# Capstone Project

# The Battle of Neighborhoods

Find the best place to open a restaurant in Milan

BY DIEGO LAZZARINI

# Summary

# Introduction

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the **Foursquare** location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

1. In Module 3, we explored **New York City** and the city of **Toronto** and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea.

2. In a city of your choice, if someone is looking to open a **restaurant**, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they set up their office?

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

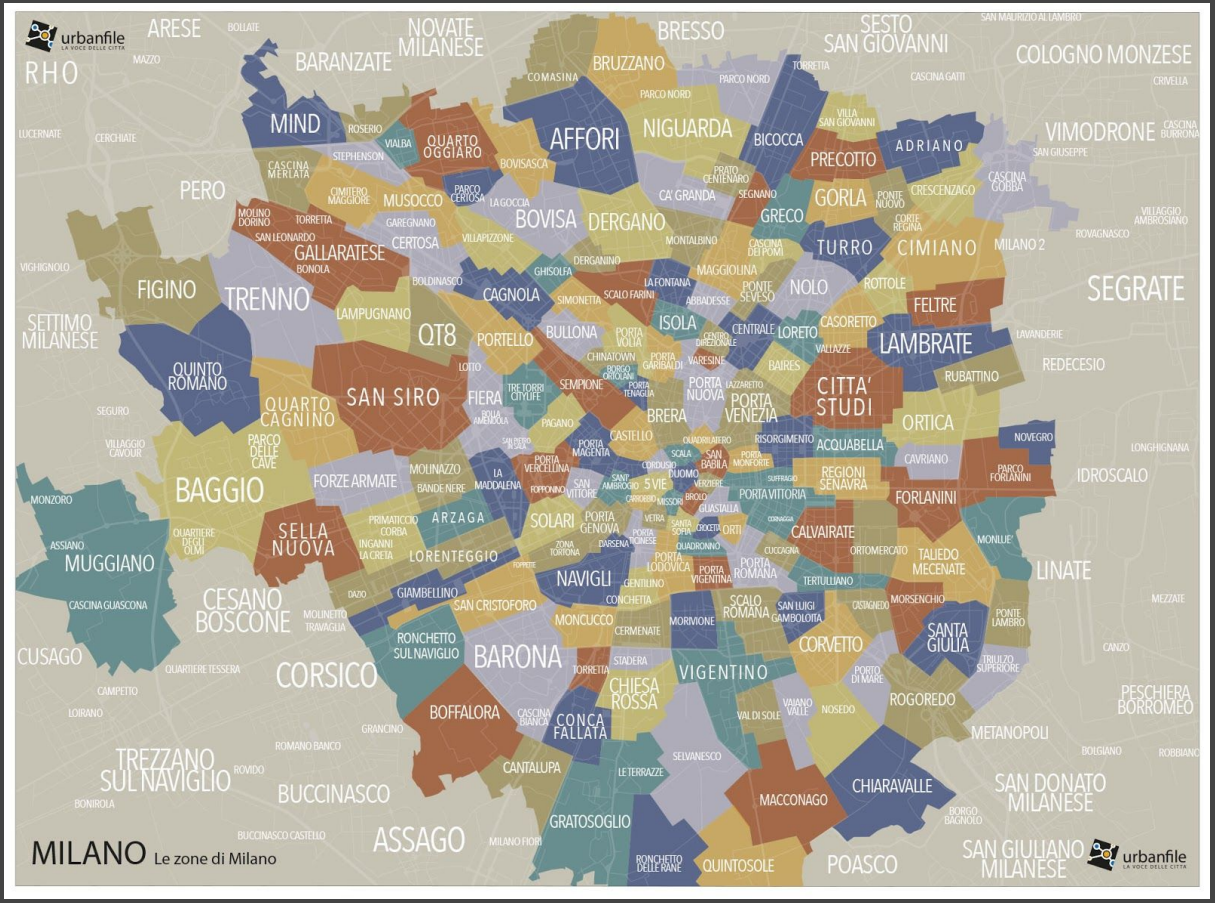
So I decided to try to answer this simple question:  where would you recommend to open a new restaurant?

# Business problem

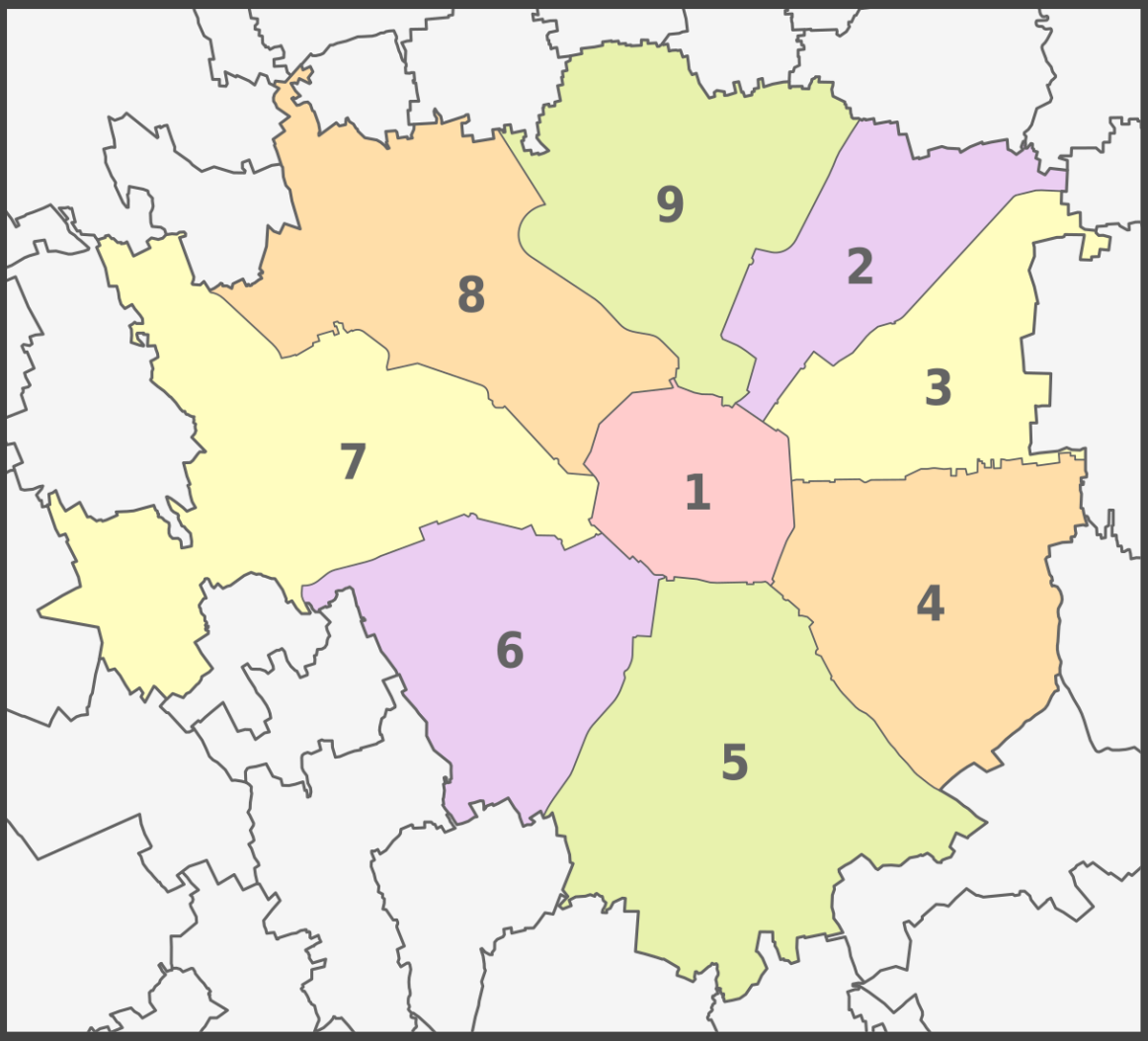
The city chosen to answer the initial question is **Milan** a city in northern Italy, capital of Lombardy, and the second-most populous city in Italy after Rome. Its continuously built-up urban area, that stretches well beyond the boundaries of the administrative metropolitan city, is the fourth largest in the EU with 5.27 million inhabitants.

Milan is considered a leading alpha global city, with strengths in the field of the art, commerce, design, education, entertainment, fashion, finance, healthcare, media, services, research and tourism. Its business district hosts Italy's stock exchange (Italian: Borsa Italiana), and the headquarters of national and international banks and companies. In terms of GDP, it has the second-largest economy among EU cities after Paris, and is the wealthiest among EU non-capital cities. Milan is also considered part of the Blue Banana and one of the "Four Motors for Europe".

Let’s see how many neighborhood there are and how they are distributed:



As you can see there are many of them, so the town is also divided in districts (municipi):




After this short **presentation**, I suppose that the city of Milan is place with a great competition, especially, if you want to **open** a **restaurant** so I would like to help a possible stakeholder to understand better the town and the market with useful insights.

## Target audience

1. A business entrepreneur that wants open a new restaurant in Milan.
2. Business Analyst or Data Scientists, who wish to analyze the neighborhoods of Milan using python, jupyter notebook and some machine learning techniques.
3. Someone curious about data that want to have an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.

# Data section

First of all we need some information about the area of Milan such as borough\districts, population, latitude\longitude etc... so I think Wikipedia is the first place to take a look:

https://it.wikipedia.org/wiki/Municipi_di_Milano

The borough are 9 with these coordinates:

| | Borough | Name | Area(km2) | Population(31/12/2018) | Population_Density(km2) | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| 0 | 1 | Centro storico | 967 | 98 531 | 10 189 | 45.471282 | 9.184999 |
| 1 | 2 | Stazione Centrale, Gorla, Turro, Greco, Cresce... | 1258 | 162 090 | 12 884 | 45.486117 | 9.203635 |
| 2 | 3 | Città Studi, Lambrate, Venezia | 1423 | 144 110 | 10 127 | 45.482506 | 9.241047 |
| 3 | 4 | Vittoria, Forlanini | 2095 | 161 551 | 7 711 | 45.431573 | 9.244738 |
| 4 | 5 | Vigentino, Chiaravalle, Gratosoglio | 2987 | 126 089 | 4 221 | 45.416987 | 9.238333 |
| 5 | 6 | Barona, Lorenteggio | 1828 | 151 291 | 8 276 | 45.440087 | 9.155924 |
| 6 | 7 | Baggio, De Angeli, San Siro | 3134 | 175 465 | 5 598 | 45.461244 | 9.089917 |
| 7 | 8 | Fiera, Gallaratese, Quarto Oggiaro | 2372 | 188 367 | 7 941 | 45.515925 | 9.140196 |
| 8 | 9 | Stazione Garibaldi, Niguarda | 2112 | 187 773 | 8 890 | 45.516888 | 9.191866 |

Now I need to find a list of all the **neighborhood** with the correspondent **borough**. Unfortunately the wikipedia tables aren't up to date so I found this paper from the official website of Milan:

https://www.pgt.comune.milano.it/sites/default/files/allegati/NIL_Intro.pdf

**Elenco schede NIL per i municipi**

**Municipio 1**
1. Duomo
2. Brera
3. Giardini Porta Venezia
4. Guastalla
7. Magenta- San Vittore
8. Parco Sempione
(5. Vigentina)
(6. Ticinese)
(68. Pagano)
(69. Sarpi)

**Municipio 2**
10. Stazione Centrale - Ponte Seveso
16. Gorla - Precotto
17. Adriano
19. Padova - Turro - Crescenzago

**Municipio 5**
5. Porta Vigentina - Porta Lodovica
6. Porta Ticinese - Conca del Naviglio
36. Scalo Romana
34. Chiaravalle
37. Morivione
38. Vigentino - Q.re Fatima
39. Quintosole
40. Ronchetto delle Rane
41. Gratosoglio - Q.re Missaglia - Q.re Terrazze
42. Stadera - Chiesa Rossa - Q.re Torretta - Conca Fallata
43. Tibaldi
85. Parco delle Abbazie
86. Parco dei Navigli
(47. Cantalupa)

**Municipio 8**
59. Tre Torri
64. Trenno
65. Q.re Gallaratese - Q.re San Leonardo - Lampugnano
66. QT8
67. Portello
68. Pagano
69. Sarpi
70. Ghisolfa
71. Villapizzone - Cagnola - Boldinasco
72. Maggiore - Musocco - Certosa
73. MIND - Cascina Triulza
74. Roserio
75. Stephenson
76. Quarto Oggiaro - Vialba - Musocco
(88. Parco Bosco in città)

Scraping the pdf file was impossible, so I created and uploaded this dataset on github: https://raw.githubusercontent.com/lazzarusd/Coursera_Capstone/master/file/Milano_Municipi_NIL.csv

This is a sample:

| | Num_NIL | NIL | Municipio | prezzo_mq |
|---|---|---|---|---|
| 1 | 17 | Adriano | 2 | € 2.800 /m² |
| 2 | 80 | Affori | 9 | € 2.350 /m² |
| 3 | 87 | Assiano | 7 | € 2.400 /m² |
| 4 | 55 | Baggio - Q.re degli Olmi - Q.re Valsesia | 7 | € 2.400 /m² |
| 5 | 52 | Bande Nere | 6 | € 3.857 /m² |
| 6 | 46 | Barona | 6 | € 3.250 /m² |

**Note**: the information about average land price is taken from those two websites (national reference points for the real estate market in Italy):
https://www.immobiliare.it/mercato-immobiliare/lombardia/milano/
https://www.mercato-immobiliare.info/lombardia/milano/milano.html

For the **final step**, I need to get the coordinates of every neighborhood.
Fortunately the statistics office of Milan created a very interesting portal about open data: https://dati.comune.milano.it/ and I found what I was looking for: a shape file (**geojson**).

https://dati.comune.milano.it/dataset/e8e765fc-d882-40b8-95d8-16ff3d39eb7c/resource/9c4e0776-56fc-4f3d-8a90-f4992a3be426/download/ds964_nil_wm.geojson

| | ID_NIL | NIL | Valido_dal | Valido_al | Fonte | Shape_Length | Shape_Area | OBJECTID | geometry |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 48 | RONCHETTO SUL NAVIGLIO - Q.RE LODOVICO IL MORO | 05/02/2020 | Vigente | Milano 2030 - PGT Approvato | 8723.368714 | 2.406306e+06 | 89 | POLYGON ((9.15422 45.43775, 9.15274 45.43887, ... |
| 1 | 64 | TRENNO | 05/02/2020 | Vigente | Milano 2030 - PGT Approvato | 3309.998800 | 4.896921e+05 | 90 | POLYGON ((9.10623 45.49016, 9.10591 45.49084, ... |
| 2 | 67 | PORTELLO | 05/02/2020 | Vigente | Milano 2030 - PGT Approvato | 3800.750663 | 9.096022e+05 | 91 | POLYGON ((9.15636 45.48785, 9.15495 45.48852, ... |
| 3 | 81 | BOVISASCA | 05/02/2020 | Vigente | Milano 2030 - PGT Approvato | 7105.469715 | 1.578028e+06 | 92 | POLYGON ((9.16803 45.52234, 9.16763 45.52272, ... |
| 4 | 84 | PARCO NORD | 05/02/2020 | Vigente | Milano 2030 - PGT Approvato | 11741.717005 | 1.532331e+06 | 93 | POLYGON ((9.20040 45.52848, 9.20028 45.52846, ... |

After some steps of data cleaning and data preparation, the final result is:

| | Id | Neighborhood | Borough | Population(31/12/2018) Borough | Average Price(€/sm) | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| 0 | 48 | Ronchetto Sul Naviglio - Q.Re Lodovico Il Moro | 6 | 151 291 | € 2.563 /m² | 45.438460 | 9.137260 |
| 1 | 64 | Trenno | 8 | 188 367 | € 2.350 /m² | 45.492822 | 9.101675 |
| 2 | 67 | Portello | 8 | 188 367 | € 4.300 /m² | 45.484490 | 9.153947 |
| 3 | 81 | Bovisasca | 9 | 187 773 | € 2.000 /m² | 45.517433 | 9.156731 |
| 4 | 84 | Parco Nord | 9 | 187 773 | € 6.800 /m² | 45.523514 | 9.184235 |

Now I'm ready to use the **foursquare API** https://developer.foursquare.com/docs/places-api/