

# 천만 영화를 만드는 조건

천만과 비천만의 차이를 중심으로

## contents

---

1. 프로젝트 개요	03
2. 프로젝트 팀 구성 및 역할	05
3. 프로젝트 수행 절차 및 방법	06
4. 프로젝트 수행 결과	09
A. 데이터 수집	
B. 데이터 전처리	
C. 파생변수 생성	
D. 그래프 분석	
E. 기술통계량	
5. 결론	48
6. 자체 평가 의견	49

# 1. 프로젝트 개요

## A. 주제 및 선정 배경

3

영화진흥회 추산 71년 ~ 23년까지 누적 관객수 1000만 달성 업적은 현재 단 29개뿐일 정도로 매우 어려운 과제.

영화 산업 발전과 연관하여 한국 영화가 황금종려상 수상, 전세계 문화적으로 큰 파장을 일으킴. 하나 OTT의 등장으로 영화 업계의 지각 변동이 예상 이러한 시장상황을 고려하여 천만영화의 조건에 대해 살펴보고 국내 영화산업의 발전에 이바지 할 수 있는 데이터를 산출하고자 함



PALME D'OR  
FESTIVAL DE CANNES



MEGABOX  
LIFE THEATER™

apple tv+ hulu tving

Disney+ WATCHA PLAY

NETFLIX

## B. 훈련내용과의 관련성

주제 선정 중 수업에 배웠던 언어를 네이버 평점 크롤링에 활용하던 기억이 나서 이를 영화에 대입하면 어떨까라는 의견을 시작으로 갈피를 잡음 관련된 요소들이 많다고 느껴짐과 동시에 데이터를 수집하고 다듬는 과정까지 훈련중 배운 내용을 활용 할 수 있다는 생각에서 시작함



## [개별 환경]

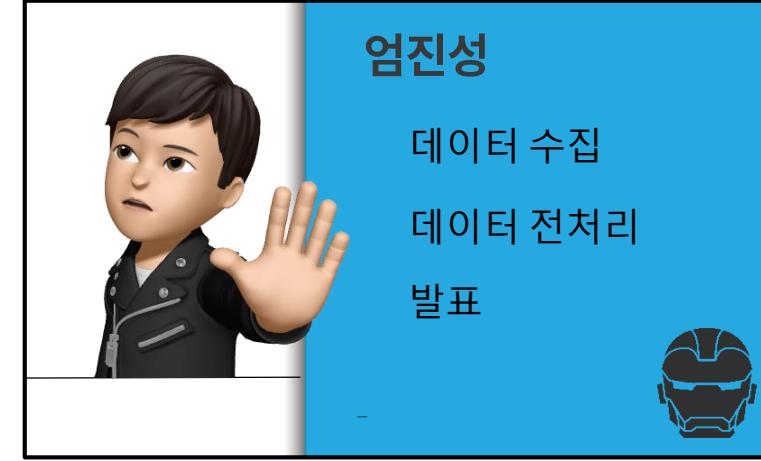
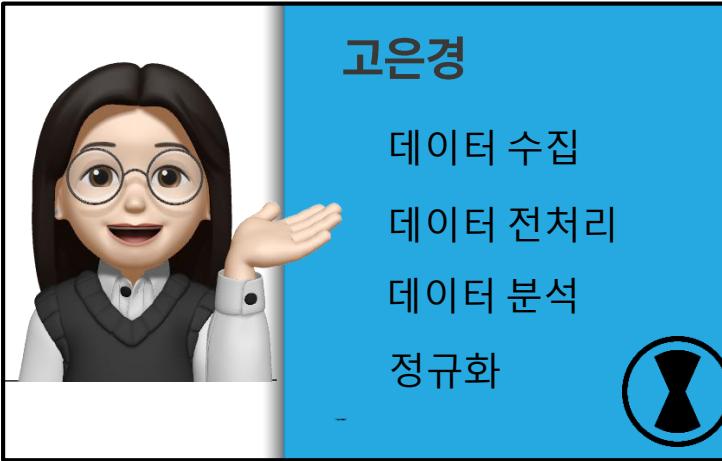
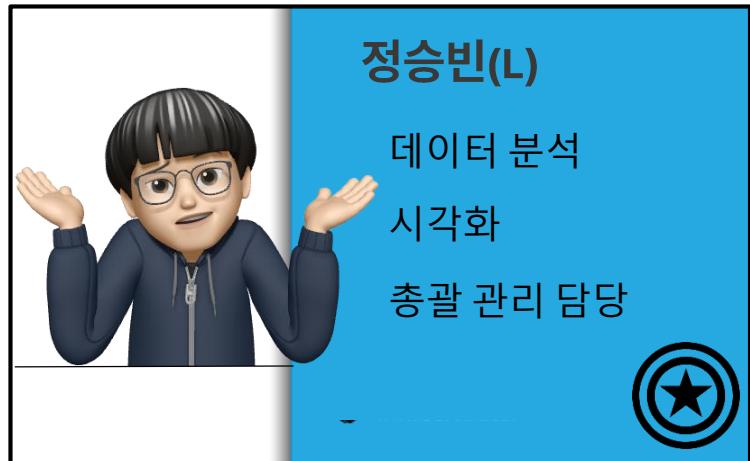


## 2. 팀 구성 및 역할

훈련생

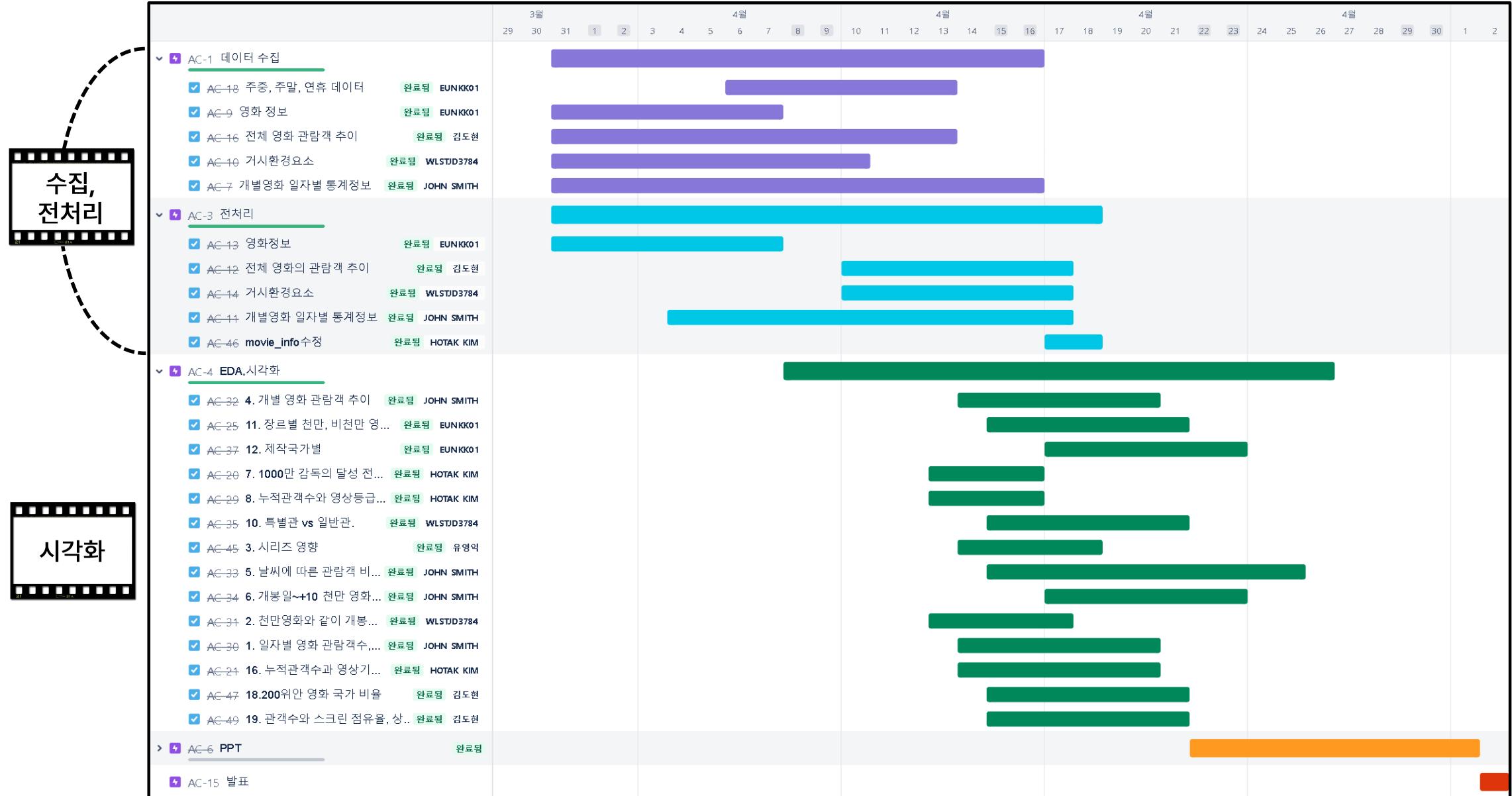
담당 업무

5



### 3. 프로젝트 수행 및 절차

6



## 4. 데이터 분석

### A. 데이터 수집

7



<https://www.kobis.or.kr>

- 겨울왕국 2 (Frozen 2)



기본정보		통계정보	박스오피스
		<input checked="" type="checkbox"/>	
		* 최종수정: 20	
코드	20197803		
ISAN	해당정보 없음		
A.K.A.	해당정보없음		
요약정보	장편   일반영화   애니메이션   103분 4:		
등급분류	ALL * 영상물등급위원회 제공		
개봉일	2019-11-21		
제작연도	2019년		
제작상태	개봉		
크랭크인/업	해당정보없음		
촬영화자	해당정보없음		
상영타입	2D : 디지털(20197803D), 디지털 더빙(3D : 3D 디지털(20197803S), 3D 더빙(4D : 4D 더빙(20197803B), 4D(201978IMAX : IMAX(20197803I), IMAX 3D(2C		

\*출처 : KOBIS(발권)통계 \*통계기준일 : 2023.04.23 엑셀

날짜	스크린수	상영횟수	매출액	관객수	누적매출액	누적관객수	순위
개봉이전	4	4	11,083,000	1,165	11,083,000	1,165	N/A
개봉1일(11/21)	2,343	12,998	4,943,457,300	606,618	4,954,540,300	607,783	1
개봉2일(11/22)	2,331	13,387	5,560,526,180	632,547	10,515,066,480	1,240,330	1
개봉3일(11/23)	2,642	① 16,220	① 14,318,356,970	① 1,661,835	24,833,423,450	2,902,165	1
개봉4일(11/24)	① 2,648	16,012	13,191,536,350	1,535,569	38,024,959,800	4,437,734	1
개봉5일(11/25)	2,190	12,415	2,896,394,520	353,767	40,921,354,320	4,791,501	1
개봉6일(11/26)	2,226	12,315	2,594,195,740	319,416	43,515,550,060	5,110,917	1
개봉7일(11/27)	2,094	10,610	3,665,481,480	599,585	47,181,031,540	5,710,502	1
개봉8일(11/28)	1,922	10,105	2,298,128,030	287,605	49,479,159,570	5,998,107	1
개봉9일(11/29)	1,983	10,757	3,760,389,730	431,912	53,239,549,300	6,430,019	1
개봉10일(11/30)	2,368	13,754	10,046,344,730	1,173,934	① 63,285,894,030	① 7,603,953	① 1

## 4. 데이터 분석

### A. 데이터 수집

8

컬럼명	[1] 영화별 정보								
Title	Title	Release	Genre	Age	Nation	Type	Director	Main_actor	
Release	0	명량	2014-07-30	사극,액션	15세이상관람가	한국	[필름, 2D, 4D]	[김한민]	[최민식(이순신), 류승룡(구루지마), 조진웅(와키자카)]
	1	극한직업	2019-01-23	코미디	15세이상관람가	한국	[2D]	[이병헌]	[류승룡(고반장), 이하늬(장혈사), 진선규(마혈사), 이동휘(영호), 공명(재훈)]
Genre	2	신과함께-죄와 벌	2017-12-20	판타지,드라마	12세이상관람가	한국	[2D, 4D]	[김용화]	[하정우(강림), 차태현(자홍), 주지훈(해원맥), 김향기(덕춘), 마동석(성주신)]
	3	국제시장	2014-12-17	드라마	12세이상관람가	한국	[필름, 2D]	[윤제균]	[활정민(덕수), 김윤진(영자), 오달수(달구), 정진영(덕수부)]
Age	4	어벤져스: 엔드 게임	2019-04-24	액션,SF	12세이상관람가	미국	[2D, 3D, 4D, IMAX]	[안소니 루소, 조 루소]	[로버트 다우니 주니어, 크리스 에반스, 마크 러팔로, 크리스 헬스워스, 스칼렛 요...]
	5	겨울왕국 2	2019-11-21	애니메이션	전체관람가	미국	[2D, 3D, 4D, IMAX]	[크里斯 베, 제니 퍼리]	[크리스틴 벨, 이디나 멘젤, 조시 게드, 조나단 그로프]
Type	6	베테랑	2015-08-05	액션,드라마	15세이상관람가	한국	[필름, 2D]	[류승완]	[활정민(서도철), 유아인(조태오), 유해진(최 상무), 오달수(오팀장)]
	7	아바타	2009-12-17	SF,액션,어드 벤처	12세이상관람가	미국	[필름, 2D, 3D, 4D, IMAX]	[제임스 카메론]	[샘 워싱턴, 조 셸다나, 시고니 위버]
Director	8	도둑들	2012-07-25	액션,범죄	15세이상관람가	한국,홍콩	[필름, 2D]	[최동훈]	[김윤석(마카오박), 김혜수(팝시), 이정재(뽀빠이), 전지현(예니콜), 임달화(전...)]
	9	7번방의 선물	2013-01-23	코미디,드라마	15세이상관람가	한국	[필름, 2D]	[이환경]	[류승룡(을구), 박신혜(큰 예술), 갈소원(어린 예술), 오달수(방장), 박원상(...)]
Main_actor	총 개수 : 200개								
	<a href="https://www.kobis.or.kr">https://www.kobis.or.kr</a>								

## 4. 데이터 분석

### A. 데이터 수집

9

컬럼명
날짜
스크린수
스크린 점유율
상영횟수
상영 점유율
좌석수
좌석점유율
좌석판매율
매출액
매출액증감(전일대비)
관객수
관객수증감(전일대비)
누적매출액
누적관객수
순위
<b>총 개수 : 8069개</b>

### [2] 영화별 일자에 따른 통계 정보

▶ 영화가 상영된 총 기간 내 통계 정보 수집(스크린 점유율, 상영횟수, 매출액, 관객수, ...)

날짜	스크린수	스크린 점유율	상영 횟수	상영점유율	좌석수	좌석점유율	좌석판매율	매출액	매출액증감(전일대비)	관객수	관객수증감(전일대비)	누적매출액	누적관객수	순위
9 2014-07-30	1159	0.336	6147	0.423	1152173	0.484	0.593	4708879000	4649246000 (-7796.4%)	682701	674628 (8356.6%)	4881110000	705201	1
10 2014-07-31	1202	0.318	6258	0.421	1176034	0.483	0.600	5243409900	534530900 (+11.4%)	705070	22369 (3.3%)	10124519900	1410271	1
11 2014-08-01	1300	0.337	6699	0.444	1255037	0.509	0.691	6918410307	1675000407 (+31.9%)	867437	162367 (23.0%)	17042930207	2277708	1
12 2014-08-02	1494	0.376	7605	0.491	1402659	0.553	0.879	9826541209	2908130902 (+42.0%)	1232529	365092 (42.1%)	26869471416	3510237	1
13 2014-08-03	1587	0.398	7963	0.521	1456645	0.582	0.863	10031212500	204671291 (+2.1%)	1257380	24851 (2.0%)	36900683916	4767617	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
124 2015-04-30	1	0.000	4	0.000	1144	0.000	0.115	393000	-273000 (-41.0%)	131	-91 (-41.0%)	135750030910	17614226	33
125 2015-05-01	1	0.000	4	0.000	1144	0.000	0.143	496000	103000 (26.2%)	164	33 (25.2%)	135750526910	17614390	30
126 2015-05-02	1	0.000	4	0.000	1144	0.000	0.093	318000	-178000 (-35.9%)	106	-58 (-35.4%)	135750844910	17614496	37
127 2015-05-03	1	0.000	4	0.000	1144	0.000	0.082	282000	-36000 (-11.3%)	94	-12 (-11.3%)	135751126910	17614590	32
128 2015-05-04	1	0.000	4	0.000	1144	0.000	0.078	267000	-15000 (-5.3%)	89	-5 (-5.3%)	135751393910	17614679	36

<https://www.kobis.or.kr>

## 4. 데이터 분석

## A. 데이터 수집

10

컬럼명	
한국	개봉편수
	상영편수
	매출액
	관객수
	점유율
외국	개봉편수
	상영편수
	매출액
	관객수
	점유율
전체	개봉편수
	상영편수
	매출액
	관객수

### [3] 일별 총 영화 관객수

▶ 일별 총 관객수 대비 특정 영화(10M 달성) 관객 수의 비중을 확인하고자 총 관객 수 데이터 수집

연도	한국				외국				전체						
	개봉편수	상영편수	매출액	관객수	점유율	개봉편수	상영편수	매출액	관객수	점유율	개봉편수	상영편수	매출액	관객수	
0	2004-01-01 00:00:00	0	4	478781500	74454	0.654	1	3	248223500	39383	0.346	1	7	727005000	113837
1	2004-01-02 00:00:00	0	4	404739500	65382	0.627	0	3	237537000	38975	0.373	0	7	642276500	104357
2	2004-01-03 00:00:00	0	4	502095000	78316	0.625	0	3	294424500	46918	0.375	0	7	796519500	125234
3	2004-01-04 00:00:00	0	4	488389000	76323	0.635	0	3	274882000	43880	0.365	0	7	763271000	120203
4	2004-01-05 00:00:00	0	4	278030000	43420	0.677	0	3	1312	date week day				7 4092	0 2003-12-01
5	2004-01-06 00:00:00	0	4	276711000	43329	0.686	0	3	1249	0	2003-12-01	0	Mon	7 4016	0 2003-12-01
6	2004-01-07 00:00:00	0	4	265123000	41509	0.687	0	3	1197	1	2003-12-02	0	Tue	7 3848	1 2003-12-02
7	2004-01-08 00:00:00	0	3	236467500	37429	0.574	0	4	1720	2	2003-12-03	0	Wed	7 4084	2 2003-12-03
8	2004-01-09 00:00:00	0	3	278072500	45356	0.551	1	4	2275	3	2003-12-04	0	Thu	7 5056	3 2003-12-04
9	2004-01-10 00:00:00	0	3	428365500	66515	0.533	0	4	3706	4	2003-12-05	0	Fri	7 7989	4 2003-12-05
										5	2003-12-06	1	Sat	...	6969 2022-12-06
										6	2003-12-07	1	Sun		6970 2022-12-07
										7	2003-12-08	0	Mon		6971 2022-12-08
										8	2003-12-09	0	Tue		6972 2022-12-09
										9	2003-12-10	0	Wed		6973 2022-12-10
														6974 rows x 14 columns	

<https://www.data.go.kr>

컬럼명
연도
전체개봉편수
전체상영편수
전체매출액
전체관객수
요일
가격
holiday
<b>총 개수 : 7036개</b>

## [4] 영화 티켓 가격 설정

▶ 주중, 주말, 공휴일에 따라 영화 티켓 가격 상이, 이에 따라 OpenAPI를 이용하여 공휴일 정보 수집

	연도	전체개봉편수	전체상영편수	전체매출액	전체관객수	요일	가격	holiday
0	2004-01-01	1	7	727005000	113837	Thu	7000	1
1	2004-01-02	0	7	642276500	104357	Fri	7000	0
2	2004-01-03	0	7	796519500	125234	Sat	7000	0
3	2004-01-04	0	7	763271000	120203	Sun	7000	0
4	2004-01-05	0	7	409269500	64180	Mon	7000	0
...	...	...	...	...	...	...	...	...
7031	2023-04-02	0	80	3423512979	336191	Sun	14000	
7032	2023-04-03	0	87	802358725	81632	Mon	13000	
7033	2023-04-04	2	91	815279367	82221	Tue	13000	
7034	2023-04-05	6	94	1347838240	134732	Wed	13000	
7035	2023-04-06	7	100	1058649412	107135	Thu	13000	

7036 rows × 8 columns

	holiday	date
0	신정	20040101
1	설날	20040121
2	설날	20040122
3	설날	20040123
4	삼일절	20040301

<https://www.data.go.kr/data/15012690/openapi.do>

컬럼명
컬럼
천만영화
동기간 상영영화
총 개수 : 58개

### [5] 동기간 상영 영화 비교

▶ 천만영화와 동기간에 상영된 영화의 관람객 수를 조회하기 위해 기간 대입, 리스트 결과 수집 JSON형식 저장

```
{
  '명량': ['명량', '해적: 바다로 간 산적'],
  '극한직업': ['극한직업'],
  '신과함께-죄와 벌': ['신과함께-죄와 벌', '1987'],
  '국제시장': ['국제시장'],
  '어벤져스: 엔드게임': ['어벤져스: 엔드게임', '기생충', '알라딘'],
  '겨울왕국 2': ['겨울왕국 2', '백두산'],
  '베테랑': ['베테랑', '암살'],
  '아바타': ['아바타'],
  '도둑들': ['도둑들', '광해, 왕이 된 남자'],
  '7번방의 선물': ['7번방의 선물', '베를린'],
  '알라딘': ['알라딘', '기생충', '엑시트', '스파이더맨: 파 프롬 홈'],
  '암살': ['베테랑', '암살'],
  '범죄도시2': ['범죄도시2', '탑건: 매버릭'],
  '광해, 왕이 된 남자': ['광해, 왕이 된 남자'],
  '신과함께-인과 연': ['신과함께-인과 연'],
  '택시운전사': ['택시운전사'],
  '태극기 휘날리며': ['태극기 휘날리며', '실미도'],
  '부산행': ['부산행', '밀정'],
  '변호인': ['변호인', '겨울왕국', '수상한 그녀'],
  '해운대': ['해운대', '국가대표'],
  '어벤져스: 인피니티 워': ['어벤져스: 인피니티 워'],
  '실미도': ['태극기 휘날리며', '실미도'],
  '괴물': ['괴물'],
  '아바타: 물의 길': ['아바타: 물의 길'],
  '왕의 남자': ['왕의 남자'],
  '어벤져스: 에이지 오브 울트론': ['어벤져스: 에이지 오브 울트론'],
  '인터스텔라': ['인터스텔라', '국제시장'],
  '기생충': ['알라딘', '기생충', '스파이더맨: 파 프롬 홈'],
  '겨울왕국': ['겨울왕국', '수상한 그녀', '변호인'],
  ...
}
```

천만영화 1~58위 동기간 상영영화.HEAD(29)

<https://www.kobis.or.kr>

컬럼명	
영화명	
영화명(영문)	
제작연도	
제작국가	
장르	
역할	
공식통계 매출액	
공식통계 관객수	
KOBIS통계매출액	
KOBIS통계관객수	
총 개수 : 129개	

## [6] 감독의 영향력

▶ 한국영화 흥행의 기준선, 천만영화감독의 영향력을 파악을 위한 감독별 필모그래피 수집

봉준호

영화명	영화명(영문)	제작연도	제작국가	장르	역할	공식통계 매출액	공식통계 관객수	KOBIS통계 매출액	KOBIS통계 관객수
0 아바타: 물의 길	Avatar: The Way of Water	2022	미국	액션, 어드벤처, SF, 스릴러	감독	1.375873e+11	10801074.0	1.376395e+11	10804506.0
1 아바타	Avatar	2009	미국	SF, 액션, 어드벤처	감독	1.284471e+11	13624328.0	1.253043e+11	13338863.0
2 타이타닉	Titanic	1997	미국	드라마, 멜로/로맨스, 액션	감독	0.000000e+00	1971780.0	1.143677e+10	883847.0
3 트루 라이즈	True Lies	1994	미국	스릴러, 액션, 어드벤처, 코미디	감독, 시나리오(각본)	0.000000e+00	874664.0	1.940000e+05	76.0
4 터미네이터 2 3D	Terminator II	1991	미국	액션, SF, 스릴러	감독	0.000000e+00	919444.0	1.353848e+08	14391.0
5 심연	The Abyss	1989	미국	SF, 드라마, 스릴러, 액션, 어드벤처	감독, 시나리오(각본)	0.000000e+00	80558.0	2.600000e+05	45.0
6 에이리언 2	Aliens	1986	미국	SF, 액션	감독, 시나리오(각본)	0.000000e+00	292436.0	1.575700e+07	2013.0
7 터미네이터	Terminator	1984	미국	SF, 액션	감독	0.000000e+00	318828.0	1.080000e+05	24.0

제임스 카메룬

영화명	영화명(영문)	제작연도	제작국가	장르	역할	공식통계 매출액	공식통계 관객수	KOBIS통계 매출액	KOBIS통계 관객수
0 기생충	PARASITE	2019	한국	드라마	감독, 음악진행, 시나리오(각본)	8.745983e+10	10313086.0	8.746397e+10	10313519.0
1 설국열차	Snowpiercer	2013	한국, 체코	SF, 액션	감독, 시나리오(각본)-각본, 스토리-스크린스토리	6.701008e+10	9349991.0	6.703502e+10	9353799.0
2 마더	Mother	2009	한국	드라마	감독, 시나리오(각본)-각본, 스토리	2.019537e+10	3013523.0	1.999011e+10	2982025.0

3 도쿄!	Tokyo!	2008	일본, 한국, 프랑스	드라마	감독, 시나리오(각본)	2.759559e+08	41198.0	2.784405e+08	41572.0
4 괴물	The Host	2006	한국	SF, 가족, 드라마	감독, 시나리오(각본)-각본	0.000000e+00	13019740.0	6.671770e+10	10917400.0
5 플란다스의 개	Barking Dogs Never Bite	2000	한국	드라마, 코미디	감독, 시나리오(각본)	0.000000e+00	57469.0	4.416200e+06	609.0

<https://www.kobis.or.kr>

컬럼명																
계정항목별																
국내총생산	계정 항목별	국내총생산(명목, 원화 표시) (십억 원)	국내총생산(명목, 원화 표시) (억달러)	국민총소득(명목, 원화 표시) (십억 원)	국민총소득(명목, 원화 표시) (억달러)	요소비용국민소득(명목) (십억 원)	국민처분가능소득(명목) (십억 원)	국민총처분 가능소득(명목) (십억 원)	가계총처분 가능소득(명목) (십억 원)	1인당 국내총생산(명목, 원화 표시) (만원)	1인당 국내총생산(명목, 원화 표시) (달러)	1인당 국민총소득(명목, 원화 표시) (만원)	1인당 국민총소득(명목, 원화 표시) (달러)	1인당 가계총처분 가능소득(명목, 원화 표시) (만원)	1인당 가계총처분 가능소득(명목, 원화 표시) (달러)	
국내총생산(\$)		2003	837365.0	7025.5	834443.2	7001.0	602692.2	689409.2	831292.4	494340.2	1748.4	14669.4	1742.3	14618.2	1032.2	8660.1
국민총소득		2004	908439.2	7936.3	906864.7	7922.5	662698.2	750340.6	904224.4	532359.3	1889.3	16505.5	1886.1	16476.9	1107.2	9672.5
국민총처분가능소득		2005	957447.8	9347.2	950685.4	9281.2	691776.2	784904.9	948273.9	556248.7	1987.0	19398.8	1973.0	19261.8	1154.4	11270.2
가계총처분가능소득		2006	1005601.5	10524.2	1002664.7	10493.5	728236.3	827011.2	999356.0	581833.5	2076.0	21727.1	2070.0	21663.7	1201.2	12571.1
1인당국내총생산		2007	1089660.2	11726.9	1086897.3	11697.1	792507.9	901343.3	1084398.3	617108.9	2238.2	24087.9	2232.6	24026.8	1267.6	13641.7
1인당국내총생산(\$)		2008	1154216.5	10468.2	1154509.7	10470.9	828123.1	946644.4	1154672.2	658676.7	2352.9	21339.9	2353.5	21345.3	1342.7	12178.0
1인당국민총소득																
1인당국민총소득(\$)																
1인당가계총처분가능소득																
1인당가계총처분가능소득(\$)																
총 개수 : 20개																

## [7-1] 국민총소득, 문화여가 지출률 데이터

▶ 영화 외적으로 관객수와 관련 요소 확인을 위한 2003. 12 ~ 2022.12 데이터 수집

<https://www.kobis.or.kr>

컬럼명
날짜
평균기온
최저기온
최고기온
강수량
<b>총 개수 : 6971개</b>

## [7-2] 날씨 데이터

▶ 영화 외적으로 관객수와 관련 요소 확인을 위한 2003. 12 ~ 2022.12 데이터 수집

	날짜	평균기온	최저기온	최고기온	강수량
0	2003-12-01	7.1	2.9	13.6	0.0
1	2003-12-02	6.6	2.6	12.0	0.0
2	2003-12-03	4.7	-0.4	11.2	0.0
3	2003-12-04	3.2	-1.8	9.9	0.0
4	2003-12-05	6.7	1.0	11.4	4.0

<https://www.data.kma.go.kr>

## [절차 및 방식]

### [1] 기초 작업(1차 전처리)

A. NULL값 확인, 연산을 위한 %, [ ] 등의 특수기호 삭제, MERGE를 위한 컬럼명 통일 및 이중 컬럼 삭제

### [2] 영화장르 분할

A. 영화마다 복합적 장르(액션, 드라마 등)를 가지기에 각각 하나의 장르로 분할, LONG형으로 데이터프레임 변환

### [3] 데이터 기간 설정

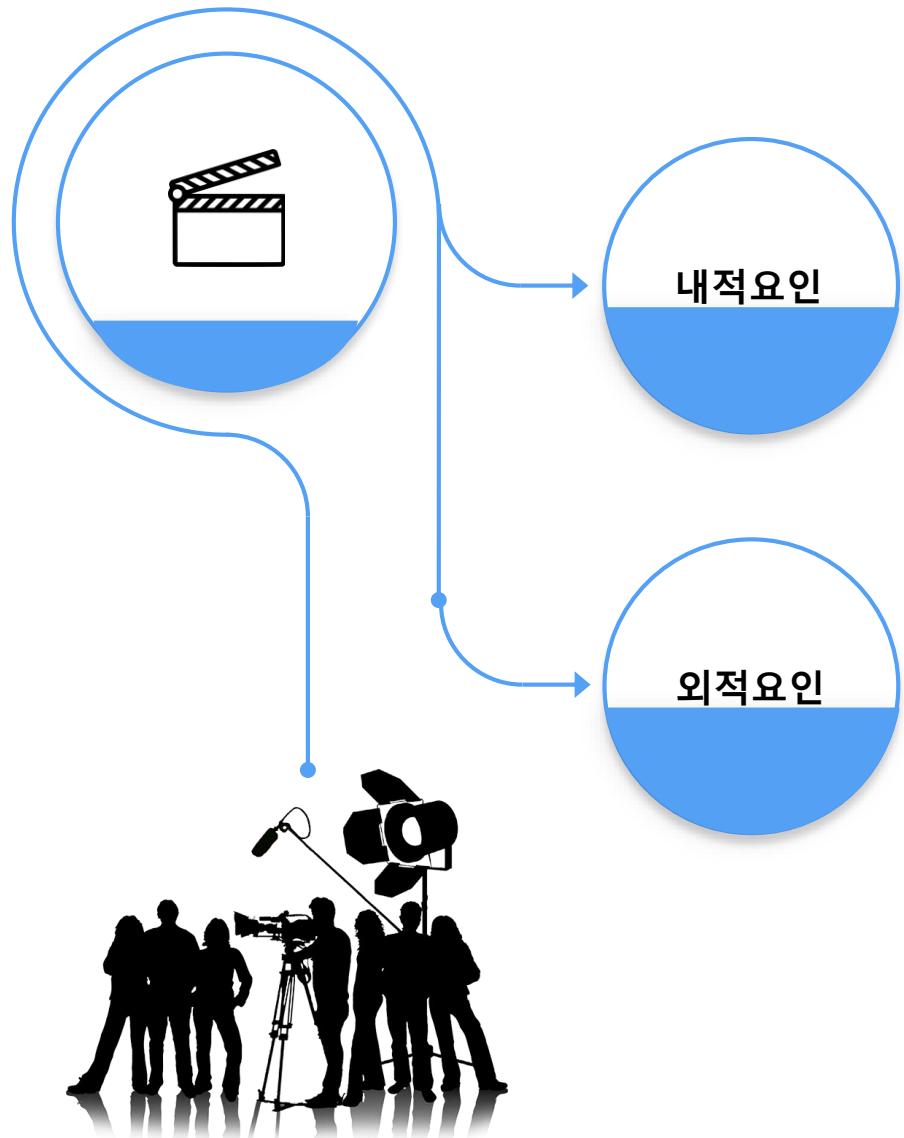
A. 영화별 정식 개봉일 전 시사회, 상영 종료 후 재개봉 등 점유율 1%미만 데이터 삭제

### [4] 영화 티켓 가격 설정

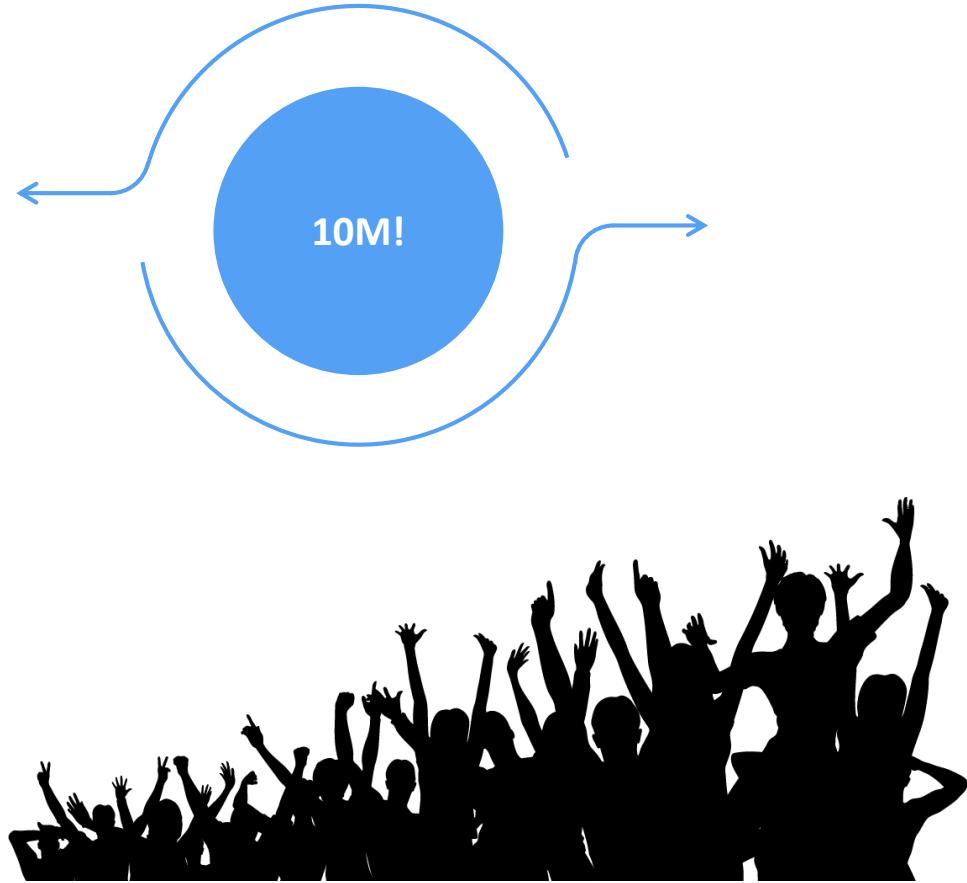
A. 시기, 멀티플렉스에 따라 가격 변동으로 인한 주중, 주말, 공휴일별 가격상승 폭 반영, 파생 변수 추가

### [5] 감독별 필모그래피

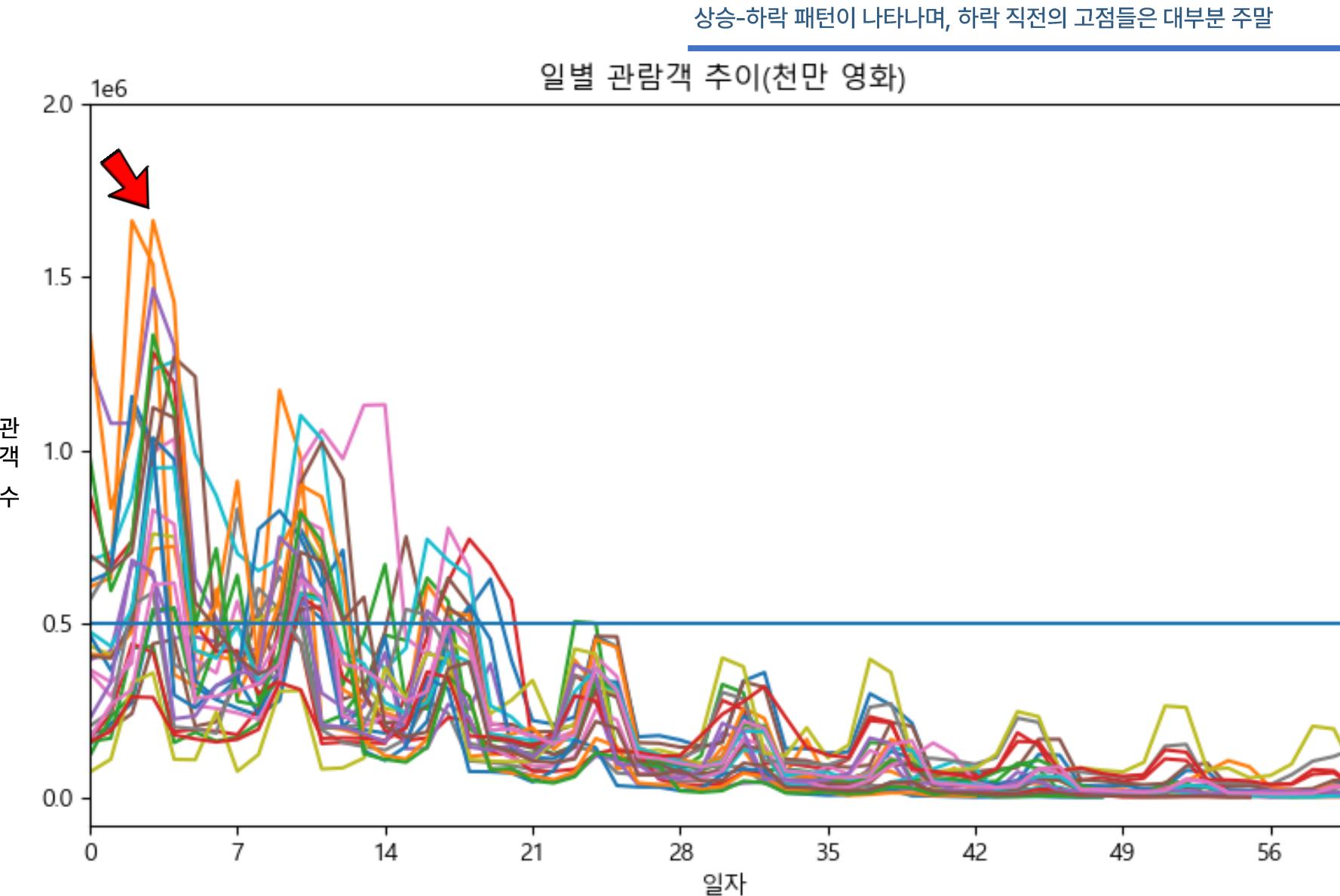
A. 데이터 상 감독에 따라 직접, 간접적으로 참여한 영화들에 대하여 간접 데이터 삭제

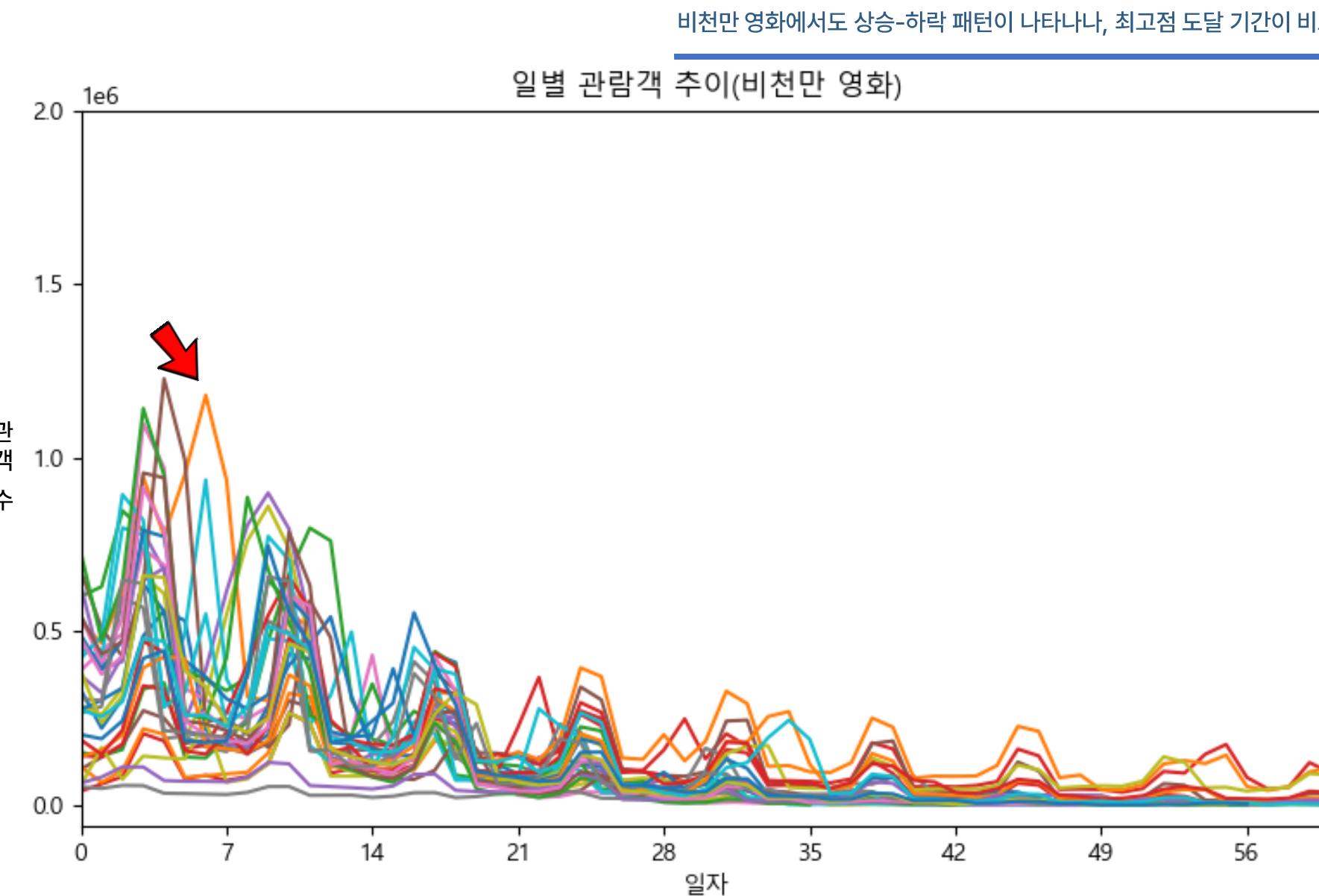


- ▶ 패 턴
- ▶ 장 르
- ▶ 연령등급
- ▶ 감 독
- ▶ 기 온 + 가 격
- ▶ 요 일
- ▶ 국민총소득



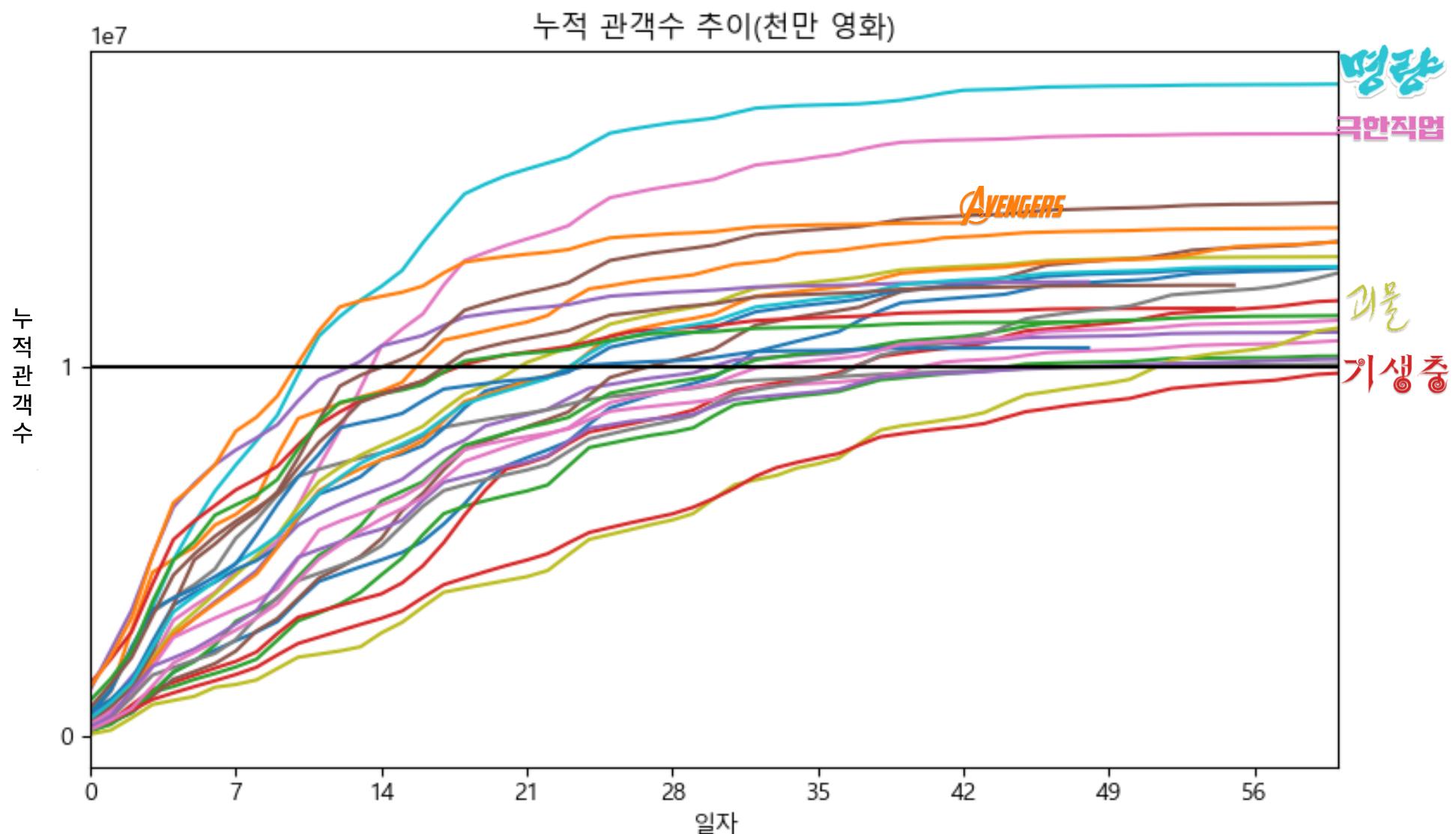


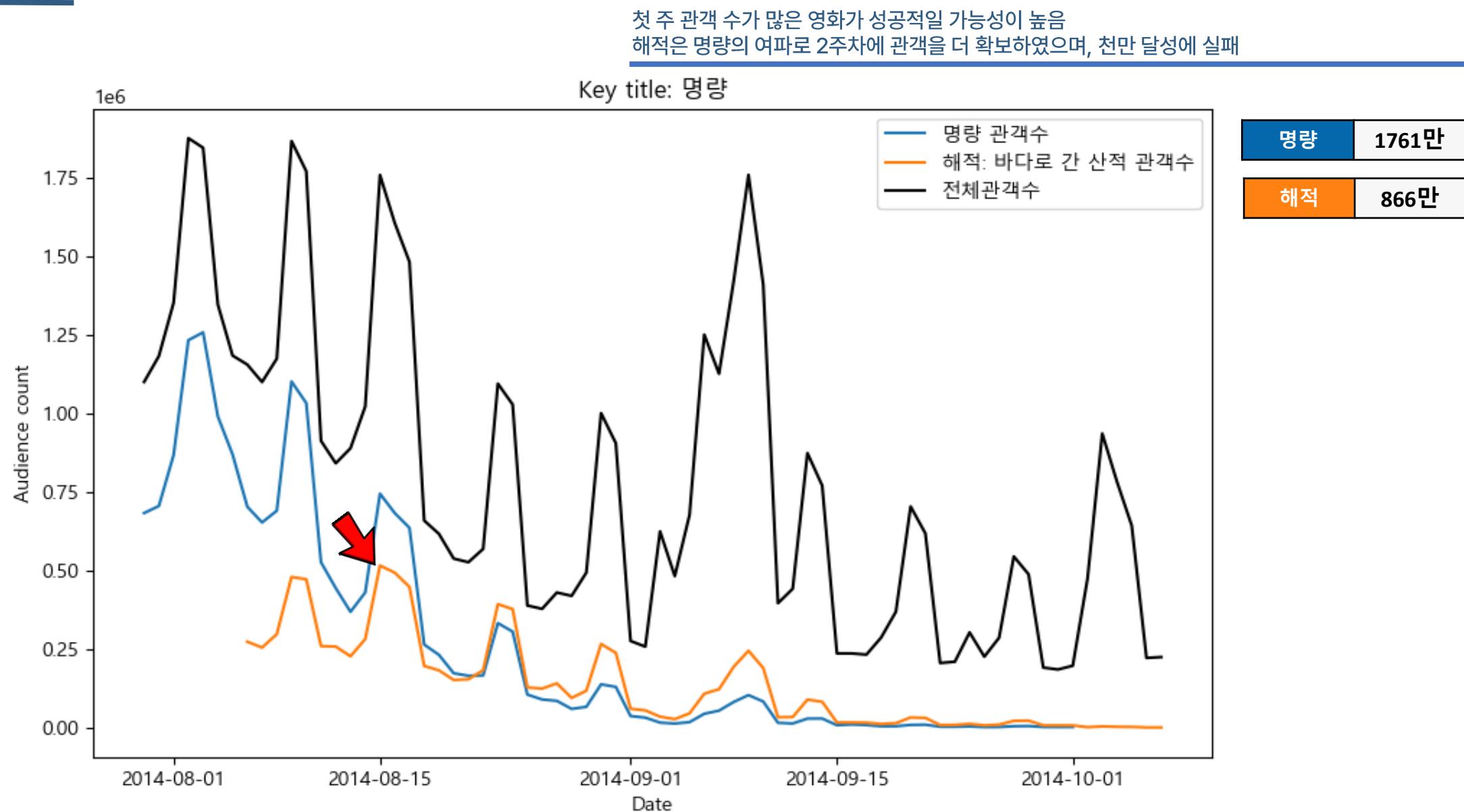






천만 영화를 달성하기까지 평균 30일, 빠르면 11일 소요  
초반 관객 수를 바탕으로 스노우볼 효과

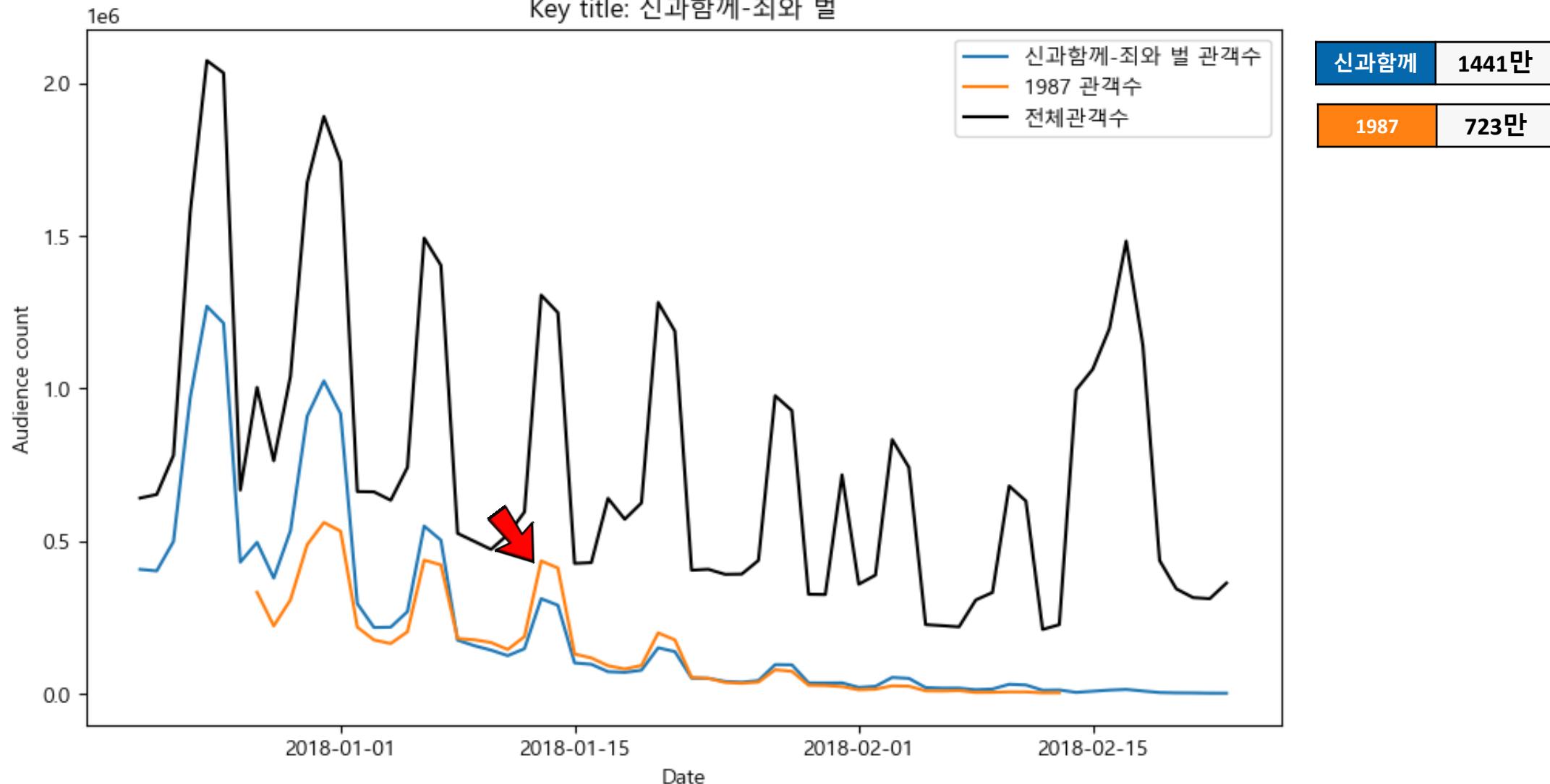




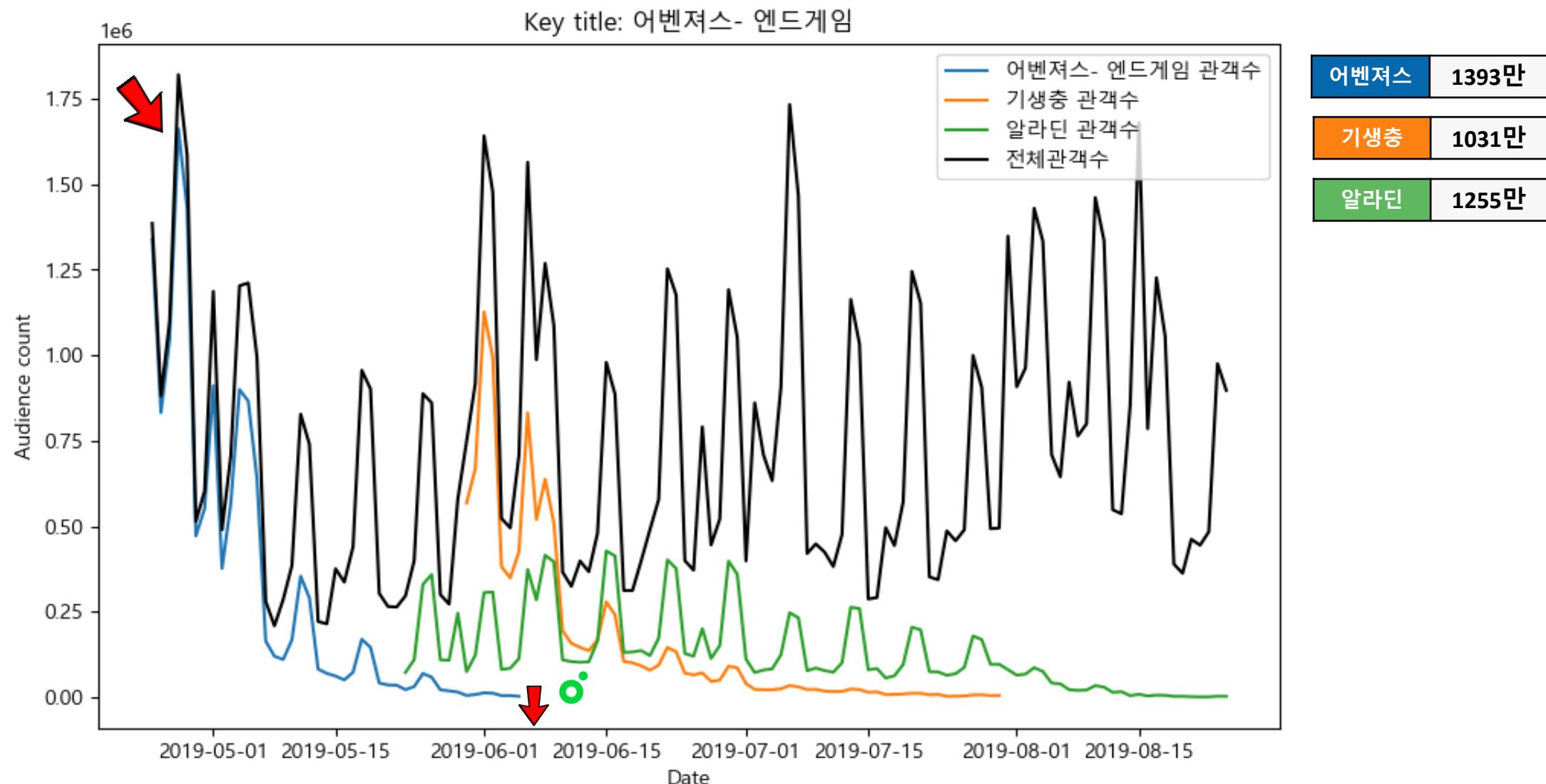


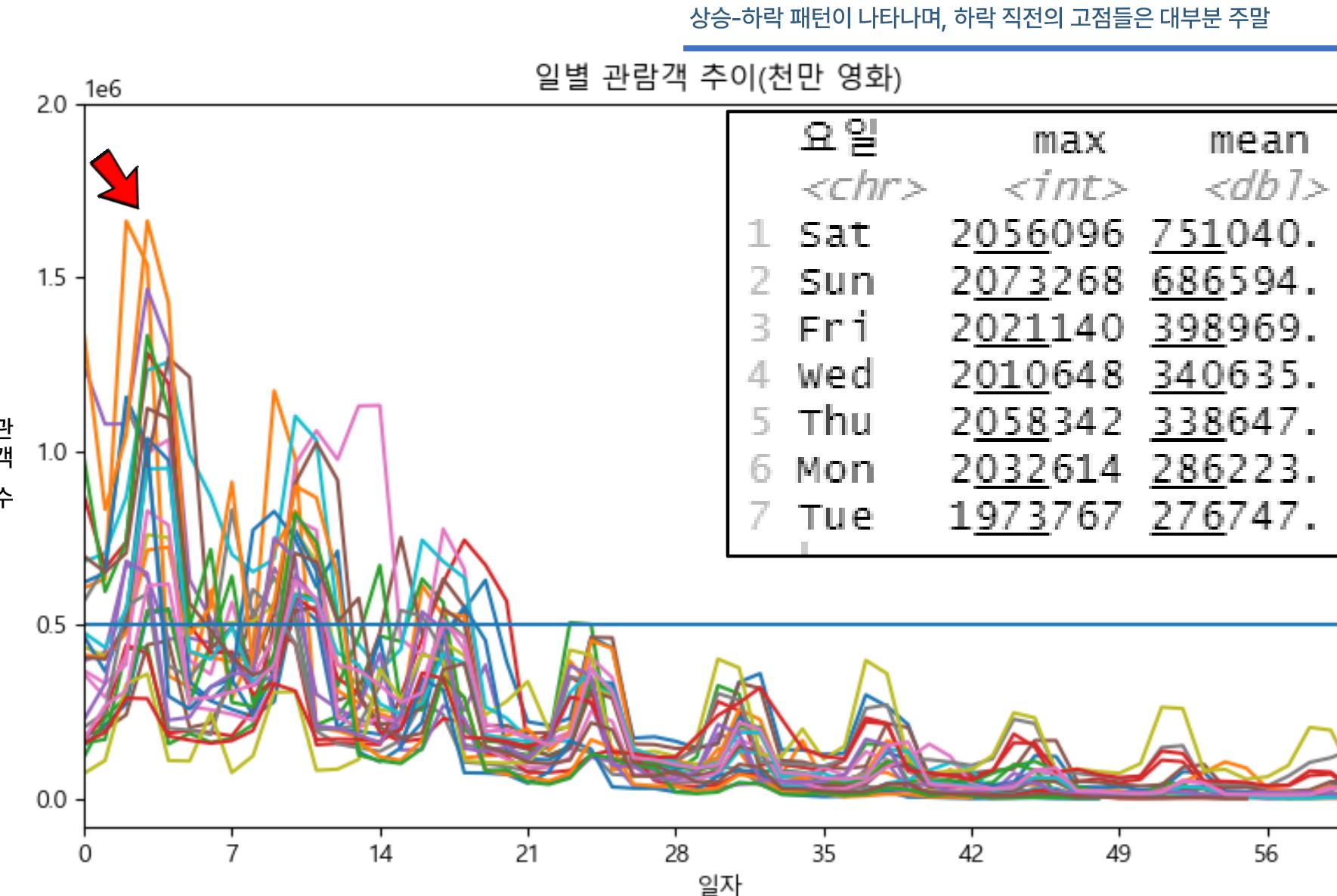
신과함께-죄와 벌, 1987 역시 유사 패턴을 보임

Key title: 신과함께-죄와 벌



신과함께	1441만
1987	723만





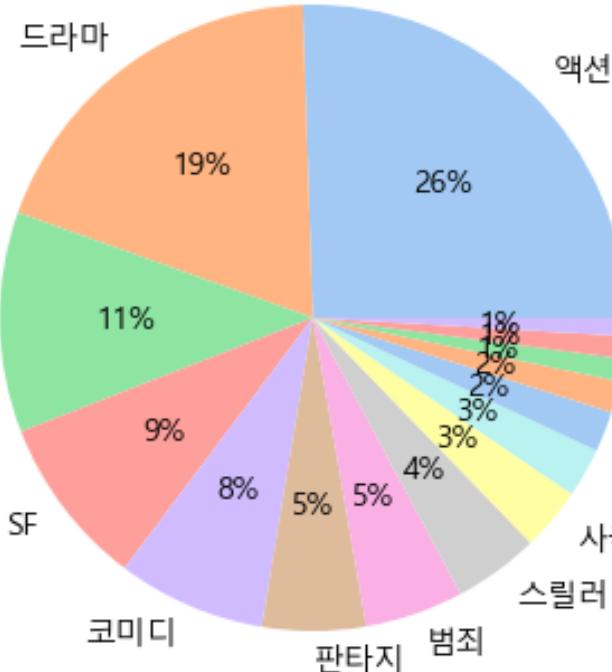




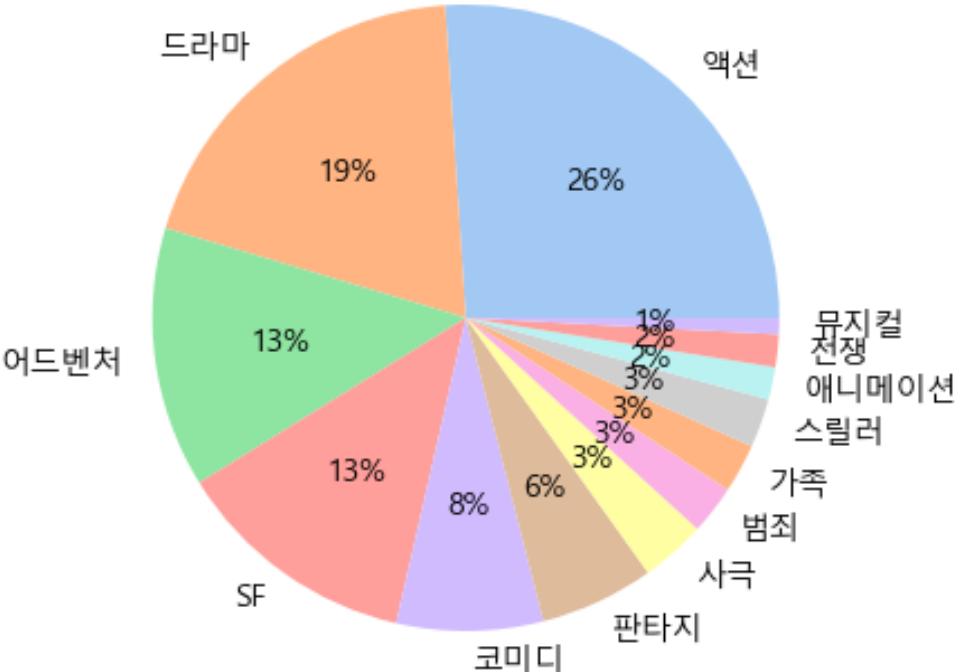
장르

관객 동원력  
대중들의 관심과 참여를  
얼마나 이끌어내는지 나타내는 지표

어드벤처



천만 &amp; 비천만 영화(1~200위)



천만 &amp; 비천만 영화(1~58위)

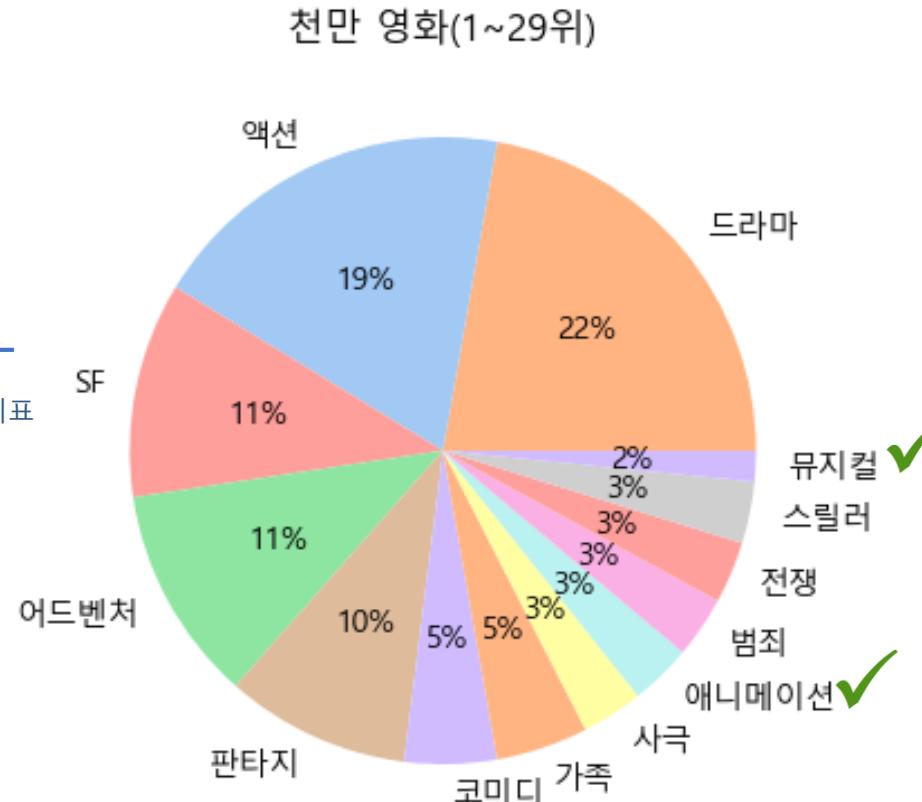
액션, 드라마, 어드벤처 장르 비율이 가장 높으며,  
상위 순위 영화에서도 장르 분포가 유사하게 나타남



장르

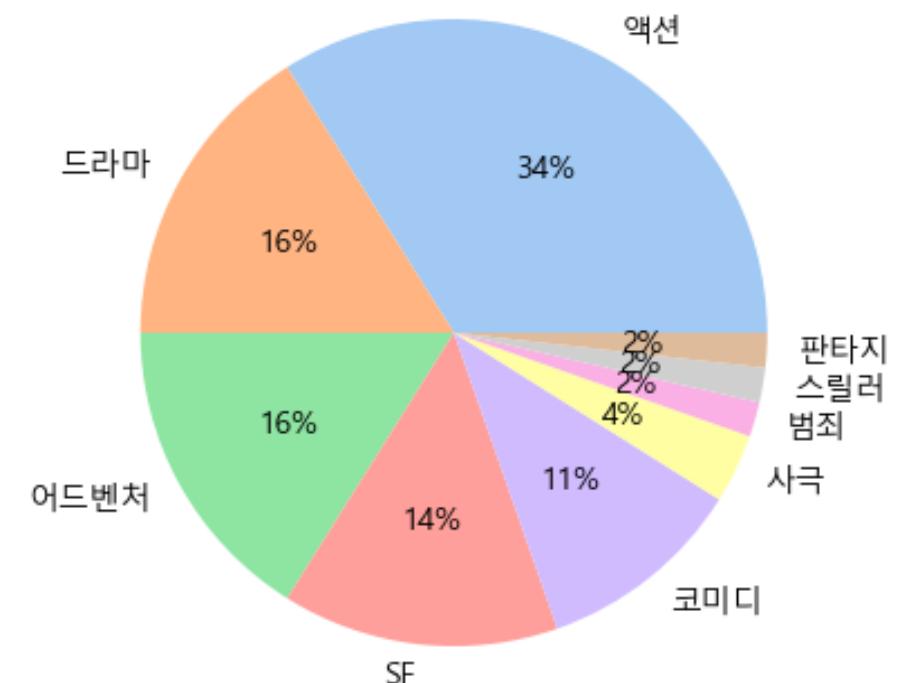
관객 동원력

대중들의 관심과 참여를  
얼마나 이끌어내는지 나타내는 지표



액션, 드라마 장르에서 가장 탄탄한 관객 수요를 보임

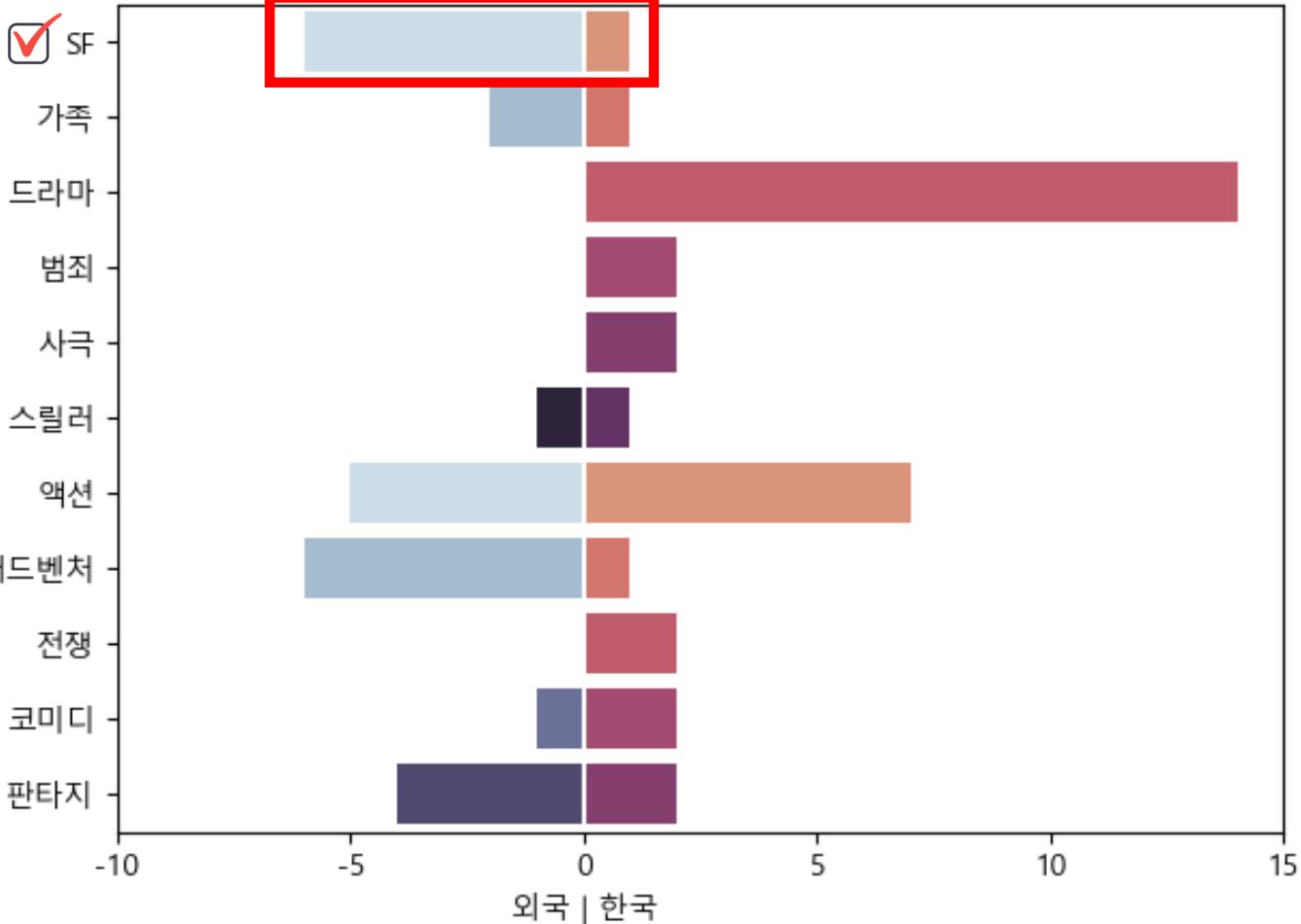
비천만 영화(30~58위)





장르

제작국가별 장르에 따른 영화수(1~58위)



SF장르의 경우 해외 제작 영화가 국내 제작 영화 대비 경쟁력 있음

해외
국내

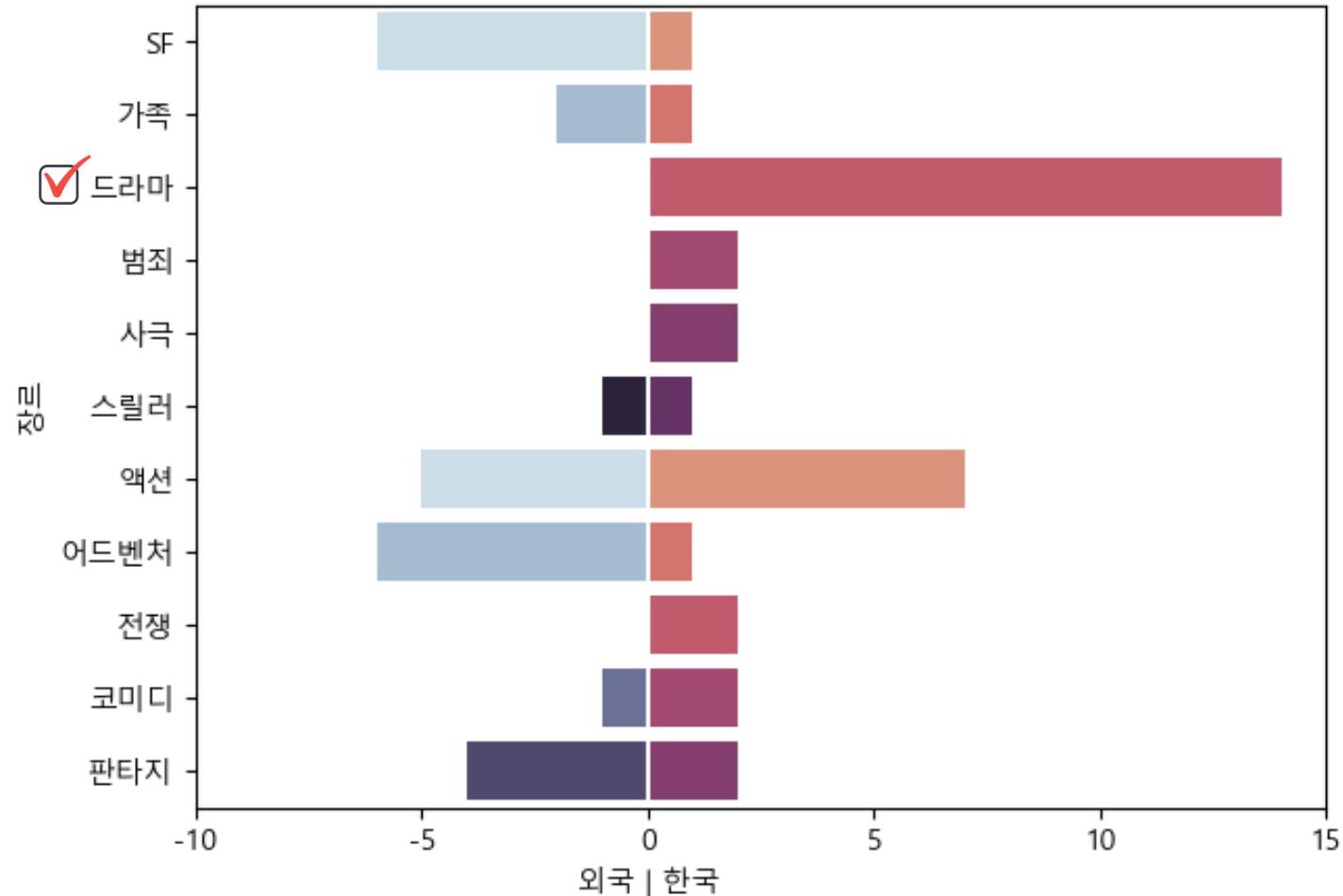
SF장르 작품 수

순위	영화	장르	제작국가	누적관객수
5	아벤져스: 엔드게임	액션,SF	미국	13977602
8	아바타	SF,액션,어드벤처	미국	13338863
21	어벤져스: 인피니티 워	액션,어드벤처,판타지,SF	미국	11233176
23	괴물	SF,가족,드라마	한국	10917400
24	아바타: 물의 길	액션,어드벤처,SF,스릴러	미국	10802350
26	어벤져스: 에이지 오브 울트론	액션,어드벤처,판타지,SF	미국	10504487
27	인터스텔라	SF	미국,영국	10342523
33	설국열차	SF,액션	한국,체코	9353799
35	아이언맨 3	액션,어드벤처,SF	미국	9001679
36	캡틴 아메리카: 시빌 워	액션,SF,스릴러	미국	8678761
43	스파이더맨: 파 프롬 홈	액션,어드벤처,코미디,SF	미국	8023606
48	스파이더맨: 노 웨이 홈	액션,어드벤처,SF	미국	7551990
52	트랜스포머	SF,액션,어드벤처	미국	7402732
53	트랜스포머: 패자의 역습	SF,액션,어드벤처	미국	7393443
56	스파이더맨: 훔 커밍	액션,어드벤처,SF	미국	7258678
61	어벤져스	액션,SF	미국	7087971
86	캡틴 마블	액션,어드벤처,SF	미국	5802810
88	쥬라기 월드: 폴른 킹덤	액션,어드벤처,SF	미국	5661128
93	쥬라기 월드	액션,어드벤처,SF	미국	5547463
95	앤틴맨과 와스프	액션,어드벤처,SF	미국	5448357
100	블랙 팬서	액션,드라마,SF	미국	5400241
101	2012 SF,액션,어드벤처	미국	5397597	
104	트랜스포머: 사라진 시대	액션,SF,어드벤처	미국	5295836
107	월드 워 Z	드라마,스릴러,SF,액션,어.	미국	5244336
120	마션	액션,어드벤처,SF	미국	4887144
121	토르: 라그나로크	액션,어드벤처,판타지,SF	미국	4858572
131	엣지 오브 투모로우	액션,SF	미국	4706213
137	스파이더맨 3	액션,SF	미국	4592309
143	연가시	공포(호러),SF	한국	4516063
	생략			
195	리얼 스틸	액션,드라마,SF	미국	3579666



장르

제작국가별 장르에 따른 영화수(1~58위)



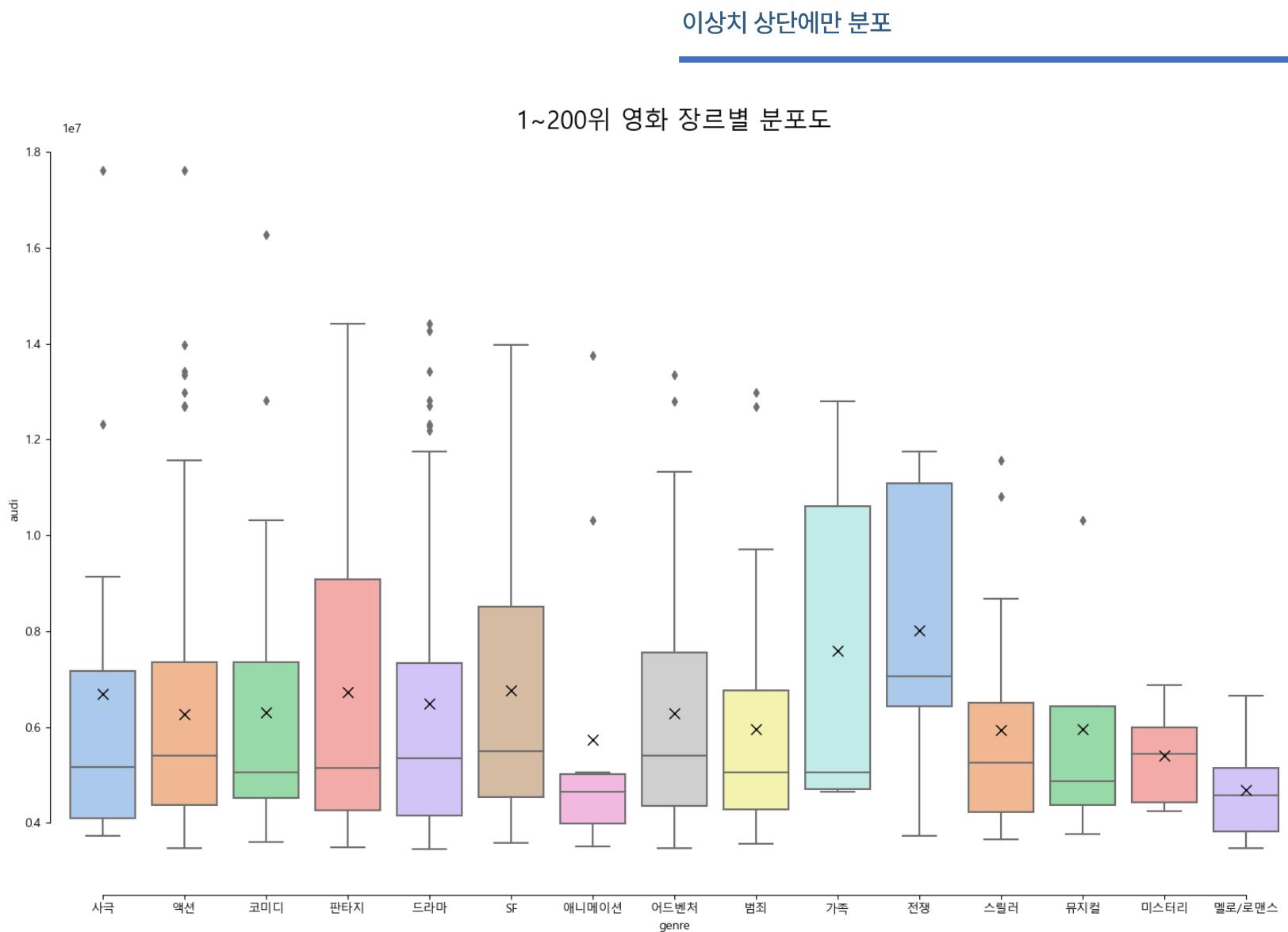
드라마 장르의 경우 반대로 국내 제작 영화가 해외 대비 경쟁력 있음

해외

국내

드라마장르 작품 수

순위	영화	장르	제작국가	누적관객수
3	신과함께-죄와 벌	판타지,드라마	한국	14414658
4	국제시장	드라마	한국	14264478
7	베테랑	액션,드라마	한국	13414484
10	7번방의 선물	코미디,드라마	한국	12812144
12	암살	액션,드라마	한국	12706947
14	광해, 왕이 된 남자	사극,드라마	한국	12324062
15	신과함께-인과 연	판타지,드라마	한국	12278010
16	택시운전사	드라마	한국	12189706
17	태극기 휘날리며	전쟁,드라마	한국	11746135
19	변호인	드라마	한국	11375399
20	해운대	액션,드라마,어드벤처	한국	11325117
23	괴물	SF,가족,드라마	한국	10917400
25	왕의 남자	드라마	한국	10514177
28	기생충	드라마	한국	10313519
30	보헤미안 랩소디	드라마	미국	9948386
생략				
57	1987 드라마		한국	7232452



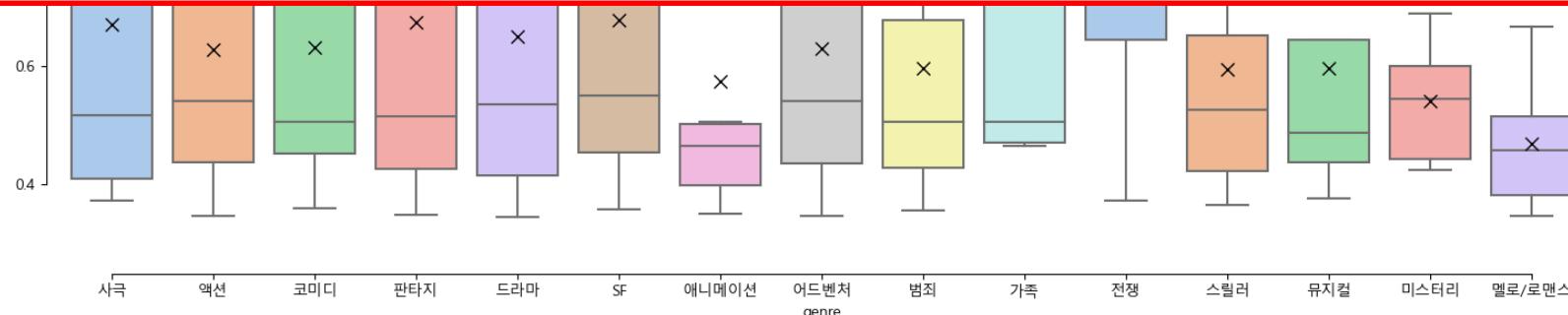


평균, 중앙값의 차이로 인하여 Shapiro test 결과 정규분포가 아님

1~200위 영화 장르별 분포도

정규분포 X

```
## 1~200위 장르별 조건 확인 ##
shapiro.test(genre_audi$audi[genre_audi$genre=="사극"])
# p-value = 0.001059
shapiro.test(genre_audi$audi[genre_audi$genre=="액션"])
# p-value = 3.658e-10
shapiro.test(genre_audi$audi[genre_audi$genre=="코미디"])
# p-value = 4.107e-05
shapiro.test(genre_audi$audi[genre_audi$genre=="판타지"])
# p-value = 0.0007708
shapiro.test(genre_audi$audi[genre_audi$genre=="드라마"])
# p-value = 6.137e-08
shapiro.test(genre_audi$audi[genre_audi$genre=="SF"])
# p-value = 0.000572
shapiro.test(genre_audi$audi[genre_audi$genre=="애니메이션"])
# p-value = 2.73e-05
shapiro.test(genre_audi$audi[genre_audi$genre=="어드벤처"])
# p-value = 0.07139
shapiro.test(genre_audi$audi[genre_audi$genre=="범죄"])
# p-value = 0.0001646
shapiro.test(genre_audi$audi[genre_audi$genre=="가족"])
# p-value = 0.02499
shapiro.test(genre_audi$audi[genre_audi$genre=="전쟁"])
# p-value = 0.5434
shapiro.test(genre_audi$audi[genre_audi$genre=="스릴러"])
# p-value = 0.004131
shapiro.test(genre_audi$audi[genre_audi$genre=="뮤지컬"])
# p-value = 0.107
```

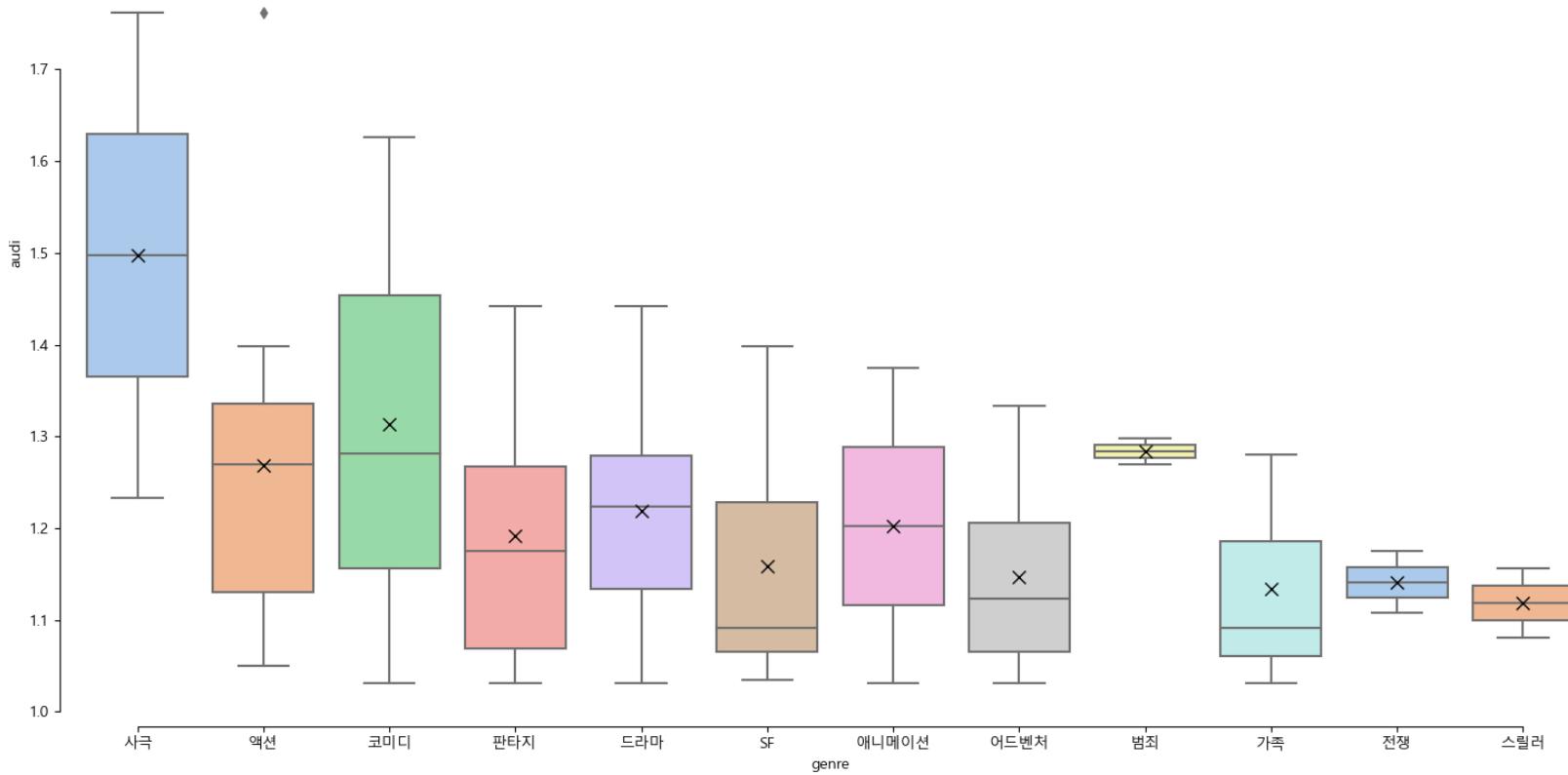




1e7

1위부터 29위까지 범위 수정 결과 이상치 하락, 정규분포 형태를 보이나

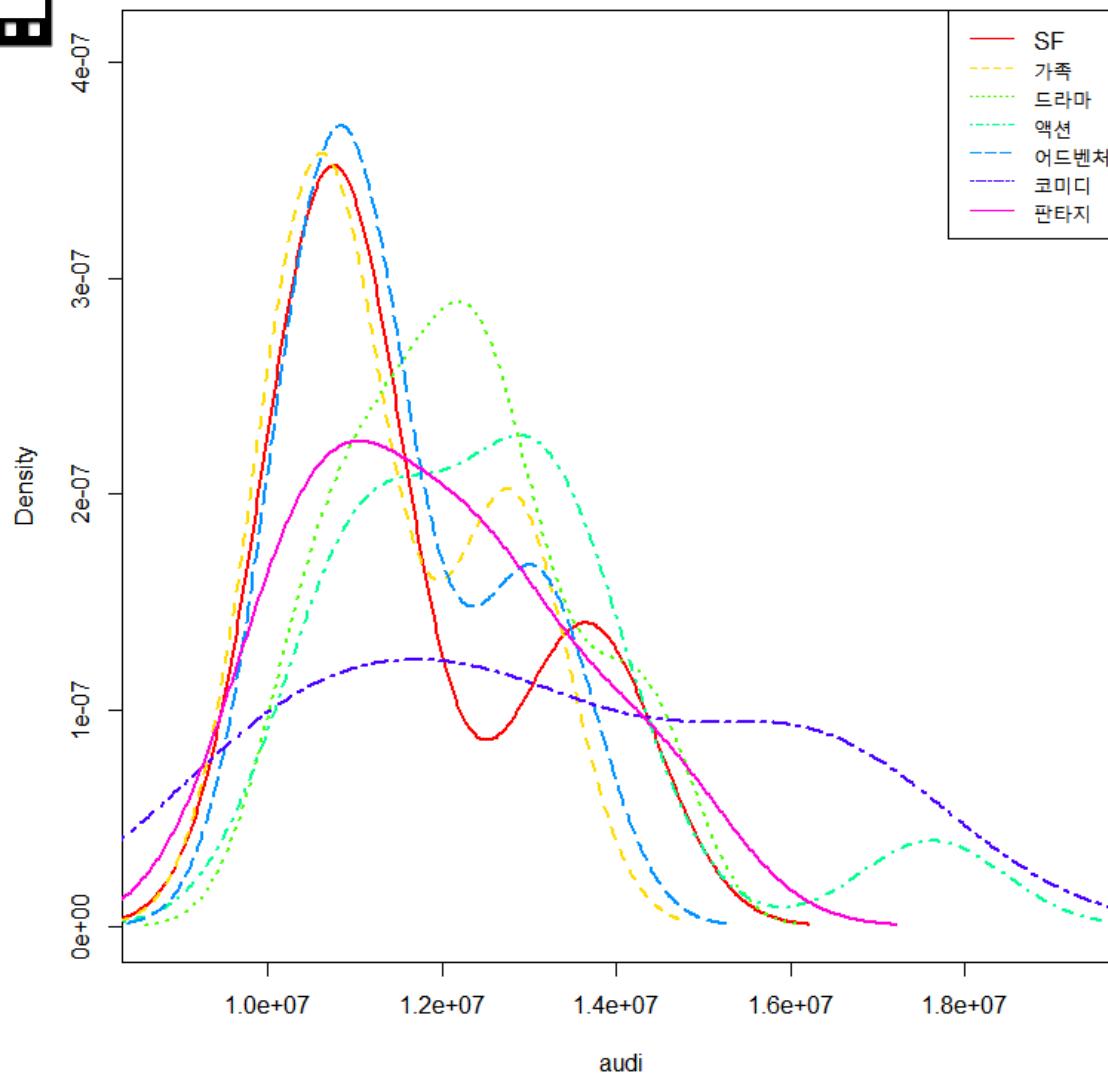
1~29위 영화 장르별 분포도





장르

1~29위 장르별 관객수분포도



천만영화의 동원 관객으로 인해 장르별 이상치를 견인하는 현상 발생  
Shapiro test, Kruskal 검증을 통해 장르 간 관객 수 차이 없음을 확인함

```
> kruskalmc(genre_audi$audi, genre_audi$genre)
Multiple comparison test after Kruskal-Wallis
alpha: 0.05
Comparisons
```

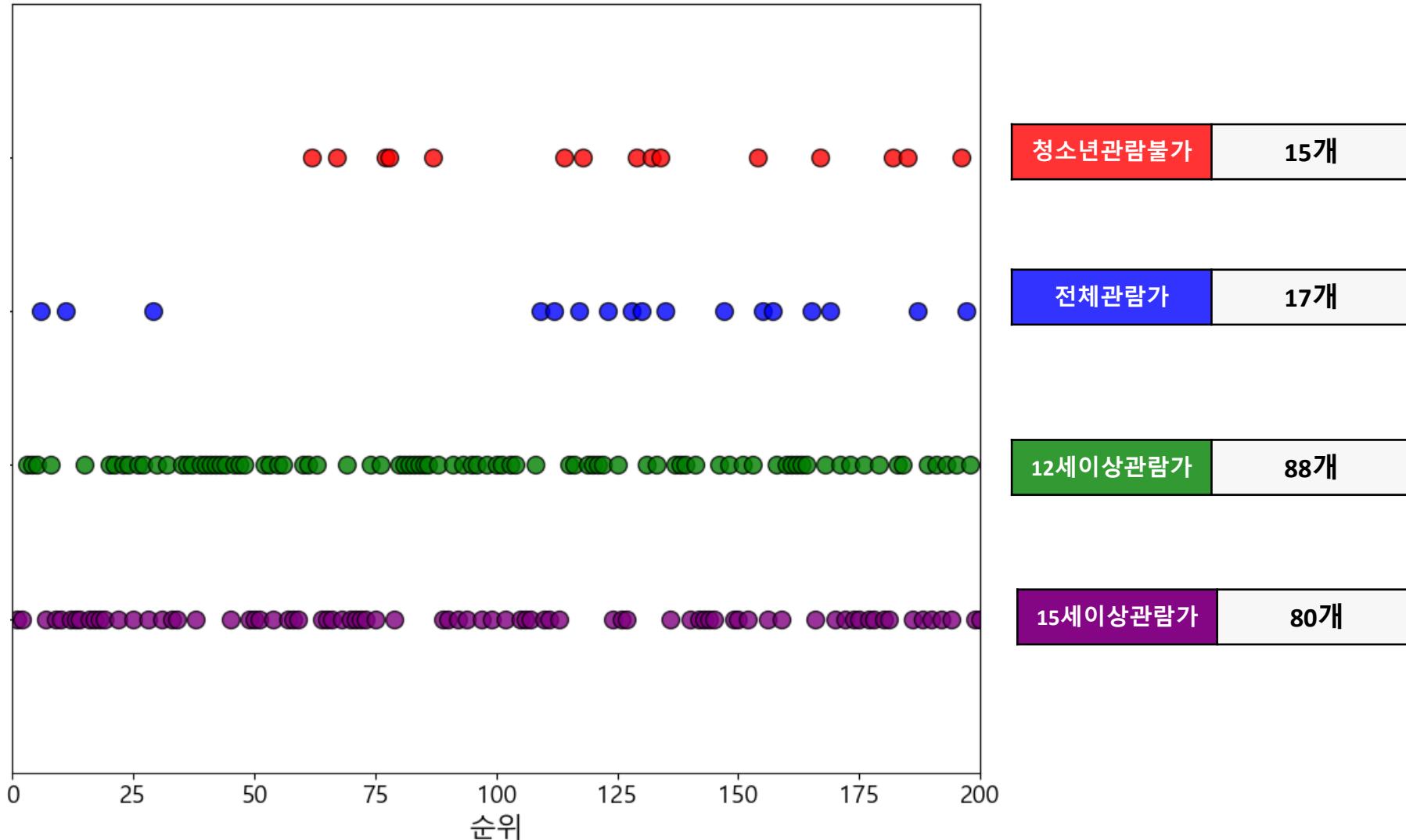
	obs.dif	critical.dif	stat.signif
SF-가족	3.476190	31.77150	FALSE
SF-드라마	7.607143	21.31297	FALSE
SF-액션	10.065476	21.89700	FALSE
SF-어드벤처	1.000000	24.61010	FALSE
SF-코미디	9.023810	31.77150	FALSE
SF-판타지	2.607143	25.61500	FALSE
가족-드라마	11.083333	29.29188	FALSE
가족-액션	13.541667	29.71952	FALSE
가족-어드벤처	2.476190	31.77150	FALSE
가족-코미디	12.500000	37.59255	FALSE
가족-판타지	6.083333	32.55610	FALSE
드라마-액션	2.458333	18.11254	FALSE
드라마-어드벤처	8.607143	21.31297	FALSE
드라마-코미디	1.416667	29.29188	FALSE
드라마-판타지	5.000000	22.46584	FALSE
액션-어드벤처	11.065476	21.89700	FALSE
액션-코미디	1.041667	29.71952	FALSE
액션-판타지	7.458333	23.02064	FALSE
어드벤처-코미디	10.023810	31.77150	FALSE
어드벤처-판타지	3.607143	25.61500	FALSE
코미디-판타지	6.416667	32.55610	FALSE





200위 내 데이터 중 12, 15세관람가 전체의 84% 차지  
전 구간에 걸쳐 오직 청소년관람불가만 정규분포를 이룸

### 순위와 영상등급의 분포도





```

Call:
lm(formula = Audi ~ dummy, data = sample1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3104012 -2117027 -1046921  1112807 11040757 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6575162   213385  30.814 <2e-16 ***
dummy       -1545826   779172  -1.984   0.0486 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2902000 on 198 degrees of freedom
Multiple R-squared:  0.01949, Adjusted R-squared:  0.01454 
F-statistic: 3.936 on 1 and 198 DF,  p-value: 0.04864

```

P-value : 0.07  
 설명력 : 0.011

청소년 관람불가 영화의 관람객수와의 선형 회귀분석 시행 허나 유의미한 데이터론 활용X  
R이 아닌 Python을 이용한 분석의 경우 표본이 20미만인 관계로 예측실패

P-value : 0.048   
설명력 : 0.0145

```

Call:
lm(formula = Audi^-1 ~ YouthNotAllowed, data = sample1)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.215e-07 -4.597e-08  5.335e-09  4.772e-08 1.098e-07 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.783e-07  4.521e-09 39.441 <2e-16 ***
YouthNotAllowed 2.969e-08  1.651e-08  1.798   0.0736 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.149e-08 on 198 degrees of freedom
Multiple R-squared:  0.01607, Adjusted R-squared:  0.0111 
F-statistic: 3.234 on 1 and 198 DF,  p-value: 0.07365

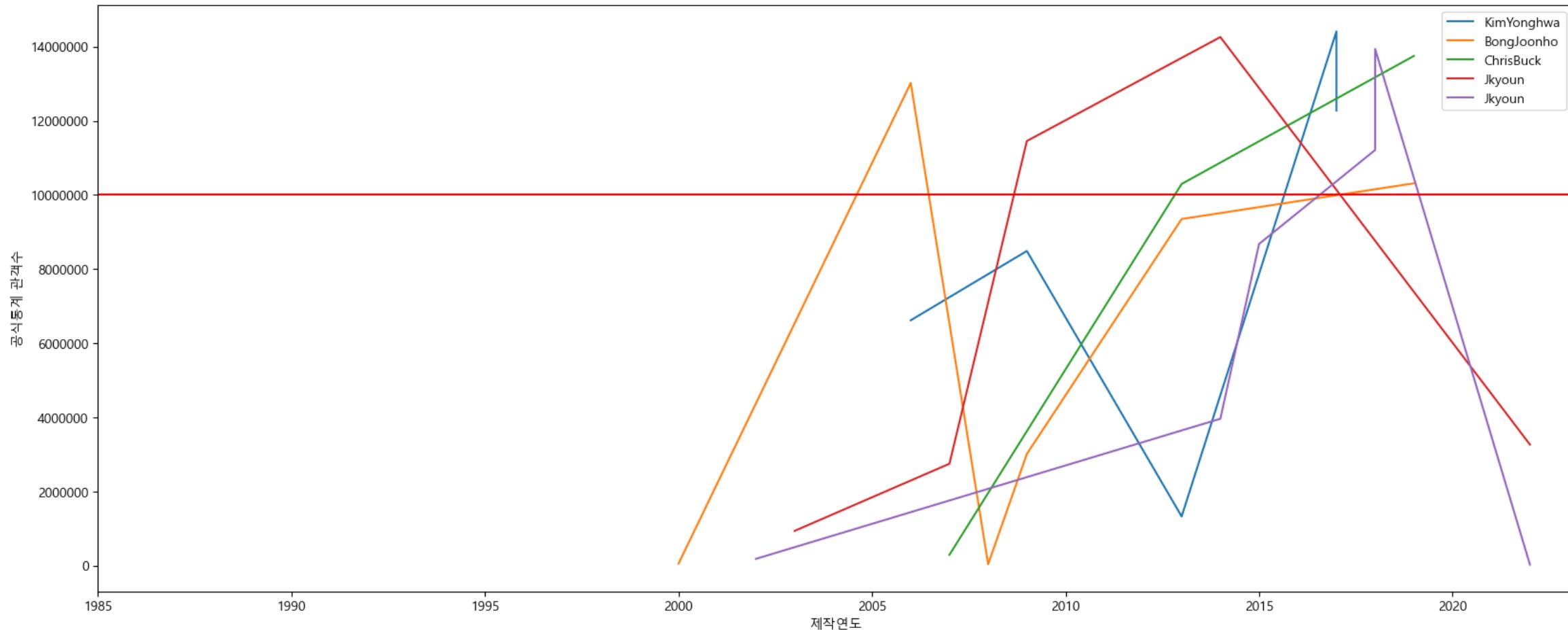
```





천만을 달성해본 감독의 차기작들은 관객 수 어드밴티지가 존재한다는 가설  
즉 후광효과 기대

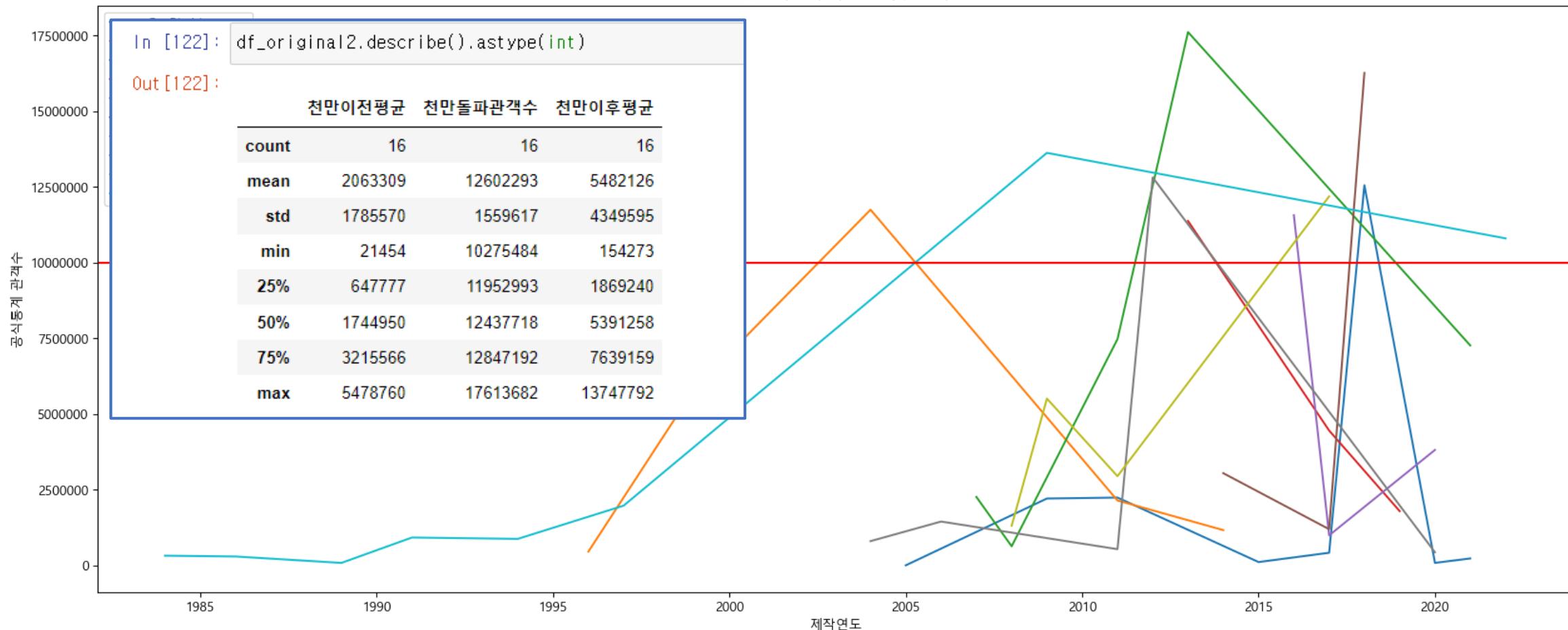
김용화, 봉준호, 크리스벅, 윤제균, 앤소니 루소 작품에 따른 관객수 추이





후광효과로 인한 많은 관객수 확보 기대에 비하여 저조하나 평균 관객 수 상승  
천만달성이라는 이벤트가 차기작에 영향을 미침

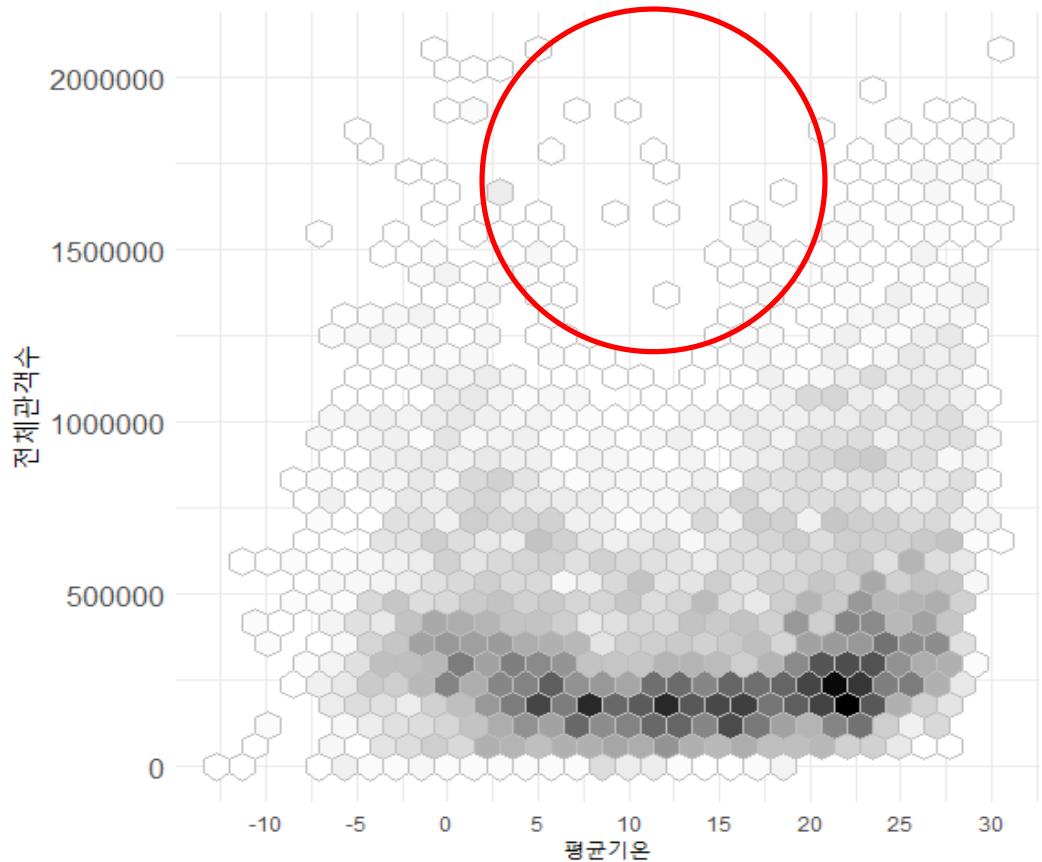
감독별 작품에 따른 관객수 추이



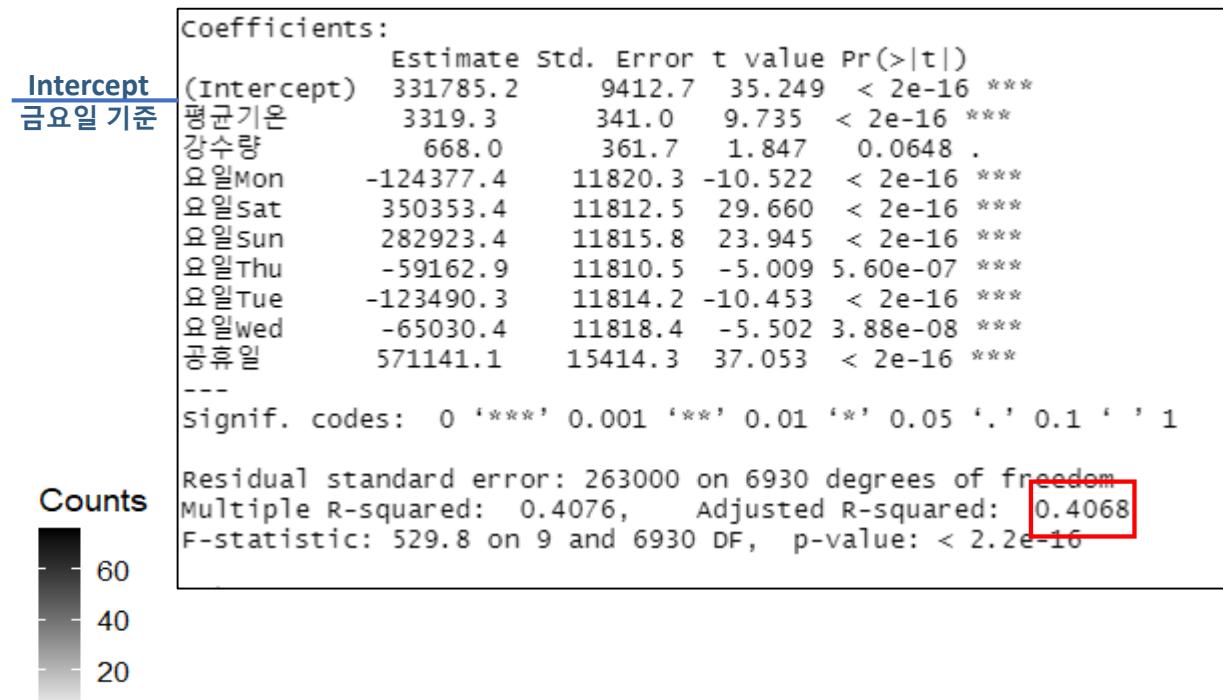




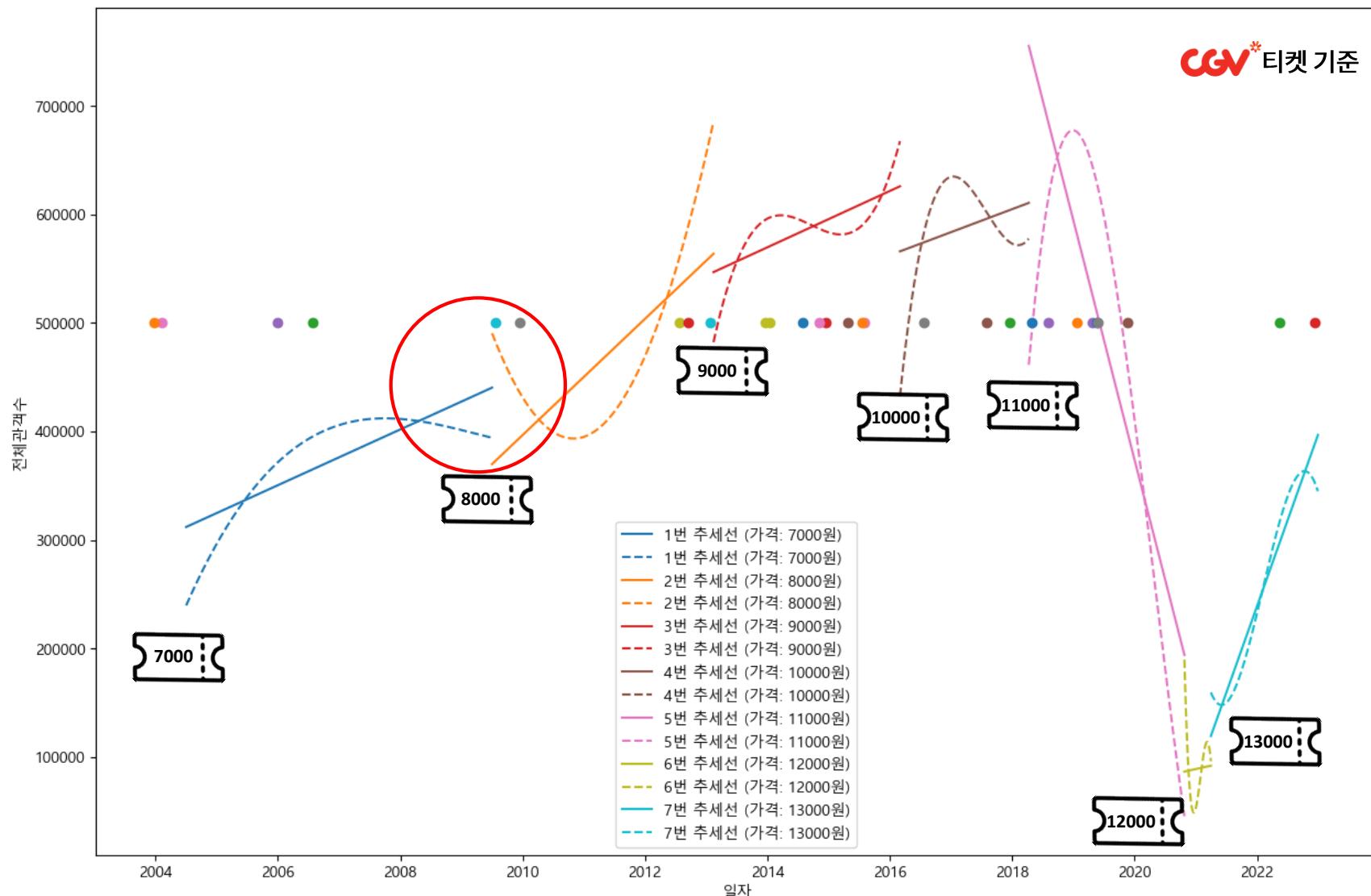
평균 기온과 전체 관객 수의 관계

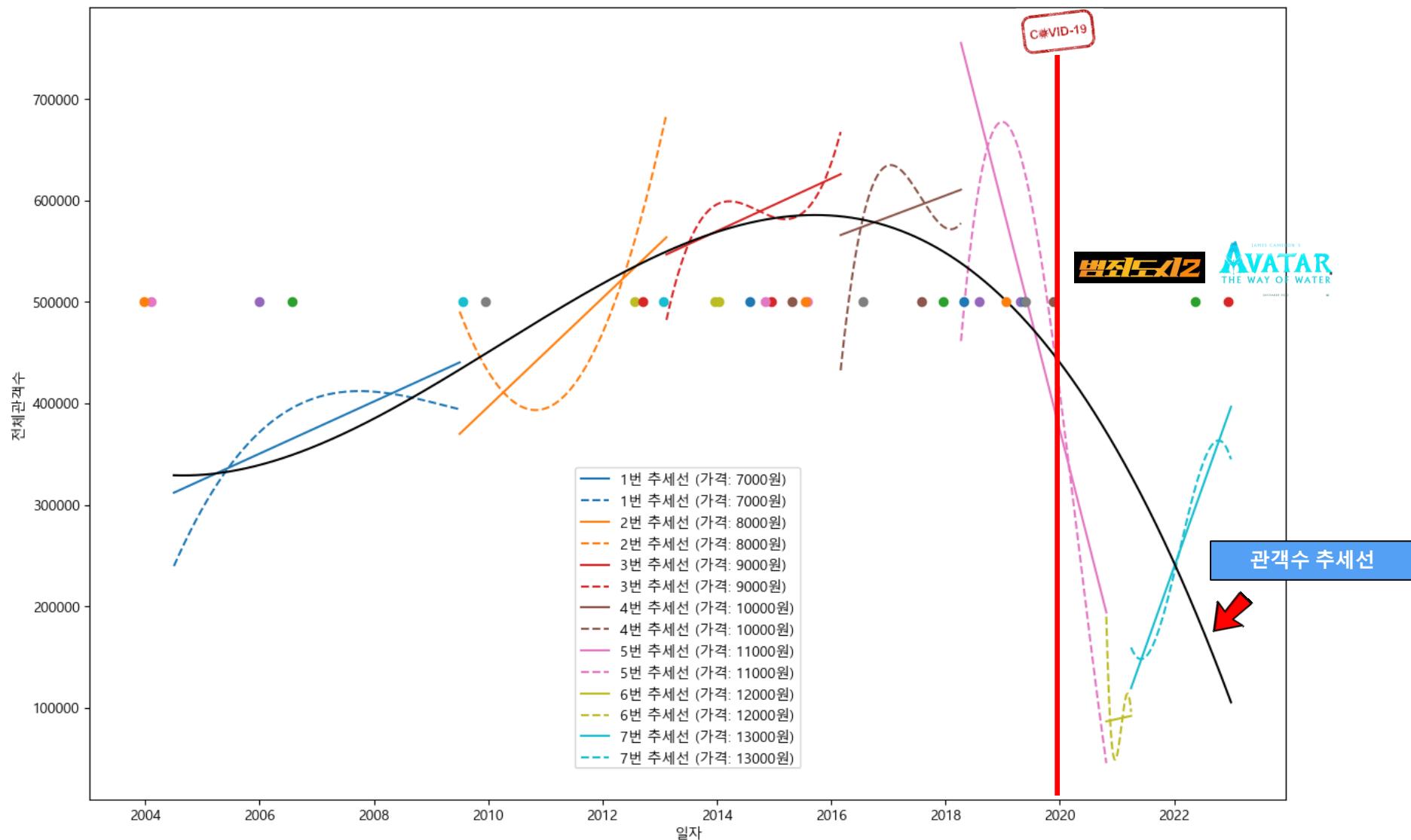


5도 이하, 10~20도 내 데이터 분포 저조 이는 영화관에 방문하지 않음을 나타냄  
이에 기온 관광객 수 연관성 UP 천만을 달성하고자 하면 영화관을 많이 찾는 5~8월 유리

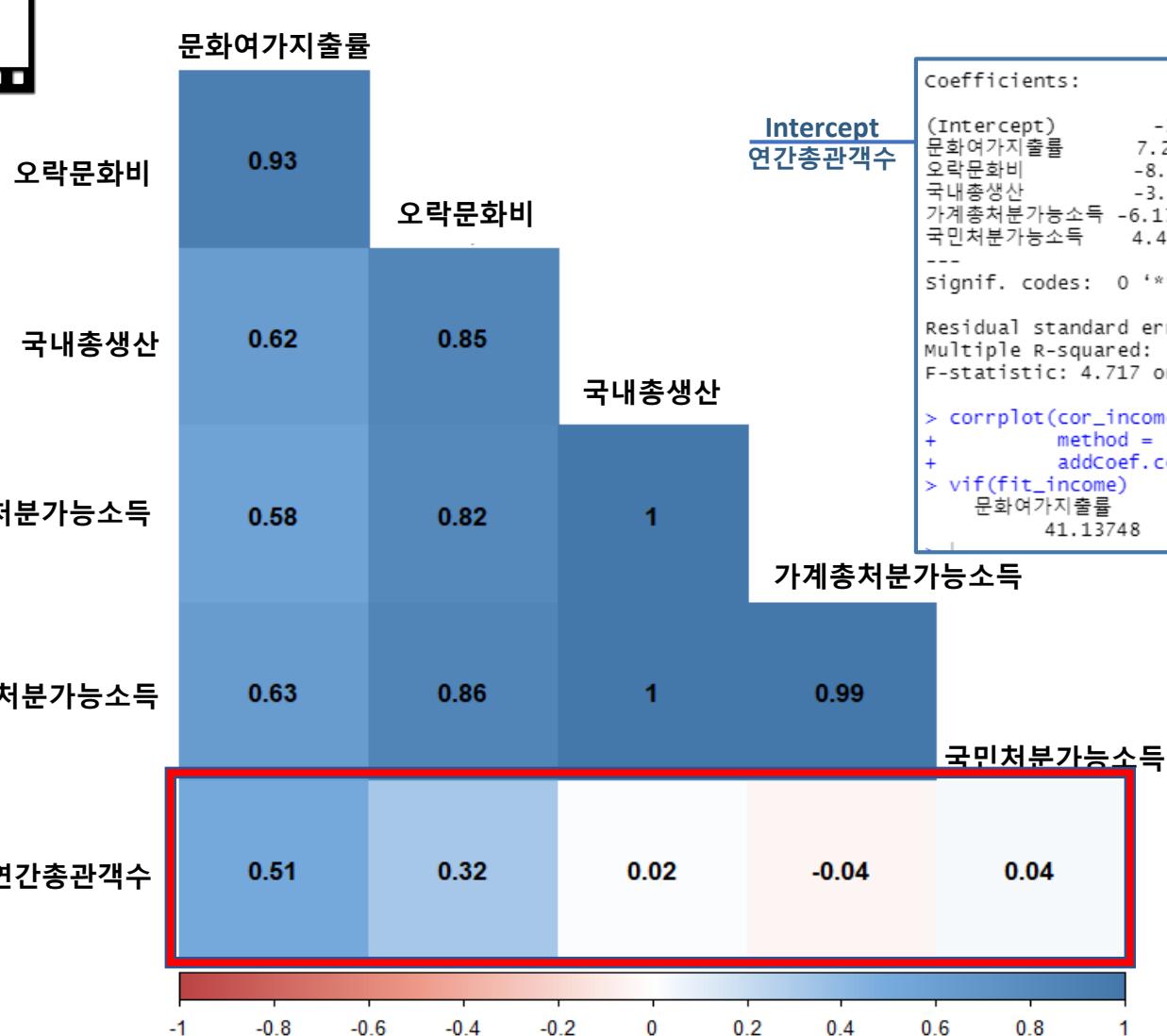












다중회귀분석 진행 결과 변수간 연관 존재, 다중 공산성UP, 설명력 55%  
문화여가 활동 내 다양한 항목이 있음을 고려, 지출률과 영화관람객의 상관관계 높음

## Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.114e+08	3.019e+08	-0.700	0.4997
문화여가지출률	7.227e+07	1.126e+08	0.642	0.5355
오락문화비	-8.590e+02	2.919e+03	-0.294	0.7746
국내총생산	-3.464e+03	2.220e+03	-1.561	0.1496
가계총처분가능소득	-6.177e+01	1.146e+03	-0.054	0.9581
국민처분가능소득	4.440e+03	2.152e+03	2.063	0.0661 .

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35900000 on 10 degrees of freedom  
Multiple R-squared: 0.7022, Adjusted R-squared: 0.5533  
F-statistic: 4.717 on 5 and 10 DF, p-value: 0.01792

```
> corplot(cor_income_notper, diag = FALSE, col = col(200), type = "lower",
+   method = "color", tl.col = "black",
+   addCoef.col = "black")
> vif(fit_income)
문화여가지출률          오락문화비          국내총생산          가계총처분가능소득          국민처분가능소득          41.13748
93.66718                 6467.18461                488.66784                3773.14895
```

문화 여가지출률 : (가계지출총액/오락문화비)

## [내적요인]

### [1] 패턴

- ▶ 개봉일 기준 2주 이내, 주간 관객 수 추이에 따른 수요일 개봉 공식을 따르는 것이 유리

### [2] 장르

- ▶ 액션, 드라마 장르에 대한 관객의 기본 관심도 UP, 해외 영화는 SF 국내 영화는 드라마 장르 유리

### [3] 상영등급

- ▶ 상영등급, 영화의 감독과 천만관객 달성 요인 상관관계 X

## [외적요인]

### [4] 기온

- ▶ 평균 기온 기준 20~25도 사이인 5 ~ 8월 개봉 유리

### [5] 가격

- ▶ 비용 대비 효용 즉 퀄리티 좋은 영화여야 함, 관객추세선 감소한 만큼 티켓 가격 조정 요망

### [6] 거시경제환경

- ▶ 경기 침체에 따른 문화활동 지양하기에 영화산업에 큰 타격은 당연지사

### [아쉬운 점]

#### [1] 선형예측 모델

- ▶ 분석하기엔 표본도 적고 정규분포도 아닌 것들 다수인 것 등 애로사항 다수

#### [2] 배포된 로우 데이터 부족

- ▶ 제작비나, OTT, 멀티플렉스관련 데이터에 접근이 제한
- ▶ 주제 선정 익일 부로 네이버의 리뷰 및 평점 데이터 서비스 종료로 인한 접근 불가

#### [3] 깃허브 미숙

- ▶ 데이터 병합과정에서 데이터 손실, 브랜치 관리 등 깃허브 숙련도 미숙

#### [4] 유의미 결과 도출 실패

- ▶ 예상했던 변수로 사용하려던 데이터를 충분히 활용치 못한 부분이 아쉬움  
ex) 일반관vs특수관, 시리즈의 영향



무엇이든 물어보세요!

내용을 입력하세요...

