Article

TOURISM ECONOMICS
The Business and Finance of Tourism and Recreation

Tourism Economics
2025, Vol. 31(3) 477–503

DOI: 10.1177/13548166241273604
journals.sagepub.com/home/teu
Sage

Using big data to measure cultural tourism in Europe

Karol Jan Borowiecki, Maja U Pedersen and Sara Beth Mitchell
University of Southern Denmark, Denmark

Abstract
International tourism statistics are notorious for being overly aggregated, lacking in detailed tourist information, not timely, and often provided only on an annual basis. We suggest a unique, complementary data-driven approach relying on big data collected from Tripadvisor. We obtain a systematic, consistent, and reliable approximation for tourism flows, with high precision, frequency, and depth of information. The approach provides also a list of all tourist attractions in a country. We validate 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 tourism economics and cultural economics.

Keywords
big data, COVID-19, cultural heritage, tourism

JEL classification
J60, L83, O1, Z11, Z3

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,

Corresponding author:
Karol Jan Borowiecki, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark.
Email: kjb@sam.sdu.dk

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 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 (2023a). 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 traveling 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 travelers 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 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 labor force (European Commission, 2023). 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., Bertacchini et al., 2023; Castillo-Manzano et al., 2021; Cuccia et al., 2016), but little is known how less established sites attract visitors. Some have asked whether the World Heritage List "makes 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.

This study introduces key conceptual innovations in the field of tourism studies. It challenges traditional perceptions of tourist behavior by integrating the concept of dynamic preference shifts, demonstrating that tourist preferences evolve in response to global events, such as the COVID-19 pandemic. This compels a rethinking of established tourism theories, especially those related to destination choice and satisfaction, suggesting a need for dynamic and adaptable models that incorporate these shifting preferences due to external factors like health crises, environmental concerns, and digital influences.

Furthermore, we reinterpret tourist attraction popularity as a multifaceted construct, influenced by online reviews and digital presence. This insight extends the conceptual framework of cultural tourism, underscoring the importance of digital engagement in determining the popularity and appeal of heritage sites. Our approach also introduces a novel perspective on sustainable tourism. By analyzing tourist flows and attraction characteristics, we offer a more nuanced understanding of sustainable tourism, emphasizing the quality and depth of tourist experiences over mere visitor numbers.

Our study also introduces, develops and tests novel supply-side measures in tourism research. Moving beyond traditional constraints of limited and aggregated data, our approach compiles a comprehensive list of attractions, including lesser-known sites, providing a more granular view of the tourism landscape. This helps in understanding how different attractions contribute to and shape tourist behavior and preferences.

Incorporating innovative metrics from user-generated content, such as attraction entries and visitor reviews, we develop new metrics that capture qualitative aspects of tourist experiences. This provides a more authentic and nuanced picture of tourist satisfaction and attraction quality, enhancing our understanding of tourist behavior and capturing evolving trends in the sector. Our methodological approach in collecting and analyzing big data in tourism marks a significant departure from traditional methods. The comprehensive nature of our data collection allows for highly disaggregated, detailed, and expansive insights into tourism patterns. We rigorously validate our approach against established tourism statistics, ensuring the credibility of our methodology and its potential for broader implementation in tourism research.

In summary, our study not only advances theoretical understanding but also paves the way for more informed and effective management of tourism resources and strategies. By incorporating notions of dynamic tourist preferences, multifaceted attraction popularity, and a nuanced view of sustainable tourism, along with a robust methodology for data collection and analysis, we set new directions for theoretical exploration and practical applications in tourism studies.

The rest of the paper is organized 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.

Literature review

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

Tourism is more than traveling 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 (Noonan and Rizzo, 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 Hagsten, 2017; Falk and Katz-Gerro, 2017). 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 (e.g., Falk et al., 2022).

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 on Flickr not only in parks but also in city centers (see Kádár, 2014; Sessions et al., 2016; Shi et al., 2017; Sinclair et al., 2020; 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.

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 (see e.g. Falk et al., 2022; Fotiadis et al., 2021; Sigala, 2020; Singh, 2021; Zenker and Kock, 2020). 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. Traveling can often be connected to some kind of perceived health risk which affects tourism behavior (Fenichel et al., 2013; Jonas et al., 2011; Lepp and Gibson, 2003; Reisinger and Mavondo, 2005) and affect both the economics of the tourism industry and the tourists behavior (Boto-García and Leoni, 2023; Milone et al., 2023; Yang et al., 2020; Zhang et al., 2020). In Kock et al. (2020) they develop a new model, the Evolutionary Tourism Paradigm, to analyze tourism behavior during pandemics. 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. Previous studies using Tripadvisor as a source of data have focused on topics such as consumers' perceived quality of attractions, incentives to create fake reviews, and rating system design. In Mayzlin et al. (2014) they compare reviews posted on Tripadvisor with those posted on the booking platform Expedia. Their findings suggest 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. Nguyen et al. (2020) use Tripadvisor data to confirm the restraint-of-expertise hypothesis which entails that reviewing experts are less willing to give extreme ratings compared to novices. Other studies have used Tripadvisor reviews to investigate tourism related to the consumption of food and wine (see Thanh and Kirova, 2018; Waldfogel, 2020). 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. Finally, a series of other studies also use Tripadvisor ratings (e.g. Banerjee and Chua, 2016; Chen et al., 2018; Hollenbeck et al., 2019). 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.

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.

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. This database has, for example, been used in Boto-García and Leoni (2023) to estimate the effects of sociodemographic and trip-related characteristics on the distance traveled by domestic tourists before and after the COVID-19 pandemic. 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. In Borowiecki et al. (2024, 2025) we use a similar approach but with a different set of countries.

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. In terms of selection of these countries, most important is to note that they were determined independently of the research question. The choice was influenced by the project's funding from the Mobile Lives Forum, based in France, and the participation of scientific members from Spain, while the authors are affiliated with an institution in Denmark. These countries - representing Southern, Western, and Northern Europe - offer a diverse geographical, cultural, and socioeconomic spectrum across the continent, enriching the comparative dimension of the research.

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, 2023b). 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 not mutually exclusive. 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.

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 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 travel distance between the user writing the review and the attraction visited, a ”foreign” dummy which equals one when a review 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 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 10 *km* as a measure of density.[1] 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 10 *km* from the attraction.[2]

In the Appendix, Table A.I Panels A-C presents descriptive statistics by attractions and users for the entire sample while descriptive statistics by country can be seen in Table A.II-A.IV. In Table A.I Panel 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 A.I, 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 A.I in the Appendix shows the summary statistics of the global sample using the travel history module without any aggregation. 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.[3] In Figure A.II 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.

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 government 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, which in most cases is representative also for the local level given that restrictions were mainly imposed at the national level. In Milone et al. (2023) they use the Stringency Index as an instrument for local Airbnb demand to explain prices and find a significant negative relationship between the Stringency Index and Airbnb demand. Along these lines, we make use of the Stringency Index and an indicator of international travel restrictions

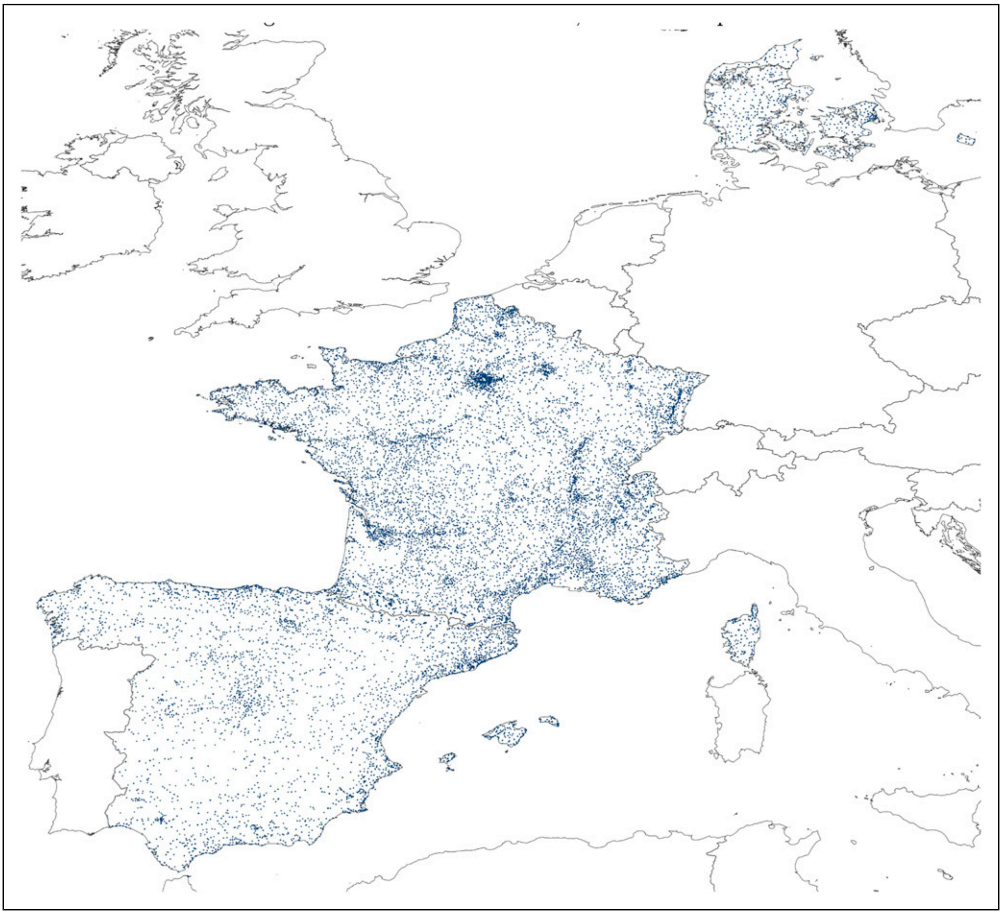

Figure 1. Attractions in Denmark, France and Spain.
Notes: This Figure shows the location of all attractions present on Tripadvisor and located in Denmark, France or Spain. *Source:* own data collected from Tripadvisor (see Section 3 for details).

to estimate the effect on different measures of tourism.[4] The stringency index is composed of nine individual indicators, I_j, each assigned with a score and re-scaled between 0 and 100.[5] 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 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.

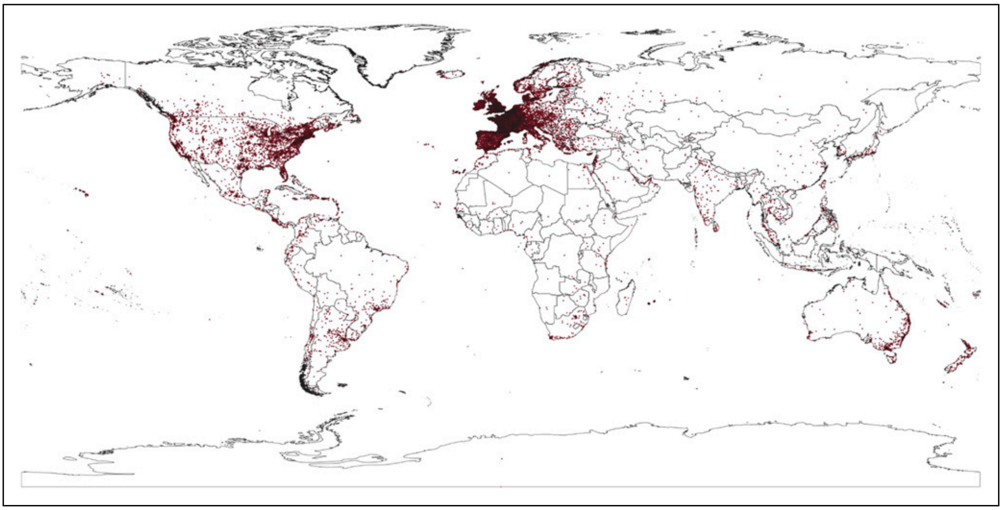

Figure 2. Location of visitors.
Notes: This Figure shows the location of visitors who have written at least one review on Tripadvisor of a Danish, French or Spanish attraction. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

Table A.I, 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 3 we show the evolution of the average number of reviews over time together with the Stringency Index measure. From this Figure it becomes clear that there is a sharp decrease in the number of reviews beginning in February 2020 when the Stringency Index goes up, and what follows is an inverse relationship between the number of reviews and the Stringency Index.[6,7]

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 traveled by all tourists. The evolution over time of the number of foreign and domestic tourists can be seen in Figure 4 while the average distance traveled can be seen in Figure 5. 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 6 when using a radius of 10 *km*.[8] 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 7 we show the change in tourist density over time when using a 10 *km* radius.[9]

In our main specifications we have three different levels of aggregation all using the attractions as 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.

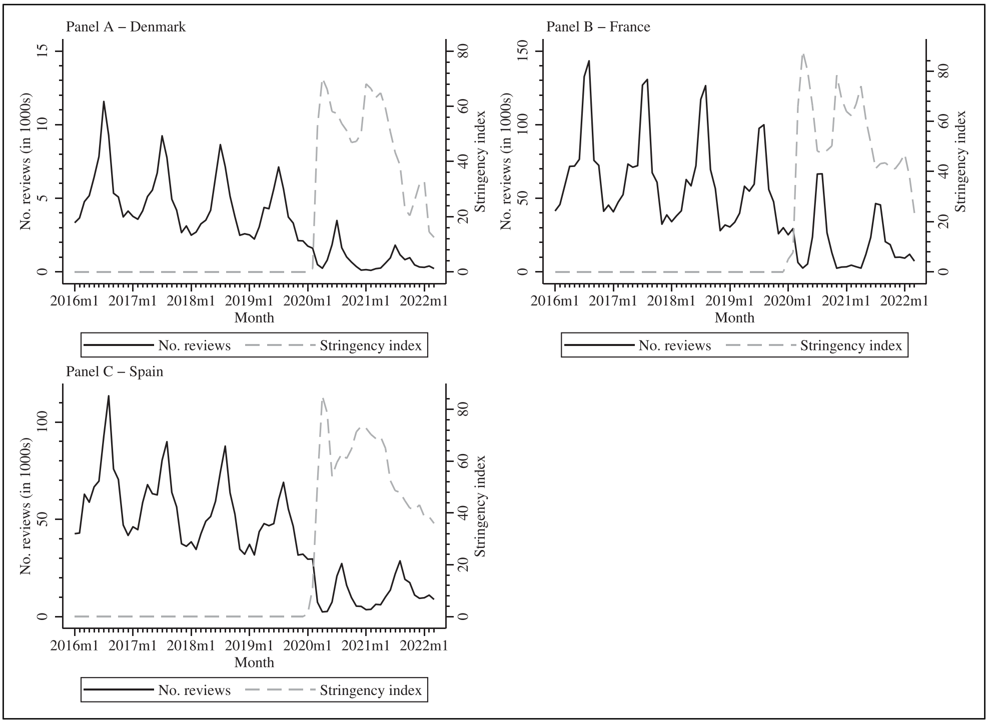

Figure 3. Number of reviews and Stringency Index over time by country.
Notes: This Figure shows the evolution of the number of Tripadvisor reviews over time together with the stringency index. Panel A shows the number of reviews for Danish attractions, panel B for French attractions and panel C for Spanish attractions. *Source:* own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

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.

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 8 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.[10]

As a second visual inspection, Figure 9 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.[11]

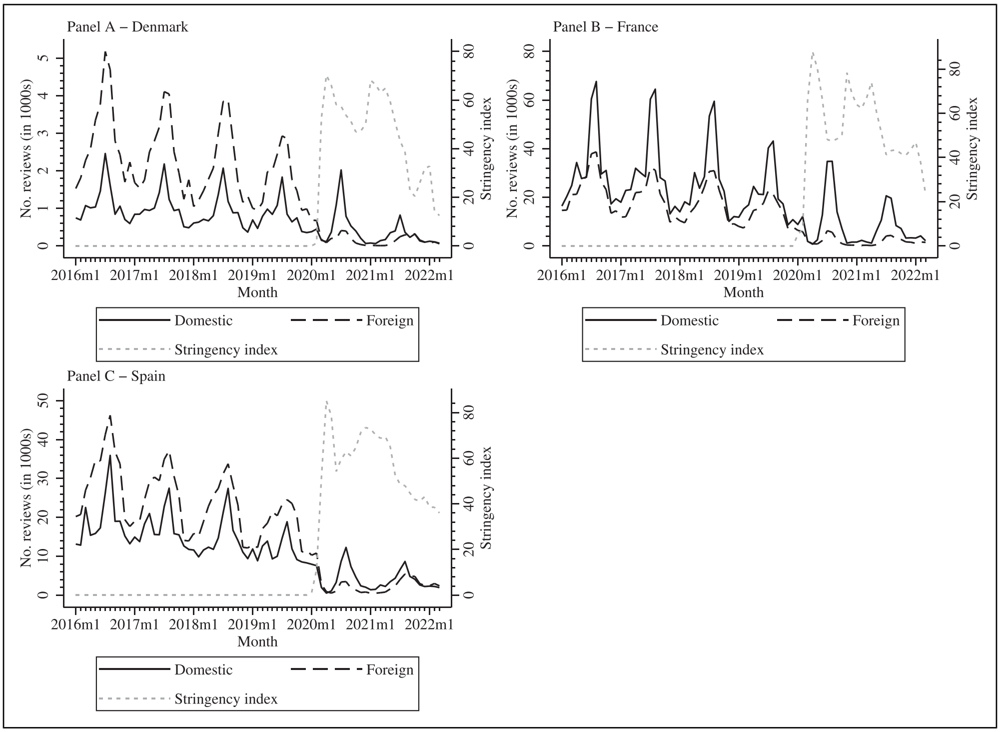

Figure 4. Number of domestic and foreign tourists by country over time.
Notes: This Figure shows the evolution of the number of reviews written by domestic and foreign tourists over time together with the stringency index. Panel A shows the number of reviews for Danish attractions, Panel B for French 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 (Hale et al., 2021).

As a more formal test we estimate how well tourism arrivals or occupancy rates can explain the number of monthly reviews from Tripadvisor.[12] The results can be seen in Table A.VII in the Appendix, 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 Figure 10. 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 A.VII we are confident that our data is a valid alternative to using official tourism statistics and we therefore proceed with our analysis.

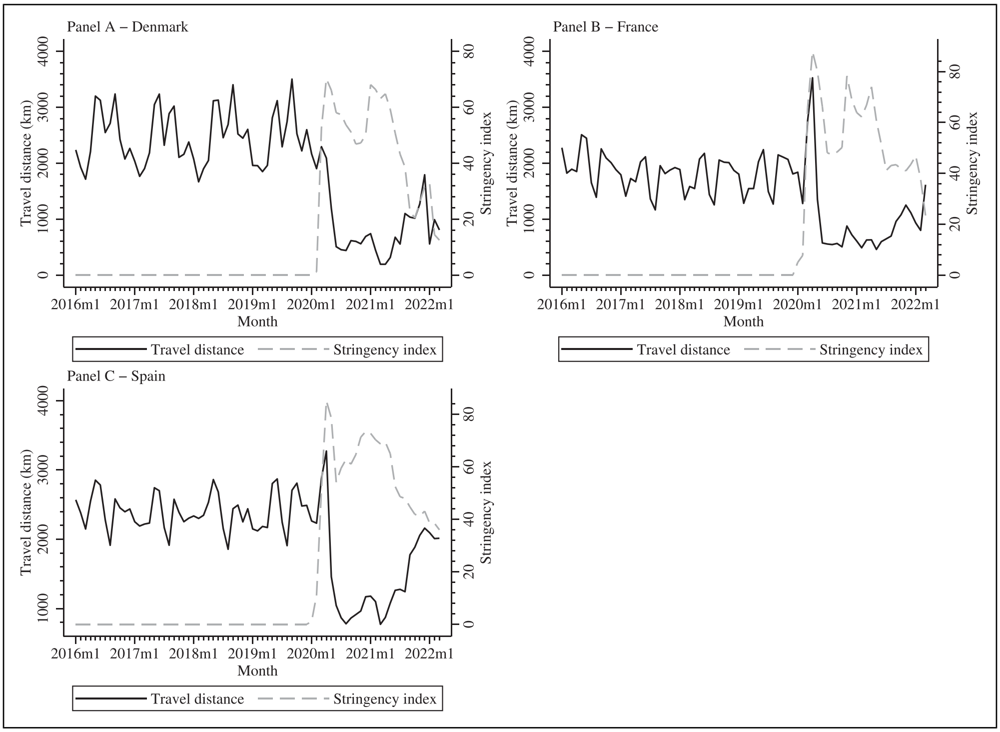

Figure 5. Travel distance to attractions and Stringency Index by country.
Notes: This Figure shows the evolution of the distance traveled to attractions together with the stringency index. Panel A shows the travel distance for Danish attractions, Panel B for French attractions and Panel C for Spanish attractions. *Source:* Own data collected from Tripadvisor (see Section 3 for details) and the stringency index from the Oxford government response tracker (Hale et al., 2021).

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 utilize an expanded dataset that includes all reviews of any attraction worldwide, written by users from our baseline dataset. Economic theory presents various models to understand tourism demand, such as gravity models (e.g., Morley et al., 2014) or the tourism attractiveness model (e.g., Dwyer and Kim, 2003). However, our study's objective differs; we aim to assess the impact of changes in the Oxford Stringency Index on tourism rather than developing a demand model to explore the effects of various explanatory variables. Consequently, our empirical strategy adopts a unique approach, which we describe below.

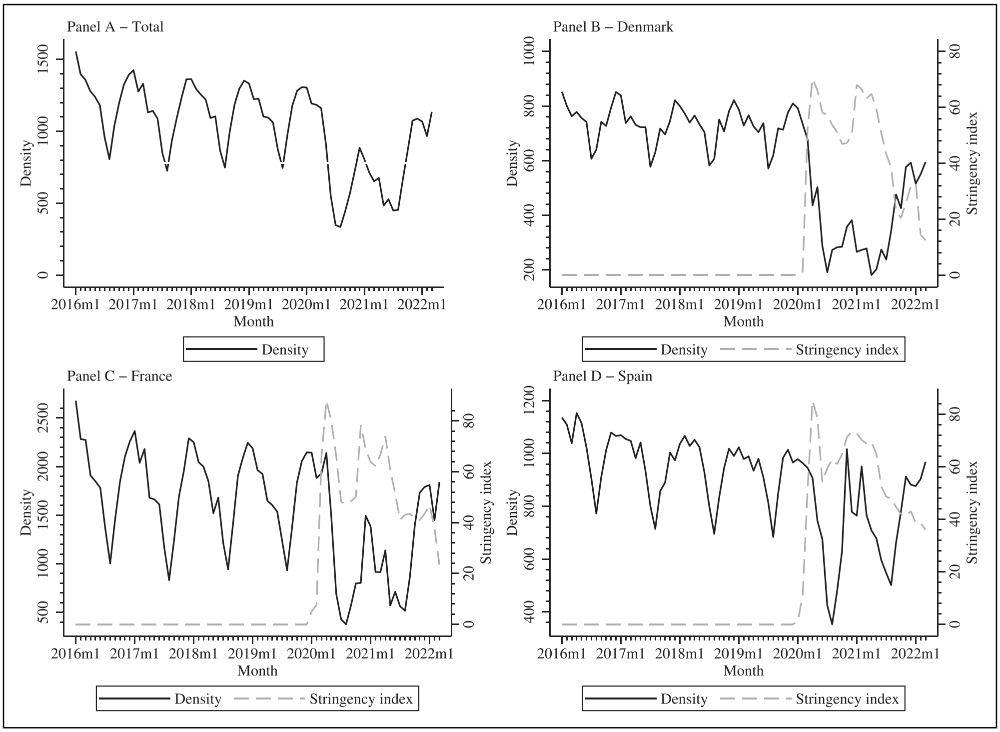

Figure 6. Attraction density of visited locations.
Notes: This Figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A shows the entire sample, Panel B shows the attraction density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. An attraction's density is measured as the number of other attractions within a radius of 10 km. 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 (Hale et al., 2021).

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

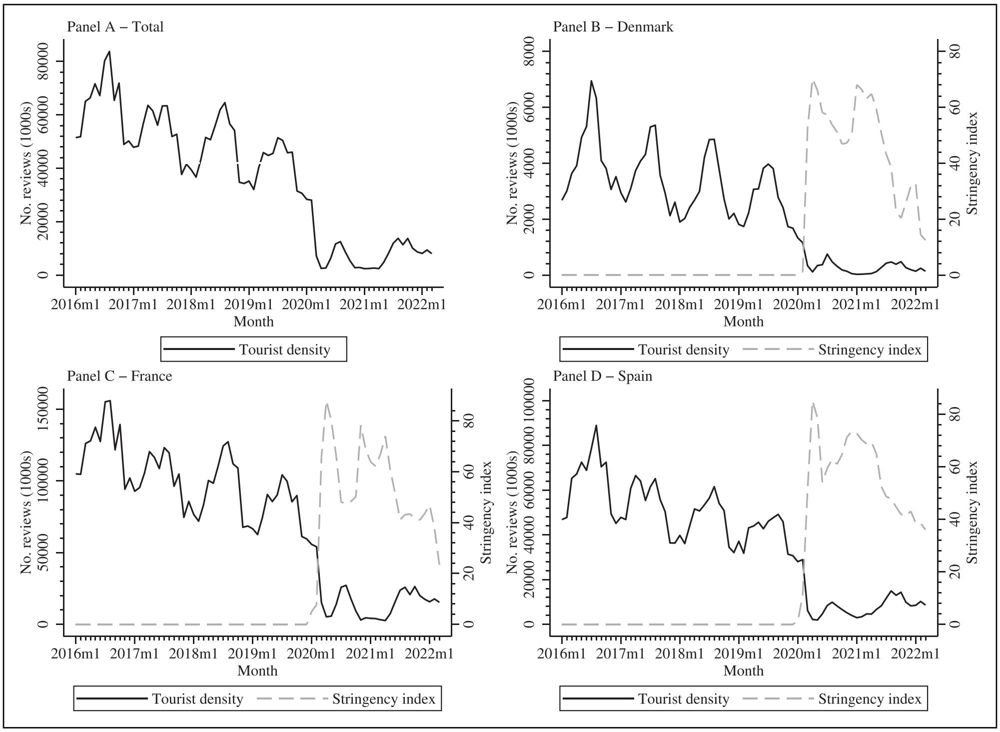

Figure 7. Tourist density of visited locations.
Notes: This Figure shows the evolution of the tourist density of visited locations together with the stringency index. Panel A shows the entire sample, Panel B shows the review density for Danish attractions, Panel C for French attractions and Panel D for Spanish attractions. The review density of an attraction is computed as the total number of reviews of all attractions within a radius of 10 km 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 (Hale et al., 2021).

& 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. As a robustness check, we also estimate the effect of the stringency index using the inter-annual first differences of the variables (e.g., January 2020 level - January 2019 level). By taking first differences, we can control for country specific seasonality patterns. Furthermore, the year-to-year changes would be small before the pandemic and the stringency index therefore captures shifts with respect to the pre-pandemic levels.

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:

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

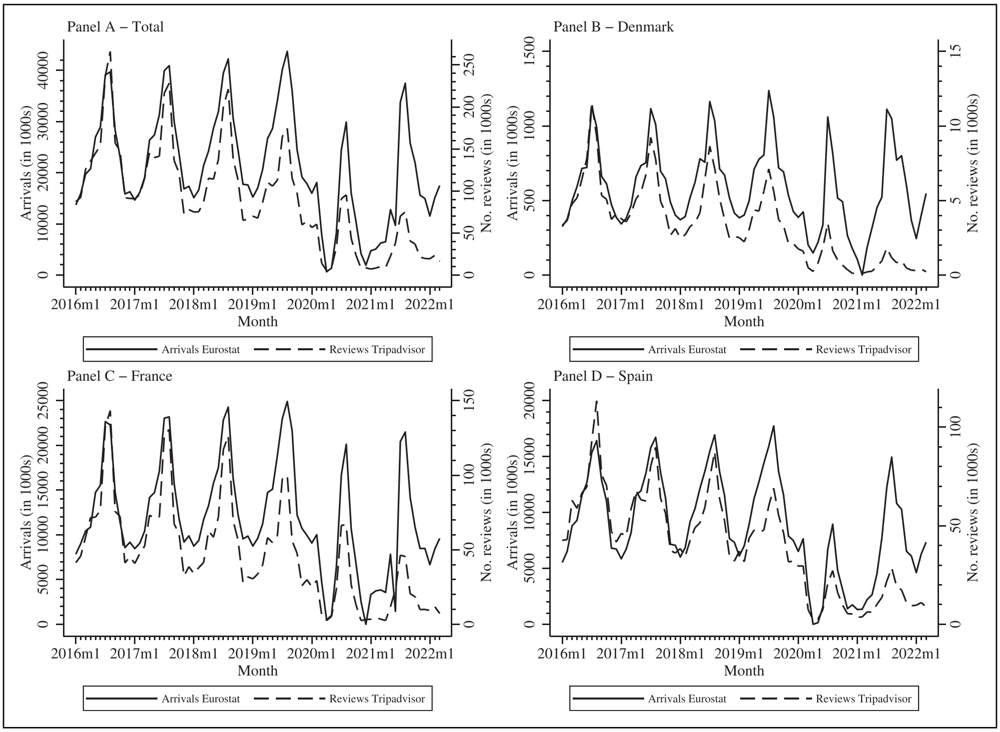

Figure 8. Validity test: tourist arrivals and number of reviews over time.
Notes: This Figure shows the number of tourism arrivals taken from Eurostat together with the total number of Tripadvisor reviews. Panel A shows the total number of arrivals and reviews for our sample, while Panels B-D show the numbers by country. *Source:* Official tourism statistics from Eurostat (2023b) and own data collected from Tripadvisor (see Section 3 for details).

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.

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

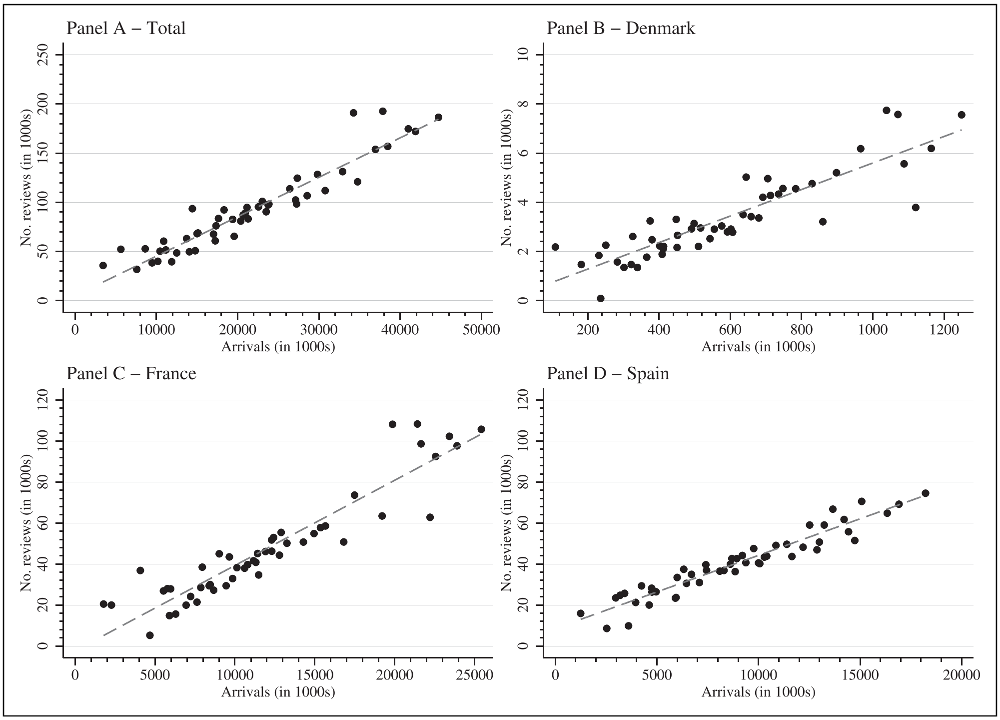

Figure 9. Validity test: monthly correlation between tourist arrivals and number of reviews.
Notes: This Figure shows binned scatter plots of the number of tourism arrivals taken from Eurostat and the number of Tripadvisor reviews. Panel A uses the entire sample, while Panels B-D by country. The correlation 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 (2023b) and own data collected from Tripadvisor (see Section 3 for details).

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 1 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. In Table A.VIII we consider the same by country. When looking at the results for each country the effect of the stringency index is similar to the overall effect, even though it seems a bit larger for France and Spain where a 10 percentage points increase in the stringency index implies a 37% and 32% decrease in the number of reviews respectively. The effect on the share of foreign tourists differs considerably between the three countries spanning from a lower 1.24 percentage points decrease in France to a 5.2 percentage points decrease in Denmark, given a 10 percentage points increase in the stringency index.

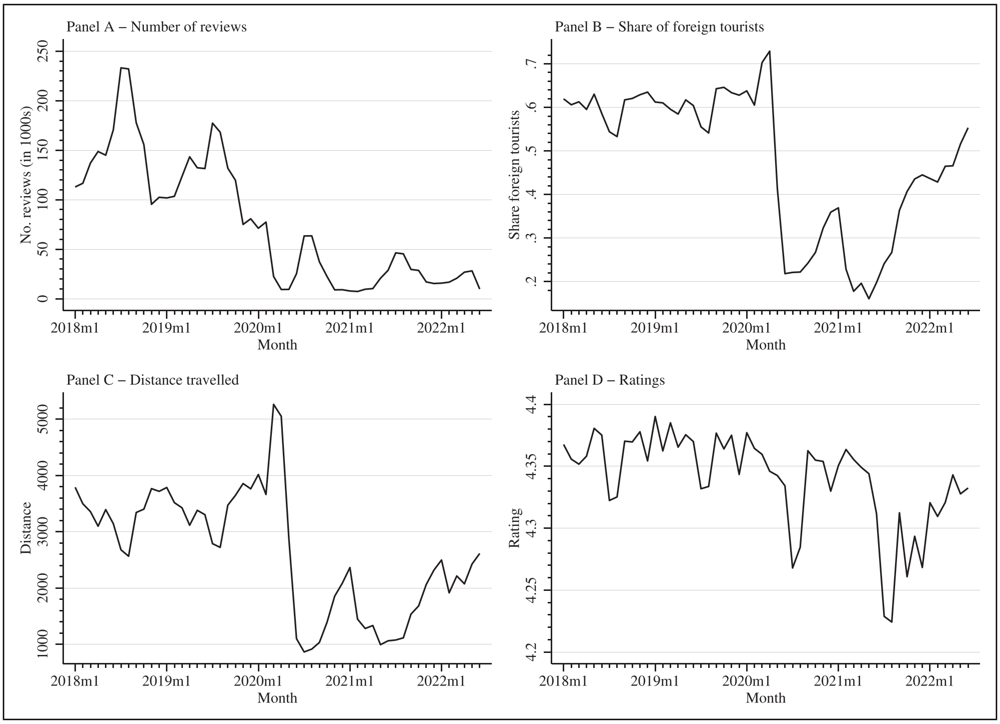

Figure 10. External validity: Global tourism activity using Tripadvisor data.
Notes: This Figure shows different measures of tourism activity using the global sample of attractions. Panel A shows the evolution in the number of reviews, panel B the share of foreign tourists, panel C the travel distance and Panel D the ratings. *Source:* Own data collected from Tripadvisor (see Section 3 for details).

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.[13]

The above results clearly shows how tourism trends have significantly changed due to policy interventions during COVID-19 and that there are some differences also between the three selected countries, especially in terms of the changes in the share of foreign tourists. 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 2 shows the estimates of equation (3) for the entire sample, while the Appendix Table A.XI shows the results by country. In Table 2 the effect of the stringency index is highly significant and similar in magnitudes to Table 1. The estimate refers to Nature & Parks, indicating a decrease in both the number of reviews (Columns 1, 3, 5) and the share of foreign

Table 1. Effect of stringency index 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
Stringency index	−0.029 (0.003)***	−0.323 (0.059)***	−0.022 (0.004)***	−0.350 (0.071)***	−0.007 (0.000)***	−0.228 (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	26,652	26,068	13,84,706	11,55,251
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 (Hale et al., 2021).

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. When looking at the results by country in Appendix Table A.XI, the magnitudes once again differ and the effect on the three heritage types Museums, Sights & Landmarks and Others are also significantly different from the baseline category in Columns 1-4.[14]

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, with differences also between the three countries, where, for example, Danish attractions in the category Sights & Landmarks have experienced a much larger drop in the share of foreign tourists than their counterparts in France and Spain.

Table 2. Effect of stringency index on tourism of different attraction categories.

	Monthly		Daily		Attraction level	
	(1)	(2)	(3)	(4)	(5)	(6)
	ln (reviews)	Tourists	ln (reviews)	Tourists	ln (reviews)	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	26,652	26,068	13,84,706	11,55,251
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 (Hale et al., 2021).

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 traveled 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 3 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 using the entire sample. 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

Table 3. Effect of stringency index on travel distance, attraction density, tourist density and ratings.

	(1)	(2)	(3)	(4)
	Distance	Attraction density 10 km	ln(Tourist density) 10 km	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 refer 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 10 *km*, 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 (Hale et al., 2021).

stringency index implies a decrease of 75 *km* in the travel distance. This implies that when the stringency index increases from 0 to 49 (the average) the travel distance decreases by 372 *km*. 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 10 *km* 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 10 *km* from each attraction. When looking at the countries separately in Appendix Table A.XIII, there are some differences. The change in the travel distance in Column 1 differs between the countries, where in France there is no significant change, while in Denmark and Spain, the decrease is larger than the overall effect, with a decrease of about 841 *km* in Denmark given an increase from 0 to 49 in the stringency index. The effect on the attraction density is not significant in France and Spain, while in Denmark it is similar to the overall. Finally the effects on tourist density and ratings are similar to the overall results in Panel A.

These results indicate that there is a change towards nearer and less crowded locations. Together with the results from the previous section Table 2, we can conclude that apart from moving towards the nature, visitors also seek more isolated attractions after the pandemic.[15] 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.[16] 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.

As a robustness check to the above analysis Table 4 shows the results using the inter-annual changes of the variables for the full sample while Appendix Table A.XVII shows the results by country. Once again the estimates are significant, showing a decrease in all variables except from ratings which is insignificant. As for Tables 1 and 3 the most relevant differences between the three countries can be found for the share of foreign tourists and the travel distance, with Denmark and Spain showing larger effects than France.

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,

Table 4. Effect of stringency index on tourism using year-to-year changes.

	(1)	(2)	(3)	(4)	(5)	(6)
	ln (reviews)	Share foreign tourists	Distance	Attraction density 10 km	ln (tourist density) 10 km	Rating
Stringency index	−0.033*** (0.002)	−0.654*** (0.080)	−21.860*** (2.983)	−7.293*** (1.784)	−0.044*** (0.003)	0.001 (0.001)
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
N	888	874	888	888	888	888
R^2	0.497	0.298	0.381	0.291	0.612	0.081

Notes: Regression results when regressing the inter-annual first differences of the number of Tripadvisor reviews (Column 1), the share of foreign tourists (Column 2), the average travel distance (Column 3), the attraction density (Column 4), the tourist density (Column 5) or the rating (Column 6) on the stringency index. 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 (Hale et al., 2021).

Table 5. Effect of stringency index on different outcomes in global sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Visits	Visits	Visits abroad	Visits abroad	Distance	Distance	Rating	Rating
Stringency index	−0.005*** (0.000)	−0.002*** (0.000)	−0.006*** (0.000)	−0.002*** (0.000)	−21.114*** (0.143)	−13.520*** (0.283)	0.000*** (0.000)	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	18,13,458	18,13,410	18,13,458	18,13,410	15,07,250	15,07,202	18,04,911	18,04,866
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 use number of visits as the outcome of interest, columns 3-4 the number of visits abroad, columns 5-6 the distance traveled 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 (Hale et al., 2021).

and rating. Both measures shows how the volume of traveling 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 5. 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 respectively, 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 130 *km* 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 restrictions indicator and the dummy variables.[17]

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.

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. These insights could prove valuable not only to others utilizing data from Tripadvisor but also to those working with other travel portals or platforms such as OpenStreetMap or Google Maps, which provide location data on points of interest. 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 re-located 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.

Building on these findings, it is crucial to also recognize the limitations and potential biases of data sourced from platforms like Tripadvisor, especially in representing diverse tourist demographics and behaviors. Our analysis, while showing a strong correlation with official data, also acknowledges concerns that depending solely on travel portals like Tripadvisor may introduce certain limitations. The Tripadvisor platform mainly captures insights from users who are active online and inclined to share their travel experiences, potentially misrepresenting certain tourist segments. For instance, senior tourists in Spain may be underrepresented on digital platforms, but not necessary in older tourists in Denmark, who tend to be digitally literate. In any case, it is unlikely that the usage of online platforms by older cohorts was affected by COVID-19 differently than younger ones. Therefore, our findings regarding the *changes* in tourism in times of COVID-19 should remain informative, despite any potential biases in measuring the overall level of tourism.

Another potential concern is that among those who do post about their travels, only a subset uses Tripadvisor. Others prefer platforms like Instagram, Twitter, and Facebook, albeit it is not clear why a user would review attractions differently depending on the platform they use. An eventual selection bias could distort metrics such as the number of reviews or the proportion of foreign tourists, influenced by the popularity of Tripadvisor in their respective home countries.

Despite those limitations, this project comes 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 complementary 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.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Mobile Lives Forum.

ORCID iD

Karol Jan Borowiecki https://orcid.org/0000-0003-4959-181X

Supplemental Material

Supplemental material for this article is available online.

Notes

1. We also compute the attraction density with an alternative radius of 5 *km* and 25 *km* as a robustness check.

2. As for the attraction density, we also compute the tourist density with an alternative radius of 5 *km* and 25 *km* as a robustness check.
3. Figure A.I shows a map of the location of our global sample of attractions.
4. 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 severe measures implemented by the government.
5. The following nine indicators are included: 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.
6. 14 in the Appendix shows the number of reviews over time for all the attraction categories included on Tripadvisor.
7. Figure A.III 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.
8. Figures A.V and A.VI use 25 *km* or 5 *km* radius.
9. Figures A.VII and A.VIII use a 25 *km* or 5 *km* radius.
10. Figure A.IX shows the same when using occupancy rates.
11. Figure A.X shows the same using the occupancy rate.
12. We also compute the simple correlation coefficients which can be seen in Table A.VI in the Appendix, showing a high correlation.
13. Table A.IX and A.X in the Appendix are similar but use travel restrictions and travel restriction dummies as the explanatory variable respectively.
14. In Table A.XII 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.
15. As a robustness check, in Table A.XIV 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.
16. In the Appendix Tables A.XV and A.XVI we show the results when using the travel restrictions indicator and the dummies.
17. The results of these two can be seen in the Appendix in Tables A.XVIII and A.XIX.

References

Adie BA, 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 BA (2020) One visitor too many: assessing the degree of overtourism in established European urban destinations. *International Journal of Tourism Cities* 6(1): 117–137.

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

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

Bertacchini EE, 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 KJ, Forbes N and Fresa A (2016) *Cultural Heritage in a Changing World*. Heidelberg, Germany: Springer.

Borowiecki KJ, Pedersen MU, Mitchell SB and Khan SA (2025) Navigating landscapes: approaches to data collection and analysis in tourism. In: Borowiecki KJ, Civantos JM and Fresa A (eds) *Innovative Cultural Tourism in European Peripheries: Visiting the Margins*. London: Routledge. chapter 5.

Borowiecki KJ, Pedersen MU and Palomeque M (2024) *Putting the Periphery on the Map: Tourism Activity Measured with Big Data*. Odense, Denmark: University of Southern Denmark. Working paper, work in progress.

Boto-García D and Leoni V (2023) Distance traveled in times of pandemic: an endogenous switching regression approach. *Tourism Economics* 29(3): 571–595.

Castillo-Manzano JI, Castro-Nuño M, Lopez-Valpuesta L, et al. (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 RB and Capone F (2016) Defining tourist clusters in europe: a micro-level data approach. In: Capone F (ed) *Tourist Clusters, Destinations and Competitiveness*. England, UK: Routledge, 95–114.

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. https://data.europa.eu/88u/dataset/ctjvyeqsbf9wsepdmmqrvw.Lastdataupdate:05/01/2023

Eurostat (2023b) *Eurostat tourism statistics*. https://ec.europa.eu/eurostat/web/tourism.Lastdataupdate.

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 & Tourism Marketing* 34(3): 369–382.

Falk MT, 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.

Fenichel E, Kuminoff N and Chowell G (2013) Skip the trip: air travelers' behavioral responses to pandemic influenza. *PLoS One* 8(3): e58249.

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

Frey BS 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.

Hale T, Angrist N, Goldszmidt R, et al. (2021) A global panel database of pandemic policies. *Nature Human Behaviour* 5: 529. (oxford covid-19 government response tracker).

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

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

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

Kock F, Nørfelt A, Josiassen A, et al. (2020) Understanding the covid-19 tourist psyche: the evolutionary tourism paradigm. *Annals of Tourism Research* 85: 103053.

Lenzen M, Sun YY, Faturay F, et al. (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, et al. (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. *The American Economic Review* 104(8): 2421–2455.

Milone F, Gunter U and Zekan B (2023) The pricing of European airbnb listings during the pandemic: a difference-in-differences approach employing COVID-19 response strategies as a continuous treatment. *Tourism Management* 97: 104738.

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 XS, Li X, et al. (2020) Reviewing experts' restraint from extremes and its impact on service providers. *Journal of Consumer Research* 47(5): 654–674.

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

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.

Sessions C, Wood SA, Rabotyagov S, et al. (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 PJ (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, et al. (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, et al. (2017) Instagram, flickr, or twitter: assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports* 7(1): 17615.

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

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

Wood SA, Guerry AD, Silver JM, et al. (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* 83: 102913.

Yoo K-H, Sigala M and Gretzel U (2016) Exploring tripadvisor. In: Egger R, Gula I and Walcher D (eds) *Open Tourism: Open Innovation, Crowdsourcing and Co-creation Challenging the Tourism Industry*. Berlin, Heidelberg: Springer Berlin Heidelberg, 239–255.

Zenker S and Kock F (2020) The coronavirus pandemic – a critical discussion of a tourism research agenda. *Tourism Management* 81: 104164.

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 (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. In: Cameron S (ed) *A Research Agenda for Cultural Economics*. Cheltenham, UK: Elgar.

Author biographies

Karol J. Borowiecki, Professor of economics at the University of Southern Denmark, is renowned for his innovative research methodologies and societal impact. He published >40 items, including in the Journal of Political Economy, a textbook with Cambridge University Press, and a co-edited volume on cultural heritage. He sits on the editorial boards of Tourism Economics and the Journal of Cultural Economics. As President- Elect of the Association for Cultural Economics International and a top cultural economist, Karol collaborates with premier European institutions, shaping policy and advancing cultural and tourism economics.

Maja U. Pedersen, PhD degree in economics, currently Assistant Professor of economics at the University of Southern Denmark. Published 6 items on different topics, including economic growth, financial history and globalization. All published in various peer-reviewed journals.

Sara Mitchell specializes in urban, labor, and cultural economics, focusing on creative worker migration, agglomeration effects, and cultural tourism. Her work leverages often historical datasets, shedding light on economic patterns. Holding a PhD from Trinity College Dublin, where she was a Grattan Scholar and Irish Research Council Fellow, Sara's career spans roles at the University of Southern Denmark, TU Dortmund, and the Institute of Public Administration, reflecting her diverse expertise and contributions to economics.