Asia Business & Service Innovation 2(1), 20-43. (2025)
https://doi.org//10.23405/ABSI.2025.2.1.02

Information available at the Journal of Asia Business and Service Innovation (http://www.absi.kr/)

Asia Business and Service Innovation

ISSN: 3092-0558 (Print) | eISSN: 3092-1112 (Online)

Big Data Exploration of India's Heritage Tourism: Understanding Visitor Experiences

Sadaf Akhtar[1], Angellie Williady[1], Handani Narariya Dita[2*]

[1] *Department of Global Business, Kyungsung University, Busan, Republic of Korea*
[2] *Wijaya Kusma University, Surabaya, Indonesia*
*Corresponding author: dita_fe@uwks.ac.id

Abstract

India's rich cultural heritage and remarkable architectural landmarks establish it as one of the leading destinations for heritage tourism, showcasing iconic sites such as the City Palace, Amber Fort in Jaipur, and Gwalior Fort. In the era of digital media, online tourist reviews have emerged as a crucial source of credible viewpoints into visitor experiences. This study examines forts and palaces, with a particular focus on Agra Fort, a prominent site frequently featured in the Incredible India Campaign and India Tourism Statistical Reports. A dataset of 1,309 recent online reviews from Google Maps was collected using the Instant Data Scraper tool to capture current visitor perspectives. The reviews were analyzed using KH Coder 3 software, allowing detailed text analysis including word frequency and co-occurrence assessments. Co-occurrence analysis revealed five primary clusters: "History & Culture," "General Infrastructure," "UNESCO," "Tourist Infrastructure," and "Heritage Experience," derived from the top 100 most mentioned words. Sentiment analysis further examined the emotional tone, highlighting the balance between positive and negative attitudes. Results indicate that visitors predominantly express positive sentiments toward India's heritage sites, with keywords like Taj Mahal, Agra Fort, and beautiful frequently associated with admiration for the architectural beauty, historical significance, and aesthetic appeal. Negative sentiments were mainly linked to terms such as ticket, maintain, and worth, pointing to issues in site management and visitor services rather than dissatisfaction with the heritage itself. Overall, the findings suggest that India's heritage attractions are highly appreciated by tourist's, underscoring their cultural and experiential value while identifying opportunities for operational improvements to enhance visitor satisfaction.

Keywords**:** Big Data; India's Heritage; Tourism; Online Reviews; Sentiment Analysis.

Received Feb 17, 2025 · Revised Apr 12, 2025 · Accepted May 26, 2025

Asia Business & Service Innovation 2(1), 20-43. (2025)
https://doi.org//10.23405/ABSI.2025.2.1.02

1. Introduction

International overnight tourist arrivals rose 5% in the first half of 2025, reaching approximately 690 million—33 million more than in 2024 (UN Tourism, 2025). This revival highlights the resilience of global tourism and the continued appeal of destinations rich in cultural heritage and unique experiences (OECD, 2025). Tourism remains one of the fastest growing economic sectors and an important driver of economic growth and development, with a significant impact on trade, job creation, investment, infrastructure development, and social inclusion (Tourism a driver for shared prosperity, UN Tourism, 2025).

India's tourism sector witnessed significant growth in 2025, as Foreign Tourist Arrivals (FTAs) reached 9.95 million, reflecting a 4.5% rise compared to the previous year. This growth reflects the country's appeal as a premier destination for cultural, spiritual, and adventure tourism. In the year 2023, India registered 9.52 million FTAs i.e., attained around 87% recovery of the previous years with a growth of 47.9% over 2022 (India Tourism Data Compendium , 2024). The top 15 countries, like the United States (18.13%), Bangladesh (17.59%), United Kingdom (10.28%), collectively contributed (77.07%) of total FTAs in 2024, slightly lower than their combined share in 2023 (78.30%) but higher than in 2019 (76.31%), suggesting a continued concentration of inbound traffic from a few key markets (India Tourism Data Compendium, Ministry of Tourism, India, 2025).

India's Incredible India campaign prominently highlights the country's rich cultural and historical heritage, which serves as a significant factor in the tourism industry (Ministry of Tourism, Government of India, 2002). The country's historic forts and palaces are highlighted by this initiative, serving as permanent archives of socio-political and cultural histories in addition to showcasing the sophistication of indigenous architectural practices. Such heritage sites are the centre of heritage tourism, drawing people from all over the world who want to learn more about India's history and material culture (Incredible India, 2025). Among a wide range of heritage sites, Agra Fort stands out as a significant attraction. Based on the India Tourism Data Compendium, Ministry of Tourism, India. (2025) report, Agra Fort consistently ranks among the top-visited Forts in India, drawing millions of domestic and international tourists annually. It serves as an optimal case study

for investigating the experiences and behaviours of tourists in heritage tourism due to its historical significance, architectural elegance, and affiliation with the Mughal dynasty.

Previous studies on heritage tourism have thoroughly studied the connection between cultural heritage sites and tourist experiences, highlighting the role of visitors motivations, perceptions, and satisfaction in shaping destination image and promoting sustainable tourism development (Timothy & Boyd, 2006). Research has also underscored the growing role of forts and palaces in representing India's architectural and historical legacy within the heritage tourism framework (Sharma & Nayak, 2018). Several empirical studies have analyzed tourist behavior at prominent Indian heritage sites such as the Taj Mahal, Fatehpur Sikri, and Jaipur's City Palace, revealing that factors such as cultural authenticity, heritage interpretation, and site management significantly influence visitor satisfaction (Baniya & Thapa, 2024; Riswanto et al., 2023)**.** However, there remains a shortage of qualitative research focusing specifically on Agra Fort, despite its high visitor volume and UNESCO World Heritage status. This gap highlights the need for a refined exploration of tourist's experiential perceptions and emotional engagement at Agra Fort, particularly through the lens of digital data such as online reviews, to generate grounded insights into visitor satisfaction and heritage interpretation.

This research investigates the qualitative dimensions of tourist behavior at Agra Fort through the systematic analysis of online reviews obtained from Google Maps. Adopting the Grounded Theory approach enables an inductive exploration of visitor's narratives to identify emergent themes and interpret underlying behavioral patterns. This method allows the study to move beyond surface-level data, uncovering the nuanced perceptions, motivations, and experiential satisfaction expressed by tourist's in their own words (Glaser & Strauss, 1967a; Papathanassis & Knolle, 2011)**.** Through this qualitative framework, the study seeks to construct a conceptual model of tourist behavior that reflects the affective and cognitive dimensions of heritage tourism experiences at Agra Fort. The affective and cognitive dimensions refer to two separate aspects of a person's experience or attitude: affective relates to emotions, feelings, and values, while cognitive relates to thought, knowledge, and belief.

2. Literature Review

2.1. Indian Heritage

In 2024, India recorded 2,948.19 million domestic tourist visits, reflecting a 17.51% rise compared to 2,508.82 million visits in 2023. Foreign tourist visits (FTVs) reached 20.94 million in 2024, representing an 8.84% increase from 19.24 million in 2023 (Ministry of Tourism Government of India, 2025)**.** This data shows that India is a country which have domestic as well as international tourist attractions because of its diversity in culture and heritage, which can be seen through its tangible heritage like Agra Fort, Taj Mahal, Amber Fort and intangible heritage like festivals, traditions, and folksongs (Government of India Ministry of Culture, 2025).

The Indian government, as well as the United Nations Educational, Scientific and Cultural Organization (UNESCO), has numerous schemes in place for the preservation of heritage (Ministry of Tourism, India, 2025; UNESCO World Heritage Convention, 2025). Tourism to and within India has undergone some important changes in recent years, as shown by the rising numbers of international tourist's and an increase in domestic tourism. This has led to the redevelopment and rebranding of many of its destinations as the Indian government has begun to recognize the potential importance of tourism to the Indian economy and has begun to invest in tourism infrastructure (Hannam & Diekmann, 2010).

Heritage tourism integrates tangible and intangible cultural assets, offering educational, aesthetic, and emotional experiences while promoting site preservation and sustainable tourism (Timothy & Boyd, 2006). Motivation and Push–Pull Theory explains tourists attraction to heritage sites through intrinsic desires and external site-related factors (Dann, 1981).

2.2 Online Review and Tourism

Information and communication technologies (ICTs) have significantly influenced the evolution of tourism on a global scale (Gretzel & Yoo, 2008). They empower consumers to search for, compare, and purchase tourism products and services with greater efficiency and transparency (Anand Bethapudi, 2013). Several studies have highlighted this digital empowerment reshaping the traditional tourism value chain, shifting greater control to

consumers while enabling service providers to engage directly with potential travelers through online platforms and data-driven marketing strategies (Buhalis & Law, 2008). In a previous study, it was found that the new generation of empowered tourist's who are more informed, discerning, and value-conscious. These travelers actively seek meaningful experiences that offer exceptional value for both time and money. Unlike traditional tourists who preferred organized group tours, contemporary travelers increasingly favor independently planned journeys tailored to their individual interests and schedules (Buhalis & Law, 2008).

2.3. Big Data Analysis

Big data analytics is the process of examining large volumes of both structured and unstructured data available on the internet (Hu et al., 2019). Research in tourism increasingly encounters large volumes of qualitative textual data, for example, online reviews, social media posts, and other user-generated content. Grounded Theory (White & Cooper, 2022) offers an inductive, theory-building approach by coding data, constant comparison, conceptual documentation, and developing categories until theoretical saturation (Papathanassis & Knolle, 2011).

Big data analytics facilitates a deeper analysis of visitor experiences and their influence on overall satisfaction**.** For example (Riswanto et al., 2023) employed big data methods to analyze increasing customer satisfaction and promoting Indian culture requires tourism businesses to understand factors influencing tourist's experiences and behavior towards heritage destinations using online reviews collected from Google Travel. Similarly, Moreno Brito et al. (2024) analyzed online reviews from fourteen eco-friendly hotels in Ecuador and found that employee politeness and site accessibility were pivotal in enhancing visitor satisfaction (Moreno Brito et al., 2024). Drawing on these insights, the present study applies big data–driven qualitative analysis to online reviews of Agra Fort, aiming to identify key themes, satisfaction drivers, and experiential patterns specific to heritage tourism, where cultural authenticity, historical significance, and visitor engagement play central roles in shaping satisfaction.

2.4. Grounded Theory

Grounded theory (Glaser & Strauss, 1967) emphasizes the inductive formulation of theory from collected data, rather than evaluating existing theories. Unlike previous study approaches, grounded theory emphasizes iterative data collection and analysis, allowing categories, relationships, and theoretical frameworks to emerge directly from the data itself (Noble & Mitchell, 2016). This method is particularly useful in areas where existing theoretical explanations are insufficient or where new behavioral patterns are emerging, such as the evolving dynamics of tourist online behavior in the digital era (Bischof & Freybe, 2022).

In tourism studies, grounded theory serves as a methodological approach to investigate travelers' intrinsic motivations, perceptions, and decision-making, based on insights from user-generated content (Kim et al., 2024). By systematically coding and categorizing online reviews, researchers can identify recurring themes and build conceptual models that explain how tourist's experience, evaluate, and interact with heritage destinations. This approach enables a deeper understanding of intangible and context-specific phenomena such as cultural perceptions, authenticity seeking, or emotional attachment that may not be captured through traditional survey-based methods. For studies focused on India's heritage tourism, grounded theory offers a powerful framework to reveal how tourist's construct meanings around heritage sites, how these meanings influence their behavior online, and how they contribute to broader patterns of destination image formation and heritage appreciation.

Several studies have adopted grounded theory approaches in the analysis of big data. A previous study mentioned that user-generated content provides rich, authentic insights into tourist's cognitive and emotional experiences, highlighting how travelers construct and communicate meanings about destinations (Papathanassis & Knolle, 2011). Another previous study mentioned that grounded theory revealed that evaluation motivations shaped by product perception, social influence, and perceived behavioral control play a critical mediating role in consumer satisfaction evaluation behavior, thereby shedding light on how online review behavior emerges in a social-commerce context (Wang et al., 2024).

2.5. Sentiment Analysis of Online Reviews

Sentiment analysis (opinion mining) is a computational technique designed to detect, categorize, and analyze subjective expressions in textual data. With the rapid growth of user-generated content on platforms such as TripAdvisor, Google Reviews, and social media, sentiment analysis has become an essential tool in tourism research for capturing travelers' perceptions, satisfaction levels, and emotional responses in many studies (Al-Otaibi et al., 2018; Lu et al., 2022). It typically involves the use of natural language processing (NLP) and machine learning algorithms to determine the polarity (positive or negative) and intensity of sentiments expressed in online reviews, however there are many available software that can count the sentiment score based on the review text (Jongeling et al., 2015).

In the domain of heritage tourism, sentiment analysis offers a means to understand visitor experiences, emphasizing key aspects of appreciation and aspects that may benefit from improvement. Positive sentiments often reflect satisfaction with authenticity, preservation, interpretation, and visitor experience, while negative sentiments may highlight challenges such as overcrowding, accessibility issues, or inadequate facilities (Zhang & Xiao, 2025). By analyzing large-scale online review data, researchers can uncover patterns of tourist behavior, preferences, and expectations at heritage destinations (Gursoy et al., 2021). Moreover, sentiment analysis complements grounded theory by providing quantitative evidence that supports or extends the qualitative themes identified, thereby enabling a more holistic understanding of tourist online behavior and its implications for destination management and marketing strategies.

Several prior studies have also applied sentiment analysis to online reviews, as demonstrated in various research papers like (Aliyah et al., 2025). This previous study shows that the sentiment analysis revealed positive guest comments were strongly associated with words like "clean" and "environment," indicating that visitors to green hotels value sustainability initiatives, while negative sentiment clustered around terms such as "price" and "bad," suggesting that high costs without perceived value can undermine guest satisfaction. Another previous study shows that negative emotions among traveler's were predominantly triggered by delays and service disruptions, whereas

positive sentiments were strongly associated with keywords like “good”, “great”, “excellent”, “clean” and “smiling”, indicating that cleanliness, friendly staff and in-flight comfort significantly enhance user satisfaction (Kwon et al., 2021).

3. Methodology

The study adopts qualitative methods to investigate the current condition of heritage tourism in India. In order to gather the customer opinions and investigate the correlation between the tourist experience and satisfaction with India’s heritage through the tourist online reviews, the research design was formulated and represented in figure 1. The study comprises the two primary sections. The first is the collection of data, and the second is the involvement of qualitative analysis, like frequency of words, co-occurrence, cluster and the sentiment analysis done among the words and between the words.

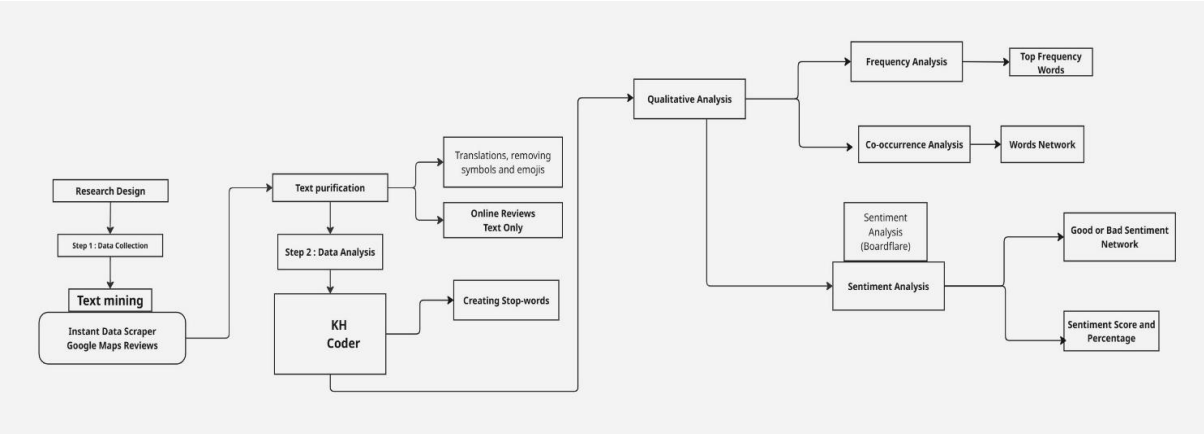

Figure 1. Research Flow

3.1. Data Collection

This research investigates the qualitative dimensions of tourist behavior at Agra Fort through the systematic analysis of online reviews obtained from Google Maps. Google Maps was chosen because, unlike platforms such as TripAdvisor or Booking.com that mainly capture pre-trip planning or accommodation feedback, it provides a larger set of location-verified, post-visit reviews that offer authentic and diverse insights into actual visitor experiences. Adopting the Grounded Theory approach enables an inductive

exploration of visitor's narratives to identify emergent themes and interpret underlying behavioral patterns. This method allows the study to move beyond surface-level data, uncovering the nuanced perceptions, motivations, and experiential satisfaction expressed by tourists in their own words (Glaser & Strauss, 1967a; Papathanassis & Knolle, 2011)**.** Through this qualitative framework, the study seeks to construct a conceptual model of tourist behavior that reflects the affective and cognitive dimensions of heritage tourism experiences at Agra Fort. The affective and cognitive dimensions refer to two separate aspects of a person's experience or attitude: affective relates to emotions, feelings, and values, while cognitive relates to thought, knowledge, and belief.

Figure 2. Online review visualization

As shown in figure2, the textual portion of the reviews will be used to examine tourists satisfaction levels in this study. In accordance with South Korean laws and institutional regulations, this study did not necessitate formal ethical or legal approval. All research procedures were designed and implemented in full compliance with data protection, confidentiality, and copyright standards to uphold the integrity and ethical rigor of the study. The analysis was conducted exclusively on publicly available data, ensuring that no

personally identifiable information was collected, disclosed, or utilized. Accordingly, the study conforms to all relevant legal and ethical requirements governing research of this nature.

3.2. Data Analysis

Qualitative textual analysis was carried out using KH Coder 3 after data collection, leveraging its capabilities for content analysis and interpretation of textual information. The software assists in the systematic interpretation of qualitative data by applying linguistic processing techniques and generating visual representations that reveal relationships among frequently used terms but the filtering out the irrelevant or insignificant words need to be done so that they don't interfere with the result.

The processed data underwent analysis using KH Coder 3 software, starting with the initial scrutiny of the top 100 most frequently used words. Afterwards the co-occurrence network analysis of these 100 words was done to see the close relationship among them. Then the sentiment analysis is done using the text review in excel with the excel extension for sentiment analysis named BoardFlare which shows the positive and negative sentiments which further tells the satisfaction level of the tourist for the Indian heritage Agra Fort. Further the visualization is done with the help of the KH Coder 3 software.

4. Analysis and Results

4.1. Frequency Analysis

A dataset comprising 1,309 tourist reviews was obtained from Google Maps via the Google Instant Data Scraper. Later they were shaped into a set of data that could be used for the research base for Indian heritage tourism. Keywords that are closely linked with this research topic were identified using a frequency table. Table 1 shows the ranking and frequency of the top 100 most used words including comprehensive topics connected to the tourist satisfaction towards India's heritage. The top few most frequent words are 'place', 'Agra Fort', 'visit', 'good', 'Taj Mahal', 'beautiful', 'Agra', 'history', 'nice', 'Red Fort'.

Table 1. Frequency Table

Rank	Words	TF	Rank	Words	TF	Rank	Words	TF	Rank	Words	TF
1	Place	408	26	wonderful	27	51	great	52	76	Huge	18
2	AgraFort	398	27	love	25	52	monument	51	77	intricate	18
3	Visit	262	28	recommend	25	53	take	51	78	Stone	18
4	Good	187	29	awesome	24	54	sandstone	47	79	Story	18
5	TajMahal	180	30	City	24	55	India	46	80	Area	17
6	Beautiful	175	31	grandeur	24	56	site	45	81	Better	17
7	Agra	152	32	Indian	24	57	ticket	43	82	Clean	17
8	History	150	33	there	24	58	well	43	83	Close	17
9	Nice	130	34	magnificent	23	59	Emperor	42	84	Look	17
10	RedFort	114	35	really	23	60	palace	42	85	Nothing	17
11	Historical	107	36	many	22	61	stunning	42	86	People	17
12	Architecture	105	37	massive	22	62	get	40	87	Truly	17
13	See	104	38	tourist	22	63	Mughal	39	88	Walk	17
14	Mughal	97	39	Era	21	64	Akbar	36	89	What	17
15	View	78	40	grand	21	65	know	36	90	Where	17
16	Experience	76	41	historic	21	66	UNESCO	36	91	Avoid	16
17	Guide	75	42	most	21	67	wall	36	92	Big	16
18	Here	66	43	who	21	68	Great	35	93	Carving	16
19	Amazing	64	44	feel	20	69	maintain	33	94	morning	16
20	Build	61	45	hour	20	70	make	33	95	Shahjahan	16
21	Best	60	46	beauty	19	71	rich	30	96	breathtaking	15
22	Time	60	47	building	19	72	worth	30	97	incredible	15
23	Heritage	59	48	enjoy	19	73	World	29	98	Delhi	14
24	Go	55	49	structure	19	74	impressive	28	99	Full	14
25	Visit	55	50	day	18	75	explore	27	100	interesting	14

4.2. Co-occurrence And Cluster Analysis

Figure 3 presents a co-occurrence network visualization highlighting the connections among the most frequently mentioned words in reviews of Agra Fort, organized into nine subgraphs. Each node in the figure represents words and the frequency is represented by the size of the node. Lines connecting signify a co-occurrence relationship that is, words that frequently appeared together in the same context.

Cluster analysis supports the research objective by revealing patterns in tourist behavior and perceptions related to Agra Fort as an Indian heritage site, consistent with previous studies using similar approaches in tourism research (Kim et al., 2023). The subgraphs

were consolidated into five thematic clusters based on a careful visual inspection of co-occurring keywords and thematic similarity. Each cluster highlights a specific dimension of the tourist experience, demonstrating how visitors evaluate the site from multiple perspectives. Table 2 shows the cluster and its significant words.

The first group is History & Culture, which includes build, Akbar, emperor, Mughal, city, Shahjahan, structure. The second group is General Infrastructure, which includes grand, walk, story, palace, intricate, carving, grandeur. The third group is UNESCO, which includes heritage, UNESCO, India, world. The fourth group is Tourist Infrastructure, which includes enjoy, people, well, maintained, go. The fifth group is Heritage Experience, which includes impressive, monument, historical, experience, visit, place, Agra, beautiful, Agra Fort, history, rich, hour, emperor, Mughal, architecture, stunning, breathtaking, view, Taj Mahal.

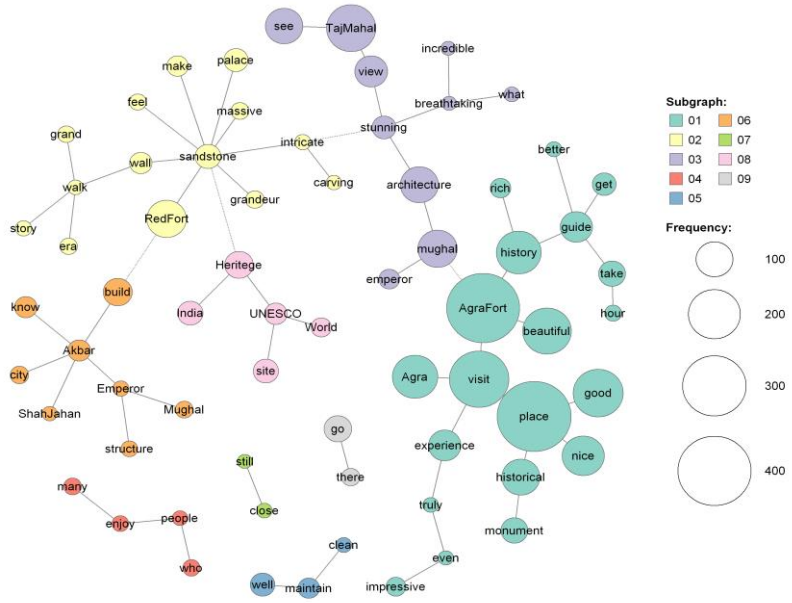

Figure 3. co-occurrence visualization

Table 2. Cluster Table

Clusters	Words	Significant Words
History &Culture	know/build/Akbar/emperor/Mughal /city/Shahjahan/structure	Know/build/Akbar/emperor/ Mughal/city/Shahjahan/ structure
General infrastructure	grand/walk/story/era/wall/feel/mak e/palace/massive/sandstone/intricat e/carving/ grandeur	Grand/walk/story/era/palace/sa ndstone/intricate/carving/ grandeur
UNESCO	heritage/UNESCO/India/world	Heritage/UNESCO/India/ world
Tourist infrastructure	Many/enjoy/people/who/still/close/ well/maintain/clean/go/there	Enjoy/people/well/maintained/ go
Heritage experience	Impressive/even/truly/monument/ historical/experience/visit/Place/ good/nice/Agra/beautiful/AgraFort/ history/rich/better/get/take/hour/ emperor/Mughal/architecture/ stunning/breathtaking/what/ incredible/view/TajMahal/see	Impressive/monuments/historic al/experience/visit/place/Agra/ beautiful/AgraFort/history/ rich/hour/emperor/Mughal/ architecture/stunning/ breathtaking/incredible/ view/TajMahal

4.3. Sentiment Analysis

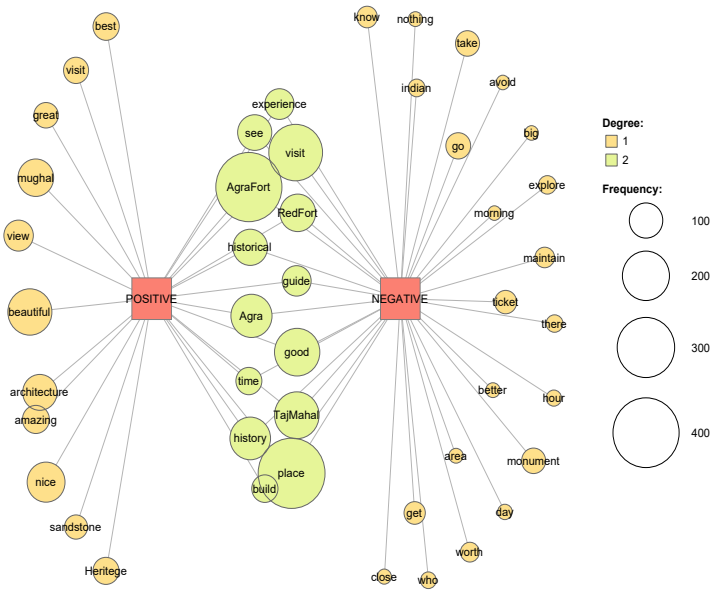

Figure 4. Sentiment Analysis Visualization

To analyze how tourists perceive Agra Fort as part of India's heritage tourism, sentiment analysis was performed on 1,309 Google Maps reviews obtained through the Instant Data Scraper Chrome tool. This approach helps in identifying the positive and negative experiences, which highlights the area that are appreciated and where they need improvement. Figure 4. Shows the visual representation of the sentiment analysis with the degree and frequency through KH Coder 3 software. The network shows strong associations between terms like Agra Fort, TajMahal, place, and historical in positive reviews, reflecting visitor appreciation for the architecture, historical significance, and scenic views. Negative reviews highlight words such as ticket, avoid, and crowd, indicating common visitor complaints regarding logistics and management. This visualization effectively captures the main themes in visitor feedback.

Table 3. Sentiment Distribution Table

Metrics	Value	Percentage
Positive Reviews	1,133	**86.55%**
Negative Reviews	176	**13.45%**
Total Views	1,309	

The sentiment analysis table (Table 3) summarizes the overall distribution of positive and negative reviews. Out of 1,310 reviews, 1,134 were positive and 176 were negative. The overwhelming majority of positive feedback confirms that visitors generally have a favorable experience, enjoying the monument's beauty, historical value, and guided tours. The smaller number of negative reviews points to specific issues such as ticketing delays, crowd management, and timing challenges, which could be addressed to enhance the visitor experience further. In the table 4 selected reviews with their corresponding sentiment scores illustrate the practical application of this analysis.

Positive reviews emphasize historical and architectural beauty, scenic views, and helpful guides, with sentiment scores close to 0.99, indicating strong confidence in the classification. Negative reviews, with scores of 0.96–0.99, focus on management-related issues like long queues or visiting times. These examples demonstrate how sentiment

analysis can quantify visitor opinions, providing actionable insights for improving the tourism experience.

Table 4. Sentiment Score and Example of Review Text

Review Text	Sentiment	Sentiment Score
Loved the place! The AgraFort is massive, clean, and full of history. Best part is the view of the TajMahal from inside. Totally worth it. Very impressive AgraFort with grand **architecture**. Easy to explore and great photo spots. Must visit if you're in Agra.	Positive	0.99
Huge Line for Tickets, No Proper Management, would prefer to visit as early **Morning** as soon as Visit Start or After 5 PM.	Negative	0.98
Average Sentiment Score 0.97		

5. DISCUSSION AND CONCLUSIONS

5.1. Discussion

This study's findings contribute to a deeper understanding of tourist perceptions and behaviors concerning Agra Fort, a major heritage site in India. The co-occurrence and sentiment analyses identified five primary clusters namely, History & Culture, General Infrastructure, UNESCO, Tourist Infrastructure, and Heritage Experience, which collectively illustrate the knowledge-based and emotion-driven dimensions of visitor experience (Poria et al., 2003). These findings defends that tourist's evaluations of heritage sites are shaped by both intellectual engagement with history and emotional appreciation of aesthetics, aligning with prior research that highlights heritage tourism as a multidimensional experience involving knowledge-based and emotional-driven, and behavioral components (Hosany & Gilbert, 2010).

The History & Culture cluster highlights tourists deep engagement with historical and cultural aspects, particularly figures such as "Akbar," "Shahjahan," and the "Mughal". Visitors actively interpret and connect with India's historical narratives, demonstrating that authenticity and storytelling form essential components of heritage appreciation (Timothy & Boyd, 2006). Such engagement supports the notion that heritage sites serve as "living

classrooms," where tourists seek to learn, connect, and reflect on the past (Chronis, 2012). Similarly, the General Infrastructure cluster captures visitor's fascination with the architectural beauty and craftsmanship of the fort, reflected in frequent mentions of "intricate," "sandstone," and "grandeur." This aligns with findings that aesthetic appreciation of tangible heritage plays a vital role in shaping positive perceptions and destination attachment (Drost, 1996).

The UNESCO cluster highlights how international acknowledgment through UNESCO designation contributes to strengthening the destination's image and enhancing visitor trust. References to "UNESCO," "world," and "heritage" indicate that international designation reinforces the perceived value and prestige of heritage sites. Similar findings by Su and Wall (2011) and Gravari-Barbas (2018) suggest that UNESCO labelling enhances tourist's expectations regarding authenticity, preservation, and global importance. Thus, international validation contributes to both intellectual and emotional satisfaction among visitors.

The Tourist Infrastructure cluster highlights visitors concerns regarding operational aspects such as "cleanliness," "maintenance," and "management". The presence of words like "clean," "maintain," and "enjoy" demonstrates that service quality is central to visitor's emotional or mental satisfaction. Conversely, negative sentiments around "ticket" and "crowd" highlight operational challenges that may detract from the overall experience. This observation corresponds with studies emphasizing that management efficiency, visitor flow, and infrastructure quality significantly influence satisfaction in heritage tourism (Gursoy et al., 2021).

The Heritage Experience cluster captures the dual nature of tourists engagement, where emotional and intellectual responses jointly influence satisfaction and the perceived significance of the heritage site. Keywords such as "beautiful," "stunning," "breathtaking," and "incredible" reveal deep emotional attachment and aesthetic admiration. This finding resonates with research noting that heritage experiences evoke strong emotional resonance and a sense of place identity (Park, 2010). The dominance of positive sentiments (approximately 86.5%) reinforces that Agra Fort delivers a fulfilling experience that

blends historical learning with emotional enrichment—an essential attribute of sustainable heritage tourism (Richards, 2018; Timothy, 2018).

5.2. Theoretical Implication

This research makes important theoretical contributions to the study of heritage tourism as well as the application of big data analytics. First, by applying co-occurrence and sentiment analysis to user-generated content, it extends the methodological scope of heritage tourism research beyond traditional survey or interview-based approaches. The integration of qualitative grounded theory with quantitative text mining demonstrates the value of hybrid methods in capturing both the cognitive and affective dimensions of visitor experience. This approach aligns with and expands upon existing literature emphasizing that heritage tourism is a multidimensional phenomenon shaped by knowledge acquisition, emotional engagement, and behavioral outcomes (Hosany & Gilbert, 2010; Poria et al., 2003).

Second, the delineation of five category-based clusters—History & Culture, General Infrastructure, UNESCO, Tourist Infrastructure, and Heritage Experience—provides an evidence-based framework for examining tourist's perceptions and evaluations of heritage sites. These clusters reflect a layered interpretation of heritage experiences that encompasses intellectual curiosity, aesthetic appreciation, and service-related satisfaction. Such a framework enriches theoretical discussions on how tourist's construct meaning at heritage sites and supports the notion that cognitive and affective components jointly drive satisfaction and attachment (Chronis, 2012; Park, 2010). Finally, by focusing on online reviews, the study contributes to the growing body of literature on digital consumer behavior in tourism contexts. It underscores the theoretical importance of user-generated content as a lens through which tourist experiences, perceptions, and expectations can be observed in real time. This highlights the evolving nature of heritage tourism theory in the digital age, where big data insights complement and deepen traditional theoretical models.

5.3. Managerial Implications

The findings also provide practical guidance for heritage site managers, destination planners, and policymakers. First, the predominance of positive sentiments which particularly linked to historical authenticity, architectural grandeur, and emotional resonance that underscores the need to preserve and promote these core attributes. Managers should design interpretive programs and storytelling initiatives that deepen visitor's cognitive engagement with the site's history while also fostering emotional connection, thereby enhancing the overall experience and encouraging repeat visitation.

Second, the analysis demonstrates that infrastructure and the quality of services are vital contributors to overall tourist satisfaction. Issues such as crowd management, ticketing efficiency, and cleanliness require strategic attention. Managers should consider implementing visitor flow control systems, improving signage and information dissemination, and adopting data-driven maintenance practices. Leveraging real-time sentiment monitoring from online platforms can enable more responsive management decisions and continuous service improvement.

Third, destination marketing strategies should incorporate the emblematic value of UNESCO recognition. Highlighting the site's global recognition in promotional materials can strengthen its image, enhance trust, and attract culturally motivated travelers. Collaboration with international heritage bodies can further enhance preservation practices and global visibility.

Finally, policymakers should integrate digital platforms into strategic approaches for effective heritage site management. Regular analysis of online reviews can serve as an early warning system for emerging issues and a valuable source of visitor feedback. This big data–driven approach can guide investment priorities, enhance visitor communication strategies, and support sustainable heritage tourism development.

5.4. Conclusion, Limitation and Future Studies

By analyzing user-generated reviews with big data analytics, this study provides deeper insights into tourist perceptions, emotional responses, and behavioral patterns related to Agra Fort. Through co-occurrence and sentiment analyses, five thematic clusters namely

History & Culture, General Infrastructure, UNESCO, Tourist Infrastructure, and Heritage Experience were identified, capturing the multifaceted nature of visitor experiences that integrate cognitive, affective, and behavioral dimensions. These findings confirm that heritage sites are valued not only for their historical authenticity and architectural beauty but also for their capacity to evoke emotional resonance and foster personal connections.

This research demonstrates how the combination of grounded theory and algorithmic text mining can link qualitative interpretation with quantitative accuracy. This hybrid approach contributes to heritage tourism scholarship by illustrating how digital data can reveal nuanced visitor insights at scale. Practically, the findings highlight the importance of both tangible (e.g., infrastructure quality, crowd management) and intangible (e.g., storytelling, symbolic recognition) elements in shaping visitor satisfaction and destination attachment.

Although the study provides significant contributions, it is important to acknowledge its limitations. First, the study's data source was restricted to Google Maps reviews, which may not fully represent the diversity of visitor opinions expressed across other platforms. Incorporating additional data sources such as TripAdvisor, Instagram, or travel blogs in future research could provide a more comprehensive view of visitor perceptions. Second, the analysis focused on a single UNESCO site, limiting the generalizability of the findings. Comparative studies involving multiple heritage destinations across different cultural and geographic contexts could offer deeper insights into how site characteristics shape visitor experiences. Third, the cross-sectional nature of the data constrains the ability to observe changes in visitor sentiment over time. Longitudinal analyses could uncover evolving expectations and trends, contributing to more adaptive heritage management strategies.

Further research might explore the role of demographic variables—like nationality, age, and travel motivation—in influencing tourist's cognitive and emotional responses. Additionally, coupling text mining with advanced analytical approaches, such as topic modeling or machine learning-based sentiment classification, could further refine the understanding of visitor behavior.

In sum, the research provides valuable contributions to heritage tourism theory and practice by highlighting the role of digital footprints in informing academic and managerial perspectives. By highlighting the central role of emotional and cognitive

engagement, the findings underscore the importance of designing heritage experiences that are not only informative but also deeply meaningful key factor in ensuring the sustainable success of heritage tourism in the digital age.

Supplementary Materials: The data supporting the findings of this study are available from the author upon request.

Author Contributions: : Conceptualization, S.A. and A.W.; Methodology, S.A. and H.N.D.; Software, S.A. and A.W.; Validation, S.A., A.W. and H.N.D.; Formal Analysis S.A. and A.W.; Data Curation, S.A., A.W., and H.N.D.; Writing—Original Draft Preparation, S.A. .; Writing—Review and Editing, A.W. and H.N.D.; Visualization, A.W.; Supervision, H.N.D.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study did not involve human participants or interventions requiring ethical approval. All data were obtained from publicly available online sources.

Informed Consent Statement: Not applicable. All data were obtained from publicly available online sources.

Data Availability Statement: The data of this study are available from the author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Aliyah, U., Williady, A., & Kim, H.-S. (2025). An evaluation of green hotels in Singapore, Sentosa Island: A big data study through online review. *Tourism and Hospitality*, *6*(1), 24. https://doi.org/10.3390/tourhosp6010024

Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N., & Albugami, A. (2018). Customer satisfaction measurement using sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *9*(2). https://doi.org/10.14569/IJACSA.2018.090216

Anand Bethapudi. (2013). The role of ICT in tourism industry. *Journal of Applied Economics and Business*, *1*(4), 67–79. https://www.aebjournal.org/articles/0104/010406.pdf

Baniya, R., & Thapa, B. (2024). Conceptualizing visitors experience in heritage hotels using big data. *Journal of Heritage Tourism*, *19*(3), 451–469. https://doi.org/10.1080/1743873X.2024.2302533

Bischof, A., & Freybe, K. (2022). Grounding Theory in digital data: A methodological approach for a reflective procedural framework. *Journal of Cultural Analytics*, *7*(4). https://doi.org/10.22148/001c.57197

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—the state of e-tourism research. *Tourism Management*, *29*(4), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Chronis, A. (2012). Tourists as story-builders: Narrative construction at a heritage museum. *Journal of Travel & Tourism Marketing*, *29*(5), 444–459. https://doi.org/10.1080/10548408.2012.691395

Dann, G.M.S. (1981). *Tourist motivation: An appraisal. Annals of Tourism Research*, 8(2), 187- 219. https://doi.org/10.1016/0160-7383(81)90082-7

Drost, A. (1996). Developing sustainable tourism for world heritage sites. *Annals of Tourism Research*, *23*(2), 479–484. https://doi.org/10.1016/0160-7383(96)83345-7

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of Grounded Theory*. Aldine Transaction A Division of Transaction Publishers. http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf

Government of India Ministry of Culture. (2025). https://culture.gov.in/ministry/world-heritage

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. In *Information and Communication Technologies in Tourism 2008*. Springer Vienna. https://doi.org/10.1007/978-3-211-77280-5_4

Gursoy, D., Akova, O., & Atsz, O. (2021). Understanding the heritage experience: A content analysis of online reviews of world heritage sites in Istanbul. *Journal of Tourism and Cultural Change*, *20*(3), 311–334. https://doi.org/10.1080/14766825.2021.1937193

Hannam, K., & Diekmann, A. (2010). *Tourism and India* (1st Edition). Routledge. https://doi.org/10.4324/9780203868782

Hosany, S., & Gilbert, D. (2010). Measuring tourists emotional experiences toward hedonic holiday destinations. *Journal of Travel Research*, *49*(4), 513–526. https://doi.org/10.1177/0047287509349267

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? text analysis using structural topic model. *Tourism Management*, *72*, 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Incredible India. (2025). *Incredible India*. https://incredible-india.org/theme/forts-palaces.html

Jongeling, R., Datta, S., & Serebrenik, A. (2015). Choosing your weapons: On sentiment analysis tools for software engineering research. *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 531–535. https://doi.org/10.1109/ICSM.2015.7332508

Kim, S., Williady, A., Wang, J., & Kim, H.-S. (2024). A mixed-method approach to Grounded Theory regarding the MICE industry at Busan exhibition & convention center. *Tourism and Hospitality*, *5*(4), 1012–1027. https://doi.org/10.3390/tourhosp5040057

Kwon, H.-J., Ban, H.-J., Jun, J.-K., & Kim, H.-S. (2021). Topic modeling and sentiment analysis of online review for airlines. *Information*, *12*(2), 78. https://doi.org/10.3390/info12020078

Lu, Q., Sun, X., Sutcliffe, R., Xing, Y., & Zhang, H. (2022). Sentiment interaction and multi-graph perception with graph convolutional networks for aspect-based sentiment analysis. *Knowledge-Based Systems*, *256*, 109840. https://doi.org/10.1016/j.knosys.2022.109840

Ministry of Tourism, Government of India. (2002). https://tourism.gov.in/

Ministry of Tourism Government of India. (2024). *India tourism data compendium*. https://tourism.gov.in/sites/default/files/2025-03/India%20Tourism%20Data%20Compendium%202024_0.pdf

Ministry of Tourism Government of India. (2025). *India tourism data compendium, Ministry of Tourism, India.* 21. https://tourism.gov.in/sites/default/files/2025-09/India%20Tourism%20Data%20Compendium%202025_1.pdf

Ministry of Tourism, India. (2025). https://tourism.gov.in/

Moreno Brito, Y. L., Ban, H.-J., & Kim, H.-S. (2024). Ecological hotels customer satisfaction through text mining of online reviews: A case of Ecuador hotels. *Journal of Hospitality and Tourism Insights*, *7*(3), 1532–1552. https://doi.org/10.1108/JHTI-09-2022-0467

Noble, H., & Mitchell, G. (2016). What is grounded theory? *Evidence Based Nursing*, *19*(2), 34–35. https://doi.org/10.1136/eb-2016-102306

OECD. (2025). *Building strong and resilient tourism destinations*. OECD. https://doi.org/10.1787/a57c343d-en

Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management*, *32*(2), 215–224. https://doi.org/10.1016/j.tourman.2009.12.005

Park, H. yu. (2010). Heritage tourism. *Annals of Tourism Research*, *37*(1), 116–135. https://doi.org/10.1016/j.annals.2009.08.001

Poria, Y., Butler, R., & Airey, D. (2003). The core of heritage tourism. *Annals of Tourism Research*, *30*(1), 238–254. https://doi.org/10.1016/S0160-7383(02)00064-6

Riswanto, A. L., Kim, S., & Kim, H.-S. (2023). Analyzing online reviews to uncover customer satisfaction factors in Indian cultural tourism Destinations. *Behavioral Sciences*, *13*(11), 923. https://doi.org/10.3390/bs13110923

Sharma, P., & Nayak, J. K. (2018). Retracted: Testing the role of tourists emotional experiences in predicting destination image, satisfaction, and behavioral intentions: A case of wellness tourism. *Tourism Management Perspectives*, *28*, 41–52. https://doi.org/10.1016/j.tmp.2018.07.004

Timothy, D. J., & Boyd, S. W. (2006). Heritage tourism in the 21st century: Valued traditions and new perspectives. *Journal of Heritage Tourism*, *1*(1), 1–16. https://doi.org/10.1016/0160-7383(81)90082-7

UN Tourism. (2025). *World tourism barometer, excerpt*. https://tourismanalytics.com/uploads/1/2/0/4/120443739/world_tourism_barometer_volume_23_issue_3_september_2025.pdf

UNESCO World Heritage Convention. (2025). *World heritage list*. https://whc.unesco.org/en/list/

Wang, R., Xu, S., Li, S., & Pang, Q. (2024). Research on influence mechanism of consumer satisfaction evaluation behavior based on Grounded Theory in Social e-commerce. *Systems*, *12*(12), 572. https://doi.org/10.3390/systems12120572

White, R. E., & Cooper, K. (2022). Grounded Theory. In *Qualitative Research in the Post-Modern Era* (pp. 339–385). Springer International Publishing. https://doi.org/10.1007/978-3-030-85124-8_9

Zhang, A., & Xiao, H. (2025). Digitalising cultural heritage through online reviews: a sentiment analysis. *Current Issues in Tourism*, 1–18. https://doi.org/10.1080/13683500.2025.2566861