



관광지 추천시스템 개발 및 체질 개선 방안 수립

SA 211



CONTENTS



서론

프로젝트 동기

본론

여행객 분류
- 데이터 통합 (R programming)

본론

여행객 분류
- 동행자/여행경비별 여행객 재분류 (Clustering analysis)

본론

만족도 및 방문수 여행지 분류 (BCG Matrix)

결론

관광지 추천시스템 (Visualizing, Matching, cost analysis)

결론

관광지 체질 개선 방안 수립 (Logistic analysis)

서론

프로젝트 동기



Simplification (단순화)

정부 · 지자체

관광상품



국내 관광객

판매자

마케팅

상품개발보완

소비자

의사결정

적합상품 선정

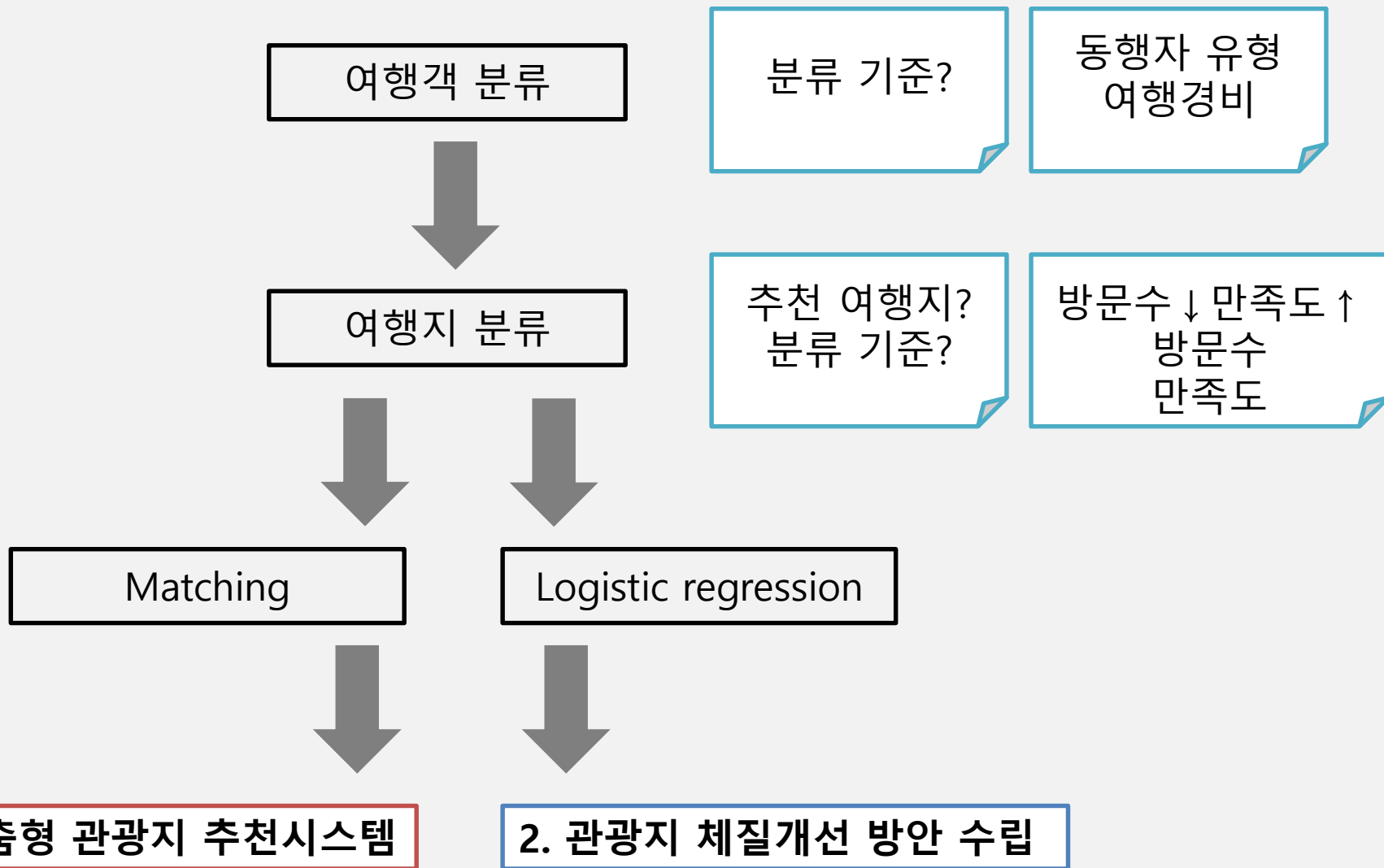
관광산업

관광지 체질개선 방안 수립

선순환 구조

맞춤형 추천시스템 개발

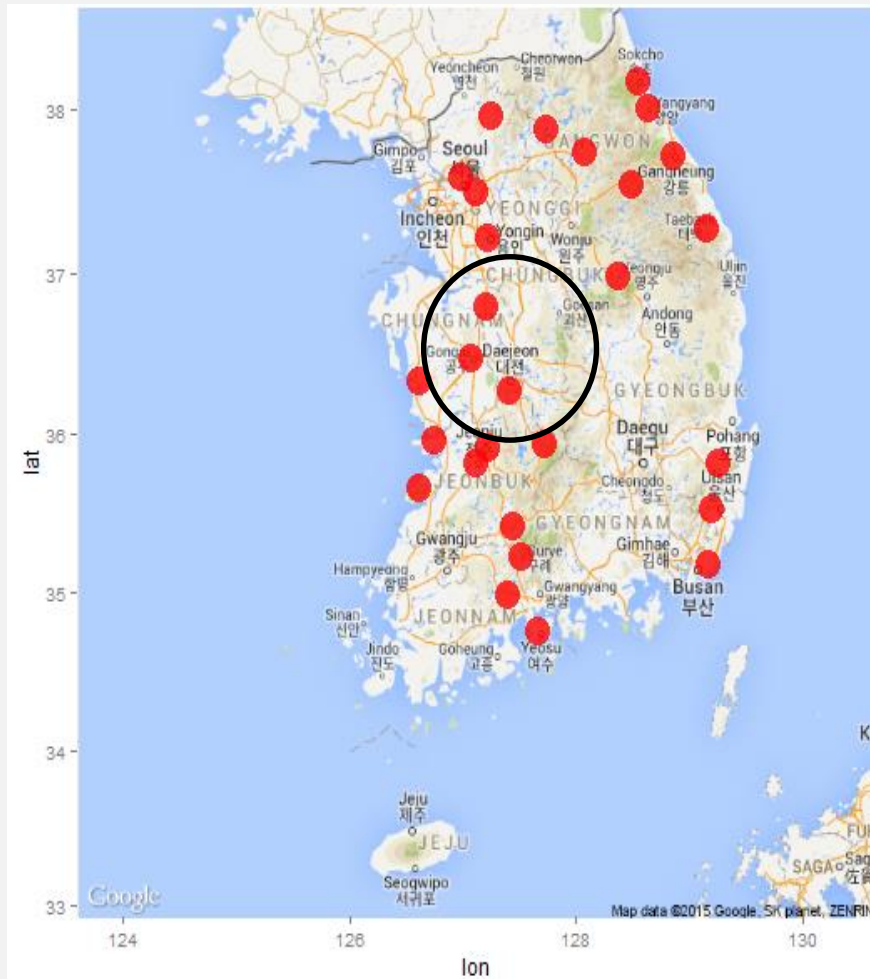
문제 정의 및 변수 설정



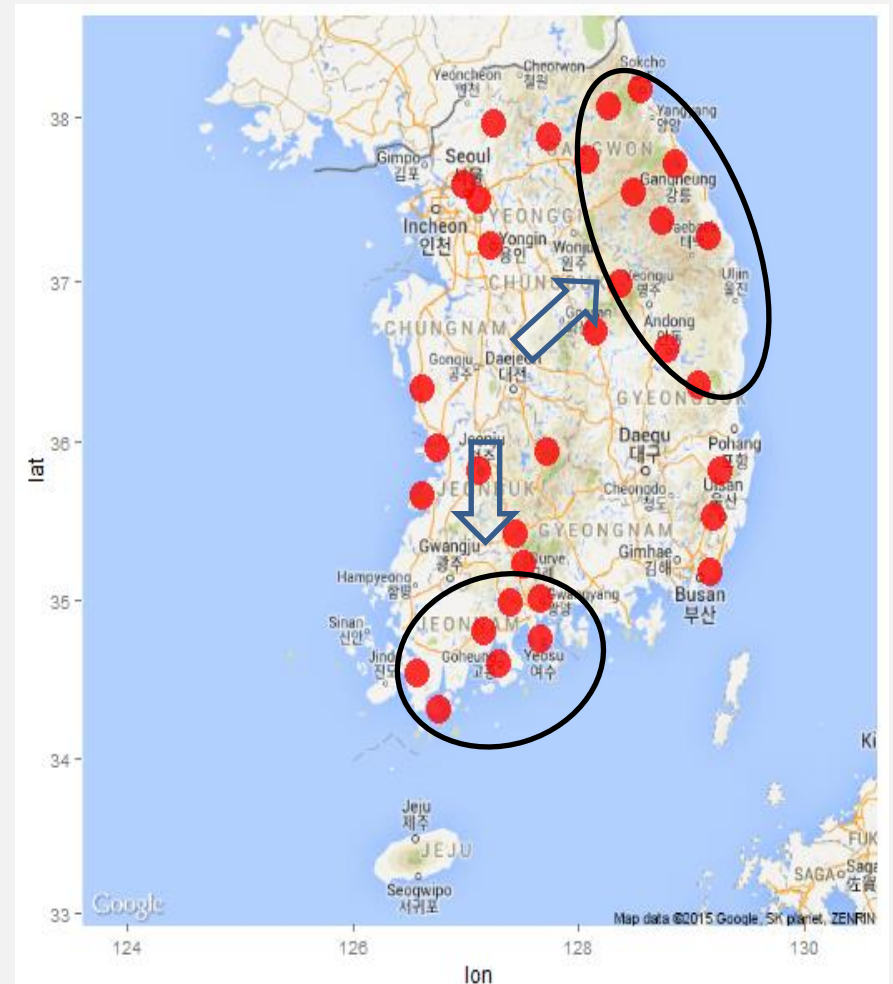
상위 20% 인기관광지

동행자 : 가족

동행자 : 친구/연인



< 충남 내륙 >

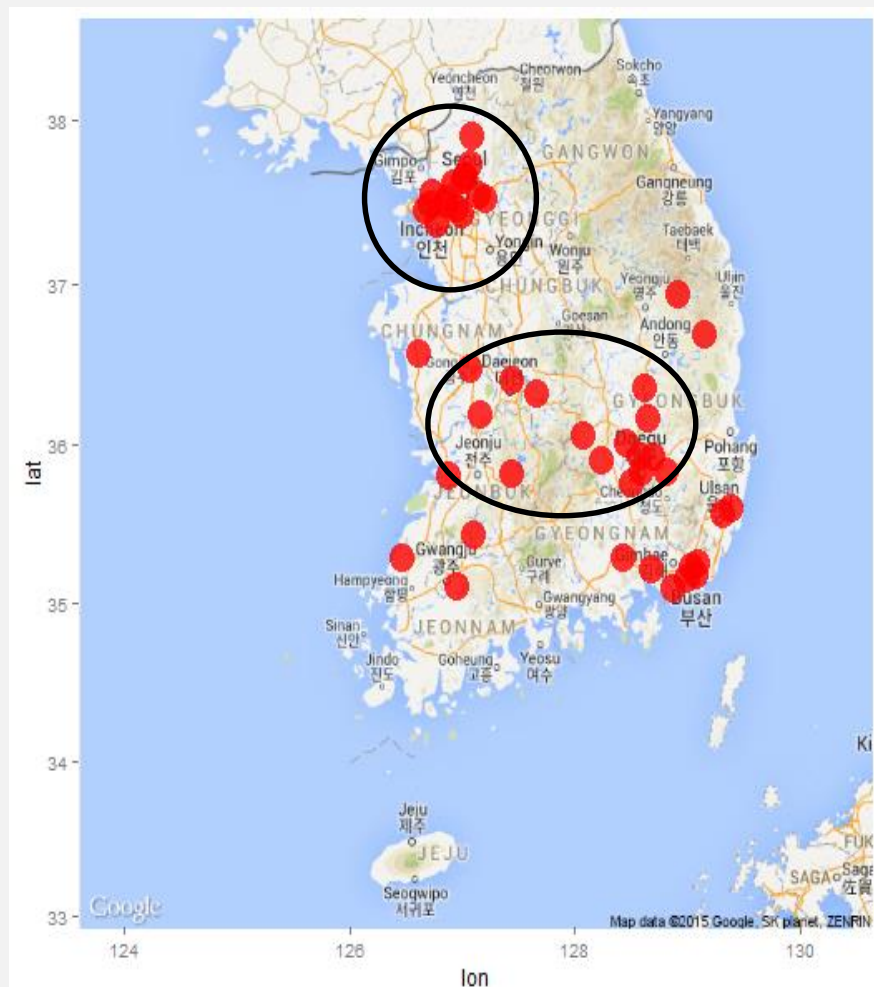


< 전남, 강원 해안 >

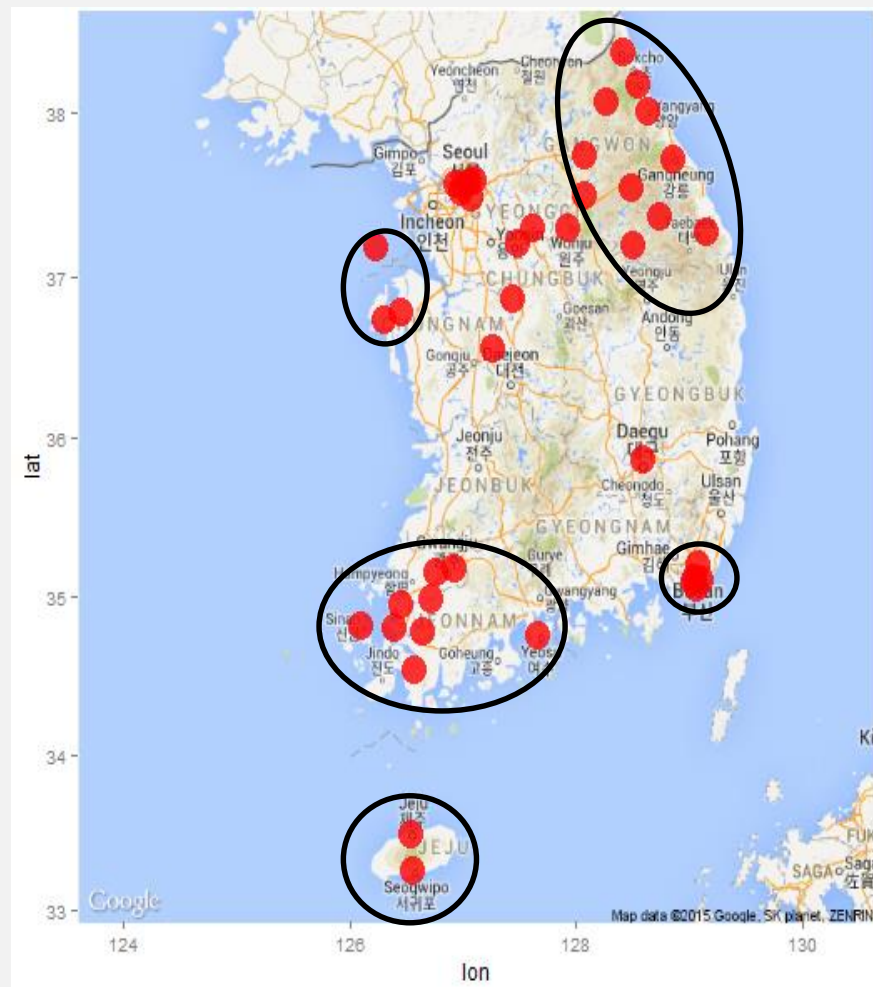
동행자 : 친구연인

저소비군 (하위 20%)

고소비군 (상위 20%)



< 대도시, 내륙 >



< 전남 강원 해안, 제주, 부산집중 >

1. 맞춤형 관광지 추천 시스템



서울, 20대

친구들과 10만원으로 여행하고 싶은데...
가평이나 인천은 가봤고...
국내에 여행할 색다른 곳 어디 없을까?

한국의 아름다운 고장!
강원 정선, 경북 문경, 충북 제천을
추천합니다!



관광지 추천 시스템



본론

여행객 분류
데이터 통합 및 재분류 (R programming)

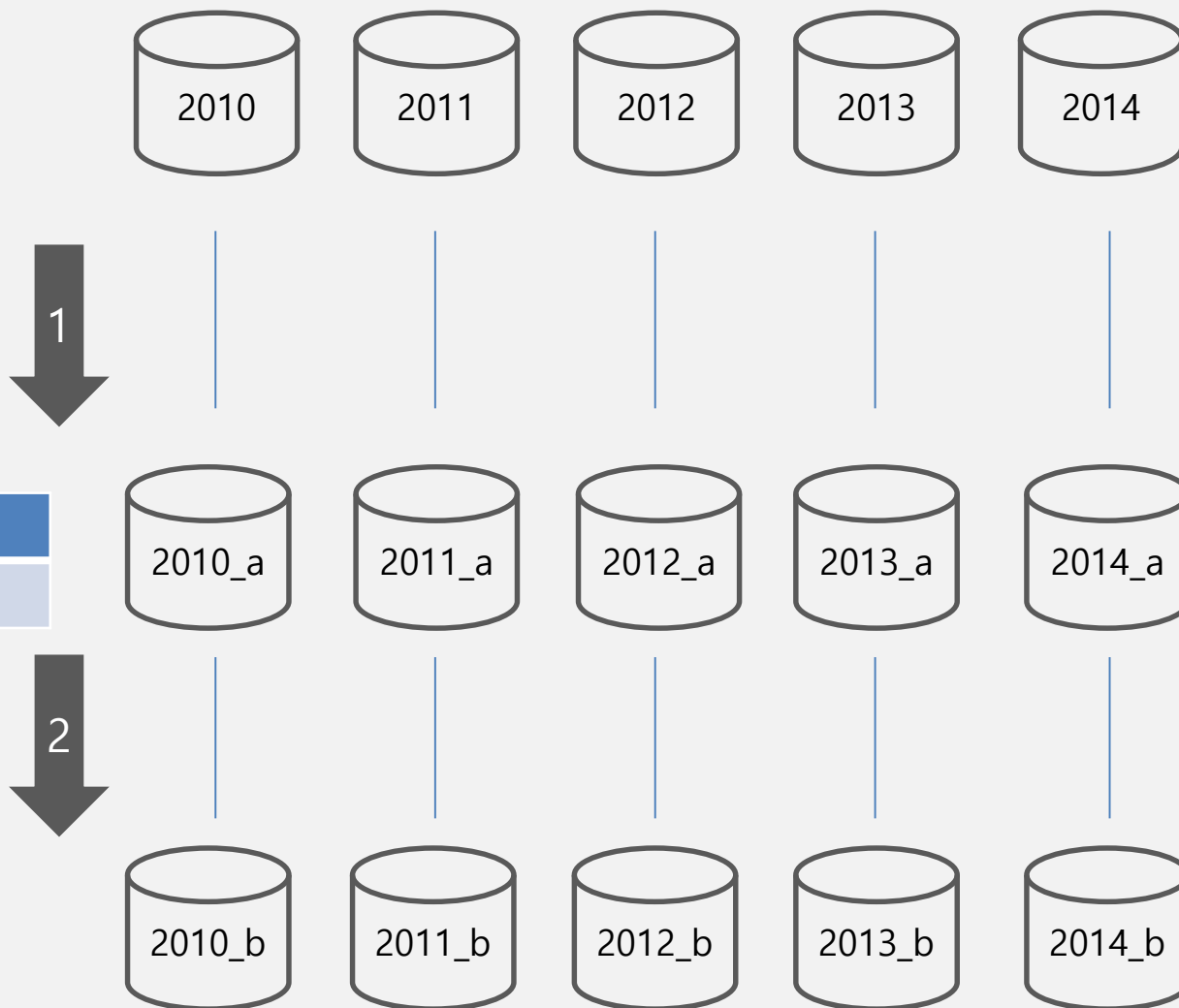
<초기 여행 데이터>

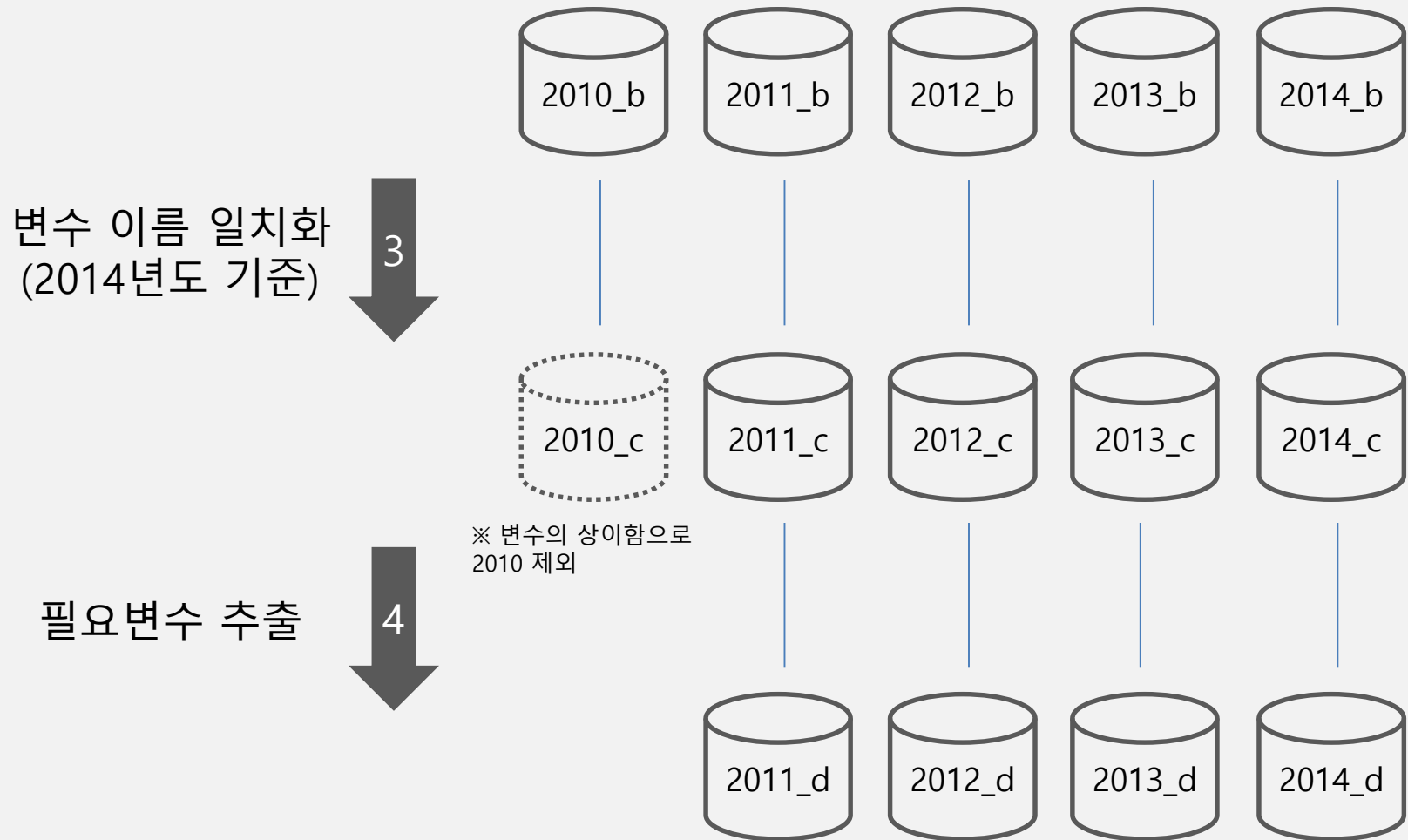
q1값 == 국내여행

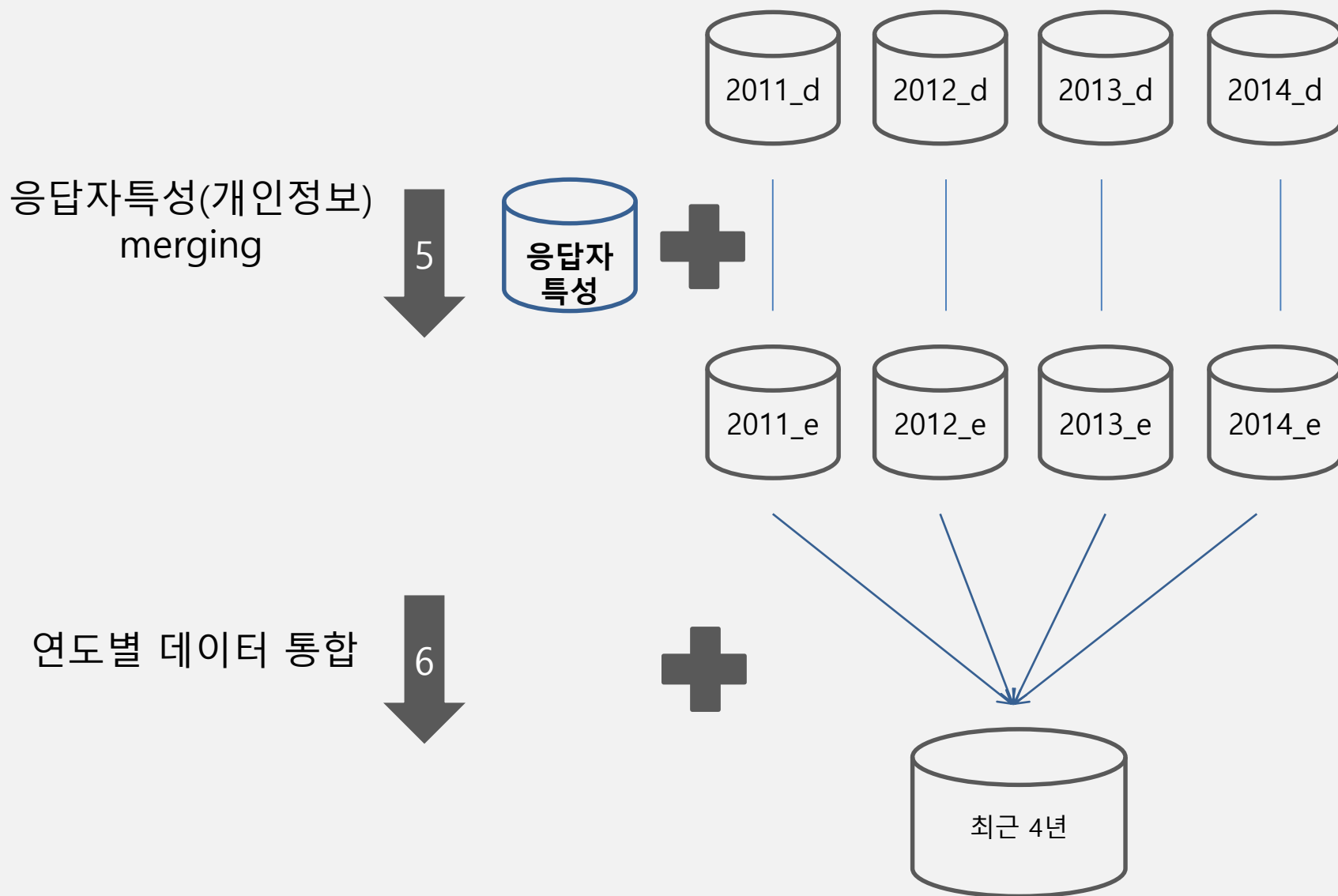
PID	1.1	1.2	2.1	2.2	...
001	3	5	4	3	...

여행별 컬럼 row화
(1행1값)

PID	1.1	2.1
001	3	4
001	5	3
...



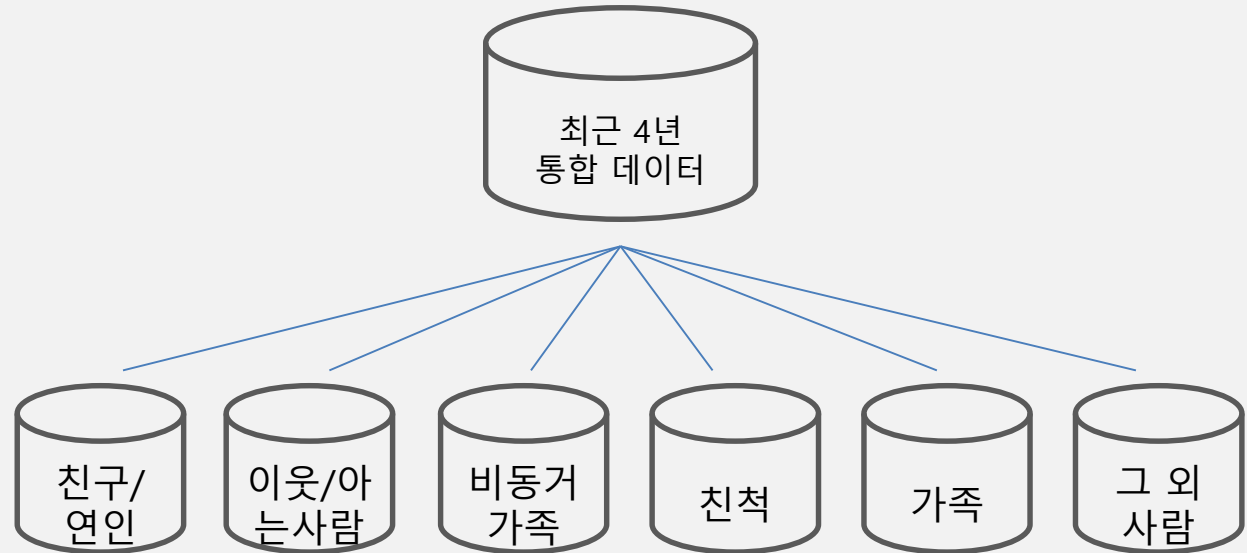




여행객 분류 - 동행자유형/여행경비별 여행객 분류 (Clustering analysis)

동행자유형별
집단 재분류

7



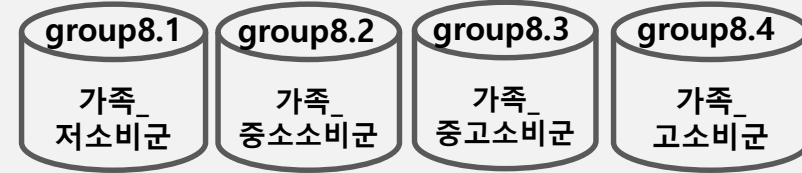
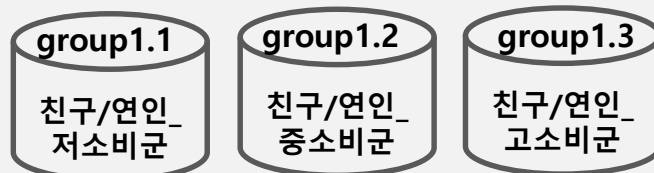
여행경비
클러스터링 분석
집단 재분류

8



K-means method
※ 표본크기 고려 2~4개 집단 분류
.....

<여행객 분류>
총 19개 그룹



본론

여행지 분류 (BCG Matrix)



BCG Matrix를 활용한 여행지 분류

추천 1순위 관광지

유명한 여행지보다 잠재적 성장가능성이 높은 hidden place

(현재 여행방문수 낮고, 만족도는 높은 여행지)
(가평,해운대,경주가 아닌 강원도 정선 or 충남 보령 등)

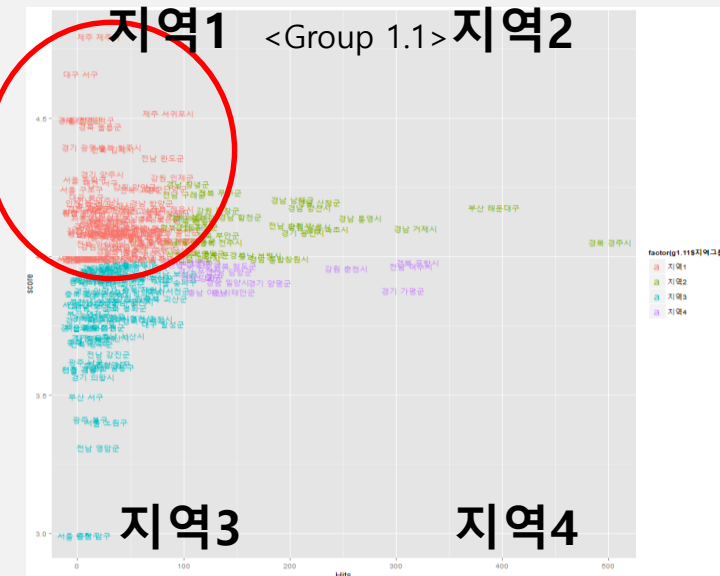
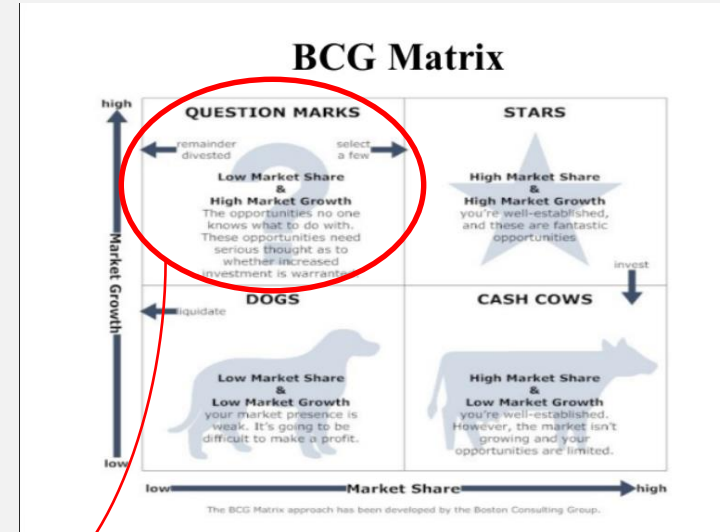
“관광지도 하나의 비즈니스 상품이다”

X 축 : 방문수 (시장점유율) , 기준 : 상위 20% (방문수 점유 60.092%)
Y 축 : 만족도 (잠재성장성) , 기준 : 4점 (만족)

각 집단 별 지역1(QUESTION MARKS) 추천

Group 1.1(친구/연인 동반, 평균 10만원 여행경비)인 경우

1. 여행객이 방문한 여행지를 BCG Matrix 활용, 4개 지역으로 분류
2. 적색 지역1 : 방문수 96.8 이하, 만족도 4 이상
2. 해당 지역 중 적색계열(지역1) 추천



결론

관광지 추천시스템
(Visualizing, Matching, cost analysis)



관광지 추천 시스템

<여행객 정보>

여행정보
누구와? 친구
여행비용은? 100000원

개인정보
사는 곳? 서울
나이? 20대
...

여행정보
(동행자, 비용)

개인정보
(나이, 지역)



<19개 여행객 그룹>

group1.1
친구/연인_저소비군

group1.2
친구/연인_중소비군

group1.3
친구/연인_고소비군

.....

group8.1
가족_저소비군

group8.2
가족_중소비군

group8.3
가족_중고소비군

group8.4
가족_고소비군

<4개 여행지 그룹>

지역1

지역2

지역3

지역4

만족도 높고 방문수 낮은 잠재유망 관광지 추천 / 홍보

관광지 추천 시스템 : 집단별 추천 여행지 및 전략 도출

그룹1.1

동행자

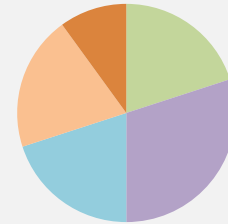
친구/연인

여행비용

저 소비 군

범위 : 148750 원 이하
대표값 : 100000 원

<주요 비용 분석>

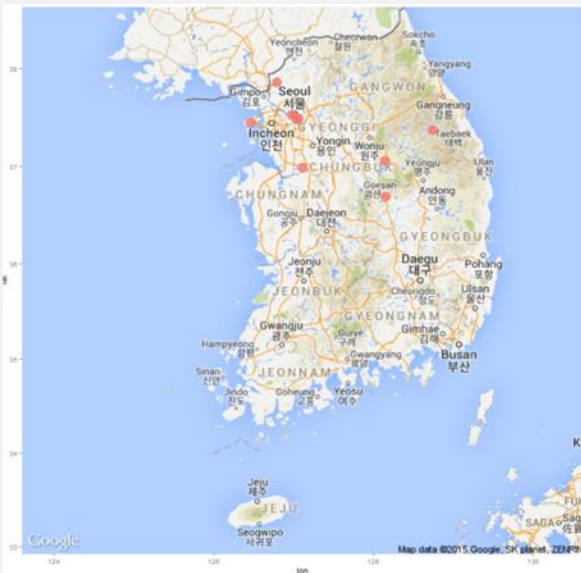
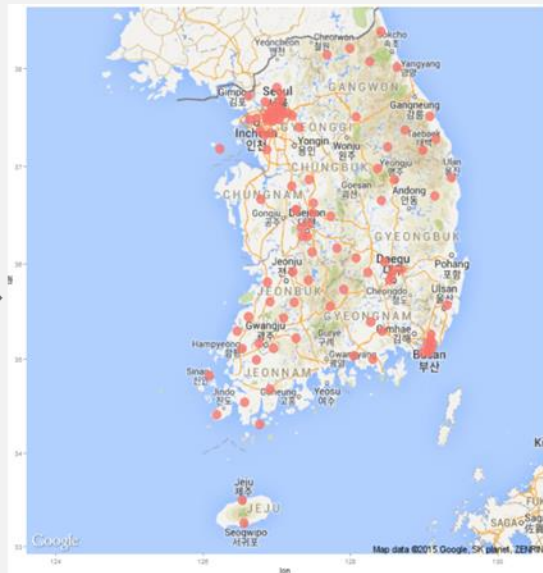
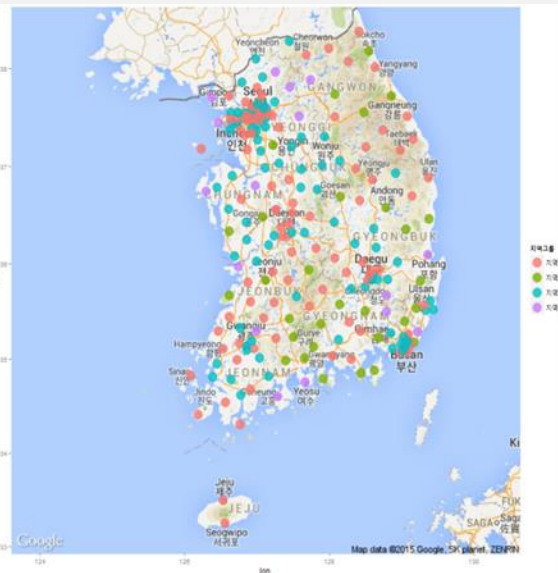


숙박비	0 원
교통비	20000 원
식비	30000 원
문화시설	20000 원
쇼핑비	20000 원
기타비용	10000 원
= 100000 원	

총 여행지 분포

지역1 여행지 선별

개인정보활용 최종 추천지 선정



여행객 그룹/지역 그룹
matching

여행객 개인정보
matching

추천 여행지 도출 여행지 전략 수립

※ 경기도 가평에 집중된 서울 20대 여행객
정선, 문경, 제천 등 지방 분산화
※ 해당 지역은 포지셔닝 전략 수립

서울, 20대

강원 정선군

경북 문경시

충북 제천시

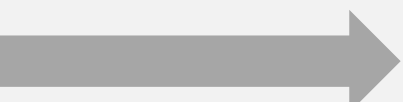
결론

관광지 체질 개선 방안 수립
(Logistic analysis)



만족도 분석의 정의 및 특성 파악

- ▶ 가정: 전반적 만족도는 항목별 만족도와 관련되어 있을 것이다.
- ▶ 목표: 모델링을 통하여 전반적 만족도와 항목별 만족도 간의 관계를 찾아내고 이를 이용해서 4그룹으로 나눈 지역에 대하여 적용하여 문제점을 찾아냄



1. 변수 설정 및 목적에 맞는 모델 선정

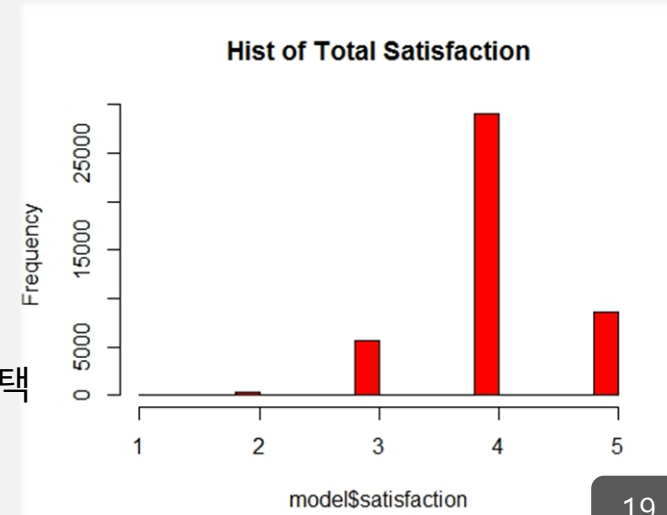
- 국민여행 실태조사 데이터 중 변수 q6.6(전반적 만족도)를 Target Variable로 설정
- 변수 q12.1~ q12.12(항목별 만족도)를 이용하여 Linear Modeling

2. Why Linear Modeling??

- 각 항목별 만족도 변수들이 전반적 만족도에 어느 정도 영향을 미치는지 알 수 있다
- 이점을 이용하여 지역 별, 변수 별 분석을 시행할 수 있다.

3. 어떤 Linear Model을 사용할까??

- 전반적 만족도 변수 q6.6의 특성을 파악
- 상대적으로 4에 치우쳐있고 1,2는 거의 없음
'여행'이라는 변수의 특성을 고려해보면 만족도
1,2,3,4: 성공적 여행이라 보기 어려움
5: 성공적 여행으로 생각
- 전반적 만족도 0~4를 class0, 5를 class1로 설정
- Binary Target 변수의 분석에 적합한 Logistic Regression의 방법을 채택



모델링 Process

- 모집단이 class0 35034개, class1 8504개로 그 비율이 치우쳐있음
→Class0에서 8504개의 변수를 임의추출을 하여 클래스 별로 데이터의 크기 차이가 없도록 맞춰줌
- 무응답(잘 모르겠음=9)의 경우 그 값을 0으로 설정해준 후 응답 변수들을 평균이 0이 되도록 처리
(Mean-centered Data)한 뒤에 분석을 시작
→무응답(데이터에서 9)한 경우 결과에 영향을 미치지 않도록 하기 위해 무응답=0 으로 설정
→평균을 0으로 설정하여 무응답 데이터가 있더라도 영향을 미치지 않도록
- "R" 의 내장함수 glm을 이용하여 Logistic Regression
→class0를 8504개씩 랜덤 추출하여 모델링 하는 과정을 100회씩 수행 후 평균 정확도를 계산

모델링 Process

분석1

- ▶ 모든 변수들 사용
- ▶ 평균 예측률: 0.7172

분석2

- ▶ 응답률 낮고 분석1의 결과에서 가중치(coefficient)가 작게 나온 변수 제거
(쇼핑 만족도, 문화유적 만족도)
- ▶ 평균 예측률: 0.7177

분석3

- ▶ 분석2의 결과에서 가중치(coefficient)가 작게 나온 변수 제거
(교통 만족도)
- ▶ 평균 예측률: 0.7174

회귀분석 모델

$$\text{Logit}(y) = \text{Constant} + c1 * X1 + c2 * X2 + c3 * X3 + c4 * X4 + c5 * X5 + c6 * X6 + c7 * X7 + c8 * X8 + c9 * X9$$

변수명	X1	X2	X3	X4	X5
실제의미	자연경관	교통	숙박	음식	안내시설
가중치(c)	1.22	0.09	0.14	0.24	0.28

변수명	X6	X7	X8	X9	Constant
실제의미	편의시설	친절성	프로그램	물가	Modeling
가중치(c)	0.18	0.24	0.30	-0.11	-0.30

* *used mean – centered data*

* Y값이 0 이상일 경우 1, 0 이하일 경우 0으로 분류

주의해야 할 변수와 특이사항

X1. 자연경관

- ▶ 가장 큰 영향을 미치는 변수

X8. 체험프로그램

- ▶ 높지 않은 응답률(30%정도 응답)에도 높은 가중치를 가짐
- ▶ 특정 지역들에 몰려있는 경향

X9. 물가

- ▶ 다른 변수들의 가중치는 (+)부호를 가지는 반면 물가에 대한 가중치는 -0.11으로 물가의 만족도가 증가할수록 여행의 만족도는 감소하는 것을 의미
- ▶ 이는 사람이 많이 몰리는 관광지 일수록 물가가 비싸고 만족도가 높은 경향을 보이는 것 때문으로 분석
- ▶ 그 외에도 숙박시설, 편의시설 등의 만족도에 복합적으로 '물가'라는 개념이 포함되어 있을 것이라 추정

정확도 개선

→ 정확도의 개선을 위해 중복되는 개념 삭제 및
뜻이 명료하지 않은 변수 세분화

X1. 자연경관

- ▶ 자연 경관이라는 변수의 의미를 보다 정확하게
 - 국립공원 등 자연경관의 관리 및 보존 상태
 - 자연경관의 아름다움 등으로 변수의 의미를 보다 잘 알 수 있도록 세분화

X9. 물가

- ▶ 물가 변수 삭제
 - 다른 모든 소비활동이 포함된 변수들은(음식, 숙박, 편의시설 등)
물가 라는 개념을 잠재적으로 포함하고 있음.
 - 따라서 물가 라는 변수 자체는 제거 하되
각 소비활동의 만족도 향을 상세하게 분류.
- Ex) 숙박시설- 비용대비 만족도, 숙박시설의 접근 편이 (예약 등) 2가지로 나눔

만족도 분석을 통한 국내관광지 체질개선 방안 수립 - 분석

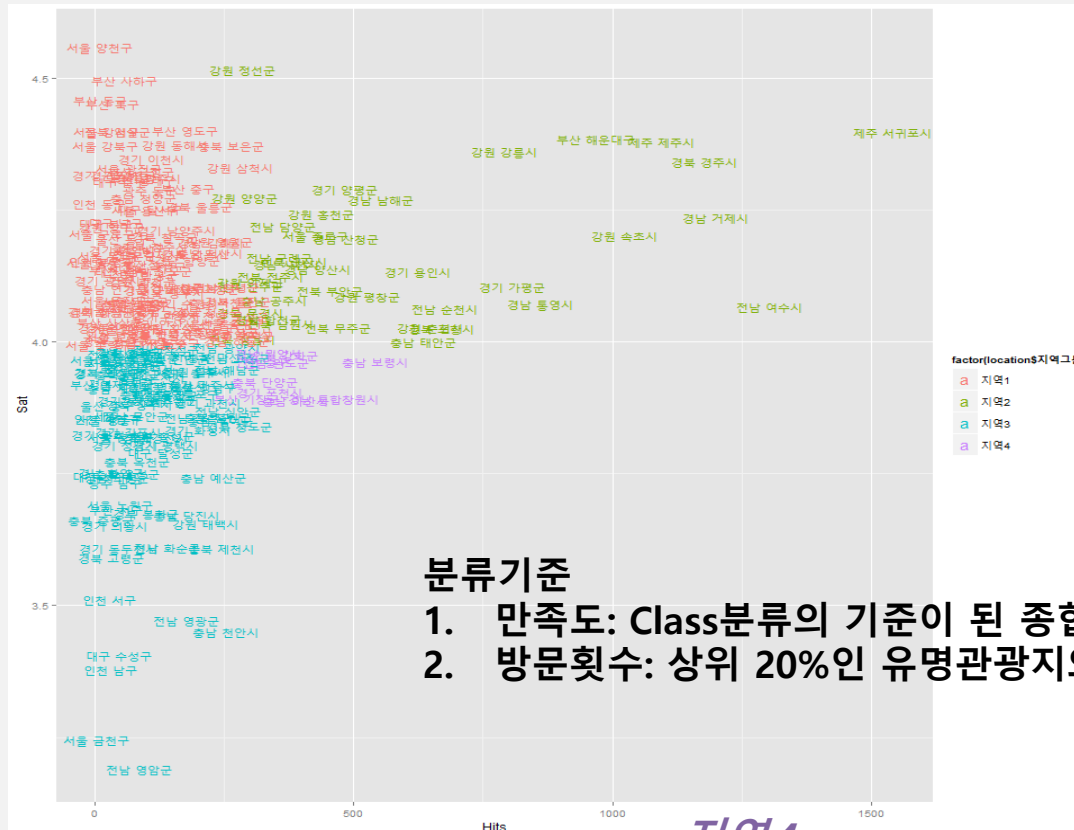
Grouping

지역1

만족도는 높지만 잘 알려지지 않은 집단

지역2

만족도 높으며 잘 알려진 집단



분류기준

1. 만족도: Class분류의 기준이 된 종합만족도 4
2. 방문횟수: 상위 20%인 유명관광지와 그 외 지역으로 분류

지역3

만족도 낮으며 인기도 낮음

지역4

잘 알려졌지만 만족도는 상대적 낮은 집단

집단별 특성에 맞춘 관광 체질 개선 전략 선택!!

만족도 분석을 통한 국내관광지 체질개선 방안 수립 - 분석

Grouping: 그룹별 지역들(방문횟수 150회 이하 지역들 생략)

지역1: 만족도 ↑ 방문자 ↓

강원 동해시	경남 김해시	경북 안동시	부산 중구	전남 보성군	충남 서천군
강원 삼척시	경남 진주시	경북 영덕군	서울 송파구	전북 고창군	충북 보은군
강원 영월군	경남 창녕군	경북 울릉군	서울 중구	전북 순창군	부산 영도구
경기 남양주시	경남 하동군	경북 울진군	울산 울주군	충남 금산군	
경기 수원시	경남 함양군	경북 청송군	전남 목포시	충남 서산시	

지역2: 만족도 ↑ 방문자 ↑

강원 강릉시	강원 정선군	경기 가평군	경북 포항시	전남 담양군	전북 전주시	경남 남해군	경남 통영시
강원 속초시	강원 춘천시	경기 양평군	부산 해운대구	전남 순천시	전북 정읍시	경남 사천시	경남 합천군
강원 양양군	강원 평창군	경기 용인시	서울 종로구	전남 여수시	제주 서귀포시	경남 산청군	경북 경주시
강원 인제군	강원 홍천군	경남 거제시	전남 구례군	전북 군산시	제주 제주시	경남 양산시	경북 문경시

지역3: 만족도 ↓ 방문자 ↓

강원 원주시	경기 화성시	전남 고흥군	전남 장흥군	충남 부여군
강원 태백시	경북 청도군	전남 광양시	전남 해남군	충남 예산군
경기 과천시	서울 강남구	전남 신안군	전북 진안군	충북 청원군
경기 파주시	서울 영등포구	전남 영광군	충남 당진시	충북 충주시

지역4: 만족도 ↓ 방문자 ↑

경기 포천시	부산 기장군	전남 완도군	충북 단양군	인천 중구
경남 밀양시	인천 강화군	충남 보령시	경남 통합창원시	충남 아산시

만족도 분석을 통한 국내관광지 체질개선 방안 수립 - 결론

지역1 : 발전잠재력이 높은 미래 관광지 찾기

- X1 변수는 가장 큰 가중치를 가지는 관광지의 만족도가 높은 직접적 원인이며
그 지역만의 고유한 자산으로 생각
- X1변수가 높은 지역들을 발전가능성이 높은 집단으로 생각하여 각 도별로 1개의 지역들을 선정
- 표본 수가 100 이하인 지역은 제외(만족도 데이터의 신뢰도 차원)

	경기 연천	강원 횡성	충북 보은	전북 김제	전남 진도	경북 울릉	경남 하동	충남 청양
인기	★★	★★★	★★★ ★	★★	★★★	★★★	★★★	★★★
음식	★★	★★★ ★	★★★ ★	★★★ ★	★★★ ★★	★★★ ★★	★★★ ★	★★★ ★
편의	★	★★★	★★★ ★	★	★★	★★★ ★	★★★ ★	★★★ ★
안내	★	★★★ ★★	★★★ ★★	★★	★	★★★ ★	★★★ ★	★★★ ★

★ : 0 ~20%
 ★ ★ : 20~40%
 ★ ★ ★ : 40~60%
 ★ ★ ★ ★ : 60~80%
 ★ ★ ★ ★ ★ : 80~100%
 %는 전국 분위기를 의미

*group1의 x1변수 값이 높은 각 도별 관광지

만족도 분석을 통한 국내관광지 체질개선- 분석

지역1 : 발전잠재력이 높은 미래 관광지 찾기

전남 진도

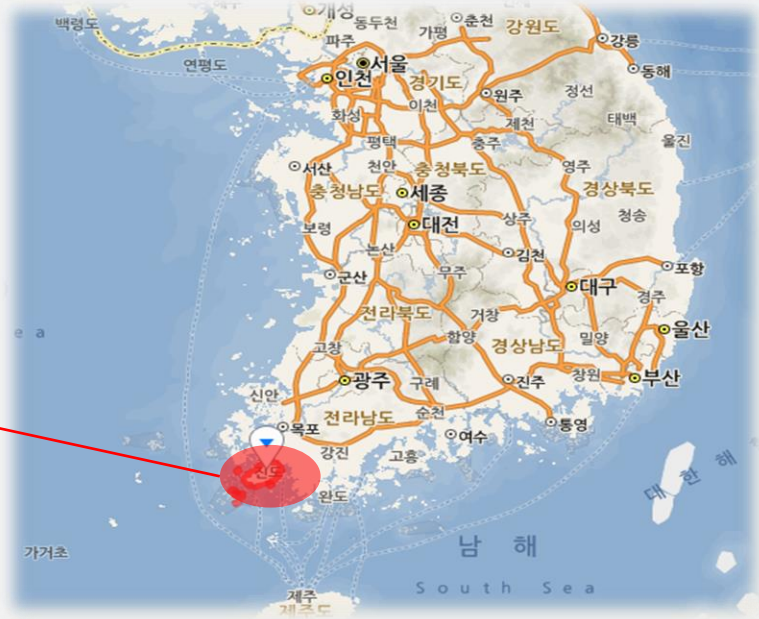
인기: ★ ★ ★

음식: ★ ★ ★ ★ ★

편의: ★ ★

안내: ★

- ▶ 훌륭한 자연경관을 가진 섬
- ▶ 평가 좋은 향토음식
- ▶ 편의시설, 관광안내 등의 관광 인프라가 전반적으로 부족



관광안내 센터, 숙박업체 설립 지원 등 관광 인프라에 투자

인기가 상대적으로 적은 지역을 사람들에게 여행지로서 인식하도록 지역 광고 등

‘섬’이라는 점을 활용한 문화체험 프로그램등 관광자원 개발

만족도 분석을 통한 국내관광지 체질개선 방안 수립 -결론

지역2 : 대한민국 관광 명소들! 더욱 완벽하게!

- 각 변수들에 대해서 가장 만족도를 저해시키는 지역별 원인 탐색 및 조사 하여 문제점 확인
- 만족도에 가중치를 곱하고 평균값을 빼주어 정량적으로 계산

자연경관	
전북 전주	-0.72
경남 합천	-0.22
충남 태안	-0.19
경기 양평	-0.11

교통	
전북 전주	-0.05
강원 강릉	-0.01
전남 여수	-0.01
경북 문경	0

숙박	
전북 전주	-0.11
경북 문경	-0.05
전북 부안	-0.05
경기 용인	-0.04

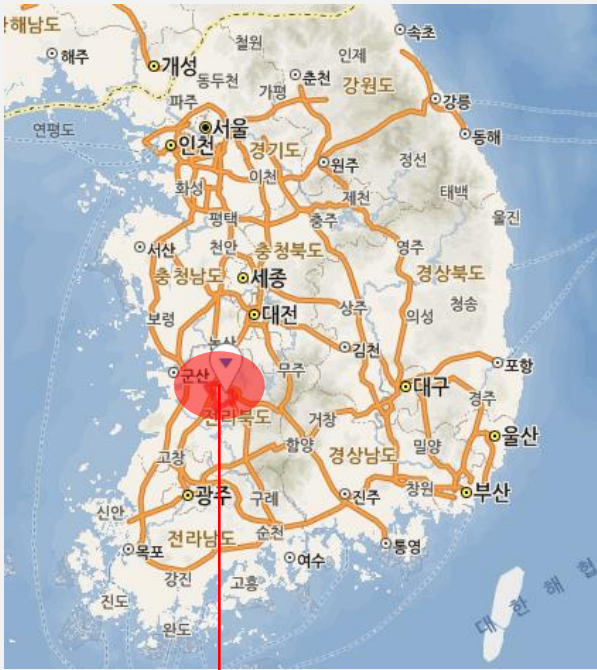
음식	
경북 경주	-0.11
전남 담양	-0.10
경남 통영	-0.09
전북 부안	-0.08

편의시설	
전남 담양	-0.09
강원 평창	-0.05
충남 태안	-0.02
경기 가평	-0.01

시설친절도	
전북 전주	-0.08
전북 부안	-0.08
전북 전주	-0.07
경북 경주	-0.04

프로그램	
전북 전주	-0.26
전북 부안	-0.20
전북 정읍	-0.13
강원 평창	-0.11

안내시설	
전남 담양	-0.19
전북 전주	-0.13
전북 부안	-0.13
경북 포항	-0.07



- ▶ 안내시설과 시설의 친절도 개선필요
- ▶ 문화유적, 공원, 공공기물 등 정비
- ▶ 체험 프로그램, 이벤트의 질적 향상

*Group2의 만족도 변수에 평균값을 빼준 후 가중치를 곱한값들

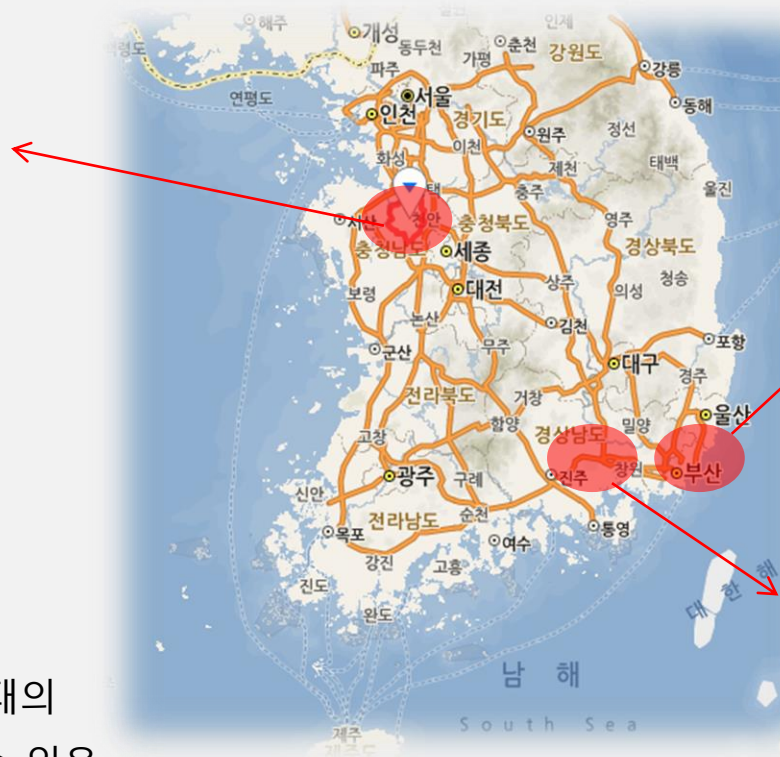
만족도 분석을 통한 국내관광지 체질개선 방안 수립 - 결론

지역4 : 아쉬운 인기 관광지들

- 사람들이 많이 찾는 유명 관광지임에도 불구하고 만족도가 상대적으로 낮음
- 각 지역별로 문제점이 하나씩은 존재할 것으로 생각
- 지역별로 분석하여 문제점들 찾기

충남 아산시	
프로그램	-0.27
자연경관	-0.26
음식	-0.12

부산 기장군	
안내시설	-0.08
음식	-0.06
편의시설	-0.05



경남 통합창원시	
자연경관	-0.11
안내시설	-0.05
안내시설친절	-0.04

- ▶ 아산시의 관광구역에 대해 검색
: 지중해마을, 세계 꽃 박물관, 함상공원, 민속마을 등
- ▶ 관광 콘텐츠 들의 만족도가 상대적으로 낮음.
: 입장료가 지나치게 비싼지, 관리상태의 문제, 내용의 부실 등 문제가 있을 수 있음

**Group4내에서도
상대적 만족도가 낮은 3지역*

지역1 만족도는 높지만 잘 알려지지 않은 지역

- ▶ 관광잠재력이 높은 지역을 찾고 부족한 부분을 개선

지역2 만족도 높고 잘 알려진 관광지

- ▶ 집단 내 랭킹 시스템을 통해 부족한 지역의 항목들을 개선

지역3 만족도 낮고 잘 알려지지 않은 지역

- ▶ 관광 잠재력 낮고 잘 알려지지 않음. 다른 지역 개발 모색

지역4 잘 알려졌지만 만족도 낮은 관광지

- ▶ 관광지 별로 문제가 있을 것으로 추정
- ▶ 각 지역별만족도 랭킹을 통해 만족도 저해요인을 찾고 개선



Thank You