The best restaurant that suits your needs

The Battle of Neighbourhoods — Capstone Project

by Carloclaudio Calvino

As part of the final project for the *IBM Data Science Professional Certificate,* I have chosen to explore italian restaurants in the city of Naples.

1. Introduction

Image yourself arriving in a beautiful city, for work or on holidays.

Here we will consider one of the most beautiful places in Italy, the city of Naples. You booked an hotel in the centre of the city, and for your launch and dinner you'll have to choose one of the restaurants, possibly close to your hotel. How could you make this choice?

We will pick up some important data about the hotel, such as the location data, the price and the quality of the restaurant.

Then we can process this data and make use of non-supervised machine learning techiques to cluster the hotels, visualize them, and make some considerations about the suitable restaurants to your needs.

2. Data

This project relies on public data from Foursquare API.

In detail we get:

- Geographic data (latitude and longitude) of the restaurants
- Class price of the restaurants (1-4, 1 cheapest)
- Average rating (1-10)
- Number of likes, another indicator of preference

This is an example of the main dataframe:

[35]:

	venue	latit	longit	rating	price	likes
0	Osteria da Antonio	40.839877	14.253767	8.1	3	45
1	Trattoria Da Nennella	40.841990	14.248700	8.3	1	336
2	Antica Capri	40.841940	14.247890	8.2	3	56

We will get all the data by using the following endpoint:

- Search, to get points of interest (POI) near our hotel
- Explore, to get detailed information about our POI

2.1 Data-Cleaning

To validate the dataset, the first thing we considered was to remove missing data. In particular, not all the restaurants had the feature **price** always populated, so we invoked the API in order to have this column always filled.

The second point was to validate all the venues returned by the API as being effective Restaurants. In fact, sometimes it is returned a Cafè, or Pizza Place,

or something else. We validate by the Category Id all the venues as being really resturants.

Third , we removed all the entries with the rating not populated.

3. Methodology

Our starting point are the restaurants near a given hotel, within a distance of 1400 meters. Let's pretend to choose the Palazzo Turchini Hotel, the geographical representation would be this one:

The dataset is the following , and we will use 3 features : price, likes, rating.

[35]:

	venue	latit	longit	rating	price	likes
0	Osteria da Antonio	40.839877	14.253767	8.1	3	45
1	Trattoria Da Nennella	40.841990	14.248700	8.3	1	336
2	Antica Capri	40.841940	14.247890	8.2	3	56
3	Antica Hosteria Toledo	40.840922	14.248122	7.9	3	33
4	L'Angolino	40.842091	14.251549	7.6	3	8
5	Ristorante Al 53	40.848668	14.250226	8.7	3	13
6	La Taverna del Buongustaio	40.845971	14.248833	8.0	3	11

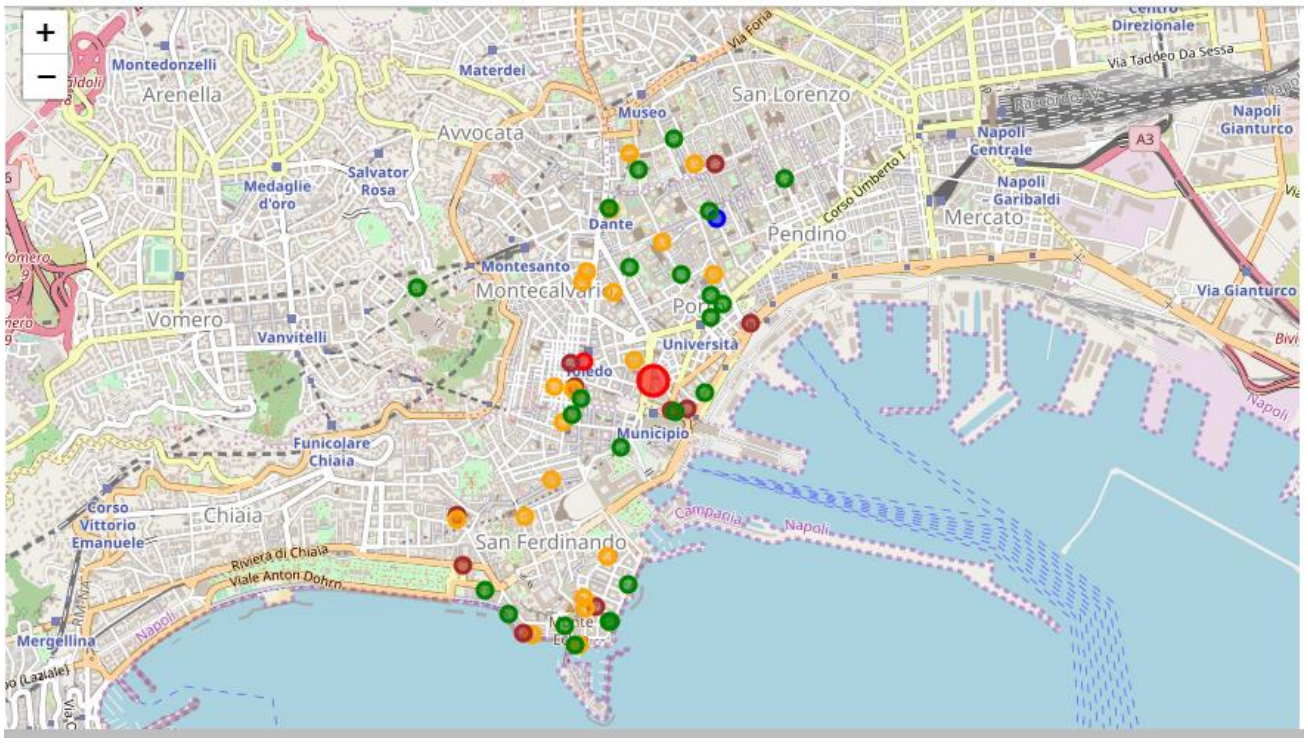
Then we will explore each feature, to understand how data is distributed.

Now we have to choose a model to help locate the favourite restaurants. We will rely on KMeans, one of the most popular machine learning algorithm for classification.

Briefly, this model groups data into clusters, where the points are somewhat similar among them. Keep in mind that the algorithm see only numerbs and doesn't know what are prices and ratings for a restaurants: we will then evaluate the clusters to see if they have any sense.

4. Results

Our final result is well rapresented by this map:



The number of restaurants have been reduced thanks to the data cleaning process. Moreover, they have been classified in 5 different clusters, each of one has a particular meaning.

5. Discussions

We can use such this approach to help make a choice. We could have a more sophisticated model, using more features then the one illustrated.

For example, we could use dataset showing dangerous streets on Naples, and using this info as a feature. Or we could use the distance in meters from the hotel as a new feature, and so on.

6. Conclusion

This is an example of classification used to help people make a choice. Without this approach, a manual choice made studying each location would have been very frustating! Thanks to ML algorithm, we can save much time to spend in the best restaurants of Naples!