

서울시 내 공연 특구 지정 정책 제안

- 비주류 공연 장르를 중심으로

1

분석주제

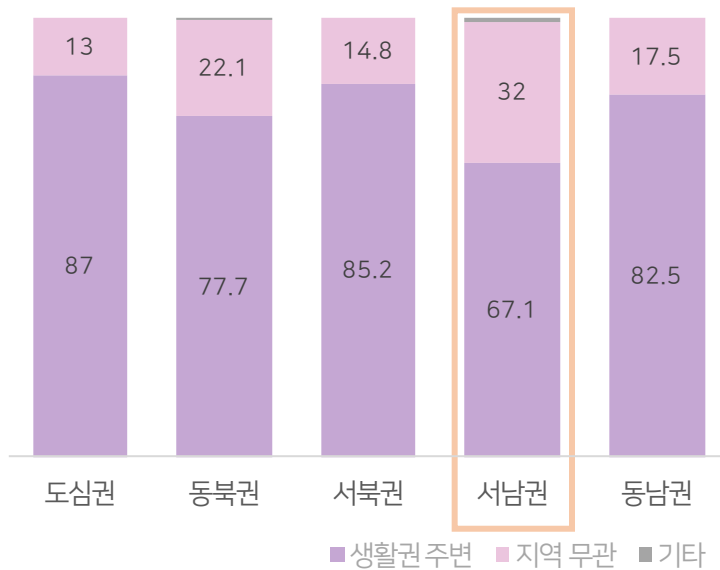
분석주제

1

문제의식

1) 서울시 내 문화격차

문화 예술 주 관람 지역 (%)

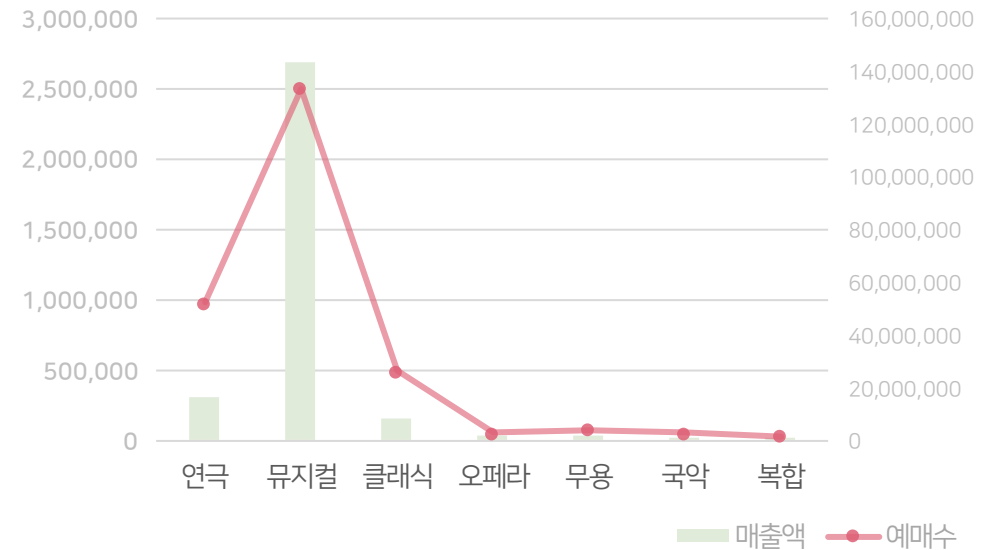


2018년 서울시민 문화향유 실태조사

문화생활이 생활권 내에서
이루어지는지에 대한 여부에 차이가 존재

2) 비주류 장르 소비 확대

2020년 예매 건수 및 매출액



KOPIS 2020년 장르별 통계

뮤지컬이 공연시장 전체 매출액의 83.4%,
예매건수를 기준으로 전체의 60.1%를 차지

2

3

4

5

분석주제

1

분석 목적

1. 공연예술 소비시장에서 **영향력**이 낮은 장르들을 **비주류 장르**로 분류
2. 공연 특구 선정을 통한 비주류 공연 시장 활성화 및 서울시 내 문화 격차 해소

비주류 장르 분류

공연예술시장을 대상으로 한 선행 연구
⇒ **단일 기준**으로 비주류 장르를 구분

ex) 장르별 매출액, 예매 건수



공연의 **다양한 특성**을 **복합적으로 고려**한 기준
⇒ 특정 장르의 소비시장에 대한 **영향력 평가**

ex) 장르별 평균 관람 비용, 객석 수

공연 특구 지정

비주류 장르 **접근성 개선**을 통한
공연예술시장의 **균형적 발전** 방안 탐색

고려 사항

✓ 시민들의 **문화예술 향유** 및 **수요 패턴**
⇒ 거주지 주변에서의 문화예술 수요 증가

서울시 내의 **지역 간 문화 격차**
⇒ 각 권역 별 문화 인프라에 대한 접근성 차이 有

2

3

4

5

분석주제

1

2

3

4

5

데이터 전처리 / 파생 변수 생성

1. 데이터 전처리

데이터 삭제

서울 이외 지역

공연일시 오류

예매 < 취소

데이터 수정

주소 표기 오류

가격 정보

2. 파생 변수 생성

공연
회차별로
그룹화

평균좌석점유율

유료관객비율

평균취소비율

평균시간차

비주류 장르 분류

1. 클러스터링

Spectral Clustering

생성된 4가지의 수치형 파생변수



2개의 클러스터로 구분

각 장르의 지배적 특징을 포함하는
클러스터를 통해 장르 특성 파악

2. 비주류 장르 분류

각 장르별 파생변수의 평균값 고려

[최종 선택된 비주류 장르]

국악

무용

* 장르 특성 파악이 어려운 복합은 제외

문화 격차 반영

1. 파생 변수 생성

공연시설 개수

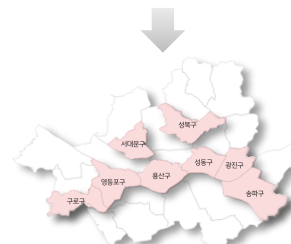
총승하차인원

K-Means Clustering

cluster	평균 공연시설 개수	평균 총승하차인원	개수
0	6.625	3097139	8
1	15.000	5803980	5
2	22.000	2998774	3

2. 8개 자치구 선정

공연시설 개수 & 총승하차인원 고려



특구 선정

1. 지표 생성

교통 상권 경제 문화

4가지 요소를 고려하여 지표 생성

Factor Analysis

Factor A

주변 상권
활성화 정도

Factor B

지역 주민
경제·문화 수준

2. 최종 특구 선정

LISA

요인 점수 별 핫스팟 탐색

KDE

공연 시설의 밀집도 고려



서울시 내
최종 공연 특구 결정

2

데이터 전처리

데이터 전처리

1

데이터 제거 및 수정

분석 방향과 맞도록 기본적인 데이터 전처리 진행 및 데이터 상의 오류라고 판단되는 일부 데이터를 삭제

기존 제공 데이터

...	공연지역명	...	입장권고유번호	공연일시	예매일시
	서울		T1563008080	2020-01-04 15:00	2019-12-17 14:01
	서울		PTD6392008032350510	2020-08-08 15:00	2020-08-03 23:52
제거	부산		153408689716000021	2020-10-08 18:00	2020-10-05 19:43
제거	대전		T1693550220	2020-12-27 15:00	2020-11-30 00:15
제거	서울		T0891491872	2020-12-24 14:00	2020-12-26 17:16

입장권고유번호를 기준으로 예매/취소 건수를 카운트한 데이터

...	입장권고유번호	총예매건수	총취소건수
	T2089218951	1	1
제거	2731894210191400017	0	2
제거	71087925_0	0	2

기존 제공 데이터

공연시설코드	기존 주소	변경한 주소
FC223420	서울특별시 종로구 대학로 57 홍익대학교대학로캠퍼스<U+00A0>	서울특별시 종로구 대학로 57 홍익대학교대학로캠퍼스
FC223420	서울특별시 종로구 대학로 57 홍익대학교대학로캠퍼스?	서울특별시 종로구 대학로 57 홍익대학교대학로캠퍼스
FC223116	서울특별시 종로구 혜화로 10-3 성인빌딩<U+00A0>	서울특별시 종로구 혜화로 10-3 성인빌딩

데이터 전처리 방향

분석 목표에 따라

서울 이외의 지역 공연 제거

공연일시보다 예매일시가 늦을 경우

예매 기록 오류라고 판단 후 제거

예매건수보다 취소건수가 많은 경우

입장권고유번호 오류라고 판단 후 제거

동일한 공연시설코드 내의

주소 표기 오류 수정

2

3

4

5

데이터 전처리

1

2

3

4

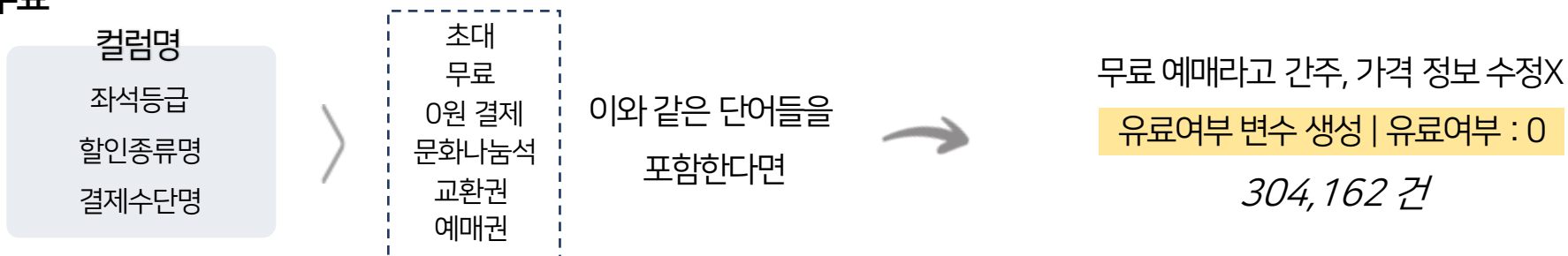
5

가격 데이터 수정

전송 사업자 별로 가격 기록 방식 및 결측치 정도가 상이하여 전처리 필요

총 1055만 건의 데이터 중 1) 가격 기록이 미흡하거나 2) "예매금액, 할인금액, 장당금액" 간의 관계가 맞지 않는 550만 여 건의 데이터 수정

무료



유료

	변경 전			변경 후			좌석등급 #설명
	장당금액	할인금액	예매금액	장당금액	할인금액	예매금액	
Case 1. 금액이 모두 0원	0	0	0	55,000	0	55,000	R석(66,000), S석(44,000) # 좌석등급의 중간값 imputation # 할인 정보 알 수 없음
Case 2. 예매금액 제외 모두 0원	0	0	22,000	22,000	0	22,000	# 할인 정보 알 수 없음
Case 3. 예매금액만 0원	44,000	11,000	0	44,000	11,000	33,000	# 장당금액과 할인금액의 관계를 맞춤
Case 4. 장당금액 = 이미 할인된 금액	55,000	29,000	55,000	84,000	29,000	55,000	# 장당금액과 할인금액의 관계를 맞춤

유료여부 변수 생성 | 유료여부 : 1
10,247,868건

3

비주류 장르 분류

1

2

3

4

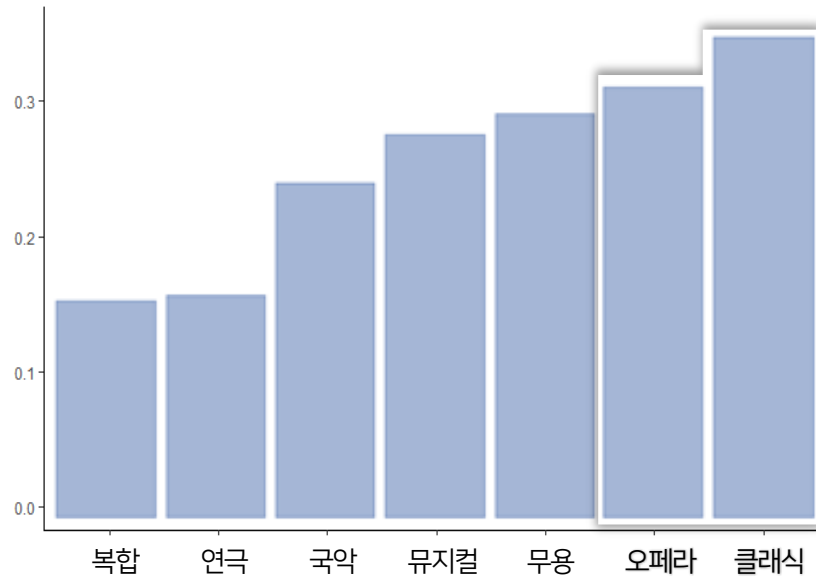
5

비주류 장르 분류

파생변수 생성을 위한 데이터 탐색

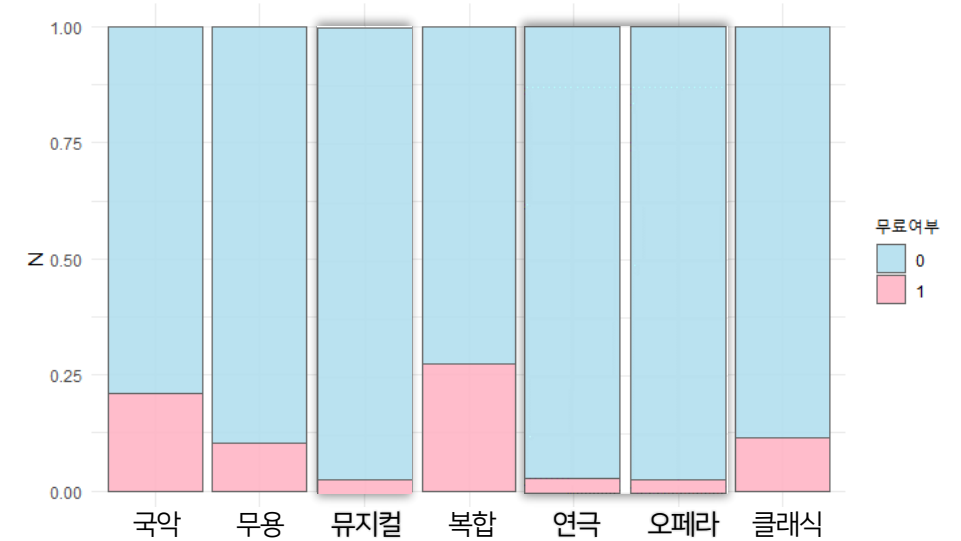
● 평균 좌석 점유율 및 유·무료 관람 비율을 통해 관객의 참여도 확인

평균 좌석 점유율



오페라와 클래식의 좌석점유율이 높음을 확인
장르 간 좌석 점유율 차이가 존재함을 인식

유·무료 관람 비율



수정된 가격 데이터를 이용해
공연 관람객 중 유료 관람객과 무료 관람객을 구분

비주류 장르 분류

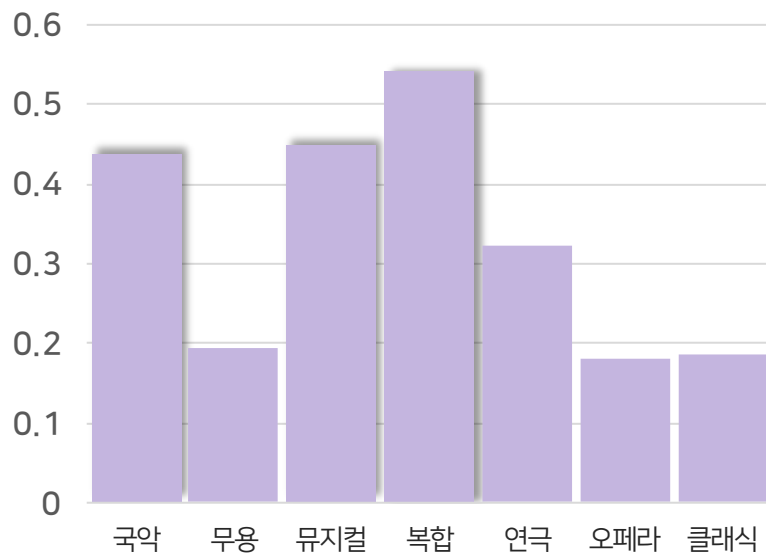
1

파생변수 생성을 위한 데이터 탐색

취소비율 및 시간차 평균을 통해 장르에 따른 관객의 특성 확인

2

취소비율

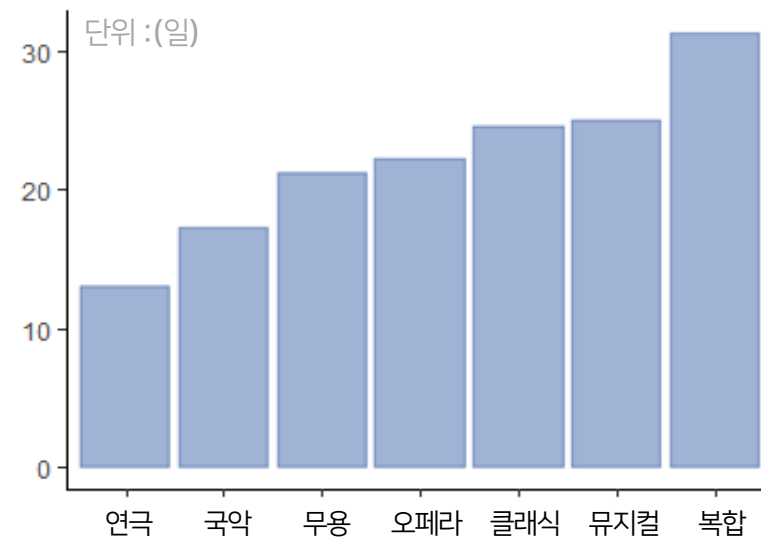


국악, 뮤지컬, 복합에서
타 장르 대비 취소율이 높음을 확인

4

5

시간차 평균 장르별 (공연일시 - 예매일시) 평균



공연일시와 예매일시의 차를 구해
예매 후 공연까지 기다리는 기간 계산

비주류 장르 분류

1

분류 기법

2

Spectral Clustering

그래프 기반 클러스터링 → 데이터 간의 상대적인 관계나 연결을 가장 중요한 정보로 이용

cf) K-Means Clustering : 데이터 값 자체의 거리 기반 클러스터링

3

Data

row: 공연회차

클러스터링 대상 변수

장르명	공연코드	공연일시	유료관객비율	좌석점유율	취소비율	시간차평균
국악	PF351178	2020-01-01T17:00:00Z	1.0	0.000000	1.000000	0.277965
국악	PF351178	2020-01-02T17:00:00Z	1.0	0.006667	0.982301	0.049748
⋮	⋮	⋮	⋮	⋮	⋮	⋮
국악	PF351178	2020-01-04T14:00:00Z	1.0	0.043333	0.814286	1.117410
국악	PF351178	2020-01-04T17:00:00Z	1.0	0.013333	0.973510	0.267280

4

5

비주류 장르 분류

1

2

3

4

5

결과

Spectral Clustering



클러스터 개수 클러스터 간 군집 정도를 나타내는 실루엣 계수
n_clusters= 2, silhouette_score=0.6545

군집	장르명	유료관객비율	좌석점유율	취소비율	시간차평균	개수
0	국악	0.944000	0.364729	0.421526	26.529299	35
	무용	0.897349	0.423901	0.357084	27.206481	83
	뮤지컬	0.982996	0.345554	0.561327	24.065373	5908
	복합	0.690787	0.127314	0.816915	38.019433	89
	연극	0.989567	0.174290	0.585364	22.983119	4253
	오페라	0.978684	0.335901	0.227982	26.885922	38
	클래식	0.974968	0.351356	0.285472	33.435912	471
1	국악	0.893298	0.214318	0.440999	4.854464	188
	무용	0.711042	0.238796	0.121393	5.609747	192
	뮤지컬	0.952934	0.231278	0.358902	6.423673	7593
	복합	0.827391	0.184270	0.327003	6.638258	115
	연극	0.980538	0.162106	0.268974	4.450460	20937
	오페라	0.948667	0.307048	0.152042	7.350059	60
	클래식	0.872438	0.350738	0.135363	6.600583	894

Cluster1 각 장르의 지배적 특징 반영

- ✓ Cluster 1 내에서 장르별 파생변수 비교
- ✓ 특정 지표에서 타 장르에 비해 부정적 수치를 보이는 장르 고려

단, 복합은 장르 특성 반영이 어려워 대상에서 제외

국악

무용

4

특구선정

특구 선정

1

2

3

4

5

문화 격차 반영

활용 변수

✓ 지하철 총 승하차인원

각 구에 위치하는 지하철역에서의 15시 - 21시
지하철 승하차객수의 총합

공연시간의 1사분위수 - 3사분위수

✓ 자치구별 공연시설개수

제공데이터 + KOPIS DB 공연장 활용

→ 자치구별 공연시설개수 count

대상 자치구 → 공연시설개수의 1사분위수 - 3사분위수
공연시설 5~23개 보유 자치구

자치구	공연시설개수	총 승하차인원
강동구	6	35472721
강북구	8	24038790
⋮	⋮	⋮
구로구	11	51609708
동대문구	9	29320029

16×3

자치구 선정

K-Means Clustering n_clusters = 3, silhouette_score = 0.5517

각 클러스터의 변수 평균 비교

cluster	평균 공연시설 개수	평균 총승하차인원	개수
0	6.625	3097139	8
1	15.000	5803980	5
2	22.000	2998774	3

총 8개의 자치구 선정



특구 선정

공연 시설별 지표 생성

요인 분석을 위한 공연 시설별 변수 생성

주변 교통 시설 수	주변 상권 발달 정도	지역 주민 경제 수준	지역 주민 문화 수준
<p>1. 수요 관련 요인</p> <p>공연 시설 주변의 접근성 ↑ ~ 공연 수요 ↑</p>	<p>1. 수요 관련 요인</p> <p>공연 시설 주변의 상권 ↑ ~ 공연 수요 ↑</p>	<p>1. 수요 관련 요인</p> <p>지역 주민들의 경제적 수준 ↑ ~ 공연 수요 ↑</p>	<p>1. 수요 관련 요인</p> <p>지역 주민들의 문화적 수준 ↑ ~ 공연 수요 ↑</p>
<p>2. 변수 생성</p> <p>주차장 개수</p> <p>지하철역 개수</p> <p>버스정류소 개수</p> <p>공연시설 기준 700m 버퍼 내의 시설 개수 계산</p>	<p>2. 변수 생성</p> <p>상권 시설 수</p> <p>일반음식점, 카페, 술집</p> <p>상권 매출</p> <p>상권 생활 인구</p> <p>공연시설 기준 700m 버퍼 내의 시설 개수 계산</p> <p>공연시설 기준 700m 버퍼와 겹치는 상권 면적 비율을 가중치로 사용</p>	<p>2. 변수 생성</p> <p>GRDP</p> <p>부동산 거래금 중앙값</p> <p>공연시설 기준 700m 버퍼와 겹치는 자치구 면적 비율을 가중치로 사용</p> <p>공연시설 기준 700m 버퍼 내의 거래금 중앙값</p>	<p>2. 변수 생성</p> <p>교육 수준</p> <p>문화단체 종사자 수</p> <p>공연시설 기준 700m 버퍼와 겹치는 자치구 면적 비율을 가중치로 사용</p>

특구 선정

1

공연 시설별 지표 생성

2

요인 분석 (Factor Analysis)

“ 변수들 간의 상관관계에 기반하여 변수들을 서로 유사한 변수 끼리 묶어 줌으로써
여러 개의 변수를 더 적은 잠재 요인으로 설명하는 방법 ”

3

KMO(Kaiser-Meyer-Olkin) Test: 요인 분석의 적절성 판단

Kaiser-Meyer-Olkin factor adequacy

Call: psych::KMO(r = fa_var)

Overall MSA = 0.77

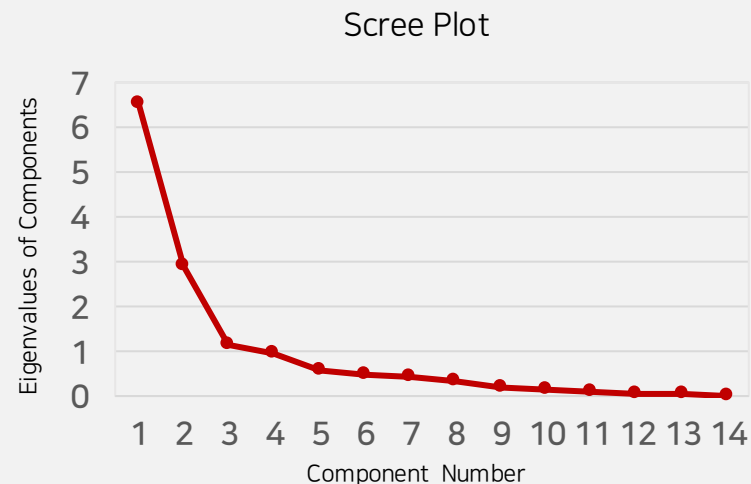
MSA for each item

버스정류장개수	지하철역개수	주차장개수	반경내_GRDP
0.91	0.92	0.82	0.81
거래금증양값	카페개수	술집개수	일반음식점개수
0.62	0.82	0.70	0.75
술집매출액	일반음식점매출액	카페매출액	총생활인구
0.83	0.72	0.79	0.85
교육수준	문화단체중사자점수		
0.52	0.56		

MSA = 0.77이므로 요인 분석 가능

4

Scree Plot: 적절한 요인 개수 도출



고윳값이 1보다 크면서 해석이 가능한 수준인 2개 요인 추출

5

특구 선정

1

2

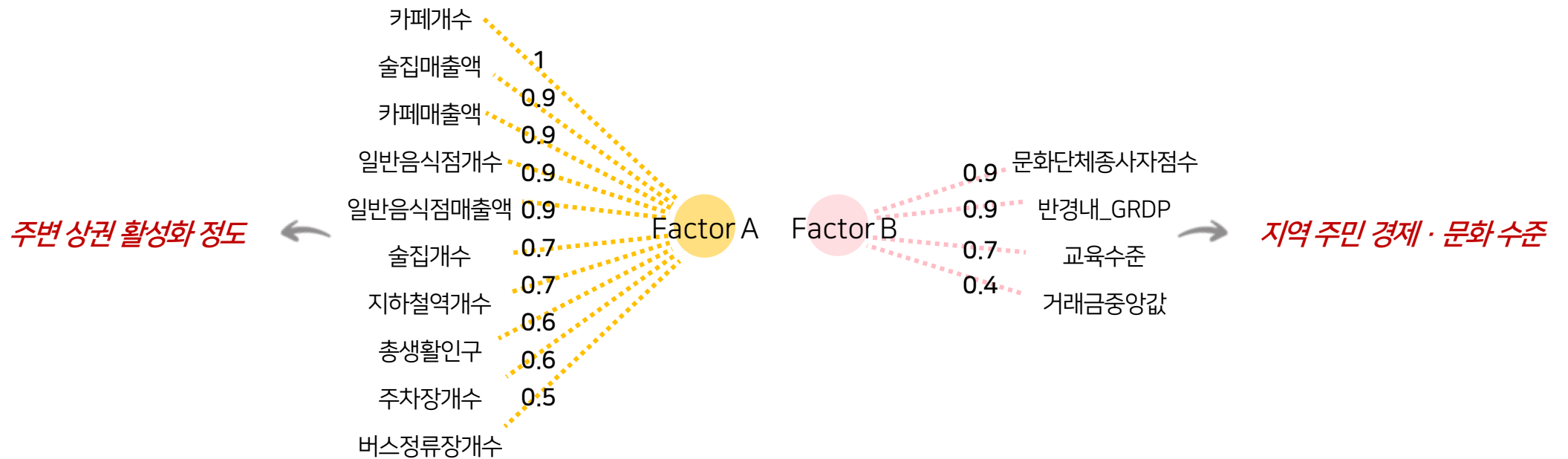
3

4

5

공연 시설별 지표 생성

요인 분석 (Factor Analysis)



특구 선정

1

공연 시설별 지표 생성

2

요인 분석 (Factor Analysis)

요인 분석 결과 & 통계적 유의성 확인

	Factor A	Factor B
SS loadings	6.23	2.67
Proportion Var	0.44	0.19
Cumulative Var	0.44	0.64
Proportion Explained	0.70	0.30
Cumulative Proportion	0.70	1.00

&

	Factor A	Factor B
Correlation of (regression) scores with factors	0.99	0.98
Multiple R square of scores with factors	0.98	0.97
Minimum correlation of possible factor scores	0.96	0.93

4

요인 개수의 적절성 확인

[Test of the hypothesis that 2 factors are sufficient]

The degrees of freedom for the null model are 91 and the objective function was 16.07 with **Chi Square of 9314.32**

The degrees of freedom for the model are 64 and the objective function was 4.58

Tucker Lewis Index of **factoring reliability = 0.601**

5

특구 선정

1

2

3

4

5

공연 시설별 지표 생성

요인 점수 계산

적재값(loadings)

변수에 곱해지는 적재값의 절댓값 ↑
~ 해당 변수의 요인에 대한 기여도 ↑

따라서, 적재값을 요인 점수를 계산하기 위한 **가중치**로 사용

표준화

각 변수들의 단위가 모두 다름
~ 각 변수들의 순수한 영향력 판단 불가

따라서, 각 변수들의 평균을 0, 분산을 1로 만드는 **표준화** 진행

공연시설별 두 개의 요인 점수 계산

주변 상권 활성화 정도 = $\sum_{i=1}^p (loading_i \times variable_i)$, p 는 Factor A에 속하는 변수의 개수, $variable_i$ 는 표준화 된 변수

주민들의 경제·문화 수준 = $\sum_{j=1}^q (loading_j \times variable_j)$, q 는 Factor B에 속하는 변수의 개수, $variable_j$ 는 표준화 된 변수

특구 선정

1

공연 시설별 지표 비교

2

Local Indicators of Spatial Association (LISA)

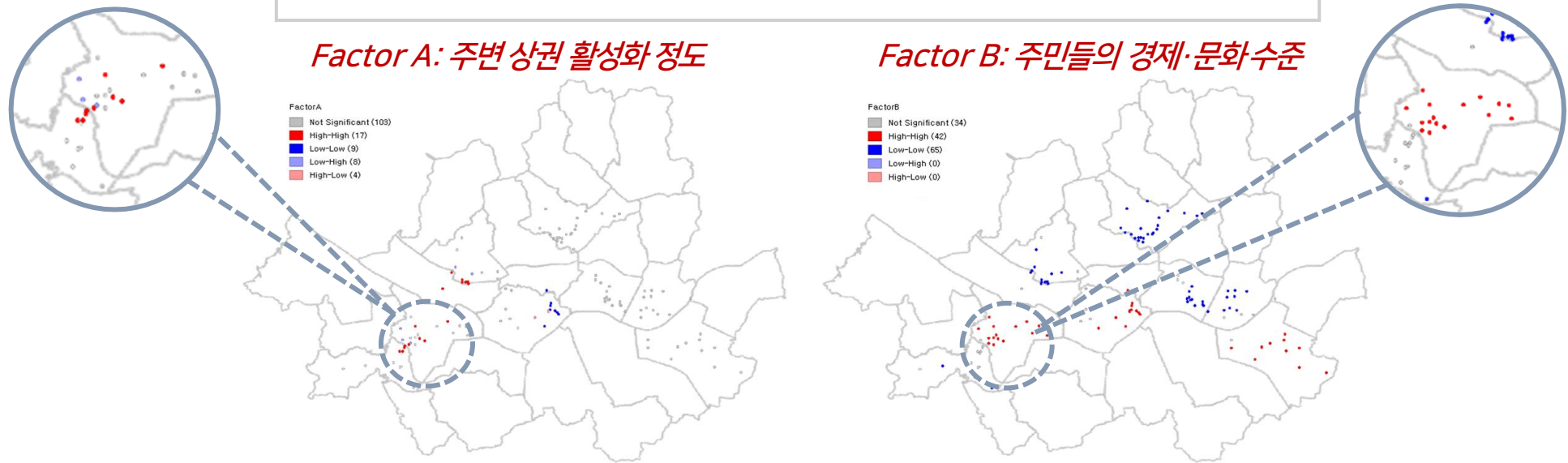
3

공연 특구 선정 기준

공연 시설 밀집 지역 높은 요인 점수를 보이는 밀집 지역 선택
따라서, 국지적 규모의 공간자기상관을 탐색하는 **LISA**를 각 요인 점수별로 적용

4

5



신도림역과 문래 창작촌 일대가 '주변 상권 활성화 정도' 및 '주민들의 경제·문화 수준' 모두 통계적으로 유의한 **핫스팟**으로 나타남

특구 선정

1

공연 시설별 지표 비교

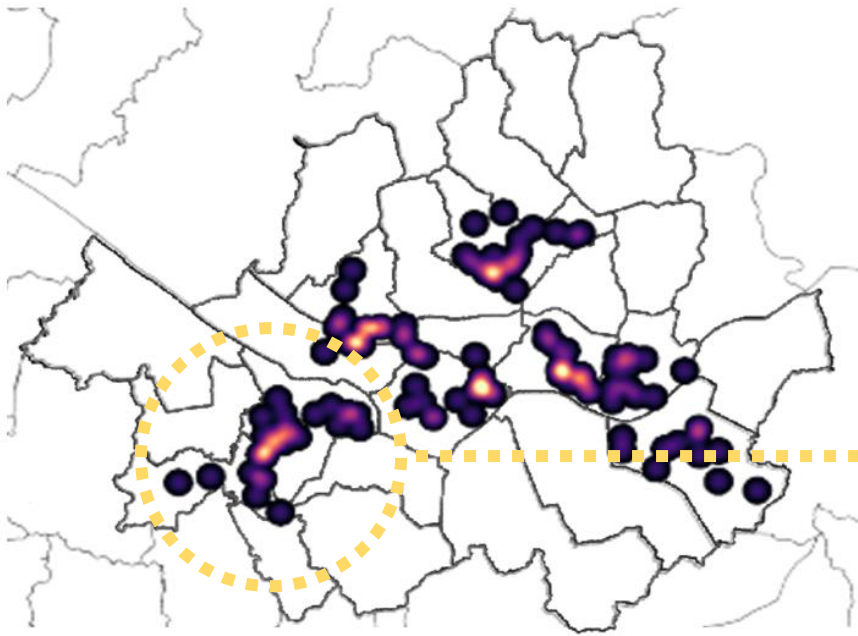
2

커널 밀도 추정 (Kernel Density Estimation)

3

4

5



커널 함수를 이용하는 비모수적 밀도 추정 방법인
커널 밀도 추정 방법을 이용하여 **공연 시설의 밀집도**를 고려



색이 밝을수록 공연 시설이 밀집된 지역으로,
두 요인 점수 모두 높게 나타난 **신도림역과 문래 창작촌 일대가**
공연 시설 밀집 지역임을 확인 가능

5

결론

결론

특구 선정 결과

두 요인 점수에 대한 LISA 결과와 공연장 밀집 지역 파악을 위한 커널 밀도 추정 결과

신도림역과 문래 창작촌 일대를 최종 특구로 선정



면적: 약 530,000m²

특구 내 시설 특성

: 국악 관련 시설 2개, 무용 관련 시설 3개,
관련 기획·제작사 2개

교통 접근성 측면의 이점

: 신도림역(1·2호선), 문래역(2호선),
영등포역(1호선·기차역)

요인 점수 관련 이점

: 주변 상권 활성화 정도 ↑
~ 공연 관람 전 후 소비 기대 가능,
주민들의 경제·문화 수준 ↑
~ 주변 거주민 공연 관심 유도 가능

발전 가능성

: 문래 창작촌 포함,
제2 세종문화회관 예정지와 인접

결론

1

활용 가능성 및 방향

2

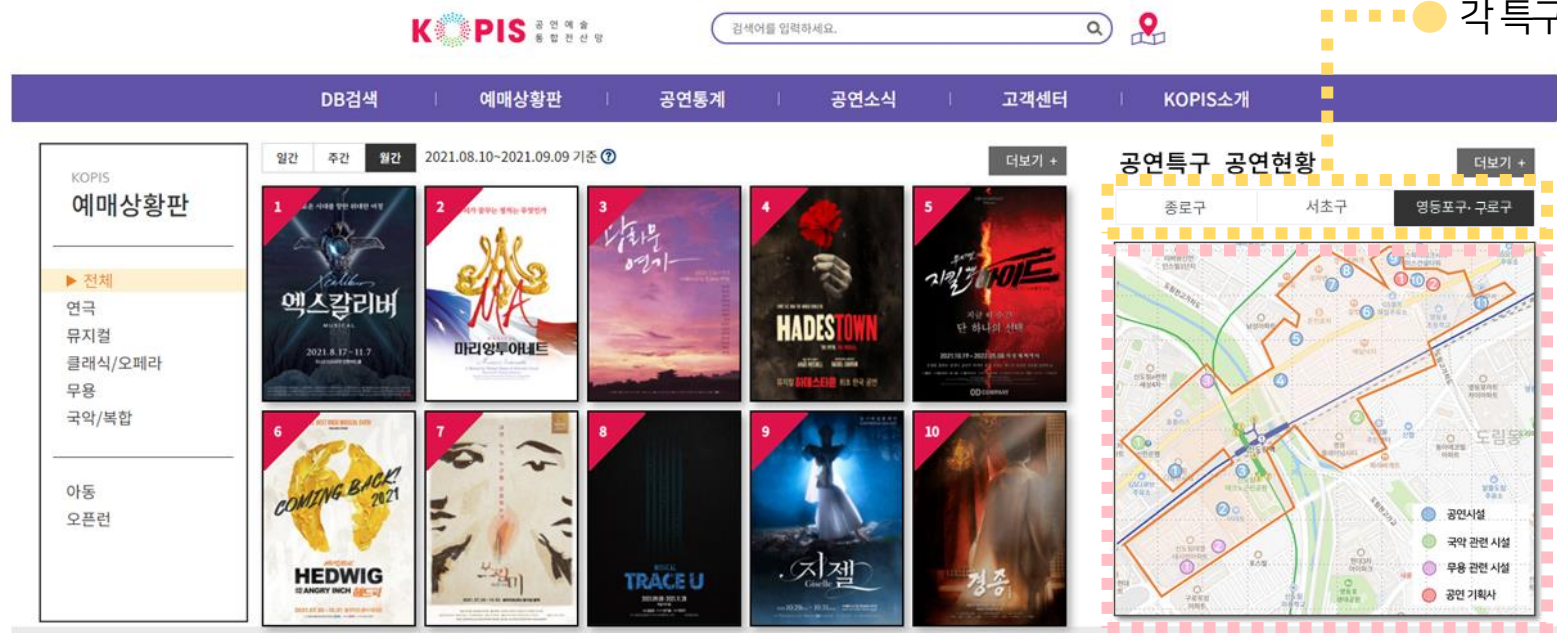
선정된 공연 특구의 활용 - '공연 특구 문화 지도' 인터페이스 구축

✓ 특구에서 열리는 공연 정보 노출을 통한 특구 내 공연 시장 활성화

3

4

5



각 특구의 공연장 및 기반시설 클릭 ~ 공연 및 연계 행사 정보 확인 가능 ●

결론

1

활용 가능성 및 방향

2

선정된 공연 특구의 활용 - 공연 특구별 공연 통계 제공

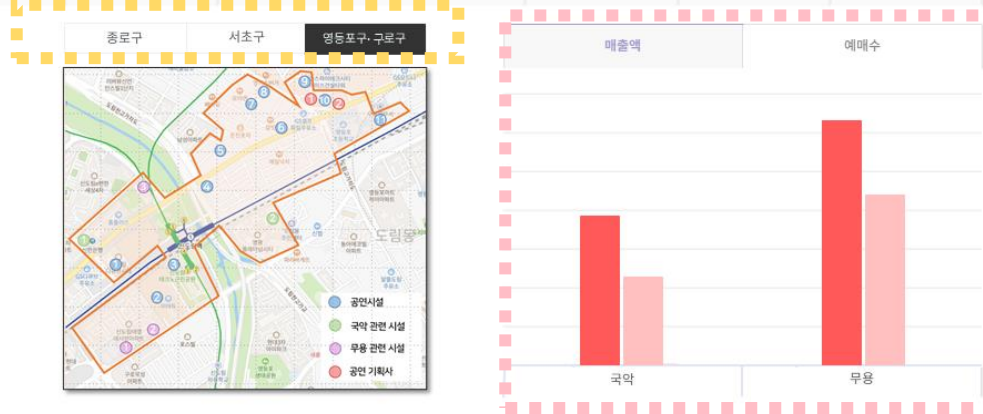
✓ 각 특구별로 진행된 공연에 대한 통계만을 따로 계산하여 개별 장르에 대한 심도 있는 정보 획득

3



● 각 특구를 돌아가며 선택 가능

4



공연 특구에 대한 공연 통계는
타 공연장에서의 공연 통계와는 다른 양상을 보일 것
→ 충분한 관측 값 및 정확한 해석이 전제된다면
해당 장르 **활성화**에 대한 중요 정보 제공 가능

5

● 각 특구에 대한 공연 통계 제공

결론

1

활용 가능성 및 방향

2

데이터 수집 방안

3

공연 특구 맞춤 데이터 수집

공연 특구 → 공연장 밀집 지역 + 공연의 성격 특정

∴ 해당 장르 열성 소비자 + 잠재적 열성 소비자에 대한 정보 획득

각 공연장 및 관련 시설마다의 네트워크 분석 가능

4

데이터 수집 방안

일반 예매 데이터 수집

금액, 예매 및 판매 일시에 대한 정보는 공연 시장 분석에 필수적

∴ 데이터의 오류 발생 지점을 찾아 이를 수정

공연 시장에 대한 더욱 정확한 정보의 획득 가능

5

결론

1

기대 효과

2

신도림역 및 문래 창작촌 일대 공연 특구 지정 시 기대 효과

3

비주류 장르 시장 활성화

특정 장르에 특화된 정책 시행



특구 내 공연장 현황 파악
+
국악 및 무용 장르 공연 특색 사업 진행



국악 및 무용 장르의 수요 증가 기대 가능

4

5

서울시 내 문화 격차 해소

서남권 → 문화 인프라 접근성 및 만족도 최하
→ 공연 수요 잠재력 확인 가능

FactorA

FactorB

주변 상권 활성화 정도	주민들의 경제·문화 수준
외부인 접근을 위한 교통 인프라	지역 내 잠재 수요

1st

2nd

∴ 투자를 통한 충분한 수요 기대 가능

지정된 특구 지역 상권 활성화

¹지역 특구와 관광 특구 동시 지정 허용 검토

- 로컬크리에이터 선정 + 지역뉴딜 벤처펀드
- 야외 공간 활용 다양화 + 활용 가능성 개선

²특구 내 문래 창작촌 포함

∴ 문래 창작촌 수요

↕ 수요의 전이

공연 특구 수요

결론

1

2

3

4

5

한계점

비주류 장르로 분류된 '복합' 장르에 대한 분석 시행의 어려움

- 장르 클러스터링에서 복합 장르는 모든 지표에서 좋지 않은 수치를 보임 ... 비주류 장르로 분류
- 복합 장르의 정의의 모호성
 - 하나의 장르로 분류하기 어려운 경우
 - 두개 이상의 장르가 혼합된 경우 ... 어떠한 장르의 혼합으로 이루어진 것인지 알 수 없음
- ✓ 복합 장르에 대한 정확한 분석을 위해서는 장르 조합에 따른 다각적 접근이 필요 ... 정확한 정보 확보 및 대중 인지도 개선 필요

공연 소비자의 인구통계적 정보 확인의 어려움

- 예매자의 연령, 성별 정보는 정보 제공에 대한 동의 필요 ... 약 절반의 관측치에서 결측치 확인, imputation 불가하여 인구통계 정보 사용 X
- 연령, 성별 정보는 '관람자'가 아닌 '예매자'에 대한 정보 ... 사실 왜곡 위험 존재

질의응답