Business Case Introduction

A big restaurant franchise wants to explore new markets open new stores in the touristic region of Lake Como, Italy.

Before proceed with their project they decided to proceed with a viability analysis considering local populations size and how they are distributed in the region and the existent restaurants that are installed in the zone.

Data Description

To proceed with the data analysis initially and give an answer to the board of the franchise I modelled the problem. I would like to investigate the price level of restaurants in the area of the province and their rating.

The first step of my analysis was scratch data from a public data set that contained all Italian cities names and postal codes. After, I retrieved statistical data from istat, that is the statistical department of Italian government. The demographic data had been download to a CSV file that is available on my github.

In the first version of the code I used foursquare database, but I saw that their database for the region in very poor, considering that the application in not very used by locals. So I decide to retrive data information regarding price level and rating from Google API. From there I obtained the list of all restaurants in the province, their location, price level, rating and other data using GEOLOCATION API and function NEARBY.

Methodology

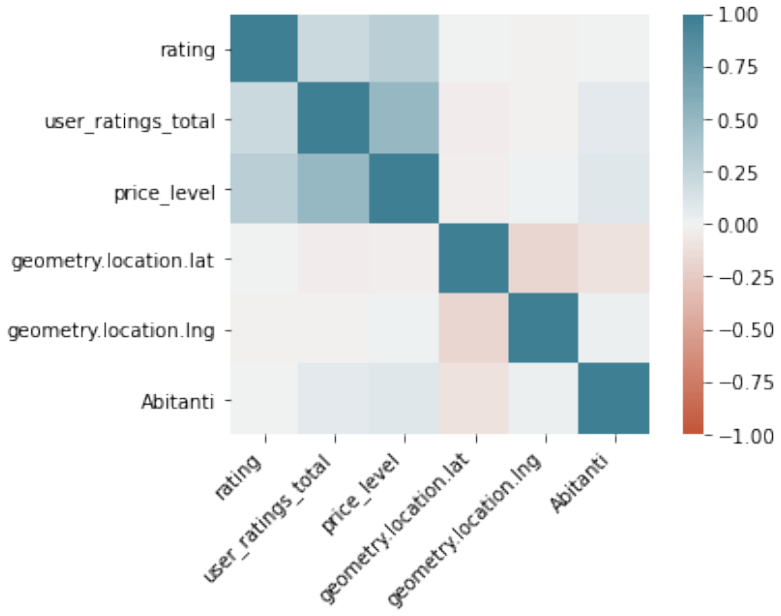
First I got the data from ISTAT, official statistical database and for Comune e Citta, a public domain that contains info about all cities in Italy. I organized the data and merged them into a dataframe.

next step was use the resultant data as an input in GOOGLE API to retrieve data about all restaurants installed in the region of Lecco and merge the result with the dataframe that contains demographic and geographic data. I cleaned the data deleting all empty and NaN values and then I normalized the data.

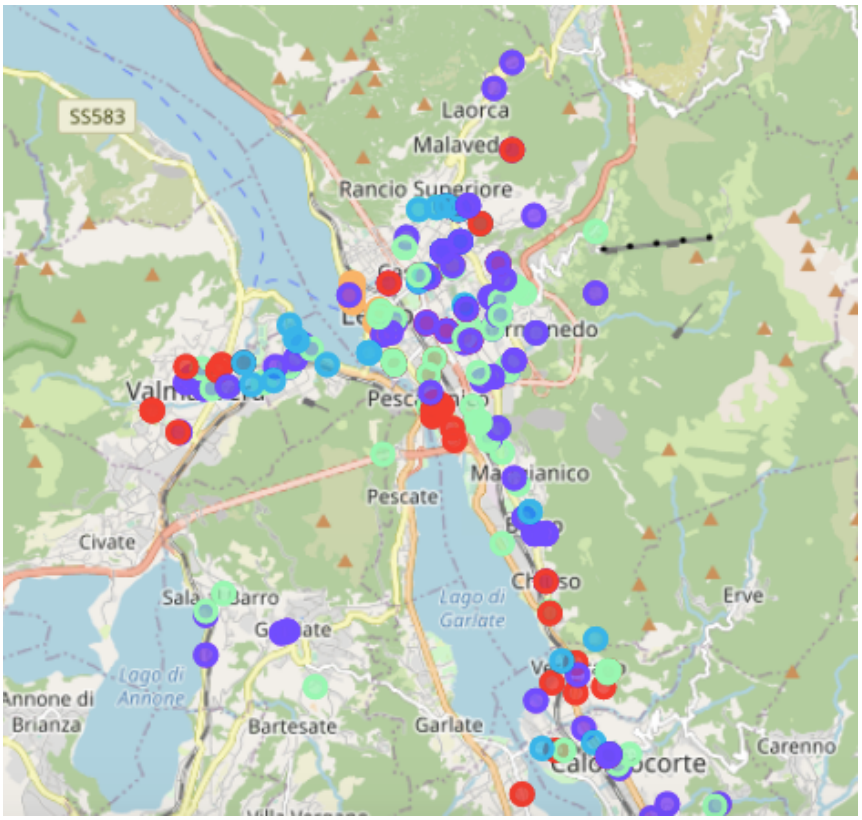
I use the resultant dataframe to cluster the restaurants in the region and created a heat map to investigate the correlation between price, rating, location and population size.

RESULTS

From the analysis evolved I got the fallowing heating map for the parameters described previously.



From clustering I obtained the Map bellow, where it is possible to see all restaurants of the region, categorized by price, rating, location and population size of each neighborhood:



Conclusion

Considering the clustered data and the heat map I would recommend a new restaurant in the biggest neighborhoods, specially those ones in neighborhoods of Lecco and Valmadrera. This will allow reach better price levels and positive ratings.