



## 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](#)

박사학위논문

관광지 만족도를 평가하는 관광지  
선택속성에 관한 연구 : 텍스트  
마이닝과 퍼지셋 질적비교분석의 활용

A research on the tourists' destination selection  
attributes and satisfaction : Using text mining  
and fuzzy-set qualitative comparative analysis

2022년 2월

세종대학교 대학원  
호텔관광경영학과  
박 지 원

관광지 만족도를 평가하는 관광지  
선택속성에 관한 연구 : 텍스트  
마이닝과 퍼지셋 질적비교분석의 활용

A research on the tourists' destination selection  
attributes and satisfaction : Using text mining  
and fuzzy-set qualitative comparative analysis

지도교수 이 형 룡

이 논문을 호텔관광경영학 박사학위논문으로 제출함

2022년 2월

세종대학교 대학원  
호텔관광경영학과  
박 지 원

박지원의 박사학위논문을 인준함

2022년 2월

심사위원장 안 윤 영 (인)

심사위원 최 충 범 (인)

심사위원 한 진 수 (인)

심사위원 임 영 희 (인)

심사위원 이 형 룡 (인)

## 국문초록

2020년 국민여행조사에 따르면, 내국인은 여행할 때 자연과 풍경을 감상하는 활동의 비중이 높은 것으로 나타났다. 실제로 자연관광에 대한 국민의 관심과 수요가 높아지고 있으며, 자연관광지는 대한민국의 관광 시장에서도 중요한 부분을 차지하는 것으로 나타났다. COVID-19 상황과 더불어 개인의 심신 회복, 스트레스 완화, 웰빙과 관련있는 자연관광은 국민의 관심사이면서, 동시에 떠오르는 연구의 대상이다. 이에 본 연구는 다양한 관광지 중에서도 국내에서 잘 알려진 자연관광지 5개소를 대상으로 관광지 선택속성과 관광지 만족도 간의 관계를 탐구하였다.

한편, 관광에 대한 동향 파악과 관광객들의 행동 원인을 규명하는 것에 유용한 단서를 제공하는 것 중 하나는 관광지에 대한 온라인 리뷰 데이터이다. 온라인 리뷰는 관광객이 여행 정보를 획득하기 위한 새로운 정보 채널로 인식되고 있으며, 평점만으로 전부 파악하기 어려운 관광지에 대한 구체적이면서 최신의 정보를 제공한다. 따라서, 관광지에 대한 전반적인 만족도를 이해하기 위해서는 방문객의 사실적인 정보와 경험이 담겨있는 온라인 리뷰 데이터를 시의적절하게 활용하는 연구가 진행되어야 한다.

관광지 만족도를 설명하기에 앞서, 우선적으로 요구되는 사항은 관광객이 고려하는 관광지 선택속성을 이해하는 것이다. 이에 본 연구는 보편화된 설문조사 방식에서 벗어나 실시간으로 축적되는 트립어드바이저 리뷰 데이터를 연구에 활용하였다. 연구의 대상이 되는 관광지는 한국관광100선에 5회 연속 선정된 자연관광지 5개소이며, 이는 모두 대한민국을 대표하는 관광지라고 할 수 있다. 여기에는 ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도가 포함된다.

본 연구의 진행 절차를 요약하면 다음과 같다.

첫 번째, 한국관광100선에 5회 연속 선정된 자연관광지를 대상으로 텍스트 마이닝을 하여 리뷰 데이터를 수집하고, 명사만 추출하여 빈도 분석을 하였다. 이때, 빈도 높은 단어들은 관광객들이 리뷰를 작성할 때 주로 언급하는 내용이라고 할 수 있다. 빈도 분석을 통해 관광객들이 전반적인 관광지 만족도와 함께 고려하는 사항에 대하여 단어 중심으로 확인하였다.

두 번째, TF-IDF 가중치를 적용한 LDA 토픽모델링을 통해 리뷰 내용에 존재하는 주요 관광지 선택속성(토픽)을 도출하였다. 각각의 관광지마다 여섯 개의 토픽을 도출하였으며, 여기에는 자연관광지 5개소에서 공통으로 도출되는 토픽도 존재하였고, 관광지마다 독특하게 발견되는 토픽도 존재하였다. LDA 토픽모델링 기법을 사용해서 리뷰에 존재하는 핵심 토픽을 파악하는 것은 관광지 방문객들의 전반적인 인식과 이들이 리뷰에서 언급하는 내용을 단어가 아닌 토픽 단위로 이해할 수 있기에 중요하다.

세 번째, FsQCA 분석을 통해 결과변수(관광지 만족도)를 평가하는 원인조건(관광지 선택속성)의 다양한 조합을 확인하였다. 예를 들어, 5점이라는 관광지 만족도 점수가 부여된 리뷰가 10개 있다고 가정했을 때, 이 10개의 리뷰 내용이 모두 동일한 토픽으로 구성되어 있다고 단정하기는 어려울 것이다. 그 이유는 각각의 리뷰 내용을 구성하는 토픽, 그리고 해당 토픽들의 조합이 리뷰마다 다르게 나타날 수 있기 때문이다. 따라서, 동일한 결과(예: 관광지 만족도 점수가 5점일 때)를 설명하는 상황에서도 서로 다른 원인조건 조합이 발생할 수 있는 상황임을 전제해야 한다. 이에 본 연구는 퍼지 셋 논리를 적용하였으며, FsQCA 분석에 대한 활용으로 보다 정교한 차원에서 관광지 선택속성과 관광지 만족도 간의 관계를 설명할 수 있었다. 우선, LDA 토픽모델링에서 얻어진 관광지 선택속성(토픽)을 원인조건으로, 관광지 만족도를 결과변수로 하여 필요조건 성립여부를 검증하였다. 다음으로 는 관광지 만족도를 결정하는 다양한 원인조건 조합을 나타내는 진리표

를 도출하였다. 진리표를 토대로 관광지 만족도를 평가하는 원인조건조합의 대표 모형을 도출하였으며, 충분조건 검증에서는 표준분석을 사용하여 최적화된 결과를 제시하는 중간 해법을 토대로 결과를 해석하였다. 그 결과, 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 확인할 수 있었다. 즉, 이러한 결과는 모형 설명력과 모형 일치도가 높은 모형을 토대로 어떤 관광지 선택속성의 조합으로 인해 관광지 만족도가 평가되는지 설명할 수 있기에 유용하다.

종합하면, 리뷰 데이터는 정성 데이터에 가까우며 만족도 점수는 정량 데이터에 가깝다고 할 수 있다. 이처럼 성격이 다른 두 가지 형태의 데이터를 동시에 활용하는 본 연구의 시도는 관광지 방문객의 만족도를 설명하는 관광지 선택속성의 다양한 조합을 탐색적으로 연구하기 위한 체계적인 연구 절차를 제시하였다는 점에서 의미가 있다. 온라인 리뷰에 드러나는 관광지 선택속성을 밝히는 것뿐만 아니라, FsQCA 분석을 통해 관광지 만족도를 평가하는 관광지 선택속성의 다양한 조합을 제시함으로써 결합된 관점에서 인과관계를 설명하였다. 특히 본 연구는 LDA 토픽모델링에 대한 결과를 FsQCA 분석으로 확장할 수 있는 발전 가능성을 보여주었다.

**주요어:** 관광지 선택속성, 관광지 만족도, 온라인 리뷰, 텍스트 마이닝, LDA 토픽모델링, 퍼지셋 질적비교분석





## 목 차

제 1 장 서 론 .....	1
제 1 절 연구의 배경 .....	1
제 2 절 연구의 목적 .....	5
제 3 절 연구의 방법 및 범위 .....	8
제 4 절 연구의 흐름도 .....	11
제 2 장 이론적 배경 .....	13
제 1 절 관광지 선택속성 .....	13
1. 관광지의 분류 .....	13
2. 관광지 선택속성의 개념 .....	15
3. 관광지 선택속성의 중요성 .....	17
4. 관광지 선택속성의 구성요인 .....	18
5. 관광지 선택속성의 선행연구 동향 .....	23
제 2 절 관광지 만족도 .....	25
1. 관광지 만족도의 개념 .....	25
2. 관광지 만족도의 중요성 .....	26
3. 관광지 만족도의 선행연구 동향 .....	28
4. 관광지 선택속성과 관광지 만족도 간의 관계 .....	30
제 3 절 온라인 리뷰 빅데이터와 텍스트 마이닝 .....	32
1. 온라인 리뷰의 개념 및 중요성 .....	32
2. 빅데이터의 개념 및 중요성 .....	34

3. 텍스트 마이닝의 개념 및 중요성 .....	36
4. 텍스트 데이터의 활용 .....	37
제 4 절 LDA 토픽모델링 .....	47
1. LDA 토픽모델링의 개념 및 중요성 .....	47
2. LDA 토픽모델링의 활용 .....	48
3. LDA 토픽모델링의 선행연구 동향 .....	49
제 5 절 FsQCA(퍼지셋 질적비교분석) .....	52
1. FsQCA의 개념 및 중요성 .....	52
2. FsQCA의 활용 .....	54
3. FsQCA의 선행연구 동향 .....	56
 제 3 장 연구 설계 .....	 60
제 1 절 연구과제 설정 .....	60
제 2 절 자료 수집의 대상 및 범위 .....	64
1. 자료 수집을 위한 커뮤니티 선정 .....	64
2. 자료 수집의 대상 .....	65
3. 자료 수집의 범위 .....	65
4. 조사설계 요약 .....	69
제 3 절 분석 절차 및 방법 .....	71
1. 리뷰 데이터 수집 .....	71
2. 데이터 전처리 및 형태소 분석 .....	72
3. 기본 분석 .....	74
4. LDA 토픽모델링 .....	75
5. FsQCA(퍼지셋 질적비교분석) .....	76

제 4 장 실증 분석 .....	81
제 1 절 기술통계 분석 .....	81
1. 온라인 리뷰의 정보 및 표본의 특성 .....	81
제 2 절 관광지 리뷰 데이터에 대한 빈도 분석 .....	83
1. 설악산 국립공원 단어 도출 빈도 분석 .....	84
2. 한라산 국립공원 단어 도출 빈도 분석 .....	85
3. 순천만습지(순천만국가정원) 단어 도출 빈도 분석 .....	86
4. 남이섬 단어 도출 빈도 분석 .....	87
5. 우도 단어 도출 빈도 분석 .....	88
6. 관광지 5개소에 대한 빈도 분석 결과 .....	90
제 3 절 LDA 토픽모델링 .....	94
1. 설악산 국립공원 토픽모델링 .....	95
2. 한라산 국립공원 토픽모델링 .....	98
3. 순천만습지(순천만국가정원) 토픽모델링 .....	101
4. 남이섬 토픽모델링 .....	103
5. 우도 토픽모델링 .....	105
제 4 절 FsQCA(퍼지셋 질적비교분석) .....	110
1. 설악산 국립공원 FsQCA .....	113
2. 한라산 국립공원 FsQCA .....	119
3. 순천만습지(순천만국가정원) FsQCA .....	123
4. 남이섬 FsQCA .....	128
5. 우도 FsQCA .....	132
제 5 장 결 론 .....	137
제 1 절 연구의 요약 .....	137

제 2 절 시사점 및 제언 .....	140
1. 학문적 시사점 .....	140
2. 실무적 시사점 .....	143
제 3 절 한계점 및 향후 연구방안 .....	148
 참 고 문 헌 .....	 151
1. 국내문헌 .....	151
2. 국외문헌 .....	162
 Abstract .....	 165

## 표 목 차

<표 2-1> 관광 유형 및 관광지 분류 .....	14
<표 2-2> 관광지 선택속성에 관한 선행연구 .....	21
<표 3-1> 한국관광100선에 5회 연속 선정된 관광지 유형별 분류 .....	67
<표 3-2> 본 연구의 분석 대상이 되는 자연관광지에 대한 기본 정보 ·	69
<표 3-3> 조사설계 요약 .....	70
<표 3-4> OKT를 활용한 정규화 처리 작업 전후 비교 .....	73
<표 3-5> 연구의 분석 방법 및 절차 요약 .....	80
<표 4-1> 관광지별 온라인 리뷰 정보 및 표본의 특성 .....	82
<표 4-2> 설악산 국립공원 단어 도출 빈도 분석(n = 1,997) .....	84
<표 4-3> 한라산 국립공원 단어 도출 빈도 분석(n = 3,028) .....	85
<표 4-4> 순천만습지(순천만국가정원) 단어 도출 빈도 분석(n = 3,938) .....	86
<표 4-5> 남이섬 단어 도출 빈도 분석(n = 1,782) .....	87
<표 4-6> 우도 단어 도출 빈도 분석(n = 6,126) .....	88
<표 4-7> 관광지별 공통된 주요 빈출 단어 .....	92
<표 4-8> 설악산 국립공원 토픽모델링 결과 .....	97
<표 4-9> 한라산 국립공원 토픽모델링 결과 .....	99
<표 4-10> 순천만습지(순천만국가정원) 토픽모델링 결과 .....	103
<표 4-11> 남이섬 토픽모델링 결과 .....	105
<표 4-12> 우도 토픽모델링 결과 .....	107
<표 4-13> 설악산 국립공원 만족도 필요조건 검증 결과 .....	113
<표 4-14> 설악산 국립공원 진리표 .....	114
<표 4-15> 설악산 국립공원 만족도 충분조건 검증 결과 .....	118

<표 4-16> 한라산 국립공원 만족도 필요조건 검증 결과 .....	119
<표 4-17> 한라산 국립공원 진리표 .....	120
<표 4-18> 한라산 국립공원 만족도 충분조건 검증 결과 .....	122
<표 4-19> 순천만습지(순천만국가정원) 만족도 필요조건 검증 결과 ..	123
<표 4-20> 순천만습지(순천만국가정원) 진리표 .....	124
<표 4-21> 순천만습지(순천만국가정원) 만족도 충분조건 검증 결과 ..	127
<표 4-22> 남이섬 만족도 필요조건 검증 결과 .....	128
<표 4-23> 남이섬 진리표 .....	129
<표 4-24> 남이섬 만족도 충분조건 검증 결과 .....	131
<표 4-25> 우도 만족도 필요조건 검증 결과 .....	132
<표 4-26> 우도 진리표 .....	133
<표 4-27> 우도 만족도 충분조건 검증 결과 .....	136

## 그 립 목 차

<그림 1-1> 연구의 흐름도 .....	12
<그림 3-1> 실제 리뷰 예시 .....	71
<그림 4-1> 관광지 및 관광지 5개소 전체에 대한 워드클라우드 시각화 결과	89

# 제 1 장 서 론

## 제 1 절 연구의 배경

2020년 국민여행조사<sup>1)</sup>에 따르면, 2020년 국내여행 경험률은 86.4%이며, 관광여행<sup>2)</sup> 경험률은 75.5%에 달하는 것으로 확인된다. 2020년 우리나라 만 15세 이상 전국민의 1인 평균 국내여행 횟수는 4.95회이며, 관광여행은 3.75회로 조사되었다. COVID-19의 영향으로 2019년 대비 국내여행 경험률, 여행 횟수, 여행 일수, 여행 지출액 등은 감소하였으나, 위드코로나 및 백신 공급 등으로 인해 관광여행 경험률, 관광여행 횟수 등이 점차 회복되는 상황이다. 한편, 관광여행을 할 때 방문객들이 여행지에서 하는 활동으로는 ‘자연 및 풍경 감상’(78.1%), ‘휴식/휴양’(57.6%), ‘음식관광(지역 맛집 등)’(53.9%)이 높은 비중을 차지하는 것으로 나타났다. 특히 자연 및 풍경 감상과 휴식/휴양, 음식관광에 대한 활동 비율은 2018년부터 증가 추세를 보이고 있다(문화체육관광부, 2020). 이를 통해 자연관광에 대한 국민의 관심과 수요가 높아지고 있으며, 자연관광지는 국내 관광 시장에서도 중요한 부분을 차지하고 있음을 유추할 수 있다.

최근 한국관광공사 보도자료에 따르면, COVID-19 상황은 한국인들의 국내여행 지형을 크게 바꿔놓았다고 소개한다. COVID-19의 영향으로 2020년에는 전국 지역 방문자가 2019년도 대비 평균 18% 감소하였다. 하지만 이런 가운데에도 잘 알려지지 않은 숨겨진 관광지, 비대면 자연관광지, 캠핑장, 수도권 공원 등에 대해서는 오히려 방문자 수가 증가한 것으로 나타났다. 상위 검색되는 관광지점을 살펴본 결과, 2019년까지는 에버랜드와 롯데

---

1) 변경되기 전의 공식 명칭은 ‘국민여행실태조사’임

2) 여행의 주된 목적이 관광·휴양인 경우를 의미함

월드가 나란히 1위, 2위를 차지했으나 2020년도에는 자연관광지인 여의도 한강공원과 을왕리 해수욕장이 1위, 2위를 차지한 것으로 보도되었다. 특히 2020년에는 공원, 바다와 같은 자연관광지가 상위 검색지점을 대다수 점유하였다(한국관광공사, 2021). 이러한 상황을 토대로 다양한 관광지 중에서도 자연관광지를 대상으로 연구하는 것에 대한 필요성을 제기할 수 있었다.

현대인들은 변화 없는 일상, 과중한 업무, 누적된 스트레스 등을 벗어나 깨끗하고 좋은 자연환경을 찾아 관광 활동에 참여하려는 현상을 보인다(김미라, 2020). 즉, 자연관광에 참여함으로써 다양한 치유를 체험하고, 심신 회복과 신체 건강유지, 삶의 질 개선 등을 추구하려는 것이다. 이러한 현상은 앞으로도 지속될 것으로 예측된다. 한편, 최근에는 관광과 일상에 대한 구분이 점차 사라짐에 따라 주 5일 근무제의 정착, 일과 삶의 균형에 대한 관심 등으로 우리 사회의 여가문화는 변화를 겪기 시작하였고 동시에 스마트폰의 보급으로 숫자, 텍스트, 이미지, 동영상 등의 다양한 형태의 데이터가 방대하게 생성되는 현상을 맞이하였다(고선영·정근오, 2021). 이러한 사회적 상황에서 관광 트렌드 파악과 관광 행동의 원인을 규명하는데 유용한 단서를 제공하는 것 중 하나는 관광지에 대한 온라인 리뷰 데이터이다.

온라인 리뷰는 온라인 구전을 발생하는 과정에서 존재하는 자연어 형태의 데이터로 관광지에 대한 정보를 파악하기 위해 다양한 부문에서 유용하게 활용된다. 관광은 소비보다 구매가 선결되는 특수성으로 인해 온라인 리뷰에 대한 의존도가 타 산업보다 크며, 실제로도 온라인 구전효과가 여행객들의 정보 탐색, 여행 계획의 수립, 소비에 이르기까지 관광 의사결정 단계에 직접적으로 개입함으로써 잠재적인 관광수요를 창출하는 것으로 알려져 있다(Litvin, Goldsmith, & Pan, 2008). 온라인 리뷰는 많은 관광객들이 여행 정보를 획득하기 위한 새로운 채널로 인식되고 있으며 관광객이 직접 방문해서 경험한 내용을 기반으로 작성되는 정보이므로 공식적으로 제공하는



정보보다 더 신뢰하는 양상을 보인다(김은미, 2021). 온라인 리뷰는 평점만으로 전부 파악할 수 없는 관광지에 대한 구체적이며 최신의 정보를 제공하고, 관광지에 대한 긍정적인 리뷰는 해당 리뷰를 감상하는 관광객의 방문욕구를 자극하므로, 이는 관광지 결정에도 중요한 영향을 미칠 수 있다. 그러므로 전반적인 관광지 만족도를 이해하기 위해서는 방문객의 사실적인 정보가 반영된 온라인 리뷰 데이터를 활용하는 연구가 선행되어야 한다.

관광과 관련된 다양한 데이터 중에서도 온라인 리뷰에 대한 데이터 수집과 활용은 앞으로 대한민국의 관광 시장을 예측하기 위해서 중요하게 고려되어야 한다. 그러나 데이터를 단순히 많이 보유하고 축적하는 시스템만을 갖추고 있는 것과, 축적된 온라인 리뷰에서 의미 있는 정보를 수집하고 결론을 도출하는 것은 또 다른 연구의 영역이다. 그러므로 온라인 리뷰 데이터의 특징을 고려해서 리뷰 데이터를 적절하게 활용하여 대한민국 주요 관광지의 개선점을 도출하고 방문객 만족도에 기여하는 관광지 선택속성을 확인하는 것은 관광 산업의 발전을 염두한 시사점을 발굴하기 위해서도 중요하며, 이러한 리뷰 데이터 활용은 시의적절하게 이루어져야 한다.

본 연구는 보편화된 데이터 수집 방식인 설문조사에서 벗어나 실시간으로 축적되는 온라인 리뷰 데이터를 활용하여 리뷰에 존재하는 유용한 정보를 도출하는 것에 초점을 두었다. 이에 본 연구는 질적 자료인 리뷰 데이터를 수집하고 분석에 활용하기 위한 구체적인 방법을 제시하고자 하였다. 이는 대부분의 리뷰 데이터가 자연어를 기반으로 하는 비정형 데이터로 이루어져 있으므로 정형 데이터의 상태를 가정하고 이루어지는 기존의 전통적인 통계 분석 방법이 적절하지 않다는 주장을 토대로 하였다(정지훈·정혜인·이준기, 2021). 따라서 본 연구는 다양한 문서로부터 비정형 또는 반정형 텍스트 데이터를 획득하여 자연어 처리 기술을 기반으로 패턴이나 관계를 도출하여 새로운 가치를 창출하고 의미 있는 정보를 파악하는 방법인

텍스트 마이닝 기법을 활용하였다. 텍스트 마이닝 기법을 통해 트립어드바이저에 공개된 온라인 리뷰 데이터를 활용하여 관광지에 대한 관광객의 전반적인 만족도 평점, 방문객의 일반적인 사항과 정보, 그리고 텍스트로 구성된 리뷰 내용을 수집하여 관광지 선택속성을 도출하고 관광지 선택속성과 만족도 간의 관계를 검증하였다.

## 제 2 절 연구의 목적

온라인 리뷰 데이터를 통해 관광지 만족도를 평가하는 관광지 선택속성을 도출하려는 이유는 다음과 같다. 관광객의 행동을 파악하기 위해서 중요한 것은 관광지 선택속성을 이해하는 것이다(봉미희 · 한주형, 2020). 관광지를 하나의 상품으로 볼 때 다른 관광지와 경쟁하기 위해서는 해당 관광지가 가지는 핵심 매력을 연구하고 강화해야 하며, 타 목적지와 차별화를 시도하기 위해서는 관광지 선택속성에 대한 연구를 통해 지역 매력성을 파악하고 해당 지역이 보유하는 여건과 실정을 충분히 반영하는 관광개발이 이루어졌을 때 지역의 정체성과 특성이 내재된 경쟁력 있는 관광자원 및 상품을 개발할 수 있다(여영숙, 2016). 이 과정에서 관광지 선택속성은 관광지에서 방문객이 경험하고 보고 느끼는 것을 포함하며, 관광과 밀접한 관련이 있는 방문객의 물리적, 사회적, 행동적 특징을 의미하므로 중요하다.

관광지 선택속성과 연관되어 분석되는 주요 변수로는 관광태도(김현철, 2018; 정균영, 2020), 관광지 이미지(김현철, 2018; 차승희 외, 2020; 진동봉 · 박현지, 2021), 행동의도(정효숙, 2017; 신형섭, 2019; 안세길, 2020; 유순호, 2020), 재방문의도(박선심, 2016; 강삼 · 추승우, 2021) 등이 존재하나 본 연구에서는 관광지 선택속성과 함께 연구되는 주요 변수로 관광지 만족도를 고려하였다. 그 이유는 본 연구에서 데이터를 수집하고자 하는 온라인 커뮤니티인 트립어드바이저에서 리뷰에 대한 내용을 제공함과 동시에 전반적인 만족도를 수치화해서 제공하기 때문이다. 기존의 선행연구에서는 어떠한 원인변수의 결과로 만족도에 영향을 미칠 것이라는 가설의 형태로 관광지 선택속성과 만족도에 관한 연구가 진행되었으나(신민정 · 원문가 · 송학준, 2018; 박혜영, 2019; 최일선 · 홍장원 · 이정아, 2019; 송옥방 · 차석빈, 2020; 안세길, 2020), 본 연구에서는 관광지 방문객의 실제 만족도 점수인 만

족도 평점을 분석에 활용하였다. 트립어드바이저는 방문객의 리뷰와 함께 전반적인 만족도를 5점 평점으로 수치화해서 제공하므로 실제적인 관광지 만족도 점수를 분석에 활용할 수 있다. 이러한 방법은 텍스트 마이닝을 통한 데이터 수집 후 분석 단계에서 설문지법을 통한 샘플링의 한계를 극복하는 대안이 되며, 국내 주요 관광 명소를 방문하는 관광객들이 해당 관광지를 경험한 후에 스스로 자유롭게 작성하는 리뷰 내용을 토대로 관광지 선택속성을 탐색하는 차원에서 폭넓은 의견을 확인할 수 있는 것이 장점이다.

온라인 리뷰는 특정 관광지를 방문하는 관광객들의 자발적인 구전 정보이며, 방문객이 자유롭게 작성한 형태의 글이다. 따라서 기존의 선행연구에서 검증되지 않았던 관광지 만족도에 기여하는 사실적인 관광지 선택속성을 파악할 수 있는 것이 특징이다. 그러므로 본 연구는 온라인 리뷰 데이터에 포함된 자연관광지의 주요 선택속성을 도출하고, 관광지 선택속성과 방문객의 실제 만족도의 관계를 검증하는 것을 연구의 주된 목적으로 하였다. 본 연구는 리뷰 데이터에 존재하는 관광지 선택속성을 도출하기 위하여 LDA 토픽모델링을 제안하였다. 또한, 관광지 만족도를 평가하는 리뷰에서 발견되는 관광지 선택속성을 도출하였으며, 이를 퍼지셋 질적비교분석(Fuzzy-set Qualitative Comparative Analysis, 이하 FsQCA)으로 확장하였다.

FsQCA 분석을 통하여 리뷰에서 반드시 언급되는 필요조건에는 어떤 관광지 선택속성이 존재하는지 밝혔으며, 관광지 만족도를 평가하는 데 기여하는 관광지 선택속성들의 조합을 검증하였다. 즉, 관광객의 만족도를 구성하는 개별 관광지 선택속성만을 규명하는 것이 아니라, 여러 속성들의 조합을 제시하면서 결합된 관점에서 관광지 선택속성과 관광지 만족도의 관계를 설명한 것이다. FsQCA 분석을 적용하면 보다 정교한 차원에서 관광지 만족도를 평가하는 관광지 선택속성을 검증할 수 있다. 따라서 본 연구는 대한민국 자연관광지에 대한 방문객의 온라인 리뷰 데이터를 시의적절하게

활용하고, 리뷰에 존재하는 정량화된 평점인 만족도 점수를 활용하여 리뷰 데이터에 드러난 관광지 선택속성과 해당 선택속성들의 조합을 확인할 수 있는 구체적인 연구 절차를 제시하였다는 점에서 다른 연구와 차별화된다.

LDA 토픽모델링 결과에 FsQCA 분석을 접목하면 관광지 만족도를 구성하는 필요조건과 충분조건을 구분하여 제시할 수 있다. 본 연구는 관광지 유지 및 개선을 위해 반드시 고려되어야 하는 관광지 선택속성과 이 속성들 간의 조합을 밝히는 것이 주된 목적이다. 따라서 계량적 척도를 통해 연구되었던 기존의 관광지 선택속성에 관한 연구와는 다르게, 본 연구는 관광객들이 실제로 방문해서 경험하고 작성한 온라인 리뷰 데이터를 활용하고, 이를 실증 분석에 적용함으로써 관광지에 대한 전반적인 이슈와 특성을 확인하여 실무적으로 적용 가능한 시사점을 제시하는 것에 목적을 두었다.

### 제 3 절 연구의 방법 및 범위

본 연구의 방법 및 범위는 다음과 같다. 연구 진행에 앞서 연구의 배경과 필요성, 연구의 목적을 제시하였으며, 주요 변수인 관광지 선택속성과 관광지 만족도에 관한 문헌을 고찰하였다. 관광지 선택속성과 관광지 만족도는 서로 밀접한 연관이 있으므로, 관련 선행연구를 통해 이들 변수 간의 관계를 검토하면서 연구에 적용된 분석 기법인 텍스트 마이닝, LDA 토픽모델링, FsQCA에 관한 내용도 이론적 배경의 고찰 범위에 포함하였다.

본 연구는 설문지법의 양적 연구가 지니는 샘플링 문제 등의 한계점을 보완하고자 「한국관광100선」에 5회 연속 선정된 자연관광지를 대상으로 텍스트 마이닝을 하여 리뷰 데이터를 수집하고, 연구 목적 달성을 위한 분석을 진행하였다. 우선 관광지 방문객의 실제 경험과 정보가 담긴 온라인 커뮤니티를 선정하고 데이터 수집에 필요한 파이썬 기반의 전용 웹크롤러를 사용하여 온라인 리뷰 데이터를 수집하였다. 여기에는 대한민국을 대표하는 관광지라고 할 수 있는 한국관광100선에 5회 연속 선정된 자연관광지 5개소가 포함된다. 한국관광100선은 한국인은 물론 한국을 찾는 외국인들이 꼭 가볼 만한 대표 관광지를 소개하는 것으로 100개소를 2년에 한 번씩 선정하여 홍보하는 사업이다. 한국관광100선은 전문가 심사와 통계 분석 등 다양한 검증을 거쳐 선정되는데(문화체육관광부, 2021), 1차 서면평가, 2차 현장점검, 3차 최종 선정위원회의 엄격한 심사를 거쳐 선정되므로 여기에 포함되는 관광지는 대한민국을 잘 나타내는 관광지라고 할 수 있으며 본 연구의 목적을 달성하기 위한 범위로도 적합하다. 관광학계와 여행업계, 여행기자·작가 등 관련 분야의 전문가가 심사에 참여하며 이동통신사, 내비게이션, 누리소통망 SNS의 빅데이터 등을 평가에 활용하므로 여기에 포함되는 관광지는 대한민국을 대표할 만한 관광지라고 하기에 무리가 없을 것

으로 판단된다. 연구의 대상이 되는 관광지 5개소는 ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도이다.

한편, 관광지와 관련된 연구에서 대상으로 삼을 수 있는 표본은 크게 내국인 방문객과 외국인 방문객으로 분류할 수 있다. 리뷰의 양만 보았을 때, 외국인이 작성한 리뷰의 비중이 대부분의 경우에서 내국인이 작성한 리뷰의 양보다 많은 것으로 나타났다. 외국인 방문객이 고려하는 관광지 선택속성을 규명하는 것도 중요하지만, 본 연구는 한국관광100선에 5회 연속 선정된 관광지를 중심으로 우선 내국인이 고려하는 관광지 선택속성을 도출하는 것이 중요하다고 판단하였다. 따라서 본 연구는 내국인 방문객이 작성한 국문 리뷰 데이터만을 본 연구의 자료 수집 및 분석의 대상으로 하였다.

자료 수집 및 분석의 대상이 정해진 상태에서 파이썬 기반의 웹크롤러를 개발하여 방문객이 작성한 온라인 리뷰 데이터를 수집하였으며, 자연어 형태로 존재하는 리뷰 데이터는 곧바로 분석에 사용될 수 없으므로 분석에 활용 가능하도록 처리하는 과정을 거쳤다. 데이터를 전처리 작업하고, 파이썬 기반의 오픈소스 한국어 분석기인 OKT를 사용하여 분석에 사용될 명사를 추출하고, 불용어 사전을 구축하여 분석에 포함하지 않을 단어들을 함께 처리하였다. 데이터 전처리 및 형태소 분석이 완료된 이후에는 연구과제 수행을 위한 기본 분석을 하였다. 우선 관광지의 기본 정보 및 표본의 특성을 확인하기 위해 기술통계 분석을 하였으며, 관광지별 리뷰에 자주 등장하는 단어를 제시하고자 도출된 명사를 중심으로 빈도 분석을 하였다.

정제된 데이터를 바탕으로 관광지 5개소 별로 리뷰에 존재하는 관광지 선택속성을 도출하기 위해 TF-IDF 가중치를 적용한 LDA 토픽모델링을 제안하고, 리뷰에 존재하는 관광지 선택속성에는 무엇이 있는지 살펴보기 위해 LDA 토픽모델링 결과를 FsQCA 분석에 활용하였다. FsQCA 분석에는 관광지 선택속성(토픽)이 원인조건으로, 관광지 만족도가 결과변수로 사용

되었다. 원인조건과 결과변수의 관계에서 필요조건 성립여부를 확인하였으며, 관광지 만족도를 평가하는 다양한 원인조건조합의 대표 모형을 도출하였다. 이로써 관광지 선택속성이 관광지 만족도를 평가하는데 있어 필요조건에 해당하는지와 충분조건에 해당하는지 여부를 함께 확인할 수 있었다.

종합하면, 리뷰 데이터는 정성 데이터에 가까우며 만족도 점수는 정량 데이터에 가까운데, 이러한 두 가지 형태의 데이터를 동시에 활용하는 본 연구의 시도는 방문객의 관광지 만족도에 영향을 미치는 관광지 선택속성의 다양한 조합을 탐색적으로 연구하는 것에 기여할 수 있다는 점에서 의미가 있다. 또한 온라인 리뷰 데이터에 반드시 나타난다고 할 수 있는 필요조건에는 어떤 관광지 선택속성이 존재하는지 밝혔으며, 관광지 만족도를 평가하는 관광지 선택속성의 다양한 조합을 제시함으로써 결합된 인과관계를 설명하고자 하였다. 연구 결과를 바탕으로 자연관광지의 특징 및 개선점을 도출하고, 방문객 만족도에 기여하는 관광지 선택속성을 확인함으로써 대한민국 주요 자연관광지가 고려할 수 있는 시사점을 제시하고자 하였다.



## 제 4 절 연구의 흐름도

제 1 장은 서론 부분이다. 제 1 절에서 연구의 배경을 언급하고, 제 2 절에서 연구의 목적을 제시하였다. 제 3 절에서 연구의 방법 및 범위를 서술하고, 제 4 절에서 연구의 흐름도를 그림으로 도식화하였다.

제 2 장은 이론적 배경 부분이다. 가장 먼저 연구의 주요 변수인 관광지 선택속성과 관광지 만족도에 관한 국내문헌과 국외문헌을 고찰하였다. 해당 변수들의 개념 및 중요성, 선행연구 동향을 주로 살펴보았으며, 제 1 절에서 관광지 선택속성을, 제 2 절에서 관광지 만족도에 관한 내용을 기술하였다. 제 3 절에서는 분석 방법과 연관 있는 온라인 리뷰 빅데이터 및 텍스트 마이닝에 관하여, 제 4 절에서는 LDA 토픽모델링을, 제 5 절에서는 FsQCA(퍼지셋 질적비교분석)에 관한 내용을 포함하였다.

제 3 장은 연구 설계 부분이다. 제 1 절에서 연구과제를 설정하고, 제 2 절에서 자료 수집의 대상 및 범위를 정하였으며, 제 3 절에서 본 연구의 전반적인 분석 절차 및 방법에 대하여 기술하였다.

제 4 장은 실증 분석 부분이다. 제 1 절에서 데이터 수집에 대한 결과로 기술통계 분석을 하였고, 제 2 절에서 관광지 리뷰 데이터에 대한 빈도 분석을 하였다. 제 3 절에서 LDA 토픽모델링을 통해 국내 자연관광지의 주요 관광지 선택속성(토픽)을 도출하였으며, 제 4 절에서 FsQCA 분석을 하였다.

제 5 장은 결론 부분이다. 제 1 절에서 연구의 요약과, 제 2 절에서 시사점 및 제언을 하였으며, 제 3 절에서 한계점 및 향후 연구방안을 제시하였다. 본 연구의 전체적인 흐름도는 <그림 1-1>과 같이 요약된다.

제 1 장 서 론			
연구의 배경	연구의 목적	연구의 방법 및 범위	연구의 흐름도



제 2 장 이론적 배경				
관광지 선택속성	관광지 만족도	온라인 리뷰 빅데이터와 텍스트 마이닝	LDA 토픽모델링	FsQCA (퍼지셋 질적비교분석)



제 3 장 연구 설계		
연구과제 설정	자료 수집의 대상 및 범위	분석 절차 및 방법



제 4 장 실증 분석			
기술통계 분석	빈도 분석	LDA 토픽모델링	FsQCA



제 5 장 결 론		
연구의 요약	시사점 및 제언	한계점 및 향후 연구방안

<그림 1-1> 연구의 흐름도

## 제 2 장 이론적 배경

### 제 1 절 관광지 선택속성

#### 1. 관광지의 분류

국립국어원 표준국어대사전에 따르면 “관광(觀光)”이란 ‘과거를 보러 감 또는 그런 길이나 과정’으로 다른 지방이나 다른 나라에 가서 그곳의 풍경, 풍습, 문물 따위의 구경을 나타내는 단어이며, “관광지(觀光地)”란 ‘경치가 뛰어나거나 사적(史跡), 온천 따위가 있어 관광할 만한 곳’이라는 뜻을 가지는 단어이다. 일반적으로 관광지는 관광객이 자신의 일상 생활공간을 떠날 수 있게 하는 자연 관광자원, 인문 관광자원, 그리고 관광시설로서의 매력성을 지니고 있다. 관광지에는 관광자원의 지역가치를 높이는 숙박시설, 식음료, 편의시설 등이 갖추어져 있으며, 관광지에 접근하기 편리한 교통시설이 설치되어 있어야 하고 관광지에 대한 정보제공 서비스가 이루어지는 일정한 장소를 의미한다(김민철·부창산·김영훈, 2008; 고선희·박은숙, 2011). 따라서 관광지는 관광객의 관광 동기 및 욕구를 만족시킬 수 있는 자원을 충분히 보유하고 있는 장소로서 관광 활동을 할 수 있는 숙박시설 및 식음료, 편의시설, 인프라 등의 제반 시설을 갖추어야 한다(유순호, 2020). TourAPI3.0<sup>3)</sup>에 따른 관광지 분류기준을 토대로 관광지 세부 유형을 제시하면 다음의 <표 2-1>와 같이 정리된다.

첫 번째, ‘자연관광’으로는 산, 계곡, 명승지, 나무, 숲, 연못·습지, 하천·강, 호수·저수지·포구, 해변·해수욕장, 섬, 휴양림·수목원, 공원·정

---

3) TourAPI: 한국관광공사의 관광정보 개방 서비스

원, 군락지·서식지, 샘·우물·약수터, 방조제·하구둑·만, 퇴적지·분지·분화구가 있다. 두 번째, ‘역사·문화재관광’으로는 불교, 왕실, 유교·양반, 인물관련, 정치·행정관련, 선사시대, 유네스코, 건축조형물, 기타, 기타 유적지가 있다. 세 번째, ‘체험관광’으로는 체험시설, 마을체험, 도보체험(길), 관광문화단지, 거리·골목, 전망대, 놀이공원·테마파크, 동·식물원·유원지, 촬영소·세트장, 온천·스파·찜질방, 휴양지·목장·농원, 레포츠가 있다. 네 번째, ‘문화시설’로는 박물관, 미술관, 기념관·역사관, 전시관, 극장·공연장, 문학관·문화공간, 생태관, 도서관·서점, 복합건물·컨벤션센터, 경기장·운동장이 있다.

<표 2-1> 관광 유형 및 관광지 분류

관광 유형	관광지 세부 유형
자연관광	산, 계곡, 명승지, 나무, 숲, 연못·습지, 하천·강, 호수·저수지·포구, 해변·해수욕장, 섬, 휴양림·수목원, 공원·정원, 군락지·서식지, 샘·우물·약수터, 방조제·하구둑·만, 퇴적지·분지·분화구
역사·문화재관광	불교, 왕실, 유교·양반, 인물관련, 정치·행정관련, 선사시대, 유네스코, 건축조형물, 기타, 기타유적지
체험관광	체험시설, 마을체험, 도보체험(길), 관광문화단지, 거리·골목, 전망대, 놀이공원·테마파크, 동·식물원·유원지, 영소·세트장, 온천·스파·찜질방, 휴양지·목장·농원, 레포츠
문화시설	박물관, 미술관, 기념관·역사관, 전시관, 극장·공연장, 문학관·문화공간, 생태관, 도서관·서점, 복합건물·컨벤션센터, 경기장·운동장

자료: TourAPI3.0 토대로 논자 재정리

## 2. 관광지 선택속성의 개념

속성이란 일반적으로 상품이 가진 유형 및 무형의 특징을 의미하며 상품은 이러한 속성들의 묶음으로 구성된다(Kotler, Makens, Bowen, & Baloglu, 2018). 선택속성은 소비자들이 특정 제품을 선택함으로써 얻는 만족을 극대화하기 위한 선택 기준이 된다(김동일 · 이승곤 · 진현식, 2012). 관광객은 의사결정 과정을 통하여 관광 목적지를 선택하며, 각 각의 의사결정 과정에서 여러 가지 개인적 · 사회적 · 문화적 요인의 영향을 받게 되는데, 관광객은 관광지의 특성에 의해 최종 목적지를 선택하며 이때 가장 큰 영향을 미치는 요소가 관광지 선택속성이다(이철진, 2008). 관광지 선택속성에 대한 정의와 개념은 선행연구마다 연구 목적에 따라 다양하게 나타나는 것을 확인할 수 있다. 주요 선행연구에 제시된 관광지 선택속성에 대한 정의와 개념을 살펴보면 다음과 같다.

관광지 선택속성은 관광지를 하나의 상품으로 지각하는 관광객들이 관광 목적지를 선택함에 있어 가장 중요하게 생각하고 선택하는 요소라고 할 수 있다(박선심, 2016). 즉 관광지 선택속성은 관광객을 유인할 수 있는 고유한 매력과 관광객의 욕구와 동기를 충족시킬 수 있는 속성의 집합체로, 관광지를 선택할 때 중요하게 고려되는 판단 기준이 된다(염수길 · 이재섭, 2016). 관광지는 관광 활동을 하기 위해 선택되는 다양한 행선지로 자연경관, 유적 및 사적지, 문화시설, 서비스, 이벤트 등으로 구성되는데 관광객의 관광 욕구와 동기를 충족시킬 수 있는 다양한 속성들로 구성되고 관광지가 지니고 있는 이러한 고유한 매력적인 요소를 통틀어 관광지 선택속성이라고 한다(송학준, 2012).

관광지 선택속성과 관련된 초창기 연구자라고 할 수 있는 Mayo & Jarvis(1981)에 따르면 관광객은 관광지 선택속성을 비교하고 이후 이를 평가

하며 선택하고 행동을 결정하는 데 관광지 선택속성이 포함된다고 주장하면서 관광지 선택속성은 관광객으로 하여금 관광지에 대한 태도를 측정하는 도구가 될 수 있다고 하였다. 즉, 관광지 선택속성은 관광객이 관광지를 선택할 때 중요하다고 생각해서 선택하는 속성이며, 방문객은 관광지가 가지는 속성을 비교하여 자신이 선호하는 관광지 선택속성을 지니는 관광지를 목적지로 선택하므로 관광지 선택속성은 그 자체로도 중요한 의미가 있는 것이다(김현 · 장호성, 2012). 따라서 관광지는 관광객을 유인할 수 있는 다양한 매력적인 관광 자원과 편리한 시설과 같은 긍정적인 속성을 갖추어야 한다(이현중, 2020). 관광지 선택속성을 설명할 때에는 관광객의 욕구 및 동기를 만족시킬 수 있는 요소를 보고 관광지 선택속성이라고 하였던 연구가 있으며(안성식 · 박효정 · 조아람, 2018), 관광객이 가지고 있는 사전 기대와 관광객의 경험 간의 상호작용 결과로서 전체적인 관광 경험에 대하여 관광객이 보이는 일련의 태도를 보고 관광지 선택속성이라고 하였던 연구도 있다(트린티하인 · 설훈구, 2018).

관광지 선택속성은 관광지를 방문하기 전 기대에 대한 숙박 · 식음료 · 편의시설 등의 다양한 관광 시설에서 이를 충족하기 위한 관광자원이 보유되어 관광객들에게 어떠한 형태로든 지속적인 선호 및 이용될 때 관광지로서의 그 가치를 인정받고 관광지 선택이나 행동에 영향을 미치는 중요한 요소가 될 수 있다(김지훈, 2020). 따라서 관광지는 관광객을 수용할 수 있는 경쟁력을 보유하고 관광객의 욕구를 충족할 수 있는 위락시설, 숙박시설, 식음료시설, 편의시설 등의 다양한 관광시설들과 서비스들로 구성된 다차원적 속성을 지닌다(김효중 · 김시중, 2012). 관광지에 대한 선택과 결정은 관광지 선택속성에 의해 결정된다고 보아야 할 것이며, 객관적인 관광 동기와 관광지 선택 행동에 대하여 정확한 측정을 하기 위해서도 관광지 선택속성에 대한 이해가 필요하다(김도훈 · 김지훈, 2015). 종합하면 다수의 선행

연구에서 밝혀진 대로 관광지는 자연경관, 역사적 장소, 시설, 서비스, 이벤트 등의 집합체이므로 관광객의 관광 욕구와 동기를 만족시킬 수 있는 다양한 속성들로 구성되며, 관광지가 지니고 있는 고유한 매력요소를 통틀어 관광지 선택속성이라고 할 수 있다(남은경 · 이승곤, 2018).

### 3. 관광지 선택속성의 중요성

관광지 선택속성은 관광에 대한 최종 목적지를 판단하는데 영향을 미치며(김동한, 2019), 관광 목적지를 둘러싼 환경이나 관광지 특성, 이와 관련된 관광객 행동을 이해하기 위해서도 관광지의 선택속성을 이해하는 것이 중요하다(봉미희 · 한주형, 2020). 관광지 선택속성에는 관광지에서 관광객이 보고, 느끼고, 경험하는 모든 요소이자 관광과 관련된 물리적, 사회적, 행동적인 모든 특성이 내재되므로 관광지는 다양한 요소들로 이루어진 하나의 복합체로서 지각되기에 관광객의 특성에 따라 선호 및 선택되는 관광지 선택속성에는 차이가 존재할 수 있다(박혜영, 2019). 실제로 관광객들이 관광지 혹은 관광 상품을 선택할 때 중요하게 여기는 특성이 관광지 선택속성이며, 이는 관광 후 관광지 방문객이 지각하는 만족도를 포함하기 때문에 중요하다(남은경 · 이승곤, 2018). 관광지가 다른 관광지와의 관계에서 경쟁우위를 확보하기 위해서는 다양한 전략적 접근이 필요하고, 관광지의 매력적인 선택속성이나 가치 있는 경험을 통해 관광객들의 높은 충성도를 구축하고 실현할 수 있어야 하며, 이러한 측면에서 관광지 선택속성은 관광객의 의사결정에 영향을 미치는 요인이므로 중요하다(김희정 · 김시중, 2012; 이익수, 2016). 즉, 관광지가 하나의 상품이라는 점에 입각했을 때, 다른 관광지와 경쟁하기 위해서는 해당 관광지가 가지고 있는 핵심 매력을 강화하고 타 관광지와 차별화하기 위한 마케팅 전략이 필요할 것인데 이를 위해서

고려해야 하는 주요 속성이 관광지 선택속성인 것이다(임화순·고계성, 2006; 장양례·김혜영, 2011). 그러므로 방문객이 선택하는 관광지는 관광지 선택속성에 의해 결정되며, 관광에 대한 원동력은 여가 소비자로 하여금 거주지를 벗어나 관광 목적지에 머무르는 동안 지출 행위를 발생시키고 지역 경제의 활성화를 도우며, 지역사회에 활기를 더하는 이점이 있으므로 관광지 선택속성이 중요하다고 할 수 있다(김연선·김영식, 2016).

#### 4. 관광지 선택속성의 구성요인

관광지 선택속성에 관한 연구는 무수히 다양하나 최근 15년 이내에 수행된 연구로써 국내 관광지 선택속성을 대상으로 하였던 관련 선행연구를 주로 고찰하였다. 본 연구는 자연관광지를 연구의 대상으로 하지만, 일반적으로 관광지 관련 연구에서 나타나는 관광지 선택속성을 살펴보고자 유형을 구분하지 않은 상태에서 관련 문헌들을 고찰하였다. 그 결과, 주요 선행 연구에서 언급된 관광지 선택속성을 제시하면 다음과 같다.

관광지 선택속성에 대한 세대별 비교 연구에서 관광지 선택속성으로 도출된 요인은 기반시설, 관광자원, 유명함, 접근성, 독특성이 있다(김성혁·오은지·권상미, 2009). 관광지 선택속성을 내부적 요인과 외부적 요인으로 구분한 연구에서는 내부요인 속성변수로 탈출, 자기실현, 휴식, 치료, 건강, 친지방문, 새로운 친구의 만남, 명성추구, 문화적 탐험, 탐험추구, 밤 생활즐김 등이 제시되었고, 외부요인 속성변수로 교통시설, 사람들의 친근감, 음식의 다양함과 질, 숙소 시설, 개인적 안정, 가격, 문화적 자원, 쇼핑하기 좋은 곳, 환경적 안정과 질, 관광지 이미지와 기대 등이 제시되어 전체 관광지 선택속성을 22개로 세분화한 연구가 있다(Hsu, Tsai, & Wu, 2009).

가족 관광객을 대상으로 한 관광지 선택속성 연구로는 가족 관광적합성,



여행비용, 안전 및 위생, 교통접근성, 관광시설, 관광매력성의 6개 요인이 제시되었고(윤정현, 2010), 강원도 동해안 지역을 방문한 관광객을 대상으로 한 연구에서는 관광지 선택속성으로 서비스품질, 물리적속성, 지각된위험, 접근성이 제시되었다(손병모, 2011). 태안 지역을 방문한 관광객을 대상으로 관광지 선택속성과 동기요인이 방문객 행동의도에 미치는 영향을 검증한 연구에서는 관광지 선택속성으로 가격 적정성, 광고 및 홍보 적정성, 구전 효과성, 관광 인프라 적정성, 자원 적정성, 편의 및 접근성 의 6개 요인이 제시되었으며(김현 · 장호성, 2012), 백제역사유적지구 방문객을 대상으로 관광지 선택속성과 경험적 가치가 행동의도에 미치는 영향을 검증한 연구에서는 관광지 선택속성으로 안락성, 서비스성, 소비성, 매력성의 4개 요인이 제시되었다(김희정 · 김시중, 2012).

해양레저스포츠 참가자를 대상으로 관광지 선택속성에 대한 중요도 및 실행도를 분석한 연구에서는 관광지 선택속성을 구성하는 요인으로 체험기반, 비용, 편의, 환경, 서비스의 5개 요인이 제시되었고(이종길 · 양재식, 2013), 익산 미륵사지 관광객을 대상으로 문화유산관광지의 선택속성을 제시한 연구에서는 관광지 선택속성으로 역사 · 문화적 가치, 관람정보, 진정성, 편리성의 4개 요인이 제시되었으며(류인평 · 최인경, 2014), 전주 한옥마을 방문객을 대상으로 관광지의 선택속성이 도시 브랜드 이미지에 미치는 영향을 알아본 연구에서는 관광지 선택속성으로 문화적 이미지, 환경의 편의성, 관광지 운영 프로그램, 주민의 친절함의 4개 요인이 제시되었다(전기홍 · 이종관 · 김남기, 2014).

관광지 유형에 따른 관광지 속성 및 제약요인이 관광지 선택에 미치는 영향을 검증한 연구에서는 관광지 선택속성으로 편의성, 지역성, 매력성, 활동성, 접근 및 친절성의 5개 요인이 제시되었으며(김범진 · 권태길 · 고희석, 2014), 감천문화마을 방문객을 대상으로 관광지 선택속성에 따른 시장세분

화를 진행한 연구에서는 관광지 선택속성으로 매력성, 경험성, 흥미성, 편의성의 4개 요인이 제시되었다(이광국 · 김경옥 · 전재균, 2016). 기후변화에 따른 국제관광산업 경쟁력 강화를 위한 관광지 선택속성에 대한 중요도 차이를 살펴본 연구에서는 국제관광지 선택속성을 관광형태, 동행자유형, 관광지 매력성, 관광유형, 접근성, 숙박부대시설의 6개 요인으로 제시되었으며(안성식 · 박효정 · 조아람, 2018), 방한 중국인 관광객을 대상으로 하는 연구에서는 관광지 선택속성이 프로그램, 서비스품질, 지각된 위험, 물리적 속성, 접근성의 5개 요인으로 제시되었다(이경숙 · 전보람, 2018). 최근에는 도시관광 선택속성 개선에 관한 연구를 통해 도시관광 목적지에 대한 선택속성을 편의 및 휴게시설, 매력성, 안전 및 청결도로 제시한 연구도 존재하며(김동한, 2019), 기장군 지역의 향토음식 이미지가 관광지 선택속성에 미치는 영향 관계를 살펴보기 위해 관광지 선택속성을 다룬 연구에서는 학습, 체험, 탐방, 쇼핑, 휴양, 시설의 6개 요인이 제시되었다(김현철, 2019).

한편, 섬 관광지 선택속성과 관련된 연구로는 관광지 선택속성으로 매력성, 연계성, 안정성, 편의성의 4개 요인이 제시되었고(한상겸, 2019), 한옥마을 선택속성의 중요도와 만족도에 관한 연구로는 관광지 선택속성으로 교통 및 편의성, 환대서비스, 오락성, 역사문화성, 프로그램의 5개 요인이 제시되었다(김연선, 2020). 전라북도 남원관광단지를 방문한 관광객을 대상으로 관광지에 대한 선택속성 및 만족도와 관광객 행동의도 간의 영향 관계를 살펴본 연구에서는 관광지 선택속성으로 매력성, 역사문화성, 프로그램, 편의성, 접근성, 서비스품질의 6개 요인이 제시되었고(김주진 · 신우진, 2020), 서울 지역을 중심으로 내국인 관광객의 도시 관광지 선택속성의 중요도 및 만족도에 관한 연구에서는 관광지 선택속성으로 접근 및 편리성, 오락성, 관광자원, 매력성의 4개 요인이 제시되었다(봉미희 · 한주형, 2020). 부산 지역 내 관광지를 방문한 경험이 있는 중국인 개별관광객을 대상으로 한 연

구에서는 관광지 선택속성으로 물리적 속성, 프로그램, 서비스품질, 접근성이 제시되었다(강삼 · 추승우, 2021).

관광지 선택속성과 관련된 선행연구를 고찰한 내용은 다음의 <표 2-2>와 같다. 종합하면 관광지 선택속성은 관광지에서 관광객이 경험하는 모든 요소로부터 시작되는 관광지만의 고유한 특성이다. 관광지 선택속성은 관광지 방문객이 지각한 사회적, 물리적, 행동적 측면에서 종합적으로 인지되는 특성을 의미한다고 할 수 있으며, 관광지는 여러 요소들로 구성된 하나의 복합체이므로 관광객의 특성에 따른 선택 및 선호에는 차이가 존재한다(박혜영, 2019). 그럼에도 편의성, 접근성, 매력성 등의 요인은 다수의 선행연구에서 관광지 선택속성과 관련하여 빈번하게 도출되는 것으로 나타나 전반적인 관광지 선택속성을 평가하는 데 있어 중요한 요인으로 확인된다.

<표 2-2> 관광지 선택속성에 관한 선행연구

연구자	관광지 선택속성
김성혁 · 오은지 · 권상미(2009)	기반시설, 관광자원, 유명함, 접근성, 독특성
Hsu, Tsai, & Wu(2009)	내부적 요인 : 탈출, 휴식, 자기실현, 치료, 건강, 새로 온 친구의 만남, 명성추구, 탐험추구, 친지방문, 문화 적 탐험, 밤 생활 즐김  외부적 요인 : 교통시설, 음식의 다양함과 질, 사람들 의 친근감, 숙소 시설, 개인적 안정, 가격, 쇼핑하기 좋은 곳, 문화적 자원, 관광지 이미지와 기대, 환경적 안정과 질
윤정현(2010)	가족 관광 적합성, 여행비용, 안전 및 위생, 교통접근성, 관광시설, 관광매력성
손병모(2011)	서비스 품질, 접근성, 물리적 속성, 지각된 위험

<표 2-2> 관광지 선택속성에 관한 선행연구

연구자	관광지 선택속성
김현 · 장호성(2012)	가격 적정성, 광고 및 홍보 적정성, 구전 효과성, 관광 인프라 적정성, 자원 적정성, 편의 및 접근성
김희정 · 김시중(2012)	안락성, 서비스성, 소비성, 매력성
이종길 · 양재식(2013)	체험기반, 비용, 편의, 환경, 서비스
류인평 · 최인경(2014)	매력성, 독특성, 편의성
전기홍 · 이종관 · 김남기(2014)	문화적 이미지, 환경의 편의성, 관광지 운영 프로그램, 주민의 친절함
김범진 · 권태길 · 고호석(2014)	편의성, 지역성, 매력성, 활동성, 접근 및 친절성
이광국 · 김경옥 · 전재균(2016)	매력성, 경험성, 흥미성, 편의성
안성식 · 박효정 · 조아람(2018)	관광형태, 관광지 매력성, 관광유형, 동행자유형, 접근성, 숙박부대시설
이경숙 · 전보람(2018)	프로그램, 서비스품질, 지각된 위험, 물리적 속성, 접근성
김동한(2019)	편의 및 휴게시설, 매력성, 안전 및 청결도
김현철(2019)	학습, 체험, 탐방, 쇼핑, 휴양, 시설
한상겸(2019)	매력성, 연계성, 안정성, 편의성
김연선(2020)	편의성, 환대서비스, 오락성, 역사문화성, 프로그램
김주진 · 신우진(2020)	매력성, 역사문화성, 프로그램, 편의성, 접근성, 서비스품질
봉미희 · 한주형(2020)	접근 및 편리성, 오락성, 관광자원, 매력성
강삼 · 추승우(2021)	물리적 속성, 프로그램, 서비스품질, 접근성

자료: 선행연구를 바탕으로 논자 재구성

## 5. 관광지 선택속성의 선행연구 동향

관광지 선택속성과 관련된 선행연구를 살펴보면 관광지에 대한 범위가 다양하므로 관광지를 지역별, 관광지 유형별, 국적별 등으로 세분화하여 특정 관광지 또는 유사한 관광지 집합에 대하여 관광지 선택속성을 도출하였던 연구가 주를 이루는 것으로 확인된다. 이러한 연구로는 방한 중국 관광객의 관광지 선택속성을 도출한 연구(김도훈·김영규, 2017; 이경숙·전보람, 2018; 강삼·추승우, 2021), 국내 도시 관광지 선택속성에 관한 연구(신형섭, 2019), 섬 관광객의 관광지 선택속성에 관한 연구(최일선·홍장원·이정아, 2019), 웰니스 관광지 선택속성에 관한 연구(김지훈, 2020), 다크 투어리즘 관광지 선택속성에 관한 연구(박상현·이미순, 2019), 온천 관광지 선택속성에 관한 연구(소네다·토루·최승담, 2019), 전주 한옥마을의 국내 전통문화 체험지에 대한 관광지 선택속성 연구(안세길, 2020)가 있다.

나아가 관광지 선택속성, 지각된 가치, 자기 효능감, 고객만족, 재방문의도 간의 복합적인 관계를 살펴본 연구도 존재한다(박선심, 2016). 남원 관광지를 대상으로 하였던 연구에서는 관광지 선택속성을 매력성, 역사문화성, 프로그램, 편의성, 접근성, 서비스품질로 구분하여 만족도 및 재방문의도와 추천의도에 미치는 영향을 검증하였으며(김주진·신우진, 2020), 해당 연구를 통하여 국내의 기존 관광지 관리와 신규 관광지 개발에 대한 방향성을 위한 시사점을 제시하였다. 이렇듯 관광지 선택속성에 관한 선행연구는 주로 관광지 선택속성이 독립변수로 사용되어 만족도, 재방문의도, 행동의도 등의 변수와 영향 관계에 있는 것으로 가정하고 연구가 진행되었다.

한편, 관광지 선택속성을 이용한 차이분석 연구도 존재한다. 세대별(베이비붐세대, X세대, N세대별)로 관광 동기와 관광지 선택속성 그리고 관광 행동의 차이를 알아본 연구가 여기에 해당하며, 해당 연구에서는 관광지 선택

속성으로 기반시설, 관광자원, 유명함, 접근성, 독특성의 5개 요인을 도출하고 세대별 관광지 선택속성의 차이를 검증하는 데 사용하였다(김성혁·오은지·권상미, 2009). 이 외에도 본 연구의 주요 변수인 관광지 선택속성과 관광지 만족도 간의 영향 관계를 검증하였던 선행연구도 확인된다. 관광지 선택속성이 만족도에 미치는 영향을 검증한 연구로는 관광지 선택속성을 만족도로 하여 행동의도에 미치는 영향을 규명한 연구(이경숙·하흠, 2018; 김주진·신우진, 2020; 안세길, 2020) 등이 있다. 이처럼 관광지 선택속성에 대한 연구는 특정 관광지에 대하여 의미 있는 속성을 도출하기 위해 연구되었을뿐만 아니라, 관광지에 대한 중요도 및 만족도 분석, 그리고 관광지 선택속성으로 시작된 방문객의 행동의도 등의 변수와 인과관계를 검증하기 위해 사용되었던 것을 확인하였다.

## 제 2 절 관광지 만족도

### 1. 관광지 만족도의 개념

관광지 만족도는 관광객의 주관적인 평가이며 관광객의 사전 기대보다 실제 관광 경험으로부터 발생하는 성과에 대한 관광객의 감정적인 긍정적 반응이다(신민정 · 원문가 · 송학준, 2018). 관광지 만족도는 관광 목적지에 대하여 관광객이 표출하거나 획득하는 긍정적인 지각 또는 기분으로, 관광 활동 및 선택에 대한 결과이며 관광 상황에 만족하고 기뻐하는 정도로 설명될 수 있다(유순호, 2020). 관광객은 저마다 관광 욕구에 따라 다양한 관광지 속성들을 탐색하며, 가장 적합한 관광 정보를 통해 관광지를 선택하고 관광 활동을 이어나간다. 이에 따라 관광지 선택과 관광 활동을 함에 있어서, 관광지 선택으로 얻는 관광 지식 정보는 관광객들의 관광지 만족도에도 중요한 영향을 미칠 수 있다(송래현, 2021). 관광지에 대한 종합적인 평가가 관광지 만족도로 나타날 수 있으며, 관광지에서 경험하는 숙박이나 편의시설, 체험 등을 중심으로 관광지 만족도가 평가될 수 있는 것이다. 이처럼 관광지 만족도는 특정 관광지에 대한 관광객의 긍정적인 지각 또는 느낌을 의미하지만 관광지 만족도를 결정하는 요소는 다양하게 존재한다.

관광지 만족도는 관광 활동 참여의 궁극적인 목적이 되는 중요한 개념이며 관광 활동을 하기 위해 지불한 경제적, 심리적, 사회적 모든 비용에 대하여 관광 활동 경험으로부터 얻은 주관적인 효용 또는 편익의 정도로 설명되므로 관광객의 만족을 명확하게 정의하거나 측정하는 것은 매우 어려운 편에 속한다(정지원, 2004). 그러나, 관광지 만족도는 다양한 이론과 접목해서 연구가 진행되고 있으며, 방문객의 관광지 만족도를 설명하기 위해 소개되는 이론 중 하나인 기대불일치 이론과 공정성 이론을 통하여 관

광지 방문객이 지각하는 전반적인 만족도를 측정할 수 있다(Jang & Feng, 2007). 기대불일치 이론은 기대와 성과를 비교하는 이론으로써 관광에 대한 만족이 높게 나타나는 것은 관광객이 기대한 것 이상의 성과를 경험하였다는 것으로 해석할 수 있다. 반면 공정성 이론은 투입된 비용과 시간 등 산출된 가치를 비교하는 이론으로 관광 만족이 높다는 것은 관광객이 투입한 비용이나 시간에 비하여 인식하는 가치가 더 높다는 것을 의미한다(김난영, 2012). 관광지 방문객의 만족도는 전반적 만족도와 구체적 만족도로 구분하여 측정할 수 있으며 전반적인 만족도는 관광 전 기대와 실제 관광 경험 간의 감정 비율로 나타낼 수 있고(Chen & Chen, 2010), 구체적인 만족도는 접근성, 숙박시설, 음식, 엔터테인먼트 요소 등의 관광지에서 경험할 수 있는 다양한 속성 또는 요소들을 종합하여 측정할 수 있다(Yi, 2009).

## 2. 관광지 만족도의 중요성

만족은 마케팅 분야에서 중요하게 다루어지는 개념으로 관광 분야에서도 매우 중요한 의미를 가진다. 이는 관광지 만족도가 관광지에 대한 서비스를 평가하고 재구매를 결정하는 중요한 판단 기준이 되기 때문이다(심우석, 2015). 관광지 만족은 전반적인 관광 경험에 대한 관광객의 평가임과 동시에 관광지에 대한 전체적인 평가가 될 수 있고, 잠재 관광객에게는 일종의 상품으로서 관광지를 선택하게 하는 요소가 된다(김현철, 2019). 따라서 관광지를 결정하거나 관광과 관련된 상품 및 서비스 구매와 재방문 의사에 직접적으로 영향을 미치는 핵심요인이 관광지 만족도라는 점에서 관광지에 대한 방문객의 만족도가 중요하다.

특정 관광지에 대하여 만족을 경험한 관광객은 해당 관광지를 다시 방문하거나 주변 사람에게 권유할 때 호의적이고 긍정적으로 전달하지만, 불



만족을 경험한 관광객은 해당 관광지를 다시 방문하지 않을뿐만 아니라, 주변 사람에게 권유하지 않으며 부정적인 구전을 형성하므로 해당 관광지에 대한 이미지 등 여러 가지 면에서 부정적인 영향을 미칠 수 있다(곡지평 · 김명종 · 유승동, 2019). 이러한 이유로 방문객의 만족도가 높은 관광지를 발굴하고 개선하는 것은 관련 여행 상품이나 해당 관광지가 속해있는 지역에 대한 관심도가 높은 고객을 유치하는 데에도 도움이 된다. 관광지 만족도는 개인의 욕구나 관광 내용 및 환경 요인에 대한 충족 정도, 관광을 위한 결과나 보상에 대한 만족에 따라서 다르게 나타나므로 개인의 지각이나 주관적인 평가가 높게 작용한다. 관광지 만족도는 주관적 관점에 따라 달라지는 특성으로 인해 어떤 관광 환경에서 만족도가 높게 나타났는지 단적으로 정의하기는 쉽지 않지만, 관광 만족은 일반적으로 관광 현장에서 관광객이 설정한 목표를 달성하여 관광을 이루는 여러 가지 요소에 대하여 만족한 정도를 나타내며 결과에 대한 성취도와 만족도 등을 기준으로 관광지의 질과 서비스 수준을 평가하는 자료로 활용된다(정용복, 2020).

관광지 만족을 높이기 위해서는 다양한 체험 프로그램, 지역 기업체 견학 프로그램, 문화 및 역사지 학습 프로그램 개발, 향토음식 거리조성 등을 통해 관광 인프라를 구축하는 것이 중요하다(김헌철, 2019). 관광 인프라를 구축함으로써 관광지에 대한 만족도를 높일 수 있으며, 이미 특정 관광지에 대하여 방문객의 관광지 만족도가 높게 나타난 것은 해당 관광지가 관광 상품을 판매하는 데 있어 적절한 인프라를 갖추었다는 것을 의미한다. 또한 여가활동에 대한 관광지 만족이 매우 주관적일지라도 해당 관광지에 대한 만족도가 중요한 이유는 관광지 방문을 통해 얻은 만족도가 방문객의 만족스러운 삶에도 긍정적인 영향을 미칠 수 있기 때문이다(윤정현, 2015).

현대인들은 바쁜 일상의 스트레스에서 벗어나 신체적 · 정신적 회복에 대한 관심이 높아지고 있기 때문에 여가시간의 증가와 더불어 개인 생활

만족을 위해 몸과 마음의 여유를 찾고 삶의 질을 증진할 수 있는 차원에서 관광에 대한 관심 또한 증대되고 있다(김지훈, 2020). 즉, 삶의 질을 높이고 건강과 행복을 추구하기 위해서도 관광에 대한 수요가 증가하고 있으며 이를 위해서도 관광지 만족도에 대한 이해가 중요하다는 것을 의미한다.

### 3. 관광지 만족도의 선행연구 동향

관련 연구로는 상품이나 서비스 만족도를 측정하기 위해 사용되는 비교·분석 기법인 IPA(Importance-Performance Analysis : 이하 IPA)를 관광 분야에 접목한 선행연구가 다수 존재한다(김도훈·김영규, 2017; 진문도·주은지·윤해진, 2019; 김지훈, 2020). IPA 분석 방법은 시각적 판단이 쉽고 고도의 학문적 지식이 없어도 일선 실무자들이 이용하기에 용이하다는 장점이 있으며, 속성의 순위를 통하여 만족도를 분석하기 위한 방법으로 적절하다고 알려져 있다(Martilla & James, 1977). 특히 실무자들이 마케팅 전략을 세우고자 할 때 비용과 시간의 한계를 가지므로 어떤 부분에 우선순위를 두고 역량을 집중할 것인가에 대하여 분석을 통해 답을 제시해 주고 있으며, 수요자가 상품 혹은 서비스 이용 전 각 요인들의 중요도와 이용 후 만족도를 스스로 평가하고 각 요인별로 중요도와 만족도를 동시에 비교분석하는 평가방법으로 널리 사용된다. 이러한 IPA를 사용한 연구로는 대표적으로 웰니스 관광지의 선택속성에 따른 중요도와 만족도를 검증한 연구(김지훈, 2020), 부산시 중구를 중심으로 도시관광 선택속성의 중요도와 만족도를 분석한 연구(이정실, 2020), 중국 개별관광객을 대상으로 관광지 선택속성의 중요도와 만족도를 검증한 연구(김도훈·김영규, 2017; 진문도·주은지·윤해진, 2019)가 있다. 관광지 선택속성에 대한 중요도 및 만족도를 검증하기 위해 IPA 기법을 활용한 연구에서는 대부분 관광지 만족도를 측

정하기 위해 이용 전에는 각 속성의 중요도를, 이용 후에는 만족도를 사용자가 스스로 평가하도록 하고 각각의 속성에 대한 상대적 중요도 및 만족도를 동시에 비교분석하는 평가 방법을 사용한다. 한편, IPA의 한계를 보완하기 위해 성취도를 자연로그(ln)로 변환하고, 전반적 만족도와 편상관분석을 통해 산출된 편상관계수를 활용하는 통계적 절차를 통해 도출된 상대적 중요도를 활용하는 수정된 IPA 활용 연구도 있다. 대표적으로 여수를 방문한 관광객을 대상으로 하여 수정된 중요도 및 실행도 분석을 통해 관광객들의 관광 목적지 선택속성을 살펴본 연구가 이에 해당한다(이현종, 2020).

이와는 다르게 방한 단체 관광객을 대상으로 여행상품의 목적지 구성에 따라 한국 여행에 대한 만족도의 차이를 분석한 연구도 있다(최승묵, 2020). 해당 연구에서는 방한 여행 특성에 따른 단체관광 여행상품의 방문 광역지 자체 수 차이, 단체관광 여행상품의 방문 광역지 자체 수별 한국 여행 만족도 차이, 단체관광 여행상품의 방문광역지 자체 구성별 한국 여행 만족도 차이 등을 분석하였으며 검증을 위한 통계적 기법으로 분산분석을 사용하였다. 이 외에도 관광지 선택속성이 관광 만족도에 미치는 영향을 검증하자 다중회귀분석을 적용한 연구가 있으며(박혜영, 2019; 최일선·홍장원·이정아, 2019), 관광지 선택속성과 관광객 만족 간의 관계뿐만 아니라 관광객 만족이 재방문의도에 미치는 영향을 검증한 연구도 있다(강삼·추승우, 2021). 이를 구조적인 관점에서 보는 연구도 있는데, 지속 가능한 관광증진 계획과 관광 분야의 연구에 기여하고자 하는 목적을 가지고 전주 한옥마을 방문객들을 대상으로 전통문화 가치인식, 관광지 선택 행동, 만족도, 및 행동의도 간의 관계를 구조모형을 통해 검증하였던 연구가 확인된다(안세길, 2020). 해당 연구에서는 전주 한옥마을 관광객의 전통문화 가치 인식에 대한 만족도를 검증하고 만족도와 행동의도 간의 관계를 규명하였다.

관광지 선택속성과 관광지 만족도를 사용하여 구조모형으로 접근한 또

다른 연구로서 도시 관광객들의 지각된 가치, 만족 및 행동의도 간의 관계를 살펴본 연구도 존재한다(신형섭, 2019). 서울을 방문해본 경험이 있는 중국 현지인을 대상으로 도시 관광 선택속성 만족요인과 전반적 만족도 간의 영향 관계를 다룬 연구도 있으며, 해당 연구를 통해 전통적인 대칭적 분석과 고객만족 세요인 모델을 활용한 비대칭적 분석을 수행함으로써 만족도 강화에 유의한 관광지 선택속성을 검증하였다(송옥방·차석빈, 2020). 이처럼 관광지 만족도는 관광지 선택속성과 관련해서 매개변수 또는 종속변수로 연구모형에 사용되기도 하며, 방문객의 전반적인 긍정적인 감정을 나타내기 위한 변수로 사용되고 있음을 확인하였다.

#### 4. 관광지 선택속성과 관광지 만족도 간의 관계

관광객은 방문하려는 관광지를 결정하기 위해 관광지가 가지는 다양한 시설 및 서비스 등의 여러 가지 제반 사항을 고려한다. 관광지에 대한 선택속성은 관광 만족도에 긍정적인 영향을 미치는 것으로 나타났다(임화순·강영순, 2008). 기장군의 향토음식 이미지가 관광지 선택속성에 미치는 영향 관계를 분석한 연구에서는 학습, 체험, 탐방, 쇼핑, 휴양, 시설 등 6개의 요인 중 학습을 제외한 모든 속성이 전반적인 만족도에 영향을 미치며 향토 음식이미지가 체험, 탐방, 휴양 속성과 만족도의 관계에서 조절효과를 보이는 것으로 밝혀졌다(김현철, 2019). 전라북도 남원관광단지를 방문한 관광객을 대상으로 하였던 연구에서는 관광지 선택속성을 매력성, 역사문화성, 프로그램, 편의성, 접근성, 서비스품질 등의 6개 요인으로 분석하였는데 그 결과, 매력성과 접근성, 서비스 품질이 만족도에 유의한 영향을 미치는 것으로 나타났다(김주진·신우진, 2020).

부산을 방문하는 중국인 관광객이 부산지역을 관광지로 선택하는 데 있

어 영향을 미치는 선택속성을 검증한 연구에서는 접근성을 제외한 나머지 물리적 속성, 서비스품질 속성, 프로그램 속성이 관광객 만족에 유의한 정(+)의 영향을 미치는 것으로 나타났다(강삼·추승우, 2021). 국내 전국 단위의 섬 관광객을 대상으로 섬 관광 선택속성과 만족도 간의 관계를 검증한 연구에서는 관광지 선택속성으로 자연·환경적 특성을 제외한 사회·문화적 특성과 여행 합리성이 관광 만족도에 유의한 정(+)의 영향을 미치는 것으로 밝혀졌다(최일선·홍장원·이정아, 2019).

관광지 선택속성과 관광지 만족도 간의 관계를 검증했던 다수의 선행연구를 고찰한 결과, 관광지 선택속성은 관광객의 의사결정이나 선택에 중요한 영향을 미치는 것으로 나타났다. 관광지를 하나의 상품으로 보았을 때 관광지가 지니고 있는 개별적인 요소를 관광지 선택속성이라고 할 수 있으며(봉미희·한주형, 2020), 관광지 선택속성은 관광지 만족도에 영향을 미치므로 관광지 만족도를 설명하기 위한 핵심 변수로서의 역할을 한다.

### 제 3 절 온라인 리뷰 빅데이터와 텍스트 마이닝

#### 1. 온라인 리뷰의 개념 및 중요성

온라인 리뷰는 소비재 시장에서 발생하는 재화나 서비스 등과 관련한 정보를 이용 후기 및 평점 등을 통하여 개인의 만족 또는 불만족의 경험을 표현하는 것으로, 온라인 플랫폼이라는 매개체를 통해 소비자들 간에 실시간으로 의사소통을 할 수 있는 정보 교환과정으로 정의된다(박정환 · 이병철, 2014). 온라인 시장의 가치만큼 중요성이 높아진 것이 있는데 이는 고객의 리뷰이다. 고객은 온라인에서 제품을 구매하고자 할 때 다양한 정보 탐색 과정을 거치는데, 이러한 탐색 단계에서 가장 쉽게 접할 수 있는 정보가 해당 제품에 대하여 이전의 사용자들이 남겨놓은 온라인 리뷰이기 때문이다(김승환 · 임혜빈 · 이병관, 2021). 온라인으로 제품을 구매하는 소비자들은 제품을 직접 시험해볼 기회가 없으므로 온라인 리뷰를 신뢰하고 의존하는 경향이 크다(최자영 · 김현아 · 김용범, 2020).

이러한 이유로 온라인 리뷰는 기업 마케팅의 주요 수단으로 활용되고 있으며, 많은 웹사이트에서 필수적으로 갖추는 구성 요소로 발전하고 있다. 이러한 현상은 관광지를 방문하려는 고객이 온라인 상에서 관광지를 탐색하는 과정에서도 유사하게 적용된다. 관광지를 결정하기 전에 관광객이 우려하는 것 중 하나는 관광지에 대한 정보 부족으로 겪는 관광지 선택의 불만족일 것이다. 실제로 관광객이 다수의 관광지를 비교 분석하여 가장 관심 있는 관광지를 선택하고자 할 때 관광객 개인의 배경지식 부족으로 인해 원하는 관광지를 찾지 못하는 상황과 직면할 수 있다(안진현 · 임동혁, 2020). 관광을 하나 이상의 장소를 방문하여 풍경, 문물 등을 구경하는 행위라고 개념화하였을 때, 특정 관광지에서 다음 관광지를 적절하게 선택하

기 위해서도 온라인 리뷰가 하나의 단서로 활용될 수 있다. 관광 분야에서는 실제 관광지를 방문한 고객이 관광지에 대한 정보나 여행경험에 대한 의견을 온라인 리뷰를 통해 제공하고 있으며, 온라인 리뷰는 잠재 고객들에게 새로운 여행 정보로 인식되고 있다(Law, Buhalis, & Cobanoglu, 2014).

만일 관광지에 대하여 긍정적인 평가가 이루어진다면 예비 방문객은 방문할 의향이 생기지만 관광지에 대하여 부정적인 평가가 이루어진다면 방문을 꺼리게 될 것이다(김은미, 2021). 온라인 리뷰는 소비자의 경험에 의해 작성된 주관적인 평가라고 할 수 있으며, 특별한 관계가 없는 모든 사람과 공유할 수 있다는 점에서 오프라인에서 발생하는 구전보다는 사실적이고 신뢰할 수 있다. 그러므로 관광객을 대상으로 하는 연구에서도 이러한 온라인 리뷰의 특성을 반영하는 연구가 진행되어야 한다.

트립어드바이저에서는 리뷰 내용과 함께 관광지 방문객의 닉네임, 해당 관광지의 방문 시기, 전반적인 만족도 평점, 동반자 유형 등을 함께 제공한다. 이러한 트립어드바이저의 리뷰 정책은 관광지에 대한 정보를 얻고자 하는 다른 관광객의 의사결정에도 유용한 정보를 제공한다. 자신과 유사한 타인으로부터 작성된 리뷰는 자신도 리뷰어와 같은 경험을 얻을 수 있을 것이라는 인상을 심어주어서 유용하고 신뢰할 수 있으며 도움 된다고 인식하는 것이다(김승환 · 임혜빈 · 이병관, 2021). 즉, 소비자들은 자신과 유사한 경험을 전제로 하는 타인이 제공하는 정보를 습득할 때 더욱 관심을 가지고 해당 리뷰에 대한 영향을 받을 수 있는 것이다. 그러므로 온라인 리뷰는 관광지에 대한 정보 탐색 및 발굴에 유용한 정보를 제공하므로 관광객의 관광지 결정에 중요한 역할을 한다고 할 수 있다. 또한 온라인 리뷰는 온라인 구전의 한 형태로 텍스트나 이미지 등을 포함하고 있으며, 상품 및 서비스에 대한 온라인 고객의 정보 및 생각을 공유하고 의견을 표출할 수 있는 매개체의 역할을 하므로 중요하다(Guy, Mejer, Nus, & Raiber, 2017).

## 2. 빅데이터의 개념 및 중요성

빅데이터는 데이터의 양, 생성 주기, 수치 데이터뿐만 아니라 문자와 같은 비정형 데이터를 포괄하는 형식을 가지고 있어 과거 데이터에 비해 규모가 크고, 형태가 다양하여 기존에 행해지던 방법으로는 수집, 저장, 검색, 분석이 어려운 방대한 크기의 데이터를 의미한다(한국콘텐츠진흥원, 2014). 빅데이터는 단순히 데이터의 양이 많은 것이 아니라 다양한 형태의 데이터가 실시간 생성되는 대용량(High Volume), 고속도(High Velocity), 고다양성(High Variety)의 특징을 지닌다. 빅데이터는 방대한 양의 데이터를 통해 기업의 방향과 조직의 다양한 문제 해결책을 제공함과 동시에 산업 전반과 사회 자체를 변화시킬 수 있는 새로운 방향을 제시할 수 있어 중요한 관심사로 떠오르고 있다(Kim, 2013). 빅데이터에 대한 사회적 관심이 깊어지는 가운데 국가적 차원에서도 관광 빅데이터에 대한 정보 제공과 관련된 서비스 개발을 앞당기고 있다.

2021년 2월에는 문화체육관광부와 한국관광공사가 관광 빅데이터 플랫폼인 ‘한국관광 데이터랩’ 서비스를 공개하였다. 관광 빅데이터 플랫폼은 기업, 지역, 국민이 실질적으로 체감할 수 있는 디지털 관광 생태계를 조성하기 위해 민간과 공공부문에 흩어져 있는 다양한 데이터를 통합하고 분석하여 개방하는 것을 중점으로 제공하는 서비스이다. 국민들이 접근하기 어려웠던 이동통신, 신용카드, 내비게이션 등의 민간 빅데이터를 비롯해 관광통계, 관광실태조사 등 공공 데이터를 일괄 확보한 후 관광 산업의 흐름을 한눈에 보여주는 서비스를 제공하는 것이다. 이러한 현상은 빅데이터에 대한 관심과 수요가 상당히 높아졌음을 반영하는 결과라고 할 수 있다.

최근 관광 업계에서는 관광객을 위한 콘텐츠를 생산하고 마케팅을 하기 위한 근거로 활용하기 위해 빅데이터를 접목하고 있으며, 관광 트렌드를 파



악하고 관광 마케팅과 업계 지원을 위한 과학적 근거로 활용하기 위해서도 반드시 필요한 것이 관광 빅데이터라고 하였다(고선영 · 정근오, 2021). 대한민국의 많은 공공기업과 지자체는 여러 산업 분야의 데이터를 활용하여 유의미한 정보를 추출하려는 가운데, 관광 산업은 그 특성상 데이터 인프라가 자연스럽게 구축되어 있는 상태로 다양한 부문에서 관광 빅데이터에 대한 접근과 활용을 앞당기고 있다(이정미 · 류미나 · 임규건, 2018). 더욱이 인터넷의 활용, 휴대전화 보급의 발달로 이동통신 수단, 신용카드 자료, SNS 보편화 등에 따라 관광에 관련된 정보를 쉽게 구할 수 있어 데이터의 생산과 습득에 대한 속도 역시 높아지고 있다. 관광 산업 분야에서도 이러한 이유로 빅데이터의 중요성을 인지하였으며 관광 빅데이터에 대한 활용방안과 정부정책 모색에 많은 관심을 보이고 있다.

2015년 8월 행정자치부는 정부 3.0 빅데이터 분석을 통해 정책효과를 높이기 위한 계획을 발표하였다. 빅데이터의 활용 및 확산을 위한 노력 강화 차원에서 공공부문 빅데이터 협의회를 통해 각 기관별 과제분석결과를 공유하고 조정함으로써 빅데이터 추진 협력체계를 강화할 것이라는 구체적인 방향을 선보인 것이다. 당시 행정자치부 보도자료에 따르면 빅데이터 선정 분석과제 중 하나는 「빅데이터 분석을 통한 지역관광 활성화」이다. 이는 지역 정보, 유동인구 및 매출자료 등을 융복합하여 축제 · 관광지에 대한 재방문을 제고하는 과제를 도출하고 지역축제 및 관광지 분석 등의 유형화를 통해 지자체별 관광 분야 빅데이터 분석 확대 방안을 마련하겠다는 계획안에 해당한다(행정자치부, 2015). 정리하면 빅데이터를 이용했을 때 실시간으로 정확하고 상세한 정보 수집이 가능하며 미래를 예측하는 효율적인 정책 수립 및 시스템 운영에 활용할 수 있다. 이와 같은 빅데이터의 이점을 통해 앞으로는 산업의 경쟁력을 제고하고 생산성 향상, 및 혁신을 위한 새로운 가치 창출이 가능해질 전망이다(권태일 · 이충희, 2017).

### 3. 텍스트 마이닝의 개념 및 중요성

최근 여러 산업 분야에서 빅데이터를 활용할 때 수치와 같은 정량 데이터는 물론이고 텍스트, 영상과 같은 비정량 데이터를 분석하려는 시도가 늘고 있다(박영욱·정규엽, 2021). 텍스트 마이닝은 구조화되지 않은 대규모의 텍스트 집합으로부터 새로운 지식을 발견하는 과정으로, 문서 내에서 유용한 정보를 찾는 기술로 대량의 정보가 담긴 문서 데이터에서 의미를 도출해내는 일련의 과정을 나타낸다(오익근·이태숙·전채남, 2015). 이러한 시도는 호텔 및 관광 분야에서도 진행되고 있다. 텍스트 마이닝 기법을 활용하면 원하는 정보 추출, 주제 분류, 감성 분석 등이 가능하다는 이점이 있다(박로운·이기훈, 2016; 김종성, 2017, 오승욱·한진욱·김민수, 2019; 임영희·김홍범, 2019; 박영욱·정규엽, 2021). 텍스트 마이닝 기법이 비정형 텍스트 데이터인 온라인 리뷰 데이터에 적용되면 자연어 처리 및 형태소 처리기술을 통해 리뷰에 존재하는 의미 있는 내용을 확인하는 데에 유용하게 사용될 수 있는 것이다. 이로써 리뷰 내용을 구성하는 내용의 전반적인 맥락을 확인함과 동시에 의미 있는 정보를 파악할 수 있다. 실제로 텍스트 마이닝을 활용하여 관광지 이미지 형성에 영향을 미치는 관광지 이미지 요소들을 파악한 연구가 존재한다(서정아·오익근, 2017). 해당 연구에서는 텍스트 마이닝을 통해 수집된 데이터들의 동시 출현 빈도 분석을 하였으며, 온라인에서 관광 목적지에 대하여 언급할 때 많이 사용되는 단어들을 확인하였다. 이를 통해 해당지역 관광에 대하여 언급할 때 가장 많이 동시에 언급되는 장소, 여행형태, 여행동기 등을 파악할 수 있었다. 그러므로 텍스트 마이닝은 분석 기법 그 자체로서 하나의 분석을 나타내는 용어로 지칭될 수 있지만, 텍스트 마이닝이라고 하면 정제되지 않은 텍스트 데이터 안에서 의미 있는 내용을 발굴하는 일련의 과정을 나타낸다고 할 수 있다. 정리하

면, 빅데이터 분야에서 다량의 텍스트형 리뷰 데이터를 수집하는 기술로 텍스트 마이닝이 널리 사용되고 있다. 텍스트 마이닝은 다양한 문서로부터 비정형 또는 반정형 텍스트 데이터를 획득하여 자연어 처리기술을 기반으로 패턴이나 관계를 도출하므로 새로운 가치를 창출하고 의미 있는 정보를 파악하는 데에 도움을 얻을 수 있으므로 중요하다.

#### 4. 텍스트 데이터의 활용

PWC(2012)는 텍스트 데이터를 활용하기 위한 방안으로 데이터 수집 단계에서 데이터 마이닝(Data Mining), 텍스트 마이닝(Text Mining), 오피니언 마이닝(Opinion Mining)이 널리 사용되고 있으며, 이에 대한 분석 방법으로 소셜 네트워크 분석(Social Network Analysis), 군집 분석(Cluster Analysis) 등이 활용되고 있다고 하였다. 텍스트 데이터에 대한 활용은 이 외에도 다양하게 존재하나, PWC(2012)가 언급하였던 주요 데이터 수집 및 분석 방법에 대한 내용을 요약해서 정리하면 다음과 같다.

첫 번째, 데이터 마이닝은 대용량의 데이터와 데이터베이스 등에서 감추어진 지식 및 새로운 규칙 등의 유용한 정보를 발견하는 기술이다. 통계 및 수학적 기술뿐만 아니라 패턴인식 기술들을 이용하여 데이터 저장소에 저장된 대용량의 데이터를 조사 및 분석함으로써 의미 있는 새로운 상관관계, 패턴, 트렌드 등을 발견하는 데 활용될 수 있다(윤홍근, 2013). 즉, 데이터 마이닝을 통해 정보의 연관성을 파악함으로써 가치 있는 정보를 만들며 의사결정에 적용하기 위해 활용되는 것이다.

두 번째, 텍스트 마이닝은 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치와 의미 있는 정보를 찾아내는 마이닝 기법이다. 이때 텍스트 마이닝은 사람들의 말하는 언어를 이해할 수 있는 자

언어 처리기법을 토대로 한다.

세 번째, 오피니언 마이닝은 평판 분석이라고도 불리는 데이터 수집 방법이다. 주로 웹사이트와 소셜 미디어에 나타난 여론과 의견을 분석하여 실질적으로 유용한 정보로 재가공하는 기술로 텍스트를 분석하여 사용자의 감성과 의견을 통계적인 수치로 변환하여 객관적 정보로 바꿀 수 있다(윤홍근, 2013).

네 번째, 소셜 네트워크 분석은 소셜 미디어에 공유되는 글과 사용자를 분석해 소비자의 흐름이나 패턴을 분석하고, 판매나 홍보에 적용하는 기법이다. 소셜 네트워크 분석은 수학의 그래프 이론에 뿌리를 두고 있으며, 소셜 네트워크 연결구조 및 연결 강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하여, 소셜 네트워크 상에서 입소문의 중심이나 허브 역할을 하는 사용자를 찾는데 주로 활용된다(신동일, 2014). 이는 마케팅 분야뿐만 아니라 사회의 흐름과 트렌드, 여론 변화의 추이를 읽어내기 위해서 사용된다.

다섯 번째, 군집 분석은 각 대상의 유사성을 측정하여 유사성이 높은 대상 집단을 분류하고, 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체 간의 상이성을 규명하는 통계기법으로 명확한 기준이 존재하지 않거나 밝혀지지 않은 상태에서 다양한 특성을 지닌 대상을 집단으로 분류할 때 사용된다(윤홍근, 2013). 군집분석은 속성이 비슷한 잠재 고객들끼리 그룹화하여 시장을 세분화하는 방법에 활용되며, 나이, 학력, 소득, 제품, 매장 등 유사한 대상끼리 집단으로 묶고자 하는 경우에도 사용할 수 있다(신동일, 2014). 앞서 제시된 텍스트 데이터 수집 및 분석 방법 이외에도 관련 선행연구를 고찰한 결과, 텍스트 데이터는 의미연결망 분석, 감성 분석, 시계열 분석 등의 방법과 접목되어 활용되는 것을 확인할 수 있었다. 텍스트 데이터와 함께 자주 활용되는 의미연결망 분석, 감성 분석, 시계열 분석에 대한 내용을 고찰한 내용은 다음과 같다.

## 1) 의미연결망 분석

의미연결망 분석은 사회의 시스템 구조를 파악할 때 행위자 사이의 관계를 중시하는 사회연결망 분석을 커뮤니케이션 메시지에 적용한 방법이다(김학선, 2017). 의미연결망 분석에서는 특정 유목을 가정하지 않은 상태에서 단어들의 빈도와 한 문장 안에서 동시에 사용되는 단어들의 관계를 통해 메시지 또는 문장에 담겨있는 의미와 패턴을 분석할 수 있다(Shi, Zhu, Yang, & Li, 2016). 이때, 정보의 단위가 되는 단어나 구를 각각의 노드 형성의 개념으로 보고 개념 간의 연결 상태를 링크(link)로 나타내는데 여기서 링크로 드러나는 단어들의 공동 출현 관계를 통해 의미를 해석한다. 온라인 리뷰 데이터를 이용해 의미연결망 분석을 수행한 연구는 다음과 같다.

관광 분야에서는 해양관광에 대한 관광객들의 전반적인 인식 분석을 하고 활성화 방안을 도출하기 위한 목적으로 다음, 네이버, 구글의 커뮤니티 및 SNS 사이트인 카페, 블로그, 페이스북을 대상으로 핵심어에 해당하는 해양관광과 함께 도출되는 동반 핵심어를 수집하여 의미연결망 분석을 수행하였다(박태수, 2020). 해당 연구에서는 주요 소셜 미디어를 통해 추출된 해양관광 핵심어에는 어떠한 특징이 있으며, 중심성 및 군집의 형성을 탐구하고 분석된 결과를 통해 해양관광에 대한 활성화 방안을 연구하였다.

의미연결망 분석은 관광 분야에서 관광객의 인식 변화를 알아보기 위해 사용되기도 한다. 사드 배치 전후 중국 관광객의 한국관광 인식 변화 비교를 위해 빅데이터를 활용하고 의미연결망 분석을 진행한 연구가 이에 해당한다(박경열 · 한혜림 · 최승담, 2019). 해당 연구에서는 사드 배치 전후 중국 관광객의 한국관광 인식 변화를 비교하는 것에 초점을 두었다. 이를 위해 2016년 7월을 기준 시점으로 사드 배치 전후 ‘한국명소’, ‘한국맛집’ 및 ‘한국여행’ 키워드와 연관된 동시 출현 단어를 수집 및 분석하고 의미연결망

분석을 한 것으로 확인된다. 이를 통해 한 국가의 관광 목적지로서 관광인식은 관광정보 등에 국한되지 않으며 정치적·사회적·환경적 요인과의 밀접한 관련성이 있으므로 정책 수립이나 실무 전략 수립 시에 관광에 대한 보다 광의적으로 확대된 시각과 접근이 필요하다는 시사점을 제시하였다.

여수관광에 대한 관광객들의 전반적인 인식과 이를 통한 관광 활성화 방안을 도출하기 위해 의미연결망 분석을 시도한 연구도 있다(한장현·서현, 2019). 해당 연구에서는 의미연결망 분석을 사용하면 연구를 진행하고자 하는 테마를 구성하고 있는 핵심어를 가지고 핵심어 사이에 구조적 배열상태 등을 통해 관광객들이 인식하는 사고를 분석할 수 있으며, 이해하기 쉽도록 핵심어들의 관계를 시각화하여 제시할 수 있는 점을 활용하였다(한장현·조윤희, 2018). 이를 위해 주요 소셜미디어를 통하여 수집하고 분석된 관광객들의 여수 관광 핵심어에 담긴 특징과 이에 따른 중심성 및 군집을 살펴보고 여수 관광 활성화에 대한 방안을 논의하였다.

음식관광을 조사하기 위해 온라인 리뷰를 활용한 연구로는 구글, 페이스북, 트위터에서 ‘음식관광’을 키워드로 포함하는 데이터를 수집하고 음식관광에 대한 다양한 정서 및 의견 분석을 분석한 연구가 확인된다(김태호·김학선, 2017). 수집된 데이터를 통해 핵심어를 도출하고, 데이터 탐색을 진행하여 데이터 수정 및 변환과정을 거친 후 텍스트 마이닝 분석 방법중 하나인 의미연결망 분석을 진행하였다. 수집된 데이터를 통해 음식관광에 대한 핵심어를 도출하였으며, 각종 포털을 활용한 빅데이터를 관광레저연구에 어떻게 활용할 것인지에 관한 방안 제시를 연구의 목적으로 하였다.

최근 1인 가구의 증가 및 COVID-19로 인한 비대면 소비 증가로 식품유통 플랫폼에 대한 이용이 급증하고 있는 연구의 배경을 토대로 5년 간의 빅데이터 자료를 이용해 식품 유통 플랫폼에 대한 핵심 키워드를 도출하고 수집된 키워드 간의 연결성 특징을 밝힌 연구가 있다(강지원·남궁영,

2021). 해당 연구에서는 비구조화된 텍스트로부터 정보를 추출하기 위한 방법으로 텍스트를 사용하였으며, 의미연결망을 바탕으로 유사한 특성을 지니는 단어들의 군집을 도출하였다. 이는 전통적인 연구 조사 방법만으로 파악하기 어려웠던 소비자의 인식을 일상생활에서 자연스럽게 누적되어 온 빅데이터를 활용하여 객관적으로 분석하여 연구를 진행한 것에 해당한다.

포털 사이트인 네이버 영화 페이지의 리뷰 데이터를 수집하여 출현 빈도가 높은 단어를 중심으로 영화 관람객의 반응을 시각화하는 작업을 수행한 연구가 있다(김슬기·김장현, 2019). 해당 연구에 따르면 의미연결망 분석을 위해 총 여섯 편의 영화를 선정하여 데이터를 수집하고 정제하는 과정을 거쳤으며, 의미연결망 분석을 통해 단어 간 관계성을 파악하였다. 구체적으로는 영화 관람객이 작성한 리뷰 데이터에 나타나는 키워드를 중심으로 소비자들의 반응을 한 눈에 확인할 수 있는 리뷰 인터페이스 구현을 탐색하였으며, 영화 리뷰를 구성하는 주요 키워드를 시각화하고 리뷰 내용에 나타나는 영화별 특성의 차이를 확인하였다.

## 2) 감성 분석

온라인 리뷰를 이용한 감성 분석은 여러 연구자를 통해 진행되어왔다. 감성 분석은 오피니언 마이닝이라고도 불리며, 온라인 리뷰의 텍스트와 같은 비정형 데이터에서 원하는 정보를 대상으로 자연어 처리를 하고, 감성사전을 활용하여 텍스트 내의 긍정과 부정을 식별하는 분석 방법이다(최자영·김현아·김용범, 2020). 감성 분석을 통해 단어를 긍정적인 단어와 부정적인 단어로 구분할 수 있으며, 온라인 리뷰를 통해 데이터를 수집하고 감성 분석을 수행한 연구들은 다음과 같다.

최근 연구로는 트립어드바이저에서 서울 관광지 10개소를 대상으로 총

29,467개에 달하는 온라인 리뷰 및 평점 등의 데이터를 수집하고 감성 분석을 통해 리뷰의 감성 값을 활용하여 감성의 강도 차이를 반영한 후, 관광수요 예측모형을 구축하기 위해 ARIMA, 인공지능 기법, 딥러닝 기법을 적용하여 관광객 수만으로 구축한 예측모형과 온라인 리뷰 정보가 반영된 예측모형의 성과를 비교한 연구가 있다(김은미, 2021).

이 외에도 제주 지역의 에어비엔비 숙소를 이용한 사용자들의 리뷰를 크롤링하여 단어 코퍼스를 구축한 후, 리뷰 문서에 나타난 긍정 또는 부정의 오피니언을 자동 분류하는 오피니언 마이닝 및 감성 분석 연구를 수행한 연구가 있다(유광훈·남지순, 2017). 해당 연구에 따르면 제주 지역의 에어비엔비 숙소 후기글을 11개 구역별로 나누어 후기글의 감성 분석을 수행하였으며, 이후 해당 문서를 중심으로 고빈도로 나타나는 주요 연관어들의 의미 자질을 분석하고 이들 사이의 연결망을 구축하는 연구를 진행하였다. 이 연구에서는 한국어 감성사전 DECO에 수록된 감성 어휘 정보와 LGG 그래프 문법으로 기술된 두 단어 이상의 구 단위 감성 표현 정보를 이용하여 극성 값을 계산하는 감성 분석을 수행함으로써, 기존의 단일 단어 기반의 감성 분석이 가지는 한계를 극복하는 연구 방법론을 제안하였다.

트립어드바이저를 통하여 데이터를 수집하고 텍스트 마이닝 기법을 이용하여 인바운드 관광객들의 특성을 분석한 연구도 확인된다(김민식·한학진·박병화, 2017). 총 19,699개의 데이터를 수집하여 지역별로 분류한 데이터를 음식, 쇼핑, 호텔이라는 키워드로 나누어 인바운드 관광객들이 관심가지는 요소를 파악하였다. 이때 인바운드 관광객들의 지역별 키워드에 대한 긍정과 부정을 파악하기 위해 감성 분석을 진행하였던 것으로 확인된다.

호텔과 관련하여서는 국내 5성급 호텔 고객들을 대상으로 온라인 리뷰 빅데이터를 통해 DMR(Dirichlet Multinomial Regression) 토픽모델링 분석을 하고, 고객들의 의견을 주요 주제별로 나누는 감성 분석을 시도한 연구가



확인된다(박영욱·정규엽, 2021). 해당 연구에 따르면 국내 5성급 호텔을 이용한 외국인 방문객들의 온라인 리뷰로부터 점수에 영향을 미치는 호텔 속성을 파악하고자 하였으며, 호텔 속성과 고객 감성 간의 관계를 호텔 리뷰와 리뷰 점수를 기반으로 확인하였다.

호텔 관련 방한 외래 관광객의 온라인 리뷰를 활용한 감성 분석 연구도 확인된다(임영희·김홍범, 2019). 해당 연구에 따르면 트립어드바이저를 통해 비즈니스 호텔, 리조트 호텔, 공항 호텔 각 5개소에서 리뷰 데이터를 수집하고 호텔에 대한 방문객의 감성 분석을 진행하였다. 분석 결과에 따르면 온라인 리뷰의 감성 극성값은 고객 평점에 긍정적인 영향을 미치는 것으로 나타났다. 빈도 분석을 통해 호텔 유형별 고객에게 중요한 단어를 추출하였으며, 감성 분석을 통해 만족과 불만족 문장에 포함된 단어들을 토대로 호텔의 품질 속성을 새롭게 확인하였다.

대부분 국내에서 진행된 감성 분석 연구는 영문 리뷰를 대상으로 하며 국문 리뷰에 대한 연구가 부족하므로, 이에 따라 국내 호텔 서비스 속성의 고객 만족도 분석방안으로 국문 온라인 리뷰에 대한 감성 분석을 제안한 연구도 확인된다(곽민정·최지유·박소현, 2019). 해당 연구에 따르면 국내 호텔을 대상으로 국문 리뷰 감성 분석 모델을 개발하였으며 서울 지역 5성급 호텔 18개소를 온라인 리뷰의 감성 분석 대상으로 하였다. 제안된 감성 분석 모델은 호텔 리뷰를 서비스 속성 수준에서 분석하였으며 각 속성에 대한 고객의 감성이 긍정인지, 혹은 부정인지를 판별하기 위해 사용되었다. 이를 통해 호텔 경영자는 고객이 호텔 서비스의 어떤 면에서 얼마나 만족하고 불만족하였는지 파악할 수 있었고, 고객의 만족 및 불만족 원인에 따른 분석을 통해 서비스 품질 향상 추진을 위한 시사점을 제시하였다.

### 3) 시계열 분석

다량으로 쌓이는 온라인 리뷰 데이터를 이용하면 특정 시기별로 구분하여 시대별 트렌드를 살펴보는 시계열 연구로 확장될 수 있다. 실제로 텍스트 마이닝 기법을 사용하여 1990년부터 2019년까지 5,141건의 신문기사에서 허위·과장광고 용어의 트렌드를 분석한 연구가 확인된다(김도희·김민정, 2021). 해당 연구에서는 전체 신문 기사를 대상으로 빈도 분석을 하여 허위·과장광고의 최빈 키워드와 추출된 키워드 간의 맥락을 확인하였고, 변화 추세를 확인하기 위해 10년 단위로 기사를 분리하여 빈도 분석을 수행하였다. 나아가 연도별 최빈 키워드를 주제로 하는 학술논문 수와 비교하였을 때 해당 시기에 이슈가 되었던 키워드가 연구로 이어지는 경향성을 파악하였다.

이 외에도 온라인 리뷰 데이터는 2000년 이후 한국 사회의 다문화 관련 논의가 어떻게 진행되어 왔는가에 대한 실증 고찰을 위해서도 활용되었다. ‘다문화’, ‘이민자’, ‘이주민’, ‘이주배경’을 키워드로 하여 지난 17년 간의 기사문 약 8만 5천 건을 수집하여 텍스트 마이닝을 하고 토픽모델링 분석 및 토픽의 시기별 비중 변화에 대한 시계열적 변화를 관찰한 연구가 이에 해당한다(양정은·노법래, 2020). 온라인 리뷰 데이터를 시계열 분석에 활용함으로써 대·내외적으로 급변하는 환경 속에서 한국 다문화 정책의 지향성에 대한 고민과 필요성을 제기하였다.

빅데이터를 활용한 도시공원 이용의 행태적 특성을 분석한 연구도 존재한다(우경숙·서주환, 2020). 해당 연구는 온라인 비정형 데이터인 블로그 데이터를 기초 자료로 하여 특정 공원과 관련된 텍스트 데이터를 수집하였으며 공원 이용행태와 이용행태 간의 네트워크 구조를 분석하였다. 이때 블로그의 데이터를 수집할 수 있는 최초 년도인 2000년부터 데이터를 수집하였으며, 분석의 효율성을 위해 5년 단위로 구간을 나누어(2000년-2004년,

2005년-2009년, 2010년-2014년, 2015-2018년) 시계열 분석에 활용하였다. 즉, 시간의 흐름에 따라 공원 이용행태는 어떻게 변화하고 있는가에 대하여 내용 행태 간의 연계성 구조를 파악할 수 있었다.

온라인 리뷰를 수집하는 방식과 유사하게 뉴스 빅데이터를 통하여 수행된 시계열 연구도 확인된다. 장기 시계열 내용 분석을 위해 자연어 처리와 의미연결망 분석이 결합된 뉴스 빅데이터 분석의 활용을 제안한 연구가 이에 해당한다(박대민, 2016). 해당 연구에 따르면 시계열 분석을 위해 26년치에 해당하는 8개 중앙지의 정치와 사회면 기사 약 100만 건의 데이터를 수집하였다. 이렇게 수집된 기사는 자연어 처리 후 기사의 정보원과 인용문 주제를 중심으로 의미연결망 분석을 하였으며, 매체별로 정보원과 주제의 시계열적 변화를 살펴보는 데 사용하였다. 이렇게 하여 2000년 전후로 가장 중요한 정보원과 인용문 주제가 시계열적으로 전면 변화하는 양상을 보였다는 결과를 발견하였다. 즉, 자동화된 장기 시계열 내용 분석 데이터를 축적하는 것에 대한 제안뿐만 아니라, 이를 바탕으로 경제 지표 등의 다양한 시계열 데이터와 정교한 비교연구를 할 수 있는 방법을 제시하였다.

#### 4) 텍스트 기반 온라인 리뷰의 다양한 활용

이상의 선행연구 고찰을 통해 텍스트 데이터 중에서도 온라인 리뷰가 실제 연구에서 어떻게 분석되고 활용되었는지 살펴보았다. 정리하면 온라인 리뷰를 텍스트 마이닝하여 크게 의미연결망 분석, 감성 분석, 시계열 분석에 활용되는 것으로 나타났으나 본 연구에서 제시한 분석 방법 이외에도 다양한 방법으로 온라인 리뷰가 활용되는 것을 확인할 수 있었다. 한 예로 소셜 네트워크에 나타난 한류에 대한 인식과 방한 의도를 살펴보기 위해 한류 영상물이 주로 업로드 되는 온라인 매체인 유튜브를 통하여 한류 영

상을 시청한 리뷰를 수집하고 사회연결망 분석을 진행한 연구가 있다(김성민 · 이훈, 2020). 해당 연구에서는 연결 중심성, 근접 중심성, 그리고 매개 중심성을 살펴보고 군집분석을 하였다. 이 외에도 LDA 토픽모델링을 적용하여 DMZ관광 소셜미디어 데이터 분석을 통해 핵심 토픽을 도출하고, 핵심 토픽과 토픽을 구성하는 키워드를 통해 현상학적으로 담론과 함의를 제시한 연구가 있으며(박경열 · 안희자, 2019), 온라인 여행 커뮤니티의 여행객 리뷰 데이터를 바탕으로 연관규칙 마이닝 기법을 활용하여 관광지 이미지의 구성개념인 인지적 이미지와 정서적 이미지 간의 연관관계를 파악하여 여행객 유형별 관광지 이미지 형성과 평가에 대한 패턴을 파악한 연구가 있다(김홍범 · 심영석, 2018).

## 제 4 절 LDA 토픽모델링

### 1. LDA 토픽모델링의 개념 및 중요성

토픽모델링은 데이터를 구성하고 있는 단어들의 관계에서 토픽, 즉 주제를 추출해내는 텍스트 분석 기법이다. LDA(Latent Dirichlet Allocation)는 토픽을 추출할 때 잠재적 디리클레 확률 분포를 사용하기 때문에 붙여진 이름이다. LDA는 토픽모델링 방법의 하나로 Blei(2012)가 제안하였다. 토픽모델링은 문서(document) 내 단어(corpus)를 통해 각 문서에 잠재된 주제를 통계적으로 파악하는 기법이다. 토픽모델링 방법으로는 LDA 방식 이외에도 LSA(Latent Semantic Analysis), PLSA(Probabilistic Latent Semantic Analysis) 등이 존재한다. 동시에 출현하는 단어들의 유사성을 바탕으로 의미를 추출하는 LSA 기법과 확률을 통해 계산되어 처리시간이 개선된 PLSA 기법은 학습 데이터에 과적합<sup>4)</sup> 되는 문제를 가지고 있어, 이를 해결하기 위해 개발된 기법이 LDA 토픽모델링이다(박예림 외, 2019).

LDA 토픽모델링은 하나의 문서에 복수의 토픽이 내재 되어있는 것을 전제로 하여 문서 내의 여러 단어들의 조합을 토대로 문서의 토픽들을 확률적으로 파악하는 통계적 분석 방법이다(Blei, Ng, & Jordan, 2003). 토픽모델링은 단순히 단어의 출현 빈도수를 기반으로 하는 텍스트 분석과는 다르며, ‘토픽’이라는 변수를 통한 확률 기반의 분석을 함으로써 과적합 문제가 적으며, 새로운 데이터가 투입되어도 분석 가능하다는 장점이 있다(박영욱 · 정규엽, 2021). LDA 토픽모델링을 통해 각 문서를 구성하는 토픽과 토픽

---

4) 과적합: 기계 학습에서 데이터에 대한 학습이 너무 많이 수행되는 현상으로, 학습 대상 데이터에 대한 오차는 감소하지만, 실제 사례에 적용할 경우 오차가 증가하는 문제를 수반함

의 키워드를 도출할 수 있으며, 키워드 분석을 통해 토픽을 분류하는 과정을 거친다.

정리하면 토픽모델링은 특정 문서의 주요 토픽이 무엇인지 확률적으로 분석하기 위해 사용되며, 비정형화된 방대한 텍스트를 분석할 때 활용 가능한 분석 기법이다. 토픽모델링 관점에서 문서는 다양한 단어들의 결합체로 구성되어 있으며, 단어들 간의 특정 관계를 군집화하면 다수의 토픽으로 유형화되고, 여러 단어들이 모여 생성된 토픽들은 하나의 문서로 구성되는 계층적 구조로 되어 있다(Kim & Baek, 2016). LDA 토픽모델링은 데이터의 단어 집단에서 특정 단어가 등장할 확률을 계산한다. 만일, 특정 리뷰에서 특정 주제에 대한 내용이 나타나면, 해당 리뷰에 사용된 단어들은 리뷰의 주제를 표현하기 위해 사용되었다고 볼 수 있는 것이다. 따라서 어떤 문서가 존재할 때, 그 문서 안에서 어떤 토픽으로 내용이 형성되고 있는지 알아보기 위해서 토픽모델링을 실행하고 리뷰 별로 자주 등장할 확률이 높은 단어들을 확인하면 해당 리뷰의 전반적인 주제를 유추할 수 있다. LDA 토픽모델링은 단순하며 의미적으로 일관성 있는 주제들을 생산하는 장점을 가지고 있으며, 텍스트 데이터를 처리하기 위한 분석 방법 중에서 활용도가 높아 인기 있는 모델로 사용되고 있다(이동혁 · 이서희, 2020).

## 2. LDA 토픽모델링의 활용

LDA 토픽모델링은 방대한 텍스트에 존재하는 사용자의 경험, 느낌이나 생각 등을 토픽 단위로 도출하기 위한 연구에 활용될 수 있다. LDA 토픽모델링은 군집 분석과는 다르게 리뷰에 존재하는 잠재적인 주제를 파악하기 위한 분석 방법이다. 문서 군집화는 문서 내 단어의 유사성을 바탕으로 문서의 집단을 나누려고 할 때 사용하는 반면, 토픽모델링은 문서 내 단어에

따라 문서 군의 주제들을 파악하고자 할 때 활용되는 것이다(Blei et al., 2003). 토픽모델링은 통계적으로 문서 내 하위 주제들을 자동으로 생성하는 특징이 있다. 이를 관광 분야에 적용하면, 온라인 리뷰에 존재하는 관광 이미지 도출, 관광객들이 관심 가지는 주요 토픽 도출, 관광지 선택속성 등을 확인하기 위한 목적으로 활용될 수 있다. 이를테면, 관광지마다 관광객들이 활발하게 언급하는 토픽에는 무엇이 있는지, 점차 관심도가 사라지는 토픽에는 무엇이 있는지 관찰할 수 있으며, 시기별 토픽의 변화 추세를 확인하는 연구에도 사용될 수 있는 것이다. 또한 토픽모델링을 할 때 가중치를 적용하면 보다 정교한 토픽을 도출할 수 있으며, 추출하고자 하는 토픽의 개수를 지정하고 샘플링을 반복 수행할 횟수 등을 지정함으로써 연구 목적에 따라 다양한 LDA 토픽모델링을 구현할 수 있다.

### 3. LDA 토픽모델링의 선행연구 동향

트립어드바이저에서 영어로 작성된 서울의 103개 장소에 대하여 21,620 개의 댓글과 평점을 수집하고 장소를 세 개의 유형으로 구분하여 BoW (Bag of Words) 방식으로 문서 집합 코퍼스를 구축하고 LDA 토픽모델링을 하여 관광지 유형별 만족도 요인을 도출한 연구가 있다(김보섭 외, 2016). 이 연구에서는 토픽모델링에 대한 결과가 관광객을 대상으로 하는 설문조사에서 문항 구성을 위한 예비조사로 활용할 수 있다고 언급하였다. 예를 들면, 관광지별로 중요한 요인을 토픽모델링으로 도출한 뒤 해당 요인을 좀 더 세분화하여 설문조사 문항 구성을 위한 작업에 활용할 수 있는 것이다.

사물 인터넷을 기반으로 하는 제품에 대한 사용자 경험을 살펴보기 위해 온라인 리뷰에 텍스트 마이닝 기법을 적용하여 토픽모델링을 시도한 연구도 확인된다(황해정 · 심혜린 · 최준호, 2016). 이 연구에서는 아마존에서

출시한 에코를 사례로 분석하였으며, 총 세 개의 하위 카테고리에 대하여 다섯 개 토픽을 도출하였다. 이렇게 도출된 LDA 토픽모델링 결과를 바탕으로 각 토픽 속성이 만족도에 미치는 영향을 검증하는 분석에 활용하였다.

관광학 분야 연구 동향을 규명하기 위해 관광학 분야 주요 학술지를 대상으로 2008년부터 2016년까지 발표된 논문 초록 1,239개를 수집하여 토픽모델링과 시계열 분석 기법을 적용하여 분석한 연구도 확인된다(김창식·곽기영·윤혜진, 2017). 이 연구에서는 관광학 분야 연구 동향을 파악하기 위해 20개의 토픽을 도출하였으며, 주요 연구 토픽의 연도별 추이를 분석하고 최근 활발하게 연구되는 토픽과 점차 연구되지 않는 토픽을 규명하였다. 해당 연구에서는 다양한 토픽모델링 알고리즘 중에서도 LSA 기법을 적용하였으며, 토픽모델링 기법과 함께 시계열 회귀분석을 적용하여 리서치 동향 분석에 활용한 것이 특징이다.

한편, LDA 토픽모델링은 사용자의 인식을 파악하기 위한 목적으로 연구에 활용되는 것을 확인하였다. 한 예로, 통일관광을 위한 비무장지대(DMZ)의 관광학적 가치에 대한 인식 변화를 확인한 연구가 이에 해당한다(노희경, 2019). 이 연구에서는 약 10년간 ‘비무장지대’ 및 ‘DMZ’와 관련된 소셜 미디어 데이터를 활용하여 토픽모델링을 하였으며, 핵심 토픽을 파악하고 DMZ와 관련된 인식, 의견 및 관심의 변화를 파악하였다. 또한, 토픽모델링 결과를 활용하여 DMZ의 관광학적 가치에 대한 인식의 변화를 현상학적으로 분석하였으며, DMZ를 주요거점으로 한반도 통일에 있어 관광의 가치와 역할에 대한 정책적 함의 제시를 목적으로 하였다.

LDA 토픽모델링은 방문 관광객의 주요 관광 요소 및 관심사를 도출하기 위해 사용된다. 소셜 네트워크 데이터인 플리커를 활용하여 서울시내 11개의 주요 관광지를 도출하고 텍스트 마이닝 후 동시출현 네트워크, 토픽모델링 기법을 사용하여 주요 토픽을 선정한 연구가 이에 해당한다(박예림



외, 2019). 해당 연구에서는 TF-IDF 가중치를 적용하여 12개의 주요 토픽을 도출하였으며, 관광지마다 주로 언급되는 단어와 토픽을 통해 관광지 이미지를 확인하였다.

관광지 특성에 따른 관광객의 인식 차이를 파악하고자, 토픽모델링을 활용하여 계량적인 분석 방법을 통해 직접적인 관광객 인식 기반의 정책적 시사점을 제안하였던 연구도 확인된다(남승주·김준환·유영준, 2020). 해당 연구에서는 관광객의 인식이 보다 직접적으로 내포되어 있는 온라인 리뷰를 기반으로 방한 관광객의 인식을 도출하기 위해 LDA 토픽모델링을 하였다. 그 결과로 관광객이 가장 중요하게 인식하는 속성과 중요도를 확인하였으며, 문화권 별로 관광객들의 인식 차이를 살펴보았다. 이 외에도, 모바일 어플리케이션에 대한 사용자들의 평가 척도를 개선하기 위하여 LDA 토픽모델링을 활용한 연구(이범국·손창호, 2020), 토픽모델링에 대한 결과로 국내 음식관광시장이 다섯 개의 세분 시장(토픽)으로 구분될 수 있음을 확인한 연구(이범준·노희경, 2020) 등이 있다.

종합하면, LDA 토픽모델링은 텍스트 마이닝에 속하는 하나의 분석 기법으로 리뷰나 댓글 등의 텍스트 데이터에 존재하는 주요 내용의 주제를 도출하기 위한 목적으로 연구에 활용되는 것을 확인하였다. LDA 토픽모델링은 단독으로 수행될 수 있지만, 연구 목적에 따라 다른 분석 방법과 접목하여 활용 가능한 것을 확인하였다.

## 제 5 절 FsQCA(퍼지셋 질적비교분석)

### 1. FsQCA의 개념 및 중요성

사회과학은 매우 복잡한 인과관계로 이루어져 있으므로 단일화된 초점으로 하나의 현상을 바라보는 것은 적절하지 않다(한영위 · 이용기 · 안성만, 2017). 사회과학 분야에서 변수 간의 관계는 특정 상황에서 동일한 원인으로 인해 서로 다른 효과가 발생할 수 있다. 이와 유사하게 동일한 결과가 발생하더라도 이에 대한 원인은 서로 다르게 나타날 수 있음을 의미한다. 따라서 사회과학 현상을 다양한 관점에서 바라보기 위해서는 다양한 접근법이 요구된다. 이에, 집합관계를 기본 개념으로 하는 질적비교분석(QCA : Qualitative Comparative Analysis, 이하 QCA)이 하나의 대안이 될 수 있다.

QCA는 어떠한 사회 현상이 발생했을 때, 이를 계량적 방법론에 기초하여 상관관계 관점에서 살펴보는 것이 아닌, 집합과 집합 간의 관계로 현상을 이해하는 것이 더욱 타당하다고 여기면서 시작된 방법론이다. Ragin(2014)은 사례연구와 변수 지향적 연구 방법의 여러 가지 문제점을 제시하면서 QCA를 제안하였다. QCA는 사례중심 접근방식과 변수중심 접근방식을 통합하는 방식이다. 즉, 양적 연구와 질적 연구의 상호보완적 개념으로 활용된다. QCA의 특징은 기존 연구 방법과 다르게 개별적인 속성 또는 변수가 중심이 되지 않고 개별적인 속성 또는 변수의 집합이 중심이 되어 여러 사례들 간의 다양한 범주화가 가능하며, 같은 결과일지라도 해당 결과를 유발하는 원인에는 다양한 인과관계가 도출될 수 있다는 것에 초점을 둔다. 따라서 긍정적 결과와 부정적 결과를 유발하는 원인조건을 탐색할 수 있으며, 부적합한 사례(contrarian case)의 조건을 함께 파악할 수 있다(Woodside, 2014).

Grofman & Schneider(2009)에 따르면 QCA는 다음의 세 가지 특성을 지닌다. 첫 번째 특성은 ‘등결과성(equifinality)’이다. 이는 여러 요인들의 다양한 조합으로 인해 동일한 결과가 발생할 수 있다는 것을 의미한다. 즉, 하나의 경로를 통해서만 특정 결과가 도출되는 것이 아니라 전혀 다른 요인의 결합된 형태를 통해서도 같은 결과가 발생할 수 있다(최영준, 2009). 그러므로 현상에 대한 결과만 보는 것이 아닌, 결과를 결정하는 다양한 원인조건을 고려해야 한다는 것을 알 수 있다. 두 번째 특성은 조건들의 ‘논리적 결합(logical conjunctions of conditions)’이다. 이는 하나의 조건이 독립적으로 결과에 영향을 미치는 것이 아니라 다른 조건과의 결합과 상호작용을 통하여 결과를 도출한다는 것을 나타낸다. 즉, 하나의 원인조건이 독립적으로는 결과에 대한 충분조건이 되지 못하더라도, 다른 원인조건을 만났을 때 그 둘의 교집합 된 결과가 충분조건이 될 수 있음을 의미한다. 이러한 논리적 결합에 의하면 어떤 조건과 결합 되는가에 따라서 다른 결과를 생산할 수 있다. 세 번째 특성은 ‘비대칭적 인과성(asymmetric causality)’이다. 이는 특정 독립변수가 결과변수에 있어 유의미한 영향을 미친다고 할 때, 그 결과가 나타나지 않았다는 사실은 곧 해당 독립변수가 존재하지 않았기 때문이라고 주장하는 대칭적 인과관계를 가정하는 양적 연구와는 달리, 어떠한 원인조건이 결과변수에 영향을 미쳤다고 할지라도, 이러한 결과변수가 나타나지 않는 것이 해당 원인조건이 존재하지 않기 때문이라고 단언할 수 없는 것에 있다(한승현 · 강민아, 2015). 이러한 이유로 현상 또는 결과변수가 발생한 것과 발생하지 않은 것에는 별개의 설명이 필요하며 충분조건과 필요조건을 구분해서 살펴볼 필요가 있다.

FsQCA는 양적 연구와 질적 연구가 가지고 있는 각각의 한계를 보완하고 장점을 발전시킨 QCA 기반의 연구 방법이다. 퍼지셋 이론은 Zadeh, Klir, & Yuan(1996)에 의해 제안되었으며, Ragin(2000)과 Kvist(1999)에 의해

구체적으로 발전되었다. 퍼지셋은 논리·집합 대수인 불리언 대수(Boolean algebra)를 통해 기본적인 개념을 이해할 수 있으며, 특히 0과 1을 이용한 퍼지셋의 기본 개념은 사회과학의 비교사례연구 중에서도 질적 비교분석에 유용하다고 할 수 있다(조경훈·박형준, 2015a). 다수의 사례에 대한 질적 변수의 존재 여부를 설명하는 불리언 대수는 이진수 자료를 활용한 진리표에 대하여 덧셈·곱셈 등의 조합으로 유형화를 형성하며, 이는 결과에 대한 원인변수의 충분조건을 제시한다는 점에서 유용하고, 여기에 0과 1만 존재하는 것이 아닌 중간값을 결정할 수 있는 퍼지셋의 방법을 적용한 FsQCA 분석은 불리언 대수만을 활용한 이분법적 변수를 극복할 수 있게 함으로써 정보의 손실을 최소화하여 질적인 정보를 양적으로 전환할 수 있도록 한다. 따라서 각각의 변수가 결과에 독립적인 영향을 끼친다고 가정하는 계량적인 분석 방법과는 다르게 FsQCA 분석을 통해서 결과에 영향을 미치는 원인조건들의 결합관계를 고려하며, 이러한 다양한 결합조건을 밝힐 수 있는 것이 장점이다(이승윤, 2014).

## 2. FsQCA의 활용

최근 국내에서는 FsQCA 방법론을 적용하는 연구가 증가하는 추세이다(김지현, 2020). 결과 집합과 원인 집합이 서로 어느 정도 속하는지에 따라 원인집합이 충분조건인지 또는 필요조건인지 판단할 수 있으며, 이를 통해 인과관계 여부를 파악할 수 있는 것이 FsQCA의 장점이라고 할 수 있다. 전통적 집합을 활용한 QCA와 달리 FsQCA는 원인조건의 조합이 결과변수의 충분조건이 되는지 여부를 판단할 수 있으므로 유용한 비교분석 방법으로 알려져 있다. 실제로 퍼지 점수를 이용한 분석은 결과변수를 구성하는 것에 있어서 원인변수의 집합이 어떤 포함관계를 가지며, 조합을 통해 이들의 설

명력을 파악한다(이승윤, 2014). 즉 사회현상은 동일한 결과를 야기하는 복수의 원인과 경로가 존재할 수 있으며, 이러한 원인조건들이 독립적인 영향력을 가지고 있다고 보는 관점보다는 다양한 방식으로 결합하여 특정한 결과에 영향을 끼친다고 보는 것이 적절하다는 것이다(민기채, 2014).

FsQCA 분석에서 충분조건을 검증하기 위한 작업으로 진리표를 도출한다. 진리표는 주어진 퍼지 점수를 일정한 기준값을 토대로 0과 1로 전환하는 것을 의미하며, 논리적 조건에서 결과변수가 어떻게 구성되는지 파악하기 위해 사용된다. 본 연구에서의 결과변수가 관광객의 전반적인 만족도라면 이를 설명하기 위해 원인조건인 관광지 선택속성들 간의 의미 있는 조합이 어떻게 구성되고 있는지 분석하는 것이다. FsQCA 분석에는 측정된 원점수를 0 이상 1 이하의 값으로 변환해야 하는데, 이 과정에서 사례가 어디에 소속되는가의 여부를 0(속하지 않음) 이상, 1(속함) 이하의 숫자로 눈금 매기기를 함으로써 정교하게 숫자로 표현할 수 있다(이승윤, 2014).

본 연구는 관광지 만족도를 구성하는 관광지 선택속성의 다양한 조합을 확인하는 것을 목적으로 한다. 즉, 관광지 만족도가 관광지 선택속성을 구성하는 설명력 높은 특정 요인에 의해서 결정된다고 보는 것이 아니라, 다소 복잡하게 얽혀 있는 여러 요인들 간의 상호작용에 의해서 관광지 만족도가 결정된다고 보는 것이다. 그러므로 본 연구에서는 중범위 사례에 대한 가설을 도출할 수 있는 장점을 가진 FsQCA 방법론을 적용하고자 하였다.

FsQCA 분석을 통해 얻을 수 있는 방법론적 이점은 크게 다음의 세 가지로 정리할 수 있다. 첫 번째, 회귀분석에서 적은 사례 수로 발생할 수 있는 자유도와 다중공선성 문제를 극복할 수 있다(조경훈·박형준, 2015a). 현재 대부분의 통계 방법은 정규분포에 기반을 두는데, 모집단의 분포가 정규분포하는지 여부와 관계없이 표본의 크기가 커질수록, 특히 30개 이상이 되면 표본 평균의 분포가 정규분포를 이루므로 정규분포가 가지는 특성들을

적용할 수 있다(박원우 외, 2010). 회귀분석에서도 종속변수와 독립변수의 인과성을 규명하는데 필요한 통계적 유의성을 확보하기 위해 표본의 수가 30개 이상이 되어야 할 것을 권고한다(조경훈·박형준, 2015a). FsQCA 분석은 표본의 수가 적은 경우, 발생하는 소수의 사례 문제를 해소하는데 효과적인 분석 도구로 평가받고 있다(Ragin, 2000). 이 외에도, 회귀분석에는 통계적인 자유도의 문제와 독립변수 간 상관관계 및 선형모형 가정에 따른 다중공선성 문제가 발생할 수 있다. 그러나 FsQCA 분석은 소수의 사례를 대상으로 원인조건 및 원인조건조합에 대한 분석이 가능하며, 각 원인조건 간 독립성 및 선형관계에 대한 가정을 사용하지 않고 인과적인 조건 및 결합에 대한 분석이 가능하다는 장점이 있다(조경훈·박형준, 2015a; 조경훈·박형준, 2015b). 두 번째, FsQCA 분석은 양적 분석과 질적 분석을 혼합한 형태의 분석 방법이므로 사례와 변수를 질적 연구 측면에서 고려하여 원인조건과 결과변수를 설정하되, 이를 양적인 자료로 전환하는 과정을 거침으로써 질적 연구의 단점을 보완하고 양적 연구가 가지는 논리적인 결과를 도출할 수 있다(오정호, 2001; 최신애·여정성, 2010; 양고운·박형준, 2013). 세 번째, 특정 원인조건이 결과변수에 미치는 영향만이 아닌 복수의 원인조건이 다양하게 결합하여 만들어내는 인과관계를 확인할 수 있다는 것이 장점이다(Rihoux, 2006; Ragin, 2009; 김종일, 2010).

### 3. FsQCA의 선행연구 동향

FsQCA 분석은 다양한 학문 분야에 적용되었으며, 결과에 대한 원인으로 다양한 원인조건조합이 존재한다면 학문 분야에 큰 제약을 받지 않고 FsQCA 분석 방법이 적용될 수 있다. 한 예로, 이산화탄소 배출에 대한 원인 조합을 파악하기 위해 FsQCA 분석을 적용한 연구가 있다(김기만·이수

연, 2021). 해당 연구에서는 FsQCA 분석을 통해 온실가스 배출의 주요 원인이 에너지 사용이며, 온실가스 배출이 높은 국가에서 에너지 사용이 전제되는 조건으로 나타나는 결과를 보고하였다. 온실가스 배출이라는 사회적 현상을 깊은 차원에서 이해하기 위해 영향 요인의 복잡성과 다양성을 고려하였으며, 원인조건의 결합적 관점에서 분석을 시도하였다. 이러한 FsQCA 분석은 국가 정책을 위한 시사점 마련에 유용하게 활용되는 분석 방법이다.

OECD 국가들 간의 출산율 차이를 설명하기 위해 일-가정 갈등을 중심으로 직무 긴장과 연관되는 삶의 만족도, 그리고 가정 내 가사노동의 균등한 분배와 출산율 간의 관계를 검증한 연구가 이에 해당한다(전승봉, 2020). 해당 연구에서는 출산율과 관련된 기존 연구들이 대체로 저출산에 대한 국가의 정책적 대응에 초점을 두고 있음을 언급하면서, 앞서 제시한 다양한 변수들이 어떤 방식으로 결합되어야 출산율에 영향을 미치는지 확인하였다. 즉, 출산율에 영향을 미칠 수 있는 요인으로 단일 원인조건을 제시하지 않고 노동시장조건과 가사노동시간의 분배, 삶의 만족도와 같은 원인조건들 간의 결합된 결과를 확인하는 것에 초점을 두었다.

이 외에도 서울시 청년 일자리 정책에 관한 연구도 FsQCA 분석을 적용한 연구에 해당한다(이혜윤·박병식·남궁금순, 2020). 해당 연구에서는 취업 지원 서비스, 맞춤형 전문인력 양성, 고용장려금 등 프로그램별로 청년 일자리 정책의 성과를 달성하기 위한 필요조건과 충분조건의 검증을 통해 서울시 일자리 정책(노동시장 프로그램)의 성공요인과 성공의 부재에 관한 요인을 결과로 제시하였다. 공공영역에서 노동시장 프로그램을 운영하고 그에 대한 성과달성의 실질적인 우선순위를 선정할 수 있는 접근방법을 제시하였으며, 정책에 관하여는 기대한 결과를 가져올 수 없는 요건들을 검증함으로써 향후 노동시장 프로그램 설계에 기여하는 시사점을 제공하였다.

또한 변수중심의 양적 연구는 인과관계를 설명할 때 독립변수와 종속변

수 간의 강력한 설명력을 전제 조건으로 하며, 사례 중심 연구는 주관적 해석에 의존하는 한계를 가지므로 혼합된 연구 방법으로 접근해야 된다는 주장에 따라 FsQCA 분석을 활용한 연구도 확인된다. 대표적으로 지역혁신시스템 이론을 바탕으로 지역혁신의 요소를 투입, 인프라, 네트워크로 구성하고, 이를 통해 16개의 지방자치단체의 지역혁신 유형을 분류하고 이를 활용하여 지역혁신의 성과인 높은 특허출원의 필요조건과 충분조건을 제시한 연구가 있다(김규환·박인권, 2018). 해당 연구에서는 종속변수에 대한 독립변수의 개별효과를 검증하는 것이 아니라 결합 요인 분석을 통해 지역혁신을 발생시키는 다양한 조건들을 규명함으로써 지역혁신 연구를 위한 새로운 방향을 제시하였다.

국내 서비스 산업은 외형적으로 많은 발전을 거듭해 왔으나, 서비스 산업의 내실화를 위해 서비스 개발 및 제공과 마케팅 분야에만 치중되고 서비스 산업에 적합한 관리시스템을 설계·운영하는 것이 미흡하다는 문제 제기를 시작으로 FsQCA 분석을 서비스 분야 관리시스템에 적용한 연구도 확인된다(강봉준·김재환, 2017). 조직의 목적을 달성하기 위해 활용되는 관리시스템은 상황별로 다르게 구성되어야 하므로, 서비스 조직에 적합한 관리시스템의 효과성 결정요인을 살펴보는 것이 필요하다고 하였다. 이에 운영성과를 결정하는 요인을 살펴보고자 FsQCA 분석을 연구에 접목함으로써 다양한 원인변수가 복합적으로 작용하는 상황을 구체적으로 살펴볼 수 있었다.

일반적으로 FsQCA 분석은 국가 정책과 관련된 연구에서 정책적 시사점 발굴을 위해 주로 활용되었으나 환대산업 분야에서 선택속성과 관련하여 FsQCA 분석이 접목되어 활용된 연구도 일부 확인된다. 대개 소비자는 제품 및 서비스 구매 과정에서 자신의 능력과 기호에 맞는 대안을 선택하며 소비자의 의사결정 과정에서 속성이 중요하게 고려되는 것을 토대로 한다고 언급한 연구가 대표적이다. 커피전문점의 점포 유형 및 커피전문점 방문 소



비자의 주요 선택속성을 원인변수로 설정하여 결과변수인 만족도 간의 관계를 FsQCA 관점에서 분석한 연구가 있다(한영위 · 이용기 · 안성만, 2017). 해당 연구에서는 커피전문점을 방문하는 고객들의 만족에 영향을 미치는 선택속성 조합을 밝히고, 커피전문점 점포유형(프랜차이즈/개인)에 따라 방문고객 만족에 영향을 미치는 선택속성 조합 간의 차이를 규명하였다.

최근에는 103개 호텔을 대상으로 트립어드바이저에서 검색되는 호텔의 인기도 즉, 평판(순위)에 미치는 영향을 알아보기 위해 온라인 리뷰 속성의 6개 요인(리뷰의 수, 긍정적인 리뷰의 비율, 리뷰에 대한 답변 비율, 리뷰의 품질, 총 객실 수, 국제 공항과의 거리)에 따른 검색 순위의 차이를 FsQCA 분석을 통해 살펴본 연구가 확인된다(황수정, 2021). 해당 연구에서는 리뷰 평판 솔루션인 TrustYou의 보고서를 분석하여 트립어드바이저 순위에 영향을 미치는 요인을 식별하였으며, 103개의 호텔 사례를 수집하여 영향요인과 성과 간의 인과관계를 FsQCA를 사용하여 분석하였다. FsQCA 관점을 연구에 접목하여 호텔의 효과적인 온라인 평판 관리에 기여하고자 하였다.

이상으로 FsQCA 분석이 정책적 시사점 발굴을 위한 연구뿐만 아니라 일부 환대산업 분야에서도 적용된 것을 확인하였다. 현재 관광 분야의 FsQCA 분석과 활용은 드물지만, 점차 적용 범위가 확장되고 있음을 관련 선행연구 고찰을 통해 확인하였다.

## 제 3 장 연구 설계

### 제 1 절 연구과제 설정

본 연구의 첫 번째 연구과제는 한국관광100선에 5회 연속 선정된 자연관광지를 대상으로 텍스트 마이닝을 하여 리뷰 데이터를 수집하고, 빈도 분석을 하여 리뷰 내에 존재하는 주요 단어들을 파악하는 것이다. 관광지에 대한 온라인 리뷰 데이터에는 관광객이 고려하는 관광지별 주요한 특성이 담겨져 있다. 리뷰는 관광지에 대한 관광객의 전반적인 인식과 경험, 관광지에 대한 느낌 등을 자유롭게 표현한 것으로 관광객들이 실제 상황에서 고려하는 관광지 선택속성을 파악하는 데에 유용한 정보를 얻을 수 있다.

많은 경우에서 이러한 관광지 선택속성 및 만족도 간의 관계를 검증하기 위해 설문지법이 사용된다. 그러나 설문지법은 설문 문항을 구성하는 과정에서 연구자의 관심이나 주관에 개입될 여지가 있으며, 조사 규모를 대폭 확장하지 않는 한 구체적인 관광지 관련 정보까지 얻는 것에 제약이 따를 수 있다. 따라서 텍스트 마이닝 기법을 활용하여 온라인 리뷰 데이터를 수집하여 자연어 처리 분석을 수행하고 온라인 리뷰에 존재하는 주요 단어들을 확인하고자 하였다. 단어는 해당 관광지를 나타내는 주요 키워드라고 할 수 있으므로, 이는 관광지를 설명하기 위한 사실적 단서의 역할을 한다.

이러한 연구 목적 달성을 위해 자료 수집 단계에서 관광지 방문객의 실제 경험과 관광지에 대한 정보가 담긴 온라인 커뮤니티를 선정하고, 프로그래밍 언어 중 하나인 파이썬(Python) 기반의 전용 웹크롤러를 개발하여 온라인 상에 존재하는 리뷰 데이터를 수집하였다. 이렇게 수집된 데이터는 자연어 형태로 데이터가 존재하므로 원문 리뷰 데이터를 별도로 백업하고, 알

파벳 대·소문자, 한글, 숫자를 제외한 특수문자, 중복된 공백, 구두점 등의 불필요한 내용을 제거하는 1차 정제화 작업을 하였다. 리뷰 데이터에 존재하는 단어들의 형태소를 파악하기 위해 국문 형태소 분석기를 사용하였으며, 리뷰에서 큰 역할을 하고 있다고 판단되는 명사만을 추출하여 분석에 활용하였다. 이렇게 최종 도출된 명사들을 대상으로 빈도 분석하여 관광객들이 실제 리뷰에서 언급하는 주요 단어들을 파악하였다.

**연구과제 1. 한국관광100선에 5회 연속 선정된 자연관광지를 대상으로 텍스트 마이닝을 통해 리뷰 데이터를 수집하고, 빈도 분석을 하여 리뷰 내 존재하는 주요 단어들을 파악한다.**

본 연구의 두 번째 연구과제는 LDA 토픽모델링 기법을 사용하여 본 연구의 대상이 되는 관광지별로 리뷰 내용에 존재하는 관광지 선택속성을 도출하는 것이다. 앞서 언급한 연구과제 1에서 빈도 분석에 대한 결과로 리뷰에 존재하는 주요 단어들을 파악하였다면, 관광지별로 어떠한 토픽으로 리뷰 내용이 형성되고 있는지 토픽 단위로 확인하기 위해 TF-IDF(Term Frequency Inverse Document Frequency) 가중치를 적용한 LDA 토픽모델링을 하였다. 토픽모델링은 텍스트 마이닝 기법 중에서 주목받고 있는 알고리즘으로서 온라인 댓글이나 SNS상의 메시지와 같은 비정형 문서에서 연관성 있는 단어들을 군집화하여 해당 주제를 추론하기 위해 사용되는 통계적 분석 방법이다(김광국·김용환·김자희, 2018). 리뷰에는 관광지 선택속성과 관련된 내용들이 자연어 형태로 존재하는데, 토픽모델링을 적용하면 관광지 방문객이 고려하는 관광지 선택속성들이 어떠한 내용의 토픽으로 형성되는지 알 수 있기에 유용하다. 토픽모델링에서 발견되는 토픽들은 해당 관광지의 만족도를 결정하는 관광지 선택속성이라고 할 수 있다. 이렇듯 리뷰에

존재하는 핵심 토픽을 파악하는 것은 관광지에 대한 방문객의 인식과 관심사에 대한 내용을 단어 기반의 빈도 분석으로 확인하는 것이 아닌, 토픽 단위로 이해할 수 있으므로 중요하다. 한편, TF-IDF란 문서 내 해당 단어에 대한 출현 빈도와 출현하는 문서의 개수를 이용하는 것으로 하나의 리뷰에서 특정 단어의 중요도는 그 리뷰 내 출현 빈도와 비례하고 전체 리뷰의 개수와는 반비례하는 특성을 활용하여 중요 핵심어를 추출하는 가중치 부여 방식이다(심영석·김홍범, 2016). TF-IDF 가중치를 적용하면 많은 리뷰 데이터에서 빈번하게 등장하는 단어들에 상대적으로 낮은 가중치를 부여한 상태에서 주요 토픽을 추출할 수 있다. 정리하면, TF-IDF 가중치를 적용한 LDA 토픽모델링 결과를 바탕으로 관광지별 관광객들이 실제로 리뷰에서 언급하는 주요 관광지 선택속성을 토픽 단위로 도출하고자 하였다.

**연구과제 2. TF-IDF 가중치를 적용한 LDA 토픽모델링 기법을 사용하여 본 연구의 대상이 되는 관광지별로 리뷰 내용에 존재하는 주요 관광지 선택속성(토픽)을 도출한다.**

본 연구의 세 번째 연구과제는 관광지 선택속성(토픽)을 원인조건으로, 관광지 만족도를 결과변수로 하여 FsQCA 분석을 하는 것이다. 앞서 언급한 연구과제 2에서 TF-IDF 가중치를 적용한 LDA 토픽모델링에 대한 결과로 리뷰에 존재하는 주요 관광지 선택속성이 토픽으로 도출되었다. 그러나, 해당 토픽만으로 관광지 선택속성과 관광지 만족도 간의 관계를 구체적으로 파악하는 것은 어렵다. 텍스트 데이터 특성상 각각의 리뷰를 구성하는 관광지 선택속성의 조합은 리뷰마다 다르게 나타날 것이기 때문이다. 예를 들어, 관광지 만족도가 5점인 리뷰가 여러 건 존재한다고 가정했을 때, 관광

지 만족도 5점을 설명하기 위해서는 관광객들마다 고려하는 관광지 선택속성이 다를 것이고, 서로 다른 관광지 선택속성의 조합으로 리뷰 내용이 구성될 것이다. 따라서 관광지 선택속성(토픽)을 원인조건으로, 관광지 만족도를 결과변수로 설정하여 결합된 인과관계를 검증하고자 하였다.

이에 적합한 분석 방법으로 FsQCA가 사용될 수 있다. 결과변수인 관광지 만족도에 대하여 필요조건에 해당하는 관광지 선택속성을 밝히며, 충분조건을 검증하기 위한 원인조건조합이 담긴 진리표를 도출하고, 진리표를 토대로 관광지 만족도를 결정하는 다양한 원인조건조합의 도출을 목적으로 하였다. 이로써 결과변수(관광지 만족도)를 설명하는 다양한 원인조건(관광지 선택속성)의 조합을 확인할 수 있다. 본 연구는 관광지 만족도를 결정하는 다양한 원인조건조합의 도출을 목적으로 하되, 관광지 선택속성과 관광지 만족도 간의 관계를 FsQCA 관점에서 설명하는 것에 초점을 두었다.

**연구과제 3. 관광지 선택속성(토픽)을 원인조건으로, 관광지 만족도를 결과변수로 설정하여 FsQCA 분석을 진행한다.**

연구과제 3-1 : 관광지 만족도(결과변수)에 대하여 필요조건에 해당하는 관광지 선택속성(원인조건)을 밝힌다.

연구과제 3-2 : 관광지 만족도(결과변수)에 대하여 충분조건을 검증하기 위한 원인조건조합 진리표를 도출한다.

연구과제 3-3 : 관광지 만족도를 결정하는 다양한 원인조건조합을 토대로 관광지 선택속성과 관광지 만족도의 관계를 FsQCA 관점에서 설명한다.

## 제 2 절 자료 수집의 대상 및 범위

### 1. 자료 수집을 위한 커뮤니티 선정

연구 진행에 앞서 관광지에 대한 실제적인 평가와 리뷰가 활발하게 공유되는 온라인 커뮤니티로 트립어드바이저를 선정하였다. 트립어드바이저는 세계적으로 매달 4억 6,000만명이 넘는 방문객에 대하여 859만 개의 숙박시설, 음식점, 체험, 항공권, 크루즈 여행에 대하여 방대한 양의 리뷰를 제공하는 여행 커뮤니티이다(Tripadvisor, 2019). 본 연구가 시작된 시점인 2021년 3월에는 트립어드바이저에 약 8억 8천 7백만 개 이상의 리뷰가 형성된 것으로 나타났다. 트립어드바이저는 대한민국 관광지에 대한 평가뿐만 아니라 특정 관광지와 연관된 호텔, 레스토랑에 대한 평가 서비스를 제공하며 49개의 지역에서 28개의 언어로 이용 가능한 서비스를 선보이고 있다.

자료 수집을 진행하기 전, 트립어드바이저 웹사이트에서 리뷰를 수집하는 것이 적절한지 여부를 우선 확인하였다. 이를 위해 트립어드바이저 웹사이트에 위치한 ‘robots.txt’ 문서를 확인하여 트립어드바이저 웹사이트가 리뷰에 대한 크롤링을 허용하고 있는지 여부를 확인하였다. robots.txt는 자동화된 검색 로봇이 사이트나 웹페이지를 수집할 수 있도록 허용하거나 제한하는 내용이 담긴 문서이다. 그 결과, 트립어드바이저에서 리뷰를 수집하는 것에 별다른 제한 사항이 존재하지 않는 것으로 나타났다. 이미 해외에서는 트립어드바이저에 게재된 온라인 리뷰 데이터를 활용한 연구가 다수 존재하며(Liu, Schuckert, & Law, 2015; Schuckert, Liu, & Law, 2015), 국내에서도 트립어드바이저를 통한 온라인 리뷰 데이터 활용 연구들이 다수 확인된다(김민식·한학진·박병화, 2017; 이수희·박득희·김맹선, 2018; 임영희·김홍범, 2019; 박영욱·정규엽, 2021; 장윤정·정수아, 2021). 이에 본 연구도

트립어드바이저에서 제공하는 온라인 리뷰 데이터를 수집 및 활용하는 것에 있어서 문제 사항이 발견되지 않는다고 판단하여 트립어드바이저를 자료 수집을 위한 커뮤니티로 선정하였다.

## 2. 자료 수집의 대상

관광객은 크게 내국인 관광객과 외국인 관광객으로 구분할 수 있다. 본 연구는 한국관광100선에 5회 연속 선정된 관광지를 중심으로 내국인이 고려하는 관광지 선택속성을 도출하기 위해 내국인 방문객이 작성한 국문 리뷰 데이터만을 본 연구의 자료 수집 및 분석의 대상으로 하였다. 한국을 대표하는 관광지 리뷰에서 발견되는 주요 관광지 선택속성을 파악하는 것은 대한민국 관광산업의 발전과 관광업으로 인한 경제 성장을 이끌기 위해 중요한 과제이다. 이에 본 연구의 모집단은 최근 10년 이내 「한국관광100선」 관광지를 방문한 내국인 관광객으로 설정하였다. 한국관광100선은 전문가 심사와 통계 분석 등 다양한 검증을 거쳐 선정되는 것으로(문화체육관광부, 2021), 여기에 선정된 관광지는 대한민국을 잘 나타내는 관광지라고 할 수 있다. 10년 이내라는 구체적인 시간적 범위를 제시한 것은 한국관광100선을 진행하기 위한 사업이 2013년에 처음 시작되었기 때문이다. 이 중에서도 5회 연속 선정된 총 19개의 관광 명소를 본 연구의 대상으로 하였으며, 이는 본 연구의 목적을 달성하기 위한 범위로도 적합하다.

## 3. 자료 수집의 범위

자연관광은 일반적으로 자연관광자원을 바탕으로 하는 관광을 의미하며, 자원관광자원으로는 산(림), 호수, 하천, 해양(안), 동·식물, 온천, 천문, 사

막, 섬, 농촌·어촌·산촌, 해수욕장, 습지, 늪지, 생태지역 등이 존재함에 따라 그 범위가 매우 넓고 포괄적이다(김미라, 2020). 이러한 자연관광은 어떤 자연관광자원을 토대로 이루어지는가 또는 자연관광자원과 연계되어 어떤 관광행위가 주로 행해지는가에 따라 세부적으로 분류될 수 있다. 자연관광지를 분류하는 기준은 선행연구마다 차이가 존재하나, 본 연구는 관광지를 분류하기 위한 기준으로 TourAPI3.0에 제시된 기준을 토대로 하였다.

TourAPI3.0에 따르면 관광지 유형은 ① 자연관광지(공원·산, 바다·강) ② 역사·문화재관광지 ③ 체험관광지 ④ 문화시설로 구분할 수 있다. 이 중에서 자연관광지는 공원·산 자연관광지와 바다·강 자연관광지로 세분화된다. 한국관광100선에 5회 연속 선정된 관광지 19개소를 TourAPI3.0가 제시한 기준에 따라 분류한 결과는 <표 3-1>에 정리되어 있다. ‘공원·산 자연관광지’에는 보성 녹차밭(대한다원), 설악산 국립공원, 한라산 국립공원이 포함되며, ‘바다·강 자연관광지’에는 순천만습지(순천만국가정원), 두물머리, 남이섬, 울릉도와 독도, 창녕 우포늪, 안면도 꽃지해변, 우도가 포함된다. ‘역사·문화재관광지’에는 서울 5대 고궁(경복궁·경희궁·창덕궁·창경궁·덕수궁), 전주 한옥마을, 수원화성, 불국사와 석굴암, 안동 하회마을, 합천 해인사가 포함되며, ‘체험관광지’에는 남산 N서울타워, 태종대유원지, 제주올레길이 포함된다. 다만, 한국관광100선에 5회 연속 선정된 관광지 19개소 중에서 ‘문화시설’에 해당하는 관광지는 없는 것으로 나타났다.



<표 3-1> 한국관광100선에 5회 연속 선정된 관광지 유형별 분류\*

관광지 유형	해당 유형에 속하는 관광지명
공원 · 산 자연관광지 (3개소)	보성 녹차밭(대한다원) 설악산 국립공원 한라산 국립공원
바다 · 강 자연관광지 (7개소)	순천만습지(순천만국가정원) 두물머리 남이섬 울릉도와 독도 창녕 우포늪 안면도 꽃지해변 우도
역사 · 문화재관광지 (6개소)	서울 5대 고궁 전주 한옥마을 수원화성 불국사와 석굴암 안동 하회마을 합천 해인사
체험관광지 (3개소)	남산 N서울타워 태종대유원지 제주올레길
문화시설	해당유형에 속하는 관광지 없음

\* : 표에 제시된 관광지는 모두 한국관광100선이 처음 시작된 때로부터 5회 연속 선정된 관광지임  
관광지 유형을 구분하는 기준은 TourAPI3.0에 제시된 기준을 토대로 하였음

본 연구는 한국관광100선에 5회 연속 선정된 관광지 19개소 중에서 가장 사례(리뷰의 개수)가 많은 자연관광지만을 자료 수집 및 분석의 대상으로 하였다. 자연관광지는 앞서 언급한대로 공원 · 산 자연관광지와 바다 · 강 자연관광지로 구분할 수 있다. 분석에 필요한 자료 수집을 위해 트립어드바이저 웹사이트에서 관광지명을 키워드로 입력하여 검색되는 결과를 활용하

였다. 예를 들면, 설악산 국립공원에 대한 리뷰 내용을 확인하고자 할 때 ‘설악산 국립공원’에 대한 검색 결과만을 참고한 것이다. 한 예로, 트립어드 바이저에서는 설악산 국립공원과 연관된 장소로 ‘금강굴’, ‘권금성’, ‘비룡폭포’, ‘신흥사’, ‘울산바위’ 등이 검색되나, 이는 국소적인 장소에 해당하므로 자료 수집의 범위에 포함하지 않았다.

대부분의 경우에서 한국관광100선이 제시한 관광지명과 동일한 결과가 검색되었으나, 일부 관광지는 한국관광100선이 제시한 관광지명과 다르게 검색되었다. 순천만습지(순천만국가정원)는 ‘순천만습지’와 ‘순천만정원’이라는 별개의 키워드로 검색되었으며, 각 관광지에 해당하는 리뷰 페이지가 별도로 생성된 것을 확인하였다. 따라서, 순천만습지(순천만국가정원)의 경우, 순천만습지에 해당하는 리뷰와 순천만정원에 해당하는 리뷰를 각각 수집하고, 이를 합산하여 분석에 활용하였다. 안면도 꽃지해변의 경우, 정확하게 일치하는 검색어가 없어 ‘꽃지해수욕장’으로 검색되는 결과를 확인하였다.

한편, 한국관광100선에 5회 연속 선정된 자연관광지에 대한 리뷰를 모두 분석에 활용하는 것은 사실상 무리가 있다고 판단하였다. 내국인이 작성한 리뷰의 개수를 확인하였을 때, 보성 녹차밭(대한다원), 두물머리, 울릉도와 독도, 창녕 우포늪, 안면도 꽃지해변에 대하여 확인되는 리뷰의 개수는 17~87건으로 모두 100건 미만이다. 관광지에 대한 리뷰 개수가 충분하게 확보되었다고 하기 어렵지만, 본 연구는 리뷰에 존재하는 관광지 선택속성을 밝히려는 탐색적 차원의 연구로써, 연구 목적 달성을 위해 100건 이상의 리뷰가 포함된 자연관광지만을 최종 분석의 대상으로 하였다. 종합하면, 한국관광100선에 5회 연속 선정된 자연관광지 중에서 리뷰의 수가 100건 이상인 관광지 5개소, ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도를 본 연구의 자료 수집 대상으로 하였다. 이 내용을 요약하면 <표 3-2>와 같다.

<표 3-2> 본 연구의 분석 대상이 되는 자연관광지에 대한 기본 정보

구분	관광지명	전체 리뷰 개수	내국인 리뷰 개수	분석 대상여부
공원·산 자연관광지	보성 녹차밭 (대한다원)	143	48	제외
	설악산 국립공원	1,015	132	○
	한라산 국립공원	1,076	202	○
	순천만습지 (순천만국가정원)*	277 (165)	149 (99)	○
바다·강 자연관광지	두물머리	115	87	제외
	남이섬	1,882	112	○
	울릉도와 독도	43	17	제외
	창녕 우포늪	25	17	제외
	안면도 꽃지해변**	32	26	제외
	우도	1,066	420	○

관광지 유형은 TourAPI3.0 기준으로 분류하였으며, 리뷰 개수는 2021.10.19를 기준으로 작성된 결과임

\* : 트립어드바이저에서 검색되는 관광지명 '순천만정원'

\*\* : 트립어드바이저에서 검색되는 관광지명 '꽃지해수욕장'

#### 4. 조사설계 요약

이상으로 자료 수집을 위한 커뮤니티 선정을 마치고, 자료 수집의 대상을 선정하였다. 자료 수집의 기간 및 범위 등을 포함하여 전반적인 조사설계에 대한 절차 및 방법을 요약하여 정리한 내용은 <표 3-3>와 같다.

<표 3-3> 조사설계 요약

구분	설명
모집단	한국관광100선 자연관광지를 방문한 내국인 관광객
표본	최근 10년 이내 한국관광100선 자연관광지를 방문하여 트립어드바이저에서 국문 리뷰를 작성한 내국인 관광객
연구의 대상 (관광지 5개소)	설악산 국립공원, 설악산 국립공원, 순천만습지(순천만국가정원), 남이섬, 우도
표본의 크기	총 1,114건의 리뷰 데이터
커뮤니티 선정 및 자료 수집	트립어드바이저에 게시된 관광객의 리뷰 데이터 및 관광지 정보를 파이썬 기반의 웹크롤러를 개발하여 수집
자료수집 기간	2021년 10월
자료의 범위	최근 10년 간 내국인 관광객이 국문으로 작성한 리뷰 데이터 (2012년 4월 ~ 2021년 8월까지의 데이터가 포함됨)

## 제 3 절 분석 절차 및 방법

### 1. 리뷰 데이터 수집

본 연구의 대상이 되는 관광지 5개소는 ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도이다. 해당 관광지를 대상으로 파이썬 기반의 전용 웹크롤러를 개발하여 자료를 수집하였다. 파이썬 웹크롤러로 수집할 수 있는 항목은 다음과 같다. 트립어드바이저에서는 <그림 3-1>과 같이 작성자명(닉네임), 국적, 포스팅 개수, 추천 수, 평점, 리뷰 제목, 관광지 방문일, (예시에는 없지만) 동반자 유형, 리뷰 내용, 리뷰 작성일, 그리고 함께 첨부한 사진을 확인할 수 있다.



<그림 3-1> 실제 리뷰 예시

이 중에서 작성자 프로필 섬네일 사진 및 첨부한 사진을 제외한 텍스트 형태로 수집 가능한 항목을 모두 수집하였다. 이때 분석의 대상이 되는 국

문 리뷰만을 수집하기 위해 트립어드바이저 대한민국 도메인(co.kr)에 접속하여 ‘한국어’로 필터링하여 검색되는 리뷰만을 수집하도록 하였다.

## 2. 데이터 전처리 및 형태소 분석

리뷰 데이터를 분석하기 위해서는 텍스트 마이닝 기법으로 수집된 자료를 분석 가능한 형태로 처리하는 단계를 거쳐야 한다. 자료의 1차 정제화 작업을 위해 정규 표현식 모듈인 `re(Regular Expression)`를 사용하여 알파벳 대·소문자, 한글, 숫자를 제외한 특수문자, 중복된 공백, 구두점 등의 불필요한 내용을 제거하였다. 다음으로는 리뷰의 띄어쓰기 상태를 점검하였다. 한글은 띄어쓰기를 기준으로 단어의 의미가 달라질 수 있으므로 띄어쓰기 상태를 미리 점검하는 것이 중요하다. 국문 리뷰를 전처리하기 위한 방법으로 띄어쓰기가 없는 문장이라면 띄어쓰기 된 문장으로 변환하는 작업을 대신하는 `PyKoSpacing` 같은 패키지를 사용할 수 있으나, 수집한 리뷰 데이터는 전반적으로 띄어쓰기가 잘 되어 있다고 판단되어 파이썬 기반의 패키지를 이용한 별도의 띄어쓰기 작업을 진행하지 않았다.

이어서 리뷰 데이터의 형태소를 파악하기 위해 국문 형태소 분석기를 사용하였다. 파이썬에서는 `KoNLPy` 패키지를 설치함으로써 다섯 가지의 국문 형태소 분석기를 사용할 수 있다. `KoNLPy`에서 호환 가능한 형태소 분석기는 다음과 같다. ① KAIST Semantic Web Research Center에서 개발한 한나눔(Hannanum), ② 서울대학교 IDS(Intelligent Data Systems) 연구실에서 개발한 꼬꼬마(Kkma), ③ Shineware에서 개발한 코모란(Komorán), ④ 일본 어용 형태소 분석기를 한국어에도 사용할 수 있도록 한 메카브(Mecab), 그리고 ⑤ 오픈소스 한국어 분석기인 OKT(Open-source Korean Text Processor)<sup>5)</sup>가 있다. 연구의 목적에 따라 두 개 이상의 형태소 분석기를 혼합해서

사용할 수 있으나 본 연구에서는 다양한 선택지 중에서 OKT를 사용하여 국문 리뷰 데이터 전처리 작업을 하였다. OKT를 사용한 이유 중 하나는 OKT를 사용했을 때 리뷰 데이터의 정규화 처리가 가능한 것이 장점이기 때문이다. 정규화 처리를 진행하면 그 결과로 <표 3-4>와 같이 정제된 형태의 데이터를 확인할 수 있다.

<표 3-4> OKT를 활용한 정규화 처리 작업 전후 비교

작업 단계	결과
정규화 처리 전	한라산에 방문했습니당 경관이 넘호 멋집니다~ 오길 잘해썬요
정규화 처리 후	한라산에 방문했습니다 경관이 너무 멋집니다~ 오길 잘했어요
정규화 처리를 거치면서 변경된 부분	방문했습니당 → 방문했습니다 넘호 → 너무 잘해썬요 → 잘했어요

연구자가 직접 구현하고 얻은 결과를 표로 정리하여 제시하였음.

또한, OKT를 사용하면 어간추출(Stemming) 기능을 통해 분석의 효율성 일 높일 수 있다. 어간추출 기능을 사용하면 하나의 단어에 문법적 요소인 접사(affix)를 분리하여 의미가 담겨있는 부분인 어간(stem)을 표시해준다. 예를 들어, ‘봤는데’, ‘봤다’, ‘봤다면’, ‘봤으면’, ‘보면’, ‘보면서’, ‘보았으면’, ‘볼 것이다’라는 단어가 있을 때, 어간을 추출하면 이 단어들은 모두 ‘보다’라는 어간으로 추출된다. 이러한 어간추출 기능은 자연어를 처리하여 분석

5) OKT 웹사이트(<https://konlpy.org/ko/v0.5.2/api/konlpy.tag/#okt-class>)

을 하고자 할 때에 매우 유용하게 사용되는 기능이다. 본 연구에서는 명사만 추출하는 함수를 호출하였으므로, 이미 어간추출 기능이 포함되어 있다고 여겨져 별도의 어간추출 작업을 진행하지는 않았다.

한편, 한국어는 명사, 대명사, 수사, 동사, 형용사, 관형사, 부사, 조사, 감탄사의 아홉 개의 품사로 나뉘며 이 중 대명사, 수사, 관형사, 조사, 감탄사는 텍스트를 통한 분석에서 활용과 기여도가 낮을 것으로 판단하여 사용하지 않았다. 그중에서도 본 연구는 분석에서 큰 역할을 하고 있다고 판단되는 명사를 추출하여 분석에 활용하였다. 명사만 추출하였음에도 분석에 불필요한 일부 단어는 불용어로 처리하였다. 불용어를 처리하기 위한 방법으로 한국어 불용어 사전<sup>6)</sup> 같은 사이트를 참고하여 분석에 필요한 불용어를 가져올 수 있으나, 본 연구에서는 불용어로 지정할 단어를 .csv(comma-separated value) 파일에 저장하여 해당 파일을 불러와 불용어를 처리하였다. 불용어로 지정한 단어에는 ‘그때’, ‘다음’, ‘앞’, ‘이번’, ‘조금’, ‘특정’ 등이 포함된다. 그러나 OKT를 사용하여 형태소 분석을 할 때 일부 단어는 한국어 품사 중에서 부사에 해당하는데 명사에 분류되는 것을 확인하였다. 따라서, ‘그냥’, ‘아주’, ‘위낙’, ‘좀’, ‘진짜’ 등의 부사도 불용어에 포함하였다. 이렇게 최종 도출된 명사들은 빈도 분석에 활용하였으며, 관광객들이 고려하는 관광지에 대한 주요 키워드 및 내용 확인을 위한 목적으로 사용하였다.

### 3. 기본 분석

본 연구의 대상인 관광지 5개소에 대한 관광지별 온라인 리뷰 정보 및 표본의 특성을 파악하기 위해 SPSS ver. 25.0을 사용하여 기술통계 분석을 하였다. 여기에는 관광지별로 분석에 사용된 리뷰의 개수, 관광객 만족도를

---

6) 불용어 웹사이트 예시(<https://www.ranks.nl/stopwords/korean>)



나타내는 평점, 동반자 유형, 방문일 등의 정보가 포함된다. 이어서 파이썬의 `collections` 모듈에서 `Counter` 클래스를 호출하여 빈도 분석을 하였다. 별도의 가중치를 적용하지 않은 상태에서 관광지별 상위 출현하는 단어의 빈도에 따라 1위부터 90위까지의 단어들을 확인하였으며, 빈도를 비율로 변환하여 해당 단어의 차지 비율도 함께 제시하였다. 이때 비율이 높을수록 빈도가 높게 나타난 단어들이며, 관광객들이 리뷰에서 자주 언급하는 단어라고 할 수 있다. 본 연구의 분석 대상인 관광지는 5개소이므로 관광지별로 다섯 개의 빈도 분석 결과가 도출되었다. 빈도 분석을 통해 관광지 5개소에 나타난 주요 단어를 확인하였으며, 파이썬의 `wordcloud` 모듈을 사용하여 시각화 작업을 하였다. 이때에는 관광지별 상위 출현하는 120개의 단어를 대상으로 워드클라우드 시각화 작업을 하였으며, 추가로 관광지 5개소 전체에 대한 워드클라우드 시각화 결과도 함께 제시하였다. 종합하면, 관광지 5개소에 대한 빈도 분석 결과를 요약하고, DTM(Document-Term Matrix)을 형성하여 최소 세 군데 이상의 관광지에서 공통으로 출현하는 단어들을 함께 확인하였다.

#### 4. LDA 토픽모델링

온라인 리뷰에는 관광지 선택속성에 대한 내용이 토픽의 형태로 나타날 것이므로 리뷰에 존재하는 관광지 선택속성을 파악하기 위해 LDA 토픽모델링을 하였다. 이는 관광지 선택속성에 대한 리뷰 내용이 어떠한 토픽으로 형성되고 있는지 알아볼 수 있기에 중요하다. 이를 위해 파이썬의 `tomotopy0.12.27)` 패키지를 사용하여 TF-IDF의 가중치를 적용한 LDA 토픽모델링을 하였다. LDA는 LSA 등의 다른 토픽모델링 방법에 비해 결과 해석

---

7) `tomotopy0.12.2` 웹사이트(<https://bab2min.github.io/tomotopy/v0.12.2/kr>)

이 용이하며(Blei, 2012), 비정형 데이터의 차원을 축소해 여러 주제로 도출해주는 장점을 지닌다. 효율적인 토픽 도출을 위해 TF-IDF 가중치 적용 방식을 택하였다. TF-IDF 가중치를 적용하면 많은 리뷰에서 상대적으로 자주 등장하는 단어는 중요도가 낮다고 여겨지며, 많은 리뷰에서 자주 등장하지는 않지만 특정 리뷰에서 자주 등장하는 단어는 중요도가 높다고 판단한다. 이는 단순히 빈도 높은 단어를 중요도 높은 단어로 판단하는 것과는 다른 방식이다. 관광지에 대한 TF-IDF 기반의 LDA 토픽모델링을 마친 다음에는 토픽명을 부여하고 각 토픽을 구성하는 주요 단어를 비중 높은 순서대로 정리하여 제시하였다. 관광지별로 도출되는 토픽의 특성을 살펴보면서 특정 관광지에서만 나타나는 독특한 토픽에는 무엇이 있는지 확인하고, 최소한 두 개 이상의 관광지에서 공통으로 도출되는 토픽도 함께 관찰하였다.

## 5. FsQCA(퍼지셋 질적비교분석)

관광지 선택속성(토픽)을 원인조건으로, 관광지 만족도를 결과변수로 설정하여 FsQCA 분석을 진행하는 것은 본 연구의 주요 목적을 달성하기 위함이다. 여기에는 윈도우 버전의 FsQCA3.0<sup>8)</sup> 분석 툴이 사용되었다. FsQCA 분석을 통해 높은 관광지 만족도를 결정하는 다양한 원인조건의 조합을 도출할 수 있으며, 관광지 선택속성과 관광지 만족도와의 관계를 FsQCA 관점에서 설명할 수 있다. 일반적인 QCA는 사례가 집합에 포함되거나 포함되지 않는 불리언 대수를 기반으로 해당 집합의 소속여부를 표현한다.

즉, QCA에서는 0(비소속)과 1(소속)만으로 집합의 상태를 나타내는 반면, FsQCA를 사용하면 퍼지집합을 구성할 때 0과 1이라는 이산 변수 사이에 다양한 정도를 넣어 표현함으로써 정보의 손실을 최소화하여 분석 가능

---

8) FsQCA 웹사이트(<https://www.socsci.uci.edu/~cragin/fsQCA/software.shtml>)

한 것이 장점이다(강봉준·김재환, 2017). FsQCA는 QCA와 마찬가지로 집합에 소속되지 않음을 나타낼 때 0을 사용하고, 집합에 소속됨을 나타낼 때 1을 사용한다. FsQCA 분석에서는 0부터 1사이의 퍼지 점수를 부여할 수 있는 것이 가장 큰 특징인데, 퍼지 점수를 부여하는 방법에는 여러 가지가 있으며 그중에서 일반적으로 활용되는 세 가지 방식을 소개하면 다음과 같다.

첫 번째, 4개의 값으로 퍼지 점수를 부여하는 방식이다. 이 경우, 0과 1 사이에 두 개의 중간값을 넣어 비소속(0), 부분 비소속(0.33), 부분 소속(0.67), 소속(1)의 값으로 표현한다. 두 번째, 6개의 값으로 퍼지 점수를 부여하는 방식이다. 이 경우, 0과 1 사이에 네 개의 중간값을 넣어 비소속(0), 거의 비소속(0.2), 약간 비소속(0.4), 약간 소속(0.6), 거의 소속(0.8), 소속(1)의 값으로 표현한다. 이러한 방식은 5점 리커트 척도가 활용된 자료에서 적용 가능한 퍼지 점수 변환 방식이다. 세 번째, 연속적인 값으로 퍼지 점수를 부여하는 방식이다. 이 경우, 0 이상 1 이하의 모든 값으로 비소속과 소속의 정도를 세밀하게 표현할 수 있다. FsQCA 분석을 하기 위해서는 연구에 사용되는 변수를 0과 1을 포함하여 0과 1 사이의 퍼지 점수로 변환하는 작업이 선행되어야 한다.

본 연구에서 사용된 퍼지 점수 부여 방식은 다음과 같다. 우선 원인조건인 관광지 선택속성(토픽)은 토픽모델링을 통해 리뷰별로 할당된 토픽의 비율을 활용하였다. 토픽모델링을 진행할 때 각 토픽에 할당되는 모든 토픽 비율의 합은 1이므로, 퍼지 점수 환산을 위해 얻어진 토픽 점수에서 토픽의 수 만큼 곱한 값을 활용하였다. 예를 들어, A 관광지에서 도출된 5개에 대한 개별 토픽 비율이 1번 토픽(0.2), 2번 토픽(0.3), 3번 토픽(0.15), 4번 토픽(0), 5번 토픽(0.35)이라고 가정할 때, 이 값에서 토픽의 수인 5를 곱한 값을 사용하는 방식이다. 이때, 퍼지 점수로 0 이상 1 이하의 값만을 사용할 수 있으므로 1 이상의 값이 계산되면 퍼지 점수로 부여할 수 있는 최대 점수

인 1을 부여하였다. 한편, 본 연구에 사용된 결과변수인 관광지 만족도는 5점 리커트 척도로 얻어진 점수이므로 앞서 제시한 두 번째 방식을 적용하여 퍼지 점수로 변환하였다. 관광지 만족도는 1점부터 5점까지 존재하므로 관광지 만족도 원점수에서 나누기 5를 하여 퍼지 점수를 구하였다. 예를 들어, 관광지 만족도에 별 다섯 개(★★★★★)를 부여한 관광객이라면 퍼지 점수로 나타낼 때  $5 \div 5 = 1$ (소속)로 변환할 수 있다. 마찬가지로 관광지 만족도에 별 한 개(★☆☆☆☆)를 부여한 관광객이라면 퍼지 점수로 나타낼 때  $1 \div 5 = 0.2$ (거의 비소속)로 변환된다.

본 연구는 퍼지 점수를 활용하여 다음의 절차에 따라 FsQCA 분석을 진행하였다. Ragin(2009)은 FsQCA 분석에 앞서 원인변수와 결과변수의 관계에서 필요조건의 성립여부를 확인할 것을 권장한다. 각 원인변수의 필요조건에 대한 일치도(consistency for necessity)가 0.75 이하로 나타나는 경우 하위집합의 관계, 즉 필요조건으로 보기 어렵다고 하였으므로(김예린, 2018 재인용 p.122), 본 연구에서는 일치도 값이 0.75 이상일 때 필요조건을 충족한다고 보았다.

한편, 충분조건에 대한 검증 또한 중요한 부분이다. FsQCA 분석에서는 독립변수라는 표현 대신 원인조건이라는 용어를 사용하는데, 그 이유는 각각의 조건이 독립적으로 결과변수에 영향을 미치는 것이 아닌, 원인조건들의 조합을 통해 결과변수와 의미 있는 인과관계를 형성한다고 전제하기 때문이다. 충분조건을 검증하기 위해서는 진리표<sup>9)</sup>(Truth Table Algorithm)를 도출해야 한다. 진리표는 원인조건 수에 따라 조합 가능한 조건의 수를 요약하여 보여주는 표이다. 이론상 본 연구에 나타날 수 있는 6개 원인조건(관광지 선택속성)에 대한 원인조건조합의 수는  $64(2^6)$ 개이다. 즉, 6개의 원인조건의 결합으로 64개의 경우의 수가 존재할 수 있는 것이다. 이 중에서 사

---

9) 진리표 대신 '진실표'라고 언급하는 연구도 있음

례의 수(해당 조건과 일치하는 리뷰의 개수)가 현저하게 적은 경우를 제외하고, 시스템 기본값<sup>10)</sup>인 일치도 0.8 이상인 조합만을 대상으로 충분조건에 사용될 사례를 선별하였다.

진리표가 도출되고 충분조건을 검증하기 위한 분석 방법으로는 일반적으로 권장되는 표준분석을 택하였다. 표준분석을 하면 결과와 함께 복잡한 해법(complex solution), 중간 해법(intermediate solution), 그리고 간명한 해법(parsimony solution)의 세 가지 해법이 도출되는데, 이 중에서 중간 해법이 가장 최적화된 결과를 제시하므로 본 연구에서는 중간 해법을 제시하고 이를 통해 분석 결과를 설명하였다(Ragin, 2009). 본 연구에서 사용한 분석 방법과 분석 진행 절차를 요약하면 <표 3-5>와 같다.

---

10) FsQCA3.0 분석 툴에서 Analyze → Truth Table Algorithm → Edit → Delete and code를 통해 진리표 도출을 위한 일치도 기준이 0.8로 설정되어 있는 것을 확인할 수 있음

<표 3-5> 연구의 분석 방법 및 절차 요약

분석 방법	진행 절차	내용 요약 및 사용된 분석 도구
텍스트 마이닝	데이터 수집	국문 리뷰 데이터 수집 (파이썬 기반의 웹크롤러 사용)
데이터 전처리 및 형태소 분석	데이터 전처리	불필요한 공백, 구두점, 특수문자 등 제거 (정규 표현식 모듈인 re 사용)
	형태소 분석	명사 추출, 불용어 처리 (오픈소스 한국어 형태소 분석기인 OKT 사용)
기본 분석 (연구과제 1)	기술통계 분석	관광지별 온라인 리뷰 정보 및 표본의 특성 확인 (SPSS ver. 25.0의 기술통계 분석)
	빈도 분석	관광지별 리뷰에 자주 등장하는 단어 도출 (파이썬 collections 모듈의 Counter 클래스 사용)
LDA 토픽모델링 (연구과제 2)	가중치 부여	TF-IDF를 적용한 단어 가중치 산출
	LDA 토픽모델링	리뷰에 존재하는 관광지 선택속성(토픽) 도출 (파이썬 tomotopy0.12.2 패키지 사용)
FsQCA (연구과제 3)	필요조건 검증	관광지 만족도를 평가하는 원인조건(관광지 선택속성) 확인 (FsQCA3.0의 Necessary Conditions 기능 사용)
	진리표 도출	논리적으로 결합 가능한 원인조건(관광지 선택속성)조합 제시 (FsQCA3.0의 Truth Table Algorithm 기능 사용)
	충분조건 검증	관광지 만족도를 결정하는 다양한 원인조건조합의 대표 모형 도출 및 표준분석결과를 통해 중간 해법 제시 (FsQCA3.0의 Standard Analyses 기능 사용)

## 제 4 장 실증 분석

### 제 1 절 기술통계 분석

#### 1. 온라인 리뷰의 정보 및 표본의 특성

분석에 앞서, 「한국관광100선」에 5회 연속 선정된 자연관광지를 중심으로 트립어드바이저에서 국문 리뷰 데이터를 수집하고 기술통계 분석을 하였다. 본 연구의 대상이 되는 관광지 5개소는 ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도이다. 수집된 리뷰 데이터는 가장 오래된 방문 시점인 2012년 4월에 해당하는 리뷰부터 2021년 8월에 해당하는 리뷰 데이터까지 최근 약 10년 간의 누적된 데이터가 포함되어 있다. 본 연구의 대상이 되는 관광지 5개소를 대상으로 리뷰 내용과 함께 부여된 전반적인 관광지 만족도 점수를 살펴본 결과, 한라산 국립공원이 평균 평점 4.76점으로 가장 높았다. 반면, 남이섬은 평균 평점 4.14점으로 상대적으로 낮은 점수인 것으로 나타났다. 동반자 유형은 가족, 개인, 비즈니스, 친구, 커플, 무응답 중에서 하나의 유형만 선택할 수 있었으며, 응답 결과로는 설악산 국립공원, 순천만습지(순천만국가정원), 우도를 대상으로 할 때에 가족과 관광하는 비율이 가장 높은 것으로 나타났다. 반면, 한라산 국립공원은 친구와 방문하는 비율이 가장 높은 것으로 나타났고, 남이섬은 커플과 방문하는 비율이 가장 높은 것으로 나타났다. 본 연구의 대상이 되는 관광지 5개소에 대하여 수집된 트립어드바이저 국문 온라인 리뷰를 대상으로 표본의 특성을 요약한 내용은 <표 4-1>와 같다.

<표 4-1> 관광지별 온라인 리뷰 정보 및 표본의 특성

분류		설악산 국립공원		한라산 국립공원		순천만습지 (순천만국가정원)		남이섬		우도	
분석에 사용된 리뷰 개수		132		202		248		112		420	
평점	최소	3		2		3		1		1	
	최대	5		5		5		5		5	
	평균	4.62		4.76		4.61		4.14		4.49	
동반자 유형	가족	46	34.8 %	41	20.3 %	70	28.2 %	25	22.3 %	114	27.1 %
	개인	17	12.9 %	41	20.3 %	31	12.5 %	3	2.7 %	67	16.0 %
	비즈니스	3	2.3 %	2	1.0 %	5	2.0 %	5	4.5 %	6	1.4 %
	친구	21	15.9 %	56	27.7 %	53	21.4 %	18	16.1 %	101	24.0 %
	커플	24	18.2 %	37	18.3 %	38	15.3 %	35	31.3 %	63	15.0 %
	무응답	21	15.9 %	25	12.4 %	51	20.6 %	26	23.2 %	69	16.4 %
	전체	132	100.0 %	202	100.0 %	248	100.0 %	112	100.0 %	420	100.0 %
방문일		2013년 10월 ~ 2021년 5월		2012년 4월 ~ 2021년 8월		2012년 10월 ~ 2021년 2월		2012년 7월 ~ 2020년 8월		2013년 11월 ~ 2021년 2월	

전체 리뷰 개수는 2021.10.19. 일을 기준으로 홈페이지에서 확인되는 수치에 기초함  
국문 리뷰 개수는 CO.KR 도메인 상에서 '한국어'로 필터링한 결과를 따름



## 제 2 절 관광지 리뷰 데이터에 대한 빈도 분석

연구과제 1을 수행하기 위해 본 연구의 대상이 되는 설악산 국립공원, 한라산 국립공원, 순천만습지(순천만국가정원), 남이섬, 우도를 대상으로 국문 리뷰 데이터를 수집하고 불용어를 처리한 다음 명사를 추출하여 도출된 단어의 빈도 분석을 진행한 결과는 <표 4-2>, <표 4-3>, <표 4-4>, <표 4-5>, <표 4-6>와 같다. 별도의 가중치를 적용하지 않은 상태에서 빈도 분석을 하였으며, 관광지별 상위 출현하는 단어의 출현 빈도에 따라 1위부터 90위까지의 단어를 비율과 함께 제시하였다. n은 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타낸다. 설악산의 경우  $n = 1,997$ 로 요약되는데, 이는 수집된 설악산 국립공원의 리뷰 데이터에서 추출된 총 단어의 개수가 1,997개임을 의미한다. 한편, 빈도 분석 시 유사어는 고려하지 않았다. 한 예로, <표 4-2>에서 ‘계절’과 ‘철’, ‘등산’과 ‘등반’은 유사한 단어인데 이같은 유사어는 빈도 분석 시 고려하지 않았다.

빈도 분석에 대한 결과를 요약하면 다음과 같다. 공원·산 자연관광지인 「설악산 국립공원」에는 케이블카, 설악산, 단풍, 산, 바위 순으로, 「한라산 국립공원」에는 코스, 한라산, 등산, 백록담, 눈 순으로 빈도 높은 단어가 도출되었다. 바다·강 자연관광지인 「순천만습지(순천만국가정원)」에는 정원, 갈대, 순천만, 순천, 생태공원 순으로, 「남이섬」에는 남이섬, 섬, 배, 사람, 메타세콰이어 순으로, 「우도」에는 우도, 섬, 제주도, 땅콩, 바다 순으로 빈도 높은 단어가 도출되었다. 빈도 분석에 대한 결과를 워드클라우드로 시각화한 결과는 <그림 4-1>과 같다. 워드클라우드에 큰 글자로 표현된 단어는 상대적으로 빈도가 높은 단어라고 할 수 있으며, 이는 해당 관광지를 방문하는 관광객들이 리뷰에서 자주 언급한 단어들이다.

## 1. 설악산 국립공원 단어 도출 빈도 분석

<표 4-2> 설악산 국립공원 단어 도출 빈도 분석(n = 1,997)\*

순위	단어	비율	순위	단어	비율	순위	단어	비율
1	케이블카	4.457	31	사진	0.551	61	바람	0.250
2	설악산	3.906	32	절경	0.551	62	분위기	0.250
3	단풍	1.652	33	관광객	0.451	63	사찰	0.250
4	산	1.652	34	레스토랑	0.451	64	산세	0.250
5	바위	1.502	35	오색	0.451	65	서울	0.250
6	가을	1.302	36	자연	0.451	66	설악동	0.250
7	사람	1.302	37	강원도	0.401	67	신홍사	0.250
8	등산	1.102	38	눈	0.401	68	예약	0.250
9	코스	1.102	39	비	0.401	69	장관	0.250
10	경치	1.052	40	여름	0.401	70	주변	0.250
11	방문	1.052	41	장소	0.401	71	차비	0.250
12	시간	1.052	42	감상	0.351	72	하이킹	0.250
13	국립공원	1.002	43	계곡	0.351	73	한계령	0.250
14	정상	0.951	44	바다	0.351	74	현금	0.250
15	속초	0.851	45	산책	0.351	75	가격	0.200
16	울산바위	0.851	46	주차	0.351	76	국내	0.200
17	공기	0.81	47	철	0.351	77	기분	0.200
18	풍경	0.81	48	공원	0.300	78	나들이	0.200
19	입구	0.751	49	길	0.300	79	날	0.200
20	입장료	0.701	50	날씨	0.300	80	대표	0.200
21	폭포	0.701	51	대청봉	0.300	81	라면	0.200
22	버스	0.651	52	대피소	0.300	82	명소	0.200
23	최고	0.651	53	등반	0.300	83	부모님	0.200
24	추천	0.651	54	등산로	0.300	84	사계절	0.200
25	한국	0.651	55	절경	0.300	85	산입	0.200
26	구경	0.601	56	여행	0.300	86	설악	0.200
27	산행	0.601	57	우리나라	0.300	87	아침	0.200
28	겨울	0.551	58	경관	0.250	88	왕성	0.200
29	계절	0.551	59	도착	0.250	89	임계	0.200
30	봄	0.551	60	멋	0.250	90	전망대	0.200

n : 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타냄

\* : 해당 표는 연구과제 1에 대한 관광지별 분석 결과를 나타냄

## 2. 한라산 국립공원 단어 도출 빈도 분석

<표 4-3> 한라산 국립공원 단어 도출 빈도 분석(n = 3,028)\*

순위	단어	비율	순위	단어	비율	순위	단어	비율
1	코스	4.954	31	준비	0.495	61	체력	0.264
2	한라산	4.789	32	중간	0.495	62	초보자	0.264
3	등산	2.246	33	봄	0.462	63	한국	0.264
4	백록담	2.114	34	왕복	0.462	64	휴게소	0.264
5	눈	1.849	35	자연	0.429	65	단풍	0.231
6	겨울	1.750	36	출발	0.429	66	라면	0.231
7	정상	1.750	37	도착	0.396	67	소요	0.231
8	등반	1.552	38	매력	0.396	68	일정	0.231
9	성판악	1.519	39	방문	0.396	69	절경	0.231
10	시간	1.420	40	등산로	0.363	70	탐방	0.231
11	산	1.288	41	여름	0.363	71	구경	0.198
12	제주도	1.222	42	가을	0.330	72	국립공원	0.198
13	산행	1.123	43	눈꽃	0.330	73	노루	0.198
14	날씨	0.826	44	시작	0.330	74	바다	0.198
15	경치	0.793	45	입구	0.330	75	바위	0.198
16	오름	0.793	46	입산	0.330	76	새벽	0.198
17	제주	0.760	47	주차장	0.330	77	웅장	0.198
18	추천	0.760	48	필수	0.330	78	주변	0.198
19	관음사	0.694	49	기분	0.297	79	하늘	0.198
20	최고	0.661	50	사진	0.297	80	해	0.198
21	날	0.627	51	아침	0.297	81	가능	0.198
22	풍경	0.627	52	안개	0.297	82	감동	0.165
23	진달래	0.594	53	장관	0.297	83	개인	0.165
24	계절	0.561	54	하산	0.297	84	겨울철	0.165
25	길이	0.561	55	계단	0.264	85	경험	0.165
26	길	0.528	56	다리	0.264	86	고지	0.165
27	바람	0.528	57	밭	0.264	87	공기	0.165
28	사람	0.528	58	비	0.264	88	기준	0.165
29	대피소	0.495	59	사계절	0.264	89	꽃	0.165
30	아이젠	0.495	60	설경	0.264	90	나무	0.165

n : 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타냄

\* : 해당 표는 연구과제 1에 대한 관광지별 분석 결과를 나타냄

### 3. 순천만습지(순천만국가정원) 단어 도출 빈도 분석

<표 4-4> 순천만습지(순천만국가정원) 단어 도출 빈도 분석(n = 3,938)\*

순위	단어	비율	순위	단어	비율	순위	단어	비율
1	정원	3.504	31	구경	0.457	61	공간	0.229
2	갈대	2.869	32	최고	0.457	62	관광	0.229
3	순천만	2.793	33	풍경	0.457	63	관리	0.229
4	순천	2.539	34	스카이큐브	0.432	64	규모	0.229
5	생태공원	2.285	35	여행	0.432	65	사계절	0.229
6	전망	1.676	36	찍기	0.432	66	산	0.229
7	방문	1.549	37	박람회	0.406	67	우리나라	0.229
8	습지	1.346	38	장소	0.406	68	주차	0.229
9	밭	1.244	39	해	0.406	69	짬뽕어	0.229
10	시간	1.143	40	날씨	0.381	70	체험	0.229
11	가을	1.117	41	입장	0.381	71	경관	0.203
12	사람	1.092	42	중간	0.381	72	도시	0.203
13	일몰	1.067	43	가족	0.356	73	버스	0.203
14	추천	1.067	44	갯벌	0.356	74	입구	0.203
15	국가	0.889	45	기분	0.356	75	철새	0.203
16	사진	0.889	46	여유	0.33	76	푸른빛	0.203
17	자연	0.889	47	노을	0.305	77	한국	0.203
18	별빛축제	0.736	48	입장권	0.305	78	힐링	0.203
19	산책	0.736	49	거리	0.279	79	경치	0.178
20	여름	0.736	50	관람	0.279	80	데크	0.178
21	하늘	0.736	51	눈	0.279	81	라면	0.178
22	용산전망대	0.66	52	불거리	0.279	82	매력	0.178
23	겨울	0.609	53	장관	0.279	83	성인	0.178
24	봄	0.609	54	다리	0.254	84	시설	0.178
25	날	0.584	55	무료	0.254	85	왕복	0.178
26	걷기	0.559	56	바다	0.254	86	인상	0.178
27	나라	0.559	57	세계	0.254	87	일정	0.178
28	계절	0.533	58	입장료	0.254	88	저녁	0.178
29	꽃	0.508	59	가격	0.229	89	관광지	0.152
30	바람	0.508	60	경험	0.229	90	관람차	0.152

n : 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타냄

\* : 해당 표는 연구과제 1에 대한 관광지별 분석 결과를 나타냄

#### 4. 남이섬 단어 도출 빈도 분석

<표 4-5> 남이섬 단어 도출 빈도 분석(n = 1,782)\*

순위	단어	비율	순위	단어	비율	순위	단어	비율
1	남이섬	4.433	31	경험	0.449	61	기분	0.281
2	섬	1.740	32	기차	0.449	62	나라	0.281
3	배	1.684	33	드라마	0.449	63	낭만	0.281
4	사람	1.571	34	볼거리	0.449	64	동남아	0.281
5	메타세콰이어	1.515	35	서울	0.449	65	방법	0.281
6	자전거	1.403	36	여름	0.449	66	봄	0.281
7	가족	1.178	37	여유	0.449	67	사계절	0.281
8	가을	1.122	38	추억	0.449	68	산책로	0.281
9	데이트	1.122	39	관광지	0.393	69	재미	0.281
10	여행	1.122	40	도시락	0.393	70	타조	0.281
11	관광객	1.010	41	시간	0.393	71	펜션	0.281
12	사진	1.010	42	식사	0.393	72	평일	0.281
13	방문	0.898	43	와이어	0.393	73	경치	0.224
14	짚라인	0.898	44	음식점	0.393	74	고향	0.224
15	겨울	0.786	45	입장	0.393	75	기억	0.224
16	산책	0.786	46	잔디밭	0.393	76	나미나라	0.224
17	연인	0.786	47	친구	0.393	77	날	0.224
18	장소	0.786	48	코스	0.393	78	날씨	0.224
19	커플	0.730	49	가격	0.337	79	내부	0.224
20	거리	0.673	50	길	0.337	80	냄새	0.224
21	나무	0.673	51	도착	0.337	81	단위	0.224
22	분위기	0.617	52	음식	0.337	82	단체	0.224
23	입장료	0.617	53	자연	0.337	83	매력	0.224
24	추천	0.617	54	주말	0.337	84	모닥불	0.224
25	계절	0.561	55	주차	0.337	85	무료	0.224
26	닭갈비	0.561	56	청설모	0.337	86	바람	0.224
27	외국인	0.561	57	체험	0.337	87	사실	0.224
28	식당	0.505	58	최고	0.337	88	상업	0.224
29	주변	0.505	59	테마	0.337	89	순간	0.224
30	겨울연가	0.449	60	힐링	0.337	90	여자친구	0.224

n : 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타냄

\* : 해당 표는 연구과제 1에 대한 관광지별 분석 결과를 나타냄

## 5. 우도 단어 도출 빈도 분석

<표 4-6> 우도 단어 도출 빈도 분석(n = 6,126)\*

순위	단어	비율	순위	단어	비율	순위	단어	비율
1	우도	8.244	31	보트	0.424	61	선택	0.229
2	섬	2.612	32	관광	0.408	62	일주	0.229
3	제주도	2.106	33	대여	0.408	63	빛	0.212
4	땅콩	1.632	34	코스	0.408	64	장소	0.212
5	바다	1.551	35	기억	0.392	65	개인	0.196
6	버스	1.485	36	도로	0.392	66	돌	0.196
7	아이스크림	1.404	37	렌트카	0.359	67	라면	0.196
8	배	1.339	38	이동	0.359	68	매력	0.196
9	자전거	1.306	39	관광객	0.343	69	물	0.196
10	해변	1.273	40	여름	0.343	70	쓰레기	0.196
11	사람	1.241	41	항구	0.343	71	음식점	0.196
12	시간	1.175	42	모래	0.326	72	전기	0.196
13	추천	1.094	43	자연	0.326	73	전기자전거	0.196
14	스쿠터	1.028	44	해수욕장	0.326	74	절경	0.196
15	여행	0.979	45	등대	0.294	75	교통	0.180
16	방문	0.947	46	비	0.294	76	물가	0.180
17	오토바이	0.881	47	여유	0.294	77	숙박	0.180
18	투어	0.849	48	가격	0.278	78	에메랄드	0.180
19	제주	0.80	49	가족	0.278	79	요즘	0.180
20	바람	0.751	50	카페	0.278	80	힐링	0.180
21	차	0.751	51	해안	0.278	81	검멀레해변	0.163
22	전기차	0.702	52	길	0.261	82	경관	0.163
23	날씨	0.686	53	최고	0.261	83	눈	0.163
24	풍경	0.62	54	경험	0.245	84	맛집	0.163
25	구경	0.555	55	기분	0.245	85	바닷가	0.163
26	사진	0.506	56	도착	0.245	86	산호	0.163
27	중간	0.506	57	바이크	0.245	87	식당	0.163
28	우도봉	0.473	58	관광지	0.229	88	운전	0.163
29	성산항	0.457	59	구석구석	0.229	89	일정	0.163
30	경치	0.441	60	렌트	0.229	90	차량	0.163

n : 해당 관광지 전체 리뷰 데이터에서 추출된 전체 명사의 개수를 나타냄

\* : 해당 표는 연구과제 1에 대한 관광지별 분석 결과를 나타냄



설악산 국립공원



한라산 국립공원



순천만습지(순천만국가정원)



남이섬



우도



관광지 5개소 전체

<그림 4-1> 관광지 및 관광지 5개소 전체에 대한 워드클라우드 시각화 결과

## 6. 관광지 5개소에 대한 빈도 분석 결과

관광지 리뷰 데이터에 대한 빈도 분석을 통해 관광지별 상위에 출현하는 단어의 빈도에 따라 1위부터 90위까지의 단어를 순서대로 제시하였다(<표 4-2>, <표 4-3>, <표 4-4>, <표 4-5>, <표 4-6>). 분석 대상인 관광지가 5개소이므로 관광지별로 다섯 개의 표가 확인되었으며, 추출된 단어들을 대상으로 문서단어행렬(DTM : Document-Term Matrix)로 변환하여 관광지 5개소에서 최소 세 군데 이상의 관광지에서 공통으로 도출되는 단어를 확인하였다. 그 결과는 <표 4-7>와 같다.

‘경치, 날씨, 바람, 방문, 사람, 사진, 시간, 여름, 자연, 최고, 추천’의 11개 단어는 관광지 5개소 리뷰에서 빠지지 않고 공통으로 언급된 것으로 밝혀졌다. 관광지별로 방문객이 고려하는 선택속성에는 차이가 존재할 수 있으며, 리뷰 데이터에 자주 등장하는 단어와 그 비중은 다르게 나타날 수 있지만, 위와 같은 결과는 공원·산·바다·강 자연관광지를 대상으로 하는 방문객이라면 일반적으로 고려하는 사항임을 확인할 수 있었다.

다음으로 ‘가격, 가을, 겨울, 계절, 구경, 기분, 도착, 매력, 바다, 사계절, 여행, 장소, 주차, 코스, 풍경’의 15개 단어는 관광지 4개소의 리뷰에서 공통으로 언급된 것으로 밝혀졌다. 이 중에서 ‘가격’은 한라산 국립공원을 제외한 4개소 관광지에서 모두 발견되었는데, 이는 한라산 국립공원은 입장료<sup>11)</sup>가 무료로 운영되기 때문에 리뷰에서도 가격에 대한 단어가 자주 등장하지 않는 것으로 추론하였다. 현재 한라산 국립공원의 경우 시설 사용료에 대하여는 요금을 부과하지만, 별도의 입장료를 징수하지 않는다. 반면, 설악산 국립공원은 성인 3,500원, 순천만습지(순천만국가정원)는 성인 8,000원, 남이섬은 일반 고객 13,000원, 우도는 입도할 때를 기준으로 성인 왕복 10,500원

---

11) 한라산 국립공원 홈페이지(<https://jeju.go.kr/hallasan>)



의 입장료를 받는 것으로 확인된다. 따라서 가격은 한라산 국립공원을 제외한 관광지 4개소를 방문하는 방문객들이 리뷰 작성시 고려하는 사안이며 실제 리뷰에서도 등장하는 내용임을 확인하였다.

‘가을, 겨울, 계절, 사계절, 주차’는 우도를 제외한 관광지 4개소에서 언급된 주요 단어로 확인되었다. 우도 방문객들은 날씨에 대한 내용을 리뷰에서 자주 언급하는 한편, 계절에 대한 내용은 상대적으로 리뷰에 자주 언급하지 않는 것으로 확인된다. 이 외에도, ‘바다’는 남이섬을 제외한 4개소 관광지에서 공통적으로 도출되었다. 설악산 국립공원 주변에는 동해안이, 한라산 국립공원 주변에는 제주 앞바다가, 순천만습지(순천만국가정원) 주변에는 남해안이, 우도 주변에도 바다가 있는 지형적 특징에 근거하여 관광객들은 주변 관광지로 바다를 체험할 수 있다. 반면, 남이섬은 지리적 특성상 남이섬을 제외한 관광지 4개소와 대비하여 상대적으로 바다를 가까이하고 있지 않으므로 리뷰에서도 ‘바다’가 나타나지 않은 것으로 추론하였다. ‘풍경’ 또한 남이섬을 제외한 4개소 관광지에서 공통적으로 도출되었다. 국립국어원 표준대사전에 따르면 “풍경”은 ‘산이나 들, 강, 바다 따위의 자연이나 지역의 모습’을 뜻하는 단어이다. 위 관광지 4개소는 모두 자연적으로 조성된 관광지이나 남이섬은 인공적으로 조성된 섬이므로 남이섬에서만 ‘풍경’이 주요 단어에 포함되지 않은 것이라고 추론하였다.

이어서 ‘가족, 경관, 경험, 관광객, 관광지, 라면, 버스, 산책, 여유, 일정, 입구, 입장료, 장관, 절경, 주변, 중간, 한국, 힐링’의 18개 단어는 관광지 3개소의 리뷰에서 언급된 것으로 밝혀졌다. ‘가족, 경험, 관광지, 여유, 힐링’은 바다·강 자연관광지인 순천만습지(순천만국가정원), 남이섬, 우도에 공통으로 나타난 반면, 공원·산 자연관광지인 설악산 국립공원, 한라산 국립공원에서는 나타나지 않았다. 즉, 이 단어들은 바다·강 자연관광지에 대한 리뷰에서 발견되는 단어들로 확인된다. 바다·강 자연관광지를 방문하는 방

문객일수록 리뷰를 작성할 때 가족과 경험을 중심으로 여유와 힐링에 목적을 두고 방문하였음을 알 수 있다. 종합하면 위 과정을 거쳐서 관광지 5개소에 나타난 빈도 높은 주요 단어를 확인하였다. 추가로 각 관광지에 공통으로 언급된 단어들을 DTM으로 확인하였으며, 관광지 방문객이 리뷰 작성시 고려하는 내용들을 도출하였다.

<표 4-7> 관광지별 공통된 주요 빈출 단어\*

관광지	단어	설악산 국립공원	한라산 국립공원	순천만습지 (순천만 국가정원)	남이섬	우도
5개소	경치	○	○	○	○	○
	날씨	○	○	○	○	○
	바람	○	○	○	○	○
	방문	○	○	○	○	○
	사람	○	○	○	○	○
	사진	○	○	○	○	○
	시간	○	○	○	○	○
	여름	○	○	○	○	○
	자연	○	○	○	○	○
	최고	○	○	○	○	○
	추천	○	○	○	○	○
4개소	가격	○	-	○	○	○
	가을	○	○	○	○	-
	겨울	○	○	○	○	-
	계절	○	○	○	○	-
	구경	○	○	○	-	○
	기분	-	○	○	○	○
	도착	○	○	-	○	○

\* : 관광지 5개소를 대상으로 빈도수 높은 상위 단어 90개 중에서 공통으로 출현한 단어들을 확인하였음

<표 4-7> 관광지별 공통된 주요 빈출 단어\* (이어서)

관광지	단어	설악산 국립공원	한라산 국립공원	순천만습지 (순천만 국가정원)	남이섬	우도
4개소	매력	-	○	○	○	○
	바다	○	○	○	-	○
	사계절	○	○	○	○	-
	여행	○	-	○	○	○
	장소	○	-	○	○	○
	주차	○	○	○	○	-
	코스	○	○	-	○	○
	풍경	○	○	○	-	○
3개소	가족	-	-	○	○	○
	경관	○	-	○	-	○
	경험	-	-	○	○	○
	관광객	○	-	-	○	○
	관광지	-	-	○	○	○
	라면	○	-	-	○	-
	버스	○	-	○	-	○
	산책	○	-	○	○	-
	여유	-	-	○	○	○
	일정	-	○	○	-	○
	입구	○	○	○	-	-
	입장료	○	-	○	○	-
	장관	○	○	○	-	-
	절경	○	○	-	-	○
	주변	○	○	-	○	-
	중간	-	○	○	-	○
	한국	○	○	○	-	-
	힐링	-	-	○	○	○

\* : 관광지 5개소를 대상으로 빈도수 높은 상위 단어 90개 중에서 공통으로 출현한 단어들을 확인하였음

### 제 3 절 LDA 토픽모델링

연구과제 2를 수행하기 위해 본 연구의 대상이 되는 관광지 5개소를 대상으로 관광지별 리뷰에는 어떠한 내용으로 토픽이 형성되고 있는지 파악하기 위해 LDA 기반의 토픽모델링을 하였다. LDA 토픽모델링 시 지정할 수 있는 파라미터로는 토픽의 수, 샘플링 반복 횟수, 샘플링 결과에서 초기 결과 중 사용하지 않을 계산의 결과 수(burn-in), 문서 내 주제의 비율(alpha), 주제의 잠재 다항분포(eta) 값이 있다. 토픽모델링은 확률에 기반하는 분석 방법이므로 토픽의 수가 주어지지 않을 때에는 샘플링 반복 횟수와 토픽 수를 연구자가 조정하여 가장 적절하다고 판단되는 수로 정하여 사용한다(황해정 · 심혜린 · 최준호, 2016). 본 연구는 문서 내 주제의 비율을 0.05~0.1 사이의 값으로, 잠재 다항분포를 0.01~0.001 사이의 값으로, 샘플링 반복 횟수를 100회로 설정하고 토픽모델링을 수행하였다.

토픽모델링 시 단어에 가중치를 부여하는 방법으로는 TF-IDF 방식을 적용하였다. 이때 최소 출현 단어 빈도는 3으로 설정하였으며, 최대 출현 단어 빈도는 지정하지 않았다. 최소 출현 단어 빈도를 설정한 것은 빈도가 현저하게 떨어지는 단어는 중요도가 낮다고 판단되므로 토픽모델링에 포함하지 않기 위함이다. 또한 토픽모델링 시 TF-IDF 가중치가 부여될 때, 매우 빈번하게 도출되는 단어는 상대적으로 낮은 가중치가 적용되어 토픽모델링이 수행될 것이므로 최대 출현 단어 빈도의 수를 지정하지 않았다.

토픽모델링 시, 토픽의 적정 개수로 5~8 사이의 값을 지정하였을 때 비교적 각각의 토픽이 상호독립적이면서 주제를 잘 나타낸다고 할 수 있었다. FsQCA 분석에 사용될 토픽의 수를 고려한다면, 토픽의 수가 5일 때 32개의 원인조건조합이, 토픽의 수가 6일 때 64개의 원인조건조합이, 토픽의 수가 7일 때 128개의 원인조건조합이, 토픽의 수가 8일 때 256개의 원인조건조합

의 경우의 수가 존재할 수 있다. 따라서, 지나치게 많은 토픽의 수를 지정하고 토픽모델링을 하면 FsQCA 분석 결과에 따른 결과 해석이 용이하지 않을 수 있다. 그러므로 관광지별 도출하고자 하는 적정한 토픽의 수를 6개로 고정하고 토픽모델링을 실시하였다.

토픽모델링은 공원·산 자연관광지인 설악산 국립공원과 한라산 국립공원, 바다·강 자연관광지인 순천만습지(순천만국가정원), 남이섬, 그리고 우도를 대상으로 각각 실시하였다. 관광지 5개소에 대한 토픽모델링을 마친 후에는 토픽명을 부여하고 토픽을 구성하는 주요 단어를 비중이 높은 순서대로 정리하여 표에 제시하였다. 토픽명은 분석을 진행할 때 자동으로 부여되는 부분이 아니므로 연구자가 분석 결과를 보고 토픽을 구성하는 단어와 단어들 간의 연관성을 고려하여 토픽명을 정하였다. 텍스트 데이터 특성상 하나의 토픽을 나타내는 주요 단어가 다른 토픽에도 발견되는 현상이 나타날 수 있다. 예를 들어, 1번 토픽은 관광자원에 관한 것이고 2번 토픽은 관광물가에 대한 내용일 때, 1번 토픽에서 ‘관광’이라는 단어가 1번 토픽을 설명하는 주요 단어로 도출됨과 동시에 2번 토픽에서도 ‘관광’이라는 단어가 중복으로 발견될 수 있는 것이다. 이러한 특징을 참고하여서, 토픽별 할당된 단어의 조합을 보고 가장 적합하다고 판단되는 토픽명을 연구자가 부여하고 해당 토픽 안에 발견된 단어들을 함께 제시하였다. 그 결과, 관광지별로 도출되는 토픽을 살펴보면서 특정 관광지에만 나타나는 독특한 토픽에는 무엇이 있는지 기술하였으며, 최소한 두 개 이상의 관광지에서 공통으로 도출되는 토픽에는 무엇이 있는지도 함께 관찰하였다.

## 1. 설악산 국립공원 토픽모델링

설악산 국립공원의 토픽모델링 결과는 <표 4-8>와 같으며 총 6개의 토

픽이 도출되었다. 도출된 토픽을 살펴보면 다음과 같다.

첫 번째 토픽은 ‘케이블카 체험’이다. 케이블카는 설악산 국립공원 소공원 내에 위치하며, 해발 700m 정상인 권금성 구간을 왕복 운행하는 시설이다. 케이블카 이용 시, 권금성 정상에서 동해바다와 울산바위, 토왕성 폭포 등의 다양한 경치를 조망할 수 있는 것으로 알려져 있다<sup>12)</sup>. 케이블카 체험을 나타내는 주요 단어로는 케이블카뿐만 아니라, 이를 통해 시각적으로 경험할 수 있는 경치, 감상 절경, 단풍 등의 단어가 포함되었다.

두 번째 토픽은 ‘사계절 매력성’이다. 이를 직접적으로 나타내는 단어로 계절, 가을 등의 단어가 포함되었다. 계절을 나타내는 단어뿐만 아니라 특정 계절에서 관찰되는 풍경(예 : 가을에는 단풍, 겨울에는 눈과 설경)이 포함되어 있어 사계절 매력성으로 토픽명을 부여하였다.

세 번째 토픽은 ‘입장료 및 경비’이다. 이를 나타내는 단어로 산행, 입장료, 차비, 비용, 현금 등이 포함되었다. 케이블카는 별도의 시설물로 이용 시 유료 입장료가 발생한다. 또한, 일부 사찰(문화재구역)은 문화재 관람료 항목으로 비용이 발생하는 것으로 보고 경비에 포함하였다. 실제로 한 신문 기사에 따르면 설악산 국립공원의 신흥사는 문화재 관람료를 징수하고 있는 것으로 보도되었다(임지혜, 2021). 그러므로 사찰은 관광객 입장에서 경비와 관련된 단어로 언급될 수 있는 것이다.

네 번째 토픽은 ‘울산바위 명소’이다. 문화재청에 따르면 울산바위는 2013년 3월 11일에 명승 제100호로 지정되었다. 설악산에는 다양한 볼거리가 존재하지만 리뷰에서는 울산바위, 정상, 명소, 폭포, 전망대 등의 울산바위를 중심으로 하는 관광 명소 및 볼거리와 관련된 단어들이 도출되었다.

다섯 번째 토픽은 ‘교통수단 접근성’이다. 이를 나타내는 단어로 버스, 주차, 방문, 입구, 진입 등의 단어가 포함되었다. 케이블카가 설악산 국립공원

---

12) 설악 케이블카 홈페이지(<https://www.sorakcablecar.co.kr/about/intro>)

소공원 내에서 권금성 구간을 왕복 운행하면서 접근성을 용이하게 하는 시설이라면, 여기서의 접근성은 설악산이 아닌 외부 지역에서 설악산 입구까지 진입할 수 있는 각종 경로 및 교통수단과 관련된 단어들을 포함한다.

여섯 번째 토픽은 ‘등산 탐방로 코스’이다. 이를 나타내는 단어로 코스, 대피소, 정상, 등산로 등이 포함되었다. 설악산 국립공원에는 용소폭포 코스, 울산바위 코스, 권금성 코스, 금강굴 코스, 양폭 코스, 백담사 코스, 수렴동 코스, 남교리 코스, 대승폭포 코스, 대청봉 코스, 공룡능선 코스 등의 다양한 탐방로 코스가 존재한다. 이 중에서 대청봉, 한계령 등의 단어는 등산 탐방로 코스와 연관 있는 단어로 확인된다.

<표 4-8> 설악산 국립공원 토픽모델링 결과\*

토픽명 (관광지 선택속성)	토픽을 구성하는 주요 단어				
케이블카 체험	케이블카	주차	왕복	관광객	경치
	감상	설악산	날씨	절경	단풍
사계절 매력성	계절	가을	경치	등반	구경
	눈	해	설경	단풍	장관
입장료 및 경비	산행	입장료	차비	케이블카	입구
	중간	비용	현금	사찰	도착
울산바위 명소	울산바위	정상	명소	폭포	전망대
	왕복	신홍사	미시령	절경	관광객
교통수단 접근성	버스	주차	예약	방문	입구
	진입	현금	산행	등산로	마을
등산 탐방로 코스	코스	대피소	대청봉	정상	한계령
	온천	설경	등산로	사진	식당

\* : 해당 표는 연구과제 2에 대한 관광지별 분석 결과를 나타냄 (토픽 개수 : 6개, 반복 수행 횟수 : 100회)

## 2. 한라산 국립공원 토픽모델링

한라산 국립공원의 토픽모델링 결과는 <표 4-9>와 같으며 총 6개의 토픽이 도출되었다. 도출된 토픽을 살펴보면 다음과 같다.

첫 번째 토픽은 ‘성판악 탐방로 코스이다’이다. 성판악 탐방로 코스는 한라산 국립공원 홈페이지<sup>13)</sup>에 소개된 대표 탐방로 코스 중 하나이다. 한라산 동쪽 코스인 성판악 탐방로는 관음사 탐방로와 더불어 한라산 정상인 백록담을 오를 수 있는 것으로 알려져 있다. 이를 나타내는 단어로 성판악, 코스, 시간, 계단, 등산로, 대피소, 길이 등의 단어들과, 성판악 탐방로 코스를 선택했을 때 관찰할 수 있는 백록담, 진달래밭 등의 장소를 나타내는 단어들이 포함되었다.

두 번째 토픽은 ‘사계절 매력성’이다. 이를 나타내는 단어로 여름, 가을, 단풍, 사계절 등의 계절과 관련된 단어와 매력, 꽃, 구경 등의 매력성을 나타내는 단어들이 포함되었다.

세 번째 토픽은 ‘관음사 탐방로 코스이다’이다. 한라산 북쪽코스인 관음사 탐방로는 성판악 탐방로와 더불어 한라산 정상인 백록담을 오를 수 있는 탐방로이다. 전문 산악인들은 물론, 성판악 코스 탐방객들도 하산할 때 주로 관음사 탐방로 코스를 이용하는 것으로 알려져 있다(한라산 국립공원 홈페이지). 해당 토픽을 나타내는 단어로는 관음사, 등반, 길, 코스, 아이젠, 고지, 정상 등의 등산 탐방로와 관련된 단어들이 포함되었다.

네 번째 토픽은 ‘위치 및 접근성’이다. 이를 나타내는 단어로 주차장, 길, 입구, 휴게소, 계단 등의 단어가 포함되었다. 앞서 제시된 단어들이 공간 또는 장소적인 접근성을 나타내는 단어라면, 시간, 새벽, 여행 등은 시간적인 접근성을 나타내는 단어로 유추된다.

---

13) 한라산 국립공원 홈페이지(<https://visithalla.jeju.go.kr/contents/contents.do?id=61>)



다섯 번째 토픽은 ‘자연 볼거리’이다. 이를 나타내는 단어로 경치, 안개, 풍경, 노루, 오름, 눈꽃, 단풍 등의 단어들이 포함되었다. 이는 한라산 국립공원에서 관찰되는 것으로 자연적인 특성에 기인한 볼거리라고 할 수 있다.

여섯 번째 토픽은 ‘편의시설 및 먹거리’이다. 편의시설을 나타내는 단어로는 화장실, 매점, 대피소 등이 포함되었고, 먹거리를 나타내는 단어로는 간단하게 식사 해결을 돕는 컵라면, 김밥 등의 단어가 포함되었다. 실제로 한라산 국립공원에서는 환경 보전을 위해 일회용 도시락 반입을 금지하고 있으나, 김밥과 햄버거 등의 간편한 음식 반입은 허용하고 있다(한라산 국립공원 홈페이지).

<표 4-9> 한라산 국립공원 토픽모델링 결과\*

토픽명 (관광지 선택속성)	토픽을 구성하는 주요 단어				
성판악 탐방로 코스	성판악 진달래(밭)	코스 계단	시간 등산로	관음사 대피소	백록담 길이
사계절 매력성	여름 제주	가을 꽃	자연 등산	단풍 구경	매력 사계절
관음사 탐방로 코스	관음사 아이젠	등반 백록담	길 체력	코스 고지	풍경 정상
위치 및 접근성	주차장 새벽	길 휴게소	국립공원 계단	입구 초보자	시간 여행
자연 볼거리	경치 노루	탐방 오름	안개 눈꽃	풍경 어승생악	병풍(바위) 단풍
편의시설 및 먹거리	화장실 중간	컵라면 김밥	부근 정비	매점 지점	대피소 가족

\* : 해당 표는 연구과제 2에 대한 관광지별 분석 결과를 나타냄 (토픽 개수 : 6개, 반복 수행 횟수 : 100회)

이로써 공원·산 자연관광지인 설악산 국립공원과 한라산 국립공원에 대하여 토픽모델링을 실시하고, 각 토픽을 구성하는 주요 단어들을 확인하

였다. 이 두 개의 관광지를 비교·대조하여 살펴본 결과는 다음과 같다.

‘케이블카 체험’과 ‘입장료 및 경비’는 설악산 국립공원에서만 발견되는 토픽으로 나타났다. 케이블카를 운영하지 않으면서 국립공원의 입장료가 무료인 한라산 국립공원과는 다르게, 설악산 국립공원은 비용을 지불하고 케이블카 체험을 할 수 있으면서 국립공원에 대한 입장료를 징수한다. 따라서, 설악산 국립공원과 한라산 국립공원의 다른 운영방식으로 인해 이러한 결과가 토픽모델링에서도 반영된 것으로 추론된다.

반면, ‘편의시설 및 먹거리’는 한라산 국립공원에서만 발견되는 토픽이다. 설악산 국립공원의 방문객은 케이블카와 국립공원 입장료에 대하여 경비로 지출하는 비용이 발생한다. 그러나, 한라산 국립공원의 방문객들은 시설 사용료(주차시설, 야영장, 샤워장)를 제외한 입장료에 대한 비용 지출이 사실상 드물다. 따라서, 설악산 국립공원의 관광객들은 주로 입장료와 케이블카 이용을 위해 비용을 지출한다는 내용으로 리뷰를 작성하는 것과는 다르게, 한라산 국립공원의 방문객들은 등산로 탐방 또는 편의시설을 이용하면서 먹거리나 필요한 물품 구매 시에 비용을 지출하는 것으로 밝혀졌다.

연구자가 동일한 토픽명을 부여하더라도 각 토픽을 구성하는 주요 단어는 리뷰 데이터 특성상 다르게 나타날 수 있다. 그러나 이 두 관광지에서 유사하게 공통으로 발견되는 토픽도 존재하였다.

첫 번째는 ‘사계절 매력성’이고, 두 번째는 ‘등산 탐방로 코스’이다. 한라산 국립공원의 경우, 성판악/관음사 탐방로 코스라는 토픽으로 도출되었다. 세 번째는 ‘접근성’이다. 설악산 국립공원에서는 교통수단 접근성으로, 한라산 국립공원에서는 위치 및 접근성으로 매우 유사한 토픽이 도출되었다. 이외에도, 설악산 국립공원에서는 ‘울산바위 명소’라는 토픽으로, 한라산 국립공원에서는 ‘자연 볼거리’라는 토픽으로 자연적인 볼거리와 관련된 토픽이 도출된 것을 확인할 수 있었다.

### 3. 순천만습지(순천만국가정원) 토픽모델링

순천만습지(순천만국가정원)의 토픽모델링 결과는 <표 4-10>와 같으며 총 6개의 토픽이 도출되었다. 도출된 토픽을 살펴보면 다음과 같다.

첫 번째 토픽은 ‘사계절 매력성’이다. 이를 나타내는 단어로 봄, 겨울, 여름 등의 계절을 나타내는 단어들이 포함되었다. ‘갈대’는 계절을 직접적으로 나타내는 단어가 아니지만, 매년 10월~11월 경 열리는 순천만갈대축제를 예상하면, 계절을 나타내는 가을과 연관된 주요 단어임을 추론할 수 있다.

두 번째 토픽은 ‘볼거리 및 이동수단’이다. 볼거리를 나타내는 단어로는 정원, 구경, 용산전망대, 일몰, 풍경, 정원 등이 포함되었다. ‘용산전망대’는 순천만습지를 대표하는 전망대로, 과거 순천만 갯벌을 한눈에 바라볼 수 있는 장소이다. 여기에서 관람차, 갈대열차, 스카이크्यू브는 볼거리 및 이동수단의 하나이다. 관람차<sup>14)</sup>는 순천만국가정원 내에서 운행하는 친환경 전기 관람차로 꿈틀정원과 중국정원 앞에서 출발하여 약 25분간 편안하게 정원을 감상할 수 있도록 마련된 하나의 이동수단이다. 순천만습지를 방문하는 관광객이 스카이크्यू브에서 내리면 바로 순천만습지에 도달하는 것은 아니므로, 약 1km 정도의 거리를 갈대열차에 탑승하여 이동할 수 있다. 스카이크्यू브<sup>15)</sup>는 자동차 배기가스 및 교통 혼잡으로부터 세계 5대 습지인 순천만을 보호하기 위해 건설된 것으로, 전기를 동력원으로 사용하는 무인자동운전시스템이다. 스카이크्यू브는 아름다운 동천과 정원을 배경으로 지상 3.5m에서 10m의 높이의 레일을 따라 운행하는 이동수단이다. 이 내용을 종합하면 두 번째 토픽에는 관광객들이 이동수단을 경험하면서 함께 관람할 수 있는 볼거리가 주요 단어로 도출되었다고 할 수 있다.

세 번째 토픽은 ‘접근 편리성’이다. 이를 나타내는 단어로 정원, 교통, 주

14) 순천만습지 홈페이지(<https://scbay.suncheon.go.kr/garden/0002/0005>)

15) 스카이크्यू브 홈페이지(<http://www.skycube.co.kr/skycube/index.php>)

차, 연결, 길 등의 단어가 포함되었다. 이러한 단어를 통해 관광객들이 접근 편리성에 대한 내용을 리뷰에서 언급하고 있음을 확인하였다.

네 번째 토픽은 ‘관광 행사’이다. 이를 나타내는 단어로 박람회, 입장, 세계, 둘러보기, 관광 등이 포함되었다. 여기에는 2013 순천만 국제정원 박람회를 포함하여 각종 박람회 및 축제, 행사 등과 관련된 단어들이 도출되었다. 토픽모델링 시 주요 단어로 도출되지는 않았지만, 매해 가을경 진행되는 순천만 갈대축제와 앞으로 열릴 2023 순천만 국제정원 박람회를 생각한다면 관광 행사에 대한 토픽도 리뷰에서 중요한 부분을 차지하는 것이다.

다섯 번째 토픽은 ‘자연관광 매력성’이다. 이를 나타내는 단어로 일몰, 철새, 찍기, 저녁, 여유 등의 단어들이 포함되었다. 특히, 일몰은 순천만습지의 관람 시간을 결정하는 중요한 요소이다. 실제 순천만습지의 관람 시간은 오전 8시부터 일몰시까지로 규정되어 있다. 따라서, 일몰은 순천만습지를 대표하는 자연관광 매력성과 관련된 단어로 간주하였다. 한편, 전망대, 사진, 사람, 철새, 저녁 등이 주요 단어로 함께 제시된 것은 전망대에 올라가 저녁 시간대에 사진을 찍기도 하며 사람들과 여유로운 시간을 가지고 일몰을 감상할 수 있기 때문인 것으로 추론하였다. 여기에는 일몰을 중심으로 자연관광에 대한 매력성을 느낄 수 있는 단어들이 다수 포함된다.

여섯 번째 토픽은 ‘먹거리 및 체험학습’이다. 이를 나타내는 단어로 짬뽕, 게, 갯벌, 습지 등의 단어들이 포함되었다. 짬뽕어와 게는 순천의 대표 먹거리를 나타내는 단어이기도 하지만, 인근에 있는 순천만 짬뽕어 마을 또는 순천만 갯벌 체험 등을 통해 체험할 수 있는 요소이기도 하다. 따라서 이러한 단어들은 먹거리 및 체험학습과 관련된 단어인 것으로 추론하였다.

<표 4-10> 순천만습지(순천만국가정원) 토픽모델링 결과\*

토픽명 (관광지 선택속성)	토픽을 구성하는 주요 단어				
사계절 매력성	봄 스카이큐브	겨울 갯벌	여름 갈대	노을 생태	장관 하늘
볼거리 및 이동수단	정원 서문	구경 풍경	용산전망대 갈대(열차)	일몰 스카이큐브	관람차 거리
접근 편리성	정원 여수	입장료 연결	무료 관광	교통 길	주차 습지
관광 행사	박람회 정원	찍기 돌러보기	입장 순천만	산책 경치	세계 관광
자연관광 매력성	일몰 철새	시간 찍기	전망대 저녁	사진 여유	사람 산책로
먹거리 및 체험학습	짬뽕어 겨울	자연 습지	게 전망대	갯벌 순천	계절 일몰

\* : 해당 표는 연구과제 2에 대한 관광지별 분석 결과를 나타냄 (토픽 개수 : 6개, 반복 수행 횟수 : 100회)

#### 4. 남이섬 토픽모델링

남이섬의 토픽모델링 결과는 <표 4-11>와 같으며 총 6개의 토픽이 도출되었다. 도출된 토픽을 살펴보면 다음과 같다.

첫 번째 토픽은 ‘접근 편리성’이다. 이를 나타내는 단어로 주차, 썰라인(썰와이어), 방문, 자전거, 배 등의 단어가 포함되었다. 썰라인과 배는 남이섬 방문객들이 주로 탑승하는 이동수단이며, 접근 편리성을 나타낸다.

두 번째 토픽은 ‘관광 명소’이다. 남이섬은 드라마 겨울연가의 촬영지로 잘 알려져 있다. 이를 나타내는 단어로 드라마, 촬영, 겨울연가, 가을, 커플, 걷기 등의 단어들이 포함되었다.

세 번째 토픽은 ‘데이트 코스’이다. 이를 나타내는 단어로 커플, 은행나

무, 코스, 자전거, 낭만, 테마, 데이트 등의 관련 단어들이 포함되었다. 표본의 특성을 살펴보았을 때, 리뷰 작성자의 동반자 유형으로는 연인과 방문하는 비율(31.3%) 가장 높은 것으로 나타났다(<표 4-1>). 따라서 데이트 코스가 주요 토픽으로 도출된 것은 이러한 방문자 유형의 특징에 근거한 결과라고 할 수 있다. 즉, 연인과 방문하는 비율이 높으므로 이와 관련된 토픽인 데이트 코스가 리뷰에서도 생성되는 것이다.

네 번째 토픽은 ‘숙박시설’이다. 이를 나타내는 단어로 정관루(호텔), 시설, 예약, 도착, 장소 등의 단어가 포함되었다. 숙박시설은 본 연구의 대상이 되는 다른 관광지에서 도출되지 않은 남이섬만의 유일한 토픽이다.

다섯 번째 토픽은 ‘음식 및 편의시설’이다. 이를 나타내는 단어로 음식, 코스, 분위기, 맛, 식사 등의 단어가 도출되었다. 한편, 상업이라는 단어가 포함된 것은 실제 리뷰를 확인한 결과 일부 시설이 상업적인 형태로 운영되고 있음을 지적인 주관적인 소견이 담긴 리뷰 내용이 있어 이러한 단어가 도출된 것으로 추론하였다.

여섯 번째 토픽은 ‘자연관광 매력성’이다. 이를 나타내는 단어로 가을, 섬, 타조, 사계절, 산책로, 나무, 추억 등의 단어가 포함되었다.

<표 4-11> 남이섬 토픽모델링 결과\*

토픽명 (관광지 선택속성)	토픽을 구성하는 주요 단어				
접근 편리성	주차 자전거	식사 배	짚라인 관광지	체험 섬	방문 나무
관광 명소 (드라마촬영)	드라마 휴일	촬영 경관	여유 가을	잔디밭 커플	겨울연가 건기
데이트 코스	커플 메타세콰이어	은행나무 낭만	코스 테마	자전거 데이트	외국인 사진
숙박시설	정관루(호텔) 한국	시설 햇살	예약 테마	도착 장소	평일 주말
음식 및 편의시설	음식 식사	코스 외국인	분위기 방문	맛 관광객	세계 상업
자연관광 매력성	가을 기분	섬 산책로	평일 종일	타조 나무	사계절 추억

\* : 해당 표는 연구과제 2에 대한 관광지별 분석 결과를 나타냄 (토픽 개수 : 6개, 반복 수행 횟수 : 100회)

## 5. 우도 토픽모델링

우도의 토픽모델링 결과는 <표 4-12>와 같으며 총 6개의 토픽이 도출되었다. 도출된 토픽을 살펴보면 다음과 같다.

첫 번째 토픽은 ‘개별 이동수단’이다. 이를 나타내는 단어로 스쿠터, 이동, 전기자전거, 대여, 일주, 바이크 등의 단어가 도출되었다. 우도를 관광하기 위해서는 관광객들이 다양한 이동수단을 선택할 수 있다. 해당 토픽에는 이동수단 중에서도 스쿠터, 전기자전거, 바이크 등의 개인이 대여하거나 소지할 수 있는 이동수단이 포함되어 개별 이동수단이라고 명하였다.

두 번째 토픽은 ‘대중교통 이동수단’이다. 앞서 제시된 개별 이동수단과는 다르게 이를 나타내는 단어로 관광객, 버스, 투어, 순환버스, 도로 등의

단어가 포함되었다. 우도 순환버스는 개별 이동수단이 준비되지 않은 관광객이 우도를 관광할 때 이용할 수 있는 하나의 수단이다. 이로써 우도는 이동수단이 개별 이동수단과 대중교통 이동수단으로 나뉘는 것을 확인하였다.

세 번째 토픽은 ‘지역 음식’이다. 이를 나타내는 단어로 땅콩, 아이스크림 등이 포함되었다. 우도 땅콩은 제주의 특산물로 알려져 있다. 땅콩을 활용한 아이스크림이 우도를 나타내는 대표 음식으로 확인되었으며, 주요 단어로 함께 제시하지 않았지만, 햄버거, 해물 칼국수, 해물 라면 등이 해당 토픽을 나타내는 단어로 함께 등장하는 것을 확인하였다. 여기에 해변, 제주도, 코스, 추천, 우도, 경치, 구경 등의 단어가 지역 음식과 함께 주요 단어로 포함된 것은 우도에서 지역 음식을 먹으며 함께 즐길 수 있는 볼거리가 형성되어 있는 관광지이기 때문으로 추론하였다. 한편, 쓰레기가 지역 음식을 나타내는 토픽에 포함된 것은 정확한 근거를 확인할 수 없으나 음식 포장으로 인해 버려지는 쓰레기, 또는 우도 관광 시 목격되는 쓰레기로 인해 해당 단어가 지역음식과 관련된 주요 단어에 포함된 것으로 추측된다.

네 번째 토픽은 ‘자연관광 매력성’이다. 이를 나타내는 단어로 해변, 바다, 빛, 등대, 에메랄드, 산호 등이 포함되었다. 이러한 단어들은 우도를 관광할 때에 관광객들이 시각적으로 보고 경험할 수 있는 자연관광 매력성에 해당한다.

다섯 번째 토픽은 ‘날씨 요소’이다. 이를 나타내는 단어로 날씨, 하늘, 비, 눈, 해 등의 단어가 도출되었다. 우도는 날씨에 따라 입도에 영향을 많이 받는 곳으로 날씨는 관광객들이 중요하게 고려하는 사안이며 리뷰에서도 등장하는 토픽이다. 우도에 대한 리뷰 내용에는 날씨에 관한 요소가 두드러지게 나타나는 반면, 계절 요소는 토픽모델링 결과에서 발견되지 않았다. 그 이유는 우도에 대한 접근성 및 관광이 계절 요소보다는 날씨 요소에 실제적인 영향을 받고 있음을 추론해볼 수 있다.

여섯 번째 토픽은 ‘접근 편리성’이다. 이를 나타내는 단어로 배, 시간



(표), 차로, 이동, 입도, 도착, 렌트카 등의 단어가 도출되었다. 우도는 섬 관광지로서 이에 접근하기 위해서는 제주도의 성산항 또는 종달항에서 우도와 이어지는 배편을 이용하여야 한다. 따라서 배, 입도 등의 단어가 도출된 것은 우도의 접근 편리성을 나타내기 위해 당연히 고려되는 요소이다.

<표 4-12> 우도 토픽모델링 결과\*

토픽명 (관광지 선택속성)	토픽을 구성하는 주요 단어				
개별 이동수단	스쿠터	방문	이동	추천	힐링
	전기자전거	코스	대여	일주	바이크
대중교통 이동수단	관광객	버스	투어	여유	기억
	순환버스	설명	전기차	도로	필수
지역 음식	땅콩	아이스크림	해변	제주도	코스
	추천	우도	쓰레기	경치	구경
자연관광 매력성	해변	바다	빛	등대	에메랄드
	추억	산호	버스	배	당일치기
날씨 요소	날씨	투어	하늘	사진	우도
	비	눈	해	올레길	돌
접근 편리성	배	시간	차로	이동	입도
	도착	날씨	바다	렌트카	해변

\* : 해당 표는 연구과제 2에 대한 관광지별 분석 결과를 나타냄 (토픽 개수 : 6개, 반복 수행 횟수 : 100회)

이로써 바다·강 자연관광지인 순천만습지(순천만국가정원), 남이섬, 우도에 대한 토픽모델링을 실시하고, 각 토픽을 구성하는 주요 단어들을 확인하였다. 이 세 개의 관광지를 비교·대조하여 살펴본 결과는 다음과 같다.

‘사계절 매력성’과 ‘관광 행사’는 순천만습지(순천만국가정원)에서 발견되는 독특한 토픽으로 나타났다. 자연관광지에서는 계절적인 요소를 배제할 수 없으나, 사계절이라는 토픽으로 계절적 요소가 두드러진 건 바다·강 자연관광지 중에서 순천만습지(순천만국가정원)가 유일하다. 또한, 박람회와

전시회등의 관광 행사 유치로 인해 ‘관광 행사’가 순천만습지(순천만국가정원)에서만 드러나는 독특한 토픽인 것으로 추론하였다.

한편, ‘관광 명소’, ‘데이트 코스’, 그리고 ‘숙박시설’은 남이섬에서 발견되는 독특한 토픽으로 나타났다. 남이섬은 드라마 겨울연가의 촬영지로 알려져 있는데, 관광객들에게는 관광 명소가 하나의 중요한 부분이므로 이와 관련된 토픽이 생성된 것으로 유추하였다. 또한, 남이섬의 경우 리뷰 작성자에 대한 동반자 유형을 확인한 결과 남이섬은 연인과 방문하는 비율이 높은 것으로 나타났다. 따라서 연인과 관련된 ‘데이트 코스’가 남이섬을 나타내는 주요 토픽으로 도출되었다고 추론하였다. ‘숙박시설’은 남이섬에서만 발견되는 독특한 토픽이다. 연구자가 직접 남이섬 대표 고객센터로 문의한 결과, 남이섬 내의 숙박시설은 호텔 정관루 한 군데가 유일하였다. 따라서 관광객 입장에서 1박 이상의 여행을 계획할 때 숙박시설에 대한 고려는 중요하므로 이러한 부분이 토픽에 반영되었을 것이라고 추론하였다.

이동수단과 관련된 토픽이 ‘개별 이동수단’과 ‘대중교통 이동수단’으로 나뉘어 도출된 것은 우도에서만 발견되는 독특한 토픽이다. 우도 안에서의 이동수단이 개별 이동수단과 대중교통 이동수단으로 나뉘는 것은 이동수단에 있어서도 다양한 선택지가 존재함을 의미한다. 우도를 관광할 때는 대중교통을 이용하는 대신, 스쿠터나 전기자전거 등의 개별 이동수단을 이용할 수 있다. 즉, 관광객 입장에서는 이동수단으로 크게 두 가지 선택지가 존재하는 것이다. 남이섬은 0.46km<sup>2</sup>로 비교적 작은 크기의 섬인 반면, 우도는 전체 면적이 6.18km<sup>2</sup>로 도보로만 이동하는 것에 제약이 따를 수 있다. 따라서 우도에서는 교통수단이 토픽의 중요한 부분을 담당한다. 이와 함께 ‘날씨 요소’ 또한 우도에서만 나타나는 독특한 토픽으로 확인된다. 특이하게 이 안에는 계절적인 요소가 거의 포함되지 않으면서 날씨에 대한 단어들로 토픽이 구성되었다. 즉, 이러한 결과는 우도를 관광할 때에 계절보다는 현지

의 날씨 상황이 더 중요하다는 것으로 추론할 수 있다. 우도에 입도하기 위해서는 선착장에 방문해야 하는데, 선착장에 게시된 안내문을 읽어보면 계절에 대한 안내보다는 조석(간만)이나 기상 상태 등으로 인해 운항이 변경될 수 있음을 고지하는 날씨와 관련된 내용들이 주로 확인된다.

‘접근 편리성’, ‘자연관광 매력성’, ‘음식과 관련된 토픽(먹거리 및 체험학습, 음식 및 편의시설, 지역음식)’은 바다·강 자연관광지 3개소에서 공통으로 도출된 토픽이다. 이 세 개의 토픽들은 리뷰에서 관광지를 불문하고 자주 등장하며 토픽 형성에 중요한 역할을 하고 있다. 나아가, 공원·산 자연관광지와 바다·강 자연관광지에서 모두 도출된 토픽도 존재하였다. 즉, 관광지 5개소에서 공통으로 도출된 토픽은 관광지의 ‘매력성’과 ‘접근 편리성’이다. 매력성으로는 공원·산 자연관광지에서 사계절 매력성이, 바다·강 자연관광지에서 자연관광 매력성이 공통으로 발견되었다. 접근성으로는 공원·산 자연관광지에서 교통수단 접근성과 위치 및 접근성이, 바다·강 자연관광지에서 접근 편리성이 도출되었다. 이를 통해 본 연구의 대상이 되는 관광지 리뷰 데이터에서 나타나는 토픽들을 제시하였으며, 특정 관광지에서만 나타나는 독특한 토픽과 공통으로 나타나는 토픽들을 확인하였다.

## 제 4 절 FsQCA(퍼지셋 질적비교분석)

연구과제 3을 수행하기 위해 관광지 5개소에 대한 LDA 토픽모델링 결과를 활용하여 FsQCA 분석을 하였다. FsQCA 분석을 본 연구에 적용하려는 이유는 다음과 같다. 예를 들어 5점이라는 관광지 만족도가 부여된 리뷰가 10개 있다고 가정했을 때, 이 10개의 리뷰 내용이 모두 동일한 토픽으로 구성되었다고 단정짓기는 매우 어려울 것이다. 그 이유는 각각의 리뷰 내용을 구성하는 토픽, 그리고 해당 토픽들의 조합이 리뷰마다 다르게 나타날 것이기 때문이다. 따라서, 동일한 결과(예: 관광지 만족도 점수가 5점일 때)를 설명하는 상황에 있어서도 서로 다른 원인조건(관광지 선택속성)의 조합이 발생할 수 있다는 것을 전제해야 한다. 이에 적합한 분석 방법이 FsQCA다. 전통적인 집합을 활용하는 QCA 분석과는 다르게 FsQCA 분석을 적용하면 결과변수를 설명하는 원인조건의 다양한 조합을 확인할 수 있다.

본 연구의 대상이 되는 관광지 5개소에 대한 FsQCA 분석은 본 연구의 제 3 장 연구 설계에 제시된 분석 절차대로 진행하였다. 요약하면, 관광지 별로 도출된 고유한 토픽명을 FsQCA 분석에서는 원인조건으로 사용하였으며, 관광지 만족도 점수를 결과변수로 사용하였다. 이후, 리뷰별 토픽의 할당 비율을 퍼지 점수로 변환 → 필요조건 검증 → 진리표 도출 → 표준분석을 통한 충분조건 검증 → 관광지 만족도를 결정하는 원인조건조합의 대표 모형 도출 순서대로 분석을 진행하였다.

한편, ‘일치도’는 필요조건 검증, 진리표 도출, 충분조건 검증에서 공통으로 확인되는 지표이다. 필요조건 검증에서는 각 원인조건이 필요조건에 대한 일치도가 0.75 이상일 때 해당 원인조건이 필요조건을 충족한다고 보았다. 즉, 필요조건은 결과가 발생하려면 원인조건을 반드시 필요로 하는 상황을 의미한다. Ragin(2006)의 경우, 엄격한 수준의 일치도가 0.8이라고 제

시하였으며, 필요조건의 일치도(일관성)가 0.75 이하로 나타나는 경우 하위 집합의 관계, 즉 필요조건으로 보기 어렵다고 하였다(김예린, 2018 재인용 p.122). 따라서 본 연구에서 원인조건이 결과변수에 대하여 필요조건을 충족하는지 여부는 일치도가 0.75 이상일 때를 기준으로 하였다. 즉, 필요조건에 대한 일치도 0.75 이상을 충족하는 원인조건이 발견될 때 이는 결과변수인 관광지 만족도를 평가하는 필요조건이라고 해석하였다.

한편, Ragin(2009)은 FsQCA 분석에서 충분조건을 검증하기 위해 합리적인 모형(또는 해법)을 도출하려면 일치도가 최소한 0.75보다 커야 하며, 일반적인 경우에서 0.8 이상을 권장하였다. 이는 원인조건조합에 따른 모형을 도출할 때 일치도가 낮은 원인조건조합으로 인해 비현실적인 경로가 제시되는 것을 방지하기 위함이다. 따라서, 충분조건 검증을 위한 사전 작업으로 진리표를 도출할 때에는 일치도 0.8 이상인 사례만을 선별하였다. 그러나 충분조건 검증에 대한 결과로 관광지 만족도를 구성하는 원인조건조합을 확인할 때에는 보다 엄격한 기준인 일치도 0.85 이상을 적용하여 충분조건을 설명하는 기준으로 하였다. 정리하면, 필요조건에 대한 일치도는 0.75 이상일 때, 진리표 도출에서는 일치도 0.8 이상일 때, 충분조건 검증에서 확인되는 각 모형별 일치도는 0.85 이상일 때를 기준으로 하였다.

충분조건을 검증하는 방법으로는 표준분석을 택하였다. 표준분석에는 복잡한 해법(complex solution), 중간 해법(intermediate solution), 그리고 간명한 해법(parsimony solution)의 세 종류 해법<sup>16)</sup>이 도출되는데, 이 중에서 가장

---

16) “복잡한 해법은 간명성 전략과 해석의 복잡성으로 인해 일반적으로 연구결과로 제시되지 않는다. 그리고 간명한 해법은 다중결합적 인과관계 전략을 강조하는 퍼지셋의 본성상 매우 간명한 1-2개의 원인조합으로 결과집합을 설명하는 단순화된 가정으로 인해 해석의 다양성에 제한을 두기 때문에 잘 활용되지 않는다. 결국 복잡성과 간명성 간의 제한된 다양성이라는 입장(Ragin & Sonnett, 2004) 하에 대부분 중간 해법이 연구결과로 보고된다. 특히 Ragin(2008, 79)은 퍼지셋 유저가이드에서 명시분석(Specify Analysis)보다 표준분석(Standard Analysis)을 추천하는데, 그 이유가 바로 표준분석에서만 중간해법을 제시해주기 때문이라고 안내한 바 있다(민기채, 2014 재인용 p.239)”

최적화된 결과를 제시하는 중간 해법을 토대로 분석 결과를 설명하였다. FsQCA 분석 결과 및 지표에 대한 설명은 첫 번째 분석 대상인 ‘설악산 국립공원’에서 비교적 상세히 기술하였다. 다른 관광지 4개소에 대하여는 분석 결과에서 결과에 따른 해석을 포함하지만, 일부 용어(예: 일치도, 설명력 등)에 대한 개념이나 설명이 중복으로 기술되는 것을 지양하고자, 반복 도출되는 용어와 설명은 간소화하고 결과에 따른 해석에 초점을 두었다. 관광지별 FsQCA 분석에 따른 결과는 다음과 같다.

## 1. 설악산 국립공원 FsQCA

설악산 국립공원 만족도에 대한 필요조건을 분석한 결과는 <표 4-13>와 같다. 일치도(Consistency for necessity) 0.75 이상을 기준으로 하였을 때 설악산 국립공원의 토픽모델링에서 도출된 6개의 원인조건(관광지 선택속성) 중에서 ‘사계절 매력성(일치도 = 0.813)’만이 관광지 만족도를 평가하는 필요조건으로 제시되었다. 즉, 설악산 국립공원을 방문한 관광객들이 남긴 온라인 리뷰 데이터에는 사계절 매력성에 관한 내용이 반드시 나타났다고 할 수 있다. 이에 비해, 다른 5개의 원인조건(케이블카 체험, 입장료 및 경비, 울산바위 명소, 교통수단 접근성, 등산 탐방로 코스)은 일치도 0.75 이상이 충족되지 않아 설악산 국립공원 관광지 만족도에 대한 필요조건이라고 할 수 없었다. 앞서 언급한 사계절 매력성이 필요조건으로써 가지는 설명력은 97.2%로 높은 수준이다. 각 원인조건이 결과변수인 설악산 국립공원 만족도에 대하여 필요조건을 충족하는지 여부를 나타내면 다음과 같다.

<표 4-13> 설악산 국립공원 만족도 필요조건 검증 결과\*

원인조건 (관광지 선택속성)	일치도	설명력	필요조건 충족 여부**
케이블카 체험	0.529	0.976	미충족
<b>사계절 매력성</b>	<b>0.813</b>	<b>0.972</b>	<b>충족</b>
입장료 및 경비	0.617	0.982	미충족
울산바위 명소	0.462	0.970	미충족
교통수단 접근성	0.360	0.985	미충족
등산 탐방로 코스	0.365	0.989	미충족

\* : 해당 표는 연구과제 3-1에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

\*\* : 결과변수에 대한 원인조건의 필요조건 충족 여부는 일치도 0.75 이상을 기준으로 하였음

이어서 설악산 국립공원 만족도에 대한 충분조건을 검증하였다. 설악산

국립공원 만족도를 설명하는 충분조건으로 6개의 원인조건이 존재할 수 있다고 가정할 때, 논리적으로 도출 가능한 원인조건의 조합은 64개이다. 이때 사례의 수가 4 미만인 조합을 제외하고 일치도 0.8 이상을 충족하는 사례만을 선별하였다. <표 4-14>는 설악산 국립공원의 관광지 만족도를 결정하는 원인조건의 조합을 나타내는 진리표이다. 진리표는 원인조건을 0과 1로 표시하여 논리적으로 결합 가능한 원인조건조합을 제시한 표이다. 진리표에 제시된 일치도는 원인조건에 해당하는 퍼지 점수가 결과의 부분집합에 일치하는 정도를 의미한다(이승윤, 2014). 그 결과 총 13개의 원인조건조합이 명시된 진리표가 도출되었다.

<표 4-14> 설악산 국립공원 진리표\*

	케이블카 체험	사계절 매력성	입장료 및 경비	울산바위 명소	교통수단 접근성	등산 탐방로 코스	사례 개수	일치도
1	1	1	1	0	0	0	12	1.000
2	0	1	1	0	0	0	11	1.000
3	1	1	0	0	0	0	9	0.984
4	0	1	1	0	0	1	7	0.999
5	1	1	0	1	0	0	7	0.991
6	0	1	0	1	1	0	7	0.978
7	0	1	0	0	0	0	5	0.994
8	0	1	1	1	0	0	4	1.000
9	1	1	1	0	1	0	4	1.000
10	0	0	1	1	1	0	4	1.000
11	1	1	1	1	1	0	4	1.000
12	0	0	1	0	0	1	4	1.000
13	1	1	1	1	0	0	4	0.993

\* : 해당 표는 연구과제 3-2에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

결과변수인 관광지 만족도(1 : 소속)에 대하여 리뷰에 나타난 원인조건조합을 사례 순으로 정리하였음

충분조건을 검증하기 위해 진리표에 나타난 13개의 원인조건조합을 대상으로 표준분석을 진행하였으며, 중간 해법을 통해 7개의 대표 모형<sup>17)</sup>을



도출하였다. <표 4-15>는 설악산 국립공원의 관광지 만족도에 대한 원인조건조합이 7개 모형으로 설명될 수 있음을 보여주는 내용이다. 모형 총 설명력은 0.593이며, 이 7개의 모형으로 관광지 만족도를 설명하는 원인조건조합의 59.3%가 설명된다고 할 수 있다. 일치도는 원인조건조합의 사례들이 결과변수의 부분집합이 되는 정도를 의미하는데, Ragin(2009)은 일치도가 0.85 이상일 때 거의 항상 충분조건이 된다고 하였다. 모형 총 일치도는 0.983으로 높게 나타났으며, 개별 모형의 일치도 역시 모두 0.85 이상으로 해당 모형들이 관광지 만족도를 구성하는 충분조건이 될 수 있음을 확인하였다.

한편, 개별 모형의 설명력은 인과적 조합의 실증적인 중요도를 나타내는 지표이다(이승윤, 2014). 개별 모형의 설명력은 해당 원인조건조합이 결과변수를 설명하는 정도를 나타내며, 회귀분석에서의 결정계수 값과 유사한 개념으로 해석될 수 있다(김지현, 2020). 설명력은 원 설명력(raw coverage)과 순 설명력(unique coverage)으로 구분되는데 원 설명력은 원인조건조합이 결과변수를 설명할 수 있는 설명력을 나타내고, 순 설명력은 다른 조합들과 중복되지 않는 해당 원인조건조합의 순수한 설명력을 의미한다. 원인조합의 소속점수보다 결과집합의 소속점수가 높은 사례가 많으면 일치도는 높게 나타나지만 설명력은 상대적으로 낮게 나타날 수 있다(민기채, 2014).

앞서 도출된 7개의 모형 중에서 결과변수인 관광지 만족도와 가장 관련이 높은 원인조건조합은 0.301의 설명력을 가지는 모형3이다. 이를 수식으로 표현하면 ‘사계절 매력성 \* 입장료 및 경비 \* ~교통수단 접근성 \* ~등산 탐방로 코스’이다. 여기서 ‘\*’는 논리 곱(logical AND)을 나타내며, ‘+’는 본 연구에서 제시된 부분이 아니지만 일반적으로 논리 합(logical OR)을 나타내고, ‘~’는 해당 원인조건의 부재(not)를 의미한다.

충분조건을 나타내는 총 7개의 원인조건조합을 모형으로 정리한 내용은

---

17) 일부 연구에서는 모형 대신 ‘해법’이라고 표기함

<표 4-15>와 같다. 해당 표에서 ‘●’ 표시는 원인조건의 존재를 의미하며, ‘○’ 표시는 원인조건의 부재를 의미한다. 그러나 ‘○’ 표시가 있다고 해서 해당 원인조건이 완전히 존재하지 않다고 해석하는 것은 적절하지 않다. 왜냐하면 본 연구에 사용된 원인조건(관광지 선택속성)은 퍼지 점수를 부여할 때 0과 1로만 부여하지 않고, 실제 토픽모델링을 통해 얻어진 연속적인 0과 1사이의 값을 활용하였기 때문이다.

예를 들어, 특정 리뷰에 존재하는 토픽 중 ‘사계절 매력성’에 대한 퍼지 점수가 0.3이라면, 이 점수를 토대로 진리표를 구성할 때에는 0과 1로만 나타내므로 사계절 매력성이 없음(0)에 표시된다. 0.3은 1보다 0에 가까우므로 존재(1)와 부재(0) 여부만 논했을 때 부재하다고 보는 것이 타당한 것이다. 따라서 원인조건조합을 나타내는 <표 4-15>는 앞서 제시된 진리표에서 얻어진 결과를 토대로 하였으므로, 사전에 진리표가 제시될 때 퍼지 점수에 따라 0과 1로 존재와 부재를 구분하고 있음에 유의하여야 한다. 만일 모형에 ‘사계절 매력성’이라는 원인조건이 제시되어 있다면 이는 사계절 매력성에 대한 내용이 해당 리뷰에 존재하는 상태, 즉 차지하는 비중이 높음을 의미한다. 반면, ‘~사계절 매력성’이라는 원인조건이 제시되어 있다면 이는 사계절 매력성에 대한 내용이 해당 리뷰에 부재하는 상태, 즉 차지하는 비중이 낮은 상태를 의미한다. 존재(●)와 부재(○)로 나타낸 표에서 빈 칸은 원인조건의 존재 혹은 부재와 상관없는 상태를 나타낸다.

충분조건 검증에서 설명력이 높게 나타난 모형들은 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 의미한다. 도출된 모형을 토대로 설악산 국립공원 관광지 만족도를 평가할 때 관광객들이 고려하는 관광지 선택속성의 조합을 두 개만 살펴보면 다음과 같다.

모형3에 따르면, 리뷰 내용에서 사계절 매력성과 입장료 및 경비에 대한 비중이 높고, 교통수단 접근성과 등산 탐방로 코스에 대한 비중이 낮을 때

관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 교통수단 접근성과 등산 탐방로 코스에 대한 내용을 상대적으로 적게 언급할지라도, 사계절 매력성과 입장료 및 경비에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 첫 번째로 많은 경우에서 설악산 국립공원에 대한 만족도를 평가하기 위해 관광객들이 사계절 매력성과 입장료 및 경비를 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형2에 따르면, 리뷰 내용에서 케이블카 체험과 사계절 매력성에 대한 비중이 높고, 교통수단 접근성과 등산 탐방로 코스에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 교통수단 접근성과 등산 탐방로 코스에 대한 내용을 상대적으로 적게 언급할지라도, 케이블카 체험과 사계절 매력성에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 두 번째로 많은 경우에서 설악산 국립공원에 대한 만족도를 평가하기 위해 관광객들이 케이블카 체험과 사계절 매력성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형2 및 모형3에 따르면, 공통으로 사계절 매력성에 대한 비중이 높고, 교통수단 접근성과 등산 탐방로 코스에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되었다. 즉, 교통수단 접근성과 등산 탐방로 코스는 리뷰에서 차지하는 부분이 적음에도 관광객들이 고려하는 관광지 선택속성이며, 사계절 매력성은 리뷰에서 차지하는 부분이 많으면서 관광객들이 중요하게 고려하는 관광지 선택속성임을 의미한다. 연구자가 대표적으로 설명한 모형 이외에도 충분조건의 검증 결과로서 나타난 7개의 원인조건조합 모형을 정리한 내용은 <표 4-15>와 같다.

<표 4-15> 설악산 국립공원 만족도 충분조건 검증 결과\*

원인조건 (관광지 선택속성)	모형1	모형2	모형3	모형4	모형5	모형6	모형7
케이블카 체험	●	●			○	○	○
사계절 매력성**	●	●	●	●	●	○	
입장료 및 경비	●		●		○	●	●
울산바위 명소				○	●	●	○
교통수단 접근성		○	○	○	●	●	○
등산 탐방로 코스	○	○	○	○	○	○	●
원 설명력	0.270	0.292	0.301	0.286	0.098	0.074	0.136
순 설명력	0.054	0.041	0.037	0.024	0.035	0.016	0.072
일치도	0.992	0.977	0.997	0.988	0.977	1.000	0.999
모형 총 설명력	0.593						
모형 총 일치도	0.983						

\* : 해당 표는 연구과제 3-3에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

진리표를 토대로 관광지 만족도를 구성하는 원인조건조합을 모형으로 표기(표준분석, 중간 해법 적용)

\*\* : 필요조건 검증에서 결과변수에 대한 필요조건을 충족한 원인조건

## 2. 한라산 국립공원 FsQCA

한라산 국립공원 만족도에 대한 필요조건을 분석한 결과는 <표 4-16>와 같다. 일치도 0.75 이상을 기준으로 하였을 때 한라산 국립공원의 토픽 모델링에서 도출된 6개의 원인조건 중에서 관광지 만족도를 평가하는 필요조건이 발견되지 않았다. 이는 공원·산 자연관광지인 설악산 국립공원에서 ‘사계절 매력성’이 필요조건으로 제시된 것과 상반된 결과이다. 즉, 필요조건에 해당하는 원인조건(관광지 선택속성)이 발견되지 않은 것은, 각 원인조건이 다른 원인조건과 결합하지 않은 상태에서 독자적으로 결과변수인 한라산 국립공원 만족도에 영향을 미치지 못하는 상황임을 의미한다. 각 원인조건이 결과변수인 한라산 국립공원 만족도에 대하여 필요조건을 충족하는지 여부를 나타내면 다음과 같다.

<표 4-16> 한라산 국립공원 만족도 필요조건 검증 결과\*

원인조건 (관광지 선택속성)	일치도	설명력	필요조건 충족 여부**
성관악 탐방로 코스	0.668	0.968	미충족
사계절 매력성	0.412	0.971	미충족
관음사 탐방로 코스	0.541	0.979	미충족
위치 및 접근성	0.582	0.972	미충족
자연 볼거리	0.704	0.969	미충족
편의시설 및 먹거리	0.343	0.979	미충족

\* : 해당 표는 연구과제 3-1에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

\*\* : 결과변수에 대한 원인조건의 필요조건 충족 여부는 일치도 0.75 이상을 기준으로 하였음

이어서 한라산 국립공원 만족도에 대한 충분조건을 검증하였다. 한라산 국립공원 만족도를 설명하는 충분조건으로 6개의 원인조건이 존재할 수 있다고 가정할 때, 논리적으로 도출 가능한 원인조건의 조합은 64개이다. 이

때 사례의 수가 5 미만인 조합을 제외하고 일치도 0.8 이상을 충족하는 사례만을 선별하였다. 그 결과 한라산 국립공원의 관광지 만족도를 결정하는 16개의 원인조건조합이 명시된 진리표가 도출되었다(<표 4-17>).

<표 4-17> 한라산 국립공원 진리표\*

	성판악 탐방로 코스	사계절 매력성	관음사 탐방로 코스	위치 및 접근성	자연 볼거리	편의시설 및 먹거리	사례 개수	일치도
1	1	0	1	1	1	0	17	0.998
2	1	0	1	0	1	0	11	1.000
3	1	0	0	1	0	0	11	0.982
4	1	0	0	1	1	1	10	0.995
5	1	0	1	1	0	0	9	1.000
6	1	0	0	0	1	0	8	1.000
7	1	0	0	0	1	1	8	1.000
8	0	1	0	0	1	0	8	0.994
9	0	1	1	0	1	0	6	1.000
10	0	1	0	1	1	0	6	0.997
11	0	0	1	0	1	1	6	0.996
12	1	1	0	1	1	0	6	0.988
13	1	0	0	0	0	0	5	1.000
14	1	0	0	1	1	0	5	1.000
15	1	1	1	1	1	0	5	0.999
16	1	1	0	0	1	0	5	0.970

\* : 해당 표는 연구과제 3-2에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)  
결과변수인 관광지 만족도(1 : 소속)에 대하여 리뷰에 나타난 원인조건조합을 사례 순으로 정리하였음

충분조건을 검증하기 위해 진리표에 나타난 16개의 원인조건조합을 대상으로 표준분석을 진행하였으며, 중간 해법을 통해 8개의 대표 모형을 도출하였다. <표 4-18>는 한라산 국립공원의 관광지 만족도에 대한 원인조건조합이 8개 모형으로 설명될 수 있음을 보여주는 내용이다. 모형 총 설명력은 0.606이며, 이 8개의 모형으로 관광지 만족도를 설명하는 원인조건조합의 60.6%가 설명된다고 할 수 있다. 원인조건조합의 사례들이 결과변수의

부분집합이 되는 정도를 나타내는 일치도를 살펴보면, 각각의 모형 일치도와 모형 총 일치도가 모두 0.85 이상으로 높게 나타나 해당 모형들이 관광지 만족도를 구성하는 충분조건이 될 수 있음을 확인하였다. 앞서 도출된 8개의 모형 중에서 결과변수인 관광지 만족도와 가장 관련이 높은 원인조건 조합은 0.272의 설명력을 가지는 모형4(성판악 탐방로 코스 \* ~사계절 매력성 \* 자연 볼거리 \* ~편의시설 및 먹거리)와 0.271의 설명력을 가지는 모형3(성판악 탐방로 코스 \* ~사계절 매력성 \* 위치 및 접근성 \* ~편의시설 및 먹거리)이다. 충분조건 검증에서 설명력이 높게 나타난 모형들은 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 의미한다. 도출된 모형을 토대로 한라산 국립공원 관광지 만족도를 평가할 때 관광객들이 고려하는 관광지 선택속성의 조합을 살펴보면 다음과 같다.

모형4에 따르면, 리뷰 내용에서 성판악 탐방로 코스와 자연 볼거리에 대한 비중이 높고, 사계절 매력성과 편의시설 및 먹거리에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 사계절 매력성과 편의시설 및 먹거리에 대한 내용을 상대적으로 적게 언급할지라도, 성판악 탐방로 코스와 자연 볼거리에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 첫 번째로 많은 경우에서 한라산 국립공원에 대한 만족도를 평가하기 위해 관광객들이 성판악 탐방로 코스와 자연 볼거리를 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형3에 따르면, 리뷰 내용에서 성판악 탐방로 코스와 위치 및 접근성에 대한 비중이 높고, 사계절 매력성과 편의시설 및 먹거리에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 사계절 매력성과 편의시설 및 먹거리에 대한 내용을 상대적으로 적게 언급할지라도, 성판악 탐방로 코스와 위치 및 접근성에 대한 내용을 상

대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 두 번째로 많은 경우에서 한라산 국립공원에 대한 만족도를 평가하기 위해 관광객들이 성판악 탐방로 코스와 위치 및 접근성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형3 및 모형4에 따르면, 공통으로 성판악 탐방로 코스에 대한 비중이 높고, 사계절 매력성과 편의시설 및 먹거리에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되었다. 즉, 사계절 매력성과 편의시설 및 먹거리는 리뷰에서 차지하는 부분이 적음에도 관광객들이 고려하는 관광지 선택속성이며, 성판악 탐방로 코스는 리뷰에서 차지하는 부분이 많으면서 관광객들이 중요하게 고려하는 관광지 선택속성임을 의미한다. 연구자가 대표적으로 설명한 모형 이외에도 충분조건의 검증 결과로서 나타난 8개의 원인조건조합 모형을 정리한 내용은 <표 4-18>와 같다.

<표 4-18> 한라산 국립공원 만족도 충분조건 검증 결과\*

원인조건 (관광지 선택속성)	모형1	모형2	모형3	모형4	모형5	모형6	모형7	모형8
성판악 탐방로 코스	●	●	●	●	●	○	○	
사계절 매력성	○		○	○		●	○	●
관음사 탐방로 코스	○	○					●	○
위치 및 접근성			●		●	○	○	
자연 볼거리	●			●	●	●	●	●
편의시설 및 먹거리		○	○	○	○	○	●	○
원 설명력	0.211	0.184	0.271	0.272	0.257	0.111	0.078	0.172
순 설명력	0.073	0.011	0.036	0.050	0.021	0.020	0.026	0.041
일치도	0.989	0.977	0.983	0.990	0.992	0.995	0.996	0.969
모형 총 설명력	0.606							
모형 총 일치도	0.977							

\* : 해당 표는 연구과제 3-3에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)  
진리표를 토대로 관광지 만족도를 구성하는 원인조건조합을 모형으로 표기(표준분석, 중간 해법 적용)



### 3. 순천만습지(순천만국가정원) FsQCA

순천만습지(순천만국가정원) 만족도에 대한 필요조건을 분석한 결과는 <표 4-19>와 같다. 일치도 0.75 이상을 기준으로 하였을 때 순천만습지(순천만국가정원)의 토픽모델링에서 도출된 6개의 원인조건(관광지 선택속성) 중에서 ‘사계절 매력성(일치도 = 0.789)’ 만이 관광지 만족도를 평가하는 필요조건으로 제시되었다. 즉, 순천만습지(순천만국가정원)를 방문한 관광객들이 남긴 온라인 리뷰 데이터에는 사계절 매력성에 관한 내용이 반드시 나타났다고 할 수 있다. 이에 비해, 다른 5개의 원인조건(볼거리 및 이동수단, 접근 편리성, 관광 행사, 자연관광 매력성, 먹거리 및 체험학습)은 일치도 0.75 이상이 충족되지 않아 순천만습지(순천만국가정원) 관광지 만족도에 대한 필요조건이라고 할 수 없었다. 앞서 언급한 사계절 매력성이 필요조건으로써 가지는 설명력은 94.6%로 높은 수준이다. 각 원인조건이 결과변수인 순천만습지(순천만국가정원) 만족도에 대하여 필요조건을 충족하는지 여부를 나타내면 다음과 같다.

<표 4-19> 순천만습지(순천만국가정원) 만족도 필요조건 검증 결과\*

원인조건 (관광지 선택속성)	일치도	설명력	필요조건 충족 여부**
<b>사계절 매력성</b>	<b>0.789</b>	<b>0.946</b>	<b>충족</b>
볼거리 및 이동수단	0.461	0.958	미충족
접근 편리성	0.252	0.940	미충족
관광 행사	0.641	0.943	미충족
자연관광 매력성	0.583	0.956	미충족
먹거리 및 체험학습	0.506	0.963	미충족

\* : 해당 표는 연구과제 3-1에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

\*\* : 결과변수에 대한 원인조건의 필요조건 충족 여부는 일치도 0.75 이상을 기준으로 하였음

이어서 순천만습지(순천만국가정원) 만족도에 대한 충분조건을 검증하였다. 순천만습지(순천만국가정원) 만족도를 설명하는 충분조건으로 6개의 원인조건이 존재할 수 있다고 가정할 때, 논리적으로 도출 가능한 원인조건 조합은 64개이다. 이때 사례의 수가 5 미만인 조합을 제외하고 일치도 0.8 이상을 충족하는 사례만을 선별하였다. 그 결과 순천만습지(순천만국가정원)의 관광지 만족도를 결정하는 20개의 원인조건조합이 명시된 진리표가 도출되었다(<표 4-20>).

<표 4-20> 순천만습지(순천만국가정원) 진리표\*

	사계절 매력성	볼거리 및 이동수단	접근 편리성	관광 행사	자연관광 매력성	먹거리 및 체험학습	사례 개수	일치도
1	1	0	0	1	0	1	17	0.994
2	1	1	0	1	0	0	16	0.984
3	1	0	0	1	1	0	15	0.973
4	1	1	0	0	1	0	14	0.999
5	1	0	0	1	1	1	14	0.990
6	1	0	0	1	0	0	12	0.987
7	1	1	0	1	0	1	11	0.997
8	1	0	0	0	1	0	10	0.993
9	1	0	0	0	0	1	9	0.995
10	1	0	0	0	1	1	8	0.999
11	1	1	0	1	1	0	8	0.984
12	1	1	0	1	1	1	6	1.000
13	1	1	0	0	1	1	6	0.999
14	1	0	1	1	1	0	6	0.998
15	1	0	1	0	1	0	5	1.000
16	0	1	0	0	1	1	5	1.000
17	1	0	1	1	1	1	5	0.997
18	1	1	0	0	0	0	5	0.996
19	1	0	1	1	0	0	5	0.986
20	1	0	1	0	0	0	5	0.967

\* : 해당 표는 연구과제 3-2에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

결과변수인 관광지 만족도(1 : 소속)에 대하여 리뷰에 나타난 원인조건조합을 사례 순으로 정리하였음

충분조건을 검증하기 위해 진리표에 나타난 20개의 원인조건조합을 대상으로 표준분석을 진행하였으며, 중간 해법을 통해 8개의 대표 모형을 도출하였다. <표 4-21>는 순천만습지(순천만국가정원)의 관광지 만족도에 대한 원인조건조합이 8개 모형으로 설명될 수 있음을 보여주는 내용이다. 모형 총 설명력은 0.721이며, 이 8개의 모형으로 관광지 만족도를 설명하는 원인조건조합의 72.1%가 설명된다고 할 수 있다. 원인조건조합의 사례들이 결과변수의 부분집합이 되는 정도를 나타내는 일치도를 살펴보면, 각각의 모형 일치도와 모형 총 일치도가 모두 0.85 이상으로 높게 나타나 해당 모형들이 관광지 만족도를 구성하는 충분조건이 될 수 있음을 확인하였다. 앞서 도출된 8개의 모형 중에서 결과변수인 관광지 만족도와 가장 관련이 높은 원인조건조합은 0.461의 설명력을 가지는 모형6(사계절 매력성 \* ~접근 편리성 \* 관광 행사)과 0.388의 설명력을 가지는 모형7(사계절 매력성 \* ~접근 편리성 \* 자연관광 매력성)이다. 충분조건 검증에서 설명력이 높게 나타난 모형들은 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 의미한다. 도출된 모형을 토대로 순천만습지(순천만국가정원) 관광지 만족도를 평가할 때 관광객들이 고려하는 관광지 선택속성의 조합을 살펴보면 다음과 같다.

모형6에 따르면, 리뷰 내용에서 사계절 매력성과 관광 행사에 대한 비중이 높고, 접근 편리성에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 접근 편리성에 대한 내용을 상대적으로 적게 언급할지라도, 사계절 매력성과 함께 관광 행사에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 첫 번째로 많은 경우에서 순천만습지(순천만국가정원)에 대한 만족도를 평가하기 위해 관광객들이 사계절 매력성과 동시에 관광 행사를 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형7에 따르면, 리뷰 내용에서 사계절 매력성과 자연관광 매력성에 대한 비중이 높고, 접근 편리성에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 접근 편리성에 대한 내용을 상대적으로 적게 언급할지라도, 사계절 매력성과 함께 자연관광 매력성에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 두 번째로 많은 경우에서 순천만습지(순천만국가정원)에 대한 만족도를 평가하기 위해 관광객들이 사계절 매력성과 동시에 자연관광 매력성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형6 및 모형7에 따르면, 공통으로 사계절 매력성에 대한 비중이 높고, 접근 편리성에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되었다. 즉, 접근 편리성은 리뷰에서 차지하는 부분이 적음에도 관광객들이 고려하는 관광지 선택속성이며, 사계절 매력성은 리뷰에서 차지하는 부분이 많으면서 관광객들이 중요하게 고려하는 관광지 선택속성임을 의미한다. 연구자가 대표적으로 설명한 모형 이외에도 충분조건의 검증 결과로서 나타난 8개의 원인조건조합 모형을 정리한 내용은 <표 4-21>와 같다.

<표 4-21> 순천만습지(순천만국가정원) 만족도 충분조건 검증 결과\*

원인조건 (관광지 선택속성)	모형1	모형2	모형3	모형4	모형5	모형6	모형7	모형8
사계절 매력성**	●	●	●	●	●	●	●	
볼거리 및 이동수단	●	○	○	○	○			●
접근 편리성	○		●	○		○	○	○
관광 행사					●	●		○
자연관광 매력성		●			●		●	●
먹거리 및 체험학습	○	○	○	●				●
원 설명력	0.219	0.206	0.123	0.253	0.248	0.461	0.388	0.110
순 설명력	0.018	0.001	0.030	0.028	0.013	0.082	0.012	0.016
일치도	0.978	0.962	0.953	0.979	0.967	0.960	0.968	0.995
모형 총 설명력	0.721							
모형 총 일치도	0.956							

\* : 해당 표는 연구과제 3-3에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

진리표를 토대로 관광지 만족도를 구성하는 원인조건조합을 모형으로 표기(표준분석, 중간 해법 적용)

\*\* : 필요조건 검증에서 결과변수에 대한 필요조건을 충족한 원인조건

#### 4. 남이섬 FsQCA

남이섬 만족도에 대한 필요조건을 분석한 결과는 <표 4-22>와 같다. 일치도 0.75 이상을 기준으로 하였을 때 남이섬의 토픽모델링에서 도출된 6개의 원인조건(관광지 선택속성) 중에서 ‘자연관광 매력성(일치도 = 0.794)’만이 관광지 만족도를 평가하는 필요조건으로 제시되었다. 즉, 남이섬을 방문한 관광객들이 남긴 온라인 리뷰 데이터에는 자연관광 매력성에 관한 내용이 반드시 나타났다고 할 수 있다. 이에 비해, 다른 5개의 원인조건(접근 편리성, 관광 명소, 데이트 코스, 숙박시설, 음식 및 편의시설)은 일치도 0.75 이상이 충족되지 않아 남이섬 관광지 만족도에 대한 필요조건이라고 할 수 없었다. 앞서 언급한 자연관광 매력성이 필요조건으로써 가지는 설명력은 85.6%로 높은 수준이다. 각 원인조건이 결과변수인 남이섬 만족도에 대하여 필요조건을 충족하는지 여부를 나타내면 다음과 같다.

<표 4-22> 남이섬 만족도 필요조건 검증 결과\*

원인조건 (관광지 선택속성)	일치도	설명력	필요조건 충족 여부**
접근 편리성	0.484	0.869	미충족
관광 명소	0.462	0.897	미충족
데이트 코스	0.717	0.850	미충족
숙박시설	0.458	0.896	미충족
음식 및 편의시설	0.531	0.878	미충족
<b>자연관광 매력성</b>	<b>0.794</b>	<b>0.856</b>	<b>충족</b>

\* : 해당 표는 연구과제 3-1에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

\*\* : 결과변수에 대한 원인조건의 필요조건 충족 여부는 일치도 0.75 이상을 기준으로 하였음

이어서 남이섬 만족도에 대한 충분조건을 검증하였다. 남이섬 만족도를 설명하는 충분조건으로 6개의 원인조건이 존재할 수 있다고 가정할 때, 논

리적으로 도출 가능한 원인조건의 조합은 64개이다. 이때 사례의 수가 3 미만인 조합을 제외하고 일치도 0.8 이상을 충족하는 사례만을 선별하였다. 그 결과 남이섬의 관광지 만족도를 결정하는 15개의 원인조건조합이 명시된 진리표가 도출되었다(<표 4-23>).

<표 4-23> 남이섬 진리표\*

	접근 편리성	관광 명소	데이트 코스	숙박 시설	음식 및 편의시설	자연관광 매력성	사례 개수	일치도
1	0	0	1	0	0	1	10	0.983
2	0	0	1	1	1	1	6	0.990
3	0	0	1	0	1	1	6	0.911
4	0	0	0	0	1	1	5	0.984
5	0	1	1	0	0	1	4	0.991
6	0	1	1	1	0	1	4	0.988
7	1	0	1	0	0	1	4	0.926
8	1	0	1	0	1	1	4	0.921
9	0	0	0	0	0	1	3	1.000
10	0	0	0	1	0	1	3	1.000
11	0	0	0	1	1	1	3	1.000
12	0	1	1	1	1	1	3	0.996
13	1	1	1	0	0	1	3	0.991
14	0	0	1	1	0	1	3	0.991
15	1	1	1	0	1	1	3	0.957

\* : 해당 표는 연구과제 3-2에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)  
결과변수인 관광지 만족도(1 : 소속)에 대하여 리뷰에 나타난 원인조건조합을 사례 순으로 정리하였음

충분조건을 검증하기 위해 진리표에 나타난 15개의 원인조건조합을 대상으로 표준분석을 진행하였으며, 중간 해법을 통해 6개의 대표 모형을 도출하였다. <표 4-24>는 남이섬의 관광지 만족도에 대한 원인조건조합이 6개 모형으로 설명될 수 있음을 보여주는 내용이다. 모형 총 설명력은 0.639이며, 이 6개의 모형으로 관광지 만족도를 설명하는 원인조건조합의 63.9%가 설명된다고 할 수 있다. 원인조건조합의 사례들이 결과변수의 부분집합이 되는 정도를 나타내는 일치도를 살펴보면, 각각의 모형 일치도와 모형

총 일치도가 모두 0.85 이상으로 높게 나타나 해당 모형들이 관광지 만족도를 구성하는 충분조건이 될 수 있음을 확인하였다. 앞서 도출된 6개의 모형 중에서 결과변수인 관광지 만족도와 가장 관련이 높은 원인조건조합은 0.395의 설명력을 가지는 모형3(∼접근 편리성 \* ∼관광 명소 \* 자연관광 매력성)이다. 충분조건 검증에서 설명력이 높게 나타난 모형들은 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 의미한다. 도출된 모형을 토대로 남이섬 관광지 만족도를 평가할 때 관광객들이 고려하는 관광지 선택속성의 조합을 살펴보면 다음과 같다.

모형3에 따르면, 리뷰 내용에서 접근 편리성과 관광 명소에 대한 비중이 낮고, 자연관광 매력성에 대한 비중이 높을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 접근 편리성과 관광 명소에 대한 내용을 상대적으로 적게 언급할지라도, 자연관광 매력성에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 앞서 진행된 필요조건 검증에서 결과변수에 대하여 필요조건을 충족한 원인조건은 자연관광 매력성이 유일하였다. 그러므로 자연관광 매력성은 남이섬 만족도에 대한 필요조건이자, 모형3에 따른 결과에서는 충분조건이 성립되는 관광지 선택속성으로 확인된다. 따라서, 첫 번째로 많은 경우에서 남이섬에 대한 만족도를 평가하기 위해 관광객들이 자연관광 매력성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다. 연구자가 대표적으로 설명한 모형 이외에도 충분조건의 검증 결과로서 나타난 6개의 원인조건조합 모형을 정리한 내용은 <표 4-24>와 같다.



<표 4-24> 남이섬 만족도 충분조건 검증 결과\*

원인조건 (관광지 선택속성)	모형1	모형2	모형3	모형4	모형5	모형6
접근 편리성	●	●	○	○	○	
관광 명소	○		○			
데이트 코스	●	●		●	●	●
숙박시설	○	○		●		○
음식 및 편의시설	○				○	○
자연관광 매력성**		●	●	●	●	●
원 설명력	0.157	0.229	0.395	0.252	0.280	0.265
순 설명력	0.029	0.043	0.139	0.014	0.003	0.000
일치도	0.935	0.885	0.932	0.959	0.959	0.942
모형 총 설명력	0.639					
모형 총 일치도	0.899					

\* : 해당 표는 연구과제 3-3에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

진리표를 토대로 관광지 만족도를 구성하는 원인조건조합을 모형으로 표기(표준분석, 중간 해법 적용)

\*\* : 필요조건 검증에서 결과변수에 대한 필요조건을 충족한 원인조건

## 5. 우도 FsQCA

우도 만족도에 대한 필요조건을 분석한 결과는 <표 4-25>와 같다. 일치도 0.75 이상을 기준으로 하였을 때 우도의 토픽모델링에서 도출된 6개의 원인조건 중에서 관광지 만족도를 평가하는 필요조건이 발견되지 않았다. 이는 바다·강 자연관광지인 순천만습지(순천만국가정원)에서 ‘사계절 매력성’이 필요조건으로 제시된 것과, 남이섬에서 ‘자연관광 매력성’이 필요조건으로 제시된 것과 상반된 결과이다. 즉, 필요조건에 해당하는 원인조건(관광지 선택속성)이 발견되지 않은 것은, 각 원인조건이 다른 원인조건과 결합하지 않은 상태에서 독자적으로 결과변수인 우도 만족도에 영향을 미치지 못하는 상황임을 의미한다. 각 원인조건이 결과변수인 우도 만족도에 대하여 필요조건을 충족하는지 여부를 나타내면 다음과 같다.

<표 4-25> 우도 만족도 필요조건 검증 결과\*

원인조건 (관광지 선택속성)	일치도	설명력	필요조건 충족 여부**
개별 이동수단	0.699	0.914	미충족
대중교통 이동수단	0.250	0.918	미충족
지역 음식	0.547	0.936	미충족
자연관광 매력성	0.705	0.916	미충족
날씨 요소	0.672	0.923	미충족
접근 편리성	0.341	0.946	미충족

\* : 해당 표는 연구과제 3-1에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)

\*\* : 결과변수에 대한 원인조건의 필요조건 충족 여부는 일치도 0.75 이상을 기준으로 하였음

이어서 우도 만족도에 대한 충분조건을 검증하였다. 우도 만족도를 설명하는 충분조건으로 6개의 원인조건이 존재할 수 있다고 가정할 때, 논리적으로 도출 가능한 원인조건의 조합은 64개이다. 이때 사례의 수가 7 미만인

조합을 제외하고 일치도 0.8 이상을 충족하는 사례만을 선별하였다. 그 결과 우도의 관광지 만족도를 결정하는 23개의 원인조건조합이 명시된 진리표가 도출되었다(<표 4-26>).

<표 4-26> 우도 진리표\*

	개별 이동수단	대중교통 이동수단	지역 음식	자연관광 매력성	날씨 요소	접근 편리성	사례 개수	일치도
1	1	0	0	1	1	0	37	0.938
2	0	0	1	1	1	0	31	0.978
3	1	0	1	1	1	0	29	0.973
4	1	0	0	1	1	1	22	0.980
5	1	0	1	1	0	0	21	0.974
6	1	0	1	0	1	0	20	0.989
7	1	0	0	1	0	0	19	0.984
8	0	1	1	0	1	0	16	0.985
9	1	0	0	1	0	1	15	0.984
10	0	0	0	1	1	0	12	0.993
11	1	0	0	0	1	1	11	0.988
12	1	0	1	1	1	1	10	0.994
13	1	0	1	1	0	1	10	0.989
14	1	0	0	0	1	0	10	0.977
15	1	1	0	1	1	0	9	0.970
16	1	1	1	0	1	0	9	0.957
17	0	0	1	0	1	0	8	0.996
18	0	0	0	1	1	1	8	0.981
19	0	0	0	1	0	1	8	0.979
20	0	0	1	1	0	0	7	0.985
21	0	0	0	1	0	0	7	0.982
22	1	1	0	1	0	0	7	0.975
23	1	1	1	1	0	0	7	0.961

\* : 해당 표는 연구과제 3-2에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)  
결과변수인 관광지 만족도(1 : 소속)에 대하여 리뷰에 나타난 원인조건조합을 사례 순으로 정리하였음

충분조건을 검증하기 위해 진리표에 나타난 23개의 원인조건조합을 대상으로 표준분석을 진행하였으며, 중간 해법을 통해 7개의 대표 모형을 도

출하였다. <표 4-27>는 우도의 관광지 만족도에 대한 원인조건조합이 7개 모형으로 설명될 수 있음을 보여주는 내용이다. 모형 총 설명력은 0.764이며, 이 7개의 모형으로 관광지 만족도를 설명하는 원인조건조합의 76.4%가 설명된다고 할 수 있다. 원인조건조합의 사례들이 결과변수의 부분집합이 되는 정도를 나타내는 일치도를 살펴보면, 각각의 모형 일치도와 모형 총 일치도가 모두 0.85 이상으로 높게 나타나 해당 모형들이 관광지 만족도를 구성하는 충분조건이 될 수 있음을 확인하였다.

앞서 도출된 7개의 모형 중에서 결과변수인 관광지 만족도와 가장 관련이 높은 원인조건조합은 0.433의 설명력을 가지는 모형2(개별 이동수단 \* ~대중교통 이동수단 \* 자연관광 매력성)와 0.425의 설명력을 가지는 모형6(~대중교통 이동수단 \* 자연관광 매력성 \* ~접근 편리성)이다. 충분조건 검증에서 설명력이 높게 나타난 모형들은 관광지 만족도를 평가하는 리뷰에서 높은 비중으로 확인되는 원인조건조합의 유형을 의미한다. 도출된 모형을 토대로 우도 관광지 만족도를 평가할 때 관광객들이 고려하는 관광지 선택속성의 조합을 살펴보면 다음과 같다.

모형2에 따르면, 리뷰 내용에서 개별 이동수단과 자연관광 매력성에 대한 비중이 높고, 대중교통 이동수단에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 대중교통 이동수단에 대한 내용을 상대적으로 적게 언급할지라도, 개별 이동수단과 함께 자연관광 매력성에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 첫 번째로 많은 경우에서 우도에 대한 만족도를 평가하기 위해 관광객들이 개별 이동수단과 동시에 자연관광 매력성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형6에 따르면, 리뷰 내용에서 대중교통 이동수단과 접근 편리성에 대한 비중이 낮고, 자연관광 매력성에 대한 비중이 높을 때 관광지 만족도가

높게 설명되는 것으로 나타났다. 이러한 결과는 관광객이 대중교통 이동수단과 접근 편리성에 대한 내용을 상대적으로 적게 언급할지라도, 자연관광 매력성에 대한 내용을 상대적으로 많이 언급함으로써 관광지 만족도가 평가되고 있음을 의미한다. 따라서, 두 번째로 많은 경우에서 우도에 대한 만족도를 평가하기 위해 관광객들이 자연관광 매력성을 고려하며, 이 내용을 토대로 리뷰를 작성한다고 할 수 있다.

모형2 및 모형6에 따르면, 공통으로 자연관광 매력성에 대한 비중이 높고, 대중교통 이동수단에 대한 비중이 낮을 때 관광지 만족도가 높게 설명되었다. 즉, 대중교통 이동수단은 리뷰에서 차지하는 부분이 적음에도 관광객들이 고려하는 관광지 선택속성이며, 자연관광 매력성은 리뷰에서 차지하는 부분이 많으면서 관광객들이 중요하게 고려하는 관광지 선택속성임을 의미한다. 연구자가 대표적으로 설명한 모형 이외에도 충분조건의 검증 결과로서 나타난 7개의 원인조건조합 모형을 정리한 내용은 <표 4-27>와 같다.

<표 4-27> 우도 만족도 충분조건 검증 결과\*

원인조건 (관광지 선택속성)	모형1	모형2	모형3	모형4	모형5	모형6	모형7
개별 이동수단	●	●	●	●			
대중교통 이동수단	○	○			○	○	
지역 음식	○			○	○		●
자연관광 매력성		●	●	●	●	●	○
날씨 요소	●		○				●
접근 편리성				○		○	○
원 설명력	0.241	0.433	0.184	0.223	0.334	0.425	0.172
순 설명력	0.042	0.042	0.013	0.012	0.031	0.070	0.098
일치도	0.933	0.926	0.955	0.927	0.926	0.927	0.965
모형 총 설명력	0.764						
모형 총 일치도	0.918						

\* : 해당 표는 연구과제 3-3에 대한 분석 결과를 나타냄(결과변수 : 관광지 만족도)  
진리표를 토대로 관광지 만족도를 구성하는 원인조건조합을 모형으로 표기(표준분석, 중간 해법 적용)

## 제 5 장 결 론

### 제 1 절 연구의 요약

본 연구는 관광지 만족도를 평가하는 관광지 선택속성에 관한 연구로서, 한국관광100선에 5회 연속 선정된 자연관광지를 대상으로 온라인 리뷰 데이터를 수집하여 연구를 진행하였다. 본 연구의 대상이 되는 자연관광지 5개소는 ① 설악산 국립공원, ② 한라산 국립공원, ③ 순천만습지(순천만국가정원), ④ 남이섬, ⑤ 우도이다. 온라인 리뷰에는 관광지에 대한 관광객의 생각과 느낌, 경험과 인식이 담겨있으므로 관광객이 관광지 만족도와 함께 고려하는 관광지 선택속성을 파악하는 데에 유용한 단서를 얻을 수 있다.

관광지에 대한 실제 평가와 리뷰가 활발하게 공유되는 여행 커뮤니티로 트립어드바이저를 선정하였으며, 관광지 5개소에 대하여 내국인 방문객이 작성한 국문 리뷰를 파이썬 기반의 웹크롤러를 통해 수집하였다. 수집된 리뷰 데이터는 분석에 활용 가능하도록 전처리 과정을 거쳤으며, 정제된 데이터를 빈도 분석, LDA 토픽모델링, FsQCA 분석에 활용하였다. 이로써 질적 자료 형태로 존재하는 리뷰 데이터를 양적 연구 및 분석에 활용할 수 있는 연구 방법을 제시하였다. 본 연구의 결과를 요약하면 다음과 같다.

첫 번째, 연구과제에 따른 분석을 수행하기에 앞서 기본 분석을 하였다. 기본 분석에는 기술통계 분석과 빈도 분석이 포함된다. 기술통계 분석으로 관광지 5개소에 대한 기본 정보(관광지명, 리뷰 개수)와 표본의 특성(관광지 만족도 점수, 동반자 유형, 방문일 등)을 확인하였다. 이와 함께 수집한 리뷰 데이터는 전처리 작업, 띄어쓰기 점검, 불용어 처리 등을 거친 후 명사만 추출하여 빈도 분석에 활용하였다. 관광지마다 도출되는 단어를 토대

로 빈도 분석을 하였으며, 관광지별 상위 출현하는 단어의 도출 빈도에 따라 1위부터 90위까지의 단어를 비율과 함께 제시하였다. 가중치를 부여하지 않은 상태에서 빈도 분석을 한 것은 일반적인 경우에서 관광객들이 리뷰에 자주 언급하는 단어가 무엇인지 확인하기 위함이며, 이때 유사어는 고려하지 않았다. 정리하면, 빈도 분석으로 관광지를 설명하는 주요 단어와 최소 세 군데 이상의 관광지에서 공통으로 도출된 단어들을 확인할 수 있었다.

두 번째, 트립어드바이저는 관광지 만족도 점수와 함께 리뷰 내용을 제공한다. 이때, 관광지 만족도를 평가하는 리뷰 내용이 어떠한 관광지 선택속성, 즉 토픽으로 구성되어 있는지 알아보기 위해 TF-IDF 가중치를 적용한 LDA 토픽모델링을 하였다. 그 결과, 관광지마다 여섯 개의 고유한 토픽(관광지 선택속성)이 도출되었다. 본 연구의 대상인 관광지 5개소는 모두 대한민국을 대표하는 자연관광지라고 할 수 있다. 관광객들의 솔직한 생각과 느낌이 담긴 리뷰를 토대로 토픽모델링을 하여 관광지 선택속성을 도출한 것은 의미있는 시도였다. 분석 결과를 통해 관광지별로 도출되는 토픽의 특성을 살펴보았으며, 특정 관광지에서만 나타나는 독특한 토픽과, 최소 두 군데 이상의 관광지에서 공통으로 도출되는 토픽을 함께 관찰하였다.

세 번째, 토픽모델링에서 FsQCA 분석으로 확장할 수 있는 구체적인 연구 방법을 제시하였다. 본 연구는 리뷰 별로 할당된 토픽의 비율 값과 관광지 만족도 점수를 퍼지 점수로 변환하여 FsQCA 분석에 활용하였다. 우선, 관광지 만족도를 평가하는 필요조건을 검증하기 위해 관광지 만족도(결과 변수)에 대하여 필요조건에 해당하는 관광지 선택속성(원인조건)을 규명하였다. 그 결과, 설악산 국립공원과 순천만습지(순천만국가정원)에서 ‘사계절 매력성’이, 남이섬에서 ‘자연관광 매력성’이 필요조건으로 밝혀졌다. 필요조건으로 제시된 관광지 선택속성은 리뷰에 반드시 언급된다고 할 수 있는 내용이므로, 관광지 만족도를 평가할 때 방문객들이 고려하는 사항이다.



한편, 동일한 관광지 만족도 점수를 부여하더라도, 관광객들마다 서로 다른 원인조건(관광지 선택속성)의 조합으로 리뷰 내용이 형성될 수 있다. 따라서, 동일한 결과(예: 관광지 만족도 점수가 5점일 때)를 설명하는 상황에 있어서도 서로 다른 원인조건 조합이 발생할 수 있음을 전제해야 한다. 이에 토픽모델링으로 얻어진 토픽(관광지 선택속성)을 원인조건으로 하여 결과변수인 관광지 만족도 간의 관계를 결합된 인과관계 관점에서 분석하였다. 관광지 만족도를 설명하는 다양한 원인조건조합을 제시하기 위해 결과변수인 관광지 만족도가 1(소속)에 포함되는 사례를 대상으로 결합 가능한 원인조건조합이 포함된 진리표를 도출하였고, 진리표를 토대로 충분조건을 검증하였다. 이때, 표준분석을 통해 제시되는 중간 해법에 근거하여 관광지 선택속성과 관광지 만족도의 관계를 설명하였다. 분석 결과와 함께 모형 설명력과 모형 일치도가 높은 경우에서 어떻게 관광지 만족도가 평가되는지 기술하였으며, 관광지 선택속성과 관광지 만족도의 관계를 FsQCA 관점에서 설명하였다.

## 제 2 절 시사점 및 제언

### 1. 학문적 시사점

본 연구의 결과를 바탕으로 학문적 시사점을 제시하면 다음과 같다.

첫 번째, 본 연구는 한국관광100선에 5회 연속 선정된 자연관광지 5개소를 대상으로 국문 리뷰 데이터를 수집하여 분석을 진행하였다. 리뷰 데이터를 분석에 활용하기 위해 텍스트 마이닝 같은 빅데이터와 관련된 문헌들을 고찰하고, 온라인 리뷰 데이터를 분석에 활용할 수 있는 구체적인 연구 방법을 제시하였다. 국문 리뷰 데이터는 내국인이 한글로 작성하였기 때문에 자연어 형태로 데이터가 존재한다. 따라서 이러한 자연어를 분석에 활용하기 위해서는 맞춤법 검사, 문장 정규화 작업, 불용어 처리 등의 다소 복잡한 처리 과정을 거쳐야 한다. 영어로 작성된 리뷰 데이터보다 전처리 및 분석 준비가 까다롭다고 할 수 있는 국문 리뷰 데이터를 일련의 작업을 거쳐 실증분석에 활용한 것은 의미 있는 시도라고 할 수 있다. 만일 국문 리뷰 데이터를 분석에 활용하려는 연구가 있다면, 본 연구는 이에 대한 구체적인 연구 설계 및 분석 방법을 제시하는 선행연구로서의 역할을 한다.

두 번째, 관광지 만족도에 대한 평점을 분석에 활용할 수 있는 연구 방법을 제시하였다. 만일 리뷰에 대한 내용만 분석 대상으로 하였더라면 리뷰에 존재하는 관광지별 특성을 파악하는 것까지 가능하지만, 리뷰 내용이 관광지 만족도 점수와 어떠한 연관이 있는지 구체적으로 확인하기 어려웠을 것이다. 이에 본 연구는 관광지 만족도 평점을 실증분석에 활용함으로써 관광객의 만족도 평점에 기여하는 개별 관광지 선택속성을 확인하고, 관광지 선택속성의 다양한 원인조건조합을 도출하였다. 본 연구에 사용된 관광지 만족도 평점은 관광객이 예상하는 만족도가 아닌 실제 관광객들이 부여한

만족도 점수이므로 사실적인 관광지 만족도를 나타낸다고 할 수 있으며, 이러한 평점을 연구에 활용한 것은 시의적절한 시도였다.

세 번째, LDA 토픽모델링을 통해 온라인 리뷰에 나타난 주요 관광지 선택속성을 도출하였다. 관광지 만족도를 결정하기에 앞서 관광지 선택속성을 규명하는 것은 매우 중요한 부분이다. 관광지 선택속성을 도출하기 위하여 관광객의 실제 생각과 느낌이 담긴 리뷰 데이터를 사용한 것은 텔과이 기법을 통해 선택속성을 도출하거나 AHP를 적용하여 선택속성의 중요도를 검증하는 연구(김승렬·김창식·곽기영, 2010; 김화룡·백우열, 2015; 신동인·나종민·곽기영, 2016; 김하나·황혜진, 2017; 김화룡, 2019; 이희경·윤병국, 2021) 및 속성에 대한 중요도와 만족도를 분석한 IPA 기반의 연구(김연선·김영식, 2016; 남은경·이승곤, 2018; 조영신, 2018; 김동한, 2019; 남승민·민하나, 2019; 김연선, 2020; 이정학 외, 2021)와는 다른 방식이다.

본 연구의 대상이 되는 관광지 5개소는 모두 자연관광지라는 공통점을 지니지만 모든 관광지에서 동일한 토픽명이 도출된 것은 아니다. 특정 관광지에만 존재하는 토픽도 있으며, 공원·산 자연관광지와 바다·강 자연관광지에 각각 도출되는 토픽도 존재하였다. 또한 관광지 5개소에서 공통으로 도출된 토픽도 있었는데, 이러한 결과는 자연관광지에 대한 관광객들의 전반적인 인식 확인에 도움이 된다. 본 연구를 통해 국문 리뷰 데이터에 존재하는 다양한 관광지 선택속성(토픽)을 확인함과 동시에 왜 이러한 토픽이 도출되었는지 추론함으로써 관광지 방문객이 고려하는 사안에 대하여 관찰할 수 있었다. 즉, 관광지마다 연구자가 고려해야 하는 선택속성에는 차이가 있을 것인데, 본 연구를 통해 텍스트 마이닝 기법 중 하나인 토픽모델링을 함으로써 관광객들이 언급하는 주요 관광지 선택속성을 규명하였다.

네 번째, FsQCA 분석에 대한 활용이다. 본 연구는 일련의 과정을 거쳐서 질적인 자료 형태로 존재하는 리뷰 데이터를 양적인 분석에 활용하기

위한 연구 방법을 제시하였다. FsQCA 분석을 하기 위해 토픽모델링을 할 때 리뷰 별로 할당되는 토픽의 비율을 퍼지 점수로 환산하여 사용하였다. 이를 통해 텍스트 형태로 존재하는 질적인 데이터도 양적인 퍼지 점수로 환산될 수 있음을 밝혔다. 이는 질적인 자료라고 할 수 있는 리뷰 내용도 일련의 과정을 거치면 양적인 자료로 전환되어 다양한 양적 연구로 확장될 수 있음을 의미한다. 실제로 토픽모델링을 통해 얻은 토픽을 변수로 활용하여 인과관계를 분석한 연구가 존재한다. 해당 연구에서는 사용자 리뷰에 대한 토픽모델링을 실시하여 구조방정식으로 연계하고 각 요인들의 중요도를 평가하는 데에 활용하였다(김광국·김용환·김자희, 2018). 그러므로 토픽모델링을 통해서 도출되는 토픽을 변수에 활용하면 회귀분석, 다중회귀분석, 구조방정식 등과 연계하여 다양한 인과관계 검증에 사용될 수 있을 것이다.

결과적으로 본 연구는 관광지 선택속성과 관광지 만족도 간의 관계를 설명할 때 단순한 인과관계를 확인하는 것에 목적을 두지 않고, 집합 이론을 적용한 결합된 인과관계를 살펴보는 것에 초점을 두었다. FsQCA 분석에서는 관광지별로 도출된 토픽이 관광지 만족도를 설명하기 위한 필요조건이 되는지 여부를 확인하였다. 다음으로 관광지 만족도를 결정하는 원인조건조합이 충분조건에 해당하는지 여부를 검증하면서 어떠한 원인조건조합을 통해 관광지 만족도가 형성되는지 확인한 것은 토픽모델링에서 한 단계 발전된 연구로 이어질 수 있는 가능성을 보여준 것이다. 이렇듯 다양한 원인조건조합과 결과변수 간의 관계를 FsQCA 관점에서 살펴본 것은 집합관계 이론을 적용하지 않고 개별 변수 차원에서 관광지 선택속성과 관광지 만족도의 인과관계를 검증하였던 일부 선행연구와 차별되는 부분이다.

## 2. 실무적 시사점

실무적 시사점은 관광지 커뮤니티가 적용할 수 있는 시사점과 관광지를 관리하는 주체가 참고할 수 있는 시사점으로 나누어 살펴볼 수 있다.

첫 번째, 빈도 분석의 결과로 기대할 수 있는 시사점은 다음과 같다. 빈도 분석 결과는 리뷰에서 관광객들이 고려하는 주요 사항이 무엇인지 직관적으로 알게 하는 단서가 된다. 일반적으로 관광객들이 고려하는 주요 단어들을 확인하기 위해 가중치를 부여하지 않은 상태에서 빈도 분석을 하였다. 빈도가 높은 단어의 출현은 곧 관광객들이 자주 언급하는 관광지만의 특성이자 할 수 있으므로, 이러한 단어들을 파악함으로써 특정 관광지에 대한 관광객의 주요 관심사를 접목하여 관광지를 홍보하는 데에 활용할 수 있다.

예를 들어, 관광지 커뮤니티를 운영하는 회사 또는 기업은 빈도 높은 단어들을 해당 관광지를 잘 나타내는 주요 키워드로 소개하는 데에 활용할 수 있다. 한 예로, 설악산 국립공원을 설명하기 위해 # 케이블카 # 설악산 # 단풍 # 바위 # 가을을 주요 키워드로 제시할 수 있다. 이를 반대 관점에서 생각해보면, 빈도 분석의 결과로 상위 도출된 단어들을 검색했을 때 이와 연관되는 관광지명이 검색 결과로 나타나도록 사이트의 검색 알고리즘을 개편할 수 있다. 예를 들어, 어떤 관광지를 여행할 지 결정하지 않은 상태에서 예비 관광객이 케이블카로 유명한 관광지를 찾고자 한다면, ‘케이블카’를 검색어에 입력했을 때 설악산 국립공원이 케이블카에 대한 검색 결과로 표시될 수 있도록 하는 것이다. 마찬가지로 ‘단풍’이라는 키워드를 검색했을 때 이와 관련된 관광지로 설악산 국립공원과 한라산 국립공원을 해당 키워드와 관련된 관광지로 보여줄 수 있다. 이는 빈도 분석의 원리를 적용한 것으로써, 리뷰에 자주 등장하는 단어를 확인하여 관광객들의 관심사와 연관되는 관광지를 검색어에 노출하기 위한 전략으로 활용될 수 있다.

두 번째, 토픽모델링 결과는 해당 관광지를 방문하는 관광객들이 고려하는 관광지 선택속성이 리뷰에서 어떠한 주제로 형성되고 있는지 알게 해준다. 하나의 토픽은 여러 개의 단어로 구성되어 있는데 그 단어들 간의 조합을 보고 토픽을 유추하는 것은 단어만으로 관광지를 이해하는 것이 아닌, 토픽 관점에서 관광객이 고려하는 전반적인 관광지 선택속성을 이해할 수 있도록 한다. 관광지 웹사이트 또는 트립어드바이저와 같은 여행 커뮤니티 운영자는 각 관광지마다 도출된 토픽(관광지 선택속성)을 토대로 관광객들에게 유용한 정보를 제공할 필요가 있다. 한 예로, 설악산 국립공원에서는 토픽모델링에 대한 결과로 ‘케이블카 체험’, ‘사계절 매력성’, ‘입장료 및 경비’, ‘울산바위 명소’, ‘교통수단 접근성’, ‘등산 탐방로 코스’라는 6개의 관광지 선택속성(토픽)이 발견되었다. 토픽모델링에 대한 결과를 실무에 적용하려면 웹사이트 관리자는 도출된 토픽에 근거하여 해당 토픽을 관광지 소개 및 해당 관광지와 관련된 정보 제공에 포함시킬 수 있는 것이다.

예를 들면, ‘케이블카 체험’에 대한 안내문과 케이블카를 통해 기대할 수 있는 시각적 요소에는 무엇이 있는지 설명하는 것, ‘사계절 매력성’을 강조하기 위해 설악산 국립공원에서 계절별로 관찰되는 특징을 언급하는 것, ‘입장료 및 경비’에 대한 내용을 사전에 고지함으로써 예비 방문객이 해당 내용을 사전에 인지하고 방문할 수 있도록 관련 정보를 제공하는 것, ‘울산바위 명소’에 대한 구체적인 설명과 해당 명소를 관광하기 위한 팁을 공유함으로써 관광 명소에 대한 관심을 높이는 것, ‘교통수단 접근성’에 대한 내용과 함께 주변 도로상황 또는 교통수단에 따른 변경사항을 업데이트하여 최신의 정보를 제공하는 것, ‘등산 탐방로 코스’에 대한 자세한 정보 및 관련 후기를 게시하여 관광객들의 흥미를 높이려는 노력 등이 이에 해당한다. 이로써 토픽모델링으로 도출된 토픽들이 해당 관광지를 방문하는 예비 관광객들이 참고할 수 있는 정보체계를 구성하는 작업에 활용될 수 있다.

토픽모델링 결과는 관광객들이 고려하는 주요 관광지 선택속성에 대한 내용이므로 이러한 내용들이 관광지 현장에서도 실무적으로 활용될 수 있다. 예를 들어, ‘케이블카 체험’에 대한 안내 문구와 방향을 표시함으로써 관광객들의 관심과 편의성을 높이는 것, ‘사계절 매력성’을 강조하는 구간을 여러 군데 지정하여 관광객들이 해당 관광지를 여행하는 동안 사계절 매력성을 충분히 느끼고 돌아갈 수 있도록 하는 것, ‘입장료 및 경비’에 대한 부분은 정확한 정보를 제공하지 않으면 민감하게 반응할 수 있는 부분이므로 관광객들의 관광지 이용에 혼란이 생기지 않도록 정확한 정보를 제공하는 것, ‘울산바위 명소’가 실제 리뷰 데이터에서 중요한 토픽으로 발견된 것을 토대로 해당 명소를 안내 책자 및 여행 구간에 소개하여 관광객들의 관심을 높이고 추가적인 관광 명소를 개발하기 위해 관광객들의 피드백을 받는 것, ‘교통수단 접근성’에 대한 내용과 방법을 구체적으로 관광객들이 알 수 있도록 게시하는 것, ‘등산 탐방로 코스’에 대한 최신 정보 제공과 함께 관광객들이 등산 탐방로 코스를 편리하게 경험할 수 있도록 관련 시설을 미리 점검하고 유지보수 하는 것 등이 이에 해당한다. 우선 설악산 국립공원에 대한 결과를 예시로 들어 설명하였으나, 토픽모델링에 대한 관광지별 실무적 시사점 및 활용은 설악산 국립공원 이외에도 본 연구의 대상인 다른 관광지 4개소에서 나타난 결과를 참고하여 동일하게 적용할 수 있다.

세 번째, 관광지 관련 커뮤니티 운영자나 관광지를 관리하는 주체자는 FsQCA 분석에서 필요조건으로 나타난 관광지 선택속성에 주목할 필요가 있다. 설악산 국립공원과 순천만습지(순천만국가정원)에서는 ‘사계절 매력성’이, 남이섬에서는 ‘자연관광 매력성’이 관광지 만족도를 평가하는 필요조건으로 제시되었다. 따라서 이러한 결과는 해당 관광지 선택속성이 실제 리뷰에 거의 반드시 나타나는 내용이라고 할 수 있으므로 관광객 입장에서 매우 중요하게 고려되는 사안임을 인지해야 한다. 이러한 관광지 선택속성

이 훼손되면 높은 만족도를 기대하기 어려울 것이다. 그러므로, 설악산 국립공원과 순천만습지(순천만국가정원)를 대상으로 할 때에는 사계절 매력성에 대한 부분을 충족시키고, 남이섬을 대상으로 할 때에는 자연관광 매력성이 충족될 수 있도록 관광지를 유지 및 관리하려는 노력이 필요하다.

한편, 여행 커뮤니티 또는 웹사이트 운영자는 주요하게 도출된 관광지 선택속성과 관련하여 이 내용을 관광객이 가능한 상세히 작성할 수 있도록 리뷰 운영정책을 개편할 수 있다. 한 예로, ‘케이블카 체험’과 ‘입장료 및 경비’는 설악산 국립공원에서 발견되는 토픽이다. 이는 관광객들이 자주 언급하면서 해당 관광지의 특성을 나타내는 내용이라고 할 수 있으므로, 가급적 최신의 정보가 제공되어야 한다. 이를테면 설악산 국립공원에 대한 리뷰를 작성할 때 주요 관광지 선택속성인 ‘케이블카 체험’과 ‘입장료 및 경비’에 대한 질문을 제시하고, 관광객들이 답변하는 형태로 리뷰를 작성하도록 유도할 수 있다. 이로써 관광지별 토픽모델링 결과를 기반으로 맞춤형 리뷰 정책을 운영할 수 있는 것이다. 이렇게 하였을 때, 관광지 선택속성에 대한 관광객의 생각과 경험을 최신 리뷰로 확인할 수 있음은 물론이며, 해당 리뷰를 읽는 예비 관광객은 해당 관광지에 대한 사실적이며 유용한 정보를 제공받을 수 있다. 이미 관광지를 다녀온 방문객은 다시 한번 해당 관광지의 중요한 선택속성을 떠올리는 계기가 될 수 있으며, 예비 관광객의 입장에서 관광지 선택속성에 대한 구체적이며 유용한 정보를 확인할 수 있다.

본 연구의 대상이 되는 관광지 5개소는 모두 한국을 대표하는 자연관광지로 구성되어 있다. 주요 관광지 선택속성을 파악하는 것은 대한민국 관광산업의 발전과 관광업으로 인한 경제 성장을 이끌기 위해 중요한 부분이다. 연구 결과에 따르면 관광지 만족도는 다양한 원인조건조합을 통해 형성되고 있는 것이 밝혀졌으므로, 여행 관련 커뮤니티와 관광지를 관리하는 주체자는 관광지 만족도를 결정하는 것에 높은 설명력을 가지는 원인조건조합



에 관심을 가질 필요가 있다. 관광지 만족도는 개별 관광지 선택속성으로 결정된다고 전제하기 보다는 다양한 관광지 선택속성의 조합으로 결정됨을 인지해야 한다. 따라서 높은 설명력과 함께 관광지 만족도를 결정하는 관광지 선택속성의 조합(FsQCA 분석에서 원인조건조합으로 제시된 부분)을 확인하고, 이들의 특징을 고려하여서 관광지 만족도를 지속적으로 유지하고 향상할 수 있어야 한다. 관광지 만족도를 결정하는 원인조건조합 특성에 기인하여 관광지 선택속성의 전반적인 품질을 개선할 수 있다면 관광객으로부터 높은 관광지 만족도와 함께 긍정적인 반응을 얻을 수 있을 것으로 예상된다.

### 제 3 절 한계점 및 향후 연구방안

본 연구는 한국관광100선에 5회 연속 선정된 자연관광지를 대상으로 트립어드바이저에서 국문 리뷰 데이터를 수집하여 연구를 진행하였다. 파이썬 언어로 코드를 작성하여 관광지 정보 및 리뷰 데이터를 수집하였으며, 수집된 리뷰는 빈도 분석, TF-IDF 가중치를 적용한 LDA 토픽모델링, FsQCA 분석에 활용하였다. 본 연구는 정성적인 자료라고 할 수 있는 리뷰 데이터를 정량적인 연구로 확장하기 위한 구체적인 방법과 연구 절차를 제시하였다는 점에서 의미가 있다. 그러나 본 연구는 다음의 몇 가지 한계점을 지닌다. 연구의 한계점을 거론하고 향후 연구방안을 제안하면 다음과 같다.

첫 번째, 관광지에 대한 리뷰 데이터 및 관련 정보는 트립어드바이저 같은 여행 커뮤니티 이외에도 블로그나 카페, SNS, 유튜브 등에서 폭넓게 찾아볼 수 있다. 그러나 이러한 매체에는 관광지 만족도 평점이 필수로 제시되어 있지 않다는 이유로 본 연구의 자료 수집 대상에서 제외되었다. 추후 자동화된 크롤링을 이용하여 합법적인 자료 수집을 할 수 있다는 전제하에, 개인 블로그나 여행 리뷰를 전문적으로 다루는 웹사이트, 커뮤니티, SNS, 유튜브 등에서 평점과 함께 리뷰 데이터를 수집할 수 있다면 트립어드바이저 이외의 다양한 매체를 활용하여 수집된 데이터로 관광지 선택속성과 관광지 만족도 간의 관계를 탐구하는 연구에 접목할 수 있을 것이다.

두 번째, 본 연구는 자연어 형태로 존재하는 리뷰 데이터를 분석에 활용하기 위해 수집하였다. 이때, 리뷰를 구성하는 내용은 긍정문과 부정문, 또는 중립문의 형태로 존재할 것인데 이러한 긍정 또는 부정 등의 감성을 고려하지 않았다. 최근에는 관광지와 관련하여 리뷰 데이터를 감성 분석으로 연계하는 선행연구도 일부 확인된다. 한 예로, 제주 복합리조트를 경험한 해외 관광객들이 직접 작성한 온라인 리뷰를 바탕으로 관광객 만족도에 영

향을 미칠 수 있는 중요한 서비스품질을 토픽모델링으로 도출하고, 온라인 리뷰에 담긴 관광객들의 감성을 긍정, 부정, 중립으로 구분하였던 연구가 이에 해당한다(신주하·임희진·이병주, 2021). 텍스트에 드러난 리뷰 작성자의 감성을 분류하거나 수치화하여 객관적인 정보로 바꾸는 감성 분석을 활용한다면, 보다 정교한 차원에서 리뷰 내용의 감성적 특성을 파악하는 연구로 이어질 수 있을 것이다. 또한 본 연구는 리뷰 데이터 수집 과정에서 글자가 아닌 사진이나 영상 등의 매체를 자료 수집 및 분석 대상에 포함하지 않았다. 이후에는 리뷰 내용과 평점 이외에도 부가 정보인 사진과 동영상 등의 자료를 수집하여 분석에 활용하는 연구로 확장되기를 기대한다.

세 번째, 본 연구에 사용된 리뷰 데이터는 약 10년간 다수의 관광객으로부터 얻어진 실제 방문 후기를 포함한다. 따라서 특정 시기에 관찰되거나 시기별로 나타나는 시계열적 특성이 존재할 수 있다. 그러나 본 연구는 시기를 구분할 만큼 리뷰의 수가 충분하지 않으므로 시기나 기간을 구분하지 않고 분석을 진행하였다. 일부 선행연구에서는 토픽모델링을 하면서 시계열 특성을 함께 고려하는 것을 확인할 수 있다(김창식·곽기영·윤혜진, 2017; 정원준, 2018; 방한솔·문호석, 2019; 양경은·노법래, 2020). 이처럼 장기간의 리뷰 데이터를 수집하여 분석을 진행하고자 한다면, 분기별 또는 시기별로 구간을 나누어 연구를 진행하는 시계열 연구로 확장될 수 있을 것이다.

마지막으로 본 연구는 자료 수집의 대상을 내국인으로 한정하였다. 그러나 지속적으로 유입되는 방한 외래 관광객의 비중을 고려한다면 외국인 관광객들이 국내 관광지를 방문하여 작성하는 리뷰의 내용 또한 관광지에 대한 전반적인 인식을 알아보기 위해 중요하다. 외국인 방문객이 작성하는 리뷰에서 발견되는 관광지 선택속성을 함께 규명할 수 있다면 내국인이 고려하는 관광지 선택속성과 외국인이 고려하는 관광지 선택속성들 간의 유사점과 차이점을 확인하는 연구로 발전될 수 있다. 나아가 내·외국인 별로

관심 가지는 관광지 선택속성에는 차이점이 존재할 수 있는데, 이를 FsQCA 분석에 적용하면 내국인과 외국인의 관광지 만족도를 결정하는 관광지 선택속성(원인조건)의 조합을 함께 규명함으로써 관광지가 주목해야 하는 관광지 선택속성 확인에 유용한 단서를 얻을 수 있을 것으로 기대한다.

## 참 고 문 헌

### 1. 국내문헌

- 강봉준·김재환 (2017). fsQCA로 살펴본 서비스업 관리시스템의 업무조정 효과성 결정요인. 『서비스경영학회지』, 18(3), 279-298.
- 강삼·추승우 (2021). 중국인 관광객의 관광지 선택속성이 만족도, 재방문의도 및 추천의도에 미치는 영향에 관한 연구: 온라인 여행사를 이용한 부산 방문 개별관광객을 중심으로. 『관광레저연구』, 33(1), 43-61.
- 강지원·남궁영 (2021). 빅데이터를 활용한 식품 유통 플랫폼에 대한 소비자 인식 분석: 텍스트 마이닝과 의미연결망 분석을 중심으로. 『호텔경영학연구』, 30(2), 37-52.
- 고선영·정근오 (2021). 지역관광 빅데이터 정책성과와 과제-제주특별자치도를 사례로. 『한국산학기술학회 논문지』, 22(3), 579-586.
- 고선희·박은숙 (2011). 관광지 매력성이 지각된 가치 및 충성도에 미치는 영향: 방한 일본인 관광객을 중심으로. 『한국콘텐츠학회논문지』, 11(7), 467-477.
- 곡지평·김명중·유승동 (2019). 관광지 이미지가 관광만족도 및 재방문의도에 미치는 영향 관계: 내국인 및 외국인 방문객의 조절효과 검증. 『호텔경영학연구』, 28(8), 231-243.
- 곽민정·최지유·박소현 (2019). 호텔 서비스 속성별 고객만족도 분석을 위한 온라인 리뷰 감성분석. 『관광경영연구』, 23(4), 1-25.
- 권태일·이충희 (2017). 관광분야 빅데이터 활용체계 및 실증분석 연구. 한국문화관광연구원.
- 김광국·김용환·김자희 (2018). 사용자 리뷰 토픽분석을 활용한 모바일 쇼핑 앱 고객만족도에 관한 연구. 『한국전자거래학회지』, 23(4), 41-62.
- 김규환·박인권 (2018). 퍼지셋 질적비교분석을 이용한 우리나라 지역혁신의 유형 및 요인 분석. 『지역연구』, 34(4), 3-18.

- 김기만·이수연 (2021). 이산화탄소 배출의 원인 조건으로서 사회경제적 요인, 기술적 요인의 역할에 관한 연구: 퍼지셋 질적 비교 분석을 활용한 OECD 국가 비교 분석. 『한국혁신학회지』, 16(1), 1-34.
- 김난영 (2012). 관광지 이미지, 관광자 만족, 관광지 충성도의 구조적 관계에 관한 연구-아산시를 중심으로. 『동북아관광연구』, 8(1), 101-122.
- 김도훈·김영규 (2017). 순수 관광목적지 가진 중국인 개별자유여행객의 한국관광지 선택속성 중요도-만족도 연구. 『관광연구』, 32(4), 219-236.
- 김도훈·김지훈 (2015). 국가 이미지와 관광목적지 선택속성, 관광객만족의 구조적관계 연구. 『관광연구』, 30(3), 415-430.
- 김도희·김민정 (2021). 텍스트마이닝 기법을 활용한 허위·과장광고 관련 기사의 트렌드 분석(1990-2019). 『한국콘텐츠학회논문지』, 21(4), 38-49.
- 김동일·이승곤·진현식 (2012). 국외여행상품 선택속성에 따른 여행객 만족이 여행사에 대한 태도, 충성도 및 전환 의도에 미치는 구조관계 분석: 여행사 규모에 따른 차이를 중심으로. 『관광연구저널』, 26(4), 309-333.
- 김동한 (2019). 도시관광 선택속성 개선에 관한 연구: 서울 코엑스몰을 중심으로. 『관광레저연구』, 31(8), 5-19.
- 김미라 (2020). 자연관광이 치유체험, 만족 및 재방문의도에 미치는 영향. 『Tourism Research』, 45(1), 59-79.
- 김민식·한학진·박병화 (2018). 트립어드바이저를 이용한 국내 목적지별 특성분석: 인바운드 관광객을 대상으로. 『관광레저연구』, 30(2), 5-19.
- 김민철·부창산·김영훈 (2008). 제주지역 내 중국 및 일본 관광객의 선택속성의 차이 분석: 구조방정식 이용. 『한국지역지리학회지』, 14(2), 126-140.
- 김범진·권태길·고호석 (2014). 관광지 유형에 따른 관광지 속성 및 제약요인이 관광지 선택에 미치는 영향. 『지역산업연구』, 37(4), 237-257.
- 김보섭·조수현·박민식·강필성 (2016). 관광지 평가 댓글을 활용한 만족도 요인 추출과 관광지 평가. 『대한산업공학회 추계학술대회 논문집』, 375-382.
- 김성민·이훈 (2020). 텍스트 마이닝 기법을 활용한 한류 영상물 매체 이용자 인식 분

- 석-잠재 한류 관광객을 중심으로. 『관광연구논총』, 32(4), 153-179.
- 김성혁·오은지·권상미 (2009). 관광지선택속성에 대한 세대별 비교: 서울지역을 중심으로. 『관광레저연구』, 21(2), 167-182.
- 김슬기·김장현 (2019). 의미연결망 분석을 활용한 영화 리뷰 시각화. 『한국정보통신학회논문지』, 23(1), 1-6.
- 김승렬·김창식·곽기영 (2010). AHP 기법을 이용한 몽골지역을 방문하는 한국인 국외관광객의 여행상품 선택에 관한 연구. 『관광레저연구』, 22(3), 451-465.
- 김승환·임혜빈·이병관 (2021). 온라인 리뷰어와의 지각된 유사성이 제품 구매 의도에 미치는 영향: 리뷰 진단성의 매개효과와 심리적 파워의 조절된 매개효과. 『한국심리학회지: 소비자·광고』, 22(1), 49-79.
- 김연선 (2020). IPA 기법을 활용한 한옥마을 선택속성의 중요도-만족도 연구. 『한국콘텐츠학회논문지』, 20(7), 585-593.
- 김연선·김영식 (2016). IPA 를 활용한 주요 관광지 선택속성의 중요도-성취도 연구: 전라북도 관광지 방문객을 중심으로. 『관광레저연구』, 28(12), 25-41.
- 김예린 (2018). 아동·청소년 정신건강서비스 네트워크의 효과성 영향요인 분석: 통합서비스의 관점에서. 『한국조직학회보』, 15(1), 101-135.
- 김은미 (2021). 온라인 리뷰의 감성분석과 순환신경망을 적용한 국내 인바운드 관광수요 예측 모형. 『관광연구저널』, 35(3), 69-79.
- 김종성 (2017). 스포츠관광 활성화와 빅데이터 활용에 관한 연구. 『한국엔터테인먼트산업학회논문지』, 11(3), 99-109.
- 김종일 (2010). 대륙복지국가의 활성화정책 추이에 관한 퍼지 집합 이념형 분석, 2000-2007. 『사회보장연구』, 26(2), 253-284.
- 김주진·신우진 (2020). 관광지에 대한 선택 속성 및 만족도와 관광객 행동의도 간 영향관계 연구: 전라북도 남원관광단지를 사례로. 『한국지역개발학회지』, 32(1), 143-166.
- 김지현 (2020). 남성 육아휴직 사용에 대한 결합요인 분석: 퍼지셋 질적비교분석을 통한 국제비교연구. 『비판사회정책』, (66), 45-92.
- 김지훈 (2020). 웰니스 관광지의 선택속성에 따른 중요도 만족도 연구: 광주·전남 지

- 역을 중심으로. 『관광연구저널』, 34(11), 21-35.
- 김창식·곽기영·윤혜진 (2017). 관광분야 연구동향 분석: 토픽모델링과 시계열분석을 중심으로. 『관광레저연구』, 29(12), 25-39.
- 김태호·김학선 (2017). 빅데이터를 활용한 음식관광의 의미연결망 분석에 관한 연구. 한국관광레저학회 학술발표대회, 372-379.
- 김하나·황혜진 (2017). AHP를 이용한 복합리조트 선택속성 우선순위도출에 관한 연구: 아웃바운드 관광객을 중심으로. 『관광레저연구』, 29(11), 163-180.
- 김학선 (2017). 빅데이터를 활용한 음식관광관련 의미연결망 분석의 탐색적 적용. 『Culinary Science & Hospitality Research』, 23(4), 22-32.
- 김현철 (2018). 관광지 선택속성과 지각된 이미지에 따른 관광태도와의 관계 연구. 『호텔관광연구』, 20(4), 28-40.
- 김현철 (2019). 관광지 선택속성이 관광만족도에 미치는 영향에 있어 향토음식이미지의 조절효과: 기장군을 중심으로. 『Tourism Research』, 44(4), 151-169.
- 김현·장호성 (2012). 관광지 선택속성과 동기요인이 방문객 행동의도에 미치는 영향: 태안지역 방문객을 중심으로. 『지방정부연구』, 16(1), 7-22.
- 김홍범·심영석 (2018). 연관규칙 마이닝 기법을 활용한 관광지 이미지 빅데이터 분석에 관한 연구. 『관광레저연구』, 30(12), 57-76.
- 김화룡 (2019). 스포츠관광 선택속성 요인 도출 및 우선순위 분석. 『관광연구』, 34(2), 133-150.
- 김화룡·백우열 (2015). AHP 기법을 이용한 캠핑장 선택속성에 대한 우선순위 분석. 『한국사회체육학회지』, 60, 363-373.
- 김효중·김시중 (2012). 대전 뿌리공원 관광지 선택속성이 관광객 만족 및 행동의도에 미치는 영향 연구-관광지 이미지를 조절효과로. 『한국경제지리학회지』, 15(4), 735-751.
- 김희정·김시중 (2012). 관광지 선택속성과 경험적 가치가 행동의도에 미치는 영향: 백제역사유적지구를 대상으로. 『국토지리학회지』, 46(2), 147-159.
- 남승민·민하나 (2019). 국내 열린관광지에 대한 인식 조사 연구: IPA 분석을 활용하여. 『관광레저연구』, 31(10), 45-57.



- 남승주·김준환·유영준 (2020). 텍스트 분석 기반의 쇼핑 관광객 인식 분석. 『신산업경영저널』, 38(2), 3-21.
- 남은경·이승곤 (2018). 수정 IPA를 활용한 역사문화관광지 선택속성 평가 연구: 대가야역사문화관광지 고령군 관광객을 대상으로. 『관광연구저널』, 32(7), 171-182.
- 노희경 (2019). 통일관광을 위한 비무장지대 (DMZ) 의 관광학적 가치에 대한 인식 변화 연구: 소셜미디어 빅데이터를 중심으로. 『관광레저연구』, 31(10), 59-78.
- 류인평·최인경 (2014). 문화유산관광지 선택속성, 만족도, 행동의도 연구: 익산미륵사지 중심으로. 『관광연구』, 29(4), 413-436.
- 문화체육관광부 (2020). 2020국민여행조사 분석편.
- 문화체육관광부 (2021). 보도자료 : 2021~2022 한국관광 100선을 소개합니다.
- 민기채 (2014). 퍼지셋 질적비교분석의 사회과학적 활용: Stata를 활용한 Y-검증과 N-검증을 중심으로. 『인문사회과학연구』, 44(-), 223-259.
- 박경열·안희자 (2019). 텍스트 마이닝을 활용한 DMZ 관광 이슈의 토픽 모델링 분석. 『관광레저연구』, 31(4), 143-159.
- 박경열·한혜림·최승담 (2019). 빅데이터를 활용한 사드 배치 전후 중국관광객의 한국관광 인식 변화 비교: 의미 연결망 분석을 중심으로. 『관광레저연구』, 31(2), 25-43.
- 박대민 (2016). 장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26 년. 『한국언론학보』, 60(5), 353-407.
- 박로운·이기훈 (2016). 빅데이터 분석을 통한 농촌관광 실태와 활성화 방안 연구: 전라북도를 중심으로. 『한국지역사회생활과학회지』, 27(S:특별호), 665-679.
- 박상현·이미순 (2019). Dark Tourism (다크투어리즘) 관광객의 선택속성과 방문동기가 관광지 만족도와 재방문 의도에 미치는 영향에 관한 연구: 부산 UN 평화공원 및 기념관을 중심으로. 『관광레저연구』, 31(4), 5-24.
- 박선심 (2016). 관광객의 관광지 선택속성이 지각된 가치, 자기효능감, 관광만족 및 재방문의도에 미치는 영향. 동신대학교 박사학위논문.
- 박영옥·정규엽 (2021). DMR (Dirichlet Multinomial Regression) 토픽모델링을 이용한 온라인 리뷰 빅데이터 기반 고객감성 분석에 관한 연구: 국내 5성급 호텔의 외

- 국민 이용객 리뷰를 중심으로. 『호텔경영학연구』, 30(2), 1-20.
- 박예림 · 강영옥 · 김동은 · 이주윤 · 김나연 (2019). 플리커 데이터의 텍스트마이닝을 통한 서울방문 외국인 관광객의 서울 이미지 분석. 『대한공간정보학회지』, 27(1), 11-23.
- 박원우 · 손승연 · 박해신 · 박혜상 (2010). 적정 표본크기 (sample size) 결정을 위한 제안. 『Seoul Journal of Industrial Relations』, 21.
- 박정환 · 이병철 (2014). 온라인과 오프라인 리뷰 비교분석-박람회 서비스 품질을 중심으로. 『관광경영연구』, 18(1), 61-79.
- 박태수 (2020). 해양관광 활성화 방안 도출을 위한 빅데이터 분석 연구. 『관광레저연구』, 32(7), 179-194.
- 박혜영 (2019). 시니어세대의 여행욕구 유형에 따른 관광지 선택속성과 관광만족에 관한 연구: 한·일 비교를 중심으로. 『MICE 관광연구』, 19(1), 85-104.
- 방한솔 · 문호석 (2019). 텍스트마이닝을 이용하여 텍스트의 주요 토픽을 시계열적으로 표현하는 방법론 연구. 『한국데이터정보과학회지』, 30(6), 1259-1276.
- 봉미희 · 한주형 (2020). 내국인 관광객의 도시 관광지 선택속성 중요도-만족도 연구: 서울지역을 중심으로. 『이벤트컨벤션연구』, 38(-), 81-100.
- 서정아 · 오익근 (2017). 사회연결망 분석을 활용한 대구의 관광지 이미지 분석: 온라인 빅데이터를 중심으로. 『한국콘텐츠학회논문지』, 17(7), 443-454.
- 소네다 · 토루 · 최승담 (2019). 한·일 관광객의 온천관광지 선택속성의 비교분석-Hofstede의 문화차원을 중심으로. 『관광연구논총』, 31(2), 27-44.
- 손병모 (2011). 관광지 선택속성이 관광객의 지각된 가치와 충성도에 미치는 영향. 『관광연구』, 26(1), 201-228.
- 송래현 (2021). 관광지 관광정보제공 인식에 따른 관광경험 만족도 및 충성도. 『호텔관광연구』, 23(3), 67-78.
- 송옥방 · 차석빈 (2020). 서울 도시 관광 선택속성이 전반적 만족도에 미치는 대칭 및 비대칭 영향 분석: 방한 경험이 있는 중국 현지인을 중심으로. 『관광레저연구』, 32(3), 41-56.
- 송학준 (2012). 문화관광지 선택속성, 지각된 가치, 만족간 구조적 관계 연구: 경복궁 관광객을 대상으로. 『호텔경영학연구』, 21(5), 219-236.

- 신동인 · 나종민 · 광기영 (2016). 카라반 야영장 활성화를 위한 우선순위 및 가중치 분석: AHP 기법의 적용. 『관광레저연구』, 28(11), 419-434.
- 신동일 (2014). 빅데이터(Bigdata)활용에 따른 제주관광의 효과 및 과제. 제주발전연구원, 『정책이슈브리프』, no:203, 1-16.
- 신민정 · 원문가 · 송학준 (2018). 동기, 선택속성, 관광만족, 재방문의도, 생활만족간 인과 관계에 대한 연구: 공정여행 여행객을 중심으로. 『호텔경영학연구』, 27(6), 1-16.
- 신주하 · 임희진 · 이병주 (2021). 국내 복합리조트 서비스품질에 대한 토픽모델링 및 감성분석. 『관광연구저널』, 35(11), 191-206.
- 신형섭 (2019). 도시관광지 선택속성에 따른 관광객의 지각된 가치가 만족 및 행동의도에 미치는 영향: 익산을 방문한 관광객을 대상으로. 『호텔관광연구』, 21(2), 45-60.
- 심영석 · 김홍범 (2016). 텍스트 마이닝을 이용한 관광지 이미지 구성요인 및 측정에 관한 연구. 『관광학연구』, 40(7), 221-245.
- 심우석 (2015). 관광자원 유형과 선택속성, 관광만족, 상품개발 선호 연구. 전주대학교 박사학위논문.
- 안성식 · 박효정 · 조아람 (2018). 기후변화에 따른 국제관광산업 경쟁력 강화를 위한 관광지 선택속성에 대한 중요도 차이 연구: 컨조인트 분석 활용을 통해. 『통상정보연구』, 20(3), 193-213.
- 안세길 (2020). 전주 한옥마을 방문객의 전통문화 가치인식과 관광지 선택속성이 만족도 및 행동의도에 미치는 영향. 『호텔리조트연구』, 19(2), 121-140.
- 안진현 · 임동혁 (2020). 이벤트와 관련된 주변 관광지 자동 추천 알고리즘 개발. 『한국산학기술학회 논문지』, 21(3), 407-413.
- 양경은 · 노법래 (2020). 한국 다문화 담론 구조와 그 시계열적 변동-언론 기사문 텍스트 마이닝 분석을 중심으로. 『한국사회복지학』, 72(3), 33-58.
- 양고운 · 박형준 (2013). 지방정부 간 자율적 행정구역 통합의 성공요인 탐색: 퍼지집합 질적비교분석(fsQCA)의 적용. 『한국지방자치학회보』, 25(1), 91-116.
- 여영숙 (2016). 중요도-성취도 분석을 이용한 담양죽녹원의 관광매력속성 평가에 관한 연구. 『관광레저연구』, 28(10), 169-189.

- 염수길 · 이재섭 (2016). 객관적 지식수준 차이에 따른 도시관광 선택속성의 중요도와 성취도에 관한 연구. 『관광레저연구』, 28(7), 37-56.
- 오승욱 · 한진욱 · 김민수 (2019). 소셜 빅데이터 분석을 활용한 수상레저스포츠관광에 대한 인식 분석. 『한국스포츠산업경영학회지』, 24(4), 83-95.
- 오익근 · 이태숙 · 전채남 (2015). 빅데이터 분석을 통한 한국관광 인식에 관한 연구. 『관광학연구』, 39(10), 107-126.
- 오정호 (2001). 성공적인 인터넷 확산의 구조적 필요조건과 충분조건-퍼지집합분석. 『언론과 사회』, 9(4), 130-172.
- 우경숙 · 서주환 (2020). 빅데이터를 활용한 도시공원 이용행태 특성의 시계열 분석. 『Journal of the Korean Institute of Landscape Architecture』, 48(1), 35-45.
- 유광훈 · 남지순 (2017). DECO 감성사전과 LGG 문법에 기반한 관광 숙박 온라인 후기글의 감성 분석 연구. 『한국사전학』, (30), 100-154.
- 유순호 (2020). 한식 이용고객 관광지선택속성이 만족 및 한류관광 행동의도에 미치는 영향. 『관광경영연구』, 96(-), 455-474.
- 윤정현 (2010). 가족관광객의 관광동기, 관광지 선택속성 및 관광만족의 영향관계. 『관광레저연구』, 22(5), 283-301.
- 윤정현 (2015). 일과 생활의 조화가 가족관광 및 생활만족에 미치는 구조적 영향관계. 『관광레저연구』, 27(5), 373-394.
- 윤홍근 (2013). 문화산업에서 빅데이터의 활용방안에 관한 연구. 『글로벌문화콘텐츠』, 10(-), 157-179.
- 이경숙 · 전보람 (2018). 관광지선택속성이 지각된 가치 및 행동의도에 미치는 영향: 방한 중국인관광객을 대상으로. 『관광경영연구』, 22(7), 583-601.
- 이경숙 · 하흠 (2018). 방한 중국 관광객의 관광지 선택속성이 행동의도에 미치는 영향. 『동북아관광연구』, 14(2), 173-196.
- 이광국 · 김경옥 · 전재균 (2016). 관광지 선택속성에 따른 시장세분화 연구. 『Journal of The Korean Data Analysis Society』, 18(6), 3239-3249.
- 이동혁 · 이서희 (2020). 정신전력 기고문의 토픽모델링 적용 탐색적 연구-국방일보 오

- 피니언 분석을 중심으로. 『정신전력연구』, (61), 103-133.
- 이범국·손창호 (2020). 사용자 리뷰 데이터를 활용한 모바일 어플리케이션 서비스 평가 척도 개선. 『한국산학기술학회 논문지』, 21(1), 380-386.
- 이범준·노희경 (2020). 토픽 모델링 기법을 활용한 음식관광 시장 세분화: 소비자 추구편익을 중심으로. 『호텔경영학연구』, 29(4), 187-204.
- 이수희·박득희·김맹선 (2018). 소셜 빅데이터 분석을 활용한 제주도 관광 인식에 관한 연구: 한한령 이전과 이후 비교. 『관광레저연구』, 30(2), 55-75.
- 이승운 (2014). 퍼지셋 질적 비교연구 방법론의 이론과 적용: Fs/QCA 입문. 고려대학교출판부.
- 이익수 (2016). 관광동기에 따른 관광지 선택속성의 차이 및 만족도에 미치는 영향: 중국인 관광객을 중심으로. 『한국콘텐츠학회논문지』, 16(2), 365-373.
- 이정미·류미나·임규건 (2018). 한국관광 실태조사 빅 데이터 분석을 통한 관광산업 활성화 방안 연구. 『지능정보연구』, 24(2), 149-169.
- 이정실 (2020). 도시관광 선택속성의 중요도-만족도 분석: 부산시 중구를 중심으로. 『관광레저연구』, 32(8), 19-34.
- 이정학·김대웅·이지혜·김민준 (2021). 지역특화 스포츠관광지 육성을 위한 스포츠관광 이미지 속성 연구: IPA 분석을 활용하여. 『한국체육과학회지』, 30(5), 639-652.
- 이종길·양재식 (2013). 해양레저스포츠 참가자의 관광지 선택속성에 대한 중요도-실행도 분석: 충남 보령지역을 중심으로. 『한국여가레크리에이션학회지』, 37(1), 89-104.
- 이철진 (2008). 리조트 관광객 심리, 선택속성과 관계지향성 연구. 경기대학교 박사학위논문.
- 이현중 (2020). 수정된 중요도-실행도 분석을 이용한 관광목적지 선택속성 연구. 『관광연구저널』, 34(12), 17-28.
- 이혜윤·박병식·남궁금순 (2020). 서울시 청년일자리 정책에 관한 연구: 구조화된 설문조사 자료에 퍼지셋 질적 분석을 적용하여. 『한국사회와 행정연구』, 31(2), 185-222.
- 이희경·윤병국 (2021). 중국인의 한국 미용성형 의료관광 선택속성에 대한 연구: 중국인 의료관광 전문가를 중심으로. 『관광연구저널』, 35(5), 37-52.
- 임영희·김홍범 (2019). 호텔 온라인 리뷰 빅데이터를 활용한 감성분석에 관한 연구. 『호텔경영학연구』, 28(7), 105-123.

- 임지혜 (2021). "등산만 할건데 왜 돈 받나"...단풍철 '사찰 입장료' 또 도마 위. URL : <https://www.kukinews.com/newsView/kuk202110130174>
- 임화순·강영순 (2008). 관광행태별 관광자원 선호도에 관한 연구: 제주지역 관광객을 중심으로. 『관광레저연구』, 20(2), 229-248.
- 임화순·고계성 (2006). 관광목적지 매력속성에 따른 중요도-성취도 연구: 일본 오키나와 (Okinawa) 중심으로. 『관광연구』, 21(1), 177-190.
- 장양례·김혜영 (2011). 개별 여행객들의 (FIT) 관광지 선택속성, 관광계약, 선호관광 활동에 관한 실증적 연구. 『관광연구』, 26(1), 331-350.
- 장윤정·정수아 (2021). 빅데이터를 활용한 외국인 관광객의 국가 그룹별 방문특성 및 유사성: 트립어드바이저 자료를 중심으로. 『관광연구저널』, 35(3), 95-109.
- 전기홍·이종관·김남기 (2014). 관광지의 선택속성이 도시 브랜드 이미지에 미치는 영향: 전주시를 대상으로. 『문화산업연구』, 14(4), 169-175.
- 전승봉 (2020). 직무긴장, 삶의 만족도, 그리고 가사노동 분배가 OECD 국가들의 출산율에 미치는 영향. 『한국콘텐츠학회논문지』, 20(8), 251-261.
- 정균영 (2020). 젠트리피케이션 관광지 선택속성, 장소애착, 관광태도와의 영향관계. 호남대학교 박사학위논문.
- 정용복 (2020). 제주도 중문관광단지 방문객의 선택속성이 관광만족과 재방문 의도에 미치는 영향 연구. 『제주도연구』, 53(-), 159-185.
- 정원준 (2018). 사드 (THAAD) 이슈를 둘러싼 한국과 중국 간 갈등 쟁점의 변화 추이 연구: 빅데이터를 이용한 시계열 토픽 모델링과 언어 네트워크 분석 기법으로. 『한국광고홍보학보』, 20(3), 143-196.
- 정지원 (2004). 관광이벤트 참가자 만족·불만족과 사후행위의도에 관한 연구. 경기대학교 박사학위논문.
- 정지훈·정혜인·이준기 (2021). 텍스트마이닝 기법과 ARIMA 모형을 활용한 배달의 민족 앱 리뷰 분석. 『한국디지털콘텐츠학회 논문지』, 22(2), 291-299.
- 정효숙 (2017). 다문화가족의 관광동기와 관광지선택속성이 행동의도에 미치는 영향 비교분석. 경희대학교 박사학위논문.

- 조경훈·박형준 (2015)a. 퍼지셋 질적비교사례 2단계 분석법의 정책분석에 활용: 입지 갈등 지속기간의 원인탐색. 『정책분석평가학회보』, 25(4), 311-333.
- 조경훈·박형준 (2015)b. 폐기물 처리시설 입지 갈등의 지속 요인 연구-퍼지셋 질적 비교연구방법을 중심으로. 『한국행정연구』, 24(2), 29-64.
- 조영신 (2018). IPA를 적용한 종교관광 선택속성 탐색. 『호텔경영학연구』, 27(7), 15-29.
- 진동봉·박현지 (2021). 관광목적지 선택속성, 관광목적지 이미지, 관광만족, 재방문의도간의 영향 관계분석: 중국 국내관광객들을 대상으로. 『관광레저연구』, 33(1), 5-26.
- 진문도·주은지·윤해진 (2019). 중국 개별관광객의 관광지 속성에 대한 중요도-만족도 분석. 『MICE 관광연구』, 19(3), 7-21.
- 차승희·김시은·최희지·김우혁 (2020). 축제선택속성이 관광지 이미지 및 재방문 의도에 미치는 영향-인적서비스의 조절효과를 중심으로. 『관광연구논총』, 32(3), 57-75.
- 최승묵 (2020). 방한 단체관광 여행상품의 목적지 구성 특성별 만족도 분석-방문 광역지자체 수 및 구성을 중심으로. 『관광학연구』, 44(2), 73-92.
- 최신애·여정성 (2010). 정책평가방법의 비교분석: 표시, 광고규제를 중심으로. 『조사연구』, 11(3), 57-79.
- 최영준 (2009). 사회과학에서 퍼지셋 활용의 모색: 퍼지 이상형 분석과 결합 요인 분석을 중심으로. 『정부학연구』, 15(3), 307-336.
- 최일선·홍장원·이정아 (2019). 섬 관광객의 관광지 선택속성과 관광제약이 관광 만족도와 재방문 의도에 미치는 영향. 『관광연구논총』, 31(4), 121-138.
- 최자영·김현아·김용범 (2020). 온라인 리뷰가 매출에 미치는 영향력 분석: 텍스트 기반 감성지수를 중심으로. 『유통연구』, 25(3), 1-21.
- 트린티하인·설훈구 (2018). 제주도 방문 베트남 관광객의 개인가치와 관광지 선택속성만족이 재방문의도에 미치는 영향. 『관광레저연구』, 30(11), 185-204.
- 한국관광공사 (2021). 코로나19가 바꿔놓은 대한민국 관광지도. URL : <https://kto.visitkorea.or.kr/kor/notice/news/press/board/view.kto?id=442962&isNotice=false&instanceId=42&num=1>
- 한국콘텐츠진흥원(2014). 빅데이터 시장 현황과 콘텐츠산업 분야에 대한 시사점. 『KOCCA포커스』, 77(-), 1-19.

- 한상겸 (2014). 롯데월드 선택속성이 만족도와 행동의도에 미치는 영향. 『관광진흥연구』, 2(2), 59-74.
- 한승현·강민아 (2015). 내전 후 평화의 달성을 위한 정부의 질 (quality of government): 질적비교연구방법론 (QCA) 의 적용. 『국제정치연구』, 18(1), 21-57.
- 한영위·이용기·안성만 (2017). 커피전문점 선택 속성과 점포유형의 결합 관계가 만족도에 미치는 영향: 퍼지셋 질적비교분석 (fsQCA)을 중심으로. 『프랜차이즈경영연구』, 8(1), 31-41.
- 한장현·서현 (2019). 소셜 빅데이터 분석을 활용한 여수관광 인식 및 관광활성화 방안에 관한 연구. 『관광레저연구』, 31(7), 81-96.
- 한장현·조윤희 (2018). 빅데이터 분석을 통한 아쿠아리움 인식에 관한 연구. 『관광레저연구』, 30(8), 341-355.
- 행정자치부 (2015). 보도자료: 정부 3.0, 빅데이터 분석을 통해 정책효과 높인다.
- 황수정 (2021). 트립어드바이저 순위의 영향요인 탐색을 통한 호텔의 효과적인 온라인 평판관리전략에 관한 연구. 경희대학교 석사학위논문.
- 황해정·심혜린·최준호 (2016). 빅데이터 분석을 활용한 사용자 경험 평가 방법론 탐색: 아마존 에코에 대한 온라인 리뷰 분석을 중심으로. 『한국콘텐츠학회논문지』, 16(8), 517-528.

## 2. 국외문헌

- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993-1022.
- Chen, C. F., & Chen, F. S. (2010). Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists. *Tourism management*, 31(1), 29-35.
- Grofman, B., & Schneider, C. Q. (2009). An introduction to crisp set QCA, with a comparison to binary logistic regression. *Political Research Quarterly*, 62(4), 662-672.



- Guy, I., Mejer, A., Nus, A., & Raiber, F. (2017, April). Extracting and ranking travel tips from user-generated reviews. In Proceedings of the 26th international conference on world wide web (pp. 987-996).
- Hsu, T. K., Tsai, Y. F., & Wu, H. H. (2009). The preference analysis for tourist choice of destination: A case study of Taiwan. *Tourism management*, 30(2), 288-297.
- Jang, S. S., & Feng, R. (2007). Temporal destination revisit intention: The effects of novelty seeking and satisfaction. *Tourism management*, 28(2), 580-590.
- Kim, D. W. (2013). Big Data Use Cases of the Sector. *Business Review of Dong-a University Business Research Center*, 34, 39-52.
- Kim, J., & Baek, S. (2016). Analysis of issues on the college and university structural reform evaluation using text big data analytics. *Asian J Educ*, 17(3), 409-36.
- Kotler, P., Makens, J. C., Bowen, J. T., & Baloglu, S. (2018). *Marketing for hospitality and tourism*. Pearson Education.
- Kvist, J. (1999). Welfare reform in the Nordic countries in the 1990s: Using fuzzy-set theory to assess conformity to ideal types. *Journal of European Social Policy*, 9(3), 231-252.
- Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*. 26(5), 727-750.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, 29(3), 458-468.
- Liu, X., Schuckert, M., & Law, R. (2015). Can response management benefit hotels? Evidence from Hong Kong hotels. *Journal of Travel & Tourism Marketing*, 32(8), 1069-1080.
- Martilla, J. A., & James, J. C. (1977). Importance-performance analysis. *Journal of marketing*, 41(1), 77-79.
- Mayor, E., & Jarvis, L. (1981). *The Psychology of Leisure Travel: Effective Marketing and Selling of Travel Services*. MI: CBI Publishing Company.
- PWC (2012). Big data - The next frontier for innovation. PricewaterhouseCooper. 1-16.

- Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.
- Ragin, C. C. (2009). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.
- Ragin, C. C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. Univ of California Press.
- Ragin, C. C., & Sonnett, J. (2004). *Between Complexity and Parsimony: Limited Diversity. Counterfactual Cases, and Comparative Analysis*.
- Rihoux, B. (2006). Qualitative comparative analysis (QCA) and related systematic comparative methods: Recent advances and remaining challenges for social science research. *International Sociology*, 21(5), 679-706.
- Schuckert, M., Liu, X., & Law, R. (2016). Insights into suspicious online ratings: direct evidence from TripAdvisor. *Asia Pacific Journal of Tourism Research*, 21(3), 259-272.
- Shi, M., Zhu, W., Yang, H., & Li, C. (2016). Applying semantic web and big data techniques to construct a balance model referring to stakeholders of tourism intangible cultural heritage. *International Journal of Computer Applications in Technology*, 54(3), 192-200.
- Tripadvisor (2019). URL : [www.tripadvisor.com](http://www.tripadvisor.com)
- Woodside, A. G. (2014). Embrace • perform • model: Complexity theory, contrarian case analysis, and multiple realities. *Journal of Business Research*, 67(12), 2495-2503.
- Yi, Y. K. (2009). Visitor satisfaction and awareness for conservation in world cultural heritage place; In case of Haeinsa (海印寺). *Journal of Korean Institute of Traditional Landscape Architecture*, 46-58.
- Zadeh, L. A., Klir, G. J., & Yuan, B. (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers* (Vol. 6). World Scientific.

## **Abstract**

A research on the tourists' destination selection attributes and satisfaction : Using text mining and fuzzy-set qualitative comparative analysis

Jiwon Park

Department of Hospitality &

Tourism Management

The Graduate School

Sejong University

According to 2020 Korean National Tourism Survey, Koreans choose activities that allow them to appreciate nature and its scenery when traveling. Not only the general public interest in, but also their demand for natural tourism is growing, hence natural attractions represent a significant part of the domestic tourism market. Research of natural tourism has been a key topic in academia along with COVID-19. The purpose of this research is to investigate how individual tourists evaluate overall satisfaction while traveling, especially in when they visit the five well-known natural tourist sites in South Korea.

Online review data for tourist destination is one of that give valuable insights in analyzing tourism trends and determining the causes of tourists' behavior while on traveling. Online review has been identified as new channels for visitors seeking traveling information. This is because it provides specific

and up-to-date information on tourist attractions that cannot be fully identified by ratings alone. Therefore, a timely study should be conducted that uses online review data that contains the visitor's real information and experience to understand the overall satisfaction of the tourist attraction.

The first step in explaining tourist behavior is to analyze the factors that affect tourist's destination choices. This study obtained Tripadvisor's actual online review data. The major destination in this study are five well-known natural tourist locations that were chosen five times in a row from among 100 must-visit tourist destination in South Korea. These can all be said to be tourist destinations that represent Korea. This includes ①Seoraksan National Park, ②Hallasan National Park, ③Suncheon Bay Nature Reserve(Suncheonman National Garden), ④Namiseom Island, and ⑤Udo Island.

The following are a summary of the methods used in this study:

First, text mining technique was used to collect review data and it was handled by Python. To improve the efficacy of frequency analysis, only nouns were retrieved during this procedure. The term 'frequently used words' include to which visitors commonly refer to when writing review and evaluating tourist attraction. Using this method, it was feasible to identify terms that visitors typically use while evaluating tourist spot with overall satisfaction score.

Second, the selection attributes which are called topics for tourist destination found in online review were derived applying LDA topic modeling with TF-IDF weights. Identifying the primary topics in review data is crucial, since visitors' impressions and their interests in tourist spots may be understood in terms of topics rather than word-based frequency analysis.

Third, FsQCA has validated a set of tourist destination selection criteria that

assess general tourists' satisfaction. The relationship between tourist destination selection attributes and satisfaction might be explained at a more elaborate level by employing FsQCA. For example, if there are ten reviews with a satisfaction score of five, it would be difficult to assume that all ten evaluations discuss the same topics. The reason for this is the topics that make up each review's content, as well as the combination of those topics, may vary for each review. As a result, it should be assumed that even in a circumstance that explains the same condition (e.g., when the satisfaction score is five in all cases), a combination of various causal conditions may arise.

FsQCA was used as an effective analytical tool in this situation. The investigation between tourist destination selection variables and their satisfaction was examined in a more advanced dimension using FsQCA. First, the establishment of necessary conditions was validated by designating tourist destination selection attributes (topics) and tourist satisfaction as causal conditions and dependent variables respectively. Following that, a truth table algorithm was created to reflect the combination of various causal conditions that determine satisfaction with tourist destination. As a result, some representative model of the causal condition combination for assessing tourist destination satisfaction were given. In this case, standard analysis was employed to verify sufficient conditions, and the analysis results were explained based on the intermediate solution.

Therefore, a few models with high coverage and high consistency were offered as a result that support this research subject. The sort of causal condition combination (tourist destination selection attributes) proven to have a high proportion in the review evaluating tourist destination satisfaction was

discovered. These findings are meaningful academically and practically, because it allows to know which combination of tourist destination selection attributes influence visitor's overall satisfaction during tourist experience.

In conclusion, the online review data is qualitative whereas the satisfaction score data is quantitative. This research attempted to use two different types of data to investigate various combinations of tourist destination selection variables that influence visitor satisfaction. The most important aspect of this study is not only investigated the tourist attraction selection attributes mentioned alone, but also explained the actual causal relationship from a combined perspective using FsQCA. According to research, improvements were presented based on the research findings, and practical ramifications were examined. This study made an academic contribution by demonstrating how LDA topic modeling may be expanded using FsQCA approach.

**Keywords : Tourist destination selection attributes, Tourist destination satisfaction, Online review, Text mining, LDA(Latent Dirichlet Allocation) Topic modeling, FsQCA(Fuzzy-set Qualitative Comparative Analysis)**