

# *DATA166*

국내 여행 인식 개선 및 여행 만족도 모델에 따른 여행 코스 제안





#### 배경 선정 이유

1. 한국인의 국내여행의 대한 부정적 인식 개선
2. 획일화된 여행 추천 서비스 → 만족도 고려한 여행 추천 서비스

#### 데이터 수집 / 데이터 처리

1. 결측치 처리
2. 변수 변환
3. 변수 생성

1. Decision Tree
  2. XGBoost
  3. AdaBoost
  4. Random Forest
- 다중 대응 분석

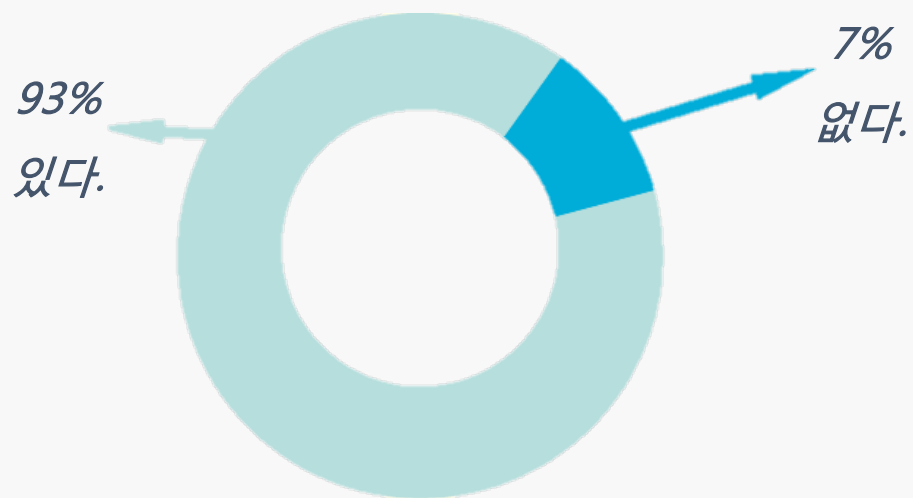
#### 개개인별 맞춤 여행 코스 제안

1. 홍길동씨
2. 최불암씨
3. 한혜진씨
4. 박찬호씨

#### Marketing 전략 제안

1. VR 체험
2. 서포터즈 창단
3. 홍보 동영상 제작

Q. 스스로를 위한 소비, 얼마나 하십니까?



A. 소비자 10명 중 9명

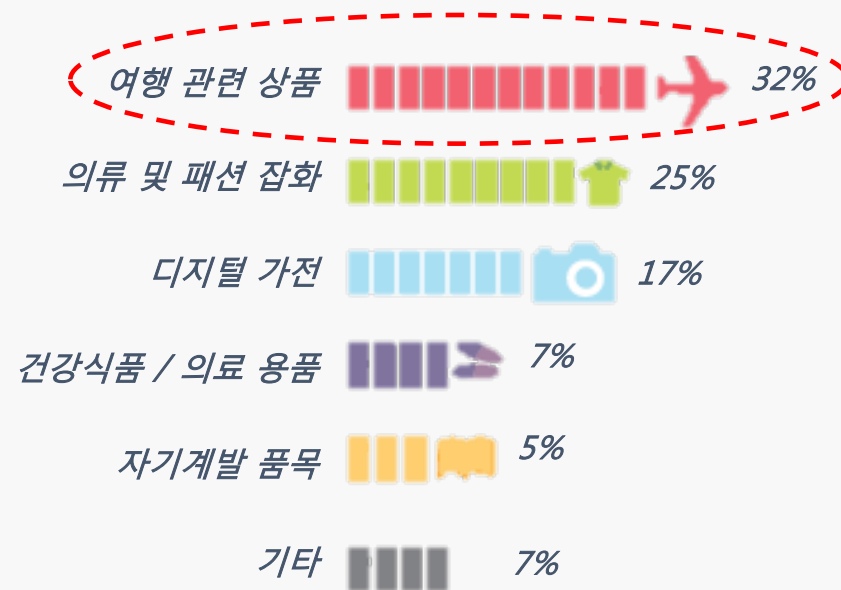
'온전히 나만을 위한 소비 한 적 있다'

스스로를 위한 소비를 많은 사람들이 하고 있으며,  
점차적으로 늘어나고 있는 추세다.

대상 : G-market을 이용하는 남녀 946명

출처 : 컨슈머와이드(<http://www.consumerwide.com>)

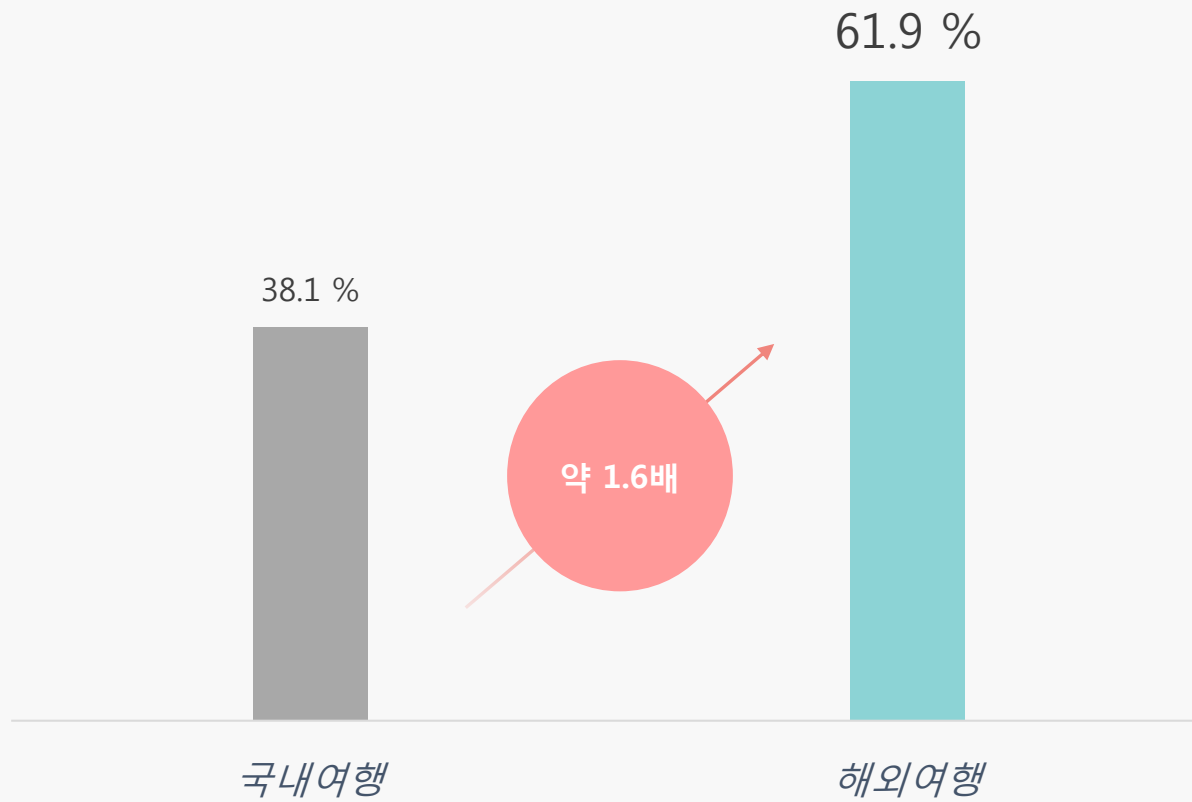
Q. 앞으로 본인을 위해 더 소비 하고 싶은 품목은?



A. 여행을 떠나고 싶다.

스스로를 위한 소비가 필요한 이유로  
'나 자신에 대한 보상 차원'(36%)이라는 응답이 가장 많았다.

출처 : 컨슈머와이드(<http://www.consumerwide.com>)



## 여행 선호도 설문 조사

Q. 여행을 떠난다면 국내 vs. 해외

어디로 가실 건가요?

대상 : 20대~50대 직장인 남녀 10,00명

출처 : 온라인 여행사 익스피디아

결론 :

현재 많은 소비자들이 여행 산업에 소비를 하려는 Needs는 있으나,  
그 여행이 국내여행이 아닌 해외여행을 선호하는 경향이 있다.

## 왜 만족도 인가?

### 국내 여행에 대한 부정적 인식

- \* 해외 여행이 더 즐길 거리가 많다.
- \* 국내는 여행 갈 곳이 마땅치 않다.
- \* 국내 여행은 언제든지 갈 수 있다.
- \* 국내 여행은 새로움이 부족하다.
- \* 국내 여행은 비용이 많이 든다.



### 기존의 여행 추천 시스템

- \* 여행의 트렌드에 따른 코스 추천.
- \* 여행사 일정에 따른 코스 추천.
- \* 개인의 특성, 여행 목적과 무관한 코스 추천.
- \* 개인의 인구통계학적 특성만을 고려한 코스 추천. (같은 40대라도, 가족과, 친구와, 혼자 여행가는 사람의 목적과 만족하는 포인트가 다름.)



### 새로운 여행 추천 시스템

- \* 부정적인 국내여행 이미지  
→ 국내 여행의 만족 의견을 통한 긍정적 이미지 구축
- \* 획일화된 여행 시스템  
→ 다중대응 분석을 통해서 목적에 따른 세세한 여행 시스템 구축

기존 데이터 처리 및 변환

1. 데이터 모으기

\* 기존 데이터 간의 Key value를 통해서 조인.

2. 복수 응답 문항

\* 이진수 형태로 변환.  
Ex) 여행을 어느 시기에 가십니까? 라는 물음에  
응답으로 주중, 주말, 방학, 휴일, 명절, 공휴일로  
복수 응답이 가능한 문항을 0, 1로 형태로 변환함.

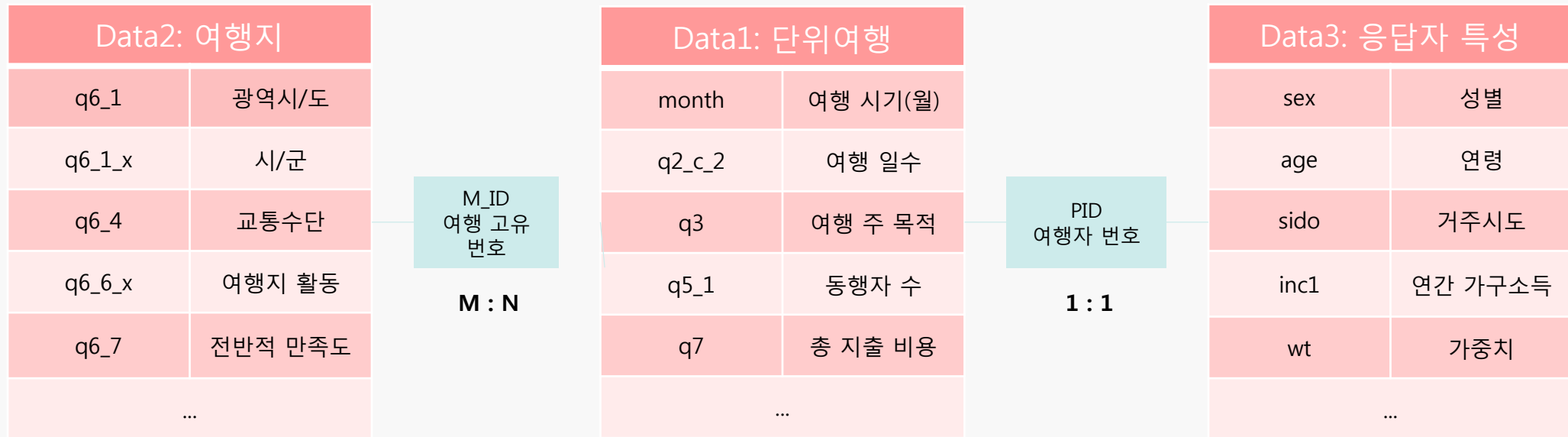
3. 범주형 변수

\* 범주형 변수는 가변수 처리( One Hot 인코딩)  
Ex) 1월부터 12월을 범주형 변수에서 12개의 컬럼 새로 생성함.

4. 목표 변수 설정하기

\* Y라는 목표 변수 설정.  
전반적 만족도, 재방문의향, 타인추천의향의 세가지 변수의 평균 값을 통해  
score라는 변수 창출 후, 범주화 시킴.

## 1-1. 주요 변수 설명



2. Understanding/ Preparation

1-2. 데이터 모으기

	gubun	HID	PID	type1	type2	month	M_ID	q1	q2_a	q2_a_1	...	q10_4	q10_5	q10_6	q10_7	q10_8	q10_9	q10_10	q10_11	q10_12
1	1.0	10001	1000101	1	2.0	7.0	101372.0	2.0	2017.0	7.0	...	4.0	4.0	9.0	4.0	4.0	4.0	4.0	4.0	6201.
2	2.0	10001	1000101	1	2.0	8.0	115439.0	2.0	2017.0	8.0	...	4.0	4.0	9.0	4.0	4.0	4.0	4.0	4.0	6201.
4	1.0	10001	1000101	1	2.0	10.0	115437.0	1.0	2017.0	10.0	...	9.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	6201.
8	1.0	10001	1000102	1	2.0	7.0	101372.0	2.0	2017.0	7.0	...	4.0	4.0	9.0	4.0	4.0	4.0	4.0	4.0	5886.
9	1.0	10001	1000102	1	2.0	8.0	101365.0	1.0	2017.0	8.0	...	9.0	4.0	9.0	4.0	4.0	4.0	4.0	3.0	5886.

5 rows x 75 columns

▲ Data1 : 여행에 대한 정보만 있음.  
Ex) 누구와 함께 갔는지, 며칠을 묵었는지.

	HID	PID	type1	type2	month	M_ID	q6_1	q6_1_1	q6_1_2	q6_1_3	...	q6_6_17	q6_6_18	q6_6_19	q6_6_20	q6_6_21	q6_7	q6_8	q6_9	...
0	10001	1000101	1	2.0	8.0	115439.0	31.0	10.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0	1000
4	10001	1000101	1	2.0	7.0	101372.0	32.0	30.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0	6500
5	10001	1000101	1	2.0	10.0	115437.0	11.0	50.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0	600
9	10001	1000102	1	2.0	7.0	101372.0	32.0	30.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0	6500
10	10001	1000102	1	2.0	10.0	115437.0	11.0	50.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0	600

5 rows x 72 columns

▲ Data2 : 여행지에 대한 정보만 있음.  
Ex) 어떤 지역으로 갔는지, 이동수단은 어떤 것인지

	gubun	HID	PID	type1	type2	month	M_ID	q1	q2_a	q2_a_1	...	q6_6_16	q6_6_17	q6_6_18	q6_6_19	q6_6_20	q6_6_21	q6_7	q6_8	q6_9
0	1.0	10001	1000101	1	2.0	7.0	101372.0	2.0	2017.0	7.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
1	2.0	10001	1000101	1	2.0	8.0	115439.0	2.0	2017.0	8.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
2	1.0	10001	1000101	1	2.0	10.0	115437.0	1.0	2017.0	10.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
3	1.0	10001	1000102	1	2.0	7.0	101372.0	2.0	2017.0	7.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
4	1.0	10001	1000102	1	2.0	8.0	101365.0	1.0	2017.0	8.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
5	1.0	10001	1000102	1	2.0	8.0	115434.0	1.0	2017.0	8.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
6	1.0	10001	1000102	1	2.0	10.0	115437.0	1.0	2017.0	10.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
7	2.0	10001	1000102	1	2.0	12.0	115441.0	1.0	2017.0	12.0	...	NaN	NaN	NaN	NaN	NaN	NaN	4.0	4.0	4.0
8	1.0	10002	1000201	1	1.0	5.0	110564.0	1.0	2017.0	5.0	...	NaN	NaN	NaN	NaN	NaN	NaN	3.0	3.0	3.0
9	1.0	10002	1000202	1	1.0	5.0	110564.0	1.0	2017.0	5.0	...	NaN	NaN	NaN	NaN	NaN	NaN	3.0	3.0	3.0

10 rows x 140 columns

▲ Data3  
Data3 = Data1 , Data2 의 두 개의 데이터를 M\_ID로 조인.

	HID	PID	sex	age	sr_type	sido	ara_size	school1	school2	occ0	occ1	occ2	fac	marry	inc1_1	inc2_1	wt
0	10001	1000101	1	40	2	1	1	5	5	NaN	2	8	4	2	4000	333.333333	6201.607951
1	10001	1000102	2	40	2	1	1	4	1	NaN	11	9	4	2	4000	333.333333	5886.547371
2	10002	1000201	1	41	1	1	1	6	1	NaN	4	3	3	1	4000	333.000000	12044.610408
3	10002	1000202	2	73	1	1	1	2	1	NaN	11	9	3	3	4000	333.000000	5436.667739
4	10002	1000203	2	46	1	1	1	4	1	NaN	9	5	3	2	4000	333.000000	12405.209670

▲ Data4 : 여행자들의 개인 정보가 있음.  
Ex) 여행자들의 나이, 교육 수준

	gubun	HID	PID	type1	type2	month	M_ID	q1	q2_a	q2_a_1	...	school1	school2	occ0	occ1	occ2	fac	marry	inc1_1	inc2_1	wt
0	1	10001	1000101	1	2	7	101372	2	2017	7	...	5	5	NaN	2	8	4	2	4000	333.333333	6201.607
1	2	10001	1000101	1	2	8	115439	2	2017	8	...	5	5	NaN	2	8	4	2	4000	333.333333	6201.607
2	1	10001	1000101	1	2	10	115437	1	2017	10	...	5	5	NaN	2	8	4	2	4000	333.333333	6201.607
3	1	10001	1000102	1	2	7	101372	2	2017	7	...	4	1	NaN	11	9	4	2	4000	333.333333	5886.547
4	1	10001	1000102	1	2	8	101365	1	2017	8	...	4	1	NaN	11	9	4	2	4000	333.333333	5886.547

5 rows x 156 columns

▲ Data5  
Data5 = Data3 , Data4 의 두 개의 데이터를 PID로 조인.

※ 특정 여행자들이 떠난 여행의 자세한 사항들을 알 수 있게 됨.



2. Understanding/ Preparation

2. 복수 응답 문항

gubun	type2	month	q1	q2_c_1	q2_c_2	q2_1_1	q2_1_2	q2_1_3	q2_1_4	...	sex	age	sido	ara_size	school1	occ1	occ2	fac	marry	inc1_1	
0	1	2	7	2	6	7	1.0	2.0	3.0	NaN	..	1	40	1	1	5	2	8	4	2	4000
1	2	2	8	2	1	2	1.0	2.0	NaN	NaN	..	1	40	1	1	5	2	8	4	2	4000
2	1	2	10	1	0	1	NaN	NaN	NaN	4.0	..	1	40	1	1	5	2	8	4	2	4000
3	1	2	7	2	6	7	1.0	2.0	3.0	NaN	..	2	40	1	1	4	11	9	4	2	4000
4	1	2	8	1	0	1	NaN	2.0	NaN	NaN	..	2	40	1	1	4	11	9	4	2	4000

5 rows × 80 columns



	gubun	type2	month	q1	q2_c_1	q2_c_2	q2_1_1	q2_1_2	q2_1_3	q2_1_4	...	sex	age	sido	ara_size	school1	occ1	occ2	fac	marry	inc1_1
0	1	2	7	2	6	7	1	1	1	0	...	1	40	1	1	5	2	8	4	2	4000
1	2	2	8	2	1	2	1	1	0	0	...	1	40	1	1	5	2	8	4	2	4000
2	1	2	10	1	0	1	0	0	0	1	...	1	40	1	1	5	2	8	4	2	4000
3	1	2	7	2	6	7	1	1	1	0	...	2	40	1	1	4	11	9	4	2	4000
4	1	2	8	1	0	1	0	1	0	0	...	2	40	1	1	4	11	9	4	2	4000

5 rows × 80 columns

◀ q2\_1 = 여행을 어느 시기에 가십니까? 라는 질문에  
응답으로  
q2\_1\_1 : 주중  
q2\_1\_2 : 주말  
q2\_1\_3 : 방학  
q2\_1\_4 : 휴일 등으로 복수 응답이 가능한 문항.

◀ q2\_1\_1 ~ q2\_1\_5 변수의 값을  
NaN에서 0, 1로 이진수 처리 함.

※ 1, 2, 3의 모두 'yes'의 의미를 갖지만, ex) '1' < '3' 으로 변수 해석의 오류를 없앴.

3. 범주형 변수

month	
0	7
1	8
2	10
3	7
4	8
5	8
6	10
7	12
8	5
9	5



	month_d_1	month_d_2	month_d_3	month_d_4	month_d_5	month_d_6	month_d_7	month_d_8	month_d_9	month_d_10	month_d_11	month_d_12
0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1
8	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	0	0

▲ month = 몇 월 달에 여행을 갔는가?

Ex) 7 → 7월 달  
10 → 10월 달

▲ 가변수 처리.  
Month\_d\_1 = 1월에 여행을 갔는가  
0 : 가지 않았다.  
1 : 갔다. 이런 방식으로 총 12개의 새로운 변수 생성 (one hot 인코딩 방법)

4. 목표 변수 설정

q6_7	전반적 만족도	1. 매우 불만족 2. 불만족 3. 보통 4. 만족 5. 매우 만족
q6_8	재방문 의향	1. 매우 낮음 2. 낮음 3. 보통 4. 높음 5. 매우 높음
q6_9	타인 추천의향	1. 매우 낮음 2. 낮음 3. 보통 4. 높음 5. 매우 높음

▲ 원본 데이터변수 설명

※ 범주화 이유 :  
설문조사의 만족도 평가 중 긍정적인 응답자  
상위 20%로 선정.  
→ 정확한 만족도와 신뢰성을 얻을 수 있음.

※ 목표 변수를 '전반적 만족도', '재방문 의향', '타인 추천의향'의  
평균 점수를 범주화하여 사용. 이를 통해 다각도의 만족도 척도  
를 통합적으로 고려하고자 함.

	gubun	type2	month	q1	q2_c_1	q2_c_2	q2_1_1	q2_1_2	q2_1_3	q2_1_4	...	age	sido	ara_size	school1	occ1	occ2	fac	marry	inc1_1	score
0	1	2	7	2	6	7	1	1	1	0	...	40	1	1	5	2	8	4	2	4000	4.0
1	2	2	8	2	1	2	1	1	0	0	...	40	1	1	5	2	8	4	2	4000	4.0
2	1	2	10	1	0	1	0	0	0	1	...	40	1	1	5	2	8	4	2	4000	4.0
3	1	2	7	2	6	7	1	1	1	0	...	40	1	1	4	11	9	4	2	4000	4.0
4	1	2	8	1	0	1	0	1	0	0	...	40	1	1	4	11	9	4	2	4000	4.0

5 rows x 81 columns

▲  $score = (q6_7 + q6_8 + q6_9) / 3$  으로 계산

	gubun	type2	month	q1	q2_c_1	q2_c_2	q2_1_1	q2_1_2	q2_1_3	q2_1_4	...	ara_size	school1	occ1	occ2	fac	marry	inc1_1	score	dist	y
1	1	2	7	2	6	7	1	1	1	0	...	1	5	2	8	4	1	4000	4.0	124091.487104	False
2	2	2	8	2	1	2	1	1	0	0	...	1	5	2	8	4	1	4000	4.0	74983.955873	False
1	1	2	10	1	0	1	0	0	0	1	...	1	5	2	8	4	1	4000	4.0	0.000000	False
1	1	2	7	2	6	7	1	1	1	0	...	1	4	11	9	4	1	4000	4.0	124091.487104	False
1	1	2	8	1	0	1	0	1	0	0	...	1	4	11	9	4	1	4000	4.0	0.000000	False

▲ y 라는 값으로 score의 값이  
- 4점 초과 인 경우 : 'True'  
- 4점 이하 인 경우 : 'False' 으로 범주화함.

#### 4가지 트리 기반 모델 선정 이유

\* 데이터 분석을 통해 여행 만족도를 높이는 요인을 찾고, 이를 여행 상품 개발에 활용하고자 함.

\* 만족도가 높은 여행과 낮은 여행을 구분하는 모델을 학습시켜, 그 모델의 분류법을 분석하여 만족도 상승 요인을 알아내고자 함.

\* 모델의 계산 과정과 변수 선정을 참고하여 실제 여행 상품을 도출하여야 하므로, 설명력이 높은 모델이 필요함.

##### \* SVM, Neural Network 등의 모델

Black box와 같음. 모델이 높은 정확도를 가지더라도 어떠한 요인이 작용하는지 알기 어려움.

##### \* 회귀분석

가격과 만족도, 동행인 수와 만족도는 꼭 선형 관계가 있다고 보기 어려움.  
대부분의 변수가 범주형 변수임.

**결론 : Tree 기반 알고리즘들을 활용.**

##### \* 의사결정 Tree

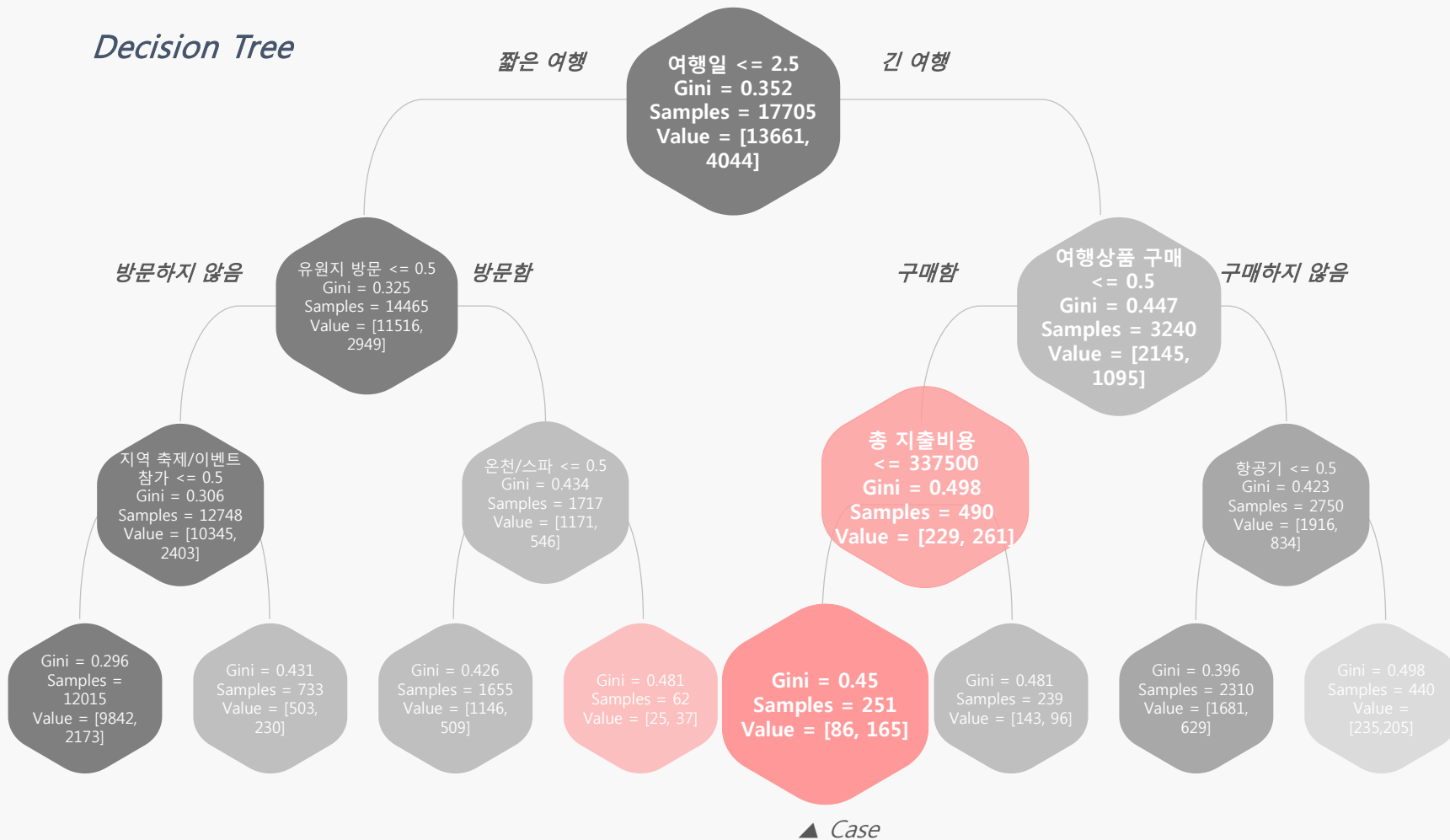
정확도는 낮으나, 만족도가 높아지는 조건의 집합을 정확히 알 수 있음.

##### \* XGBoost, AdaBoost, RandomForest 등 Ensemble method

각 요인의 기준을 정확히 알기는 다소 어렵으나 변수 중요도를 통해 설명력을 확보할 수 있고 높은 정확도를 가짐.

### 3. Modeling

#### Decision Tree



#### ◀ 티켓 층 : 전체 data set



홍길동(성별 나이 무관)

#### \* 가장 만족도가 높은 Case

긴 여행(2박 3일 이상)이면서 패키지 여행이며 여행경비를 약 30만원 이하로 사용한 경우 만족도가 높아진다.

#### 조건의 통계적 유의미함을 검정함

◀ Decision Tree 상세과정(1번 트리):

사용한 툴 = Python 3.6, scikit-learn 0.19.1

train : test = 3 : 1

Parameters: max\_depth=3, min\_samples\_split=0.05

선정한 군집과 전체 군집의 모비율 비교:

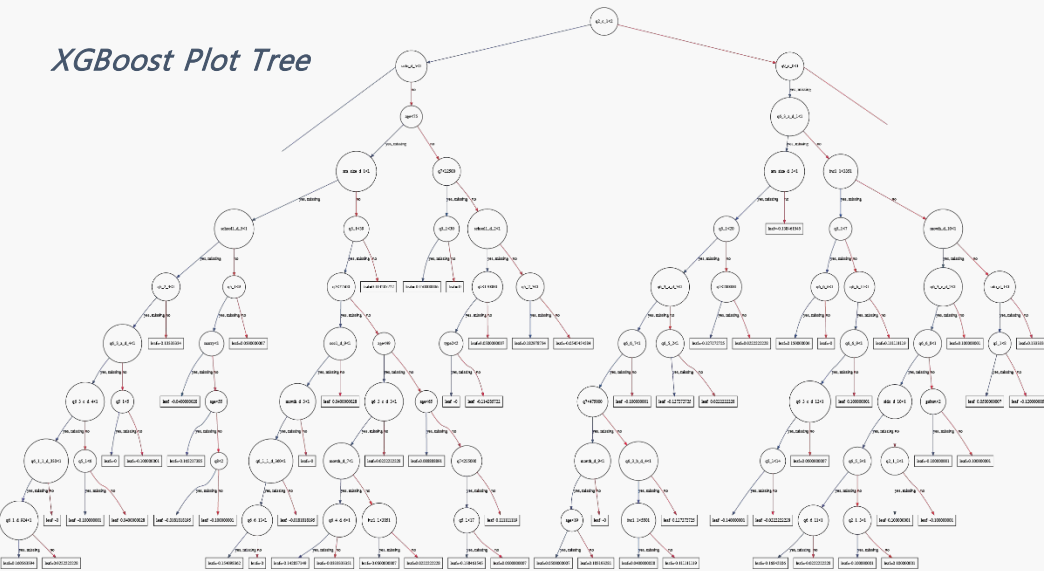
z-score= 15.930

p-value= 1.964e-57

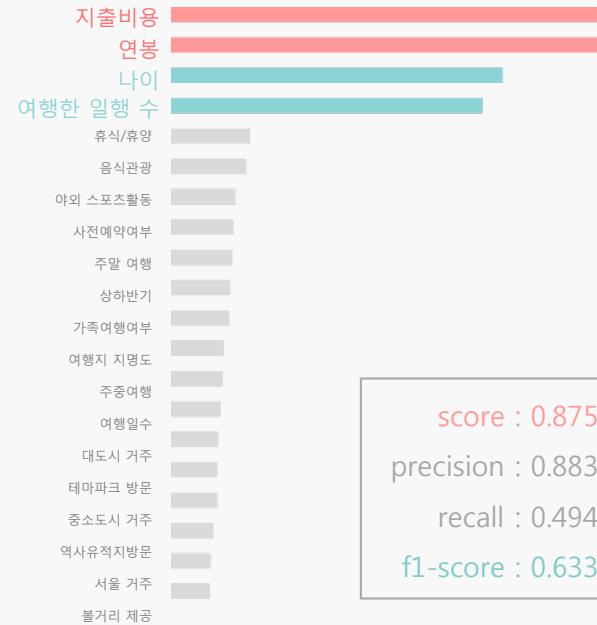
### 3. Modeling

## Xgboost & AdaBoost & RandomForest

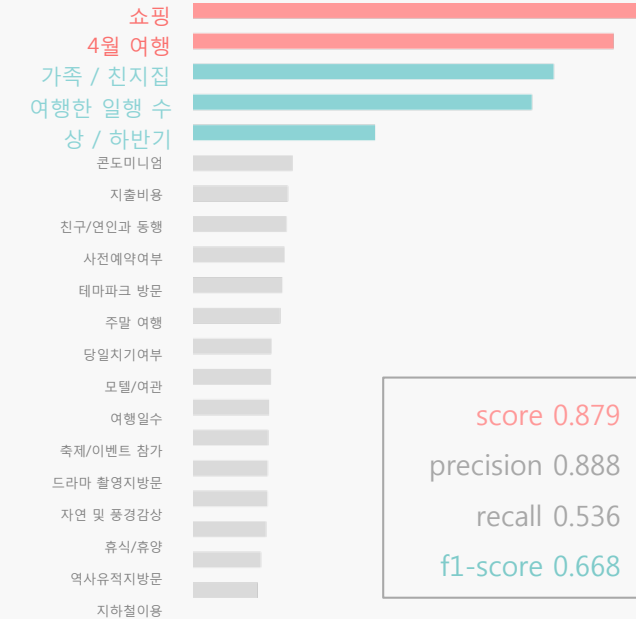
XGBoost Plot Tree



XG Boost 변수 중요도 그래프



Ada Boost 변수 중요도 그래프



\* 앙상블 기반의 알고리즘들은 높은 정확도와, 이상치에 영향을 더 적게 받는다는 장점이 있음.

\* 실제 모델링 후 모델 간의 평가에서 상대적으로 높은 예측 정확도를 얻음.

\* 데이터 기초분석 결과와의 교차 검증을 통해 확인한 결과,  
의사결정나무 모델이 만족도 향상에 더 효과적인 조건 집합을 도출해 내는 것으로 드러나  
**의사결정나무 모델을 기준으로 사용하기로 결정함.**

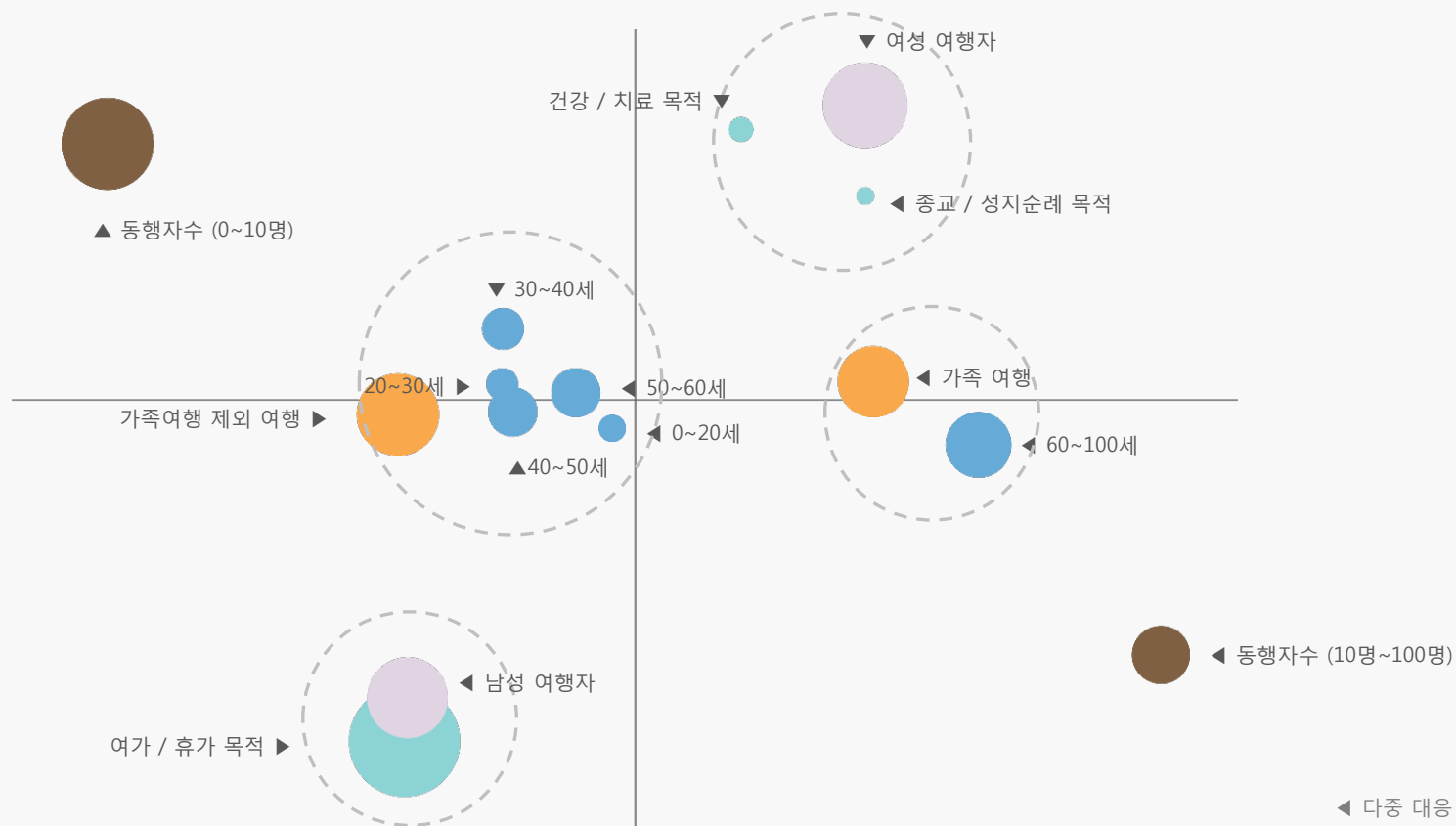
▲ XGBoost 상세과정:

사용한 툴 = Python 3.6, xgboost 0.80  
train : test = 3 : 1  
Parameters: max\_depth=10,  
learning\_rate=0.1, n\_estimators=100,  
objective="binary:logistic"

▲ AdaBoost 상세과정:

사용한 툴 = Python 3.6,  
scikit-learn 0.19.1  
train : test = 3 : 1  
Parameters: base\_estimator=(Decision Tree),  
algorithm='SAMME', n\_estimators=150

#### After Modeling - 다중 대응 분석



\* 60~100세 나이대의 여행자, 가족 여행.

\* 가족여행 X, 60~100 세 X.

\* 남성 여행자, 여행 목적은 여가 / 휴가.

다중 대응 분석으로 ,  
세가지 Target 층을 선정 함.

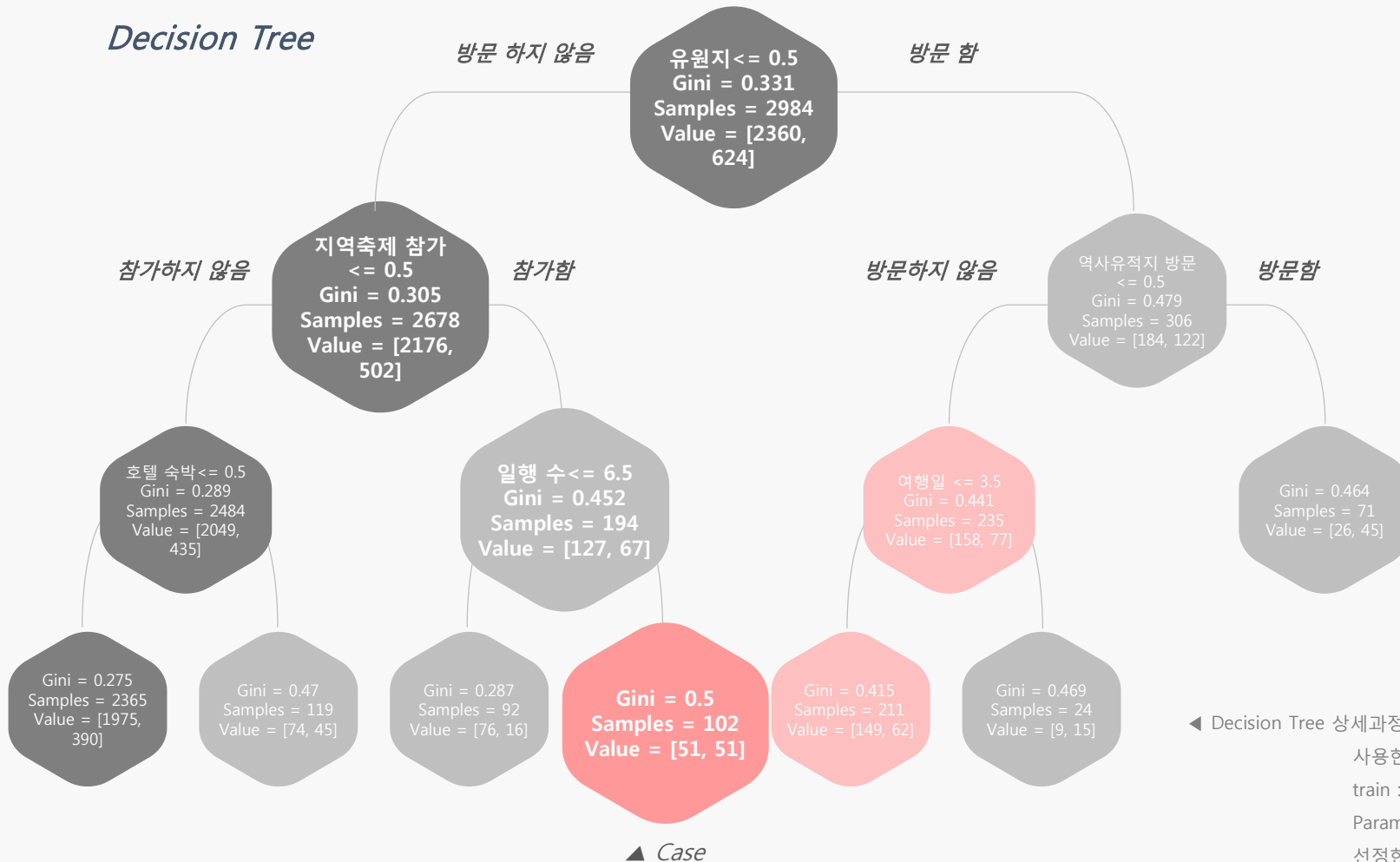
◀ 다중 대응 분석 상세 과정:

사용한 툴 = Python 3.6, scikit-learn 0.19.1


여행 데이터에서 5개 항목(성별, 나이, 여행목적, 가족여행여부, 동행자수) 선정하여,  
정규화(평균 0, 분산 1) 후 주성분 분석 -> 상위 2개의 성분의 계수를 분석

### 3. Modeling

#### Decision Tree



#### ▲ 티켓 층 :

 최불암 ( 79세, 남 )  
60~100대 나이 대, 가족 여행 위주

#### \* 가장 만족도가 높은 Case

주로 지역축제에는 참가하며, 본인을 제외한 일행 수가 7명 이상인 경우 만족도가 높아진다.

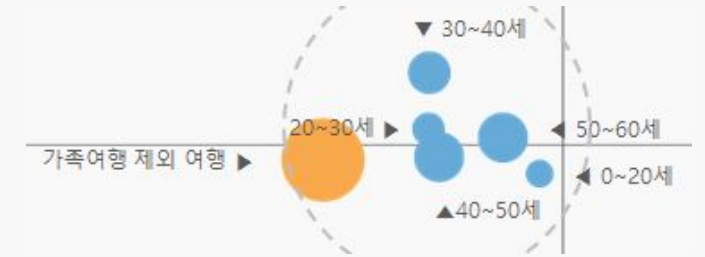
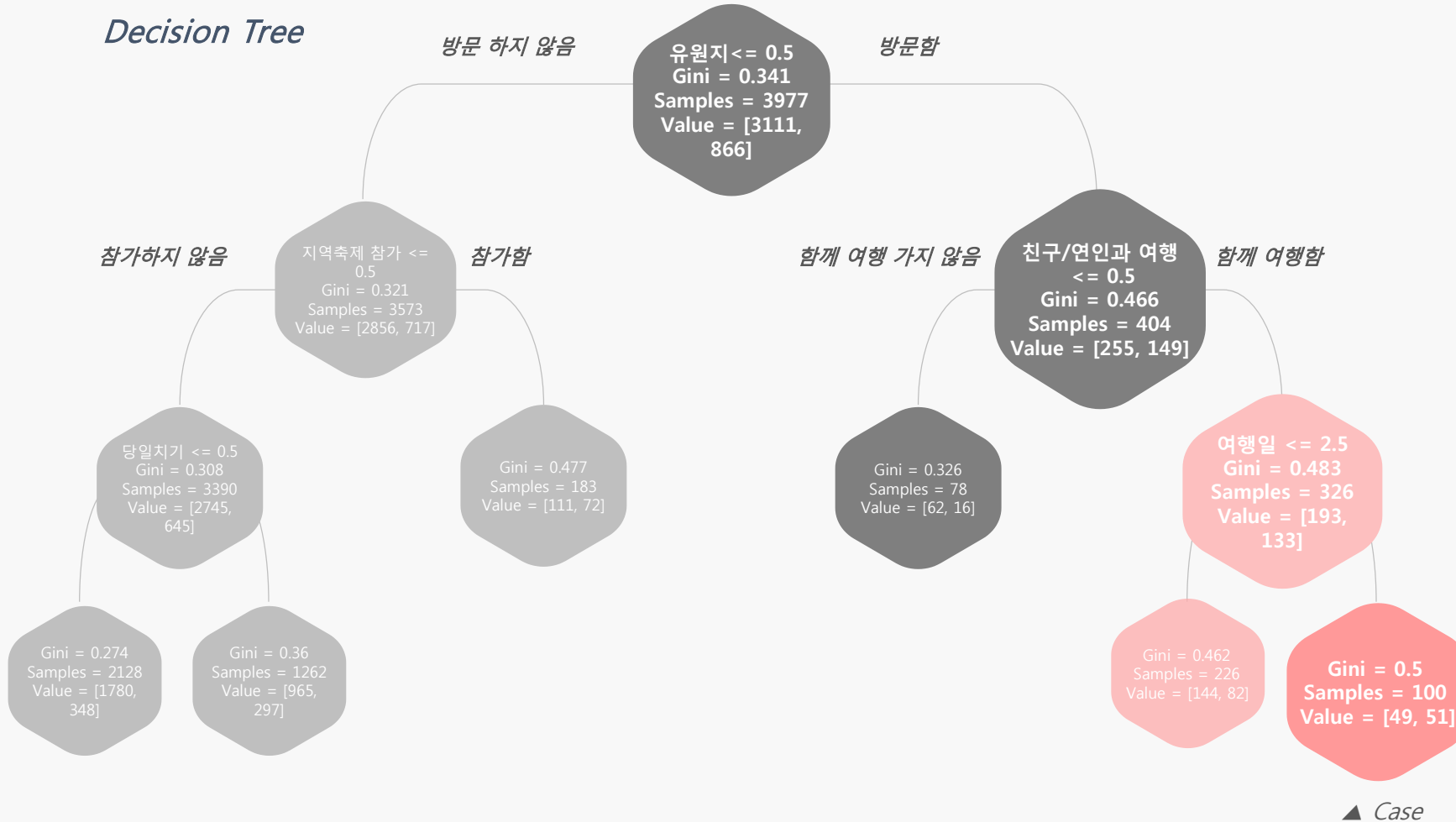
#### ◀ Decision Tree 상세과정(2번 트리):

사용한 툴 = Python 3.6, scikit-learn 0.19.1  
train : test = 3 : 1  
Parameters: max\_depth=3, min\_samples\_split=0.05  
선정한 군집과 전체 군집의 모비율 비교:  
z-score= 6.988  
p-value= 1.392e-12



### 3. Modeling

#### Decision Tree



#### ▲ 타겟 층 :



전현무 (40세, 남), 한혜진 (35세, 여)



젊은 나이 대, 가족 여행 외 다른 목적의 여행 선호

#### \* 가장 만족도가 높은 Case


유원지를 방문하고 친구 / 연인과 함께하며, 여행 일이 2박 3일 이상일 때 만족도가 높아진다.

#### ◀ Decision Tree 상세과정(3번 트리):

사용한 툴 = Python 3.6, scikit-learn 0.19.1  
 train : test = 3 : 1  
 Parameters: max\_depth=3, min\_samples\_split=0.05  
 선정한 군집과 전체 군집의 모비율 비교:  
 z-score= 6.913  
 p-value= 2.371e-12

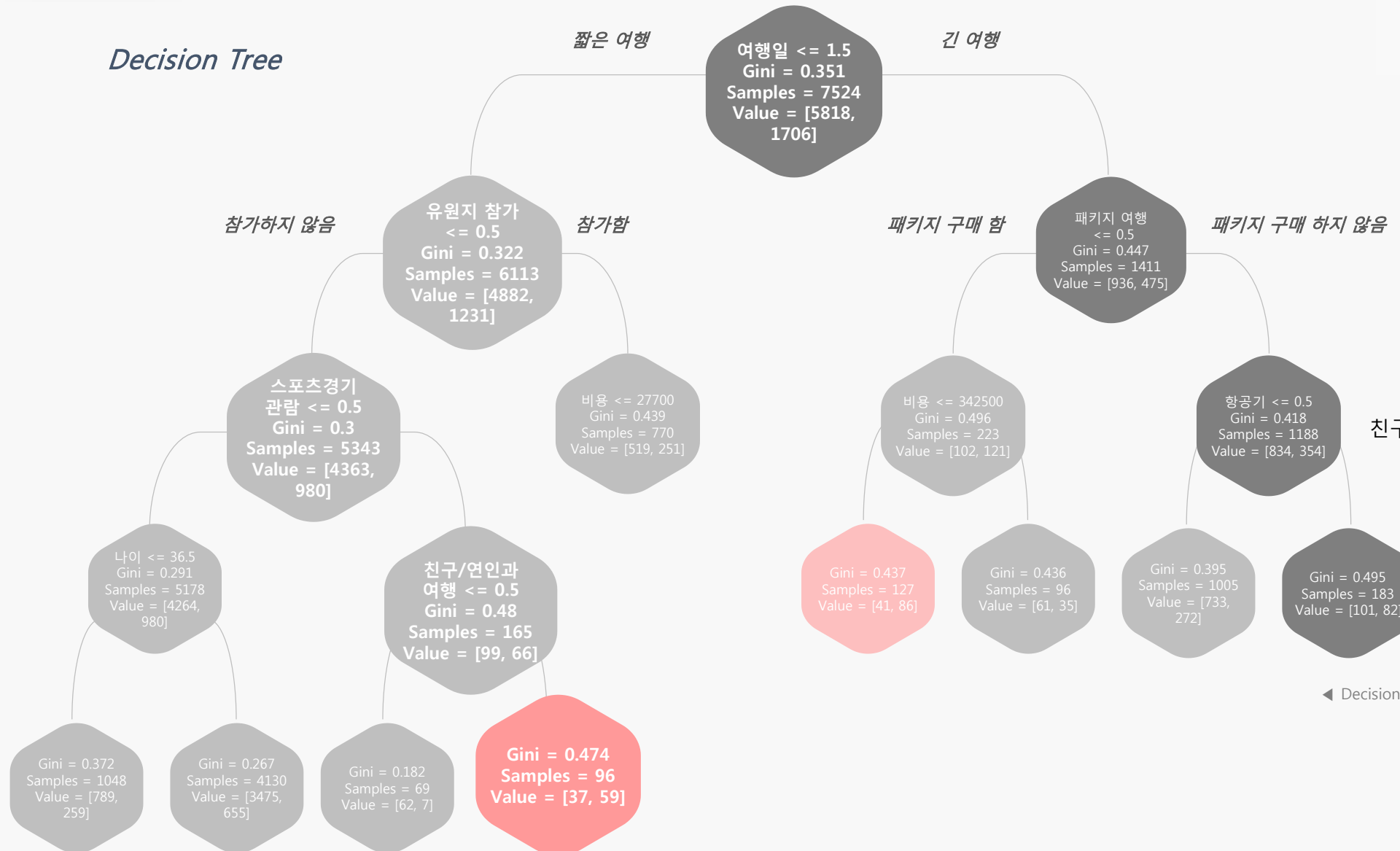
## Decision Tree



▲ **타겟 층 :**  
 박찬호 (45세, 남성)  
 남성 여행자, 여행 목적은  
 여가 / 휴가를 선호.

### \* 가장 만족도가 높은 Case

친구 / 연인과 함께 당일치기, 1박 2일로 스포츠 경기를 관람할 때 만족도가 높아진다.



▲ Case 4

◀ Decision Tree 상세과정(4번 트리):

사용한 툴 = Python 3.6, scikit-learn 0.19.1

train : test = 3 : 1

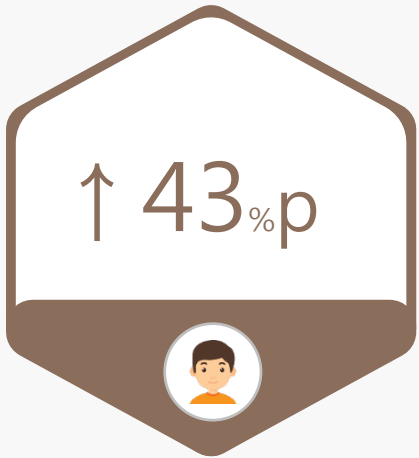
Parameters: max\_depth=4, min\_samples\_split=0.05

선정한 군집과 전체 군집의 모비율 비교:

z-score= 8.951

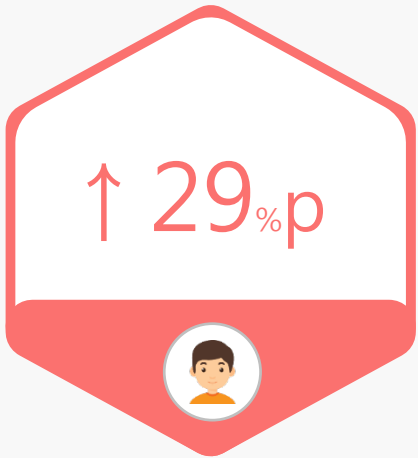
p-value= 1.766e-19

맞춤 여행 코스로 진행할 경우 기대할 수 있는 만족도 향상 수치



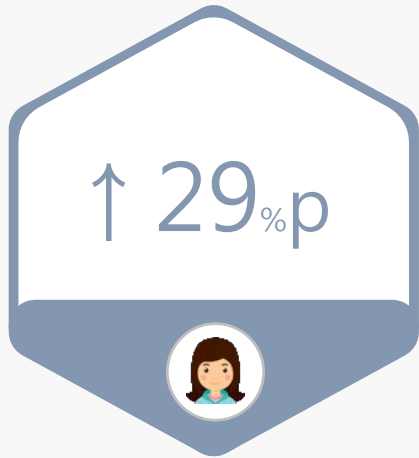
홍길동(나이무관)

" 나는 패키지 여행으로 2박 3일  
이상이면서 여행경비를 약 30만원  
이하인 여행을 떠나고 싶어"



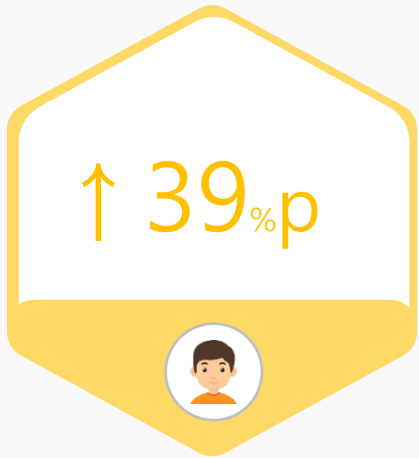
최불암(79세)

" 나는 지역축제에는 참가하고, 일행 수  
가 7명 이상인 여행을 떠나고 싶어 "



한혜진(35세)

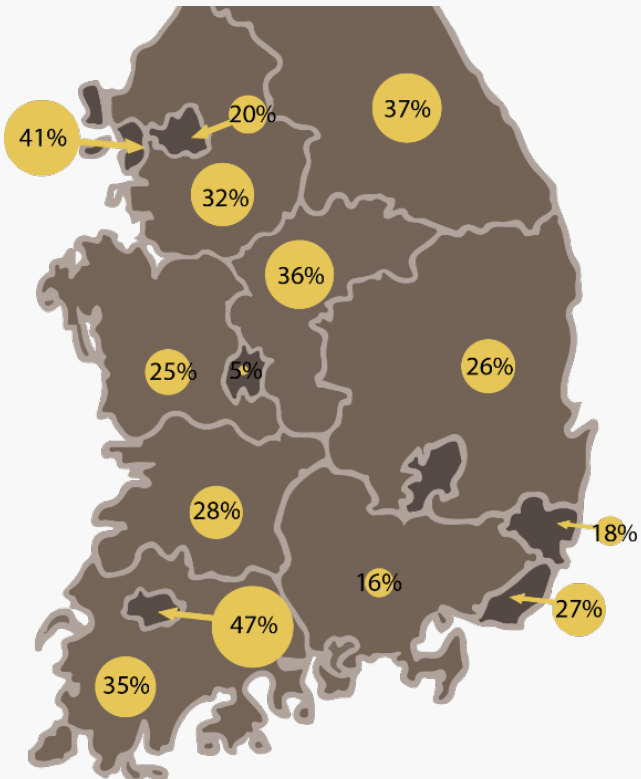
" 나는 친구 / 연인과 함께 유원지를  
방문하고 2박 3일 이상인 여행을  
떠나고 싶어 "



박찬호(45세)

" 나는 친구 / 연인과 함께 스포츠 경기를  
관람하는 짧은 여행을 떠나고 싶어 "

## 지역 선정 이유



※ 제주도 선정 이유 ?

- 국민여행실태조사 기초분석에 기반한 2박 3일 이상의 여행객의 가장 만족도가 높은 지역

- 주요 관광지 입장객통계에 기반한 인기 지역 선정

### '홍길동'씨를 위한 맞춤 추천 코스



## 1일차

A코스 : 제주도 민속자연사 박물관 ▶ 점심식사 ▶ 만장굴 ▶ 비자림

B코스 : 제주도 민속자연사 박물관 ▶ 점심식사 ▶ 한라산 국립 공원 ▶ 제주 자연 휴양림

## 2일차

A코스 : 성산일출봉 ▶ 아침식사 ▶ 한화 아쿠아 플래넷 ▶ 점심식사 ▶ 일출 랜드

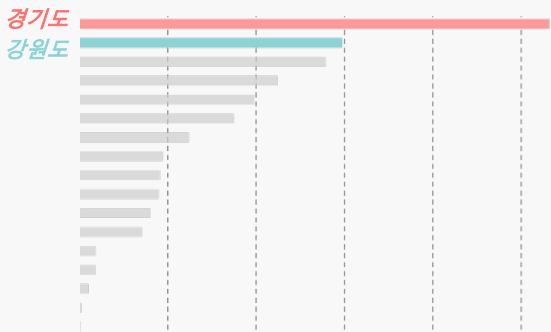
B코스 : 정방 폭포 ▶ 아침식사 ▶ 올레길 ▶ 점심식사 ▶ 서귀포 휴양림 ▶ 대포 주상 절리

### 3일차

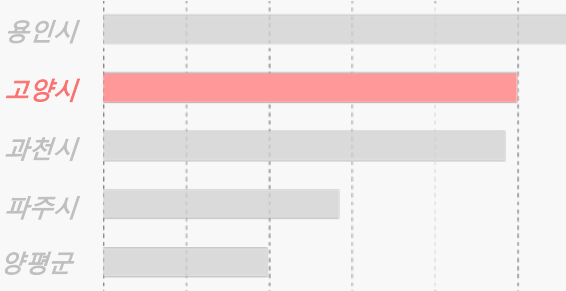
A코스 : 정방폭포 ▶ 아침식사 ▶ 올레길 ▶ 서귀포 휴양림 ▶ 점심식사 ▶ 대포 주상 절리

B코스 : 천제연 폭포 ▶ 아침식사 ▶ 산방산 ▶ 점심식사 ▶ 마라도 해양 공원

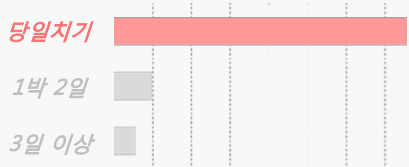
지역 선정 이유



▲ 상반기 상위 만족 지역 - 경기도



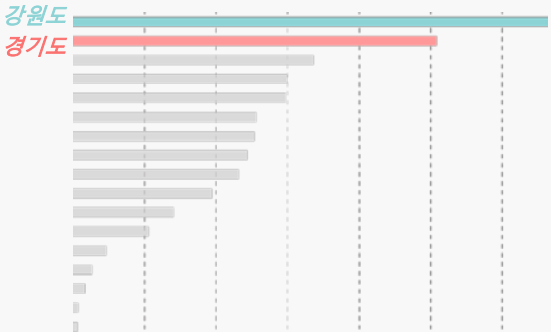
▲ 상반기 - 경기도 고양시



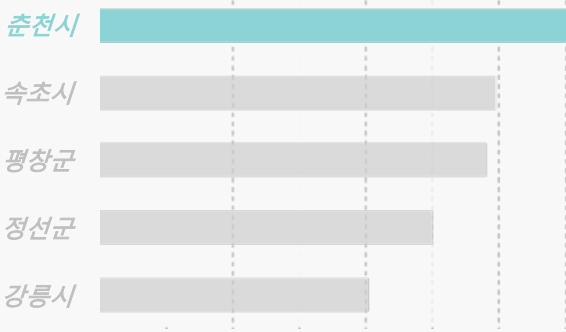
▲ 고양시 평균 여행일수

※ 경기도 고양시 당일치기 선정 이유 ?

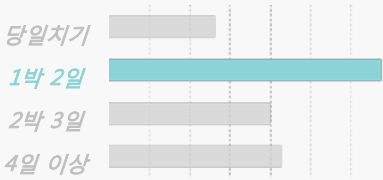
- 국민여행실태조사에서 만족한 사람들 데이터의 상반기 상위 만족 지역을 선정 → 관광객 입장 통계에서 '경기도' 지역의 상위 방문객 지역을 인기 지역으로 선정
- 경기도 고양시 관광객의 평균 여행일수 결과 반영



▲ 하반기 상위 만족 지역 - 강원도



▲ 하반기 - 강원도 춘천시



▲ 춘천시 여행일수 만족도

※ 강원도 춘천시 1박 2일 선정 이유 ?

- 국민여행실태조사에서 만족한 사람들 데이터의 하반기 상위 만족 지역을 선정 → 관광객 입장 통계에서 '강원도' 지역의 상위 방문객 지역을 인기 지역으로 선정
- 강원도 춘천시 관광객의 평균 여행일수와 만족도 결과 반영

'최불암' 씨를 위한 맞춤 추천 코스

\* 상반기 - 경기도 고양시



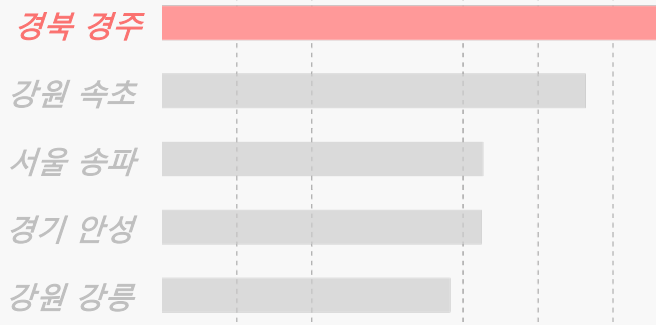
당일치기  
행주산성 기념제 ▶ 점심식사 ▶ 배다골 테마파크 ▶ 저녁식사 ▶ 메타세콰이어길

\* 하반기 - 강원도 춘천시

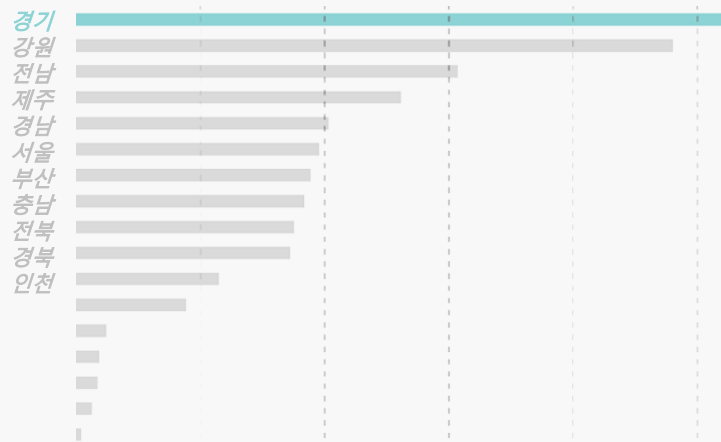


1일차 : 의암호 스카이워크 ▶ 점심식사 ▶ 소양강 문화제 ▶ 막국수거리(저녁식사)  
2일차 : 소양호 유람선 ▶ 청평사

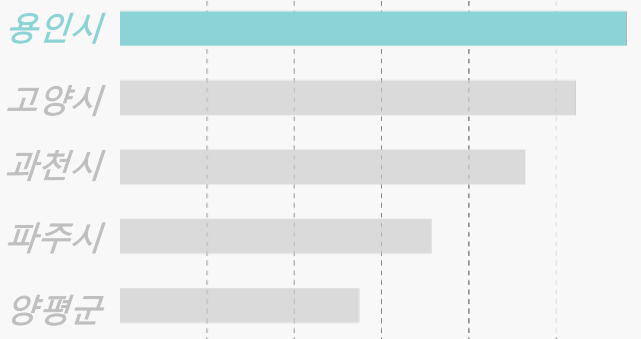
## 지역 선정 이유



▲ 유원지 방문자의 만족도



▲ 상하반기 통합 만족도 상위 지역



▲ 경기도 지역 상위 방문 지역

### ※ 경상북도 경주시 선정 이유 ?

- 국민여행실태조사의 유원지 방문자 중 만족한 사람들의 데이터의 상위 만족 지역을 선정

• 단, 다양한 여행지 선정 위해 제주 지역 제외

### ※ 경기도 용인시 선정 이유 ?

- 국민여행실태조사의 만족한 사람들 데이터의 상하반기 통합 상위 만족 지역을 선정 → 관광객 입장 통계에서 '경기도' 지역의 상위 방문객 지역을 인기 지역으로 선정한 결과 반영

'한혜진' 씨를 위한 맞춤 추천 코스

\* 경상북도 경주시



- 1일차 : 경주 월드 ▶ 점심식사 ▶ 경주 문화 엑스포 공원 ▶ 저녁식사  
2일차 : 아침식사 ▶ 석굴암 ▶ 불국사 ▶ 저녁식사  
3일차 : 통일전 ▶ 아침식사 ▶ 경주동궁원 ▶ 동궁과월지 ▶ 점심식사 ▶ 경주 양동 마을

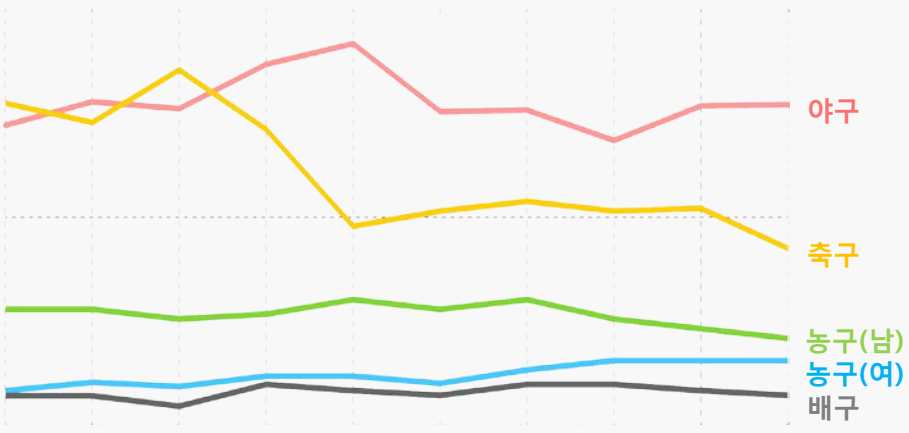
\* 경기도 용인시



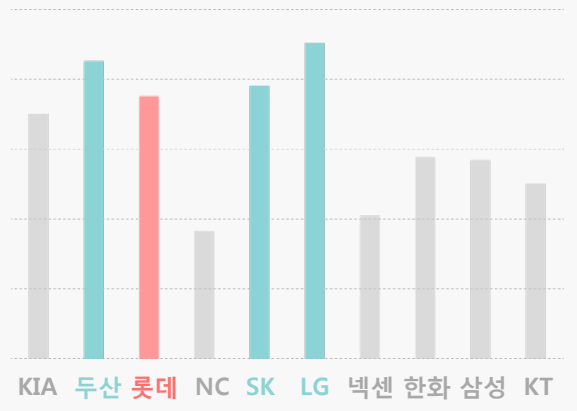
- 1일차 : 캐리비안베이 ▶ 점심식사 ▶ 에버랜드  
2일차 : 용인자연휴양림 ▶ 양지파인리조트 아쿠아펀 ▶ 용인농촌테마파크  
3일차 : 한국민속촌 ▶ 백남준아트센터 ▶ 호암미술관



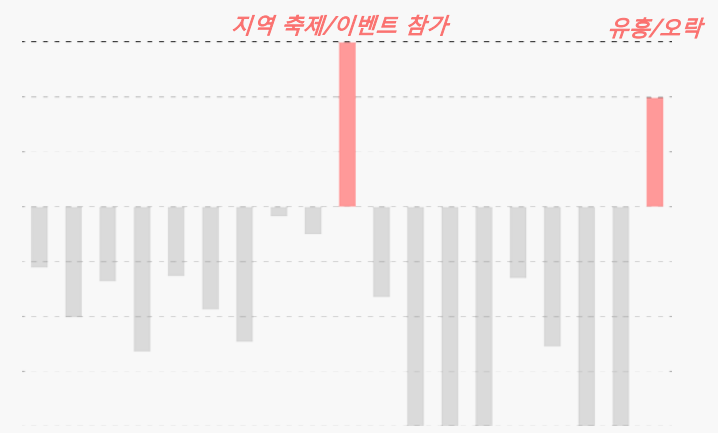
지역 선정 이유



▲ 연도별 프로스포츠 평균관객수 추이  
출처 문화체육관광부



▲ 야구 구단별 총관객수  
출처 문화체육관광부



▲ 여행지에서의 활동 상관 분석

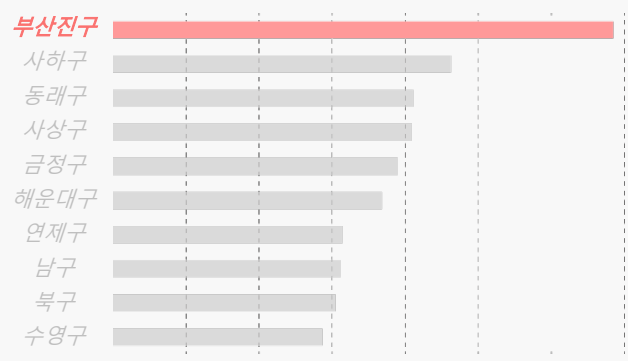
※ 부산, 인천 및 서울 선정 이유?

- 스포츠 평균관객수 추이를 통해 '야구' 종목의 인기가 가장 많은 것을 반영
- 구단별 총관객수를 통해 '롯데' 구단이 있는 부산과 '두산', 'LG' 구단이 있는 서울, 'SK' 구단이 있는 인천을 선정.

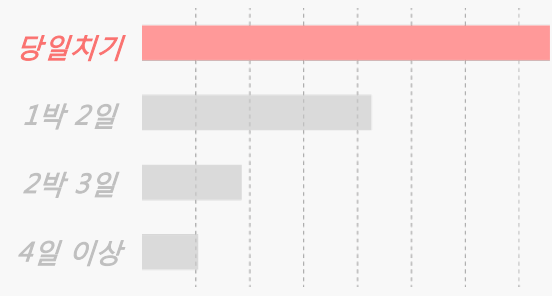
※ 주요 여행 코스 선정 고려 사항

- 스포츠 경기 관람과 다른 여행지에서의 활동을 상관 분석한 결과 '지역 축제/이벤트 참가'와 '유흥/오락'이 상관성이 있다는 것을 발견

지역 선정 이유



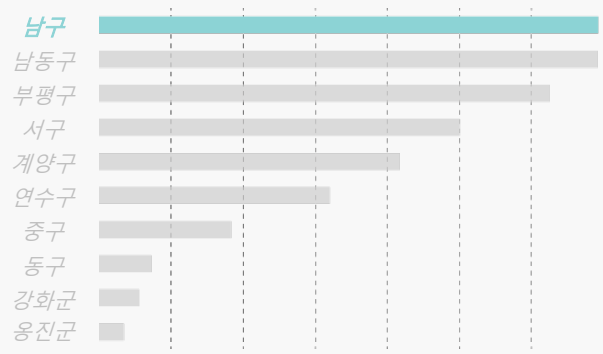
▲ 부산광역시 구별 유흥/오락 발달 지역



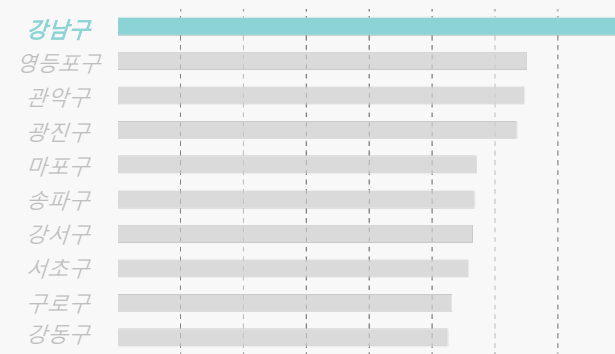
▲ 부산광역시 평균 여행일수

※ 부산광역시 당일치기 선정 이유 ?

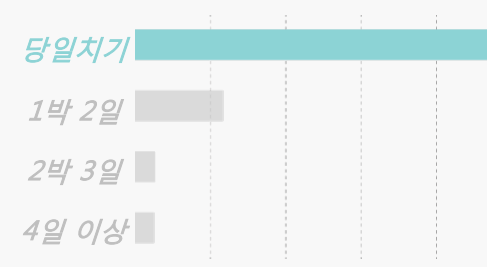
- 소상공인 상권정보 상가업소 데이터에서 '관광/여가/오락'과 '유흥주점'의 구별 개수를 반영하여 발달 지역을 선정.
- 국민여행실태조사의 부산광역시 관광객의 평균 여행일수 결과 반영.



▲ 인천시 구별 유흥/오락 발달 지역



▲ 서울시 구별 유흥/오락 발달 지역



▲ 서울 + 인천 평균 여행일수

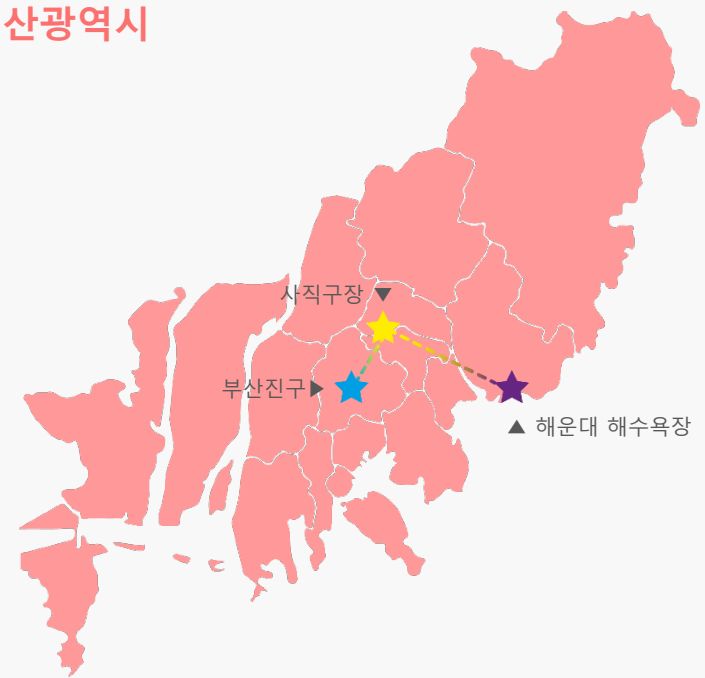
※ 인천 및 서울 당일치기 선정 이유?

- 인천시와 서울시의 소상공인 상권정보 상가업소 데이터에서 '관광/여가/오락'과 '유흥주점'의 구별 개수를 반영하여 발달 지역을 선정.
- 국민여행실태조사의 서울 및 인천의 관광객 평균 여행일수 결과 반영.

'박찬호'씨를 위한 맞춤 추천 코스

\* 비수도권 추천

\* 부산광역시



당일치기  
부산바다 축제(해운대) ▶ 점심식사 ▶ 사직구장 ▶ 부산진구(유흥/오락)

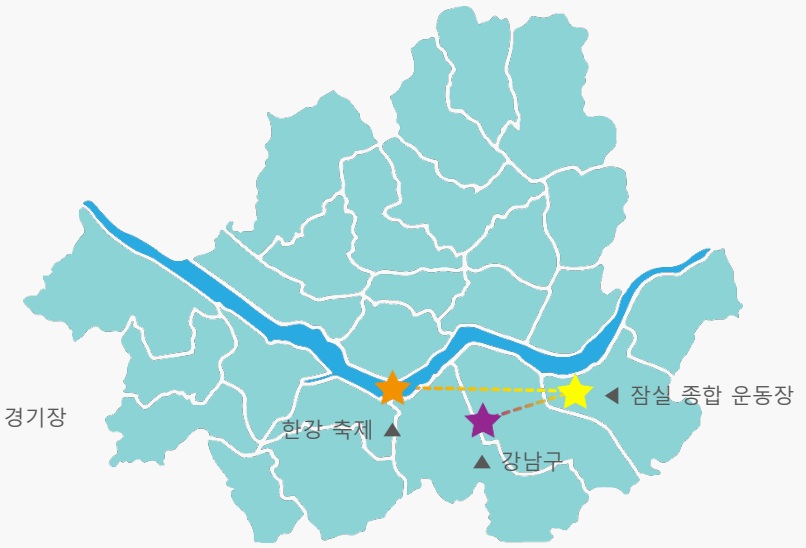
\* 수도권 추천

\* 인천광역시



당일치기  
인천코스 : 송도 센트럴파크 ▶ 인천 문학 경기장 ▶ 미추홀구(유흥/오락)

\* 서울특별시



서울코스 : 한강 축제(반포 및 여의도) ▶ 잠실 종합 운동장 ▶ 강남구(유흥/오락)

## 1. VR 체험



◀ VR로 메타세콰이어 길 체험



◀ VR로 불국사 체험

### ※ VR 체험 기대효과

1. 국내 관광 홍보 기능
2. 미리 여행 코스를 체험 함으로서,  
국내 여행의 부정적 인식(다양한 볼거리의  
부재) 개선
3. 본인의 취향에 따른 맞춤형 여행 코스 선택지  
기능.

## 2. 서포터즈 창단

**제 1기 국내여행 서포터즈 모집!**  
**SNS 홍보단 적극환영!**

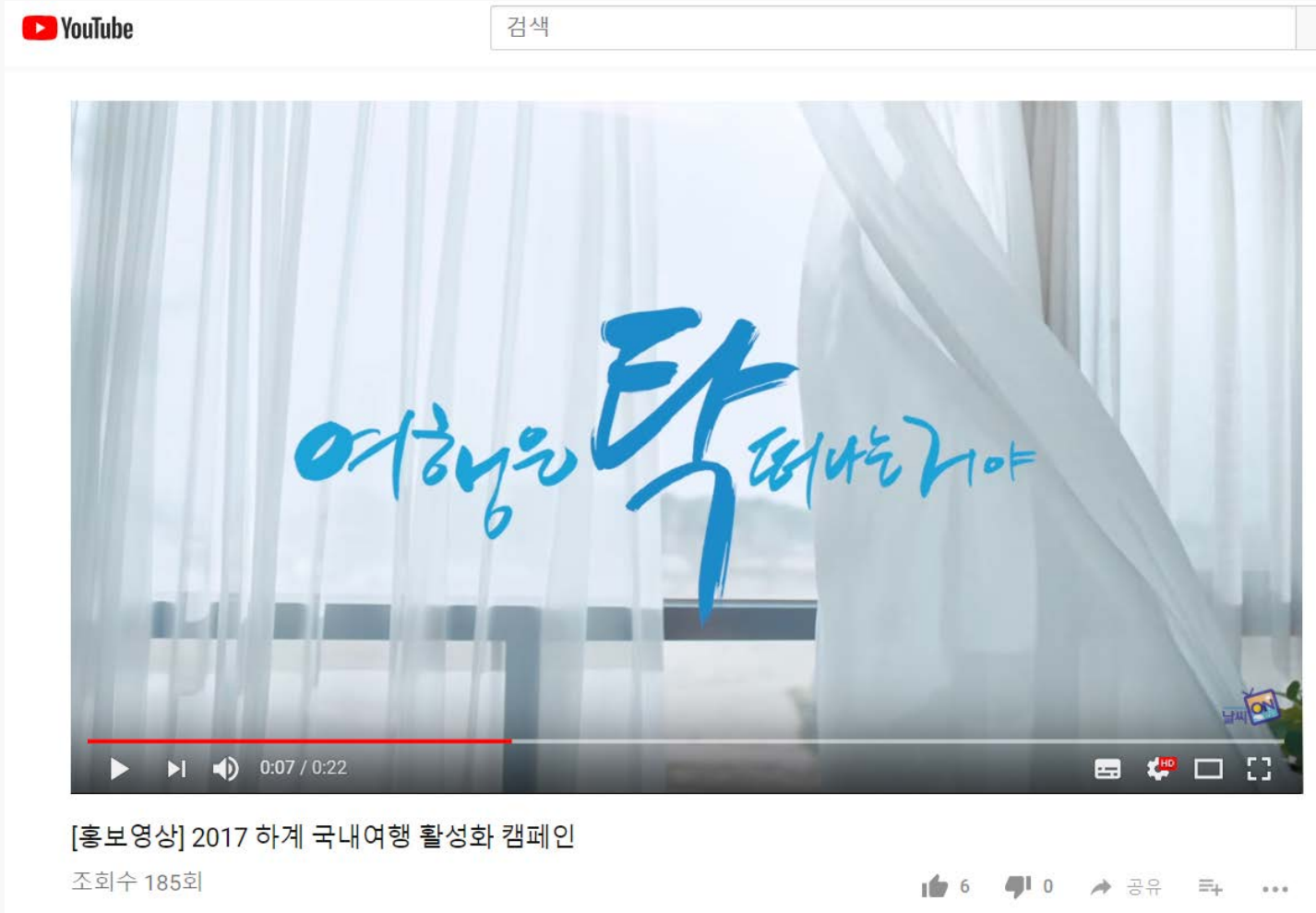


**모집기간** 2018. 08. 27 ~ 2018. 09. 07  
**모집대상 및 인원** 여행코스 그룹별 10명  
**홍보대상** 국내외 여행객  
**활동내용** 개인 SNS 홍보, 콘텐츠 제작 등  
**홍보단 혜택** 소정의 활동비, 수료증, 봉사시간 지급 등  
**지원방법** 공지사항 참조  
**문의** 010 - 1234 - 5678

※ 국내 여행 서포터즈 창단의 기대효과

1. 서포터즈들의 여행 후기를 담은 SNS 홍보 기능
2. 서포터즈를 통해 미리 여행 코스를 체험 → 여행 코스의 미흡한 점 개선 가능.
3. 지역 발전 이바지 및 내수 산업 활성화.

### 3. 홍보 동영상 제작



※ 현재 국내 여행 활성화를 위한 높은 퀄리티의 동영상이 제작되고 있으나, 다소 아쉬운 점이 있다.

이에 몇 가지 제안을 하고자 한다.

1. 국내 여행자들의 리얼한 후기 추가.
2. 그룹별 맞춤 동영상 제작.  
→ 다양한 버전 출시.

<b>SWOT Analysis</b>	<b>Strength</b> 강점	만족도를 높일 수 있는 요인을 데이터 기반으로 선정. 그룹으로 나눠서 각각 그룹에 특성에 맞는 여행 코스를 제안. 기존 국내 여행의 부정적 인식 및 이미지 타파
	<b>Weakness</b> 약점	가격을 고려 하지 않음.
	<b>Opportunity</b> 기회	관광지 개발을 할 경우, 위 프로젝트의 방법론을 사용 → 보다 많은 사람들이 만족함.
	<b>Threat</b> 위협	해외여행의 색다른 문화 체험. 저가 항공.

※ *SWOT 분석*

*여행지에서 사용하는 금액은 경우의 수가 많기 때문에 고려 하지 않았으나,  
실제 여행에서 굉장히 중요한 요소임.  
때문에, 위 코스에서 가격에 대한 부분을 추가하면 좀 더 효과적인 여행 추천 시스템이 됨.*

## 제주도에 간 사람에게 또 제주도를 추천?

### 협업 필터링 도입

개인이 다녀왔던 여행에 대한 만족도를 기반으로 취향에 맞게 다른 여행지를 추천해주는 방법.

단, 협업 필터링 기반의 추천 모델의 문제점은 아무 정보가 없는 사람에 대해서는 추천하기 어려움.

(Cold Start Problem)

→ 군집 기반 추천 모델로 보완 가능

당신에 대해 알려주세요!

성별 ☒ 남성 ☐ 여성

나이대  
20대

동행인  
4~10명

☐ 가족과 ☒ 친구와 ☒ 연인과

기억에 남는 여행 +

여행 1 ▼ 여행 2 ▼

제주도 여행은 어땠나요?

음식	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
교통	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
경치	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
가격	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
시설	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
체험	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

전체적으로

■■■■■■■■■■ 6

나에게 맞는 여행 찾기 ▶



## 출처

### 사용 데이터

국민 여행 실태조사 – 관광지식정보시스템 ([www.tour.go.kr](http://www.tour.go.kr))

주요 관광 지점 입장객 통계 – 관광지식정보시스템 ([www.tour.go.kr](http://www.tour.go.kr))

전국 문화 축제 표준 데이터 – 공공데이터포털 ([www.data.go.kr](http://www.data.go.kr))

소상공인 상권정보 상가업소 데이터 – 공공데이터포털 ([www.data.go.kr](http://www.data.go.kr))

### 국내 여행 선호도 설문조사 출처

[http://news.tongplus.com/site/data/html\\_dir/2017/08/24/2017082401150.html](http://news.tongplus.com/site/data/html_dir/2017/08/24/2017082401150.html)

### 스스로를 위한 소비 설문조사 출처

<http://www.consumerwide.com/news/articleView.html?idxno=15897>

### VR 사진 출처

<https://m.blog.naver.com/PostView.nhn?blogId=ulsan-namgu&logNo=221173508342&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2F>

### 메타 세콰이어 길 사진 출처

[https://www.popco.net/zboard/view.php?id=photo\\_gallery&no=33016](https://www.popco.net/zboard/view.php?id=photo_gallery&no=33016)

### 동영상 출처

<https://www.youtube.com/watch?v=VJZzrgR23Zc>