

국내 개봉 영화 데이터 분석을 통한

“개봉 예정 영화 매출액 예상”

— 봉준호 감독이 만들면 영화의 매출액이 변화할까? —

01. 주제 / 02. 데이터

- 주제
 - 국내 개봉 영화 특징 기반, 개봉 영화들의 매출액 예측
- 사용 데이터
 - 영화관입장권통합전산망 KOBIS
 - 플랫폼 내 영화 정보 크롤링 내용

연도별 박스오피스

- 출처: 2020-10-21

- 출처: 영화진흥위원회 공공데이터 (<http://www.kobis.or.kr>)

▶ 집계조건 (조회기간: 2019년, 월간구분: 월간별, 단위: 천원)

순위	영화명	개봉일	개봉일	매출액 점유율	관객수	스크린수	상영횟수	대표국적	국적	배급사
1	연말특급	2019-01-23	139,851,845,516	7.5%	16,265,618	2,003	292,554	한국	한국	(주)세제이이엔엘
2	이벤트스 맥드게일	2019-04-24	122,182,694,160	6.6%	13,934,592	2,835	242,201	미국	미국	볼트디즈니윌리엄스과와 윌리엄스윌리엄스
3	가을동화 2	2019-11-21	111,396,240,725	6.0%	13,969,084	2,648	282,557	미국	미국	볼트디즈니윌리엄스과와 윌리엄스윌리엄스
4	결혼식	2019-05-20	106,955,130,359	5.9%	12,552,281	1,405	266,469	미국	미국	볼트디즈니윌리엄스과와 윌리엄스윌리엄스
5	가을동화	2019-05-30	85,883,963,645	4.6%	10,089,275	1,540	192,855	한국	한국	(주)세제이이엔엘
6	연식	2019-07-31	79,232,012,162	4.3%	9,426,011	1,662	202,223	한국	한국	(주)세제이이엔엘
7	스피리드인: 더 리틀	2019-07-02	69,010,005,100	3.7%	8,021,145	2,142	180,474	미국	미국	스피리드인: 더 리틀 (주)세제이이엔엘
8	배우	2019-12-19	52,905,799,770	2.9%	6,296,502	1,871	99,915	한국	한국	(주)세제이이엔엘 (주)세제이이엔엘
9	결혼식	2019-03-06	51,507,486,723	2.8%	5,802,810	2,100	196,362	미국	미국	볼트디즈니윌리엄스과와 윌리엄스윌리엄스

03. 전처리

결측치 처리, 중복값 병합

재개봉작, 장기상영작 제거

감독, 배급사

기준에 의해 가중치 부여

개봉월

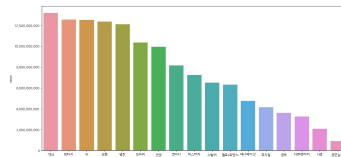
개봉일자 month 추출

계절, 장르 등

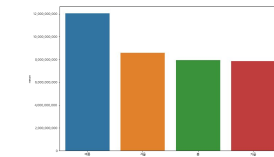
원 핫 인코딩 사용

04. EDA / 05. 모델

장르별 평균 매출액



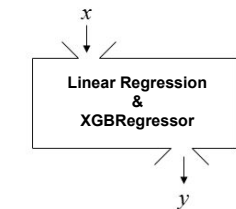
계절별 평균 매출액



스크린수 - 관객수/상영횟수 상관 관계



모델



06. 결론

- 핵심 요소 : 상영 횟수, 관객 수, 스크린 수
- 서브 요소 : 배급사, 대표 국적 등

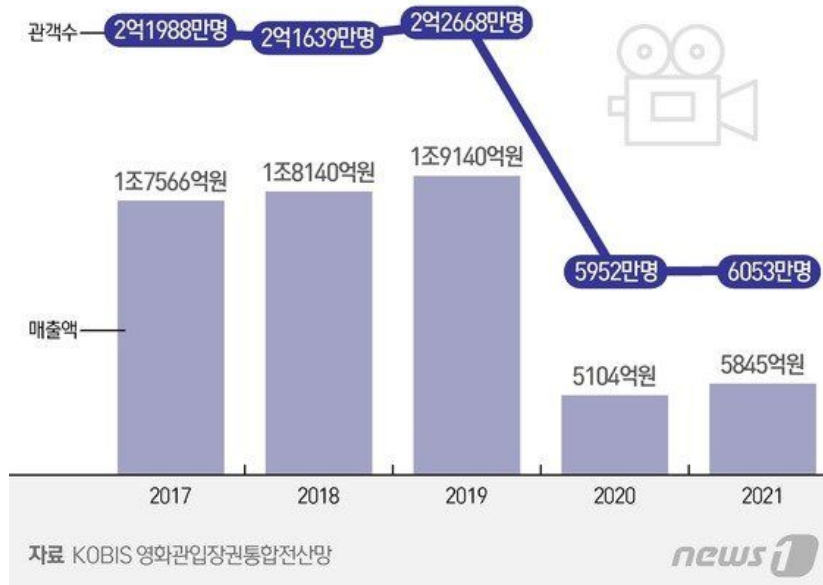
- 질문 : 봉준호 감독이 '자전차왕 엄복동'을 감독했다면 매출액에 변화가 있을까?
- 답변 : 매출액은 800% 이상 증가했을 것이다.

Contents

1. 주제
2. 데이터
3. 전처리
4. EDA
5. 모델
6. 결과
7. 결론

01 주제

전국 영화관 총 관객수 및 매출액



2019년 말부터 시작된 코로나 19 팬데믹으로 인해
얼어붙은 영화 산업은 **4분의 1** 수준으로 급감

국내 개봉 영화들의 **특징**들을 **분석**하고, 회귀 머신러닝을 통해
앞으로 개봉할 영화들의 매출액을 **예측**한다.

02 데이터



영화관입장권통합전산망(KOBIS:KOREA Box-office Information System)은 전국영화관 입장권 발권 정보를 실시간으로 집계처리하는 시스템(서비스 플랫폼)으로, 신속하고 다양한 박스오피스 정보와 각종 영화산업 통계정보를 제공

데이터 - 순위, 영화명, 개봉일, 매출액, 매출액 점유율, 관객수, 스크린수, 상영횟수, 대표국적, 국적, 배급사

연도별 박스오피스

- 조회일: 2022-12-21

- 출처: 영화진흥위원회 통합전산망 (<http://www.kobis.or.kr>)

▶ 검색조건 [조회기간: 2019 영화구분: 일반영화 국적: 전체 지역: 전체]

순위	영화명	개봉일	매출액	매출액 점유율	관객수	스크린수	상영횟수	대표국적	국적	배급사
1	극한직업	2019-01-23	139,651,845,516	7.5%	16,265,618	2,003	292,584	한국	한국	(주)씨제이이엔엠
2	어벤져스: 엔드게임	2019-04-24	122,182,694,160	6.6%	13,934,592	2,835	242,001	미국	미국	월트디즈니컴퍼니코리아 유한책임회사
3	겨울왕국 2	2019-11-21	111,596,248,720	6.0%	13,369,064	2,648	282,557	미국	미국	월트디즈니컴퍼니코리아 유한책임회사
4	알라딘	2019-05-23	106,955,138,359	5.8%	12,552,283	1,409	266,469	미국	미국	월트디즈니컴퍼니코리아 유한책임회사
5	기생충	2019-05-30	85,883,963,645	4.6%	10,085,275	1,948	192,855	한국	한국	(주)씨제이이엔엠
6	엑시트	2019-07-31	79,232,012,162	4.3%	9,426,011	1,660	202,223	한국	한국	(주)씨제이이엔엠
7	스파이더맨: 파 프롬 홈	2019-07-02	69,010,000,100	3.7%	8,021,145	2,142	180,474	미국	미국	소니픽처스엔터테인먼트코리아주식회사극장배
8	백두산	2019-12-19	52,905,789,770	2.9%	6,290,502	1,971	99,915	한국	한국	(주)씨제이이엔엠,(주)텍스터스튜디오
9	캡틴 마블	2019-03-06	51,507,488,723	2.8%	5,802,810	2,100	186,382	미국	미국	월트디즈니컴퍼니코리아 유한책임회사

02. 데이터

데이터 추가(연령등급, 장르, 러닝타임) - 크롤링 사용

영화 기생충 정보

통합 VIEW 이미지 지식iN 인플루언서 동영상 쇼핑 뉴스 어학사

dl.info.txt_4

622 x 138.5

Color ■ #000000

Font 13px -apple-system, BlinkMacSystemFon...

Padding 0px 0px 0px 15px

ACCESSIBILITY

Name

Role

Keyboard-focusable

무비클럽 | 포토 | 리뷰 | 명대사

개봉 2019.05.30.

등급 15세 관람가

장르 드라마

국가 대한민국

러닝타임 131분

배급 CJ ENM

```
<script type="text/javascript">var g_uau = false, </script>
<script type="text/javascript"></script>
<link rel="stylesheet" type="text/css" href="https://ssl.pstatic.net/sstatic/keypage/outside/scui/cs_common_module/css/cs_common_module_221230.css">
<link rel="stylesheet" type="text/css" href="https://ssl.pstatic.net/sstatic/keypage/outside/scui/cs_common_module/css/cs_common_color_221208.css">
<div class="sc_new cs_common_module case_emphasis color_10 _au_movie_content_wrap" data-dss-logarea="x0a">
  <div class="cm_top_wrap_sticky_custom_select_header"></div>
  <div class="cm_content_wrap">
    <div class="cm_content_area_cm_content_area_info">
      <div class="cm_info_box">
        <div class="detail_info">
          <a nocr onclick="return goOtherCR(this, 'a-nco_x0a*A.tabposter&r=18i=18000009D_000000566D2B&u=' + urlencode(this.href));" href="https://movie.naver.com/movie/bi/mi/basic.nhn?code=161967" class="thumb_i tem"></a>
          <dl class="info txt 4"> == $0
            <div class="info_group"></div>
            <div class="info_group"></div>
            <div class="info_group"></div>
            <div class="info_group"></div>
            <div class="info_group"></div>
            <div class="info_group"></div>
          </dl>
```

02. 데이터

데이터 추가(감독, 감독 참여작수) - 크롤링 사용

영화 기생충 출연진

통합 VIEW 이미지 지식인 인플루언서 동영상 쇼핑 뉴스 어학사전 지도 ...

기생충

영화 PARASITE | 2019

전체 | 기본정보 | 감독/출연 | 평점 | 무비클립 | 포토 | 리뷰 | 명대사 | 편성표 | ...

감독/배우

제작진

감독



봉준호
감독

연관 검색어

영화기생충출연진 과



코로나19 현황
확진 현황 및 백



신속항원검사
내 주변에서 신



당장 떠나고 싶
지금 가장 핫한



2023년 첫 해
지금 가장 인기

```
Elements Console Sources Network Performance Memory
<h3 class="title_numbering">감독</h3>
<div class="list_image_info_button_scroller" data-button-scroller-
  applied="true">
  <div class="type_scroll_scroller" style="touch-action: pan-y; us
    er-select: none; -webkit-user-drag: none;">
    <div>
      <ul class="cast_list">
        <li>
          <a nocr onclick="return goOtherCR(this, 'a=nco_x0a*A.tabca
            sting&r=1&i=1800009D_000000566D2B&u=' + urlencode(this.hre
              f));" href="?where=nexearch&sm=tab_etc&mra=bjky&pkid=1&os=
              95234&qvt=0&query=%E8%B4%B9%EC%A4%B0%ED%98%B8" class="inne
                r">
          <div class="item">
            <div class="thumb">...</div>
            <div class="title_box">
              <strong class="name" style="max-height: 4rem;" data-
                kgs-ellipsis data-ellipsis-applied="true">
                <span class="text" style="font-size: 1em;">봉준호</span>
              </strong>
              <div class="sub_text" style="max-height: 4rem;"
                data-kgs-ellipsis data-ellipsis-applied="true">...
              </div>
            </div>
          </li>
        </ul>
      </div>
    </div>
    <a nocr onclick="goOtherTCR(this, 'a=nco_x0a*A.tabcastingtableft&
      r=1&i=1800009D_000000566D2B');return false;" href="#" class="pre_
        visual_prev">이전</a>
    <a nocr onclick="goOtherTCR(this, 'a=nco_x0a*A.tabcastingtbrigr&
      r=1&i=1800009D_000000566D2B');return false;" href="#" class="next
```

03 전처리

결측치 처리, 중복값 병합

중복된 영화 = 한달 이상의 상영 기간, 재개봉

```
all_movies.isnull().sum()
```

✓ 0.1s

순위	0
영화명	0
개봉일	1
매출액	0
매출액 점유율	0
관객수	0
스크린수	0
상영횟수	0
대표국적	0
국적	0
배급사	2
연령등급	1
장르	0
러닝타임	1
감독	0
개봉월	1
계절	0

dtype: int64

all_movies[all_movies['영화명']=='황금 나침반']

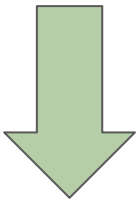
Python

	순위	영화명	개봉일	매출액	매출액 점유율	관객수	스크린수	상영횟수	대표국적	국적	배급사	연령등급	장르	러닝타임	감독	개봉월	계절	개봉년
14	15	황금 나침반	2007-12-18	18336681000	0.016	2906958	785	52708	미국	미국	롯데쇼핑롯데엔터테인먼트	전체 관람가	판타지	113.0	크리스 웨이츠	12.0	겨울	2007
211	62	황금 나침반	2007-12-18	18336681000	0.004	2906958	785	52708	미국	미국	롯데쇼핑롯데엔터테인먼트	전체 관람가	판타지	113.0	크리스 웨이츠	12.0	겨울	2007



그 외 데이터 처리

감독, 배급사



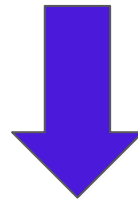
기준에 의해 가중치 부여

개봉월



개봉일자의 month 추출

계절, 장르, 연령등급, 대표국적

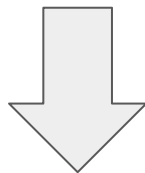


원 핫 인코딩 사용

이상치 확인

이상치 제거 전

	매출액	관객수	스크린수	상영횟수	러닝타임	개봉월	le_감독	le_배급사	감독_참여작수
매출액	1.000	0.991	0.687	0.913	0.326	0.050	0.381	0.043	0.032

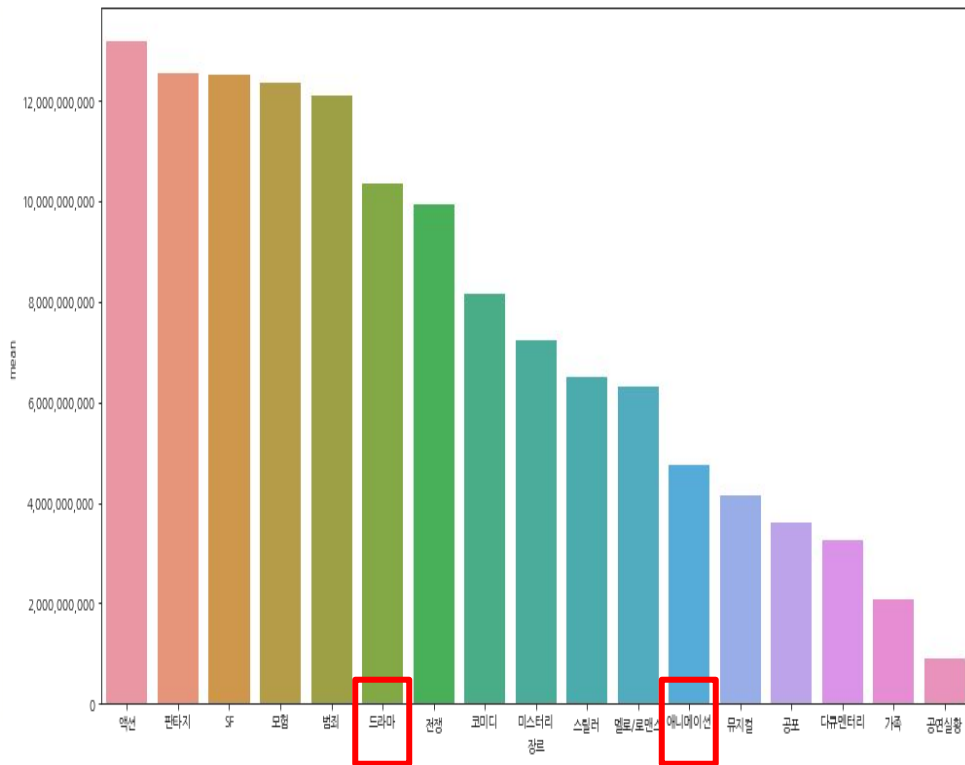
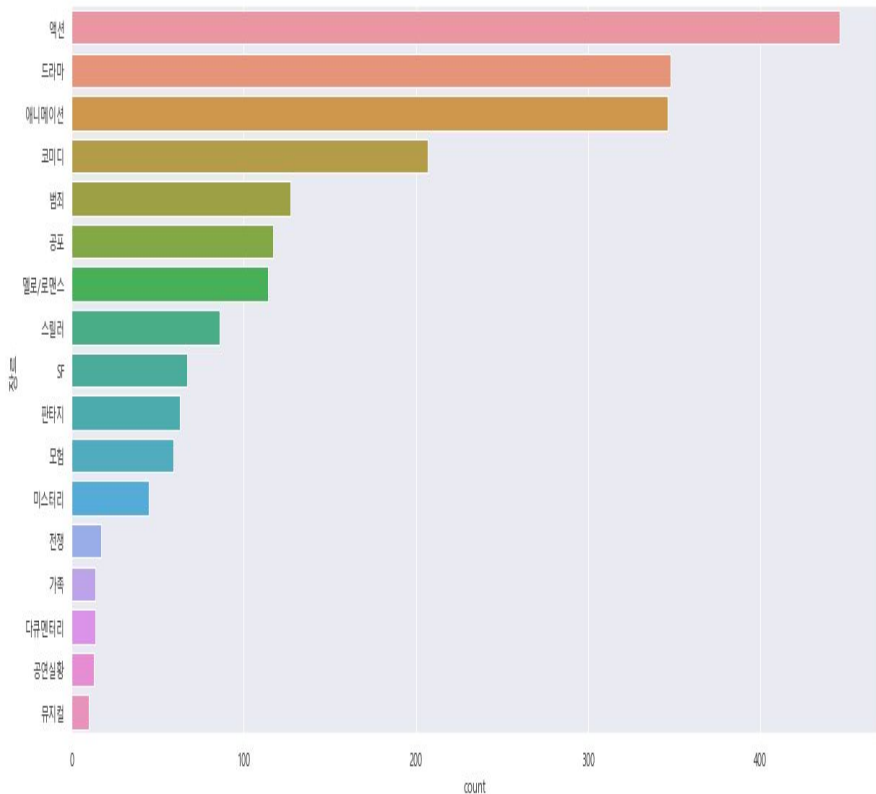


이상치 제거 후

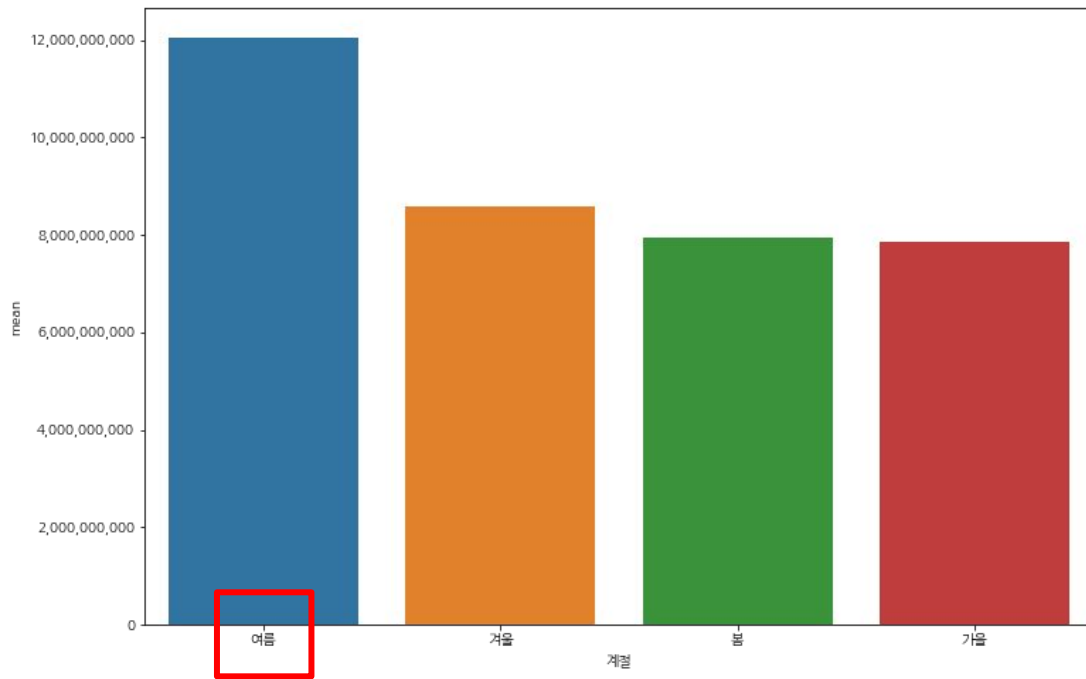
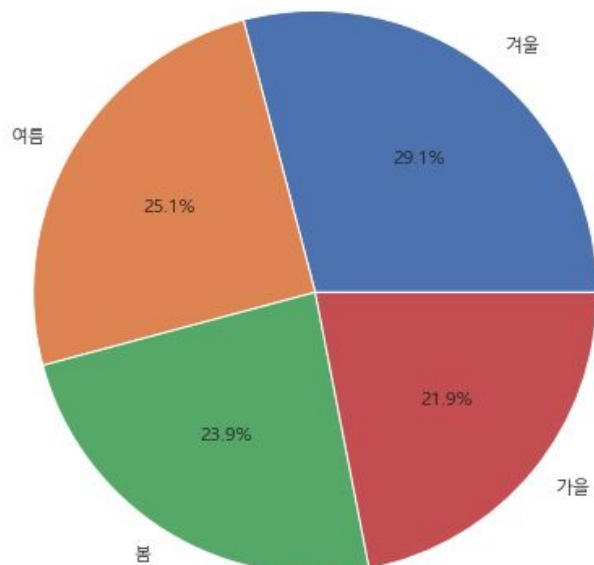
	매출액	관객수	스크린수	상영횟수	러닝타임	개봉월	le_감독	le_배급사	감독_참여작수
매출액	1.000	0.989	0.582	0.869	0.255	0.047	0.381	0.064	0.022

04 EDA

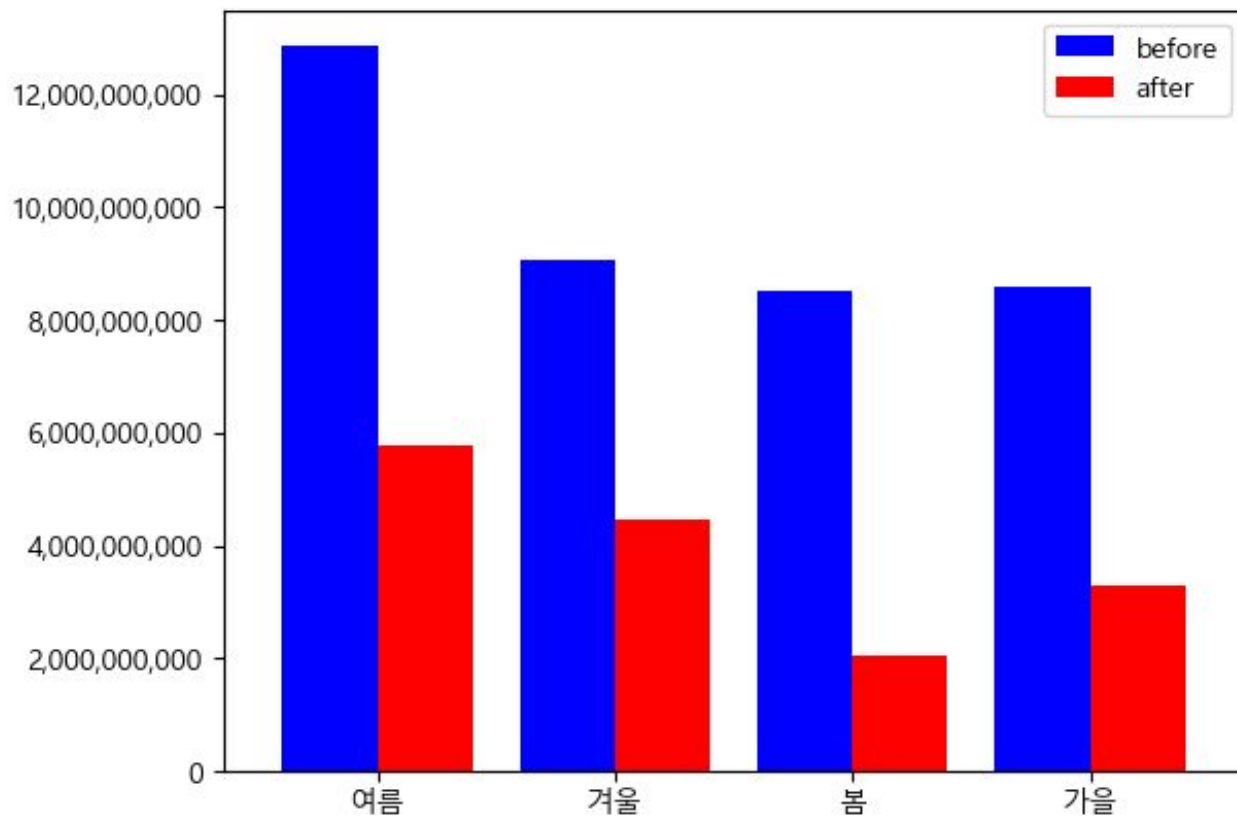
장르별 영화 개수 & 장르별 평균 매출액



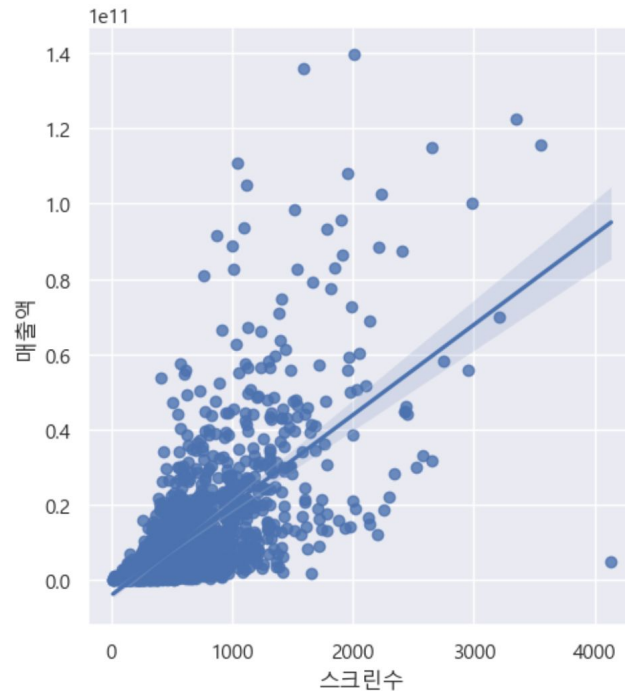
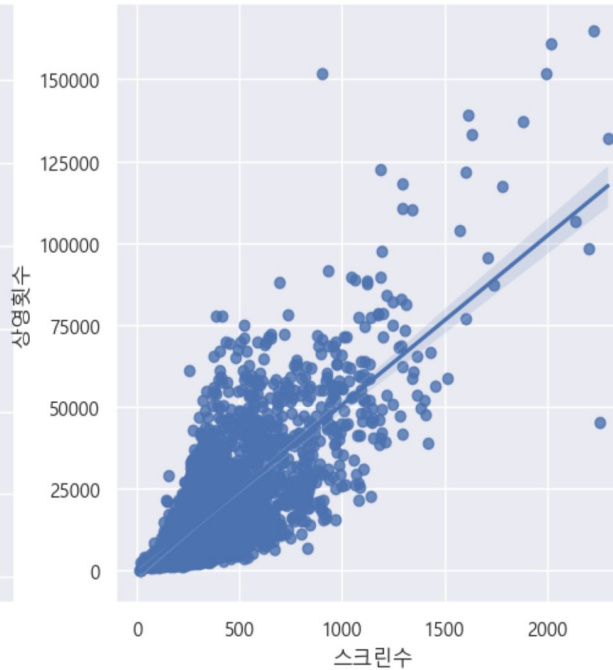
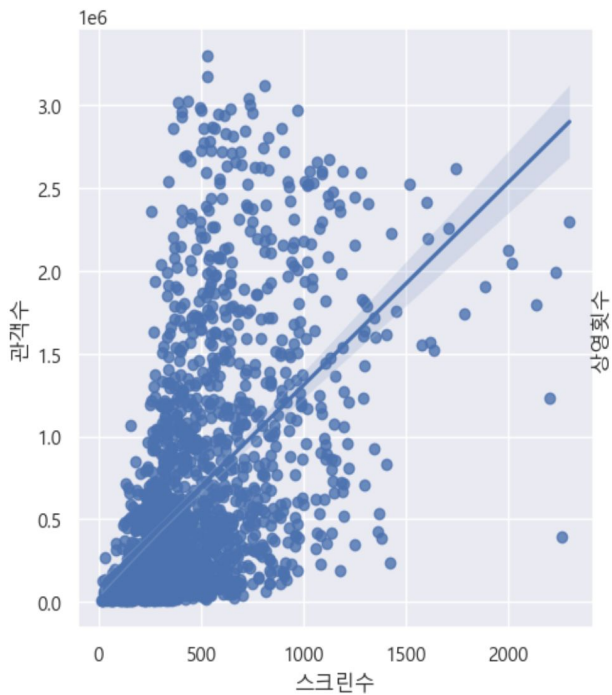
계절별 영화 개봉 개수 비율 & 계절별 평균 매출액



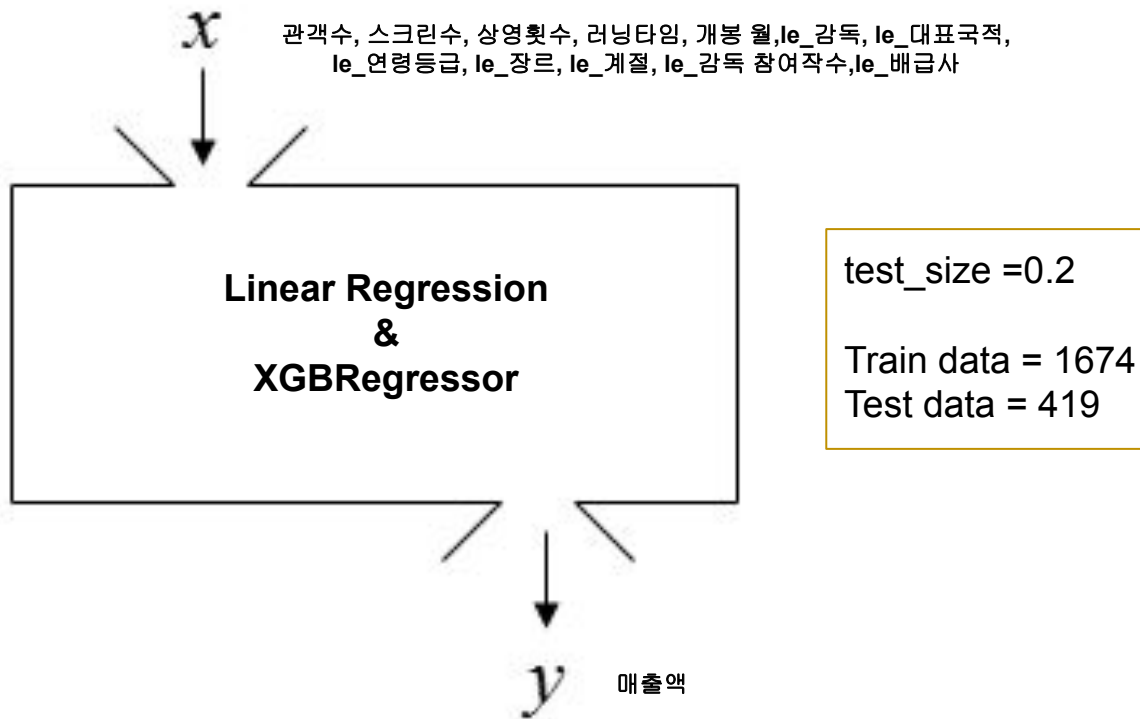
코로나 전후 매출액 평균



스크린수와 관객수, 스크린수와 상영횟수, 스크린수와 매출액의 산점도 & 선형회귀선



05 모델



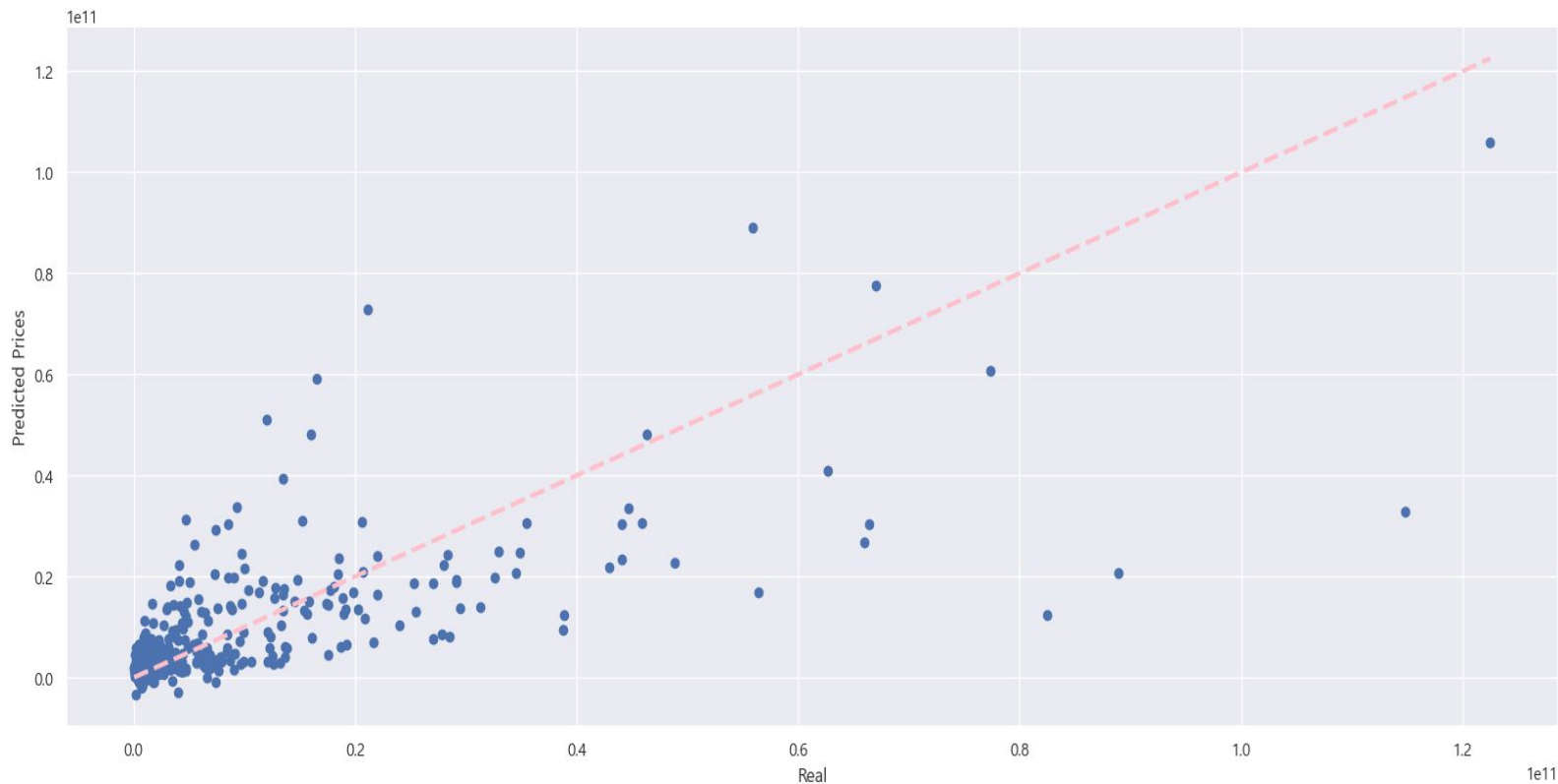
06 결과

	Features	스케일링 여부	알고리즘	RMSE
1	관객수, 스크린수, 상영횟수, 러닝타임, 개봉 월, le_감독, le_대표국적, le_연령등급, le_장르, le_계절, le_감독 참여작수	X	Linear Regression	1,135,966,885
2	관객수, 스크린수, 상영횟수, 러닝타임, 개봉 월, le_감독, le_대표국적, le_연령등급, le_장르, le_계절, le_감독 참여작수	Standard scaler	Linear Regression	1,135,966,885
...				
7	스크린수, 관객수 , 상영횟수, le_감독, 개봉 월, 러닝타임, le_대표국적, le_연령등급, le_장르, le_계절, le_감독 참여작수	X	XGBRegressor	9,329,308,381

06. 결과

머신러닝 예측 값 VS 실제 값

XGBRegressor



07 결론

상영 횟수

관객수

스크린 수



배급사

대표국적

연령 등급

...

봉준호 감독이 ‘자전차왕 엄복동’을 감독했다면 매출액에 변화가 있을까?



자전차왕 엄복동

Race to Freedom : Um Bok Dong, 2018

개봉 2019.02.27

평점 ★4.6

장르 드라마

누적관객 172,212명

국가 한국

수상내역 39회 황금촬영상, 2019

등급 12세 이상 관람가

러닝타임 116분

영화정보 !

★☆☆☆☆ 1

충격적인건 기생충보다 제작비 더 들었음

2020.12.13.22:01

 13 2

★★★★★ 0

최악 그 이상 이하도 아님, 배우들의 과도한 오버 액션, 끝내주는 CG, 뜬금없는 독백운동, 더 뜬금 없는 비와 강소라의
춤, 자연계 시합때의 잔인한 긴박감, 처음 시합 10분만에 밀친 드러난 스토리 구성 등은 이쯤에서 엄청난 대학을
탄생시킬, 시대가 어찌때면디 3월 1일날 개개생가 뜬금없이 독백운동 넣고 너무나 전형적인 이비스에다가 신작까지
짬뽕 시키면 잘 될지 알았나보네, 얼마나 관객들을 무시한건지... 마지막으로 엄복동의 미화 등등 최소한 대학생들의
영화에도 사전 조사는 하고 영화를 만드는데 150억이라는 돈은 어디에 썼는지... 그냥 이비스구 인맥으로 유명한 조연들만
쓰면 될줄 알았으니... 정말 하신 하고 관객을 철저히 무시한 영화다.

7.02:11

 12 2

영화명	개봉일	매출액	매출액 점유율	관객수	스크린 수	상영횟수	대표국적	국적	배급사	연령등급	장르	러닝타임	감독	개봉월	계절
자전차왕 엄복동	2019-02-27	1304861670	0.001	172212	624	11245	한국	한국	(주)셀트리온엔터테인먼트	12세 관람가	드라마	116.0	김유성	2.0	겨울

모델명	예상 매출액	봉준호	액션	여름(7월)
XGB Regressor	5,195,403,000 (298% ▲)	13,027,546,000 (898% ▲)	5,195,403,000 (298% ▲)	6,756,180,000 (417% ▲)

감사합니다