

Tourist Satisfaction in Ice And Snow Tourism Based on Web Text Data Analysis

Lin Shen^{1, a}

¹Sports Marketing, College of Art and Sports, Dong-A University, Busan, 49315, Korea

^ashenlinwww@gmail.com

Abstract

This study examined 50 travelogues to explore tourists' concerns and emotions regarding ice and snow tourism in Ma Hive. The study identified five categories of high-frequency words and used social network relationship clustering and multiple regression analysis to identify areas that require improvement. Tourists were primarily concerned about infrastructure, high costs, and inadequate amenities and facilities. The study found that most tourists had positive emotions, but some were not entirely satisfied. To enhance the tourist experience, attractions should provide better warming equipment, more visitor participation programs, and improve hawker management. The study recommends incorporating more Chinese culture into the scenic spots to enhance cultural diversity and overall experience for tourists.

Keywords

Tourist satisfaction, Snow tourism, Text data.

1. Need and Purpose of the Study

Pizam (1978) first introduced the concept of tourist satisfaction, which is the result of comparing a tourist's trip experience with their expectations. However, Churchill and Surprenant (1982) found that attribute performance can also impact satisfaction. Hence, scholars have focused on enhancing tourist satisfaction, as it is an evaluative emotion. Chen (2015) found that the tourism industry values being tourist-oriented and service-oriented, making it crucial to prioritize and improve tourism services.

Scholars have developed various methods to measure tourist satisfaction over time. This began with market survey questionnaires and gradually evolved to more scientific and rational techniques. For example, Dunn (1991) proposed evaluation criteria for tourists' motivation and satisfaction. Various methods have been used to analyze tourist satisfaction, including online reviews, questionnaires, and data mining techniques. Scholars have applied techniques such as the applied-factors-analysis method (FAMM), the fuzzy-comprehensive evaluation method (FCEM), and the importance-performance analysis (IPA) to evaluate tourist satisfaction. Brejla (2014) analyzed visitor reviews and satisfaction, and Cheng (2020) used word frequency analyses, semantic network analyses, and sentiment analysis methods to analyze visitor satisfaction.

2. Research Methodology

2.1. Object of research

Scholars have used Web Text Data Analysis Methods to analyze online travelogues, which are becoming increasingly important as tourists share their experiences on various online platforms. Sullivan (2001) considered these methods as Content Analysis Methods, which extract feature words from text to represent information, and can be used to investigate ice and

snow tourism satisfaction. Similarly, Zhang and Zhao (2018) found that online travelogues are a valuable source of information for understanding travel perceptions and preferences. Fang and Liu (2016) used content analysis to explore tourists' perceptions of Shaoshan as a destination. This study uses the Web Text Analysis Method to collect and analyze data from the top three ice and snow tourist attractions in China, selecting only the best travelogues for analysis. A total of 50 travelogues were collected for this study.

2.2. Research Methodology

The study collected data from five popular Chinese travel websites, and 50 travelogues were analyzed using various text analysis tools. The data was visualized using Gephi software, and the texts were coded and summarized using Nvivo according to Grounded Theory.

2.3. Data Processing

1) High frequency word selection

After text mining, a list of 7500 high-frequency words was created. Redundant words were excluded, and similar words were combined to avoid duplication. This resulted in 3500 high-frequency words, and 50 of these were selected.

2)Text de-duplication

To ensure reliability, we carefully screened texts and eliminated any with irrelevant content or advertisements. We then selected 50 high-quality texts with clear and informative content, high readership volume, and likes/retweets for further analysis.

3) Analysis of emotion words

This study analyzed the emotional attributes of 2811 sentences in 50 travelogues using an emotion dictionary. Values were assigned to high-frequency words and emotional sentences based on their lexical nature and emotional attributes. Comments, retweets, likes, and reads of each travelogue were also considered. The data was analyzed using SPSS26.0 and regression equation analysis was conducted to reflect the emotional analysis.

2.4. Categorization of This Paper

Previous studies by Cheng et al. (2020), Ao et al. (2020) analyzed high-frequency words related to ski and ice/snow tourism. They classified words into different categories, such as destination, services, and tourist feelings. In this study, Nvivo software was used to apply Grounded Theory's three-level coding to summarize the dimensions that affect tourist satisfaction.

3. Research Results

Table 1. High frequency table

HFW	WF	DF	HFW	WF	DF	HFW	WF	DF	HFW	WF	DF
specialized	41	14	gallery	28	17	Train station	29	3	Locke	20	4
Sophia	39	17	snow scene	27	12	double board	27	13	architecture	20	13
forest	39	21	plank road	27	9	Glasses	26	13	Villa	20	9
train	38	15	mushroom	31	14	temperature	26	7	Inn	20	12
cost	38	2	cable car	31	20	enjoy	25	15	style street	20	9
clothing	37	17	bus	31	15	Food	25	8	pot helmet	19	12
popsicle	37	22	bonfire	30	14	birch forest	24	15	spring snow	19	5
play	36	11	Room	30	17	snow rhyme	24	17	Wanda	16	10
home	34	13	dream	30	13	weather	24	14			
speed	34	10	feature	30	11	snow gear	24	13			
sledge	34	10	dumpling	30	5	Stimulate	22	12			
snow	32	15	beginner	29	15	ice sculpture	22	18			
breakfast	32	16	landscape	29	15	down jacket	22	8			
Dining room	29	14	view	29	17	child	22	12			

3.2. Sentiment analysis

Table 2. Sentiment analysis of sentences

Type of Emotion		Quantity (bars)	Percentage of(%)
Positive emotions		1931	68.69%
Neutral mood		551	19.12%
Negative emotions		329	11.70%
Positive Mood Segment Statistics	General(0—9)	229	11.85%
	Moderate(10—20)	314	16.26%
	Height(20 or more)	589	42.30%
Negative mood segmentation statistics	General(-10—0)	276	83.89%
	Moderate(-20—-11)	40	12.15%
	Height(Below -20)	13	3.90%

The majority of tourists have a high level of satisfaction with ice and snow tourism in Harbin, as per Table 2. However, a significant number of tourists expressed neutral or negative sentiments. To improve overall satisfaction, the industry should focus on enhancing services, facilities, providing engaging activities, and creating a welcoming atmosphere. Addressing specific issues can turn neutral or negative sentiments into positive ones, ensuring the industry's success in attracting tourists.

3.3. Analysis of Relevant Variables

Table 3. Text-related data

Total Score	Reading volume	Compliments	Collection	Total Score	Reading volume	Compliments	Collection
114	278957	3755	1528	31	393	9	6
104	71740	1344	1044	30	385	9	6
98	40293	320	8867	29	293	9	9
88	662	17773	3747	28	271	16	7
84	638	6856	232	25	267	15	12
72	604	4168	152	23	246	13	10
69	22121	234	99	18	380	12	5
65	14442	189	19	18	376	10	5
62	13160	155	19	16	351	10	4
60	1463	49	71	16	344	19	4
58	908	43	39	14	190	19	4
57	892	29	86	14	185	18	4
56	4550	29	76	14	169	9	4
51	2961	119	20	13	156	7	3
47	2925	109	18	12	136	7	2
42	548	38	18	11	135	7	4
41	531	36	17	10	67	3	3
39	440	30	17	10	63	3	4
39	415	62	14	8	116	7	4
39	234	61	14	8	116	6	2
36	222	26	12	8	83	4	2
35	199	26	11	8	130	7	1
34	194	24	11	3	118	6	1
33	407	21	13	0	51	2	0
32	400	20	13	-3	22	1	0

Table 3 presents data on reading volume, compliments, and collection of each of the 50 travelogues, along with their corresponding affective words scores. These scores were used to

assess visitors' travel experiences. The dependent variable was the total score, which generally correlated with higher reading volume, compliments, and collection. However, some travelogues had high scores in these three indicators but a low total score. The study used multiple regression analysis with SPSS 26.0 to explore the impact of independent variables on the dependent variable. Figures 3 and 4 were used to illustrate this relationship.

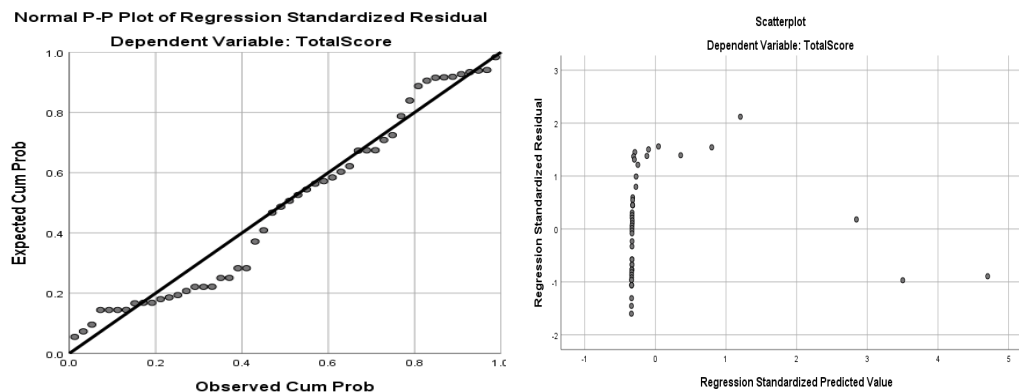


Figure 3. Scatterplot

It is evident from Figure 3 that there is some degree of bias in the residuals, but it is not severe enough to significantly impact the stability of the regression outcomes. Next, we investigate whether there is a significant linear relationship between the independent and dependent variables, as shown in Figure 4 and Figure 5.

Table 4. ANOVA analysis

ANOVAa					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	20321.267	3	6773.756	16.411	.000b
Residual	18986.513	46	412.75		
Total	39307.78	49			

a. Dependent Variable: TotalScore

b. Predictors: (Constant), Collection, ReadingVolume, Compliments

The results show that the F-test yielded a value of 16.411, with a p-value less than 0.001, indicating that the regression model passed the F-test at a 95% confidence level and the fitted equation was deemed statistically significant. Table 5. Correlation coefficient analysis.

Table 5. Correlation coefficient analysis

Coefficientsa							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	29.491	3.04		9.702	0		
ReadingVolume	0	0	0.422	3.962	0	0.927	1.079
Compliments	0.003	0.001	0.306	2.774	0.008	0.862	1.16
Collection	0.006	0.002	0.282	2.496	0.016	0.823	1.216

a. Dependent Variable: TotalScore

Based on the results, it can be concluded that all independent variables have statistically significant mean P values. Among them, ReadingVolume has the smallest P value, which indicates that it has the strongest impact on the Total Score. Moreover, all the VIF (variance inflation factor) values are less than 10, indicating that there is no significant covariance problem between the independent variables.

3.4. Dimensional Analysis

Table 6. Dimensional analysis

Dimensionality	Number of high frequency words	Frequency of high frequency words	Percentage of high frequency words
Ice and Snow Tourism Resources and Attractions category	14	502	28.00%
Ancillary Services	13	385	26%
Visitor Experience Category	8	207	16.00%
Visitor behaviour category	8	180	16.00%
Ice and Snow Tourism Image Category	7	134	14.00%

In this paper, the range of high-frequency words was redefined according to their attributes, resulting in five dimensions. As shown in Table 6.

4. Conclusion

The study analyzed text data from 50 travelogues of selected areas in Ma Hive, Ctrip, Go Where, Poor Travel, and Tuniu Travel using Gooseeker for text analysis. The study identified 50 high-frequency words, which were categorized into 5 dimensions of tourism resources, facilities and services, emotional perception, behavioral perception, and tourism image. The dimensions of tourists' concerns were classified into primary and secondary orders to facilitate targeted adjustments by scenic spots.

The most concerning topics for tourists were represented by word cloud maps, while social network relationship clustering analysis was performed using Gephi to highlight the most concerning parts of each cluster. The study found that reading volume was the main factor affecting the total score, indicating that the recognition of the ice and snow tourism places represented by the travelogue was strongly influenced by the number of reads.

The study also analyzed the emotional attributes of the text data sentences and found that most tourists had positive emotions (68.69%), followed by neutral emotions (19.12%), and negative emotions (11.70%). However, certain issues such as long waiting times, inadequate infrastructure for service convenience, high prices, and low temperatures negatively impacted tourists' experience and satisfaction.

Furthermore, the study revealed that high-frequency words related to architecture, Locke, and churches appeared more frequently, as did natural landscapes such as birch forests. However, there was little mention of Chinese cultural characteristics, indicating an area for improvement in promoting the uniqueness of the area.

Overall, the study provides valuable insights into tourists' concerns and emotions in the ice and snow tourism places, highlighting areas for improvement in the tourism industry.

5. Suggestions

1. Tourists have concerns about infrastructure and services in the ice and snow tourism areas. Long waiting times, inadequate service areas, and the need for better amenities and facilities were highlighted as areas for improvement to enhance the tourists' experience.

2. The high cost of visiting the scenic spot was found to be a negative factor impacting tourists' satisfaction. To improve satisfaction, differential pricing strategies, discounts for multiple purchases, and family travel packages can be implemented to stimulate visitors' willingness to revisit and attract more tourists.

3. The study revealed that tourists' emotions were mostly positive, but some areas require improvement to enhance their satisfaction. To improve the overall experience, tourist attractions should focus on providing adequate warming equipment, more visitor participation programs, and strengthening management to enhance the integrity of hawkers' operations.

4. Finally, the study suggests that promoting the uniqueness of the area's Chinese cultural characteristics could enhance the area's appeal to tourists. Therefore, it is recommended to incorporate more elements of Chinese culture into the scenic spots to promote the area's cultural diversity and enhance tourists' overall experience.

References

- [1] Pizam, A. (1978). Tourism's impacts: The social costs to the destination community as perceived by its residents. *Journal of Travel Research*, 16(4), 8-12.
- [2] Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460-469.
- [3] Churchill, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of Marketing Research*, 19(4), 491-504.
- [4] Dunn Ross, E. L., & Iso-Ahola, S. E. (1991). Sightseeing tourists motivation and satisfaction. *Annals of Tourism Research*, 18(2), 226-237.
- [5] Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management*, 26(1), 45-56.
- [6] Kang, T., Yu, C., Bao, X., et al. (2012). A study of theme park visitor satisfaction based on fuzzy comprehensive evaluation: Wuhu Fantasy Fun World as an example. *Resource Science*, 34(5), 973-980.
- [7] Brejla, P., & Gilbert, D. (2014). An exploratory use of web content analysis to understand cruise tourism services. *International Journal of Tourism Research*, 16(2), 157-168.
- [8] Fang, W., Wang, L., & Xu, Y. (2016). Research on factors influencing tourists' ecotourism satisfaction. *Business Research*, (11), 168-176.
- [9] Zhang, C., & Bai, K. (2018). Research on the classification of elements of tourist destination satisfaction attributes based on Tetra-class model - taking Shanxi Taibai Mountain National Forest Park as an example. *Economic Management*, (5), 123-140.
- [10] Liu, J., & Jin, M. Y. (2018). Analysis of tourist satisfaction of Beijing historical monuments scenic spots based on IPA method. *Journal of Beijing Union University (Humanities and Social Sciences Edition)*, (1), 38-45.
- [11] Li, Z. G., & Sun, Y. (2020). Research on tourist satisfaction of tourism public services under the new crown pneumonia epidemic based on IPA analysis. *Journal of Tianjin University of Commerce*, (5), 17-22.

- [12] Sun, X., & Ni, R. (2018). Product perceptions, emotional expressions and brand image perceptions of Chinese cruise tourists - a content analysis based on online reviews. *Geography Research*, (6), 1159-1180.
- [13] Zhao, C., & Chen, M. (2019). Analysis of factors influencing tourists' satisfaction based on web text analysis. *Statistics and Decision Making*, (13), 115-118.