A case of study on Italian Cities and Traffic Risk Analysis

Riccardo Pollo

IBM Data Science Professional Certificate

Problem Definition

The goals are:

- Understand traffic situation in Italy
- Define the related incident risk
- Suppose someone want to organize a vacation in Italy, add to previous results some informations about geographical position and venues for some interesting Provinces.

Data Description

Data are retrieved from:

Provinces in Italy

https://it.wikipedia.org/wiki/Province_d%27Italia

List of active drive licences per region

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

Incidents locations from 2004 to 2018

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

Top 5 Provinces with less driver percentage

|	Province	% of driver
27	crotone	32.08
18	caltanissetta	34.31
29	enna	34.37
0	agrigento	34.39
21	catania	34.85

- Crotone, Caltanisetta, Enna, Agrigento and Catania are cities with the less percentage of driver in Italy.

- Less drivers doesn't mean less risks...

Incident Trend (1)
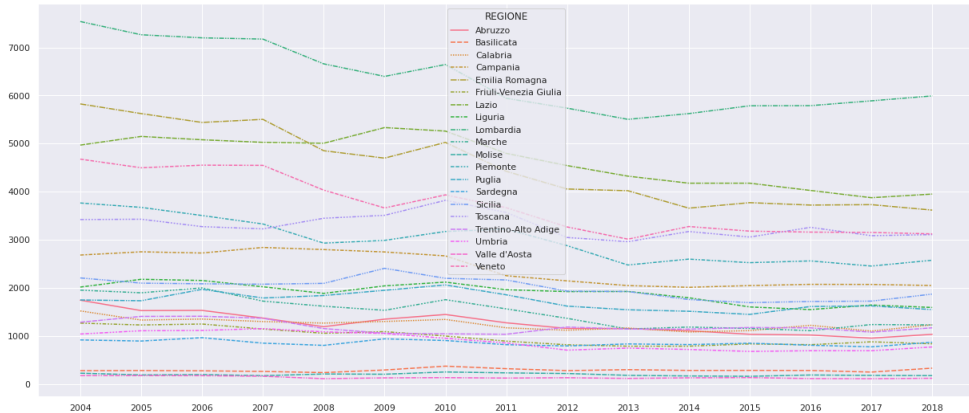
- Number of incident goes down during last years.
- Umbria, Molise and Basilicata has the lower number of incidents.
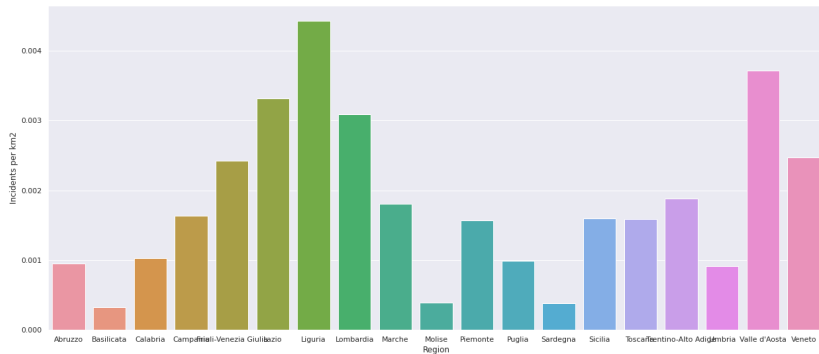
Incident Trend (2)
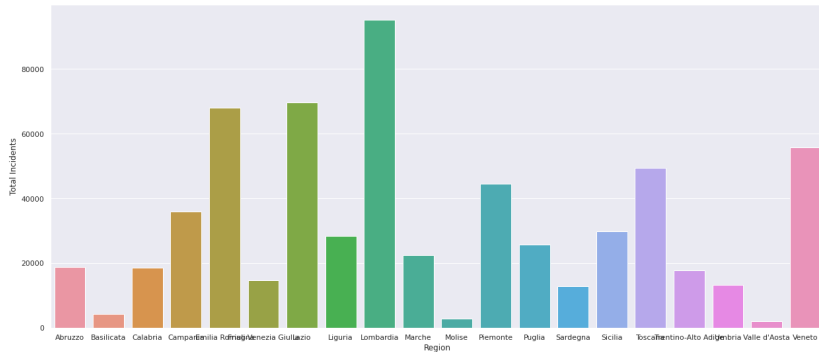
- Lombradia has the largest number of incident in last years sum over the following formula: $avg_{inc} = count(incident)/(2018 - 2004)$...
- ...but this doesn't take into account of the density of population
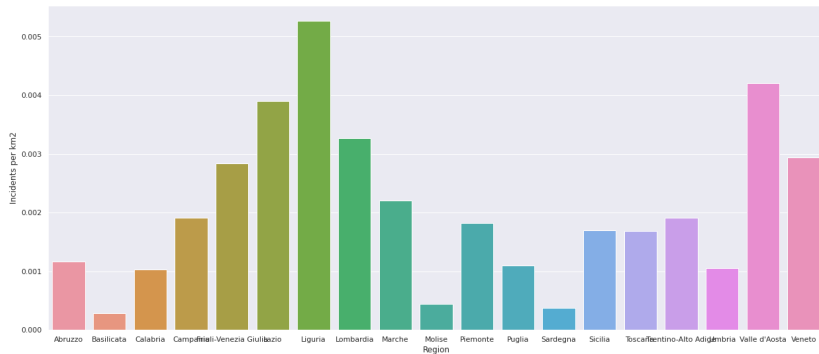
- The new plot divide the previous formula for the density of the population.

- Here's the same plot, showing only last year: 2018.
- Average values are confirmed during this.

Drive Licences and Incident correlation

- Our intuition suggest that the higher the percentage of driver over the all population is, the highest the number of incidents in a Region is.

- We still consider only data regarding 2018, evaluating the Pearson Correlation Coefficient between the two measures. Results is -**0:58**.

- This result suggests that, using our data, we must be careful on considering the number of drive licences, because seems not to be an meaningful feature.

Nearest Provinces
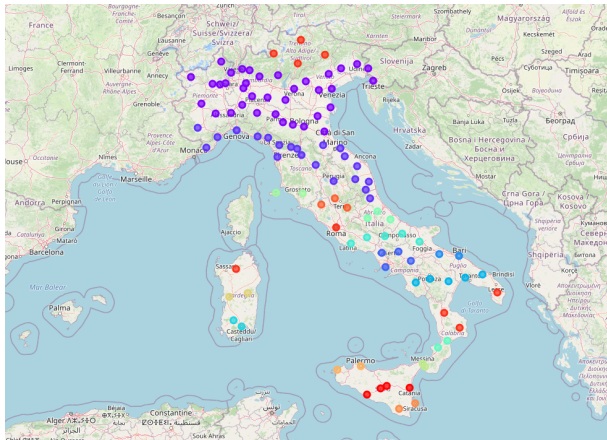
- Lot of Provinces can be considered as "easily reachable". A client could organize a tour to visit all the provinces in cluster.

- Only few province are not connected to any one other: one of these is Rome, which is bigger enough to be visited alone.

Traffic risk segmentation

- What if we want to recognize exactly 3 type of region: low risk, moderate and higher.

- According to the map:
  - Violet Regions have an higher risk,
  - Cyan Regions have a moderate risk,
  - Red ones are the safest.

Venues and Population behaviour per provinces

Province	count	categories
cagliari	3	Bar, Ice Cream Shop
catanzaro	3	Pizza Place
nuoro	3	Mountain
oristano	3	Church, Pizza Place
reggio calabria	5	Café
sud sardegna	3	Miscellaneous Shop
vibo valentia	5	Pizza Place

- According to previous analysis Sardinia, Calabria and Puglia are the safest ones. Which are the top venues for these?
- Let's suppose a good reasoning:
  - Suppose in our vacation we are interested in Pizza Places.
  - Catanzaro has 3 Pizza Places, and Vibo Valentia 5. Catanzaro and Vibo Valentia are also easily reachable as we discovered in previous segmentation.
  - They are both in Calabria, one of the region with lower number of incidents per area as shown in previous bar chart.
  - One of these should be our choice!

Thank you for listening!