Monza e Brianza restaurant choice

"Applied Data Science Capstone Project"
Alessandro Fumagalli
December, 2020

The problem

Monza is the capital of the "*Monza e Brianza*" province (MB). Its territory is placed between Milan and the Lake of Como district

- It hosts different fairs including the Italian Formula 1 GP

- It attracts many tourists and business men

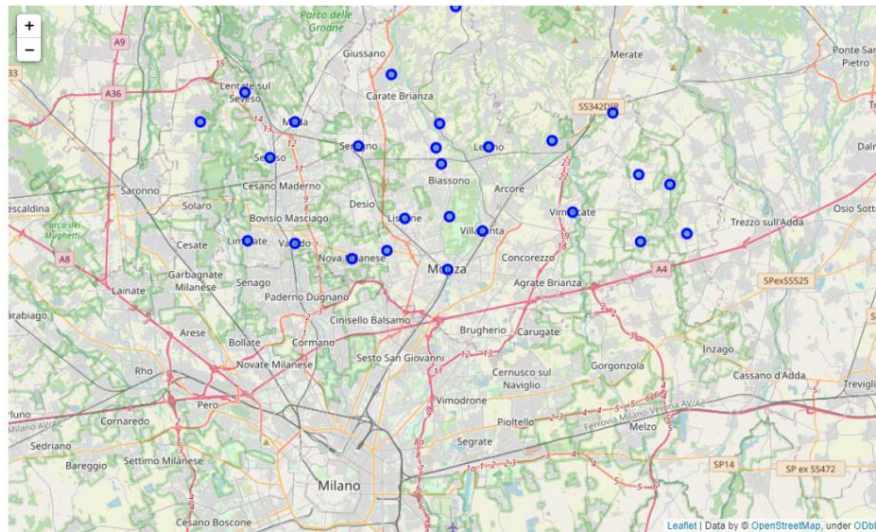*Part of MB province as seen from ISS by astronaut Reid Wiseman*

<u>The business question is</u>: how are distributed the restaurants types in MB province, so can meet better foreign people taste and gastronomy palates?

Data retrieval & cleaning

Retrieved the list of all 27 municipalities that made up MB province from [here](#)

Consider geographic data with municipality name.
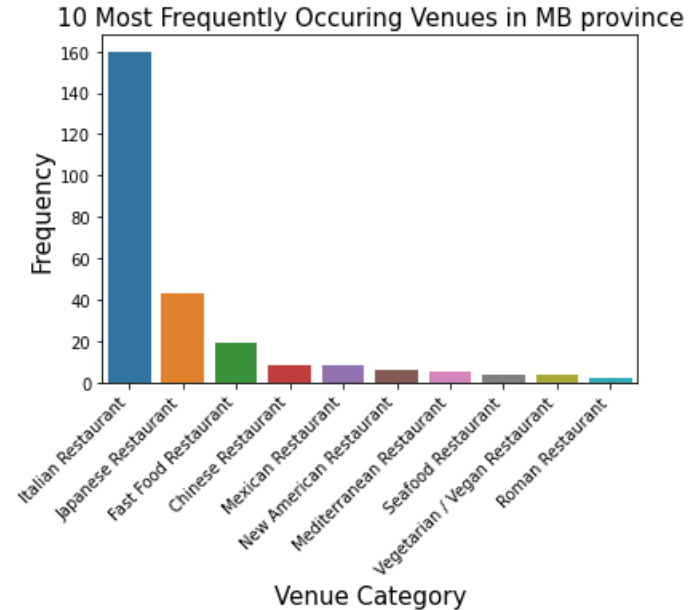
Remaining info has been drop

With [Foursquare](#) retrieved all venues in MB province and filtered by restaurant types.

- A deeper inspection figure out that 50 restaurants has been misclassified (like Japanese, Australian, Argentinian, and others)
- Data has been post processed and the type has been assigned accordingly

Classification Model

A first classification considered the most frequently occurring restaurant types in all MB province
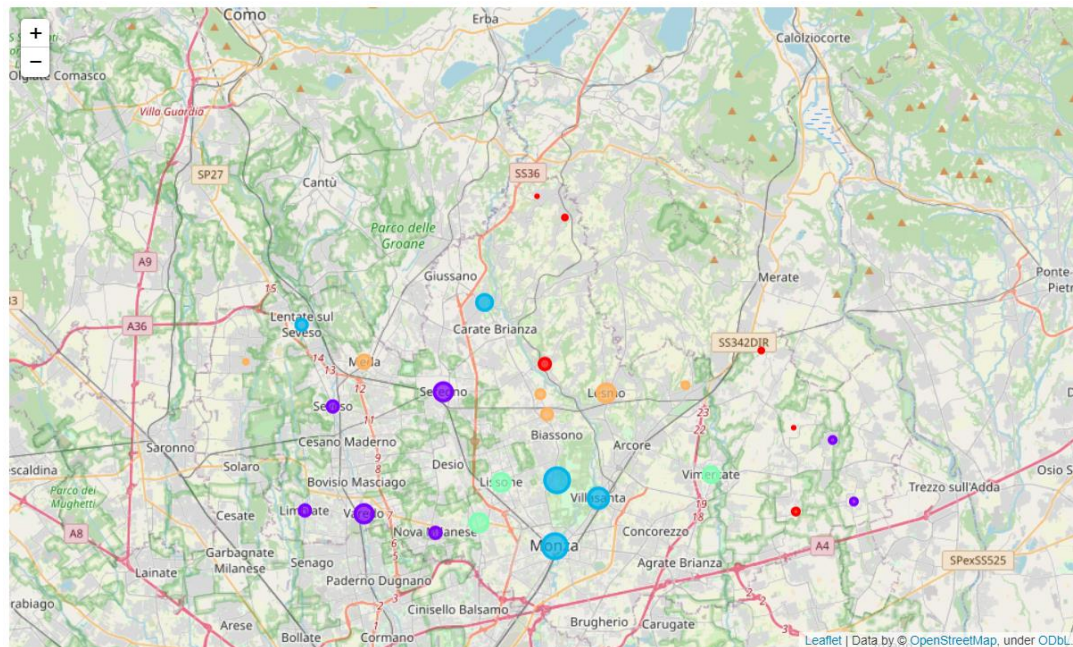
At first sight the Italian restaurants type goes for the most, but also Japanese and fast food are the runners up.

10 Most Frequently Occuring Venues in MB province

How are they distributed an clustered in the province?

Clustering analysis

A *second classification* proposed is to extract from data the frequency of each category of restaurants in each city district.

	Neighborhood	American Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant	Italian Restaurant	Japanese Restaurant	Kebab Restaurant	Lo... Rest...
0	Lentate sul Seveso	0.111111	0.000000	0.000000	0.000000	0.000000	0.222222	0.000000	0.000000	0.555556	0.111111	0.000000	0.0
1	Lesmo	0.000000	0.000000	0.000000	0.133333	0.000000	0.000000	0.000000	0.000000	0.866667	0.000000	0.000000	0.0
2	Limbiate	0.000000	0.000000	0.000000	0.222222	0.000000	0.111111	0.000000	0.000000	0.333333	0.333333	0.000000	0.0
3	Lissone	0.000000	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.400000	0.200000	0.000000	0.0
4	Macherio	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.800000	0.000000	0.000000	0.0
5	Meda	0.000000	0.000000	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000	0.727273	0.181818	0.000000	0.0
6	Mezzago	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.250000	0.500000	0.000000	0.0
7	Misinto	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.666667	0.000000	0.000000	0.0
8	Monza	0.000000	0.000000	0.000000	0.047619	0.000000	0.047619	0.000000	0.000000	0.571429	0.190476	0.000000	0.0
9	Muggió	0.000000	0.000000	0.000000	0.000000	0.066667	0.066667	0.000000	0.000000	0.266667	0.333333	0.000000	0.0
10	Nova Milanese	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.222222	0.333333	0.000000	0.0
11	Ornago	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0
12	Renate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0
13	Roncello	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.250000	0.500000	0.000000	0.0
14	Ronco Briantino	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0
15	Seregno	0.000000	0.066667	0.000000	0.000000	0.000000	0.133333	0.066667	0.000000	0.333333	0.266667	0.066667	0.0
16	Seveso	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.666667	0.333333	0.000000	0.0
17	Sovico	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.857143	0.000000	0.000000	0.0
18	Sulbiate	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0
19	Triuggio	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0
20	Usmate Velate	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000	0.833333	0.000000	0.000000	0.0
21	Varedo	0.000000	0.000000	0.000000	0.066667	0.000000	0.200000	0.000000	0.000000	0.400000	0.133333	0.000000	0.0
22	Vedano al Lambro	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.545455	0.136364	0.000000	0.0
23	Veduggio con Colzano	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.0

Finally a k-Means algorithm has been implemented (k=5) to identify clusters.
- ***See the map aside***

Results of the model

Cluster 1: Italian, Vegetarian and Argentinian

	Neighborhood	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
11	Ornago	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
12	Renate	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
14	Ronco Briantino	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
18	Sulbiate	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
19	Triuggio	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
23	Veduggio con Colzano	Italian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant

Cluster 2: Italian, Japanese and Fast Food

	Neighborhood	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
2	Limbiate	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Fast Food Restaurant	Argentinian Restaurant	Australian Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant	Greek Restaurant
6	Mezzago	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant	Greek Restaurant
10	Nova Milanese	Japanese Restaurant	Italian Restaurant	Fast Food Restaurant	New American Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant
13	Roncello	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant	Greek Restaurant
15	Seregno	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	Kebab Restaurant	Argentinian Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Lombard Restaurant	Mexican Restaurant
16	Seveso	Italian Restaurant	Japanese Restaurant	Falafel Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
21	Varedo	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	New American Restaurant	Chinese Restaurant	Mediterranean Restaurant	Seafood Restaurant	German Restaurant	Argentinian Restaurant	Australian Restaurant

Cluster 3: Italian, Fast Food and Japanese

	Neighborhood	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	Lentate sul Seveso	Italian Restaurant	Fast Food Restaurant	Japanese Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant	Greek Restaurant
8	Monza	Italian Restaurant	Japanese Restaurant	Seafood Restaurant	Chinese Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Roman Restaurant	Mexican Restaurant	Mediterranean Restaurant	Lombard Restaurant
22	Vedano al Lambro	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Seafood Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Modern European Restaurant	New American Restaurant	Lombard Restaurant	Kebab Restaurant
24	Verano Brianza	Italian Restaurant	Fast Food Restaurant	Japanese Restaurant	Mexican Restaurant	Seafood Restaurant	Greek Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant
25	Villasanta	Italian Restaurant	Chinese Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Argentinian Restaurant	Japanese Restaurant	New American Restaurant	Mediterranean Restaurant	Lombard Restaurant

Results of the model

Cluster 4: Italian, Japanese and Mexican

	Neighborhood	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
3	Lissone	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	New American Restaurant	Mediterranean Restaurant	Lombard Restaurant	Comfort Food Restaurant	Falafel Restaurant	German Restaurant	Argentinian Restaurant
9	Muggiò	Japanese Restaurant	Italian Restaurant	Mexican Restaurant	Roman Restaurant	Mediterranean Restaurant	Comfort Food Restaurant	Fast Food Restaurant	German Restaurant	Argentinian Restaurant	Australian Restaurant
26	Vimercate	Italian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Australian Restaurant	Chinese Restaurant	German Restaurant	Kebab Restaurant	American Restaurant	Lombard Restaurant	Mexican Restaurant

Cluster 5: Italian and Mediterranean (Spanish, Greek)

	Neighborhood	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	Lesmo	Italian Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
4	Macherio	Italian Restaurant	Paella Restaurant	New American Restaurant	Vegetarian / Vegan Restaurant	Greek Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant
5	Meda	Italian Restaurant	Japanese Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant
7	Misinto	Italian Restaurant	Greek Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant
17	Sovico	Italian Restaurant	Paella Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Australian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant
20	Usmate Velate	Italian Restaurant	Australian Restaurant	Vegetarian / Vegan Restaurant	Argentinian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant

Conclusion and future directions

In all clusters Italian restaurant is at first place: there are other choices where other restaurants types goes for more, in particular Japanese ones.

- The openings of these last types of restaurants have significant increase in MB province and the analysis confirm that.

The distribution of the restaurant types split the province in four macro areas: South, East-West, Center and North-Est. Your choice depends on where you are going to stay.

Here are some ideas for further improvement:

- include the ratings of the restaurants to shape the cluster.
- split the generic Italian restaurant into regional Italian ones (e.g. North, Center and South).
- Try to identify any correlation between restaurant type and local communities.