

## KHÁM PHÁ HÀNH VI DU KHÁCH SỬ DỤNG XỬ LÝ NGÔN NGỮ TỰ NHIÊN: BẰNG CHỨNG TỪ CÁC ĐÁNH GIÁ TRÊN TRIPADVISOR

Nguyễn Phước Hoàng<sup>1\*</sup>, Lương Tiến Vinh<sup>2\*</sup>, Phạm Minh Triết<sup>3</sup>

<sup>1</sup>Trường Đại học Tây Đô, số 68 Trần Chiên, phường Lê Bình, quận Cái Răng, thành phố Cần Thơ, Việt Nam

<sup>2</sup>Trường Đại học Quang Trung, số 327 Đào Tấn, phường Nhơn Phú, thành phố Quy Nhơn, tỉnh Bình Định, Việt Nam.

<sup>3</sup>Phòng Hậu Cần, Công an tỉnh Bình Dương, số 681, đường Cách Mạng Tháng Tám, phường Chánh Nghĩa, thành phố Thủ Dầu Một, tỉnh Bình Dương, Việt Nam

\* Tác giả liên hệ: [nphoang@tdu.edu.vn](mailto:nphoang@tdu.edu.vn), [tvinh@qtu.edu.vn](mailto:tvinh@qtu.edu.vn)

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận:	13/2/2025	Nghiên cứu này sử dụng phương pháp xử lý ngôn ngữ tự nhiên (NLP) và Phân tích Ma trận Không âm (NMF) để phân tích các đánh giá trên TripAdvisor về các tour du lịch ngắn ngày tại miền Nam Việt Nam. Các chủ đề chính phản ánh trải nghiệm và động lực của du khách được xác định dựa trên lý thuyết "push and pull". Các yếu tố "push" bao gồm nhu cầu tìm kiếm kiến thức và sự tiện nghi, trong khi các yếu tố "pull" như các hoạt động độc đáo và sự hòa mình vào văn hóa địa phương có ảnh hưởng đáng kể đến hành vi của du khách. Kết quả nghiên cứu đưa ra các gợi ý hữu ích dành cho các nhà tổ chức tour, nhằm thiết kế những trải nghiệm đáp ứng cả nhu cầu tri thức và cảm xúc, góp phần nâng cao sự hài lòng của khách hàng và xây dựng mối liên kết sâu sắc hơn với điểm đến.
Ngày hoàn thiện:	5/3/2025	
Ngày chấp nhận:	11/3/2025	
Ngày đăng:	15/3/2025	
TỪ KHÓA		
Xử lý ngôn ngữ tự nhiên;		
Mô hình chủ đề;		
Trải nghiệm khách hàng;		
Ma trận không âm;		
Phân rã.		

## EXPLORING TOURIST BEHAVIOR USING NATURAL LANGUAGE PROCESSING: EVIDENCE FROM TRIPADVISOR REVIEWS

Nguyen Phuoc Hoang<sup>1\*</sup>, Luong Tien Vinh<sup>2\*</sup>, Pham Minh Triet<sup>3</sup>

<sup>1</sup>Tay Do University, No. 68 Tran Chien Street, Le Binh Ward, Cai Rang District, Can Tho City, Vietnam.

<sup>2</sup>Quang Trung University, No. 327 Dao Tan Street, Nhon Phu Ward, Quy Nhon City, Binh Dinh Province, Vietnam.

<sup>3</sup>Logistics Department, Binh Duong Provincial Police, No. 681, Cach Mang Thang Tam Street, Chanh Nghia Ward, Thu Dau Mot City, Binh Duong Province, Vietnam

\*Corresponding Author: [nphoang@tdu.edu.vn](mailto:nphoang@tdu.edu.vn), [tvinh@qtu.edu.vn](mailto:tvinh@qtu.edu.vn)

ARTICLE INFO		ABSTRACT
Received:	Feb 13 <sup>th</sup> , 2025	This study applies natural language processing (NLP) and Non-negative Matrix Factorization (NMF) to analyze TripAdvisor reviews of short-day tours in Southern Vietnam. Key topics representing traveler experiences and motivations are identified, framed within the push and pull theory. Push factors, like the desire for knowledge and convenience, and pull factors, such as unique activities and cultural immersion, influence tourist behaviors. The findings provide actionable recommendations for tour operators to create experiences that meet both intellectual and experiential needs, enhancing customer satisfaction and deepening connections with the destination.
Revised:	Mar 5 <sup>th</sup> , 2025	
Accepted:	Mar 11 <sup>st</sup> , 2025	
Published:	Mar 15 <sup>th</sup> , 2025	
KEYWORDS		
Natural language processing;		
Top Modelling;		
Customer experience;		
Non-negative Matrix		
Factorization.		

Doi: <https://doi.org/10.61591/jslhu.20.630>

Available online at: <https://js.lhu.edu.vn/index.php/lachong>

## 1. INTRODUCTION

Tourist attractions are important components for numerous fields of study about domestic and international tourism [1]. Traveler reviews are one of the noticeable resources for managers, researchers and travelers used to evaluate the quality of a tour management and understand customer decision-making [2]. In the dynamic context of the tourism industry, understanding and adapting to traveler behaviors and preferences is crucial for destinations aiming to enhance their appeal and competitiveness. The advent of digital platforms, especially review sites like TripAdvisor, has opened a treasure trove of user-generated content, offering deep insights into tourists' experiences and expectations [3]. Against this backdrop, this study employs topic modeling—a text analysis method for identifying themes within large textual datasets—to distill the essence of tourists' experiences on short-day tours in Southern Vietnam. Utilizing NLP techniques and NMF, it aims to systematically dissect TripAdvisor reviews to unearth core topics reflective of tourist experiences [4, 5]. Southern Vietnam's tourist attractions, with their rich cultural heritage, historical significance, and natural beauty, attract a global audience [6]. To maintain and increase this allure, the local tourism industry must remain attuned to the evolving desires of tourists. This necessitates an understanding of the motivations behind travel behaviors, well-captured by the push and pull theory [7, 8]. This theory distinguishes between push factors—internal motivations that compel individuals towards travel, such as the quest for knowledge, relaxation, or personal growth—and pull factors, which are the destination's attributes that draw visitors, including cultural sites, natural landscapes, and recreational opportunities [9].

In the dynamic landscape of modern consumer behavior, user-generated content (UGC) presents a potent tool for researchers to delve into traveler expectations and evolve tourism strategies [10]. UGC, encompassing online reviews, social media posts, and other digital artifacts created by consumers themselves, offers a real-time window into the dynamic world of traveler sentiment, desires, and experiences [11]. By using the power of UGC, The Southern Vietnam tours through the lens of traveler expectations and current strategies from Vietnam Adventure Tours can provide an updated understanding about consumer behaviors [6]. By combining UGC with TripAdvisor, related information can be the foundation for optimizing operational and marketing strategies in Southern tourism. Additionally, identifying thematic threads within reviews can unlock opportunities for niche marketing campaigns that resonate with specific traveler segments, fostering a more tailored and effective approach to attracting and engaging visitors [5].

The push and pull theory was introduced by Dann in 1977 to understand consumer motivation and their decisions in the travel industry [8]. The study articulated travel motivation as a profound mental disposition

prompting individuals or groups to embark on travel. This concept was clarified through the distinction between two motivational forces, termed push and pull factors [7]. From this perspective on travel motivation, push and pull factors elucidate the reasons behind individuals' decisions to travel and their selection of specific destinations over others [16, 17].

Given this framework, this research seeks to address two pivotal questions:

(1) What key topics emerge from TripAdvisor reviews of Southern Vietnam's short-day tours, and what do they reveal about travelers' experiences?

(2) How do these topics align with push and pull factors, shedding light on the motivations and satisfaction of travelers?

By answering these questions, the study aims to offer actionable insights for tour operators and the tourism sector to refine and enhance the quality of tourist experiences. This effort seeks to not only elevate the attractiveness of Southern Vietnam as a travel destination but also contribute to the sustainable growth of the tourism industry in the region.

## 2. METHODOLOGY

### 2.1 Analysis method

Topic modeling is an analytical technique that mines underlying themes from extensive text data, with various algorithms offering unique advantages. Latent Semantic Analysis (LSA) is an early technique that uses singular value decomposition to identify relationships between words and topics [12]. Probabilistic Latent Semantic Analysis (PLSA), an advancement over LSA, incorporates a probabilistic framework for topic discovery [13]. Latent Dirichlet Allocation (LDA) further refines this process by assigning probability distributions to both documents and topics [14]. Each method offers different pathways to extract thematic structures within text corpora, useful for analyzing complex datasets like customer reviews.

The study uses topic modeling, particularly the NMF algorithm, to analyze TripAdvisor reviews for its ability to efficiently distill large texts into distinct, interpretable topics by identifying themes with non-negative weights [4]. This feature makes NMF ideal for analyzing TripAdvisor reviews, allowing for accurate alignment with the push and pull framework. Through this method, the research successfully extracted 10 distinct topics that represent the core themes discussed by travelers in their reviews of Vietnam Adventure Tours [22].

Following the extraction of these topics, the study applies the push and pull theory as a theoretical framework to map and interpret the topics. This dual approach enables a nuanced understanding of the motivations behind travel decisions and overall satisfaction with the travel experiences provided by Vietnam Adventure Tours. By linking the identified topics to push factors (internal motivations to travel) and pull factors (destination's attractions that draw visitors) [7, 8], the research offers valuable insights into the dynamics of traveler behavior and destination choice [21].

Travel choices are not solely influenced by push or pull factors; rather, they interact in complex ways to shape travelers' decisions. Push factors spark the desire to travel, while pull factors determine the choice of destination [20]. Beyond guiding destination selection, the push-pull framework highlights the role of travel in fostering personal growth and social connections, with shared experiences often leading to self-discovery and a sense of well-being. Table 1 outlines key push and pull factors in tourism.[17]. Table 1 below shows some main push and pull factors in the tourism industry.

**Table 1.** Experiment results on each Algorithms Clustering

Factors	Tourism choice of destination
<b>Push</b>	Cultural values; Convenience of facilities; Family togetherness; Knowledge; Interest; Relaxation
<b>Pull</b>	Safety; Budget; Leisure; Activity; Nature; Religion

## 2.2 Dataset Description

This research conducted a comprehensive evaluation of 17,337 TripAdvisor reviews concerning Vietnam Adventure Tours, collected in November 2023. The extraction process utilized the powerful programming language Python, specifically employing Selenium and BeautifulSoup for web scraping, alongside relevant libraries [21]. These tools were crucial for collecting and parsing extensive web data into a dataset rich in traveler feedback, laying the groundwork for our analysis of customer experiences on short-day tours in Southern Vietnam. The comprehensive range of reviews forms a solid foundation for using advanced NLP techniques to reveal insights into traveler sentiments and preferences, as detailed in Fig. 1.



**Figure 1.** The process of analyzing data

The dataset includes TripAdvisor review attributes such as reviewer IDs ('NameReviewer'), origins ('ReviewerCountry'), travel types ('TypeTravellers'), likes ('Like'), review dates ('date\_review'), detailed reviews ('FullReview'), ratings ('Rating'), sentiments ('Sentiment'), and review lengths ('ReviewLength'), as shown in Table 2. It focuses on 'FullReview' for topic modeling, offering a detailed insight into customer feedback. Some popular words that appear in the travelers' feedback are shown in Fig. 3.



**Figure 2.** The word cloud from tourists' reviews

## 3. FINDINGS AND DISCUSSION

### 3.1 Topic Modeling with NMF

By using NMF analysis, the present study reveals a spectrum of latent topics. These 10 distinct categories encompass elements of visitor experience on short trips organized by Vietnam Adventure Tours. These themes highlight the diverse aspects of visitor enjoyment, revealing the key factors contributing to positive sentiment and satisfaction.

A detailed examination of the themes drawn upon the insights gleaned from visitor reviews is presented below.

**Table 2.** Top 10 visitor reviews:

Topic	RF
1	tour, recommend, highly, guide, informative, amazing, knowledgeable, definitely, friendly, funny
2	war, jackie, vietnam, tuan, experience, story, veteran, tour, guide, personal
3	mekong, delta, day, boat, local, river, lunch, ride, coconut, island
4	good, english, nice, guide, ele, food, tour, time, friendly, spoke
5	great, day, time, experience, sense, guide, humour, thanks, knowledge
6	trip, guide, enjoyable, ele, amazing, day, thank, excellent, fun, friendly
7	tunnel, chi, history, half, visit, chu, went, visiting, interesting, bus
8	lan, unicorn, highly, recommended, recommend, funny, experience, guy, knowledgeable, guide
9	really, enjoyed, nice, funny, interesting, interesting, interesting, tour, john, guide, helpful, explained
10	lot, information, interesting, history, fun, vietnam, learned, gave, culture, tour

The extensive analysis of TripAdvisor reviews has yielded critical insights into the travelers' experiences using Vietnam Adventure Tours. By applying NMF topic modeling to the review content, this study not only uncovers the core themes within traveler feedback but also provides definitive answers to the research questions initially posed. The extracted topics from the reviews shed light on the multifaceted nature of travelers' experiences [23]. The findings illustrate that travelers' discussions revolve around a desire for enriching experiences (topics #1, #8, #9), comfort and reliability of facilities (topics #2, #4, #7), and engaging activities (topics #3, #5, #6). Based on the exploration, the suggestions are to increase the activities related to Vietnamese culture and history during *sau từ này còn thông tin về thời gian bị bỏ sót?* Travelers can participate in making traditional food such as steamed cassava with coconut milk and understand the reasons why Vietnamese people have this food in this area. In addition, understanding Southeast history through hearing about the war and unique activities in specific locations such as Cu Chi and Ben Tre. These activities can strengthen knowledge and the intellectual and emotional engagement with the destination.

The alignment of these topics with push and pull factors illuminates the motivations behind travelers' decisions and their satisfaction levels [24]. The intrinsic push factors, such as knowledge acquisition and quality facilities, drive travelers toward seeking out these tours. Simultaneously, the extrinsic pull factors, like unique activities and the opportunity for cultural immersion, play a crucial role in destination choice, confirming the theoretical framework's applicability in the digital era of travel reviews.

### 3.2 Mapping Non-negative Matrix Factorization topics to push & pull theory

Through this innovative approach, ten topics are grouped into a push and pull framework. These factors reveal some main reasons that influence tourist decision-making and travelers' satisfaction.

(1) Push Factors – Knowledge: Among the identified topics, topics #1, #8, #9 highlight the inherent human desire for intellectual enrichment and cultural understanding. These topics resonate with tourists seeking to delve into the rich tapestry of Vietnamese history and customs, suggesting a strong intrinsic motivator beyond mere sightseeing.

(2) Push Factors – Facilities: Topics #2, #4, #7 focusing on the importance of tangible conveniences. These topics reveal the critical role that reliable transportation, comfortable accommodation, and well-maintained facilities play in shaping a positive travel experience, emphasizing the role of perceived risk reduction in influencing travel decisions. Pull Factors – Activities: Topics #3, #5, #6 show a preference for enriching activities that promote cultural immersion and personal growth, reflecting a shift towards authentic, educational travel experiences in Vietnam. Push-pull analysis

underscores a demand for knowledge and high-quality experiences, as highlighted by the influence of informed guides (Fig. 4) and superior facilities (Fig. 5) on tourist satisfaction. These points stress the value of focusing on educational content and excellence in operations to meet and exceed short-day tour expectations, as shown in Fig. 6.

Oct 2023 • Friends  
We had the best time with Duc Phan. Duc is so friendly so professional so on time with great sense of humor too. He had a cut during coconut island tour while demonstrating how to cut open a coconut but he just kept going, that's some professionalism. He went above and beyond for our mates who had different needs as well. Overall it was an excellent experience, definitely book with Duc. Thanks again Duc for a great time!

**Figure 3.** A traveler's review about professional attitude of a tour guide [26]

Mar 2023 • Friends  
we spent a whole day with a this company and our gorgeous tour guide Van. the bus was very comfortable and air condition was our salvation during hot weather. Our guide was very nice, positive and active. the trip was very balanced and Van told us many interesting stories. the river trip, folk music and traditional, special snake wine were the most amazing 🍷

**Figure 4.** A traveler's review about the facilities of the tour [26]

Mar 2023 • Friends  
we spent a whole day with a this company and our gorgeous tour guide Van. the bus was very comfortable and air condition was our salvation during hot weather. Our guide was very nice, positive and active. the trip was very balanced and Van told us many interesting stories. the river trip, folk music and traditional, special snake wine were the most amazing 🍷

**Figure 5.** A traveler's review their positive experience about the knowledge they have from a tour [26]

Beyond the push factors, the research illuminates the potent pull factors that draw tourists to heritage experiences. The analysis reveals a strong desire for meaningful activities directly related to Vietnamese culture and history (#3, #5, #6). These topics showcase the growing preference for experiential tourism, where travelers seek to engage in activities that offer deeper understanding and a transformative connection with the local culture (Fig. 7). This trend creates great potential for tourism stakeholders to construct unique and real experiences that connect with today's sensitive visitors.

Nov 2023 • Friends  
Our amazing guide (Van) and bus driver picked us up in a comfortable, clean bus at 7:40am. From there we spent the day exploring a famous pagoda, explored many islands - Dragon, tortoise, phoenix and unicorn - as well as local music, food and boat. Lunch provided was of excellent quality. we were then able to ride around the island on bicycles, before visiting yet another island where we learnt how coconut candy is produced. The strong stomachache had the opportunity to try snake wine and banana wine. All in all an excellent full day tour - great highlights, terrific activities and excellent service. A definite 5 stars and would highly recommend for anyone visiting the region - irrespective of the length of their stay. Thank you Van and co.!

**Figure 6.** A traveler's review of their positive experience about the knowledge gained from a tour [26]

The application of push and pull factors in understanding tourist behavior presents a nuanced framework that enriches the theoretical discourse on travel motivation [24]. The study's findings affirm the relevance of Dann's push and pull theory within the context of modern tourism, particularly in the realm of digital consumer feedback. By demonstrating how distinct topics are extracted from online reviews map onto push and pull factors, this research contributes to a deeper understanding of the dynamic interplay between a traveler's internal desires and the external attributes of a destination [25]. It underscores the theory's adaptability in deciphering the complex motivations behind why people travel and how they select destinations, enriching the academic dialogue on tourism studies.

## 4. CONCLUSION

This study on Vietnam Adventure Tours, while insightful, is limited by the length of the research. Therefore, its reliance on TripAdvisor reviews and NLP techniques potentially misses the full spectrum of tourist experiences and subtler insights. It also overlooks broader aspects of tourist behavior, such as decision-making and satisfaction. Future research should expand using other techniques such as LDA and LSA and employ advanced analytical methods for a more thorough understanding of Southern Vietnam tourism, improving destination management.

## 5. REFERENCES

- [1] J. A. S. Fernández, et al., "Determinants of tourism destination competitiveness in the countries most visited by international tourists: Proposal of a synthetic index," *Tourism Management Perspectives*, 33, pp. 100582, **2020**.
- [2] A. Mardani, et al., "Application of multiple-criteria decision-making techniques and approaches to evaluating of service quality: a systematic review of the literature," *Journal of Business Economics and Management*, 16(5), pp. 1034-1068, **2015**.
- [3] V. Nelson, "Liminality and difficult heritage in tourism," in *Liminality in Tourism*: Routledge, **2021**, pp. 80-100.
- [4] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, 13, **2000**.
- [5] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short- text data: A comparative analysis," *Frontiers in artificial intelligence*, 3, pp. 42, **2020**.
- [6] N. H. Tien, N. M. Ngoc, and D. B. H. Anh, "Sustainability of coastal tourism development: comparative analysis of Vietnam's Northern and Southern provinces," *Journal of southwest jiaotong university*, 55(6), **2020**.
- [7] G. M. Dann, "Tourist motivation an appraisal," *Annals of tourism research*, 8(2), pp. 187-219, **1981**.
- [8] G. M. Dann, "Anomie, ego-enhancement and tourism," *Annals of tourism research*, 4(4), pp. 184-194, **1977**.
- [9] D. Getz and S. J. Page, "Progress and prospects for event tourism research," *Tourism management*, 52, pp. 593-631, **2016**.
- [10] B. Garner and D. Kim, "Analyzing user-generated content to improve customer satisfaction at local wine tourism destinations: an analysis of Yelp and TripAdvisor reviews," *Consumer behavior in tourism and hospitality*, 17(4), pp. 413-435, **2022**.
- [11] R. K. Mishra, et al., "Knowledge based topic retrieval for recommendations and tourism promotions," *International Journal of Information Management Data Insights*, 3(1), pp. 100145, **2023**.
- [12] S. Deerwester, et al., "Indexing by latent semantic analysis," *Journal of the American society for information science*, 41(6), pp. 391-407, **1990**.
- [13] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, **1999**, pp. 50-57.
- [14] D. M. Blei, et al., "Latent dirichlet allocation," *Journal of machine Learning research*, 3, pp. 993-1022, **2003**.
- [15] K. Nithin Goud, et al., "User preference multi-criteria recommendations using neural collaborative filtering methods," in *Proceedings of the Sixth International Conference on Mathematics and Computing: ICMC 2020*, 2021: Springer, pp. 55-64.
- [16] A. D. Bright, "Motivations, attitudes, and beliefs," in *Handbook of hospitality marketing management*: Routledge, **2009**, pp. 239-265.
- [17] A. Correia, P. O. do Valle, and C. Moço, "Modeling motivations and perceptions of Portuguese tourists," *Journal of Business Research*, 60(1), pp. 76-80, **2007**.
- [18] J. L. Crompton and S. L. McKay, "Motives of visitors attending festival events," *Annals of tourism research*, 24(2), pp. 425-439, **1997**.
- [19] G. Erol, A. Deliormanli, and E. Düşmezkalender, "A tourism accommodation perspective in a post-pandemic nature-based tourism development context," in *The Routledge Handbook of Nature Based Tourism Development*: Routledge, **2023**, pp. 474-489.
- [20] I. Cifci, O. Atsiz, O. C. Kahraman, and F. Istanbulu Dincer, "Recreational fishers' motivations at wetland destinations: The push and pull theory approach," *Leisure Sciences*, pp. 1-23, **2021**.
- [21] S. Han and C. K. Anderson, "Web scraping for hospitality research: Overview, opportunities, and implications," *Cornell Hospitality Quarterly*, 62, (1), pp. 89- 104, **2021**.
- [22] S. Singh, T. Chauhan, V. Wahi, and P. Meel, "Mining tourists' opinions on popular Indian tourism hotspots using sentiment analysis and topic modeling," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, **2021**: IEEE, pp. 1306-1313.
- [23] M. J. Sánchez-Franco and S. Rey-Tienda, "The role of user-generated content in tourism decision-making: an exemplary study of Andalusia, Spain," *Management Decision*, **2023**.
- [24] R. P. Sastre and P. Phakdee-Auksorn, "Examining tourists' push and pull travel motivations and behavioral intentions: The case of British outbound tourists to Phuket, Thailand," *Journal of Quality Assurance in Hospitality & Tourism*, 18(4), pp. 437-464, **2017**.
- [25] M. Carvache-Franco, M. M. Solis-Radilla, W. Carvache-Franco, and O. Carvache- Franco, "Push and pull motivations as demand predictors of coastal and marine destinations: a study in Acapulco, Mexico," *Tourism Planning & Development*, 20(6), pp. 1000-1018, **2023**.
- [25] TripAdvisor (**2023**). Vietnam Adventure Tours.