

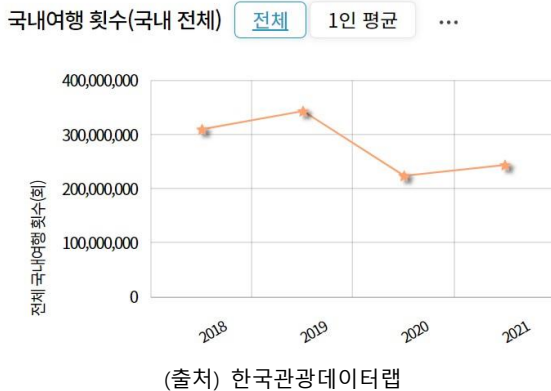
전국 지역축제 성패 요인 분석 및 군집분석

- 성패 요인 기반 축제 군집별 프로파일링 및 개선 방안 제안 -

1) 분석 주제

본 분석의 주제는 '전국 지역축제 성패 요인 분석 및 군집분석'으로, 2022년 지역 축제 데이터를 기반으로 지역 축제의 성패에 영향을 미치는 주요한 요인을 분석한다. 나아가 분석을 통해 찾은 성패 요인을 활용하여 총 123개의 지역 축제를 군집화한다. 각 축제 군집을 성공과 실패로 나누어 군집별 성패 원인에 대한 인사이트를 찾고, 성과가 부진했던 축제 군집에 대한 개선 방향을 제시하고자 한다.

2) 분석의 배경 및 목적



관광지식정보시스템에서 발표한 2022년 전체 국내여행은 283,722,000건으로 코로나 완화 이후 국내 관광은 점차 회복되고 있지만, 2019년 이전만큼의 관광산업 회복을 위해서는 더 많은 노력이 필요한 것으로 보인다.

문화체육관광부 역시 올해 국정과제 중 하나로 <여행으로 행복한 국민, 관광으로 발전하는 대한민국>를 선정하여, 시장 회복 및 재도약, 지역특화 관광개발을 위해 힘쓰고 있다.

우리는 이를 위한 노력으로써 지역축제 개선을 통한 국내 관광객 유입에 주목했다. 지역축제 성패 요인 분석을 통해 데이터에 기반한 개선 방향을 제시함으로써, 보다 효과적인 관광시장 회복 및 국내 관광 활성화에 기여하고, 동시에 지역 특색을 살린 균형 있는 지역 발전과 더 효율적인 지역축제의 개선 방안을 제시하고자 한다.

3) 활용 데이터 선정

문화체육관광부

- 2023 년 지역축제 개최 계획

문화빅데이터플랫폼

- 국내 여행지역 관심도
- 전국 축제 트렌드 지도
- 국내 지역별 관광명소데이터(시군구 행정도 100 미터격자코드 포함)

KOSIS

- 교통문화지수

한국관광 데이터랩

- 문화관광축제 주요 지표

- 목적지 검색순위
- 연도별 방문자 추이
- 성/연령별 내국인 방문자
- 관광 소비액
- 업종별 소비액

지역축제의 성공 및 실패에 영향을 주는 요인을 찾기 위해, 축제의 구성 및 특징을 나타낼 수 있는 축제의 '요인'과 관련된 지표를 수집하였다. 수집된 변수는 개최일정, 개최지, 예산, 방문객, 여행지역 관심도, 축제별 검색량 등이다. 나아가 전국 지역축제의 군집별 분석을 위해, 한국 관광 데이터랩으로부터 연령별 방문객, 업종별 소비액 등을 담은 축제 '결과'에 관한 데이터도 수집하여 활용하였다.

4) 분석 내용 및 결과

4.1) 데이터 전처리

성패 요인 분석을 위한 변수 중요도 확인과 군집분석을 위한 데이터셋을 구성하였다. '2023 년 지역축제 개최 계획' 데이터셋을 중심으로 데이터를 병합하였다. 각 데이터셋을 축제명을 기준으로 병합한 다음, 결측치가 있는 행은 전부 제거했다.

데이터의 일관성을 위해 '축제유형'과 '개최일정' 변수는 수기로 전처리를 진행하였다. '개최일정' 일 경우 직접 2022 년 축제 개최 일정을 확인하여 수정했다. '축제유형' 변수의 경우, 축제의 테마에 따른 분류가 일관적이지 않아서, [문화예술, 특산물, 역사, 전통, 자연, 체험, 공연] 7 가지 분류로 각 지역축제를 재분류했다. 이 때, '축제유형' 변수의 재분류에 대한 세부적인 기준은 아래와 같다.

- 전통 vs 역사 : 민속극, 한복 등 문화와 관련된 축제는 전통으로, 백제, 조선 등 우리 민족의 역사와 관련된 축제를 역사로 구분
- 공연 vs 문화예술 : '록 페스티벌', '댄싱 카니발' 등 공연이 주제인 축제를 공연으로 분류하고, 다양한 즐길 거리를 포함한 축제는 문화예술로 구분
- 체험 vs 자연 : 직접 참여하는 활동이 주요한 축제는 체험으로, 나머지를 자연으로 구분

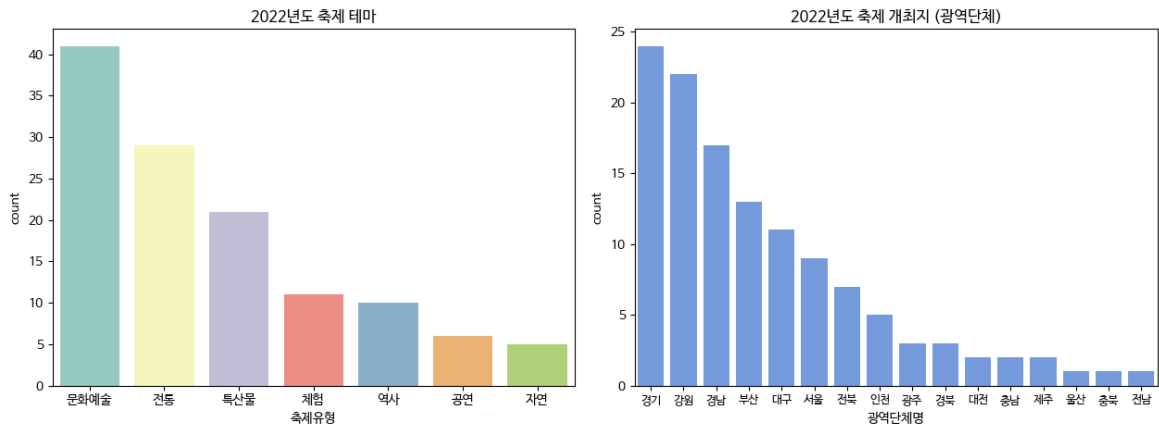
	축제명	광역단체명	기초단체명	축제유형	예산합계	국비	지방비	방문객_내국인	검색량	교통접수	...	군구	주변관광명소개수	공연	문화예술	역사	자연	전통	체험	특
0	종로한복축제	서울	종로구	전통	100	0	100	157541	4770	82.94	...	0 1	325	0	0	0	0	1	0	
1	구로G페스티벌	서울	구로구	공연	868	0	868	100000	30470	81.23	...	0 1	58	1	0	0	0	0	0	
2	한성백제문화제	서울	송파구	역사	1967	175	1172	249000	43560	79.15	...	0 1	90	0	0	1	0	0	0	
3	4.19혁명국민문화제	서울	강북구	역사	770	300	470	57000	12	78.05	...	0 1	79	0	0	1	0	0	0	
4	서울장미축제	서울	종로구	자연	600	0	600	730000	760	78.74	...	0 1	51	0	0	0	1	0	0	
...
117	천령문화제	경남	함양군	전통	500	0	500	25000	390	83.51	...	1 0	315	0	0	0	0	1	0	
118	함진배강스축제	경남	함천군	체험	260	0	260	51212	60	85.72	...	1 0	409	0	0	0	0	0	1	
119	대야문화제	경남	함천군	체험	610	0	610	30000	980	85.72	...	1 0	409	0	0	0	0	0	1	
120	월대천축제	제주	제주시	체험	20	0	20	1000	20	84.74	...	0 0	667	0	0	0	0	0	1	
121	이호테우축제	제주	제주시	체험	30	0	30	29500	35	84.74	...	0 0	667	0	0	0	0	0	1	

122 rows × 31 columns

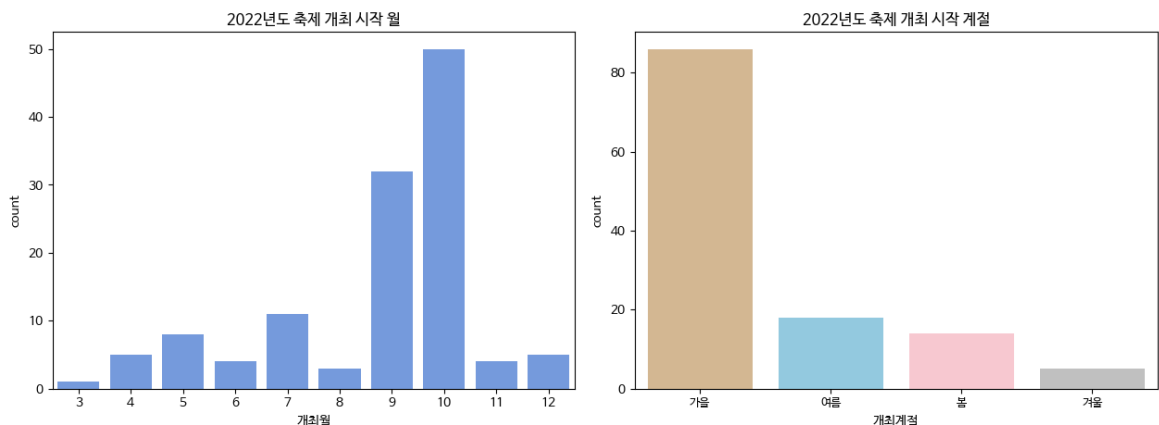
최종적으로 구성한 변수 중요도, 군집분석을 위한 데이터셋은 위와 같다. 총 122 개의 전국의 지역축제에 대한 데이터들로 구성하였다. 기본 변수는 축제명, 광역단체명, 기초단체명, 축제유형, 예산합계, 국비, 지방비, 방문객_내국인, 검색량, 교통접수, 개최기간, 개최월, 20 대지역관심도, 30 대지역관심도, 40 대지역관심도, 50 대지역관심도, 60 대지역관심도, 주변관광명소 개수이다. 파생변수로 일평균방문객, 방문객/예산을 추가하였고, 범주형 변수인 광역단체명을 수도권, 광역시, 도로, 기초단체명을 시, 군, 구로 더미변수화 하였고, 7 가지 범주의 축제유형 변수도 더미변수로 만들었다.

4.2) 탐색적 데이터 분석 (EDA)

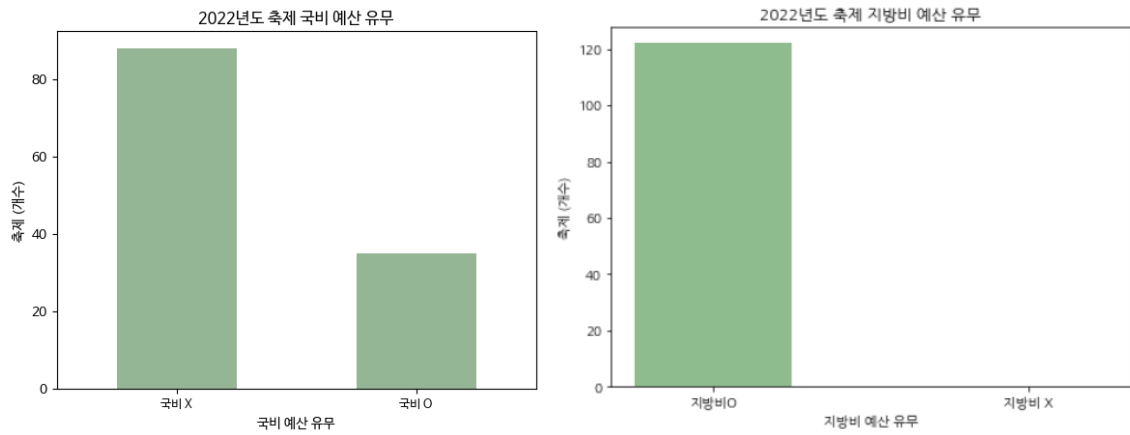
최종 데이터셋에 대해 탐색적 데이터 분석(EDA)을 진행하였다.



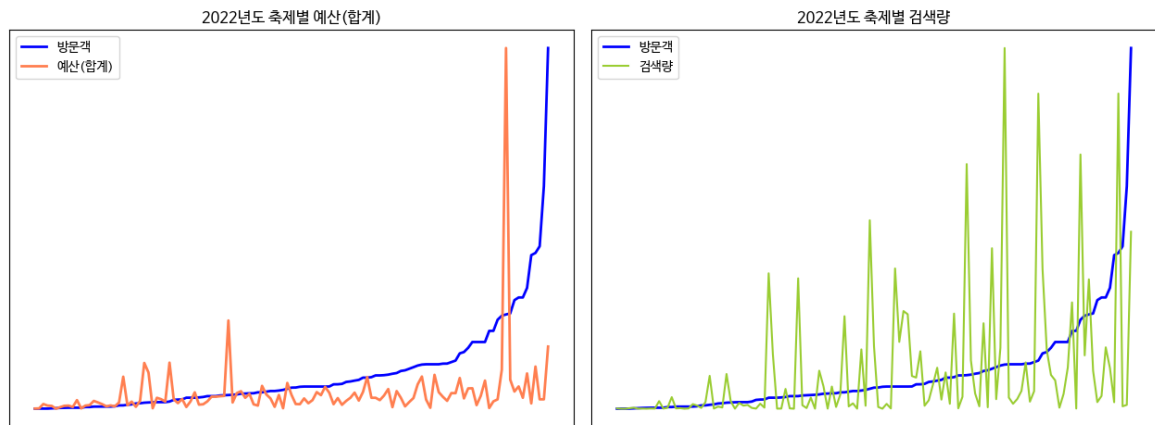
가장 많은 축제 유형은 '문화예술'이고, '전통'과 '특산물'도 높은 수를 기록하였다. 축제의 개최지는 경기도, 강원도, 경상남도 등이 상위를 차지했다. 인구가 많은 서울은 비교적 적은 수의 지역축제가 개최됨을 확인할 수 있다.



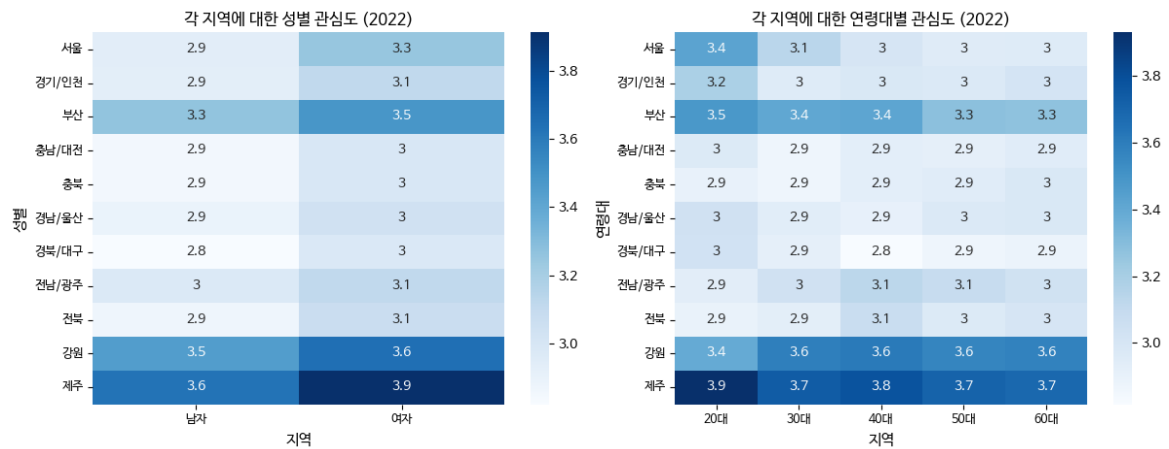
개최 일정을 시각화한 결과, 많은 축제가 9, 10 월의 가을에 개최되고 있음을 확인할 수 있다. 가을에 개최되는 경우가 대부분이고, 여름과 봄에도 일부 축제가 개최되지만 겨울에 개최되는 사례는 많지 않았다.



예산 지원 여부를 시각화한 결과이다. 국비를 지원받지 않는 지역축제는 국비 지원이 있는 축제의 두 배 이상이다. 하지만 지방비의 경우, 분석 대상인 모든 지역축제가 지방비 예산을 할당 받고 있음을 확인할 수 있다.



위 그래프는 축제 방문객과 예산, 검색량을 시각화한 것이다. 예산은 방문객 수와의 유의미한 상관관계를 보이지 않았다. 검색량의 경우, 방문객 수가 늘어남에 따라 함께 커지는 경향성을 보이긴 하지만, 유의미한 상관관계로 보기는 어렵다.



각 축제 개최지역에 대한 여행지로서의 관심도를 시각화한 결과이다. 제주, 강원, 부산은 모든 경우에서 높은 지역 관심도가 나타났다. 20, 30 대는 수도권에 대한 관심도도 높게 나타났다.

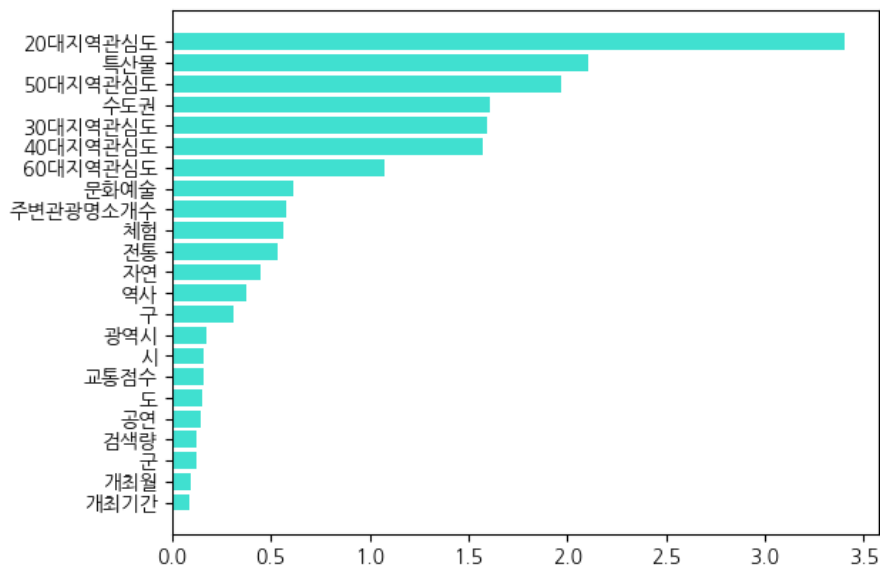
4.3) 지역축제 성패 요인 분석

Feature Importance(변수 중요도)를 활용해 지역축제의 성패에 영향을 미치는 주요한 요인을 분석하였다. y 변수(종속 변수)에 대하여 여러 회귀 모델을 적합한 후, 도출된 변수 중요도를 종합하여 y 변수에 영향을 미치는 주요한 성패 요인을 찾고자 했다.

이 때, 모델링에 사용되는 y 변수는 '방문객/예산'으로 결정하였다. 투자 대비 지역축제의 성공 정도를 평가하기 위해, 단순 방문객 수가 아닌 예산 대비 방문객 수를 y 변수로 지정하였다.

다음으로 LGBM Regressor, XGB Regressor, Decision Tree Regressor, Random Forest Regressor, CatBoost Regressor, Linear Regressor, Ridge Regressor 의 회귀 모형들을 위의 데이터셋을 활용해 모델링하고, 각각의 feature importance 를 도출하였다. 이 때, MSE(평균제곱오차)가 200 이상을 기록한 Linear, Ridge Regressor 는 성능이 좋지 않다고 판단하여 고려하지 않았다. 선별된 각 모델의 feature importance 결과를 minmax scaling 한 후, 전부 더해 그래프로 표현한 것은 아래와 같다.

```
1 sorted = lgbm_sorted['importance'] + xgb_sorted['importance'] + dt_sorted['importance'] + rf_sorted['importance'] + cat_sorted['importance']
2 final = pd.DataFrame(zip(lgbm_sorted['feature'], sorted), columns=['feature', 'importance'])
3 final_rank = final.sort_values(by=['importance'])
4 final_rank
```



종합한 Feature importance 결과에 따르면, 여행 지역에 대한 관심도와 축제 유형(축제의 주제), 개최 지역이 축제의 성과에 많은 영향을 주는 주요한 성패 요인으로 나타났다. 특히, **20 대**의 지역 관심도가 지역축제의 성과에 비교적 큰 영향을 준다는 것을 유추할 수 있다. 20 대 다음으로는 **50 대**의 관심도가 큰 영향을 주는 것으로 나타났다. 또한, 축제 유형 중에서도 **특산물**이 지역축제의 성과에 큰 영향을 미치는 것으로 나타났다. 개최 지역과 관련해서는, **수도권 여부**에 따라 지역축제의 성과가 달라질 것을 추측할 수 있다. 구체적으로 이러한 성패

요인이 지역축제의 성과에 긍정 혹은 부정적으로 작용하는지에 대해서는 이어지는 군집분석에서 더 자세히 살펴볼 것이다.

위 그래프에서 feature importance 의 합계가 급격히 감소하는 지점인 광역시 이하의 변수들은 y 변수인 '방문객/예산'에 미치는 영향이 크지 않다고 판단하여 버리기로 결정했다. 검색량과 개최 시기가 축제의 성과에 큰 영향을 미치지 않는 것은 예상과는 다른 의외의 결과였다.

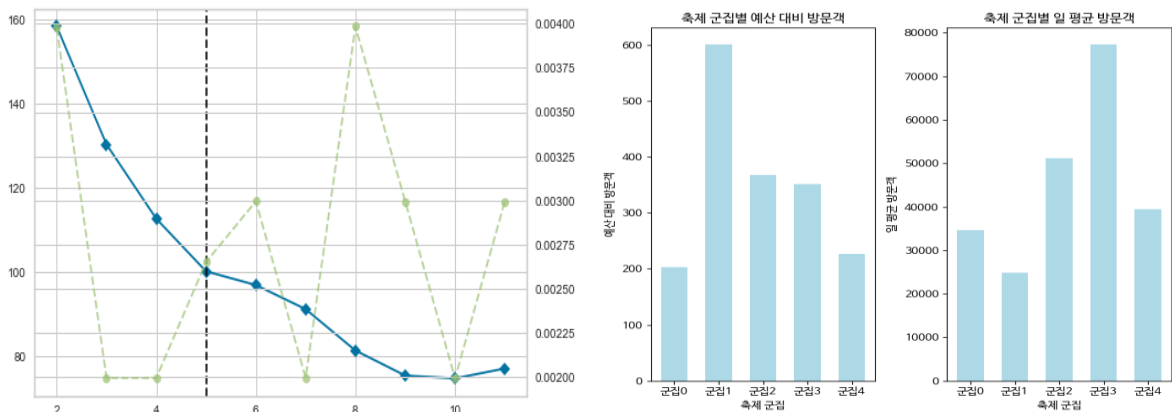
결론적으로, 최종 선정된 변수는 '20 대지역관심도', '특산물', '50 대지역관심도', '수도권', '30 대지역관심도', '40 대지역관심도', '60 대지역관심도', '문화예술', '주변관광명소개수', '체험', '전통', '자연', '역사', '구'로 총 14 개이다. 선택된 14 개의 변수는 이어지는 전국 지역축제에 대한 클러스터링, 즉 군집분석에 활용된다.

4.4) 전국 지역축제 군집분석

병합 군집, KMeans, KMedoids 중 elbow point 가 명확한 KMedoids 알고리즘을 이용하여 군집분석을 진행하였다. 이 때, n=5 가 최적의 군집 개수로 나타났다.

```
1 from sklearn_extra.cluster import KMedoids
2 from yellowbrick.cluster import KElbowVisualizer
3
4 pam = KMedoids()
5 visualizer = KElbowVisualizer(pam, k=(2, 12), random_state=2023)
6 visualizer.fit(dsc)

1 from sklearn_extra.cluster import KMedoids
2 from scipy.stats import mode
3
4 pam = KMedoids(n_clusters=5, random_state=42)
5 pam.fit(dsc)
6 kmed_clst = pam.labels_
```



예산 대비 방문객과 일 평균 방문객을 기준으로, 군집 1,2,3 을 성공적인 지역축제로, 군집 0 과 4 를 성과가 저조했던 지역축제로 구분하였다. 이어서 군집별 특징을 자세히 살펴보자.

4.4.1) 지역축제 군집별 특징 – 성공 지역축제

<군집 1>

축제 유형		광역단체명		군집 1은 일평균 방문객 수는 가장 낮지만 예산 대비 방문객이 1위인 군집 으로, 축제 개최지역이 강원, 부산, 제주 인 점이 군집 1의 대표적인 특징이다. 이 지역들의 공통점은 여행지로서 지역 관심도가 전 연령대에서 높다. 즉, 이미 <u>관광지로서 유명하고 선호도가 높은</u> 점이 <u>지역축제에 긍정적인 영향을 미친다고</u> 판단된다. 축제 유형은 전통, 체험, 문화예술 등 다양한 유형이 혼재하는 것으로 보아, 축제유형에 따른 유의미한 특징은 보이지 않는다.
전통	7	강원	22	
체험	6	부산	5	
문화예술	4	제주	2	
특산물	4			
역사	3			
공연	3			
자연	2			

<군집 2>

광역단체명		축제 유형		군집 2는 예산대비 방문객, 일 평균 방문객 모두 2위인 클러스터로, 개최지 중 수도권에 없는 점이 특징이다. 또 축제유형의 대부분이 특산물 인 것을 볼 수 있다. 해당 클러스터에 속하는 축제들은 '순천장류축제', '봉화은어축제'와 같이 지역 특산물이 축제명에 포함되어, <u>축제 이름만으로도 지역 특산물이 주가 되는 축제임을 직관적으로 유추가 가능한 경우가 많았다.</u>
경기	5	특산물	17	
전북	5	역사	1	
경남	3			
경북	2			
부산	1			
인천	1			
충남	1			

<군집 3>

광역단체명		축제 유형		군집 3은 예산대비 방문객은 3위, 일평균 방문객은 1위인 클러스터이다. 축제 지역이 대구, 부산, 인천 등 주로 광역시인 점 과 문화예술 (공연, 전시, 예술 등 융복합적 테마) 주제의 축제가 대부분인 것이 특징이다.
대구	10			
부산	6			
인천	2			
대전	2			
경남	2	문화예술	23	
광주	1	자연	1	

4.4.2) 지역축제 군집별 특징 - 실패 지역축제

<군집 0>

광역단체명		축제 유형		군집 0은 예산 대비 방문객 5위 , 일 평균 방문객 4위인 클러스터이다. 군집 0에 속하는 지역축제는 수도권 지역인 경기, 서울에서 개최한 경우가 많은 점 이 특징적이다. 축제 유형에 따른 유의미한 차이는 보이지 않았다.
경기	14	문화예술	14	
서울	7	역사	5	
전북	2	공연	3	
경남	2	체험	3	
전남	1	자연	1	

<군집 4>

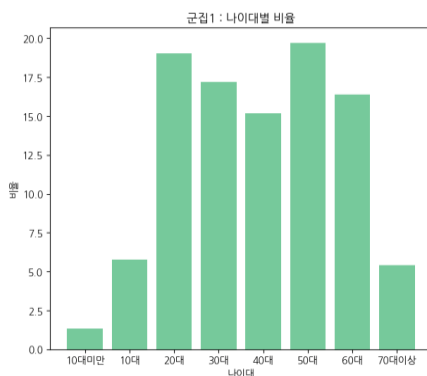
광역단체명		축제 유형	
경남	10	전통	22
경기	5	체험	2
서울	2	역사	1
인천	2	자연	1
광주	2		
부산	1		

군집 4 는 예산대비 방문객 4 위, 일평균 방문객 3 위인 클러스터이다. 군집 4 의 대표적인 특징은 **전통을 주제로한 축제**가 많다는 점이다.

4.5) 성공 축제 군집별 주요 방문객 프로파일링

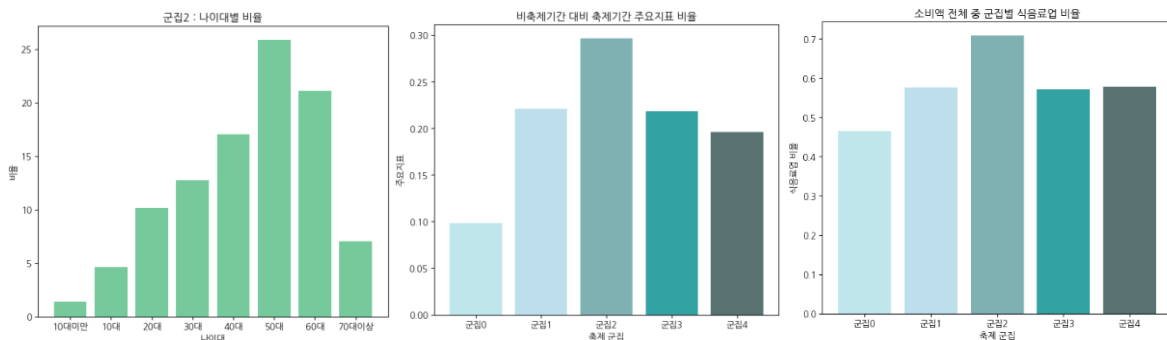
앞서 분석한 각 성공 축제 군집 중 각 군집을 대표할 수 있는 일부 축제들을 바탕으로, 군집별 주요 방문객 프로파일링을 진행하였다. 연령별 방문객, 업종별 소비 등의 데이터를 제공하는 '한국관광 데이터랩'의 문화관광축제분석 데이터를 활용하였다.

4.5.1) 군집 1 : <쇠뿔도 "간" 김에>



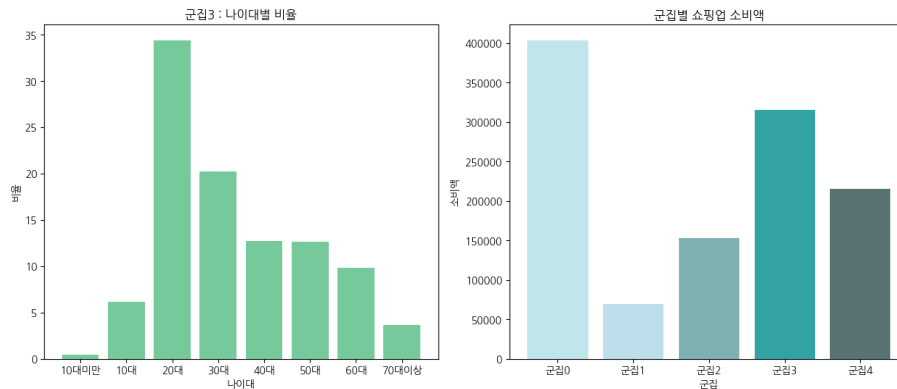
군집 1 의 주요 방문객은 <쇠뿔도 '간' 김에>이다. 이들은 지역축제가 주된 방문 목적은 아니지만, **관광**을 위해 강원/부산/제주 지역을 방문한다. 이 때, 취향에 맞는 축제가 있다면 **놀러간 김에 축제**도 함께 즐긴다. 수도권 외 지역에서 여행을 즐기며, 일정에 없던 다양한 축제들이 선호도에 맞는 경우 폭넓게 방문할 수 있는 **다양한 연령층의 여행객**을 주요 방문객으로 정의할 수 있다.

4.5.2) 군집 2 : <금강산도 식후경>



군집 2 의 주요 방문객은 <금강산도 식후경>이다. 비축제기간 대비 축제기간의 주요 지표가 높은 것으로 보아, 인지도가 높은 축제를 즐기는 50 대와 60 대가 군집 2 의 주요 방문객이다. 군집 2 는 대부분이 특산물을 주제로 한 축제이다. 축제에서 판매하는 식음료, 즉 **특산물에 많은 지출**을 할 수 있는 **소비력 있는 중장년층**의 특징을 확인할 수 있다.

4.5.3) 군집3 : <맛 따라 멋 따라 2030>



군집3의 주요 방문객은 <맛 따라 멋 따라 2030>이다. 이들은 **쇼핑**에 비교적 많은 소비를 하는 **20대 청년층**이 주요 방문객이다. 군집3이 대체로 광역시에서 개최되는 문화예술 주제의 축제임을 감안하면, 2030 세대의 특성이 잘 드러난 결과임을 알 수 있다.

4.6) 실패 축제 군집별 개선 방안

지금까지의 분석 결과를 바탕으로, 우리가 제안하는 성과가 저조했던 축제에 대한 군집별 개선 방안은 아래와 같다.

4.6.1) 군집 0 : <축제의 매력을 끌어~올려~>

목적지 검색순위 ① ☐ 현지인 ☒ 외국인 ☐ 전체 2022 ...

순위	이름	주소	구분
1	시흥갯골생태공원	경기 시흥시 동서로 2...	도시공원
2	솔트베이GC	경기 시흥시 마유로 9...	육상레저스...
3	시흥갯골축제	경기 시흥시 동서로 2...	기타문화관...
4	매곡공원	경기 시흥시	도시공원
5	시흥갯골생태공원둘레전망대	경기 시흥시 설말길 9...	랜드마크관...

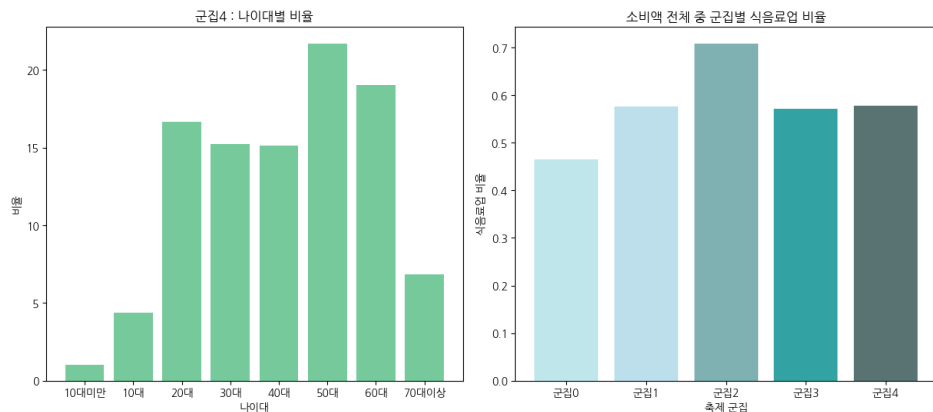
(출처) 한국관광데이터랩 - '시흥갯골축제' 내비게이션 목적지 검색 순위 데이터

군집 0 은 축제가 아닌 다른 장소를 목적으로 해당 지역을 방문한 경우가 많다. 이런 점에서, 군집 1 과 공통점이 있다.

하지만 군집 0 과 군집 1 의 지역축제를 비교했을 때, 군집 1 의 축제가 더 우수한 성과를 냈다. 타지로 여행을 떠난 이들이 놀러간 김에 근처의 지역축제도 함께 방문하는 경우가 많은 반면에, 군집 0 의 개최지역인 수도권에서는 방문 목적을 마치고 근처의 축제를 구경하지 않고 돌아가버리는 사례가 많을 것으로 예상된다.

이를 바탕으로 제안하는 군집 0 지역축제의 개선 방안은 다음과 같다. 첫째, **축제의 정체성을 명확하게 정립**해야 한다. 남녀노소 전부를 타겟팅하기보다, 특정 연령/성별 집단을 타겟팅하여 축제의 특색을 강화하고 인지도를 올려야 한다. 둘째, 백화점 혹은 캠핑 등 **다른 목적을 가진 주변 방문객이 해당 축제에 유입되도록 홍보** 방안을 마련할 필요가 있다. 특히, 군집 0 은 대부분 수도권에서 개최되는 축제이므로, 젊은 세대를 위한 타겟팅이 필요하다. 셋째, **예산의 효율적인 활용**이 필요하다. 수익이 큰 식음료 관련 예산 배치를 늘리는 등의 노력이 필요하다.

4.6.2) 군집 4 : <볼 거리, 즐길 거리, 살 거리!>



군집 4는 50대, 60대 관광객이 비교적 많으며, 소비력이 높은 식음료 부문에서의 소비 또한 높은 편이다. 군집 2와 비교해보면, 방문자의 다수가 5060 중장년층이고 이들의 식음료업에서의 소비 또한 높다는 공통점이 있다.

하지만 군집 2 축제의 주요 주제는 특산물이고 군집 4 축제의 주요 주제는 전통이다. 군집 2와의 비교를 바탕으로 군집 4의 축제들은 지역 특산물과 연계하여 축제 활성화 방안을 도모할 필요가 있다. 전통과 더불어 매력적인 지역 특산물을 통해 축제를 홍보한다면, 주요 방문객인 5060 세대의 마음을 더욱 사로잡을 수 있을 것이다. 축제 이름에서 전통과 특산물을 함께 강조하여 관심을 모으고 지역축제를 홍보하는 등의 노력이 필요하다.

5) 시사점 및 기대효과

본 분석은 데이터에 근거하여 지역축제의 성패에 기여하는 주요 요인을 제시하고, 이를 바탕으로 전국의 지역축제를 군집화하였다. 군집별 프로파일링을 통해 성공 축제 군집에 대한 인사이트를 찾고, 성과가 부진했던 지역축제 군집에 대한 개선 방향을 제시하였다.

이를 통해 도출한 주요 인사이트 및 시사점은 다음과 같다.

- ✓ 지역축제의 성공과 관련된 주요 요인으로는 **지역 관심도, 축제 유형, 비수도권 여부**가 있다. 특히, 축제 개최지역에 대한 20대 또는 50대의 관심도가 높거나, 축제의 주제가 특산물인 경우는 성공에 미치는 영향이 비교적 크게 나타났다.
- ✓ 모든 연령대에서 여행지역 관심도가 높은 **강원, 부산, 제주**에서 개최되는 축제는 **예산 대비 방문객이 많았다**. 반면, 수도권에서 개최되는 축제는 대체로 **예산 대비 방문객이 저조했다**.
- ✓ 지역 **특산물**을 주제로 한 지역축제는 대체로 **높은 성과**를 보였다. 반면, **전통**을 주제로 한 지역축제는 비교적 **성과가 저조했다**.
- ✓ 지역축제의 **정체성이 명확하고 주 방문객 층이 특정될수록**, 축제가 **성공한** 사례가 많은 것으로 보인다.

이를 통해 도출한 인사이트는 국내 관광 산업의 회복을 위한 효과적이고 현실적인 개선 방안 수립에 활용될 수 있다. 또한, 지역축제를 주관하는 지자체 및 정부는 지역축제 개선 및 발전을 위한 가이드라인으로써 본 분석 결과를 참고할 수 있다.

