WHIOCE PUBLISHING PTE. LTD.
PROVIDING FIRST-CLASS SCIENTIFIC INFORMATION FOR TOP SCHOLARS

2025,Volume 3, Issue 2
Lecture Notes in Education, Arts, Management and Social Science
Online ISSN: 2705-053X

Research on the Tourism Image Perception of Sanya Scenic Spots based on Big Data Text Mining

Jun Yu[1], Hongjuan Wang[2]*, Chaonan Jia[3]

[1]School of Tourism, Huangshan University, Huangshan 245000, Anhui, China

[2]Pusan National University, Busan 47545, Republic of Korea

[3]School of Business, Anyang Normal University, Anyang 455000, Henan, China

**Author to whom correspondence should be addressed.*

Abstract: The online travel evaluations of tourists about Sanya scenic spots were collected through Python crawler technology. Then, ROSTCM6.0 software was used to perform text analysis on the collected online travel evaluations. The high-frequency words, word units, and visual cloud maps of the network semantics of online travel evaluations were analyzed and sorted out. Finally, from the data processing results, it was concluded that the three dimensions of the perception of the tourism image of Sanya scenic spots are the company service dimension, the team atmosphere dimension, and the natural scenery dimension. From the unique perspective of group tourists, the tourists' perception of the tourism image of Sanya scenic spots was analyzed, and then the tourists' perception characteristics of the electronic word-of-mouth (e-WOM) of Sanya scenic spots were studied and analyzed. The research results provide reference significance for tourists' travel to Sanya scenic spots, and at the same time provide data support for the decision-making of tourism companies.

Keywords: Big data text mining; Electronic word-of-mouth (e-WOM); Group tourists; Tourism image perception

Online publication: March 26, 2025

1. Introduction

With the progress of the times and the development of technology, electronic word-of-mouth (e-WOM) formed by online tourist comments plays a crucial role in customers' tourism decision-making [1]. In the era of big data and information technology, new technologies such as 5G and artificial intelligence have brought convenience to tourists' travel experiences [2]. The emergence of various multimedia software applications has provided a platform for tourists to share their stories and travel experiences [3]. With the widespread use of smartphones, more and more people enjoy posting their travel insights on various social media platforms, forming a type of e-WOM [4]. Since the end of the pandemic in 2023, the tourism industry has rapidly recovered. According to the "2023 Cultural and Tourism

Development Statistics Bulletin of the Ministry of Culture and Tourism of the People's Republic of China" issued by the Ministry of Culture and Tourism, the number of domestic tourist trips in 2023 reached 4.89 billion, an increase of 93.3% compared to 2022 [5]. This article mainly analyzes tourists' perceptions of the tourism image of Sanya scenic spots from the unique perspective of group tourists, thereby investigating the characteristics of their e-WOM perceptions of Sanya tourism.

2. Literature review

2.1. Electronic word-of-mouth

The concept of e-WOM originated from the term "online word-of-mouth." With the continuous development of electronic technology and the progress of digitization, the term e-WOM emerged, giving rise to research on the network semantics of online customer electronic word-of-mouth [6]. In the tourism industry, e-WOM largely consists of tourists' personal experiences and travel stories, playing a crucial role in shaping the image of tourism destinations. Most tourists prepare travel guides beforehand, relying primarily on user-generated online travelogues and UGC (user-generated content) to gather various information about their intended destinations. This information encompasses multiple aspects such as local cuisine, scenic spots, travel itineraries, and hotel details along the route. UGC not only assists tourists in planning their trips but also provides valuable insights for destination managers to identify areas for improvement and development [7]. Tourist e-WOM mainly consists of spontaneous personal insights and evaluations generated by tourists after their travels. It plays a decisive role in shaping the image of tourism destinations. For instance, the rising popularity of Harbin's cultural tourism and the booming situation of Zibo's barbecue are positive impacts of online e-WOM on the promotion of tourism destinations. Therefore, tourist e-WOM holds significant importance in shaping the image of tourism destinations and influencing tourists' personal decision-making processes.

2.2. Tourism image perception

The concept of tourism destination image was initially introduced by scholar JD Hunt and quickly became a focal point of research in various disciplines such as tourism studies and psychology [8]. Tourism image perception refers to the overall personal evaluation of a destination by tourists based on their experiences during their visit. It encompasses multiple aspects, including food, accommodation, transportation, shopping, entertainment, and more. Tourists' perception of a destination's image is influenced by various factors such as the cultural significance of the destination, natural scenic beauty, and online reviews from other tourists [9].

Research methods for studying tourism destination image perception have gradually diversified, with a wide range of quantitative and qualitative analysis techniques emerging. These include quantitative methods like survey questionnaires and IPA analysis, as well as qualitative methods such as text analysis, content analysis, and grounded theory [10]. These methodologies provide decision-makers and managers with valuable data-driven insights and theoretical support to assess the image perception of tourism destinations. This article primarily utilizes big data text analysis and network semantic visualization techniques to analyze and study the online reviews of group tourists regarding their experiences at Sanya tourist attractions. The aim is to explore the perception structure of group tourists towards the tourism image of Sanya.

3. Research design and methodology

3.1. Sample selection and processing

This article primarily employs Python web scraping technology to extract online tourism evaluation data from group tourists regarding Sanya's attractions from the Ctrip website and app. Through this data collection process, a total of 12,763 pieces of raw network data were obtained.

3.2. Research methods and processes

As shown in **Figure 1**, the article mainly uses two major research methods: data collection and data analysis to analyze and study the image perception of Sanya scenic spots. Firstly, data scraping of online customer reviews was performed using Python software. Then, data preprocessing was carried out by importing the scraped raw data into an Excel spreadsheet. Through data preprocessing, meaningless adverbs, modal particles, etc., in the raw data were deleted, and similar adverbs were merged. Finally, after rigorous screening, 12,000 valid reviews were obtained. Through ROSTCM6.0 research and analysis of high-frequency words and word clouds, three dimensions of group tourists' image perception of Sanya scenic spots were obtained.

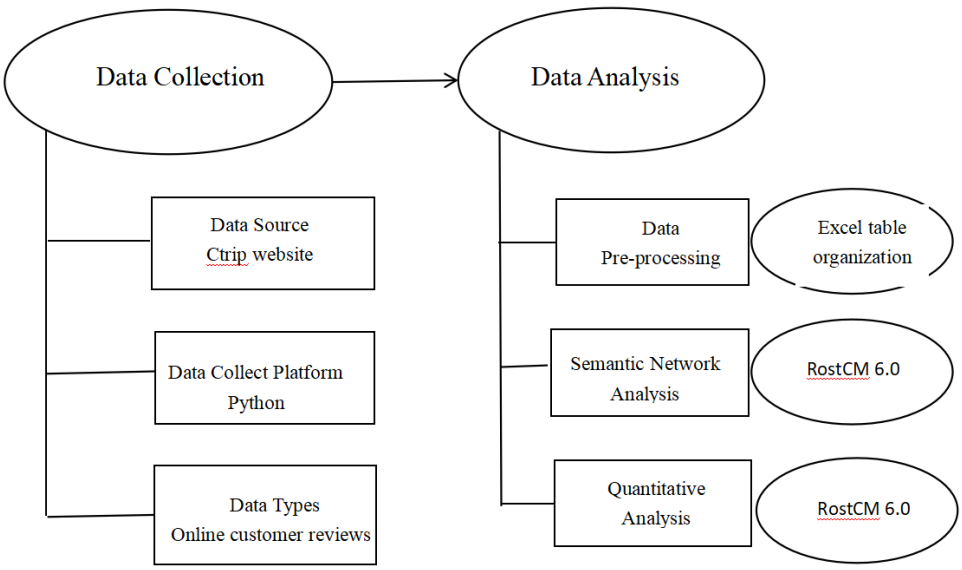

Figure 1. Flowchart of the research process

4. Research analysis

4.1. Analysis of high-frequency words in tourists' online reviews

Based on the analysis of high-frequency words in Table 1, group tourists' perception of the image of Sanya scenic spots is mainly manifested in three dimensions. These three dimensions are named the company service dimension, group atmosphere dimension, and natural scenery dimension.

Table 1. High-frequency words in group tourists' online reviews (top 50)

Serial number	High-frequency word	Word frequency	Serial number	High-frequency word	Word frequency
1	Guide	9,061	26	experience	641
2	journey	3,548	27	child	628
3	Sanya	3,380	28	care	616
4	arrange	3,160	29	intimate	594
5	Serve	2,671	30	shopping	576
6	hotel	2,629	31	first	560
7	satisfy	2,435	32	serious	549
8	explain	2,400	33	manner	516
9	happy	2,090	34	play	510
10	enthusiasm	2,001	35	funny	506
11	travel	1,767	36	mangroves	476
12	Hainan	1,305	37	worth	469
13	travel	1,303	38	elder	466
14	grateful	1,294	39	weather	462
15	Play	1,282	40	free	423
16	time	1,264	41	environment	422
17	Attractions	1,207	42	reviews	417
18	thoughtful	1,167	43	Knowledge	398
19	Reasonable	1,049	44	beauty	363
20	joy	845	45	parents	357
21	humor	808	46	scenery	355
22	family	785	47	room	339
23	patience	771	48	pretty	329
24	landscape	658	49	Consumption	320
25	careful	651	50	seafood	292

From the high-frequency words in Table 1, it can be seen that "tour guide" ranks first with a frequency of 9,061 times, indicating that "tour guide" is crucial for tourism companies. Secondly, "arrangement" ranks fourth with a frequency of 3,160 times, and "time" ranks 16th with a frequency of 1,264 times. These high-frequency words suggest that group tour companies need to make relevant schedule arrangements and travel route planning ahead of time, such as "two days and one night" or "three days and two nights" schedules. These high-frequency words reflect that tourists value the scheduled arrangement and travel route planning.

It can also be seen from the high-frequency words in Table 1 that tourists attach great importance to the group atmosphere. Among them, "satisfied" ranks seventh with a frequency of 2,435 times, "enthusiastic" ranks tenth with a frequency of 2,001 times, "happy" ranks 20th with a frequency of 845 times, "humorous" ranks 21st with a frequency of 808 times, and "witty" ranks 35th with a frequency of 506 times. The appearance of these high-frequency words indicates that group tour tourists value the atmosphere very much. A good group atmosphere can make tourists relax and enhance their stickiness. The word cloud derived from high-frequency words is presented

in Figure 2 below.

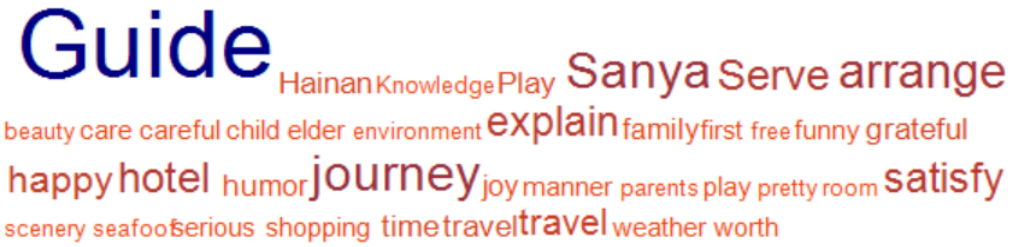

Figure 2. Word cloud of high-frequency words in online reviews of Sanya scenic spots

4.2. Social network semantics

Social network semantics is a visual representation of tourists' online review texts, where the connections between words reflect their mutual relationships [11]. The words "tour guide," "Sanya," "happy," "arrangement," and "service" are closely connected to other words, indicating their strong centrality and prominence in the network semantics. As shown in Figure 3 below, the social network semantic analysis diagram of group tourists' online reviews of Sanya scenic spots visualizes the tourists' online text reviews.

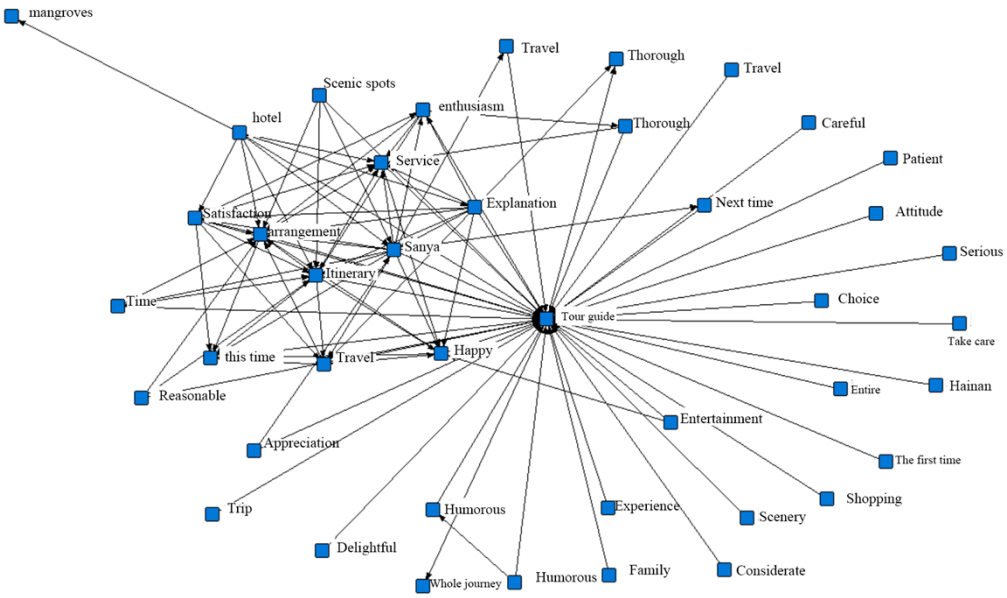

Figure 3. Semantic analysis diagram of online reviews social network for tourists

5. Results and discussion

5.1. Results

Through the visualization of high-frequency words, word clouds, and social network analysis diagrams of online tourist reviews, three dimensions of group tourists' cognitive image of Sanya scenic areas can be identified. These three dimensions are named the company service dimension, group atmosphere dimension, and natural scenery

dimension.

In the cognition of the company service dimension, it can be seen from the high-frequency words “tour guide,” “thoughtful,” and “careful” that group tourists value the service quality of tourism companies very much. It can also be seen from Figure 3, the semantic analysis diagram of the online review social network for group tourists, that the words “tour guide” and “arrangement” have strong centrality and are closely connected with other words. At the same time, the high-frequency words “elderly” and “children” appear in Table 1, indicating that family group tours and parent-child group tours are important components of group tours. Therefore, tourism companies should make advance preparations for service systems and training for special groups, with a “people-oriented” approach that takes care of tourists’ emotional value and experience. For group tours of special groups, tourism companies need to be equipped with relevant doctors to accompany them, so that tourists have positive evaluations of the company’s service dimension.

In the cognition of the group atmosphere dimension, it can be seen from Figure 2, the high-frequency word cloud diagram of online reviews for Sanya scenic areas, that words such as “happy,” “enthusiastic,” and “satisfied” are related to the group atmosphere dimension. Therefore, it can be seen that the group atmosphere is also a critical part for group tour tourists. A relaxed and pleasant group atmosphere during the trip plays a significant role in enhancing tourists’ satisfaction, increasing their intention to revisit, and enhancing their loyalty.

In the cognition of the natural scenery dimension, it can be seen from high-frequency words such as “scenic spots,” “scenery,” and “mangroves” that natural scenery is an indispensable part of the tourist experience. Appreciating the natural scenery of the tourist destination has a significant impact on tourists’ mood and perception of the destination. It can be seen from Figure 3 that “Sanya” and “scenic spots” have strong centrality and are closely connected with other words.

5.2. Discussion

This article mainly collects tourism evaluation data of group tourists on the Ctrip travel website through Python web scraping software, from the unique perspective of group tourists. It uses the text analysis software ROST-CM6.0 to study the tourism image perception of Sanya scenic spots through two aspects: word frequency and network semantic visualization.

Through analysis and research, three dimensions of group tourists’ cognitive image of Sanya scenic spots are identified, namely the company service dimension, group atmosphere dimension, and natural scenery dimension. Firstly, from the analysis of high-frequency words, word clouds, and network semantics, it is found that the word “tour guide” appears most frequently in group tour companies. Therefore, training for tour guides is very important and should be prioritized. Among them, “explanation” and “culture” also appear frequently. Therefore, the training of tour guides should not only focus on service etiquette but also strengthen their cultural knowledge of tourist destinations and enhance their ability to explain tourist destinations.

Secondly, the appearance of high-frequency words such as “elderly” and “children” indicates that there are more and more tours for the elderly, families, and parent-child groups. Therefore, tourism companies need to be equipped with relevant medical staff ahead of time, so that they can solve any physical problems that may arise for the elderly or children during the trip. This can provide a sense of security for group tourists and increase user loyalty.

Finally, the group atmosphere of group tourists is also particularly important. As mentioned in the three

dimensions of cognitive image, the group atmosphere can increase the bond between group members. A pleasant, humorous, and relaxed group is more conducive to experiencing and appreciating natural scenery. It can leave a good impression and enhance tourists' loyalty.

5.3. Research limitations

Firstly, the article only adopts online tourism evaluation data from group tourists to the Sanya scenic area, and the data collected is relatively homogeneous. Future research can increase the diversity of data channels, such as offline questionnaire surveys and destination interview studies.

Secondly, the analysis and verification methods used in the article are limited to big data text analysis and network semantic visualization. Future analyses can utilize relevant software such as SPSS to add confirmatory factor analysis, making the overall analysis more scientific, reasonable, and complete. This will enable a more comprehensive understanding of group tourists' perception of the tourism image of Sanya scenic areas.

Funding

School Scientific Research Startup Project (2023xskq011), Huangshan University School-Level Projects; Anhui Provincial Key Laboratory Project of Regional Culture and Smart Tourism Integration Effect (WLSYS202402)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Liu S, Tao Z, Xia J, 2024, Research on the Reform Path of "Tourism Reception Industry" Course from the Perspective of Integration of Specialization and Innovation. Scientific Consulting (Education Research), (03): 138–142.

[2] Zhang J, Chen Z, 2024, Exploring Human Resource Management Digital Transformation in the Digital Age. Journal of the Knowledge Economy, 15(1): 1482–1498.

[3] Pencarelli T, 2020, The Digital Revolution in the Travel and Tourism Industry. Information Technology & Tourism, 22(3): 455–476.

[4] Yang W, Zhang Y, So KKF, 2021, Tourism Experiences vs. Material Purchases: Effects of Eudaimonic Consumption Motive on Consumers' Reactions to Invidious Comparisons. Tourism Management, 83: 104247.

[5] Hong T, 2024, Thoughts on Adjusting Consumption Policies in the New Cycle to Cope with the "Two Temperature Differences." Journal of Hebei University of Economics and Business, 45(03): 31–40.

[6] Chen S, Li P, Hui Q, 2022, Research on the Experience Perception of Tourists Entering Hainan from the Perspective of Big Data—Based on the Network Text Analysis of Famous Scenic Spots in Sanya. Special Zone Economy, 2022(12): 157–160.

[7] Liao Q, 2023, Research on Coastal Tourism Attractiveness Based on Tourist Perceived Value—Taking Sanya City as an Example, dissertation, Hainan Normal University.

[8] Hunt JD, 1975, Image as a Factor in Tourism Development. Journal of Travel Research, 13(3): 1–7.

[9] Huang C, Huang Y, Yu H, et al., 2022, Research on Perceived Image of Guilin Tourism Destination Based on LDA and Grounded Theory. Areal Research and Development, 41(05): 91–97.

[10] Shi R, 2022, Research on the Impact of the Particularity of Sports Event Tourism on Tourists' Intention to Revisit—An Empirical Analysis Based on Structural Equation Modeling. Price Theory and Practice, (02): 192–195 + 203.

[11] Zhu Q, Yu J, Wang M, et al., 2015, Visualization of Social Network Data. Journal of Jilin University (Information Science Edition), 33(05): 584–587.