Capstone project IBM Professional Data Science:

Open A pizza Restaurant in Salerno (South of Italy)

OPEN A PIZZA RESTAURANT IN SALERNO (SOUTH OF ITALY)

Background

In Salerno, one of the most popular meal is the pizza and many people are interested in stat a business opening a “pizzeria” ("restaurant that make pizza"). Nevertheless, a lot of "pizzeria" are already open in my city and some of them are very close to each other and those lead to a high lever of competition.

Business Problem is to understand where in Salerno is a good choise to open a pizzeria. I'm going to address the abovementioned business problem translate it in a question that is representative for the business, that is: which are the area that are densely located the pizza restaurant?.

Interest

The investor that are planning to open a pizza restaurant are very interested in understand in which borough of Salerno the business will have the less market competition and, potentially the higher income.

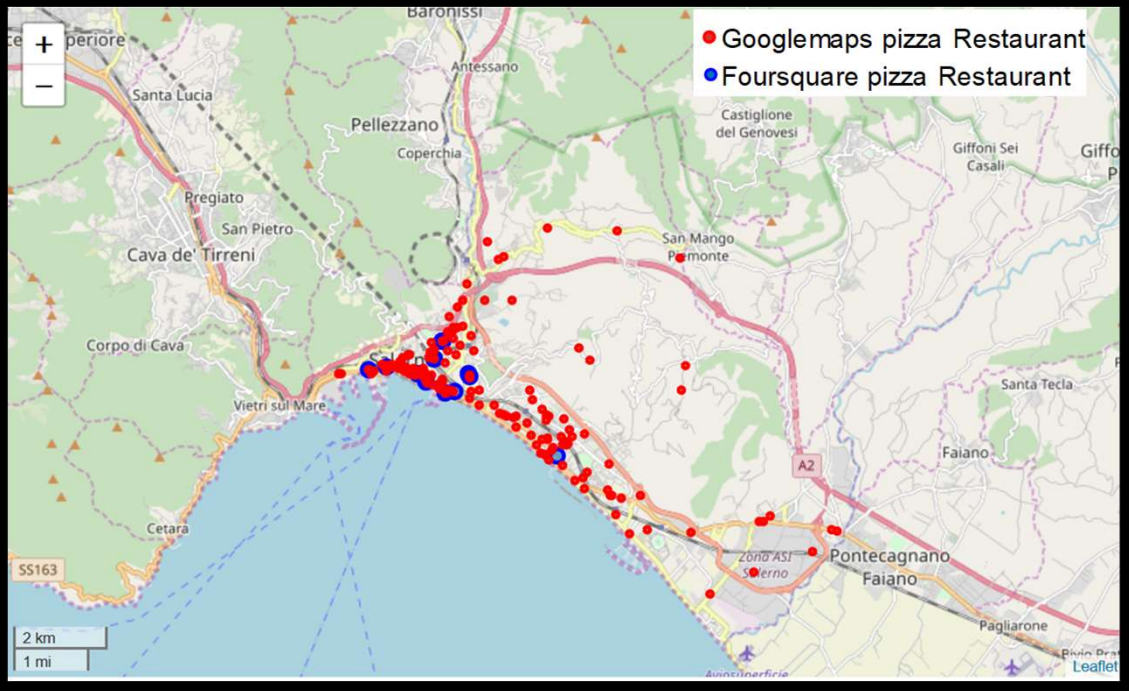
DATA ACQUISITION

I made the data acquisition of the pizza Restaurants respectively from:

- google maps: I use an executive for the web scarping of the acquisition of the pizza restaurant in Salerno in a *.csv file;
- Foursquare (using the civic code of the Borough of Salerno)

DATA CLEANING

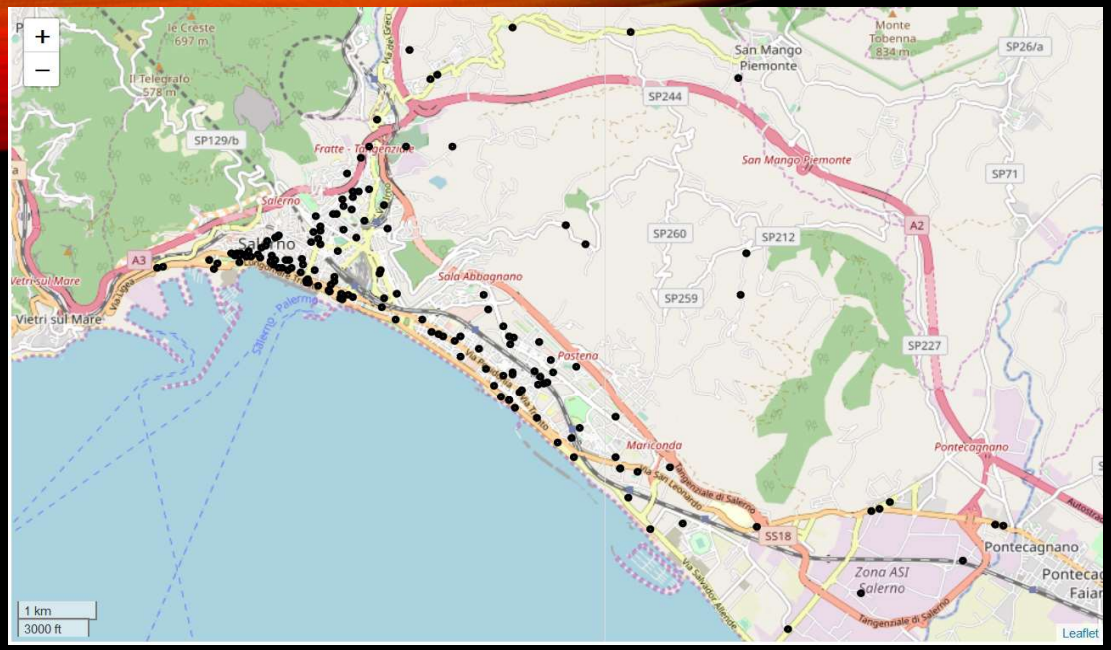
- For the database of Foursquare I considered only the vanue named «pizza place»
- I compared both the databases and I created the database named UNIQUE database contaning all the pizza place to analyze



FEATURE SELECTION

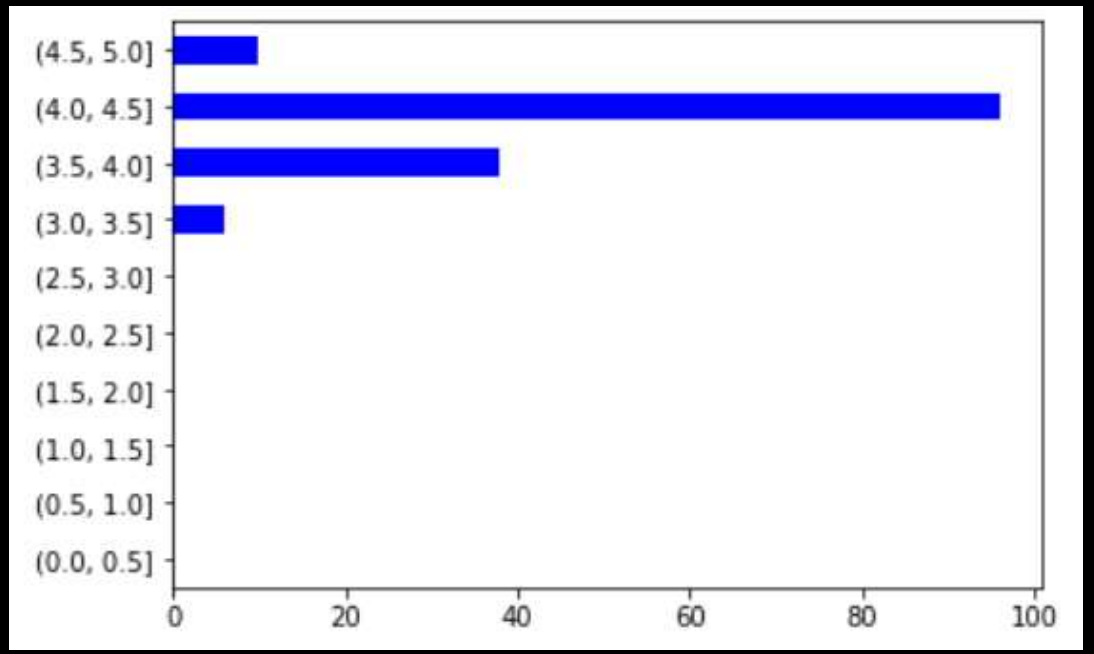
- For the analysis I had to choose between two options:
- The “completeness of the database” in terms of number of pizza place: the Google Maps database have about 150 entities, the Foursquare has about 15 entities;
- The feature associated to the Foursquare entities are more than the feature associated to Google maps entities (I only have the geolocation and the rating)
- As my aims is a geospatial analysis, I decide to consider two feature that are:
- Geolocation
- Google maps Rating

DATA EXPLORATION



Geolocation of pizza restaurant according to the UNIQUE database

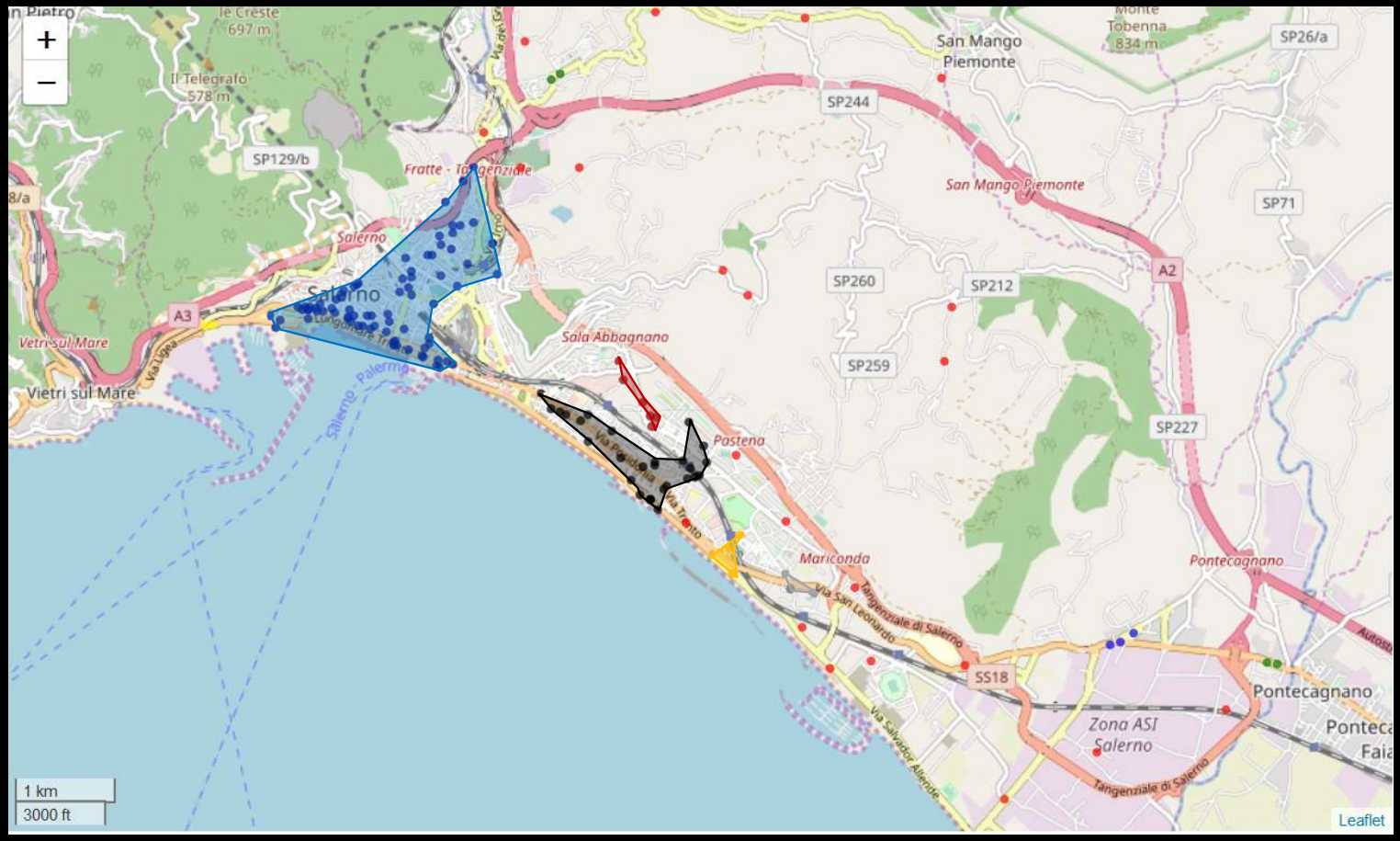
Rating of pizza restaurant according to Googlemaps rating



CLUSTER ANALYSIS

- My objective is to identify the areas where each pizza place has al least one pizza place in a radius pf 300 metres.
- I considered the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for the cluster analysis because the machine learning algorithms performs density based analysis that are useful for my objective. In fact, density-based clustering identify regions of high density that are separated from one another by regions of low density. Density in this context is defined as the number of points within a specified radius.

CLUSTER ANALYSIS



In these areas there are oizza restaurant having distance less than 300 m
So, if someone whant to open a pizza restaurant should avoid this areas

CONCLUSIONS

- I have created a database of pizza using both Google maps and fourspare API
- I have applied the DBSCAN machine learning algoritm fo the avauation of the most densely populated area in salerno (less than 300 m between two pizza resyauran)
- I found the area of Salerno in wich is better to aope a pizza reastant