Open a Pizza restaurant in Salerno (Southern Italy)

Claudia Sacco

June 28, 2019

1. Introduction

1.1 Background

HI! my name is Claudia and I come from Salerno a nice city in the south of Italy. In Salerno, one of the most popular meal is the pizza and many people are interested in start a business opening a Pizza restaurant. 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.

1.2 Problem

So, my Business Problem is to understand in which area of Salerno is a good choice to open a pizzeria. I'm going to address the abovementioned business problem by translating it in a question that is representative for the business, that is: which area are the most densely populate by pizza restaurant? Avoiding this area will be the right way to open a pizza restaurant that have more chance to have high income .

1.3 Interest

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

2. Data acquisition and cleaning

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 (I used the civic code in order to search the pizza restaurant that are in the area contained in the circle having center in the coordinates of each postal code).

2.2 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 havong all the pizza place to analyze.

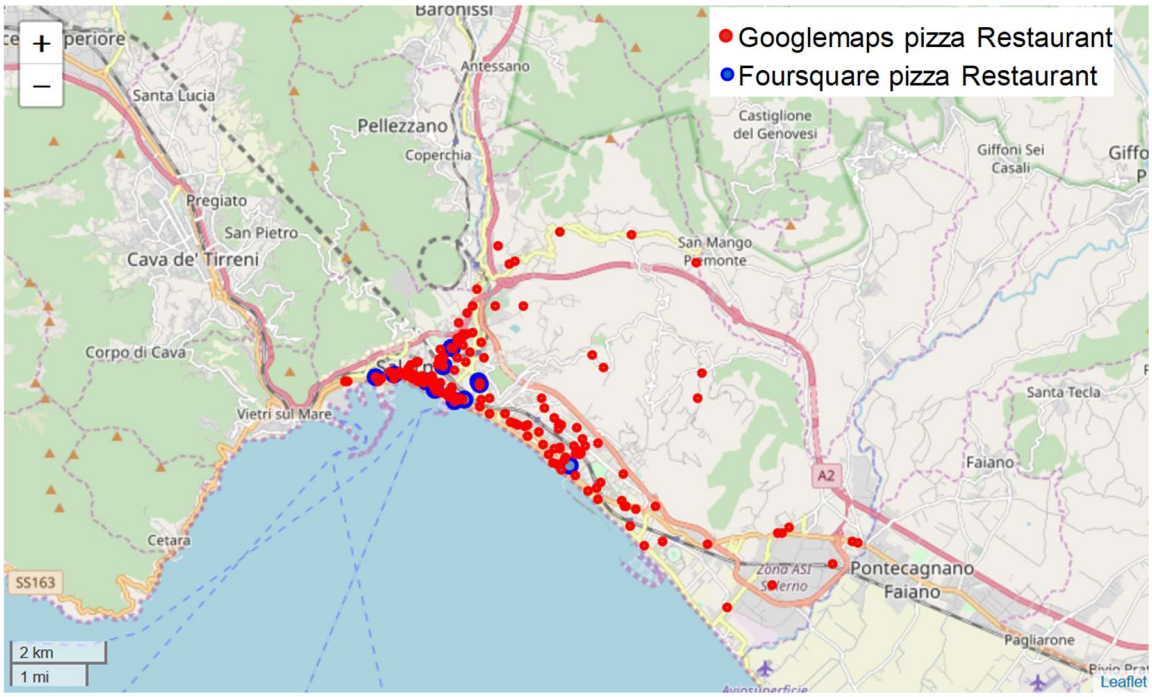


Figure 1: Foursquare and google maps database.

2.3 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

3. Exploratory Data Analysis

In figure 2 is illustrated the database containing all the pizza restaurant collected for the Salerno City. As it is possible to understand analyzing the figure 3 containing the rate of pizza restaurant the quality of is pizza not so significative as information as the large amount of the pizza restaurant have an higth rating (spanning from 4 to 4,5).

So the best indication that It is possible to give to an investor is to avoid the area where the pizza restaurant are densely located.

In the next section I'm going to identify these area using machine learning algorithms.

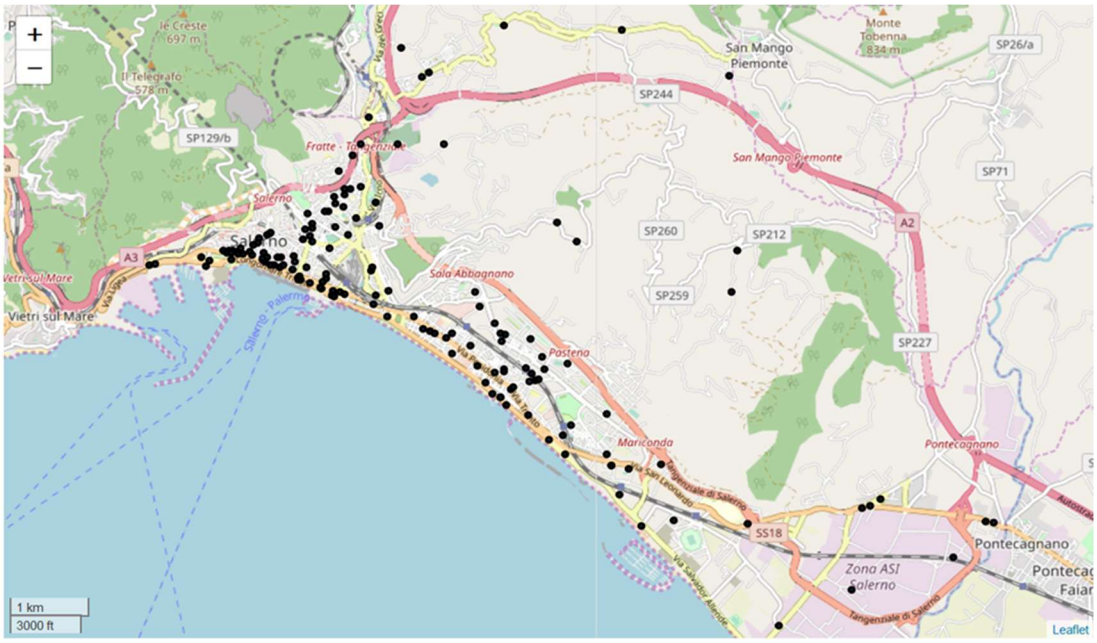


Figure 2. Unique Database.

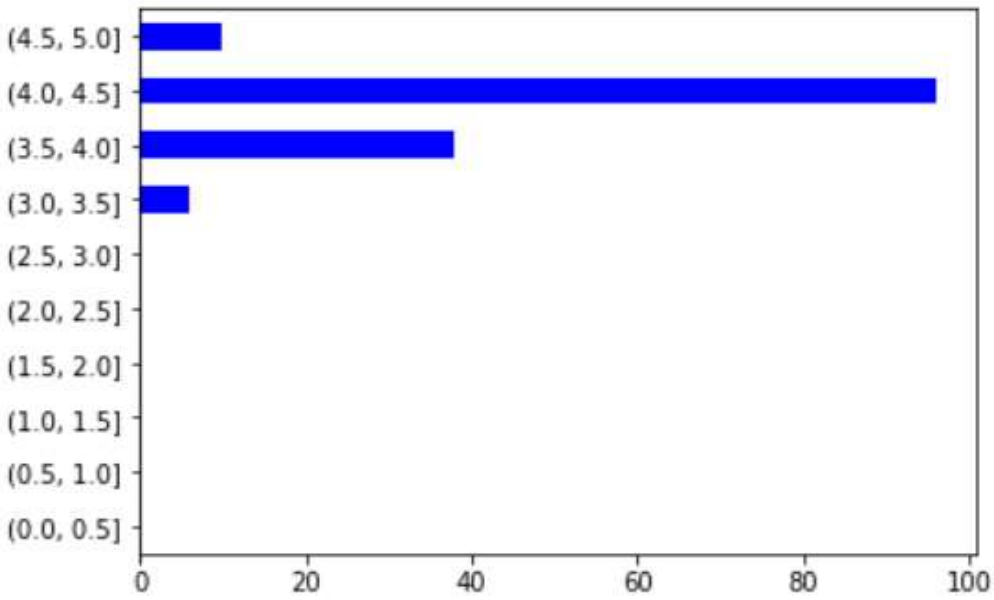
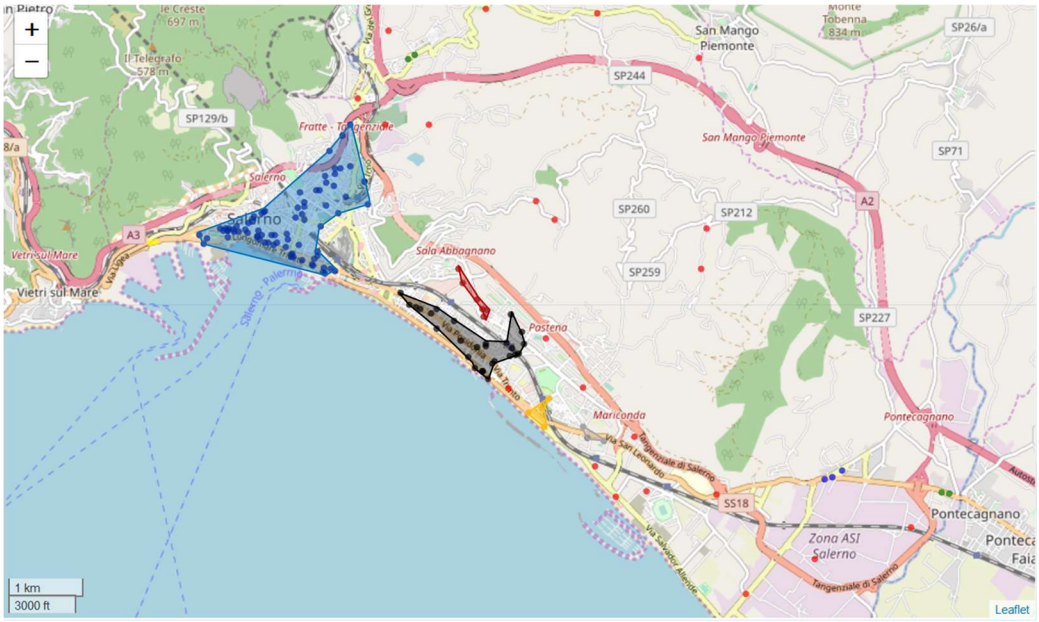


Figure 3. Rating of Pizza restaurant.

4. Machine learning modelling

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.



5. Conclusions

I have created a database of pizza using both Google maps and Fourspare API and I have cleaned this database. I have applied the DBSCAN machine learning algorithm for the evaluation of the most densely populated area in Salerno (less than 300 m between two pizza restaurant). I found the area of Salerno in which is better to open a pizza restaurant.