



코로나 이전 이후 국내영화 시장 분석

1조

박한빈, 박재현, 손보영, 전영욱, 이봉학

Table of Contents.



주제

주제 및 주제 선정 이유



데이터 분석

- 데이터 수집(크롤링)
- 데이터 분석
- 데이터 시각화



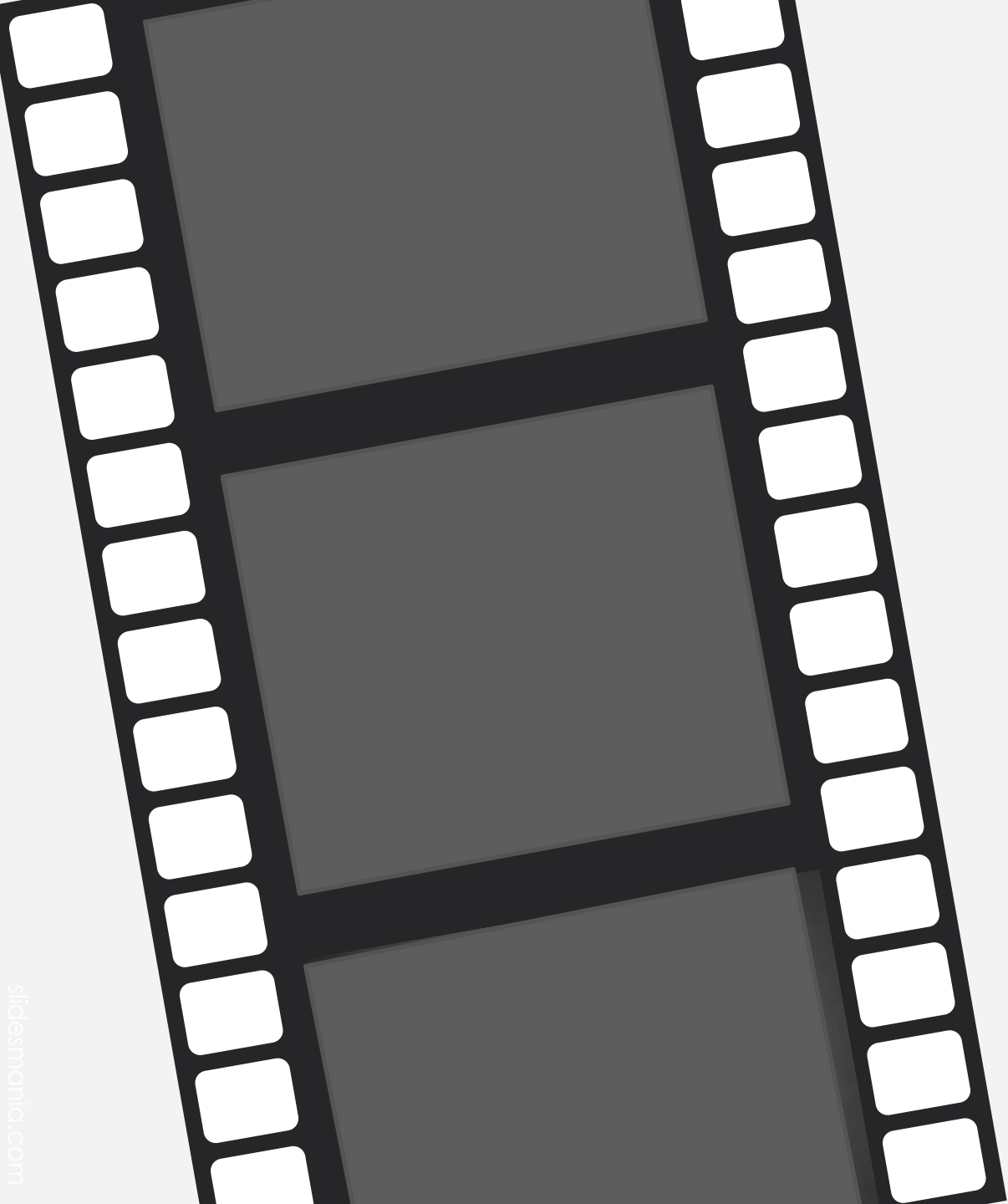
과정 공유

- 코드 공유
- 애로사항 공유



분석 결과

- 영화시장 연도별 총 매출액 및 관람객 변화
- 연도별 주요 선호 장르 변화/관람객(장르별 변화)
- 월별 관객 분포도 (최근 5개년)
- 국가별 매출액 분포 비율
- 국내외 영화 점유율
- 코로나의 영화 산업에 영향력 비교
- '18-'20 영화 개봉편수 비교
- 영화 티켓 가격 변동
- OTT 시장과의 연관성 비교



1. 주제

2. 데이터 분석

3. 데이터 분석 결과

4. 과정 공유

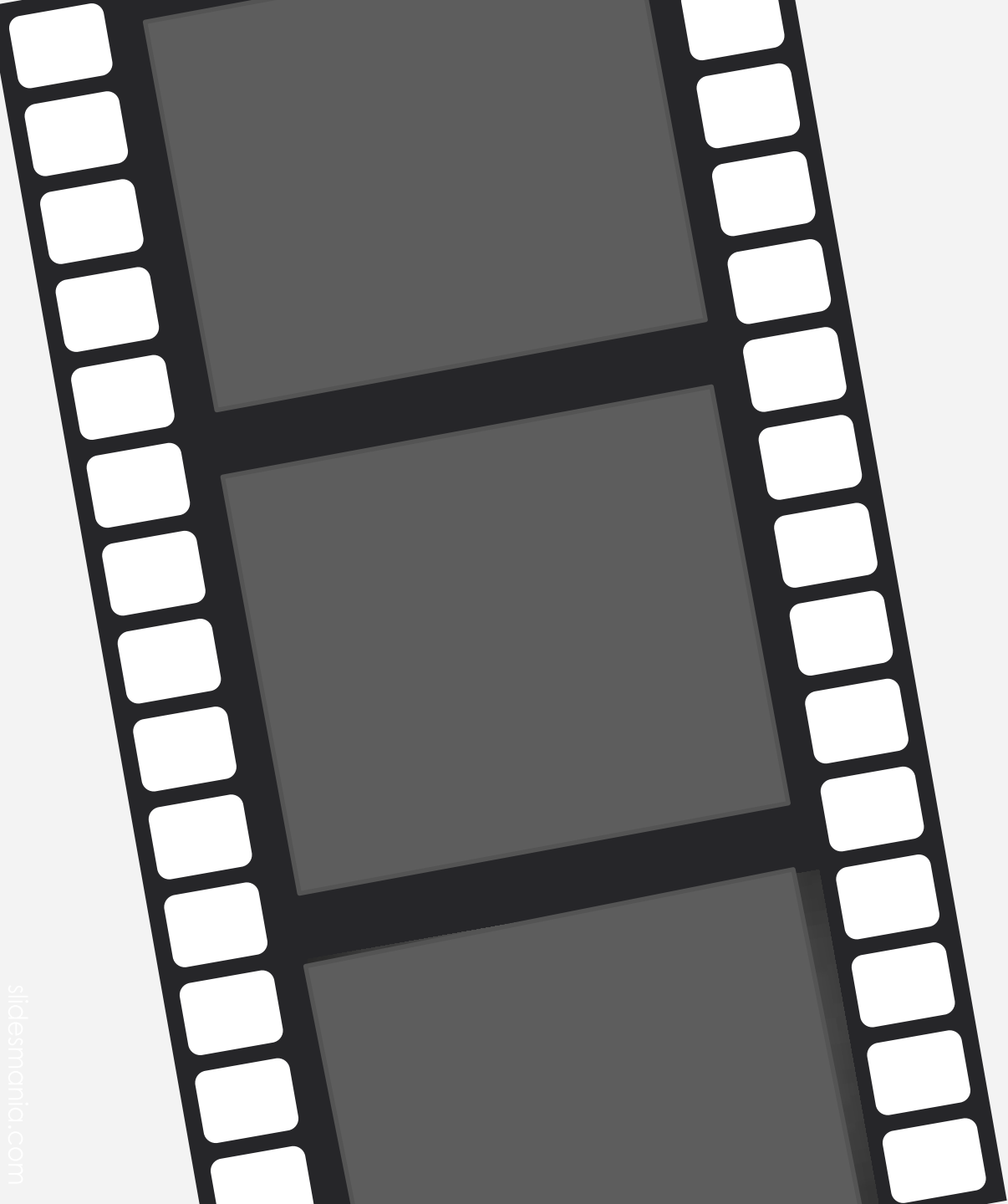


국내 영화 시장, 그 경쟁력

국내 영화시장의 현황은 어떠한가?

국내 영화시장에서의 출구전략은 무엇인가?

코로나 시기를 겪은 영화 산업, 어떤 방향성을 갖는 것이 좋을까?



1. 주제

2. 데이터 분석

3. 데이터 분석 결과

4. 과정 공유

1. 수집대상 (1)

- ❖ 영화관입장권통합 전산망 정보를 통해 2004-2022년 간 매출액 기준 상위 50개 영화 정보 추출
- ❖ 연도별 상위 50개 영화에 대한 매출 관련 정보 / 연도별 상위 50개 영화에 대한 기본 정보 등

· 조회기간	2022	· 영화구분	--전체--
· 국적	--전체--	· 지역	--전체--

총 50건

순위	영화명	개봉일	매출액 ▲ ▼	매출액 점유율 ▲ ▼	관객수 ▲ ▼
1	스파이더맨: 노 웨이 홈	2021-12-15	19,489,875,950	16.3%	1,990,269
2	해적: 도깨비 깃발	2022-01-26	12,500,410,990	10.5%	1,334,995
3	더 배트맨	2022-03-01	9,287,122,000	7.8%	899,040
4	썬2게터	2022-01-05	8,181,778,050	6.8%	883,236
5	퀸메이커	2022-01-26	7,308,159,110	6.1%	774,538

kofic

KOBIS

영화관입장권통합전산망

스파이더맨: 노 웨이 홈 (Spider-Man: No Way Home)

영화상영권상영종

기본정보

통계정보

상영현황정보

* 최종수정: 2021-12-15 16:14:00

수정요청

공

트

코드

20210028

해당정보 없음

ISAN

스파이더맨: 노 웨이 홈 (SpiderMan No Way Home)

A.K.A.

장편 | 일반영화 | 액션, 어드벤처, SF | 148분 1주 | 12세이상관람가 | 미국

요약정보

개봉일

2021-12-15

제작연도

2021년

제작상태

개봉

크랭크인/업

해당정보없음

촬영회차

해당정보없음

상영타입

2D : 다지컬(20210028D)

4D : 4D(20210028F)

IMAX : IMAX(20210028I)

ScreenX : ScreenX(20210028J)

DOLBYCINEMA : DOLBYCINEMA(20210028G)

포스터

(Total : 8건)

스크린수

2,162

1,701

2,370

1,251

1,103

1. 수집대상 (1)

❖ 20*11 행렬의 데이터

```
1 # concat 활용 - 데이터 합치기
2
3 cols=['연도', '한국개봉편수', '한국상영편수', '한국매출액', '한국관객수', '한국점유율', '외국개봉편수', '외국상영편수', '외국매출액', '외국관객수', '외국점유율']
4 df_movies_all = pd.concat([korea_movies, overseas_movies], axis=1, ignore_index=True)
5 df_movies_all.columns=cols
6 df_movies_all
```

	연도	한국개봉편수	한국상영편수	한국매출액	한국관객수	한국점유율	외국개봉편수	외국상영편수	외국매출액	외국관객수	외국점유율
0	2004	74	95	239,143,250,406	37,741,433	54.5%	206	205	201,585,655,800	31,513,193	45.5%
1	2005	82	116	451,707,494,794	71,346,379	57.8%	224	298	328,657,484,673	52,005,680	42.2%
2	2006	110	159	568,090,373,200	91,745,620	63.6%	241	322	324,351,764,800	52,510,415	36.4%
3	2007	111	183	479,858,559,900	75,791,003	49.9%	282	428	485,628,064,638	76,231,989	50.1%
4	2008	110	252	407,327,224,301	62,047,324	42.1%	270	552	554,126,072,950	85,381,315	57.9%
5	2009	119	211	526,482,459,500	75,644,847	48.7%	243	503	556,713,545,450	79,753,807	51.3%
6	2010	142	275	508,426,689,450	68,843,175	46.6%	288	521	648,828,048,800	78,916,039	53.4%
7	2011	152	334	613,722,968,100	82,868,189	51.9%	290	609	622,076,496,100	76,856,276	48.1%
8	2012	176	389	838,238,171,234	114,911,591	59.0%	465	842	616,902,183,201	79,978,996	41.0%
9	2013	183	491	909,932,880,105	127,291,781	59.7%	724	1184	641,419,104,616	86,059,249	40.3%
10	2014	217	515	820,639,569,302	107,703,922	50.1%	878	1437	843,592,979,858	107,365,036	49.9%
11	2015	257	584	879,658,901,660	112,941,882	52.0%	946	1556	835,842,670,018	104,357,641	48.0%
12	2016	337	702	927,865,027,723	116,555,889	53.7%	1237	1913	815,329,396,647	100,470,293	46.3%
13	2017	495	887	902,667,356,338	113,907,067	51.8%	1272	2059	853,910,494,775	105,969,160	48.2%
14	2018	661	1096	912,746,432,308	110,149,287	50.9%	1210	2088	901,217,376,930	106,235,982	49.1%
15	2019	697	1126	970,793,408,451	115,621,862	51.0%	1246	1963	943,195,671,617	111,056,915	49.0%
16	2020	782	1133	350,411,877,790	40,462,297	68.0%	1115	1865	159,962,840,401	19,061,670	32.0%
17	2021	817	1178	173,420,607,692	18,219,745	30.1%	1039	1762	411,116,948,708	42,310,977	69.9%
18	2022	178	312	44,052,815,960	4,649,125	37.5%	277	527	75,458,280,076	7,759,756	62.5%
19	합계	5,700	10,038	11,525,186,068,214	1,548,442,418	52.5%	12,453	20,634	10,819,915,080,058	1,403,794,389	47.6%

1. 수집대상 (1)

- ❖ 연도별 상위 50개 영화에 대한 매출 관련 정보
- ❖ 연도별 상위 50개 영화에 대한 기본 정보

1 df_movies									
	연도	순위	영화명	개봉일	매출액	매출액 점유율	관객수	스크린수	상영횟수
0	2004	1	태극기 휘날리며	2004-02-05	15,687,180,500	3.6%	2,544,911	115	22,375
1	2004	2	트로이	2004-05-21	12,777,346,000	2.9%	2,001,293	177	21,245
2	2004	3	내 머리 속의 지우개	2004-11-05	11,919,939,500	2.7%	1,885,827	206	29,532
3	2004	4	귀신이 산다	2004-09-17	11,898,748,500	2.7%	1,875,936	174	27,936
4	2004	5	투모로우	2004-06-03	11,658,584,500	2.6%	1,830,767	164	22,881
...
945	2022	46	도쿄 리벤저스	2022-01-19	128,160,240	0.1%	13,119	216	2,337
946	2022	47	프리! 더 파이널 스트로크 전편	2022-01-20	127,488,230	0.1%	12,996	77	1,704
947	2022	48	벨파스트	2022-03-23	116,632,770	0.1%	12,848	272	3,270
948	2022	49	장민호 드라마 최종회	2022-01-24	223,662,200	0.2%	11,687	81	839
949	2022	50	픽그	2022-02-23	106,708,970	0.1%	11,615	113	1,600

950 rows × 9 columns

1 movies_info			
	장르	러닝타임	제작국가
0	전쟁,드라마	148분0초	한국
1	사극	136분0초	미국
2	멜로/로맨스	117분0초	한국
3	코미디,공포(호러),판타지	123분0초	한국
4	SF	123분0초	미국
...
945	드라마,액션	120분11초	일본
946	애니메이션	86분29초	일본
947	드라마	98분7초	미국
948	공연	106분10초	한국
949	드라마,스릴러	91분4초	미국

950 rows × 3 columns

1. 수집대상 (1)

❖ 950*14 행렬의 데이터

1 top50

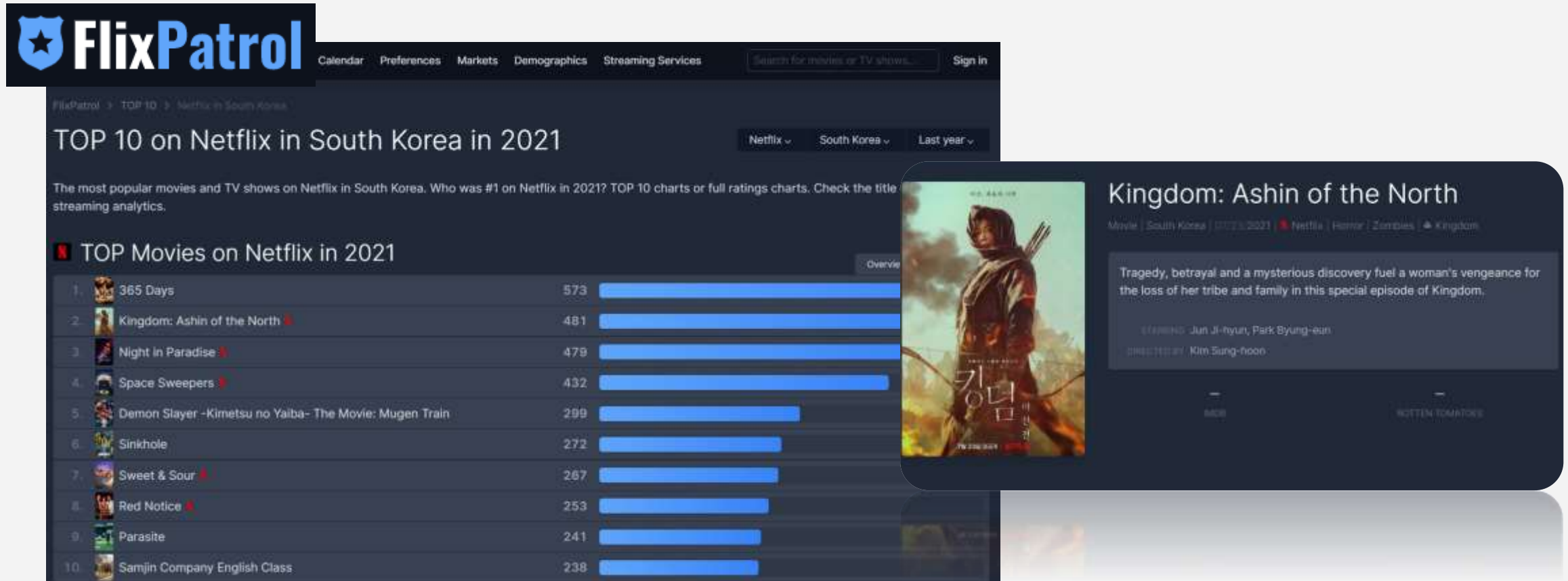
	연도	순위	영화명	개봉일	매출액	매출액 점유율	관객수	스크린수	상영횟수	연도	코드	장르	러닝타임	제작국가
0	2004	1	태극기 휘날리며	2004-02-05	15,687,180,500	3.6%	2,544,911	115	22,375	2004	20030371	전쟁,드라마	148분0초	한국
1	2004	2	트로이	2004-05-21	12,777,346,000	2.9%	2,001,293	177	21,245	2004	20040555	사극	136분0초	미국
2	2004	3	내 머리 속의 지우개	2004-11-05	11,919,939,500	2.7%	1,885,827	206	29,532	2004	20040673	멜로/로맨스	117분0초	한국
3	2004	4	귀신이 산다	2004-09-17	11,898,748,500	2.7%	1,875,936	174	27,936	2004	20040649	코미디,공포(호러),판타지	123분0초	한국
4	2004	5	투모로우	2004-06-03	11,658,584,500	2.6%	1,830,767	164	22,881	2004	20040490	SF	123분0초	미국
...
945	2022	46	도쿄 리벤저스	2022-01-19	128,160,240	0.1%	13,119	216	2,337	2022	20211238	드라마,액션	120분11초	일본
946	2022	47	프리! 더 파이널 스트로크 전편	2022-01-20	127,488,230	0.1%	12,996	77	1,704	2022	20223350	애니메이션	86분29초	일본
947	2022	48	벨파스트	2022-03-23	116,632,770	0.1%	12,848	272	3,270	2022	20223583	드라마	98분7초	미국
948	2022	49	장민호 드라마 최종회	2022-01-24	223,662,200	0.2%	11,687	81	839	2022	20223398	공연	106분10초	한국
949	2022	50	픽그	2022-02-23	106,708,970	0.1%	11,615	113	1,600	2022	20211423	드라마,스릴러	91분4초	미국

950 rows × 14 columns



1. 수집대상 (2)

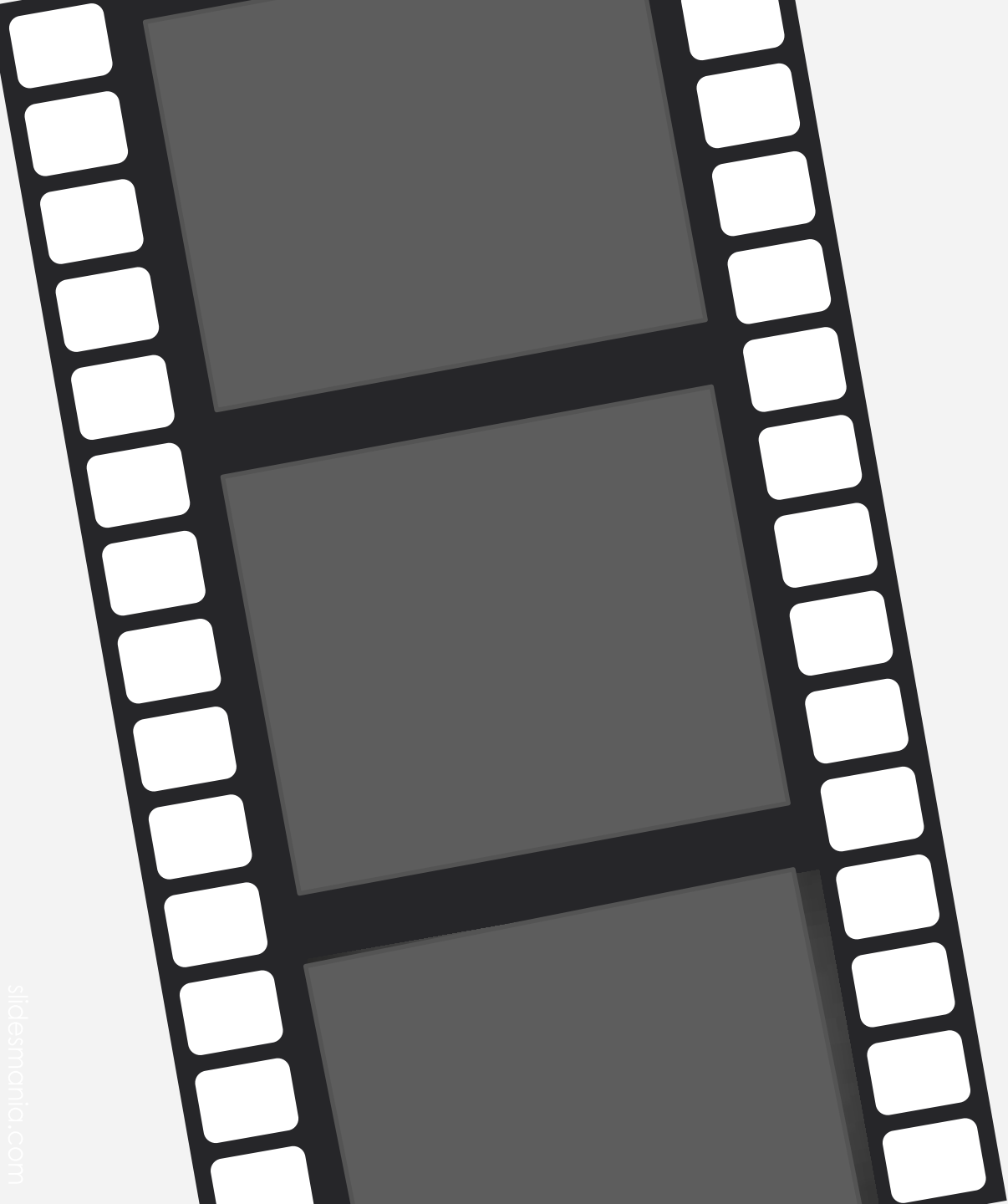
- ❖ 글로벌 OTT 서비스 제공자 별 선호도를 집계하여 정보를 제공하는 'Flix Patrol' 사이트를 통해, 2021년 넷플릭스 유저의 선호 영화 정보 추출



1. 수집대상 (2)

	국가	영화제목	포인트	개봉일	장르	개봉연도
0	Poland	365 Days	573	06/07/2020	Romance	2020
1	South Korea	Kingdom: Ashin of the North	481	07/23/2021	Horror	2021
2	South Korea	Night in Paradise	479	09/03/2020	Crime	2020
3	South Korea	Space Sweepers	432	02/05/2021	Science Fiction	2021
4	Japan	Demon Slayer -Kimetsu no Yaiba- The Movie: Mug...	299	10/16/2020	Animated	2020
...
196	Taiwan	Hospital	2	06/01/2020	Horror	2020
197	United States	Animals on the Loose: A You vs. Wild Interacti...	2	02/16/2021	Adventure	2021
198	South Korea	Default	1	11/28/2018	Drama	2018
199	United States	Get Out	1	02/24/2017	Horror	2017
200	South Korea	The Outlaws	1	10/03/2017	Crime	2017

201 rows × 6 columns



1. 주제

2. 데이터 분석

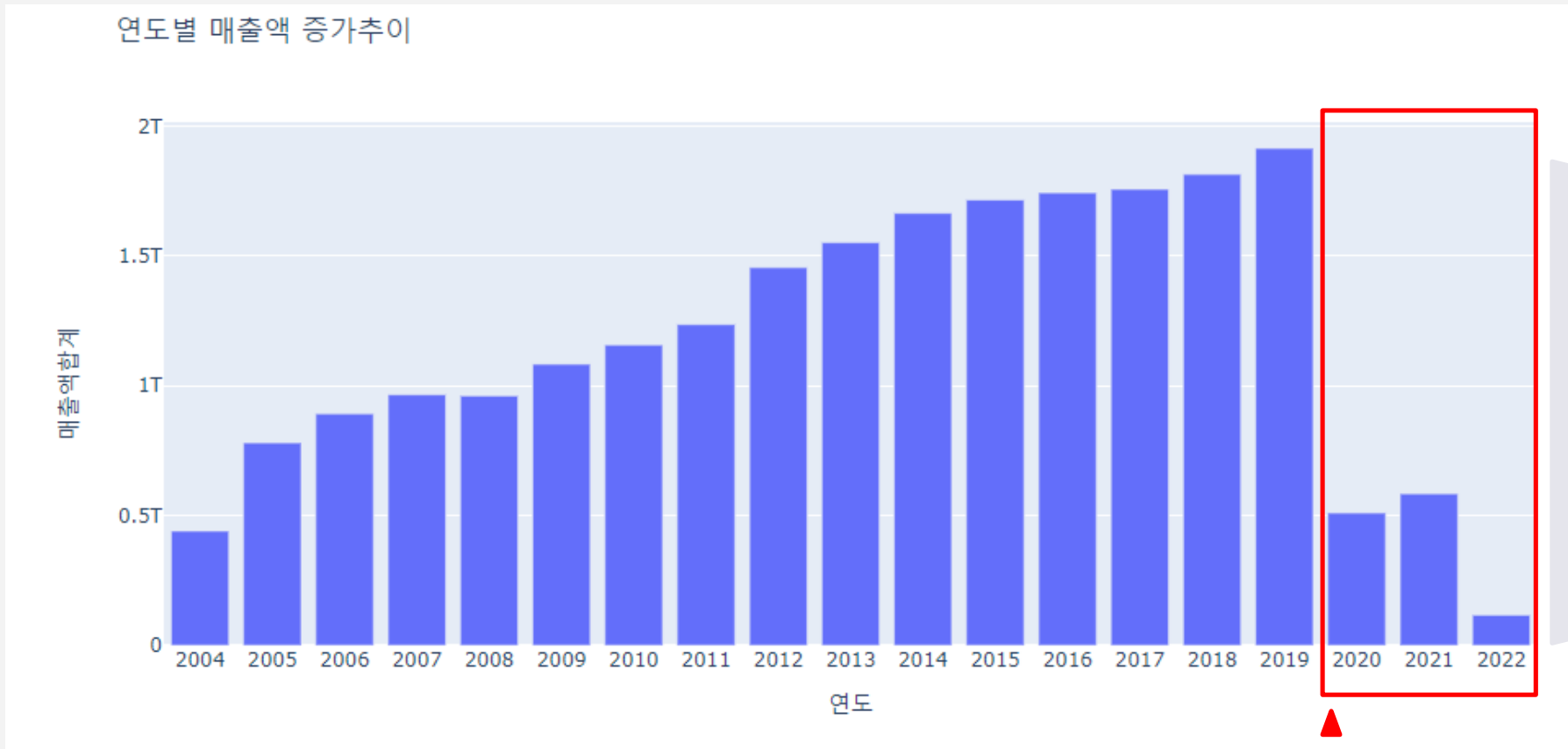
3. 데이터 분석 결과(1) 기본

4. 과정 공유

1. 영화 시장의 연도별 매출액 변화

추이

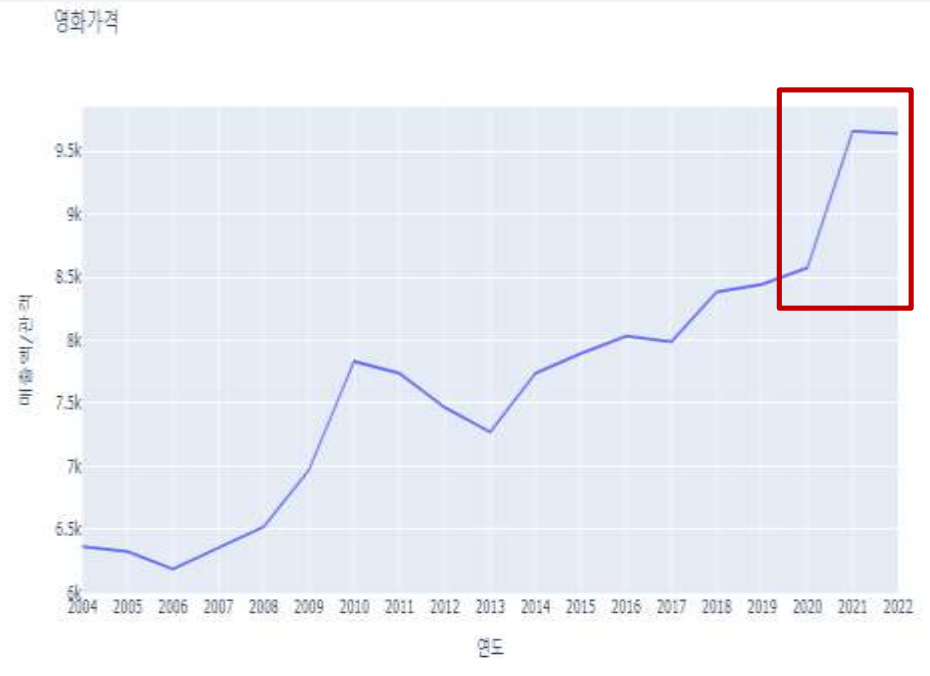
- ❖ 우리나라 영화시장의 매출액은 2004년 이후로 꾸준히 증가 추세였으며, 코로나 이전 3년 간 2조 원에 육박하기에 이르렀음.
- ❖ 중국 우한에서 최초 코로나 발생 이후, 우리나라에 '20.1.20. 첫 확진자가 발생하고 사회적 거리두기(영화관 등 집합금지) 조치 시행으로 모두가 예측할 수 있듯 매출은 급하락 하였으며, 20년은 전년 대비 75% 이상 매출액이 감소함. 이에 대한 반작용으로 OTT 시장이 급격히 성장.



2020.1.20. 한국 코로나 첫 확진자 발생

2. 영화 시장의 영화 가격 변화

❖ 우리나라 영화 티켓 가격은 꾸준히 증가하다, 코로나 이후 영업시간, 띄어 앉기, 주요 기대작들의 개봉 지연에 따른 타격으로 '20,'21,'22년 영화 가격을 1000원 씩 인상하였으며, 최근 3년 간 인상률 평균은 10%를 상회하는 수준으로 이는 기존과 다른 급격한 변화임.



2020년 코로나 피해로 인한 가격 인상, 좌석 차등제 폐지
2021년 코로나로 인한 가격 재인상

코로나 직격탄에 CGV 관람료 인상... 주중·주말 1000원씩

머니S 한영선 기자 | 조회수 : 1,776 | 입력 : 2022.04.05 06:13

5일 CGV에 따르면 지난 4일부터 성인 2D 영화관람료가 1000원 상향 조정돼 주중 1만4000원, 주말 1만5000원으로 올랐다. IMAX를 비롯한 4DX, ScreenX, SPHEREX, 스타리움 등 기술 특별관은 2000원 인상됐다. 씨네드쉐프, 템퍼시네마, 골드클래스 등 고급관은 5000원씩 올랐다. 무비머니(영화관람권)도 동일하게 인상됐다.

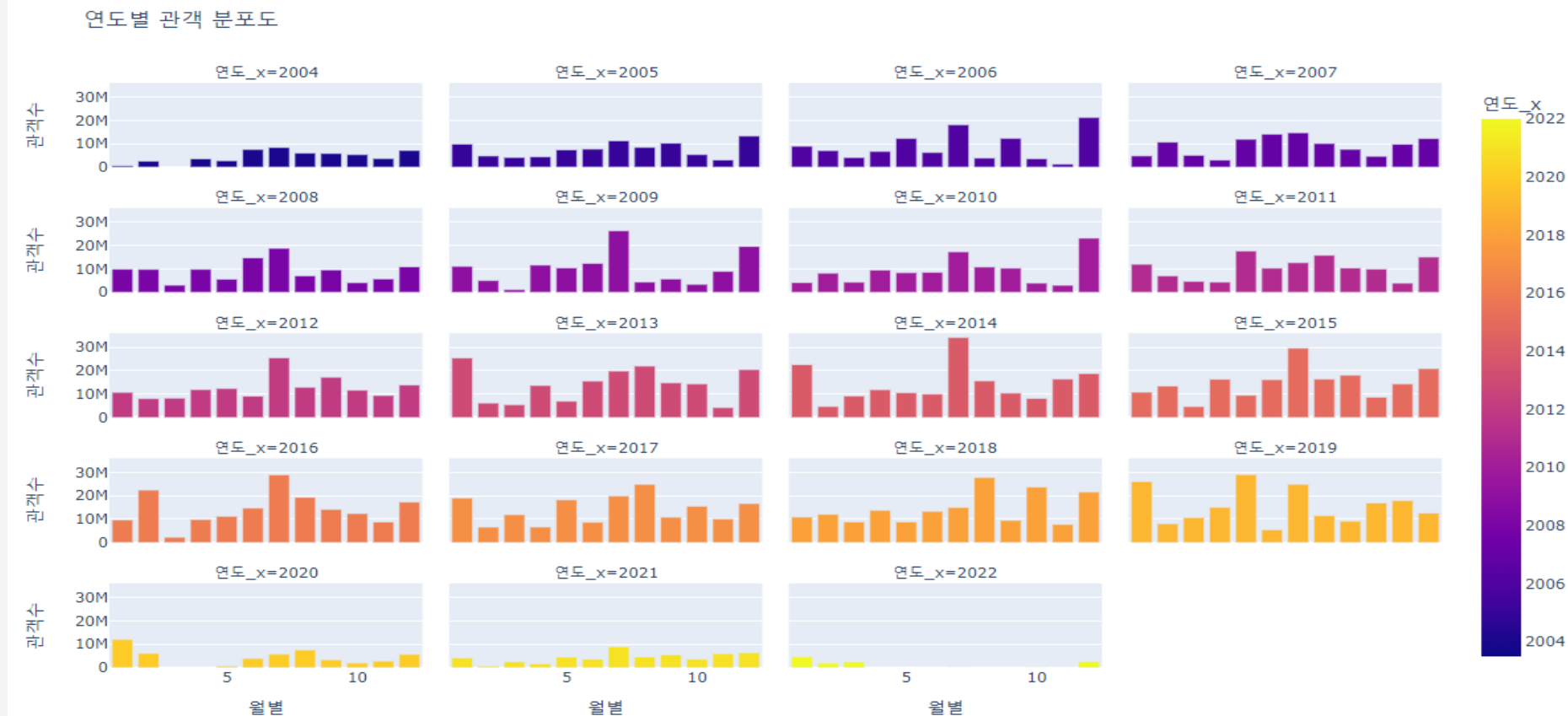
일반관	2D		3D		경로	우대
	일반	청소년	일반	청소년		
평일	모닝	10,000	8,000	11,000	10,000	2D: 7,000 2D: 5,000
	브런치	12,000	10,000	13,000	11,000	3D: 10,000 3D: 8,000
	일반	14,000	11,000	15,000	12,000	
주말	모닝	11,000	8,000	12,000	10,000	
	브런치	15,000	12,000	16,000	13,000	
	일반	15,000	12,000	16,000	13,000	

폭탄 맞은 오프라인 콘텐츠...영화관 매출 3년새 75% ↓

3. 연월별 관객수

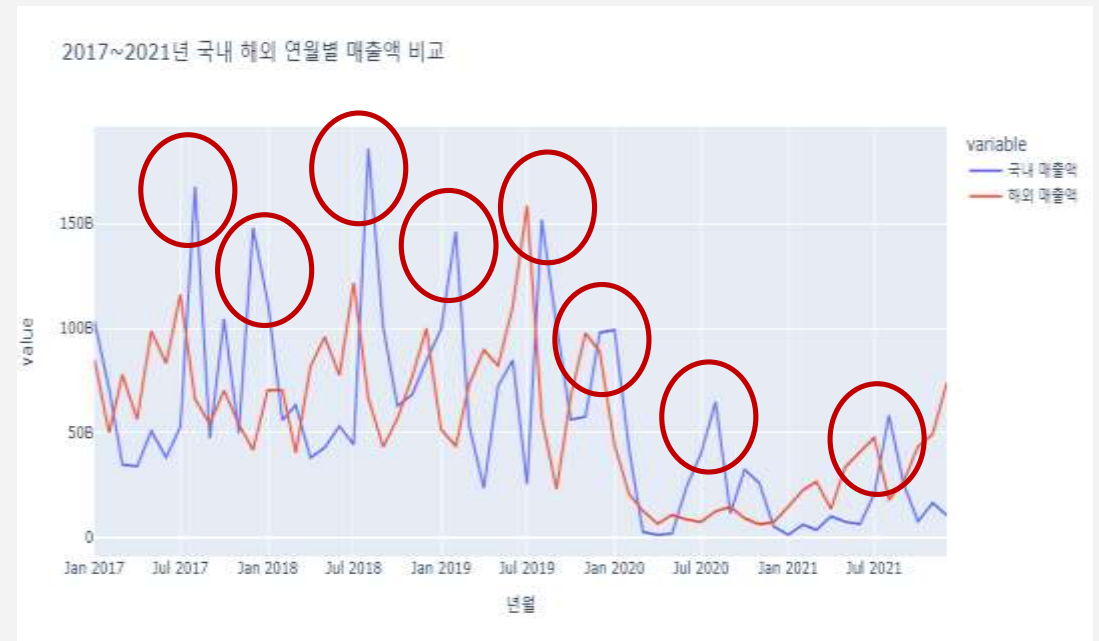
분포도

❖ 주로 여름과 연말에 많은 관객이 영화를 본 것으로 보임



4. 5년간(17~21년)의 월별 관객 분포도 (1)

- ❖ 우리나라 영화 관람객 수를 월별로 확인한 바, 관람객 수는 3~5월(봄)에 적은 편이며, 7~9월(여름) 경에 가장 많이 분포.(8월이 피크)
- ❖ 국내와 해외 영화의 연월별 매출액 비교하였을 때도, 동일하게 연초에는 매출이 크지 않다가 여름 경 매출액이 급증하는 모습을 보임.



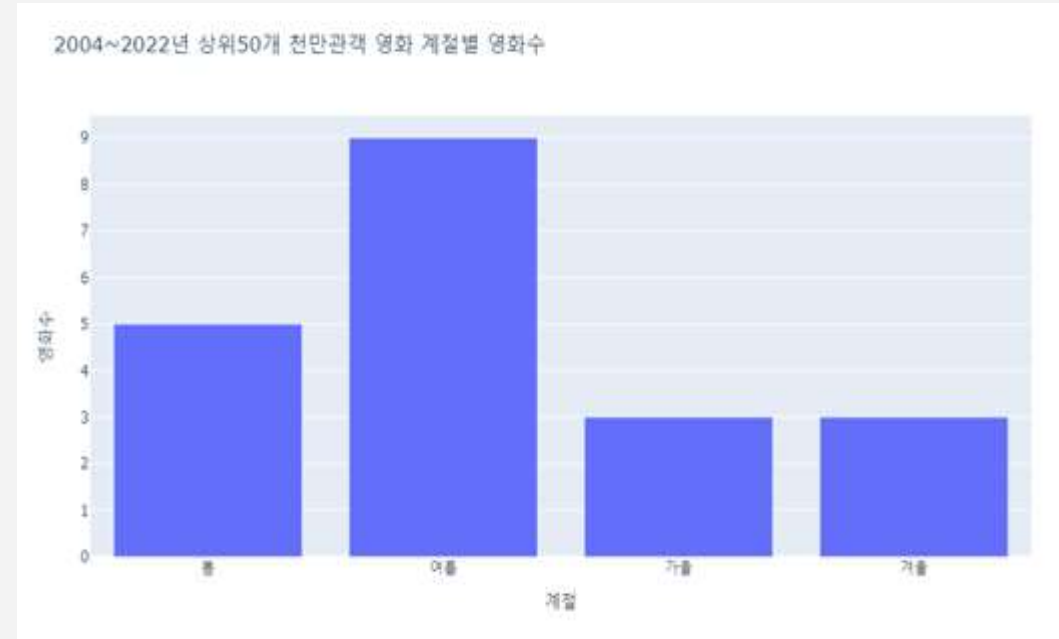
※여름에 관객들이 더 몰리는 원인과 분석 기사

[김형호의 영화시장] 찜통더위 길수록 영화시장 커진다

기상청 날씨데이터 서비스
기상자료개방포털

4. 5년간(17~21년)의 월별 관객 분포도 (2)

- ❖ 그러나 월별 개봉 편수 자료를 확인했을 때, 여름에 관객수와 매출액이 높지만 개봉 편수는 상대적으로 적은 것을 확인.
 - > 여름에는 영화 그 자체의 특성뿐만 아니라 영화관의 쾌적한 환경을 찾기 위해 영화를 본다는 것을 확인 가능.
 - > 천만을 넘은 영화 또한 여름에 개봉시기가 집중 되고 있음을 확인할 수 있음.



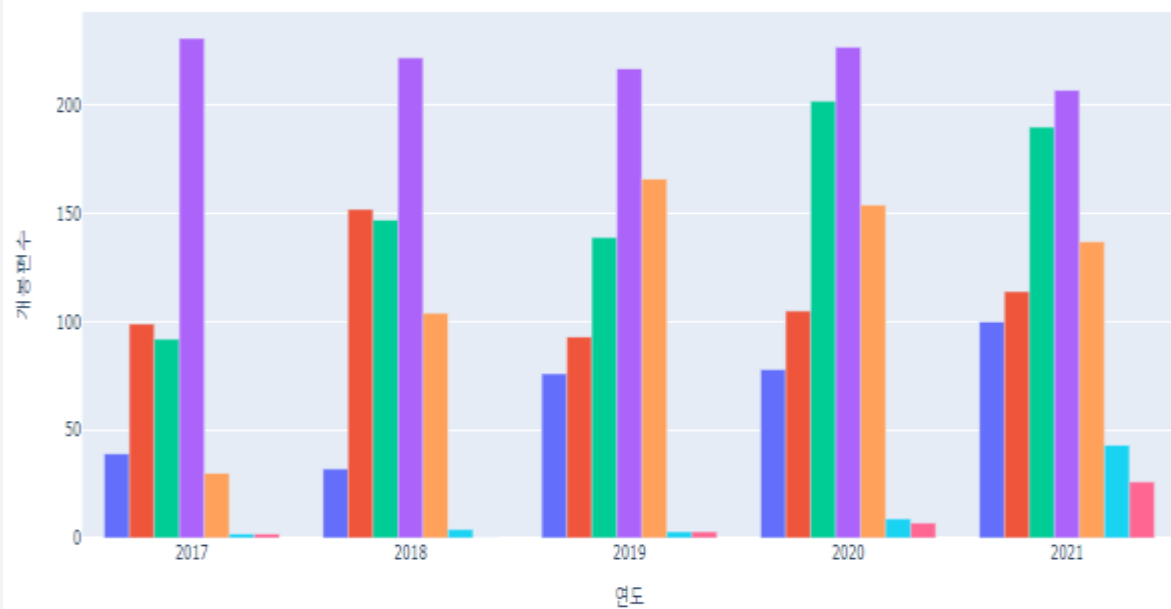
괴물(2006.07.27)	암살(2015.07.22)	신과함께 -인과 연(2018.08.01)
해운대(2009.07.22)	베테랑(2015.08.05)	
도둑들(2012.07.25)	부산행(2016.07.20.)(2020.07.01)	
명량(2014.07.30)	택시운전사(2017.08.02)	

5. 17~21년 전체 영화 연도/요일별

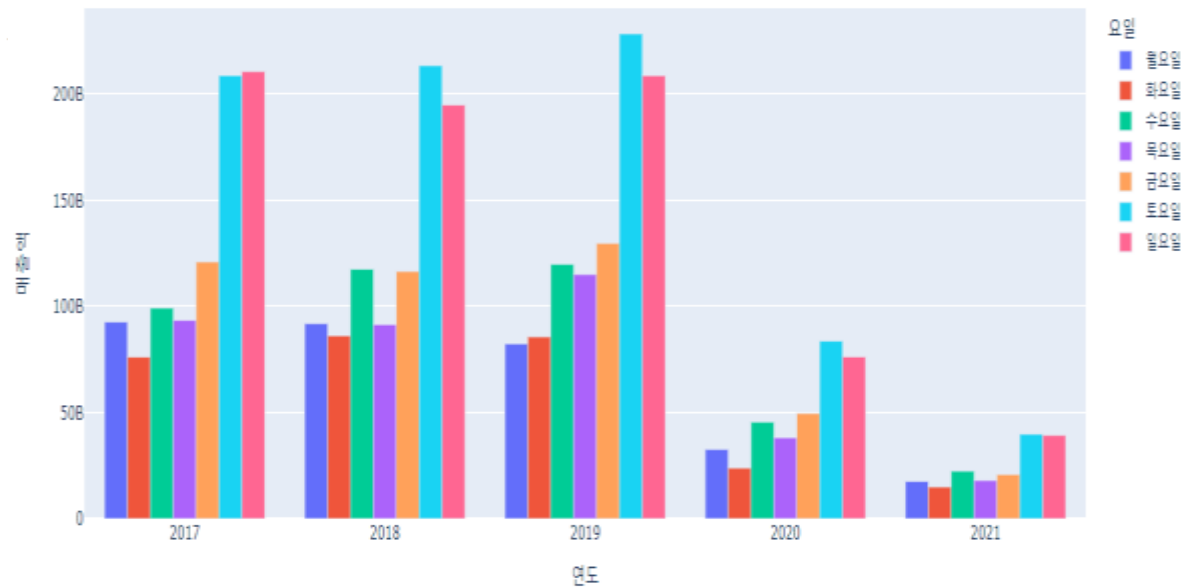
개봉편수/매출액

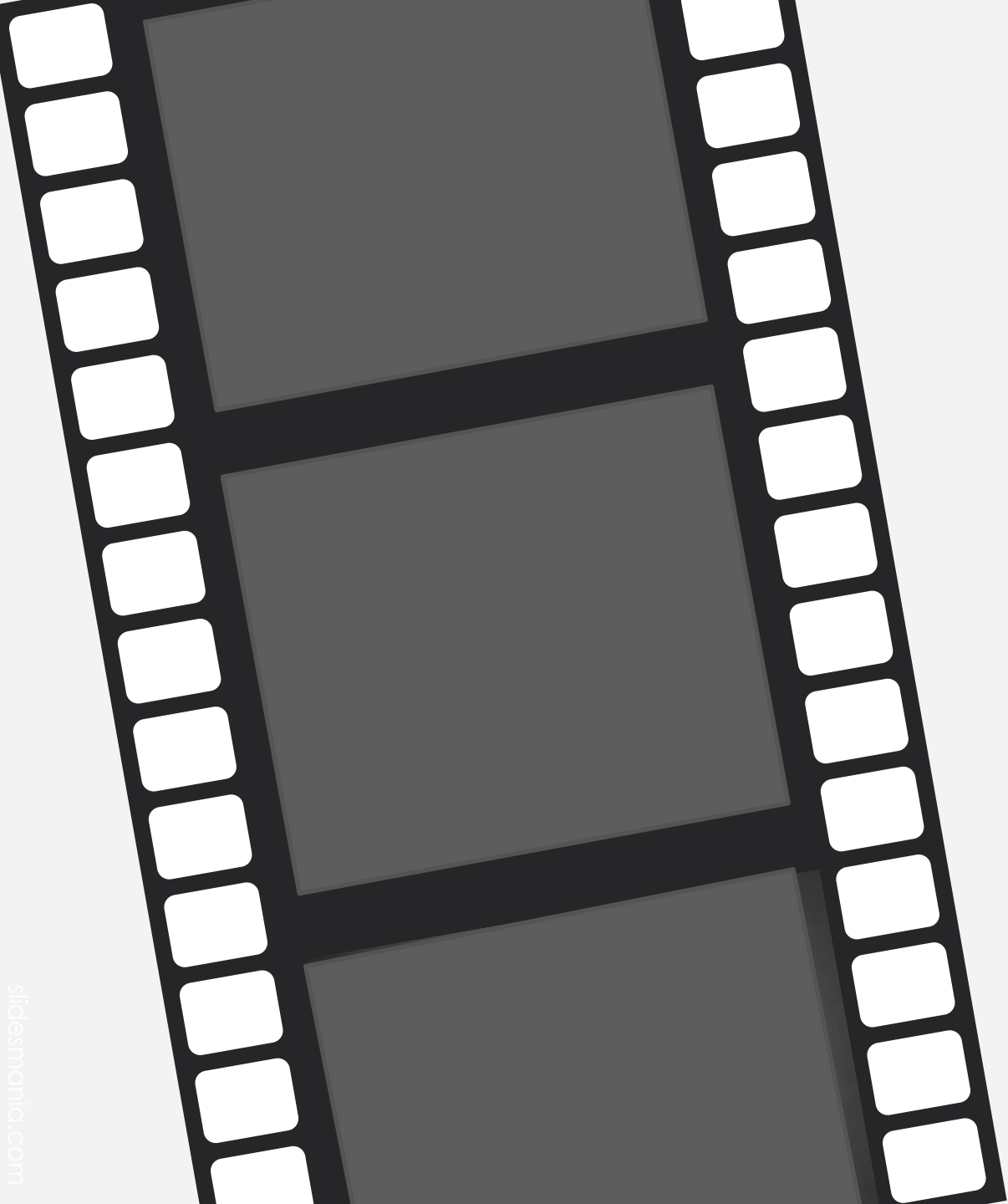
❖ 목요일에 가장 많이 개봉하는 것으로 나타남.

연도, 요일별 개봉편수



연도, 요일별 매출액 분포도





1. 주제

2. 데이터 분석

3. 데이터 분석 결과(2) 추가분석

4. 과정 공유

1. 연도별 상위 50개 영화의 국가별 영화수

비교

- ❖ 연도별 상위 50개에 있는 영화들만큼 매출액과 인지도, 흥행 모든 면에서 높은 평가를 받음.
 - ❖ 순위 내 국내 영화들이 많다면 우리나라의 영화시장이 경쟁력이 강하다고 판단의 근거가 될 수 있음.
 - ❖ 시각화 결과, **외국영화를 우세하다가 비등한, 경쟁에 뒤쳐지지 않는 모습을 보여주며 성장한 영화시장을 볼 수 있었으나 코로나 발생후 21년에 급격하게 줄어드는 모습을 볼 수 있음.**
 - ❖ 이는 코로나로 개봉을 미룬 외국영화들이 21년에 줄지어 개봉함과 한국영화의 흥행 부진이 맞물린 결과.
- > 동일한 조건에서 영화 콘텐츠 자체는 한국 영화도 외국 영화 못지 않게 경쟁력이 있지만, 코로나라는 외부 변수에는 급격히 경쟁력이 저하됨을 확인.

연도별 상위50개 국내/해외 개봉한 영화 갯수



연도별 상위50개 국내/해외 개봉한 영화 갯수



https://www.chosun.com/culture-life/culture_general/2021/10/26/KVXYPMYT3NBFTIQZ5NW3SVU7CQ/

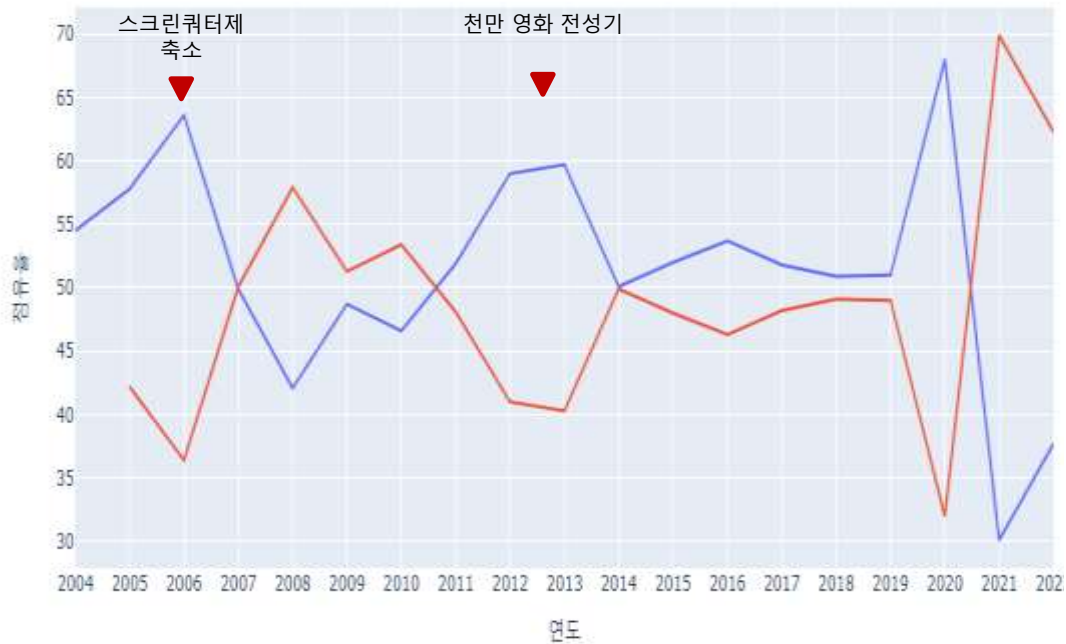
2. 영화 시장의 연도별 국내외 영화 점유율 추이

- ❖ 우리나라 영화시장은 스크린쿼터제 등의 보호조치로 국내 영화의 근소한 우세 속에 점유율을 유지해 왔음.
- ❖ 한국영화 점유율이 2011년 이후로 50% 이상이 계속 유지 되고 있으며 충분히 외국영화와 경쟁하고도 남는 상황이 됨.
- ❖ 코로나 직후인 '20년 해외 영화 점유율은 상당 부분 감소하였으나, '21년, '22년 해외 영화의 역전 현상이 발생하였음.

✓ 스크린쿼터제(한국영화 의무상영제)

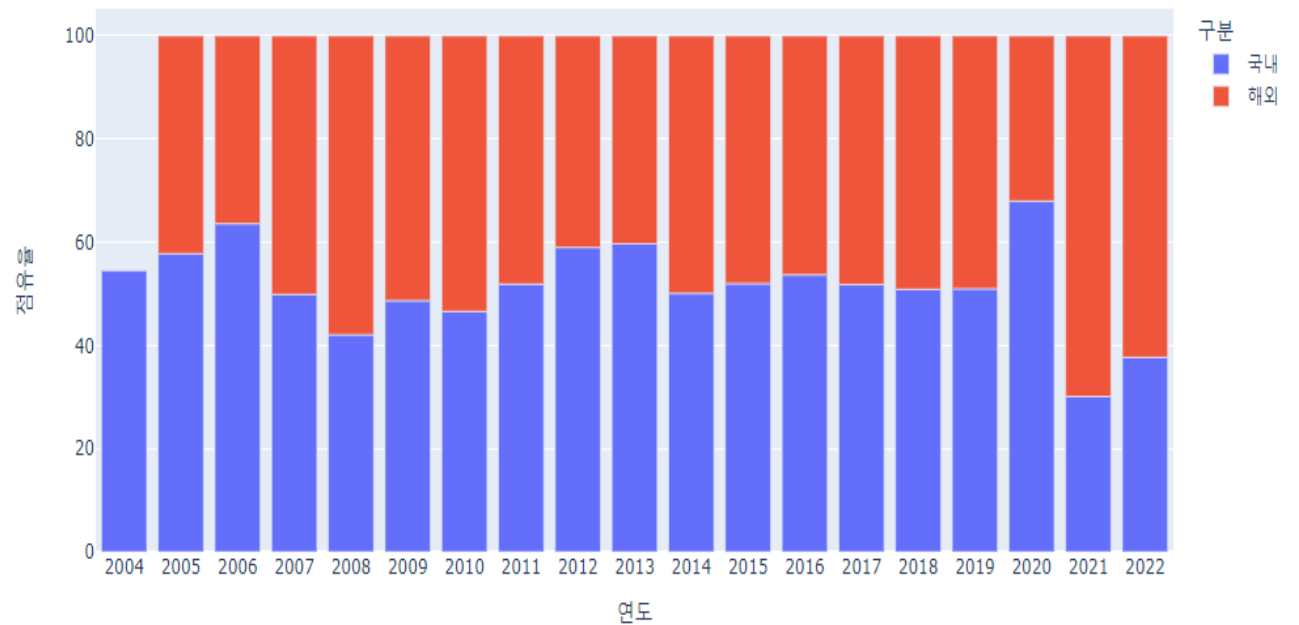
외국영화의 무차별 시장 잠식을 견제해 자국 영화시장을 보호하기 위해 연간 상영일수의 일정기준 이상을 의무적으로 자국 영화를 상영하도록 하는 제도. 우리나라는 1967년 시행되었고, 한·미 FTA 4대 선결 조건으로 2006년 7월 1일부터 한국영화 상영일수를 146일에서 발표했고 영화인들의 거센 반발이 일어난 바 있음.

연도별 국내/해외 영화 점유율 추이



연도별 국내/해외 영화 점유율 추이

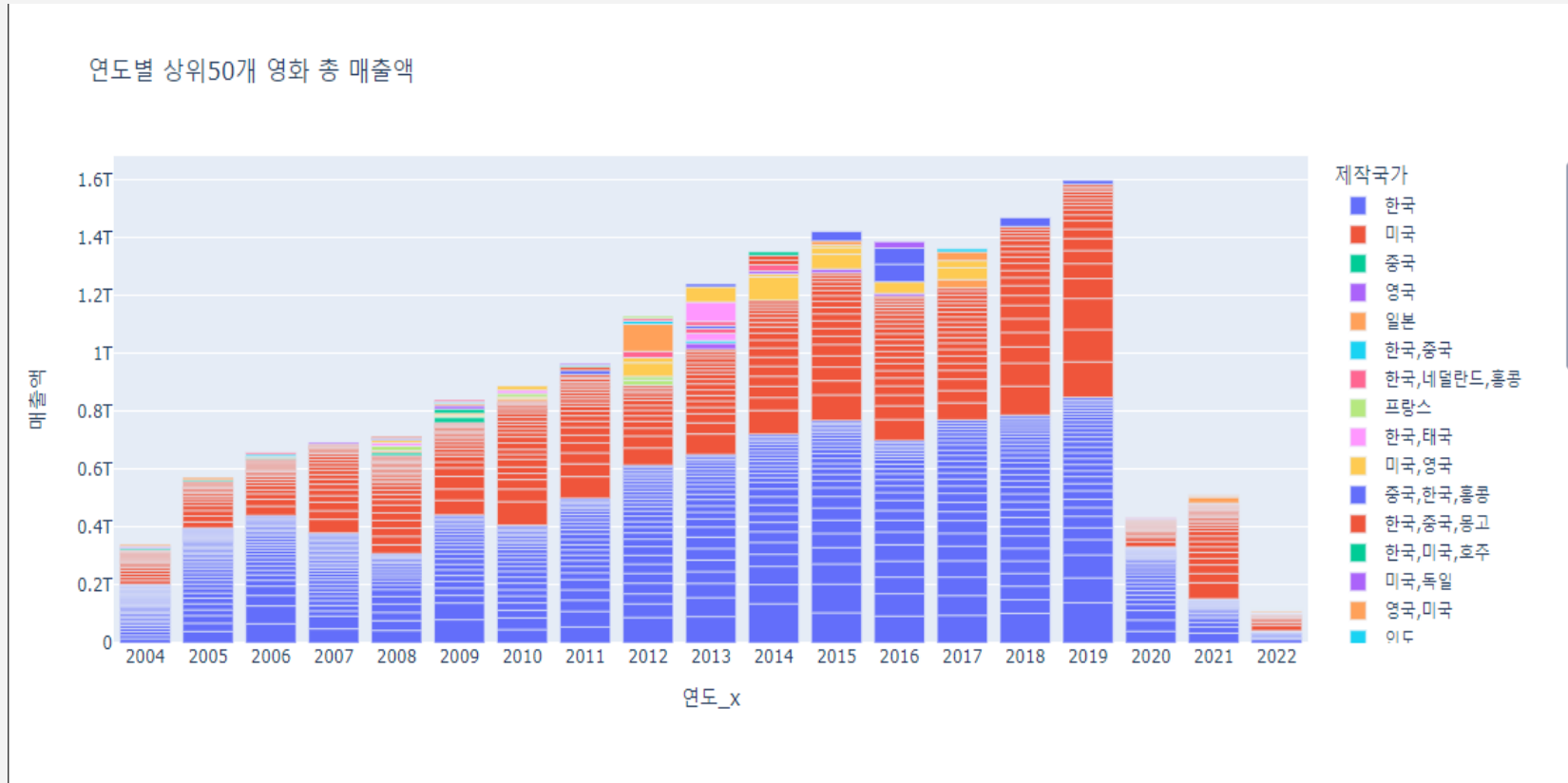
점유율 평균 국내: 51.58% 해외: 48.42%



3. 연도별 상위 50개

영화

- ❖ 연도별 매출액 기준 상위 5개 영화에 대한 제작국가를 추려보니 너무 많은 나라가 있음
- ❖ 카테고리의 정리를 위해 한국, 외국, 한국이 포함된 공동제작 이 3가지로 전처리 실시

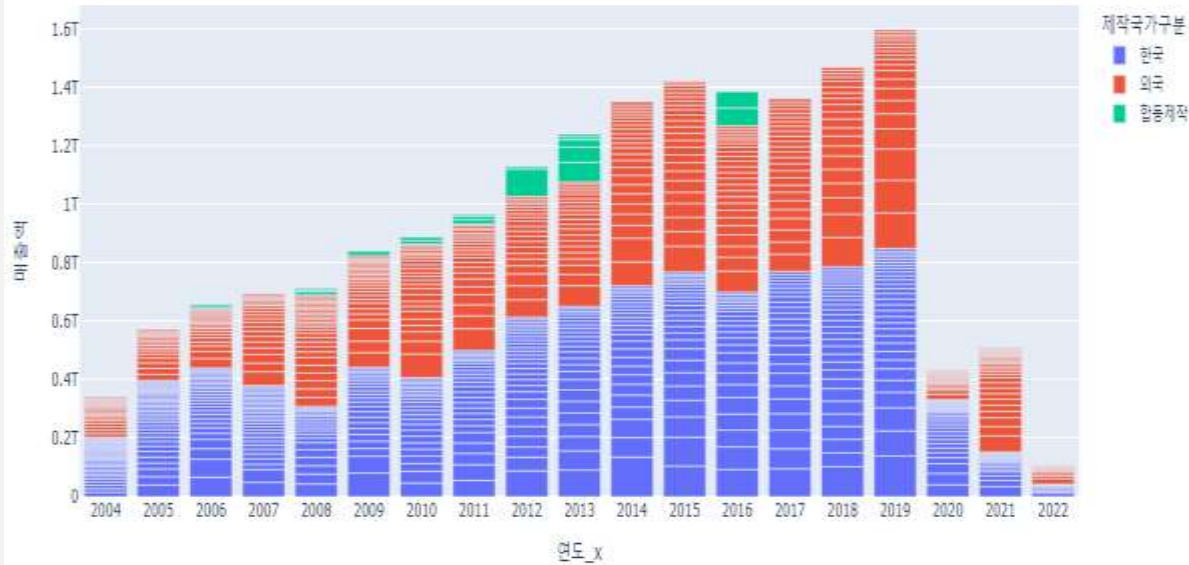


3. 연도별 상위 50개

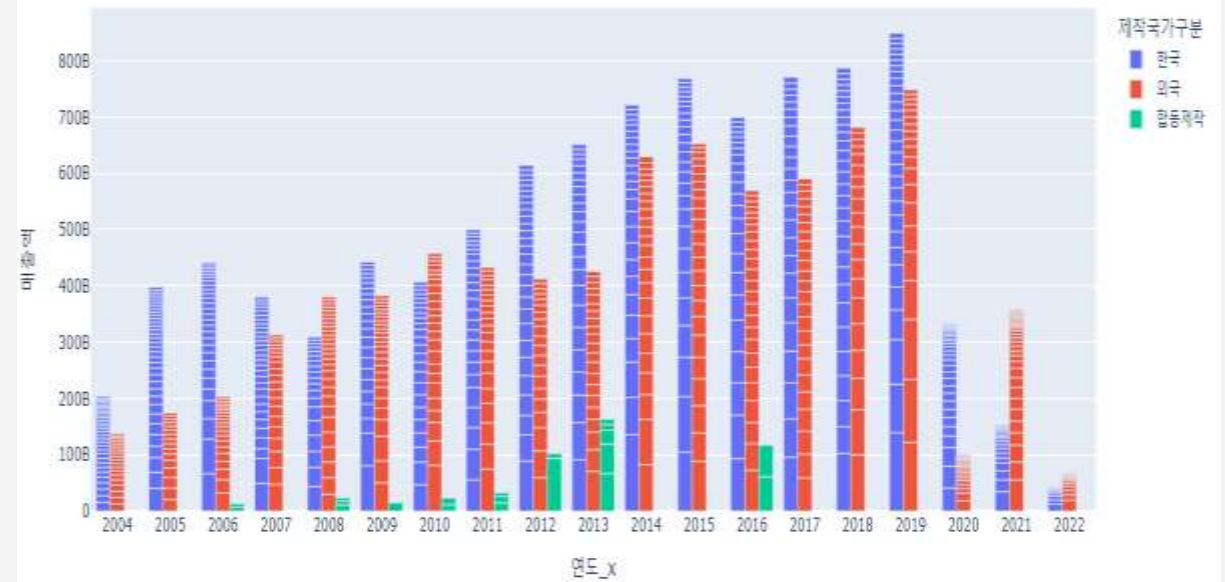
영화

- ❖ 카테고리를 전처리한 후 나온 결과
- ❖ 앞서본 전체 연도별 매출액 점유율에서와 마찬가지로 2020년엔 국내영화가 압도적이지만 2021년엔 역전된 상황을 확인할 수 있음

연도별 상위50개 해외 국내 합동제작 총 매출액



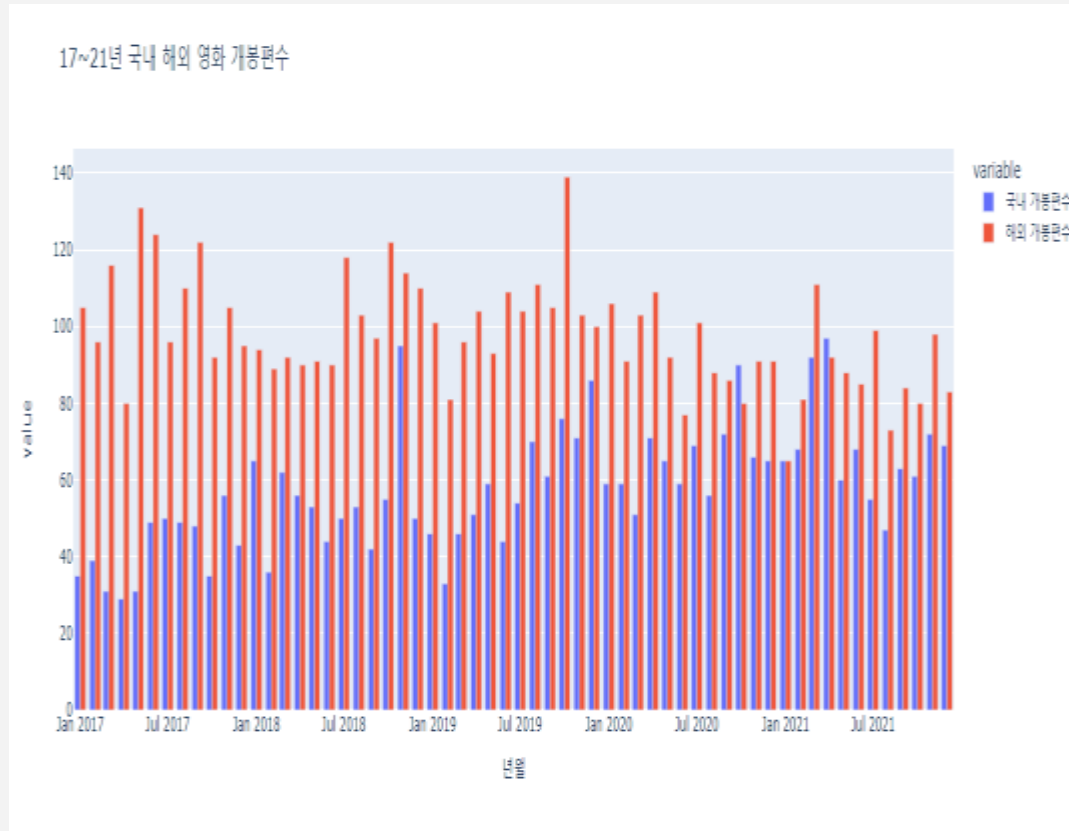
연도별 상위50개 해외 국내 합동제작 총 매출액



4. 17~21년 국내/해외 영화 개봉 편수

비교

- ❖ 해외 영화가 국내 영화보다 더 많이 개봉하고 있음



5. OTT 시장 분석 - 개요

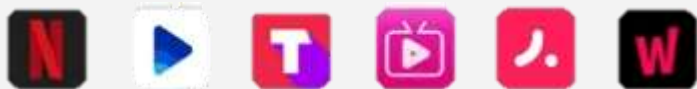
- ❖ 코로나 이후 OTT 시장을 급격히 성장하였으며, 사회적 거리두기 등으로 영화관에 가지 못한 사람들이 자연스레 OTT 시장으로 유입되었음을
- ❖ 추측할 수 있음. 영화산업 매출액 추이와 OTT 시장 성장률은 정확히 반비례 함.

- OTT(Over The Top)란?



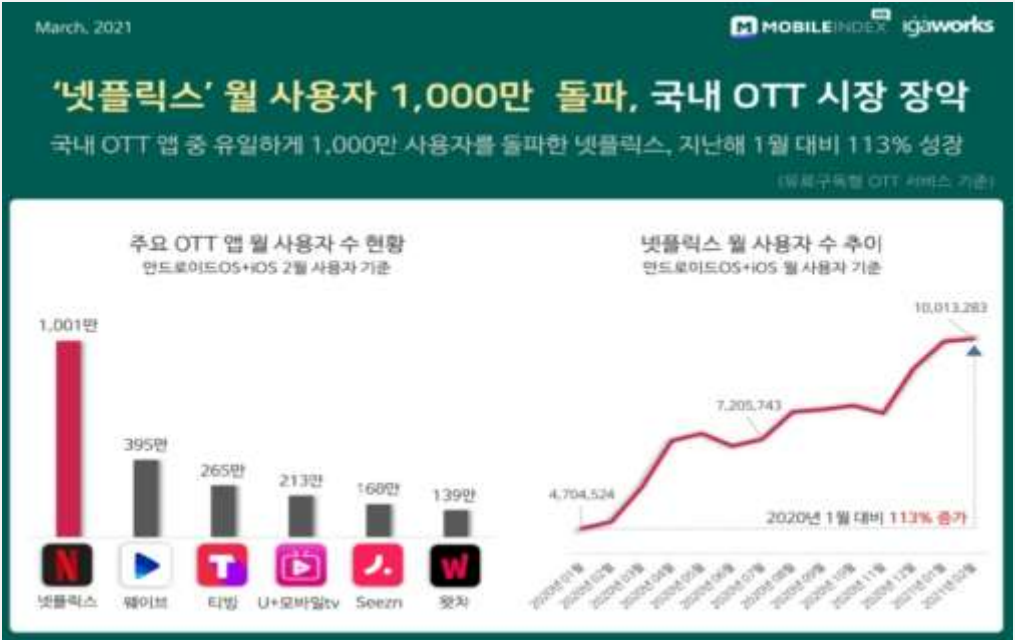
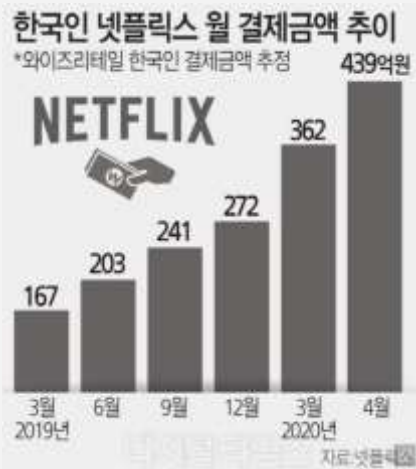
- 인터넷을 통해 영화, 드라마, TV방송 등 각종 영상을 제공하는 서비스.
- 초고속 인터넷의 발달과 보급이 영향을 줌.

-대표적인 OTT 업체
미국의 넷플릭스, 유튜브, 구글TV를 비롯해
국내의 Wavve, 티빙, 왓챠플레이 등이 있다.



- 코로나 이후 급격히 성장한 OTT 시장

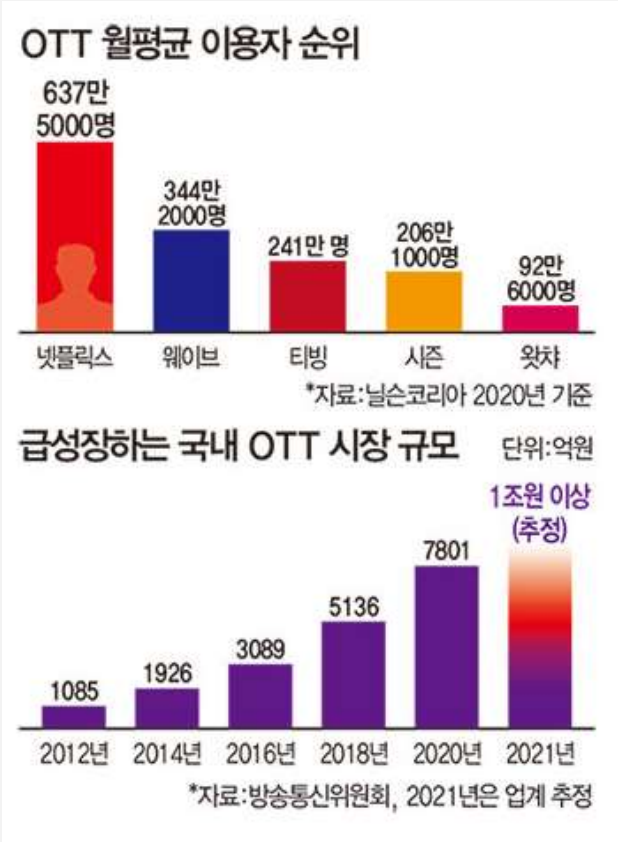
'21.2월 넷플릭스' 월 사용자 수는 1001만3283명
돌파
'20. 1월 470만4524명에 비해 113% 증가한 규모다.



모바일 빅데이터 플랫폼 기업인 '아이지에이웍스'가 15일 '국내 OTT 앱 시장 분석' 리포트 출처 : 미디어오늘(<http://www.mediatoday.co.kr>)

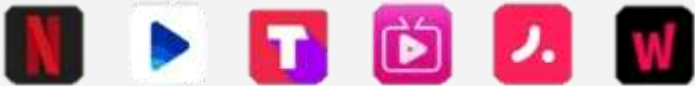
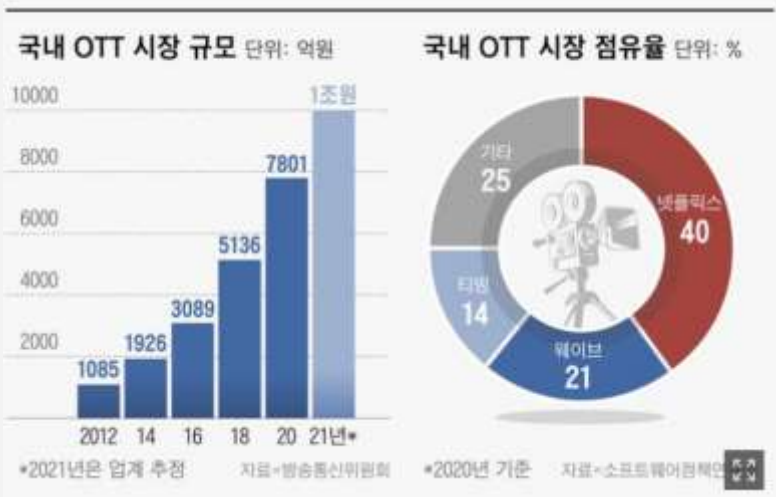
5. OTT 시장 분석 - 개요

- ❖ 코로나 이후 OTT 시장을 급격히 성장하였으며, 사회적 거리두기 등으로 영화관에 가지 못한 사람들이 자연스럽게 OTT 시장으로 유입되었음을
- ❖ 추측할 수 있음. 영화산업 매출액 추이와 OTT 시장 성장률은 정확히 반비례 함.



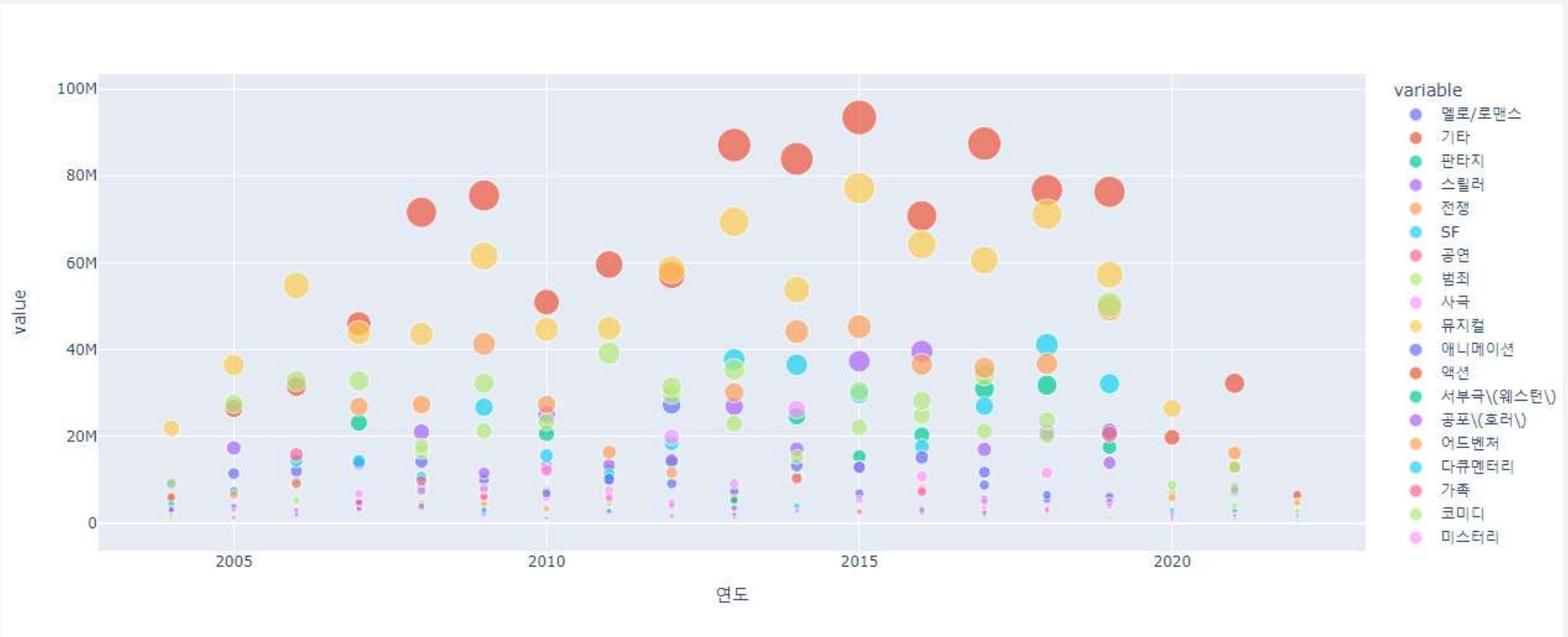
주요 OTT 전략

넷플릭스	올해 5500억원 투자해 오리지널 콘텐츠 13편 제작
디즈니	올 하반기 디즈니플러스 앱 한국 출시 · 국내 통신사와 제휴해 IPTV 서비스 제공
SK텔레콤	웨이브에 2025년까지 총 1조원 투자 · 콘텐츠 제작 스튜디오 설립
KT	콘텐츠 전문법인 설립 · 2023년까지 5000억원 상당 투자해 드라마 100편 제작
CJENM	2023년까지 티빙 오리지널 콘텐츠 제작에 4000억원 투자 · 네이버와 제휴해 멤버십 상품 출시
쿠팡	쿠팡 와우 멤버십과 연계 · 스포츠 경기중계 서비스 제공



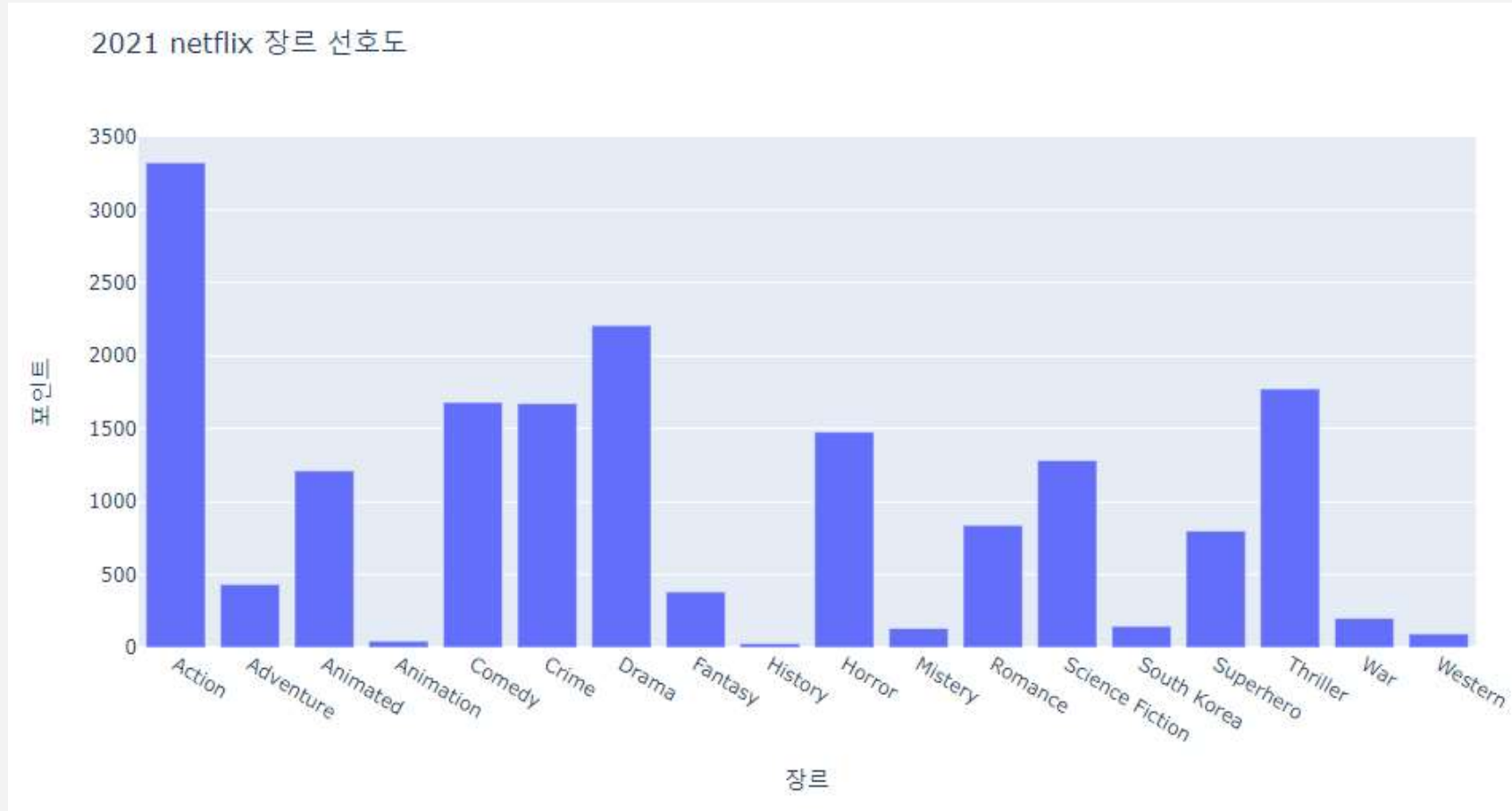
6. 연도별 상위 50개 영화의 장르 분석

- ❖ 상위 50개 영화의 장르에 대한 분석
- ❖ 액션, 드라마가 주를 이루고 있고 그 뒤로 SF, 어드벤처가 뒤따르고 있음



6. OTT 시장 - 한국 넷플릭스 장르별 선호도 분석

❖ 글로벌 OTT 서비스 제공자 별 선호도를 집계 하는 'Flix Patrol' 사이트를 통해 2021년 가장 많은 선호도를 보인 장르를 크롤링하여 분석함.

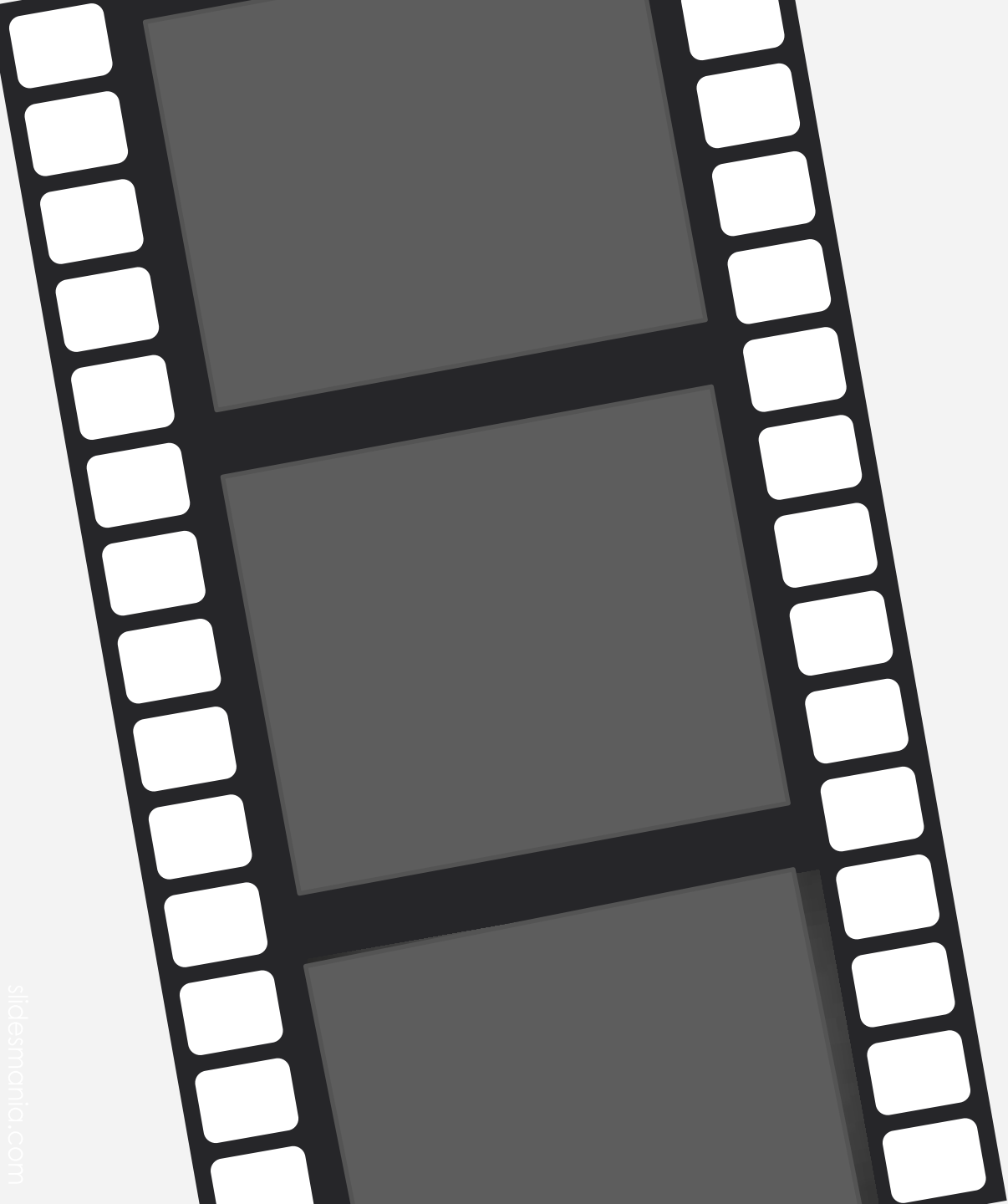


- ✓ 액션>드라마>스릴러>코미디, 범죄>호러 순으로 기존의 영화관에서의 선호와 유사함.
- ✓ 다만, 영화관의 고화질 및 높은 몰입감이 중요한 '어드벤처' 영화의 경우 영화관 보다 선호가 낮다는 것을 확인할 수 있음.



결론

1. 개봉 시기는 목요일이나 금요일을 노리는 것이 좋다.
2. 상업적 성과를 위해선 액션과 드라마 위주의 장르로 영화를 만들어야 한다.
3. OTT 시장 진출을 통해 수익의 극대화를 노려야 하며, 기존 극장 시장과 다른 전략을 취하여야 한다.



1. 주제
2. 데이터 분석
3. 데이터 분석 결과
4. 과정 공유

1. 크롤링 데이터

❖ 데이터 수집의 연도에 관한 변수 자동화

▶ 1. 국내/해외 연도별 통계

```
[ ] 1 # 연도별 통계 범위 자동화
    2
    3 url = 'https://www.kobis.or.kr/kobis/business/stat/them/findYearlyTotalList.do'
    4
    5 headers = {'Referer': 'https://www.kobis.or.kr/kobis/business/stat/them/findYearlyTotalList.do',
    6 | | | | 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.60 Safari/537.36 Edg/100.0.1185.29'
    7 }
    8
    9 data = {'CSRFToken': 'MpMHqnkVhwXicWxDDFFwcr7a_DaPPYQixLtK510v8Co',
   10 | | | | 'loadVal': 0,
   11 | | | | 'searchType': 'search'}
   12
   13 resp = requests.post(url, headers=headers, data=data)
   14 soup = BeautifulSoup(resp.text, 'lxml')
   15
   16 page_tags=soup.select("tbody tr td")
   17
   18 page_tags[::15][:-1]
```


1. 크롤링 데이터 .

❖ 변수 설정 후 크롤링 실시

```
1 # 연도별 통계 변수 생성
2
3 # 한국 영화
4 korea_movies = []
5     '연도' : [],
6     '개봉편수' : [],
7     '상영편수' : [],
8     '매출액' : [],
9     '관객수' : [],
10    '점유율' : []
11 ]
12
13 # 외국 영화
14 overseas_movies = {
15     '개봉편수' : [],
16     '상영편수' : [],
17     '매출액' : [],
18     '관객수' : [],
19     '점유율' : []
20 }
```

```
1 # 연도별 통계 크롤링
2
3 url = 'https://www.kobis.or.kr/kobis/business/stat/them/findYearlyTotalList.do'
4
5 headers = {'Referer': 'https://www.kobis.or.kr/kobis/business/stat/them/findYearlyTotalList.do',
6            'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.60 Safari/537.36 Edg/100.0.1185.29'}
7
8
9 data = {'CSRFToken': 'MpmHqnkVhwXicWxDOFFwcr7a_DaRPYQixLtK510v8Co',
10         'loadVal': 0,
11         'searchType': 'search'}
12
13 resp = requests.post(url, headers=headers, data=data)
14 soup = BeautifulSoup(resp.text, 'html')
15
16 year_tags=soup.select("tbody tr td")
17
18 for i in range(int(len(page_tags[:15][:1]))+1):
19     korea_movies['연도'].append(year_tags[:15][i].text)
20     korea_movies['개봉편수'].append(year_tags[1:15][i].text)
21     korea_movies['상영편수'].append(year_tags[2:15][i].text)
22     korea_movies['매출액'].append(year_tags[3:15][i].text)
23     korea_movies['관객수'].append(year_tags[4:15][i].text)
24     korea_movies['점유율'].append(year_tags[5:15][i].text)
25
26     overseas_movies['개봉편수'].append(year_tags[6:15][i].text)
27     overseas_movies['상영편수'].append(year_tags[7:15][i].text)
28     overseas_movies['매출액'].append(year_tags[8:15][i].text)
29     overseas_movies['관객수'].append(year_tags[9:15][i].text)
30     overseas_movies['점유율'].append(year_tags[10:15][i].text)
31
32 korea_movies = pd.DataFrame(korea_movies)
33 overseas_movies = pd.DataFrame(overseas_movies)
```


2. POST

방식


- ❖ Requests Method가 POST 방식

× Headers Payload Preview Response Initiator Timing Cookies

▼ General

Request URL: `https://www.kobis.or.kr/kobis/business/stat/them/findShowTypeShareList.do`

Request Method: POST

Status Code:  200 OK

Remote Address: `210.104.16.53:443`

Referrer Policy: `strict-origin-when-cross-origin`

2. POST

방식

- ❖ Requests Method가 POST 방식

▼ Request Headers View source

```
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Encoding: gzip, deflate, br
Accept-Language: ko-KR,ko;q=0.9,en-US;q=0.8,en;q=0.7
Cache-Control: max-age=0
Connection: keep-alive
Content-Length: 152
Content-Type: application/x-www-form-urlencoded
Cookie: ACEUCI=1; ACEUACS=1648913529775; JSESSIONID=3XfxvSQVY
Host: www.kobis.or.kr
Origin: https://www.kobis.or.kr
Referer: https://www.kobis.or.kr/kobis/l
sec-ch-ua: " Not A;Brand";v="99", "Chromium";v="100", "Google Chrome";v="100"
sec-ch-ua-mobile: ?0
sec-ch-ua-platform: "Windows"
Sec-Fetch-Dest: document
Sec-Fetch-Mode: navigate
Sec-Fetch-Site: same-origin
Sec-Fetch-User: ?1
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.0.0 Safari/537.36
```

× Headers Payload Preview Response

Form Data view source view URL-encode

```
CSRFToken: qwOI7Wa9WifX2Ux-fBsm3GES2je5a
searchType: search
loadEnd: 0
sSearchFrom: 2022-04-01
sSearchTo: 2022-04-09
sRepNationCd:
sWideAreaCd:
```

```
url = 'https://www.kobis.or.kr/kobis/business/stat/boxs/findYearlyBoxOfficeList.do'

headers = {
    'Referer' : 'https://www.kobis.or.kr/kobis/business/stat/boxs/findYearlyBoxOfficeList.do',
    'User-Agent' : 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.0.0 Safari/537.36'
}

data = {
    'CSRFToken' : 'XCQ-5se9TqMzP8jykvstvj9096rrEnyb25iWmmMoYVKc',
    'loadEnd' : 0,
    'searchType' : 'search',
    'sSearchYearFrom' : page_tag,
    'sMultiMovieYn' : '',
    'sRepNationCd' : '',
    'sWideAreaCd' : ''
}

resp = requests.post(url, headers=headers, data=data)
soup = BeautifulSoup(resp.text, 'lxml')
```

3. 회피

방법

- ❖ 회피용법 사용
- ❖ 원랜 아래와 같이 해당 장르를 포함하는 모든 영화가 나와야하지만 2가지에서 에러가 생김
- ❖ 장르 리스트 안에 () 값이 문제가 되어 '서부극(웨스턴)', '공포(호러)'의 값이 출력되지 않음
- ❖ 파이썬에서 ()는 의미를 갖고 있기 때문에 문자로 인식할 수 있게 회피용법을 적용해야함.

장르
전쟁,드라마
사극
멜로/로맨스
코미디,공포(호러),판타지
SF
...
드라마,스릴러
애니메이션,판타지,멜로/로맨스
판타지,액션,어드벤처
애니메이션
멜로/로맨스,코미디

```
['가족',  
'SF',  
'스릴러',  
'뮤지컬',  
'어드벤처',  
'판타지',  
'사극',  
'범죄',  
'서부극(웨스턴)',  
'멜로/로맨스',  
'드라마',  
'코미디',  
'다큐멘터리',  
'기타',  
'공포(호러)',  
'전쟁',  
'공연',  
'애니메이션',  
'액션',  
'미스터리']
```

```
9 for i in top50_genre :  
10  
11     temp = pd.DataFrame(top50[top50['장르'].str.contains(i)].groupby(['연도']).size())  
12
```

Unnamed: 0	연도	코드	장르	러닝타임	제작국가
4	4	2004	20040490	SF	123분0초 미국
9	9	2004	20040491	SF	126분0초 미국
16	16	2004	20040597	SF	110분0초 미국
54	54	2005	20050144	SF, 드라마, 액션, 스릴러	135분0초 미국
60	60	2005	20050137	드라마, 액션, SF, 스릴러	116분0초 미국
...
897	897	2021	20211281	액션, SF	147분45초 미국
900	900	2022	20210028	액션, 어드벤처, SF	148분1초 미국
914	914	2022	20223676	액션, 드라마, 스릴러, SF, 어드벤처	130분24초 미국
927	927	2022	20210087	SF	155분15초 미국
936	936	2022	20211281	액션, SF	147분45초 미국

[105 rows x 6 columns]

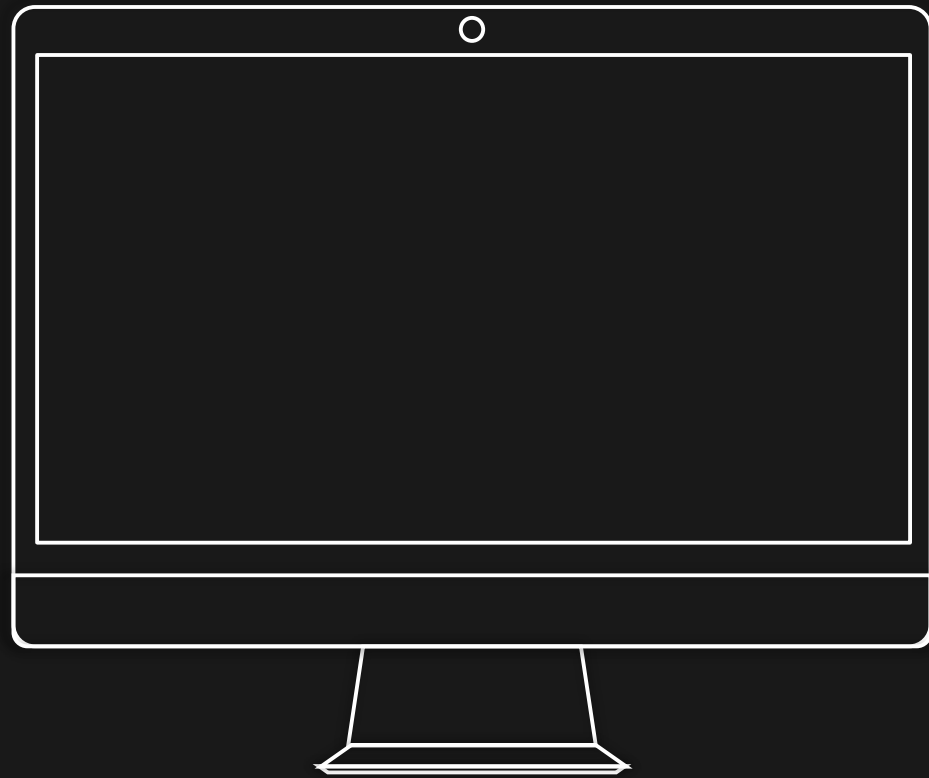
3. 회피 용법

- ❖ replace를 이용, () 앞에 모두 \값을 추가하여 회피용법 설정
- ❖ 오류 없이 값이 출력 됨

```
1 # ()회피용법 적용
2
3 top50_genre = ','.join(list(set(list(top50['장르'].sort_values()))))
4 top50_genre = top50_genre.replace('(','\(').replace(')','\)') # 회피용법 적용
5 top50_genre
6
```

```
['가족',
 '서부극(웨스턴)',
 '다큐멘터리',
 '미스터리',
 '뮤지컬',
 '기타',
 '멜로/로맨스',
 '스릴러',
 '어드벤처',
 '드라마',
 '액션',
 '공연',
 '전쟁',
 '코미디',
 '사극',
 '공포(호러)',
 '범죄',
 'SF',
 '애니메이션',
 '판타지']
```

	가 족	서부극(웨스턴)	다큐멘터 리	미스터 리	뮤지 컬	기 타	멜로/로맨 스	스릴 러	어드벤 처	드라 마	액 션	공 연	전 쟁	코미 디	사 극	공포(호러)	범 죄	SF	애니메이 션	판타 지
연도																				
2005	1	0	0	2	0	0	6	8	3	15	15	0	2	18	0	1	2	3	3	4
2006	4	0	0	1	0	0	8	5	4	22	17	0	0	17	1	2	4	3	3	5
2007	3	0	0	4	0	0	9	8	9	24	14	0	0	18	0	2	2	3	3	7
2008	5	2	0	3	1	0	8	10	12	22	28	0	4	8	4	3	6	6	4	6
2009	4	0	1	6	0	0	7	9	12	25	25	0	2	13	3	1	13	8	3	4
2010	7	0	0	6	0	0	6	16	10	19	19	0	1	11	3	1	5	4	4	10
2011	2	0	0	2	0	0	6	7	4	19	20	0	3	15	1	1	3	5	5	5
2012	1	0	0	2	0	0	9	5	7	18	17	0	0	13	3	1	6	5	6	4



THANK YOU