Marshall Plan Research Paper

Analysis of crowd-sourced geodata to study park visitation patterns

Author

Innocensia Owuor

Geomatics Sciences, University of Florida, Fort Lauderdale Research & Education Center

Supervisors

Prof. Dr. Hartwig H. Hochmair

Geomatics Sciences, University of Florida, Fort Lauderdale Research & Education Center

Prof. Dr. Gernot Paulus

School of Spatial Information Management, Carinthia University of Applied Sciences

October 2022

Contents

Acknowledgements

I am grateful to my PhD advisor at the University of Florida, Dr. Hartwig Hochmair for conceptualizing this project and excellent tireless guidance during the project proposal, execution and final report writing.

I wish to acknowledge the efforts of Dr. Gernot Paulus who was my second advisor in this project for bringing new perspectives to the research. His great hospitality, in addition to the prompt support by Alexandra Maurer from the international office at the Carinthia University of Applied Sciences made my stay at the host university a pleasure.

This project completion was only possible through the financial support of the Austrian Marshall Plan Foundation through the Marshall Plan scholarship program and enabled me to build collaborative research networks with my counterparts in Austria.

Abstract

This study analyzes user contribution data to TripAdvisor, Yelp, Google Maps, Twitter, Flickr, and Wikipedia for various attractions in Florida (U.S.) and Carinthia (Austria) with respect to the two followings tasks.

1) comparison between counts of user contributions (reviews, photos, or messages) to selected online platforms for specific locations, and visitation counts of these locations from reference data to determine which online platform data provide the best match with reference visitation numbers; 2) assess the change of reviewer and visitor home regions between before and during the COVID-19 pandemic using review data, and more specifically to determine a change in distance from home to location, and also the localness of visitors and reviewers of the examined locations. Techniques employed in the study include correlation analysis, multiple regressions, time series analysis and origin-destination travel comparison. With regards to task 1, Wikipedia pageviews and reviews from TripAdvisor and Google Maps had the highest significant positive correlations with the reference visitation counts for both regions and thus best reflect state park and sight visitation numbers. Regression and time series analysis revealed that Google Maps reviews had the strongest explanatory power to the variations detected in the reference visitation counts and they feature most similar seasonal patterns to those from reference visitation datasets in both study areas. With regards to task 2, analysis of home origins of reviewers derived for TripAdvisor and Yelp, and of likely visitors (based on photos shared on Flickr) showed a close match with the reference dataset, i.e., SafeGraph home region data, as well as a significant increase in the proportion of local users during the pandemic in 2020 compared to the pre-pandemic era in 2019. The latter finding underscores the effect of COVID-19 related travel restrictions on travel mobility.

1.Introduction

Ubiquity of social media platforms for communication has resulted in a surge in volume, and variety and rapid spread of shared information. Researchers have used this big data to analyze tourist experiences, travel behavior, destination marketing strategies and the tourism sector's effect on local culture (Zeng & Gerritsen, 2014).

As the tourism industry is a significant contributor to employment and the gross domestic product in many countries around the world (World Travel & Tourism Council, 2022), information about tourism visitation patterns is vital for planning and development of sustainable tourism management.

In protected areas, such as parks which offer valuable ecosystem services such biodiversity, water, and air purification (Mexia et al., 2018), information about visitation patterns is crucial as overcrowding can have adverse effects on these ecosystems, such as pollution, soil erosion and land degradation (Milman, 2019). During the global health crisis caused by the COVID-19 pandemic in 2020, research found parks to be positive contributors to the maintenance of physical and mental well-being (Cheng et al., 2021). Therefore, understanding visitation trends in such areas highlights the value of these areas and the need for continuous investment in their management and their incorporation in urban planning.

Conventional data collection methods such as surveys, ticket sales, visitor counts, and traffic monitoring stations are labor intensive and not spatially and temporally exhaustive (Hale, 2018). In addition, some parks lack specific entry points or may also be in remote locations which makes conducting visitor counts using these traditional methods challenging (Ziesler & Pettebone, 2018).

Previous tourism research has found social media data to contain a wealth of information for both quantitative and qualitative studies. Geotagged multimedia content such as text on Twitter, or photos on Flickr and Instagram has been used to get insights into the dynamics of park tourism such as traffic flow to parks, popular visitation times, prevalence of various recreational activities and satisfaction levels of park patrons (Hausmann et al., 2018). These findings are valuable for monitoring crowds and consequently helps in the sustainable management of parks (Tenkanen et al., 2017). Social media data has been demonstrated to be cost effective and be useful as a proxy for visitation patterns in tourist facilities (Ma & Kirilenko, 2021). However, further research is needed to find the platform that offers the best match to tourism patterns as the applications are known to vary in popularity and data availability (Morstatter et al., 2013; Owuor & Hochmair, 2020).

This research demonstrates how big data from social media platforms, online encyclopedias, and non geo-social datasets such as SafeGraph can be jointly used to study visitation patterns in various parks and attractions in Florida and Carinthia between 2019 and 2021. It has been found that the combination of diverse sources can provide an improved match to official visitation counts compared to a single data source when forecasting tourism patterns (Li, et al., 2021). Additionally, the popularity of social media varies between regions, for example, based on the fact that certain social media applications being banned in some countries (Hou et al., 2018), which is the reason that this research uses two study regions (Carinthia, Florida).

Digital online data sources, especially some social media platforms, offer near real time data which is critical when analyzing the effects of global events. Examples, are the COVID-19 pandemic which led to worldwide travel restrictions (Jiang et al., 2021), or Brexit which led to uncertainties among travelers to the UK (Dutta et al., 2021). Parks were one of the few spaces that the public could use while still following COVID-19 mitigation protocols such as social distancing when many indoor recreational facilities such as gyms were closed. Various studies identified increased visitation numbers to parks during the COVID-19 pandemic as people used these venues for physical exercise and the maintenance of mental health care (Volenec et al., 2021). Florida and Carinthia are both premier tourist destinations in the U.S. and Austria, respectively, with Florida welcoming 121 million visitors in 2021 (VISIT FLORIDA, 2022) and Carinthia recording 13 million overnight stays annually (Landesregierung, 2019). This research determines the usability of various big data sources for the extraction of certain travel behavioral patterns in 46 Florida state parks in Florida and three tourist attractions in Carinthia. The locations of the Florida state parks are mapped in Figure 1.

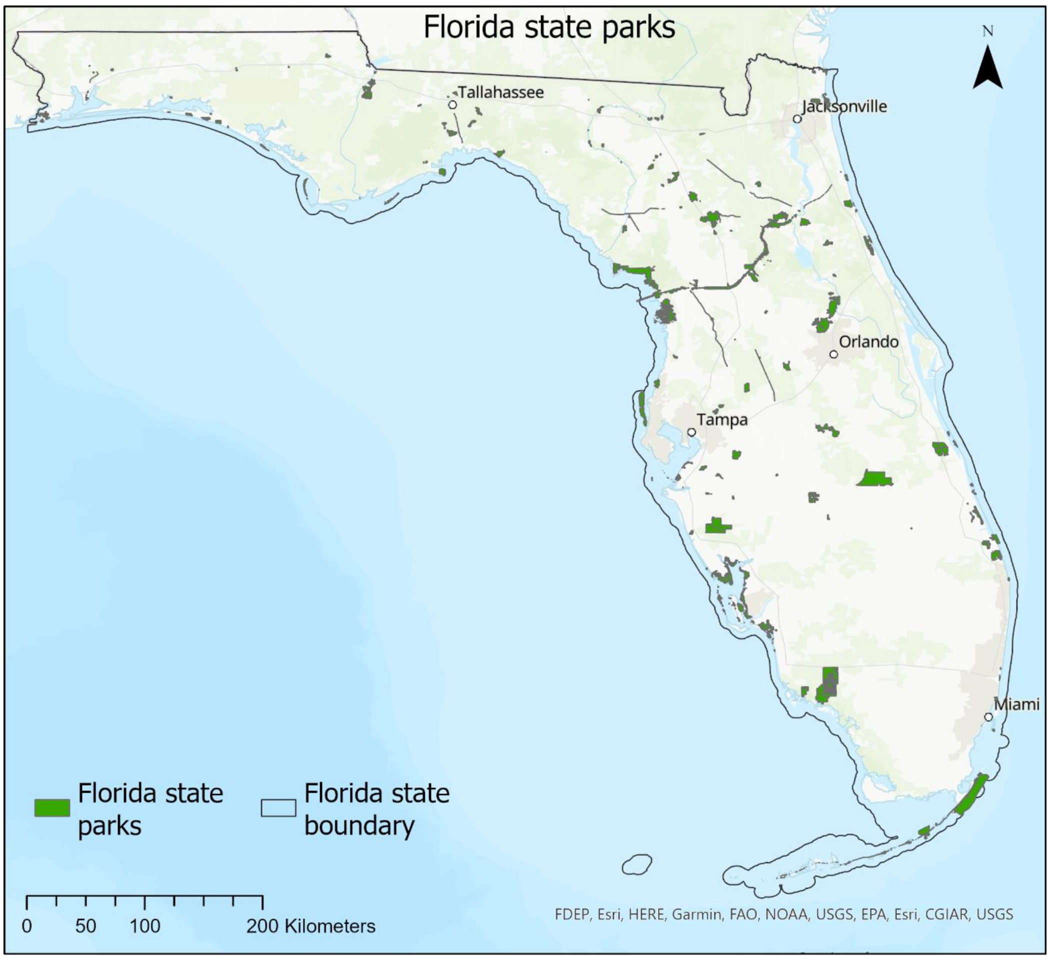

Figure 1. Florida state parks.

Based on the availability of official visitation counts, the following tourist destinations instead of parks (Figure 2) were analyzed in Carinthia: Grossglockner High Alpine Road (Grossglockner Hochalpenstrasse), Pyramidenkogel, and Villach Alpine Road (Villacher Alpenstrasse). In Florida, the study was conducted for data collected between January 2019 and December 2021 whereas in Carinthia the study dates included May through October for year 2019, 2020 and 2021, given that mountain roads are closed during winter season. The following data are used for the different analyses in this study: Reviews from TripAdvisor, Yelp and Google Maps; spatio-temporal contribution patterns from tweets, Flickr photos, Wikipedia page edits, and Wikipedia page views; SafeGraph POI visitation patterns. Yelp, Twitter and SafeGraph datasets were only used for Florida where they were available.

Figure 2. (a) Grossglockner High Alpine Road, (b) Pyramidenkogel and (c) Villach Alpine Road.

The main objective of this research was to assess the quality of data from different social media and other platforms for modeling and replicating various aspects of visitation patterns to Florida and Carinthia sights. This is achieved through the following specific tasks:

1. Correlate monthly contribution counts for included parks and attractions from the different platforms with the official visitation statistics.
2. Conduct multiple regression analysis with monthly counts from different platforms as predictors of published visitation numbers.
3. Compare temporal patterns exhibited in the different platforms to those reflected in official visitation numbers by analyzing their respective seasonal patterns.
4. Determine the data quality from TripAdvisor, Yelp, and Flickr through comparison of the origin (home regions) of visitors to parks/attractions inferred from each platform to the

origin derived from SafeGraph reference data and assess the change in home regions during the pandemic compared to pre-pandemic years.

2. Literature review

Big data sources from travel based social media applications (TripAdvisor, Yelp, Google Maps) and online data sources such as Wikipedia pages have become the go-to resources for tourists when making decisions about travel. Potential tourists rely on the travel experiences of others who document their experiences on these platforms. TripAdvisor (TripAdvisor, 2022) and Yelp (Statista, 2021) are popular online review platforms for travel planning, boasting over one billion reviews globally and a quarter billion reviews respectively while Google Maps is used by a billion people (Rusell, 2019). Current studies that employ big data in tourism related research have leveraged it in tourism demand forecasting (Peng et al., 2021), understanding tourist experiences through sentiment analysis (Alaei et al., 2019) and assessing tourists' mobility by analyzing spatial and temporal attributes (Yan et al., 2020). The nexus between the increase in activity on social media applications and tourism patterns have been uncovered by numerous studies. For example, Twitter data was used to derive traffic flows to certain attractions (Hu et al., 2019) and Flickr/ Panoramio images were used to identify popular attractions (Majid et al., 2013). A research study found that edits to Wikipedia pages led to an increase in the number of tourists in the UK (Hinnosaar et al., 2017) and Wikipedia page views were used to predict tourist flows to various attractions (Signorelli et al., 2016). TripAdvisor reviews can infer tourism visitation patterns when cross-validated with cellphone data and official visitation statistics at various attractions in Florida (Ma & Kirilenko, 2021).

A paper which reviewed social media platforms prominently used to monitor outdoor recreational areas found Flickr to be the most popular source with 36 papers, followed by Panoramio which shut down in 2019 with 10 papers, Instagram (6 papers), and Twitter and TripAdvisor with three and one paper (Teles da Mota & Pickering, 2020). Twitter and Flickr geodata were also found to be useful indicators for park visitation and equitable park access in a New York City study (Hamstead et al., 2018). A similar study used Twitter, Flickr, and Instagram to model visitation patterns to parks and natural areas in Finland and South Africa, indicating that Instagram data best matched the official visitation statistics (Tenkanen et al., 2017). This research serves as an update to this preceding study since Instagram restricted access to their data in 2018 (Schroepfer, 2018) and Flickr introduced a paid membership account called Flickr Pro and limited the photo storage of users with free accounts (Stadlen, 2018).

Tourism flows in outdoor attractions are heavily shaped by weather patterns. There are vast differences in the climatic conditions between the two study areas of Florida and Carinthia under investigation. For instance, areas in the Alpine climate such as Austria have been found to have a surge in tourism during summer season while those in tropical humid climate similar to Florida experience an uptick in tourism demand during winter and spring seasons (Falk, 2014). Studies have found weather seasons to have an effect on a tourist choice of destination (Buhalis & Foerste, 2015) and among different age groups, older seniors are significantly more likely to factor in weather when travel planning (Pestana et al., 2020).

Prior research has used time series analysis of the contribution patterns of Flickr images to reveal tourist seasonality in the Alpine regions (Schirpke et al., 2018). Google Trends data has been used to analyze the seasonal patterns of visits to various museums in London where derived weekly and monthly seasonal patterns were used to forecast demand (Volchek et al., 2019). Tourism forecasting studies typically used monthly seasons and less often daily and weekly frequencies (Wu et al., 2017).

One aspect of tourist movement patterns is their origins and destinations, which is important to the industry players for drawing destination marketing strategies, sustainable management of tourist attractions, and for analyzing the impact of events on travel (Cirer-Costa, 2017). This can be obtained through mobile network data (Alexander et al., 2015) or travel surveys. The latter data is limited in terms of spatial and temporal resolution whereas mobile phone data is expensive (Ma & Kirilenko, 2021) and access is heavily regulated due to privacy concerns (Kubo et al., 2020). Therefore, social media data is a potential alternative data source which allows to infer a traveler's origin either from contribution data or user profiles (Yang et al., 2022). Hometowns and country of origin of visitors to various Italian cities have been inferred from Flickr to identify attractions that are particularly popular with foreign tourists (Giglio et al., 2019). A comparison of Yelp reviews of Japanese restaurants by Westerners (North American and European origins) and Japanese found that cultural differences in the respective regions influenced the sentiments portrayed in the reviews. That is, Westerners prioritized quality of service whereas for the Japanese customer satisfaction is largely impacted by menu pricing (Nakayama & Wan, 2019).

An assessment of origin-destination studies that leverage social media data, such as Twitter, indicated that the locations data extracted from the platform may not be an accurate presentation of a user's origins (home) location (Hecht et al., 2011). Therefore, these locations should be validated, through methods such as analyzing a Twitter user posting history and network (Zheng et al., 2018). Research pertaining data from Twitter, Flickr, Swarm found the assumption that local users can be identified from the geotagged content they posted in an area is only valid in 75% cases (Johnson et al., 2016). With respect to Flickr and Wikipedia data, a study found that only 53% and 23% of their users respectively posted content that was within 100 km of their self-defined home location (Hecht & Gergle, 2010).

3.Methodology

3.1 Data sources

The data sources and analysis methods (numbered 1 to 4) applied in this research are summarized in Table 1. A combination of geo-social data from social media apps (Twitter, Flickr, TripAdvisor, Yelp, and Google Maps) and non-geo-social sources (Wikipedia page views, Wikipedia page edits, SafeGraph patterns and official visitation counts) were analyzed as shown in Table 1 for the corresponding study area.

The spatial granularity of the geo-social data was inconsistent and occurred at various administrative levels such as city, state, province, and country in the two study areas. For the origin-destination analysis (method 4 in Table 1), for the Florida state park visitation analysis, user contributions from out-of-state were aggregated to U.S. state level centroids, while for Carinthia, home regions from visitors and reviewers were aggregated to country centroids.

The reference data (official visitation counts) had a monthly temporal resolution. Therefore, for matching count data from geo-social and other platforms (Wikipedia page views, Wikipedia page edits and SafeGraph patterns data with reference data) for methods 1, 2, and 3 in Table 1, daily counts data from these platforms were aggregated to monthly activity counts.

Table 1. Research methods with the respect data sources

Analysis Task	Analysis method	Florida	Study period	Carinthia	Study period
1	**1.Correlation analysis**	Google Maps TripAdvisor Flickr Twitter Wikipedia page views Wikipedia page edits Official visitation counts	Jan 2019 to Dec 2021	Google Maps TripAdvisor Flickr Wikipedia page views Wikipedia page edits Official visitation counts	May to Oct of 2019, 2020 and 2021
1	**2.Multiple regression**	Google Maps Wikipedia page views Official visitation counts	Jan 2019 to Dec 2021	Google Maps Wikipedia page views Official visitation counts	May to Oct of 2019, 2020 and 2021
1	**3.Time series analysis**	Google Maps Official visitation counts	Jan 2019 to Dec 2021	Google Maps Official visitation counts	May to Oct of 2019, 2020 and 2021
2	**4.Origin-destination analysis**	TripAdvisor Yelp SafeGraph	Mar to Dec of 2019 and 2020 respectively March 1 to April 30, 2019 (SafeGraph)	TripAdvisor Flickr	March to Dec of 2019 and 2020

3.1.1. Geo-social sources

a) TripAdvisor

This website allows its registered users to share reviews about their experiences in restaurants, hotels, and attractions such as public parks. The review data contains attributes such as reviewer's name and self-defined location, review text, rating of the experience, date of the review and occasionally photos of the attraction. TripAdvisor has an Application Programming Interface (API) but only allows the retrieval of five reviews per attraction, therefore, to get sufficient data sample for research, a python script was used to obtain all the reviews in the languages available. The crawler stored the reviews in comma-separated values (csv) files for the respective park/attraction. A total of 4857 reviews were recovered for Florida state parks whereas for the three attractions in Carinthia, a total of 465 reviews were found respectively for the periods specified in Table 1 .

b) Google Maps

Google Maps API allows for the retrieval of only five reviews per place except for owners of a location. The reviews associated with Florida state parks and Carinthia attractions were obtained from a third-party provider outscraper.com/, resulting in 69442 reviews for Florida parks and 9738 reviews Carinthia attractions which were downloaded in csv file format.

c) Twitter

The full archive search endpoint on the Twitter API enabled the retrieval of tweets that were sent out within the boundaries of the Florida state parks and Carinthia attractions. Only tweets with exact coordinates were considered in the study. Twitter data was only used in the study with respect to Florida parks. For attractions in Carinthia, only tweets posted within Grossglockner High Alpine Road were found and therefore Twitter was excluded from analysis related to this study area. The data was retrieved in JavaScript Object Notation (JSON) format and stored into a PostgreSQL database. A total of 1140 tweets were used in the analysis.

d) Flickr

The Flickr API was used to retrieve Flickr photos with geotags that were taken within the boundaries of Florida state parks and Carinthia attractions. This resulted in 6934 photos in Florida state parks and 863 photos for Carinthia attractions. The data which was in JSON format was converted to a tabular format and stored in the PostgreSQL database.

e) Yelp

This app contains attributes such as reviewer's name and self-defined location, review text, rating of the experience, date of the review and occasionally photos of the attraction. The Yelp fusion API limits search results to three reviews per business. Therefore, a python script was developed to scrape the 520 reviews which were used in this study.

3.1.2. Non geo-social sources

a) Wikipedia

Wikipedia page views and edit counts for Florida state parks and Carinthia attractions were retrieved in JSON format through the Wikimedia REST API (https://wikimedia.org/api/rest_v1/#/). For Florida parks, the en.wikipedia.org domain was used while for Carinthia attractions, the de.wikipedia.org domain was used. The JSON data was converted to a tabular format compatible with PostgreSQL database where it was stored.

b) SafeGraph

SafeGraph is a company that, based on data from mobile phone carriers and social media companies provides three datasets: (1) Core places which contains over six million points on interest (POI) location and attribute data (address, name); (2) Geometry which give the geometric representation of these POIs, for instance, the building outline that hosts a coffee shop; (3) Patterns which provide information about visitation patterns of POIs. The data is updated once a month and is available for free to academic researchers. SafeGraph Places has over eleven million POIs in 200 countries, including foot traffic data to POIs in the U.S., Canada, and the UK. Data for the attractions in Carinthia are not available from this dataset.

c) Reference data

Reference data containing official visitor statistics needed during the analysis method 1, 2, and 3 in Table 1 for validating the quality of visitation patterns derived from the geo-social and non geo-social data in Table 1 were obtained from the Florida Department of Environmental Protection (FDEP) and from the respective visitor management offices of the Carinthia attractions. FDEP provided us with monthly visitation counts of Florida state parks for the period between January 2019 and December 2021in tabular format.

Visitation counts for Pyramidenkogel contained actual counts of visitors while for the other two road attractions, traffic count data containing number of cars, buses, trucks, and motorcycles were provided including conversion factors that were used to convert vehicle numbers to visitor counts. The reference datasets for Carinthia were obtained for the period between May and October for years 2019 through 2021.

A summary of contribution counts from each source is shown in Table 2 for both study areas.

Table 2 Total counts of contributions for different data sources

Data source	Florida state parks	Carinthia tourist sites
TripAdvisor	4,857	465
Google maps	69,442	9,738
Twitter	1,140	Not applicable
Flickr	6,934	863
Yelp	520	Not applicable
Wikipedia page views	1,167,691	90,974
Wikipedia page edits	368	49
SafeGraph	10,398	Not applicable
Official visitation counts	39,975,932	3,179,117

3.2. Data analysis

3.2.1. Pearson correlation coefficient

The linear relationship between the monthly official visitation statistics for Florida parks and Carinthia attractions versus the monthly counts from each respective data source (TripAdvisor, Google Maps, Wikipedia page views, Wikipedia page edits, Flickr, and Twitter) was investigated using Pearson Product Moment Correlation (Sedgwick, 2012). As Pearson's correlation is a parametric test (Sedgwick, 2012), the variables were assessed for normality using Shapiro’s Walk test (Coppack, 1990) and monthly aggregated counts of the datasets for the two study areas were found to be normally distributed as they had p values greater than the 0.05 threshold. Data sources that were not highly correlated were used as candidates in the regression models as the independent variables.

3.2.2. Multiple regression

The multiple regression model was used to determine if the combination of data from the numerous sources provides a more accurate forecast of visitation patterns for parks and attractions than a single data source. The variance inflation factor (VIF) which tests for multicollinearity was checked in this step, where those that were highly correlated with VIF values of over ten were excluded from the analysis (O’brien, 2007). Official visitor counts were used as the dependent variable. For Florida parks, three regression models were used for each respective year with 12 observations, while for Carinthia, the data counts for the study period in 2019, 2020 and 2021 were combined in the regression analysis due to the availability of only six-monthly observations each year.

3.2.3. Time series analysis

To expose the underlying patterns inherent in the time series data of the official visitation counts and the big data sources, Google Maps was singled out by the regression model for having the highest explanatory power to the variation in official visitation counts in the two study areas. Its temporal patterns were therefore compared to the reference data. The analysis involved splitting their respective original time series into three components, i.e., trend, seasonality, and residuals. The additive decomposition method was used as the original time series for the datasets under study were discovered to have a seasonal component with a constant magnitude. Time series decomposition is valuable as it splits the data into the predictable (trend and seasonality) and random (residuals) parts which is useful when forecasting. The trend component captures the average increase and decline pattern of a time series typically using values calculated using moving averages based on the seasonal pattern in the datasets. The seasonal component shows regular variations that occur after specific time intervals. The residuals are the random components which remain after removal of trend and seasonal parts. The original time series $\mathrm{Y_t}$ is constructed using components in the additive (Equation 1) decomposition approach, where $\mathrm{S_t}$, $\mathrm{T_t}$ and $\mathrm{R_t}$ are the seasonal, trend and residual components of the original time series $\mathrm{Y_t}$ at time t:

$$\mathrm{Y_t} = \mathrm{S_t} + \mathrm{T_t} + \mathrm{R_t} \tag{1}$$

Time series analysis of the data (official visitation counts and Google Maps review numbers) between January 2019 and December 2021 for Florida and between May and October for the year 2019, 2020 and 2021 for Carinthia were analyzed and their main characteristics highlighted in charts. The time series data for the two areas were analyzed using Fourier transformations (Bloomfield, 2004) to extract their frequency which was used to quantify their seasonality using the periodogram function in R (R Core Team and contributors worldwide, 2013).

3.2.4. Origin-destination analysis

Home origins of visitors to parks in Florida and attractions in Carinthia, were inferred from the user-defined home locations from select social media applications (TripAdvisor, Flickr, and Yelp) and from the SafeGraph dataset. These locations were analyzed to investigate whether the travel restrictions introduced during the COVID-19 pandemic affected the origins of park visitors. For the Florida study, TripAdvisor, Yelp and SafeGraph were used, while for the Carinthia due to data availability, only TripAdvisor and Flickr were applied. The periods of before and during the pandemic in both study areas for TripAdvisor, Flickr and Yelp were defined as (March 20 to December 31, 2019) and (March 20 to December 31, 2020) respectively. For SafeGraph, due to data abundance, the before and during the pandemic were chosen to be shorter, namely as (March 1 to April 30, 2019) and (March 1 to April 30, 2020).

The spatial granularity of the origin information provided in the datasets varied in both study areas. Therefore, to retain a sufficient sample size, the distances were calculated between the reviewer/visitor origin state centroid coordinates and the Florida state centroid location in the Florida study. For Carinthia, the distances were between reviewer/visitor origin country centroid and the Austria country centroid.

The specific steps in the origin-destination analysis were as follows:

i. The spatial distribution of home origin of tourists' pre-pandemic (2019) vs. during pandemic (2020) were determined. For Florida, origins of visitors were mapped using U.S. state level centroids coordinates while for Carinthia they were mapped based on their respective country level centroid coordinates.

ii. Distances between the visitor origins and Florida state and Austria country centroid coordinates respectively for tourists pre-pandemic (2019) vs. during pandemic (2020) periods were measured.

iii. The statistical association between the proportion local and foreign park/attraction visitors, pre-pandemic (2019) vs. during pandemic (2020) periods for Florida and Austria were analyzed using Chi-square test.

4. Results

4.1. Correlation analysis between official visitor counts and different platforms

For Florida state parks, monthly counts of Wikipedia page views, Wikipedia page edits, reviews from Google Maps and TripAdvisor, Flickr photos, and tweets for the year 2019, 2020 and 2021 were correlated with official monthly visitation counts. All data sources had positive correlations apart from Twitter, Wikipedia page edits and Flickr which had non-significant negative correlations due to a small sample size and were therefore discarded from this part of the analysis. Table 3 shows correlation results between visitor counts and the remaining data sources, based on 12 data points for each year.

Overall, a high fluctuation in correlations can be observed between years. The only dataset that was significantly correlated with visitor counts in all three years was TripAdvisor. 2020 stands out as the only year with all three sources showing significant correlations with reference visitor counts. 2019 reveals the highest correlation for the combined dataset ($r = 0.94$) and two individual data sources, i.e., Wikipedia page views ($r = 0.86$) and TripAdvisor ($r = 0.82$). Years 2020 and 2021 show a smaller match, possibly due to perturbations in visitor patterns and review activities during and right after the pandemic.

Table 3. Pearson correlation between official visitor counts and study datasets for Florida state parks.

Data source	**2019**	**2020**	**2021**
Wikipedia page views	0.86***	0.58*	0.22
Google Maps	0.45	0.78**	0.67**
TripAdvisor	0.82***	0.61*	0.50*
Combined sources	0.94***	0.56*	0.34
p < .001 ‘***’, p < .01 ‘**’, p < .05 ‘*’			

The correlation plot consisting of a scatter plots and histograms, highlights the correlations between official visitor counts, Wikipedia page views, TripAdvisor, and Google Maps for 2019 (Figure 3). Wikipedia page views and TripAdvisor are significantly correlated at ($r = 0.93$, $p < 0.001$), which means that these two data sources cannot be jointly used in multivariate regression analysis. The first row reflects the first three values under the 2019 column in Table 3.

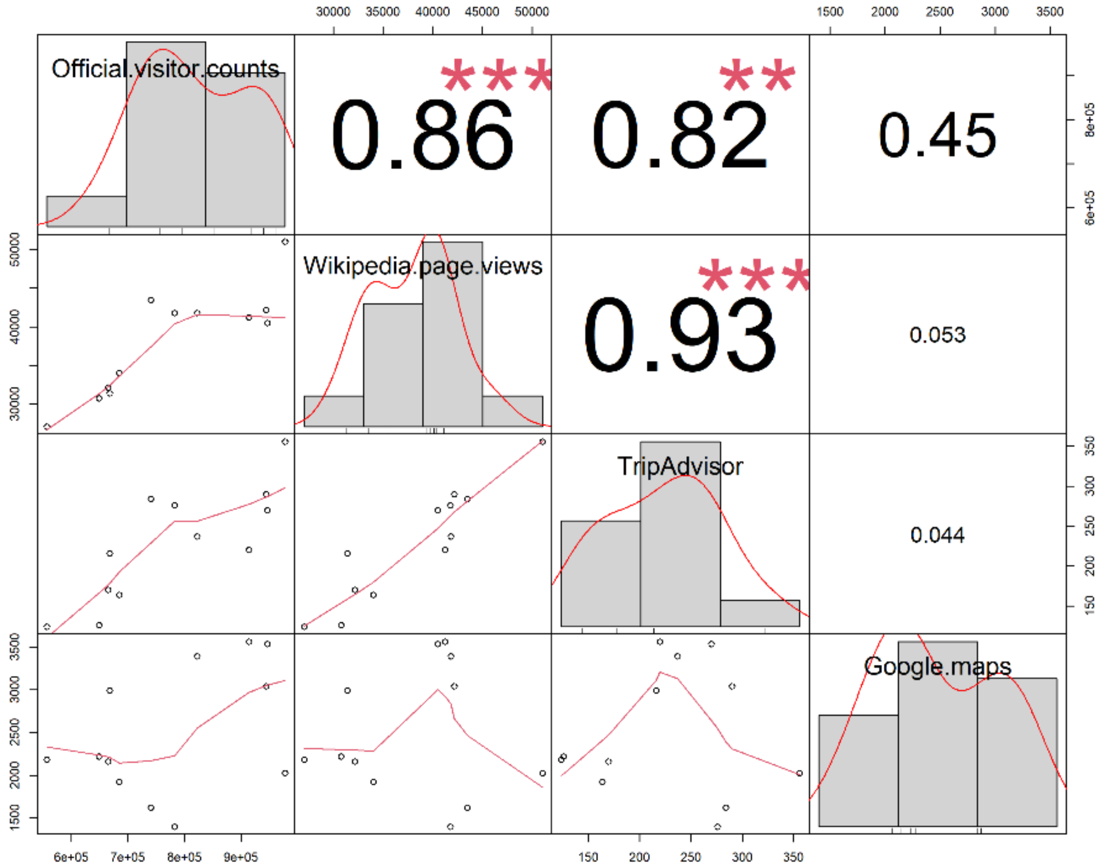

Figure 3. Correlation plot of official visitation counts for Florida parks and related data from different platforms (2019).

In the Carinthia study, data points for the three years were combined leading to 18 observations. Due to low sample size, data from Wikipedia page edits, and Flickr were excluded from this part of the analysis.

Correlation results for the remaining data sources are shown in Table 4, showing high correlations for all three datasets, and their combination.

Table 4. Pearson correlations between official visitor counts and study datasets (Carinthia)

Data source	**2019 -2021**
Wikipedia page views	0.89***
Google Maps	0.90***
TripAdvisor	0.72***
Combined sources	0.88*
p < .001 ‘***’, p < .01 ‘**’, p < .05 ‘*’	

Figure 4 shows the correlation plot between visitor counts and social media activity counts for the combined three-year data from 2019 through 2021. Significant correlations can be observed between Wikipedia page views, TripAdvisor, and Google Maps besides their correlation with visitor counts. The prior shows some redundancy when considering all three datasets.

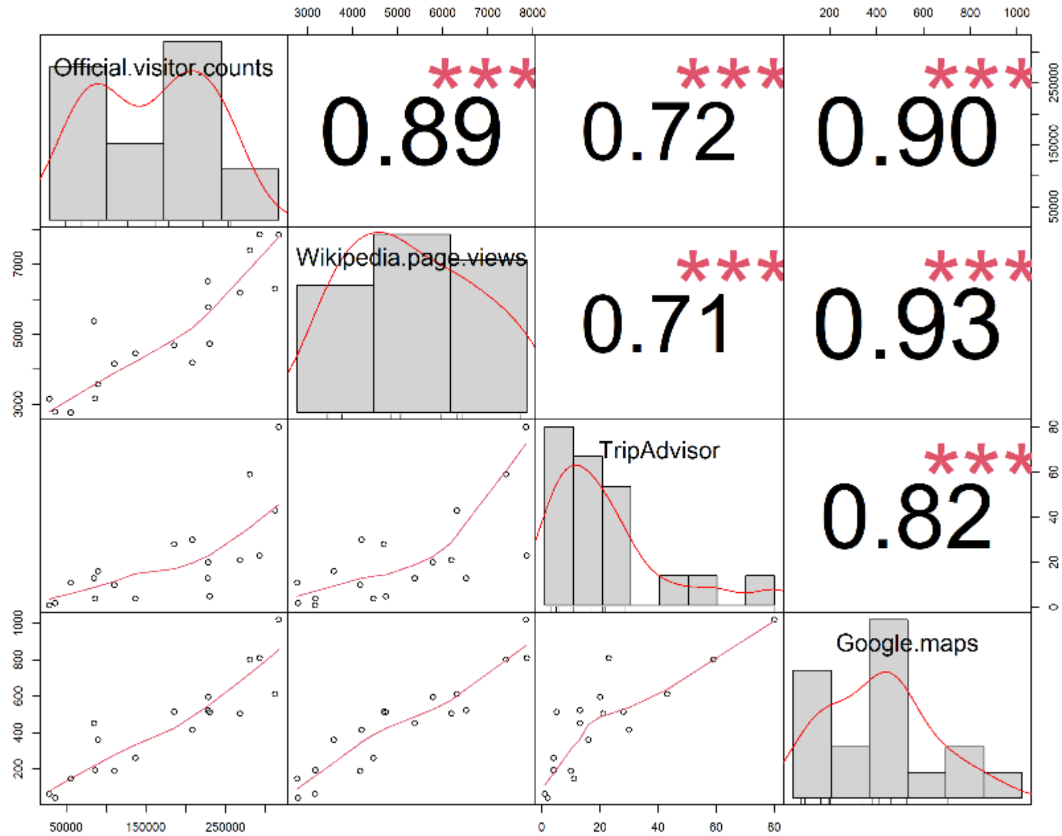

Figure 4. Correlation plot of official visitation counts for Carinthia attractions and related data from different platforms for 2019, 2020 and 2021.

4.2. Multiple regression analysis

When building best-fitting multiple regression models to predict monthly visitor counts, multicollinearity was observed when including TripAdvisor, Google Maps and Wikipedia page views, resulting in a VIF of above 10, which led to the exclusion of TripAdvisor for both regions.

For Florida, three regression models based on monthly count data were built for the years 2019, 2020 and 2021 using Wikipedia page views and Google Maps as candidate predictors. Results show that only in year 2019 a combination of Wikipedia page views and Google Maps reviews lead to the highest adjusted R^2. (Table 5), whereas for year 2020 (Table 6) and year 2021 (Table 7) models with use of Google Maps reviews as the only predictor result in the highest adjusted R^2.

Table 5. Regression model for Florida park visitation patterns (2019)

Independent variables	**Model 1**	**Model 2**	**Model 3**
Wikipedia page views (log)	1028711.26** (141970.42)		1009356.69*** (110536.98)
Google Maps (log)		220219.56 (200505.97)	181789.19 ** (66102.03)
N	12	12	12
R^2	0.84	0.11	0.91
Adjusted R^2	0.82	0.02	0.89
p< .001 '***', p < .01 '**', p < .05 '*' Standard errors in parentheses			

Table 6. Regression model for Florida park visitation patterns (2020)

Independent variables	**Model 1**	**Model 2**	**Model 3**
Wikipedia page views (log)	551712.72 (253276.31)		84828.18 (172570.16)
Google Maps (log)		299919.03 ** (47031.80)	2833538.52*** (59401.29)
N	12	12	12
R^2	0.32	0.80	0.81
Adjusted R^2	0.25	0.78	0.77
p < .001 '***', p < .01 '**', p < .05 '*' Standard errors in parentheses			

Table 7 Regression model for Florida park visitation patterns (2021)

Independent variables	**Model 1**	**Model 2**	**Model 3**
Wikipedia page views (log)	272117.47 (236901.45)		-76910.66 (214284.58)
Google Maps (log)		717956.80 ** (208371.45)	770066.97 * (261994.95)
N	12	12	12
R^2	0.12	0.54	0.55
Adjusted R^2	0.03	0.50	0.45
p < .01 '**', p < .05 '*' Standard errors in parentheses			

For Carinthia, the regression analysis was conducted using the combination of 18 observations across three years.

Three models were explored using Wikipedia page views and Google Maps reviews as candidate predictors and year as a categorical control variable (Table 8). The highest adjusted R^2 value was found for a model which includes both Wikipedia page views and Google Maps reviews as predictors.

Table 8 Regression model for Carinthia attractions visitation patterns (2019-2021)

Independent variables	**Model 1**	**Model 2**	**Model 3**
Wikipedia page views (log)	178051.06 ** (62897.40)	156875.27 ** (54390.57)	161349.02 * (53117.41)
Google Maps (log)	20048.04 (32344.50)	34643.14 (37710.98)	56997.01 ** (22350.02)
Year 2020		61636.47 * (25573.72)	67859.49 * (23744.19)
Year 2021		55209.49 (35024.38)	35697.47 (22778.26)
N	18	18	18
R^2	0.84	0.90	0.90
Adjusted R^2	0.80	0.83	0.86
p < .01 ‘**’, p < .05 ‘*’ Standard errors in parentheses			

4.3. Temporal patterns

The charts in Figure 5. Time series decomposition results for (a) official statistics and (b) Google Maps reviews for Florida parks between January 2019 and December 2021. The decomposed (i) original timed series was split into the following components (ii) trend, (iii) seasonality and (iv) residuals are shown for the respective data sources in Figure 5.

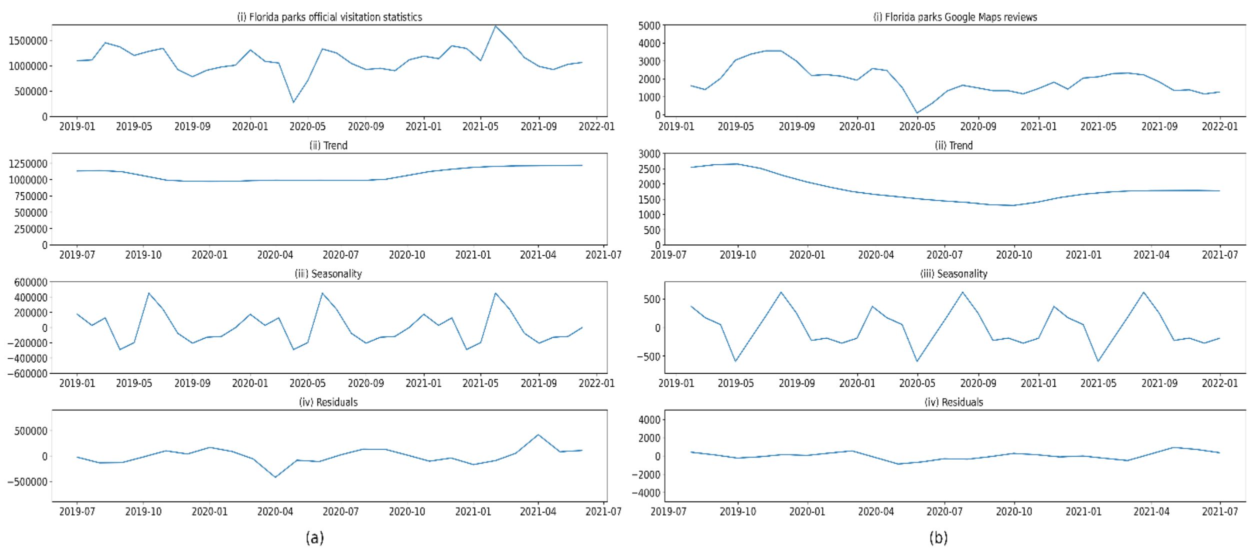

Figure 5. Time series decomposition results for (a) official statistics and (b) Google Maps reviews for Florida parks.

The results of the seasonality detection using the periodogram function in R, found six months to be a dominant frequency, indicating a semi-annual seasonality for both official visitation counts and Google Maps reviews. During the semi-annual season, the following peaks were detected from both data sources for each respective year between 2019 and 2021, winter (January), spring (March) and summer (June) which had highest visitation peak. In June, there was a 19%, 33% and 46% increment in visitation counts above the annual average for 2019, 2020 and 2021 respectively. The year 2020 and 2021 both had seasonal percentage increments that were higher than 2019.

A 72% dip in visits was recorded in official visitor counts in April 2020, indicating the effect of the COVID-19 lock down while a 74% decrease was recorded from the Google Maps data.

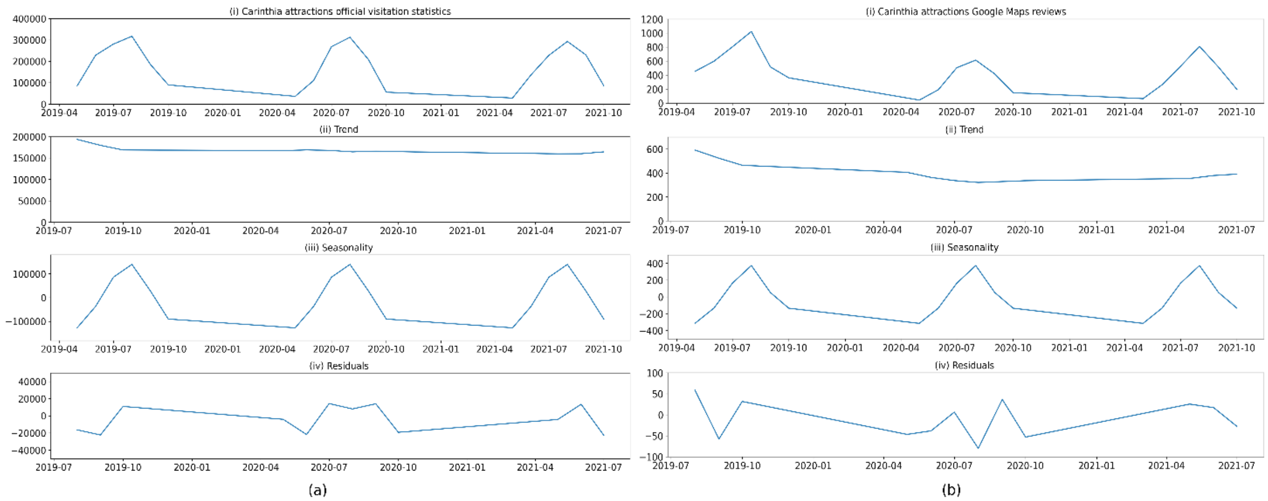

Figure 6. Time series decomposition results for (a) official statistics and (b) Google Maps reviews for Carinthia attractions.

For Carinthia attractions, temporal patterns from Google Maps (Figure 6b) were compared to those from the official visitation counts (Figure 6a) as well. A semi-annual seasonality was detected from both sources. There was one main high peak identified in the month of August where an increase of 60%, 89% and 75% above the seasonal average was detected for the year 2019, 2020, 2021 respectively from the analysis of the official visitation counts. Seasonal percentage increments for the year 2020 and 2021 were higher than 2019.

4.4. Analysis of changes of social media home locations

TripAdvisor reviewers from all states in the contiguous U.S. except for Idaho reviewed Florida parks in 2019 (Figure 7a) whereas in the year 2020 (Figure 7b), no TripAdvisor reviews were recorded in seven states. The mean distance between a reviewer's origin state centroid and the Florida state centroid reduced from 845 km before the pandemic in 2019 to 634 km during the pandemic in 2020.

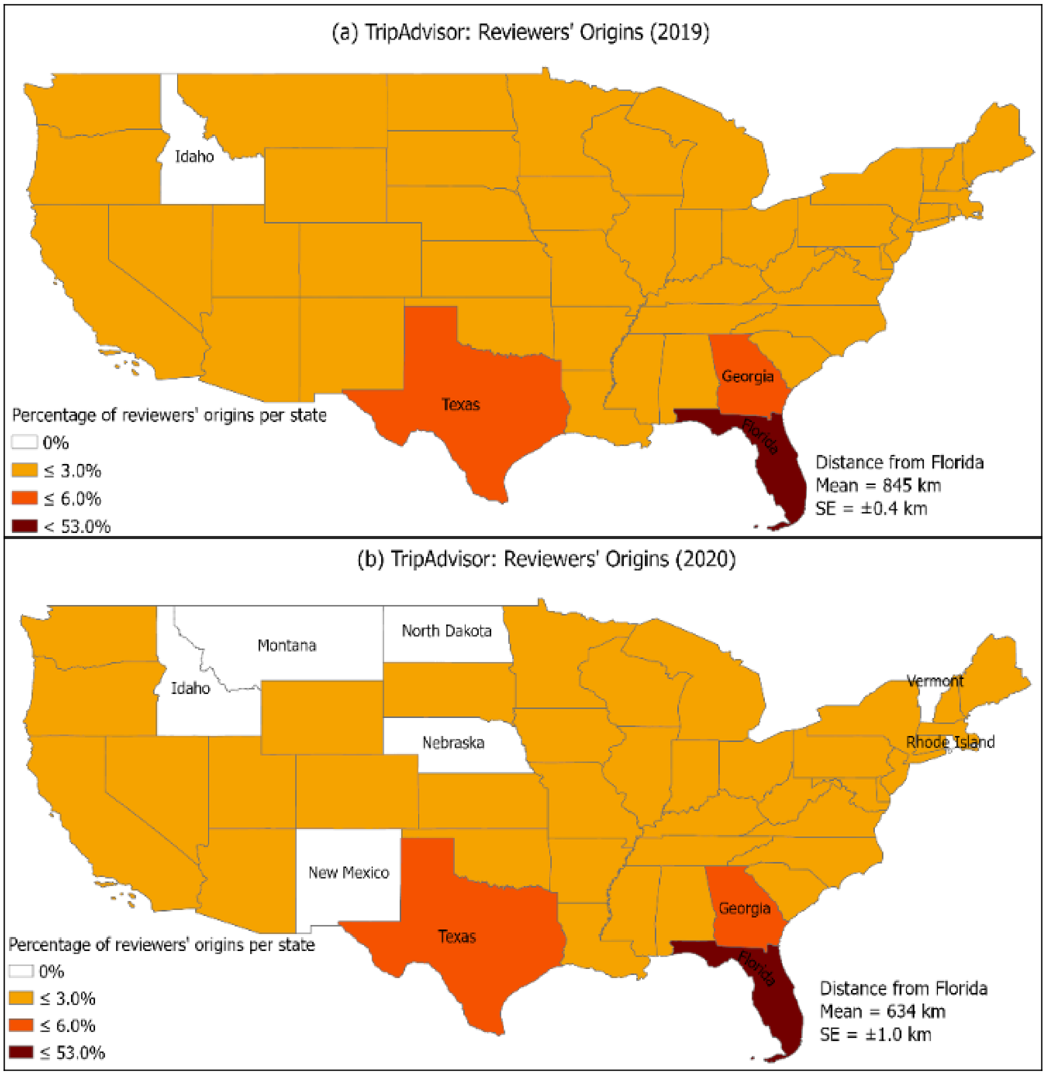

Figure 7. Origin of Florida state park reviewers on TripAdvisor in 2019 (a) and 2020 (b).

In Yelp, state park reviews were posted from users in 41 states in 2019 (Figure 8a) but from only 34 states in 2020 during the pandemic (Figure 8b). The mean distance between reviewer/visitor origin state centroid and Florida dropped from 753 km before the pandemic in 2019 to 522 km during the pandemic in 2020.

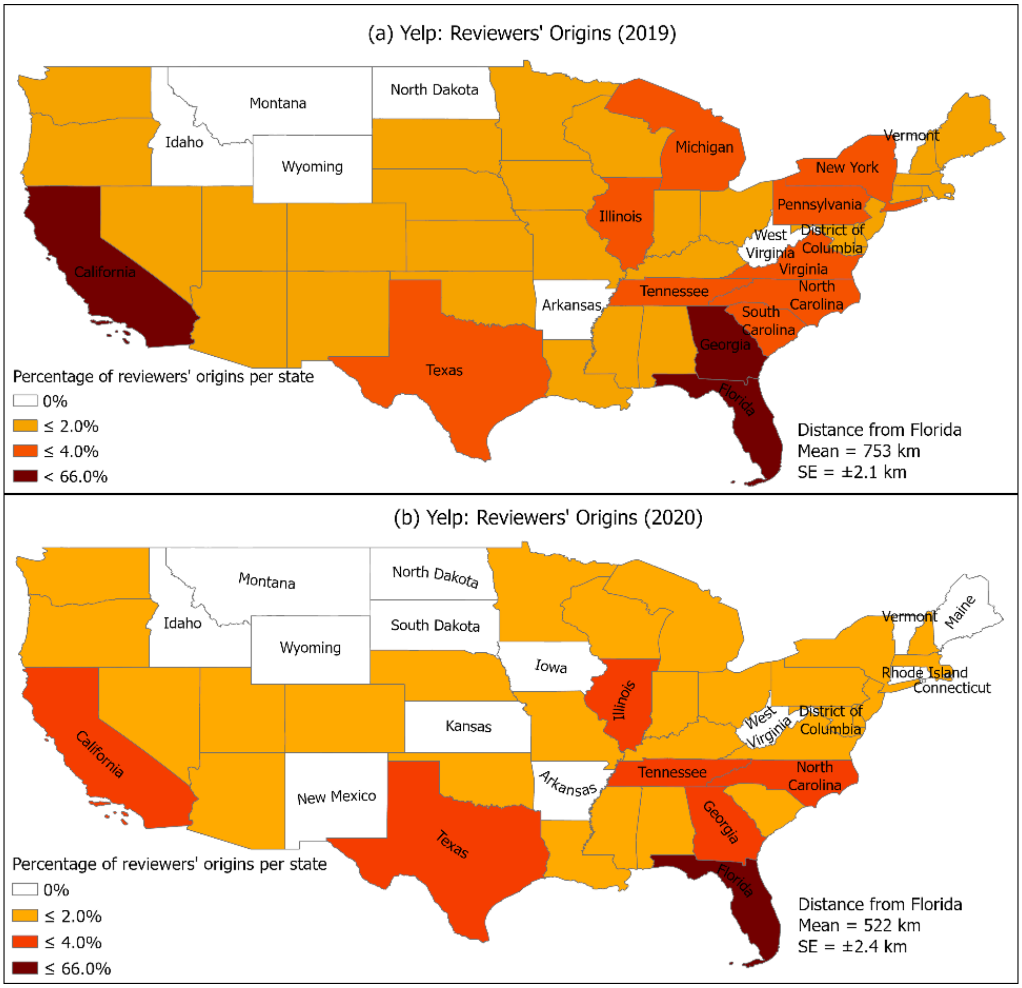

Figure 8. Origin of Florida state park reviewers on Yelp in 2019 (a) and 2020 (b).

Florida state park visitors from all 49 states in the contiguous U.S. were detected from the SafeGraph dataset both for 2019 (Figure 9a) and 2020 (Figure 9b). However, the portion of local (Florida) residents among all visitors of Florida state parks increased from 51.1% (pre-pandemic) to 59.4% (during pandemic), which was a statistically significant increase (Table 9). The mean distance between reviewer/visitor origin state centroid and Florida dropped from 673 km before the pandemic in 2019 to 568 km during the pandemic in 2020.

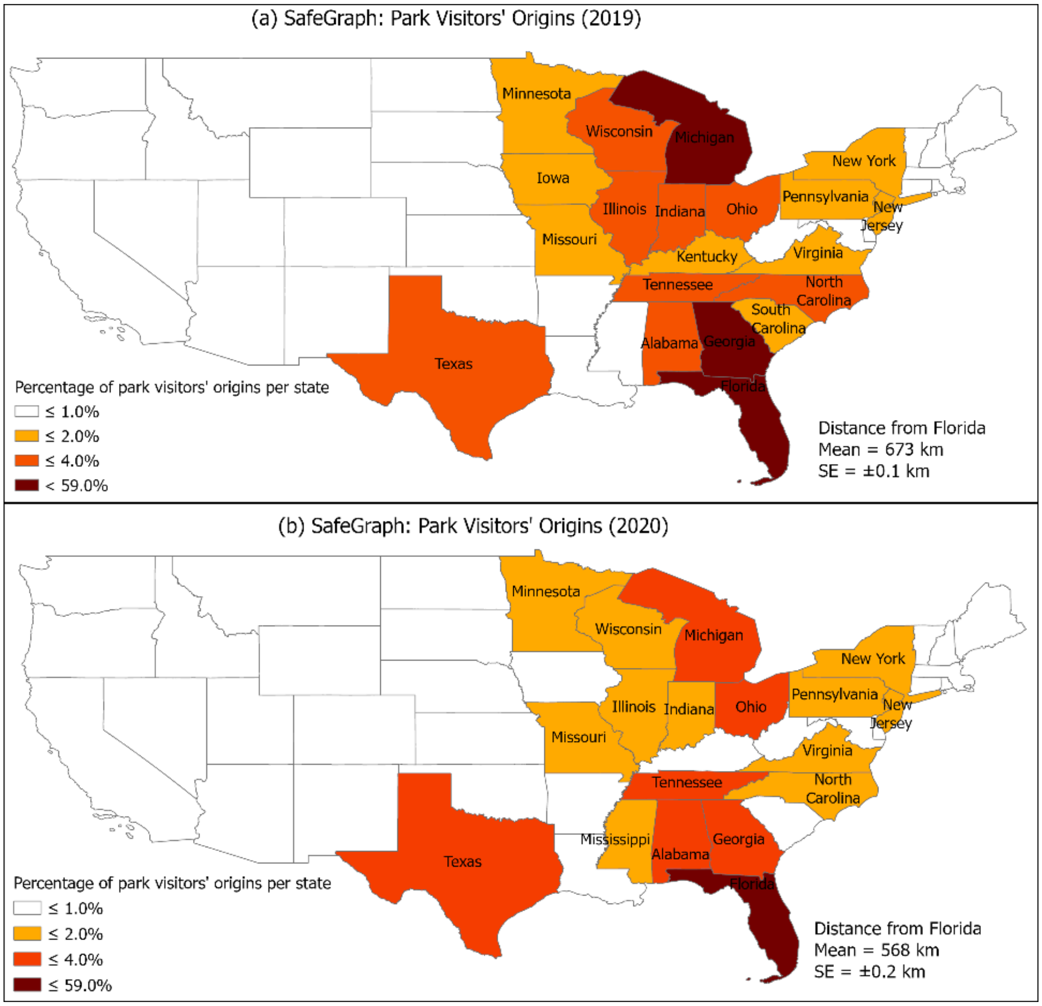

Figure 9. Origin of Florida state park visitors based on SafeGraph data in 2019 (a) and 2020 (b).

A Chi-square test of independence showed that for each data source there was a significant association between the proportion of local vs. out-of-state reviewers/visitors and the analysis period (Table 9), hence social media support the trend identified in reference SafeGraph data.

Table 9 Chi-square test cross-table for local versus out-of-state reviewers/visitors for Florida

Data source	**2019**		**2020**			
	Local visitors/ reviewers	**Out of state visitors/ reviewers**	**Local visitors/ reviewers**	**Out of state visitors/ reviewers**	**χ^2 (df= 1)**	**p**
TripAdvisor	1109 (42.7%)	1486 (57.3%)	462 (53.3%)	405 (46.7%)	29.2	< 0.001
Yelp	288 (55.7%)	232 (44.3%)	253 (66.2%)	131 (33.8%)	10.1	0.001
SafeGraph	3092 (51.1%)	2956 (48.9%)	2578 (59.4%)	1763 (40.6%)	69.6	< 0.001

In Austria, during the pandemic, the number of countries of origin for TripAdvisor reviewers dropped from nine to eight, and the mean distance between reviewer origin country and Austria dropped from 408 km in 2019 (Figure 10a) to 261 km 2020 (Figure 10b). The percentage of reviewers from Austria increased from 28% (pre-pandemic) to 45% (during the pandemic).

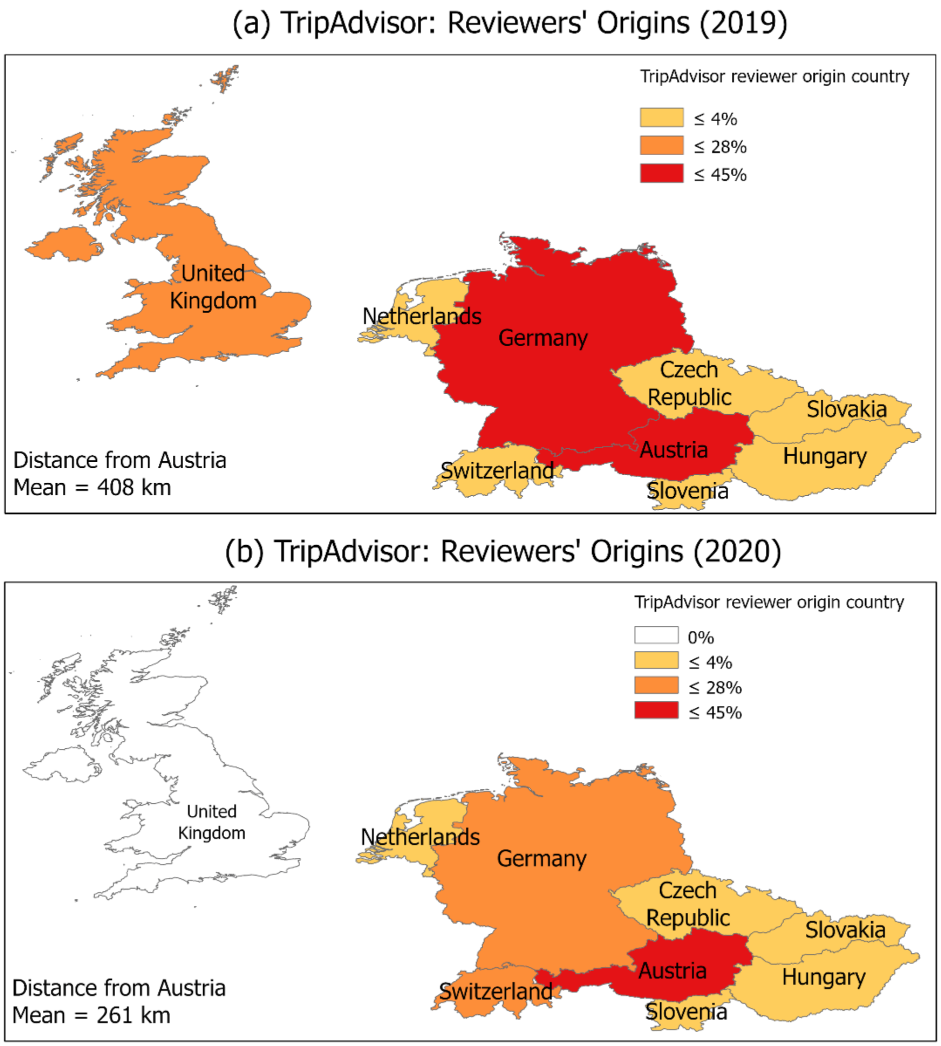

Figure 10 Carinthia attractions reviewers' origins on TripAdvisor in 2019 (a) and 2020 (b).

During the pandemic, the number of countries of Flickr user origins dropped from three to two. The mean distance between a Flickr user's origin country and Austria dropped from 246 km 2019 (Figure 11a) to 62 km in 2020 (Figure 11b). The percentage of reviewers from Austria increased from 51% (pre-pandemic) to 78% (during pandemic).

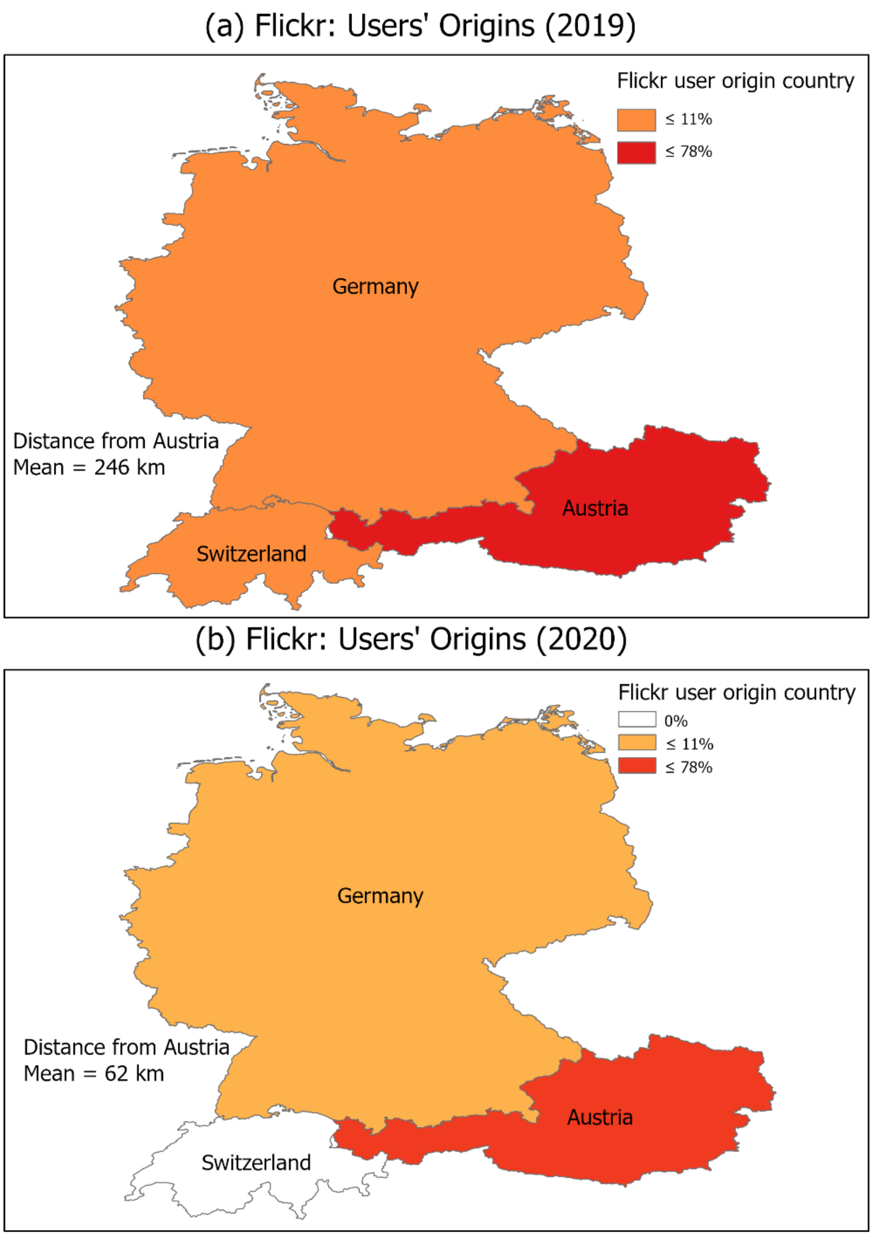

Figure 11 Carinthia attractions Flickr users' origins on TripAdvisor in 2019 (a) and 2020 (b).

Chi-square test results in Table 10 showed that the proportion of local vs. foreign TripAdvisor reviewers and Flickr users was significantly associated with the analysis period. A larger percentage of reviewers and Flickr users were local (Austria based) during the pandemic than before.

Table 10 Chi-square test cross-table for local versus out-of-state reviewers/visitors for Carinthia

Data source	**2019**		**2020**			
	Austria	**Outside of Austria**	**Austria**	**Outside of Austria**	χ^2 **(df= 1)**	**p**
TripAdvisor	41 (28%)	105 (72%)	40 (45%)	49 (55%)	6.9	<0.05
Flickr	103 (51%)	98 (49%)	14 (78%)	4 (22%)	4.6	<0.05

5. Discussion

With respect to research task 1, this research found strong significant positive correlations between contribution counts based on online user-generated information from Wikipedia page views, Google Maps reviews and TripAdvisor reviews, and official visitation statistics, respectively. Wikipedia page views were confirmed to be a reliable indicator for estimating tourist demand by earlier studies (Alis et al., 2015; Donovan et al., 2017). Visitation patterns extracted from TripAdvisor reviews of various attractions in Florida were found to be strongly correlated to those derived from mobile phone tracking data and official visitation survey data (Ma & Kirilenko, 2021). Studies involving Google Maps user reviews predominantly use sentiment analysis to assess tourist experiences in various tourist facilities such as restaurants (Mathayomchan & Taecharungroj, 2020). Therefore, our study is the first to the of our knowledge that has investigated its linear relationship with official tourism visitation statistics.

Correlations between official tourism visitation statistics and Wikipedia page edits, Flickr, and Twitter were not significant, which may be attributed to the small sample size of data provided in these platforms. Twitter is widely used in tourism research, but our study found there was no significant correlation revealed between Twitter post numbers and the official tourism visitation counts in Florida state parks. In the Carinthia study, the sample size tweets detected in only one of the attractions under investigation led to the exclusion in the Austria study. The analysis of only tweets with exact coordinates that lie within the boundaries of the attractions in our study further reduced the sample size of tweets available for research. Studies have also found the change in Twitter policy in 2019 (Twitter Support [@TwitterSupport], 2019) which removed the ability of users to tag their tweets with precise locations through the app led to users posting fewer tweets with exact coordinates and this affects the data sample available for research (Cao et al., 2022). A study which compared results from ordinary least squares, negative binomial and quantile regressions found there is a positive relationship between Facebook penetration and tourist arrivals in over 100 countries (Asongu & Odhiambo, 2019). Therefore, when using social media applications, the app usage in an area should be taken into consideration as it will impact the data availability as demonstrated in this study.

Flickr data is also predominantly used in tourism studies, but our research finds that it is a weak indicator for tourism visitation patterns in parks and attractions in Florida and Carinthia respectively, with our analysis not revealing significant correlations in both study areas. Flickr Pro a paid subscription which was introduced in 2018 enables the upload of more than 1000 photos or videos to the platform. Old content on Flickr accounts under free subscriptions that had exceeded the free account 1000 item limit were automatically deleted which may have affected the data sample size from the platform (Stadlen, 2018). As our study starts right after Flickr Pro was introduced, we theorize that it might have influenced the sample size available for our research and consequently our results. Flickr and Twitter were found to have lower correlations to official visitor counts in parks compared to Instagram in a previous study (Tenkanen et al., 2017). Correlations between Wikipedia page edits and official visitor counts were not significant as the number of edits per month was less than ten with majority of the time periods having zero number of edits in both study areas, thus our research notes that it is not a reliable indicator for tourism demand in parks.

In both study areas, the regression models indicate that Google Maps reviews are the strongest predictor of official visitation counts. In the multiple regression model, our study employed Google Maps reviews and Wikipedia page views. A previous study found that multisource datasets comprising of online review platforms and nonsocial media data sources can provide more accurate results when tourist demand forecasting than when a single platform is used (Li et al., 2020). Various regression models have employed Wikipedia page views to infer tourism demand (Ashouri et al., 2022; Hinnosaar et al., 2021).

Seasonal patterns are inherent in the tourism sector (Chen & Pearce, 2012) as they are often impacted by factors such as weather conditions and occurrences such as holidays. The visitation patterns in recreational outdoor spaces in areas with humid tropical climate such as Florida and the Alpine region of Carinthia particularly vary with seasons. In the Florida study, official visitation counts, and Google Maps reviews had peaks in the winter season (January) with a higher peak in March which coincide with the spring break which is a period in which Florida known to be a popular travel destination. The highest peak was noted in the summer season (June). Similar temporal patterns have been found in Florida amusement parks (Juhász & Hochmair, 2020). Official visitation counts and Google Maps reviews in the Carinthia time series analysis exhibited one major peak in the summer season (August) which is concurrent with researchers who found an uptick of mountain bike tourism in the southern region of the country during the summer season (Pröbstl-Haider et al., 2018). A significant drop in both official visitation counts and Google Maps reviews related to Florida parks was detected in April 2020 which comes after travel restrictions in the U.S. were imposed from March 2020 during the COVID-19 pandemic (CDC, 2022).

Analysis results of task 2 show that there were changes in the distribution of the origins of online reviewers and visitors of Florida state parks and Carinthia attractions from before to during the COVID-19 pandemic. Results indicate a higher share of local visitors during the pandemic (2020) than before the pandemic (2019) for both study areas implying that the parks were mostly utilized by locals during the pandemic. The difference in percentage between local and foreign visitors between before and during the pandemic periods for Florida (TripAdvisor, Yelp and SafeGraph data) and Carinthia (TripAdvisor and Flickr) were statistically significant for the analyzed year. The number of foreign visitors and visitor trip distances to Florida state parks and Carinthia attractions also significantly dropped in the year 2020 for both places, revealing how the COVID-19 pandemic impacted travel behavior. Tourism activities in Indonesia during the pandemic in 2020 were as a result of domestic travelers (Pramana et al., 2022) while in Czechia international tourist reduced significantly but domestic tourism in rural areas increased substantially (Vaishar & Šťastná, 2022). Research that used online ticket sales for attractions to study tourism patterns, found that local tourism increased as a result of the COVID-19 pandemic (Li, et al., 2021). The integrity of social media geodata is affected by a range of factors such as automated content generated through bots is estimated to be at least 9% of its accounts on Twitter (Varol et al., 2017). Data vandalism, and user privacy restrictions may also lead to incorrect origins (home) locations provided in a user profile. The reviewer origins collected from TripAdvisor, Yelp and Flickr user profiles were not validated. Therefore, validation methods, such as analyzing the number of likes a reviewer received or checking whether a user's reviews are primarily posted for POIs near the user's origin location, will be applied as part of future work.

6. Conclusions

The study employed correlation, multiple regression, time series analysis and origin-destination analysis to find the digital data platform that provided the best match to the reference data used in the respective case. Our results agree with previous studies which found Wikipedia page views and TripAdvisor reviews to have strong positive linear relationship with official visitation statistics and highlights new findings that Google Maps reviews also has a strong positive linear dependence with the reference data. The regression results found Google Maps reviews to have the strongest significant explanatory power over the official tourism statistics in both study areas and this further underscores the role of Google Maps as a popular resource for tourists when travel planning. Seasonal temporal patterns extracted from Google Maps reviews were found to best match the patterns exhibited from the official tourism statistics, which further fosters its value in tourism studies. The principal findings from TripAdvisor, Yelp and SafeGraph in Florida and TripAdvisor and Flickr in Austria origin-destination studies revealed the impacts of COVID-19 on travel, with both sites recording significant increments in local reviewers/visitors from the data platforms, which emphasizes the effects that travel mitigation strategies put in place during the pandemic had on the tourism sector.

References

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research, 58(2), 175–191. https://doi.org/10.1177/0047287517747753

Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. Transportation Research Part C: Emerging Technologies, 58, 240–250. https://doi.org/10.1016/j.trc.2015.02.018

Alis, C. M., Letchford, A., Moat, H. S., & Preis, T. (2015). Estimating tourism statistics with Wikipedia page views. Proceedings of the ACM Web Science Conference, 1–2. https://doi.org/10.1145/2786451.2786925

Ashouri, M., Hyndman, R. J., & Shmueli, G. (2022). Fast Forecast Reconciliation Using Linear Models. Journal of Computational and Graphical Statistics, 31(1), 263–282. https://doi.org/10.1080/10618600.2021.1939038

Asongu, S., & Odhiambo, N. M. (2019). Tourism and social media in the world: An empirical investigation. Journal of Economic Studies, 46(7), 1319–1331. https://doi.org/10.1108/JES-07-2018-0239

Bloomfield, P. (2004). Fourier analysis of time series: An introduction. John Wiley & Sons.

Buhalis, D., & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. Journal of Destination Marketing & Management, 4(3), 151–161.

Cao, J., Hochmair, H. H., & Basheeh, F. (2022). The Effect of Twitter App Policy Changes on the Sharing of Spatial Information through Twitter Users. Geographies, 2(3), Article 3. https://doi.org/10.3390/geographies2030033

CDC. (2022, August 16). CDC Museum COVID-19 Timeline. Centers for Disease Control and Prevention. https://www.cdc.gov/museum/timeline/covid19.html

Chen, T., & Pearce, P. L. (2012). Research Note: Seasonality Patterns in Asian Tourism. Tourism Economics, 18(5), 1105–1115. https://doi.org/10.5367/te.2012.0163

Cheng, Y., Zhang, J., Wei, W., & Zhao, B. (2021). Effects of urban parks on residents' expressed happiness before and during the COVID-19 pandemic. Landscape and Urban Planning, 212, 104118. https://doi.org/10.1016/j.landurbplan.2021.104118

Cirer-Costa, J. C. (2017). Turbulence in Mediterranean tourism. Tourism Management Perspectives, 22, 27–33. https://doi.org/10.1016/j.tmp.2017.01.004

Coppack, S. W. (1990). Correspondence Section: Limitations of the Pearson Product-Moment Correlation. Clinical Science, 79(3), 287–287.

Donovan, C., Flaherty, E. T., & Quinn Healy, E. (2017). Using big data from Wikipedia page views for official tourism statistics. Statistical Journal of the IAOS, 33(4), 997–1003. https://doi.org/10.3233/SJI-160320

Dutta, A., Mishra, T., Uddin, G. S., & Yang, Y. (2021). Brexit uncertainty and volatility persistence in tourism demand. Current Issues in Tourism, 24(16), 2225–2232. https://doi.org/10.1080/13683500.2020.1822300

Falk, M. (2014). Impact of weather conditions on tourism demand in the peak summer season over the last 50years. Tourism Management Perspectives, 9, 24–35. https://doi.org/10.1016/j.tmp.2013.11.001

Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. Tourism Management, 72, 306–312. https://doi.org/10.1016/j.tourman.2018.12.007

Hale, B. W. (2018). Mapping potential environmental impacts from tourists using data from social media: A case study in the Westfjords of Iceland. Environmental Management, 62(3), 446–457.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access.

Computers, Environment and Urban Systems, 72, 38–50. https://doi.org/10.1016/j.compenvurbsys.2018.01.007

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. Conservation Letters, 11(1), e12343.

Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 237–246.

Hecht, B. J., & Gergle, D. (2010). On the "localness" of user-generated content. Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, 229–232. https://doi.org/10.1145/1718918.1718962

Hinnosaar, M., Hinnosaar, T., Kummer, M. E., & Slivko, O. (2017). Does Wikipedia matter? The effect of Wikipedia on tourist choices. The Effect of Wikipedia on Tourist Choices (September 25, 2017). ZEW-Centre for European Economic Research Discussion Paper, 15–089.

Hinnosaar, M., Hinnosaar, T., Kummer, M., & Slivko, O. (2021). Wikipedia matters. Journal of Economics & Management Strategy.

Hou, J., Ndasauka, Y., Pan, X., Chen, S., Xu, F., & Zhang, X. (2018). Weibo or WeChat? Assessing preference for social networking sites and role of personality traits and psychological factors. Frontiers in Psychology, 9, 545.

Hu, F., Li, Z., Yang, C., & Jiang, Y. (2019). A graph-based approach to detecting tourist movement patterns using social media data. Cartography and Geographic Information Science, 46(4), 368–382. https://doi.org/10.1080/15230406.2018.1496036

Jiang, Y., Huang, X., & Li, Z. (2021). Spatiotemporal Patterns of Human Mobility and Its Association with Land Use Types during COVID-19 in New York City. ISPRS International Journal of Geo-Information, 10(5), Article 5. https://doi.org/10.3390/ijgi10050344

Johnson, I. L., Sengupta, S., Schöning, J., & Hecht, B. (2016). The Geography and Importance of Localness in Geotagged Social Media. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 515–526. https://doi.org/10.1145/2858036.2858122

Juhász, L., & Hochmair, H. H. (2020). Studying spatial and temporal visitation patterns of points of interest using SafeGraph data in Florida.

Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., & Shirayama, Y. (2020). Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. Tourism Management, 77, 104010. https://doi.org/10.1016/j.tourman.2019.104010

Landesregierung, A. der K. (2019, November 7). Tourism. Tourism. https://www.carinthia.com/en/business-technology/tourism

Li, H., Hu, M., & Li, G. (2020). Forecasting tourism demand with multisource big data. Annals of Tourism Research, 83, 102912. https://doi.org/10.1016/j.annals.2020.102912

Li, X., Gong, J., Gao, B., & Yuan, P. (2021). Impacts of COVID-19 on tourists' destination preferences: Evidence from China. Annals of Tourism Research, 90, 103258. https://doi.org/10.1016/j.annals.2021.103258

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. Tourism Management, 83, 104245. https://doi.org/10.1016/j.tourman.2020.104245

Ma, S., & Kirilenko, A. (2021). How reliable is social media data? Validation of TripAdvisor tourism visitations using independent data sources. In Information and communication technologies in tourism 2021 (pp. 286–293). Springer.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 27(4), 662–684. https://doi.org/10.1080/13658816.2012.696649

Mathayomchan, B., & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. International Journal of Hospitality Management, 90, 102641. https://doi.org/10.1016/j.ijhm.2020.102641

Mexia, T., Vieira, J., Príncipe, A., Anjos, A., Silva, P., Lopes, N., Freitas, C., Santos-Reis, M., Correia, O., Branquinho, C., & Pinho, P. (2018). Ecosystem services: Urban parks under a magnifying glass. Environmental Research, 160, 469–478. https://doi.org/10.1016/j.envres.2017.10.023

Milman, A. (2019). Visitor management in highly-visited attractions: Lessons that practitioners can learn from the U.S. theme park industry and the national parks. In Overtourism. Routledge.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. (2013). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. Proceedings of the International AAAI Conference on Web and Social Media, 7(1), Article 1.

Nakayama, M., & Wan, Y. (2019). Same sushi, different impressions: A cross-cultural analysis of Yelp reviews. Information Technology & Tourism, 21(2), 181–207. https://doi.org/10.1007/s40558-018-0136-5

O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity, 41(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Owuor, I., & Hochmair, H. H. (2020). An Overview of Social Media Apps and their Potential Role in Geospatial Research. ISPRS International Journal of Geo-Information, 9(9), Article 9. https://doi.org/10.3390/ijgi9090526

Peng, T., Chen, J., Wang, C., & Cao, Y. (2021). A Forecast Model of Tourism Demand Driven by Social Network Data. IEEE Access, 9, 109488–109496. https://doi.org/10.1109/ACCESS.2021.3102616

Pestana, M. H., Parreira, A., & Moutinho, L. (2020). Motivations, emotions and satisfaction: The keys to a tourism destination choice. Journal of Destination Marketing & Management, 16, 100332. https://doi.org/10.1016/j.jdmm.2018.12.006

Pramana, S., Paramartha, D. Y., Ermawan, G. Y., Deli, N. F., & Srimulyani, W. (2022). Impact of COVID-19 pandemic on tourism in Indonesia. Current Issues in Tourism, 25(15), 2422–2442. https://doi.org/10.1080/13683500.2021.1968803

Pröbstl-Haider, U., Lund-Durlacher, D., Antonschmidt, H., & Hödl, C. (2018). Mountain bike tourism in Austria and the Alpine region – towards a sustainable model for multi-stakeholder product development. Journal of Sustainable Tourism, 26(4), 567–582. https://doi.org/10.1080/09669582.2017.1361428

R Core Team and contributors worldwide. (2013). periodogram function—RDocumentation. https://www.rdocumentation.org/packages/TSA/versions/1.3/topics/periodogram

Rusell, E. (2019, September 30). 9 things to know about Google's maps data: Beyond the Map. Google Cloud Blog. https://cloud.google.com/blog/products/maps-platform/9-things-know-about-googles-maps-data-beyond-map/

Schirpke, U., Meisch, C., Marsoner, T., & Tappeiner, U. (2018). Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. Ecosystem Services, 31, 336–350. https://doi.org/10.1016/j.ecoser.2017.11.017

Schroepfer, M. (2018, April 4). An Update on Our Plans to Restrict Data Access on Facebook. Facebook.Com. https://about.fb.com/news/2018/04/restricting-data-access/

Sedgwick, P. (2012). Pearson's correlation coefficient. BMJ, 345, e4483. https://doi.org/10.1136/bmj.e4483

Signorelli, S., Reis, F., & Biffignandi, S. (2016). What attracts tourists while planning for a journey? An analysis of three cities through Wikipedia page views.

Stadlen, A. (2018, November 1). Why we're changing Flickr free accounts. Flickr Blog. https://blog.flickr.net/en/2018/11/01/changing-flickr-free-accounts-1000-photos/

Statista. (2021, February 11). Yelp: Cumulative number of reviews 2020. Statista. https://www.statista.com/statistics/278032/cumulative-number-of-reviews-submitted-to-yelp/

Teles da Mota, V., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. Journal of Outdoor Recreation and Tourism, 30, 100295. https://doi.org/10.1016/j.jort.2020.100295

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific Reports, 7(1), Article 1. https://doi.org/10.1038/s41598-017-18007-4

TripAdvisor, T. (2022, February 1). Travelers Push Tripadvisor Past 1 Billion Reviews & Opinions! MediaRoom. https://tripadvisor.mediaroom.com/2022-02-01-Travelers-Push-Tripadvisor-Past-1-Billion-Reviews-Opinions

Twitter. (2022, September 11). Advanced filtering for geo data. https://developer.twitter.com/en/docs/tutorials/advanced-filtering-for-geo-data

Twitter Support [@TwitterSupport]. (2019, June 18). Most people don’t tag their precise location in Tweets, so we’re removing this ability to simplify your Tweeting experience. You’ll still be able to tag your precise location in Tweets through our updated camera. It’s helpful when sharing on-the-ground moments. [Tweet]. Twitter. https://twitter.com/TwitterSupport/status/1141039841993355264

Vaishar, A., & Šťastná, M. (2022). Impact of the COVID-19 pandemic on rural tourism in Czechia Preliminary considerations. Current Issues in Tourism, 25(2), 187–191. https://doi.org/10.1080/13683500.2020.1839027

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Eleventh International AAAI Conference on Web and Social Media.

VISIT FLORIDA. (2022). About Us. https://www.visitflorida.com/about-us/

Volchek, K., Liu, A., Song, H., & Buhalis, D. (2019). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. Tourism Economics, 25(3), 425–447. https://doi.org/10.1177/1354816618811558

Volenec, Z. M., Abraham, J. O., Becker, A. D., & Dobson, A. P. (2021). Public parks and the pandemic: How park usage has been affected by COVID-19 policies. PLOS ONE, 16(5), e0251799. https://doi.org/10.1371/journal.pone.0251799

World Travel & Tourism Council. (2022). WTTC/Oxford Economics 2022 Travel & Tourism Economic Impact Research Methodology.

https://wttc.org/Portals/0/Documents/Reports/2022/WTTC%20methodology%20report%202022.pdf?ver=2022-06-13-202957-567

Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. International Journal of Contemporary Hospitality Management.

Yan, Y., Chen, J., & Wang, Z. (2020). Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. Applied Geography, 123, 102306. https://doi.org/10.1016/j.apgeog.2020.102306

Yang, Y., Zhang, X., & Fu, Y. (2022). Foreign tourists' experiences under air pollution: Evidence from big data. Tourism Management, 88, 104423. https://doi.org/10.1016/j.tourman.2021.104423

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. Tourism Management Perspectives, 10, 27–36. https://doi.org/10.1016/j.tmp.2014.01.001

Zheng, X., Han, J., & Sun, A. (2018). A Survey of Location Prediction on Twitter. IEEE Transactions on Knowledge and Data Engineering, 30(9), 1652–1671. https://doi.org/10.1109/TKDE.2018.2807840

Ziesler, P. S., & Pettebone, D. (2018). Counting on visitors: A review of methods and applications for the National Park Service's visitor use statistics program. Journal of Park and Recreation Administration, 36(1).