

개인 과제 계획서 (중간보고)

4조 박준민

1. 통합 데이터 소개

Team Project를 통해 공연시설 data, 공연시설 주변 시설 data, 공연 data를 통합했습니다. 공연 예술 통합전산망(www.kopis.or.kr)에서 공연시설 data와 공연 data를 크롤링하였습니다. 카카오 맵 open API를 이용해 공연시설 주변의 편의점, 음식점, 카페 등의 개수를 크롤링했습니다. 공연시설 data에는 공연시설의 이름과 공연시설에 속한 공연장의 이름이 있고 이를 key로 이용해 해당 공연장에서 공연한 공연의 정보를 column으로 통합했습니다. 공연시설의 주소를 key로 사용하여 주변 시설의 개수를 column으로 통합했습니다. 최종 통합된 데이터의 모습은 다음과 같습니다.

name	n_theater	n_seats	theater_name	n_seats2	attr	open
(재)경기문화재단	1	154	다산홀	154	공공(기타)	2001
2001 아울렛키즈홀 [구로]	1	100	2001 아울렛키즈홀	100	민간(대학로 외)	2017
2001 아울렛키즈홀 [구로]	1	100	2001 아울렛키즈홀	100	민간(대학로 외)	2017

open	city	gu	address	pyuneul	jangaein
2001	경기	수원시	경기도 수원시 팔달구 인계로 178 (인계동)	1	1
2017	서울	구로구	서울특별시 구로구 중앙로1길 36 (고척동)	1	1
2017	서울	구로구	서울특별시 구로구 중앙로1길 36 (고척동)	1	1

jangaein	jucha	CS2	PK6	OL7	SW8	CT1	AT4	AD5	FD6	i
1	1	25	27	1	0	12	2	25	464	
1	1	14	16	2	0	1	0	2	237	
1	1	14	16	2	0	1	0	2	237	

FD6	CE7	show	genre	start	end
464	78	경기인형극제, 어린왕자	연극	20180717	20170718
237	38	크리스마스 가족음악회	클래식	20191223	20191223
237	38	춤추는 호랑이와 해님달님	뮤지컬	20191231	20190105

2. Research Question 탐색

Team project를 진행하며 각각의 데이터에 대해서 간단한 EDA를 진행했습니다. 그러던 중 흥미로운 점을 발견할 수 있었습니다.

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

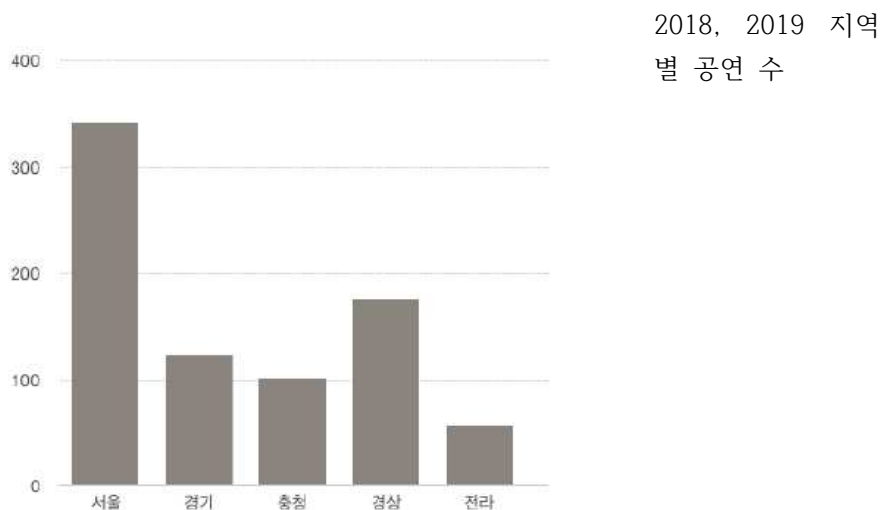
지역별 공연시설 개수

위 그림은 지역별 공연시설 개수입니다. 전국 총 1,846개의 공연시설 중, 958개의 공연시설이 서울, 경기 수도권에 있습니다. 약 51.9%의 공연시설이 수도권에 있습니다. 우리나라 인구의 약 50%가 수도권에 집중된 것을 생각하면 지역별 격차가 그리 크지 않다고 생각할 수 있습니다.

city	강원	경기	경남	경북	광주	대구	대전	부산	서울	세종	울산	인천	전남	전북	제주	충남	충북
n_seats	95025	1032427	100166	73607	74830	78981	35837	115734	1823744	4948	20403	32826	45676	116501	17056	50971	28855

지역별 공연 좌석 총 합

하지만 위의 표를 보면 알 수 있듯이 좌석의 총합을 따지면 전국 좌석의 약 76.21%가 수도권에 집중되어있습니다.



2018년, 2019년 지역별 공연 수의 분포를 보아도 서울 경기 지역에 집중된 것을 확인할 수 있었습니다.

이를 바탕으로 ‘지역별 문화 격차가 존재하는 것이 아닐까?’라는 의문점이 생겼습니다.

개인 과제로 <지역별 문화 격차는 존재하는가>에 관하여 탐구해보고자 합니다.

3. 개인 과제 진행 계획

1) EDA 추가 진행

앞서 진행한 EDA 외에도 지역별로 공연시설과 공연, 공연장의 주변 시설 등에 관하여 다방면으로 시각화를 진행하여 문화 격차가 존재하는지 살펴볼 예정입니다.

- 지역별 장르별 공연 수
- 지역별 최근 공연시설 신규 개관 수
- 지역별 공연시설 규모 (공연장 수, 좌석 수의 평균, 중간값, 기타 통계량 등)
- 지역별 공연 기간 (얼마나 오래 공연하는지)
- 기타 등등

2) 지역별 문화 격차에 관하여

지역별 문화 격차에 대한 도메인 지식이 필요합니다. 인터넷에 지역별 문화 격차를 검색하면 많은 문서들이 나옵니다.

- 지역 간 문화격차 실태 및 개선방안 (국토연구원 - 박태선, 이미영, 한우석)
(policy.nl.go.kr > cmmn > FileDown)
- [Opinion] 우리 모두가 주목해야 할 '문화 격차' [문화 전반] (아트 인사이드 - 고지희 에디터)

- '지역문화 실태조사 및 종합지수' 결과 발표 (문화체육관광부

(https://www.mcst.go.kr/kor/s_notice/press/pressView.jsp?pSeq=17158)

등의 문서들을 참고하여 문화 격차에 관하여 정확히 파악하며, 참고하여 Team project를 통해 얻은 통합된 data를 이용하여 새로운 지역별 문화 지수를 개발할 수 있을지 탐구해 볼 계획입니다.

3) 추가적인 data 탐색

지역별 공연시설이나 공연의 수는 지역별 인구와도 관계가 있을 것으로 보입니다. 국가통계포털(KOSIS)에서 '연령 및 성별 인구 - 시군구'

(https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_11N1603)

data를 추가적으로 사용하여 지역별 인구수, 지역별 연령별 인구와 공연시설 등이 어떤 관계가 있는지 탐구해 볼 계획입니다.

4. 예상되는 어려움 및 문제점

1) 상관관계 인과관계 문제

지역별 인구와 공연시설의 관계를 살펴볼 때, data로는 상관관계만 알 수 있다는 것을 주의해야 합니다. ‘인구 분포 차이로 인해 공연시설의 분포 차이가 발생한다.’ 또는 ‘공연시설을 늘리면 젊은 층의 유입을 확대할 수 있을 것이다.’ 등의 선부른 결론을 주의해야 할 것입니다.

선부르게 결론 짓는 것을 주의하다 보면 분석을 진행하며 어떠한 결과도 도출할 수 없을까 걱정됩니다.

2) 통합된 data의 형식 변경

Team project를 통해 통합된 data는 각각의 공연이 행으로 들어가 있습니다. 지역별 장르별 개수를 새로운 column으로 형태를 변경할 필요가 있을 것 같습니다. 또한, 지역별 공연 기간을 살펴보기 위해서는 종료일과 시작일의 차이를 구해서 새로운 column으로 만들어야 할 것입니다.

분석을 위해 새로 전처리를 할 필요가 있어 보입니다.