

# Analysis of the COVID-19 Pandemic Evolution in Europe

Data Visualization Report – Group M

NOVA Information Management School

Master's Degree in Data Science and Advanced Analytics

2020/2021

**Authors:** Filipe Lourenço (R20170799), Guilherme Neves (R20170749), Rui Monteiro (R20170796), Vasco Pestana (R20170803)

**Dashboard:** Click [here](#).

## 1. Introduction

The main goal of this Data Visualization project is to create an interactive dashboard, that can provide meaningful insights into the evolution of the COVID-19 pandemic in Europe, since its beginning until March 21<sup>st</sup> 2021.

The project was developed in Microsoft Power BI and it is available both in the link where it was deployed (in “Dashboard”, at the beginning of this report) and in the delivered Power BI file (.pbix).

In order to produce a quality application, we implemented slicers (to select specific time frames and European countries), cards with many relevant Key Performance Indicators (KPIs), and several different visualizations, such as a map, treemap, bubble chart, bar chart, line charts, gauge chart, pie chart, donut chart, as well as an infographic chart.

To produce meaningful visualizations, it is essential to have good data from trustworthy sources. This will be discussed in the next section of the report.

## 2. Dataset Description

The data used for our project comes from the *Our World in Data*'s repository, available in this [hyperlink](#). These datasets “are updated daily and include data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest” (Our World in Data, 2021).

The Excel sheet that contains our data has 59 variables, though we did not use all of them. The complete list of descriptions and sources for each variable is accessible [here](#). In fact, the majority of the data gathered by this organization comes from a COVID-19 data repository made by Johns Hopkins University, the European Centre for Disease Prevention and Control, official reports, and from a variety of trustworthy sources (e.g., United Nations, World Bank), among others.

In its raw state, the dataset was not ready to be used for our project. Thus, initially, we used Python to perform data manipulations and transformations, such as: remove

the non-European countries, change some negative values that did not make sense to NaN, put the date-time variables in the correct type, remove some territories/states with many missing values (e.g., Faroe Islands, Gibraltar, Isle of Man), and also remove some small countries (e.g., San Marino, Vatican City, Liechtenstein), due to some limitations when building the map on a map designer tool (it was not possible to select these countries on the map). After this stage, the data was ready to be imported into Power BI.

### 3. Visualization and Interaction Choices

The two main inspirations for this work were two visualizations available [here](#), as well as [here](#). We got some ideas regarding the background, layout and graphics. Nevertheless, we used different colors and visualizations, elaborating the interactivity so who is visualizing can have an immersive experience with our dashboard.

Besides the previous mentioned inspirations, as we live in a pandemic context, we decided that it would be interesting and useful to have a dashboard showing data for the European continent, as well as data for each specific country. The first tab contains a colored map representing the total COVID-19 cases per million people in each country, a tree map showing the total COVID-19 cases for each European country, a visualization representing the top 10 European Countries with most COVID-19 deaths, a bubble chart showing the relationship between COVID-19 cases per million inhabitants, human development index and COVID-19 deaths per million and some relevant KPIs related with COVID-19.

Furthermore, by choosing a country in the colored map of the first tab, it is possible to move into a second tab where all the data displayed is related with the chosen country. In this second tab, it is possible to visualize the same KPIs, but regarding the chosen country. It is also possible to visualize the evolution of daily cases, a comparison between the COVID-19 patients in ICU and off ICU, the stringency index (quantifies the COVID-19 measures' strictness), a comparison between the fully vaccinated people and total population, as well as another one between hospital beds and hospital patients, and the average percentage of hospital patients in ICU for a given time period.

In both tabs, the user can choose a specific time period, where all the displayed data and visualizations will be adjusted for the selected period, enabling to understand in more detail what happened in a time frame, for example, to notice an increase or a decrease in COVID-19 cases or deaths between two dates.

### 4. Technical Aspects

We decided to develop and to produce the dashboard using Microsoft Power BI, publishing it on the web and making it available through a link. However, we opted to perform the data preparation phase using Python, since it is the tool we have been using for this in other projects. Both files, the python notebook and the Power BI file are in a GitHub repository, available [here](#).

Regarding Power BI, we had to use the query editor to prepare the data to be used and we also had to construct some DAX measures, such as averages, sums and measures to calculate the values displayed in the KPIs in the corresponding time period.

## 5. Discussion

With this work, we had the opportunity to develop our skills in Power BI and develop a very interactive dashboard, where we can observe and analyze the evolution of the pandemic situation in which we find ourselves, both from a general European perspective and from the individual perspective of each country, through visualizations of easy understanding, so that the user can comprehend and use the dashboard to its full potential.

During the realization of this project, we had some limitations, namely in the part related to data. This limitation has to do with the fact that we are dependent on the data published by each country and the periodicity with which these same countries publish it (for instance, some countries report their respective COVID-19 statistics on a daily basis, while others report them weekly).

As a perspective of future work, there is room for some improvements, such as expanding the scale of the work from a European perspective to a global perspective, thus opening more horizons in terms of possible analyses and adding more information, and consequently more visual elements, about the vaccines and the impacts they had on the evolution of the pandemic situation of COVID-19. In the future, it could also be possible to enable the dashboard to be fed with up-to-date data through a pipeline which would connect the original data source containing the COVID-19 dataset already mentioned with the Power BI dashboard, in order to enable the end user to have the most updated information about the pandemic evolution.

## References

Our World in Data (2021), *Coronavirus Pandemic (COVID-19)*

Available at: <https://ourworldindata.org/coronavirus>

Retrieved on 22/03/2021