

석사학위 논문

공공데이터를 활용한 코로나 시대  
여행행태 연구:제주특별자치도를  
중심으로

A Study on Travel Behavior in  
the Corona Era Using open data  
: Focusing on Jeju Special  
Self-Governing Province

2021년 6월

송실대학교 정보과학대학원

소프트웨어공학과

이 규 남



석사 학위 논문

공공데이터를 활용한 코로나 시대  
여행행태 연구:제주특별자치도를  
중심으로

A Study on Travel Behavior in  
the Corona Era Using open data  
: Focusing on Jeju Special  
Self-Governing Province

2021년 6월

숭실대학교 정보과학대학원

소프트웨어공학과

이 규 남

석사학위 논문

공공데이터를 활용한 코로나 시대  
여행행태 연구:제주특별자치도를  
중심으로

지도교수 신 언 석

이 논문을 석사학위 논문으로 제출함

2021년 6월

송실대학교 정보과학대학원

소프트웨어공학과

이 규 남

이 규 남 의 석 사 학 위 논 문 을 인 준 함

심 사 위 원 장 박 재 표 인

---

심 사 위 원 신 언 석 인

---

심 사 위 원 장 주 팔 인

---

2021년 6월

숭실대학교 정보과학대학원

## 목 차

국문초록 .....	vi
영문초록 .....	viii
제 1 장 서론 .....	1
1.1 연구 배경 및 목적 .....	1
1.2 연구 범위 및 방법 .....	2
제 2 장 선행연구 및 이론적 배경 .....	4
2.1 선행연구 .....	4
2.2 이론적 배경 .....	6
2.2.1 공공데이터 정의 및 활용 .....	6
2.2.2 여행행태 연구 .....	9
2.2.3 단변량 선형회귀분석 .....	12
제 3 장 여행행태 분석 방법 .....	14
3.1 분석 절차 .....	14
3.1.1 데이터 수집 .....	14
3.1.2 데이터 전처리 .....	17
3.1.3 데이터 분석 .....	22
3.2 분석 대상 데이터 .....	24
제 4 장 여행행태 분석 결과 .....	26

4.1 여행행태 비교분석 및 시각화 .....	26
4.1.1 이동통신 방문자수 .....	26
4.1.2 신용카드 지출금액 .....	27
4.1.3 내비게이션 검색건수 .....	36
4.2 코로나 확진자수 단변량 선형회귀분석 .....	44
 제 5 장 결론 .....	46
 참고문헌 .....	48

## 그 립 목 차

[그림 2-1] 연구의 방법 및 주요 내용 .....	6
[그림 3-1] 여행행태 분석 절차 .....	14
[그림 3-2] 이동통신 방문자수 월 데이터 형태 .....	18
[그림 3-3] 신용카드 지출금액 월 데이터 전처리 결과 .....	19
[그림 3-4] 내비게이션 검색건수 월 데이터 전처리 결과 .....	20
[그림 3-5] 단변량 선형회귀분석용 일 데이터 전처리 결과 .....	22
[그림 4-1] 이동통신 방문자수 비교분석 결과 .....	26
[그림 4-2] 이동통신 방문자수 시각화 결과 .....	27
[그림 4-3] 신용카드 총소비 지출금액 비교분석 결과 .....	28
[그림 4-4] 신용카드 총소비 지출금액 시각화 결과 .....	28
[그림 4-5] 신용카드 숙박업 지출금액 비교분석 결과 .....	29
[그림 4-6] 신용카드 여행업 지출금액 비교분석 결과 .....	29
[그림 4-7] 신용카드 대중교통 지출금액 비교분석 결과 .....	30
[그림 4-8] 신용카드 렌터카 지출금액 비교분석 결과 .....	30
[그림 4-9] 신용카드 항공사 지출금액 비교분석 결과 .....	31
[그림 4-10] 신용카드 레저스포츠 지출금액 비교분석 결과 .....	31
[그림 4-11] 신용카드 면세점 지출금액 비교분석 결과 .....	32
[그림 4-12] 신용카드 문화서비스 지출금액 비교분석 결과 .....	32
[그림 4-13] 신용카드 쇼핑(면세점 제외) 지출금액 비교분석 결과 .....	33
[그림 4-14] 신용카드 식음료 지출금액 비교분석 결과 .....	33
[그림 4-15] 내비게이션 전체 검색건수 비교분석 결과 .....	36
[그림 4-16] 내비게이션 전체 검색건수 시각화 결과 .....	37



[그림 4-17] 내비게이션 자연관광지 검색건수 비교분석 결과 .....	37
[그림 4-18] 내비게이션 역사관광지 검색건수 비교분석 결과 .....	38
[그림 4-19] 내비게이션 휴양관광지 검색건수 비교분석 결과 .....	38
[그림 4-20] 내비게이션 공연/행사 검색건수 비교분석 결과 .....	39
[그림 4-21] 내비게이션 문화시설 검색건수 비교분석 결과 .....	39
[그림 4-22] 내비게이션 레포츠 검색건수 비교분석 결과 .....	40
[그림 4-23] 내비게이션 쇼핑 검색건수 비교분석 결과 .....	40
[그림 4-24] 내비게이션 식음료 검색건수 비교분석 결과 .....	41
[그림 4-25] 내비게이션 숙박 검색건수 비교분석 결과 .....	41
[그림 4-25] 내비게이션 기타관광지 검색건수 비교분석 결과 .....	42

## 표 목 차

[표 1-1] 관광산업 BSI 현황 .....	2
[표 2-1] 개별 법령에서의 공공데이터 유사용어와 정의 .....	6
[표 2-2] 공공데이터 활용 우수 사례 .....	8
[표 2-3] 외래 관광객 여행 현황 .....	10
[표 3-1] 수집 데이터 현황 .....	15
[표 3-2] 코로나 확진자수 분석 활용 선정 기준 .....	21
[표 3-3] 신용카드 업종 유형 .....	25
[표 3-4] 내비게이션 목적지 유형 .....	25
[표 4-1] 신용카드 지출금액 여행행태 변수별 구성비 집계 - 1 .....	35
[표 4-2] 신용카드 지출금액 여행행태 변수별 구성비 집계 - 2 .....	35
[표 4-3] 내비게이션 검색건수 여행행태 변수별 구성비 집계 - 1 .....	43
[표 4-4] 내비게이션 검색건수 여행행태 변수별 구성비 집계 - 2 .....	43
[표 4-5] 코로나 확진자수 기준 여행행태 변수별 단변량 선형회귀분석 결과 .....	45

국문초록

## 공공데이터를 활용한 코로나 시대 여행행태 연구:제주특별자치도를 중심으로

이규남

소프트웨어공학과

승실대학교 정보과학대학원

현대사회에서 여행행태는 시대가 흐르면서 계속해서 변하고 있다. 2017년에는 ‘인생은 한번 뿐이기에 현재를 즐겨야 한다’를 의미하는 율로(YOLO) 열풍으로 즉흥 여행이나 혼자 여행이 증가하였고, 2018년에는 호캉스(호텔에서 바캉스를 보내다의 신조어)라는 신조어가 생길 만큼 쾌적한 공간에서 자신만의 시간을 즐기는 20~30대가 늘어났다. 시대의 흐름에 맞게 여행 트렌드가 급변하면서 관광지에 따른 여행행태를 분석 및 예측하여 관광객들의 욕구와 수요를 충족시킬 수 있는 맞춤형 서비스가 중요한 상황에서 2020년도에 발생한 코로나바이러스 감염증(이하 코로나)은 관광산업의 위기와 동시에 또 다른 여행행태의 변화를 주고 있다.

이에 본 연구에서는 국내 제주특별자치도를 중심으로 코로나 시대의 여행행태를 분석하여 코로나 시대 관광산업 맞춤형 서비스 제공 모형개발의 기초 자료로 활용하고자 한다. 본 연구는 인터넷에서 제공하는 공공데이터를 활용하였으며 분석기간은 코로나가 국내에 처음 유입된 2020년 1월을 기점으로 전년 대비 비교를 위해 2019년 1월부터 2020년 12월 까지로 한정한다. 분석 방법은 한국관광데이터랩에서 제공하는 이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수 데이터를 활용하여

전년 대비 증감률과 구성비를 포함한 여행행태의 비교분석과 시각화를 진행하였고, 코로나 확진자수 데이터를 기준으로 단변량 선형회귀분석을 실시했다.

분석 결과에 의하면 코로나 시기에 제주특별자치도를 방문한 관광객의 여행행태는 인파가 물리지 않는 안전하고 쾌적한 장소를 찾아 이동하고 레저스포츠와 같이 폐쇄된 실내 관광지보다 자연 중심의 스포츠를 선호하는 것으로 확인됐다. 반면에 코로나로 인해 이용 제한이 있는 문화시설이나 인파가 물리는 관광지에 대한 거부감은 증가한 것으로 나타났다.

본 연구의 분석 결과는 향후 포스트 코로나 시대에 대응하여 관광산업 맞춤형 서비스를 제공하는 기초 자료로 활용될 수 있고, 유의미한 여행행태 분석 데이터를 제공하여 효율적인 관광산업 대응 방안을 제시했다는 점에서 의미가 있다.

## ABSTRACT

# A Study on Travel Behavior in the Corona Era Using open data : Focusing on Jeju Special Self-Governing Province

LEE, KYU-NAM

Department of Software Engineering  
Graduate School of Information Sciences  
Soongsil University

Travel behavior in modern society is constantly changing as the times pass. In 2017, due to the YOLO craze, which means ‘you have to enjoy the present because you only have one life’ improvisational travel and solo travel increased, and in 2018, a new word called hocance (a new word for spending a vacation at a hotel) emerged. As such, more people in their 20s and 30s enjoy their own time in a comfortable space. As travel trends change rapidly in line with the times, it is important to analyze and predict travel behavior according to tourist destinations to meet the needs and demands of tourists. The coronavirus infection that occurred in 2020 is giving another change in travel behavior at the same time as the crisis in the tourism industry. Therefore, this study aims to analyze travel behavior in the Corona era, centering on Jeju Special Self-Governing

Province in Korea, and use it as basic data for the development of a customized service provision model for the Corona era tourism industry. This study utilized public data provided by the Internet, and the analysis period is limited from January 2019 to December 2020 for comparison compared to the previous year from January 2020, when Corona first entered Korea. As for the analysis method, comparative analysis and visualization of travel behavior including year-on-year increase and decrease rate and composition ratio were conducted using data on the number of mobile communication visitors, credit card expenditure, and number of navigation searches provided by the Korea Tourism Data Lab. As a standard, univariate linear regression analysis was performed.

According to the analysis results, it was confirmed that the travel behavior of tourists who visited Jeju Special Self-Governing Province during the coronavirus period was to move in search of a safe and comfortable place where no crowds were gathered, and to prefer nature-centered sports to closed indoor tourist attractions such as leisure sports. On the other hand, it was found that resistance to cultural facilities with limited use or tourist spots crowded due to the corona has increased.

The analysis results of this study are meaningful in that they can be used as basic data to provide customized services for the tourism industry in response to the post-corona era in the future, and provide meaningful travel behavior analysis data to suggest efficient measures to respond to the tourism industry.

# 제 1 장 서론

## 1.1 연구 배경 및 목적

2019년 12월 중국 우한에서 최초 발생한 코로나바이러스 감염증(이하 코로나)은 세계적 대유행(Pandemic)을 불러일으켰다. 우리나라에도 2020년 1월 20일 최초 중국인 입국자의 확진 판정을 시작으로 확산하였다. 2021년 2월 25일 00시 기준으로 우리나라의 확진자수는 88,516명, 사망자수는 1,581명을 기록하며 전국적으로 확산되고 있다[1].

코로나의 빠른 확산은 현대사회의 모든 분야에 정신적, 경제적 고통과 함께 많은 변화를 주고 있으며 관광산업도 예외는 아니다. 한국문화관광연구원에서 발표한 보고서를 살펴보면[2] [표 1-1]과 같이 관광산업의 2020년 BSI<sup>1)</sup>는 2019년에 대비하여 모든 업종에서 하락하였다. 특히, 코로나가 국내에 유입된 시기인 2020년 1분기에는 관광산업 전체를 포함하여 대부분의 업종별 BSI가 전분기 대비 크게 하락한 것으로 나타났다. 관광산업의 전반적 업황은 2020년 1분기 31.3을 시작으로 2020년 4분기에는 56.2로 상승하였으나, 모든 업종에서 BSI가 전분기 대비 하락하여 코로나 시기의 관광산업 경기에 대한 부정적 인식을 보여주고 있다.

이에 따라 코로나 시대의 관광산업을 중심으로 관광지 방문객의 여행 행태를 분석하여 코로나 시대에 맞는 관광산업 맞춤형 서비스를 제공하기 위한 기초 활용 데이터가 필요하다.

---

1) BSI(Business Survey Index) : 기업의 매출, 생산 등 주요경영활동의 결과와 전망에 대해 설문조사하여 그 결과를 지수화함으로써 산출한 통계로 산식은  $BSI = (\text{긍정응답업체수} - \text{부정응답업체수}) / \text{전체응답업체수} \times 100 + 100$  이다.

[표 1-1] 관광산업 BSI 현황

구분	2019				2020			
	1분기	2분기	3분기	4분기	1분기	2분기	3분기	4분기
전체	74.9	87.3	85.1	88.6	31.3	53.8	69.2	56.2
관광숙박업	75.8	90.6	90.3	83.0	27.1	46.6	66.8	42.3
관광식당업	68.0	79.4	77.8	98.1	28.9	55.1	50.4	37.7
여행사 및 관광운수업	79.8	92.1	78.8	84.5	17.7	32.9	48.9	48.9
문화오락 및 레저산업	100.7	110.5	110.4	98.2	64.3	85.5	83.9	73.7
관광쇼핑업	93.5	87.3	79.1	114.8	29.9	51.4	58.1	37.7
국제회의 및 전시업	99.2	104.0	97.9	104.5	39.8	52.7	73.1	85.6

※출처: 한국문화관광연구원[2] 자료를 연구자가 재편집함

본 연구는 코로나 시대의 국내 관광지 여행행태를 분석하여 코로나가 여행행태에 주는 영향도를 파악하고, 향후 관광산업의 유의미한 분석 데이터를 제공하는 것에 목적을 둔다.

## 1.2 연구 범위 및 방법

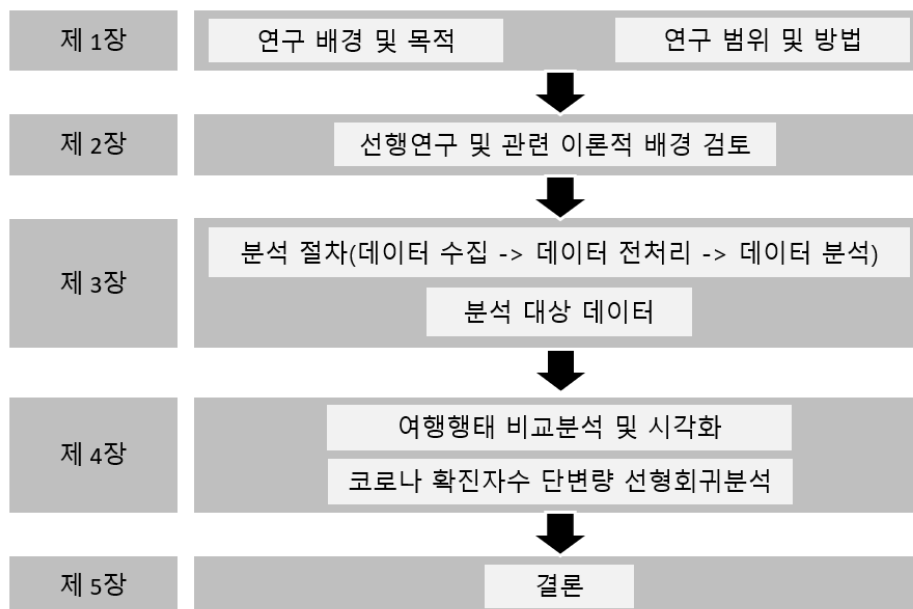
본 연구의 시간적 범위는 코로나 시대의 전년 대비 비교를 위하여 2019년 1월부터 2020년 12월까지 설정하였으며, 공간적 범위는 제주특별자치도로 한정하였다. 여행행태 분석에 활용되는 데이터는 인터넷에서 제공하는 공공데이터를 활용하였다.

연구의 방법은 코로나 시대의 관광지 여행행태 비교분석과 코로나 확



진자수를 기준으로 회귀분석을 실시하기 위해 본 연구에 맞게 수정 및 보완하여 통계 분석 도구인 R 프로그램을 사용하였다.

본 연구의 구성과 내용은 [그림 1-1]과 같이 모두 5장으로 구성되어 있다. 1장은 서론 부분으로 연구의 배경과 목적, 연구의 범위 및 방법에 관해 서술하고, 2장에는 선행연구와 본 연구에 활용되는 정보의 이론적 배경을 살펴보았다. 3장에서는 분석 절차와 분석에 사용되는 데이터에 관한 내용을 서술하고, 4장에서는 가공된 데이터를 기준으로 코로나 시대의 여행행태 비교분석과 시각화를 실시하고 코로나 확진자수가 여행행태 변수에 미치는 영향을 파악하기 위해 진행한 단변량 선형회귀분석 결과를 서술하였다. 5장에서는 분석 결과를 기준으로 연구의 결론을 서술하였다.



[그림 1-1] 연구의 방법 및 주요 내용

## 제 2 장 선행연구 및 이론적 배경

### 2.1 선행연구

코로나 시대의 여행행태와 관련된 선행연구는 관광산업에 큰 영향을 주는 만큼 다양하게 이루어져 왔다. 박근화(2020)는 특정 조사업체를 대상으로 코로나로 인한 국민의 국내 여행에 대한 행태와 향후 여행에 대한 인식을 조사하였다[3].

특정 국가를 선정하여 관광산업을 연구한 경우에는 안동, 유효강(2020)이 코로나의 발생이 중국 관광산업에 미치는 영향과 대응책을 연구 하였다. 위기관리 체계에서 관광산업은 피해의 감소와 신속한 회복을 도모하기 위해 정부의 지원정책(금융지원, 세금감면, 재정지원, 취업유지 등), 협회의 대응책, 기업의 대응책(산업융합, 내부 위기관리 등)을 제시하였다[4].

국내에 특정 지역을 선정하여 여행행태를 분석한 연구로는 안필균, 엄성준, 조숙영, 김상범(2020)이 포스트 코로나 시대에 관광 트렌드를 반영한 전라북도 완주군 소양면 위봉마을의 농촌체험마을에 대한 조성방안 연구를 하였다. 농촌체험마을의 프로그램 구성에 있어 소규모 체험 및 안전이 보장된 형태의 관광 선호도가 증가하는 트렌드에 대응하고, 지역자원을 활용한 체험코스를 계획하여 안전한 농촌관광이 지속될 수 있는 방법을 제안하기 위해 수행하였으며, 연구 결과를 토대로 관광, 휴식, 경제, 다원적 기능 측면에서 증가하고 있는 농촌에 대한 수요를 충족시킬 것이라 하였다[5].

코로나 시대에 관광객의 관광욕구가 어떠한 감정으로 표출되는지를 감정 분석을 통해 연구한 사례도 있다. 홍민정, 오문향(2020)은 코로나 확산

에 대한 국내 잠재 관광객의 감정 반응 연구를 하였다. 트위터를 활용한 의미론적 네트워크 분석 결과, 국내 잠재 관광객들은 코로나 확산에 대한 주요 이슈를 중심으로 여행에 대한 불안감과 억눌린 감정, 여행 후 감염 뉴스에 대한 부정적 놀라움, 각국의 입국금지 조치에 대한 화나는 감정들을 텍스트를 통해 소셜 미디어에 공유한다는 것을 밝혔다.[6] 또한, 김성태(2020)도 코로나 확산에 따른 관광위험지각과 감정반응이 행동 의도에 미치는 영향을 연구하였으며, 연구의 결과는 관광위험지각의 이론적 확장에 기여할 것으로 예상했다[7].

본 연구와 같은 지역인 제주특별자치도를 선정하여 코로나 시대의 관광산업을 연구한 김영남, 홍성화(2020)는 제주 방문 관광객을 중심으로 코로나 위험 인식과 낙관적 편향에 관한 탐색적 연구를 진행하였다. 감염 위험에도 불구하고 관광목적지로 이동하여 관광 행동을 하는 관광객의 관광 행동에 관한 심리적 요인을 낙관적 편향으로 보고, 선행연구를 바탕으로 실증 연구를 실시하였다. 연구 결과, 코로나 기간 제주를 방문한 관광객들은 코로나 위험에 대해 낙관적 편향을 보이는 것으로 나타났으며, 이러한 현상은 남성보다는 여성에서, 그리고 60대 이상의 연령보다 20대 이하의 연령에서 유의한 차이로 더 높게 나타났다. 이 연구의 결과를 통해 전염병, 자연재해 등의 관광 위협 요소에 대한 관광객의 인식과 행동 의도를 낙관적 편향을 바탕으로 설명이 가능하다고 밝혔다[8].

이번 연구는 선행연구들과 비교하여 공공데이터를 활용하여 여행행태를 분석한다는 점에서 크게 차이가 있다. 코로나가 발생한 시점의 2020년 데이터를 활용하여 전년 대비 비교분석을 하고, 신용카드 지출금액, 내비게이션 검색건수 등 여행행태의 새로운 변수를 선정하여 코로나 확진자 수 기준으로 회귀분석을 실시하여 코로나 시대의 효율적인 관광산업 대응 방안을 제시하는 기초 자료로 활용하고자 한다.

## 2.2 이론적 배경

### 2.2.1 공공데이터 정의 및 활용

공공데이터는 공공기관이 보유한 관광, 여행, 교육, 교통, 인구, 안전 등 여러 분야의 풍부한 데이터를 국민에게 공유하여 누구나 차별 없이 자유롭게 사용할 수 있는 데이터를 의미한다. 특히, 공공데이터는 여러 데이터와의 융합 분석을 통해 다각도의 가치 있는 데이터를 추출하고 향후 데이터 기반의 맞춤형 서비스를 제공할 수 있다는 점에서 의미가 있다.

공공데이터는 민간이 생산하는 데이터보다 광범위 하며, 지속적이고 안정적으로 신뢰성 있는 데이터를 생산하고 있다. 그러나 현행 공공데이터의 제공 및 이용 활성화에 관한 법률이 제정되기 전까지 직접적으로 공공데이터로 지칭한 용어와 정의는 없었으며 (디지털)정보, 행정정보, 지식자원정보, 공공정보, 기록물 등 유사한 용어를 사용하고 있었다[9].

현재 공공데이터는 [표 2-1]에 정의된 공공정보 의미와 다르게 제공 및 이용 활성화에 관한 법률 제 2조에 의해 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하는 전자적 방식의 처리된 자료 또는 정보라고 정의한다.

[표 2-1] 개별 법령에서의 공공데이터 유사용어와 정의

관련법령	용어	법적 정의
국가정보화기본법 제3조(정의)	정보	특정 목적을 위하여 광(光) 또는 전자적 방식으로 처리되어 부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식
	지식정보자원	국가적으로 보존 및 이용 가치가 있는 자료로서 학술, 문화, 과학기술, 행정 등에 관한 디지털화된 자료나 디지털화의 필요성이 인정되는 자료

관련법령	용어	법적 정의
공공기관의 정보공개에 관한 법률 제2조 (정의)	정보	공공기관이 직무상 작성 또는 취득하여 관리하고 있는 문서(전자문서를 포함한다. 이하 같다)·도면·사진·필름·테이프·슬라이드 및 그 밖에 이에 준하는 매체 등에 기록된 사항
공공기록물 관리에 관한 법률 제3조 (정의)	기록물	공공기관이 업무와 관련하여 생산하거나 접수한 문서·도서·대장·카드·도면·시청각물·전자문서 등 모든 형태의 기록정보 자료와 행정박물(行政博物)
전자정부법 제2조 (정의)	행정정보	행정기관 등이 직무상 작성하거나 취득하여 관리하고 있는 자료로서 전자적 방식으로 처리되어 부호, 문자, 음성, 음향, 영상 등으로 표현된 것
이러닝(전자학습)산업 발전 및 이러닝 활용 촉진에 관한 법률 제2조(정의)	공공정보	공공기관이 직무상 작성하거나 취득하여 관리하고 있는 문서·도면·사진·필름·테이프·슬라이드 및 컴퓨터에 의하여 처리되는 매체 등에 기록된 사항
콘텐츠산업진흥법 제11조(공공정보의 이용 활성화)	공공정보	그 공공기관이 보유·관리하는 정보 중 공공기관의 정보공개에 관한 법률 제9조에 따른 비공개대상 정보를 제외한 정보

※출처: 법제처 국가법령정보센터[10]

공공데이터를 개방한 시점 이후로 공공데이터를 활용하면서 국민의 편의를 향상시키고 신규 일자리나 비즈니스 창출에도 기여하는 많은 사례들이 있다. [표 2-2]는 중앙부처, 지방자치단체 및 공공기관에서 제공하는 데이터를 활용하여 기술 및 제작한 우수 사례들이다. 공공데이터를 통한 국가와 국민의 동반성장 및 사회 가치 창출을 위해 지속적으로 공공데이터가 민간에 개방되어야 할 것이다.

[표 2-2] 공공데이터 활용 우수 사례

분야	기관/소속	실무자용 제목	내용 요약
국민안전	전라북도	골든타임 확보를 위한 구급차 배치 운영최적화 모델	구급차 구급 요원 취약 지역 시간대 분석을 통한 배치 최적화
	경기도	CCTV 설치지역 분석 및 모델 고도화	범죄예측지수 모델링 감시취약지수 분석을 통한 우선 설치지역 도출
	경찰청 국가정보 자원관리원	인공지능을 활용한 과학적 민생치안	임장일지 기반 유사사건 여죄분석
국민생활 개선	국민건강 보험공단	국민건강 알람서비스	기상 환경 변수 등 활용 실시간 국민관심질병 예측 서비스
	광주광역시	빅데이터 분석을 통한 시내버스 효율적 운영	대중교통 사각지대 분석을 통한 취약지역 분석 개선
	대구광역시	전기차 충전 인프라 설치 입지선정	충전인프라 지속적인 설치와 편리한 접근성을 위한 입지 지역 도출
경제활성 화	남양주시 국민연금공 단	잠재적 사회취약계층 일자리 창출 자립지원	사회취약계층을 선제적으로 지원하기 위해 우선지원 취약 계층 선정
	전주시	전주 한옥마을 관광분석 을 통한 경제활성화	방문객의 동선 및 상권분석을 통한 마케팅 활용 관광정책 수립
	기상청 농촌진흥청	기상데이터와 농산물 생산성 예측	기상 병충해 발생자료 등의 분석을 통한 생산량 예측모델 개발
스마 트행 정	고용노동부	근로감독 사업장 선정 과학화	근로기준법 위반 사업장 적발을 위한 데이터 분석
	경기도 국토교통부	공동주택관리비 빅데이터 분석	관리비 비리패턴 추출 및 민원분석을 통한 관리비 부당사용 방지
	한국도로공 사 ICT센터	인공지능기반 도로포장 파손 실시간 탐지시스템	교통하중 기후환경 등의 분석을 통한 효율적 도로유지 관리체계 구축
	포항시	민원분석을 통한 갈등 문제 해결	콜센터 민원 데이터 분석을 통한 통합적 민원관리 시스템 구축

※출처: 행정안전부[11] 자료를 연구자가 재편집함

## 2.2.2 여행행태 연구

정보 의존도나 환경변화에 민감한 여행행태는 시대가 흐르면서 계속해서 변하고 있다. 2017년에는 1인 가구의 비중이 증가하면서 혼자 여행이 여행 트렌드가 되었고, 2018년에는 쾌적한 공간에서 자신만의 시간을 즐기는 20~30세대가 늘어났다. 여행의 인식도 특별한 날에만 떠나는 행사가 아닌 일상에서도 즐길 수 있는 취미로 자리 잡기 시작하면서 즉흥적으로 떠나는 여행이 트렌드가 되었고, 2019년에는 인스타그램, 페이스북 등 SNS가 활성화되면서 SNS 기반의 여행을 추구하는 세대도 증가하였다. 2020년에는 코로나 팬데믹(세계적 대유행)으로 인해 관광산업에 위기를 맞으며 이에 따른 여행수요 급감과 함께 안전을 중요하게 여기는 여행 트렌드로 변하였다. 여행행태를 분석하고 예측하여 관광객들의 욕구와 수요를 충족시킬 수 있는 맞춤형 서비스가 중요해진 만큼 주요 여행행태에 대해 살펴보려고 한다.

### 1) 혼자 여행

혼자 여행을 떠나는 여행 트렌드가 사회 전반적으로 확산되면서 나 홀로 여행족이 급증하고 있지만, 아직은 관광학 및 여가측면에서 혼자 여행자에 대한 정확한 정의나 연구는 부족한 실정이다. 혼자 여행자에 대한 정의가 정확히 확립되지 않았지만, 일반적으로 ‘가족, 친구 등 동행인 없이 혼자서 하는 여행을 즐기는 사람’을 의미한다[15].

혼자 여행이 증가한 요인 중에서 1인 가구의 증가는 많은 부분을 차지한다. 혼밥(혼자 먹는 밥), 혼술(혼자 마시는 술), 혼행(혼자 하는 여행), 혼영(혼자 보는 영화), 1인과 경제를 뜻하는 1코노미(1conomy) 등과 같은 신조어의 등장은 국내 1인 가구의 증가에 따른 라이프 스타일 및 경

제구조의 변화를 단적으로 보여주고 있다[12].

통계청 자료에 의하면 우리나라의 2019년 기준 전체 가구 10가구 중 3가구(30.2%)가 1인 가구라고 밝혔다. 세대별로는 20대가 전체 1인가구의 18.2%로 가장 큰 비중을 차지했으며, 다음은 30대(16.8%), 50대(16.3%), 60대(15.2%) 순이었다[13]. 1인 가구 내에서도 20~30대가 차지하는 비중이 급증하면서 혼자 여행의 여행 트렌드를 주도하는 세대가 되었다.

1인 가구의 증가는 우리나라뿐만 아니라 전 세계적으로 해당하는 현상이다. 한국관광 데이터랩에서 제공하는 자료 [표 2-3]에 의하면 우리나라를 방문하는 외래 관광객의 여행 형태는 개별여행이 2015~2019년까지 가장 높은 비율을 차지하면서 혼자 여행을 떠나는 관광객도 증가하고 있다는 것을 알 수 있다.

[표 2-3] 외래 관광객 여행 형태

구분	2015년	2016년	2017년	2018년	2019년
개별여행	67.92%	67.35%	82.81%	79.93%	77.08%
단체여행	26.2%	25.01%	11.6%	12.36%	15.12%
Air-tel	5.88%	7.64%	5.58%	7.7%	7.8%

※출처: 한국관광 데이터랩[14] 자료를 연구자가 재편집함

## 2) 가족 여행

가족 여행은 가족 간의 행복을 얻는 효과적인 방법으로 관광시장에서 수요가 계속 커지고 있다. 가족 여행과 관련된 국내 학술논문을 살펴보면, 가족여행객을 대상으로 가족 간의 애정 교류행태와 체험 가치의 관계를 분석한 논문이 있다. 연구 결과로는 가족애에 대한 상호작용은 가족 관광객의 기능적, 인지적, 감정적 체험가치에 대해서 모두 긍정적인 것으로 나타났다. 또한, 사회적 지지도 기능적, 인지적, 감정적 등의 전반



적인 체험가치에 대해서 모두 긍정적인 것으로 나타났다. 본 연구는 여행의 본질적인 측면에서 가족 간의 상호작용 행동과 체험가치를 이해할 수 있는 근거를 제공하였다[17].

2019년 국내 가족 여행의 분석 자료에 의하면, 가족 여행의 1회 평균 일수는 1.8일로 여행 전체 1회 평균 여행 일수(1.72일)보다 높게 나타났다. 또한, 자녀와 함께 사는 가구의 가족 여행이 자녀가 없는 가구의 가족 여행보다 1회 평균 여행 일수가 높게 나타났다고 밝혔다[18]. 자녀를 둔 부모의 관점에서 아이들과 공동 경험을 쌓고 평소 부족했던 대화의 시간을 통해 가족 간의 화합을 중요하게 생각한 결과이다.

### 3) SNS 기반 여행

IT기술과 인터넷을 활용한 정보검색이 증가하면서 페이스북, 인스타그램을 비롯한 SNS(소셜네트워크서비스)로 여행 정보를 얻을 기회가 많아지고 있다. 또한, 스마트폰의 보편화로 인해 언제 어디서나 정보를 손쉽게 획득할 수 있는 SNS의 특성은 온라인 관광시장의 변화를 주도하고 다양한 이벤트 및 여행상품 정보와 여행 체험후기 등의 콘텐츠를 제공하고 있다[19].

SNS 활용을 통해 어떤 요인들이 여행 선택에 영향을 미치는지에 대한 분석도 증가하고 있다. 김문정(2020)은 최근 1년 이내 여행을 다녀온 대학생들을 대상으로 설문조사를 하였다. 분석 결과, 관광객들이 관광지를 선택할 때 영향을 받는 SNS 관광정보의 특성 순위가 생생함, 상호작용, 디자인, 지각된 평판 순이라고 나타났다. 또한, 관광객들이 실제 경험한 것처럼 느끼고 정보가 매우 구체적일 때 SNS 관광정보의 생생함이 가장 크게 느껴지는 것으로 나타났다[20].

### 2.2.3 단변량 선형회귀분석

회귀분석은 하나 또는 그 이상의 변수들이 다른 변수에 미치는 영향을 파악하는 분석 방법으로, 영향을 주는 독립변수의 변화에 따라 영향을 받는 종속변수에 어떠한 변화가 있는지를 분석하는 통계기법이다.

단변량 선형회귀분석은 위에서 언급한 독립변수와 종속변수가 하나일 때 연속적인 두 변수 간의 인과관계를 파악 및 분석하여 값을 예측하기 위한 방법이다. 따라서, 단변량 선형회귀분석은 변수 하나로부터 다른 변수에 대한 예측이 가능하다는 점에서 변수사이의 관계 강도나 유무를 분석하는 상관분석과는 차이가 있다.

단변량 선형회귀분석은 독립변수와 종속변수 간의 모형을 수립하고 이를 선형회귀직선으로 나타내면  $Y = a + bX$  이다. 여기서  $X$ 는 독립변수를,  $Y$ 는 종속변수를 의미한다.  $b$ 는 기울기로  $X$ 와  $Y$ 간의 선형 상관 관계를 의미하는 상관계수이다. 상관계수는 두 변수 사이의 연관성을 나타내는 계수로서 수치는 -1과 1 사이의 값을 지닌다. 상관계수를 제공한 값이 결정계수이며 결정계수는 회귀식이 얼마나 데이터를 잘 설명하고 있는지를 나타낸 계수이다. 즉, 회귀식에 사용된 독립변수가 종속변수를 얼마나 설명하고 있는지를 의미한다.  $a$ 는  $y$ 의 절편으로 직선이  $y$ 축과 만나는 점을 의미한다. 회귀직선에서 최적의 선은 모형에 의해 예측된 종속변수의 값과 실제 측정된 종속변수 값의 차이인 잔차가 가장 작은 경우이다.

단변량 선형회귀분석을 사용하여 선형 모델을 적용하기 전에 독립변수와 종속변수가 얼마나 적합한 관계를 맺고 있는지 확인 절차가 필요하며, 이는 산점도와 상관관계 검정으로 확인할 수 있다. 산점도는 독립변수와 종속변수 사이의 관계를 보기 위하여 관측점을 만드는 통계 그래프

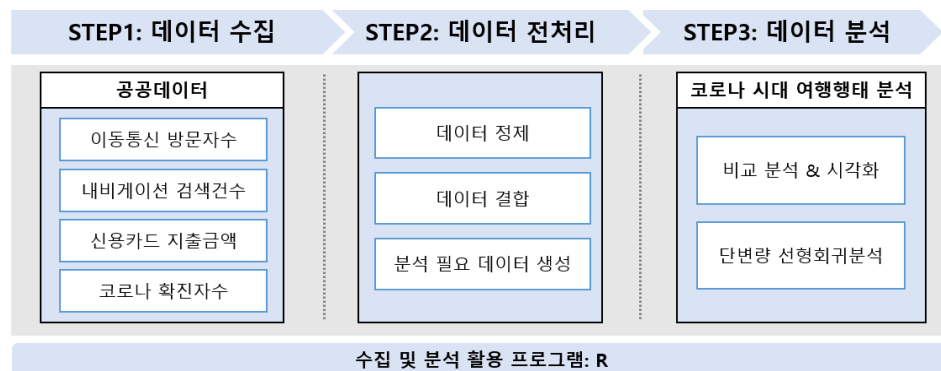
로 산점도에서 점의 위치를 통해 적합한 선형회귀 모델의 적용 여부를 결정하게 된다. 상관관계 검정은 독립변수와 종속변수 간에 관련성을 파악하는 통계 분석 방법으로 상관계수는 0과 1 사이의 값을 지닌다. 통계적으로 유의미한 결과인지 확인하기 위해서는 귀무가설은 상관계수가 0 이고, 대립가설은 상관계수가 0이 아니라고 놓고 유의확률(p-value)을 검증한다. 0에 가까울수록 상관관계가 낮은 것을 의미하며, 유의확률(p-value)의 값이 0.05보다 작으면 귀무가설이 기각되어 두 변수 간에 상관관계가 존재한다고 볼 수 있다.

본 연구에서는 여행행태 분석에 필요한 공공데이터를 수집하여 독립변수로는 코로나 확진자수를 선정하고, 종속변수에는 이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수 데이터를 선정하여 각각 단변량 선형회귀분석을 진행하였다. 두 개 이상의 독립변수를 사용하는 다중 회귀분석은 독립변수들 사이에 상관관계 나타나는 다중공선성 문제가 발생하여 진행하지 않았다. 또한, 본 연구에 사용되는 단변량 선형회귀분석은 사회과학 관점에서 코로나 확진자수에 따른 인간의 여행행태를 분석하고 변수간의 관계를 추정하고 설명을 위한 목적으로 사용하였다. 따라서, 독립변수와 종속변수 간의 관계를 설명하는 결정계수의 수치를 모형에 평가하는 절대적인 기준으로 두지 않았다.

## 제 3 장 여행행태 분석 방법

### 3.1 분석 절차

본 연구에서는 코로나 시대의 제주특별자치도의 여행행태를 분석하기 위해 [그림 3-1]과 같이 데이터 수집 및 전처리를 진행하고 마지막으로 데이터 분석을 통해 결과를 도출하는 절차와 방법으로 연구를 진행하였다. 분석을 위한 데이터 수집과 정제작업은 분석 프로그램 R을 사용하여 처리하였다.



[그림 3-1] 여행행태 분석 절차

#### 3.1.1 데이터 수집

여행행태 분석에 사용되는 공간적 범위를 제주특별자치도로 한정하였으며, 분석에 필요한 관련 데이터를 [표 3-1]와 같이 수집하였다. 이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수는 한국관광 데이터랩 홈페이지에서 데이터를 CSV 파일로 저장 및 수집하였고, 코로나

확진자수는 공공데이터포털에서 제공하는 API 연동과 위키백과에서 검색한 자료를 기준으로 데이터를 수집하였다. 데이터 수집에 활용된 프로그램은 R 프로그램을 사용하였다.

[표 3-1] 수집 데이터 현황

구분	여행행태 분석 변수명	단위	수집 주기	수집 방법	수집 기간	변수 활용 내용
이동통신	방문자수	명	일/월	CSV 파일 적재	월 데이터 2019.01.01 ~ 2020.12.31	비교분석 및 시각화 -> 월 데이터 활용
신용카드	총소비	금액 (원)				
	대중교통					
	레저스포츠					
	렌터카					
	면세점					
	문화서비스					
	쇼핑					
	숙박업					
	식음료					
	여행업					
	항공사					
내비게이션	전체	검색 건수 (건)			일 데이터 2020.01.20 ~ 2020.12.31	단변량 선형회귀 분석 종속변수 -> 일 데이터 활용
	공연행사					
	기타관광지					
	레포츠					
	문화시설					
	쇼핑					
	숙박					
	식음료					
	역사관광지					
	자연관광지					
	휴양관광지					
코로나	확진자수	명	일	API 호출 및 위키 백과	2020.01.20 ~ 2020.12.31	단변량 선형회귀 분석 독립변수

※ 데이터 출처

1) 이동통신, 신용카드, 내비게이션: <https://datalab.visitkorea.or.kr>

(한국관광데이터랩)

2) 코로나: <https://data.go.kr> (공공데이터포털), 위키백과

본 연구에서 데이터를 수집한 CSV 파일은 , (콤마) 구분자로 데이터를 구분해서 (.csv) 확장자로 저장하였다. 저장된 파일은 R 프로그램 내에 read.csv 함수를 사용하여 데이터 수집을 진행하였다.

코로나 확진자수의 경우, 공공데이터포털에서 제공하는 오픈 API와 위키백과에서 제공하는 자료를 활용하여 데이터를 수집하였다. 오픈 API만으로 수집이 어려운 점은 공공데이터포털에서 코로나 확진자수에 대한 자료를 2020년 3월 10일 이후부터 제공했기 때문이다. 따라서, 국내에 코로나가 처음 발생한 시점인 2020년 1월 20일부터 2020년 3월 9일까지의 코로나 확진자수 데이터는 위키백과[16] 자료를 활용하였다. 데이터 수집에 대한 처리는 R 프로그램에서 제공하는 함수를 활용하였으며, 상세 과정은 아래와 같다.

첫째, API를 호출할 수 있는 URL 변수를 공공데이터포털에서 제공하는 서비스 URL, 인증키, 요청변수(데이터의 개수 및 조건) 등을 사용하여 생성한다.

둘째, 생성된 URL 변수를 xmlTreeParse 함수의 인자로 적용하여 데이터를 XML 포맷 형태로 저장한다.

셋째, XML 데이터를 처리하기 위해 xmlRoot 함수를 사용하여 최상위 노드 정보를 획득한다.

넷째, 최상위 노드 정보를 기준으로 xmlToDataFrame 함수를 적용하여 실제 데이터가 저장되어 있는 item Node의 모든 데이터를 추출한다.

다섯째, 위키백과에서 API 호출만으로 데이터 수집이 불가능한 시기의 자료를 CSV 파일 형태로 작성하여 저장한다.

마지막으로, 위에서 추출한 API 연동 데이터와 CSV 파일에 저장된 데이터를 rbind함수로 적용하여 최종 분석에서 활용되는 데이터프레임 구조의 변수에 적재한다.

이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수 데이터는 코로나 시대의 비교분석과 시각화에 사용하기 위해 월 데이터를 수집하였고, 회귀분석에는 일 데이터를 수집하여 사용하였다. 월 데이터는 코로나 시대의 전년 대비 증감률과 구성비를 분석하기 위해 데이터 수집 기간을 2019년 1월부터 2020년 12월까지로 선정하였다. 일 데이터는 독립변수인 코로나 확진자수가 종속변수에 해당하는 여행행태 분석용 변수에 미치는 영향을 파악하기 위해 코로나가 국내에 처음 발생한 시점인 2020년 1월 20일부터 2020년 12월 31일까지 데이터를 수집하였다.

### 3.1.2 데이터 전처리

데이터 전처리는 수집된 데이터에서 이상치나 결측치에 대한 값을 확인하여 불필요한 데이터를 삭제하거나 분석에 필요한 데이터를 결합 및 수정하는 단계로서 데이터 품질을 올리는 중요한 과정이다. 본 연구에서 사용하는 데이터는 추후 분석 방법에 따라 여행행태 분석 변수별로 데이터 전처리가 진행된다. 분석방법은 코로나 시대 여행행태 비교분석과 코로나 확진자수에 영향받는 변수를 파악하기 위한 단변량 선형회귀분석이 있다.

#### 1) 여행행태 비교분석

한국관광 데이터랩에서 수집한 데이터는 여행행태 비교분석에 맞게 다듬는 작업이 필요하다. 한국관광 데이터랩에서 CSV 파일 형태로 수집한 이동통신 방문자수 데이터를 분석용 데이터 형태에 맞게 R 프로그램을 사용하여 [그림 3-2]와 같이 수정 및 정제작업을 진행하였다.

비교분석 및 시각화 작업에 맞게 월 항목 기준으로 2020년, 2019년 방

문자수 데이터를 생성하였으며, 증감률에 대한 순위 항목도 추가 생성하였다.

월	2019년 방문자수 (단위: 십만명)	2020년 방문자수 (단위: 십만명)	2020년 기준 전년대 비 증감률(단위: %)	증감률 순위
1	4266230	4718360	10.60	12
2	3919934	3129308	-20.17	4
3	4446690	2928020	-34.15	2
4	4080300	2553394	-37.42	1
5	4266714	3470128	-18.67	5
6	4080814	3630736	-11.03	7
7	4713832	4600592	-2.40	10
8	5349022	5108012	-4.51	8
9	4199822	3522309	-16.13	6
10	4742090	4560640	-3.83	9
11	4348533	4443582	2.19	11
12	4481918	3518081	-21.51	3

[그림 3-2] 이동통신 방문자수 월 데이터 전처리 결과

신용카드 지출금액은 관광업종별로 여행행태 분석에 활용되는 변수가 상세하게 구분 지어진다. 따라서, 신용카드 데이터는 여행행태 비교분석에 활용되는 자료를 수집하여 [그림 3-3]과 같이 수정 및 추가하였다.

수집된 데이터의 분석 활용 변수명 중에서 카지노는 코로나의 영향으로 인해 2020년 3월부터 지출금액이 0원이어서 분석 활용 변수에서 제외했다. 월별 분석 활용 변수명에 대한 지출금액, 전년 지출금액, 전년 대비 증감률, 구성비 항목을 생성하여 여행행태 비교분석과 시각화로 표현할 때 활용하였다.

구성비 항목은 월 기준으로 분석활용 변수명 항목이 전체 지출금액에서 차지하는 비율을 의미한다. 이 항목은 코로나 발생 시기에는 전체적으로 지출금액이 감소할 것이라는 일반적인 사고 전제하에 지출금액과 같이 추가로 비교할 수 있는 항목이다.



기준 연월	연도	월	시도명	분석활용변수명	지출 금액	전년지출 금액	전년대비 증감률	구성비
202003	2020	3	제주특별자치도	총소비	1.63e+10	4.27e+10	-61.89	100.000
202003	2020	3	제주특별자치도	숙박업	9.60e+08	3.45e+09	-72.17	5.896
202003	2020	3	제주특별자치도	여행업	2.03e+07	7.88e+07	-74.30	0.124
202003	2020	3	제주특별자치도	대중교통	1.50e+06	1.85e+06	-18.94	0.009
202003	2020	3	제주특별자치도	렌터카	2.88e+08	5.48e+08	-47.42	1.770
202003	2020	3	제주특별자치도	항공사	1.73e+08	2.99e+09	-94.21	1.063
202003	2020	3	제주특별자치도	레저스포츠	1.92e+09	2.06e+09	-6.87	11.794
202003	2020	3	제주특별자치도	면세점	2.94e+09	1.50e+10	-80.34	18.051
202003	2020	3	제주특별자치도	문화서비스	1.14e+08	5.45e+08	-79.11	0.699
202003	2020	3	제주특별자치도	쇼핑(면세점 제외)	8.69e+08	2.34e+09	-62.82	5.334
202003	2020	3	제주특별자치도	식음료	9.00e+09	1.57e+10	-42.68	55.260

[그림 3-3] 신용카드 지출금액 월 데이터 전처리 결과

내비게이션도 데이터도 신용카드와 동일하게 관광지 유형별로 분석에 활용되는 변수가 있으므로 [그림 3-4]과 같이 검색건수, 전년 검색건수, 전년 대비 증감률, 구성비 항목을 생성하였다.

기준 연월	연도	월	시도명	분석활용변수명	검색 건수	전년검색 건수	전년대비 증감률	구성비
202001	2020	1	제주특별자치도	전체	732249	562828	30.10	100.000
202001	2020	1	제주특별자치도	자연관광지	79846	71022	12.42	10.904
202001	2020	1	제주특별자치도	역사관광지	6515	6651	-2.04	0.890
202001	2020	1	제주특별자치도	휴양관광지	23595	18515	27.44	3.222
202001	2020	1	제주특별자치도	공연/행사	4330	3310	30.82	0.591
202001	2020	1	제주특별자치도	문화시설	32935	26932	22.29	4.498
202001	2020	1	제주특별자치도	레포츠(육상/해상/항공)	21142	14446	46.35	2.887
202001	2020	1	제주특별자치도	쇼핑	21440	19968	7.37	2.928
202001	2020	1	제주특별자치도	식음료	310491	227693	36.36	42.402
202001	2020	1	제주특별자치도	숙박	140565	102339	37.35	19.196
202001	2020	1	제주특별자치도	기타관광지	91390	71952	27.02	12.481

[그림 3-4] 내비게이션 검색건수 월 데이터 전처리 결과

## 2) 단변량 선형회귀분석

단변량 선형회귀분석은 코로나 확진자수에 의해 영향받는 변수를 파악하기 위해 사용한 분석 방법으로 비교분석에서 사용한 월 데이터와 다르게 일 데이터를 수집하여 진행하였다. 이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수의 일 데이터는 비교분석에서 수집한 월 데이터 형태와 동일하게 자료를 수집하여 데이터 전처리를 진행하였다.

코로나 확진자수 데이터는 오픈 API 연동을 통해 수집한 데이터 구조를 분석에 활용할 수 있도록 데이터 전처리를 진행하였다. 오픈 API로 수집한 데이터 기준 일자는 생성일시(createDt) 항목이 기준이 된다. 데이터가 하루에 한 번 이상 들어올 수 있기 때문에 생성일자 기준으로 가장 최근에 적재되어 있는 생성일시(createDt)에 해당하는 데이터를 추출하였다. 구분(gubun) 항목에는 국내 시도명과 전국에 해당하는 합계 데이터가 존재하지만 본 연구에서는 전국 코로나 확진자수 기준으로만 분석하기 때문에 합계에 해당하는 데이터만 활용하였다.

단변량 선형회귀분석에서 독립변수에 해당하는 코로나 확진자수는 [표 3-2]와 같이 정부의 거리두기 방침 내용을 참고하여 일일 코로나 확진자수가 50명 이상에 해당하는 일자만 선정하여 분석을 진행하였다.

[표 3-2] 코로나 확진자수 분석 활용 선정 기준

기간	정부 방침 내용	독립변수 선정 여부
2020.03.22 ~ 2020.05.05	사회적 거리두기 대책 발표 및 시행 집단감염 위험시설 운영제한 조치 통보	O
2020.05.06 ~ 2020.08.15	10명 안팎의 확진자수로 생활 속 거리두기라는 이름으로 생활 방역 시행	X
2020.08.16 ~ 2020.10.11	2020.08.16: 수도권에 한하여 2단계로 격상 2020.08.23: 전국으로 확대	O
2020.10.12 ~ 2020.11.23	거리두기 단계 1단계로 하향 *2020.11.01: 정밀 방역 형태로 세분화하여 5단계 체제 변경	X
2020.11.24 ~ 2020.12.31	- 2020.11.24: 연속 300명대 코로나 확진자수로 수도권 거리두기 단계 2단계 시행 - 2020.12.01: 비수도권도 거리두기 단계 1.5단계 시행 - 2020.12.08: 수도권 거리두기 2.5단계, 비수도권 2 단계 시행	O
공통사항	2020년 5월 6일에 생활방역으로 전환하면서 해당 시점의 거리 두기 2단계 기준인 일일 코로나 확진자수 50명 이상을 독립변 수 선정 기준으로 적용함.	

※출처: 위키백과[22] 자료를 연구자가 재편집함

이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수, 코로나 확진자수에 대한 데이터 전처리 작업이 완료되면 최종적으로 단변량 선형회귀분석에 활용이 가능한 형태로 데이터를 결합하여 [그림 3-5]와 같은 구조로 생성한다.

기준 일자	방 문 객 수	대 중 교 통 지 출 금 액	레저스 포조지 출금액	엔터 카지 출금액	면세점 지출금 액	문화 서비스 지출금 액	소형지 출금액	숙박업 지출금 액	식음료 지출금 액	여행 업지 출금 액	카지 노지 출금 액	항공 사지 출금 액	공연 행 사 검 색 건 수	기 타 관 람 지 검 색 건 수	레 포 즈 검 색 건 수	문 화 시 설 검 색 건 수	소 형 검 색 건 수	숙 박 검 색 건 수	식 음 료 검 색 건 수	역 사 관 람 지 검 색 건 수	자 연 관 람 지 검 색 건 수	휴 양 관 람 지 검 색 건 수	확 진 자 수
20200311	90124	0	64454950	7728405	81316910	3223355	23954550	25290285	298749208	187000	0	4488600	6	1492	520	351	358	2093	4499	77	2151	262	242
20200312	95206	0	68724705	8011960	1525510	1940260	27706130	34557973	301282738	2524255	0	8373800	20	1526	590	309	362	2054	4486	70	2300	239	114
20200313	116578	80500	57223390	11234055	107296686	3043010	7522480	33496576	324673522	1077690	0	6546600	19	1686	570	384	413	2417	5005	83	2109	251	110
20200314	105410	57000	91763301	9517930	108870990	5705550	43144720	35103974	376373476	0	0	3780900	17	2004	739	484	604	2699	6714	117	2831	330	107
20200315	100027	0	92621847	5338860	127734590	7492340	39098430	38305752	344929320	29500	0	5532200	27	2052	657	475	488	2350	6722	145	2751	304	76
20200316	99308	0	49952510	12078278	103009840	3243850	28599900	31933354	274338449	287300	0	7825100	15	1690	519	369	372	2220	4829	87	2265	237	74
20200317	98618	0	62503215	7004170	100358470	2781450	25954990	27541860	280579207	6364500	0	6056500	13	1619	550	388	449	2104	4497	74	2392	240	84
20200318	99614	145000	65513070	8611565	85402100	2766450	27282800	20916182	291583085	8732400	0	7160114	10	1639	519	369	361	2068	4565	80	2372	242	93
20200319	100955	0	50689907	11243659	90761130	3467200	24647020	18786365	300863866	0	0	5399400	3	1531	430	444	425	2028	4616	91	1988	255	152
20200320	104538	0	66358100	13961015	91676330	2898270	29249470	33273566	312786223	442610	0	5594400	17	1737	653	339	373	2216	4961	83	2446	265	87
20200321	106959	156000	105136629	10285375	94465970	7944700	44756600	33874105	393543962	0	0	6390500	5	2079	856	387	456	2685	6531	134	3270	327	147
20200322	104468	223000	87163448	9857975	141162760	6375180	38701230	70603081	348401677	48000	0	4793000	15	1852	632	459	575	2300	6004	116	2607	276	98
20200323	105013	0	53607520	13688610	110889200	4966050	30969390	59651389	301808301	75700	0	6470100	7	1738	516	332	376	2179	4924	118	2423	272	64

[그림 3-5] 단변량 선형회귀분석용 일 데이터 전처리 결과

### 3.1.3 데이터 분석

본 연구에서는 사용하는 분석방법은 여행행태에 대한 비교분석 및 시각화와 단변량 선형회귀분석이다. 비교분석은 코로나가 우리나라에 처음 유입된 2020년과 전년 대비 비교분석을 위한 2019년 여행행태 데이터를 월별로 비교분석하는 것을 의미하며, 비교분석에 대한 데이터를 시각화로 표현하여 데이터를 효과적으로 전달하였다. 데이터 시각화는 많은 양의 데이터를 한눈에 볼 수 있게 하여 데이터의 변화량을 쉽게 파악할 수 있게 해주며 이를 통해 데이터 인사이트를 발견할 수 있게 도와주는 방법으로 R 프로그램에서 제공하는 plot 함수를 사용하였다.

단변량 선형회귀분석은 하나 또는 그 이상의 변수들이 다른 변수에 미치는 영향을 파악하는 분석방법으로 R 프로그램에서 제공하는 cor.test 함수를 사용해 상관관계 검정을 먼저 진행한 후 회귀분석은 lm 함수를 사용하였다.

cor.test 함수는 두 변수 사이의 상관 관계를 측정하기 위해 사용하며 추출한 상관계수 값은 0에서 1사이의 값을 지닌다. 결과 값이 0에 가까우면 두 변수 간의 관계가 낮고 1에 가까우면 관계가 높다는 의미이다.

lm 함수를 사용하여 추출한 회귀분석 결과는 summary 함수를 통해 상세한 분석 결과의 내용을 확인할 수 있으며, 분석 결과를 평가하는 방법은 다음과 같다.

첫째, 회귀 모형의 F 통계량과 p-value를 확인하여 통계적으로 유의미한 결과인지를 평가한다. p-value가 0.05 보다 작을 때 보통 회귀 모형이 유의하다고 볼 수 있으며, 이는 귀무가설을 기각하고 대립가설을 채택한다는 의미와 동일하다. 따라서, 대립가설은 독립변수가 종속변수에 영향을 미치는 것과 같다.

둘째, 회귀 모형에 적용된 독립변수의 B(Estimate: 비표준화 회귀계수), SE(std. Error: 표준오차), t(t-value: 검정통계량), p(Pr: p-value)의 출력값을 확인한다. B값은 회귀식의 기울기를 의미하는 통계량으로 독립변수가 1 증가할 때 종속변수가 얼마나 증가하는지를 나타낸다. SE값은 종속변수의 추정값이 얼마나 정확한지 측정할 때 표준오차를 통해 해당 모형의 정확성을 파악한다. t값과 p값은 회귀계수에 대한 가설검정을 하는 통계량으로 단변량 선형회귀분석에서는 회귀 모형의 결과가 유의하면 적용된 독립변수는 동일하게 유의미하다.

셋째, 결정계수를 통해 회귀 모형에서 사용한 독립변수가 종속변수를 얼마만큼 설명해 주는지를 확인한다. 결정계수 지표값이 높을수록 독립변수가 종속변수를 많이 설명한다고 볼 수 있다.

### 3.2 분석 대상 데이터

본 절에서는 분석에 사용되는 이동통신 방문자수, 신용카드 지출금액, 내비게이션 검색건수 데이터에 대한 의미를 한국관광 데이터랩 홈페이지에서 제공하는 자료[21] 기준으로 살펴보고자 한다.

이동통신 방문자수는 일상생활권(거주, 통근, 통학 등)을 벗어나 관광 등의 목적으로 한 장소에 일정 시간 이상 체류한 사람을 의미한다. 이동통신 방문자수는 KT의 분석기준에 의한 이동통신 서비스 인구를 의미하며 방문객을 판단하기 위해서는 관광지에 방문한 사람이 주기적이고 빈번한 방문이 아니어야 하고 일정한 시간 이상을 머물러 있어야 한다. 일정시간 이상은 해당 지자체 내 특정 기지국에 30분 이상 체류하는 것을 의미하며 동시에 방문한 해당 기간 이전의 1개월 동안 3회 이하 방문자만 데이터로 활용되었다. 또한, 본 연구에서는 내국인 방문자수에 대한 분석을 위해 해당 지자체의 거주자나 근무자를 의미하는 현지인과 외국인 데이터는 제외하였다.

신용카드 지출금액 데이터는 고객이 사용한 신용카드의 소비 정보를 익명화하여 처리한 데이터로 비씨카드에서 제공한 데이터를 활용했다.

신용카드 지출금액은 업종별로 [표 3-3]과 같이 상세하게 분류하고 있으며, 업종 중에서 카지노는 코로나 시기에 지출금액이 존재하지 않아 제외하였다. 카드 발급자의 주소지와 분석 대상의 지자체가 동일한 대상을 현지인으로 보고 있으며, 업종 중에서 쇼핑, 식음료만 현지인 제외 데이터를 제공하여 해당 데이터를 활용하였다. 그 외 업종별 지출금액은 비씨카드 이용고객의 소비액을 활용한 데이터로 현지인과 외부방문자 등 상세 데이터를 구분할 수 없다는 한계가 있다.

[표 3-3] 신용카드 업종 유형

구분	업종 유형
신용카드	숙박업(호텔, 콘도), 여행업(관광여행업), 대중교통(철도, 고속버스, 여객선), 항공사, 렌터카, 면세점, 카지노, 레저스포츠(종합레저타운, 골프 경기장), 식음료(한식, 중식, 주점, 제과점), 문화서비스(영화관), 쇼핑(면세점 제외)

내비게이션 검색건수 데이터는 SKT의 내비게이션 서비스인 티맵 데이터를 활용하였다. 내비게이션 검색건수는 [표 3-4]와 같이 목적지 유형별로 카테고리를 분류하고 있으며, 사용자가 목적지를 조회하고 해당 목적지를 1km 이상 이동한 대상에 대해서 건수를 집계했다. 내비게이션 검색건수 데이터의 한계로는 내비게이션을 사용하는 대상이 외부방문자의 관광목적이 아닌 일상적인 방문목적으로 사용하는 현지인도 포함되어 있을 수 있다.

[표 3-4] 내비게이션 목적지 유형

구분	목적지 유형
내비게이션	자연관광지(강, 공원, 섬, 자연휴양림, 폭포, 호수 등), 역사관광지(고궁, 절, 탑), 휴양관광지(동식물원, 온천, 수족관, 놀이시설, 테마파크), 기타관광지(관광단지, 공항, 전망대), 식음료(한식, 중식, 일식, 카페), 숙박(호텔, 펜션), 공연/행사(공연장, 극장, 영화관) 문화시설(미술관, 박물관, 전시관), 레포츠(골프장, 낚시, 경마장), 쇼핑(종합시장, 면세점, 재래시장)

따라서, 본 연구에서 활용하는 데이터는 정확한 수치보다는 빅데이터의 관점에서 다양한 여행행태를 분석하는 것에 의미를 두고 정확성과 절대값 수치 보다는 추세치를 중심으로 데이터를 활용했다.

## 제 4 장 여행행태 분석 결과

### 4.1 여행행태 비교분석 및 시각화

이번 절에서는 R 프로그램을 활용하여 코로나 시기의 여행행태를 비교분석하고 그에 해당하는 데이터를 시각화한 결과이다.

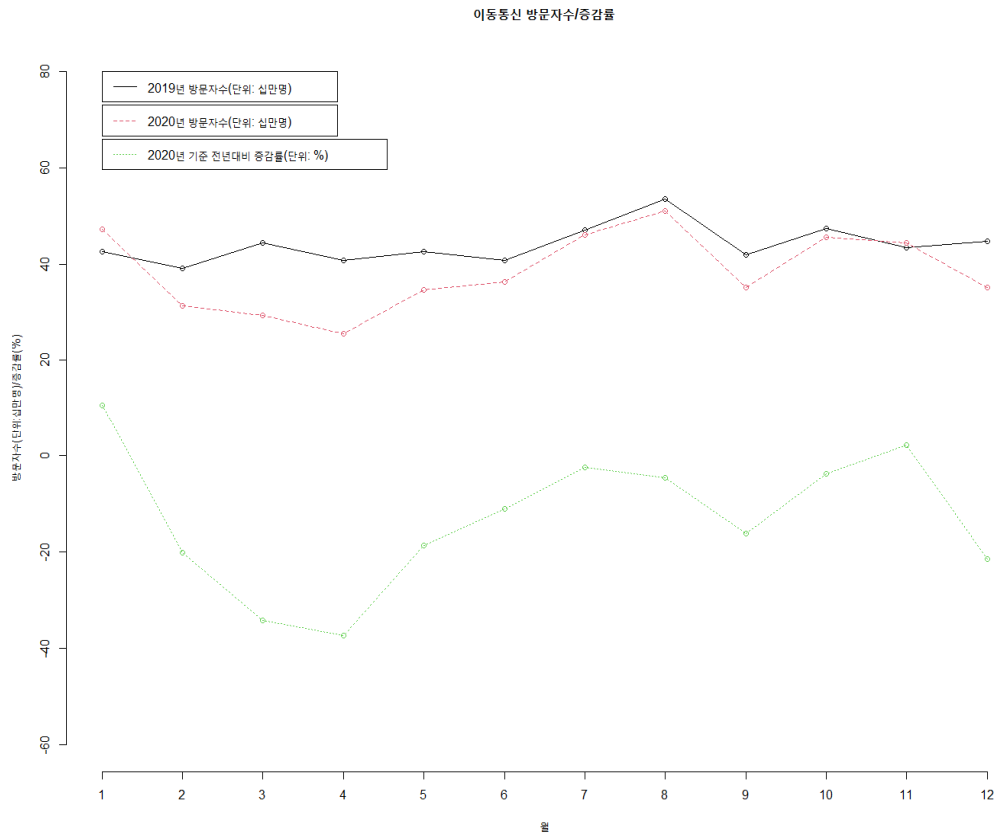
#### 4.1.1 이동통신 방문자수

[그림 4-1]은 이동통신 방문자수에 대한 비교분석 결과이다. 코로나가 국내에 유입되기 시작한 2020년 1월과 거리두기 단계가 1단계로 하향되었던 11월을 제외하고는 2020년의 이동통신 방문자수는 전체적으로 감소하였다. 특히, 2020년 기준 전년 대비 증감률이 가장 높은 1, 2순위가 각각 4월과 3월인 것은 정부가 2020년 3월 22일에 발표한 사회적 거리두기 대책과 위험시설의 운영제한 조치가 반영된 결과로 보인다. 비교분석 결과를 시각화한 화면은 [그림 4-2]이다.

월	2019년 방문자수 (단위: 십만명)	2020년 방문자수 (단위: 십만명)	2020년 기준 전년대 비 증감률(단위: %)	증감률 순위
1	4266230	4718360	10.60	12
2	3919934	3129308	-20.17	4
3	4446690	2928020	-34.15	2
4	4080300	2553394	-37.42	1
5	4266714	3470128	-18.67	5
6	4080814	3630736	-11.03	7
7	4713832	4600592	-2.40	10
8	5349022	5108012	-4.51	8
9	4199822	3522309	-16.13	6
10	4742090	4560640	-3.83	9
11	4348533	4443582	2.19	11
12	4481918	3518081	-21.51	3

[그림 4-1] 이동통신 방문자수 비교분석 결과





[그림 4-2] 이동통신 방문자수 시각화 결과

#### 4.1.2 신용카드 지출금액

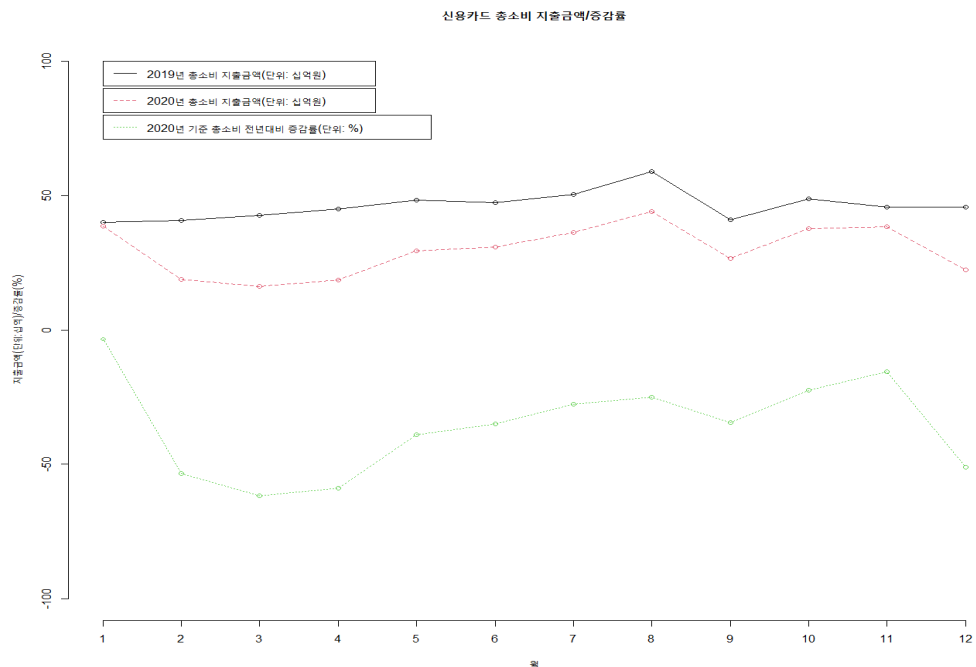
신용카드 지출금액은 총소비 항목을 포함한 11개의 여행행태 변수별로 분류하여 분석하였다. 총소비의 경우 2020년 기준 전년 대비 지출금액과 증감률에 대해 비교분석을 진행하였고, 그 외에 10개의 여행행태 변수별 비교분석에는 지출금액 대신 구성비로 대체하여 진행하였다. 구성비는 해당 월에 총소비 대비 분석하는 여행행태 변수가 차지하는 비율을 의미한다.

## 1) 총소비

신용카드 총소비 지출금액은 [그림 4-3]과 같이 2020년 기준 전년 대비 모든 월에서 감소하였다. 2, 3, 4월에는 전년 대비 총소비 50% 이상이 하락하였다. 코로나 확진자수가 줄어든 5월부터는 점차 회복세를 보였으나, 12월에 다시 총소비 지출금액이 급격히 감소하였다. 비교분석 내용을 시각화한 결과는 [그림 4-4]이다.

월	2019년 총소비 지출금액(단위: 원)	2020년 총소비 지출금액(단위: 원)	2020년 기준 총소비 전년대비 증감률(단위: %)	증감률 순위
1	40030727101	38590647097	-3.60	12
2	40834024061	18966282346	-53.55	3
3	42737673183	16288050121	-61.89	1
4	45108074492	18515120190	-58.95	2
5	48364148497	29455327617	-39.10	5
6	47484758749	30821841047	-35.09	6
7	50401951057	36401231316	-27.78	8
8	59000045887	44083368537	-25.28	9
9	41039373873	26747992746	-34.82	7
10	48868830428	37747414539	-22.76	10
11	45667063418	38449331436	-15.81	11
12	45754408556	22349214046	-51.15	4

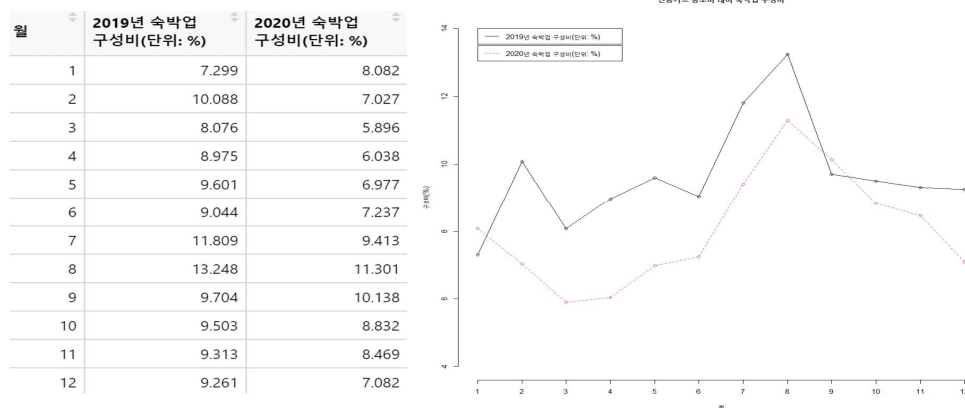
[그림 4-3] 신용카드 총소비 지출금액 비교분석 결과



[그림 4-4] 신용카드 총소비 지출금액 시각화 결과

## 2) 숙박업

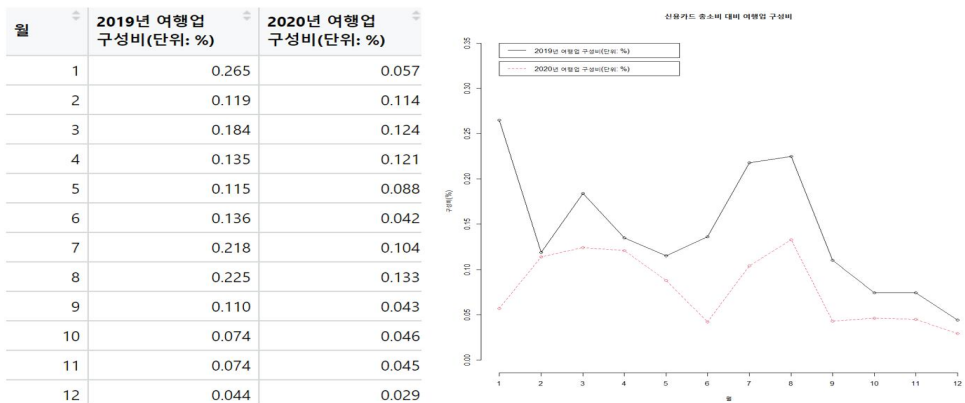
[그림 4-5]은 월별 신용카드 총소비 대비 숙박업이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 1, 9월을 제외한 모든 월에서 감소하여 숙박업소보다 캠핑이나 차박(차 안에서 하룻밤 자는 것)에 대한 관심이 증가한 것을 알 수 있다.



[그림 4-5] 신용카드 숙박업 지출금액 비교분석 결과

## 3) 여행업

[그림 4-6]은 월별 신용카드 총소비 대비 여행업이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 감소하여 여행업이 위기에 처한 상황을 알 수 있다.

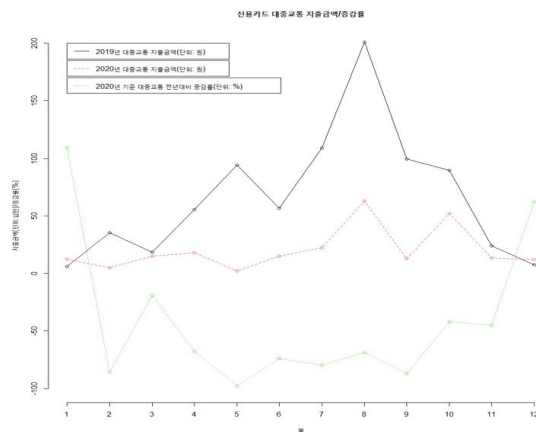


[그림 4-6] 신용카드 여행업 지출금액 비교분석 결과

#### 4) 대중교통

[그림 4-7]은 월별 신용카드 총소비 대비 대중교통이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 1, 12월을 제외한 모든 월에서 감소하였다.

월	2019년 대중교통 구성비 (단위: %)	2020년 대중교통 구성비 (단위: %)
1	0.001	0.003
2	0.009	0.003
3	0.004	0.009
4	0.012	0.010
5	0.020	0.001
6	0.012	0.005
7	0.022	0.006
8	0.034	0.014
9	0.024	0.005
10	0.018	0.014
11	0.005	0.003
12	0.002	0.005

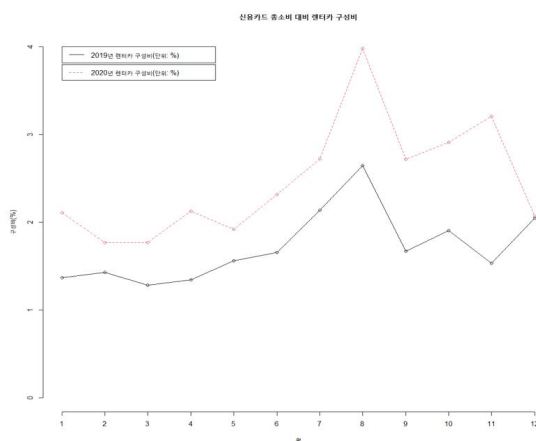


[그림 4-7] 신용카드 대중교통 지출금액 비교분석 결과

#### 5) 렌터카

[그림 4-8]은 월별 신용카드 총소비 대비 렌터카가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 증가하여 렌터카에 지출하는 비율이 높다는 것을 확인할 수 있다.

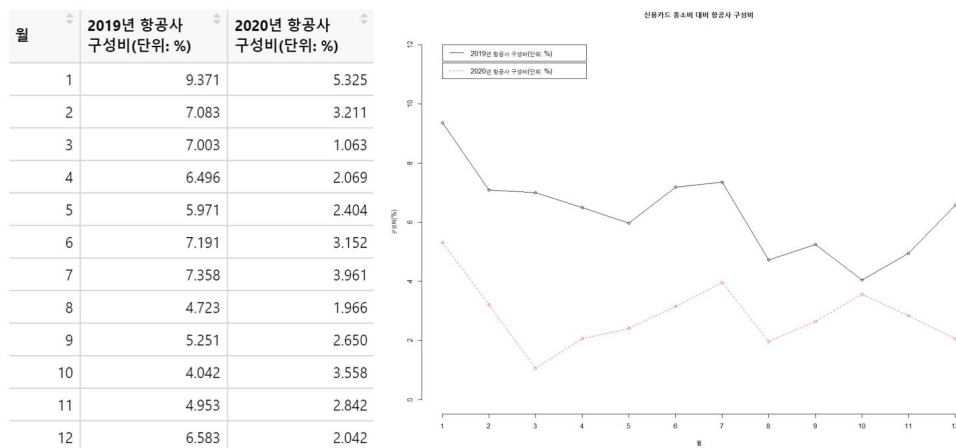
월	2019년 렌터카 구성비(단위: %)	2020년 렌터카 구성비(단위: %)
1	1.369	2.106
2	1.430	1.768
3	1.283	1.770
4	1.342	2.129
5	1.559	1.919
6	1.655	2.314
7	2.135	2.720
8	2.646	3.984
9	1.670	2.715
10	1.904	2.910
11	1.530	3.207
12	2.045	2.070



[그림 4-8] 신용카드 렌터카 지출금액 비교분석 결과

## 6) 항공사

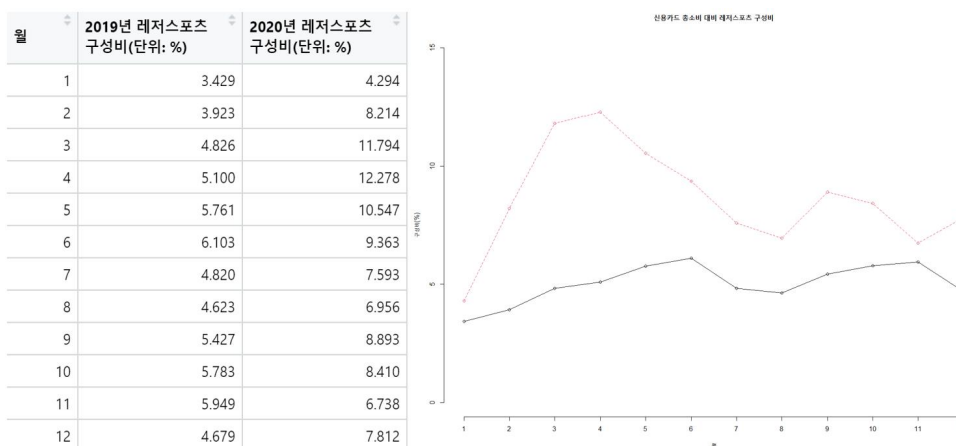
[그림 4-9]은 월별 신용카드 총소비 대비 항공사가 차지하는 구성비를 비교분석한 결과이다. 여행업과 동일하게 2020년 기준 전년 대비 모든 월에서 감소하였다.



[그림 4-9] 신용카드 항공사 지출금액 비교분석 결과

## 7) 레저스포츠

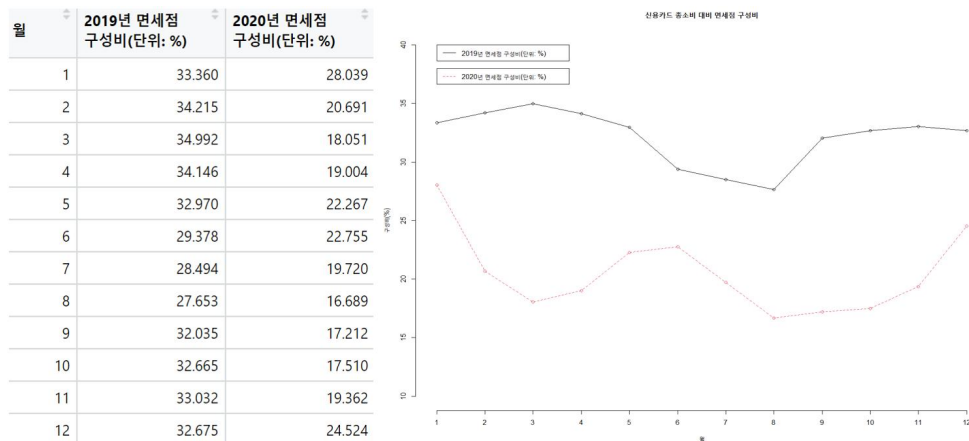
[그림 4-10]은 월별 신용카드 총소비 대비 레저스포츠가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 증가하였으며 2, 3, 4월에는 전년 대비 구성비가 2배 이상 증가하였다.



[그림 4-10] 신용카드 레저스포츠 지출금액 비교분석 결과

## 8) 면세점

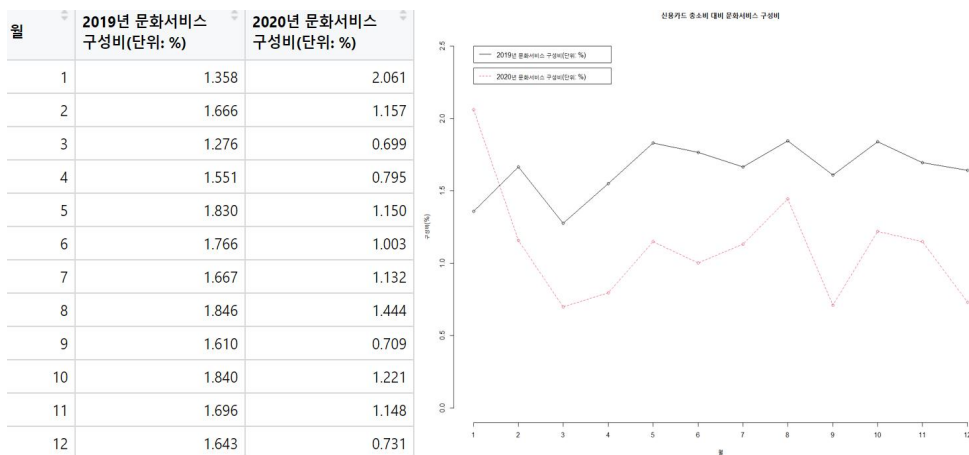
[그림 4-11]은 월별 신용카드 총소비 대비 면세점이 차지하는 구성비를 비교분석한 결과이다. 여행업, 항공사와 동일하게 2020년 기준 전년 대비 모든 월에서 감소하였다.



[그림 4-11] 신용카드 면세점 지출금액 비교분석 결과

## 9) 문화서비스

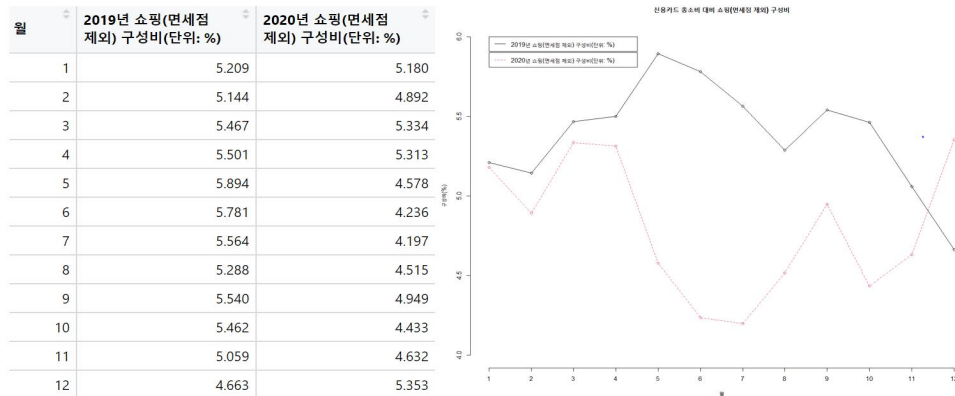
[그림 4-12]는 월별 신용카드 총소비 대비 문화서비스가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 1월을 제외한 모든 월에서 감소하였다.



[그림 4-12] 신용카드 문화서비스 지출금액 비교분석 결과

#### 10) 쇼핑(면세점 제외)

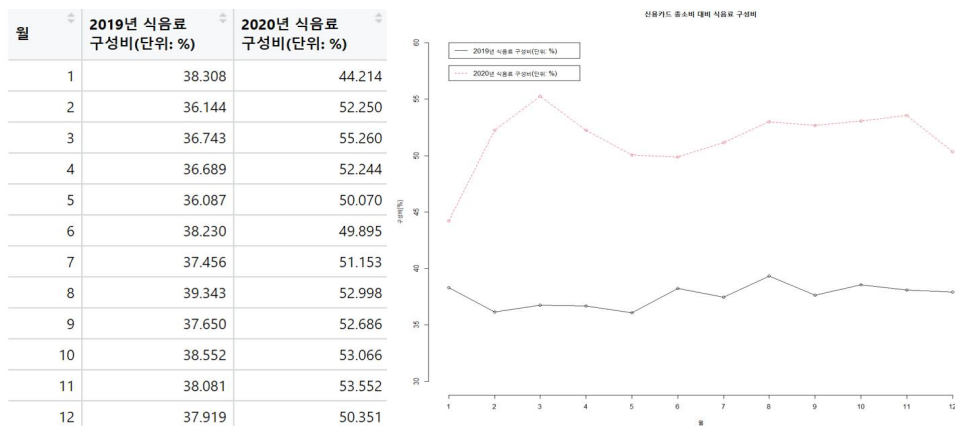
[그림 4-13]은 월별 신용카드 총소비 대비 쇼핑(면세점 제외)이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 12월을 제외한 모든 월에서 감소하였다.



[그림 4-13] 신용카드 쇼핑(면세점 제외) 지출금액 비교분석 결과

#### 11) 식음료

[그림 4-14]는 월별 신용카드 총소비 대비 식음료가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 증가하였고 1, 6월을 제외한 월에서는 구성비가 50% 이상을 차지하여 식음료에 지출하는 비율이 여행행태 변수 중에서 가장 높다는 것을 확인할 수 있다.



[그림 4-14] 신용카드 식음료 지출금액 비교분석 결과

신용카드 지출금액에서 총소비 항목을 제외한 10개의 여행행태 변수별 구성비 결과를 집계 요약하여 [표 4-1], [표 4-2]와 같이 작성하였다. 집계 계산식은 ‘2020년 구성비 - 2019년 구성비’로 결괏값이 (+)인 경우 해당 변수의 2020년 구성비가 2019년에 비해 증가한 것을 의미한다. 반면, (-)인 경우는 해당 변수의 2020년 구성비가 2019년에 비해 감소한 것을 의미한다.

2020년 기준 전년 대비 모든 월에서 구성비가 증가한 변수는 렌터카, 레저스포츠, 식음료이다. 다르게 말하면 3개 변수를 제외한 숙박업, 여행업, 대중교통, 항공사, 면세점, 문화서비스, 쇼핑(면세점제외)에 해당하는 7개 변수의 구성비는 코로나 영향으로 감소하였다.

이러한 결과는 제주특별자치도를 방문한 관광객의 지출 여행행태 변화를 확인할 수 있다. 대중교통보다는 자가 차량을 통해 인파가 몰리지 않는 안전하고 쾌적한 장소를 찾아 이동하는 관광객이 증가하고, 레저스포츠와 같이 폐쇄된 실내 관광지보다 자연 중심의 스포츠를 선호하는 것으로 확인된다. 또한, 코로나로 인해 단체 관광이 감소하거나 문화시설을 포함한 이용 제한 관광지가 증가하면서 자연스럽게 지출금액의 많은 부분이 식음료 위주로 지출하는 변화를 볼 수 있다.



[표 4-1] 신용카드 지출금액 여행행태 변수별 구성비 집계 - 1

월	여행행태 변수명				
	숙박업	여행업	대중교통	렌터카	항공사
1월	0.783	-0.208	0.002	0.737	-4.046
2월	-3.061	-0.005	-0.006	0.338	-3.872
3월	-2.18	-0.06	0.005	0.487	-5.94
4월	-2.937	-0.014	-0.002	0.787	-4.427
5월	-2.624	-0.027	-0.019	0.36	-3.567
6월	-1.807	-0.094	-0.007	0.659	-4.039
7월	-2.396	-0.114	-0.016	0.585	-3.397
8월	-1.947	-0.092	-0.02	1.338	-2.757
9월	0.434	-0.067	-0.019	1.045	-2.601
10월	-0.671	-0.028	-0.004	1.006	-0.484
11월	-0.844	-0.029	-0.002	1.677	-2.111
12월	-2.179	-0.015	0.003	0.025	-4.541

[표 4-2] 신용카드 지출금액 여행행태 변수별 구성비 집계 - 2

월	여행행태 변수명				
	레저스포츠	면세점	문화서비스	쇼핑 (면세점 제외)	식음료
1월	0.865	-5.321	0.703	-0.029	5.906
2월	4.291	-13.52	-0.509	-0.252	16.106
3월	6.968	-16.94	-0.577	-0.133	18.517
4월	7.178	-15.14	-0.756	-0.188	15.555
5월	4.786	-10.7	-0.68	-1.316	13.983
6월	3.26	-6.623	-0.763	-1.545	11.665
7월	2.773	-8.774	-0.535	-1.367	13.697
8월	2.333	-10.96	-0.402	-0.773	13.655
9월	3.466	-14.82	-0.901	-0.591	15.036
10월	2.627	-15.16	-0.619	-1.029	14.514
11월	0.789	-13.67	-0.548	-0.427	15.471
12월	3.133	-8.151	-0.912	0.69	12.432

### 4.1.3 내비게이션 검색건수

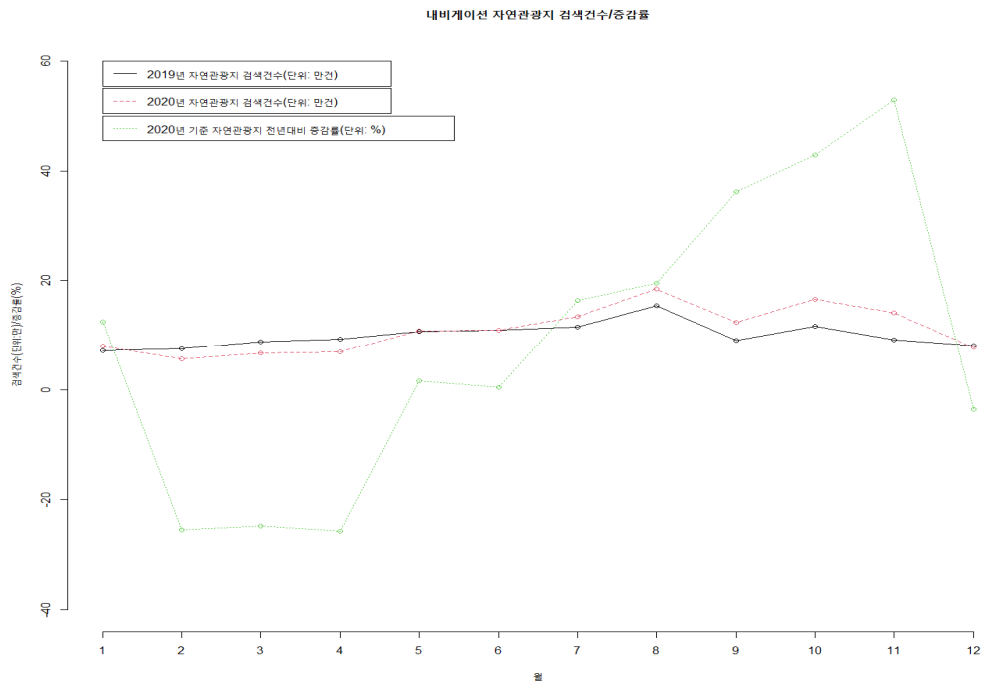
내비게이션 검색건수도 신용카드 지출금액 분석방법과 동일하게 전체 항목을 포함한 10개의 여행행태 변수별로 분류하여 분석하였다.

#### 1) 전체

내비게이션 전체 검색건수는 [그림 4-15]와 같이 2020년 기준 전년 대비 2, 3, 4, 12월에만 감소하였다. 이는 코로나가 국내에 유입되고 본격적으로 확산하기 시작한 2~4월에는 감소하고, 확진자수가 급격히 줄어든 5월부터 관광객들의 여행 소비 심리가 적용한 결과로 볼 수 있다. 또한, 이동통신 방문자수와 신용카드 지출금액이 2020년 기준 전년 대비 모든 월에서 감소했지만, 내비게이션 검색건수가 5~11월에 증가한 이유는 코로나 시기에 방문이나 영업을 가능한 곳을 찾거나 유명 관광지가 아닌 다양한 관광지를 찾고자 하는 심리가 반영된 결과로 보인다. 비교분석 내용을 시각화한 결과는 [그림 4-16]이다.

월	2019년 전체 검색건수 (단위: 건)	2020년 전체 검색건수 (단위: 건)	2020년 기준 전체 전년대비 증감률(단위: %)	증감률 순위
1	562828	732249	30.10	10
2	560136	449450	-19.76	3
3	601279	459118	-23.64	1
4	624579	478043	-23.46	2
5	701107	741688	5.79	5
6	744430	807145	8.42	6
7	785815	970457	23.50	7
8	995651	1270370	27.59	8
9	664294	858603	29.25	9
10	836393	1167318	39.57	11
11	725963	1115207	53.62	12
12	766380	660961	-13.76	4

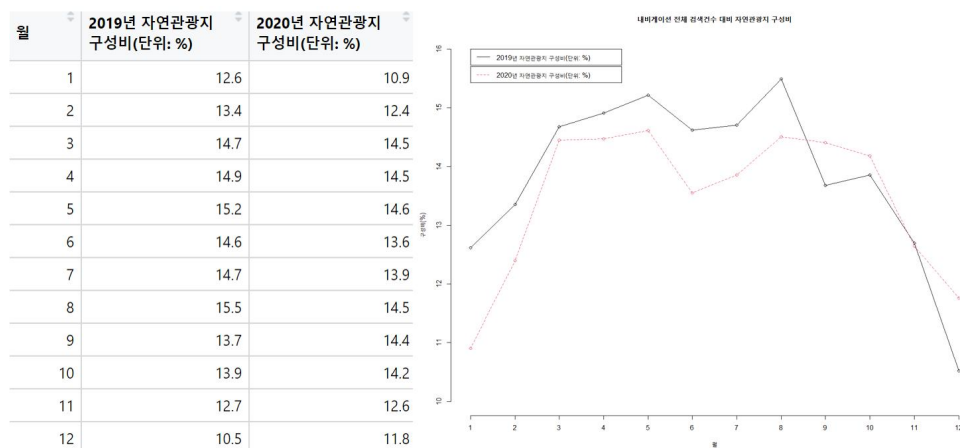
[그림 4-15] 내비게이션 전체 검색건수 비교분석 결과



[그림 4-16] 내비게이션 전체 검색건수 시각화 결과

## 2) 자연관광지

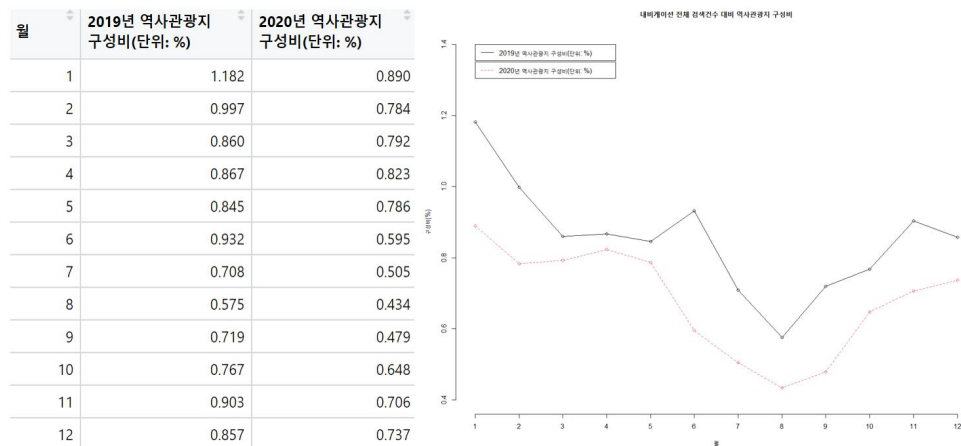
[그림 4-17]은 월별 내비게이션 전체 검색건수 대비 자연관광지가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 9, 10, 12월을 제외한 모든 월에서 감소하였다.



[그림 4-17] 내비게이션 자연관광지 검색건수 비교분석 결과

### 3) 역사관광지

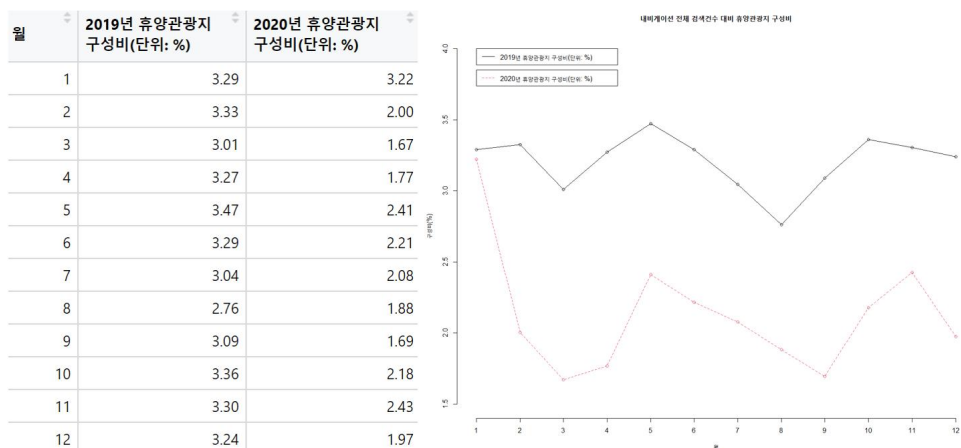
[그림 4-18]은 월별 내비게이션 전체 검색건수 대비 역사관광지가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 감소하였다.



[그림 4-18] 내비게이션 역사관광지 검색건수 비교분석 결과

### 4) 휴양관광지

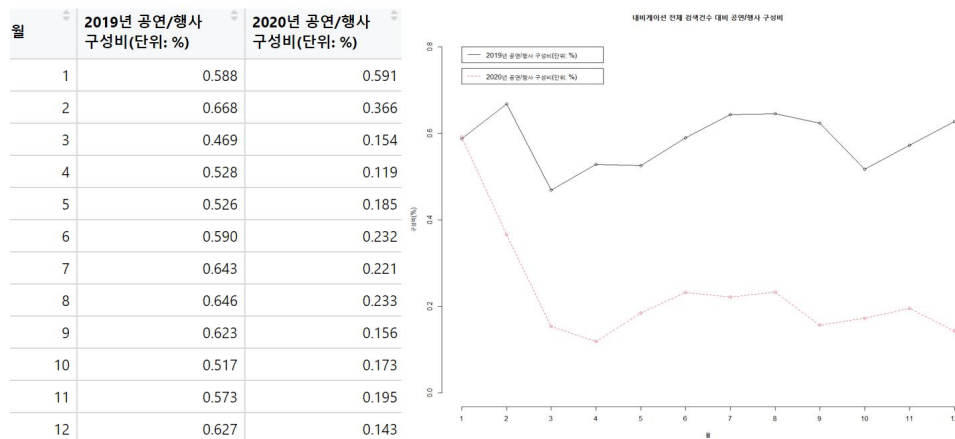
[그림 4-19]는 월별 내비게이션 전체 검색건수 대비 휴양관광지가 차지하는 구성비를 비교분석한 결과이다. 역사관광지와 동일하게 2020년 기준 전년 대비 모든 월에서 감소하였다.



[그림 4-19] 내비게이션 휴양관광지 검색건수 비교분석 결과

## 5) 공연/행사

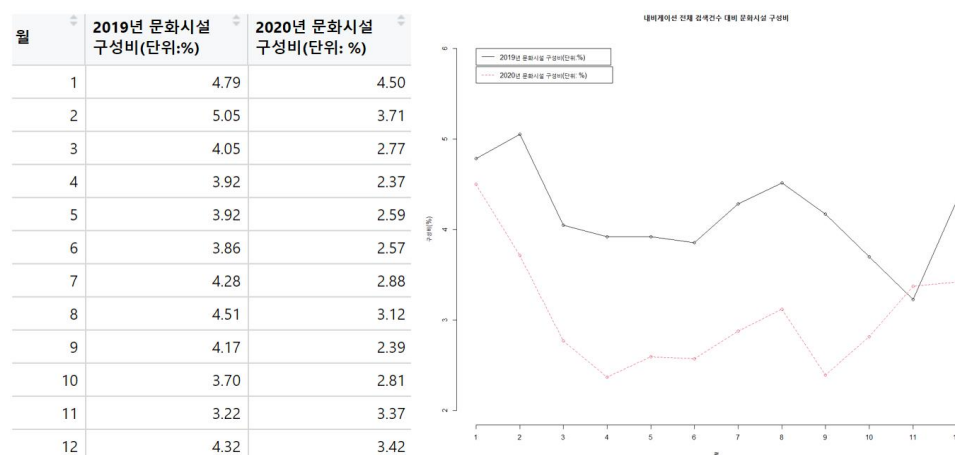
[그림 4-20]은 월별 내비게이션 전체 검색건수 대비 공연/행사가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 감소하여 공연과 행사 관련 업종이 위기에 처한 상황을 알 수 있다.



[그림 4-20] 내비게이션 공연/행사 검색건수 비교분석 결과

## 6) 문화시설

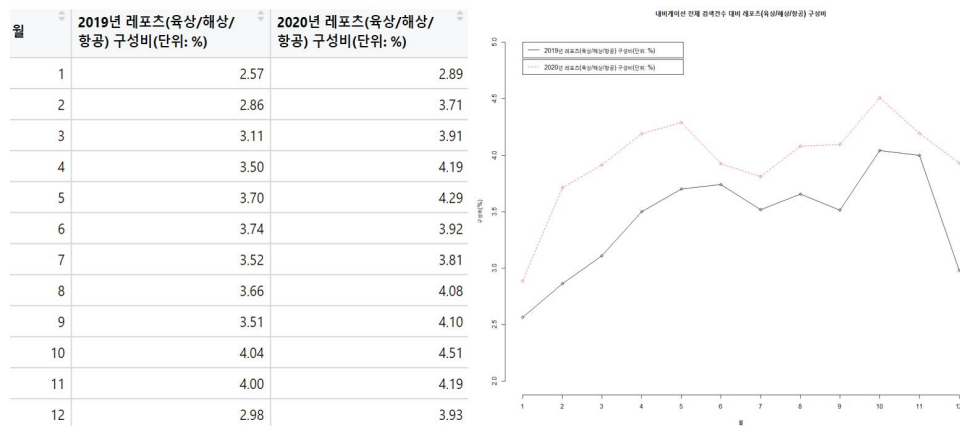
[그림 4-21]은 월별 내비게이션 전체 검색건수 대비 문화시설이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 11월을 제외한 모든 월에서 감소하였다.



[그림 4-21] 내비게이션 문화시설 검색건수 비교분석 결과

## 7) 레포트즈

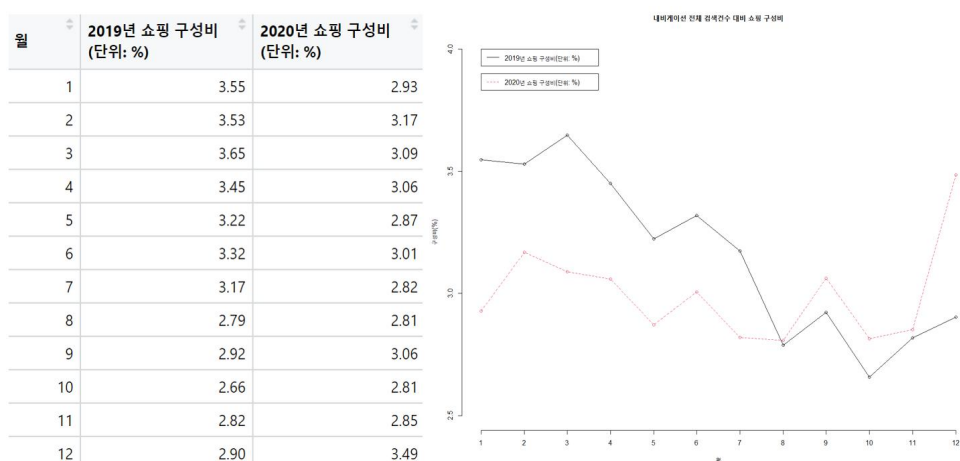
[그림 4-22]은 월별 내비게이션 전체 검색건수 대비 레포트즈가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 증가하였다.



[그림 4-22] 내비게이션 레포트즈 검색건수 비교분석 결과

## 8) 쇼핑

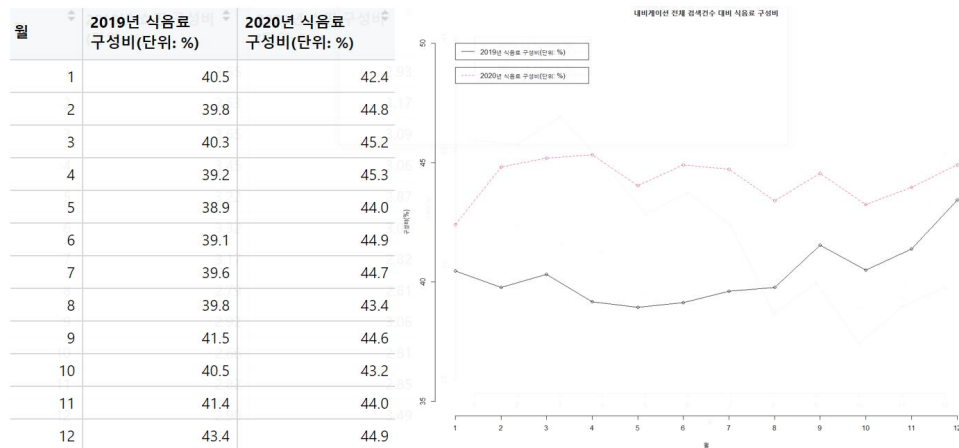
[그림 4-23]은 월별 내비게이션 전체 검색건수 대비 쇼핑이 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 1~7월까지의 감소하다가 8~12월에는 증가하였다.



[그림 4-23] 내비게이션 쇼핑 검색건수 비교분석 결과

## 9) 식음료

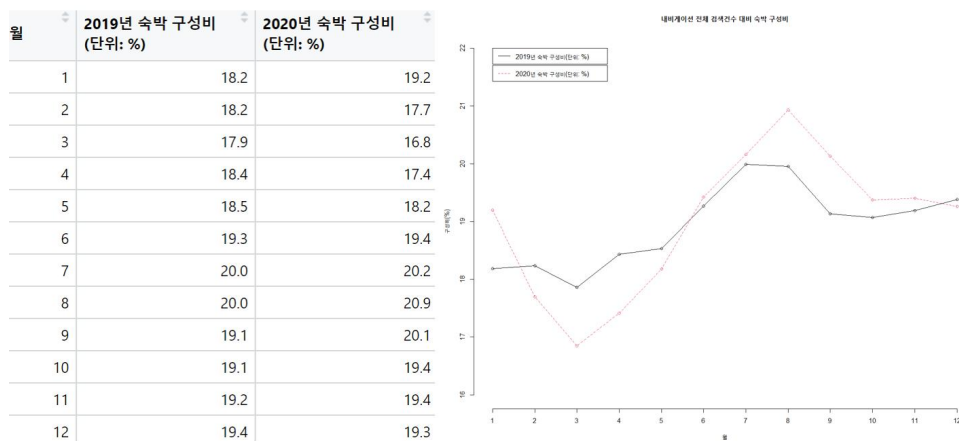
[그림 4-24]은 월별 내비게이션 전체 검색건수 대비 식음료가 차지하는 구성비를 비교분석한 결과이다. 2020년 기준 전년 대비 모든 월에서 증가하였다.



[그림 4-24] 내비게이션 식음료 검색건수 비교분석 결과

## 10) 숙박

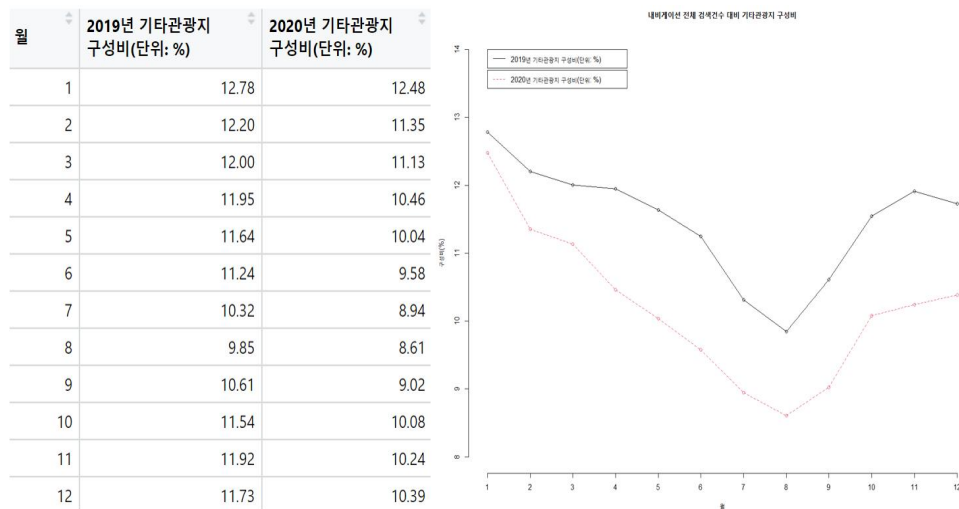
[그림 4-25]은 월별 내비게이션 전체 검색건수 대비 숙박이 차지하는 구성비를 비교분석한 결과이다. 코로나가 국내에 유입되어 확진자수가 증가하기 시작한 2020년 2~5월이 전년 동월과 대비하여 감소하였다.



[그림 4-25] 내비게이션 숙박 검색건수 비교분석 결과

## 11) 기타관광지

[그림 4-26]는 월별 내비게이션 전체 검색건수 대비 기타관광지가 차지하는 구성비를 비교분석한 결과이다. 역사관광지, 휴양관광지와 동일하게 2020년 기준 전년 대비 모든 월에서 감소하였다.



[그림 4-26] 내비게이션 기타관광지 검색건수 비교분석 결과

내비게이션 전체 검색건수 항목을 제외한 10개의 여행행태 변수별 구성비 결과를 신용카드 구성비 집계 방식과 동일하게 요약하여 [표 4-3], [표 4-4]와 같이 작성하였다.

2020년 기준 전년 모든 월에서 구성비가 증가한 변수는 레포츠, 식음료 항목이다. 반면에, 코로나의 영향으로 문화시설, 공연/행사와 대부분의 관광지 검색건수는 감소하였다.

신용카드 지출금액 비교분석과 동일하게 레포츠의 구성비가 전년 대비 모든 월에서 증가하고 자연관광지의 구성비가 9, 10, 12월에 증가한 결과에 따라 개방된 자연 중심의 관광지나 스포츠 중심으로 선호하는 결과를 볼 수 있다.



[표 4-3] 내비게이션 검색건수 여행행태 변수별 구성비 집계 - 1

월	여행행태 변수명				
	자연관광지	역사관광지	휴양관광지	공연/행사	문화시설
1월	-1.715	-0.292	-0.067	0.003	-0.287
2월	-0.956	-0.214	-1.323	-0.302	-1.339
3월	-0.226	-0.068	-1.341	-0.315	-1.280
4월	-0.439	-0.044	-1.506	-0.410	-1.549
5월	-0.601	-0.059	-1.061	-0.341	-1.324
6월	-1.070	-0.337	-1.074	-0.358	-1.284
7월	-0.850	-0.204	-0.967	-0.422	-1.405
8월	-0.983	-0.141	-0.877	-0.413	-1.396
9월	0.734	-0.240	-1.396	-0.467	-1.778
10월	0.328	-0.120	-1.183	-0.344	-0.883
11월	-0.052	-0.197	-0.877	-0.378	0.145
12월	1.237	-0.120	-1.265	-0.484	-0.902

[표 4-4] 내비게이션 검색건수 여행행태 변수별 구성비 집계 - 2

월	여행행태 변수명				
	레포트즈	쇼핑	식음료	숙박	기타관광지
1월	0.321	-0.620	1.947	1.013	-0.303
2월	0.850	-0.362	5.036	-0.538	-0.852
3월	0.803	-0.559	4.865	-1.009	-0.869
4월	0.690	-0.390	6.160	-1.021	-1.490
5월	0.589	-0.353	5.100	-0.353	-1.598
6월	0.184	-0.314	5.769	0.153	-1.669
7월	0.292	-0.356	5.116	0.168	-1.373
8월	0.423	0.020	3.631	0.977	-1.240
9월	0.581	0.139	3.019	0.995	-1.588
10월	0.468	0.157	2.743	0.300	-1.467
11월	0.193	0.034	2.596	0.211	-1.675
12월	0.950	0.582	1.463	-0.121	-1.340

## 4.2 코로나 확진자수 단변량 선형회귀분석

이번 절에서는 코로나 확진자수의 변화가 여행행태 변수에 얼마나 영향을 미치는지 파악하기 위해 단변량 선형회귀분석을 진행한 결과이다. 분석 결과는 [표 4-5]와 같다. 독립변수인 코로나 확진자수에 유의한 종속변수를 살펴보면 신용카드 지출금액에서는 레저스포츠(0.000), 숙박업(0.001), 항공사(0.001), 식음료(0.002), 문화서비스(0.025), 렌터카(0.025), 대중교통(0.046) 항목이 유의하였다. 또한, 내비게이션 지출금액에서는 자연관광지(0.000), 레포츠(0.002), 식음료(0.004), 숙박(0.008), 공연/행사(0.015), 기타관광지(0.022) 항목의 유의확률이 0.05보다 작아 유의한 결과를 보였다. 반면에, 이동통신 방문자수( $1.945e-05$ )와 신용카드 지출금액에서의 여행업(0.059), 면세점(0.255), 쇼핑(0.279) 그리고 내비게이션 검색건수에서의 휴양관광지(0.0768), 쇼핑(0.17), 문화시설(0.387), 역사관광지(0.712) 항목은 상관계수 결과에 따라 코로나 확진자수와의 상관관계가 낮은 항목으로 나왔으며, 회귀식에서 제시한 회귀계수를 통계적으로 인정하기 어려워 유의미한 회귀결과를 도출하지 못하였다.

회귀식이 얼마나 데이터를 잘 설명하고 있는지에 대한 결정계수( $R^2$ ) 통계량을 살펴보면 신용카드 지출금액 중에 레저스포츠가 0.194로 가장 높게 나왔다. 결정계수 기준은 학자들마다 차이가 있지만<sup>1)</sup> 본 연구의

---

1) Cohen(1988)이 제시한 결정계수( $R^2$ )의 기준은 small ( $0.02 \leq R^2 < 0.13$ ), middle ( $0.13 \leq R^2 < 0.26$ ), large( $0.26 \leq R^2$ )로 분류했다. Falk와 Miler(1992)는 결정계수( $R^2$ )가 0.10 이상 넘으면 모형의 설명력을 충족시키는 것으로 판단했다. 본 연구에서는 신용카드 지출금액 중에 레포츠가 0.194이고 내비게이션 검색건수 중에 자연관광지가 0.153으로 위에서 언급한 학자들이 제시한 결정계수의 기준을 충족하여 해당 변수가 모형의 설명력을 충족시키는 유의미한 결과로 볼 수 있다.

단변량 선형회귀분석은 코로나 확진자수와 여행행태 변수간의 관계를 추정하고 유의미한 영향력 분석이 목적이므로, 결정계수보다 여행행태 변수별 회귀계수의 유의성 결과가 중요하다고 볼 수 있다.

[표 4-5] 코로나 확진자수 기준 여행행태 변수별 단변량 선형회귀분석 결과

구분	종속변수	비표준화 회귀계수(B)	유의확률 (p-value)	결정계수 (R <sup>2</sup> )	상관계수
이동통신	방문자수	-3.409e-01	0.964	1.945e-05	-0.004
신용카드	레저스포츠	-38190	1.46e-06 ***	0.1941	-0.44
	숙박업	-41929	0.00081 ***	0.099	-0.314
	항공사	-10324	0.00092 ***	0.0972	-0.311
	식음료	-140387	0.00219 **	0.08357	-0.29
	문화서비스	-4120	0.0248 *	0.04575	-0.21
	렌터카	-9399	0.025 *	0.0455	-0.213
	대중교통	-93.3	0.046 *	0.0365	-0.191
	여행업	-781	0.059 .	0.0327	-0.181
	면세점	25614	0.255	0.012	-0.110
	쇼핑 (면세점 제외)	-4886	0.279	0.01085	-0.104
내비 게이션	자연관광지	-1.8207	2.44e-05 ***	0.1527	-0.391
	레포츠	-0.39	0.00189 **	0.0859	-0.293
	식음료	-2.98	0.00376 **	0.07512	-0.274
	숙박	-1.3372	0.00807 **	0.06319	-0.251
	공연/행사	-0.016649	0.0151 *	0.05345	-0.231
	기타관광지	-0.534	0.0218 *	0.04778	-0.218
	휴양관광지	-0.10246	0.0768 .	0.0287	-0.169
	쇼핑	-0.08801	0.17	0.01734	-0.131
	문화시설	0.0681	0.387	0.006926	0.0832
	역사관광지	-0.004993	0.712	0.001266	-0.036

## 제 5 장 결 론

본 연구는 국내 제주특별자치도를 중심으로 코로나 시대의 여행행태를 분석하여 코로나 시대 관광산업 맞춤형 서비스 제공 모형개발의 기초 자료로 활용되는 것을 목표로 하였다. 분석에 필요한 자료는 공공데이터를 활용하여 한국관광데이터랩, 공공데이터포털 등에서 제공하는 이동통신 방문자수, 신용카드 지출금액 관련 변수 11개, 내비게이션 검색건수 관련 변수 11개, 코로나 확진자수 데이터를 수집하여 사용하였다. 분석방법은 R프로그램을 사용하여 코로나가 국내에 유입된 2020년 기준으로 전년 대비 증감률과 구성비가 포함된 여행행태 비교분석과 시각화를 진행하였고, 코로나 확진자수가 여행행태 변수에 미치는 영향을 파악하기 위해 단변량 선형회귀분석을 실시하였다.

비교분석 결과에 의하면, 이동통신 방문자수는 2020년 기준 전년 대비 1월, 11월을 제외한 모든 월에서 감소하였고, 신용카드 총소비 지출금액은 2020년 기준 전년 대비 모든 월에서 감소하였다. 신용카드 지출금액과 관련된 여행행태 변수별 구성비 결과를 살펴보면 렌터카, 레저스포츠, 식음료 항목이 전년 대비 모든 월에서 증가하였다. 내비게이션 전체 검색건수는 코로나가 국내에 유입되고 본격적으로 확산하기 시작한 2020년 2~4월이 전년과 대비하여 감소하였고, 코로나 확진자수가 급격히 줄어든 5월부터 관광객들의 여행 소비 심리가 적용하여 5~11월에는 증가하였다. 내비게이션 검색건수와 관련된 여행행태 변수별 구성비 결과를 살펴보면 레포츠, 식음료 항목이 전년 대비 모든 월에서 증가했지만, 코로나의 영향으로 문화시설, 공연/행사와 관광지와 관련된 검색건수는 감소하였다. 단변량 선형회귀분석 결과를 살펴보면, 코로나 확진자수의 변화

에 영향을 받는 변수는 유의확률과 결정계수에 의해 신용카드 지출금액에서는 레저스포츠(0.000), 숙박업(0.001), 항공사(0.001), 식음료(0.002), 문화서비스(0.025), 렌터카(0.025), 대중교통(0.046) 항목 순으로 유의하였고 내비게이션 지출금액에서는 자연관광지(0.000), 레포츠(0.002), 식음료(0.004), 숙박(0.008), 공연/행사(0.015), 기타관광지(0.022) 항목 순으로 유의한 결과를 나타냈다. 비교분석과 단변량 선형회귀분석 결과에 의하면 코로나 시기에 제주특별자치도를 방문한 관광객의 여행행태는 인파가 물리지 않는 안전하고 쾌적한 장소를 찾아 이동하고 폐쇄된 실내 관광지보다 레저스포츠와 같은 자연 중심의 스포츠를 선호하는 것으로 확인된다. 반면에 코로나로 인해 이용 제한이 있는 문화시설이나 인파가 물리는 관광지에 대한 거부감이 확대된 것으로 보인다.

본 연구는 공공데이터를 활용하여 여행행태를 분석했다는 점에서 의미가 있지만 수집한 여행행태의 데이터가 빅데이터의 관점에서 추세치를 분석하기에 용이한 데이터로 정확한 수치를 제공하는 데는 한계가 존재한다. 또한, 독립변수인 코로나 확진자수 기준으로 단변량 선형회귀분석을 실시하여 다양한 변수를 고려하지 않았다. 따라서 향후 연구에서는 공공데이터를 활용한 다양한 변수가 제공된다면 방문객의 유형(외국인, 현지인, 내국인 방문자)과 성별, 연령대, 방문 시간대, 기상청 데이터 등과 같은 변수를 투입한 다변량 선형회귀분석을 실시하여 좀 더 세밀하고 의미 있는 분석이 이루어질 수 있을 것이다.

본 연구에서 활용한 여행행태 분석 결과는 향후 포스트 코로나 시대에 대응하여 관광산업 맞춤형 서비스를 제공하는 기초 자료로 활용될 수 있고 유의미한 여행행태 분석 데이터를 제공하여 관광 수요에도 도움이 될 것이라 기대한다. 또한, 코로나 사태와 같은 전염병 유행에 대비하여 효과적인 관광산업의 대책 수립에 기여할 것으로 기대된다.

## 참고문헌

- [1] 질병관리청, “코로나 바이러스 감염증-19”, <http://ncov.mohw.go.kr/>
- [2] 한국문화관광연구원, “(보고서)2020년 문화체육관광동향조사\_BSI\_보고서(4분기)”, <https://www.kcti.re.kr/web/board/boardContentsView.do>
- [3] 박근화, “코로나19로 인한 국민의 국내 여행 행태 변화”. 학술저널, 138-144, 2020.12
- [4] 안동, 유효강, “코로나19의 발생이 중국관광산업에 미치는 영향 및 대응책”, 한국관광산업학회, 309-325, 2020.06
- [5] 안필균, 엄성준, 조숙영, 김상범, “포스트 코로나 시대 관광 트렌드를 반영한 농촌체험마을 조성방안 연구- 전라북도 완주군 소양면 위봉마을을 사례로 -”, 한국농촌계획학회, 27-39, 2020.11
- [6] 홍민정, 오문향, “코로나19 확산에 대한 국내 잠재 관광객의 감정 반응 연구:의미론적 네트워크 분석의 활용”, 대한관광경영학회, 47-65, 2020.04
- [7] 김성태, “코로나19 확산에 따른 관광위험지각과 감정반응이 행동의도에 미치는 영향”, 한국호텔리조트학회, 135-150, 2020.10
- [8] 김영남, 홍성화, “코로나 19 위험 인식과 낙관적 편향에 관한 탐색적 연구: 제주 방문 관광객을 중심으로”, 한국관광학회, 79-101, 2020.11
- [9] 정국환, 문정욱, 이시직, 유지연, 한은영, 왕재선, 서혁준, “공공데이터 개방·활용 성과측정을 위한 평가모델 연구”, 정보통신정책연구원, 17-18, 2013.12
- [10] 법제처 국가법령정보센터, “공공데이터의 제공 및 이용 활성화에 관한 법률 시행령”, <http://www.law.go.kr>
- [11] 행정안전부, “공공 빅데이터 우수사례집”,

[https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR\\_000000000015&nttId=61876](https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000015&nttId=61876)

[12] 이미경, “소셜 빅데이터를 이용한 한국인의 여행트렌드 분석 - ‘가족여행’과 ‘나홀로여행’을 중심으로 -”. 관광학연구, 111-134, 2018.11

[13] 통계청, “2020 통계로 보는 1인가구”,  
[http://kostat.go.kr/portal/korea/kor\\_nw/1/1/index.board?bmode=read&aSeq=386517](http://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=386517)

[14] 한국관광 데이터랩, “관광실태조사 외래관광객 조사”,  
<https://datalab.visitkorea.or.kr/datalab/portal/tae/getOseaTourExmn.do>

[15] 이재민, “새로운 관광콘텐츠로서 나홀로 여행족의 사회문화적 배경 및 특성”, 한국콘텐츠학회, 241-242, 2015

[16] 위키백과, “대한민국의 코로나19 범유행의 경과”, 2021년 3월 5일 접속,  
[https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%EC%9D%98\\_%EC%BD%94%EB%A1%9C%EB%82%9819\\_%EB%B2%94%EC%9C%A0%ED%96%89%EC%9D%98\\_%EA%B2%BD%EA%B3%BC](https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%EC%9D%98_%EC%BD%94%EB%A1%9C%EB%82%9819_%EB%B2%94%EC%9C%A0%ED%96%89%EC%9D%98_%EA%B2%BD%EA%B3%BC)

[17] 왕월, 심재연, 김형호, “가족여행에서 가족 간의 상호작용 행동이 체험 가치에 미치는 영향에 관한연구”, 한국융합학회논문지. 한국융합학회, 101 - 108, 2019

[18] 김상태, “2019년 국내 가족여행 분석”, 한국문화관광연구원, 150-156, 2020.7

[19] 나한울, “여행사의 SNS활용특성이 소비자 온라인여행상품 구매의도에 미치는 영향.”, 국내석사학위논문 호남대학교 대학원, 2013

[20] 김문정, “SNS 관광정보 특성이 관광지 선택에 미치는 영향”, 디지

털문화아카이브지, 2020.8

[21] 한국관광데이터랩, “[발간보고서] 관광 빅데이터 분석 가이드라인  
수립 연구(최종보고서)”,

<https://datalab.visitkorea.or.kr/site/portal/ex/bbs/View.do?cbIdx=1129&bcIdx=295807&pageIndex=1&tgtTypeCd=&searchKey=&searchKey2=&cateCont=tl02>

[22] 위키백과, “사회적 거리두기”, 2021년 3월 27일 접속,

[https://ko.wikipedia.org/wiki/%EC%82%AC%ED%9A%8C%EC%A0%81\\_%EA%B1%B0%EB%A6%AC%EB%91%90%EA%B8%B0](https://ko.wikipedia.org/wiki/%EC%82%AC%ED%9A%8C%EC%A0%81_%EA%B1%B0%EB%A6%AC%EB%91%90%EA%B8%B0)

#### <공공데이터 활용 사이트>

한국관광 데이터랩: <https://datalab.visitkorea.or.kr>

공공데이터포털: <https://data.go.kr>