

데이터마이닝 기법을 이용한
외래관광객 재방문의도 예측
및 관광마케팅 전략



목차

1. 분석 개요
2. 분석단계 정의(SEMMA 방법론)
3. Sampling
4. Exploration/Modification
5. Modeling
6. Assessment
7. 결과 분석
8. 추가 분석(회귀/군집/연관성 분석)
9. 마케팅 전략 및 효과

분석 개요

분석 배경

- 관광산업은 고부가가치 국가전략산업으로 정부에서 많은 투자를 진행중이며, 향후 21세기 최대 산업으로 전망되고 있음.
- 한국을 방문하는 외래관광객의 규모는 이제 1천만명 수준으로 확대되어 관광수지 흑자 시대로 진입할 가능성이 높아짐. 외래관광객 1천만명 시대를 맞이함에 따라 '관광 강국'으로 도약하고 있음.

분석 목적

- 외래관광객 실태조사에서 수집된 데이터를 기반으로 단순 통계 분석이 아닌 데이터마이닝 기법을 이용하여 외래관광객 재방문의도를 예측하는 모형을 구축하고자 함.
- 이를 통해 본 분석보고서에서는 관광산업의 발전을 유도할 수 있는 활용방안과 관광마케팅 전략을 제안함.

분석 개요

데이터 현황

- 한국문화관광연구원에서는 매년 방한 외래관광객의 한국 여행실태, 한국 내 소비실태 및 한국 여행 평가를 조사하여 관광정책을 수립하는 데 있어서 기초자료로 활용함.
- 2009년~2011년의 데이터를 데이터마이닝 모형 구축을 위하여 사용함.

연도별 데이터 수

연도	데이터 수
2009	11,912
2010	11,995
2011	12,038
총합	35,945



분석단계 정의 : SEMMA 방법론

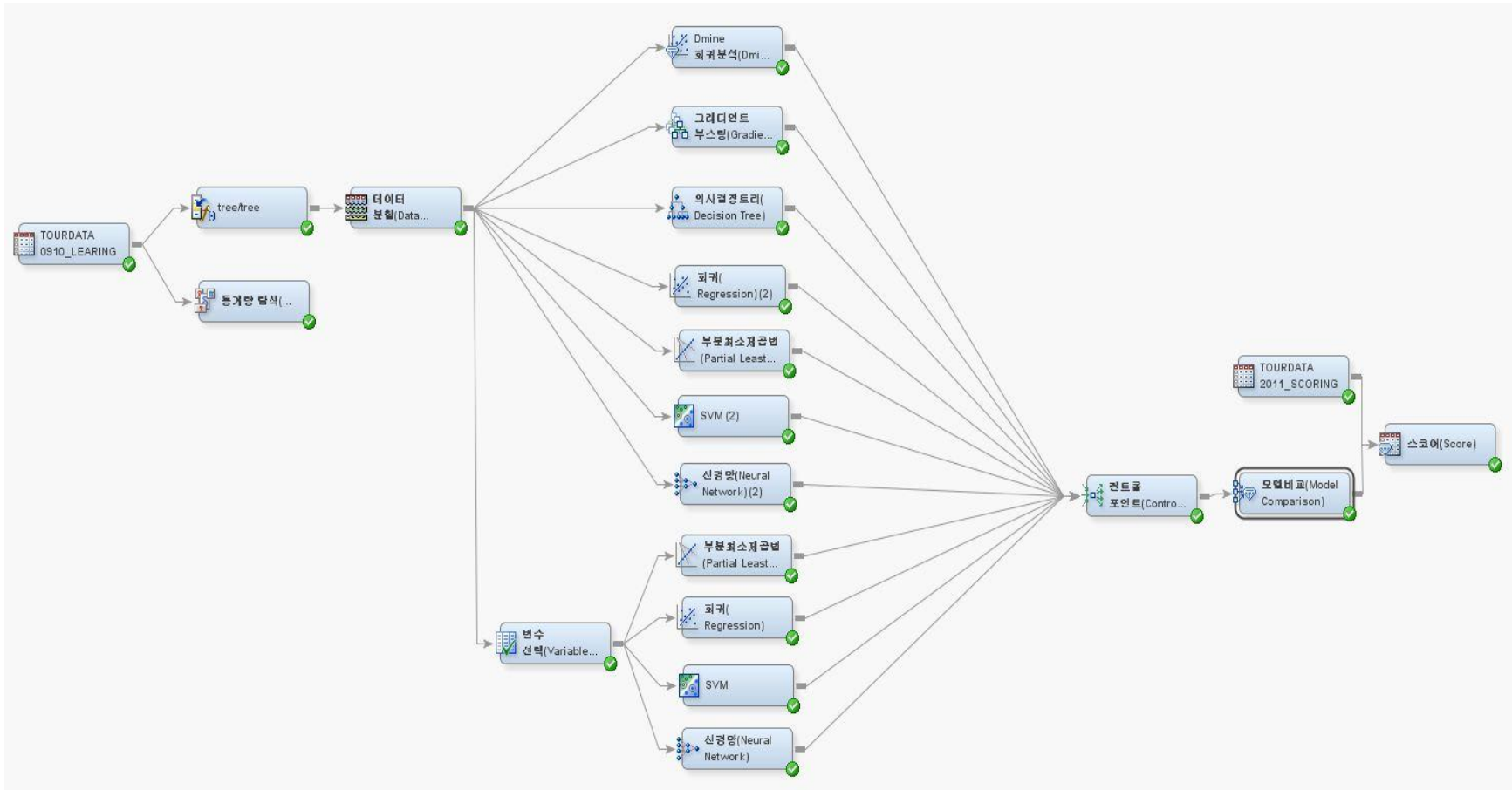
SEMMA 방법론

- SAS의 Enterprise Miner는 데이터를 잘 처리하고 분석하여 유용한 정보의 창출을 얻는 전 과정을 SEMMA라는 과정의 방법론에 의하여 지원함.
- SEMMA의 단계는 Sampling, Exploration, Modification, Modeling, Assessment의 5단계로 이루어짐.

Sampling	Exploration	Modification	Modeling	Assessment
분석 데이터 생성 <ul style="list-style-type: none"> ▶ 입력 데이터 <ul style="list-style-type: none"> - 데이터 역할 - 데이터 척도 ▶ 데이터 분할 <ul style="list-style-type: none"> - Train - Validation - Test 	분석 데이터 탐색 <ul style="list-style-type: none"> ▶ 통계량 탐색 ▶ 그래프 탐색 ▶ 변수 선택 ▶ 클러스터링(추가분석) ▶ 연관성 분석(추가분석) 	분석 데이터 수정/변환 <ul style="list-style-type: none"> ▶ 결측값 처리 <ul style="list-style-type: none"> - 명목형 변수:Tree - 연속형 변수:Tree 	모형 구축 <ul style="list-style-type: none"> ▶ 회귀 ▶ Dmine회귀분석 ▶ 부분최소제곱법 ▶ 신경망 ▶ 의사결정트리 ▶ 그래디언트 부스팅 ▶ SVM 	모형 평가/검증 <ul style="list-style-type: none"> ▶ 모형 비교 ▶ 스코어

분석단계 정의 : SEMMA 방법론

프로세스 플로우 다이어그램



Sampling

Sampling

Exploration

Modification

Modeling

Assessment

- 대량의 데이터에서 유용한 정보를 얻을 수 있도록 분석용 데이터를 추출하는 단계
- 데이터 준비 과정: 분석 데이터 구성, 데이터 분할, 데이터 통합

1) 분석 데이터 구성

연도	칼럼 수	데이터 개수
2009	191	11,912
2010	191	11,995
2011	238	12,038

2) 데이터 분할

역할	비율	데이터개수
Training	50	11,927
Validation	30	7,157
Test	20	4,771
Score	-	12,038

- 2개년 (2009년 및 2010년) 데이터를 예측 모형 구축에 사용함
- 2011년 데이터는 스코어 데이터로 사용함
- 3개년 데이터 중 동일 항목이지만 칼럼명이 다른 경우 변수명과 변수값을 통일하여 통합함
- 입력데이터 노드의 옵션을 사용하여 결측 비율이 60 이상인 칼럼을 제거함

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

- 그래프와 통계량을 통하여 탐색적으로 데이터를 분석하고, 효과적인 모형 구축을 위하여 분석에 적합하도록 변수를 변형, 선택, 생성하는 단계
- 결측값 처리 및 주요 변수 선정 과정: 데이터의 결측값을 처리하고 데이터의 모형 개발을 위해 주요 변수 선정

결측값 처리 - Tree imputation

- 방법:
 - ① 결측값을 처리하고자 하는 변수를 Target으로 지정하고 기각된 변수를 포함한 모든 Input 변수들을 Input으로 지정하여 의사결정트리 모형을 구축함.
 - ② 모형의 예측값이 해당 변수의 결측값을 대체
 - 사용 이유:
 - ① 결측치의 수가 비교적 많을 때 사용하기 적합함
 - ② 다른 변수들을 기반으로 결측값을 처리하기 때문에 단순하게 분포의 중심성을 사용하는 결측값 처리 방법보다 정확도가 높음
- ✓ Enterprise Miner의 **결측값 처리(Impute) 노드**를 사용하여 결측값을 처리함
- ✓ Interval 및 Class 변수에 모두 Tree imputation을 적용함

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

파생변수 생성

방법	설명	생성 변수	
		설명	생성 로직
계산	두 개 이상의 변수를 계산하여 새로운 변수를 도출하는 방법	<ul style="list-style-type: none"> 인상깊은 점 개수 불편했던 점 개수 최근 방문 경과월 	Count(한국 여행에서 인상 깊은 점) Count(한국 여행에서 불편했던 점) 조사년월 - 최근 한국 방문년월
통합	하나의 항목에 대한 응답이 여러 칼럼에 나누어 존재하는 경우 하나의 칼럼에 통합하는 방법	<ul style="list-style-type: none"> 항목별 비용 변수 방문 동반자 현황 직전/직후 방문예정 	개별여행 지출금액 & 단체여행 지출금액 방문 동반자현황 1 & 2 & ...& 6 한국만 & 직전 & 직후

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

목표변수 정의

- 외래 관광객의 한국에 대한 '재방문 의도'를 목표변수로 선정함:
 5점 척도로 측정된 '재방문 의도'를 긍정적(4점, 5점)과 부정적(1점~3점)으로 범주화
 - **정의** : 재방문 의도는 해당 서비스 장소에 재방문 가능성이 있거나 타인에게 해당 서비스 장소의 우수성을 자발적으로 추천할 가능성이 높은 상태를 의미함. 즉, 한국을 방문한 외래 관광객이 다시 한국을 방문할 가능성과 다른 사람들에게 한국을 추천하는 가능성이 있을 때 그 관광객은 재방문 의도가 긍정적인 것으로 볼 수 있음¹⁾.
 - **이유** : **1)** 5개의 class를 가진 변수를 2개의 class로 변환했을 때 인지적으로 이해하기 쉬우며, **2)** 일반적으로 이진분류(Binary class classification) 문제가 다진분류(Multi-class classification) 문제보다 예측 성과가 더 높다고 알려져 있음.

1) 최나리(2005), "해수욕장의 방문동기, 관광만족 및 재방문 의도 연구: 부산지역 해수욕장을 중심으로", 「관광연구」, 20(2), pp. 165~181.

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

목표변수 정의

공격적 마케팅
전략 중심
- 신규고객 유치
- 시장 점유율 확대



방어적 마케팅
전략 중심
- 기존고객 유지
- 고객충성도 제고

- 충성도가 높은 기존 관광객은 한국을 빈번하게 방문할 뿐만 아니라, 긍정적 구전효과를 통해 신규 잠재 관광객들을 창출하여 관광 산업 이익증대에 큰 역할을 할 수 있음.
- 신규고객을 유치하는 비용은 기존 고객을 유지하는 비용의 5배임¹⁾. 따라서 관광객의 재방문을 유도하는 것은 관광산업이 최소한의 비용으로 수익을 증대시킬 수 있는 효율적인 방법 중 하나로 볼 수 있음.

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

설명변수 선정 (R 제곱)

- SAS Enterprise Miner 탐색 단계의 변수 선택 노드를 이용하여 주요 변수 선정함.
- 결정계수(R 제곱) 이용하여 목표변수와 관련성이 높은 설명변수 선택함. 임계값 (최소 R-제곱)과 결정계수를 비교하여 임계값보다 작은 결정계수를 가지면 제거함. 설명변수 연속형인 경우와 이산형인 경우 각각에 대해서 적절한 변환을 통해 복잡한 비선형 관계 파악함.

R 제곱 옵션(default)

구분	옵션
최대 변수 개수	3000
최소 R-제곱	0.0050
R-제곱 중지	5.0E-4
AOV16 변수 사용	YES
그룹 변수 사용	YES
교호작용	NO

GROUP 변수

각 클래스 변수의 레벨을 타겟과의 관계에 기반하여 그룹 변수로 축소

AOV16 변수

구간 변수를 16개의 동일한 간격의 그룹으로 나눠서 타겟과의 비선형관계 확인

Exploration/Modification

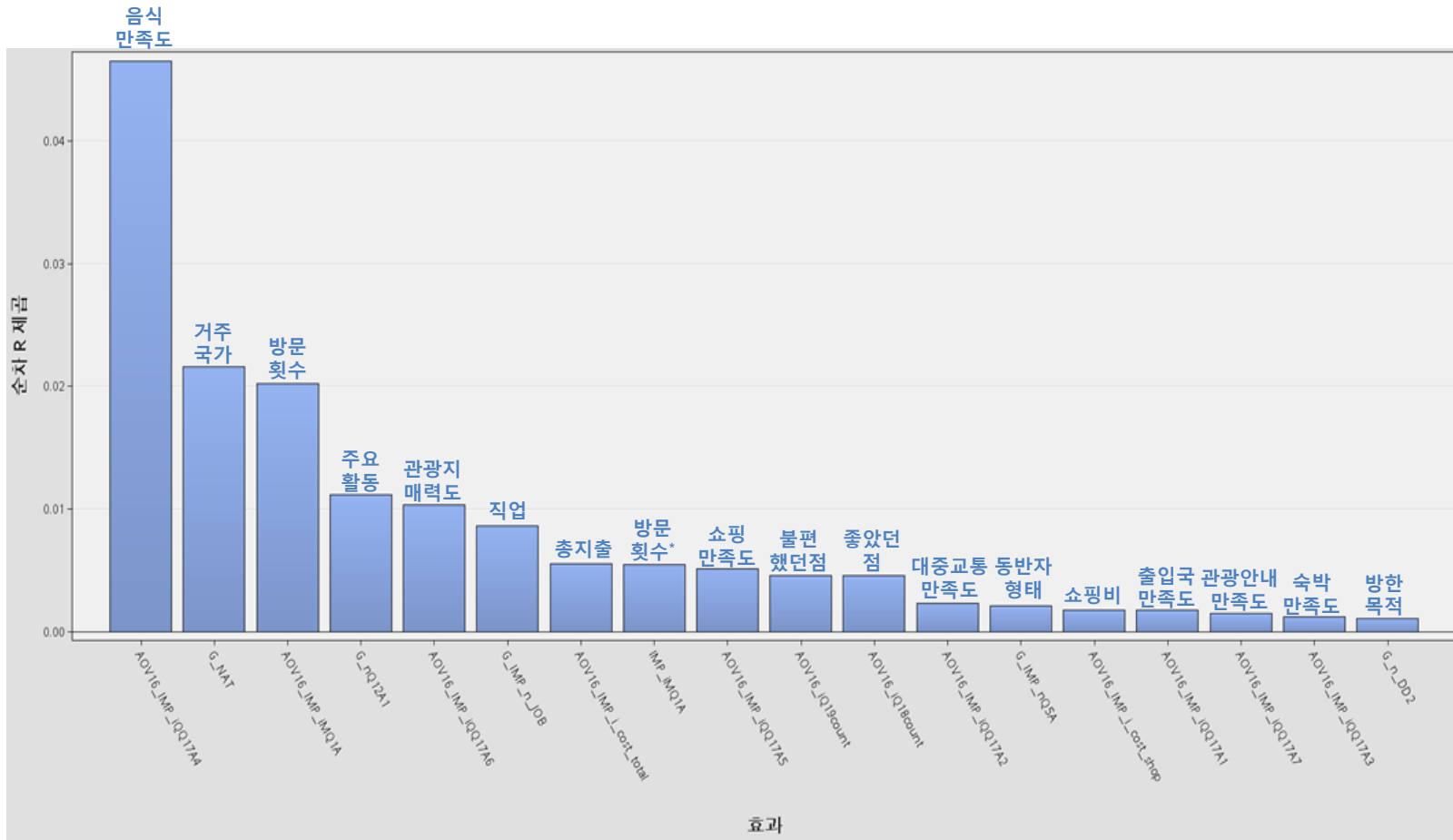
선정된 설명변수

변수명	설명	측도레벨	df	R ²	F
AOV16_IMP_iQQ17A4	음식 만족도	Ordinal	10	0.046549	58.1756 ^{***}
G_NAT	거주국가	Nominal	6	0.02156	45.9237 ^{***}
AOV16_IMP_iMQ1A	최근 3년 한국방문횟수	Ordinal	12	0.020216	22.958 ^{***}
G_nQ12A1	주요활동(1순위)	Nominal	6	0.01118	24.0909 ^{***}
AOV16_IMP_iQQ17A6	관광지 매력도	Ordinal	10	0.010313	13.4733 ^{***}
G_IMP_n_JOB	직업	Nominal	3	0.008639	37.9713 ^{***}
AOV16_IMP_i_cost_total	총지출	Ordinal	15	0.005537	4.8917 ^{***}
AOV16_IMP_iQQ17A5	쇼핑 만족도	Ordinal	9	0.005145	7.8302 ^{***}
AOV16_iQ19count	불편했던점(개수)	Ordinal	10	0.004592	6.3461 ^{***}
AOV16_iQ18count	좋았던점(개수)	Ordinal	12	0.004561	5.2284 ^{***}
AOV16_IMP_iQQ17A2	대중교통 만족도	Ordinal	12	0.002292	2.6442 ^{***}
G_IMP_nQ5A	동반자형태	Nominal	3	0.002082	9.6286 ^{***}
AOV16_IMP_i_cost_shop	쇼핑비	Ordinal	15	0.001753	1.6226 [*]
AOV16_IMP_iQQ17A1	출입국절차 만족도	Ordinal	9	0.001735	2.68 ^{***}
AOV16_IMP_iQQ17A7	관광안내서비스 만족도	Ordinal	12	0.001475	1.71 [*]
AOV16_IMP_iQQ17A3	숙박 만족도	Ordinal	10	0.001207	1.6799 [*]
G_n_DD2	방한목적	Nominal	4	0.001072	3.7348 ^{***}

p < 0.01: ^{***}, p<0.05: ^{**}, p<0.1: ^{*}

Exploration/Modification

선정된 설명변수의 중요도



*: 기존 방문횟수 변수는 AOV16로 변환된 방문횟수에 대한 변수가 R²이 더 높아 기각되었음.

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

설명변수 선정 (의사결정트리)

- 의사결정트리는 규칙과 함께 목표변수에 영향을 주는 변수를 제시해줌.
- 트리로부터 생성되는 결정변수(각 마디를 형성하는 변수)들을 파악하여 신경망, SVM, 회귀 모형의 설명변수로 적용함.

■ 분리 규칙 및 노드

구분	옵션
Interval 기준	ProbF ¹⁾
Nominal 기준	ProbChisq ²⁾
Ordinal 기준	Entropy ³⁾
최대 가지	2
최대 깊이	6
Leaf 크기	500



■ 의사결정트리 결정변수

변수명	설명
IMP_iMQ1A	최근3년한국방문횟수
iQQ17A4	음식 만족도
iQQ17A6	관광지 매력도
NAT	거주국가

1) 분산분석에서 F 통계량: p-값이 가장 작은 예측변수와 그 때의 최적분리에 의해서 자식마디를 형성

2) 카이제곱 통계량의 p-값: p-값이 가장 작은 예측변수와 그 때의 최적분리에 의해서 자식마디를 형성

3) 엔트로피 지수: 다항분포에서 우도비 검정통계량을 사용하는 것과 같은 것으로 알려져 있음. 이 지수가 가장 작은 예측변수와 그 때의 최적분리에 의해서 자식마디를 형성

Exploration/Modification

Sampling

Exploration

Modification

Modeling

Assessment

선정된 설명변수 비교

R제공

변수명	설명
AOV16_IMP_iQQ17A4	음식 만족도
G_NAT	거주국가
AOV16_IMP_iMQ1A	최근 3년 한국방문횟수
G_nQ12A1	주요활동(1순위)
AOV16_IMP_iQQ17A6	관광지 매력도
G_IMP_n_JOB	직업
AOV16_IMP_i_cost_total	총지출
AOV16_IMP_iQQ17A5	쇼핑 만족도
AOV16_iQ19count	불편했던점(개수)
AOV16_iQ18count	좋았던점(개수)
AOV16_IMP_iQQ17A2	대중교통 만족도
G_IMP_nQ5A	동반자형태
AOV16_IMP_i_cost_shop	쇼핑비
AOV16_IMP_iQQ17A1	출입국절차 만족도
AOV16_IMP_iQQ17A7	관광안내서비스 만족도
AOV16_IMP_iQQ17A3	숙박 만족도
G_n_DD2	방한목적

의사결정트리

변수명	설명
IMP_iMQ1A	최근3년한국방문횟수
iQQ17A4	음식 만족도
iQQ17A6	관광지 매력도
NAT	거주국가

- 변수선택노드를 사용한 결과 17개의 변수가 선택됨
- 의사결정트리에 의해 4개의 변수가 선택됨
- R제공 변수는 의사결정트리 변수를 모두 포함하고 있음 (의사결정트리 C R제공)
- 즉, 중복된 4개의 변수는 목표변수인 재방문의도를 설명하는 데 매우 적절함

Modeling

Sampling

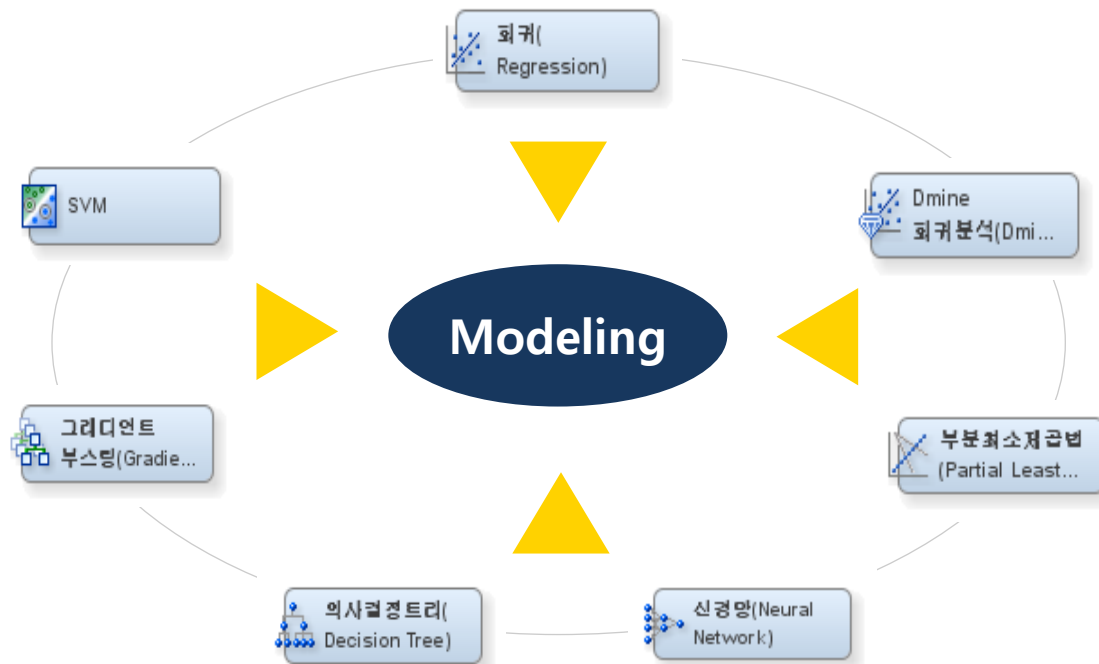
Exploration

Modification

Modeling

Assessment

- 목적에 맞는 데이터마이닝 기법을 중심으로 다양한 모델링을 수행하는 단계
- 모형 개발 과정: 탐색적/예측적 모형 개발 시 설정한 옵션 및 근거



Modeling

Sampling

Exploration

Modification

Modeling

Assessment

구축 모형 옵션

구분	내용
회귀분석	<ul style="list-style-type: none"> ▪ 목표변수가 이진(binary)이기 때문에 로지스틱 회귀분석(Logistic regression) 수행 ▪ 연결함수(Link function)로 로짓(Logit) 사용
Dmine 회귀분석	<ul style="list-style-type: none"> ▪ AOV16변수 및 그룹변수를 포함하여 분석 ▪ 전진 단계별 최소제곱회귀계산 방법을 이용
부분최소제곱법	<ul style="list-style-type: none"> ▪ 잠재변수(Latent variable)를 사용 ▪ 여러 알고리즘 중 PLS요인분석을 사용
신경망	<ul style="list-style-type: none"> ▪ 1개의 은닉층, 은닉마디수는 입력변수의 수와 동일하게 설정 ▪ 출력층의 결합함수는 Linear, 활성화함수는 Logistic, 오차함수는 Logistic으로 설정
의사결정트리	<ul style="list-style-type: none"> ▪ 분류규칙을 위해 Interval은 ProbF, Nominal은 ProbChiq, Ordinal은 Entropy로 설정 ▪ 노드의 최대 가지는 2, 최대 깊이는 6, Leaf 크기는 500으로 설정(과적합 방지)
그래디언트 부스팅	<ul style="list-style-type: none"> ▪ 시리즈 옵션에서 N 반복은 50, 분석용 관측치 비율 60으로 설정 ▪ 트리의 최대 가지는 2, 최대 깊이는 2로 설정
SVM	<ul style="list-style-type: none"> ▪ 추정을 위해 DQP(Decomposed Quadratic Programming) 방법 사용 ▪ Regularized 모수는 조정(Tuning) 방법 선택(조정 범위: 0.1~1.0, 증가값 0.1) ▪ 커널 유형으로 선형을 이용(다항식, Radial Basis, Sigmoid 사용 가능)

Assessment

Sampling

Exploration

Modification

Modeling

Assessment

- 어떤 모형이 가장 우수한 예측력을 제공하고 있는지 등을 비교, 분석, 평가하는 단계
- 모형 결과 비교 및 평가 과정: 정량적/정성적 모형 비교 및 평가

모형 결과
비교 및 평가



정량적 평가
: 분류정확도



정성적 평가
: 활용성

Assessment

Sampling

Exploration

Modification

Modeling

Assessment

정량적 평가(정확도 측면)

- 컨트롤포인트를 생성한 후 모델비교 노드를 사용함.
- **모델 선택 통계량:** 검증용 테이블의 오분류비율
- **결과:** 그래디언트 부스팅 (26.43%) 이 선택됨.

사용 모델	사용변수*	오분류 비율		
		평가	분석	검증
그래디언트 부스팅(Gradient Boosting)	전체	25.31	25.53	26.43
Dmine 회귀분석(Dmine Regression)	전체	25.02	26.21	26.70
부분최소제곱법(Partial Least Squares)	변수선택	25.40	25.90	27.21
회귀(Regression)	변수선택	25.35	25.90	27.27
회귀(Regression) (2)	DT변수	26.35	26.90	27.31
신경망(Neural Network) (2)	DT변수	25.64	26.51	27.42
부분최소제곱법(Partial Least Squares) (2)	DT변수	26.99	27.20	27.48
신경망(Neural Network)	변수선택	26.72	26.58	27.77
의사결정트리(Decision Tree)	전체	26.19	26.60	27.79
SVM (2)	DT변수	28.24	28.24	28.23
SVM	변수선택	30.61	31.97	33.25

사용변수*	설명	변수 개수
전체	결측비율 고려하여 선정된 Data set의 모든 Input 변수	46
변수선택	변수선택노드를 사용하여 선정된 Input 변수	17
DT변수	의사결정나무에서 선택된 Input 변수	4

Assessment

Sampling

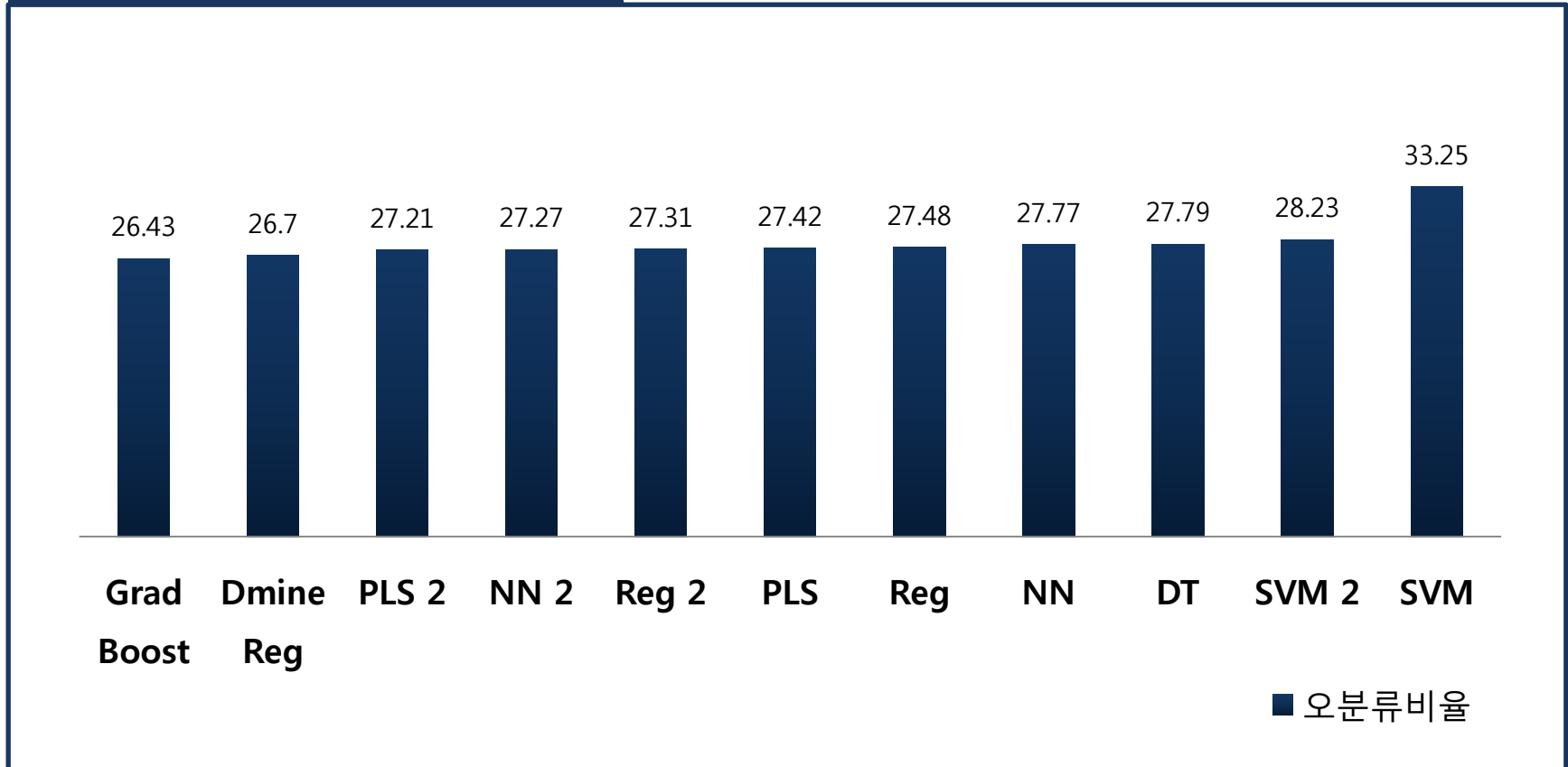
Exploration

Modification

Modeling

Assessment

정량적 평가(정확도 측면)



Assessment

Sampling

Exploration

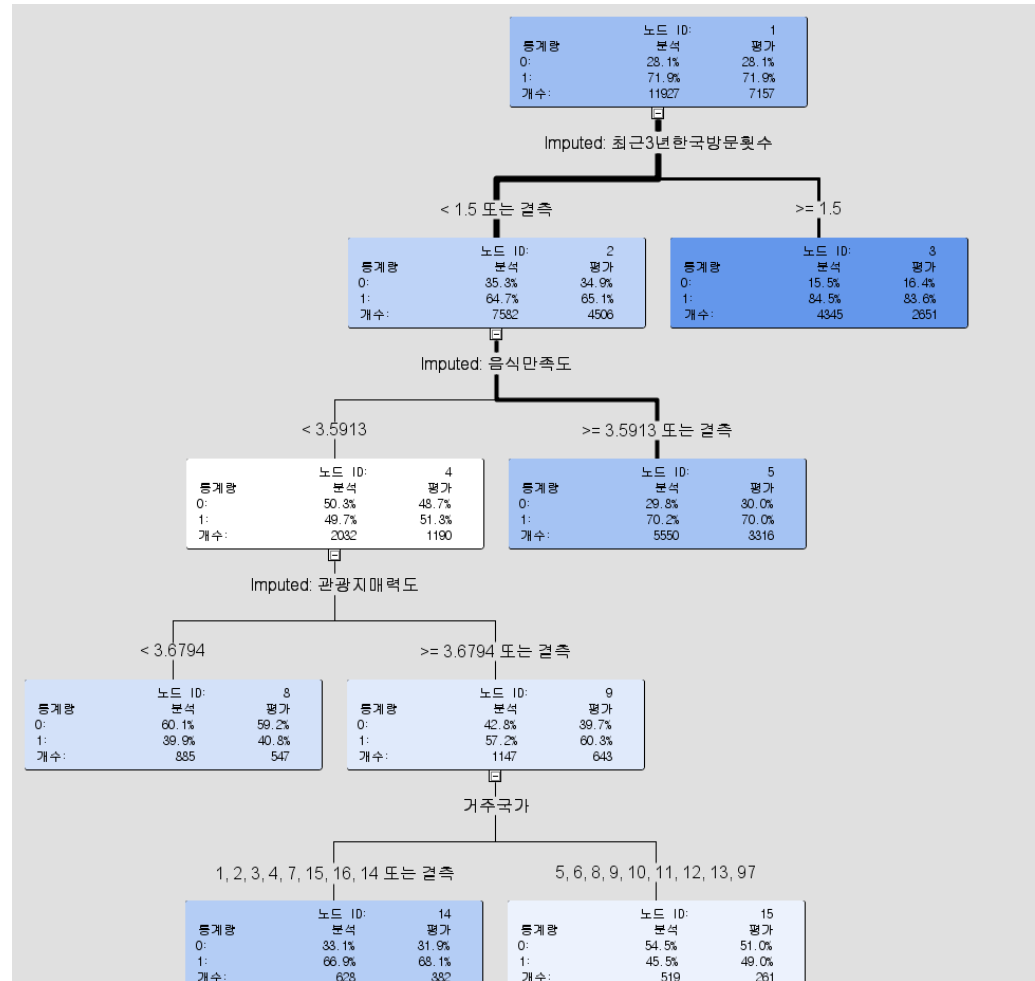
Modification

Modeling

Assessment

정성적 평가(활용성 측면)

- 의사결정트리 기법에 의해 구축된 모형은 신경망이나 SVM 등에 의한 방법과는 달리 명확히 표현된 규칙들이 도출되기 때문에 어떠한 경우에 외래관광객들이 한국을 재방문 하는지 쉽게 파악할 수 있음.
- 즉, 데이터로부터 나무구조의 일반화된 지식을 추출할 수 있어 활용성이 높음.

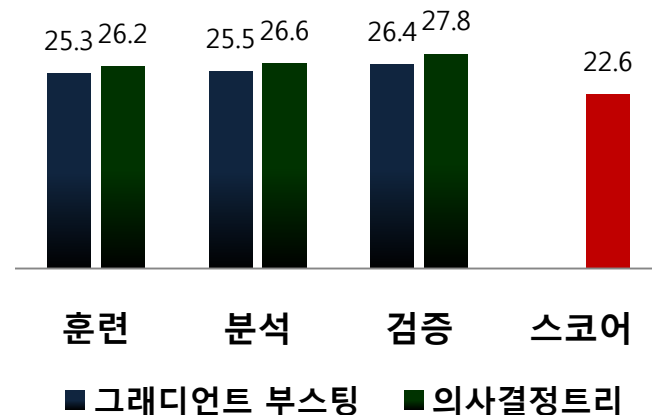


결과 분석

최종 모형 선택 - 의사결정트리

- 사용변수 개수가 많지 않아 모형의 효율성이 높음
- 그룹변수나 AOV16변수를 사용하지 않아서 스코어 데이터에 적용하기 쉬움
- 일반화된 지식을 추출하여 비즈니스 활용성이 높음
- 명확히 표현된 규칙이 도출됨

⇒ 따라서 의사결정트리 모형을 최종 예측모형으로 선정함



결과 분석

Sampling

Exploration

Modification

Modeling

Assessment

외래관광객의 한국 재방문의도에 관한 규칙

번호	규칙
Rule 1	IF 최근3년한국방문횟수 ≥ 1.5 THEN 한국 재방문의도=있음(긍정적, 85%)
Rule 2	IF 최근3년한국방문횟수 < 1.5 AND 음식만족도 ≥ 3.5913 THEN 한국 재방문의도=있음(긍정적, 70%)
Rule 3	IF 최근3년한국방문횟수 < 1.5 AND 음식만족도 < 3.5913 AND 관광지매력도 < 3.6794 THEN 한국 재방문의도=없음(부정적, 60%)
Rule 4	IF 거주국가 IS ONE OF: 일본, 중국, 홍콩, 싱가포르, 말레이시아, 중동, 인도, 러시아 AND 최근3년한국방문횟수 < 1.5 AND 음식만족도 < 3.5913 , 관광지매력도 ≥ 3.6794 THEN 한국 재방문의도=있음(긍정적, 67%)
Rule 5	IF 거주국가 IS ONE OF: 대만, 태국, 호주, 미국, 캐나다, 영국, 독일, 프랑스 AND 최근3년한국방문횟수 < 1.5 AND 음식만족도 < 3.5913 , 관광지매력도 ≥ 3.6794 THEN 한국 재방문의도=없음(부정적, 55%)

추가 분석

- 회귀분석
- 군집분석
- 연관성 분석

추가 분석_회귀분석



회귀분석

- 한국을 방문한 외래관광객이 평가한 한국여행의 항목별 만족도(출입국절차, 대중교통, 숙박, 음식, 쇼핑, 관광지매력도, 관광안내서비스)를 설명변수로 설정하고 재방문의도(5점 척도)를 목표변수로 설정함(3개년 데이터 사용).
- 일반선형회귀를 이용하였고 옵션으로 2요인 교호작용과 기본 최적화기법을 선택함.
- 분석결과는 **관광정책 및 마케팅 전략 수립 시 어떤 부분에 더 집중해야 하는지 기준을 제시해 줌.**

사용된 변수 및 역할 지정

변수명	설명	역할
iQQ17A1	출입국절차 만족도	Input
iQQ17A2	대중교통 만족도	Input
iQQ17A3	숙박 만족도	Input
iQQ17A4	음식 만족도	Input

변수명	설명	역할
iQQ17A5	쇼핑 만족도	Input
iQQ17A6	관광지매력도	Input
iQQ17A7	관광안내서비스 만족도	Input
iQ20	재방문의도	Target

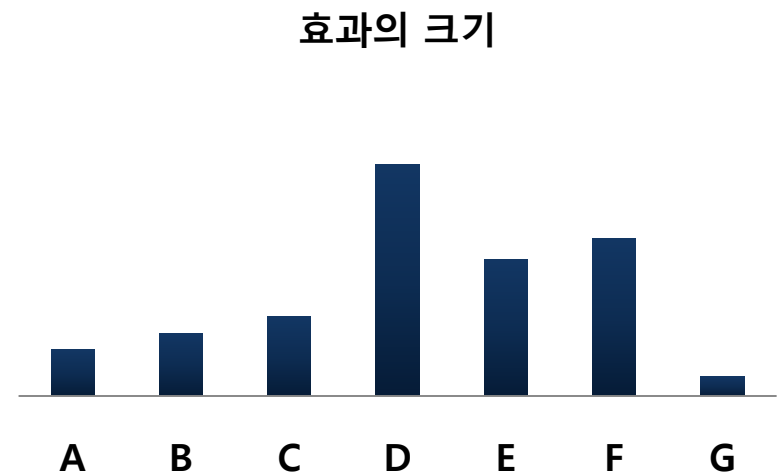
추가 분석_회귀분석

결과

- 도출된 회귀 모형은 유의수준 0.01에서 통계적으로 유의하여 도출된 모형이 목표변수인 재방문의도를 설명하기 적합하다고 할 수 있음 ($F=466.79$, $d.f.=7$)
- 회귀계수의 t-검정 결과, 관광안내서비스 만족도를 제외한 모든 변수가 목표변수인 재방문의도를 설명하기에 적합함.
- 설명변수의 효과:
음식만족도 > 관광지매력도 > 쇼핑만족도 > 숙박만족도 > 대중교통만족도 > 출입국절차만족도
- 따라서 관광 마케팅 정책 및 전략 수립 시 각 항목의 중요도를 고려해야 함

변수	Estimate	t-value
A 출입국절차 만족도	0.0340	4.51***
B 대중교통 만족도	0.0412	5.42***
C 숙박 만족도	0.0518	6.81***
D 음식 만족도	0.1519	22.60***
E 쇼핑 만족도	0.0893	10.99***
F 관광지매력도	0.1032	11.97***
G 관광안내서비스 만족도	0.0129	1.64

*** $p<0.01$



추가 분석_군집분석

군집분석

- 목표변수에 대한 사전 정보가 없는 경우, 주어진 관측 값들 사이의 거리 또는 유사성을 기반으로 전체를 몇 개의 집단으로 그룹화하여 각 집단의 성격을 파악할 수 있음
- 방문객의 과거 방문패턴(RFM)을 분석하여 **높은 가치를 지닌 고객군**을 찾을 수 있음
- SAS Enterprise Guide의 군집분석을 이용함

RFM 분석

- RFM분석을 관광마케팅에 적용함
 - ✓ RFM(Recency, Frequency, Monetary): 미래의 구매행위를 예측하기 위해 과거의 구매 패턴, 즉 구매 시점, 특정 기간 동안의 구매 빈도, 구매액을 분석하는 방법
- 설명변수를 RFM으로 설정하고 K가 3인 K-means 군집분석을 실시함

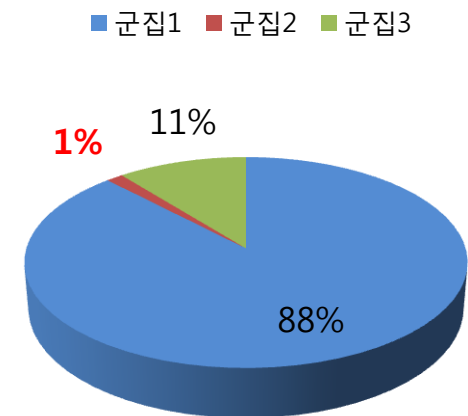
역할	사용변수
R	최근 방문 경과 (월)
F	최근 3년간 한국 방문 횟수
M	총 지출 금액

추가 분석_군집분석

결과

- 군집분석 결과 방문객의 과거 방문패턴에 따라 고유의 특성을 가진 세 그룹으로 나누어짐
- 현재 우리나라의 관광 수지는 적자이기 때문에 우리나라에 대한 충성도가 높은 VIP 관광객의 특성을 파악하고 그들을 대상으로 타겟 마케팅을 실시하는 것이 바람직함

군집	군집평균			개수	특성
	R	F	M		
1	5.81	2.50	975.65	31,574	오래 전 방문 / 가끔 방문 / 적은 지출
	H	L	L		
2	4.70	4.41	8,171.37	501	최근 방문 / 자주 방문 / 많은 지출
	L	H	H		
3	5.70	2.93	3761.10	3,870	보통 방문 / 보통 방문 / 보통 지출
	M	M	M		



추가 분석_군집분석

VIP 방문객 특성 요약

구분	VIP 방문객	전체 방문객
국적	중국	일본
연령	20대	20대
직업	자영업자	경영/사무/기술직
여행형태	개별여행	개별여행
서울방문지	동대문시장	동대문시장
지방방문지	부산	설악산
주요활동	쇼핑	쇼핑
쇼핑품목	의류	의류
장소	공항면세점	공항면세점

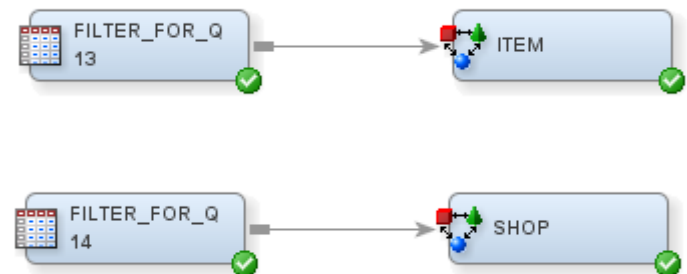
추가 분석_연관성 분석

연관성 분석

- 한국을 방문한 외래관광객의 '쇼핑 장소' 및 구매한 '쇼핑 품목' 변수를 대상으로 연관성 분석을 수행함.
- 연관성 분석을 통해 도출된 규칙은 관광상품 패키지 개발을 위한 기초 자료로 활용 가능하고, 여행안내서나 광고에 **연관성 높은 항목**을 가까이 배치할 수 있음.
- 사전에 최대 항목(Maximum items), 최소 신뢰도 수준(Minimum confidence level), 지지도 백분율(Support percentage)을 지정하였음.

사용된 변수 및 역할 지정

변수명	설명	역할
ID	외래관광객 번호	ID
Q13item	한국여행 시 구입한 주요 쇼핑 품목	Target
Q14place	한국여행 시 주로 쇼핑한 장소	Target



추가 분석_연관성 분석

한국여행 시 주로 쇼핑한 장소에 대한 연관 규칙

Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Description
2	17.93	27.93	2.58	1.56	858	이태원 시장 ==> 남대문 시장
2	30.73	40.35	12.86	1.31	4,268	명동 ==> 동대문 시장
2	30.73	39.51	7.08	1.29	2,352	남대문 시장 ==> 동대문 시장
2	10.36	12.48	2.24	1.2	743	남대문 시장 ==> 인사동
3	30.73	35.95	2.22	1.17	738	명동 & 남대문 시장 ==> 동대문 시장
2	31.86	35.73	5.31	1.12	1,764	시내 면세점 ==> 명동
2	31.86	34.74	3.6	1.09	1,195	인사동 ==> 명동
2	31.86	34.49	6.18	1.08	2,053	남대문 시장 ==> 명동
2	30.73	31.29	4.65	1.02	1,545	시내 면세점 ==> 동대문 시장
2	28.25	28.03	5.61	0.99	1,863	소규모 상점 ==> 할인점(이마트, 홈플러스 등)

(최대 항목=4, 최소 신뢰도 수준=10, 지지도 백분율=5)

추가 분석_연관성 분석

한국여행 시 구입한 주요 쇼핑 품목에 대한 연관 규칙

Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Description
2	40.22	62.21	9.93	1.55	2,336	신발류 ==> 의류
3	40.22	54.46	2.23	1.35	525	향수, 화장품 & 신발류 ==> 의류
2	20.25	27.33	4.36	1.35	1,025	김치 ==> 인삼, 한약재
2	31.84	36.27	7.34	1.14	1,727	인삼, 한약재 ==> 향수, 화장품
3	15.97	17.8	2.23	1.11	525	향수, 화장품 & 의류 ==> 신발류
2	31.84	32.59	5.2	1.02	1,222	김치 ==> 향수, 화장품
2	29.38	29.08	3.3	0.99	775	술 ==> 식료품
2	40.22	39.41	12.55	0.98	2,950	향수, 화장품 ==> 의류
3	31.84	30.57	2.02	0.96	475	인삼, 한약재 & 의류 ==> 향수, 화장품
2	40.22	38.07	2.97	0.95	699	피혁제품 ==> 의류

(최대 항목=4, 최소 신뢰도 수준=10, 지지도 백분율=5)

마케팅 전략 및 효과

방문객 특성 파악 및 마케팅 전략 수립

회귀분석: 집중할 분야의 선정 기준 제시

재방문의도에 큰 영향을 주는 음식, 관광지, 쇼핑의 만족도를 향상 시킬 수 있는 정책을 수립하고 세 항목을 중심으로 한정된 예산을 할당한다

군집분석: 집중할 고객의 선정 기준 제시

방문객의 과거 방문패턴(RFM)을 분석하여 높은 가치를 지닌 고객군을 선정한 후 타겟 마케팅을 할 수 있다

- VIP고객을 대상으로 한 프로모션 - 항공사, 백화점 등

연관성분석: 추천 항목의 선정 기준 제시

관광지, 쇼핑품목, 주요활동 등의 연관성분석을 통해 관련있는 항목을 파악하고 추천할 수 있다

- 추천 관광상품 패키지 구성
- 여행안내서나 광고에 연관성 높은 항목을 가까이 배치

의사결정
트리를
이용한
재방문의도
예측

효과

▪ 규칙 도출

이해하기 쉬운 규칙을 도출함으로써 관광 정책이나 마케팅 전략 수립 및 변경이 용이하다

▪ 관광산업발전

정확한 목표를 설정하여 재방문율을 높임으로써 관광산업을 발전시킬 수 있다

▪ 관광강국

재방문율 상승은 구전 효과와 국가 이미지 향상에 긍정적 영향을 줄 것으로 기대할 수 있다

감사합니다