



IBM/COURSERA Applied Data Science Capstone Project

THE BATTLE OF NEIGHBORHOODS: SUSHI RESTAURANTS IN MILAN

RICCARDO ZANNONI

BUSINESS PROBLEM

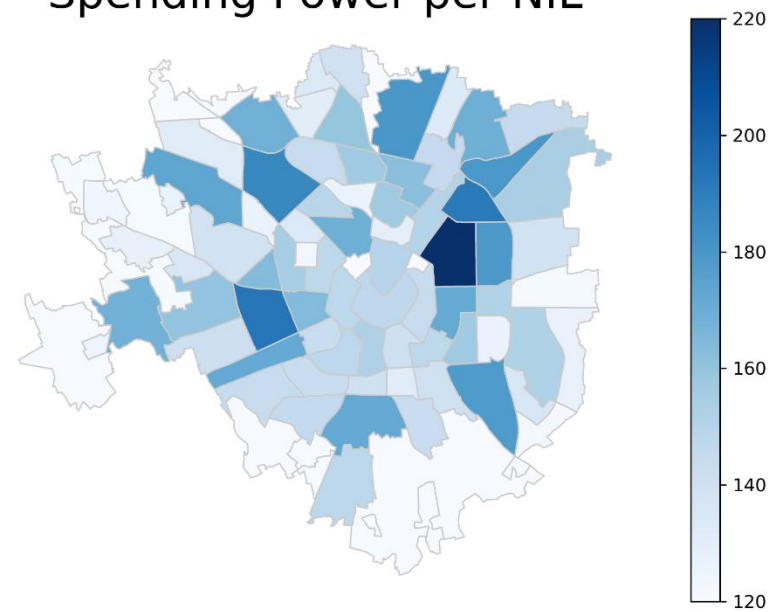
UNDERSTANDING OF THE
CONTEXT

- ▶ 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?

DATA

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

Spending Power per NIL



Source: Elaboration on Open Data Milan 2018

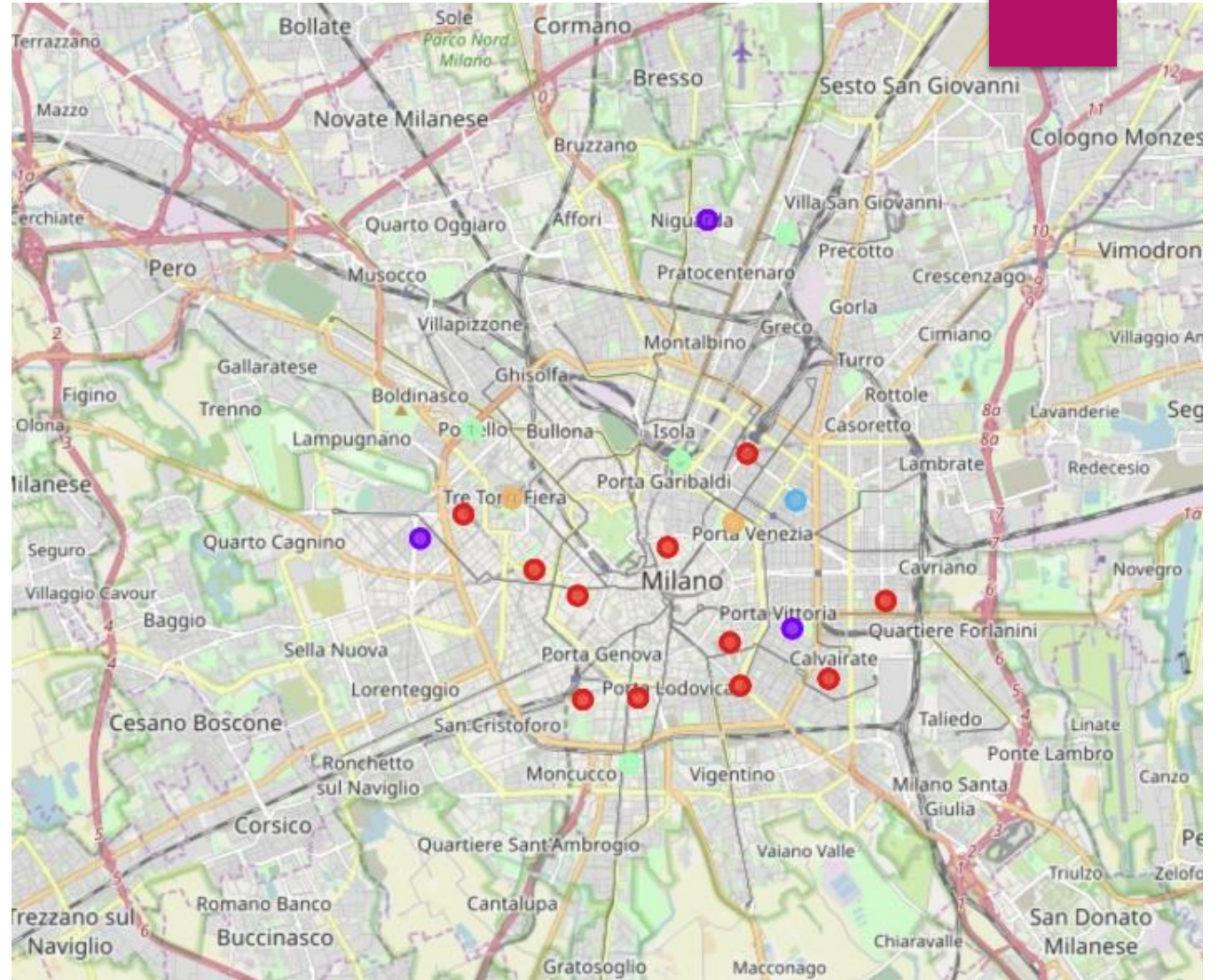
METHODOLOGY

Main techniques used for this analysis:

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

SOLUTION

- ▶ For this purpose it is possible to run machine learning models (clustering) in order to make an informed decision about potential neighborhoods where to open a new sushi restaurant



CONCLUSION

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

Main techniques used for this analysis:

- 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**