Research on Tourist Satisfaction and Improvement Measures of Wuzhizhou Island Based on Data Mining

Ting Lin*, Qiumi Chen, Ling Wang, Mengyao Chen

Hainan Tropical Ocean University, Sanya, Hainan, 572000, China

*865336416@qq.com

Abstract. As the main service subjects in the tourism industry, the deviation between tourist expectations and the experiential value will directly affect the development of tourist destinations. By using the LDA topic analysis method, 10 tourist attention topics were extracted from tourist comments on Wuzhizhou Island, and the attention degree of each topic was measured. Then, sentiment analysis tools were used to obtain the emotional tendency of tourists from the attention topics, and then to objectively analyze the views of the attention subjects. Finally, the IPA model was used to study the relationship between the weight of each topic and tourist satisfaction. The research results show that the three topics of scenic beauty, transportation within the scenic area, and aesthetic value performed well in terms of tourist attention and satisfaction, highlighting their importance in the tourism of Wuzhizhou Island. However, satisfaction with shopping consumption, route arrangement, and hotel accommodation is relatively low, and improvement and enhancement of service quality are needed.

Keywords: Wuzhizhou Island; data mining; sentiment analysis; LDA topic analysis

1   Introduction

In recent years, many scholars have used big data methods to study user satisfaction and its influencing factors in multiple fields. In domestic research, Ouyang Enshan[1] optimized the perception satisfaction of videos through DPI big data analysis method in 2022; Xia Yingying[2] conducted a study on improving outpatient satisfaction in hospitals using big data in 2020, fully reflecting the progress and promotion of doctor-patient relationships and hospital reforms; Zhu Linqi[3], in 2020, constructed a big data platform for tourist satisfaction based on Hadoop and analyzed the attitude of tourists towards folk villages. Based on satisfaction and dissatisfaction, suggestions for the future development direction of the folk village were proposed; Gao Bo[4] conducted a study on 4S shop service satisfaction and personnel management based on big data in 2021, established the relationship between satisfaction scores and dealership operating indicators based on big data analysis, and provided management measures for the personnel to improve service quality and service output from the personnel aspect.In foreign research, Gianluca Elia[5] assessed online courses through big data in

2019 and provided improvement methods, supplementing the shortcomings of collecting feedback (such as questionnaires or surveys), pointing out that the application of big data in online learning (CL) is a recent research area, which may play an important role in future online education;

Looking at the research on tourist satisfaction at home and abroad, it can be found that research on satisfaction has shifted from qualitative to quantitative research, and there is still relatively little research on the evaluation system of tourist satisfaction and influencing factors of tourist destinations. With the help of data mining technology, this study deeply investigates the topics that tourists are concerned about and their emotional tendencies. The research aims to provide data and analytical conclusions on tourist experiences for national island managers and provide targeted improvement suggestions for the managers of Wuzhizhou Island.

2   Research Methods and Data Processing

2.1   Data Collection and Processing

The research data comes from Ctrip, Mafengwo, and Qunar, which are ranked among the top three comprehensive rankings by Alexa. Data collection was carried out using the automated web data collection tool "Octopus". Since the comment data is unstructured data, there are a large number of redundant words, emoticons, and invalid comments. Direct analysis would result in poor text mining effects, thus requiring data preprocessing such as data deduplication, deletion of short comments, custom word library creation, text segmentation, removal of stop words, and synonym replacement. After deleting meaningless, repeated, and automatically posted comments, 3256 valid comments were collected. Subsequently, the Jieba word segmentation tool in Python was used, and stop words were set to remove stop words and process the research data.

2.2   Topic Modeling and Sentiment Analysis

This study uses Latent Dirichlet Allocation (LDA) to extract latent topics from the texts. Since the number of LDA topics cannot be directly obtained, the study uses perplexity data to determine the best number of topics for LDA. Through the "Word Cloud" tool, a perplexity line graph is drawn to obtain the best number of topics, and then the best number of topics is input into the LDA model for testing, followed by visual analysis to obtain the most probable feature words covered under each topic. At the same time, the "Word Cloud" sentiment analysis tool quickly obtains the overall sentiment analysis situation of the texts.

2.3   Tourist Satisfaction Analysis

This study uses the Importance Performance Analysis (IPA) model to analyze tourist satisfaction. The IPA model is widely used to analyze tourist satisfaction, and the measurement criteria include two aspects: the performance level of each indicator and

the importance level. The IPA model architecture uses importance as the abscissa and performance, namely satisfaction, as the ordinate. The average values of both are taken as the dividing points of the X-Y axis, dividing the space into four quadrants. This model helps operators understand customer satisfaction and can clearly show the specific situation of the satisfaction and importance of each attribute through the IPA quadrant map, quickly grasping which aspects should be improved and maintained, and is simple and practical.

3   Research Results

3.1   Word Frequency Analysis

Word frequency analysis is a content analysis method for discovering the focus of text and finding key words in text. The higher the frequency of a word, the higher the perception of tourists towards that item, and it is easier for this item to influence the overall evaluation of Wuzhizhou Island by tourists.

From the distribution of high-frequency words, tourists are most concerned about the scenery of Wuzhizhou Island, with high attention to keywords such as "Wuzhizhou Island", "sea water", and "scenery". Secondly, for the island's attractions, keywords such as "electric car", "project", and "tour around the island" are also very concerned. Emotion words in the word cloud such as "good", "acceptable", and "not bad" are all high-frequency words, indicating that the overall evaluation of tourists is satisfactory."

3.2   Sentiment Analysis

The "Micro Word Cloud" sentiment analysis tool provides three scales for sentiment analysis: positive, negative, and neutral. According to the data results, it is calculated that 75.42% of the comments are positive, 13.56% are neutral, and 11.02% are negative. The analysis results also conform to the research results of individual tourist memory emotional responses, showing an overall positive trend, consistent with the above word frequency analysis results. As shown in Fig 1.
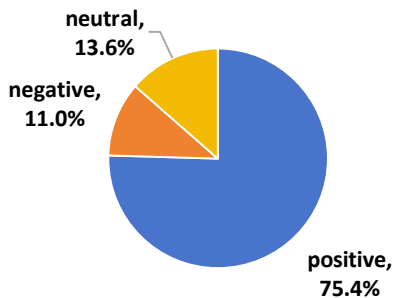

Fig. 1. Text Sentiment Analysis Situation

3.3   Topic Extraction

Topic extraction requires the use of the LDA topic model to cluster the pre-processed content of ethnic museum reviews and explore related topics and key feature words. The LDA model needs to determine the optimal number of topics K to ensure the effectiveness of topic extraction. Using the "Micro Word Cloud" LDA topic analysis tool, when the number of topics K=10, local minimum perplexity is obtained, indicating better clustering effectiveness.

Afterwards, we will input the number of topics into the LDA model and select the top 10 keywords with the highest appearance probability from the generated topic keywords. This will help us determine the core features of each topic. Subsequently, we will identify the meanings of the topic keywords contained in each topic and analyze the logical relationships between these words. This will help us name each topic and understand the concepts they represent. At the same time, we will conduct sentiment analysis to associate the emotional tendency of each comment with the corresponding topic. This will help us understand the positive or negative nature of each topic and the satisfaction of visitors under each topic. Finally, through statistical analysis, we will obtain the frequency of appearance of each topic in all texts and the proportion of positive comments. This will provide a more comprehensive understanding, helping us identify which topics have positively affected visitor satisfaction. The specific results are shown in Table 1.

Table 1. Topic Classification and Topic Keywords

Topic Category	Topic Words	Topic Importance	Good Comments Rate
Scenic views	Wuzhizhou Island, sea water, scenery, landscape, attractions, Sanya, beach	38.20%	84.50%
Entertainment projects	Electric car, project, play, diving, on the island, circular island tour	21.30%	67.30%
Aesthetic value	Electric car, project, play, diving, on the island, circular island tour	13.30%	78.20%
Scenic transportation	Electric car, circular island, driver, tourist car, circular island tour, getting off, driving	11.30%	90.80%
Scenic services	Explanation, service, driver, queuing, staff, service area, guide, coach	6.00%	87.60%
Hotel accommodation	Hotel, accommodation, environment, service, service area	1.60%	68.90%
Recognition	China, Hainan, Sanya, Maldives, beach	4.20%	87.10%
Shopping consumption	Project, service, ticket, shopping	1.30%	50.0%
Scenic facilities	Queue, staff, service area, guide	1.20%	79.40%
Route arrangement	Queue, circular island, guide	1.60%	64.40%

3.4   IPA Model Analysis

To measure the impact of each topic on tourists' satisfaction, this study uses the IPA model to analyze the two indicators of topic importance and satisfaction. Satisfaction is represented by the good comments rate of each topic. The importance indicator is represented by the frequency of appearance of each topic in the corpus, with higher appearance frequency indicating higher attention and importance of the topic to tourists. By entering the data of topic importance and satisfaction rate into SPSS, the four-quadrant scatter plot is drawn. Fig 2 shows the data results:
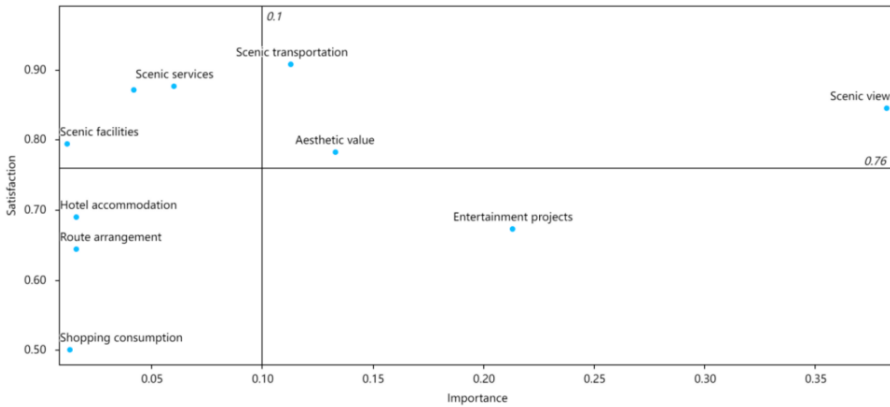

Fig. 2. Quadrant Diagram of Importance and Satisfaction

The first quadrant is the dominant area, where both the importance and satisfaction of the topics are relatively high. Scenic views, aesthetic value, and scenic transportation are located in the first quadrant, which are strengths of Wuzhizhou Island and should continue to be maintained. These three topics play an important role in promoting tourist satisfaction, as Wuzhizhou Island has beautiful scenery, clear seawater, and convenient transportation due to its smaller island size. Therefore, the conclusion is as expected.

In the second quadrant, all topics have relatively high tourist satisfaction but relatively low importance. This quadrant includes recognition, scenic services, and scenic facilities. Although tourists perceive these topics as less important, they have had a good experience with these topics during the tour. For topics in this area, scenic managers should pay attention to maintaining these topics and further explore them to transform them into competitive advantages for the scenic area.

The topics in the third quadrant have both lower than average tourist satisfaction and perceived importance. This quadrant includes hotel accommodation, route arrangement, and shopping consumption. In this quadrant, the satisfaction of shopping consumption lags behind noticeably. The topics in this quadrant have lower priority and do not require major development, but they are not completely insignificant. With limited resources, they are not suitable for prioritized development, but their disadvantages

need improvement as the future market changes. At the same time, hotel accommodation and route arrangement are factors that need to be considered.

The topics in the fourth quadrant have higher scores in terms of importance but lower than average tourist satisfaction, indicating areas that scenic managers urgently need to improve. This quadrant mainly consists of scenic entertainment projects. Many comments reflect issues such as expensive charges for entertainment projects and long waiting times, which are areas that the scenic area should focus on improving.

4   Conclusion and Recommendations

4.1   Conclusion

Based on the LDA topic model, this study extracts 10 topics from the online text reviews corpus of Wuzhizhou Island tourism, finding that tourists perceive scenic views, entertainment projects, aesthetic value, and scenic transportation as topics of relatively high importance during their visit to Wuzhizhou Island.

Based on sentiment analysis and topic extraction results, in-depth exploration of the impact mechanisms of various topics on tourist satisfaction was conducted using the IPA model. The study found that scenic views, aesthetic value, and scenic facilities are strengths of Wuzhizhou Island, which has a high resource endowment and, due to early development, has provided tourists with a good tourist experience through its leisure activities. The scenic area should pay attention to maintaining the ecological environment of Wuzhizhou Island and maintaining the quality of tourist products. In addition, shopping consumption is the topic with the lowest satisfaction feedback for Wuzhizhou Island and an area where scenic managers urgently need to improve.

4.2   Recommendations

4.2.1 Adjust Scenic Area Consumption Pricing and Change Marketing Strategies.
Due to the high ticket prices of Wuzhizhou Island's 5A scenic spot, many tourists find it unaffordable. Therefore, the scenic area can adjust pricing strategies in the low and peak seasons. In the low season, the scenic area can moderately reduce ticket prices, especially as Wuzhizhou Island is the leading scenic spot in Sanya. Lowering ticket prices can stimulate more tourists to visit Wuzhizhou Island's 5A scenic spot during the off-season and increase the number of off-season tourists. In the peak season, ticket prices can be adjusted appropriately based on past years' data analysis, combined with the development of local folk culture to produce practical, artistic, precious, commemorative, and modern products. When tourists enter the exit shops, provide valuable merchandise and introduce these souvenirs. Active promotion activities can stimulate the desire to purchase, direct the attention of tourists to other tourist products, and promote the sale of tourist goods. According to big data reviews, over 73% of tourists are dissatisfied with the dining prices and taste at Wuzhizhou Island's scenic spot, as internal transport and storage issues on the island have led to higher dining prices. Moreover, differences in dietary habits are problems faced by scenic spots across the country. Therefore, the scenic area needs to reasonably adjust the pricing of dining

prices, enrich the variety of meals, and meet the dining needs of visitors from different regions, making them feel a sense of belonging.

4.2.2 Standardize Entertainment Project Management and Rational Layout of Projects.

Enrich entertainment projects: Introduce more diversified entertainment projects, such as water sports, surfing, sailing, yachting, etc., to meet the needs of different tourists; Plan and launch themed activities, such as night lights show, cultural festivals, music concerts, and film festivals, allowing tourists to experience more diverse cultural and entertainment activities during their visits; Reasonable pricing strategy: In-depth understanding of market conditions and tourist needs, develop more reasonable pricing strategies to avoid high prices affecting tourists' willingness to participate; Provide different levels of packages and promotional activities, allowing tourists to choose according to their budgets and needs; Improve service quality: Strengthen employee training and management, improve service quality and professionalism, ensure that tourists have a good experience while enjoying entertainment projects. Establish a sound customer feedback mechanism, collect and process customer opinions and suggestions in a timely manner, and continuously improve and optimize services.

4.2.3 Enhance Hotel Accommodation Quality and Offer Customized Services.

Ensure room cleanliness, comfort, and good hygiene conditions. The quality of bedding and toiletries should also be emphasized, ensuring that guests have a good sleep and pleasant bathing experience; Provide high-speed stable wireless networks and premium TV programs to meet guests' entertainment needs. Introduce intelligent devices to speed up check-in and check-out processes, reducing guest waiting time; Prior to guest check-in, understand their special requirements, such as breakfast preferences and room layout requirements, and provide corresponding customized services during guest stays; Provide personalized services, such as providing customized travel information and advice based on guest identities and needs. Adjust the types and quality of bedding and toiletries based on seasonal changes to meet different guest needs. Diverse hotel facilities and activities: Provide various facilities, such as restaurants, gyms, swimming pools, and spas, to meet the needs of different customers. Organize various outdoor activities and tourism projects based on the natural landscapes and cultural characteristics of Wuzhizhou Island, such as diving, snorkeling, whale watching, etc., allowing guests to fully experience the charm of Wuzhizhou Island.

Acknowledgment

References

1. Ouyang Enshan, Fang Yungen. Video perception satisfaction optimization based on DPI big data analysis[J]. Mobile Communications, 2022, 46(8):6.
2. Xia Yingying, Cui Xiao, Meng Lingying, et al. The role of hospital patient satisfaction survey in hospital management under the background of big data[J]. Special Health, 2020, Issue 29, page 111, 2020.
3. Zhu Linqi, Zou Wei. Research on the construction of big data platform for tourist satisfaction based on Hadoop[J]. Information Recording Materials, 2020, 21(7):2.
4. Gao Bo, Wang Xiaofeng, Guo Sibo, et al. Analysis of 4S store service satisfaction and personnel management based on big data[J]. Automobile and Driving Maintenance: Maintenance Edition, 2021.
5. Elia G ,Solazzo G,Lorenzo G , et al. Assessing learners' satisfaction in collaborative online courses through a big data approach.2019.