

Covid-19: Italy's phase 2



Emanuele Lanzani
09/05/2020

Which provinces are the safest to open?

Our problem:

- differentiate provinces based on infected, density, population, risky venues
- is it possible to get the economy going while still waiting in the riskier provinces?
- we don't want to sacrifice lives for the sake of the economy
- we also don't want to destroy the economy in the process

Our goal

- We will collect different data from different sources
- We will create different provinces clusters to understand which ones are the safest
- We will also try to understand if there are some provinces which may possibly require harsher restrictions

Data

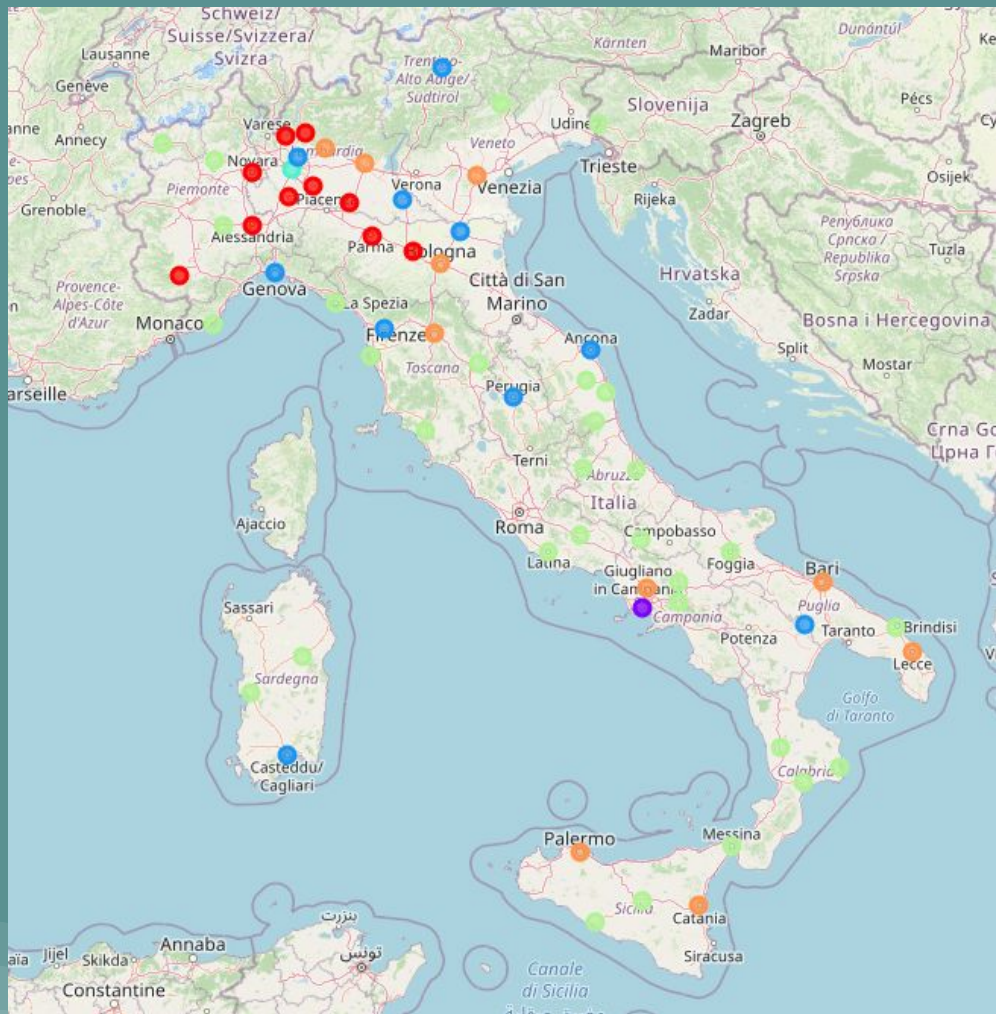
- Data from wikipedia Italy's provinces
- Data from the italian health ministry website, with up to date infected numbers
- Venues Data from FourSquare

Methodology

- Get provinces data
- Wrangle it to clean some format errors
- Get infected data (this proved difficult and required a csv conversion)
- Wrangle it to get the correct dataframe
- Merge the two dataframes together
- Retrieve provinces' coordinates with geocoder package
- Create folium map
- Retrieve foursquare venues
- Merge the dataframe with the other two
- Standardize infected, venues, population and density fields
- K-cluster the new standardized dataframe
- Project the cluster on the folium map

Results

- Cluster 0 (red circles): infected are in the thousands. Venues, density and population are average
- Cluster 1 (purple circle near the center): infected in the thousands, but outlier density
- Cluster 2 (light blue): high number of venues, compared to other clusters
- Cluster 3 (turquoise in the north): high density, high population, high venues and extremely high infected
- Cluster 4 (light green): mid low density, mid low infected, mid low venues
- Cluster 5 (orange): high population, average density, mid high venues



Conclusions and Recommendations

- The **safest group** to undertake some risks with is **cluster 4**: they present a relatively low density, a relatively low number of venues, and a relatively low number of infected. Definitely areas where the virus is not widely spread and the density is not that high.
- On the other hand, **cluster 2**, which includes just Milan, presents the **highest risks** and therefore further restrictions should be considered.

Thank you!