Applied Machine Learning for opening a new pizzeria in Naples

Riccardi Luca, Data Analyst @ Mopar (F.C.A.)

Introduction

Nowadays, opening a new activity is an investment that needs a deep study in order to not lose money. In this presentation, we will show a study focused on fininding the optimal place where opening a new pizzeria in the city of Naples, Italy.

This study has been done as Capstone project for the Data Science specialization by IBM. Report and notebook can be found on my GitHub page.

GitHub Repository

Report

Notebook

1. Pizza and Naples

An historic combination

Why pizza in Naples?

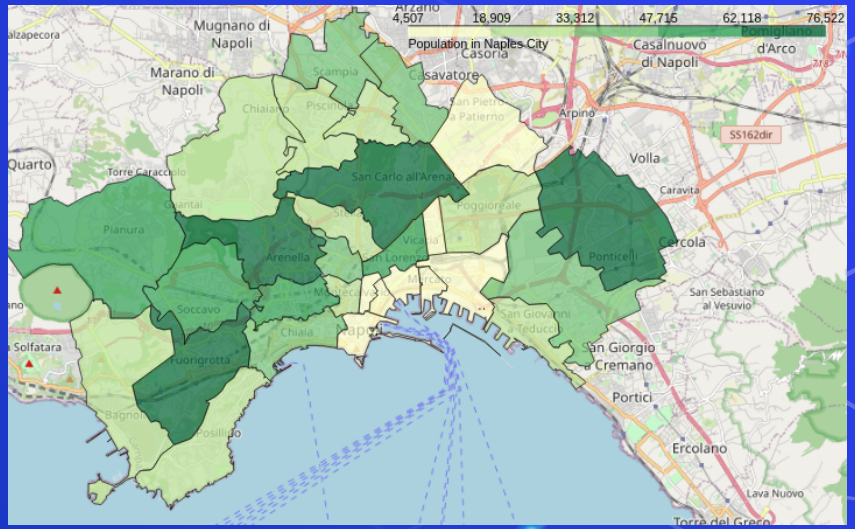
One of the most popular meal in the world. Someone says that it comes from Turkey, others from Italy. Pizza has radically changed the world, in particular in Naples, the city considered as the birthplace of that disc of dough with tomato souce, mozarella and fresh basil.

On the side you can see a beatuful pizza margherita with fresh basil and a small mozzarella on top.

2. Overview of the city

Why pizza in Naples?

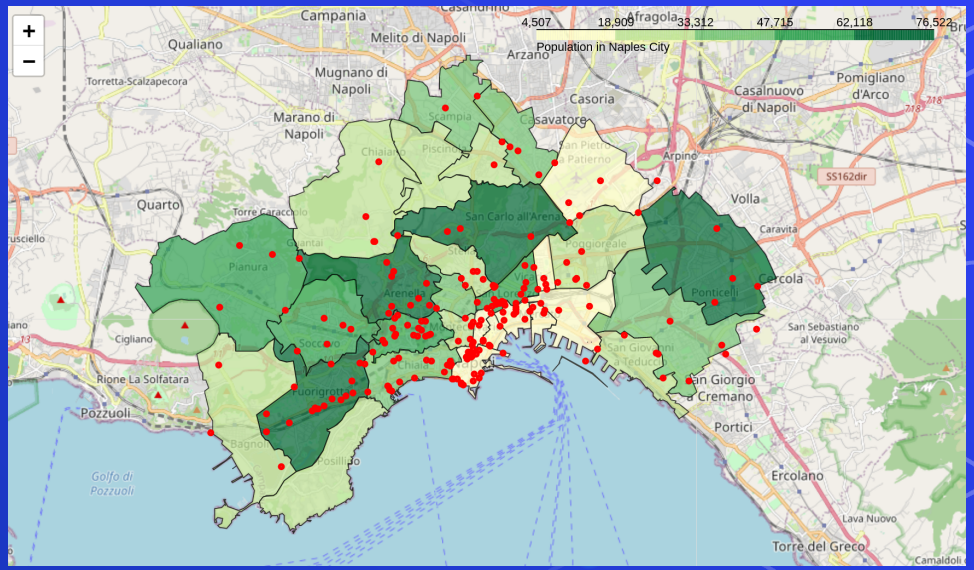
Naples is the the third city in Italy for number of people, with almost one million of residents. It is the capital of Campania region and it is one of the most important cultural center all over the Italy. The city is mainly divided in 10 municipalities, with in total 30 boroughs.



Why pizza in Naples?

Pizzeria are very popular in the city of Naples. You can sit in, or take-away a pizza and eat it on the seafront. There is also a special version so called wallet-pizza, that is a small folded up version.

On the side you find recognized pizza using Foursquare API



3. Methodology

K-means for optimal places

We will use k-means algorithm to find the best places where it is worth for opening a new pizzeria. We are going to clusterize the pizzeria in clusters that share similarities on nearby venues.

Feature engineering

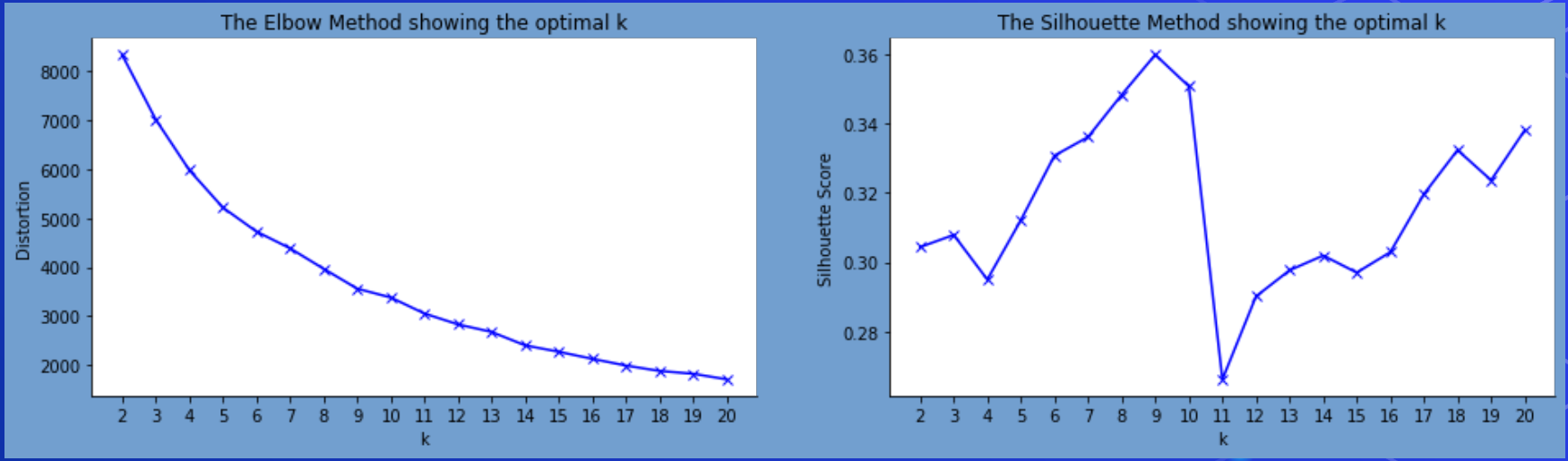
One hot encoding is performed as pre-processing of data.

	Center	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
0	4c191cad834e2d7f50e42980	Pizzeria La Notizia	4c191cad834e2d7f50e42980	40.835749	14.209840	Pizza Place
1	4c191cad834e2d7f50e42980	pizzeria gaetano genovesi	55a6c55b498e4f5ac2760547	40.835365	14.211542	Pizza Place
2	4c191cad834e2d7f50e42980	Carrefour Market	4e58d7ce2271886714ea723e	40.837883	14.214211	Supermarket
3	4c191cad834e2d7f50e42980	Krugel	4e43932e1fc752e9103fad70	40.838380	14.214197	Sandwich Place
4	4c191cad834e2d7f50e42980	Pizzeria La Caraffa	4d58427136d1721eceabf4b1	40.838522	14.214109	Pizza Place

Venue Category	Airport	Airport Terminal	Aquarium	Art Gallery	Art Museum	Asian Restaurant	Auto Garage	BBQ Joint	Bakery	Bar	Basketball Stadium	Bed & Breakfast
Center												
4b3b7359f964a520057425e3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4b519aa5f964a5208f5027e3	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
4b7d8f92f964a520d3c62fe3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
4b8ecf2cf964a5207f3833e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	1.0
4b92fe3af964a520382d34e3	0.0	0.0	0.0	1.0	2.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0

Model selection

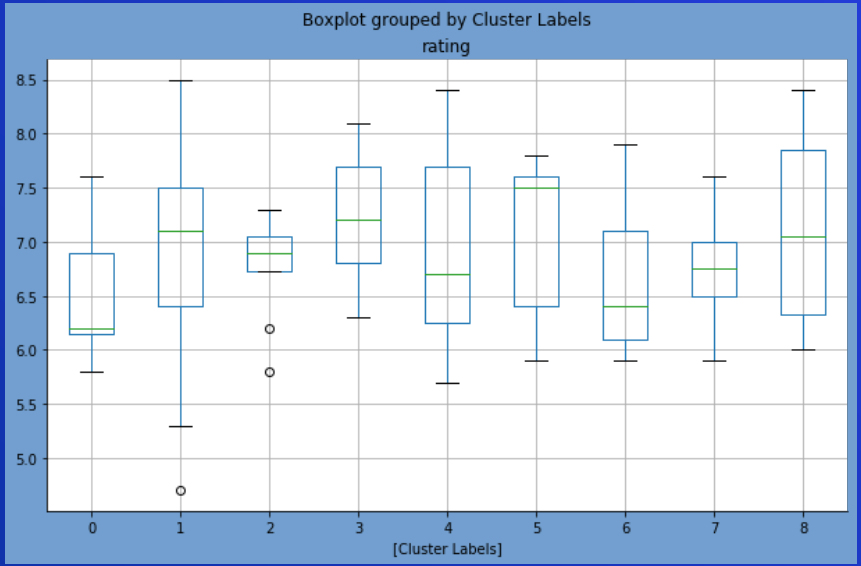
The parameter k is unknown, so both elbow and silhouette method are shown.



Best k is 9 since maximize silhouette score (elbow method does not give a clear indication).

Median rating for cluster

We can build a cluster rating, for example using median value

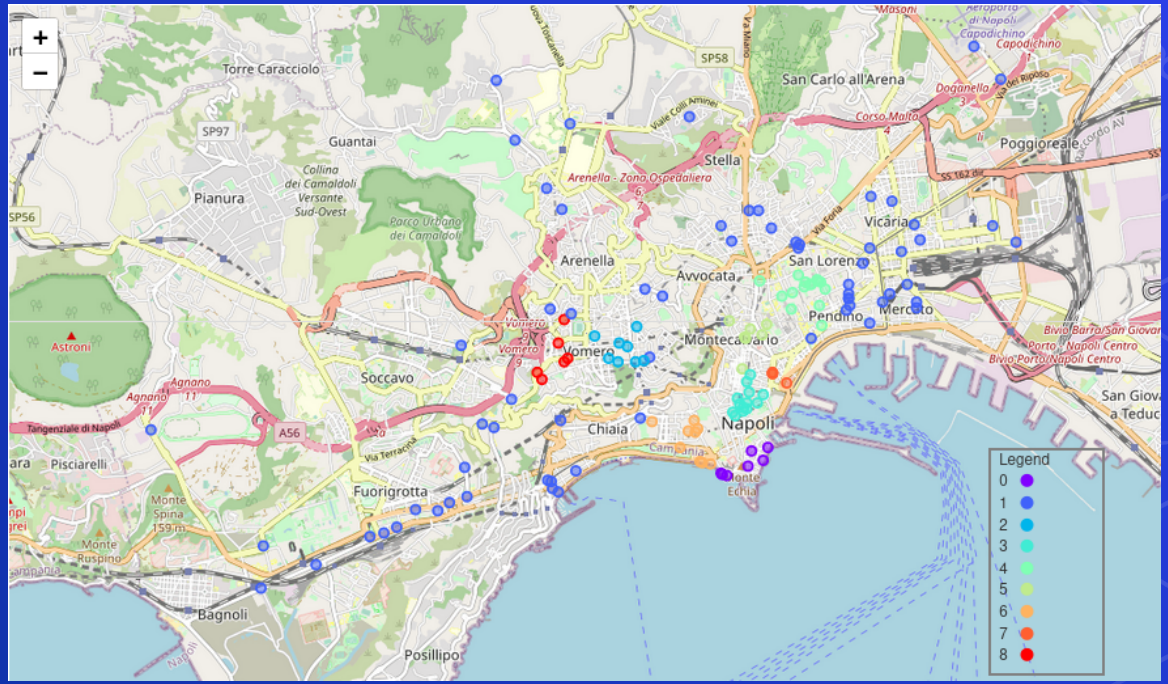


Cluster Labels	rating
5	7.50
3	7.20
1	7.10
8	7.05
2	6.90
7	6.75
4	6.70
6	6.40
0	6.20

Accordingly with our criteria, best pizzeria are in cluster 5

Cluster map

Montecalvario have the highest median rating.



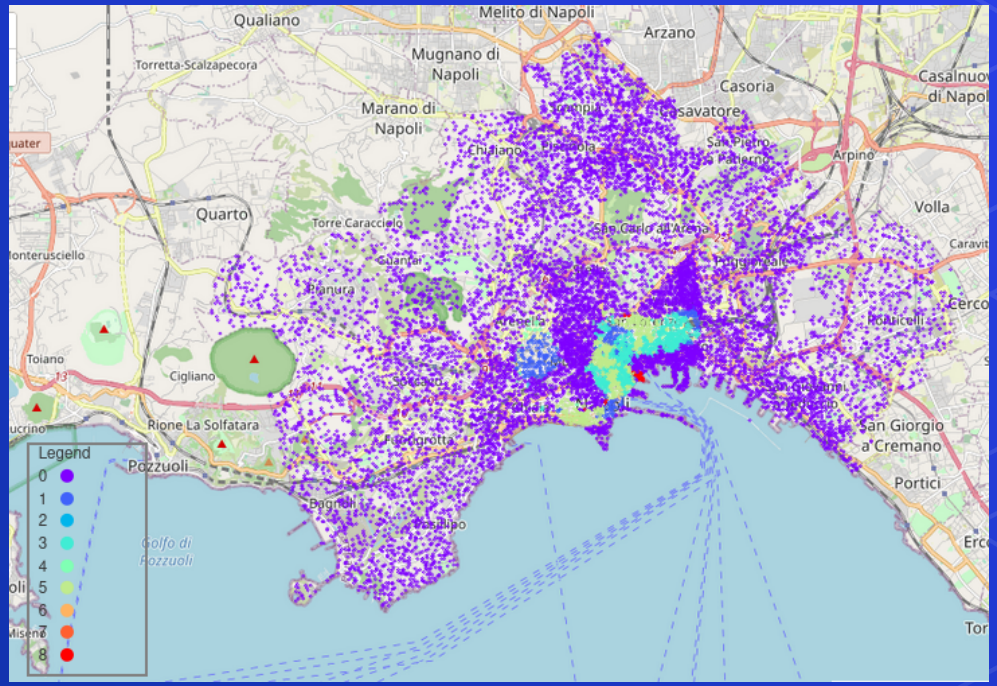
Prediction

We have now built the model that is capable of clusterize venues based on the neighborhood similarity. Next steps are:

- Sample point from borough
- Explore nearby venues
- Classify with previously trained model

Results

Accordingly to the algorithm, cluster 5 areas are the best where opening a new pizzeria



Conclusion

Power of ML

Machine Learning can be applied in various field.

In this report we gave an example of application in finding the optimal place based on acquired experience on active pizzeria.

This can be generalized for opening bar, pub and also restaurants.

Thanks!

Any questions?

You can find me at:

luca.riccardi@live.it