

Article

A Multi-Model Text Mining Approach to Tourism Image Analysis of the Historic Centre of Macao Based on User-Generated Content

Xiao Xu ^{1,2}  and Qiaoyun Zhang ^{1,*} 

¹ Faculty of Humanities and Social Sciences, Macao Polytechnic University, Macao 999078, China; p2413438@mpu.edu.mo

² Faculty of Art and Media, Xianda College of Economics & Humanities Shanghai International Studies University, Shanghai 202162, China

* Correspondence: qiaoyunzhang@mpu.edu.mo

Abstract

User-generated content (UGC) offers large-scale, naturalistic data for examining tourism destination image. Focusing on the Historic Centre of Macao (HCM), a World Cultural Heritage site, this study collected 3781 tourist reviews from Rednote, Ctrip, Dianping, and TripAdvisor and developed a multi-model text-mining framework integrating TF-IDF, BERTopic, and RoBERTa. The results show that HCM's online tourism image comprises four dimensions: perceptions of history, culture, and heritage value; perceptions of spatial landmarks and urban landscapes; modes of travel behavior and embodied experience; and emotional evaluation and tourism experience quality. The TF-IDF results indicate that terms such as architecture, history, Portuguese, church, Ruins of St. Paul's, and Senado Square constitute the core elements of tourists' cognitive image. BERTopic further identified 18 valid topics and revealed three interrelated semantic clusters: heritage-space cognition, landmark and district experiences, and integrated tourism experiences. The RoBERTa-based sentiment analysis shows that tourists' overall evaluations are dominated by positive emotions, while crowding, high visitor density, and gaps between expectations and actual experiences remain important sources of negative evaluations. This study demonstrates the applicability of the proposed framework in the Historic Centre of Macao and provides a methodological reference for tourism image research in other cultural heritage destinations.

Keywords: World Cultural Heritage tourism; Historic Centre of Macao; user-generated content; TF-IDF; BERTopic; RoBERTa



Academic Editor: Nikos A. Salingeros

Received: 24 June 2026

Revised: 18 July 2026

Accepted: 20 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The formation of tourism destination image (TDI) is a complex cognitive and affective process involving the interaction among tourists' perceptions of destination attributes, emotional responses, and behavioral intentions [1]. With the rapid development of internet platforms and mobile social media, tourists increasingly express their travel experiences through online reviews, travel notes, photo-text travel narratives, and social media posts. User-generated content (UGC) has become an important information source influencing the formation and dissemination of destination image [2]. Relevant reports indicate that social media and online travel platforms have become major channels through which tourists

obtain destination information [3]. Online reviews not only record individual travel experiences but also contribute to the construction of a collective cognitive image of destinations [4]. Consequently, UGC-based text mining methods provide a new data foundation and technical pathway for TDI research [5].

World Heritage sites are not only cultural preservation areas but also significant tourist destinations. The Historic Centre of Macao (HCM) was inscribed on the UNESCO World Heritage List in 2005. According to UNESCO, the property, with its historic streets, religious buildings, public buildings, and residential architecture, bears witness to one of the “earliest and longest-lasting encounters between China and the West” [6]. Twenty-two historic buildings of diverse architectural styles and eight distinctive squares are organically connected through a traditional street network, together constituting Macao’s unique historic urban landscape. Catholic churches stand adjacent to Taoist temples, Portuguese-style cobblestone pavements echo Lingnan arcade architecture, and European building façades coexist with Chinese courtyard layouts. These spatial juxtapositions make HCM an important heritage space for understanding the interaction between East Asia and the wider world, cultural exchange between China and the West, and the coexistence of multiple cultures [7]. Therefore, HCM is not only a core attraction of urban tourism in Macao, but also a representative case for examining the construction of tourism image of world heritage sites.

However, existing studies on the tourism image of HCM still have certain limitations. Ge et al. identified the static cultural genes of HCM from the perspectives of tourists, residents, and experts, but did not further examine their respective priorities and emotional responses [8]. Meng et al.’s visual analysis of HCM as a World Heritage tourism destination relied solely on online reviews from Chinese online travel platforms [9]. Chen et al. compared differences in the image of HCM projected through Chinese- and English-language comments on Weibo and Instagram, as well as the extent to which these images aligned with official narratives [10]. Wong and Qi employed text-mining techniques to examine the evolution of Macao’s destination image based on TripAdvisor reviews. However, their study focused on Macao as an overall tourism destination and relied on data from a single platform [11]. Collectively, these studies have not systematically captured the tourism image of HCM as represented in cross-platform and cross-cultural UGC. To address these limitations, this study takes HCM as its research object and draws on online review data from four platforms: Rednote, Ctrip, Dianping and TripAdvisor. A multi-model text mining framework integrating TF-IDF, BERTopic, and RoBERTa is developed to systematically analyze user-generated content. This analytical framework corresponds closely to the cognitive, affective, and overall dimensions of destination image theory [12], enabling a comprehensive examination of the online tourism image of HCM from three perspectives: cognitive attributes, semantic themes, and emotional evaluations. Compared with tourism image analyses relying on a single model, the integration of these three analytical levels makes it possible to identify not only what tourists notice, but also how these perceptions are semantically organized and emotionally evaluated, thereby providing more actionable evidence for heritage communication, visitor flow management, and experience optimization. Accordingly, this study seeks to address the following research questions:

RQ1: What are the key cognitive elements of tourists’ perceived image of HCM?

RQ2: What latent semantic topics can be identified from multi-platform UGC using BERTopic?

RQ3: What sentiment patterns and experience-related issues are revealed by RoBERTa-based sentiment analysis?

2. Literature Review

This section reviews the theoretical and methodological foundations of this study. It first discusses the relationship between tourism image and user-generated content, then reviews text mining approaches in tourism research, and finally explains the relevance of sentiment analysis for understanding heritage tourism experiences.

2.1. Tourism Destination Image and Cultural Heritage Tourism

Tourism destination image (TDI) is an important concept in tourism management, destination marketing, and tourist behavior research. Hunt [13] was one of the first scholars to introduce the concept of image into tourism research, arguing that potential tourists' perceptions and impressions of a destination influence their travel decisions. Crompton [14] defined destination image as the sum of beliefs, ideas, and impressions that an individual holds about a particular destination. Based on a systematic review of definitions related to TDI, Lai and Li [15] noted that destination image is generally understood as the mental impression or overall perception formed by tourists toward a destination, although its conceptual boundaries remain somewhat ambiguous. In terms of destination image measurement, Echtner and Ritchie [16] proposed that a complete destination image should include both attribute-based images and holistic impressions, while also covering multiple dimensions such as functional, psychological, common, and unique characteristics. Bramwell and Rawding [17] further distinguished between projected image and perceived image: the former refers to the image communicated by destination marketing organizations through official media, whereas the latter refers to the subjective impression formed by tourists through information exposure and actual experience. With the development of social media and online tourism platforms, tourists' perceived images are increasingly constructed, communicated, and reproduced through online reviews, travel narratives, and social media content. Regarding image structure, the cognitive-affective-overall image model proposed by Baloglu and McCleary [12] has had a significant influence on subsequent research. According to this model, cognitive image refers to tourists' beliefs, knowledge, and evaluations of destination attributes and characteristics, including natural landscapes, cultural resources, infrastructure, and service quality. Affective image refers to tourists' emotional responses toward a destination, such as pleasure, affection, excitement, or disappointment. Together, these two components influence tourists' overall evaluations and their behavioral intentions. Beerli and Martín [18] further pointed out that destination image is shaped by multiple factors, including tourism resources, infrastructure, leisure and entertainment, culture and history, social environment, and tourists' personal characteristics.

The tourism image of cultural heritage sites differs from that of general tourism destinations. Richards [19] pointed out that cultural tourism is centered on learning, discovering, experiencing, and consuming the tangible and intangible cultural resources of a destination, involving multiple elements such as architecture, heritage resources, culinary traditions, art, lifestyles, and value systems. For cultural heritage sites, tourist perception involves not only landscape aesthetics and service experience, but also understanding of historical narratives, cultural authenticity, and place meanings, through which tourists identify heritage value [20,21]. Existing studies indicate that the tourism image of cultural heritage sites usually contains both cognitive and affective dimensions. On the one hand, tourists perceive heritage value through historical knowledge, architectural landscapes, and cultural symbols; on the other hand, they form affective evaluations through spatial atmosphere, aesthetic experience, and cultural resonance [22,23].

As a World Heritage site characterized by the convergence of Chinese and Portuguese cultures, the Historic Centre of Macao (HCM) provides a representative case for research

on the tourism image of cultural heritage sites. Existing studies on Macao have addressed issues such as tourist experience, visual landscape, and the tensions between heritage conservation and tourism development [24–29]. However, existing research still lacks a systematic analysis of user-generated content (UGC) across multiple platforms. In particular, few studies have comprehensively identified the online tourism image of HCM across the three dimensions of cognitive elements, semantic topics, and sentiment tendencies.

2.2. User-Generated Content and the Digital Construction of Tourism Image

UGC refers to content independently created and publicly posted by internet users on digital platforms, including online reviews, social media posts, travel narratives, travel guides, images, and short videos [30]. In tourism research, UGC has become an important data source for examining tourist perceptions, destination image, and tourism experiences. Compared with traditional questionnaire surveys, UGC is characterized by its large volume, naturalistic nature, real-time availability, and rich contextual information, thereby revealing tourists' spontaneous expressions and experiential narratives in authentic travel settings [31].

The influence of UGC on destination image can be understood from two perspectives. First, UGC serves as an important source of travel information and shapes potential tourists' expectations, perceptions, and travel decisions regarding destinations. Narangajavana et al. [32] classified social media information sources into three categories: strong-tie sources, weak-tie sources, and tourism-specific sources, and argued that UGC from different sources may exert differentiated effects on tourists' destination expectations and satisfaction. Second, UGC functions as an important medium through which tourists express their travel experiences, evaluate services, and contribute to the dissemination of destination images after their visits. Existing studies have shown that UGC can indirectly influence tourists' revisit intention and word-of-mouth communication through perceived destination image and tourist satisfaction [33]. From the perspective of cognitive–affective information processing, UGC shapes tourists' imagination of destinations through both cognitive and affective processing pathways, thereby further influencing their intention to visit [34]. With the ongoing transformation of the digital media environment, UGC provides researchers with a valuable textual corpus for examining the digital construction of tourism destination image.

2.3. Applications of TF-IDF, BERTopic Model, and RoBERTa-Based Sentiment Analysis in Tourism Image Research

Tourists continuously generate reviews, travel narratives, and social media content across various platforms, producing large-scale tourism-related textual data. Alaei, Becken, and Stantic [35] noted that technological advances have transformed the production and consumption of tourism information. As massive volumes of online text have become difficult to process manually, text mining and sentiment analysis have emerged as important tools in tourism big-data research.

Term Frequency–Inverse Document Frequency (TF-IDF) is a widely used term-weighting method in information retrieval and text mining. Its underlying logic is to consider both the frequency of a term within a specific corpus and its distinctiveness across the corpus as a whole [36]. In tourism image research, TF-IDF is commonly used to extract high-weight keywords from tourist reviews, thereby identifying the core cognitive attributes of a destination, such as landscapes, culture, cuisine, transportation, services, and activities [37]. Compared with simple word-frequency statistics, TF-IDF reduces the influence of common high-frequency terms and is therefore better suited to identifying keywords with strong thematic distinctiveness.

In terms of topic identification, Latent Dirichlet Allocation (LDA) has long been a mainstream model in tourism text analysis. LDA assumes that documents consist of multiple latent topics, each represented by a specific probability distribution of words, thereby enabling the automatic extraction of topic structures from large-scale textual data [38]. Existing studies have applied LDA to the analysis of tourist attraction reviews, tourist experiences, and destination image, achieving promising results [39,40]. However, as LDA is based on the bag-of-words assumption, it ignores word order and contextual semantic relationships, which creates limitations when dealing with short texts, colloquial reviews, and heritage tourism texts containing complex cultural meanings. BERTopic is a semantic topic modeling method that has gained considerable attention in recent years. It employs pre-trained language models such as BERT to generate text embeddings, applies UMAP for dimensionality reduction and HDBSCAN for clustering to identify semantically similar text groups, and finally uses c-TF-IDF to extract topic keywords [41,42]. Compared with LDA, BERTopic can better preserve contextual semantic information and improve the semantic coherence and interpretability of topics [43]. Therefore, it is particularly suitable for analyzing the latent experiential structures and semantic topics embedded in tourist reviews of cultural heritage sites. Chen et al. [44] employed a combined LDA and BERT-BiLSTM approach to analyze Chinese- and English-language tourist reviews of Xiamen and found that cultural differences significantly influence tourists' perceived image of destinations. This finding suggests that deep learning-based text models can offer new methodological possibilities for cross-cultural destination image research. For HCM, the combined use of TF-IDF and BERTopic enables the identification of both the explicit cognitive elements and latent semantic topics embedded in tourist reviews, thereby providing a more comprehensive analytical framework for investigating the tourism image of cultural heritage sites.

2.4. RoBERTa-Based Sentiment Analysis and Evaluation of Heritage Tourism Experience

Sentiment analysis is an important direction in tourism text mining and is primarily used to identify emotional tendencies in tourist reviews. Chinese sentiment analysis tools represented by SnowNLP have been widely applied in tourism review studies, but their results are often affected by the limitations of fixed dictionaries and the adaptability of training corpora. In recent years, sentiment analysis methods based on deep learning and Transformer architectures have attracted increasing attention in tourism research. RoBERTa improves upon BERT by optimizing its pre-training strategy, enabling it to better capture contextual dependencies, semantic nuances, and long-range textual information. Studies comparing the sentiment classification performance of traditional machine learning, deep learning, and Transformer-based models using large-scale TripAdvisor reviews have shown that RoBERTa performs particularly well when handling imbalanced datasets [45]. In tourism destination image research, topic modeling can identify the specific content that tourists focus on, while sentiment analysis can further determine their evaluative tendencies toward such content. The review by Bi et al. [46] further confirms the value of text analytics in tourism and hospitality research. In particular, the integration of topic modeling and sentiment analysis can reveal the relationships between tourists' sentiment patterns and the attributes of tourism products and services.

For cultural heritage sites, tourists' sentiment patterns not only reflect levels of satisfaction but also reveal the tensions among heritage conservation, tourism development, and experience quality. Antonio et al. combined text mining and sentiment analysis to examine 8638 online reviews of Coimbra and Salamanca, demonstrating the value of UGC analytics for identifying visitor perceptions and informing heritage-destination management [47]. Existing studies have shown that the uniqueness, authenticity, and cultural atmosphere of heritage sites are often important sources of positive sentiment, whereas com-

mercialization, crowding, overtourism, and the loss of authenticity may generate negative sentiment [22,23]. Therefore, RoBERTa-based sentiment analysis can not only quantify tourists' overall sentiment toward HCM but also help identify experience-related issues embedded in tourist reviews.

Recent AI-assisted studies on architectural perception also provide a useful methodological reference for interpreting affective evaluations of historic built environments. Salingaros used generative AI to assess building façade geometry through cognitive-emotional descriptors such as beauty, calmness, coherence, comfort, visual pleasure, and well-being, and found that traditional and geometrically coherent window compositions were consistently associated with more positive emotional responses [48]. This suggests that aesthetic and affective qualities of architecture can be examined through computational methods rather than relying solely on subjective description. Accordingly, the present study uses RoBERTa-based sentiment analysis not only to measure tourists' overall emotional tendencies toward HCM, but also to examine how positive and negative expressions reflect visitors' affective responses to historic architecture, traditional streetscapes, and the cultural atmosphere of the heritage site.

3. Materials and Methods

This section introduces the research materials and methodological procedures used in this study. It first explains the data sources and collection process for user-generated content (UGC) related to the Historic Centre of Macao (HCM), followed by the procedures for data cleaning and text preprocessing. It then presents the multi-model text mining framework, including term frequency–inverse document frequency (TF-IDF), BERTopic, and RoBERTa-based sentiment analysis, which was used to identify cognitive elements, semantic topics, and emotional evaluations in tourists' online reviews.

3.1. Research Object

The Historic Centre of Macao is located in the central to southwestern part of the Macao Peninsula, Macao Special Administrative Region, China. It is an urban heritage area composed of multiple historic buildings, religious spaces, public squares, and a traditional street network. The heritage site comprises 22 historic buildings and eight squares, which are organically connected through a continuous system of streets and lanes, forming a complex cultural heritage site characterized by the convergence of Chinese and Portuguese cultures, a historic urban landscape, and everyday living spaces. In 2005, HCM was officially inscribed on the World Heritage List, becoming one of China's important World Cultural Heritage Sites.

3.2. Research Design

All text processing and model analyses were conducted in the Python 3.14.3 environment. Prior to analysis, a dataset suitable for text mining was prepared. During the text analysis stage, the TF-IDF method was first employed to extract high-weight keywords from tourist reviews, with the aim of identifying the key cognitive elements of tourists' perceived image of HCM. Subsequently, the BERTopic model was applied to perform topic modeling on the review texts in order to identify latent semantic topics and their underlying structures. Finally, sentiment analysis based on a Chinese RoBERTa pre-trained language model was applied to identify the sentiment tendencies of the review texts, thereby determining tourists' evaluative attitudes toward HCM (Figure 1).

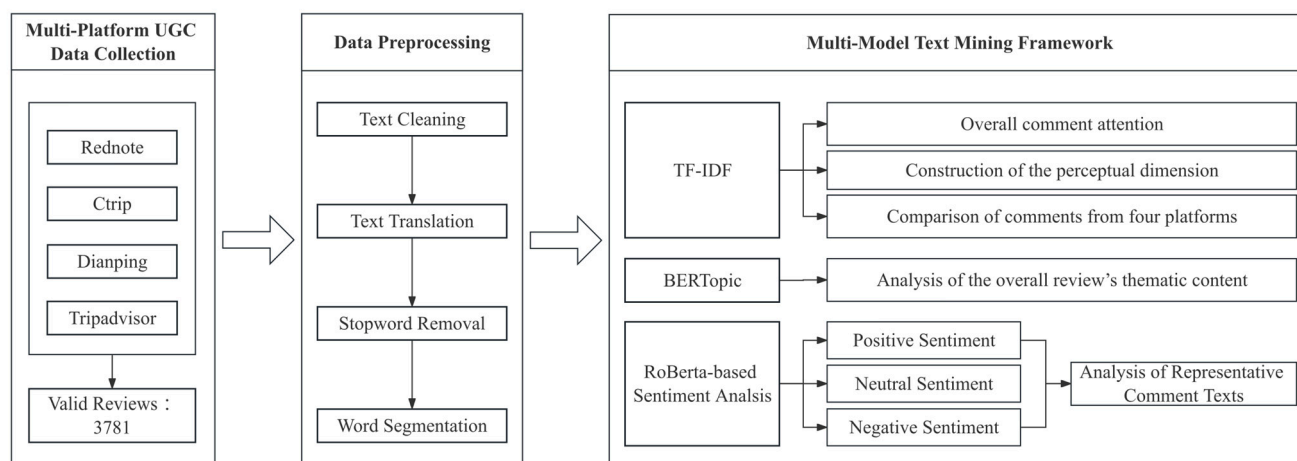


Figure 1. Research Design Flowchart. Created by authors.

3.3. Data Sources

This study selected four platforms, Rednote, Ctrip, Dianping, and TripAdvisor, as sources of UGC data. Rednote is a social media platform characterized by lifestyle sharing and the dissemination of image-text content. Its users tend to post travel guides, photography experiences, route recommendations, and check-in (打卡) content, which can reflect tourists' visualized and socialized perceptions of HCM in the context of social media. Ctrip is a major online travel service platform in mainland China, where reviews are usually related to scenic-site visits, overall satisfaction, and tourism service experiences, thus presenting domestic tourists' comprehensive evaluations of HCM. Dianping is mainly characterized by local life-service reviews, with content often involving dining, shopping, transportation, consumption, and everyday local experiences. It is therefore useful for identifying how HCM functions as both a tourism space and an everyday consumption space. TripAdvisor, by contrast, has a strong international tourism platform orientation. Its users come from more diverse backgrounds, and its reviews can supplement understandings and evaluations of HCM from tourists with different cultural backgrounds.

Because the four platforms differ in user composition, content-production mechanisms, review conventions, and language environments, the combined corpus was not treated as a completely homogeneous dataset. The combined corpus was used to identify the general online tourism image of HCM, while platform-stratified TF-IDF analysis was conducted to examine cross-platform consistency and platform-specific emphases. This design makes it possible to distinguish relatively stable components of HCM's tourism image from patterns that may be influenced by particular platform environments. Previous research has similarly shown that tourists' experiences of UGC platforms can influence the cognitive and affective formation of destination image, highlighting the importance of distinguishing between platform contexts [49].

3.4. Data Collection

This study used “澳门历史城区” (the Chinese name for “Historic Centre of Macao”) and “Historic Centre of Macao” as keywords, and the data collection focused on publicly available online reviews, travel notes, and tourism evaluation texts posted by tourists. The collected fields mainly included platform source, posting time, user ID, review title, and review body. User IDs were used only for data deduplication and sample identification. The data collection period ranged from 21 November 2011 to 21 May 2025. After preliminary screening, deduplication, and the removal of invalid texts, a total of 3781 valid review texts were obtained. Table 1 shows differences in the number of reviews across the four platforms, with TripAdvisor accounting for a relatively high proportion of

the total reviews. Therefore, in addition to the overall corpus analysis, this study further conducted a platform-stratified comparison to avoid the excessive influence of data from a single platform on the overall interpretation.

Table 1. Basic Information of the Dataset. Created by authors.

Platform	Platform Characteristics	Main Language(s)	Number of Reviews
Rednote	Travel guides, check-ins, photos	Chinese	489
Ctrip	Comprehensive reviews, tourism services	Chinese	190
Dianping	Dining, consumption, local experience	Chinese	341
TripAdvisor	International tourists, multilingual reviews	Chinese/English/ Portuguese/Japanese	2761
Total			3781

3.5. Data Preprocessing

To ensure the accuracy of subsequent text mining and sentiment analysis, this study conducted standardized preprocessing of the raw review data. The preprocessing procedure mainly consisted of four steps: text cleaning, language conversion, stop-word processing, and Chinese word segmentation.

First, in the text cleaning stage, content irrelevant to the research topic was removed, including advertisements, platform-generated default reviews, duplicate texts, short sentences without substantive semantic meaning, URL links, HTML tags, garbled characters, irrelevant symbols, and some emojis.

Second, English, Portuguese, and Japanese reviews were translated into simplified Chinese texts to facilitate subsequent processing using Chinese word segmentation, Chinese semantic topic modeling, and Chinese sentiment analysis models. During the translation process, particular attention was paid to verifying Macao place names, heritage building names, and common tourism expressions, so as to avoid inconsistencies in the translation of proper nouns that might affect keyword identification and topic clustering results.

Third, in the stop-word processing stage, this study used a general Chinese stop-word list as the basis and further expanded it according to the corpus characteristics of tourist reviews of HCM. High-frequency words with limited analytical significance, such as “this”, “that”, and “place”, were removed. At the same time, words with clear heritage and spatial references, such as “history”, “architecture”, “Portuguese”, “church”, “Ruins of St. Paul’s”, and “Senado Square”, were retained.

Fourth, in the Chinese word segmentation stage, this study used the Jieba segmentation tool to segment the review texts and constructed a customized dictionary related to HCM to improve the accuracy of proper noun recognition. The customized dictionary included place names and attraction names such as “Lou Kau Mansion” (卢家大屋), “A-Ma Temple” (妈祖阁/妈阁庙), and “Dom Pedro V Theatre” (岗顶剧院). It also incorporated a broader range of colloquial names for these sites. For example, the Ruins of St. Paul’s are commonly referred to as “大三巴牌坊”, “大三巴”, or “圣保禄教堂”, and such informal expressions are more widely used among tourists and the general public. After the above procedures, a standardized textual dataset suitable for TF-IDF keyword extraction, BERTopic topic modeling, and RoBERTa-based sentiment analysis was established.

4. Results

This section presents the main empirical findings of the study. The analysis first identifies the core cognitive elements of the online tourism image through TF-IDF, then examines the semantic structure of tourist perceptions using BERTopic, and finally evaluates tourists’

emotional attitudes toward the Historic Centre of Macao (HCM) through RoBERTa-based sentiment analysis.

4.1. Extraction of Tourists' Core Cognitive Elements Based on TF-IDF

4.1.1. Top TF-IDF Keyword Analysis of Tourist User-Generated Content (UGC)

To identify tourists' core cognitive elements of HCM, this study first applied TF-IDF to extract high-weight keywords from 3781 UGC review texts. Table 2 presents the top 50 keywords, together with their word frequencies and total TF-IDF weights. Overall, the highest-ranking keywords include core terms with high weights, such as "Macao", "architecture", "history", and "Portuguese"; specific landmark terms, such as "Ruins of St. Paul's" and "Senado Square"; activity-related terms, such as "walking", "walk", "visit", "strolling", and "photography"; and evaluative terms, such as "worthwhile", "like", "good", "excellent", "beautiful", and "pretty". This indicates that tourists' online expressions about HCM are not concentrated on a single object, but simultaneously involve multiple aspects, including city recognition, historic architecture, cultural characteristics, landmark spaces, modes of visitation, and experiential evaluation. It provides a basis for further identifying and summarizing the dimensions of tourist perception in the subsequent analysis.

Table 2. Top 50 Keywords Ranked by Cumulative TF-IDF Score. Created by authors.

Rank	Word	Frequency	Cumulative TF-IDF Score	Rank	Word	Frequency	Cumulative TF-IDF Score
1	Macao	5377	203.881	26	Senado Square	492	46.488
2	Architecture	1981	125.677	27	Visit	364	43.471
3	History	1782	120.593	28	Stroll	278	43.237
4	Portugal	1129	102.871	29	Nice	252	42.044
5	Streets	754	72.273	30	Hong Kong	400	41.171
6	China	785	71.454	31	Hotels	523	40.401
7	Centre	635	67.942	32	A lot of	328	39.550
8	Shops	500	64.872	33	Small	395	39.154
9	City	791	62.298	34	Squares	416	38.528
10	Worth (visiting)	472	60.239	35	Historic	344	38.075
11	Churches	773	60.006	36	Taking photos	458	37.836
12	Tourists	609	59.938	37	Wonderful	210	37.729
13	Like	466	58.946	38	Beautiful	241	37.691
14	Walking	708	57.475	39	Crowded	244	37.458
15	Interesting	382	55.412	40	Ruins of St. Paul's (大三巴牌坊)	493	37.251
16	Historic Centre of Macao	639	54.834	41	World Heritage	379	36.964
17	Culture	634	53.871	42	Tour	295	36.875
18	Scenic spots	613	52.717	43	Historic Area	329	36.724
19	Area	443	52.539	44	Ruins	328	36.453
20	See	430	48.135	45	All	285	36.363
21	Europe	353	47.761	46	Shopping	264	35.681
22	Time	556	47.551	47	Ruins of St. Paul's (大三巴)	545	35.229
23	Feel	368	47.125	48	Style	334	34.806
24	Casinos	427	46.713	49	Asia	206	34.148
25	Walk	519	46.577	50	Pretty	215	34.038

Note: The score represents the cumulative TF-IDF value of each keyword, calculated by summing its L2-normalized document-level TF-IDF values across all reviews. No additional min-max normalization was applied.

4.1.2. Analysis of Online Tourism Image Dimensions Based on TF-IDF Keywords

Based on the top 50 keywords ranked by cumulative TF-IDF score and their specific contexts in the original tourist reviews, two authors independently conducted a manual classification according to clearly defined criteria. Specifically, keywords referring to historical narratives, cultural attributes, World Heritage status, cross-cultural characteristics, and heritage significance were classified under perceptions of historical culture and heritage value. Keywords referring to heritage buildings, landmarks, streets, squares,

historic districts, and the physical or visual characteristics of space were classified under perceptions of spatial landmarks and urban landscapes. Keywords describing walking, visiting, touring, viewing, photographing, shopping, and other forms of embodied tourism activity were classified under modes of tourist behavior and embodied experience. Keywords expressing aesthetic judgments, emotional attitudes, satisfaction, perceptions of crowding or discomfort, and overall experience quality were classified under emotional evaluation and tourism experience quality. For keywords with multiple meanings or the potential to fall into more than one category, the two authors jointly re-examined their contexts in the original UGC and determined their final classification through discussion based on their predominant semantic usage. The resulting four dimensions of the online tourism image are shown in Table 3.

Table 3. Classification of TF-IDF Keywords into Four Dimensions of Online Tourism Image. Created by authors.

Four Dimensions	High-Frequency Words
Perceptions of Historical Culture and Heritage Value	Macao, Historic Centre of Macao, History, Culture, World Heritage, Ruins, Historic, Portugal, China, Europe, Asia, Hong Kong
Perceptions of Spatial Landmarks and Urban Landscapes	Architecture, Streets, City, Centre, Area, Churches, Squares, Scenic spots, Senado Square, Historic Area, Ruins of St. Paul's (大三巴牌坊), St. Paul's Ruins (大三巴), Style
Modes of Tourist Behavior and Embodied Experience	Tourists, Walking, Walk, Stroll, Visit, Tour, Taking photos, See, Time, Shopping
Emotional Evaluation and Tourism Experience Quality	Worth (visiting), Like, Interesting, Nice, Feel, Beautiful, Pretty, Wonderful, Crowded, Too many

In the dimension of perceptions of historical culture and heritage value, terms such as “history”, “culture”, “world heritage”, “ruins”, and “historic” have relatively high word frequencies and total TF-IDF weights, indicating that tourists’ basic cognition of HCM is grounded in its historical continuity, cultural accumulation, and World Heritage value. “Portugal” shows particularly prominent word frequency and total TF-IDF weight, indicating that Portuguese historical remains and related architectural styles are highly visible in tourists’ perceptions. This also gives the tourism image of HCM a clear characteristic of Chinese–Western cultural convergence.

In the dimension of perceptions of spatial landmarks and urban landscapes, the frequent occurrence of terms such as “architecture”, “streets”, “churches”, “squares”, “Senado Square”, “Ruins of St. Paul’s”, “historic area”, and “style” shows that tourists’ spatial perceptions are mainly based on the iconic buildings, public squares, street-and-lane textures, and overall urban landscape within the Historic Centre. As highly recognizable spatial nodes, the Ruins of St. Paul’s and Senado Square constitute important landmark symbols through which tourists perceive HCM.

In the dimension of tourist behavior and embodied experience, terms such as “walking”, “walk”, “stroll”, “visit”, “tour”, “taking photos”, and “see” reflect the pronounced slow-mobility characteristics of tourism experiences in HCM. Tourists do not merely treat individual attractions as isolated objects of visitation; rather, they form an overall perception of the Historic Centre through walking, viewing, stopping, taking photos, and moving across spaces. This indicates that the tourism appeal of HCM derives not only from individual heritage buildings or scenic spots, but also from the embodied, processual, and immersive experiences created by its continuous street-block spaces.

Finally, in the dimension of emotional evaluation and tourism experience quality, positive evaluative terms such as “worth”, “like”, “interesting”, “nice”, “beautiful”, “pretty”,

and “wonderful” occupy an important position, suggesting that tourists generally hold a relatively positive emotional attitude toward HCM. Its aesthetic value, cultural atmosphere, and visitor experience are highly recognized. At the same time, however, terms such as “crowded”, and “too many” also indicate that high visitor density, spatial crowding, and the concentrated distribution of commercial spaces may affect tourists’ comfort and the quality of their experience to some extent.

It can be seen that the online tourism image of HCM is not simply determined by its officially promoted identity as a world heritage site. Rather, it is continuously constructed through the interaction among recognition of heritage value, perception of landmark spaces, slow-mobility tourism practices, and comprehensive experiential evaluation.

4.1.3. Platform-Stratified Comparison and Potential Platform Bias

Different platforms vary in their user groups, content production mechanisms, and usage contexts, which may influence how tourists express their perceptions of the same object. This is also an issue that must be recognized before conducting multi-platform data analysis. To further examine differences in tourist perceptions across platforms, this study performed a platform-stratified comparison of the top-ranked keywords based on cumulative TF-IDF scores. The results show that “Macao”, “architecture”, and “history” consistently rank among the most prominent keywords across all four platforms (Figure 2). This convergence indicates that architectural heritage and historical narratives constitute relatively stable components of the cognitive image of HCM across different platform contexts.

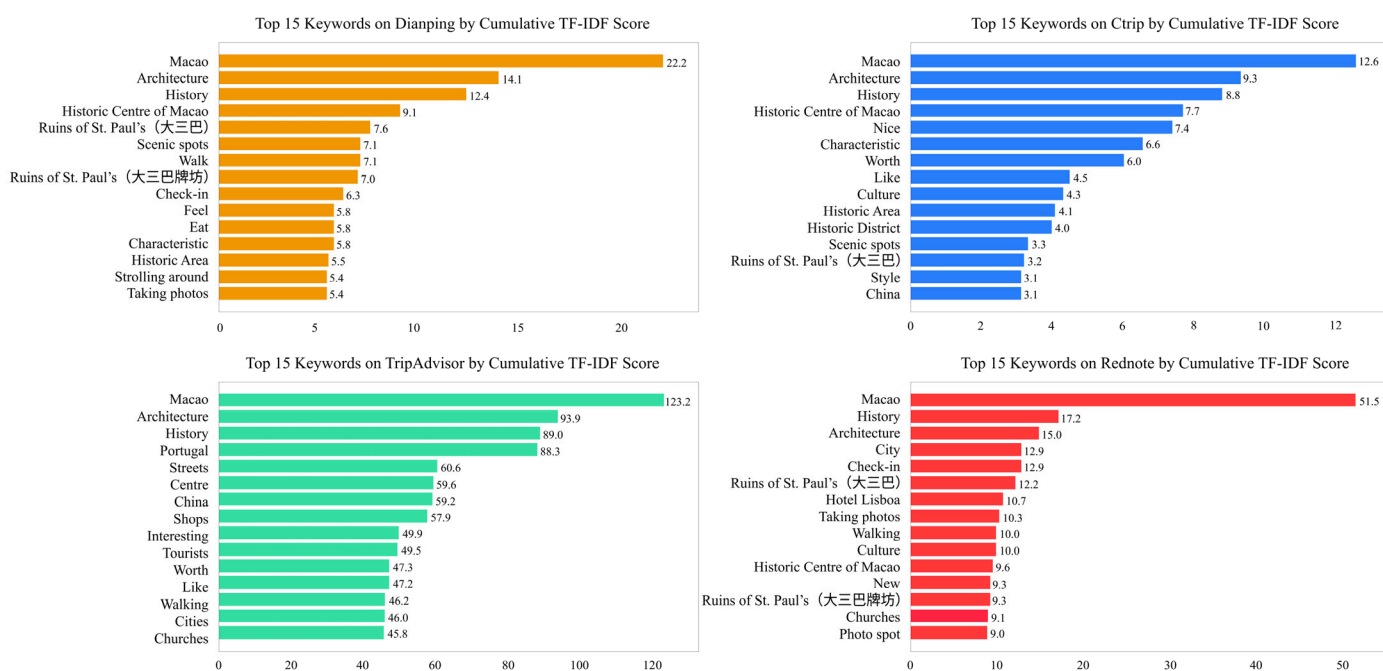


Figure 2. Top 15 keywords ranked by cumulative TF-IDF score within each platform. TF-IDF scores were calculated separately for each platform. Therefore, the figure is intended to compare the ranking and semantic composition of high-weight keywords within each platform, rather than the absolute score values across platforms. Created by authors.

TripAdvisor reviews place greater emphasis on cultural comparison and the heritage gaze. Terms such as “Portugal”, “China”, “streets”, “centre”, and “churches” rank prominently within the TripAdvisor corpus, indicating that its users tend to interpret HCM through Chinese–Western cultural encounters, historical remains, and cross-cultural comparison. As TripAdvisor primarily serves an international tourism audience, this

pattern may be associated with the comparative perspectives of users from diverse cultural backgrounds.

Rednote, by contrast, presents distinct features of visualized and socialized expression. Terms such as “check-in”, “taking photos”, and “photo spot” are more prominent in the Rednote corpus, suggesting that users of this platform are more likely to perceive HCM as a tourism space that can be viewed, photographed, and shared. Compared with conventional scenic-site reviews, Rednote comments place greater emphasis on travel routes, visual presentation, and the effects of social media dissemination, reflecting the influence of social media platforms on the production of destination image.

The top-ranked keywords on Dianping more strongly reflect local life and consumption experiences. In addition to terms such as “architecture”, “history”, and “Historic Centre of Macao”, words such as “walk”, “eat”, “strolling around”, “feel”, and “characteristic” are relatively prominent. This indicates that Dianping users’ perceptions of HCM are not limited to heritage landscapes, but also include dining, shopping, street walking, and everyday consumption scenarios.

Ctrip reviews, meanwhile, highlight overall evaluation and destination-level summarization. Terms such as “nice”, “characteristic”, “worth”, and “like” suggest that Ctrip users are more inclined to make overall satisfaction judgments about their visiting experience. Terms such as “culture”, “historic district”, “historic area”, “style”, and “China” further indicate that, while expressing their travel experiences, Ctrip reviews also provide general descriptions of the cultural attributes, urban landscape, and historical space of the Historic Centre.

4.2. BERTopic-Based Semantic Topic Analysis

This study employed BERTopic to conduct topic modeling on all tourist reviews, identifying a total of 18 valid semantic topics (Topics 0–17) (Figure 3). Based on the distribution of topic keywords, these topics cover the major semantic content of online reviews of HCM, including history and culture, architectural heritage, spatial landmarks, district-based sightseeing, local consumption, festive atmosphere, night-time landscapes, transport connections, urban entertainment, and crowding experiences. Specifically, Topics 0, 2, 8, and 16 mainly include keywords such as “Portugal”, “architecture”, “China”, “history”, “culture”, and “Historic Centre of Macao”, reflecting tourists’ attention to historical and cultural value, architectural heritage, and the characteristics of Sino-Western cultural encounters. Topics 1, 6, 12, and 17 highlight words such as “Ruins of St. Paul’s (大三巴)”, “church”, “St. Paul (圣保禄)”, “Senado Square”, “historic district”, and “attractions”, indicating that tourists’ perceptions of the historic centre rely heavily on iconic landmarks, religious buildings, and district spaces. In addition to heritage value and landmark spaces, the topic model also reveals multiple scenarios embedded in tourists’ experiences. Keywords in Topic 5 and 7, such as “walking”, “visiting”, “on foot”, and “streets”, suggest that slow-paced walking tours constitute an important mode of experiencing HCM. Words in Topic 15, including “night”, “lights”, “beautiful”, and “romantic”, show that night-time landscapes and visual atmosphere are also important components of tourists’ evaluations. Topics 3, 4, 9, 10, and 13 involve words such as “shops”, “food”, “shopping”, “New Year”, “Christmas”, “casino”, “tourists”, and “crowded”, suggesting that tourists’ perceptions are not limited to heritage buildings themselves, but also extend to consumption, festivals, urban entertainment, and visitor density. About tourist experience evaluation, Topic 11 contains a concentration of words such as “crowded”, “too many”, “tourists”, “unable”, “crowds”, and “taking photos”, indicating that high-density tourist flows have become an important factor affecting the quality of tourist experiences. By contrast, Topic 14 is dom-

inated by positive evaluative expressions such as “nice” and “happy”, reflecting tourists’ favorable overall assessments of their visits.

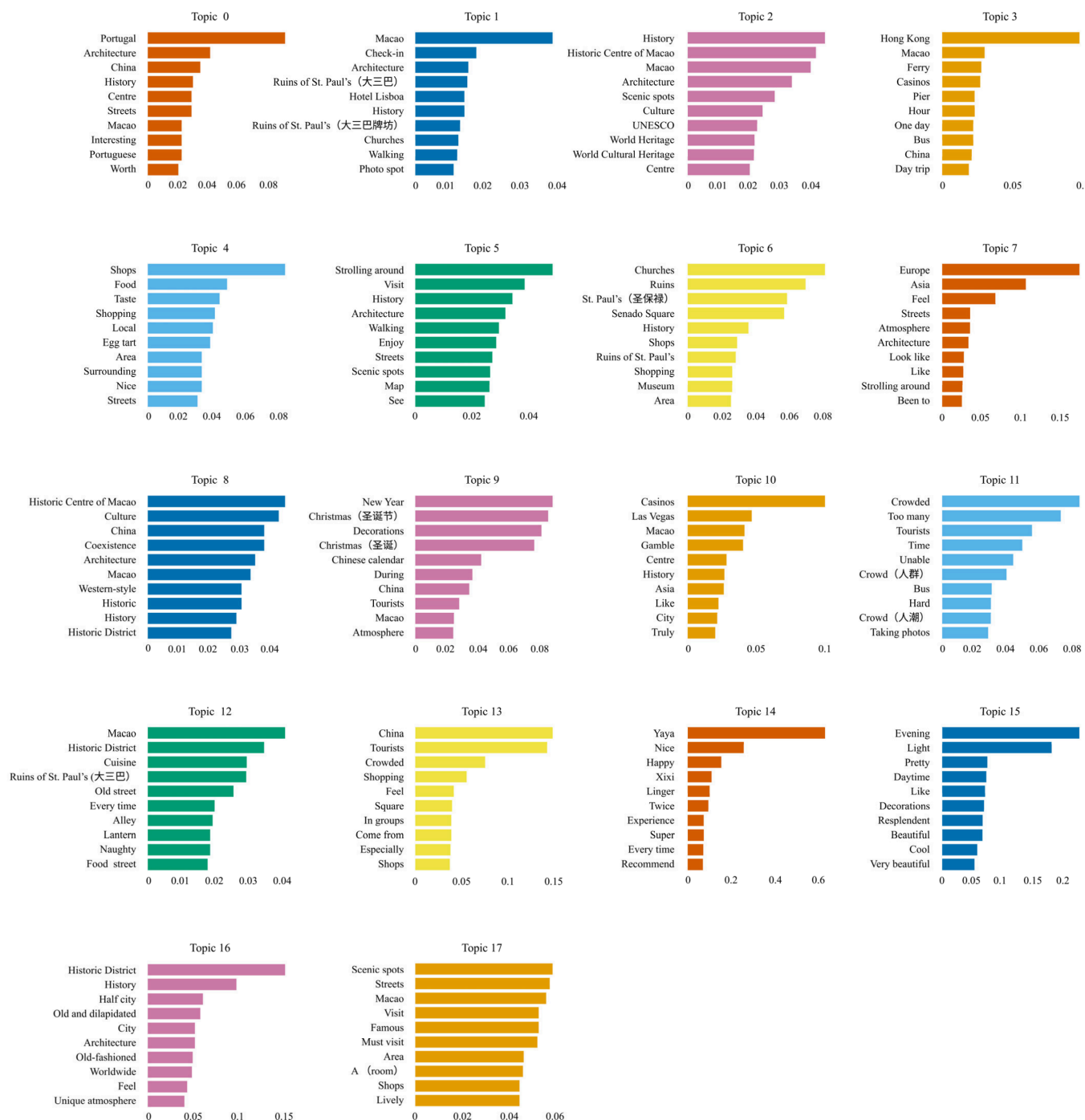


Figure 3. Topic keywords and their weight distribution in UGC related to HCM. Created by authors.

The two authors independently examined the top c-TF-IDF keywords and the representative review documents associated with each topic and proposed labels reflecting their predominant semantic content. Disagreements were resolved through discussion and reference to the original review contexts. To improve the transparency and interpretability of the topic-naming process, Table 4 presents the topic groupings, thematic labels, representative c-TF-IDF keywords, and typical review excerpts derived from the 18 topics.

The further hierarchical clustering results show that the 18 topics generally form three semantic clusters (Figure 4). The first cluster, represented by Topics 14 and 15, mainly re-

flects immediate emotional evaluations, night-time landscapes, and aesthetic experiences. The second cluster covers integrated tourism experiences related to festive atmosphere, commercial consumption, religious architecture, transport connections, cross-cultural comparison, and urban entertainment. The third cluster mainly revolves around history and culture, architectural heritage, old-town landscapes, iconic attractions, and cultural identity, reflecting tourists' heritage-space cognition and landmark-based perception of HCM.

Table 4. Topic Labels, Top c-TF-IDF Keywords, and Representative Review Excerpts Identified by BERTopic. Created by authors.

Topic	Topic Label	Top c-TF-IDF Keywords	Representative Review Text
0, 2, 8, 16	Historical and cultural value	history, culture, Portugal, China, heritage, historic district, historic centre of Macao	Macao is small, but it has a distinctive charm. Most of its major attractions are concentrated in this area and can easily be reached on foot. There are many churches, most of which are legacies of the Portuguese colonial period. They are beautiful and give the city a unique character. The famous Ruins of St. Paul's are also located here, along with a street lined with local food stalls. It is truly a great place for sightseeing and enjoying delicious food.
1, 6	Ruins of St. Paul's and religious heritage	Ruins of St. Paul's, churches, St. Paul	For anyone visiting Macao, this is probably a must-see attraction. Macao is home to numerous Western-style Catholic churches, among which St. Paul's Church is one of the oldest and most famous. It is commonly known as the Ruins of St. Paul's. When visiting the façade, you should not only admire its imposing and magnificent appearance but also take a closer look at the intricate reliefs and consider their symbolic meanings. From the top downward, there is first a large cross, followed by three levels of niches, each containing a bronze statue cast by one of Macao's early cannon foundries. Directly beneath the cross is a bronze statue in the shape of a dove.
5, 7	Walking and street experience	walking, on foot, streets, visit	It is very pleasant to stroll around here, and you can see architecture that is completely different from elsewhere!
15	Night-time landscape and atmosphere	night, lights, beautiful, romantic	Late at night, when everything is quiet, Macao's historic district enters its most peaceful time of day. I really enjoy this sense of tranquility.
3, 4, 9, 10, 13	Consumption, Entertainment, and Festival Experiences	casinos, shops, food, New Year, Christmas, tourists	I visited during the Chinese New Year holiday, when the square was decorated with various festive displays and filled with a lively holiday atmosphere. It is one of Macao's most popular attractions and was extremely crowded with tourists. When visiting Macao, it is well worth stopping by, especially since it can easily be included along the way.
11, 14	Tourist Experience Evaluation	crowded, nice, happy	There is plenty to see, but the large number of tourists makes it difficult to explore at a leisurely pace. It was hard enough to take photos in the crowd, and even moving around was difficult.

The topic similarity matrix further indicates that history and culture, architectural heritage, and the overall perception of the historic centre are strongly associated, while spatial landmarks, district-based sightseeing, local consumption, and check-in experiences also show relatively high similarity (Figure 5). Overall, the BERTopic results suggest that the online tourism image of HCM is not shaped by a single heritage label, but is formed through the interweaving of three semantic clusters: heritage-space cognition, landmark and district experiences, and integrated tourism experiences.

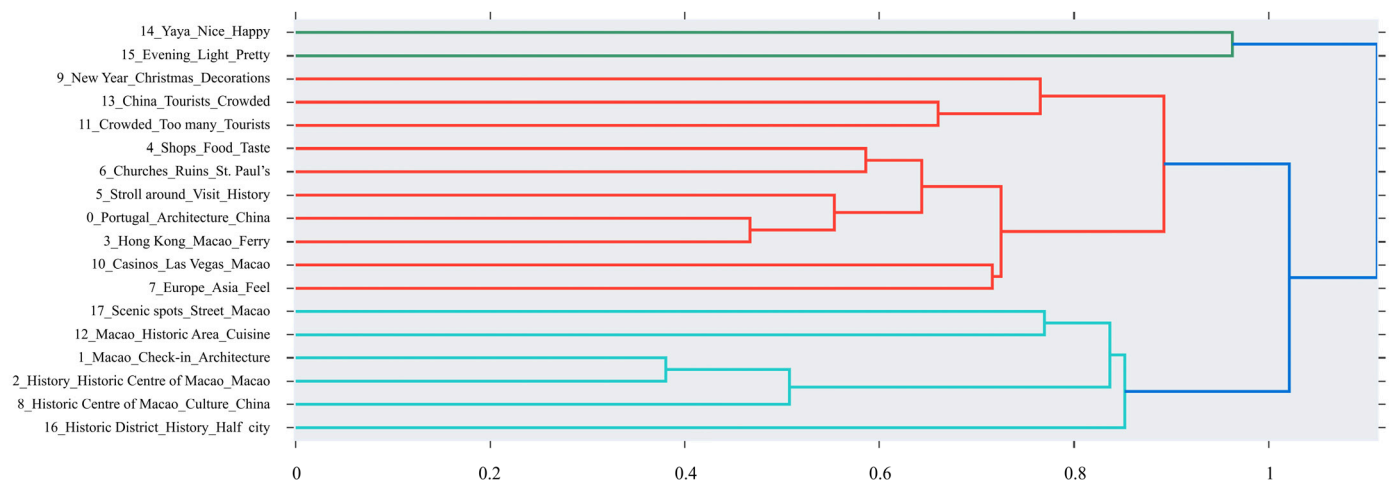


Figure 4. Hierarchical clustering dendrogram of the 18 BERTopic topics based on inter-topic semantic distances. The colored branches indicate the three broader topic clusters discussed in the text (green, red, and cyan). The blue branches indicate the higher-level hierarchical merging of these clusters. Created by authors.

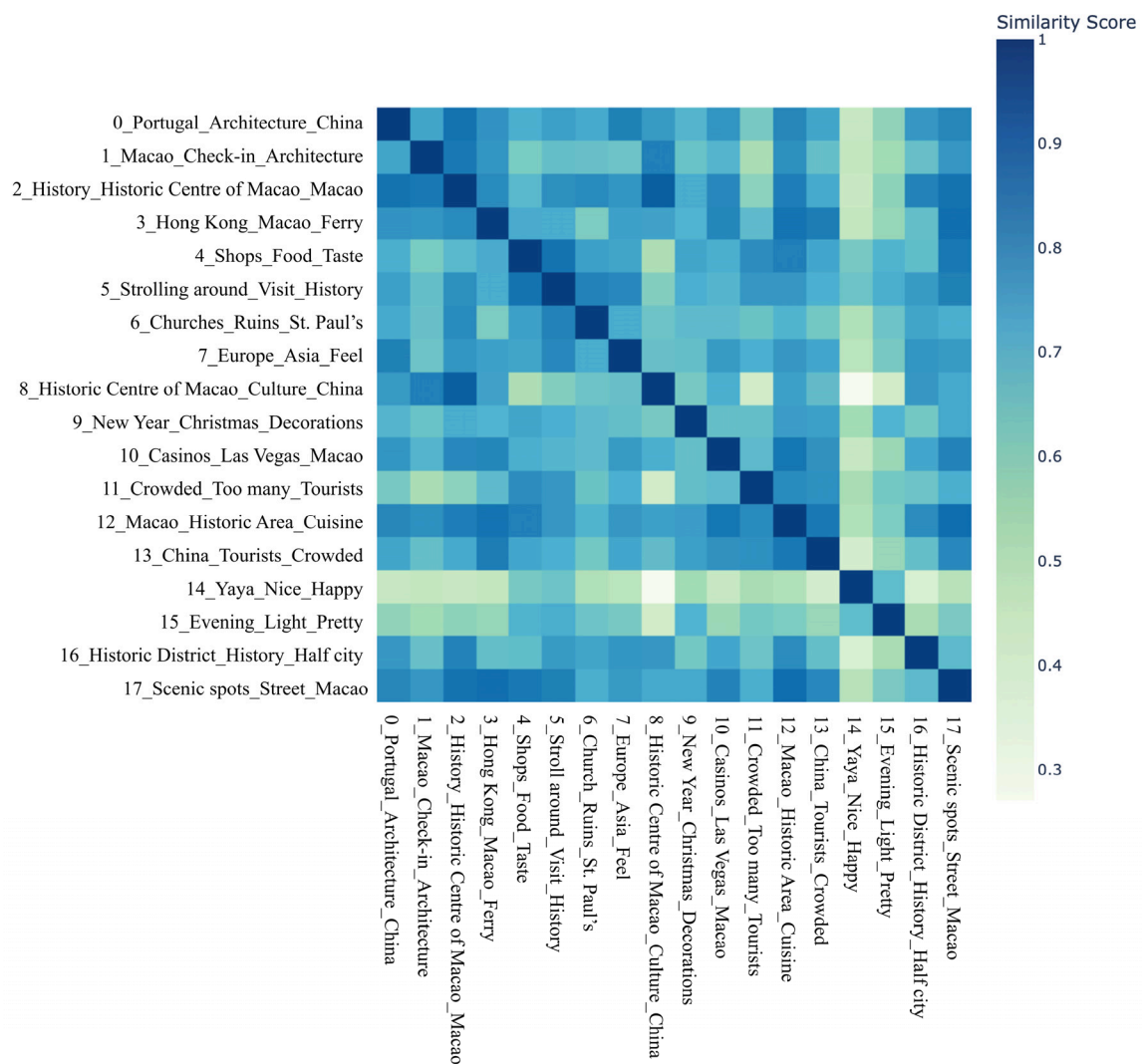


Figure 5. Pairwise semantic similarity matrix of the 18 BERTopic topics. Darker cells indicate higher semantic similarity between topic representations. Created by authors.

4.3. RoBERTa-Based Sentiment Analysis

Sentiment analysis was used to identify the emotional tendencies of user reviews and to further examine keyword differences among positive, negative, and neutral comments. As shown in Table 5 among the 3781 reviews, positive comments account for the largest proportion, with 3593 entries. There are 97 neutral comments and 91 negative comments, with relatively similar numbers. Table 6 presents the distribution of sentiment across the four review platforms. Positive sentiment predominates on all platforms, indicating that tourists generally hold favorable perceptions of HCM. Nevertheless, slight cross-platform differences can be observed. Ctrip (98.4%) and Rednote (98.0%) exhibit relatively higher proportions of positive evaluations, whereas Dianping (3.2%) and TripAdvisor (2.9%) contain comparatively more negative comments. TripAdvisor also shows the highest proportion of neutral reviews (2.8%). These variations suggest that differences in platform characteristics and user-generated content contexts may influence the way tourists express their travel experiences. However, the overall sentiment distribution remains highly consistent across platforms, indicating a broadly consistent pattern of predominantly positive sentiment across the four platforms. Representative examples of positive and negative reviews are provided in Table 7.

Table 5. Distribution of Sentiment Categories. Created by authors.

Emotional Types	Positive	Neutral	Negative
Number of Comments	3593	97	91
Percentage	95.03%	2.57%	2.41%

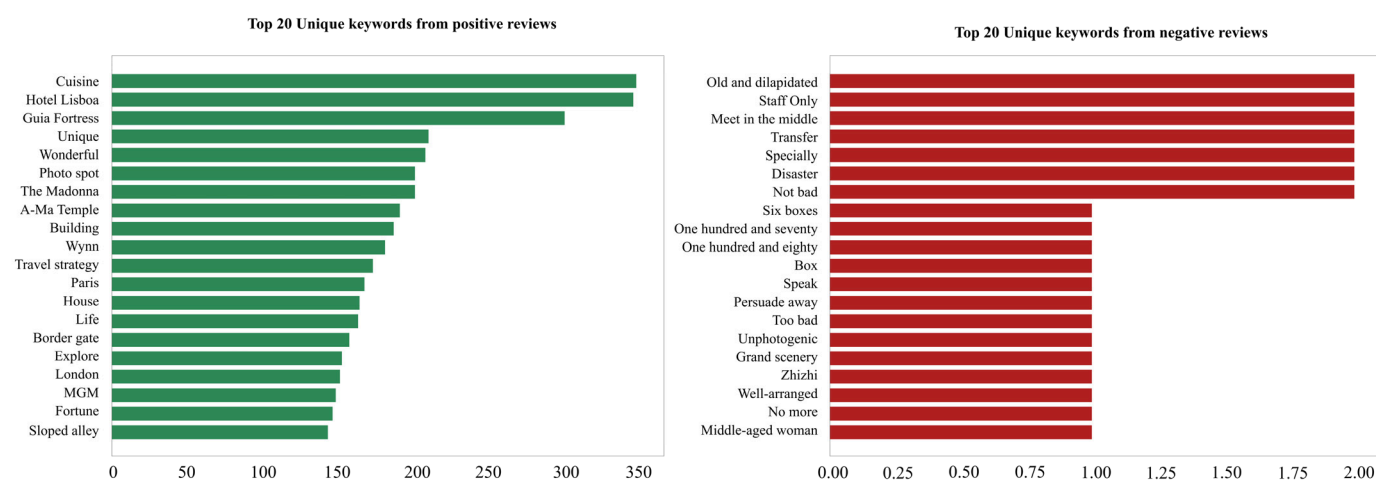
Table 6. Sentiment Distribution Across Different Review Platforms. Created by authors.

Platform	Positive	Neutral	Negative
Rednote	479 (98.0%)	10 (2.0%)	0 (0.0%)
Ctrip	187 (98.4%)	2 (1.1%)	1 (0.5%)
Dianping	323 (94.7%)	7 (2.1%)	11 (3.2%)
TripAdvisor	2604 (94.3%)	78 (2.8%)	79 (2.9%)

As shown in Figure 6, the keywords in positive comments mainly relate to urban experience, landscape appreciation, food consumption, and travel recommendations, including words such as “cuisine”, “wonderful”, “photo spot”, “life”, “sloped alley”, and “explore”. This indicates that tourists’ positive evaluations are largely derived from the visual appeal, photography experience, food experience, and urban walking experience of HCM and its surrounding spaces. By contrast, the keywords in negative comments are generally low in frequency and semantically dispersed. Words such as “Staff Only”, “persuade away”, “too bad”, and “no more” indicate that some tourists experienced dissatisfaction during their visits. These negative expressions may be related to spatial restrictions, service experiences, gaps between expectations and reality, or inconvenience in specific scenarios. However, judging from the keyword distribution, negative comments do not form a highly concentrated thematic structure, but instead appear as scattered and individualized experiential feedback. These findings provide reference points for the subsequent optimization of tourist flow management, spatial guidance, and visitor experience services.

Table 7. Representative Text Examples of Positive and Negative Reviews. Created by authors.

Positive Comments
HCM is one of the city’s major tourist attractions. It embodies Macao’s history and development. When visiting the area, I found that many places still preserve the distinctive features of the last century, especially the fusion of arcade architecture and Portuguese architectural styles. For me, travelling is about walking around, observing, and learning about history, so this place suited my interests very well.
It is a wonderful architectural site and a great place to visit, with a wide variety of shops, restaurants, and entertainment venues to explore. There are also many different local markets along the streets.
I really enjoyed walking around the Historic Centre and looking at the different buildings influenced by Portuguese architecture. The area has a very nostalgic atmosphere. There is no admission fee, and street snacks and restaurants can be found everywhere.
Walking through HCM was very refreshing. The architecture is quite old-fashioned, yet it blends very well with modern design. The streets are very clean, and the people are very friendly. Walking around the city felt very comfortable and safe.
When visiting Macao, this is definitely a place worth walking around. HCM, with its yellow-and-white building façades, has a distinctive charm and is also a great spot for taking memorable photos. There are shops and restaurants along the way, and visitors can walk from here to another attraction, the Ruins of St. Paul’s.
It is a must-visit when sightseeing in Macao. Many places have information signs explaining the city’s history. Everything is within walking distance, which is simply wonderful.
Negative Comments
The weather was poor that day, hot and humid, and there were an excessive number of Chinese tourists, which made the sightseeing experience rather unpleasant. The attraction itself was also not particularly appealing.
With McDonald’s, clothing stores, cosmetics shops, and pharmacies, the Historic Centre seems to have been somewhat overwhelmed by commercialized businesses.
Come and see it for yourself, you may be left speechless. I would not recommend visiting during the summer holiday, as there are far too many people. It might be better to come in the evening.
This place was indeed far too crowded. We visited on a weekday, but it was still packed with tourists, and even finding a decent restaurant was difficult. Visiting once is enough!
I was curious to see what Macao might be like, but I have to say I was somewhat disappointed. The famous Historic Centre is nothing more than a group of Portuguese-style buildings, which are very ordinary and poorly maintained. For me, the city was not worth visiting.
It was not that spectacular. I had hoped there would be more activities related to culture and art. Shopping in the city centre was rather boring.

**Figure 6.** Top 20 Keywords in Positive and Negative Reviews. Created by authors.

5. Discussion

This section discusses the main findings in relation to the online tourism image of the Historic Centre of Macao (HCM). Based on the results derived from user-generated content (UGC), the discussion focuses on how tourists perceive heritage value, spatial landmarks, embodied travel practices, and emotional experience. It further explains how these dimensions jointly shape the representation of HCM as both a World Cultural Heritage site and an everyday tourism destination.

5.1. *The Multidimensional Structure of the Online Tourism Image of HCM*

The prominence of terms related to architecture, history, churches, and Portuguese culture is consistent with previous studies that identified architectural heritage and Sino-Portuguese cultural hybridity as central components of Macao's tourism image [7]. However, the present study further shows that the online tourism image of HCM is mainly composed of four dimensions: perceptions of history, culture, and heritage value; perceptions of spatial landmarks and urban landscapes; sightseeing behaviors and embodied experiences; and emotional evaluations and the quality of tourism experiences. Tourists first identify the heritage attributes and cross-cultural characteristics of HCM through words such as "history", "culture", "World Heritage", "Portugal", and "China". They then form landmark-based perceptions through specific spatial symbols such as the "Ruins of St. Paul's", "Senado Square", "churches", "streets", and "squares". Finally, they complete the overall construction of the destination image through walking, photography, visiting, shopping, and emotional evaluation. This suggests that HCM is both a cultural heritage space centered on history, architecture, and its World Heritage identity, and a tourism experience space where visitors engage in urban walking, visual consumption, social sharing, and emotional evaluation. Heritage value, urban landscapes, consumption activities, and embodied experiences are intertwined to form tourists' comprehensive perceptions. Unlike previous studies that primarily emphasized architectural heritage and visual landmarks, the present findings show that tourists' online image of HCM also incorporates embodied walking practices and affective evaluations.

5.2. *The Joint Effects of Landmark-Based Perception and Cross-Cultural Imagination*

Spatial nodes such as the Ruins of St. Paul's, Senado Square, churches, and streets exhibit high visibility in tourist reviews, a finding consistent with the high-frequency word analysis reported in previous studies [9]. Among them, the Ruins of St. Paul's and Senado Square constitute representative places through which tourists identify HCM. This indicates that the online tourism image of HCM demonstrates a clear tendency toward landmark-based perception. Shi et al. also found that hotspots of perceived historical ambience in the Historic Centre of Macao were concentrated around Senado Square and the steps leading to the Ruins of St. Paul's [50]. However, the landmarkization of heritage may also create certain problems. When tourist experiences are overly concentrated around a few popular nodes, such as the Ruins of St. Paul's, other heritage buildings, street textures, and everyday living spaces may be overlooked. High-frequency words such as "Portugal", "China", "Europe", "Asia", and "Hong Kong" indicate that tourists often construct the uniqueness of HCM through cross-regional and cross-cultural comparisons. Especially, users from all over the world on TripAdvisor place greater emphasis on cross-cultural communication and the interpretation of cultural heritage.

5.3. *Embodied Practices and Emotional Evaluation as Co-Constructors of the Tourism Experience*

The tourism experience of HCM has clear embodied and processual characteristics. Words such as "walking", "strolling", "visiting", "sightseeing", "taking photos", and

“seeing” indicate that tourists mainly experience, understand, and evaluate the historic centre through slow mobility, viewing, photography, and continuous spatial movement. The compact scale of the district, dense concentration of historic buildings, and continuous public spaces make it suitable for walking-based experiences. Therefore, tourists’ perceptions of HCM are not simply evaluations of individual attractions but are gradually formed through movement along streets and lanes, viewing of landmarks, visual recording, and spatial pauses. It reflects the embodied process value of heritage tourism [51]. The sentiment analysis results show that tourists generally hold a positive attitude toward HCM. Positive comments are mainly related to cultural heritage value, urban landscapes, sightseeing convenience, photography experiences, and urban walking. At the same time, expressions such as “crowded”, “too many” and “unable” in negative comments indicate that high-density tourist flows, spatial crowding, commercialization, and gaps between expectations and actual experiences may still weaken the quality of tourist experiences. These findings indicate that visitor-flow pressure is a central constraint on the otherwise positive walking-based heritage experience of HCM.

5.4. Implication for Heritage Communication and Tourism Management of HCM

Based on the text-mining results, several targeted management implications can be proposed. First, the TF-IDF results show that tourists’ perceptions are centered on architectural heritage, historical narratives, and Sino-Portuguese cultural characteristics. Heritage interpretation should therefore extend beyond a few iconic landmarks and connect buildings, squares, streets, and everyday living spaces through integrated digital and on-site narratives. Second, the BERTopic results indicate that tourists perceive HCM as an interconnected heritage landscape, suggesting the need for thematic walking routes linking major heritage sites with surrounding streets, commercial spaces, and local neighborhoods. Third, as negative evaluations are mainly associated with crowding, commercialization, and expectation gaps, visitor management should strengthen real-time crowd monitoring, peak-period guidance, and the promotion of alternative heritage nodes. Finally, cross-platform differences suggest the value of differentiated communication strategies: visually oriented and route-based content for Rednote, local-life information for Dianping, concise travel guidance for Ctrip, and multilingual heritage interpretation for TripAdvisor.

6. Conclusions

Taking the Historic Centre of Macao (HCM) as a study case, this study developed and applied a multi-model text-mining framework integrating TF-IDF, BERTopic, and RoBERTa to 3781 user-generated content (UGC) reviews collected from Rednote, Ctrip, Dianping, and TripAdvisor. The findings show that the online tourism image of HCM is constructed through four interrelated dimensions: perceptions of history, culture, and heritage value; perceptions of spatial landmarks and urban landscapes; sightseeing behaviors and embodied experiences; and emotional evaluations and tourism experience quality. Across the four platforms, architecture, history, Portuguese cultural elements, churches, the Ruins of St. Paul’s, and Senado Square constitute a relatively stable cognitive core, while platform-specific differences reflect variations in user groups and content-production contexts. BERTopic further identifies three semantic clusters centered on heritage-space cognition, landmark and district experiences, and integrated tourism experiences, indicating that tourists perceive HCM not only through iconic buildings but also through walking, food, shopping, festivals, and street-level encounters. RoBERTa-based sentiment analysis shows that tourists’ overall evaluations are predominantly positive, but crowding, dense tourist flows, commercialization, and gaps between expectations and

actual experiences remain important factors affecting the quality of tourist experiences. Together, these results provide a case-specific account of how the online tourism image of HCM is cognitively organized, semantically structured, and emotionally evaluated.

This study still has several limitations. First, the four platforms differ in user composition, review conventions, content-production mechanisms, and language environments, while TripAdvisor accounts for a substantially larger proportion of the corpus. Although the platform-stratified analysis identified a relatively stable cross-platform core centered on Macao, architecture, and history, the combined results may still give greater weight to the cross-cultural heritage perspectives represented in TripAdvisor reviews. Second, the translation of English, Portuguese, and Japanese reviews into simplified Chinese may reduce certain culturally specific expressions and emotional nuances. Third, some review texts are short and colloquial, which may limit the model's ability to identify deeper cultural meanings. Future research could further incorporate multilingual reviews, image data, tourist interviews, and spatial trajectory data to conduct multimodal and cross-group comparative analyses of the tourism image of HCM.

Author Contributions: Methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, visualization, X.X.; conceptualization, writing—review and editing, supervision, project administration, funding acquisition, Q.Z. All authors agree to be accountable for all aspects of the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Macao Polytechnic University Research Fund, grant number RP/FCHS-04/2025, under the project “A Study on the Multi-Stakeholder Participation Mechanisms in the Safeguarding of Intangible Cultural Heritage in Macao”.

Data Availability Statement: The social media dataset analyzed in this research was retrieved from open-access platforms (Ctrip, Rednote, TripAdvisor, Dianping). The processed and anonymized data supporting the findings of this study may be made available by the corresponding author upon reasonable request, subject to platform terms of service and ethical considerations.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UGC	User-Generated Content
TF-IDF	Term Frequency–Inverse Document Frequency
BERTopic	BERT-based topic modeling
LDA	Latent Dirichlet Allocation
RoBERTa	Robustly Optimized BERT Pretraining Approach
HCM	Historic Centre of Macao

References

1. Pike, S. Destination Image Analysis—A Review of 142 Papers from 1973 to 2000. *Tour. Manag.* **2002**, *23*, 541–549. [[CrossRef](#)]
2. Molinillo, S.; Liébana-Cabanillas, F.; Anaya-Sánchez, R.; Buhalis, D. DMO Online Platforms: Image and Intention to Visit. *Tour. Manag.* **2018**, *65*, 116–130. [[CrossRef](#)]
3. Hinton, T. Social Media Has the Greatest Influence on Travel Destination Choices. Available online: <https://www.statista.com/chart/30135/media-influences-on-travel-destination> (accessed on 5 April 2026).
4. Marine-Roig, E.; Ferrer-Rosell, B. Measuring the Gap between Projected and Perceived Destination Images of Catalonia Using Compositional Analysis. *Tour. Manag.* **2018**, *68*, 236–249. [[CrossRef](#)]
5. Wei, Y.; Chen, M. Tracing the Evolution of Tourist Perception of Destination Image: A Multi-Method Analysis of a Cultural Heritage Tourist Site. *Sustainability* **2025**, *17*, 5476. [[CrossRef](#)]

6. UNESCO. Historic Centre of Macao. Available online: <https://whc.unesco.org/en/list/1110/> (accessed on 5 April 2026).
7. Amaro, V.; Simpson, T. The Macau Arraial: Portuguese Heritage, Serious Games, and Postcolonial Identity in a Chinese Tourist City. *Tour. Stud.* **2024**, *24*, 385–409. [\[CrossRef\]](#)
8. Ge, Y.; Wei, K.; Wang, Z.; Huang, Y.; Zhu, R. Identifying and Evaluating Cultural Genes in the Historic Centre of Macao: A Multi-Stakeholder Perspective. *Buildings* **2026**, *16*, 1517. [\[CrossRef\]](#)
9. Jia, M.; Feng, J.; Chen, Y.; Zhao, C. Visual Analysis of Social Media Data on Experiences at a World Heritage Tourist Destination: Historic Centre of Macau. *Buildings* **2024**, *14*, 2188. [\[CrossRef\]](#)
10. Chen, X.; Zhang, H.; Zuo, T.; Tang, X. When Language Shapes Landscape: Quantifying Tourism Image Discrepancy of Macao Historic Centre through Bilingual Lens. *J. Tour. Cult. Change* **2025**, *23*, 741–760. [\[CrossRef\]](#)
11. Wong, C.U.I.; Qi, S. Tracking the Evolution of a Destination's Image by Text-Mining Online Reviews—The Case of Macau. *Tour. Manag. Perspect.* **2017**, *23*, 19–29. [\[CrossRef\]](#)
12. Baloglu, S.; McCleary, K.W. A Model of Destination Image Formation. *Ann. Tour. Res.* **1999**, *26*, 868–897. [\[CrossRef\]](#)
13. Hunt, J.D. Image as a Factor in Tourism Development. *J. Travel Res.* **1975**, *13*, 1–7. [\[CrossRef\]](#)
14. Crompton, J.L. An Assessment of the Image of Mexico as a Vacation Destination and the Influence of Geographical Location Upon That Image. *J. Travel Res.* **1979**, *17*, 18–23. [\[CrossRef\]](#)
15. Lai, K.; Li, X. Tourism Destination Image: Conceptual Problems and Definitional Solutions. *J. Travel Res.* **2016**, *55*, 1065–1080. [\[CrossRef\]](#)
16. Echtner, C.M.; Ritchie, J.R.B. The Measurement of Destination Image: An Empirical Assessment. *J. Travel Res.* **1993**, *31*, 3–13. [\[CrossRef\]](#)
17. Bramwell, B.; Rawding, L. Tourism Marketing Images of Industrial Cities. *Ann. Tour. Res.* **1996**, *23*, 201–221. [\[CrossRef\]](#)
18. Beerli, A.; Martín, J.D. Factors Influencing Destination Image. *Ann. Tour. Res.* **2004**, *31*, 657–681. [\[CrossRef\]](#)
19. Richards, G. Cultural Tourism: A Review of Recent Research and Trends. *J. Hosp. Tour. Manag.* **2018**, *36*, 12–21. [\[CrossRef\]](#)
20. Timothy, D.; Boyd, S. *Heritage Tourism*; Prentice Hall: Harlow, UK, 2003.
21. Wang, N. Rethinking Authenticity in Tourism Experience. *Ann. Tour. Res.* **1999**, *26*, 349–370. [\[CrossRef\]](#)
22. Chen, C.-F.; Chen, F.-S. Experience Quality, Perceived Value, Satisfaction and Behavioral Intentions for Heritage Tourists. *Tour. Manag.* **2010**, *31*, 29–35. [\[CrossRef\]](#)
23. Poria, Y.; Butler, R.; Airey, D. The Core of Heritage Tourism. *Ann. Tour. Res.* **2003**, *30*, 238–254. [\[CrossRef\]](#)
24. Suntikul, W.; Jachna, T. Profiling the Heritage Experience in Macao's Historic Center. *Int. J. Tour. Res.* **2016**, *18*, 308–318. [\[CrossRef\]](#)
25. Cai, P. World Heritage in Macau and Its Blocked View Problems. *Front. Humanit. Soc. Sci.* **2025**, *5*, 143–148. [\[CrossRef\]](#)
26. Wang, P.; Fu, H. The Influence of Different Visual Elements of High-Density Urban Observation Decks on the Visual Behavior and Place Identity of Tourists and Residents. *Appl. Sci.* **2025**, *15*, 3875. [\[CrossRef\]](#)
27. Chen, Z.; Wan, P.Y.K. Interdependence between Community Participation and Sustainable Livelihood Development: Community Perspectives on Heritage Conservation in Macao. *J. Herit. Tour.* **2023**, *18*, 1–17. [\[CrossRef\]](#)
28. Zhao, S.; Cai, R.; Wang, X.; Liu, X. Study on the Spatial Equilibrium of Cultural and Tourism Resources in Macao, China. *Open Geosci.* **2025**, *17*, 20250874. [\[CrossRef\]](#)
29. Zhang, Y.; Wang, B. Study on the Activation, Utilization and Conservation of Historic Buildings in Macao: A Case of the Development of Tourist Routes in Macao. *Urban Archit.* **2022**, *19*, 166–170. [\[CrossRef\]](#)
30. Kaplan, A.M.; Haenlein, M. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Bus. Horiz.* **2010**, *53*, 59–68. [\[CrossRef\]](#)
31. Wang, X.; Mou, N.; Zhu, S.; Yang, T.; Zhang, X.; Zhang, Y. How to Perceive Tourism Destination Image? A Visual Content Analysis Based on Inbound Tourists' Photos. *J. Destin. Mark. Manag.* **2024**, *33*, 100923. [\[CrossRef\]](#)
32. Narangajavana Kaosiri, Y.; Callarisa Fiol, L.J.; Moliner Tena, M.Á.; Rodríguez Artola, R.M.; Sánchez García, J. User-Generated Content Sources in Social Media: A New Approach to Explore Tourist Satisfaction. *J. Travel Res.* **2019**, *58*, 253–265. [\[CrossRef\]](#)
33. Sánchez-Franco, M.J.; Rey-Tienda, S. The Role of User-Generated Content in Tourism Decision-Making: An Exemplary Study of Andalusia, Spain. *Manag. Decis.* **2024**, *62*, 2292–2328. [\[CrossRef\]](#)
34. Aboalghanam, K.M.; AlFraihat, S.F.; Tarabieh, S. The Impact of User-Generated Content on Tourist Visit Intentions: The Mediating Role of Destination Imagery. *Adm. Sci.* **2025**, *15*, 117. [\[CrossRef\]](#)
35. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [\[CrossRef\]](#)
36. Aizawa, A. An Information-Theoretic Perspective of TF-IDF Measures. *Inf. Process. Manag.* **2003**, *39*, 45–65. [\[CrossRef\]](#)
37. Wang, Z.; Udomwong, P.; Fu, J.; Onpium, P. Destination Image: A Review from 2012 to 2023. *Cogent Soc. Sci.* **2023**, *9*, 2240569. [\[CrossRef\]](#)
38. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

39. Huang, C.; Wang, Q.; Yang, D.; Xu, F. Topic Mining of Tourist Attractions Based on a Seasonal Context Aware LDA Model. *Intell. Data Anal.* **2018**, *22*, 383–405. [[CrossRef](#)]
40. Ali, T.; Omar, B.; Soulaïmane, K. Analyzing Tourism Reviews Using an LDA Topic-Based Sentiment Analysis Approach. *MethodsX* **2022**, *9*, 101894. [[CrossRef](#)] [[PubMed](#)]
41. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186.
42. Egger, R.; Yu, J. Identifying Hidden Semantic Structures in Instagram Data: A Topic Modelling Comparison. *Tour. Rev.* **2021**, *77*, 1234–1246. [[CrossRef](#)]
43. Egger, R.; Yu, J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Front. Sociol.* **2022**, *7*, 886498. [[CrossRef](#)] [[PubMed](#)]
44. Chen, Q.; Liu, R.; Jiang, Q.; Xu, S. Exploring Cross-cultural Disparities in Tourists' Perceived Images: A Text Mining and Sentiment Analysis Study Using LDA and BERT-BiLSTM Models. *Data Technol. Appl.* **2024**, *58*, 669–690. [[CrossRef](#)]
45. Srianan, S.; Nanthamornphong, A.; Phucharoen, C. Advancing Tourism Sentiment Analysis: A Comparative Evaluation of Traditional Machine Learning, Deep Learning, and Transformer Models on Imbalanced Datasets. *Inf. Technol. Tour.* **2025**, *27*, 1011–1045. [[CrossRef](#)]
46. Bi, J.-W.; Zhu, X.-E.; Han, T.-Y. Text Analysis in Tourism and Hospitality: A Comprehensive Review. *J. Travel Res.* **2024**, *63*, 1847–1869. [[CrossRef](#)]
47. Antonio, N.; Correia, M.B.; Ribeiro, F.P. Exploring User-Generated Content for Improving Destination Knowledge: The Case of Two World Heritage Cities. *Sustainability* **2020**, *12*, 9654. [[CrossRef](#)]
48. Salingaros, N.A. Façade Psychology Is Hardwired: AI Selects Windows Supporting Health. *Buildings* **2025**, *15*, 1645. [[CrossRef](#)]
49. Lam, J.M.S.; Ismail, H.; Lee, S. From Desktop to Destination: User-Generated Content Platforms, Co-Created Online Experiences, Destination Image and Satisfaction. *J. Destin. Mark. Manag.* **2020**, *18*, 100490. [[CrossRef](#)]
50. Shi, J.; Yan, Y.; Li, M.; Zhou, L. Measuring the Convergence and Divergence in Urban Street Perception among Residents and Tourists through Deep Learning: A Case Study of Macau. *Land* **2024**, *13*, 345. [[CrossRef](#)]
51. Svensson, M. Walking in the Historic Neighbourhoods of Beijing: Walking as an Embodied Encounter with Heritage and Urban Developments. *Int. J. Herit. Stud.* **2021**, *27*, 792–805. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.