

한국인터넷전자상거래학회

2022 추계학술대회

머신러닝을 이용한 고객 만족 요인 분석 : 공유 경제 플랫폼을 중심으로



배준호, 김성수, 김희웅
연세대학교 정보대학원

I. 서론 : 연구 배경

- 유연근무의 일환으로 장기간 여행지에 머무르며 휴가와 업무를 병행하는 워케이션(Workcation)에 대한 관심이 확산
- 전통 호텔 업계뿐만 아니라 숙박 플랫폼 또한 여행객이 머물 수 있는 장기 숙박 상품을 개발
- 장기간 여행지에 머무는 “체류형 관광”은 국내에서는 2010년대 중반부터 국내에서 나타나기 시작(김소윤 외, 2016)
- 일반 관광보다 숙박 기간이 긴 체류형 관광은 숙박 시설의 고정 비용을 줄여주어 높은 수익성과 직접 관련이 있음(Barros et al., 2010)

트렌드로 부상한 워케이션의 지속 성장 기대

1 코로나19 이후 원격근무, 재택근무 등 일하는 방식 변화에 따라 국내·외 워케이션 관심 증가 및 지속적 성장 전망
 (글로벌) 19년 대비 20년 워케이션 소셜버즈량 200% 증가
 (국내) 원격근무 확산 된 20년 7월 소셜버즈량 전년 대비 100% 증가 워케이션 도입 본격화 된 21년 소셜버즈량 전년대비 300% 증가



출처: 데이터 구글 트렌드 및 네이버 트렌드

호텔과 플랫폼의 장기 숙박 상품 개발

LONG TERM STAY, JEJU
**장기 숙박
 할인 이벤트!**
 2021.06.21 - 2021.08.31
 최대 이용수익률은 "일주일(7박7일)"에서
 나타날 수 있는 특별한 혜택

일주일 (6박7일) 스탠다드 199,900-	보름 (14박15일) 슈퍼리어 (90㎡방) 399,900-	한달 (29박30일) 디럭스 (100㎡방) 649,900-
--------------------------------	--	--

장기숙박 이벤트 혜택
 - 세탁실 무료 이용 가능
 - 생수 무료 제공
 - 한달살이 예약 고객에 한하여 주1회 1인 무료 조식 이용

‘한 달 살기도
 ‘워케이션’도
 여기어때에서
 #연박예약시
 #더블혜택

2박 이상 예약하실 분!
**최대 20% 특가 +
 2만원 추가 할인**
 중독할만 적용 가능
 전 숙소
20,000원
 30만원 이상 결제 시

여기어때.

출처: 엠스테이호텔 제주, 여기어때

I. 서론 : 연구 동기

- 체류형 관광객은 민박 또는 공유 숙소를 주로 이용하며 숙박 시설에 대한 정보의 불확실성으로 어려움을 겪고 있음(김송이, 2019)
- 호텔 서비스의 경우 일반적으로 수용되는 서비스 품질 수준이 있으므로 상품에 대한 불확실성이 낮지만, 민박이나 공유형 숙소의 경우 상품에 대한 불확실성으로 인해 소비자가 선택한 결과로 발생할 수 있는 기대 손실이 높음(Yi et al., 2020)
- 비교적 오랜 기간 숙소에 머무는 체류형 관광에서 숙소 서비스 불확실성은 고객 불만족으로 이어질 수 있음



출처: 제주여행소비자권익증진센터, 미디어뉴스

I. 서론 : 연구 목적

□ Research Gap

- 체류형 관광객은 숙박 시설에 대한 정보의 불확실성으로 어려움을 겪고 있으며 이는 숙소 서비스의 불만족은 고객 불만으로 이어질 수 있음
- 공유형 숙박 서비스 품질과 고객 만족에 관한 연구는 Airbnb를 중심으로 한 일반적인 공유 숙박 경험을 바탕으로 진행한 연구가 대다수
- 체류형 관광의 공유 숙박 경험에 대한 고려가 부족하다는 한계가 존재

□ Research Objective

- 체류형 관광에서 공유 숙박 서비스 이용 고객의 서비스 만족 요인을 파악

□ Research Question

- (1) RQ 1 : 체류형 공유 숙소의 서비스 품질 요인은 무엇인가?
- (2) RQ 2 : 고객 만족에 영향을 미치는 체류형 공유 숙소의 서비스 품질 요인이 무엇인가?

Ⅱ. 개념적 배경 : 체류형 관광

□ 체류형 관광이란?

- 체류형 관광은 체류 기간을 기준으로 목적지에 오랜 기간 머무르는 형태의 관광행위라 정의할 수 있음(Anantamogkolkul et al., 2019)
- 체류 기간에 대한 기준은 국가에 따라 서로 다르게 정의할 수 있음(Gössling et al., 2018)
- 한국의 경우 대부분의 여행 수요가 일주일 미만이기 때문에 장기 체류의 기준을 일주일로 정의(김송이 외, 2020).

□ 체류형 관광객과 일반적 관광객의 구분되는 특성

- 동행인과 함께 관광활동을 하면서 휴식을 취하는 활동을 하는 데 주안점을 둠(이상훈 and 류시영, 2016)
- 체류형 관광객은 긴 시간 동안 지역 교류를 하면서 보다 진정성 있는 지역 경험을 함(Anantamongkolkul et al., 2019)

□ 체류형 관광에서의 숙박 시설

- 비교적 오랜 기간 동안 체류하기 때문에 체류 기간동안 머물게 되는 숙박 시설에 대한 선택이 중요(이상훈 and 류시영, 2016)
- 체류형 관광객은 민박 또는 공유 숙박을 주로 이용하며 숙박 시설에 대한 정보의 불확실성으로 어려움을 겪고 있음(김송이, 2019)
- 숙박 서비스 품질은 체류 기간과 밀접한 관계가 있으며(Barros et al., 2010), 체류 기간에 따라 고객이 중시하는 서비스 품질이 달라짐

II. 개념적 배경 : 공유 숙박에서의 서비스 품질 속성과 사용자 만족

- 공유 숙소에 대해 고객 만족 요인에 영향을 미치는 서비스 품질 속성을 찾기 위해 많은 연구가 이루어짐
- 대다수의 연구가 체류형 관광을 고려하지 않은 일반적인 공유 숙박 경험을 바탕으로 진행
- 서비스 품질에 대한 사전 기반의 감성 분석을 적용해 리뷰의 전후 맥락을 반영하지 못했다는 한계가 있음 (Ding, et al., 2021)

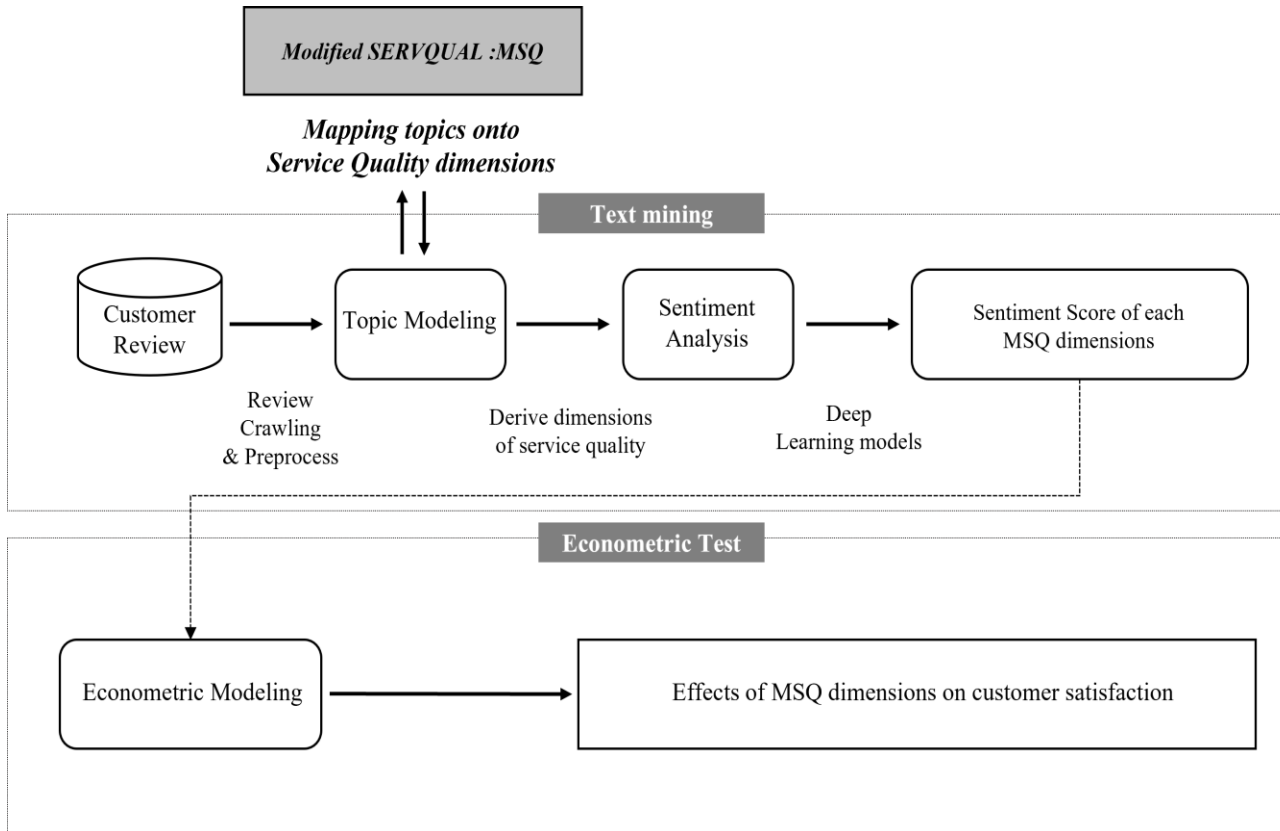
Study	Method	Key Finding
Priporas et al. (2017)	Survey Methods	1) Airbnb guests are interested in lodgings which have access to certain tourist sights, and in easily accessible information and efficient resolution of problems during their stay 2) Airbnb guests have very low expectations of the amenities and services available at the lodgings.
Tussyadiah and Zach. (2017)	Text mining and regression analysis	Attributes frequently mentioned in guest reviews are associated with location (proximity to point of interest and characteristics of neighborhood), host (service and hospitality), and property
Ju et al. (2019)	Text mining and factor analysis	Service quality attributes associated with website, host, and facility that produce distinctive effects on customer satisfaction
Ding et al. (2021)	Text mining and Sentiment analysis	tangibility are the major sources of Airbnb users' dissatisfaction. attributes under the dimension of convenience are the major source of user satisfaction
Amat-Lefort et al. (2022)	Text mining and Sentiment analysis	a significant positive relation to the guest's sentiment: apartment views, host tips and advice, location and host friendliness. negatively correlated with user sentiment: sleep disturbance, website responsiveness, thermal management and hygiene issues

II. 개념적 배경 : 배경 이론 Modified SERVQUAL

- 5개의 차원으로 구성된 (tangibles, reliability, response, assistance, and empathy) SERVQUAL은 숙박 산업에서 가장 많이 활용되는 서비스 품질 측정도구(Ladhari,2012).
- 하지만 SERVQUAL은 공유 숙박과 같은 새로운 비즈니스에 대한 서비스 품질을 측정하기에 일반적이라는 한계가 존재(Adbholkar et al., 2000)
- Akbaba(2006)이 수정한 Modified SERVQUAL : MSQ 는 전통 호텔 산업과 공유 플랫폼의 서비스 품질 측정에도 검증 및 활용(Ding et al., 2020)
=> Akbaba(2006), Ding et al.(2020)의 MSQ 관점에서 서비스 품질 속성을 도출하고 어떠한 속성이 고객 만족에 영향을 미치는 분석

MSQ Dimension	정의
Tangibles	the environment in which the service is delivered. It also includes any tangible commodities that facilitate the service's performance or communication (Zeithaml and Bitner, 1996:111)
Adequacy in Service supply	the ability to deliver the promised service consistently and dependably (Abu Alroub et al., 2012).
Understanding and Caring	caring individualized attention, the firm provides its customer(Zeithaml et al., 2006, 120) through which the customer is treated as if he is unique and special
Assurance	the employees' knowledge and courtesy and the service provider's ability to inspire trust and confidence" (Zeithaml et al., 2006, 119)
Convenience	Getting information about the facilities and services of the property is easy. It is easy to access property(Ding et al.,2021)

Ⅲ. 연구 방법론 : 연구 절차



Data Collecting and Preprocessing

- ✓ 체류형 공유 숙박 플랫폼 Mr,mention 숙박 Review 활용
- ✓ 2017 ~ 2022년 5월 7박 이상 숙박한 고객 리뷰 데이터 4883건 활용
- ✓ 불용어 처리 및 토큰화

Text mining

- ✓ GSDMM 활용 리뷰 Topic modeling 진행
- ✓ 도출된 Topic과 Modified SERVQUAL :MSQ 연결
- ✓ KoBERT 활용 MSQ Dimension 별 감성 스코어 도출

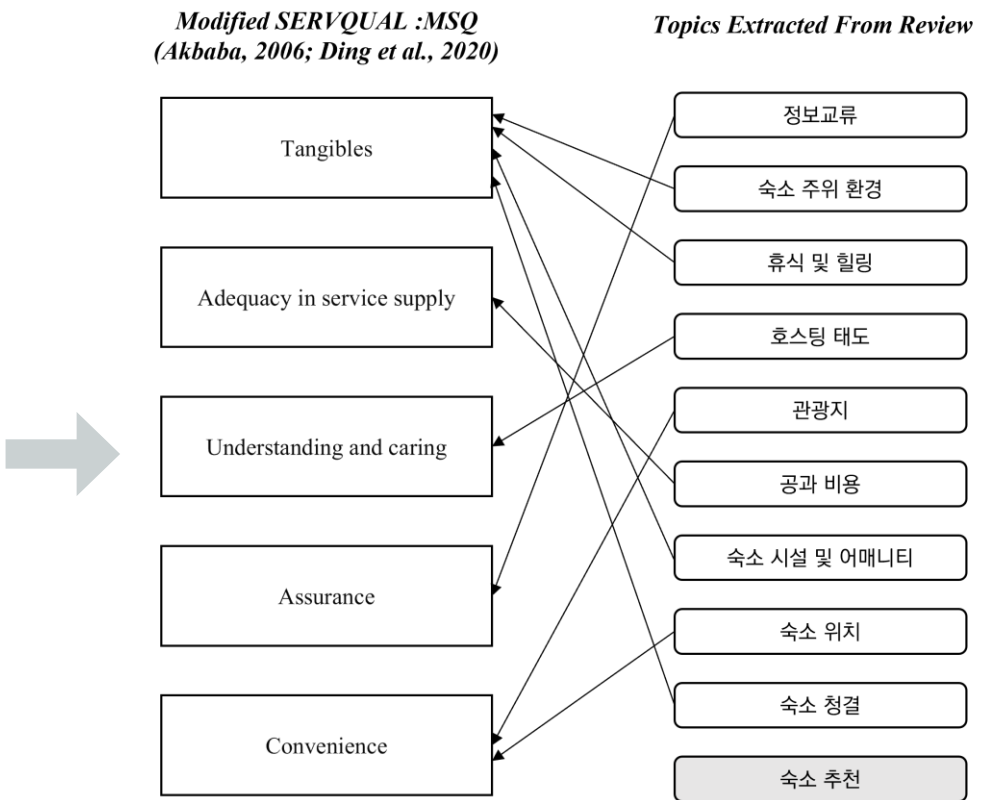
Econometric Model Test

- ✓ Tobit 모델을 활용한 Econometric Testing

III. 연구 방법론 : 토픽모델링(RQ1)

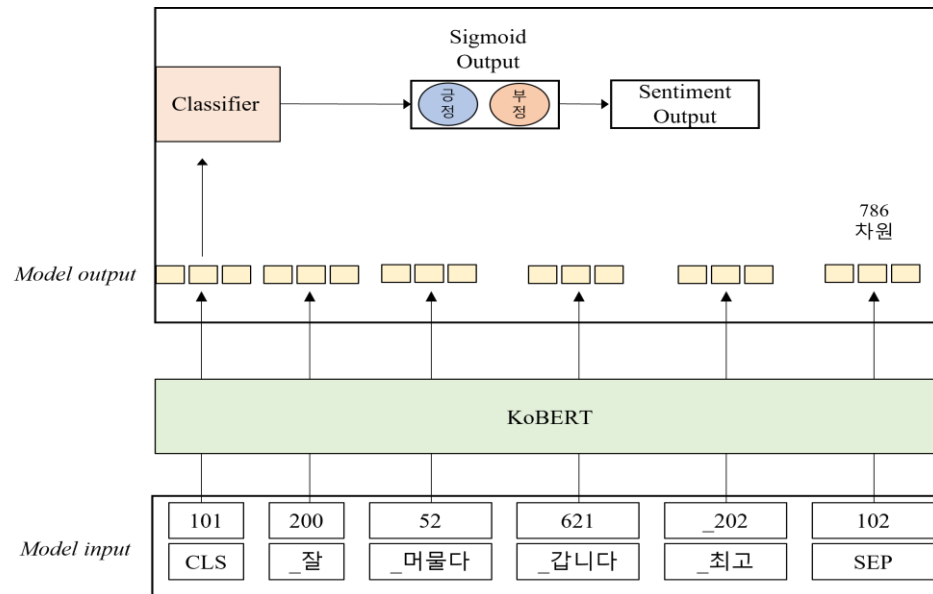
- 온라인 리뷰에서 서비스 품질 속성을 도출하기 위해 Gibbs Sampling for Dirichlet Multinomial Mixture : GSDMM 기법 적용
- GSDMM 모델은 하나의 문서가 하나의 주제와 연관되어 있다고 가정해 짧은 문서에 주제를 도출하기 적합(Lossio-Ventura et al. 2021)
- 리뷰에서는 짧은 텍스트와 긴 텍스트가 혼용되어 있으므로 본 연구에서는 GSDMM을 활용
- 토픽의 개수를 산정하는 방법으로 Top20 Word의 Coherence Score를 활용. 클러스터의 개수가 10개 일 때 Coherence가 높아져(0.384) 10개로 선정

Topic	Topic Label	Relevant Keywords
1	정보교류	사장, 친절, 맛집, 여행, 카페, 직접, 추천, 정보, 안내
2	숙소 주위 환경	바다, 산책, 근처, 거리, 숙소, 주변, 해변, 카페, 조용
3	휴식 및 힐링	숙소, 조용, 지내다, 깨끗, 편안, 힐링, 한적, 만족, 아늑, 따뜻
4	호스팅 태도	친절, 사장, 지내다, 감사, 숙소, 주인, 호스트, 불편, 신경, 배려, 관리
5	관광지	바다, 근처, 맛집, 올레길, 애월, 해수욕장, 한라산, 비자림
6	공과 비용	난방, 따뜻, 사용, 관리, 보일러, 전기, 가스, 공과금, 보증금
7	숙소 시설 및 어매니티	깨끗, 사용, 수건, 세탁기, 침구, 주방, 시설, 구비, 냉장고, 식기
8	숙소 위치	위치, 마트, 숙소, 거리, 편의점, 주변, 근처, 버스 정류장, 버스
9	숙소 청결	화장실, 냄새, 벌레, 침대, 청소, 창문, 바닥, 곰팡이, 환기, 이불
10	숙소 추천	숙소, 여행, 만족, 편안, 가족, 기회, 방문, 추억, 추천, 가성비

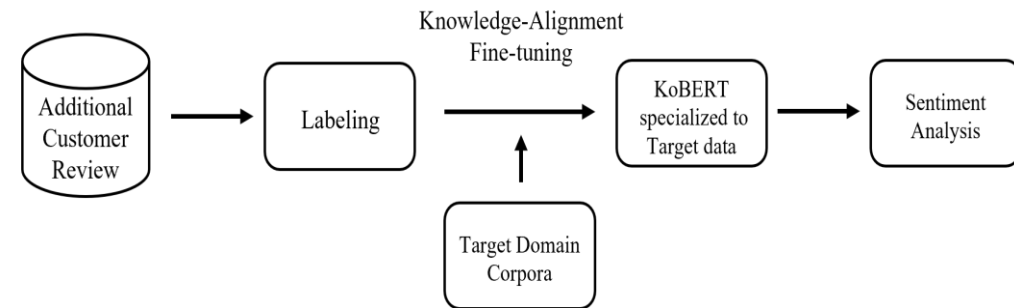


Ⅲ. 연구 방법론 : 감성분석(RQ2)

- 감성분석은 사전 기반과 기계학습 기반의 감성 분석으로 나뉨(Gilbert and Hutto, 2014)
- 사전 기반 분석은 기존에 분류해 놓은 단어 긍·부정 사전을 활용해, 문장의 긍정단어와 부정 단어의 비율로 문서의 감성 정도를 구함
- 사전 기반 분석은 문장의 맥락을 고려하지 못한다는 단점이 존재
- 한국어 사전학습 모델(Pretrained Language Model)인 KoBERT 모델을 활용해 각 서비스 품질 속성에 해당하는 고객 감성 수준을 도출
민박, 공유형 숙소에 대한 리뷰 및 만족도 여부를 Airbnb 및 국내 OTA 서비스에서 추가 수집해 타겟 도메인 미세조정 진행



KoBERT를 활용한 감성 수준 도출



데이터 추가 수집 및 타겟 도메인 미세조정(Fine-tune)

Ⅲ. 연구 방법론 :계량 모델(RQ2)

- 서비스 품질에 대한 감성수준이 고객 만족에 미치는 영향을 확인하기 위해 토빗(Tobit)모델을 활용
- 종속변수인 고객 리뷰의 평점은 1점에서 5점 사이의 상한값과 하한값이 존재하므로 OLS로 추정하게 되면 계수 과소추정 우려 존재(Amemiya, 1984)
- 고객의 여행정보와 관련된 통제변수를 추가함(체류 기간, 동행 인원)

$$\begin{aligned} Rating_{ij} = & \beta_0 + \beta_1 Senti_정보교류_{ij} + \beta_2 Senti_숙소주위환경_{ij} + \beta_3 Senti_휴식및힐링_{ij} + \\ & \beta_4 Senti_호스팅태도_{ij} + \beta_5 Senti_관광지_{ij} + \beta_6 Senti_공과비용_{ij} + \beta_7 Senti_숙소시설및어매니티_{ij} + \\ & \beta_8 Senti_숙소위치_{ij} + \beta_9 Senti_숙소청결_{ij} + \beta_{10} Senti_숙소추천_{ij} + LengthOfStay_{ij} + \\ & NumberOfGuest + \epsilon_{ij} \end{aligned}$$

where

$i = 1, 2, \dots, i^{th} review$

$j = 1, 2, \dots, j^{th} Sharing Accommodation$

* *Rating* – Customer Overall satisfaction

* *Senti* = Sentiment Score

IV. 분석 결과

MSQ Dimension	Variable	Tobit Model 1	Tobit Model 2	Ordered logit regression Model 3
Tangibles	숙소 주위 환경	0.407***(-0.064)	0.408***(-0.062)	0.423***(-0.061)
	휴식 및 힐링	0.362***(-0.053)	0.359***(-0.053)	0.379***(-0.05)
	숙소시설 및 어매니티	0.328***(-0.054)	0.328***(-0.042)	0.345***(-0.042)
	숙소 청결	0.432***(-0.11)	0.430***(-0.044)	0.546***(-0.046)
Adequacy in Service supply	공과 비용	0.447***(-0.095)	0.448***(-0.083)	0.455***(-0.081)
Understanding and caring	호스팅 태도	0.663***(-0.059)	0.656***(-0.054)	0.661***(-0.052)
Assurance	정보 교류	0.205*(-0.103)	0.205***(-0.044)	0.330***(-0.049)
Convenience	관광지	0.038(-0.058)	0.042(-0.06)	0.043(-0.056)
	숙소 위치	0.139*(-0.055)	0.135*(-0.053)	0.129*(-0.05)
-	숙소 추천	0.436***(-0.05)	0.441***(-0.035)	0.401***(-0.033)
	Constant	5.394***(-0.084)	5.423***(-0.138)	-
Control variables	숙박 일수	-	0.009*(-0.004)	0.009*(-0.004)
	숙박 인원 수	-	-0.086**(-0.032)	-0.079**(-0.029)
Log Likelihood		-3,714.51	-3,707.92	-3169.92
Observations		4883	4883	4883

V. 토의 및 시사점 : 토의

(1) 체류형 관광 공유 숙소에서의 주요 서비스 품질 요인 탐색

첫째, 체류형 관광객은 공유 숙박 시설을 이용하며 Tangibles 속성에 대해 높은 관심을 나타냄

리뷰 텍스트 마이닝 결과 숙소 주위 환경, 휴식 및 힐링, 숙소시설 및 어매니티, 숙소 청결 등의 Tangibles 관련 주제어가 가장 많이 도출
체류형 관광객은 숙소에서 오랫동안 머물며 지역을 경험하는 경향을 보인다는 것을 의미(Anantamongkolkul et al., 2019)

둘째, 체류형 관광객은 공유 숙박 시설을 이용하며 Adequacy in Service supply의 “공과 비용”에 대한 관심을 나타냄

체류형 관광에서 오랫동안 한 숙소에 머물면서 숙소에 대한 추가적인 관리비가 발생(김소윤, 2016).

계약 불이행 등 서비스의 신뢰와 관련된 “공과 비용 이슈”는 체류형 공유 숙소에서의 주요 고객 경험 요소 중 하나임을 의미(김송이, 2019)

(2) 서비스 품질과 고객 만족의 관계

첫째, 체류형 공유 숙소에서 Tangible 속성이 고객 만족에 미치는 영향

Tangible 속성에 대한 감성 정도는 모두 숙소의 고객 만족과 정(+)의 영향을 미침.

Tangible 속성이 일반적인 공유 숙박에서 고객 만족에 중요한 역할을 하지 않는다고 밝힌 연구와 차이가 있음 (Priporas et al., 2018)

오랜 기간 숙소에 지역에 머무는 체류형 관광객은 숙박 시설 및 편의 시설의 품질을 중요시한다고 분석할 수 있음 (Lyu et al., 2021)

둘째, 체류형 공유 숙소에서 Adequacy in Service supply 속성이 고객 만족에 미치는 영향

Adequacy in Service supply에 해당하는 “공과 비용”에 대한 감성 정도는 숙소의 고객 만족과 정(+)의 영향을 미침

추가 비용요구 및 계약 불이행 등의 문제를 겪는다는 점을 조사한 선행연구의 주장을 뒷받침(김송이, 2019)

V. 토의 및 시사점 : 시사점

□ 학술적 시사점

- 체류형 관광분야에서 Modified SERVQUAL의 개념을 적용해 고객 숙박 만족에 영향을 미치는 서비스 품질 속성을 도출하여 기존 연구 범위를 확장
- 표본조사에 그치지 않고 실제 체류형 관광객의 리뷰를 바탕으로 텍스트 마이닝 기법을 활용하여 고객 숙박 만족 요인 분석
- 체류형 관광 고객의 리뷰의 감성 수준을 도출하기 위해 전통적인 사전 기반 대신 딥러닝 기법을 활용한 연구라는 점에서 차별점 존재

□ 실무적 시사점

- 본 연구의 분석 결과를 바탕으로 체류형 관광 분야의 서비스 품질 속성 중 집중해야 할 지표에 대한 기준 제시
- 공유 성과를 극대화하고 싶은 호스트에게도 숙박 공유 전략에 대한 실질적인 조언을 제공

□한계점 및 향후 연구 방향

- 체류형 관광의 기준인 체류 기간을 여행 정보 변수로 활용했지만, 체류 기간에 따라 변화하는 서비스 품질의 차이를 확인하지 못함
- 체류 기간에 따른 각 서비스 품질의 조절 효과에 대한 비교 및 분석으로 향후 연구 범위 확장
- 토픽 도출과 MSQ 차원의 연결에 객관성 확보를 위해 전문가와 교차검증을 진행 한다면 결과를 보다 객관적으로 확인할 수 있음

Thank YOU

Appendix 1

□ Modified SERVQUAL Dimensions 및 해당 속성

Dimension	Attribute
Tangibles	1. The property has visually appealing buildings and facilities 2. The service units of the property have adequate capacity (dining rooms, meeting rooms, swimming pools, business center facilities, etc.) 3. The property has modern-looking equipment (air conditioner, furniture, elevator, communication devices, etc.) 4. The atmosphere and equipment are comfortable and appropriate for purpose of stay (beds, chairs, rooms, etc. comfortable, clean, and tranquil) 5. The equipment of the property works properly without causing breakdowns 6. Materials associated with the services are adequate and sufficient (soap, shampoo, towel, etc.) 7. Food and beverages served are hygienic, adequate, and sufficient
Adequacy in Service supply	8. The Airbnb host provides prompt service 9. The Airbnb host provides the services at the time it promises to do so 10. The Airbnb host is always available when needed 11. The Airbnb host keeps accurate records (reservations, guest records, bills, orders, etc.) 12. The Airbnb host resolves guest complaints and compensate for the inconveniences guests go through
Understanding and caring	13. The Airbnb host provides flexibility in services according to guest demands 14. The Airbnb host always treat guests in a friendly manner 15. The Airbnb host understands the specific needs of guests
Assurance	16. The property and its facilities have operating hours convenient to all their guests.
Convenience	17. It is easy to access to the property (transportation, car parking area, etc.) 18. Getting information about the facilities and services of the property is easy (reaching information via phone, internet, etc., direction signs, etc.)

Note: Adopted from Akbaba (2006). Ding et al.(2021)

Appendix 2

□ 변수 기초 통계

	별점	공과 비용	숙소 주위환경	숙소시설 및 어매니티	숙소 청결	정보 교류	관광지	휴식 및 힐링	숙소 추천	숙소 위치	호스팅 태도	Year	NumOf Guest	Length OfStay
mean	4.62	-0.10	0.31	0.14	-0.16	0.14	0.29	0.42	0.91	0.30	0.52	2020.43	2.34	19.79
std	0.81	0.88	0.72	0.90	0.95	0.56	0.66	0.77	1.21	0.73	0.80	1.19	1.20	9.13
min	1.00	-17.28	-3.99	-10.96	-22.25	-3.99	-3.72	-3.22	-9.11	-4.68	-5.51	2017.00	1.00	7.00
25%	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2020.00	2.00	14.00
50%	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	2021.00	2.00	17.00
75%	5.00	0.00	0.97	0.00	0.00	0.00	0.70	1.00	1.98	0.97	1.00	2021.00	3.00	29.00
max	5.00	3.99	8.95	4.54	2.99	7.94	6.67	5.97	10.93	4.98	4.98	2022.00	13.00	82.00

Appendix 2

□ 변수 기초 통계 (상관 관계)

	별점	공과 비용	숙소 주위 환경	숙소 시설 및 어매니티	숙소 청결	정보 교류	관광지	휴식 및 힐링	숙소 추천	숙소 위치	호스팅 태도	Year	NumOfGuest	LengthOfStay
별점	1.00													
공과 비용	0.37	1.00												
숙소 주위 환경	0.16	0.09	1.00											
숙소 시설 및 어매니티	0.30	0.27	0.09	1.00										
숙소 청결	0.40	0.45	0.07	0.30	1.00									
정보 교류	0.16	0.16	0.09	0.08	0.12	1.00								
관광지	0.07	0.06	0.06	0.06	0.05	0.03	1.00							
휴식 및 힐링	0.16	0.11	0.03	0.02	0.10	0.01	0.01	1.00						
숙소 추천	0.30	0.15	0.10	0.16	0.17	0.11	0.06	0.01	1.00					
숙소 위치	0.10	0.07	0.00	0.07	0.10	0.01	0.04	-0.01	0.03	1.00				
호스팅 태도	0.27	0.13	0.10	0.08	0.16	0.10	0.02	0.03	0.12	0.01	1.00			
Year	-0.12	-0.04	0.05	-0.01	-0.07	0.01	0.06	-0.03	0.03	-0.02	-0.03	1.00		
NumOfGuest	-0.02	-0.01	0.01	0.01	-0.02	-0.00	0.00	-0.02	0.04	-0.03	-0.02	0.04	1.00	
LengthOfStay	0.03	-0.02	0.03	-0.01	-0.02	-0.02	0.01	0.01	-0.02	0.02	0.05	-0.04	-0.13	1.00