

FROM TEXT TO INSIGHT: A COMPARATIVE ANALYSIS OF SENTIMENTS AND EMOTIONS IN RURAL AND URBAN TOURISM EXPERIENCES IN SPAIN

Maria Fernanda BERNAL SALAZAR¹¹ PhD Students

Elisa BARAIBAR-DIEZ¹² PhD

Jesús COLLADO AGUDO¹³ PhD

ORCID: 0000-0002-6069-082X, 0000-0003-4677-3255, 0000-0002-4152-4439

Abstract: *In an era where the digital environment plays a pivotal role in shaping consumer choices, sentiment analysis emerges as a crucial tool for unravelling tourists experiences. This study employs sentiment and emotion analyses to delve into tourist reviews in rural and urban areas in Spain. Employing advanced text mining techniques with Python programming, we scrap 173,369 reviews from rural areas and 347,040 reviews in urban areas from the website TripAdvisor from the period 2012-2022. Results show significant differences in sentiments and emotions by area and by review category (environmental, social, economic, technological, and cultural), offering valuable insights for destination management and strategic tourism development.*

Key words: *Sentiment analysis, emotion analysis, tourism, sustainability, engagement, text mining.*

JEL: M14 Corporate Culture • Diversity • Social Responsibility.

1. Introduction

User-generated content (UGC) in the field of tourism has emerged as an essential component to understand and improve travel experiences. This type of content, created spontaneously by travelers themselves, offers an authentic and valuable perspective on destinations, accommodation, and activities. The UGC is considered an effective tool that travelers use to obtain essential information when making travel decisions (Ukpabi & Karjaluoto, 2017). Given the intangible nature and experiential approach of tourism, the UGC diminishes the uncertainty of travelers (Kim & Kim, 2020) and can be fostered by the desire to help others (altruistic concern) or as social interaction from a virtual community (Chen et al., 2011; Tian, 2013).

Some studies confirm the relationship between user ratings and the performance of a product or service (Dellarocas et al., 2007; Geetha et al., 2017; Sharma & Aggarwal, 2021; Zhang et al., 2010). The online popularity of an establishment can benefit significantly by increasing the number of positive reviews and high scores. The study by Tsao et al., (2015) showed that around 80% of travelers read hotel reviews before starting a trip, and 53% say they will avoid booking a hotel without comments. Both companies and academics analyze the UGC to understand social phenomena and passenger behavior (Mehra, 2023), which allows them to improve marketing strategies, develop competitive advantages and useful models for decision-making.

¹¹ maria-fernanda.bernal@alumnos.unican.es, University of Cantabria

¹² elisa.baraibar@unican.es, University of Cantabria

¹³ jesus.collado@unican.es, University of Cantabria

Understanding the differences in perception between tourists in rural and urban destinations can have important implications for tourism professionals and government policymakers. While, in rural regions, tourism can be an essential driving force in combating poverty and preserving local culture, in urban areas, tourism is seen as one of several social and economic influences. Understanding these differences allows us to develop specific strategies that adapt to the needs and characteristics of each environment, ensuring sustainable and equitable development in the tourism industry in all its manifestations.

This study focuses on analyzing the feelings and emotions associated with tourist attractions in rural and urban areas of Spain, to understand the complexity of visitors' experiences from the perspective of sustainability and technology. First, a review of the literature is done. Then, the research objective and questions are set. The methodology used and the results are detailed. Finally, the conclusions are presented.

2. Literature Review and research questions

UGC analysis can include text analysis methods, sentiment evaluation and spatial analysis (Mirzaalian & Halpenny, 2019). Some previous research has focused on understanding the impact of positive experiences and emotions on behavioral intentions (Hosany & Gilbert, 2009; Lee & Kyle, 2011). From psychology there are two approaches to examine emotions: 1) Dimensional: Based on valence, which conceptualizes emotions using few dimensions (positive, neutral, and negative). 2) Categorical: Emotions are conceived as a series of unique and personalized affective states, such as joy, anger, sadness, and surprise (idiosyncratic affective states). Previous studies in the field of tourism show that: 1) emotions and their intensity vary depending on the products and stages of the trip. 2) emotions influence the satisfaction of tourist experiences and behavioral intentions 3) tourist experiences remembered by travelers are associated with positive emotions (Prayag et al., 2017).

Tourists have a favorable attitude towards sustainable options during the trip, but a gap is evident between those positive feelings and the desire of the tourist to experience a pleasant and high quality holiday (Budeanu, 2007; Dolnicar, 2020; Dolnicar & Grün, 2009). Some research in tourism has begun to analyze sustainability and corporate social responsibility (CSR) from the UGC (Brazytė et al., 2016; D'Acunto et al., 2020; Ettinger et al., 2018; Väisänen et al., 2023). The study by Väisänen et al. (2023) proposes as a future line of research to carry out these analyses comparing data from rural and urban destinations. From sustainability (economic, social, environmental and cultural) can be identified notable differences between rural and urban tourist areas (Rasoolimanesh et al., 2017). Rural tourism is seen as an essential tool to combat poverty, preserve the natural environment and promote local culture (Ryan & Huimin, 2009). On the contrary, in the urban context the contribution of tourism can be considered as part of a broader and diversified panorama of forces that shape the life and economy of the city (Edwards et al., 2008). On the other hand, it is essential to examine the way in which tourists use and appreciate technologies within industry, as digitization and technological innovations are transforming travelers' behavior, which poses new challenges for businesses and destinations.

In this way, the main objective of this study is: "Analyze the feelings and emotions of rural and urban areas of the tourist destination of Spain from a sustainable and technological perspective". The following research questions are asked:

Q1: What are the perceptions and themes that tourists reflect in the comments of the TripAdvisor platform in rural and urban areas?

Q2: Has the sustainable and technological perspective of tourists changed over time in rural and urban areas?

Q3: Do the dimensional and categorical emotions of tourists vary in rural and urban areas from a sustainable and technological perspective?

3. Research methodology

Research objectives are addressed through an inductive approach, analyzing the perception of tourists in attractions in urban and rural areas. To do this, the classification between rural and urban areas has been carried out according to Law 45/2007 on Sustainable Development of the Rural Environment in Spain, which consider a rural municipality as one with less than 100 inhabitants per km². In total, 554 rural municipalities and 400 urban municipalities have been considered.

Data collection, pre-processing and analysis of data are executed using the Python programming language, taking advantage of its capabilities and libraries for efficient handling of large datasets. The implementation of advanced techniques in natural language processing and machine learning contribute to a comprehensive understanding of tourist sentiments, opinions, and preferences, enhancing the rigor and depth of the study. Tourists' reviews have been sourced from the TripAdvisor website, chosen for its comprehensive coverage and widespread use among both users and businesses in the tourism sector. Data collection occurred from April to June 2023, employing web scraping techniques to extract reviews and comments from key tourist attractions across the 17 Autonomous Communities of Spain. The sample comprises 173,369 reviews from rural areas and 347,040 reviews from urban areas, spanning the period from 2012 to 2022.

Data pre-processing tasks are essential to enhance the quality of the data. This phase involves data cleaning (i.e., handling missing data, correcting errors and duplicities), normalization and tokenization (breaking down text into smaller units or tokens).

Data analysis is carried out in three stages. First, topic modelling in tourists' perceptions is carried out with BERTopic (Grootendorst, 2021) topic modelling machine learning technique, which can capture the semantic context of the text. To do this, Spanish reviews were translated into English for superior performance. Secondly, with the aim of exploring tourists' perceptions by area, sentiment analysis and emotion analysis are carried out. Opinion mining is the process of determining the sentiment or attitude expressed in a piece of text, so that the goal is to identify whether the expressed sentiment is positive, negative, or neutral. Emotion analysis goes beyond sentiment analysis by identifying specific emotions such as anger, disgust, fear, joy, sadness, or surprise. In order to investigate potential differences in sentiments and emotions across areas and dimensions, mean difference tests are conducted. Initially we examine the normal distribution of the groups using Kolmogorov-Smirnov and Anderson-Darling tests. Given that normal distributions were not observed, non-parametric test are employed to assess significant differences: the Wilcoxon test for comparisons between two groups (rural/urban) and the Kruskal-Wallis test for comparisons involving three or more groups (dimensions).

4. Results

The evolution of tourists' perceptions in rural and urban areas is shown in Figure 1. Using text mining methodologies, reviews were first classified into five dimensions (environmental, social, cultural, economic, and technological), associating keywords with each of them and assigning the review to

the most relevant dimension. The technological dimension is the least highlighted by the reviews in both areas, while the cultural dimension is the most notable, for the entire period considered. While in the rural environment the social dimension has taken on more relevance since 2020, although without surpassing the environmental dimension, in the urban area the social and environmental dimensions become less relevant than the economic one. The importance of these dimensions is consistent to the topics identified in each area. In the rural area, the five main topics identified by BERTopic (which considers both the relevance and frequency of topics) are heritage and culture, exploring villages, routes, nature, coasts and beaches, and oenology and gastronomy. In the urban area, main topics are heritage and culture, exploring cities, coasts and beaches, oenology and gastronomy, parks and routes, and entertainment (i.e., escape rooms).

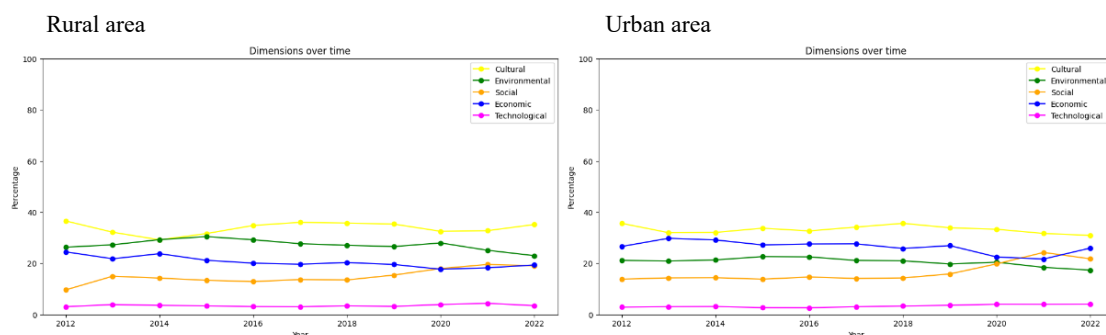


Figure 1. Evolution of reviews by dimension and by area. Period 2012-2022.

Source: Authors.

Reviews' sentiments are mostly positive, across all dimensions and in both rural and urban areas (Figure 2). While the differences are subtle, statistical significance is observed in sentiment between rural and urban areas, encompassing all reviews. When considering reviews based on dimensions, significant differences are also observed among all groups (Table 1).

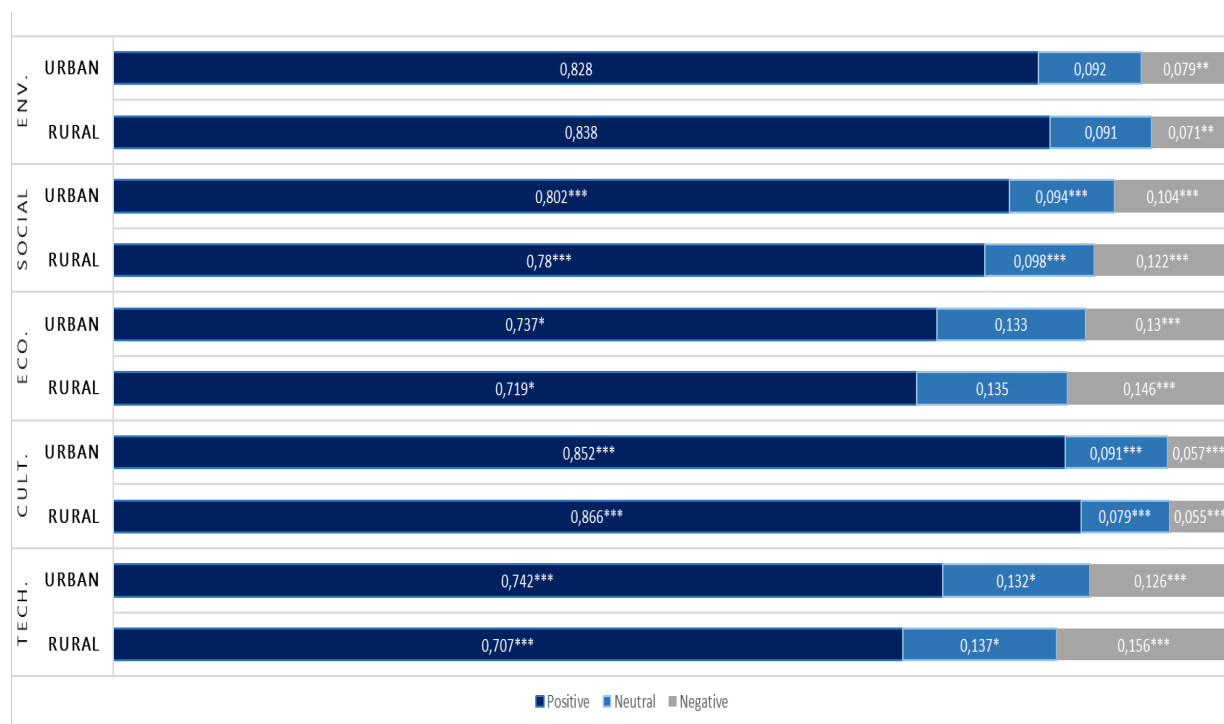
Table 1. Averages of sentiment scores by area and by dimension.

Area/Dimension	Positive	Neutral	Negative
Total	0,816853	0,101005	0,082142
Rural	0,8153301	0,09918843	0,085481143
Urban	0,8176057	0,10190195	0,0804924
p-value ^{1,3}	***	***	***
Cultural	0,8555247	0,08723674	0,05723860
Economic	0,7382728	0,13393691	0,12779027
Environmental	0,7990192	0,10392839	0,09705236
Social	0,7658777	0,10701800	0,12620435
Technological	0,7293921	0,13388624	0,13672170
NA	0,8394791	0,09439431	0,06612657
p-value ^{2,3}	***	***	***

Source: Authors.

Notes: ¹Wilcoxon test (for non-normal distributions). ²Kruskal-Wallis test (for non-normal distributions). ³Significance levels are denoted as *** for p-value < 0,01, ** for p-value < 0,05, and * for p-value < 0,1.

Differences between areas (urban/rural) in each dimension are shown in Figure 2. Reviews categorized as social, economic, and technological are statistically more positive in urban areas. This may be due to better technological infrastructures in urban areas, so that tourists value access to digital services. The wide range of economic activities (i.e., shopping, dining, leisure...) and entertainment options can explain the positivity in the economic dimension. The positivity in the social dimension could be due to the cosmopolitan atmosphere and exposure to different cultures and diversity. However, reviews categorized as cultural are statistically more positive in rural areas (as are environmental, but not statistically significant). Limitations of nature-based activities in urban areas and their environmental challenges (such as traffic or pollution) mean that positivity is somewhat lower than in rural environments.



Note: Given the non-normal distributions, the Wilcoxon test was employed to test differences between areas for each dimension. Significance levels are denoted as *** for p -value $< 0,01$, ** for p -value $< 0,05$, and * for p -value $< 0,1$.

Figure 2. Distribution of sentiments (positive, neutral, negative) by dimension and by area.

Source: Authors.

Figure 3 graphically represents the value of negative reviews by area and dimension at a national level, showing darker colors the more negative the reviews are in that autonomous community. The reviews are slightly more negative in the social, economic, and technological dimensions, with higher values in rural areas.

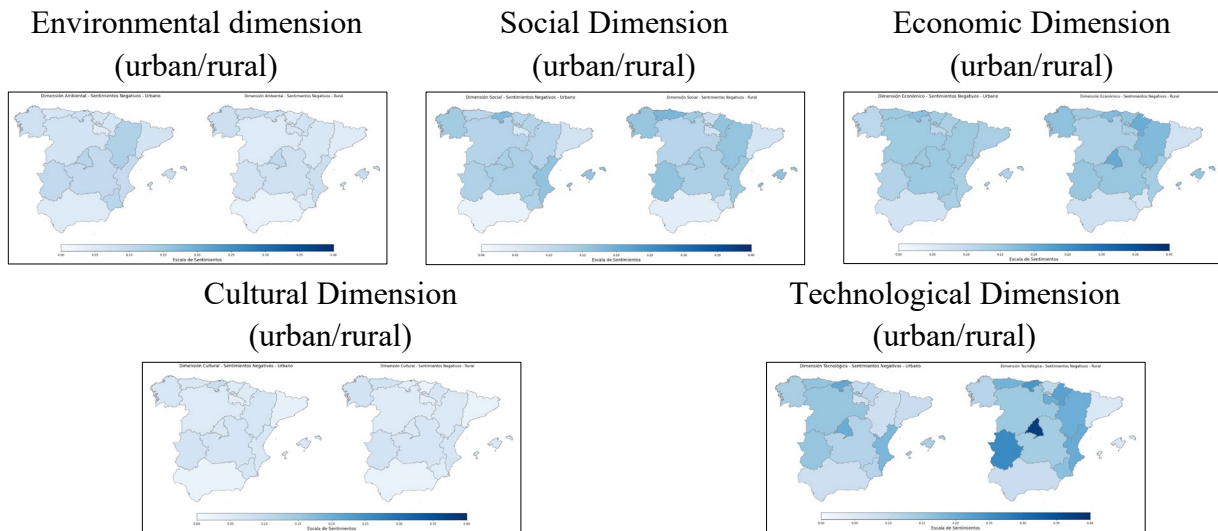


Figure 3. Distribution map of negative sentiments by dimension and by area.

Source: Authors.

Emotions are statistically different across dimensions and areas (Table 2). In rural areas, the predominant emotion is joy, followed by neutral emotions. Conversely, in urban areas, the prevalence of joy is lower than in rural counterparts, with neutral emotions and surprise following closely. Notably, urban areas experience higher levels of anger, disgust, and surprise in their reviews compared to rural areas, whereas joy and sadness are more prevalent in rural environments. The stress or faster pace of urban areas can lead to more visceral emotions, whereas a more leisurely pace might lead to the highlighting of joy in rural settings (Korpela et al., 2002).

Within rural areas, reviews categorized as economic are associated with higher levels of anger, disgust, and sadness, which may be due to greater difficulties in obtaining resources or the state of certain infrastructures. Fear is more pronounced in reviews categorized as cultural, and surprise is higher in reviews categorized as technological, explained by the fact that it is an emotion linked to the experimental aspect (Luo and Tang, 2019). In urban areas, anger is more prevalent in cultural and environmental dimensions, explained perhaps by environmental challenges or cultural clashes, while joy is heightened in reviews categorized as social. Reviews categorized as technological exhibit neutrality, and sadness and surprise are equally prevalent.

Table 2. Average of emotions by dimension and by area.

Dimension	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Rural	0,02450517	0,00334560	0,00186997	0,5997364	0,2876212	0,0719468	0,01097492
Urban	0,01880265	0,04511802	0,04143998	0,4522577	0,2878526	0,02920524	0,12532381
p-value ^{1,3}	***	***	***	***	***	***	***
Rural areas	Cultural	0,01001603	0,0020316	0,001156141	0,6476772	0,2516732	0,0792711
	Economic	0,06105162	0,00657664	0,001480065	0,4863087	0,3480369	0,08781419
	Environmental	0,03252017	0,00430677	0,002476473	0,5744618	0,2910572	0,08489994
	Social	0,04900918	0,00481837	0,00167626	0,5735614	0,2750448	0,00887172
	Technological	0,05156777	0,00438185	0,001992264	0,4629426	0,3809871	0,08663185
	NA	0,01296009	0,00240001	0,001974231	0,627846	0,2829568	0,05909843
	p-value ^{2,3}	***	***	***	***	***	***

Urban areas	Cultural	0,1681707	0,04164329	0,03828950	0,4230990	0,3120873	0,02645610	0,14160774
	Economic	0,02492892	0,05562108	0,03063869	0,3898931	0,3334170	0,03534864	0,13015247
	Environmental	0,1925329	0,06208555	0,04299324	0,4382652	0,3119718	0,03087387	0,09445714
	Social	0,02311569	0,05696678	0,04367805	0,4574377	0,2828742	0,02880065	0,10712701
	Technological	0,02113305	0,04897505	0,03678569	0,3469241	0,3488814	0,04018720	0,15711351
	NA	0,01706061	0,03813584	0,04396839	0,4767569	0,2669100	0,02791849	0,12924879
	p-value ^{2,3}	***	***	***	***	***	***	***

Source: Authors.

Notes: ¹Wilcoxon test (for non-normal distributions). ²Kruskal-Wallis test (for non-normal distributions). ³Significance levels are denoted as *** for $p\text{-value} < 0,01$, ** for $p\text{-value} < 0,05$, and * for $p\text{-value} < 0,1$.

5. Conclusion and discussion

This research contributes to explain the sustainable and technological perceptions of tourists of Spain's tourist destinations, analyzing the similitudes and differences between urban and rural areas. Considering TripAdvisor reviews along 10 years, more than 500,000 tourist reviews of rural and urban destinations were analyzed. The cultural dimension was the most notable in both rural and urban samples, whereas the technological dimension was the least highlighted by the reviews in both areas. The principal difference between rural and urban areas is related to the environmental and economic dimensions. The environmental dimension was more frequently mentioned in the comments in the rural area, while the economic dimension became more relevant in the urban context. Secondly, the results point out several differences between rural and urban areas in the level of sentiments and emotions across the sustainable and technological perceptions. On the one hand, reviews categorized as cultural and environmental are more positive in rural areas. In contrast, the social, economic, and technological dimensions are strongly associated with urban areas. On the other hand, emotions are quite different between the two analyzed contexts. Higher levels of anger, disgust, and surprise are associated to urban context, while joy and sadness are more relevant in rural environments.

From a managerial perspective, the mentioned results could be useful to destination managers in urban and rural areas. Depending on the context, destinations could match their key competitive features in terms of sentiments and emotions and their relationship with a specific dimension analyzed (cultural, environmental, social, economic, or technological). Moreover, knowing the perceptions of tourists in rural and urban areas allow companies and institutions to manage the expectations of users, to make effective investments that respond to real perceptions and to plan more effective communication campaigns.

Finally, this research has some limitations which deserve future venues. Related to the Spanish context, it is only used reviews for urban and rural attractions in Spain. Future research could amplify the range and compare the results with other competitive country destinations, such as France, Italy, or Greece. Another limitation is linked to the overall rating made in the review, not analyzed in this research. The consideration of the review rating as a variable to be explained would substantially enrich this research and, therefore, it would make it possible to analyses the effect of sentiments and emotions between urban and rural areas.

Acknowledgments

This work is part of the Research Project "TED2021-131314B-I00 SOSTENIBILIDAD CORPORATIVA Y TURISMO INTELIGENTE EN COMUNIDADES RURALES: INFLUENCIA EN EL DESARROLLO ECONÓMICO Y SOCIAL DEL TERRITORIO", funding by MICIU/AEI/10.13039/501100011033 and European Union (EU) “NextGenerationEU”/PRTR.

References

1. Budeanu, A. (2007). Sustainable tourist behaviour – a discussion of opportunities for change. *International Journal of Consumer Studies*, 31(5), 499–508. <https://doi.org/10.1111/j.1470-6431.2007.00606.x>
2. Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45. <https://doi.org/10.1002/dir.20087>
3. Dolnicar, S. (2020). Designing for more environmentally friendly tourism. *Annals of Tourism Research*, 84, 102933. <https://doi.org/10.1016/J.ANNALS.2020.102933>
4. Dolnicar, S., & Grün, B. (2009). Environmentally friendly behavior: Can heterogeneity among individuals and contexts/environments be harvested for improved sustainable management? *Environment and Behavior*, 41(5), 693–714. <https://doi.org/10.1177/0013916508319448>
5. Edwards, D., Griffin, T., & Hayllar, B. (2008). Urban Tourism Research: Developing an Agenda. *Annals of Tourism Research*, 35(4), 1032–1052. <https://doi.org/10.1016/J.ANNALS.2008.09.002>
6. Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis. *Tourism Management*, 61, 43–54. <https://doi.org/10.1016/j.tourman.2016.12.022>
7. Grootendorst, M. (2021). MaartenGr/BERTopic: Fix embedding parameter (v0.4.2). Zenodo. <https://doi.org/10.5281/zenodo.4430182>
8. Hosany, S., & Gilbert, D. (2009). Measuring Tourists' Emotional Experiences toward Hedonic Holiday Destinations. [Http://Dx.Doi.Org/10.1177/0047287509349267](http://Dx.Doi.Org/10.1177/0047287509349267), 49(4), 513–526. <https://doi.org/10.1177/0047287509349267>
9. Kim, M., & Kim, J. (2020). The Influence of Authenticity of Online Reviews on Trust Formation among Travelers. *Journal of Travel Research*, 59(5), 763–776. https://doi.org/10.1177/0047287519868307/ASSET/IMAGES/LARGE/10.1177_0047287519868307-FIG2.JPEG
10. Korpela, K. M., Klemetilä, T., & Hietanen, J. K. (2002). Evidence for Rapid Affective Evaluation of Environmental Scenes. *Environment and Behavior*, 34(5), 634–650. <https://doi.org/10.1177/0013916502034005004>
11. Lee, J., & Kyle, G. T. (2011). Recollection Consistency of Festival Consumption Emotions. [Http://Dx.Doi.Org/10.1177/0047287510394197](http://Dx.Doi.Org/10.1177/0047287510394197), 51(2), 178–190. <https://doi.org/10.1177/0047287510394197>
12. Luo, Y., Tang, R. (2019). Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *International Journal of Hospitality Management*, 80, 144–154.
13. Mehra, P. (2023). Unexpected surprise: Emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, 45, 101063. <https://doi.org/10.1016/J.TMP.2022.101063>
14. Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology*, 10(4), 764–790. <https://doi.org/10.1108/JHTT-08-2018-0078/FULL/XML>
15. Prayag, G., Hosany, S., Muskat, B., & Del Chiappa, G. (2017). Understanding the Relationships between Tourists' Emotional Experiences, Perceived Overall Image, Satisfaction, and Intention to Recommend. *Journal of Travel Research*, 56(1), 41–54. https://doi.org/10.1177/0047287515620567/ASSET/IMAGES/LARGE/10.1177_0047287515620567-FIG1.JPEG

16. Rasoolimanesh, S. M., Ringle, C. M., Jaafar, M., & Ramayah, T. (2017). Urban vs. rural destinations: Residents' perceptions, community participation and support for tourism development. *Tourism Management*, 60, 147–158. <https://doi.org/10.1016/J.TOURMAN.2016.11.019>
17. Ryan, C., & Huimin, G. (2009). *Tourism in China: Destination, cultures and communities*. Routledge.
18. Sharma, H., & Aggarwal, A. G. (2021). The influence of user generated content on hotel sales: an Indian perspective. *Journal of Modelling in Management*, 16(4), 1358–1375. <https://doi.org/10.1108/JM2-07-2020-0201>
19. Tsao, W. C., Hsieh, M. T., Shih, L. W., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management*, 46, 99–111. <https://doi.org/10.1016/J.IJHM.2015.01.008>
20. Ukpabi, D. C., & Karjaluoto, H. (2017). Consumers' acceptance of information and communications technology in tourism: A review. In *Telematics and Informatics* (Vol. 34, Issue 5, pp. 618–644). Elsevier Ltd. <https://doi.org/10.1016/j.tele.2016.12.002>
21. Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694–700. <https://doi.org/10.1016/j.ijhm.2010.02.002>