

2024 관광 데이터 분석 공모전



목차



주제 선정



활용 데이터
및 분석 과정



데이터 분석



정책 제안 및
기대효과

1. 주제 선정 - (1) 제안 배경

① 고물가로 인한 관광 산업 불황

지속적으로 증가하고 있는
소비자 물가지수



출처 : 컨슈머인사이트(2023.09.26)

알뜰여행에서 초긴축여행으로…
여행산업 위축 시작됐다

- ‘주례 여행기획조사’ … 국내여행 비용지출 추이 비교
- 국내여행 1인당 23만3000원, 하루 7만6600원 지출
 - 작년 8만원대 중반에서 1년 새 10% 이상 늘어들어
 - ’19년 대비 4.3% 늘었지만 물가상승률 크게 밀돌아
 - 숙소 선택 때 가격 최우선 고려 등 지출 최소화 고민

고물가로 인해 여행 기간과 1일 경비 모두 감소하는 추세
여행 비용에 민감하게 반응하는 경향 증가
→ 여행 소비 위축

1. 주제 선정 - (1) 제안 배경

② 지역별 관광의 불균형

출처 : 여행신문(2023.07.31)

위기에 처한 인구감소지역... 여행으로 달라질 수 있는 이유

지방 도시들은 인구가 소멸할 수 있다는 위기마저 맞닥뜨렸다. 2023년 현재 우리나라는 89개 지역을 인구감소 지역으로 지정했다. 그리고 사라질 위기에 처한 지방 소도시들을 살리기 위한 방안으로 관광 산업을 주목하고 있다. 한국관광공사와 문화체육관광부는 인구소멸위기 지역으로 관광 수요를 유도하는 사업을 지속적으로 진행하고 있다.

출처 : channelPNU(2023.11.02)

각양각색 과잉 관광 대응법... 지속가능한 해법이 필요해

‘과잉 관광’(오버투어리즘, overtourism)이란 수용 가능 인원보다 많은 관광객이 지역을 방문하는 현상과 그로 인한 부작용을 말한다. 코로나19 팬데믹으로 한동안 잠잠했던 과잉 관광 문제가 세계 곳곳에서 터지고 있다. 한국문화관광연구원 박주영 관광산업연구실장은 <채널 PNU>와의 인터뷰에서 (...중략...) “관광객이 특정 시기에 너무 몰리지 않도록 숨겨진 명소를 찾고 질서를 지키는 등 관광객 분산이 필요하다”고 덧붙였다.

관광 비활성화 지역
관광을 통한 지역 발전 노력

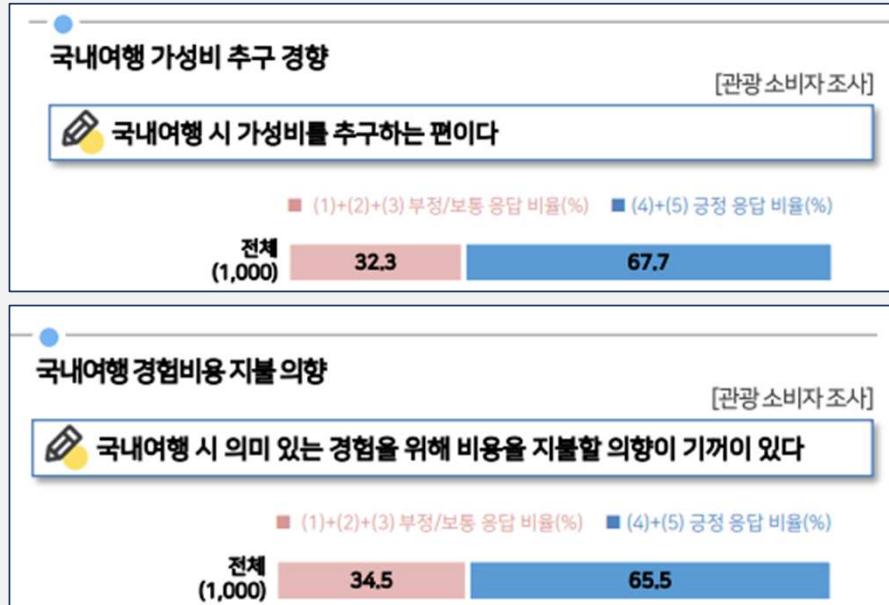
관광 활성화 지역
과잉 관광 해결 위한 관광 분산 필요

VS

1. 주제 선정 - (1) 제안 배경

③ 여행 소비 트렌드 : 가심비 VS 가성비

출처 : 한국관광공사(2023)



국내여행시 가성비를 추구하는 비율 및 의미 있는 경험을 위해 비용을 추가 지불할 의향 비율이 동시에 높게 나타남

출처 : 데일리안(2023.02.15)

**‘여행도 양극화’ 여행업계,
가성비·가심비 투트랙 본격화**

여행업계가 ‘가심비(가격 대비 심리적 만족도)’와 ‘가성비(가격 대비 성능)’를 모두 잡는 투트랙 전략을 본격화한다. 고물가·고금리 기조가 이어지면서 상대적으로 저렴한 가성비 여행상품을 찾는 고객들도 있지만 그동안 얹눌렸던 소비 심리가 폭발하면서 여행 비용을 아끼지 않는 여행객들 까지 늘어나고 있어서다.

가심비 + 가성비를 동시에 잡는 전략 필요

1. 주제 선정 - (2) 분석 목표



관광 발전이
필요한 지역 파악
관광이 활성화되지
않은 지역 찾기



관광 활성화
잠재력이
높은 지역 파악
실제 방문객들이
만족한 지역 찾기



가성비 + 가심비
가성비와 가심비를
모두 만족 할 수 있는
지역 찾기

잠재적 관광 활성화 지역 선정

→ 맞춤형 전략 수립을 통한 관광 불균형 해소

2. 활용 데이터 및 분석 과정

공간군집분석



관광 활성화 지수(2023)

데이터 전처리

- 일부 하위 행정구 통합
- 각 지역 별 방문객 비율을 이용한 가중 평균

공간군집분석

Local Moran's I 이용
공간자기상관성 분석

콜드스팟 & 핫스팟 도출

데이터 전처리



국민여행조사(2023)

- 만족도 결측치 삭제
- '방문지역' 변수 생성

분류분석

데이터 전처리

- 파생변수 생성
- 1인 지출 비용 총합
 - 여행 일 수
 - 1인 1일 지출 비용
 - 지역별 비용&만족도 가중평균

분류분석

비용 및 만족도 사분면
→ 지역 분류

공간군집분석과 결합

잠재적 관광 활성화 지역 선정

EDA

주된 활동 & 연령대를 고려한
맞춤형 관광 전략 제시

회귀분석

데이터 전처리

- ① 잠재적 관광 활성화 지역의 데이터만 추출
- ② 명목형 독립변수
차원축소, 원-핫 인코딩

회귀분석

단계적 변수 선택법 이용
독립변수 선택

종속변수 : 만족도

OLS 회귀분석 실시



만족도에 영향을 끼치는
요인 파악 (긍정적/부정적)

→ 관광 정책, 전략 제안

3. 데이터 분석 - (1) 데이터 전처리



2023 국민여행조사 데이터

만족도 결측치 제거

: 만족도(A10)가 존재하지 않는 데이터 제거



전처리 전

`korea_travel_2023_국내.shape`

(52111, 2368)

↓
전처리 후

`df_k.shape`

(24282, 2368)

총 27,829개
행 삭제

방문지역 변수 생성

: 두 지역 이상 방문한 데이터 삭제



`D_TRA_1~6_1~17_SPOT` : 여행 1차~6차 1~17번째 방문지역

- ① 각 행의 지역코드 최빈값을 대입한 'spot' 변수 생성
- ② 한 지역만 여행한 데이터 추출 → 4,092개 행 삭제
- ③ 지역코드에 한글지역명 매핑한 'ko_spot' 변수 생성

최종 '방문 지역' 변수 : `ko_spot`

spot
37330
32010
31220
32410

ko_spot
경상북도 청송군
강원도 춘천시
경기도 안성시
강원도 양양군

3. 데이터 분석 - (2) 공간군집분석



2023 관광 활성화지수 데이터 전처리

관광 활성화 지수란? 관광빅데이터(신용카드, 이동통신, 내비게이션) 활용, 시군구별 방문객의 관광관련 활동을 통해 관광산업이 얼마나 원활하게 유지되고 성과로 이어지는지 보여주는 지수

시군구 통합

: 국민여행조사 지역 기준과 통일하기 위해 일부 하위 행정구를 기초지자체로 통합

X

ex) 경기도 안산시

한국관광 데이터랩

안산시 단원구

안산시 상록구

코드북

경기도

090

안산시

→ 안산시 단원구, 안산시 상록구를 안산시로 통합

하위 행정구 별 방문객 수 비율을 이용한 가중평균

시	구	방문자수	비율	활성화지수
안산시	상록구	28,026,250	35%	37.3
안산시	단원구	51,764,290	65%	77.1
안산시		79,790,540		63.1

안산시 활성화 지수

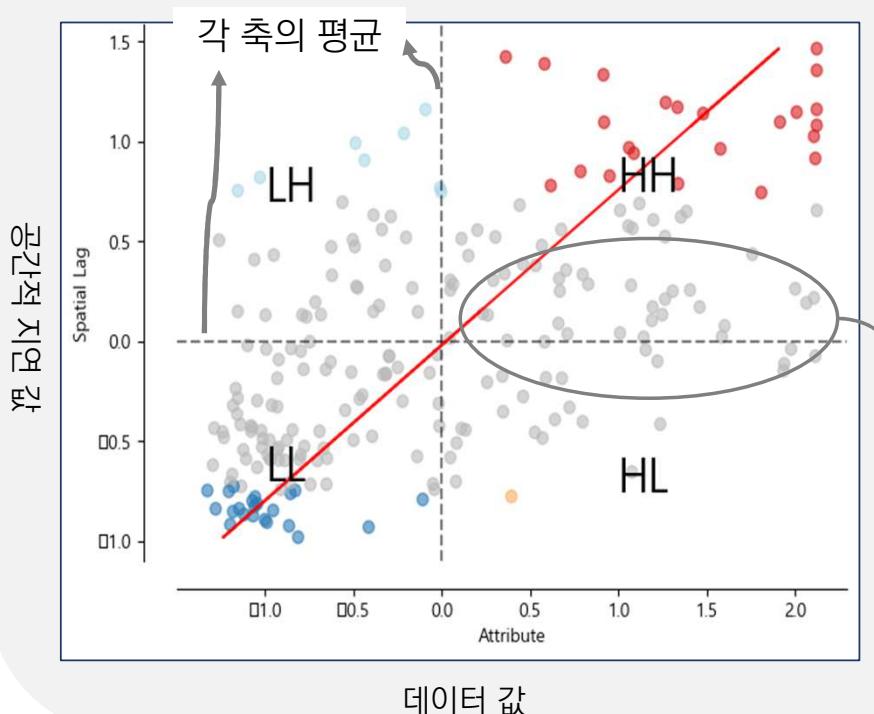
$$\begin{aligned} &= 37.3 * 0.35 + 77.1 * 0.65 \\ &= 63.1 (\%) \end{aligned}$$

3. 데이터 분석 - (2) 공간군집분석

국지적 모란 지수(Local Moran's I) 분석이란?

- 전체 공간에서 특정 지역과 그 주변 지역을 비교하여 현저하게 차이가 나는 지역을 찾아내는 방법
- 통계량이 **높은 양의 값**이 나올 경우 **지역 간 유사성**이 있음
- 모란 산점도를 통해 공간적 연관성을 네 가지로 분류

[모란 산점도]



- 회귀선 : X축(데이터 값)과 Y축(공간적 지연 값)간의 선형 관계
- 기울기 : 공간적 자기 상관의 정도 표현

HH : **높은 값** 주변의 **높은 값** → 핫스팟
LL : **낮은 값** 주변의 **낮은 값** → 콜드스팟

p-value > 0.05 이상인 지역 : 통계적 유의성을 만족하지 않음

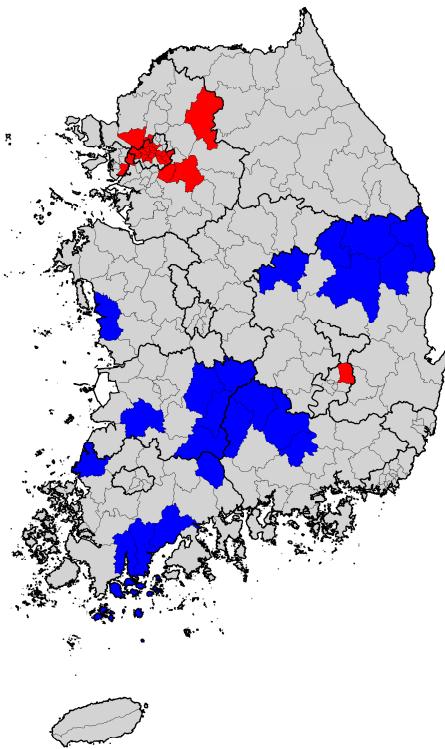
▶ KNN(K-Nearest Neighbors) 방식으로 공간 가중치 부여

```
wq = lps.weights.KNN.from_dataframe(df_map, k=3)
wq.transform = 'r'
```

3. 데이터 분석 - (2) 공간군집분석

활성화지수 콜드스팟&핫스팟

- Hot Spot
- Cold Spot



```
y = df_map['활성화지수']
mo =esda.moran.Moran_Local(y, wq)

sig = 1 * (mo.p_sim < 0.05)
hotspot = 1 * (sig * mo.q==1) # 핫스팟(높은 값의 클러스터)
coldspot = 3 * (sig * mo.q==3) # 콜드스팟(낮은 값의 클러스터)
```

유의수준 5% 하에서 통계적 유의성을 만족하는 지역 수 : 54개

핫스팟	서울	종로, 중구, 용산, 성동, 광진, 성북, 은평, 서대문, 마포, 양천, 강서, 영등포, 동작, 강남, 송파, 강동
	대구	동구
	인천	남동구
	경기	부천, 광주, 가평, 고양, 성남
콜드스팟	충남	보령
	전북	정읍, 남원, 진안, 무주, 장수
	전남	구례, 보성, 장흥, 강진, 영광, 완도
	경북	안동, 영주, 문경, 영양, 봉화, 울진
	경남	의령, 함양, 거창, 합천

→ 관광 활성화가 필요

3. 데이터 분석 - (3) 분류분석

2023 국민여행조사 데이터

분류분석을 위한 추가적인 전처리

1인 총합 지출 비용 변수 생성



사용한 비용 변수

변수명	설명
NA9C	1인 숙박비 총액
NA9D	1인 음식점비 총액
NA9E	1인 식음료비 총액
NA9G	1인 여행활동비 총액
NA9H	1인 쇼핑비 총액

1인 총합 지출 비용 = 비용의 총합

NA9H	NA9C	NA9E	NA9D	NA9G	cost
100000	0	15000	0	0	115000
0	0	0	15000	15000	30000
150000	0	0	0	0	150000
0	50000	0	50000	0	100000
0	0	0	0	0	0



‘여행 일 수’ 변수 생성



D_TRA1~6_S_Day: 여행 1차~6차 숙박일수

여행 일 수 = 숙박 일 수 + 1

D_TRA1_S_Day	D_TRA2_S_Day	D_TRA3_S_Day	D_TRA4_S_Day	D_TRA5_S_Day	D_TRA6_S_Day
2	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	0

tour_count
3
1
1
2
1

3. 데이터 분석 - (3) 분류분석

지출 비용 및 만족도에 따른 지역 분류



지역별 1인 1일 총합 지출 비용

① IQR방식으로 지역별 비용 이상치 제거

```
def outlier(df, col):
    Q1 = df[col].quantile(0.25) # 1분위 수
    Q3 = df[col].quantile(0.75) # 3분위 수
    IQR = Q3 - Q1             # iqr = 3분위 수 - 1분위 수

    # 최댓값 : 3분위 수 + 1.5 * iqr
    # 최솟값 : 1분위 수 - 1.5 * iqr
    # 최솟값 미만 or 최댓값 초과 -> 이상치로 판단
    idx = ((Q1 - 1.5*IQR) <= df[col]) & ((Q3 + 1.5*IQR) >= df[col])
    df = df.loc[idx]
    return df
```

② 1인 1일 지출 비용 (= 1인 지출 비용 / 총 여행 일 수) 변수 생성

```
cst_e['cost_day'] = cst_e['cost'] / cst_e['tour_count']
```

③ 지역별 비용의 가중 평균 도출

가중 평균 계산 코드 일부

```
cst_2 = cst_e.groupby('spot')[['cost_day']].sum().reset_index()
cst_3 = pd.merge(cst_e, cst_2,
                  on = 'spot', how = 'left')
cst_3['cost_rate_cost_day'] = cst_3['cost_day_x'] / cst_3['cost_day_y'] # 가중치
cst_3['cost_day_mean_w'] = cst_3['cost_rate_cost_day'] * cst_3['cost_day_x']
cost = cst_3.groupby(['spot', 'ko_spot',
                      'count'])['cost_day_mean_w'].sum().reset_index()
```

지역명	데이터 수	비용 가중 평균	
		spot	ko_spot count
0 11010	서울특별시 종로구	90	66533.84
1 11020	서울특별시 종구	50	88812.20
2 11030	서울특별시 용산구	21	46982.12
3 11040	서울특별시 성동구	27	58320.49
4 11050	서울특별시 광진구	26	36327.15
...

3. 데이터 분석 - (3) 분류분석

지출 비용 및 만족도에 따른 지역 분류



지역별 만족도

· 지역별 만족도 가중 평균 도출

※ 대부분의 응답이 4점, 5점
→ 이상치가 존재하지 않음

A10	count
1	0
2	24
3	611
4	15102
5	4453

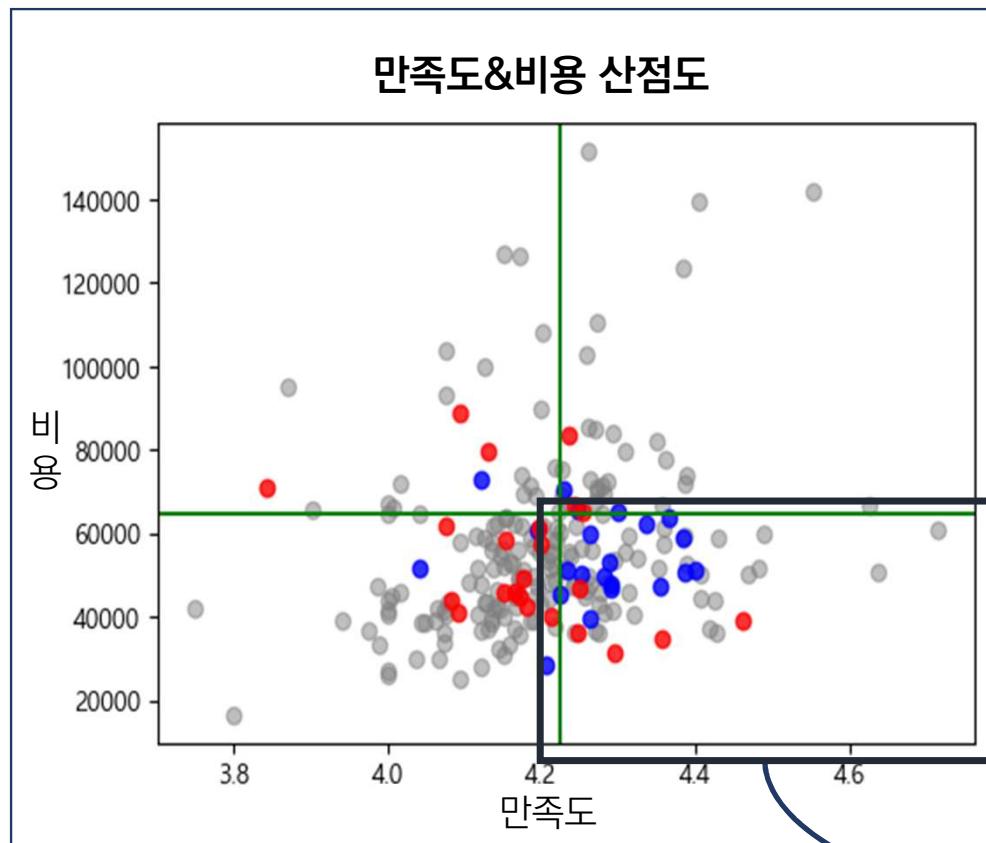
가중 평균 계산 코드 일부

```
sat_2 = sat_1.groupby('spot')[['A10']].sum().reset_index()
sat_3 = pd.merge(sat_1, sat_2,
                  on = 'spot', how = 'left')
sat_3['sat_rate_A10'] = sat_3['A10_x'] / sat_3['A10_y'] # 가중치
sat_3['A10_mean_w'] = sat_3['sat_rate_A10'] * sat_3['A10_x']
satisfaction = sat_3.groupby(['spot','ko_spot',
                               'count'])['A10_mean_w'].sum().reset_index()
```

spot	ko_spot	count	지역명	데이터 수	만족도 가중 평균
			A10_mean_w		
0 11010	서울특별시 종로구	90	4.24		
1 11020	서울특별시 중구	50	4.09		
2 11030	서울특별시 용산구	21	4.25		
3 11040	서울특별시 성동구	27	4.15		
4 11050	서울특별시 광진구	26	4.25		
...

3. 데이터 분석 - (3) 분류분석

지출 비용 및 만족도에 따른 지역 분류



① 만족도(X축)

대부분의 응답이 ‘만족(4)’, ‘매우 만족(5)’으로 구성
→ 지역 간 편차가 크게 나타나지 않음

평균을 기준으로 고만족 vs 저만족 분류

② 비용(Y축)

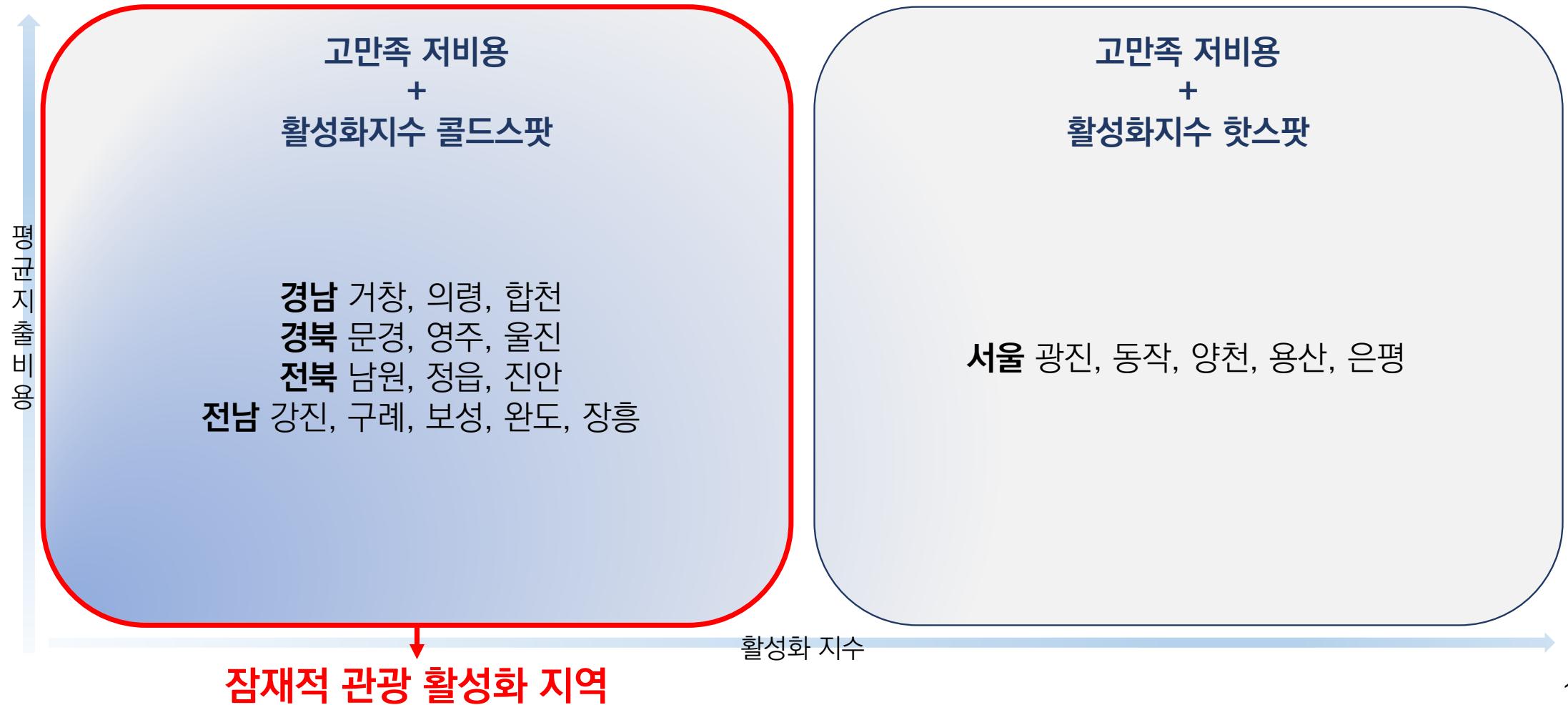
높은 일부 지역을 제외하고 대부분 지역이
20,000원 ~ 80,000원 구간에 분포

3분위 수(상위 25%) 기준으로 고비용 vs 저비용 분류

고만족 저비용 지역 대상으로 지역을 두 그룹으로 분류

3. 데이터 분석 - (3) 분류분석

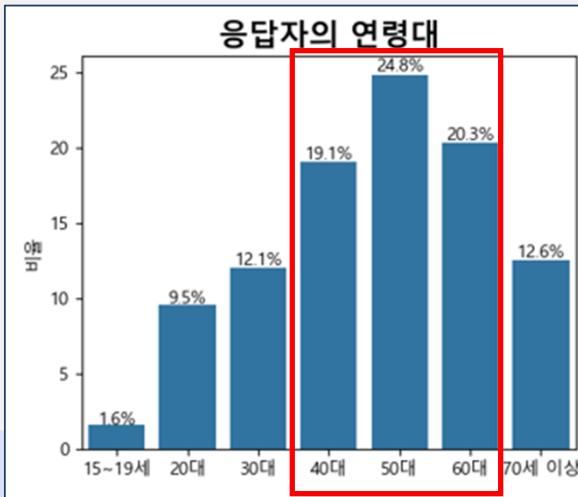
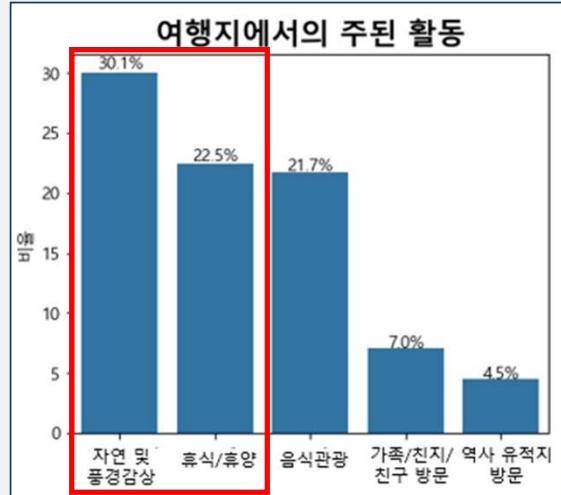
지역 분류



3. 데이터 분석 - (3) 분류분석



EDA



‘자연경관 감상’, ‘휴식’, ‘휴양’ 등
‘힐링’ 관련 활동 선호

40대~60대 관광객의
비중이 높음

맞춤형 테마 관광
구축 필요

3. 데이터 분석 - (4) 회귀분석

경남 거창, 의령, 합천

경북 문경, 영주

전북 정읍, 진안

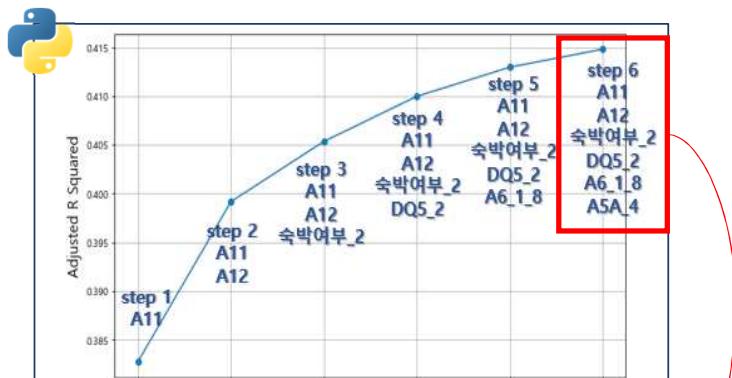
전남 강진, 보성, 장흥

잠재적 관광 활성화 지역 대상으로 **다중 회귀분석** 실시
→ 만족도에 영향을 미치는 요인 분석

단계적 변수 선택법을 통한 독립변수 결정

단계적 변수 선택법이란?

모든 독립 변수를 한 번에 투입하는 대신, 통계적으로 유의미한 변수들을 단계적으로 선택하고 유의미하지 않은 변수는 제거



설명력이 가장 높을 때의 변수들 선택

OLS 회귀분석 실시

```
import statsmodels.api as sm
X = df_reg[selected_variables]
X_ad = sm.add_constant(X)
y = df_reg['A10']

ols = sm.OLS(y,X_ad).fit()
print(ols.summary()) # 회귀분석 결과
```

3. 데이터 분석 - (4) 회귀분석

회귀분석 결과

		Adj. R-squared : 0.415					
		② coef	std err	t	③ P > t	[0.025	0.975]
const		1.5628	0.096	16.254	0.000	1.374	1.751
A11		0.4982	0.023	21.809	0.000	0.453	0.543
A12		0.1607	0.024	6.746	0.000	0.114	0.207
국내_숙박 여부_2		-0.0838	0.022	-3.738	0.000	-0.128	-0.040
DQ5_2		0.0971	0.027	3.555	0.000	0.044	0.151
A6_1_8		-0.0708	0.026	-2.685	0.007	-0.123	-0.019
A5A_4		0.1394	0.060	2.320	0.020	0.022	0.257

- ① R-squared : 회귀모형의 설명력
→ 6개의 독립변수들이 종속변수의 변동을 약 41.7% 설명함
- ② coef : 회귀 계수
→ 양(+)의 값이면 긍정적인 영향, 음(-)의 값이면 부정적인 영향
- ③ P > |t| : 유의확률
→ 모든 변수들이 0.05 미만으로 통계적 유의성 만족

4. 정책 제안(1)

웰니스 관광이란?

건강 유지 및 증진을 목적으로 자연과 휴식을 중심으로 하는 관광 트렌드 (Mueller&Kaufmann, 2001)

웰니스 관광 테마



뷰티 · 스파



힐링 · 명상



자연 · 숲치유



한방



푸드



스테이

주 방문 연령층(40~60대)이 선호할만한 한방 테마 활성화

4. 정책 제안(1)

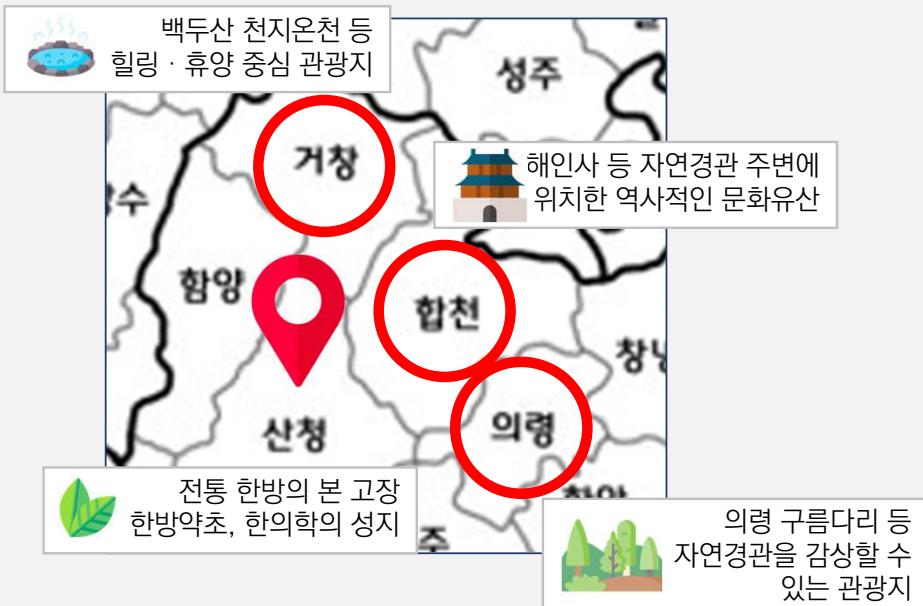
한방 중점 관광 클러스터 구축

① 경상북도 문경, 영주, 울진



한방을 중점으로 한
경북지역 관광 클러스터 구축

② 경상남도 거창, 의령, 합천 + 경상남도 산청



경상남도 산청과 연계하여
힐링·자연 + 한방을 중점으로 한
경남지역 관광 클러스터 구축

4. 정책 제안(1)

한방 중점 웰니스 관광 코스 구축

③ 전라남도 강진, 보성, 장흥, 완도



④ 전라북도 진안, 정읍, 남원, 전라남도 구례



각 지역의 한방 웰니스 프로그램을 하루씩 체험하는
한방 체험 코스

4. 정책 제안(2)

긍정적인 요소 강화하기



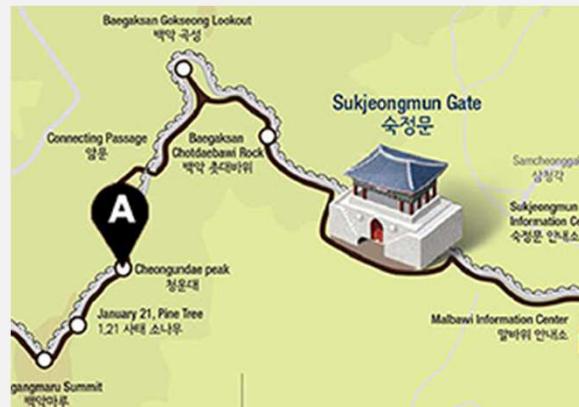
역사 관광 투어 활성화

주 활동으로 역사 유적지를 방문한 경우 만족도 ↑



주요 역사 유적지를 연결하는
탐방 코스 개발

사례) 서울 한양 도성 스탬프투어



출처 : 서울 한양 도성

- 여러 역사 유적지를 경유하는 도보 코스
- 특정 장소에서 스탬프를 발급
- 완주 시 인증서 발급

4. 정책 제안(2)

긍정적인 요소 강화하기



재방문 관광객에게 인센티브 제공

재방문 의향, 타인 추천 의향이 높을 수록 만족도 ↑



관광객의 관심도를 파악한
인센티브 제공

과거 여행 이력을 고려한 맞춤형 할인 쿠폰 제공

4. 정책 제안(2)

부정적인 요소 보완하기



관광지 정보 제공 강화

여행에 대한 사전 정보 없이 방문한 경우 만족도 ↓



지역 홍보 계정을 통해 **사전에 쉽게 접근** 가능한
다양한 **미디어 콘텐츠** 제작



현장에서 쉽게 정보를 얻을 수 있게
관광 안내 센터 활성화

4. 정책 제안(2)

부정적인 요소 보완하기



숙박 여행 장려

당일 여행인 경우 만족도 ↓



역사 유적지, 자연 휴양림 등
숙박옵션 + 다양한 프로그램 결합



숙박시설의 질 개선 및
주변 인프라, 교통 개선

4. 기대효과

1

각 지역의 특색 있는 한방 체험 프로그램을 통한 지역 브랜드 가치와 관광지로서의 인지도 상승

2

자연과의 조화를 강조한 한방 웰니스 프로그램을 통한 지속 가능한 관광 개발에 기여

3

잠재적 관광 활성화 지역이 웰니스 관광지로 부상함으로써 지역간 균형 발전 촉진

4

관광객 만족도 향상 및 지속적인 수요 창출을 통한 지역 관광 수익 증가

5

관광 산업에서의 소비 양극화 완화

Q&A