

# Final Assignment Report

## A case of study on Italian Cities Analysis

Riccardo Pollo

May 2, 2021

# 1 Introduction and Problem definition

Italian peninsula is a beautiful place where spend vacation. There are lot of interesting venues to visit, lot of history, lot of things to do. Not only culture: in Italy there are lot of mountain in which it is possible to hiking, beaches where sunbathing and entertainment.

A travel agency need to reply to a clients about some question about traffic in Italy. Our clients want to know more about traffic risks on italian roads.

Once we find some low-risk-cities, clients will be able to decide in which city spend their vacations.

Some of them will go by car, but some other want to rent a car in place and they want to know if they can reach their city by air or by boat: is there any port or airport near to selected low-risk-cities?

As data scientist we don't have any specific goal. Instead, it is asked to acquire information about risks create a report to explain the traffic situation in Italy to our clients, and they will decide autonomously based on our informations.

We scheduled some points to follow in order to successfully carry out the analysis.

- First let's see how many people can drive, according to the population of each city and the total amount of active drive licenses.
- Plot a line chart on the total number of incident cases in last year (from 2004 to 2018).
- Which the region with the larger number of incident from 2004 and 2018? Plot a bar chart.
- Since we want to go now, let's check the result according to the last year available (2018)
- Is there any correlation between number of incident and (number of drive licence/population) per province? let's check with a simple regression, checking if relation is linear or not, changing the exponent.
- Suppose we want to divide 3 cluster: minimum risk, maximum risk, average risk. So we perform a k-means cluster on number of incident (and eventually other features) per k=3. Let's plot it with Folium.
- From minim-risk spot we want to choose only places that are province (if there is at least one, otherwise we will chose places with higher population or density), supposing there will be more interesting things.
- Suppose we choose 5 cities between which we need to choose our destination. Let's ask to fourstroke which are the top 100 venues for this cities. Check the type of venue trying to understand which kind of behaviour the population can have according to the entertainment.

## 2 Available Data

In order to acquire data we explored lot of italian government sites. We have found lot of dataset.

Even if not all dataset will be completely useful, we list all site we visited.

### **List of active drive licences per region**

<http://dati.mit.gov.it/catalog/dataset/patenti>

### **Provinces in Italy**

[https://en.wikipedia.org/wiki/Provinces\\_of\\_Italy](https://en.wikipedia.org/wiki/Provinces_of_Italy)

### **Population infos of italian cities**

<https://simplemaps.com/data/it-cities>

### **High Density Population infos**

<https://data.humdata.org/dataset/italy-high-resolution-population-density-maps-demographic-estimates>

### **Incidents locations from 2004 to 2018**

<http://dati.mit.gov.it/catalog/dataset/localizzazione-incidenti-stradali-anni-2004-2018>

### **Incident infos per region from 2004 to 2018**

<http://dati.mit.gov.it/catalog/dataset/localizzazione-incidenti-stradali-anni-2004-2018>