

데이터사이언스(1) : 데이터통합과 정보보호

# 개인 최종 보고서

박                      준                      민  
2 0 1 5 1 2 2 0 5 4

# 목차

**01 Data Structure**

**02 Research Question**

**03 Analysis**

**04 Conclusion**

시설 명	소공연장명	시설 정보		공연 정보	예매 정보
		시설 내부 정보	시설 주변 정보		
key	key	KOPIS (www.kopis.or.kr) 제공 공연시설 excel 및 크롤링  시설 총 좌석수, 소공연장 좌석수, 시설 특성, 개관 연도, 주소	kakao map API 크롤링  시설 주변 편의점, 주차장, 주요소, 지하철, 문화 시설, 관광 명소, 숙박 업소, 음식점, 카페 총 개수, 위도, 경도	KOPIS (www.kopis.or.kr) 제공 공연 excel (2018.01.01 ~ 2019.12.31)  해당 시설에서 공연한 국악, 무용, 뮤지컬, 복합, 연극, 오페라, 클래식 공연 수  해당 시설에서 공연한 총 공연 수	KOPIS (www.kopis.or.kr) 제공 공연시설 통계 API 활용 (2018.01.01 ~ 2019.12.31)  해당 극장 총 예매 수

시설 명	소공연장	시설 정보	국악	무용	...	총공연 수	총예매 수
시설A	공연장 a	시설정보 A, a	n_a1 개	n_a2 개	...	N_a1 개	N_a2 개
	공연장 b	시설 정보 A, b	n_b1 개	n_b2 개	...	N_b1 개	N_b2 개
...	...	...	...	...	...	...	...

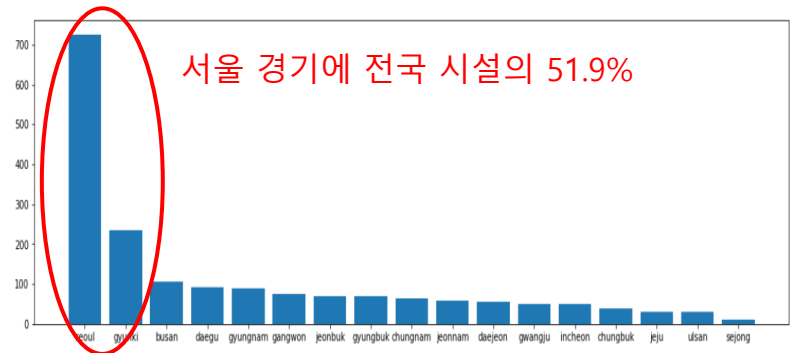
- 공연 시설 이름과 소공연장 이름을 key로 4개의 개별 data를 통합
- 공연 정보 data는 공연을 기준이던 것을 공연 시설 기준으로 재정렬
- DB 입력 과정의 오류에 따른 NA 확인 후 수정, 재결합
- 공연하지 않음으로 생긴 NA 0으로 채움

## &lt; 지역별 공연 시설 수 &gt;

	서울	경기	부산	대구	경남	강원	전북	경북	충남
city	724	234	107	92	89	75	70	70	65
	전남	대전	광주	인천	충북	제주	울산	세종	
	57	55	50	49	40	30	29	10	

## &lt; 지역별 공연 좌석 수 총 합 &gt;

city	서울	경기	강원	부산	경남	경북	대구	전북	충남
n_seats2	382960	271105	60671	52572	40517	38566	37242	34798	27310
	전남	충북	대전	광주	인천	제주	울산	세종	
	26496	20399	19752	17281	16828	12444	10565	3974	



수도권과 타 지역 간 공연 시설의 문화 격차가 존재하는 것으로 보인다.

**But, 인구 수와 인구 구조를 고려하더라도 문화 격차가 존재한다고 할 수 있을까??**

**존재한다면 그 원인은 무엇이며 해결방안은 무엇일까?**

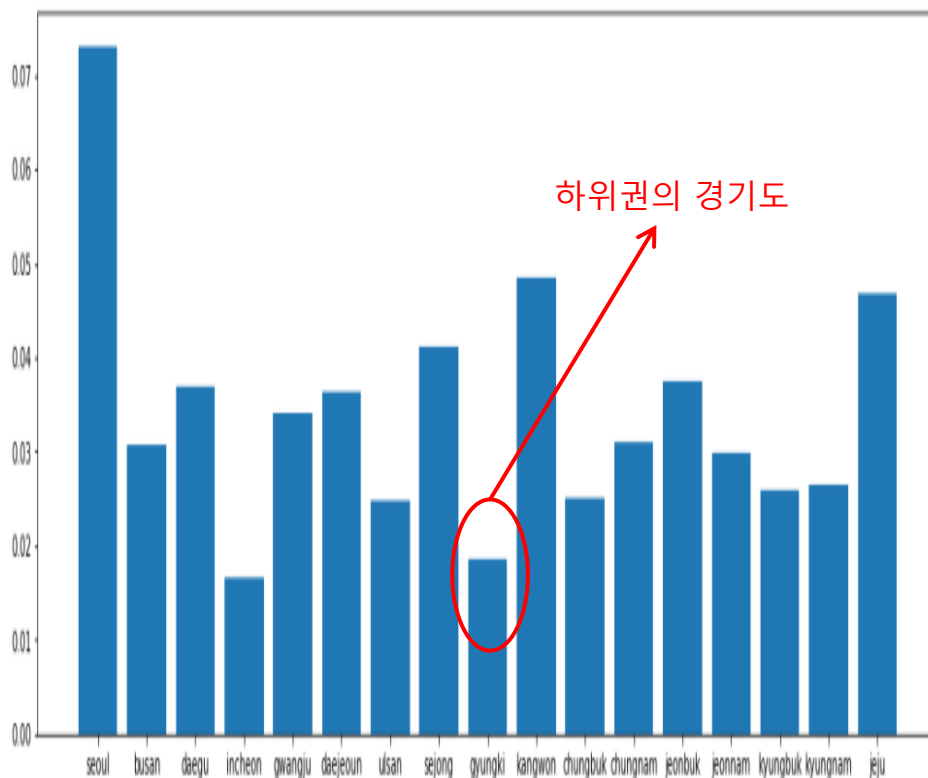
- 지역별 인구와 인구구조를 반영한 문화 격차 여부를 알아보기 위해, Integrated Data를 지역별 공연 시설 수, 좌석 수, 예매 평균을 pivot table로 정리
- 국가통계포털(KOSIS)에서 지역별 인구, 연령별 인구 data를 가져와 두 data를 통합

지역	전체 인구	성별 인구 비율	연령별 인구 비율	공연시설 정보
key 서울, 경기, 부산, 충남 등 전국 17 개 광역시, 도, 특별자 치도	국가통계포털 KOSIS (kosis.kr)  지역별 인구	국가통계포털 KOSIS (kosis.kr)  지역별 성별 인구를 지역 인수로 나눈 비율	국가통계포털 KOSIS (kosis.kr)  지역별 연령별 인구를 지역 인수로 나눈 비율  15세 미만, 15~19, ..... 60~64, 65세 이상	from Integrated data  공연시설 수, 좌석 수, 예매수를 지역별 인구 수로 나눔

- 지역별 인구 구조를 알아보기 위해 성별 연령별 인구 수를 지역별 인구로 나누어 비율 구함
- 인구 수에 따른 공연 시설 수 등을 반영 하기 위해 지역별 인구 수로 나눔

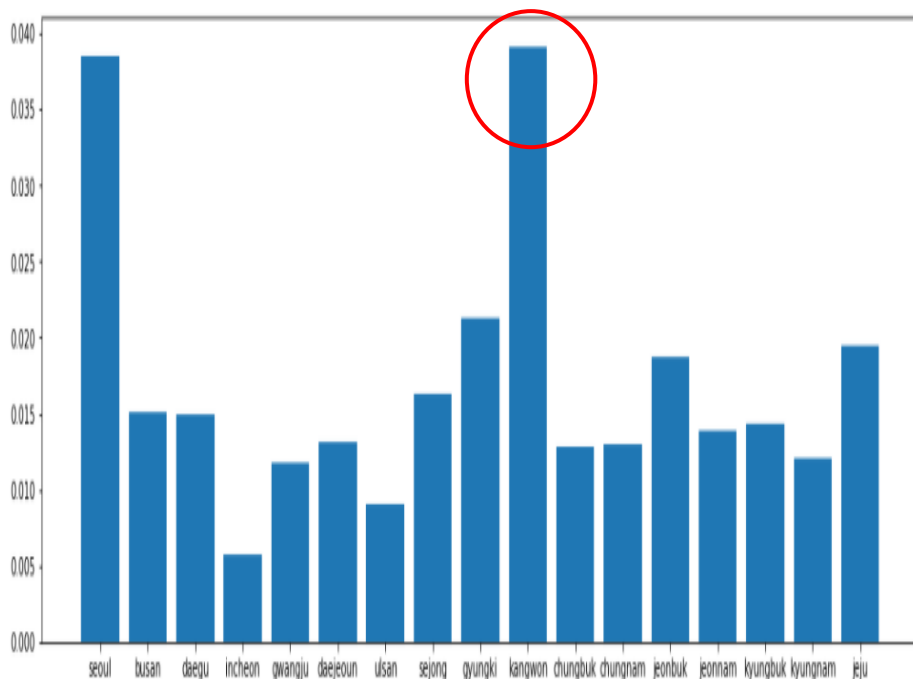
< 분석을 위한 추가 데이터 통합 및 structure 변경 >

&lt; 지역별 공연 시설 수 / 지역별 인구 &gt;



- 서울과 경기도에 전국 공연 시설 수의 50% 이상이 분포해 있다. 하지만, 인구수 대비 공연 시설 수의 비율의 경우, 지역별 격차가 줄어든 것을 확인
- 특히 경기도의 인구 대비 공연 시설 수는 최하위권
- 인천의 경우, 공연 시설 수, 인구대비 공연 시설 수 모두 하위권
- 특이하게 강원도의 경우, 인구 대비 시설 수 상위 2번째 인 것을 확인
- 다만 서울의 경우, 인구 대비 공연 시설 수가 여전히 현저히 높은 것을 확인

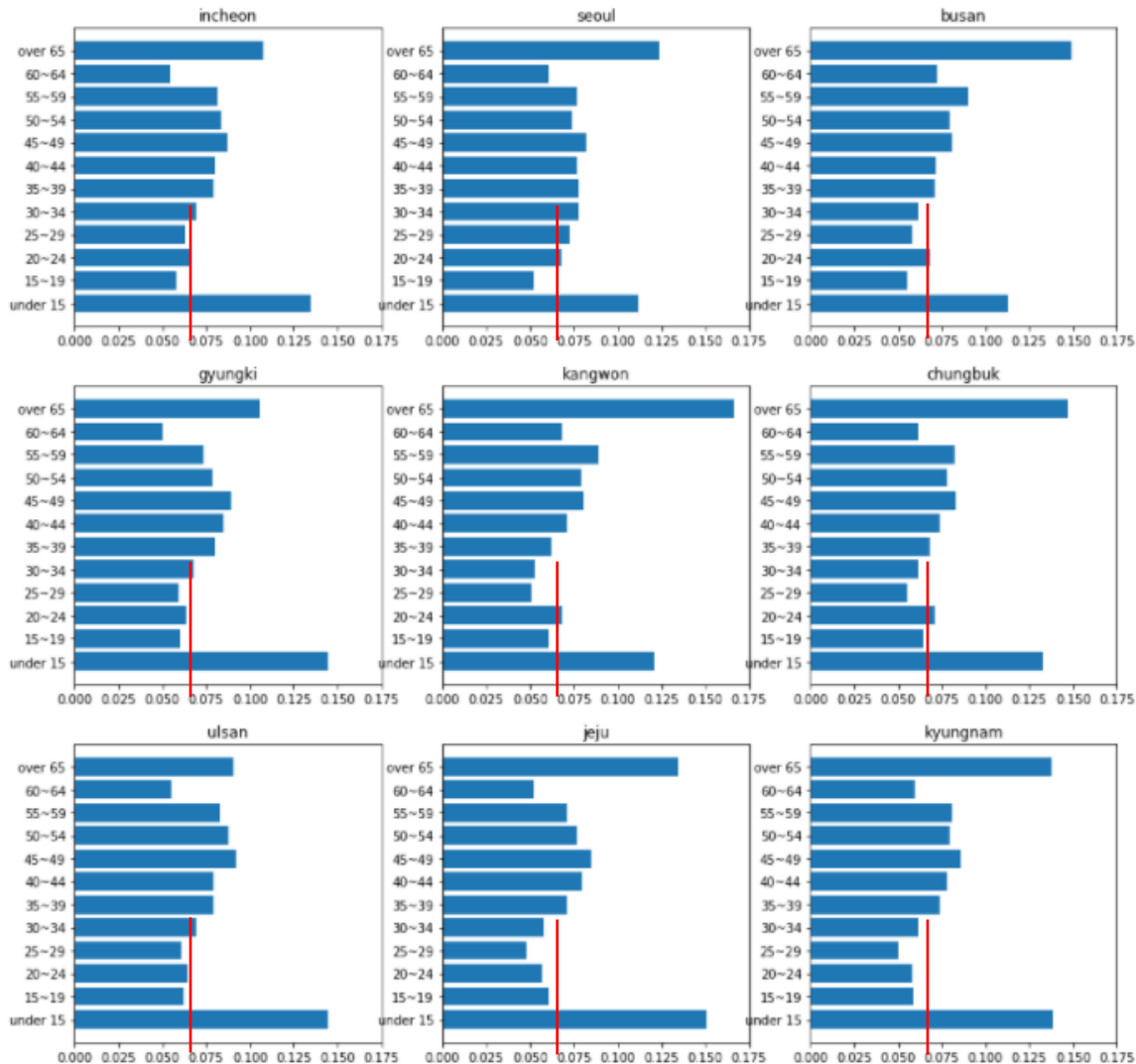
&lt; 지역별 공연 좌석 수 / 지역별 인구 &gt;



- 서울과 경기도에 전국 공연 좌석 수의 60% 이상이 분포해 있다. 하지만, 인구수 대비 공연 좌석 수의 비율 경우, 서울, 강원, 인천, 울산을 제외하면 지역별 격차가 크게 줄어든 것을 확인
- 인천의 경우, 인구대비 공연 시설 수, 좌석 수 모두 최하위
- 특이하게 강원의 경우, 인구 대비 좌석 수가 서울과 비슷한 수준인 것을 확인
- 서울의 경우, 인구 대비 공연 좌석 수가 여전히 크게 높은 것을 확인

인구 대비 공연 시설 수, 좌석 수의 지역별 격차의 원인 분석 필요

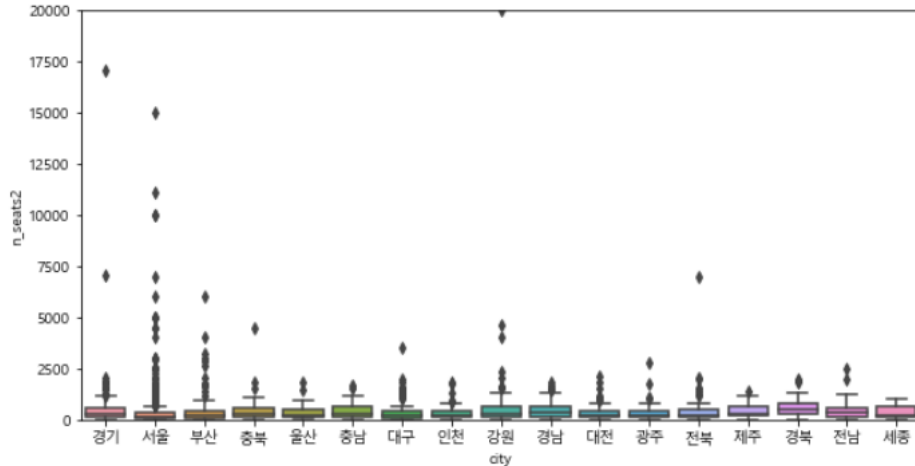
특히, 서울과 강원도의 특잇점 파악 필요  
인천의 원인 분석 필요



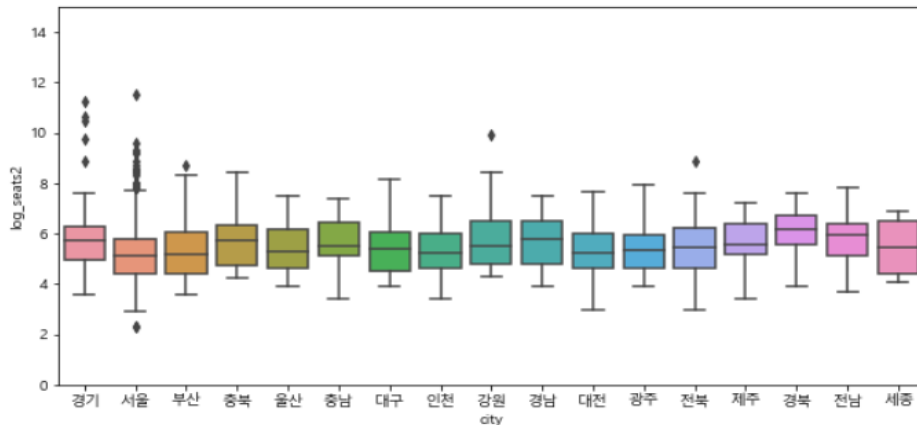
- 17개 전국 광역시, 도, 특별자치도 중 9개 지역의 인구 연령별 구조
- 20~24, 25~29, 30~34 세 모두 0.063을 넘는 지역은 서울이 유일
- 2030세대가 많은 서울의 인구구조가 공연 시설 및 좌석 수의 차이를 나타낸 것으로 보임
- 낮은 인천의 인구 대비 공연 시설, 좌석 수의 원인은 연령별 인구 구조와는 관계 없는 것으로 보인다



&lt; 지역별 공연 좌석 분포 &gt;



로그 변환



- 지역별 각 공연 시설의 좌석 수의 분포를 box-plot을 통해 살펴보았다

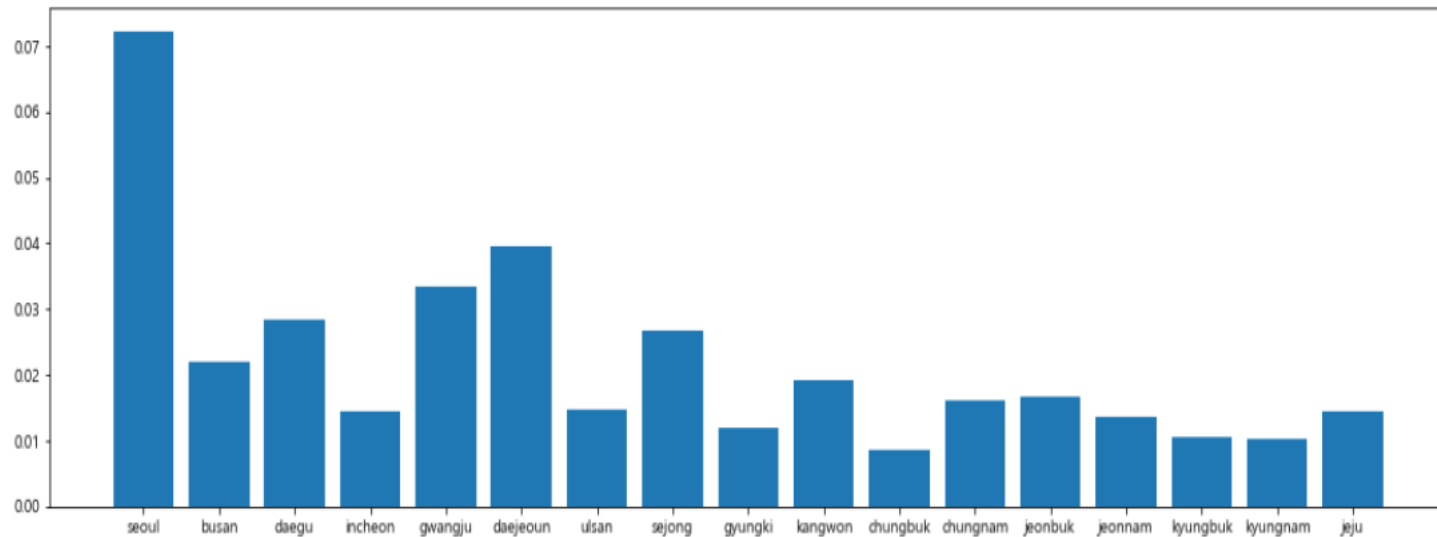
- Outlier가 많아 로그 변환

- 강원도의 경우, 다른 지역에 비해 좌석 수가 많은 공연장이 많은 것을 확인

- 강원도는 우리나라에서 2번째로 면적이 넓은 지역인 반면, 인구는 밑에서 6번째

⇒ 토지 면적이 넓어 큰 규모의 공연 시설이 많으나, 인구가 적어 인구 대비 시설 수와 좌석 수가 높음

&lt; 지역별 인구 대비 공연 수 (2018~2019) &gt;



- 지역별 인구 대비 2018, 2019년 공연 수는 시설, 좌석 수와 다른 특징을 갖는다
- 서울은 여전히 타 지역에 비해 비율이 높으나, 경기도는 하위권에 속하며, 강원도는 뚜렷한 특징이 없다
- 인천은 인구 대비 시설, 좌석, 공연 수 모두 하위권에 속한다

1. 수도권과 타 지역간 문화 격차가 크게 존재할 것이라는 예상과 다르게 **인구를 고려했을 때**, 서울을 제외한 다른 지역간에는 문화 격차가 존재하지 않았다
2. 서울에서 타 지역보다 높은 인구 대비 공연 시설 수, 좌석 수, 공연 수를 보이는 원인은 **2030세대의 높은 비율**로 인한 것이다
3. 경기도의 경우, 인구 대비 공연 시설 수는 적으나, 좌석 수는 많은 것을 보아 큰 규모의 공연 시설이 있으나 **접근성이 떨어지는 것**으로 보인다
4. 인천과 울산의 경우, 연령별 인구 구성 비에 특잇점이 없음에도, 인구 대비 공연 시설 수, 좌석 수, 공연 수가 전구 최하위권으로, **공연 시설의 확대**가 시급해 보인다
5. 인구 대비 공연 수가 서울과 타 지역간 격차가 가장 심하므로, 특별한 지원을 통한 **지방 공연의 확대**가 필요하다