IBM Capstone project assignment – Week 5

April 2020
By
Fen Jin

Table of Contents

Introduction

With the spreading of COVID-19 in the world, every nation has been dealing with this virus in any way possible. To mitigate the effect of the spreading is it important to understand the fast diffusion, both in biological and social ways.
Here, we will simply analyse how the virus has been spreading in Italy, according to the available data. We will try to show the progress of the spreading with a temporal analysis.
It is important to understand the current situation, everyone should remain alert and take the best precautions.

Data

We collected the data from the Protezione Civile website, which is officially the institution in charge of the COVID-19 emergency in Italy. They collect daily the situation on the Italian territory.
The data is reported in csv files, divided in different levels: national situation, region situation and province situation.
The last one contains more geographical information and the total number of COVID-19 cases by province daily.
On region level, the data contain more information on numbers of positive-death-recovered people. For example, there are numbers of people in the hospitals, positive cases, people with symptoms, number of swabs.
The national level data contain the summary of the region level.
We will start analysing the province data.

Method

According to the data available we will proceed analysing the progress of the virus spreading.
First, we need to clean the data and filter only the ones we want/need. More detailed information is explained in the related Python notebook.
Then, we need to visualize the data in simple plots to see the temporal progress of the situation.

Results and Discussion

Here we show some of the plots we created to visualize the processed data.

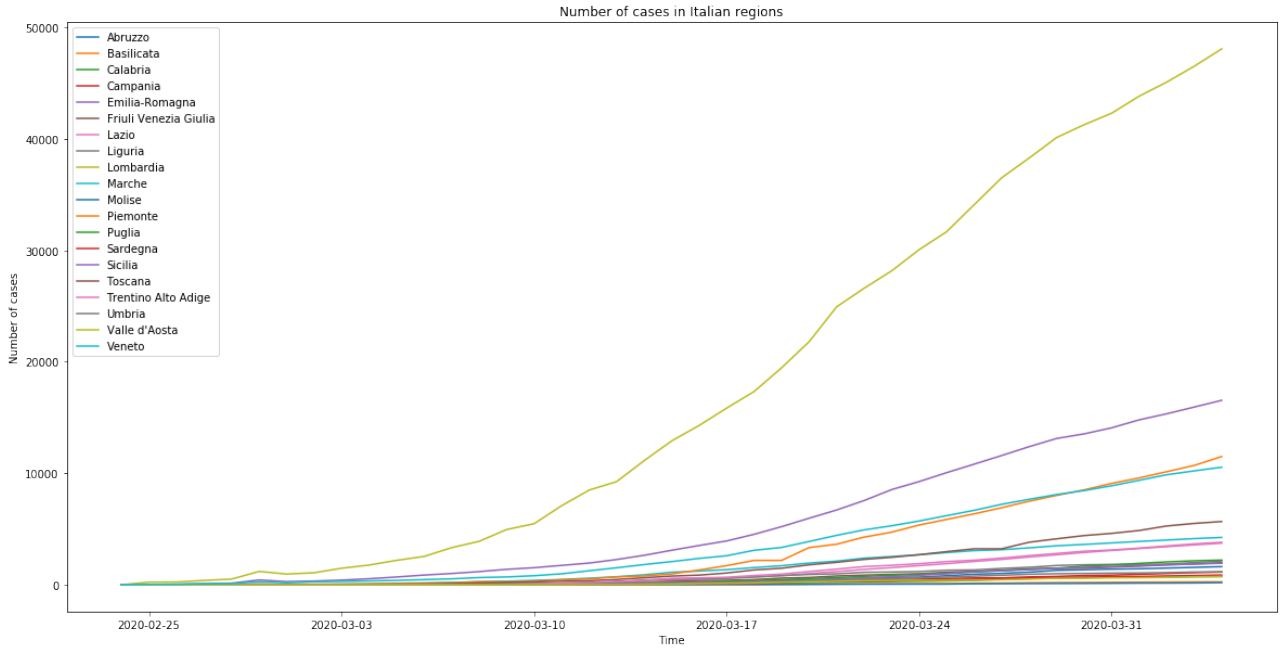

Figure 1. Linear plot of the number of COVID-19 cases in the different Italian regions.

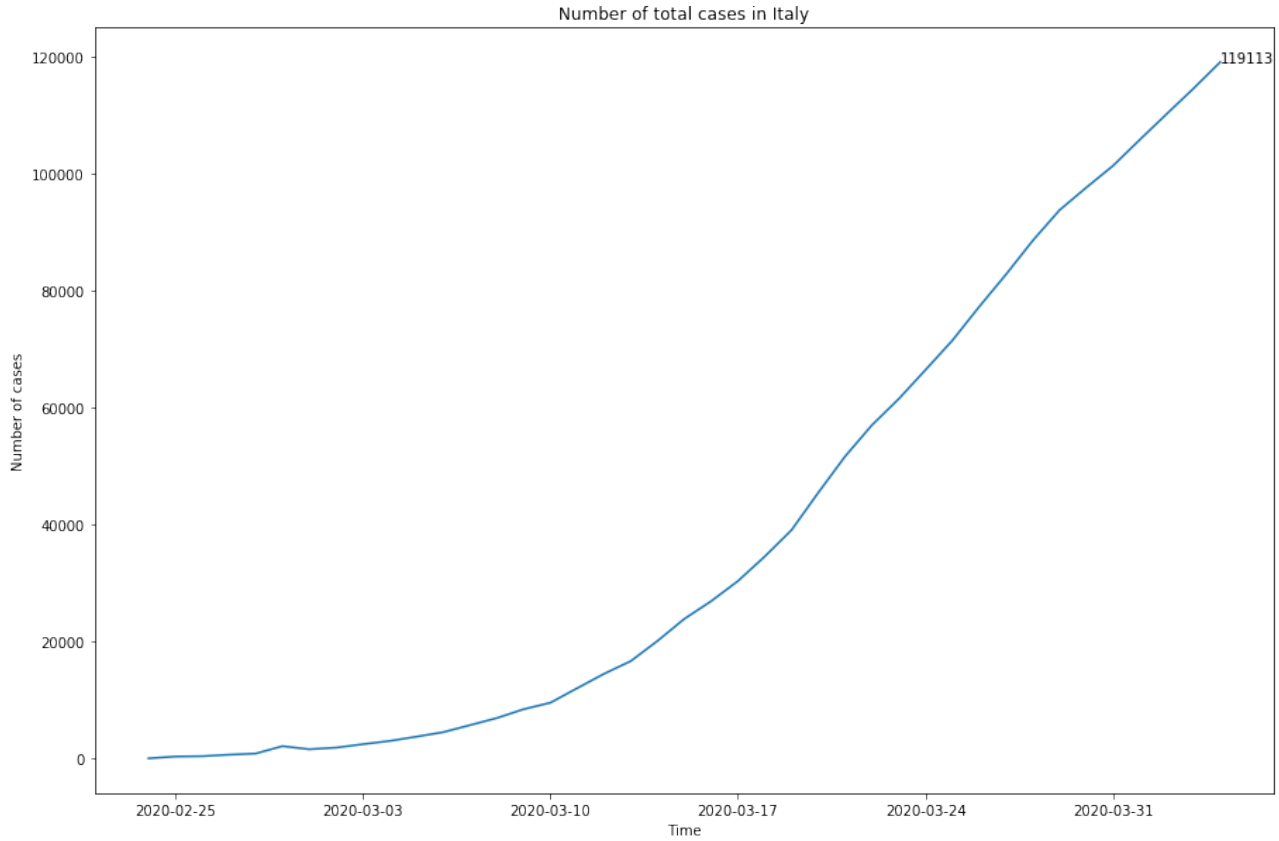

Figure 2. Linear plot of the total number of COVID-19 cases in Italy.

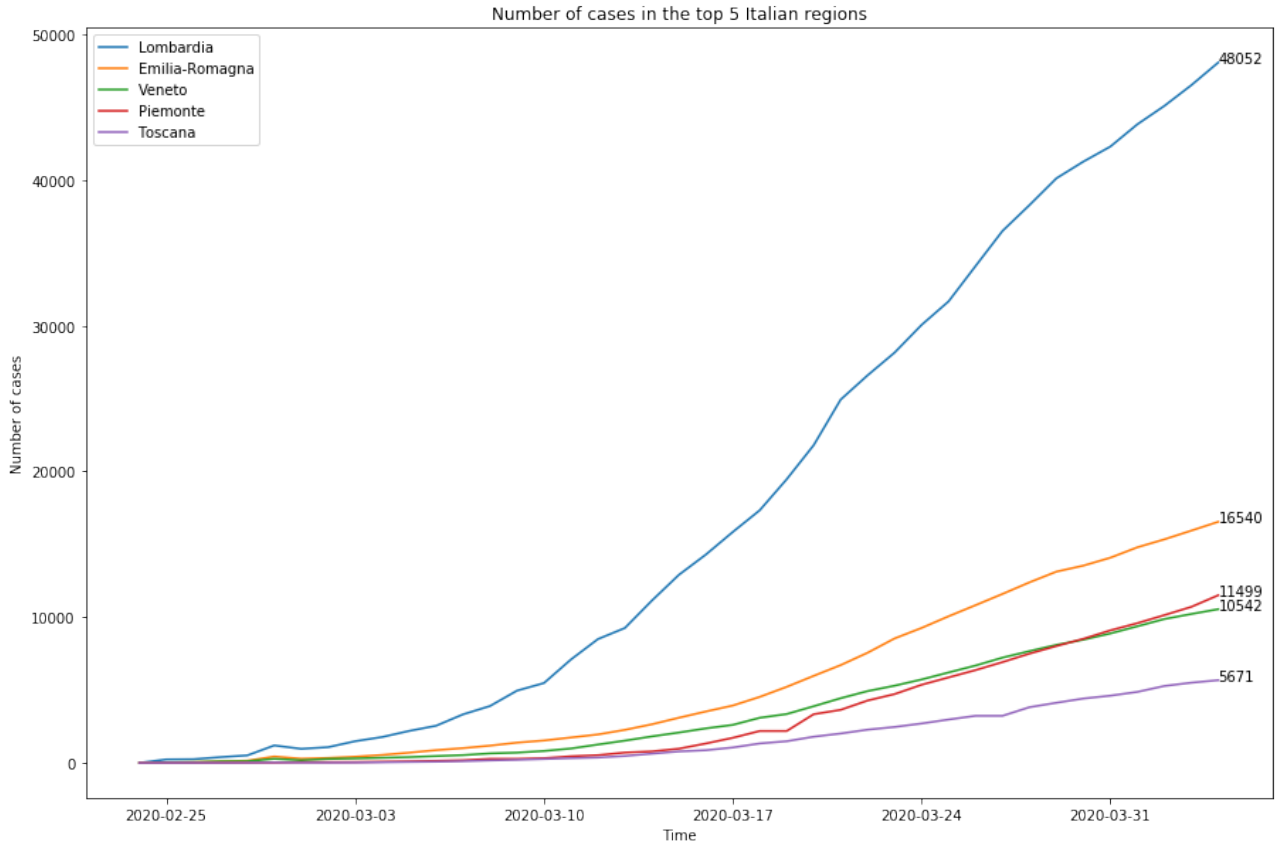

Figure 3. Linear plot of the 5 Italian regions with the most COVID-19 cases.

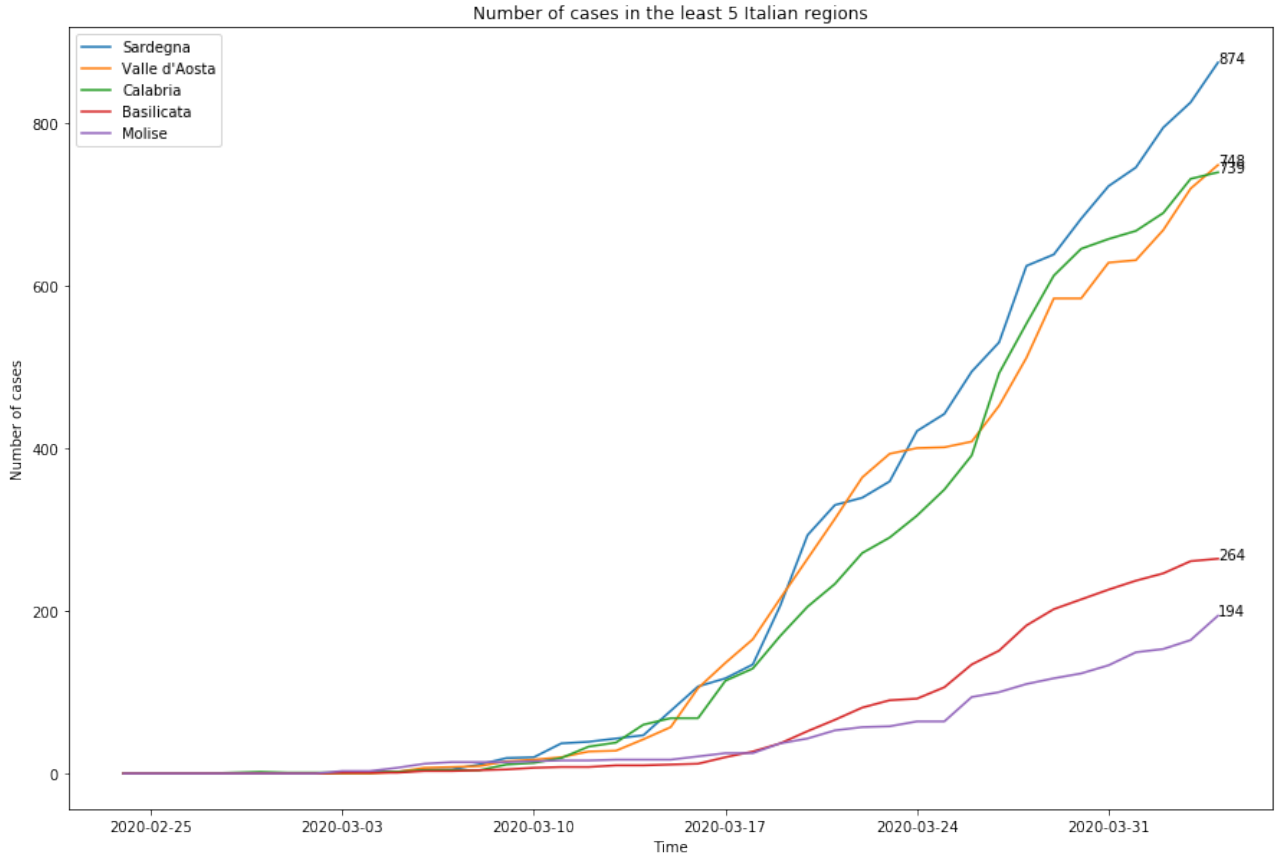

Figure 4. Linear plot of the 5 Italian regions with the least COVID-19 cases.

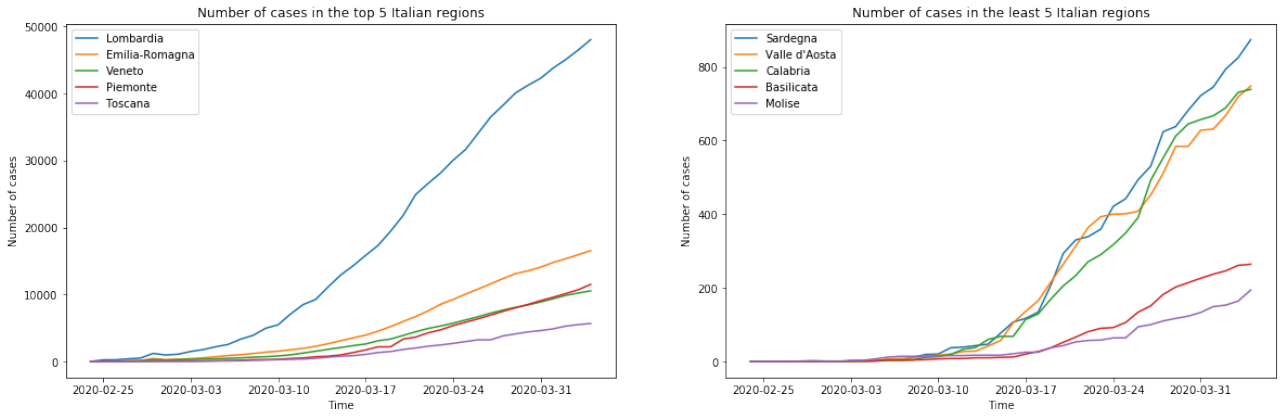

Figure 5. Comparison of the regions with most and least cases.

As we can see from these last figures, the spread of the virus has been different in the various regions of Italy. The most infected regions are the ones in the north while the least infected are in the south (and islands) with the only exception of 'Valle d'Aosta'. The number infections are about 10 times more in the northern regions, with the peak in 'Lombardia'.

Conclusion

The work done is still insufficient to extrapolate useful information in order to mitigate the virus spreading but it could be useful to understand the progress of the COVID-19 in Italy.