

Covid-19 phase two: Italy case

In this project, we're going to understand how we can use a data driven approach to the beginning of phase two.

Actually, Italy has been hit hard by covid-19, in such a way that for eight consecutive weeks the country has been put in a strict lockdown, where people were forbidden to get outside, unless it is for strict necessity or for emergencies.

Right now, Italy is beginning its phase two, meaning that some restrictions have been lifted and some categories can get back to work, people can go visit relatives, and so on.

Actually, the restrictions are the same for all the countries, even though some regions have decided to take a more restrict or more open approach to it.

The question is: can we leverage the data we currently possess to understand where restrictions should be applied more firmly and where we can lose them up a little bit more? in other words, is it possible to tell the difference between a province where the risk of infection is low, compared to those where the risk is much higher?

The category that would be interested in such a study would be the political one, since politicians are the ones required to take decisions, balancing the need to reopen companies and other stuff and the need to preserve people's lives.

Considering this is just an exercise, we will not take into account all possible factors, but we will set the basis for further investigations which will include other factors.

The data we will be using is:

- provinces' population density. The higher the density, the riskier to reopen activities, since it will help the virus spread faster. Data about provinces' density can be freely found on [wikipedia](#);
- provinces number of infected people. The more people in a province are infected, the more the virus in that province is spread. Information about infected per province is provided by the health ministry [website](#), which releases daily reports with updated numbers, such as the one on [this](#) page. The language is italian, but the reports are basically just returning numbers, so it is fairly understandable;
- **only to include some foursquare data**, as requested by the exercise, but actually making it just more complex than it is supposed to be, we will find on foursquare data about aggregation venues, such as bars, restaurants and parks. The higher the density, the riskier it is to reopen, since it will incentivize people to aggregate in such places. This is of course an approximation, but it is needed to comply with the exercise.