

# 관광객 2000만 시대? 지방 관광객을 유치하라!

외래 관광객 실태 데이터 마이닝 분석 결과

팀명 SA 141



# 관광객 2000만 시대, “지방 관광객” 유치가 관건

“양적 성장보다 질적 성장이 우선이다”

## I. 과제 파악

“프로젝트 FRAMEWORK 파악”

## II. 사전 조사

## III. 분석 방향 설정

## IV. 기본분석

## V. 추가분석

## VI. 최종 관광마케팅 방안

## VII. 결론 및 제언



# 1. FRAMEWORK 개괄

## I. 과제 파악 1.FRAMEWORK

### 주어진 데이터

2009, 2010, 2011년  
외래관광객 설문조사

2009년도 외래관광객 설문조사 설문지

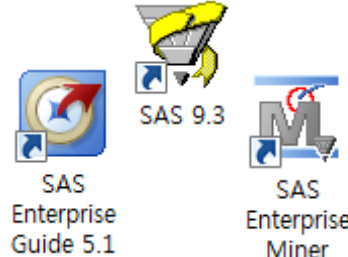
2010년도 외래관광객 설문조사 설문지

2011년 외래관광객 설문조사

### 분석 과제

- 기본 분석-
- 1. 목표변수 정의
- 2. 예측모형 구축

+추가분석



### 최종 목표

분석결과 ⇒ 창의적 전략  
“관광산업의 발전 유도”



수행방향

- ✓ ‘관광 활성화’를 설명할 수 있는 목표변수
- ✓ 주어진 데이터 안에서 분석
- ✓ Predictive modeling: 어떻게 각 요소들이 연결되는지를 파악해야 함

## 2. 설문 데이터에 대한 이해

주어진 데이터의 체계화 필요: 2가지 기준으로 분류

	방문 전	방문 중	방문 후
DEMO-GRAPHIC	성별, 나이, 학력, 직업, 거주도시, 거주국, 소득	(해당 없음)	(해당 없음)
행동	타국 방문여부 방문 동반자 여행형태 한국방문횟수 방문 목적 여행정보 입수경로	주된 활동 쇼핑품목 쇼핑장소 주 이용 숙박시설 방문지	UNKNOWN
태도	여행동기 여행 전 방문 검토 국가	좋았던 방문지 좋았던 점 불편했던 점	항목별 만족도 전반적 만족도 3년 내 재방문 의사 한국 여행 추천의사 이미지 변화



영향 및 관계



최종적으로 알고 싶은 것

# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

**II. 사전 조사**

“분석 테마와 방향 설정을 위한  
Brainstorming 과정”

III. 분석 방향 설정

IV. 기본분석

V. 추가분석

VI. 최종 관광마케팅 방안

VII. 결론 및 제언

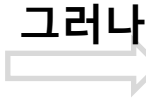


# 1. 문제 인식 과정 : “왜 지방인가”

## WHAT

무엇이  
문제인가?

정부  
2020년  
외국인 관광객  
2000만명 목표



한국 관광산업,  
양적 향상보다  
**질적 측면의 개선**  
이 우선시됨

- 관광객 만족도
- 재방문율  
⇒ 4년 최저치 갱신 (2012)

## WHY

왜  
문제인가?

서울  
과도한 편향



지방 관광객 유치의 초보적 대응수준

1. 구체적이지 않은 전략
2. 대안의 부재
3. 모호한 TARGET

입력: 20121214 16:28 / 수정: 20121214 16:28

'서울을 벗어나야, 관광객 2000만명 시대 열 수 있다'

'부·울·경 방문의 해' 등 국내관광 활성화

해 관광공사는 이같은 숙박시설 부족 문제를 해결하기 위해 관광인프라편드를 조성, 숙박시설을 확대해 나갈 계획이다.

## HOW

어떻게  
해결하나?

“지방관광객 유치 성공요소” 탐색이 필요

- 이를 위해 기존 데이터를 바탕으로 한 예측모형 구축  
(⇒기본분석 과제)

# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

II. 사전 조사

**III. 분석 방향 설정**

“선정된 주제를 바탕으로  
기본 분석 방향 설정”

IV. 기본분석

V. 추가분석

VI. 최종 관광마케팅 방안

VII. 결론 및 제언



- 기본 분석 방향 설정

### 1. 집단 분리

알고 싶은 것: '지방 방문 관광객'과 그렇지 않은 관광객의 차이

집단 분리 후 개별 모델링을하기로 결정

- 서울 지방 모두 방문, 서울만 방문, 지방만 방문
- 각 집단 간의 차이점, 장단점  
⇒ 지방 관광 활성화 방안

### 2. 예측 vs 설명

- 주어진 데이터에서 "예측 요소"를 찾기 어려움
- Target을 "설명"하는 변수에 대한 탐색  
(Descriptive Modeling)

### 3. 모델링 방법

- '설명'의 용이함을 위해 2가지 모델링 방법만 사용
  - 회귀분석
  - 의사결정나무



# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

II. 사전 조사

III. 분석 방향 설정

**IV. 기본분석**

1. 사전처리
2. SEMMA
3. 결과해석
4. 추가분석진행방향

V. 추가분석

VI. 최종 관광마케팅 방안

VII. 결론 및 제언



# 1. 사전처리 \_원자료 가공 작업

원자료를 분석용 데이터로 만들기 위한 사전작업 수행



### <요약>

'Sample' 단계에서 필요한 4가지 작업을 수행

#### 1. 목표 변수 설정 (Setting the Target Variable)

한국 여행에 대한 전반적 '불'만족도

#### 2. 데이터 선택 (Data Selection for Modeling)

2009~2011 자료를 **통합**,  
서울 및 지방 방문여부에 따라 3개 집단으로 **분리**

#### 3. 데이터 분할 (Data Partition)

Train : Validate : Test = 40 : 30 : 30

#### 4. 과대표본추출 (Oversampling)

Target : Non-Target = 50 : 50

### 1. 목표 변수 설정

(Setting the Target Variable)

한국 여행에 대한 전반적 '불'만족도

### 후보 선정

- ① 한국 재방문 의사(문항19, 2009)
- ② 관광지로서 한국을 추천 여부 (문항20, 2009)
- ③ 한국에 대한 이미지 변화 (문항21, 2009)
- ④ 한국에 대한 전반적 만족도 (채택) (문항 17, 2009)

Y= 전반적으로 불만족



①매우 불만족 ②불만족 ③보통 ④만족 ⑤매우 만족

#### • 왜 '불만족도'?

- 응답자 중 '④만족', '⑤매우 만족'의 빈도수가 60% 육박  
⇒ 면접조사, 설문조사라는 특성
- '문제점'을 파악하려면 '불만족'한 사람들에게서 원인을 찾아야 함  
⇒ 만족한 사람들만 주시할 경우, 특이점을 찾기가 힘들 것으로 예상

## 2. SEMMA \_(1) Sample \_1) 데이터 선택

### 2. 데이터 선택

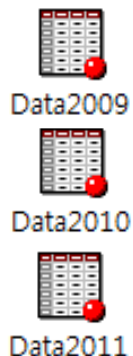
(Data Selection for Modeling)

2009~2011 자료를 **통합**,  
서울 및 지방 방문여부에 따라 3개 집단으로 **분리**

### 데이터 통합 (Data Integration)

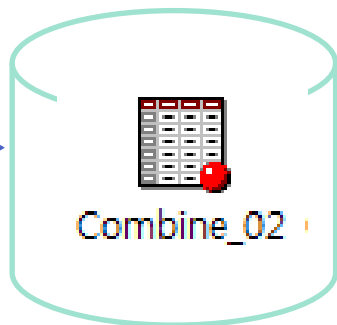
- 3년간의 전반적인 경향
- 2009, 2010, 2011년 데이터 통합
- 설문지에서 공통된 항목만 반영

원자료



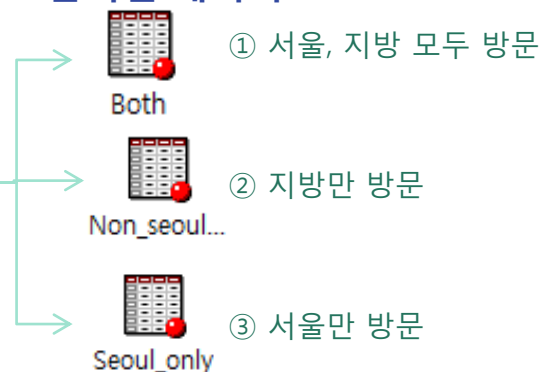
사전처리

통합된 데이터



Segmentation

분리된 데이터

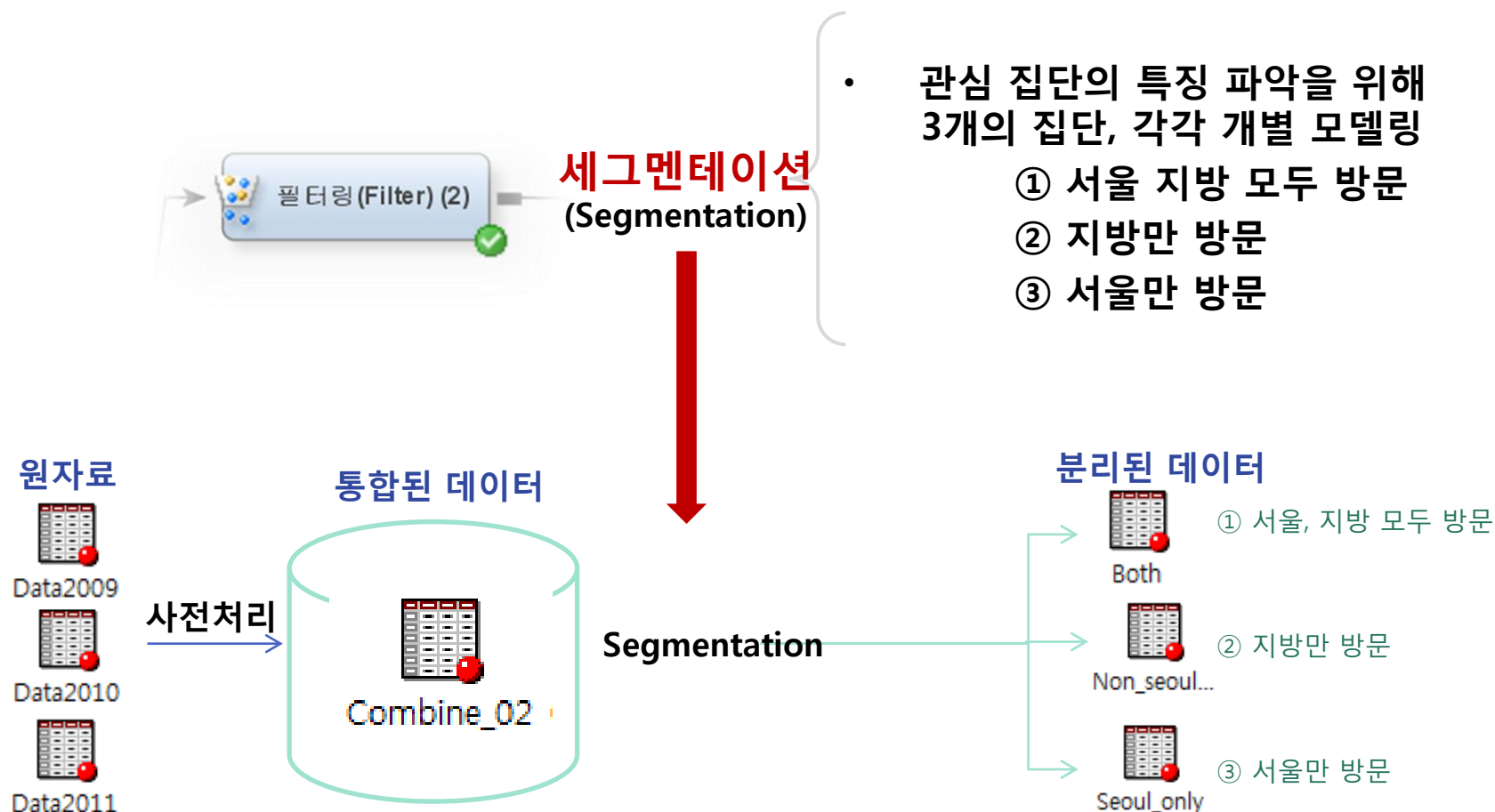


## 2. SEMMA \_(1) Sample \_1) 데이터 선택

### 2. 데이터 선택

(Data Selection for Modeling)

2009~2011 자료를 **통합**,  
서울 및 지방 방문여부에 따라 3개 집단으로 **분리**



## 2. SEMMA \_(1) Sample \_1) 데이터 선택

### 2. 데이터 선택

(Data Selection for Modeling)

2009~2011 자료를 **통합**,  
서울 및 지방 방문여부에 따라 3개 집단으로 **분리**

- **목적:** “집단을 Segment하기 전, 목적을 우선 설정해야 함”

<서울, 지방 **모두** 방문 관광객>  
우리가 점점 숫자가 많아졌으면 하는 집단

“이들의 (불)만족 요인을 찾아라”

<**지방만** 방문 관광객>  
지방 관광객 유치 부진의 원인을  
보여주는 열쇠

“지방이 개선해야 할 점”

<**서울만** 방문 관광객>  
지방이 수도권으로부터 가질 수 있는  
비교우위는?

“서울의 결핍요소를 지방에서 충족하도록”

### 2. 데이터 선택

(Data Selection for Modeling)

2009~2011 자료를 **통합**,  
서울 및 지방 방문여부에 따라 3개 집단으로 **분리**

#### • 근거: 왜 **그룹화**, 왜 **개별 모델링**을 하는가?

- 동질성 검정 (카이제곱) 수행
- 세 집단간 목표변수(전반적 (불)만족도)에 대한 분포가 다름

통계량	자유도	값	Prob
카이제곱	10	348.4453	<.0001
우도비 카이제곱	10	345.8387	<.0001
Mantel-Haenszel 카이제곱	1	70.1503	<.0001
파이 계수		0.1008	
우발성 계수		0.1003	
크래머의 V		0.0713	

P-value가 유의수준(0.05)보다 작음  
귀무가설인 '집단 간 동질성'을 기각



### 3. 데이터 분할 (Data Partition)

- Train : Validate : Test = 40 : 30 : 30

RAW DATA (100%)



**학습 데이터 (Train) 40%**  
모델 구축시 사용

**검증 데이터 (Validate) 30%**  
모델 성능 비교시

**평가 데이터 (Test) 30%**  
새로운 데이터 대한 성능

분할 근거

- default setting 준수
- 표본 수가 충분함
- 모델 적합 결과의 당위성을 확보

### 4. 과대표본추출 (Oversampling)

Target : Non-Target = 50 : 50

### Target의 비율: 0.54%

- 모델 예측력이 낮아질 것으로 예상  
⇒ Train Data에 과대표본추출을 적용

### “Rule of Thumb(엄지의 법칙)”

Target 변수 비율 및 관측치 숫자에 따라-

- 5%미만, 10,000 관측치 이상: 50/50 (해당케이스)

### <요약>

#### 1. 변수간 관계

(Relationships Between the Variables)

- Target과 Input간의 관계 비교
- Input들과 Target에 관한 Histogram과 빈도표 작성

#### 2. 결측치, 이상값

(Abnormalities)

- 결측치에 대한 특별한 조치를 취하지 않음
- 그외/기타/무응답: 분석에서 제외

#### 3. 데이터 시각화

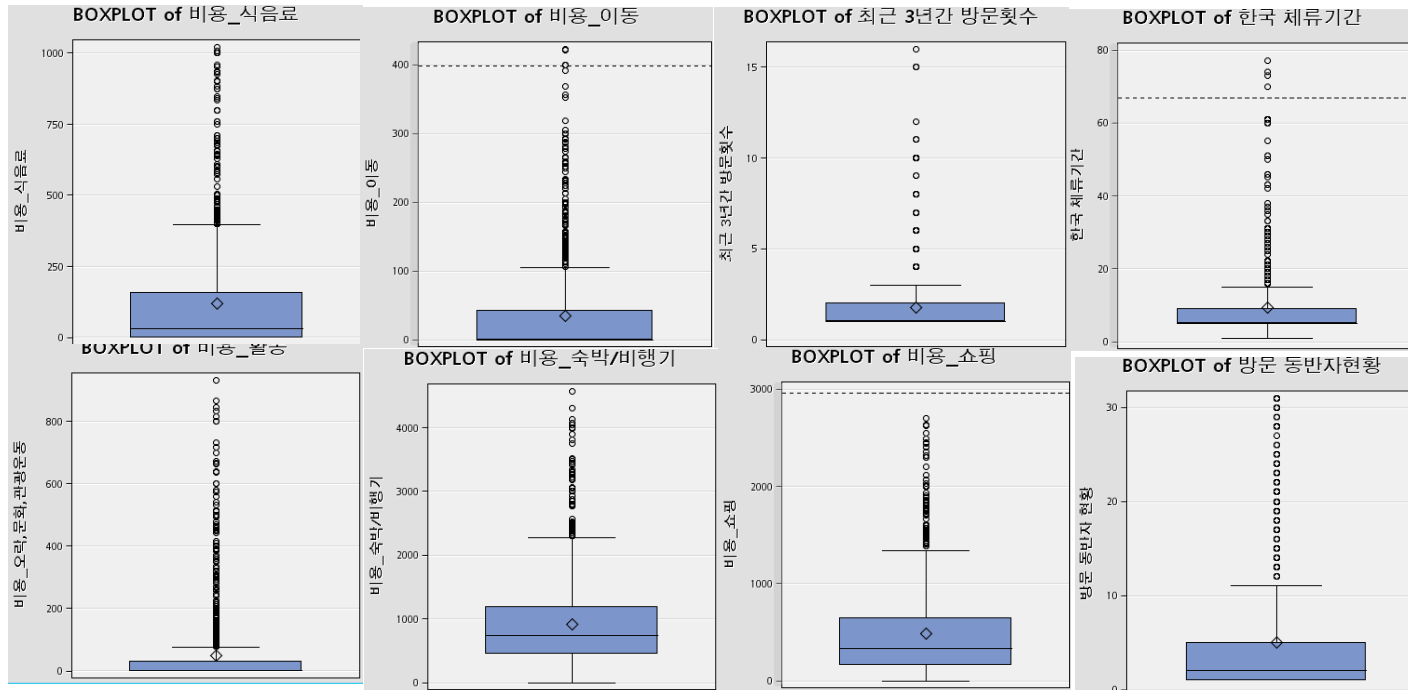
(Data Visualization)

- 연속형 변수: 히스토그램, Boxplot
- 범주형 변수: 히스토그램, 빈도표

### 3. 데이터 시각화 (Data Visualization)

- 연속형 변수: 히스토그램, Boxplot
- 범주형 변수: 히스토그램, 빈도표

Interval 변수가 대체로 다수의 outlier와 쓸림 현상이 있음.



### <요약>

#### 1. 변수 변형

(Variable Modification)

- 연속형 변수를 범주화 변수화
- 범주형 변수의 재범주화

#### 2. 변수 선택

(Variable Selection)

- 필요한 변수만 선택
  - (1차) 자료에 대한 사전 지식
  - (2차) 통계적 탐색
- 변수 재범주화

### 2. 변수 선택

(Variable Selection)

- 필요한 변수만 선택
- 변수 재범주화

#### 1. 사전지식을 통한 임의 선정

- 해석의 용이함, 관심 있는 변수만 고려
- 최종 선정 변수 내역

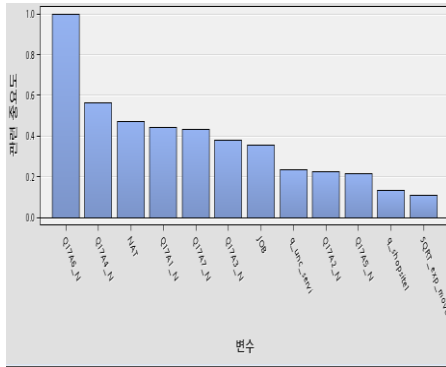
	방문 전	방문 중	방문 후
DEMO-GRAPHIC	성별, 나이, 학력, 직업	(해당 없음)	(해당 없음)
행동	방문 동반자 수 여행형태 방문 목적 여행정보경로 1순위	주된 활동 1순위 쇼핑 품목 1순위 쇼핑 장소 1순위 지출금액	(해당 없음)
태도	여행동기	좋았던 방문지 1순위 좋았던 점 불편했던 점	항목별 만족도 <b>전반적 (불)만족도</b> (Target)

## 2. 통계적 근거로 선정

- Explore 단계의 데이터 시각화 과정에서 판단하여 임의 제거
- 변수 선택 노드를 사용하여 자동적으로 변수 선택 OPTION을 최종적으로 사용

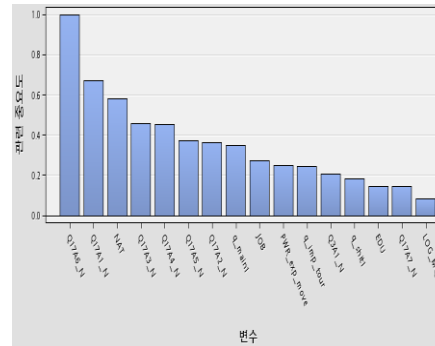
### 변수선택 내역

#### 1. 둘 다 방문



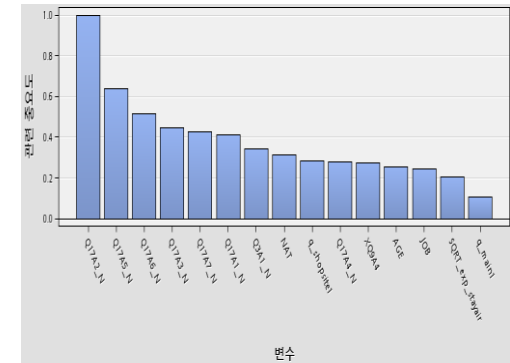
- 항목별 만족도
- 직업, 국적
- 여행비용\_이동
- 쇼핑장소 1위
- 불편했던 점\_서비스

#### 2. 서울만 방문



- 항목별 만족도
- 교육수준, 직업, 국적, 동반자 수
- 여행비용\_이동
- 쇼핑장소 1위
- 여행정보 입수경로
- 인상깊었던 점\_볼거리/관광요소
- 주된활동 1위
- 쇼핑아이템 1위

#### 3. 지방만 방문



- 나이, 직업, 국적
- 항목별 만족도
- 여행비용\_숙박/비행기
- 동반자 현황
- 주된활동1
- 주된 쇼핑장소

### 모델링 방식 선택

(Modeling Method)

#### 로지스틱 회귀 분석

(Logistic Regression)

- Target을 불만족(=1), 만족(=0)으로 설정
- 교호작용 인정, 절편 포함 고정 옵션
- 총 6가지 방법으로 적합
  - 모델선택방식(전진, 후진, 단계별)
  - 링크(Logit, Probit)

#### 의사결정나무

(Decision Making Tree)

- Target을 불만족(=1), 만족(=0)으로 설정
- Interval 기준: ProbF로 고정
- Nominal 기준:  
ProbChisq, Entropy, Gini



### 최종 선택 모델 (Final Selection)

세 집단 모두 회귀모형이 채택되었으며, 상세 내역은 다음과 같음

선정 기준: 가장 우선적으로 Train 데이터의 RoC 인덱스 고려  
(오분류율, 반응률, Test 데이터의 오분류율 등을 부차적으로 고려)

해당 기준을 바탕으로 각 segment 당 1개의 모델 채택

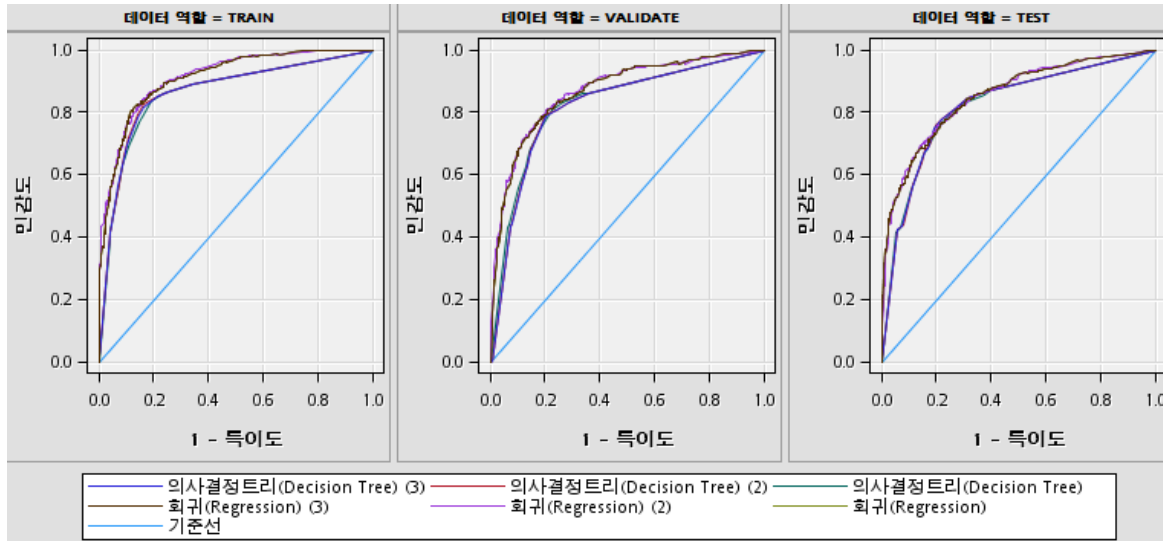
### 채택결과

1. 서울과 지방 둘 다 간 경우 (VISIT=1)  
(Logit & Stepwise)
2. 서울만 간 경우 (VISIT=2)  
(Probit & Stepwise)
3. 지방만 간 경우 (VISIT=3)  
(Probit & Backward)

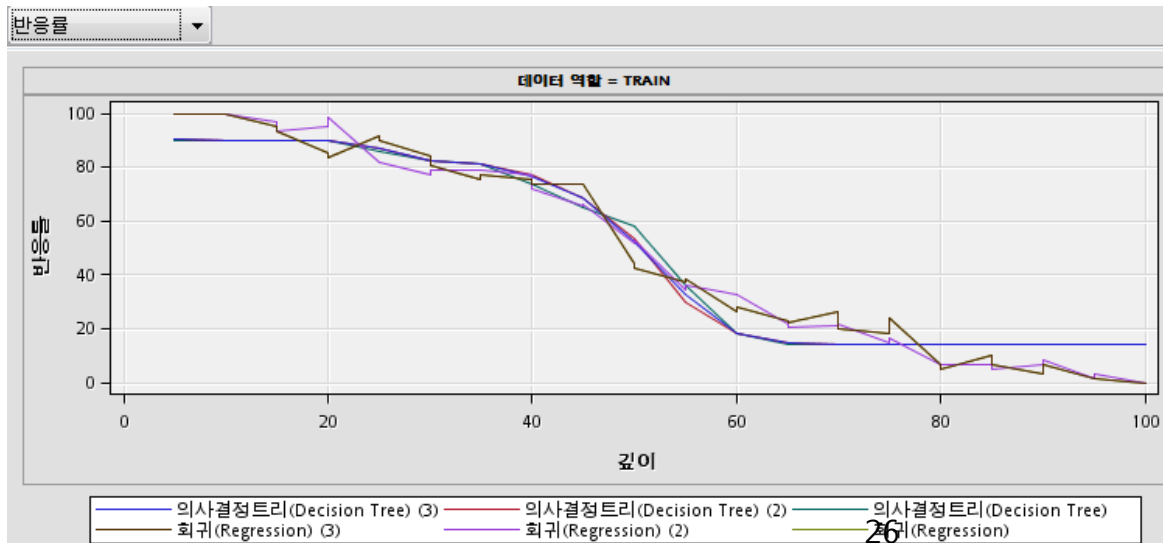
## 2. SEMMA\_(5) Assess

### 1. 서울 지방 둘 다 간 경우

#### <그래프탐색>



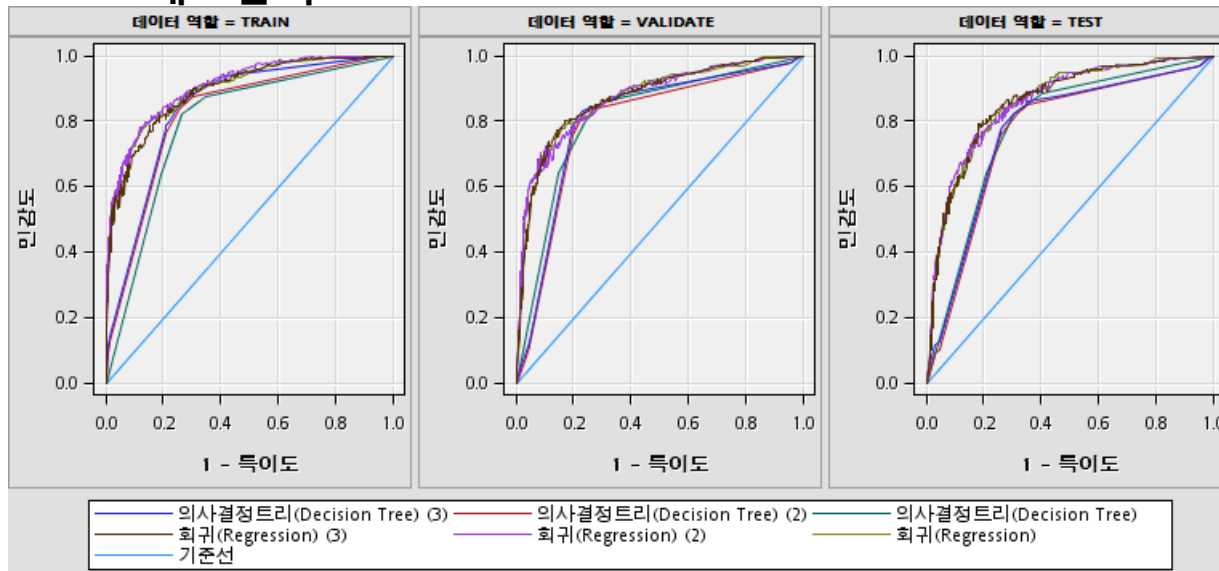
⇒ ROC 커브의 민감도가 전반적으로 우수,  
채택된 회귀모형이 가장 우수한 민감도



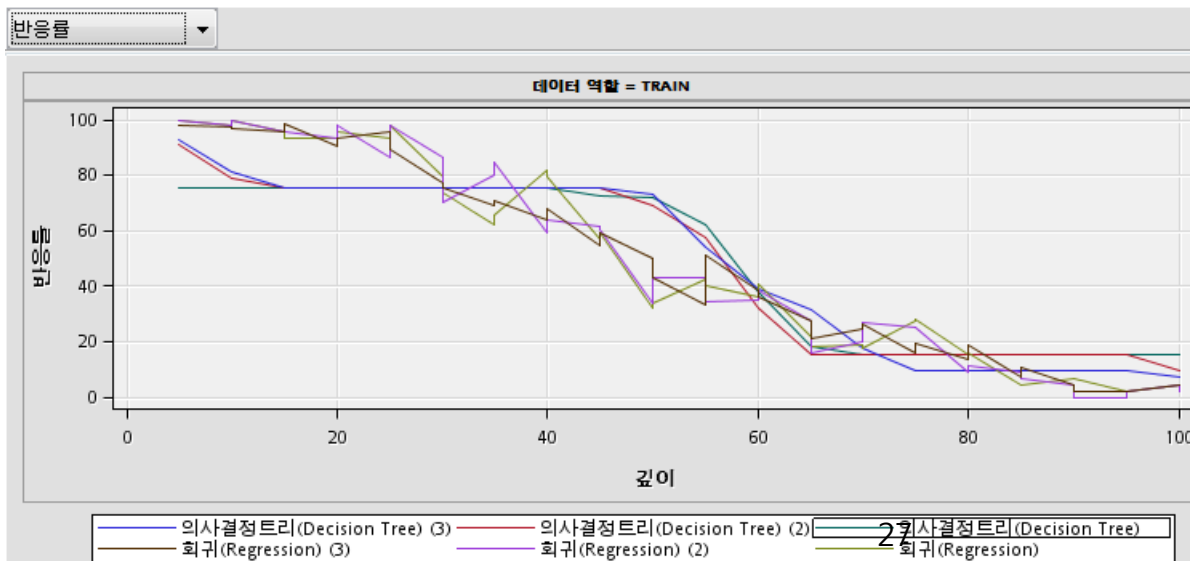
⇒ 20%의 수준에서 거의 100%의 반응률

⇒ 곡선이 울퉁불퉁한 이유:  
CLASS VARIABLE이 다수

## 2. 서울만 간 경우 <그래프탐색>



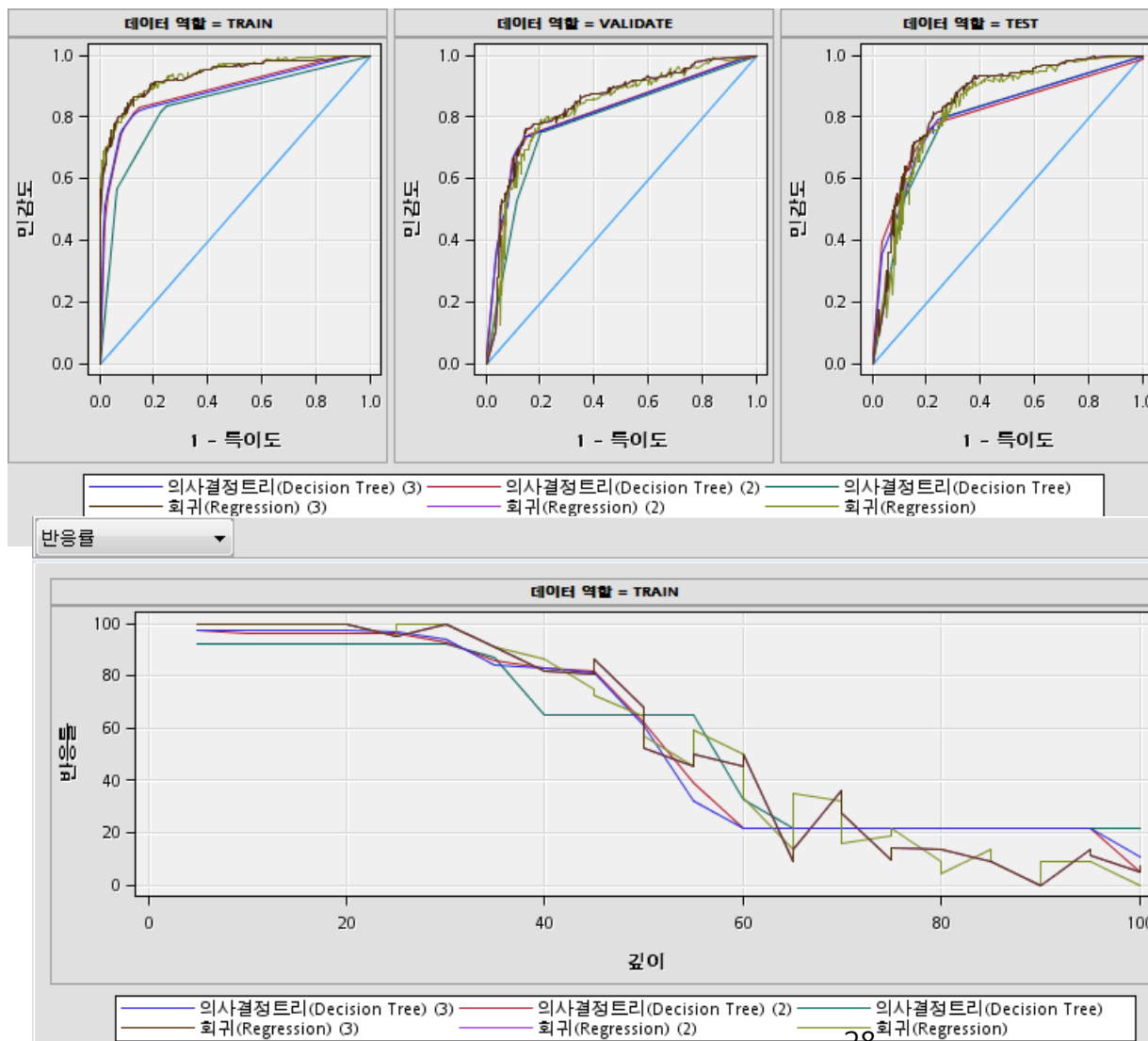
⇒ ROC 커브의 민감도가 전반적으로 우수,  
채택된 회귀모형이 가장 우수한 민감도



⇒ 20%의 수준에서 거의 100%의 반응률

⇒ 곡선이 울퉁불퉁한 이유:  
CLASS VARIABLE이 다수

## 3. 지방만 간 경우 <그래프탐색>



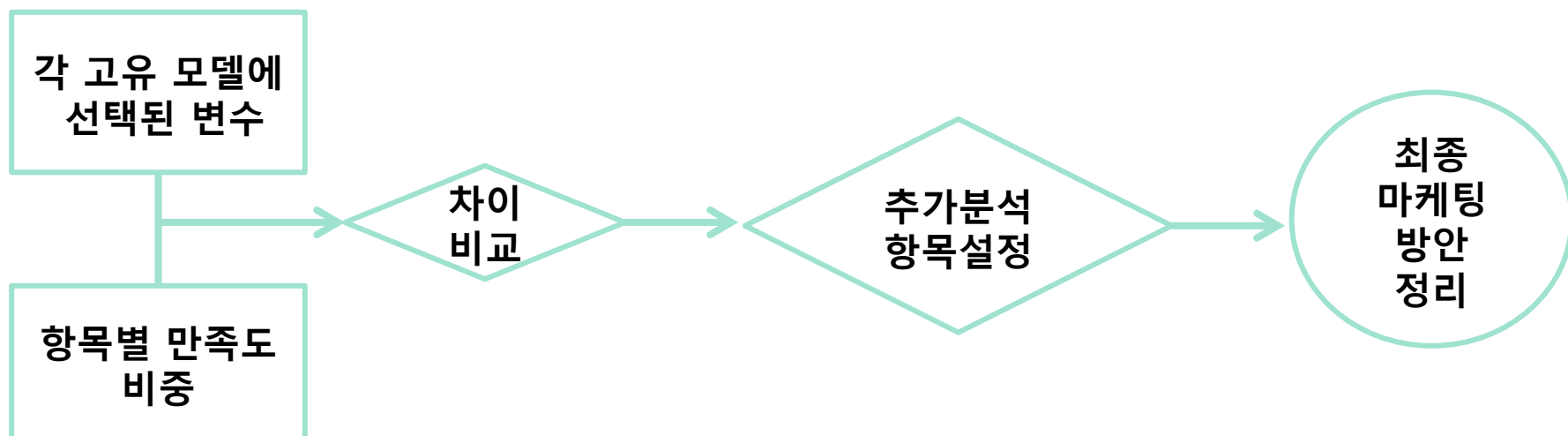
⇒ ROC 커브의 민감도가 전반적으로 우수,  
채택된 회귀모형이 가장 우수한 민감도

⇒ 20%의 수준에서 거의 100%의 반응률

⇒ 곡선이 울퉁불퉁한 이유:  
CLASS VARIABLE이 다수

- 1 { • 서로 다른 변수들이 각 모델에서 채택  $\Rightarrow$  집단을 셋으로 나눈 효과가 있었다 }
- 2 { • 다변량 회귀분석  $\Rightarrow$  기울기: 해당 변수 고유의 효과를 측정할 수 있다  
• 로지스틱 회귀분석  $\Rightarrow$  직관적인 해석은 어렵지만, 영향도는 파악 가능 }

#### <추가분석 방향 설정>



### 3. 결과해석 \_(1) 서울 지방 모두 방문

#### (1) 서울 지방 모두 방문 집단의 불만족도 설명요인

유의미한 변수	변수 설명	유의한 항목	절편값	해석
NAT	거주국	중국, 싱가포르 대만, 태국	0.8440 0.9845 0.8733 0.6782	아시아권 출신의 관광객이 불만족도에 상대적으로 큰 영향을 미치는 것을 알 수 있다.
Q17A1_N	항목별 만족도 _출입국절차	만족 매우 만족	-0.742 -0.9665	출입국 절차에서 만족하면 불만족과 상반되는 결과를 가져올 수 있다
Q17A4_N	항목별 만족도 _음식	만족 매우 만족	-0.8180 -1.2415	음식에서 만족하면 불만족과 상반되는 결과를 가져올 수 있다
q_shopsite1	주요쇼핑장소1위	면세점	0.4230	불만족을 가져올 수 있는 요소가 있음

### 3. 결과해석 \_(2) 서울만 방문

#### (2) 서울만 방문 집단의 불만족도 설명요인

유의미한 변수	변수 설명	유의한 항목	절편값	해석
JOB	직업	사무기술직 전문직	0.3187 -0.4538	사무기술직군의 관광객, 전문직의 불만족도 영향력
Q17A2_N	항목별 만족도 _대중교통	불만족	0.6823	대중교통에서 불만족이 전반적 불만족 을 가져올 수 있다.
Q17A3_N	항목별 만족도 _숙박	매우 불만족 만족	1.17011 -0.5116	숙박에서 불만족이 전반적 불만족을 가져올 수 있다.
Q17A6_N	항목별 만족도 _관광지 매력도	불만족	1.1067	관광지에서 매력도를 느끼지 못하면 전반적 불만족을 가져올 수 있다.
Q17A7_N	항목별 만족도 _관광안내서비 스	만족 매우 만족	-0.4780 -0.5722	관광안내 서비스에서 만족은 불만족과 상반되는 결과를 가져올 수 있다.

#### (3) 지방만 방문 집단의 불만족도 설명요인

유의미한 변수	변수 설명	유의한 항목	절편값	해석
q_main1	주요활동1위	관광	-1.3638	치료 > 관광 > 놀거리 > 연구/교육/업무 순으로 불만족도와 상반되는 방향으로 영향을 미친다.
		치료	-4.0224	
		연구,교육,업무	-0.9712	
		놀거리	-1.22903	
Q_shopsite	쇼핑장소1위	백화점, 할인마트	0.6823	백화점 및 할인마트가 불만족도에 영향을 줄 수 있다.



### 3. 결과해석 \_마케팅의 실마리 찾기

각 집단 별 모델링 결과 시사점:

#### <서울, 지방 모두 방문 관광객>

-우리가 점점 숫자가 많아졌으면 하는 집단

그들이 인천 국제공항에서 출국할 때,  
음식과 쇼핑에서 만족도가 결정되었다  
“음식, 쇼핑에 대한 항목 강화 필요”

#### <서울만 방문 관광객>

-지방이 수도권으로부터 가질 수 있는 비교우위는?

대중교통, 숙박, 관광지 매력도, 관광안내서  
비스에서 불만을 느꼈던 그들에게  
“지방에서 돌파구를 찾게 하라”

#### <지방만 방문 관광객>

-지방 관광객 유치 부진의 원인을 보여주는 열쇠

그들의 주요 활동을 파악  
관광, 치료, 놀거리에 개선할 점이 있다면  
“당장 개선”

### 3. 결과해석 \_추가분석 진행 방향

#### 1. '지방'은 너무 광범위 하다!

- ⇒ 어느 지방을 어떻게 나눌 것인가?
- ⇒ 단순히 지리적 근접성으로 패키지를 구성하는 것만이 답인가?
- ⇒ 현재 방문 성향 파악 필요

#### 2. 지방관광활성화를 위한 테마 상품 개발

- ⇒ 지방 방문객이 주목하는 '치료'

#### 3. 불만족도 요인 관련 항목: 관심 집단 별 세부 빈도 분석

# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

II. 사전 조사

III. 분석 방향 설정

IV. 기본분석

**V. 추가분석**

“구체적 마케팅 방안을 위한  
추가분석”

VI. 최종 관광마케팅 방안

VII. 결론 및 제언



장바구니 분석(Market Basket Analysis): 연관성 분석을 기반으로 주요 방문지를 파악

조건:

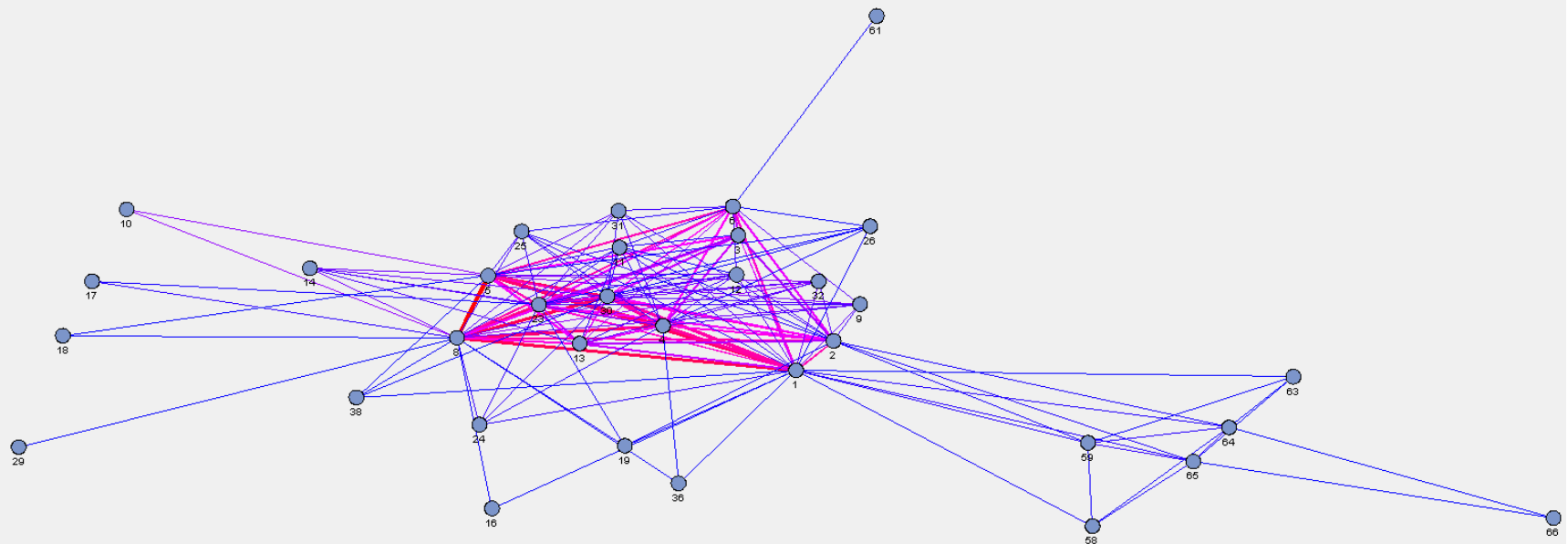
- 서울과 지방 모두 방문
- 체류기간 3~5일
- 확실히 만족한 사람 (만족, 매우 만족)

기준: 지지도

- If X then Y 라는 규칙에 대한 지지도

[물건]

MBA\_Places Visited



## V. 추가분석\_ 1. 방문지 분석

## II. 기본분석 3. 결과해석

서울	인천	강원	경기	충청	제주	경상
동대문시장	송도/인천대교	춘천/남이섬	수원화성	대전	제주 민속촌	부산
명동	영종도/공항주변	용평리조트/ 피닉스파크	대장금 테마파크	단양	성산일출봉	대구
고궁	월미도/ 차이나타운/ 개항장	속초	민속촌	금산인삼타운	용두암	경주
남산/서울N타워	강화도	강릉	서울랜드		중문관광단지/ 서귀포	울산

Support(%)	Confidence(%)	Lift	Item 1	Item 2	Rule	관계	규칙 전치
38.9200	65.0960	1.22938	5		8 ==> 5		2Y
38.9200	73.4975	1.22935	8		5 ==> 8		2Y
36.5723	61.1693	1.11698	1		8 ==> 1		2Y
36.5723	66.7778	1.11691	8		1 ==> 8		2Y
35.0463	74.7011	1.24946	8		6 ==> 8		2Y
35.0463	58.6169	1.24948	6		8 ==> 6		2Y
32.2682	60.9360	1.11265	1		5 ==> 1		2Y
32.2682	58.9188	1.11261	5		1 ==> 5		2Y
30.7943	58.1527	1.23955	6		5 ==> 6		2Y
30.7943	65.6380	1.23956	5		6 ==> 5		2Y
30.4552	55.6085	1.43211	2		1 ==> 2		2Y
30.4552	78.4347	1.43212	1		2 ==> 1		2Y
30.4161	64.8318	1.18386	1		6 ==> 1		2Y
30.4161	55.5370	1.18381	6		1 ==> 6		2Y

- 규칙 파악
- 규칙 별 지지도 파악
- 선정된 방문지 파악
- 방문지 간 거리 파악

# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

II. 사전 조사

III. 분석 방향 설정

IV. 기본분석

V. 추가분석

**VI. 최종 관광마케팅 방안**

VII. 결론 및 제언

1. 코스 개발: 힐링 관광
2. 현재 관광 서비스에서 개선할 점



# 1. 핵심 테마 상품 개발\_“힐링관광”

V. 최종 마케팅 방안  
1. 핵심테마상품개발



"건강검진도 받고 쇼핑, 관광도 하는 패키지 의료"

# 1. 핵심 테마 상품 개발\_“힐링관광”



**서울:** 종합 관광 도시, 특화 의료 서비스  
(성형, 피부, 한방, 척추 여성 질환)

**경기:** 풍부한 관광 자원 발굴 가능,  
지방으로의 '가교' 역할



# 1. 핵심 테마 상품 개발\_“힐링관광”

V. 최종 마케팅 방안  
1. 핵심테마상품개발

## 코스 1. 청정+레포츠형

강원

경기

서울

제주



숲, 산림, 동해, 청정한 공기



스키, 보드 등 레포츠

- 환동해 국제 헬스케어권역
- 특화된 의료서비스 (국내 유일 1일 인플란트 가능)

# 1. 핵심 테마 상품 개발\_“힐링관광”

V. 최종 마케팅 방안  
1. 핵심테마상품개발

## 코스 2. 역사+젊음형



• 의료메카 부산, 메디시티 대구

# 1. 핵심 테마 상품 개발\_“힐링관광”

V. 최종 마케팅 방안  
1. 핵심테마상품개발

## 코스 3. 휴양+재활형

장기휴양, 재활치료

강원

경상

경기

서울

제주

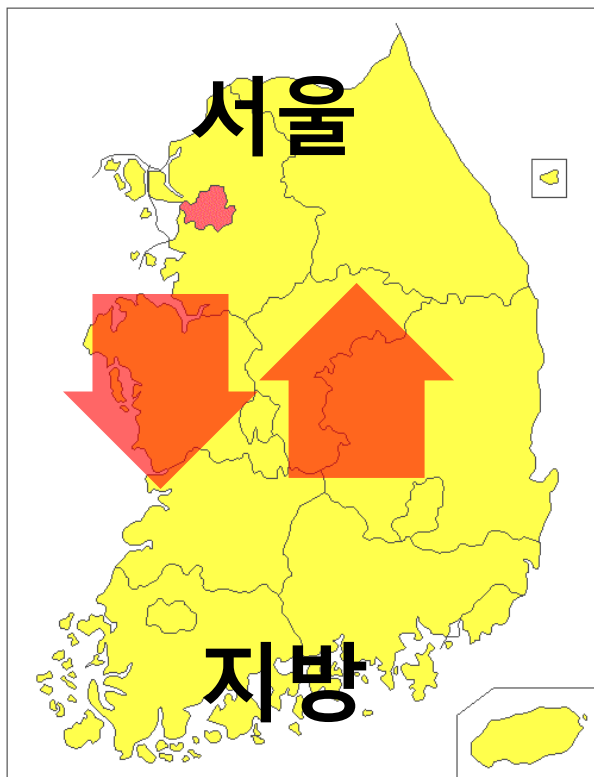


휴양지, 파라다이스

•의료 R&D 센터, 재활센터

## 2. 개선 필요 영역

도출점	시사점	
기본분석	<div> <div>둘다 방문</div> <ul style="list-style-type: none"> <li>아시아권 출신 관광객에 특화된 서비스 필요</li> <li>출입국 절차에서 개선 요소 파악</li> <li>음식에 대한 추가 설문 또는 자료 분석 필요</li> <li>대형 유통 업체(백화점, 할인 마트, 면세점) ⇒ 고객의 만족도 향상 위한 방안 마련</li> </ul> </div>	
	<div> <div>서울만 방문</div> <div>↓ ↑</div> <div>지방만 방문</div> </div> <div> <ul style="list-style-type: none"> <li>지방 관광객 유치 돌파구: 대중교통, 숙박, 관광지 매력도, 관광 안내 서비스에서</li> </ul> </div>	
	<div> <div>지방만 방문</div> </div> <div> <ul style="list-style-type: none"> <li>지방의 백화점, 할인마트 서비스 향상 필요</li> <li>지방 특화 관광 상품 개발 필요</li> </ul> </div>	
추가분석	<ul style="list-style-type: none"> <li>소외지역: 전라도(호남)권, 기타 도서지방 ⇒ 유인 요소 강화 필요</li> <li>선정된 방문지와 실제 대중교통 연계 여부 확인 필요</li> </ul>	



대중교통,  
숙박,  
관광지 매력도,  
관광 상품 다양성

셔틀, 프리패스 등으로  
교통 연계 활성화

백화점,  
할인마트 등의 쇼핑 불편함

## 2. 개선 필요 영역

도출점	시사점	
기본분석	둘다 방문	<ul style="list-style-type: none"> <li>아시아권 출신 관광객에 특화된 서비스 필요</li> <li>출입국 절차에서 개선 요소 파악</li> <li>음식에 대한 추가 설문 또는 자료 분석 필요</li> <li>대형 유통 업체(백화점, 할인 마트, 면세점) ⇒ 고객의 만족도 향상 위한 방안 마련</li> </ul>
	서울만 방문	<ul style="list-style-type: none"> <li>지방 관광객 유치 돌파구: 대중교통, 숙박, 관광지 매력도, 관광 안내 서비스에서</li> </ul>
	지방만 방문	<ul style="list-style-type: none"> <li>지방의 백화점, 할인마트 서비스 향상 필요</li> <li>지방 특화 관광 상품 개발 필요</li> </ul>
추가분석	<ul style="list-style-type: none"> <li>소외지역: 전라도(호남)권, 기타 도서지방 ⇒ 유인 요소 강화 필요</li> <li>선정된 방문지와 실제 대중교통 연계 여부 확인 필요</li> </ul>	

# 관광객 2000만 시대 “지방 관광객” 유치가 관건

I. 과제 파악

II. 사전 조사

III. 분석 방향 설정

IV. 기본분석

V. 추가분석

VI. 최종 관광마케팅 방안

**VII. 결론 및 제언**

1. 분석의 한계
2. 주어진 자료의 한계 및 제언
3. 향후 조사 방향 제언



- 범주형 변수 및 Dummy Variable이 과다  
⇒ 반응률 곡선 등이 제대로 나오지 않음
- Logistic Regression 채택:  
⇒ 직관적인 절편값 해석이 어려움
- 분석 주제인 예측모형("Predictive" Model)의 의미를 제대로 살리지 못함

## 아쉬운 점

- 현실적으로 주어진 데이터 속에서 다양한 마케팅 방안 도출이 어려웠음
- 목표변수가 지정되지 않아서 요구사항 정확히 파악 어려웠음
- 시간 부족, 관리 허술



**‘빅 데이터 시대’, 진정한 데이터 마이닝을 위해서**

### 1. ‘설문지’ 특성의 한계

- 1) 객관성 부족  $\Rightarrow$  실증적 데이터 필요
- 2) 5지 선다형의 Scale의 적절성?

### 2. CODING ISSUE:

- 1) 다중응답항목
- 2) 분류 및 계층화  
EX. 방문지
- 3) 연도별 CODING 통일화 : 분석자의 편의 위해

- Demographic에 대한 체계적인 접근이 필요
- 설문 데이터 이외의 활용 가능한 자료:

#### 위치정보

1. 여행자용 교통카드를 제작?  
⇒ 방문지, 시간 등의 체계적 집계
2. 유관기관과 연계:  
Korail, 시내외버스, 공항

#### 거래 정보

1. 기존에 유통업체가 시행중인  
데이터 마이닝과 연계:
  - 해외 신용카드 거래 내역 정보
  - 민원 사항

#### 관광 안내 정보

1. 관광안내소 통화 내용 및 상담내용:
  - 텍스트로 변환 ⇒ Text Mining
2. 웹 사이트 검색어 텍스트 마이닝

#### 의료정보

1. 성형외과, 대형 병원들과  
외국인 환자 자료 공유
2. 원내 해외마케팅팀과의 협력

Q&A