Big Data Promotes the Balanced Development Path of Rural Tourism in Small Areas – An Empirical Study on Dali City Based on K-Means Cluster Analysis

Wei Sun[1(✉)] and Yaqiao Zhao[2]

[1] School of Economics and Management, Yunnan Agricultural University, Kunming, China
15195865589@qq.com
[2] Yunnan Plateau Characteristic Agricultural Industry Research Institute, Kunming, China
yaqiao@163.net

Abstract. The balanced development of rural tourism plays an important role in realizing common prosperity for local people. Taking Dali City as an example, this paper studies how big data can promote the balanced development of rural tourism in small areas. Through field research in Dali City, 217 questionnaires were collected, and the analysis found that tourists to Dali City can be divided into the following five categories: traditional elderly tourists; young tourists with high potential for leisure punching in; middle-aged tourists with quality service type; young and middle-aged tourists with high freedom of experience; power middle-aged tourists. According to the characteristics of different groups, this paper puts forward different big data drainage methods, so as to provide ideas for realizing the balanced development of rural tourism in small regions.

Keywords: Rural tourism · The balanced development · Big data · K-Means cluster analysis · Tourist Satisfaction

1 Introduction

Rural tourism is a new industrial integration development mode. Its most basic feature is to promote the development of tourism based on the unique local resource endowment in rural space, so as to drive rural development and increase farmers' income. However, the spatial distribution of the content and quality of Chinese existing tourism resources is unbalanced. From the perspective of micro, the significance of the unbalanced distribution of rural tourism resources in a small region decreases, which makes it possible to shorten the economic development gap in a small region and promote the common prosperity of local people. However, how to promote the balanced development of rural tourism through specific ways has become an urgent problem to be solved. In recent years, the development of networking and information technology has continuously promoted the evolution of tourism informatization. The classification and exploration of big data on tourists' social and economic basic characteristics, tourism needs and tourism motivations are an important exploration to guide regional balanced development from the perspective of meeting diverse tourists' needs.

The causes of unbalanced development of tourism have been discussed a lot. It is generally believed that regional differences are a universal spatial feature of human social and economic activities, including rural tourism [1], and attributed to a series of natural and human factors such as regional resource conditions, market source of tourists, regional social and economic development level and other factors [2]. In summary, this study believes that natural resource endowment and historical and cultural factors are the "innate factors" that lead to regional differences, while others are the "acquired" human behavior choices on this basis. At the same time, existing researches on regional unbalanced development are more focused on the medium macro perspectives, based on national statistical data, inter-provincial statistical data or statistical data within a certain river basin. For example, Yanqin Lv made a quantitative study on the distribution, gap and dynamic evolution of the development level of tourism facilities in China based on the inter-provincial panel data of China from 2006 to 2019 by means of entropy weight method, Dagum Gini coefficient decomposition method, Kernel kernel density estimation method and Markov chain [2]. Based on the time series data from 2006 to 2016, Zhu Haiyan used primacy and frequency analysis to judge the rank-scale rule and fit the applicability of regional tourism scale structure with high primacy [3]. In order to better distinguish the existing research scope and scale, this kind of research scope and scale based on a certain city or county or the whole tourism landscape area as the research object from a micro perspective is called "small area".

2 Present Situation and Problems of Unbalanced Development of Rural Tourism in Dali City

The unbalanced development of rural tourism in Dali City has four manifestations. The first level is the imbalanced distribution of rural tourism resource points in city of Dali. The second level shows that the number of tourists in a small area of Dali City is more dense than other areas in this area, which means a certain area has stronger tourism attraction. The third level shows that the small area of Dali City has obvious low and high season in terms of the number of tourists received, which means the tourist attraction of a certain period of a year is far more than that of other periods. The fourth level is based on the above three levels. Due to the stronger attraction of time and space in a certain region, the local tourism resources have stronger market competitiveness, thus bringing more tourism income to the local area, that is, the local tourism income far exceeds that of other regions.

3 Big Data Guide Tourism Balanced Development Path Exploration

In order to solve this problem, we conducted a field survey and collected 217 questionnaires. By analyzing the characteristics of tourists and their preferences, we explored specific ways for big data to promote the balanced development of rural tourism in Dali City.

Table 1. Characteristics of clustering results

| | Cluster analysis results | | | | |
	1	2	3	4	5
Age	Age 61 and above	Age 30 and below	31–40 years old	Age 30 and below	31–40 years old
Tourism-generating region	Other Provinces	Other regions of Yunnan Province	Other Provinces	Other regions of Yunnan Province	Other Provinces
Occupation	Retiree	students	else	else	employees of the enterprises
Educational background	Junior high school, secondary school	Colleges and Universities	Colleges and Universities	Colleges and Universities	Colleges and Universities
Annual income level	50,000–80,000 yuan	<50,000 yuan	>150,000 yuan	50,000–80,000 yuan	100,000–150,000 yuan
Playmate	Travel with Family	Travel with Friends	Travel with Friends	Travel with Friends	Travel with Family
Popular scenic spots Attention	Barely paying attention	Occasional attention	Occasional attention	Occasional attention	Occasional attention

Table 2 Note: Not following means "never looking at travel-related information"; Little attention means "almost no travel-related information"; Occasional attention indicates that "the network platform will read the push, but will not take the initiative to fin

3.1 Cluster Analysis of Tourist Characteristics

Tourists are the main participants of tourism activities and the direct experience in tourism services. Their attitudes and behaviors are directly related to the development of rural tourism. K-means clustering analysis method uses K as the parameter to divide the n objects into K classes, so that the objects within the cluster have as much similarity as possible, while the similarity between clusters is as small as possible [4]. In order to better classify the internet-famous site in Dali City and S Bay according to their demographic characteristics, we divided them into five categories after multiple attempts and comparison of their significance (see Table 1).

3.2 Motivation Preference Analysis of Different Cluster Members

Tourism decision-making can be generally divided into three aspects: "pre-tourism decision-making, in-tourism decision-making, and post-tourism decision-making". This study adopts a five-point Likert scale to measure the importance degree of agricultural tourism product features of tourists at different clustering levels before they go out (see No. RR1–RR2 in Table 2).

According to the ranking results of the basic characteristics of tourists and the importance of various factors of tourism, the tourists in cluster 1 are mainly 61 years old and above, and they have a demand for the natural landscape, unique culture and efficient service mechanisms, while they have a low demand for catering. The tourists in cluster 2 and cluster 4 are both young groups under 30 years old and have certain similarities in their specific tourism demand ranking. Both have a high demand for efficient service mechanisms and unique cultures. The difference lies in that cluster 2 is dominated by students

Table 2. Scale of the importance of agricultural tourism product features

Serial number	Issue	Options
RR1	Ecological environment, natural scenery, landscape pastoral landscape	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC1	Reasonable and cost-effective price level (accommodation, catering and other activities)	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC2	Activities for different groups of people	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC3	With local characteristics, more selective catering services	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC4	Attractive local products or souvenirs	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC5	Convenient and varied transportation conditions	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RC6	Service staff can provide timely and effective service (e.g. complaint handling, etc.)	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5
RR2	Unique traditional folk customs and cultural atmosphere	Very unimportant = 1 Unimportance = 2 Of general importance = 3 Important = 4Very important = 5

with low-income levels, while cluster 4 is dominated by high occupational freedom with high-income levels. Cluster 2 has a higher demand for specialty catering, while cluster 4 has a higher demand for diverse activities. Cluster 3 and Cluster 5 are mainly tourist groups between 31 and 40 years old. In terms of their specific tourism needs, efficient service mechanisms, characteristic catering and convenient transportation conditions are the first three. The difference lies in their specific ranking.

3.3 Analysis of Sources of Tourists' Cognitive Information

The cognitive behavior of rural tourism refers to that after the tourism consumption motivation is generated, tourists actively or passively accept relevant tourism product information, form a preliminary understanding of tourism products, and form an abstract product outline, which provides psychological and information basis for the occurrence of the next decision-making behavior and experience behavior [5] (Table 3).

From the perspective of tourists' information acquisition channels, tourists' information acquisition methods are diversified and concentrated. On the one hand, different

Table 3. Statistical table of cognitive sources of tourist information

Specific information channels	Dali City		S bay	
	Frequency	Percentage	Frequency	Percentage
Tik Tok	84	38.7%	108	49.8%
Quora	5	2.35%	10	4.6%
RED	70	32.3%	78	35.9%
moments	22	10.1%	15	6.9%
Micro Blog	19	8.8%	7	3.2%
Recommended by relatives and friends	47	21.7%	43	19.8%
Package Tour Recommendation	4	1.8%	2	0.9%
Other	75	34.6%	36	16.6%

types of information acquisition channels can attract tourists to a certain extent; On the other hand, compared with traditional information acquisition methods, modern online platforms are more able to attract tourists and are more inclined to Douyin and XiaoHongshu platforms.

4 Conclusions

According to the above research, big data drainage can be carried out from the perspective of different tourist groups' characteristics and preferences, so as to make regional development more balanced.

For Cluster 1 (traditional elderly tourists), outdoor health experience routes can be built to meet their tourism needs and health needs. For example, relying on the existing pastoral scenery, carrying out "pastoral+" activities, etc.

For Cluster 2 (young tourists with high potential for leisure punching in), due to they are mainly students and have limited economic foundation but hope to enjoy the most beautiful scenery and taste the most distinctive local catering, we can build Internet celebrity punching-in lines for them and build another punching in points based on building existing Internet celebrity points.

For Cluster 3 (middle-aged tourists with quality service type), because they attach more importance to the service quality and service experience in tourism, they can be more directed to niche and high-end tourism resources, such as Bai intangible cultural heritage experience and high-end health vacation.

For Cluster 4 (young and middle-aged tourists with high freedom of experience), because they have high freedom, and pay attention to tourism experience but have limited economic foundation, we can focus on launching medium and long-term service experience projects with relatively favorable prices. For example, we can build rural families into B&B experience areas in a planned way, so that they can live at a close distance from farmers and at the same time be relatively independent.

For Cluster 5 (power middle-aged tourists), which is similar to Cluster 3 but has higher economic strength, has stricter requirements on tourism service quality. Customized tourism service construction can be carried out for this group of people, which is required to meet the beautiful scenery, first-class service and reflect the unique charm of Cangshan and Erhai and the unique cultural flavor of Bai culture. To prevent tourists from congregating in the same area at the same time.

References

1. LI Tao, WANG Zhao, TAO Zhuo-min, SHEN Zhong-liang. Regional differences and mechanism of rural tourism development in Zhejiang and Shanxi provinces: A study from the perspective of industrial investment[J]. Journal of Natural Resources, 2022, 37(08):2051–2064.
2. Cao Fangdong, Wu Jiang, Xu Min, Jiang Haining. Exploration of Regional Tourism Difference Based on New System Economy[J]. Journal of Industrial Technological Economics, 2011, 30(03):25–31.
3. Lyu Yanqina, Wang Yijinb, Miruibangguli Maimaitib, Wang WeibSpatial Disequilibrium and Dynamic Evolution of Tourism Facilities Development Level in China[J]. Statistics & Decision, 2021, 37(12):52–56.
4. ZHU Haiyan, SUN Gennian. Has the domestic tourism of Hubei province been distributed balancedly: An analysis by applying the rank-size rule to high priority regions[J]. Journal of Zhejiang University (Science Edition), 2020, 47(2):244–252, 260
5. LI Ming. The Research and Application of Text Clustering Based on Improved K-means Algorithm[D]. Jiangnan University, 2018.
6. SUN Gen-nian, FENG Mao-e. Inbound tourism market competitive State and the relationship with resources and location factor in the West of China[J]. Journal of Northwest University (Natural Science Edition), 2003(04):459–464.