서울시 외국인 관광객의 주요 관광지 방문 경험 분석
- Tripadvisor 리뷰 데이터와 BERTopic 모형의 적용

Analyzing Foreign Tourists Experiences at Major Attractions in Seoul; Korea
- Application of BERTopic Model Using Tripadvisor Reviews Data

김나연* · 김혜빈** · 이수기***
*한양대학교 도시공학과 석사과정 / **한양대학교 도시공학과 석사과정 / ***한양대학교 도시공학과 교수
Kim, Nayeon* · Kim, Hyebin** · Lee, Sugie***

●●● 국문요약

본 연구는 서울시 외국인 관광객의 주요 관광지 방문 경험을 분석하기 위해 Tripadvisor 리뷰 데이터를 활용하고, BERTopic 모델을 적용하여 주요 관광 경험의 특징을 도출하였다. 연구 대상은 서울시의 주요 관광지의 영어 리뷰이며, 시계열 분석을 통해 관광객의 관심사 변화를 살펴보았다. 분석 결과, 외국인 관광객의 경험은 역사·문화적 요소, 혼잡함에 대한 경험, 자연경관, 쇼핑과 체험 활동 등의 다양한 요인에 영향을 받는 것으로 나타났다. 또한 COVID19 팬데믹 전후로 관광객의 관심사의 변화가 있었다. 쇼핑 중심의 관광 활동에서 새로운 관광 콘텐츠 및 문화 체험 요소에 대한 관심이 증가하는 경향이 나타났다. 또한, 내국인 리뷰에서 도출한 토픽과 비교하여, 관광 경험에서 중요시 여기는 요소의 차이가 있음을 확인하였다. 음식, 쇼핑, 역사적 장소 등이 내외국인의 공통적인 관심사로 나타났지만, 내국인 관광객은 일상 속 여가 활동과 정서적 경험에 대한 관심이 더 높게 나타났다. 본 연구는 온라인 리뷰 데이터의 활용 가능성을 제시하고, 서울시 관광 정책 수립 및 관광지 개선 방향 설정에 유용한 기초 자료를 제공한다는 점에서 학술적·실무적 의의를 가진다.

●●● Abstract

This study analyzes the visiting experiences of foreign tourists in Seoul's major tourist attractions using Tripadvisor review data and applies the BERTopic model to identify key characteristics of their tourism experiences. The study focuses on English-language reviews of major tourist sites in Seoul and examines changes in tourists' interests through a time-series analysis. The results reveal that foreign tourists' experiences are influenced by various factors, including historical and cultural elements, perceptions of congestion, natural landscapes, shopping, and experiential activities. Additionally, there were notable shifts in tourist interests before and after the COVID-19 pandemic, with an increasing preference for new tourism content and cultural experiences over shopping-centric activities. A comparison with topics extracted from domestic tourist reviews further highlights differences in key tourism experiences. While food, shopping, and historical sites emerged as common interests among both domestic and foreign tourists, domestic tourists showed a greater emphasis on leisure activities and emotional experiences in their everyday lives. This study underscores the potential of online review data for tourism research and provides valuable insights for the development of tourism policies and the improvement of tourist destinations in Seoul.

주제어 : 관광 경험, 외국인 관광객, 리뷰 데이터, 토픽모델링, BERTopic
Keywords : Tourism Experience, Foreign Tourists, Review Data, Topic Modeling, BERTopic

Corresponding Author : Sugie Lee, Department of Urban Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763l, Korea. Tel: +82-2-2220-4545, E-mail: sugielee@hanyang.ac.kr

https://doi.org/10.38195/judik.2025.02.26.1.51

1. 서론

1.1. 연구의 배경

관광지 경쟁력은 도시의 경제적 목표 달성에 중요한 역할을 한다(Crouch & Ritchie, 1999). 외국인 관광객은 외화 수입 증가와 고용 창출을 통해 경제 성장을 촉진한다(Lee & Chang, 2008). 2023년 문화체육관광부의 보고서에 따르면, 한국 방문 외국인 입국자는 COVID-19로 인한 충격 이전 수준으로 점차 회복 중이며, 전년 대비 245% 증가한 1,103만 명이다. 이 중 80.3%가 서울을 방문했는데, 서울은 다양한 관광 인프라와 많은 관광객을 수용할 수 있는 능력을 갖추고 있으며, 문화·역사적 가치를 바탕으로 관광지 경쟁력을 갖추고 있다.

사용자 생성 콘텐츠(User Generated Content: UGC)는 관광객의 여행 결정에 핵심적인 요인이다. 정보통신기술의 발전으로 사용자 생성 콘텐츠가 증가하고 있으며, 이를 통해 생성된 빅데이터는 개인의 디지털 흔적을 수집할 수 있다. 개인은 SNS, 웹사이트, 블로그 등을 통해 여행 경험을 공유하고 서비스 품질을 평가하며, 이러한 사용자 생성 콘텐츠는 관광산업에서 중요한 자료로 활용된다(Lu & Stepchenkova, 2015; Cenni, 2024). Tripadvisor, Booking.com과 같은 글로벌 여행 정보 웹사이트에 작성된 리뷰는 온라인 입소문(Electronic Word Of Mouth: eWOM)을 확산시키는 사용자 생성 콘텐츠로서, 온라인에서의 큰 평가 영향력을 발휘하며, 이는 평판 관리 측면에서 관련 이해관계자의 운영에 중요한 역할을 한다(Baka, 2016). 특히, 대다수의 관광객들은 여행 결정을 내릴 때 다른 사람의 후기를 참고하며 이를 신뢰할 수 있는 정보 소스로 여긴다(Lu & Stepchenkova, 2015; Nilashi et al, 2018).

이처럼 사용자 생성 콘텐츠는 관광객이 직접 경험한 인식과 감정을 반영하며, 관광지 이미지 형성에 중요한 역할을 한다. 관광지 이미지(Tourism Destination Image, TDI)는 목적지에 대해 방문자가 접하는 정보를 바탕으로 형성되는 인식과 반응의 집합이다(Crompton, 1979; Lai & Li, 2016). 이는 사람들의 목적지 방문 행동에 영향을 미치며, 도시 관광지의 홍보 및 마케팅 핵심 요소로 작동할 수 있다(Echtner & Ritchie, 1993). 사람들이 관광지에서 인상 깊게 느낀 요소와 그곳에서 이루어진 활동을 기반으로 형성된 관광 경험은, 관광객의 시각에서 관광지를 어떻게 인식하고 경험하는지를 보여준다.

따라서, 글로벌 여행 정보 웹사이트에서 작성되는 리뷰 데이터를 통해 서울시 관광지에서 나타나는 관광 경험을 분석하는 것은 관광지 마케팅과 운영에서 효과적인 관광지 이미지 전달에 중요한 정보를 제공할 수 있다. 관광지 리뷰 데이터는 사용자의 의견 분석과 고객 만족도 평가에 중요한 정보를 제공하며, 명소에 대한 평가, 리뷰, 사진 등 정량적, 정성적 피드백을 확인할 수 있다(O'connor, 2010). 그러나 국내 관광지 관련 연구에서는 주로 설문조사를 중심으로 관광객의 경험을 분석하였으며, 데이터 수집의 범위가 제한적이고 특정 시점에 국한된다는 한계를 가진다.

Tripadvisor는 전 세계 관광객을 위한 대규모 온라인 플랫폼으로, 2018년 말까지 4억 9천만 명 이상의 월평균 방문자와, 7억 3천만 건 이상의 관광지 리뷰 데이터를 기록한 바 있다(Taecharungro & Mathayomchan, 2019). 이용자는 호텔, 즐길 거리, 레스토랑 등의 정보를 탐색할 수 있으며, 관광객 리뷰, 평점, 관광지 인기도에 따라 관광지 추천 순위를 제공한다. 이용자는 자발적으로 100자 이상의 리뷰 텍스트를 기록하고, 제목, 평점, 방문일, 동행인 정보, 사진 등을 게시한다. 한편, Xiang(2017)는 Tripadvisor와 Expedia, Yelp와 같은 주요 온라인 리뷰 플랫폼의 호텔 리뷰를 비교하면서, Tripadvisor 리뷰가 더 높은 유용성을 보였다고 밝혔다. 또한, Ruytenbeek et al.(2021)은 Booking.com과 Tripadvisor의 부정적인 호텔 리뷰를 비교한 결과, Tripadvisor 리뷰가

더 길고 구체적이며 명시적인 경향을 보였다.

이처럼 Tripadvisor의 리뷰는 다른 플랫폼보다 구체적인 내용을 담고 있어 더 유용하다는 평가를 받는다. Tripadvisor 관광지 리뷰가 다른 사람들의 여행 의사결정에 영향을 미친다는 점을 바탕으로, 관광객의 경험을 분석하는 연구가 진행되고 있다. 리뷰 데이터를 활용해 관광지 이미지, 네트워크 분석, 감성 분석 등이 진행되었으며, 이와 같은 연구는 관광객의 경험을 이해하고 실무에 적용하는 데 중요하게 기여한다(Nilashi et al, 2018; Taecharungroj & Mathayomchan, 2019; Hong, 2020; Cenni, 2024). 한편, 이러한 온라인 플랫폼의 리뷰를 바탕으로 관광객이 느끼는 인식과 감정뿐만 아니라, 관광지 유형별로 형성되는 관광지 이미지를 도출하는 연구는 부족한 상황이다.

1.2. 연구의 목적

최근 관광 분야에서 빅데이터를 활용한 관광객의 경험 분석이 중요해지고 있으며, 특히 글로벌 여행 정보 웹사이트의 리뷰 데이터는 관광객의 실제 방문 목적과 경험을 상세히 반영한다. 리뷰 데이터에는 관광객이 경험한 다양한 측면이 담겨져 있다. 전시, 문화, 쇼핑, 인적 서비스, 음식, 역사 등에 대한 인지적 특성과 경관에 대한 감정, 관광 후 기억, 장소의 전통과 활력, 그리고 방문 경험에서 느낀 감정 등 정서적 특성이 나타난다.

외국인 관광객이 서울시 도시 관광지를 방문하면서 작성한 리뷰 데이터는 관광 목적과 경험을 텍스트로 생생하게 기록한다. 이러한 리뷰는 다른 잠재적 관광객에게 영향을 미칠 수 있기 때문에, 리뷰에 담긴 경험을 도출하는 것은 관광지 운영에 매우 중요하다. COVID-19 이후 국내를 방문하는 관광객이 다시 증가하는 현 상황에서, 리뷰 데이터 분석을 통해 관광지별 주요 토픽을 파악하고 지역사회의 관광 마케팅 전략 수립 및 관광지 경쟁력 강화를 위한 기초 자료를 제공하고자 한다.

본 연구는 서울시를 대상으로 2012년부터 2023년까지 Tripadvisor의 리뷰 데이터를 수집하여 국내 관광지를 방문한 외국인의 관광 경험을 분석한다. BERTopic 토픽 모델링 모형을 적용하여 국내 관광지의 주요 관심사의 시계열적 변화와 관광지 유형별 토픽을 확인하고, 관광지에 대한 이미지 특성을 구조화하여 제시한다. 더불어 외국인과 내국인 관광객의 리뷰를 비교 분석함으로써 관광 경험의 차이점을 살펴본다. 본 연구의 결과는 장기적인 관점에서 관광객의 주요 관심사 변화를 살피고, 지역사회의 관광 관련 정책 수립에 기여하고자 한다.

2. 선행연구 검토

2.1. 관광 경험 분석

관광 경험은 주관적인 특성으로 인해 다양한 동기로 설명될 수 있으며(Uriely, 2005), 기억에 남는 관광 경험(Memorable Tourism Experiences: MTE)은 관광지의 고객 유치와 충성도에 결정적인 역할을 한다(Kandampully et al., 2011; Kim et al., 2012). Chen & Chen(2010)은 서비스 품질이 관광객 만족도와 행동 의도에 큰 영향을 미친다고 강조했으며, 만족스러운 관광 경험은 잠재 고객 증가로 이어진다(전형규 외, 2010). 이러한 경험이 축적됨에 따라 관광지에 대한 이미지가 형성될 수 있는데, 관광지 이미지는 관광객이 접하는 다양한 정보와 인상의 모음으로 정의된다(Crompton, 1979). 이는 개인의 지각과 미래 관광 선택에 영향을 미친다(Echtner & Ritchie, 1991; Qu et al., 2011). 최근에는 정보통신기술의 발달로 온라인상에서 형성되는 정보를 계속 접할 수 있으며, 실제 관광지를 방문하지 않아도 관광지에 대한 이미지가 형성될 수 있다.

Baloglu & McCleary(1999)는 다른 관광 목적지와 차별화할 수 있는 고유 이미지의 중요성을 강조했으나, 이를 효과적으로 도출하는 방법에는 기존의 전통적인 방법론으로는 한계가 있었다.

기존 연구들은 주로 설문조사와 인터뷰를 통해 관광지에 대한 인식과 경험을 분석해 왔다(곽은진 · 김이태, 2017; 전정연 · 전현모, 2022). 윤유식 외(2017)는 관광 경험이 관광지 이미지와 재방문 의도에 미치는 영향을 연구했으며, 정규엽 · 김한선(2009)은 외래관광객을 대상으로 한 설문을 통해 서울의 도시 이미지가 관광객 만족도에 미치는 영향을 조사했다. 그러나 이러한 연구 방법은 표본의 제한성과 설문 구성 범위 내에서만 관광 경험 구성 요소를 도출할 수 있다는 한계가 있다.

2.2. 사용자 생성 콘텐츠의 활용 연구

사용자 생성 콘텐츠는 관광 연구에서 핵심 자료로 활용된다. 온라인에서 공유되는 개인의 경험과 불만 사항, 후기 등의 정보는 관광지의 서비스 품질이나 도시 이미지, 평판 등에 영향을 미친다(Lu & Stepchenkova, 2015). 빅데이터로서의 사용자 생성 콘텐츠는 관광객 만족도 측정 및 수요 예측 등의 분석에 활용될 수 있으며(Baka, 2016), 시계열 분석, 계량 경제학, 인공지능, 감성 분석 등 다양한 방법론을 통해 미래 관광 수요 예측에 중요한 역할을 한다(Alaei et al., 2019; Li et al., 2021).

김나연 · 강영옥(2019)은 소셜미디어에 게시된 관광지 사진의 지오태깅 데이터를 활용해 주요 관광지 군집을 도출하고 관광 패턴을 분석했으며, 장윤정 · 정수아(2021)는 Tripadvisor 리뷰를 통해 관광 목적지 방문 특성과 만족도를 국적별로 비교 분석하였다. 또한, 네트워크 분석을 활용해 관광지 간 영향력을 파악하고, 적절한 여행 경로를 추천하는 연구도 진행되고 있다(박득희 외, 2015; Park & Liu, 2022). 이러한 사용자 생성 콘텐츠는 여행 결정에 영향을 미치며, 특정 주제를 포함한 게시글을 키워드 기반으로 수집하여 실제 관광객이 느끼는 도시의 편익을 분석하는 데 활용될 수 있다(곽은진 · 김이태, 2017; 박경열 · 안희자, 2019). 그러나 국내 연구에서는 외국인 관광객의 사용자 생성 콘텐츠 분석이 특정 관광지에 제한되거나 분석의 범위가 좁아, 보다 다양한 관광 경험을 반영할 필요가 있다(이현주, 2017; 신주하 외, 2021; 윤희정, 2024).

2.3. 텍스트 마이닝을 활용한 관광 경험 분석

온라인 빅데이터의 가용성이 높아짐에 따라 텍스트 마이닝을 활용한 관광 경험 분석 연구가 활발히 진행되고 있다. 그중 토픽 모델링은 문서 집합의 추상적 주제를 도출하는 통계적 분석 기법으로, 텍스트 데이터의 숨겨진 구조를 발견하는 데 효과적이다(Koltringer & Dickinger, 2015; 한지연 · 김홍범, 2017; Wang & Hao, 2018).

텍스트 마이닝 분석은 지속적으로 발전해왔다. TF-IDF 분석은 단어의 출현 빈도를 바탕으로 문서 간 유사도를 측정하는 데 유용하지만, 단어의 의미적 관계를 고려하기 어렵다는 한계가 존재했다. Latent Semantic Analysis(LSA)와 Latent Dirichlet Allocation(LDA) 등의 방법이 도입되며 단어의 잠재적인 의미를 파악하고자 하였다. 최근에는 BERT 임베딩 기반 BERTopic이 등장했다. BERTopic은 클래스 기반 TF-IDF(cTF-IDF)를 활용해 최적의 토픽 개수를 자동 결정하고, 보다 조밀한 클러스터링을 제공할 수 있다(Grootendorst, 2022). Kirilenko & Stepchenkova(2025)는 LDA, Chat GPT 등과 비교 연구를 진행한 결과, BERTopic은 대규모 데이터에서 노이즈에 강한 장점을 가진다고 분석했다. 또한, 긴 리뷰의 경우 문장 단위로 나누어 다양한 토픽을 효과적으로 도출하는 데 뛰어난 성능을 보였다.

관광지에 대한 인식은 공식 플랫폼, 여행 정보 웹사이트, 개인 블로그 등 다양한 데이

터 소스를 활용해 분석할 수 있다(Choi et al., 2007; Liu et al., 2021; Luo et al., 2023). 심영석 · 김홍범(2016)과 장민희 · 김남조(2022)는 수집한 텍스트 데이터를 통해 관광지 이미지를 형성하는 핵심 키워드를 도출하였다. 이러한 연구들은 관광지 이미지가 관광 매력 속성에도 영향을 미친다는 점을 확인하였다. Liu et al(2021)은 이용자 그룹별로 관광지 인식의 차이를 분석해, 관광지가 제공하고자 하는 이미지와 실제 관광객 인식 간의 차이를 파악하였다. 이러한 분석은 향후 관광지 이미지의 효과적인 관리와 경쟁력 강화를 위한 중요한 시사점을 제시한다.

국내 연구에서는 Tripadvisor 리뷰 데이터를 적극적으로 활용하며 핵심 관광 경험 요소를 측정하고 있다. 박홍제 외(2021)는 Tripadvisor에 게시된 서울시 관광지 리뷰 데이터를 활용하여 감성 점수를 산출하고, 외국인의 생활 이동 데이터와 결합하여 외국인 관광객 수에 영향을 미치는 요인을 분석하였다. 한지연 · 김홍범(2017)은 Tripadvisor 리뷰 데이터를 이용해 네트워크 중심성 분석을 수행함으로써 서울시 관광지 이미지의 핵심 요소를 인지적, 정서적 차원에서 측정해 긍정적인 이미지 형성에 기여하는 주요 요소를 도출하였다. 이처럼 직접적인 관광객 경험을 반영하는 내제적 차원에서의 접근 필요성이 제기되고 있다. 또한, 관광지 유형별로 경험의 인식 및 느끼는 감정의 차이를 도출할 필요가 있다.

2.4. 연구의 차별성

국내외 관광 연구에서는 설문조사, 인터뷰, 온라인 리뷰 데이터, POI 정보, 위치 데이터 등을 활용하여 관광객의 만족도 요인을 도출해 왔으나, 대부분 특정 관광지에 한정되거나 분석 범위가 제한적인 경우가 많았다. 이에 본 연구는 서울을 방문한 외국인 관광객의 관심사가 시간에 따라 어떻게 변화하는지를 시계열적으로 분석하고자 한다. 이를 위해 2012년부터 2023년까지 Tripadvisor에 작성된 서울 관광지 리뷰를 활용하여, 10년 이상의 데이터를 종합적으로 분석함으로써 관광객 관심사의 변화를 보다 체계적으로 파악한다.

또한, 온라인상에서 생성된 사용자 생성 콘텐츠를 활용하여 관광객이 경험한 관광지에 대한 인식을 내제적 차원에서 도출한다. 기존의 설문조사 기반 연구는 미리 제시된 선택지 내에서 응답이 이루어져 한계가 있었던 반면, 본 연구는 토픽 모델링 기법을 적용하여 리뷰 데이터에 포함된 다양한 주제를 보다 폭넓게 도출한다.

특히, 기존 연구에서 주로 사용한 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)과 같은 전통적인 모델은 텍스트의 문맥을 충분히 반영하지 못한다는 한계를 가진다. 이를 보완하기 위해 본 연구에서는 임베딩 기반 BERTopic 모델을 적용하여 관광객의 리뷰 데이터를 분석한다. BERTopic은 문맥을 보다 정확하게 반영하여, 짧고 구조화되지 않은 텍스트에서 주요 패턴과 주제를 효과적으로 도출할 수 있다는 장점이 있다. 이를 통해 대규모 텍스트 데이터를 효과적으로 분석하고, 관광 경험에 대한 새로운 인사이트를 제공하고자 한다.

3. 연구 방법

3.1. 연구의 범위

본 연구는 서울특별시의 관광지를 대상으로 2012년부터 2023년까지 작성된 방문객 리뷰를 분석하였다. 2024년 5월 기준으로 Tripadvisor에 등록된 서울시 관광지는 총 2,598개에 달하지만, 리뷰가 없거나, 작성된 리뷰 수가 100개 이하인 경우는 제외하였다. Tripadvisor에 등록된 관광지 유형 및 관광지명을 기준으로 표1과 같이 총 60개의

관광지를 분석 대상으로 설정하였다. Tripadvisor는 관광지에 대한 리뷰와 평점, 방문일 등을 수집할 수 있는 영향력 있는 온라인 플랫폼이다(O'connor, 2010). Tripadvisor는 다양한 언어로 리뷰를 제공하고 있으며, 이용자는 원하는 언어로 리뷰를 선택해 조회할 수 있다. 본 연구는 외국인 관광객의 관광 경험을 파악하기 위해 영어로 작성된 리뷰를 분석 대상으로 하며, 방문일이 확인되지 않는 데이터는 분석에서 제외하였다. 또한, 내국인 관광객의 관광 경험과 비교하기 위해 한글로 작성된 내국인 리뷰를 수집하였다.

〈표 1〉 리뷰 수집 대상

관광지 유형	관광지 목록
역사적 장소 Historic Sites	Changdeokgung Palace, Changgyeonggung Palace, Deoksugung, Dongdaemun Gate(Heunginjimun), Gwanghwamun Gate, Gyeongbokgung Palace, Gyeonghuigung Palace, Jongmyo Shrine, Seolleung&Jeongneung Royal Tomb, Seoul City Wall, Sungnyemun Gate, Unhyeongung
지역 관광지 Neighborhoods	Bukchon Hanok Village, Ewha Womans University, Garosu-gil, Hongik University Street, Ihwa Mural Village, Insadong, Itaewon, Myeongdong Shopping Street, Namsangol Hanok Village, Seoul Daehangno, Sinchon
POI 및 랜드마크 Points of Interest & Landmarks	Bongeunsa Temple, Dongdaemun Design Plaza(DDP), Gwanghwamun Square, Lotte World Tower&Mall, N Seoul Tower, Naksan Public Art, Samcheongdong-gil Road, Samcheonggak, Seoul City Hall, Seoul Plaza
공원 Parks	Dongdaemun History&Culture Park, Haneul Park, Hangang Park, Naksan Park, Namsan Park, Seoul Forest, Seoullo7017, Tapgol Park, Yeouido Hangang Park, Yeouido Park
전통시장 Flea & Street Markets	Dongdaemun Market, Grilled Fish Street in Dongdaemun Market, Hongdae Art Market Free Market, Jungbu Market, Kwangjang Market, Majang Meat Market, Namdaemun Market, Noryangjin Fisheries Wholesale Market, Tongin Market
특산품 및 기념품 가게 Speciality & Gift Shops	Apgujeong Rodeo Street, Gangnam Underground Shopping area, Kyobo Book Store Gwanghwamun, Lotte Duty Free Main Store, Smurf Magic Forest, Ssamzigil, The Shilla Duty Free Seoul

3.2. 데이터 수집

데이터 수집을 위해 Python의 Selenium 4.1.0 패키지와 Beautiful Soup 라이브러리를 사용하여 웹 크롤링을 진행했다. 각 관광지 페이지에서 영어로 작성된 리뷰의 방문 시기, 평점, 본문을 수집하였으며, 이에 따라 수집한 1,522개 관광지의 리뷰 총합은 251,291건이다. 다만, 절반 이상의 관광지 리뷰 수는 7건 이하로 작성되어 있어, 관광객의 경험을 볼 수 있는 장소를 중심으로 분석을 진행하였다. 외래관광객의 주요 참여 활동인 쇼핑, 식도락 관광, 자연경관 감상, 고궁/역사 유적지 방문 등을 반영하여(외래관광객조사, 2023) 리뷰 수를 기준으로 상위 10개 유형 중 6개 유형을 선정했다. 투어 상품이나 대중교통 시스템, 공연 등의 관광지 유형은 제외하였다. 선정된 유형으로는 'Historic Sites', 'Neighborhoods', 'Points of Interest & Landmarks', 'Parks', 'Flea & Street Markets', 'Speciality & Gift Shops'로, 세부 관광지 중에서도 영어 리뷰 수가 10개 미만인 경우는 제외하였다. 연도별 수집된 데이터는 표2와 같다. 데이터 분석 결과, 2015년과 2017년에 리뷰 수가 가장 많았으며, 2020년부터 2021년에는 COVID-19 확산으로 인해 리뷰 수가 급격히 감소하였다. 2022년부터는 COVID-19의 영향력이 점차 감소하면서 외국인 관광객의 입국 증가와 함께 리뷰 수도 증가하는 추세를 보였다.

〈표 2〉 연도별 수집된 리뷰 수

연도	영어 리뷰	한글 리뷰
2012	1,711	34
2013	2,247	58
2014	3,717	252
2015	6,405	2,089
2016	8,403	1,215
2017	6,443	1,123
2018	4,992	511
2019	3,645	434
2020	247	105
2021	42	20
2022	249	11
2023	443	15
합계	38,544	5,867

3.3. 데이터 전처리

대량의 텍스트 데이터에서 의미 있는 패턴을 추출하기 위해 텍스트 마이닝 기법을 적용하였다(Sánchez-Franco & Rey-Moreno, 2022). 리뷰 데이터를 정제하고 분석하기 위해 다음과 같은 전처리 과정을 거쳤다. 먼저, 문장 단위로 본문을 분리하였으며, 다양한 주제를 포함하는 리뷰 본문을 문장 수준에서 분석하여 다양한 주제를 도출하고자 했

다(박홍제 외, 2021). NLTK 라이브러리에서 제공하는 불용어를 활용했고, 텍스트 전처리는 (1) 구두점 제거 (2) 소문자 변환 후 단어 토큰화 (3) 불용어 제거 및 표제어 추출 과정을 진행했다. 또한, 리뷰 문장에서 5단어 미만의 짧은 문장을 제거하여 분석에 적합한 데이터를 확보하였다.

3.4. BERTopic 모형

BERTopic은 BERT와 같은 사전 훈련된 트랜스포머 모델을 활용하여 문서 임베딩을 생성하고, 이를 기반으로 토픽을 도출하는 최신 토픽 모델링 기법이다. 기존의 LDA와 같은 전통적인 토픽 모델링 기법들은 단어 빈도에 기반하여 토픽을 도출하는 반면, BERTopic은 문장의 문맥을 반영하여 보다 의미 있는 토픽을 도출하고 문서간 의미적 유사성을 더 정확하게 반영한다(Grootendorst, 2022).

BERTopic의 기본 작동 원리는 네 단계로 나눌 수 있다. 먼저, 문서를 임베딩하는 데 SBERT를 사용한다. 그 다음, UMAP을 이용하여 임베딩된 차원을 축소하고, HDBSCAN을 통해 문서를 클러스터링 한다. 마지막으로, 클래스 기반 TF-IDF(cTF-IDF)를 적용해 단어의 점수를 부여한다. 이 과정에서 BERTopic은 연구자의 주관성을 배제하고 최적의 토픽 개수를 도출할 수 있게 한다.

본 연구에서는 영어의 경우 “all-MiniLM-L6-v2” 문장 임베딩 모델을 사용하며, 이를 통해 문서 간의 의미론적 유사성을 포착할 수 있다(Grootendorst, 2022). 한글 리뷰는 다국어 모델 “sentence-transformers/distiluse-base-multilingual-cased-v1”를 사용하였다. 또한, UMAP 모델의 파라미터를 설정하여 확률적 특성을 제어하고 일관된 토픽 결과를 도출하도록 하였다. 문서 임베딩의 차원을 줄이고 결과 임베딩에 보관된 정보를 최대한 활용하기 위해 기본값으로 설정하였으며, 이 UMAP 모델을 BERTopic에 사용하였다.

본 연구에서는 토픽의 키워드와 중요성을 바탕으로 객관적인 중심 주제를 도출하기 위해 OpenAI의 ChatGPT 4o를 사용했다. BERTopic에서 생성되는 기본 토픽 이름은 ‘토픽번호_키워드_키워드_키워드’ 형식으로 구성되므로, 이를 보다 직관적으로 이해할 수 있도록 주요 키워드와 중요도를 반영하여 토픽의 중심 주제를 표현하는 제목을 생성하였다(김혜빈 · 이수기, 2024).

이러한 과정을 통해 본 연구는 관광 부문에서의 사용자 생성 콘텐츠 분석에 BERTopic을 활용하고자 한다. 특히, 다양한 표현과 주제가 포함된 관광지 리뷰 데이터를 분석하는 데 있어, BERTopic은 기존의 토픽 모델링 기법보다 더 깊이 있는 통찰을 제공하여 관광객의 경험을 보다 정확하게 파악할 수 있다(Egger & Yu, 2022).

4. 결과

4.1. 연도별 주요 토픽 변화

2012년부터 2023년까지의 전체 데이터를 대상으로 BERTopic을 활용하여 토픽 모델링을 진행한 결과, 총 159개 토픽이 도출되었다. 표3과 같이 상위 토픽 10개를 살펴본 결과, 맛집과 카페 탐방, 쇼핑과 가격, 서울의 랜드마크 및 역사적 장소 등이 주요 토픽으로 나타났다. 이는 Tripadvisor에서 서울시 관광지와 관련해 작성된 리뷰에서 도출된 토픽 중 전체 기간 중 가장 빈도가 높은 순위이며, 각 토픽 번호의 중심 주제를 반영한 한글 제목으로 정리하였다.

특히, ‘맛집과 카페 탐방’ 토픽이 전체 리뷰에서 가장 많이 언급된 주요 토픽으로 나타나 외국인 관광객이 서울을 방문할 때 음식과 카페 경험을 중요한 요소로 인식하고

있음을 시사한다. 실제로 외래관광객조사(2023)에서도 한국을 방문하는 외국인 관광객의 주요 방문 요인으로 '식도락 관광(59.8%)'과 '쇼핑(57.9%)'이 높은 비율을 차지하는 만큼, '맛집과 카페 탐방' 및 '쇼핑과 가격'이 상위 토픽으로 도출된 결과와 맥락이 일치한다.

반면, '관광지 혼잡도' 관련 토픽이 상위 토픽에 포함된 점은 관광객들이 서울에서 경험한 활동뿐만 아니라, 혼잡함에 대한 인식과 불편함 또한 주요한 경험으로 인식되고 있음을 보여준다. 관광지의 혼잡도는 관광의 질과 매력도에 영향을 미치는 요소로 평가되며(Yin et al., 2020), 이는 기존의 공식적인 관광 통계에서는 잘 드러나지 않는 부분이다. 따라서, '관광지 혼잡도' 토픽과 함께 등장하는 키워드 및 관광지 유형별 관련 내용을 면밀히 분석하면, 관광지 운영 및 개선 방향을 도출하는 데 활용할 수 있다.

〈표 3〉 외국인 리뷰 상위 토픽의 키워드와 각 중요도

토픽	키워드1	키워드2	키워드3	키워드4	키워드5
0. 맛집과 카페 탐방	food(음식)	restaurant(식당)	cafe(카페)	street(거리)	coffee(커피)
	0.0355	0.0249	0.0242	0.0228	0.0187
1. 관광지 혼잡도	crowded(혼잡함)	crowd(군중)	time(시간)	hour(시간)	weekend(주말)
	0.0260	0.0204	0.0187	0.0182	0.0153
2. 쇼핑과 가격	price(가격)	shopping(쇼핑)	brand(브랜드)	sample(샘플)	bargain(싼 물건)
	0.0302	0.0259	0.0179	0.0171	0.0169
3. 서울타워와 케이블카	tower(타워)	view(경치)	lock(자물쇠)	city(도시)	love(사랑)
	0.0293	0.0185	0.0148	0.0105	0.0097
4. 궁궐과 역사	palace(궁)	huge(큰)	history(역사)	beautiful(아름다운)	royal(왕족)
	0.0557	0.0142	0.0102	0.0094	0.0092
5. 케이블카 타기	top(꼭대기)	view(경치)	elevator(엘리베이터)	lock(자물쇠)	queue(대기줄)
	0.0197	0.0196	0.0136	0.0114	0.0102
6. 전통 기념품	korean(한국의)	souvenir(기념품)	shopping(쇼핑)	traditional(전통적인)	craft(공예)
	0.0258	0.0228	0.0209	0.0115	0.0114
7. 한국 음식과 식당	korean(한국의)	food(음식)	restaurant(식당)	kimchi(김치)	chicken(치킨)
	0.0350	0.0312	0.0182	0.0127	0.0104
8. 서울 필수 방문지	seoul(서울)	must(반드시)	visit(방문하다)	place(장소)	definitely(분명히)
	0.0581	0.0298	0.0252	0.0186	0.0155
9. 명동 쇼핑	shopping(쇼핑)	hotel(호텔)	food(음식)	cheaper(저렴한)	compared(비교하다)
	0.0272	0.0195	0.0114	0.0098	0.0093

표4는 연구 기간의 서울시 관광지 관련 온라인 리뷰 데이터를 기반으로 BERTopic을 활용하여 도출된 주요 토픽들의 연도별 변화를 나타낸다. 이를 통해 관광객의 관심사가 시간의 흐름에 따라 어떻게 변화해왔는지를 확인할 수 있다. 다만, 2021년의 경우 전체 리뷰 수가 100개 미만으로 COVID-19의 영향을 크게 받은 해이며, 다른 연도와 비교해 도출된 토픽의 특성이 다소 차이를 보이므로 시계열 분석에서는 제외하였다.

연도별 분석 결과, '맛집과 카페 탐방', '쇼핑과 가격', '서울타워와 케이블카', '궁궐과 역사' 등의 토픽이 지속적으로 등장하며, 관광객들이 서울을 방문할 때 주요 명소와 쇼핑 활동에 꾸준히 관심을 가지고 있음을 보여준다. 또한, '관광지 혼잡도' 토픽이 계속해서 포함된 것은 서울 방문객들이 관광지의 인구 과밀과 관련된 경험을 중요하게 여기고, 반복적으로 언급되고 있음을 시사한다.

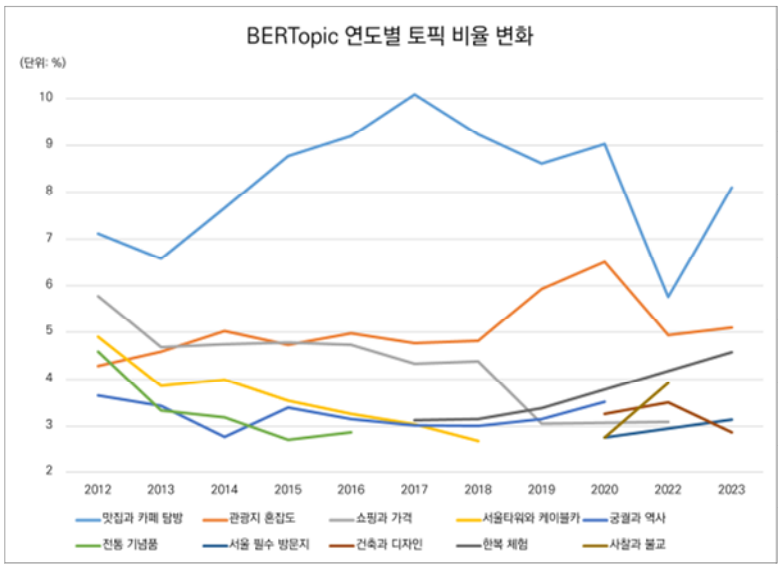
특히, 2017년 이후 '한복 체험'이 새로운 주요 토픽으로 등장하는데, 이는 한복 대여 서비스의 확산과 SNS를 통한 전통문화 체험이 외국인 관광객에게 인기를 얻었음을 반영하는 결과로 볼 수 있다. 한편, 2020년 이후에는 COVID-19의 영향으로 서울을 방문하는 관광객 수가 급격히 감소하였으며, '건축과 디자인' 및 '서울 필수 방문지'와 같은 보다 포괄적인 관광 관심사가 나타나는 등 전반적인 관심사의 변화가 감지되었다. 이처럼 서울 관광객의 관심사가 전통적인 명소와 소비 중심에서 문화 체험 및 자연경관을 즐기는 방향으로 변화하는 모습을 보여주며, 특정 시기의 사회·환경적 요인이 관광 패턴에 영향을 미칠 수 있음을 시사한다.

〈표 4〉 BERTopic 연도별 토픽 도출 결과

연도	토픽별 세부 내용과 빈도				
	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
2012	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=205)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=166)	3. 서울타워와 케이블카 tower, view, cable, seoul, car, lock, top, city, bus, love (n=141)	6. 전통 기념품 korean, souvenir, shopping, traditional (n=132)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=123)
2013	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=273)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=194)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=191)	3. 서울타워와 케이블카 tower, view, cable, seoul, car, lock, top, city, bus, love (n=160)	4. 궁궐과 역사 palace, huge, history, beautiful, royal (n=143)
2014	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=501)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=329)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=311)	3. 서울타워와 케이블카 tower, view, lock, city, love (n=261)	6. 전통 기념품 korean, souvenir, shopping, traditional (n=208)
2015	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=901)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=491)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=486)	3. 서울타워와 케이블카 tower, view, lock, city, love (n=363)	4. 궁궐과 역사 palace, huge, history, beautiful, royal (n=349)
2016	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=1,270)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=686)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=651)	3. 서울타워와 케이블카 tower, view, lock, city, love (n=449)	4. 궁궐과 역사 palace, huge, history, beautiful, royal (n=434)
2017	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=1,035)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=487)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=443)	12. 한복 체험 'hanbok', wearing, rent, free, picture (n=319)	5. 케이블카 타기 top, view, elevator, lock, queue (n=318)
2018	0. 맛집과 카페 탐방) food, restaurant, cafe, street, coffee (n=739)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=389)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=350)	5. 케이블카 타기 top, view, elevator, lock, queue (n=298)	12. 한복 체험 'hanbok', wearing, rent, free, picture (n=252)
2019	0. 맛집과 카페 탐방 'food, restaurant, cafe, street, coffee (n=505)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=349)	12. 한복 체험 'hanbok', wearing, rent, free, picture (n=198)	4. 궁궐과 역사 'palace, huge, history, beautiful, royal (n=185)	2. 쇼핑과 가격 'price, shopping, brand, sample, bargain (n=179)
2020	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=36)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=26)	4. 궁궐과 역사 palace, huge, history, beautiful, royal (n=14)	11. 건축과 디자인 design, modern, structure, old, restoration, (n=13)	8. 서울 필수 방문지 must, visit, place, visiting, definitely (n=11)
2022	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=28)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=24)	17. 사찰과 불교 temple, buddha, prayer, bongeunsa, monk (n=19)	11. 건축과 디자인 design, modern, structure, old, restoration, (n=17)	2. 쇼핑과 가격 price, shopping, brand, sample, bargain (n=15)
2023	0. 맛집과 카페 탐방 food, restaurant, cafe, street, coffee (n=62)	1. 관광지 혼잡도 crowded, crowd, time, hour, weekend (n=39)	12. 한복 체험 'hanbok', wearing, rent, free, picture (n=35)	8. 서울 필수 방문지 must, visit, place, visiting, definitely (n=25)	10. 한강과 공원 river, bike, picnic, along, cycling (n=24)

그림1은 연도별 각 토픽의 비율 변화를 요약해서 보여주며, 특히 2년 이상 지속적으로 상위에 나타난 토픽들의 추세를 파악할 수 있도록 하였다. 먼저, '맛집과 카페 탐방' 토픽은 지속적으로 높은 비율을 차지하고 있으며, '관광지 혼잡도' 토픽도 비율이 점차 증가하여 2016년부터는 꾸준히 두 번째로 높게 나타난다. 반면, '쇼핑과 가격'이나 '전통 기념품' 등 쇼핑과 관련된 토픽과 '서울타워와 케이블카' 토픽은 점차 비중이 감소하는 추세로 보인다. 새로이 등장하는 토픽으로는 '한복 체험'은 2017년부터, '사찰과 불교'는 2020년에 상위 토픽에 나타나며 일정 기간 상승하는 추세를 보였다. 이는 관광객의 관심이 문화적 체험 요소로 변화하고 있음을 시사한다.

관광지의 리뷰는 대체로 해당 지역에서 느낄 수 있는 문화 역사적 요소들이 많이 언급된다(Yu & Egger, 2021). 본 연구에서는 '관광지 혼잡도'가 연구 범위 동안 점차 증가하여 상위 토픽으로 유지되었다. 이는 많은 수의 관광객이 관광지에서 느낀 혼잡함을 꾸준히 언급하고 있음을 반영한다. 이러한 현상은 외국인 관광객의 관심사가 문화적 경험뿐만 아니라 관광지 혼잡함에도 영향을 받고 있다고 해석할 수 있다.



〈그림 1〉 BERTopic 연도별 토픽 비율 변화

4.2. 관광지 유형별 토픽

관광 경험을 파악하고자 할 때 개별 리뷰 수준에서의 분석도 중요하지만, 관광명소 차원에서 주요 요소를 식별하는 것도 관광객 경험을 종합적으로 이해하는 데 필수적이다(Taecharungroj & Mathayomchan, 2019). 본 연구에서는 Tripadvisor의 6개 관광지 유형을 기준으로 외국인 관광객 리뷰에서 도출된 상위 토픽을 분석하여 관광지별 주요 경험 요소를 정리하였다(표5 참조). 이를 통해 각 관광 유형에서 방문객들이 강조하는 경험과 특징을 파악할 수 있다.

역사적 장소(Historic Sites) 유형에서는 '궁궐과 역사'뿐만 아니라 관광지의 혼잡도, '비밀의 정원 투어', '한복 체험' 등 문화 체험 요소가 주요하게 언급되었다. 이는 해당 관광지의 역사적 가치와 건축적 특성이 방문객들에게 중요한 경험으로 작용했음을 보여준다. 또한, 영어로 진행되는 무료 가이드 투어에 대한 언급이 많아, 외국인 관광객들이 서울의 역사적인 건축물에 대한 설명을 듣는 기회를 중요하게 여김을 확인할 수 있다.

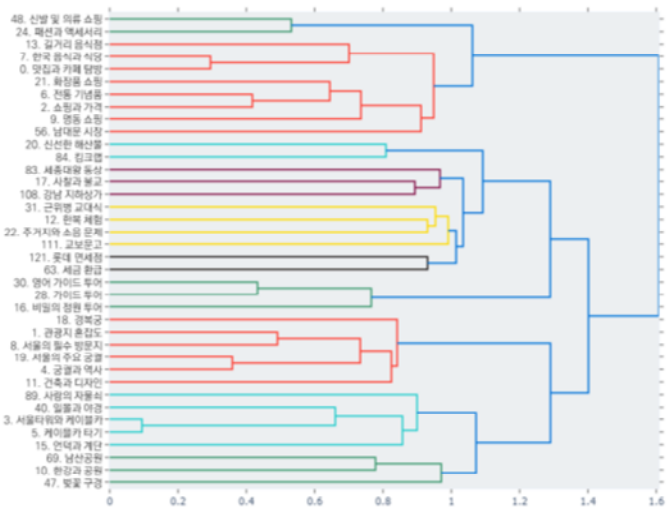
반면, 지역 관광지(Neighborhoods) 유형에서는 '맛집과 카페 탐방', '쇼핑과 가격', '전통 기념품' 등 음식 및 쇼핑과 관련된 언급이 대다수를 차지하고 있다. 이는 지역의 식문화 경험과 쇼핑이 관광객들에게 주요한 관심사임을 시사한다. 그러나 관광지 혼잡도 및 주거지에서의 소음 문제가 상위 토픽에 포함된 점을 고려할 때, 인근 거주민과의 갈등 가능성도 존재함을 알 수 있다.

POI 및 랜드마크 유형에서는 '서울타워와 케이블카', '케이블카 타기' 등의 토픽이 가장 빈번하게 등장했으며, 서울타워에서의 경험과 전망 감상이 핵심 관심사로 나타났다. 이러한 분석을 통해 관광지 유형별로 외국인 관광객이 기대하는 경험과 선호도를 파악할 수 있으며, 이는 도시 관광지 운영 및 서비스 개선에 중요한 시사점을 제공한다.

〈표 5〉 관광지 유형별 상위 토픽

역사적 장소 Historic Sites	지역 관광지 Neighborhoods	POI 및 랜드마크 Points of Interest & Landmarks	공원 Parks	전통시장 Flea & Street Markets	특산품 및 기념품 가게 Speciality & Gift Shops
4. 궁궐과 역사	0. 맛집과 카페 탐방	3. 서울타워와 케이블카	10. 한강과 공원	0. 맛집과 카페 탐방	2. 쇼핑과 가격
1. 관광지 혼잡도	2. 쇼핑과 가격	5. 케이블카 타기	5. 케이블카 타기	20. 신선한 해산물	111. 교보문고
16. 비밀의 정원 투어	1. 관광지 혼잡도	17. 사찰과 불교	47. 벚꽃 구경	7. 한국 음식과 식당	6. 전통 기념품
12. 한복 체험	6. 전통 기념품	11. 건축과 디자인	15. 언덕과 계단	13. 길거리 음식점	0. 맛집과 카페 탐방
19. 서울의 주요 궁궐	9. 명동 쇼핑	0. 맛집과 카페 탐방	3. 서울타워와 케이블카	2. 쇼핑과 가격	1. 관광지 혼잡도
18. 경복궁	7. 한국 음식과 식당	1. 관광지 혼잡도	1. 관광지 혼잡도	1. 관광지 혼잡도	121. 롯데 면세점
11. 건축과 디자인	8. 서울 필수 방문지	40. 일몰과 야경	0. 맛집과 카페 탐방	6. 전통 기념품	24. 패션과 액세서리
31. 근위병 교대식	21. 화장품 쇼핑	8. 서울 필수 방문지	69. 남산공원	84. 킹크랩	108. 강남 지하상가
28. 가이드 투어	13. 길거리 음식점	2. 쇼핑과 가격	40. 일몰과 야경	24. 패션과 액세서리	48. 신발 및 의류 쇼핑
30. 영어 가이드 투어	22. 주거지와 소음 문제	83. 세종대왕 동상	89. 사랑의 자물쇠	56. 남대문 시장	63. 세금 환급

이와 같이 도출된 관광지 유형별 상위 토픽을 BERTopic의 계층적 클러스터링(Hierarchical Clustering)을 통해 분류해 총 10개의 클러스터가 형성되었다(그림2 참조). BERTopic은 단어 임베딩을 활용해 유사한 주제를 클러스터링하는 방식으로 작동한다. 각 토픽 간의 연관성을 보다 효과적으로 파악할 수 있다.



〈그림 2〉 BERTopic 기반 계층적 클러스터링

표6은 각 클러스터별 토픽과 주요 키워드를 요약한 것으로, OpenAI의 ChatGPT 4o를 활용하여 클러스터 이름을 도출하였다. 총 10개의 클러스터는 '패션 및 액세서리 쇼핑', '거리 쇼핑과 식도락 체험', '신선한 해산물 시장', '도시 랜드마크', '전통문화 체험', '면세 쇼핑', '영어 가이드 투어', '서울의 궁궐과 현대 건축물', 'N서울타워 체험', '공원의 자연경관'으로 확인할 수 있다.

먼저, 쇼핑 관련 클러스터가 가장 다양한 형태로 나타났으며, 패션, 화장품, 면세품, 기념품, 음식점, 시장 등 여러 분야의 쇼핑 활동이 포함되었다. 또한, '전통문화 체험'과 관련하여 한복 체험 및 근위병 교대식이 주요 관심 요소로 나타났으며, 궁궐에서 제공되는 무료 영어 가이드 투어가 '영어 가이드 투어' 클러스터를 형성할 만큼 높은 언급 빈도를 보였다.

‘서울의 궁궐과 현대 건축물’ 클러스터는 Tripadvisor에서 가장 인기 있는 서울의 관광지인 경복궁을 포함해 여러 궁궐과 현대적인 건축물 관련 토픽을 포함하고 있다. ‘공원의 자연경관’ 클러스터에는 남산공원과 한강공원이 포함되며, ‘도시 랜드마크’ 클러스터는 서울 내 주요 랜드마크에 대한 언급을 중심으로 구성되었다. 특히, ‘N서울타워 체험’ 클러스터에서는 케이블카, 사랑의 자물쇠, 서울의 야경 감상 등의 토픽이 포함되었으며, 이는 N서울타워가 관광객들에게 높은 관심을 받는 장소임을 보여준다.

〈표 6〉 계층적 클러스터링 분석 결과

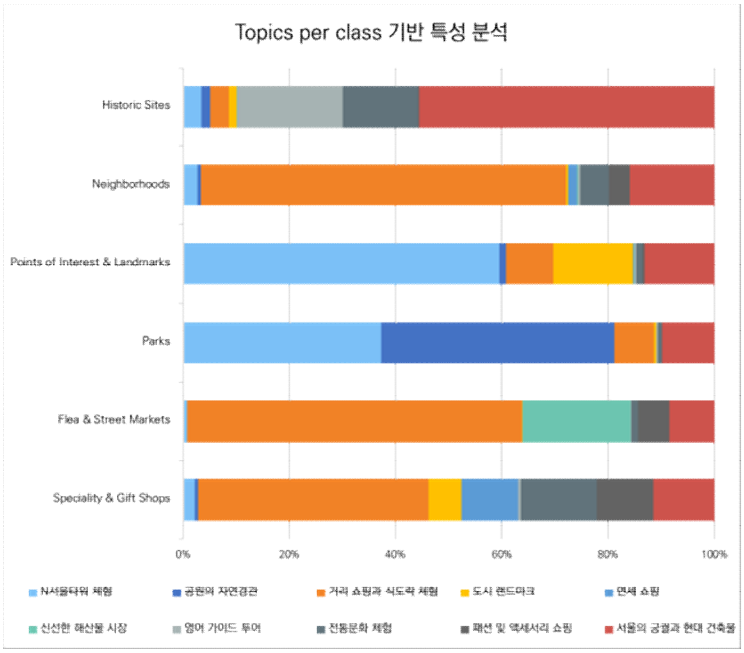
구분	Cluster	Topic	Keyword
1	패션 및 액세서리 쇼핑	신발 및 의류 쇼핑	shoe, sock, clothes
		패션과 악세서리	clothes, bag, clothing
2	거리 쇼핑과 식도락 체험	길거리 음식점	stall, food, selling
		한국 음식과 식당	korean, food, restaurant
		맛집과 카페 탐방	food, restaurant, cafe
		화장품 쇼핑	cosmetic, brand, korean
		전통 기념품	korean, souvenir, shopping
		쇼핑과 가격	price, shopping, brand
		명동 쇼핑	shopping, hotel, street
		남대문 시장	market, cheaper, compared
3	신선한 해산물 시장	신선한 해산물	seafood, fish, fresh
		킹크랩	crab, abalone, king
4	도시 랜드마크	세종대왕 동상	sejong, statue, admiral
		사찰과 불교	temple, buddha, buddhist
		강남 지하상가	gangnam, underground, station
5	전통문화 체험	근위병 교대식	guard, changing, change
		한복 체험	hanbok, wearing, rent
		주거지와 소음 문제	noise, residential, quiet
		교보문고	book, bookstore, kyobo
6	면세 쇼핑	롯데 면세점	duty, lotte, free
		세금 환급	tax, refund, passport
7	영어 가이드 투어	영어 가이드 투어	english, tour, language
		가이드 투어	tour, guided, guide
		비밀의 정원 투어	secret, garden, tour
8	서울의 궁궐과 현대 건축물	경복궁	gyeongbokgung, palace, changdeokgung
		관광지 혼잡도	crowded, crowd, time
		서울 필수 방문지	seoul, must, visit
		서울의 주요 궁궐	palace, seoul, one
		궁궐과 역사	palace, ground, history
		건축과 디자인	building, architecture, modern
9	N서울타워 체험	사랑의 자물쇠	lock, padlock, love
		일몰과 야경	sunset, light, night
		서울타워와 케이블카	tower, view, lock
		케이블카 타기	top, view, elevator
		언덕과 계단	stair, steep, hill
10	공원의 자연경관	남산공원	namsan, park, tower
		한강과 공원	river, bike, picnic
		벚꽃 구경	cherry, blossom, april

관광지 유형별로 나타나는 클러스터의 속성에는 차이가 있으며, 이는 각 관광지의 특성과 방문객의 경험 및 인식이 반영된 결과이다. 이러한 클러스터는 방문자가 직접 작성한 리뷰 텍스트를 분석하여 도출된 결과이다. 특정 관광지 유형에서 높은 비중을 차지하는 토픽은 해당 관광지의 주요 특성을 반영하며, 이를 파악하는 것은 빠르게 변화하는 온라인 환경에서 관광지의 이미지를 이해하고 개선 및 강화하는 데 중요한 정보를 제공할 수 있다. 그림3은 관광지 유형별 클러스터의 특성을 파악하기 위해 각 유형에서 도출된 토픽의 정규화된 빈도를 시각화한 것이다.

이는 BERTopic의 Topics per class 기능을 활용하여 관광지 유형별로 토픽을 분석한 결과이며, 각 유형 내에서 특정 토픽이 차지하는 상대적 비중을 나타냈다. 다만, 본 시각화는 각 관광지 유형 간 특정 토픽의 중요도를 직접 비교하는 것이 아니라, 각 관광지 유형 내부에서 해당 토픽이 차지하는 비중을 보여주는 방식으로 구성되었다. 따라서, 유형 간 클러스터의 상대적인 중요도를 비교하기보다는, 각 관광지 유형의 특징을 시각적으로 표현하는 데 초점을 맞추었다.

우선, 역사적 장소(Historic Sites)에서는 '서울의 궁궐과 현대 건축물' 클러스터의 중요도가 가장 높았으며, 그 다음으로 '영어 가이드 투어', '전통문화 체험' 순으로 나타났다. 지역 관광지(Neighborhoods)에서는 '거리 쇼핑과 식도락 체험'이 압도적으로 높은 비중을 차지하였으며, 지역 관광지에서 쇼핑 및 음식점 관련 활동이 집중적으로 이루어지고 있음을 확인할 수 있다. 이는 외국인 관광객이 지역 관광지인 명동, 인사동, 대학로 등을 방문 시 식사 및 쇼핑이 주 활동임을 보여준다. 실제로 외래관광객조사(2023)에서도 명동 및 로드샵에서 쇼핑한다는 응답 비율이 높게 나타났으며, 본 연구의 결과 역시 이러한 조사 결과와 맥락을 같이한다.

POI 및 랜드마크 유형에서는 'N서울타워 체험' 클러스터가 상대적으로 높은 중요도를 보였다. 이는 서울을 대표하는 랜드마크로서 N서울타워의 역할이 크며, 방문객들의 관심이 집중된다는 점을 반영한다. 공원(Parks) 유형에서도 N서울타워가 위치한 남산공원이 포함되면서 유사한 분포를 보였다. 공원 유형에서는 특히 '공원의 자연경관' 클러스터가 가장 높은 비중을 차지하고 있다.



〈그림 3〉 Topics per class 기반 특성 분석

전통시장(Flea & Street Markets) 유형에서는 '거리 쇼핑과 식도락 체험'이 가장 큰 비중을 차지하였으며, 쇼핑 및 음식과 관련된 활동에 대한 관심이 두드러졌다. 그 외에도 '신선한 해산물 시장', '패션 및 액세서리 쇼핑'과 같은 클러스터도 형성되었다. 특산품 및 기념품 가게(Speciality & Gift Shops) 유형에서는 다른 유형에 비해 클러스터 간 중요도의 차이가 상대적으로 낮게 나타났다. 이는 다른 관광지 유형에 비해 방문객들이 특정한 활동에 집중하는 경향이 낮음을 시사한다.

4.3. 내국인 관광 경험 비교

내국인 관광 경험과의 비교를 위해, 앞서 외국인 리뷰 분석과 마찬가지로 Tripadivsor에 등록된 서울시 주요 관광지 리뷰 중, 사용자 국적이 '대한민국'으로 등록된 리뷰만을 수집하였다. BERTopic을 통해 분석한 결과, 총 52개의 토픽이 도출되었으며, BERTopic의 계층적 클러스터링을 통해 유사한 주제를 묶어 총 14개의 클러스터로 분류하였다. 이를 통해 서울시를 방문한 내국인 관광객의 주요 관심사를 파악하고, 외국인 관광객과 이를 비교할 수 있다. 클러스터별 세부 내용은 표7에 나타나 있다.

먼저, 내국인 관광객이 도심에서 주로 즐기는 활동을 중심으로 클러스터가 형성되었다. '전망과 야경 명소', '궁궐 후원과 전통 체험', '골목과 소규모 문화 공간', '박물관과 전시 공간' 등 다양한 문화 체험 관련 클러스터가 나타났다. 또한 '쇼핑과 상점 거리 탐방', '전통시장과 먹거리 체험'처럼 외국인 관광객과 공통적으로 쇼핑 및 음식 관련 활동도 중요한 관심사로 확인되었다. '교통수단 이용' 클러스터는 내국인 관광객이 대중교통을 통한 접근성을 주요 요소로 고려할 수 있음을 보여준다. 또한, '도시 변화와 감상'과 '도심 속 힐링 공간'과 같은 클러스터에서는 도심 속에서의 정서적 경험에 대한 관심이 반영되었다.

'나들이 코스 추천' 클러스터에서는 가족, 친구, 연인 등과의 동행이 자주 언급되며, 내국인 관광객의 관광 활동이 사회적 활동과 밀접하게 연계되어 있음을 보여준다. 또한, '자연 감상', '공원과 나들이' 클러스터에서는 계절별 자연경관 및 도심 속 자연환경 요소에 대한 관심이 높게 나타났다.

마지막으로, '외국인 관광객과 혼잡함', '외국인 관광객 밀집 지역'과 같은 클러스터에서는 내국인 관광객이 느낀 혼잡함에 대한 경험을 담고 있다. 혼잡함과 외국인을 함께 언급하며, 외국인 관광객이 많은 장소에서 혼잡함을 더욱 민감하게 인식할 가능성이 있음을 시사한다.

〈표 7〉 계층적 클러스터링 분석 결과(내국인)

구분	Cluster	Topic	Keyword
1	전망과 야경 명소	서울의 타워와 전망 명소	타워, 서울, 야경
		남산타워와 전망대의 매력	타워, 올라가다, 전망대
2	교통수단 이용	지하철 이용 및 접근성	지하철, 호선, 편리하다
		순환버스 운행	버스, 타고, 올라가다
		케이블카 이용 경험과 비용	케이블카, 올라가다, 타고
3	나들이 코스 추천	강력 추천 및 리뷰	추천, 강력, 참고
		연인을 위한 데이트 코스 추천	데이트, 코스, 연인
		가족 및 친구와의 나들이	가족, 친구, 연인
4	궁궐 후원과 전통 체험	궁궐 후원과 투어의 예약	예약, 미리, 인터넷
		한복 무료 입장 혜택 정보	무료, 한복, 입다
		사진 촬영 명소와 포토스팟	사진, 찍다, 나오다
5	골목과 소규모 문화 공간	재미와 감정이 담긴 순간들	재밌다, 이루다, 줄이다
		테마형 놀이 및 탈출 게임	테마, 엘리베이터, 주차장
		재미와 역사 속 골목 여행	재미있다, 쏠쏠하다, 해설

6	도시 변화와 감상	문화 행사와 전시 · 공연 경험	행사, 전시, 이벤트
		현대 건축 디자인	건물, 디자인, 동대문
		야간 개장과 서울의 밤	야간, 저녁, 야경
		여행 경험과 개인적 감상	가보다, 한번, 없다
		과거와 현재의 변화된 거리	예전, 지금, 변하다
		대학가 분위기와 캠퍼스 문화	학교, 대학로, 캠퍼스
		서울의 고궁과 역사적 명소	경복궁, 창덕궁, 창경궁
		서울의 역사와 궁궐	역사, 조선시대, 궁궐
7	도심 속 힐링 공간	조용한 거주 지역	조용하다, 거주, 부탁드리다
		도심 속 휴식과 여유	도심, 시내, 휴식
		북촌 및 전통 마을 관광	북촌, 마을, 관광
8	자연 감상	벚꽃 축제와 계절별 자연경관	벚꽃, 여의도, 축제
		가을 단풍과 자연의 아름다움	가을, 단풍, 특히
9	공원과 나들이	계절 변화와 야외 활동 경험	여름, 겨울, 계절
		맥주와 치킨, 한강의 밤	맥주, 치킨, 한잔
		한강과 서울의 공원	공원, 한강, 주차
		한강 자전거 문화	자전거, 한강, 빌리다
		걷기 좋은 나들이 장소	좋다, 걷기, 나들이
		산책과 운동하기 좋은 공간	산책, 운동, 좋다
		날씨와 계절별 야외활동	날씨, 좋다, 산책
10	외국인 관광객 밀집 지역	외국인과 함께하는 관광	친구, 외국인, 데려가다
		외국인 관광객과 인기 방문지	외국인, 관광객, 많다
		외국인과 한국인이 모이는 장소	외국인, 한국인, 많다
11	전통시장과 먹거리 체험	한국적 정취 및 감성	한국, 느끼다, 대한민국
		전통문화와 쇼핑 명소	전통, 외국인, 쇼핑
		전통 문화 체험 및 유산	문화, 전통, 체험
12	외국인 관광객과 혼잡함	주말의 인파	주말, 평일, 사람
		혼잡한 장소와 인파 경험	사람, 많다, 불편하다
		관광명소 방문과 여행객 경험	방문, 관광객, 많다
		서울의 외국인 관광객	중국인, 관광객, 일본인
13	쇼핑과 상점 거리 탐방	아기자기한 소품 가게 탐방	아기자기하다, 소품, 골목
		쇼핑 문화	쇼핑, 가게, 브랜드
		외국인을 위한 명동 쇼핑 거리	외국인, 쇼핑, 외국
		카페 문화	카페, 커피한잔, 많다
		다양한 음식과 길거리 미식 탐방	음식, 식당, 길거리
		전통시장과 먹거리	시장, 광장시장, 김밥
14	박물관과 전시 공간	음식 가격과 길거리 음식	가격, 비싸다, 음식
		가격 비교 및 흥정 문화	비싸다, 저렴하다, 흥정

5. 토의 및 결론

5.1. 토의

본 연구는 서울의 주요 관광지에서의 방문객 경험을 분석하고, 주요 클러스터를 도출하여 유형별 핵심 경험 요소를 탐색하였다. 연구 결과, 시간의 흐름에 따라 관광 트렌드가 변화하고 있으며, 관광지 유형별로 방문객이 중시하는 경험이 상이하게 나타났다. 특히, 외국인 관광객과 내국인 관광객의 관광 경험에 영향을 미치는 요소 및 감정에서 차이가 나타났다.

관광 경험의 형성은 관광지의 문화적 · 물리적 자원에 영향을 받으며, 적절한 환경이 조성될 때 더 활발한 방문이 이루어질 수 있다(정은혜, 2018). 본 연구 결과는 서울특

별시(2022)의 '서울관광 활성화 계획(2022~2026)'에서 제시된 관광자원 개발 방향과도 일맥상통한다. 서울시는 기존의 쇼핑 중심 관광에서 벗어나, 수변공간, 등산과 같은 자연환경 및 야간 경관을 적극 활용하고, 역사 · 문화 · 자연이 조화를 이루는 관광자원을 확대하고자 한다. 분석 결과, 초기에는 쇼핑, 케이블카, 궁궐 등의 주요 관심사였으나, 이후 다양한 관광 경험이 부각되었다. 특히, N서울타워와 관련된 언급이 시간이 지남에 따라 상대적으로 낮은 비중을 차지하는 반면, 새로운 관광 콘텐츠 및 문화 체험에 대한 관심이 증가하는 경향이 나타났다. 이는 서울시가 추진하는 도심 관광 활성화 정책과도 맞닿아 있다.

그러나 본 연구에서는 '관광지 혼잡도'가 지속적으로 상위 토픽으로 도출되었으며, 이는 외국인 관광객이 서울시 방문에 혼잡함을 주요 경험으로 인식하고 있음을 시사한다. 기존 설문조사 연구에서 충분히 반영되지 않았던 '혼잡함'에 대한 경험이 온라인 리뷰 데이터 분석을 통해 높은 비중을 차지하고 있음을 확인하였으며, 이는 관광 빅데이터 분석이 도심 관광 경험을 보다 심층적으로 파악하는 데 기여할 수 있음을 보여준다.

이러한 분석 결과는 관광지별 맞춤형 마케팅 전략 수립에도 실질적인 시사점을 제시할 수 있다. 예를 들어, 역사적 장소에서는 영어로 제공되는 무료 가이드 투어에 대한 언급이 많았으며, 외국인 관광객의 관심도가 높게 나타났다. 이에 따라 해당 서비스에 대한 홍보를 강화하고, 제공 범위를 확대할 필요가 있다. 반면, 지역 관광지에서는 관광객과 지역 주민 간의 갈등을 반영하는 토픽이 도출되었다. 특히, 혼잡함뿐만 아니라 주거지 인근에서의 소음 문제가 주요 이슈로 나타났으며, 이는 관광객 증가로 인한 오버투어리즘(overtourism)과 관련이 있다. 이러한 문제는 지역 주민과 방문객 간의 갈등을 초래할 수 있으며, 궁극적으로 관광 경험에도 부정적인 영향을 미칠 수 있다. 서울특별시(2022)에서도 관광객과 지역 주민 간의 긍정적인 상호작용을 강조하며 지속 가능한 관광 환경을 조성하고자 하고 있다. 기존 연구에서도 오버투어리즘이 관광지의 부정적 이미지 형성에 영향을 미칠 수 있다고 지적된 바 있으며(정혜진 · 이충기, 2020), 본 연구에서도 이러한 문제를 확인할 수 있었다.

또한, POI 및 랜드마크 유형 관광지의 주요 토픽에서 나타난 바와 같이, N서울타워가 여전히 외국인 관광객에게 높은 관심을 받고 있으나, 다른 랜드마크 관광지는 상대적으로 언급 빈도가 낮은 것으로 나타났다. 이는 서울의 관광 트렌드가 특정 관광지에 집중되는 경향을 반영하는 것으로 볼 수 있으며, 이에 따라 다양한 랜드마크 관광지를 활성화하기 위한 전략이 필요하다.

마지막으로, 내국인 관광객은 일상 속 여가 활동과 정서적 경험에 대한 관심이 높아, 서울 도심에서의 힐링과 나들이와 관련된 토픽이 다수 나타났다. 반면, 외국인 관광객은 서울의 역사적 · 전통적인 요소를 탐색하고, 대표적인 관광지(궁궐, N서울타워 등)를 중심으로 언급하는 경향을 보였다. 두 그룹 모두 쇼핑 및 음식 관련 경험을 중요하게 여긴다는 점에서 공통점이 있으나, 내국인 관광객은 가족, 친구, 연인과의 동행을 언급하며 사회적 활동과 연계하는 패턴이 강하게 나타났다. 또한, 외국인 관광객과 혼잡함을 함께 언급하는 경우가 많아, 이들이 공간적으로 충돌하는 경험을 하고 있음을 시사한다. 이는 향후 서울의 도심 관광 정책에서 국내 · 외 관광객이 공존할 수 있는 관광 동선 및 시설을 분산하는 전략이 필요함을 보여준다.

향후 서울시는 지속가능한 도심 관광 운영을 위해 랜드마크에서 새로운 문화 체험적 요소를 추가하거나, 도심 내 자연경관을 활용한 관광 프로그램을 개발하는 방식이 효과적으로 작용할 수 있다. 또한, 전통시장과 같은 공간을 활용하여 서울시 도심의 역사 · 문화 · 자연 자원을 활용한 관광 모델의 확대도 지속가능한 방문객 유치로 이어질 수 있다.

5.2. 결론

본 연구는 Tripdvisor의 리뷰 데이터 44,411개를 바탕으로 BERTopic을 활용해 서울시 주요 관광지별 관광 경험을 분석했다. 구체적으로 첫째, 2012년부터 2023년까지 작성된 영어 리뷰를 통해 외국인 관광객의 전반적인 주요 관심사의 시계열 변화를 분석했다. 둘째, 특정 관광지 유형에서 나타난 토픽을 살펴보았다. 셋째, 영어 리뷰와 내국인 리뷰 데이터를 바탕으로 관광객의 관심사 차이를 비교했다. 연구 결과, 서울시의 관광 경험은 쇼핑 중심에서 벗어나 보다 다양한 콘텐츠가 강조되는 방향으로 변화하고 있으며, 이는 서울시 관광 정책의 지향점과도 일치함을 확인하였다. 또한, 구조화되지 않은 리뷰 데이터에서 도출된 '관광지 혼잡도' 이슈가 관광 경험에서 큰 비중을 차지함을 확인하였으며, 이는 서울시 관광환경 개선을 위한 주요 과제로 고려될 필요가 있다.

기존 연구들이 설문조사를 통해 관광 경험을 유형화하고 만족도와의 관계를 분석한 반면, 본 연구는 리뷰 데이터를 활용하여 설문조사에서 포착하기 어려운 다양한 경험 요소를 도출하였다. 특히, 특정 관광지에 국한되지 않고 서울시 전반을 대상으로 관광 경험의 공통 요소와 유형별 특성을 분석함으로써, 관광 흐름을 이해할 수 있었다. 또한, 10년 이상 축적된 리뷰 데이터를 BERTopic 모델을 통해 지속적으로 유지되는 관광 경험과 새롭게 등장하는 경험을 시계열적으로 분석하였으며, 이를 통해 변화하는 관광 정책과의 일치성을 확인하였다. 이러한 접근은 장기적인 도시관광 연구에서 온라인 빅데이터의 활용 가능성을 보여주며, 보다 정밀한 정책 수립에 기여할 수 있음을 시사한다.

그러나 본 연구에는 몇 가지 한계가 있다. 먼저, 영어로 작성된 리뷰만을 대상으로 분석을 진행하여, 다른 언어를 사용하는 관광객들의 경험을 충분히 반영하지 못했다. 다만, 본 연구에서 활용한 Tripadvisor 데이터의 경우, 대표적인 서울 관광지인 경복궁에서도 전체 11,659개 리뷰 중 영어 리뷰가 6,410개로 절반 이상을 차지하며, 영어 리뷰가 높은 비중을 보였다. 따라서 본 연구에서는 분석의 일관성을 유지하기 위해 영어 리뷰와 내국인의 경험을 담고 있는 한글 리뷰만을 사용하였다. 그러나 우리나라의 주요 방문객이 중국, 일본 등 동아시아 관광객이라는 점을 고려하면, 이들의 경험이 충분히 반영되지 않은 것은 한계로 작용할 수 있다. 향후 연구에서는 보다 다양한 플랫폼과 언어의 리뷰 데이터를 수집하여 분석할 필요가 있다. 이를 위해 다국어 모델과 번역 기능을 활용한 연구를 수행한다면, 보다 포괄적인 관점에서 서울시의 관광 경험을 분석할 수 있을 것이다.

그럼에도 불구하고, 본 연구는 온라인상에 축적된 관광 빅데이터를 활용하여 서울시의 시계열적인 관광 관심사 변화를 분석하고, BERTopic 모델을 통해 주요 관광 경험 요소를 유형별로 도출하여 개선점을 탐색했다는 점에서 학술적 의의가 있다. 이러한 연구 결과는 서울시가 지향하는 관광 정책 방향을 설정하는 데 유용한 자료로 활용될 수 있으며, 도심 자원을 효과적으로 활용한 관광 활성화 전략 수립에 기여할 수 있을 것이다.

참고문헌

1. 곽은진 · 김이태 2017, "지역 음식 관광의 경험가치, 지역이미지 및 관광행동의도에 관한 연구", 『관광레저연구』, 29권, 4호, pp.41-58.
2. 김나연 · 강영옥 2019, "지오태깅된 사진 데이터를 활용한 서울방문 관광객의 주요 관광지 분석", 『한국지도학회지』, 19권, 1호, pp.35-46.
3. 김혜빈 · 이수기 2024, "COVID-19 전 · 후 우울감 완화를 위한 도시공간 활용의 변화 분석-소셜미디어 빅데이터와 KoBERTopic 모형의 적용", 『도시설계』, 25권, 2호, pp.109-125.

4. 박경열 · 안희자 2019, “텍스트 마이닝을 활용한 DMZ 관광 이슈의 토픽 모델링 분석”, 『관광레저연구』, 31권, 4호, pp.143-159.
5. 박득희 외 3인 2015, “사회네트워크 분석권, Social Network Analysis)을 이용한 방한 외래 관광객의 관광목적지 선택 연구: 서울 · 경인지역 관광목적지를 중심으로”, 『관광연구저널』, 29권, 9호, pp.5-19.
6. 박홍제 · 이도현 · 김경옥 2021, “외국인 관광객 리뷰 분석을 통한 외국인 관광객 수에 영향을 끼치는 요인 분석”, 『대한산업공학회지』, 47권, 1호, pp.45-55.
7. 신주하 · 임희진 · 이병주 2021, “국내 복합리조트 서비스품질에 대한 토픽모델링 및 감성분석”, 『관광연구저널』, 35권, 11호, pp.191-206.
8. 심영석 · 김홍범 2016, “텍스트 마이닝을 이용한 관광지 이미지 구성요인 및 측정에 관한 연구”, 『관광학연구』, 40권, 7호, pp.221-245.
9. 윤유식 · 김은진 · 송래헌 2017, “기억할만한 관광경험권, MTE) 세분 집단에 따른 관광목적지 이미지, 만족도 및 행동의도 차이연구”, 『관광연구저널』, 31권, 1호, pp.19-34.
10. 윤희정 2024, “해안도서형 자연유산에 대한 외국인 관광객의 관광경험구조 - TripAdvisor의 제주 온라인 리뷰 및 평점을 대상으로”, 『한국경관학회지』, 16권, 1호, pp.80-91.
11. 이현주 2017, “빅데이터를 활용한 경복궁 방문 경험 분석: Tripadvisor.com 온라인 리뷰의 활용”, 『관광연구』, 32권, 2호, pp.297-318.
12. 장민희 · 김남조 2022, “데이터마이닝 기법을 활용한 관광목적지 이미지와 온라인 관광정보의 지각된 유용성과의 관계분석”, 『관광레저연구』, 34권, 1호, pp.5-28.
13. 장윤정 · 정수아 2021, “빅데이터를 활용한 외국인 관광객의 국가 그룹별 방문특성 및 유사성: 트립어드바이저 자료를 중심으로”, 『관광연구저널』, 35권, 3호, pp.95-109.
14. 전정연 · 전현모 2022, “역사문화관광지에서의 관광경험과 관광만족, 삶의 질, 재방문의도 간 관계 - MZ 세대를 중심으로”, 『한국외식산업학회지』, 18권, 3호, pp.243-256.
15. 전형규 · 강인호 · 조원섭, 2010, “관광지의 특성, 관광객의 지각된 가치, 만족, 행동의도 간의 관계연구”, 『관광연구』, 25권, 3호, pp.237-258.
16. 정규엽 · 김한선 2009, “외래관광객의 만족도에 영향을 미치는 서울의 도시관광요소와 도시이미지”, 『호텔경영학연구』, 18권, 2호, pp.201-219.
17. 정은혜 2018, “도시재생을 통한 도시관광 논의: 서울로 7017을 사례로”, 『한국도시지리학회지』, 21권, 1호, pp.53-64.
18. 정혜진 · 이충기 2020, “북촌 한옥마을의 진정성이 관광행동에 미치는 영향연구: 계획행동이론 적용”, 『관광연구저널』, 34권, 5호, pp.5-18.
19. 한지연 · 김홍범 2017, “빅데이터 기반의 사회연결망 분석을 이용한 관광지 이미지 인식에 관한 연구”, 『관광학연구』, 41권, 8호, pp.91-119.
20. Alaei, AR, Becken, S & Stantic, B 2019, ‘Sentiment analysis in tourism: capitalizing on big data’, *Journal of Travel Research*, vol. 58, no. 2, pp.175-191.
21. Baka, V 2016, ‘The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector’, *Tourism Management*, vol. 53, 148-162.
22. Baloglu, S & McCleary, KW 1999, ‘A model of destination image formation’, *Annals of Tourism Research*, vol. 26, no. 4, pp.868-897.
23. Cenni, I 2024, ‘Sharing travel experiences on TripAdvisor: A genre analysis of negative hotel reviews written in French, Spanish and Italian’, *Journal of*

Pragmatics, vol. 221, 76-88.
24. Chen, CF & Chen, FS 2010, 'Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists', *Tourism Management*, vol. 31, no. 1, pp.29-35.
25. Choi, S, Lehto, XY, Morrison, AM 2007, 'Destination image representation on the web: Content analysis of Macau travel related websites', *Tourism Management*, vol. 28, no. 1, pp.118-129.
26. Crompton, JL 1979, 'Motivations for pleasure vacation', *Annals of Tourism Research*, vol. 6, no. 4, pp.408-424.
27. Crouch, GI & Ritchie, JB 1999, 'Tourism, competitiveness, and societal prosperity', *Journal of Business Research*, vol. 44, no. 3, pp.137-152.
28. Echtner, CM & Ritchie, JB 1991, 'The meaning and measurement of destination image', *Journal of Tourism Studies*, vol. 2, no. 2, pp.2-12.
29. Echtner, CM & Ritchie, JB 1993, 'The measurement of destination image: An empirical assessment', *Journal of Travel Research*, vol. 31, no. 4, pp.3-13.
30. Egger, R & Yu, J 2022, 'A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts', *Frontiers in Sociology*, vol. 7, pp.1-16.
31. Grootendorst, M 2022, *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*, arXiv preprint (in press).
32. Hong, JE 2020, 'Qualitative analysis of users' negative reviews on TripAdvisor: International tourists' reviews on Gyeongbokgung Palace in Seoul, Korea', *International Journal of Geospatial and Environmental Research*, vol. 7, no. 1, pp.5.
33. Kandampully, J, Juwaheer, TD & Hu, HH 2011, 'The influence of a hotel firm's quality of service and image and its effect on tourism customer loyalty', *International Journal of Hospitality & Tourism Administration*, vol. 12, no. 1, pp.21-42.
34. Kim, JH, Ritchie, JB & McCormick, B 2012, 'Development of a scale to measure memorable tourism experiences', *Journal of Travel Research*, vol. 51, no. 1, pp.12-25.
35. Kirilenko, AP & Stepchenkova, S 2025, 'Facilitating topic modeling in tourism research: Comprehensive comparison of new AI technologies', *Tourism Management*, vol. 106, 105007.
36. Költringer, C & Dickinger, A 2015, 'Analyzing destination branding and image from online sources: A web content mining approach', *Journal of Business Research*, vol. 68, no. 9, pp.1836-1843.
37. Lai, K & Li, X 2016, 'Tourism destination image: Conceptual problems and definitional solutions', *Journal of Travel Research*, vol. 55, no. 8, pp.1065-1080.
38. Lee, CC & Chang, CP 2008, 'Tourism development and economic growth: A closer look at panels', *Tourism Management*, vol. 29, no. 1, pp.180-192.
39. Li, X et al. 2021, 'Review of tourism forecasting research with internet data', *Tourism Management*, vol. 83, 104245.
40. Liu, MT et al. 2021, 'Using text mining to track changes in travel destination image: The case of Macau', *Asia Pacific Journal of Marketing and Logistics*, vol.

33, no. 2, pp.371-393.
41. Lu, W & Stepchenkova, S 2015, 'User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software', *Journal of Hospitality Marketing & Management*, vol. 24, no. 2, pp.119-154.
42. Luo, Y et al. 2023, 'Exploring destination image through online reviews: An augmented mining model using latent Dirichlet allocation combined with probabilistic hesitant fuzzy algorithm', *Kybernetes*, vol. 52, no. 3, pp.874-897.
43. Nilashi, M et al. 2018, 'Travelers decision making using online review in social network sites: A case on TripAdvisor', *Journal of Computational Science*, vol. 28, 168-179.
44. O'Connor, P 2010, 'Managing a hotel's image on TripAdvisor', *Journal of Hospitality Marketing & Management*, vol. 19, no. 7, pp.754-772.
45. Park, ST & Liu, C 2022, 'A study on topic models using LDA and Word2Vec in travel route recommendation: Focus on convergence travel and tours reviews', *Personal and Ubiquitous Computing*, vol. 1-17.
46. Qu, H, Kim, LH & Im, HH 2011, 'A model of destination branding: Integrating the concepts of the branding and destination image', *Tourism Management*, vol. 32, no. 3, pp.465-476.
47. Ruytenbeek, N, Verschraegen, M & Decock, S 2021, 'Exploring the impact of platforms' affordances on the expression of negativity in online hotel reviews', *Journal of Pragmatics*, vol. 186, pp.289-307.
48. Sánchez-Franco, MJ & Rey-Moreno, M 2022, 'Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings', *Psychology & Marketing*, vol. 39, no. 2, pp.441-459.
49. Taecharungroj, V & Mathayomchan, B 2019, 'Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand', *Tourism Management*, vol. 75, pp.550-568.
50. Uriely, N 2005, 'The tourist experience: Conceptual developments', *Annals of Tourism Research*, vol. 32, no. 1, pp.199-216.
51. Wang, R & Hao, JX 2018, 'Gender difference on destination image and travel options: An exploratory text-mining study', *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, pp.1-5.
52. Xiang, Z et al. 2017, 'A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism', *Tourism Management*, vol. 58, pp.51-65.
53. Yin, J et al. 2020, 'Tourists perceived crowding and destination attractiveness: The moderating effects of perceived risk and experience quality', *Journal of Destination Marketing & Management*, vol. 18, pp.1-18.
54. Yu, J., & Egger, R. 2021, 'Tourist experiences at overcrowded attractions: A text analytics approach', in *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference*, Springer, Cham. pp.231-243.
55. 문화체육관광부 2023, "2023년 외래관광객조사 보고서".
56. 서울특별시 2022, "서울관광 활성화 계획(2022~2026)".

투고일 2024.11.10
심사일 2024.12.20
게재확정일 2025.02.19