INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Analysis of the influence of non-pharmaceutical interventions and cultural differences on the evolution of COVID-19

Margarida Martins Duarte

Master in Computer Science

Supervisor:

PhD Catarina Eufémia Domingues Alves Ferreira da Silva, Associate Professor with Habilitation

ISCTE-IUL

Co-Supervisor:

PhD Sérgio Miguel Carneiro Moro, Associate Professor with Habilitation

ISCTE-IUL

Novembro, 2021

Department of Information Science and Technology

Analysis of the influence of non-pharmaceutical interventions and cultural differences on the evolution of COVID-19

Margarida Martins Duarte

Master in Computer Science

Supervisor:

PhD Catarina Eufémia Domingues Alves Ferreira da Silva, Associate Professor with Habilitation

ISCTE-IUL

Co-Supervisor:

PhD Sérgio Miguel Carneiro Moro, Associate Professor with Habilitation

ISCTE-IUL

November, 2021

“Look up at the stars and not down at your feet. Try to make sense of what you see, and wonder about what makes the universe exist. Be curious.”

- Stephen Hawking

Acknowledgments

I want to express my gratitude to a set of people who were my support for the realization of this dissertation.

First of all, I would like to thank Professor Dr. Catarina Ferreira da Silva and Professor Dr. Sérgio Moro for the opportunity and vote of confidence. I do not have enough words to show my gratitude for your help and motivation during this journey which allowed me not only to achieve a goal but also my personal growth.

I also want to thank from the bottom of my heart my boyfriend Miguel who was my pillar and who always believed in me. He was the one who gave me strength in the most difficult moments and never let me give up, thank you for your patience and all the love.

Last but not least, my many thanks to my family in particular, my mother for her dedication and love that allowed me to fly higher and my father and sister who supported me throughout my academic journey.

Resumo

Desde o início da pandemia COVID-19 que os países do mundo inteiro implementaram um conjunto de intervenções não farmacêuticas (NPIs) como forma de combater a evolução da pandemia.

Nesta dissertação propomos uma análise detalhada da evolução da pandemia que considera o grau de restrição das NPIs desde Março de 2020 a Maio de 2021 e a taxa de reprodução em cinco países: India, Brasil, Reino Unido, Israel e Portugal. Além disto, analisamos o impacto que as dimensões culturais de Hofstede podem ter entre as implementações de vários graus de restrição de NPIs e a taxa de reprodução aplicando modelos de *machine learning* para compreender se as características culturais são informações úteis para melhorar as previsões da taxa de reprodução. Para concretizar estes objetivos, nós seguimos a metodologia do CRISP-DM sendo que, reunimos dados de *Our World in Data COVID-19*, *Oxford COVID-19 Government Response Tracker* e o *website Hofstede Insights*.

Mostramos uma análise aprofundada e extensa ao longo destes meses da pandemia a qual mostra diferenças entre os cinco países que implementaram as mesmas NPIs em diferentes graus e em que a cultura desempenha um papel importante na resposta de cada país às várias NPIs implementadas.

Palavras-chave: COVID-19, NPIs, intervenções não farmacêuticas, dimensões culturais de Hofstede, machine learning

Abstract

Since the beginning of the COVID-19 pandemic, countries worldwide have implemented a set of Non-Pharmaceutical Interventions (NPIs) as a way to face the evolution of the pandemic.

In this dissertation we propose a detailed analysis of the evolution of the pandemic that considers the degree of restriction of NPIs from March 2020 to May 2021 and the reproduction rate in five countries: India, Brazil, United Kingdom, Israel and Portugal. In addition to this, we analyse the impact that Hofstede's cultural dimensions may have between implementations of various degrees of restriction of NPIs and the reproduction rate by applying machine learning models to understand whether cultural characteristics are useful information to improve reproduction rate predictions. To achieve these objectives, we follow the CRISP-DM methodology being that we gather data from Our World in Data COVID-19, Oxford COVID-19 Government Response Tracker and Hofstede Insights website.

We show an in-depth and extensive analysis over these months of the pandemic which shows differences between the five countries that have implemented the same NPIs to different degrees and where culture plays an important role in each country's response to the various NPIs implemented.

Keywords: *COVID-19*, NPIs, Non-Pharmaceutical Interventions, Hofstede’s cultural dimensions, machine learning

Contents

Figures

Tables

Acronyms

WHO	World Health Organization
COVID-19	Coronavirus Disease 2019
NPIs	Non-Pharmaceutical Interventions
R_t	Reproduction rate
SIR	Susceptible Infection Recovery
UK	United Kingdom
ICU	Intensive Care Units
CRISP-DM	CRoss Industry Standard Process for Data Mining
DSR	Design Science Research
EBSCOhost	Elton B. Stephens COmpany host
GDP per capita	Gross Domestic Product per capita
SEIR	Susceptible Exposure Infection Recovery
CIG	Confirmed Infection Growth
IRTCCPM	Increase Rate of the Total COVID-19 Cases per Million
OxCGRT	Oxford COVID-19 Government Response Tracker
C	Containment and Closure policies
E	Economic policies
H	Health System policies
EDA	Exploratory Data Analysis
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
BRICS	The group of economies of Brazil, Russia, India, China and South Africa
SARS-CoV-2	Severe Acute Respiratory Syndrome Corona Virus 2
PDI	Power Distance

INV	Individualism
MAS	Masculinity
UAI	Uncertainty Avoidance
LTO	Long Term Orientation
ING	Indulgency
MI	Mutual Information
LASSO	Least Absolute Shrinkage and Selection Operator
LARS	Least-Angle Regression
SGD	Stochastic Gradient Descent
KNN	K Nearest Neighbours
SVR	Support Vector Regression
XGBoost	EXtreme Gradient Boosting

Chapter 1 Introduction

On the 30th of January of 2020, the World Health Organization (WHO) declared the coronavirus disease-2019 (COVID-19) as a Public Health Emergency of International Concern (Bo et al., 2021; González-Bustamante, 2021; Sohrabi et al., 2020). Since then, the world faces a global health crisis in which the development of new variants of the coronavirus that cause COVID-19 further aggravates the infection rate (Fontanet et al., 2021). Worldwide, governments mobilized their resources to fight the COVID-19 pandemic through a set of Non-Pharmaceutical Interventions (NPIs) (Brauner et al., 2021; Fontanet et al., 2021) that, allied to the increase of vaccinations, is considered crucial to the decrease of its propagation until herd immunity is achieved (Fontanet et al., 2021).

Each country adopted different measures over time with different restriction levels and studying the effectiveness of these measures can help in understanding which NPIs are the most influential regarding the virus' propagation. In this context, the R_t, also known as reproduction rate indicator "is defined by the expected number of secondary cases arising from a primary case infected at time t" (Li et al., 2020), i.e. the number of effective reproduction. Hence, it measures how many people an infected person can infect. It constitutes an important measure in infectious diseases' epidemiology by reflecting several factors such as control measures, contact rates and climatic conditions (Arroyo-Marioli et al., 2020).

By identifying the most effective NPIs, it is possible for stakeholders to implement a selection of key measures in a timely manner to fight COVID-19 or other future pandemics (Haug et al., 2020). In addition to the impact of the NPIs, considering socioeconomic and cultural factors can be an asset to identify vulnerable cultural minority groups that become a hotbed of infection by COVID-19 in order to address their needs and then also to mitigate the spread of the virus (Waitzberg et al., 2020).

1.1 Motivation

Since the beginning of the COVID-19 pandemic, it became important to study its impact in different countries. Empirical evidence on the effectiveness of specific policy interventions has been limited (Liu et al., 2020). Therefore, our study proposes to make a detailed analysis in which it considers the degree of restriction of the NPIs over time as well as other aspects not considered in previous studies such as the vaccination policy NPI. In addition, in this dissertation cultural characteristics were considered because cultural attributes account for some of the global disparities in COVID-19 attributed health outcomes (Erman & Medeiros, 2021).

During the early stages of the spread of a virus, contagion risk models (e.g. Susceptible Infection Recovery models – SIR) were used, however, because contagion depends on individual and social behaviour, these models are difficult to calibrate at a relatively small scale and their spatiotemporal predictions are highly uncertain (Azevedo et al., 2020). In addition to this, the machine learning models made in previous studies only considered data from the year 2020 (Güss & Tuason, 2021; Luu & Huynh, 2020; Wang, 2021; Yeung et al., 2021).

For these reasons we first chose five different countries that we considered interesting to analyse and compare. In this sense we chose to analyse the impact of the NPIs on the R_t for India, Brazil, the United Kingdom (UK), Israel and Portugal, considering the various degrees of NPI restrictions until May 2021. We intend to analyse and understand the impact of the NPIs on the pandemic control and the response of each country over time. Brazil, India and the UK are countries from different continents, populous and had a high number of cases of infection by COVID-19 at the beginning of the pandemic. Israel and Portugal are also considered because they are small countries with less population. We chose two countries from the European continent (Portugal and UK) and two from the Asian continent (India and Israel) so that we could compare their dimensional and populational differences in the evolution of COVID-19. All countries are culturally different, which makes it even more interesting to analyse.

Next, in this dissertation we use data from 101 countries until May 2021 to test several machine learning models with and without the cultural dimensions so that we can compare the results.

1.2 Objectives

The objectives of this work are the following:

- Analyse the degrees of restriction of the NPIs, the R_t and other variables, such as new cases, new deaths, hospitalized patients and Intensive Care Units (ICU), from the beginning of the pandemic until May, 2021;
- Analyse whether there are differences in the implementations of various degrees of restriction of NPIs and their corresponding R_t impact in countries, as well as the influence of cultural dimensions to improve the prediction of the R_t evolution.

Therefore, we analyse the impact of the NPIs on the R_t for India, Brazil, the UK, Israel and Portugal, considering the various degrees of NPI restrictions, in which through visually appealing graphics we can observe the evolution of the R_t and explain the existing differences through the underlying culture of these countries, based on the Hofstede framework. Additionally, several machine learning models were analysed to understand the influence of cultural dimensions on R_t prediction.

1.3 Research Questions

The research questions we propose to answer with our research are:

- How can cultural characteristics explain the introduction and consequent impact of the NPIs in different countries?
- Can cultural characteristics be useful information for predicting the R_t?

Culture refers to a system of values that are shared and the collective practices that distinguish one society from another (Hofstede, 1984). Each country has its own cultural characteristics which can have a positive or negative impact on compliance with the NPIs implemented.

Through our research questions we try to understand if, for instance, countries with higher power distance (e.g., India) tend to not comply with the established NPIs. Additionally, through machine learning models we understand whether the use of cultural dimensions together with the implemented NPIs and socio-economic characteristics improves the prediction of the R_t.

1.4 Research Method

In this dissertation the methodology used is the CRoss-Industry Standard Process for Data Mining (CRISP-DM). According to Moro (2015) CRISP-DM can be an instance of the Design Science Research (DSR) methodology.

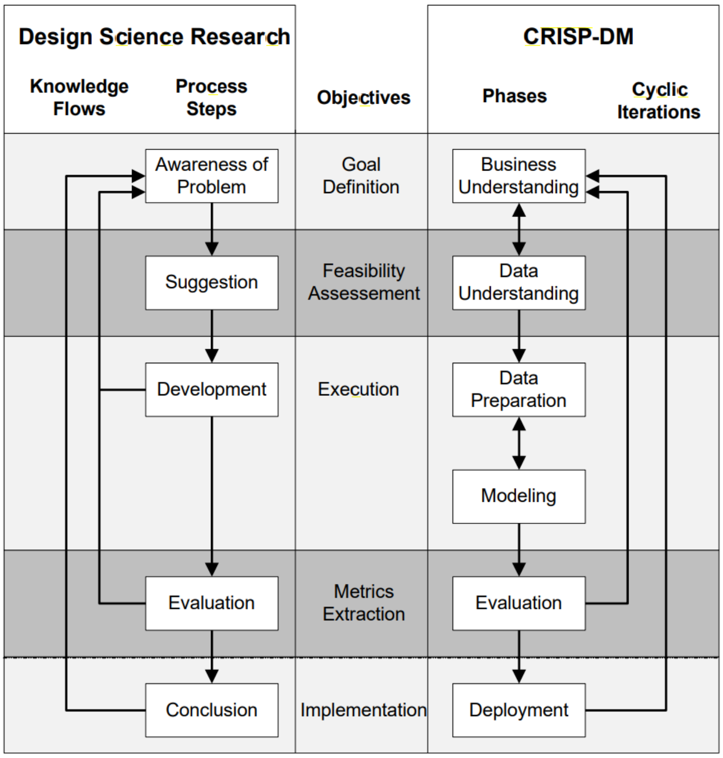

Figure 1: The CRISP-DM methodology according to Moro (2015)

Figure 1 shows the similarities between the DSR and CRISP-DM methodologies in which the phases of each can be compared since they present identical objectives. The CRISP-DM methodology defines a non-rigid sequence of six phases, which allow the construction and implementation of a Data Mining model to be used in a real environment (Moro et al., 2011).

According to Wirth & Hipp (2000) this methodology includes six phases. The first phase is business understanding, which consists of defining the problem by understanding the business objectives and requirements. The second phase is called data understanding which includes initial data collection and exploration to gain initial insights into the data, or to detect hidden information. This is followed by the data preparation phase which covers all activities to build the final dataset from the initial raw data, e.g., selecting features, creating new features, and cleaning missing values. In the modelling phase various modelling techniques are selected, applied and tuned which is followed by the evaluation phase where the built model is evaluated to make sure it correctly achieves the goals defined in the first phase. The deployment corresponds to the final phase of the project where the knowledge acquired in the previous phases will be presented in order to proceed to its implementation.

1.5 Document Structure

This dissertation is structured as follows:

- **Chapter 2 Literature Review** presents the existing studies in the research area namely, on the influence of NPIs and Hofstede's cultural differences on the evolution of COVID-19;
- **Chapter 3 Methodology** explains the different phases of application of the CRISP-DM methodology;
- **Chapter 4 Results** shows the results and analysis of the implementation of the various degrees of NPIs over the pandemic months through heatmaps and line graphs as well as the results of the predictions of the machine learning models;
- **Chapter 5 Discussion** compares the implementation of the restriction degrees of the countries' NPIs considering their cultural characteristics. The results of the models in predicting R_t with and without cultural dimensions are also analysed;
- **Chapter 6 Conclusions** summarizes the main conclusions, in which the research questions proposed in 1.3 are answered and limitations and future work are presented.

Chapter 2 Literature Review

We present a review of the existing studies about the impact of NPIs throughout the evolution of the pandemic and the influence of cultural dimensions. To conduct this research, we used Google Scholar and EBSCOhost as search engines in which we used the following search terms: COVID-19, non-pharmaceutical interventions, cultural dimensions, Hofstede’s cultural dimensions, and machine learning. This search was conducted between March and October 2021.

Thus, this chapter is divided into the following sections:

- **Section 2.1** includes the studies of a set of countries or focusing on the studies of a particular country in which the impact of NPIs in relation to the evolution of COVID-19 are analysed.
- **Section 2.2** includes the studies that consider the cultural factor with COVID-19 variables.

2.1 COVID-19 and NPIs implementation worldwide

At the end of 2019 in China, the first cases of infection by COVID-19 (Bo et al., 2021; Castex et al., 2020) were identified, generating unprecedent socioeconomic disturbances on a global scale throughout the following months (González-Bustamante, 2021). In this context, mathematicians developed a measure to calculate the number of effective reproduction, i.e., to measure the rate of transmission of an infectious disease in a population (Liu et al., 2020). In this sense, a new method was developed that manages to estimate the R_t in real time from three steps (Arroyo-Marioli et al., 2020): in a first phase, they constructed a time series that quantified the infected individuals at a given moment, then estimated the growth rate through the Kalman filter (Arroyo-Marioli et al., 2020) and, finally, used a SIR model to obtain the R_t from the estimated growth rate.

Since the beginning of the COVID-19 pandemic, several measures have been implemented by countries worldwide with the aim of lowering new cases and new fatalities, these measures are called NPIs (Castex et al., 2020). NPIs refer to a wide range of top-down measures, which are governmental and bottom-up measures that refer to self-initiated behaviours (of each citizen) that aim to interrupt the chains of infection by changing key aspects of human behaviour (Perra, 2021). In this study, we only look at NPIs from the top-down point of view which corresponds to the implementation of NPIs established by the government for each country throughout these pandemic months.

The R_t of individuals depends on contact mixing patterns between the infected individual, age-specific susceptibility to infection, region-based behavioural susceptibility, and potential seasonal forcing (Yechezkel et al., 2021). Therefore, two fundamental strategies to control the transmission of COVID-19 can be considered: suppression and mitigation. The suppression strategy consists of implementing and maintaining NPIs to eliminate virus transmission or until the population is vaccinated, aiming to reduce R_t to less than 1, while the mitigation strategy has the function of only reducing the speed of virus transmission until herd immunity is achieved (González-Bustamante, 2021; Neil M et al., 2020).

Through our analysis of the existing literature, we verify that certain studies scrutinize the impact of NPIs in a wide range of countries (Bo et al., 2021; Castex et al., 2020; Duhon et al., 2021; Fuller et al., 2021; Haug et al., 2020; Islam et al., 2020; Jardine et al., 2020; Koh et al., 2020), while others focus on a small set of countries or just in one country (Auger et al., 2020; Chernozhukov et al., 2021; González-Bustamante, 2021; Jefferies et al., 2020; White & Hébert-Dufresne, 2020).

The existing studies refer to the initial period of the pandemic when the implementation of the NPIs seems to have been crucial to control its evolution (Bo et al., 2021; Fuller et al., 2021; Islam et al., 2020). Nevertheless, there are some controversies regarding the impact of these measures, for example according to Duhon et al., (2021) the NPIs did not have a strong association with the initial growth of the pandemic. Through an analysis by time series, the effects of community related NPIs on the value of the R_t were evaluated (Islam et al., 2020), for which no evidence was found of the benefit of the lockdown policy in conjunction with the closure of schools and workplaces, restrictions on mass meetings and confinement in place.

In contrast, Haug et al., (2020) identified lockdowns, limits on social gathering, remote working and school closures as effective measures to decrease the R_t. Also, Koh et al., (2020) highlighted strict border controls, working from home, and total confinement as effective measures for the same purpose.

More restrictive social distance NPIs such as school closures and workplace closures are more effective in countries with lower population density, lower surface area, less air pollution, higher Gross Domestic Product per capita (GDP per capita), lower employment rate, higher health expenses and lower proportion of the elderly population, i.e. over 65 years old (Castex et al., 2020).

According to Bo et al., (2021) the social distancing NPI is more effective in comparison with the mandatory face mask in public, isolation or lockdown, and traffic restriction NPIs.

At continent level, some studies were carried out to understand the effectiveness of NPIs in containing the pandemic, namely in Europe, where countries that implemented stricter mitigation policies when they reached a mortality threshold earlier, tended to report fewer deaths associated with COVID-19 compared to countries that implemented similar policies later (Fuller et al., 2021). In the same sense, according to the results of Koh et al., (2020) a combination of physical distance measures, if implemented early, can be effective in containing COVID-19.

In Muslim Majority Countries despite taking different measures, those with functional democracies were able to contain the pandemic significantly better than those which are not democratic (Jardine et al., 2020).

In the South American continent, some countries like Uruguay and Paraguay, have managed to contain the pandemic relatively successfully, while others appear to be overwhelmed by COVID-19, like Brazil and Peru (González-Bustamante, 2021).

The United States were the target of some studies in order to understand the NPIs that had the most effectiveness in reducing infections. In Auger et al., (2020) the school closure NPI was temporally associated with decreased COVID-19 incidence and mortality. However, according to Chernozhukov et al., (2021) it was difficult to identify the effect of this NPI due to the lack of variation in the timing of school closures across states.

In India, lockdown, if implemented correctly, can reduce the total number of cases in the short term and save time to prepare the healthcare system and monitor diseases (Ray et al., 2020). The results of Douglas et al., (2020) indicate that the combined use of NPIs can significantly improve the impact of the COVID-19 pandemic, although containment of dissemination is difficult without serious community participation.

China has also developed some studies in this area, the main conclusions being that the cities which implemented NPIs pre-emptively reported fewer cases in the first week of their outbreaks compared with cities that started later to control and suspending intracity public transport, closing entertainment venues and banning public gatherings were associated with reductions in case incidence (Tian et al., 2020). According to Pan et al., (2020) the NPIs were temporally associated with improved control of the COVID-19 outbreak.

In relation to New Zealand, through the analysis of a study Jefferies et al., (2020) concluded that the rapid action resulted in the low incidence of new daily cases and in low levels of population disparities by COVID-19.

Some previous studies followed a time series analysis (e.g., Auger et al., 2020; Haug et al., 2020; Islam et al., 2020; Kozlakidis et al., 2020; Li et al., 2020; Liu et al., 2020) to assess the impact of the NPIs on the change in COVID-19 incidence in the period of their pre-implementation and post-implementation. Nevertheless, there is a delay in the effect between an NPI's introduction and withdrawal of 1-3 weeks (Li et al., 2020) because the virus' detection occurs after an incubation period in which symptoms take on average eight days to manifest (Castex et al., 2020).

Compartmental models are recurrently used to describe the evolution of infectious diseases because they require little data about the population under study (Perra, 2021). These epidemiological models include variations of the SIR model guided by a set of differential equations relating the number of people susceptible, the number of people infected, and the number of people recovered or who died at a given time (Ray, 2020). The Susceptible Exposure Infection Recovery (SEIR) model is an extension of the SIR model that incorporates an additional compartment of exposed people that is a latent variable (Ray, 2020).

In Caetano et al., (2021) the authors used the SEIR model in order to estimate the effect of implementing NPIs throughout the pandemic in Portugal in which they took into account the mix of contacts between different age groups depending on contact patterns at work, home, school and other locations as well as, asymptomatic transmission, hospitalization, recovery and death. In this study they found that the implementation of NPIs decreases the number of infections and consequently the number of hospitalizations, ICU and deaths. Similarly, in Silva et al., (2021) the authors built a deterministic mathematical model to test several scenarios of real evolution of COVID-19 through the analysis of the Portuguese population's readiness to follow NPIs collected in social media. In this study the authors realized that pandemic evolutions over the past months in different regions of the world demonstrated that population behaviour is of crucial influence and that the same control policies, implemented in different regions and at different times, resulted in different outcomes, since the public opinion also changes over time.

Yechezkel et al., (2021) developed a dynamic model for the progression and transmission of the COVID-19 virus in Israel considering variables of age, risk, and region in which they found out that impoverished areas were associated with high transmission rates and the NPI of elderly protection was the most effective in reducing mortality while, the NPI of school closure had an adverse effect. Ray (2020) used a time-varying transmission rate eSIR, which described a series of time-varying changes caused by external variations (e.g., NPIs) and internal variations (e.g., mutations and evolutions of the pathogen) in order to predict the number of infected and recovered people or deaths at a given point in time. This model showed that implementing NPIs has a high probability of reducing the number of COVID-19 cases in the short term and allowing time to prepare a healthcare system.

There are several factors that can affect the risk of transmission of COVID-19 including demographic, economic, and environmental characteristics (Auger et al., 2020; Bo et al., 2021; Castex et al., 2020; Chernozhukov et al., 2021; Duhon et al., 2021; Fuller et al., 2021; Islam et al., 2020; Koh et al., 2020; White & Hébert-Dufresne, 2020).

In addition to these factors the underlying culture of each society can also influence the impact of NPIs since, it can influence the collective responses of a group of individuals in their environment (Hofstede, 1984). According to Guan et al., (2020), collective culture-oriented actions and norms in response to the pandemic of COVID-19 will serve as a top-down influence on the individual behaviours of its members, and governments for each country must carefully balance collective and individual interests as well as effectively manage the trade-offs of social control and individual freedom in order to achieve the best outcomes for the public. According to Silva et al., (2021) the social and clinical experience with COVID-19 will leave lasting marks on society and in the healthcare systems, from Latino cultural habits (e.g. proximity, kissing) to changes in the configuration of the healthcare systems, leaving hospitals for more complex clinical situations and providing healthcare institutions with diagnostic and therapeutic means that avoid systematic recourse to hospital emergencies.

2.2 NPIs and culture dimensions impact in COVID-19

Culture refers to a system of values that are shared and the collective practices that distinguish one society from another (Hofstede, 1984). Hofstede (2011) presents six dimensions that represent a country's culture:

1. **Power Distance**, referring to the different solutions to the basic problem of human inequality;
2. **Uncertainty Avoidance**, related to the level of stress in a society in the face of an unknown future;
3. **Individualism versus Collectivism**, which relates to the integration of individuals into primary groups;
4. **Masculinity versus Femininity**, related to the division of emotional roles between women and men;
5. **Long Term versus Short Term Orientation**, relating to the choice of focus for people's efforts: the future or the present and past;
6. **Indulgence versus Restraint**, related to the gratification versus control of basic human desires related to enjoying life.

Lu et al., (2021) states culture fundamentally shapes how people respond to crises such as the COVID-19 pandemic and plays an important role in mask usage. According to Gelfand et al., (2021) people in loose cultures, which have weaker norms and are much more permissive, had much less fear of COVID-19 than people in tight cultures, which have stricter rules and punishments for deviance.

Erman & Medeiros (2021) affirm that during the initial period of the pandemic policy makers should more explicitly consider the cultural attributes of a society along with other important parameters such as demographic characteristics and healthcare system constraints in order to develop more tailored and effective policy responses.

Some studies have already added Hofstede's cultural dimensions along with NPIs in order to study their impact on the pandemic. In particular, they have focused on data from the first wave of the pandemic.

Wang (2021) combined government policies and national culture to explain their influence on social distancing which showed that, the stringency index, a composite measure based on nine response indicators such as school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100, where 100 represents the strictest response, and Hofstede's cultural dimensions long-term orientation and indulgence significantly and negatively affect social distancing. On the other hand, in Gokmen et al., (2021) national cultural dimensions of 31 European countries were related to the rise in the total number of COVID-19 cases per million and obtained that power distance has a significant and negative effect while, individualism and indulgence have a positive and significant effect. The later result diverges from the previous study in which indulgence decreased social distance.

Guss & Tuason (2021) shows that the higher the GDP per capita is, the higher the number of COVID-19 deaths and also, countries with high individualistic values and high intellectual autonomy were associated with a higher number of COVID-19 deaths, while countries with higher collectivistic values were associated with fewer COVID-19 deaths.

Luu & Huynh (2020) studied the impact of the cultural factor on respect for social distancing measures and obtained that countries with high uncertainty avoidance predict the lowest proportion of people gathering in public (e.g., retail and recreation, grocery and pharmacy, parks, transit stations, workplaces).

Yeung et al., (2021) used machine learning techniques with the goal of predicting COVID-19 confirmed infection growth (CIG) over 14 days considering NPIs and Hofstede's cultural dimensions. This study showed that Random Forest regression and AdaBoost regression may be able to predict CIG_{14} with a considerable degree of accuracy.

Although previous studies indicate that culture impacts NPIs, for Wang (2021) governmental strictness has more impact on social distancing than national culture since, people may drop their guard due to exhaustion with social distancing rules after a long time living under the pandemic of COVID-19.

Table 1 shows the studies found during the literature analysis of this dissertation in which the authors consider cultural dimensions to understand their impact on the evolution of the pandemic and the implementation of the NPIs.

Table 1: Previous studies

Reference	Period	Countries	Objective	Datasets	Features types used	Method	Results
Gelfand et al., (2021)	Beginning of the pandemic to Oct 16th, 2020	57 countries	Examine the association between the strength of social norms and the number of cases and deaths in COVID-19	• Our World in Data • OxCGRT • Hofstede Insights	• COVID-19 measures • Socioeconomic measures • Hofstede dimensions	Ordinary least squares regression	Tight groups cooperate much faster under threat and have higher survival rates than loose groups
Erman & Medeiros (2021)	First wave of the pandemic	73 countries	Explore the global variability of public health outcomes during the first wave of the pandemic	• WHO Global Health Observatory Data Repository • Hofstede Insights	• Socio-demographics • Health system capacity • COVID-19 measures • Cultural dimensions • Political dimensions	Random-effects meta-analysis models	Cultural attributes do account for some of the global disparities in COVID-19 attributed health outcomes.
Gokmen et al., (2021)	N/A	31 European countries	Investigate the impact of national cultural dimensions on the increasing rate of the total COVID19 cases per million (IRTCCPM)	• Our World in Data; • VHD European countries; • Hofstede Insights	• COVID-19 measures: cases; • Hofstede dimensions: power distance and collectivism	Explanatory modeling	Power distance has a significant and negative effect on IRTCCPM whereas, individualism and indulgence have significant and positive effects on IRTCCPM
Wang (2021)	First wave of the pandemic	80 countries	Assess the impact of national culture and government policies on social distancing to fight COVID-19	• OxCGRT • Hofstede Insights • Google mobility reports	• COVID-19 measures • Hofstede culture dimensions • Google Community Mobility Reports • Socio-economic measures	Linear regression	Social distancing decreases with long-term orientation and the opposite is true for indulgence
Yeung et al., (2021)	April 1st to September 30th, 2020	114 countries	Predict the national growth of confirmed infections growth in COVID-19 across 14 days	• OxCGRT • Hofstede Insights • Johns Hopkins University	• NPIs measures • Current infection numbers (CIG) • Hofstede cultural dimensions	Ridge, decision tree, random forest, AdaBoost and SVR regressions	Non–time series machine learning models can predict future CIG to an appreciable degree of accuracy
Huynh (2020)	February 16th to March 29th 2020	58 countries	Examine the role of the cultural dimensions in practising social distancing across the world	Google	• Social distancing taking place in various locations • Hofstede cultural dimensions	Ordinary Least Squares regression	Countries with higher uncertainty avoidance index predict the lower proportion of people gathering in public
Güss & Tuason (2021)	July 14th to December 29th, 2020	75-76 countries	Study culture values impact in deaths by COVID-19	• Johns Hopkins University • World Health Organization	• COVID-19 measures • Socioeconomic and demographical measures • Hofstede cultural dimensions	Regression analyses	Countries with high GDP per capita, high individualistic values and high intellectual autonomy were significantly and consistently associated with high numbers of COVID-19 deaths

Chapter 3 Methodology

This chapter describes each of the phases of the methodology applied in this dissertation. Figure 2 schematizes the phases of the CRISP-DM methodology applied in this dissertation, showing the information and activities we considered for each step of the methodology.

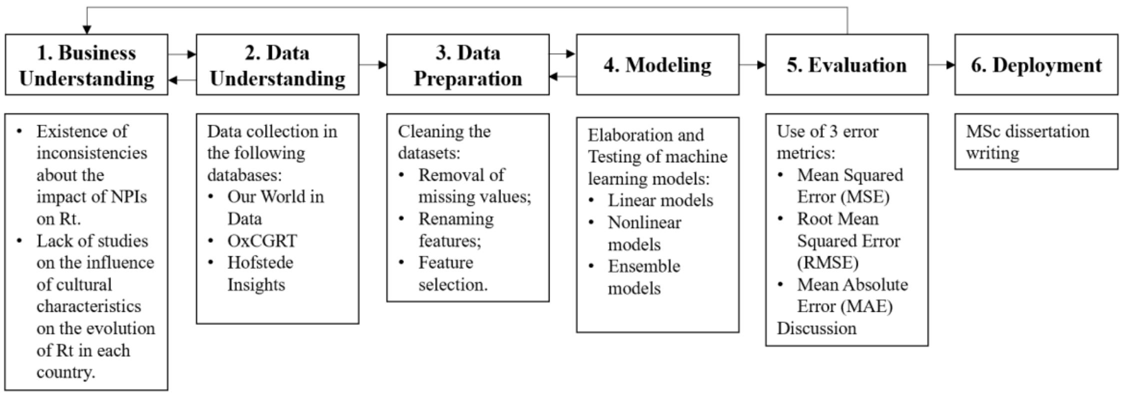

Figure 2: CRISP-DM methodology

Initially, we started from the identification of the existence of a problem recognised during the literature review (Chapter 2) in the scope of the impact of the NPIs. In countries considering their cultural differences more specifically, throughout our literature review we did not find consensus on the impact of implementing NPIs to control the reproduction rate. In addition to this fact, to date we found no studies that apply machine learning models to compare the use of NPIs, socioeconomic and demographic characteristics, and cultural dimensions in predicting the reproduction rate. Of the existing studies, these consider the impact of NPIs and cultural dimensions with COVID-19 cases, deaths and CIG_{14}. Thus, we proceeded to the second phase, data understanding, in which we search for datasets that contained the data needed for our analysis. Next, in the data preparation phase, we performed all the necessary activities for data cleaning, and exploratory data analysis (EDA). After this phase, we applied several machine learning models in order to predict the value of the R_t for each of the countries using the implemented NPIs and Hofstede's cultural dimensions. To evaluate our work we compare our results with other studies in Chapter 5. Also, we evaluate the models developed using three error metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Finally, the Deployment phase is done through this dissertation.

The rest of this chapter contains the following sections:

- **Section 3.1** describes the data sources from which we extracted the data needed to answer our research questions;
- **Section 3.2** presents the procedure we followed to perform the data cleaning, joining and processing;
- **Section 3.3** describes the types of visualization chosen to show the data in order to identify patterns and correlations for the five countries we analysed.
- **Section 3.4** describes the features selected for the models and the models used to predict R_t.
- In **section 3.5** the model validation technique is described, as well as the errors used to evaluate the predictions of the machine learning models.

3.1 Data Understanding

After the business understanding phase, we started by collecting the data at its sources where we gathered daily data of COVID-19 and NPIs implemented considering their degree of restriction throughout the pandemic from March 2020 to May 2021. Data on cultural dimensions were also considered to understand the cultural differences of the countries that were chosen for this analysis.

The data related to COVID-19 was obtained from the Our World in Data COVID-19 dataset[1]. More specifically, we collected new cases per million, new deaths per million, reproduction rate, ICU patients per million and hospitalized patients per million.

NPIs data for COVID-19 was obtained using the Oxford COVID-19 Government Response Tracker (OxCGRT)[2]. From this dataset NPI related information of COVID-19 was collected. The NPIs’ policy values are ordinal measures which associated values are integers between 0 and 5, which represent the restriction degree of the given measure. Measures from three different areas were analysed:

- **Containment and Closure policies (C)** which include school closing, workplace closing, cancel public events, restrictions on gatherings, close public transport, stay at home requirements, restrictions on internal movement and international travel controls.

[1] https://ourworldindata.org/coronavirus-source-data
[2] https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/interpretation_guide.md

- **Economic policies (E)** that contain income support and debt / contract relief.
- **Health System policies (H)** with public information campaigns, testing policy, contact tracing, facial coverings, vaccination policy and protection of elderly people.

The cultural characteristics' data were obtained from the Hofstede Insights[3] website which includes a score ranging from 0 to 100 for each of the six cultural dimensions: power distance, individualism, masculinity, uncertainty avoidance, long term Orientation and indulgency.

3.2 Data Preparation

After extracting and loading the data, we proceed to its preparation by cleaning its null values and removing variables which are not relevant to this study. The tools chosen to perform these activities were Python[4] as the programming language being used and the Jupyter Notebook[5] platform.

In the Our World in Data dataset (Table 2) we removed some features such as the ISO code because the name of the country is the variable that identifies the COVID-19 data origin and is the feature used to merge the datasets. We decided to remove features that represented total, rounded and weekly values in order to allow for cross-country comparisons. We also removed female smokers, male smokers, extreme poverty, cardiovascular death rate and diabetes prevalence features since these do not exist in most of the countries datasets. We also decided to rename the feature location to country in order to standardize this feature for later merging the datasets. Additionally, there were R_t negative values which should be positive so we replaced with their absolute value. Regarding the missing values, we decided to remove the rows that had no value associated such as population, population density, median age, people aged 65 or older, GDP per capita, hospital beds per thousand, life expectancy, human development index and stringency index. In relation to the COVID-19 variables - new cases, new deaths, ICU and hospitalized patients - the initial values are null because we assumed there was no data for each country yet, so we replaced them with 0 until there was data, whereas after the first appearance of a value the next null values that appear are replaced by the average of the values for that country.

[3] https://www.hofstede-insights.com/country-comparison/
[4] https://www.python.org/
[5] https://jupyter.org/

Additionally, we have removed the global scale values from the dataset since we want to do an analysis per country.

Table 2: Cleaning Our World in Data Dataset

Cleaning Our World in Data Dataset	Features
Removed features	iso_code, tests_units, total_cases, new_cases_smoothed, total_deaths, new_deaths_smoothed, new_cases_smoothed_per_million, new_cases_smoothed_per_million, icu_patients, hosp_patients, weekly_icu_admissions, weekly_icu_admissions_per_million, weekly_hosp_admissions, weekly_hosp_admissions_per_million, new_tests, total_tests, new_tests_smoothed, new_tests_smoothed_per_thousand, tests_per_case, total_vaccinations, people_vaccinated, people_fully_vaccinated, new_vaccinations, new_vaccinations_smoothed, new_vaccinations_smoothed_per_million, total_tests_per_thousand, new_tests_per_thousand, new_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, new_tests_per_thousand, new_cases, new_deaths, positive_rate, tests_per_case, aged_70_older, female_smokers, male_smokers, extreme_poverty, cardiovasc_death_rate, diabetes_prevalence
Renamed features	location → country
Negative values	Replaced negative values with abs()
Missing values	Replaced initial missing values with 0 and after the first non-null value replace them with the mean of those values: • new_cases_per_million • new_deaths_per_million • reproduction_rate • icu_patients_per_million • hosp_patients_per_million Removed countries where socioeconomic and demographical features are missing: • population • population_density • median_age • aged_65_older • gdp_per_capita • life_expectancy • human_development_index

In the Oxford dataset (Table 3) we removed features that were of no interest for the analysis such as the country code, flag and initial note. The "CountryName" feature was renamed to country for the purpose of later merging the datasets. In this dataset there were missing values in the "end date" and in the "policy value". In this sense, after understanding the data, the lines that had the NPI end date ("EndDate") set to null refer to NPIs that were implemented and are still in force and, therefore, these missing values were replaced by the current date of the dataset extraction. Regarding the "policy values" with no value we decided to remove them as they would not be useful data for our analysis.

Table 3: Cleaning Oxford Dataset

Cleaning Oxford Dataset	Features
Removed features	CountryCode, Flag, InitialNote
Renamed features	CountryName → country
Missing values	Removing PolicyValue missing values EndDate filled with the current date

In the Hofstede Insights dataset (Table 4) we removed some features with metadata related information such as the "adjective" and the "slug" as well as features with useless information to our work such as the "id" and "ivr". We also decided to rename the "title" to "country" and the cultural dimensions by their long name instead of their short name.

Table 4: Cleaning Culture Dataset

Cleaning Culture Dataset	Features
Removed features	id, slug, adjective, ivr
Renamed features	title : country, pdi : power_distance, idv : individualism, mas : masculinity, uai : uncertainty_avoidance, lto : long_term_orientation, ind : indulgence
Removed rows with missing values	Qatar and Tunisia

After cleaning and combining the data we obtained a dataset of 101 countries with data between March 2020 and May 2021. This dataset is composed of variables related to COVID-19 and the NPIs in which each row represents the variable’s value per day. The variables related to the socioeconomic and demographic characteristics, and also Hofstede’s cultural dimensions in which each row represents the cultural values for that specific country. We can observe these variables in Table 5.

Table 5: The characterization of the datasets with their final features

Features Type	Features
COVID-19	New cases per million Deaths cases per million Reproduction rate ICU patients per million Hospitalized patients per million
Socioeconomic and demographic	Population Population density Median age Aged 65 older GDP per capita Life expectancy Human development index
NPIs	Stringency index Testing policy School closing International travel controls Public information campaigns Protection of elderly people Contact tracing Workplace closing Stay at home requirements Cancel public events Restrictions on gatherings Restrictions on internal movement Debt/contract relief Income support Facial Coverings Vaccination policy Close public transport
Hofstede's cultural dimensions	Power distance Individualism Masculinity Uncertainty avoidance Long term orientation Indulgence

3.3 Exploratory Data Analysis

In a first stage we made an exploratory data analysis using correlation matrixes and line graphs where the main goals were to understand the relationship between restriction levels of the NPIs and COVID-19 evolution and compare the NPIs between countries as well as understand what differences may explain the evolution of the R_t for each of them.

The correlation matrix allows observing the correlations between the variables. The closer to 1 or -1 the stronger the correlation is, and if the correlation is positive, this means that the two variables grow in the same direction, that is, while one variable increases the other also increases, whereas, a negative correlation means the opposite (while one variable increases, the other decreases). It is important to highlight that the values portrayed on the correlation matrix do not implicate causality between the variables.

Pearson's correlation is used, which evaluates the linear relation between two continuous variables and we consider associations from moderate to very strong correlation (Table 6).

Table 6: Correlation table according to (Schober & Schwarte, 2018)

Correlation	Positive	Negative
Negligible	0.00 to 0.10	0.00 to -0.10
Weak	0.10 to 0.39	-0.10 to -0.39
Moderate	0.40 to 0.69	-0.40 to -0.69
Strong	0.70 to 0.89	-0.70 to -0.89
Very Strong	0.90 to 1.00	-0.90 to -1.00

In this sense, we considered five countries for the analysis, India, Brazil, the UK, Israel, and Portugal. The former two are among the most populous countries in the world, with high-density cities such as Mumbai and São Paulo concentrating over 10M people, which facilitates airborne infectious disease dissemination.

Furthermore, India and Brazil belong to the so-called BRICS emerging economies (Stuenkel, 2020), with both presenting significant inequality society layers in education and purchasing power (Alvarez Candido et al., 2021; Kumar et al., 2019), which affects access to healthcare services and increases the pressure over lower income layers to work, despite imposed labour restrictions during the COVID-19 pandemic. The UK is a developed populous country in Europe which, like India and Brazil, was heavily affected by the COVID-19 pandemic. The UK is an interesting case study since it was the first to approve COVID-19 vaccines and start administering them, being currently (as of May 2021) the second country with the largest percentage of its population vaccinated (OurWorldInData, 2021). In comparison, Israel is a country with a smaller area and less population, but since the beginning it has implemented several NPIs with a high degree of restrictions. Israel is also the country with the largest vaccinated percentage of its population, as of May 2021 (OurWorldInData, 2021). Finally, Portugal represents an European country that has a similar population to Israel and that at the beginning of the pandemic was highly praised in the first containment because after the first cases of SARS-CoV-2 infection appeared on March 18th, 2020 the government declared the first containment (Costa et al., 2020).

Figure 3 portrays Hofstede's six cultural dimensions (Hofstede Insigths, 2021) for each country in analysis. Through Hofstede's cultural dimensions Brazil is characterized as a collective society (INV=38) with a high power distance, meaning an unequal distribution of power (PD=69), with a need to establish many rules to live in society (UAI=76) and that has a positive and optimistic attitude towards life (ING=59). While India is also a society that is divided by hierarchies (PD=77), it is driven by competition, achievement and professional success (MAS=56), as its people is used to deal with ambiguous or unknown situations (UAI=40) and present a pessimistic and restrictive attitude as the majority of the population do not, or cannot, put much emphasis on leisure time (ING=26).

In contrast, Israel, in which the power is considered to be more equally distributed, and its people believe in equal rights (PD=13), they have a strong need to create rules and devise legal systems in order to structure life in society (UAI=81), showing a preference for normative thinking and exhibiting great respect for traditions (LTO=38). In the indulgence dimension, Israel has no score attributed, which represents a limitation for the analysis of this dimension for this country.

In the UK, there is a moderate power distance, meaning a moderate distribution of power (PD=35). The UK's population is highly individualistic (INV=89), success-oriented in their work (MAS=66), and also comfortable dealing with ambiguous situations (UAI=35), tending to be optimistic and giving great importance to leisure time (ING=69).

Finally, Portugal has a collective society (IND=27) where the hierarchical distance is accepted and individuals closer to the top have more privileges (PDI=63). Like most European countries, the society fosters strong relationships where everyone takes responsibility for their group members and excessive competitiveness is not appreciated (MAS=31). The Portuguese people is more restraint (ING=33), avoiding uncertainty (UAI=99) and preferring normative over pragmatic thinking (LTO=28).

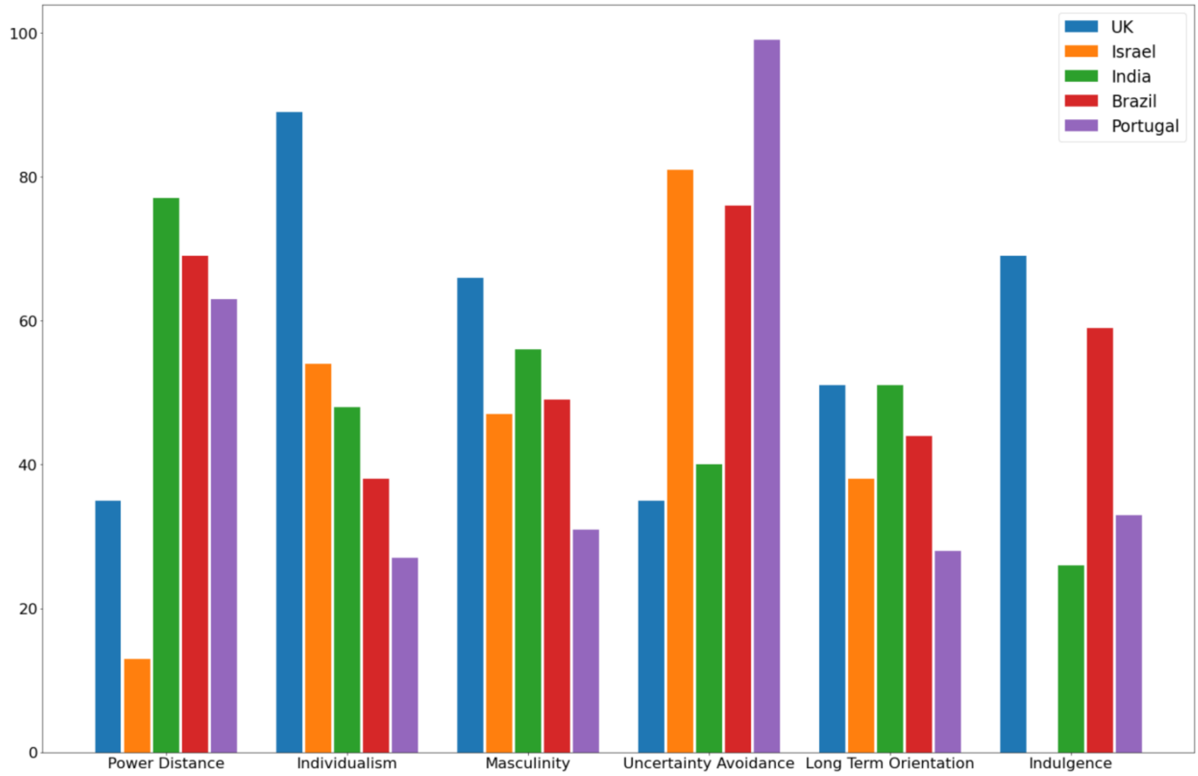

Figure 3: Hofstede's Cultural Dimensions

In a second stage we used the data from all the countries of our dataset which is composed of 101 countries to train several machine learning models so that we could verify the impact of cultural dimensions in the prediction of the reproduction rate. Thus, the 101 countries considered for the models are: Albania, Algeria, Argentina, Australia, Austria, Azerbaijan, Bangladesh, Belarus, Belgium, Bhutan, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, Burkina Faso, Canada, Cape Verde, Chile, China, Colombia, Costa Rica, Croatia, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kuwait, Latvia, Lebanon, Libya, Lithuania, Luxembourg, Malawi, Malaysia, Malta, Mexico, Moldova, Morocco, Mozambique, Nepal, Netherlands, New Zealand, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Saudi Arabia, Serbia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sri Lanka, Suriname, Sweden, Switzerland, Tanzania, Thailand, Trinidad and Tobago, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, Vietnam and Zambia.

3.4 Modelling

In a first phase analysis we only considered socioeconomic and demographic characteristics of each country as well as the degree of restriction of the NPIs implemented in the 101 countries for predicting the R_t, over which we run a set of machine learning models. In a second phase, we ran the same machine learning models, adding Hofstede's cultural dimensions to the independent variables.

Through these models we aim to assess whether there is an improvement of the models with the addition of Hofstede's cultural characteristics in predicting the R_t.

For data normalization we used the StandardScaler procedure which is one of the most used techniques in machine learning in which the features are normally distributed with a mean value close to 0 and a standard deviation of 1 (Raju et al., 2020). After normalizing the data, we selected the features for the machine learning models from the dataset we obtained after the data preparation phase. Thus, we tested two data selection methods: correlation and mutual information (MI).

Figure 4 shows the correlations of all independent variables with the R_t. We can observe that one of the most relevant correlations with the reproduction rate we obtained was a weak negative correlation with the facial coverings NPI (-0.16). We also obtained other correlations close to 0.10 and -0.10 such as, human development index (0.09), vaccination policy (-0.09), new deaths per million (-0.08), median age (0.08), life expectancy (0.08) and uncertainty avoidance (0.07).

Although individually each of these variables did not obtain a relevant linear correlation with the target R_t, this does not mean that all these variables are not important to train the predictive models. For this reason, it was decided to explore nonlinear correlations using MI.

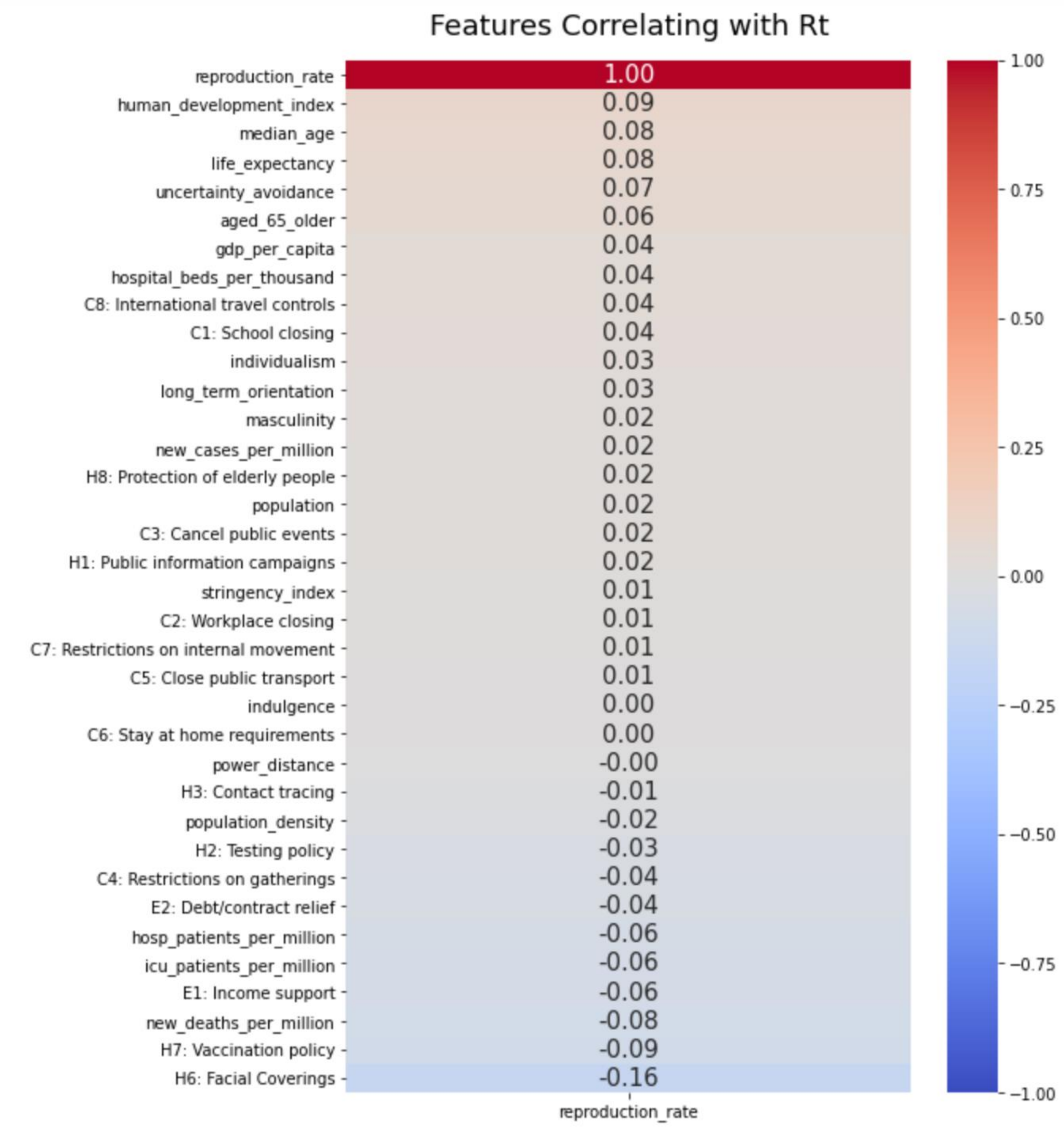

Figure 4: Features correlating with R_t

MI is used as a criterion to select the best input variables (from a set of possible variables) for the long-term forecasting objective, in other words, between two variables, say X and Y, is the amount of information obtained from X in the presence of Y, and vice versa (Sorjamaa et al., 2005).

This dependency measure can capture linear as well as nonlinear dependencies but does not capture multivariate dependencies or indicate collinearity between features (Yeung et al., 2021). Accordingly, we use the Python package ennemi[6] based on MI (Laarne et al., 2021). In Table 7 we can observe that all features obtained a mutual information score above 0.10. The features that obtained the highest correlation were the socioeconomic and demographic ones such as population size, GDP per capita, portion of the elderly population (older than or equal to 65 years), population density, average life expectancy, human development index and average age with values ranging from 0.25 to 0.29. Then the features that obtained the highest correlation were the cultural dimensions of long-term orientation, indulgence, power distance, individualism, masculinity, and uncertainty avoidance with values between 0.18 and 0.21. Finally, NPIs such as close public transport, facial coverings, income support and the restriction index obtained correlations between 0.10 and 0.27.

Through the analysis of this graphic, we can highlight that in addition to the socioeconomic and demographic dimensions, Hofstede's cultural dimensions are relevant to help predict the R_t value. Regarding the NPIs, these values can be explained due to the volatile nature of their values, as there are countries that rarely change them (e.g., Brazil) as well as countries that do it regularly (e.g., Portugal).

Yeung et al., (2021) state that all features are important to be considered in the machine learning models because features are selected if they achieve mutual information greater than 0.10. Therefore, all NPIs, socioeconomic and demographical features and Hofstede's cultural dimensions were considered.

All variables present in Table 7 were considered as input for several linear, non-linear and ensemble models. The models used were imported from the sklearn[7] library.

[6] https://github.com/polsys/ennemi
[7] https://scikit-learn.org/stable/

Table 7: Mutual Information score

Feature	Score
Stringency index	0.27
Population	0.29
Population density	0.28
Median age	0.26
Aged 65 older	0.29
GDP per capita	0.29
Hospital beds per thousand	0.25
Life expectancy	0.28
Human development index	0.27
H2: Testing policy	0.12
C1: School closing	0.11
C8: International travel controls	0.12
H1: Public information campaigns	0.17
H8: Protection of elderly people	0.10
H3: Contact tracing	0.13
C2: Workplace closing	0.13
C6: Stay at home requirements	0.11
C3: Cancel public events	0.14
C4: Restrictions on gatherings	0.11
C7: Restrictions on internal movement	0.10
E2: Debt/contract relief	0.11
E1: Income support	0.12
H6: Facial Coverings	0.15
H7: Vaccination policy	0.15
C5: Close public transport	0.11
Power distance	0.18
Individualism	0.19
Masculinity	0.18
Uncertainty avoidance	0.18
Long term orientation	0.21
Indulgence	0.21

3.5 Evaluation

The validation technique used was the hold-out technique as it is one of the most popular and faster validation techniques (Moro, 2015). According to Han et al., (2011) the hold-out technique consists of randomly dividing the dataset into two independent sets, a training set and a test set. Usually, two-thirds of the data are assigned to the training set, and the remaining third is assigned to the test set (Han et al., 2011). Thus, we randomly assigned two thirds of the dataset for the training set and one third for the test set.

According to Steurer et al., (2021), in order to evaluate the applied models, we apply three error metrics:

- The **Mean Absolute Error or MAE** measures the average of the sum of absolute differences between observation values and predicted values.
- The **Mean Squared Error or MSE** measures the average squared difference between the estimated and true values.
- The **Root Mean Squared Error or RMSE** is a monotonic transformation of the MSE that corresponds to its root.

Chapter 4 Results

This chapter is organized into the following sections:

- **Section 4.1** includes the presentation of heatmaps and line graphs made for each of the five countries under analysis: India, Brazil, the UK, Israel and Portugal;
- **Section 4.2** contains the results of the various machine learning models trained on the 101 countries with the aim of predicting the evolution of the R_t, i.e., the pandemic propagation.

4.1 Heatmap and Graph analysis for the five countries

We analysed five countries individually to understand the NPIs that most influenced the evolution of the pandemic over time, as well as to observe differences in the degree of the restriction of the NPIs implemented for each of the selected countries. Such approach enables to discuss each country's success or failure in controlling the pandemic at different key moments and in light of each's context, including cultural, political and economic dimensions.

During the initial period of the pandemic, there were studies that identified the importance of some NPIs, such as the restrictions of social distancing (Haug et al., 2020), travel bans (Haug et al., 2020; Koh et al., 2020) and mandatory face masks (Chernozhukov et al., 2021).

Our study covers a more recent period (mid-March 2020 until May 2nd 2021) when countries have already begun their vaccination plan, with the UK and Israel leading the way, we also plot visually appealing graphics which are easily interpreted to discuss the effects of the NPIs that influence most of the pandemic propagation (the R_t factor) in different countries' context. It is important to note that the effectiveness of the NPIs' imposition is arduous without serious community participation (Douglas et al., 2020).

While we are focused on the reproduction rate and thus, the R_t factor becomes the most important dependent variable for identifying the influence of NPI measures, we also considered the number of new daily cases per million and the number of new daily deaths per million, as these three are metrics provided in the original dataset which enable to also understand the direct impact of the NPIs. To assess the correlation between each NPI and the above-mentioned variables, we plot a correlation heatmap and also introduce the evolution of the R_t in easy-to-read graphics.

4.1.1 India

According to Douglas et al., (2020) and Ray et al., (2020), which report to the period until August 2020, suppression measures in India can significantly reduce the impact of the COVID-19 pandemic. We further analyse our results in detail here after.

For India, the analysed data reports the daily situation from March 15th 2020 to May 2nd 2021.

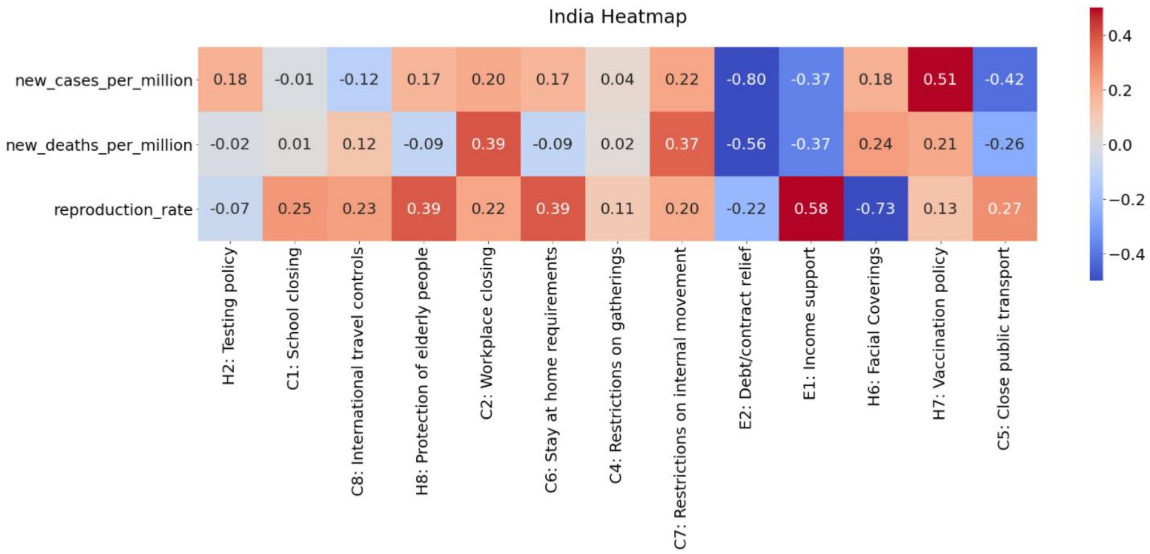

Figure 5: India NPIs Heatmap

From the heatmap (Figure 5) we can see that the facial coverings (H6) NPI has a strong negative correlation with the reproduction rate (-0.73), which implies that, as the mandatory use of masks became more restrictive, the R_t value decreased. However, the income support (E1) NPI has a moderate positive correlation (0.58).

The latter result hints that government support to society in general (families and businesses) might have an opposite effect to expected, with the R_t increasing as income support increases, which would deserve a deeper analysis. Finally, in India there are no visible results of the vaccination policy on the reduction of the R_t, and it has a moderate positive association (0.51) with new cases per million, which may indicate that the increase in vaccination coincided with the period when there were more new daily cases of infection. As for the remaining dependent variables, we can see that the application of the debt / contract relief (E2) NPI coincides with the reduction of the new cases and deaths of COVID-19. While the increase of the restriction of the close public transport (C5) NPI seemed to decrease the number of new cases of COVID-19. These results lead us to believe that in such a heavily populated country, where the population faces some economic difficulties, the application of these NPIs can be effective in controlling contagions.

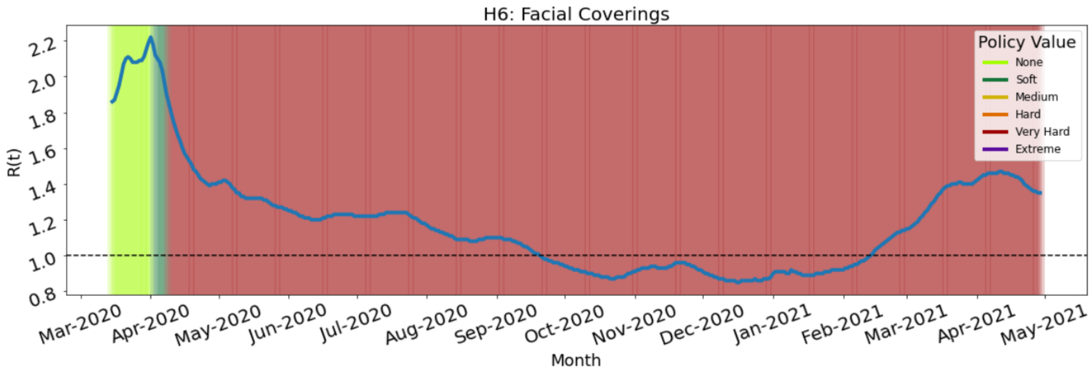

Figure 6: Facial Coverings NPI in India

In Figure 6 we can see that the implementation of the NPI facial coverings (H6), resulted in R_t plummet, which stayed below 1 until mid-February 2021 when it started to slightly rise again. We think that such increase can be explained by the withdrawal / relief of the degree of restriction of the NPI close public transports (C5) and restrictions on internal movement (C7), which was in demand by a society desperate for activities reopening (Yu et al., 2021) as the little income support (E1) NPI became none in July 2020 (Figure 7). These results show that while facial coverings by itself is a really effective measure in a large and populous country, it is not enough in the medium / long term if not accompanied by other measures that restrict mobility.

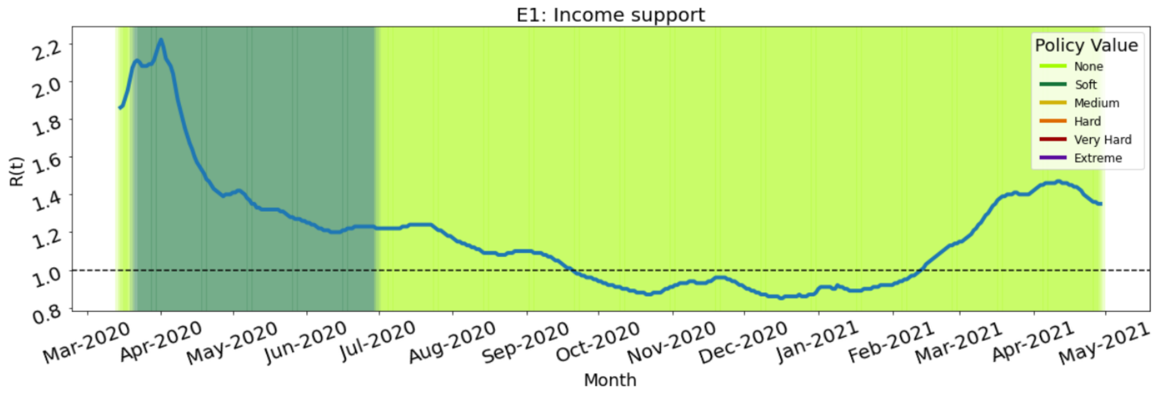

Figure 7: Income Support NPI in India

Unlike the facial coverings NPI (H6), the income support NPI (E1) (Figure 7) obtained a moderate association with the increase in R_t.

Apparently, the application of this policy was soft from the beginning of its implementation until the end of June 2020, when the soft measures ceased entirely, which coincided with the decrease of the R_t. However, during the same period of time other NPIs, such as protection on elderly people and stay at home requirements, were applied with a greater degree of restriction, which influenced the R_t decrease after July 2020. The moderate positive correlation observed in Figure 5 is the result of the steep increase of the R_t after February 2021 when no income support NPI was in place, which remained steady and above the values of May-June 2020, when soft income support was implemented. We can hypothesize that India's little to no efforts in implementing income support has led families and businesses to take advantage of the relaxation of mobility NPIs to resume pre-COVID-19 regular activities, which increased contacts, as reflected by the R_t increase. However, this is a tough call by Indian authorities, in a 1.3B population carved by significant income inequalities, which makes it very difficult to implement every citizen's income support policy. The balance between mobility resume and contagion propagation control will likely continue a subject of debate until we can bring the pandemic under control at a worldwide scale (Han et al., 2020).

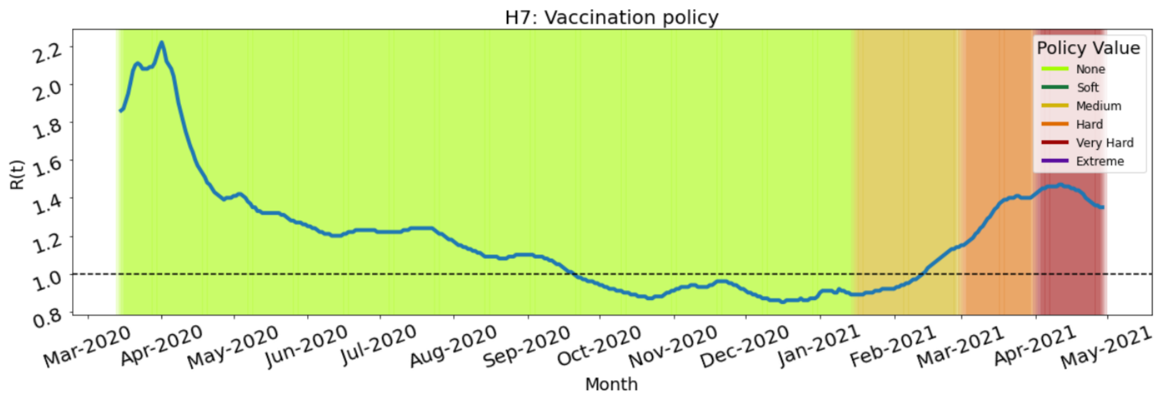

Figure 8: Vaccination Policy NPI in India

Although the vaccination policy (H7) does not show a moderate / strong association with R_t, we can observe (Figure 8) that this NPI started in mid-January and, over time, its implementation has been increasing and it is beginning to observe its effect on the R_t, which has started to decrease in mid-April 2021. While this is still a preliminary result when the vast majority of the population has not yet been vaccinated, it should be observed with caution, as the relaxation of other measures might imply that the R_t can still increase, with drastic effects in a country already severely affected by COVID-19.

4.1.2 Brazil

In pandemic times Brazil becomes an interesting country to analyse because it has a different population in each region in terms of social behaviour, genetics and economic origins, which increases the need for different medical and social management in each area (Marson & Ortega, 2020). We analysed data from daily reports for the period from March 14th 2020 to May 2nd 2021.

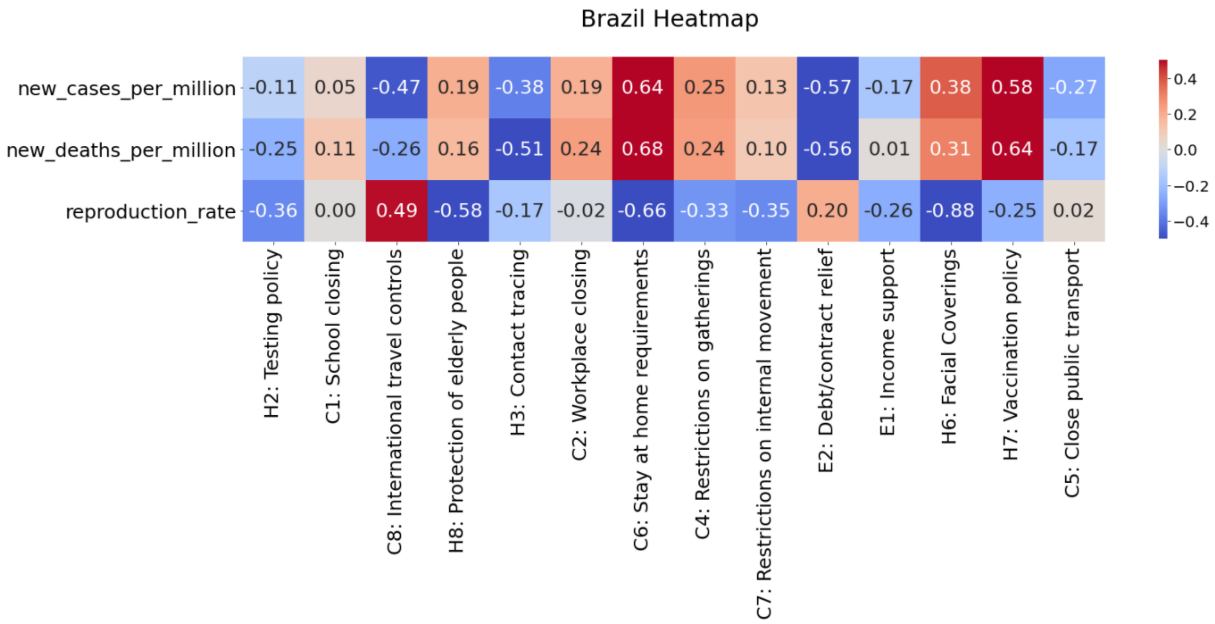

Figure 9: Brazil NPIs Heatmap

Through the observation of Brazil's NPIs heatmap (Figure 9), as the protection of elderly people (H8), stay at home requirements (C6) and facial coverings (H6) NPIs became more restrictive over time, the R_t decreased. On the opposite, NPI international travel controls (C8) showed a moderate positive association with R_t. The latter finding leads us to hypothesize that, when international travel restrictions were imposed, the epidemic was already well disseminated in Brazil and the contagion was local / regional, rather than imported from abroad. Such rationale is supported by Laroze et al., (2021), who argue that the coronavirus continued to be disseminated among local communities in UK districts, despite regional lockdown.

As for the remaining COVID-19 variables we analysed, there were associations that stood out. The debt / contract relief (E2) is negatively correlated with the new cases (-0.57) and deaths per million (0.56). Nevertheless, unexpectedly, stay at home requirements (C6) and vaccination policy (H7) NPIs have a positive correlation (Figure 9). One reason for this result might be that during the first months of the pandemic, physical distancing measures were implemented (Rafaell et al., 2020) such as, the NPI stay at home requirements in which corresponded to the period when the number of new cases and deaths per million reached very high values.

Contact tracing (H3) NPI has a negative correlation with new deaths per million. Previous studies pointed out that tracing contacts through technologies such as smartphones and other mobile devices are ineffective due to the lack of adoption by the wider population and to inaccurate tracing technology (Hernández-Orallo et al., 2020). At the beginning of the pandemic, world experiences pointed to the need to control the speed of the curve's progression through measures of social physical isolation (Rafaell et al., 2020). When the pandemic was first disseminated in Brazil, some NPIs were established, such as: restrictions on internal movement (C7), restrictions on gatherings (C4) and stay at home requirements (C6) that portray a rising degree of restriction.

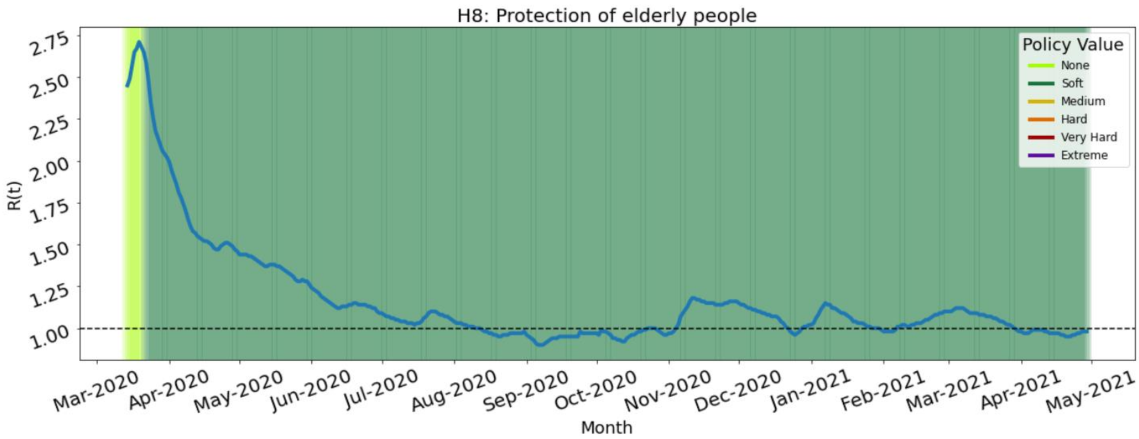

Figure 10: Protection of elderly people NPI in Brazil

In Figure 10 we can see that the protection of elderly people (H8) NPI since mid-March 2020 has not changed in its degree of restriction, which has remained soft. Although the heatmap shows a moderate negative correlation between this NPI and the R_t (Figure 9), this is due to the sharp reduction in the R_t when the NPI is implemented, and it does not change during the following months. It should be noted that during the time period in which the NPI is implemented several other NPIs have been implemented, for example, close public transport (C5), restrictions on gatherings (C4) and workplace closing (C2).

The weak association between economic NPIs and R_t may be justified due to the weak implementation of economic support NPIs during these months of the pandemic. The social, political and economic challenges that the country has been facing due to the withdrawal of labour rights and guarantees, and the increasing loss of the purchasing power of families in recent years are strong factors that contribute to the vulnerability of the poorest populations for coping with the pandemic (Rafaell et al., 2020).

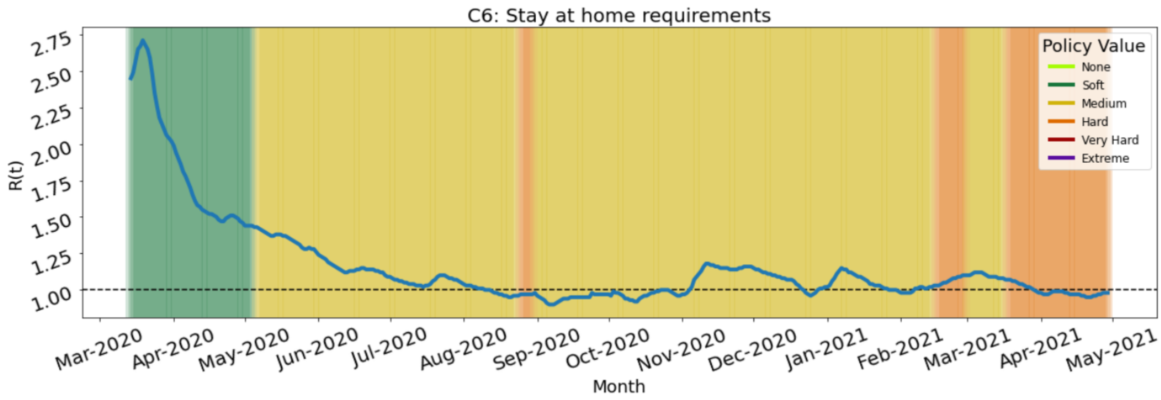

Figure 11: Stay at Home Requirements NPI in Brazil

The stay at home requirements (C6) NPI appears to have been an effective measure throughout these months of the pandemic which varied between different degrees of restriction (Figure 11). According to Rafaell et al., (2020) at the beginning of the pandemic in Brazil, social isolation measures such as the NPI stay at home requirements were imposed, which may have contributed to the reduction of R_t, which is in line with our results.

Among all NPIs for containment and closure, this was the one that seemed to have the greatest impact in reducing R_t, which is aligned with previous studies findings e.g., Koh et al., (2020). This type of NPIs tends to reduce the need for ventilators and hospitalization in intensive care units in a short period of time, adapting the need of the health system's assistance capacity (Rafaell et al., 2020). According to Marson & Ortega (2020) there are no official data on the number of beds in the ICU beyond which the same number is divided between public and private institutions which is in accordance with the data sources used in our work which do not contain this information.

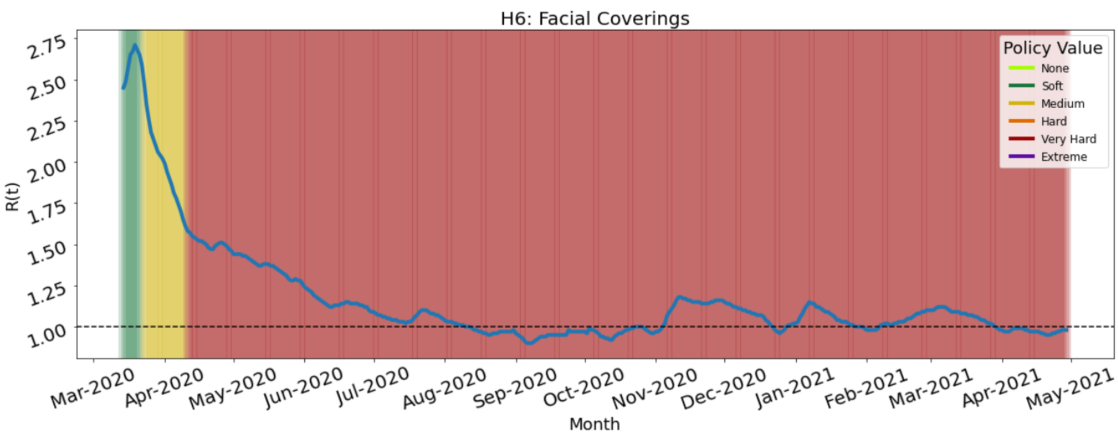

Figure 12: Facial Coverings NPI in Brazil

In Figure 12 we were able to observe the decrease of R_t after a more restrictive implementation of the facial coverings (H6) NPI which seems to confirm that this NPI is very important to control the transmissibility of the virus (van der Westhuizen et al., 2020).

The resistance and the denial of the seriousness of the pandemic by the president of the Brazilian republic may have had consequences in terms of the population's understanding of what is the guideline to be adopted, in failures in social isolation and damage in the implemented sanitary barriers (Rafaell et al., 2020) which may justify the fact that the NPIs restrictions on gatherings (C4) and restrictions on internal movements (C7) did not have a moderate / strong association with the R_t decrease, despite high restrictions having been implemented over the months of the pandemic.

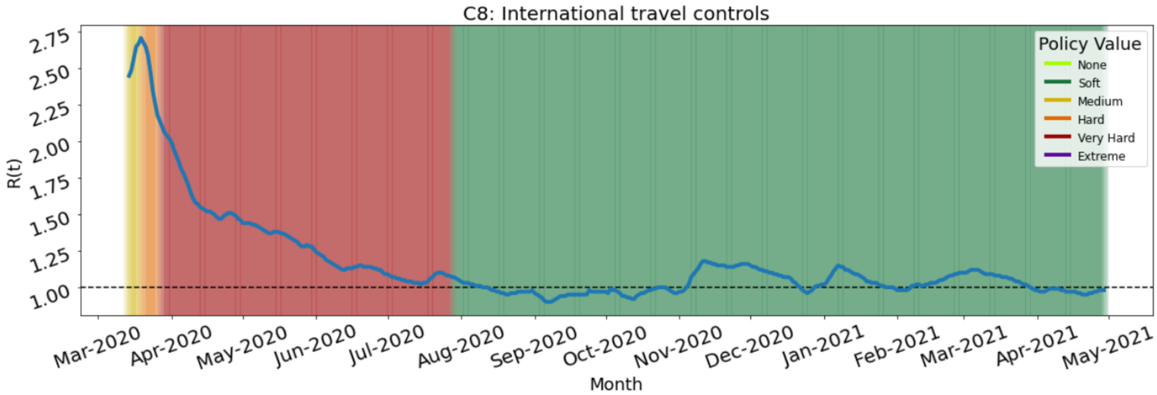

Figure 13: International Travel Controls NPI in Brazil

The implementation of the international travel controls (C8) NPI (Figure 13) shows an increase in the R_t, which may indicate that this measure was implemented when the value of the R_t was on the rise when the pandemic started.

The results by González-Bustamante (2021) show that Brazil has just implemented the international travel controls NPI instead of implementing key integrated suppression policies or clear infection detection and tracking strategies to monitor the virus dissemination throughout the community. However, in our analysis during the first three months of the pandemic, Brazil implemented suppression NPIs such as restrictions on gatherings (C4), stay at home requirements (C6) and close public transport (C5) as well as NPIs linked to the detection and tracking of infections such as contact tracing (H3) and testing policy (H2), which may have had a role in avoiding further dissemination during the first wave.

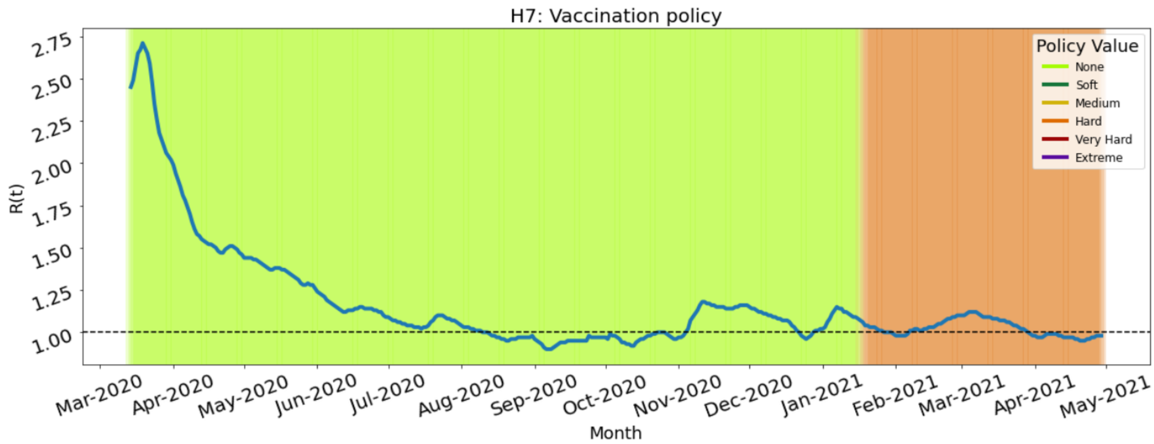

Figure 14: Vaccination Policy NPI in Brazil

As in India, the vaccination policy (H7) (Figure 14) in Brazil started in January 2021 and seems to be having positive results because since April 2021 the R_t has not exceeded the value 1.

4.1.3 United Kingdom

In the United Kingdom, at the beginning of the pandemic, the implementation of NPIs was assessed as the most viable option to control the outbreak of COVID-19 (Neil et al., 2020). Next, we will analyse UK daily data of COVID-19 reports from March 3rd 2020 to May 2nd 2021.

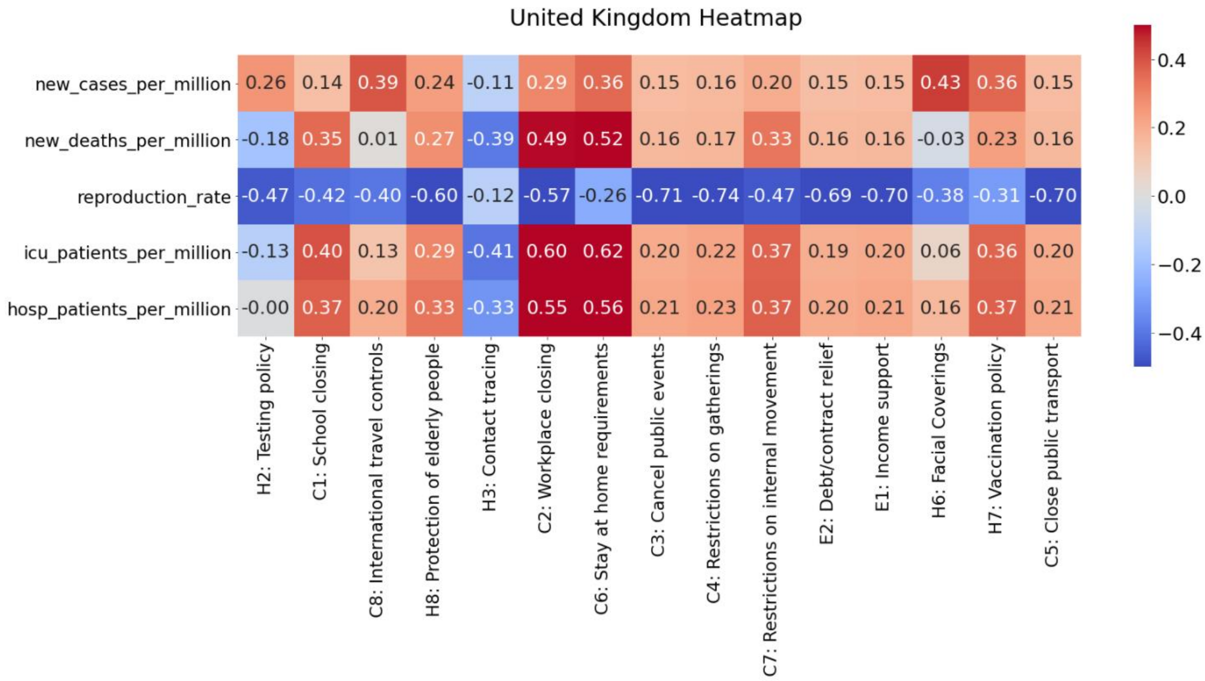

Figure 15: UK NPIs Heatmap

By analysing the UK's NPIs heatmap (Figure 15) we can identify several strong associations of the NPIs cancel public events (C3), restrictions on gatherings (C4), income support (E1) and close public transport (C5) with the reduction of the R_t, as these NPIs have correlation values ranging from -0.74 to -0.70 with the COVID-19 reproduction rate. Also, there are also moderate associations between -0.40 to -0.69 that stand out such as, NPIs testing policy (H2), school closing (C1), international travel controls (C8), protection of elderly people (H8), workplace closing (C2), restriction on internal movements (C7) and debt / contract relief (E2).

In general, the NPIs had significant negative associations with the R_t which can be explained by a high R_t value at the beginning of the pandemic and after the implementation of the various NPIs coincided with a decrease in the R_t over the months.

The NPIs contact tracing (H3), stay at home requirements (C6), facial coverings (H6) and vaccination policy (H7) were the measures that obtained the least correlation with the R_t. It should be noted that the lack of association of these NPIs with the R_t may be related to the relief of restriction of other NPIs during a more restrictive level of the former.

Furthermore, it can be observed that the NPIs workplace closing (C2) and stay at home requirements (C6) have a positive correlation with the new deaths, ICU patients and hospitalized patients per million. There are subpopulations at higher risk of contracting COVID-19 and developing its more severe forms namely, elderly people (Prats-Uribe et al., 2020). According to Neil et al., (2020) the combination of case isolation, home quarantine and social distancing of people at higher risk of severe outcomes are the most effective policy combination for mitigating the epidemic.

Although NPIs cancel public events (C3), restrictions on gatherings (C4), income support (E1) and close public transport (C5) have had a greater degree of association in reducing the R_t, they have not varied in their degree of implementation since the beginning of the pandemic. In a deeper look, the NPI restrictions on gatherings (C4) was the only NPI to which a greater degree of restriction was assigned as it was ascribed a level of hard restriction. Then, the NPIs cancel public events (C3) and income support (E1) were given an average degree of implementation. Only the NPI close public transport (C5) was imposed the less severe restriction.

In contrast, the NPIs that obtained a moderate association with the decrease in R_t are those that had varying degrees of restriction implementations throughout the pandemic with exception to NPIs testing policy (H2) and debt / contract relief (E2). The NPI testing policy (H2) since the beginning of COVID-19 had only two variations, from March 2020 to mid-May 2020 it had a soft implementation and in the months that followed a medium implementation, similarly to NPI debt / contract relief (E2) which in the middle of March 2020 started to be applied with a soft degree and soon afterwards, medium.

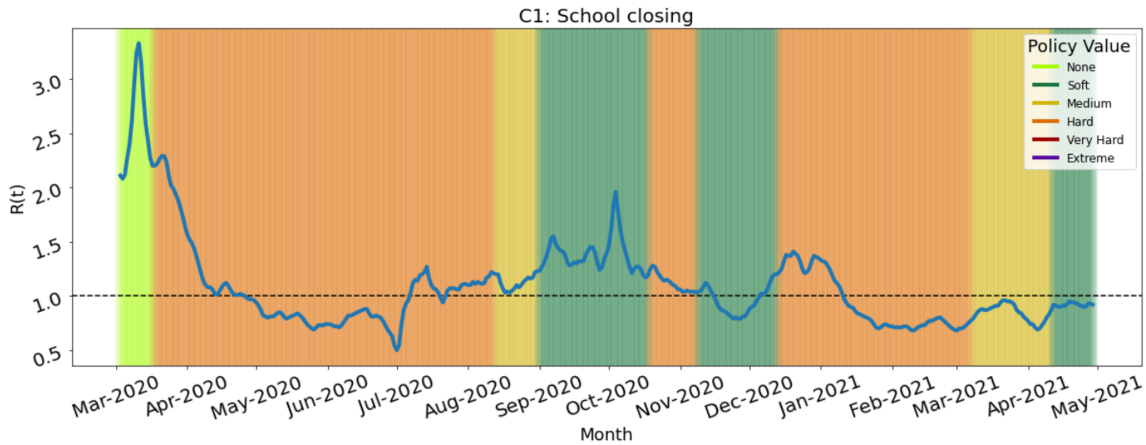

Figure 16: School Closing NPI in the UK

According to Neil et al., (2020) at the beginning of the pandemic, the implementation of NPIs related to the distancing of the entire population, home isolation from cases and home quarantine among family members should have been complemented with school closures because educational institutions can play a large role in transmission, despite the security measures implemented when the degree of NPI restriction decreases (Brauner et al., 2021).

We can see that the NPI school closing (C1) (Figure 16) had several degrees of restriction over time, when there was a relief of its degree of restriction, the R_t began to increase and a high degree of restriction was again imposed. Although NPI school closing (C1) is important, its more restrictive implementation has the disadvantage of being able to contribute to the increase in absenteeism (Neil et al., 2020).

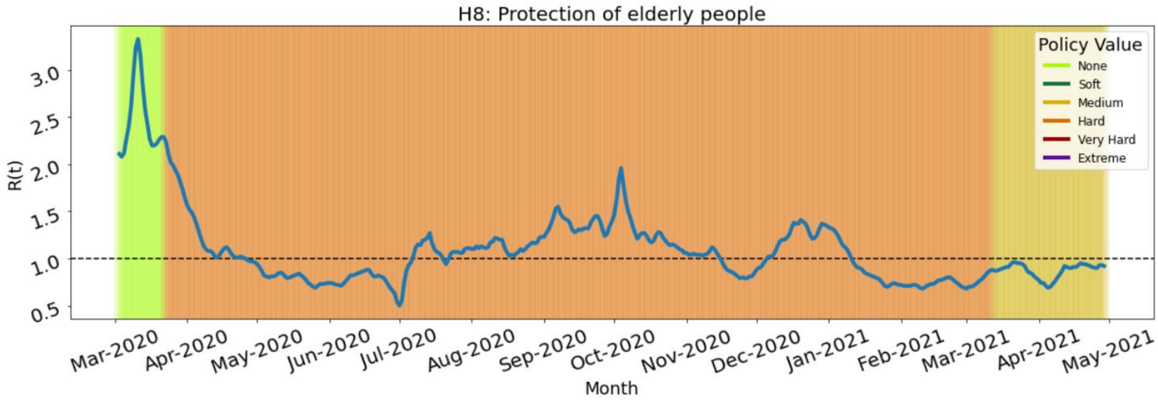

Figure 17: Protection of Elderly People NPI in the UK

The NPI protection of elderly people (H8) (Figure 17) had a hard degree of implementation from mid-March 2020 to mid-March 2021. After this time there was a decrease in the degree of restriction of this NPI to medium, which coincided with the time in which the NPI vaccination policy increased its degree of implementation to hard. This NPI seems to be important, considering that the UK has a rate of 28.52% of people over 65 (OurWorldInData, 2021).

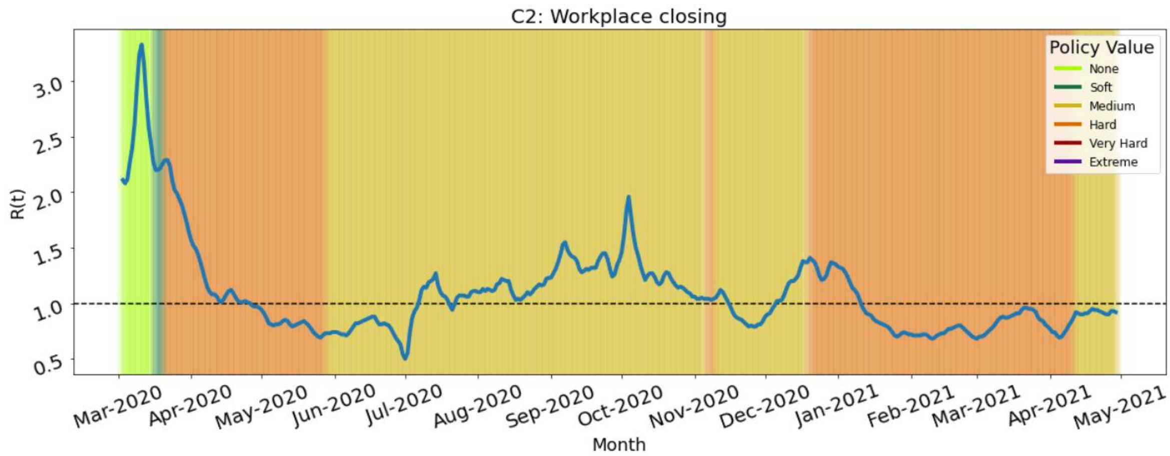

Figure 18: Workplace Closing NPI in the UK

The NPI workplace closing (C2) (Figure 18) involves a substantial number of employees that can represent a significant containment measure on the incidence of COVID-19 (Wong et al., 2020). The UK implemented the NPI workplace closing (C2) in March 2020 and as we can see, the increase in the restriction of this NPI, combined with others, has a positive impact on the evolution of the pandemic, more specifically on the value of the R_t because with some of the population working remotely the influx of public transport usage and the number of individuals performing their work duties in office closed spaces decreases.

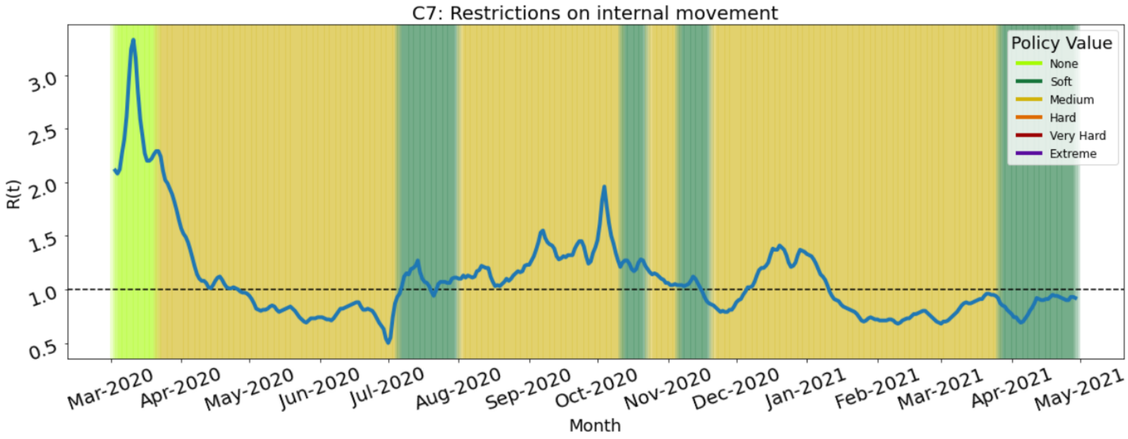

Figure 19: Restrictions on Internal Movement NPI in the UK

The NPI restrictions on internal movement (C7) (Figure 19) began to be implemented in mid-March 2020 with a medium degree of restriction. Over the months, there were some reliefs of this NPI, however, when the rise of the R_t occurred, it was reimplemented more restrictedly, with a medium policy value. It was the case in the beginning of August 2020, middle of October 2020 and middle of November 2020.

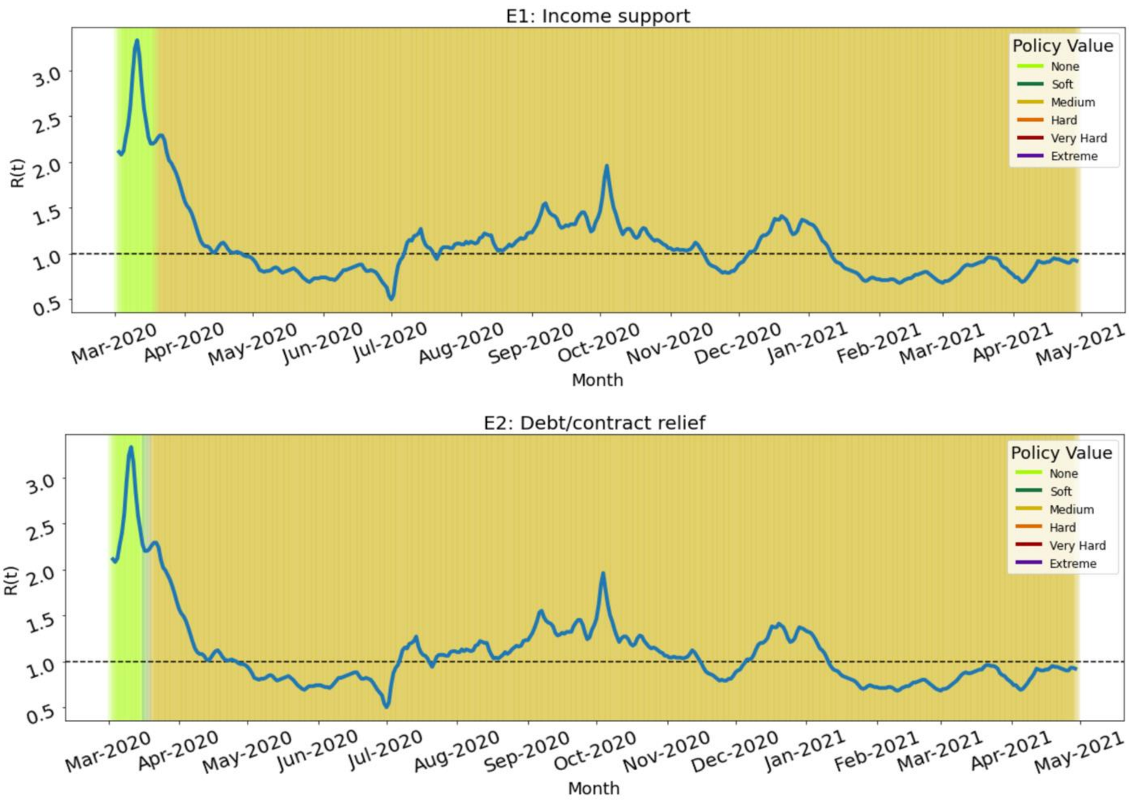

Figure 20: Economic policies NPIs in the UK

In the UK it was observed that economic NPIs (Figure 20) were associated with a decrease in R_t (E1 = -0.70 and E2 = -0.69) which were applied with a medium degree since the beginning of the pandemic. Given the fact that in the first three months of the pandemic in the UK, an association was observed between socioeconomic deprivation and the risk of infection by COVID-19 (Prats-Uribe et al., 2020), the application of economic NPIs may have played an important role in controlling the R_t for the most deprived individuals.

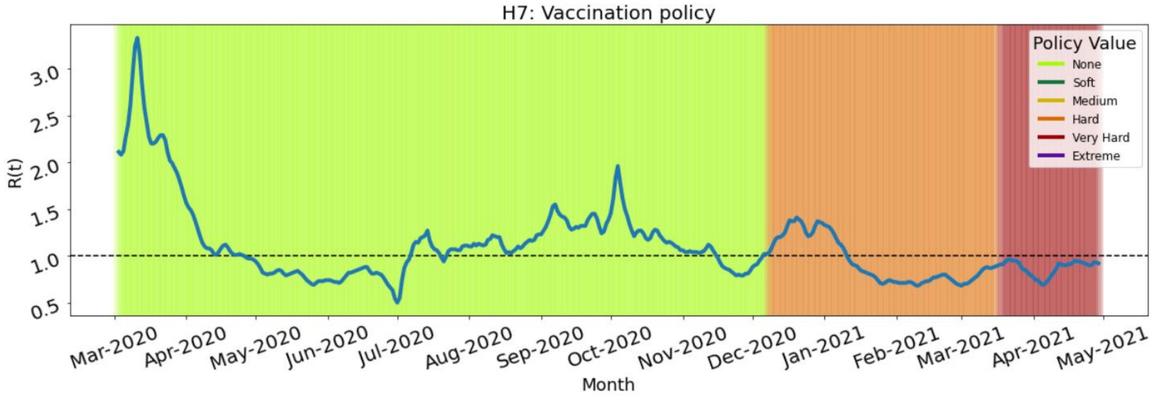

Figure 21: Vaccination Policy NPI in the UK

The UK was one of the first countries to start the vaccination campaign (OurWorldInData, 2021). In Figure 21 we can see that this was applied on a large scale increasing its degree in mid-March 2021.

The NPI vaccination policy (H7) obtained a weak association with the R_t (-0.31) (Figure 15), which may be due to the fact that it started to be implemented in mid-December, coinciding with the arrival of winter, and that its effects are not immediately visible. Although, we can highlight the vaccination policy (Figure 21), which has evolved considerably since the beginning, with the R_t below 1.

4.1.4 Israel

Israel is a multi-ethnic and multicultural country which has adopted a combination of rules of stay at home and social detachment, encouraging people to avoid leaving home and to keep a safe distance from people outside their own home (Saban et al., 2021). In this way, the data analysed for Israel refer to the period from March 14th 2020 to May 2nd 2021.

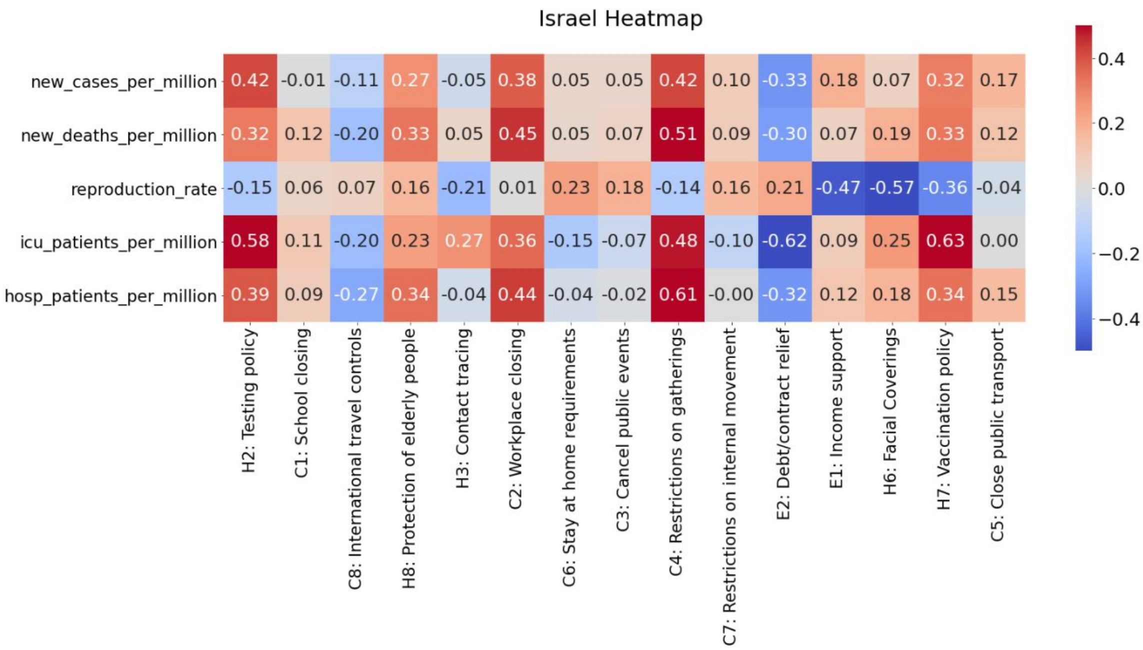

Figure 22: Israel NPIs Heatmap

In Israel's heatmap (Figure 22), we observed that the NPIs income support (E1) and facial coverings (H6) had a moderate negative correlation with the R_t.

We can also highlight the economic NPI debt / contract relief (E2), which has a negative relation with ICU patients per million. However, the restrictions on gatherings (C4) NPI increase as new deaths, ICU patients and hospitalized patients increase. ICU patients increase as NPIs testing policy (H2), cancel public events (C3) and vaccination policy (H7) increase their restriction degree. Previous studies in Israel have shown that banning meetings was effective, with a significative effect to limit meetings to 10 people or less (Brauner et al., 2021). Finally, the debt / contract relief (E2) NPI has a negative correlation with ICU patients.

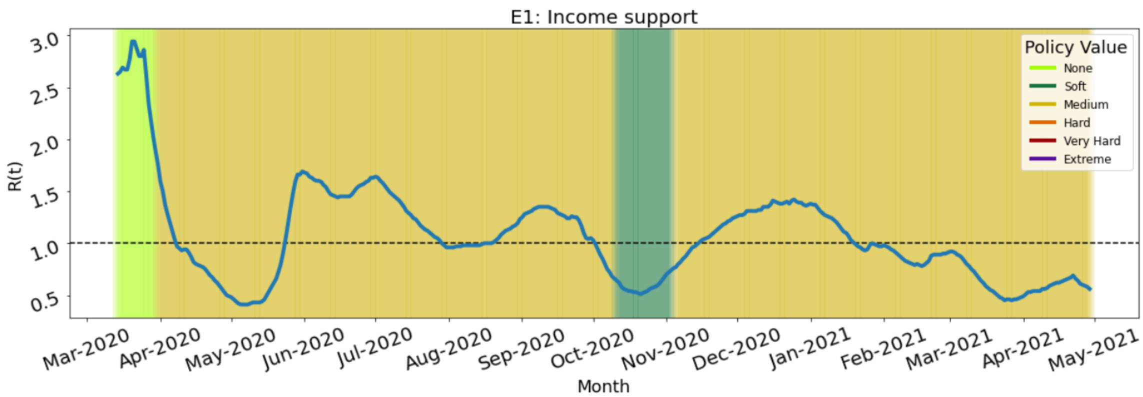

Figure 23: Income Support NPI in Israel

In Figure 23 we can see that the NPI income support (E1) was implemented throughout the pandemic consistently, having been eased between October 2020 and early November 2020.

Bearing in mind that, Israel is characterized by great social disparities and a high proportion of people living in poverty (Saban et al., 2021). Providing people with income guarantees during periods of absence from their workplaces seems to be an important component in meeting public health regulations (Bodas & Peleg, 2020).

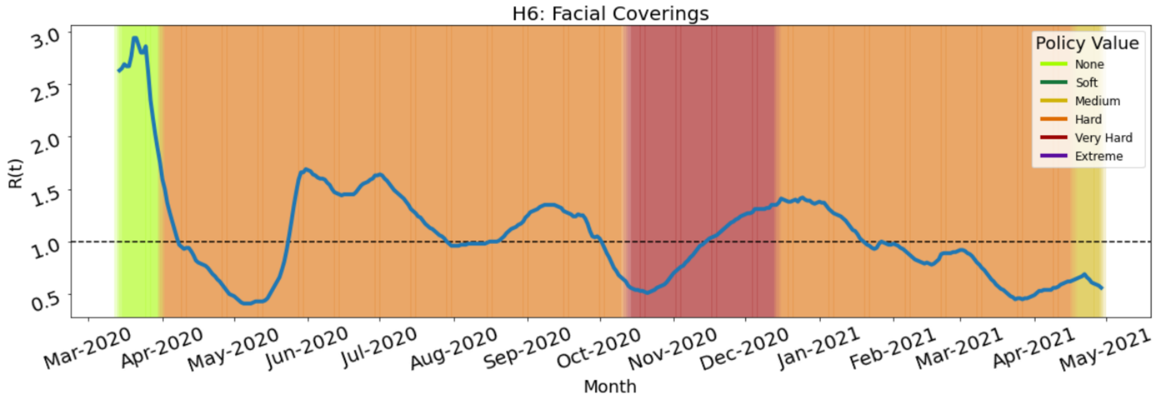

Figure 24: Facial Coverings NPI in Israel

The association of the facial coverings NPI (H6) with the R_t was due to the sharp decrease of the R_t at the very beginning of its implementation (Figure 24). We must consider that during the initial period of the pandemic in March 2020 other NPIs were implemented, namely, restrictions on gatherings (C4) and international travel controls (C8). Throughout the implementation of the facial coverings (H6) NPI this had an increase in restriction from mid-October 2020 to mid-December which coincides with the increase in the R_t. This may be due to the easing of some other restrictions during this period, for example, the lifting of the close public transport (C5) NPI. By mid-April 2021 it is noted that the government has eased the restriction on this NPI which can be justified with the advancement of vaccination.

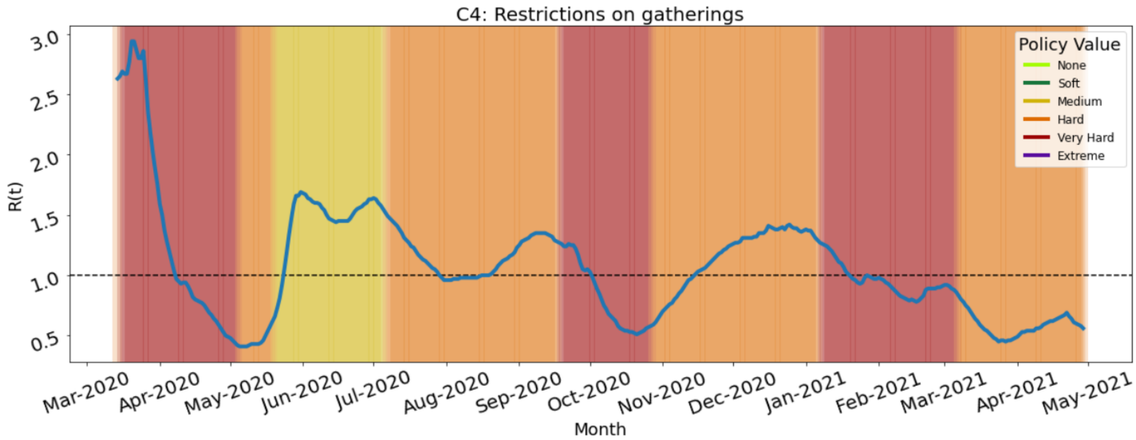

Figure 25: Restrictions on gatherings NPI in Israel

Our results only show association between the R_t and the NPIs income support (E1) and facial coverings (H6) however, these did not get much variation in degree of restriction over these months of the pandemic in contrast to NPI restrictions on gatherings (C4) (Figure 25) which, at the beginning of the pandemic, was implemented with a very high degree of restriction. In Israel, holidays are critical occasions characterized by family gatherings, increased travel and more frequent religious services, and to curb these activities, the government has enacted special measures e.g., during the Passover period (April 8th to April 15th), such as restrictions on gatherings and stay at home requirements (Maor et al., 2020; Waitzberg et al., 2020).

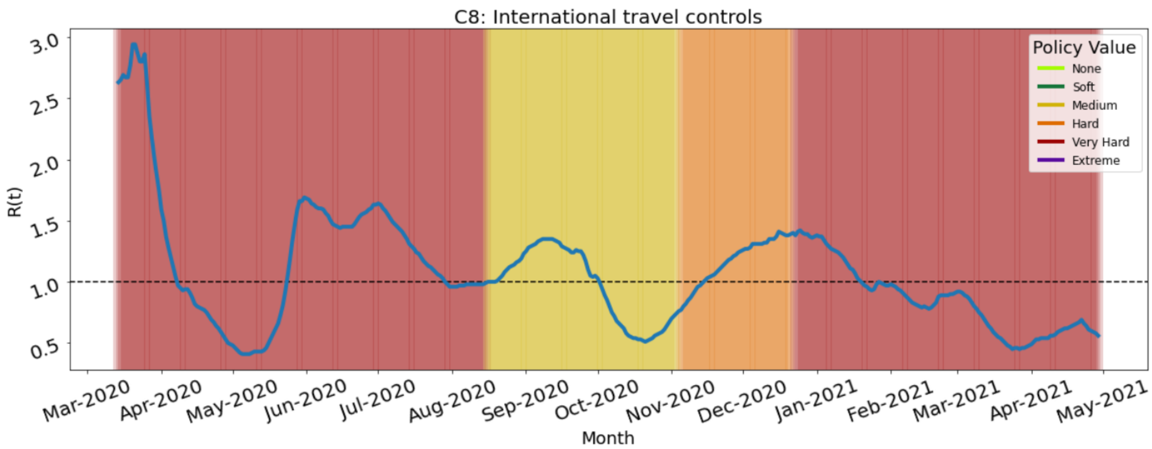

Figure 26: International Travel Controls NPI in Israel

Just the restrictions on gatherings (C4) NPI (Figure 25) and the international travel controls (C8) NPI (Figure 26) were implemented at the beginning of the pandemic with a high degree of restriction. Although these NPIs did not have enough association with the R_t (Figure 22), they seem to have been important measures as quarantines were imposed on travellers from Ben-Gurion Airport, which functions as a major entry point with 24 million travellers in 2019 (Maor et al., 2020).

We can assume that the results regarding the restrictions on gatherings (C4) and international travel controls (C8) NPIs were due to the relief of the restriction of the NPIs cancel of public events (C3), stay at home requirements (C6) and restrictions on internal movement (C7), during the months of August 2020 and mid-September 2020, as well as, the move to cold seasons and the withdrawal of NPI close public transports (C5), between late October 2020 and January 2021.

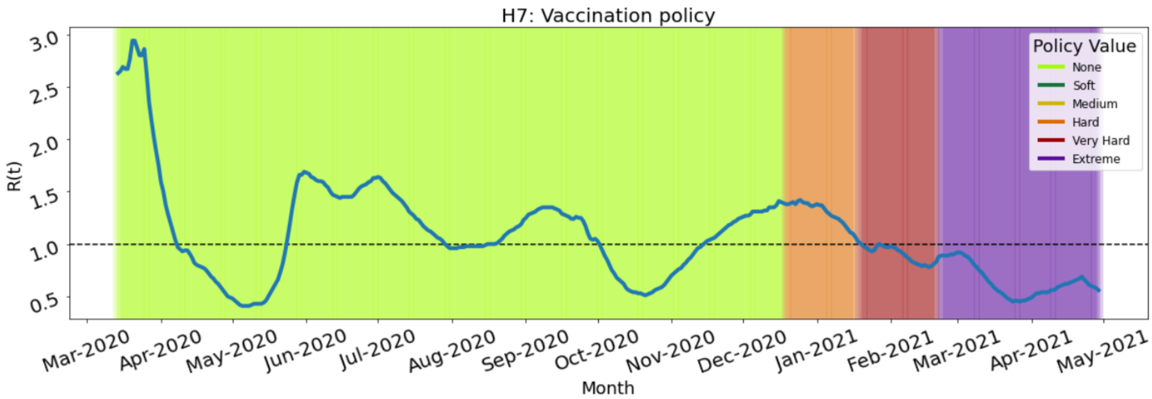

Figure 27: Vaccination Policy NPI in Israel

In Figure 27, it should be noted that the vaccination policy (H7) is the NPI that reached its maximum level of implementation since the end of March having vaccinated 29.2% of the population by January 25th 2021 (Jabal et al., 2021).

Despite the NPI vaccination policy (H7) having a weak association with the R_t, Israel employs a rapid and robust vaccination effort and as of January 29th 2021 had the highest percentage of vaccinated population (Caspi et al., 2021).

4.1.5 Portugal

Portugal declared a state of emergency on March 18th, 2020 (Costa et al., 2020) and as a result of the lockdown a variety of NPIs were adopted namely, a ban on events with 100 or more people; a ban on drinking alcoholic beverages in public outdoor spaces, except in outdoor dining areas and drinking lounges, duly licensed for that purpose; document control of people at borders; air traffic to and from Portugal was banned for all flights to and from non-EU countries, with some exceptions (Caetano et al., 2021; Silva et al., 2021).

According to Costa et al., (2020) in a first phase of the pandemic, the evolution of the number of confirmed cases until April 1st 2020, shows a strong relation with population density, with the urban structure and the location of the main concentrations of employment and then the phenomenon expanded with strong expression to the territories of the north and central interior of the mainland territory, counties with lower population density and an aging population. In the following analysis we consider data from March 14th 2020 to early May 2nd 2021.

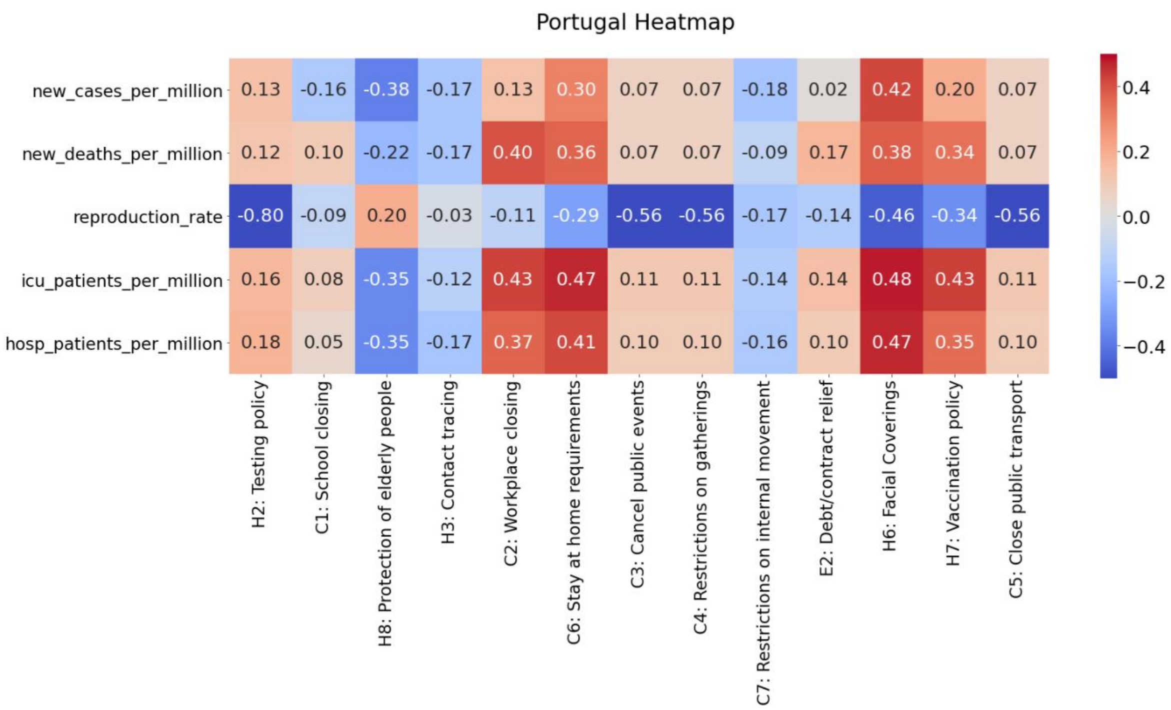

Figure 28: Portugal NPIs Heatmap

From Figure 28 we can see that the NPI testing policy (H2) has a strong negative association with the reproduction rate. Additionally, the NPIs cancel public events (C3), restrictions on gatherings (C4), close public transport (C5) and facial coverings (H6) had moderate negative associations with the R_t.

The degree of restriction is higher in the NPIs workplace closing (C2), stay at home requirements (C6), facial coverings (H6) and vaccination policy (H7) when the number of ICU patients increases. The restriction degree in NPI workplace closing (C2) is greater when new deaths increase. The rise of the restriction degree in NPI stay at home requirements (C6) is higher when the number of ICU and hospitalized patients is higher.

The rise in NPI facial coverings (H6) restriction coincides with the increase in new cases, ICU patients and hospitalized patients. Finally, the NPI vaccination policy (H7) started to be implemented to a greater extent when the number of ICU patients were increasing.

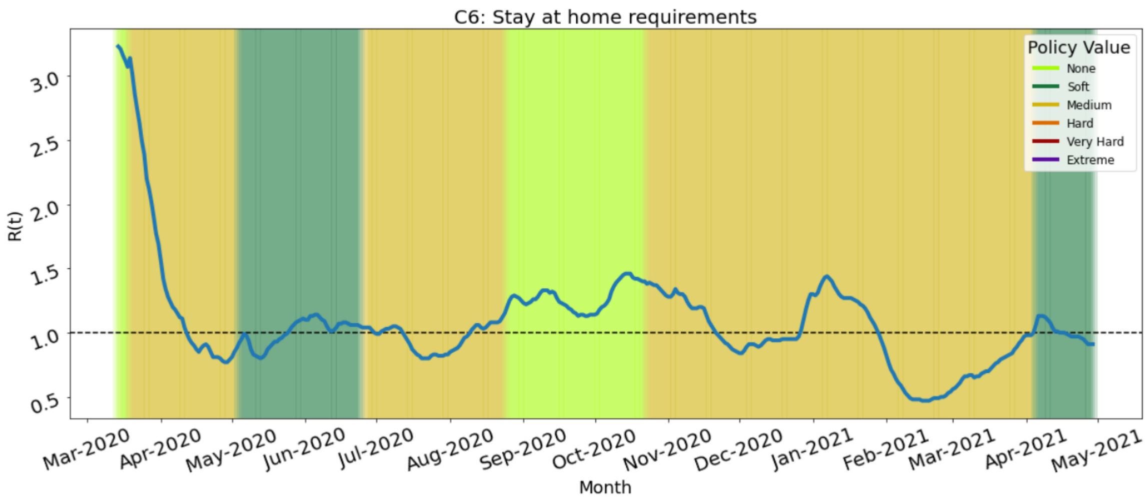

Figure 29: Stay at Home Requirements NPI in Portugal

The NPI stay at home requirements (C6) (Figure 29) was among the first measures to be implemented at the very beginning of the pandemic in March 2020 with a high degree of restriction during the first phase of containment in Portugal. This lockdown severely restricted the movement of individuals within the country, as well as instituted a mandatory stay at home order for the population (Caetano et al., 2021). After 45 days of the state of emergency, the government progressively established measures for the reopening of the economy, but with rules for controlling the spread of the virus (Silva et al., 2021).

Between May 2020 and mid-June 2020 the government lowered the degree of the restriction and between mid-August and October 2020 they completely lifted this NPI. After this period and until early April 2021, the NPI stay at home requirements (C6) was implemented with a medium degree of restriction and from April 2021 they lowered the degree of restriction to soft.

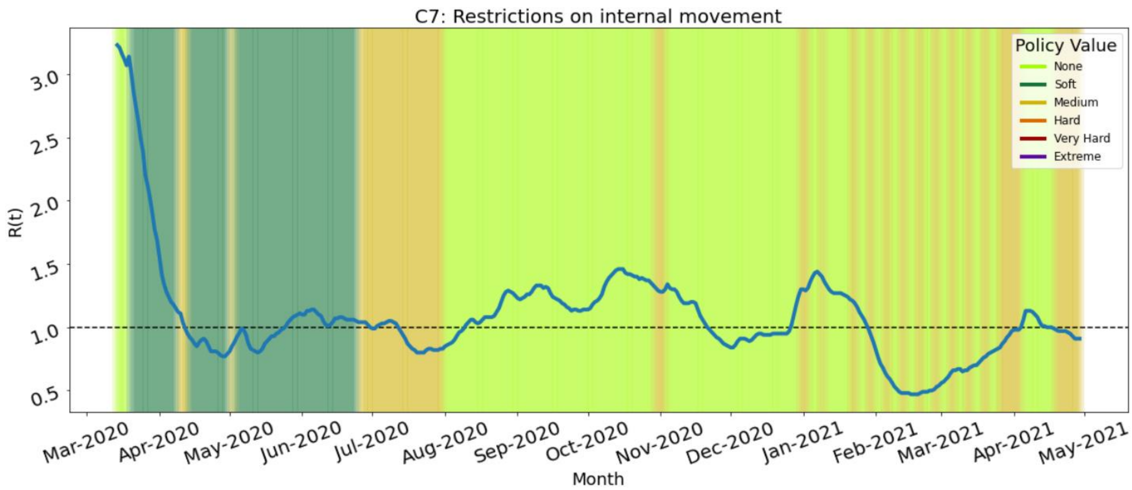

Figure 30: Restrictions on Internal Movement NPI in Portugal

Initially the NPI restrictions on internal movement (C7) (Figure 30) began to be implemented in March 2020 with a low degree of restriction interspersed with a higher degree of restriction around Easter weekend in April 2020 and in early May 2020. According to Silva et al., (2021), health literacy should be a central goal to achieve as it prevented the breakdown of the national health system in this early period of the pandemic.

Subsequently, in late June 2020 to late July 2020 the degree of restriction of this NPI increases again, however, after this period the government lifts this restriction. Later on, this NPI was implemented again between October and November 2020 and then it was reinstated in several intervals, with medium restriction level, successively from late December 2020 during the holidays and New Year's Eve. Despite the increase in restriction during the Christmas season there was a rise in the R_t.

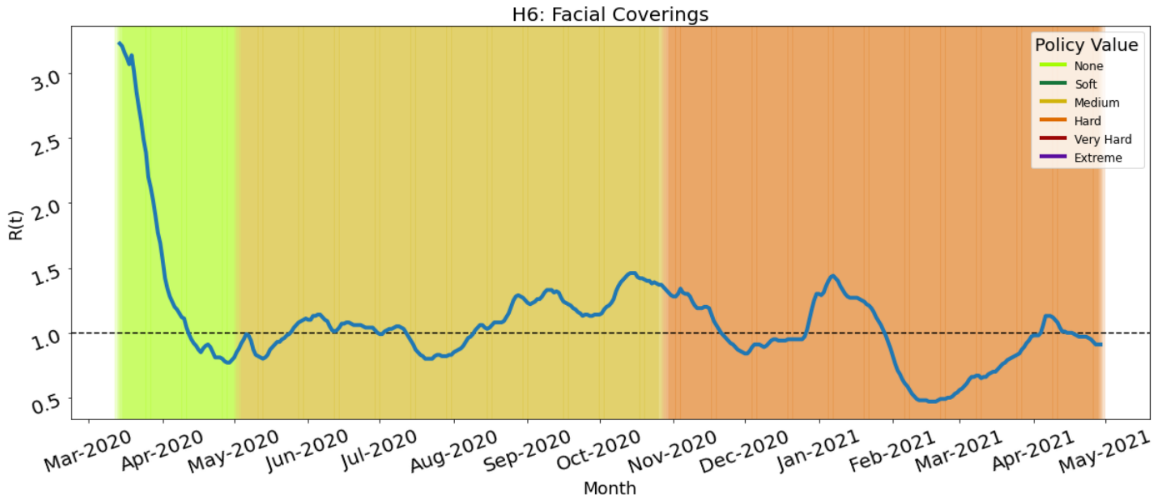

Figure 31: Facial Coverings NPI in Portugal

The NPI facial coverings (H6) (Figure 31) started to be implemented at the beginning of May 2020 with a medium restriction degree and from the end of October 2020 with a hard restriction degree. Similar to other countries in Europe, the end of summer led to an increase in the number of new cases which was probably the result of the low perception of the risk of infection during summer and the movement of the population during the vacations, leading to Portugal declaring a state of calamity on October 15th, 2020 and also the mandatory use of face masks outdoors on November 4th, 2020 (Caetano et al., 2021).

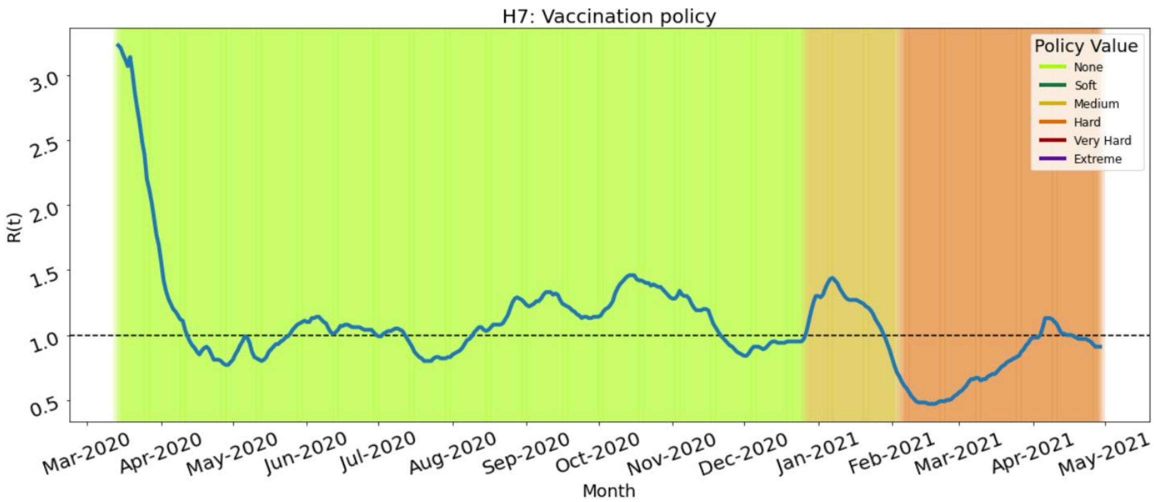

Figure 32: Vaccination Policy NPI in Portugal

The NPI vaccination policy (H7) (Figure 32) started to be implemented in late December 2020 and coincided at the beginning of a new wave for Portugal after the Christmas season. Subsequently, vaccination is applied on a larger scale starting in February 2021.

4.2 Predictive machine learning models

We tested several machine learning models in order to compare the results of data from 101 countries where each row contains the value of the R_t per day in a particular country. Table 7 shows all the features used for the models' input. Three types of models were used for predicting the R_t: linear models, nonlinear models, and ensemble models.

Due to their simplicity and low complexity linear regression models are among the most fundamental and widely used techniques to model and predict the behaviour of massive data flows (Akgün & Öğüdücü, 2015).

The linear regression models that were applied were Linear Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Ridge Regression, Elastic Net, Huber Regression, Least-Angle Regression (LARS), Passive Aggressive Regressor, and the Stochastic Gradient Descent (SGD) Regressor.

In general, nonlinear regression models are similar to linear regression models, where an important difference from nonlinear regression models is that the number of regression parameters is not necessarily directly related to the number of independent variables in the model (Crawley, 2012).

In this dissertation, we apply four nonlinear models: K Nearest Neighbours Regression (KNN), Decision Tree Regression, Extra-Tree Regression and Support Vector Regression (SVR).

Ensemble models aim to combine several models in order to improve the prediction accuracy in learning problems with a numerical target variable (Mendes-Moreira et al., 2012). The ensemble models used were AdaBoost, Random Forest Regressor, Extra-Trees Regressor, Gradient Boosting and Extreme Gradient Boosting (XGBoost).

Table 8 shows that the results of the models improved with the addition of Hofstede's cultural dimensions. Comparing the results of all the models we can state that the linear models were the ones with the highest errors. This is explained by the fact that the data does not follow a linear trend, which can be seen through the correlations obtained in the heatmap shown in Figure 4.

Although the linear models were those with the worst results, the ones that applied the Huber regression obtained the lowest MAE error of 0.22292 with the use of Hofstede's six cultural dimensions and an error of 0.22379 without them. Linear and Ridge regression both showed the lowest MSE of 0.11174 with cultural dimensions and an error of 0.11329 without them.

Also, the previous models both showed the lowest RMSE of 0.33427 with cultural dimensions and 0.33659 without the cultural dimensions. In the nonlinear models the error values obtained are significantly better which confirms the nonlinearity of the data, justifying the use of MI for the feature selection stage.

Comparing the nonlinear models, the Decision Tree regressor was the method that obtained the best results for the three metrics: a MAE of 0.08628 with cultural dimensions and an error of 0.08603 without them; a MSE of 0.02122 with cultural dimensions and an error of 0.02127 without them; and a RMSE of 0.14569 with cultural dimensions and an error of 0.14584 without them.

Finally, we tested the ensemble models which performed even better compared to the nonlinear models. This is because these methods are an ensemble of several models, for example, a Random Forest is an ensemble of Decision Trees.

The model with the lowest error was the Random Forest regressor with a MAE of 0.08659, a MSE of 0.02101 and a RMSE of 0.14495 without considering the use of the cultural dimensions. These results further improve with the use of the cultural dimensions being the Random Forest regressor the best model with a MAE of 0.08575, a MSE of 0.02077 and a RMSE of 0.14412.

Table 8: Machine learning models results

Predictive Model \ Error	Without the cultural dimensions			With the cultural dimensions		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Linear Models						
Linear Regression	0.22795	0.11329	0.33659	0.22758	0.11174	0.33427
LASSO	0.22903	0.12197	0.34924	0.22903	0.12197	0.34924
Ridge	0.22795	0.11329	0.33659	0.22758	0.11174	0.33427
Elastic Net	0.22903	0.12197	0.34924	0.22903	0.12197	0.34924
Huber Regression *	0.22379	0.11693	0.34195	0.22292	0.11554	0.33992
LASSO LARS	0.22903	0.12197	0.34924	0.22903	0.12197	0.34924
Passive Aggressive Regressor	0.34407	0.22909	0.47863	0.29025	0.16326	0.40405
SGD Regressor	0.22842	0.11414	0.33785	0.22773	0.11261	0.33557
Non-linear Models						
KNN Regressor	0.09761	0.02753	0.16593	0.09643	0.02700	0.16433
Decision Tree Regressor	0.08603	0.02127	0.14584	0.08628	0.02122	0.14569
Extra-Tree Regressor	0.08673	0.02175	0.14748	0.08642	0.02261	0.14931
SVR	0.12457	0.03815	0.19532	0.12099	0.03511	0.18738
Ensemble Models						
AdaBoost Regressor	0.29101	0.13456	0.36682	0.27001	0.11759	0.34281
Random Forest Regressor	0.08659	0.02101	0.14495	0.08575	0.02077	0.14412
Extra-Trees Regressor	0.08596	0.02088	0.14451	0.08587	0.02084	0.14437
Gradient Boosting Regressor	0.18593	0.07183	0.26802	0.18566	0.07004	0.26466
XGBoost Regressor	0.10535	0.02552	0.15976	0.10443	0.02504	0.15824

In order to optimize the two best ensemble models, their hyperparameters were adjusted using the GridSearch tool. Table 9 shows the best hyperparameters used for the Extra Trees and Random Forest.

Table 9: Hyperparameter tuning for the three best models

Models	Hyperparameters
Extra Trees Regressor	Estimators: 50 Maximum depth: 100 Bootstrap: False Minimum samples split: 2 Maximum features: auto Minimum samples leaf: 1
Random Forest Regressor	Estimators: 200 Maximum depth: 100 Bootstrap: False Minimum samples split: 2 Maximum features: sqrt Minimum samples leaf: 1

In general, the ensemble models obtained low R_t prediction errors with the addition of the features of the six cultural dimensions along with the features of the socioeconomic, demographic, and various degrees of restriction levels of the NPIs.

By testing the best parameters of each of the two best ensemble models it was concluded that the Random Forest regressor obtained the lowest error. Thus, this one was selected to predict the R_t for the five countries analysed in detail in this dissertation. Table 10 shows the results obtained for predicting the R_t for each of these countries. It can be observed that India was the country that obtained the best result with a MAE of 0.02637, MSE of 0.00163 and a RMSE of 0.04046. Whereas the UK was the worst performing country with a MAE of 0.09994, a MSE of 0.02941and a RMSE of 0.17151.

Table 10: Results of the errors for the random forest regressor regarding the data of the five considered countries

Errors / Countries	MAE	MSE	RSME
India	0.02637	0.00163	0.04046
Brazil	0.03083	0.00225	0.04749
UK	0.09994	0.02941	0.17151
Israel	0.07953	0.01677	0.12951
Portugal	0.07347	0.01142	0.10687

Chapter 5 Discussion

This chapter is divided into the following sections:

- In **section 5.1** a comparison is made between the NPIs and the underlying culture in each of the countries;
- In **section 5.2** an analysis of the results of the machine learning models is performed;
- **Section 5.3** analyses the impact of cultural dimensions on the evolution of the pandemic.

5.1 Comparison between NPIs and culture in the five countries

Our results show the importance of imposing NPIs from the beginning of the pandemic where there is a marked decrease in the R_t and successful response results from the early identification of effective NPIs taking into consideration the geopolitical, social and health system differences in each country (Leshem et al., 2020). Furthermore, it should be taken into consideration that, the effect of the introduction and withdrawal of NPIs has a delay of 1-3 weeks (Li et al., 2020) which may have an impact on getting associations in NPIs.

In this analysis we address the response of five distinct countries India, Brazil, the UK, Israel and Portugal in which we compare the various NPIs implemented and their degrees of restriction from March 2020 to early May 2021, relating them to Hofstede's six cultural dimensions. Hereafter, we relate our results to the culture of each country under analysis.

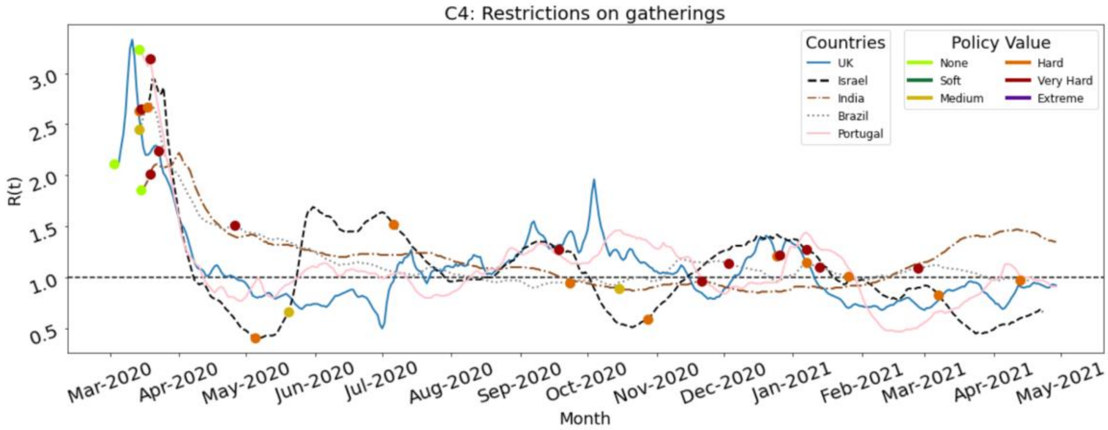

Figure 33: Restrictions on Gatherings NPI for the five countries

The NPI restrictions on gatherings (Figure 33) was one of the most restrictive measures implemented by the countries under analysis and that had a higher degree of restriction since its implementation, the UK and Portugal were the only countries in which we obtained a relevant association for the NPI restrictions on gatherings. According to Luu & Huynh (2020) countries with higher uncertainty avoidance (in our case, Portugal, Brazil and Israel) predict the lowest proportion of people meeting in public however, in our results Portugal was the only country that has this characteristic and obtained correlation with decreased R_t.

In India, this NPI was also implemented with no change in restriction level during the first months of the pandemic; however, no relevant association with the R_t was found. Although the UK and India are cultures in which there is no need to establish many rules, in India the rules that do exist are not always obeyed by the society while in the UK the rules that exist are respected by the population (Hofstede Insigths, 2021).

Portugal has a culture with a very high preference for uncertainty avoidance, which had a great influence on the beginning of the pandemic. The climate of uncertainty regarding the emergence of a new virus was conducive to the implementation and enforcement of new rules. On the other hand, India is culturally different from Portugal in this aspect, as they do not mind unexpected situations and rules are often in place only to be circumvented, which may have caused the uncontrolled evolution of the number of new cases of COVID-19 in this country.

In Israel and Brazil, it was verified that the NPI restrictions on gatherings alternated between hard and very hard throughout the months of the pandemic, however, there was no association between the R_t and this NPI. In the case of Brazil, it may have been due to the fact that they are a highly indulgent society as they like to enjoy their leisure time combined with a low level of obedience to the rules that are established for the society. In Israel, despite being a society that sets a variety of rules, these do not always seem to work, and the Israelis are a society that has a great respect of traditions, which may be possible reasons for the lack of association.

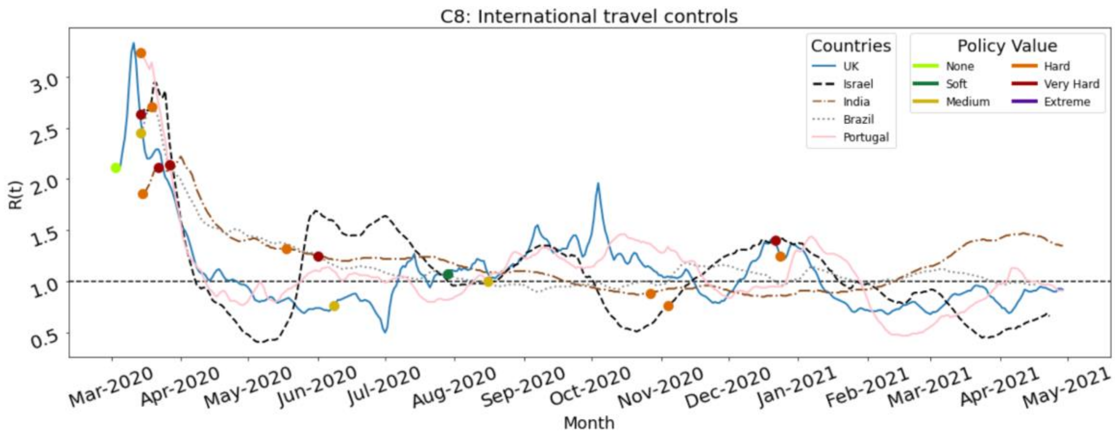

Figure 34: International Travel Controls NPI for the five countries

The NPI international travel controls (Figure 34) was among the first NPIs to be implemented in mid-March 2020 by Israel, Portugal, Brazil and India except in the UK which only implemented this NPI in late May 2020. The implementation of this NPI in the UK seems to have been later, however, in the countries under analysis, it was the only country where this NPI had a relevant association with the R_t, contrary to the analysis of Liu et al., (2020) where no evidence was found of the implementation of this NPI in decreasing the R_t because it only makes sense in preventing the introduction of infections (Russell et al., 2021).

In Israel, Portugal, Brazil, and India the governments have implemented a variety of NPIs such as international travel controls since the beginning of the pandemic. In countries like Brazil, Israel and Portugal there is a great need to create rules, which may explain why so many NPIs were implemented at the beginning of the pandemic.

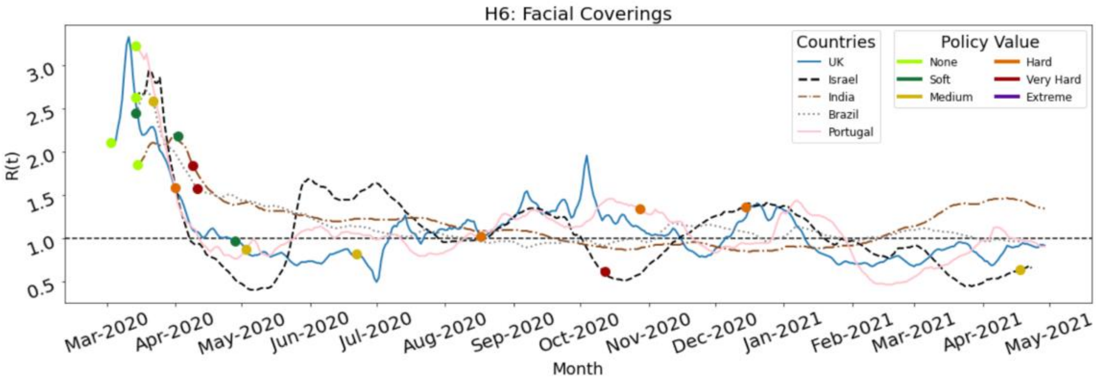

Figure 35: Facial Coverings NPI for the five countries

In India, Brazil, Israel and Portugal the facial coverings NPI (Figure 35) had a relevant association with the R_t. In the UK, this NPI was only implemented in late April 2020. According to Raina et al., (2021) despite its later implementation this NPI had a strong adherence to its use and it did not reduce its compliance with other NPIs. In opposition, in the study of Lu et al., (2021) people in more individualistic countries are less likely to wear masks and if we look at Hofstede's cultural dimensions it is noteworthy that the UK is a highly individualistic society centered on the "I" as opposed to India, Brazil, Israel, and Portugal which have collectivism associated to their culture (Hofstede Insigths, 2021) and, therefore, the implementation of the NPI facial coverings is seen as an important measure right at the beginning of the pandemic for the control of the R_t in these countries where the population is inserted in cohesive groups.

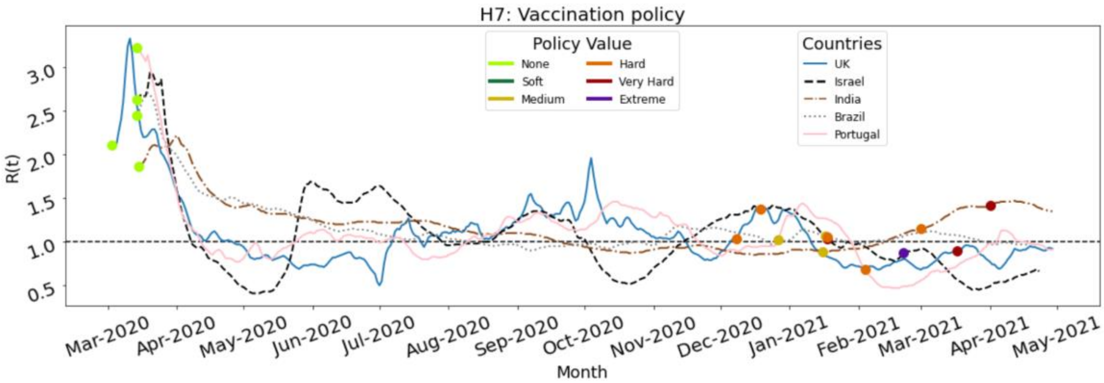

Figure 36: Vaccination Policy NPI for the five countries

Mass vaccination in combination with other NPIs is seen as one of the central elements for pandemic control (Jabal et al., 2021). Israel, Portugal and the UK were the first countries to implement vaccination in December 2020 while in Brazil and India population vaccination started later in January 2021(OurWorldInData, 2021).

Despite having no association with the R_t at the beginning of May 2021 we have already been able to observe the R_t below 1 for Brazil, the UK, Israel, and Portugal (Figure 36).

According to the authors Brauner et al., (2021) and Fontanet et al., (2021) vaccination coupled with other NPIs would be crucial to decrease the spread of COVID-19 however with the emergence of new COVID-19 variants the range of immunity among the population is not so linear.

5.2 Analysis of the Predictive Models

According to Yeung et al., (2021) machine learning models can predict the Confirmed Infection Growth in 14 days (CIG_{14}) with an appreciable degree of accuracy. In this study, the authors aimed to predict CIG_{14} using NPIs and Hofstede's cultural dimensions, for which they tested five machine learning models: Ridge regression, Decision Tree Regression, Random Forest Regression, AdaBoost Regression and Support Vector Regression.

In this dissertation we used machine learning models to predict the R_t, we considered different NPIs (16 ordinal and one index) than the ones chosen on the above-mentioned study (13 ordinal and 4 indexes). Besides these, we also considered socioeconomic and demographic features, namely population, population density, median age, population aged 65 older, GDP per capita, life expectancy, and human development index. From our results we can see that the use of Hofstede's cultural dimensions improved the models compared to the models without the cultural features.

Regarding the evaluation metrics in Yeung et al., (2021), MAE and median percent error were used while, in this dissertation MAE, MSE and RMSE were used. Also, in the study of Yeung et al., (2021) three validation methods were used, in-distribution, out-of-distribution and cross-validation whereas, in this dissertation the hold-out technique was applied.

We can see that in both studies the ensemble models were the ones that obtained better results for the R_t prediction because these are the combination of several predictions coming from a set of models to make the final prediction through averaging, voting so that the ensemble model is better than any of the individual models (Ganaie et al., 2021). The Random Forest model was the one that obtained better results in both, being that in Yeung et al., (2021) the authors obtained a MAE of 0.031 for predicting the CIG_{14} and in this dissertation we obtained an MAE of 0.085 for predicting the R_t.

Table 11 shows a comparison between the study of Yeung et al., (2021) and this dissertation in order to understand the main differences of both.

Table 11: Comparison of this dissertation with the work of Yeung et al., (2021)

	Yeung et al., (2021)	This dissertation
Objective	Predict the national growth of confirmed COVID-19 infections across 14 days	Understand if cultural dimensions are useful data to improve the R_t predictions
Date range	April 1st to September 30th, 2020	March 3rd 2020 to May 2nd 2021
Countries	114	101
Target	Confirmed Infection Growth in 14 days (CIG_{14})	Reproduction rate (R_t)
Features	17 NPIs: 13 ordinal + 4 indexes 6 Hofstede cultural dimensions	17 NPIs: 16 ordinal + 1 index 7 Socioeconomic and demographic 6 Hofstede cultural dimensions
Evaluation metrics	MAE Median Percent Error	MAE MSE RMSE
Validation techniques	In-distribution Out-of-distribution Cross-validation	Hold out
Best models	Random Forest AdaBoost	Random Forest Extra Trees

5.3 Culture´s impact on the R_t

The R_t depends on several factors such as contact with infected individuals, age-specific susceptibility to infection, behavioural susceptibility based on region, and seasonality (Yechezkel et al., 2021). As a way to control and reduce the number of infections by COVID-19 several NPIs have been imposed however, their effectiveness depends on social compliance, namely the extent to which people reduce their daily contacts following government restrictions that are difficult to measure prospectively (Liu et al., 2020).

The UK and Israel were countries that stood out positively throughout the pandemic which can be explained by their cultural, social, economic and political context. In Israel which is a highly regulated society (Hofstede Insigths, 2021) its population tends to conform to government regulations as Israelis exhibit a relatively high level of trust in the instructions of health authorities and comply with a variety of regulations, including reporting disorderly people who violate quarantine regulations, and support compliance with quarantine regulations and criminal actions against those who do not comply (Bodas & Peleg, 2020).

In the case of the UK, at the start of the pandemic, the importance of research was emphasized by medical directors, scientific advisors and the UK government. The WHO convened scientists in mid-February 2020 and also, the UK research funding was quickly made available to carry out the research needed to ensure the long-term health and well-being of the population (Oyebode et al., 2021).

As opposed to Brazil and India which face different realities. In Brazil scientific research has been greatly affected due to the reduction of its financial support and grants. In addition, approximately 2 millions of São Paulo's population live in "slums" where people live in poor conditions, without access to health, social and financial support (Marson & Ortega, 2020). In India, health inequalities, increasing economic, social disparities, and distinct cultural values affect the control of COVID-19 contagions (The Lancet, 2020).

These countries have been dealing with serious difficulties since the beginning of the pandemic, and although they have implemented restrictions for several NPIs, their cultural dimensions seem to exert a strong influence on their behaviour. Other factors may have an important influence, for example in Brazil´s politics and in India´s socio-demographic factors, as the latter has a population of 1.3 billion people (The Lancet, 2020).

Nevertheless, the UK and Israel also had some difficulties at the beginning of the pandemic namely, Israel's healthcare system faced a chronic shortage of healthcare resources. However, as Israel shifted from containment to mitigation, the structural features were leveraged to improve the response (Leshem et al., 2020) and the UK had to deal with a high rate of infections during the first few months of the pandemic (OurWorldInData, 2021). Despite their difficulties they managed to control the pandemic better than Brazil and India.

The main difference between the UK and Israel relative to India and Brazil relates to power distance (PDI). While the former are cultures with lower power distance meaning that they prefer a participatory decision-making process, in India and Brazil a centralized decision-making process dominated by top leaders is seen as a legitimate way to formulate and implement collective coping strategies to COVID-19 (Guan et al., 2020).

Israel and the UK are countries that show to value equal rights (Hofstede Insigths, 2021) which may be one of the reasons to justify their actions for coping with the current pandemic. For example, these countries are among the top three that applied more daily tests and vaccinations during the pandemic (OurWorldInData, 2021).

In the case of Portugal, we can say that it had a very positive performance, for example, in the beginning of the pandemic Portugal showed a rapid response since March 3rd, 2020, when the first infected Portuguese patient was declared, until March 18th, 2020, when a State of Emergency was declared in Portugal (Costa et al., 2020). Portugal was also one of the first countries to implement vaccination in December 2020, like the UK and Israel. It should be noted that Portugal differs from Brazil and India in several aspects, namely in its socioeconomic and political context already described above. Compared to India and Brazil, Portugal is a country with a lower population density, a higher human development index and also a higher GDP per capita.

Chapter 6 Conclusion

Since the start of the pandemic several NPIs have been implemented with varying degrees of restriction by the five countries analysed in this dissertation. Previous researchers have focused on analysing the impact of the NPIs implemented over the first few months of the pandemic, but we have done a more extensive analysis from the start of the pandemic to early May 2021 for India, Brazil, the UK, Israel and Portugal and we tested several machine learning models, with and without Hofstede's six cultural dimensions, to compare and understand whether these improve the prediction of the reproduction rate.

In summary, this dissertation is an in-depth and extensive analysis over these months of the pandemic which includes the impact of the restriction degree of the NPIs implemented, and additionally the underlying cultural component in these countries which we believe plays an important role in each country's response to the several NPIs implemented.

6.1 Main Contributions

This dissertation set out to answer the following research questions:

How can cultural characteristics explain the introduction and consequent impact of the NPIs in different countries?

For this question, in this dissertation we found that the facial coverings NPI had a strong association with the R_t in India, Brazil and Israel but in the UK had a weaker association, which may be justified by its later implementation, and that the importance of this NPI for R_t control is remarkable because COVID-19 is a virus that is easily transmitted through the air. Other NPIs have a prominent place such as, in collective countries as in the case of Brazil the stay at home requirements NPI, and on the other hand, in the UK and Israel the income support NPI which seems to have been an important measure since, with the pandemic came the increase of unemployment (Oyebode et al., 2021).

Our results show differences between the five countries that implemented the same NPIs to varying degrees during this period and our analysis suggests a look at the underlying Hofstede’s cultural dimensions indicating that India and Brazil are countries with marked inequalities as opposed to the UK and Israel which may be reflected in the higher number of tests performed and the latter's advanced state of immunization. Portugal also differs from countries such as India and Brazil in socio-economic and demographic characteristics. Overall, the UK, Portugal and Israel stand out as having implemented vaccination by December 2020.

Can cultural characteristics be useful information for predicting the R_t?

To answer this question, we used data from 101 countries to train several machine learning models to compare the results between the models with and without Hofstede's cultural dimensions. Our results showed that the use of cultural dimensions helped improve the models, and that the ones that obtained a better prediction of the R_t were the ensemble models, especially the Random Forest which without the cultural dimensions obtained a MAE of 0.08659 and with the cultural dimensions obtained an improved error of 0.08575. Additionally, we used two models: Random Forest Regression and Extra-Trees Regression to predict the value of the R_t for the five countries considered in this dissertation, where we found a better result for India (MAE=0.02637) and Brazil (MAE=0.03083).

6.2 Limitations

The main limitation of this study was the lack of more specific variables on the number of hospitalizations and ICU patients per million for India and Brazil as well as the number of new tests per thousand for Brazil in order to better compare these two countries with the UK, Israel and Portugal.

6.3 Future Research

Regarding the vaccination policy NPI, we started a first analysis which considered the beginning of its implementation between mid-December 2020 and January 2021 until May 2021 where it is possible to observe at the beginning of May 2021 the R_t value is below 1 for Brazil, the UK and Portugal. In this regard, we suggest future studies that follow the trend of the reproduction rate as vaccination progresses over the coming months. In addition, with the emergence of new variants (e.g., Delta variant) we suggest as future research a deeper analysis that includes these COVID-19 variants.

We also tested several machine learning models such as linear, nonlinear and ensemble models to predict the R_t value. The hold-out validation technique was used to split the data, so we propose for future studies the use of other validation techniques.

References

Akgün, B., & Öğüdücü, Ş. G. (2015). Streaming linear regression on spark MLlib and MOA. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, 1244–1247. https://doi.org/10.1145/2808797.2809374

Alvarez Candido, S. E., Sacomano Neto, M., & Côrtes, M. (2021). How social inequalities shape markets: Lessons from the configuration of PET recycling practices in Brazil. *Business & Society*. https://doi.org/10.1177/0007650321989063

Arroyo-Marioli, F., Bullano, F., Kučinskas, S., & Rondón-Moreno, C. (2020). Tracking $\mathscr{R}$ of COVID-19: A New Real-Time Estimation Using the Kalman Filter. *MedRxiv*, 2020.04.19.20071886. https://doi.org/10.1101/2020.04.19.20071886

Auger, K. A., Shah, S. S., Richardson, T., Hartley, D., Hall, M., Warniment, A., Timmons, K., Bosse, D., Ferris, S. A., Brady, P. W., Schondelmeyer, A. C., & Thomson, J. E. (2020). Association between Statewide School Closure and COVID-19 Incidence and Mortality in the US. *JAMA - Journal of the American Medical Association*, *324*(9), 859–870. https://doi.org/10.1001/jama.2020.14348

Azevedo, L., Pereira, M. J., Ribeiro, M. C., & Soares, A. (2020). Geostatistical COVID-19 infection risk maps for Portugal. *International Journal of Health Geographics*, *19*(1), 1–8. https://doi.org/10.1186/s12942-020-00221-5

Bo, Y., Guo, C., Lin, C., Zeng, Y., Li, H. B., Zhang, Y., Hossain, M. S., Chan, J. W. M., Yeung, D. W., Kwok, K. O., Wong, S. Y. S., Lau, A. K. H., & Lao, X. Q. (2021). Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. *International Journal of Infectious Diseases*, *102*, 247–253. https://doi.org/10.1016/j.ijid.2020.10.066

Bodas, M., & Peleg, K. (2020). Self-isolation compliance in the COVID-19 era influenced by compensation: Findings from a recent survey in Israel. *Health Affairs*, *39*(6), 936–941. https://doi.org/10.1377/hlthaff.2020.00382

Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick, M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., & Kulveit, J. (2021). Inferring the effectiveness of government interventions against COVID-19. *Science*, *371*(6531). https://doi.org/10.1126/science.abd9338

Caetano, C., Morgado, M. L., Patrício, P., Pereira, J. F., & Nunes, B. (2021). Mathematical Modelling of the Impact of Non-Pharmacological Strategies to Control the COVID-19 Epidemic in Portugal. *Mathematics*, *9*(10), 1084. https://doi.org/10.3390/math9101084

Caspi, G., Dayan, A., Eshal, Y., Taub, S., Twig, G., Shalit, U., Lewis, Y., Shina, A., & Caspi, O. (2021). Socioeconomic Disparities and COVID-19 Vaccination Acceptance: Experience from Israel. *MedRxiv*.

Castex, G., Dechter, E., & Lorca, M. (2020). COVID-19: The impact of social distancing policies, cross-country analysis. *Economics of Disasters and Climate Change*, 135–159. https://doi.org/10.1007/s41885-020-00076-x

Chernozhukov, V., Kasahara, H., & Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *Journal of Econometrics*, *220*(1), 23–62. https://doi.org/10.1016/j.jeconom.2020.09.003

Costa, E. M., Pinto and, & Costa, N. M. (2020). A Pandemia COVID-19 em Portugal Continental – Uma Análise Geográfica da Evolução Verificada nos Meses de Março e Abril. *Angewandte Chemie International Edition, 6(11), 951–952.*, 72–79.

Crawley, M. J. (2012). Applied Linear Regression Models:Non-linear Regression. *Book*, 661–684.

Douglas, M., Katikireddi, S. V., Taulbut, M., McKee, M., & McCartney, G. (2020). Mitigating the wider health effects of covid-19 pandemic response. *The BMJ*, *369*(April), 1–6. https://doi.org/10.1136/bmj.m1557

Duhon, J., Bragazzi, N., & Kong, J. D. (2021). The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Science of the Total Environment*, *760*, 144325. https://doi.org/10.1016/j.scitotenv.2020.144325

Erman, A., & Medeiros, M. (2021). Exploring the Effect of Collective Cultural Attributes on Covid-19-Related Public Health Outcomes. *Frontiers in Psychology*, *12*(March), 1–21. https://doi.org/10.3389/fpsyg.2021.627669

Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., & Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, *397*(10278), 952–954. https://doi.org/10.1016/S0140-6736(21)00370-6

Fuller, J. A., Hakim, A., Victory, K. R., Date, K., Lynch, M., Dahl, B., & Henao, O. (2021). Mitigation Policies and COVID-19-Associated Mortality - 37 European Countries. *MMWR. Morbidity and Mortality Weekly Report*, *70*(2), 58–62. https://doi.org/10.15585/mmwr.mm7002e4

Ganaie, M., Hu, M., Tanveer, M., & Suganthan, P. (2021). *Ensemble deep learning: A review*.

Gelfand, M. J., Jackson, J. C., Pan, X., Nau, D., Pieper, D., Denison, E., Dagher, M., Van Lange, P. A. M., Chiu, C. Y., & Wang, M. (2021). The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. *The Lancet Planetary Health*, *5*(3), e135–e144. https://doi.org/10.1016/S2542-5196(20)30301-6

Gokmen, Y., Baskici, C., & Ercil, Y. (2021). The impact of national culture on the increase of COVID-19: A cross-country analysis of European countries. *International Journal of Intercultural Relations*, *81*, 1–8.

González-Bustamante, B. (2021). Evolution and early government responses to COVID-19 in South America. *World Development*, *137*. https://doi.org/10.1016/j.worlddev.2020.105180

Guan, Y., Deng, H., & Zhou, X. (2020). Understanding the impact of the COVID-19 pandemic on career development : insights from cultural psychology. *Journal of Vocational Behavior.*, *119*, 103438. http://dro.dur.ac.uk/30766/

Güss, C. D., & Tuason, M. T. (2021). Individualism and Egalitarianism Can Kill: How Cultural Values Predict Coronavirus Deaths Across the Globe. *Frontiers in Psychology*, *12*(June), 1–13. https://doi.org/10.3389/fpsyg.2021.620490

Han, E., Tan, M. M. J., Turk, E., Sridhar, D., Leung, G. M., Shibuya, K., Asgari, N., Oh, J., García-

Basteiro, A. L., & Hanefeld, J. (2020). Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *The Lancet*.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S., & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, *4*(12), 1303–1312. https://doi.org/10.1038/s41562-020-01009-0

Hernández-Orallo, E., Manzoni, P., Calafate, C. T., & Cano, J.-C. (2020). Evaluating how smartphone contact tracing technology can reduce the spread of infectious diseases: the case of COVID-19. *Ieee Access*, *8*, 99083–99097.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (Vol. 5). sage.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1), 919–2307.

Hofstede Insigths. (2021). *Compare Countries*. https://www.hofstede-insights.com/

Islam, N., Sharp, S. J., Chowell, G., Shabnam, S., Kawachi, I., Lacey, B., Massaro, J. M., D'Agostino, R. B., & White, M. (2020). Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *The BMJ*, *370*, 1–10. https://doi.org/10.1136/bmj.m2743

Jabal, K. A., Ben-Amram, H., Beiruti, K., Batheesh, Y., Sussan, C., Zarka, S., & Edelstein, M. (2021). Impact of age, ethnicity, sex and prior infection status on immunogenicity following a single dose of the BNT162b2 MRNA COVID-19 vaccine: Real-world evidence from healthcare workers, Israel, December 2020 to January 2021. *Eurosurveillance*, *26*(6), 1–5. https://doi.org/10.2807/1560-7917.ES.2021.26.6.2100096

Jardine, R., Wright, J., Samad, Z., & Bhutta, Z. A. (2020). Analysis of covid-19 burden, epidemiology and mitigation strategies in muslim majority countries. *Eastern Mediterranean Health Journal*, *26*(10), 1173–1183. https://doi.org/10.26719/emhj.20.120

Jefferies, S., French, N., Gilkison, C., Graham, G., Hope, V., Marshall, J., McElnay, C., McNeill, A., Muellner, P., Paine, S., Prasad, N., Scott, J., Sherwood, J., Yang, L., & Priest, P. (2020). COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study. *The Lancet Public Health*, *5*(11), e612–e623. https://doi.org/10.1016/S2468-2667(20)30225-5

Koh, W. C., Naing, L., & Wong, J. (2020). Estimating the impact of physical distancing measures in containing COVID-19: An empirical analysis. *International Journal of Infectious Diseases*, *100*, 42–49. https://doi.org/10.1016/j.ijid.2020.08.026

Kozlakidis, Z., Ribeiro, S. P., Dutta, P. S., Kaur, T., Sarkar, S., Chowdhury, S., Sinha, S. K., & Jolly, M. K. (2020). Anticipating the Novel Coronavirus Disease (COVID-19) Pandemic. *Frontiers in Public Health | Www.Frontiersin.Org*, *8*, 569669. https://doi.org/10.3389/fpubh.2020.569669

Kumar, A. S., Yazir, P., & Gopika, G. G. (2019). Consumption inequality in India after liberalization: A caste based assessment. *The Singapore Economic Review*, *64*(01), 139–155.

Laarne, P., Zaidan, M. A., & Nieminen, T. (2021). ennemi: Non-linear correlation detection with mutual information. *SoftwareX*, *14*, 100686. https://doi.org/10.1016/j.softx.2021.100686

Lancet. (2020). India under COVID-19 lockdown. *Lancet (London, England)*, *395*(10233), 1315. https://doi.org/10.1016/S0140-6736(20)30938-7

Laroze, D., Neumayer, E., & Plümper, T. (2021). COVID-19 does not stop at open borders: spatial contagion among local authority districts during England's first wave. *Social Science & Medicine, 270*, 113655.

Leshem, E., Afek, A., & Kreiss, Y. (2020). Buying Time with COVID-19 Outbreak Response, Israel. *Emerging Infectious Diseases*, *26*(9), 2251–2253. https://doi.org/10.3201/eid2609.201476

Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., Nair, H., Network, U., & Evidence, C.-. (2020). The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2 : a modelling study across 131 countries. *The Lancet Infectious Diseases*, *21*(2), 193–202. https://doi.org/10.1016/S1473-3099(20)30785-4

Liu, Y., Morgenstern, C., Kelly, J., Lowe, R., & Jit, M. (2020). The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Medicine*, 1–12. https://doi.org/10.1101/2020.08.11.20172643

Lu, J. G., Jin, P., & English, A. S. (2021). Collectivism predicts mask use during COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(23). https://doi.org/10.1073/pnas.2021793118

Luu, T., & Huynh, D. (2020). Does culture matter social distancing under the COVID-19 pandemic? *Safety Science*, *130*, 104872. https://doi.org/https://doi.org/10.1016/j.ssci.2020.104872

Maor, M., Sulitzeanu-Kenan, R., & Chinitz, D. (2020). When COVID-19, constitutional crisis, and political deadlock meet: the Israeli case from a disproportionate policy perspective. *Policy and Society*, *39*(3), 442–457. https://doi.org/10.1080/14494035.2020.1783792

Marson, F. A. L., & Ortega, M. M. (2020). COVID-19 in Brazil. *Pulmonology*, *26*(4), 241–244. https://doi.org/10.1016/j.pulmoe.2020.04.008

Mendes-Moreira, J., Soares, C., Jorge, A. M., & De Sousa, J. F. (2012). Ensemble approaches for regression: A survey. *ACM Computing Surveys*, *45*(1). https://doi.org/10.1145/2379776.2379786

Moro, S. (2015). *Feature Selection Strategies for Improving Data-Driven Decision Support in Bank Telemarketing*. ISCTE-Instituto Universitario de Lisboa (Portugal).

Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the CRISP-DM methodology. *ESM 2011 - 2011 European Simulation and Modelling Conference: Modelling and Simulation 2011*, *Figure 1*, 117–121.

Neil M, F., Daniel, L., Gemma, N.-G., Natsuko, I., Ainslie, K., Marc, B., Sangeeta, B., Adhiratha, B., Zulma, C., Gina, C.-D., Amy, D., Ilaria, D., Han, F., Katy, G., Will, G., Arran, H., Wes, H., Okell, L. C., Sabine, van E., … Azra C, G. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College COVID-19 Response Team, March*. https://doi.org/https://doi.org/10.25561/77482

OurWorldInData. (2021). *Coronavirus (COVID-19) Vaccinations*. https://ourworldindata.org/covid-vaccinations

Oyebode, O., Ramsay, S. E., & Brayne, C. (2021). Public health research in the UK to understand and mitigate the impact of COVID-19 and COVID-19 response measures. *Journal of Epidemiology and Community Health*, *75*(3), 209–212. https://doi.org/10.1136/jech-2020-214997

Pan, A., Liu, L., Wang, C., Guo, H., Hao, ; Xingjie, Wang, Q., Huang, J., He, N., Yu, H., Lin, ; Xihong, Wei, S., Wu, T., & Lin, X. (2020). Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China Supplemental content. *JAMA*, *323*, 1915–1923. https://doi.org/10.1001/jama.2020.6130

Perra, N. (2021). Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports*, *913*, 1–52. https://doi.org/10.1016/j.physrep.2021.02.001

Prats-Uribe, A., Paredes, R., & Prieto-Alhambra, D. (2020). Ethnicity, comorbidity, socioeconomic status, and their associations with COVID-19 infection in England: A cohort analysis of UK Biobank data. *MedRxiv*. https://doi.org/10.1101/2020.05.06.20092676

Rafaell, R. de M. R., Neto, M., Carvalho, M. M. B. de, David, H. M. S. L., Acioli, S., & Faria, M. G. de A. (2020). Epidemiologia, políticas públicas e pandemia de Covid-19: o que esperar no Brasil? *Rev. Enferm. UERJ*, 2–7.

Raina, C., Nguyen, P., Ahmad, A., Trent, M., Gerber, B., Steinhofel, K., & Seale, H. (2021). International Journal of Infectious Diseases Mask use , risk-mitigation behaviours and pandemic fatigue during the COVID-19 pandemic in fi ve cities in Australia , the UK and USA : A cross-sectional survey. *International Journal of Infectious Diseases*, *106*(June 2020), 199–207. https://doi.org/10.1016/j.ijid.2021.03.056

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, *Icssit*, 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Ray, D. et. al. (2020). Predictions, Role of Interventions and Effects of a Historic National Lockdown in India’s Response to the the COVID-19 Pandemic: Data Science Call to Arms. *Harvard Data Science Review*, *2020*(Suppl 1), 1–45. https://doi.org/10.1162/99608f92.60e08ed5

Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Mohammed, S., Purkayastha, S., Halder, A., Rix, A., Barker, D., Kleinsasser, M., Zhou, Y., Song, P., Bose, D., Banerjee, M., Baladandayuthapani, V., Ghosh, P., & Mukherjee, B. (2020). Predictions, role of interventions and effects of a historic national lockdown in India’s response to the COVID-19 pandemic: Data science call to arms. In *medRxiv*. https://doi.org/10.1101/2020.04.15.20067256

Russell, T. W., Wu, J. T., Clifford, S., Edmunds, W. J., Kucharski, A. J., & Jit, M. (2021). Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *The Lancet Public Health*, *6*(1), e12–e20. https://doi.org/10.1016/S2468-2667(20)30263-2

Saban, M., Myers, V., Shachar, T., Miron, O., & Wilf-Miron, R. R. (2021). Effect of Socioeconomic and Ethnic Characteristics on COVID-19 Infection: the Case of the Ultra-Orthodox and the Arab Communities in Israel. *Journal of Racial and Ethnic Health Disparities*. https://doi.org/10.1007/s40615-021-00991-z

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation.

Anesthesia and Analgesia, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Silva, C. J., Cruz, C., Torres, D. F. M., Muñuzuri, A. P., Carballosa, A., Area, I., Nieto, J. J., Fonseca-Pinto, R., Passadouro, R., Santos, E. S. dos, Abreu, W., & Mira, J. (2021). Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-83075-6

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). In *International Journal of Surgery* (Vol. 76, pp. 71–76). Elsevier Ltd. https://doi.org/10.1016/j.ijsu.2020.02.034

Sorjamaa, A., Hao, J., & Lendasse, A. (2005). Mutual information and k-Nearest Neighbors approximator for time series prediction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3697 LNCS*, 553–558. https://doi.org/10.1007/11550907_87

Steurer, M., Hill, R. J., & Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*, *38*(2), 99–129. https://doi.org/10.1080/09599916.2020.1858937

Stuenkel, O. (2020). *The BRICS and the future of global order*. Lexington Books.

Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U. G., Li, B., Cai, J., Xu, B., & Yang, Q. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, *368*(6491), 638–642.

van der Westhuizen, H.-M., Kotze, K., Tonkin-Crine, S., Gobat, N., & Greenhalgh, T. (2020). Face coverings for covid-19: from medical intervention to social practice. *Bmj*, *370*.

Waitzberg, R., Davidovitch, N., Leibner, G., Penn, N., & Brammli-Greenberg, S. (2020). Israel's response to the COVID-19 pandemic: tailoring measures for vulnerable cultural minority populations. *International Journal for Equity in Health*, *19*(1), 1–5.

Wang, Y. (2021). Government policies, national culture and social distancing during the first wave of the COVID-19 pandemic: International evidence. *Safety Science*, *135*(October 2020), 105138. https://doi.org/10.1016/j.ssci.2020.105138

White, E. R., & Hébert-Dufresne, L. (2020). State-level variation of initial COVID-19 dynamics in the United States. *PLoS ONE*, *15*(10 OCTOBER), 1–13. https://doi.org/10.1371/journal.pone.0240648

Wirth, R., & Hipp, J. (2000). *CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39. 24959*.

Wong, M. C. S., Huang, J., Teoh, J., & Wong, S. H. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: An analysis of 139 countries. *The Journal of Infection*, *81*(3), e70.

Yechezkel, M., Weiss, A., Rejwan, I., Shahmoon, E., Ben-Gal, S., & Yamin, D. (2021). Human mobility and poverty as key drivers of COVID-19 transmission and control. *BMC Public Health*, *21*(1), 1–13. https://doi.org/10.1186/s12889-021-10561-x

Yeung, A. Y. S., Roewer-Despres, F., Rosella, L., & Rudzicz, F. (2021). Machine Learning–Based

Prediction of Growth in Confirmed COVID-19 Infection Cases in 114 Countries Using Metrics of Nonpharmaceutical Interventions and Cultural Dimensions: Model Development and Validation. *J Med Internet Res*, *23*(4), e26628. https://doi.org/10.2196/26628

Yu, X., Qi, G., & Hu, J. (2021). Analysis of second outbreak of COVID-19 after relaxation of control measures in India. *Nonlinear Dynamics*, *106*(2), 1149–1167.