# Enhanced BI towards tourist activity in Fano Italy

Matheus Rychescki jacobs

July 2021

## 1 Introduction

Turism is one of the main drivers of a lot of italian cities during high season, sometimes being the main income source of a lot of families and small business. Thus, the location of a business can be of relevant importance due to the high tourist activity in certain parts of the city. From the tourist perspective, it can be a struggle for some tourists to find the right neighborhood in a way that they can explore the gastronomic scenario of the region. On the other hand, from the business perspective, knowing in which part of the city one should invest and where they may be the highest tourist activity can further enhance the business focus/restaurants of the city in order to enhance the local economy.

### 1.1 Problem

Data that might contribute to determine the best location for opening a restaurant/hotel in the city of fano, Italy might include the latitude and longitude of each restaurant and hotel in the database.

## 2 Data acquisiton and cleaning

### 2.1 Data sources

Most data about the Restaurants  Hotels can be found on the Foursquare database, such as: Name, addresses, latitude, longitude, city.

### 2.2 Data Cleaning

Data downloaded from the Foursquare database was assembled in a data frame. There were quite a bit of data missing from City location, postal code, due to the lack of registration by the users in the database. First, the data that was not relevant to the analysis were dropd such as Distance, Country and duplicated address, restaurant name. Second, We polish the data by removing Nan values

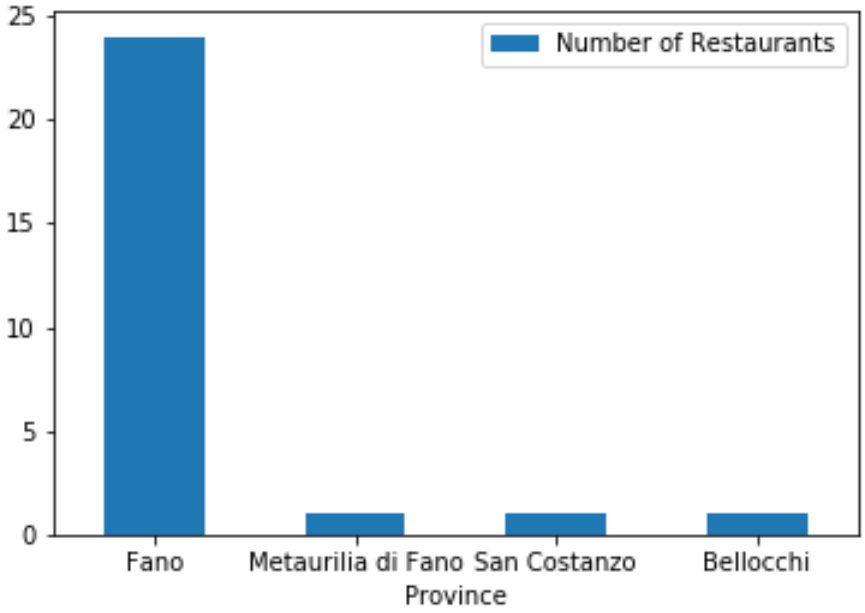


Figure 1: Number of restaurants in the data base for each City/village in the metropolitan area.

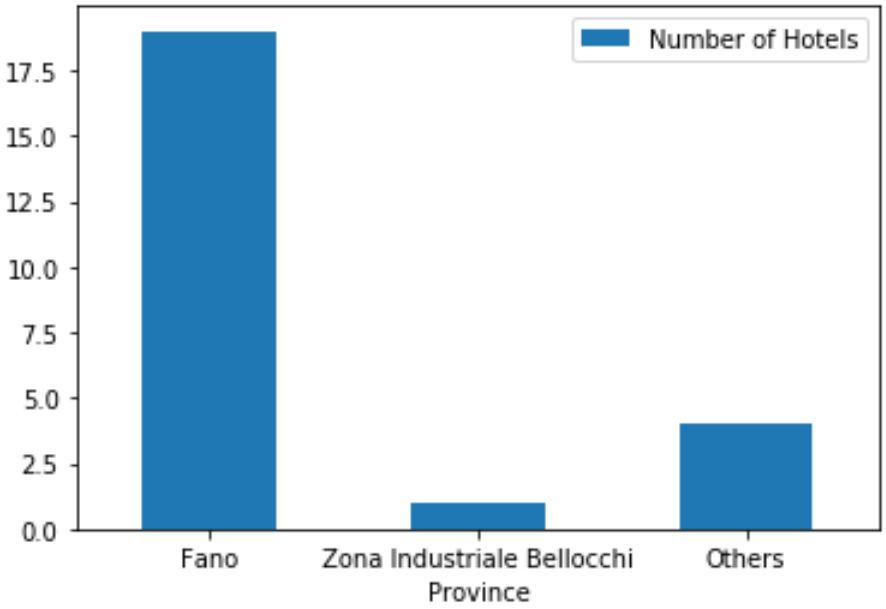


Figure 2: Number of hotels in the data base for each City/village in the metropolitan area.

for zero and renaming the columns of the data in order to provide a better understand of the data frame.

### 2.3 Feature selection

After the data cleaning, the restaurants data frame were 36 samples and 7 features in the data, as the one for the hotel there were 24 samples and also 7 features.

## 3 Exploratory Data Analysis

First we start by looking at the amount of restaurants available in the data base for each city in the region (Figure 1). Clearly as one could expect, most of the restaurants are located in the main city of Fano. Furthermore, we look also at the number of hotels in same region (Figure 2) with it we could have an initial idea of where to focus our analysis.

Moreover, we combined on the map of the region all the restaurants and hotels available in the data base as can be seen in Figure 3.

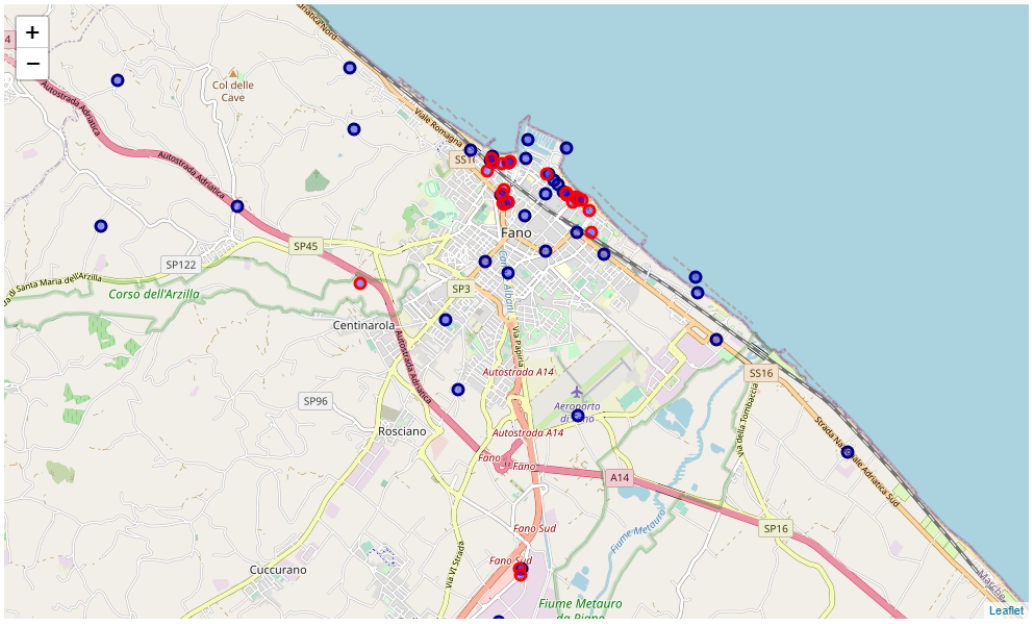


Figure 3: Number of restaurants and hotels in the data base for each City/village in the metropolitan area. Blue points represent restaurants and red the hotels

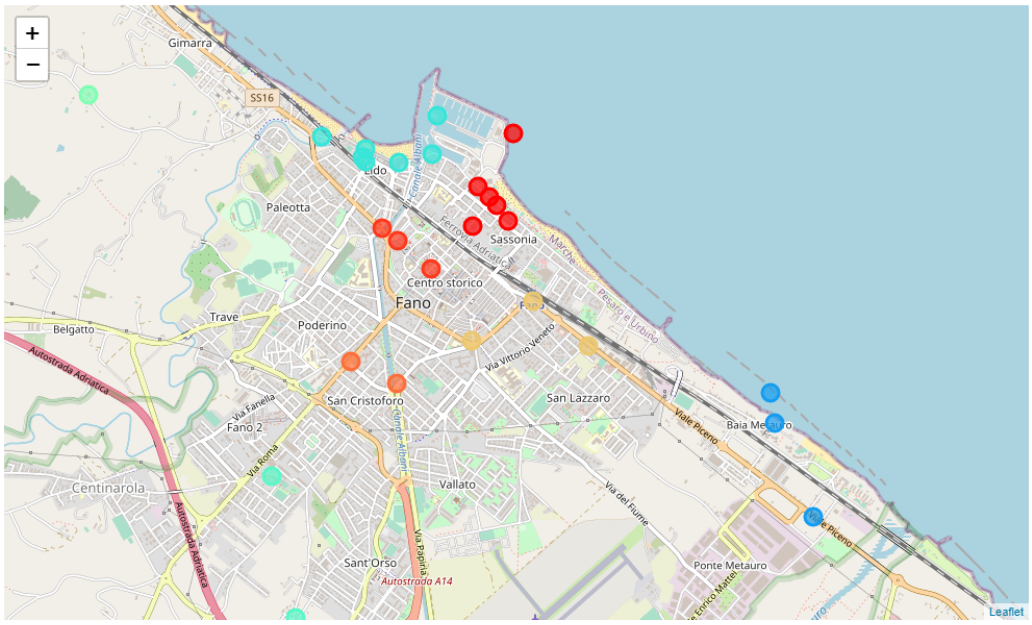


Figure 4: Clustering of the restaurants based on the number of restaurants and neighborhoods in the city

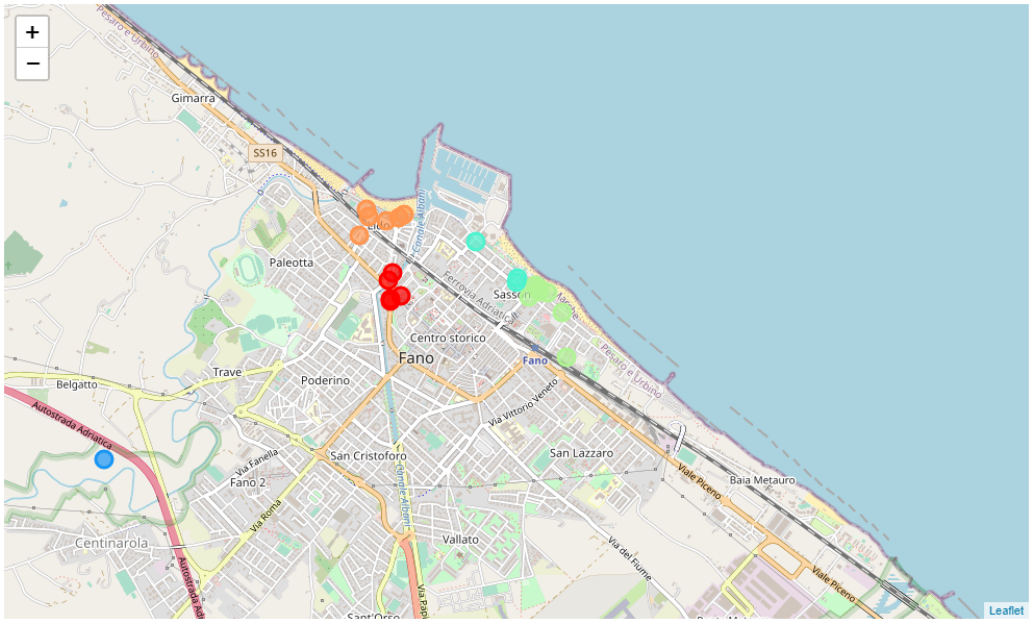


Figure 5: Clustering of the hotels based on the density of restaurants in the city

# 4 Clustering

For further analysis we use K-means clustering algorithm. k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. k-means clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more difficult Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. For instance, better Euclidean solutions can be found using k-medians and k-medoids.

## 4.1 Solution of the problem

Since the city of Fano has officially 15 neighborhoods, we start our analysis by group the number of restaurants into clusters represented by the each neighborhood. We applied the above methodology until convergence was reached. As it can be seen in Figure 4 the clusters with the highest density of restaurants are the one closet to the Harbor/beach of the city, represented in the picture by the light blue and red points. As one could expect from a coastal city, as further one gets to the sea, less restaurants one finds it.

Furthermore, by clustering the hotels based on the density of restaurants in

the city, one can see that the main location for hotels are either the historical center of the city or the ones right in from of the beach, represented in Figure 5 by red and orange points, respectively.

# 5 Conclusions

In this study, we analysed the relationship between the density of restaurants and hotels in the city of Fano (IT). It was identifies that the regions with the highest/loest number of restaurants and hotels. This supports a possible opening of restaurants in a high density hotel region with low restaurant density and suggest also that opening a restaurant close to the beach and harbor may not be the most profitable location due to high competitively. Possible limitations of this study are the number of registered restaurants and hotels in the database and also the lack of the rating of such locations. This can be used to further enhance the model and give a better prediction of possible more profitable locations.