



도농상생을 위해 XAI를 활용한

# 금산군 축제 성매 예측 모델링 및 제안

금산군 데이터 기반 행정 시각화 아이디어 공모전

국준호, 권동구, 유소현



# 목차

## 주제 선정 이유

배경 및 필요성

제안 목표

## 데이터 분석

크롤링 및 감성분석

데이터 수집

탐색적 데이터 분석(EDA)

데이터 전처리

모델링

## 개선안 제안

분석 결과 해석 및 제안

시각화 인사이트

대시보드 시연



## 배경 및 필요성

### 금산군이 직면한 문제 상황



#### 인구 감소와 농촌 빈곤

- 이 상황에서 '도농상생' 개념은 중요성을 갖고 있으며, 지역 축제는 이를 실현하기 위한 중요한 전략으로 간주된다.
- 즉, 축제는 단순한 문화 행사를 넘어 지역 경제 활성화와 도시민의 농촌 방문을 촉진하는 중요한 수단으로 인식되고 있다.

### 문제 해결의 필요성



#### 다양한 요인에 의해 성패가 결정되는 지역 축제

- 지역 축제는 지역적 특성, 프로그램, 방문객과의 상호작용 등이 중요한 요소로 작용한다.
- 성공적인 축제는 지역 경제적 이익과 문화적 활성화를 촉진하며, 지속 가능한 관광 산업을 유도한다.
- 결과적으로 축제 운영은 치열한 경쟁 속에서 진행되며, 효과적인 데이터 기반 분석과 전략적인 관리가 성공을 결정짓는 중요한 요소로 부각된다.

## 제안 목표

# XAI 기술을 활용한 금산군 축제 성패 예측 및 전략적 방안 모색

## XAI(Explainable AI)란?

인공지능(AI) 시스템의 의사결정 과정과 결과를 인간이 이해할 수 있도록 설명하는 기술. 모델의 결과를 해석하는 데 활용되며, SHAP나 LIME 같은 기법이 주로 사용된다.

## XAI를 활용해 무엇을 할 것인가?

XAI를 통해 축제 성패에 영향을 미치는 요소들을 명확하게 분석하여 축제 기획과 운영에 있어 데이터 기반의 전략을 수립하고, 특정 변수들이 축제 성공에 미치는 정량적인 영향을 평가한다. 최종적으로 금산군 지역 축제 발전에 기여할 수 있는 전략적 방안을 모색하고자 한다.

## 크롤링 및 감성분석

### Track 1) 축제 현황 분석을 위한 네이버 보도자료 크롤링 및 제목 감성분석

#### 네이버 뉴스기사 크롤링

크롤링(Crawling)은 웹에서 자동으로 정보를 추출하는 과정으로, 데이터를 직접 수집하기 위해 활용한다. 본 연구에서는 축제 현황 파악을 위한 감성 분석 데이터를 구하고자 HTML파일을 쉽게 파싱하고 추출을 돋는 BeautifulSoup으로 네이버 뉴스기사 크롤링을 진행했다.

#### 검색 키워드 및 크롤링 기간 설정

ex)

검색 키워드 : 금산인삼축제

크롤링 기간 : 2023.07.14 ~ 10.13  
(축제 시작 전 3개월 치)

#### URL 크롤링 함수 정의

The NAVER logo is displayed in its signature green color with a bold, sans-serif font.

검색 결과 페이지 순차적 요청  
네이버 뉴스 URL만 필터링

#### 뉴스 기사 정보 크롤링

The BeautifulSoup logo is shown in a large, stylized black font where the letter 'f' has a musical note-like shape above it.

Beautifulsoap으로 html구조 파싱  
각 기사의 신문사, 제목 본문 크롤링

## 데이터 분석

# 크롤링 및 감성분석

company	url	title	content
0 아시아경제 <a href="https://n.news.naver.com/mnews/article/277/000...">https://n.news.naver.com/mnews/article/277/000...</a>	'바가지 없는 지역축제 조성'...관광공사, 백종원과 지역관광 활성화 나선다	더본코리아와 음식 활용 지역관광 활성화 MOU한국관광공사가 백종원 대표와 손잡고 ...	
1 뉴스1 <a href="https://n.news.naver.com/mnews/article/421/000...">https://n.news.naver.com/mnews/article/421/000...</a>	[오늘의 주요 일정] 대전·충남(25일, 월)	충남도청 전경,/뉴스1(대전 · 충남=뉴스1) = ◇대전시-이장우 시장 10:00 대...	
2 주간조선 <a href="https://n.news.naver.com/mnews/article/053/000...">https://n.news.naver.com/mnews/article/053/000...</a>	백종원, 지역 축제 '바가지 요금' 잡으러 간다	백종원 더본코리아 대표. photo 뉴시스지역 먹거리 축제에서의 '바가지 요금'을...	
3 뉴시스 <a href="https://n.news.naver.com/mnews/article/003/001...">https://n.news.naver.com/mnews/article/003/001...</a>	금산군, 슬로·트렌드 푸드 만들기 등 주민 재능계발 교육	제과제빵, 쌀베이킹 활용 등 8개 교육 추진 금산군 재능계발 전문교육이 주민들로부...	
4 뉴시스 <a href="https://n.news.naver.com/mnews/article/003/001...">https://n.news.naver.com/mnews/article/003/001...</a>	[오늘의 주요일정] 대전·세종·충남 (9월 20일)	[세종=뉴시스] ▲오전 7시 20분, 대전 대덕해윰포럼, 디딤돌플라자▲ 오전 7시 ...	

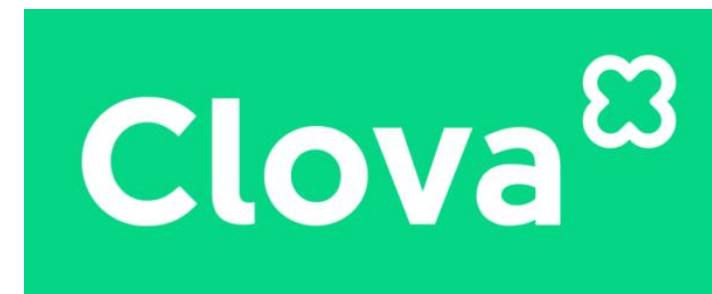
[그림1 크롤링한 네이버 뉴스기사 데이터]

## 크롤링 및 감성분석

### 네이버뉴스 제목 감성분석

감성 분석(Sentiment Analysis)은 텍스트 데이터를 분석하여 그 속에 담긴 감정이나 감정을 파악하는 자연어 처리(NLP)의 한 분야이다. 크롤링한 뉴스 기사 제목은 핵심적인 정보를 담고 있으며, 해당 기사에서 전달하고자 하는 주요 감정을 반영한다. 본 연구에서는 타겟 축제인 '금산인삼축제'와 '금산천벚꽃축제'에 대한 긍정, 부정, 중립적인 감정을 뉴스기사 제목을 바탕으로 분석해, 축제에 대한 전반적인 여론과 인식을 파악했다.

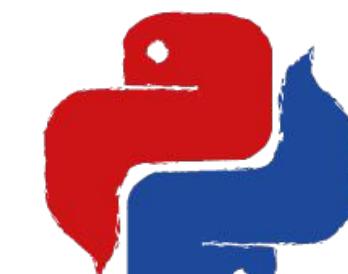
#### 네이버 CLOVA Sentiment API



뉴스 제목에 대해 감성분석 진행

감성 라벨, 신뢰도, 주요 감정 구간 추출

#### 주요 감정 구간 TF-IDF 벡터화



KoNLPy

KoNLPy Okt를 활용한 형태소 분석

제목별 최고 기여 단어 추출

#### 데이터 정리

주요 감정 구간 형태소 단어 카운팅

중립 단어 제외한 긍부정 단어 필터링

## 데이터 분석

# 크롤링 및 감성분석

133	노컷뉴스	<a href="https://n.news.naver.com/mnews/article/079/000...">https://n.news.naver.com/mnews/article/079/000...</a>	충남 금산군, 전 직원 금산세계인 삼축제 티셔츠 입고 홍보	금산세계인삼축제 홍보 포스터. 금산군 제공충남 금산군(군수 박범인)은 전 직원이 ...	neutral	{'negative': 0.03224166, 'positive': 0.0307524...}	[{{'offset': 8, 'length': 23}}]	전 직원 금산세계인삼축제 티셔츠 입고 홍보
134	노컷뉴스	<a href="https://n.news.naver.com/mnews/article/079/000...">https://n.news.naver.com/mnews/article/079/000...</a>	박범인 금산군수, 제41회 금산세계인삼축제 준비 철저 당부	박범인 금산군수. 금산군 제공충남 금산군은 박범인 군수가 14일 군청 상황실에서 ...	neutral	{'negative': 0.08886535, 'positive': 0.0471470...}	[{{'offset': 10, 'length': 22}}]	제41회 금산세계인삼축제 준비 철저 당부
135	뉴시스	<a href="https://n.news.naver.com/mnews/article/003/001...">https://n.news.naver.com/mnews/article/003/001...</a>	금산세계인삼축제 국제인삼교역전, 400만달러 수출협약	총 7건 협약지난해 130만달러 대비 270만달러 증가 국제인삼교역전 *재판매 및...	neutral	{'negative': 0.06320516, 'positive': 0.1220618...}	[{{'offset': 18, 'length': 11}}]	400만달러 수출협약
136	뉴스1	<a href="https://n.news.naver.com/mnews/article/421/000...">https://n.news.naver.com/mnews/article/421/000...</a>	충남도, 지역 대형축제 앞두고 '고위험병원체 선제검사'	탄저균·페스트균·브루셀라균·코로나19 등 충남도청 전경./뉴스1(내포=뉴스1) 이...	neutral	{'negative': 0.5416572, 'positive': 0.11676115...}	[{{'offset': 5, 'length': 24}}]	지역 대형축제 앞두고 '고위험병원체 선제검사'
137	연합뉴스	<a href="https://n.news.naver.com/mnews/article/001/001...">https://n.news.naver.com/mnews/article/001/001...</a>	한국타이어, 2023 금산세계인삼축제 후원금 기탁	2023 금산세계인삼축제에 후원금 기탁하는 한국타이어박범인 금산군수(좌)와 황성하	neutral	{'negative': 0.073257424, 'positive': 3.414640...}	[{{'offset': 7, 'length': 20}}]	2023 금산세계인삼축제 후원금 기탁

[그림2 감성분석 결과]

top_word	count	label	negative	positive	neutral
백종원	27	positive	0.02	85.5	14.48
활성화	6	positive	0.02	99.61	0.37
요금	3	negative	85.71	0.02	14.27
만끽	3	positive	0.02	99.95	0.03
전달	3	positive	0.02	98.49	1.49
문화	3	positive	0.04	85.45	14.51
기원	3	positive	0.03	95.7	4.27
구경	2	positive	0.03	99.88	0.1
경제	2	negative	99.97	0.01	0.02
인기	2	positive	0.02	99.82	0.16

[그림3 형태소 분석 및 데이터 필터링]

## 데이터 분석

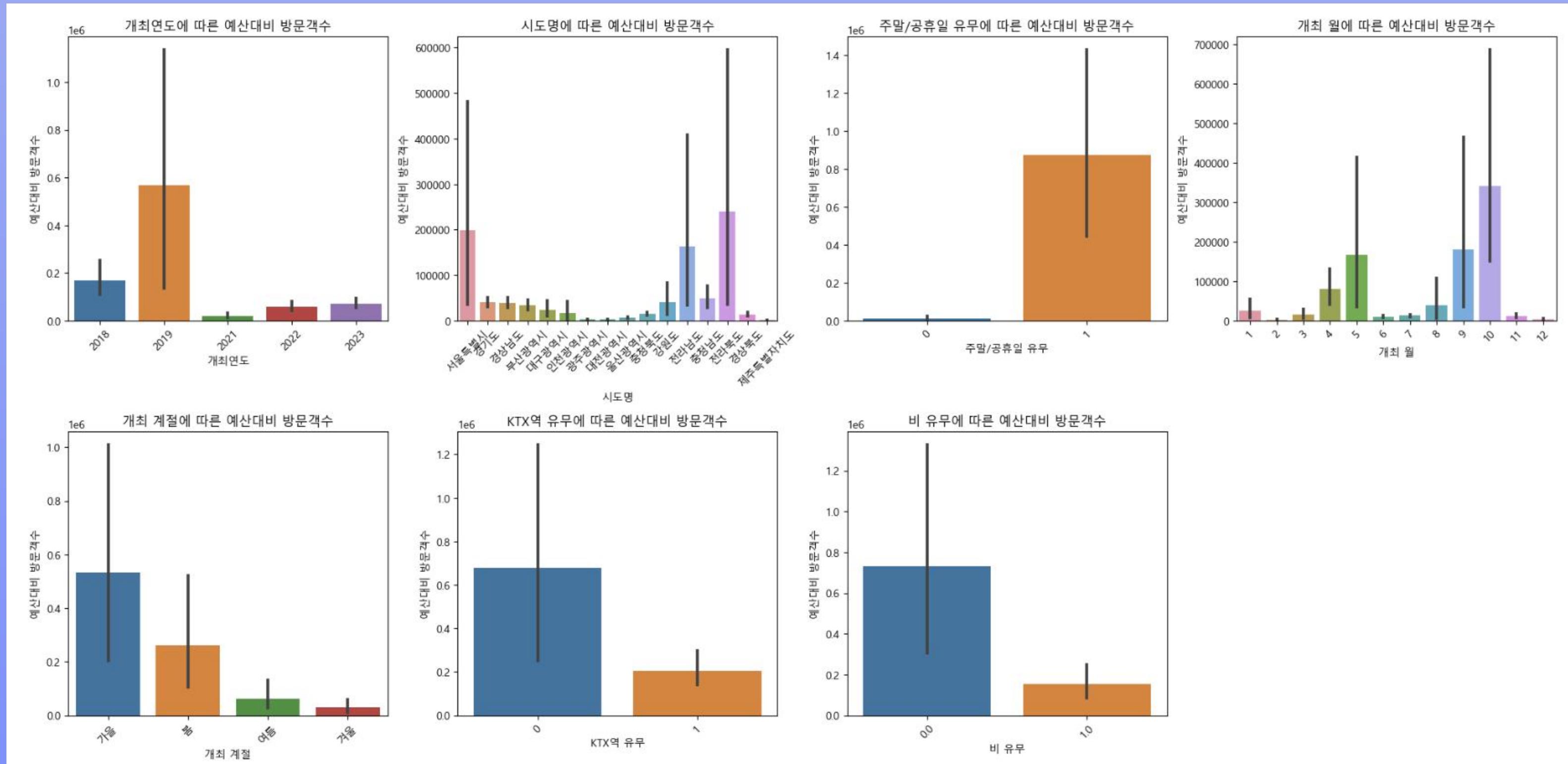
# 데이터 수집 (2018년~2023년 데이터, COVID-19 2020년 제외)

## Track 2) XAI를 활용한 축제 성패 예측 모델링

1. 지역 축제 정보	3. 전국 문화축제 표준데이터	4. 행정동별 인구수	5. 기상청 종관기상관측 데이터	6. 한국철도공사 역 위치 데이터
<p> 문화체육관광부</p> <p>Data Cleansing</p> <p>모델링에 사용할 종속 변수인 방문객 수, 예산 데이터 단위 통일</p> <p>개최 기간</p> <p>최초 개최년도 대비 축제의 개최 기간 파생변수를 생성해, 축제의 역사, 인지도, 성장 추세를 확인</p> <p>동일 기간 축제 수</p> <p>동일 기간 축제 수를 통해 방문객 수 분산 정도, 타 축제와의 경쟁 강도가 시장 성패에 미치는 영향을 파악</p>	<p> 문화 빅데이터 플랫폼</p> <p>축제 일수</p> <p>축제 지속 기간의 영향 파악</p> <p>개최 월, 개최 계절</p> <p>시계열 패턴, 계절적 요인에 따른 영향을 파악하기 위한 파생변수 생성</p> <p>서울과의 거리, 광역시와의 거리</p> <p>i) 하버사인 거리</p> <p><math>d = 2r \arcsin \left( \sqrt{\text{hav}(\varphi_2 - \varphi_1) + \cos \varphi_1 \cos \varphi_2 \cdot \text{hav}(\lambda_2 - \lambda_1)} \right)</math></p> <p><math>= 2r \arcsin \left( \sqrt{\sin^2 \left( \frac{\varphi_2 - \varphi_1}{2} \right) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right)</math></p> <p>cities = {     "seoul_station": (37.554623, 126.970609),     "Busan": (35.1796, 129.0756),     "Daegu": (35.8714, 128.6014),     "Incheon": (37.4563, 126.7052),     "Gwangju": (35.1595, 126.8526),     "Daejeon": (36.3504, 127.3845),     "Ulsan": (35.5384, 129.3114)}</p> <p>위도와 경도를 사용하여 거리를 계산하는 하버사인 거리를 활용해 축제 접근성의 영향을 파악</p> <p>주말/공휴일 유무</p> <p>주말 및 공휴일 방문객 참여도를 파악</p>	<p> 행정안전부</p> <p>연령대별 인구수</p> <p>10대 미만 인구수 10대 인구수 20대 인구수 ⋮ 60대 이상 인구수</p> <p>축제 별 타겟 관객 분석 연령대별 인구수가 축제에 미치는 경제적 효과와 소비패턴을 파악</p> <p>평균 나이</p> <p>방문객의 특성 파악</p> <p>고령 지수</p> <p>고령 지수 = 60대 인구수 / 총 인구수 행정동별 인구 고령화 정도를 파악하여 축제 방문객의 특성 파악</p>	<p> 기상청</p> <p>비 유무</p> <p>축제 기간 동안 비 유무 파악</p> <p>평균 기온</p> <p>축제 기간 동안 평균 기온 파악</p> <p>평균 풍속</p> <p>축제 기간 동안 평균 풍속 파악</p> <p>평균 상대습도</p> <p>축제 기간 동안 상대습도 파악</p> <p>↓</p> <p>날씨가 축제 성패에 미치는 영향을 평가</p> <p>기상 관측 데이터의 경우 기상청에서 관측을 하는 관측 지점이 있어, 축제 지역과 하버사인 거리가 가장 가까운 관측 지점의 데이터를 맵핑</p>	<p> KORAIL 한국철도</p> <p>KTX 역 유무</p> <p>축제 위치 5km 이내의 KTX 역 유무 파생변수 생성</p> <p>접근성이 축제 성패에 미치는 영향 확인</p>
<p>7. 소득 구간별 주민 비율</p> <p> 문화 빅데이터 플랫폼</p> <p>소득 구간별 주민 비율</p> <p>소득 2천만원 주민 비율 소득 3천만원 주민 비율 소득 4천만원 주민 비율 ⋮ 소득 7천만원 이상 주민 비율</p> <p>지역 주민의 소비력을 파악하여, 소비력이 축제 성패에 주는 영향을 확인</p>				

## 데이터 분석

# 탐색적 데이터 분석(EDA)



[그림4 막대그래프 분석]

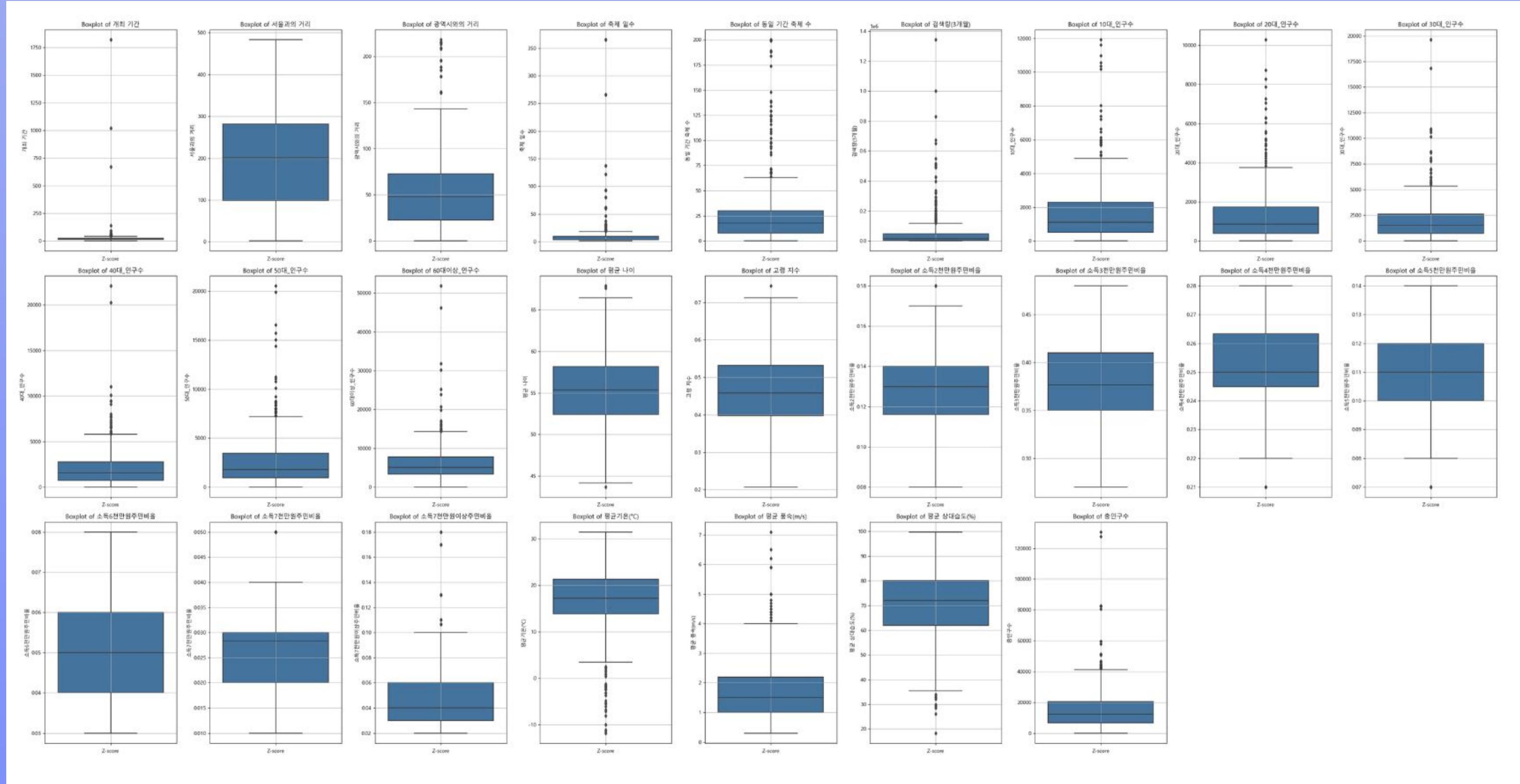
## 막대그래프 분석

범주형 변수들에 대한 예산 대비 방문객 수를  
막대그래프로 시각화

## 막대그래프 분석 결과

시각화 결과를 토대로 각 변수들이 예산 대비 방문  
객 수에 미치는 영향을 확인  
이를 바탕으로 예산 대비 방문객 수 예측 모델링에  
해당 변수들을 모두 고려하기로 결정

# 데이터 전처리 - 이상치 제거



[그림5 박스 플롯 분석]

## 박스 플롯(Box-plot) 분석

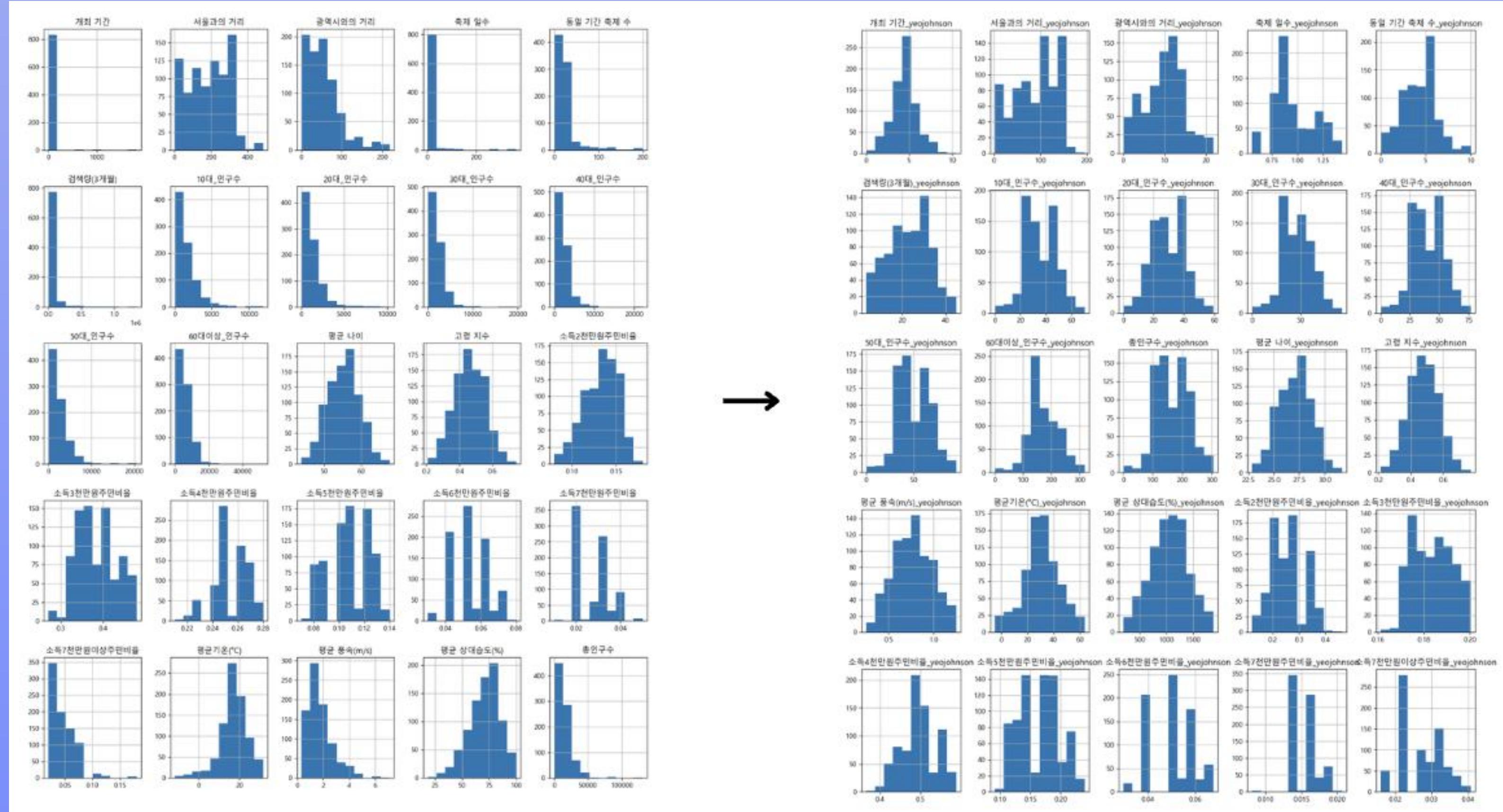
범주형 변수들에 대한 예산 대비 방문객 수를  
수치형 변수들에 대한 박스 플롯을 통해 데이터의 분  
포와 이상치를 시각적으로 확인

## Z 점수 기반 이상치 처리

데이터에서 Z 점수의 절댓값이 3.5 이상인 이상치를  
제거한 후, 초기 851개의 데이터 중 764개를 최종적  
으로 선택하여 분석에 활용

## 데이터 분석

# 데이터 전처리 - 변환



[그림6 Yeo-Johnson 변환]

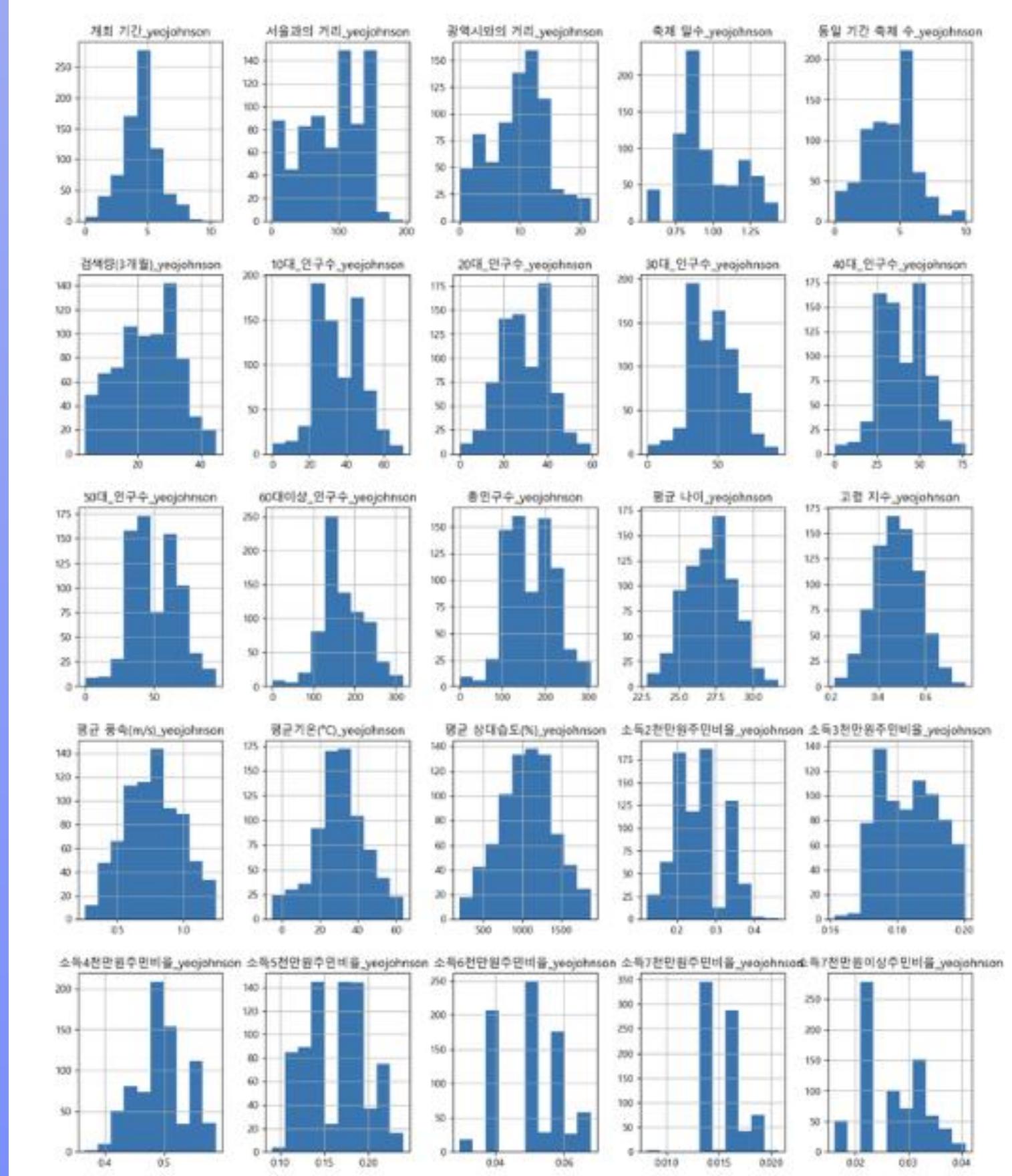
## Yeo-Johnson 변환 적용

데이터의 정규성을 높이기 위해 수치형 변수들에  
Yeo-Johnson 변환 적용

Yeo-Johnson 변환은 제곱근, 로그 변환, Box-Cox  
변환 등 다양한 left-skewed 데이터에 적합한 변환  
방법 중 가장 좋은 분포를 보였으며, 다른 변환과 달리  
음수 값을 포함한 데이터에도 적용 가능

# 데이터 분석

# 데이터 전처리 - 변환



변수	$\lambda$
개최 기간	0.27227138802165446
서울과의 거리	0.832827273508824
광역시와의 거리	0.4507695696107665
중략	
평균기온( $^{\circ}\text{C}$ )	1.2760701163288268
평균 상대습도(%)	1.7567306676253238

# Yeo-Johnson 변환 후 $\lambda$ 값

각 변수에 대해 최적의  $\lambda$  값을 계산한 결과는 왼쪽 표와 같으며, 이는 Yeo-Johnson 변환을 통해 얻어진 값임.

해당 값을 이용하여 역변환을 통해 실제 값 계산 가능

## 모델링 - 모델 비교

선형 회귀 모델	비선형 회귀 모델
Ridge Regression	Support Vector Regression
Lasso Regression	Gradient Boosting
ElasticNet Regression	RandomForest
	XGBoost
	LightGBM
	CatBoost

[그림8 모델 비교]

### 선형 회귀 모델 비교

Ridge, Lasso, ElasticNet Regression 선형 회귀  
모델들은 비선형적인 데이터 관계를 잘 반영하지 못해  
 $R^2$  값이 낮아 분석에서 제외

### 비선형 회귀 모델 비교

Support Vector Regression, Gradient Boosting,  
RandomForest, XGBoost, LightGBM, CatBoost  
중 좋은 성능을 보인 RandomForest, XGBoost,  
CatBoost를 최종 후보 모델로 선정

## 모델링 - 최적화

최적화 방법
Grid Search
Bayesian Optimization
AutoML - AutoGluon
AutoML - PyCaret

[그림9 모델 최적화 기법 비교]

### 하이퍼파라미터 비교

RandomForest, XGBoost, CatBoost 모델에 대하여 최적화 방법을 비교한 결과 Bayesian 최적화를 사용하기로 결정  
AutoML 기법의 경우 실행에 많은 시간이 소요되며 최적 모델이 Ensemble인 경우 SHAP Value와 같은 XAI 기법을 적용하는 것이 어려움

### Bayesian 최적화

알려지지 않은 목적 함수를 최대(혹은 최소)로 하는 최적해를 찾는 기법

- Surrogate Model: 현재까지 조사된 입력값과 목적 함수 값 쌍들을 기반으로 목적 함수의 확률적 추정을 수행
- Acquisition 함수: Surrogate Model을 활용하여 다음 입력값 후보를 추천하는 함수

## 데이터 분석

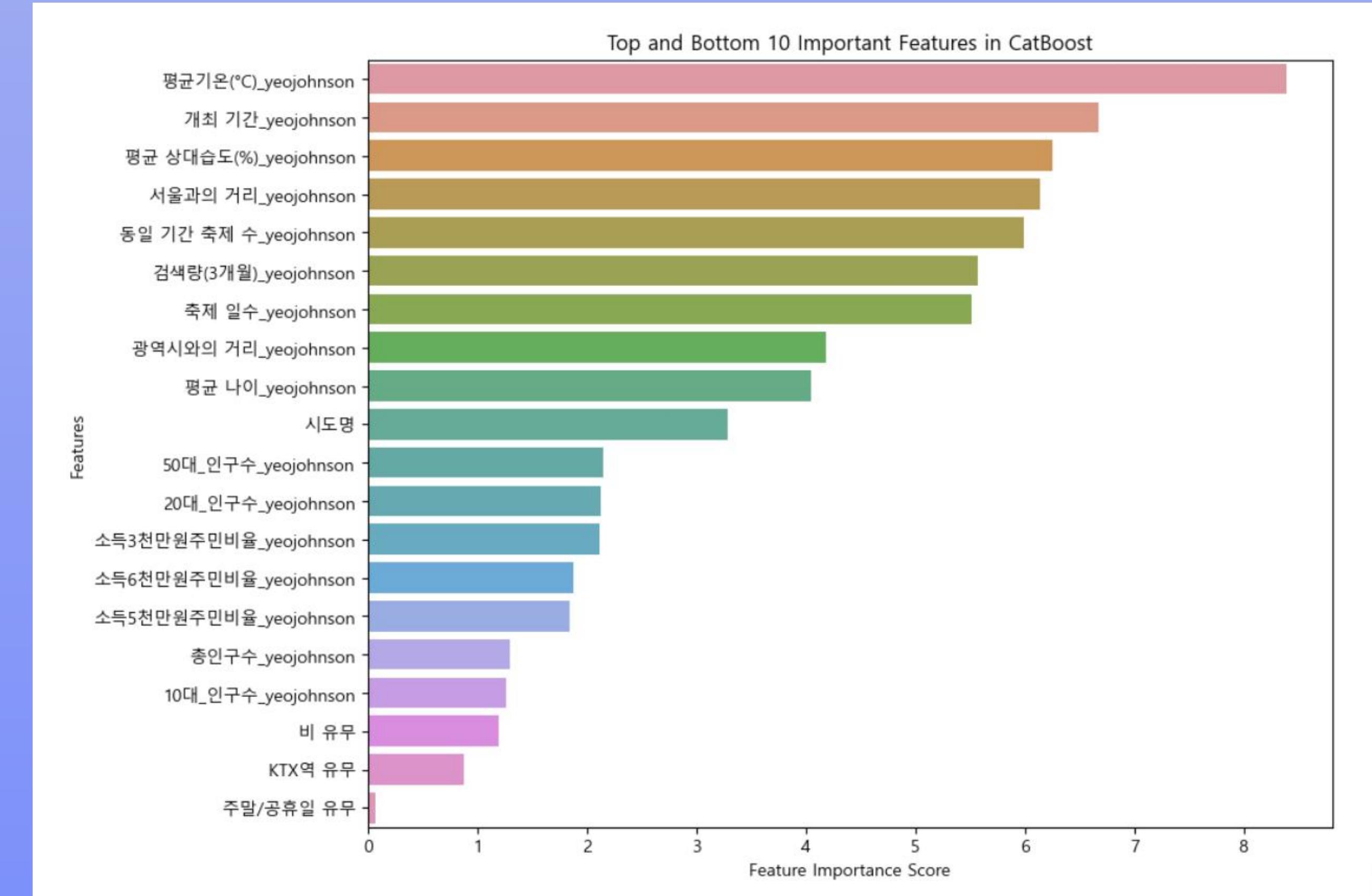
# 모델링 - 최적화

최종 모델	파라미터	RMSE	MAE
CatBoost	<ul style="list-style-type: none"><li>• depth=8</li><li>• iterations=300</li><li>• l2_leaf_reg=1</li><li>• learning_rate=0.</li></ul> 0537061911601 7978	1.2774	0.9239

[그림10 최종 모델 파라미터 및 성능]

## Feature Importance

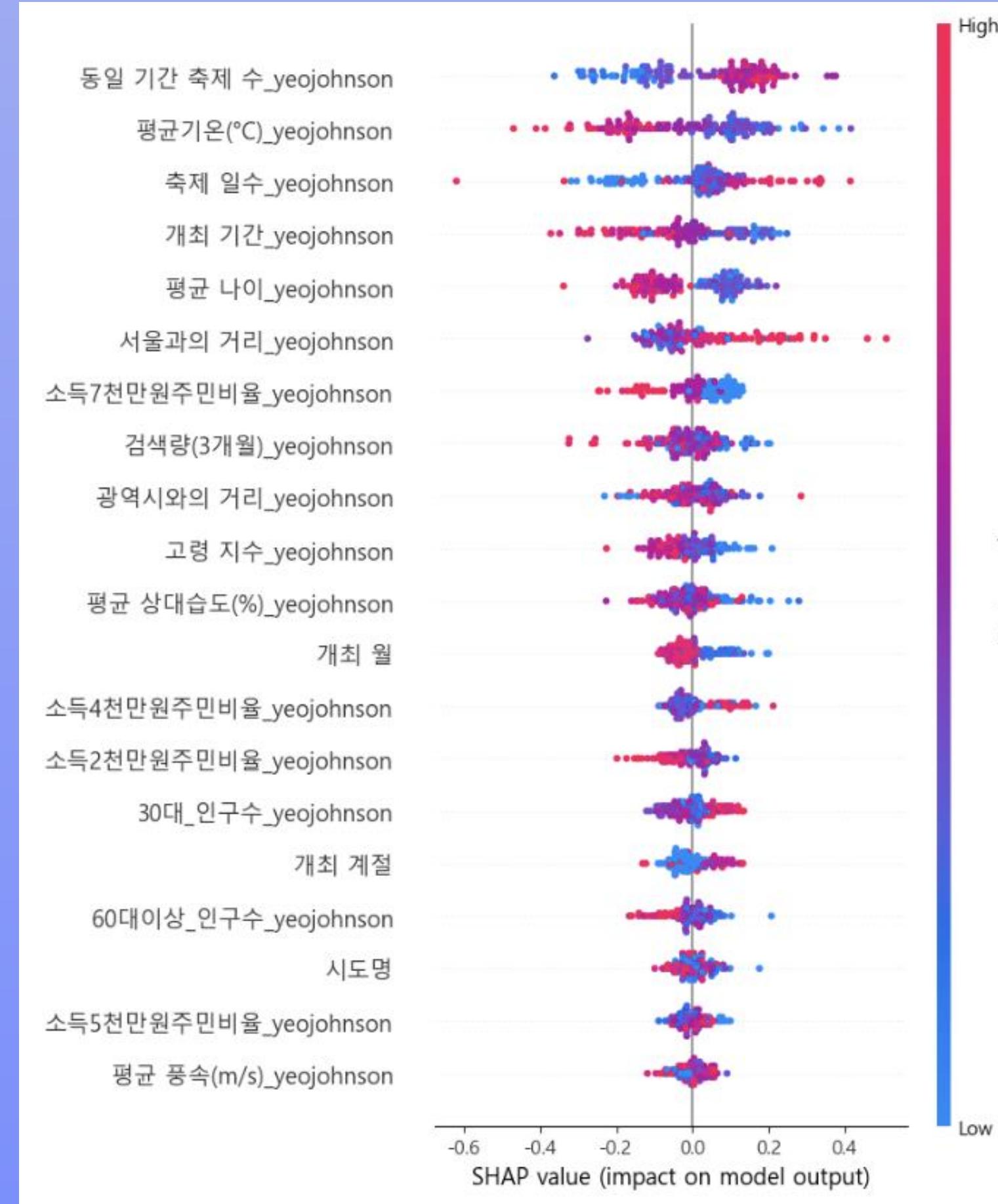
변수 중요도 시각화 결과, 평균기온(°C), 평균 상대습도(%) 등의 날씨 변수와 개최 기간, 동일 기간 축제 수, 축제 일수 등의 축제 관련 변수가 예측에 중요한 영향을 미침을 확인할 수 있음



[그림11 Feature Importance]

## 데이터 분석

# 모델링 - SHAP



[그림12 SHAP Value 시각화]

## SHAP

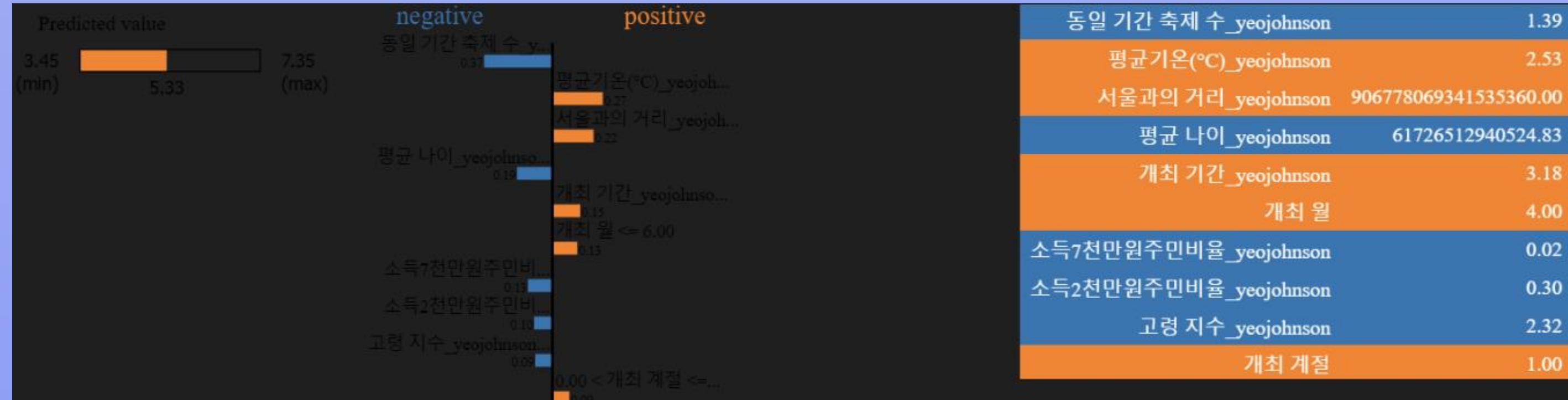
- SHAP은 각 변수가 예측값에 어떻게 영향을 미치는지를 설명하는 수치로, 각 열은 하나의 특성을 나타내며, 각 행은 하나의 데이터 샘플을 나타냄
- SHAP Value 값이 양수인 경우 해당 특성이 예측값을 증가시키는 방향으로 작용하고, 음수인 경우 예측값을 감소시키는 방향으로 작용
- SHAP 값의 절댓값이 클수록 해당 특성이 예측값에 큰 영향을 미침

시도명	주말/공휴일 유무	개최 월	개최 계절	KTX역 유무	비 유무	개최 기간 yeojohnson	서울과의 거리 yeojohnson	광역시와의 거리 yeojohnson	축제 일수 yeojohnson	동일 기간 축제 수 yeojohnson
0.019747	-0.000899	-0.010847	-0.029212	0.005573	-0.016493	0.034737	-0.032548	0.047090	0.323694	-0.022721
0.049283	-0.000690	0.006812	-0.013686	0.017231	-0.016734	0.002039	-0.048603	-0.041997	-0.077029	0.173518
0.007914	-0.004051	0.125993	0.091300	-0.013711	0.000097	0.107838	0.012053	-0.044840	0.060616	-0.120783
0.003906	-0.000255	-0.028617	-0.002693	-0.023517	-0.001846	0.207845	0.021405	-0.022439	-0.149908	0.187253
0.016937	-0.001249	0.099547	0.009189	-0.013904	0.086714	0.093368	-0.078308	0.071789	0.057170	-0.060733

[그림13 SHAP Value 표]

## 데이터 분석

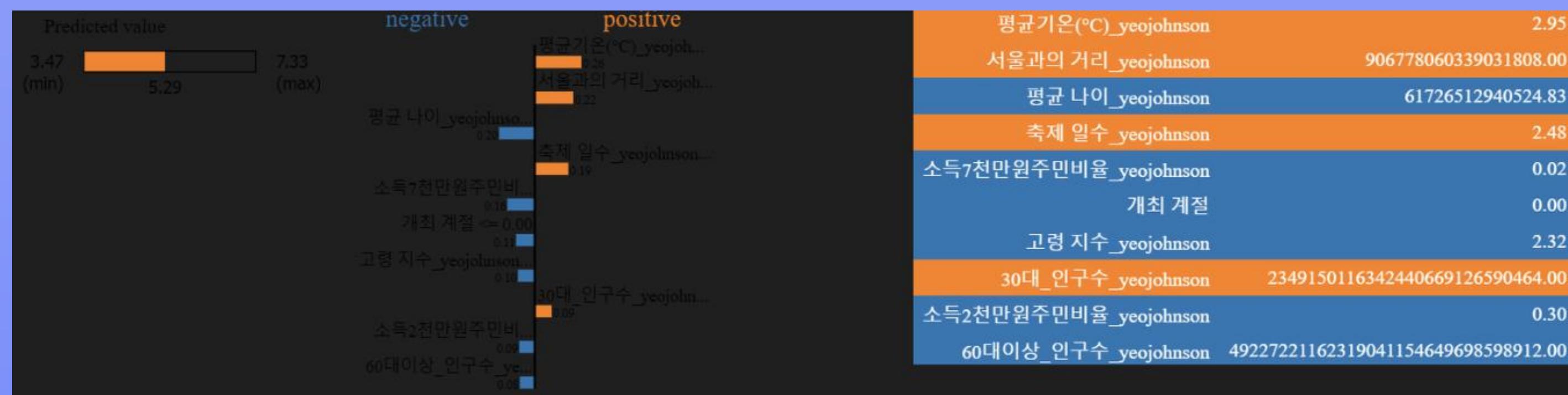
# 모델링 - LIME



[그림14 금산인삼축제 LIME Value 시각화]

## LIME

- LIME은 개별 예측을 해석하는 데 사용되며, 각 변수의 영향력을 설명하는 간단한 수치
- LIME 값이 양수인 경우 해당 특성이 예측값을 양의 방향으로 증가시키고, 음수인 경우 음의 방향으로 감소시킴
- LIME 값이 클수록 해당 특성이 예측값에 큰 영향을 미침
- 왼쪽 LIME 시각화 결과에 따르면, 2023 금산인삼축제의 경우 평균 기온(°C)이 증가할 수록 예측값이 증가하는 방향으로 작용함을 확인할 수 있음



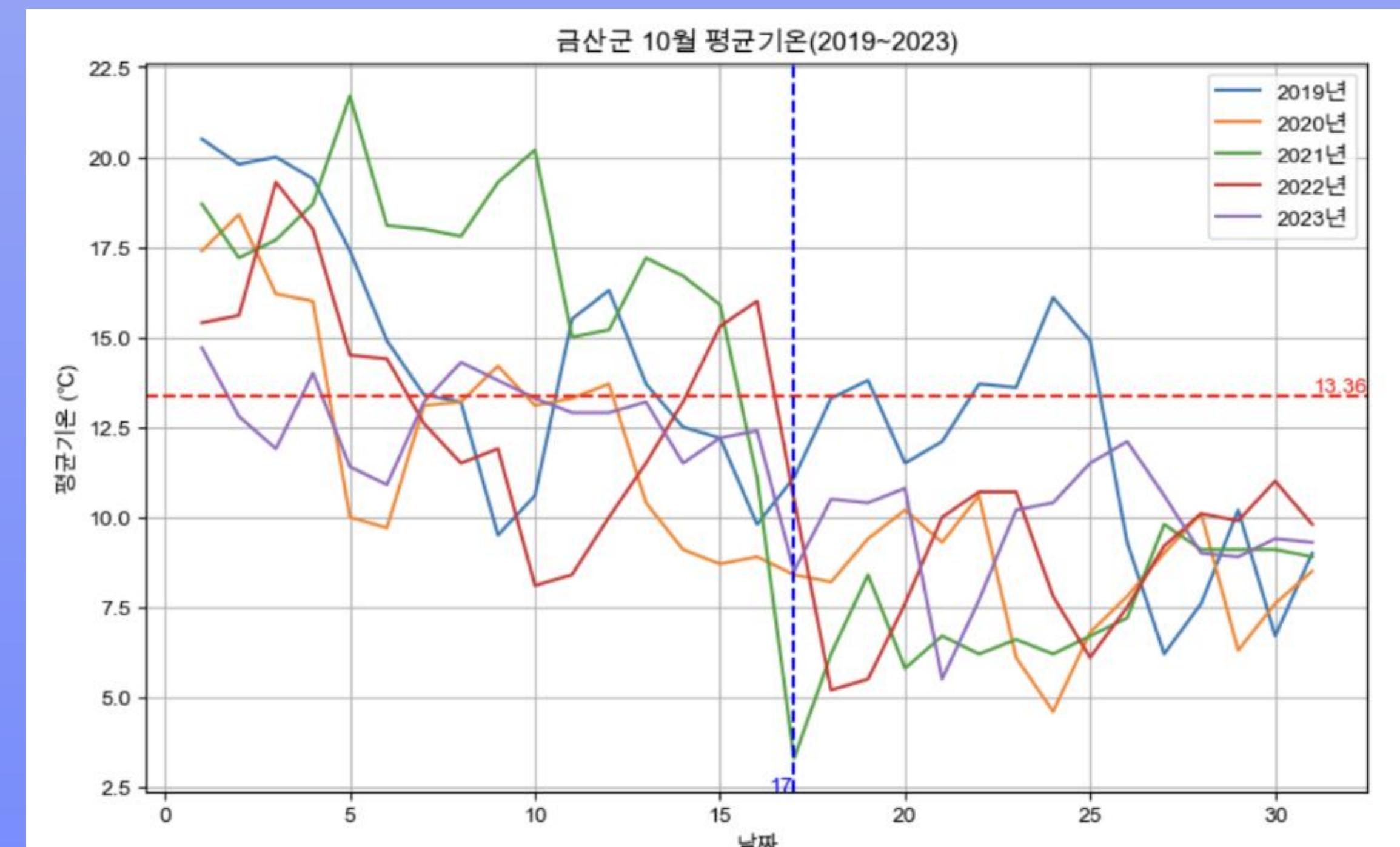
[그림15 금산천벚꽃축제LIME Value]

## 분석 결과 해석 및 제안

(XAI를 통한 주요 변수들이 불가항력 요인일 경우 개선안 제안에서 제외)

### [금산 인삼 축제 개선안 1]

축제 성패 예측 모델링을 통해 평균 기온이 13.36도 이하일 때, 축제의 성공 확률이 증가함을 확인했다. 금산군 인삼 축제의 경우, 10월 초에 11일 가량 진행 된다. 인삼은 9월 중순부터 11월 말이 수확 적기 이기에 평균기온이 13.36도 밑으로 내려오는 17일 이후로 축제 기간을 개편한다면 축제의 성공 확률을 높일 수 있을 것으로 기대된다.



[그림16 금산군 연도별 10월 평균 기온]

## 분석 결과 해석 및 제안

### [금산 인삼 축제 개선안 2]

- 축제 성패 예측 모델링을 통해 30대의 인구수가 많을 수록, 60대 이상의 인구수가 적을 수록, 축제의 성공 확률이 증가함을 확인했다.
- 축제 성패 예측 모델링을 통해 소득 7천만원 이상 주민 비율, 소득 2천만원 이상 주민 비율이 적을수록, 소득 4천만원 이상 주민 비율이 높을 수록 축제의 성공 확률이 증가함을 확인했다. 해당 분석 결과 해석에 따른 제안은 아래와 같다.

#### 3040세대 맞춤형 전략 필요

인삼의 주소비층은 부모님에게 선물을  
준비하는 3040세대, 주소비층인 50대



실질적 소비력을 지닌 3040세대  
맞춤형 마케팅 전략이 중요하다.

X세대의 경우, 중년 마케팅을 지향하고 '**영포티(Young forty): 나이에 비해 젊게 사는 4050세대를 의미한다**'들을 위한 마케팅을 진행한다. 전 세대 중 브랜드 충성도가 가장 높은 세대로 금산군 인삼에 대한 고급화 전략을 통해 로열티 구축에 힘써야 한다.

M세대의 경우, 가치소비를 지향한다. 자신이 지향하는 가치나 만족에 기반하여 제품이 합리적인지를 판단하고자 하기에 타 제품 대비 품질의 우월성, 가격에 합리성을 어필하는 마케팅을 진행한다. M세대의 경우 직접 소비보다 선물의 비중이 높기에 높은 수준의 패키징을 어필하는 것도 좋은 방법 중 하나이다.

# 시각화 인사이트

금산군

도농상생을 위해 XAI를 활용한

# 금산군 축제 성매 예측 모델링 및 제안

금산군 축제 분석

Month	Participation Rate (%)
Jan	??%
Feb	??%
Mar	??%
Apr	??%
May	??%
Jun	??%
Jul	??%

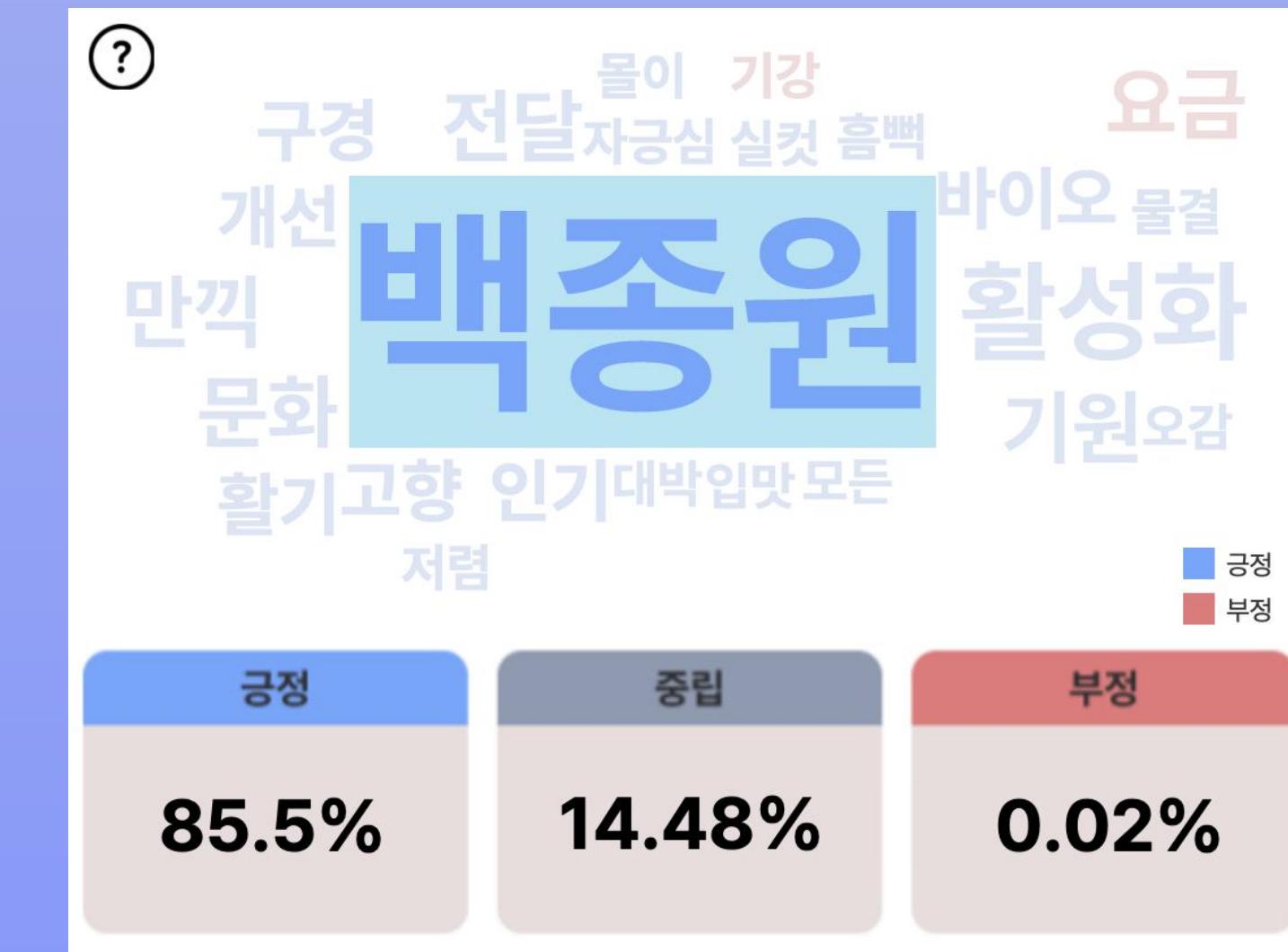
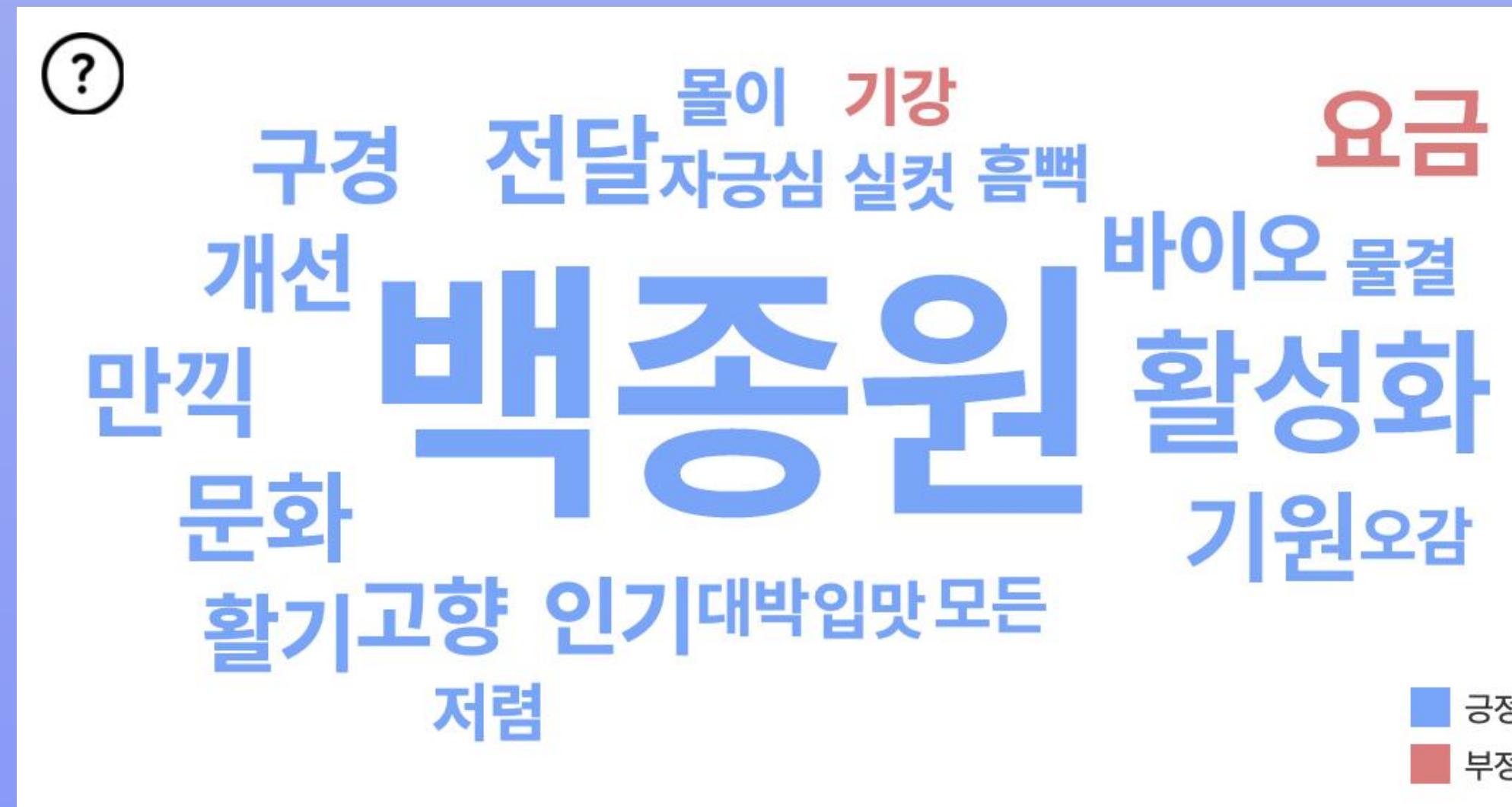
금산군 축제 성매 성과는 다음과 같습니다.

[성매 예측 보러가기](#)

금산군 축제 분석

● 현지인 ● 외지인

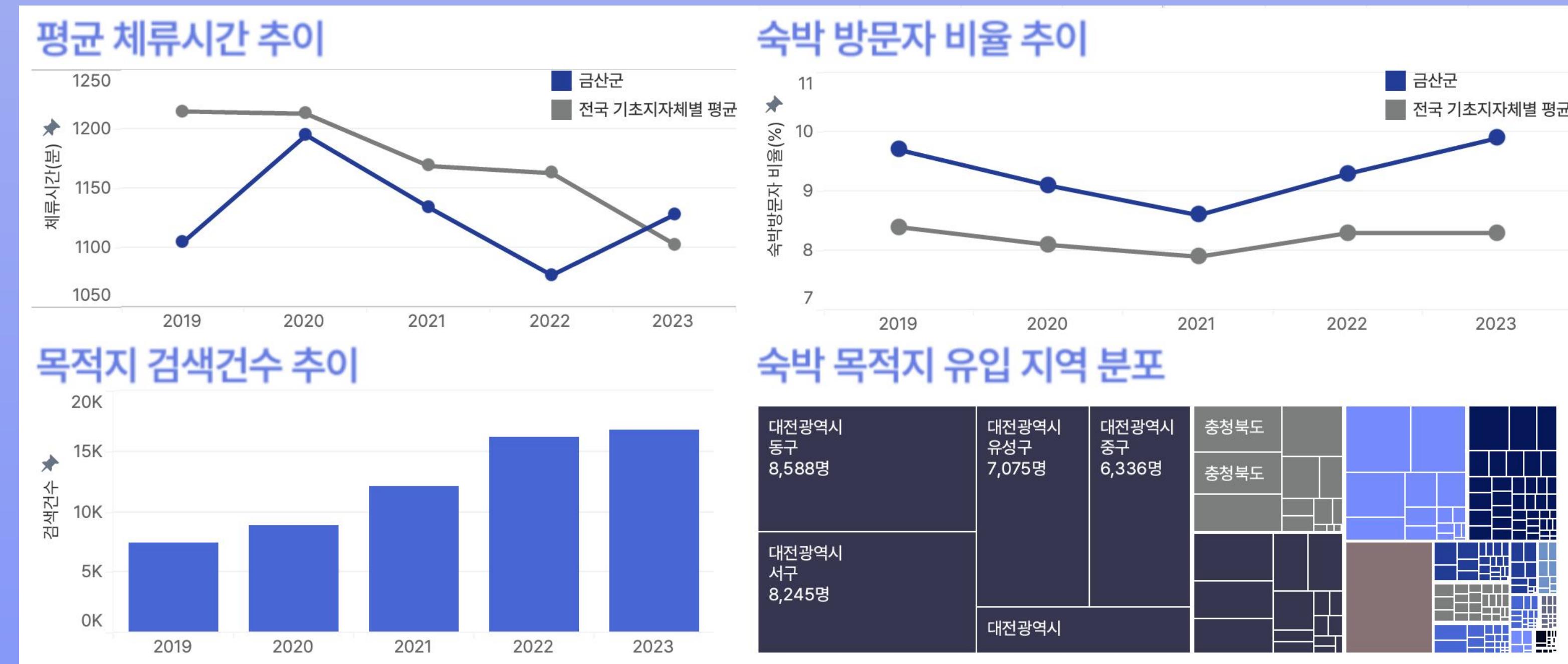
# 시각화 인사이트



# 워드클라우드 시각화

- 분석의 타겟 축제인 '금산인삼축제'와 '금산천벚꽃축제'의 감성분석 결과를 워드클라우드로 시각화
  - 각 단어 선택 시 해당 단어가 포함된 뉴스기사의 감성분석 결과 확인 가능

# 시각화 인사이트

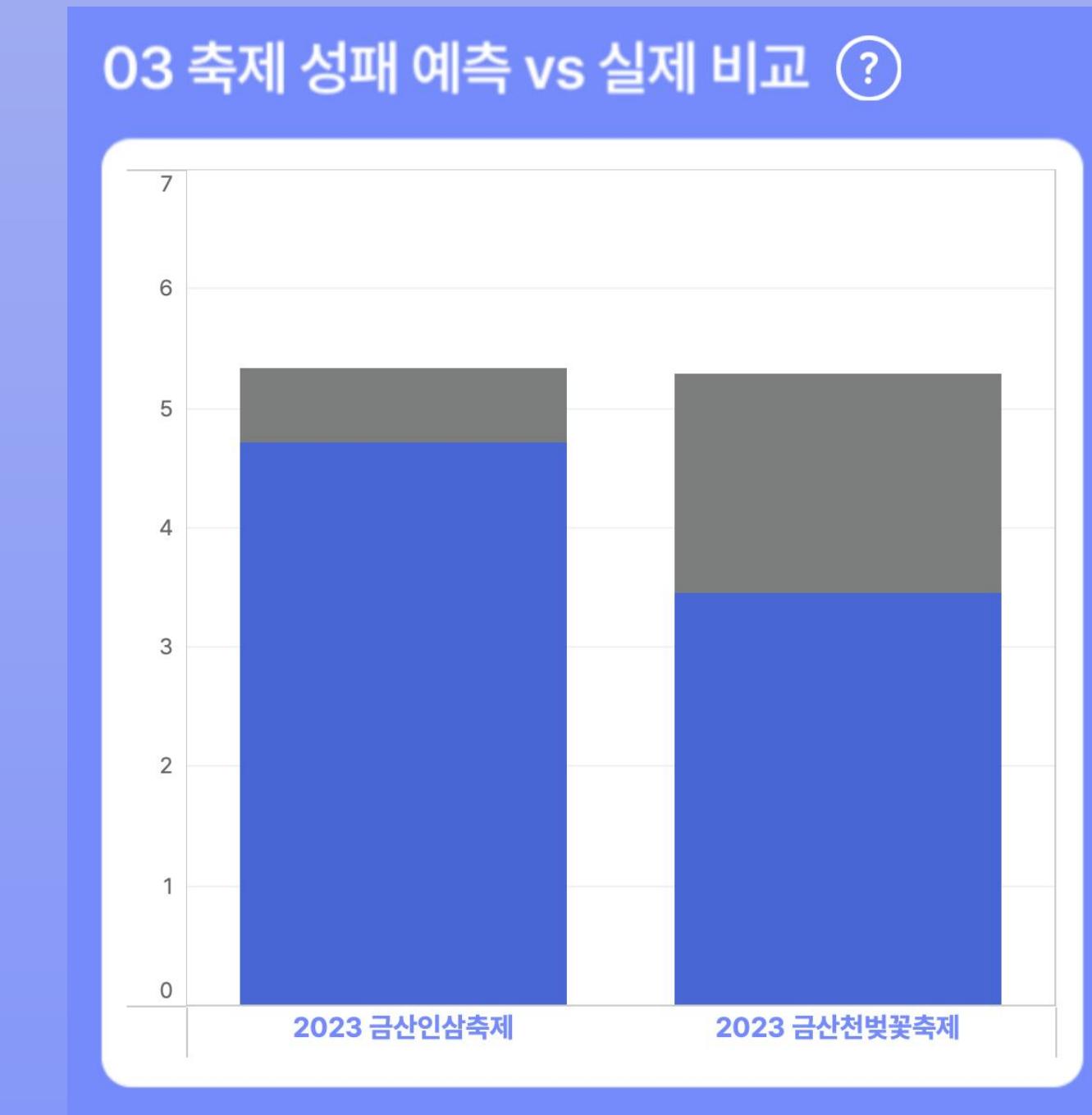


라인차트, 바 차트, 트리맵, 도넛 차트 등

금산군 축제와 관련된 방문객, 숙박, 관광/소비, 소셜미디어 데이터를 현황 파악에 용이하도록 다양한 차트로 시각화

개선안 제안

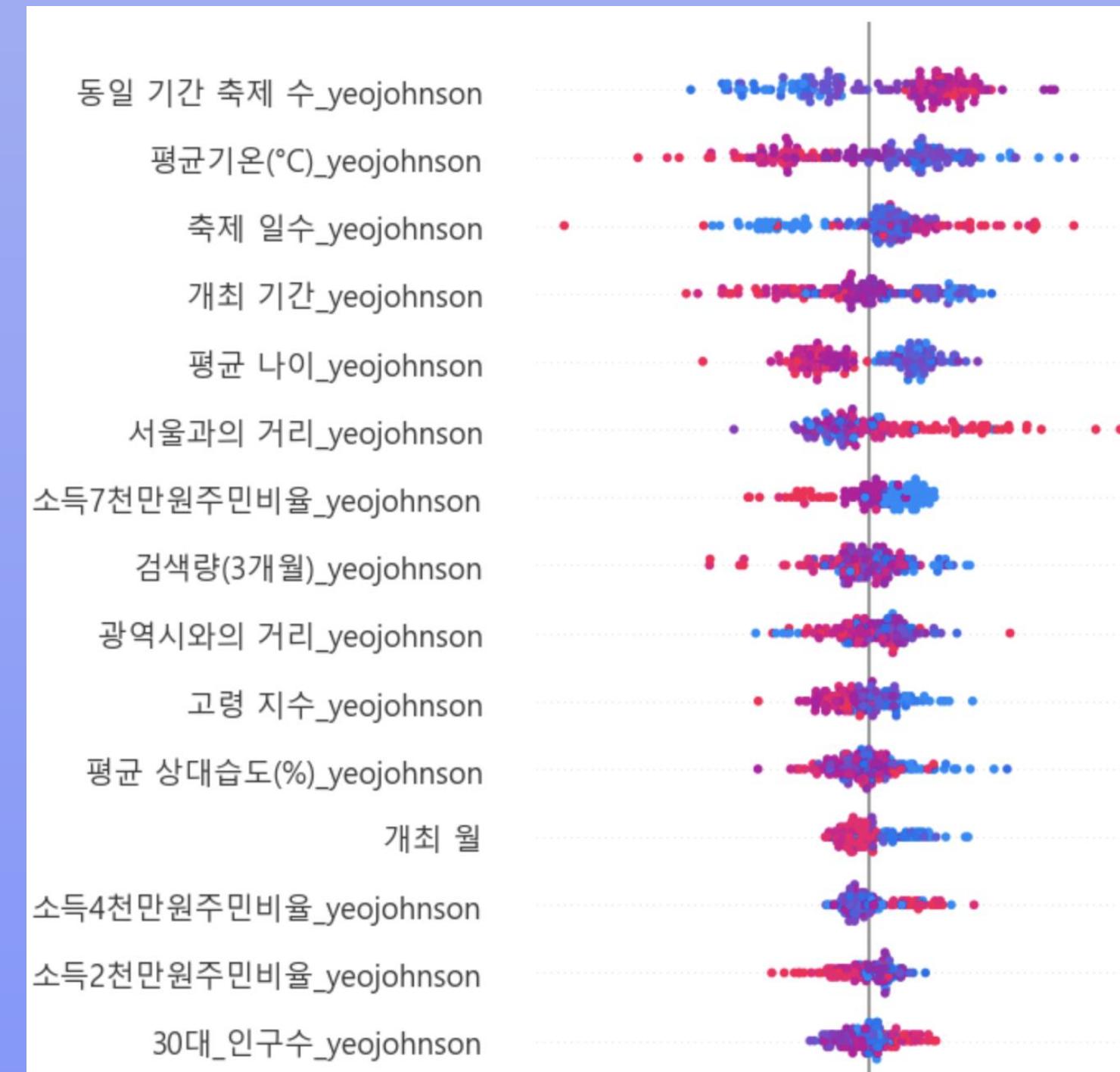
## 시각화 인사이트



## 텍스트 차트, 바 차트

축제 성패 예측 모델에 대한 성능값과 변수중요도, 2023 금산인삼축제와 2023 금산천벚꽃축제에 대한 종속변수(예산 대비 방문객 수) 예측값, 실제값 비교를 효과적으로 시각화.

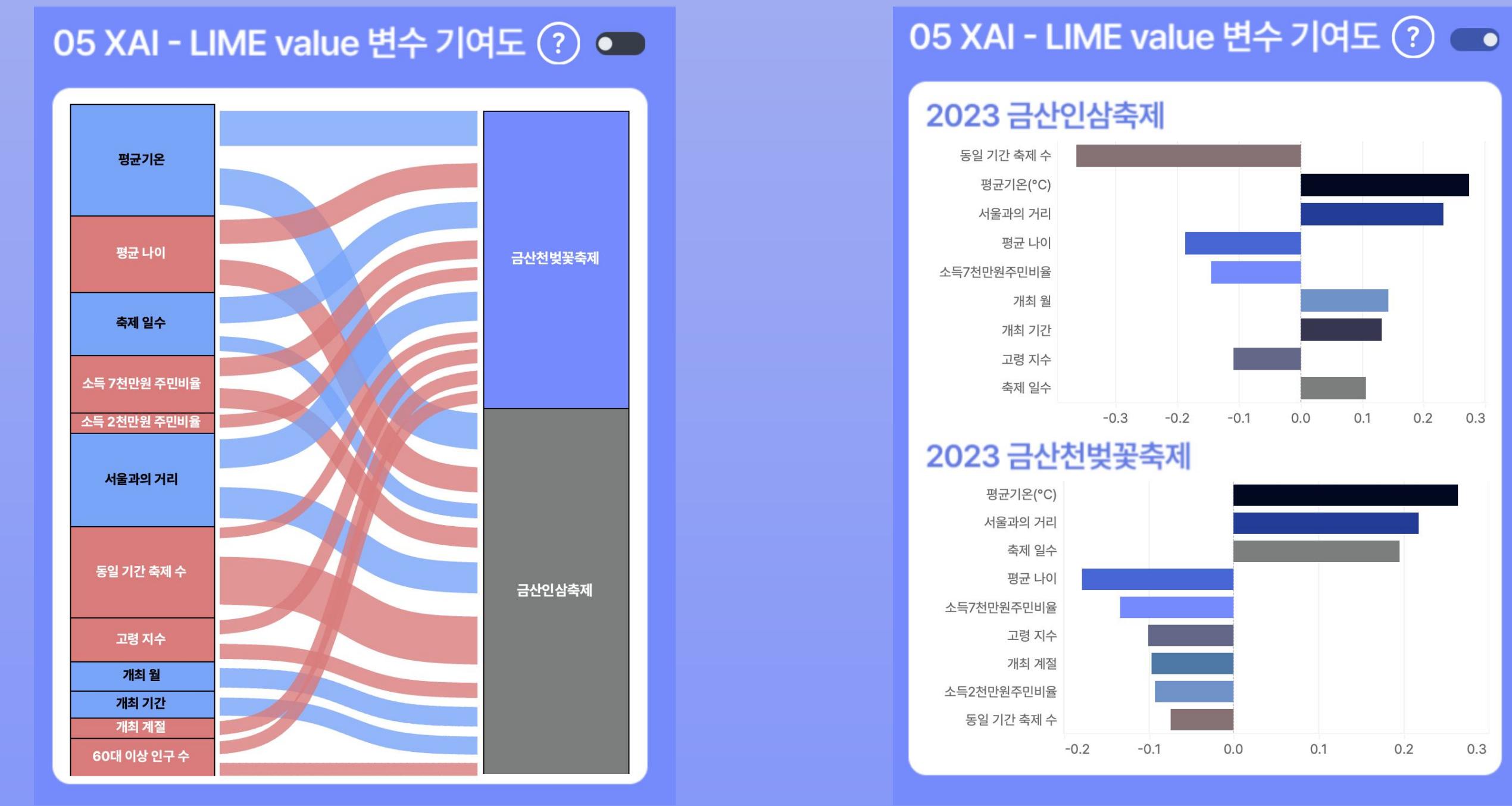
# 시각화 인사이트



## Summary plot

축제 성패 예측에 대한 각 Feature의 중요성과 영향을 계산하는 SHAP Value를 시각화할 때 주로 사용하는 Summary plot을 태블로 상에서 구현

# 시각화 인사이트



## 생키 차트, 바 차트

- 축제 성패 예측에 대한 각 데이터포인트별 Feature의 중요성과 영향을 계산하는 LIME Value를 시각화할 때 주로 사용하는 바 차트로 2023 금산인삼축제와 금산천벚꽃축제에 대한 LIME Value 를 표현
- 각 축제 성패 예측에 대한 Feature의 기여도를 한 눈에 파악할 수 있도록 돋는 생키 차트로 고도화.

개선안 제안

## 시각화 인사이트



### 인덱스 디자인, 스위치 on/off 디자인

많은 양의 차트를 효과적으로 배치하고자 데이터 카테고리별 인덱스와 스위치를 디자인해 클릭 시 해당 페이지로 넘어가도록 동작 가능 구현. 원하는 시각화를 볼 수 있도록 동적 대시보드 구현

## 시각화 인사이트

변수중요도는 해당 모델에서 각 Feature가 예측 결과에 얼마나 중요한지를 평가하는 방법으로, CatBoost 모델의 변수중요도 측정 방법인 PredictionValueChange방법을 사용해 계산합니다.

각 Feature의 값을 제외하고 모델을 다시 학습했을 때의 예측값 변화를 비교하여 변수중요도를 평가하는 방식입니다.

LIME은 특정 예측에 대한 지역적 해석을 제공합니다. 모든 Feature의 조합을 고려하여 예측 기여도를 계산하고, 특정 데이터 포인트에 대한 설명에 장점을 보입니다.

장점에 따라 성패 예측을 진행하기 위해 테스트 데이터로 활용한 '2023년 금산인삼축제'와 '2023 금산천벚꽃축제'의 LIME값을 생기 차트로 시각화했습니다. 라인의 굵기가 굵을수록 Feature 기여도 점수가 높음을 의미하며, 양의 영향을 끼치는 Feature일 경우 파란색, 음의 영향을 끼치는 Feature일 경우 빨간색으로 표시하였습니다.

## 도움말 아이콘

사용자에게 각 차트에 대한 설명을 돋는 물음표 모양의 도움말 아이콘 배치

개선안 제안

# 시각화 인사이트

## Font

Pretendard 프리텐다드

Pretendard 프리텐다드

Gmarket Sans 지마켓 산스

Gmarket Sans 지마켓 산스

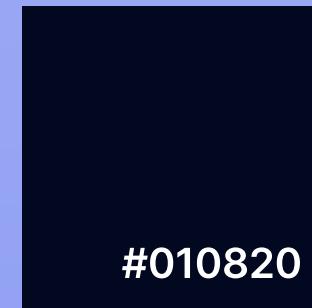
**Bold**

Semi Bold

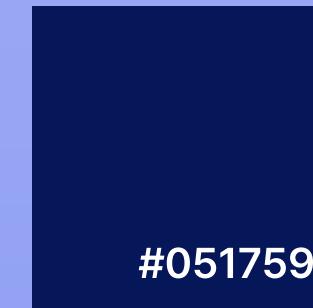
**Bold**

Medium

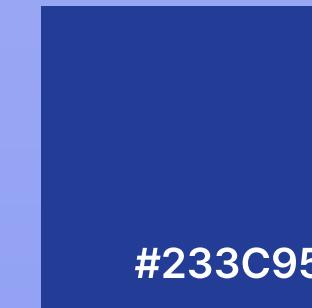
## Color



#010820



#051759



#233C95



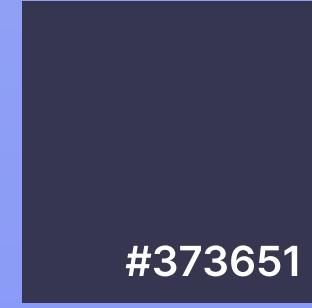
#4866D4



#758AFD



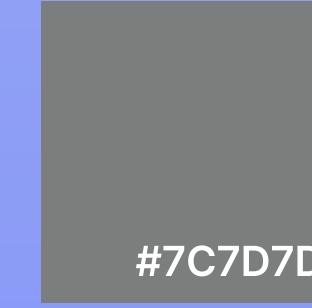
#6F93C9



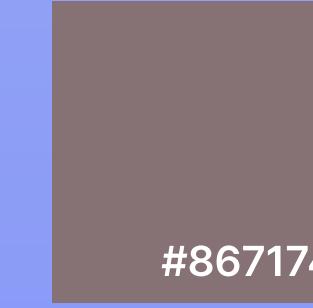
#373651



#646788



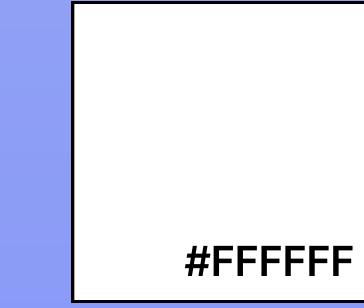
#7C7D7D



#867174



#F25F74



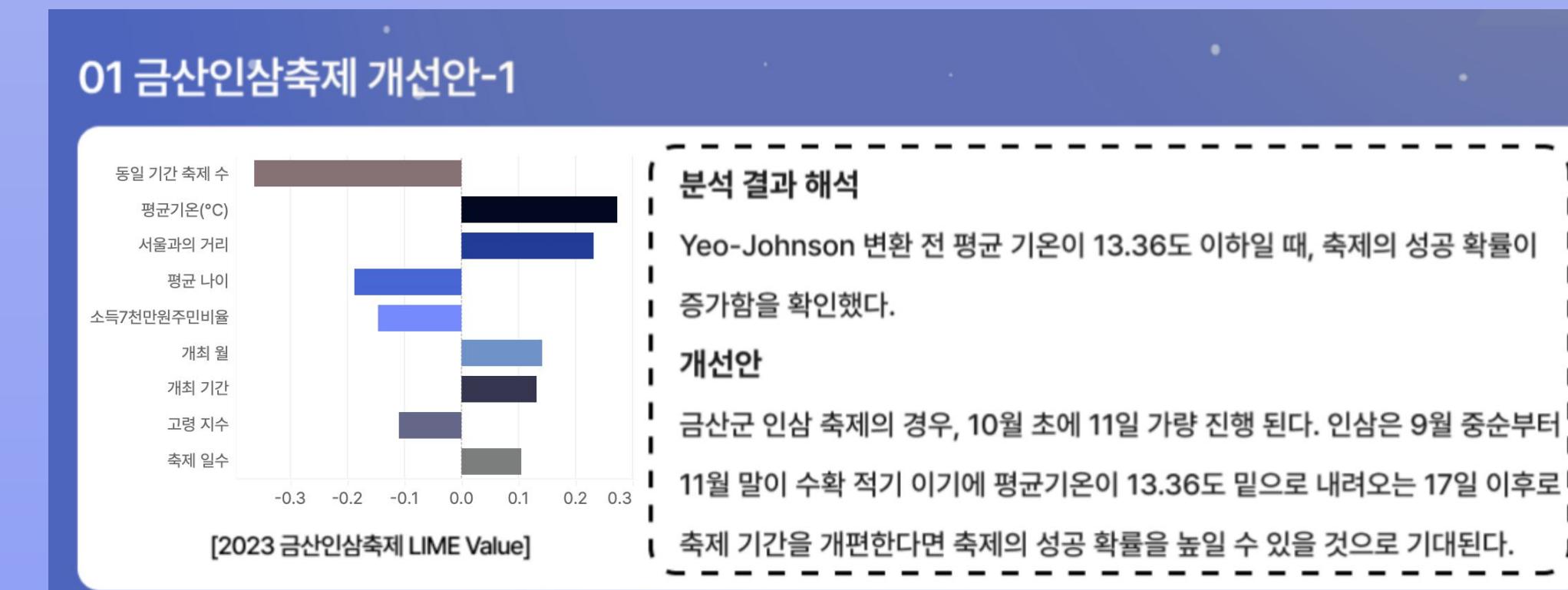
#FFFFFF

## 컬러 및 폰트 통일

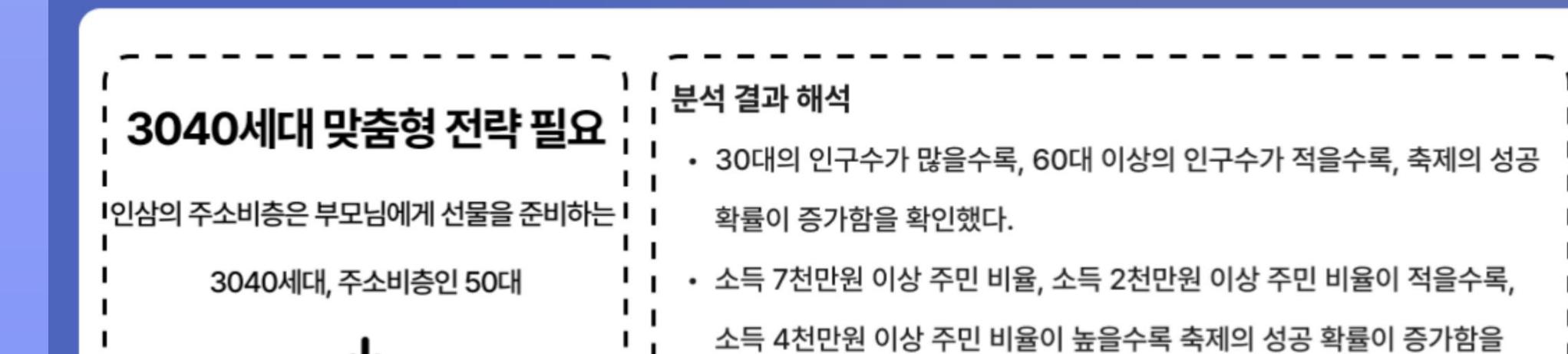
축제 분위기 어울리는 색상 및 폰트를 제목 및 차트에 동일하게 적용해 대시보드의 전체적인 통일감 강조

## 개선안 제안

# 시각화 인사이트



### 02 금산인삼축제 개선안-2



## 개선안 제안

분석 결과 해석과 개선안을 담아 현황 지표 → 모델링 지표 → 개선안 파악으로 연결되는 유기적인 대시보드 구성

개선안 제안

## 대시보드 시연

금산군  
도농상생을 위해 XAI를 활용한  
**금산군 축제 성패  
예측 모델링 및 제안**



금산군 축제 분석

금산군 축제

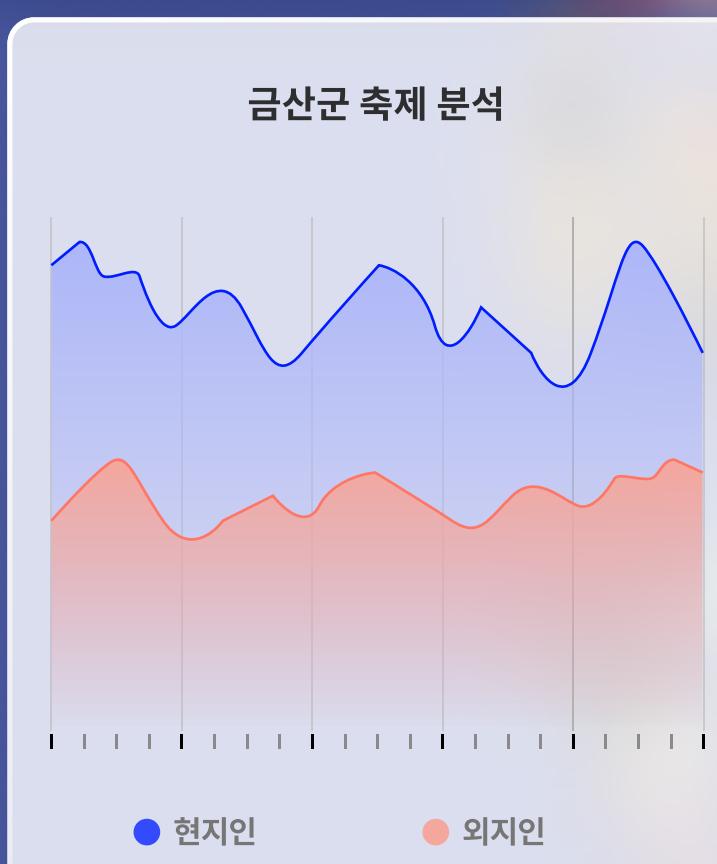
Month	Value
Jan	100
Mar	120
May	110
Jul	90

??% 금산군 축제 성패 측정결과는 다음과 같습니다.

성패 예측 보러가기

금산군 축제 분석

● 현지인 ● 외지인



대시보드 시연 영상

Tableau Public



# 감사합니다

금산군 데이터 기반 행정 시각화 아이디어 공모전

국준호, 권동구, 유소현

