



영화 흥행에 영향을 미치는 요인 분석

INDEX

01

주제 설명

02

방법론

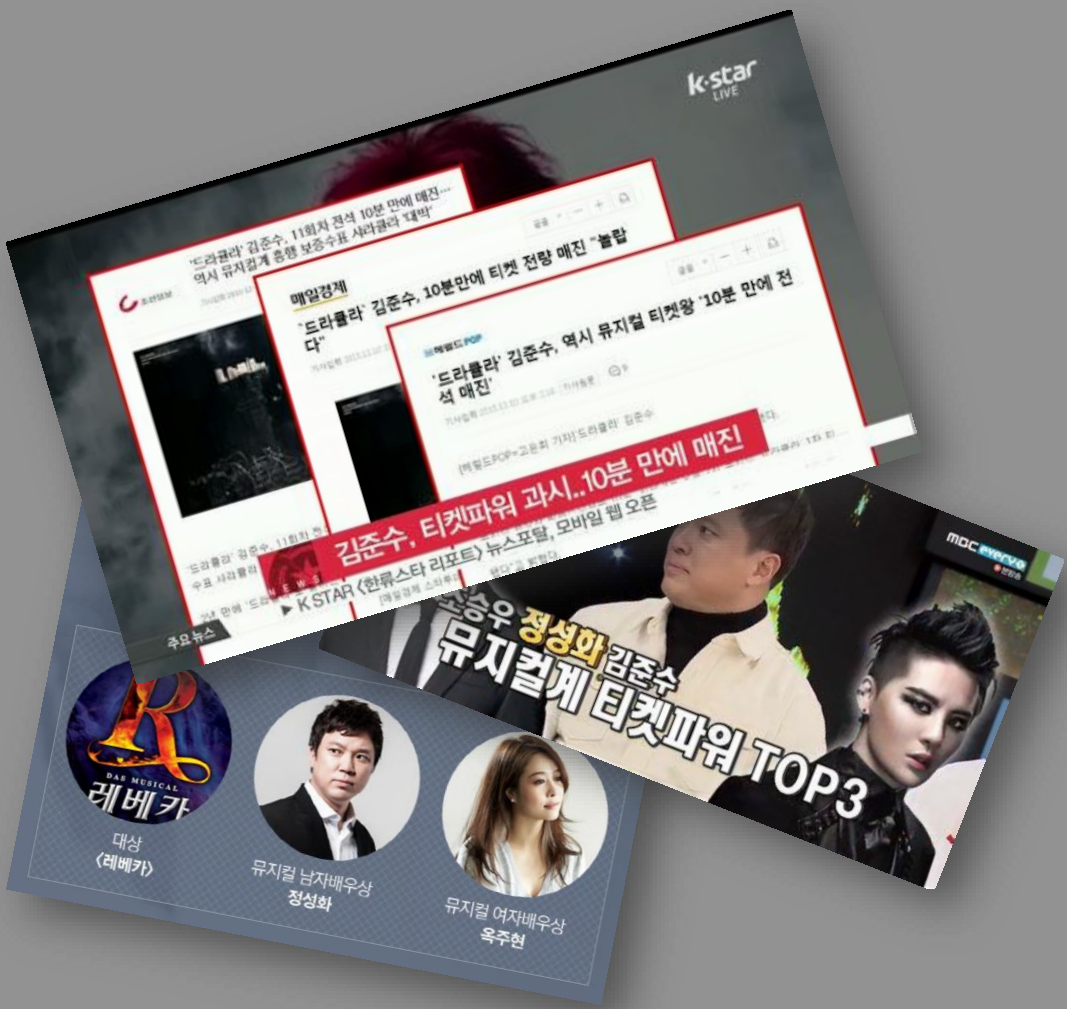
03

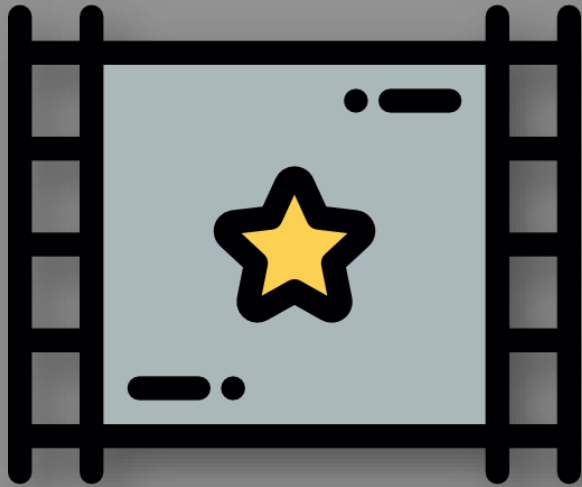
분석 결과

04

향후 계획 및 Q&A

뮤지컬 흥행 요인
티켓파워





영화계에도 배우들의
티켓파워가 있지 않을까?



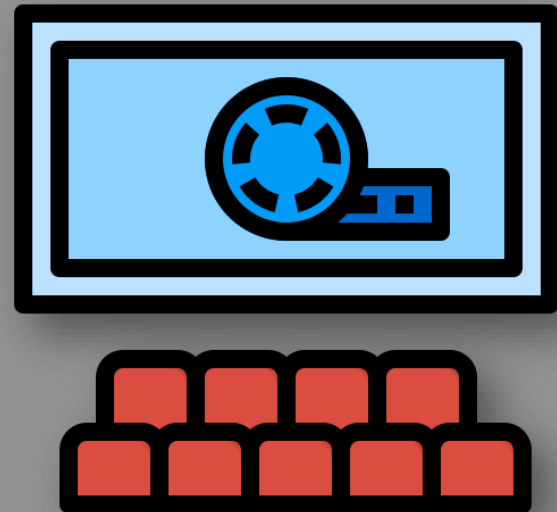
스타 파워

스타가 사람들의 관심을 이끌어 낼 수 있는 힘. 또는 그런 영향력
- 국어사전

“투자자와 배급사들은 스타들의 면면에 따라 계약 여부를 결정한다.
지방 흥행 업자들에게 **스타 파워**는 더욱 위력적이다.” - << 한국경제 >>



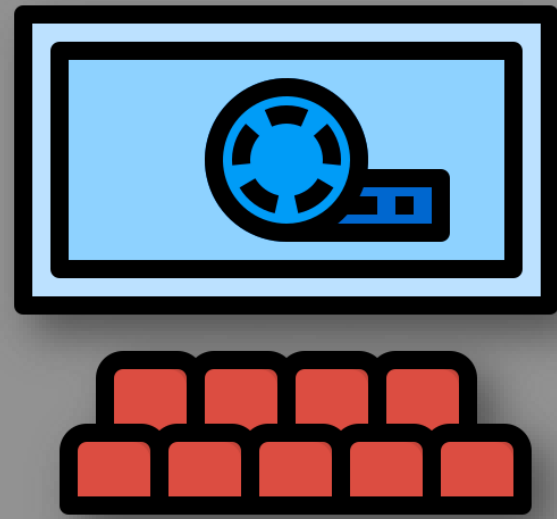
- 01 제작비
- 02 스타
- 03 장르
- 04 관람 연령 등급
- 05 속편
- 06 배급사 파워
- 07 배포 시기
- 08 광고
- 09 홍보
- 10 비평가, 수용자 평가



출처 : 영화 흥행 요인 - 장병희



- 01 ✓ 제작비
- 02 ✓ 스타
- 03 장르
- 04 관람 연령 등급
- 05 ✓ 속편
- 06 배급사 파워
- 07 배포 시기
- 08 광고
- 09 ✓ 홍보
- 10 비평가, 수용자 평가



출처 : 영화 흥행 요인 - 장병희



영화관 입장권 통합 전산망 오픈 API

- 국내 영화 500편의 관객수, 배급사, 감독, 배우

씨네21 & 네이버 영화

- 각 영화들의 네티즌, 평론가 평점



데이터 - 영화관 입장권 통합 전산망 오픈 API



KOFIC 사이트에서 key 발급
미리 다운 받은 500개의 영화의 이름으로
영화 코드를 찾아 필요한 자료 조회

 1 박스오피스	▪ <u>일별 박스오피스</u> ▪ 주간/주말 박스오피스
 2 공통코드조회	▪ 공통코드 조회
 3 영화정보	▪ 영화목록 ▪ 영화 상세정보
 4 영화사정보	▪ 영화사목록 ▪ 영화사 상세정보
 5 영화인정보	▪ 영화인목록 ▪ 영화인 상세정보

<http://www.kobis.or.kr/kobisopenapi/homepg/main/main.do>

```
for (i in 1:dim(df)[1]){  
  name <- df$영화명_공백X[i]  
  
  url <- paste('https://search.naver.com/search.naver?sm=tab_hyt.top&where=nexearch&query=영화', name, sep = '')  
  
  grade <- read_html(url) %>% html_nodes('.r_grade em') %>% html_text()  
  
  df$네티즌 평점[i] <- grade[1]  
  df$평론가 평점[i] <- grade[3]  
  print(i)  
}
```

```
for (i in 1:dim(df)[1]){  
  name <- URLencode(df$영화명_공백X[i])  
  
  url <- paste('http://www.cine21.com/search/?q=', name, sep = '')  
  
  grade <- read_html(url) %>% html_nodes('.num') %>% html_text()  
  
  df$씨네21_네티즌 평점[i] <- grade[1]  
  df$씨네21_평론가 평점[i] <- grade[3]  
  print(i)  
}
```

**Rvest를 이용해 KOFIC에서 가져온 영화명으로 해당 사이트에서
각 영화의 네티즌 평점과 평론가 평점을 파싱한다.**



df

▲	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	↕	
	영화명	개봉 일	매출액	관객 수	스크 린 수	대 표 국 적	배급사	영화명_공백X	영화 코드	장르	main_genre	네 티 즌 평 점	로 딩 가 평 점
1	명량	20140730	135748398910	17613682	1587	한국	씨제이이엔엠(주)	명량	20129370	사극 액션	사극	8.87	6.29
2	극한직업	20190123	139647979516	16264944	1978	한국	씨제이이엔엠(주)	극한직업	20182530	코미디	코미디	9.20	6.80
3	신과함께-죄와 벌	20171220	115698654137	14410754	1912	한국	롯데쇼핑롯데엔터테인먼트	신과함께-죄와벌	20150976	판타지 드라마	판타지	8.73	5.92
4	국제시장	20141217	110913469630	14257115	966	한국	씨제이이엔엠(주)	국제시장	20137048	드라마	드라마	9.16	5.81
5	베테랑	20150805	105168155250	13414009	1064	한국	씨제이이엔엠(주)	베테랑	20148048	액션 드라마	액션	9.24	7.50

cast

↑	영화명	↕	배우
1	명량		최민식 류승룡 조진웅 김명곤 진구 이정현 권율
2	극한직업		류승룡 이하늬 진선규 이동휘 공명 양현민 장진희
3	신과함께-죄와 벌		하정우 차태현 주지훈 김향기 마동석 김동욱 도경수
4	국제시장		황정민 김윤진 오달수 정진영 장영남 라미란 김슬기
5	베테랑		황정민 유아인 유해진 오달수 장윤주 김시후 오대환

영화코드

```
<movie>
  <movieCd>20129370</movieCd>
  <movieNm>명량</movieNm>
  <movieNmEn>Roaring Currents</movieNmEn>
  <prdtYear>2013</prdtYear>
  <openDt>20140730</openDt>
  <typeNm>장편</typeNm>
  <prdtStatNm>개봉</prdtStatNm>
  <nationAlt>한국</nationAlt>
  <genreAlt>사극, 액션</genreAlt>
  <repNationNm>한국</repNationNm>
  <repGenreNm>사극</repGenreNm>
  <directors>
    <director>
      <peopleNm>김한민</peopleNm>
    </director>
  </directors>
  <companys>
    <company>
      <companyCd>20123063</companyCd>
      <companyNm>(주)빅스톤픽처스</companyNm>
    </company>
  </companys>
</movie>
```

```
<movie>
  <movieCd>20157431</movieCd>
  <movieNm>명량: 회오리 바다를 향하여</movieNm>
  <movieNmEn>Roaring Currents: The Road of the Admiral</movieNmEn>
  <prdtYear>2015</prdtYear>
  <openDt>20150507</openDt>
  <typeNm>장편</typeNm>
  <prdtStatNm>개봉</prdtStatNm>
  <nationAlt>한국</nationAlt>
  <genreAlt>다큐멘터리</genreAlt>
  <repNationNm>한국</repNationNm>
  <repGenreNm>다큐멘터리</repGenreNm>
  <directors>
    <director>
      <peopleNm>정세교</peopleNm>
    </director>
    <director>
      <peopleNm>김한민</peopleNm>
    </director>
  </directors>
  <companys>
    <company>
      <companyCd>20123063</companyCd>
      <companyNm>(주)빅스톤픽처스</companyNm>
    </company>
  </companys>
</movie>
```

KOFIC의 오픈 API에 영화명으로 검색할 경우 그 단어가 들어간 모든 영화 데이터가 조회

Df의 **영화명**과 **<movieNm>**의 값이 같는지,

Df의 **대표국적**과 **<repNationNm>**의 값이 같은지를 판단하고 영화 코드를 추출한다.

	영화명	배우	최민식	류승룡	조진웅	김명곤	진구	이정현	권율	이하늬	진선규	이동휘	한효주	양현민
1	명량	최민식 류승룡 조진웅 김명곤 진구 이정현 권율	1	1	1	1	1	1	1	0	0	0	0	0
2	극한직업	류승룡 이하늬 진선규 이동휘 공명 양현민 장진희	0	1	0	0	0	0	0	1	1	1	1	1
3	신과함께-죄와 벌	하정우 차태현 주지훈 김항기 마동석 김동욱 도경수	0	0	0	0	0	0	0	0	0	0	0	0
4	국제시장	황정민 김윤진 오달수 정진영 장영남 라미란 김슬기	0	0	0	0	0	0	0	0	0	0	0	0
5	베테랑	황정민 유아인 유해진 오달수 장윤주 김시후 오대환	0	0	0	0	0	0	0	0	0	0	0	0
6	괴물	송강호 변희봉 박해일 배두나 고아성 오달수 박노식	0	0	0	0	0	0	0	0	0	0	0	0
7	도둑들	김윤석 김혜수 이정재 전지현 임달화 김해숙 오달수	0	0	0	0	0	0	0	0	0	0	0	0
8	7번방의 선물	류승룡 박신혜 갈소원 정진영 오달수 박원상 김정태	0	1	0	0	0	0	0	0	0	0	0	0
9	암살	전지현 이정재 하정우 오달수 조진웅 이경영 최덕문	0	0	1	0	0	0	0	0	0	0	0	0
10	광해, 왕이 된 남자	이병헌 류승룡 한효주 장광 김인권 심은경 김명곤	0	1	0	1	0	0	0	0	0	0	0	0

각 영화에 출연한 배우들의 리스트 중 포스터에 이름이 올라가는 7명 정도의 배우들을 추출
모든 영화에 출연한 배우들을 하나의 열로 분리 한 뒤 해당 영화에 출연한 배우 컬럼을 1로 아니면 0으로 변환

4의 관객수와 결합하여 회귀모델을 만들어 스타파워를 측정



	영화명	관객 수	복합장르	대표장르
1	명량	17613682	사극 액션	사극
2	극한직업	16264944	코미디	코미디
3	신과함께-죄와 벌	14410754	판타지 드라마	판타지
4	국제시장	14257115	드라마	드라마
5	베테랑	13414009	액션 드라마	액션
6	괴물	13019740	SF 가족 드라마	SF
7	도둑들	12983330	액션 범죄	액션
8	7번방의 선물	12811206	코미디 드라마	코미디
9	암살	12705700	액션 드라마	액션
10	광해, 왕이 된 남자	12319542	사극 드라마	사극

각 영화의 대표장르와 복합장르로 그룹화 한 뒤 관객수의 평균수치를 구한다.

배급사 파워

	영화명	관객 수	배급사	씨제이이엔엠	롯데쇼핑롯데엔터테인먼트	소박스	넥스트엔터테인먼트월드	시네마서비스	롯데컬처웍스롯데엔터테인먼트	씨제이엔터테인먼트	위너브러더스코리아
1	명량	17613682	씨제이이엔엠(주)	1	0	0	0	0	0	0	0
2	극한직업	16264944	씨제이이엔엠(주)	1	0	0	0	0	0	0	0
3	신과함께-죄와 벌	14410754	롯데쇼핑롯데엔터테인먼트	0	1	0	0	0	0	0	0
4	국제시장	14257115	씨제이이엔엠(주)	1	0	0	0	0	0	0	0
5	베테랑	13414009	씨제이이엔엠(주)	1	0	0	0	0	0	0	0
6	괴물	13019740	(주)소박스	0	0	1	0	0	0	0	0
7	도둑들	12983330	(주)소박스	0	0	1	0	0	0	0	0
8	7번방의 선물	12811206	(주)넥스트엔터테인먼트월드(NEW)	0	0	0	1	0	0	0	0
9	암살	12705700	(주)소박스	0	0	1	0	0	0	0	0
10	광해, 왕이 된 남자	12319542	씨제이이엔엠(주)	1	0	0	0	0	0	0	0
11	왕의 남자	12302831	(주)시네마서비스	0	0	0	0	1	0	0	0
12	신과함께-인과 연	12274996	롯데컬처웍스(주)롯데엔터테인먼트	0	0	0	0	0	1	0	0
13	택시운전사	12186684	(주)소박스	0	0	1	0	0	0	0	0

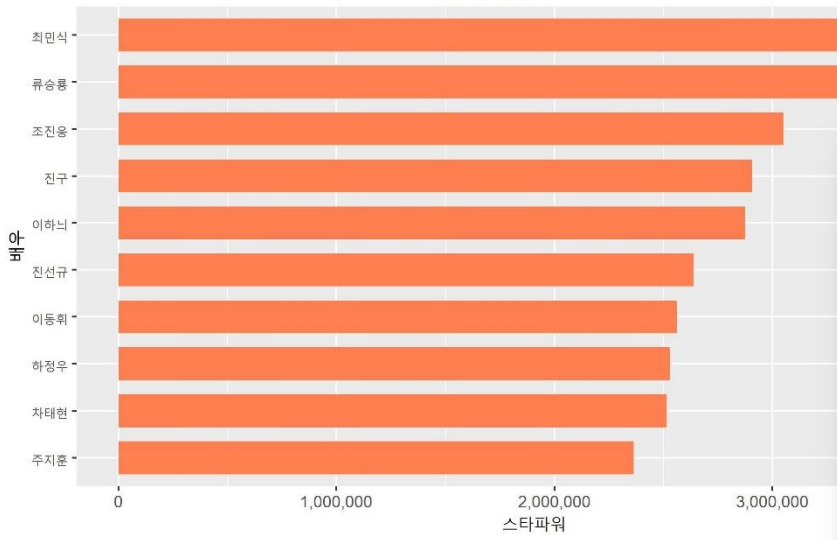
스타와 마찬가지로 방법으로 데이터셋을 구성한 뒤 **회귀모델**을 통해 배급사 파워를 측정한다.

비평가, 수용자의 평가

	영화명	관객 수	네티 즌평 점	평론 가평 점	씨네 21_ 네티 즌평 점	씨네 21_ 평론 가평 점
1	명량	17613682	8.87	6.29	8.62	6.54
2	극한직업	16264944	9.20	6.80	8.95	7.05
3	신과함께-죄와 벌	14410754	8.73	5.92	8.48	6.17
4	국제시장	14257115	9.16	5.81	8.91	6.06
5	베테랑	13414009	9.24	7.50	8.99	7.75
6	괴물	13019740	8.62	8.00	8.37	8.25
7	도둑들	12983330	7.62	7.05	7.37	7.30
8	7번방의 선물	12811206	8.83	6.58	8.58	6.83
9	암살	12705700	9.10	6.57	8.85	6.82
10	광해, 왕이 된 남자	12319542	9.24	7.27	8.99	7.52

revest의 read_html을 이용해 네이버 영화와 씨네21에서 해당 영화의 네티즌 평점과 평론가 평점을 파싱
각 평점 별 관객수의 산점도를 통해 어떠한 상관관계가 있는지 조사한다.

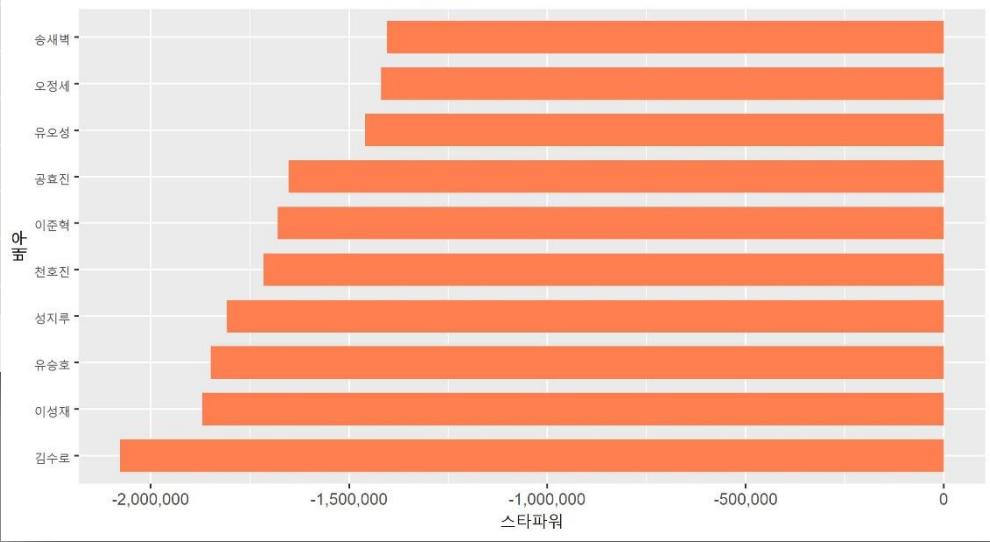
스타파워 TOP 10



lm (관객수 ~ 배우1 + 배우2 ..)

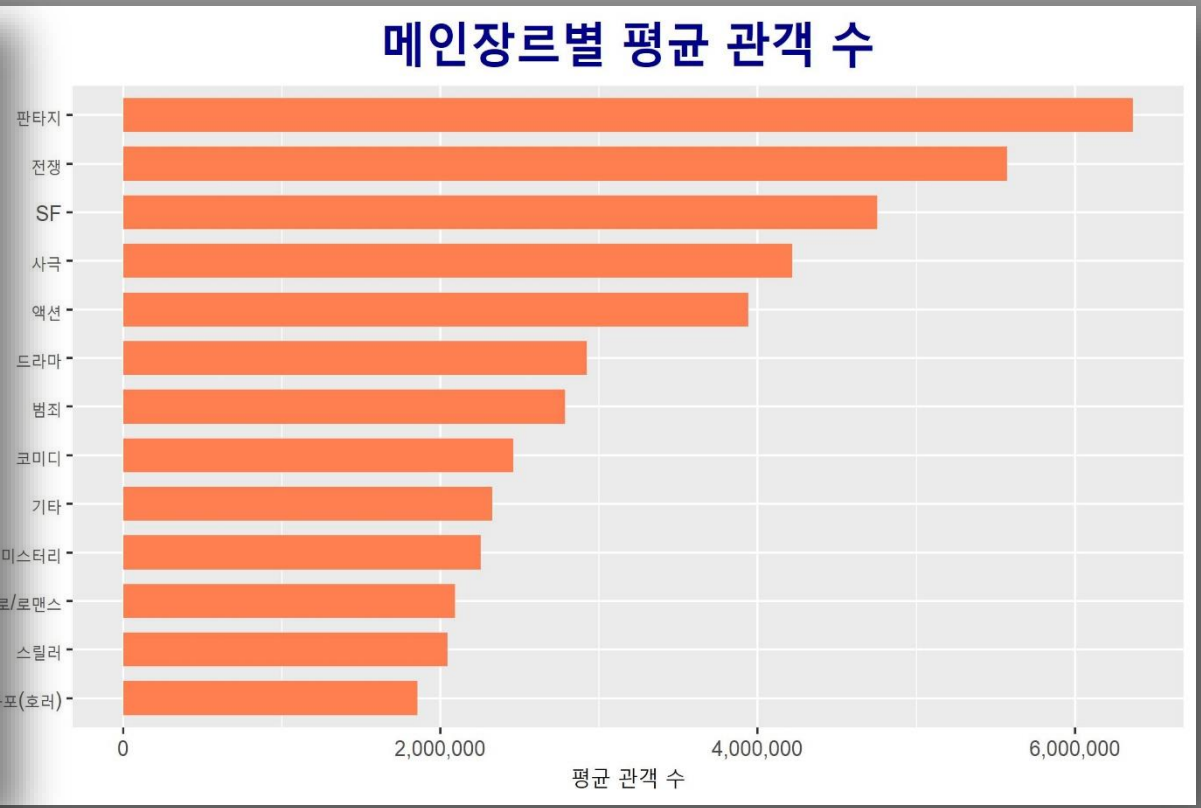
R-squared 0.53 / Adjusted R-squared 0.30
p-value 0.05 이하

스타파워 BOTTOM 10



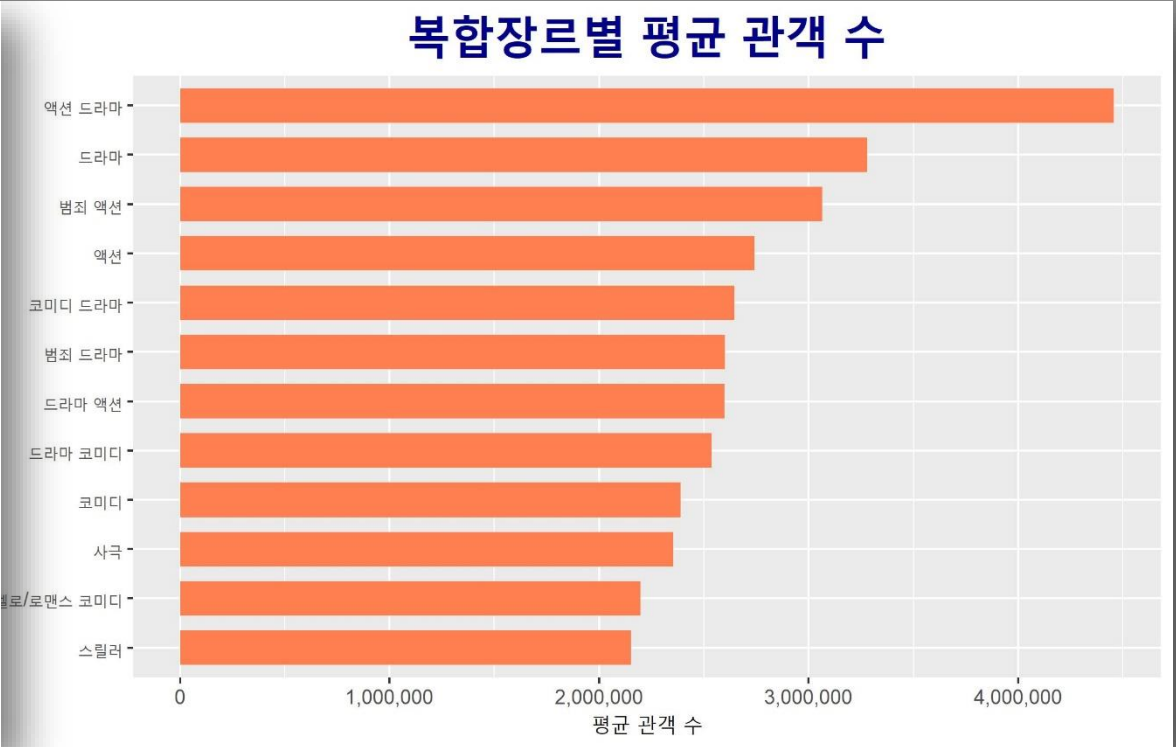
5번 이상의 영화에 출연한 배우들로 회귀모델을 생성
독립변수의 수가 늘어가 결정계수값은 0.53으로 올라갔으나 수정된 결정계수는 0.3으로 낮은 수치
각 변수의 회귀계수 값을 스타파워로 정의해 상위 10명, 하위 10명을 조사

	관객수	min	max	count
main_genre				
판타지	6365105	1168625	14410754	5
전쟁	5571933	1140876	11746135	8
SF	4755124	856150	13019740	6
사극	4219329	958775	17613682	28
액션	3941687	904802	13414009	53
드라마	2921824	844103	14257115	130
범죄	2784205	855980	9707158	50
코미디	2460522	856090	16264944	104
기타	2323996	1408882	4583645	5
미스터리	2253209	1063421	6879908	16
멜로/로맨스	2092611	848148	6654837	36
스릴러	2042448	870785	5604104	32
공포(호러)	1856235	867386	4515833	17



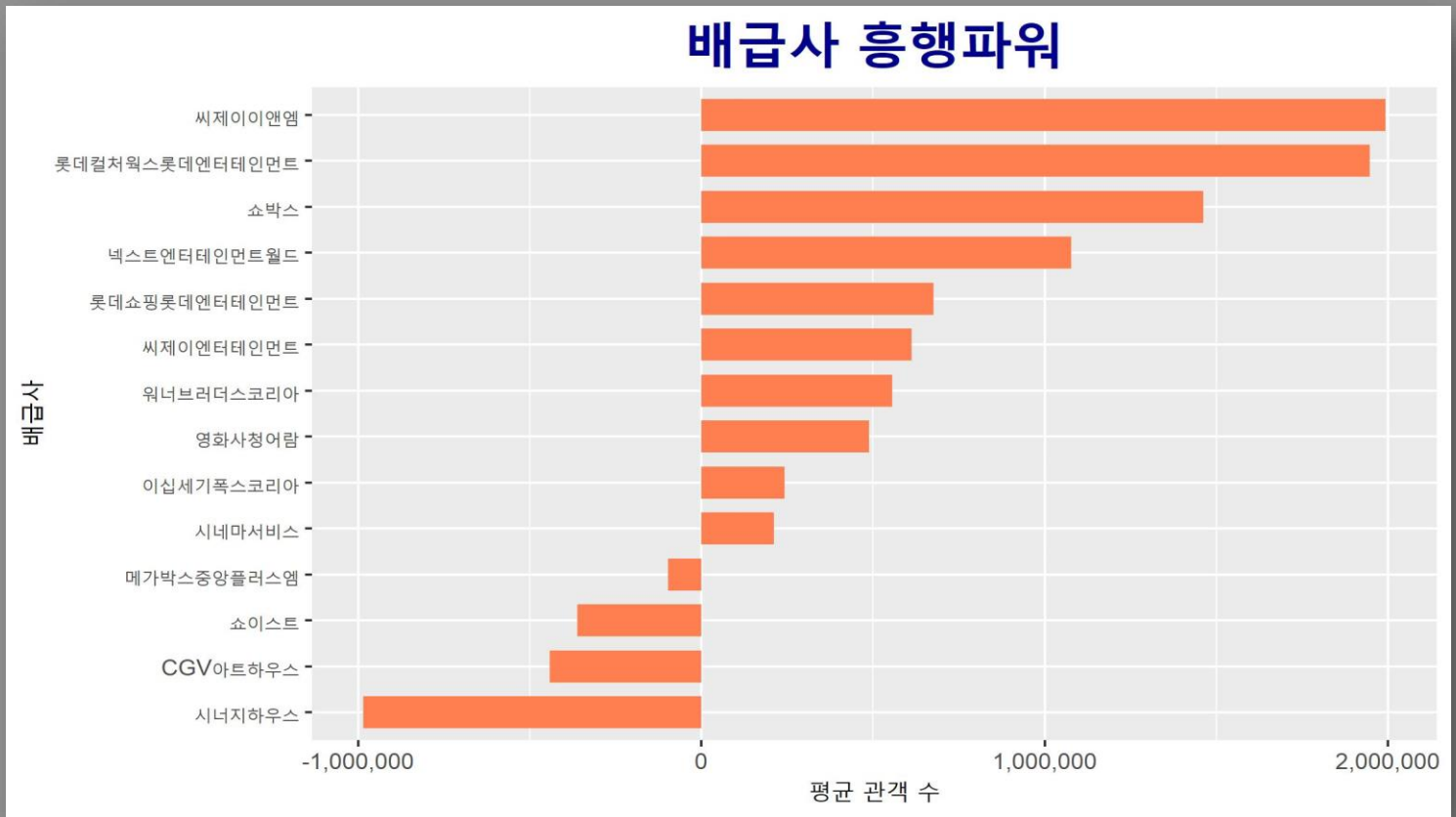
드라마와 코미디 장르의 영화가 100편 이상 개봉했다.
평균 관객수는 판타지, 전쟁, SF장르 순으로 높게 나타났고,
공포, 스릴러, 멜로/로맨스 장르 순으로 낮게 나타났다.

관객수	min	max	count
장르			
액션 드라마	4457687	1048280 13414009	13
드라마	3279878	844103 14257115	81
범죄 액션	3065146	1128288 6879841	16
액션	2742868	989573 7817446	19
코미디 드라마	2645935	970373 12811206	18
범죄 드라마	2600992	986923 7072021	16
드라마 액션	2598920	857148 5507106	10
드라마 코미디	2538554	917782 7362467	11
코미디	2389978	856090 16264944	59
사극	2352942	958775 6246849	10
멜로/로맨스 코미디	2196988	848148 6619498	12
스릴러	2151121	870785 5604104	20



복합 장르로는 **코미디/드라마** 장르의 영화가 가장 많이 개봉했다.
평균 관객 수는 **액션/드라마**, **드라마**, **범죄/액션** 순으로 높게 나타났고,
스릴러, **로맨스/코미디**, **사극** 순으로 낮게 나타났다.

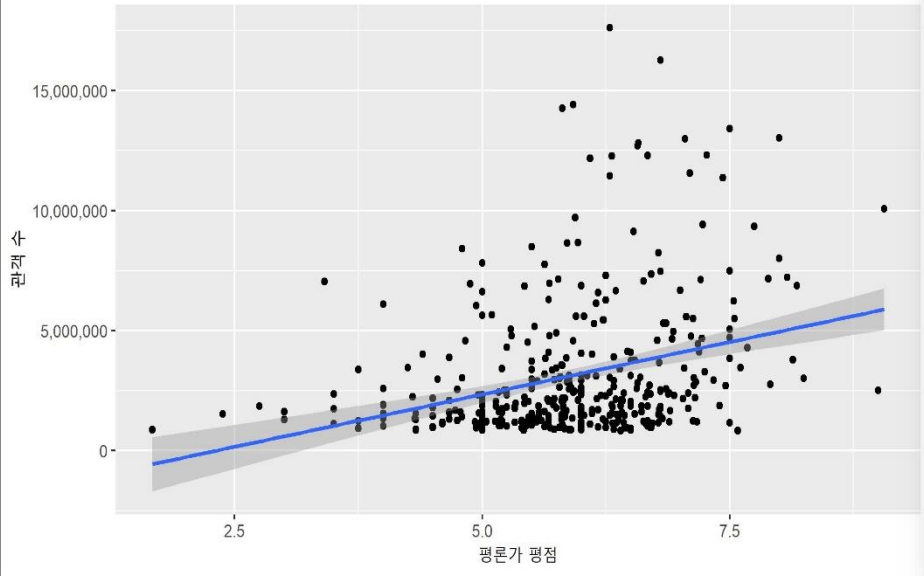
배급사 파워



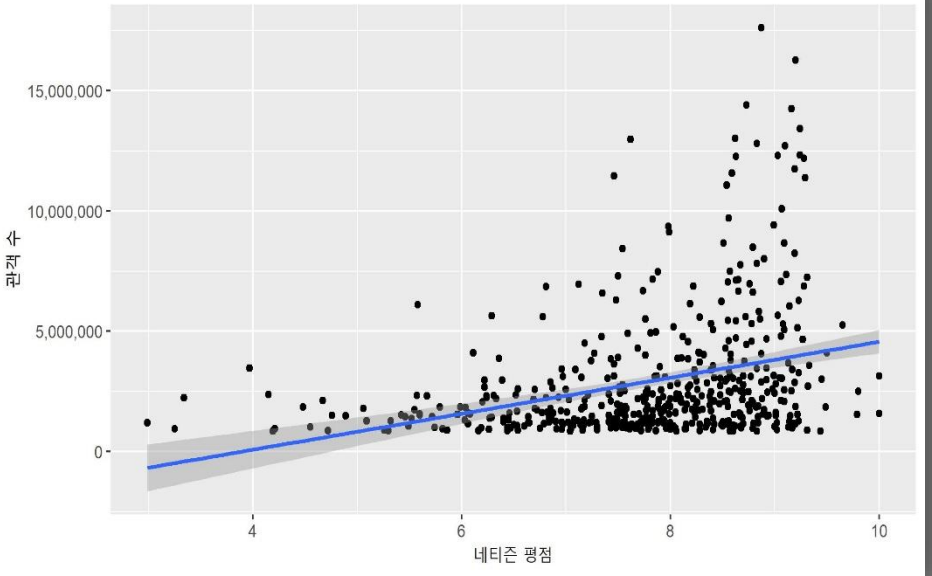
5번 이상의 영화를 배급한 배급사로 회귀모델을 생성
결정계수와 수정된 결정계수의 값이 **0.1~0.2** 사이의 값을 갖는 모델로
배급사 회귀모델은 스타의 회귀모델보다 더 **적은 설명력**을 가진다.

비평가, 수용자의 평가 - 네이버

평론가 평점별 관객 수



네티즌 평점별 관객 수

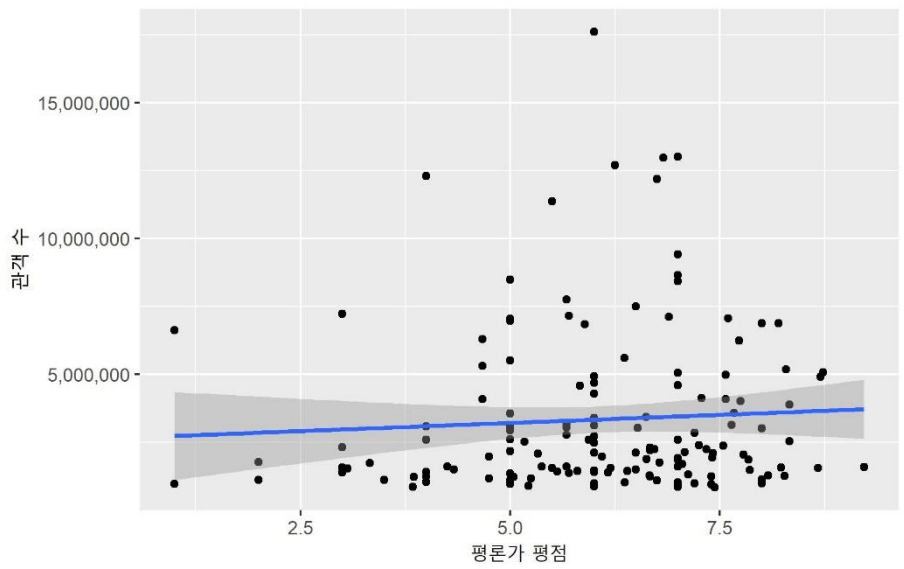


평점과 관객수의 산점도 그래프를 그려보니 **우상향**하는 그래프를 가진다.
평점과 관객수의 상관 관계는 **0.31** 정도의 **양의 상관관계**를 가진다.

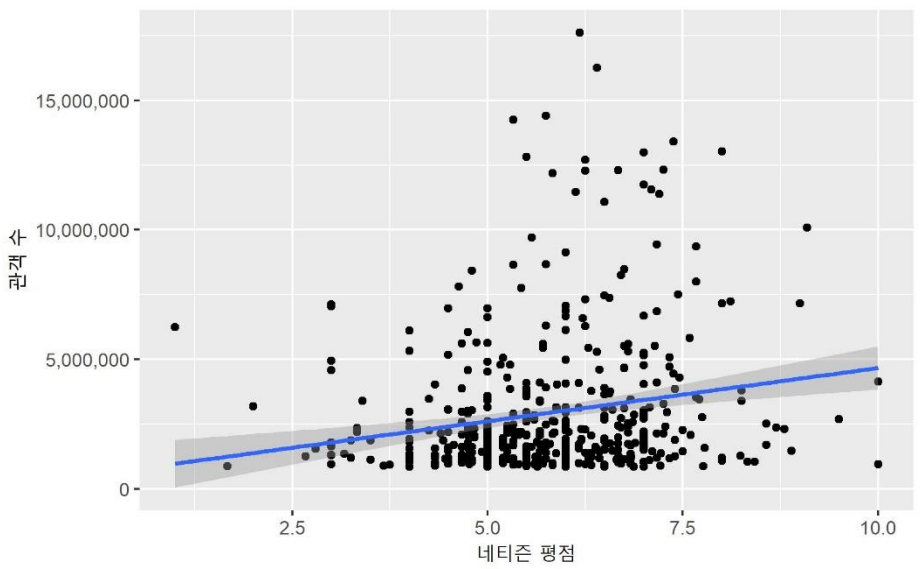
하지만, 평점의 경우 **평점을 올린 시기**가 중요한데 평균 평점을 가지고 하다보니 반영되지 않았다.

비평가, 수용자의 평가 - 씨네21

씨네21 평론가 평점별 관객 수



씨네21 네티즌 평점별 관객 수



씨네21 평점의 경우 네이버 영화 평점의 그래프보다 기울기가 **작은 것**을 볼 수 있다.
하지만, 평론가와 네티즌평점 모두 미세하나 **양의 상관관계**를 가진다.



1

평점의 경우 동명 영화 처리에 문제가 있는 것을 발견,
그 문제를 해결해야 더 정확한 결과가 나올 것

2

각 회귀모델의 설명력을 어떻게하면 더 높일 수 있을까?

3

영화 평점에 함께 올라오는 코멘트를 처리해 어떤 단어가
많이 나오면 평점과 흥행에 좋은 영향을 미치는지 분석



Thank you

Made by CHANUNE