

공연예술통합전산망(KOPIS) 빅데이터 분석 공모전

온라인 사전설명회

2021. 06. 04. (금) 14:00 ~ 15:00



목차

I.

KOPIS(공연예술통합전산망) 소개

II.

「공연예술통합전산망(KOPIS)
빅데이터 분석 공모전」 소개

III.

KOPIS 데이터 소개

IV.

내·외부 데이터 활용 사례

V.

데이터 수집 가능 사이트 안내

VI.

질의응답(Q&A)

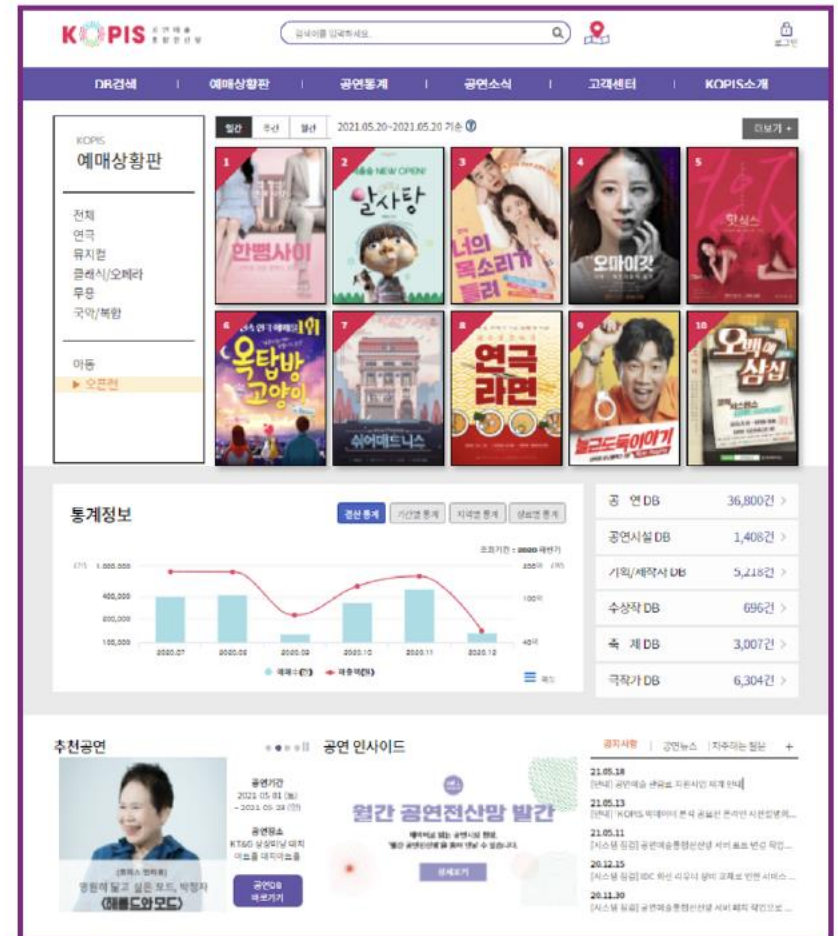
KOPIS(공연예술통합전산망) 소개

KOPIS

Korea Performing arts box office Information System

공연예술통합전산망(KOPIS)이란?

분산된 공연 입장권 예매/취소 정보를 집계하여 정확하고 신뢰성 있는 공연 정보와 통계 서비스를 제공하는 시스템 (공연법 제4조, 제43조에 의거 운영)



공모전 목적

- 빅데이터 활용 사업 추진을 통해 공연시장 활성화 및 고도화에 기여
- 공연시장 97%를 포괄하고 있는 KOPIS 데이터와 신기술 결합을 통해 혁신적 분석 결과 및 모형 산출
- 공연예술통합전산망 데이터 활용 가능 범위 탐색 및 새로운 시도의 장 마련

공모명

공연예술통합전산망(KOPIS) 빅데이터 분석 공모전

참가 자격

- 개인 또는 팀 (최대 4인)
- 학적부상 전공 무관한 대학(원)생

-
1. 1인 단독 지원 가능 - 주제 선정부터 분석 및 결과 발표까지 홀로 완수하셔야 합니다.
 2. 휴학생 지원 가능 - 휴학증명서 제출 필수입니다.
 3. 특수 대학원 재학생 지원 가능 - 소속 대학원 형태와 무관하게 지원 가능하며, 직장인 대학원생도 지원 가능합니다.
 4. 팀원 전체 학생으로 구성되어야 합니다.

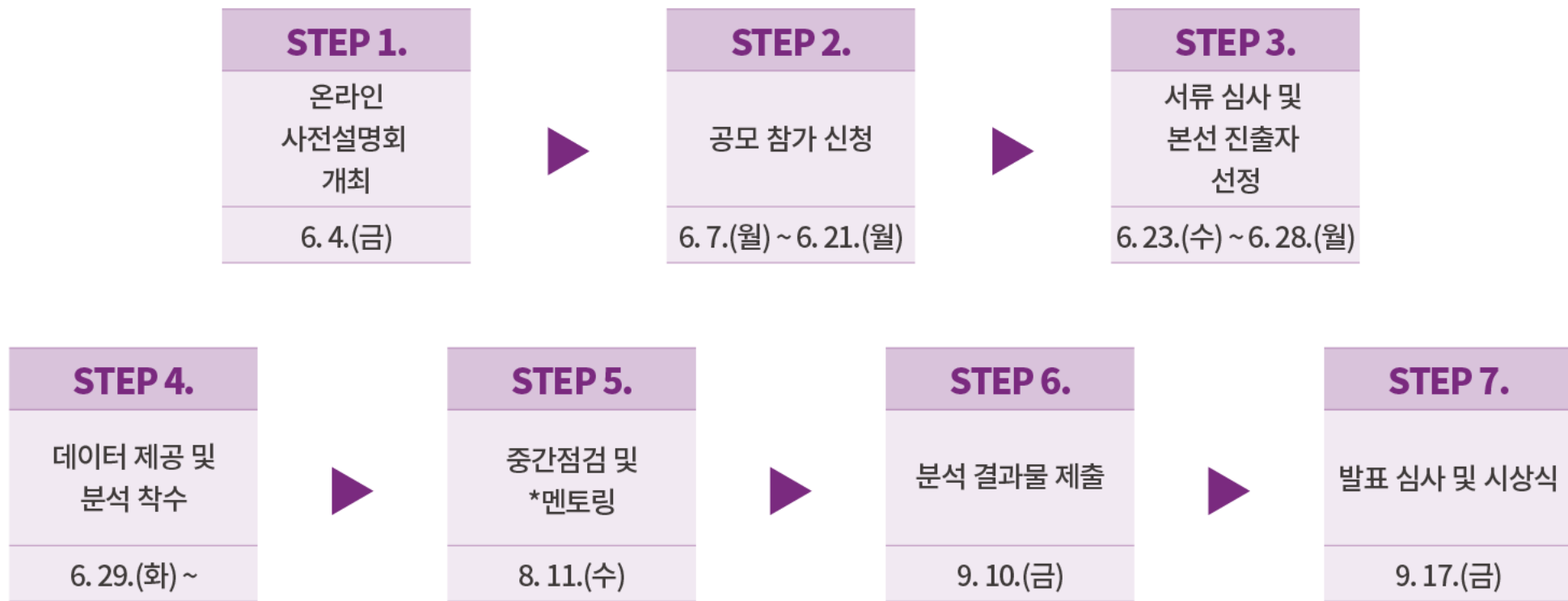
공모 주제

공연 소비, 유통 등의 측면에서 시장 문제점을 해결 및 발전시킬 수 있는 범주 내 자유 주제

주제 예시
공연 관람 수요 예측 모델 개발
공연 매출액에 영향을 미치는 주요 요인 도출
코로나19로 인한 공연계 현황 분석 및 뉴노멀 시대 예측
장르별 공연 추천 시스템 개발
공연예술계 영상 플랫폼 이용 현황 및 성과 분석
공연소비 패턴 분석

- 1차 제안서 주제와 2차 분석 주제는 상이하셔도 됨 (※ 주제는 데이터 수령 및 멘토링 후, 변경 가능)
- 공모 주제에서 벗어나거나 제안서 양식에 맞지 않는 작품은 심사에서 제외됨
- 순수 창작 아이디어가 아닌 표절 주제 및 결과물임이 적발될 경우 발생하는 모든 민·형사상의 책임은 참가자 본인에게 있음
- 타 공모전 수상작, 출판물을 도용 및 제출한 경우, 부정행위로 간주하며 수상 시, 취소 및 상금 전액을 환수 조치함
- **동일한 아이디어의 작품이 접수될 경우 도착 시점을 기준으로 우선 접수된 작품만 인정함**
(단, 수정 내용이 접수될 경우 수정 접수된 시점을 기준으로 함)

공모 일정



공모 일정 Q&A

Q1. 온라인 사전설명회 사전 접수 및 참석은 필수인가요?

A1. 아니요. 온라인 사전설명회는 공모 내용 및 접수 방법 등을 상세히 설명하여 참가자들의 이해도를 향상시키기 위한 목적일 뿐 필수사항은 아닙니다. 물론, 사전설명회 참석 여부가 서류 및 발표 심사에 영향을 주지도 않습니다.

Q2. 공모 접수 기간 외 접수는 가능한가요?

A2. 공모 접수 기간 전·후에 접수하신 경우는 심사 대상에서 제외됩니다. 반드시 기간 내에 접수해 주세요!

Q3. 중간점검 및 멘토링에서는 무얼 하나요?

A3. 지금까지의 진행 상황을 공유하고, 분석 시 막혔던 부분이나 조언이 필요했던 사항들에 대해 전문가에게 도움을 요청할 수 있습니다.

Q4. 데이터 제공 방식은 어떻게 되나요?

A4. 데이터는 서류 합격자를 대상으로 데이터 보안 유지 약정 체결 후, 메일로 제공될 예정입니다.

접수 방법 및 제출물

접수 방법: 이메일 접수(kopis@gokams.or.kr)

제출 방식

대상	제출 서류	제출 규격	기한
참가자 전체	참가신청서	제출 서류 붙임서식에 작성	6월 7일 ~ 6월 21일
	제안서		
	참가서약서		
	보안서약서		
	개인정보 수집이용 동의서		
서류 합격자	결과보고서 1부	붙임서식에 작성(한글(hwp)파일 제출)	6월 29일 ~ 9월 10일
	발표용 PPT 1부	별도 양식은 없으나, 결과보고서 내 항목이 누락되지 않도록 작성 (PDF파일 30페이지 이내 분량)	
	활용 데이터 및 코드 등	· Zip 파일 압축하여 제출 · 코드 전체(package 및 Function 포함)를 제출해야 함	
	활용 데이터 출처	붙임서식에 작성(한글(hwp)파일 제출)	
수상 팀	재학증명서	수상 팀 발표 후, 학적부상 학생 여부 확인을 위해 추가 요청 및 별도 제출	9월 17일 ~ 9월 24일

※ 재학증명서는 시상식 후 7일 이내에 제출하셔야 하며, 응모 자격 요건인 대학(원)생 여부 확인용으로만 사용됩니다.

분석도구 및 활용 데이터

분석 도구 R, Python, SQL 등 오픈소스

활용 데이터 KOPIS 데이터 사용은 필수이며, 외부 데이터와의 결합은 선택사항

별도 제공 데이터

1) KOPIS 보유 데이터 전반 (필수)

- 제공 데이터: '20년 1~12월 데이터
- 데이터 양: 약 1,600만 건 / 2.8~3.0 GB

2) 한국문화예술위원회 보유 데이터 일부 (선택)

- ※ 별도 제공 데이터는 서류 합격자에 한해 데이터 보안유지 약정서 체결 후 별도 제공 예정
- ※ 별도 제공 데이터 외의 모든 외부 데이터는 개별 수집 및 활용

「공연예술통합전산망(KOPIS)빅데이터 분석 공모전」 소개

첨부1. 제 안 서	
① 참가자 정보	
개 인 · 팀 명	
주 제(제목)	
② 기획서 작성	
• 개요	
1) 분석 목적	
2) 배경 및 필요성	
3) 차별성 및 독창성	
4) 주요 내용	
• 세부 내용	
1) 분석 방법	
- 활용 데이터, 분석 사용 도구, 프로세스, 분석 모형, 방법론, 알고리즘 등을 중심으로 구체적으로 작성 2) 분석 가능성 - 제안한 주제에 대해 분석/모델링 구현이 가능할 것이라고 생각하는 이유에 대해 논리적이고 구체적으로 작성 3) 적용 방향 - 제안한 주제 및 내용을 향후 KOPIS에 적용할 수 있는 방향성 - 활용 및 적용 가능한 부문 등 4) 기대효과 - 분석 결과물을 적용하였을 때 예측되는 효과, 적용 가능 부문 및 활용 가능성 등 • 기타 - 추가 기재하고 싶은 항목 및 내용 - 활용 데이터 출처 - 참고문헌	
※ 기획서 작성 시 유의사항	
- 분량은 10페이지 이내로 작성 - 추가 이미지 파일 등은 별도 첨부 가능 (PDF파일로 변환 후 제출) - 하단에 제시한 목차 외 추가 내용이 있을 경우 별도 타이틀을 기재하여 작성	

심사 및 시상

심사 방법: 외부 심사위원 5인 구성을 통한 평가

서류 심사: '21. 6. 23. (수) ~ 6. 25. (금) / 합격자 발표: '21. 6. 28. (월)

발표 심사: '21. 9. 17. (금) ※ 최종 수상 발표 및 시상식도 당일 진행

시상 규모: 총 6개 팀(또는 개인) 선정

*수상자 혜택: 시상금(200~500만 원) 지급 + 예술경영지원센터 대표 상장 수여

*입상자 혜택: 수상자 6팀 외, 1차 서류 합격자(=본선 진출자)에게는 입상금 30만 원 지급

1차 서류 합격자(최대 18팀) = 수상자(총 6팀) + 입상자(최대 12팀)

즉, 1차 서류 합격자만 되어도 분석 결과물 완수 및 제출 시, 입상금 30만 원 수령 가능

구분		대상	시상금
수상	최우수	1팀	500만 원
	우수	2팀	300만 원
	장려	3팀	200만 원
입상	입상	12팀	30만 원

※ 포상금 등 관련한 제세공과금은 수상자 본인 부담
 ※ 적격자(팀)가 없을 경우, 시상 규모는 축소될 수 있음

예매처 (위탁판매시스템)



공연시설 및 기획제작사 (자체예매시스템)



전송사업자코드	전송사업자명	공연시설코드	공연장코드	공연코드	공연일시	공연회차	입장권고유번호
SE000201	인터파크	20000360	20000360	21000894	2021/05/13 19:30:00	049	T0892054918
SE000101	예스24	2290	9799	38461	2021/04/24 15:00:00	1129278	1577271758400006
SE000307	멜론티켓	34	34	205772	2021/04/22 19:30:00	100071	M217538253101
SE000303	네이버예약	95942 126.8345321	임누리 어울림극장	502163_3850548	2021/05/23 14:00:00	0	143803259_0
SE000303	네이버예약	95942 126.8345321	임누리 어울림극장	502163_3850548	2021/05/23 14:00:00	0	143803259_1
SE000303	네이버예약	95942 126.8345321	임누리 어울림극장	502163_3850548	2021/05/23 14:00:00	0	143803259_2
SE000201	인터파크	18000580	18000580	21001625	2021/04/17 14:00:00	025	T1723060170
SE000201	인터파크	19001399	19001399	21002454	2021/05/22 14:00:00	003	T1725231140
SE000201	인터파크	19001399	19001399	21002454	2021/05/22 14:00:00	003	T1725231141

보유항목

[공연예술통합전산망 보유 데이터]

전송사업자코드	전송사업자명	공연시설코드	*공연시설명	시설특성	개관연도	주소	편의시설
장애시설	주차시설	주차시설세부	공연장코드	*공연장명	좌석수	무대시설	공연코드
*공연명	공연일시	입장권고유번호	예매일시	예매/취소구분	예매/취소매수	예매금액	예매/취소방식코드
예매/취소방식명	결제수단코드	결제수단명	할인금액	할인종류코드	할인종류명	장당금액	성별
연령	공연시작일자	공연종료일자	소요시간	장르명	세부장르명	공연지역명	출연진내용
제작진내용	기획제작사명	극작가명	수상실적	관람연령	아동공연여부	축제여부	내한공연여부
오픈런여부	단독판매여부	판매시작일시	판매종료일시	좌석등급명	판매페이지URL		

*제외 데이터

비식별화 코드

컬럼 정의서

필드명	구분	필드값 예시	세부 설명	비고
전송사업자코드	코드	SE000001	공연전산망 코드 분류 기준으로 'SE#####'로 표기	공연전산망 기준
전송사업자명	텍스트	인터파크	공연전산망과 연계되어 매일 데이터를 전송하는 업체를 뜻함	
공연시설코드	코드	FC000001	공연전산망 코드 분류 기준으로 'FC#####'로 표기	
시설특성	텍스트	민간(대학로)	해당 시설의 특성 국립, 공공(문예회관), 공공(기타), 민간(대학로), 민간(대학로 외), 민간(비공연장)으로 구분 [※ 붙임 0. 시설특성 분류 기준 참고]	
개관연도	숫자	2001	해당 시설의 개관년도 표기	
주소	텍스트	서울특별시 종로구 동숭길 41(동숭동)	해당 시설의 주소 표기	
편의시설	Y/N	Y	해당 시설 내 레스토랑/카페/편의점/놀이방/수유실 여부에 따라 Y/N값으로 표기	
장애시설	Y/N	Y	해당 시설 내 장애인 주차장/화장실/경사로/전용엘레베이터 여부에 따라 Y/N값으로 표기	
주차시설	텍스트	공영	자체/공영/NULL 여부에 따라 그 값에 표기	
주차시설 정보	텍스트	서울특별시 종로구 대학로 86 한국방송통신대학교 주차장	자체일 경우, 해당 시설 주소 표기/공영일 경우, 해당 시설 주변 공영주차장 주소 표기/NULL일 경우, NULL 표기	
공연장코드	코드	FC00001-01	공연전산망 코드 분류 기준으로 'FC#####-##'로 표기(시설의 하위 코드부여)	
좌석수	숫자	110	해당 공연장의 좌석수(객석수)	
무대시설 여부	Y/N	Y	해당 공연장 내 오케스트라 피트/연습실/분장실/무대넓이/장애인석 여부에 따라 Y/N값으로 표기	
공연코드	코드	PF000001	공연전산망 코드 분류 기준으로 'PF#####'로 표기	

비식별화 코드

컬럼 정의서

필드명	구분	필드값 예시	세부 설명	비고
공연일시	숫자	2021-03-01 10:00:00	결제한 공연의 일시로 'YYYY-MM-DD HH:MM:SS' 로 표기	전송사업자 기준
입장권고유번호	텍스트+숫자	T1726142390	결제 데이터당 다른 티켓(예매)번호를 부여(연계기관별로 각기 다름)	
예매일시	숫자	2021-01-01 18:01:02	결제한 일시로 'YYYY-MM-DD HH:MM:SS' 로 표기	
예매/취소건수	숫자	1	예매 : 1, 취소 : 2로 구분	
예매/취소매수	숫자	1	해당 데이터는 결제 건당의 데이터로 모두 1로 처리되어 있음	
예매금액	숫자	46000	결제/취소 시, 결제한 최종금액(할인 포함)	
예매/취소방식코드	숫자	01/02/03/04/99	01 : 웹, 02 : 모바일, 03 : 현장, 04 : 전화, 99 : 기타로 구분	
예매방식명	텍스트	인터파크 앱	예매/취소 방식 명칭(연계기관별로 각기 다를 수 있음)	
결제수단코드	숫자	02	01 : 현금, 02 : 카드, 03 : 상품권, 04 : 무통장, 05 : 다중결제, 99 : 기타로 구분	
결제수단명	텍스트	신용카드	결제수단 명칭(연계기관별로 각기 다를 수 있음)	
할인금액	숫자	20000	결제/취소 시, 받은 할인 금액	
할인종류코드	숫자	99	01 : 신용카드사, 02 : 통신사, 03 : 자체할인, 04 : 다중할인, 99 : 기타로 구분	
할인종류명	텍스트	조기예매할인	할인종류의 명칭(연계기관별, 예매를등록한 기획제작사별로 각기 다를 수 있음)	
장당금액	숫자	66000	권종가격(할인 등을 제외한 정가)	
성별	숫자	0	1 : 남, 2 : 여, 0 : 알 수 없음(제3자 동의를 받은 데이터만 남여구분 됨)	
연령	숫자	1990	출생연도로 'YYYY'로 표기(제3자 동의를 받은 데이터만 작성되어 있음. 미상이면 '0000'으로 표기)	
공연시작일자	숫자	2021-01-01	해당 공연의 시작일자 'YYYYMMDD'로 표기(시작일이 없을 경우 티켓등록일로 표기)	
공연종료일자	숫자	2021-01-31	해당 공연의 종료일자 'YYYYMMDD'로 표기(오픈런 등 종료일이 미지정일 경우, 99991231로 표기)	

컬럼 정의서

필드명	구분	필드값 예시	세부 설명	비고
소요시간	텍스트+숫자	1시간 20분	‘00시간00분’으로 기입	공연전산망 기준
장르명	텍스트	연극	연극, 뮤지컬, 클래식, 오페라, 무용, 국악, 복합으로 표기	
세부장르명	텍스트	마당극	장르의 세부장르로 표기 [※ 붙임 0. 장르분류 기준 참고]	
공연지역명	텍스트	강원도	시·도로 구분하여 표기	
출연진내용	텍스트	박정자, 김준수, 박은태, 김기민등	상세페이지 내 출연진 최대 7명 작성. 7명 이상일 시, 등으로 기입	
제작진내용	텍스트	강병원, 민새롬, 강수진 등	상세페이지 내 출연진 최대 3명 작성. 3명 이상일 시, 등으로 기입	
기획제작사명	텍스트	(재)예술경영지원센터(주관), 극단 미미(제작)	상세페이지 내 주최, 주관, 기획, 제작으로 구분하여 기입	
극작가명	텍스트	알베르 카뮈(원작), 정진세(각색)	원작자, 창작·원안·각색·윤색으로 구분하여 작성(번역은 작성하지 않음)	
수상실적	텍스트+숫자	2021 제12회 서울연극제 대상	상세페이지 내 표기된 수상작 기준으로 회차 / 수상명 / 수상실적 기입(제7회 2017 서울연극제 금상 등)	
관람연령	텍스트+숫자	만 7세 이상	만00세 이상, 00개월 이상, 전체 관람가 등으로 기입	
아동공연여부	Y/N	N	Y/N으로 구분	
축제여부	Y/N	Y	Y/N으로 구분	
내한공연여부	Y/N	N	Y/N으로 구분	
오픈런여부	Y/N	N	Y/N으로 구분	
판매시작일시	숫자	2021-01-01 18:01:02	해당 연계기관에서 판매 시작일시 ‘YYYYMMDD HH:MM:SS’로 표기	
판매종료일시	숫자	2021-01-31 23:01:02	해당 연계기관에서 판매 종료일시 ‘YYYYMMDD HH:MM:SS’로 표기	
좌석등급명	텍스트+숫자	R석-60000, S석-30000	좌석등급에 따른 정가 작성(전석-00,000, R석-00,000 등)	
판매페이지URL	URL	https://www.ntok.go.kr/kr/Ticket/Performance/Details?performanceId=265777	전송처에서 판매하고 있는 페이지URL	

내·외부 데이터 결합 분석 사례

1. 온라인 라이브 성과 및 온·오프라인 연관성 분석

활용 데이터

- KOPIS 데이터 (공연 예매, 취소 데이터)
- 라이브 공연 관련 크롤링 데이터

주요 내용

- 동일 작품의 온·오프라인 공연 성과 비교 (조회수/관객수)
- 온라인 공연 이후 오프라인 공연의 매출액 및 관객수 증감 추이 분석

〈표 1 온·오프라인 공연예매 성과지표 비교 결과〉

작품명	장르	오프라인 (회당 평균 관객수)	온라인 (1회 공연 조회수)	관객수 대비 조회수
비공개	국악	82	510	6.2배
	국악 평균	82	510	6.2배
	무용	168	682	3.6배
	무용	174	784	4.5배
	무용	125	629	5.0배
	무용 평균	162	698	4.4배
	유지철	5,670	11,395	2.0배
	유지철	1,697	19,679	12.0배
	유지철	355	95,515	269.1배
	유지철	597	2,256,487	3,781.3배
	유지철	595	3,783	6.3배
	유지철	420	8,832	21.0배
	유지철	325	24,715	76.0배
	유지철	325	80,636	248.0배
	유지철	87	5,094	58.9배
	유지철	228	14,656	64.2배
	유지철	151	14,402	95.6배
	유지철	224	21,747	97.1배
	유지철	155	13,635	88.0배
	유지철	134	6,545	51.2배
	유지철	54	4,134	75.7배
	유지철	75	25,804	344.5배
	유지철	27	6,235	232.6배
	유지철	6	3,555	592.5배
	유지철	2	1,121	494.6배
	유지철 평균	582	137,804	347.9배
	연극	206	2,970	14.4배
	연극	142	6,442	45.3배
	연극	87	5,434	62.4배
	연극	43	1,562	36.4배
	연극	41	1,992	48.9배
	연극	37	3,389	91.3배
	연극	13	4,191	322.2배
	연극	6	2,443	434.3배
	연극	4	3,359	876.3배
	연극 평균	64	3,530	215.7배
	오페라	51	784	15.4배
	오페라	3	1,821	507.0배
	오페라 평균	27	1,153	261.2배
	전체 평균	361	78,087	267.5배

〈표 4 온라인 공연 이후 오프라인 공연 매출액 및 관객수 증감률 (공연 간 비교)〉

장르	공연명	온라인	오프라인				매출액 증감률
		공연일시	공연일시	지역 (공연시설)	최대 평균 관객수	최대 평균 매출액	
유지철	비공개	2020-03-02	비공개				-
		2020-08-17					△32.1%
		2019-12-23					25.2%
유지철	비공개		비공개				-
							△73.5%
							△77.1%
연극	비공개	2019-12-11	비공개				-
							-
							6.65%

내·외부 데이터 결합 분석 사례

2. 대학로 공연 수요예측 모델 개발

활용 데이터

- KOPIS 데이터 (공연 관객수, 매출액, 좌석 점유율, 장르 등)
- 기상 정보 데이터
- 공연 관람 후기 관련 크롤링 데이터
- 휴일 정보 데이터
- 공연 및 출연진 인지도 관련 크롤링 데이터 등

주요 내용

- 공연 수요와 유의미한 상관관계를 보이는 인자 탐색
- 공연 특성 및 외부 환경 요인에 따른 수요 예측 모델 개발



4. 모델 개발 ○ 모델 성능 평가

종속 변수	평가 지표	연립 부패스트	XGBoost	Light GBM
관객수	결정계수 (R-Squared) 평균제곱근오차 (RMSE)	비공개		
매출액	결정계수 (R-Squared) 평균제곱근오차 (RMSE)			
좌석점유율	결정계수 (R-Squared) 평균제곱근오차 (RMSE)			

가장 결정계수가 높은 모델을 최적 모델로 선정
관객수, 매출액 예측을 위한 최적 모델은 연립 부패스트,
좌석 점유율 예측을 위한 최적 모델은 XGBoost 채택





kopis.or.kr

공연정보 DB, 공연 통계 데이터



culture.go.kr

문화예술, 문화유산, 문화산업,
관광, 체육 데이터



data.go.kr

정부기관 보유 데이터



data.seoul.go.kr

서울시 관련 수집 데이터



data.gg.go.kr

경기도 관련 수집 데이터