

CAPSTONE PROJECT: SUSHI RESTAURANT IN MILAN_ WEEK5

Riccardo Zannoni

05/04/2020

Table of Content:

1.Introduction & Business Understanding

2.Data

3.Methodology

4.Results

5.Discussion & Conclusion

1.Introduction & Business Understanding

Milan is probably the most international city in Italy and for few years one of the most trendy restaurant type is Sushi. For this reason the idea behind this problem is: if an investor wants to open a new Sushi restaurant, where would be the optimal location/neighborhood?

In such a problem many variables could play a vital role; then, in this notebook the key assumptions are:

- 1) We are considering 88 NIL (Nuclei di Identità Locale) which are the statistica neighborhood of Milan;
- 2) The locations (Latitude & Longitude) are going to be retrived from Geopy
- 3) Total Spending Power is the KPI we are interested as a proxy of the overall wealth of the NIL;
- 4) We will use Foursquare API to register key locations for each NIL;
- 5) To compare every NIL and for the clustering processes we are going to use the % of Sushi Restaurants on the overall restaurants

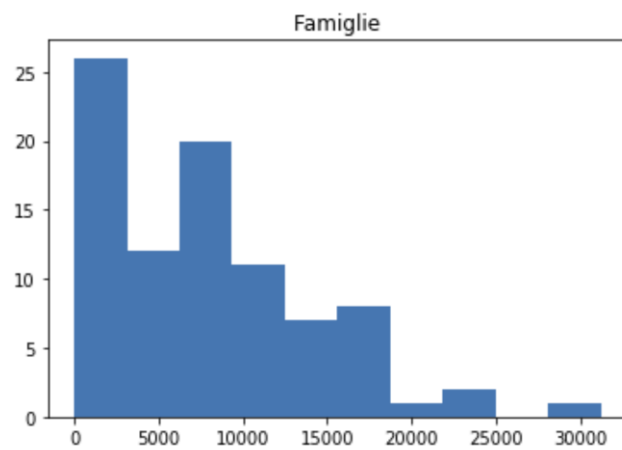
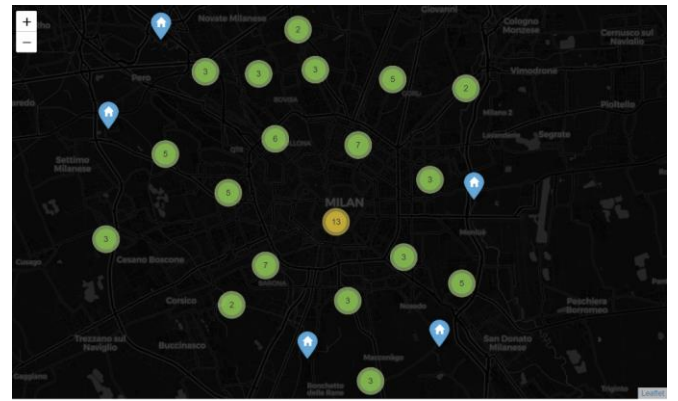
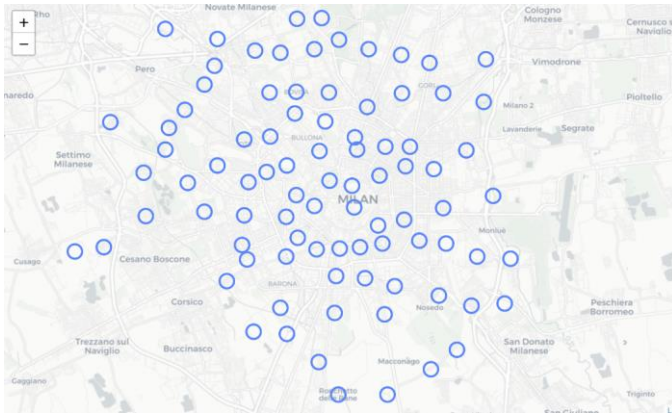
2.Data

For all the above assumption, I will deep dive regarding data sources, needs & missing information:

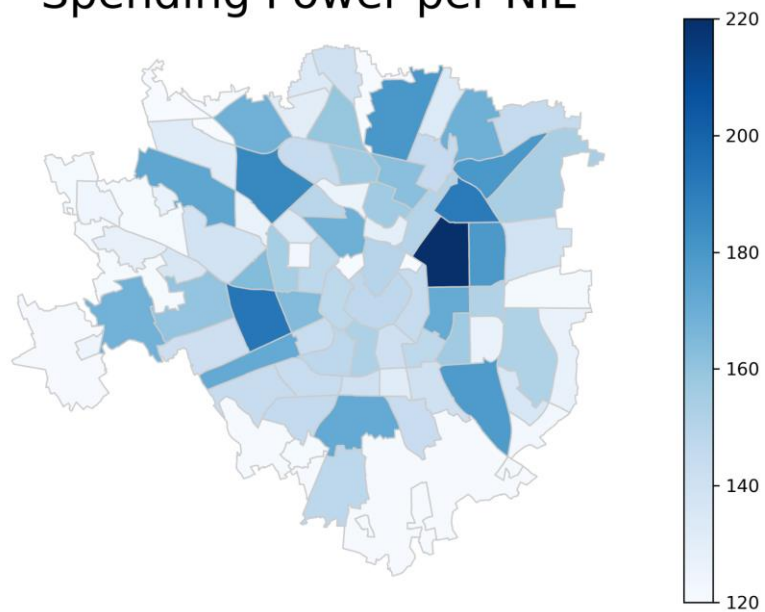
1. The municipal entity of Milan offer an amazing portal ([Open Data Milan](#)). From this, it was possible to retrieve the list of all the existing Neighborhoods. I discovered that this new classification is relatively new and substituted the older Municipal Districts. Thus, I believe that this notebook could be a first analysis exploring this classification;
2. The geographical Information were missing and couldn't be extrapolated somewhere. For this reason, I used Geopy library to retrieve Latitude & Longitude for each Neighborhood
3. Total Spending Power: this KPI was calculated considering the total numeber of Family per NIL & Consumption per Family for Grocery & Other Goods. The data were retrived again from ([Open Data Milan](#)), respectevly from ([Consumptions](#)) & ([Families](#))
4. During this Course, I created a developer Account at Fourquare which I used to retrieve all the necessary information. Following in this notebook, I cancelled all the lines which included my account info for Privacy Issues
5. Also for this KPI, I used all the info retrived from Fourquare and I indicated all the necessary steps in this Notebook

It follows a summary of the dataset, together with come visualisation and graphs:

	Municipio	Totale Abitanti	Famiglie	Comp. Fam	Total Spending Power	latitude	longitude	altitude
count	88.000000	88.000000	88.000000	88.000000	8.800000e+01	88.000000	88.000000	88.0
mean	5.295455	15855.386364	7902.636364	2.533498	2.901058e+08	45.470978	9.174152	0.0
std	2.609263	12944.493916	6546.487671	4.150605	2.403216e+08	0.031461	0.047447	0.0
min	1.000000	2.000000	2.000000	1.000000	7.342000e+04	45.401867	9.056185	0.0
25%	3.000000	4252.500000	2087.750000	1.970755	7.664130e+07	45.450038	9.141482	0.0
50%	5.000000	14750.000000	7432.000000	2.023845	2.728287e+08	45.472564	9.175946	0.0
75%	8.000000	23135.750000	11537.000000	2.083899	4.235233e+08	45.495764	9.208850	0.0
max	9.000000	62438.000000	31230.000000	40.761905	1.146453e+09	45.527369	9.262690	0.0



Spending Power per NIL



Source: Elaboration on Open Data Milan 2018

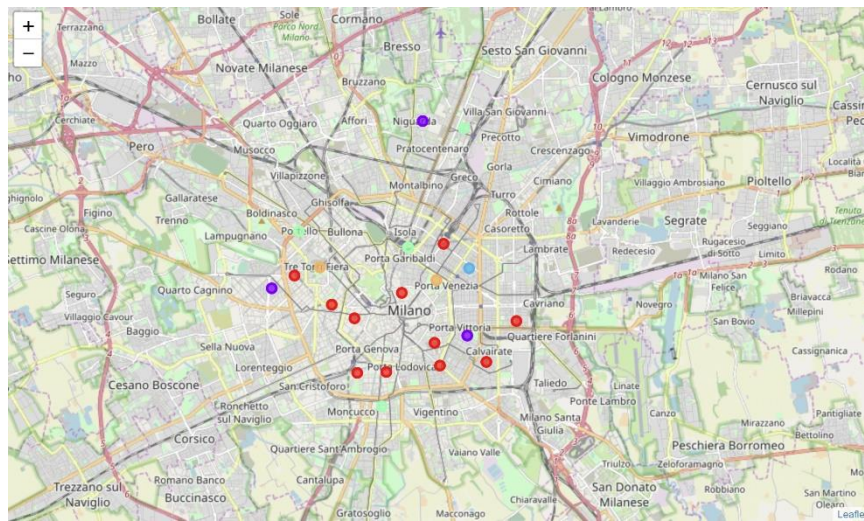
3.Methodology

Main techniques used for this analysis:

- Geocoding
- Venues Extraction (Foursquare API)
- One Hot Encoding
- Algorithm: K-means clustering

4.Results

It follows a screenshot of the final clustering result, together with a summary table of the clusters.



CLUSTER	TYPE	COLOR
0	This first cluster is characterized by Medium-Low Spending Power & Medium % of Sushi Restaurant	Red
1	This second cluster is characterized by Medium-High Spending Power & Medium-High % of Sushi Restaurant	Purple
2	This third cluster is characterized by High Spending Power & Low % of Sushi Restaurant. This Case is particularly interesting: due to its High Spending Power (it is the max among the 88 NIL), this Neighborhood represents itself a cluster	Blue
3	This cluster is characterized by Low Spending Power & Medium % of Sushi Restaurant	Green
4	This cluster is characterized by Low Spending Power & Medium-High % of Sushi Restaurant	Orange

5. Discussion & Conclusion

So we arrived at the end of this interesting (and fun) analysis. In this section, my goal is twofold: to further discuss the overall assumption of the model and some further development points; to wrap up and arrive at some kind of conclusion.

1. In this analysis, the overall assumption is that Total Spending Power is a proxy of the wealth of the people living in the Neighborhood and thus, the higher the better. As it is a Proxy, it could be improved: some examples that came to my mind (if info are available): total spending in restaurants per capita, average usage of restaurants, addition information regarding the type of families etc.
2. Let's recap the identified clusters: a) **CLUSTER 0**: Medium-Low Spending Power & Medium %; **CLUSTER 1**: Medium-High Spending Power & Medium-High %; **CLUSTER 2**: High Spending Power & Low %; **CLUSTER 3**: Low Spending Power & Medium %; **CLUSTER 4**: Low Spending Power & Medium-High %; In this case, we can make the assumption that it is necessary to have at least Medium in both Spending Power & % of Sushi Restaurants. The first part it is straightforward: the more the Spending power, the more likely people in that hood are going out to restaurants etc. Instead, the second part embodies the fact that if there are already some sushi restaurants in the area, this has a positive impact because people have already tested this type of cuisine and thus it is more likely that are going again (from a business point of view, this assumptions replicate the logics behind "followers" strategy rather than "first-movers").

Thus, narrowing down the analysis, it is possible to exclude Cluster 0, 2,3,4 and we can identify Cluster 1 as the best fit

Concluding, this analysis led to the identification of a Cluster which represents a first group of areas where to open a Sushi Restaurant. From this analysis is then necessary to deep dive in this areas to better understand the context, real estate prices, consumers habits etc. which are out of scope for this analysis