A novel application of big data to measure trends in tourism: France, Spain and Denmark, January 2016 – March 2022

Report prepared by:

Karol J. Borowiecki
(University of Southern Denmark)

Maja U. Pedersen
(University of Southern Denmark)

March 11, 2024

This project is funded and scientifically supervised by the Mobile Lives Forum, as part of the mobility transition research program. The Mobile Lives Forum is a research institute created by SNCF.

Contents

Executive summary

Executive Summary

This report presents an analysis of cultural and nature tourism in Europe, utilizing an innovative, big data approach based on user-generated content from a leading travel portal. By moving beyond traditional data sources, the study offers deeper insights into the dynamics of the tourism sector, particularly in the context of the Covid-19 pandemic.

Research Scope and Methodology

The research leverages a vast dataset from a leading travel portal, encompassing ca. 7 million reviews, to understand changes in tourist behavior and preferences. Covering the period from January 2016 to March 2022, the study focuses on three key European countries: Denmark, France, and Spain. The research uses computer-science techniques and econometric methods, such as regression analysis, to dissect and interpret large-scale tourism data. The approach is validated in a comparison of the novel data used here with existing tourism statistics. The data collected will be available to other researchers upon request to the Mobile Lives Forum. Furthermore, the study also uses the Oxford Government Response Tracker to assess how trends in tourism were affected by national policies to limit mobility during the pandemic.

Key Findings

The study confirms shifts in European cultural and nature tourism noted during the Covid-19 pandemic. It documents and measures a pronounced preference for outdoor and less crowded destinations, signifying a substantial change in tourist behavior. During the pandemic when they were able to travel, tourists increasingly favored nature-oriented experiences and destinations off the beaten path, diverging from pre-pandemic trends that favored popular and often congested tourist spots. If these trends are sustained beyond the period studied, they may be indicative of broader changes in tourist behavior. In this report we show how big data can help us understand the magnitude and evolution of these changes.

The research also sheds light on the popularity of various tourist attractions.

This research is particularly adept at monitoring broad trends in tourism activity and provides insights at international, national, and regional levels, making the pursued approach a valuable tool for understanding large-scale changes in the tourism sector.

Additionally, the study provides a unique perspective by leveraging big data from a leading travel portal. A comparison with official tourism data, such as Eurostat statistics, validates the reliability

of the novel approach pursued here, highlighting its potential as a complementary tool for tourism studies.

Limitations of the Approach

While innovative, our study acknowledges certain limitations inherent to using big data extracted from user-generated content. These include potential gaps in data where users (tourists) do not leave reviews for every trip, and the absence of detailed user demographics. Additionally, the study employs the Oxford Covid-19 Government Response Tracker's stringency index and categorizes attractions as indoor or outdoor, which, while effective for broad analysis, may not capture every nuance of some sites. However, these limitations do not significantly detract from the overall validity and value of the findings.

Conclusion

This research provides a nuanced understanding of the European cultural and nature tourism landscape, particularly during the Covid-19 pandemic. The use of big data from a leading travel platform offers a new perspective in tourism studies, albeit with certain limitations. These insights are valuable for comprehending the current state and potential future trajectories of the tourism industry in Europe.

1 Introduction

International tourism statistics, such as those provided by Eurostat, are appreciated and used by scholars and practitioners alike. However, they also come with a number of notable shortcomings: they are over-aggregated (usually at the country level), available with a lag of many months or more, often only at the annual level, and the information about the tourist may be partial.

In this report, we suggest a complementary, computer-science-driven approach to measure tourism that relies on big data collected from a leading travel portal. This novel approach enables us to obtain a systematic, consistent, and reliable approximation of tourism flows in different countries, offering unprecedented precision, frequency, and depth of information. Compared to mainstream tourism statistics, our approach delivers 1) information on tourism flows at the attraction-level (not just at the 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 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 (2023). 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 at different levels of aggregation.

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 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. 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. Finally, the results suggest that the imposition, by governments, of restrictive measures has no impact on visitors' satisfaction of a destination.

The scope of this paper is motivated by the vast and strategic role of tourism, particularly of cultural and nature tourism, in Europe.[1] 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

[1]For an early empirical study of cultural tourism, refer to Borowiecki and Castiglione (2014), who investigate the association between participation in cultural activities and tourism flows in Italian provinces.

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). 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 also provides unique indicators on the satisfaction of a visit, including the rating given or various indicators derived from over 3 million 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 perceptions 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 have 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 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 society in a post-Covid-19 Europe.

The rest of the paper is organized as follows: Section 1 provides an introduction. Section 2 delves into data and validation, explaining the methodology and data sources used. Section 3 discusses findings from the analysis using data aggregated at the country level. Section 4 offers an in-depth look at findings from analysis of regions, localities, and tourists. Section 5 addresses the limitations of the approach, acknowledging the constraints and challenges encountered in the research. The paper concludes with Section 6.

2 Data and validation

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.

2.1 Measuring tourism using big data

International tourism statistics have several significant shortcomings such as over-aggregation and are lacking in important information about the tourist. National statistics in some countries provide additional information, however, each country's statistics are 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 by 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. Part of the analysis presented in this report is based on Borowiecki et al. (2023).[2] Furthermore, a similar approach to collect data on tourism is also used in Borowiecki et al. (2024a,b) with a different set of countries, less detailed data and with a focus on peripheral locations.

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, 2023). 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.[3]

[2]The data collection process and descriptive statistics of the raw data are also available in Borowiecki and Mitchell (2024).

[3]It should here be noted that the classification system is not mutually exclusive and hence some attractions can be

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.

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 locations 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 "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 a given location is for tourists. For each attraction we count the number of other attractions located within a radius of $10km$ as a measure of density.[4] 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.[5]

In the appendix we show descriptive statistics by attractions and users for the entire sample in Table A1 and by country in Tables A2-A4. Overall, for our analysis we have 6,847,931 reviews of which

classified in multiple of the first three categories at the same time. 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.

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

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

about 3.5 million for France, 3 million for Spain and 0.2 million from Denmark.[6] In Table A5 we show summary statistics for the global sample. In the tables with summary statistics, we also include additional variables on tourism from Eurostat (Eurostat, 2023) which we use in subsection 2.2. Finally, we also use the Oxford Covid-19 Government Response Tracker's stringency index (Hale et al., 2021) in the analysis of the impact of restrictions imposed during Covid-19 on tourism.

2.2 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, 2023) regarding tourist arrivals aggregated monthly at the country level.

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 1 shows the evolution of all Eurostat arrivals and all Tripadvisor reviews over time. Panel A uses all data, while panels B-D show the patterns individually for each country: Denmark, France and Spain. It becomes fairly clear that the time-series follow each other very closely in magnitude and seasonality.

As a second visual inspection, Figure 2 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 which is an indication that they correlate well.

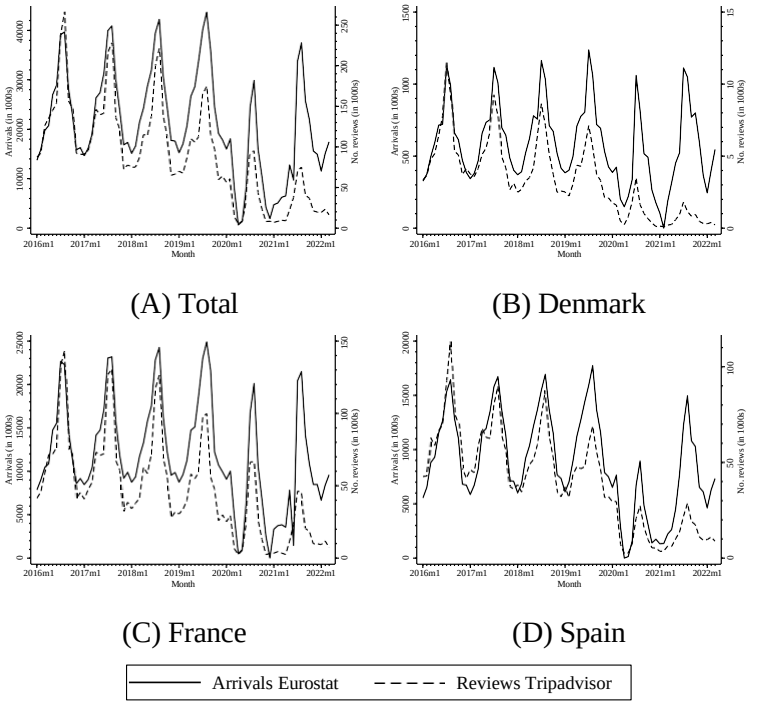
As a more formal test we also compute the correlation coefficients between the number of reviews from Tripadvisor and the number of arrivals from Eurostat. The closer the correlation coefficient is to one, the closer the variables correlate. The results can be seen in Table A6 in the Appendix, both when using the entire sample and when concentrating on each country separately. In all cases, the correlation coefficients are very close to one and also significantly different from zero.

Finally it is possible to validate the data by estimating how well tourism arrivals can explain the number of monthly reviews from Tripadvisor in a simple regression design. The results can be seen in Table A7 in the Appendix, using $ln(Arrivals)$ as the explanatory variable. We show the results for all countries together in column 1 and then individually for each of our three countries, Denmark, France and Spain, in columns 2-4. In all models we include country fixed effects, year fixed effects and month fixed effects. All models have a high explanatory power considering the high R^2 values 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.

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

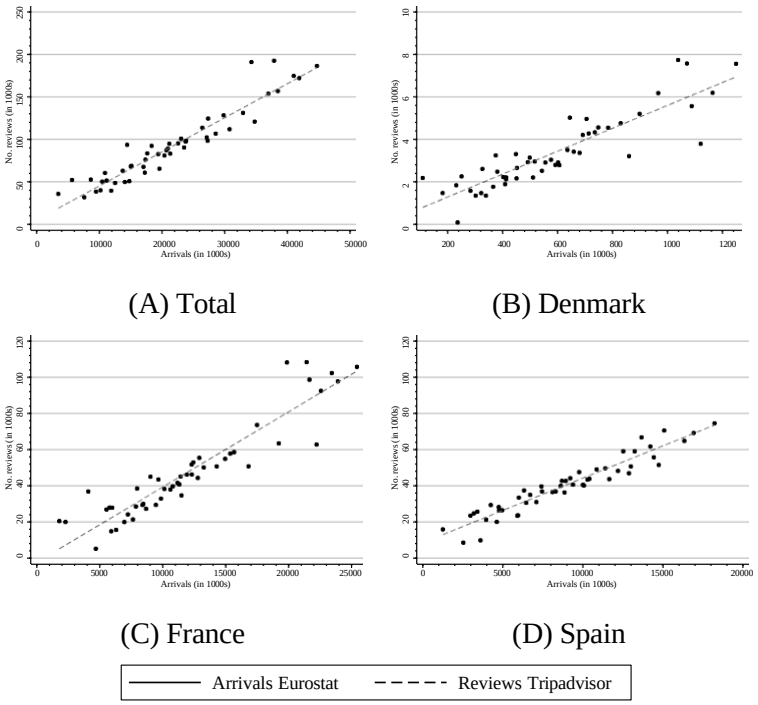
[6]Due to missing information about the location of some attractions and users in the raw reviews data file, the final number of user reviews included in our analysis is a bit lower than in the raw data.

Figure 1: 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, 2023) and own data collected from Tripadvisor (see Section 2 for details).

Figure 2: 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, 2023) and own data collected from Tripadvisor (see Section 2 for details).

3 Findings of analysis using data aggregated at the country level

In this section we present results of the impact of Covid-19 restrictions on tourism flows using data aggregated at the country and attraction category level. In this part of the analysis the attractions are classified into one of the four different categories: 1) Museums, 2) Sights & Landmarks, 3) Nature & Parks, and 4) Others and the data aggregated accordingly. The unit of observation is hence the attraction category by country and month. We start by describing our measures of restrictions and then we present figures illustrating travel flows. Finally, we show results from a formal analysis estimating the impact of restrictions on tourism flows.

3.1 Stringency index and travel controls

As explained above, in addition to our data from Tripadvisor, we make use of 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 continue to be 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 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.[7] Whenever one of the nine included indicators change, the stringency index also changes 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

[7]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*.

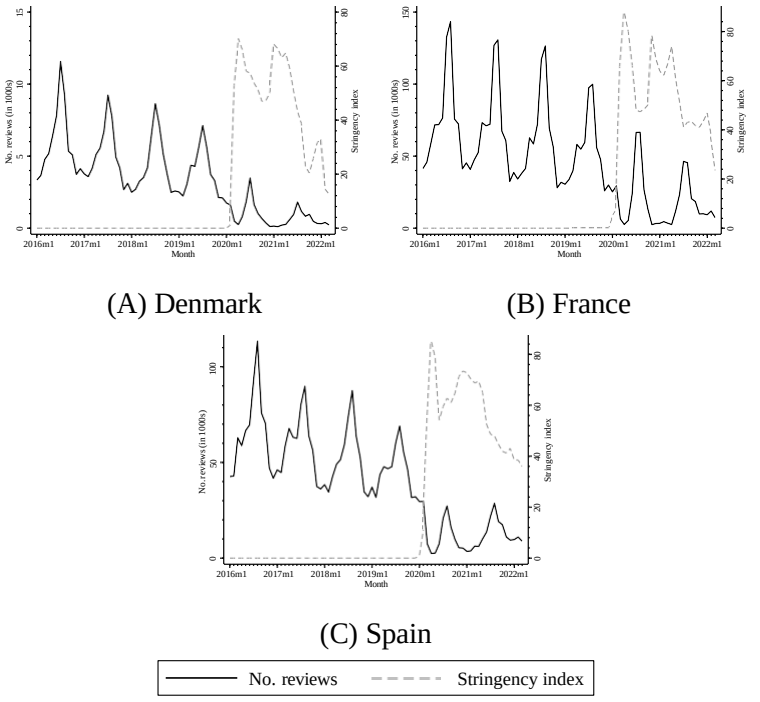
$$SI = \frac{1}{9} \sum_{j=1}^{9} I_j \tag{1}$$

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. In the analysis we consider the international travel restrictions in a version where we create a dummy indicator for each level of restrictions.

3.2 Change in tourism flows over time

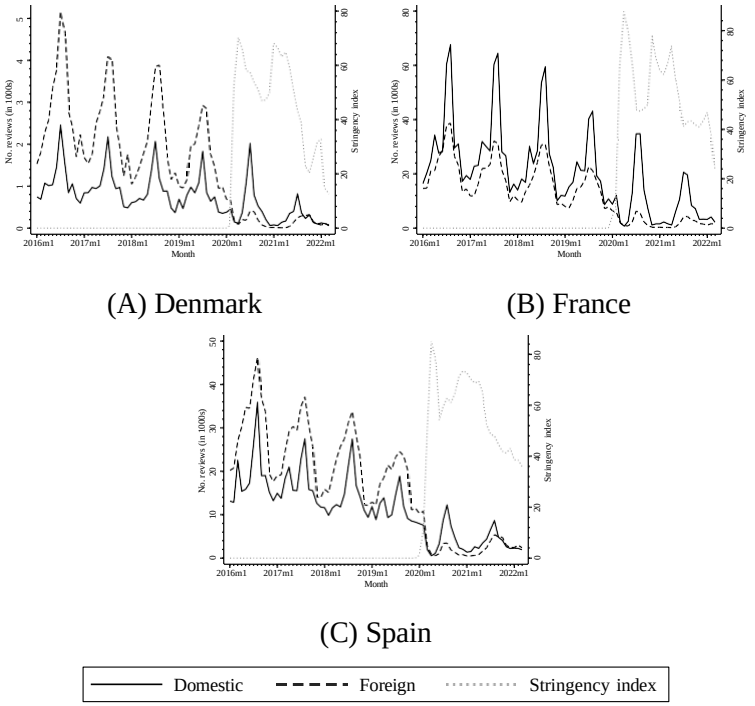
In this section we look at two measures of tourism flows to see they are impacted by the introduction of restrictions. The two measures are the number of reviews and the share of foreign tourists in each of the three countries. In Figure 3 we show the evolution of the total number of reviews over time together with the Stringency Index measure. From Figure 3 it is clear that there is a sharp decrease in the number of reviews beginning in February 2020 when the first restrictions are imposed, where after the reviews follow a clear inverse relationship with the Stringency Index. Apart from looking at the total number of reviews, we also looked specifically at the number of foreign and domestic tourists visiting the attractions in our sample. The evolution over time of the number of foreign and domestic tourists can be seen in Figure 4 and once again there is a clear break around February 2020.

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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Figure 4: Number of domestic and foreign tourists by country over time



(A) Denmark (B) France

(C) Spain

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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Finally, it is also possible to look at the number of reviews for different categories of attractions. In Figure 5 we show the total number of reviews in Denmark, France and Spain over time in each of the attraction categories identified by Tripadvisor. For all categories there is a clear pattern of seasonality, with more reviews during the high season in the summer. Similar to the overall number of reviews in Figure 3, there is a sharp decrease in the number of reviews around February 2020, which appears more pronounced for some categories than others. Especially the ”Events” and ”Fun and Games” categories experience large decreases. From Figure 5 it is also possible to see how big a share each category constitutes out of the total number of reviews by comparing the number of reviews in each category. Especially the categories in Panels A and B represent the largest groups of attractions.

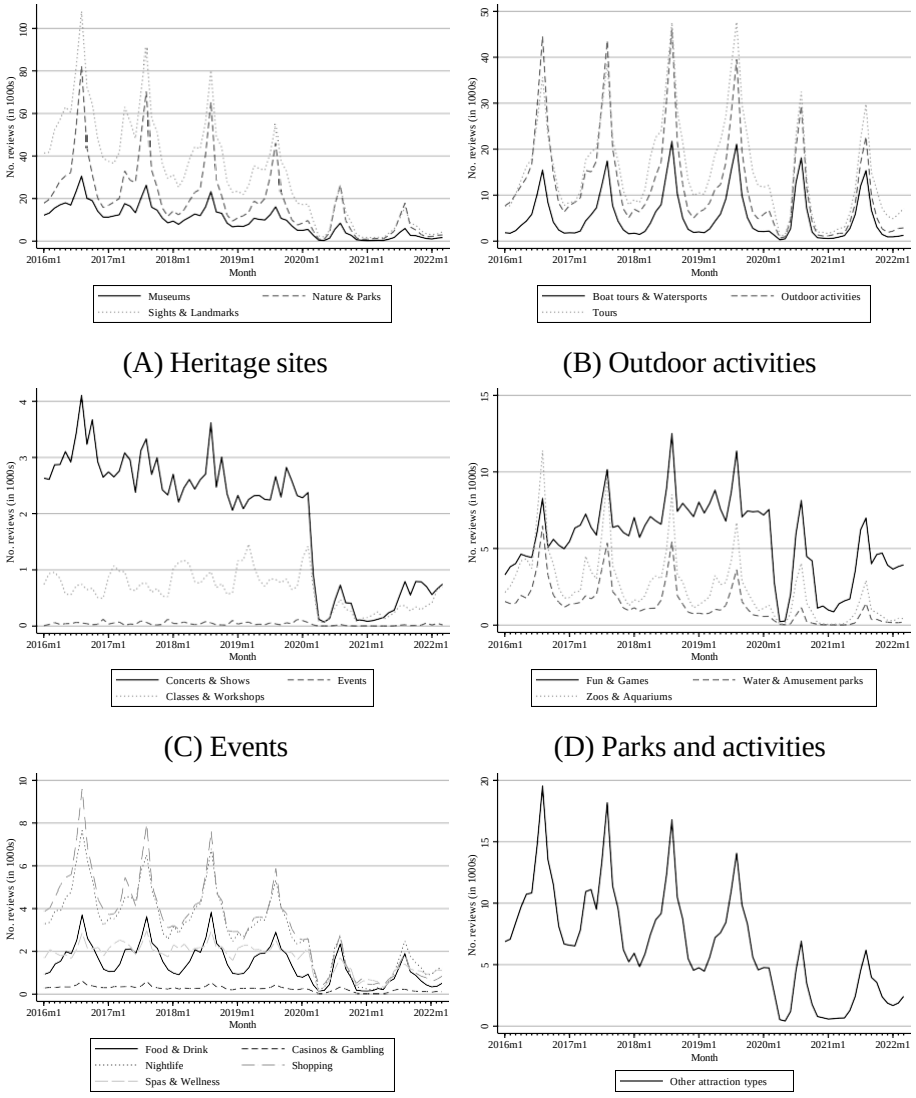
3.3 Impact of Covid-19 on tourism flows

To measure more formally the impact of the restrictions on the two measures of interest, we perform a regression analysis.[8]

[8]We estimate the impact using the equation:

$$y_{ct} = \beta_1 SI_{ct} + \beta_0 + \Gamma + \varepsilon_{ct} \quad (2)$$

Figure 5: Number of reviews over time by different attraction categories



(A) Heritage sites (B) Outdoor activities

(C) Events (D) Parks and activities

(E) Various (F) Other attraction types

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 categorized 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 2 for details).

Given that we have observations from different countries which are potentially different in many ways, we include country fixed effects in the analysis. The country fixed effects control for characteristics specific to each country but constant over time. To control for characteristics that change over time but are the same across countries, we also include two kinds of time fixed effects. The

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.

first are monthly fixed effects 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 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 regression design we use the two different measures of restrictions, i.e., the stringency index and the travel restriction dummies to obtain a measure of the impact of the restrictions. When using the stringency index, the estimated value tells how a 1% change in the stringency index affects the outcome variable, i.e., number of reviews or the share of foreign tourists. When using the travel restriction dummies we obtain four different estimated impacts, one for each level of travel restrictions. These estimates indicate the difference between no restrictions and one of the four different levels of travel restrictions. In this section we concentrate on the results using data aggregated monthly at the country and attraction category level, however, in the appendix we also show the results when using data aggregated daily and when using the single attraction as the unit of observation instead of aggregating at the country level.

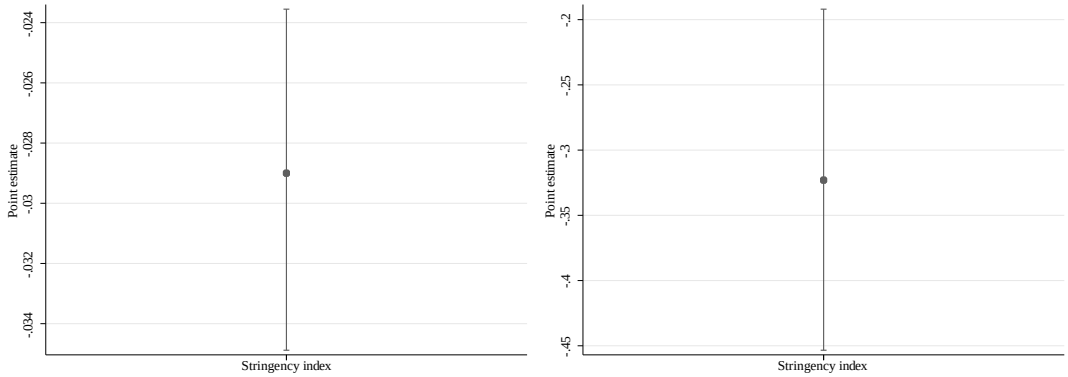
In Figure 6 we show the point estimates: the estimated value of the impact using the stringency index as the explanatory variable and the natural logarithm of the number of reviews (Panel A) or the share of foreign tourists as the outcome of interest (Panel B). The point estimates are shown together with the 95% confidence intervals. The confidence intervals show how precise the estimated values are, where smaller intervals indicate more precise estimates. The estimates are significantly different from zero whenever the confidence intervals do not cross the horizontal line around zero. The illustrated point estimates are based on the regression results in Table A8, columns 1-2 in the appendix, where it is also possible to see the estimates when considering the data aggregated daily and at the attraction level.

In both cases there is a clear negative impact of the stringency index on the two measures, given the negative point estimates. To have an idea about the magnitudes of the effects, Figure 7 shows the effects on the two outcomes of interest, considering different percentage changes in the stringency index measured along the x-axis. For example, in Panel A, a 10 percentage points increase in the stringency index implies a 29% decrease in the number of monthly reviews. In Panel B we find that a 10 percentage points increase in the stringency index implies a 3.2 percentage points decrease in the share of foreign tourists.

Next, we look at the effect of the travel restriction dummies on the two outcomes of interest. The point estimates can be seen in Figure 8 and are based on the results in Table A9, columns 1-2 in the appendix.[9] The point estimates refer to the effect when going from no restrictions to one of the four levels of restrictions. From Panel A it is clear that the effect on the number of reviews becomes larger with the introduction of more severe travel restrictions, and the decrease is only significant with the introduction of bans at arrival. Panel B shows similar results, indicating that the share of foreign tourists only significantly decreases when a partial ban or more is introduced, in which case

[9]In the same table we also show the results when aggregating daily and at the attraction level.

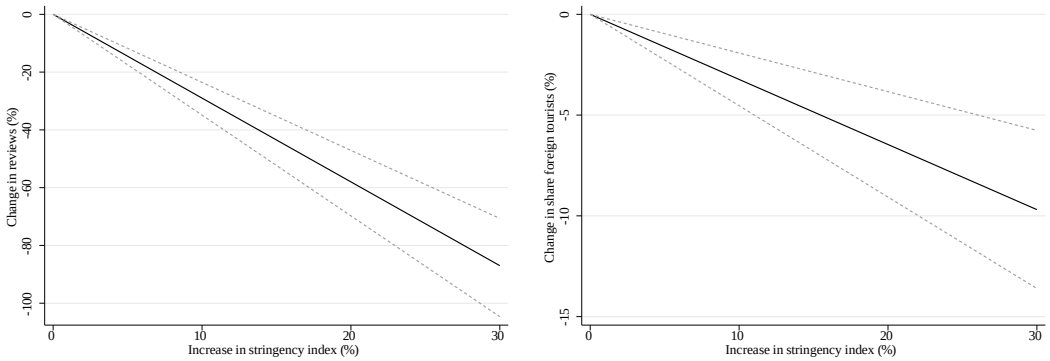
Figure 6: Point estimates - effect of stringency index on reviews and share of foreign tourists



(A) Reviews (B) Share foreign tourists

Notes:. This figure shows the point estimates from a regression using the stringency index as the explanatory variable and data aggregated monthly. In panel A the number of reviews is the outcome of interest and in Panel B the share of foreign tourists is the outcome of interest. The point estimates are shown together with the 95% confidence intervals. Estimates are based on results in Table A8, columns 1-2 in the appendix. *Source:* Own data collected from Tripadvisor (see Section 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Figure 7: Effect of a change in the stringency index on reviews and share of foreign tourists



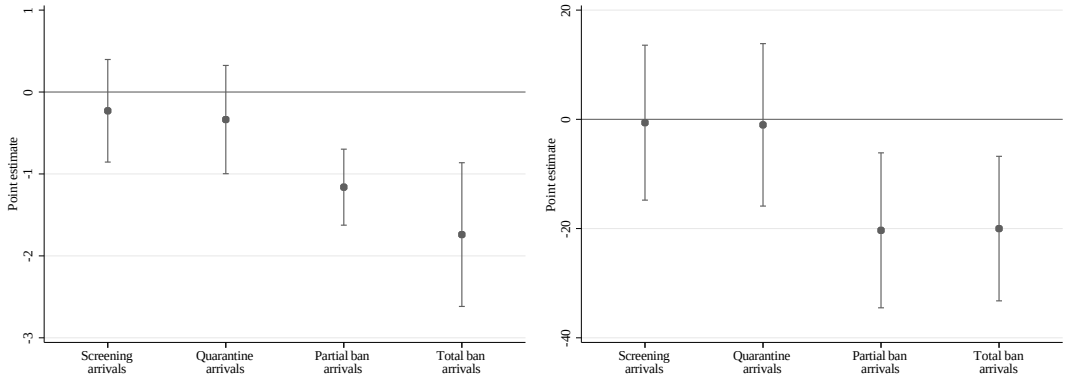
(A) Reviews (B) Share foreign tourists

Notes:. This figures shows the effect of a percentage change in the stringency index on the number of reviews or the share of foreign tourists, considering data aggregated monthly at the country and attraction category level. The dashed lines indicate the 95% confidence intervals. *Source:* Own calculations based on the results from Table A8 in the appendix.

the share of foreign tourists decreases by about 20% with respect to no restrictions.

Finally, we also conduct a regression analysis where we can compare the impact of restrictions on different attraction categories, concentrating on attractions related to cultural and natural heritage, i.e., Museums, Nature & Parks and Sights & Landmarks. We use the stringency index as the explanatory variable to estimate the impact on each of the categories. To obtain more variation in the outcomes of interest, we show the results using data at the attraction level and aggregated

Figure 8: Point estimates - effect of travel restriction dummies on reviews and share of foreign tourists



(A) Reviews (B) Share foreign tourists

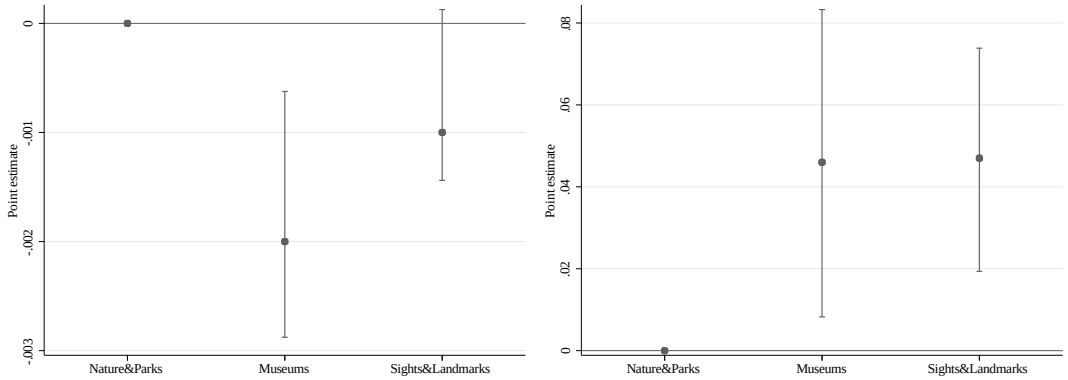
Notes: This figure shows the point estimates from a regression using travel restriction dummies as the explanatory variables. In panel A the number of reviews is the outcome of interest and in Panel B the share of foreign tourists is the outcome of interest. The point estimates are shown together with the 95% confidence intervals. Estimates are based on results in Table A9, columns 1-2 in the appendix. *Source:* Own data collected from Tripadvisor (see Section 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

monthly.[10] In our analysis, we use Nature & Parks as the reference category, and therefore, the obtained estimates measure the differential impact between Nature & Parks and the other categories. The point estimates can be seen in Figure 9 and are based on the results in Table A10, columns 5-6 in the appendix.

From Panel A, it is possible to see that the effect on Museums is negative and significant, indicating that an attraction classified as a museum receives 2% less reviews with respect to an attraction classified as Nature & Parks, given a 10 percentage points increase in the stringency index. 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 Panel B, 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.

[10]We also conducted the analysis using data aggregated at the country level monthly and daily with similar results.

Figure 9: Point estimates - effect of the stringency index on different attraction categories



(A) Reviews (B) Share foreign tourists

Notes:. This figure shows the point estimates from a regression using the stringency index as the explanatory variable. The point estimates show the differential effect between Nature & Parks and the other two attraction categories Museums and Sights & Landmarks. In panel A the number of reviews is the outcome of interest and in Panel B the share of foreign tourists is the outcome of interest. The point estimates are shown together with the 95% confidence intervals. Estimates are based on results in Table A10, columns 5-6 in the appendix. *Source:* Own data collected from Tripadvisor (see Section 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

From the above analysis we can conclude that the introduction of restrictions during Covid-19 had a negative impact on tourism flow measures by both number of Tripadvisor reviews and share of foreign tourists. We also find an indication that the introduction of restrictions moved tourists towards outdoor attractions.

4 Findings of analysis of regions, localities and selected individuals

The use of our novel data from Tripadvisor allows us to look at tourism at disaggregated levels. We start this section by showing new maps of tourism activity for our sample followed by an analysis of individual tourists, that is the users, who post reviews on Tripadvisor. In the second part of the section we turn to other novel measures illustrating changes in tourism. These measures are based on the unique features of our disaggregated data, and include travel distance, density of attractions/tourists and ratings. We also exploit the disaggregated data by showing novel insights on where the reviewers came from (e.g., locals from same region, domestic from any other region, etc.). The final part of this section is dedicated to a more detailed analysis of specific locations in the three countries. For each country we have selected a popular travel destination area such as a capital, and compare these to more peripheral destinations.

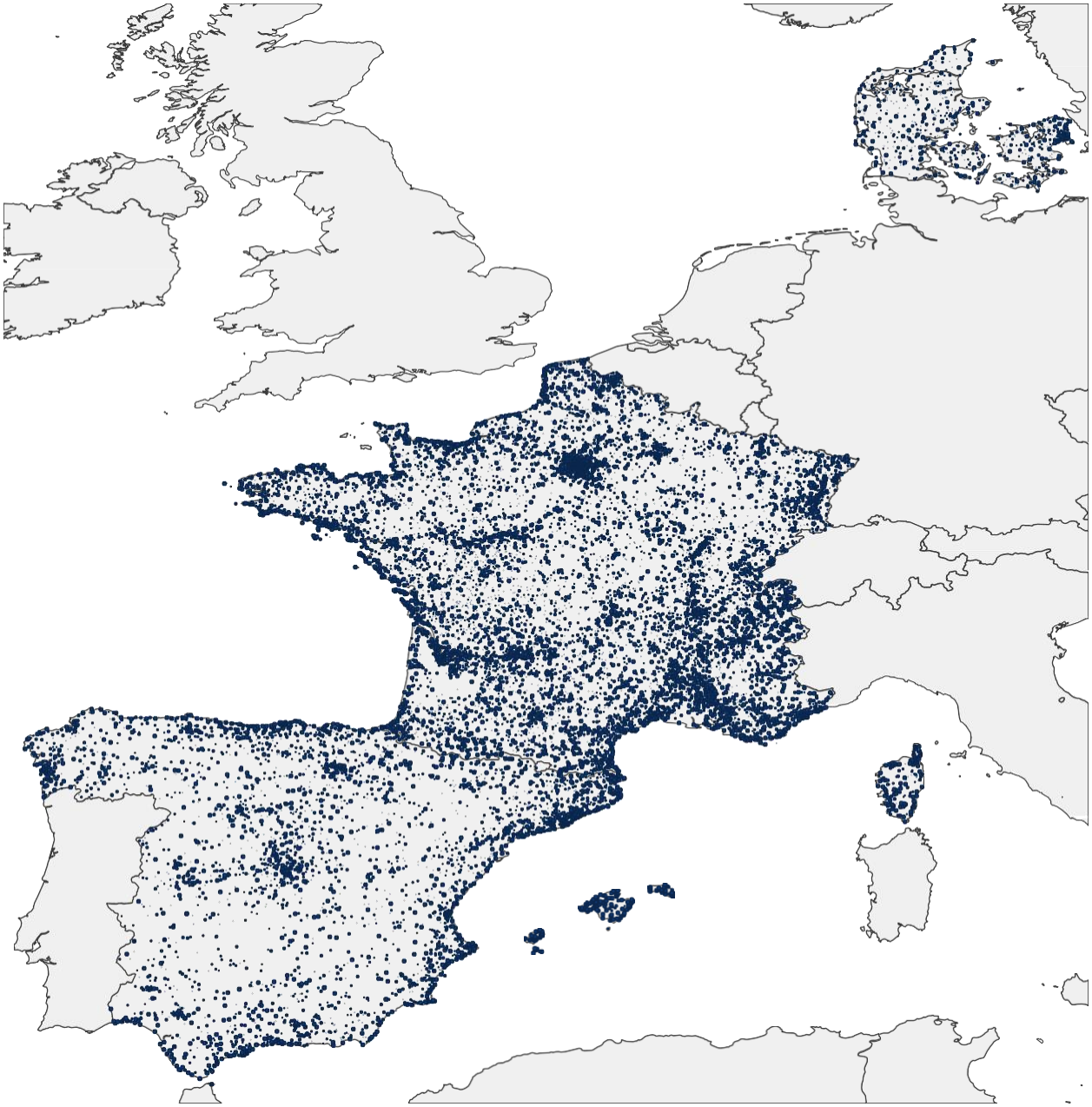
4.1 Mapping attractions and users

Figure 10 shows a map of the location of all the attractions in Denmark, France and Spain. To better illustrate the locations, the size of the blue dots are weighted using the total number of reviews in a location. In this way, the more popular travel destinations have larger dots to show that in these places there is more activity in terms of tourism. From Figure 10 it clearly appears that there is a higher concentration of attractions around bigger cities and coastal locations. It is also fascinating to observe how well the locations reflect the geo-historical landscape of a country. For example, for France just by looking at the attractions, it becomes apparent where lies the Loire Valley or the rivers and Rhone Dordogne.

In Figure 11 we show the location of users who have provided information about their location in our sample. Here the size of the red dots are weighted according to the number of users located in a specific location to better illustrate more frequent locations. There is clearly a large number of users originating from Europe, but also from other and more distant locations.

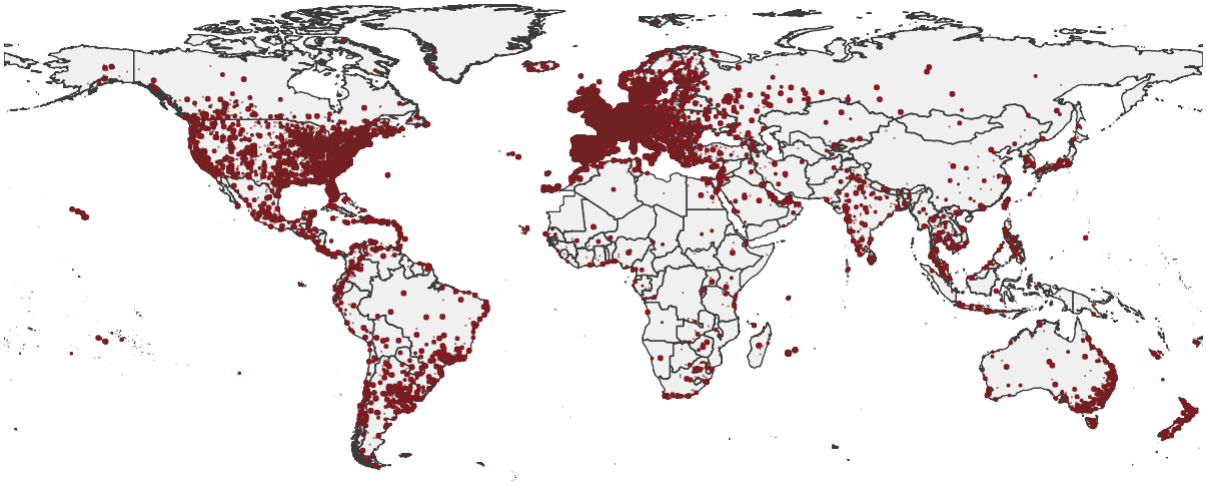
Finally, Figure A1 in the appendix shows a map of the location of our global sample of attractions. Even though the global sample does not include all attractions outside of Denmark, France and Spain, it still appears representative, given the distribution of attractions all over the world, with more frequent destinations having more visitors.

Figure 10: 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 2 for details).

Figure 11: 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 2 for details).

Apart from looking at the locations of attractions and users, we can also look at the travel patterns of the users by combining the user with the attraction visited through lines on a map. To do this, we have divided our sample into three periods. The first period spans from six months before the outbreak of the Covid-19 pandemic, i.e., from July 2019 to December 2019. The second period represents the easing after the first lockdown, which we have identified to be from June 2020 to September 2020.[11] Finally, the third period is representative of the second round of easing of restrictions which we have identified to be from September 2021 to December 2021. The travel patterns from the six month period before the pandemic can be seen in Figure 12, while the two other periods can be seen in Figures A2 - A3 in the appendix. Panel A in each figure shows travel patterns for tourists all over the world, Panel B for tourists within Europe, Panel C domestic tourists and Panel D local tourists.[12] In terms of origin of tourists and their destination, the three maps are not very different from one another. On the other hand, there are some differences in terms of the number of users in the three periods, which is especially apparent considering users from outside Europe. In general, the three countries covered in this research attract tourists from all 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 other's countries. In general, apart from urban destinations, coastal locations attract the most activity. Domestic tourism (Panel C Domestic) is very capital-centered, but

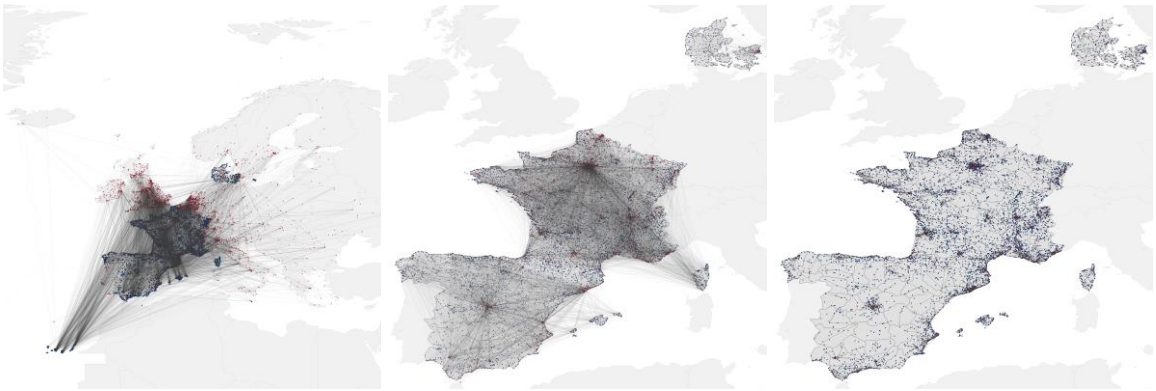
[11]We have identified the periods by looking at the changes in the Oxford stringency index and determined when the index was lower.

[12]The local tourists are identified by comparing the NUTS3 region in which the user is located and the NUTS3 region of the attraction.

also towards other cities as well as coastal locations. Madrid and Barcelona appear to be well connected with their 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. Traveling within regions is often directed to cities, but also towards the coast or mountains (Panel D Local).

Figure 12: Travel patterns - Six months before the outbreak of the Covid-19 pandemic

(A) World



(B) Europe (C) Domestic (D) Local

Notes: This figures shows travel patterns of all reviews written by tourists visiting attractions in the period before the outbreak of the Covid-19 pandemic, September 2019 to December 2019. The blue dots represent the location of attractions and the red dots the location users. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

4.2 Analysis of Tripadvisor users

Before turning to an analysis of the direction of tourism and tourism in specific locations, we dedicate this section to a more detailed description of the users posting reviews on Tripadvisor. This

analysis is based on the information provided by Tripadvisor regarding the origin of the user and when the user joined Tripadvisor. The information is supplemented with patterns observable in the data regarding when a review is posted, how many reviews per user, the time elapsed between reviews and the time elapsed from the month when a user joined and January 2016 when our data starts. Summary statistics can be seen in Table 1. Overall, we have almost 3 million users posting on average 2.28 reviews. Of all users, about 1 million posted more than one review, and the average time elapsed between two posted reviews is about 147 days. Finally, the first users joined Tripadvisor 80 months before January 2016, while on average users registered about 5 months after 2016, indicating that a large share of the users joined before our data starts.

Table 1: Summary statistics of users and their reviews

Variable	Mean	Std.dev.	Observations
Reviews per user	2.28	8.31	2924872
Months from 2016	5.12	40.64	2924872
Days between reviews	147.02	261.9	955407

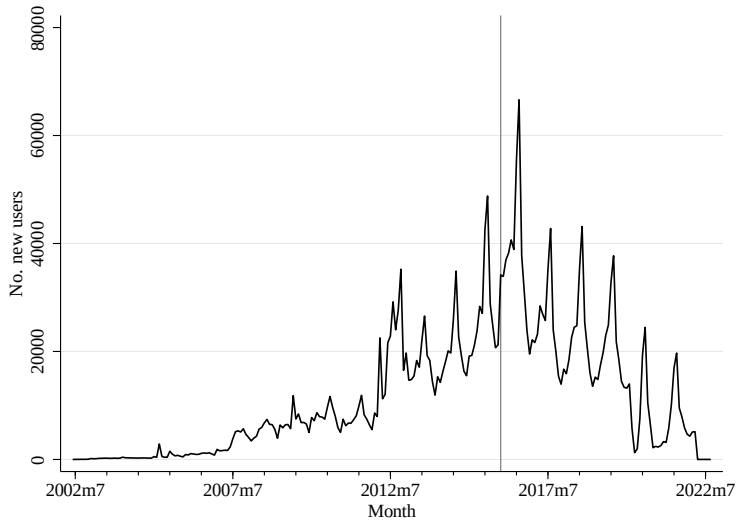
Notes: This table shows summary statistics for the users on Tripadvisor. *Source:* Own data collected from Tripadvisor (see Section 2 for details).

In Figure 13 it is possible to see the number of new users joining Tripadvisor each month. The first users joined Tripadvisor in July 2002 while the peak of new users is experienced during 2016. After reaching a peak in 2016, the number of new users starts to decrease. From Figure 13 it is also possible to notice some degree of seasonality, with more users joining during the high season. This is in line with what is observed for the number of posted reviews by country.

By using the information about the origin of the users, we can find their nationality. In Figure 14 it is possible to see the share of users from the most frequent nationalities. French users constitute the largest share with more than 30% followed by users from Spain, the United Kingdom and United States. The continents refer to all other countries in the specific continent, i.e. excluding the countries already in the figure. Overall, most users come from Europe followed by America.

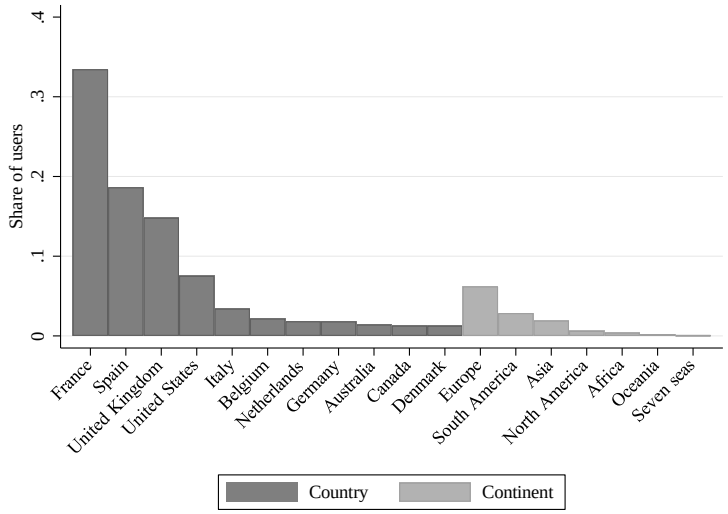
Finally, we look at variables explaining the tendencies of users when they post reviews. In Panel A of Figure 15 it is possible to see the share of users writing 1 review, 2-3 reviews, 4-5 reviews 6-10 reviews or more than 10 reviews on Tripadvisor. More than 60% of the users have posted only one review followed by about 20% posting 2-3 reviews. About 5% of users have written between 4-5 reviews, 4% between 6-10 reviews and 2.5% of users have written more than 10 reviews. Overall it can be concluded that even though many users only post one review, there is a good share posting more. In panel B of Figure 15 we divide users into intervals based on the number of months between

Figure 13: Number of users joining Tripadvisor over time



Notes: This Figure shows the number of new users joining Tripadvisor over time. The vertical line indicates January 2016 when our data starts. *Source:* Own data collected from Tripadvisor (see Section 2 for details).

Figure 14: Nationality of users on Tripadvisor visiting Denmark, France and Spain

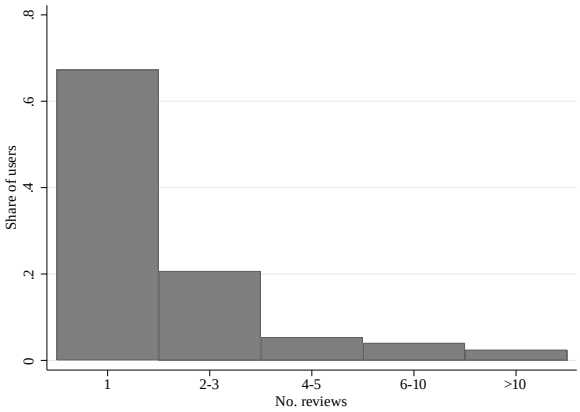


Notes: This Figure shows the nationality of users on Tripadvisor. Users refer to users posting reviews of Danish, French and Spanish attractions. Countries already included in the figure and not counted in the continents, therefore, continents refer to all other countries in that continent. *Source:* Own data collected from Tripadvisor (see Section 2 for details).

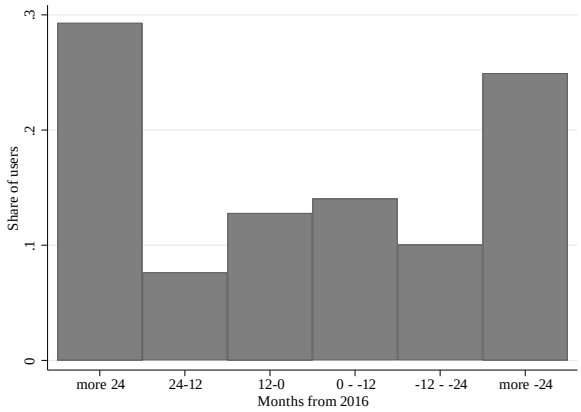
when they joined Tripadvisor and January 2016. About 30% of the users in our sample signed up more than two years before January 2016, while about 25% joined more than two years after. The remaining almost 50% of users joined between less than two years before and two years after. From Panel B of Figure 15 it is clear that users are continuously joining, but that more than half of them have already joined before our sample starts. Panel C of Figure 15 shows how active users are in posting reviews. Considering only users posting at least two reviews, we can compute the average

number of days elapsed between posting two reviews. About half of these users, on average, post another review within 30 days from the previous one and more than 85% post another review within a year. Finally, in Panel D of Figure 15 we categorize users based on whether they posted more than one review and, if they did, whether the reviews regarded different attractions. As already illustrated in Panel A, more than half of the users only posted one review, but in the vast majority of cases in which they have posted more than one, the reviews were for different attractions.

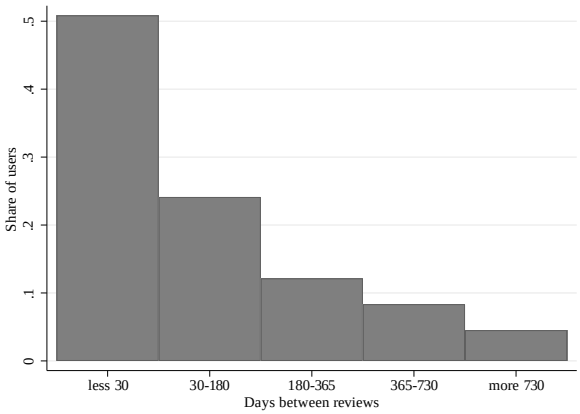
Figure 15: Users and their reviews



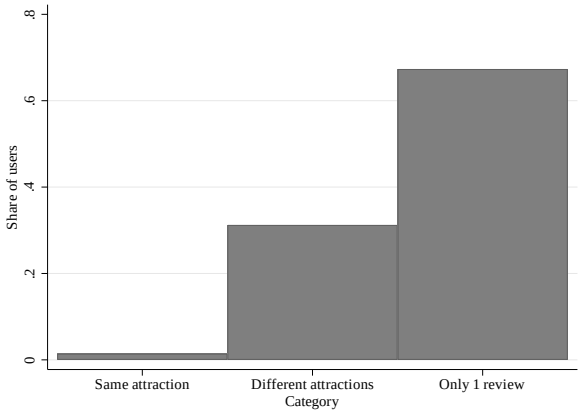
(A) Number of reviews per user



(B) Months from 2016



(C) Days between reviews



(D) Share of users reviewing one or more than one attraction

Notes: This Figure shows different variables describing the users' tendencies when posting reviews. Users refer to users posting reviews of Danish, French and Spanish attractions. Panel A shows the share of users within different intervals of the number of reviews written. 67% of users post one review, 21% post 2-3 reviews, 5% post 4-5 reviews, 4% post 6-10 reviews, and 2.5% post more than 10 reviews. Panel B shows the share of users within different intervals of month from January 2016. Panel C shows the share of users within different intervals of the average time between two reviews. Panel B shows the share of users reviewing different attractions. *Source:* Own data collected from Tripadvisor (see Section 2 for details).

4.3 The direction of tourism

In this section we look at measures that can explain the direction and intensity of tourism. In particular, we concentrate on the distance traveled by tourists, the density of attractions, the density of tourists and the average rating of attractions, all measured as explained in Section 2.[13] These are micro-level measures calculated using the unique disaggregated features of the database, including the geo-coded location from where the tourist traveled to (the origin) and the geo-coded location of an attraction (the destination).

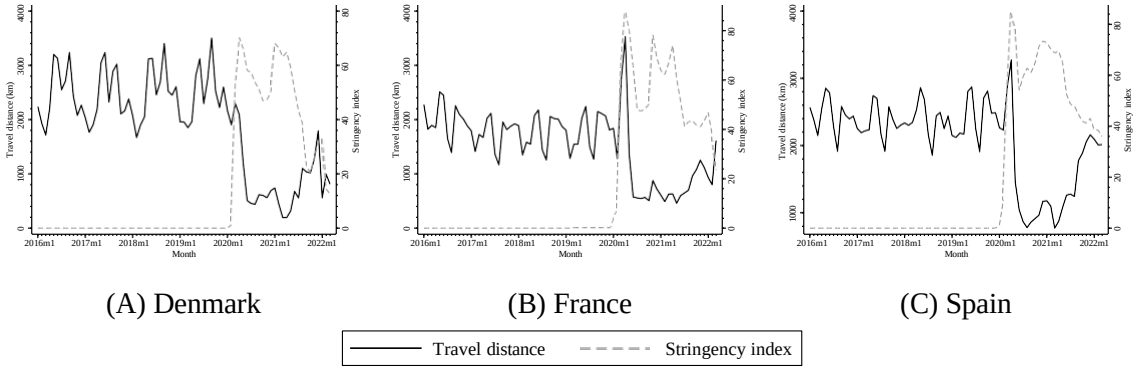
We start with a visual inspection by illustrating how the variables change over time together with the stringency index by country and month in Figures 16 - 19. For both the average travel distance and the two measures of density, there is a clear inverse relationship between the variables and the stringency index. When the stringency index is at its highest the three measures reach their minima. The large drops observed in all three measures with the introduction of restrictive measures are quite large, and by the end of our analysis in the beginning of 2022, most of them have not reached their pre-pandemic levels. The travel distance seems to be rebounding after 2021, especially for France and Spain, while the attraction density follows a similar trend. Turning to the tourist density, there is less indication of a rebound for all three countries, given that no clear increase can be observed after 2021.

In Figure 19 it is possible to see the average ratings of attractions by country. For the ratings the inverse relationship with the stringency index is less clear, even though for Denmark there is some indication that the average rating decreases, when the stringency index increases. For France and Spain the ratings almost follow the opposite trend, i.e. they increase after the introduction of restrictive measures. Looking at the ratings there also appears to be seasonality as for the other variables, but this time the lowest ratings are related to the high season, indicating that visitor satisfaction is lower during the most popular periods of the year.

To formally assess the impact of restrictive measures on the four variables above, we conduct a regression analysis similar the one conducted in Section 3.3. We use the Oxford stringency index as the explanatory variable, to find the estimated effect on the four outcomes of interest. The formal regression results can be seen in Table A11 in the appendix while Figure 20 summarizes the estimated impacts by showing how the variables change given changes in the stringency index. From Table A11 it appears that there is no significant impact on ratings of attractions due to the introduction of restrictions, we therefore concentrate on the other three measures in Figure 20. All three variables decrease when the stringency index increases. In terms of magnitude, a 10 percentage point increase in the stringency index implies about a 4% decrease in the travel distance while if the stringency index increases from 0 to 49 (the average) the travel distance decreases by almost 22%. In Panel B, a 10 percentage point increase in the stringency index also implies about a

[13] In this section we concentrate on the density measures using a 10km radius, but to make sure that our results and conclusions are not based on this choice, we show the corresponding results in the appendix Figures A4-A7.

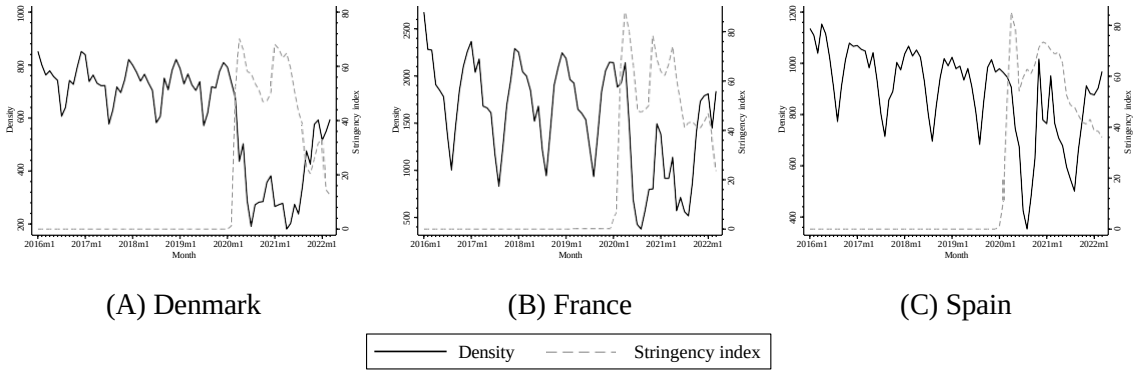
Figure 16: Travel distance to attractions and Stringency Index by country



(A) Denmark (B) France (C) Spain

Notes: This Figure shows the distance travelled to attractions together with the stringency index over time. 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Figure 17: Attraction density of visited locations



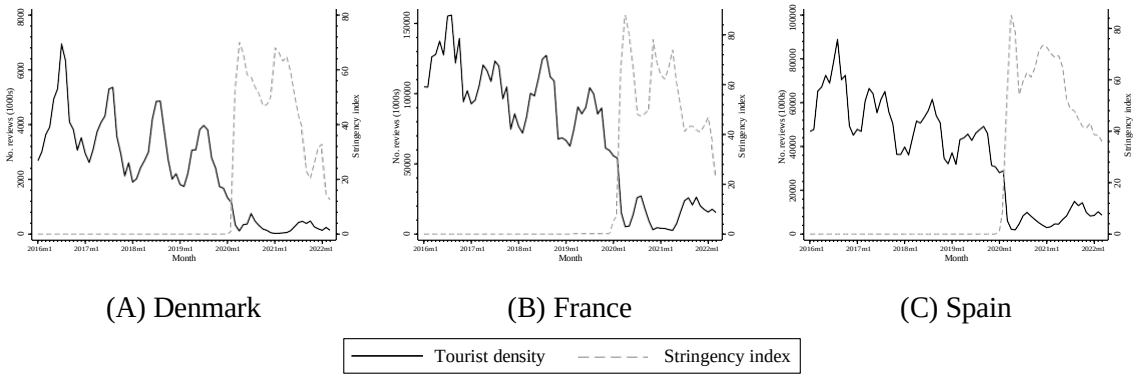
(A) Denmark (B) France (C) Spain

Notes: This Figure shows the attraction density of visited locations together with the stringency index over time. Panel A shows the attraction density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

4% decrease in the attraction density. In numbers this means 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. Finally in Panel C we show the changes in the tourist density, where, for example, a 10 percentage point increase in the stringency index implies a 34% decrease in the tourist density.

Overall, we can conclude that not only the number of reviews are significantly impacted by the introduction of restrictive measures, but the direction of tourism has also experienced decreases. These results indicate that there is a change towards nearer and less crowded locations. Together with the results from the previous section, we can conclude that apart from moving towards the nature, visitors also seek more isolated attractions after the pandemic. In conclusion, this section has provided evidence that individuals choose destinations closer to their home and also that they choose destinations that are less crowded.

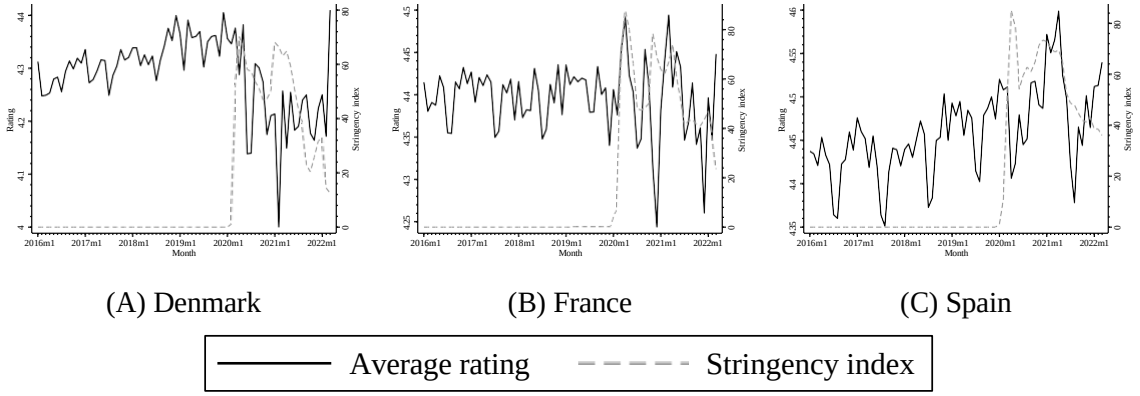
Figure 18: Tourist density of visited locations



(A) Denmark (B) France (C) Spain

Notes: This Figure shows the tourist density of visited locations together with the stringency index over time. Panel A shows the review density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Figure 19: Ratings and Stringency Index by country



(A) Denmark (B) France (C) Spain

Notes: This Figure shows the evolution of the average ratings of visited locations together with the stringency index. Panel A shows the average ratings of Danish attractions, Panel B of French attractions and Panel C of Spanish attractions. *Source:* Own data collected from Tripadvisor (see Section 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

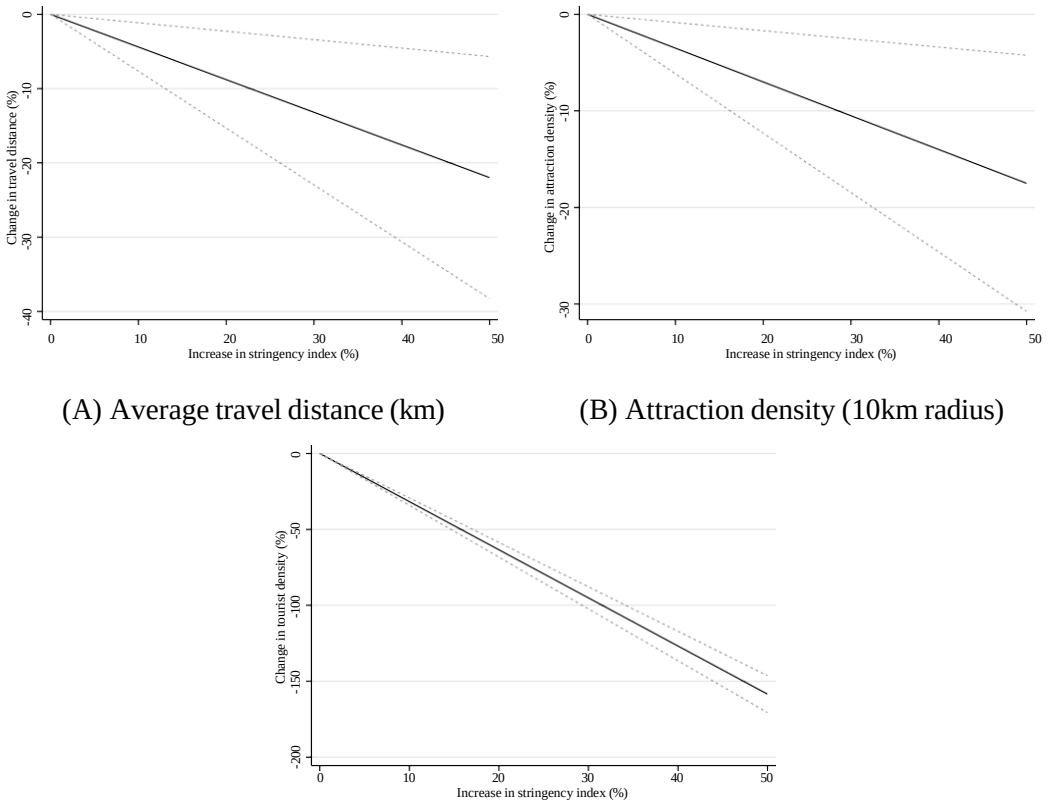
4.4 Analysis of regions and selected areas

In the rest of this section we concentrate on selected locations, to compare patterns both between countries, but also between popular destinations and the periphery.

4.4.1 Tourism activity in NUTS3 regions

To have an idea about how tourism activity is distributed within Denmark, France and Spain, for each attraction we identify the corresponding NUTS3 region. With this information at hand, we can compute the total number of reviews received by attractions within each NUTS3 region. The

Figure 20: Effect of a change in the stringency index on travel distance, attraction density and tourist density



(A) Average travel distance (km)

(B) Attraction density (10km radius)

(C) Tourist density (10km radius)

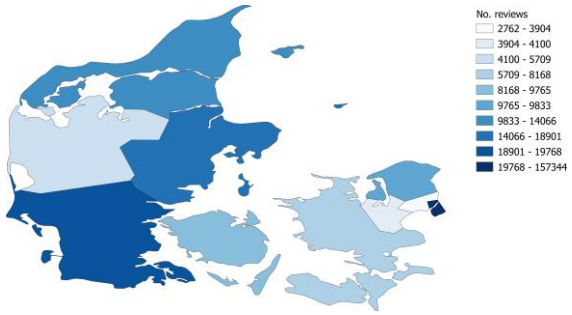
Notes: These figures shows the effect of a percentage change in the stringency index on travel distance (Panel A), attraction density (Panel B), and tourist density (Panel C), considering data aggregated monthly at the country and attraction category level. The dashed lines indicate the 95% confidence intervals. *Source:* Own calculations based on the results from Table A11 in the appendix.

resulting maps with the number of reviews can be seen in Figure 21 Panels A, C and E. Darker blue shaded regions have more reviews, and hence more tourism activity. We have also computed the percentage change in the number of reviews between yearly average before the pandemic and yearly average after the pandemic which is illustrated in Panels B, D and F. Darker reds indicate bigger drops in the number of reviews.

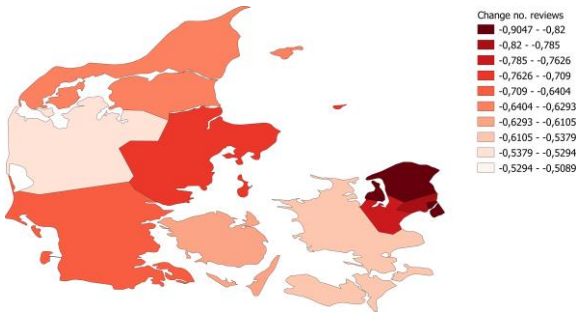
In Panel A, it is clear that the NUTS3 regions where Copenhagen is located receives most reviews followed by Southern- and Northern Jutland where some very popular travel destinations are also located close to the coasts. In Panel B, the largest drops can be observed in the NUTS3 regions around Copenhagen, while the regions experiencing the smallest drops are also the ones with less tourism activity.

In Panel C, the French NUTS3 regions with most activity are the coastal regions and Paris. The same regions are, in most cases, also the regions experiencing the largest drops in Panel D. Also

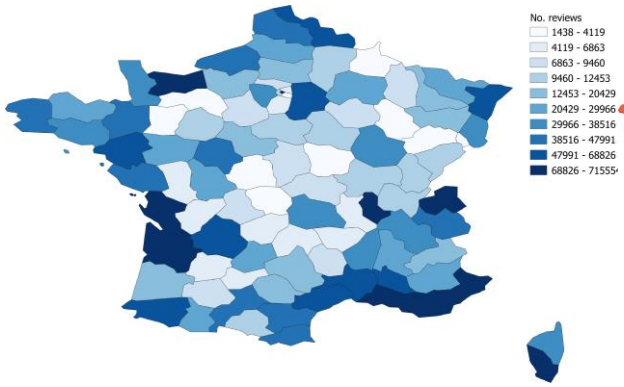
Figure 21: Number of reviews and change



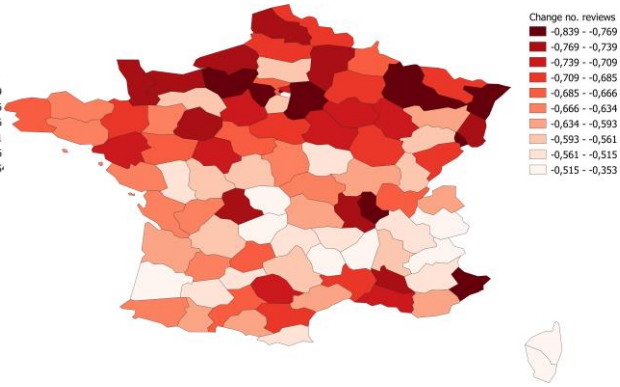
(A) Number of reviews - Denmark



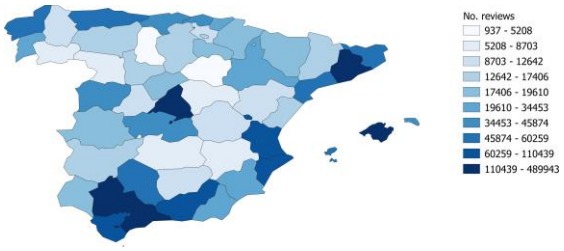
(B) Change no. reviews (%) - Denmark



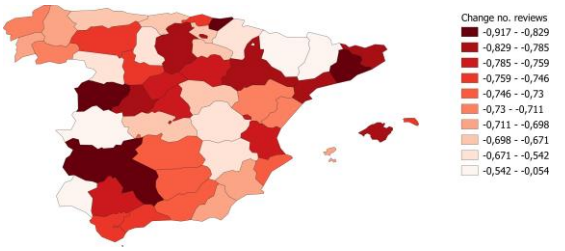
(C) Number of reviews - France



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



(E) Number of reviews - Spain



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

Notes: These 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, C 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 2 for details).

border regions in Northern France experience large drops in the number of reviews, which could be a sign of less traveling across the borders. On the other hand, Corsica, which has a high number of reviews, does not experience the same large drop as other regions with high activity.

In Panel E, Spain has a similar pattern to France, with most activity along the coasts and the bigger

cities such as Madrid and Barcelona. Once again, it seems that regions with more tourism activity also experience larger drops, with the exception of some of the regions along the Portuguese border which experience large drops, but are not among the regions with most activity. Regions along the French border do not experience the same drops.

4.4.2 Tourism in selected areas

Finally, we look at tourism activity in some selected areas which differ in size and characteristics. We concentrate on seven different areas, two in Denmark and France and three in Spain. The areas are selected to be representative for different levels of tourism activity, from very popular destinations to more peripheral ones. They are also selected to show differences between smaller and larger places. In Denmark we look at the NUTS3 region of Copenhagen as an example of a popular destination and the NUTS3 region of the island of Funen as an example of a periphery. The NUTS3 region of Copenhagen corresponds also to the area of the city of Copenhagen. In France we concentrate on the NUTS3 regions of Paris and Nice. They can both be considered popular destinations, as can also be seen from Figure 21, Panel C. However, they are somewhat different in nature and the purpose of the travel, given that Paris is a big inland city, while Nice is a coastal locality with a different set of possible activities. Given this, the two locations can show differences in how they are impacted by the pandemic. Similarly to Copenhagen, the NUTS3 region of Paris corresponds to the area of the city of Paris. Finally, for Spain we look at the NUTS3 region of Asturias together with the seaside resort of Denia and the city of Valencia. Valencia and Denia can be expected to be similar, given that they are located close to each other and they are both more popular destinations close to the sea. However, Denia differs from Valencia, since it is a relatively small place on the coast with no major indoor leisure facilities. Hence, comparing Valencia and Denia shows differences between smaller and larger locations. On the other hand, comparing Valencia to Asturias can show differences between a large popular destination and a more peripheral one, where the size of tourism is smaller and the activities are different in nature.

The analysis is divided into two different parts, both based on a visual inspection through maps and figures. First we create maps with travel patterns similar to the ones illustrated in Subsection 4.1. We create four maps for each of the seven selected areas as explained above, to show patterns in international, European, domestic and local tourism, covering the entire period from 2016 to 2022 for which we have data.

In Figures A8-A14 we show the four maps for each of the seven areas. For Denmark, it is clear that the number of visitors from around the world is much larger for the region of Copenhagen than for Funen (see Figures A8 and A9 Panel A). Most of the visitors to Funen are from North America, while for Copenhagen visitors come from all around the world. The same is true for European tourists, even though it seems that for both Copenhagen and Funen they come from all over Europe with higher shares from the UK and Germany. Finally, regarding domestic and local tourism, in

both cases this is distributed from all over Denmark with no clear pattern.

For France, Paris and Nice are similar when looking at tourists from around the world, with the main difference that, in numbers, there are more visitors to Paris (Figures A10 and A11). European visitors come from all around Europe and once again, especially from the UK in both cases. Looking at domestic travel, many visitors travel from Paris to Nice, but also from all other regions. At the local level, many visitors come from the outer parts of Paris to visit the city center while for Nice there are no clear patterns.

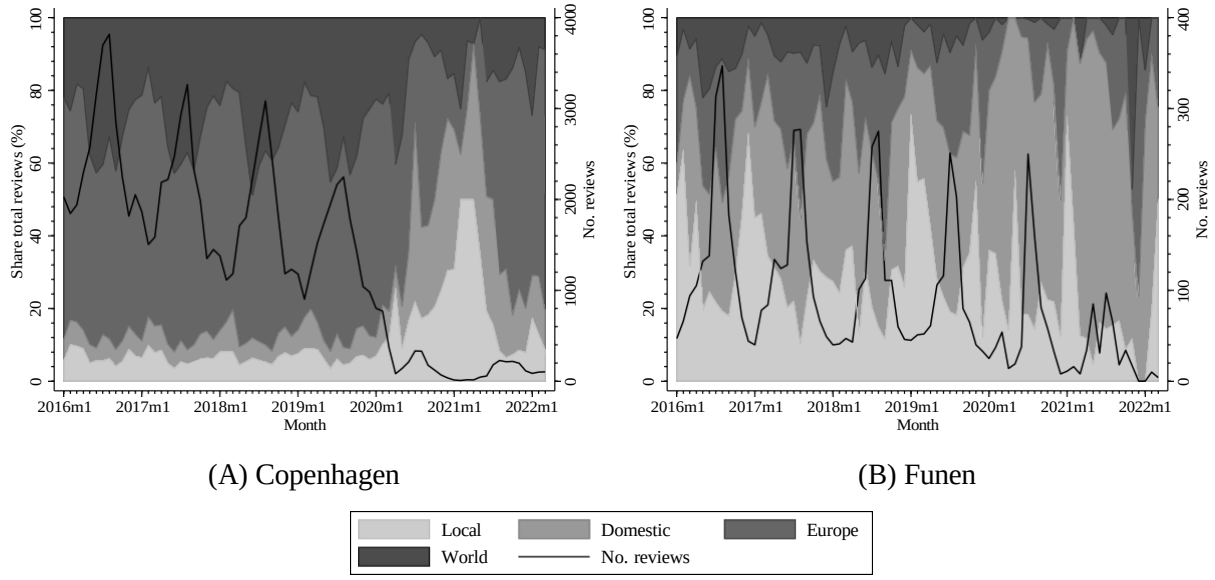
Finally, in Spain the patterns for visitors from around the world are similar in all three selected areas (Figures A12-A14). Once again there are many visitors from the UK in all three cases, and no clear patterns when looking at domestic and local travel.

For the final part of the analysis, we construct time series to see how the shares of the different travel categories and the total number of reviews change over time in each of the areas. In Figure 22 it is possible to see the share for the NUTS3 regions of Copenhagen (Panel A) and Funen (Panel B) in Denmark. The two panels show clear differences in the patterns, with Copenhagen having a much higher share of visitors from both Europe and the rest of the world. When domestic tourism is at its highest in Copenhagen before 2020, it only reaches about 20% of all tourism during the low season. For Copenhagen, the composition of tourists changes after the outbreak of the pandemic, where suddenly the number of tourists drops, together with the shares of foreign tourists. From summer 2021 foreign tourism increases again, but without reaching the levels from before 2020. When looking at Funen in Panel B, the share of tourists from Europe and other countries in the world is much lower, while both domestic and local tourism represent large shares of the overall numbers, reaching together more than 80% during the low season. At the same time there is also no change in the composition after the outbreak of the pandemic, even though the total number of reviews decreases. Interestingly, there is no substantial difference in the number of reviews when comparing summer 2019 and summer 2020 which followed a period of Covid-19 related restrictions. This could be explained by the fact that a big tourism campaign in Denmark took place in the spring of 2020, encouraging Danish tourists to explore their own country, hence boosting domestic tourism especially in the periphery.

In Figure 23 we show the composition of tourists in the NUTS3 regions of Paris (Panel A) and Nice (Panel B). In terms of the composition before the outbreak of Covid-19, the two localities are very similar, with a high share of visitors from outside France. They also both experience a drop in the number of reviews from 2020, but, as expected, they differ in their composition. In Paris the share of both domestic and local tourists increases, reaching almost 80% following the first lockdown. In Nice the share of tourists from outside Europe decreases, while domestic tourism increases very much. The main difference between Paris and Nice after 2020 is the share of local tourists, which suddenly becomes much larger in Paris and domestic tourists which increases in Nice.

The composition of tourists in the three selected areas in Spain are illustrated in Figure 24. In Panel A, the city of Valencia is similar to both Copenhagen and Paris, with a high share of international

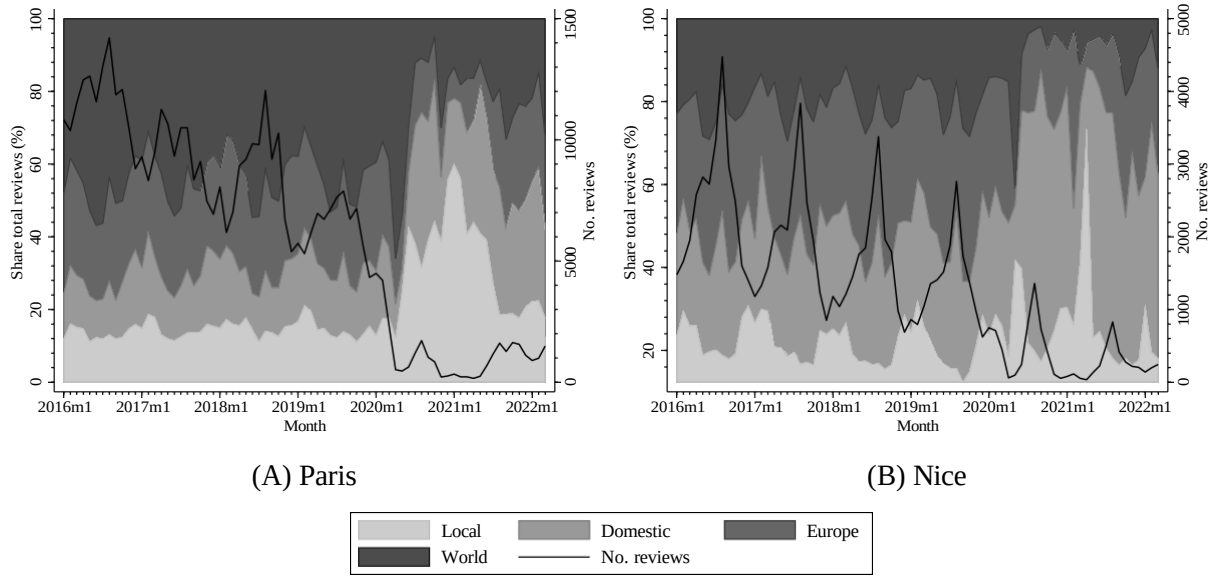
Figure 22: Share of reviews for different travel categories over time at selected areas - Denmark



(A) Copenhagen (B) Funen

Notes: This Figure shows the share of reviews out of the total in the following four travel categories: local, domestic, Europe and world. Regions refer to the NUTS3 regions of Copenhagen (Panel A) and Funen (Panel B). *Source:* Own data collected from Tripadvisor (see Section 2 for details).

Figure 23: Share of reviews for different travel categories over time at selected areas - France

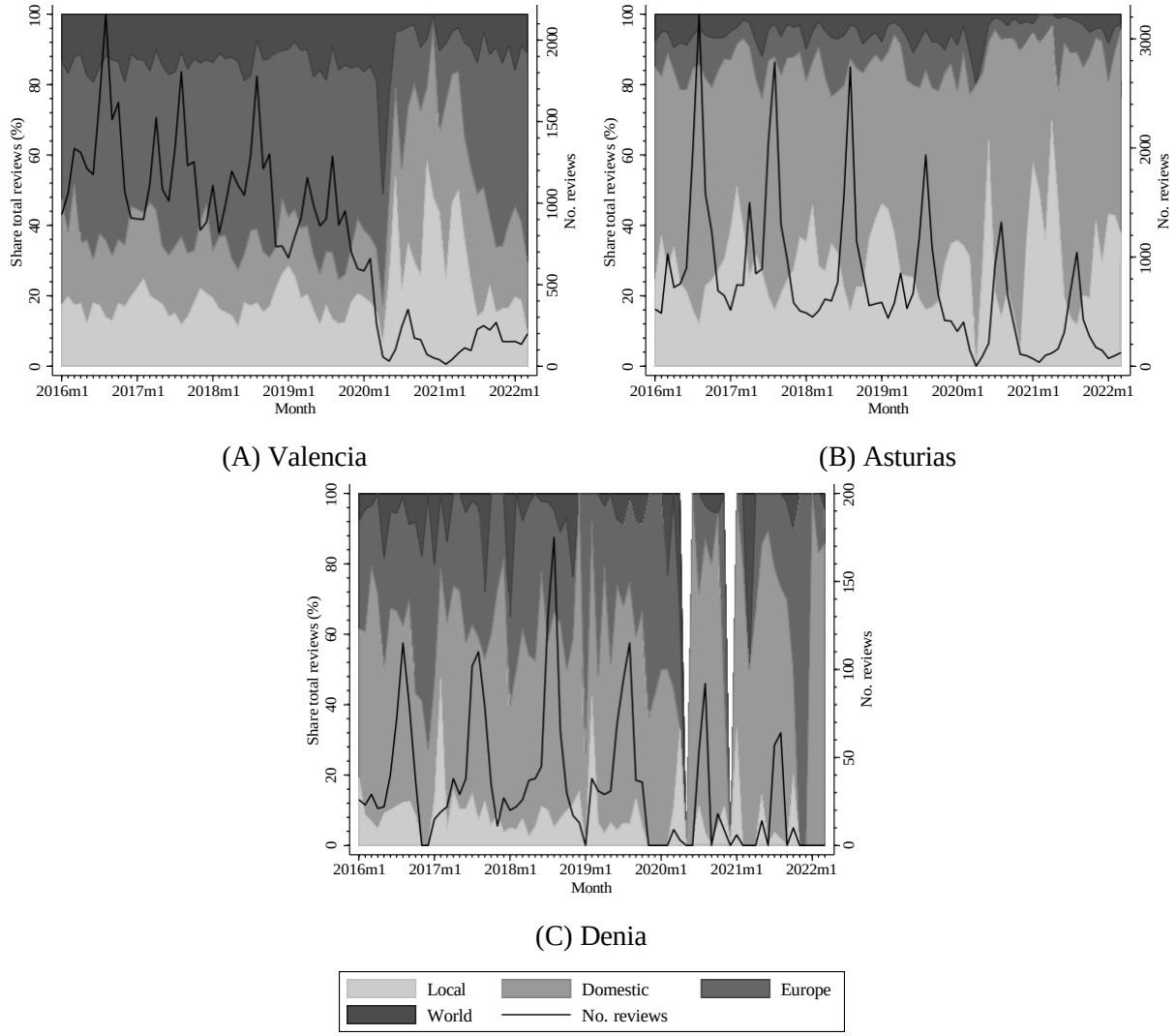


(A) Paris (B) Nice

Notes: This Figure shows the share of reviews out of the total in the following four travel categories: local, domestic, Europe and world. Regions refer to the NUTS3 regions of Paris (Panel A) and Nice (Panel B). *Source:* Own data collected from Tripadvisor (see Section 2 for details).

visitors before 2020 followed by a drop and a share of domestic and local tourists which increases after the outbreak of the pandemic. On the other hand, the region of Asturias is very different in the composition of tourists, with more than 80% of them being local or domestic. As for Funen in Denmark, Asturias does not experience any clear change in composition after 2020, even though

Figure 24: Share of reviews for different travel categories over time at selected areas - Spain



Notes: This Figure shows the share of reviews out of the total in the following four travel categories: local, domestic, Europe and world. The three panels refer to the following areas: city of Valencia (Panel A), NUTS3 region of Asturias (Panel B) and the city of Denia (Panel C). *Source:* Own data collected from Tripadvisor (see Section 2 for details).

there is a large drop in the total number of reviews. The city of Denia in Panel C is somewhat in-between Valencia and Asturias with about 60% of local and domestic tourists and 40% from outside Spain. However, there is also an indication that the composition of visitors in Denia changes slightly after 2020 similarly to Valencia, but less pronounced. In two places (in summer 2020 and January 2021), the graph disappears indicating that there were no reviews.

Common for all the selected areas is that they exhibit seasonality. At the same time the share of domestic/local tourists is higher during the winter indicating that locals choose to visit their own country when the number of foreign tourists is at its lowest. From Figures 22-24 it is also clear that the top destinations and the peripheral destinations were affected differently by the pandemic and the related restrictions and lockdowns. The popular destinations such as Paris experienced big

changes in tourist composition, with a sudden increase of the domestic, while the periphery did not experience the same changes, even though the number of reviews decreased in all cases.

5 Limits of the approach

This project presents a novel approach to understanding tourism dynamics using digital data sources. However, several limitations are inherent in this study, which must be considered when interpreting its findings and conclusions, or when extending the approach to other contexts.

Representation of Trips through Reviews

A limitation of the study is the assumption that one review on Tripadvisor equates to one trip. This correlation might not always be accurate, as there are likely missing observations where travelers do not leave reviews for every trip they make. Moreover, the study's ability to understand complete travel histories is constrained, as not all tourists write reviews for places they visit. This results in an incomplete representation that may skew the data, particularly understating the frequency or popularity of certain destinations or attractions. To illustrate patterns in how users post reviews, we have shown the distribution of users writing different numbers of reviews. We found that more than 60% posted only one review followed by about 20% posting 2-3 reviews. Furthermore, approximately 5% of users have written between 4 and 5 reviews, 4% between 6 and 10 reviews and 2.5% of users have written more than 10 reviews.

In addressing the limitation that one review on Tripadvisor may not accurately represent one trip, we presented a range of validity tests to substantiate the reliability of the data. These tests were designed to assess how well our dataset from Tripadvisor reflects the overall travel flows, movements, and dynamics of tourism.

Firstly, we compared our Tripadvisor data with official tourism statistics, such as those provided by Eurostat. This comparison allowed us to evaluate whether the trends and patterns observed in the Tripadvisor data were in alignment with established, official tourism data sources. The high correlation found between these datasets provided a strong indication that, despite the potential for missing observations in the Tripadvisor data, it still offers a valuable approximation of overall tourism activity.

Furthermore, we conducted analyses to understand the seasonal and yearly trends in tourism as reflected in our data, comparing these trends with known patterns in the tourism industry. These analyses helped in affirming that the data captured by Tripadvisor reviews closely mirrored the expected fluctuations and movements in the tourism sector.

It is also important to consider that while a small number of tourists leave reviews regularly, the sheer volume of data available through Tripadvisor — encompassing millions of reviews — allows for a comprehensive overview. The vast amount of data compensates to a significant extent for individual missing observations, as it captures a wide cross-section of tourist experiences and behaviors.

Additionally, our approach to analyzing the data involved advanced statistical techniques that account for potential biases and gaps in the data. By employing these robust analytical methods, we were able to extract meaningful insights that are reflective of the broader trends in tourism, despite the inherent limitations of the data source.

In summary, while the assumption that one review equates to one trip may apply only to a small number of tourists, the range of validity tests and analytical methods employed in our study demonstrate that the Tripadvisor data provides a reliable approximation of the overall travel flows and dynamics in the tourism sector. The data, therefore, remains in our assessment a valuable asset for understanding tourism trends, especially when supplemented with additional sources and rigorous analytical approaches.

Understanding User Demographics and Data Disaggregation

The research faces limitations in gaining comprehensive knowledge about users who contribute reviews on Tripadvisor. This affects the depth of understanding regarding the demographics and preferences of the tourists, potentially impacting the study's conclusions about tourist behaviors. Disaggregating the data to a granular level presents its own set of challenges. While it offers detailed insights, over-disaggregation may lead to less reliable or meaningful conclusions, as the data might become too sparse to provide a representative overview.

We mitigate these limitations in several ways. While comprehensive demographic details of each user are not available, our research maximizes the use of available data to extract significant trends. We examine user behavior patterns, encompassing factors such as the frequency of reviews, time intervals between posts, and - for a sample of users - their comprehensive global travel histories. This approach allows for a nuanced understanding of user profiles and travel behaviors, compensating for the lack of specific demographic information.

To address the risks of data over-disaggregation, we carefully balance the level of detail with the need to maintain data representativeness. Our methodology ensures that while the data is dissected to provide detailed insights, it remains robust and reflective of broader tourism trends. Advanced statistical methods, including robustness checks, further reinforce the validity of our findings despite the granularity of the data.

Additionally, the study supplements the Tripadvisor data with external tourism statistics for cross-validation, which strengthens the overall reliability and applicability of our research outcomes.

Use of Oxford Covid-19 Government Response Tracker's Stringency Index

The study's reliance on the Oxford Covid-19 Government Response Tracker's stringency index to gauge the impact of the pandemic on tourism introduces certain limitations. The primary concern is that the index may not capture the nuances of local or regional variations in movement and socialization restrictions within countries. These localized measures can have a substantial impact on tourism patterns and behaviors, which the more generalized stringency index might not fully reflect. For instance, a region may have specific restrictions that differ significantly from the national policies, thus affecting local tourism trends in ways that the stringency index does not capture.

However, it is important to acknowledge the Oxford stringency index's role as a widely recognized and utilized tool in global research for tracking government responses to the pandemic. Its standardization across different countries and time periods provides a valuable, consistent metric for large-scale analysis and comparison of tourism trends across Europe.

To address potential shortcomings of the index, we supplement the stringency index with other measures such as travel restrictions, to check for robustness in our results. Furthermore, the drop and relocation in tourism activity is so significant in the months of the Covid-19 pandemic, that just a raw time-series is instructive and delivers significant results.

Coverage and Comprehensiveness of Attraction Listings

The list of attractions used in the study is extensive but not exhaustive. It primarily includes attractions mentioned in user reviews, potentially omitting lesser-known sites. Additionally, attractions that do not require economic transactions, like free, open spaces, may be underrepresented in the data.

Although the list is derived from user reviews on Tripadvisor and may not be exhaustive, encompassing primarily those attractions that are reviewed by users, it represents one of the most extensive collections of tourist attraction data used in academic research to date.

While the dataset might underrepresent attractions that do not require economic transactions, including some lesser-known locations, the extensive dataset developed here marks a significant step forward in tourism research. It provides a foundation for future studies to build upon, enhancing the overall understanding of tourism trends and contributing to the development of more informed tourism policies and strategies.

Classification of Attractions as Indoor or Outdoor

Classifying attractions as indoor or outdoor is complex and not always clear-cut. For example, a well-known cultural heritage site in France or Spain might have both indoor and outdoor elements,

complicating its categorization. This classification also does not account for indoor activities that might have implemented contagion-minimizing measures, such as visitor limits or enhanced ventilation.

The study's approach to classifying attractions as indoor or outdoor, while complex, represents a pragmatic method to analyze tourism patterns under pandemic conditions. It is acknowledged that certain attractions, like many cultural heritage sites, encompass both indoor and outdoor elements, making classification challenging.

In our study, we have adopted a nuanced approach to categorizing attractions, moving away from a simplistic binary classification of indoor or outdoor. Instead, we categorized attractions as Museums, Sights & Landmarks, and Nature & Parks. On average, these categories correspond to an ordinal scale based on a gradation from the most indoor to the most outdoor types of attractions. Museums, typically enclosed spaces, are considered the most indoor-oriented attractions. Sights & Landmarks, often a mix of indoor and outdoor elements, are less indoor than Museums but more so than Nature & Parks. Finally, Nature & Parks are the most outdoor-oriented attractions, representing the other end of the spectrum.

Representation of Local Tourism

Local tourism is possibly underrepresented in the data, a limitation shared with conventional tourism statistics. This underrepresentation can skew the understanding of tourism dynamics, especially in understanding the nuances of domestic travel patterns.

The potential underrepresentation of local tourism in our data presents a limitation, but it is a characteristic it shares with traditional tourism statistics. It is important to emphasize that despite this limitation, our research provides valuable disaggregated insights into broader tourism trends. The depth and breadth of the dataset from Tripadvisor still allows for a detailed analysis of various aspects of tourism, including regional and international travel patterns.

Tripadvisor's Evolving Usage

The declining popularity of Tripadvisor raises questions about the long-term value of the database for studying the evolution of tourism. This trend might be reflective of broader shifts in social media platform usage. This trend, which reflects broader shifts in social media and digital platform usage, is an important consideration. However, it is crucial to note that the extensive historical data available through Tripadvisor remains a valuable resource for understanding past and present tourism trends.

Moreover, the inclusion of month and year controls in our statistical analyses enables us to account for seasonality effects as well as any changes in the platform's popularity over time. This approach

allows us to adjust for potential shifts in user behavior and ensures that our findings remain relevant and reflective of actual tourism activities, despite fluctuations in the platform's usage.

Limitations of Location Data

In our data we have identified coordinates of users' reported locations and the location of attractions, based on all available information on Tripadvisor. However, in almost all cases, Tripadvisor only reports the administrative areas and not exact addresses. This feature of Tripadvisor restricts the accuracy of assessing visitor density and travel distances, particularly for short trips such as local tourism within one's own city. This is a critical issue in the context of a pandemic, where precise location data is crucial for understanding tourism patterns.

The study acknowledges the limitation presented by the location data of attractions, primarily confined to administrative areas (e.g., a city), which can impact the precision in assessing visitor density and travel distances. This aspect is particularly relevant for short-distance travel, such as local tourism within one's own city.

However, it is important to highlight the broader context of our research, which focuses on larger-scale tourism trends, including across cities, regions or countries. The use of administrative area-based data still provides valuable insights into regional and national tourism flows.

External Influences and Temporal Limitations in Tourism Study

External factors unrelated to tourism, such as economic changes, political events, or technological advancements, could have influenced the study's findings but were not fully accounted for. These external factors might confound the interpretation of the data. Also, the study is constrained by its temporal scope. Changes in tourism trends over time, especially post-pandemic, might not be fully captured, affecting the relevance and applicability of the findings, particularly in the rapidly evolving post-Covid-19 era.

While these broader elements were not exhaustively examined, our focus was to analyze tourism trends in the context of the Covid-19 pandemic based on Tripadvisor data. This approach provides targeted insights into the pandemic's immediate impact on tourism, acknowledging that the sector is influenced by a myriad of interconnected factors. Furthermore, our statistical analysis include various controls for unobserved events and variables. For example, the inclusion of year fixed effects accounts for unobserved events that are specific to a given year).

Regarding the temporal scope, the study captures a specific period, primarily during the pandemic. It is acknowledged that the evolving nature of tourism trends, especially in the post-pandemic era, may extend beyond the timeframe of our research. However, our study's methodologies and findings offer a valuable framework for future research. Other researchers can build upon our

approaches to explore the long-term and post-pandemic effects on tourism, thereby expanding the understanding of this dynamic sector's ongoing evolution.

6 Conclusion

The underlying research marks a significant advance in understanding the complexities and evolving dynamics of the tourism sector, especially in the wake of the Covid-19 pandemic. This study has not only shed light on the changing patterns of tourist behavior and preferences but also underscores the importance of adaptive and sustainable tourism strategies.

The use of big data in this research validates and demonstrates the power of digital tools in capturing tourism flows and tourist preferences. This potentially offers a valuable resource for policymakers and industry stakeholders.

The use of big data to measure cultural and nature tourism in Europe reveals significant insights that can guide policy-making in the tourism sector, particularly in the post-pandemic landscape. A key finding is the change in tourist behavior due to the Covid-19 pandemic, with a noted preference for outdoor activities and less crowded destinations. The pandemic has also led to a change in the types of attractions tourists prefer, with an inclination towards lesser-known sites that are located nearer the visitor.

This study also sets a new direction for tourism scholarship, by blending traditional approaches with innovative big data analysis, to more effectively understand and respond to the dynamic nature of global tourism. The future of tourism in Europe lies in its capacity to adapt, innovate, and sustainably grow, leveraging the rich cultural and natural heritage that it possesses.

7 Acknowledgments

We are grateful for the Runner-up *Best Paper Award* received at the 9th Advances in Tourism Marketing Conference at the University of Bordeaux in 2023. We extend our sincere gratitude to Ilde Rizzo, Vincent Kaufmann, Guillaume Drevon, Martin Falk, Tiziana Cuccia, and Enrico E. Bertacchini for their invaluable contributions and insights. Our thanks also go to the participants of the Association for Cultural Economics International Conference 2023 in Bloomington, the 9th Advances in Tourism Marketing Conference, and seminar attendees at the University of Southern Denmark for their constructive feedback and support.

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.

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 over-tourism in established european urban destinations: Assessing the impact of tourism in european urban destinations. *International Journal of Tourism Cities*, 6(1):117–137.

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 Mitchell, S. B. (2024). A novel application of big data to measure trends in tourism - Data Manual and Description. Data manual, University of Southern Denmark.

Borowiecki, K. J., Pedersen, M. U., and Mitchell, S. B. (2023). Using big data to measure cultural tourism in Europe with unprecedented precision. Discussion Papers on Economics, Working Paper No 5/2023, University of Southern Denmark.

Borowiecki, K. J., Pedersen, M. U., Mitchell, S. B., and Alam Khan, S. (2024a). Navigating cultural landscapes: Approaches to data collection and analysis in tourism. In Borowiecki, K. J., Civantos, J. M., and Fresa, A., editors, *Visiting the Margins: Innovative Cultural Tourism in European Peripheries*, Routledge, chapter 5. London, forthcoming.

Borowiecki, K. J., Pedersen, M. U., and Palomeque, M. (2024b). Putting the Periphery on the Map: Tourism Activity measured with Big Data. Working paper, work in progress, University of Southern Denmark.

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.

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.

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

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

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

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*.

A Appendix

Tables

Table A1: Summary statistics

Panel A - Reviews 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 (2023) and own data collected from Tripadvisor (see Section 2 for details).

Table A2: Summary statistics Denmark

Panel A - Reviews 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 (2023) and own data collected from Tripadvisor (see Section 2 for details).

Table A3: Summary statistics France

Panel A - Reviews 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 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 (2023) and own data collected from Tripadvisor (see Section 2 for details).

Table A4: Summary statistics Spain

Panel A - Reviews 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 (2023) and own data collected from Tripadvisor (see Section 2 for details).

Table A5: 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: This table shows summary statistics at the individual level for the global attraction data. *Source:* Own data collected from Tripadvisor (see Section 2 for details)

Table A6: Correlation coefficients between Tripadvisor reviews and Eurostat arrivals

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)
Sample	All	Denmark	France	Spain

Notes: Simple correlation coefficients between number of reviews from Tripadvisor and number of arrivals from Eurostat. *Source:* Official tourism statistics from Eurostat (2022) and own data collected from Tripadvisor (see Section 2 for details).

Table A7: Validity Test: Regression results

	(1) ln(Reviews)	(2) ln(Reviews)	(3) ln(Reviews)	(4) ln(Reviews)
ln(Arrivals)	0.637***	1.130***	0.674***	0.653***
	(0.050)	(0.118)	(0.090)	(0.035)
Country FE	Yes	No	No	No
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Sample	All	Denmark	France	Spain
N	222	74	74	74
R^2	0.961	0.953	0.889	0.974

Notes: Regression results when estimating the number of Tripadvisor reviews on the number of arrivals from Eurostat. Column 1 uses the entire sample and columns 2-4 show 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 2 for details).

Table A8: 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***	-0.323***	-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 2 for details) and the stringency index from the Oxford Government Response Tracker.

Table A9: 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 2 for details) and the international travel restrictions indicator from the Oxford Government Response Tracker.

Table A10: 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 2 for details) and the stringency index from the Oxford Government Response Tracker.

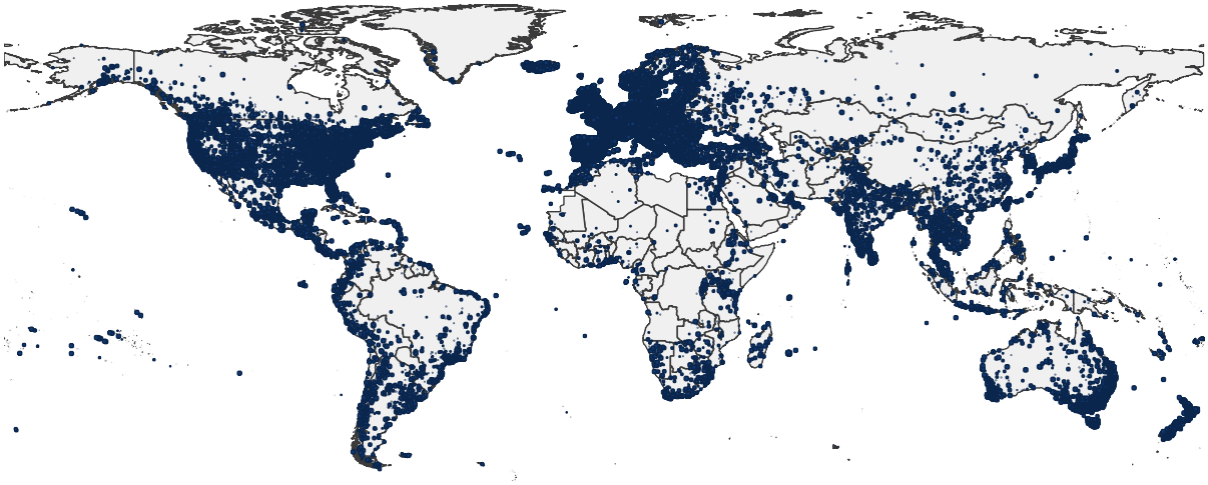
Table A11: Effect of stringency index on travel distance, attraction density, tourist density and ratings

	(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 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 $10km$, 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 2 for details) and the stringency index from the Oxford Government Response Tracker.

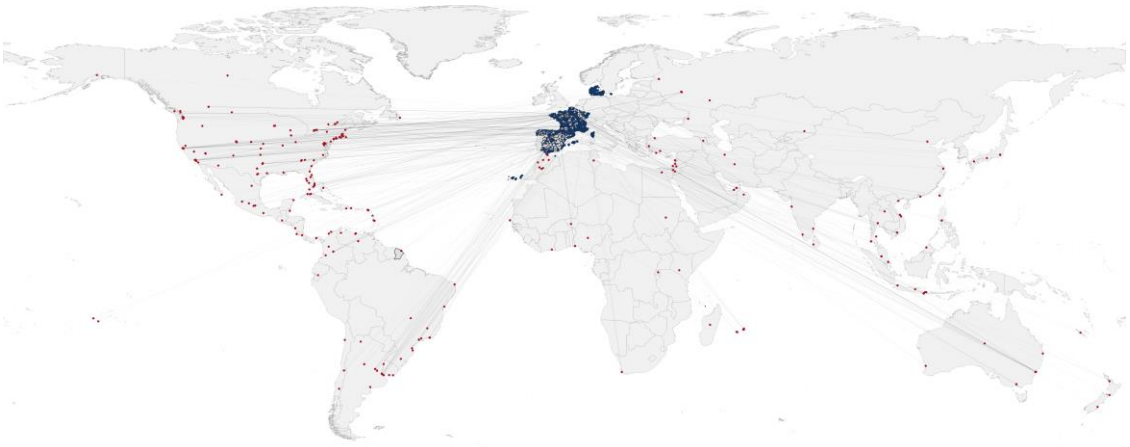
Figures

Figure A1: Attraction locations using global sample

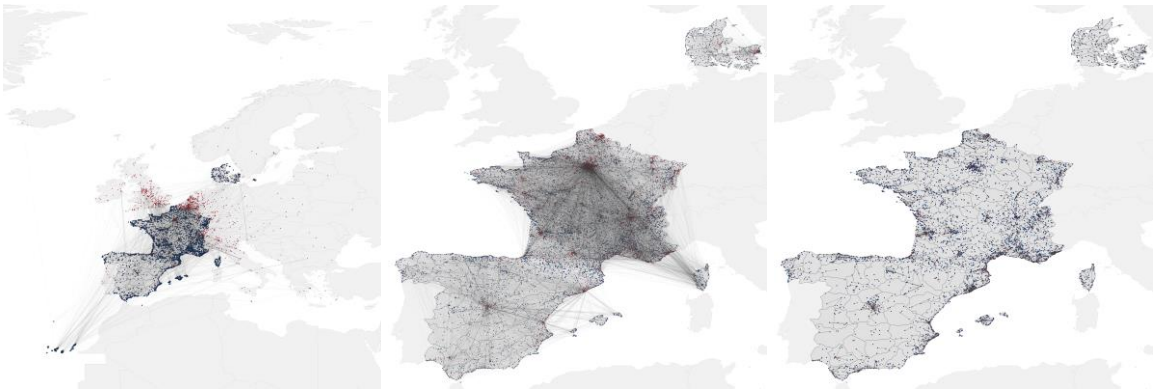


Notes: This figure shows the location of attractions present all over the world reviewed by visitors who also reviewed at least one attraction in Denmark, France or Spain. *Source:* Own data collected from Tripadvisor (see Section 2 for details).

Figure A2: Travel patterns - Easing of first lockdown



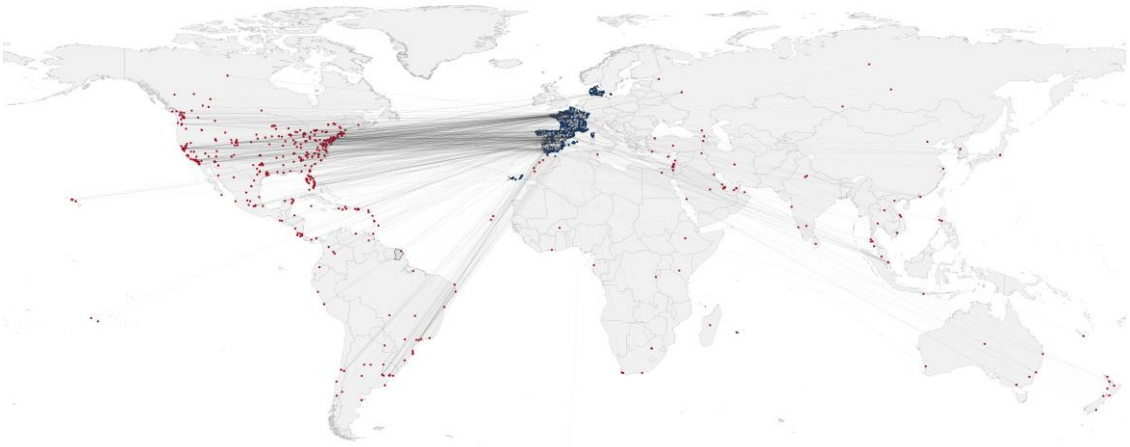
(A) World



(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the period after the first lockdown, from June 2020 to September 2020. The blue dots represent the location of attractions and the red dots the location users. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A3: Travel patterns - Second easing of restrictions

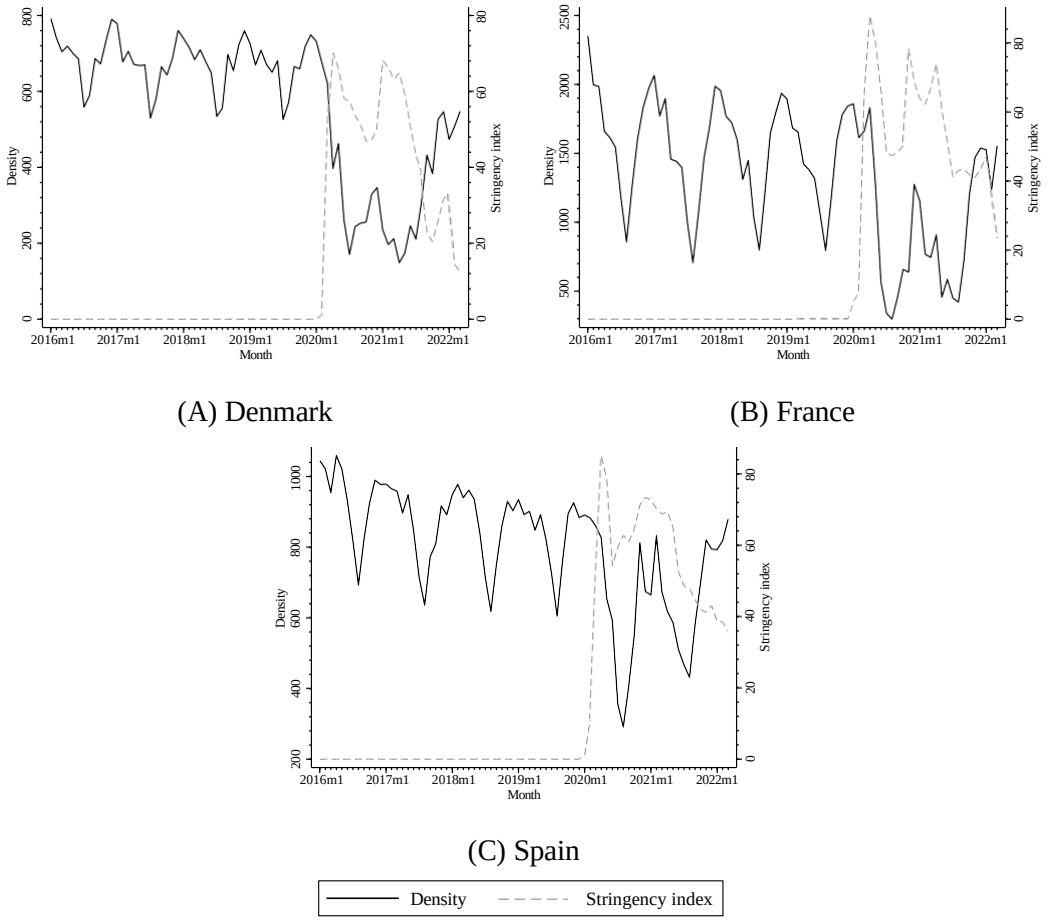


(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the period following the second easing of restrictions, from September 2021 to December 2021. The blue dots represent the location of attractions and the red dots the location users. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A4: Attraction density of visited locations - 5km radius



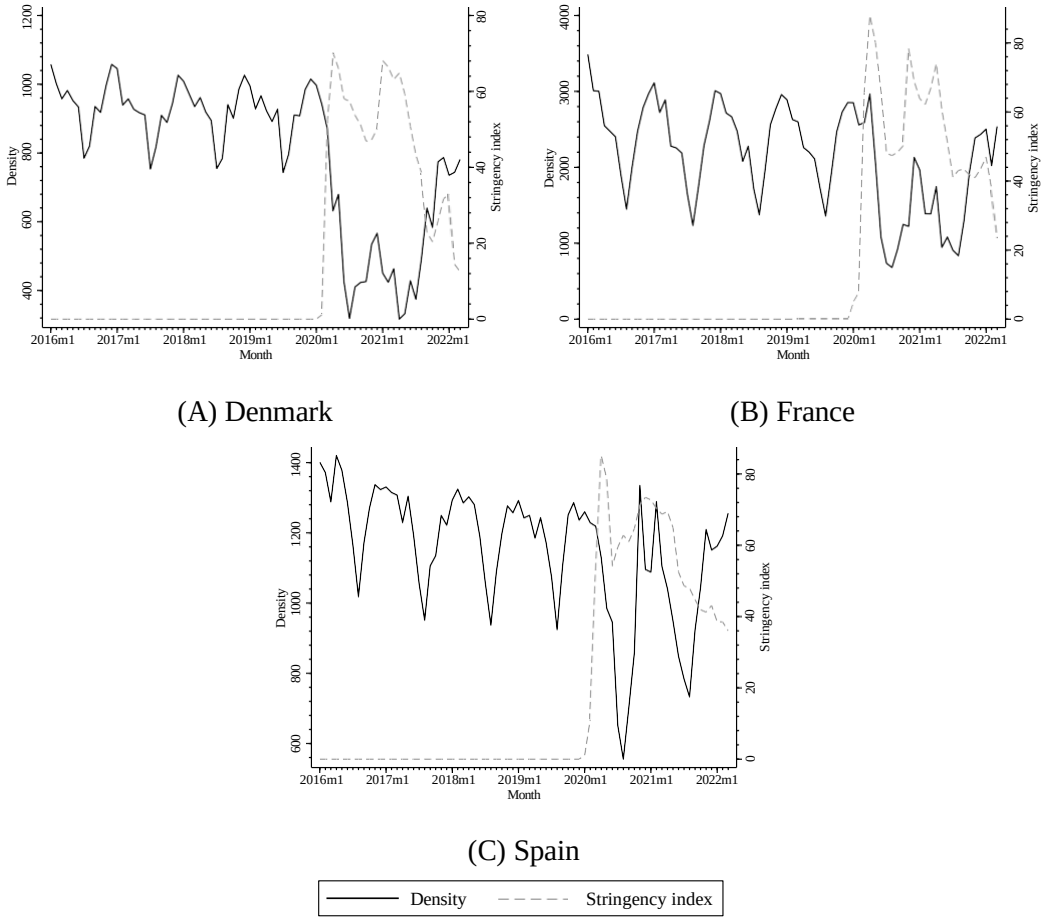
(A) Denmark

(B) France

(C) Spain

Notes: This figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A shows the attraction density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

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

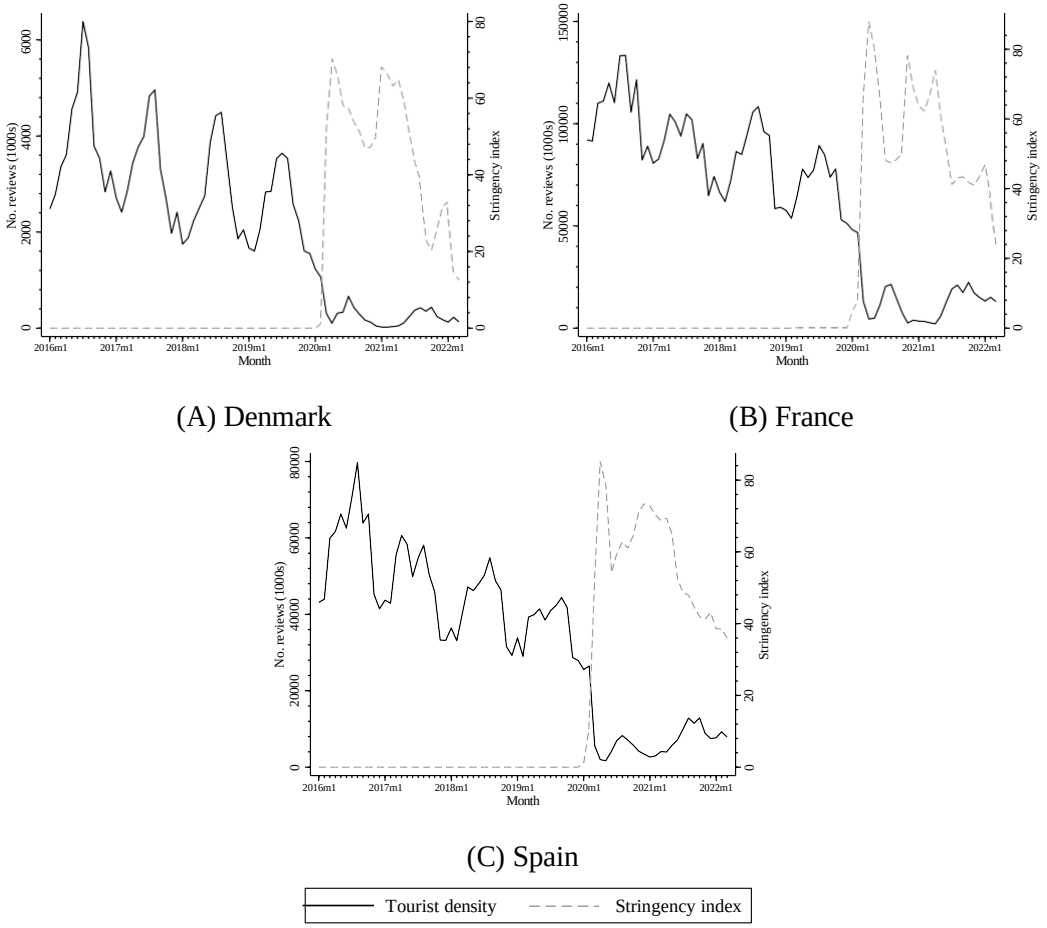


(A) Denmark (B) France

(C) Spain

Notes: This figure shows the evolution of the attraction density of visited locations together with the stringency index. Panel A shows the attraction density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

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

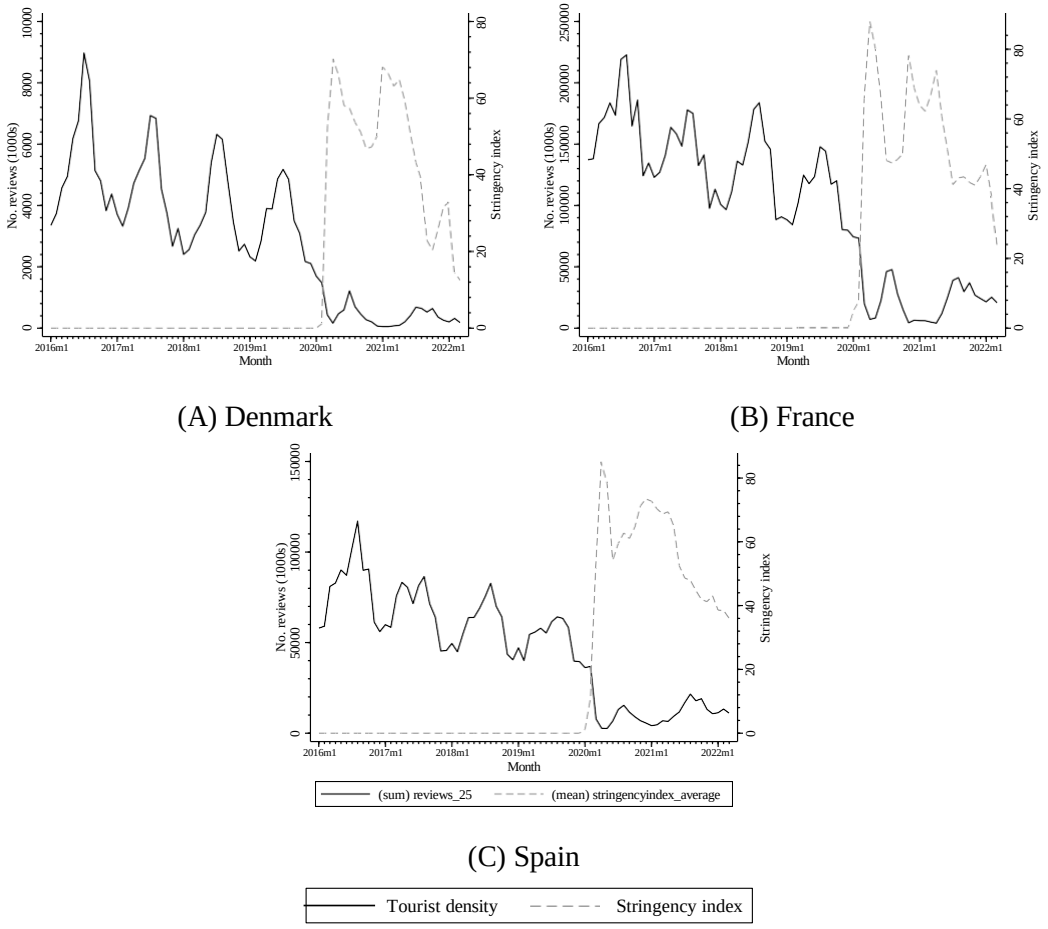


(A) Denmark (B) France

(C) Spain

Notes: This figure shows the change in the tourist density of visited locations together with the stringency index. Panel A shows the review density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

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 review density for Danish attractions, Panel B for French attractions and Panel C 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 2 for details) and the stringency index from the Oxford Government Response Tracker (Hale et al., 2021).

Figure A8: Travel patterns - Copenhagen

(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the NUTS3 region of Copenhagen. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A9: Travel patterns - Funen

(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the NUTS3 region of Funen. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A10: Travel patterns - Paris

(A) World

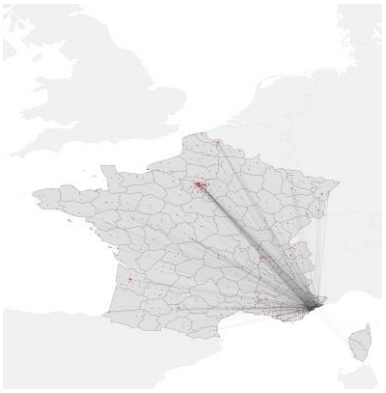
(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the NUTS3 region of Paris. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

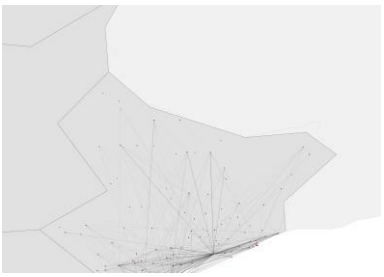
Figure A11: Travel patterns - Nice

(A) World

(B) Europe



(C) Domestic



(D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the NUTS3 region in which Nice is located. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A12: Travel patterns - Valencia

(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the city of Valencia. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region of Valencia in which the city of Valencia is located. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A13: Travel patterns - Asturias

(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the NUTS3 region of Asturias. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region. *Source*: Own data collected from Tripadvisor (see Section 2 for details).

Figure A14: Travel patterns - Denia

(A) World

(B) Europe (C) Domestic (D) Local

Notes: This figure shows travel patterns of all reviews written by tourists visiting attractions in the city of Denia. 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 considers domestic tourists. Panel D shows travel patterns at the very local level, i.e. within NUTS3 region of Alicante in which Denia is located. *Source*: Own data collected from Tripadvisor (see Section 2 for details).