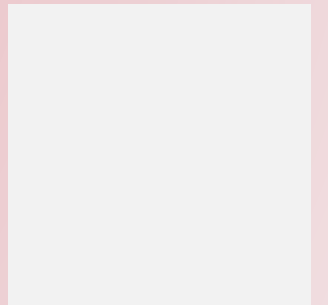


# 클러스터 모델링 분석 기반 AI 전주 여행 테마 추천 서비스



# CONTENTS

01



분석 목적

분석 배경 및 필요성

분석 주제 선정

데이터 소개

02



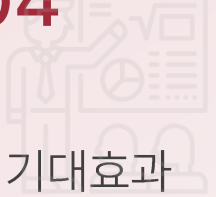
데이터 전처리

03



분석 결과

04



결론 및 기대효과

## \* 전주 한옥마을 관광 키워드 검색(1)

전북일보

지난해 전주한옥마을 관광객 1129만 명, 역대 최다

전북일보

전주한옥마을 1500만 관광시대 '청신호'

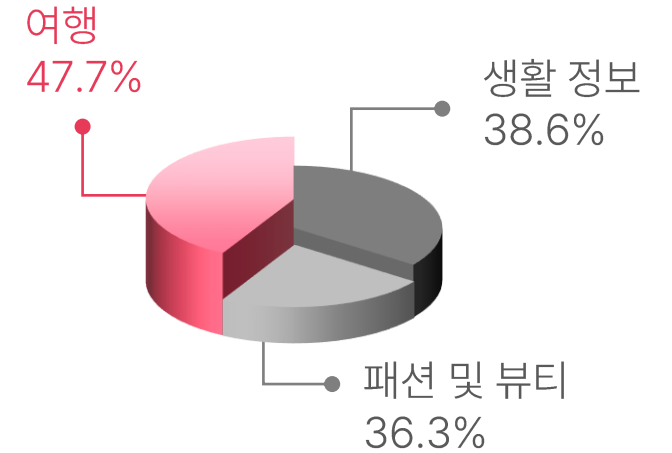
전라일보

[전북]전주 한옥마을, 볼거리 없는 당일치기 '노잼도시'

## \* 네이버 블로그 '전주' 키워드 데이터 크롤링



## \* 2020년 전주시 추천 서비스 필요 분야(2)



이용자 성향에 따른 여행 테마 추천 서비스 → 전주시 관광 활성화(문화 관광 발전에 기여)

전주의 다양한 여행 테마 분석 및 제공 → 전주시 지역관광발전의 도모 이룸

## 이용자의 취향에 맞춘 장소를 추천하는 ' AI 모델 기반 코스 추천 서비스 '

01 . 여행 테마 추천 서비스 수요에 따른 모델 개발

02. 잠재적 니즈 분석에 따른 관광객 유치 활성화

03. 독창적인 역사문화도시, 지속가능한 관광도시 발전 도모

데이터수집



환경분석

설문조사

데이터분석



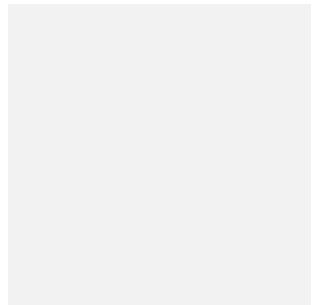
응답자 클러스터링

그룹 선호도 기반 코스 선정

최종산출물



개인 특성 맞춤 코스 추천



데이터수집

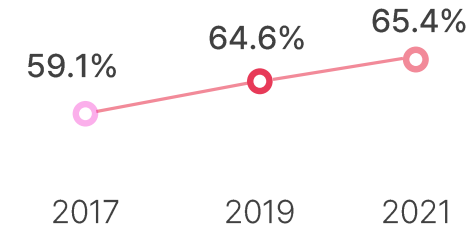
환경분석

## 전통문화에 대한 관심도



\* 전주시 사회조사 전통문화에 대한 관심, 통계청(3)

## 전통문화에 대한 무관심



Q. 전주시 전통문화 관광지를 알고 있나요?

A. 아무 장소도 모른다.  
한옥마을만 알고 있다.

→ 관심 증대를 위한  
전통문화를 포함하는 장소 추가

## 데이터수집

## 설문조사

## &lt;일반 설문조사 항목&gt;

1. 귀하의 나이는 어떻게 되십니까? \*

- ☐ 10대
- ☐ 20대
- ☐ 30대
- ☐ 40대
- ☐ 50대
- ☐ 60대 이상

2. 귀하의 성별은 어떻게 되십니까? \*

- ☐ 남
- ☐ 여

⋮

나이

성별

평균 연간 여행 횟수

과거 전주시 방문 이력과 목적

여행정보를 얻는 매체

전주시의 한옥마을과 같은 관광 명소나 축제 인지

같이 여행하는 일행과의 관계

전주 여행 시 교통수단

## &lt;카노 설문조사 항목&gt;

9. 전주시에 어린이를 위한 프로그램이 \*

|                       | 마음에 든다                | 당연하다                  | 아무 느낌 없다              | 할 수 없다                | 마음에 안든다               |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 마련되어 있다면<br>어떠시겠습니까?  | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 마련되지 않았다면<br>어떠시겠습니까? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

⋮

키즈 프로그램

실버 프로그램

가족 프로그램

맛집

커플 프로그램

낮 액티비티

저녁 문화 활성화

주차공간 충분 여부

대중교통

당일 여행

관광지 평균 물가

한국 전통체험 공간 유무

자연을 즐길 수 있는 프로그램

한옥마을 투어

한옥 숙소

비싸지만 좋은 숙소

## 데이터분석

## 응답자 클러스터링

## \* 클러스터링 구현 코드

```
sns.scatterplot(x=scaled_df[0], y=scaled_df[1], hue = scaled_df['cluster'], legend="full")
plt.title('Kmeans {} clusters'.format(len(scaled_df.cluster.unique())))
plt.show()
```

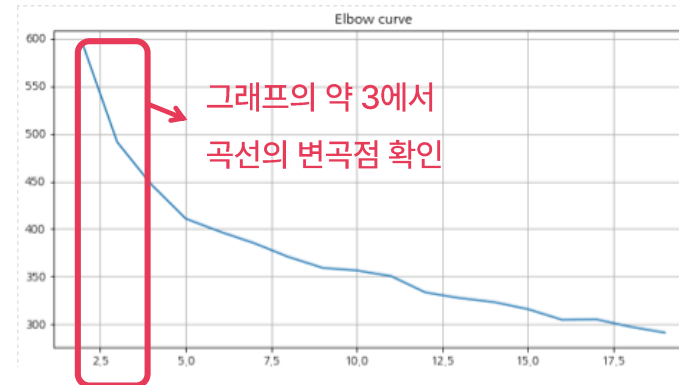
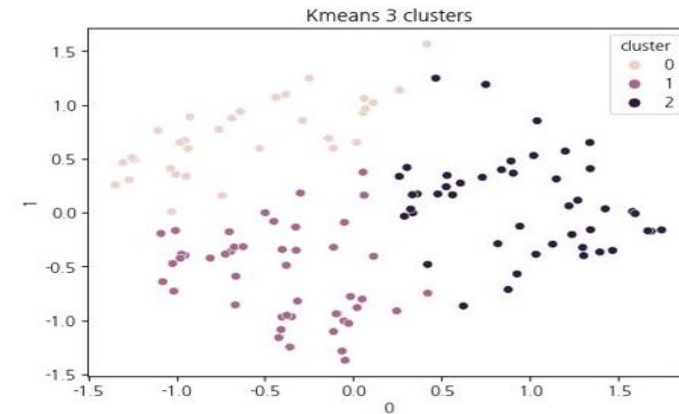
```
print('Silhouette Coefficient: {:.4f}'.format(metrics.silhouette_score(df2.iloc[:, :-1], df2['cluster'])))
print('Davies Bouldin Index: {:.4f}'.format(metrics.davies_bouldin_score(df2.iloc[:, :-1], df2['cluster'])))
```

```
distortions = []
for k in range(2, 20):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(df2)
    distortions.append(kmeans.inertia_)

fig = plt.figure(figsize=(10, 5))
plt.plot(range(2, 20), distortions)
plt.grid(True)
plt.title('Elbow curve')
plt.show()
```

```
kmeans = KMeans(n_clusters=3, random_state=7)
clusters = kmeans.fit(df3)
df2['cluster'] = clusters.labels_
```

## \* 클러스터링 시각화



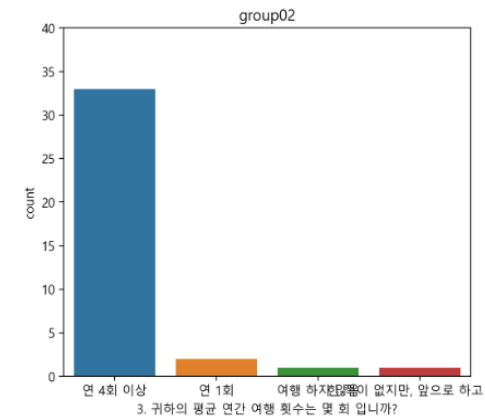
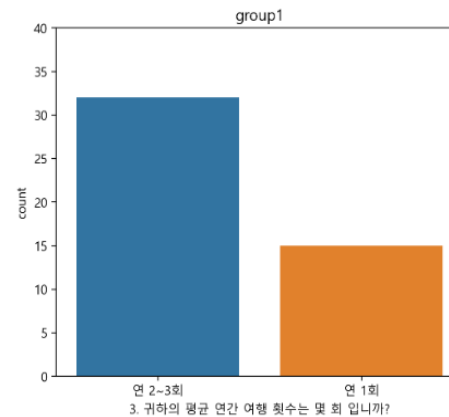
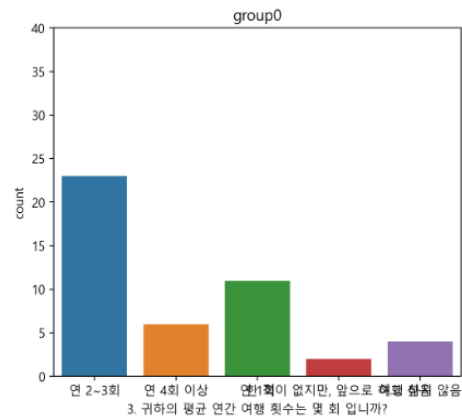
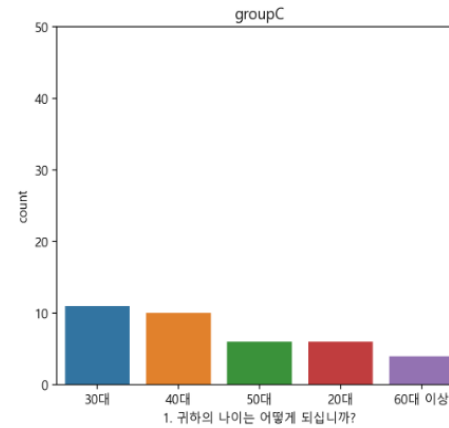
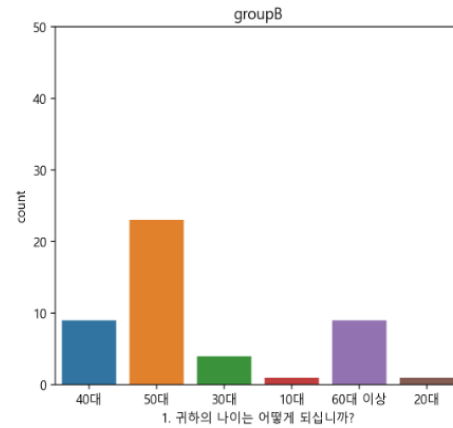
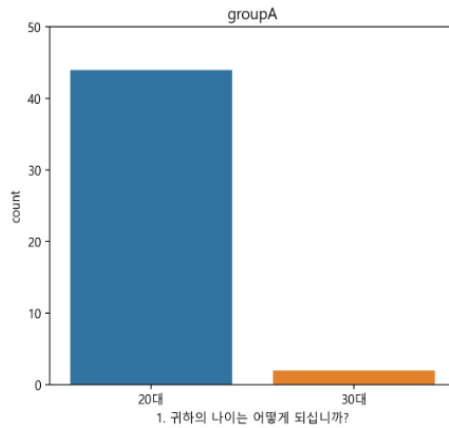


## 데이터분석

그룹 선호도 기반 코스 선정

설문 1 ~ 8번 문항별로 잘 나누어짐을 알 수 있음

이 특성을 기반으로, 서비스 이용자는 자신에게 맞는 코스를 추천 받음

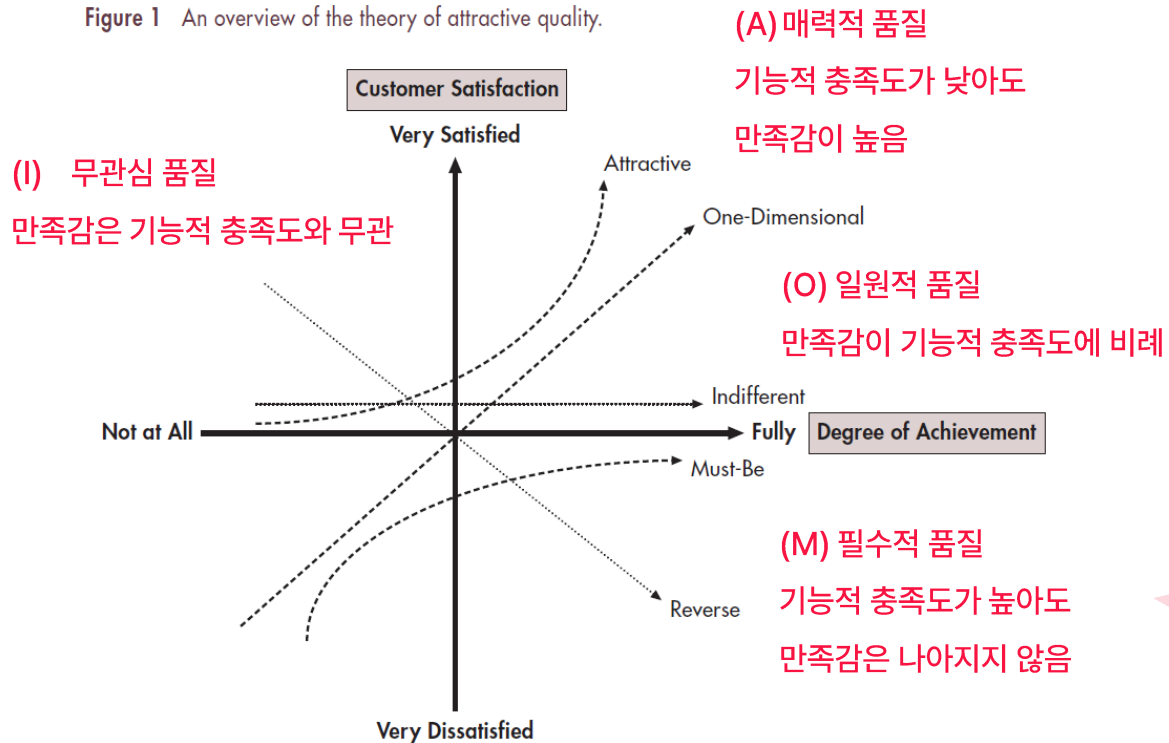


카노 설문 : 긍정 문항과 부정 문항이 쌍으로 제시, 응답자는 문항에 각각 대답함

- 설문 결과에 따라 문항별 품질 속성 정의

### <카노 설문지 예시>

Figure 1 An overview of the theory of attractive quality.



| Kano question   | Answer   |
|---|--|
| Functional form of the question<br>(e.g., if the car has air bags, how do you feel?)              | <input type="checkbox"/> I like it that way<br><input type="checkbox"/> It must be that way<br><input type="checkbox"/> I am neutral<br><input type="checkbox"/> I can live with it that way<br><input type="checkbox"/> I dislike it that way |
| Dysfunctional form of the question<br>(e.g., if the car does not have air bags, how do you feel?) | <input type="checkbox"/> I like it that way<br><input type="checkbox"/> It must be that way<br><input type="checkbox"/> I am neutral<br><input type="checkbox"/> I can live with it that way<br><input type="checkbox"/> I dislike it that way |

Q. Xu, R. J. Jiao, X. Yang and M. Helander, "An Analytical Model for Customer Need Analysis," Design Studies, Vol. 30, pp. 87-110, 2009 [4]

기능적 충족도가 높을 경우  
고객의 만족이 많이 상승하는  
(A) 매력적 품질을 중심으로 코스 추천

R. Florez-Lopez and J. M. Ramon-Jeronimo, "Managing Logistics Customer Service under Uncertainty: An Integrative Fuzzy Kano Framework," Information Sciences, Vol. 202, pp. 41-57, 2012 [5]

### ① 응답결과를 카노 테이블에 기록한다

| Quality attribute →<br>↓ |              | Dysfunctional |            |            |              |            |
|--------------------------|--------------|---------------|------------|------------|--------------|------------|
|                          |              | 1. like       | 2. must-be | 3. neutral | 4. live with | 5. dislike |
| Functional               | 1. like      | Q             | A          | A          | A            | O          |
|                          | 2. must-be   | R             | I          | I          | I            | M          |
|                          | 3. neutral   | R             | I          | I          | I            | M          |
|                          | 4. live with | R             | I          | I          | I            | M          |
|                          | 5. dislike   | R             | R          | R          | R            | Q          |

A = attractive  
 O = one-dimensional  
 M = must-be  
 I = indifferent  
 R = reverse  
 Q = questionable

M. Lofgren and L. Witell, "Kano's Theory of Attractive Quality and Packaging," Quality Management Journal, Vol. 12, pp. 7-20, 2005 [6]

### ② 모든 결과를 합산하여 A,O,M,I Score를 구한다

Ordinary Kano Matrix(OKM)

|   |   |   |    |    |
|---|---|---|----|----|
| 0 | 3 | 8 | 14 | 11 |
| 0 | 0 | 0 | 1  | 6  |
| 1 | 0 | 2 | 0  | 2  |
| 0 | 0 | 0 | 1  | 0  |
| 0 | 0 | 0 | 0  | 1  |



$(A, O, M, I) = (25, 11, 8, 4)$

### ③ 설문 응답자의 모호함을 조정하는 GMF를 적용한 Fuzzy Kano Matrix

|   |   |   |   |   |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 |
|   |   |   |   |   |
|   |   |   |   |   |

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 1     | 2     | 3     | 4     | 5     |
| 0.798 | 0.108 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.108 | 0.798 | 0.108 |
|       |       |       |       |       |
|       |       |       |       |       |

OKM



FKM

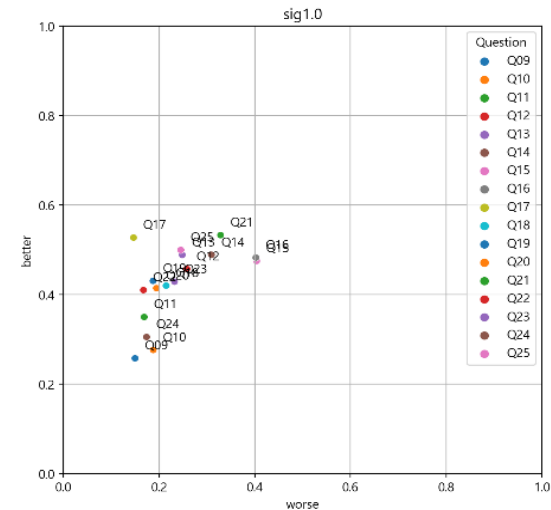
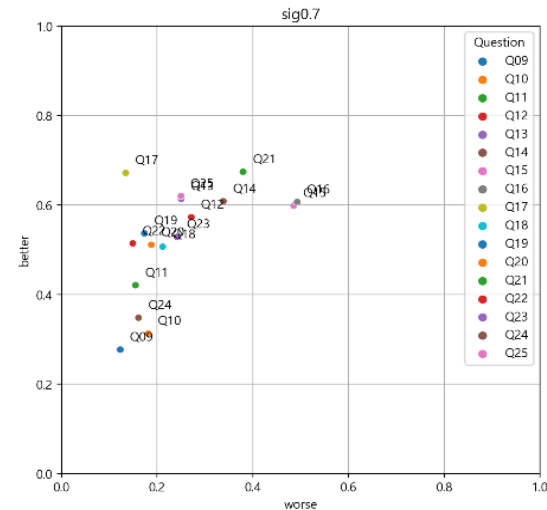
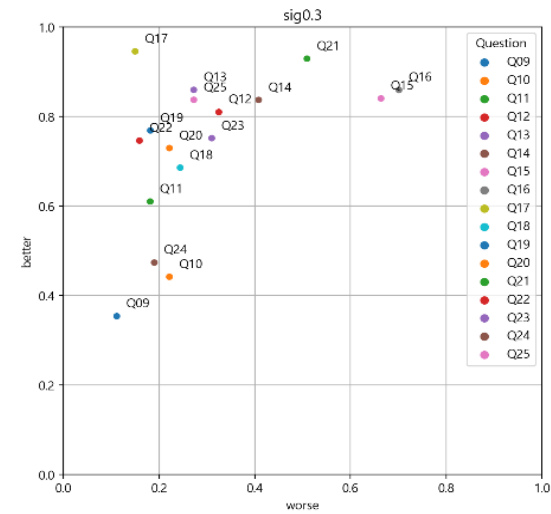
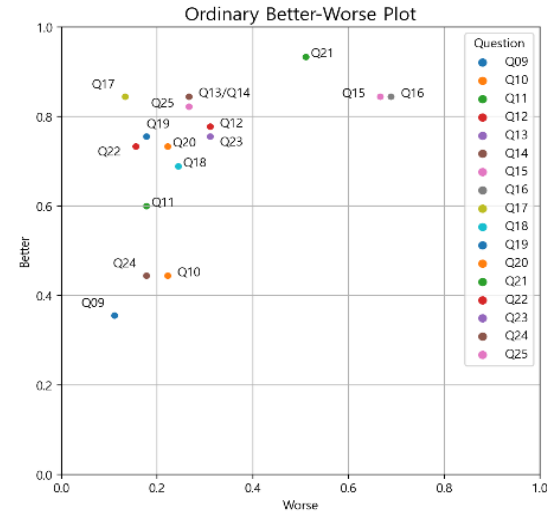
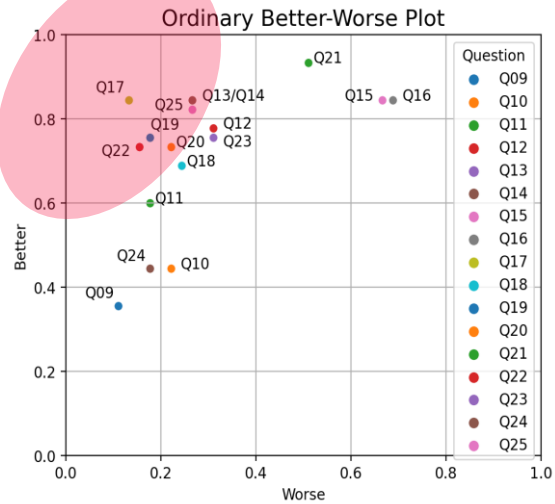
Gaussian  
Membership  
Function(GMF)

### ④ A, O, M, I 스코어를 이용해 Better Score와 Worse Score를 구한다

$$\text{Better} = (A+O)/(A+O+M+I) \quad (1)$$

$$\text{Worse} = (O+M)/(A+O+M+I) \quad (2)$$

매력적 품질



&lt;그룹 A의 선호도&gt;

1순위 Q17 평균물가

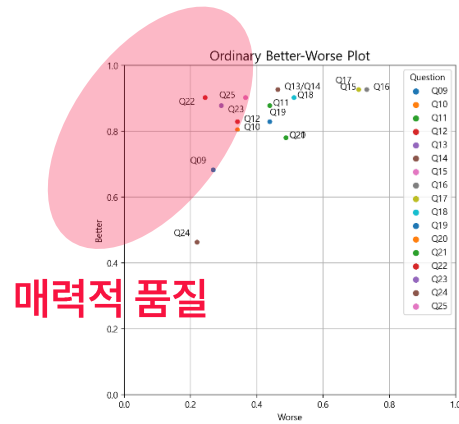
2순위 Q13 액티비티

Q19 자연

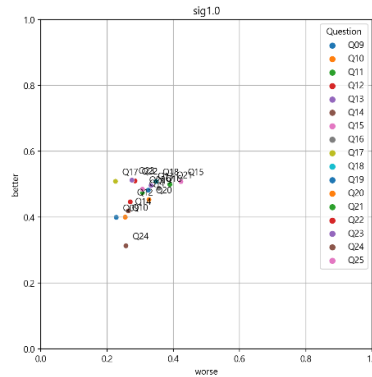
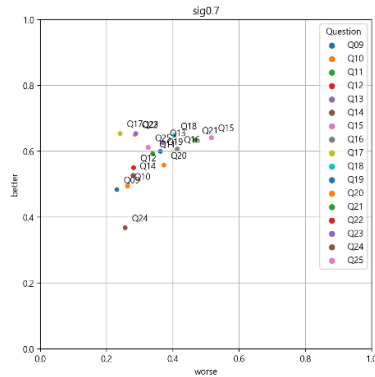
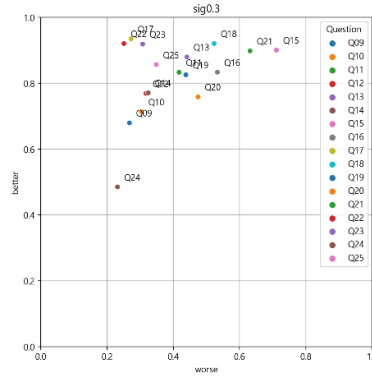
Q22 당일여행

Q25 한옥

# 카노 데이터 기반 그룹별 선호도 분석



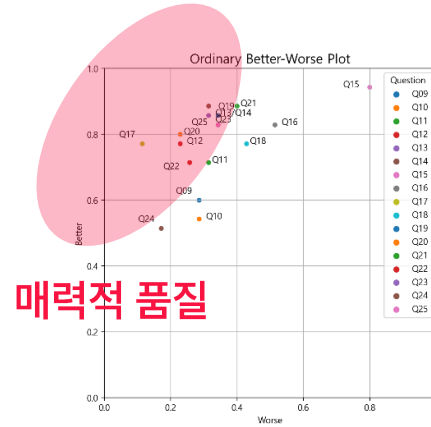
매력적 품질



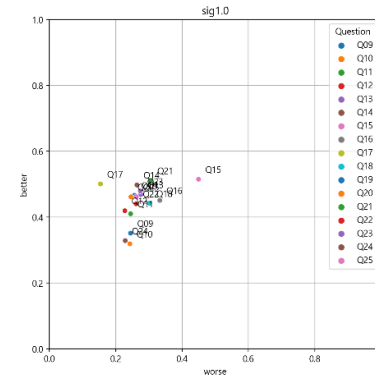
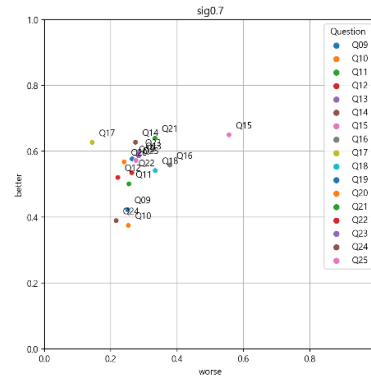
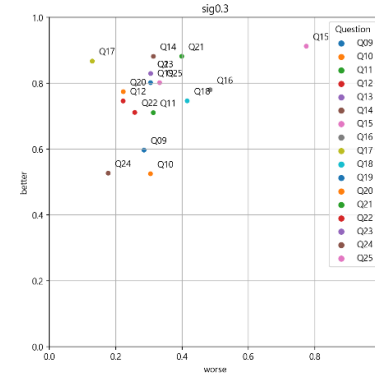
<그룹 B의 선호도>

1순위 Q22 당일여행 · Q23 1박 이상 여행

2순위 Q25 한옥 · Q12 연인 · Q10 노인 · Q17 평균물  
가



매력적 품질



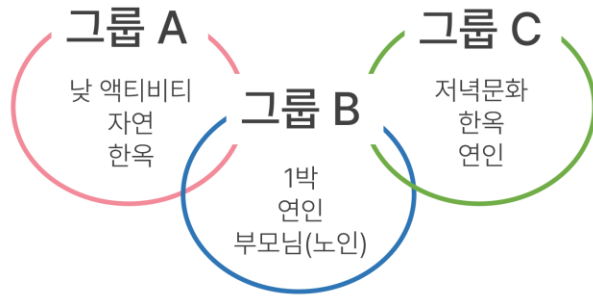
<그룹 C의 선호도>

1순위 Q17 평균물가

2순위 Q14 저녁문화 · Q20 한옥마을 · Q12연인

\* Better-Worse 그룹 선호도 통계

## 군집 클러스터링



## SNS 및 콘텐츠 키워드 기반



키워드 분석을 참고,  
전주시 장소 별 속성 값을  
그룹 간 특징으로 분류

## 키워드 분석을 통한 속성값

| 장소명          | 낮,저녁,숙,속성 |      |
|--------------|-----------|------|
| 전주한옥레일바이크    | 낮         | 액티비티 |
| 꽃실이대여사전거     | 낮         | 액티비티 |
| 전주월드컵경기장     | 낮         | 액티비티 |
| 전주남부시장야시장    | 저녁        | 저녁문화 |
| 하늘정원달빛버스킹    | 저녁        | 1박   |
| 전주 문화재 야행    | 저녁        | 저녁문화 |
| 완산골 게스트하우스   | 숙소        | 숙소   |
| 24 게스트하우스 전주 | 숙소        | 숙소   |

## &lt; 추천 알고리즘 결과 도출 값 &gt;

아침 활동: ['전주레이싱', '연인']  
 저녁 활동: ['전주신시가자', '저녁문화']  
 숙소: ['도원 게스트하우스', '숙소']

## 군집별키워드 기반 추천

```

if k == 0 :
    df1 = df.loc[(df['속성']=='액티비티')|(df['속성']=='자연')|(df['속성']=='한옥')|(df['속성']=='숙소'),['장소명','낮,저녁,숙소','속성']].copy()
    for i in range(3):
        if i == 0:
            df2 = df1.loc[df1['낮,저녁,숙소']=='낮',['장소명','속성']].copy()
            print("아침 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[0]-1),:].values)
        elif i == 1:
            df2 = df1.loc[df1['낮,저녁,숙소']=='저녁',['장소명','속성']].copy()
            print("저녁 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[2]-1),:].values)
        else:
            df2 = df1.loc[df1['낮,저녁,숙소']=='숙소',['장소명','속성']].copy()
            print("숙소:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[1]-1),:].values)
    elif k == 1 :
        df1 = df.loc[(df['속성']=='1박')|(df['속성']=='연인')|(df['속성']=='노인')|(df['속성']=='숙소'),['장소명','낮,저녁,숙소','속성']].copy()
        for i in range(3):
            if i == 0:
                df2 = df1.loc[df1['낮,저녁,숙소']=='낮',['장소명','속성']].copy()
                print("아침 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[0]-1),:].values)
            elif i == 1:
                df2 = df1.loc[df1['낮,저녁,숙소']=='저녁',['장소명','속성']].copy()
                print("저녁 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[2]-1),:].values)
            else:
                df2 = df1.loc[df1['낮,저녁,숙소']=='숙소',['장소명','속성']].copy()
                print("숙소:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[1]-1),:].values)
    else:
        df1 = df.loc[(df['속성']=='저녁문화')|(df['속성']=='한옥')|(df['속성']=='연인')|(df['속성']=='숙소'),['장소명','낮,저녁,숙소','속성']].copy()
        for i in range(3):
            if i == 0:
                df2 = df1.loc[df1['낮,저녁,숙소']=='낮',['장소명','속성']].copy()
                print("아침 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[0]-1),:].values)
            elif i == 1:
                df2 = df1.loc[df1['낮,저녁,숙소']=='저녁',['장소명','속성']].copy()
                print("저녁 활동:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[2]-1),:].values)
            else:
                df2 = df1.loc[df1['낮,저녁,숙소']=='숙소',['장소명','속성']].copy()
                print("숙소:",df2.iloc[random.randint(0,df1['낮,저녁,숙소'].value_counts().sort_index()[1]-1),:].values)

```

### 군집 클러스터링

사용자의 그룹별 니즈를 예측하기 위한 머신러닝 및 딥러닝

-> 각 변수는 설문에 사용되었던 기본 질문

### SNS 및 콘텐츠 키워드 기반

RandomForest와 GridSearch를 통한  
XGBoost, LightGBM, 그리고 딥러닝 성능을 비교

그 결과 test 데이터에 대한 정확도가 유사하게 나와,  
모델이 빠르고 가벼운 lightgbm 모델로 사용

### LightGbm 모델링 code

```
from lightgbm import LGBMClassifier
lgb_model = LGBMClassifier(random_state=2023)
params = {'learning_rate': [0.01, 0.05, 0.1, 0.2, 0.3], 'max_depth': [2, 3, 4, 5]}
gs_lgb_model = GridSearchCV(lgb_model, params, cv=3, verbose=2)
gs_lgb_model.fit(x_train, y_train)
```

### LightGbm 성능 결과

| 모델           | 정확도      |
|--------------|----------|
| RandomForest | 0.974359 |
| XGBoost      | 0.974359 |
| LightGBM     | 0.974359 |
| DNN          | 0.717949 |

사용자가 자신의 인적사항 입력

자신이 속한 그룹의  
선호도에 따른 코스 추천

11:05

전주  
테마 여행 코스 추천

1. 귀하의 나이는 어떻게 되십니까?  
10대 ☐ 20대 ☒ 30대 ☐ 40대 ☐ 50대 ☐ 60대 이상 ☐

2. 귀하의 성별은 어떻게 되십니까?  
남성 ☒ 여성 ☐

3. 귀하의 평균 연간 여행 횟수는 몇 회입니까?  
여행 하지 않음 ☐ 한 번 이하지만, 앞으로 하고 싶음 ☐  
연 1회 ☐ 연 2~3회 ☒ 연 4회 이상 ☐

4. 전주시를 방문한 적이 있으십니까?  
있다면, 어떤 목적이었습니까?  
방문한 적이 없음 ☐ 거주 ☐ 관광 ☒ 업무 ☐ 기타:

11:05

전주  
테마 여행 코스 추천

5. 전주시의 한옥마을과 같은 관광 명소나 축제에 대해 알고 계십니까?  
여우 장소도 모름 ☐ 한옥 마을만 알고 있음 ☒  
그 외 관광지나 축제에 대해 알고 있음 ☐ 기타:

6. 여행 정보는 주로 어디서 얻으십니까?(복수)  
인터넷 ☒ TV ☐ 광고 ☐ 여행사 ☐ 관광 안내 서적 ☐ SNS ☒  
과거 방문 경험 ☐ 정보 없이 방문 ☐ 기타:

7. 같이 여행하는 일행과의 관계는 어떻게 되십니까?(복수)  
혼자 ☒ 가족 ☐ 친구 ☐ 연인 ☐ 동료 ☐ 친목 단체 ☐  
학교 단체 ☐

8. 전주 여행 시, 어떤 교통수단을 이용하시겠습니까?  
자차 ☐ 렌트 ☒ 대중교통 ☐ 기타:

테마여행 추천  
알고리즘



11:05

전주  
테마 여행 코스 추천

맞춤 테마 여행을 찾았습니다!  
여행자님에게 추천드리는 여행은 "꽃상아타고 돌레길 탐방" 여행입니다

다른 테마 여행 보기

한옥 마을 < 한옥 체험
 전주 마시당 투어

상세한 여행 무트를 원하시나요?  
AI 전주 여행 플래너에게 물어보기



## >> 관광객의 취향에 맞는 코스 추천 앱

전주시에서 열리는 행사 정보 및 계절별 장소 등  
API 데이터 업데이트를 통한 최신화 유지

## >> 전주시 관광 수요 예측

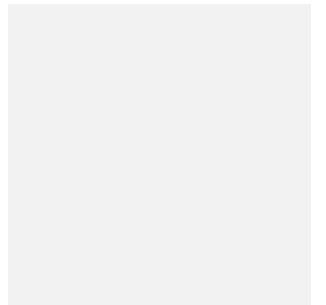
추후 그룹별 방문 횟수 조사를 통한  
재방문 그룹의 예측 모델을 제작하여  
관광 모델의 방향성 제시 가능

## >> 역사 문화도시 이미지 강화

전주시의 새로운 매력 발굴 및  
기존 역사 문화 도시로서의 이미지 강화

## >> 지속가능성

고객의 니즈를 충족하며  
향후 전주시 니즈 충족에도 지속적으로 활용 가능



## 정보 습득



Q. 전주시 전통문화 관광지를 알고 있나요?

A. 아무 장소도 모른다.  
한옥마을만 알고 있다.

전주시에서 제공 하는 데이터를 제공받아  
신뢰 있는 정보 습득 가능

## 높은 여행 만족도



정확도 높은 클러스터링을 통한 관광지  
추천으로 만족도 높은 여행 가능  
→ 관광객들이 손쉽게 취향에 맞는 정보를  
얻어 여행 기대감 상승

## 매출 증대



시설 현대화에 이어 시장홍보로 매출 증대 도움  
다양한 여행지 추천으로 n박 여행객 증가

## 도시 이미지



빅데이터 분석 기반의 새로운 e-서비스의 도입  
→ 차세대 스마트 디지털 도시 이미지 구축

## 신규 관광객



전주시 관광지 노출량 상승으로 인한  
신규 관광 목적인 외지인들의 유입 증대  
→ 관광지 검색 시간 단축 및 접근성 향상

## 재방문



방문하기 편한 도시 서비스 제공으로  
재방문을 증가

# 분석 데이터 및 레퍼런스

[1] 추천서비스유형에 따른 소비자 인식과 구매의도에 미치는 영향. 국내석사학위논문 성균관대학교 일반대학원, 2020. 서울, 피주영

[2] 통계청\_전주시 사회조사: 전통문화에 대한 관심

[3] Q. Xu, R. J. Jiao, X. Yang and M. Helander, "An Analytical Model for Customer Need Analysis," Design Studies, Vol. 30, pp. 87-110, 2009

[4] R. Florez-Lopez and J. M. Ramon-Jeronimo, "Managing Logistics Customer Service under Uncertainty: An Integrative Fuzzy Kano Framework," Information Sciences, Vol. 202, pp. 41-57, 2012

[5] M. Lofgren and L. Witell, "Kano's Theory of Attractive Quality and Packaging," Quality Management Journal, Vol. 12, pp. 7-20, 2005

[6] <https://blackkiwi.net/> (검색일자: 2023.06.16)

[7] 통계청\_전주시 사회조사: 매출 증대에 도움이 되는 시장지원사업

[8] 통계청\_전주시 사회조사: 전통문화에 대한 관심

[9] '2023 전주시 데이터 분석 공모전'을 위한 설문조사(일반 문항과 카노 문항으로 구성)

감사합니다