Using big data to measure cultural tourism in Europe with unprecedented precision

by

Karol Jan Borowiecki, Maja U. Pedersen, and Sara Beth Mitchell

Discussion Papers on Economics
No. 5/2023

FURTHER INFORMATION
Department of Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55, DK – 5230 Odense M
Denmark
Email: astni@sam.sdu.dk

ISSN 2596-8157
https://www.sdu.dk/en/om_sdu/institutter_centre/oekonomiskinstitut

Using big data to measure cultural tourism in Europe with unprecedented precision

Karol Jan Borowiecki *
University of Southern Denmark

Maja U. Pedersen
University of Southern Denmark

Sara Beth Mitchell
University of Southern Denmark

November 2, 2023

Abstract:

International tourism statistics are notorious for being over-aggregated, lacking information about the tourist, available with a lag, and often provided only at the annual level. In response to this, we suggest a unique complementary approach that is computer-science driven and relies on big data collected from a leading travel portal. The novel approach enables us to obtain a systematic, consistent, and reliable approximation for tourism flows, and this with unparalleled precision, frequency, and depth of information. Our approach delivers also an unprecedented list of all tourist attractions in a country, along with data on the popularity and quality of these attractions. We provide validity tests of the approach pursued and present one application of the data by illuminating the patterns and changes in travel flows in selected European destinations during and after the Covid-19 pandemic. This project opens a range of new research questions and possibilities for cultural economics, in particular related to cultural heritage and tourism.

JEL Classification: J60; L83; O1; Z11; Z3

Keywords: Tourism, Cultural heritage, Big data, Covid-19

*Address: Department of Economics, University of Southern Denmark, 5230 Odense, Denmark; email: kjb@sam.sdu.dk.

We are grateful for the **Runner-up Best Paper Award** received at the 9th ATMC Conference. We thank also Ilde Rizzo, Christophe Gay, Javier Caletrío, Vincent Kaufmann, Guillaume Drevon, Enrico E. Bertacchini, and audiences at the Association for Cultural Economics International Conference 2023 in Bloomington, 9th Advances in Tourism Marketing Conference and seminar participants at the University of Southern Denmark for many helpful comments. This project is funded and scientifically supervised by the Mobile Lives Forum, as part of its research program on the mobility transition. The Mobile Lives Forum is a research institute created by SNCF.

1 Introduction

International tourism statistics, such as those provided by Eurostat, are appreciated and used by scholars and practitioners alike. However, they come also with a number of notorious shortcomings: they are over-aggregated (usually at the country level), lack information about the tourist (at the best, the data records whether the tourist is domestic or foreign), available with a lag of many months or more, and often only at the annual level.

In this paper we suggest a complementary approach to measure tourism that is computer-science driven and relies on big data collected from a leading travel portal. The novel approach enables us to obtain a systematic, consistent, and reliable approximation for tourism flows in different countries, and this with unprecedented precision, frequency, and depth of information. In comparison with mainstream tourism statistics, our approach delivers 1) information on tourism flows at the attraction-level (not country-level), 2) detailed information about the tourist, including the rating given (a proxy for visitor satisfaction), city of origin, and the travel history for several previous years, 3) data as good as in real-time, and 4) at a daily frequency. The approach opens a range of new research questions and possibilities for cultural economics and tourism scholarship, in particular related to cultural heritage and tourism.

We evaluate critically the approach developed here and conduct a range of validity tests. Among others, we show that our data, when aggregated to the country-month-level, correlates at >90% with official tourism statistics from Eurostat (2023b). We then map and describe the data by illuminating the patterns and changes in travel flows in three European countries since 2016. Finally, we present one application of the data and explore tourism flows during and after the Covid-19 pandemic.

In particular, we explore to what degree has tourism activity decreased due to the pandemic, how has the travelling distance changed due to the pandemic, or whether the pandemic has pushed tourism to the nature and/or periphery, that is away from over-crowded top-destinations? To answer these questions, we measure the responses to policy restrictions due to the pandemic and estimate their effect on various outcomes describing tourism. We find that tourism decreased significantly with the introduction of several measures imposed by national governments during the pandemic. Furthermore, we document increases in domestic tourism and a decrease in travel distances along with a redirection towards less crowded destinations. Finally, by considering the global travel history of 3 million travellers in the years since the beginning of 2018, we reconfirm the external validity of the results.

The scope of this paper is motivated by the the vast and strategic role of tourism, particularly of

cultural tourism, in Europe. The travel and tourism sector contributed 3.9% to the gross domestic product in Europe in 2018 and accounted for 5.1% of the total labour force (European Parliament, 2019). The value of the project is visible also through the lens of Europe's cultural and natural heritage attractions, which not only drive tourism, but are also invaluable treasures that offer insights into our past, contribute to environmental conservation, and foster economic growth (Borowiecki et al., 2016). However, these sites face threats from various challenges, including climate change and unsustainable tourism practices. Thus, conducting a research project that provides comprehensive data on all these attractions is crucial for their preservation and serves as a foundation for informed decision-making to safeguard Europe's rich heritage for future generations.

A key novelty is the usage of unique data measuring tourist flows by several million tourists to more than 100.000 tourist attractions in three selected European countries (Denmark, France, and Spain). The attractions covered are the population of all attractions (not a sample anymore) and the data provides also unique indicators on the satisfaction of a visit, including the rating given or various indicators derived from over 3 millions reviews. This project thus pushes the boundaries forward beyond previous studies that measure tourism or visitor density on the basis of tourist arrivals or overnight stays (Amore et al., 2020) or on the basis of the perception of overtourism of cultural sites among locals (Adie et al., 2020), which is subjective and difficult to validate from the outside. It also provides new insights into the geography of tourism activity, which has been previously approximated by the location of enterprises from the tourism industries (Domenech and Capone, 2016). With our data, we are able to show not only the location of attractions, but also that of the tourists and hence illustrate actual travel patterns. Finally, there exists a large and interdisciplinary literature on the role of UNESCO sites for tourism (e.g., Cuccia et al., 2016; Bertacchini et al., 2023; Castillo-Manzano et al., 2021), but little is known how lesser known sites attract visitors. Some have asked whether the World Heritage List "make sense?" (Frey and Steiner, 2011); if it does not, our approach opens new horizons for scholarship to cover any cultural or natural heritage, not just those selected by UNESCO.

This research provides four main contributions. First, it demonstrates and validates the possibility to construct a large dataset on tourism activity and tourist attractions from a leading travel portal. Second, it provides novel insights into the tourism mobility in selected European countries with unprecedented depth and precision. Third, it obtains a unique database of the population (not a sample) of cultural and natural heritage attractions. Fourth, it contributes new insights on tourism activity at the attraction level and with daily frequency during the onset of Covid-19, as well as after the gradual re-opening of the society in a post-Covid-19 Europe.

The rest of the paper is organised as follows: Section 2 presents an overview of the exiting literature, Section 3 describes our data, Section 4 presents the results and Section 5 concludes.

2 Literature review

This paper contributes to different strands of the literature both within tourism, cultural/natural heritage and economics.

Tourism is more than travelling and consuming and it has a great potential for sustainable development if it focuses on culture, nature, knowledge and experiences (Zieba, 2017). It has become one of the most important industries and economic activities in the world (Rizzo and Noonan, 2017), with implications for culture, social relations and the environment, and it can be considered one of the key elements of globalization (Zieba, 2020). It is thus not surprising that the European Commission puts a significant focus on tourism (European Commission, 2023), as it contributes to growth and value creation throughout Europe. Perhaps particularly desired are cultural tourists who spend more than other tourists and are generally more educated (Falk and Katz-Gerro, 2017; Falk and Hagsten, 2017).[1] However, with the outbreak of the global Covid-19 pandemic and the associated lockdown measures, the tourism sector is facing massive repercussions in Europe and around the world. Not only is it the hardest hit sector; it could be the one slowest to recover from the upcoming economic recession, affecting Europe's business and working-life in unprecedented ways.

Until the outbreak of the pandemic, tourism grew strongly across cities and rural areas. This has not been uncontroversial and over-tourism has been extensively discussed in the literature and is, together with the carbon footprint of tourism (Lenzen et al., 2018), one of the most important sustainability concerns (Adie et al., 2020; Amore et al., 2020). One challenge faced in the literature is related to the lack of quantitative information on mobility to tourist destinations and attractions. Standard indicators such as the number of visitors per population and the perception of local residents towards overtourism, measured as likert scale variables, have strong disadvantages. New, alternative measures emerge, such as that by Bertacchini et al. (2021) who use transaction data from museum cards to track tourism flows to specific attractions or Tenkanen et al. (2017) who point out that instagram posts for parks are a good approximation of official visitor statistics and thus reflect their popularity. Others have implemented a similar approach using photographs posted

[1] We recognize the challenges in defining *cultural tourism* and its elusive nature, and refer the reader to the discussions provided by Rizzo and Noonan (2017). For the purpose of our paper, we understand as cultural tourists those who visited a cultural site covered by our data.

on Flickr not only in parks but also in city centres (see Sessions et al., 2016; Sinclair et al., 2020; Kádár, 2014; Shi et al., 2017; Wood et al., 2013). For a comprehensive overview of the existing literature using big data in tourism research, see Li et al. (2018). However, to the best of our knowledge, there are no studies available that explicitly focus on the popularity of cultural sites based on social media platforms to study the extent of overtourism and congestion.

This is a point of reflection that enables to observe also the negative sides of tourism, which has also become in many places a problem due to the negative impacts at different levels. The process by which a place is transformed once it becomes object of tourist consumption - the process of Touristification - is one of the consequences in some historical cities or protected natural areas across Europe. It also reduces the quality of the visitor experience. Crowded historical city centres or natural sites result in long waiting times and do not allow visitors to enjoy the visit, particularly in places, where tourism is seasonal and taking particularly place between the months of May and August. The shock of the Covid-19 pandemic may be used to accelerate the trends in innovation and cooperation, promoting sustainable economic development in local communities, their participation and their specific contexts, and contrasting the negative effects of touristification.

Relative to this strand of literature we contribute by showing new trends in tourism using detailed data able to describe not only the volumes of tourism but also the direction. With our newly collected data we can identify different trends for different categories of attractions and see where individuals choose to go, e.g., crowded vs. less crowded destinations.

We also contribute to the tourism literature about the impact of pandemics on tourism and especially that of the impact of Covid-19. A series of studies have been written in the time following the onset of the pandemic, trying to identify what the impact could be on the tourism sector. Travelling often can be connected to some kind of perceived health risk which affects tourism behaviour (Jonas et al., 2011; Lepp and Gibson, 2003; Reisinger and Mavondo, 2005). The more specific effects on both the economics of the tourism industry and the tourists behaviour (Kuo et al., 2008; Yang et al., 2020; Zhang et al., 2020). For example, it has been found that tourists would perform protective behaviours when travelling during the swine flu in 2009 (Fenichel et al., 2013). In Kock et al. (2020) they develop a new model, the Evolutionary Tourism Paradigm, to analyse tourism behaviour during pandemics. Sigala (2020) reviews the emerging literature about Covid-19 and tourism, discussing the impacts and opportunities created by the pandemic and similarly, in Zenker and Kock (2020) they look at the existing literature about the pandemic and in addition they suggest how a research agenda could look like. In Falk et al. (2022) they investigate the change in domestic summer tourism demand following the pandemic while Fotiadis et al.

(2021) forecast different scenarios of international tourism demand computing the expected drops in tourist arrivals. Finally, Singh (2021) explores exotic natural or cultural heritage destinations, including UNESCO World Heritage sites. They argue that these destinations are safer from Covid infection due to fewer tourists and can therefore be expected to flourish in the short and middle term. In this paper we contribute to this literature by showing the actual effects on tourism caused by the pandemic. With our new dataset we are able to look at tourism both before and after to estimate the causal effect.

Finally, we contribute to the literature regarding the use of Tripadvisor as a source for data collection. In (Yoo et al., 2016) they describe Tripadvisor's business model and how Tripadvisor represents open innovation in tourism by summarizing findings of the associated literature. They argue that user-generated content, such as online reviews, is important in influencing destination awareness and selection for trip planners. Additionally, they argue that the success of Tripadvisor is based on the continuous added value to its services arising from its co-creation ecosystem. Previous studies using Tripadvisor as a source of data has focused on topics such as consumers' perceived quality of attractions, incentives to create fake reviews, and rating system design. Regarding the incentives to post fake reviews, Mayzlin et al. (2014) compare reviews posted on Tripadvisor with those posted on the booking platform Expedia. While Tripadvisor has no restrictions on who can post a review, Expedia only allows actual consumers (those who booked via Expedia) of the attraction to review it. Their findings suggests that hotels with a high incentive to fake their reviews are rated more positively on Tripadvisor relative to Expedia. However, according to Glazer et al. (2021) a platform such as Tripadvisor is best off by reporting all reviews in order to "filter out" the fake reviews. Other studies using Tripadvisor as a data source are for example Nguyen et al. (2020), who uses Tripadvisor data to confirm the restraint-of-expertise hypothesis which entails that reviewing experts are less willing to give extreme ratings compared to novices. Additionally, the study finds that experts provide significantly higher ratings than novices for service providers who generally provide mediocre experiences. In Waldfogel (2020) they investigate trade patterns of an important cultural product, food prepared at restaurants. Particularly, the study investigates trade patterns of global cuisines in 52 countries by using Tripadvisor data of restaurants in hundreds of cities around the world and their respective cuisine categories (e.g., Italian, Greek, Japanese) and Euromonitor data for expenditures on fast-food cuisines (e.g., hamburger, chicken, pizza). Related to this is Thanh and Kirova (2018) where they use Tripadvisor reviews to investigate wine tourism in France. Grewal and Stephen (2019) use Tripadvisor data to examine how online reviews written on mobile devices can influence consumers' perception of online reviews and their purchase intentions. The study finds that reviews posted from mobile devices influences consumers' purchase intentions

positively. The authors explain this result by the fact that consumers attach greater credibility to reviews made on mobile devices since they are more physically effortful to create. Wuepper and Patry (2016) use reviews from Tripadvisor to create an index that shows the extent to which World Heritage sites are actually branding themselves as such. A series of studies also uses Tripadvisor ratings. For example, in Hollenbeck et al. (2019) they investigate the relationship between online ratings and advertising spending within the hotel industry using Tripadvisor hotel reviews. Their results suggest that hotels with higher ratings spend less on advertising than lower rated hotels. Chen et al. (2018) investigates how multidimensional rating systems affect consumer satisfaction using suggesting suggest that a multidimensional rating system enhances rating informativeness and help consumers match their preferences with product attributes. Finally, Banerjee and Chua (2016) makes use of Tripadvisor data to investigate hotel rating patterns among different types of travelers. Their findings indicates that travelers' rating patterns for independent and chain hotels vary across profiles. Our contribution to this literature is the collection of a new and large database which contains all available information from Tripadvisor regarding reviews, users and attractions in three selected European countries. We introduce a novel data approach using alternative sources to obtain detailed data.

3 Data

In this section we describe and present our data and their validation. We first present our novel data set about tourism and thereafter we briefly explain the auxiliary data used in our analysis.

3.1 Measuring tourism using big data

International tourism statistics have several significant shortcomings such as being over-aggregated and lacking important information about the tourist. National statistics in some countries provide additional information, for example, the "Familitur" database in Spain includes data on the age and profile of the tourist. However, each country's statistic is unique and international comparisons are not possible. Furthermore, to the best of our knowledge nobody has been able to track multiple moves of a single tourist over several years.

We try to overcome these issues implementing a novel approach based on computer-science and big data collected from a leading travel portal, Tripadvisor. We obtain a systematic, reliable and consistent approximation for tourism flows with unprecedented precision, frequency, and depth of information.

Apart from the detailed information about the tourist, we also collect information about the individual attractions and split them into different categories. This enables us to concentrate on specific attraction types, particularly cultural sites, and also to study tourism flows individually for these different attraction types.

The data collected covers all reviews posted for attraction sites in three selected countries: Denmark, France and Spain. The data collection covers reviews starting from January 2016 and spans up to March 2022. We include reviews in a total of 22 different languages including French, English, Spanish, Italian, Portuguese, German, Dutch, Danish, Russian, Japanese, Mandarin (Chinese Simplified), Taiwanese Mandarin, Swedish, Polish, Norwegian, Korean, Turkish, Greek, Finnish, Czech, Hungarian and Slovakian. With these we cover >96% tourist arrivals to the three countries, according to Eurostat statistics on tourist arrivals by country (Eurostat, 2023a). We used a purpose-built Python web scraping program to collect data from Tripadvisor.com dividing it into four different data entities: list of attractions, attraction reviews, user profiles, and user travel history.

The list of attractions is a complete list of all attractions located in one of our three selected countries and present on Tripadvisor. This module contains information about the attraction, such as the name, the within-country ranking, overall rating, number of reviews, attraction location and the attraction type. The attraction type is based on Tripadvisor's own classification covering 20 different categories. In our analysis we concentrate on the following four: 1) Museums, 2) Nature & Parks, 3) Sights & Landmarks, 4) Others. The "Others" category includes all attractions which cannot be classified in one of the first three. It should here also be noticed that the classification system is not mutually exclusive and hence some attractions can be classified in multiple of the first three categories at the same time. Since a tourist during a pandemic may particularly differentiate between indoor and outdoor activities, we have chosen to only classify an attraction as "Nature & Parks" whenever this is the only category. In cases where the attraction is also classified as Museum or Sights & Landmarks one of these two categories have been used.[2]

The attraction reviews module contains a list of the reviews of each of the attractions included in the attraction module. The module contains the title and text of the reviews, the date the review was published, the rating and a unique and anonymous identifier of the user who published the review. This latter can be used to link the review to the user profile module to obtain additional

[2] An alternative classification would be to categorize each attraction within a given category, if this category is listed. In this case, many attractions would appear more than once in our analysis, leading to a potential bias in the results towards larger and more diverse attractions. However, in robustness tests, we have estimated regressions with the alternative classification (allowing an attraction to appear more than once) and have arrived at qualitatively similar results.

information about the user such as the user location.

The user profile module contains basic information about the users who wrote at least one review for at least one attraction in our sample of countries. It reveals information about the user such as the user location.

Finally, the user travel history module reports all reviews written by the users in the user profile module. This last module therefore extends our data to attractions outside our three selected countries and can therefore be considered a global sample of attractions. However, it should be noticed here, that this global sample does not represent a complete list of all attractions present on Tripadvisor, but only those visited by the users in the user profile module. The data collected in this module covers a period spanning from January 2018 to March 2022. In our analysis we use this module to conduct an analysis at the individual level and as a robustness check to confirm the external validity of our main results.

With the first three modules at hand we can combine their information to obtain a big panel containing information about both the users, the reviews and the attractions. The information included here is at the individual and daily level and hence highly dis-aggregated. To obtain additional variables, we geocode the location of attractions and users to identify their latitudes and longitudes. In addition to the variables already explained above, we add the the travel distance between the user writing the review and the attraction visited, a "foreign" dummy which equals one when a reviews is written by a user who is not from the same country as where the attraction is located. We also include two measures of density, one measuring attraction density and one measuring tourist density. The travel distance is measured for the individual and it is computed using the existing information about user location and attraction location present in the list of attractions and the user profiles. The "foreign" dummy is also at the individual level and simply compares the user country and attraction country. The attraction density, which is measured at the attraction level, is an approximation of the supply of attractions in a given location; in other words, this density measure proxies for how appealing is a given location for tourists. For each attraction we count the number of other attractions located within a radius of $10km$ as a measure of density.[3] Finally, the tourist density, which is also measured at the attraction level, is computed as the total number of reviews in a given month within a radius of $10km$ from the attraction.[4]

Table 1 Panels A-C presents descriptive statistics by attractions and users for the entire sample while descriptive statistics by country can be seen in the appendix Tables A1-A3. In Table 1 Panel

[3] We also compute the attraction density with an alternative radius of $5km$ and $25km$ as a robustness check.

[4] As for the attraction density, we also compute the tourist density with an alternative radius of $5km$ and $25km$ as a robustness check.

A, we show the overall numbers of users, reviews and attractions. We also show the number and share of attractions within each of the four attraction categories. Our data includes about 6.8 million reviews written by 3111105 users covering 102423 attractions. Of the attractions, 7.2% are classified as Museums, 11.2% as Nature & Parks and 30.7% are Sights & Landmarks. The category covering all other attraction types consists of about 50% of the attractions. When looking at the countries individually, it appears that the share of both Museums, Nature & Parks, and Sights & Landmarks are somewhat higher in Denmark with the respect to the overall, while for the French and Spanish attractions the share are more similar to the overall. In Panel B of Table 1, we show summary statistics with the attractions as the unit of observation, while in Panel C we use the individuals as the unit of observation. Table 2 shows the summary statistics of the global sample using the travel history module without any aggregation. The global sample consists of about 5 million reviews with an average rating of 4.3 out of 5 and an average distance travelled of $3024km$ per visit. In the four years and three months covered by the global sample, the average number of reviews per individual is six and the average number of visits abroad is 2.8. Our detailed data allows us to show the geographical distribution of both - the attractions and users. Figure 1 shows a map of the location of all the attractions in Denmark, France and Spain, while Figure 2 shows a map with the location of users who have provided information about their location. Finally, Figure 3 shows a map of the location of our global sample of attractions. In Figure A1 in the appendix, we show a map of the travel patterns of a sample of reviews for which a user location is provided. The map shows how tourists move both internationally, nationally and locally to reach their destination, including their origin.[5]

3.2 Additional variables and aggregation of the data

In order to estimate the impact of Covid-19 on tourism flows, we use the Oxford Covid-19 Government Response Tracker by Hale et al. (2021), to trace the severity of Covid-19 related lockdowns and policy responses made by governments in Europe during the pandemic. The dataset includes indicators on travel restrictions, school closures, and vaccination policy, as well as an overall govern-

[5]The three countries covered in this research attract tourists from around the world (Panel A World), including the Americas, Australia and New Zealand, and parts of Asia. The outgoing travels from Spain, France or Denmark are equally global, and directed particularly at the Americas and many parts of Asia. Within Europe (Panel B Europe), Spain is one of the most important tourism destinations for the United Kingdom, but attracts also visitors from other parts of Europe, including France. The cross border tourism is marked with, for example, Spaniards and Portuguese visiting each others countries. In general, apart urban destinations, coastal locations attract the most activity. Domestic tourism (Panel C Domestic) is very capital-centred, but also towards other cities as well coastal locations. Madrid and Barcelona appear to be well connected with the surroundings, attracting regional tourism outwards. Paris is a major destination of tourists, but those living in its surroundings tend to travel away from it rather than towards it. Travelling within regions is often directed to cities, but also towards the coast or mountains (Panel D Local).

ment response index which attempts to record the degree of government response to the Covid-19 pandemic. The indicators have been tracked since 1 January 2020 and are still updated. The indicators are measured at the national level, and could appear to be less powerful to explain changes in tourism at a less aggregated level. However, in most cases countries were affected by nation-wide restrictions in which case there is no variation within smaller units of observation.

Specifically, in our analysis we make use of the Stringency Index and an indicator of international travel restrictions. The stringency index is composed of the following nine individual indicators: school closure, workplace closure, cancellation of public events, restrictions on gatherings, closure of public transportation, stay at home requirements, internal movement restrictions, international travel restrictions and public information campaigns. Each of the included indicators I_j has been assigned a score and re-scaled between 0 and 100. The scores have then been averaged according to equation 1 to obtain the composite stringency index, *SI*.

$$SI = \frac{1}{9}\sum_{j=1}^{9} I_j \tag{1}$$

Whenever one of the nine included indicators change, the stringency index will also change accordingly. The travel restrictions indicator records restrictions on international travel. It is measured on an ordinal scale from 0 to 4. A value of zero means no restrictions on international travel while moving from 1 through 4 implies increasingly more severe measures implemented by the government. 1 indicates that there has been implemented a screening at the arrival and 2 a quarantine at arrival from some or all regions. The two most severe categories, 3-4, indicate a partial or total ban from all regions respectively. In all cases the measures are available only from January 2020. For our analysis, we assume the indicator is equal to zero for the earlier years, but our results are robust if we consider only the period for which the indicators are available.

To validate our data, we make use of official tourism statistics from (Eurostat, 2023a) which are aggregated monthly at the country level.

Table 1, Panel D, shows summary statistics for the two main tourism indicators from Eurostat, i.e. number of arrivals and occupancy rates, the stringency index and travel restrictions from the Oxford Covid-19 Government Response Tracker and for the key variables of our main data. Given that data from Eurostat is only monthly, we have aggregated all variables at the country and monthly level.

In Figure 4 we show the evolution of the number of the average number of reviews over time together with the Stringency Index measure. From Figure 4 it is clear that there is a sharp decrease in the

number of reviews beginning in February 2020 when the Stringency Index starts where after the reviews follow a clear inverse relationship with the Stringency Index. In the appendix Figure A2 we also show the number of reviews each month for the years 2020-2022 together with the averages of the years 2016-2019 for each of our three countries and the total. Figure A2 shows a clear pattern of an increase in the number of reviews during the high season. Figure A3 in the appendix, illustrates maps with the total number of reviews and the percentage change in the number of reviews between the years 2016-2019 and 2020-2021. Finally, to have an understanding of the distribution of reviews between different attraction categories, Figure A4 in the appendix shows the number of reviews over time for all the attraction categories included on Tripadvisor.

Apart from looking at the total number of reviews, we also look specifically at the number of foreign and domestic tourists visiting the attractions in our sample and the average distance travelled by all tourists. The evolution over time of the number of foreign and domestic tourists can be seen in Figure 5 while the average distance travelled can be seen in Figure 6. In both figures there is a clear break around February 2020. Finally to look at how tourism has changed over time, we look at the attraction and tourist density. The raw attraction density averaged over time can be seen in Figure 7 when using a radius of $10km$ and Figures A5 and A6 when using a $25km$ or $5km$ radius. The variation in the attraction density measure is here given by the change in the number of reviews of each attraction over time. In Figure 8 we show the change in tourist density over time when using a $10km$ radius, while Figures A7 and A8 uses a $25km$ or $5km$ radius.

Throughout our analysis, we will make use of different levels of aggregation of our data. In our main specifications we have three different levels of aggregation all using the attractions at the main unit of observation. In the most highly aggregated version we aggregate by month, country and attraction type leaving us with a balanced panel with four different attraction types for three countries covering the period 2016-2022. The second level of aggregation is similar but uses daily observations instead of monthly. In this case the panel is unbalanced, given that some attraction types do not receive any reviews on some days. Our most detailed version aggregates directly at the attraction level. In this case we aggregate by month, to avoid too many zeros, given that many minor attractions might receive only a few reviews over a longer period of time.

Finally, when we use our global sample of attractions, we use the individual as the unit of observation. Here we create a panel with the users aggregated monthly and covering the period 2018-2022.

3.3 Validity tests

Before presenting our main results we perform various tests to show the validity of using the data from Tripadvisor as a way to measure tourism flows. We use the data from Eurostat regarding tourism as presented in the former section.

We start with a visual inspection of our data aggregated at the monthly level and compare this to the number of arrivals as given by Eurostat. Figure 9 shows the evolution of all Eurostat arrivals and all Tripadvisor reviews over time. Panel A uses all data, while panels B-D shows the patterns individually for each country: Denmark, France and Spain. It becomes fairly clear that the time-series follow each other very closely in its magnitude and seasonality. We do the same using the occupancy rates which can be seen in Figure A9.

As a second visual inspection, Figure 10 shows a binned scatterplot of arrivals and reviews. This shows the simple correlation between arrivals and reviews. In all four panels it is very clear that they are well aligned. Again, Figure A10 shows the same using the occupancy rate.

As a more formal test we also compute the correlation coefficients between the number of reviews from Tripadvisor and the number of arrivals or the occupancy rate from Eurostat. The results can be seen in Table A4 in the Appendix. Both when using the entire sample and when concentrating on each country separately, the correlation coefficients are very high and also significantly different from zero.

As a final test of the validity we estimate how well tourism arrivals or occupancy rates can explain the number of monthly reviews from Tripadvisor. The results can be seen in in Table 3 where columns 1,3,5 and 7 use $ln(Arrivals)$ as the explanatory variable and columns 2,4,6 and 8 use *Occupancy rate* as the explanatory variable. We show the results for all countries together in columns 1-2 and then individually for each of our three countries, Denmark, France and Spain, in columns 3-8. In all models we include country fixed effects, year fixed effects and month fixed effects. All models have a high explanatory power and the estimates are all statistically significant. In column 1, for example, a 1% increase in the number of arrivals corresponds to a 0.63% increase in the number of reviews. The correlation between the occupancy rate and reviews is somewhat smaller but still significant. When using the entire sample a 1% increase in the occupancy rate implies about a 0.4% increase in the number of reviews.

Given the results in Table 3 we are confident that our data is a valid alternative to using official tourism statistics and we therefore proceed with our analysis.

4 Empirical strategy and results

In this section we demonstrate one application of the Tripsadvisor data by exploring the effect of the Oxford Stringency Index on different measures of tourism. We first use all reviews from attractions in Denmark, France and Spain during the period 2016-2022. In the appendix we also present the results when using the international travel restrictions indicator as the explanatory variable. We show this in two different versions: 1) using the ordinal scale proposed by the Oxford Government Response Tracker and 2) creating a dummy for each of the four levels of restrictions. In the second part of the analysis, we use the different set of reviews taken from our Travel History module representing a global sample of attractions. Economic theory suggests different ways to model the demand of tourism such as gravity models (e.g., Morley et al., 2014) or the tourism attractiveness model (e.g., Dwyer and Kim, 2003). The aim of our analysis is different, as we want to estimate the impact of a change in the Oxford Stringency Index on tourism rather than estimating a demand model to identify the impact of different explanatory variables. Therefore, our empirical strategy takes a different approach.

4.1 Empirical approach

We conduct our analysis using a fixed effects panel data model to show the effect of the stringency index on different outcomes of interest. We present the results using the three different levels of aggregation explained in Section 3. A first set of regressions are estimated using the following model:

$$y_{ct} = \beta_1 SI_{ct} + \beta_0 + \Gamma + \varepsilon_{ct} \tag{2}$$

where y_{ct} is our outcome of interest, SI_{ct} is the stringency index, or alternatively the travel restrictions and ε_{ct} is the error term. Γ is a vector of fixed effects included in the regressions. We include country fixed effects to control for country specific characteristics that do not change over time. We also include two kinds of time fixed effects. The first is monthly fixed effect that controls for seasonality in our data and the other is year fixed effects which control for characteristics that are constant across countries but change over time. Finally we include two types of attraction fixed effects depending on the data used in the estimation. In the regressions using the data aggregated by country and attraction type, we include heritage type fixed effects, to control for characteristics that are constant across the different categories of attractions, i.e. Museums, Nature & Parks, and Sights & Landmarks, and Others. In the detailed data at the attraction level, we also include attraction fixed effects to control for characteristics specific to each attraction. The parameter β_1

is our estimate of interest and tells how a 1% change in the stringency index affects the outcome variable of interest. We use different outcomes of interest in the analysis. We start by looking at tourism flows using the number of reviews and the share of foreign tourists as the dependent variables. Subsequently, we also look at the direction of tourism where we include the travel distance, attraction density, tourist density, and ratings as the outcomes of interest.

Apart from the overall effect of the stringency index estimated by equation 2, we also estimate a model to establish the differential effect on the different categories of attractions:

$$y_{ct} = \beta_1 SI_{ct} + \sum \beta_i SI_{ct} \times HeritageType_{ct} + \beta_0 + \Gamma + \varepsilon_{ct} \tag{3}$$

where y_{ct} is again the outcome of interest, SI_{ct} is the stringency index, ε_{ct} is the error term and Γ the set of fixed effects as described above. In this case our outcomes of interest are the number of reviews and the share of foreign tourists. $HeritageType$ is a set of dummy variables, one for each of the four attraction categories. The β_i, with $i = (Museums, Nature\&Parks, Sights\&Landmarks, Others)$ are specific to each of the four categories with Nature & Parks as the reference category. It estimates the additional effect of the stringency index on Museums, Sights & Landmarks, and Others with respect to Nature & Parks. A significant estimate of β_i indicates a significantly different effect between the reference category and each of the other three categories.

4.2 Effect on tourism flows

We start our analysis by showing the effect of the stringency index on two simple measures of tourism. The first is the natural logarithm of the number of reviews, $\ln(Reviews)$, and the second is the share of foreign tourists, $Share\ foreign\ tourists$. The share of foreign tourists is measured between 1 and 100 and measures the share of tourists originating from a different country than that of the attraction reviewed. The number of reviews can be seen as a measure of the volume of tourism, i.e. an alternative to the number of tourists in a destination while the share of foreign tourists tells something about the origin of the tourism. Both variables can tell something about how Covid-19 and the policy measures implemented by the national governments have impacted trends in tourism.

Table 4 shows the results when estimating equation 2 using the two above mentioned measures as the outcome of interest and for the three different levels of aggregation of the data. All columns include a list of fixed effects: Country, Year, Month, Heritage type and Attraction fixed effects as explained above. Columns 1-2 shows the results when aggregating by country, attraction category and month.

In column 1, a 10 percentage points increase in the stringency index implies a 29% decrease in the number of monthly reviews. In column 2 we find that the stringency index also affects negatively the share of foreign tourists, where a 10 percentage points increase in the stringency index implies a 3.2 percentage points decrease in the share of foreign tourists. The results when using the data aggregated at the daily level in columns 3-4 are very similar while in columns 5-6 when using the attraction level they are somewhat smaller in magnitude but still highly significant. This is what we would expect, given the much smaller level of aggregation and hence a much smaller number of reviews by unit of observation. The results in columns 5-6 illustrates the impact at the individual attraction level and indicates that a 10 percentage points increase in the stringency index implies a 7% decrease in the number of reviews and a 2.28 percentage points decrease in the share of foreign tourists.

Table A5 in the Appendix is similar but uses travel restrictions as the explanatory variable. In Column 1 going from one level of restriction to the next (e.g. going from quarantine at arrival to a ban from some regions) implies about a 37% decrease in the the number of reviews and in Column 2, a 7 percentage points decrease in the share of foreign tourists. Again using daily aggregated data shows similar results while at the attraction level the estimates are smaller but still highly significant with a 10% decrease in the number of reviews (Column 5) and a 4.6 percentage points decrease in the share of foreign tourists given a move from one level of restrictions to the next. A more detailed view of these results from Table A5 can be seen in Table A6 which shows the results assigning a dummy variable to each of the levels of restrictions with no restrictions imposed as the baseline category. The results clearly shows how more severe restrictions have an increasingly larger effect on the number of reviews. Furthermore, it also appears clearly that the effect is much larger once a ban on arrivals is imposed while the effect of screenings or quarantines is much smaller. The effect on the share of foreign tourists follows a similar pattern.

The above results all clearly shows how tourism trends have significantly changed due to policy interventions during Covid-19. Once this relationship has been established, we can move forward to explore more in detail how different categories of attractions have been affected. To this end we estimate equation 3 on the same outcomes of interest. It is an empirical question whether tourism has been affected differently depending on the type of heritage attraction.

Table 5 shows the estimates of equation 3. The effect of the stringency index is highly significant and similar in magnitudes to Table 4. The estimate refers to Nature & Parks, indicating a decrease in both the number of reviews (Columns 1, 3, 5) and the share of foreign tourists (Columns 2, 4, 6) for the three different levels of aggregation. Turning to the different interactions between heritage types

and the stringency index, there is no significant difference with respect to the reference category when looking at the results using data aggregated at the attraction category level. However, the results in Columns 5-6 using data aggregated monthly at the attraction level, indicates a decrease in Museums and and increase in Others. In other words, a 10 percentage points increase in the stringency index implies a 2% decrease in the number of Museum reviews with respect to Nature & Parks. Even though not large, this is an indication that visitors substitute visits to museums, which are predominantly indoors with open spaces, preferring outdoor activities. On the other hand, in Column 6 the share of foreign tourists is significantly higher for both Museums and Sights & Landmarks. A 10 percentage points increase in the stringency index implies about a 0.5 percentage points higher share of foreign tourists. The effect is not very large in magnitude, but an indication that the movement towards open spaces is mainly driven by the domestic tourists. In Table A7 in the appendix, we show the results when using the travel restrictions indicator as the explanatory variable. The results are very similar in terms of significance, showing once more that there is a significantly different effect only at the attraction level.

These results are one indication that tourism has changed due to the pandemic. It has not only decreased in total, but also shifted from museums, and sights and landmarks (albeit less so) towards nature and parks.

4.3 Direction of tourism

We have estimated a significant impact of government interventions on tourism volumes, but what happens to underlying outcomes of tourism such as the distance travelled by visitors and their ratings of the attractions they visit? With our detailed new data we can shed light on these alternative outcomes and go beyond simple measures of tourism, to say something about the direction that tourism has been taking after the pandemic. Furthermore, to investigate further the choice between crowded and less crowded attractions we introduce the attractions density and the natural logarithm of tourist density as described in Section 3 as two ways to measure this. If the number of reviews of attractions with a lower density increases, the average density for each category of attractions will decrease. Similarly, if tourists chose to visit locations with less reviews within a given area the tourist density measure will decrease. Therefore, a significant negative impact of the stringency index on both density measures, will indicate that visitors move towards less crowded places. Table 6 shows the results when estimating equation 2 using travel distance, attraction density, tourist density and ratings as the outcome of interest and the data aggregated at the monthly and attraction category level. The effects on the travel distance, the attraction density and the

tourist density are all highly significant while the effect on ratings is not significantly different from zero. In Column 1, a 10 percentage points increase in the stringency index implies a decrease of $75km$ in the travel distance. This implies that when the stringency index increases from 0 to 49 (the average) the travel distance decreases by $372km$. The attraction density also decreases when the stringency index increases, as seen in Column 2. The obtained point estimate of -3.4 implies that, given a change in the stringency index from 0 to 10, the average number of attractions within a $10km$ radius of a visited attraction is lower by 34. Furthermore, in Column 3, a 10 percentage points increase in the stringency index implies a 32 percentage points decrease in the total number of reviews within a radius of $10km$ from each attraction. These results indicate that there is a change towards nearer and less crowded locations. Together with the results from the previous section Table 5, we can conclude that apart from moving towards the nature, visitors also seek more isolated attractions after the pandemic. As a robustness check, in Table A8 we show the results when using different versions of the attraction density and tourist density measures. Reassuringly, the estimates do not change, indicating that the results are not sensitive to the choice of the radius chosen in the computation. When it comes to the ratings, the results indicate that there is no significant change to be attributed to an increase in the stringency index. In the Appendix Tables A9 and A10 we show the results when using the travel restrictions indicator and the dummies. The results are very similar, showing no effect on ratings but significant negative effects on both travel distance and attraction density. Interestingly, in Table A10 it is only a partial or total ban of all arrivals that has an impact on attraction density, while for the tourist density the and the travel distance the effects for these two categories are also larger. In conclusion this section has provided evidence that individuals chose destinations closer to their home and also that they chose destinations that are less crowded.

4.4 External validity using a sample of global tourism

In this section, we present results using our global sample of attractions. The analysis is conducted at the individual level, where we follow the users from Tripadvisor throughout the years 2018-2022 aggregated monthly. Once again we estimate equation 2, this time with the following four outcomes of interest: number of reviews domestically, number of reviews abroad, travel distance, and rating. Both measures shows how the volume of travelling has been affected. In addition, the travel distance and the rating once again are indications of the direction that tourism is taking after the pandemic. The results can be seen in Table 7. In all columns we include individual fixed effects to control for characteristics that are specific to each individual and constant over time. Additionally, in Columns 2,4,6,8 we also include country fixed effects to control for destination country specific

characteristics, and year and month fixed effects to control for time specific characteristics and seasonality. In all specifications the estimated parameters are highly significant and an increase in the stringency index negatively affects most of the outcomes. in Columns 2 and 4 repsecetively, a 10 percentage points increase in the stringency index implies a 2% decrease in the number visits and in the number of visits abroad. In Column 6 the travel distance decreases with about $130km$ given a 10 percentage points increase in the stringency index. Finally, the estimate on ratings is significantly different from zero but very small in magnitude. As for the rest of the analysis we also estimate using the travel restictions indicator and the dummiy variables. The resulst of these two can be seen in the appendix in Tables A11 and A12.

From this subsection we can confirm our main results and hence conclude that they are not specific to Denmark, France and Spain, but can apply to all destinations worldwide. Furthermore, the we also show that, at all levels of aggregation, all the way to the individual level, there is a significant effect not only on the volume of tourism but on other factors as well.

5 Conclusion

The findings delivered in this project push the knowledge frontier in several directions. First, we demonstrate and validate that a large dataset on tourist attractions and tourism flows can be collected from travel portals like Tripadvisor, and the novel approach is validated. Second, we map, measure and summarize tourism activity with unprecedented depth and precision. Third, we provide unique insights on tourism activity at the daily and attraction level during the onset of Covid-19, as well as after the gradual re-opening of the society in a post-Covid-19 Europe.

We find that an increase in restrictive measures by national governments implies a decrease in tourism volumes and a decrease in the share of foreign tourists. Furthermore, we show by how much travel distances decrease due to the imposed measures and document that tourism activity is relocated to less dense locations in the periphery. The destinations chosen by visitors are located in less crowded places and there is also a move towards outdoor activities such as visiting a nature park.

This project comes also with policy and societal relevance. The analytical findings are useful for developing strategies and policies at different scales too (e.g., visits to high-density sites vs. periphery). For example, given that Covid-19 increased tourism in the periphery and at nature sites, it should be contemplated on how these trends could be strengthened for the future. Furthermore, since attractions benefit from the proximity to other attractions, it is recommended that locations

in the periphery cross-promote in order to benefit from network effects.

Finally, the data presented in this paper opens a range of new possibilities for future research. One possible extension of or analysis is to look at different attraction types and investigate complementarity between different categories. Another open question is what type of attractions are most conducive towards tourism or the unanswered question on what is the significance or fraction of cultural tourism. The data collected on review texts provides also the potential for studies related to sentiment or simply the experience. With the provided information about both the location of users and attractions, it is also possible to value attractions based on the travel cost method. Last but not least, this project produced the first database of the population (not a sample) of cultural and natural heritage attractions, which pushes the boundaries of scholarship on the heritage.

Tourism should be part of complex socio-economic systems with the capacity to adapt, including the capacity to avoid, limit or reallocate touristic flows or negative impacts within a combination of activities, a network and a community with common interests and expectations, and especially in times after a national pandemic-related lockdown. But this kind of tourism should be also based on a development model aligned with the Sustainable Development Goals, taking into account current global change and its consequences. This paper provides a data-supported reflection on these dimensions and opens paths for further interdisciplinary scholarship in this area.

References

Adie, B. A., Falk, M., and Savioli, M. (2020). Overtourism as a perceived threat to cultural heritage in europe. *Current Issues in Tourism*, 23(14):1737–1741.

Amore, A., Falk, M., and Adie, B. A. (2020). One visitor too many: assessing the degree of overtourism in established european urban destinations: Assessing the impact of tourism in european urban destinations. *International Journal of Tourism Cities*, 6(1):117–137.

Banerjee, S. and Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in tripadvisor. *Tourism Management*, 53:125–131.

Bertacchini, E. E., Nuccio, M., and Durio, A. (2021). Proximity tourism and cultural amenities: Evidence from a regional museum card. *Tourism Economics*, 27(1):187–204.

Bertacchini, E. E., Revelli, F., and Zotti, R. (2023). Lord, how i want to be in that number! on the blessing of unesco world heritage listing. cesifo WORKING PAPERS, no. 10293.

Borowiecki, K. J. and Castiglione, C. (2014). Cultural participation and tourism flows: an empirical investigation of italian provinces. *Tourism Economics*, 20(2):241–262.

Borowiecki, K. J., Forbes, N., and Fresa, A. (2016). *Cultural Heritage in a Changing World*. Springer, Heidelberg.

Borowiecki, K. J. and O'Hagan, J. (2012). Historical patterns based on automatically extracted data: the case of classical composers. *Historical Social Research (Section 'Cliometrics')*, 37(2):298–314.

Castillo-Manzano, J. I., Castro-Nuño, M., Lopez-Valpuesta, L., and Álvaro Zarzoso (2021). Assessing the tourism attractiveness of world heritage sites: The case of spain. *Journal of Cultural Heritage*, 48:305–311.

Chen, P.-Y., Hong, Y., and Liu, Y. (2018). The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments. *Management Science*, 64(10):4629–4647.

Cuccia, T., Guccio, C., and Rizzo, I. (2016). The effects of unesco world heritage list inscription on tourism destinations performance in italian regions. *Economic Modelling*, 53:494–508.

Domenech, R. B. and Capone, F. (2016). *Defining tourist clusters in Europe: A micro-level data approach*, pages 81–100.

Dwyer, L. and Kim, C. (2003). Destination competitiveness: Determinants and indicators. *Current Issues in Tourism*, 6(5):369–414.

European Commission (2023). Tourism satellite accounts in europe, 2023 edition. *Statistical Reports, Eurostat*.

Eurostat (2023a). Arrivals at tourist accommodation establishments by country of origin of the tourist. Online data code: TOUR_OCC_ARNRAW. http://data.europa.eu/88u/dataset/ctjvyeqsbf9wsepdmmqrvw. Last data update: 05/01/2023.

Eurostat (2023b). Eurostat tourism statistics. https://ec.europa.eu/eurostat/web/tourism. Last data update: 05/01/2023.

Falk, M. and Hagsten, E. (2017). Measuring the impact of the european capital of culture programme on overnight stays: evidence for the last two decades. *European Planning Studies*, 25(12):2175–2191.

Falk, M. and Katz-Gerro, T. (2017). Modeling travel decisions: Urban exploration, cultural immersion, or both? *Journal of Travel and Tourism Marketing*, 34(3):369–382.

Falk, M. and Scaglione, M. (2023). Effects of regulations on airbnb supply in geneva. 13548166231175049.

Falk, M. T., Hagsten, E., and Lin, X. (2022). Domestic tourism demand in the north and the south of europe in the covid-19 summer of 2020. *The Annals of Regional Science*, 69(2):537–553.

Falk, M. T. and Yang, Y. (2021). Hotels benefit from stricter regulations on short-term rentals in european cities. *Tourism Economics*, 27(7):1526–1539.

Fenichel, E., Kuminoff, N., and Chowell, G. (2013). Skip the trip: Air travelers' behavioral responses to pandemic influenza. *PLoS One*, 8(3).

Fotiadis, A., Polyzos, S., and Huan, T.-C. T. (2021). The good, the bad and the ugly on covid-19 tourism recovery. *Annals of Tourism Research*, 87:103117.

Frey, B. S. and Steiner, L. (2011). World heritage list: does it make sense? *International Journal of Cultural Policy*, 17(5):555–573.

Glazer, J., Herrera, H., and Perry, M. (2021). Fake reviews. *Economic Journal*, 131(636):1772–1787. Date revised - 2021-11-18; Availability - URL:https://academic.oup.com/ej/issue] Publisher's URL; Last updated - 2021-11-24.

Grewal, L. and Stephen, A. T. (2019). In mobile we trust: The effects of mobile versus nonmobile reviews on consumer purchase intentions. *Journal of Marketing Research*, 56:791–808.

Gunter, U., Önder, I., and Zekan, B. (2020). Modeling airbnb demand to new york city while employing spatial panel data at the listing level. *Tourism Management*, 77:10400.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., and Tatlow, H. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour*.

Hollenbeck, B., Moorthy, S., and Proserpio, D. (2019). Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science*, 38(5):793–811.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24.

Jonas, A., Mansfeld, Y., Paz, S., and Potasman, I. (2011). Determinants of health risk perception among low-risk-taking tourists traveling to developing countries. *Journal of Travel Research*, 50(1):87–99.

Kock, F., Nørfelt, A., Josiassen, A., Assaf, A. G., and Tsionas, M. G. (2020). Understanding the covid-19 tourist psyche: The evolutionary tourism paradigm. *Annals of Tourism Research*, 85:103053.

Kuo, H. I., Chen, C. C., Tseng, W. C., Ju, L. F., and Huang, B. (2008). Assessing impacts of sars and avian flu on international tourism demand to asia. *Tourism Management*, 29(5):917–928.

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*, 16(1):88–104.

Lenzen, M., Sun, Y. Y., Faturay, F., Ting, Y. P., Geschke, A., and Malik, A. (2018). The carbon footprint of global tourism. *Nature Climate Change*, 8(6):522–528.

Lepp, A. and Gibson, H. (2003). Tourist roles, perceived risk and international tourism. *Annals of Tourism Research*, 30(3):606–624.

Li, J., Xu, L., Tang, L., Wang, S., and Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68:301–323.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

Morley, C., Rosselló, J., and Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, 48:1–10.

Nguyen, P., Wang, X. S., Li, X., and Cotte, J. (2020). Reviewing Experts' Restraint from Extremes and Its Impact on Service Providers. *Journal of Consumer Research*, 47(5):654–674.

Oklevik, O., Gössling, S., Hall, C. M., Steen Jacobsen, J. K., Grøtte, I. P., and McCabe, S. (2019). Overtourism, optimisation, and destination performance indicators: A case study of activities in fjord norway. *Journal of Sustainable Tourism*, 27(12):1804–1824.

Reisinger, Y. and Mavondo, F. (2005). Travel anxiety and intentions to travel internationally: Implications of travel risk perception. *Journal of Travel Research*, 43(3):212–225.

Rizzo, I. and Noonan, D. (2017). Economics of cultural tourism: issues and perspectives. *Journal of Cultural Economics*, 41:95–107.

Sessions, C., Wood, S. A., Rabotyagov, S., and Fisher, D. M. (2016). Measuring recreational visitation at us national parks with crowd-sourced photographs. *Journal of environmental management*, 183:703–711.

Shi, B., Zhao, J., and Chen, P. J. (2017). Exploring urban tourism crowding in shanghai via crowdsourcing geospatial data. *Current Issues in Tourism*, 20(11):1186–1209.

Sigala, M. (2020). Tourism and covid-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, 117:312–321.

Sinclair, M., Mayer, M., Woltering, M., and Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in germany's national parks. *Journal of Environmental Management*, 263:110418.

Singh, S. (2021). 'quixotic' tourism? safety, ease, and heritage in post-covid world tourism. *Journal of Heritage Tourism*, 16(6):716–721.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., and Toivonen, T. (2017). Instagram, flickr, or twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports*, 7(1):1–11.

Thanh, T. V. and Kirova, V. (2018). Wine tourism experience: A netnography study. *Journal of Business Research*, 2(1):30–37.

Vaske, J. J. and Shelby, L. B. (2008). Crowding as a descriptive indicator and an evaluative standard: Results from 30 years of research. *Leisure Sciences*, 30(2):111–126.

Waldfogel, J. (2020). Dining out as cultural trade. *Journal of Cultural Economics*, 44(2):309–338.

Wood, S. A., Guerry, A. D., Silver, J. M., and Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports*, 3:2976.

Wuepper, D. and Patry, M. (2016). The world heritage list: Which sites promote the brand? a big data spatial econometrics approach. *Journal of Cultural Economics*, 41:1–21.

Yang, Y., Zhang, H., and Chen, X. (2020). Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Annals of Tourism Research*.

Yoo, K.-H., Sigala, M., and Gretzel, U. (2016). *Exploring TripAdvisor*, pages 239–255. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zenker, S. and Kock, F. (2020). The coronavirus pandemic – a critical discussion of a tourism research agenda. *Tourism Management*, 81.

Zervas, G., Proserpio, D., and Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Journal of Marketing Research*, 54(5):687–705.

Zhang, K., Hou, Y., and Li, G. (2020). Threat of infectious disease during an outbreak: Influence on tourists' emotional responses to disadvantaged price inequality. *Annals of Tourism Research*, 84:102993.

Zieba, M. (2016). Tourism flows and the demand for regional and city theatres in austria. *Journal of Cultural Economics*, 40:191–221.

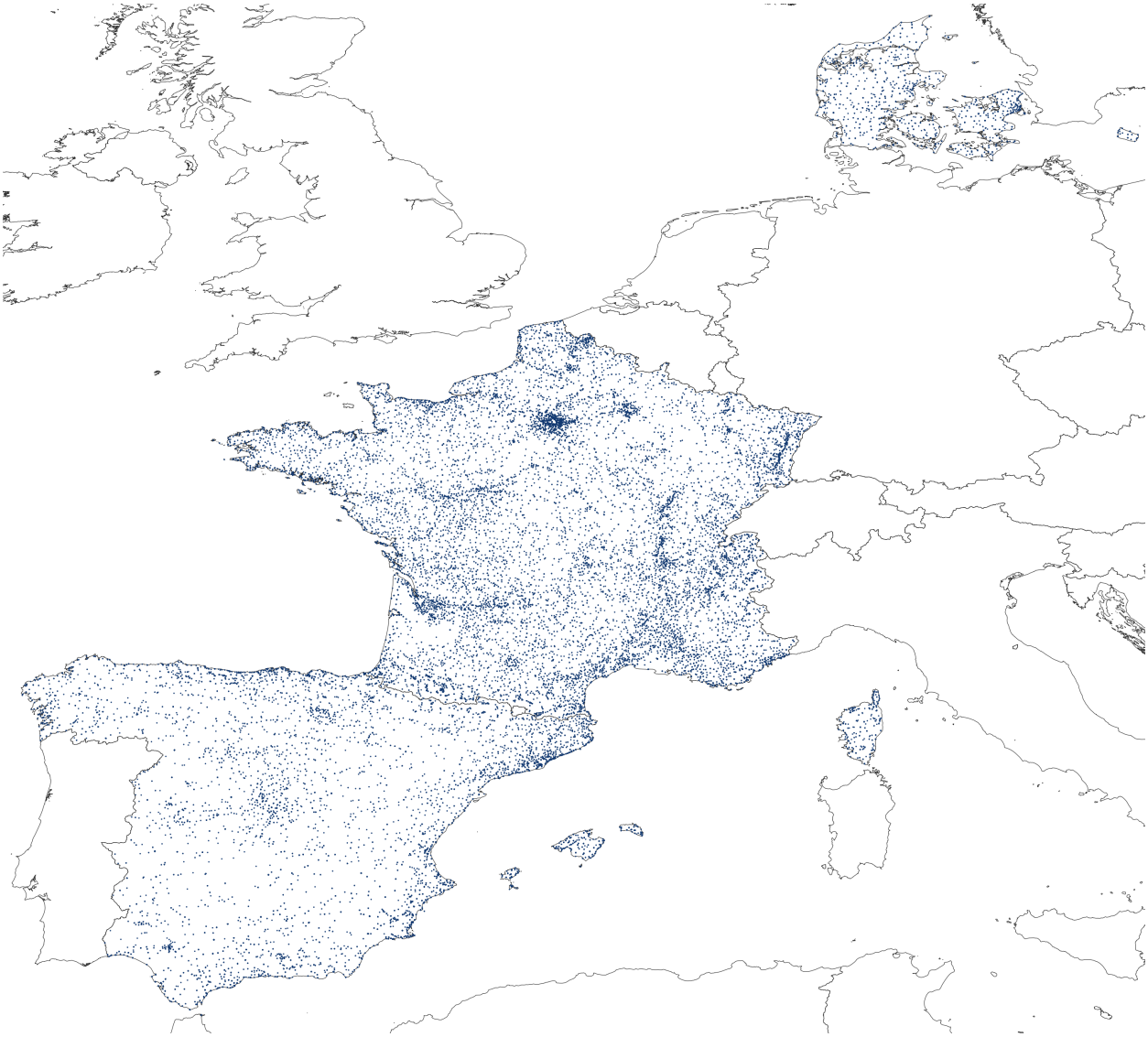
Zieba, M. (2017). Cultural participation of tourists – evidence from travel habits of austrian residents. *Tourism Economics*, 23(2):295–315.

Zieba, M. (2020). *Global trade in cultural tourism services*. Elgar.

Önder, I., Koerbitz, W., and Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints: The case of austria. *Journal of Travel Research*, 55(5):566–573.

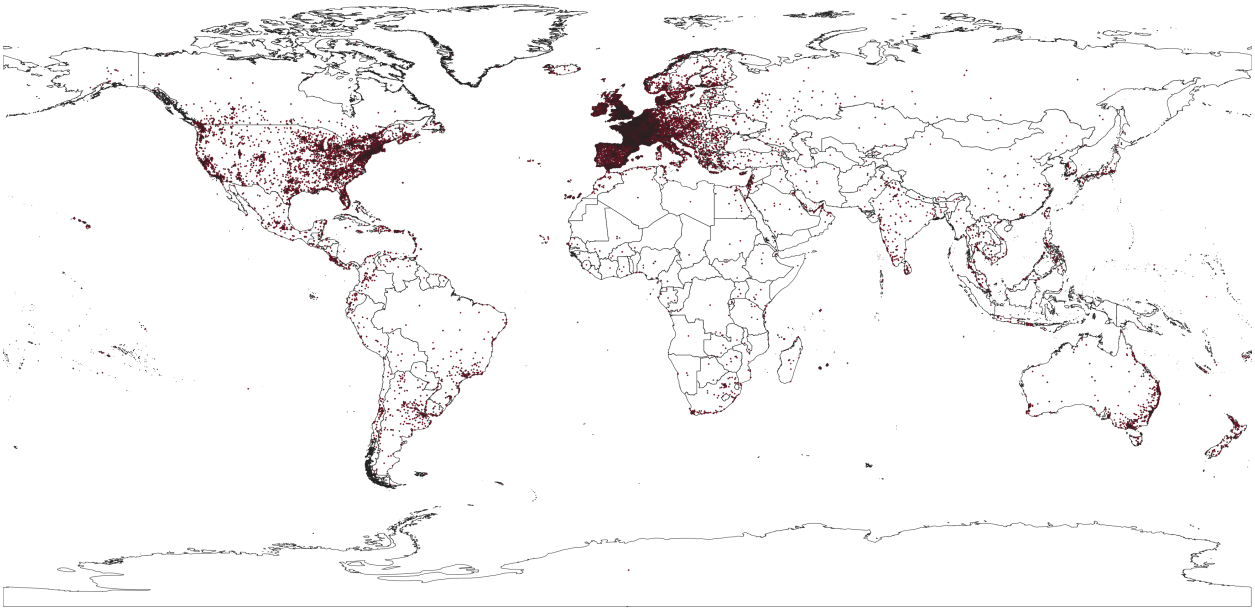
6 Figures

Figure 1: Attractions in Denmark, France and Spain



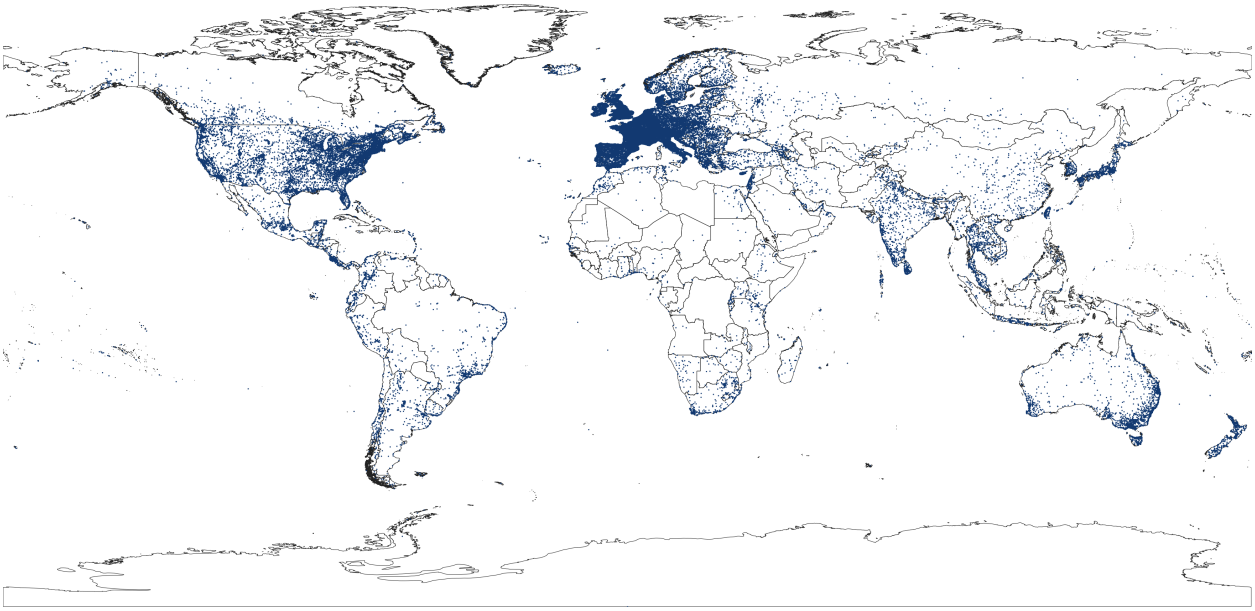
Notes: This Figure shows the location of all attractions present on Tripadvisor and located in Denmark, France or Spain. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

Figure 2: Location of visitors



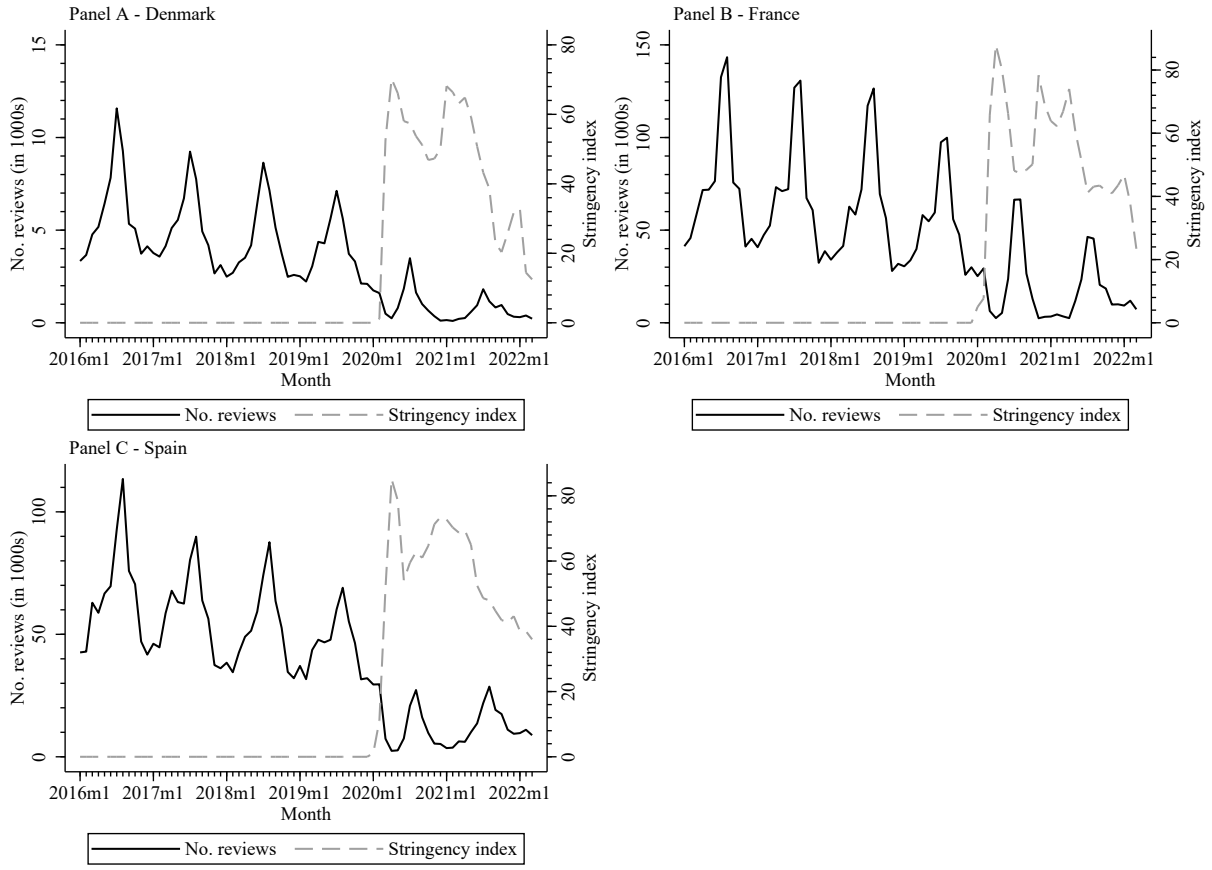
Notes: This Figure shows the location of visitors who have written at leat one review on Tripadvisor of a Danish, French or Spanish attraction. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

Figure 3: Attraction locations using global sample



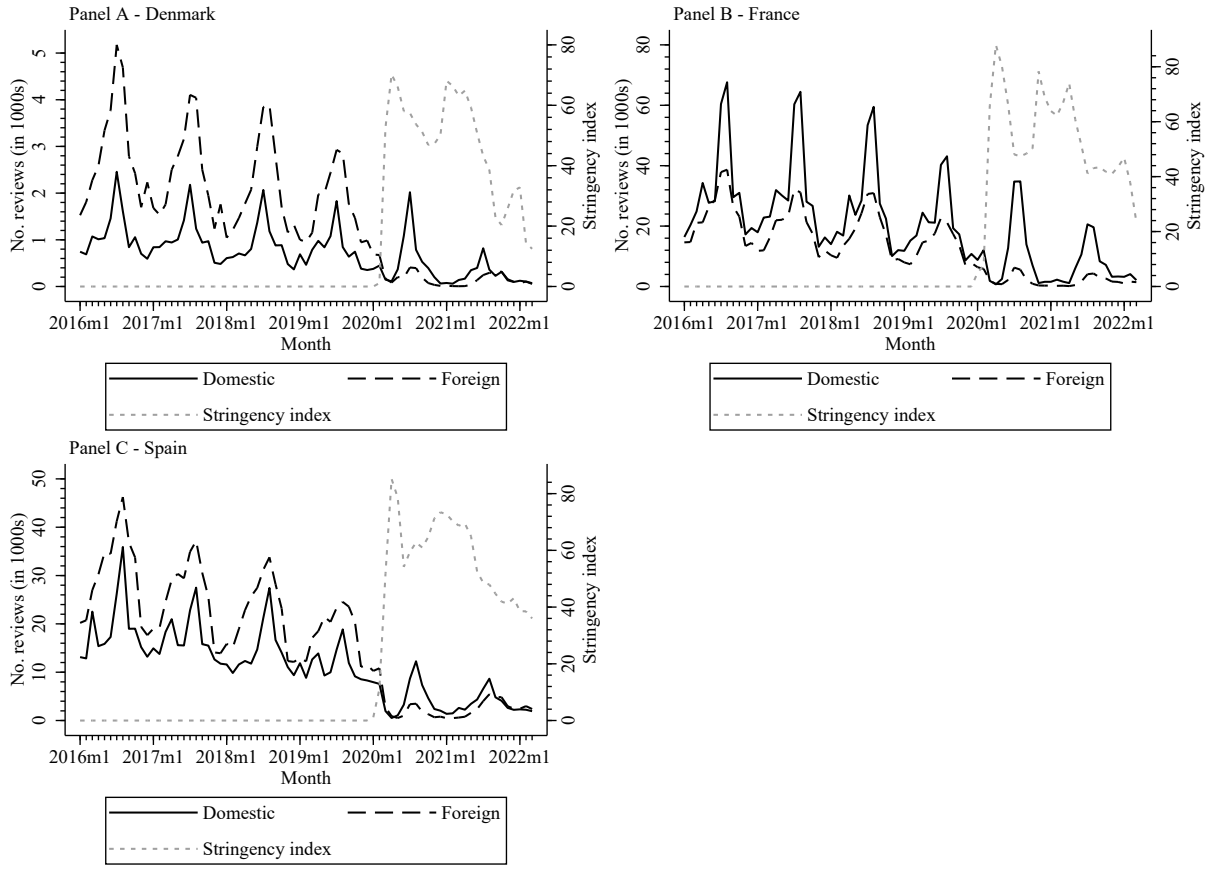
Notes: This Figure shows the location of attractions present all over the world reviewed by visitors who also reviewed at least one attraction in Denmark, France or Spain. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

Figure 4: Number of reviews and Stringency Index over time by country



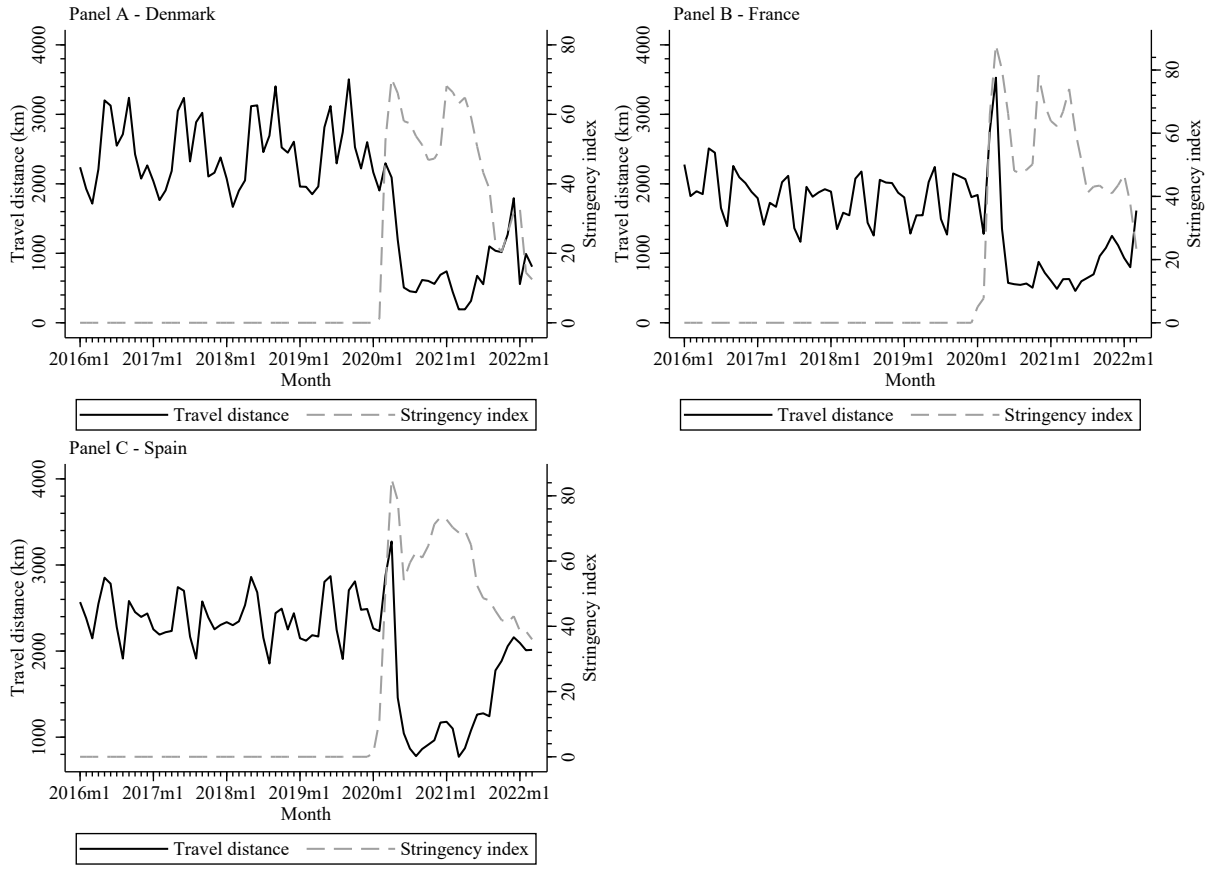
Notes: This Figure shows the evolution of the number of Tripadvisor reviews over time together with the stringency index. Panel A, shows the number of reviews for Danish attractions, Panel B for French attractions and Panel C for Spanish attractions. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure 5: Number of domestic and foreign tourists by country over time



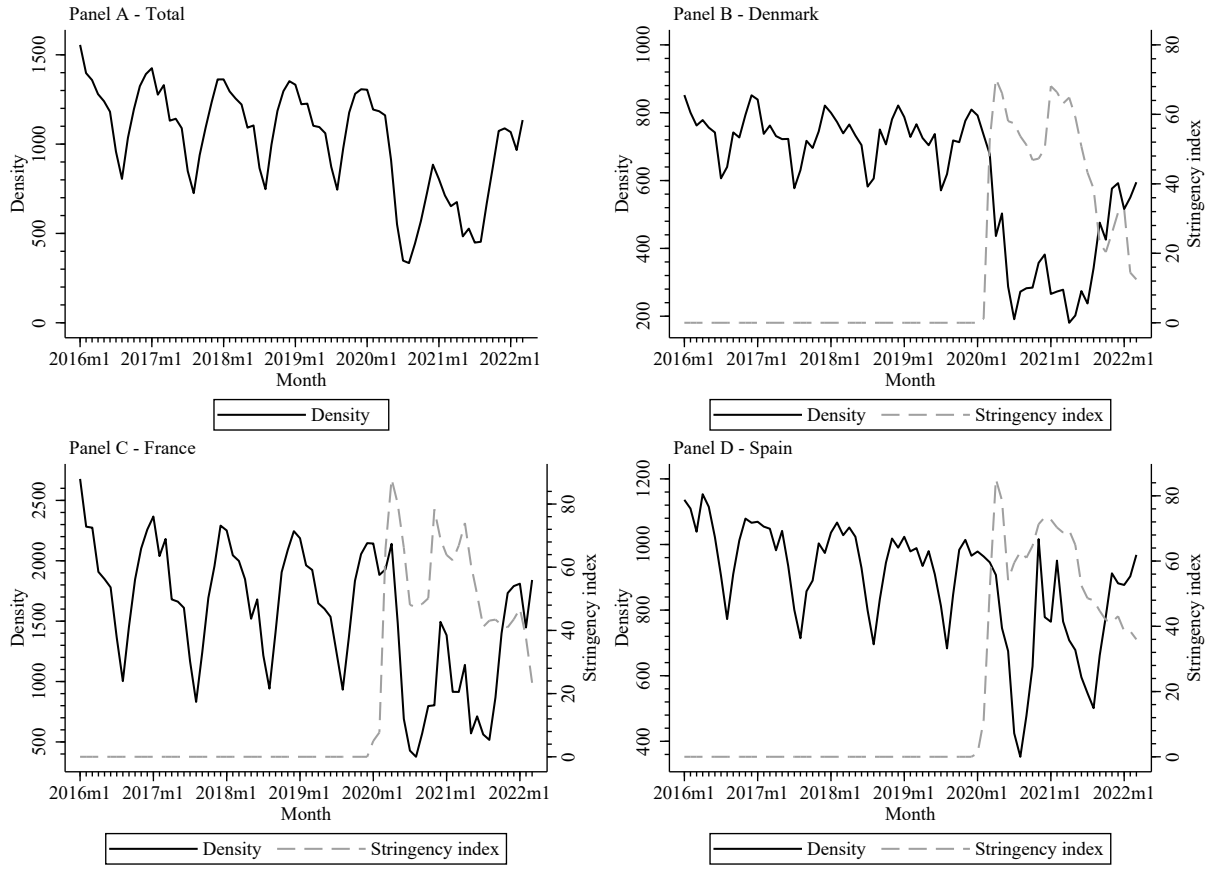
Notes: This Figure shows the evolution of the number of reviews written by domestic and foreign tourists over time together with the stringency index. Panel A, shows the number of reviews for Danish attractions, Panel B for French attractions and Panel C for Spanish attractions. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure 6: Travel distance to attractions and Stringency Index by country



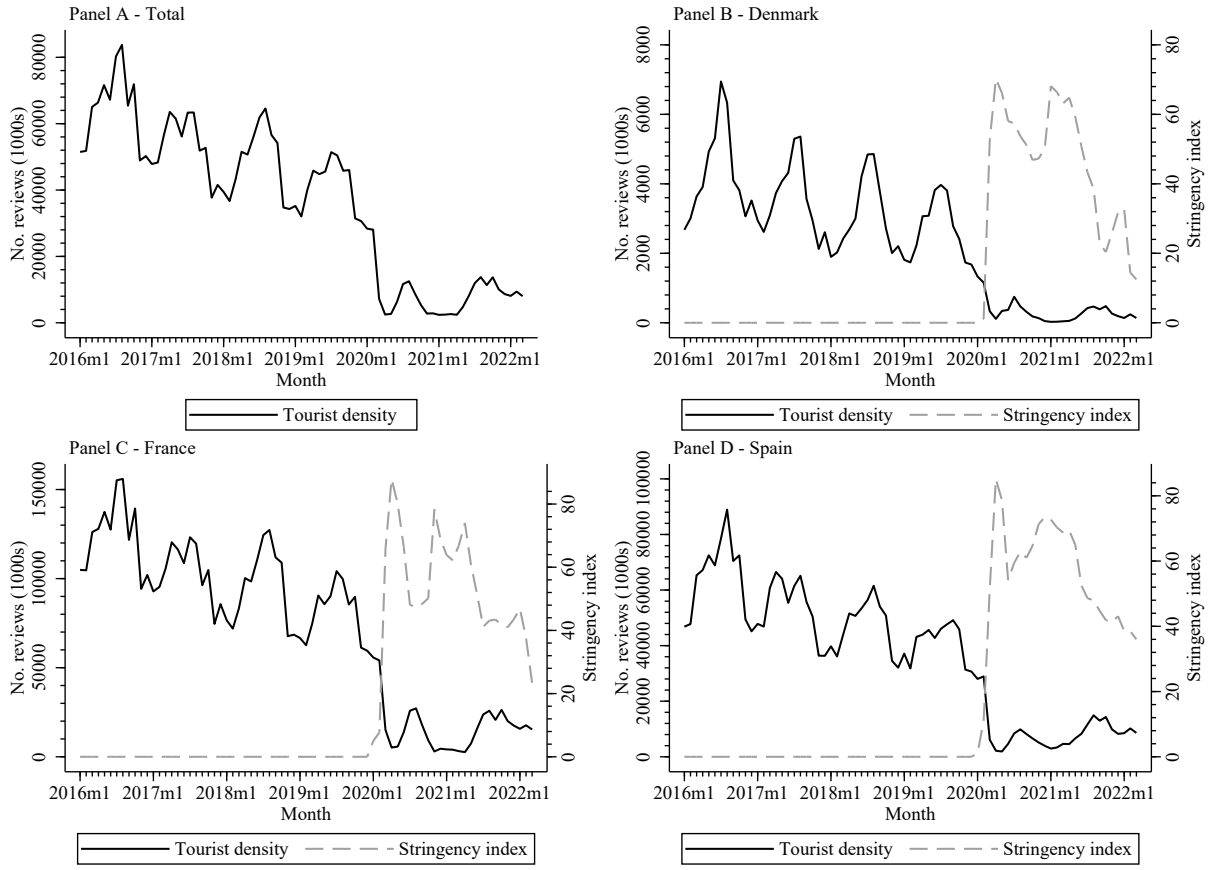
Notes: This Figure shows the evolution of the distance travelled to attractions together with the stringency index. Panel A, shows the travel distance for Danish attractions, Panel B for French attractions and Panel C for Spanish attractions. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure 7: Attraction density of visited locations



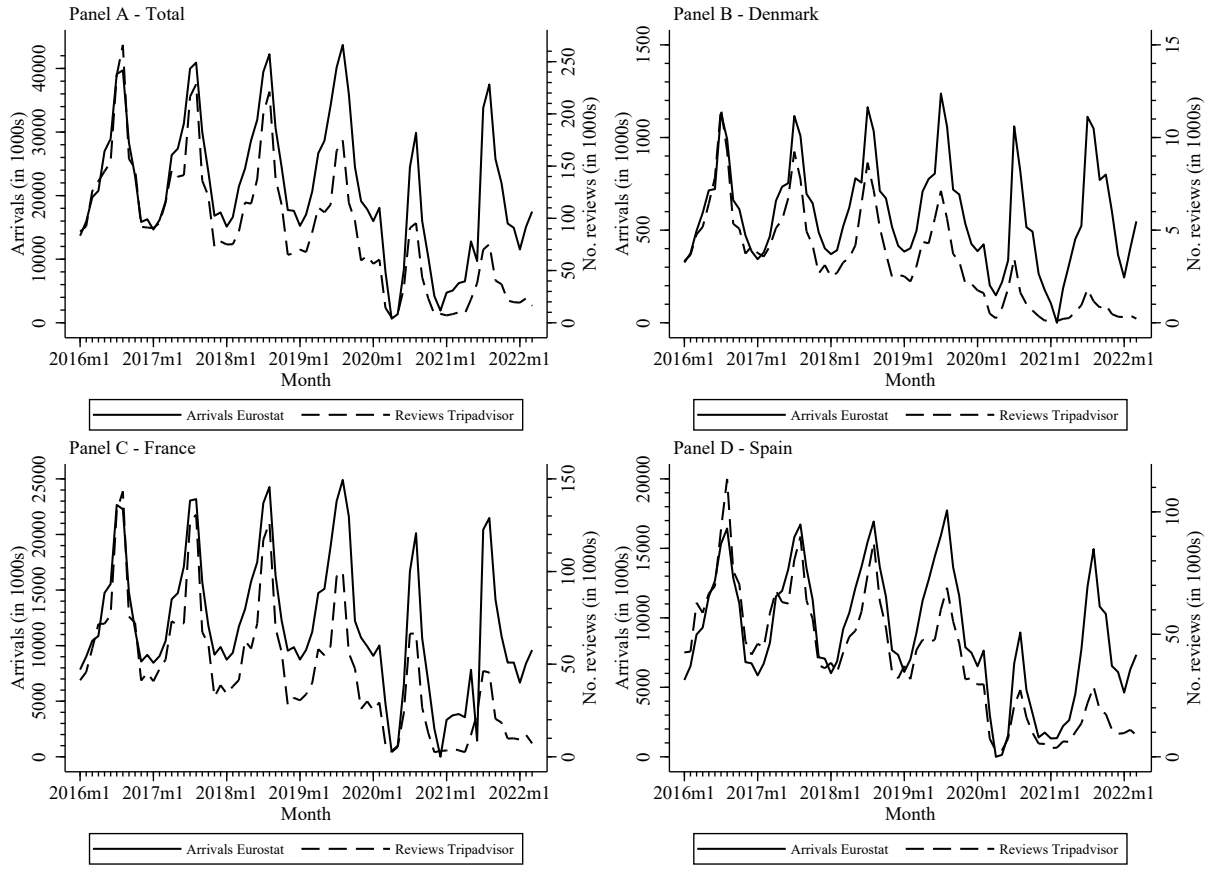
Notes: This Figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the attraction density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. An attraction's density is measured as the number of other attractions within a radius of 10km. The overall density is the average of all attractions' densities in a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure 8: Tourist density of visited locations



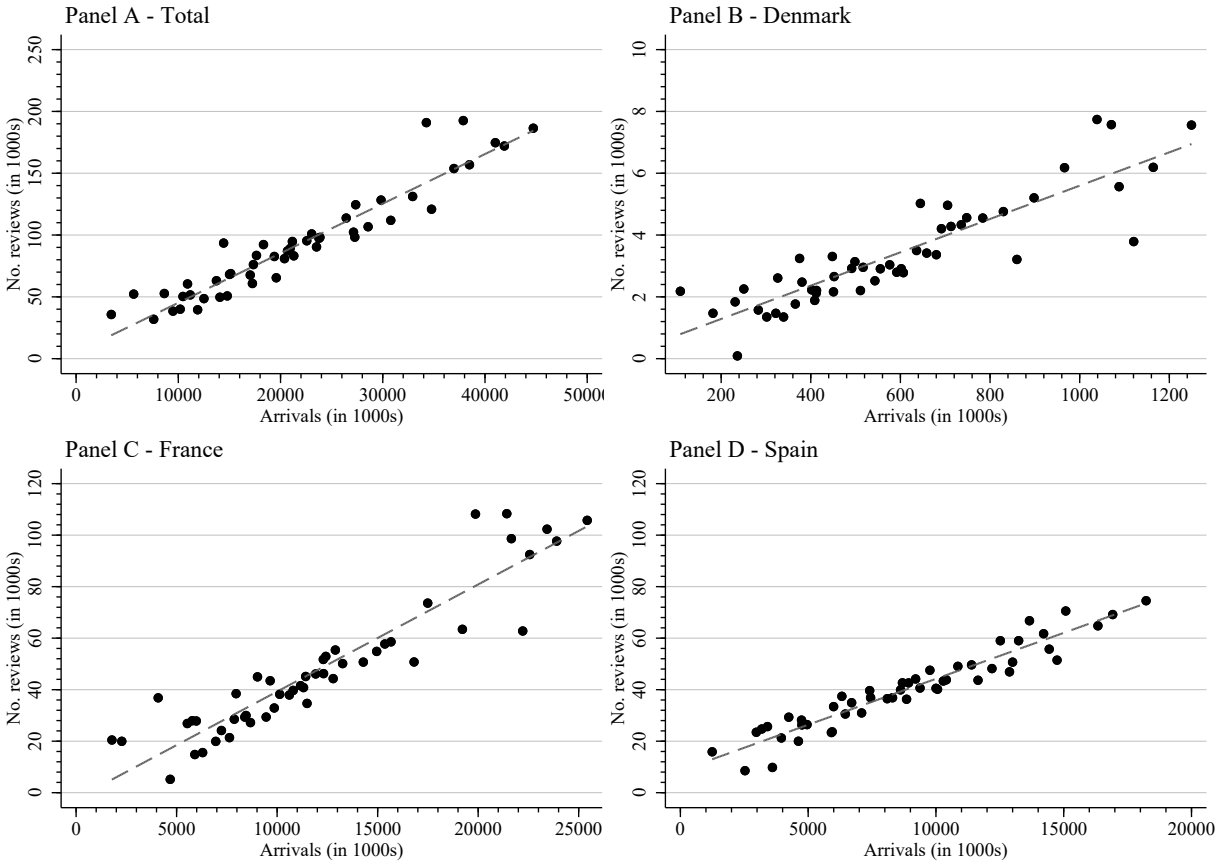
Notes: This Figure shows the evolution of the tourist density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the review density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. The review density of an attraction is computed as the total number of reviews of all attractions within a radius of 10km within a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure 9: Validity test: Tourist arrivals and number of reviews over time



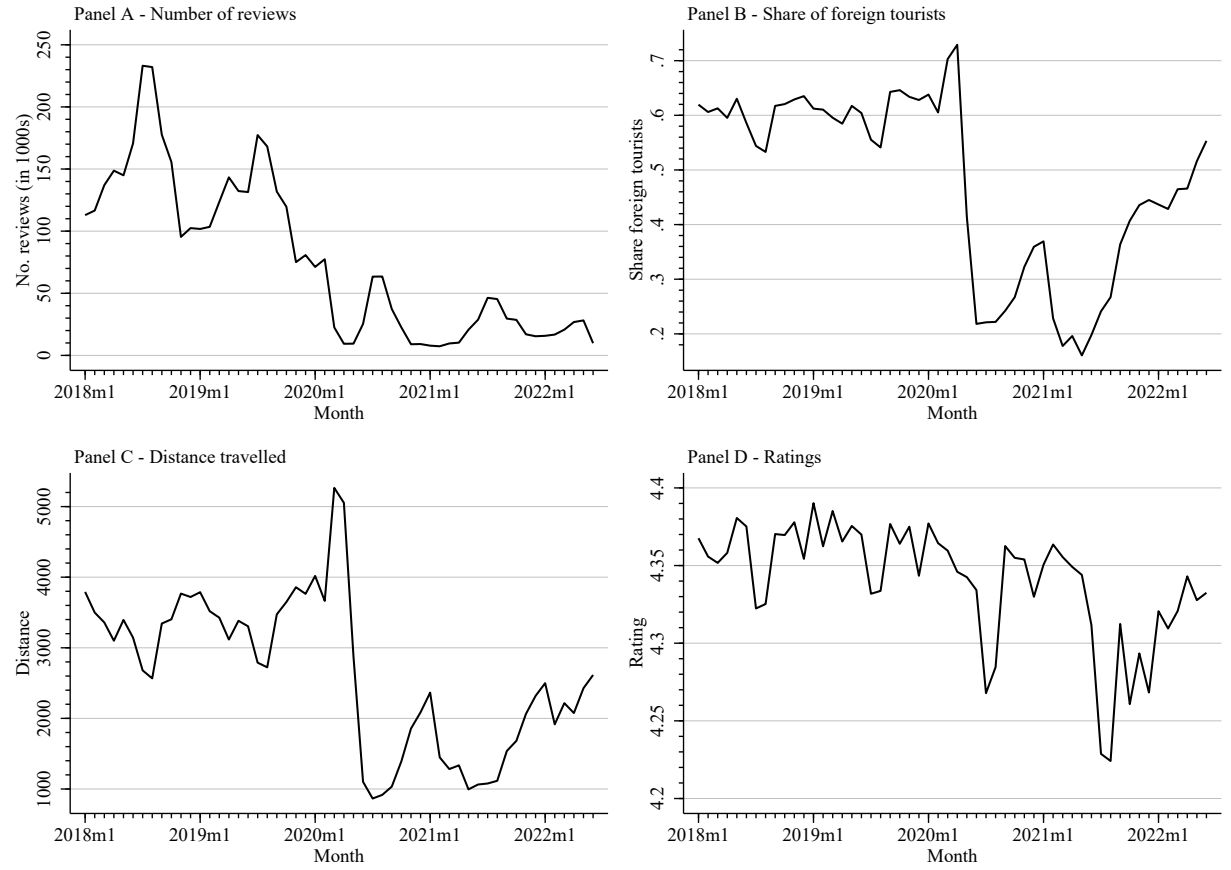
Notes: This Figure shows the number of tourism arrivals taken from Eurostat together with the total number of Tripadvisor reviews. Panel A shows the total number of arrivals and reviews for our sample, while Panels B-D shows the numbers by country. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Figure 10: Validity test: Monthly correlation between tourist arrivals and number of reviews



Notes: This Figure shows binned scatter plots of the number of tourism arrivals taken from Eurostat and the number of Tripadvisor reviews. Panel A uses the entire sample, while Panels B-D by country. The correlation coefficient corresponding to the correlation in Panel A is 0.637, in panel B it is 1.130, in Panel C it is 0.674 and in Panel D it is 0.653. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Figure 11: External validity: Global tourism activity using Tripadvisor data



Notes: This Figures shows different measures of tourism activity using the global sample of attractions. Panel A shows the evolution in the number of reviews, Panel B the share of foreign tourists, Panel C the travel distance and Panel D the ratings. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

7 Tables

Table 1: Summary statistics

Panel A - Rewiews and attractions

	Reviews	Users	Attractions	Museums	Nature & parks	Sights & landmarks	Others
Number	6847931	3111105	102423	7417	11504	31397	53086
Share				0.072	0.112	0.307	0.518

Panel B - Attractions and their visitors

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per attraction	66.859	494.115	1	52021	102423
Foreign visitors of attractions	22.518	282.314	0	32534	102423
Share foreign visitors at attraction	0.287	0.330	0	1	97088
Travel distance to attraction	1073.681	1613.977	0	19655.756	97028
Rating of attraction	4.244	0.718	1	5	102416

Panel C - Visitors and reviews

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per visitor	2.201	8.085	1	5552	3111105
Visits abroad	0.788	6.480	0	5552	3111105
Travel distance of visitor	2111.928	3348.705	0	19664.193	1802312
Rating of visitor	4.465	0.947	1	5	3110689

Panel D - Monthly data aggregated by country

Variable	Mean	Std.dev.	Min	Max	Observations
No. reviews (in 1000s)	30.291	31.421	0.099	143.339	225
Travel distance	1845.167	765.432	193.203	3527.532	225
Share foreign tourists	48.145	18.944	4.762	78.938	225
Rating	4.382	0.090	4.000	4.599	225
Attraction density (within 5km radius)	870.594	452.395	127.029	2173.517	225
Attraction density (within 10km radius)	984.210	526.123	162.634	2478.469	225
Attraction density (within 25km radius)	1315.907	695.942	303.809	3228.884	225
Tourist density (in 1000s) (within 5km radius)	32.024	35.405	0.022	133.448	225
Tourist density (in 1000s) (within 10km radius)	36.514	40.778	0.027	155.970	225
Tourist density (in 1000s) (within 25km radius)	48.923	55.412	0.049	222.763	225
No. arrivals (in 1000s) (Eurostat)	7129.415	6512.297	0	24913.408	225
Occupancy rate (Eurostat)	44.242	16.575	0	83.620	225
Stringency index	49.615	20.305	0.000	87.960	81
Travel restrictions	2.494	1.097	0.000	4.000	81

Notes: This table shows summary statistics for the data and different units of observation. Panel A shows overall numbers. Panel B shows summary statistics using attractions as the unit of observation. Panel C shows summary statistics using the individual users as the unit of observation. Panel D uses monthly aggregated data at the country level. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table 2: Summary statistics for the global sample

Variable	Mean	Std.dev.	Min	Max	Observations
Rating	4.350	0.904	1	5	4937372
Distance	3023.906	4097.554	0	19955.283	4106620
Visits	6.158	19.972	1	3138	805690
Visits abroad	2.839	13.517	0	2628	805690

Notes:Notes: This table shows summary statistics at the individual level for the global attraction data. *Source:* Own data collected from Tripadvisor (see Section 3 for details)

Table 3: Validity Test: Regression results

	(1) ln(Reviews)	(2) ln(Reviews)	(3) ln(Reviews)	(4) ln(Reviews)	(5) ln(Reviews)	(6) ln(Reviews)	(7) ln(Reviews)	(8) ln(Reviews)
ln(Arrivals)	0.637***		1.130***		0.674***		0.653***	
	(0.050)		(0.118)		(0.090)		(0.035)	
Occupancy Rate		0.039***		0.041***		0.063***		0.038***
		(0.003)		(0.006)		(0.006)		(0.004)
Country FE	Yes	Yes	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	Denmark	Denmark	France	France	Spain	Spain
N	222	225	74	75	74	75	74	75
R^2	0.961	0.957	0.953	0.930	0.889	0.919	0.974	0.933

Notes: Regression results when estimating the number of Tripadvisor reviews on the number of arrivals or the occupancy rate from Eurostat. Columns 1-2 uses the entire sample and columns 3-8 shows estimates by country. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table 4: Effect of stringency index on tourism flows

	Monthly		Daily		Attraction level	
	(1) ln(Reviews)	(2) Share foreign tourists	(3) ln(Reviews)	(4) Share foreign tourists	(5) ln(Reviews)	(6) Share foreign tourists
Stringency index	-0.029***	-0.323***	-0.022***	-0.350***	-0.007***	-0.228***
	(0.003)	(0.059)	(0.004)	(0.071)	(0.000)	(0.005)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes
Attraction FE	No	No	No	No	Yes	Yes
N	900	893	26652	26068	1384706	1155251
R^2	0.842	0.346	0.594	0.235	0.192	0.017

Notes: Regression results when regressing the number of Tripadvisor reviews or the share of foreign tourists on the stringency index. Columns 1-2 show the results using monthly aggregated data at the country and attraction category level. Columns 3-4 show the results using daily data aggregated at the country and attraction category level. Columns 5-6 show the results using monthly aggregated data at the attraction level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Table 5: Effect of stringency index on tourism of different attraction categories

	Monthly		Daily		Attraction level	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists
Stringency index	-0.025***	-0.359***	-0.026***	-0.375**	-0.007***	-0.251***
	(0.006)	(0.105)	(0.007)	(0.123)	(0.000)	(0.011)
Museums × Stringency index	-0.009	0.081	0.004	0.033	-0.002***	0.046**
	(0.008)	(0.151)	(0.013)	(0.212)	(0.001)	(0.019)
Sights & Landmarks × Stringency index	-0.006	0.105	0.014	-0.002	-0.001	0.047***
	(0.008)	(0.140)	(0.009)	(0.183)	(0.000)	(0.014)
Others × Stringency index	-0.003	-0.034	-0.001	0.071	0.001**	0.010
	(0.007)	(0.151)	(0.008)	(0.154)	(0.000)	(0.013)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes
Attraction FE	No	No	No	No	Yes	Yes
N	900	893	26652	26068	1384706	1155251
R^2	0.842	0.345	0.597	0.235	0.192	0.017

Notes: Regression results when estimating the number of Tripadvisor reviews or the share of foreign tourists on the stringency index interacted with attraction category dummies. The reference category is Nature & Parks. Columns 1-2 show the results using monthly aggregated data at the country and attraction category level. Columns 3-4 show the results using daily data aggregated at the country and attraction category level. Columns 5-6 show the results using monthly aggregated data at the attraction level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Table 6: Effect of stringency index on travel distance, attraction density, tourist density and ratings

	(1) Distance	(2) Attraction density 10km	(3) ln(Tourist density) 10km	(4) Rating
Stringency index	-7.509**	-3.446**	-0.032***	0.001
	(2.530)	(1.186)	(0.001)	(0.001)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes
N	900	900	900	900
R^2	0.433	0.514	0.908	0.128

Notes: Regression results when estimating the travel distance, attraction density, tourist density or ratings on the stringency index. The results refers to the monthly data aggregated at the country and attraction category level. Column 2 shows the attraction density and column 3 shows the tourist density both using a radius of $10km$, see the text for an explanation of how it has been computed. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Table 7: Effect of stringency index on different outcomes in global sample

	(1) Visits	(2) Visits	(3) Visits abroad	(4) Visits abroad	(5) Distance	(6) Distance	(7) Rating	(8) Rating
Stringency index	-0.005***	-0.002***	-0.006***	-0.002***	-21.114***	-13.520***	0.000***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.143)	(0.283)	(0.000)	(0.000)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	No	Yes	No	Yes	No	Yes	No	Yes
N	1813458	1813410	1813458	1813410	1507250	1507202	1804911	1804866
R^2	0.417	0.427	0.364	0.384	0.397	0.522	0.216	0.218

Notes: Regression results using the global sample of attractions and different outcomes of interest. The data is at the individual level aggregated monthly. Columns 1-2 uses number of visits as the outcome of interest, columns 3-4 the number of visits abroad, columns 5-6 the distance travelled by individuals and columns 7-8 the rating given to an attraction by an individual. Columns 1,3,5,7 only includes individual fixed effects while columns 2,4,6,8 also add country, year and month fixed effects. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

A Appendix

A.1 Tables

Table A1: Summary statistics Denmark

Panel A - Rewiews and attractions

	Reviews	Users	Attractions	Museums	Nature & parks	Sights & landmarks	No. others
Number	254272	109513	6062	1014	758	2506	1854
Share				0.167	0.125	0.413	0.306

Panel B - Attractions and their visitors

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per attraction	41.945	437.268	1	20568	6062
Foreign visitors of attraction	19.022	273.165	0	13310	6062
Share foreign visitors at attraction	0.260	0.328	0	1	5662
Travel distance to attraction	761.596	1417.364	0	16203.868	5654
Rating of attraction	4.126	0.718	1	5	6062

Panel C - Visitors and reviews

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per visitor	2.322	5.851	1	907	109513
Visits abroad	0.513	5.250	0	907	109513
Travel distance of visitor	2651.336	3852.741	0	18139.723	62691
Rating of visitor	4.390	0.886	1	5	109494

Panel D - Monthly data aggregated by country

Variable	Mean	Std.dev.	Min	Max	Observations
No. reviews (in 1000s)	3.362	2.597	0.099	11.574	75
Travel distance	1914.461	915.327	193.203	3504.575	75
Share foreign tourists	56.789	21.484	4.762	78.938	75
Rating	4.289	0.074	4.000	4.410	75
Attraction density (within 5km radius)	529.813	178.049	127.029	772.284	75
Attraction density (within 10km radius)	579.227	187.561	162.634	833.021	75
Attraction density (within 25km radius)	761.350	207.650	303.809	1041.803	75
Tourist density (in 1000s) (within 5km radius)	2.110	1.646	0.022	6.384	75
Tourist density (in 1000s) (within 10km radius)	2.287	1.783	0.027	6.946	75
Tourist density (in 1000s) (within 25km radius)	2.919	2.260	0.049	8.974	75
No. arrivals (in 1000s) (Eurostat)	585.205	279.233	0	1237.441	75
Occupancy rate (Eurostat)	39.547	16.410	0	71.000	75
Stringency index	43.326	20.813	0.000	70.247	27
Travel restrictions	2.593	1.248	0.000	4.000	27

Notes: This table shows summary statistics for the data and different units of observation and only for Danish reviews and attractions. Panel A shows overall numbers. Panel B shows summary statistics using attractions as the unit of observation. Panel C shows summary statistics using the individual users as the unit of observation. Panel D uses monthly aggregated data at the country level. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table A2: Summary statistics France

Panel A - Rewiews and attractions

	Reviews	Users	Attractions	Museums	Nature & parks	Sights & landmarks	No. others
Number	3578698	1544564	57247	3929	6066	16827	31010
Share				0.069	0.106	0.294	0.542

Panel B - Attractions and their visitors

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per attraction	62.513	468.152	1	52021	57247
Foreign visitors visitors of attraction	16.654	255.088	0	32534	57247
Share foreign visitors at attraction	0.229	0.291	0	1	54583
Travel distance to attraction	894.492	1556.975	0	19210.844	54510
Rating of attraction	4.263	0.697	1	5	57240

Panel C - Visitors and reviews

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per visitor	2.317	8.330	1	5552	1544564
Visits abroad	0.998	7.037	0	5552	1544564
Travel distance of visitor	1775.086	3387.683	0	19647.875	892527
Rating of visitor	4.428	0.975	1	5	1544287

Panel D - Monthly data aggregated by country

Variable	Mean	Std.dev.	Min	Max	Observations
No. reviews (in 1000s)	47.434	34.511	2.439	143.339	75
Travel distance	1535.584	631.749	458.622	3527.532	75
Share foreign tourists	34.846	11.966	7.240	54.488	75
Rating	4.396	0.043	4.244	4.494	75
Attraction density (within 5km radius)	1305.357	488.221	316.621	2173.517	75
Attraction density (within 10km radius)	1515.608	548.553	397.261	2478.469	75
Attraction density (within 25km radius)	2076.635	665.808	698.078	3228.884	75
Tourist density (in 1000s) (within 5km radius)	61.068	39.629	2.094	133.448	75
Tourist density (in 1000s) (within 10km radius)	70.923	45.701	2.629	155.970	75
Tourist density (in 1000s) (within 25km radius)	97.291	62.272	4.046	222.763	75
No. arrivals (in 1000s) (Eurostat)	11963.248	6163.543	0.000	24913.408	75
Occupancy rate (Eurostat)	42.886	12.427	10.000	65.000	75
Stringency index	51.669	19.915	5.004	87.960	27
Travel restrictions	2.296	0.912	1.000	3.000	27

Notes: This table shows summary statistics for the data and different units of observation and only for French reviews and attractions. Panel A shows overall numbers. Panel B shows summary statistics using attractions as the unit of observation. Panel C shows summary statistics using the individual users as the unit of observation. Panel D uses monthly aggregated data at the country level. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table A3: Summary statistics Spain

Panel A - Rewiews and attractions

	Reviews	Users	Attractions	Museums	Nature & parks	Sights & landmarks	No. others
Number	3014961	1457028	39114	2474	11504	12064	20222
Share				0.063	0.120	0.308	0.517

Panel B - Attractions and their visitors

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews of attraction	77.08	537.312	1	33795	39114
Foreign visitors per attraction	31.642	319.016	0	23694	39114
Share foreign visitors at attraction	0.376	0.362	0	1	36843
Travel distance to attraction	1386.511	1674.350	0	19655	36864
Rating of attraction	4.235	0.745	1	5	39114

Panel C - Visitors and reviews

Variable	Mean	Std.dev.	Min	Max	Observations
Reviews per visitor	2.069	7.964	1	4959	1457028
Visits abroad	0.586	5.916	0	3868	1457028
Travel distance of visitor	2426.917	3229.284	0	19964.193	847094
Rating of visitor	4.511	0.918	1	5	1456908

Panel D - Monthly data aggregated by country

Variable	Mean	Std.dev.	Min	Max	Observations
No. reviews (in 1000s)	40.076	25.709	2.423	113.461	75
Travel distance	2085.455	613.220	771.970	3272.933	75
Share foreign tourists	52.799	14.306	19.192	69.697	75
Rating	4.459	0.050	4.351	4.599	75
Attraction density (within 10km radius)	857.796	182.166	350.272	1161.803	75
Attraction density (within 25km radius)	1109.735	195.904	536.825	1434.111	75
Attraction density (within 5km radius)	776.612	176.715	298.141	1069.838	75
Tourist density (in 1000s) (within 5km radius)	32.895	21.416	1.669	79.696	75
Tourist density (in 1000s) (within 10km radius)	36.332	23.547	1.913	88.935	75
Tourist density (in 1000s) (within 25km radius)	46.560	29.829	2.599	117.125	75
No. arrivals (in 1000s) (Eurostat)	8839.790	4500.776	0.000	17738.068	75
Occupancy rate (Eurostat)	50.294	18.616	0.000	83.620	75
Stringency index	53.849	19.357	1.111	85.190	27
Travel restrictions	2.593	1.118	0.000	4.000	27

*Notes:*This table shows summary statistics for the data and different units of observation and only for Spanish reviews and attractions. Panel A shows overall numbers. Panel B shows summary statistics using attractions as the unit of observation. Panel C shows summary statistics using the individual users as the unit of observation. Panel D uses monthly aggregated data at the country level. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table A4: Correlation coefficients between Tripadvisor reviews and Eurostat tourism measures

Variable	No. reviews	No. reviews	No. reviews	No. reviews
No arrivals	0.904	0.698	0.861	0.835
	(0.000)	(0.000)	(0.000)	(0.000)
Occupancy rate	0.617	0.831	0.826	0.873
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	Denmark	France	Spain

Notes: Simple correlation coefficients between number of reviews from Tripadvisor and number of arrivals or occupancy rate from Eurostat. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Table A5: Effect of travel restrictions on tourism flows

	Monthly		Daily		Attraction level	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists
Travel restrictions	-0.470***	-7.176***	-0.368***	-9.390***	-0.114***	-4.633***
	(0.053)	(1.047)	(0.052)	(0.800)	(0.002)	(0.086)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes
Attraction FE	No	No	No	No	Yes	Yes
N	900	893	26652	26068	1384706	1155251
R^2	0.818	0.358	0.582	0.256	0.190	0.017

Notes: Regression results when regressing the number of Tripadvisor reviews or the share of foreign tourists on international travel restrictions. Columns 1-2 show the results using monthly aggregated data at the country and attraction category level. Columns 3-4 show the results using daily data aggregated at the country and attraction category level. Columns 5-6 show the results using monthly aggregated data at the attraction level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A6: Effect of travel restriction dummies on tourism flows

	Monthly		Daily		Attraction level	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists
Screening arrivals	-0.229	-0.610	-0.247**	-5.092***	-0.276***	-9.437***
	(0.284)	(6.443)	(0.097)	(1.415)	(0.010)	(0.463)
Quarantine arrival	-0.336	-1.008	-0.403***	-5.975**	-0.277***	-12.730***
	(0.300)	(6.757)	(0.084)	(1.933)	(0.009)	(0.479)
Ban arrivals partly	-1.161***	-20.322***	-0.970***	-27.688***	-0.418***	-19.309***
	(0.211)	(6.444)	(0.155)	(3.302)	(0.009)	(0.409)
Ban arrivals total	-1.740***	-20.003***	-1.482***	-28.607***	-0.638***	-9.476***
	(0.399)	(6.010)	(0.233)	(3.307)	(0.013)	(0.559)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes
Attraction FE	No	No	No	No	Yes	Yes
N	900	893	26652	26068	1384706	1155251
R^2	0.821	0.375	0.583	0.265	0.191	0.018

Notes: Regression results when regressing the number of Tripadvisor reviews or the share of foreign tourists on international travel restriction dummies. Columns 1-2 show the results using monthly aggregated data at the country and attraction category level. Columns 3-4 show the results using daily data aggregated at the country and attraction category level. Columns 5-6 show the results using monthly aggregated data at the attraction level. All specifications include a series of fixed effects, for more details see the text. No restrictions imposed is the reference category. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A7: Effect of travel restrictions on tourism of different attraction categories

	Monthly		Daily		Attraction level	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists	ln(Reviews)	Share foreign tourists
Travel restrictions	-0.404***	-9.355***	-0.396***	-10.383***	-0.103***	-5.076***
	(0.094)	(1.552)	(0.120)	(1.468)	(0.006)	(0.207)
Museums × Travel restrictions	-0.165	2.709	0.057	1.423	-0.035***	0.513
	(0.120)	(1.990)	(0.169)	(2.214)	(0.009)	(0.343)
Sights & Landmarks × Travel restrictions	0.018	5.836**	0.121	-0.179	-0.022***	0.465*
	(0.153)	(1.995)	(0.142)	(2.307)	(0.007)	(0.255)
Others × Travel restrictions	-0.116	0.643	-0.063	2.795*	-0.004	0.561**
	(0.099)	(2.032)	(0.147)	(1.512)	(0.006)	(0.233)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes
Attraction FE	No	No	No	No	Yes	Yes
N	900	893	26652	26068	1384706	1155251
R^2	0.819	0.361	0.583	0.257	0.190	0.017

Notes: Regression results when estimating the number of Tripadvisor reviews or the share of foreign tourists on international travel restrictions interacted with attraction category dummies. The reference category is Nature & Parks. Columns 1-2 show the results using monthly aggregated data at the country and attraction category level. Columns 3-4 show the results using daily data aggregated at the country and attraction category level. Columns 5-6 show the results using monthly aggregated data at the attraction level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A8: Effect of stringency index on other attraction and tourist density measures

	(1) Attraction density 5km	(2) Attraction density 25km	(3) ln(Tourist density) 5km	(4) ln(Tourist density) 25km
Stringency index	-3.430***	-3.553**	-0.033***	-0.030***
	(1.024)	(1.395)	(0.001)	(0.001)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes
N	900	900	900	900
R^2	0.536	0.489	0.907	0.910

Notes: Regression results when regressing the two density measures on the stringency index. Columns 1-2 show the results for the attraction density using $5km$ and $25km$ radiuses respectively. Similarly, columns 3-4 show the results for the tourist density measure. The results refers to the monthly data aggregated at the country and attraction category level. See the text for an explanation of how the density measures have been computed. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Table A9: Effect of travel restrictions on travel distance, attraction density, tourist density and ratings

	(1) Distance	(2) Attraction density) 5km	(3) Attraction density 10km	(4) Attraction density 25km	(5) ln(Tourist density) 5km	(6) ln(Tourist density) 10km	(7) ln(Tourist density) 25km	(8) Rating
Travel restrictions	-231.659***	-67.786***	-68.506***	-73.224***	-0.665***	-0.646***	-0.606***	0.008
	(26.708)	(16.567)	(18.117)	(20.788)	(0.020)	(0.018)	(0.014)	(0.022)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	900	900	900	900	900	900	900	900
R^2	0.452	0.536	0.515	0.490	0.915	0.916	0.915	0.121

Notes: Regression results when estimating the travel distance, attraction density, tourist density or ratings on international travel restrictions. The results refers to the monthly data aggregated at the country and attraction category level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A10: Effect of travel restriction dummies on travel distance, attraction density, tourist density and ratings

	(1) Distance	(2) Attraction density 5km	(3) Attraction density 10km	(4) Attraction density 25km	(5) ln(Tourist density) 5km	(6) ln(Tourist density) 10km	(7) ln(Tourist density) 25km	(8) Rating
Screening arrivals	-244.416	46.904	54.865	70.103	-0.420**	-0.431**	-0.434**	0.047
	(137.160)	(105.231)	(116.545)	(136.085)	(0.154)	(0.154)	(0.152)	(0.125)
Quarantine arrival	-495.557**	112.818	135.340	164.054	-0.607***	-0.613***	-0.615***	0.019
	(171.510)	(108.152)	(120.131)	(141.698)	(0.147)	(0.145)	(0.142)	(0.138)
Ban arrivals partly	-975.575***	-170.160***	-171.038***	-177.080**	-1.666***	-1.635***	-1.540***	0.070
	(165.533)	(45.668)	(52.246)	(60.383)	(0.018)	(0.022)	(0.025)	(0.115)
Ban arrivals total	-666.345***	-130.476*	-120.005	-118.188	-2.564***	-2.503***	-2.380***	0.022
	(140.543)	(67.374)	(74.184)	(85.711)	(0.062)	(0.062)	(0.075)	(0.146)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heritage type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	900	900	900	900	900	900	900	900
R^2	0.474	0.548	0.527	0.501	0.921	0.921	0.920	0.123

Notes: Regression results when estimating the travel distance, attraction density, tourist density or ratings on the international travel restrictions dummies. No travel restrictions imposed is the reference category. The results refers to the monthly data aggregated at the country and attraction category level. All specifications include a series of fixed effects, for more details see the text. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A11: Effect of travel restrictions on different outcomes in the global sample

	(1) ln(Visits)	(2) ln(Visits)	(3) ln(Visits abroad)	(4) ln(Visits abroad)	(5) Distance	(6) Distance	(7) Rating	(8) Rating
Travel restrictions	-0.094***	-0.030***	-0.098***	-0.026***	-381.321***	-248.302***	0.004***	0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(3.092)	(5.626)	(0.001)	(0.001)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	No	Yes	No	Yes	No	Yes	No	Yes
N	1813458	1813410	1813458	1813410	1507250	1507202	1804911	1804866
R^2	0.416	0.427	0.363	0.384	0.395	0.522	0.216	0.218

Notes: Regression results using the global sample of attractions and different outcomes of interest. The data is at the individual level aggregated monthly. Columns 1-2 uses number of visits as the outcome of interest, columns 3-4 the number of visits abroad, columns 5-6 the distance travelled by individuals and columns 7-8 the rating given to an attraction by an individual. Columns 1,3,5,7 only includes individual fixed effects while columns 2,4,6,8 also add country, year and month fixed effects. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

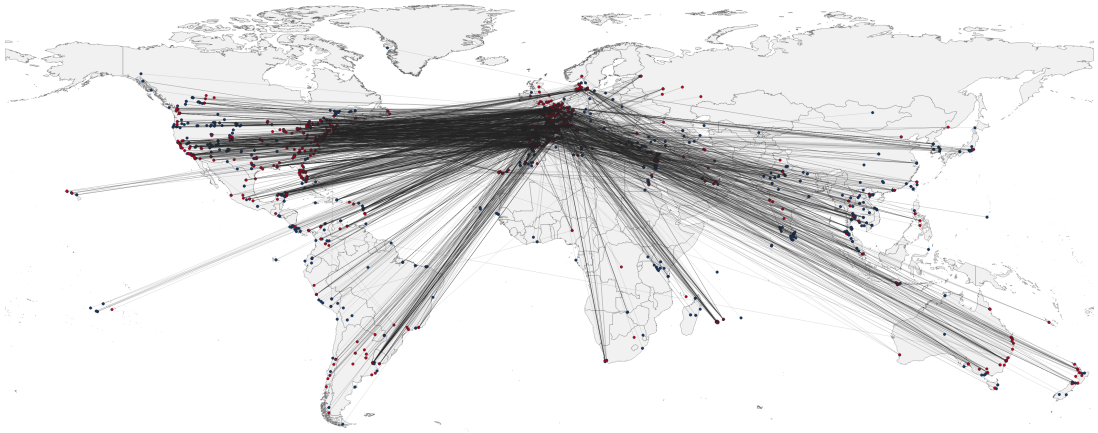
Table A12: Effect of travel restriction dummies on different outcomes in travel history module

	(1) ln(Visits)	(2) ln(Visits)	(3) ln(Visits abroad)	(4) ln(Visits abroad)	(5) Distance	(6) Distance	(7) Rating	(8) Rating
Screening arrivals	-0.292***	-0.011**	-0.323***	-0.011	-542.951***	-135.357***	0.010***	-0.006
	(0.003)	(0.005)	(0.006)	(0.009)	(12.814)	(18.461)	(0.004)	(0.006)
Quarantine arrival	-0.258***	-0.056***	-0.302***	-0.034***	-888.747***	-302.154***	-0.011**	-0.016***
	(0.004)	(0.005)	(0.009)	(0.011)	(12.662)	(18.315)	(0.005)	(0.006)
Ban arrivals partly	-0.262***	-0.069***	-0.274***	-0.076***	-1094.116***	-528.661***	0.008***	-0.003
	(0.002)	(0.004)	(0.005)	(0.008)	(8.666)	(17.251)	(0.003)	(0.005)
Ban arrivals total	-0.367***	-0.190***	-0.241***	-0.099***	-1654.140***	-1687.142***	0.054***	0.027***
	(0.008)	(0.009)	(0.015)	(0.016)	(49.081)	(48.201)	(0.007)	(0.008)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	No	Yes	No	Yes	No	Yes	No	Yes
N	1813458	1813410	1813458	1813410	1507250	1507202	1804911	1804866
R^2	0.417	0.427	0.364	0.384	0.395	0.523	0.216	0.218

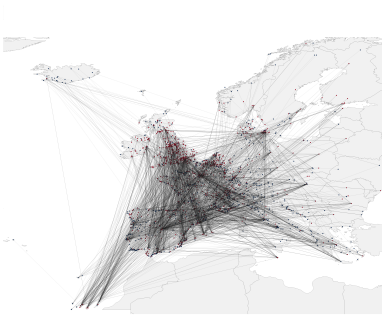
Notes: Regression results using the global sample of attractions and different outcomes of interest. The reference category is no travel restrictions imposed. The data is at the individual level aggregated monthly. Columns 1-2 uses number of visits as the outcome of interest, columns 3-4 the number of visits abroad, columns 5-6 the distance travelled by individuals and columns 7-8 the rating given to an attraction by an individual. Columns 1,3,5,7 only includes individual fixed effects while columns 2,4,6,8 also add country, year and month fixed effects. Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

A.2 Figures

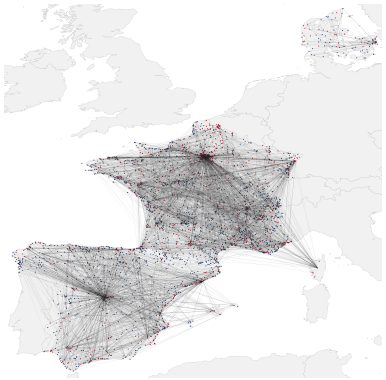
Figure A1: Travel patterns



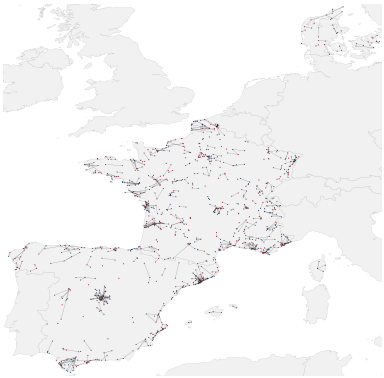
(A) World



(B) Europe



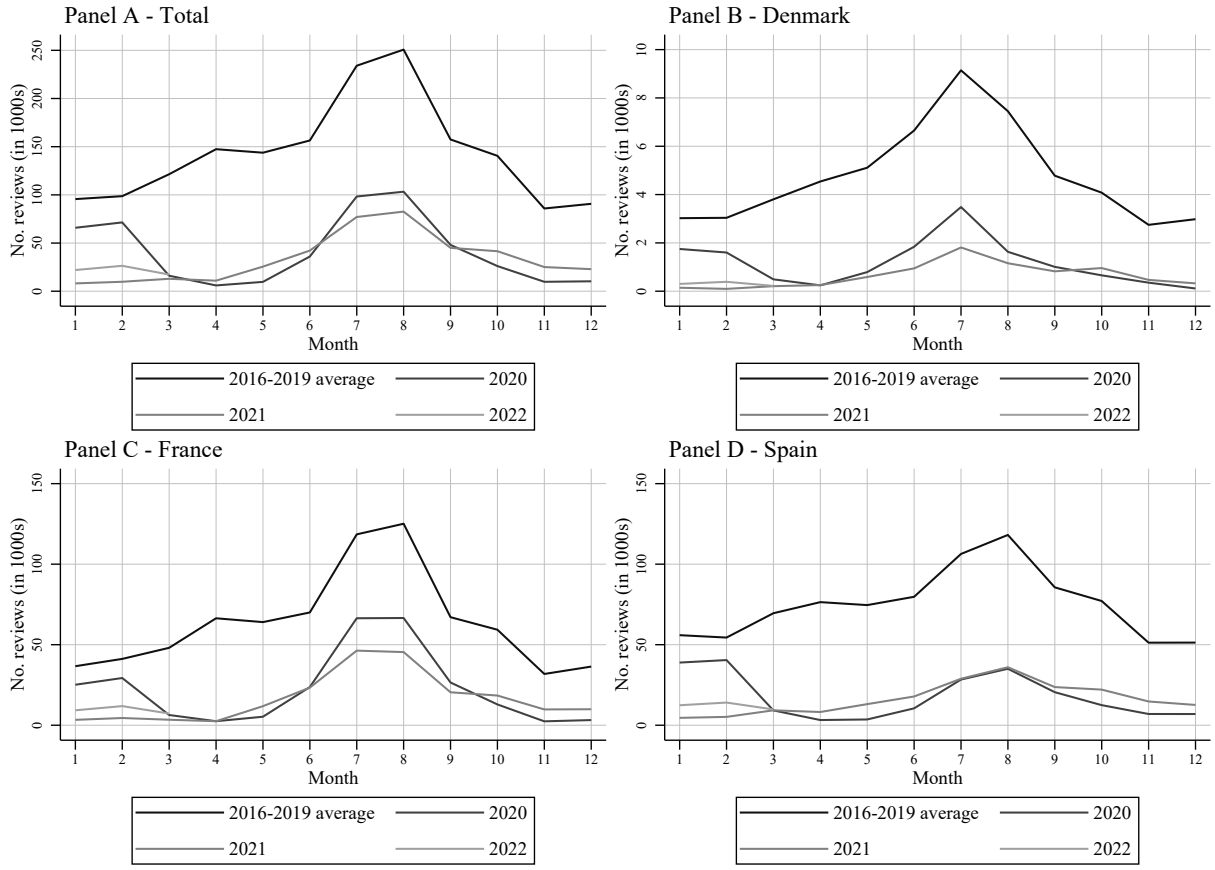
(C) Domestic



(D) Local

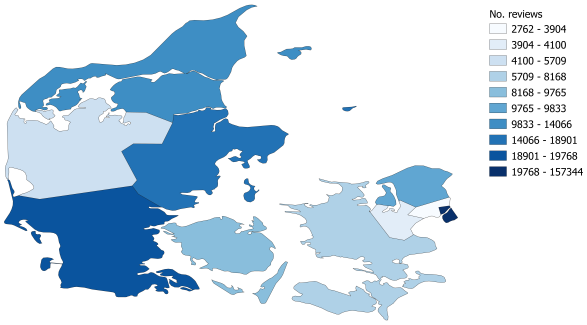
Notes: This figures shows travel patterns of a 1% sample of all reviews written by tourists visiting attractions in Denmark, France and Spain. The red dots indicate the location of origin of the tourist, while the blue dots indicate the location of an attraction. Panel A illustrates the travel patterns of tourists from outside Europe, Panel B shows the travel patterns of tourists within Europe and Panel C consider domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 regions. *Source*: Own data collected from Tripadvisor (see Section 3 for details)

Figure A2: Number of reviews by month and country

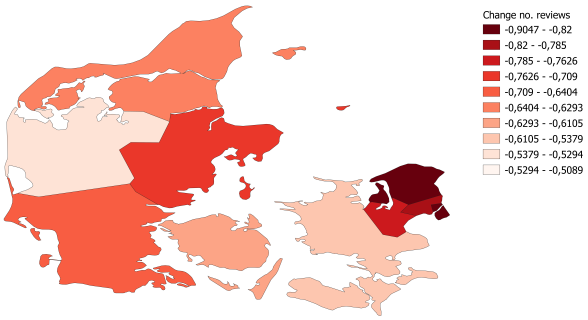


Notes: This Figure shows number of Tripadvisor reviews by country and month. For the years 2016-2019 the average has been computed. Panel A, shows the entire sample, Panel B for Denmark, Panel C for France and Panel D for Spanin. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

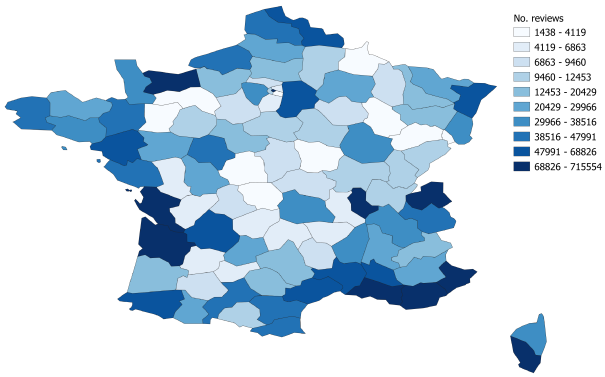
Figure A3: Number of reviews and change



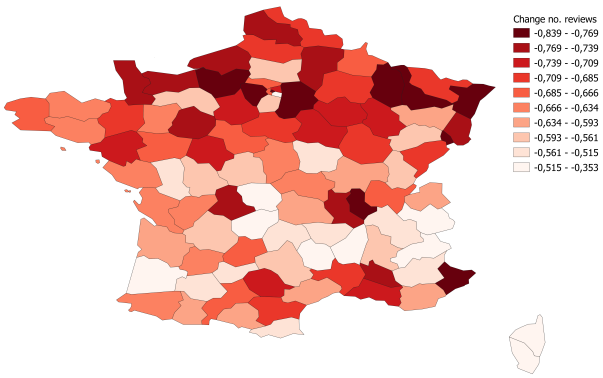
(A) Number of reviews - Denmark



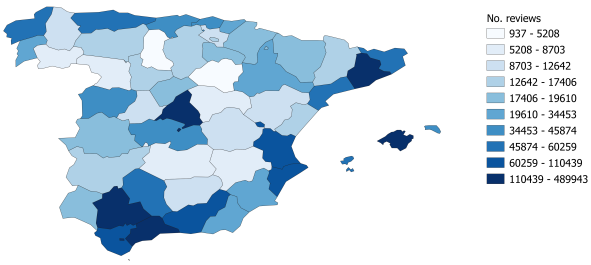
(B) Change no. reviews (%) - Denmark



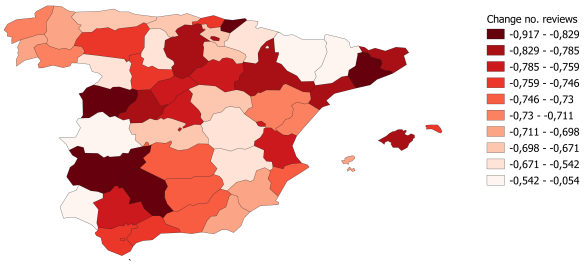
(C) Number of reviews - France



(D) Change no. reviews (%) - France



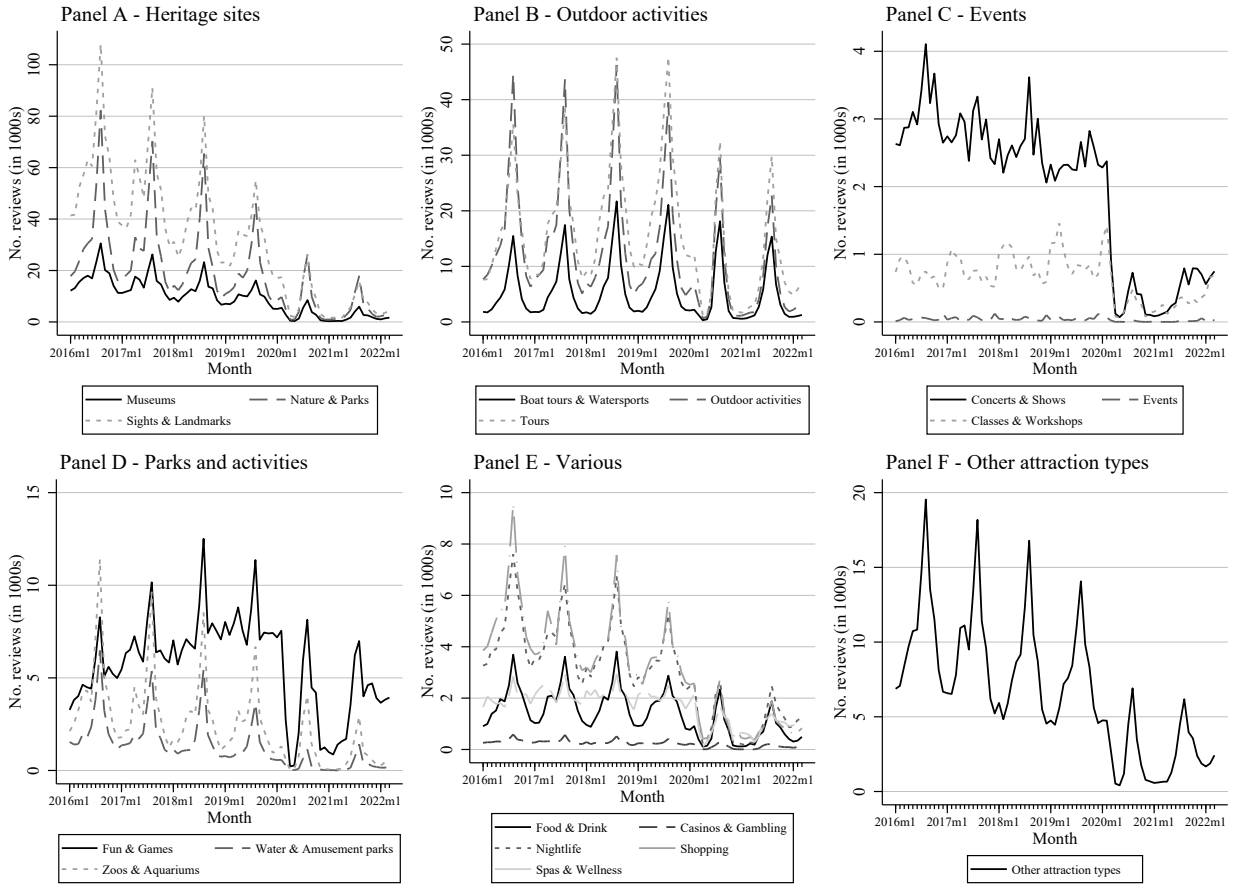
(E) Number of reviews - Spain



(F) Change no. reviews (%) - Spain

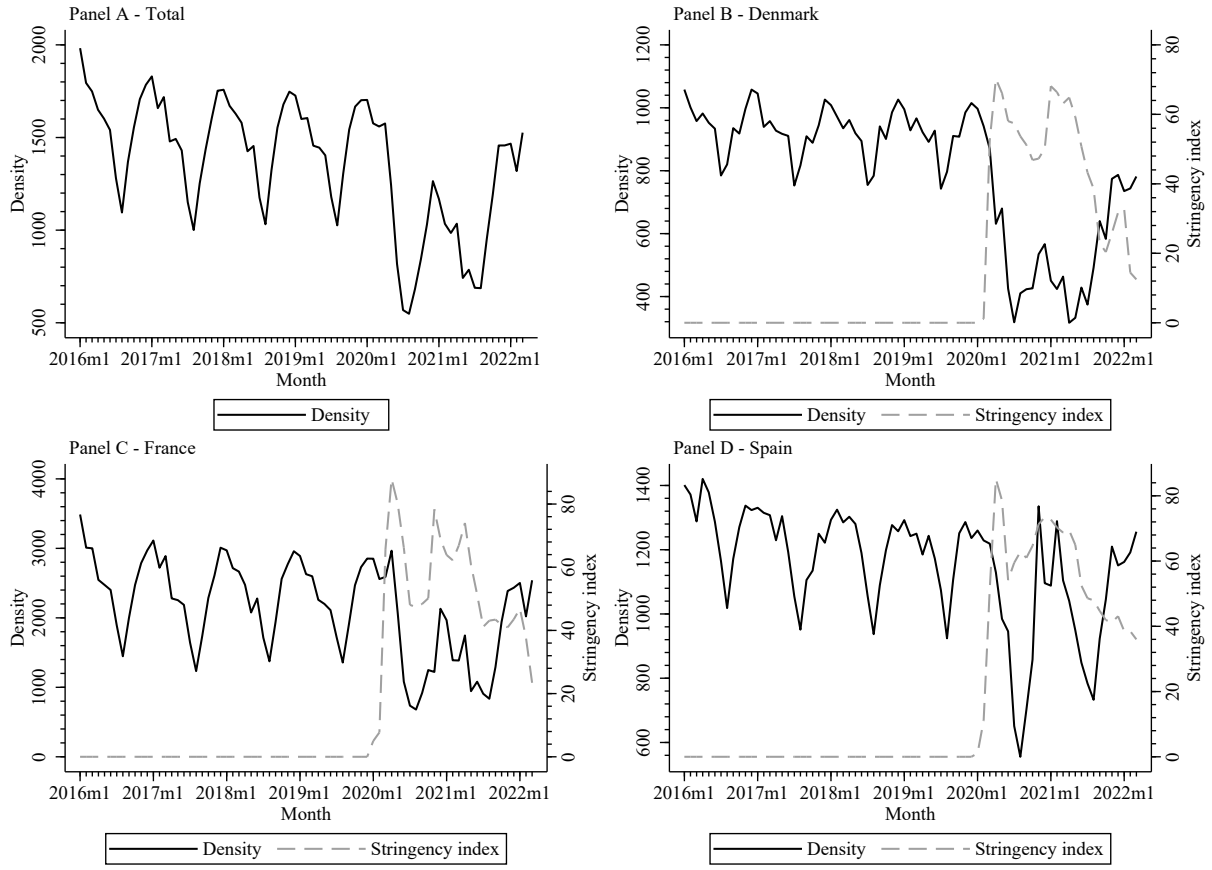
Notes: This figures shows the total number of reviews together with the percentage change in the number of reviews between before and after 2020 by NUTS3 regions. The percentage change is obtained by taking the yearly average of the number of reviews for the two periods 2016-2019 and 2020-2021 and computing the percentage change. Panels A, A and E show the total number of reviews while Panels B, D and F show the percentage change. *Source*: Own data collected from Tripadvisor (see Section 3 for details)

Figure A4: Number of reviews over time by different attraction categories



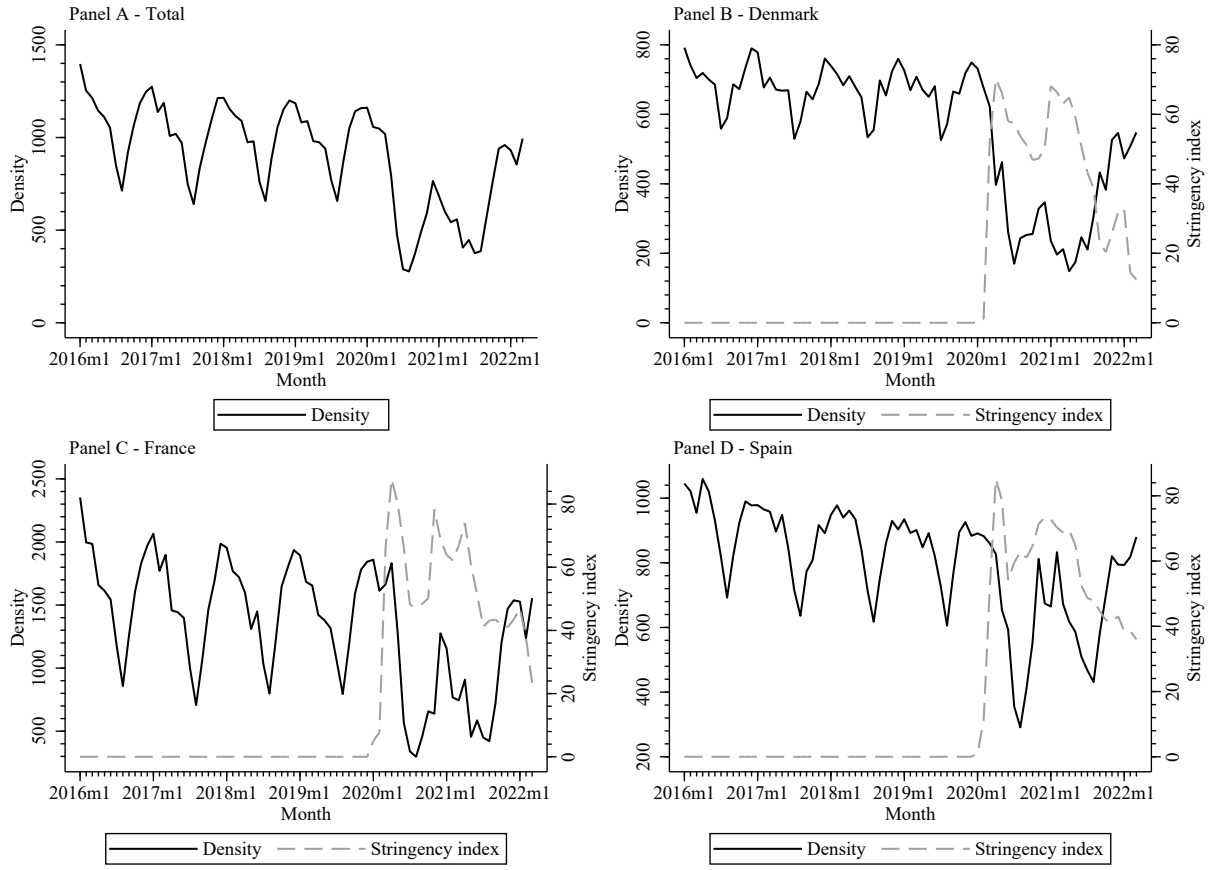
Notes: This Figure shows the change in the number of Tripadvisor reviews over time for different attraction categories defined by Tripadvisor. Each category is not mutually exclusive, and hence the same review can appear in multiple attraction categories. Panel A, shows the number of reviews for different attraction types of cultural and natural heritage. Panel B shows different outdoor attractions, Panel C attractions categorised as various events, Panel D different activities including amusement parks and Panel E shows attraction categories of other various kinds. Panel F shows the remaining category of attractions categorized as "Other" including transportation and travel resources. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

Figure A5: Attraction density of visited locations - 25km radius



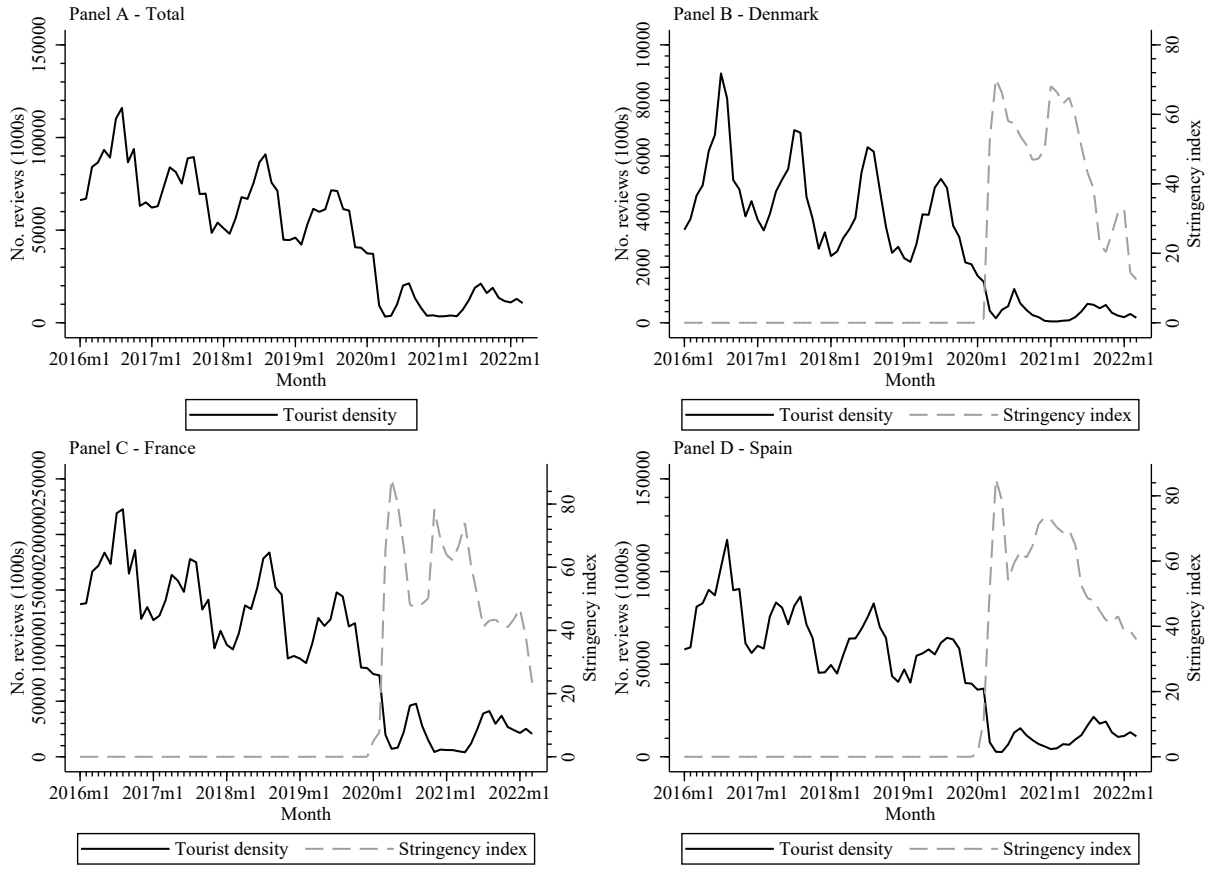
Notes: This Figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the attraction density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. An attraction's density is measured as the number of other attractions within a radius of 25km. The overall density is the average of all attractions' densities in a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure A6: Attraction density of visited locations - 5km radius



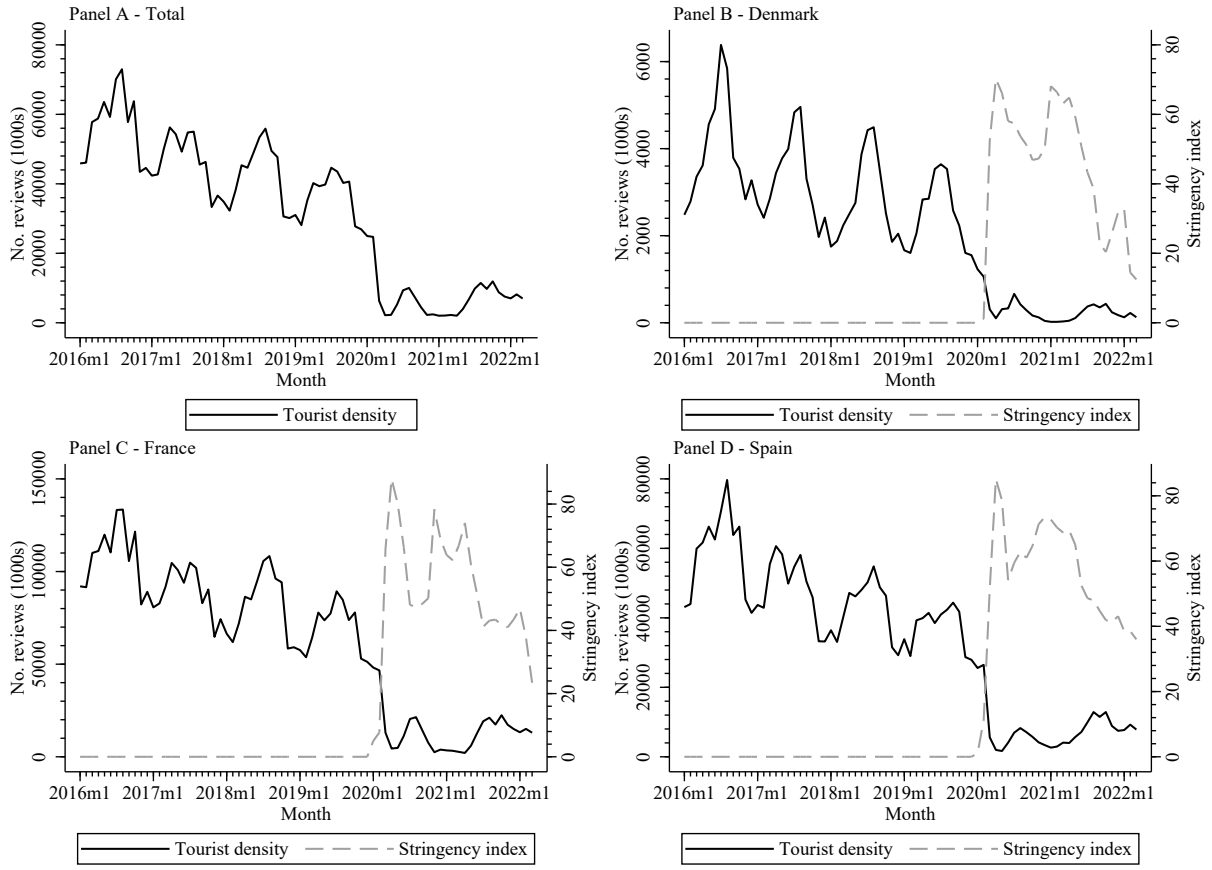
Notes: This Figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the attraction density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. An attraction's density is measured as the number of other attractions within a radius of 5km. The overall density is the average of all attractions' densities in a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure A7: Tourist density of visited locations - 25km radius



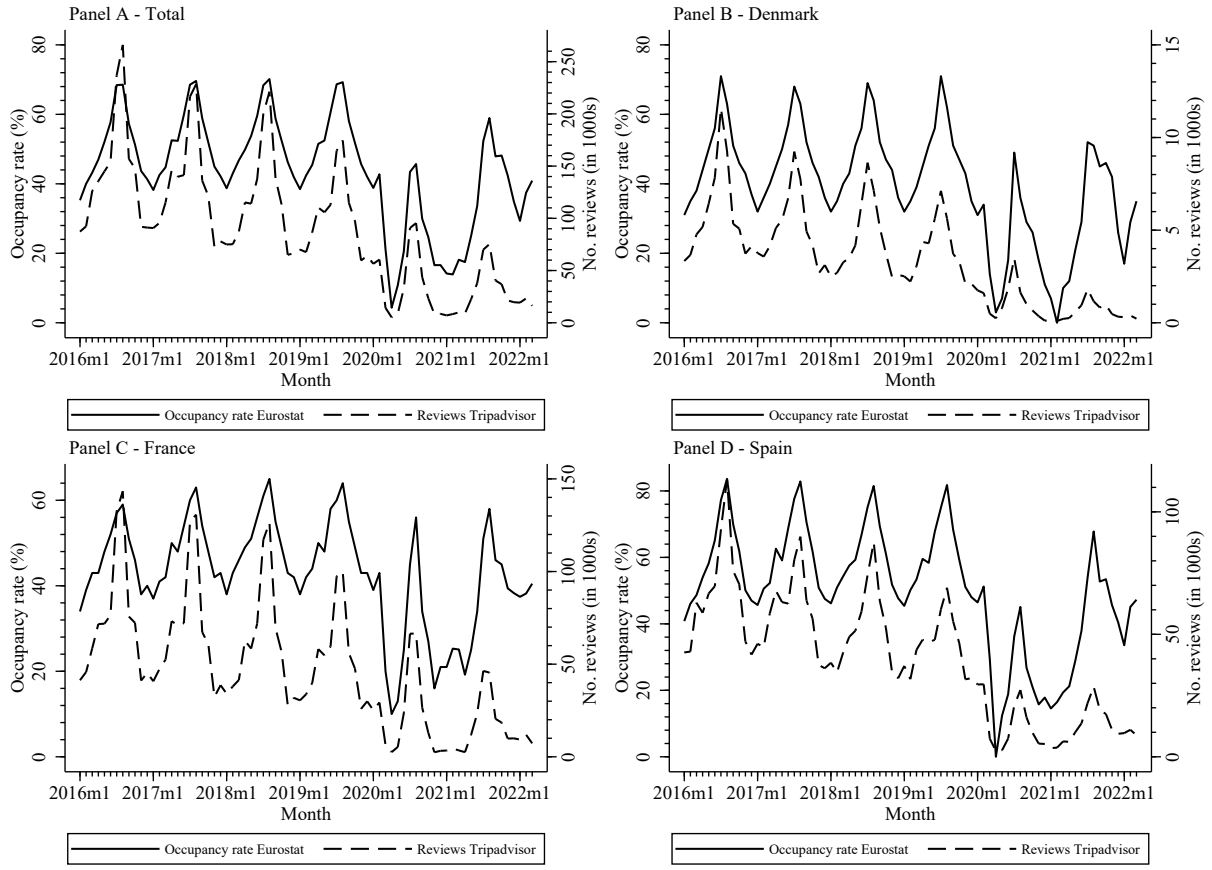
Notes: This Figure shows the change in the tourist density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the review density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. The review density of an attraction is computed as the total number of reviews of all attractions within a radius of 25km within a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure A8: Tourist density of visited locations - 5km radius



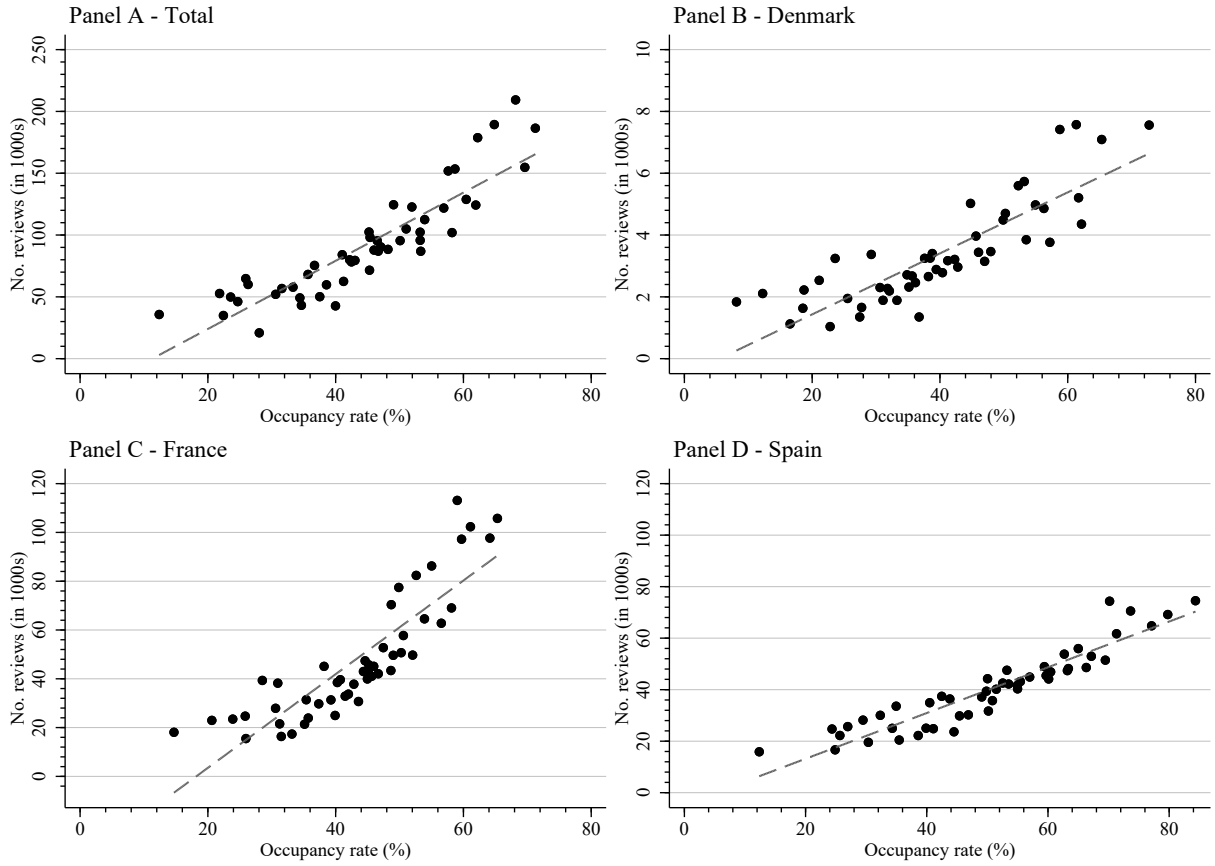
Notes: This Figure shows the evolution of the tourist density of visited locations together with the stringency index. Panel A, shows the entire sample, Panel B, shows the review density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. The review density of an attraction is computed as the total number of reviews of all attractions within a radius of 5km within a given month. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker.

Figure A9: Validity test: Occupancy rates and number of reviews over time



Notes: This Figure shows the occupancy rates taken from Eurostat together with the total number of Tripadvisor reviews. Panel A for the entire sample, and Panels B-D by country. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).

Figure A10: Validity test: Monthly correlation between occupancy rates and number of reviews



Notes: This Figure shows binned scatter plots of the occupancy rates taken from Eurostat and the number of Tripadvisor reviews. Panel A uses the entire sample, while Panels B-D by country. The correlation coefficient corresponding to the correlation in Panel A is 0.039, in panel B it is 0.041, in Panel C it is 0.063 and in Panel D it is 0.038.*Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 3 for details).