Received: 13 Mar. 2026 | Revised: 14 Mar. 2026 | Accepted: 18 Mar. 2026 | Published online: 24 Apr. 2026
RESEARCH ARTICLE

Journal of Global Humanities and Social Sciences
2026,Vol. 7(2)95-103
DOI: 10.61360/BoniGHSS262019740203

Understanding China's Destination Image through TikTok Comments: Evidence from Sentiment and TF–IDF Analysis

Peng Gao[1] & Zong-Yi Zhu[2,*]

[1]*Global Management, Graduate School, Kook-Min University, Republic of Korea*

[2]*Graduate School of Business Administration, Kook-Min University, Republic of Korea*

Abstract: In recent years, China has implemented a series of policies to promote the development of inbound tourism, and "China travel" has become a widely discussed topic on social media platforms. Using Python-based text mining and sentiment analysis techniques, this study analyzes user comments under the "#chinatravel" topic on TikTok to examine the emotional tendencies of international users toward tourism in China. The results indicate that positive comments significantly outnumber negative ones, suggesting that international users generally hold favorable attitudes toward traveling in China. TF–IDF analysis further shows that positive sentiments emphasize aesthetic appreciation, emotional resonance, and overall positive impressions, whereas negative sentiments are relatively limited and mainly associated with critical or politicized expressions. Overall, the findings reveal a distinct emotional bias in TikTok comment sections. User-generated comments serve as an important source of informational and emotional cues for potential tourists and provide empirical evidence of how online discourse shapes international perceptions of Chinese tourism destinations.

Keywords: inbound tourism, TikTok, user-generated content, sentiment analysis, destination image, China travel

1. Introduction

In 2025, international tourist arrivals worldwide are projected to reach 1.52 billion, representing a 4% increase compared with 2024 and marking a new post-pandemic record. In addition, international tourism receipts are expected to total approximately USD 1.9 trillion, reflecting a year-on-year growth of 5% (UN Tourism, 2026). Amid the strong recovery of the global tourism market, the Chinese government has introduced a series of policies to promote inbound tourism, including the continuous expansion of visa-free entry, the gradual improvement of travel convenience, and the enhancement of tourism infrastructure. As a result, the label "China travel" has attracted widespread attention on online platforms (China Daily, 2025), and an increasing number of international tourists are choosing to visit China (China Daily, 2026). The comprehensive recovery of China's inbound tourism market has also contributed new momentum to the global tourism recovery.

As tourists increasingly rely on online information for travel planning, digital content plays an increasingly important role in shaping their perceptions, attitudes, and behavioral intentions toward tourist destinations (An et al., 2022; Wang & Yan, 2022). Among the various forms of online information, online reviews have attracted particular attention because they provide rich informational and emotional cues. These characteristics also influence the decision-making processes of potential tourists (Wu & Yang, 2023; Mou & Wang, 2025). Therefore, in the context of China's inbound tourism recovery, it is essential to understand how international tourists process tourism-related information embedded in online reviews. Such understanding is particularly

Corresponding Author: Zong-Yi Zhu
Graduate School of Business Administration, Kook-Min University, Republic of Korea

important for advancing theoretical development within the digital tourism information environment and for supporting the sustainable development of tourist destinations.

Early studies on online reviews primarily focused on traditional text-based review platforms, such as TripAdvisor, Booking.com, and online travel forums, where tourists express their evaluations and feelings through written comments (Orea-Giner et al., 2022; Rita et al., 2022). Using information-processing techniques, prior research has demonstrated that the emotional polarity and evaluative tone embedded in online reviews significantly influence destination image, travel intentions, and tourists' perceptual responses (Guo et al., 2022; Wu & Yang, 2023). These studies have established a solid methodological foundation for understanding tourists' perceptions through textual data. However, with the rapid evolution of the social media environment, tourism-related communication has gradually expanded from traditional review platforms to emerging short-video social media platforms (Liu et al., 2024; Serrano-Malebrán et al., 2025). In particular, TikTok has become one of the most popular short-video-sharing platforms worldwide, featuring billions of user-generated videos and comments that attract global audiences across diverse age groups (Corso et al., 2024). TikTok comments not only reflect users' immediate reactions and attitudes but also constitute an important data source for research on the sustainable development of tourist destinations (Wengel et al., 2022). Owing to its visually engaging, emotionally rich content and rapid dissemination, TikTok-based user interactions have increasingly been employed in academic research to explore tourists' perceptions, motivations, and behavioral intentions (Liu et al., 2024). Therefore, it is essential to understand users' viewpoints on tourism-related information within platforms such as TikTok, in order to further examine how short-video content and user interactions shape destination image and tourists' perceptual responses.

From a methodological perspective, previous studies on sentiment analysis of large-scale social media comments have typically relied on lightweight text mining techniques, as short texts often exhibit a fragmented nature (Yue et al., 2019; Cui et al., 2023). Among these techniques, the Term Frequency–Inverse Document Frequency (TF–IDF) algorithm has been widely used for keyword extraction and for identifying key emotional expressions in short, unstructured text data. Prior research has demonstrated that TF–IDF-based methods are effective for analyzing large-scale social media text, offering a practical and interpretable approach for emotion-related research (Durachman et al., 2024; Lestari & Hutagalung, 2025).

Building on these research trends, the study follows the existing progress in online review sentiment analysis and shifts the focus from traditional text-based review platforms to short-video-oriented social media platforms, such as TikTok. Specifically, the study analyzes TikTok comments under the topic tag #chinatravel, applying the TF–IDF method to extract frequent keywords and identify the emotional tendencies expressed in user comments. Using this approach, the study aims to reveal the dominant emotional patterns related to Chinese tourism on TikTok and to assess the applicability of lightweight text mining techniques for analyzing large-scale, short social media comments. The findings provide empirical evidence regarding the public's emotional expressions toward Chinese tourism on TikTok and enrich our understanding of how tourism destinations are emotionally constructed and disseminated in the contemporary social media environment.

2. Methodology

2.1. Data sources

To capture online public perceptions and attitudes toward "China travel," this study conducted a search on the TikTok platform on June 26, 2025, using the hashtag #chinatravel. The search was performed using TikTok's built-in search function without restrictions on geographical location or language, aiming to collect user-generated video content related to Chinese tourism worldwide. The

search returned approximately 208,900 short videos associated with the hashtag, covering a wide range of content creators and audiences from different countries and regions. To ensure the representativeness and research value of the data sample, this study selected the top-ranked #chinatravel short video from the recommended list as the primary object of analysis(Link: *https://www.tiktok.com/@beautifulworld.1901/video/7482372660882607382?is_from_webapp=1&sender_device=pc&web_id=7514865925394138625*). This video has a high number of views, likes, and comment interactions, which to some extent reflect users' attention and attitudinal tendencies toward the "chinatravel" theme on the TikTok platform.

The data acquisition process employed a Python-based script to crawl and extract all publicly visible comment data from the aforementioned videos. To ensure data quality, the raw data underwent cleaning procedures, including the removal of duplicate comments, advertising content, comments without substantive textual information (e.g., those containing only emojis), and noise data unrelated to the research theme. After data cleaning, a total of 32,943 valid comments were retained, forming the basic dataset for subsequent sentiment analysis and text mining. All data used in this study were obtained from publicly accessible online content and did not contain any identifiable personal information. The data collection process strictly adhered to TikTok's terms of use and relevant academic research ethics, and all data were used solely for academic research purposes.

2. 2. Analysis method

Date processed and analyzed the collected comments using the Python programming language. The entire analysis process was divided into three main stages: data reprocessing, sentiment analysis, and text mining.

During the data reprocessing stage, the original data were imported into a structured format using the *pandas* library, and duplicate records were removed based on the unique identifiers of the comments. Subsequently, all text was converted to lowercase to ensure case insensitivity. On this basis, regular expressions were applied to remove punctuation marks, numbers, hyperlinks, and non-alphabetic characters.

To enhance interpretability and visualization, keyword word clouds were generated using the *wordcloud* library, allowing intuitive observation of salient terms and their relative importance. Furthermore, a word co-occurrence network was constructed to examine the semantic relationships among high-frequency keywords. In the co-occurrence network, nodes represent keywords, and edges indicate their co-occurrence relationships within the same comments, thereby revealing the structural patterns and thematic associations embedded in the discourse. All visualizations were generated using the *matplotlib* library.

During the sentiment analysis stage, the VADER (Valence Aware Dictionary for Sentiment Reasoning) model from the nltk.sentiment module was used to calculate the sentiment polarity score for each comment. Based on the threshold values of the compound score, comments were classified into three categories: positive ($> +0.05$), negative (< -0.05), and neutral (between the two). VADER was selected because it is a lexicon- and rule-based sentiment analysis tool. Specifically, it assigns sentiment scores to individual words using a predefined sentiment lexicon and then computes an overall sentiment score for each sentence by considering sentence-level features, such as degree adverbs, negation terms, and punctuation (e.g., exclamation marks). In particular, VADER has been optimized for short, informal, and social media texts, making it highly suitable for the sentiment analysis of TikTok comments.

During the text mining stage, the Term Frequency–Inverse Document Frequency (TF–IDF) method implemented using the *scikit-learn* library was employed to identify high-weight keywords in the comment corpus. To capture sentiment-related lexical differences, TF–IDF scores were calculated separately for positive and negative comment subsets based on the sentiment classification results.

2. 3. TF－IDF algorithm

The Term Frequency–Inverse Document Frequency (TF–IDF) algorithm is a widely used statistical method for evaluating the importance of words in textual data. Its fundamental principle is that a word is considered to have strong discriminative ability if it appears frequently in a specific document while occurring infrequently in the overall corpus (Havrlant and Kreinovich, 2017). Owing to its simplicity, computational efficiency, and interpretability, TF–IDF has been extensively applied in social media text analysis and opinion mining.

Term Frequency (TF) refers to the frequency with which a word appears in a document or a set of comments. It reflects the local importance of a word within a given text and is calculated as follows:

$$TF_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (1)$$

In Equation (1), TF_{ij} represents the frequency with which the term t_i appears in document d_j. The numerator indicates the number of times the term t_i occurs in document d_j, while the denominator represents the total number of terms in document d_j.

Inverse Document Frequency (IDF) measures the global importance of a term across the document corpus. It is calculated by dividing the total number of documents by the number of documents containing the term and then taking the logarithm of the resulting ratio:

$$IDF_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (2)$$

In Equation (2), |D| represents the total number of documents, and $|\{j:t_i \in d_j\}|$ denotes the number of documents containing the term t_i. A higher IDF value indicates that the term appears in fewer documents and thus has stronger discriminative power.

The TF–IDF value of a term is obtained by multiplying its Term Frequency (TF) and Inverse Document Frequency (IDF):

TF–IDF = Term Frequency (TF) × Inverse Document Frequency (IDF).

A higher TF–IDF score indicates that a term is both frequent in a specific document subset and relatively rare in the overall corpus, making it particularly informative for identifying key concepts and evaluative expressions. In this study, TF–IDF was used to extract salient keywords from TikTok comments related to #chinatravel, providing an empirical basis for subsequent keyword analysis.

3. Results

3. 1. Word cloud analysis

To provide an intuitive overview of the lexical features of TikTok comments related to the #chinatravel topic, a word cloud was generated based on word frequency. As shown in Figure 1, several terms appear with particularly high frequencies, including “China,” “Beautiful,” “Country,” “Amazing,” and “Love.” These high-frequency words indicate that users’ discussions are largely centered on China as a destination.

3. 2. Word co-occurrence network analysis

To further explore the semantic structure of the comments, a word co-occurrence network analysis was conducted. As shown in Figure 2, “China” is closely connected with terms such as “strong,” “different,” and “interesting,” indicating a generally positive and distinctive national image. In addition, words such as “you,” “they,” and “America” occupy central positions in the network, reflecting frequent comparisons between China and other countries and highlighting themes related to national image, cultural differences, and international comparison.

Figure 1. Word Cloud Based on Word Frequency

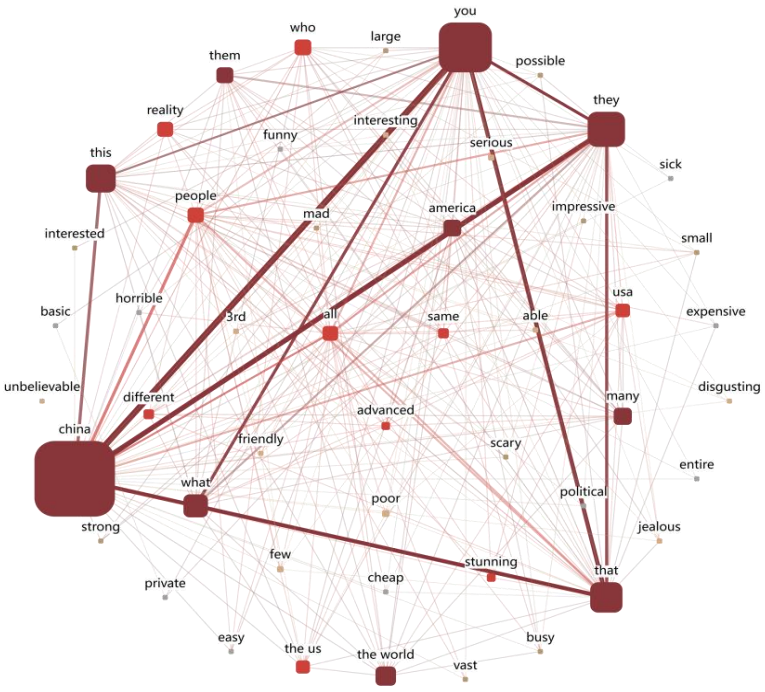

Figure 2. Word Co-occurrence Network

3. 3. Sentiment analysis

The results of the sentiment analysis indicate that the overall emotional tendency of TikTok comments related to the #chinatravel topic is predominantly neutral. Specifically, among the collected comments, 5,691 were classified as positive, accounting for 23.4% of the total, while 1,360 comments were identified as negative, representing only 5.6%. In contrast, a substantial majority of the comments—17,276 in total—were categorized as neutral, comprising 71% of the dataset (see Table 1).

Table 1. Sentiment Analysis Data Summary

Sentiment Category	Count	Percentage
Neutral	17,276	71.0%
Positive	5,691	23.4%
Negative	1,360	5.6%
Total	24,327	100.0%

3. 4. TF-IDF keyword analysis

Following the sentiment classification, a TF-IDF–based keyword analysis was conducted to further identify representative terms in the positive and negative comment subsets. By calculating TF-IDF weights for each term, this analysis highlights words that are not only frequently used within a specific sentiment category but also exhibit strong discriminative power across the overall comment corpus. Table 2 presents the top-ranked keywords extracted from positive and negative comments based on their TF-IDF values.

The results indicate that positive comments are predominantly characterized by affective and evaluative expressions such as "beautiful," "wow," "love," and "amazing," reflecting users' favorable emotional responses and admiration toward China-related travel content. In contrast, negative comments are mainly associated with terms such as "no," "bad," "propaganda," and "war," which suggest skepticism, political associations, or negative perceptions held by a small subset of users.

4. Discussion

This Study processed and analyzed comments collected under the "chinatravel" topic on TikTok using the Python to explore how international TikTok users expressed their opinions about China travel in the comment section. The findings indicate that TikTok comments related to China travel exhibit a clear emotional bias, suggesting that social media comment sections serve as an important space for the public to express their attitudes toward tourist destinations. This finding is consistent with previous tourism research, which indicates that user-generated comments provide valuable informational and emotional cues for potential tourists (Liu et al., 2024; Wijaya et al., 2025; Aboalganam et al., 2025).

Table 2. Top 10 TF-IDF Keywords in Positive and Negative Comments

Sentiment	Keyword	Word Frequency	TF-IDF Value	Rank
Positive	beautiful	1005	0.00879	1
	wow	694	0.00680	2
	love	476	0.00515	3
	like	470	0.00513	4
	amazing	405	0.00454	5
	want	277	0.00339	6
	good	249	0.00315	7
	please	242	0.00308	8
	nice	226	0.00289	9
	well	209	0.00274	10
Negative	no	643	0.00660	1
	bad	104	0.00156	2
	die	91	0.00143	3
	propaganda	82	0.00129	4
	war	49	0.00083	5
	hate	44	0.00076	6
	sin	44	0.00076	7
	fake	35	0.00062	8
	crazy	34	0.00061	9
	wrong	31	0.00057	10

A word cloud generated based on word frequency shows that “China”, “beautiful”, “country”, “amazing”, and “love”, indicating that discussions primarily revolve around national image and emotional evaluations. The co-occurrence network analysis revealed a closely knit semantic structure in which “China” is strongly associated with descriptive terms such as “strong”, “different”, and “interesting”. This pattern indicates that perceptions of China are multifaceted and often constructed through processes of comparison and contrast. The central positions of pronouns such as “you” and “they”, along with references to other countries such as “America”, suggest that commenters frequently situate China within an international comparative framework.

The sentiment analysis results indicated that neutral comments accounted for the majority, while the number of positive comments was significantly higher than that of negative comments. This distribution pattern suggests that positive attitudes toward China travel are more prevalent than negative ones. Further analysis of high-frequency words and TF–IDF features revealed the underlying content associated with these sentiments. Positive comments were primarily characterized by evaluative terms such as “beautiful”, “amazing”, “love”, and “good”, emphasizing aesthetic appreciation, emotional resonance, and favorable overall impressions. In contrast, negative comments were relatively limited in number and mainly consisted of critical or politically charged terms such as “bad”, “hate”, and “propaganda”. These results suggest that negative perceptions tend to concentrate on specific issues rather than being broadly distributed across multiple evaluative dimensions.

4.1. Theoretical implications

This study aligns with existing research trends in the sentiment analysis of online reviews by shifting the research focus from traditional text-based review platforms to TikTok, a social media platform dominated by short videos (Liu et al., 2024; Serrano-Malebrán et al., 2025). The findings support previous research indicating that user-generated content on social media platforms carries significant emotional value and informational characteristics (Wijaya et al., 2025; Aboalganam et al., 2025).

Secondly, by applying the TF-IDF method to identify key words and sentiment tendencies, this study demonstrates the applicability of lightweight text-mining techniques for analyzing large volumes of short and informal user comments (Durachman et al., 2024; Lestari & Hutagalung, 2025). This offers important methodological insights for future research to translate sentiment analysis results into experimental stimuli and to further examine, through experimental methods, how emotional expressions in social media information influence tourists’ perceptual responses.

4. 2. Managerial implications

The results of this study provide valuable insights for destination marketing organizations and tourism practitioners. Monitoring emotional tendencies in TikTok comments can help practitioners better understand public perceptions of destinations and identify emerging topics of interest or concern, thereby facilitating more effective communication with tourists.

Additionally, insights derived from keyword and co-occurrence analyses can guide short-video content creators in designing videos that better resonate with the audience’s emotions and expectations, ultimately enhancing destination attractiveness and stimulating travel intentions.

4. 3. Limitations and future research

This study has several limitations, which also provide directions for future research. First, the analysis focuses solely on TikTok comments related to the “chinatravel” topic, which may limit the generalizability of the findings to other social media platforms and tourism destinations. Future research could adopt a comparative approach by examining destination-related comments across multiple social media platforms.

Second, this study employs keyword extraction and sentiment analysis based on the TF–IDF approach. Although this lightweight text-mining technique is effective for analyzing large volumes of short text, it may not fully capture subtle emotional

expressions. Future studies may introduce more advanced natural language processing techniques, such as deep learning–based sentiment models (Singgalen, 2024), to better identify nuanced emotional patterns in unstructured text data and enhance analytical depth.

Finally, future research could build upon the high-frequency emotional words identified in this study to design emotional stimuli and adopt experimental methods to further examine the causal relationship between comment sentiment and tourists' perceptual responses.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

References

Aboalganam, K. M., AlFraihat, S. F., & Tarabieh, S. (2025). The impact of user-generated content on tourist visit intentions: The mediating role of destination imagery. *Administrative Sciences, 15*(4), 117. https://doi.org/10.3390/admsci15040117

An, S., Kim, W., Lee, B., & Suh, J. (2022). A study on the tourism-related information consumption process of tourists on social networking sites. *Sustainability, 14*(7), 3980. https://doi.org/10.3390/su14073980

China Daily. (2026, January 7). Polished inbound tourism sector sparkles brightly again. *China Daily*. https://www.chinadaily.com.cn/a/202601/07/WS695d9c6ea310d6866eb32574.html

China Daily. (2025, May 6). Exotic charm: East meets Middle East. *China Daily*. https://www.chinadaily.com.cn/a/202505/06/WS6819fc49a310a04af22bdbd6.html

Corso, F., Pierri, F., & De Francisci Morales, G. (2024, May). What we can learn from TikTok through its Research API. *In Companion Publication of the 16th ACM Web Science Conference* (pp. 110-114). https://doi.org/10.1145/3630744.3663611

Cui, J., Wang, Z., Ho, S. B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. Artificial Intelligence Review, 56(8), 8469-8510. https://doi.org/10.1007/s10462-022-10386-z

Durachman, Y., Putra, S. J., Nanang, H., & Sukmana, H. T. (2024, August). Analysis sentiment of public opinion on social media using naïve Bayes and TF-IDF algorithms. *In 2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCIT62134.2024.10701191

Guo, X., Pesonen, J., & Komppula, R. (2022). Analysing online travel reviews to identify temporal changes of a destination image. *European Journal of Tourism Research, 32*, 3209-3209. https://doi.org/10.54055/ejtr.v32i.2447

Havrlant, L., & Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems, 46*(1), 27-36. https://doi.org/10.1080/03081079.2017.1291635

Lestari, V. B., & Hutagalung, C. A. (2025). Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. J-KOMA: *Jurnal Ilmu Komputer dan Aplikasi, 8*(1), 36-44. https://doi.org/10.21009/j-koma.v8i1.05

Liu, C., Jiang, M., & Muhammad, Z. A. (2024). The impact of TikTok short video factors on tourists' behavioral intention among Generation Z and Millennials: The role of flow experience. *Plos one, 19*(12), e0315140. https://doi.org/10.1371/journal.pone.0315140

Mou, T., & Wang, H. (2025). Online comments of tourist attractions combining artificial intelligence text mining model and attention

mechanism. *Scientific Reports, 15*(1), 1121. https://doi.org/10.1038/s41598-025-85139-3

Orea-Giner, A., Fuentes-Moraleda, L., Villacé-Molinero, T., Muñoz-Mazón, A., & Calero-Sanz, J. (2022). Does the implementation of robots in hotels influence the overall TripAdvisor rating? A text mining analysis from the industry 5.0 approach. *Tourism Management, 93*, 104586. https://doi.org/10.1016/j.tourman.2022.104586

Rita, P., Ramos, R., Borges-Tiago, M. T., & Rodrigues, D. (2022). Impact of the rating system on sentiment and tone of voice: A Booking. com and TripAdvisor comparison study. *International Journal of Hospitality Management, 104*, 103245. https://doi.org/10.1016/j.ijhm.2022.103245

Serrano-Malebrán, J., Campos-Núñez, F., Vidal-Silva, C., Veneros-Alquinta, D., Zuñiga-Contreras, F., & Ahumada-Gutierrez, D. (2025). From search engines to social influence: a stimulus–organism–response model of travel influencers on TikTok. *Frontiers in Communication, 10*, 1649647. https://doi.org/10.3389/fcomm.2025.1649647

Singgalen, Y. A. (2024). Sentiment Analysis and Trend Mapping of Hotel Reviews Using LSTM and GRU. *Journal of Information Systems and Informatics, 6*, 2814-2836. https://doi.org/10.51519/journalisi.v6i4.926

UN Tourism. (2026, January). *World tourism barometer* (Vol. 24, No. 1, excerpt). https://pre-webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2026-01/World_Tourism%20Barometer_Jan26_excerpt_v2.pdf?VersionId=u75u9KWPa6Dzc2CUHld7AvQ49FYrDTQC

Wang, H., & Yan, J. (2022). Effects of social media tourism information quality on destination travel intention: Mediation effect of self-congruity and trust. *Frontiers in Psychology, 13*, 1049149. https://doi.org/10.3389/fpsyg.2022.1049149

Wengel, Y., Ma, L., Ma, Y., Apollo, M., Maciuk, K., & Ashton, A. S. (2022). The TikTok effect on destination development: Famous overnight, now what?. *Journal of Outdoor Recreation and Tourism, 37*, 100458. https://doi.org/10.1016/j.jort.2021.100458

Wijaya, C. O., Wijaya, S., & Jaolis, F. (2025). The influence of social media content on attitude, destination image and intention of female Muslim travelers to visit halal destinations: comparison between UGC and FGC. *Journal of Islamic Marketing, 16*(2), 402-427. https://doi.org/10.1108/JIMA-08-2023-0235

Wu, J., & Yang, T. (2023). Service attributes for sustainable rural tourism from online comments: Tourist satisfaction perspective. *Journal of Destination Marketing & Management, 30*, 100822. https://doi.org/10.1016/j.jdmm.2023.100822

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and information systems, 60*(2), 617-663. https://doi.org/10.1007/s10115-018-1236-4

How to Cite: Gao, P., & Zhu, Z. Y. (2026). Understanding China's Destination Image through TikTok Comments: Evidence from Sentiment and TF–IDF Analysis. *Journal of Global Humanities and Social Sciences*, *7*(2), 95-103
https://doi.org/ 10.61360/BoniGHSS262019740203