

# 데이터 분석 프로젝트 : 영화산업 데이터 분석

지능형 소프트웨어  
기말고사 대체 과제

협성대학교 컴퓨터공학과

20170677 오윙택

# 목차

- • 선정 동기
- • 데이터 수집 방법
- • 데이터 분석 목표
- • 데이터 분석 방법

# 선정 동기

- • 최근에 영화를 한 편 관람하였는데 어떤 영화는 기대받지 못했지만 실제로 개봉했을 때 예상외의 성적을 거두어 관람객의 평가가 좋았고, 어떤 영화는 좋은 성적을 기대했지만 예상했던 것 보다 좋지 못한 결과를 보여주고 있습니다.
- 이에 영화의 예매율과 관객 리뷰가 실제 영화 매출에 미치는 영향을 알아보고 싶어 주제를 선정하게 되었습니다,

# 데이터 수집 방법

- • 영화 상영회사 3사(CJ CGV, 롯데시네마, 메가박스)의 예매율, 평점/리뷰, 성별과 연령 분포 데이터를 수집한다.
- • 파이썬을 기반으로 Selenium과 BeautifulSoup 라이브러리를 통해 동적웹 크롤링을 통해 데이터를 수집한다.

# 분석 목표

- • 극장 위치, 상영 날짜에 따라 영화 예매율과 실제 관람객 수의 차이에 영향이 있는지를 분석한다.
- • 관람객의 성별, 나이 등에 따라 선호하는 영화 장르를 비교하고 이 것이 영화의 예매에 미치는 영향을 분석한다.
- • 관람객이 영화를 관람한 뒤 작성한 리뷰가 추후 예매율에 미치는 영향을 분석한다.

# 분석 방법

- • 리뷰데이터는 감정 분석과 빈도 분석을 통해 자연어 처리를 먼저 수행한 뒤 긍정 리뷰와 부정 리뷰로 다시 분류한다.
- • 리뷰를 제외한 다른 데이터들은 우선 군집화를 통해 특징을 추출해낸다.
- 추출된 데이터들이 서로에게 미치는 영향을 파악하기 위해 회귀분석을 수행한다.

## 결론 도출

- • 분석을 수행한 뒤 matplotlib을 사용하여 분석한 데이터들의 상관계수를 시각화 하여 영화 예매율에 영향을 미치는지 확인한다.
- 만약 회귀분석 결과로 높은 상관계수( $-1 \sim -0.5$  or  $0.5 \sim 1$ )가 도출된다면 실제로 리뷰가 영화 예매율에 유의미한 영향을 미친 것이니 유의미한 데이터 분석임을 알 수 있다.
- 반대의 경우( $-0.5 \sim 0.5$ ) 영화 리뷰가 예매율에 영향을 미치지 않은 것이니 가설을 폐기한다.



# DataCrawl.py

영화 3사의 데이터를 크롤링하는 파일

3개의 리스트(CGTV, lottecinema, megabox)를 생성하여  
각 리스트의 요소 안에

[[영화 제목, 평점, 전체 예매율, 장르], [영화 제목, 리뷰들 (30개)]]를  
저장하고 csv로 다시 저장합니다.



# 결과화면

CGV

C:\Users\dbdxo\anaconda3\envs\chatbot\python.exe C:\Workspace\Python\FinalTest\DataCrawl.py

<http://www.cgvl.co.kr/movies>

From <http://www.cgvl.co.kr/movies>, got 10 midx values.

['87045', '87055', '87059', '87035', '87063', '87062', '87061', '87065', '86883', '87064']

{'title': '범죄도시3', 'reservation\_rate': '16.9%', 'genre': None, 'gender\_distribution': {'남': '47%', '여': '53%'}, 'age\_distribution': {'10대': '5.2', '20대': '28.5', '30대': '31.3', '40대': '21.4', '50대': '13.6'}}

{'title': '트랜스포머-비스트의 서막', 'reservation\_rate': '3.4%', 'genre': None, 'gender\_distribution': {'남': '55.3%', '여': '44.7%'}, 'age\_distribution': {'10대': '2.2', '20대': '17.1', '30대': '27.2', '40대': '36.3', '50대': '17.2'}}

{'title': '탑건', 'reservation\_rate': '1.2%', 'genre': None, 'gender\_distribution': {'남': '38.9%', '여': '61.1%'}, 'age\_distribution': {'10대': '3.9', '20대': '34.8', '30대': '31.7', '40대': '18.9', '50대': '10.7'}}

{'title': '남은 인생 10년', 'reservation\_rate': '1.1%', 'genre': None, 'gender\_distribution': {'남': '39.3%', '여': '60.7%'}, 'age\_distribution': {'10대': '20.3', '20대': '42.4', '30대': '15.4', '40대': '11.9', '50대': '10'}}

{'title': '매그놀리아', 'reservation\_rate': '0.8%', 'genre': None, 'gender\_distribution': {'남': '48.8%', '여': '51.2%'}, 'age\_distribution': {'10대': '0.2', '20대': '39.3', '30대': '32.1', '40대': '15.7', '50대': '12.7'}}

{'title': '뱀파이어와의 인터뷰', 'reservation\_rate': '0.8%', 'genre': None, 'gender\_distribution': {'남': '27.6%', '여': '72.4%'}, 'age\_distribution': {'10대': '0.1', '20대': '29.6', '30대': '40.2', '40대': '20.4', '50대': '9.8'}}

{'title': '어 퓨 굿 맨', 'reservation\_rate': '0.7%', 'genre': None, 'gender\_distribution': {'남': '35.5%', '여': '64.5%'}, 'age\_distribution': {'10대': '2.6', '20대': '31.9', '30대': '33.8', '40대': '20.2', '50대': '11.6'}}

{'title': '바닐라 스카이', 'reservation\_rate': '0.7%', 'genre': None, 'gender\_distribution': {'남': '36%', '여': '64%'}, 'age\_distribution': {'10대': '0.4', '20대': '35.8', '30대': '35.7', '40대': '16.5', '50대': '11.6'}}

{'title': '가디언즈 오브 갤럭시-Volume 3', 'reservation\_rate': '0.6%', 'genre': None, 'gender\_distribution': {'남': '49.6%', '여': '50.5%'}, 'age\_distribution': {'10대': '2.7', '20대': '27.8', '30대': '37.4', '40대': '21.8', '50대': '10.3'}}

{'title': '아이즈 와이드 섯', 'reservation\_rate': '0.4%', 'genre': None, 'gender\_distribution': {'남': '48.3%', '여': '51.7%'}, 'age\_distribution': {'10대': '0.2', '20대': '37.3', '30대': '33.2', '40대': '16.4', '50대': '12.9'}}

[[{'title': '범죄도시3', 'reviews': ['마동석표 액션과 개그가 3편까지 오면서 조금씩 식상해지는 듯 하다. 악역의 카리스마도 전편들에 비해 조금 약한 듯. 그래도 마동석표 액션&개그가 아작은 먹히는 편.', '다른 사람들보다 조금 일찍 보게 된 범죄도시 3편. 이번에도 초

<https://www.lottecinema.co.kr/NLCHS/Movie/List?flag=1>

# 결과화면

LotteCinema

<https://www.lottecinema.co.kr/NLCHS/Movie/List?flag=1>

로딩로딩

From <https://www.lottecinema.co.kr/NLCHS/Movie/List?flag=1>, got 10 midx values.

['19803', '19850', '19764', '19776', '19874', '19786', '19898', '19481', '19742', '19814']

{'title': '범죄도시3', 'grade\_point': '9.2', 'reservation\_rate': '32.6%', 'genre': '범죄, 액션', 'gender\_distribution': {'남성': '48.8%', '여성': '51.2%'}, 'age\_distribution': {'10대': '3.5%', '20대': '31.8%', '30대': '38.3%',

{'title': '트랜스포머: 비스트의 서막', 'grade\_point': '8.4', 'reservation\_rate': '5.0%', 'genre': '액션, SF', 'gender\_distribution': {'남성': '56.9%', '여성': '43.1%'}, 'age\_distribution': {'10대': '3.2%', '20대': '28.5%', '30대': '31.8%',

{'title': '포켓 몬스터 DP: 아르세우스 초극의 시공으로', 'grade\_point': '9.3', 'reservation\_rate': '1.1%', 'genre': '애니메이션, 어드벤처, 판타지', 'gender\_distribution': {'남성': '36.3%', '여성': '63.8%'}, 'age\_distribution': {'10대': '3.2%', '20대': '28.5%', '30대': '31.8%',

{'title': '가디언즈 오브 갤럭시: Volume 3', 'grade\_point': '9.6', 'reservation\_rate': '1.0%', 'genre': '액션', 'gender\_distribution': {'남성': '49.1%', '여성': '50.9%'}, 'age\_distribution': {'10대': '3.3%', '20대': '39.9%', '30대': '31.8%',

{'title': '인어공주', 'grade\_point': '8.0', 'reservation\_rate': '0.4%', 'genre': '뮤지컬, 가족, 판타지', 'gender\_distribution': {'남성': '30.8%', '여성': '69.2%'}, 'age\_distribution': {'10대': '4.9%', '20대': '50.0%', '30대': '31.8%',

{'title': '분노의 질주: 라이드 오어 다이', 'grade\_point': '9.1', 'reservation\_rate': '0.3%', 'genre': '액션', 'gender\_distribution': {'남성': '47.5%', '여성': '52.5%'}, 'age\_distribution': {'10대': '2.4%', '20대': '40.7%', '30대': '31.8%',

{'title': '부기맨', 'grade\_point': '8.9', 'reservation\_rate': '0.1%', 'genre': '공포(호러)', 'gender\_distribution': {'남성': '69.1%', '여성': '31.0%'}, 'age\_distribution': {'10대': '13.8%', '20대': '44.8%', '30대': '27.6%', '40대': '13.6%',

{'title': '스즈메의 문단속', 'grade\_point': '9.2', 'reservation\_rate': '0.0%', 'genre': '애니메이션', 'gender\_distribution': {'남성': '56.1%', '여성': '43.9%'}, 'age\_distribution': {'10대': '2.9%', '20대': '47.1%', '30대': '34.3%', '40대': '15.7%',

{'title': '슬픔의 삼각형', 'grade\_point': '8.6', 'reservation\_rate': '0.0%', 'genre': '코미디, 드라마', 'gender\_distribution': {'남성': '66.7%', '여성': '33.3%'}, 'age\_distribution': {'10대': '0%', '20대': '66.7%', '30대': '0%', '40대': '33.3%',

{'title': '이웃과 바다에 달다', 'grade\_point': '7.5', 'reservation\_rate': '0.0%', 'genre': '드라마', 'gender\_distribution': {'남성': '83.3%', '여성': '16.7%'}, 'age\_distribution': {'10대': '0%', '20대': '25.0%', '30대': '75.0%',

No more 'Show more' buttons.

# 결과화면

Megabox

<https://megabox.co.kr/movie>

로딩로딩로딩

```
[ '23041900', '23034900', '23035600', '23028800', '23030000', '23037300', '23018400', '22103500', '23028100', '23038800', '23030200', '23037800', '23000600', '23040100', '22075200', '23029200', '23040200', '23029900', '23012200' ]
{ 'title': '플래시', 'grade_point': '9.1', 'reservation_rate': '29.2%', 'genre': '액션' }
{ 'title': '엘리멘탈', 'grade_point': '0', 'reservation_rate': '22.4%', 'genre': '피터 손' }
{ 'title': '범죄도시3', 'grade_point': '9', 'reservation_rate': '18.4%', 'genre': '이상용' }
{ 'title': '스파이더맨: 어크로스 더 유니버스', 'grade_point': '0', 'reservation_rate': '9%', 'genre': '저스틴 톰슨, 조아킴 도스 산토스, 캠프 파워' }
{ 'title': '귀공자', 'grade_point': '8.4', 'reservation_rate': '7.4%', 'genre': '박훈정' }
{ 'title': '트랜스포머: 비스트의 서막', 'grade_point': '8.2', 'reservation_rate': '3.2%', 'genre': '스티븐 케이플 주니어' }
{ 'title': '가디언즈 오브 갤럭시: Volume 3', 'grade_point': '9.5', 'reservation_rate': '1.3%', 'genre': '제임스 건' }
{ 'title': '더 퍼스트 슬램덩크', 'grade_point': '9.5', 'reservation_rate': '0.9%', 'genre': '이노우에 다케히코' }
{ 'title': '슬픔의 삼각형', 'grade_point': '8.3', 'reservation_rate': '0.7%', 'genre': '루벤 외스틀룬드' }
{ 'title': '2023 베를린 필하모닉 발트뷔네 콘서트', 'grade_point': '0', 'reservation_rate': '0.7%', 'genre': '공연' }
{ 'title': '극장판 포켓몬스터DP : 아르세우스 초극의 시공으로', 'grade_point': '9.3', 'reservation_rate': '0.5%', 'genre': '애니메이션' }
{ 'title': '[존 크랑코's 발레 3부작] 오네긴', 'grade_point': '9.9', 'reservation_rate': '0.4%', 'genre': '공연' }
{ 'title': '스즈메의 문단속', 'grade_point': '9.1', 'reservation_rate': '0.4%', 'genre': '신카이 마코토' }
{ 'title': '자우림, 더 원더랜드', 'grade_point': '9.2', 'reservation_rate': '0.3%', 'genre': '김지환, 허민' }
{ 'title': '슈퍼 마리오 브라더스', 'grade_point': '9.1', 'reservation_rate': '0.3%', 'genre': '애니메이션, 어드벤처, 코미디' }
{ 'title': '드림캐리스', 'grade_point': '8.1', 'reservation_rate': '0.3%', 'genre': '가성문' }
{ 'title': '[시네도스톤 - 유럽 아트 투어] 네덜란드', 'grade_point': '9.7', 'reservation_rate': '0.3%', 'genre': '12세이상관람가' }
{ 'title': '분노의 질주: 라이드 오어 다이', 'grade_point': '8.9', 'reservation_rate': '0.3%', 'genre': '루이스 리터리어' }
{ 'title': '인어공주', 'grade_point': '7.6', 'reservation_rate': '0.2%', 'genre': '롭 마샬' }
{ 'title': '하나님의 마음', 'grade_point': '0', 'reservation_rate': '0.2%', 'genre': '데이비드 헬링' }
[[ { 'title': '범죄도시3', 'reviews': [ '마동석표 액션과 개그가 3편까지 오면서 조금씩 식상해지는 듯 하다. 악역의 카리스마도 전편들에 비해 조금 약한 듯. 그래도 마동석표 액션&개그가 아직은 먹히는 편.', '다른 사람들보다 조금 일찍 보게 된 범죄도시 3편. 이번에도 초대' ] }
```

Topooback (most recent call last):