



대한민국 테마 여행 10선 문제점 개선을 통한 지역관광 활성화 전략

DATA 129

목차

1. 정책소개 및 주제도출

- 대한민국 테마 10선 여행 소개
- 대한민국 테마 10선 여행의 필요성
- 대한민국 테마 10선 여행의 문제점 및 주제도출

2. 데이터 분석

- 원데이터 병합과정
- 데이터 전처리
- 전체적인 분석방향
- 백제문화권역 문제점 분석: 권역특성, 다중대응 분석
- 남도바닷길권역 문제점 분석: 권역특성, web crawling, 계층군집 분석, k-means clustering, 의사결정나무, 다중선형회귀분석

3. 문제해결 전략과 활용

- 백제문화 권역 마케팅
- 남도 바닷길 권역 마케팅
- 분석결과의 활용방안



대한민국 테마여행 10선 - 소개

‘대한민국 테마여행 10선’육성은 우리나라 지역관광의 수준을 높이기 위해 3~4개의 지방자치단체 (이하 지자체)를 하나의 관광권역으로 묶어, 이를 집중적으로 발전시키는 5개년 프로젝트이다.



• 정책 목표 “만족도 높은 분산형·체류형 선진 관광지 육성”

- 만족도 높은 내, 외국인이 다시 찾는
- 분산형 여행공간을 복수의 지자체로 연계 및 확장
- 체류형 관광객들의 체류시간 증대

• 추진 방향

‘점’ 단위의 관광명소/콘텐츠 개발	→ 개별 자원의 ‘선형(관광객동선)’ 연계
관광지 위주의 단순 소개	→ 지역의 스토리 등 ‘인문자원’ 융합
공모사업별 분절지원	→ 볼거리/즐길거리/이동망 등 종합 지원
개별 기초지자체 대상 균분 지원	→ ‘유망 권역(3~4개 지역)’ 대상 집중 지원

대한민국 테마여행 10선 - 필요성

관광수요 환경

국민 국내관광

- 양적 성장 둔화, 수도권 점유율 증대

외국인 국내관광

- 수도권 및 제주도 집중 현상, 저가 여행상품으로 인한 한국관광 품질 저하 문제 발생

국민 해외관광

- 양적 성장 가속화

지역관광 환경

지역의 경제·사회 활성화 수단으로 관광을 적극적으로 활용하는 정책 증가



지역관광의 질적 성장을 견인할 수 있는 정책 패러다임 변화 필요



“대한민국 테마여행 10선은 관광수요 환경의 과제와 지역관광 환경의 요구를 동시에 해결하여 지역관광의 질적 성장을 이루어 낼 수 있음”

대한민국 테마여행 10선 - 문제점 및 주제도출

테마여행 10선의 10개 권역 중, 모범 사례인 평창로드와 비교해 보았을 때 권역 내 도시 별 방문객 편차가 심하고, 평균 숙박일수가 짧아 정책 목표인 "분산형, 체류형 관광지 육성"을 실현하는 데 문제가 있는 권역이 존재한다.

"이를 위해 권역 내 도시간 방문객 편차가 가장 심한 '남도바닷길'과 숙박일수가 짧은 '백제문화' 권역을 선정해 각각의 문제를 해결하는 것을 프로젝트 주제로 한다"

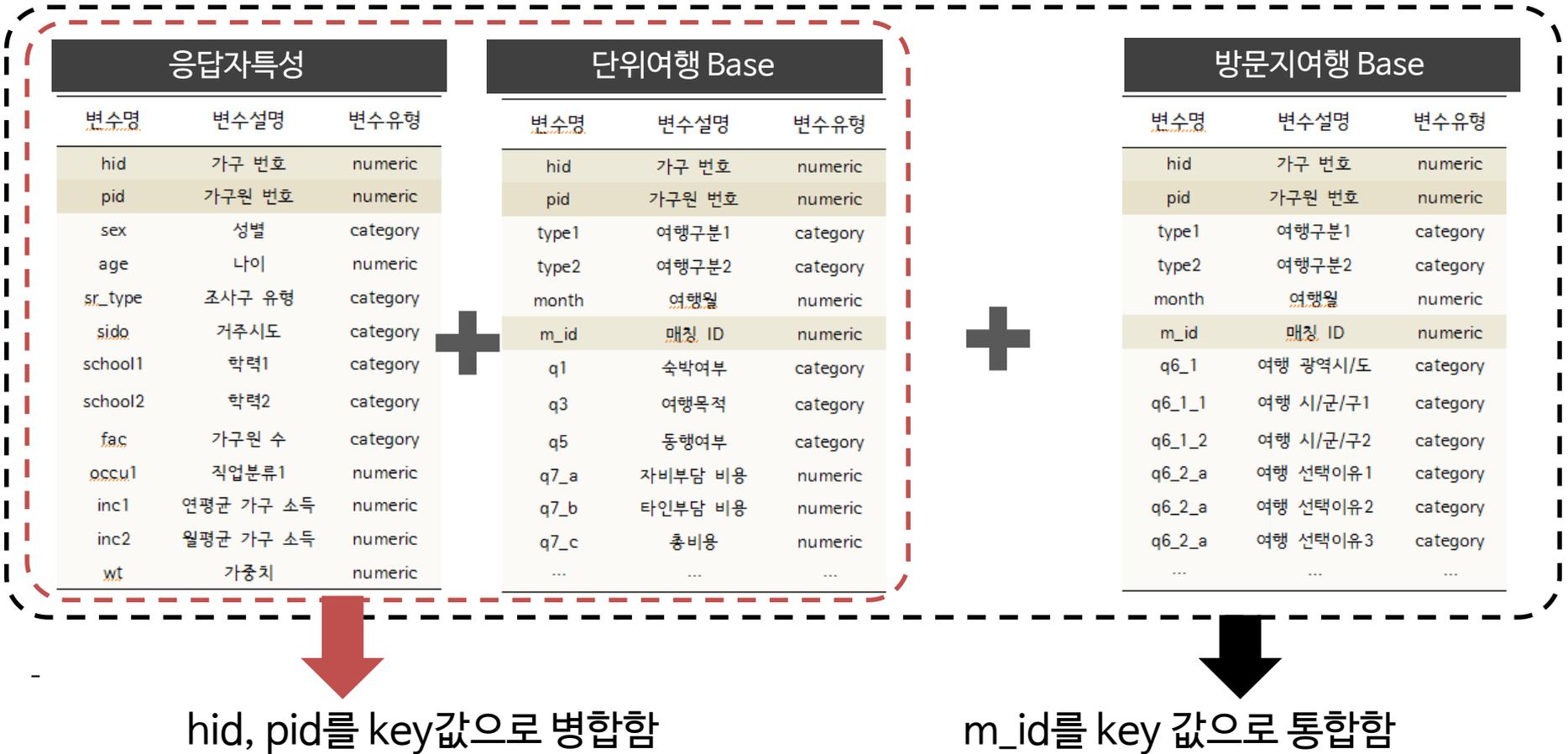


	도시간 방문객 편차	숙박일수
백제문화	적음	짧음
남도바닷길	큼	보통
평창로드	적음	깊

- 도시간 방문객 편차는 주요 관광지점 입장객 통계를 이용함
- 권역별로 관광지 개수가 다르므로 감안하여 판단할 필요 있음
- "숙박일수=체류기간" 이라고 볼 수 있음
- 인천경기는 관광객편차 문제가 공존하여 정확한 문제분석을 위해 백제문화권역을 선정함

원 데이터 병합과정

- m_id는 매칭id로 단위 베이스와 방문지 베이스 데이터를 매칭할 수 있는 고유번호임
- 단위여행 베이스와 방문지 베이스 데이터를 융합할 때는 pid와 m_id를 매칭하고 개인응답자특성과 매칭할 때는 hid와 pid를 사용함
- 개인화 데이터와 개인여행 데이터 두 데이터셋에 대해 동일한 방법으로 데이터를 융합함



데이터 전처리

분석목적에 적합한 방식으로 데이터를 정제함

불필요한 데이터 삭제

통합된 38275개의 레코드중 type1 항목을 기준으로 1(국내 관광여행), 2(기타여행) 이외의 응답을 한 4696개의 여행데이터를 삭제함

변수명	변수 설명	항목
Type1	여행구분	1. 국내관광여행
		2. 해외여행
		3. 개인 국내 관광여행/해외관광여행 없음
		4. 기타여행
		5. 개인 기타 여행 없음

새로운 변수 정의

q6_1, q6_1_1로 부터 region이라는 변수를 새롭게 정의하였고, 이로부터 남도 바닷길테마와 백제문화권 테마의 데이터를 추출함

변수명	변수 설명	변수명	변수 설명	값
q6_1	여행 광역시/도	region	권역	1. 남도바닷길
q6_1_1	여행 시/군/구			2. 백제문화권

변수 변환

- 항목별 만족도 q12_1~q12_12 중 결측치가 75% 이상인 q12_7~q12_12 번 항목을 삭제함
 - 여행의 특성상 매우 만족(5점) 미만의 점수는 사실상 만족으로 보기 어려움. 또한 항목별 만족도 항목 간의 비교가 중요하다고 판단하여 0~4, 9(무응답)→'0', 5→'1' 로 분류하여 binary value 로 변환함
- 연령층을 청년, 장년, 중년, 노년의 4계층 으로 구분함. 각 계층의 구분은 삶의 패턴이 달라질 것으로 예상되는 나이를 기준으로 함
(-청년층: 결혼 이전 -장년층: 결혼후 자녀 독립이전 -중년층: 자녀의 독립이후 -노년층: 은퇴이후)

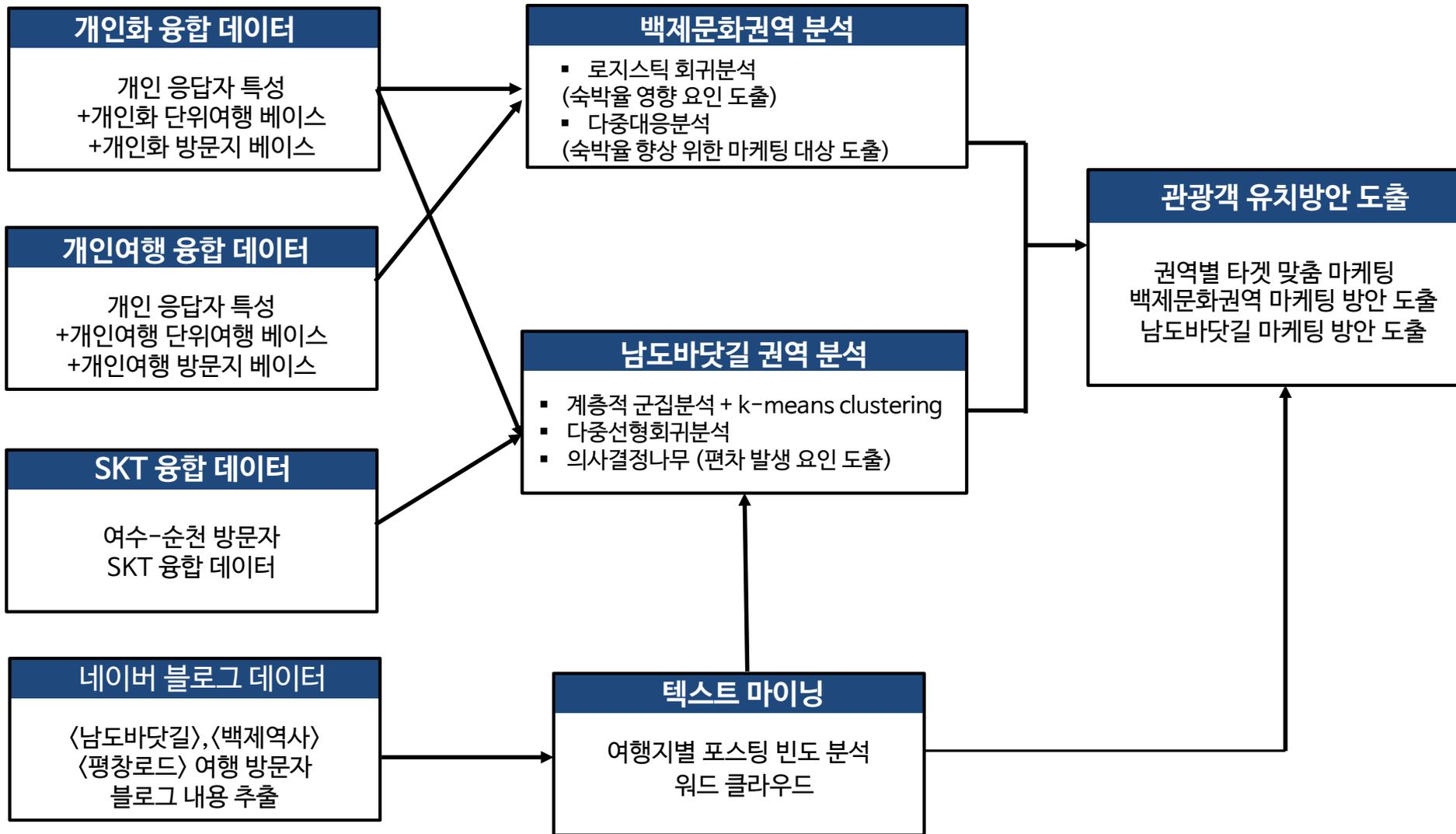
1.

변수명	변수 설명	값
q12_1~q12_6	항목별 만족도	1. (1~4,9) → 0
		2. (5) → 1

2.

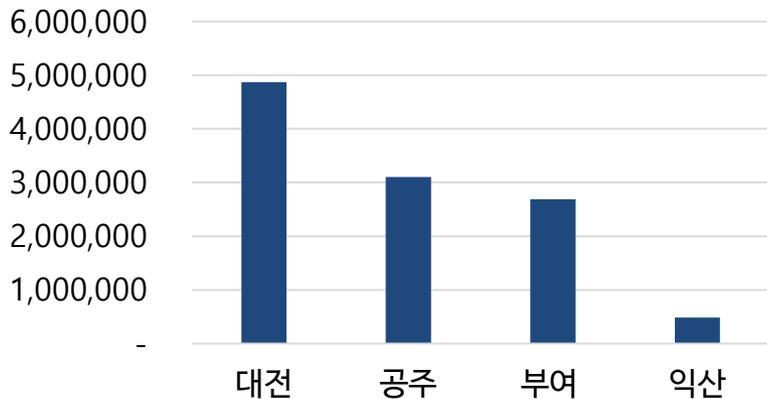
변수명	변수 설명	값
Age	나이	1. (15~32) → 청년층
		2. (33~52) → 장년층
		3. (53~65) → 중년층
		4. (65~) → 노년층

전체적인 분석 방향

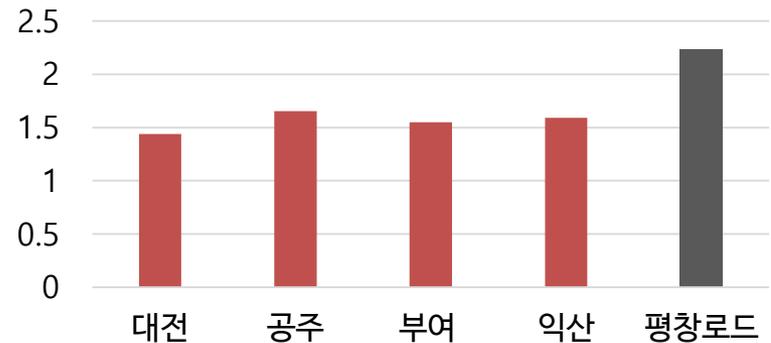


백제문화 : 권역 특성 분석

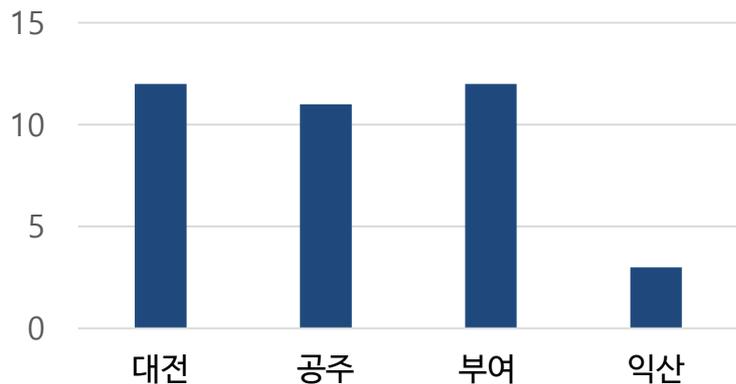
도시별 관광객 수



평균 숙박 일수



관광지 개수



'대한민국 테마여행 10선' 백제문화 코스



자료: 문화체육관광부

백제문화 : 짧은 체류기간 문제

짧은 체류기간은 저조한 숙박율에 기인한다

분석단계

1. “외래관광객 체류기간의 결정요인” (2010, 박상곤 외) : 외래관광객의 여행목적이 숙박율에 영향을 미침을 확인
2. 국내관광객의 여행목적도 숙박율에 영향을 미친다는 것을 로지스틱 회귀분석으로 재확인
3. 대전/공주/부여/익산 방문자들의 방문 목적을 평창권역 및 전체 관광객과 비교하여 숙박율이 낮은 이유 도출
4. 다중대응분석을 통해 세대로 관광목적, 숙박여부, 동행여부 등 여행패턴의 차이를 규명해 마케팅 타겟 도출

데이터 전처리

데이터 융합

: 개인응답자특성+개인화 단위여행+개인화 방문지

변수 추출 후 로지스틱 회귀분석에 적합하게 변수 변환

- R에서 nominal 변수를 interval 변수로 혼동하지 않도록 자료를 변환함
- 여행월보다 여행계절이 통계적으로 유의미하므로 분류하여 변환함
- tado : 거주시도와 여행시도가 다를 경우 1, 아니면 0 값을 부과함
- relation : 연계여행여부, 방문지 베이스에 매칭ID 하나 당 여러 개의 여행이 존재하거나 q6_1_2가 NA가 아니라면 1, otherwise 0 값을 부과함

변수명	변수설명	변수유형	변수명	변수설명	변수유형
pid	가구원 번호	nominal	pid	가구원 번호	nominal
m_id	매칭 ID	nominal	m_id	매칭 ID	nominal
month	여행 월	nominal	season	여행계절	categorical
q1	숙박여부	nominal	stay	숙박여부	categorical
q3	여행목적	nominal	purpose	여행목적	categorical
q5	일행여부	binary	with	일행여부	categorical
q6_1	여행한광역시	nominal	q6_1	여행한광역시	nominal
q6_1_1	여행한시군1	nominal	q6_1_1	여행한시군1	nominal
q6_1_2	여행한시군2	nominal	q6_1_2	여행한시군2	nominal
sex	성별	binary	sex	성별	binary
age	나이	interval	gen	세대	categorical
sido	시도	nominal	tado	타도여행여부	binary
inc2	월평균 가구소득	numerical	relation	연계여행여부	binary
			inc2	월평균 가구소득	numerical

백제문화 : 짧은 체류기간 문제

변수명	변수설명	회귀계수(표준편차)	p-value
intercept		-1.0005(0.7901)	<2e-16 ***
with=1	일행여부	0.415(0.0496)	<2e-16 ***
tado1=1	타도여행여부	1.2107(0.0357)	<2e-16 ***
purpose=business	여행목적=사업	-0.4707(0.1169)	5.71e-05***
purpose=ect	여행목적=기타	-1.0251(0.1756)	5.26e-09***
purpose=edu	여행목적=교육	0.2043(0.1023)	0.0472*
purpose=health	여행목적=건강치료	0.0940(0.1171)	0.4226
purpose=kins	여행목적=친지방문	-0.3676(0.0312)	<2e-16 ***
purpose=religion	여행목적=종교	-0.7368(0.1359)	5.91e-08***
purpose=shopping	여행목적=쇼핑	-3.2811(0.2641)	<2e-16 ***
season=spring	계절=봄	-0.2485(0.0382)	7.73e-11***
season=summer	계절=여름	0.4798(0.0400)	<2e-16 ***
season=winter	계절=겨울	0.4673(0.0396)	<2e-16 ***
relation	연계여행여부	1.7995(0.0802)	<2e-16 ***
age	나이	-0.0194(0.0008)	<2e-16 ***

종류	McFadden	R2ML	R2CU
pseudo R2 값	0.12470	1.56017	2.098693

로지스틱 회귀분석

- 타겟변수 : 숙박(숙박여행=1, 당일여행=0)
- 독립변수 : 여행목적(default=여가/위락/휴가)
- 통제변수 : 일행여부, 타도여행여부, 여행계절, 연계여행여부, 나이 (stepwise 방식으로 통제변수 추출)
- 회귀식
: stay ~ purpose + with + tado1 + season + relation + age

회귀분석 결과

- ✓ ‘여가/위락/휴가’ 목적의 여행보다 숙박율을 높이는 목적 변수는 없음
- ✓ ‘교육’, ‘건강치료’ 목적의 여행은 숙박율 증가요인이지만, 통계적으로 유의미하지 않음
- ✓ ‘쇼핑’, ‘종교’, ‘친지방문’, ‘사업’ 목적의 여행은 ‘여가/숙박/휴가’ 여행보다 숙박 가능성이 낮음

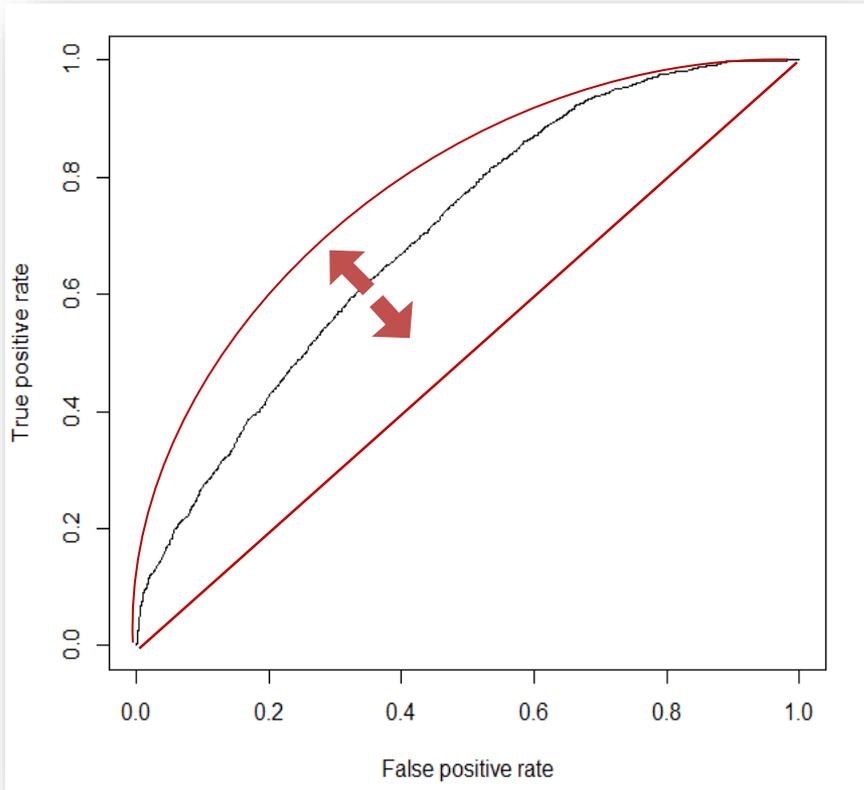
-Pseudo R Squared 값

: 회귀모형으로 예측할 수 있는 데이터셋 분산의 크기

-R2ML = R Squared Maximum Likelihood

: 약 15.6%의 데이터셋의 분산을 설명 가능

백제문화 : 짧은 체류기간 문제



모형의 평가

AUC = 0.7032248

- AUC는 ROC커브 아래 영역의 크기
- 일반적으로 AUC > 0.7 이면 유의미한 것으로 추정한다.
- 곡률이 클수록 예측 정확성이 크고 직선에 가까울수록 예측의 정확성이 작다

일반적인 AUC 평가기준

Excellent = 0.9~1.0

Good = 0.8~0.9

Fair = 0.7~0.8

Poor = 0.6~0.7

Fail = 0.5~0.6

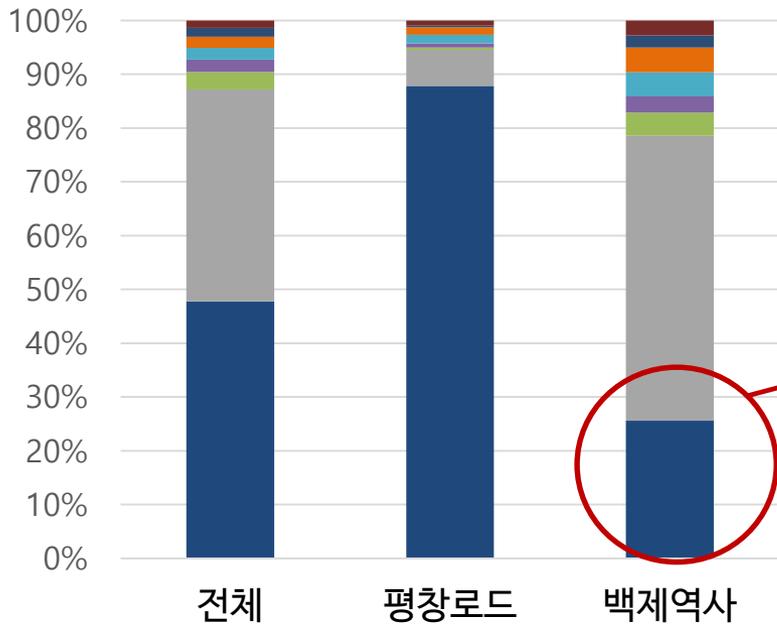
따라서 여행목적이 숙박율에 통계적으로 유의미한 영향을 미친다는 것을 알 수 있음

회귀분석을 통해 여행 목적은 숙박율에 영향을 미친다는 것을 증명하였으며, '여가/숙박/휴가' 여행일 경우에 숙박율이 가장 높다는 것을 알 수 있음

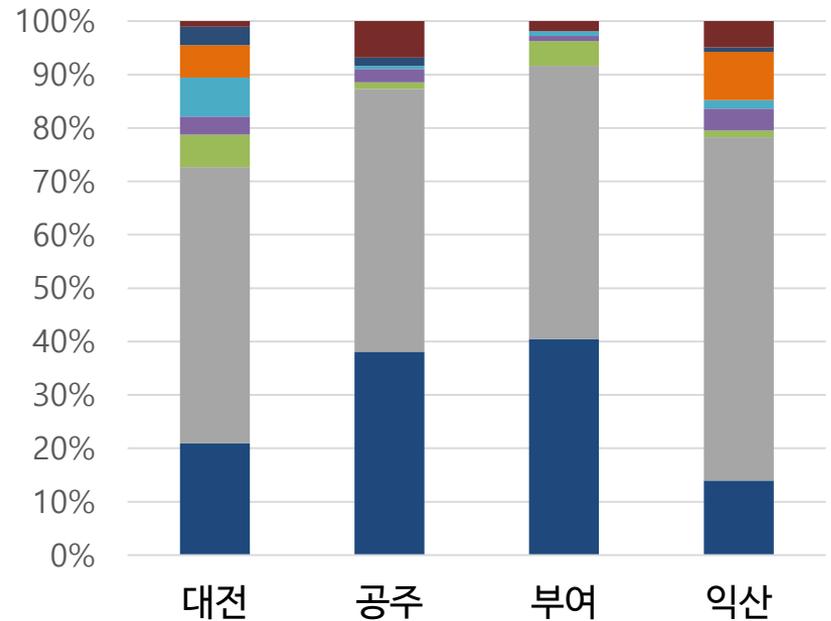
백제문화 : 짧은 체류기간 문제

- 여가/위락/휴가 ■ 친지방문 ■ 쇼핑
- 사업 및 전문활동 ■ 교육/훈련/연수 ■ 건강/치료
- 기타 ■ 종교/성지순례

목적별 방문자 비율



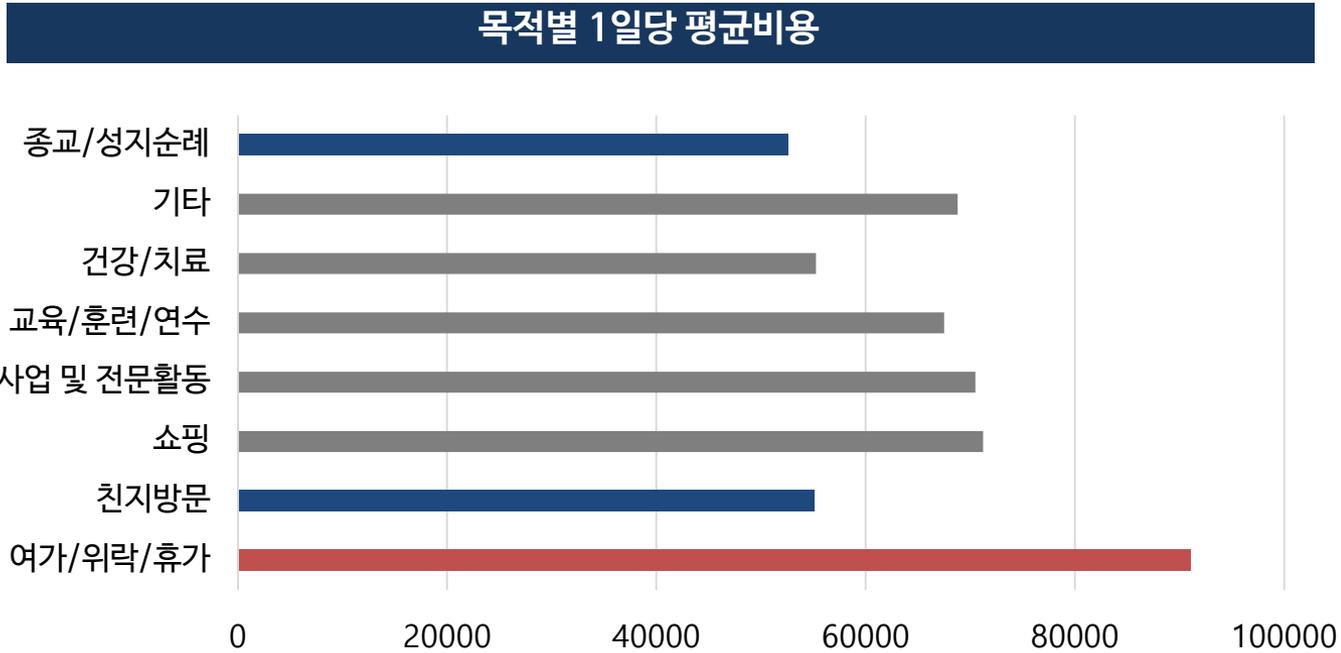
목적별 방문자 비율



백제역사권역 (대전/공주/부여/익산)의 친지방문 목적의 비율이 다른 지역에 비해 월등히 높은 반면, 여가 목적의 방문이 평창이나 전체 관광지 평균에 비해 현저히 낮다

앞선 로지스틱 회귀분석 결과를 적용할 때, 친지방문 목적은 여가/위락/휴가 목적보다 숙박율이 낮음을 알 수 있다. 따라서 백제문화권역의 숙박율이 낮은 이유를 알 수 있음

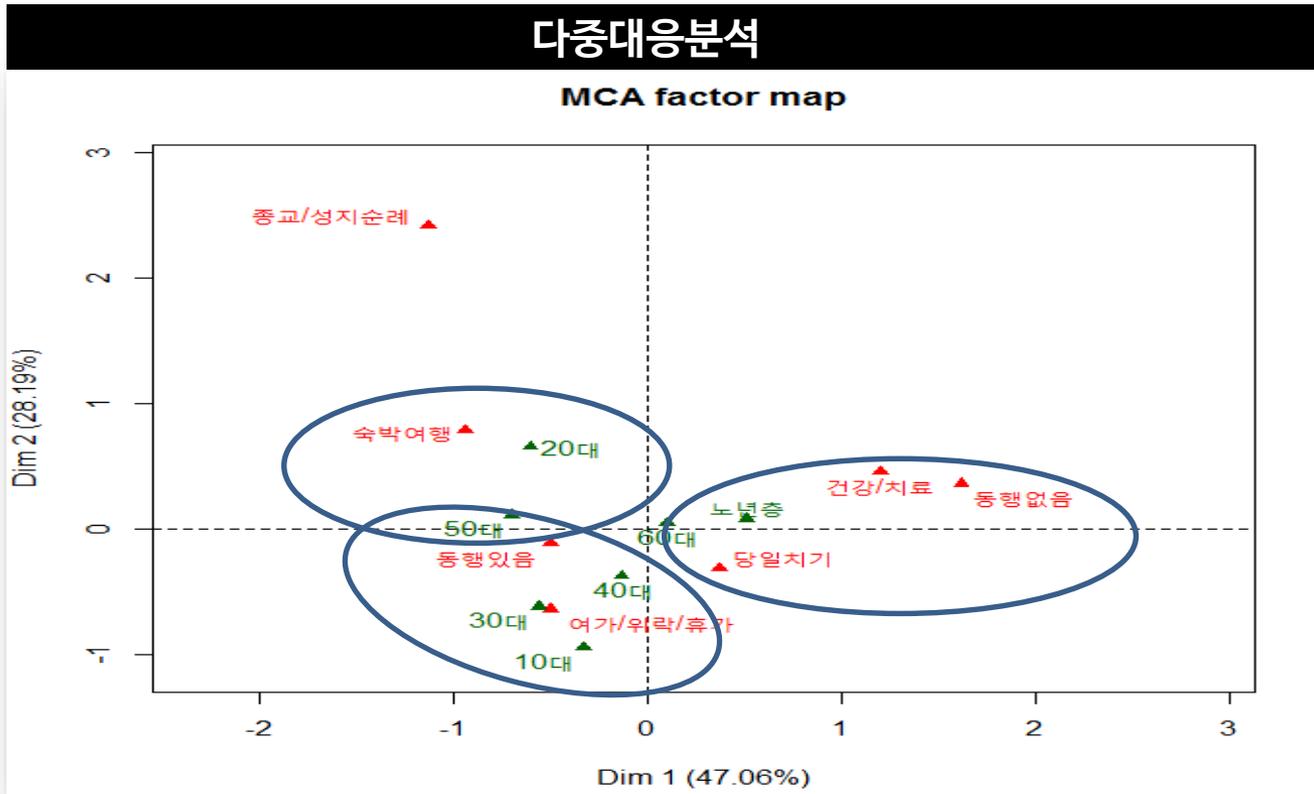
백제문화 : 짧은 체류기간 문제



- 여가/위락/휴가 목적의 여행이 1일당 평균 지출비용이 큼
- 친지방문, 건강/치료, 종교/성지순례 목적의 여행은 1일당 평균비용이 낮음

따라서 경제적 효과 또한 증대시킬 수 있는 여가/위락/휴가 목적의 여행을 늘릴 수 있는 방법을 찾아야 함

백제문화 : 짧은 체류기간 문제



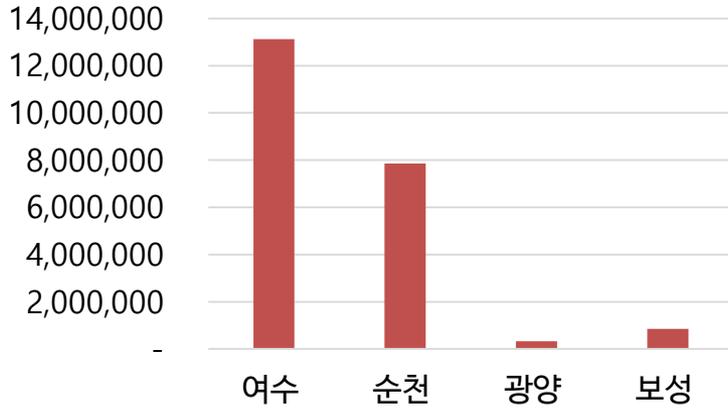
다중대응 분석 결과

- ✓ 60대와 노년층 : 당일여행, 동행없음, 건강치료 목적
- ✓ 20대와 50대 : 숙박여행
- ✓ 20대 제외한 10대부터 50대 : 여가/위락/휴가 목적, 동행 있음

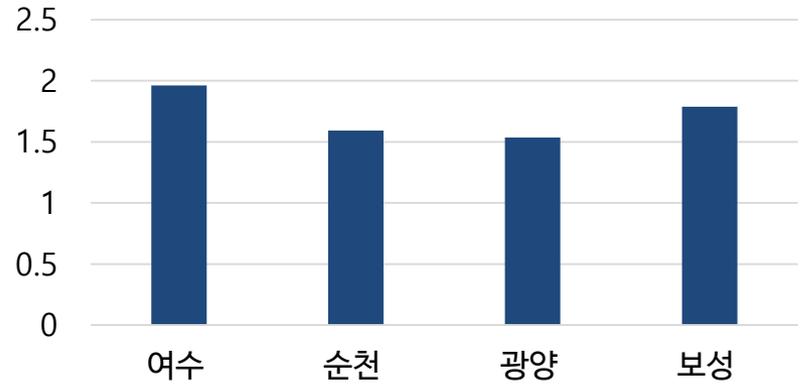
숙박율이 가장 높은 여가/위락/휴가 목적의 여행을 늘리기 위해서는 30~50대와 10대의 가족여행을 중심으로 한 상품 개발이 필요함

남도바닷길 : 권역 특성 분석

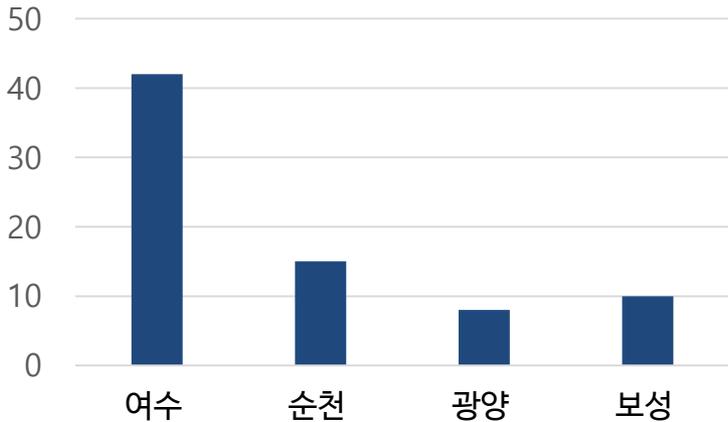
도시별 관광객 수



평균 숙박 일수

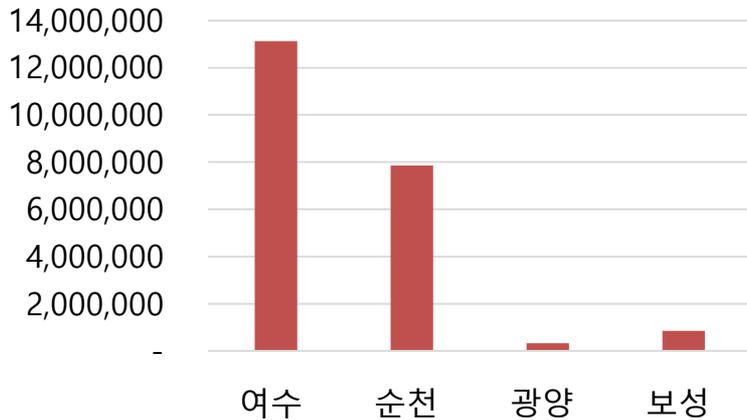


관광지 개수



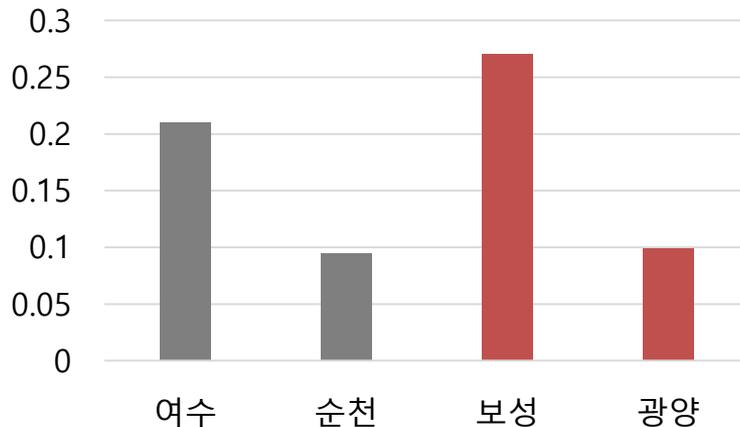
남도바닷길 : 권역 특성 분석

도시별 관광객 수



4개의 지역이 하나의 권역임에도 여수,순천/보성,광양 지역간의 방문객 편차가 극심하나, 전반적 만족도는 오히려 보성,광양이 더 높다.

전반적 만족도



보성, 광양은 낮은 방문객 숫자에도 만족도가 여수/순천지역에 뒤쳐지지 않으므로 관광지로서의 성장가능성이 충분하다고 판단하였다. 따라서 분석을 통해 보성,광양지역 관광객 유치방안을 마련할 필요가 있음

남도바닷길 : 권역 특성 분석

Web Crawling

```
String location = "전남+여행";
String page = "1";
for (int j = 1; j < 101; j++) {
    Document document = Jsoup
        .connect("https://search.naver.com/search.naver?where=post&sm=tab_pge&query=" + location
            + "&st=sim&date_option=0&date_from=&date_to=&dup_remove=1&post_blogurl="
            + "&post_blogurl_without=&srchby=all&nso=&ie=utf8&start="
            + page)
        .get();
}
```

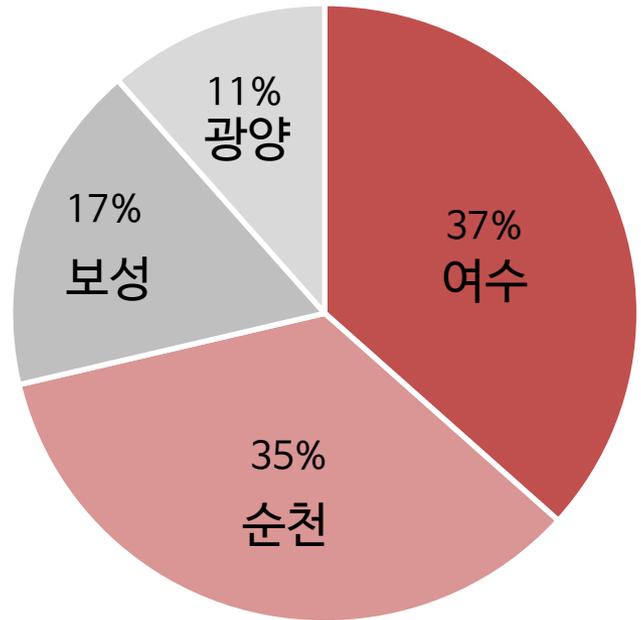


블로그 1-10 / 170,423건



Jsoup 라이브러리를 이용해 JAVA로 전남+여행 으로 검색한 네이버 블로그 1000개를 Web Crawling 한 후 4개 지역 단어의 출현 빈도를 조사하였다

포스팅 빈도수



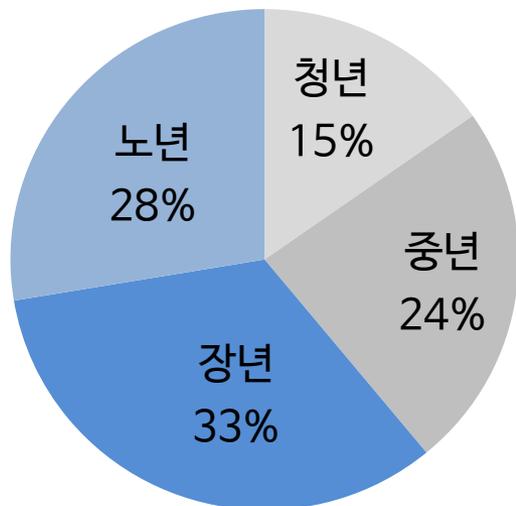
포스팅 빈도는 지역간 방문객의 관심도 편차를 뒷받침하는 자료이므로, 상대적으로 낮은 보성,광양의 인지도 현실을 보여줌

남도바닷길 : 도시간 방문객 편차 문제

분석단계

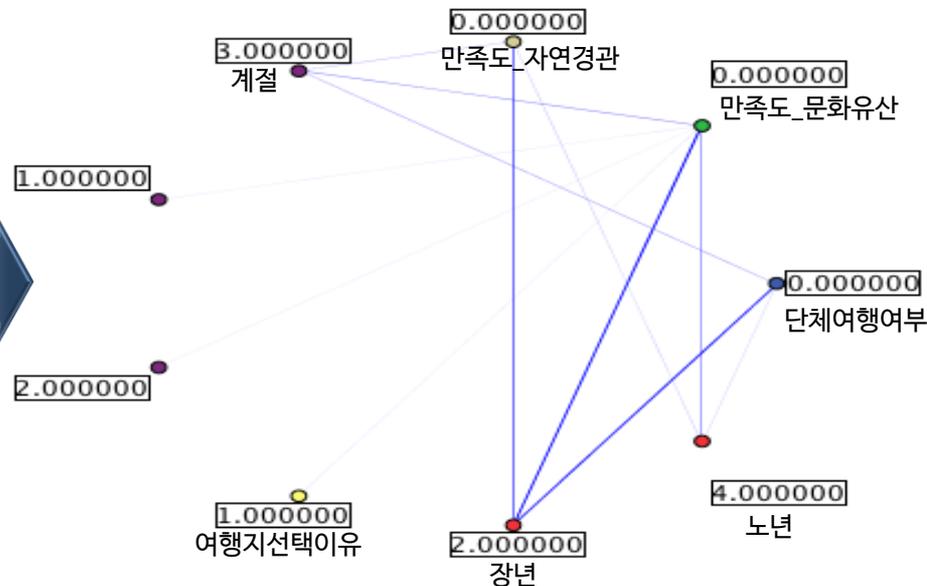
1. 여행객 연령별 빈도분석 및 속성간 연관성 분석을 통해 여행객의 특성 확인 및 분석 대상 변수 선별
2. 계층적 군집분석과 k-means clustering을 모두 활용한 최적의 군집화로 타겟층에 대한 타당성 확보
3. 여행목적 및 만족도를 설명변수로 갖는 의사결정나무를 통해 타겟층을 유형화한 뒤 마케팅 방안 도출
4. 총지출액을 목표변수로 다중선형회귀분석을 실시하여 마케팅 방안에 대한 근거 마련

남도바닷길 관광객 연령별 비율



여행객비율을 살펴보면 장년/중년/노년 층의 비율이 대부분임을 알 수 있음.

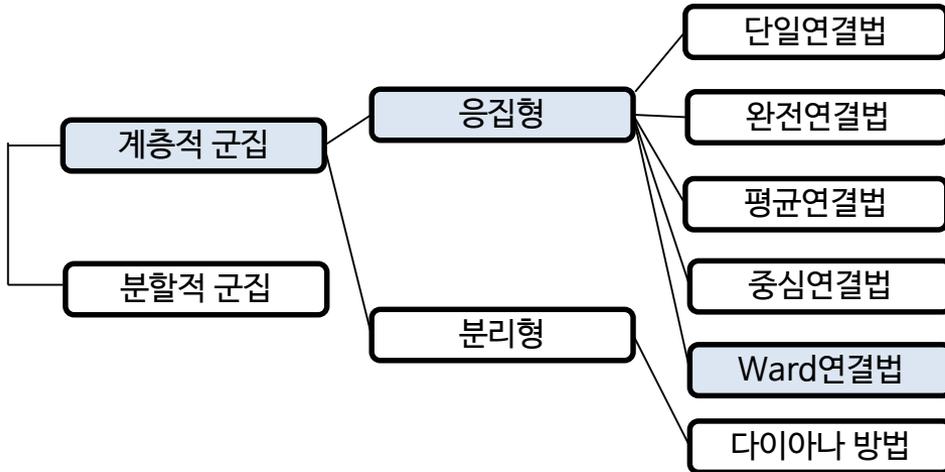
연관성 분석



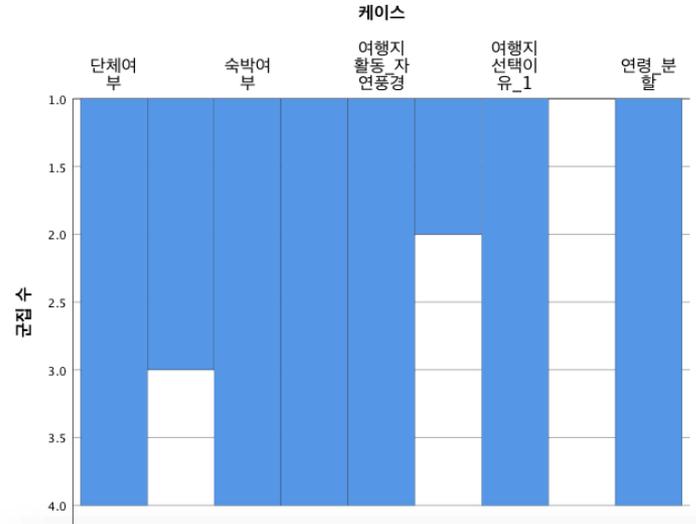
연관성분석을 통해 장년층과 노년층을 중심으로 특정한 여행 패턴이 형성되고 있음을 확인함. 이때 각 연령층과 연관이 두텁게 나오는 속성들을 이용하여 k-means 분석을 계획함

남도바닷길 : 도시간 방문객 편차 문제

계층적 군집분석



Ward 연결법

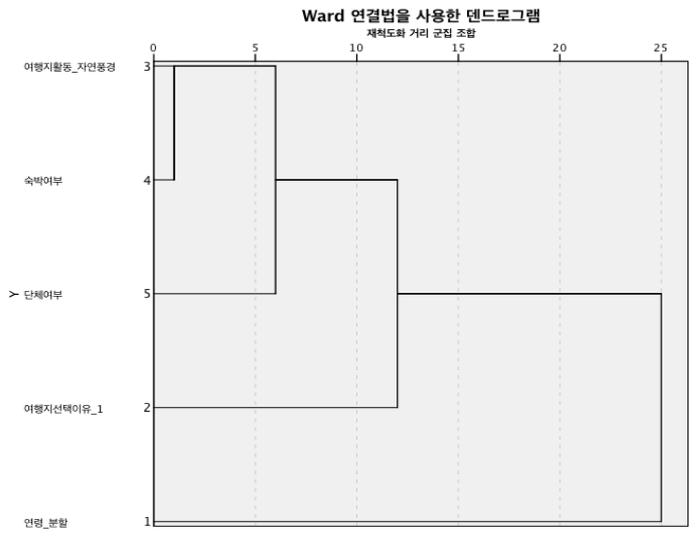


Ward 연결법 사용 이유

K-means 군집분석에 앞서, 어떠한 속성으로 몇개의 군집을 생성할지 결정하기 위해 계층적 군집분석을 실시함.

초기 38000여개 표본에서 테마별 분석으로 세분화하며 데이터의 갯수가 1000개 남짓으로 줄어들었기 때문에 이상치의 영향력에 대한 위험도가 증가했다고 판단하였음. 따라서 이상치에 민감하지 않은 ward 연결법을 사용함.

그 결과 장년, 노년 모두 4개의 군집으로 분류됨. 따라서 4개의 그룹으로 k-means clustering하기로 결정함

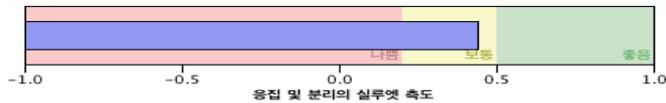


남도바닷길 : 도시간 방문객 편차 문제

여수 순천

크기	37.9% (293)	33.2% (257)	19.3% (149)	9.6% (74)
입력	숙박여부 2.000 (100.0%)	숙박여부 1.000 (100.0%)	숙박여부 1.000 (100.0%)	숙박여부 1.000 (56.8%)
	단체여부 0.000 (100.0%)	단체여부 0.000 (100.0%)	단체여부 0.000 (67.1%)	단체여부 1.000 (100.0%)
	연령_분할 2.000 (36.2%)	연령_분할 2.000 (41.2%)	연령_분할 4.000 (100.0%)	연령_분할 3.000 (45.9%)
	여행지선택이유_1 1.000 (80.8%)	여행지선택이유_1 1.000 (70.6%)	여행지선택이유_1 1.000 (57.7%)	여행지선택이유_1 1.000 (57.5%)
	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)

군집 품질



보성 광양

크기	51.0% (130)	31.8% (81)	11.4% (29)	5.9% (15)
입력	단체여부 0.000 (100.0%)	단체여부 0.000 (100.0%)	단체여부 1.000 (100.0%)	단체여부 1.000 (100.0%)
	숙박여부 1.000 (100.0%)	숙박여부 2.000 (100.0%)	숙박여부 1.000 (100.0%)	숙박여부 2.000 (100.0%)
	연령_분할 2.000 (40.8%)	연령_분할 2.000 (50.6%)	연령_분할 3.000 (44.8%)	연령_분할 2.000 (46.7%)
	여행지선택이유_1 1.000 (63.0%)	여행지선택이유_1 1.000 (57.1%)	여행지선택이유_1 1.000 (50.0%)	여행지선택이유_1 1.000 (42.9%)
	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)	여행지활동_자연풍경 1.000 (100.0%)

군집 품질



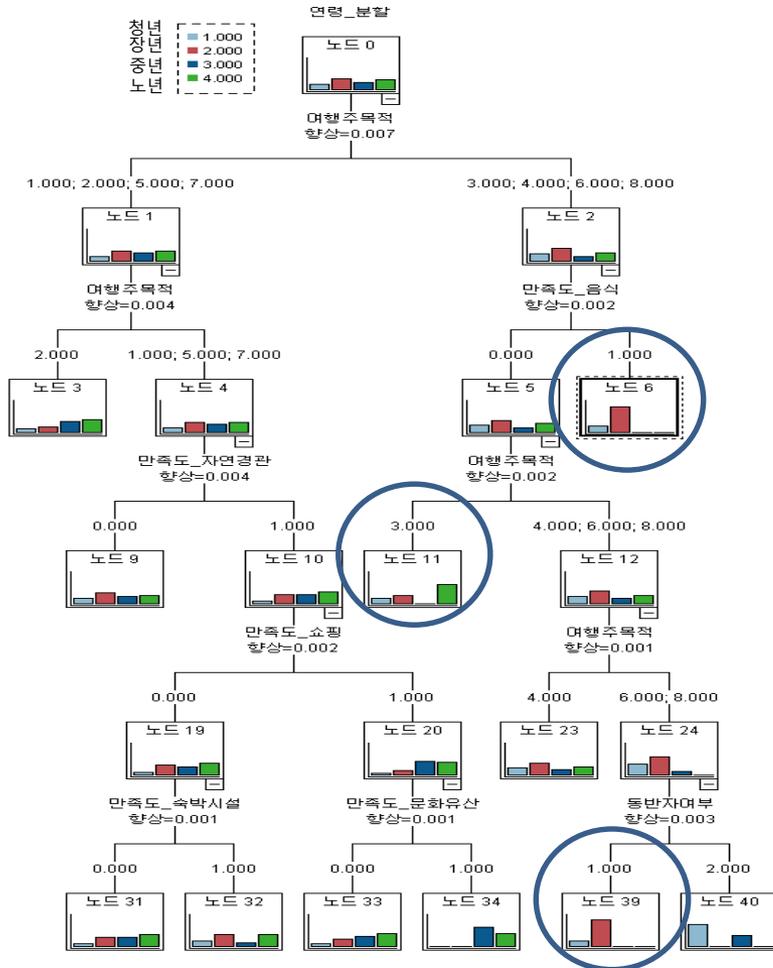
K-means 클러스터링

여수와 순천을 한 권역으로, 광양과 보성을 한 권역으로 묶어 k-means 클러스터링을 실행함
두 권역 모두 장년/노년층을 중심으로 군집이 형성되었음을 확인함.

따라서 남도 바닷길 권역을 방문한 장년, 노년들이
여행지를 선택할 때 중요하게 생각하는 변수를 파악하기로 함

남도바닷길 : 도시간 방문객 편차 문제

의사결정나무



변수선택

타겟변수 : 연령_분할 (age_divided)

예측변수 : 여행 주 목적 (q3), 항목별 만족도 (q12_1~q12_6)

변수선택이유

장년층과 노년층의 여행 목적과 여행지에서의 활동에 대한 선호도가 다를 것으로 예상하여 두 변수에 따른 연령분포를 확인하고자 함

의사결정나무 분석

- 남도 바닷길 권역 전체를 대상으로 의사결정나무를 활용함
- 모집단을 연령층으로 구분한 뒤 여행 목적에 따라 crt-tree를 그려 본 결과, 장년/노년층이 두드러지는 노드 (6, 11, 39) 를 살펴보면 이들의 특성을 살필 수 있음
- 장년층 - 여가/위락 목적의 여행보다는 종교/가족방문/쇼핑 목적의 여행이 눈에 띄었다. 음식에대한 만족도가 높았음
- 노년층 - 여가/위락보다는 종교/성지순례 목적의 여행이 눈에 띄었다. 음식에 대한 만족도는 낮은 것으로 나타남

남도바닷길 : 도시간 방문객 편차 문제

타겟층의 이해를 위한 또 다른 근거 확보를 위해 총 지출액을 타겟변수로 설정하여 회귀분석을 실시함

<장년>

변수명	변수설명	변수유형	사용여부	이유
q11_b_1	여행상품 구입비	numeric	X	결측치 과다
q11_b_2	숙박비	numeric	O	
q11_b_3	교통비	numeric	O	
q11_b_4	식비	numeric	O	
q11_b_5	문화오락비	numeric	X	다중공선성(쇼핑비)
q11_b_6	스포츠 관람비	numeric	X	결측치 과다
q11_b_7	쇼핑비	numeric	O	
q11_b_8	여행용품비	numeric	X	결측치 과다
q11_b_9	기타비용	numeric	X	결측치 과다

<노년>

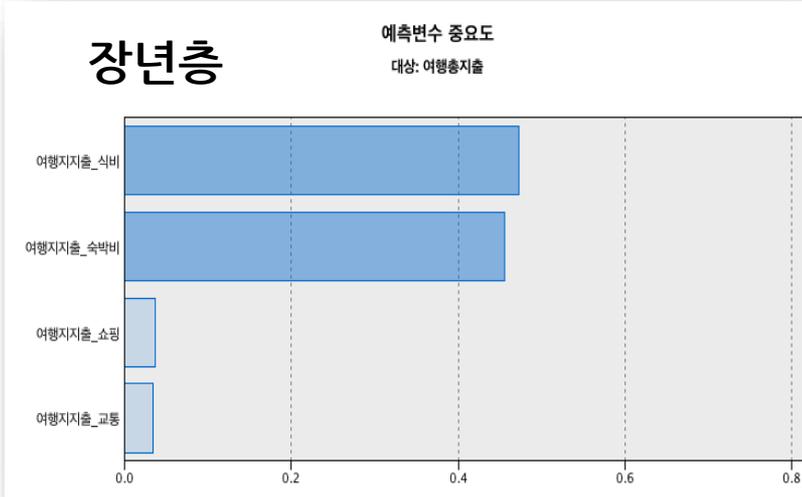
변수명	변수설명	변수유형	사용여부	이유
q11_b_1	여행상품 구입비	numeric	X	결측치 과다
q11_b_2	숙박비	numeric	O	
q11_b_3	교통비	numeric	O	
q11_b_4	식비	numeric	O	
q11_b_5	문화오락비	numeric	X	결측치 과다
q11_b_6	스포츠 관람비	numeric	X	결측치 과다
q11_b_7	쇼핑비	numeric	X	결측치 과다
q11_b_8	여행용품비	numeric	O	
q11_b_9	기타비용	numeric	O	

회귀분석 종속변수 설정

- 결측치 75% 이상은 원 데이터로서의 의미가 없다고 판단하여 변수에서 제외함
- 종속변수간의 상관관계가 뚜렷하여 다중공선성 문제가 발생할 것으로 보이는 변수를 제거함

남도바닷길 : 도시간 방문객 편차 문제

다중선형회귀분석



신뢰도 검증

- ANOVA 분석 결과 F-test 유의확률은 0.00으로 본 회귀모형은 신뢰할 만 하다.
- R제곱값으로 미루어보아 식비/숙박비/쇼핑비/교통비가 총 지출의 77% 이상을 설명하는것으로 나타났다. R제곱값과 수정된 R제곱값이 거의 차이하지 않는것으로 보아 독립변수의 설명력을 신뢰할만 하다.
- 유의확률은 쇼핑과 교통을 제외한 변수에서 0.05이하로 나타났다. 따라서 나머지 숙박비, 식비 두 가지 속성에 대한 추정은 신뢰할만 하다.

ANOVA

모형	제공함	자유도	평균제곱	F	유의확률
1	회귀	4	3.620E+11	67.909	.000 ^b
	잔차	73	5.330E+9		
	전체	77			

b. 예측자: (상수), 여행지지출_쇼핑, 여행지지출_숙박비, 여행지지출_교통, 여행지지출_식비

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.888 ^a	.788	.777	73009.4672

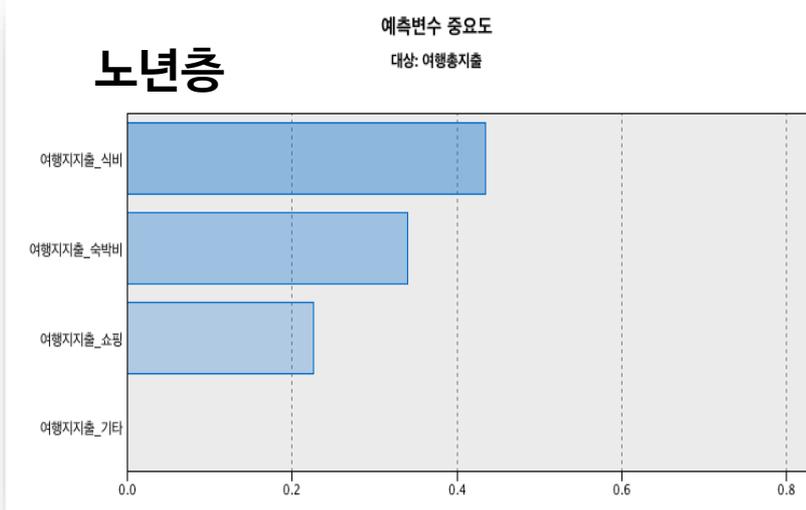
a. 예측자: (상수), 여행지지출_쇼핑, 여행지지출_숙박비, 여행지지출_교통, 여행지지출_식비

계수

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	37790.978	11877.094		3.182	.002
	여행지지출_숙박비	1.073	.185	.422	5.793	.000
	여행지지출_교통	.932	.509	.119	1.829	.071
	여행지지출_식비	1.471	.245	.462	6.014	.000
	여행지지출_쇼핑	.560	.502	.064	1.116	.268

남도바닷길 : 도시간 방문객 편차 문제

다중선형회귀분석



ANOVA

모형	제곱합	자유도	평균제곱	F	유의확률
1 회귀	3.133E+11	5	6.266E+10	94.236	.000 ^b
잔차	3.790E+10	57	664906466		
전체	3.512E+11	62			

b. 예측자: (상수), 여행지지출_쇼핑, 여행지지출_식비, 여행지지출_교통, 여행지지출_기타, 여행지지출_숙박비

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.933 ^a	.870	.861	28009.7519

a. 예측자: (상수), 여행지지출_기타, 여행지지출_숙박비, 여행지지출_쇼핑, 여행지지출_식비

신뢰도 검증

- ANOVA 분석 결과 F-test 유의확률은 0.00으로 본 회귀모형은 신뢰할 만 하다.
- R제곱값으로 미루어보아 식비/숙박비/쇼핑비/교통비/기타 비용이 총 지출의 86% 이상을 설명하는것으로 나타났다. R 제곱값과 수정된 R제곱값이 거의 차이가 나지 않는것으로 보아 독립변수의 설명력을 신뢰할만 하다.
- 유의확률은 상수를 제외한 변수에서 모두 0.05이하로 나타났다. 따라서 나머지 식비, 숙박비, 쇼핑비, 교통비, 기타비용 다섯 가지 속성에 대한 추정신뢰할만 하다.

계수

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	5155.454	5895.454		.874	.386
	여행지지출_숙박비	.690	.177	.208	3.896	.000
	여행지지출_교통	1.728	.511	.178	3.382	.001
	여행지지출_식비	2.085	.246	.376	8.483	.000
	여행지지출_기타	1.049	.057	.810	18.323	.000
	여행지지출_쇼핑	1.147	.159	.319	7.222	.000

남도바닷길 : 도시간 방문객 편차 문제

다중선형회귀분석 결과 모델 및 해석

$$Y = b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 + b_5 * x_5 + C$$

- 회귀분석 결과 유의수준이 과하게 높아 신뢰할 수 없는 변수는 제외
- 연령층별로 다른 독립변수를 설정하여 회귀분석모델확립

<장년>

변수	x1	x2	x3	x4	x5
속성	숙박비	교통비	식비	쇼핑비	상수
계수(b)	1.073	0.9316	1.471	0.5601	37791.0

<노년>

변수	x1	x2	x3	x4	x4	x5
속성	숙박비	식비	쇼핑비	교통비	기타지출	상수
계수(b)	0.6895	2.085	1.147	1.728	1.049	5155.5

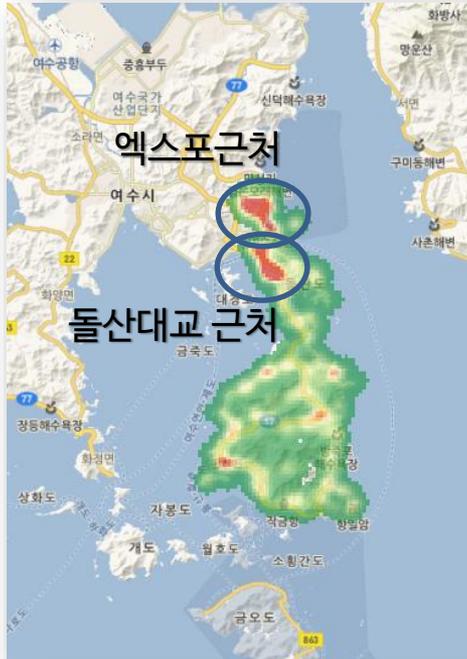
두 계층 모두 식비의 영향력이 가장 큰 것으로 나타남

노년층은 장년층에 비해 총지출에 대한 교통비의 영향력이 높게 나타남

X3(식비)가 한 단위 늘면 Y(총지출)이 1.471 단위 만큼 증가함.
따라서 식비 지출이 늘수록 경제적 효과가 큼

경제적 효과를 극대화 할 수 있도록 보성, 광양 지역에 장년, 노년층을 대상으로 한 음식 마케팅을 통해 방문객을 유인할 전략이 필요함

남도바닷길 : 도시간 방문객 편차 문제



여수, 순천의 통신사 데이터 분석 결과, 실제 장년과 노년층의 관광객 비율이 크에도 연령별로 구분되는 인구 밀도가 나타나지 않고 비슷한 패턴을 보임



여수, 순천 또한 소수의 유명 관광지에 대부분의 관광객이 밀집되는 것으로 판단되므로, 잘 알려지지 않은 관광지를 연령별로 맞춤 홍보하여 관광객을 분산할 필요가 있음

백제문화권 마케팅 : 숙박을 제고 방안

“친지방문 위주의 가족단위 방문자를 여가/위락/휴가 목적의 관광객으로 전환”



공주: 한옥마을



익산: 응포캠핑장



부여: 롯데부여리조트 (워터파크)



대전: 유성호텔(은천)



WordCloud를 통한 텍스트마이닝 결과

- “지역+가족여행” 키워드로 1000개의 네이버 블로그 포스팅을 크롤링
- “부여리조트”와 “물놀이”, “백제문화”, “부소산성” 등의 검색어의 빈도수 높음

1. 주요 역사 관광지 야간개장 : 서동요 테마파크, 공남지, 공산성
 - 역사교육 효과가 있어 아이가 있는 가족단위 관광객 유인 가능
 - 야간 관광 활동으로 숙박 유도 가능

2. 도시별 주요 숙박시설을 중심으로 관광코스 개발
 - 지역별로 특색 있고, 다양한 체험이 가능한 숙박시설 홍보로 숙박율 증대
 - 친지방문 목적을 여가/위락/휴가 목적으로 전환 가능

남도바닷길 마케팅 : 지역간 방문객 편차 해결 방안

“장년, 노년층의 만족도를 높이는 맞춤 관광”



순천: 회정식



광양: 불고기



보성: 꼬막정식



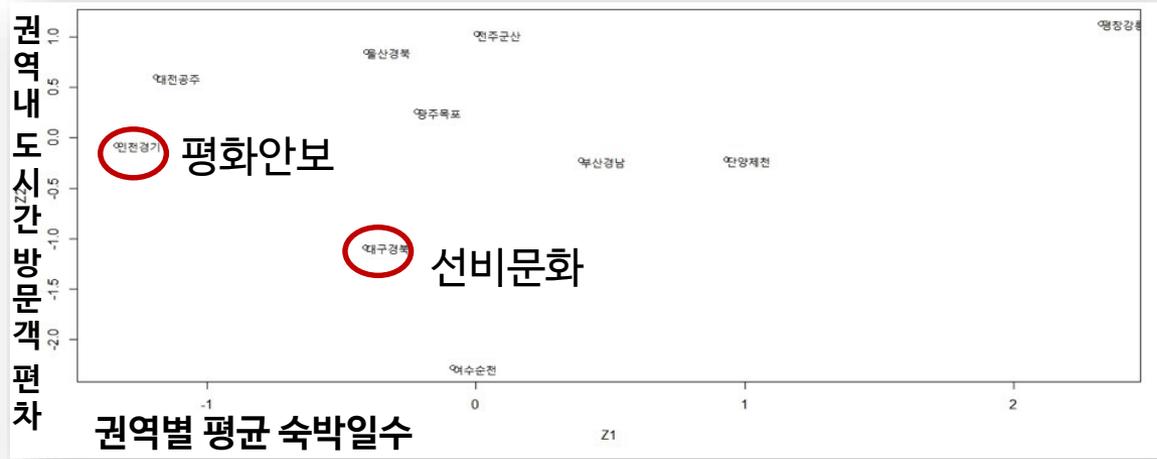
여수: 간장게장

1. 각 지역의 대표적인 음식을 특화하여 집중적으로 홍보
 - 음식 소비 만족도가 높은 장년층이 다양한 도시로 향하도록 유도
 - 음식 소비 만족도가 낮은 노년층의 만족도 향상
 - 식비지출을 늘려 경제적 효과 유발 (회귀분석결과)
2. 자연감상, 종교목적 방문이 가능한 '사찰'을 관광지 코스에 추가
 - 잘 알려지지 않은 관광지로 관광객 분산 가능
 - 장년, 노년층의 방문목적과 남도바닷길 테마에 부합
 - 교통비에 민감한 노년층을 고려해 관광 투어버스 신설

- (1일차) 서울→여수엑스포 →박람회장 → 오동도 → **항일암** → 여수밤바다 1박(여수)
- (2일차) 이순신대교 → 구봉산전망대 → 망덕포구 → 매화마을 → **옥룡사지** 2박(광양)
- (3일차) 순천만자연생태공원 → 순천만국가정원 → 낙안읍성 → **선암사** 3박(순천)
- (4일차) 태백산맥 문학관 → 보성차밭 → 제암산 휴양림 → **대원사** → 귀가

분석 결과의 활용 방안

평화안보권역(파주, 인천, 수원, 화성)과 선비문화권역(영주, 안동, 대구, 문경) 역시 각각 체류기간이 짧고, 도시간 방문객 편차가 크다는 문제점이 있다. 이를 해결하기 위해 앞서 제시한 분석방법을 활용해 문제를 해결할 수 있다.



평화안보권역

로지스틱 회귀분석 결과	목적별 방문자비율 확인	다중대응 분석	마케팅 전략
“여행목적이 숙박율에 영향을 미친다”	평화안보권역에 “여가/위락/휴가” 이외의 다른 목적 여부 파악	여가/위락/휴가목적의 여행을 늘릴 수 있는 타겟층 도출	타겟층을 대상으로 적절한 전략 구상

선비문화권역

계층적 군집분석 및 K-means clustering	의사결정나무	다중선행회귀분석	마케팅 전략
타겟층 도출	타겟층 유형화	항목별 지출액 분석으로 타겟층의 니즈 파악	타겟층을 대상으로 적절한 전략 구상

감사합니다