PRINCIPAL ACADEMIC TUTOR

Prof. Stefano Ceri, DEIB,
Politecnico di Milano

OTHER ACADEMIC TUTORS

Prof. Elena Baralis - DAUIN,
Politecnico di Torino

Prof. Fabio Pammolli - DIG,
Politecnico di Milano

EXTERNAL INSTITUTIONS

CERN, One Health Center
of Excellence , ISI Foundation

EXTERNAL TUTORS

Prof. Ilaria Capua - University
of Florida

TEAM MEMBERS

Christian Candedda
Politecnico di Torino

Federico Capello
Politecnico di Milano

Alessio Cappellato
Politecnico di Torino

Luigi Maninchedda
Politecnico di Milano

Leonardo Meacci
Politecnico di Milano

Sofia Peracchi
Politecnico di Milano

Claudia Salerni
Politecnico di Milano

Covid@Lombardy

Executive summary

At the end of 2019, a new disease called **COVID-19** hit the entire world. Firstly identified in Wuhan (China), it spread around the globe, infecting and killing people belonging to the most diverse spheres. However, it did not affect all regions in the same way. Therefore, the goal of this research is to understand **why Lombardy** has been so drastically hit during the first two waves of the outbreak, investigating the **external contributing factors** that may have affected the density of COVID-19 cases, and the effectiveness of the **anti-contagion policies** employed in Europe.

To achieve this objective, an aggregated **open-source dataset** was created, to include a heterogeneous set of twenty-two variables, related to six macro areas of interest: economy, healthcare, population, primary sector, mobility and education. This dataset, together with **machine learning techniques**, revealed that the same external contributing factors, that were important for the prediction of the risk-category of COVID-19 density for a specific region, were also the ones displaying Lombardy as an outlier when compared to other European regions. The main predictive contributing factors were **life expectancy** (positively correlated), the amount of **working hours** (positive correlation).

On the other hand, the analysis of the policies and their effectiveness was based on the data available in the CoronaNet dataset, which collects the worldwide application of **non-pharmacological regulations**. The application of **statistical descriptive methods** computed the effectiveness of each policy type, by comparing the contagion curve trend between when the policy is enacted and 7-20 days after, when its effects should become evident. These analyses showed that the **restriction of businesses and interregional transit** and the enactment of **lockdowns** have been the most impactful policies. It is important to underline that these results must be considered as a general indication, since the implementation of similar policies can be enacted with varying degrees of strictness, which are not captured by the considered dataset.

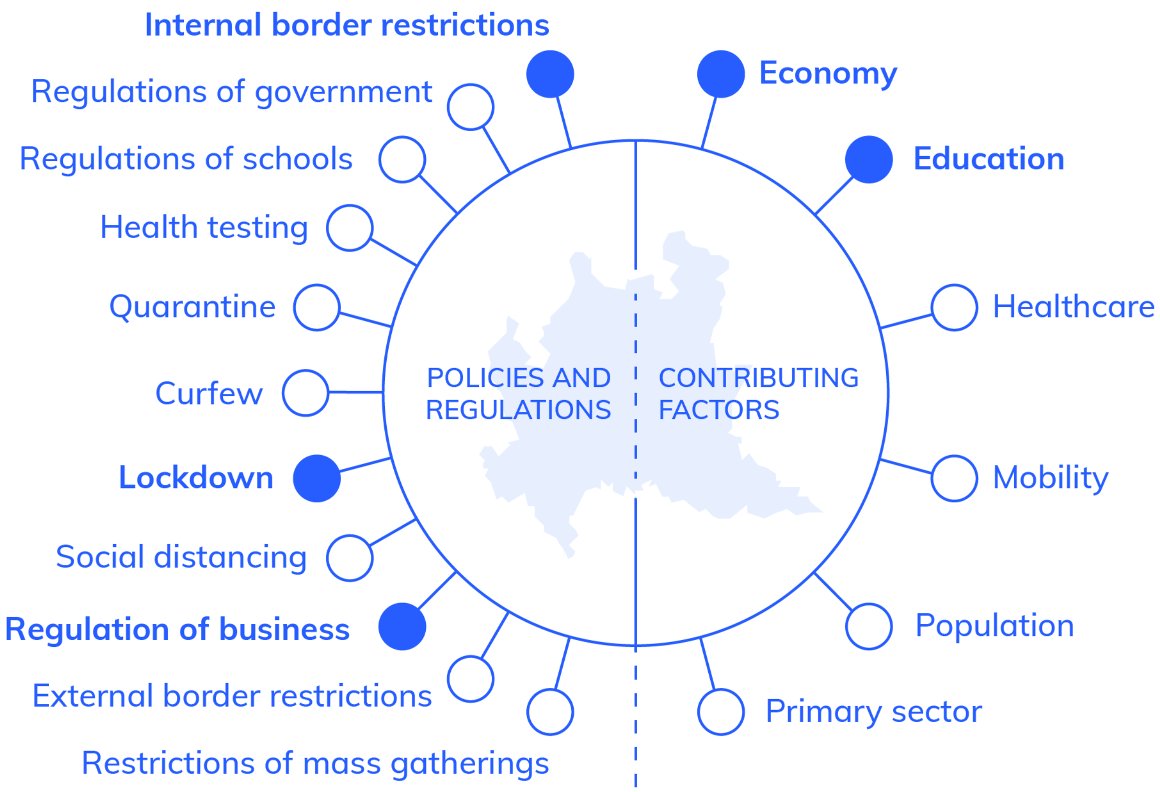
Finally, since this pandemic revealed the need for accurate data representation, these analyses are published on a **public website**, named **CovidatLombardy**, available to anyone who is interested in the topic. The website will not only provide a detailed description of the analysis and its results, but it will also allow the audience to freely navigate the data through interactive maps. The website will have both an **author-driven** and a **reader-driven approach**, creating a meeting point between scientific and information visualisations.

TAGS

COVID-19 - LOMBARDY - CLASSIFICATION - EXTERNAL FACTORS - POLICIES

Covid@Lombardy

Our project aims to investigate the reasons behind the spread of the virus, particularly in Lombardy, which appeared to be more heavily hit during the pandemic than other European areas.



RESULTS

Regulations of businesses proved to be the most effective policy

Higher work hours lead to a higher risk of infection

Regulations of schools did not show particular efficiency in limiting infections

Number of students enrolled and **early leavers from education**, facilitate people mobility and contact

Life expectancy is directly correlated with the number of COVID-19 cases: **the older the population, the higher the risk**

Lockdown and **Internal border restrictions** were effective against the virus. Limiting circulation lowers the chances of infection.

Project description written by the Principal Academic Tutor

The project has the aim of addressing the issues related to COVID-19 in Lombardy, in relationship with the **Interdisciplinary Convergence Initiative launched by Ilaria Capua** and colleagues from the One Health Center of Excellence at the University of Florida, entitled Is Lombardy an outlier or the first of many? Students' and advisor's interest, as well as data availability, will drive the selection of the most relevant issues to be addressed by the project. A multi-disciplinary approach, based upon the **"circular health" concept**, will be used. Beneficiaries are present and future generations, in the hope that **a better understanding of the causes of the virulent COVID-19 spread in Lombardy** will teach us some lessons. The activity in the project will be data-driven, counting on availability of data from collaborating institutions. For what concerns the specific plan of action, students will use a **data science approach**, going through the phases of: data collection and preparation, data exploration, model generation and eventually decision making as a critical analysis of the obtained results. **Findings will be shared with local governments and all interested institutions** to empower their decision process and enable the identification of the best behaviors for dealing with the considered problem. All data and results produced within the study will be public and accessible to general public through **user-friendly interfaces**.

Team description by skill

The Covid@Lombardy team consists of seven members with multifaceted backgrounds: **Christian Cancedda** and **Alessio Cappellato** are two data science engineers from Politecnico di Torino, **Federico Capello** is a mathematical engineer from Politecnico di Milano, **Luigi Maninchedda** and **Leonardo Meacci** are two management engineers from Politecnico di Milano, **Sofia Peracchi** is a communication designer from Politecnico di Milano and **Claudia Salerni** is a biomedical engineer from Politecnico di Milano. Christian was in charge of the technical analysis, in particular regarding the contributing factors, together with Federico. Alessio developed the technical analysis, in particular regarding the policies, and of the elaboration of visual data representations of the results. Luigi and Leonardo maintained all the relations with tutors and other partners, supporting other team members in various tasks, controlling project development. Sofia was in charge of the presentation designing, website development, data visualisation and overall communication strategy. Claudia was responsible for supervising and participating in the whole analytical process, supporting the creations of communication artefacts.

Goal

This study has two main objectives. Firstly, it aims to identify what are the **contributing factors that may have affected the spreading of COVID-19** in European regions during both waves, split in three different periods:

1 - First period defined as starting from March 1st, 2020 through August 20th, 2020 (first wave)

2 - Second period defined as starting from March 1st, 2020 through April 31st, 2020 (initial part of the first wave)

3 - Third period defined as starting from August 20th, 2020 through February 20th, 2021 (second wave).

To perform the analysis, we have considered different aspects, both directly related to SARS-CoV-2 and not: these were the basis to understand which factors are connected to a higher risk of spread. Morevore this project also aims to discover whether these **contributing factors were particularly relevant in Lombardy** with respect to other European regions, in order to understand **why Lombardy** was hit more.

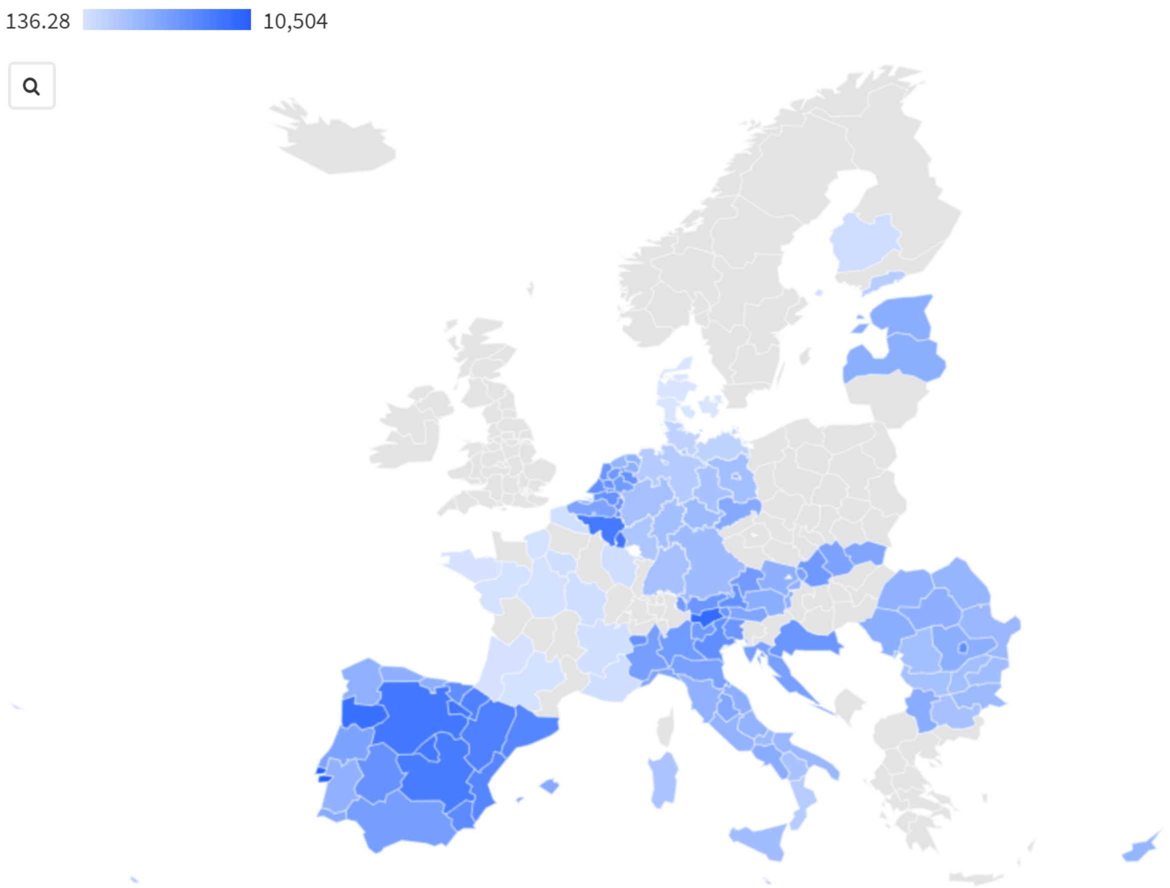


Image 01: The map represents che density of COVID-19 cases in European regions, the focal variable of the study.

The second objective is to inspect how different European regions have implemented anti-COVID-19 regulations, highlighting the main differences and similarities. The ultimate aim is to discover which have been the **most effective policies in limiting the spread** of COVID-19 in Europe.

To sum up, both information are crucial to further developing and assessing models for discovering the hidden pathways behind the spread of the virus, and this contribution aims to offer **new knowledge and insights** to the new, yet ever-growing academic field born around the COVID-19 pandemic.

Understanding the problem

Before starting the analysis, it was necessary to **study the problem and the previous related works**, in order to understand which possible variables could have affected the spread of COVID-19 and could consequently explain the differences of cases density among European regions.

Ciotti et al. (2020) reported that **population density** should be an interesting parameter to take into consideration when investigating a virus outbreak, in particular considering the high percentage of asymptomatic carriers. Moreover, it is a proxy for other variables of interest: Aabed et al. (2021) pointed out that higher population density corresponds to higher **pollution** (%CO2), which could be linked to the spread of a respiratory disease like COVID-19. Hossain (2020) reported that several **social and economic variables** may have contributed to the outbreak, also because they influenced the government interventions. Allel et al. (2020) stated that factors such as **healthcare resources** and their costs are correlated with diseases' spread, in particular in older populations; Kumar et al. (2021) additionally related these parameters to the level of e**conomic wellness, education and mobility**, since more dynamic communities increase the risk of spread.

Hartl et al. (2020) suggested that **non-pharmaceutical interventions** like border restrictions, quarantine and isolation could have helped the reduction of the spread of the pandemic. Cowling et al. (2020) also proposed **social distancing** and changes in the population behavior as limiting measures. However, Hsian et al. (2020) underlined the **social and economical consequences** due to an unsustainable implementation of such regulations.

Exploring the opportunities

By adopting a data-driven analysis, we explored possible ways to carry out studies that would involve heterogeneous data sources, in order to obtain a holistic view of the pandemic.

Thanks to unprecedented availability of **open data**, there are many opportunities for search of new knowledge on epidemic events. From these, we established that data-driven studies could be carried out at the biological level [4] as well as the macro level by means of data about **regional population characteristics** [1,5,6] and **policy makers' decisions** [2].

To this end, a variety of studies can be carried out in order to assess: **correlation of the features of a geographical region with its epidemic condition**; a posteriori **effectiveness of the adopted policies** in containing the spread of the virus[2]; the correlation between the virus and people's mobility[5,6];

By means of such studies, there are opportunities to **understand how the economic, healthcare, social conditions and population characteristics relate to the spread** in a geographic area. This is fundamental, as the availability of large amounts of data allowed us to assess if biases from studies of the past (e.g. limiting mobility is the only measure to counter epidemics) hold or if there are more effective ways to manage the epidemics.

In order to satisfy the requirements initially stated by our research direction, we decided to focus on the study of: the relationship between macro scale features and the spread of COVID-19 [1] and epidemic management policies [2]. These are the two analysis frameworks that, based on the outcome of the study, can allow us to find insights related to the Covid@ Lombardy research questions.

Generating a solution

Our solution consists of two studies aimed at answering separately to the research questions of our project:

1 -the relationship between **macro scale features and the spread of COVID-19** [1];

2 -the **effectiveness of epidemic management policies** [2].

The first focuses on the search for insights on the main reasons for the high or low epidemic spread in the European regions. This allows to define a framework of comparison by means of which the features of Lombardy can be compared to all other regions to determine whether Lombardy is an outlier.

The first study required a dataset capable of characterizing a geographic area(such as Lombardy) at the regional level (**NUTS2**), in terms of **economic, social, demographic and healthcare features**. For each region, its number of cases per hundred thousand inhabitants, in each pandemic wave, has been used as target variable to determine **low and high risk classes of exposure to the virus**. A **machine learning model** has then been employed in order to assess which of the features used by the model characterize a given risk class, and if they are **positively or negatively correlated** with each category.

For the purpose of the second study, also in this case, we built a **regional level dataset** (NUTS2) from the Coronanet policy usage dataset [2]. This allowed us to characterize European regions based on their adopted policies, both in terms of usage and effectiveness in reducing the spread of the virusr. Then, we used this construction to observe if similar responses to the adoption of certain policies are shared between different regions.

We clustered regions based on their policy usage and effectiveness. By means of this framework, we observe by means of a **descriptive methodology**, the outcome of the adoption of certain policies with respect to the pandemic condition of the regions under study.

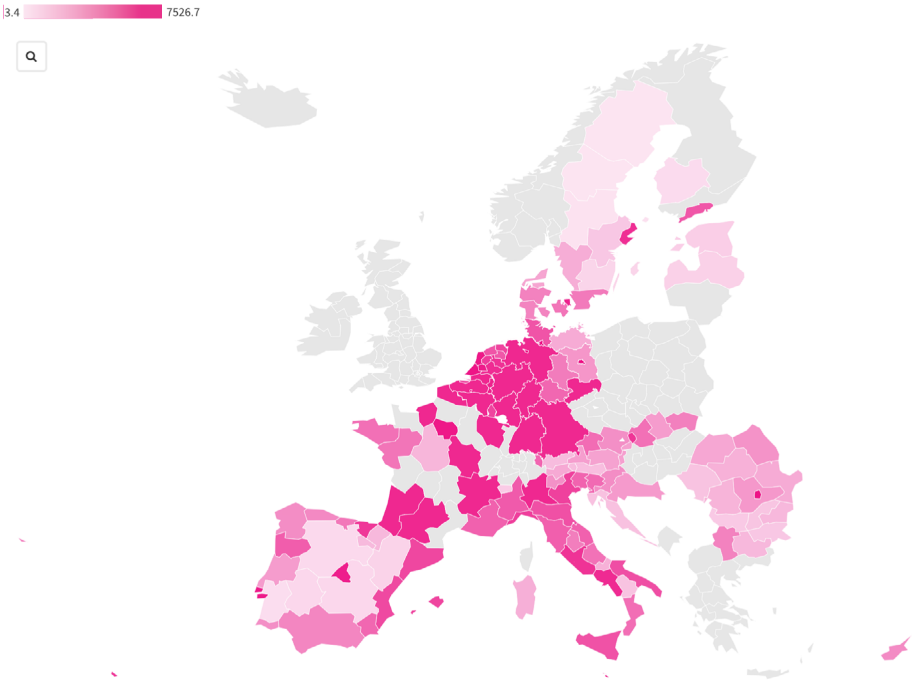


Image 02: The map represents the population density in European regions

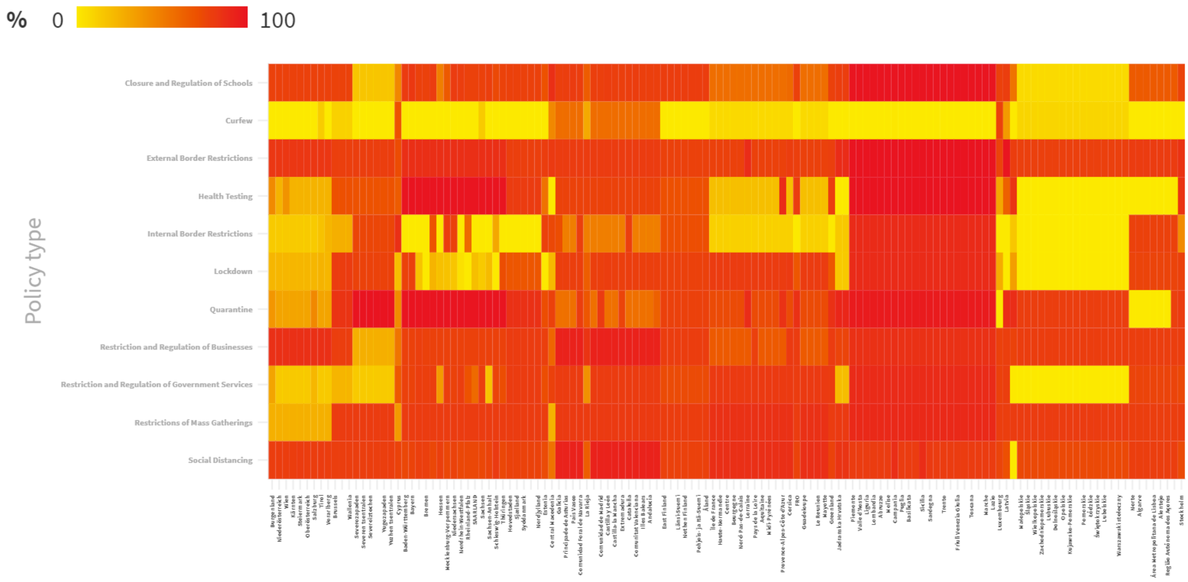


Image 03: The heatmap represents the percentage of utilization of restrictive policies in European regions

At the end of the analysis and the models application, we identified a set of features that characterize the susceptibility of certain European regions to the spread of the virus. We determined how **Lombardy is characterized in terms of such features**.

Life expectancy, labor force and educational indicators have been found as key discriminating factors between European regions at low or high risk.

These features also assume extreme values in Lombardy, that is:

→ **high life expectancy,**

→ **high number of work hours**

→ **high number of students enrolled in tertiary education and high number of early leavers from school**

Lombardy showed to possess features that are also shared by other high risk regions.

Image 04: The bubble graph shows which are the most relevant factors influencing the spread of the virus.

From the point of view of the policy interventions for the limitation of the spread of COVID-19, the exploratory analysis revealed that regional approaches heavily cluster according to their country of belonging both for the type and temporal extent of policy activation. For this reason, **a national aggregation** of this type of information achieves a good approximation. The application of clustering techniques to the resulting data points shows that European countries can be categorized depending on their similarity in the use of preventive regulations. Furthermore, a descriptive analysis was performed to derive the effectiveness of the various types of policies with a data-driven approach: the obtained results indicate that the **regulation of businesses and of local transit** achieve the best deceleration of the COVID-19 cases growth during the second week after their introduction, while the **enactment of lockdowns** requires more time to reach their full effects.

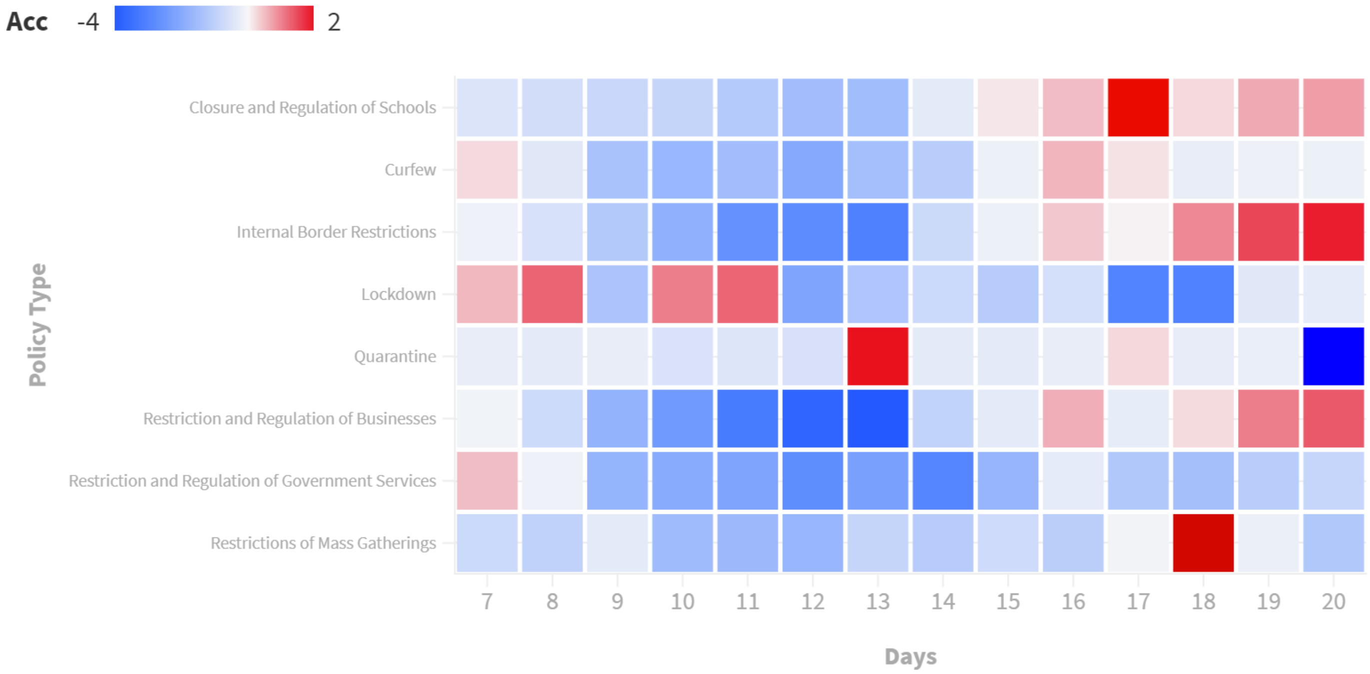


Image 05: The heatmap shows the average impact of every policy 7-20 days after its introduction on the acceleration of the COVID-19 curve.

References

[1]Eurostat open dataset : https://ec.europa.eu/eurostat/data/database

[2]Coronanet database: https://www.coronanet-project.org/

[3]Coronavirus open data: https://github.com/ec-jrc/COVID-19

[4]Coronavirus sequences data: https://www.ncbi.nlm.nih.gov/sars-cov-2/

[5]Google mobility data: https://www.google.com/covid19/mobility/

[6]Facebook mobility data: https://data.humdata.org/dataset/movement-range-maps?

[7] Cdc past epidemics https://www.cdc.gov/flu/pandemic-resources/basics/past-pandemics.html

Other main bibliographic references

Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. B., & Bernardini, S. (2020). The COVID-19 pandemic. Critical reviews in clinical laboratory sciences, 57(6), 365-388.

Aabed, K., & Lashin, M. M. (2021). An analytical study of the factors that influence COVID-19 spread. Saudi Journal of Biological Sciences, 28(2), 1177-1195.

Hossain, M. A. (2020). Is the spread of COVID-19 across countries influenced by environmental, economic and social factors?. medRxiv.

Allel, K., Tapia-Muñoz, T., & Morris, W. (2020). Country-level factors associated with the early spread of COVID-19 cases at 5, 10 and 15 days since the onset. Global Public Health, 15(11), 1589-1602.

Kumar, R., Pandey, A., Ibsa, R. G., Sinwar, D., & Dhaka, V. S. (2021). Study of social and geographical factors affecting the spread of COVID-19 in Ethiopia. Journal of Statistics and Management Systems, 24(1), 99-113.

Hartl, T., K. Wälde and E. Weber (2020a). Measuring the impact of the German public shutdown on the spread of Covid-19. Covid Economics, Issue 1: April 03 2020, 25-32.

Cowling, B.J., S.T. Ali, T.W.Y. Ng, T.K. Tsang, J.C.M. Li, M.W. Fong, Q. Liao, M.Y.W. Kwan, S.L. Lee, S.S. Chiu, J.T. Wu, P. Wu and G.M. Leung (2020). Impact assessment of nonpharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. The Lancet Public Health, April 17 2020, pp. 1-10

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L.Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng and T. Wu (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature