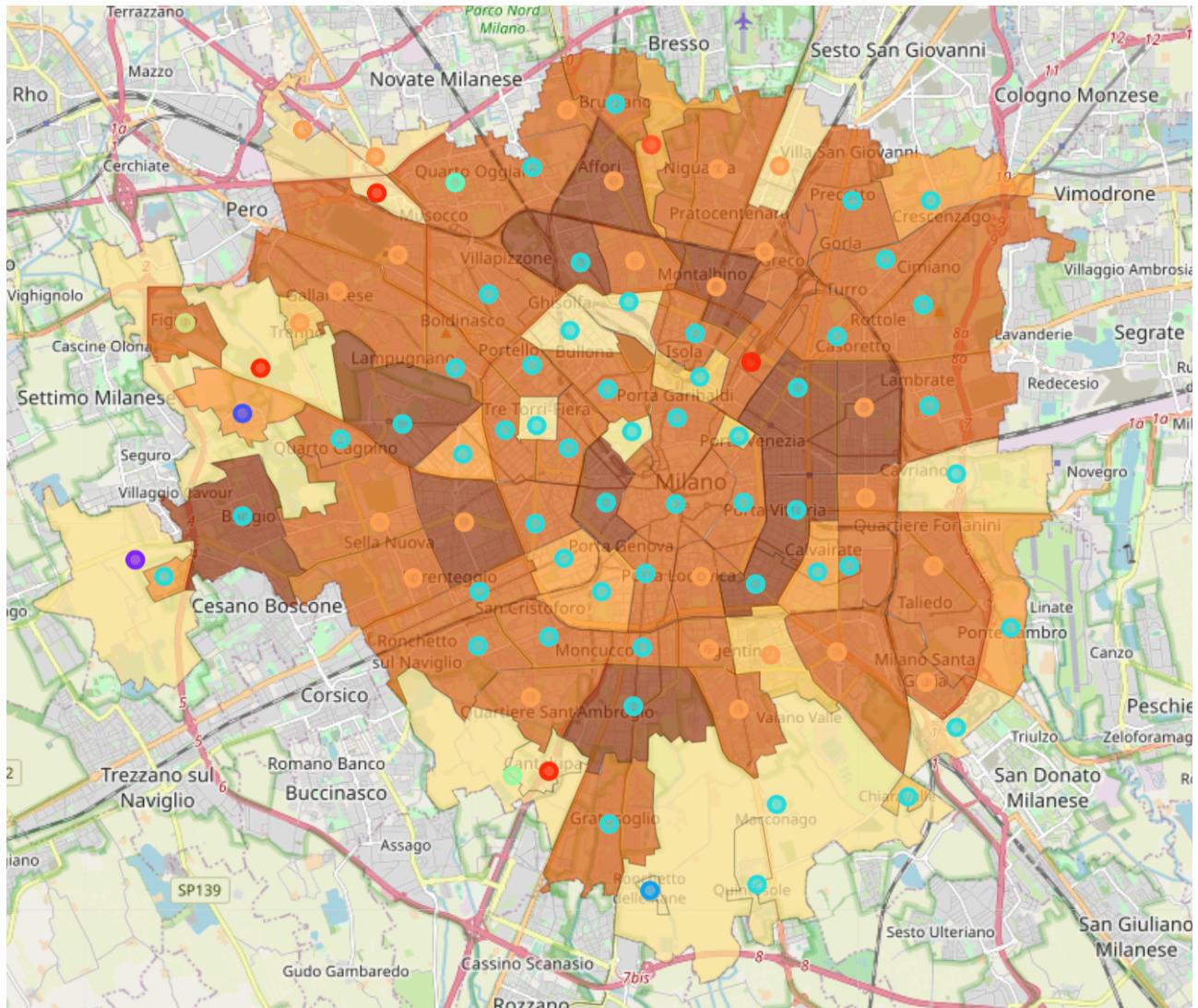


	NIL	ID_NIL	Latitude	Longitude
0	ADRIANO	17	45.511567	9.244967
1	AFFORI	80	45.514848	9.172868
2	BAGGIO	55	45.461216	9.088377
3	BANDE NERE	52	45.460334	9.138853
4	BARONA	46	45.432494	9.154109
5	BICOCCA	15	45.517181	9.210563
6	BOVISA	77	45.501838	9.165442
7	BOVISASCA	81	45.516937	9.154213
8	BRERA	2	45.476953	9.187232
9	BRUZZANO	83	45.526961	9.173317

TOP 10 NEIGHBOURHOOD

We can notice that the top 10 neighbourhoods are located around the city center, but close to it. Thats quite normal in a city where the city center is commercial-oriented.

Analysing which venues is present in a given NIL, we obtained 10 clusters:



The top 10 neighbourhoods in term of Family-Friendly Index fall in only two clusters: 8 (turchese) and 4 (orange).

The cluster 4 will be chosen by parents who prefer to live in a place characterised by an high presence of Supermarkets, Parks, Gardens and easy Italian restaurant such as Pizza Place.

The cluster 8 will best fits people who prefer a multicultural environment, with a high presence of Ethnic restaurants, Pubs and Café.

# CONCLUSION

We obtained good results, but some improvements are needed in order to get a more detailed analysis:

- Enrichment of Family-Friendly Index with more data such as:
  - Pharmacies
  - Playgrounds
  - Metro, Bus, Tram station
  - Shops for kids (Toys, Cloths)
- Clustering algorithm investigation: we used K-Means, but probably a density approach would give better results (for example DBSCAN).