Capstone Project - The Battle of Neighborhoods

Finding the Best Place to Open A Restaurant in Milan

KHALID ALRIFAI

Table of Contents

INSTRUCTION

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.

INTRODUCCTION

I am currently living in the Long Island City neighborhood in New York City, which is in the Queens borough. I like my current neighborhood very much because of its multiculturalism, its venues and because of the density of venues. I have been living there for several years and really enjoy the surrounding.

Recently, I have been given a job offer in Milan, Italy. To work in this place, I must move to Brera Town in the Downtown of Milan. I have never been in Italy before. Therefore, it is a huge step for me to move there. I also have family. So, I want to make sure that there are enough high rated high schools in Brera Town. In addition to that, my family really like to eat Italian food.

I once lived in the Allapattah neighborhood in Miami, USA. I did not like it at all. So, before I move to Milan, Italy, I want to be sure, that the Brera Town neighborhood is more like the Long Island City neighborhood and not like Allapattah. If Brera Town is more like Long Island City and not like Allapattah, I will move to Milan, Italy. However, If Brera Town turns out to be more like Allapattah, which I did not like at all, I will decline the job offer and stay in Long Island City.

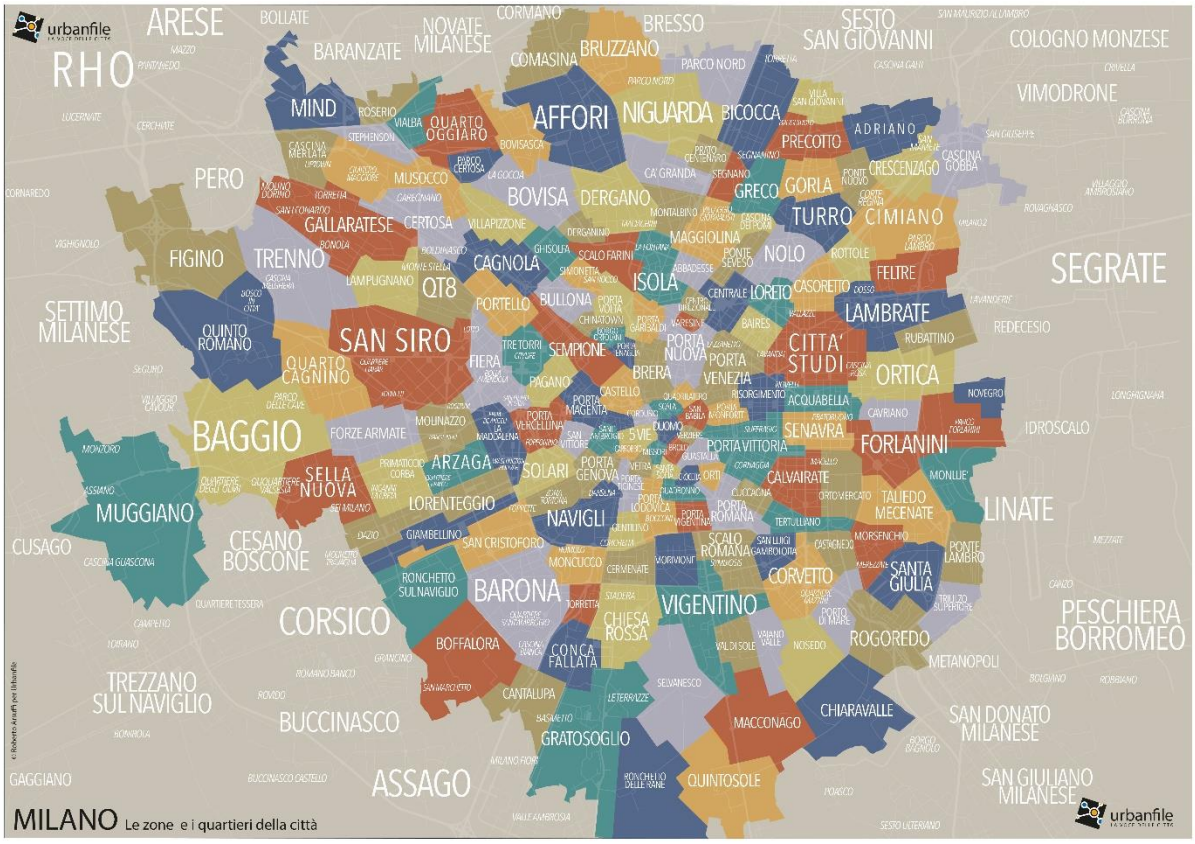
Going with the choice where I will be moving to Italy. I would like to start a business there and the question is: Where would you recommend opening a new restaurant?

BUSINESS PROBLEM

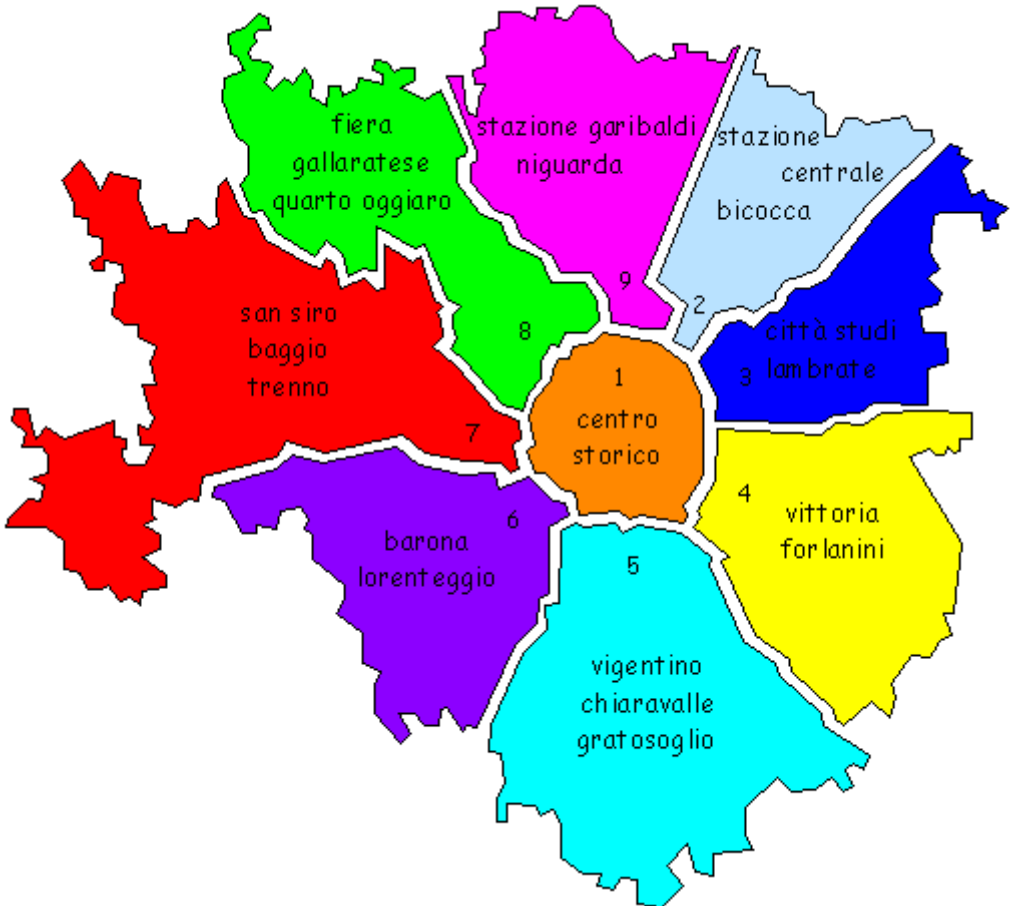
The city chosen to answer the initial question (Where would you recommend to opening a new restaurant?) is **Milan** a city in northern Italy, capital of Lombardy, and the second-most populous city in Italy after Rome. Its continuously built-up urban area, that stretches well beyond the boundaries of 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 neighbourhoods 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.

Tools:

This project will make use of the programming language Python. Therefore, the following API's and libraries will be used:

1. **Foursquare API:**

   This API will be the primal source of data. The Foursquare compony gathers data of a huge variety of places and stores them in a database. The Foursquare API will be used to connect to the Foursquare database and get the needed information and data.

2. **Clustering techniques:**

   To compare different neighborhoods, this project will use different clustering approaches. By building models, we can therefore categorize the St. James Town neighborhood and compare it to other neighborhoods. K-Means will be one of the used algorithms.

3. **Libraries:**

   Pandas: For creating Data Frames to store / manipulate data and information.
   NumPy: As fundamental package for scientific computing.
   Folium: Python's visualization library to generate interactive leaflet maps.
   Scikit Learn: For developing different clustering / classification models.
   Matplotlib: For visualization of data.
   Geocoder: To retrieve locational Coordinates.
   JSON: To handle JSON files.

Data Section:

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

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

The boroughs 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 are not 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

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
(11. Isola)
(12. Maciachini-Maggiolina)
(13. Greco - Segnano)
(20. Loreto - Casoretto - NoLo)

Municipio 3
18. Cimiano - Rottole - Q.re Feltre
21. Buenos Aires - Porta Venezia - Porta Monforte
22. Città studi
23. Lambrate - Ortica
(20. Loreto)
(24. Parco Forlanini - Cavriano)

Municipio 4
25. Corsica
26. XXII Marzo
28. Umbria - Molise - Calvairate
29. Ortomercato
30. Taliedo - Morsenchio - Q.re Forlanini
31. Monluè - Ponte Lambro
32. Triulzo Superiore
33. Rogoredo - Santa Giulia
35. Lodi-Corvetto
(27. Porta Romana)

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 6
44. Porta Ticinese - Conchetta
45. Moncucco - San Cristoforo
46. Barona
47. Cantalupa
48. Ronchetto sul Naviglio - Q.re Lodovico il Moro
49. Giambellino
50. Porta Genova
52. Bande Nere
53. Lorenteggio
86. Parco dei Navigli
(51. Washington)

Municipio 7
51. Porta Magenta
54. Muggiano
55. Baggio - Q.re degli Olmi - Q.re Valsesia
56. Forze Armate
57. San Siro
58. De Angeli-Monte Rosa
60. Stadio - Ippodromi
61. Quarto Cagnino
62. Quinto Romano
63. Figino
87. Assiano
88. Parco Bosco in città
(68. Pagano)

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)

Municipio 9
9. Porta Garibaldi - Porta Nuova
11. Isola
14. Niguarda - Ca' Granda - Prato Centenaro - Q.re Fulvio Testi
15. Bicocca
77. Bovisa
78. Farini
79. Dergano
80. Affori
81. Bovisasca
82. Comasina
83. Bruzzano
84. Parco Nord
(12. Maciachini-Maggiolina)
(13. Greco)

Scraping the pdf file was impossible, so I created and uploaded this dataset on GitHub: https://raw.githubusercontent.com/khalidalrifai/IBM-Data-Scientists-Professional/master/Milano_Municipi_NIL.csv

This is a sample:

	Num_Neighborhood	Neighborhood	Borough	Average_price_sm
0	17	Adriano	2	€ 2.800 /m²
1	80	Affori	9	€ 2.350 /m²
2	87	Assiano	7	€ 2.400 /m²
3	55	Baggio - Q.re degli Olmi - Q.re Valsesia	7	€ 2.400 /m²
4	52	Bande Nere	6	€ 3.857 /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 result is:

	Id	Neighborhood	Borough	Population(31/12/2018) Borough	Average Price(€/sm)	Latitude	Longitude	Geometry
0	48	Ronchetto Sul Naviglio - Q.Re Lodovico Il Moro	6	151 291	€ 2.563 /m²	45.438460	9.137260	POLYGON ((9.154221025150809 45.43775166985864,...
1	64	Trenno	8	188 367	€ 2.350 /m²	45.492822	9.101675	POLYGON ((9.10622707748691 45.4901620817868, 9...
2	67	Portello	8	188 367	€ 4.300 /m²	45.484490	9.153947	POLYGON ((9.156361803387014 45.48785426899354,...
3	81	Bovisasca	9	187 773	€ 2.000 /m²	45.517433	9.156731	POLYGON ((9.168034631377399 45.5223394320893, ...
4	84	Parco Nord	9	187 773	€ 6.800 /m²	45.523514	9.184235	POLYGON ((9.200403184652478 45.5284767228649, ...

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