Smart Tourism 2025, 6(1), 3357.
https://doi.org/10.54517/st3357

Article

Visitor perceptions of over tourism: Text-mining insights from Seoul and Bali

Myrza Rahmanita[1], Surya Fadjar Boediman[2], Sri Mariati[2], Doni Muhardiansyah[2], Sofian Lusa[2], Rahmat Ingkadijaya[2], Mohammad Sayed Noor[3], Narariya Dita Handani[4,5], Rianmahardhika Sahid Budiharseno[4,*]

[1] Doctoral Program in Tourism, Faculty of Tourism, Trisakti Institute of Tourism, Jakarta 12330, Indonesia
[2] Master Program in Tourism, Faculty of Tourism, Trisakti Institute of Tourism, Jakarta 12330, Indonesia
[3] Department of Tourism Management, Dong-A University, Busan 49315, Korea
[4] School of Global Studies, Kyungsung University, Busan 48434, Korea
[5] Department of Management, Faculty of Economics and Business, Universitas Wijaya Kusuma Surabaya, Surabaya 60225, Indonesia
*** Corresponding author:** Rianmahardhika Sahid Budiharseno, riansb@ks.ac.kr

CITATION

Rahmanita M, Boediman SF, Mariati S, et al. Visitor perceptions of over tourism: Text-mining insights from Seoul and Bali. Smart Tourism. 2025; 6(1): 3357. https://doi.org/10.54517/st3357

ARTICLE INFO

Received: 28 February 2025
Accepted: 8 May 2025
Available online: 16 June 2025

Abstract: This study examines the impact of overtourism on visitor experiences in Bali, Indonesia, and Seoul, South Korea, by analyzing online user-generated reviews from popular tourist locations. The selected sites in Seoul include Myeongdong Shopping Street, Bukchon Hanok Village, and Hongdae Street, while Bali's sites comprise Penglipuran Village, Kuta Beach, and Seminyak. A total of 1080 reviews per city (360 per location) were collected to ensure a robust and balanced dataset for comparative analysis. To extract visitor perceptions related to congestion, crowding, and other overtourism indicators, reviews were sourced using Outscraper.com, a web-based data extraction tool. A keyword-based filtering approach was applied, focusing on terms such as "noisy," "traffic," and "tourist" to identify relevant reviews that reflect visitor concerns about overcrowding and its associated effects. To maintain objectivity and representativeness, 360 reviews per location were randomly selected from the filtered dataset, ensuring a balanced sample of visitor opinions across both urban and island-based tourism destinations. By applying text-mining techniques, this study uncovers key themes and sentiment patterns that define visitor experiences in over-touristed areas. The findings offer valuable insights for tourism stakeholders seeking to mitigate the negative effects of overtourism while improving destination sustainability and visitor satisfaction.

Keywords: online reviews; over tourism; sentiment analysis; text-mining analysis; tourist congestion; visitor perceptions

1. Introduction

It has become increasingly evident in recognized global locations that overtourism represents a critical issue related to the influx of tourists beyond the capacity of the destination. Overtourism has emerged as one of the most pressing issues in global tourism, as the rapid increase in visitor numbers has led to significant environmental, social, and economic consequences for many popular destinations. The term "overtourism" refers to the excessive concentration of tourists in particular areas, surpassing the capacity of the local infrastructure and negatively affecting the quality of life for residents [1].

Overtourism also refers to the excessive concentration of visitors in certain destinations, leading to environmental degradation, reduced quality of life for residents, and compromised visitor experiences [1,2]. According to United Nations

World Tourism Organization (UNWTO) [3], global tourist arrivals reached 1.4 billion, with hotspots like Bali and Seoul experiencing unprecedented surges in tourism, straining local resources and infrastructure.

In Seoul, South Korea, popular tourist sites such as Myeongdong Shopping Street, Bukchon Hanok Village, and Hongdae Street have witnessed severe overcrowding, causing environmental degradation and reducing both resident well-being and tourist satisfaction [4]. For example, Bukchon Hanok Village, a heritage site, received approximately 6.6 million visitors in 2023—far exceeding its residential capacity—leading to concerns about noise pollution, waste accumulation, and loss of cultural authenticity [5]. Similarly, in Myeongdong, once a thriving shopping district, reports indicate that an overwhelming influx of tourists has pushed rental prices up while displacing local businesses [6].

A similar challenge is faced by Bali, Indonesia, revered for its natural beauty and cultural heritage, in Penglipuran Village, Kuta Beach, and Seminyak, where excessive tourism has adversely impacted local resources, caused severe traffic congestion, and diluted cultural identity [7]. Bali saw a record-breaking 5.3 million international arrivals in 2023, exceeding its projected target of 4.5 million. While this growth has bolstered the local economy, it has also placed immense pressure on infrastructure, particularly in tourist-heavy areas like Kuta and Seminyak, where water shortages and waste disposal issues have escalated [8]. Moreover, in Penglipuran Village, increased commercialization has led to the gradual loss of traditional Balinese customs, as tourism-driven businesses replace cultural practices once upheld by the local community [9].

Overtourism has long-term effects beyond immediate congestion, threatening the sustainability of local communities. The phenomenon exacerbates environmental issues such as increased waste production, pollution, and loss of biodiversity, while also straining local infrastructure such as ports and other public facilities and reducing accessibility for residents [4,10,11]. Seoul has begun implementing measures such as restricted access to certain heritage sites and curfews in Bukchon Hanok Village to control visitor flow, yet challenges persist in maintaining an equilibrium between economic gains and cultural preservation [5]. In contrast, Bali faces structural limitations, where rapid tourism-driven development has outpaced infrastructure growth, leading to road congestion, flooding risks, and overdevelopment of beachfront areas in Kuta and Seminyak [9]. Despite efforts to introduce sustainable tourism initiatives, such as Bali's 2024 tourism tax policy aimed at curbing mass tourism, its long-term effectiveness remains uncertain [12].

Despite the growing focus on overtourism, nuanced understandings of visitor perceptions in affected areas remain lacking. Most research on overtourism has focused on its economic and environmental consequences, yet there is limited scholarship on how tourists themselves experience and respond to congestion, crowding, and environmental degradation [11,13]. Visitor sentiment plays a crucial role in shaping destination management strategies, as negative perceptions can deter future travelers and impact long-term tourism sustainability [14,15]. Moreover, research comparing urban and rural destinations has been sparse, leading to a fragmented understanding of overtourism in different geographical and cultural contexts. For instance, while previous studies of Dodds and Butler [16] have analyzed

overtourism in European cities such as Venice and Barcelona, comparative research focusing on Asian destinations remains limited. To address these gaps, this study examines the sentiments and narratives expressed in online reviews to provide an in-depth analysis of the phenomenon of overtourism.

To uncover hidden themes in visitor feedback, this study employs advanced text-mining techniques. KH Coder, a widely used tool in text analysis, is applied to extract key patterns through co-occurrence and topic modeling, offering an innovative approach to understanding tourist perceptions [17,18]. By comparing two distinct but overtourism-affected destinations—Bali and Seoul—this study provides valuable insights into the varying dynamics of overtourism across different cultural and geographical contexts. Findings from this research aim to inform policymakers, tourism stakeholders, and local communities in developing more effective strategies for mitigating overtourism while enhancing visitor experiences and preserving destination sustainability.

2. Conceptual framework

2.1. Overtourism

Overtourism refers to the excessive influx of tourists to a destination, leading to negative impacts on the environment, local residents, and the overall visitor experience. This phenomenon has garnered significant attention in recent years, particularly in urban and rural settings. Milano et al. [14] highlight that overtourism often results in local dissatisfaction, environmental degradation, and the erosion of cultural heritage. Dodds and Butler [16] further argue that unmanaged tourism growth and ineffective governance exacerbate these adverse effects.

In urban areas, overtourism manifests through increased air pollution, traffic congestion, and the displacement of residents. [11,19] note that cities experiencing high tourist volumes often face challenges related to infrastructure strain and reduced quality of life for locals. For instance, the European Parliament's study on overtourism indicates that social impacts are more pronounced in urban destinations, affecting housing affordability and public space accessibility [20].

Conversely, rural and island destinations like Bali encounter issues such as biodiversity loss, water shortages, and the erosion of cultural traditions due to overtourism. Jenkins [21] discusses how the influx of tourists to rural areas can lead to environmental concerns, including pollution and soil degradation. Additionally, a study by ResearchGate highlights that many rural areas, due to their limited resources and infrastructure, are more susceptible to the negative impacts of overtourism, which can disrupt local ecosystems and traditional ways of life [22].

Addressing overtourism requires context-specific management strategies that consider the unique challenges of each destination. Inclusive governance models that involve local communities in decision-making processes have been suggested as effective approaches to mitigate the adverse effects of tourism. Such models can help balance economic benefits with the preservation of environmental and cultural integrity, ensuring sustainable tourism development.

2.2. Online reviews and visitor perceptions

The advent of digital platforms has transformed how tourists share their experiences, with online reviews becoming a pivotal source of information for both potential visitors and destination managers [23,24]. Tourist-generated content offers authentic insights into visitor perceptions, shedding light on satisfaction levels and areas needing improvement. A prior study emphasizes that analyzing online travel reviews provides valuable insights into tourist experiences and destination image [17,25].

Online reviews serve as a rich repository of data, reflecting the sentiments and opinions of tourists regarding various aspects of their visits. Xiang et al. [26] highlight that sentiment analysis of online reviews can reveal patterns in tourist satisfaction and dissatisfaction, offering actionable insights for destination management. By examining the language and expressions used in reviews, researchers can gauge the emotional responses of visitors, which is crucial for understanding the impact of overtourism on tourist experiences.

Moreover, online reviews can highlight specific issues related to overtourism, such as overcrowding, long wait times, and environmental degradation. For instance, a study analyzing TripAdvisor reviews found that tourists frequently mention overcrowding as a significant detractor from their experience in popular destinations [27]. These insights underscore the importance of monitoring online feedback to identify and address the negative consequences of overtourism on visitor perceptions.

2.3. Text mining and sentiment analysis

Text mining and sentiment analysis have emerged as powerful tools for extracting meaningful information from large volumes of unstructured data, such as online reviews [17]. These methodologies enable researchers to systematically analyze textual content to discern patterns, trends, and sentiments expressed by tourists. A comprehensive review in the previous research discusses the application of text analysis in tourism and hospitality, highlighting its utility in understanding tourist perceptions and experiences [28].

Sentiment analysis, in particular, focuses on determining the emotional tone behind words, allowing for the classification of opinions as positive, negative, or neutral [29]. This technique has been applied to assess tourist satisfaction and identify critical areas affecting the visitor experience [30]. For example, a study utilizing sentiment analysis on Google reviews identified key factors contributing to negative tourist sentiments, such as overcrowding and environmental concerns in the context of overtourism [29].

The integration of text mining and sentiment analysis facilitates a deeper understanding of the nuanced experiences of tourists, especially in the context of overtourism [31]. By analyzing large datasets of online reviews, researchers can uncover latent themes and sentiments that may not be immediately apparent, providing a more comprehensive view of tourist perceptions. This approach enables destination managers to make data-driven decisions aimed at enhancing tourist satisfaction and mitigating the adverse effects of overtourism.

3. Methodology

Google Reviews, one of the most frequently utilized platforms for business reviews, provides a comprehensive dataset that accurately represents the sentiments and experiences of actual customers [32]. Google Reviews are known for their accessibility, detailed content, and user authenticity, making them an ideal data source for analyzing tourist perceptions in popular destinations. **Figure 1** displays an illustration of online reviews based on Google, highlighting the user interface, ratings, and feedback provided by customers.

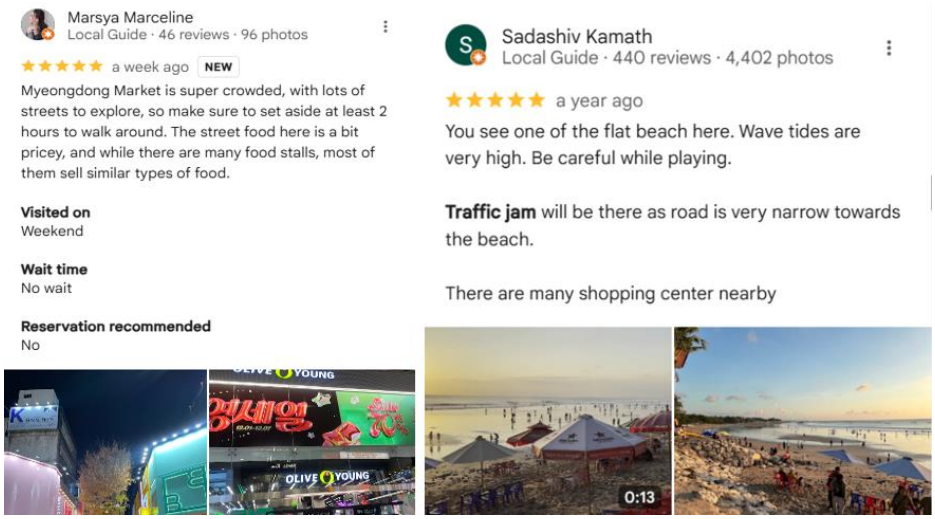

Figure 1. Review sample.

Furthermore, the review data was scraped using Outscraper, a third-party program designed for web crawling and data extraction. Outscraper is a widely used tool for extracting large volumes of web-based data, ensuring efficiency and accuracy during the data collection process. By using Outscraper, we were able to collect extensive review data for both Bali and Seoul while ensuring the data quality remained consistent. Frequently, crawling data is used to investigate human behavior, visitor satisfaction, and overall perceptions by analyzing network reviews [33]. This method guarantees a diverse and representative collection of data while simultaneously preserving manageable processing volumes. To ensure compliance with ethical standards, no personally identifiable information (PII) was extracted during the data collection process.

This research data was collected through a Google review of three specific areas identified in previous studies and reputable travel websites as high-traffic tourist locations in both Bali and Seoul. In this study, 360 reviews were collected for each area, amounting to a total of 1080 reviews per city. While the initial dataset for each location contained more than 10,000 reviews, the application of filters using keywords such as ‘noisy,’ ‘traffic,’ and ‘tourist’ reduced the sample size to hundreds or thousands per location. To ensure consistency and comparability, the authors decided to take a uniform sample of refined data, which was 360 reviews per location, providing a balanced and representative dataset for robust comparative analysis. The review content, including both textual feedback and numerical ratings, was included to ensure a comprehensive understanding of visitor experiences. The data was

collected from January 2022 to December 2024 to ensure that the research focused on post-COVID-19 tourism trends. This timeframe allows for capturing the resurgence of tourism activity and its associated impacts on overtourism dynamics. **Table 1** presents the list of selected areas and the number of reviews collected for each location.

Table 1. Review data.

	Seoul		**Bali**	
	Area	**Review number**	**Area**	**Review number**
1	Myeongdong Shopping Street	360	Penglipuran Village	360
2	Bukchon Hanok Village	360	Kuta Beach	360
3	Hongdae Street	360	Seminyak	360
	Total	1080	Total	1080

In order to conduct the analysis, we used three primary methods using KH-Coder, a powerful text mining tool. KH-Coder was selected for this study due to its ability to handle large-scale textual data and provide comprehensive analytical capabilities tailored to linguistic and statistical research [18,34]. Initially, co-occurrence analysis was conducted in order to identify any relationships or patterns among the top 100 most frequently mentioned words in the reviews. By using KH-Coder's advanced co-occurrence network visualization, we were able to identify dominant themes and key issues raised by tourists, providing a clear understanding of how terms relate to one another. Additionally, sentiment analysis was performed using customer ratings (1 to 5 stars) in order to explore how customer sentiment and experiences differed at different levels of satisfaction [35,36]. By segmenting data based on numerical attributes, such as ratings, the software was able to provide a more nuanced understanding of the relationship between word frequency and visitor satisfaction. Finally, topic modeling was used to uncover hidden themes within the reviews. Latent Dirichlet Allocation (LDA) capabilities of KH-Coder enabled the extraction of underlying topics, thereby providing deeper insights into the factors contributing to the perceptions of overtourism at various destinations [18,30]. KH-Coder was employed to ensure a robust, data-driven approach to text analysis, improving both the reliability and interpretability of the findings. Thus, **Figure 2** shows the research flow of this research.

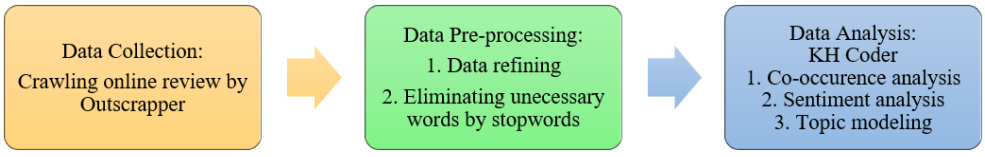

Figure 2. Research flow.

4. Result

4.1. Bali data analysis results

Using KH Coder software, this study first derived insights from the co-occurrence analysis of visitor reviews in Bali. **Figure 3** provides a visualization of co-

occurrence networks, illustrating the relationships between frequently cited keywords and revealing distinct thematic clusters. In the green cluster, cultural and traditional themes dominate, with keywords such as "village," "Balinese," and "local," indicating tourists' appreciation of Bali's cultural heritage and authenticity. Similarly, the purple cluster, containing terms like "beach," "beautiful," "sunset," and "surfing," highlights Bali's appeal as a coastal destination known for its scenic beauty and recreational activities.

However, challenges related to infrastructure and transportation are evident in the red cluster, where terms such as "traffic," "parking," and "road" indicate tourists' dissatisfaction with congestion and accessibility issues. The yellow cluster focuses on shopping and dining experiences, with keywords like "shopping," "mall," and "restaurant," showcasing the commercial aspects of tourism in Bali. Additionally, the blue cluster emphasizes nature-based attractions, featuring terms like "bamboo" and "forest," which suggest tourists' interest in serene, eco-friendly experiences.

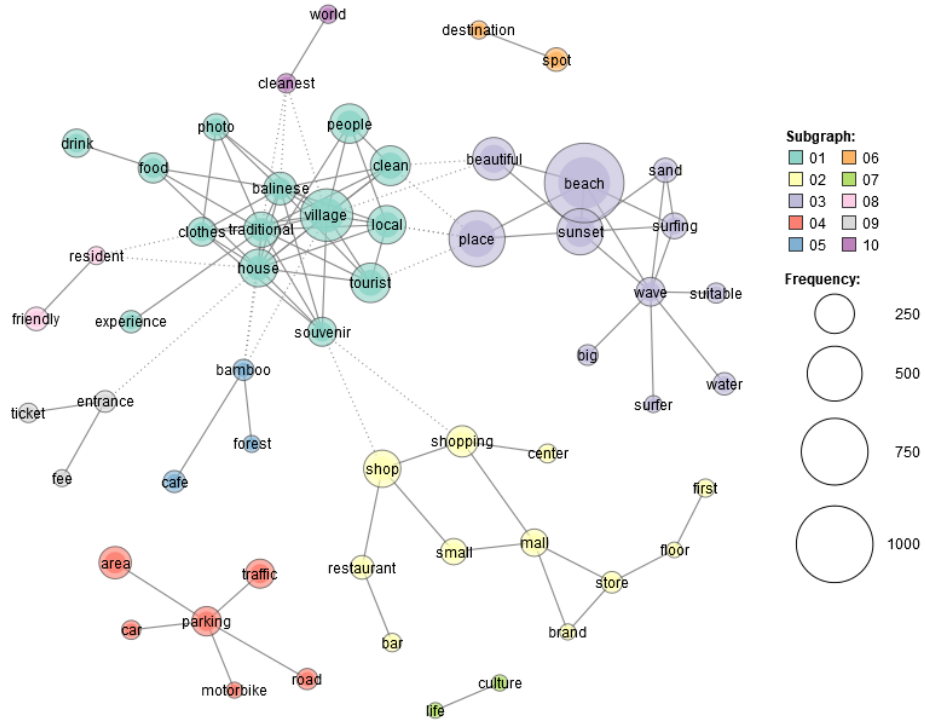

Figure 3. Co-occurrence networks analysis in Bali.

The visualization underscores the dual nature of tourism in Bali; while cultural and natural attractions remain key draws, issues such as overtourism and commercialization continue to challenge sustainable tourism development. This highlights the necessity of implementing effective tourism management strategies to balance economic growth with environmental and cultural preservation.

Moving to the next part, this network visualization illustrates how key terms are interconnected in tourist reviews from Bali, as well as their associations with different satisfaction levels, as shown in **Figure 4**. This co-occurrence rating analysis examines tourist satisfaction in Bali based on review ratings, where ratings of 5 and 4 indicate satisfaction, ratings of 1 and 2 reflect dissatisfaction, and rating 3 represents a neutral sentiment. In the figure, the red squares represent major nodes that categorize tourist

sentiment, with 1 indicating the most dissatisfied tourists, 2 reflecting dissatisfaction, 3 representing neutral opinions, 4 signifying satisfaction, and 5 denoting the most satisfied visitors.

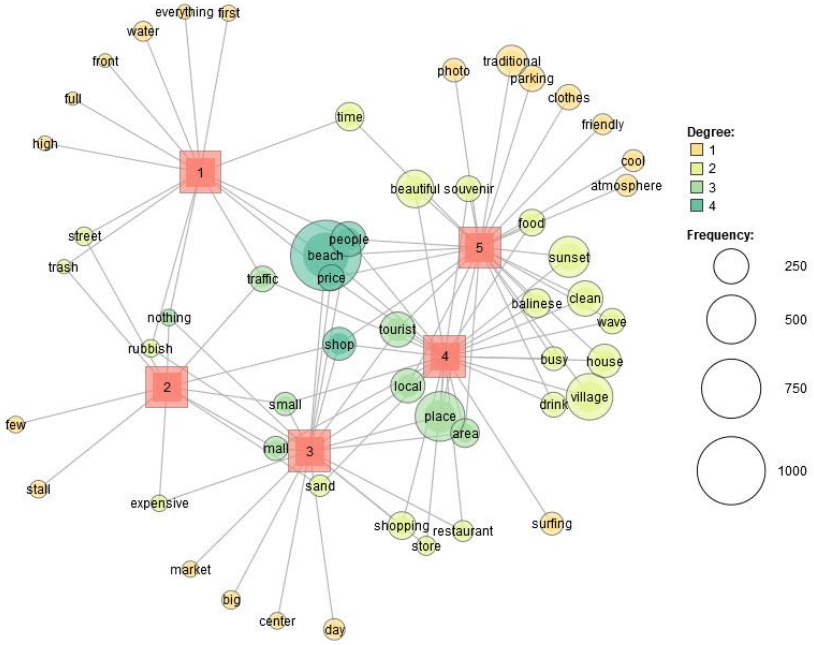

Figure 4. Co-occurrence rating network analysis in Bali.

Rating 1 primarily reflects general dissatisfaction, with words such as “trash,” “street,” and “rubbish” indicating tourists’ concerns about environmental cleanliness in Bali. Rating 2 focuses on cost-related complaints, as seen in terms like “expensive” and “few,” emphasizing tourists’ perceptions of high prices and limited value for money. Rating 3 highlights discussions related to commercial tourism, with words like “mall,” “market,” and “shop,” indicating that urban shopping and retail experiences play a significant role in visitors’ experiences. Rating 4 captures interactions between visitors and the local culture and natural environment, with words such as “beach,” “local,” “place,” and “area” representing engagement with Bali’s landscape and traditions. Finally, Rating 5 conveys appreciation for cultural and aesthetic aspects, reflected in terms such as “traditional,” “souvenir,” “friendly,” and “beautiful,” which suggest positive sentiments toward Balinese heritage and hospitality.

This visualization portrays the dual narrative of overtourism in Bali, where cultural and natural attractions continue to draw visitors, yet challenges such as environmental degradation, commercialization, and rising costs persist. These findings underscore the importance of sustainable tourism management to balance economic benefits with the preservation of environmental and cultural integrity in Bali.

In **Figure 5**, the topic modeling analysis categorizes themes in tourist reviews from Bali into ten distinct groups, each comprising ten words. This classification aligns with the co-occurrence analysis, providing a structured method for identifying latent patterns and recurring themes within a large volume of textual data. KH Coder’s topic modeling feature, which employs Latent Dirichlet Allocation (LDA), extracts these themes by analyzing word co-occurrences and probability distributions,

assigning coefficients to indicate the strength of association within the dataset. A higher coefficient signifies a more dominant theme in shaping visitor perceptions and experiences in Bali.

Topics

#1	_	#2	_	#3	_	#4	_	#5	_	#6	_	#7	_	#8	_	#9	_	#10	_
shop	0.161	price	0.123	beach	0.557	tourist	0.229	area	0.136	place	0.436	beautiful	0.222	village	0.328	beach	0.406	clean	0.212
shopping	0.139	food	0.103	surf	0.078	local	0.167	parking	0.122	people	0.147	sunset	0.208	house	0.160	sunset	0.117	photo	0.091
mall	0.105	drink	0.088	long	0.041	visitor	0.067	traffic	0.087	souvenir	0.056	time	0.118	traditional	0.143	wave	0.096	busy	0.086
small	0.096	day	0.079	bar	0.040	experience	0.065	entrance	0.057	atmosphere	0.055	surfing	0.083	balinese	0.123	water	0.067	cool	0.079
store	0.058	best	0.054	club	0.038	souvenir	0.050	road	0.048	comfortable	0.052	cafe	0.051	bamboo	0.046	dog	0.043	clothes	0.070
street	0.057	people	0.053	suitable	0.038	spot	0.046	ticket	0.047	family	0.051	first	0.040	cleanest	0.037	sea	0.035	friendly	0.068
center	0.047	stall	0.049	shop	0.030	destination	0.038	car	0.047	spot	0.031	place	0.038	temple	0.034	surfer	0.034	afternoon	0.048
few	0.045	snack	0.043	restaurant	0.029	culture	0.029	free	0.038	amazing	0.019	sand	0.034	clothes	0.027	sun	0.031	morning	0.045
big	0.042	local	0.042	hour	0.029	home	0.028	location	0.036	large	0.016	evening	0.023	forest	0.026	sand	0.026	beautiful	0.043
brand	0.030	nothing	0.039	activity	0.027	attraction	0.028	child	0.034	vendor	0.015	trash	0.017	world	0.013	atmosphere	0.018	resident	0.040

Figure 5. Topic modeling analysis in Bali.

The most prominent topic is "shopping and urban experiences" (Topic #1), with key terms such as "shop" (0.161), "shopping" (0.139), and "mall" (0.105), highlighting the significance of retail tourism in Bali. Topic #2 focuses on "cost and pricing," with words like "price" (0.123), "food" (0.103), and "drink" (0.088), reflecting frequent concerns about affordability and value for money. Topic #3, centered on "beach and surfing," includes "beach" (0.557), "surf" (0.078), and "bar" (0.040), demonstrating the appeal of Bali's coastal activities and nightlife.

Tourist engagement with local communities and culture is represented in Topic #4, which features "tourist" (0.229), "local" (0.167), and "souvenir" (0.050), indicating that cultural interactions are integral to visitor experiences. Similarly, Topic #8 reinforces cultural heritage, with words like "village" (0.328), "traditional" (0.143), and "Balinese" (0.123), emphasizing an appreciation for Bali's authenticity and historical richness.

Challenges related to infrastructure emerge in Topic #5, where "traffic" (0.087), "parking" (0.126), and "road" (0.048) highlight visitor frustrations with congestion and mobility issues. Conversely, Topic #7 captures positive perceptions of "scenic beauty," with key terms like "beautiful" (0.222), "sunset" (0.208), and "time" (0.118), reflecting admiration for Bali's natural landscapes. Topic #10 also highlights aesthetic and atmospheric factors, with "clean" (0.212), "photo" (0.091), and "cool" (0.079), indicating that visual appeal significantly influences visitor satisfaction.

Topic #6 emphasizes "atmosphere and relaxation," with terms such as "place" (0.436), "people" (0.147), and "souvenir" (0.108), suggesting that comfort and ambiance contribute to positive tourist experiences. Meanwhile, Topic #9 touches on "coastal and water-related experiences," including "beach" (0.406), "wave" (0.036), and "sea" (0.034), reinforcing the central role of Bali's marine environment in tourism appeal.

These findings highlight the dual narrative of tourism in Bali, where cultural and natural attractions continue to draw visitors, yet concerns such as pricing, infrastructure strain, and commercialization persist. This underscores the need for sustainable tourism strategies that enhance positive visitor experiences while mitigating the negative effects of overtourism.

4.2. Seoul data analysis results

Based on the clustering of tourist reviews, **Figure 6** visualizes distinct themes that emerge from visitors' experiences in Seoul. The blue cluster prominently features keywords such as "shop," "food," "restaurant," and "market," highlighting urban tourism activities centered around nightlife, retail experiences, and dining options. This suggests that shopping and food culture are central to visitors' experiences in Seoul's commercial districts.

In contrast, the orange cluster emphasizes cultural heritage and residential settings, with key terms such as "village," "traditional," "architecture," and "Korean." These words indicate that many tourists appreciate Seoul's historical and cultural identity. However, the presence of terms like "noisy" and "residential" also suggests dissatisfaction among visitors who find certain urbanized areas overcrowded or disruptive.

The green cluster captures aesthetic and sensory elements, with words such as "beauty" and "product," highlighting the visual and experiential appeal of specific locations. Meanwhile, the red cluster focuses on pricing concerns, where dominant terms like "price" and "expensive" reflect tourists' dissatisfaction with the cost of goods and services in Seoul.

Additionally, smaller clusters provide further insights into visitor perceptions. The teal and yellow clusters emphasize aspects such as "atmosphere," "experience," and "fashion," suggesting that Seoul offers a vibrant and dynamic environment that appeals to visitors seeking diverse cultural and lifestyle experiences.

Overall, this visualization underscores the dual nature of urban tourism in Seoul, where the appeal of cultural heritage, shopping, and vibrant atmospheres coexists with challenges such as overcrowding and high costs. These findings highlight the need for balanced tourism strategies that enhance visitor satisfaction while addressing concerns related to congestion and affordability.

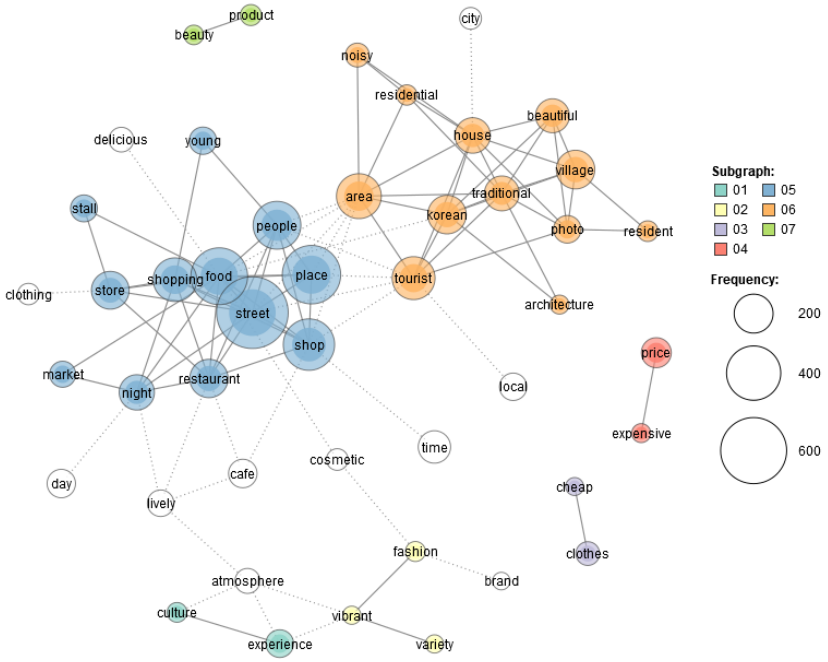

Figure 6. Co-occurrence network analysis in Seoul.

In **Figure 7**, the co-occurrence network visualization presents key themes derived from tourist reviews in Seoul, categorized based on review ratings to reflect varying levels of satisfaction. This analysis assigns ratings of 5 and 4 to positive experiences, ratings of 1 and 2 to dissatisfaction, and a rating of 3 to neutral opinions. The red squares represent major nodes that indicate the most frequently occurring themes across different rating levels.

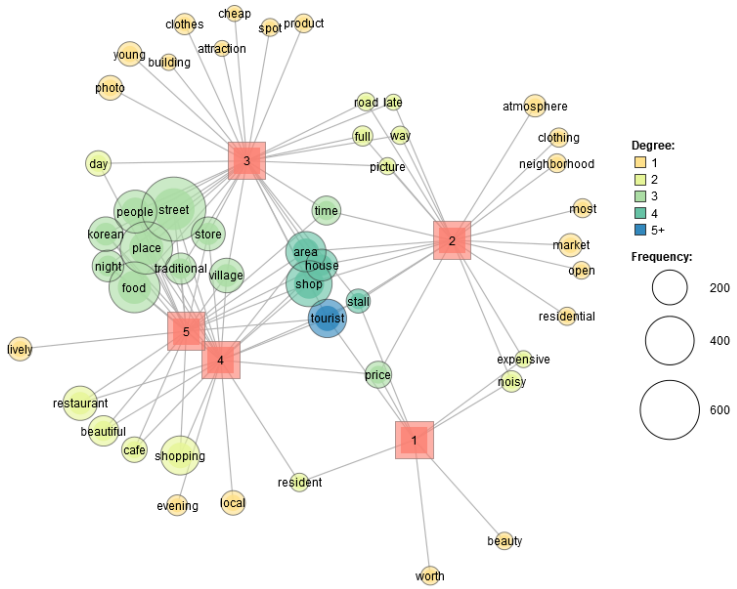

Figure 7. Co-occurrence rating network analysis in Seoul.

Rating 1 primarily captures dissatisfaction with pricing concerns, with words such as “price,” “expensive,” and “worth” highlighting tourist frustration over high costs. Similarly, Rating 2 focuses on residential and environmental concerns, where terms like “neighborhood,” “residential,” “atmosphere,” and “noisy” suggest dissatisfaction with tourism’s impact on local communities and quality of life.

Rating 3 reflects a neutral stance, predominantly centered around shopping and attractions. Keywords such as “cheap,” “clothes,” “spot,” and “product” indicate that tourists acknowledge affordability and variety but do not strongly associate them with either positive or negative experiences.

Rating 4 highlights themes related to accessibility and cultural interaction, with words like “area,” “house,” “shop,” and “tourist” reflecting visitors’ engagement with local spaces and heritage. Accessibility and authenticity are emphasized as contributing factors to satisfaction at this level.

Rating 5 conveys strong appreciation for cultural heritage and local experiences. Terms such as “village,” “traditional,” “Korean,” “food,” “people,” “street,” and “store” suggest that tourists highly value social interactions, cultural immersion, and authentic culinary experiences as key highlights of their trip.

Positive themes are further emphasized in clusters containing words such as “shopping,” “beautiful,” “restaurant,” and “cafe,” reinforcing the appeal of Seoul’s dynamic urban environment. Overall, the visualization illustrates the dual nature of tourism in Seoul, where cultural richness, food experiences, and vibrant atmospheres coexist with challenges like high costs, noise, and the impact on residential areas.

These findings highlight the need for balanced tourism management strategies to enhance visitor satisfaction while mitigating negative effects on local communities.

In **Figure 8**, the topic modeling analysis categorizes key themes from tourist reviews in Seoul into ten distinct topics, each comprising frequently mentioned words. This classification aligns with the co-occurrence analysis, providing structured insights into the diverse aspects of urban tourism. KH Coder's topic modeling feature, which employs Latent Dirichlet Allocation (LDA), extracts these themes by analyzing word co-occurrences and probability distributions, assigning coefficients that indicate their significance in shaping visitor perceptions.

Cultural heritage and local identity emerge as dominant themes in Topic #4, where words such as "Korean" (0.188), "village" (0.178), "traditional" (0.165), and "architecture" (0.046) highlight tourists' appreciation for historical landmarks and authentic local experiences. Similarly, scenic appeal and unique experiences are reflected in Topic #8, which includes terms such as "beautiful" (0.152), "photo" (0.128), and "experience" (0.120), indicating that visual aesthetics and immersive travel moments are significant factors in visitor satisfaction.

Urban tourism and social engagement are central to Topic #2 and Topic #7, where "street" (0.427), "food" (0.186), "fashion" (0.083), and "restaurant" (0.130) reflect Seoul's vibrant lifestyle, emphasizing shopping, dining, and nightlife experiences. In addition, Topic #6, with words like "store" (0.223), "price" (0.134), and "restaurant" (0.031), showcases tourists' engagement with commercial districts and their concerns about affordability.

However, pricing and economic concerns are notable in Topic #9, where "food" (0.305), "expensive" (0.042), and "market" (0.061) reflect dissatisfaction with the cost of goods and services. Additionally, Topic #1 captures residential strain, with terms like "area" (0.334), "tourist" (0.257), "noisy" (0.087), and "residential" (0.060), indicating challenges related to overtourism in local neighborhoods.

Social interactions and cultural engagement appear in Topic #10, where words like "people" (0.384), "young" (0.118), and "street" (0.048) highlight the significance of dynamic, community-driven experiences. Meanwhile, aspects of accessibility and transportation are reflected in Topic #3, which includes "place" (0.510), "shop" (0.370), and "tourist" (0.362), reinforcing the importance of well-connected locations in shaping visitor mobility and convenience.

These findings reinforce the dual nature of tourism in Seoul, where cultural richness, urban vibrancy, and social engagement continue to attract visitors, yet challenges such as high costs, noise pollution, and residential strain persist. This highlights the need for balanced tourism management strategies that enhance visitor experiences while addressing overtourism-related concerns.

Topics

#1	_	#2	_	#3	_	#4	_	#5	_	#6	_	#7	_	#8	_	#9	_	#10	_
area	0.334	street	0.427	place	0.510	korean	0.188	street	0.315	store	0.223	shop	0.330	beautiful	0.152	food	0.305	people	0.384
tourist	0.257	food	0.186	fun	0.075	village	0.178	shopping	0.283	price	0.134	restaurant	0.197	photo	0.128	night	0.151	young	0.118
noisy	0.087	fashion	0.053	shop	0.071	house	0.167	evening	0.070	time	0.130	lively	0.089	experience	0.120	stall	0.104	full	0.080
residential	0.060	performance	0.051	tourist	0.054	traditional	0.165	district	0.057	clothes	0.072	street	0.059	worth	0.070	day	0.101	cafe	0.057
local	0.056	various	0.040	cosmetic	0.054	city	0.056	station	0.050	product	0.051	open	0.055	spot	0.062	market	0.097	picture	0.056
attraction	0.044	variety	0.037	everything	0.037	culture	0.046	atmosphere	0.039	clothing	0.046	plenty	0.044	local	0.057	delicious	0.070	street	0.048
foreigner	0.044	bustling	0.028	best	0.031	architecture	0.045	brand	0.038	snack	0.046	souvenir	0.034	way	0.056	expensive	0.040	place	0.043
building	0.039	interesting	0.028	bar	0.026	alley	0.036	shop	0.027	best	0.041	cafe	0.033	place	0.045	beauty	0.030	busy	0.041
day	0.013	different	0.027	time	0.021	small	0.033	beautiful	0.023	cheap	0.040	hour	0.027	resident	0.042	late	0.030	easy	0.036
atmosphere	0.012	beauty	0.023	beautiful	0.018	neighborhood	0.030	souvenir	0.015	restaurant	0.031	unique	0.025	road	0.036	atmosphere	0.015	night	0.028

Figure 8. Co-occurrence rating network analysis in Seoul.

5. Discussion

The findings of this study confirm that overtourism manifests in multifaceted ways, differing based on destination characteristics, as observed in Seoul and Bali. Consistent with previous studies by Boháč and Drápela [14] and Milano et al. [22], overtourism-related challenges include cultural preservation, environmental degradation, strain on infrastructure, and economic concerns. In Seoul, the overtourism phenomenon is primarily characterized by urban congestion, infrastructure strain, and environmental degradation. Popular sites such as Bukchon Hanok Village and Myeongdong Shopping Street exhibit severe overcrowding, with visitors reporting negative impacts on both their experience and the local community. Specifically, the presence of high traffic congestion, noise pollution, and the commercialization of cultural sites significantly diminish the quality of the visitor experience. These urban issues align with findings from prior studies, which highlight the economic and social pressures faced by densely populated cities with high tourist inflows [16].

In contrast, Bali's overtourism challenges are rooted in environmental degradation, loss of cultural authenticity, and infrastructural limitations. While the island remains a major attraction for its scenic beaches, cultural heritage, and natural beauty, sites such as Kuta Beach and Seminyak are overwhelmed by tourist traffic, leading to severe congestion, pollution, and resource depletion. Unlike Seoul, Bali's challenges are more related to environmental sustainability, including water shortages and waste management, compounded by the commercialization of cultural sites such as Penglipuran Village. Tourists express concerns about the dilution of traditional Balinese customs and practices, underscoring the need for more effective management of cultural heritage in tourist-heavy areas [37].

Despite these differences, both Seoul and Bali share common themes in visitor perceptions. Tourists in both cities express appreciation for cultural heritage and natural beauty, but they also report dissatisfaction with overcrowding, environmental degradation, and rising costs. This dual nature of tourism, where visitors enjoy the benefits of the destinations but also face significant drawbacks, mirrors findings in other popular destinations suffering from overtourism, such as Barcelona and Venice [16].

The study also highlights the dual nature of tourist experiences, reinforcing prior research on tourism's paradoxical effects [38]. While visitors express positive sentiments toward cultural heritage, scenic beauty, and vibrant atmospheres, they simultaneously report dissatisfaction with infrastructure strain, high costs, and overcrowding. These findings are in line with studies suggesting that while tourism fosters economic benefits, it also exacerbates socioeconomic inequalities and environmental burdens in popular destinations [39].

The methodological approach of text mining and topic modeling provided nuanced insights into visitor perceptions, revealing dominant themes such as shopping, dining, and recreational activities, alongside concerns about environmental sustainability and affordability. This aligns with previous research indicating that data-driven sentiment analysis can effectively capture shifts in tourist expectations and satisfaction levels [27]. Furthermore, by comparing an urban destination (Seoul) and

a mixed urban-rural destination (Bali), this study addresses research gaps in the comparative analysis of overtourism across different spatial contexts, as identified by [13].

These findings underscore the urgency of implementing destination-specific tourism management strategies. For Seoul, urban planning initiatives focusing on mitigating congestion, regulating commercial tourism zones, and preserving cultural integrity may enhance sustainability. For Bali, stronger environmental policies, improved waste management, and cultural conservation programs could help alleviate the adverse effects of overtourism. Future studies could expand on this research by incorporating longitudinal data analysis to track sentiment changes over time and integrating demographic insights to better understand visitor behavior patterns.

6. Implications

The findings of this study provide critical insights for policymakers, tourism authorities, and business stakeholders aiming to mitigate the negative consequences of overtourism while maximizing its economic and cultural benefits. For Seoul, urban planning strategies should focus on regulating commercial tourism zones, implementing congestion management policies, and integrating technology-driven visitor flow control systems, similar to successful models in other global cities [40]. Additionally, destination marketing organizations should promote alternative attractions beyond high-traffic areas such as Bukchon Hanok Village and Hongdae to distribute visitor density more evenly.

For Bali, tourism management should prioritize environmental sustainability through improved waste management, stronger enforcement of conservation policies, and stricter zoning regulations to preserve cultural authenticity in areas like Penglipuran Village [37]. Local businesses and hospitality operators can benefit from these findings by adapting pricing models, improving service quality, and fostering more community-based tourism initiatives that encourage cultural immersion without exacerbating residential strain [38]. Both destinations should consider leveraging real-time data analytics to monitor tourist sentiment dynamically, allowing authorities to implement adaptive tourism policies based on emerging visitor concerns.

From an academic perspective, this study contributes to the growing body of research on overtourism by applying advanced text mining and sentiment analysis techniques to compare visitor perceptions across different destination types. By integrating co-occurrence analysis, topic modeling, and sentiment classification, this research extends previous studies Capocchi et al. [13] and Milano et al. [14]. by demonstrating the effectiveness of big data methodologies in assessing tourist sentiment at scale.

Additionally, this study bridges a research gap by comparing overtourism effects in an urban setting (Seoul) versus a mixed urban-rural destination (Bali), offering valuable insights into how destination characteristics shape tourism experiences and challenges. Future research should explore cross-regional comparisons with other major tourism hotspots to generalize findings further and examine how socioeconomic and demographic factors influence visitor perceptions of overtourism [16]. Longitudinal studies incorporating evolving tourist sentiments over time would

provide a deeper understanding of how tourism patterns shift post-pandemic and in response to destination management interventions.

By emphasizing the application of big data analytics in tourism research, this study encourages future scholars to adopt computational techniques such as machine learning-based sentiment classification and network analysis to refine the measurement of tourism impacts. Expanding these methodologies to include multimodal data sources, such as social media images and geospatial analytics, would further enhance the precision of overtourism assessments.

7. Conclusion

This study provides significant insights into overtourism by analyzing visitor perceptions in two distinctly different tourism destinations—Seoul and Bali—through the lens of online reviews. The findings contribute to the growing body of research on overtourism by employing advanced text-mining techniques, such as co-occurrence analysis and topic modeling, to uncover patterns in tourist sentiment and experience. These methods offer a novel approach to understanding the nuanced effects of overtourism, revealing not only the environmental and infrastructural challenges but also the deeper, often overlooked, emotional responses of tourists.

In comparison to previous studies, such as those by [14,16], which primarily focus on economic and environmental aspects of overtourism, this research highlights the significant role of visitor perceptions and the psychological impacts of overcrowding. It aligns with findings from research in European destinations but extends the understanding by incorporating Asian urban and rural perspectives, filling a gap in the comparative analysis of overtourism in different geographical and cultural contexts.

The implications of this study are far-reaching. For urban destinations like Seoul, there is a clear need for strategies that regulate tourism in high-density areas and mitigate issues such as congestion and cultural degradation. In contrast, Bali's challenges are tied more to environmental sustainability and the preservation of cultural heritage, which requires stricter policies to manage visitor numbers and maintain local traditions.

Furthermore, the study underscores the importance of incorporating text mining and sentiment analysis into tourism management practices. By analyzing large volumes of unstructured data, this research provides a more granular understanding of visitor sentiment, which can inform real-time decision-making for destination managers. This approach not only enhances the accuracy of visitor experience assessments but also contributes to the development of sustainable tourism strategies that are informed by actual tourist feedback.

Finally, this research opens avenues for future studies to expand on these findings by exploring other tourism hotspots, incorporating longitudinal data to track shifts in tourist sentiment over time, and integrating demographic factors to understand how different visitor groups perceive overtourism. By extending the use of big data analytics in tourism research, this study encourages further exploration of innovative methodologies to refine our understanding of overtourism and its long-term impacts.

8. Limitation and future research

8.1. Limitation

The results of this study should be interpreted within the context of several limitations. A first limitation of this analysis is that it relies solely on online reviews, which may fail to reflect the perspectives of all stakeholders, such as local residents and policymakers, which could result in bias in the data collected. However, it would be more original if surveys could be done with the stakeholders with real questions prepared by the author(s) to get results that are more realistic. Secondly, the study is limited to two destinations, Seoul and Bali, thereby limiting the generalizability of the findings to other urban and rural contexts throughout the world. Thirdly, the cross-sectional nature of the data prevents us from gaining a full understanding of overtourism's long-term dynamics and trends. Moreover, as a result of relying on keyword filtering during the data collection process, relevant reviews may have been excluded from the analysis. Further, although text mining techniques such as topic modeling and co-occurrence analysis provide valuable insights, they do not take into account nuances in a given context that require qualitative or mixed-methods approaches to be explored in depth.

8.2. Future research

Consequently, future research should incorporate broader, more diverse data sources, such as interviews, surveys, and focus groups, in order to capture the perspectives of different stakeholders. These stakeholders include residents, business owners, and policymakers. A broader geographic scope that includes destinations across a variety of cultural, economic, and environmental contexts would enhance the generalizability of the results. Moreover, longitudinal studies may be useful for examining the long-term effects of over tourism and assessing the impact of interventions implemented over time. The use of mixed-methods approaches can also provide a more comprehensive understanding of overtourism dynamics by combining quantitative insights gained through text mining with qualitative insights. Further research could explore new technologies, such as artificial intelligence and real-time monitoring systems, as a means of reducing overtourism and promoting sustainable tourism. A better understanding of overtourism and its possible solutions will result from these directions.

Author contributions: Conceptualization, RSB and MR; methodology, NDH; software, NDH; validation, SFB, SM and DM; formal analysis, SL and RI; investigation, MSN; resources, NDH; data curation, MR and SFB; writing—original draft preparation, NDH; writing—review and editing, RSB; visualization, RSB; supervision, MR; project administration, MR; funding acquisition, SFB. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Agyeiwaah E. Over-tourism and sustainable consumption of resources through sharing: the role of government. International

Journal of Tourism Cities. 2019; 6(1): 99-116. doi: 10.1108/ijtc-06-2019-0078
2. Perkumienė D, Pranskūnienė R. Overtourism: Between the Right to Travel and Residents' Rights. Sustainability. 2019; 11(7): 2138. doi: 10.3390/su11072138
3. UNWTO. World Tourism Barometer. Available online: https://www.unwto.org/un-tourism-world-tourism-barometer-data (accessed on 23 January 2025).
4. The Korea Times. Seoul Faces Challenges to Become Sustainable City. The Korea Times; 2024.
5. Chen H, Rhee S. A Seoul neighborhood is so crowded with tourists that the government is taking over. CNN; 2024.
6. Hong S. "Timewalk Myeongdong" sale resumes due to revitalization of commercial district. Available online: https://www.mk.co.kr/en/economy/11016136 (accessed on 23 January 2025).
7. Responsible Travel. Overtourism in Bali. Available online: https://www.responsibletravel.com/holidays/bali/travel-guide/overtourism (accessed on 23 January 2025).
8. Antara. Bali Foreign Tourist Arrivals Recover at 88 Percent: Association. Antara; 2023.
9. The Converstation. Bali gives a snapshot of what overtourism looks like in the developing world. Avilable online: https://theconversation.com/bali-gives-a-snapshot-of-what-overtourism-looks-like-in-the-developing-world-235550 (accessed on 10 January 2025).
10. The Star. Seoul Faces Challenges to Become Sustainable City. The Star; 2024.
11. Lousada S, Castanho RA. The Role of Ports in Tourism: Porto Santo Harbour. Water. 2022; 14(19): 3176. doi: 10.3390/w14193176
12. Wilson N. Bali to Italy: 39 destinations cracking down on overtourism. The Independent; 2025.
13. Capocchi A, Vallone C, Amaduzzi A, et al. Is 'overtourism' a new issue in tourism development or just a new term for an already known phenomenon?. Current Issues in Tourism. 2019; 23(18): 2235-2239. doi: 10.1080/13683500.2019.1638353
14. Milano C, Novelli M, Cheer JM. Overtourism and degrowth: a social movements perspective. Journal of Sustainable Tourism. 2019; 27(12): 1857-1875. doi: 10.1080/09669582.2019.1650054
15. Gössling S, Scott D, Hall CM. Pandemics, tourism and global change: a rapid assessment of COVID-19. Journal of Sustainable Tourism. 2020; 29(1): 1-20. doi: 10.1080/09669582.2020.1758708
16. Dodds R, Butler R. The phenomena of overtourism: a review. International Journal of Tourism Cities. 2019; 5(4): 519-528. doi: 10.1108/ijtc-06-2019-0090
17. Lim VC, Handani ND. Study of Tourist Satisfaction At Busan'S Temples: A Quantitative Text Mining Approach. International Journal of Business and Social Science. 2024; 25(3): 930-946. doi: 10.33736/ijbs.8516.2024
18. Noor MS, Handani ND. Exploring factors and customer perceptions of airport services: A quantitative textual analysis. Smart Tourism. 2024; 5(1): 2485. doi: 10.54517/st.v5i1.2485
19. Gössling S, Scott D. The decarbonisation impasse: global tourism leaders' views on climate change mitigation. Journal of Sustainable Tourism. 2018; 26(12): 2071-2086. doi: 10.1080/09669582.2018.1529770
20. Peeters P, Gössling S, Klijs J, et al. Overtourism: Impact and Possible Policy Responses. Brussels, Belgium; 2018.
21. Jenkins CL. Tourism and political change. Journal of Policy Research in Tourism, Leisure and Events. 2012; 4(3): 382-383. doi: 10.1080/19407963.2011.613231
22. Boháč A, Drápela E. Overtourism Hotspots: Both a Threat and Opportunity for Rural Tourism. European Countryside. 2022; 14(1): 157-179. doi: 10.2478/euco-2022-0009
23. Noor MS, Handani ND, Kim HS. Exploring user generated content for beach resorts in Cox's Bazar, Bangladesh: A pre- and post-pandemic analysis. Environment and Social Psychology. 2023; 8(2). doi: 10.54517/esp.v8i2.1771
24. Zhong Y, Williady A, Handani ND, et al. Big Data Insights into Coastal Tourism: Analyzing Customer Satisfaction at Egyptian Red Sea Dive Resorts. Tourism and Hospitality. 2024; 5(4): 996-1011. doi: 10.3390/tourhosp5040056
25. Handani ND, Riswanto AL, Kim HS. A Study of Inbound Travelers Experience and Satisfaction at Quarantine Hotels in Indonesia during the COVID-19 Pandemic. Information. 2022; 13(5): 254. doi: 10.3390/info13050254
26. Xiang Z, Schwartz Z, Gerdes JH, et al. What can big data and text analytics tell us about hotel guest experience and satisfaction?. International Journal of Hospitality Management. 2015; 44: 120-130. doi: 10.1016/j.ijhm.2014.10.013
27. Yu J, Egger R. Tourist Experiences at Overcrowded Attractions: A Text Analytics Approach. In Information and Communication Technologies in Tourism 2021, Proceedings of the ENTER 2021 eTourism Conference. Springer International Publishing; 2021. doi: 10.1007/978-3-030-65785-7.
28. Bi JW, Zhu XE, Han TY. Text Analysis in Tourism and Hospitality: A Comprehensive Review. Journal of Travel Research.

2024; 63(8): 1847-1869. doi: 10.1177/00472875241247318
29. Handani ND, Williady A, Kim HS. An Analysis of Customer Textual Reviews and Satisfaction at Luxury Hotels in Singapore's Marina Bay Area (SG-Clean-Certified Hotels). Sustainability. 2022; 14(15): 9382. doi: 10.3390/su14159382
30. Kwon HJ, Ban HJ, Jun JK, et al. Topic Modeling and Sentiment Analysis of Online Review for Airlines. Information. 2021; 12(2): 78. doi: 10.3390/info12020078
31. Bachtiar FA, Paulina W, Rusydi AN. Text Mining for Aspect Based Sentiment Analysis on Customer Review : A Case Study in the Hotel Industry. In: Proceedings of the 5th International Workshop on Innovations in Information and Communication Science and Technology; 2020.
32. Noerhartati E, Handani ND, Dhaneswara N. Exploring The Influence of Online Reviews on Sales. JWEE; 2023. doi: 10.28934/jwee23.34.pp48-74
33. Wei S, Kim HS. Online Customer Reviews and Satisfaction with an Upscale Hotel: A Case Study of Atlantis, The Palm in Dubai. Information. 2022; 13(3): 150. doi: 10.3390/info13030150
34. Jurkus E, Povilanskas R, Taminskas J. Current Trends and Issues in Research on Biodiversity Conservation and Tourism Sustainability. Sustainability. 2022; 14(6): 3342. doi: 10.3390/su14063342
35. Lee M, Lee S, Koh Y. Multisensory experience for enhancing hotel guest experience. International Journal of Contemporary Hospitality Management. 2019; 31(11): 4313-4337. doi: 10.1108/ijchm-03-2018-0263
36. Raharjo JH, Budiharseno RS, Quy TM. Hitting the Right Notes: Exploring the Relationship between YouTube Music Videos and Digital User Engagement. Global Business and Finance Review. 2024; 29(1): 44-56. doi: 10.17549/gbfr.2024.29.1.44.
37. Chin WL, Haddock-Fraser J, Hampton MP. Destination competitiveness: evidence from Bali. Current Issues in Tourism. 2015; 20(12): 1265-1289. doi: 10.1080/13683500.2015.1111315
38. Gössling S, Zeiss H, Hall CM, et al. A cross-country comparison of accommodation manager perspectives on online review manipulation. Current Issues in Tourism. 2018; 22(14): 1744-1763. doi: 10.1080/13683500.2018.1455171
39. Sujarwo W. Sasak Traditional Villages: A Potential Tourism and Portrait of Conservation Efforts for Culture and Plants. Jurnal Masyarakat dan Budaya; 2019.
40. Socratous M, Thrasyvoulou G, Stylianou C. Technology-driven sustainable tourism destinations: perspectives from key policymakers. Worldwide Hospitality and Tourism Themes. 2025; 17(1): 77-90. doi: 10.1108/whatt-12-2024-0298