



6

# 서울 관광지 혼잡도 예측

## : 방한 외국인의 효율적 여행을 위한 데이터 인사이트

오윤진 김효정 박성준 박태호 서동완 신진수 윤채희

# Contents

## 01 Introduction

- 분석 배경
- 분석 방향

## 02 데이터 수집 및 전처리

- 데이터 수집
- 데이터 전처리 및 변수 선정

## 03 모델링

- 랜덤 포레스트
- XGBoost
- SVR

## 04 결론 및 기대효과

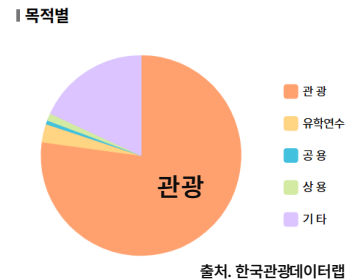
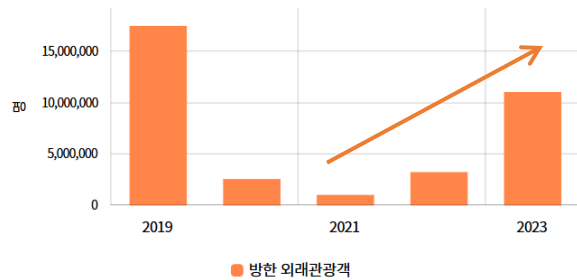
- 모델링 결과
- 기대효과
- 기술적 보완점
- 비기술적 보완점

## 01. Introduction

## 주제

외국인을 위한 서울 관광지 7곳에 대한 혼잡도 인사이트를 제공해 최적의 여행 경험을 돕는다

## 프로젝트 배경



코로나 이후 관광객이 크게 **증가**하고 있음

-> 관광지 방문객 수를 예측해 혼잡도를 줄일 필요성 대두

## 프로젝트 목적

실시간 관광지 혼잡도 정보와 관광객의 행동반응

Real-time destination crowding information and tourists' behavioral responses

장재협

한국관광학회 · 2021.07

한국관광학회 국제학술발표대회집 · vol. 90 · 480-491(12pages)

관광객에 내몰리는 주민들, 지속 가능한 여행은

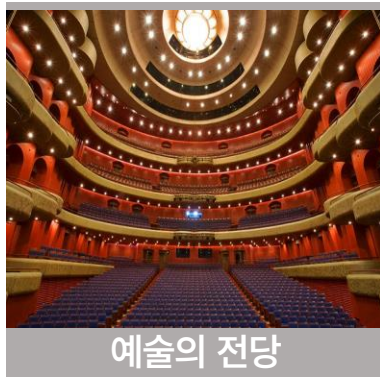
조인우 기자 | 입력 2024.09.02 01:27

관광지 혼잡도를 줄임으로써,

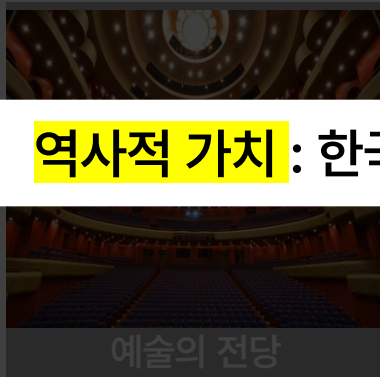
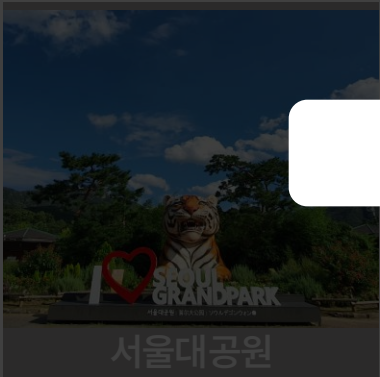
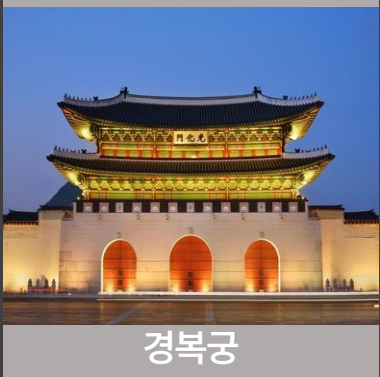
-> **한국 여행의 긍정적인 이미지** 강화

-> 명소의 밀도를 분산시켜 인근 거주민의 편안한 일상 보장

## 서울 명소 7곳 선정 이유



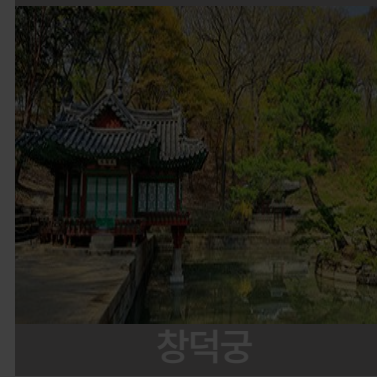
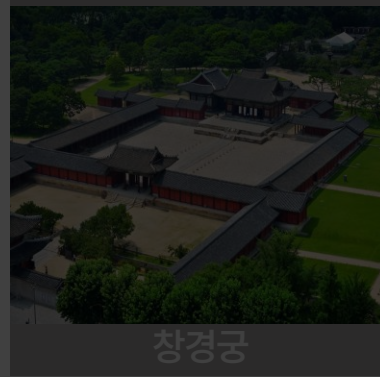
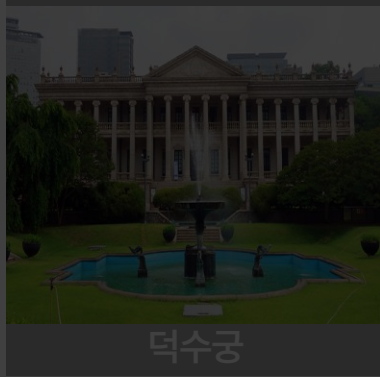
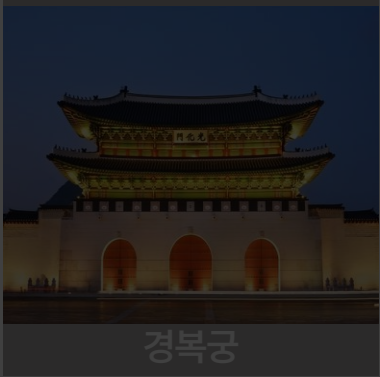
## 서울 명소 7곳 선정 이유



**역사적 가치**: 한국을 대표하는 5대 궁을 통한 한국의 전통문화 소개

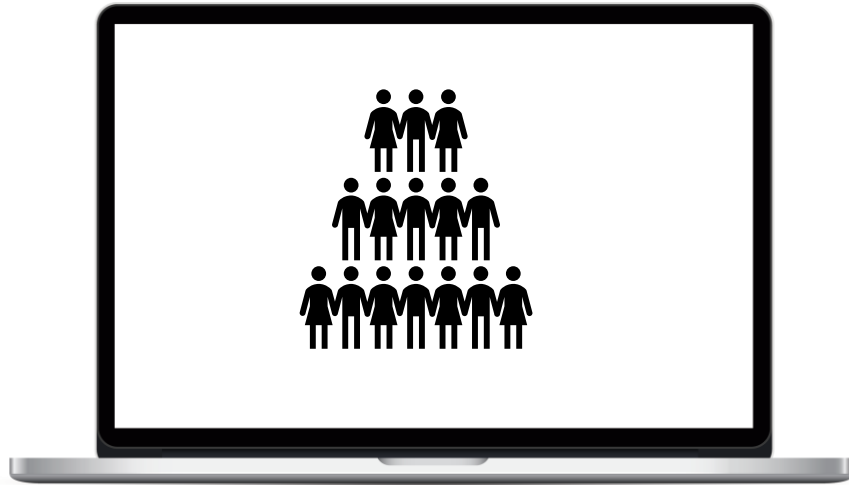


## 서울 명소 7곳 선정 이유



**문화적 가치**: 한국의 공연과  
액티비티, 고유종 경험

## 종속 변수 소개



1 관람객 수

2 혼잡도

$$\text{혼잡도}(\%) = \frac{\text{예측 관람객 수}}{\text{수용 인원}} \times 100$$

$$\text{수용인원} = \frac{\text{바닥면적 합계}}{4.6m^2} + \text{관람석(고정식 의자) 수} + \frac{\text{긴 의자의 정면너비}}{0.45m}$$

\*소방법률 제15조



## 02. 데이터 수집 및 전처리

# 데이터 출처



## 공공데이터

- 4대궁 관람객 수-국가유산청 공능유적본부
- 예술의 전당 입장객
- 환율 변동 현황-한국산업은행



## 국가 통계

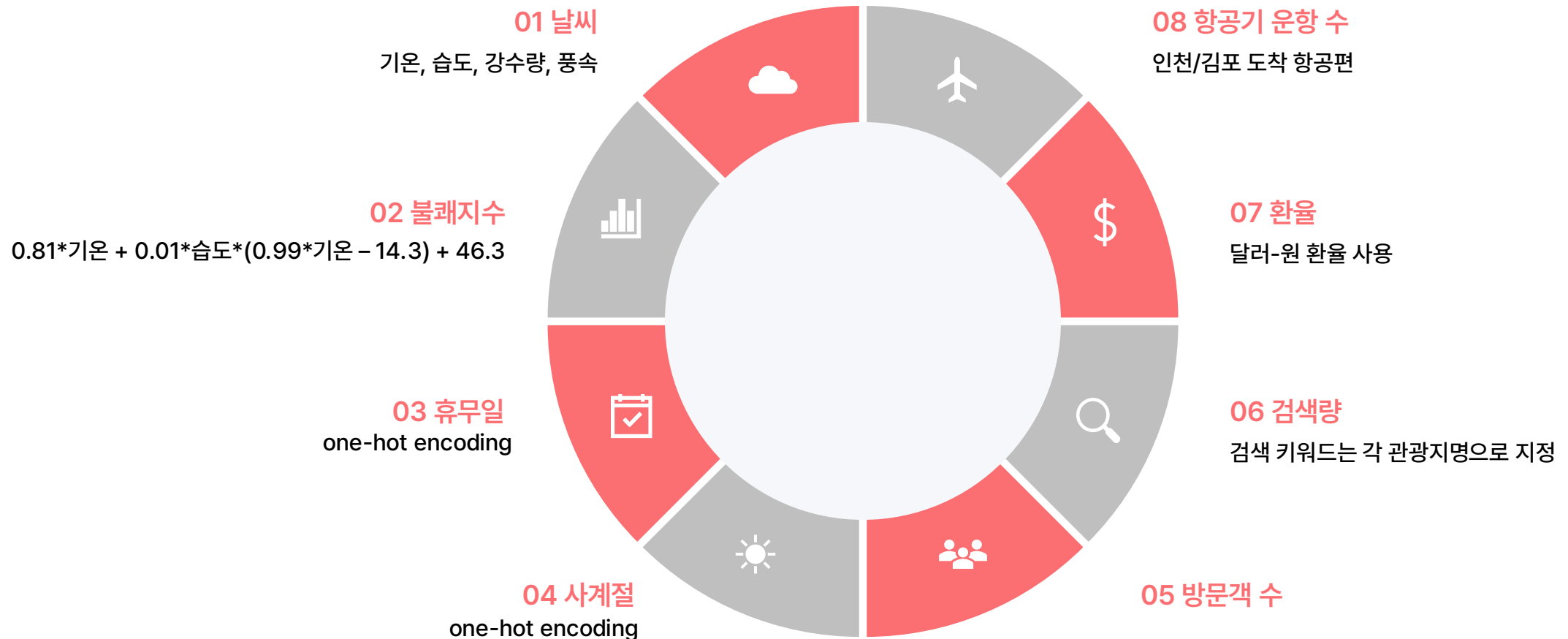
- 서울대공원 일일입장객수
- 미세먼지, 강수량, 기온, 풍속  
- 실시간 도시데이터



## API

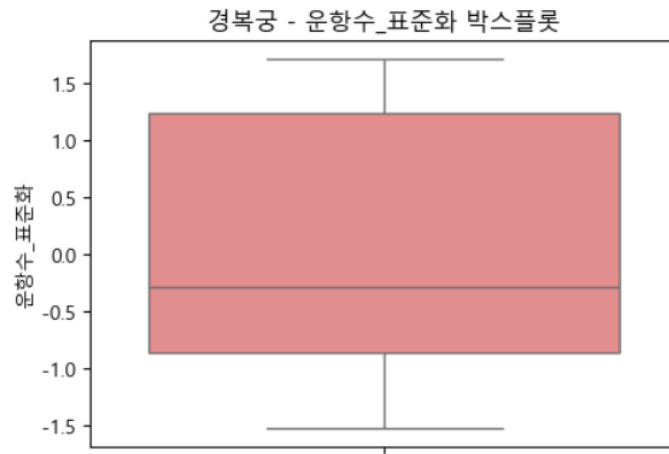
- 검색량-네이버 데이터랩, 구글 트렌드
- 운항정보-항공정보포털시스템

# 독립변수

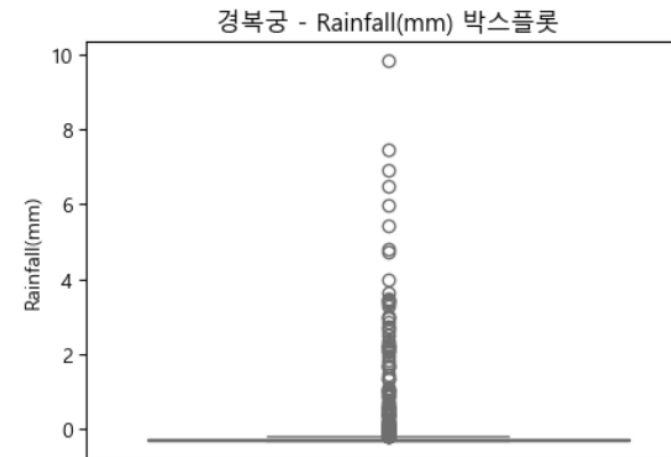


## 경복궁 기준 이상치 확인

## 이상치 X



## 이상치 O

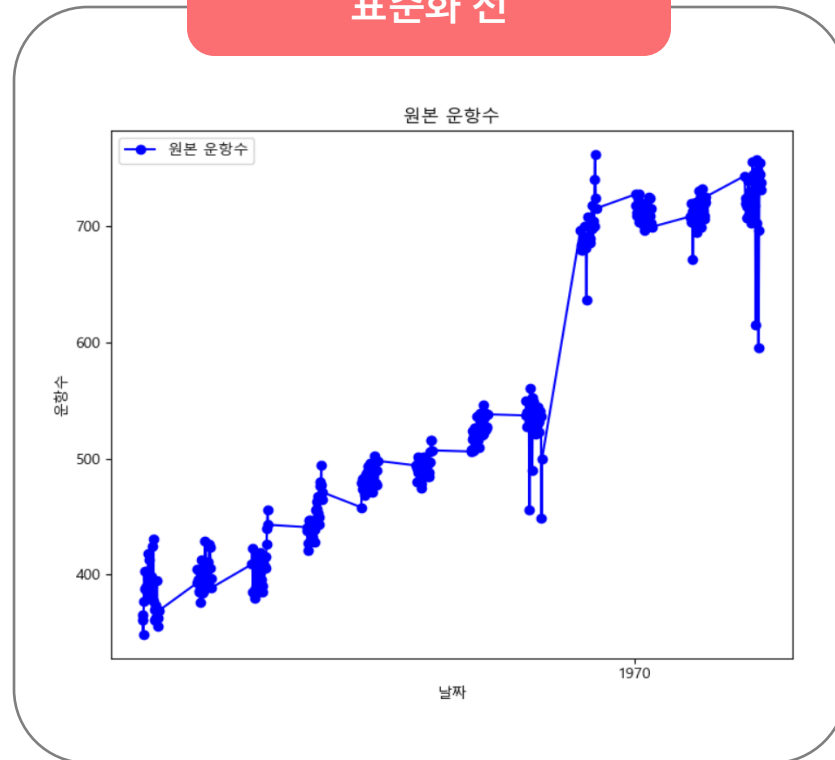


➔ 강수량 데이터는 본질적으로 불균형하기에  
현실적인 변동성을 반영했다고 판단

표준화 진행 변수(항공기 운항 수, 달러환율, 불래지수, 날씨 데이터)

`StandardScaler()`

표준화 전

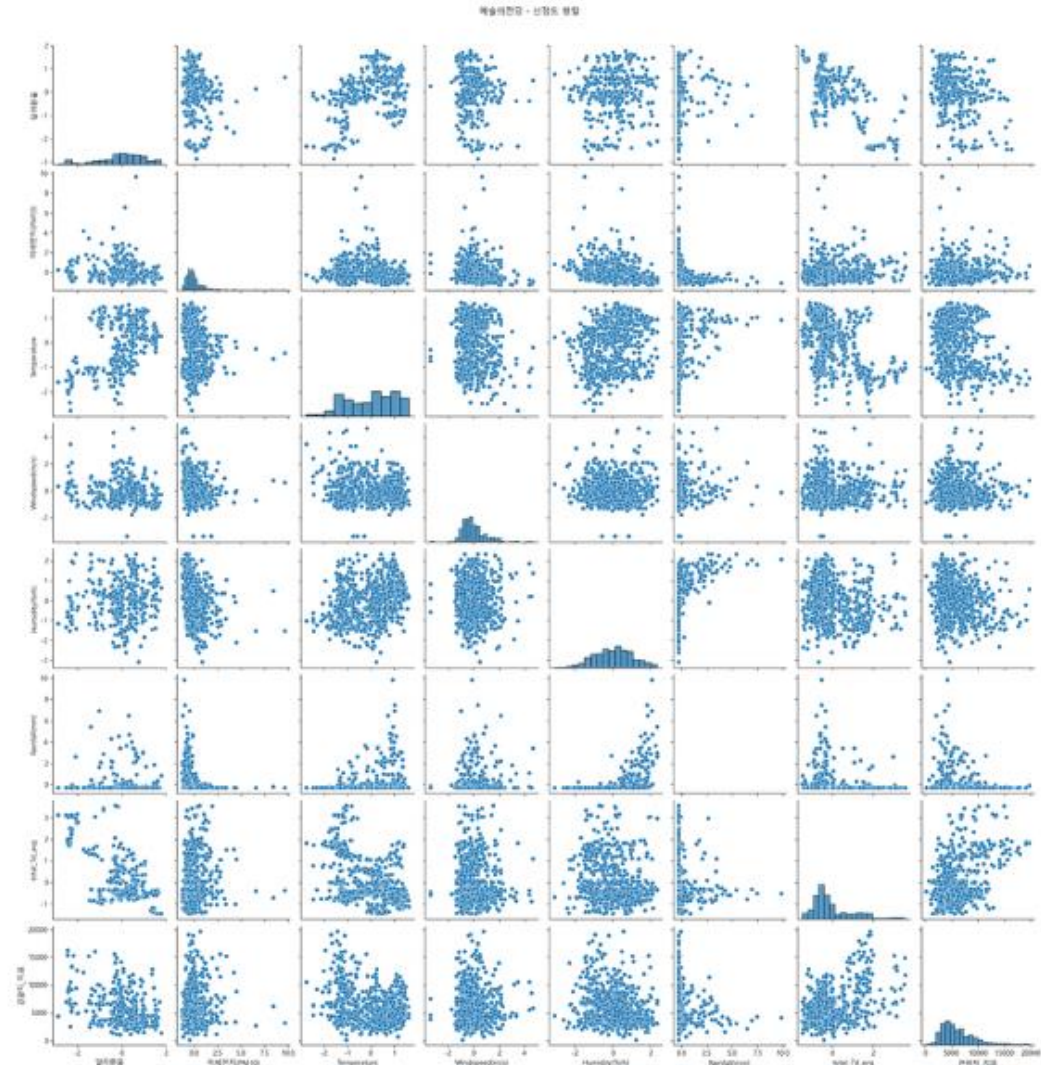
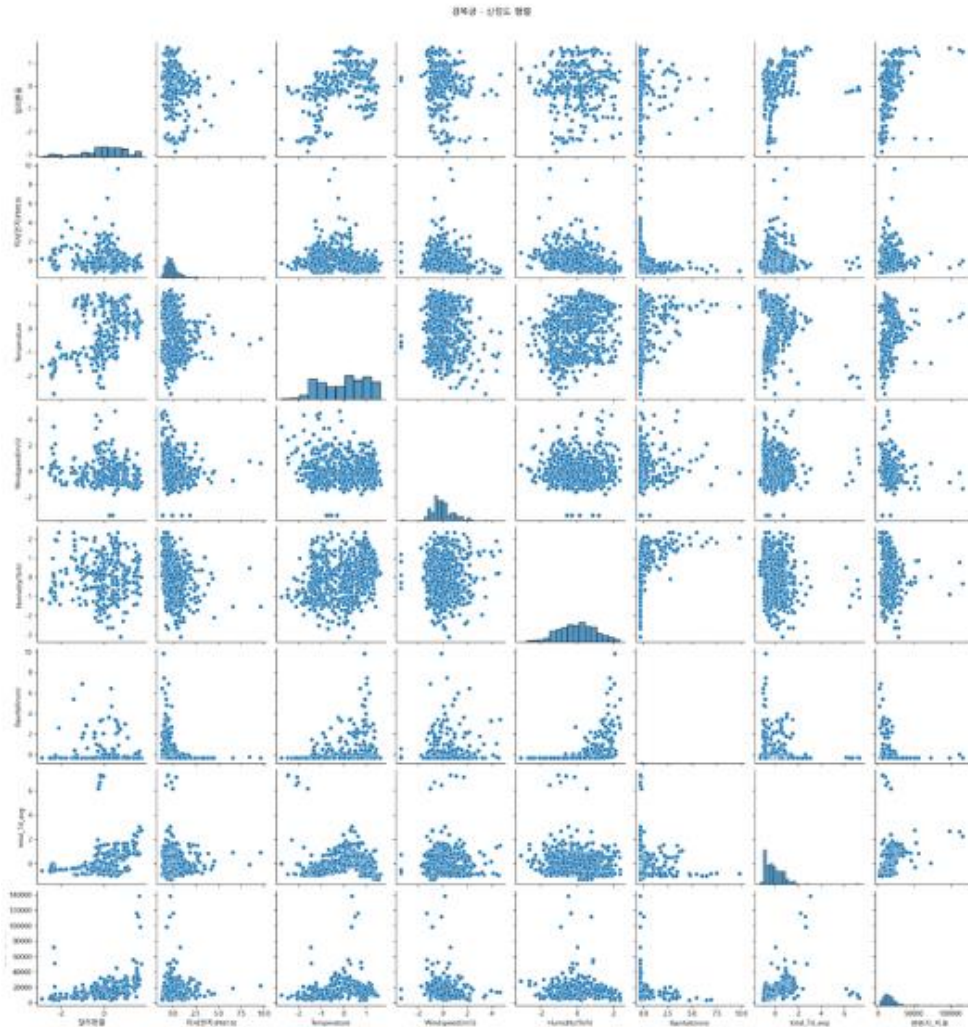


표준화 후



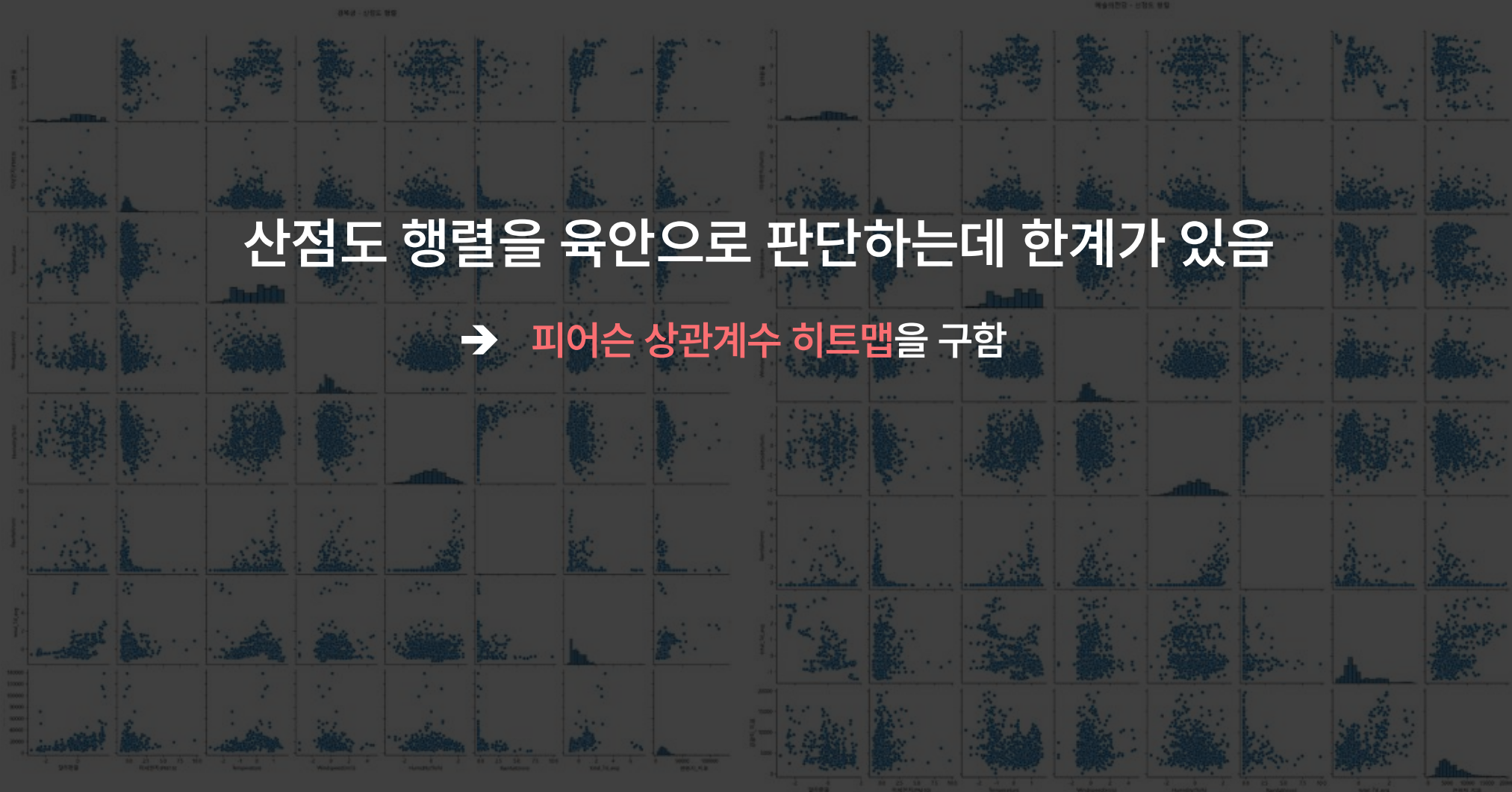


## 변수 선정 과정\_산점도 행렬

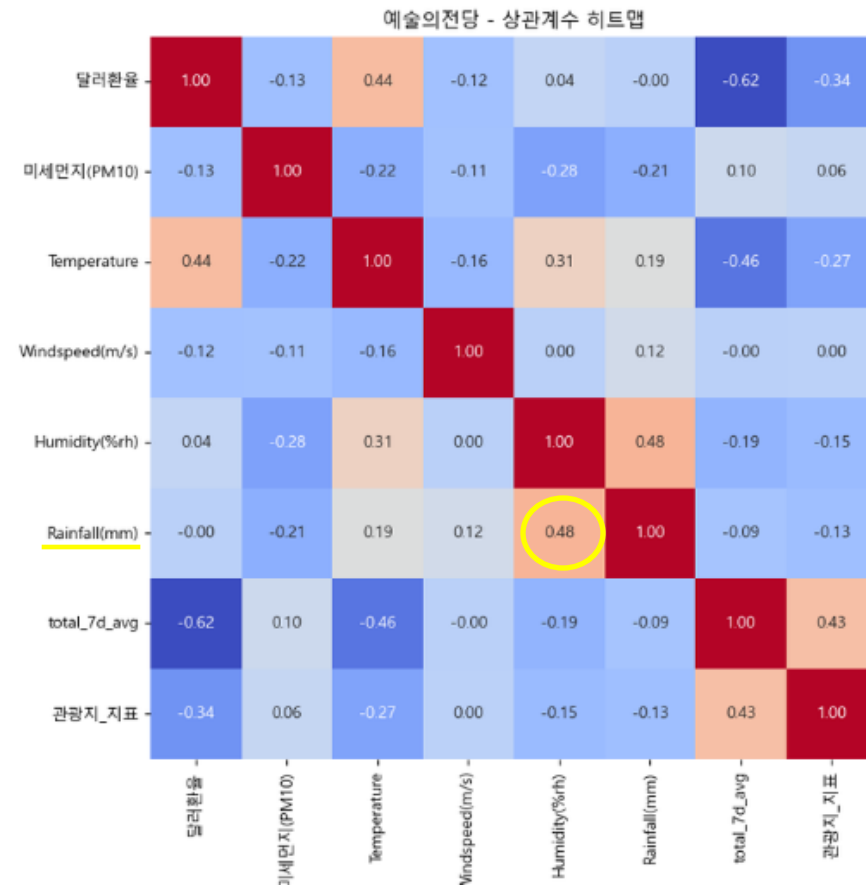
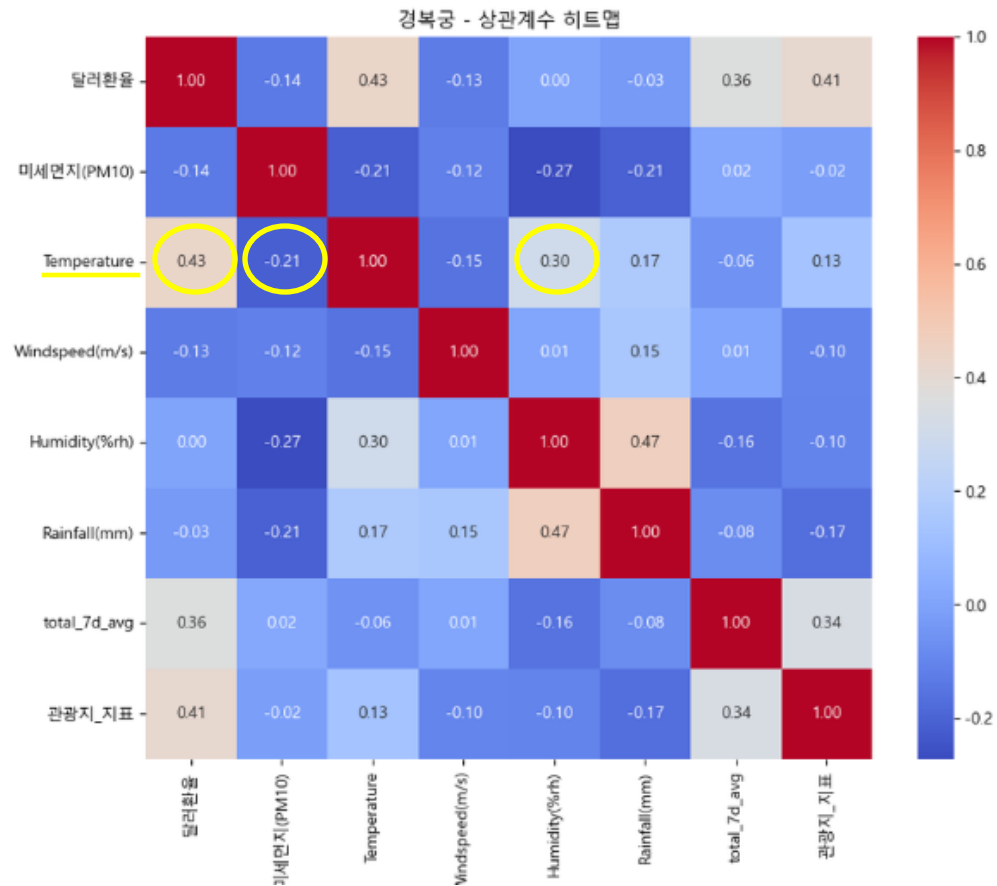




## 변수 선정 과정\_산점도 행렬



## 변수 선정 과정\_피어슨 상관계수 히트맵



r 값	관계
+0.7 ~ +1.0	강한 양적 상관관계
+0.3 ~ +0.7	뚜렷한 양적 상관관계
+0.1 ~ +0.3	약한 양적 상관관계
-0.1 ~ +0.1	상관관계 거의 없음
-0.3 ~ -0.1	약한 음적 상관관계
-0.7 ~ -0.3	뚜렷한 음적 상관관계
-1.0 ~ -0.7	강한 음적 상관관계

## 변수 선정 과정\_피어슨 상관계수 히트맵



## 변수 변경 전후 모델 비교

## 기존 모델

RMSE

	덕수궁	종묘	서울대공원	예술의 전당
RF	1965	742	3655	2130
XGBoost	2084	876	4387	3126
SVR	2414	960	5227	2608

R<sup>2</sup> score

	덕수궁	종묘	서울대공원	예술의 전당
RF	0.8743	0.5872	0.544	0.5349
XGBoost	0.8169	0.4442	0.2969	0.2292
SVR	0.8155	0.1365	0.1943	0.2105

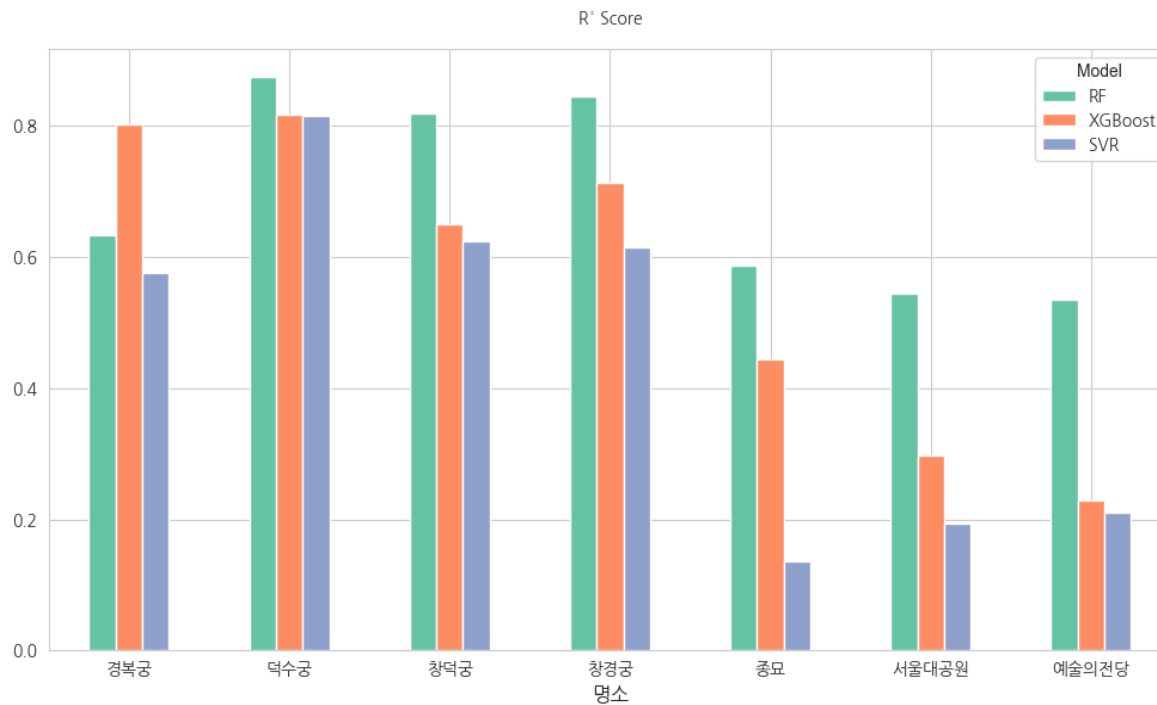
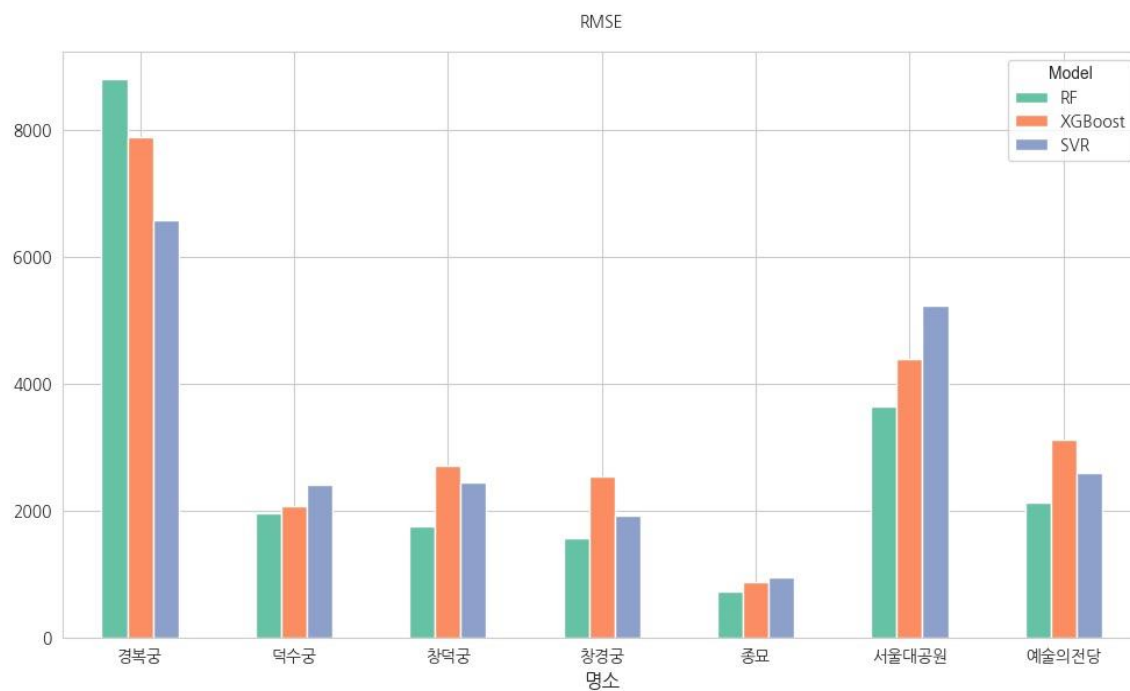
## 변경된 변수 사용 모델

	덕수궁	종묘	서울대공원	예술의 전당
RF	1424	545	2293	2210
XGBoost	2610	835	3290	4444
SVR	2837	1165	5402	2505

	덕수궁	종묘	서울대공원	예술의 전당
RF	0.934	0.777	0.8205	0.4997
XGBoost	0.7129	0.4952	0.2786	0.1462
SVR	0.7584	0.1725	0.3057	0.4637

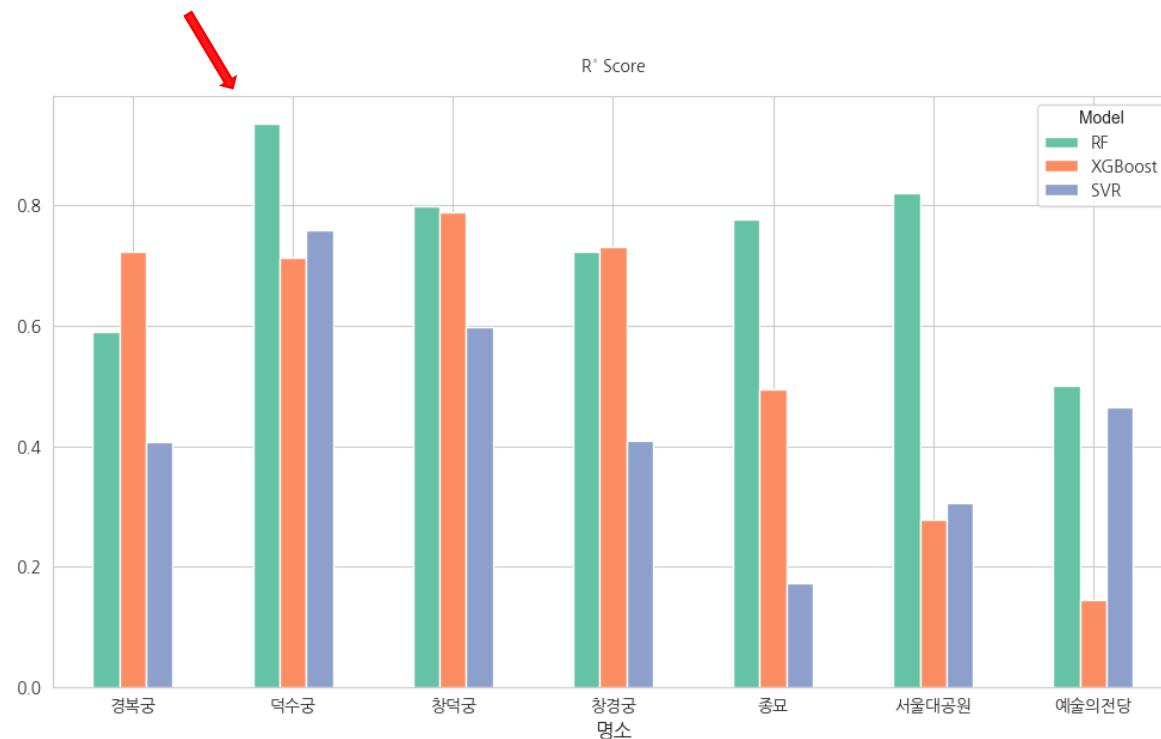
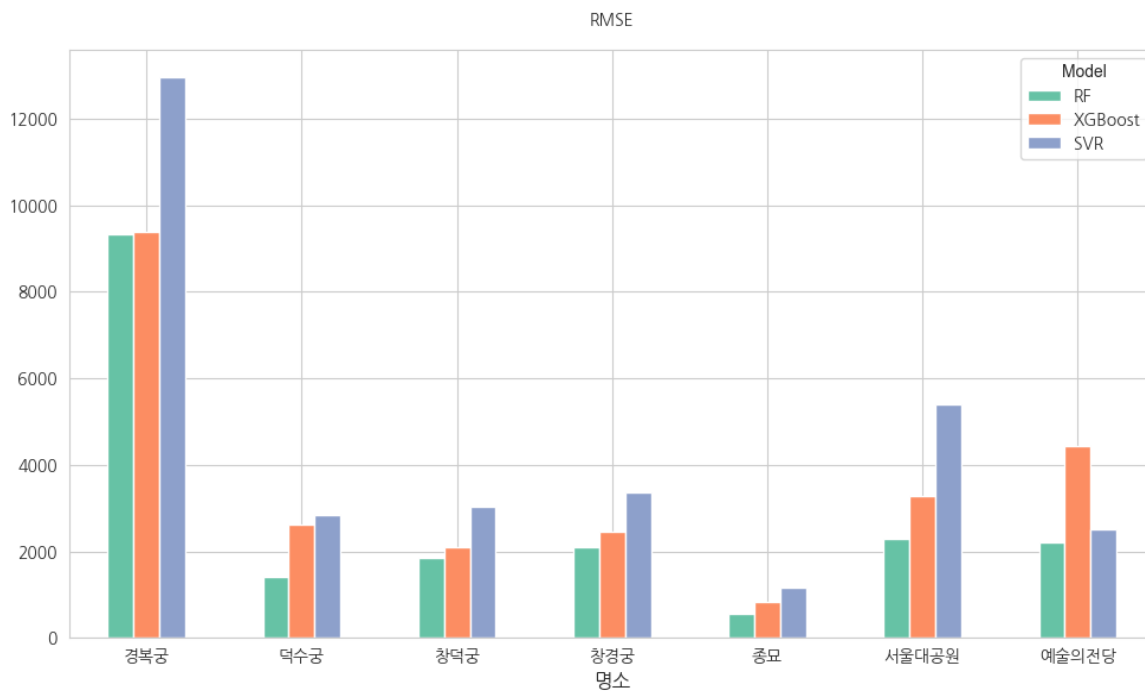
### 03. 모델링

Feature : 달러 환율, 미세먼지, 풍속, 강수량, 검색량 키워드

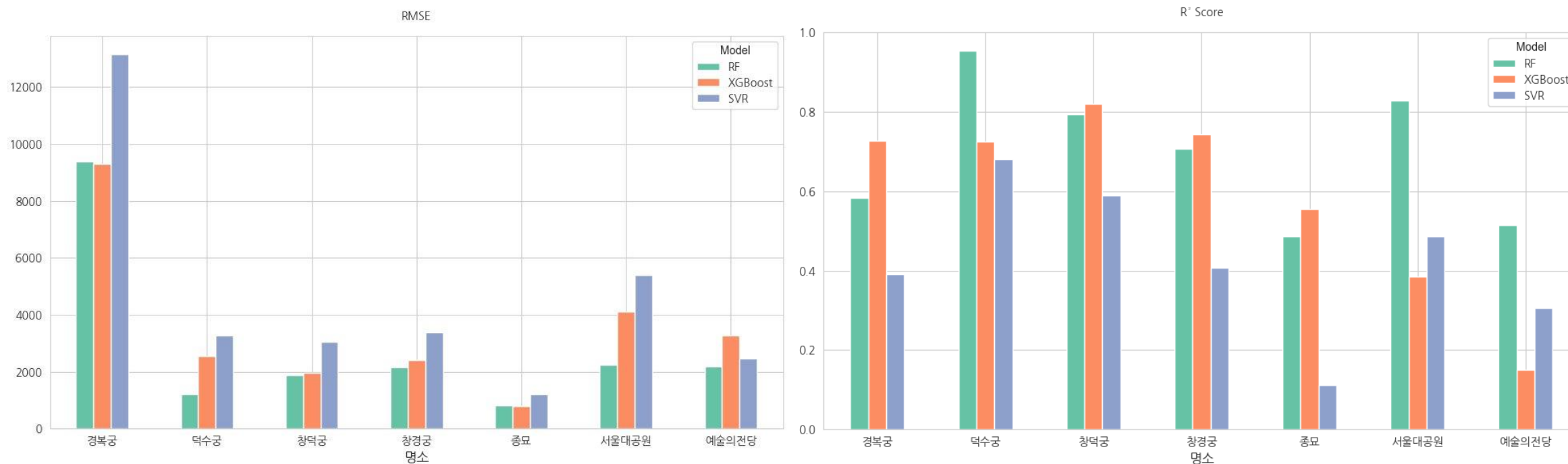




Feature : 달러 환율, 미세먼지, 풍속, 불쾌지수, 강수량, 검색량 키워드, 요일 (7일), 계절 (4계절)

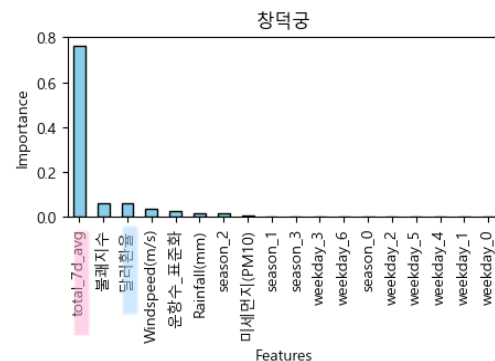
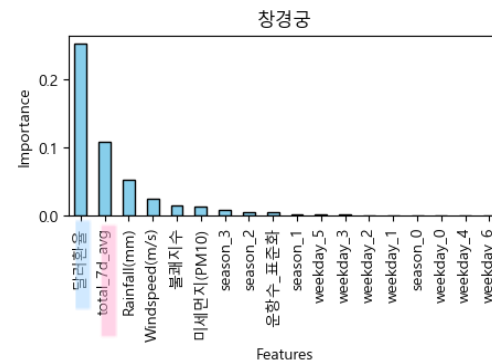
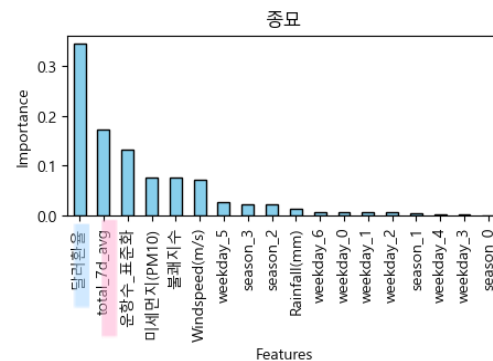
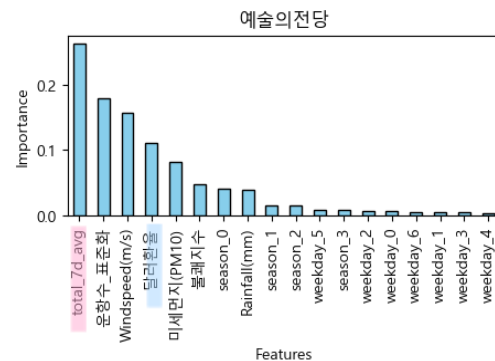
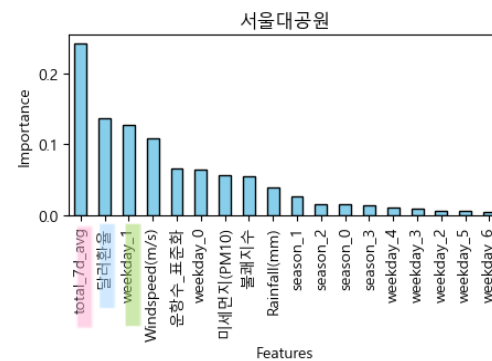
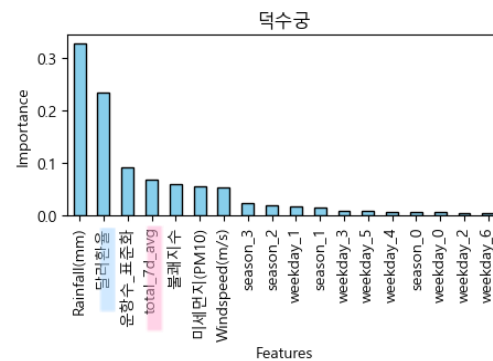
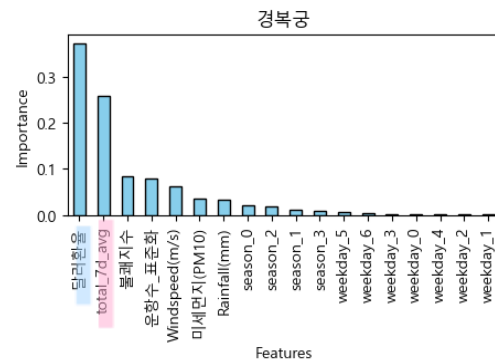


Feature : 달러 환율, 미세먼지, 풍속, 불쾌지수, 강수량, 검색량 키워드, 요일 (7일), 계절 (4계절), 운항수\_표준화



## RandomForest

## Feature importance



## Random Forest

적용 이유:

- 여러 변수(예: 요일, 계절, 날씨 등)가 방문자 수에 영향을 미치는 복잡한 환경에서 변수 중요도를 확인하며 예측함
- 이상치에 덜 민감
- 과적합 방지에 좋음

## XGBoost

적용 이유:

- 관광지 방문자수는 시간에 따른 복잡한 비선형 패턴이 존재하며, 계절성, 요일 등 여러 복합적인 요인이 영향을 미침. XGBoost는 이러한 비선형 관계를 잘 학습함

## SVM

적용 이유:

- 특정 요일이나 특정 시즌에 극단적인 방문자 수 변화가 있는 경우, 이를 잘 구분할 수 있는 결정 경계를 찾기 위해 사용함
- 그러나 하이퍼 파라미터에 민감
- 이상치에 매우 민감

## 04. 결론 및 기대효과

## 모델 최종 결과

사용자는 웹 상에서 혼잡도 기반 안내형 대시보드를 확인하여 관광에 필요한 정보를 얻습니다.

## 서울 관광지 혼잡도 예측 대시보드

관광지 선택

서울대공원

요일 선택

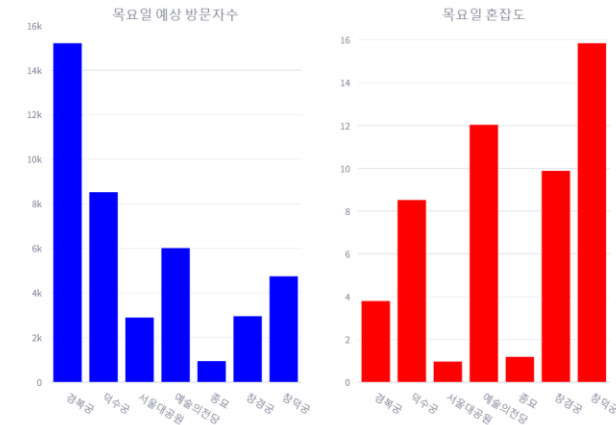
목요일

예상 방문자수: 2909명

혼잡도: 0.97%

## 혼잡도 랭킹 (낮은 순)

- 1. 서울대공원 - 혼잡도: 0.97% (여유)
- 2. 종묘 - 혼잡도: 1.19% (여유)
- 3. 경복궁 - 혼잡도: 3.80% (여유)
- 4. 덕수궁 - 혼잡도: 8.53% (여유)
- 5. 창경궁 - 혼잡도: 9.89% (여유)
- 6. 예술의전당 - 혼잡도: 12.05% (여유)
- 7. 창덕궁 - 혼잡도: 15.86% (여유)

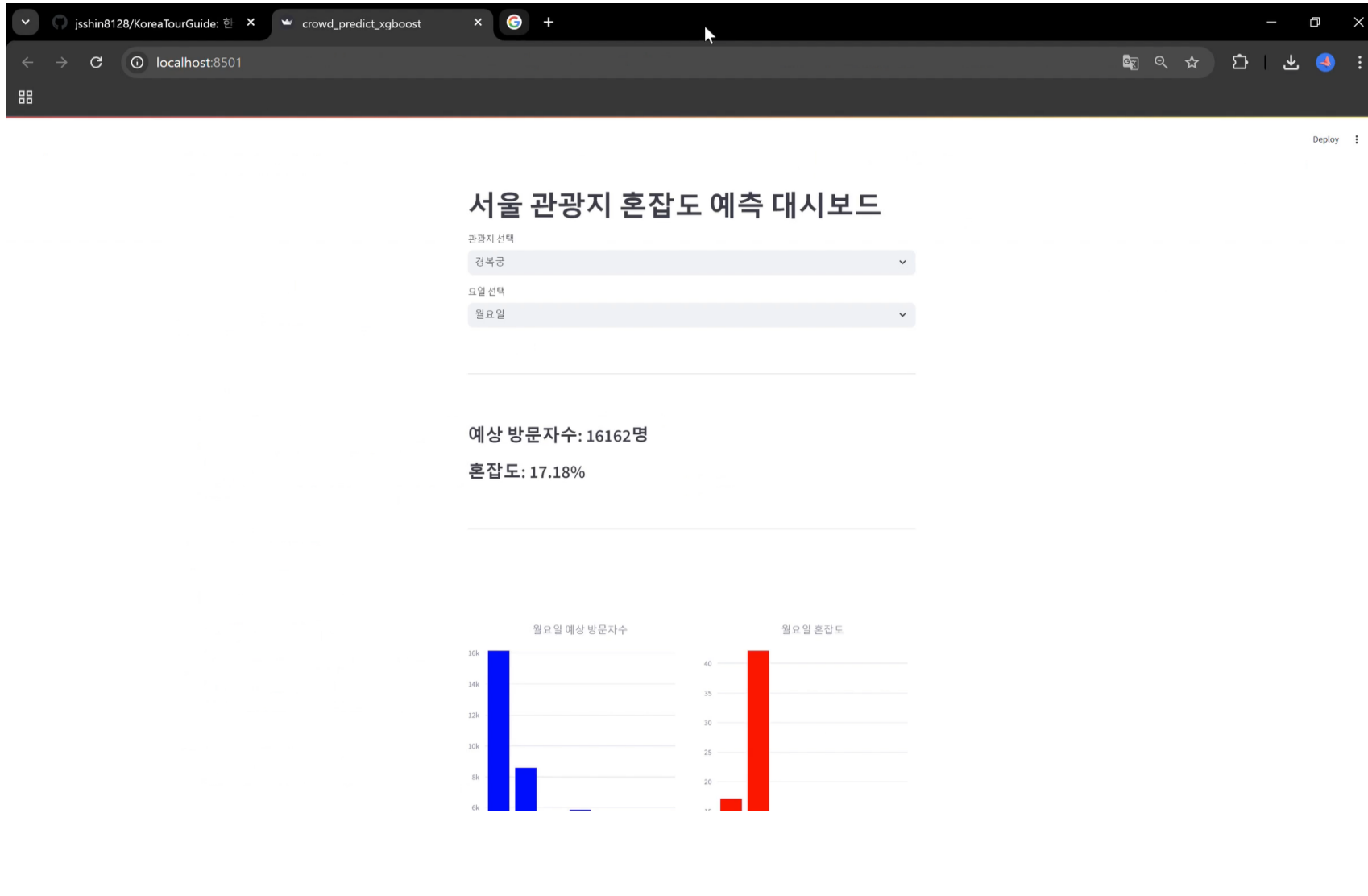


## 대안 추천 관광지 🔥 ⇄

여기는 어때요? 서울대공원 🍦  
 예상 방문자수: 2909명  
 혼잡도: 0.97%



## 시연 영상



## 데이터 기반 스마트 관광 기술

외국인 관광객의 "기분 좋은 경험"

쾌적한 관광을 통한 긍정적 피드백



관광객 분산 유도

운영 효율 증대

관광지 주변 거주인의 생활 환경 개선

→ 지속 가능한 여행지

장기적으로 관광 편의성이 높은 나라라는 국가 브랜드 형성

## Pain Points

일별 데이터라는 한계로 인해  
더 많은 변수를 고려하지 못함



혼잡도에 영향을 미치는 **피쳐 추가** 구상

2023년 초반 코로나로 인해 항공기 운  
항 수 데이터가 비정상적으로 적음



**코로나 이후**의 경향을 반영할 때까지 충분한 기간이 필요

수동으로 조정된 하이퍼파라미터



그리드 서치 등을 이용해 **최적의 하이퍼파라미터**를 찾기

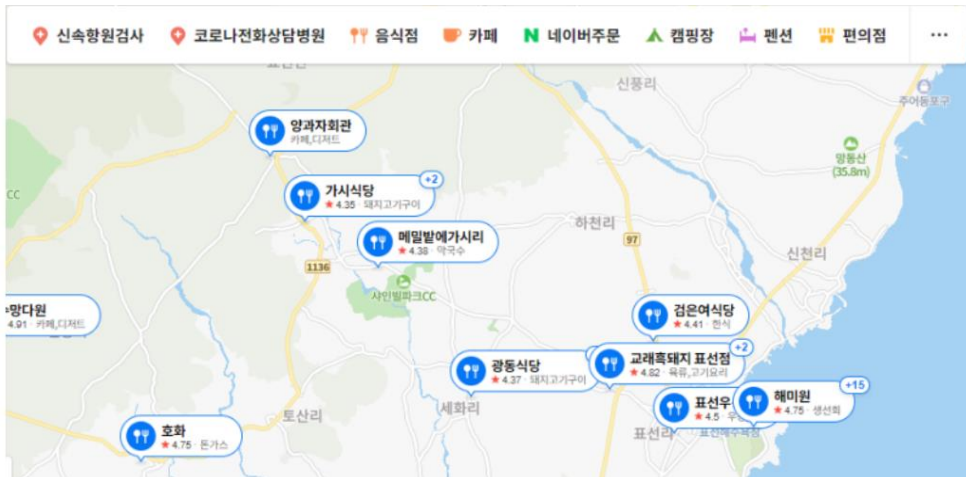
# 서울 관광지의 혼잡도를 줄이기 위한 또 다른 방법

## 데이터 기반 명소 발굴

- 소셜 미디어 분석 : 해시태그 및 방문 후기분석  
-> 인기 상승 명소 파악
- 지도 검색 데이터 활용 : 검색량 및 리뷰 증가 추이  
-> 잠재 관광지 발굴

## 테마별 여행 코스 기획

- 자연 휴양, 문화 체험, 맛집 투어, 액티비티 등  
테마별로 큐레이션 된 관광 명소를 SNS나  
커뮤니티를 통해 홍보



6

Thank you :)