

# 교통·인구 통계 및 공연 예매 데이터 기반 지역 클러스터링

1

우리동네 공연대장

# 목차

---

01. 연구 배경

02. 분석 프로세스

04. 분석 내용

05. 분석 결과물

06. 활용 방안 및 기대 효과

# 연구 배경

## ■ 공연시장의 성장세

구분	2022년 상반기	21년 대비 22년 증감률	20년 대비 22년 증감률
공연건수(건)	5,663	▲(48%)	▲(228%)
공연회차(회)	35,484	▲(51%)	▲(41%)
티켓판매수(건)	5,351,313	▲(109%)	▲(131%)
티켓판매액(원)	231,564,300,368	▲(98%)	▲(136%)

2022년 상반기 동향 분석 보고서(KOPIS 제공)

→ 공연시장이 회복세를 넘어 성장세를 보이고 있음

## ■ 서울과 그외 지역 간 공연시장 불균형 심화

### 공급 및 소비 편중 현상

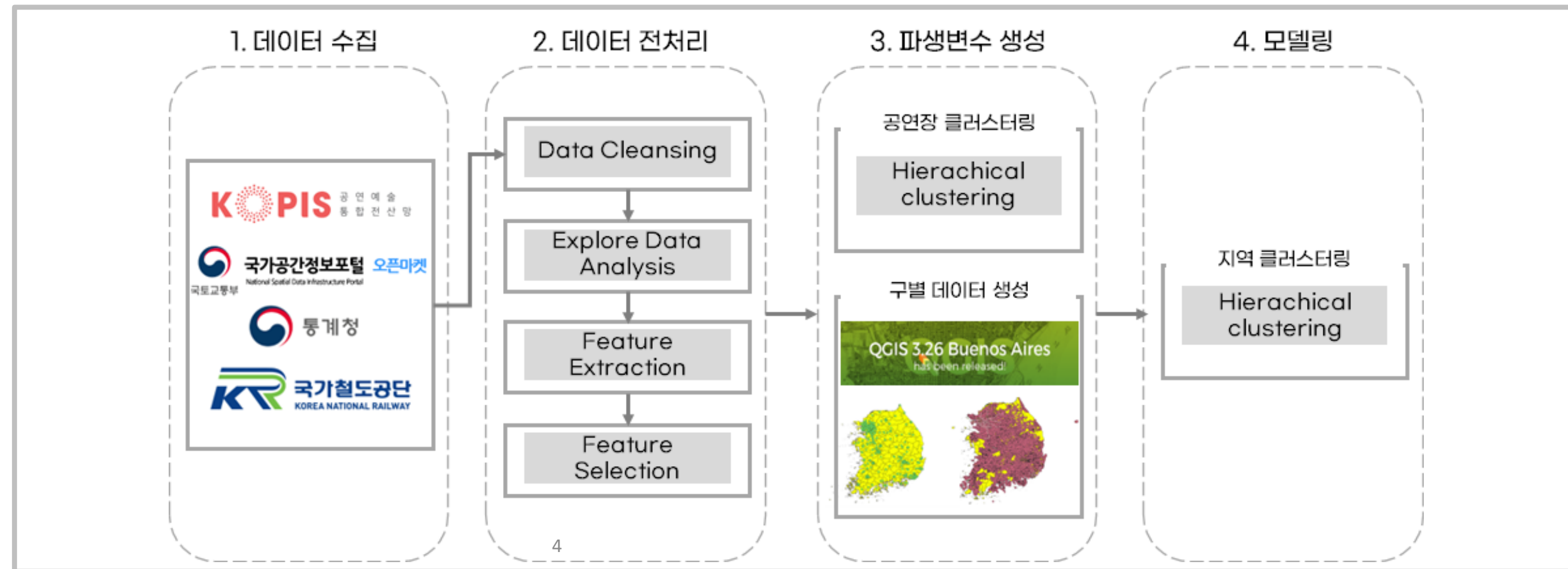
공연실적	서울의 비율
공연건수	약 47%
티켓판매수	약 68%
티켓판매액	약 73%

- 뮤지컬을 제외한 모든 장르에서 서울의 공연건수가 전체의 50% 이상 차지
- 22년 8월 기준, 서울에서 활동하는 예술인 비율 37%

데이터에 기반한,  
서울과 지방 간의 공연시장 불균형 해소 방안 모색 필요

# 분석 프로세스

## ■ 전체 프로세스

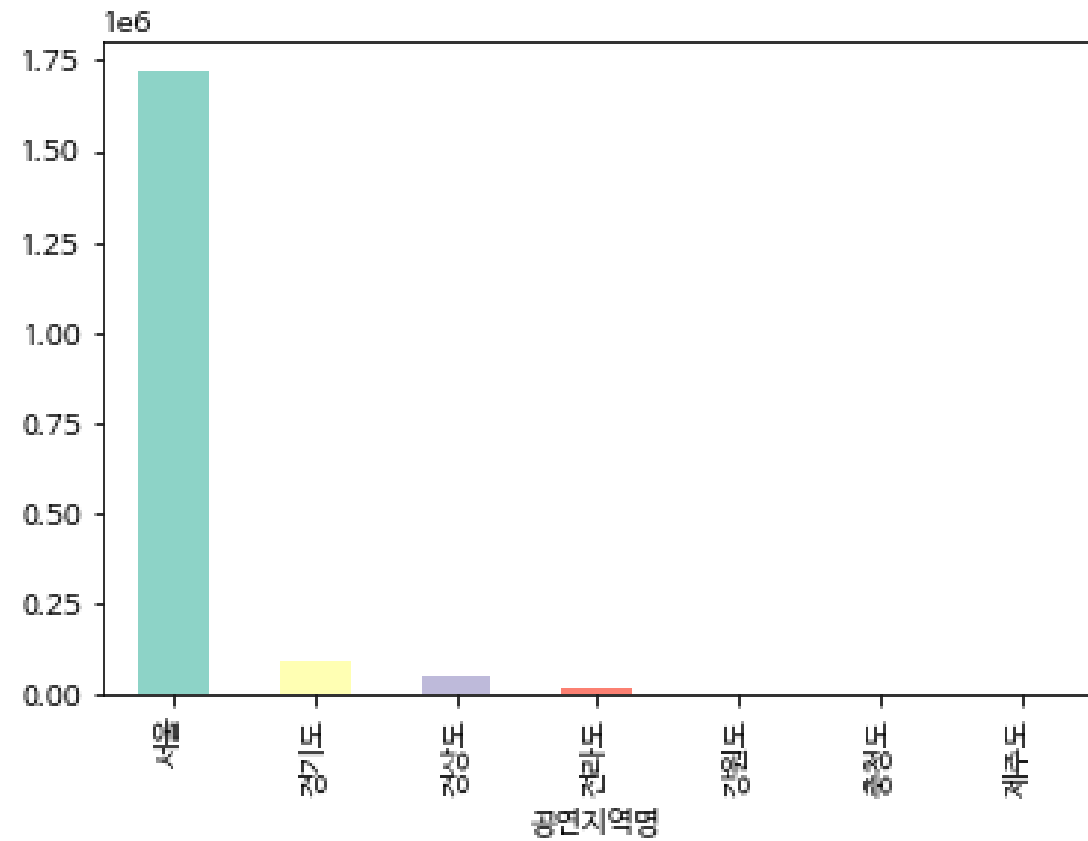


## ■ 활용 데이터

- 인구 데이터
  - : 성별·연령별 인구수, 경제활동참여율(%)
- 교통 데이터
  - : 지하철역 개수 + 버스정류장 개수
- 시군구 경계 데이터
  - : 공연장·교통 위치 정보 → 시군구코드
- KOPIS 데이터
  - : 공연장 관련 변수

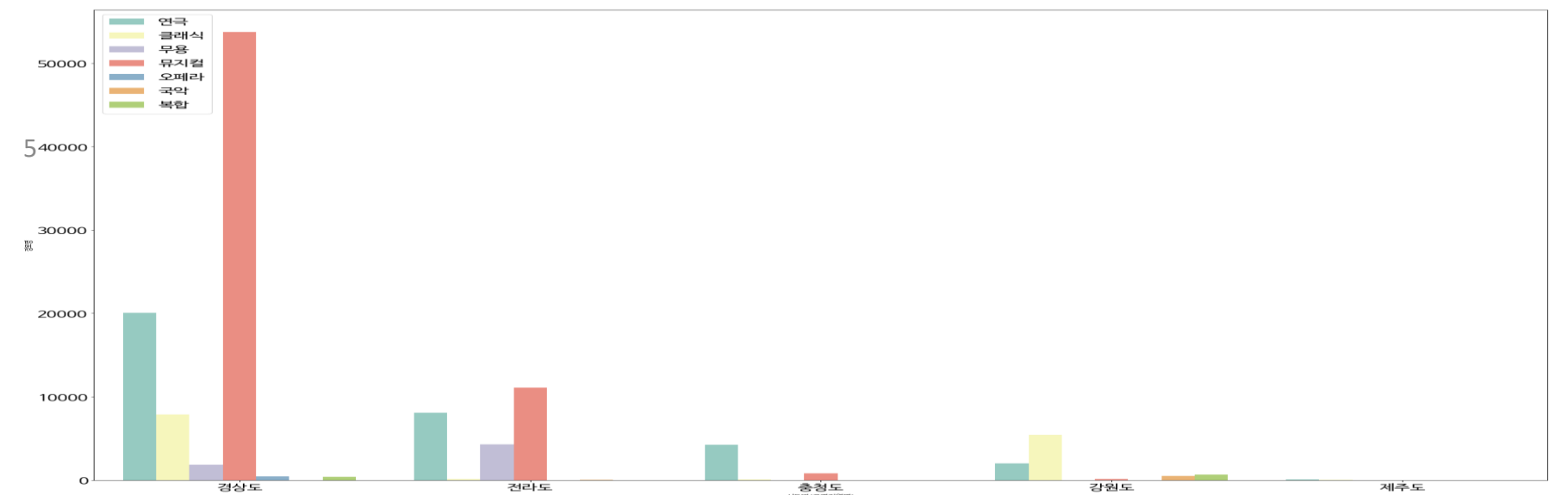
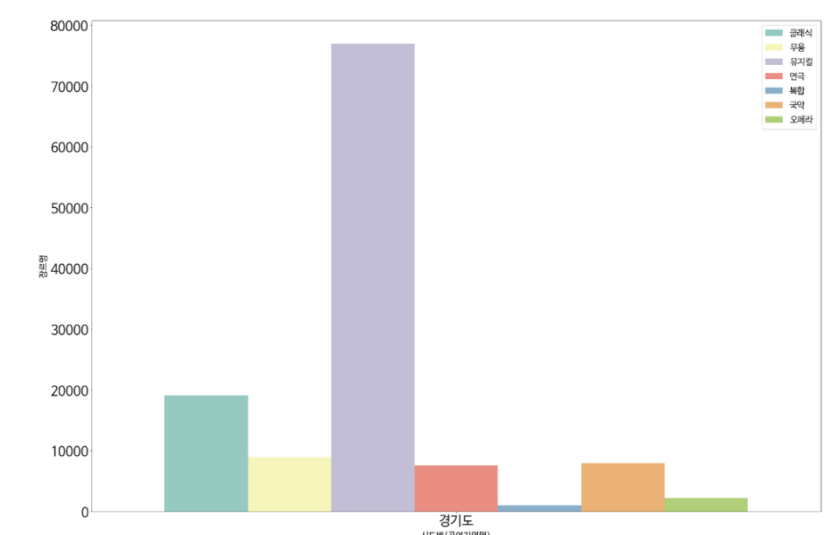
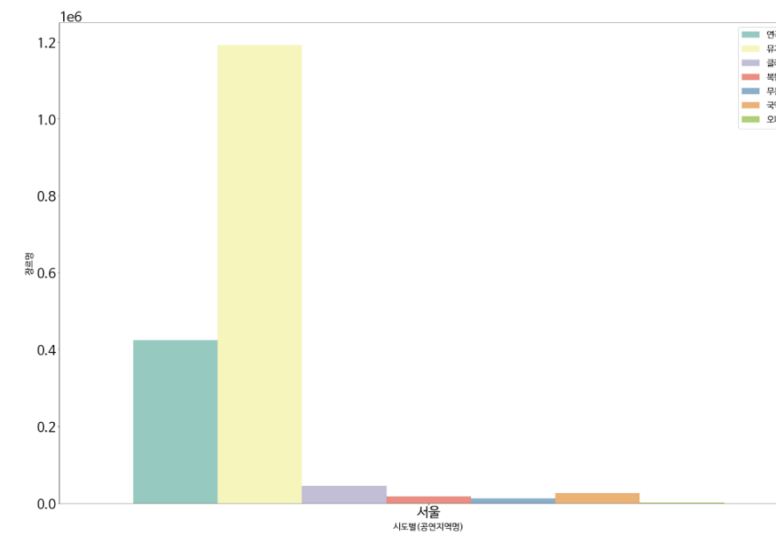
# 데이터 탐색

## ■ 개요



- 서울과 다른 지역의 공연장수, 공연횟수 차이가 큼  
→ 서울 / 서울 외 지역 그래프를 나눠서 탐색함

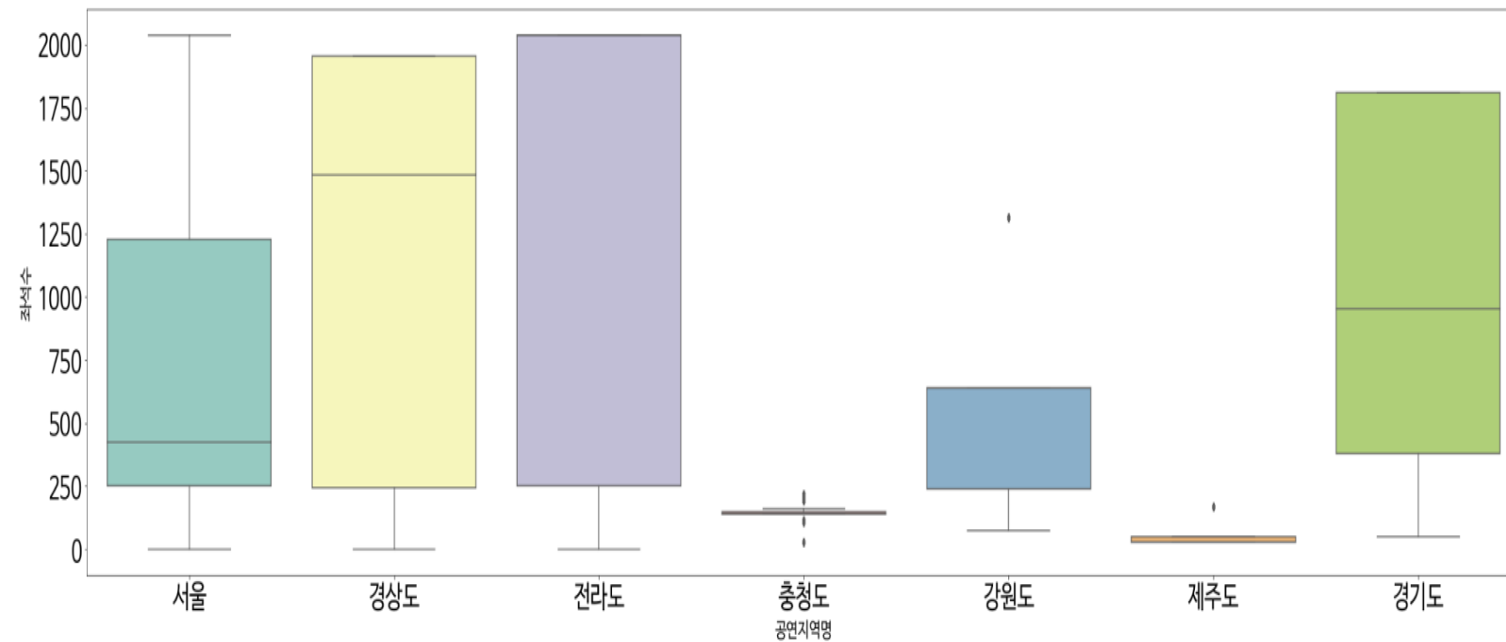
## ■ 장르명



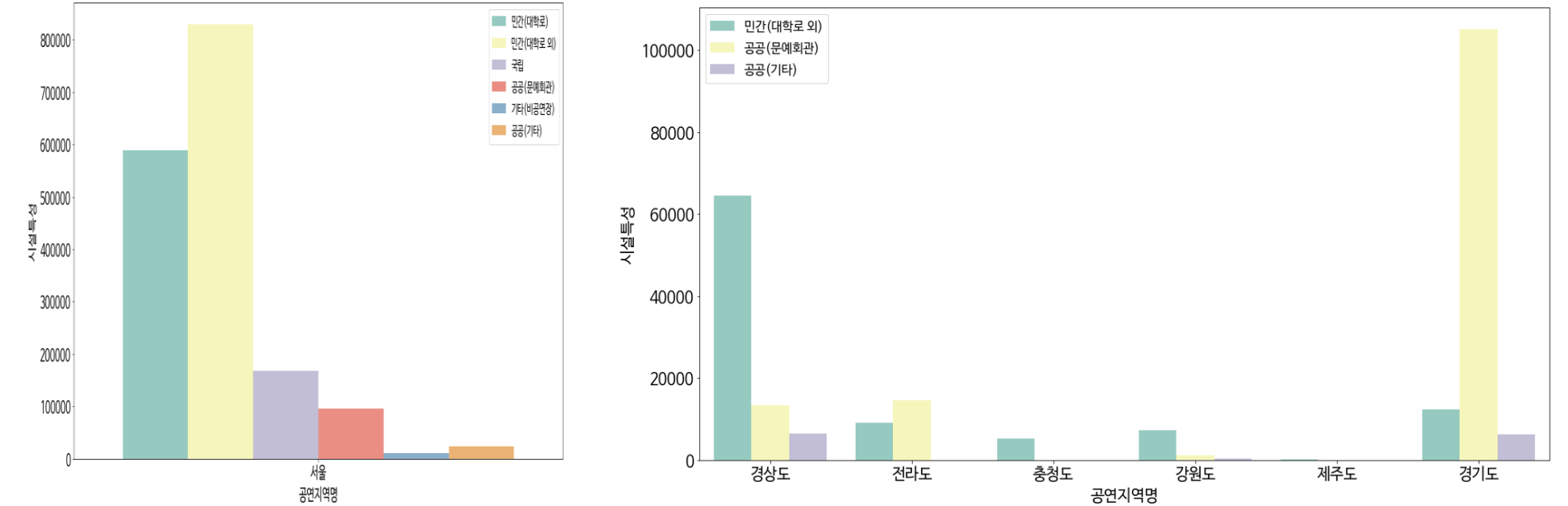
- 서울 : 뮤지컬 비율이 압도적으로 많았음
- 경기도·경상도, 전라도·충청도·강원도, 제주도에서 유사한 분포를 보임

# 데이터 탐색

## 좌석수



## 시설특성



6

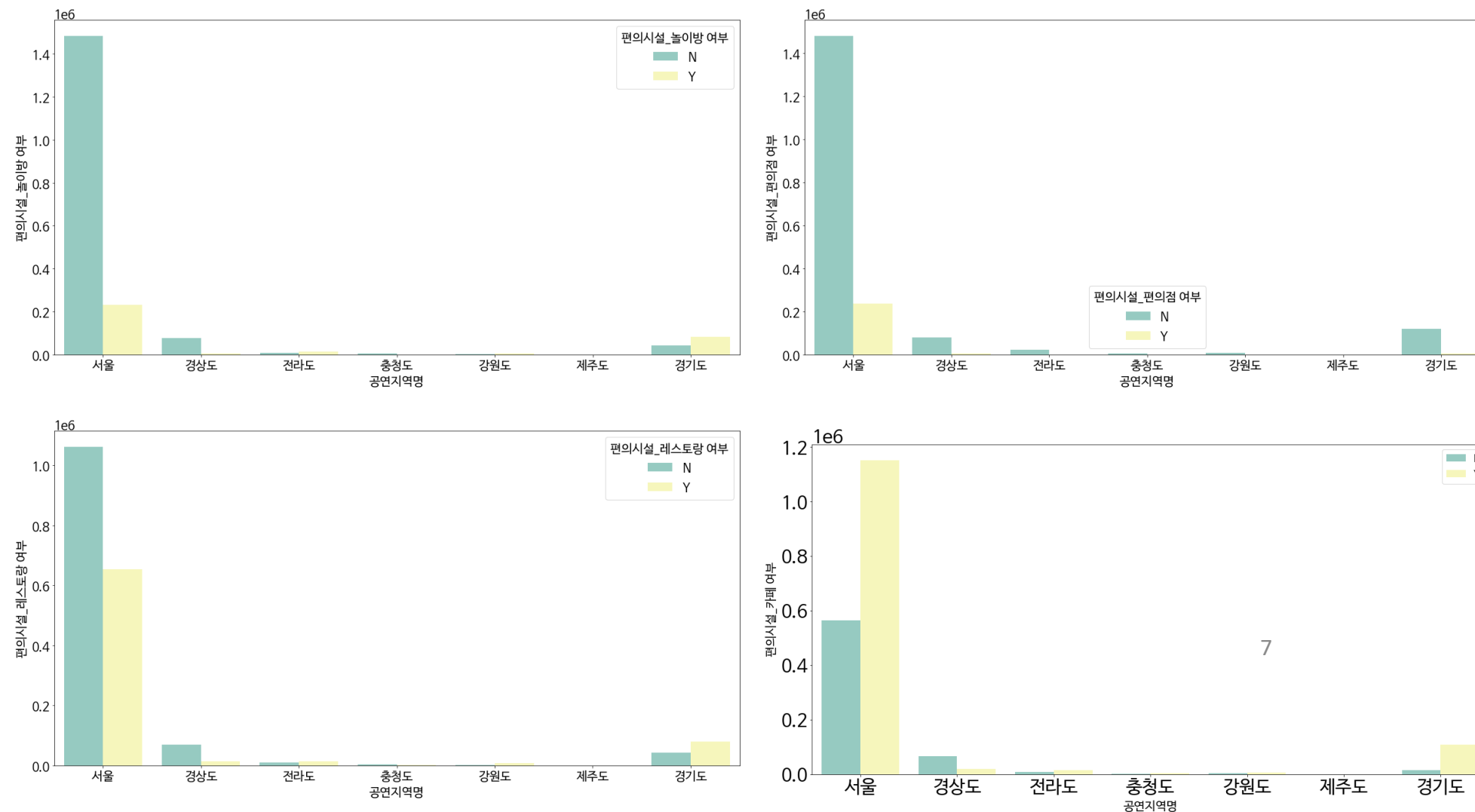
- 전라도 : 좌석수 분포가 넓음
- 제주도 : 평균 규모 공연장, 아주 작은 공연장
- 충청도 : 규모가 매우 차이나게 크거나 작은 공연장 존재

좌석수  
분포 균일

- 서울 : 민간(대학로 외) > 민간(대학로) > 공공(문화예관)
- 경기도 : 대부분 공공시설
- 경상도 : 민간시설 > 공공시설 (2.5배)
- 전라도 : 민간시설 > 공공시설 (미세한 차이)
- 충청도 : 대부분 민간시설
- 제주도 : 민간시설 X

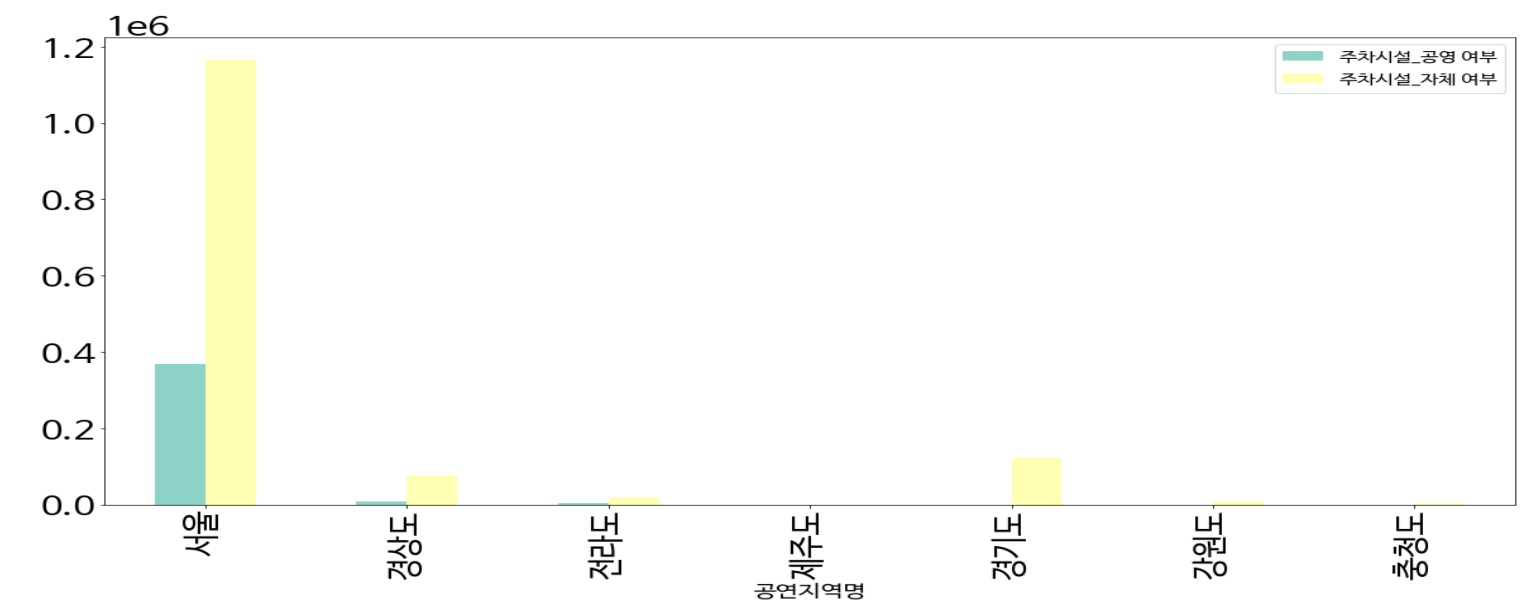
# 데이터 탐색

## ■ 편의시설여부



- 서울: 값이 눈에 띄게 큼  
→ 서울의 공연장 수가 다른 지역을 압도하기 때문
- 서울 외 지역: 편의시설끼리 유의미한 경향성 X  
→ 시도라는 큰 단위로 분석했기 때문이라고 판단

## ■ 주차시설여부

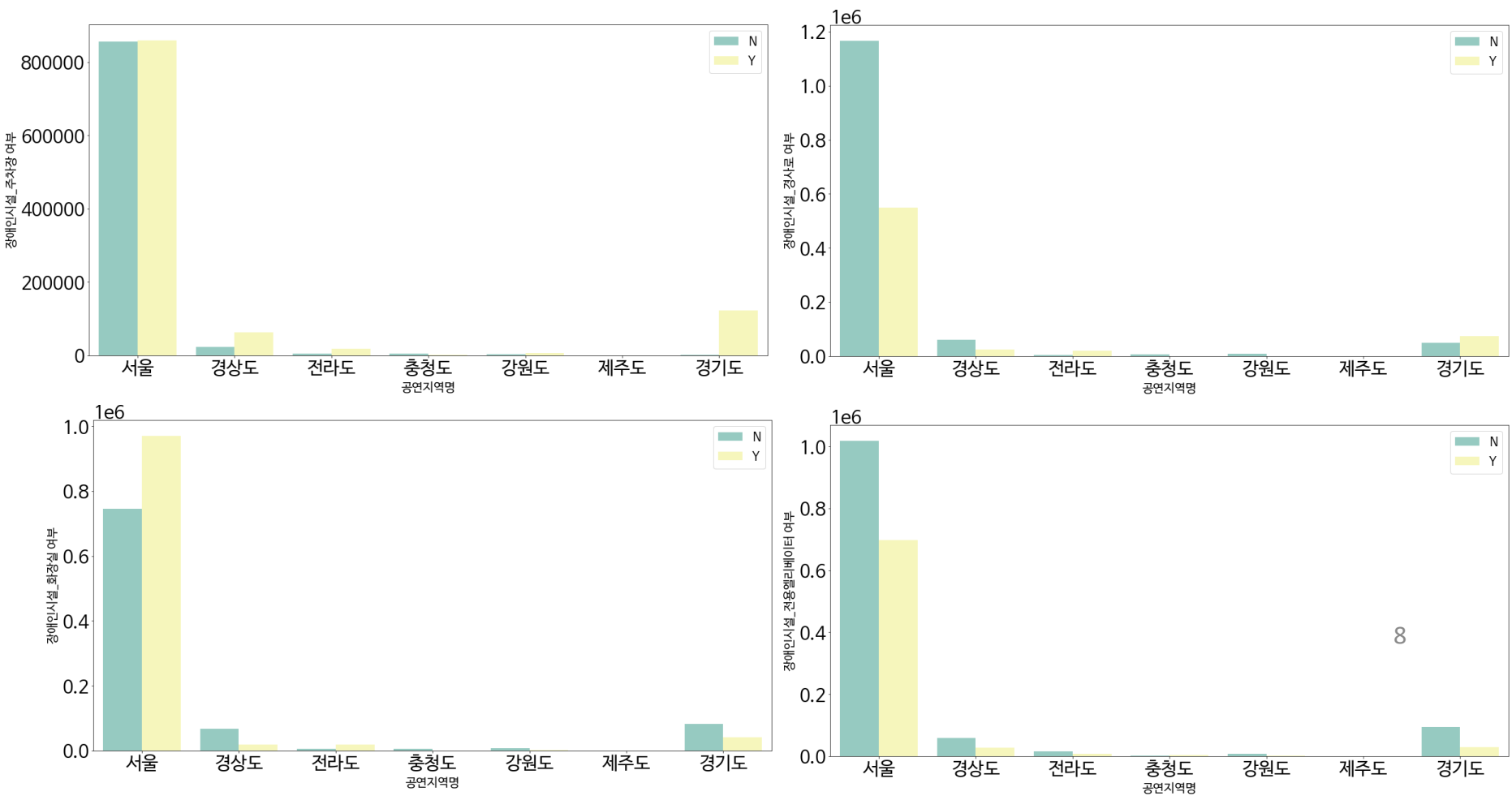


- 분포 : 주차시설\_자체 여부 ↔ 주차시설\_공영 여부
- 편의시설 여부 변수들과 유사한 결과

분석 내용

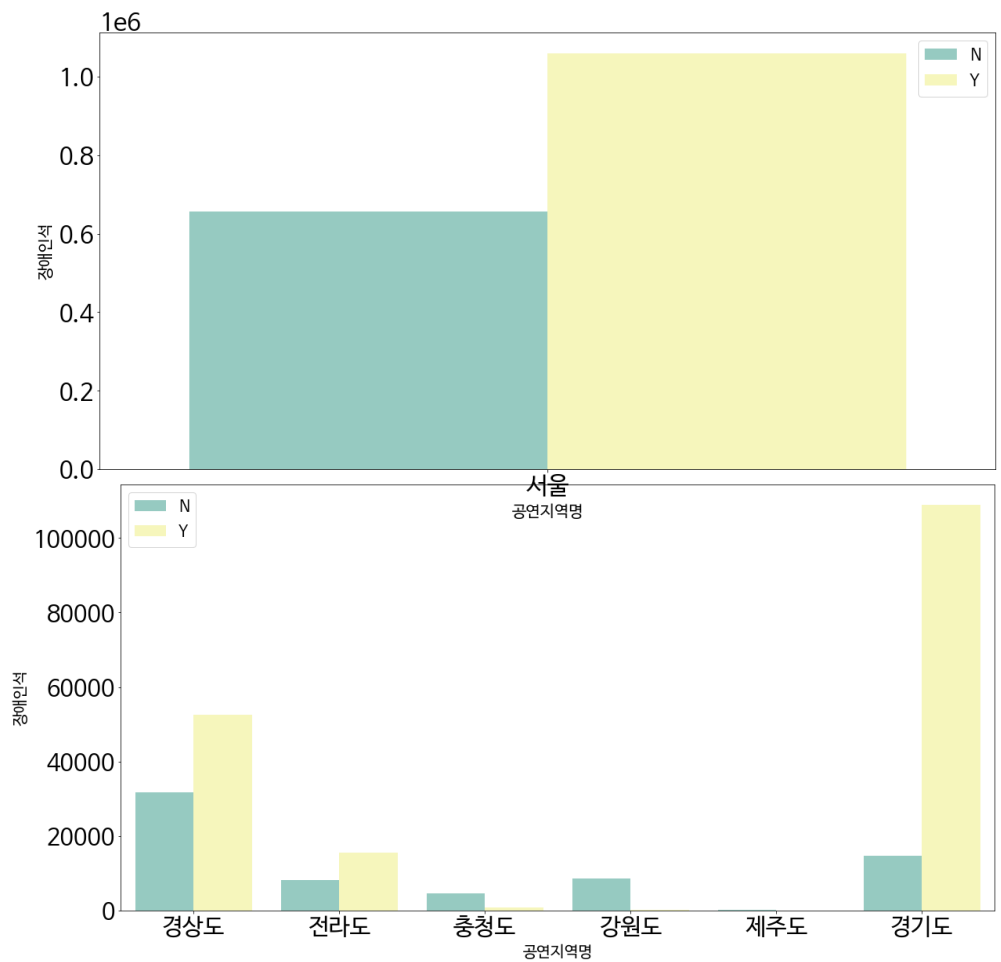
# 데이터 탐색

■ 장애인시설여부



- 편의시설 여부 변수들과 유사한 결과를 보임

■ 장애인석

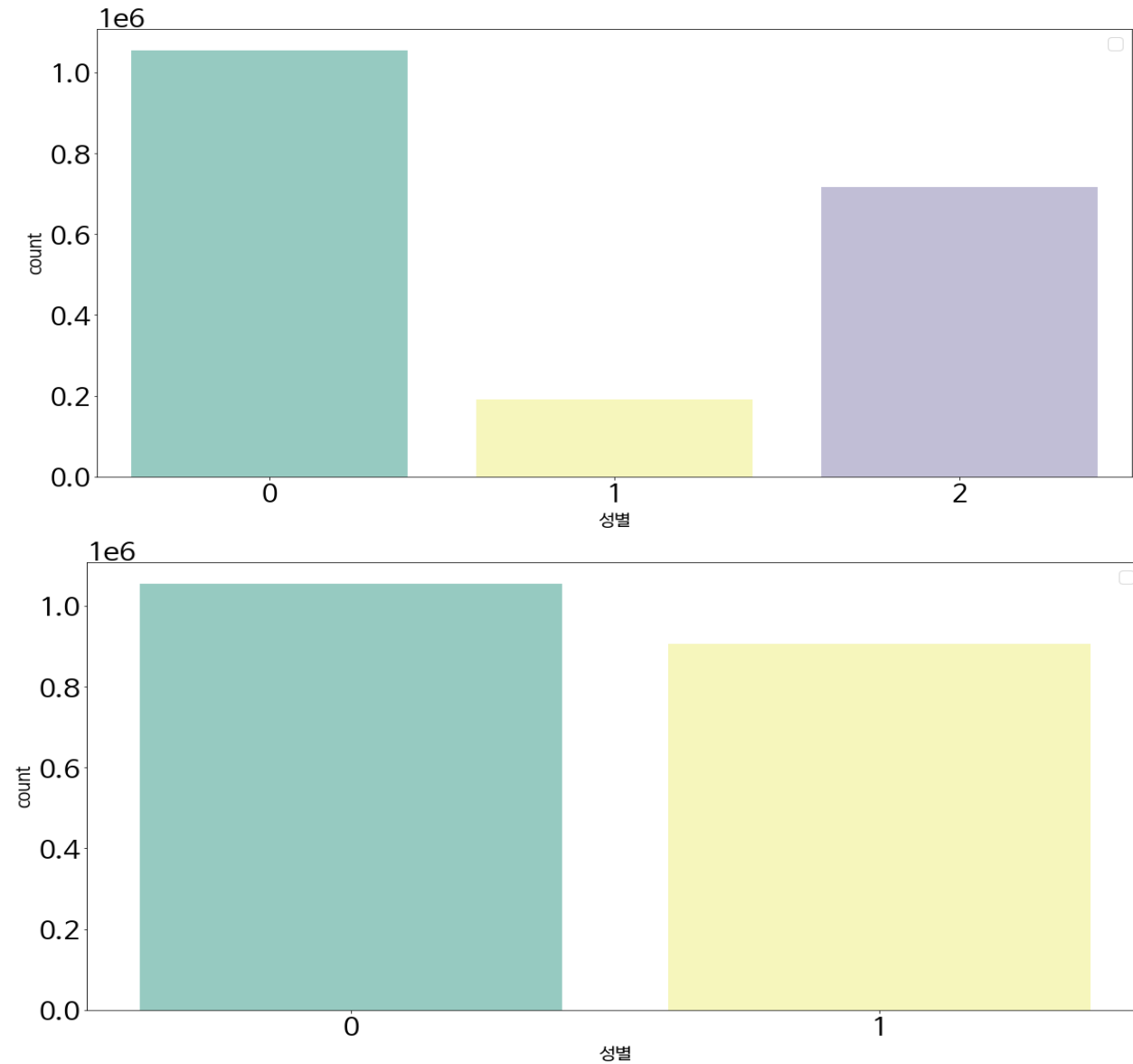


- 대부분의 지역이 장애인석을 갖춤
- 충청도, 강원도 : 장애인석 X > 장애인석 O



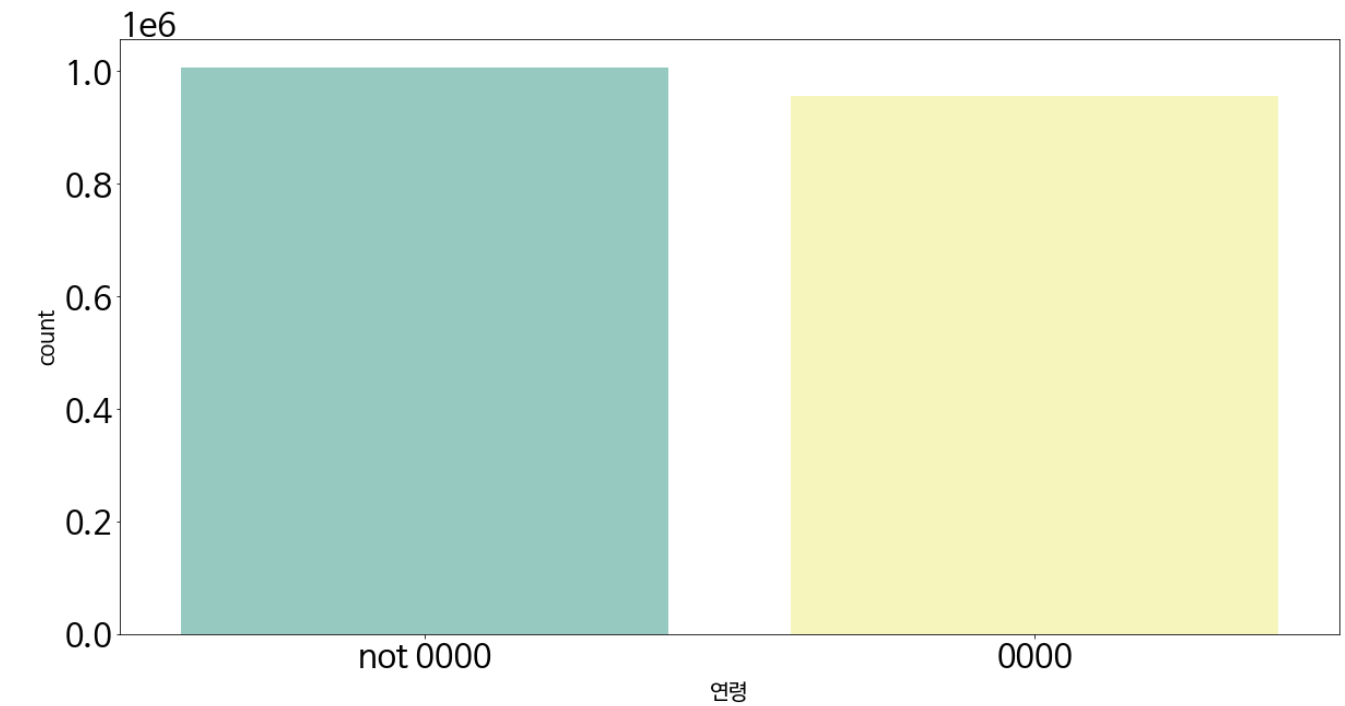
# 데이터 탐색

■ 성별



- 성별의 값이 0 → 결측치를 의미
- 결측치 비율 多 ⇒ 성별 변수 삭제

■ 연령



- 연령의 값이 0000 → 결측치를 의미
- 결측치 비율 多 ⇒ 연령 변수 삭제

# 데이터 전처리(인구통계)

■ 행정구역명 → 시군구 행정동코드

시도	시군구	시군구코드
경기도	광주시	31250
경기도	양주시	31260
...	...	...
경상남도	의창구	38111
...	...	...

■ 연령대별 인구수

- 70세 이상 인구수 통합
- 각 연령대별 인구수 ÷ 총 인구수

■ 성별 인구수

- 여성인구수 ÷ 총 인구수

인구통계 테이블

10										
시군구코드	경제활동참가율	여성인구비율	10세이하	10대	20대	30대	40대	50대	60대	70세이상
21010	0.567	0.513163	0.026424	0.045211	0.145866	0.110938	0.123644	0.165072	0.194710	0.188135
21030	0.520	0.511733	0.044547	0.054265	0.115851	0.120040	0.131313	0.162340	0.183341	0.188302
21050	0.550	0.518734	0.054923	0.067980	0.143892	0.132718	0.141035	0.159245	0.163243	0.136963
21060	0.558	0.515084	0.072585	0.092065	0.113450	0.120321	0.160614	0.160606	0.155451	0.124908
21070	0.545	0.511602	0.060696	0.086873	0.129239	0.105334	0.153472	0.161284	0.159124	0.143978
21080	0.541	0.506903	0.062928	0.078763	0.121821	0.119043	0.139855	0.183176	0.175605	0.118808
21110	0.524	0.511261	0.050760	0.075228	0.137889	0.104433	0.136069	0.170507	0.176203	0.148912

# 데이터 전처리(시군구 행정동코드)

GEOSERVICE-WEB

도로명주소 → 위·경도 위치 정보 산출 & 결측치 처리

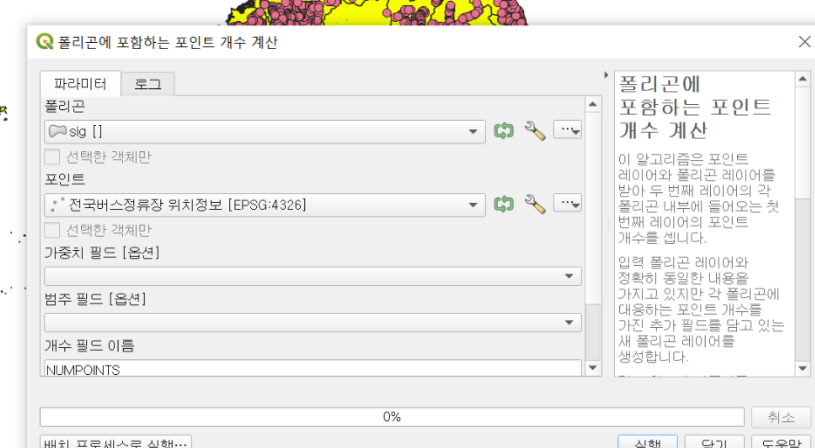
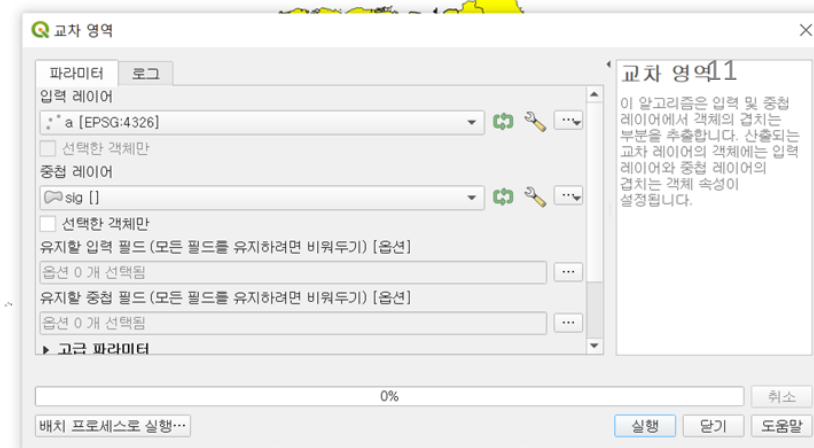
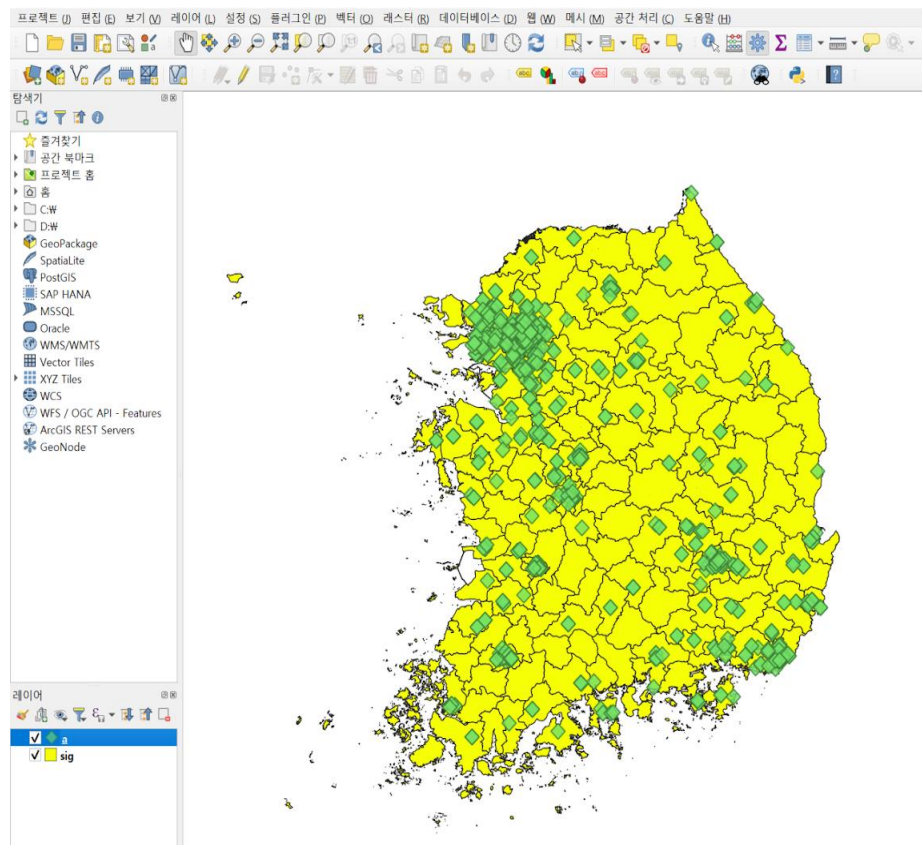
공연장

QGIS

교통

- 시군구경계 shp 파일 → 시군구 레이어 추가
- 공연장위치 레이어 중첩
- 두 레이어의 교차 영역 계산 ⇒ 시군구 행정동코드 컬럼 추가

- 시군구경계 shp 파일 → 시군구 레이어 폴리곤 설정
- 도시철도·버스정류장 위치 레이어 중첩
- 위경도 결측치 처리 및 위치 포인트 교차 영역 계산



시군구별 대중교통 수 ÷ 면적

대중교통밀도	
시군구코드	
21010	1.245028
21030	0.462196
21050	0.564682
21060	1.203111
21070	0.043807

# 데이터 전처리(KOPIS)

- 편의시설 · 장애인시설 · 주차시설 여부 변수 : Y, N → 1, 0
- '노후화정도' : 현재 연도(2022) - '개관연도'

- 무대시설 여부 : 결측치 多
- 공연 관련 변수 : 시군구 단위 결합 불가

⇒ 미사용

## 클러스터링

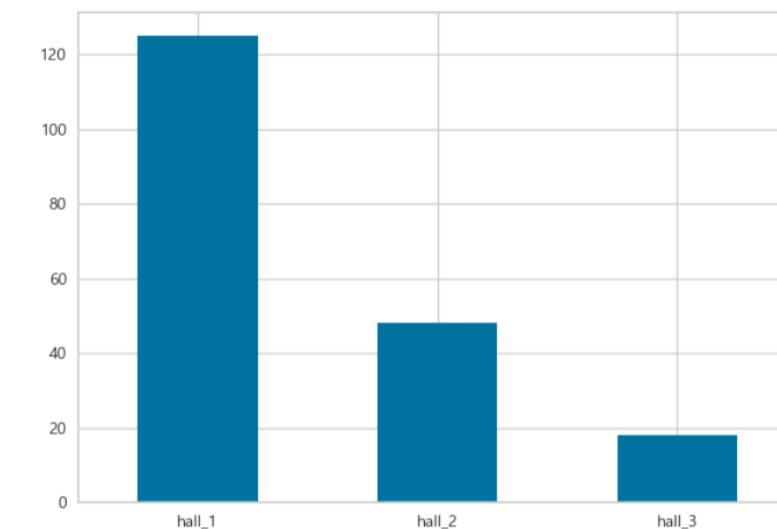
### ■ 공연장 특성 파생변수 생성

- 구별로 결합된 데이터  
→ 공연장 관련 변수 모두 고려 시 좌우로 긴 테이블 생성
- 공연장 관련 변수로 공연장 특성을 그룹화하는 파생변수를 정의

12

## 공연장 테이블

	hall_1	hall_2	hall_3
구코드			
21010	1	0	0
21030	1	0	0
21050	1	0	0
21060	0	1	0
21070	9	0	0



총 3개의 집단

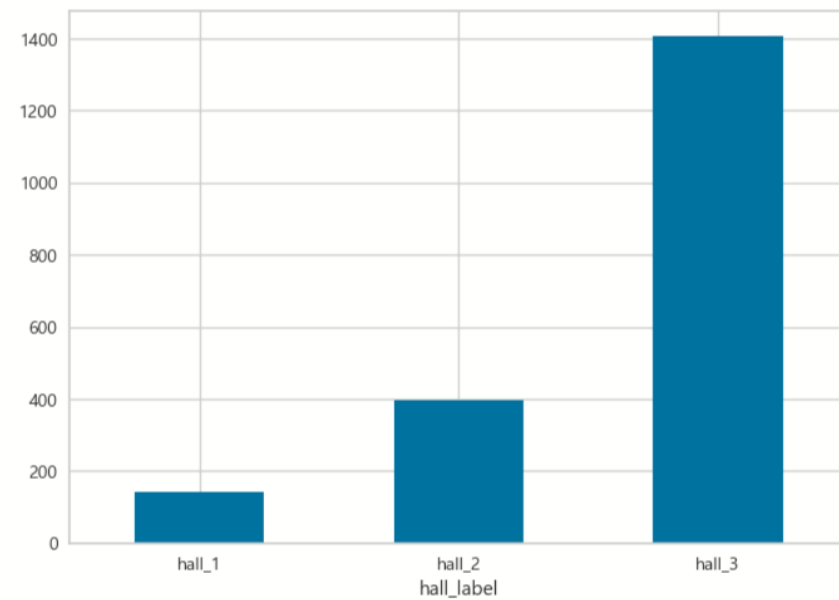
- 공연장 특성과 교통, 인구통계를 반영한 지역 클러스터링
  - 서울 외 지역
  - 파생변수 'hall\_1', 'hall\_2', 'hall\_3' : 공연장 특성 군집에 해당되는 공연장의 수
  - 교통, 인구통계 데이터 결합

최종 활용 데이터

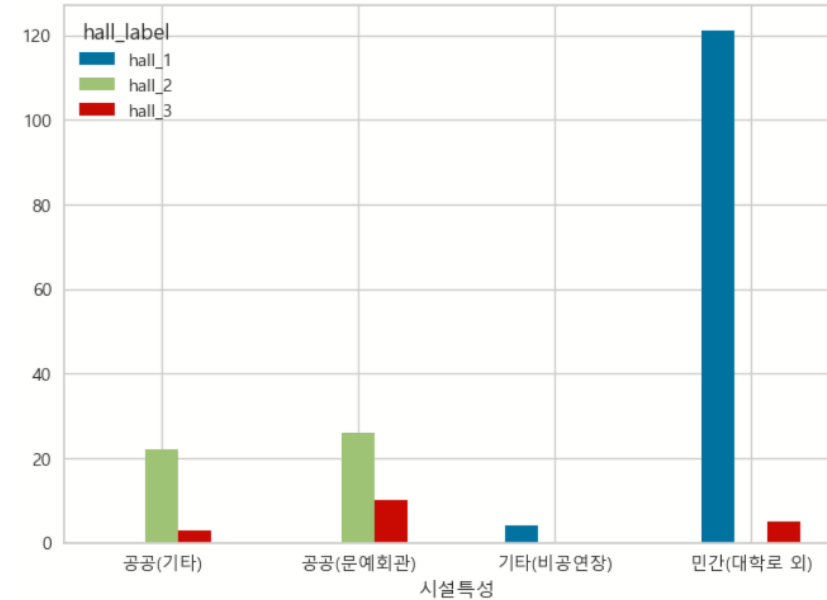
공연장 파생변수				교통	인구통계									
시군구코드	hall_1	hall_2	hall_3	대중교통밀도	경제활동참가율	여성인구비율	10세이하	10대	20대	30대	40대	50대	60대	70세이상
21010	1	0	0	1.245028	0.567	0.513163	0.026424	0.045211	0.145866	0.110938	0.123644	0.165072	0.194710	0.188135
21030	1	0	0	0.462196	0.520	0.511733	0.044547	0.054265	0.115851	0.120040	0.131313	0.162340	0.183341	0.188302
21050	1	0	0	0.564682	0.550	0.518734	0.054923	0.067980	0.143892	0.132718	0.141035	0.159245	0.163243	0.136963
21060	0	1	0	1.203111	0.558	0.515084	0.072585	0.092065	0.113450	0.120321	0.160614	0.160606	0.155451	0.124908
21070	9	0	0	0.043807	0.545	0.511602	0.060696	0.086873	0.129239	0.105334	0.153472	0.161284	0.159124	0.143978
21080	0	1	0	0.728389	0.541	0.506903	0.062928	0.078763	0.121821	0.119043	0.139855	0.183176	0.175605	0.118808
21110	3	1	0	2.031002	0.524	0.511261	0.050760	0.075228	0.137889	0.104433	0.136069	0.170507	0.176203	0.148912

# 공연 클러스터링

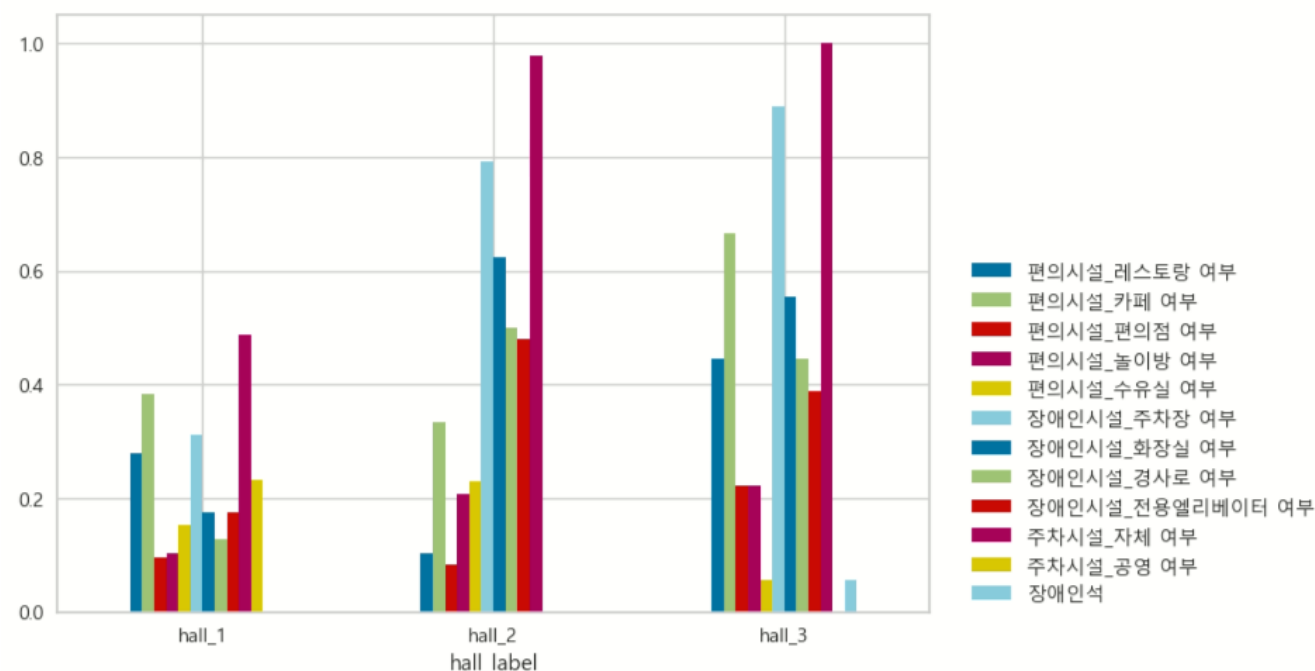
## 좌석수



## 시설특성



## 편의·주차·장애인 시설



14

## 집단별 분석

### Hall 1

- 민간시설이 대부분
  - 좌석수가 가장 적음
  - 전반적인 시설 부족
- ⇒ 소규모 민간공연장, 시설은 주변 상권에 의존

### Hall 2

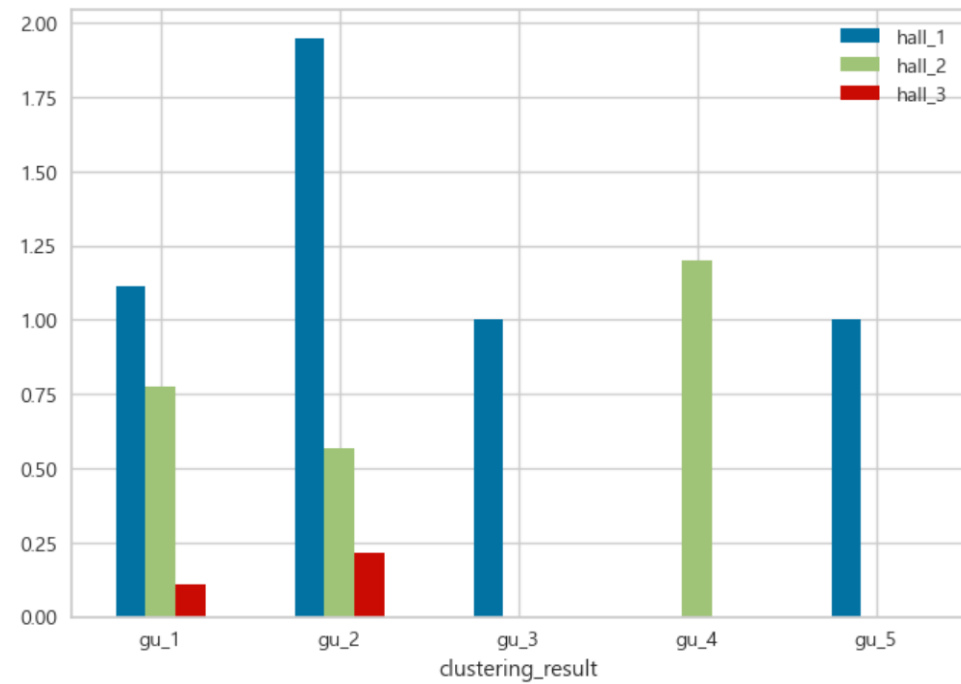
- 공공 문예회관·기타 균일 분포
  - 대부분 자체 주차시설
  - 자체 수유실·놀이방이 다른 집단에 비해 많음
- ⇒ 공공기관 특성

### Hall 3

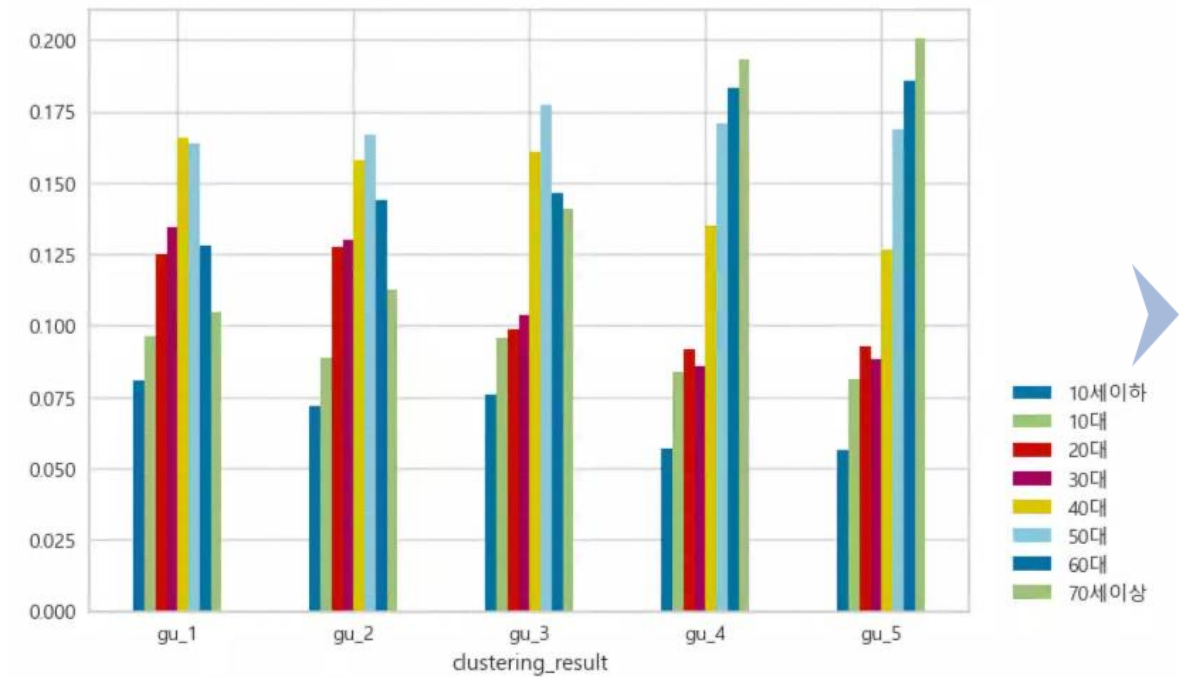
- 시설특성 X & 좌석수 압도적
  - 자체 레스토랑·카페가 다른 집단에 비해 많음
  - 장애인석 비율이 눈에 띄
- ⇒ 대규모 공연장 특성

# 지역 클러스터링

■ 공연장 파생변수

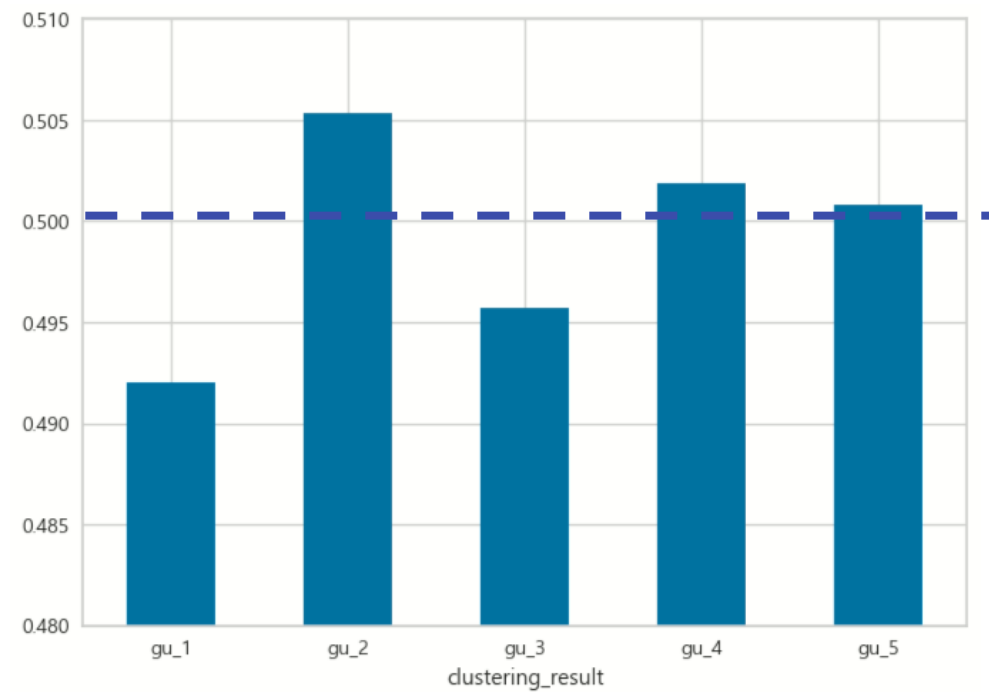


■ 연령대별 인구수

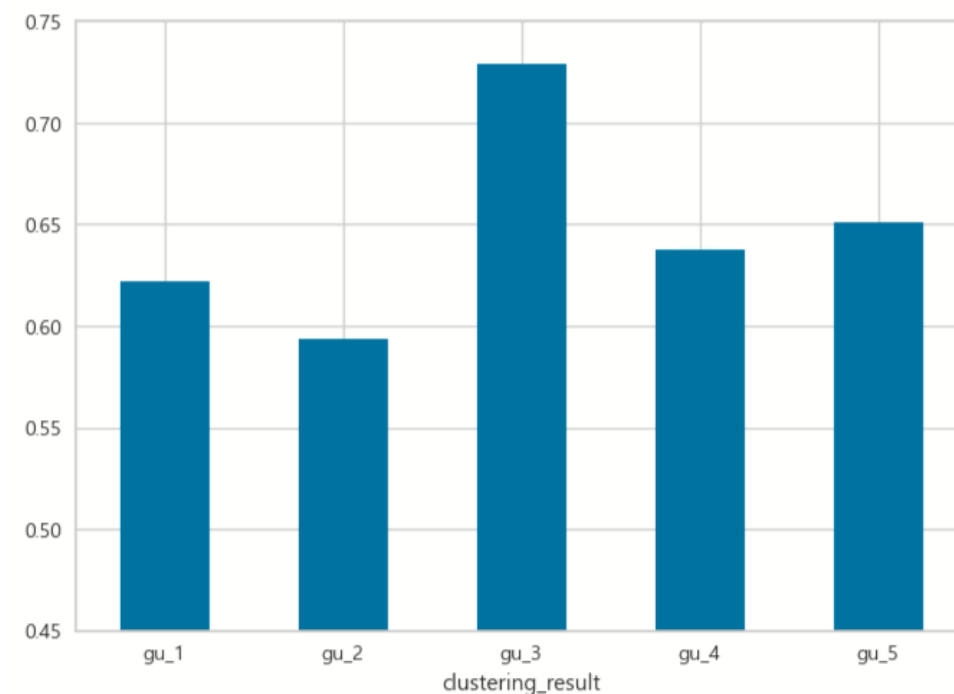


- gu\_1 : 30대 이하의 비율은 높고 60대 이상의 비율은 낮은, 젊은 연령대의 집단
- gu\_2 ≡ gu\_1 : 20대 비율이 높음
- gu\_3 : 40~50대 비율이 높음
- gu\_4 ≡ gu\_5 : 연령대가 높은 집단

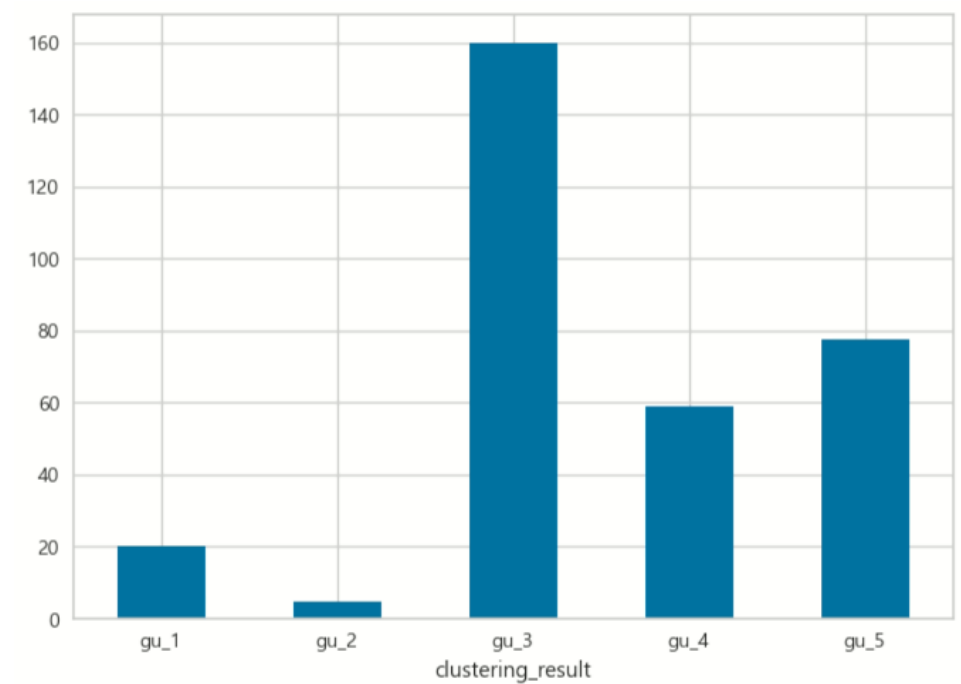
■ 여성인구비율



■ 경제활동참가율



■ 대중교통밀도





# 지역 클러스터링

## ■ 집단별 분석

\* Hall 1: 자체 보유 시설이 부족한 소규모 민간 공연장  
Hall 2 : 평균적인 크기의 공공 공연장  
Hall 3 : 자체적으로 카페·레스토랑을 가지는 대규모 공연장

### 구 집단 1

- Hall 1·2·3 모두 존재 가능하지만 소규모 민간 공연장과 공공 공연장이 각 1개씩 존재할 확률이 큼
- 2-40대 비율 가장 큼 → 가장 젊은 집단
- 남성인구 > 여성인구

### 구 집단 2

- 집단 1과 유사하지만 소규모 민간 공연장이 더 많음
- 대규모 공연장 존재 확률이 가장 큼
- 20대 비율 & 여성 비율 가장 큼
- 지역의 면적 대비 대중교통 수 가장 적음

16

### 구 집단 3

- 소규모 민간 공연장만 1개씩 존재
- 남성인구 > 여성인구
- 4,50대 비율 & 경제활동참가율 가장 큼
- 대중교통 밀집 지역

### 구 집단 4

- 공공 공연장만 1개씩 존재
- 60세 이상 비율 두 번째로 큼
- 여성인구 > 남성인구

### 구 집단 5

- 소규모 민간 공연장만 1개씩 존재
- 연령대 가장 높음
- 여성인구 > 남성인구

차이점



# 활용방안 및 기대효과

## ■ 활용방안



지역문화재단

유연한 정책 필요



공연 공급자

지역 특성 고려한

공연 유치 필요

공연 지역 클러스터링

- 지역 특성 기반의 적실성 높은 정책 연구 개발 가능
- 지역 특성 기반 공연 기획 및 유치 가능

## ■ 기대효과

시군구 특성 → 지역 맞춤형 공연 활성화 정책 제언 가능



- '공연예술 지원 및 투자 → 공연 공급 → 공연 관계자 소득 및 상권 활성화 → 공연 관람 → 공연 추가적 공급 → 추가적 지원 및 투자' 순환 유도
- 고정 관객 유치

# 참고문헌

---

- [1] 배수민. (2021). 서울에 집중된 문화예술 인프라, 지역 간 문화 격차 문제.  
<https://www.iworldtoday.com/news/articleView.html?idxno=406359>
- [2] 송광호. (2019). [디지털스토리] '지방러'는 문화생활 한번 하기도 힘드네요.  
<https://www.yna.co.kr/view/AKR20190517043800797>
- [3] 한국예술인복지재단. (2022). 예술활동증명현황.  
[https://www.kawfartist.kr/kawf\\_dash/dash/html/p01.html](https://www.kawfartist.kr/kawf_dash/dash/html/p01.html)
- [4] 김담희, 안가경. (2018). 머신러닝을 이용한 고객세분화에 관한 연구. 융복합지식학회논문지, 6(2), 115-120.
- [5] 송우영. (2019). 지역별 공연예술 수요지수개발에 대한 연구. 한국지식정보기술학회 논문지, 14(6), 635-643.
- [6] 함윤주, 김연준, 홍근석. (2020). 지방자치단체 문화시설 수요추정 개선방안 연구: 문화예술회관을 중심으로. 한국정책과학학회보, 24(2), 121-163.
- [7] 김하연, 서대호. (2019). 토픽 모델링 기반 국내 공연 흥행 요인 분석. 한국무용과학회지, 36(3), 99-111.
- [8] 주민등록 인구통계 성/연령별 인구 데이터.
- [9] 시군구 경제활동인구 총괄 데이터.
- [10] 전국도시철도역사정보표준데이터(국가철도공단 제공).

# 참고문헌

---

[11] 전국 버스정류장 위치정보(국토교통부 제공).

[12] (센서스경계) 시군구경계 - 국가공간정보포털 제공.

[13] KOPIS. (연도미상). KOPIS 공연시설DB.

[https://www.kopis.or.kr/por/db/prfplc/prfplc.do?menuId=MNU\\_00021&searchWord=&searchType=total](https://www.kopis.or.kr/por/db/prfplc/prfplc.do?menuId=MNU_00021&searchWord=&searchType=total)

[14] 임성복. (2003). 지역자금의 역외유출 방지방안. 지역자금의 역외유출 방지방안 연구 과제.

[15] 이수민.(2020). [기획] 예술극장, 관객을 이끄는 전략은?.

<http://www.ithemove.com/news/articleView.html?idxno=1383>

감사합니다