

DATA 171

지역 특성에 따른 한달살기 프로그램 활성화 전략

관광 비인기 지역, 한달살기로 살아나다

목차



주제 선정

- 프로젝트 배경
- 프로젝트 목적
- 전체 로드맵



데이터 설명

- 활용 데이터
- 데이터 전처리



데이터 분석

- 한달살기 활성화 지역 선정
- 지역 클러스터링
- 클러스터 설명



정책 제안

- 한달살기 지역 추천 서비스
- 지자체별 개선점 제안



기대효과

프로젝트 배경

프로젝트 목적

전체 로드맵



주제 선정

여전한 오버투어리즘 문제

오버투어리즘이란?

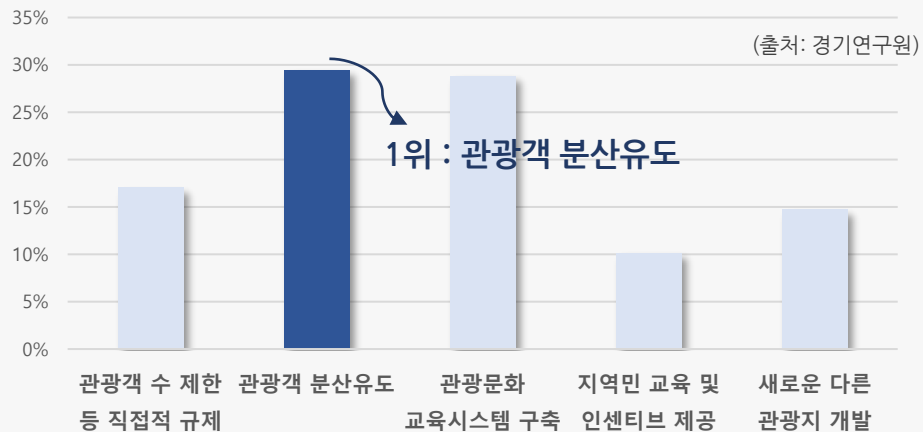
수용 가능한 범위를 넘어서는 관광객이 관광지에 모여들면서
주민들의 삶을 침범하는 현상

[사설] 경주시 오버투어리즘 미리 방책 찾아야

오버투어리즘 사회적 문제...관광객 분산해야

전주 꽃동산 '오버투어리즘'... 꽃이 반갑지 않다

오버투어리즘의 바람직한 정책방향



한달살기 프로그램도 제주도로?

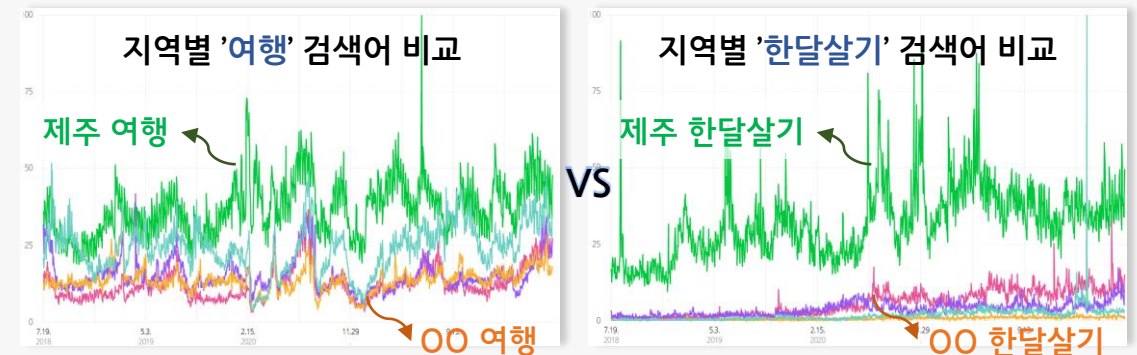
한달살기 프로그램이란?

기존 여행패턴에서 벗어나 한 곳에서 좀 더 여유롭게
생활하는 여행 형태로, 4-5년 전부터 유행하기 시작

☑ '한달살기' 단어 자체가 명사화가 된 것으로,
실제 기간은 일주일, 한달, 3개월.. 등으로 다양함

지자체들 이 때다 싶어 한달살기 프로그램 홍보해보지만..
이 마저도 기존 여행 인기 지역으로 집중

NAVER DataLab. 네이버 데이터랩 분석



☑ 한달살기에 있어서 제주는 다른 지역보다 압도적인 검색량

한달살기 프로그램을 통해 관광 비인기 지역을 활성화시키자!

왜 한달살기 프로그램인가?

관광객 분산유도 측면에서 한달살기 프로그램의 장점

- ① 지역 특성에 따라 타겟을 다르게 삼아 그에 맞는 경쟁력을 갖출 수 있음

관광지 개수가 부족한 지역
휴식/힐링을 타겟으로



기존에는 관광지로서의 경쟁력이
사실상 없던 지역

VS

관광지 개수가 충분한 지역
체험/관광객을 타겟으로



부산, 제주 등의 관광 인기 지역에
인지도가 밀려 빛을 발하지 못한 지역

- ② 관계인구의 증가에 따른 장기적인 효과

지역에 새롭게 이주하는 '정주인구'나
여행이나 관광으로 방문하는 '교류인구'가 아닌,
어느 정도 지속적 관계를 맺고 체류하는 사람들을 일컫는 개념

현재 한달살기 프로그램의 미흡한 점

- ① 정책 목적과는 맞지 않는 지원책

귀농 타겟팅

짧은 기간, 박한 혜택
무료 농촌 봉사가 되어버림

여행 타겟팅

이용자에 대한 고려 ↓
해당 지역 여행 정보 부족

- ② 프로그램 홍보 부족과 정보 제공 미흡

- ☑ 지자체별 홈페이지에서만 확인 가능한 프로그램 정보들
- ☑ 한달살기 프로그램이 진행되고 있는 지역, 기간, 지원요건 등을 한눈에 확인하기 힘들어 블로그에서 이웃끼리 정보를 공유하는 상황

- ③ 숙박시설의 미흡

- ☑ 숙박시설은 한달살기 만족도에 큰 영향을 미침

프로젝트 배경

여전한 오버투어리즘 문제 BUT 한달살기 프로그램도 제주도로?

프로젝트 목적

관광객 분산유도 측면 한달살기의 장점

한달살기 프로그램의 미흡한 점

한달살기 프로그램으로 관광 비인기 지역을 활성화시키자!



신한카드 신한카드 내국인 데이터 전처리

데이터 전처리



문화체육관광부 국민여행조사 데이터 전처리

한달살기 프로그램 활성화 지역 선정



신한카드 내국인 지역별 여행건수 데이터 이용

관광 비인기 지역 선정

1. 수도권, 광역시, 제주 제외
2. 그 외 지역 중 여행 결제건수 TOP 5 제외

지역 클러스터링



한국관광 데이터랩 네비게이션 유형별 검색 데이터 이용

검색 유형들을 K-Means 알고리즘을 통해 군집화

분류 모델을 이용해 군집화 중요 변수 시각화

클러스터 설명



휴식형



맛집형



자연관광형



레저스포츠형



역사관광형

한달살기 지역 추천 서비스



네이버 여행정보 크롤링 데이터 이용

- (1) 클러스터와 (2) 키워드에 따라 한달살기 지역을 추천

Persona 1

Persona 2

지자체별 개선점 제안

- 군집 특성에 따라 -

AS-IS 분석과 TO-BE 제안

휴식형 클러스터 사례
강원도 평창군

레저스포츠형 클러스터 사례
충청북도 충주시

지자체별 개선점 제안

- 지역별 여행 만족도 회귀분석 결과에 따라 -



국민여행조사 데이터 이용

AS-IS 분석과 TO-BE 제안

사례
강원도 고성군

사례
전라북도 고창군

기대 효과

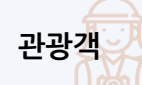


지자체

기존 관광 비인기 지역의 경쟁력 강화
지역별 특성을 반영한 차별화
...등등

대한민국의 균형적인
관광 발전에 이바지

개인별 선호 유형에 맞는 지역에서의 프로그램
양질의 숙소 이용으로 치안과 위생 걱정 최소화
...등등



관광객

→ 활용 데이터

→ 데이터 전처리 →



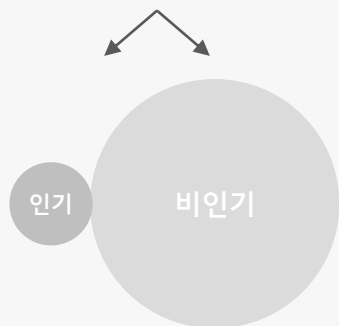
데이터 설명



신한카드 내국인 데이터

18년 1월~22년 4월 내국인
개인별 카드 소비 데이터

→ 각 지역에 대해
여행 목적 방문자 수 집계
인기 / 비인기 지역으로 분류



➡ 비인기 지역의
한달살기 여행 활성화 방안은 ?



국민여행조사 데이터

2019년 한국문화관광연구원
국민여행조사 데이터

→ 지역별 회귀분석
지역별로 만족도를
높이는/낮추는 요인 파악

변수명	coef	std err
주요이용교통수단9	1.4922	0.046
여행정보획득경로6	1.4552	0.288
여행지선택이유9	-0.4254	0.103
여행활동12	-1.1074	0.276



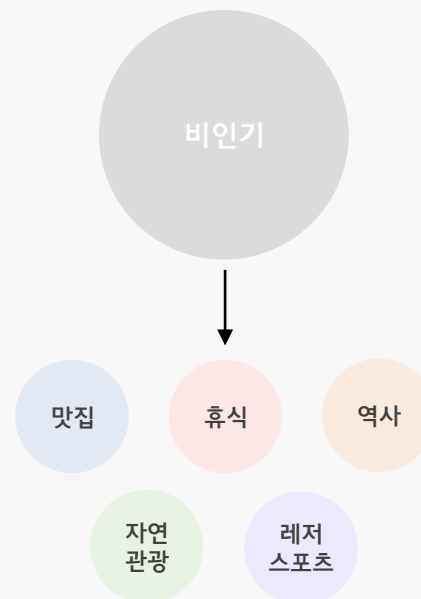
변수명	coef	std err
주요이용교통수단9	1.2117	0.02
여행지선택이유12	-0.5682	0.136
비용	1.14E-06	3.51E-07
주요이용교통수단5	0.9174	0.359



한국관광 데이터랩 데이터

최근 1년간 전국 내비게이션
검색 유형별 비율 데이터

→ 클러스터링
5개 유형으로 군집화



네이버 크롤링 데이터

네이버 여행정보
(travel.naver.com) 데이터

→ 지역별 데이터 크롤링
지역 추천 서비스에 이용

전라남도 함평군

자연과 조화를 이룬 생태관광의 메카
함평군. 유채꽃 물결 사이로 날갯짓하
는 나비들의 모습을 볼 수 있는 '함평
나비 대축제'가 ... 그의 생가 옆에 용천
사 꽃무릇공원은 ...

#놀이동산 #생태공원 #유희비빔밥 #해변

충청북도 옥천군

시인 정지용의 고장 충북 옥천. 그의 생
가 옆에 마련된 '정지용 문학관'에서는
현대시의 흐름을 엿볼 수 있고, ... 장령
산 자연휴양림은 물놀이와 취사도 가능
해 여름 캠핑 장소로 제격이며, 일출과

#이당국수 #유원지 #벽화마을 #휴양림



① 신한카드 내국인 데이터

사용할 변수 선정

V1	이용회원 거주지
✓ V2, V3	카드이용 가맹점 주소
✓ gb3	가맹점 업종 대분류
gb2	가맹점 업종 소분류
SEX_CCD	성별
CLN_AGE_R	연령대
...	
TA_YM	이용년월
✓ VLM	취급액 (단위 : 원)

	v1	v2	v3	gb3	gb2	sex_ccd	cln_age_r	ta_ym	daw_ccd_r	apv_ts_dl_tm_r	vlm	usec
0	세종	세종	.	취미오락	종합쇼핑	F	30	202108	WHITE	휴식	325008700	12639
1	세종	세종	.	취미오락	종합쇼핑	M	30	202112	WHITE	휴식	599872600	20805
2	서울	세종	.	취미오락	종합쇼핑	F	50	202106	WHITE	활동	30363400	1075

V2, V3 : 카드이용 가맹점 주소 (V2: 시도 / V3: 시군구)

☑ 시도는 다르지만 시군구 이름은 동일한 변수가 있음 (Ex.강원 고성군, 경남 고성군)

→ V2와 V3를 합친 'travel_address' 변수 생성

df_travel['travel_address']=df_travel['v2']+' '+df_travel['v3']

travel_address

부산 남구

대전 동구

대전 동구

gb3 : 가맹점 업종 대분류 (문화예술, 스포츠, 여행, 취미·오락)

☑ 스포츠, 문화예술, 취미오락의 경우, 여행 중 사용인지의 여부를 구별할 수 없음

→ 여행 목적의 데이터가 필요하므로 gb3가 '여행' 인 데이터만 추출

df_travel=data.loc[data['gb3']=='여행']

gb3 gb2

여행 숙박

여행 교통

여행 교통



② 국민여행조사 데이터

	ID	MON_EXP_1	MON_EXP_2	MON_EXP_3	MON_EXP_4	MON_EXP_5	MON_EXP_6	D_TRA1_CHECK	D_TRA1_CASE	
0	11010550341_1000000	1	2	2	2	2	2	Y	1.0	...
1	11010550341_1000005	1	2	2	2	2	2	Y	1.0	
2	11010550341_1000007	1	2	2	2	2	1	Y	1.0	

사용할 데이터 선정

코로나 팬데믹 시기인 2020년, 2021년 데이터는 특이한 여행 패턴으로 **정확한 인사이트**를 도출하기 어려울 거라는 판단



2019년 데이터 이용

☑ 한달살기 붐이 시작된 건 2017-2018년 무렵이므로 문제 없음

☑ 그 이전 데이터는 설문조사 형식이 다르므로 이용 불가

사용할 변수 선정

ID	아이디
...	...
✓ D_TRA1_1_SPOT	(1차 여행지 1차 방문지의) 방문지역
✓ D_TRA1_CASE	여행유형
✓ D_TRA1_S_Day	여행기간
✓ D_TRA1_ONE_COST	1인 지출비용
...	...
✓ A1, A2_1~A2_3 등	여행지상품구매여부, 여행지 선택이유, 여행활동 등
✓ A12	전반적 만족도
...	...
✓ BSEX, BAGE 등	성별, 연령 등 개인정보
...	...

D_TRA1_1_SPOT : (1차 여행지 1차 방문지의) 방문지역

- ☑ 데이터에는 1~8차 여행지, 각 여행지마다 1~10차 방문지 존재
 - ☑ 전체 데이터 중 1차 여행만 기입한 데이터: 17860개 중 17080개
1차 여행 중 1차 방문지만 기입한 데이터: 17080개 중 16769개
 - ☑ 따라서 1차 여행지 1차 방문지의 데이터로만 분석해도 대표성을 가질 수 있다고 판단
- 지역코드로 되어있기 때문에 지역명으로 변환한 '여행지역' 변수 생성

```
for i in df.index :
    for j in region.index :
        if df.loc[i, 'D_TRA1_1_SPOT'] == region.loc[j, 'code'] :
            df.loc[i, '여행지역'] = region.loc[j, 'region']
```

D_TRA1_S_Day : 여행기간

- ☑ 1이면 1박2일을 뜻함
- 1을 더해준 '여행기간' 변수 생성

```
df = df.rename(columns={'D_TRA1_S_Day': '여행기간'})
df['여행기간'] = df['여행기간'] + 1
```

D_TRA1_ONE_COST: 1인지출비용

- ☑ 여행일수에 따라 달라짐
- 여행일수로 나뉘준 '비용' 변수 생성

```
df['비용'] = df['D_TRA1_ONE_COST'] / df['여행기간']
```



② 국민여행조사 데이터

	ID	MON_EXP_1	MON_EXP_2	MON_EXP_3	MON_EXP_4	MON_EXP_5	MON_EXP_6	D_TRA1_CHECK	D_TRA1_CASE
0	11010550341_1000000	1	2	2	2	2	2	Y	1.0
1	11010550341_1000005	1	2	2	2	2	2	Y	1.0
2	11010550341_1000007	1	2	2	2	2	1	Y	1.0

...

D_TRA1_CASE: 여행유형

- ☑ 1: 국내 관광,휴양 여행
- 2: 국내 가족,친지,친구 방문 여행 (관광,휴양 O)
- 3: 국내 가족,친지,친구 방문 여행 (관광,휴양 X)
- 4: 국내 출장,업무 여행 (관광,휴양 O)
- 5: 국내 출장,업무 여행 (관광,휴양 X)
- 6: 해외 여행

- ☑ 관광,휴양이 포함되지 않거나 출장,업무 목적의 여행은 분석 목적에 맞지 않다고 판단

→ 여행유형이 1,2인 데이터만 이용

```
df=df.loc[(df['MON_EXP_1'] == 1)|(df['MON_EXP_2']==1)]
```

A1: 여행사 상품 구매여부

- ☑ 구매 O: 1 / 구매 X: 2

→ 2를 No(0)으로 변경 & 카테고리형으로 변경

```
df.loc[df['여행사상품구매여부'] == 2, '여행사상품구매여부'] = 0
```

A3_1~A3_21: 여행활동
A6_1~A6_13: 숙박시설
A8_1~A8_7: 사전예약 서비스 종류

- ☑ 복수선택 변수들
(Ex. A6_1: 호텔 / A6_2: 여관 ... 中 해당되는 숙박 시설 모두 선택)

- ☑ 해당 항목을 이용했을 경우 해당 숫자
(Ex. 호텔:1/여관:2)로 표시되어 있음

→ 숫자가 있으면 Yes(1) / Null이면 No(0)으로 변경

```
for j in range(0,21):
    col_name="A3_{}".format(j+1)
    for i in df.index:
        if pd.isnull(df.loc[i,col_name]):
            df.loc[i,col_name]=0
        else:
            df.loc[i,col_name]=1
```

A2_1~A2_3: 여행지 선택이유
A4_1~A4_3: 여행정보 획득경로
A5_1~A5_3: 주요 이용 교통수단

- ☑ 1~3순위까지 선택하는 변수들
(Ex. A2_1: 여행지 선택이유 1순위)

- ☑ 1,2,3순위에 해당되는 항목이 숫자로 표시되어 있음 (Ex. A2_1: 4 -> 4(이동거리)가 1순위)

→ 항목별 칼럼 만들고,
(Ex.여행지 선택이유_1 ~ 여행지 선택이유_14)

1~3순위 중 해당 항목이 있으면 Yes(1) / 없으면 No(0)로 지정

```
for i in range(0,14):
    df['여행지선택이유{}'.format(i+1)]=0
    for j in df.index:
        if (df.loc[j,'A2_1']==i+1) or (df.loc[j,'A2_2']==i+1) or (df.loc[j,'A2_3']==i+1):
            df.loc[j,'여행지선택이유{}'.format(i+1)]=1
```

————→ 한달살기 활성화 지역 선정

————→ 지역 클러스터링

————→ 클러스터 설명



데이터 분석

한달살기 프로그램 활성화가 필요한 관광 비인기 지역을 선정

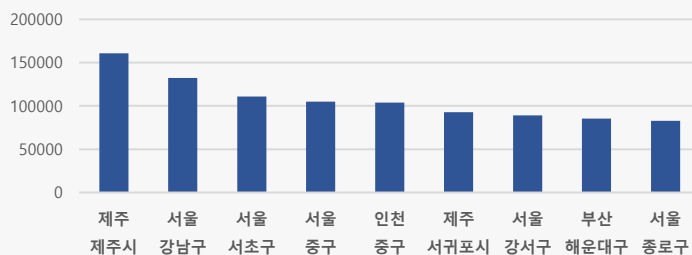


신한카드 데이터 이용

STEP 1. 수도권, 광역시, 제주도 제외

지역별 여행 데이터 건수 비교

상위권에는 모두 수도권, 광역시, 제주도가 포진



→ 수도권, 광역시, 제주도 제외

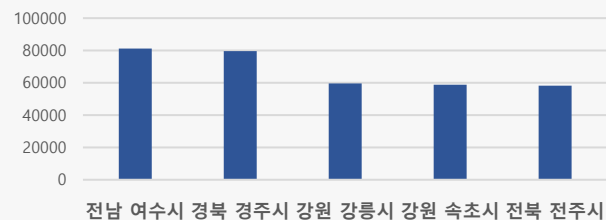


```
df_travel = df_travel[df_travel['v2'].
isin(['경북', '강원', '전남', '충남',
      '경남', '전북', '충북'])]
```

STEP 2. 그 외 지역 중 TOP 5 제외

그 외 지역 여행 데이터 건수 비교

관광지로 유명한 여수, 경주, 강릉, 속초, 전주가 TOP 5에 위치



→ 여수, 경주, 강릉, 속초, 전주 제외



```
df_travel = df_travel[df_travel['v3'].
isin(['여수시', '경주시', '강릉시',
      '속초시', '전주시']) == False]
```

대상 지역 시각화



지역 특성에 따라 지역 클러스터링



한국관광 데이터랩

Data Lab

데이터 이용

한국관광 데이터랩 네비게이션
유형별 검색 비율 데이터

- ☑ 지역별 검색건수의 평균이 약 1,000,000개로
데이터가 충분히 많음

유형별 검색 비율이
유의미한 데이터라고 판단

Ex.) 강원도 철원군



자연관광 16.2%
문화관광 15.2%
음식 24.4%
역사관광 9.9%
레저스포츠 6.7%
숙박 9.2%
체험관광 0.1%
쇼핑 2.3%
기타관광 16.1%



클러스터링 과정

네비게이션 유형별 검색 비율 데이터로
K-Means 알고리즘을 이용해 클러스터링지역 특성에 따른
클러스터 생성

군집화(클러스터링)란?

주어진 데이터 집합을 유사한 데이터들의
그룹, 즉 군집으로 나누는 것

클러스터링의 종류 :



K-means



GMM



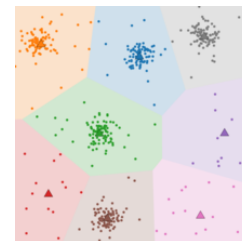
DBSCAN



Hierarchical

유형별 검색 비율을 바탕으로 군집화
→ 지역의 관광 특성 파악

K-means 알고리즘이란?

군집 중심점(centroid)을 선택하여
해당 중심에 가장 가까운 점들을
선택하는 군집화 기법

빠르고 효율적

데이터에 대한
사전정보 불필요대용량 데이터에
도 적용이 쉬움

지역 특성에 따라 지역 클러스터링



한국관광 데이터랩

Data Lab

데이터 이용

① 군집화 평가를 통한 클러스터 개수 선정

전체 실루엣 계수와 개별 군집 실루엣 계수 평균값의
편차를 종합적으로 고려해 클러스터 개수 선정



실루엣 분석이란?

대표적인 군집화 평가 방법으로, 각 군집 간의 거리가
얼마나 효율적으로 분리돼 있는지를 나타냄

실루엣 계수 $s(i) = \frac{(b(i) - a(i))}{(\max(a(i), b(i)))}$ 로 판단

- (1) 실루엣 계수가 높을수록 효율적인 군집화
- (2) 하지만 개별 군집 실루엣 계수의 평균값의
편차가 크지 않아야 함



둘을 종합적으로 판단하기 위해 그래프로 확인!

클러스터 개수별 실루엣 계수 시각화

클러스터링 개수를 입력 받아 그에 따른
실루엣 계수를 면적으로 시각화한 함수 생성



```
def visualize_kmeans_plot_multi(cluster_lists, X_features):
    from sklearn.cluster import KMeans
    from sklearn.decomposition import PCA
    import pandas as pd
    import numpy as np

    n_cols = len(cluster_lists)
    fig, axes = plt.subplots(figsize=(4*n_cols, 4), nrows=1, ncols=n_cols)

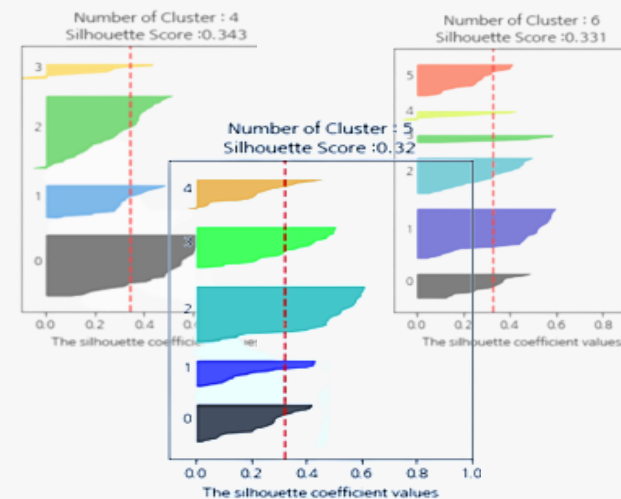
    pca = PCA(n_components=2)
    pca_transformed = pca.fit_transform(X_features)
    dataframe = pd.DataFrame(pca_transformed, columns=['PCA1', 'PCA2'])

    # 리스트에 기재된 클러스터링 개수들을 차례로 iteration 수행하면서 KMeans 클러스터링
    for ind, n_cluster in enumerate(cluster_lists):
        # KMeans 클러스터링으로 클러스터링 결과를 dataframe에 저장.
        clusterer = KMeans(n_clusters = n_cluster, max_iter=500, random_state=0)
        cluster_labels = clusterer.fit_predict(pca_transformed)
        dataframe['cluster'] = cluster_labels

        unique_labels = np.unique(cluster_labels_)

        # 클러스터링 결과값 별로 scatter plot 으로 시각화
        for label in unique_labels:
            label_df = dataframe[dataframe['cluster'] == label]
            if label == -1:
                # ... (rest of the code for plotting)
```

클러스터 개수 선정



- ✓ 실루엣 계수는 군집 개수 4개, 6개일 경우가
좀 더 높지만, 그래프를 고려해봤을 때
군집 개수 5개가 적합하다고 판단

지역 특성에 따라 지역 클러스터링



한국관광 데이터랩

Data Lab

데이터 이용

② 클러스터링에 영향을 미친 변수 확인

클러스터 컬럼이 추가된 데이터를
분류 모델인 RandomForest로 모델링 후
feature_importances_ 속성으로
분류에 영향을 미친 중요 변수 확인



```
X_train, X_test, y_train, y_test = train_test_split(data, target, test_size=0.2, random_state=34)
from sklearn.model_selection import GridSearchCV

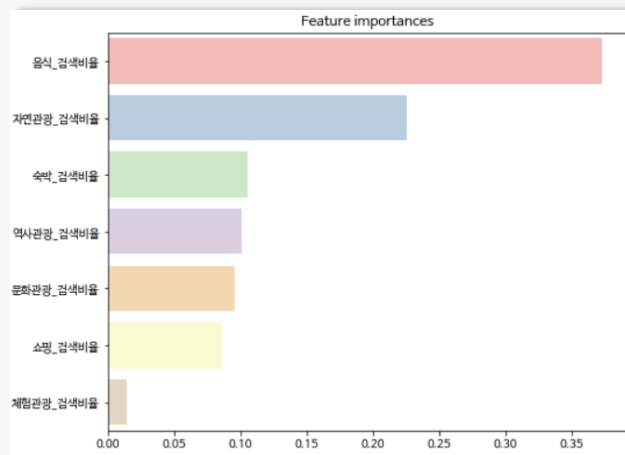
params = {
    'n_estimators': [100],
    'max_depth': [6, 8, 10, 12],
    'min_samples_leaf': [8, 12, 18],
    'min_samples_split': [8, 16, 20]
}

# RandomForestClassifier 객체 생성 후 GridSearchCV 수행
rf_clf = RandomForestClassifier(random_state=0, n_jobs=-1)
grid_cv = GridSearchCV(rf_clf, param_grid=params, cv=2, n_jobs=-1)
grid_cv.fit(X_train, y_train)

print('최적 하이퍼 파라미터: ', grid_cv.best_params_)
print('최고 예측 정확도: {0:.4f}'.format(grid_cv.best_score_))
rf_clf1 = RandomForestClassifier(n_estimators=300, max_depth=10, min_samples_leaf=8, min_samples_split=8, random_state=0)
rf_clf1.fit(X_train, y_train)
pred = rf_clf1.predict(X_test)
print('예측 정확도: {0:.4f}'.format(accuracy_score(y_test, pred)))
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

ftr_importances_values = rf_clf1.feature_importances_
ftr_importances = pd.Series(ftr_importances_values, index=X_train.columns)
ftr_top20 = ftr_importances.sort_values(ascending=False)[:20]
```

중요 변수 시각화



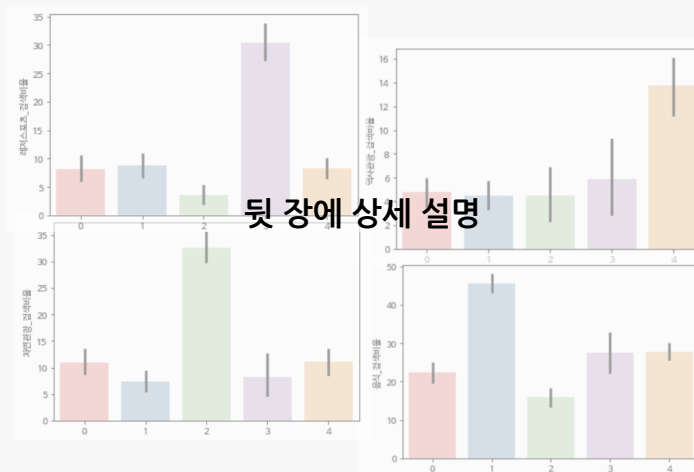
한 가지 변수에만 중요도가 집중되지 않고,
대부분의 변수에서 어느 정도의 중요성을
가지고 있는 것을 볼 수 있음

③ EDA를 통해 클러스터별 특성 파악

중요 변수를 중심으로 군집별 시각화



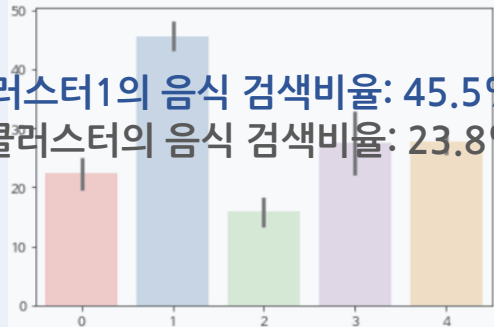
```
for i in range(0, len(column)):
    sns.barplot('cluster', column[i],
                data=df, palette='Pastel1')
    plt.show()
```





Cluster 1 맛집형

클러스터1의 음식 검색비율: 45.5%
타 클러스터의 음식 검색비율: 23.8%

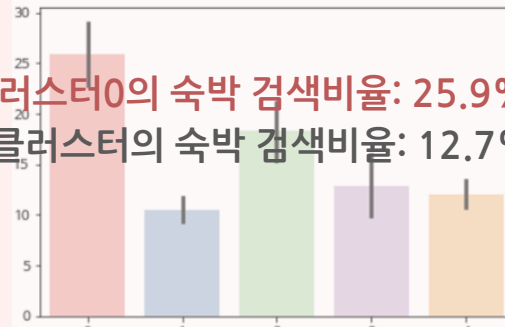


진주 천안 청주 구미



Cluster 0 휴식형

클러스터0의 숙박 검색비율: 25.9%
타 클러스터의 숙박 검색비율: 12.7%



진도 정선 평창 영월



Cluster 4 역사관광형

클러스터4의 역사관광 검색비율: 13.8%
타 클러스터의 역사관광 검색비율: 4.7%

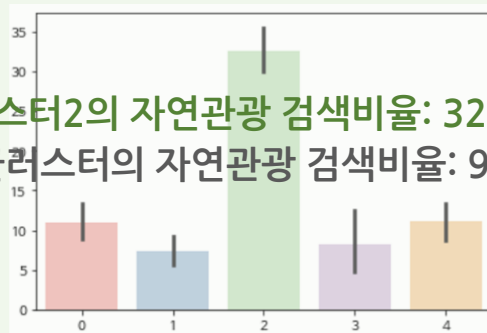


공주 안동 구례 부여



Cluster 2 자연관광형

클러스터2의 자연관광 검색비율: 32.6%
타 클러스터의 자연관광 검색비율: 9.5%

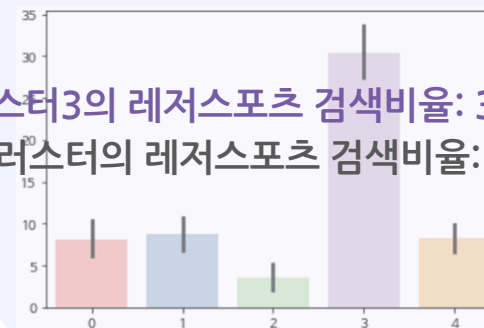


청송 보령 태안 울진



Cluster 3 레저스포츠형

클러스터3의 레저스포츠 검색비율: 30.3%
타 클러스터의 레저스포츠 검색비율: 7.5%



음성 진천 의성 횡성

→ 한달살기 지역 추천 서비스

→ 지자체별 개선점 제안



정책 제안

클러스터와 키워드에 따라 한달살기 지역을 추천해주는 서비스



크롤링 데이터 이용

네이버 여행정보(travel.naver.com)에서 지역별 설명, 태그 등 스크래핑

강원도 고성군

Goseong-gun

위치 길찾기정보 · 지도

날씨 흐림, 22° 30° 날씨정보

강원도 최북단에 위치한 고성. 고성 통일 전망대는 군상할 수 있다. DMZ 박물관은 우리나라 분단의 역사해수욕장과 철새 전망시설을 갖춘 송지호 해수욕장!

고성군 > 하이라이트

[#스킨스쿠버](#) [#에번](#) [#데디베어](#) [#바다낚시](#)

지역별 태그



```
# 지역별 이름 스크래핑
name = driver.find_element_by_xpath('//*[@id="__next"]/div/div[2]/main/div[2]/div[1]/div/b').text
region_name.append(name)
driver.find_element_by_tag_name('body').send_keys(Keys.PAGE_DOWN)
driver.implicitly_wait(3)

# 지역별 하이라이트 태그 스크래핑
highlights = []
try:
    driver.find_elements_by_class_name('keyword-31-0m')
except:
    region_highlight.append('하이라이트 태그 없음')
else:
    for j in driver.find_elements_by_class_name('keyword-31-0m'):
        highlights.append(j.text)
        region_highlight.append(highlights)
```

서비스 알고리즘



```
print('가장 중요시하는 여행 형태를 선택해주세요!')
choice = int(input('1. 휴식 2. 음식 3. 자연 4. 레저 5. 역사'))
print('선택하신 여행 형태는 다음과 같은 항목이 있습니다.')

cluster = data[data['군집'] == choice-1]
cluster_tag = set(cluster['태그'])
cluster_tag.update(set(cluster['태그2']), set(cluster['태그3']), set(cluster['태그4']), set(cluster['태그5']))
print('선택하신 여행 형태는 다음과 같은 항목이 있습니다.')
cnt = 0
for i in cluster_tag:
    cnt += 1
    if cnt % 7 == 0 or cnt == len(cluster_tag):
        print(f'({i})')
    else:
        print(f'({i}), ', end='')
print()
tag_list = input('이중 원하는 여행지 특징을 선택해주세요!').split(',')
tag_choice_list = list()

for i in range(0, len(cluster)):
    cnt = 0
    for j in tag_list:
        if j in cluster.iloc[i][1:5]:
            cnt += 1
    tag_choice_list.append(cnt)
print('선택하신 (태그, 군집)과 태그 중 최대 (태그, 군집)에 일치합니다.')
index_list = list(filter(lambda x: tag_choice_list[x] == max(tag_choice_list), range(len(tag_choice_list))))
choice_place = cluster.iloc[index_list]
choice_place.sort(key=lambda x: x['태그'], reverse=True)
```

운영 방안 제안

yanolja **여기어때.**

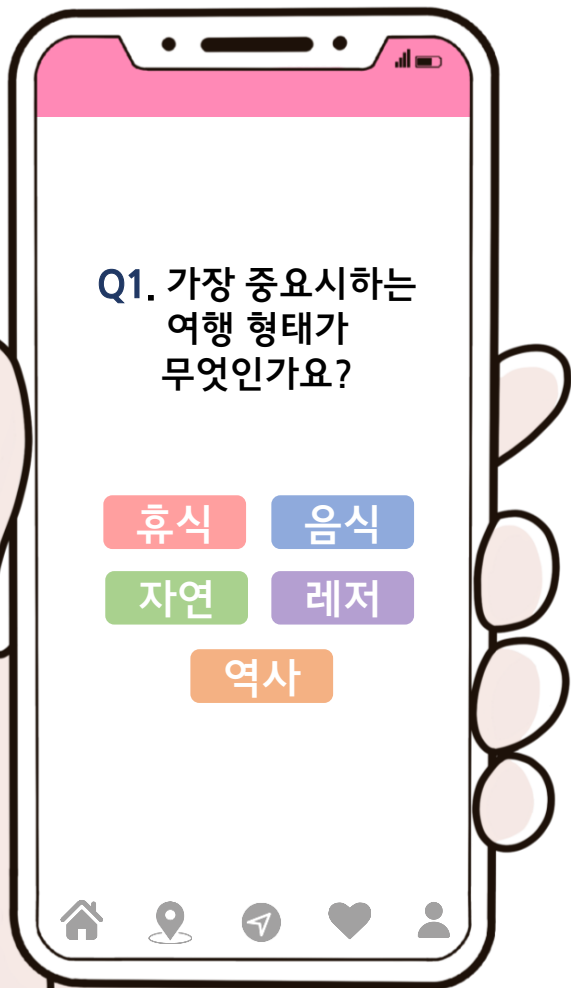
야놀자나 여기어때 같은
여행 어플리케이션과 협업해
추천 서비스 페이지 제작

☑ 사용자수가 많은 앱이니
홍보 효과도 누릴 수 있음

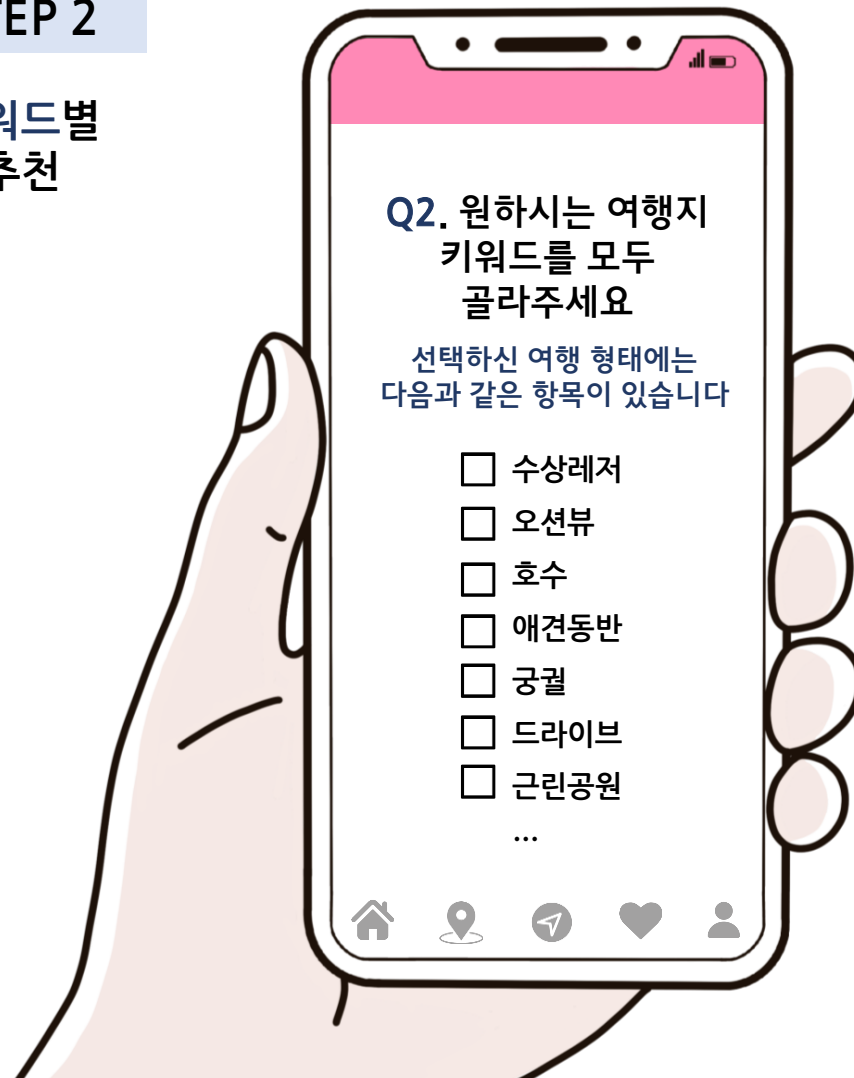


클러스터와 키워드에 따라 한달살기 지역을 추천해주는 서비스

STEP 1

클러스터별
추천

STEP 2

키워드별
추천

클러스터와 키워드에 따라 한달살기 지역을 추천해주는 서비스

Persona 1



삭막한 도시에서 벗어나 자연환경을 만끽하고 싶은 30대 남성
 서울에서는 즐기기 힘든 자연 체험활동을 원함

Q1. 가장 중요시하는 여행 형태가 무엇인가요?

자연

Q2. 원하시는 여행지 키워드를 모두 골라주세요
 선택하신 여행 형태에는 다음과 같은 항목이 있습니다

갯벌체험

수상체험

낙시체험

폭포

스노클링

해수욕장

해변

래프팅

시군구	만족도	키워드	설명	태그1	태그2	태그3	태그4	태그5
완도군	4.25	세트장, 공룡알, 유적지, 전복구이	시간이 느리게 흘러가는 완도군....	해변	해수욕장	전망대	근린공원	일몰
보령시	4.18	유람선, 맛조개, 수산시장, 휴양림	모세의 기적이 존재하는 보령....	포토존	해변	아이와	낙시체험	해수욕장

Persona 2



외국인 친구와 함께 전통적인 곳에서 머물고 싶은 20대 여성
 한국 문화와 역사를 느낄 수 있는 곳을 원함

Q1. 가장 중요시하는 여행 형태가 무엇인가요?

역사

Q2. 원하시는 여행지 키워드를 모두 골라주세요
 선택하신 여행 형태에는 다음과 같은 항목이 있습니다

브런치

궁궐

산책

전통가옥

한옥카페

루프탑

빵투어

시군구	만족도	키워드	설명	태그1	태그2	태그3	태그4	태그5
부여군	4.0	연꽃, 유적지, 연잎밥, 야경	백제의 역사가 잠들어있는 부여...	포토존	궁궐	산책	데이트	아이와

① 군집 특성에 따른 지자체별 개선점 제안

예시 사례

강원도 평창군



Cluster 0 휴식형

Cluster 0 휴식형의 특징

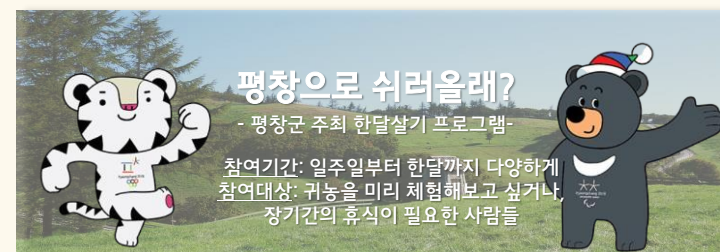
- ☑ 다른 클러스터보다 관광지 개수와 방문자 수가 모두 적은 편
- ☑ 따라서 체험, 관광보다는 휴식과 힐링에 초점을 두는 것을 추천

AS-IS

강원도 한 달 살기(평창) 신청 안내	
신청기간	• '20.6.15.(월) ~ 6.30.(화)
운영기간	• '20.7.1.(수) ~ 11.30.(월)
참가대상	• 강원도 평창에 관심이 있고, 평창군 귀농·귀촌을 희망하는 도시민으로 평창군 거주 희망 하는 자 • 주민등록지 기준(만18세 이상, 도시지역인 자)
참가기간	• 한 달(30일)
참가비용	• 숙박, 농촌체험프로그램·교육 등 제공(무료) • 다만 식사, 교통비 등 생활비 부담
주요내용	• 주민만남, 문화관광지 탐방, 귀농귀촌교육 참여, 선도·귀농 농가 견학, 일손 돕기, 재능기부, 귀농귀촌 정보습득 등 • 참가자는 체험마을에서 운영하는 프로그램 및 교육 의무적으로 참여 (1일당 1회 이상)

- ☑ 평창으로의 귀농, 귀촌을 희망하는 사람들만을 대상으로 함
- ☑ 프로그램 기간이 한 달로 정해져 있음 (이보다 짧게는 신청이 불가능함)
- ☑ 체험마을 프로그램에 1일당 1회 이상 의무적으로 참여해야 함

TO-BE



- ☑ 휴식형 클러스터는 귀농, 귀촌을 미리 체험해보기 위한 사람들을 타겟으로 하는 것이 적절하긴 함
하지만 귀농, 귀촌 희망자들은 대부분 고연령층. 지역 활성화와 홍보를 위해서는 젊은 층이 필요!
따라서 단순히 휴식을 원하는 사람도 참가대상에 포함시키면 SNS 홍보 등을 통해 한달살기의 효과를 극대화할 수 있을 것으로 보임
- ☑ 휴식형 클러스터는 관광지 개수가 적으므로, 한 달이라는 시간은 길게 느껴질 수 있음
일주일~한 달까지 참가기간을 다양화해야 함
- ☑ 휴식이 목적인 사람들에게 1일 1회 이상의 프로그램 참가는 부담일 수 있음.
주당 2-3회로 변경하는 것이 적절해 보임

① 군집 특성에 따른 지자체별 개선점 제안

예시 사례 충청북도 충주시



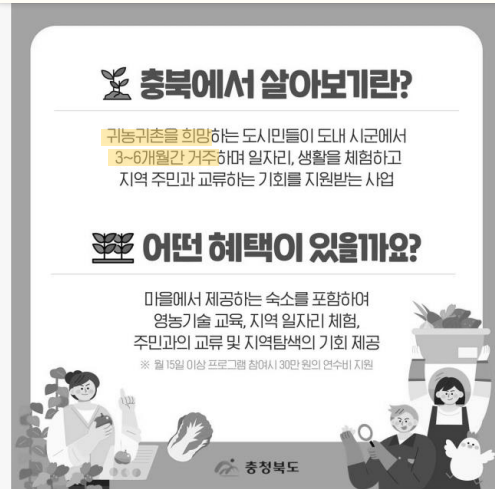
Cluster 3 레저스포츠형

Cluster 3 레저스포츠형의 특징

- ☑ 컨트리클럽(CC) 골프장으로 유명한 지역들이 이 클러스터에 해당
- ☑ 다른 클러스터에 비해 만족도의 평균이 4.04로 낮은 편

AS-IS

충북도, '충북에서 살아보기' 사업 본격 운영



- ☑ 충청북도 차원의 일괄적인 한달살기 프로그램
각 시군마다 특성이 다른데 그 점을
전혀 반영하지 못하고 있음
- ☑ 귀농, 귀촌이 목적인 사람들만을 대상으로 함
- ☑ 참여기간이 3-6달로 너무 길어서 부담스러움

TO-BE



- ☑ 충청북도는 시군별로 한달살기 프로그램을
다르게 기획하는 게 적합해 보임.
(Ex. 충주는 충주대로, 제천은 제천대로)
- ☑ MZ 세대의 새로운 취미가 되어가는 골프!
골프를 중심으로 한달살기 프로그램을 기획한다면
중장년층 뿐만 아니라 MZ 세대도
관광객으로 유입할 수 있을 것이라 예상
- ☑ 한달 집중 골프 레슨 등의 골프 프로그램을
중심으로 하되, 충주호 유람선 체험, 충주 고구려비
탐방 등의 골프 외 프로그램도 일부 추가해
만족도도 높이고, 다른 관광지도 자연스럽게 홍보

② 지역별 만족도 회귀 분석을 통한 개선점 제안



국민여행조사 데이터 이용

지역별 만족도 OLS 회귀분석



```
## 전진 단계별 선택법
variables = col ## 설명 변수 리스트

y = df['만족도'] ## 반응 변수
selected_variables = [] ## 선택된 변수들
sl_enter = 0.05
sl_remove = 0.05

sv_per_step = [] ## 각 스텝별로 선택된 변수들
adjusted_r_squared = [] ## 각 스텝별 수정된 결정계수
steps = [] ## 스텝
step = 0
while len(variables) > 0:
    remainder = list(set(variables) - set(selected_variables))
    pval = pd.Series(index=remainder) ## 변수의 p-value
    ## 기준에 포함된 변수와 새로운 변수 하나씩 돌아가면서
    ## 선택 모형을 적당한다.
    for col in remainder:
        X = df[selected_variables+col]
        X = sm.add_constant(X)
        model = sm.OLS(y,X).fit()
        pval[col] = model.pvalues[col]
```



회귀분석 중요지표

R²: 회귀모형의 설명력을 나타냄Adj-R²: R²의 문제점을 보완한 통계량

(유의미한 독립변수만 반영)

DW-통계량: 1.5~2.5 사이의 값이 나오면

자기상관이 없다고 볼 수 있음.

관광 비인기 지역의 전반적인 시사점

관광 비인기 지역

자전거

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.2965	0.019	15.469	0.000	0.258	0.335
주요이용교통수단9	1.1865	0.077	15.469	0.000	1.032	1.3
여행지선택이유13	-1.0855	0.187	-5.811	0.000	-1.461	-0.71
여행활동21	0.6419	0.246	2.609	0.012	0.148	1.136
직업	0.0313	0.010	3.015	0.004	0.010	0.052
숙박시설11	0.9295	0.263	3.534	0.001	0.400	1.459
혼인상태	-0.2789	0.078	-3.580	0.001	-0.436	-0.122
학력	-0.1561	0.066	-2.352	0.023	-0.290	-0.023

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.2472	0.013	19.099	0.000	0.221	0.273
주요이용교통수단9	0.9887	0.052	19.099	0.000	0.886	1.0
여행활동14	0.9235	0.193	4.775	0.000	0.539	1.308
여행활동8	0.3883	0.119	3.263	0.002	0.151	0.625
여행지선택이유8	-0.3434	0.096	-3.585	0.001	-0.534	-0.15
연령	0.0505	0.023	2.185	0.032	0.004	0.096
사전예약1	0.2909	0.068	4.301	0.000	0.156	0.426
여행활동5	0.3145	0.108	2.919	0.005	0.100	0.529

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.7738	0.033	23.495	0.000	0.709	0.839
주요이용교통수단9	1.5476	0.066	23.495	0.000	1.417	1.678
여행지선택이유14	0.3161	0.071	4.436	0.000	0.175	0.457
여행정보획득경로1	-0.3731	0.143	-2.609	0.010	-0.575	-0.171
여행지선택이유7	-1.3593	0.420	-3.237	0.002	-2.190	-0.529
숙박시설2	-0.9934	0.294	-3.377	0.001	-1.575	-0.412
개인소득	0.1268	0.029	4.385	0.000	0.070	0.184

VS

관광 인기 지역

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.0228	0.056	71.333	0.000	3.910	4.136
숙박시설1	-0.8629	0.194	-4.442	0.000	-1.252	-0.474
여행활동16	-1.4342	0.233	-6.144	0.000	-2.081	-0.787
여행지선택이유14	0.4117	0.238	1.729	0.088	0.236	0.587
사전예약2	1.0989	0.238	4.616	0.000	0.622	1.576
여행지선택이유9	0.5203	0.171	3.035	0.004	0.177	0.864
여행지선택이유8	-0.3095	0.111	-2.796	0.007	-0.531	-0.088

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.3400	0.135	24.667	0.000	3.071	3.609
여행활동5	0.7879	0.156	5.046	0.000	0.478	1.098
사전예약2	0.9802	0.298	3.291	0.001	0.389	1.572
개인소득	0.1093	0.033	3.301	0.001	0.044	0.174
여행지선택이유11	-0.2686	0.108	-2.496	0.014	-0.482	-0.055
여행활동11	1.0543	0.390	2.700	0.008	0.279	1.830
숙박시설9	1.2209	0.391	3.125	0.002	0.445	1.997



비유명 지역은 지자체별로 특성이 많이 달라서 만족도에 영향을 미치는 공통요인이 거의 없음.



그럼에도 불구하고, 인기 지역과 달리 비인기 지역에서 자전거는 만족도에 큰 영향을 미침. 교통 인프라가 부족한 비유명 지역에서 자전거는 이를 보완할 수 있는 수단이기 때문

② 지역별 만족도 회귀 분석을 통한 개선점 제안

예시 사례 강원도 고성군



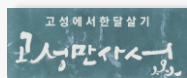
고성군 만족도 회귀분석 결과

- ☑ 통계량을 봤을 때, 유의미한 결과
- ☑ **만족도 긍정요인**
자전거, 자연휴양림 내 숙박
- ☑ **만족도 부정요인**
여행지선택이유 중 교육성

+ Cluster 0의 특성도 함께 고려!

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.2966	0.019	15.469	0.000	0.258	0.335
주요이용교통수단9	1.1865	0.077	15.469	0.000	1.032	1.341
여행지선택이유13	-1.0855	0.187	-5.811	0.000	-1.461	-0.710
여행활동21	0.6419	0.246	2.613	0.012	0.148	1.133
직업	0.0313	0.010	3.015	0.004	0.010	0.052
숙박시설11	0.9295	0.263	3.534	0.001	0.400	1.459
혼인상태	-0.2789	0.078	-3.580	0.001	-0.436	-0.122
학력	-0.1561	0.066	-2.352	0.023	-0.290	-0.023

AS-IS



고성에서 한달살기 '고성만사성'

○ 참가대상 : 분야별 인원수 제한 없음

참여 그룹	대상
디스커버 Discover 그룹 : 자연+위라벨	강원도 고성에서 살아보고 싶은 일반 청년, 사회적 환경 또는 일상에 지친 청년 등
그로우업 Grow up 그룹 : 창의+도전	귀촌을 꿈꾸거나 디지털 노마드(랜선 청년)의 삶을 경험해보고 싶은 디자이너, 문화기획, 예술분야 등 인터넷 연결만으로도 프로젝트를 실행할 수 있는 직업적 전문성을 가진 청년

○ 각종 체험지원

- 화요일 로컬투어 : 고성8경 중심의 **역사·문화** 투어
- 목요일 명상투어 : 고성의 **자연경관**에서 산책하는 명상 투어

○ 숙박비 : 문암진리 마을 **민박시설** 숙박 지원

☑ 참가 대상의 선정은 적절(휴식을 원하는 청년들)

☑ 역사,문화 체험과 자연경관 체험 모두 제공

☑ 숙소로 마을 민박시설만 제공

TO-BE

☑ 자연 속의 휴양이 만족도가 높아지는 요인이고, 지역의 교육성 측면이 부족하기 때문에, 체험 중 **역사/문화** 관련 프로그램을 **줄이고**, **자연 속 휴식 프로그램**을 늘리는 것이 좋을 듯함

☑ 만족도 긍정요인에 '자연휴양림 내 숙박'이 있고, 휴식형 클러스터는 숙소가 중요하기 때문에, **합리적인 가격의 숙박시설을 발굴·지원**해야 함

참여자
30대 신00씨

"프로그램 후 귀촌에 대한 마음이 더 확고해졌는데 객관적인 주거와 부동산 정보가 없어서 불편해요. 지자체가 신경썼으면 좋겠어요."

☑ 만족도에 큰 영향을 미치는 교통수단인 **공용 자전거** 활성화 필요

"원래 제주에 가려고 생각했지만, 고성에 와보니 제주만큼이나 좋았어요. 다만 대중교통이 편리해졌으면 좋겠어요. 공용 자전거를 더 설치하면 좋을 거 같아요."

참여자
20대 최00씨

② 지역별 만족도 회귀 분석을 통한 개선점 제안

예시 사례 전라북도 고창군



고창군 만족도 회귀분석 결과

- ☑ 통계량을 봤을 때, 유의미한 결과
- ☑ 만족도 긍정요인
자전거, 여행활동 중 지역축제
- ☑ 만족도 부정요인
인터넷을 통한 여행 정보 획득,
기대보다 못한 쇼핑·교통,
숙소 중 모텔·여관

+ Cluster 4의 특성도 함께 고려!

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.7738	0.033	23.495	0.000	0.709	0.839
주요이용교통수단9	1.5476	0.066	23.495	0.000	1.417	1.678
여행지선택이유14	0.3161	0.071	4.436	0.000	0.175	0.457
여행정보획득경로1	-0.3731	0.102	-3.648	0.000	-0.575	-0.171
여행지선택이유7	-1.3593	0.420	-3.237	0.002	-2.190	-0.528
숙박시설2	-0.9934	0.294	-3.377	0.001	-1.575	-0.411
개인소득	0.1268	0.029	4.385	0.000	0.070	0.184
여행활동11	0.3013	0.105	2.866	0.005	0.094	0.509

AS-IS

역사문화관광광도시 고창군에 머물며 휴식, 체험, 힐링과 더불어 추억을 쌓을 수 있는 더없이 좋은 기회로 그동안 패키지여행에 아쉬움을 느끼는 분들께 가고픈 날에 가고픈 곳곳의 숨은 명소를 찾아 전라북도 고창의 맛·멋과 함께 여유로움을 만끽하는「전라북도 고창에서 한 달 여행하기」참여자들을 추가 모집하오니 많은 참여 바랍니다.

○ 숙박비 : 1일당 5만원(가족은 1실)

※ 고창군에서 지정한 시설(7개소)을 이용하며 비용은 참여자에게 지급
(영수증과 결과물 제출시 사후정산)

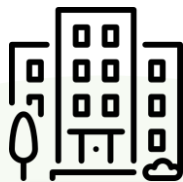
- ☑ 숙소 지원은 양호함. 지정 숙소 7개 중 품질인증업소로 지정된 호텔도 존재
- ☑ 프로그램 포스터나 홍보자료의 부재
- ☑ 역사문화관광광도시라고 소개하고 있지만, 막상 프로그램에 역사문화 체험은 거의 없음

TO-BE

- ☑ 지원 숙소 중 품질인증업소로 지정된 숙소의 비중을 높여, 만족도 부정요인으로 나타난 모텔/여관 등의 저가 숙박 시설 이용으로 인한 불만족을 최소화해야 함
- ☑ 만족도 부정요인 중 '인터넷을 통한 여행 정보 획득'이 있음. 고창군은 전라도 타 군 대비 공식 소통채널 인지도도 2배 이상 부족하고, 한달살기 프로그램에 대한 홍보가 미흡함
지자체에서 여행 관련 공식 자료를 배포해 이용객의 정보 부족으로 인한 불편함을 최소화하는 것이 필요해 보임
- ☑ 만족도 긍정요인 중 '지역축제'가 있고, '국내 지역별 관광명소데이터'에서 확인해본 결과 고창군 관광명소의 27%가 지역축제임
고창의 '청보리밭 축제'는 2022년 대한민국 대표브랜드로 선정
한달살기 프로그램을 지역축제와 연결시키면 만족도가 향상될 것으로 예상함



기대 효과



지자체

지역별로 맞춤형 정책 실시

- ☑ 기존 관광 비인기 지역의 경쟁력 강화
- ☑ 지역별 특성을 반영한 차별화
- ☑ 관계인구의 증가로 인한 장기적 효과
- ☑ 귀농 유도 효과와 지역 홍보 효과
동시 발생



관광객

발전된 한달살기 프로그램 이용

- ☑ 개인별 선호 유형에 맞는 지역에서의 프로그램
- ☑ 품질인증 받은 양질의 숙소 이용으로
치안과 위생 걱정 최소화
- ☑ 지자체로부터 양질의 정보 받아
더 편리해진 한달살기

대한민국의 균형적인 관광 발전에 이바지