March 2020: 31 days that will reshape tourism

Nuno António & Paulo Rita
NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisbon, Portugal CITUR – Centro de Investigação, Desenvolvimento e Inovação em Turismo, Universidade do Algarve, Faro, Portugal

This is the Accepted Author Manuscript of the article published by Taylor & Francis in *CURRENT ISSUES IN TOURISM**:***

António, N., & Rita, P. (2020). March 2020: 31 days that will reshape tourism. Current Issues in Tourism. [Advanced online publication on 25 december, 2020]. https://doi.org/10.1080/13683500.2020.1863927

March 2020: 31 days that will reshape tourism

Nuno António [a, b]* and Paulo Rita [a] *

[a]*NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal*

[b]*CITUR – Centro de Investigação, Desenvolvimento e Inovação em Turismo, Universidade do Algarve, Faro, Portugal;*

* nantonio@novaims.unl.pt * prita@novaims.unl.pt

NOVA Information Management School (NOVA IMS)

Universidade Nova de Lisboa

Campus de Campolide

1070-312, Lisbon

Portugal

March 2020: 31 days that will reshape tourism

Lessons from previous endemics/pandemics show which type/timing of public health measures had a significant influence on the impact of diseases. However, those show that public health measures and travel restrictions represent a significant burden on countries´ economies, especially in the tourism industry. This study aims to investigate whether a country's dependence on tourism might influence the time/nature of pandemic mitigation measures and the impact of the pandemic on tourism, particularly in the hospitality sector. To achieve a comprehensive/multidimensional perspective, twelve European countries were studied based on the collection of data from six different sources: cases/deaths caused by the disease, economic indicators, public health measures, rooms supply/demand, reservation/cancellation rates, demographic and healthcare system characteristics. Using data science techniques/methods allowed to verify that the dependence of some countries on tourism did not make them to have a different behaviour in terms of the application of measures. Despite the differences in the timings/types of measures implemented, tourism was highly affected in all countries.

Keywords: covid-19; pandemics; epidemics; crisis; data science; hospitality; tourism

Introduction

On 11th of March 2020, more than 118,000 cases and 4,291 deaths related to COVID-19 had already been reported in 114 countries, leading the WHO to recognise that the disease was no longer occurring only at the level of a region or community (epidemic), but at a global level (pandemic) (World Health Organization, 2020).

Due to the acknowledged lack of studies focusing on the economic effects of pandemics, in particular of those taking a transnational perspective (Gössling et al., 2020; Novelli et al., 2018), this paper takes advantage of Data Science methods and multiple data sources to contribute to fulfilling this gap. This research studies the impact of COVID-19 on twelve European countries and how it affected the tourism industry, in particular the hospitality sector. The reasons for focusing this study in hospitality have to do with the relevance of this sector in the tourism industry (Eurostat, 2020), and also because in hospitality performance data is available in a timeliness and relevant way.

This study is centred on March 2020, the month where most European countries officially recognized COVID-19 as a threat to public health and the economy. This research aims to interpret what really happened on March 2020 in order to draw valuable lessons regarding the balance between advantages and disadvantages of a country's economy exposure to tourism and how countries should react to health-related crises, considering the impact that mitigation measures can have on tourism. To achieve the previously described objective, this study addressed two main goals. First, to understand if the countries' dependency on tourism may have influenced both the timing and nature of the pandemic mitigation measures. Second, to understand the impact of the pandemic on tourism, and in particular, on the hospitality industry.

Methods and data

Data from multiple sources were used to obtain a multidimensional understanding of the pandemic impact on the hospitality sector and to derive knowledge from that understanding. The following sub-sections detail the data used and how the research was executed.

Research background and data

Several authors have studied the impact of economic crises on tourism (e.g., Ritchie et al., 2010; Song & Lin, 2010). However, except from new studies involving COVID-19, not many have analysed the impact of health-related crises on the tourism industry or tourism sectors (Novelli et al., 2018). Among the new hospitality related COVID-19-centered studies, the focus seems to be on the impact on hotel operations, in particular how hotels were affected and what will change in the future in hotel operations (Anguera-Torrell et al., 2020; Hao et al., 2020; Lai & Wong, 2020).

From the authors' best knowledge, the few studies found on the impact of health-related crises on tourism made use of multiple datasets to understand the phenomena. Although some studies shared the origin of some data sources, such as using WHO health related data, most differed on the health-related data to the region under research (Gössling et al., 2020; Haque & Haque, 2018; Novelli et al., 2018; Page et al., 2012).

In order to obtain a broader perspective of the impact of COVID-19 on the hospitality sector, and based on the type of data sources included in the topic's literature, it was decided to use a mix of public and non-public datasets from six sources:

- ACAPS (ACAPS, 2020): public dataset with country, daily details of COVID-19 key measures and events (social distancing, movement restrictions, public health, social and economic, and lockdowns). ACAPS is a non-profit, non-governmental project with the aim of providing independent information for humanitarian analysis;
- D-EDGE (D-EDGE, 2020): non-public dataset with hotel bookings daily data (new and canceled bookings), country, and distribution channel. D-EDGE is the European number 1 and world number 3 in hotel distribution technology. The D-EDGE platform is used by more than 10,000 properties to distribute their inventory across multiple channels;
- European Centre for Disease Prevention and Control (ECDC, 2020): public dataset with worldwide, daily statistics on COVID-19 cases and deaths, including the total population per country in 2019. ECDC is the European

agency responsible for strengthening Europe's defenses against infectious diseases;

- Our World in Data (OWID) (Our World in Data, 2020b): public dataset with worldwide, daily statistics on COVID-19, including tests, cases, and deaths, but also countries' demographic, economic, and health indicators. The data is compiled from various sources. OWID is an organization based at the University of Oxford with a mission to present "research and data to make progress against the world's largest problems" (Our World in Data, 2020a);
- STR (STR, 2020): non-public dataset with daily hotel performance data per country. STR is the global leader in providing data for benchmarking, analytics and marketplace insights for the hospitality sector;
- World Travel & Tourism Council (WTTC) (World Travel & Tourism Council, 2020a): non-public dataset with 2019 economic performance data by country. WTTC is an organization that represents the private sector of travel and tourism.

Considering the previously mentioned lack of studies on the impact of health-related crises on multiple regions, the current study analysed its effect on European countries. In 2019 Europe continued to be the largest region in the world in terms of international visitor spending (US$619 billion), a value that accounted for 37% of the global tourist expenditure. In Europe, travel and tourism generated a revenue of US$2.0 trillion and represented 9.1% of the Gross Domestic Product (GDP) (World Travel & Tourism Council, 2020b). However, two obstacles existed in analysing the data for all European countries. The size of the task and the inexistence of data for several countries in some of the datasets. For these reasons, it was decided to select 12 countries that could represent different regions in Europe (north, south, west and east) as well as where tourism had different shares of contribution to the economy (as seen in Figure 1).

Consequently, the 12 European countries selected for this study were (in alphabetic order): Austria (AUT), Belgium (BEL), Czech Republic (CZE), France (FRA), Germany (DEU), Italy (ITA), Ireland (IRL), Poland (POL), Portugal (PRT), Spain (ESP), Switzerland (CHE), and the United Kingdom (GBR/UK). Only two of these countries, Switzerland and the United Kingdom, are not members of the European Union. Regarding the contribution to GDP, it is clear the existence of a higher dependency on tourism by four countries, namely Portugal, Spain, Italy and Austria, showing between 12% and 16%, with the two Iberian countries (PRT and ESP) also exhibiting quite high contribution to exports. Five countries present a medium level (between 6% and 9%), specifically UK, Germany, France, Switzerland and Czech Republic. Lower tourism contributions (between 4% and 5%) are portrayed by the remaining three countries, explicitly Poland, Ireland and Belgium.

<Insert figure 1 here>

Research process

Due to the use of multiple datasets and the empirical character of this research, the process to conduct the study was based on the CRoss-Industry Process Model for Data Mining (CRISP-DM) (Chapman et al., 2000).

The Python programming language and its wide-range of standard analytics and scientific packages were used, including standard packages such as pandas, numpy, scikit-learn, scipy, matplotlib, plotly, and seaborn. Other not so standard packages were also used, like the DTAIDistance package (Meert, 2019). DTAIDistance implements the Dynamic Time Warping (DTW), an algorithm used to measure the similarity of time sequences, i.e., if the time series are aligned, if time series are similar, but slightly shifted, or if time series have a different shape, but with a different scale.

Since not all providers make the data available in an adequate format for analytical tools, manual collection and pre-processing were performed. The final datasets covered 89 variables, more than 3.8 million observations, 24 million cells, and 1.3 Giga bytes of data. The description (dictionary) of the public datasets is provided in the respective references. The authors are available to describe the non-public datasets (D-EDGE and STR) under consultation.

After the initial exploration of data in the data understanding phase, data was prepared for modelling. The preparation involved reformatting data, removing out-of-scope data, aggregating data, merging data, and generating new features. Reformatting data was essentially the transformation of string variables as categorical or renaming variables. The removal of out-of-scope data was done in the D-EDGE, ECDC, OWID, and STR, as these datasets included data from countries or dates outside the scope of the study. Aggregation was required in the D-EDGE dataset since the base unit of analysis is the distribution channel. Merging datasets was needed to create normalized key variables in all datasets. Merging was also needed to create multidimensional analysis, such as analysing booking performance or room occupation performance and implementing measures in time. The generation or derivation of new features from the existing variables is a fundamental component in the development of any analytical model, due to the improvement in information power they can add to models. New features were added in two datasets, D-EDGE and ECDC. The feature *BookingIndex* was added in the D-EDGE dataset. This new feature represents the proportion of new bookings versus cancelled bookings. The feature was calculated with the formula:

$$x = \begin{cases} \frac{New\ Bookings}{Canceled\ Bookings}, & New\ Bookings \geq Canceled\ Bookings \\ \frac{-Canceled\ Bookings}{New\ Bookings}, & Canceled\ Bookings > New\ Bookings \end{cases}$$

Therefore, the new feature is positive if the number of new bookings is equal or superior to the number of cancellations; alternatively, it is negative if the number of cancellations is higher than the number of new bookings. The booking index is calculated based on the date of the creation of new bookings and the cancellation date, in the case of cancellations. The arrival date is not taken into consideration. In the ECDC dataset, several features were created: *cumCases* and *cumDeaths*, to store the cumulative cases and deaths per country; *cumCasesPer100K* and *cumDeathsPer100K*, to save the cumulative cases and deaths per 100,000 of the population of the country; *casesPer100K* and *deathsPer100K*, to store the cases and deaths per 100,000 of the population of the country. The reason to create features normalized by 100,000 of the population was to facilitate cross-country comparisons. Otherwise, comparisons between countries with very different population sizes would not be accurate (ECDC, 2020; Johns Hopkins University of Medicine, 2020).

The analyses were done using Data Science techniques (Rita et al., 2018), including Data Mining and statistical methods, correlation analysis, DTW measurement, k-means clustering, and hierarchical clustering. Data visualization methods were primarily employed, as data visualization enables the display of a large amount of information in small spaces, making complex information easier to interpret and compare, and encourages the detection of patterns, anomalies, and tendencies (Few, 2013). The achieved results are presented and discussed in the following section.

Results and discussion

As presented in Figure 2, the pandemic's progression during March 2020 differed substantially between countries, both in cumulative number of cases and in mortality. In cumulative number of cases per 100,000 of the population, Switzerland (CHE) overtakes Italy and clusters also with Spain in the top three countries, but in terms of deaths, unlike those two, CHE comes down significantly. In the case of the cumulative number of deaths, it is possible to verify that there seems to be a clear pattern, separating Italy and Spain from the remaining countries. This pattern is even more evident when the daily mortality by country is compared using DTW. As presented in Figure 3, there are two base clusters in terms of mortality: one with Italy and Spain, and another with the remaining countries. In the latter cluster, it is possible to identify two sub-clusters, one with more impacted countries: Belgium, France, UK, and Switzerland, and one with countries where the pandemic had a less dramatic impact in terms of mortality: Austria, Ireland, Portugal, Czech Republic, Poland, and Germany.

<Insert figure 2 here>

<Insert figure 3 here>

The difference in cases and deaths by 100,000 of the population may be explained by factors such as the differences in the proportion of the population tested, demographics (e.g., older population), or the characteristics of the healthcare system (Johns Hopkins University of Medicine, 2020).

Although the 12 studied countries share numerous values and characteristics, there are differences between demographic and health-related characteristics. K-Means clustering was employed to study the similarity between countries. The elbow and the silhouette methods were used to define K (the number of clusters). While the elbow method pointed to a value of 3 or 4 groups, the silhouette method undoubtedly pointed to 4. The analysis, per cluster, of each variable mean, as shown in Table 1, allowed for the qualitative classification of countries as follows:

- Cluster 0 (Austria and Germany): Germanic countries. Countries better prepared in terms of hospital beds per 100,000 of the population. Although not the oldest cluster, it is the second cluster in terms of the percentage of the population with over 65 years old and median age.

- Cluster 1 (Italy, Portugal, and Spain): Southern countries. Countries worse prepared in terms of hospital beds per 100,000 of the population (less than half of cluster 0). However, these countries have a higher life expectancy, which may explain the higher median age and percentage of the population over 65. This cluster also presents one of the smallest averages of cardiovascular death rates.
- Cluster 2 (Belgium, France, Ireland, Switzerland, and the UK): Western and Central European Countries. Countries with the youngest population have the best indicators in terms of cardiovascular death rate and diabetes prevalence, but not so good indicators in terms of hospital beds per 100,000 of the population.
- Cluster 3 (the Czech Republic and Poland): Eastern European countries. In terms of age, very similar to cluster 2 countries. Though having the second-best number of hospital beds per 100,000 of the population, the cluster has a higher prevalence of diabetes, cardiovascular death rate, and the shortest life expectancy.

As depicted in Figure 2, although some countries started to apply measures with an impact on tourism at the beginning of March, the vast majority of the measures was implemented between 9th and 24th of March, with a peak on the 16th of March. Although tourism is highly vulnerable to the mobility restrictions and social distance imposed by the implemented measures to mitigate the pandemic effects (Anderson et al., 2020; Gössling et al., 2020), it seems that the weight of tourism in the country's economy (see Figure 1), was not the only factor influencing the timing and type of measures implement by country. For example, Italy, Portugal, Austria and Spain, the countries where tourism has a higher contribution to the national GDP, closed restaurants, on

11th, 14th and 16th (the last two) of March, respectively, which does not differ much from other countries where tourism is not so relevant in the economy, such as Germany or Belgium, who closed restaurants, respectively on 12th and 13th of March. Other countries had a different approach on how to deal with the impact of the pandemic, such as the UK and Ireland, which closed restaurants only later on 21st and 24th of March. Moreover, it seems that most countries did not want to close hotels, a sector vital to tourism. Indeed, solely five of the twelve studied countries closed hotels. Besides the Czech Republic, Ireland, and Poland, which closed hotels on 16th, 24th and 31st of March, two other countries where tourism has an essential contribution to the GDP, did close hotels, Austria and Spain. However, whereas Austria closed hotels on 16th of March 16 only in a region where an outbreak occurred, Spain closed all hotels on 24th of March.

Along with the rapid spread of the disease during the month of March, another factor that seems to have influenced the timing and type of measures implemented by each country was the characteristics of the country's health system. Being COVID-19 a disease that can lead to hospitalization in intensive care units (ICU), it was likely that lesser prepared European countries such as Portugal (with 4.2 ICU beds per 100,000 of the population) took a different approach when compared to better-prepared countries, such as Germany (with 29.2 ICU beds per 100,000 of the population (Rhodes et al., 2012). For example, Spanish authorities alleged that the reason for closing hotels was to contain the pandemic and, at the same time, to fit some hotels with medical equipment and to use them in the park phase (Hospitality Net, 2020).

As illustrated in Figure 4, in the beginning of March the pandemic effect was already having a negative impact on occupation rates of hotels all over Europe. If in some countries, like Belgium, Ireland, Portugal, or the UK, until the second weekend of

March (8th of March was a Sunday), the drop in the occupation rate was lighter, in others, such as Switzerland and Italy, the decline was already more substantial. In the case of Italy, the first European country where COVID-19 cases were reported, on 30th of January (Corriere della Sera, 2020), this drop in the hotels' occupation rate is explained by the number of cases in the country and the lockdown measures imposed in the northern region already on 8th of March. In the beginning of the week starting on 9th of March, occupation rates presented a pattern similar to the previous year, with occupation rates increasing from the weekend to weekdays. As events unfold, with the number of cases rapidly growing in Italy, travel restrictions started to be implemented. On 16th of March, the European Commission recommended EU member states to restrict non-essential travel (European Comission, 2020a). From then on, occupation rates plunged. This inverse relationship between travel restrictions and hotel occupation rates is even more evident when the Stringency index is analysed against the room occupation rate. The Stringency index is a measure created by researchers from the Blavatnik School of Government (University of Oxford) that measures government policies' number and strictness (Blavatnik School of Government - University of Oxford, 2020).

<Insert figure 4 here>

Even though only five countries required hotels to close down and two of them only required that in specific regions, the room occupation rate dropped to below 10% in most countries in the last two weeks. Only in Ireland and the UK did the occupation rate not fall to such low values, but the decrease left occupation rates between 10% and 20%. Notwithstanding, if the supply number of rooms present in STR's daily samples are taken into consideration, as shown in Figure 5, it is possible to see that the supply of rooms decreased sharply from 16th of March in most countries. The exceptions were Italy and the UK. In Italy, hotels started earlier to close down, with the increase in the number of cases and the first lockdown measures. In the UK, it was not until the general lockdown imposition on 24th of March that hotels started to close down in large numbers. As detailed in Table 2, the drop in the number of rooms in STR's samples, which can be seen as a proxy for the number of hotels operating during the month of March, suffered a negative variation from 60.7% in Germany to mesmerising 98.2% in Spain. If these hotels had not closed down, the room occupation rate would have been even much lower than it was. Overall and on average the supply of rooms in these twelve European countries was down a staggering -79% by the end of March vis-à-vis the beginning of the month.

<Insert figure 5 here>

<Insert table 2 here>

As expected, travel restrictions did not impact only the March performance of hotels, but also their results in the following months. As detailed in Figure 6, the pandemic situation generated a massive number of cancellations. During March 2019, the daily mean booking index, in all countries, oscillated between 5 and 10. These values indicate that the number of new bookings was superior in 5 to 10 times the number of cancellations. However, one year later, in March 2020 the situation changed dramatically. From the beginning of the month, the booking index was already lower than in the previous year. Despite being positive in most countries during the first week of that month, meaning that the number of new bookings was still superior to cancellations, as travel restrictions and other measures were implemented, the situation inverted. After 9th of March, the number of cancellations received by hotels was higher than the number of new reservations in most countries. The fact that cancellations increased, or new bookings decreased, even before the significant travel restrictions applied on 16th of March, can be explained by the pandemic declaration. In fact, from the moment the WHO declared the pandemic, on 11th of March, many travellers were left without travel insurance, as most insurance policies do not cover pandemic situations (Elliott, 2020). Therefore, it is understood that many travellers decided to cancel their reservations or at least postpone them, making new later bookings since 11th of March.

In Austria and Italy, the number of cancellations in a day reached over 30 times the number of new bookings. The peaks in cancellations appear to be associated to the application of travel-related restrictions, as most peaks coincided or happened immediately after those types of measures were implemented. In Austria and Italy, the more prominent peaks appear to be related to the imposition of stricter border controls, respectively, on 19th and 28th of March.

<Insert figure 6 here>

What looks somewhat surprising is the relation between cancellations and the impact of the disease in different countries. It seems there is not a higher correlation as one would expect. A hierarchical cluster by the daily booking index value of all the countries (see Figure 7) showed that countries where the disease caused a high number of cases and mortality, such as Spain, were not distant from countries with a much lower number of cases and mortality, like Portugal. This situation could be attributed to the fact that ten EU countries, including France, Italy, Poland, and Portugal, announced or published new legislation to try to diminish cancellations, going against the EU law to protect consumers and travellers' rights. These countries are now facing a EU infringement package (European Commission, 2020b).

When analysed in the form of a hierarchical clustering (Figure 7), hotel booking cancellations are grouped into three major clusters: one with Belgium, a second one with Italy and Austria, and another one with the remaining studied countries. On the one hand, these results seem to indicate that in countries where tourism has a higher contribution to GDP and exports, there were more cancellations, which is the case of Italy and Austria. On the other hand, that is not the case for Belgium or even the UK, which was not too distant from Italy and Austria. In fact, the hierarchical clustering confirms the finding that despite implementing travel restrictions later, the UK suffered hotel cancellations in a pattern not much different from other countries.

The pattern of cancellations vs. cumulative deaths vs. room occupation rate during March can be seen in the interactive animation in the supplementary file InteractivePlot1.html. In this animation, the room occupation rate is represented by the dimension of the data point. Furthermore, it is possible to see that in general, excluding Spain and Italy, there were not very big differences on the evolution of the disease over the month of March, despite having an impact on health and in the hospitality sector different between countries, with some more affected than others.

<Insert figure 7 here>

Hierarchical clustering by the end of March values (cumulative cases per 100,000 of the population, cumulative deaths per 100,000 of the population, rooms occupation rate, booking index, and stringency index), as shown in Figure 8, do not show a clear relation between tourism contribution to the country's economy, disease impact, measures stinginess, and the impact on the hospitality sector. More specifically, two main clusters were found. One with two countries, Spain and Italy, where travel and tourism contributes significantly to the economy and that belong to cluster 1 in terms of demographics and healthcare system characteristics (see table 1). In the other main cluster, with the remaining countries, it is possible to see that the different levels were composed of countries where tourism had very different contributions to the country's GDP and with different demographics and healthcare system characteristics, such as Portugal and Belgium, or Austria and Ireland.

<Insert figure 8 here>

However, a correlation analysis of the end of March values with the tourism contribution to economy measures per country, presented in Figure 9, revealed some interesting findings:

(1) There is a high positive correlation between travel and tourism contribution to the country's GDP and the contribution to exports (0.87). The fact that these values did not present an even higher correlation, like the correlation between the contribution of travel and tourism to the country's GDP and the contribution to jobs (0.98), shows that in some countries domestic tourism has a higher degree of importance than in others. These countries, such as Germany, where domestic tourism has a higher weight taking into consideration the contribution of tourism to the GDP, have higher probabilities of being less affected by the tourism crisis caused by the pandemic. Nevertheless, as it happened in the UK, with the swine flu in 2009 (Page et al., 2012), domestic tourism is insufficient to avoid significant losses on revenue. As such, even countries not so dependent on international tourism need to take into serious consideration the negative impact of the restriction measures on the tourism industry.

(2) There is a high positive correlation between the number of cases and the number of deaths per 100,000 of the population (0.75). Nevertheless, the fact that the correlation is not higher can be due to the differences between demographic and healthcare system characteristics (as shown in Table 1), as suggested by the Johns Hopkins University of Medicine (2020).

(3) There is a moderate positive correlation between the disease impact (cases and deaths per 100,000 of the population) and the contribution of travel and tourism to the country's GDP (respectively 0.45 and 0.48). Additionally, due to the expected high correlation between a country's travel and tourism contribution to GDP and the contribution to the number of jobs and exports, disease impact is also moderately correlated to those two economic indicators. On the one hand, this correlation between the disease impact and the contribution to GDP suggests that countries with more tourism tend to have more cases and deaths, which is expected, due to the higher exposure to people from other regions. On the other hand, the correlation between the disease impacts and the contribution to exports (a proxy for the weight of international tourism) is smaller, suggesting the opposite. These contradictions reinforce what Chinazzi et al. (2020) stated that if not combined with public health measures and behavioral changes, travel limitations exert a moderate effect in reducing the spread of the disease. Travel limitations by themselves will not prevent the pandemic, as such, policy makers should focus in inducing behavioral changes.

(4) There is a high/very high positive correlation between the occupancy rate of rooms on March and the booking index (0.89). This correlation indicates that countries with fewer cancellations had higher room occupancy rates, which makes sense for March. However, since cancellations were not only for March

arrivals but for future months, it seems to indicate that countries which maintained a certain occupation level were less impacted with cancellations. In March, the country with a higher occupancy rate was the UK, with a room occupancy rate of 39%. From the studied countries, the UK was the only where the monthly booking index was positive (1.08), meaning that new bookings were superior 1.08 times cancellations. That metric was negative in all the other countries. Once more, results confirm we are living in a global, connected world. Even though the measures taken by each country differed in terms of rigidness and timing, all countries were affected.

(5) There is a moderate correlation between the stringency index, the booking index and the rooms' occupancy rate. This correlation is positive with the deaths per 100,000 of the population (0.51), but negative with booking index and the rooms occupancy rate (respectively -0.62 and -0.48). These correlations indicate that, to some extent, the measures were stricter in countries with a higher number of deaths. It also stated that in these countries, when the number of deaths tends to increase, cancellations tend to grow and rooms' occupancy rate decreases. This relation seems to indicate that the number of deaths could be seen as a proxy to measure both policy makers and hotel customers perception of the pandemic level.

(6) There is a weak positive correlation between the stringency index and the measures related to the travel and tourism contribution to the country’s economy (GDP, jobs, and export), respectively 0.25, 0.16, and 0.26. This weak correlation seems to indicate that there was not a clear relationship between the importance of tourism to the country’s economy and the severity of the measures implemented by the country. Thus, one could extrapolate that policy makers,

when defining mitigation measures, were above all concerned with the population health and not so much with the impact on tourism.

(7) There is a moderate negative correlation between the booking index and the disease impact (cases and deaths per 100,000 of the population), respectively, -0.49 and -0.52. A low/moderate negative correlation also exists between the room occupancy rate and the disease impact on the country (-0.37 and -0.46, respectively). This correlation suggests to some extent that the higher the number of cases and deaths is in a country, the higher the impact on the hospitality sector, with room occupation decreasing and cancellations increasing. This seems to be connected to point 5, and how the disease impact (cases and deaths) is used by hotel customers to measure the disease dissemination and spread. Thus, when the number of cases and deaths increase, tourism decreases and less rooms are occupied.

<Insert figure 9 here>

Conclusion

From a methodological point of view, this study presents several contributions to the tourism literature. First, it shows how Data Science methods, particularly data mining and data visualization, can be instrumental in transforming information to knowledge from tourism data. Second, it introduces a set of data sources, mostly public sources that other authors can use in future research on the impact of COVID-19 on tourism. Third, it presents the Booking index, a measure that can be used in hotels and other services which depends on reservations to calculate the proportion of new reservations vs. cancelled reservations in a given period. It is a measure that can be used, for example, in Revenue Management. The study's empirical results present a broader and multidimensional perspective of the impact of the pandemics on tourism, particularly in the hospitality sector, when on March 2020 more extreme measures had to be implemented by European countries, when Europe was the epicentre of COVID-19 across the world. Cases and deaths vary per country, depending on different factors affecting each country. Travel restrictions by itself are not sufficient to stop the spreading of the disease. It is essential that each country is aware of its demographic and public health system to define which public health-related measures to take and when. Tourism and the economy will suffer the impact, one way or the other. A more evident example was that almost all studied countries required restaurants to close, but not hotels. Nevertheless, hotels closed by themselves, and the ones that did not close had meagre occupancy rates. Although globalization and travel mobility facilitate tourism, it also facilitates virus spreading. So, even though some countries did not implement so strict public health-related measures and travel restrictions, they could not avoid a high impact on tourism. Therefore, in the presence of pandemics countries' governments priority should be the application of public health measures to mitigate the pandemic. This is because, even if a country decides to take a lighter approach, compared to other countries, its economy, and in particular tourism, will be affected the same way. Therefore, in future pandemics it is fundamental that policy makers of different countries work together with scientists and academics to define global mitigation strategies taking into account both health and economy protection measures.

This study also has managerial implications. First, it shows that even if a country is capable of mitigating the pandemic effects without applying stringent measures, the travel restrictions and measures imposed in other countries will impact tourism-related businesses, like hotels. Therefore, tourism-related businesses should have contingency plans for such occasions, such as special insurances or emergency funds. Second, since tourism, particularly international incoming tourism, is very important in many countries, tourism-business representative organizations should collaborate with the governments of those countries to promote the country as a safe destination, as soon as travel restrictions start to be lifted.

Like most studies, this study is not without limitations. First, data is not available for all European countries. For example, D-EDGE data did not have a significant sample for Greece and Sweden, countries that had opposite approaches on how to mitigate the pandemics. This limitation led to the exclusion of these two countries from the study. Second, some variables exist in more than one dataset with different values, depending on the source where the value was compiled from. For example, the countries' population was available both in the ECDC and in the OWID datasets. Being ECDC, an official source, the values considered in the study were from the ECDC. Third, this study focuses on European countries, which, although having different demographics and public healthcare systems, are better prepared for these types of crises than many other countries in the world. As such, conclusions from this study may not be applied to countries in different regions worldwide.

With the resume of the tourism activity, hotel managers are reporting: 1) changes in customer behaviour patterns, such as low demand for non-refundable rates or rates with rigid cancellation policies; 2) customers preference for horizontal properties, boutique hotels, private villas, and other small hotels; 3) extended stays; 4) shorter lead-time; 5) peaks on new bookings and cancellations, depending on travel restrictions lifting or implementation. Future studies on this topic of the COVID-19 should explore these changes in customer patterns and the impact of the COVID-19 in a broader period, for example, from March 2020 to December 2020. This study is necessary to understand if the countries' different timing of decisions and measures made any difference and to understand if these patterns will be temporary or permanent. Future studies should also address if any lessons learned from the 1st wave were implemented in the following waves. Namely, future studies should evaluate at what point did measures with direct impact on tourism, such as travel restrictions, restaurant opening restrictions or hotel closures were eased in the following waves due to their economic impact in the 1st wave. Future research could also compare different regions of the globe, for example, comparisons involving countries with a less rigid approach, such as the United States of America and Brazil.

Acknowledgements

The authors would like to thank D-EDGE and STR for allowing the authors to use their data in this research.

References

ACAPS. (2020). *COVID19 Government Measures Dataset*. ACAPS. https://www.acaps.org/covid19-government-measures-dataset

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, *395*(10228), 931–934. https://doi.org/10.1016/S0140-6736(20)30567-5

Anguera-Torrell, O., Aznar-Alarcón, J. P., & Vives-Perez, J. (2020). COVID-19: Hotel industry response to the pandemic evolution and to the public sector economic measures. *Tourism Recreation Research*, 1–10. https://doi.org/10.1080/02508281.2020.1826225

Blavatnik School of Government - University of Oxford. (2020). *Coronavirus Government Response Tracker*. University of Oxford.

https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. The Modeling Agency. https://the-modeling-agency.com/crisp-dm.pdf

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Piontti, A. P. y, Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400. https://doi.org/10.1126/science.aba9757

Corriere della Sera. (2020). *Coronavirus, primi due casi in Italia: Sono due turisti cinesi*. Corriere Della Sera. https://www.corriere.it/cronache/20_gennaio_30/coronavirus-italia-corona-9d6dc436-4343-11ea-bdc8-faf1f56f19b7.shtml

D-EDGE. (2020). *About us*. D-EDGE. https://www.d-edge.com/about-us/

ECDC. (2020). *COVID-19 situation update worldwide*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

Elliott, C. (2020). *What Travel Insurance Does Not Cover After The Pandemic*. Forbes. https://www.forbes.com/sites/christopherelliott/2020/06/07/what-travel-insurance-does-not-cover-after-the-pandemic/

European Comission. (2020a). *Communication from the comission to the european parliament, the european council and the council: COVID-19: Temporary Restriction on Non-Essential Travel to the EU*.

https://ec.europa.eu/transparency/regdoc/rep/1/2020/EN/COM-2020-115-F1-EN-MAIN-PART-1.PDF

European Commission. (2020b). *July infringements package: Key decisions*. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/inf_20_1212

Eurostat. (2020). *Tourism industries—Economic analysis—Statistics Explained*. Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_industries_-_economic_analysis#Analysis_by_subsectors

Few, S. (2013). Data visualization for human perception. In *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* (Second). Interaction Design Foundation. http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: A rapid assessment of COVID-19. *Journal of Sustainable Tourism*, *0*(0), 1–20. https://doi.org/10.1080/09669582.2020.1758708

Hao, F., Xiao, Q., & Chon, K. (2020). COVID-19 and China's Hotel Industry: Impacts, a Disaster Management Framework, and Post-Pandemic Agenda. *International Journal of Hospitality Management*, *90*, 102636. https://doi.org/10.1016/j.ijhm.2020.102636

Haque, T. H., & Haque, M. O. (2018). The swine flu and its impacts on tourism in Brunei. *Journal of Hospitality and Tourism Management*, *36*, 92–101. https://doi.org/10.1016/j.jhtm.2016.12.003

Hospitality Net. (2020). *Spain To Close All Hotels, Help Nursing Homes As Coronavirus Deaths Climb*. Hospitality Net. https://www.hospitalitynet.org/news/4097652.html

Johns Hopkins University of Medicine. (2020). *Mortality Analyses*. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/data/mortality

Lai, I. K. W., & Wong, J. W. C. (2020). Comparing crisis management practices in the hotel industry between initial and pandemic stages of COVID-19. *International Journal of Contemporary Hospitality Management*, *32*(10), 3135–3156. https://doi.org/10.1108/IJCHM-04-2020-0325

Meert, W. (2019). *DTAIDistance* (1.0.0) [Computer software]. https://dtaidistance.readthedocs.io/en/latest/index.html

Novelli, M., Gussing Burgess, L., Jones, A., & Ritchie, B. W. (2018). 'No Ebola…still doomed' – The Ebola-induced tourism crisis. *Annals of Tourism Research*, *70*, 76–87. https://doi.org/10.1016/j.annals.2018.03.006

Our World in Data. (2020a). *About*. Our World in Data. https://ourworldindata.org/about

Our World in Data. (2020b). *Data on COVID-19 (coronavirus) by Our World in Data*. GitHub. https://github.com/owid/covid-19-data

Page, S., Song, H., & Wu, D. C. (2012). Assessing the Impacts of the Global Economic Crisis and Swine Flu on Inbound Tourism Demand in the United Kingdom. *Journal of Travel Research*, *51*(2), 142–153. https://doi.org/10.1177/0047287511400754

Rhodes, A., Ferdinande, P., Flaatten, H., Guidet, B., Metnitz, P. G., & Moreno, R. P. (2012). The variability of critical care bed numbers in Europe. *Intensive Care Medicine*, *38*(10), 1647–1653. https://doi.org/10.1007/s00134-012-2627-8

Rita, P., Rita, N., & Oliveira, C. (2018). Data science for hospitality and tourism. *Worldwide Hospitality and Tourism Themes*. 10(6), 717-725. https://doi.org/10.1108/WHATT-07-2018-0050

Ritchie, J. R. B., Amaya Molinar, C. M., & Frechtling, D. C. (2010). Impacts of the World Recession and Economic Crisis on Tourism: North America. *Journal of Travel Research*, *49*(1), 5–15. https://doi.org/10.1177/0047287509353193

Song, H., & Lin, S. (2010). Impacts of the Financial and Economic Crisis on Tourism in Asia. *Journal of Travel Research*, *49*(1), 16–30. https://doi.org/10.1177/0047287509353190

STR. (2020). *About STR*. STR. https://str.com/about

World Health Organization. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020*. World Health Organization. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

World Travel & Tourism Council. (2020a). *About Us | World Travel & Tourism Council (WTTC)*. https://wttc.org/About/About-Us

World Travel & Tourism Council. (2020b). *Global Economic Impact & Trends 2020*. World Travel & Tourism Council. https://wttc.org/Research/Economic-Impact/moduleId/1445/itemId/91/controller/DownloadRequest/action/QuickDownload

Table 1. Means per variable, per cluster

Cluster	Countries	Population density	Median age	Aged 65 or older (%)	Cardiovascular death rate	Diabetes prevalence (%)	Hospital beds per 100,000 of the population	Life expectancy
0	2	171.883	45.5	20.33	150.661	7.33	7.685	81.44
1	3	137.112	46.5	21.32	113.465	7.27	3.180	83.04
2	5	211.031	41.3	17.83	109.859	4.44	4.330	82.34
3	2	130.602	42.6	17.90	227.408	6.37	6.625	79.06

Table 2. Rooms supply in STR's daily samples

Country	March 1	March 31	Variation (%)
Austria	31,062	6,927	-77.7
Belgium	23,766	5,196	-78.1
Czech Republic	20,127	3,005	-85.1
France	47,144	8,982	-80.9
Germany	259,743	102,161	-60.7
Ireland	33,503	6,187	-81.5
Italy	75,761	9,946	-86.9
Poland	38,439	10,109	-73.7
Portugal	27,531	3,552	-87.1
Spain	143,654	2,556	-98.2
Switzerland	36,213	13,727	-62.1
United Kingdom	420,183	68,640	-83.7
Total	1,157,126	240,988	-79.2

Figure 1. 2019 travel and tourism contribution in the country’s economy. Data source: WTTC.

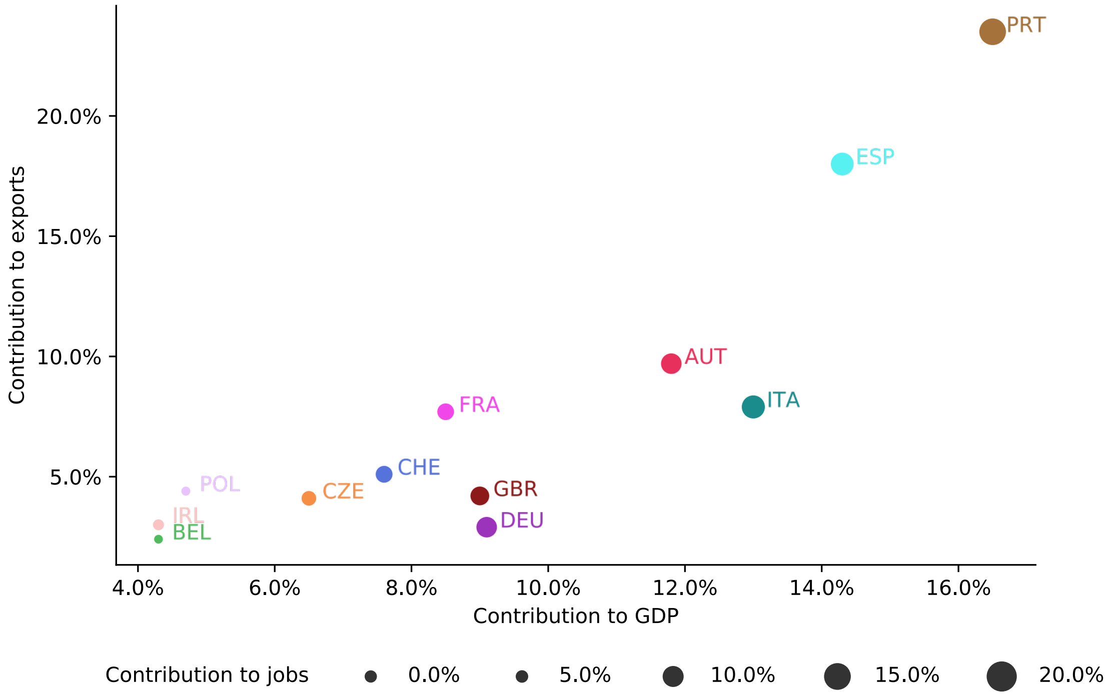

Figure 2. Cumulative cases and deaths per 100,000 of the population, together with measures implemented in March 2020. Data sources: ECDC and ACAPS.

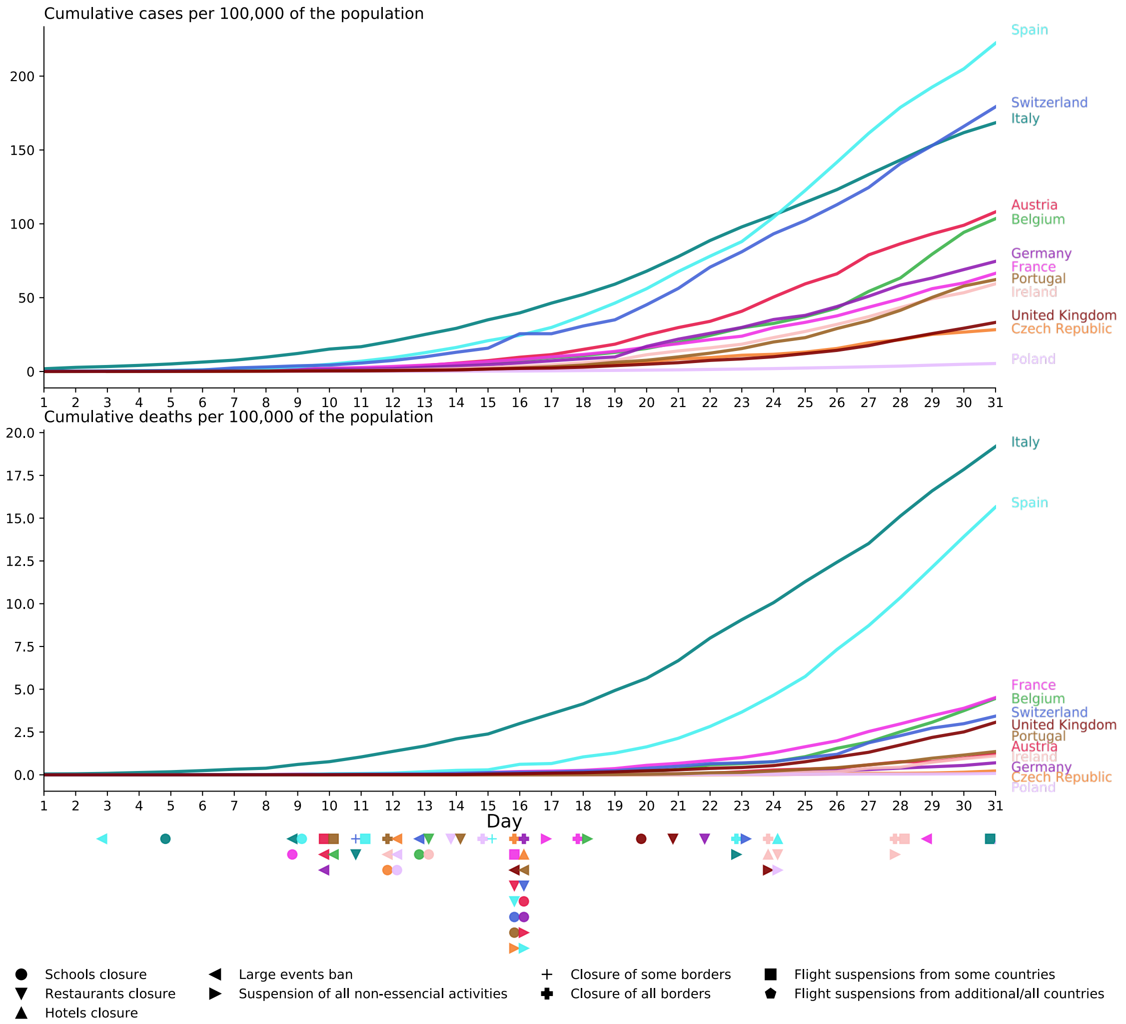

Figure 3. Hierarchical clustering and time series of countries by daily deaths per 100,000 of the population. Data source: ECDC.

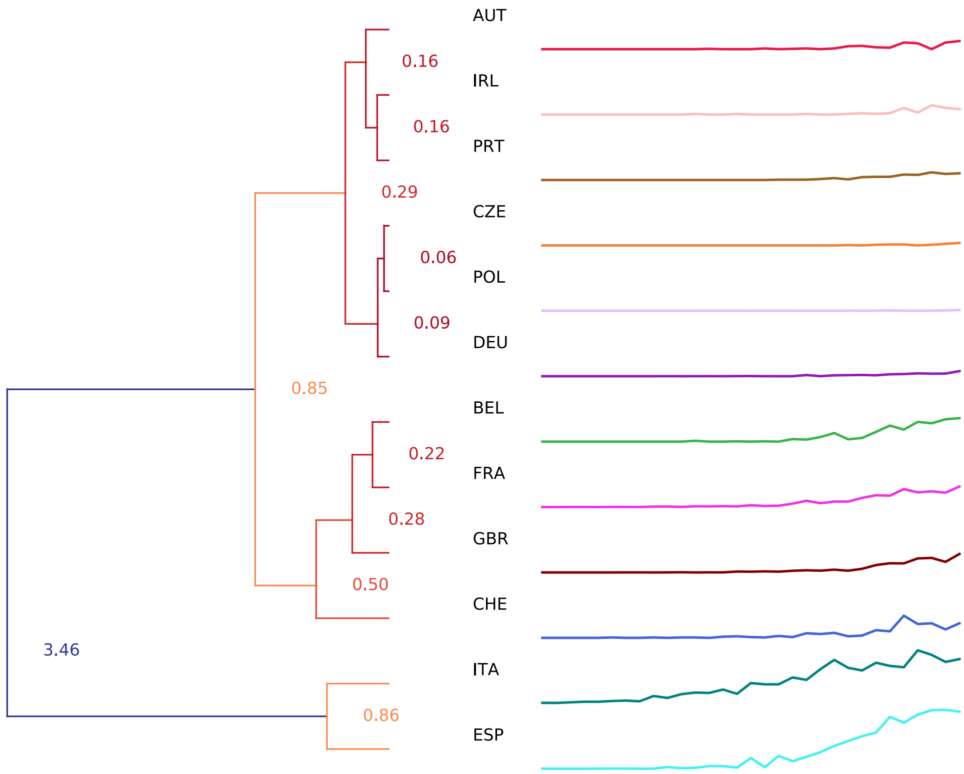

Figure 4. Daily room occupation rate vs. comparable date in 2019, together with measures implemented in March 2020. Data source: STR, OWID, and ACAPS.

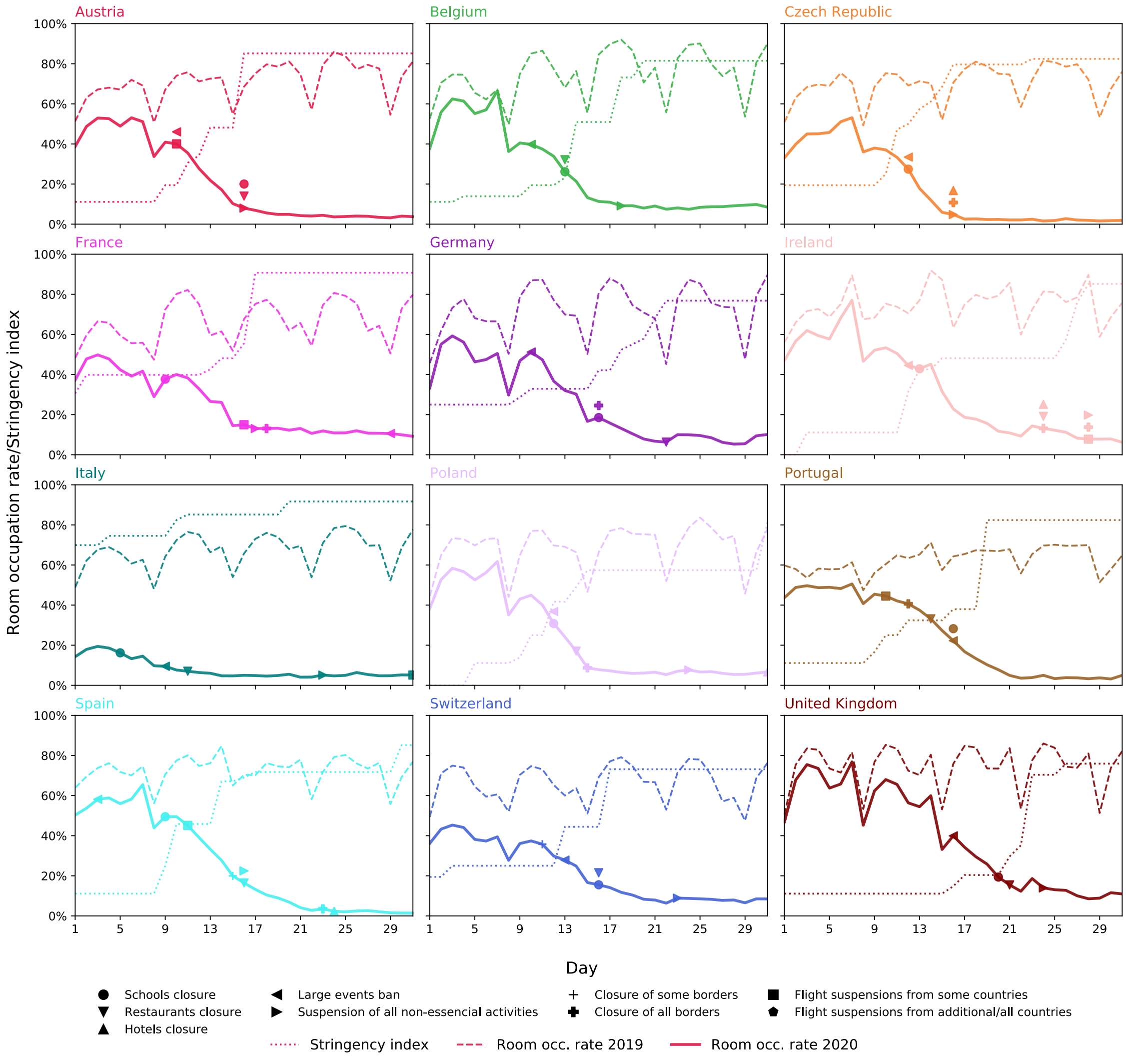

Figure 5. Daily rooms count present in STR's sample, together with measures implemented in March 2020. Since the number of rooms differs per country, numbers are normalized per country using the min-max normalization. Data sources: STR and ACAPS.

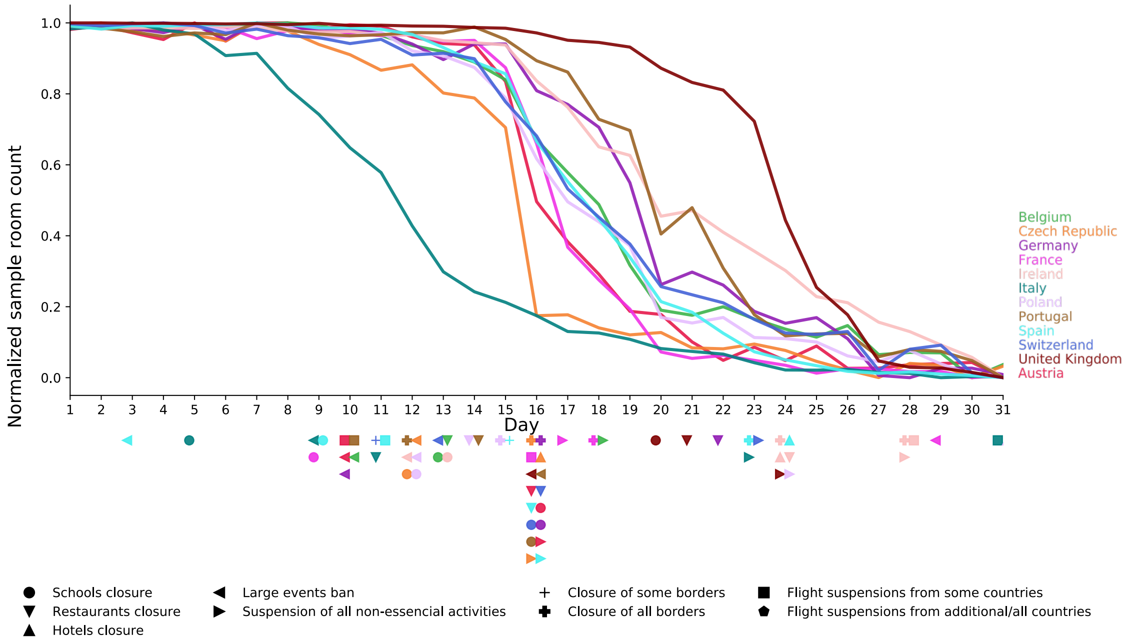

Figure 6. Daily booking index vs. comparable date in 2019, together with measures implemented in March 2020. Data sources: D-EDGE and ACAPS.

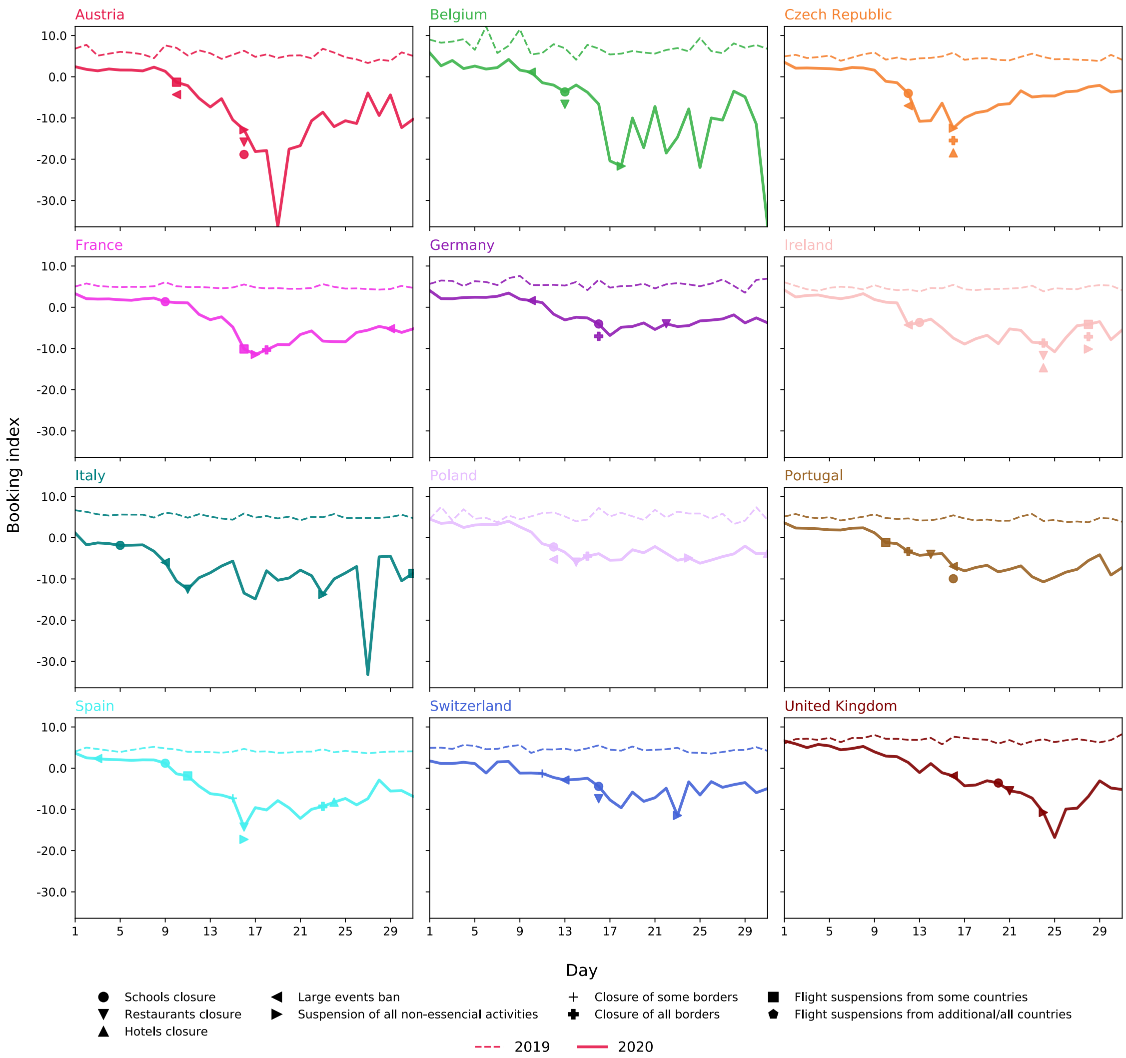

Figure 7. Hierarchical clustering and time series of countries by daily booking index. Data source: D-EDGE.

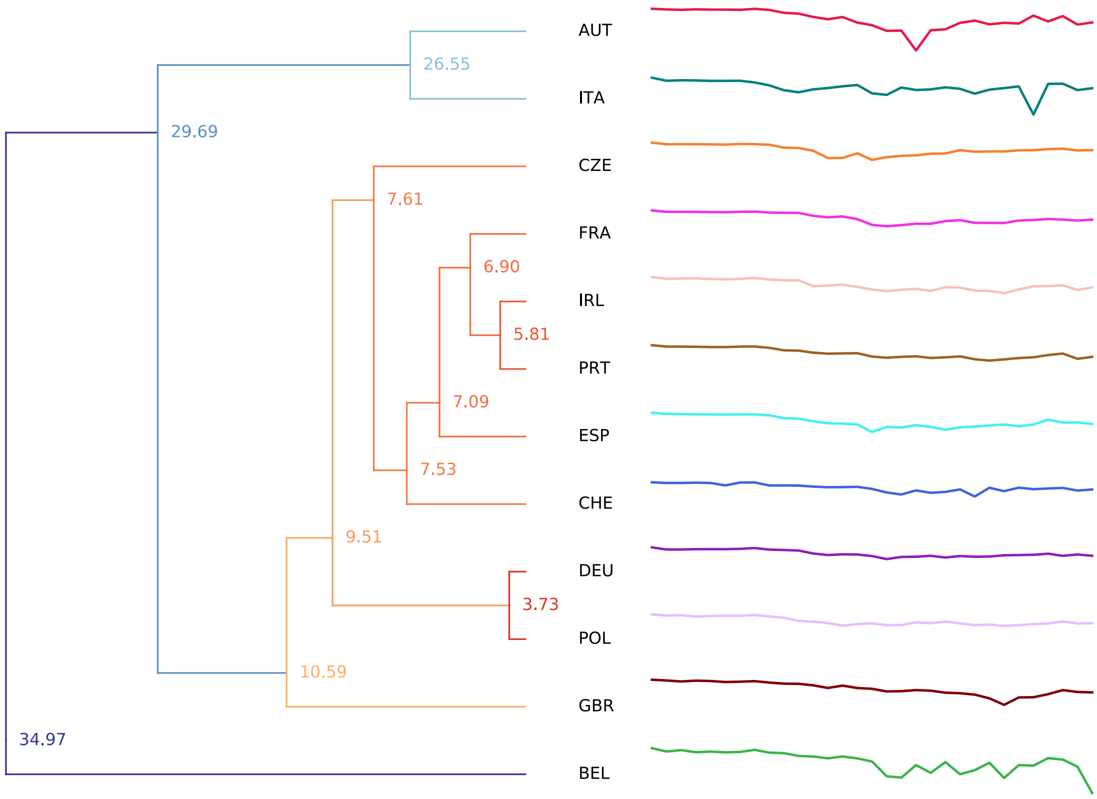

Figure 8. Hierarchical clustering of countries based on the end of March values. Values in parenthesis are to show the travel and tourism contribution to the country's GDP. Data sources: D-EDGE, ECDC, OWID, STR, and WTTC.

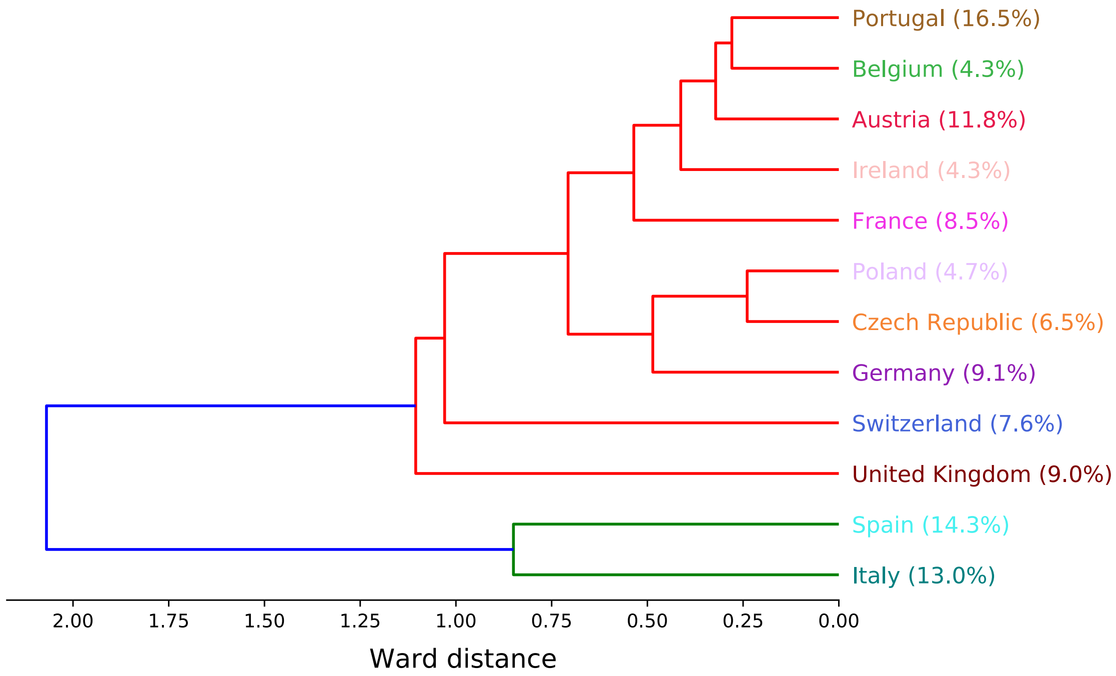

Figure 9. Pearson correlation heatmap based on the end of March values by country. Data sources: D-EDGE, ECDC, OWID, STR, and WTTC.