

Article

Research on Consumer Reviews of Sports Service Complexes Transformed from Old Industrial Plants Based on Topic Mining and Sentiment Analysis

Lixin Jia *, Du Jiang  and Junchao Wang

Physical Education Institute, Xi'an University of Architecture and Technology, Xi'an 710055, China; jiangdu@xauat.edu.cn (D.J.); wangjc0216@xauat.edu.cn (J.W.)

* Correspondence: jialx@xauat.edu.cn; Tel.: +86-189-9184-7800

Abstract

As urban renewal progresses, transforming old industrial plants into sports service complexes has become a prevalent strategy to revitalize urban stock space. To evaluate the actual performance of this spatial reproduction from the users' perspective, this study analyzes 3702 consumer reviews from five typical Chinese sports service complexes using an interdisciplinary text-mining framework. We employ the LDA topic model to extract core dimensions of consumer concern and utilize the DistilBERT model for fine-grained, sentence-level sentiment computation. The results reveal that consumer attention spans five key topics: venue services, cultural business districts, sports training, industrial integration, and heritage utilization. Critically, the sentiment analysis uncovers a structural paradox: while topics associated with cultural and heritage utilization trigger highly positive emotions and strong place identity, core functional modules like sports training and venue services generate substantial negative feedback. This disparity highlights a profound friction between rigid historical industrial architectural structures and the flexible demands of modern servicescapes, reflecting a clear path dependency in spatial transformation. Based on these theoretical and empirical findings, we propose targeted optimization strategies, including the structural reconstruction of training services and the flexible upgrading of basic venue facilities, to transition these complexes from initial physical construction toward long-term, service-oriented operation.

Keywords: sports service complex; renovation of old industrial plants; topic mining; sentiment analysis

1. Introduction

In the post-industrial era, the adaptive reuse of industrial heritage has emerged as a global strategy for urban renewal and economic revitalization. Revitalizing idle urban stock space and stimulating emerging consumption sectors are critical to sustainable development. As the world's largest emerging economy, China has proactively integrated sports consumption into its national urban regeneration agenda. Recent macro-policies, such as "The Strategic Plan for Expanding Domestic Demand (2022–2035)" [1] and "Opinions on Releasing the Potential of Sports Consumption to Further Promote the High-quality Development of the Sports Industry" [2], underscore the strategic transition toward accelerating the construction of a modern sports industry system. Within this context, sports service complexes transformed from obsolete industrial facilities have become vital spatial



Received: 17 June 2026

Revised: 19 July 2026

Accepted: 29 July 2026

Published: 1 August 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.

carriers for mass sports consumption. These sports service complexes serve as essential environments for consumer behavior. Their diversity and appeal act as primary drivers for promoting mass sports participation and continuously expanding the market. In 2025, the General Office of the State Council issued the “Urban Renewal Action” to encourage and promote the renewal and transformation of old factories, revitalize the use of idle and inefficient industrial facilities, and implant new formats and functions [3]. Transformed industrial sports service complexes have integrated novel business formats and functions, unlocking significant consumption potential. Given the immense consumption potential of the sports market, identifying strategies to release this potential and promote the high-quality development of the sports industry has become a primary priority for both government agencies and academic researchers.

2. Literature Review

As global deindustrialization progresses alongside urban micro-regeneration strategies, the adaptive reuse of old industrial plants has become a core strategy for revitalizing urban stock space and stimulating community vitality [4]. The transformation of industrial heritage into sports service complexes is not only the physical restoration of abandoned buildings but also the reconstruction of social space interwoven with capital, power, and mass consumption demands. Examined through the lens of Henri Lefebvre’s “production of space” theory, this transformation profoundly reflects the recoding of historical industrial symbols and modern leisure culture [5]. While consumer-centric evaluations have gained significant traction in general urban studies, research specifically addressing the adaptive reuse of industrial heritage into sports facilities still traditionally emphasizes macro-level property rights replacement and top-down planning orientations. Consequently, there remains a need for a more systematic evaluation of the actual performance of this specific spatial reproduction from the micro-perspective of “space users” (i.e., consumers) [6]. Since the ultimate goal of spatial reproduction is to serve the public, the subjective perceptions and emotional feedback expressed in online reviews serve as a crucial proxy for evaluating its practical effectiveness. These textual reviews represent the “lived space”—the daily reality and genuine interactions of users—effectively bridging the gap between top-down physical renovation and bottom-up consumer experience. Therefore, analyzing large-scale review text provides an indispensable tool for measuring how well industrial renovations align with modern service demands.

The classic “Servicescape” theory emphasizes that the layout, facility conditions, and spatial atmosphere of the physical environment directly determine consumers’ emotional experiences and behavioral intentions [7]. Sports consumption is highly experiential and immersive, but the renovation of old industrial heritage is often accompanied by a strong “path dependency” effect [8]. When the large-span spatial structures and complex movement lines, originally designed for heavy industrial production, are transformed into modern sports service complexes, they inevitably generate structural contradictions in spatial comfort, functional zoning, and traffic micro-circulation. This mismatch between “industrial historical heritage” and “modern service demands” constitutes the core pain point limiting the consumer experience [9]. Nevertheless, previous studies mostly adopted qualitative case descriptions or a one-way supply-side perspective, lacking systematic quantitative measurement and deep attribution of consumers’ genuine pain points.

Accurately capturing consumers’ multidimensional needs and emotional feedback regarding sports service complexes is the prerequisite for formulating scientific spatial optimization strategies. Previous evaluation studies on sports venue operations or urban renewal projects heavily relied on traditional questionnaire surveys or qualitative interviews [10], which are often constrained by limitations such as limited sample sizes,

significant subjective recall biases, and lagging timeliness. With the development of natural language processing technology, large-scale online user-generated content (UGC) provides a new path for capturing consumers' spontaneous and authentic feedback. The Latent Dirichlet Allocation (LDA) topic model and advanced natural language processing (NLP) techniques—particularly recent applications of transformer-based sentiment analysis and Large Language Models (LLMs)—have been widely and successfully applied in various fields, including the tourism sector [11], urban planning [12], and sports management [13], proving their powerful ability to bottom-up mine potential dimensions of concern from unstructured massive texts. While these modern text analysis methods are well-established, applying them to the distinct interdisciplinary context of “transforming industrial heritage into sports service complexes” requires a more nuanced approach. First, unlike general public spaces, these transformed complexes possess unique spatial and functional complexities. Previous applications often treat such sites as a homogenized whole, which may not fully deconstruct consumers' differentiated demands across distinct operational modules at a fine-grained level. Second, within these specific hybrid spaces (industrial aesthetics meeting sports functionality), traditional word frequency analysis or basic sentiment analysis often struggles to accurately identify complex, topic-specific emotions, making it particularly difficult to locate the specific service breakpoints that trigger negative emotions, thus failing to directly guide refined operations [14]. To fill the aforementioned gaps, this study innovatively introduces a large-scale online review text mining framework, adopting the LDA topic model to extract core dimensions of concern, and combining the DistilBERT deep learning model to calculate sentiment tendencies at a fine-grained sentence level [15]. Methodologically, this study overcomes the limitations of traditional surveys; theoretically, it achieves the alignment of macro-spatial renewal policies with micro authentic consumer experiences, providing a solid data-driven basis for breaking the path dependency of industrial heritage spatial transformation and releasing sports consumption potential.

3. Research Design

This study proposes a consumer review analysis framework based on LDA topic mining and DistilBERT sentiment analysis for the sports service complex in the transformation of old industrial plants. Consumer reviews from Meituan and Dianping—China's leading location-based lifestyle and consumer review platforms (analogous to Yelp)—were utilized as the primary data sources to analyze the topic content and emotional tendency of consumers. The research framework is shown in Figure 1.

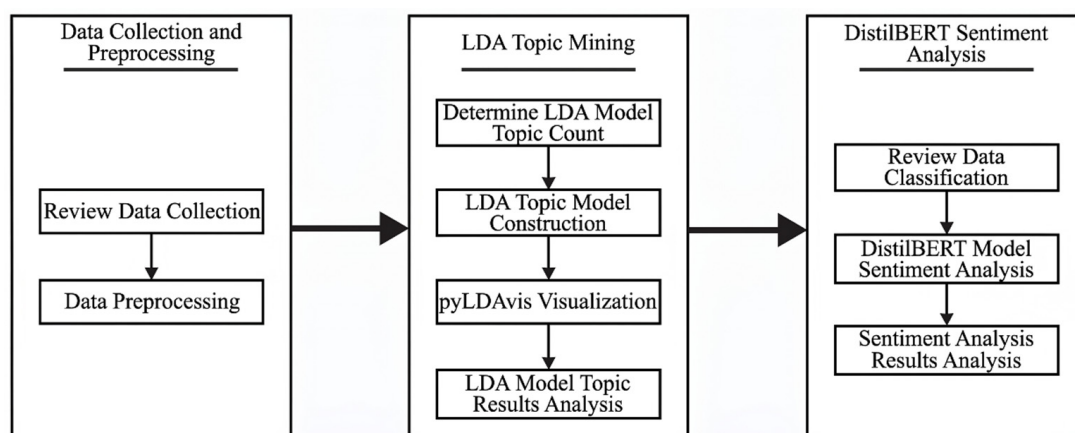


Figure 1. Research framework.

3.1. Data Sources and Data Preprocessing

3.1.1. Data Source

This study draws its research subjects from the authoritative “Public Notice on the Selection Results of Typical Cases of Sports Service Complexes” [16] issued by the General Administration of Sport of China. Within this national list, five sports service complexes belong to the specific category of being transformed from old industrial plants: Borui Sports Complex, Sanlinqiao Sports Culture Park, New Shougang High-end Industrial Comprehensive Service Area, Xiang Cube Sports Culture Complex, and Yuedong·Xinmenxi Sports Culture Industrial Park. Rather than conducting subjective sampling, this study includes all five of these relevant cases from the authoritative list as the research objects.

By including the complete set of old industrial-transformation cases from this national list, the research objects possess strong inherent representativeness. Furthermore, these five sites naturally capture significant quantitative and spatial diversity, ensuring the empirical findings possess broad generalizability. To explicitly illustrate this diversity, Table 1 presents a comparative overview of their key quantitative characteristics, including site area, post-reconstruction opening year, and regional demographics. Geographically, they capture diverse urban consumer contexts and varying visitor capacities, ranging from dense residential zones with a population exceeding 700,000 within a 3-kilometer radius (e.g., Sanlinqiao) to distinct cultural districts adjacent to historical ruins (e.g., Yuedong·Xinmenxi). In terms of scale, the cases span from mega-scale heavy industry regeneration projects covering over 8.6 million square meters (e.g., New Shougang) to community-embedded light industry micro-renewals occupying approximately 15,000 square meters (e.g., Borui). By encompassing these quantifiable variations, these five cases provide a robust foundation for evaluating the spatial reproduction of industrial heritage.

Table 1. Key quantitative characteristics of the selected sports service complexes.

Case Name	Former Identity (Industrial Heritage Type)	Year of Post-Reconstruction Opening	Regional Characteristics	Site Area (sq.m.)
Borui Sports Complex	Zhaofeng Ceramics Factory (Ceramics Manufacturing Facility)	July 2019	Located in the southeastern urban district, with dense surrounding residential areas	15,000
Sanlinqiao Sports Culture Park	Shanghai Rixiao Vacuum Flask Liner Factory (Vacuum Flask Liner Manufacturing Facility)	May 2019	Residential population exceeding 700,000 within a 3 km radius; adjacent to the economic radiation zone of higher education institutions	36,000
New Shougang High-end Industrial Comprehensive Service Area	Original Factory Area of Shougang Group (Iron and Steel Industrial Base)	Gradually opened post-2018	The only large-scale, contiguous undeveloped area within the central six districts	8,630,000
Xiang Cube Sports Culture Complex	Shanghai Machine Tool Plant, Shanghai Dongfeng Pharmaceutical Plant, etc. (Machinery Manufacturing/Pharmaceutical Facilities)	August 2016	Nearly 200,000 residents across 82 residential communities within a 3 km radius	30,000
Yuedong·Xinmenxi Sports Culture Industrial Park	Nanjing Printing and Dyeing Mill, No. 1 Cotton Textile Mill (Textile Printing and Dyeing Facilities)	2018	Adjacent to the Ming City Wall and Outer Qinhuai River; bordered to the east by Fenghuang Terrace and Waguan Temple ruins	84,000

As important platforms for consumer groups to express their opinions on consumption scenarios, Meituan and Dianping serve as crucial channels for merchants to interact with consumers in the digital era, providing rich data support for the study of consumer demand. Python 3.12.9 (hereinafter referred to as Python) was used to crawl the consumer review data of the aforementioned five sports service complexes on the Meituan and Dianping applications.

3.1.2. Data Preprocessing

Because the obtained raw data is not standardized, it cannot be directly analyzed [17]. Online review text contains noise data such as web address, emojis, etc., and is unstructured text [18]. If not handled, it may affect the results of subsequent topic mining, as well as the accuracy of text classification and sentiment analysis. Therefore, it is necessary to further process the original data to obtain the required actual data. Through data cleaning, incomplete data, erroneous data, duplicate data, and irregular data are transformed into data that can be directly substituted into the model. In this study, the original data is preprocessed in four steps: removing noise, performing word segmentation, removing stop words, and merging synonyms.

3.2. LDA Topic Mining

3.2.1. LDA Topic Model

This study utilizes Latent Dirichlet Allocation (LDA), a generative three-layer Bayesian probability model introduced by Blei et al. [19] to discover latent thematic structures in unstructured review texts. As an unsupervised machine learning approach, LDA is highly effective at clustering text data and capturing multi-dimensional consumer concerns by establishing document-topic and topic-word probability distributions. The principle of the LDA topic model is shown in Figure 2.

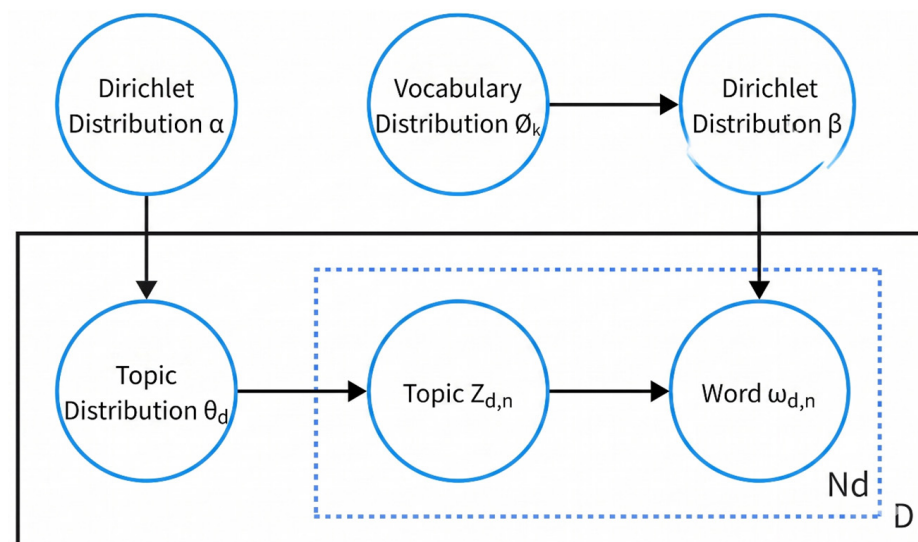


Figure 2. Principle of LDA model.

3.2.2. LDA Model Topic Number Selection

Selecting the appropriate number of topics is crucial to ensure the effectiveness of the LDA model. Too many topics will lead to semantic repetition between topics, which makes the model unable to effectively distinguish similar topics. If the number of topics is too small, the topics with different semantics may be classified into one category, which makes it difficult to accurately reveal the diversity in the text. This study uses the two indicators of Perplexity and Coherence to determine the number of topics.

The perplexity of the model is calculated by using the `log_perplexity` function in the Gensim 4.4.3 library in Python. The mathematical expression of the perplexity is shown in Formula (1). The calculation of the perplexity is based on the likelihood function, and the performance of the model is determined by measuring the accuracy of the model's prediction of unseen documents. At the same time, the topic coherence can be calculated by the `CoherenceModel` function in the Gensim library, and the coherence is evaluated by

calculating the mutual information value between the words in the topic. The higher the mutual information value, the stronger the correlation between words.

$$\text{Perplexity} = \exp \left(- \frac{\sum_{d=1}^{|D|} \log(\omega_d)}{\sum_{d=1}^{|D|} Nd} \right) \quad (1)$$

Usually, the smaller the perplexity value is, the greater the coherence is, indicating that the model predicts the document more accurately. As the number of topics increases, the perplexity will decrease because the increased number of topics improves the flexibility of the model to capture data structures. When the number of topics is too large, the perplexity may decrease slightly, but it may lead to overfitting of the model, thus losing versatility and interpretability. Therefore, the inflection point in the perplexity and coherence curve is usually regarded as the optimal number of topics, because the decline in perplexity before the inflection point indicates the improvement of the model, and the decline after the inflection point tends to be gentle.

3.3. Sentiment Analysis

3.3.1. Review Data Classification and Granularity

Firstly, the original data is processed by word segmentation and filtering stop words according to the data preprocessing method introduced above. Secondly, the text is classified according to the topic of the text. Finally, this study adopts granular sentiment analysis at the sentence level. The content of a cell in the original data Excel file is called a text, and the text is divided into multiple sentences according to the punctuation marks in the text. Examples include Chinese periods, half-width or full-width exclamation marks, tab characters, etc. [20]. Processing the original data can effectively improve the accuracy of the analysis results.

3.3.2. DistilBERT Model

The current methods used in sentiment analysis are mainly divided into two types. One is based on machine learning, which uses various classification methods of machine learning to identify emotions. Second, the method based on a semantic dictionary first constructs a dictionary or list of emotional words, and uses the dictionary to judge the tendency of emotions [21]. The lexicon-based method constructs a dictionary of emotional words to judge sentiment tendencies. However, it heavily relies on predefined dictionaries and struggles to capture context-dependent emotions or domain-specific nuances (e.g., complex expressions regarding industrial aesthetics). Meanwhile, classical machine learning models, while more flexible, often lack deep semantic understanding and fail to accurately capture the structural dependencies within long, unstructured consumer reviews.

To overcome the limitations of these simpler approaches and achieve higher contextual accuracy, this study uses the DistilBERT-Base-Multilingual-Cased-Sentiments-Student model (hereinafter referred to as the DistilBERT model) to perform sentiment analysis on the review data. By employing knowledge distillation technology [22], this architecture efficiently compresses the complex mDeBERTa-v3 “teacher” model into a lightweight “student” model without significantly sacrificing multilingual processing accuracy. Utilizing zero-shot learning capabilities [23], DistilBERT bypasses the need for massive labeled datasets and efficiently outputs probability scores for positive, neutral, and negative sentiments at the sentence level.

4. Empirical Research

4.1. Data Collection and Preprocessing

This study uses Python to crawl the review data of five old industrial plants in Meituan and Dianping applications: Borui Sports Complex, Sanlinqiao Sports Culture Park, New Shougang High-end Industrial Comprehensive Service Area, Xiang Cube Sports Culture Complex and Yuedong·Xinmenxi Sports Culture Industrial Park. On 18 November 2025, the most recent consumer reviews were extracted in reverse chronological order. If the total number of reviews was less than 500, the maximum number of reviews was selected.

An examination of the crawled text reveals that there are some repetitive reviews and irrelevant reviews. To improve reproducibility and prevent potential selection bias, strict objective criteria were established for the manual elimination of ‘irrelevant reviews’. Specifically, the excluded texts strictly consisted of: (1) system-generated default reviews (e.g., ‘The user did not leave a text comment’); (2) reviews containing only emojis, meaningless punctuation, or random strings without substantive semantic content; and (3) commercial spam advertisements or reviews mistakenly describing completely unrelated sports service complexes. After filtering out repeated reviews and those violating these predefined rules, a highly valid dataset of 3702 reviews was obtained, as shown in Table 2.

Table 2. Sports service complex review data quantity list.

	Meituan		Dianping		Data Validity Rate
	Before Cleaning	After Cleaning	Before Cleaning	After Cleaning	
Borui Sports Complex	58	54	47	47	96.2%
Sanlinqiao Sports Culture Park	504	499	532	505	96.9%
Xiang Cube Sports Culture Complex	462	426	401	397	95.4%
New Shougang High-end Industrial Comprehensive Service Area	502	490	511	495	97.2%
Yuedong·Xinmenxi Sports Culture Industrial Park	414	398	395	391	97.5%

This study uses the Jieba library in Python for Chinese text segmentation of all the original data. The stop word list of Harbin Institute of Technology is used to screen out the stop words, such as words with a text length of 1, URL, English, numbers, adverbs and pronouns with high word frequency but low correlation with the research content. In order to improve the accuracy of word segmentation, this study uses the Harbin Institute of Technology Synonym Dictionary to construct a synonym lexicon. For example, “country” and “China” are merged into “China”, and words like “kids”, “little friends”, and “children” are merged into the unified term “children”.

4.2. Topic Mining of Reviews

4.2.1. Selection of the Number of Topics in LDA Model

After preprocessing the original data, the LDA topic model is used to mine the potential topic features of the review text, and the LDA topic model is constructed using Python’s Gensim library. When the number of topics $K \in [1, 10]$, the perplexity and coherence of the LDA topic model are observed. It is found that when the number of topics is 5, the coherence score is as high as 5.99, the perplexity score is low, and it is at the inflection point of the coherence curve, as shown in Figures 3 and 4. Beyond these statistical metrics, the semantic interpretability of the extracted topics was evaluated as a crucial criterion. A qualitative examination confirmed that setting $K = 5$ yields highly coherent and functionally distinct thematic dimensions (e.g., successfully separating basic facility attributes from specialized service experiences and surrounding commercial environments) without excessive semantic overlap. Tests with fewer topics forced distinct operational aspects to merge, thereby obscuring critical consumer friction points, whereas a larger number of

topics resulted in fragmented and overly granular sub-topics (e.g., awkwardly splitting specific types of sports equipment or isolated spatial features into independent topics). Therefore, the optimal number of topics is determined to be 5.

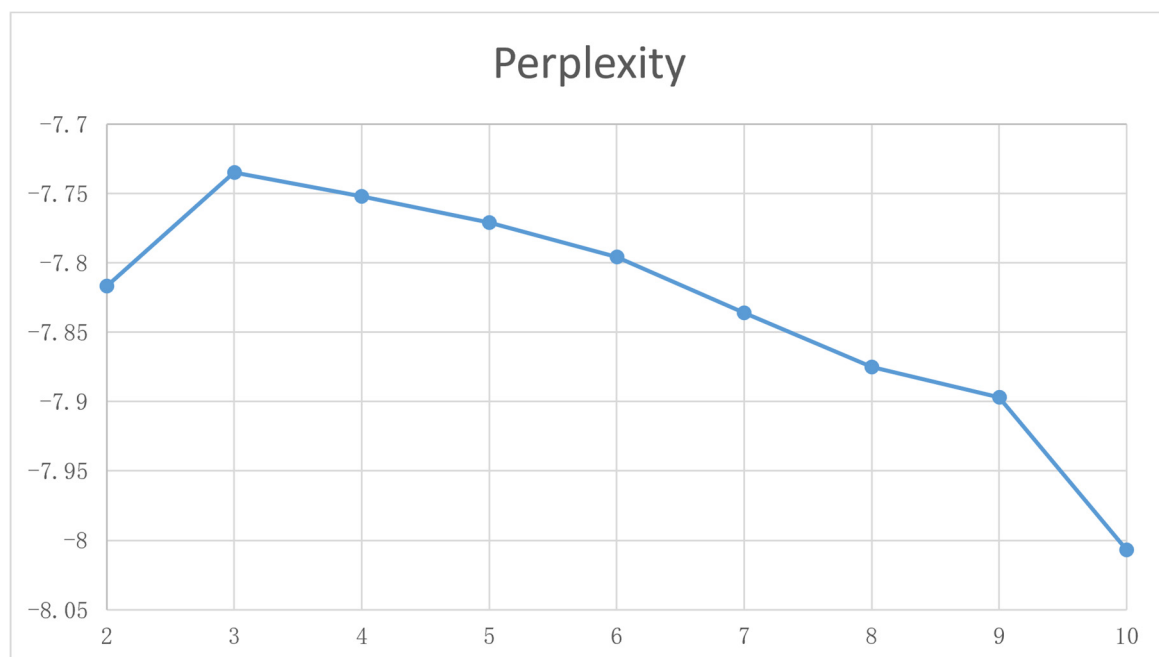


Figure 3. LDA topic perplexity.

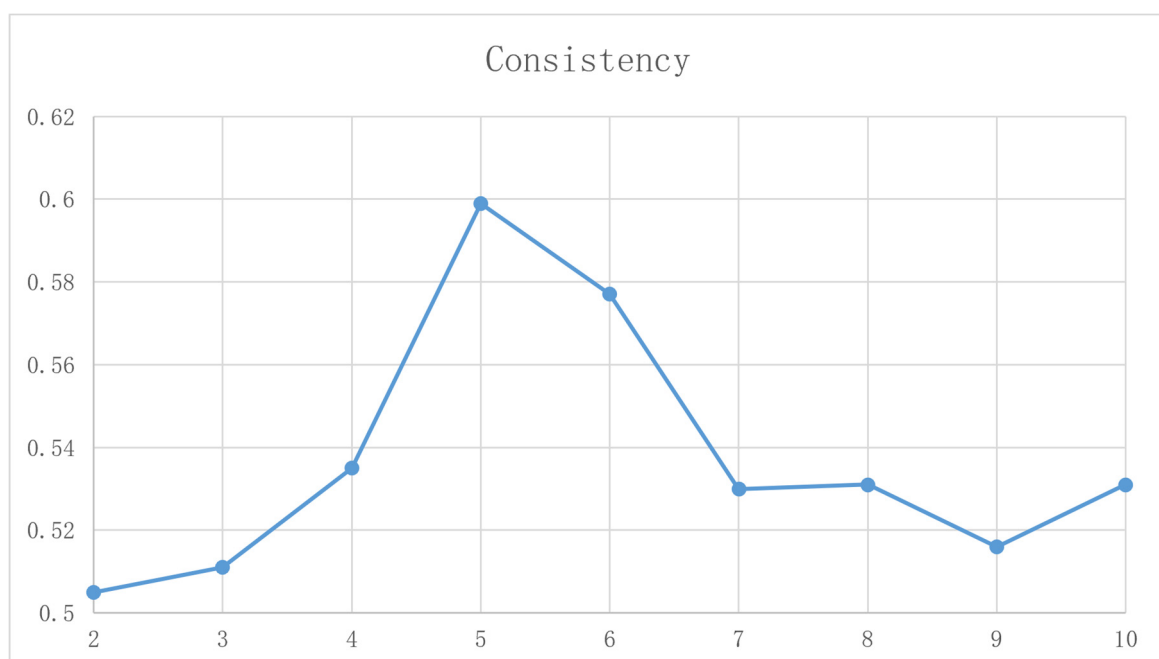


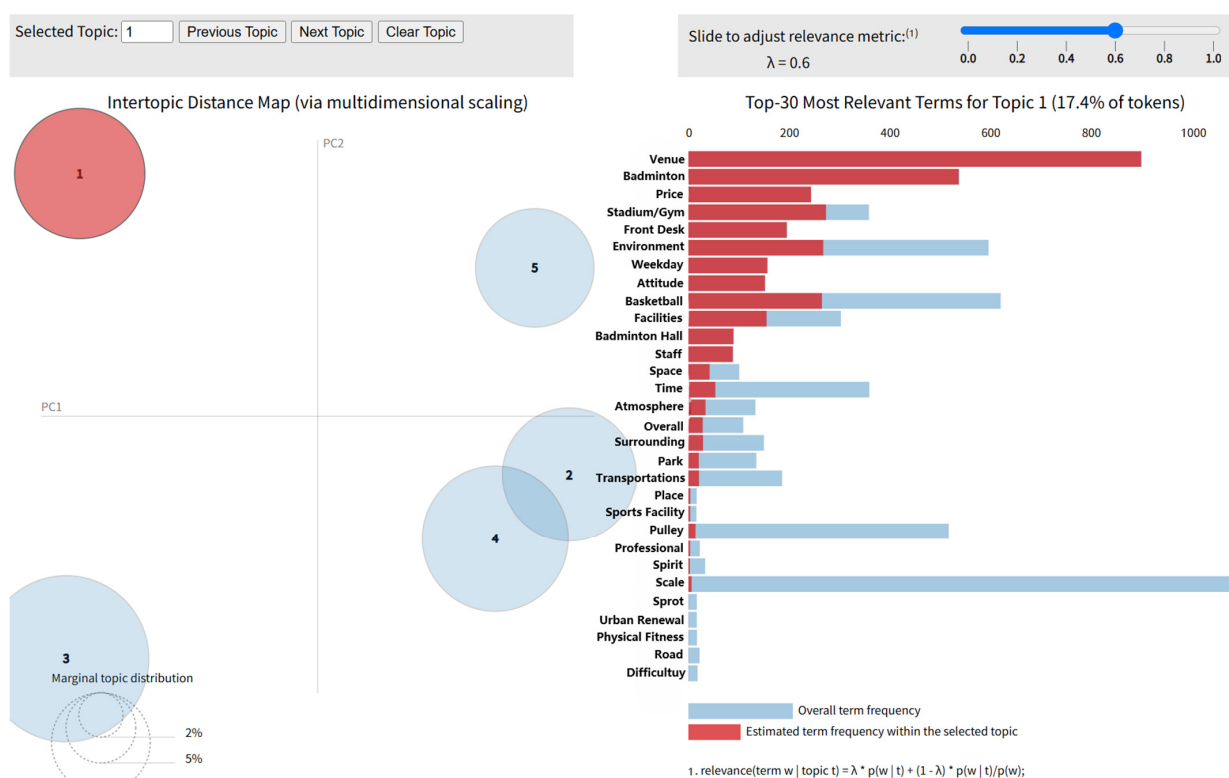
Figure 4. LDA topic coherence.

4.2.2. LDA Model Topic Recognition

After the analysis of the LDA topic model, Table 3 presents the top 15 relevant terms for each of the five topics. The pyLDAvis library is used to visualize the analysis results of the LDA topic model with a topic number of 5, and the bubble diagram is evenly distributed, indicating that the number of topics is 5 is more reasonable, as shown in Figure 5.

Table 3. LDA topic identification and key topic word distribution.

Topic	Top Relevant Terms				
Venue Services	Venue	Badminton	Price	Stadium/Gym	Front Desk
	Environment	Working Day	Attitude	Basketball	Facilities
	Badminton Hall	Staff	Space	Time	Atmosphere
Cultural Business Districts	Sports	Culture	Night Market	Catering	Residents
	Features	Location	Surrounding	Park	Workshop
	Chimney	Place	Commercial	Center	Facilities
	Children	Sports	Course	Professionalism	Coach
Sports Training	Parents	Physical Fitness	Teaching	Environment	Interest
	Swimming Pool	Institution	Time	Basketball	Approachability
	Industry	Industrial Park	Rock Climbing	Park	Basketball Hall
Industrial Integration	City	Cultural Industry	Basketball	Culture	Location
	History	Vitality	Landmark	Style	Gym
	Park	Parking Lot	Blast Furnace	Ski Jumping Platform	Subway
Heritage Utilization	Shopping Mall	Architecture	Shijingshan	Market	Walkway
	Attractions	Atmosphere	Park	Scenery	Transportation

**Figure 5.** pyLDAvis visual bubble diagram.

Regarding the interpretation of the visual bubble map, the right panel displays the ranking of the keyword scores in the entire corpus set and each topic. The higher the ranking of the key subject words extracted by the LDA model under each topic, the higher the score, indicating that the greater the probability that the subject word belongs to this topic. The visual bubble map on the left side of the topic is divided into two aspects: the bubble size of each topic and the bubble distribution position of each topic. First of all, it is the size of each topic bubble. The larger the topic bubble, the greater the importance of the topic in the text data, and vice versa. The topics are sorted by bubble size: topic 3, topic 4, topic 2, topic 1, topic 5, indicating that sports training is the topic of greatest concern to the public. Secondly, it is the location of bubble distribution, and the distance between bubbles reflects the similarity of the topic. Bubbles close to each other indicate that the topic similarity is higher, and vice versa. Among them, topic 1 and topic 3 are in a remote position from the center and are far away from other topics, indicating that topic 1 and topic 3 have high independence. The distance between Topic 2 and Topic 4 is close, indicating that the two topics have a certain relevance.

Furthermore, to provide robust quantitative evidence of the relationships between these thematic dimensions, we calculated a topic co-occurrence matrix (Table 4) based on the document-topic distribution probabilities. The matrix quantifies the frequency with which two distinct topics appear within the same consumer review. As shown in Table 4, significant structural correlations exist among specific modules. Notably, Topic 1 (Venue Services) and Topic 3 (Sports Training) exhibit the highest cross-topic co-occurrence frequency (297 times), statistically confirming that consumers evaluate basic facility conditions alongside professional training quality. Additionally, Topic 2 (Cultural Business Districts) serves as a central connecting hub, demonstrating strong linkages with Topic 1 (168 times), Topic 3 (168 times), and Topic 5 (Heritage Utilization, 161 times). This quantitative matrix substantiates that consumers perceive the spatial reproduction of industrial heritage as a highly integrated servicescape rather than a collection of isolated functional zones.

Table 4. Topic co-occurrence matrix based on consumer reviews.

	Topic 1 (Venue Services)	Topic 2 (Cultural Business Districts)	Topic 3 (Sports Training)	Topic 4 (Industrial Integration)	Topic 5 (Heritage Utilization)
Topic 1	1026	168	297	45	72
Topic 2	168	806	168	132	161
Topic 3	297	168	1391	45	60
Topic 4	45	132	45	680	60
Topic 5	72	161	60	60	672

4.2.3. LDA Model Topic Result Analysis

(1) For Topic 1 (Venue Services), the top 10 keywords are “venue, badminton, price, stadium/gym, front desk, environment, weekday, attitude, basketball, facilities”. This directly reflects that for basic venue services, consumers are highly focused on the daily consumption experience, particularly the balance between physical hardware and service software. The actual consumer reviews explicitly highlight the friction in service delivery. For example, a highly representative user review states: “The badminton venue has a great environment and reasonable prices on weekdays, but the front desk staff’s attitude is quite impatient during peak hours, and the supporting facilities lack humanized design.” Such representative feedback indicates that while the physical site supply function of the sports service complex transformed from old industrial plants is well-received, significant operational bottlenecks remain. Currently, there is a clear contradiction between supply and demand during peak hours, and many sports service complexes tend to prioritize physical infrastructure over service quality, resulting in a lack of user-centric amenities. Therefore, the optimization of venue services must transition from mere physical space supply to refined, service-oriented operations, directly addressing the micro-level friction in spatial reproduction.

(2) For Topic 2 (Cultural Business Districts), the top 10 keywords are “sports, culture, night market, catering, residents, characteristics, location, surrounding, park, factory”. This reflects that consumers highly value the integration of sports and cultural leisure in urban renewal. Instead of viewing these spaces purely as athletic facilities, a representative user review notes: “After exercising, it is great to directly walk into the night market and catering area within the park; the historical factory features and surrounding commercial facilities create a fantastic atmosphere.” Based on these authentic comments, operators should recognize that cultural IP mining and business coordination are crucial. Currently, some sports service complexes pay more attention to attracting initial foot traffic than ensuring long-term retention. Therefore, the operational focus must shift from mere physical space aggregation to creating immersive ‘culture + sports + business’ experiences to increase consumer stickiness.

(3) For Topic 3 (Sports Training), the top 10 keywords are “children, sports, curriculum, professionalism, coach, parents, physical fitness, teaching, environment, interest”. This reveals that youth sports training and professional coaching are major focal points for families. A typical parent’s review illustrates the current service friction: “The venue environment is good for children, but the sports curriculum is quite homogeneous, and it is hard for parents to verify the actual professionalism of the coaches during class time.” This feedback indicates a significant supply-demand mismatch where some institutions focus heavily on physical construction but neglect teaching depth and operational transparency. Therefore, to build a successful “education + sports + interest” ecosystem, operators must shift from generic venue provision to enhancing core teaching professionalism and creating tailored curricula.

(4) For Topic 4 (Industrial Integration), the top 10 keywords are “industry, industrial park, rock climbing, park, basketball hall, city, cultural industry, basketball, culture, location”. The text mining results objectively show that consumers frequently associate specific sports activities (e.g., rock climbing, basketball) with cultural and industrial contexts. A representative user review illustrates this connection: “The industrial park’s location in the city is highly accessible, and perfectly integrates the raw industrial culture with modern rock climbing facilities.” Based on these objective textual patterns, our expert interpretation suggests that consumers do not view these old industrial plants merely as isolated athletic venues, but rather as agglomerations of cultural and sports industries. The core competitiveness of these sports service complexes lies in creating a “culture + sports” ecosystem. However, the presence of negative sentiment in some reviews implies issues such as homogeneous competition and superficial cultural extraction. Therefore, future operational strategies should prioritize deep industrial synergy rather than basic spatial aggregation.

(5) For Topic 5 (Heritage Utilization), the top 10 keywords are “park, parking lot, blast furnace, ski jumping platform, subway, shopping mall, architecture, Shijingshan, market, walkway”. The objective co-occurrence of terms related to magnificent industrial structures (e.g., blast furnace, ski jumping platform) alongside basic urban amenities (e.g., subway, parking lot, shopping mall) reveals a dual demand from consumers. A typical review highlights this contrast: “Seeing the giant blast furnace and the magnificent architecture in Shijingshan is breathtaking, but the parking lot is insufficient and the walkway connecting to the subway needs improvement.” From an analytical perspective, we infer that while the unique aesthetic of industrial heritage (e.g., the spatial reproduction of blast furnaces) generates significant cultural attraction and place identity, the practical success of these sports service complexes heavily depends on modern supporting infrastructure. The tension between preserving historical atmosphere and fulfilling contemporary functional requirements remains a critical friction point. Operators must transition from purely facility-driven architectural preservation to comprehensive, service-oriented ecosystem management.”

4.3. Sentiment Analysis of Reviews

4.3.1. Classification of Review Data

After preprocessing the text data, classifying and cutting into sentences according to the topic of each text in the text data, a total of 10,650 sentences are obtained. Among them, the topic 1 venue services contains 2071 sentences, the topic 2 cultural business districts contains 1898 sentences, the topic 3 sports training contains 1678 sentences, the topic 4 industrial integration contains 2964 sentences, and the topic 5 heritage utilization contains 2039 sentences.

4.3.2. Model Validation and Performance

To ensure the reliability of the sentiment classification results before scaling up the empirical analysis, a rigorous validation procedure was conducted on a randomly selected sample of 300 review sentences. Given the total corpus size of 10,650 sentences, a random sample of 300 is statistically sufficient to achieve a 95% confidence level with an approximate 5.6% margin of error, which represents an accepted standard for manually validating NLP models in management research. To effectively mitigate potential annotation bias, this validation dataset was independently annotated by three researchers specializing in sports management.

To evaluate the reliability of the reference sample, we calculated the interrater agreement using the Fleiss' Kappa coefficient (appropriate for three or more raters). The resulting Fleiss' Kappa score was 0.84, indicating an 'almost perfect agreement' among the annotators. In rare cases of initial disagreement, the final sentiment label was determined through a consensus discussion to establish an objective gold standard. The classification performance of the DistilBERT model was then evaluated against this gold standard using four standard metrics including accuracy, precision, recall, and the F1 score. Table 5 presents the detailed evaluation results of the model across different sentiment categories.

Table 5. Performance metrics of the DistilBERT sentiment classification model.

Emotion Category	Precision	Recall	F1 Score	Support
Positive	0.8841	0.9137	0.8986	218
Neutral	0.7583	0.7231	0.7403	39
Negative	0.8493	0.8284	0.8387	43
Macro Average	0.8306	0.8217	0.8259	300
Overall Accuracy			0.8631	300

The model demonstrated robust classification performance with an overall accuracy of 0.8631. For the dominant positive class, the F1 score reached 0.8986. More importantly, the macro average F1 score across all three sentiment classes was 0.8259, indicating that the model maintains high sensitivity and predictive stability even when dealing with imbalanced text distributions. This statistical verification confirms that the automated sentiment computation is sufficiently rigorous to support the subsequent analysis of space reproduction and consumer behavioral patterns in sports service complexes.

4.3.3. DistilBERT Model Results Analysis

The sentences under each topic are analyzed using the DistilBERT model for sentiment analysis. The analysis results include the number of reviews and their proportions under the five topics, as shown in Figures 6 and 7. The results showed that the five topics had mostly positive reviews, followed by negative reviews and neutral reviews. From the perspective of the proportion of positive reviews, the five topics are ranked as topic 4 industrial integration, topic 5 heritage utilization, topic 2 cultural business districts, topic 1 venue services, topic 3 sports training. Among them, the proportion of positive reviews on topic 4 industrial integration, topic 5 heritage utilization and topic 2 cultural business districts exceeds the total proportion of positive reviews. From the perspective of the proportion of negative reviews, the five topics are ranked as topic 3 sports training, topic 1 venue service, topic 2 cultural business district, topic 5 heritage utilization, and topic 4 industrial integration. Among them, the proportion of negative reviews on topic 3 sports training and topic 1 venue service exceeds the total proportion of negative reviews. To statistically validate these observed differences in sentiment distributions among the five topics, a Pearson's Chi-square test of independence was conducted. The results yielded a

highly significant difference ($\chi^2 = 584.287$, $df = 8$, $p < 0.001$), robustly confirming that consumer sentiment tendencies are structurally dependent on the specific service topics.

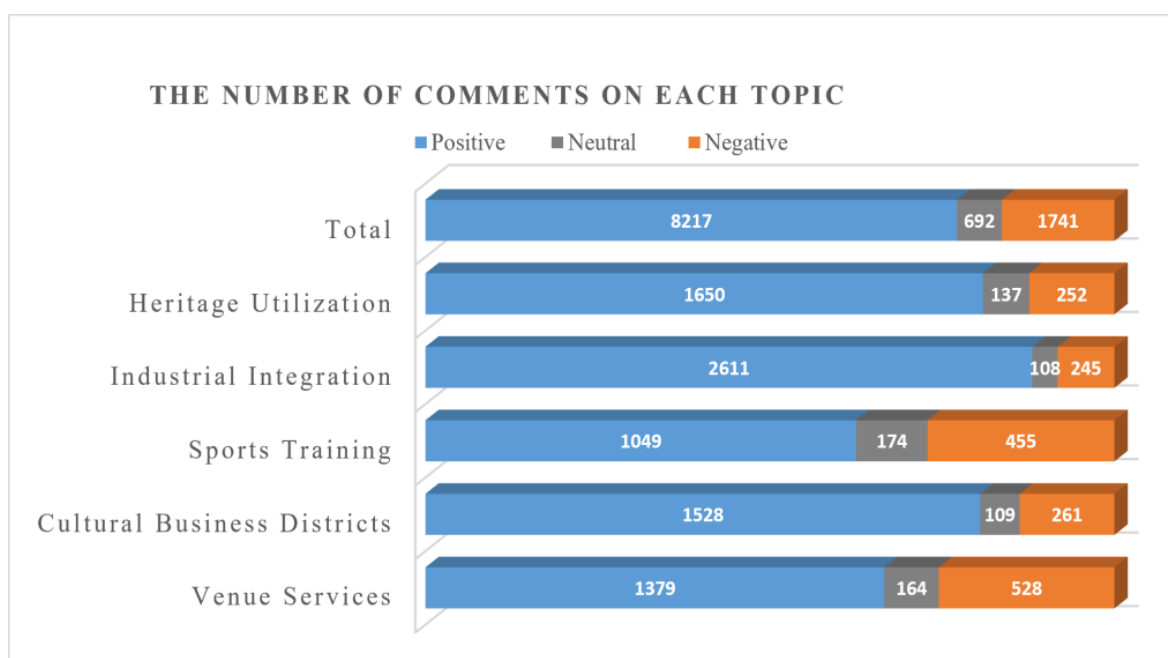


Figure 6. The number of reviews on each topic.

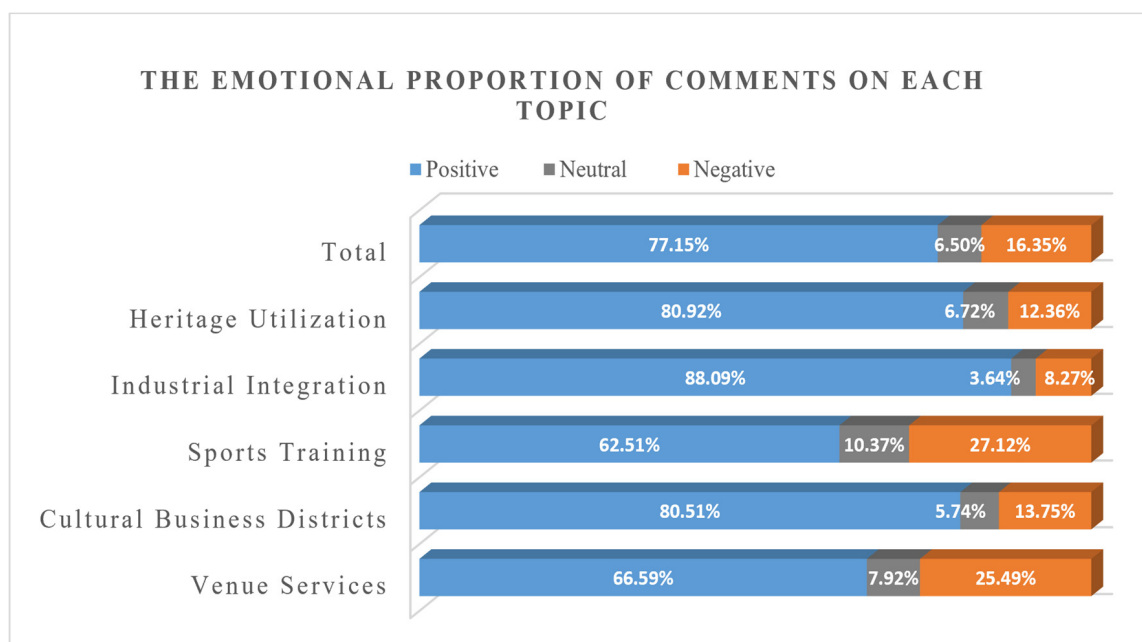


Figure 7. The emotional proportion of reviews on each topic.

In summary, the following conclusions can be drawn from the LDA model topic analysis and DistilBERT sentiment tendency analysis:

(1) Across the five different topics, consumers focus on different aspects. From the above analysis, it can be seen that for the topic of venue services, consumers pay more attention to the site and comprehensive experience, that is, hardware facilities and service quality. Venue quality directly impacts sports safety and exercise outcomes. Meanwhile, environmental conditions, pricing, and front-desk efficiency significantly influence consumer convenience and psychological satisfaction. For the topic of cultural business districts,

consumers pay more attention to cultural atmosphere and commercial amenities. Cultural atmosphere fosters consumers' emotional resonance and place identity, while surrounding commercial facilities enhance consumption convenience and leisure experiences; both are primary factors influencing consumer satisfaction. For the topic of sports training, consumers are more concerned about coaching professionalism and curriculum design. Coaching professionalism directly impacts personal safety and training outcomes, whereas curriculum rationality affects psychological engagement and learning efficiency, serving as key determinants for consumers. For the topic of industrial integration, consumers focus on industrial formats and the park environment. The richness of industrial formats dictates functional experience and participation value, while spatial design determines scene immersion and convenience, both of which are critical factors. For the topic of heritage utilization, consumers emphasize heritage characteristics and basic service facilities. The uniqueness of industrial heritage drives place identity and landscape experiences, and the improvement of basic supporting facilities ensures transportation and parking accessibility. These aspects constitute the main factors affecting consumers.

(2) Furthermore, these five dimensions interact dynamically to shape the consumers' holistic experience. Building upon the quantitative topic co-occurrence matrix established in Section 4.2.2, we further visualized the micro-level semantic relationships through a term co-occurrence network (Figure 8). In this network, nodes represent individual keywords (color-coded by their affiliated LDA topics), and edge thickness explicitly reflects the co-occurrence strength between specific terms.

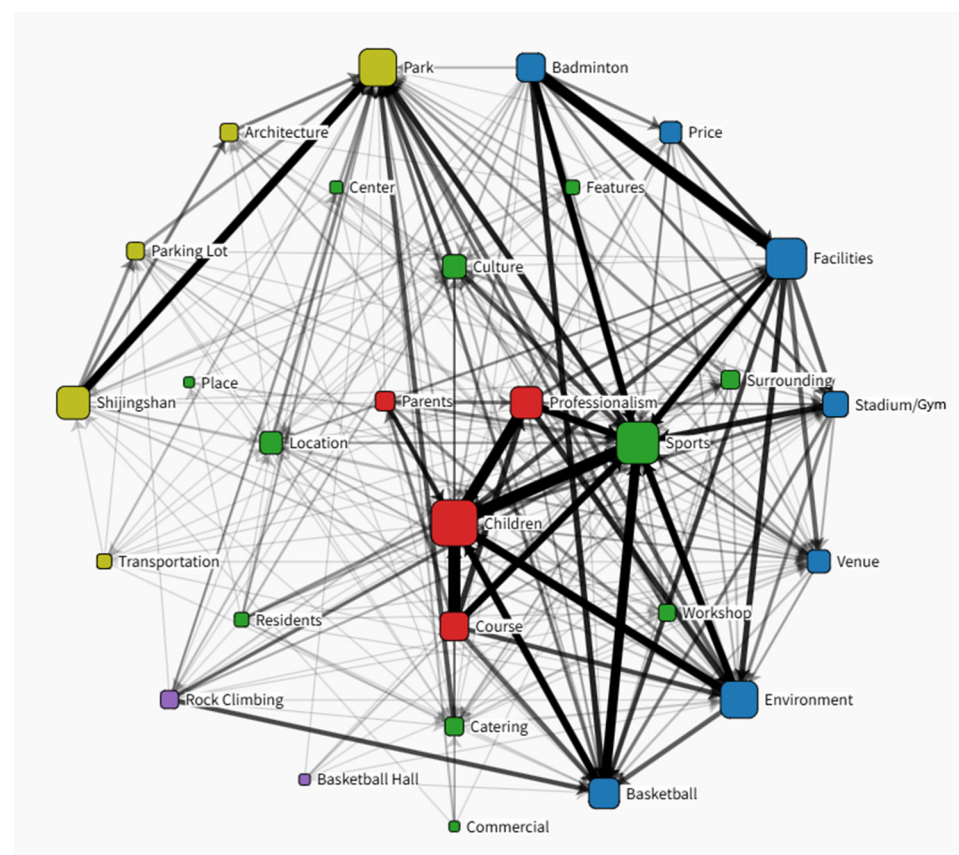


Figure 8. Semantic co-occurrence network of the top relevant terms across topics.

By examining this topological structure, clear cross-domain mechanisms emerge that explain the consumer evaluations identified above. At the micro-level, prominent thick edges directly connect user-centric terms (e.g., ‘Children’ and ‘Parents’ from Topic 3) with

infrastructural attributes (e.g., ‘Environment’ and ‘Facilities’ from Topic 1). This visualizes exactly why negative sentiments regarding basic venue services directly compromise the evaluation of sports training, corroborating the high topic-level co-occurrence (297 times) observed in Table 4. At the macro-level, the dense structural clustering among ‘Park’, ‘Culture’, and ‘Sports’ reinforces the matrix findings: consumers view preserved industrial elements not merely as isolated historical monuments, but as essential anchors driving cultural vitality. Ultimately, this integration of quantitative matrix data and semantic network visualization proves that mitigating service frictions requires a synergistic approach, addressing both physical venue constraints and cultural scene construction simultaneously.

5. Discussion

Our sentiment analysis highlights distinct variations in consumer feedback across the five extracted topics. These differences provide an empirical basis for understanding how spatial reproduction and path dependency affect the actual performance of sports service complexes.

5.1. Structural Constraints and Path Dependency

The data shows an interesting contrast: while cultural aspects received mostly positive feedback, core functions like sports training and venue services generated the highest negative review rates (27.12% and 25.49%, respectively). This disparity can be largely explained by the path dependency effect during spatial transformation. Industrial buildings are originally characterized by large spans and heavy load-bearing structures, which impose strict physical limitations. When these historical production spaces are repurposed for modern sports training, they often struggle to meet current requirements for temperature control, noise reduction, and fine-grained functional zoning. Consequently, the negative feedback regarding parental waiting areas and environmental comfort directly reflects the friction between rigid historical structures and the demands of a modern servicescape.

5.2. Spatial Reproduction and Consumer Identity

On the other hand, topics related to industrial integration and heritage utilization showed highly positive sentiments (88.09% and 80.92%). From the perspective of spatial production theory, the renovation of old factories involves more than simple physical repair; it represents the recoding of spatial meaning. By preserving industrial elements such as blast furnaces and steel trusses, these projects successfully convert obsolete production spaces into cultural consumption hubs. This transformation turns industrial memory into cultural capital, fostering a strong sense of place identity among visitors. The aesthetic appeal of these industrial symbols satisfies consumers’ psychological needs for novel experiences, which appears to compensate for some of the physical limitations mentioned earlier.

5.3. Shift in Demand and Operational Focus

Overall, these findings point to a clear shift in consumer expectations. The overlapping interests found in the venue services and cultural district topics suggest that visitors view these complexes not merely as exercise facilities, but as integrated spaces for social interaction and leisure. However, current management practices often exhibit a tendency to prioritize initial physical construction over long-term service maintenance. While the architectural transformation initially attracts foot traffic through industrial aesthetics, sustaining consumer loyalty requires a different approach. To break the constraints of path dependency, operators need to move beyond physical upgrades and systematically improve the participatory and service-oriented elements of the environment.

6. Optimization Strategy

Based on the empirical findings from topic mining and sentiment distribution, this section proposes targeted optimization strategies. These strategies aim to address the immediate operational issues identified in the consumer reviews, while also providing broader suggestions for the long-term development of these transformed industrial sites.

6.1. Sports Training Business Reconstruction

A comprehensive evaluation of the extracted topics and sentiment data reveals a severe supply-and-demand mismatch in the sports training business. Specifically, this operational dimension generated a negative sentiment proportion of 27.12%. Within these negative reviews, the high frequency of specific terms points directly to consumer dissatisfaction regarding coaching transparency, homogeneous curricula, and poor parental waiting experiences. Furthermore, the robust topic-level co-occurrence (297 instances) between Sports Training and Venue Services confirms that consumers inherently evaluate teaching quality alongside basic infrastructural comfort. To directly resolve these data-identified problems, management must implement immediate operational improvements. This includes standardizing coaching evaluations to build trust, diversifying training programs to include niche urban sports, and fundamentally transforming the parental waiting experience by integrating multi-functional zones, such as co-working spaces or cafes, to reduce waiting friction.

Moving beyond these immediate service fixes, we offer broader suggestions for long-term digital and structural upgrading. Operators should integrate wearable technology to provide quantitative performance tracking, thereby transforming subjective physical training into a premium, data-driven curriculum. Furthermore, actively fostering “Sports + Education” partnerships with local schools and community organizations can secure stable off-peak traffic, ensuring the sustainable utilization of the large-scale industrial training spaces.

6.2. Basic Service Optimization and Upgrading

Synthesizing the topic distribution and sentiment results demonstrates that basic venue services critically impact the overall consumer experience while generating substantial negative feedback. The data shows that this specific module accounts for a 25.49% negative review rate. The semantic clustering within this negative feedback highlights precise operational bottlenecks: price dissatisfaction, front-desk inefficiency during peak hours, and inadequate physical amenities. To address these empirically identified bottlenecks, operators must execute immediate rectifications. Direct recommendations include deploying self-service access control systems (e.g., smart gates) to bypass front-desk bottlenecks, implementing off-peak pricing discounts to address cost concerns, and specifically upgrading HVAC (heating, ventilation, and air conditioning) systems to resolve the unique climate control challenges inherent to high-clearance factory buildings.

To achieve a sustainable and modernized service environment, we propose several forward-looking suggestions utilizing intelligent facility management. Sports service complexes should introduce Internet of Things (IoT) systems for smart energy management, dynamically adjusting lighting and climate control based on real-time occupancy to reduce operational costs. Additionally, adopting flexible spatial partitioning—such as retractable seating and folding acoustic walls—will allow operators to seamlessly switch the venue from daily sports training modes to weekend event-hosting modes, maximizing spatial efficiency.

6.3. Integration of Cultural and Sports Business Formats

The combined evidence from the textual data robustly indicates that consumers view these transformed sites as integrated consumption ecosystems anchored by industrial aesthetics, rather than isolated exercise facilities. This conclusion is supported by the overwhelmingly positive sentiment toward Heritage Utilization (80.92%) and the fact that Cultural Business Districts and Venue Services exhibit the highest cross-topic co-occurrence (1310 instances). These quantitative metrics demonstrate a strong consumer preference for seamless functional integration. Based on these data findings, operators should prioritize the preservation of core industrial aesthetic elements. More importantly, management must implement unified membership or loyalty programs that explicitly incentivize cross-consumption between sports training, catering, and retail, preventing these interdependent modules from operating as isolated silos.

To translate this strong place identity into long-term commercial success, we outline broader recommendations for future site evolution. Operators can capitalize on the rugged industrial aesthetic to introduce trendy urban sports (e.g., rock climbing on old factory walls, parkour, or frisbee), creating a unique sports tourism destination. Moreover, utilizing outdoor peripheral spaces to develop a “Night Economy”—such as hosting sports-themed night markets, outdoor screenings, and community fitness festivals—can effectively extend operating hours and transform these industrial complexes into vibrant, 24 h urban landmarks.

7. Conclusions

The purpose of this study is to reveal the supply-and-demand dynamics in the operation of sports service complexes transformed from old industrial plants from the consumers’ perspective, providing a micro-empirical basis for the high-quality development of the industry. Based on Python web crawler technology, this study collected online consumer review data from Meituan and Dianping regarding five typical sports service complexes in China. After preprocessing, an unsupervised LDA topic model and a DistilBERT sentiment analysis model were combined for text mining. The study identified five core themes—venue services, cultural business districts, sports training, industrial integration, and heritage utilization—and quantitatively analyzed the volume and distribution of emotional tendencies under each topic. By systematically analyzing consumers’ focus, core needs, and negative feedback, this study deconstructs the structural contradictions and spatial performance bottlenecks in current sports consumption spaces.

By evaluating consumers’ multidimensional needs and emotional feedback, this study provides a clear empirical framework for understanding the structural contradictions in current sports consumption spaces. Rather than addressing macro-level urban policies or broader industry development, these findings strictly focus on the micro-level user experience, offering targeted operational strategies across venue optimization, business integration, and training service reconstruction. Ultimately, this research provides a micro-empirical, data-driven reference for facility operators to break the constraints of spatial path dependency. It highlights that the sustainable reproduction of urban stock spaces relies not merely on initial architectural preservation, but critically on continuous, service-oriented operational improvements that directly address consumers’ daily service frictions.

Despite its contributions, this study has several limitations. First, relying predominantly on online textual reviews to assess the effectiveness of spatial reproduction carries inherent constraints. User-generated content is fundamentally subjective and susceptible to self-selection bias, meaning consumers with extreme experiences (highly positive or highly negative) are more likely to post reviews, which may not represent the silent majority. Additionally, because all primary data originate from two specific Chinese online platforms

(Meituan and Dianping), there are inevitable demographic and platform biases. Demographically, the user bases of these digital platforms typically skew toward younger, urban, and more digitally active populations, potentially underrepresenting the voices of the elderly or young children who also frequent these sports service complexes. Platform-wise, since Meituan and Dianping are primarily lifestyle and consumption-oriented applications, users may inherently overemphasize commercial and service attributes (e.g., catering, parking, prices) while underreporting specific architectural or sports-technical metrics. Furthermore, review text analysis cannot fully substitute for objective physical and engineering assessments of the built environment. Second, there are limitations regarding the geographic scope of data collection and the domain-specific accuracy of model training. Future research should expand the sample coverage to more diverse industrial transformation cases and adopt mixed-method approaches. Combining objective physical space assessments (e.g., spatial flow tracking) with multimodal data fusion analysis (text, images, and spatial data) will deepen the understanding of consumer behavior and spatial perception characteristics.

Author Contributions: Conceptualization, L.J.; writing—original draft preparation, D.J. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grants from the Social Science Fund of Shaanxi Province of China (2023Q011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.32725542> (accessed on 15 June 2026).

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

References

1. Chinese Government. Outline of the Strategy for Expanding Domestic Demand (2022–2035). Available online: https://www.gov.cn/zhengce/2022-12/14/content_5732067.htm (accessed on 5 June 2025).
2. Jiang, T.; Yang, J. A Review of the “14th Five-Year Plan” and Prospect of the “15th Five-Year Plan” for China’s Sports Industry. *J. Tianjin Univ. Sport* **2025**, *40*, 132–139. [CrossRef]
3. General Office of the State Council. Opinions on Continuing to Promote the Urban Renewal Action. Available online: https://www.gov.cn/zhengce/202505/content_7023880.htm (accessed on 15 May 2025).
4. Günay, Z.; Yaşar, S.S. Industrial Heritage as a Catalyst for Urban Regeneration: The Case of Istanbul. *Cities* **2021**, *119*, 103395. [CrossRef]
5. Lefebvre, H. *The Production of Space*; Blackwell: Oxford, UK, 1991.
6. Oevermann, H.; Mieg, H.A. *Industrial Heritage Sites in Transformation: Clash of Discourses*; Routledge: London, UK, 2014.
7. Bitner, M.J. Servicescapes: The Impact of Physical Surroundings on Customers and Employees. *J. Mark.* **1992**, *56*, 57–71. [CrossRef]
8. Sorensen, A. Taking Path Dependence Seriously: An Historical Institutional Research Agenda in Planning History. *Plan. Perspect.* **2015**, *30*, 17–38. [CrossRef]
9. Yoshida, M. Consumer Experience Quality: A Review and Extension of the Sport Management Literature. *Sport Manag. Rev.* **2017**, *20*, 427–442. [CrossRef]
10. Kim, M.; Trail, G.T. A Conceptual Framework for Understanding Relationships between Sport Consumers and Sport Organizations. *J. Sport Manag.* **2011**, *25*, 57–69. [CrossRef]
11. Guo, Y.; Barnes, S.J.; Jia, Q. Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation. *Tour. Manag.* **2017**, *59*, 467–483. [CrossRef]
12. Dai, X.; Zhou, X.; Dong, Q.; Zhou, K. AI-Driven Exploration of Public Perception in Historic Districts Through Deep Learning and Large Language Models. *Buildings* **2026**, *16*, 437. [CrossRef]
13. Zou, Y.; Zhao, Q.; Wang, B.; Chen, G. Sentiment mining of online comments of sports venues: Consumer satisfaction and its influencing factors. *PLoS ONE* **2025**, *20*, e0319476. [CrossRef] [PubMed]

14. Ciani, F.; Dell’Olmo, L.; Foggi, B.; Lippi, M.M. The effect of urban green areas on pollen concentrations at ground level: A study in the city of Florence (Italy). *Urban For. Urban Green.* **2021**, *59*, 127045. [[CrossRef](#)]
15. Mishev, K.; Gjorgjevikj, A.; Vodenska, I.; Chitkushev, L.T.; Trajanov, D. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access* **2020**, *8*, 131662–131682. [[CrossRef](#)]
16. Department of Economic Affairs, General Administration of Sport. Announcement on the Selection Results of Typical Cases of Sports Service Complexes. Available online: <https://www.sport.gov.cn/n315/n20001395/c20005561/content.html> (accessed on 10 November 2025).
17. Zhou, Y.; Zhang, X.; Yu, X. Consumer Preference Analysis Based on Product Review Mining. *Inf. Sci.* **2022**, *40*, 58–65. [[CrossRef](#)]
18. Guo, Y.; Yao, X. Research on User Comments in Online Health Communities Based on Topic Mining and Sentiment Analysis. *Mod. Inf.* **2025**, *45*, 135–145. [[CrossRef](#)]
19. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
20. Xu, H.; Jiang, Y. Code Quality Attribute Judgment for Complex User Comments. *J. Softw.* **2021**, *32*, 2183–2203. [[CrossRef](#)]
21. Lucini, F.R.; Tonetto, L.M.; Fogliatto, F.S.; Anzanello, M.J. Text Mining Approach to Explore Dimensions of Airline Customer Satisfaction Using Online Customer Reviews. *J. Air Transp. Manag.* **2020**, *83*, 101760. [[CrossRef](#)]
22. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. *arXiv* **2015**, arXiv:1503.02531.
23. Li, H.; Wei, B. Zero-Shot Recognition Method with Attribute Distillation. *Comput. Eng. Appl.* **2024**, *60*, 219–227.

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.