



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

경영학박사 학위논문

숙박공유서비스의 온라인 리뷰를 활용한 전반적 만족도 결정 요인 분석

- 에어비앤비(Airbnb) 호스트 등급에 따른 조절효과를 중심으로 -



2020년 8월

부경대학교 대학원

경영컨설팅협동과정

권 혜 진

경영학박사 학위논문

숙박공유서비스의 온라인 리뷰를 활용한 전반적 만족도 결정 요인 분석

- 에어비앤비(Airbnb) 호스트 등급에 따른 조절효과를 중심으로 -

지도교수 전 재 균

이 논문을 경영학박사 학위논문으로 제출함

2020년 8월

부경대학교 대학원

경영컨설팅협동과정

권 혜 진

권혜진의 경영학박사 학위논문을 인준함.

2020년 8월

위원장 경영학 박사 배 상 욱



위원 관광레저학 박사 양 위 주



위원 호텔외식경영학 박사 김 학 선



위원 경영학 박사 이 준 혁



위원 관광경영학 박사 전 재 균



< 목 차 >

제 1 장 서 론	1
제 1 절 연구의 배경 및 문제제기	1
제 2 절 연구의 목적 및 필요성	6
제 3 절 연구의 내용 및 범위	9
제 4 절 연구의 구성 및 체계	12
제 2 장 이론적 배경	14
제 1 절 숙박공유서비스 에어비앤비	14
1. 에어비앤비	14
2. 호스트 등급(일반 호스트 vs 슈퍼 호스트)	18
3. 에어비앤비 선택 속성	20
4. 에어비앤비 관련 선행연구	21
제 2 절 전반적 만족도	24
1. 전반적 만족도(평점)의 개념	25
2. 전반적 만족도 관련 선행연구	27
제 3 절 온라인 리뷰	29
1. 온라인 리뷰	29
2. 온라인 리뷰 관련 선행연구	31
제 4 절 텍스트 마이닝	35
1. 텍스트 마이닝	35
2. 텍스트 마이닝 기법 종류	37

제 3 장 연구 조사 설계	45
제 1 절 연구 절차 및 방법	45
제 2 절 자료수집 대상 선정	51
제 3 절 자료수집 및 전처리	53
1. 웹 크롤러(web crawler)를 통한 자료수집	53
2. 자료 전처리(data preprocessing)	56
제 4 장 연구 결과	61
제 1 절 에어비앤비 분포 특성	61
1. 데이터 분포 특성	61
제 2 절 연구 과제 분석	64
1. 언급 빈도 분석	65
2. 워드 클라우드	68
3. 토픽 모델링	71
4. 감성분석	73
5. 다중회귀분석	77
6. 위계적 회귀분석	78
제 5 장 결 론	83
제 1 절 연구의 요약	83
제 2 절 연구의 시사점	85
1. 이론적 시사점	85
2. 실무적 시사점	88
제 3 절 연구의 한계 및 향후 연구 방향	91
참 고 문 헌	93
부 록	111

< 표 목 차 >

<표 2-1> 에어비앤비 국내·외 파급 현황	17
<표 3-1> 연구 절차의 요약	50
<표 3-2> 측정변수의 정의	59
<표 4-1> 에어비앤비 데이터 수집 결과	61
<표 4-2> 상위 100위 언급 빈도 분석 결과	66
<표 4-3> 연관 키워드 분류	67
<표 4-4> 사용자 기반 에어비앤비 온라인 리뷰 감성사전	75
<표 5-1> 숙박선택 속성이 전반적 만족도에 미치는 영향 분석	77
<표 5-2> 호스트 등급의 조절효과에 대한 위계적 회귀분석	79

< 그림 목 차 >

<그림 1-1> 연구구성 흐름도	13
<그림 2-1> 에어비앤비 비즈니스 모델	15
<그림 2-2> LDA 토픽 모델링 모형	41
<그림 3-1> 온라인 리뷰 분석 연구 절차	47
<그림 3-2> 연구 흐름에 따른 목적 및 방법	49
<그림 3-3> 에어비앤비 리뷰 및 평점 형태	52
<그림 3-4> 에어비앤비 웹 크롤링 수집 샘플 데이터	54
<그림 3-5> 최초 수집된 raw data 형태	55
<그림 3-6> 데이터 분석 전 최종 데이터 구성	56
<그림 3-7> 에어비앤비 한국어 텍스트 형태소 분석 과정	57
<그림 3-8> ‘말뭉치(corpus)’ 데이터로 구조 변환 R 함수식	58
<그림 3-9> 수치 데이터 파일	60
<그림 4-1> 나무지도(tree map)분석 과정	62
<그림 4-2> 나무지도를 활용한 서울 분포표	63
<그림 4-3> 리스트(list) 형태로 분리하는 과정	65
<그림 4-4> 온라인 리뷰 전체 워드 클라우드 결과	69
<그림 4-5> 일반 호스트 워드 클라우드 결과	70
<그림 4-6> 슈퍼 호스트 워드 클라우드 결과	70
<그림 4-7> 토픽 모델링 분석 과정	71
<그림 4-8> 토픽 모델링 분석 결과	73
<그림 4-9> 감성분석 결과	76

<그림 5-1> 청결성과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과	80
<그림 5-2> 소통과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과	81
<그림 5-3> 체크인과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과	82



Analysis of Factors Affecting on Overall Satisfaction of Accommodation Sharing Service through Online Review -Moderating Effect according to Host Grade of Airbnb-

Hye-Jin Kwon

Department of Business Administration
Graduate School of Pukyong National University

Abstract

With the Fourth Industrial Revolution and the rapid development of new technologies, the environment is changing rapidly, and the online market is growing. Consumers use the Internet and mobile platforms to find products or services they need. The culture of exchanging surplus resources with neighbors in the past has been expanded to exchange resources and services with strangers due to the development of information technology. It created economic benefits by utilizing resources by sharing unused or less-used products or services with others. As these new forms of trading increased, the sharing economy began to gain attention.

Airbnb is the most representative business of P2P sharing economy in the area of accommodation between individuals. The Airbnb is a lodging platform company that provides a peer-to-peer (P2P) service

that connects hosts and tenants directly so that hosts can share or rent their own empty rooms or houses. The service method features a form of community platform that allows people to share and book accommodation around the world on online, mobile and tablet.

According to the Airbnb, the economic effects of Korean Airbnb market have resulted in 1.37 trillion won in 2018. In 2017, 1.888 million guests visited, of which 1.23 million Koreans were found to be using the Airbnb, 65% of the total. As the use of Koreans increases, it is time to study Koreans. Also, the subject of the study is timely at a time when COVID-19 is now turning to domestic travel rather than overseas travel.

Recent prior studies targeting the Airbnb locals can be seen, and there are limitations in research by following existing survey research methods or by conducting research on specific topics. The overall lack of exploratory approaches to Koreans makes this study absolutely necessary.

Online reviews left by customers using Airbnb themselves have various meanings.

Research is needed to analyze online reviews and ratings by showing their needs frankly. Existing prior studies related to the Airbnb have also been conducted from various perspectives. However, it is time to actively analyze real-time data in this rapidly changing era to discover the inherent meaning and study the perception of Koreans about the Airbnb. Therefore, this study analyzed the meaning of the reviews recorded online by the Airbnb customers to derive meaningful results

and also analyzed the relationship between overall satisfaction level according to customer ratings. Text mining techniques for reviews were used to identify these relationships.

In order to achieve the research purpose, it has been explored the Airbnb selective attributes, online reviews and overall customer satisfaction through domestic and international literature review, and the literature related to web crawling, text mining techniques, statistical techniques, etc.

First of all, the Airbnb online review was selected and analyzed to collect data that would be the most important basic data of this study. Online reviews of host accommodations in Seoul were collected on Airbnb's official site. Various text mining techniques were used to collect reviews, ratings, and prices left by customers using selected hosts. Using online review data, frequency analysis of the mentioned words was conducted and the meaning of the words (keywords) was understood through word cloud visualization of the derived top words. The online review document was then converted into a single document to derive the topic between the document and the word to confirm the meaning of the entire document. Finally, after the researcher identified the entire online review, a user-based emotional dictionary was established, and the percentage of positives and negatives in the entire document was identified by the researcher. Through the meaning of the exploratory results identified here, the Airbnb selective attribute factors determined by the researchers were selected and the relationship with overall satisfaction was confirmed.

To collect the data needed for this study, the data were collected using deep learning techniques, and R and SPSS25.0 were used for text mining and statistical analysis.

The analysis data examined regional distribution characteristics with 281 local data, and conducted text mining analysis and statistical analysis with 1,545 review data.

The results of the study on the research tasks presented in this study are summarized as follows:

First, after checking the characteristics of the regional distribution, it was confirmed that many accommodations were distributed around the area where tourist attractions were located. Second, the exploratory study of online reviews first identified keywords associated with the top keyword in the reference frequency analysis, and the difference in elicitation keywords between 'general host' and 'super host' through the top frequency keyword was identified. The topic modeling analysis then extracted the six topics (topic) of the overall review of the Airbnb to find the underlying meaning and, as a result of the sentiment analysis, identified the probability of positivity and negation of the entire review document. For the sentiment analysis, the researcher established a user-based sentiment dictionary, which resulted in 62% of the accuracy of the sentiment analysis, which proved to be significant.

As a result of eliciting the meaning of online review data through various text mining analysis methods, we can see that online reviews contain a lot of customer opinions. The review analysis confirmed that the limitations that ratings alone cannot explain all the implications

should be assessed and re-elected.

Third, 'clean', 'communication' and 'smooth check-in' in the relationship of the host class (general host vs. super host) between the select property of the Airbnb accommodation and the overall satisfaction. The following results of host-level control effectiveness verification show that the general host affects overall satisfaction depending on the presence or absence of 'cleanness', 'communication', and 'round check-in'. Through this, the host needs to maintain cleanliness, prompt feedback with the guest, and smooth check-in.

Through text mining techniques and statistical analysis, this study has shown that the Airbnb online review is very important information that can grasp the customer's sensibility and opinions, which is deeply related to overall satisfaction.

Therefore, online reviews of various opinions and emotions have been shown to be important information for the Airbnb, and the scientific methods provided through this study will be the basis for further extended research.

Key Words: Airbnb, Online Review, Text Mining, Host Grade, Rating, Hierarchical Regression

제 1 장 서 론

제 1 절 연구의 배경 및 문제제기

신기술의 급속한 발전으로 우리를 둘러싼 모든 환경이 과거 어느 때보다 빠르게 변화하고 있다. 과거의 시장이 대면 접촉에 의한 오프라인 시장이 대부분이었다면 현재에는 유무선 네트워크에 의한 온라인 시장이 급속하게 성장하고 있다(최유성·안혁근·박정원·심우현, 2017). 최근 소비자들은 인터넷, 모바일 플랫폼을 활용해 자신이 필요한 상품이나 서비스를 찾는 추세이다. 과거 이웃들과 남는 자원을 교환하던 문화가 정보기술(IT: information technology)발전으로 낯선 사람과의 자원 및 서비스 교환으로 확장되었다. 사용하지 않거나 적게 쓰는 제품 또는 서비스 등을 다른 사람과 공유하고 자원을 활용함으로써 경제적 이익을 창출하게 된 것이다. 이러한 새로운 형태의 거래가 증가하면서 공유경제(sharing economy)가 주목받기 시작하였다(김지희·이건웅, 2019). 그 중에서도 에어비앤비(airbnb)는 여행산업의 성장과 함께 현지국의 주거 및 생활 문화를 직접 체험할 수 있는 특성으로 세계적으로 성장하고 있다. 최근 에어비앤비에서는 호스트의 신분확인이나 전문성에 따른 등급 구분을 두어 안정적인 고객 유치를 위한 노력을 하고 있다. 예를 들어, 공식적으로 입증된 호스트의 신분, 호스트 등급 등의 정보와 같이 이용고객의 에어비앤비 숙소 선택 의사결정에 도움을 주기 위한 서비스들을 지속적으로 추가하고 있다(이의준·원형식·이새롬, 2020).

최근 온라인과 모바일 환경이 급속하게 발전함에 따라 수많은 정보탐색과 소통이 온라인 모바일을 통해 이루어지고 있다. 온라인 환경에서의 구

전활동은 SNS, 개인 블로그(blog), 고객 게시판, 전문가 및 고객 리뷰를 통해 급속도로 퍼져가고 있으며 특히, SNS상의 구전 활동은 과거 오프라인 구전에 비해 훨씬 강력하게 소비자의 정보 탐색, 구매 결정, 기업 이미지 형성에 영향을 주고 있다. 기업들도 이러한 SNS상의 온라인 구전이 소비자들에게 미치는 영향력을 인식하여 소비자로부터 SNS를 활용하여 자발적인 입소문과 추천을 하도록 적극적인 마케팅 활동을 전개하고 있다(최영민·배상욱, 2016). 한편, 최영민·배상욱(2016)은 온라인상에서 이루어지는 구전 정보의 특성이 구전효과에 영향이 있는지를 실증분석하였는데, 구전 정보가 생생하고, 부정적이고, 사실적일수록 정보수용 의도에 긍정적인 영향을 미치는 것을 확인했다. 이는 온라인상의 리뷰 텍스트들이 영향력을 가지고 있다는 것을 입증한 결과라 할 수 있다.

온라인 리뷰와 관련된 호텔 관광 분야의 선행연구에서는 예비 고객들은 객실 예약을 온라인으로 할 때, 가장 많이 영향을 받는 부분은 호텔을 경험한 고객의 리뷰와 평점이다(Li, & Hitt, 2008)라고 하여 온라인 리뷰와 평점의 영향력에 대한 결과를 입증하였고, 온라인 리뷰나 평점 등 온라인 정보는 예비 고객들에게 지원 도구가 되고 있다고 했다(Wang, & Fesenmaier, 2004). 온라인 리뷰를 이용한 연구는 예전부터 꾸준히 진행되어 왔다(Bickart, & Schindler, 2001; Li, & Hitt, 2008; Crotts, Mason, & Davis, 2009; Menner, Höpken, Fuchs, & Lexhagen, 2016; 도해용·류기상, 2014; 반현정·김학선, 2019). 그리고 에어비앤비 온라인 리뷰를 활용하여 진행한 연구들도 있다(Guttentag, 2015; Bridges, & Vásquez, 2018; Cheng, & Jin, 2019; Lawani, Reed, Mark, & Zheng, 2019; 류일한·주재현·김동연·안재현, 2017). 하지만, 기존 선행연구들은 영문 리뷰를 중심으로 진행되거나 해외 지역을 중심으로 연구가 진행되어왔다. 아직 국문 리뷰나 국내 지역을 중심으로 한 연구들은 미흡한 실정이다.

에어비앤비 숙박서비스는 제공하는 공급자가 전문적인 숙박업체가 아니라 유희 주거공간을 확보한 개인(호스트)이기 때문에 서비스 공급자가 불특정 다수(호스트들)라는 점에서 기존 숙박 예약서비스와 차이점을 가지는데(Guttentag, 2015), 국내 선행연구의 경우 이 호스트가 가지는 특성, 호스트의 신뢰, 호스트 피드백 등이 구매 의도 또는 재이용 의도에 미치는 영향에 관한 연구가 진행되어왔다(류일한 등, 2017; 권혁진·김진후·이충훈, 2017). 그리고 호텔서비스와 에어비앤비 숙박서비스의 차이점을 검증하는 연구도 진행되었고(김진한·강우성, 2018), 소비자 행동을 중심으로 에어비앤비 이용자 연구도 있었다(안성숙·양수진, 2018). 반면 에어비앤비의 부정적인 요소인 위험요소에 관한 연구도 꾸준히 진행되어 오고 있는데, 이혜미(2018)는 지각된 위험 인식 연구를 통해 에어비앤비의 문제점에 대해 지적하였고, 박홍식·김주일(2019)은 에어비앤비 이슈 발생 전후 감성분석을 실시하여 텍스트의 변화를 확인하였다. 최근에는 관광 분야뿐만 아니라, 입지 특성(박소민·양승우, 2020)과 공간 분포(허지정·노승철, 2018)를 서울시를 대상으로 분석하였는데, 서울지역의 경우 에어비앤비 숙박서비스 호스트(숙소)가 많이 분포되어있는 지역은 주요 관광지가 있는 곳으로 확인되었다. 이 결과는 본 연구의 대상인 서울지역의 호스트(숙소) 지역분포를 확인함으로써 선행연구와의 차이점이나 공통점을 검증할 수 있을 것이라 판단된다.

지난 5월 ‘제5차 국가관광전략회의’에서 정세균 국무총리는 국내 여행 활성화를 위해 다양한 계획을 발표하였는데 관광 내수시장 활성화 대책으로 에어비앤비 ‘외국인관광 도시민박업’의 규제개혁에 대해 언급하였다(손덕호, 2020). 전 세계는 ‘코로나19’로 해외를 이동하는 것이 자유롭지 못하게 되었으나, 국내는 사회적 거리두기가 완화되면서 국내 여행을 통해 지친 마음을 달래길 원하는 여행객들이 많아지기 시작했다. 이렇게 국내 여행에

대한 관심이 높아지는 지금의 시점이야말로 내국인을 대상으로 하는 연구는 아주 시기적절하다고 판단된다. 언론인과 관광 분야 학자들을 대상으로 한 ‘전문가 포커스 그룹 설문조사’에서 에어비앤비의 숙소 활용에 대해 80~90%가 빈집을 활용할 수 있게 허용해야 한다 했고, 외국인 입국이 제한되어있는 지금 ‘외국인관광 도시민박업’에 따른 내국인 규제 정책도 완화해야 한다는 목소리가 나오고 있다(에어비앤비, 2020a). 에어비앤비의 국내 경제효과는 2018년 1년 동안 1조 3,700억 원의 효과를 가져왔고(에어비앤비, 2019), 2017년도 국내 에어비앤비 숙소서비스를 이용한 고객들을 기준으로 전체 방문 게스트 중 내국인이 123만 1,500명으로 65%를 차지했다(안갑성, 2018). 이런 성장세에도 불구하고 에어비앤비 숙박서비스에 대한 국내 연구는 아직 미흡하고 내국인을 대상으로 다양한 연구가 필요하다고 할 수 있다. 코로나19로 전 세계 관광산업은 많은 어려움을 겪고 있다. 하지만 최근 에어비앤비 숙박서비스의 경우 6월 5일~7일 숙박 예약액이 지난 2월 이후 처음 성장세를 보였고, 국내에서의 숙박 예약 역시 빠른 회복세를 보이고 있다(에어비앤비, 2020b). 내수시장의 활성화를 위한 이 시대의 흐름에 맞는 내국인의 니즈와 의견을 통해 마케팅적 시사점을 제언하기 위해 본 연구에서는 온라인 리뷰를 활용하여 분석하였고, 에어비앤비 숙박서비스 선택 속성을 파악하고자 하였다.

호텔 관광 분야와 관련된 선행연구에서는 선택 속성에 대해 살펴보는 연구들을 많이 찾아볼 수 있다(이광국·김경옥·전재균, 2016; 김도경·김인신, 2017; 광민정·최지유·박소현, 2019). 에어비앤비의 경우 최근 연구가 진행되어 오고 있는데, 에어비앤비의 선택 속성에 관한 연구로 김상태·Tsolmon·김남조(2017)는 웹 사이트에서 ‘숙박가격’, ‘후기수’, ‘인테리어 수준’에 관한 데이터를 수집해 독립변수로 설정하고 예약률을 종속변수로 하여 분석하였다. 정은별·이형룡(2020)은 국내 에어비앤비를 이용하는 내국인 고객의 숙

박시설 선택 속성의 중요도에 대한 연구를 진행하였고, 선택 속성은 에어비앤비 웹 사이트를 통해 12개를 도출하여 분석한 결과, 가격이 가장 중요도가 높았으며 그 다음으로 ‘청결’, ‘위치’, ‘정확성’ 등으로 나타났다. 그리하여 본 연구에서는 선행연구들을 바탕으로 에어비앤비 숙박서비스 선택 속성을 도출하고자 하였다.

본 연구는 기존 국내 에어비앤비 숙박서비스 선행연구에서 진행하지 않았던 빅데이터의 텍스트 마이닝 기법들을 통해 리뷰 텍스트를 분석하여 의미 있는 선택 속성을 도출하였고 선행연구 고찰을 바탕으로 전반적 만족도에 미치는 영향을 호스트 등급을 조절효과로 차이 검증하여 에어비앤비 숙박서비스의 활성화 방안을 마련하는데 도움을 줄 수 있는 계기를 마련하였다.



제 2 절 연구의 목적 및 필요성

본 연구는 숙박공유서비스 대표 플랫폼인 에어비앤비를 이용한 내국인들의 온라인 리뷰를 수집하여 한국어 텍스트를 중심으로 텍스트 마이닝 기법인 언급 빈도 분석, 워드 클라우드, 토픽 모델링 그리고 감성분석을 통해 텍스트에 숨겨져 있는 의미를 찾아내어 고객의 욕구를 파악하고 에어비앤비 숙박서비스에 대한 전반적 만족도의 중요한 핵심을 찾아내고자 하고 에어비앤비 서비스 개선 및 마케팅 전략 수립에 대한 기초자료를 마련하는 것이 본 연구의 목적이다.

에어비앤비의 선행연구를 고찰하면 해외의 경우 그 나라의 대표지역을 대상으로 에어비앤비를 선택하는 요인들에 대해 연구가 진행되어왔다(Ert, Fleischer, & Magen, 2016; Lawani et al, 2019; Cheng, & Jin, 2019). 국내 경우에도 대표지역인 서울을 중심으로 연구가 진행 되었는데(허지정·노승철, 2018; 박소민·양승우, 2020), 아직은 관광 분야에서 서울지역을 대상으로 내국인 관광객들의 니즈에 대해 많은 자료를 가지고 있지는 못하며 연구도 부족한 실정이다. 이에 에어비앤비는 본 연구의 다양한 분석 기법을 통해 고객의 온라인 리뷰에 대한 분석으로 어떠한 요인들이 전반적 만족도에 영향을 미치는지를 파악할 수 있다. 그러므로 본 연구는 내국인 관광객들이 에어비앤비 숙소서비스 이용에 있어 만족도를 높이고 지속적인 이용이 가능할 수 있도록 마케팅에 도움이 되는 에어비앤비 숙박선택 속성에 대해 고찰하고 이를 위해 온라인 리뷰 및 평점, 가격(숙박료), 숙소 지역 등의 유의한 데이터를 추출하여 분석하는 빅데이터 분석 방법을 통해 시사점을 제공하였다.

기존 연구에서는 대부분 해외사례 연구나 영문 리뷰 중심으로 연구가 진

행되어왔고, 내국인을 대상으로 국문 리뷰를 연구한 사례는 부족한 실정이다. 기존 선행연구에서는 밝힐 수 없었던 보다 깊이 있는 분석을 통해 고객들이 필요로 하는 요소들이 무엇인지 인식할 필요가 있다. 온라인 리뷰 데이터에는 대상자를 평가함은 물론 고객의 이용 경험과 인지된 상황 및 인식 경험 등을 포함하고 있다(심영석·김홍범, 2016). 그러므로 이를 수집하고 구조화하여 다양한 분석 방법을 통해 고객들이 인식하고 느낀 평가를 탐색적으로 규명하는 연구를 실시하여 내국인 관광객들이 에어비앤비 숙박서비스를 이용함에 불편함이나 개선할 점들을 파악하는 계기를 마련하고 고객을 이해하는 데에 도움이 되는 연구를 하고자 하였다. 그리고 기존 연구들은 정성적 분석 또는 정량적 분석 어느 하나를 집중적으로 연구하였다면 본 연구는 온라인 리뷰 텍스트에서 도출된 키워드를 정성적 분석 기법으로 결과를 도출하고, 보다 다양한 의미의 결과값을 찾아내기 위해 확장된 연구방식으로 데이터를 정량적으로 변환시켜 통계분석을 실시하였는데, 호스트 등급별 조절효과가 전반적 만족도에 영향력 있는 요인들 간의 관계를 검증하고자 하였다.

연구 배경에서 제시된 연구의 필요성을 충족시키기 위해 본 연구에서는 다음의 연구 목적으로 연구를 진행한다.

첫째, 온라인 리뷰에서 수집한 숙소 지역 데이터를 활용하여 서울지역의 호스트 숙소 분포표를 확인하여 내포되어있는 의미를 파악한다.

둘째, 온라인 리뷰 텍스트에 숨겨져 있는 의미를 도출하기 위해 다양한 텍스트 마이닝 기법을 활용하여 에어비앤비 숙박서비스를 이용한 내국인 고객들의 인식에 대해 고찰한다.

셋째, 텍스트 마이닝 기법들을 통해 도출된 에어비앤비 숙박서비스 결정(선택) 요인들과 전반적 만족도 간의 영향관계를 검증한다.

넷째, 에어비앤비 숙박서비스만의 특징인 호스트 등급에 따른 조절효과

를 분석하기 위해 위계적 회귀분석을 실시하여 차이를 검증한다.

이와 같이 에어비앤비를 이용하는 결정 요인을 파악하는 것은 내국인 관광객이 에어비앤비를 선택할 때 영향을 미치는 요인으로서 효과적으로 에어비앤비를 마케팅할 수 있는 수단이 될 것이다. 또한, 분석의 대상을 내국인 관광객으로 설정하였기 때문에 내국인 활성화 방안을 마련하는 중요한 연구자료로 사용될 것이다. 그리고 관광 분야의 연구들은 조사 목적을 고려하지 않고 대부분 연구들이 설문조사에 의한 계량적 연구방법들 위주로 진행되어왔지만, 본 연구에서는 방대한 규모의 정보가 쏟아져 나오는 온라인 환경에서 대표적인 커뮤니케이션 수단인 SNS(social networking service)나 온라인 리뷰 데이터를 분석하기에 적합한 기술적 방법인 텍스트 마이닝 기법을 온라인 리뷰 데이터에 적용한 탐색적 연구를 실시함으로써 향후 관광산업 연구에 또 다른 방법론적 접근을 제시할 수 있을 것이다.

제 3 절 연구의 내용 및 범위

본 연구를 진행하기 위하여 선행된 정성적 연구와 정량적 연구를 병행하였다. 정성적 연구에서는 국내·외 문헌을 중심으로 호텔 선택 속성과 에어비앤비 숙박선택 속성을 비교 고찰하고, 호스트 등급, 온라인 리뷰, 전반적 만족도, 텍스트 마이닝 등에 관련된 선행연구를 탐구하고 구성개념 간 관계들을 고찰하였고, 공학적 접근을 위해 웹 크롤링, R을 활용한 텍스트 마이닝 기법들에 관련된 문헌 고찰도 하였다. 이러한 선행연구 고찰을 통하여 연구 과제를 설정하고, 이를 분석하기 위하여 자료수집과 분석을 실시하였다.

본 연구는 선행연구 고찰, 에어비앤비 숙박서비스 정보, 호스트 정보, 숙소 지역 정보, 가격(숙박료) 및 온라인 리뷰 데이터 수집, 리뷰에 대한 언급 빈도 분석, 호스트 등급별 워드 클라우드 시각화, 리뷰 전체 문서의 주제 분류 토픽 모델링, 사용자 기반 감성사전 구축과 긍정·부정 비율 검증의 감성분석, 전반적 만족도 영향관계 검증의 다중회귀분석 및 요인들의 호스트 등급별 조절효과 검증을 위해 위계적 회귀분석을 진행하였다. 단계별 연구 내용은 다음과 같다.

첫 번째, 에어비앤비 숙박서비스 서울지역 숙소 분포 특성을 확인하기 위해 R을 사용하였고 숙소 지역 분포를 시각화하여 검증하였다. 두 번째, 기존 계량적 설문지 방식을 통해서 연구되었던 선택 속성에 대해 탐색적으로 접근하여 결과를 검증하였다. 세 번째, 방법론적 접근을 위하여 본 연구에서는 최근 빅데이터 분석 기법으로 많이 활용되고 있는 텍스트 마이닝 기법들인 언급 빈도 분석, 워드 클라우드, 토픽 모델링과 감성분석의 개념과 이들을 이용한 선행연구를 고찰하였다. 네 번째, 에어비앤비 내국인 고

객들이 경험하고 남긴 온라인 리뷰를 수집하기 위해 웹 크롤러를 개발하여 데이터를 수집하였다. 데이터는 서울지역의 임의의 기간을 예약일로 선정 후 노출된 숙소 308개를 수집하고, 데이터 정제 과정을 거쳐 본 연구에 활용하기 위한 구조형태로 변환하였다. 다섯 번째, 구조화된 텍스트 데이터를 형태소 분석을 거쳐 리뷰에 언급된 빈도수에 따른 영향력을 확인하기 위해 언급 빈도 분석을 실시하였다. 여섯 번째, 도출된 상위 빈도 단어(키워드) 워드 클라우드를 활용하여 호스트 등급별로 분류하여 시각화한 후 비교 분석하였다. 일곱 번째, 문서를 말뭉치(corpus) 형태로 변환하고 문서×단어 형태의 행렬(dtm)형태로 변환시켜 전체 리뷰 문서의 주제를 분류하여 잠재되어있는 의미를 도출하는 토픽 모델링을 실시하였다. 여덟 번째, 전체 리뷰 문서에 사용자 기반 감성사전을 구축하여 감성분석을 실시하였고, 결과로 전체 문서의 긍정·부정 비율을 검증하였다. 아홉 번째, 다양한 텍스트 마이닝의 결과에 도출된 의미와 고찰된 선행연구를 바탕으로 영향력이 있는 요인들을 선정하였고, 그 요인들과 전반적 만족도 간의 영향관계를 다중회귀분석을 통해 검증하였다. 열 번째, 호스트 등급별 조절효과로 전반적 만족도 차이 검증을 실시하였다.

본 연구의 범위는 다음과 같다.

첫째, 모집단은 서울지역 내 에어비앤비 숙박서비스 호스트들의 상품을 이용한 내국인이며 표본집단은 에어비앤비 서울지역의 호스트 308곳을 수집하고 그 중 리뷰가 없는 숙소를 제외하고 281곳을 이용해 숙소 지역 분포 특성을 확인하고, 외국어 리뷰만 있는 숙소를 제외한 후 한국어 텍스트 리뷰가 있는 숙소 146곳에 대한 온라인 리뷰를 작성한 내국인으로만 한정하였다.

둘째, 데이터 수집 대상을 에어비앤비 숙박서비스 서울지역으로 임의로 설정한 예약 날짜에 예약이 가능한 숙소의 내국인 리뷰 데이터를 수집하였

다.

셋째, 수집된 데이터를 정제하고 전처리하여 텍스트 마이닝을 실시하고, 온라인 정보에서 수집한 평점(rating)과 가격(숙박료)의 수치 데이터와 텍스트 마이닝 결과에서 도출된 숙박선택 속성 요인들을 수치 데이터로 변환시켜 통계분석에 활용하였다.



제 4 절 연구의 구성 및 체계

본 연구의 목적을 달성하기 위해 가장 먼저 에어비앤비 온라인 리뷰를 수집하기 위한 사이트를 선정하였다. 에어비앤비 웹 사이트에서 ‘리뷰’, ‘평점’, ‘가격(숙박료)’, ‘숙소 지역’을 수집하고 리뷰 내용에 대해 다양한 텍스트 마이닝을 실시하였다. 그리고 리뷰 데이터를 수치 데이터로 구조화하여 전반적 만족도(평점)과의 영향관계를 검증하여 에어비앤비 선택 속성의 영향력을 깊이 있고 다양한 방법으로 결과를 도출하였다.

본 연구는 다음과 같이 구성하였다.

제 1 장 서론은 연구의 배경 및 문제제기, 연구의 목적 및 필요성, 연구의 내용 및 범위, 연구의 구성 및 체계 등에 대해 서술하였다.

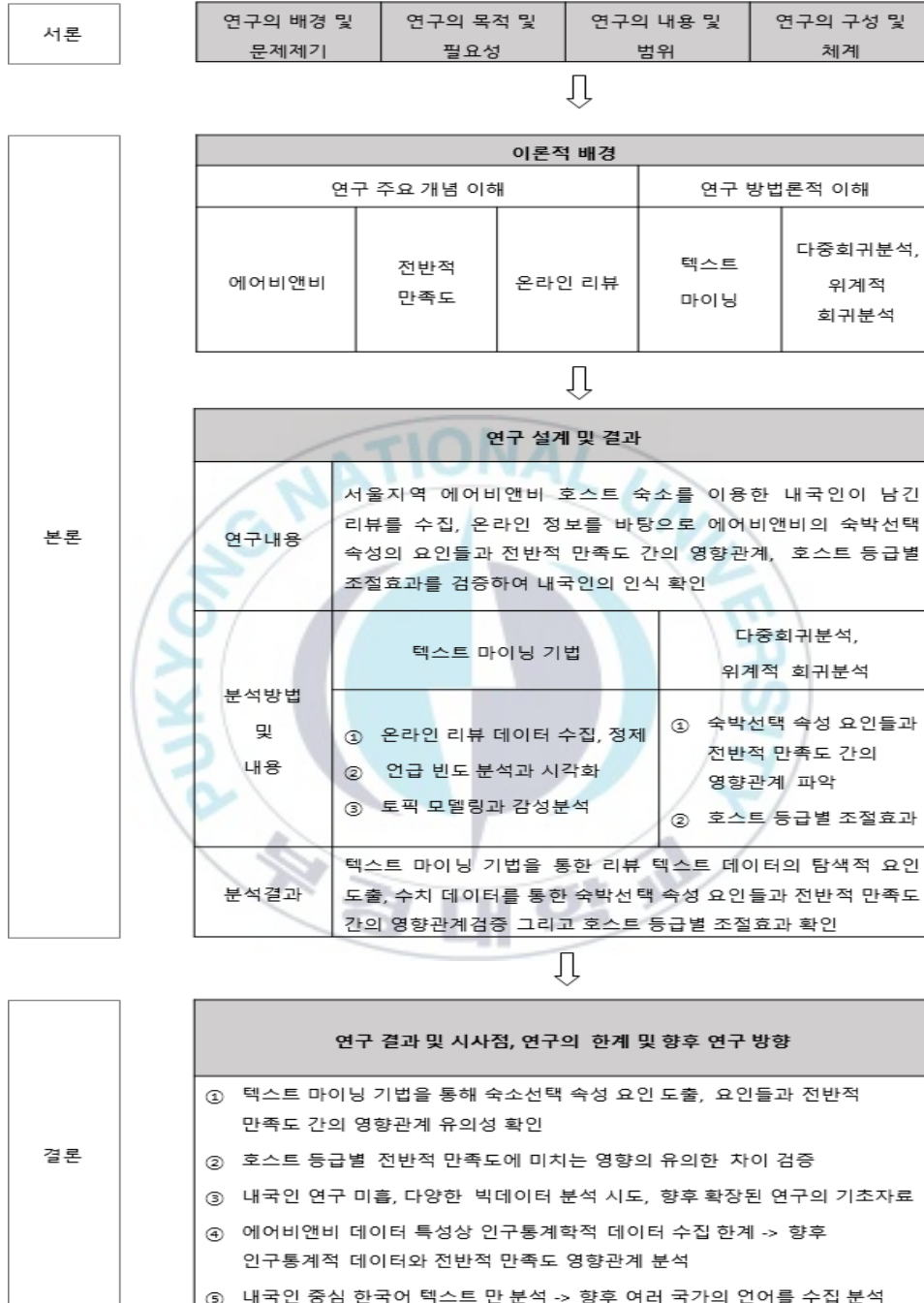
제 2 장 이론적 배경은 1절은 에어비앤비에 대해, 2절은 전반적 만족도에 대해, 3절은 온라인 리뷰에 대해, 4절은 텍스트 마이닝의 특징과 기법의 종류에 관하여 선행연구를 검토하였다.

제 3 장 연구 조사 설계에서 연구 절차 및 방법, 자료 수집 대상 선정, 자료수집 및 전처리에 대해 서술하였다.

제 4 장 연구 결과 부분에서는 먼저 1절에서는 수집한 숙소 지역 데이터 분포 특성을 보여주고, 2절에서는 연구 과제 분석 결과를 검증하였다.

제 5 장 결론 부분은 연구에 대한 요약과 이론적 시사점과 실무적 시사점 및 연구의 한계, 향후 연구방법에 대해 제시하였다.

본 연구 구성 흐름도는 다음 <그림 1-1>과 같다.



〈그림 1-1〉 연구구성 흐름도

제 2 장 이론적 배경

제 1 절 숙박공유서비스 에어비앤비

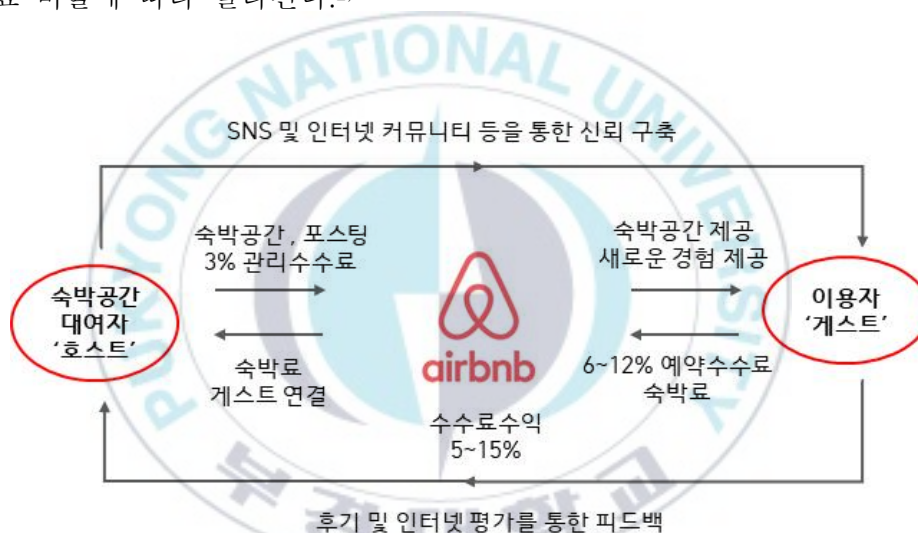
1. 에어비앤비

일반적으로 여행을 할 때 전문적인 숙박서비스를 제공하는 시설을 이용하는 것이 일반적인 형태였다. 하지만 최근 들어 공유경제 개념을 숙박서비스에 적용한 에어비앤비가 등장하면서 전통적인 숙박 이용방법에 변화가 나타나게 되었다(김진한·강우성, 2018).

에어비앤비(airbnb)는 2008년 8월 미국 샌프란시스코에서 Joe Gabbia(chief product officer), Brian Chesky(chief executive officer), Nathan Blecharczyk(chief technology officer)가 “남는 침대를 여행객들에게 임대해 주고 아침밥을 같이 먹는다”라는 슬로건을 내걸고 공동으로 창업하였다. 2007년 월세 고민을 안고 있던 로드아일랜드 디자인 스쿨 재학생이던 Joe Gabbia와 Brian Chesky는 국제 디자인컨퍼런스 기간 동안 손님에게 간이용 침대(air bed)를 제공하면서 아이디어를 얻어 호스트와 게스트를 연결하는 웹 사이트를 개설하게 되고 서비스를 시작하였다. 현재 에어비앤비는 온라인과 모바일을 통해 자기 주택 및 숙소를 업로드하거나 예약하는 숙박 중개 플랫폼으로써 아파트, 빌라, 성(成) 등 다양한 주택 유형을 숙소로 제공하고 있다(정원준·김가운, 2016).

에어비앤비는 협력적 커뮤니티를 기반으로 재화를 거래하는 양면시장 플랫폼으로 비즈니스 모델은 <그림 2-1>과 같이 구성되어 있다. 플랫폼 내에서 유희자원 대여자는 숙박공간을 제공하는 호스트(host)에 해당하며, 이

용자는 게스트(guest)로 대표된다. 에어비앤비는 호스트에게 호스트 서비스 수수료를 부과하고, 게스트에게 게스트 서비스 수수료(숙박예약수수료)를 부과하는 것을 수익모델로 한다. 호스트 서비스 수수료는 예약이 완료될 때마다 호스트에게 부과되는 수수료이며, 게스트의 결제 금액 처리 비용으로 사용된다. 이는 호스트가 게스트에게 부과할 수 있는 청소비를 포함하여 적용되며, 관할 지역의 세법에 따라 부가세가 서비스 수수료에 추가될 수 있다. 게스트 서비스 수수료는 세금 부과 전 예약 대금 소계 적용되는 수수료 비율에 따라 달라진다.¹⁾



〈그림 2-1〉 에어비앤비 비즈니스 모델

에어비앤비는 2008년 8월에 창립되어 샌프란시스코에 본사를 두고 있다. 자신의 공간 일부를 다른 사람에게 빌려주는 세계 최대 숙박공유서비스이며, 현재 2020년 5월 기준으로, 전 세계 191개국 국가 81,000개 도시에서 하루 평균 500만 개의 빈방을 연결해주고 있다. 에어비앤비 숙소는 아파트, 성, 빌라, 주택 등 전 세계 현지인의 다양한 주거공간을 보유하고 있다. 숙

1) 에어비앤비 수수료, <http://www.airbnb.co.kr/help/article/63/what-are-host-service-fees>

박유형은 크게 개인실, 다인실, 집 전체(독채)로 분류하고 있다. 최근 아시아 시장에서도 활발하게 성장하고 있는데, 그 중 한국의 에어비앤비 커뮤니티의 성장은 정부의 혁신성장과 관광산업 성장에 직접적으로 기여하고 있다. 에어비앤비가 내부 데이터와 설문조사 결과를 근거로 ‘2018년 한국경제 파급효과’에 대한 분석을 실시한 결과 2018년 한 해 동안 1조 3,700억 원(12억 5,000달러)의 경제적 효과를 보았고, 이것은 전 세계 에어비앤비 운영 지역 중 16위에 해당하는 성과이다(에어비앤비, 2019). 2018년 10월 10일 ‘세계지식포럼’의 연사에서 에어비앤비 아시아·태평양 정책 총괄 대표는 한국의 경우 2018년 1월 1일 기준 전국 에어비앤비 숙소는 3만 7,100곳이라고 했다. 2017년 에어비앤비를 방문한 게스트는 188만 8,000명에 달했고, 그 중 내국인은 123만 1,500명으로 전체의 65%나 차지한다고 했다(안갑성, 2018). 이처럼 에어비앤비를 이용하는 내국인 고객 비중이 증가하는 상황에서 국내 수요를 꾸준히 창출하기 위해 숙소 공급을 양적으로 늘릴 뿐 아니라, 내국인 고객은 어떤 기준으로 숙소를 선택하는지, 숙소를 예약할 때 중요하게 고려하는 것은 무엇인지 살펴보기 위해 이를 반영한 탐색적 연구를 진행하였다. 내국인 고객 수요는 외래 관광객 수요보다 국제 사회에서 급변하는 외교, 질병, 테러, 자연재해 등 다양한 거시 환경 변수 영향에 덜 민감하기 때문에(문지호·한상린, 2016; 정은별·이형룡, 2020), 지금과 같은 코로나19로 인한 어려움을 이겨내기 위해서는 내국인 고객을 중심으로 다양한 정책과 연구가 진행되어야 한다.

다음은 <표 2-1>과 같이 국내·외 에어비앤비의 파급 현황을 정리하였다.

〈표 2-1〉 에어비앤비 국내·외 파급 현황

일시	국외	국내	자료출처
2008년	8월 샌프란시스코에서 시작	-	송순영(2015)
2013년 1월	-	에어비앤비 한국진출, 한국 숙박 공간 725개 새로 등록 한국 10월 숙소 2,000개로 시작	이혜미(2018)
2015년	8월 전세계 190개국, 34,000개 이상의 도시, 150만 개 이상의 숙소, 1,400개 이상의 성을 보유, 누적 이용객 수 4,000만 명 기록	한국 숙소 10,000개 넘어섬 12월 기준 국내 이용객 2014년 대비 약 235% 증가 숙소 또한 13,000개로 117% 성장률	송순영(2015) 이혜미(2018)
2016년	창립 8년만 기업가치 약 255억 달러(약 30조 690억 원) 2016년 5월 15,000개 넘어섬 에어비앤비 기업가치 세계 3위를 차지, 약 300억 달러(약 33조원) 가치 상승, 2015년 대비 6배 상승 191개국, 34,000개 이상의 도시, 200만 개 이상의 숙소 보유, 누적 이용객 수 8천만 명 기록	한국은 2월까지 국내 방문 이용자가 전년 동기보다 219% 증가, 13,000개 숙소 확보	김지희·이건웅(2019) 이혜미(2018) 에어비앤비(2020)
2017년	191개국, 65,000개 이상 도시, 게스트 2억 명 이상, 숙소는 300만 개 이상 등록	-	최유성 등 (2017)
2018년	190여 개국에 60만 개의 숙소등록, 9월 전 세계 191개국 81,000개 도시 500만 개 이상의 숙소 정보제공, 매일 평균 2백만 명 이상 이용, 현재까지 1억 6천만 명의 사람들이 이용 전세계 191개국, 81,000개의 도시, 500만 개 이상의 숙소 중개, 시장가치 380억 달러	한국은 2018년 1년간 1조 3,700억원의 경제효과, 에어비앤비 이용 방문객 수 총 294만명	김지희·이건웅(2019) 김진한(2019) 윤동원·이성립(2019) 에어비앤비(2019)
2019년	전 세계 700만 개 이상의 숙소, 5만여 개의 체험, 지금까지 7억 5천만 명 이상의 게스트 이용, 220여 개 국가와 지역에서 62개 언어 이용 가능 전 세계 34개 도시에 지사 설립	-	에어비앤비(2020)

자료출처: 선행연구를 바탕으로 연구자가 재구성

2. 호스트 등급(일반 호스트 vs 슈퍼 호스트)

에어비앤비를 이용하고자 하는 사람들은 호스트(host)가 제공하는 숙소를 선택하는 경우 일반적인 상업형태의 숙소를 선택할 때와는 다른 관점에서 숙소를 결정하게 된다. 대표적인 상업형 숙박서비스인 호텔을 예로 들면 글로벌 호텔 체인의 경우 브랜드만으로도 숙박시설 및 서비스에 대해 상당히 많은 부분을 이용하지 않은 상태에서도 충분히 가늠할 수가 있다. 호텔 객실의 품질, 온라인 웹 사이트에서 제공하는 사진과 실제 시설과 일치 정도, 웹 사이트에서 제공하는 정보에 대한 신뢰, 온라인에서 결제할 때 보안 및 안전성, 숙소로 도착해서 이용하는데 수반되는 상당히 많은 내용을 이미 사전에 파악하고 예상에 따른 결과와 일치하여 대부분 예측할 수가 있다. 하지만 에어비앤비 숙박공유서비스의 경우는 상업 서비스를 이용할 때 기본 전제로 간주하던 상당히 많은 내용들을 경험하기 전에는 진실성 여부 또는 정도를 확인할 수 없기에 전체적으로 확인해야 할 항목을 빠뜨리지 않고 점검하고 최대한 각각의 특성을 주의 깊게 살펴봐야 한다(한희정·구철모·정남호, 2016).

호스트 특성으로 새로운 유형의 서비스인 숙박공유서비스는 서비스 이용에 있어 호스트가 이용자의 태도와 이용 의도 형성에 중요한 영향을 미친다. 일반적인 온라인에서 물건을 거래하는 경우는 전문적인 판매 기업이거나 개인이 제공자가 될 것이며 대부분 판매자가 누구인지 보다는 어떤 물건을 판매하는지가 주요 관심 대상일 것이다. 하지만 낯선 사람의, 때로는 국가가 다른 나라 사람이 현재 이용하고 있는 개인 공간을 이용하는 숙박공유서비스의 경우 호스트의 특성이 서비스 이용에 지대한 영향을 미친다고 볼 수 있다(김진한, 2019). Crosby, Evan, & Cowles(1990)의 연구에서

는 호스트의 전문성이 이용자의 신뢰를 얻는 가장 중요한 요인으로 밝혀졌다. 신지용·박성용·김대호(2006)은 호스트의 특성들이 서비스 이용 의도에 영향을 주고 이용 결정에 결정적인 영향을 미치는 중요한 요인이라고 하였다. 특히 이런 특성들은 온라인 상품 구매에 관련된 연구를 통해서도 구매 의도에 영향을 주는 요인이라는 것이 연구 결과에서 밝혀졌다.

슈퍼 호스트(super host)는 에어비앤비 자체에서 연 4회 호스트의 활동을 직접 평가함으로써 자동 지정되는 등급으로 까다로운 조건을 요구받는다. 총 100박 이상의 호스팅 경험, 게스트 중 50% 이상의 후기 작성, 취소율 0%, 그리고 4.8점 이상의 전체 평점 등을 조건으로 정하며, 슈퍼 호스트로 지정될 시 각종 혜택을 제공받는다. 슈퍼 호스트는 숙소의 운영 및 관리가 잘 됨을 인증받은 호스트이며 고객에게 일정 수준 이상의 서비스 품질을 예상 가능하게 한다. 따라서 게스트는 슈퍼 호스트 숙소의 서비스 품질에 대해 더욱 신뢰성을 갖고 해당 숙소를 예약, 이용할 수 있게 된다(이의준 등, 2020). 즉 에어비앤비 사이트로부터 공식적으로 인증받은 슈퍼 호스트 여부는 게스트에게 숙박 시설과 호스트에 대한 신뢰성을 입증할 수 있는 신호가 될 수 있다. 권혁진 등(2017)에 따르면 호스트의 진실성과 호감성 그리고 전문성이 호스트 신뢰와 재이용 의도에 유의한 영향을 미친다고 한다. 에어비앤비 플랫폼 내 슈퍼 호스트라는 등급은 앞서 언급한 여러 조건을 충족한 호스트이므로 사용자들은 해당 호스트의 숙박시설에 대해 일정 수준 이상의 서비스 품질을 기대하게 된다.

Gunter(2018)는 슈퍼 호스트가 될 수 있는 조건에 대한 연구를 수행하였다. 슈퍼 호스트는 에어비앤비 서비스를 성공적으로 이끈 호스트들에게만 주어지는 특권에 해당하기 때문에, 슈퍼 호스트 배지를 받아 게스트들에게 선호를 얻는 것뿐만 아니라 에어비앤비 운영 시에도 지원의 우선권을 받을 수 있으며, 더 많은 예약건을 확보할 수 있다는 장점이 있다. Liang,

Schuckert, Law, & Chen(2017)은 게스트는 슈퍼 호스트 배지가 있는 호스트에 대해 예약 시 더 높은 금액을 지불하고자 하는 의향이 있음을 밝혔다. Ert et al(2016)은 에어비앤비에서 호스트의 사진이 호스트에 대한 신뢰에 미치는 영향에 대하여 분석하였다. 이에 본 연구에서 호스트 등급에 따른 에어비앤비 숙소에 대한 전반적 만족도에 미치는 영향의 차이가 있을 것으로 판단하여 ‘일반 호스트’와 ‘슈퍼 호스트’로 따로 분류하여 차이검증을 진행하였다.

3. 에어비앤비 선택 속성

속성이란 어떤 한 상품이 가진 유형 및 무형적인 특징을 의미한다. 숙박 선택 속성은 숙박시설이 제공하는 유형, 무형의 요소 중에서 소비자가 숙박상품을 구매할 때 중요하게 여기는 특징적인 요소라고 할 수 있다. 이와 마찬가지로, 숙박공유서비스의 선택 속성은 게스트가 플랫폼에 등록되어있는 숙소(상품)를 비교 및 검토하여 자신이 중요하게 생각하는 기준에 따라 숙소를 예약하게 하는 유형·무형의 특성이라고 할 수 있다(김상태 등, 2017). 선택 속성은 구매 의사결정 과정에서 상품이나 서비스에 대한 고객 선호를 나타내며 고객이 상품이나 서비스에 내재한 여러 가지 속성을 평가하며 최적의 대안을 선택하는 데 관련된 속성을 의미한다(조민호·최인호, 1999).

숙박시설의 다양한 속성 중에서 이용자의 선택에 영향을 미치는 요소를 파악하는 것은 숙박공유서비스의 효과적인 설계와 호스트의 영업 전략을 수립하는 데 있어 필수적이다. 그동안 호텔 등과 같은 숙박시설의 선택 속성에 관한 선행연구는 활발하게 진행되어 왔다(Lewis, 1984; Chow, Garretson, & Kurtz, 1994; Shanhan, & Hyman, 2007; Mao, Yang, &

Wang, 2018; 조민호·최인호, 1999; 손은미·정란수·정철, 2014; 윤석영·이무용, 2015; 정은별·이형룡, 2017; 광민정·최지유·박소현, 2019). 그러나 숙박 공유서비스 이용자의 선택 속성에 대한 실증적 연구는 미비한 상황이지만, 최근 정은별·이형룡(2020)이 에어비앤비를 이용한 내국인 고객의 숙박시설 선택 속성에 대한 실증적 연구를 진행하였는데, 선택 속성을 도출하는 방법으로 에어비앤비 웹 사이트에서 제공하는 정보를 바탕으로 도출하고, 설문조사를 실시하였다. 분석 방법은 상대적 중요도를 계산할 수 있는 BWS(best-worst scaling)을 사용하여 서로 다른 속성을 비교한 결과, 내국인이 가장 중요하게 고려하는 속성으로 ‘가격’으로 나타났고, 그 다음으로 ‘청결’, ‘위치’, ‘정확성’ 순으로 유의하게 나타났다. 선행연구 고찰 결과, 에어비앤비가 온라인으로 거래가 이루어지는 플랫폼이라는 특성을 고려하여 에어비앤비 웹 사이트에서 호스트와 게스트가 서로 제공하고 확인할 수 있는 정보인 평점(rating) 측정 항목과 사진 정보, 가격 정보 등을 바탕으로 에어비앤비 선택 속성을 도출하였다.

이에 본 연구에서는 위 선행연구의 선택 속성을 참고하여 본 연구에 선택 속성 요인 도출에 활용하였다.

4. 에어비앤비 관련 선행연구

에어비앤비는 2008년 샌프란시스코에서 처음 시작되어, 국내는 2013년에 진출하면서 급속한 성장을 이루었고 그에 따라 에어비앤비에 관한 연구도 점차 증가하고 있다. 연구 동향을 살펴보면 2016년 시점부터 국내에서도 다양한 연구가 진행되어왔다. 선행연구는 대체로 에어비앤비의 사용자 측면에서 호스트 또는 게스트의 속성으로 구성된 독립변수가 서비스 이용에 미치는 영향에 관해 이루어졌다. 그 중에서도 호스트의 비명시적 요소에

관한 연구로 호스트의 ‘진실성’, ‘전문성’, ‘호감성’, ‘유사성’이 게스트의 호스트에 대한 신뢰와 재이용 의도에 미치는 영향에 관한 연구(권혁진 등, 2017)가 있었고, 류일한 등(2017)은 호스트의 리뷰에 대한 피드백이 구매 의도에 미치는 영향에 대해 연구하였다. 에어비앤비와 같은 온라인 플랫폼은 상호 신뢰를 기반으로 하는 특성이 있기 때문에 호스트의 신뢰성, 사회적 자본과 관련된 연구도 진행되었다(주덕, 2016; 최연태·박상인, 2018).

또한, 명시적 요소에 관한 연구는 숙박시설 자체의 특성인 관광지와의 거리, 입지 분포 등에 관해 분석을 진행하였는데 허지정·노승철(2019)과 박소민·양승우(2020)이 대표적인 연구이다. 그리고 건물의 유형, 객실 구성, 부가 서비스 등과 가격의 상관관계에 관한 연구(박나희·유윤희, 2018)가 있다. 최근에는 공급자의 명시적 요소에 관한 연구에 관심이 높아지고 있다. 이유는 에어비앤비의 가장 큰 특징인 호스트에 대한 검증이 숙박시설의 안전성과 신뢰를 구축하기 위해 신분증을 요구하거나, 호스트의 사진, 숙소 내부 사진, 숙소의 위치, 사용자 후기와 평점을 공개하고 있기 때문이다. Gunter(2018)는 미국의 특정 지역을 대상으로 슈퍼 호스트가 평점에 미치는 영향관계를 검증하였는데, 이렇게 호스트의 정보는 게스트가 호스트를 선택하는 과정에서 활용되어 플랫폼 간의 신뢰를 구축할 수 있다고 했다. Tussyadiah(2016)는 숙박공유서비스가 만족도에 영향을 주는 관계에서 객실의 타입, 호스트 의사소통이 유의한 영향을 주는 것으로 나타났고, 주위예치·김소연(2019)은 정보 품질이 소비자의 정보 만족에 미치는 연구에서 ‘신뢰성’, ‘용이성’, ‘유희성’이 유의하게 나타났다. 김진한·강우성(2018)의 연구에서는 소비가치가 이용 의도에 유의한 영향을 미치는 것으로 나타났다.

국내 에어비앤비를 이용하는 이용자의 선택 속성을 밝혀내는 연구가 최근에 진행되었는데, 김상태 등(2017)은 ‘후기수’, ‘인테리어’, ‘숙박가격’이 예약률에 미치는 영향관계를 분석한 결과 ‘숙박가격’을 제외한 나머지는 유의

하게 나타났고, 정은별·이형룡(2020)은 12개의 선택 속성들 간의 중요도 분석을 실시 하였는데, ‘가격’이 가장 중요하게 나타났고 그 다음으로 ‘청결’, ‘위치’, ‘정확성’이 유의한 영향을 미치는 것으로 나타났다. 이렇게 실증 연구 중심으로 진행되어 오던 연구가 최근에는 빅데이터 분석 기법을 활용한 연구도 진행되고 있다(박홍식·김주일, 2019).

따라서 본 연구에서는 기존 선행연구의 고찰을 통해 검증된 속성을 참고 하여 에어비앤비 온라인 리뷰에 빅데이터 분석 기법을 활용하여 선택 속성에 대한 탐색적 검증을 진행하고자 한다.



제 2 절 전반적 만족도

그동안 기업들은 고객의 니즈(needs)와 요구에 대해 적극적으로 수용하고 상품개발과 연구에 반영하기 위한 고객의 소리(VOC: voice of customer)를 운영해왔다. 고객들은 상품이나 서비스 관련하여 의견이나 개선사항, 문제에 대해 보상과 해결을 요구하기도 하였는데 기본적으로 고객의 소리는 고객과의 소통을 전제로 하고, 고객의 상품 서비스 품질에 대해 평가하는 소통창구가 되었다(박정기·이동명, 2018). 이런 고객평가제도는 서비스제공자와의 관계, 나아가 기업과의 관계를 지속할지를 결정하는데 중요한 요소였다(박재범·김경애·김상희, 2013). 고객은 서비스제공자가 자신들이 원하는 것을 얼마나 정확하고 성실히 제공하는지, 그리고 고객에게 얼마나 친근하고 호의적이며 따뜻하게 서비스를 전달하는가를 중심으로 평가가 되어진다(Busch, & Wilson, 1976; Sundaram, & Webster, 2000).

마케팅 측면에서 고객의 전반적 만족도는 제품이나 서비스를 경험한 후에 고객 행동을 예측하기 위한 주요 판단변수가 되기에 고객의 니즈에 부응하기 위한 결과로 제품이나 서비스의 재구매가 이루어지는 마케팅 사고의 핵심 개념이라고 할 수 있다(임영희, 2019). 그러므로 고객의 전반적 만족도는 고객의 기대를 충족하는 것을 의미하는 것으로 고객의 다양한 욕구에 대한 만족된 정도로 총체적인 만족도의 측정이라고 할 수 있다(박재관·조태영·이준엽, 2011).

본 연구에서는 에어비앤비 숙소서비스를 경험한 후에 그 결과로 나타난 고객의 전반적 만족도를 파악하기 위하여 온라인 리뷰를 분석하는 것이며, 고객의 기대에 얼마나 부응하고 어떤 요인들이 만족을 형성하는 것인지 분석하기 위해 전반적 만족도에 대한 선행 연구를 고찰하였고 에어비앤비 숙

소 선택 속성의 요인들과 전반적 만족도의 영향관계, 호스트 등급별 차이에 대한 연구를 고찰하였다. 그리고 에어비앤비에서 게스트들에게 요구하는 전체 평점은 전체 만족을 의미하므로 본 연구에서는 전체 평점을 전반적 만족도와 동일시하였다.

1. 전반적 만족도(평점)의 개념

고객의 전반적 만족도에 대하여 Hunt(1977)는 소비 경험이 최소한 기대되었던 것보다 좋았다는 평가라고 정의하였으며, Oliver(1980)는 소비자가 경험에 의해서 지니게 된 결과와 불일치된 기대 감정이 복합적으로 결합하여 발생하게 되는 종합적인 심리상태라고 정의하였다. 만족에 대한 연구는 대부분 고객들이 느끼는 만족을 이끌어 내는 선행 변수에 초점이 맞춰져 있는데, 이는 만족을 이끌어 내는 구성요인이 무엇인지 파악하는 것이 핵심이고 상품에 대한 속성, 선호도, 기대심리 등과 같은 심리적인 측면에서 다루어져 왔다(전영록, 2015). 만족은 특정 상품에 대해 소비 경험에서 갖게 되는 감정적 반응 혹은 고객이 지니고 있는 기대를 특정한 수준 이상으로 충족시키는 것이다(Westbrook, & Reilly, 1983; Cronin, & Taylor, 1994; 박상준, 2017). 또한, 만족을 제품이나 서비스 이용에 대한 전반적인 평가라고 말하며 현명한 선택, 올바른 결정, 전반적 만족, 기대 달성으로 고객 만족 항목으로도 구성된다(한진수·현경석, 2009).

예전부터 서비스산업에서는 고객의 만족에 중점을 두는 연구들이 진행되어왔다(Gunderson, Heide, & Olsseon, 1996). 고객의 만족은 고객의 성취 반응으로 정해진 수준 이상으로 고객의 기대를 충족하는 것을 의미한다. 만족은 보통 경험되는 것이므로 개개인의 감정적 과정에 의해 좌우된다. 객관적으로 낮은 혹은 높은 품질로도 개인에 따라 만족을 얻거나 얻지 못

할 수도 있다(김윤민, 2009). 그리고 고객은 상품이나 서비스 수준이 기대 이상일 경우 일반적으로 만족을 하고 기대보다 훨씬 넘어서면 고객은 매우 만족하여 기업에 충성하게 된다. 높은 품질의 서비스를 제공하고 만족도를 향상시키는 것은 호텔, 숙박, 외식, 관광산업에서 기업의 성공을 이끄는 중요한 요인으로 널리 인식된다(Stevens, Knutson, & Patton, 1995; 이성각·조현진·임병호, 2017).

만족의 측정은 여러 학자들에 의해 다양한 견해가 제시되나, 만족 여부를 판단하려면 에어비앤비 숙박서비스를 구성하는 다양한 서비스의 평가가 종합적으로 이루어져야 하는데, 본 연구의 전반적 만족도는 평가대상 속성들의 종합적 평가 결과라는 관점(김화진·이승곤·한진수, 2011)을 바탕으로 한다.

고객의 만족은 상품 구매 동기가 되고 긍정적 구전이나 추천, 재구매 등의 행동 의도로 이어져 기업 매출 및 경영 성과를 높이기 때문에 기업 입장에서는 매우 중요하다(Anderson, Fornell, & Lehmann, 1994; Bolton, 1998). 고객 만족도 결과 변수 및 효과는 기업의 지속적인 성장과 수익성에 매우 중요한 요소라고 볼 수 있다. 만족한 고객은 재구매를 하며 고정 고객이 되고, 만족한 고객이 전하는 구전효과는 신규 고객을 창출할 수 있다(임영희, 2019).

고객의 전반적 만족도는 본질적으로 간접적인 측정 방법을 통하여 소비자의 최종 행위만을 조사하는 것으로 구매 행동 결과를 보여주게 되는데 그 과정은 매우 다양하다. 그러므로 본 연구는 에어비앤비를 경험한 게스트들이 칭찬과 감사, 만족과 불만족 등의 감정과 사실적 경험을 글과 평점으로 남겨 둔 온라인 데이터를 연구하였다. 경험의 글과 평점은 수많은 잠재 고객들에게 전파되어 영향을 미치게 된다. 이와 같이 에어비앤비 숙소와 호스트 서비스 전반에 걸쳐 선택 속성들의 내용이 있는 온라인 리뷰를

통해 전반적 만족도(평점)에 미치는 영향관계에 대해 연구하고자 한다.

2. 전반적 만족도 관련 선행연구

고객의 전반적 만족도에 대한 선행연구들 중 먼저 호텔 숙박서비스 분야를 중심으로 살펴보면 호텔 객실 부서의 서비스 품질과 고객의 만족과의 관계를 규명하는 연구에서 객실 서비스 품질의 요인으로 ‘인적서비스’, ‘객실 시설’, ‘객실 내 편리성’ 및 ‘객실의 안정성’으로 구분하였는데, ‘객실 시설’을 제외하고는 만족도에 유의미한 영향을 미친다고 했다(구정대·박태수, 2005). 그리고 호텔 속성에 따른 만족도를 외부환경 부문과 객실 부문으로 분류하여 비교 분석한 결과, ‘호텔의 규모’, ‘등급’, ‘위치(접근성)’, ‘객실 청결’, ‘편의시설’ 등이 호텔 만족도에 영향을 미치는 것으로 나타났으며(공효순, 2008), 호텔의 선택 속성 중 ‘객실 및 시설환경’, ‘식음료’, ‘직원 서비스’ 및 ‘부대 서비스’가 호텔 만족 등에 미치는 영향이 유의미하게 나타난 연구도 있다(정지혜·한지수·김진휘, 2015). 또한, 호텔 선택 속성 연구에서 호텔 선택 속성 평가에 따라 고객의 만족도가 달라진다는 연구도 있고(엄수진·윤태환, 2010), 그 밖에도 외식산업 분야에서 선택 속성이 소비자 만족에 긍정적인 영향을 미친다고 했고(이상미·정현영, 2009), 베이커리 선택 속성이 고객 만족에 영향을 미친다는 연구도 있었다(김은주·김천서·신용경, 2006).

에어비앤비의 고객 만족도와와의 관계연구를 살펴보면, Tussyadiah(2016)는 P2P 숙소 이용 경험이 있는 고객을 대상으로 만족 요인들과의 영향관계를 확인하는 연구에서 ‘즐거움’, ‘가치’, ‘편의시설’이 만족도에 영향을 준다고 하였고, 주위에치·김소연(2019)는 정보제공주체별 정보 품질이 정보만족에 미치는 영향관계에서 ‘신뢰성’, ‘용이성’, ‘유희성’이 정보만족도에

유의한 영향을 미친다고 했다. 그리고 이의준 등(2020)은 최근 에어비앤비의 서울지역을 중심으로 데이터를 수집하여 호스트의 특성이 사용자의 만족감에 미치는 연구를 실시하였는데, 여기서 사용자 만족감의 지표로 에어비앤비에서 수집한 대표 점수를 사용하였다.

따라서 본 연구에서는 전반적 만족도를 숙소 경험의 전반적 평가(평점)로 정의하고 내국인이 에어비앤비를 이용한 후 남긴 온라인 리뷰의 텍스트 마이닝 결과로 선택 속성 요인들을 판단하고자 한다.



제 3 절 온라인 리뷰

1. 온라인 리뷰

온라인 리뷰(online review)는 온라인 구전(eWOM: electronic word of mouth)이라 불리며, 서비스를 경험해 본 적이 있는 고객이 자유롭게 평가를 서술하거나 점수를 선택하는 것으로 기업이 제공하는 정보보다 신뢰성과 객관성을 갖춘 정보로 인식되고 있다(Sparks, & Browning, 2011). 고객 리뷰는 마케팅 분야에서 주로 연구되어온 구전(word of mouth)의 한 종류로 이 구전이 소비자들의 구매 의사결정에 중요한 역할을 한다는 것은 이전부터 밝혀져 왔다(Swan, & Oliver, 1989). 온라인 고객 리뷰는 인터넷상에서 구전을 하기 위하여 소비자들이 직접 작성한 인터넷 문서로서, 이미 문서화가 되어 있으므로 연구자가 개입하지 않고서도 현상을 있는 그대로 관찰 및 연구할 수 있는 좋은 조건을 제시하고 있다(김근형·오성열, 2009). 사람들의 입에서 입으로 전해지는 ‘말’의 의미인 구전(口傳)에서 시작되어 지금은 온라인상에서의 구전이 ‘글’이라는 형태로 생성되어 전파되는데, 이것은 소비자의 의사결정 및 제품 구매 행동에 결정적인 영향을 줄 수 있는 중요한 요인이다(Richins, & Root-Shaffer, 1988; 변숙은·조은성, 2015). 고객이 구매한 제품에 대한 의견을 웹 사이트에 기록할 수 있도록 한 곳은 1995년 아마존닷컴이 처음이었고, 그 이후 현재까지 아마존닷컴은 판매하는 모든 제품에 대해 약 1억 개 이상의 온라인 리뷰를 보유하고 있는데 이와 같은 온라인 리뷰를 아마존닷컴의 성공 요인 중 하나라는 견해가 있다(Chen, & Xie, 2008).

Bickart, & Schindler(2001)는 온라인 리뷰는 마케터의 일방적 전달형태가 아니라 소비자의 관점에서 작성되어 사용 전 고객에게 간접경험의 기회를 제공하기 때문에 영향력이 더 있는 것으로 인식된다고 하였고, Smith, Menon, & Sivakumar(2005)는 마케터가 제공하는 정보보다 더 신뢰를 한다고 하였다. 온라인 리뷰에는 판매자가 제공하는 제품정보에서는 얻을 수 없는 정보들을 잠재 구매자들에게 제공하기 때문이다(Li, & Hitt, 2008). 소비자가 생성하고 소비자로 전달되는 온라인 리뷰는 반영구적 지속성을 가지고 문자화되어 불특정 다수의 사람들에게 빠르고 쉽게 노출되는 특성이 있다(이정현, 2013).

관광 분야에서는 관광 정보탐색, 여행 계획수립, 소비에 이르기까지의 일련의 관광 행동 의사결정단계에서 온라인 리뷰 정보는 구전효과를 통해 수요창출의 원천으로 안착하였다(심영석·김홍범, 2016). 이런 웹 사이트상의 온라인 리뷰 정보의 양적 증가로 인하여 여행객들이 방문한 관광지에 대해 어떤 제약에도 얽매이지 않고 그들의 의견을 솔직담백하게 표현한 정보를 과학적 증거기반의 빅데이터 분석을 통해 산업의 전반적인 흐름을 파악하는 유용한 자료로 활용되고 있다(Xiang, Gretzel, & Fesenmaier, 2009).

온라인 리뷰는 호텔서비스를 경험한 고객들이 다른 소비자의 선택을 돕기 위해, 또는 서비스 품질향상을 촉구하기 위해 자신의 경험을 서술한 것으로, 경제적 대가나 외부의 자극 없이 자발적으로 게시하는 것이 일반적이다(최지유·박소현·김상호·이규범·곽민정, 2018). 호텔 산업분야에서도 온라인 리뷰나 평점 등 온라인 정보는 예비 소비자들에게 생활에 필요한 여가 및 관광활동, 특히 호텔 예약 등 중요한 의사결정의 지원 도구가 되고 있다(Wang, & Fessenmaier, 2004). 따라서 리뷰에서 언급하고 있는 서비스에 대한 만족 및 불만족 사항들은 호텔서비스를 구성하는 여러 속성 중 고객이 주목한 속성이 무엇이고 고객이 지각한 서비스 품질 수준이 어떠한

지에 대한 보다 객관적이고 신뢰할 수 있는 정보를 제공한다(Decker, & Trusov, 2010; 권진영·이문용, 2012). 이러한 온라인 리뷰를 대량으로 수집하여 빠르게 분석하는 온라인 리뷰 분석은 질문내용, 표본선정, 응답자의 태도, 연구자의 편향 등에 크게 영향받는 기존 설문지나 인터뷰 방식을 보완하는 유용한 대안이 된다(송보미·윤병운·박용태, 2011; 한주형·정남호, 2017).

2. 온라인 리뷰 관련 선행연구

정보화로 인한 인터넷 사용이 급증하면서 소비자가 제품 및 서비스를 평가하고 구매하는 방식이 변화하고 있다. 온라인 리뷰를 통한 소비 경험의 공유는 이러한 변화의 가장 큰 특징 중 하나이다(곽민정 등, 2019). 온라인 리뷰는 소비자들이 상품의 어떤 속성에 주목했고 그 품질을 어떻게 지각했는지, 소비자 관점의 상품품질과 고객 만족도에 대한 대량의 정보를 포함한다(최지유 등, 2018). 특히 호텔 산업분야에서는 고객의 온라인 리뷰 과 급력과 활용 가능성이 높이 평가되고 있다(Saleh, & Ryan, 1991; 이병철·변효정, 2014; 박경배·하성호, 2015; 유양호·강상묵, 2016; Philips, Barnes, Zigan, & Schegg, 2016). 이에 따라 최근 리뷰의 평점과 서술내용의 분석을 통해 고객의 니즈를 파악하기 위한 온라인 리뷰 분석 연구가 국내·외에서 지속적으로 확대되고 있다. 이들 연구는 크게 리뷰 평점을 분석한 연구, 서비스 속성의 리뷰 내 출현 빈도(언급 빈도)를 분석한 연구, 리뷰 내용의 긍정·부정 감성을 분석한 연구로 나누어 살펴볼 수 있다(곽민정 등, 2019).

온라인 리뷰의 평점에 주목한 연구들 중 먼저 호텔서비스 분야에서 이용숙(2013)은 유럽지역 호텔을 대상으로 리뷰 평점이 높은 호텔들이 어떠한 특징을 가지고 있는지를 분석하는 참여관찰 후 후기 내용과의 비교분석을

수행하였고, 이정현(2013)은 온라인상에 남겨진 호텔 이용 후기에 대해 질적 내용분석을 실시하였고, 한주형·정남호(2017)는 서울지역 호텔을 대상으로 리뷰 평점과 호텔 특성(호텔등급 및 외국인 리뷰수)간의 관계를 검증하였다. 관광 분야에서는 심영석·김홍범(2018)이 관광지 개성과 온라인 정보의 평점과의 영향관계를 검증하였다. 최근 에어비앤비에 대한 온라인 정보를 수집하여 호스트의 특성이 사용자의 만족감에 영향을 미치는 연구가 진행되었는데, 여기서 사용자 만족감을 평점으로 하여 분석하였다(이의준 등, 2020).

한편 리뷰 내용에 주목한 연구는 더욱 활발히 진행 중이다. 리뷰 평점만으로는 소비자의 만족도를 높이는 핵심요인을 세부적으로 파악하는 데 한계가 있으므로 내용분석의 필요성이 강조되고 있으며(Archak, Ghose, & Ipeirotis, 2011), 다수 연구가 텍스트 마이닝(text mining)을 활용해 리뷰 내에서 많이 출현하는 명사 형태의 속성을 추출하는 언급 빈도 분석을 제안하였다. 이정현(2013)은 호텔이 고객에 제공하는 가치를 ‘기능적 가치’, ‘정서적 가치’, ‘가격대비 가치’ 등으로 나누고, 200여 개 온라인 리뷰의 내용을 분석하여 각 가치 관련 속성들의 언급 빈도와 상세 내용을 확인하였다. 심영석·김홍범(2016)은 관광객들이 인식하는 관광지 이미지의 구성요인을 밝히기 위해 여행 커뮤니티 트립어드바이저닷컴(tripadvisor.com)의 서울 관광지를 방문한 외국인 관광객의 온라인 리뷰를 텍스트 마이닝 기법을 활용하여 분석하였다. 임영희·김홍범(2019)도 외래관광객의 온라인 리뷰를 활용하여 빈도 분석을 실시하고, 추가적인 텍스트 마이닝 기법을 활용한 감성분석을 통해 만족과 불만족의 단어들을 추출하는 분석을 실시하였다. 최지유 등(2018)은 서울·제주·부산지역 호텔에 대한 8,791개 리뷰를 대상으로 온라인 리뷰를 분석하였다. 10개의 대표속성, 42개의 세부속성 및 하위단어로 구성된 호텔서비스 품질 속성사전을 구축하고 속성별 언급 비율 분

석을 수행하여 호텔서비스 주요속성을 파악하였다. 특히 지역·호텔등급·고객 유형을 동시에 고려하는 세분화를 제안하고 세분시장별 언급 비율 분석을 수행하였으며, 지역별, 호텔등급별, 고객유형별 공통점과 차이점을 논하였다.

Chevalier, & Mayzlin(2006)는 booksellers와 Amason.com의 고객 리뷰를 바탕으로 연구한 결과 ‘가격’, ‘평점’, 그리고 각각의 평점에 해당하는 리뷰를 갖고 분석하였는데, 많은 양의 리뷰 추천이나 피드백은 매출액을 증가시킨다는 것을 발견했다. 숙박공유서비스 에어비앤비의 경우 대부분의 고객들은 숙소 예약 전 호스트에 대해 실제 이용 경험이 있는 사람들의 리뷰와 평점을 미리 읽고 숙소를 선택하게 된다. 고객과 호스트는 예약 이행이 완료된 후에만 서로에 대해 후기를 남길 수 있기 때문에 남겨진 후기는 실제 경험을 바탕으로 한 생생한 정보이다. 이용 후기에 사용자는 ‘숙소정보의 정확성’, ‘숙소의 위치’, ‘호스트와의 의사소통’, ‘체크인’, ‘청결도’, ‘숙소 이용가치(현재는 가격대비 만족도)’에 대해 종합적으로 평가하여 평점을 매기거나 문장형태로 리뷰를 작성한다. 류일한 등 (2017)은 게스트가 남긴 리뷰에 호스트 피드백이 구매 의도에 영향을 미치는지를 분석하여 의미 있는 결과를 도출하였는데, 이 연구를 통해 게스트의 리뷰뿐만 아니라 호스트의 리뷰 피드백의 중요성을 인식하게 되었다. 최근 해외연구에서도 에어비앤비 숙박서비스 온라인 리뷰의 중요성에 대한 연구가 진행되었다(Bridges, & Vásquez, 2018; Lawani et al, 2019; Cheng, & Jin, 2019). 리뷰를 통해 에어비앤비를 이용하는 고객에 대한 인식 연구가 필요해지는 시점에서 국내연구 윤동원·이성림(2019)은 에어비앤비 온라인 커뮤니티 클리앙(www.clien.net)에서 에어비앤비와 관련된 게시글을 수집하여 텍스트 마이닝 분석을 통해 빈도 분석, 군집 분석, 사회연결망 분석을 실시하여 소비자의 긍정적·부정적 인식의 관한 연구를 진행하였다.

따라서 본 연구에서는 호텔 관광 분야의 선행연구를 고찰하고, 최근 에어비앤비의 온라인 리뷰 연구 동향을 파악하여 온라인 정보를 수집한 데이터를 통해 다각도의 방법론을 적용하여 고객이 인식하는 요인들에 관한 탐색적인 연구를 진행하고자 한다.



제 4 절 텍스트 마이닝

1. 텍스트 마이닝

오늘날 텍스트 마이닝(text mining)은 빅데이터 분석 또는 데이터 과학의 대표적인 한 분야이다(김수현·이영준·신진영·박기영, 2019). 다양한 분야에서 생성되는 데이터의 대부분은 정돈되지 않은 비정형 데이터의 형태를 이룬다. 이를 이용한 분석에 대한 관심도 점점 증가하고 있다. 텍스트 마이닝은 텍스트(text)를 이용한 데이터 마이닝(data mining)이라고도 말할 수 있는데, 텍스트를 사용하여 패턴이나 관계를 추출하고 그 안에 의미 있는 정보나 가치를 발굴하여 해석하는 일련의 과정을 통칭한다. 즉, 텍스트 마이닝은 다양한 포맷(웹 콘텐츠, PDF, 마이크로소프트 오피스 파일, XML, 텍스트 파일 등)의 문서로부터 데이터를 획득해 이를 문서별 단어의 행렬(matrix)로 만들어 추가분석이나 데이터 마이닝 기법을 적용해 통찰(insight)을 얻거나 의사결정을 지원하는 방법이라고 할 수 있다(한국데이터산업진흥원, 2014). 텍스트 마이닝은 비구조화된 텍스트로부터 흥미 있고 유용한 패턴을 찾아내는 과정으로, 정보 추출·정보 검색·자연어 처리·텍스트 요약·자동 분류 등에서 사용되는 기법을 결합하여 사용하는 것으로 텍스트로부터 이전에 알려지지 않은 숨겨진 지식을 찾아내는 것이 핵심이다(김수현, 2006).

IBM 보고서(2015)²⁾에 따르면 전 세계 데이터의 80%가 텍스트와 같은 비정형 데이터(unstructured data)로 이루어져 있다. 우리가 매일 접하는 언어와 텍스트에는 풍부한 정보가 내재 되어 있으나, 하루에도 수백만 건

2) IBM보고서: <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/>

의 문서가 생산되는 오늘날에는 인공지능력만으로 방대한 텍스트 자료를 처리하는 데는 한계가 있다. 이제는 텍스트 마이닝으로 컴퓨터의 방대한 텍스트 데이터를 자동으로 분석하게 함으로써 유익한 정보를 효율적으로 얻을 수 있게 되었다.

이런 텍스트 데이터는 4차 산업과 빅데이터 분석 기술 발전 등에 힘입어 최근 새로운 정보의 원천으로 인식되고 있다. 텍스트 마이닝은 연산이 어려운 비구조적 텍스트 데이터를 수치화 또는 벡터(vector)화하여 구조화된 자료로 변환시킴으로써 정보의 원천으로 활용하게 하는 기술이다(김수현 등, 2019). 텍스트 마이닝은 텍스트 기반 데이터로부터 자연어 처리 기법(NLP: natural language processing)을 근간으로 기존에 알려지지 않은 유용한 지식을 발견하기 위한 목적으로 언어(텍스트)와 기계적 알고리즘에 의해 사용자가 관심을 가지는 정보를 자동적으로 추출하는 프로세스를 의미한다(Hotho, Nürnberger, & Paaß, 2005). 텍스트 마이닝 기법을 활용하여 과거에는 분석이 어려웠던 질적 데이터(텍스트)의 분석이 가능해졌고, 실시간으로 생성되는 엄청나게 많은 데이터를 분석할 수 있는 기술이 현실화되고 있다. 새롭게 생성되는 질적 데이터를 이해하고 유용한 지식을 추출하여 현재의 주요 이슈를 탐지하고 모니터링하여 미래를 예측하기 위한 연구도 활발히 진행되고 있다(백영민, 2017).

본 연구에서 활용한 분석 기법 툴(tool)인 R은 텍스트 마이닝의 함수 패키지를 설치하여 사용하는 방식으로, 텍스트 마이닝을 위한 함수 라이브러리로 텍스트 마이닝의 첫 글자를 따서 만든 ‘tm 라이브러리’는 방대한 양의 텍스트를 효과적이고 효율적으로 처리할 수 있다. 텍스트 마이닝에서는 절차에 따른 다양한 용어를 사용하고 있는데, 텍스트 마이닝이 적용되는 전체 텍스트 데이터를 ‘말뭉치(corpus)’ 또는 ‘텍스트 말뭉치(text corpus)’라고 부른다. 말뭉치는 “대용량의 정형화된 텍스트 집합(large and

structured set of text)”으로 정의된다(Miner, Elder, & Hill, 2012). 그리고 데이터를 전처리(preprocessing)하는 과정에서 많은 자연어 처리 알고리즘이 단어 수준의 텍스트 분석에 근거하는데, 단어를 식별해내는 것을 토큰화(tokenization)라고 한다. 이는 텍스트가 주어졌을 때 이를 하나의 유용한 의미 단위들로 분해(parsing)하는 과정이다.³⁾ 그러나 토큰화된 모든 문자열을 분석에 사용할 수 있는 것은 아니다. 언어에는 자주 사용되지만 별 의미가 없이 관용적으로 사용되는 단어들이 상당수 존재한다. 이런 단어들은 제거해 줌으로써 분석의 효율을 높일 수 있다. 이런 단어를 불용어(stop words)라고 한다(김수현 등, 2019).

텍스트 마이닝을 활용한 연구가 다양한 분야에서 활발히 진행되고 있다. 학술논문의 연구 동향 분석(김성근·조혁준·강주영, 2016; 황서이·황동열, 2018; 김희섭·강보라, 2019b)이나, 트위터와 같은 SNS의 데이터를 활용한 사회적 이슈 분석(곽수정·김현희, 2019; 이해진·최자영, 2019), 온라인 리뷰를 활용한 고객 유형 분석 등의 연구(심영석·김홍범, 2016; 류일한 등, 2017; 객민정 등, 2019; 임영희·김홍범, 2019)에서 활용되고 있다.

2. 텍스트 마이닝 기법 종류

텍스트 마이닝 방법에는 분류(classification), 군집화(clustering), 정보 추출(information extraction), 네트워크 분석(network analysis), 감성분석(sentiment analysis), 토픽 분석(topic tracking), 개념분석(concept linkage) 등이 있는데(Fan, Wallace, Rich, & Zhang, 2006), 이런 분석 기법들의 사용 의도는 데이터의 성격, 연구자의 판단과 연구의 목적에 따라 달라질 수

3) 하나의 의미를 나타내는 일련의 문자열은 영어에서는 단어가 되겠지만 한국어에서는 조사가 붙게 되므로 이를 제외한 어근 등의 형태소가 된다. 형태소는 일정한 의미가 있는 가장 작은 말의 단위로 한국어 텍스트를 분석할 때는 형태소를 기준으로 분석하므로 영어에 비해 훨씬 복잡한 알고리즘으로 토큰화 작업이 필요하다(김수현 등, 2019).

있다. 텍스트 마이닝 과정은 크게 데이터 수집, 데이터 전처리, 단어(키워드) 추출, 정보 분석으로 나눌 수 있다. 다시 크게 3단계로 나누면, 텍스트 데이터를 수집한 후 전처리 과정을 거쳐 비정형에서 정형의 형태로 변경하는 과정이 1단계이고, 전처리 과정을 통해 형태소 단위로 분해하고 불용어를 제거하거나 문법적 오류를 교정하는 작업을 통해 텍스트 분석이 가능한 형태로 변환하는 것이 2단계이다. 마지막 3단계는 전처리를 거친 데이터는 정보 추출, 요약, 군집화 등의 분석이 가능해진다. 다음은 본 연구에서 활용한 텍스트 마이닝 기법의 종류이다. 가장 먼저, 언급 빈도 분석이 있고, 그다음으로 워드 클라우드, 토픽 모델링, 감성분석을 진행하였다.

가. 언급 빈도 분석

언급 빈도 분석(term frequency)은 전체 문서에서 특정 단어가 등장하는 빈도수를 확인하는 분석이다. 문서 내에서 자주 언급되는 단어가 무엇인지를 확인할 수 있는 분석 기법으로 텍스트 마이닝 분석 과정에서 가장 먼저 도출되는 결과값이다. 본 연구에서는 사용자 후기인 온라인 리뷰 텍스트를 분석하기 위해 가장 기본적인 문서 내 단어 언급 빈도를 사용하였다.

텍스트 마이닝을 활용하는 연구에서 반드시 사용하는 분석이 언급 빈도 분석인데, 이정현(2013)은 호텔이 고객에 제공하는 가치를 ‘기능적 가치’, ‘정서적 가치’, ‘가격대비 가치’ 등으로 나누고, 200여 개 온라인 리뷰의 내용을 분석하여 각 가치 관련 속성들의 언급 빈도와 상세 내용을 확인하였다. 이선미·배광현·이규범·이진희·곽민정(2016)은 국내 400여 개 호텔에 대한 1만여 개 리뷰를 대상으로 언급 빈도 분석을 수행하였다. 이들은 ‘고객 유형별’, ‘호텔등급별’, ‘계절별’, ‘연도별’ 언급 빈도에 차이를 검증하였다.

나. 워드 클라우드

문서의 전처리 과정을 거쳐 단어의 품사별로 빈도를 계산하게 되면, 이를 활용하여 다양한 시각화를 할 수 있다. 이러한 시각화 과정은 문서의 주제, 특징 등을 보다 직관적으로 표현할 수 있어 보편적으로 사용되는 방법이다. 빈도를 기준으로 그래프를 그릴 수도 있겠으나, 워드 클라우드(word cloud)는 빈도에 비례하여 단어의 크기를 나타내어 말뭉치를 시각화하는 방법으로 문서의 특징을 보다 직관적으로 나타내어준다(김수현 등, 2019).

박홍식·김주일(2019)은 워드 클라우드를 활용하여 사회적 사건 전후를 비교 분석하였는데, 상위 빈도 단어들의 차이를 직관적으로 확인할 수 있었다. 곽수정·김현희(2019)도 역시 사회적인 현상에 대한 빈도 단어를 도출하여 워드 클라우드로 시각화하였다. 이은영·주경희·이두희(2019)는 최근 10년간 소비자학 연구 동향을 파악하기 위해 워드 클라우드 기법을 활용하였다.

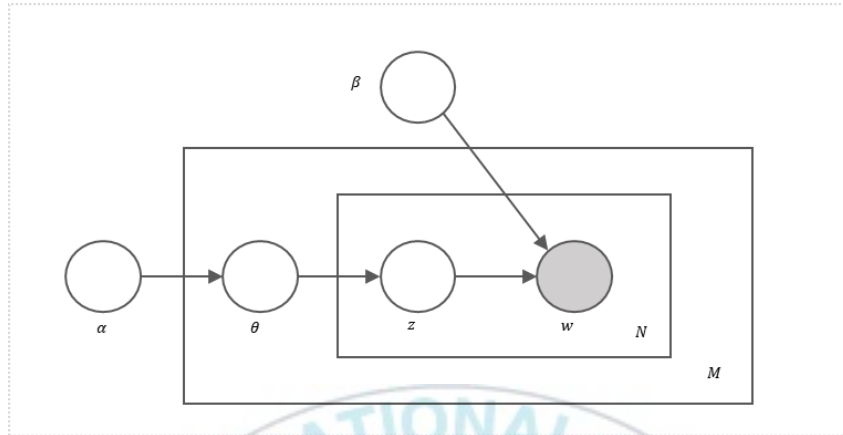
다. 잠재디리클레할당(LDA: latent dirichlet allocation) 기반 토픽 모델링

토픽 모델(topic models)은 머신 러닝 분류 기법과 유사하게 데이터의 차원을 축소하거나 이를 바탕으로 데이터를 분류하는 기법이다(김수현 등, 2019). 토픽 모델링(topic modeling)은 주로 비구조화된 텍스트 문서에서 새로운 정보를 추출하기 위한 연구에 많이 활용되고 있으며, 특정 분야의 주제 및 동향을 파악하는데 효과적인 수단으로 제안되고 있다(이금실·이인주·이영경, 2018). 토픽 모델링은 전체 문서 집합에서 나타나는 의미론적 구조를 탐색하기 위해 사용되는 텍스트 마이닝 기법으로 문서 집합 내에

숨겨져 있는 주제(topic)를 다양한 통계적 방법을 활용하여 표현하는 것을 말한다. 이러한 토픽 모델링은 어떤 문서의 토픽이 여러 개의 추상적인 토픽으로 구성되어 있으며, 각각의 토픽은 여러 단어의 집합으로 구성되어 있다는 것을 전제하고 있다. LDA는 문서 내의 잠재된 토픽을 추정하는 것으로 어떤 단어들을 어떠한 토픽에서 선택하여 배치할 것인지 각각의 파라미터(parameter: α, β)로 모델링하는 기법이다(박종도, 2019).

본 연구에서는 R을 활용하여 토픽 모델링을 실시하였다. 토픽 모델링은 문서와 단어로 구성된 행렬(dtm)을 기반으로 문서에 잠재되어(latent)있다고 가정된 토픽의 등장확률을 추정하는 일련의 통계적 텍스트 처리기법을 일컫는다(Blei, & Lafferty, 2009; Blei, Ng, & Jordan, 2003). R의 'topicmodels' 라이브러리를 이용하면 LDA를 추정할 수 있다. 말뭉치를 사전처리한 후 dtm으로 구성했을 경우, dtm의 가로줄에는 문서가, 세로줄에는 단어가 배치된다. 따라서 dtm은 M 개의 문서(d_m)와 N개의 단어(w_n)로 구성된 행렬이다. 말뭉치는 일련의 문서들의 집합이다. 다시 말해 말뭉치를 측정의 최상의 단위라고 가정할 때 특정 문서(d_m)는 말뭉치에 배속되어(embedded)있다. 이를 수학적으로 표현하자면, 말뭉치를 D라고 할 때 말뭉치는 $D = \{ d_1, d_2, d_3, \dots, d_m \}$ 과 같다. 문서는 일련의 단어들의 집합이다. 문서와 단어의 관계는 말뭉치와 문서의 관계와 본질적으로 동일하다. 즉, 하위 단위인 문서들이 상위 단위인 말뭉치를 구성하듯, 하위단위인 단어들은 상위 단위인 문서를 구성한다. 따라서 단어는 문장을 구성하고, 다시 문서는 말뭉치를 구성한다. 이를 수학적으로 표현하자면, 특정 문서(d_m)는 $d_m = \{w_1, w_2, w_3, w_4, \dots, w_n\}$ 으로 표현된다(문서에서 나타나지 않은 단어의 경우 $w_m = 0$ 이라고 할 수 있다). 그리고 단어는 잠재된 특정 토픽을 나타내는 관측이다. 토픽은 관측된 단어들을 통해서 추론할 수 있다고 간주한다. 이때 단어들의 집합인 문서에서 특정한 잠재토픽이 등장할 확률은 다음 <그림

2-2>와 같이 표현할 수 있다.



자료 출처: Blei, Ng, & Jordan(2003: 997).

M : 말뭉치 공간, N 은 문서 공간, w 단어, z 는 토픽

β 는 특정단어가 특정 토픽에서 생성될 때 개입되는 확률값

z 를 생성하는 θ 라는 이름의 모수(parameter)로서 토픽의 개수(k)만큼의 차원을 갖는 디리클레 랜덤 변수이며,

이는 인간이 사용하는 전체 언어에서 분석 대상인 말뭉치의 토픽들을 생성시키는 역할을 함

〈그림 2-2〉 LDA 토픽 모델링 모형

토픽 모델링 기법의 장점은 특정 비정형 텍스트 자료에 대한 맥락적 이해를 돕는데 활용할 수 있다. 주어진 문서에서 단어 수를 분석하고 해당 문서가 어떤 주제들과 함께 출현하고 연결되는지에 따라 분류하여 예측하는 알고리즘이다. 주제를 특정하기가 어려운 문서들의 조합에서 등장하는 확률을 반복적으로 계산하여, 출현 확률이 높은 단어들의 조합과 연결을 통해 주제의 가능성이 높은 단어들을 그룹화하는 방법이다. 이것은 특정 문서집단이 어떤 맥락과 범주에서 기술되었으며, 초점이 맞춰져 있는지 알아내는 데 사용한다.

토픽 모델링의 중심 아이디어는 문서는 다양한 단어들의 결합체이고, 단어 간의 특정 관계에 따라 군집화하면 몇몇 토픽들로 유형화될 수 있으며, 단어들이 모여 생성된 토픽들을 조합하면 또 하나의 문서로 구성되는

계층적 구조가 개발되는 것이다(김지은·백순근, 2016). 토픽 모델링을 수행하기 위해서는 연구자들이 토픽의 개수를 미리 지정해야 하는데, 일반적으로 토픽을 임의로 여러 개로 정한 다음, 도출되는 결과 중 타당해 보이는 수를 토픽의 수로 정하거나 복잡계(complexity system)⁴⁾개념을 활용하여 토픽의 개수를 결정할 수 있다(Hornik, & Grün, 2011).

이러한 토픽 모델링은 사회과학 분야에서 널리 활용될 뿐만 아니라 다양한 분야의 연구 동향을 분석하는 데 자주 활용된다(김성근 등, 2016; 한성수·양동우, 2017; 김희섭·강보라, 2019a; 윤정현·박태정, 2019).

라. 감성분석

감성분석은 텍스트에 나타난 감정을 분류하거나 수치화하여 객관적인 정보로 바꾸는 기술을 뜻한다(Jang, Park, & kim, 2015). 긍정과 부정 등 감성에 대한 분석을 ‘sentiment analysis’라 하며 국내에서는 ‘감성분석’이라는 용어로 번역되어 사용되고 있다(오연주·채수환, 2015). 인간은 언어를 이용해 자신의 생각과 느낌을 전달한다. 앞에서 소개했던 토픽 모델링이 ‘텍스트가 다루는 대상’을 파악하는 텍스트 마이닝 기법이었다면, 감성분석은 ‘텍스트에 담겨져 있는 태도’를 추정하는 텍스트 마이닝 기법이다. 토픽 모델링이 텍스트에 내재 되어 있을 것으로 가정되는 토픽을 구현하는 단어들을 추출하여 토픽을 추정하듯, 감성분석 역시 텍스트에 사용된 단어가 내포하는 감정을 추정한다. 예를 들어 ‘선거(election)’라는 단어가 많이 등장하는 논문이 ‘정치(politics)’토픽을 다룰 것으로 예측하듯, ‘좋은(good)’이라는 단어를 많이 사용하는 텍스트는 ‘긍정적(positive)’ 감정을 전달한다고

4) 복잡계(complexity): 수많은 복잡한 자연현상들이 특정한 상호작용에 의해 연결되어 있으며 이러한 상호작용은 비선형적이며 특정한 관계가 있다고 정의한다. 복잡하고 다양한 자연 및 사회현상을 이해하기 위하여 창발현상(emergent behavior)이나 상호작용하는 현상으로 접근할 필요가 생긴다. 여기서 창발 현상은 상호관계에 의해 새로운 현상과 질서가 나타나는 현상을 말한다.

추정할 수 있다. 감성분석이나 토픽 모델링이나 사실 텍스트에 사용된 단어나 표현을 이용해 문서의 특성을 추정한다는 점은 동일하다.

본 연구에서 감성분석을 위해 사용한 R은 전 세계에서 통용되는 툴(tool)로 다양한 언어를 분석할 수 있다. 다만 각각의 언어를 분석하는데 사용하는 패키지는 다르기 때문에 유의해야 한다. R에서 한국어 감성분석을 할 경우 설치해야 하는 패키지는 'KoNLP'라는 함수 패키지고, 한국어 분석을 가능하게 해준다. 그리고 영어 같은 경우는 감성분석을 하는 경우 대중적으로 사용 가능한 감성사전과 텍스트 정제에 필요한 불용어(stopwords) 사전도 있지만, 한국어의 경우는 사용자 기반으로 사전을 구축하여 분석을 하는 경우가 일반적이다. 따라서 본 연구에는 내국인 온라인 리뷰를 분석하기 위한 한국어 텍스트 마이닝을 실시하기 위해 불용어 사전과 감성사전을 사용자 기반으로 구축하고 분석에 사용하였다.

최근에는 리뷰 텍스트에 나타난 사람들의 의견, 감정을 분석하는 감성분석이 주목을 받고 있다. 송보미 등(2011)은 영문 리뷰에 대한 감성분석을 바탕으로 고객이 기대한 품질 수준과 고객이 인지한 품질 수준 사이의 차이를 측정하는 서비스 품질 갭 분석 방법론을 제안하였다. 박경배·하성호(2015)는 영문 리뷰를 대상으로 감성분석을 통한 호텔 평판관리 프레임워크를 제안하였다. 문서(리뷰) 단위 감성분석을 통해 리뷰를 긍정감성 리뷰와 부정감성 리뷰로 구분하고, 부정감성 리뷰에 대해서는 토픽 모델링을 추가 실시해 소비자들의 불평 및 불만사항 유형을 추출하는 방향을 제시하였다. 홍태호·니우한잉·임강·박지영(2018)은 세계 6개 도시에 위치한 호텔들의 온라인 영문 리뷰를 수집하고 R을 활용하여 토픽 모델링과 감성분석을 결합한 모형을 제안하였다. 임영희·김홍범(2019)도 외래관광객의 온라인 영문 리뷰를 활용하여 빈도 분석을 실시하고, 추가적인 텍스트 마이닝 기법을 활용한 감성분석을 통해 만족과 불만족의 단어들을 추출하는 분석을

실시하였다.

이처럼 대부분의 국내 감성분석 연구는 그 수나 활용사례 측면에서 아직 초기단계에 있을 뿐만 아니라 대부분 영문 리뷰를 대상으로 하고 있어 국내호텔 국문 리뷰에 대한 연구 필요성이 제기된다. 공효순·송은지(2014)는 예외적으로 국내호텔 국문 리뷰를 대상으로 감성분석을 수행하였다. 호텔 만족도 평가요소를 외부환경·객실·프런트·식음료·기타 부대시설의 5가지로 나누고 국내 3개 특급호텔에 대한 평가요소별 감성분석을 진행하였다. 최근 에어비앤비 관련 해외연구에서도 감성분석을 활용하였는데, Lawani et al(2019)은 미국 보스턴지역의 에어비앤비를 대상으로 리뷰와 가격 데이터를 활용하여 감성분석을 실시하였다. 따라서 본 연구에서는 아직 연구가 미흡한 국문 리뷰 데이터인 내국인 데이터를 활용하여 감성분석을 실시하고자 한다.

제 3 장 연구 조사 설계

제 1 절 연구 절차 및 방법

이론적 배경을 바탕으로 본 연구는 웹 크롤링 방식으로 수집한 수치 데이터와 텍스트 데이터를 저장하였고 이들 데이터 중 리뷰 데이터는 텍스트 마이닝 기법을 통하여 자연어 처리 분석을 수행하였다. 자연어 처리 분석이 완료된 후, 수치화된 언급 빈도가 높은 단어들을 중심으로 호스트 등급별로 시각화하였고, 전체 리뷰 데이터를 통해서 주제를 도출할 수 있는 토픽 모델링과 리뷰 문서 전체의 감정을 확인할 수 있는 감성분석을 실시하였다. 감성분석에 사용한 감성사전은 사용자 기반 감성사전을 구축하여 에어비앤비 내국인의 리뷰 전체의 긍정과 부정의 비율을 확인할 수 있었다. 그리고 연구자가 선정한 에어비앤비 숙박서비스 선택 속성요인과 전반적 만족도(평점)와의 영향관계를 호스트 등급별 차이로 검증하였다.

기존의 연구들을 살펴본 바와 같이 에어비앤비 숙박서비스의 선택 속성요인들이 전반적 만족도에 유의미한 영향관계가 있다는 것을 확인하였다(권혁진 등, 2017; 류일환 등, 2017; 김진한·강우성, 2018; 김지희·이건웅, 2019; 주위에치·김소연, 2019; 정은별·이형룡, 2020). 기존 에어비앤비에 관한 연구들은 다양한 주제들로 다루어져 왔으며 그 중 ‘선택 속성’, ‘리뷰의 수’, ‘가격의 영향’, ‘호스트의 특성’ 등(권혁진 등, 2017; 김진한, 2019; 정은별·이형룡, 2020)의 각각의 주제들로 기존 설문조사 방식의 계량적 연구 위주로만 이루어져 왔다. 정은별·이형룡(2020)은 최근 에어비앤비를 이용하는 내국인 고객의 선택 속성에 대해 연구를 진행하였는데, 기존 설문조사 방식으로만 진행함에 따른 한계점으로 빅데이터 분석 기법을 활용하여 확장

된 연구의 시도가 필요하다고 판단하였다.

본 연구에서는 이러한 선행연구를 토대로 온라인 데이터에서 수집된 리뷰, 평점, 가격(숙박료), 숙소 지역 등의 자료에서 도출될 수 있는 의미들을 텍스트 마이닝을 통해 확인하고 탐색적 연구에서 밝혀낸 에어비앤비 숙박서비스 이용고객의 온라인 리뷰 선택 속성요인이 전반적 만족도에 미치는 영향과 호스트 등급별 차이가 있는지를 확인하기 위해 다음과 같은 연구과제를 설정하였다.

연구 과제 1: 수집된 숙소 지역 데이터를 활용하여 서울지역 숙소 분포 특성을 확인한다.

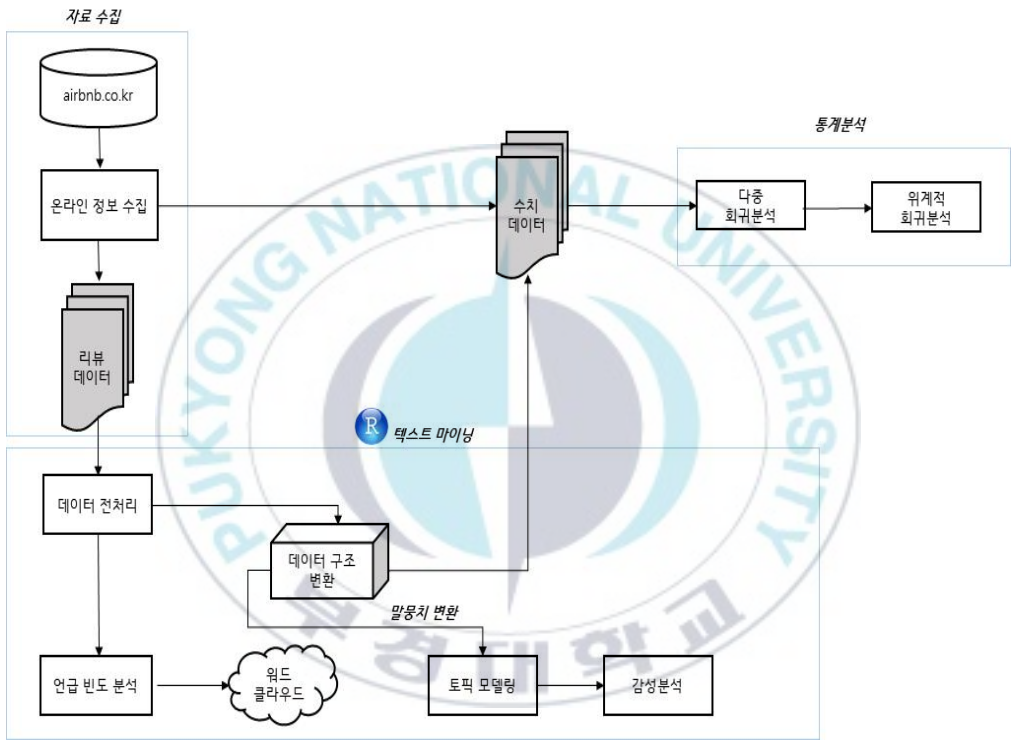
연구 과제 2: 리뷰 텍스트 데이터 내 숨겨진 의미를 도출하기 위해 다양한 텍스트 마이닝 기법을 활용하여 에어비앤비 숙박서비스를 이용한 내국인 고객들의 인식에 대해 고찰한다.

연구 과제 3: 텍스트 마이닝 기법들을 통해 도출된 에어비앤비 숙박서비스 결정(선택) 요인들과 전반적 만족도 간의 영향관계를 검증한다.

연구 과제 4: 호스트 등급에 따른 조절효과를 통해 전반적 만족도 간의 차이를 검증한다.

연구 과제를 분석하기 위하여 기존의 선행연구를 토대로 리뷰 데이터를 수집하였다. 리뷰의 내용들은 여러 에어비앤비 숙박서비스 선택 속성에 관하여 기록되어 있으나 본 연구에서는 에어비앤비가 운영하는 홈페이지에 있는 측정 항목을 참고하여 연구자가 선정한 선택 속성을 리뷰에서 찾아내어 수치 데이터로 변환하였다. 이 수치 데이터를 전반적 만족도 영향관계 분석에 활용하였다.

위와 같은 연구 과제를 해결하기 위해 본 연구의 절차를 <그림 3-1>과 같은 체계로 진행한다.



<그림 3-1> 온라인 리뷰 분석 연구 절차

자료 수집 단계에서는 에어비앤비의 온라인 정보를 웹 크롤러를 통해 수집하여 자료 분석이 가능한 형태로 데이터 전처리한 후 저장한다[①,②].

다음으로 텍스트 마이닝 단계에서는 저장된 리뷰 데이터를 텍스트 마이닝 기법을 활용하여 자연어 처리(NLP: natural language processing) 및 형태소 분석을 실시하여 데이터 전처리(data preprocessing)를 수행하고[③, ④], 가장 먼저 실시하는 언급 빈도 분석(term frequency)[⑤]과 상위 빈도 단어를 활용하여 전체와 호스트 등급별로 분류하여 시각화한 워드 클라우드(word cloud) [⑥]을 실시하고, 형태소 분석으로 분류된 리뷰 데이터를 다시 말뭉치(corpus)형식으로 변환하는데 비구조화 텍스트 데이터를 구조화 자료형태인 문서×단어 행렬(DTM: document-term matrix)로 변환하여 분석에 적합한 형태로 데이터를 가공한다. 이후, 텍스트 마이닝 분석 기법 중 토픽 모델링[⑦]과 감성분석[⑧]을 실시한다.

다음으로 온라인 리뷰 정보의 수치 데이터를 활용하여 서울지역의 에어비앤비 숙소 분포 특성을 파악[⑨]한다. 그리고 수치 데이터로 변환된 선택 속성 데이터를 통해 전반적 만족도와 영향관계를 분석하기 위해 다중회귀분석을 실시한다[⑩]. 그리고 호스트 등급별 조절효과를 검증하기 위해 위계적 회귀분석을 실시한다[⑪].

<그림 3-2>는 위의 연구 흐름에 따른 목적 및 방법에 관한 설명을 그림으로 작성한 것이다.

진행 단계	주요 과제	연구 방법 및 분석
1단계 문헌 고찰	①구성개념 범위 고찰	- 에어비앤비에 대한 이해 - 전반적 만족도 개념 이해 - 온라인 리뷰에 대한 이해 - 텍스트 마이닝 기법 종류의 이해
2단계 자료 수집	②자료 수집 대상 선정	- 에어비앤비 관련 기사와 선행연구 동향 파악 - 에어비앤비 서울지역 내국인 온라인 리뷰
	③자료의 수집	- 웹 크롤러를 통한 자료수집 - 리뷰, 평점(rating), 가격(숙박료), 숙소 지역, 호스트 등급
3단계 텍스트 마이닝	④데이터 전처리	- 텍스트 마이닝 기법을 활용한 온라인 리뷰 자연어처리 및 형태소 분석 -구두점 제거 -공백제거 -숫자 제거 -불용어 제거
	⑤언급 빈도 분석	개별 문서 형태소 분석 후 빈도 추출
	⑥워드 클라우드	- 언급 빈도 단어 상위 100개 추출 시각화 - 호스트 등급 별 상위 20위 워드 클라우드 비교
	⑦토픽 모델링	- 리뷰 문서의 구조를 말뭉치로 변환 - 말뭉치 문서를 DTM행렬로 변환 - 문서 전체의 그룹별 토픽 주제 도출
	⑧감성분석	- 사용자 기반의 감성사전 구축 - 말뭉치 문서에 감성사전을 포함하여 분석 - 긍정과 부정의 비율 검증
4단계 통계 분석	⑨에어비앤비 분포 특성	서울 지역의 에어비앤비 숙소 분포 특성 파악
	⑩다중회귀분석	에어비앤비 숙박선택 속성 요인들과 전반적 만족도 간의 영향관계 검증
	⑪위계적 회귀분석	호스트 등급별 조절효과 검증

〈그림 3-2〉 연구 흐름에 따른 목적 및 방법

자료수집과 선택 요인측정을 위한 표본 설계와 분석 방법 요약은 <표 3-1>과 같다.

〈표 3-1〉 연구 절차의 요약

분석 과정	데이터 수집	에어비앤비 숙박선택 속성요인 도출	선택 속성요인이 전반적 만족도 간에 미치는 영향, 호스트 등급별 차이검증
분석에 사용된 데이터	에어비앤비 공식 사이트 (airbnb.co.kr)	수집한 리뷰 데이터에서 추출한 단어(keyword)	• 변환된 수치 데이터 • 평점(rating)
유효표본	• 리뷰: 1,545개	2020.06.03.~2020.06.04. 1박 성인 기준으로 설정하고 홈페이지에 노출된 308개의 호스트 숙소에서 리뷰가 없는 숙소를 제외하여 281개의 데이터로 숙소 지역 분포 특성 확인 281개 숙소에서 국문 리뷰만 수집하기 위해 다른 외국어만 있는 숙소를 제외하여 수집, 총 146개의 내국인이 이용한 호스트 숙소에서 한국어 텍스트 리뷰 1,545개 수집	
활용 분석	JAVA를 이용한 웹 크롤링	텍스트 마이닝 기법 • 언급 빈도 분석 • 워드 클라우드 • 토픽 모델링 • 감성분석	• 다중회귀분석 • 위계적 회귀분석
결과분석에 사용된 프로그램	• Java, Deep Learning • Excel	• R • Excel	• SPSS 25.0 • Excel

<표 3-1>를 살펴보면 서울의 에어비앤비 숙박서비스를 이용한 호스트 숙소 146곳의 리뷰 1,545개의 표본을 수집하였다. 분석에는 R과 excel을 이용하여 텍스트 마이닝 기법을 통해 에어비앤비 숙박선택 속성요인을 도출하고, 도출된 요인을 연구자가 자체 코딩하여 수치 변수화하고 SPSS 25.0을 활용하여 다중회귀분석을 실시하고, 호스트 등급별 조절효과를 검증하기 위해 위계적 회귀분석을 실시한다.

제 2 절 자료수집 대상 선정

자료수집 대상 선정으로 에어비앤비의 내부 데이터와 설문조사 결과에서 ‘2018년 한국경제 파급효과’에 대한 분석을 실시한 결과 2018년 한 해 동안 1조 3,700억 원(12억 5,000달러)의 경제적 효과를 보았고, 이것은 전 세계 에어비앤비 운영 지역 중 16위에 해당하는 성과이다(에어비앤비, 2019). 그리고 2018년 10월 10일 ‘세계지식포럼’의 연사에서 에어비앤비 아시아·태평양 정책 총괄 대표는 한국의 경우 2018년 1월 1일 기준 전국 에어비앤비 숙소는 3만 7,100곳이라고 했다. 2017년 에어비앤비를 방문한 게스트는 188만 8,000명에 달했고, 그 중 내국인은 123만 1,500명으로 전체의 65%나 차지한다고 했다(안갑성, 2018). 위의 조사 데이터 결과와 같이 에어비앤비는 한국경제에 큰 영향력을 갖추고 있다고 판단하고 충분히 연구할 가치가 있어 본 연구에서의 자료수집 대상을 에어비앤비로 선정한다.

다음으로 본 연구에 사용될 데이터를 수집하기 위해 수집지역을 선정한다. 최근 에어비앤비 주제의 해외 연구사례를 보면 Cheng, & Jin(2019)는 호주 시드니의 에어비앤비의 특징에 대해 연구하였고, Lawani et al(2019)는 미국 보스턴 지역을 대상으로 온라인 리뷰를 수집하여 리뷰와 가격에 대한 감성분석을 실시하였다. 이와 같이 최근 해외 연구사례에서 보면 특정 지역을 연구 대상으로 선정하는데 유명한 관광지역이라는 특징을 알 수 있다. 국내 선행연구 사례를 보면 허지정·노승철(2018)은 에어비앤비 서울 지역을 대상으로 에어비앤비 특성과 공간분포 분석을 실시하였고, 박소민·양승우(2020)은 서울지역의 에어비앤비 입지적 특성에 대한 연구를 진행했다. 두 연구 모두 지리적 특성을 중심으로 연구가 진행된 것이 특징인데, 아직 서울지역을 대상으로 한 온라인 리뷰 분석은 거의 없는 것을 확인할

수 있다. 이에 본 연구에서는 해외사례를 참고하여 유명 관광지를 가장 많이 분포되어있는 ‘서울’을 수집지역으로 선정한다.

최근 해외연구에서는 에어비앤비의 온라인 리뷰 관련정보를 바탕으로 한 연구들이 꾸준히 이루어지고 있다(Ert et al, 2016; Bridges, & Vásquez, 2018; Lawani et al, 2019; Cheng, & Jin, 2019). 이와 같이 에어비앤비 온라인 리뷰를 활용한 연구가 다양한 주제로 이루어지고 있지만, 국내에서는 에어비앤비 온라인 리뷰를 활용한 연구들은 해외 지역의 리뷰를 사용한 연구(류일한 등, 2017)로, 국내 내국인 중심의 국문 리뷰를 분석한 사례는 거의 없는 실정이다. 내국인 연구가 최근 이루어졌지만, 이 역시도 기존 설문지 방식의 연구로 진행했다(정은별·이형룡, 2020).

따라서 본 연구에서는 에어비앤비의 온라인 리뷰에서 내국인 관광객이 호스트 숙소를 이용한 후 작성된 온라인 리뷰를 조사 수집 대상으로 한다.

아래 <그림 3-3>은 에어비앤비의 온라인 리뷰 형태이며 텍스트와 평점(rating)으로 이루어진 구조를 확인할 수 있다.



<그림 3-3> 에어비앤비 리뷰 및 평점 형태

제 3 절 자료수집 및 전처리

1. 웹 크롤러(web crawler)를 통한 자료수집

웹 크롤러(web crawler)는 리뷰, 상품평 등과 같은 방대한 웹 문서를 제공하는 웹에서 특정 사이트의 웹 문서를 자동으로 수집하는 기술을 뜻한다(김광영·이원구·이민호·윤화목·신성호, 2011). 웹 크롤러를 통해 자료를 수집하여 이를 활용한 연구로 도해용·이애주·강한훈(2012)은 레스토랑 평가 사이트인 ‘윙스푼’의 리뷰를 수집하여 레스토랑 만족도 평가에 관한 연구를 하였고, 심영석·김홍범(2016)은 웹 크롤러를 통해 서울의 관광지 63곳의 외래관광객의 온라인 리뷰를 수집하여 관광지 이미지 구성요인을 검증하였다. 류일한 등(2017)은 에어비앤비 공식 페이지에서 리뷰를 수집하여 호스트의 리뷰 피드백과 위시 리스트를 종속변수로 하는 구매 의도와의 영향관계 연구를 하였다. 임영희·김홍범(2019)은 웹 크롤러를 활용하여 트립어드 바이저 리뷰 자료를 비즈니스호텔, 리조트호텔, 공항호텔별로 수집하여 감성분석과 회귀분석에 활용하였다.

따라서 본 연구에서도 웹 크롤러를 통하여 자료를 수집하였으며, 아래 <그림 3-4>는 에어비앤비 서울 숙소 1개를 선정하여 웹 크롤링하고, 수집된 리뷰를 확인할 수 있다. 홈페이지에서 보여지는 정보는 리뷰들, 게스트 이름, 호스트 숙소 지역, 슈퍼 호스트인지 일반 호스트인지 외에도 다양한 정보가 수집 가능하지만, 분석 전 데이터를 전처리할 때 더 많은 시간을 할애해야 하기 때문에 가능한 한 효율적으로 전처리하기 위해서 연구자는 연구 목적에 맞게 계획하고 주제에 맞는 데이터 정보만을 수집한다. 그래서

본 연구는 숙소 지역, 슈퍼 호스트 유무, 텍스트 리뷰, 가격(숙박료)과 홈페이지 검색 상에는 노출되어 있지 않지만, 웹 크롤링을 한 결과에는 게스트들의 평점이 수집된 것을 확인하여 본 연구에서는 표본을 각각의 게스트의 리뷰와 평점이 주어진 데이터 정보로 분석에 활용한다. <그림 3-5>는 수집된 데이터의 정제 전 raw data 자료이다.

The image shows a screenshot of an Airbnb listing page for 'Seochon Night' with a blue arrow pointing to a JSON data structure on the right. The listing shows reviews from users 'Joonyuk' and 'Su Kyung' with their respective comments. The JSON data on the right contains details about the listing, including location, price, and reviews.

```

{
  "url": "https://api.airbnb.com/v2/pdp_listing_details/399967157_format=for_native",
  "name": "Seochon Night",
  "stars": 4.5,
  "numberOfGuests": 2,
  "address": "Chebu-dong, Jongno-gu, Seoul, South Korea",
  "roomType": "Entire house",
  "location": {
    "lat": 37.57881,
    "lng": 126.96999
  },
  "reviews": [
    {
      "author": {
        "firstName": "Joonyuk",
        "hasProfilePic": true,
        "id": 200318043,
        "pictureUrl": "https://a0.muscache.com/im/pictures/user/147d0762-5913-4711-bd4a-c580981fe48d.jpg?aki_policy=profile_x_medium",
        "smartName": "Joonyuk",
        "thumbnailUrl": "https://a0.muscache.com/im/pictures/user/147d0762-5913-4711-bd4a-c580981fe48d.jpg?aki_policy=profile_small"
      },
      "authorId": 200318043,
      "canBeEdited": false,
      "comments": "개끗하구 이쁜어요. 위치는 정말로 최고여서 주변에 맛집이나 이쁜 카페들이 정말로 많아서 놀아다니기 막 좋았어요.",
      "createdAt": "2020-01-22T07:57:57Z",
      "id": 596239271,
      "idStr": "596239271",
      "listingId": "39996715",
      "recipientId": "202192157",
      "collectionTag": null,
      "listing": {
        "id": "39996715",
        "name": "Seochon Night",
        "rating": 5
      },
      "recipient": {
        "firstName": "Su Kyung",
        "hasProfilePic": true,
        "id": 202192157,
        "pictureUrl": "https://a0.muscache.com/im/pictures/user/643722ec-be5f-4dbc-a279-80388aa1b109.jpg?aki_policy=profile_x_medium",
        "smartName": "Su Kyung",
        "thumbnailUrl": "https://a0.muscache.com/im/pictures/user/643722ec-be5f-4dbc-a279-80388aa1b109.jpg?aki_policy=profile_small"
      }
    }
  ]
}

```

<그림 3-4> 에어비앤비 웹 크롤링 수집 샘플 데이터

Listing ID	Name	Location	Other Details
19227515	BRIGGS HOSTEL YEONNAM - DOUBLE	Seoul, South Korea	Entire house, 37.55903
19227516	Jiang's Han River view Donggug	Donggug, Seoul, South Korea	Entire apartment, 37.51388
19227517	Like your home	Seoul, South Korea	Entire house, 37.54449
19227518	N Seoul Tower's Modern Chic 'HISTAR' in Yongsang-gu	Yongsang-gu, Seoul, South Korea	Entire house, 37.5183
19227519	Private duplex terrace house near Namsan	Seoul, South Korea	Entire house, 37.551
19227520	Jamul's Comestay NZ - Double bedroom	Songpa-gu, Seoul, South Korea	Private room, 37.53445
19227521	이강호의 서울 (강남구)	Jongno-gu, Seoul, South Korea	Entire house, 37.57948
19227522	#1 Connected to Maple St. Juhui River view Hongdae	Mapo-gu, Seoul, South Korea	Entire apartment, 37.54274
19227523	수리터의 1st 레스텔 & 부엌은 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227524	OPEN SPACE in the heart of Hongdae	Donggug-dong, Mapo-gu, Seoul, South Korea	Entire house, 37.55666
19227525	STUDIO, Gangnam #1 SUPER CLEAN OPEN SPACE	Seoul, South Korea	Entire house, 37.55455
19227526	Smart location in the heart of Seoul	Seoul, South Korea	Entire apartment, 37.55449
19227527	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227528	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227529	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227530	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227531	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227532	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227533	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227534	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227535	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227536	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227537	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227538	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227539	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227540	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227541	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227542	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227543	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227544	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227545	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227546	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227547	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227548	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227549	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392
19227550	이강호의 서울 (강남구) - 1인만인 신축 건물	Seoul, South Korea	Entire apartment, 37.47392

<그림 3-5> 최초 수집된 raw data 형태

이러한 웹 크롤링 기법을 통해 서울지역의 308곳의 호스트 숙소를 수집하였고, 그 중 리뷰가 없는 숙소를 제외하고 총 281개의 숙소의 데이터 표본을 수집, 저장 후 R을 사용하여 지역별 데이터 분포 특성을 확인한다. 그 후 국문 리뷰를 제외한 외국어 리뷰만 있는 숙소는 최종 표본에서 제외하여 내국인 국문 리뷰만을 선정하여 최종 숙소는 146개이고, 총 게스트의 리뷰 수는 1,545개로 선정하고 리뷰에서 수집한 수치 데이터와 연구자가 변수화한 수치 데이터를 합한 새로운 최종 파일을 엑셀에 저장한다. 저장한 파일의 형식은 아래 <그림 3-6>과 같다.

A	B	C	D	E	F	G	H	I	J
	review	score	host	price	clean	commu	checkin	accuracy	location
1	포근한 침구가 너무 좋았습니다. 하지만 주택이 아닌 곳 개조된 공간이기 때문에 공기가 찻어요. 침구가 자단을 잘 해워서 감기에 안 걸린것 같은데, 창문 틈으로 들어오는 바람 등은 조금의 보온이 더 필요한 것 같아요. 화장실이나 개수구를 통해 들어오는 하수구 냄새가 아쉬웠습니다. 하지만 모던한 공간의 분위기는 너무 좋았습니다.	4	1	75	0	0	0	0	0
2	여태까지 다녀온 에어비엔비 중 최고! 숙소였어요 물리 나가기에는 시간이 없고 서울에서 새해 특별한 추억을 만들어보고 싶어 찾아보면 1층에서 전시도 하고 문화특화 공간으로 되어있어 핫한 곳 같았어요 지어진지 얼마 되지않아 굉장히 깔끔했구요 사진이랑 똑같았어요 화이트와 우드톤이 따뜻한 느낌을 주고 방 곳곳에 있는 조명과 거실 을 코티지가 또 듯하게 감싸주어 너무 인상적이었어요. 미국 친구집 놀러온 듯한 느낌이 들고 편안했 습니다. 지금 또 생각나네요 ㅎㅎ 호스트님의 세심한 정성과 배려가 돋보였고 친절하셔서 감사했습니다 연날등 걸지 않아서 눈이 안보이고 골목 구경하기에 시간가는줄 몰랐고 저 네에는 불지지도 조종해서 좋았습니다. 가짜역 근처라 기차 소리가 종종 들리긴했지만 밤에 침실에서는 들리지 않아서 예민 한 저도 포근하니 꿀잠갔어요~ 서울에서 특별한 추억과 숙박을 해보고싶다면 꼭! 추천합니다 만족스러워 일게 남기고 갑니다 ~ 다음 또 올릴게요^^	5	1	75	1	1	0	1	1
3	연날등 조립과는 다르게 떨어져있어서 그런지 여유로우며 아기자기한 샵과 카페들 이 주변에 있어서 이용하기에도 편리했습니다. 뿐만 아니라 정원이나 주변을 걸으며 보이는 오래된 동네의 모습도 함께 있어서 그런지 옛 동네의 정겨움을 느낄수도 있 었습니다. 울 소재로 만들어진 조명이나 베딩, 쿠션 등이 돋보였습니다. 특히 너무나도 따뜻하게 아늑했던 코타즈와 너무나도 촉감이 부드러웠던 슬리퍼가 유독 기억에 남았네요. 사고 싶더라고요!								
	화장실, 침구 모두 깨끗했고, 호스트분의 친절하고도 세심한 안내 또한 좋았습니다. 주변에 개찰구, 식당들이 있어서 여행보았을 때도 면 코레일 추천해주시어서 편하게 잘								

〈그림 3-6〉 데이터 분석 전 최종 데이터 파일 구성

2. 자료 전처리(data preprocessing)

텍스트 마이닝에서 전처리는 일반적인 텍스트 데이터들을 컴퓨터가 처리하기 쉽도록 변환하는 작업으로 특정 단어와 관련된 문서들을 신속하게 검색할 수 있도록 인덱스(index) 파일을 만드는 과정이며 수집한 데이터를 분석하기 위해서는 전처리가 필수적이다(김근형·오성열, 2009). 자연어 처리는 텍스트 마이닝한 데이터에 대하여 불용어 처리, 구두점 제거, 숫자 제거, 공백 제거를 실시하는 것이다(Hotho et al, 2005).

본 연구는 내국인을 중심으로 한국어 텍스트를 분석하는 것이 목적으로, 이전 선행연구의 영문 리뷰에서 사용했던 R 함수 패키지는 적용이 불가하여, 한국어 텍스트 분석 툴을 사용해 전처리 과정과 형태소 분석을 실시하였다. R에서는 한국어 텍스트를 돕기 위한 패키지가 있는데, ‘KoNLP’라는 패키지를 다운로드 받아야 하는데, 최근 R버전이 업그레이드 되면서 구버

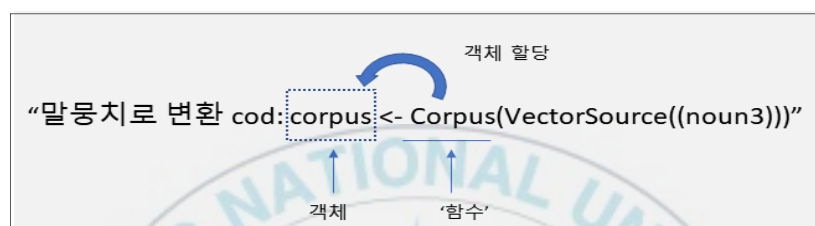
전인 ‘KoNLP’설치를 수동으로 처리하여야 하고, 반드시 이 패키지를 설치하기 전 ‘Java’를 설치해야 하는 번거로움이 있지만, 현재는 이 틀 말고는 방법이 없다. 이 의미는 R에서 한국어 텍스트를 분석하는 방법은 영어 텍스트에 비해 조금 번거로운 단점이 있다. 그리고 한국어 텍스트를 분석할 때 불용어를 제거할 때는 영어와 같이 ‘SMART’나 ‘en’과 같은 불용어 사전이 없어서 본 연구에서는 연구자가 1,545개의 리뷰 전체를 이해하고 파악한 후 ‘사용자 불용어 사전’을 구축하여 전체 데이터에서 불용어를 먼저 제거하고, 구두점, 공백 등을 제거 후 형태소 분석을 실시한다. ‘KoNLP’ 패키지 안에는 한국어 텍스트에서 명사만을 추출해주는 함수 ‘extractNoun’과 품사 분석을 해주는 함수 ‘SimplePos09’를 활용하여 형태소 분석을 간단히 할 수 있다. 여러 가지 방법으로 전처리 후 추가로 단어를 수정할 필요가 있을 때는 R 내장함수 ‘gsub’을 사용하여 수정하거나 통합하거나 삭제할 수 있다.

아래 <그림 3-7>은 본 연구 에어비앤비 한국어 텍스트를 전처리하고 형태소 분석하는 과정이다.

```

1 library(KoNLP)
2
3 getwd()
4
5 setwd("C:/Users/sooyha/Documents")
6
7 df_acsent <- readLines("airbnbreviews.txt")
8
9 save(df_acsent, file = "df_acsent.rda")
10
11 load("df_acsent.rda")
12 df_acsent
13
14 comments <- gsub("<div>.*</div>", "", df_acsent)
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
```

형태소 분석이 끝난 데이터는 언급 빈도 분석과 워드 클라우드 시각화가 가능하다. 그리고 본 연구에서는 단어 각각의 의미를 도출하는 연구와 함께 문서 전체 ‘말뭉치(corpus)’라고도 하는 1,545개 각각의 리뷰를 하나의 전체 말뭉치로 변환해 주어야 이후 진행될 분석이 가능하다. 아래 <그림 3-8> 함수식은 본 연구에서 사용한 말뭉치로 변환하는 함수식이다.



<그림 3-8> ‘말뭉치(corpus)’ 데이터로 구조 변환 R 함수식

위의 cod에서처럼 연구자가 분석 과정에서 객체 할당한 오브젝트(object)를 괄호안에 넣고 ‘corpus’ 함수를 사용하여 말뭉치로 다시 데이터를 변환해 준다. 그리고 하나의 문서로 변환된 데이터를 문서×단어 행렬로 다시 변환해 주어 토픽 모델링과 전체 데이터의 긍정과 부정의 비율을 확인하는 감성 분석을 진행할 수 있다. 아래의 cod에서는 말뭉치 ‘corpus’ 객체 데이터를 다시 불러오고, 구두점을 바로 제거하는 정규 표현식 ‘removePunctuation’과 사용자 불용어 사전을 불러오는 함수 ‘stopwords = stw’ 설정해 주고, 숫자 제거 ‘removeNumbers’ 하고 가중치는 단어 빈도(tf)값을 설정한 것을 ‘DocumentTermMatrix’ 함수를 적용해 처리하고 마지막 ‘dtm’객체에 할당하는 과정이다. R에서의 dtm을 만들 수 있는 cod는 다음과 같다.

```
“dtm <- DocumentTermMatrix(corpus, control=list(
  removePunctuation=TRUE, stopwords = stw,
  removeNumbers=TRUE, weighting=weightTf))”
```

자연어 처리된 텍스트 데이터는 비계량 텍스트 형태이기 때문에, 분석이 가능하도록 계량 자료로 수치화 변환시켜주어야 한다. 본 연구에서는 선행 연구 고찰과 텍스트 마이닝에서 도출된 결과를 바탕으로 다음 6개의 요인으로 선정하였다. 선정된 6개의 요인은 ‘가격(숙박료)’, ‘청결성’, ‘소통’, ‘체크인’, ‘정확성’, ‘위치’가 전반적 만족도 간에 미치는 영향을 다중회귀분석을 실시하였으며, 분석을 위해 선정된 요인들의 데이터를 수치 데이터로 변환하였다. 기존 수집된 가격, 평점은 수집된 값을 그대로 사용하였다. 그리고 위계적 회귀분석을 통해 ‘일반 호스트’와 ‘슈퍼 호스트’의 조절효과에 따른 선택 속성요인들이 전반적 만족도에 영향을 주는지 검증하기 위해 본 연구에서 사용한 측정변수의 조작적 정의를 정하고 통계분석을 할 수 있도록 엑셀 파일 형식으로 저장한다.

아래 <표 3-2>는 에어비앤비 내국인의 전반적 만족도에 영향을 미치는 측정변수의 정의이다.

<표 3-2> 측정변수의 정의

변수	선정기준	측정 범위	척도
호스트	에어비앤비는 호스트 체계를 등급화하여 ‘일반 호스트’와 ‘슈퍼 호스트’로 분류하고 있음	연구자가 0 과 1 로 설정 ‘일반 호스트’ : ‘0’ ‘슈퍼 호스트’ : ‘1’	명목척도
청결성	sample review: “화장실, 침구 모두 깨끗했고, 호스트분의 친절하고도 세심한 안내 또한 좋았습니다. 주변에 괜찮은 식당들이 있는지 여쭙보았을 때도 몇 군데를 추천해주셔서 편하게 잘 이용할 수 있었구요. 식당 또한 너무 만족스러웠구요! 세심한 노고와 정성이 온전히 느껴지는 공간과 호스트였습니다. 다음에 또 방문하고 싶습니다^^ ”	리뷰 내용에 청결성과 관련된 단어 유무에 따른 ‘0’, ‘1’ 코딩 있으면 ‘1’, 없으면 ‘0’	명목척도
가격(숙박료)	웹 크롤링으로 수집한 호스트 숙소의 링크를 개별 접속하여 각각의 호스트의 1박 기준 가격 확인	숙소별 가격 책정으로 같은 숙소의 개별 리뷰들은 같은 가격으로 코딩	비율척도
위치	sample review: “호스트님의 빠른응대가 정말 감사했습니다. 롯데월드몰1분거리인 위치가 정말 좋았고, 들어서는 순간부터 별이 잘드는 기분이 좋아지는 곳이었습니 다. 특히 탁트인 롯데타워 뷰와 감각적인 인테리어가 훌륭합니다.”	리뷰 내용에 위치와 관련된 단어 유무에 따른 ‘0’, ‘1’ 코딩 있으면 ‘1’, 없으면 ‘0’	명목척도

변수	선정기준	측정 범위	척도
전반적 만족도(평점)	웹 크롤링을 통해 수집된 데이터로 개별 리뷰어들이 측정한 호스트 숙소 이용 후 작성한 점수 예상보다 훨씬 좋았어요 예상보다 약간 나았어요 예상했던 것과 비슷했어요 예상보다 약간 별로였어요 예상보다 훨씬 별로였어요	에어비앤비 자체 설문지 측정 항목 5점 척도	등간척도
소통	sample review: “한달 넘게 장기로 지냈는데 깨끗하고(이불 완전 깨끗) 위치도 좋아서 대중교통 이용하기 편했어요.방음도 잘되고 호스트님의 신속한 피드백과 배려 덕분에 불편한거 하나 없이 잘 지내다가 갑니다. 기회가 있다면 재방문 할 의사 100%예요.”	리뷰 내용에 소통과 관련된 단어 유무에 따른 ‘0’, ‘1’ 코딩 있으면 ‘1’, 없으면 ‘0’	명목척도
정확성	sample review: “사진처럼 깔끔하고 따뜻하기도 했습니다! 집도 어렵지 않게 찾아갈 수 있었고, 부족한 것 없이 아늑한 숙소였습니다:)”	리뷰 내용에 정확성과 관련된 단어 유무에 따른 ‘0’, ‘1’ 코딩 있으면 ‘1’, 없으면 ‘0’	명목척도
체크인	“체크인 절차가 간단하게 편했어요”	리뷰 내용에 체크인과 관련된 단어 유무에 따른 ‘0’, ‘1’ 코딩 있으면 ‘1’, 없으면 ‘0’	명목척도

그리고 다음 <그림 3-9>와 같이 수치 데이터를 작성한다.

	A	B	C	D	E	F	G	H	I	J
	rating	revisit	host	price	clean	commu	checkin	accuracy	location	satisfaction
1										
2	4	0	1	75	0	0	0	0	0	0
3	5	1	1	75	1	1	0	1	1	1
4	5	1	1	75	1	1	0	0	0	1
5	2	0	1	75	0	1	0	1	1	1
6	5	1	1	75	0	0	0	0	0	1
7	5	1	1	75	0	0	0	0	0	0
8	4	0	1	75	1	0	0	0	0	1
9	5	0	1	75	1	0	0	0	0	1
10	5	1	1	75	0	0	0	0	0	0
11	5	0	1	75	0	0	0	0	0	0
12	4	0	1	75	0	0	0	0	0	0
13	4	0	1	75	0	0	0	1	1	1
14	5	1	1	75	0	1	0	1	0	0
15	5	0	1	75	0	0	0	0	0	1
16	5	0	1	75	1	0	0	0	0	0
17	5	1	1	75	0	0	0	0	0	1
18	5	1	1	75	1	1	0	0	0	1
19	5	1	1	75	0	0	0	0	0	1
20	5	0	0	25	0	1	0	0	0	0
21	5	0	0	25	0	1	0	0	0	0
22	4	0	0	25	1	0	0	0	0	0
23	5	0	0	25	1	0	0	0	0	0
24	5	0	0	25	0	0	0	0	0	0

<그림 3-9> 수치 데이터 파일

제 4 장 연구 결과

제 1 절 에어비앤비 분포 특성

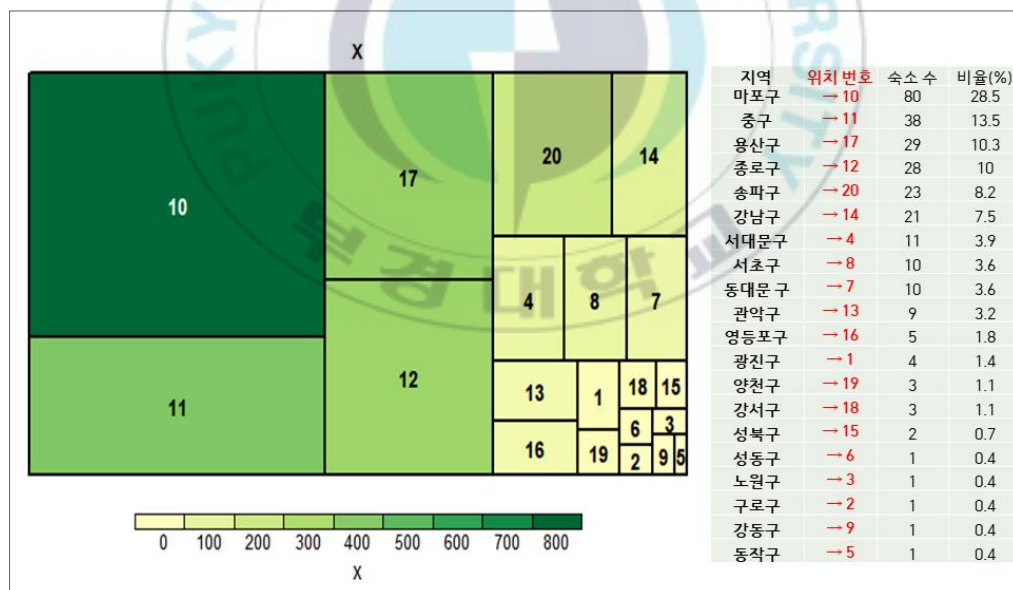
1. 데이터 분포 특성

본 연구에서 사용된 에어비앤비 데이터는 웹 크롤러를 이용하여 에어비앤비 공식 사이트(airbnb.co.kr)을 통해 수집하였다. 수집 기간은 서울의 전체 내·외국인 이용자 호스트 숙소 중 2020년 6월 3일부터 2020년 6월 4일까지 성인 1박 기준 숙소 예약 설정 후 홈페이지에 노출된 숙소를 내국인이 이용하고 남긴 리뷰가 작성되어있는 호스트 숙소만을 선정하였으며, 수집된 전체 308개의 숙소에서 리뷰가 없는 숙소를 제외하여 총 281개의 호스트 숙소를 대상으로 서울지역 분포 특성을 살펴보고, 그 중 외국어 리뷰만 있는 곳은 제외하고 남은 숙소는 총 146개의 숙소로 리뷰 분석에 사용하였다. 이 146개의 내국인 리뷰 데이터는 총 1,545건으로 본 연구의 분석을 수행하는 데이터로 활용하였다. 본 연구의 수집 결과는 <표 4-1>과 같다.

<표 4-1> 에어비앤비 데이터 수집 결과

	전체 데이터	분석 데이터
수집 대상 사이트	www.airbnb.co.kr	
수집 기간 설정	예약 설정: 2020년 6월 3일 ~ 2020년 6월 4일 성인 1박 기준 예약 기간 설정에 따른 예약이 가능한 호스트 숙소 노출 노출된 숙소 전체 수집	
수집 방법	웹 크롤링	
호스트 수	308곳	281곳 / 146곳
리뷰 수	-	1,545

분석 대상은 총 281개의 호스트 숙소이며, 281개의 서울 분포표<그림 4-2>의 결과를 살펴보면 ‘마포구(80개)’에 가장 많은 28.5%의 분포율을 확인할 수 있었다. 그 다음으로 ‘중구(38개, 13.5%)’, ‘용산구(29개, 10.3%)’, ‘종로구(28개, 10%)’, ‘송파구(23개, 8.2%)’, ‘강남구(21개, 7.5%)’ 등의 순으로 결과가 나타났다. 이는 박소민·양승우(2020)의 선행연구와 비슷한 결과를 확인할 수 있었는데, 에어비앤비 숙소 분포표를 보면 상위지역으로 선정된 지역들이 관광객들이 선호하는 관광지역에 많이 분포되어있는 것을 확인할 수 있었다. 관광을 목적으로 서울을 방문하였거나, 이전 관광경험으로 재방문을 할 경우(한경·양위주, 2015)와 같은 관광을 중심으로 에어비앤비 숙소가 많이 분포되어있는 것은 서울지역의 에어비앤비는 관광지의 영향을 많이 받는 것으로 판단할 수 있다.



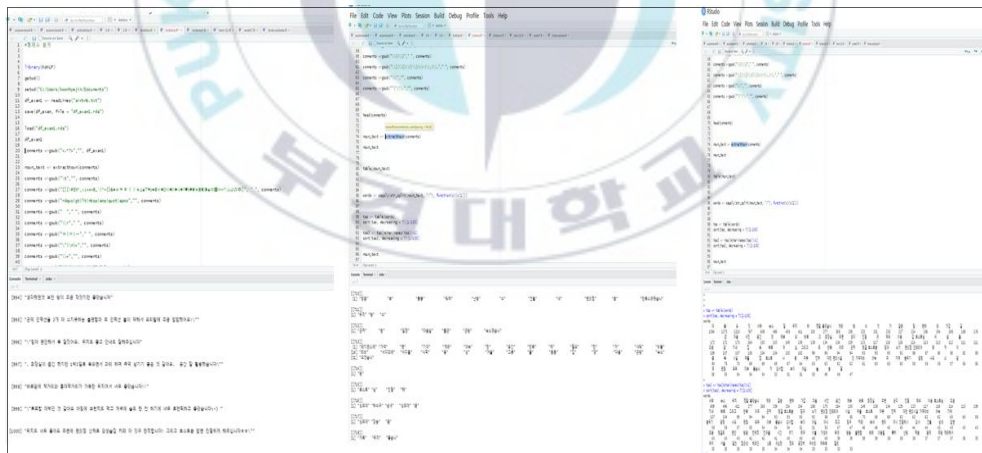
〈그림 4-2〉 나무지도를 활용한 서울 분포표

제 2 절 연구 과제 분석

다음은 연구 과제를 검증하기 위한 분석 절차를 진행하였다. 웹 크롤러를 통해 수집된 1,545건의 온라인 리뷰의 텍스트 데이터에서 상위 빈도 키워드를 도출하여 그 단어들의 영향력을 파악하기 위해 언급 빈도 분석과 워드 클라우드 분석을 실시하였고, 분리된 리뷰 문서를 하나의 말뭉치(corpus) 형태로 변환 후 전체 리뷰 문서의 주제를 파악하기 위한 토픽 모델링과 전체 문서가 의미하는 긍정과 부정을 파악하기 위한 감성분석을 실시하였다. 그리고 선행연구의 고찰과 탐색적인 연구를 통해 도출된 결과를 바탕으로 선정된 에어비앤비 숙소의 선택 속성요인들과 전반적 만족도 간의 영향관계를 다중회귀분석을 통해 검증하고, 위계적 회귀분석을 활용하여 호스트 등급별 조절효과를 검증하였다.

1. 언급 빈도 분석

언급 빈도 분석을 하기 위한 데이터 형태의 변환 과정을 <그림 4-3>에서 확인할 수 있다. raw data의 경우 하나의 리뷰는 여러 문장으로 구성되어 있는데 그것을 한 문장 단위로 분리 후 리스트(list) 형태로 만들어 보여주고, 형태소 분석을 통해 구두점 제거, 공백 제거, 숫자를 제거하여 문장 단위에서 단어 단위로 쪼개 준 후 상위 빈도를 추출한다. 처음 빈도 분석결과에는 불용어가 많이 포함되어 있어 이를 연구자 판단에 따라 의미 있는 형태로 수정하거나 제거하는 단계를 거치는데, 이때 사용한 내장함수는 ‘gsub’ 함수를 사용하였다. 그 결과 아래 <표 4-2>과 같이 총 1,545개의 리뷰 안에서 여러 번 언급된 단어의 상위 100위 언급 빈도 분석 결과이다.



<그림 4-3> 리스트(list) 형태로 분리하는 과정

〈표 4-2〉 상위 100위 언급 빈도 분석 결과

순위	단어(키워드)	빈도	순위	단어(키워드)	빈도	순위	단어(키워드)	빈도
1	위치	450	35	시간	57	69	부분	32
2	깨끗	316	36	필요	55	70	지하철	32
3	깔끔	313	37	답변	54	71	걱정	31
4	호스트	279	38	불편	54	72	도구	31
5	친절	220	39	건물	52	73	숙박	31
6	이용	205	40	냄새	51	74	아침	31
7	공간	170	41	최고	51	75	쾌적한	31
8	청결	176	42	방문	50	76	기분	30
9	사진	170	43	식기	49	77	내부	30
10	만족	146	44	침구	48	78	마음	30
11	생각	138	45	카페	47	79	완전	30
12	화장실	134	46	전체	44	80	침구류	30
13	주변	133	47	이불	43	81	완벽	29
14	따뜻	110	48	훌륭한	43	82	지하철역	29
15	침대	102	49	거리	42	83	폭신	29
16	감사	101	50	구비	41	84	파티	28
17	아늑함	101	51	난방	41	85	안내	27
18	친구	100	52	음식	41	86	여행	27
19	근처	99	53	홍대	41	87	골목	26
20	추천	93	54	연락	40	88	부족	26
21	인테리어	90	55	응답	40	89	창문	26
22	편의점	90	56	주방	40	90	거실	25
23	시설	76	57	가성비	39	91	소리	25
24	조용한	75	58	방음	39	92	야경	25
25	편리	73	59	배려	39	93	전반	25
26	세심	71	60	불편함	39	94	조명	25
27	편안함	67	61	요리	39	95	보일러	24
28	가격	66	62	문제	38	96	욕실	24
29	사용	65	63	주차	38	97	칫솔	24
30	편의시설	65	64	체크인	38	98	수건	23
31	가격대비	60	65	접근성	36	99	역	23
32	예약	60	66	버스	35	100	응대	23
33	준비	59	67	하루	35			
34	분위기	58	68	청소	33			

언급 빈도 분석에서 언급된 상위 100위 키워드를 도출했다. 그 결과 호스트 관련, 위치 관련, 가격 관련, 청결성 관련, 소통 관련, 리뷰 내용에 ‘사진과 같아요’에 내포되어있는 의미로 사진과 실물의 갭(gap)의 차이가 적

을수록 이런 리뷰들을 쓰는 경향을 파악했다. ‘사진’이라는 키워드는 정확성과 관련이 있는 표현으로 판단했다. 상위 대표 키워드와 연관 키워드를 분류하여 아래 <표 4-3>와 같이 정리하였다.

<표 4-3> 연관 키워드 분류

상위 대표 키워드	연관 키워드
위치	위치/ 주변/ 근처/ 편의점/ 건물/ 거리/ 카페/ 홍대/ 주차/ 접근성/ 버스/ 지하철/ 지하철역/ 골목/ 역
청결성	깨끗/ 깔끔/ 청결/ 화장실/ 냄새/ 이불/ 쾌적한/ 침구류/ 청소/ 수건
호스트와 소통	호스트/ 친절/ 세심/ 답변/ 응답/ 배려/ 응대/ 연락
정확성	사진/ 공간/ 인테리어/ 거실/ 편의시설
가격	가격/ 가격대비/ 가성비

가장 높은 빈도를 차지한 단어는 ‘위치(450)’로 나왔다. 이는 에어비앤비는 호텔숙소처럼 찾기 쉬운 위치에 있을 수도 있지만, 일반 가정집을 호스트 숙소로 예약한 경우는 숙소의 위치를 사전에 미리 검색하거나 호스트와 의사소통을 통해 조금 더 편하게 찾을 수도 있다. 이때 호스트의 응대 수준에 따라 게스트는 에어비앤비의 숙소에 대한 만족감이 더 커질 수 있다. 그리고 두 번째로 많이 언급된 단어는 ‘깨끗(316)’, 그 다음 3위로 ‘깔끔’과 같이 청결 관련 표현이 연이어 상위로 언급이 된 것은 에어비앤비 숙소는 내국인들에게 청결이라는 요인이 아주 중요하다는 것을 내포하고 있으며 에어비앤비가 아직은 고객들에게 호텔만큼 청결이 보장되는 수준의 신뢰를 보여주고 있지 못하기 때문이다. 그리고 내국인들은 에어비앤비 호스트 숙소를 검색하고 선택할 때 청결을 어떻게 유지하는지를 궁금해할 것이다. 그 다음으로 에어비앤비의 가장 큰 특징인 ‘호스트’가 279번 언급되었고, 연이어 ‘친절(220)’이 도출된 것은 호스트와의 소통에서 친절성이 숙소 이용 만족도에 영향을 끼친다고 할 수 있다.

연구 결과와 같이 이전 설문지 조사 방식으로는 알 수 없는 텍스트가 갖고 있는 의미를 분석함으로써 고객이 중요하게 생각하는 요인이 무엇인지 확인할 수 있다.

2. 워드 클라우드

다음은 텍스트마이닝 결과를 제시할 때 언제나 등장하는 워드 클라우드(word cloud)를 실시하였다. 워드 클라우드를 그리기 위해서는 ‘wordcloud’ 라이브러리 함수를 사용하고 하부 라이브러리 함수인 ‘RColorBrewer’를 사용하여 워드 빈도별 워드 크기, 배치와 색상의 기능을 추가하여 분석하였다.

다음은 워드 클라우드를 그리기 위해 연구자가 사용한 R cod이다.

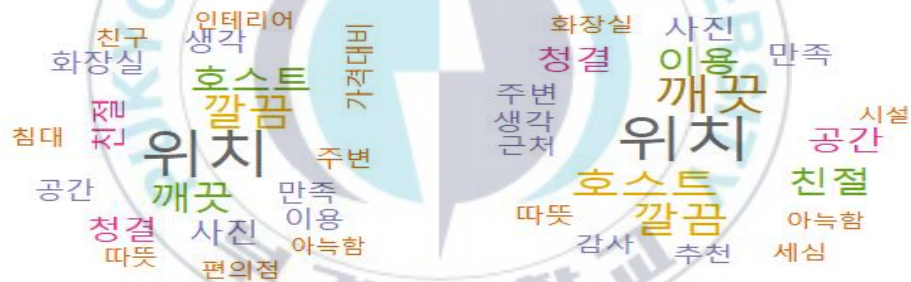
```
“ pal <- brewer.pal(8, "Dark2") # 컬러를 만들어주는 함수,  
wordcloud(words = names(temp), #고유 단어의 열  
freq = temp, #단어 빈도 파일을 freq에 할당  
min.freq = 20, # 표시된 단어의 최소 빈도  
max.words = 500, #빈도 순서로 표시된 500개의 단어  
random.order = FALSE, #중앙에 위치한 최다 빈도 단어들  
rot.per = 0.1, #플롯에서 회전하는 단어의 비율  
scale = c(3, 0.8), #단어의 크기 범위  
color = pal) #단어 색상 ”
```

워드 클라우드는 상위 단어들을 시각화하여 보여주는 기법이다. 표에서 한눈에 들어오기 힘든 키워드들이 시각화를 통해 중요한 의미의 단어들을

[illegible]

와 세 번째 워드 클라우드 분석은 ‘일반 호스트’와 ‘슈퍼 호스트’의 키워드를 따로 분류하여 각각 워드 클라우드 분석을 실시하였다. 등급에 따른 비교분석을 통해 공통점과 차이점을 도출하였다. 차이점은 각각의 호스트의 특징을 설명할 수 있을 것으로 기대된다. 워드 클라우드 결과 일반 호스트와 슈퍼 호스트에서 가장 많이 나타나는 단어(키워드)는 ‘위치’, ‘호스트’, ‘깨끗’, ‘깔끔’, ‘청결’로 나타났다. 이는 에어비앤비에서 긍정적인 역할을 하는 것이 위치적인 장점과 청결에 관련되어 있다는 것을 많이 표현하고 있다는 것을 의미한다. 그리고 일반 호스트의 차이점으로 나타나는 단어(키워드)로 ‘가격대’가

‘침대’, ‘친구’, ‘인테리어’ 단어는 일반 호스트에서 언급 빈도가 높은 것으로 나타났다. 이는 일반 호스트 숙소를 고려할 때 가격에 대한 민감도 높고, 편의시설이 가까이 있어 좋았다는 의미로 해석할 수 있고, 내부 시설에 대한 관심이 높으며 여행목적이 ‘친구’와 함께 하는 경우가 많았다고 볼 수 있다. 슈퍼 호스트의 경우 ‘추천’, ‘감사’, ‘세심’, ‘근처’, ‘시설’이라는 단어의 언급 빈도가 높은 것으로 나타났다. 이는 슈퍼 호스트 숙소 만족도에 영향을 주는 호스트의 응대 수준이 일반 호스트 보다 중요하게 생각하며 높은 만족도는 지인에게 추천하고 싶은 의사를 나타나게 한다고 볼 수 있다. 아래 <그림 4-5> , <그림 4-6>은 호스트 등급별 상위 20위 빈도 결과이다.

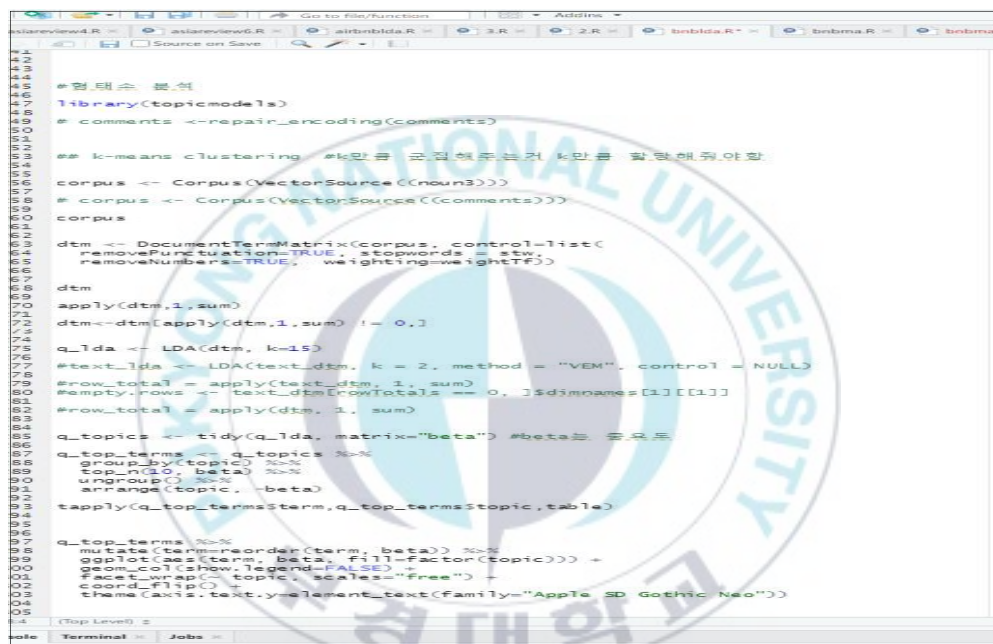


〈그림 4-5〉 일반 호스트 워드 클라우드
결과(n=762)

〈그림 4-6〉 슈퍼 호스트 워드 클라우드
결과(n=783)

3. 토픽 모델링

먼저, 아래 <그림 4-7>은 R studio에서 토픽 모델링을 위해 작성한 코드이다.



```

#토픽 분석
library(topicmodels)
# comments <- repair_encoding(comments)

## k-means clustering #k만큼 군집해주는거 k만큼 할당해줘야함

corpus <- Corpus(VectorSource((noun3)))
# corpus <- Corpus(VectorSource((comments)))
corpus

dtm <- DocumentTermMatrix(corpus, control=list(
  removePunctuation=TRUE, stopwords = stopwords,
  removeNumbers=TRUE, weighting=weightTf))

dtm
apply(dtm,1,sum)
dtm<-dtm[apply(dtm,1,sum) != 0,]

q_lda <- LDA(dtm, k=15)
#text_lda <- LDA(text_dtm, k = 2, method = "VEM", control = NULL)
#row_total = apply(text_dtm, 1, sum)
#empty_rows <- text_dtm[rowTotal == 0, ]$dimnames[1][1]]
#row_total = apply(dtm, 1, sum)

q_topics <- tidy(q_lda, matrix="beta") #beta는 중요도
q_top_terms <- q_topics %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

tapply(q_top_terms$term, q_top_terms$topic, table)

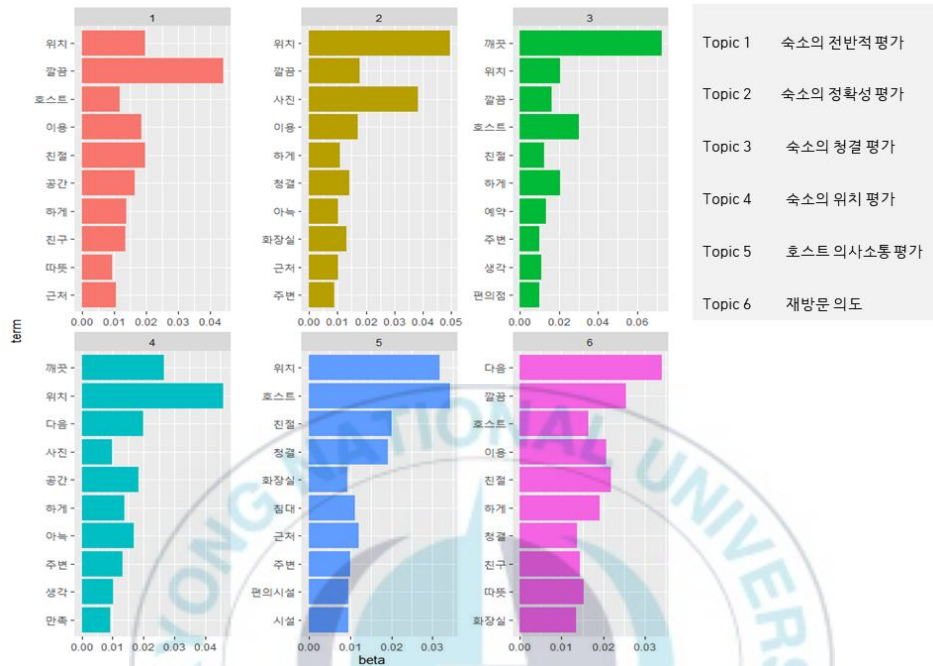
q_top_terms %>%
  mutate(term=reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill=factor(topic))) +
  geom_col(show.legend=FALSE) +
  facet_wrap(~ topic, scales="free") +
  coord_flip()
  theme(axis.text.y=element_text(family="Apple SD Gothic Neo"))
  
```

<그림 4-7> 토픽 모델링 분석 과정

토픽 모델링 분석을 위해 문서의 핵심적인 주제를 찾아 내국인 국문 리뷰 전체의 문서에 주제별로 분류하였는데, 분석 결과 총 6개의 주제로 분류하였다. 주제별 내용은 ‘topic 1: 숙소의 전반적 평가’, ‘topic 2: 숙소의 정확성 평가’, ‘topic 3: 숙소의 청결 평가’, ‘topic 4: 숙소의 위치 평가’, ‘topic 5: 호스트 의사소통 평가’, ‘topic 6: 재방문 의도’로 연구자가 에어비앤비의 평가 항목을 참고하여 도출된 키워드의 의미와 beta 값이 큰 키워드를 중

심으로 주제를 분류하였다. 결과의 의미를 살펴보면 전체적으로 숙소를 이용한 후 평가하는 요소별로 분류되었는데, 결과는 앞서 진행한 언급 빈도 분석, 워드 클라우드 분석과 비슷한 결과의 의미를 포함하고 있음을 확인할 수 있다. 이 방법은 연구자가 여러 번 토픽 수를 반복하여 시도하고 그 중 가장 설명력이 높은 그룹으로 분류되는 토픽 수를 연구자가 선택한다. 주제 그룹이 전체 문서를 가장 잘 설명할 수 있다고 판단하는 토픽 수를 추출해 내는 것이다. 분석 결과, 첫 번째 토픽 주제로 정한 ‘위치’는 언급 빈도 분석에서 1위 빈도수를 가진 단어로 양적인 영향과 함께 중요도도 높은 ‘위치’라는 키워드는 에어비앤비를 이용하는 내국인들이 빈번히 언급하면서도 핵심 키워드라고 할 수 있다. 두 번째 토픽에서 ‘사진’이라는 키워드가 중요하다는 것은 ‘정확성’에 대한 언급이라고 판단하였고, 그리고 ‘호스트’ 역시 빈번히 언급되고 중요도도 높은 것은 호스트와 게스트의 의사소통에 대한 언급이 많다고 판단하여 ‘호스트 의사소통’으로 정했고, ‘청결’과 관련된 키워드들은 계속해서 반복적으로 많이 도출되는 것은 에어비앤비의 주제토픽으로 ‘청결’의 중요성을 강조하지 않을 수 없다고 평가할 수 있다. 마지막 여섯 번째 토픽은 ‘친구’, ‘다음’, ‘이용’, ‘호스트’, ‘친절’ 등으로 ‘재방문 의사’를 표현하는 리뷰에 주로 나오는 키워드들이다. 연구자는 topic 6을 ‘재방문 의도’로 명명하였다.

아래 <그림 4-8>은 R을 사용하여 분석한 토픽 모델링 결과이다.



<그림 4-8> 토픽 모델링 분석 결과($k = 6$)

4. 감성분석

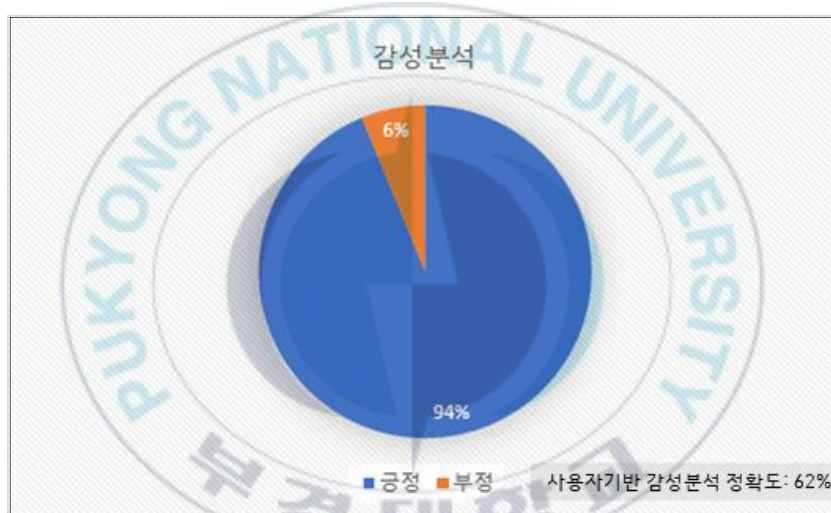
감성분석은 에어비앤비를 이용하는 내국인 게스트들이 서울지역의 숙소에 대해 어떻게 평가하는지를 긍정과 부정의 비율로 결과를 확인할 수 있는 감성분석을 실시하였다. 분석을 위해서는 말뭉치(corpus)에 사용자 기반 감성사전을 대입시킨다. 영문 리뷰 같은 경우는 R에서 사용할 수 있는 감성사전 함수가 있어 쉽게 감성분류를 할 수 있지만, 한국어 텍스트 같은 경우는 사용자 기반의 감성사전을 구축해서 비지도 학습(unsupervised learning) 방식으로 긍정·부정 두 종류로 분류할 수 있고, 또 하나의 방식

은 머신 러닝 지도학습(supervised learning)으로 학습데이터를 생성한 후 분석하는 방식이다. 2가지 모두 장단점이 있지만, 본 연구에서는 사용자 기반 감성사전을 먼저 구축한 후 전체 말뭉치 데이터에 대입시켜 긍정과 부정의 비율을 검증하고, 사용자 기반의 사전이 실제 정확도가 얼마나 되는 지도 분석결과로 확인할 수 있다. 이 사용자 기반 감성사전은 이후 같은 주제의 연구에 반복해서 사용할 수 있다. 한번 구축하여 정확도를 파악하고 저장해 둔 후 재사용할 수 있다. 아래 <표 4-4> 은 연구자가 긍정단어, 부정단어를 분류하여 만든 감성사전이다. 구축하는 방법은 리뷰 전체를 읽고 반복적으로 나오는 긍정단어와 부정단어를 사전에 포함시켜 분석한다. 감성분석 사전을 구축하는 방식은 전체 말뭉치 문서를 품사 분석을 통해서 단어를 확인한다. 본 연구에서 사용한 품사 분석 함수는 한국어 분석 함수 패키지 'KoNLP'의 'SimplePos09' 함수를 통해 전체 문서 내 품사를 9가지로 분류해 주는 방법으로 여기서 'N'은 체언에 해당하며 '수준에서나마' 자연어 상태에서는 띄어쓰기가 정확하지 않기도 해서 '수준에서나마'를 '수준/N+에서나마/J(관계어)', '보았다'를 '보/P(용언)+아다/E(어미)'로 자동분류시켜 앞의 N이나 P에 해당하는 부분을 감성사전에 추가하는 방법으로 사전을 구축하였다. 그리고 긍정 표현과 부정 표현을 나눈 기준은 연구자가 1,545개의 리뷰를 모두 읽고 문서에서 의미하고 있는 표현이 부정적으로 사용하였으면 '부정', 긍정적인 표현으로 사용하였다면 '긍정'으로 분류하였다. 예를 들어 '넷플렉스'는 호스트가 게스트를 위해 넷플렉스에서 영화나 드라마를 볼 수 있도록 서비스를 제공하였는데, 사용 후 만족의 표현으로 넷플렉스를 많이 언급했고, '와이파이' 같은 경우는 요즘은 국내에서는 어디는 '무료 와이파이 서비스'를 제공하고 있는데, 일부 호스트 숙소에서 무료 와이파이 서비스를 제공하지 않는 경우도 있었다. 이런 기본적인 서비스를 제공하지 않을 경우 부정적인 의미를 내포해 언급하게 된다.

〈표 4-4〉 사용자 기반 에어비앤비 온라인 리뷰 감성사전

구분	추출 단어	긍정표현	부정표현	구분	추출 단어	긍정표현	부정표현
1	최고	√		46	주차		√
2	추천드립	√		47	냄새		√
3	핫한	√		48	청결상태		√
4	깔끔	√		49	답답		√
5	똑같	√		50	작았		√
6	편안	√		51	무서울		√
7	세심한	√		52	와이파이		√
8	친절, 친절하네, 친절하	√		53	하수구		√
9	위치	√		54	개수구		√
10	단독	√		55	불친절해서		√
11	폭신	√		56	비싸		√
12	편의시설	√		57	안간		√
13	편찮	√		58	더럽		√
14	아늑했습니	√		59	가까	√	
15	따뜻했	√		60	완벽한	√	
16	만족	√		61	뷰	√	
17	깨끗	√		62	향	√	
18	행복한	√		63	주변	√	
19	사진	√		64	식당	√	
20	요리	√		65	소품	√	
21	내부인테리어	√		66	비품	√	
22	피드백	√		67	빔	√	
23	넷플릭스	√		68	예쁜	√	
24	호스트	√		69	카페	√	
25	응답	√		70	의사소통	√	
26	바로	√		71	넓적	√	
27	넉넉	√		72	좋았습니	√	
28	이용할	√		73	분위기	√	
29	재방문	√		74	적당한	√	
30	방음	√		75	가격대	√	
31	구비	√		76	파티하기	√	
32	연락	√		77	구조	√	
33	접근성	√		78	어렵	√	
34	청결한	√		79	외관	√	
35	응대	√		80	편하	√	
36	답장	√		81	교통	√	
37	수건	√		82	잘쉬고	√	
38	대중교통	√		83	다음	√	
39	아쉽		√	84	저렴	√	
40	생각		√	85	소음	√	√
41	낡은		√	86	배수		√
42	지저분		√	87	코로나		√
43	불편		√	88	먼지		√
44	보일러		√	89	주변소리		√
45	작동		√	90			

본 연구의 결과는 아래 <그림 4-9> 감성분석 결과와 같다. 감성분석은 비지도 학습(unsupervised learning)방식으로 연구자가 사용자 기반 감성사전을 구축한 후 전체 문서에 사전을 넣고 분석한 결과값을 긍정과 부정의 비율로 나타낸 것이다. 전체 문서에 사용자 기반 감성사전을 넣고 분석한 결과 긍정 비율이 94%이고, 부정 비율이 6%로 분석되었다. 이는 에어비앤비 숙소에 대한 내국인 온라인 리뷰는 전체적으로 사용 후 좋은 경험을 가지고 후기를 작성한다고 판단할 수 있다.



<그림 4-9> 감성분석 결과

5. 다중회귀분석

본 연구에서는 에어비앤비 숙박서비스의 선택 속성이 전반적 만족도(평균)에 미치는 영향을 분석하기 위해 다중회귀분석(multiple regression analysis)을 실시하였다. 이를 위해 가격, 청결성, 소통, 체크인, 정확성, 위치의 6개 선택 속성을 독립변수로 하고 전반적 만족도를 종속변수로 설정하여 회귀분석을 실시한 결과 <표 5-1>와 같이 나타났다. 선택 속성의 전반적 만족도에 대한 설명력(R^2)은 6.3%(Adjusted R^2 =5.9%)로 나타났으며, 회귀모형의 F값은 17.198($p<.001$)로 회귀분석은 유의한 것으로 나타났다. 또한 다중공선성을 확인하기 위해 공차한계(tolerance)가 1에 가깝고 분산 확대지수(variance inflation factor: VIF)가 10 미만인지를 확인한 결과 <표 5-1>에 나타난 것처럼 다중공선성에는 문제가 없는 것으로 판단된다. 유의성 검증결과 선택 속성의 6개 가운데 체크인, 정확성, 위치를 제외한 나머지 속성은 전반적 만족도에 유의한 영향을 미치는 것으로 나타났다. 즉 가격, 청결성, 소통이 높을수록 에어비앤비 숙박서비스에 대한 전반적 만족도는 높아진다고 판단할 수 있다. 구체적으로 살펴보면, 청결성($\beta=.193$, $t=7.778$)이 가장 높게 나타났고, 그 다음으로 소통($\beta=.101$, $t=4.063$), 가격($\beta=.092$, $t=3.716$)의 순으로 전반적 만족도에 영향력을 미치는 것으로 나타났다.

〈표 5-1〉 숙박선택 속성이 전반적 만족도에 미치는 영향 분석

종속변수	독립변수	B	S.E.	β	t-value	p	Tolerance	VIF
전반적 만족도 (평균)	상수	4.354	.049		88.348	.000		
	가격	.004	.001	.092	3.716	.000	.993	1.007
	청결성	.264	.034	.193	7.778	.000	.994	1.006
	소통	.162	.040	.101	4.063	.000	.989	1.011
	체크인	.053	.107	.012	.493	.622	.993	1.007
	정확성	.083	.065	.032	1.291	.197	.992	1.008
	위치	.041	.034	.030	1.205	.229	.992	1.008
	$R^2=.063$, Adjusted $R^2=.059$, $F=17.198$, $p<.001$							

6. 위계적 회귀분석

본 연구에서는 호스트 등급(일반 호스트 vs 슈퍼 호스트)이 에어비앤비 숙박서비스의 선택 속성과 전반적 만족도 간의 관계를 조절할 것이라는 연구 과제를 검증하기 위해 위계적 회귀분석을 실시하였다. 이를 위해 6개 선택 속성을 독립변수로, 전반적 만족도를 종속변수로, 호스트 등급을 조절변수로 설정하였다. 위계적 회귀분석을 실시한 결과는 <표 5-2>과 같다. 호스트 등급과 선택 속성의 상호작용항을 투입한 결과 R^2 변화량($\Delta R^2=.014$, $p=.001$)이 유의한 것으로 나타났다. 가격, 정확성, 위치의 3개 변수를 제외한 나머지 선택 속성 변수는 조절효과가 유의한 것으로 나타났다. 구체적으로 살펴보면, 청결성과 호스트 등급 간의 상호작용항($\beta=-.140$, $p=.001$), 소통과 호스트 등급 간의 상호작용항($\beta=-.095$, $p=.015$), 체크인과 호스트 등급 간의 상호작용항($\beta=-.077$, $p=.043$) 등이 유의한 것으로 나타났

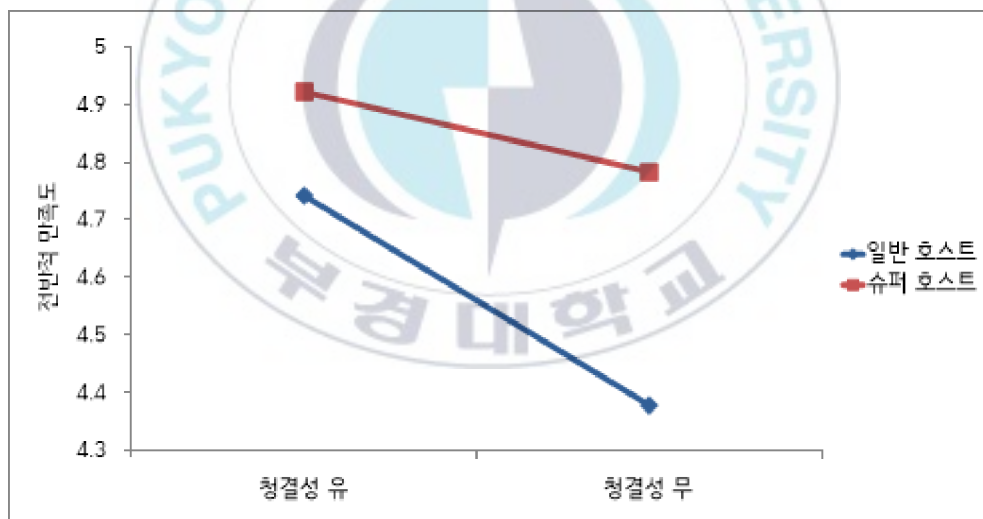
다.

〈표 5-2〉 호스트 등급의 조절효과에 대한 위계적 회귀분석

변수	2 step			3 step		
	B	β	t	B	β	t
가격	.002	.045	1.822	.002	.049	1.514
청결성	.240	.175	7.207**	.358	.261	7.484**
소통	.145	.091	3.725**	.250	-.005	-.019
체크인	.030	.007	.282	.278	.008	.015
정확성	.076	.029	1.205	.113	.115	.940
위치	.016	.012	.477	.035	.676	7.164**
호스트 등급	.285	.209	8.353**	.493	-.054	-.127
가격×호스트 등급				-.001	-.031	-.415
청결성×호스트 등급				-.223	-.140	-3.356**
소통×호스트 등급				-.190	-.095	-2.435*
체크인×호스트 등급				-.429	-.077	-2.022*
정확성×호스트 등급				-.067	-.019	-.525
위치×호스트 등급				-.041	-.025	-.611
F		25.368**			15.643**	
ΔR^2		.041**			.014**	
R^2		.104			.117	

*p<.05, **p<.01

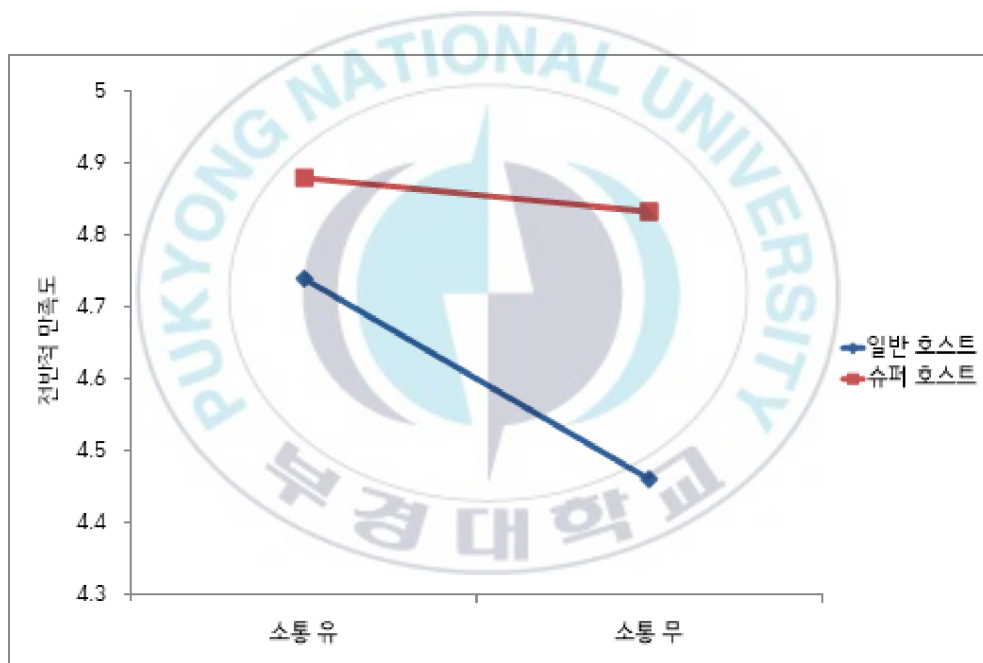
조절효과가 유의한 선택 속성 변수들을 중심으로 이원분산분석을 실시하여 그 결과값으로 그래프를 제시하여 구체적인 조절효과를 살펴보았다. 먼저 청결성과 호스트 등급의 상호작용 그래프를 보면 <그림 5-1>과 같다. 분석 결과 청결성 유무에 따라 호스트 등급별 전반적 만족도의 값이 차이가 나는 것으로 나타났다. 슈퍼 호스트의 경우 청결성 유무의 차이를 보면, 청결성 없음의 전반적 만족도는 4.783이고 청결성 있음의 전반적 만족도는 4.922로 나타나 유의한 차이를 보이지 않고 있다. 그러나 일반 호스트의 경우, 청결성 없음의 전반적 만족도는 4.378인 반면, 청결성 있음의 전반적 만족도는 4.742로 나타나 큰 차이가 있음을 보이고 있다. 따라서 슈퍼 호스트와는 달리 일반 호스트 숙소의 경우 청결성이 유무가 전반적 만족도에 미치는 영향은 유의함을 알 수가 있다.



<그림 5-1> 청결성과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과

둘째, 원활한 소통과 호스트 등급의 상호작용 그래프를 보면 <그림 5-2>와 같다. 분석 결과 원활한 소통 유무에 따라 호스트 등급별 전반적

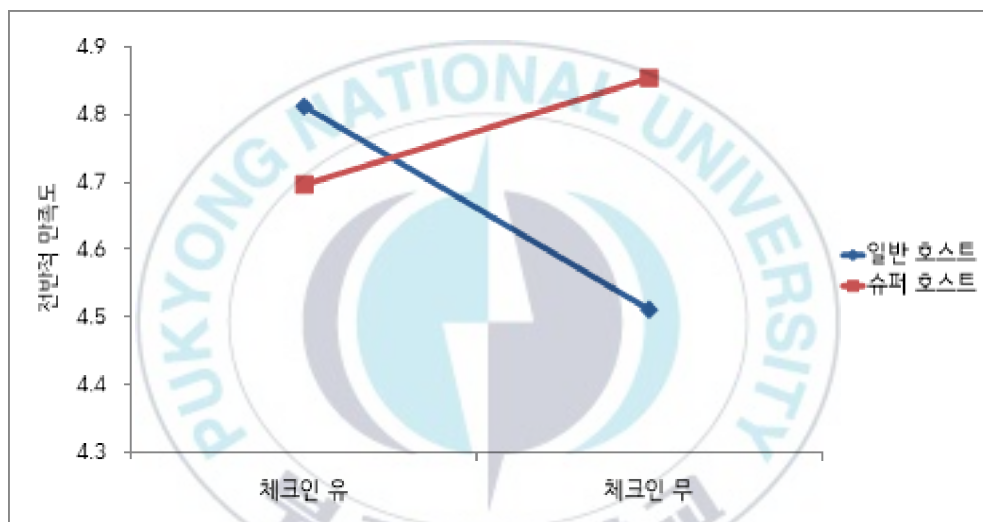
만족도의 값이 차이가 나는 것으로 나타났다. 슈퍼 호스트의 경우 소통 유무의 차이를 보면, 소통 없음의 전반적 만족도는 4.832이고 소통 있음의 전반적 만족도는 4.897로 나타나 유의한 차이를 보이지 않고 있다. 그러나 일반 호스트의 경우, 원활한 소통 없음의 전반적 만족도는 4.460인 반면, 소통 있음의 전반적 만족도는 4.739로 나타나 큰 차이가 있음을 보이고 있다. 따라서 슈퍼 호스트와는 달리 일반 호스트 숙소의 경우 원활한 소통 유무가 전반적 만족도에 미치는 영향은 유의함을 알 수가 있다.



〈그림 5-2〉 소통과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과

셋째, 원활한 체크인과 호스트 등급의 상호작용 그래프를 보면 <그림 5-3>과 같다. 분석 결과 원활한 체크인 유무에 따라 호스트 등급별 전반적 만족도의 값이 차이가 나는 것으로 나타났다. 슈퍼 호스트의 경우 체크

인 유무의 차이를 보면, 원활한 체크인 없음의 전반적 만족도는 4.854이고 원활한 체크인 있음의 전반적 만족도는 4.696로 나타나 유의한 차이를 보이지 않고 있다. 그러나 일반 호스트의 경우, 원활한 체크인 없음의 전반적 만족도는 4.511인 반면, 원활한 소통 있음의 전반적 만족도는 4.812로 나타나 큰 차이가 있음을 보이고 있다. 따라서 슈퍼 호스트와는 달리 일반 호스트 숙소의 경우 원활한 체크인 유무가 전반적 만족도에 미치는 영향은 유의함을 알 수가 있다.



〈그림 5-3〉 체크인과 전반적 만족도 간의 영향관계에 대한 호스트 등급의 조절효과

제 5 장 결 론

제 1 절 연구의 요약

본 연구는 서울지역을 중심으로 에어비앤비를 이용한 내국인들의 온라인 리뷰 데이터를 분석하여 탐색된 선택 속성 요인들과 전반적 만족도 간의 영향관계와 호스트 등급별 조절효과의 차이검증에 대한 고객들의 인식을 파악하고자 하였다.

본 연구를 위해 먼저 선행된 문헌 연구를 고찰하였다. 에어비앤비의 속성 및 전반적 만족도, 온라인 리뷰에 관한 선행연구와 공학적 접근을 위해 텍스트 마이닝의 기법들에 관한 문헌 고찰도 하였다. 이를 통하여 연구 과제 및 연구 범위 등을 설정하였고 분석을 위해 서울지역 에어비앤비 숙소를 방문한 내국인들이 남긴 온라인 리뷰 및 정보를 수집하였다. 수집된 리뷰는 텍스트 마이닝을 거쳐 통계분석이 가능한 형태로 변환하였다. 단어(키워드)에 대한 언급 빈도 분석을 통해 에어비앤비를 의미하는 특이점을 찾을 수 있었고, 워드 클라우드 분석을 통해 ‘일반 호스트’와 ‘슈퍼 호스트’에서 언급되는 빈도 단어들의 차이점을 시각화를 통해 확인할 수 있었다. 그리고 토픽 모델링을 통해 전체 리뷰 문서의 주제를 도출하였고, 에어비앤비 사용자 기반 감성사전을 구축하여 감성분석을 진행하였다. 감성분석 결과로 전체 문서가 의미하는 긍정과 부정의 감정에 대한 평가를 통해 고객들의 인식을 확인할 수 있었다. 텍스트 마이닝에서 도출한 탐색적 연구 결과를 바탕으로 연구자가 수치 데이터로 변환하여 도출된 요인들과 게스트들이 부여한 전반적 만족도(평균) 간의 영향관계를 호스트 등급별 차이와 함께 분석하였다. 본 연구 결과를 요약하면 다음과 같다.

첫째, 기존에 연구되었던 에어비앤비 선택 속성에 대한 범위를 확장하여 내국인들이 지각하는 에어비앤비에 대한 의견과 감정을 담은 리뷰에 대하여 연구하였다. 현재 에어비앤비에 대한 연구가 다양한 분야에서 진행해 왔지만, 빅데이터, 다양한 텍스트 마이닝 기법들, 온라인 리뷰 데이터 등을 활용 연구가 아직은 미흡한 편이다. 국내 사회과학 분야의 연구에서 텍스트 마이닝을 하는 경우에는 직접 프로그램을 개발하여 다양한 방식으로 접근하는 것이 아닌 텍스트 마이닝을 도와주는 응용 소프트웨어 도구를 이용하기 때문에 분석에 한계가 있어 전체 텍스트의 의미를 파악하기에는 부족한 면이 있다고 볼 수 있다. 그러나 본 연구에서는 deep learning 방식과 R 등을 이용하여 프로그램을 직접 개발함으로써 연구 목표에 도달할 수 있었고 다양한 텍스트 마이닝 기법을 활용함으로써 고객의 의견을 다각도로 파악할 수 있게 되었다. 이를 통해 고객의 의견에 대한 핵심을 파악하고 개선점을 찾을 수 있는 계기를 마련할 수 있게 되었다.

둘째, 리뷰에 대해 자연어 처리 분석을 통하여 불용어 등 분석에 불필요한 단어들을 전처리하였다. 그러나 텍스트 마이닝 기법들에 따라 한국어 텍스트를 분석하기 위한 전처리 방법을 고려하여 실행시키고, 분석의 종류에 따라 자연어 처리 방법을 다르게 하였다. 그리고 도출된 키워드들이 갖고 있는 의미를 분석하였다.

셋째, 서울지역의 숙소 분포 특성을 살펴보고 숙소가 형성되어 있는 지역이 갖는 의미를 도출해 냈다.

넷째, 선정된 에어비앤비 숙박 장소의 선택 속성 요인들이 전반적 만족도(평균)에 유의한 영향을 미치고 있음을 검증하였다. 그리고 그 요인들이 호스트 등급별 조절효과 차이가 나타나는지 분석한 결과 유의한 차이를 보였다.

제 2 절 연구의 시사점

본 연구에서는 위와 같은 연구 결과를 통하여 다음과 같은 이론적, 실무적 시사점을 도출하였다.

1. 이론적 시사점

본 연구의 이론적 시사점은 다음과 같다.

첫 번째, 기존의 많은 연구에서 진행하는 설문조사 방법의 한계점을 보완할 수 있도록 에어비앤비의 온라인 리뷰를 이용하여 고객의 의견과 감정을 표현한 텍스트들을 통해 좀 더 고객의 니즈를 명확하게 규명할 수 있는데, 온라인 리뷰의 활용한 연구는 매우 바람직하다.

두 번째, 텍스트 마이닝에서 기존 선행연구 기법들과 함께 한 단계 더 진보된 형태의 분석들도 함께 활용하였다. 기존의 빈도 분석과 시각화에서 더 확장된 연구를 제시하기 위해 다양한 기법들을 활용하여 온라인 리뷰 문서에서 찾을 수 있는 잠재되어있는 의미를 도출해 내고 전체 문서의 감정상태가 전반적으로 긍정적인 것을 확인하였는데, 이는 선행연구에서 같은 결과로 나타났다. Bridges, & Vásquez(2018)는 미국의 여러 지역별로 평점에 대해 분석한 결과에서 전체적으로 평점이 높고 긍정적인 리뷰가 많은 것은 에어비앤비에는 호스트가 피드백을 해줘야 하는 시스템이 있어 게스트 관리를 적극적으로 하는 것이 리뷰에 긍정적인 영향을 미치는 된 것이라고 했다. 토픽 모델링 기법은 최근 가장 많이 활용되고 있는 분석기법(이수상, 2016; 홍태호 등, 2018; 박종도, 2019)이지만, 아직 관광 분야에서는 분석 방법을 활용한 연구가 미흡하다. 전체 온라인 리뷰 문서에 의미를

주제로 분류하는 방식으로 기존 선행연구에서 사용한 선택 속성 요인과 관련된 결과를 도출할 수 있었다. 물론, 토픽 수는 연구자가 정하는 것으로 주관적 개입이 있는 부분은 있지만, 새로운 기법을 통해 의미를 찾아낸 부분에 대한 의의는 있다고 판단된다.

세 번째, 내국인에 대한 다양한 접근 방식으로 연구 진행이 필요한 시점에서 본 연구는 아주 시기적절하다고 판단된다. 그리고 최근 가장 떠오르는 빅데이터 분석 프로그램 R을 활용하여 다양한 텍스트 마이닝을 시도한 것은 앞으로 호텔 관광 분야에서도 적극적으로 활용될 수 있도록 하는 기초자료로 활용될 것이다.

네 번째, 기존 분석방식으로부터 확장하여 깊이 있게 데이터에 접근을 시도하였다. 이는 방법론적인 이해를 통해 향후 확장된 연구 계획을 수립할 수 있다.

다섯 번째, 본 연구에서는 사용자 기반 감성분석을 실시하였는데, 영어 텍스트 분석의 경우 R에서는 다양한 감성분석 사전이 이미 구축되어 있어 편리하게 사용이 가능하지만, 안타깝게도 한국어 텍스트의 경우는 여러 곳에서 연구가 진행되고 있지만, 공식적인 사전이 현재는 없다. 따라서 감성분석 한국어 텍스트를 연구할 경우 사용자 기반의 감성사전을 먼저 구축하는 것이 중요하고 분야마다 사전을 달라질 수 있다. R을 활용하여 감성분석을 하는 방법은 두 가지가 있지만 본 연구에서는 에어비앤비 온라인 리뷰에 사용이 가능한 사용자 기반 감성사전을 구축한 것에 의의를 두며 감성분석 결과로 긍정과 부정의 비율을 검증하면서 전체 문서가 의미하는 감정을 표현할 수 있었다.

여섯 번째, 텍스트 마이닝 분석에서 가장 많이 사용하는 분석이 워드 클라우드이다. 키워드의 빈도에 따른 크기를 비교하여 핵심 키워드를 직관적으로 확인할 수 있는 방식이다. 본 연구에서는 이 분석을 ‘일반 호스트’와

‘슈퍼 호스트’에 관련된 상위 20위 단어를 비교 분석하여 내포되어있는 키워드의 의미를 찾기 위해 활용하였다. 기존 에어비앤비 연구에서는 시도되지 않은 분석 형식으로 새로운 결과를 통찰할 수 있었다.

일곱 번째, 기존 연구에서도 텍스트 마이닝 기법들에서 통계분석을 함께 분석하는 시도들이 되고 있다. 에어비앤비의 경우 국내연구에서는 거의 최초라고 볼 수 있어 연구의 큰 차별점이라고 할 수 있다. 기존 호텔 분야의 선행연구를 고찰하였고 연구자가 텍스트 마이닝에서 도출된 의미와 선행연구 선택 속성에 대한 고찰, 호텔 연구방법론에 대한 고찰, 에어비앤비 공식 사이트의 호스트 평가 항목 6가지를 고찰하여 선택 속성요인 6개를 선정하였고, 요인들과 전반적 만족도 간의 영향력 관계에서 유의한 결과를 통계적으로 이끌어 낼 수 있었다. 이는 향후 빅데이터 연구의 확장성을 시사함으로써 다양한 시도에 필요한 기초자료가 될 것이다.

여덟 번째, 본 연구는 에어비앤비 숙박서비스의 가장 큰 특징인 호스트 등급에 주안점을 두어 둘의 차이가 고객의 전반적 만족도에 어떻게 영향력이 있는지를 요인들 간의 관계에서 조절효과로 검증하였는데, 아주 흥미로운 결과를 도출하였다. 이는 향후 호스트의 특성에 대한 연구를 진행할 때의 방향성을 제시할 수 있을 것이라 판단한다.

아홉 번째, 호스트 등급의 조절효과의 검증 결과 본 연구에서는 ‘일반 호스트’가 ‘청결성’, ‘소통’, ‘체크인’의 속성 요인들에서 전반적 만족도에 더 영향을 미치는 결과를 나타냈는데, 이 결과는 기존 선행 연구인 김지희·이건웅(2019)과는 다른 결과를 나타냈다. 본 연구에서는 슈퍼 호스트의 영향력에 대해 입증하지는 못했다. 하지만 기존 연구에서 검증하지 않았던 일반 호스트의 영향력에 대한 새로운 견해를 발견할 수 있었다.

2. 실무적 시사점

본 연구는 빅데이터 분석 기법인 텍스트 마이닝을 활용하여 에어비앤비 게스트들이 남긴 온라인 리뷰에 대한 의견이나 생각들을 단어로 도출하여 해석하고 의미를 고찰하였다. 이렇게 도출된 선택 속성요인들이 에어비앤비를 이용하는 전반적 만족도에 미치는 영향을 통계분석을 실시하여 에어비앤비 숙박서비스 산업의 발전에 도움을 주고자 과학적 분석 방법을 제시하였다. 이를 통해 내국인 관광객의 의견에 대해 분석하고 개선점 등을 찾고 더 나은 전략을 수립하는 데 도움이 되고자 하였다. 본 연구 결과, 다음과 같은 실무적 시사점을 도출하였다.

첫 번째, 데이터 분포 특성에서 관광지과 숙소 분포 지역과 매우 밀접한 연관성이 있는 것을 확인하였다. 이는 에어비앤비가 현재 추구하고 있는 방향과 일치하는데, ‘경험’을 적극적으로 활용해야 한다는 것을 더욱 암시하고 있다. 숙소뿐만 아니라 관광지 상품과 연계하면 더욱 만족도도 높아질 수 있을 것이다.

두 번째, 일반 호스트와 슈퍼 호스트의 리뷰 데이터를 분류하여 워드 클라우드 분석을 실시하였는데, 상위 20위 단어들을 중심으로 의미를 도출하여 보면, 일반 호스트는 슈퍼 호스트에서는 보이지 않는 키워드들이 몇 가지 도출되었는데, ‘가격대비’가 의미하는 것은 일반 호스트를 이용하는 게스트들은 가격 민감도가 높다고 할 수 있다. 이는 숙소를 운영하는 일반 호스트들이 가격 책정 시 적절한 가격을 유지하여야 게스트들의 예약율을 높일 수 있을 것이라고 판단할 수 있다. 그리고 ‘침대’, ‘인테리어’, ‘친구’에서 볼 수 있듯이 누구와 이 여행을 하려는지를 짐작할 수 있다. 친구와 주로 에어비앤비를 이용하는 것으로 판단할 수 있는데 내부 시설의 의미인

인테리어와 침대는 역시 친구와 여행을 가려고 선택하는 호스트 숙소 타입을 고려할 때 내부 인테리어나 침대의 타입이 중요한 변수가 될 것이라고 판단할 수 있다. 다음은 슈퍼 호스트에만 나타난 키워드를 살펴보면 ‘추천’, ‘세심’, ‘감사’ 등의 키워드가 도출되었는데, 슈퍼 호스트는 호스트와의 소통이 그 무엇보다 중요한 요인이라고 판단할 수 있다. 친절한 서비스를 경험하고 호스트와 좋은 경험을 교환했다면 주위 지인들에게 추천할 의향이 있다는 의미로 볼 수 있다. 이는 슈퍼 호스트는 특히 게스트와의 소통을 원활히 하는 것이 중요하다. 무엇보다 게스트가 요구하는 문의사항에 최대한 바른 답변을 하고, 처리할 사항은 빠르게 처리해주는 것으로 고객 유치와 수익창출을 유지할 수 있을 것이다.

세 번째, 다중회귀분석 결과에서 ‘가격’, ‘청결성’, ‘소통’, ‘체크인’, ‘정확성’, ‘위치’ 중 ‘가격’, ‘청결성’, ‘소통’이 전반적 만족도에 유의한 영향을 미치는 것은 에어비앤비는 가격이 호텔과는 다르게 ‘적절성’을 유지해야 할 필요가 있고, 청결이 중요한 것은 텍스트 마이닝 분석 결과를 통해서도 확인이 되었지만, 기존 호텔서비스는 이용 경험이 많고 어느 정도의 고객의 기대감과 신뢰가 형성되어 있지만, 아직 에어비앤비는 그런 부분에서는 부족한 부분이 있다고 평가할 수 있고, 호스트 숙소는 개인이 운영하기 때문에 호스트별 차이를 크게 느껴질 수 있어, 이 부분을 체계적으로 관리 할 필요가 있다. 그리고 최근에는 코로나19로 인해 더욱 청결성이 중요하게 부각되고 있는데, 에어비앤비에서는 ‘청결강화프로그램’을 최근에 운영하기 시작했고, ‘숙소 청소 10계명’이 담긴 ‘숙소 안전 가이드 라인’을 제작하여 배포하고 있다(에어비앤비, 2020b). 이처럼 시대의 상황에 맞는 적절한 시스템과 개선안이 빠르게 투입할 수 있는 사업이 에어비앤비의 특징이다. 최근에는 일부 호스트는 손 소독제 배치 등을 홈페이지 소개들에 추가하여 ‘청결성’을 부각시키고 있다. 이와 같이 청결에 관해서는 앞으로도 고객에

게 알려드리는 일은 소홀히 하지 말고, 청결을 위한 노력을 어떻게 체계적으로 진행하고 있는지를 알려주어야 한다.

네 번째, ‘가격’, ‘청결성’, ‘소통’, ‘체크인’, ‘정확성’, ‘위치’의 요인들과 전반적 만족도에 대한 호스트 등급의 조절효과에 대한 위계적 회귀분석 결과 ‘청결’, ‘소통’, ‘체크인’이 유의한 차이를 보였는데, 흥미로운 것은 일반 호스트가 더 많은 영향을 미치는 것으로 나타났다. 이는 일반 호스트를 운영할 때, ‘청결’, ‘호스트와의 소통’, ‘원활한 체크인’이 잘 유지되면 고객들은 전반적 만족도(평점)를 높게 인식하게 되는데, 만약 이 부분이 미흡하다면 전반적 만족도(평점)에 부정적인 영향을 미칠 수 있다고 판단할 수 있다. 이점을 유의하여 운영한다면 고객들에게 좋은 평가를 받을 수 있고, 슈퍼 호스트 배지를 얻는데 도움이 될 것이다. 그에 반해 슈퍼 호스트는 유의한 차이를 보이지 않은 것은 기존 슈퍼 호스트를 선택하고, 그에 따른 전반적 만족도에 대한 평가를 하는데, 특정한 요인이 크게 작용한다고 보기 힘들다. 이는 슈퍼 호스트 제도는 에어비앤비에서 매우 까다롭게 운영되는 제도로 슈퍼 호스트 배지를 유지하기 위해서는 모든 평가 요인들에 대해 높은 만족도를 보여야 하는 것이고, 슈퍼 호스트 배지를 유지하는 숙소는 기본적으로 고객들에게 신뢰를 주기 때문에 일반 호스트보다는 전반적 만족도가 높게 형성되어 있다. 단지 개별 요인이 일반 호스트와 비교했을 때 큰 영향의 차이가 없어 전반적 만족도에 미치는 영향이 낮은 것은 슈퍼 호스트 특정 요인에 집중하기 보단 전반적으로 모든 부분에서 우수해야지만 높은 만족도를 유지할 수 있다고 할 수 있다.

제 3 절 연구의 한계 및 향후 연구 방향

본 연구에는 다음과 같은 한계점이 있다.

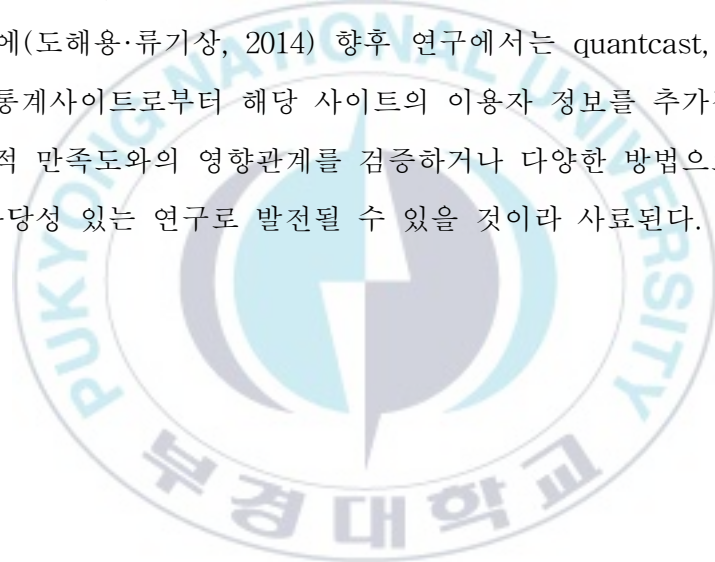
첫 번째, 본 연구는 웹 크롤링 방식을 통해 서울지역의 에어비앤비 호스트 숙소와 리뷰들을 수집하여 분석한 결과를 통해 서울지역 전반적인 견해를 일반화하고자 하였으나, 수집하는 단계에서 전체 호스트 숙소와 리뷰를 수집하는 것은 개인 연구자가 진행하기에는 매우 큰 양의 데이터이고, 시간도 많이 소요되어 전체 데이터를 수집하지 못하는 한계를 들어낸다. 하지만 에어비앤비의 특징이 설정한 날짜에 예약이 가능한 숙소만을 노출시키고 보여주는 개수가 한정되어 있는 것이 특징으로 본 연구에서 수집한 308개는 그 설정값에서는 전체 데이터와 마찬가지로 볼 수 있지만, 차후 연구에선 서울지역 전체 숙소 데이터를 통해 더 다양한 의미를 도출해 낼 수 있을 것이다.

두 번째, 본 연구에서는 한국어로 남긴 리뷰만 수집하여 분석하였으나 다른 언어들도 모두 수집하여 언어별로 분류하여 비교하는 것도 향후 좋은 연구가 될 것이다. 영어로 리뷰를 남긴 고객들은 호스트와 어느 정도의 의사소통이 가능하여 불편함이 많이 나타나지 않을 수 있으나 영어를 구사하지 못하고 자국의 언어만으로 호스트와 대화에 어려움을 겪는 고객들은 많은 불편과 불만을 겪었을 수 있다. 그러므로 이들을 분석하여 개선점을 찾는다면 영어 구사가 어려운 다른 나라에서 온 고객들에 대해 더 좋은 서비스를 제공할 수 있으며 우리나라의 이미지를 향상시키는데 도움이 될 방법을 제시할 수 있을 것이다.

세 번째, 본 연구에서는 우리나라 에어비앤비 숙소서비스를 제공하는 지역 중 특정 ‘서울지역’ 만을 선정하여 분석하였다. 향후 연구에서는 에어비

앤비의 지역적 특성을 살피기 위해 여러 지역을 함께 분석함으로써 각 지역의 특성에 맞는 마케팅 시사점을 제시할 수 있을 것이다.

마지막으로 본 연구의 한계점은 웹상의 정보에 대해 많은 연구자들이 지적하는 인구통계학적 정보의 누락이다(조성겸·주영수·조은희, 2005; Sparrow, & Curtice, 2004). 이러한 문제는 최근 인터넷이나 SNS, 모바일 등의 사용자가 증가하면서 인터넷 사용자라는 표본이 점차 전체 모집단화되고 있어 기존의 표본 추출과 비슷하거나 우수하다는 연구들이 있지만(Coderre, Mathieu, & St-Laurent, 2004), 한계점이 완전히 해결된 것은 아니기 때문에(도해용·류기상, 2014) 향후 연구에서는 quantcast, alexa 등과 같은 인구통계사이트로부터 해당 사이트의 이용자 정보를 추가적으로 수집하여 전반적 만족도와 의 영향관계를 검증하거나 다양한 방법으로 활용한다면 보다 타당성 있는 연구로 발전될 수 있을 것이라 사료된다.



참 고 문 헌

<국 내 문 헌>

- 공효순(2008). 비즈니스 여행자를 위한 국내 특급호텔과 중저가호텔 선택 속성에 따른 만족도 비교분석 I, 『한국콘텐츠학회논문지』, 8(12), 414-423.
- 공효순·송은지(2014). 빅데이터 기반 호텔고객 평판 분석에 관한 연구, 『한국디지털콘텐츠학회논문지』, 15(2), 219-225.
- 곽수정·김현희(2019). 텍스트 마이닝과 토픽 모델링을 기반으로 한 트위터에 나타난 사회적 이슈의 키워드 및 주제 분석, 『정보처리학회논문지』, 소프트웨어 및 데이터 공학, 8(1), 13-18.
- 곽민정·최지유·박소현(2019). 호텔 서비스 속성별 고객만족도 분석을 위한 온라인 리뷰 감성분석, 『관광경영연구』, 23(4), 1-25.
- 구정대·박태수(2005). 호텔 객실 서비스 품질이 사후평가에 미치는 영향, 『관광연구』, 20(2), 87-102.
- 권진영·이문용(2012). 온라인 리뷰 유용성의 결정 요인에 대한 연구, 『한국지능정보시스템학회』, 15(2), 219-225.
- 권혁진·김진후·이충훈(2017). 에어비앤비 호스트 특성이 호스트 신뢰 및 재이용 의도에 미치는 영향, 『호텔경영연구』, 26(7), 55-70.
- 김광영·이원구·이민호·윤화목·신성호(2011). 웹 자원 아카이빙을 위한 웹 크롤러 연구 개발, 『한국콘텐츠학회논문지』, 11(9), 9-16.
- 김근형·오성열(2009). 온라인 고객 리뷰 분석을 통한 시장세분화에 텍스트 마이닝 기술을 적용하기 위한 방법론, 『한국콘텐츠학회논문지』, 9(8), 272-284.

- 김도경·김인신(2017). 텍스트 마이닝을 이용한 온라인 리뷰의 호텔 선택 속성 분석, 『관광학연구』, 41(9), 109-127.
- 김지희·이건웅(2019). 공유숙박 서비스 성공에 미치는 요인에 대한 실증연구, 『Information System Review』, 21(2), 69-89.
- 김수연(2006). 텍스트 마이닝 기법을 이용한 연관용어 선정에 관한 실험적 연구, 『정보관리학회지』, 23(3), 147-165.
- 김수현·이영준·신진영·박기영(2019). 경제분석을 위한 텍스트마이닝, 『BOK경제연구』, 서울:한국은행, 단행본(다권본), 제2019-18호.
- 김상태·Tsolmon.b·김남조(2017). 숙박공유 플랫폼 이용자의 선택속성에 관한 연구: 에어비앤비 이용자를 중심으로, 『관광연구논총』, 29(2), 51-69.
- 김성근·조혁준·강주영(2016). 학술 연구에서의 텍스트 마이닝 활용 현황 및 주요 분석 기법, 『정보화연구』, 13(2), 317-329.
- 김지은·백순근(2016). 텍스트 빅데이터 분석 기법을 활용한 대학구조개혁 평가의 쟁점 분석, 『아시아교육연구』, 17(3), 409-436.
- 김진한(2019). 공유형 숙박서비스의 제품 특성, 플랫폼 특성, 호스트 특성이 태도와 이용의도에 미치는 영향: 이용경험, 상호의존적 성향, 불확실성 회피성향, 네트워크 효과의 조절효과를 중심으로, 동국대학교 대학원 박사학위논문.
- 김진한·강우성(2018). 공유형 숙박서비스와 호텔서비스의 소비가치가 이용의도에 미치는 영향: 네트워크 효과와 서비스 유형의 조절된 매개효과를 중심으로, 『관광연구저널』, 32(9), 33-49.
- 김윤민(2009). 뷔페레스토랑의 선택속성이 고객 만족과 재방문 의도에 미치는 영향, 『관광연구저널』, 23(4), 329-347.

- 김은주·김천서·신용경(2006). 베이커리 유형에 따른 소비자 선택속성 차이에 관한 연구, 『관광레저연구』, 18(4), 263-280.
- 김화진·이승곤·한진수(2011). 항공사 좌석등급별 서비스품질 차원이 고객 가치와 고객 만족에 미치는 영향 관계 연구, 『관광연구저널』, 25(1), 235-252.
- 김희섭·강보라(2019a). 토픽 모델링을 활용한 도서관, 기록관, 박물관간의 연구주제 분석, 『한국도서관정보학회지』, 50(4), 339-358.
- 김희섭·강보라(2019b). 국내외 허위정보 연구 동향 비교분석, 『한국문헌정보학회지』, 53(3), 291-315.
- 도해룡·류기상(2014). 빅데이터를 이용한 레스토랑 브랜드 개성 측정: 미국 레스토랑 평가 사이트 yelp.com 리뷰 데이터 기반, 『관광학연구』, 38(1), 249-271.
- 도해룡·이애주·강한훈(2012). 웹 데이터 마이닝 기법을 이용한 레스토랑 만족도 평가에 관한 연구, 『외식경영연구』, 15(2), 115-135.
- 류일환·주재현·김동연·안재현(2017). 에어비앤비 호스트 리뷰 피드백이 게스트의 구매의도에 미치는 영향, 『한국경영과학회 학술대회논문집』, 79-106.
- 문지효·한상린(2016). 외부 사건 충격으로 인한 해외관광객 감소가 서울시 유통산업에 미치는 영향: 2015년 메르스 사태로 인한 경제적 파급효과 분석, 『유통연구』, 21(1), 129-152.
- 박경배·하성호(2015). 상품의 구매 후기 분석을 통한 결점발견: 호텔 산업을 중심으로, 『한국경영정보학회 학술대회논문집』, 571-576.
- 박나희·유윤희(2018). 헤노닉가격모형을 이용한 에어비앤비 숙박가격 결정요인 분석, 『관광경영연구』, 22(6), 425-442.

- 박상준(2017). 고객 만족과 고객 충성도의 관계에 있어서 소비자 관여도의 예측력 관점에서 조절효과, 『고객만족경영연구』, 19(4), 1-15.
- 박소민·양승우(2020). 서울시 숙박공유업체 에어비앤비 입지특성에 관한 연구, 『한국도시설계학회지 도시설계』, 21(1), 25-34.
- 박재관·조태영·이준엽(2011). 지역축제의 환경단서와 관광체험이 축제 개최 효과에 미치는 영향연구, 『호텔관광연구』, 40, 22-35.
- 박재범·김경애·김상희(2013). 서비스제공자 커뮤니케이션이 서비스제공자에 대한 평가와 고객 행동 의도에 미치는 영향: 서비스 유형의 조절효과를 중심으로, 『마케팅논집』, 21(2), 119-157.
- 박정기·이동명(2018). VOC 분류와 판단평가 결과에 따른 직원의 개인특성 차이 분석, 『유라시아연구』, 15(1), 81-98.
- 박종도(2019). 토픽 모델링을 활용한 다문화 연구의 이슈 추적 연구, 『한국문헌정보학회지』, 53(3), 273-289.
- 박홍식·김주일(2019). 공유숙박 서비스 기업의 감성분석을 통한 이미지 변화, 『Cullinary Science & Hospitality Research』, 25(5), 67-74.
- 백영민(2017). R을 이용한 텍스트 마이닝, 경기도: 한울아카데미.
- 반현정·전재균(2019). 빅데이터 분석을 통한 부산의 특급호텔과 비즈니스호텔의 의미연결망 분석에 관한 연구, 『Culinary Science & Hospitality Research』, 25(1), 18-28.
- 반현정·김학선(2019). 온라인 리뷰 분석을 통한 부산지역 레스토랑 인식에 관한 연구2: 국문 리뷰를 중심으로, 『산업혁신연구』, 35(2), 185-207.
- 변숙은·조은성(2015). 빅데이터를 이용한 온라인 구전 문화 분석, 『문화산업연구』, 15(2), 1-8.

- 송보미·윤병운·박용태(2011). 고객 리뷰 분석을 통한 서비스품질 진단: 감성분석 및 겹 분석 접근, 『대한산업공학회 추계학술대회논문집』, 737-742.
- 송순영(2015). 에어비앤비(Airbnb) 사례를 통해 본 공유경제 관련 법 제정 현황 분석, 『소비자정책동향』, 64, 19-36.
- 손은미·정란수·정철(2014). 웰니스 관광시설 선호도 및 선호선택: 건강 리조트 잠재 관광객을 대상으로, 『관광연구논총』, 26(3), 51-77.
- 신지용·박성용·김대호(2006). 온라인 쇼핑 구매의도 결정요인에 관한 연구, 『대한경영학회지』, 19(2), 529-560.
- 심영석·김홍범(2016). 텍스트 마이닝을 이용한 관광지 이미지 구성요인 및 측정에 관한 연구, 『관광학연구』, 40(7), 221-245.
- 심영석·김홍범(2018). 온라인 리뷰 빅데이터 기반의 Word2Vec 기법을 활용한 관광지 개성과 여행객 평점 간 구조적 관계 분석, 『관광학연구』, 42(8), 165-1.
- 안성숙·양수진(2018). 가치-태도-행동 모델에 기반한 공유숙박에 대한 소비자 행동의 이해: 에어비앤비 서비스를 중심으로, 『소비문화연구』, 21(2), 97-119.
- 이광국·김경옥·전재균(2016). 관광지 선택속성에 따른 시장세분화 연구: 감천문화마을을 중심으로, 『Journal of the Korean Data Analysis Society』, 18(6), 3239-3249.
- 이금실·이인주·이영경(2018). LDA(latent dirichlet allocation) 기반 토픽 모델링 기법을 활용한 관광 분야의 가상현실(virtual reality) 연구 동향 분석, 『한국관광레저학회 학술발표대회 논문집』, 425-432.
- 이병철·변효정(2014). 온라인 리뷰가 관광상품 구매행동에 미치는 영향: 호텔리조트를 중심으로, 『관광레저연구』, 26(7), 208-225.

- 이상미·정현영(2009). 국내거주 외국인의 한식 레스토랑 선택과 만족도 및 재방문에 관한 연구, 『외식경영연구』, 12(1), 293-313.
- 이수상(2016). 독후감 텍스트의 토픽모델링 적용에 관한 탐색적 연구, 『한국도서관정보학회지』, 47(4), 1-18.
- 이성각·조현진·임병호(2017). 비즈니스 호텔의 선택 속성에 따른 고객 만족이 행동 의도에 미치는 영향에 관한 연구, 『관광레저연구』, 29(10), 221-237.
- 이선미·배광현·이규범·이진희·곽민정(2016). 고객후기 분석을 통한 국내 호텔 서비스 품질 핵심요인 연구, 『대한산업공학회 추계학술대회논문집』, 208-225.
- 이의준·원형식·이새롭(2020). 숙박공유 플랫폼 서비스의 서비스 공급자 특성이 사용자의 만족감에 미치는 영향: 에어비앤비(Airbnb) 호스트 주체를 중심으로, 『정보시스템연구』, 29(1), 1-22.
- 이용숙(2013). 소비자 평점이 높은 유럽 호텔의 특징에 대한 참여관찰 결과와 후기내용 비교, 『소비자학연구』, 24(2), 253-283.
- 이은영·주경희·이두희(2019). 워드 클라우드 기법을 이용한 최근 소비자학 연구 트렌드 분석, 『상품학연구』, 37(1), 1-7.
- 임영희(2019). 온라인 리뷰 빅데이터를 활용한 외래관광객의 국내호텔 이용 만족도에 관한 연구: 감성분석을 중심으로, 세종대학교 대학원 박사학위논문.
- 임영희·김홍범(2019). 호텔 온라인 리뷰 빅데이터를 활용한 감성분석에 관한 연구, 『호텔경영학연구』, 28(7), 105-123.
- 이정현(2013). 온라인 호텔 이용 후기의 질적 내용분석에 의한 고객가치 연구, 『한국콘텐츠학회논문지』, 13(10), 533-546.
- 이혜미(2018). 빅데이터를 활용한 공유숙박의 지각된 위험 인식에 관한 연

- 구, 경희대학교 대학원 박사학위논문.
- 이혜진·최자영(2019). 텍스트 마이닝 기법을 활용한 편의점 이용 고객의 트
위터 리뷰에 대한 감성분석, 『글로벌경영학회지』, 16(4), 143-164.
- 엄수진·윤태환(2010). 레지던스 호텔 선택 속성의 중요도와 만족도에 관한
연구: 부산지역을 중심으로, 『관광레저연구』, 22(5), 93-110.
- 윤동원·이성림(2019). 에어비앤비에 대한 소비자의 긍정적·부정적 인식: 온
라인커뮤니티 게시글 내용에 대한 텍스트 분석을 중심으로, 『소비자
학연구』, 30(2), 173-192.
- 윤석영·이무용(2015). 숙박시설의 장소성 체험 요인 연구: 담양의 숙박시설
죽향문화체험 마을과 명가해를 사례로, 『관광연구논총』, 27(2),
123-144.
- 유양호·강상묵(2016). 호텔 레스토랑의 온라인 구전 정보 특성과 행동 의도
간에 영향관계에서 신뢰전이 효과, 『관광경영연구』, 20(6), 309-328.
- 윤정현·박태정(2019). 텍스트 마이닝 기반 토픽 모델링을 통한 대학 교수자
의 교수(Teaching)관련 연구 동향 분석, 『교육혁신연구』, 29(4),
507-529.
- 오연주·채수환(2015). 영화평가 평점을 이용한 감성문장구축을 통한 영화
평점 추론, 『인터넷정보학과논문지』, 16(2), 41-048.
- 주위에치·김소연(2019). 공유숙박서비스의 정보주체별 정보품질과 소비자
정보 만족에 관한 연구: 에어비앤비 사례를 중심으로, 『소비문화연
구』, 22(4), 51-77.
- 주덕(2016). 관광부문 공유경제의 실험: 공유숙박 이용 의도에 따른 숙박시
설 선택 속성, 『관광레저연구』, 28(4), 43-58.
- 전영록(2015). 관광객 방문 동기에 따른 관광 만족 및 행동 의도 간의 차이
분석: 안동 월영교를 중심으로, 『관광연구』, 20(7), 355-377.

- 정은별·이형룡(2020). 국내 에어비앤비를 이용하는 내국인 고객의 숙박시설 선택 속성에 관한 연구: Best-Worst Scaling을 이용한 상대적 중요도 분석과 비교를 중심으로, 『관광학연구』, 44(4), 49-69.
- 정원준·김가운(2016). 에어비앤비 숙박공유 서비스의 비즈니스 현황 및 규제 동향, 『주간기술동향』, 1743, 2-13.
- 정지혜·한지수·김건휘(2015). 호텔 이용객의 체면 민감성과 호텔선택 속성, 고객 만족도 간의 관계, 『관광레저연구』, 27(7), 131-149.
- 조성겸·주영수·조은희(2005). 인터넷 여론조사의 정확도 관련 요인, 『조사연구』, 6(2), 51-74.
- 조민호·최인호(1999). 호텔선택속성 평가의 비교 문화 연구: Hofstede 이론의 적용, 『관광학연구』, 23(1), 122-137.
- 최유성·안혁근·박정원·심우현(2017). 4차 산업혁명에 대응하는 규제개혁 연구: 공유(共有)경제와 디지털 헬스케어 분야를 중심으로, 『한국행정연구원』, 1-713.
- 최영민·배상욱(2016). 온라인 구전 정보의 특성이 구전효과에 미치는 영향에 관한 연구, 『한국산학기술학회』, 17(2), 307-314.
- 최지유·박소현·김상호·이규범·곽민정(2018). 온라인 리뷰 마이닝을 통한 국내호텔 서비스품질 주요속성 분석, 『관광경영연구』, 22(5), 1069-1093.
- 최연태·박상인(2018). 신뢰가 공유경제 플랫폼 이용에 미치는 영향에 관한 탐색적 연구, 『한국지역정보화학회지』, 21(4), 103-125.
- 한경·양위주(2015). 관광경험 결과에 따른 관광객의 재방문 의도 차이 분석, 『동북아관광연구』, 11(1), 183-197.
- 한국데이터산업진흥원(2014). 데이터 분석 전문가 가이드, 서울: 한국데이터산업진흥원.

- 황서이·황동열(2018). 토픽 모델링과 의미연결망 분석을 위한 예술경영 연구 동향 분석, 『예술경영연구』, 47, 5-29.
- 한진수·현경석(2009). 패밀리레스토랑의 고객가치, 고객 만족, 행동 의도 간의 구조적 관계에 관한 연구: 서울지역 대학생을 중심으로, 『호텔경영학연구』, 18(1), 135-150.
- 한주형·정남호(2017). 글로벌 마케팅 활성화를 위한 빅데이터 효용성 연구, 『관광레저연구』, 29(7), 233-251.
- 한성수·양동우(2017). 텍스트마이닝을 이용한 창업 관련 연구 동향 분석, 『벤처창업연구』, 12(5), 1-12.
- 한희정·구철모·정남호(2016). 공유경제 숙박상품 선택에 영향을 미치는 상대적 정보 신호 분석, 『한국경영정보학회 추계학술대회』, 80-85.
- 허지정·노승철(2018). 서울시 숙박공유업체 에어비앤비(Airbnb)의 특성과 공간분포 분석, 『한국도시지리학회지』, 21(1), 65-76.
- 홍태호·니우한잉·임강·박지영(2018). LDA를 이용한 온라인 리뷰의 다중 토픽별 감성분석: TripAdvisor 사례를 중심으로, 『정보시스템연구』, 27(1), 89-110.

<국 외 문 헌>

- Anderson, E. W., Fornell, C., & Lehmann, D. R.(1994). Customer satisfaction, market share, and profitability: Findings from Sweden, *Journal of Marketing*, 53-66.
- Archak, N., Ghose, A., & Ipeirotis, P. G.(2011). Deriving the pricing power of product features by mining consumer reviews, *Management Science*, 57(8), 1485-1509.
- Bickart, B., & Schindler, R. M.(2001). Internet forums as influential sources of consumer information, *Journal of Interactive Marketing*, 15(3), 31-40.
- Blei, D. M., Ng, A. Y., & Jordan, M. I.(2003). Latent dirichlet allocation, *Journal of Machine Learning Research*, 3(2003), 993-1022.
- Blei, D. M., & Lafferty, J. D.(2009). Topic models, Text mining: classification, clustering and applications, 10(71).
- Bolton, R. N.(1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction, *Marketing Science*, 17(1), 45-65.
- Bridges, J., & Vásquez, C.(2018). If nearly all airbnb reviews are positive, does that make them meaningless?, *Current Issues in Tourism*, 21(18), 2057-2075.
- Busch, P., & Wilson, D. T.(1976). An experimental analysis of a salesman's expert and referent bases of social power in the buyer-seller dyad, *Journal of Marketing Research*, 13(1), 3-11.
- Chen, Y., & Xie, J.(2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix, *Management*

Science, 54(3), 477-491.

- Cheng, M., & Jin, X.(2019). What do airbnb users care about? an analysis of online review comments, *International Journal of Hospitality Management*, 76, 58-70.
- Chevalier, J., & Mayzlin, D.(2006). The effect of word of mouth on sales: Online book reviews, *Journal of Marketing Research*, 43, 345 - 354.
- Chow, K., Garretson, J., & Kurtz, D.(1994). An exploratory study into the purchase decision process used by leisure travellers in hotel selection, *Journal of Hospitality and Leisure Marketing*, 2(4), 53-71.
- Coderre, F., Mathieu, A., & St-Laurent, N.(2004). Comparison of the quality of qualitative data obtained through telephone, postal and email surveys, *International Journal of Market Research*, 46(3), 347-358.
- Cronin, J. J., & Taylor, S. A.(1994). SERVPERF versus SERVQUAL: Reconciling performance based and perceptions minus expectations measurement of service quality, *Journal of Marketing*, 58(1), 125-131.
- Crosby, L. A., Evans, K.R. & Cowles, D.(1990). Relationship quality in services selling: An interpersonal influence perspective, *Journal of Marketing*, 54(3), 68-81.
- Crotts, J. C., Mason, P. R., & Davis, B.(2009). Measuring guest satisfaction and competitive in the hospitality and tourism industry an application of stance-shift analysis to travel blog

- narratives, *Journal of Travel Research*, 48(2), 139–151.
- Decker, R., & Trusov, M.(2010). Estimating aggregate consumer preferences from online product reviews, *International Journal of Research in Marketing*, 27(4), 293–307.
- Ert, E., Fleischer, A., & Magen, N.(2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb, *Tourism Management*, 55, 62–73.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z.(2006). Tapping the power of text mining, *Communications of the ACM*, 49(9), 76–82.
- Gunderson, M. G., Heide, M., & Olsson, U. H.(1996). Hotel guest satisfaction among business travelers, *Cornell hotel and restaurant administration quarterly*, 37(2), 72–81.
- Gunter, U.(2018). What makes an airbnb host a superhost? Empirical evidence from San Francisco and the bay area, *Tourism Management*, 66, 26–37.
- Guttentag, D.(2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector, *Current Issues in Tourism*, 18(12), 1192 - 1297.
- Hotho, A., Nurnberger, A., & Paaß, G.(2005). A brief survey of text mining, *In Ldv Forum*, 20(1), 19–62.
- Hornik, K., & Grün, B.(2011). Topicmodels: An R package for fitting topic models, *Journal of Statistical Software*, 40(13), 1–30.
- Hunt, H. K.(1977). CS/D-Overview and future research directions, in Hunt, H. K. (Eds.), *Conceptualization and measurement of consumer satisfaction and dissatisfaction*. Cambridge, MA:

Marketing Science Institute.

- Jang, G. E., Park, S. H., & Kim, W. J.(2015). Automatic construction of a negative/positive corpus and emotional classification using the internet emotional sign, *Journal of KIISE*, 42, 512-521.
- Kim, H. S., & Noh, Y.(2019). Elicitation of design factors through bigdata analysis of online customer reviews for washing machines, *Journal of Mechanical Science and Technology*, 33(6), 2785-2795.
- Lawani, A., Reed, R. M., Mark, T., & Zheng.(2019). Review and price on online platforms: Evidence from sentiment analysis of airbnb reviews in boston, *Regional Science and Economics*, 75, 22-34.
- Lewis, R. C.(1984). Isolating differences in hotel attributes, *The Cornell Hospitality Quarterly*, 25(3), 64-77.
- Liang, S., Schukert, M., Law, R., & Chen, C. C.(2017). Be a Superhost: The importance of badge systems for peer-to-peer rental accommodation, *Tourism Management*, 60, 454-465.
- Li, X., & Hitt, L. M.(2008). Self-selection and information role of online product reviews, *Information Systems Research*, 19(4), 456-474.
- Mao, Z., Yang, Y., & Wang, M.(2018). Sleepless nights in hotels? Understanding factors that influence hotel sleep quality, *International Journal of Hospitality Management*, 74, 189-201.
- Menner, T., Höpken, W., Fuchs, M., & Lexhagen, M.(2016). Topic detection: Identifying relevant topics in tourism reviews, *Information and Communication Technologies in Tourism*, 411-423.

- Miner, G., Elder, J., & Hill, T.(2012). Practical text mining and statistical analysis for non-structured text data applications, Academic Press.
- Oliver, R. L.(1980). A cognitive model of the antecedents and consequences of satisfaction decisions, *Journal of Marketing Research*, 17(4), 460-469.
- Phillips, P., Barnes, S., Zigan, K., & Schegg, R.(2016). Understanding the impact of online reviews on hotel performance: An empirical analysis, *Journal of Travel Research*, 56(2), 235-249.
- Richins, M. L., & Root-Shaffer, T.(1988). The Role of Involvement and Opinion Leadership in Consumer Word-of -Mouth: An Implicit Model Made Explicit, *Advanced in Consumer Research*, 15, 32-35.
- Saleh, F., & Ryan, C.(1991). Analysis service quality in the hospitality industry usingthe SERVQUAL model, *Service Industries Journal*, 11(3), 324-343.
- Schindler, R. M., & Bickart, B.(2002). Characteristics of online consumer comments valued for hedonic and utilitarian shopping tasks, *Advances in Consumer Research*, 29, 428-429.
- Shanhan, K., & Hyman, M. R.(2007). An exploratory study of desired hotel attributes for American tourists vacationing in China and Ireland, *Journal of Vacation Marketing*, 13(2), 107-118.
- Smith, D., Menon, S., & Sivakumar, K.(2005). Online peer and editorial recommendations, trust, and choice in virtual markets, *Journal of Interactive Marketing*, 19(3), 15-37.
- Sparrow, N., & Curtice, J.(2004). Measuring the attitudes of the general

- public via internet polls: An evaluation, *International Journal of Market Research*, 46(1), 23-45.
- Sparks, B. A., & Browning, V.(2011). The impact of online reviews on hotel booking intentions and perception of trust, *Tourism Management*, 32(6), 1310-1323.
- Stevens, P., Knutson, B., & Patton, M.(1995). DINESERV: A tool for measuring service quality in restaurants, *The Cornell Hotel and Restaurant Administration Quarterly*, 36(2), 56-60.
- Sundaram, D. S., & Webster, C.(2000). The role of nonverbal communication in service encounters, *Journal of Service Marketing*, 14(5), 378-391.
- Swan, J. E., & Oliver, R. L.(1989). Postpurchase communication by consumer, *Journal of retailing*, 65(4), 516-533.
- Tapscott, D., & Williams, A. D.(2008). Wikinomics: How mass collaboration changes everything. Penguin.
- Tapscott, D., & Williams, A. D.(2010). Innovating the 21st-century university: It's time, *Educase review*, 45(1), 16-29.
- Taylor, S. A., & Baker, T. L.(1994). An assessment of the relationship between service quality and customer satisfaction in the formation of consumer's purchase intention, *Journal of Retailing*, 70(2), 163-178.
- Telles Jr, R.(2016). Digital matching firms: A new definition in the sharing economy space, *ESA issue brief(01-16)*, 1-27.
- Toni, M., Renzi, M. F., & Mattia, G.(2018). Understanding the link between collaborative economy and sustainable behavior: An

- empirical investigation, *Journal of Cleaner Production*, 172, 4467-4477.
- Tussyadiah, L. P.(2016). Factors of satisfaction and intention to use peer-to-peer accommodation, *International Journal of Hospitality Management*, 55, 70-80.
- Wang, Y., & Fessenmaier, D.(2004). Towards understanding members' general participation in and active contribution to an online travel community, *Tourism Management*, 25(6), 709-722.
- Westbrook, R. A., & Reilly, M. D.(1983). Value-percept disparity: An Alternative to the disconfirmation of expectations theory of consumer satisfaction in R. P. Bagozzi and A. M. Tybout (Eds.), *Advances in consumer research*, 10.
- Xiang, Z., Gretzel, U., & Fesenmaier, D. R.(2009). Semantic representation of tourism on the internet, *Journal of Travel Research*, 47(4), 440-453.

<웹 기사>

- 김상훈 (2013.01.30.). 호텔 묵는다는 것, 이제 구식이죠... 현지인 집으로 가 보세요”, <http://www.donga.com/news/article/all/20130129/52678945/1>
- 손덕호(2020.05.26.). 4만원 숙박쿠폰 100만장·DMZ일부개방...관광 내수 살리기 총력전, https://biz.chosun.com/site/data/html_dir/2020/05/26/2020052603605.html?utm_source=naver&utm_medium=original&utm_campaign=biz
- IBM 2015년 보고서. <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/>
- 안갑성(2018.10.26.). 숙박에 체험 더했더니...손님과 주인 경계가 사라졌다, MK뉴스, <https://www.mk.co.kr/news/print/2018/667902>
- 안상희(2013.01.29.). 에어비앤비 창업자, 한국 진출 본격화, https://biz.chosun.com/site/data/html_dir/2013/01/29/2013012901428.html
- 에어비앤비(2019.07.01.). 에어비앤비가 한국에 미친 경제적 파급효과, 작년 한 해 1조 4,000억원”, <https://news.airbnb.com/ko/%ec%97%90%ec%96%b4%eb%b9%84%ec%95%a4%eb%b9%84%ea%b0%80-%ed%95%9c%ea%b5%ad%ec%97%90-%eb%af%b8%ec%b9%9c-%ea%b2%bd%ec%a0%9c%ec%a0%81-%ed%8c%8c%ea%b8%89%ed%9a%a8%ea%b3%bc-%ec%9e%91%eb%85%84-%ed%95%9c/>
- 에어비앤비 수수료. <http://www.airbnb.co.kr/help/article/63/what-are-host-service-fees>
- 에어비앤비(2020.06.09.a). 언론계·학계 10명 중 8명, 빈집 숙박용으로 쓸 수 있게 허용해야, <http://news.airbnb.com/ko/expert-polling-result/>

에어비앤비(2020.06.11.b). 에어비앤비, 국내 관광 활성화 위한 이제, 여행은 가까운 곳에서 캠페인 시작, <http://news.airbnb.com/ko/airbnb-launches-campaign-to-support-domestic-travel-local-economic-growth/>

오대석(2016.02.24.). 돈 되는 판(Platform)이 온다...더 플랫폼 개최, <https://www.etnews.com/20160224000243>



<온라인 리뷰 수집데이터: 정제 전 raw data>

- 111 -

<에어비앤비 국내 선행연구>

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
정은별·이형룡	2020	관광학연구	국내 에어비앤비를 이용하는 내국인 고객의 숙박시설 선택 속성에 관한 연구:Best-Worst Scaling을 이용한 상대적 중요도 분석과 비교를 중심으로	2020년 2월 24일부터 27일까지 지난 1년간 국내 에어비앤비 이용 경험이 있는 내국인을 대상으로 318부 설문조사 수집	SPSS 22.0을 활용하여 기술적 통계분석을 실시, Excel을 이용하여 집계법으로 BWS집수를 분석	내국인은 에어비앤비를 이용하는데 가장 중요하게 생각하는 속성이 가격으로 나타났고, 그 다음으로 청결, 위치, 정확성, 가치, 다양한 사진 순으로 나타남. 그에 반해 후기 수, 사진 수, 의사소통 평점, 체크인 평점, 호스트 응답속도 등은 상대적으로 낮은 중요도 결과로 나타남	에어비앤비 외국인 속성 중요도를 조사하여 내국인과 외국인 고객의 선택 속성 중요도를 비교하는 연구 필요
윤동원·이성립	2019	소비자학연구	에어비앤비에 대한 소비자의 긍정적·부정적 인식: 온라인커뮤니티 게시물 내용에 대한 텍스트분석을 중심으로	일일 데이터 수집 개수와 접근 제한이 없는 클리앙 온라인 커뮤니티 사이트에서 2013년 1월 7일부터 2018년 5월 23일까지 전체 게시물과 댓글 12,188개 수집	R 프로그램을 사용하여 텍스트마이닝 실시하여 빈도분석, 군집분석(LDA), 감성분석, 사회연결망 분석, 텍스트시각화를 위한 워드 클라우드 실시	소비자들은 새로운 소비형태에서 느끼는 즐거움, 성취감, 타인과의 관계에서 얻는 친밀감 등의 감성적 니즈를 중시하는 것, 그리고 소비자들은 제도적 이슈에 관련된 키워드가 부정적 요인으로 작용.	위계적 군집분석을 위한 덴드로그램, 토픽 모델링, 중심성 분석

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
김지희·이건웅	2019	Information System Review	공유숙박 서비스 성공에 미치는 요인에 대한 실증연구	에어비엔비 플랫폼 데이터 수집, 2018년 6월 26일부터 27일까지 미국 뉴욕 맨해튼 지역 1인(성인) 기준 1박이 가능한 숙소를 대상으로 2018년 4월 3일부터 4일까지 숙박기간 설정하여 숙소 2,596건의 정보를 수집 중 결측치를 제거하고 1,673 숙소로 연구 진행	고객 평점, 숙박비, 편의시설 개수, 숙소이용규칙, 예약취고 규정, 고객 요구 응답률/ 시간, 슈퍼호스트의 여부, 지역 정보 등을 포함하여 분석 리뷰 평점(만족도)을 종속변수로 하고, 독립변수로 숙소의 경제성, 편리성, 신뢰성을 요인으로 선정 후 기술통계 실시, 상관분석, 로지스틱 회귀분석 실시	환불제도가 엄격할수록 고객평점에 부정적 영향을 줌. 편의시설의 비율이 높을수록 고객 평점(만족도)에 긍정적인 영향을 줌. 호스트의 응답률이 높으면 고객의 만족도가 높은 것을 확인, 슈퍼호스트 여부에 따라 신뢰성에 영향을 주는 것을 알 수 있었음 슈퍼호스트일 경우 아닐 경우보다 약 11배 정도 예약이 성공할 확률이 높아졌음	고객 리뷰, 경제성, 편리성, 신뢰성
안효인·최유리·오래은·송종우	2019	The Korean Journal of Applied Statistics	에어비엔비(Airbnb) 웹 로그 데이터를 이용한 고객 행동 예측	웹사이트 Kaggle에서 제공하는 Airbnb 고객 데이터 수집, 총 20,559명의 개별 데이터를 수집	고객들의 성별, 연령 등의 기본 정보 및 웹 로그가 포함된 데이터 셋을 이용하여 첫 숙소 예약까지 걸리는 개인의 의사결정 시간을 예측	랜덤 포레스트(random forest) 분류모형을 통해 고객 행동 예측 결과 모형을 사용하지 않은 것 보다 예측 정확도가 두 배가 높아짐	웹 로그 활동의 순서적 흐름에 대한 정보 반영 고려

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
김진한·강우성	2018	관광연구저널	공유형 숙박서비스와 호텔서비스의 소비가치가 이용의도에 미치는 영향: 네트워크 효과와 서비스 유형의 조절된 매개효과를 중심으로	호텔과 에어비앤비의 이용경험자들을 대상, 2018년 4월 21일 미국 아마존의 서베이 서비스인 엠펙 온라인 서베이를 통한 설문조사를 시행, 미국인들을 대상으로 호텔이용경험자 159명, 에어비앤비 이용경험자 142명 으로 총 301명 설문지를 수집 지난 2년간 여행 경험이 비정상적으로 많은 응답자를 제외하여 총255(호텔132명, 에어비앤비 123명)을 분석	255명의 설문조사 응답지를 분석 통계분석을 위해 SPSS 22와 Hayes의 PROCESS macro를 사용, 호텔과 에어비앤비 이용경험자를 두 집단으로 나누어 설문조사 실시, 숙박서비스의 소비가치와 이용의도와의 관계를 네트워크 효과를 매개로 분석과 서비스 유형의 차이가 소비가치 및 네트워크 효과가 이용의도에 미치는 영향 분석	에어비앤비 이용자는 유희적 가치에, 호텔 이용자는 경제적 가치 때문에 숙박시설에 대한 이용의도가 높은 것과 네트워크 효과에서 에어비앤비(Airbnb)와 호텔 모두 이용의도에 유의한 영향을 미치는 것으로 나타났다. 이는 곧 에어비앤비 서비스를 이용하는데 타인과의 교감, 어울림인 네트워크 효과가 더 강하게 작용하고 있다는 것을 확인	소비가치가 이용의도에 미치는 영향을 검증하기 위해 Bootstrapping 과 Sobel Test 방법 이용, 매개효과 확인을 위해 SPSS PROCESS 59번 모델과 Bootstrapping 신뢰구간을 이용하여 검증

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
안성숙·양수진	2018	소비문화연구	가치-태도-행동 모델에 기반한 공유숙박에 대한 소비자행동의 이해: 에어비앤비 서비스를 중심으로	서울대학교 경영학 수업을 듣는 대학생들을 대상, 에어비앤비 모바일 어플리케이션 사용 경험을 부여하고 이용자 294명을 대상으로 온라인 설문지 수집	이용 경험이 없는 소비자들을 대상으로 공유경제와 공유숙박 에어비앤비에 대해 설명하고 15분 정도 에어비앤비 어플리케이션 서핑의 시간을 부여한 후 설문조사 실시, 공유경제에 대한 지각된 가치와 브랜드 인지도, 태도, 충성도, 지각된 위험, 사용경험 등의 요인들의 확인적 요인분석 실시	공유경제 참여 동기를 다차원으로 고려하여 그 중 어떠한 동기가 최신 IT서비스에 대한 태도를 형성하고 지속적으로 사용하고자 하는 추진력으로 작용하는가를 살펴본 결과 정서적 가치와 사회적 가치가 소비자들의 태도 형성에 유의한 영향을 주고, 지각된 위험 즉 에어비앤비의 보안수준에 대한 소비자 인식이 사용 경험이 있는 대학생 소비자에게 있어 에어비앤비에 대한 태도를 결정하는 데 중요한 변수로 작용함을 규명했다는 것	VAB모델을 적용하여 공유숙박에 대한 소비자 행동 의도 분석, AMOS 22.0 구조방정식 모델의 분석 실시

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
허지정·노승철	2018	한국도시지리학 회지	서울시 숙박공유업체 에어비앤비(Airbnb) 의 특성과 공간분포 분석	서울시를 대상으로 숙박공유업체 에어비앤비의 특성과 공간적 분포 패턴을 분석, Tom Slee 박사가 공유한 세계 주요 도시의 에어비앤비 자료를 활용, 118개 도시 중 대륙별로 관광객이 많은 6개 도시(싱가포르, 뉴욕, LA, 런던, 파리, 바르셀로나)와 에어비앤비가 시작된 샌프란시스코 그리고 국내비교를 위해 부산을 선정하여 총 8개 도시의 숙소와 공유자 특성을 서울시와 비교	총 8개 도시의 숙소와 공유자 특성을 서울시와 비교, 에어비앤비의 공간적 분포 특성을 보기 위해 계층군집방법(NNHC:Nearest Neighbor Hierarchical Clustering)을 이용해 클러스터 추출	여러 개의 숙소를 운영하는 다중 공유자가 큰 비중을 차지, ‘집 전체’임대가 약 절반을 차지, 에어비앤비는 도심, 강남역, 홍대 입구 등 주요 관광지와 대중교통이 편리한 지역에 집중, 호텔과 달리 주거 지역에도 다수의 에어비앤비가 입지하고 있음	

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
권혁진·김진후 ·이충훈	2017	호텔경영학연구	에어비앤비 호스트 특성이 호스트 신뢰 및 재이용 의도에 미치는 영향	2016년~2017년 최근 1년 이내 이용경험이 있는 고객을 대상으로 150부 설문 조사 수집	신생 공유숙박서비스 비즈니스 모델인 에어비앤비에서 이용자가 지각한 호스트 속성이 이용자의 호스트에 대한 신뢰와 재이용 의도에 어떠한 영향을 미치는지 실증적 검증	에어비앤비의 재이용 의도를 높이기 위해서는 호스트들은 게스트들에게 약속이행을 반드시 지키고 정확한 정보를 제공하고 전문적인 지식과 경험을 공유하는데 있어 의사소통이 원활해야 하는 것이 중요한 결과로 제시	변수의 조작적 정의 SPSS 22, 신뢰도 분석, 상관분석, 변수간 영향관계 다중회귀분석, 다중매개효과분 석 위계적 회귀분석 실시

<에어비앤비 국외 선행연구>

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
Ulrich Gunter	2018	Tourism Management	What makes an airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area	Airbnb listings from San Francisco and the Bay Area scraped by AirDNA and aggregated for the time period from September 2014 until August 2016, the present study quantified the marginal contributions of the four super host criteria	to obtain the Airbnb super host status. In order to quantify the marginal contributions of the four criteria, different index models of binary response (logit, probit, and IV probit, which allows for the endogeneity of Airbnb demand) are applied	the most important criterion, followed by reliable cancellation behavior of the host, host responsiveness, and sufficient Airbnb demand. Moreover, commercial Airbnb providers are more likely to obtain the super host status	-
Eyal Ert, Aliza Fleischer, Nathan Magen	2016	Tourism Management	Trust and reputation in the sharing economy: The role of personal photos in Airbnb	data set consists of snapshot of all listings available in the city of Stockholm as of June 25, 2014. includes 395 posted listings, of which 185 received guest	focuses only on the listings that were reviewed. regression of log of price on median score of trustworthiness of host's photo and review	trustworthy the host is perceived to be from host's photo, the higher the price of the listing and the probability of its being chosen. host's reputation, communicated by online review scores,	Electronic word of mouth, decision making

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
				reviews.	scores. comparison of average review score booking.com and Airbnb.	has no effect on listing price or likelihood of consumer booking.	
Iis P. Tussyadiah	2016	International Journal of Hospitality Management	Factors of satisfaction and intention to use peer-to-peer accommodation	via AMT in June 2015 to capture response from adult travelers residing in the US who had used P2P accommodation within the past six months	Respondents reported using P2P accommodation for both lesiure and business trips. 5-point Likert-type scale, pilot study, questionnaire was via AMT(Amason Mechanical Turk), September 2014 targeting adults(aged 18 years or older)	significant positive effects of enjoyment and economic benefits(cost-savings) on satisfaction and intention to use P2P accommodation were identified. Amenities also significantly influence satisfaction, but not intention. the s spirit of ground well and sustainable consumption in collaborative consumption were found to be associated with negative effects. guest who shared	pilot study, AMOS

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
						space with the hosts, an emphasis on sustainable lifestyle or environmentally-friendl y practices in the property seems to result in a less satisfying stay.	
Daniel Guttentag	2015	Current Issues in Tourism	Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector	data collected on 26 January 2013 and refer to one night of accommodation for one adult on Friday.	analysis about rates for Airbnb and Hotel	Airbnb retails are actually illegal due to short-term rental regulations.	peer-to-peer accommodati on, disruptive innovation, informal economy

<온라인 리뷰 국내 선행연구>

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
반현정· 김학선	2019	산업혁신연구	온라인 리뷰 분석을 통한 레스토랑 인식에 관한 연구 2: 국문 리뷰를 중심으로	2013년 5월부터 2017년 8월까지 트립어드바이저 사이트 내 부산지역 레스토랑 리뷰 711개의 리뷰와 21,405개의 단어들을 수집	리뷰 데이터를 통해 정성적 분석과 정량적 분석 실시, 정성적 분석으로 텍스트 마이닝, UCINET6.0을 활용한 중심성 분석과 구조적 등위성분석을 실시하였고 정량적 분석으로 SPSS 21.0을 이용하여 온라인 사이트에서 제시하는 서비스품질평가요인 4가지(‘음식’, ‘서비스’, ‘가치’, ‘환경’)과 만족도의 관계를 회귀분석 실시	구조적 등위성 분석 결과 4개의 그룹(‘환경’, ‘가치’, ‘음식’, ‘장소’)이 생성되었고, 서비스 품질 평가요인(‘음식’, ‘서비스’, ‘가치’, ‘환경’)과 만족도 분석 결과 ‘음식’, ‘서비스’, ‘환경’은 만족도에 영향을 미친다.	Text mining, UCINET, SPSS
임영희· 김홍범	2019	호텔경영학연 구	호텔 온라인 리뷰 빅데이터를 활용한 감성분석에 관한 연구	트립어드바이저에서 외래 관광객을 중심으로 국내호텔(비즈니스호텔 5개, 제주 리조트호텔 5개, 인천공항호텔 5개) 15곳을 선정, 데이터는 2010년부터 2018년까지의 영문리뷰	R을 이용하여 빈도분석, 엘리스틱 넷 회귀분석, 선형 회귀분석을 실시, 텍스트 마이닝 실시하고, 호텔에 대한 감성사전을 구축,	온라인 리뷰의 극성값이 고객 평점에 긍정적인 영향을 미치는 것으로 나타났고, 온라인 리뷰 텍스트 마이닝 분석과	Python, R, Sentiment analysis, Text mining, Elastic net regression

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
				데이터 수집, 2018년 12월 1일부터 약 한달 간 웹 크롤링을 통해 데이터 수집	온라인 리뷰 감성 극성값과 고객 평점을 활용한 회귀분석 실시	감성분석을 실시하여 만족과 불만족 단어를 추출하여 호텔 품질 속성을 새롭게 파악하였다. 외래 관광객들이 남긴 온라인 리뷰를 수치화하여 통계 분석한 감성 극성값은 전체 평점과 관계가 있었다.	
곽민정· 최지유· 박소현	2019	관광경영연구	호텔 서비스 고객만족도 분석을 위한 온라인 리뷰 감성분석	트립어드바이저를 활용하여 데이터 수집은 자바 기반의 웹 크롤러(JAVA-based Wed Crawler)와 파싱(parsing)을 도와주는 전용오픈 소스 라이브러리 jsoup 1.10.2를 사용하였고, 서울 5성급 호텔을 대상으로 2010년 8월부터 2018년 12월까지 작성된 총 4,821개의 국문 리뷰를	2010년 8월부터 2014년 12월까지 리뷰 283개는 감성분석을 위한 속성사전을 구축하고 분석방법 검증하는 데 사용, 2015년 1월부터 2018년 12월까지의 리뷰 4,538개는 서울 5성급 호텔 31개 중 비중 1%미만의 호텔(13개)들은	10개의 대표속성별 개인 리뷰의 속성 언급비율과 긍·부정비율, 만족지수, 주요 세부속성, 주요 연관어를 도출하였다. 언급비율에서 중요도가 높게 나타난 것은	

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
				수집, 개별 리뷰 데이터는 호텔명, 호텔지역(도시), 호텔등급, 리뷰일자, 고객유형, 리뷰 평점, 그리고 리뷰 내용의 7가지 항목을 포함	제외시켜, 리뷰 수 상위 18개의 호텔만 대상으로 선정하고 총 4,203개의 리뷰를 분석, ‘KNU 한국어 감성사전’을 기반으로 감성사전 구축, 개별 리뷰의 세부속성 연급유무분석	‘객실환경’, ‘서비스’, ‘호텔시설’, ‘식음료’순으로 나타났고, 긍·부정비율은 속성을 언급한 리뷰 중 각각의 비율로 나타났고, 만족지수는 객실환경에서는 ‘룸컨디션/청결’, ‘전망’이 긍정적인 세부속성으로 나타났고, ‘냄새’, ‘방음’ 등은 부정적인 세부속성으로 나타났다. 대표사례인 ‘호텔 A’를 통해 감성분석 모델이 호텔별 서비스 특성과 품질수준을 잘 보여줄 수 있으며 품질관리 및 서비스 마케팅에 기여	

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
심영석· 김홍범	2018	관광학연구	온라인 리뷰 빅데이터 기반의 Word2Vec기 법을 활용한 관광지 개성과 여행객 평점 간 구조적 관계 분석	트립어드바이저에서 2004년 7월 10일부터 2017년 2월 28일까지 누적된 여행객 리뷰를 웹크롤러를 통해 수집	데이터 전처리 Word2Vec, TF-IDF 텍스트 가중치 산출, 12개 카테고리의 117개소의 세부 목록 작성, OLS회귀분석	코사인 유사도 분석 실시한 결과 코사인 유사도 분석을 통해 관광지 개성과 기존 브랜드 개성 구성개념 간의 큰 차이를 확인, 향후 척도 개발에 유용하게 활용 가능, WLS를 통한 모형 추정 결과, 관광지 개성 5가지(진실성, 흥미로움, 유능한,세련됨)은 소비자의 평점에 유의한 영향을 미치는 것으로 확인	web crawling, Word2Vec, Python, OLS, WLS
류일한· 주재현· 김동연· 안재현	2017	한국경영과학 회 학술대회논문 집	에어비앤비 호스트의 리뷰 피드백이 게스트의 구매 의도에 미치는 영향	2016년 10월 15일부터 2016년 10월 25일까지 10일간 수집 2016년 12월 10일, 1인 기준에 1박으로 숙박 기간 설정하여 그에 해당하는 숙소를 대상으로 수집,	숙소당 고객의 위시 리스트 클릭 수를 통해 게스트 구매의도 정도를 파악, 텍스트 분석을 실시하여 TF-IDF 값 확인, LSA를 이용한	호스트가 피드백 코멘트를 많이 남기면 남길수록 잠재적 게스트의 구매 의도가 늘어남 호스트 피드백이 많을수록	text analysis(TF- IDF, LSA)

연구자	년도	학회지	연구주제	수집기간	방법론	주요도출결과	주요 고려사항
				영어권 국가(미국, 영국, 캐나다, 호주)의 주요도시의 27,966개 숙소 대상으로 65,536쌍의 리뷰와 피드백 코멘트 수집, 에어비앤비 공식 페이지에서 웹크롤러를 이용하여 수집, Python Beatifulsoup을 이용하여 파싱(Parsing), 리뷰 데이터는 Json-API 가공	유사도 분석,	숙소판매에 정(+)의 영향을 줌 반복적인 어휘를 사용하여 피드백을 할 경우 호스트 신뢰도는 떨어지고, 구매의도 부(-)의 영향을 줌	

<온라인 리뷰 국외 선행연구>

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
Mingming Cheng & Xin Jin	2019	International Journal of Hospitality Management	What do Airbnb users care about? An analysis of online review comments	by December 2016, online review was obtained Airbnb website with a total of 181,263 review comments from Sydney	analysis a 'big data' set of online review comments through the process of text mining and sentiment analysis	Three main attributes, location, amenities and host form the majority of content of the written comments on Airbnb user's online review portal, Sentiment analysis result is negative sentiments are 'noise', 'price' is not identified as a key influencer	unsupervised learning, compared with existing attributes reported in the hotel literature
Abdelaziz Lawani, Michal R. Reed, Tyler Mark, Yuqing Zheng	2019	Reginal Science and Urban Economics	Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston	Airbnb platform for Boston and retrieved from Inside Airbnb during the month of September 2016,	reviews are cleaned of punctuation, number, extra spaces and non-textual contents, remove stop words,	prices are strategic complements and are influenced by the review score, the characteristics of the room, and the features of the neighborhood. marketing	Review, Price competition, Sentiment analysis, ols regression

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
				2051 individual hosts	N-gram-based, average of the review score per room, used sentiment analysis methodology, used AFINN dictionary, word cloud, Moran's I statistic used to test for the price data	implication in that consumers respond to the contents of online review. in addition to customer rating.	
Hak-Seon Kim & Yoojeong Noh	2019	Journal of Mechanical Science and Technology	Elicitation of design factors through big data analysis of online customer reviews for washing machines	the data on customer reviews were collected using a web crawling from March 10, 2014 to June 21, 2017. the total number of reviews	using the smart crawling and text mining(SCTM), the frequency analysis of review and percentage value of the brand ratings, semantic network	displays the results of the regression analysis between the independent variable customers' experience (CE) and the dependent variable customers' satisfaction (CS). The overall variance explained by the	to derive the engineering specifications for products, a more accurate model of the customer satisfaction may be needed, so that it might be necessary to

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
				collected from the Bestbuy is 14,785.	analysis, the degree centrality and eigenvector centrality, CONCOR analysis, factor analysis of quantitative analysis, regression analysis, This study found the main design factors for the front loading washer design by the big data analysis, factor analysis, and regression analysis of data collected from customer reviews.	seven predictors was 12.5 %($R^2 = 0.125$). In social and behavioral science field, regression models related to opinion mining are impossible to include all relevant input variables to predict output variables such as opinion from text data, and thus, they often have low R^2 values . Also, the model for customer satisfaction is intended to obtain the relative weights of the	further refine the data in the data preprocessing and use machine learning models with higher accuracy than the regression models used in this study.

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
						major product attributes affecting customer satisfaction, but it is not intended to quantitatively predict the customer satisfaction. The ANOVA result shows that F and its corresponding p-value ($F(7,144776) = 302.41, p < 0.001$) confirm the significance of the model. Thus, the research model can be still accepted	
Judith Bridges, Camilla	2018	Current Issue in Tourism	If nearly all Airbnb reviews are positive, does that make them	It was conducted in two phases.	textual analyses to better understand the	Airbnb reviews tend to comprise a very restricted set	eWOM, discourse analysis, online

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
Vásquez			meaningless?	The initial data set collected for this study consists of 400 publicly available reviews posted on Airbnb.com. It was collected manually in August and September 2015 and then samples reviews from four large cities located in each quadrant of the USA	travel reviews written by consumers. initial data were collected, used a concordancing software AntConc, provides information about wrd frequencies and other pattern of word co-occurrence.	of linguistic resource, establishing the site's norm of highly positive commentary, simultaneously responding to hospitality scholar's calls for analyses which extend beyond the star ratings and also take into account consumers' constructions of experience in the review texts themselves.	consumer reviews

<텍스트 마이닝 국내 선행연구>

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
박종도	2019	한국문헌 정보학회지	토픽 모델링을 활용한 다문화 연구의 이슈 추적 연구	한국연구재단에 등재 및 등재 후보 학술지를 대상으로 2007년부터 2018년까지 '다문화' 키워드 포함, 논문 5,345개 수집	토픽모델링 분석	총 5개의 토픽을 구성, 다문화 연구 트렌드를 연구동향 파악가능, 최근 중요 관심사를 식별	-
이혜진· 최자영	2019	글로벌 경영학회지	텍스트 마이닝 기법을 활용한 편의점 이용 고객의 트위터 리뷰에 대한 감성분석	트위터에서 2018년 한 해 동안의 '씨유' 또는 '지에스'를 언급한 트윗을 수집 18,622개 수집	파이썬 3.7을 이용하여 웹 크롤링, 텍스트 마이닝 감성단어추출, SPSS를 이용하여 집단 간 차이 분석과 회귀분석 실시	온라인 텍스트를 이용하여 텍스트와 감성의 관계를 분석, 편의점 음식과 관련해선 '맛'의 영향력이 가장 높게 나타남	제품의 양과 가격보다 맛을 우선시 해야 한다는 결론
홍태호· 니우한잉· 임강· 박지영	2018	정보시스템 연구	LDA를 이용한 온라인 리뷰의 다중 토픽별 감성분석: tripadvisor 사례를 중심으로	R을 활용하여 'rvest'함수 패키지를 이용하여 온라인 데이터 수집, 2010년 1월부터 2017년 7월까지 세계 7개 도시(런던, 파리, 방콕,	웹크롤링을 통해 평가점수, 리뷰, 고객 유형, 속성별 평가 점수 활용하여 토픽모델링 분석과 감성분석	LDA를 통해 추출된 키워드를 바탕으로 6가지 속성과 관련된 최종 키워드를 도출	다른 언어연구 기반으로 일반화 고려

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
				베이징, 홍콩, 서울, 싱가포르)의 호텔들의 온라인 리뷰 수집	실시		
강윤수· 김민수· 홍창환· 김성백· 권상철	2018	Asia-pacific Journal of Multimedia Service Convergent with Art, Humanities and Sociology	빅 데이터 분석 도구 R언어를 이용한 교육 자료 시각화	-	R 프로그래밍 언어를 사용하여 여러 가지 기법들의 시각화 자료를 교육자료로 사용하는 것에 대한 통찰 연구	시각화 방안으로 '파이차트', '워드 클라우드', '바차트', '마인드 맵', '구글 맵' 등을 사용	초고등 교육 개편과 4차산업혁명 에 따른 R 교육과정 진행
길호현	2018	우리말글	텍스트마이닝을 위한 한국어 불용어 목록 연구	국립국어원 제공하는 말뭉치 통계 2002년부터 2006년까지 80만 어절, 7만 3천 문장 을 연구 대상으로 선정	R을 활용, 'KoNLP'함수 패키지를 이용, 한국어 텍스트 마이닝 분석 실시, 'SimplePos09'함 수를 사용하여 형태소 분석과, 불용어 제거	형태소 10,000개를 추출, 불용어 목록 제시	영역별, 자연어 처리에 대한 한계
이수상	2016	한국도서관· 정보학회지	독후감 텍스트의 토픽 모델링 적용에 관한 탐색적 연구	<가족의 두얼굴> 2012년 출판된 책의 독후감 23편(일반부	독후감의 문서를 토픽 모델링하고 추출된 주제	독후감 문서에서 총 16개의 토픽을 추출하였고, 그	향후 토픽을 추출하고 감성분석을

연구자	년도	학회지	연구주제	수집 기간	방법론	주요도출결과	주요 고려사항
				8편, 고등부 9편, 중등부 6편)으로 구분 수집	범주를 언어네트워크의 노드로 사용하여 관계연구를 진행, R을 사용하고, ‘ topicmodels’ 함수를 사용	16개의 토픽을 활용하여 언어네트워크 분석하였다. 각각의 노드들의 중심성값을 토대로 노드 간의 연관성 정도를 파악했다.	진행을 제언

