

예술의 전당 콘서트홀의 효과적 가격 모델 수립

팀명 | DAz





1. 문제 배경 및 필요성
2. 활용 데이터
3. 전처리 및 EDA
4. 분석 과정 및 결과
5. 활용 방안
6. 결론

01 문제 배경 및 필요성





현재 공연예술계에 나타나고 있는 2가지 현상

코로나19 완화로 공연예술 수요 증가

클래식 장르 간 수요 불균형

2019~2022년 3분기 장르별 관람객 수

	2019년	2020년	2021년	2022년
전체	274만 3304명	91만 7504명	176만 1862명	375만 4448명
연극	68만 6594명	21만 8461명	33만 2015명	82만 2311명
뮤지컬	138만 9823명	56만 9307명	99만 3052명	204만 4498명
클래식	41만 1225명	10만 7235명	34만 1581명	63만 4828명
무용	11만 6386명	1만 0189명	4만 9187명	11만 6696명
국악	6만 2244명	4773명	2만 9636명	9만 2132명
복합	7만 7032명	7539명	1만 6391명	4만 3983명

〈자료: 한국문화관광연구원 '코로나19 전후 공연시장 변화 비교 분석'〉

코로나19로 인해 2020년, 2021년 공연 관람객 감소

2022년 이후 코로나19의 완화로

공연예술에 대한 수요가 이전 수준을 넘어설 만큼 다시 회복

기악

성악

전년 대비

티켓 예매 건수를 제외하고

모든 분야에서 증가

각 분야 전년 대비 감소

〈2023년 공연시장 티켓판매 현황 분석 보고서 (상반기)〉

기악과는 다르게 성악 장르에 대한 수요가 뚜렷하게 부족

수익 창출이 필수인 기관의 입장에서는

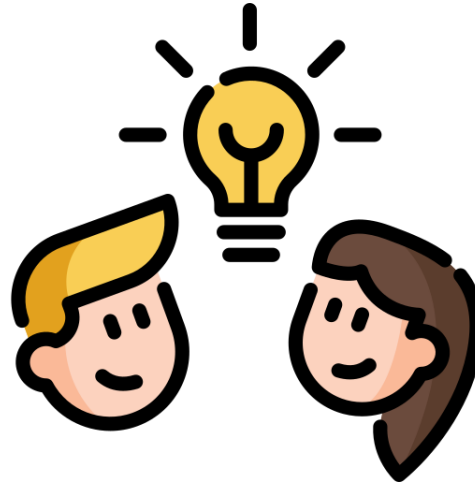
수요가 적은 장르보다 수요가 많은 장르의 공연을 유치

-> 다양한 장르의 공연을 접할 수 있는 기회 감소



증가하는 공연 수요에 대응하여
관객 유치를 위한 판매 전략 필요

장르별 선호도 차이로 나타나는
공연 집중 및 불균형 현상 해결



합리적인 가격 산정 과정을 통해
공연 장르별로 새로운 좌석 등급을 제시

02 활용 데이터

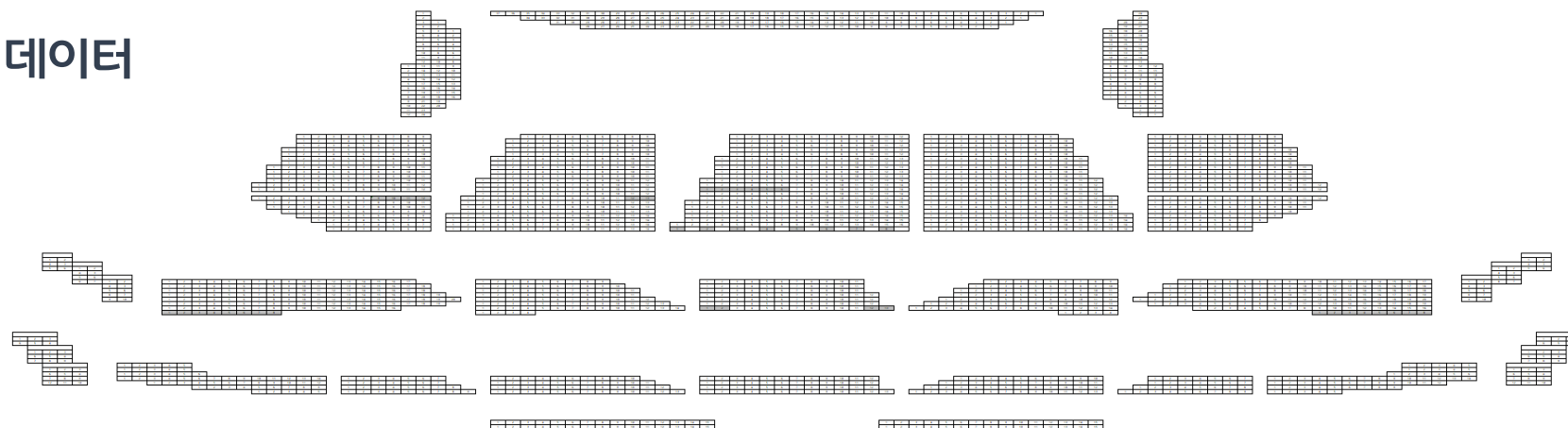




[제공 데이터]

age	gender	seat	price	ticket_cancel	discount_type
membership_type_ 1	membership_type_ 2	membership_type_ 3	membership_type_ 4	membership_type_ 5	membership_type_ 6
tran_date	tran_time	play_date	play_st_time	pre_open_date	open_date
performance_code	genre	place	runnig_time	intermission	member_yn

[외부 데이터] – 생성 데이터



예술의 전당 홈페이지에서 제공하고 있는 **콘서트홀 좌석배치도** 엑셀 xlsx를 참고하여 생성한 데이터
총 **2505개의 좌석**에 대한 상대적인 거리, 위치적 특징, 좌석별 특징을 참고하여 생성한 데이터

03 전처리 및 EDA



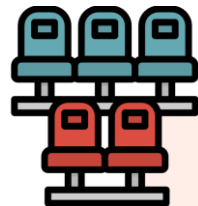


[제공 데이터] – 4가지 전처리 수행



데이터 변경

- price가 0이면 할인가입을 '초대권'으로 변경
- 할인가입이 '초대권'이면 price를 0으로 변경
- 자료 개수가 많지 않은 ['재즈', '크로스오버', '가죽극', '복합장르']를 '기타' 장르로 변경



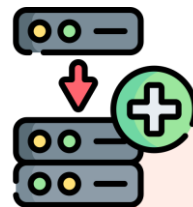
좌석 삭제 및 고유 정보 생성

- 판매하지 않는 공연진행석, 휠체어석 좌석 삭제
- 실제 좌석이 아닌 허상의 자리 삭제
- 각 좌석의 정보를 분리하여 [층, 블록, 박스, 열, 좌석 번호]로 표현



변수 선택

- 선호도가 반영된 변수를 선택
- 고객 개인 정보가 반영된 변수는 선택X
- 날짜, 시작시간, 러닝타임 등 공연의 진행 정보와 관련된 변수는 선택X



파생 변수 생성

- 선예매 여부를 나타내는 pre_yn
- 회원 상태를 나타내는 member_status
- 가입한 유료 멤버십의 개수를 나타내는 member_count
- 정가를 나타내는 real_price



[외부 데이터] – 생성된 5개의 변수

- x축: 3층의 수평적 전체 좌석 중 가운데 좌석을 기준으로 하여 양쪽으로 멀어질수록 +1씩 증가하는 변수
- y축: 1층 블록의 1열을 기준으로 하여 뒤로 멀어질수록 +1씩 증가하는 변수, 단 합창석은 -1씩 감소
- floor: 해당 좌석이 속하는 층을 값으로 하는 변수, 단 합창석은 4로 지정
- COL/ROW: 콘서트홀 좌석배치도 xlsx에서 해당 좌석이 위치한 셀의 열/행 위치를 지정

[외부 데이터] – 2가지 전처리 수행

파생 변수 생성

- 블록을 1로, 박스를 0으로 구분하는 block_box 변수 생성
- 판매 빈도가 적은 합창석과 박스는 0, 그 외는 1인 always 변수 생성
- 셀의 행, 열 위치를 하나로 나타내는 cell 변수 생성

사용한 변수 제거

- cell 변수를 생성하고 남은 COL, ROW 변수를 삭제

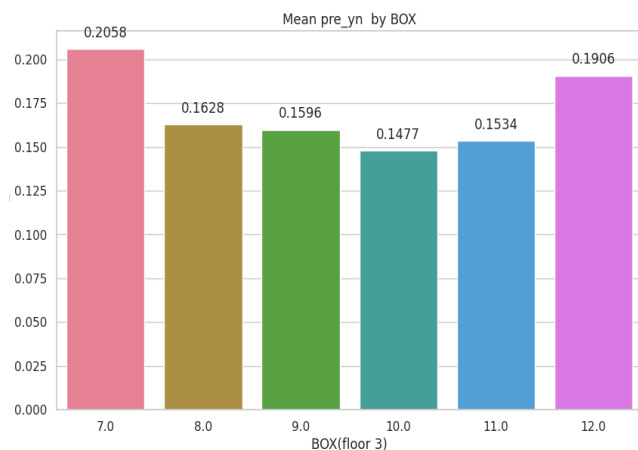
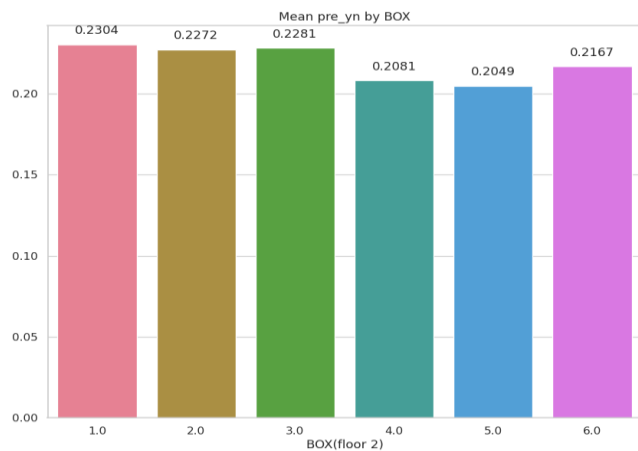
=> 총 2465개의 좌석에 대한 7개 변수

	seat	x	y	floor	block_box	always	cell
0	1층 A블록1열 1	34	1	1	1	1	U31
1	1층 A블록1열 2	33	1	1	1	1	V31
2	1층 A블록1열 3	32	1	1	1	1	W31
3	1층 A블록1열 4	31	1	1	1	1	X31
4	1층 A블록1열 5	30	1	1	1	1	Y31

	seat	x	y	floor	block_box	always	cell
2460	3층 BOX12 2	52	47	3	0	0	DA77
2461	3층 BOX12 3	53	47	3	0	0	DB77
2462	3층 BOX12 4	53	48	3	0	0	DB78
2463	3층 BOX12 5	52	48	3	0	0	DA78
2464	3층 BOX12 6	51	48	3	0	0	CZ78

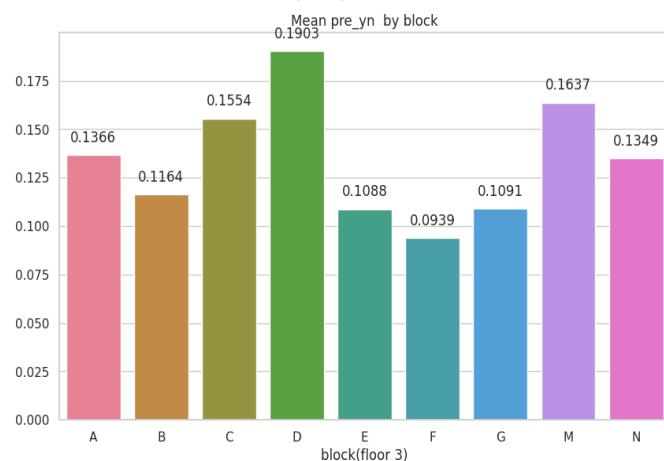
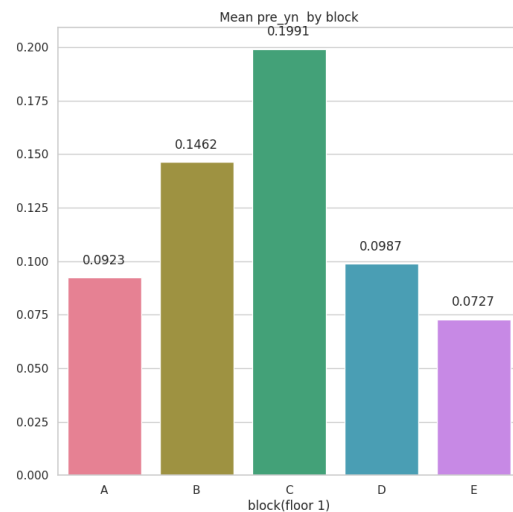


선예매는 1, 일반 예매는 0로 선예매 여부를 나타내는 변수



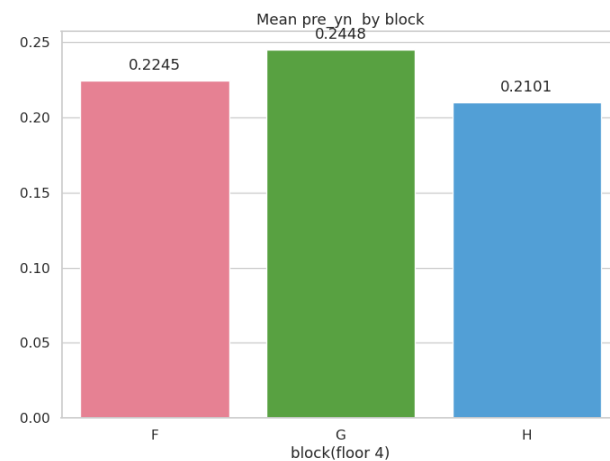
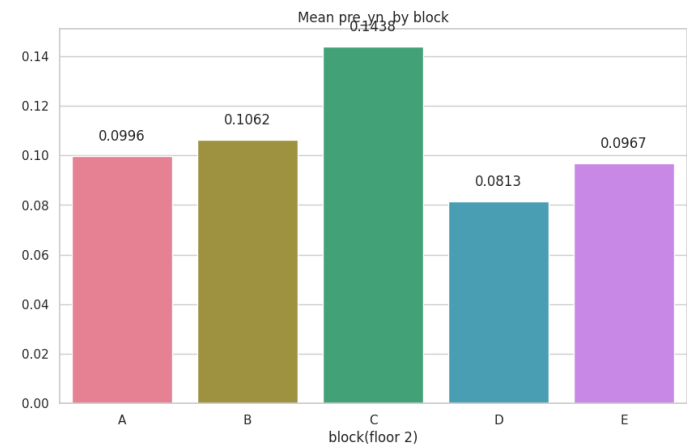
각 BOX 별 pre_yn 평균은

- 2층 왼쪽의 BOX 1,2,3이 4,5,6보다 높음
- 3층의 양끝 BOX 7,12가 다른 BOX에 비해 높음



각 block 별 pre_yn 평균은

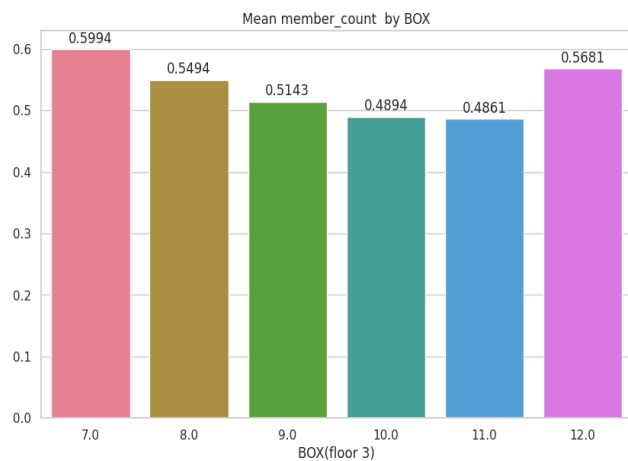
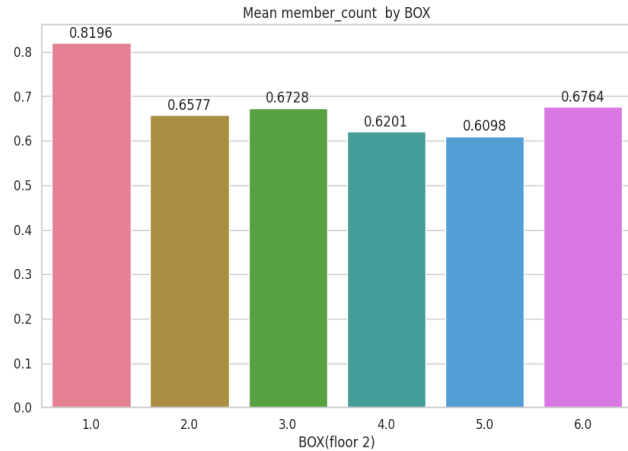
- 각 층의 중앙인 1층 C블록, 2층 C블록이 다른 블록에 비해 높음
- 각 층의 중앙인 3층 D블록, 합창석 G블록이 다른 블록에 비해 높음



pre_yn에 대하여 좌석 위치에 따른 차이가 존재!

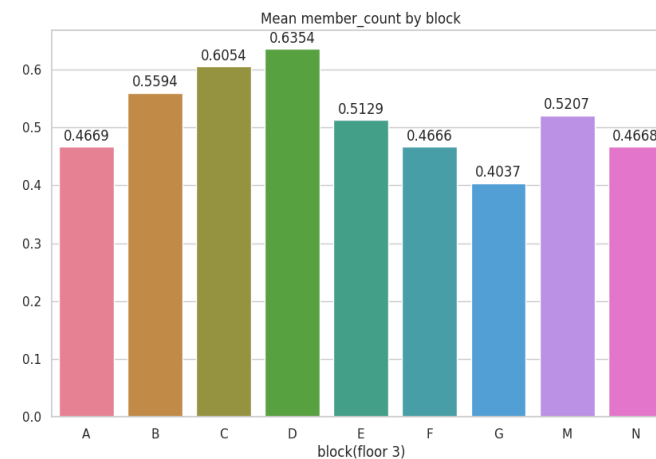
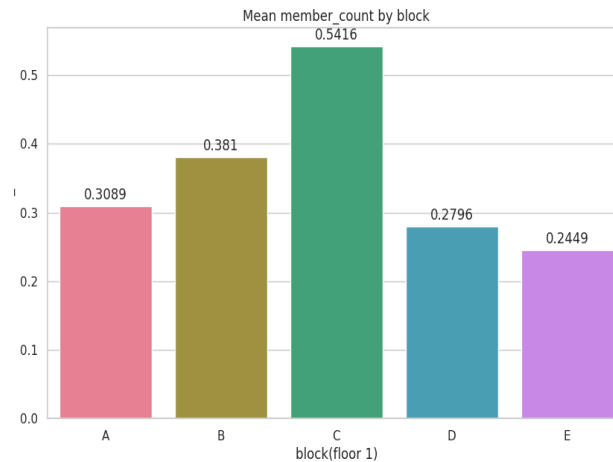


0부터 3까지 가입한 유료 멤버십 개수를 나타내는 변수



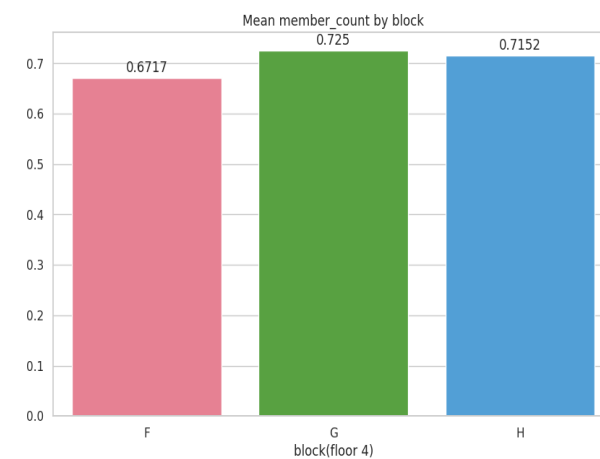
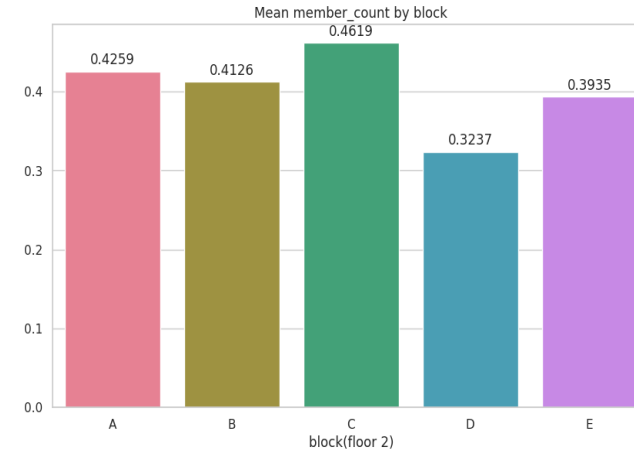
각 BOX 별 member_count 평균은

- 2층 가장 왼쪽의 BOX 1이 다른 BOX에 비해 높음
- 3층의 양끝 BOX 7, 12가 다른 BOX에 비해 높음



각 block 별 member_count 평균은

- 각 층의 중앙인 1층 C블록, 2층 C블록이 다른 블록에 비해 높음
- 각 층의 중앙인 3층 D블록, 합창석 G블록이 다른 블록에 비해 높음



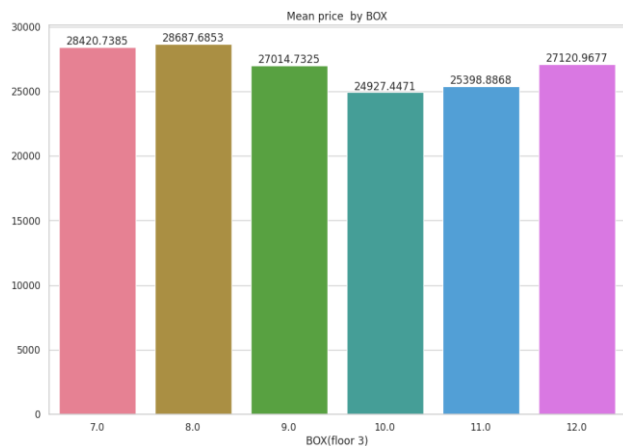
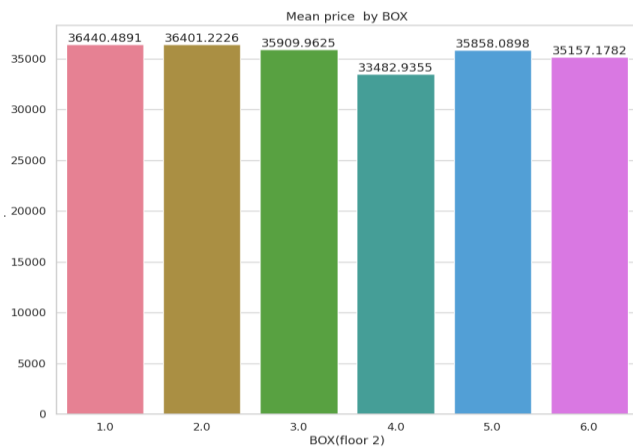
member_count에 대하여 좌석 위치에 따른 차이가 존재!

**BIGCONTEST**

전처리 및 EDA

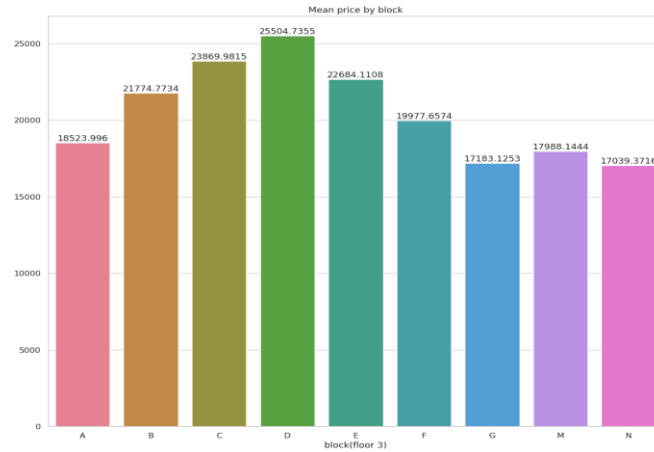
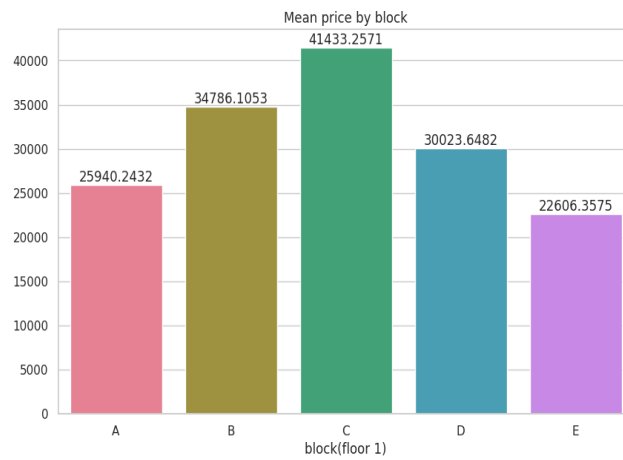
price

고객이 실제로 지불한 금액을 나타내는 변수



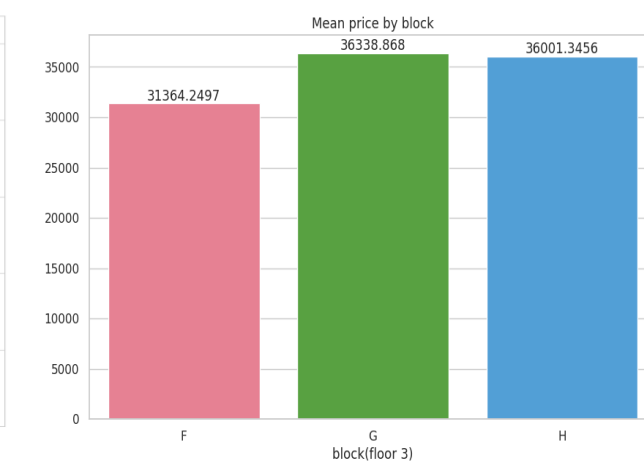
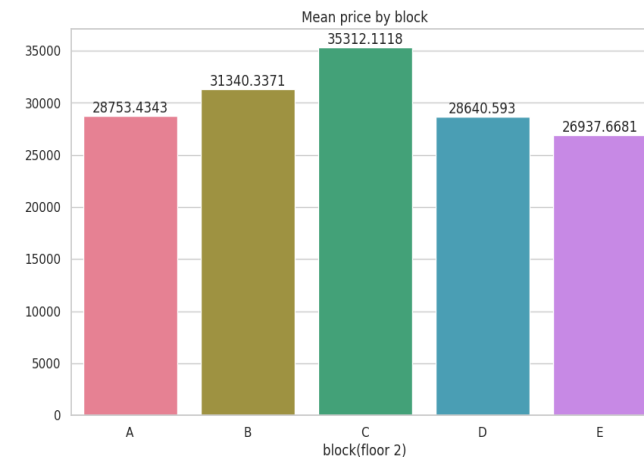
각 BOX 별 price 평균은

- 2층 왼쪽의 BOX 1,2,3이 4,5,6에 비해 높음
- 3층 왼쪽의 BOX 7,8이 다른 BOX에 비해 높음



각 block 별 price 평균은

- 각 층의 중앙인 1층 C블록, 2층 C블록이 다른 블록에 비해 높음
- 각 층의 중앙인 3층 D블록, 합창석 G블록이 다른 블록에 비해 높음



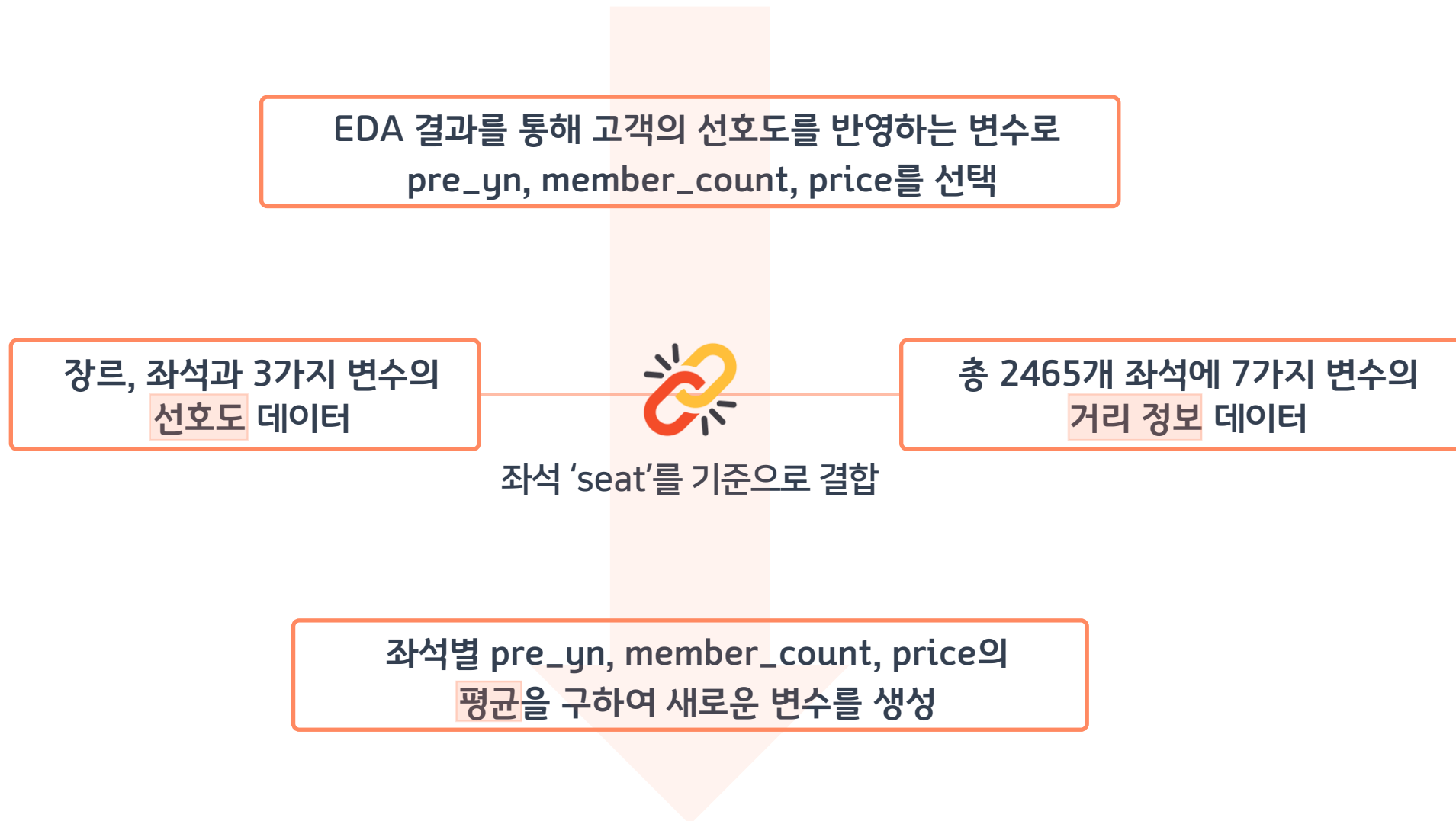
price에 대하여 좌석 위치에 따른 차이가 존재!

04 분석 과정 및 결과





[최종 데이터] – 3가지 과정 수행





[최종 데이터]

ex) 클래식 장르를 선택한 경우

	seat	x	y	floor	always	block_box	cell	genre	price	pre_yn	member_count	price_m	pre_yn_m	member_count_m
0	1층 A블록1열 1	34	1	1	1	1	U31	클래식	72000	NaN	0	37975.9036	0.0769	0.3253
1	1층 A블록1열 1	34	1	1	1	1	U31	클래식	99000	0.0	1	37975.9036	0.0769	0.3253
2	1층 A블록1열 1	34	1	1	1	1	U31	클래식	0	0.0	0	37975.9036	0.0769	0.3253
3	1층 A블록1열 1	34	1	1	1	1	U31	클래식	50000	0.0	0	37975.9036	0.0769	0.3253
4	1층 A블록1열 1	34	1	1	1	1	U31	클래식	30000	0.0	0	37975.9036	0.0769	0.3253
...
233515	3층 BOX12 6	51	48	3	0	0	CZ78	클래식	54000	NaN	2	31113.6364	0.2982	0.7955
233516	3층 BOX12 6	51	48	3	0	0	CZ78	클래식	0	0.0	0	31113.6364	0.2982	0.7955
233517	3층 BOX12 6	51	48	3	0	0	CZ78	클래식	18000	0.0	0	31113.6364	0.2982	0.7955
233518	3층 BOX12 6	51	48	3	0	0	CZ78	클래식	0	0.0	0	31113.6364	0.2982	0.7955
233519	3층 BOX12 6	51	48	3	0	0	CZ78	클래식	20000	0.0	0	31113.6364	0.2982	0.7955

거리 정보 데이터

선호도 데이터

선호도와 거리 정보를 잘 반영할 수 있는
하나의 변수로 각각 차원 축소





[표준화 스케일링 및 차원 축소] : 차원 축소 이전, StandardScaler를 사용하여 각 변수를 표준화

ex) 클래식 장르

	0	1	2
0	0.090131	-0.859309	-0.458923
1	0.090131	-0.859309	-0.458923
2	0.090131	-0.859309	-0.458923
3	0.090131	-0.859309	-0.458923
4	0.090131	-0.859309	-0.458923
...
233515	-0.460259	1.540146	1.633591
233516	-0.460259	1.540146	1.633591
233517	-0.460259	1.540146	1.633591
233518	-0.460259	1.540146	1.633591
233519	-0.460259	1.540146	1.633591

선호도 데이터

3가지 변수

↓
1가지 변수

	PCA1
0	-0.743012
1	-0.743012
2	-0.743012
3	-0.743012
4	-0.743012
...	...
233515	1.653689
233516	1.653689
233517	1.653689
233518	1.653689
233519	1.653689

	0	1	2	3
0	1.074944	-1.018555	0.342724	-0.767375
1	1.074944	-1.018555	0.342724	-0.767375
2	1.074944	-1.018555	0.342724	-0.767375
3	1.074944	-1.018555	0.342724	-0.767375
4	1.074944	-1.018555	0.342724	-0.767375
...
233515	2.418952	1.226089	-2.917802	1.368609
233516	2.418952	1.226089	-2.917802	1.368609
233517	2.418952	1.226089	-2.917802	1.368609
233518	2.418952	1.226089	-2.917802	1.368609
233519	2.418952	1.226089	-2.917802	1.368609

거리 정보 데이터

4가지 변수

↓
1가지 변수

	PCA2
0	-0.448931
1	-0.448931
2	-0.448931
3	-0.448931
4	-0.448931
...	...
233515	3.913163
233516	3.913163
233517	3.913163
233518	3.913163
233519	3.913163

차원 축소 데이터

	PCA1	PCA2
0	-0.743012	-0.448931
1	-0.743012	-0.448931
2	-0.743012	-0.448931
3	-0.743012	-0.448931
4	-0.743012	-0.448931
...
233515	1.653689	3.913163
233516	1.653689	3.913163
233517	1.653689	3.913163
233518	1.653689	3.913163
233519	1.653689	3.913163



[군집화]: K-Means 알고리즘을 사용하여 차원 축소 데이터 군집화

군집 3개

	PCA1	PCA2	ccluster
0	-0.743012	-0.448931	1
1	-0.743012	-0.448931	1
2	-0.743012	-0.448931	1
3	-0.743012	-0.448931	1
4	-0.743012	-0.448931	1
...
233515	1.653689	3.913163	2
233516	1.653689	3.913163	2
233517	1.653689	3.913163	2
233518	1.653689	3.913163	2
233519	1.653689	3.913163	2

	cluster	n
0	1	136815
1	0	67338
2	2	29367

군집 4개

	PCA1	PCA2	ccluster
0	-0.743012	-0.448931	2
1	-0.743012	-0.448931	2
2	-0.743012	-0.448931	2
3	-0.743012	-0.448931	2
4	-0.743012	-0.448931	2
...
233515	1.653689	3.913163	1
233516	1.653689	3.913163	1
233517	1.653689	3.913163	1
233518	1.653689	3.913163	1
233519	1.653689	3.913163	1

	cluster	n
0	0	102591
1	2	90436
2	1	25647
3	3	14846

군집 5개

	PCA1	PCA2	ccluster
0	-0.743012	-0.448931	2
1	-0.743012	-0.448931	2
2	-0.743012	-0.448931	2
3	-0.743012	-0.448931	2
4	-0.743012	-0.448931	2
...
233515	1.653689	3.913163	3
233516	1.653689	3.913163	3
233517	1.653689	3.913163	3
233518	1.653689	3.913163	3
233519	1.653689	3.913163	3

	cluster	n
0	2	90436
1	1	59965
2	0	42626
3	3	25647
4	4	14846



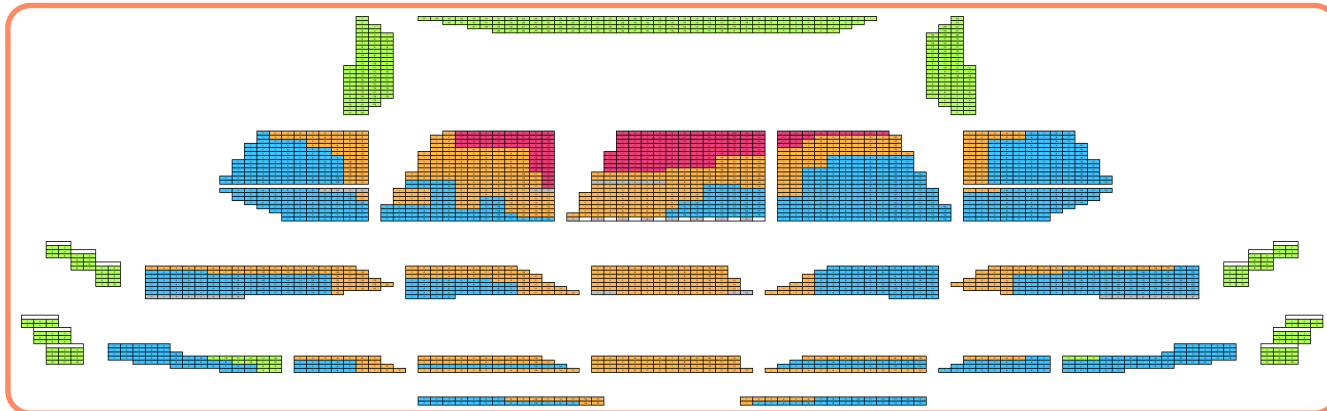
BIGCONTEST

분석 과정 및 결과

[결과 시각화]

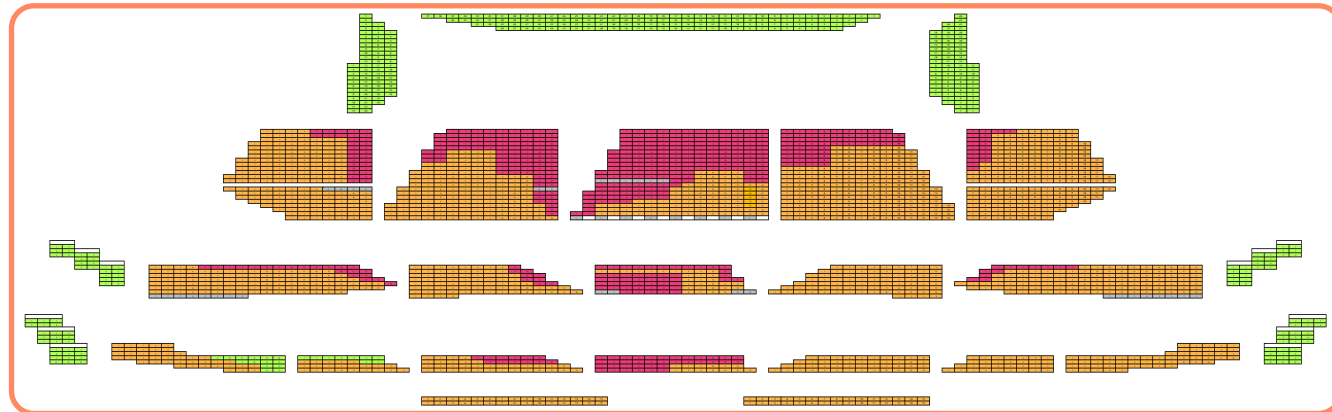
: 예술의 전당 콘서트홀 좌석배치도 엑셀 xlsx을 이용하여
군집화 결과를 좌석배치도로 시각화

클래식 장르의 좌석배치도

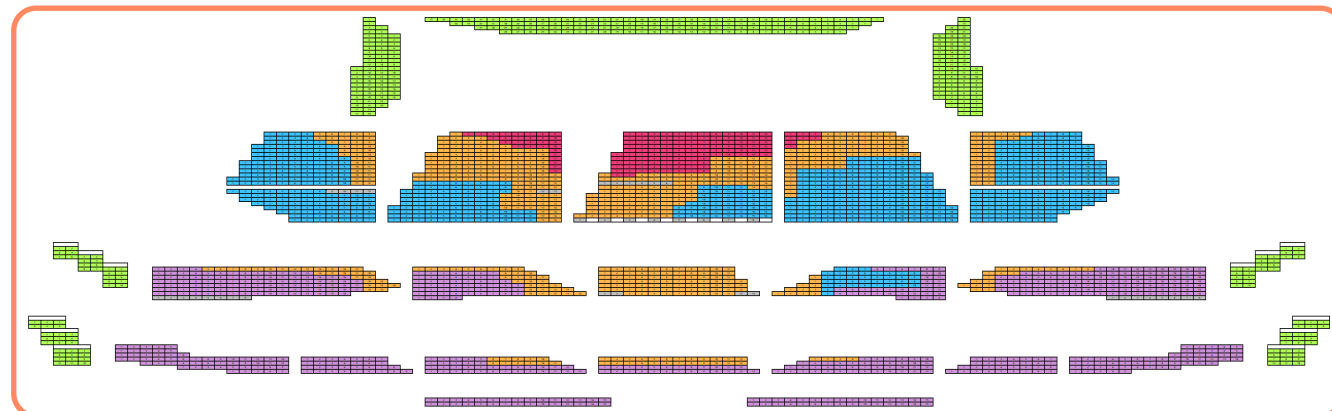


▶ 좌석 등급 4개

▶ 좌석 등급 3개



▶ 좌석 등급 5개



※ 도출된 시각화 결과의 이상치 및 특이치를 조정한 결과



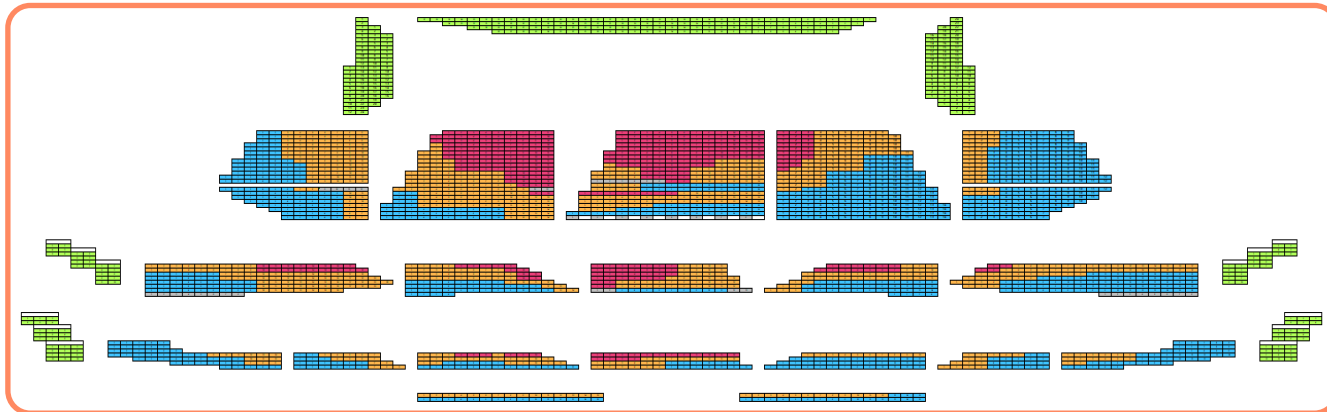
BIGCONTEST

분석 과정 및 결과

[결과 시각화]

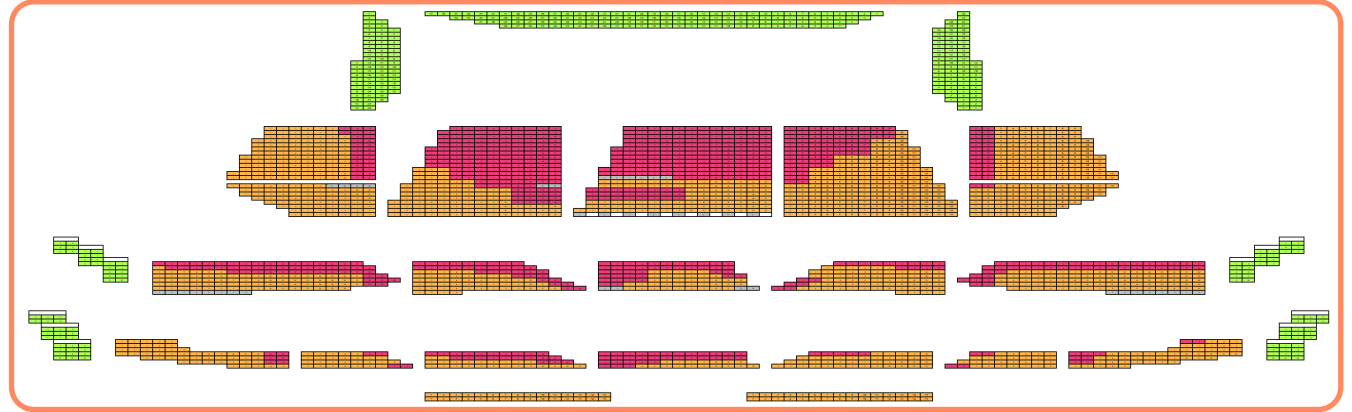
: 예술의 전당 콘서트홀 좌석배치도 엑셀 xlsx을 이용하여
군집화 결과를 좌석배치도로 시각화

독주 장르의 좌석배치도

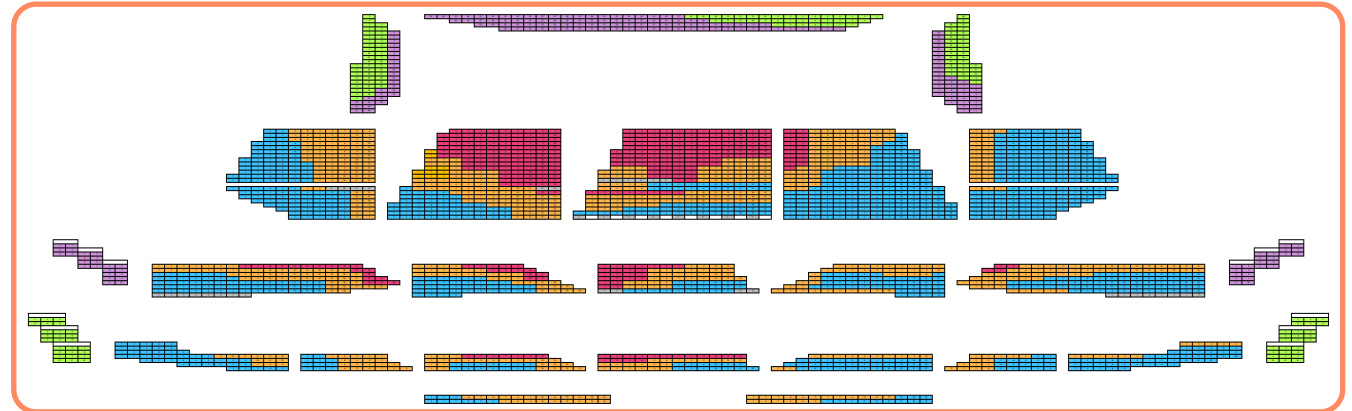


▶ 좌석 등급 4개

▶ 좌석 등급 3개



▶ 좌석 등급 5개



※ 도출된 시각화 결과의 이상치 및 특이치를 조정한 결과



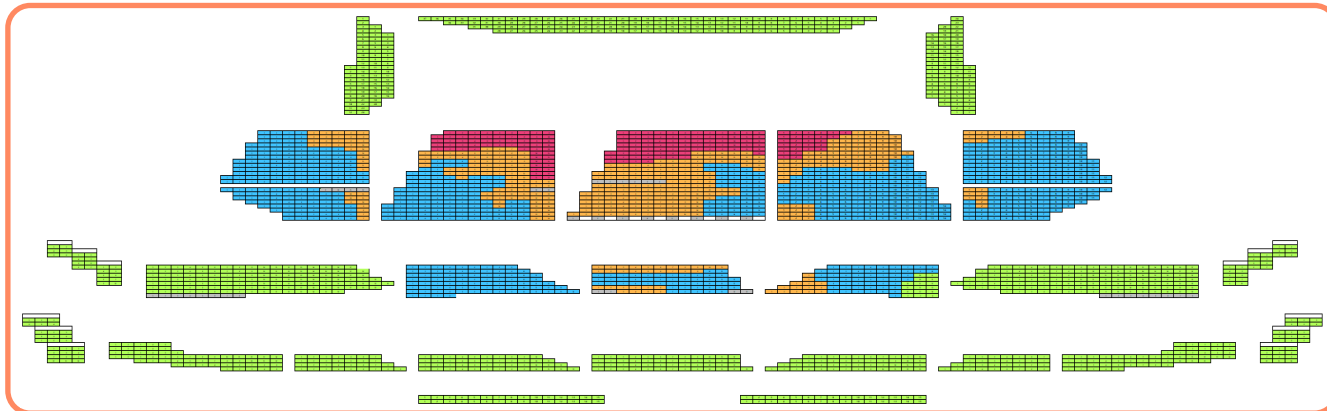
BIGCONTEST

분석 과정 및 결과

[결과 시각화]

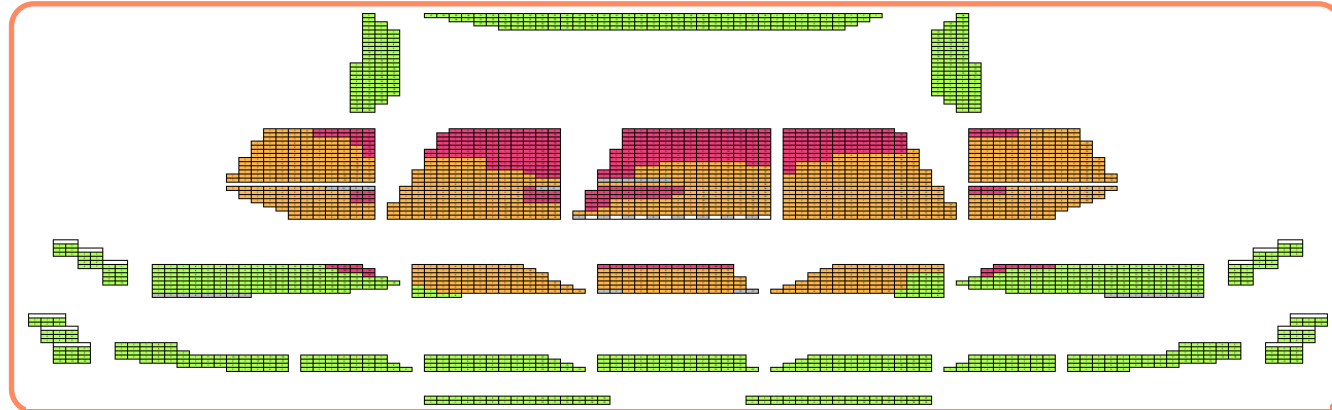
: 예술의 전당 콘서트홀 좌석배치도 엑셀 xlsx을 이용하여
군집화 결과를 좌석배치도로 시각화

성악 장르의 좌석배치도

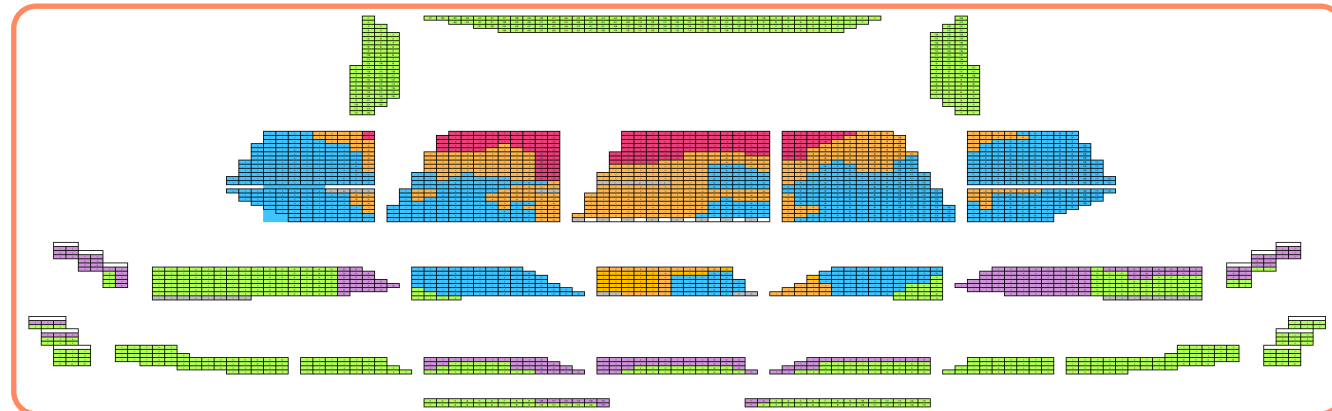


▶ 좌석 등급 4개

▶ 좌석 등급 3개



▶ 좌석 등급 5개



※ 도출된 시각화 결과의 이상치 및 특이치를 조정한 결과

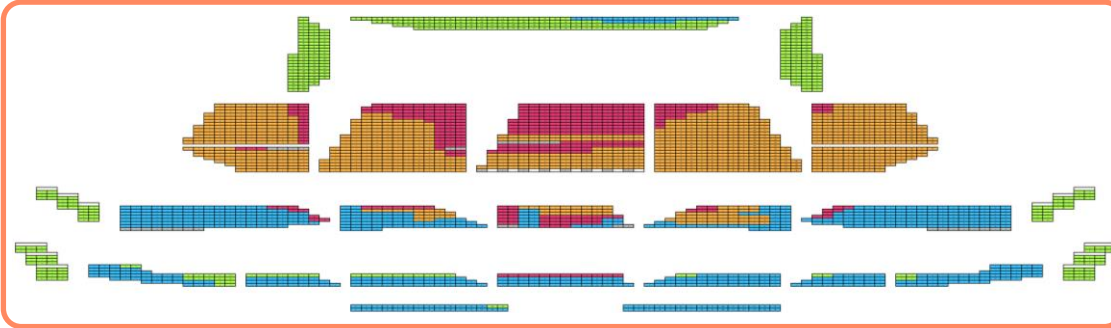


BIGCONTEST

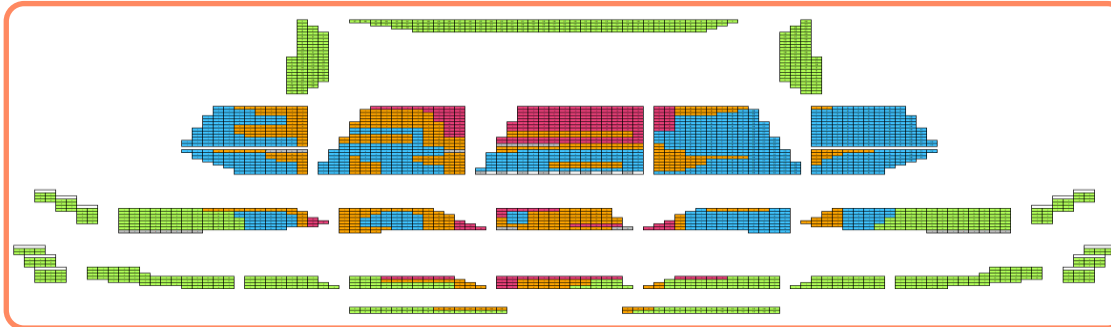
분석 과정 및 결과

[결과 시각화] : 좌석 등급 4개인 그 외 장르

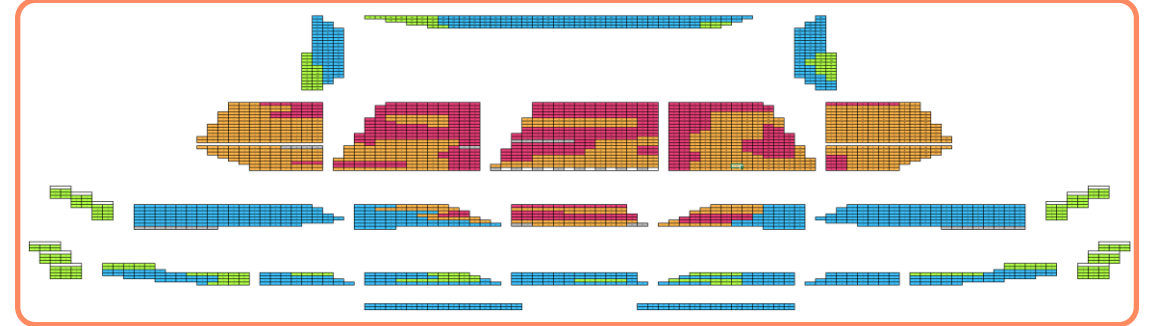
▶ 교향곡



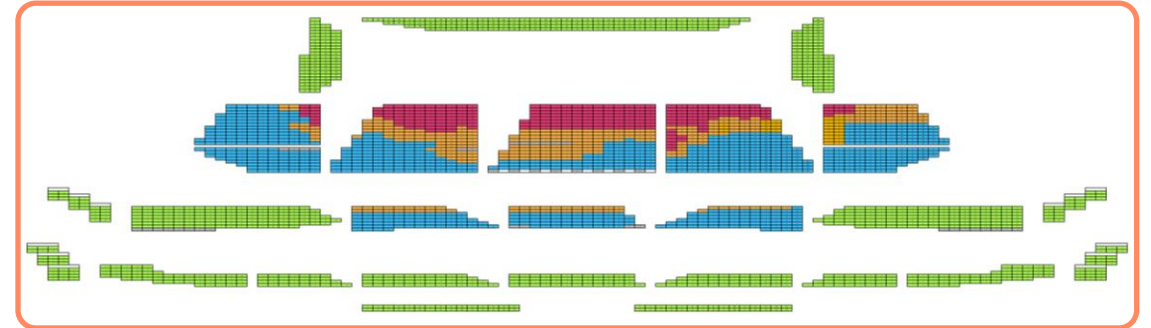
▶ 합창



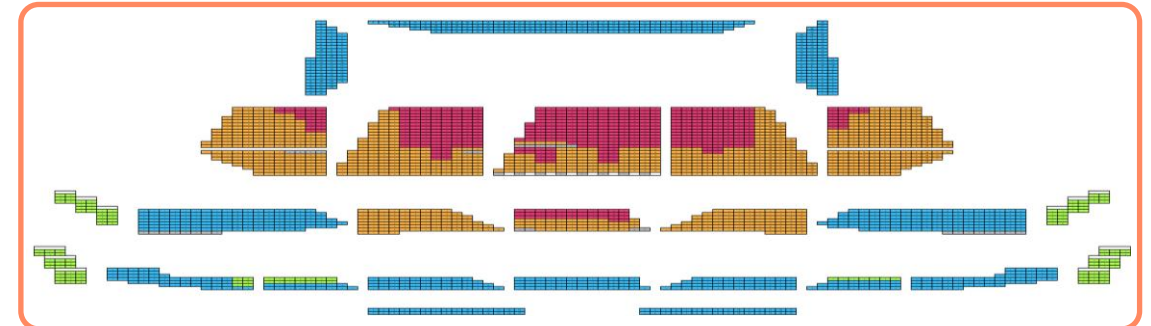
▶ 콘서트



▶ 오페라

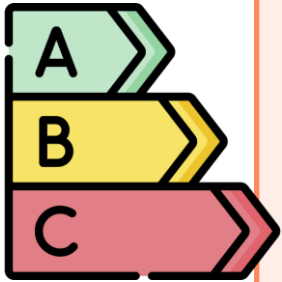


▶ 기타



05 활용 방안





장르별 자유로운 좌석 등급 및 가격 결정

- 본 프로젝트에서 구현한 알고리즘은 장르와 군집 개수를 지정할 수 있음
⇒ 공연 장르의 특수성 및 관객의 선호도를 고려하여 적절하고 자유로운 좌석 등급 개수 선택 가능
- 군집화 결과로 얻은 레이블은 군집을 구분할 뿐, 순서적인 의미를 가지고 있지 않음
⇒ 좌석배치도 시각화를 통해 좌석 등급의 구분을 확인하여 합리적인 가격을 책정 가능



다양한 프로모션 및 이벤트

- 합리적인 가격 정책을 제시하여 기관에 대한 관객의 신뢰도 상승 기대
- 새로운 좌석 등급 제도 도입 시, 개선 부분의 홍보와 걱정 유예 기간으로 관객의 혼선을 방지
- 무료 예매권을 제공하여 새로운 좌석 등급 제도에 대한 관객의 거부감 완화 및 참여 독려
- 다양한 장르에 대한 추천을 제시하여 상대적으로 수요가 낮은 장르의 접근성 향상
- 혜택을 받은 관객들을 대상으로 설문조사를 시행하여 새로운 좌석 등급과 가격에 대한 인식 조사

06 결론





[시사점]

- 관객의 주관적 요소인 선호도 변수와 좌석의 객관적 요소인 거리 변수를 사용하여 좌석 등급 분류를 위한 분석 과정과 결과의 타당성을 확보
- 기관에서 좌석 등급을 분류하고자 하는 장르와 등급 개수에 대응하는 다양하고 유연한 결과를 도출
- 구분된 각각의 좌석 등급을 좌석배치도에 다양한 색상으로 시각화하여 결과의 가독성 증가
- 좌석 등급에 맞는 가격 책정에 대한 자율성을 기관에게 부여하여 관객 유치를 유도할 수 있는 효과적인 판매 전략의 제안 및 합리적인 가격 산정 가능



[한계]

- 특정 장르에 대한 제공된 데이터 부족으로 장르를 통일하여 사용함으로써 구체적인 분석이 어려움
⇒ 더 많은 공연 예매 정보 데이터를 수집하여 장르별로 고유한 특성을 더욱 반영한 결과 도출
- 거리 정보 데이터를 2차원 좌석배치도를 통해 생성하여 현실 공간에 존재하는 실제 정보의 사용 부족
⇒ 3차원 실제 정보를 측정하여 반영함으로써 더 정확한 분석 가능
- 범주형 변수의 특성을 반영하는 알고리즘을 사용하면 오히려 이상치가 증가하는 현상이 발생
⇒ 여러 알고리즘을 결합함으로써 이상치를 줄이면서 성능을 향상시킬 수 있는 개선의 여지 존재



감사합니다!

