

제작사 중심의

영화 흥행 요인 분석

박창현

00. 목차

01
요구사항 시나리오

02
데이터 확인

03
데이터 전처리

04
데이터 분석 &
인사이트 도출

05
분석 결과 &
결론 제시

06
아쉬운점

요구사항 시나리오

창현 엔터테인먼트는 최근 몇 년간 다양한 장르의 영화를 제작했지만, 기대만큼의 상업적 성공을 거두지 못했다. 이를 해결하기 위해 **영화 흥행에 영향을 미치는 주요 요인을 분석**하여 향후 제작 전략에 반영하고자 한다. 아래는 요구사항 목록이다.

1. 흥행 요소 분석

- 1) 장르, 예산, 개봉 시기 등을 사용하여 흥행에 유의미한 영향을 미치는 요소들 파악.
- 2) 경쟁 제작사들의 장르별, 감독별 현황 파악.

2. 트렌드 분석

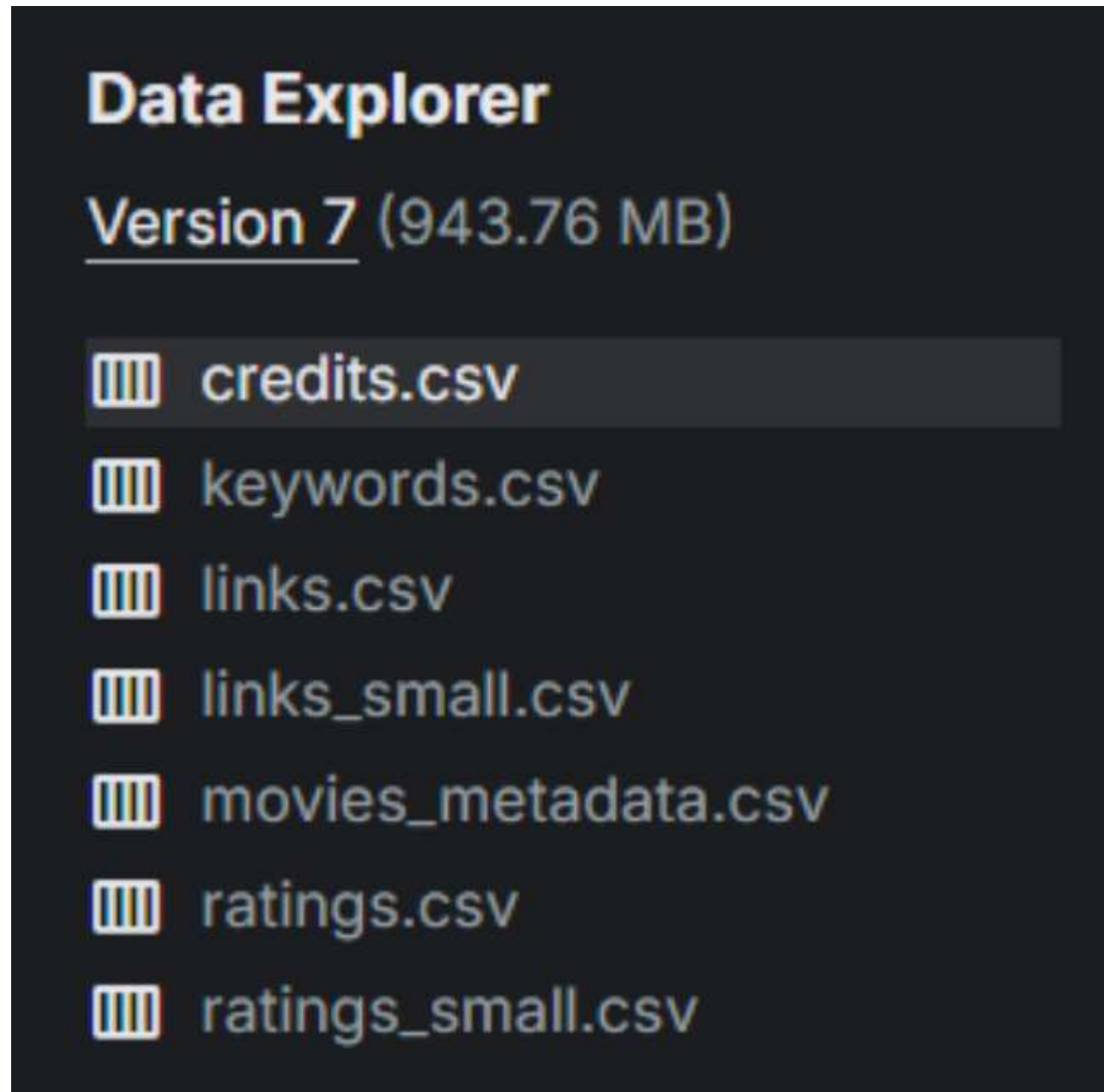
최근 영화 시장에서 흥행했던 장르, 테마, 출연진 등에 대한 트렌드 확인

3. ROI 분석

- 1) 각 영화의 투자 대비 수익률(ROI)계산, ROI가 높은 요소들 파악
- 2) 예산, 개봉 시기, 장르 등 여러 요소들로 인한 영화 흥행 실패 가능성 분석

02. 데이터 확인

The Movie Dataset (<https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>)



credits.csv: 모든 영화에 대한 출연진 및 제작진 정보를 포함하고 있다. 문자열화된 JSON 객체 형식으로 제공.

keywords.csv: 영화의 줄거리 키워드 파일. 문자열화된 JSON 객체 형식으로 제공.

links.csv: 모든 영화의 TMDB 및 IMDB ID를 포함하고 있는 파일.

links_small.csv: 전체 데이터셋의 소규모 하위 집합에 해당하는 9,000편의 영화에 대한 TMDB 및 IMDB ID를 포함하고 있는 파일.

movies_metadata.csv: 영화 메타데이터 파일. 45,000편의 영화에 대한 정보를 포함하고 있다. 포스터, 배경 이미지, 예산, 수익, 개봉일, 언어, 제작 국가 및 제작사 등이 포함된다.

ratings_small.csv: 9,000편의 영화에 대해 700명의 사용자가 남긴 100,000개의 평점이 포함된 소규모 하위 집합.

데이터 확인 및 전처리 계획

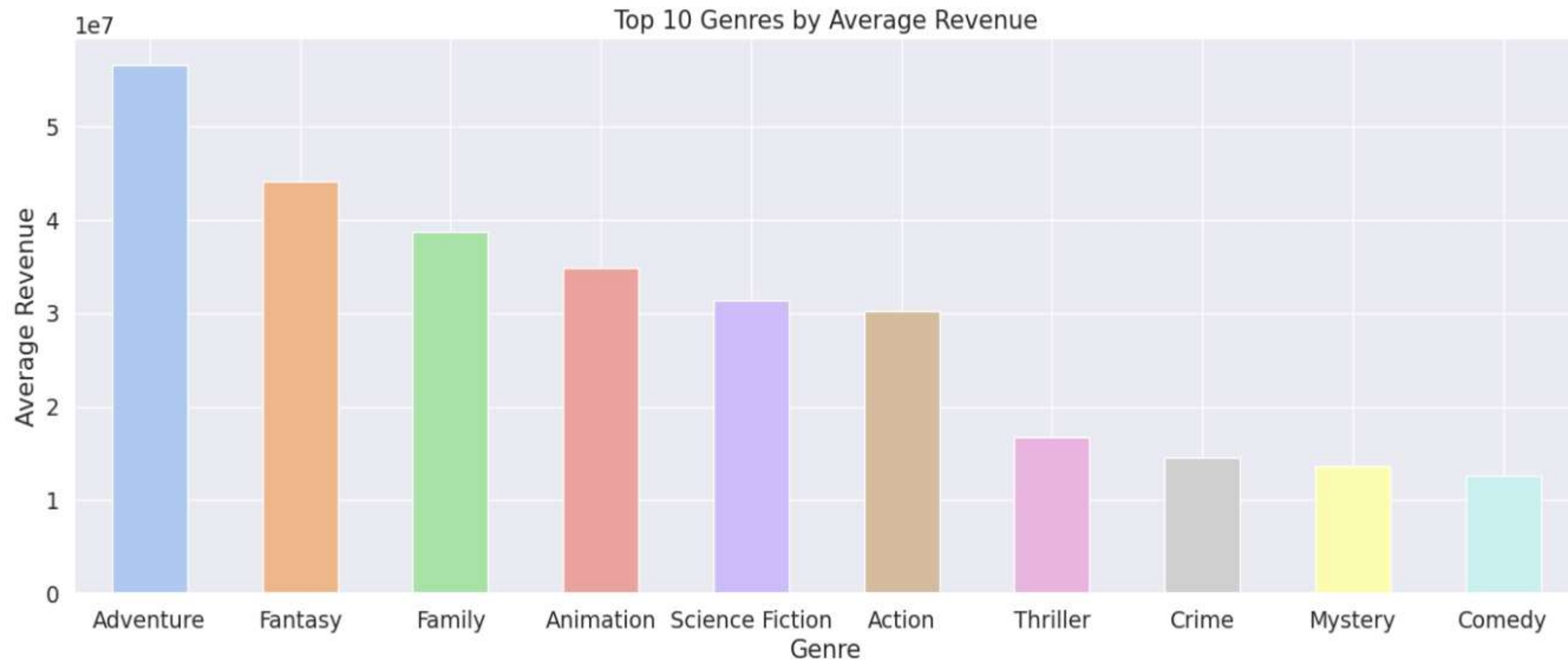
1. 전처리 목표

- movies_metadata.csv: 예산, 수익, 장르, 개봉 날짜 등 주요 정보를 포함하는 파일로, 흥행 요소 분석, 트렌드 분석, ROI 분석에 필요한 데이터를 정리.
- credits.csv: 배우와 감독 정보를 포함하는 파일로, 주요 인물들의 성과 분석에 사용될 데이터를 정리.
- keywords.csv: 영화와 관련된 키워드를 포함하는 파일로, 특정 테마와 트렌드 분석에 사용될 데이터를 정리.

2. 전처리 계획, 과정

데이터 파일	Column 명	전처리 목표
movies_metadata.csv	12개의 columns 제거	불필요한 columns 제거
	budget, revenue	float64형으로 변환 / 결측치를 0으로 대체
	genres	JSON 형식의 문자열을 리스트 형식으로 변환
	release_date	개봉 연도만 추출하여 새로운 컬럼 release_year 생성
credits.csv	crew	감독 정보 추출하여 새로운 컬럼 director 생성
	cast	주요 배우 상위 5명을 추출하여 리스트로 저장
keywords.csv	id, keywords	JSON 형식의 문자열을 리스트 형식으로 변환 id, keywords 제외 나머지 컬럼 삭제

1) 흥행 요인 분석 - 상위 10개의 인기있는 장르 확인



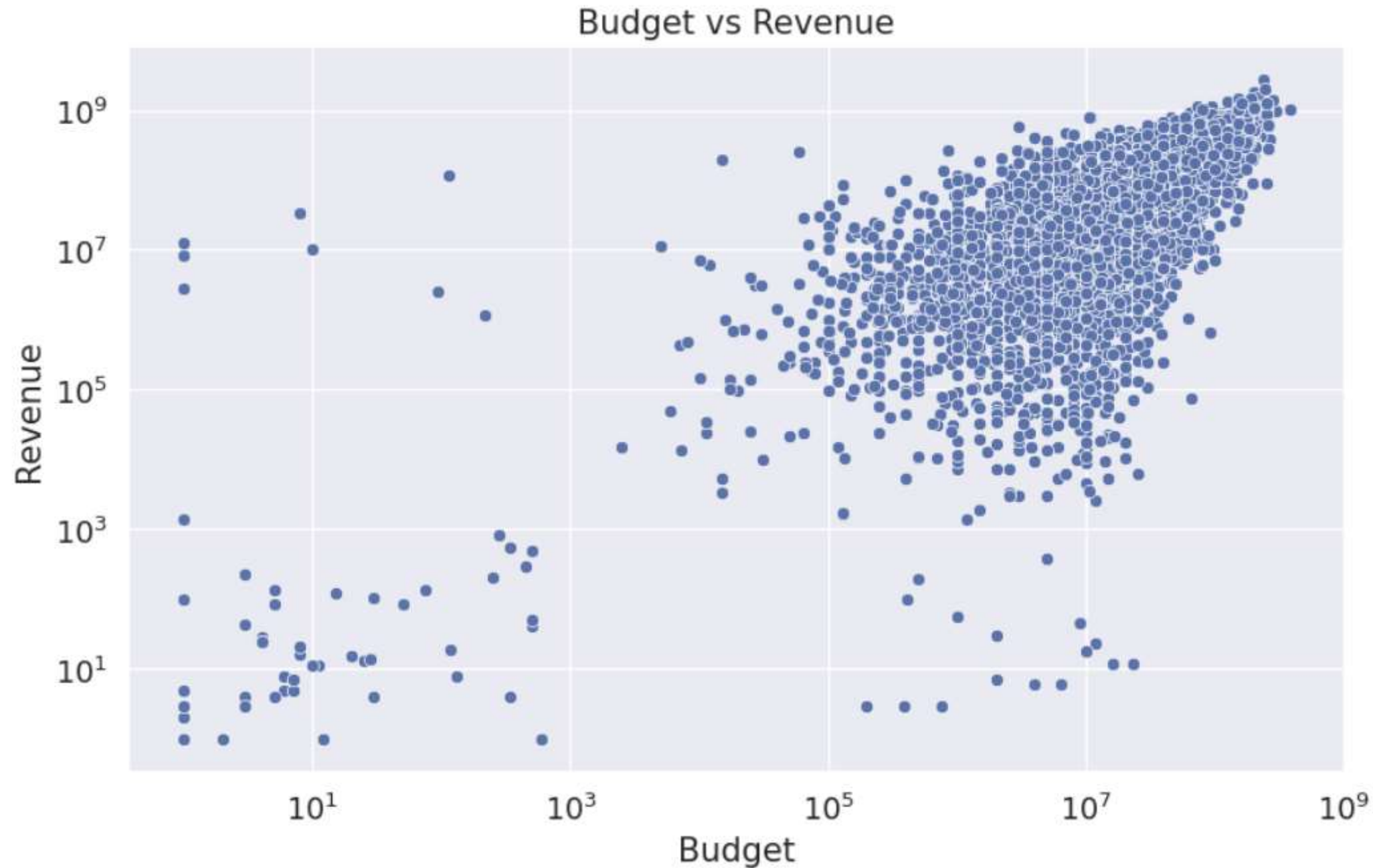
수익이 높은 상위 10개의 인기 장르를 확인해 본 결과
Adventure, Fantasy, Family, Animation 등의 장르가
대체적으로 인기 있는 것을 확인해볼 수 있다.



온 가족이 볼 수 있는 영화가 대체적으로 수익이 높을 수 있다.

04. 데이터 분석

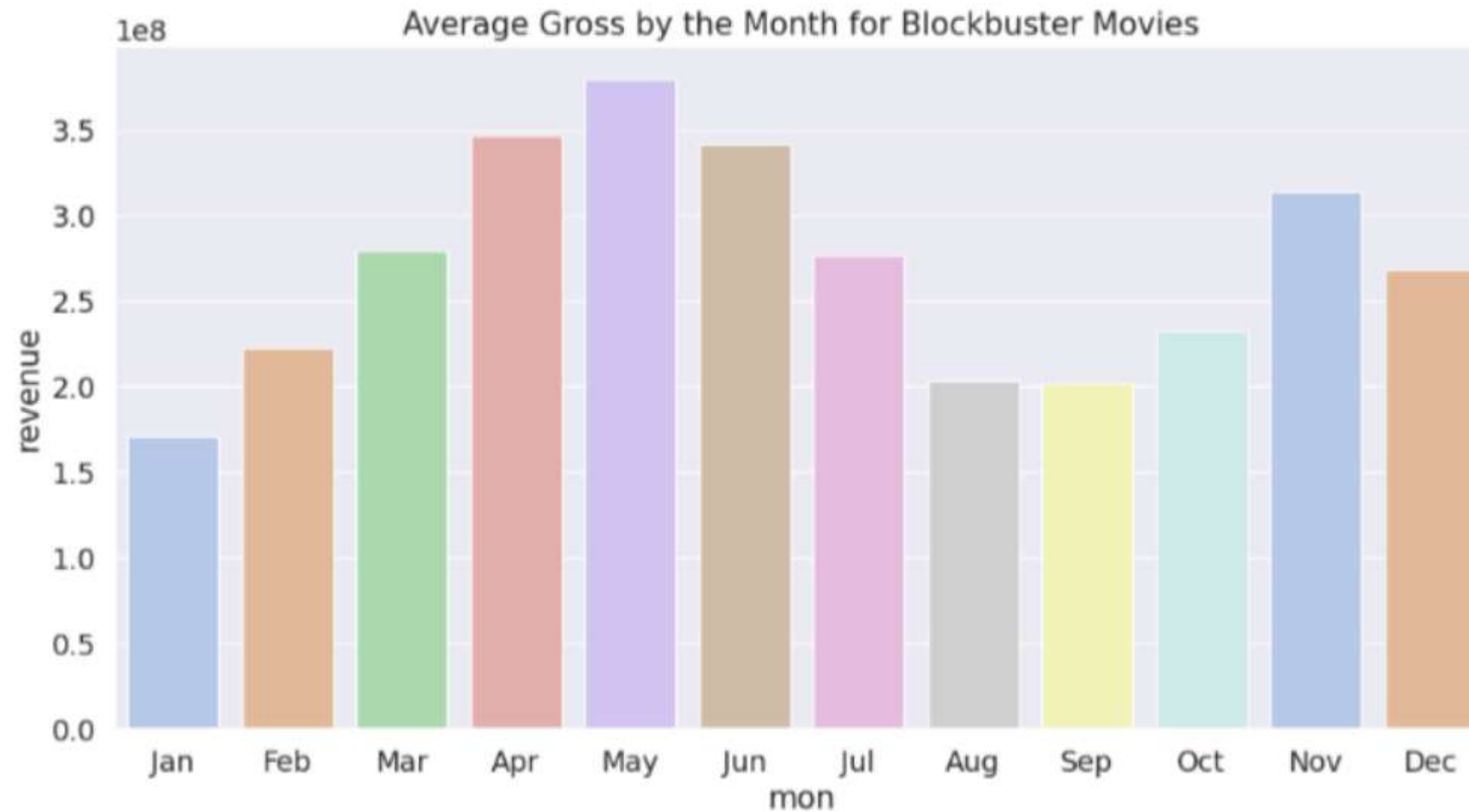
1) 흥행 요인 분석 - 예산(Budget)과 수익(Revenue)의 Scatter Plot



예산과 수익의 산점도를 확인해봤을 때 대체적으로 예산이 클 수록 영화의 수익이 증가할 확률이 높다. 하지만 이는 절대적이라고 보기는 어렵다.

04. 데이터 분석

1) 흥행 요인 분석 - 1억 달러 이상의 수익을 올린 영화들의 평균 총 수익



대체적으로 높은 수익을 기록한 영화들의 개봉 시기는 4,5,6월에 분포되어있는것을 확인 할 수 있다.



1. 여러 인기 영화들이 개봉하는 4,5,6월에 영화를 개봉하여 대중들에게 인지도를 쌓는 방법
2. 1,2,8,9,10월과 같이 상대적으로 인기 영화들이 덜 개봉하는 월을 선택하여 경쟁 상대를 줄이는 방법 두 가지를 고려해 볼 수 있다.

2) 트렌드 분석 - WordCloud를 통한 트렌드 확인



영화 제목(title) WordCloud

"Love", "Girl", "Day", "Man" 등의 단어가 자주 등장
이는 영화에서 로맨스의 보편적인 존재감을 보여준다.



영화 개요(overview) WordCloud

"Life", "One", "Find"는 영화 개요(overview)에서 자주 등장하는 단어
"Love", "Man", "Girl"과 함께, 영화에서 가장 인기 있는 테마를 확인할
수 있다.

04. 데이터 분석

3) ROI 분석 - revenue(수익)과 budget(예산)을 활용

ROI (%) = (revenue - budge) / budget * 100 으로 계산

```
1 merged_data['ROI (%)'] = ((merged_data['revenue'] -  
2 merged_data['budget']) / merged_data['budget']) * 100  
  
1 merged_data['ROI (%)'].describe()
```

ROI (%)	
count	1.124400e+04
mean	inf
std	NaN
min	-1.000000e+02
25%	-1.000000e+02
50%	2.560873e+01
75%	6.334259e+02
max	inf

dtype: float64

단순 계산시 mean = inf(무한대) , std = NaN이 나오는것을 확인

```
1 # budget이 0인 경우 확인  
2 zero_budget = merged_data[merged_data['budget'] == 0]  
3  
4 # budget이 0인 경우 ROI는 계산할 수 없으므로 해당 행들을 제거  
5 merged_data = merged_data[merged_data['budget'] > 0]
```



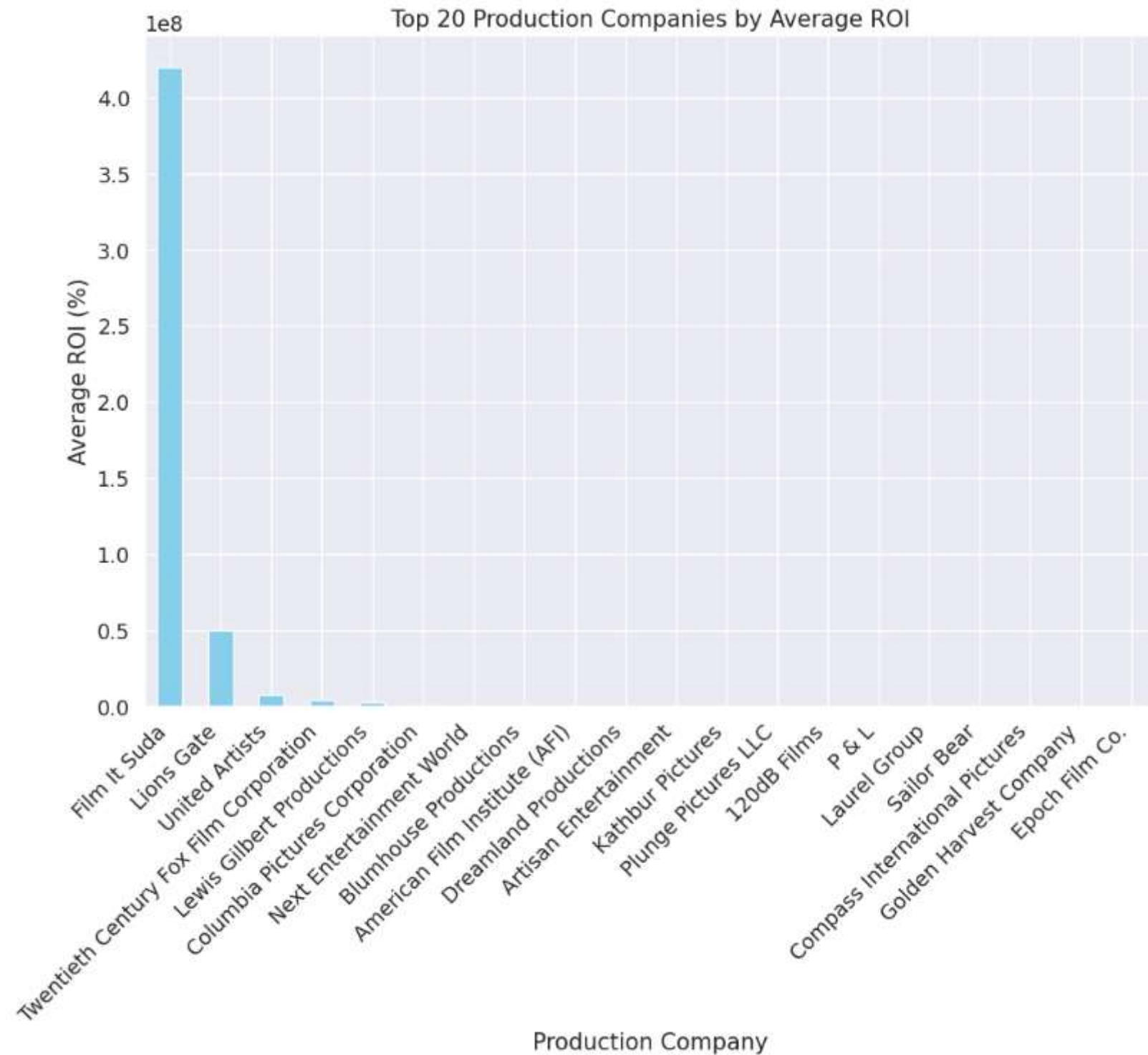
```
1 merged_data['ROI (%)'].describe()
```

ROI (%)	
count	9.133000e+03
mean	3.278364e+05
std	1.664714e+07
min	-1.000000e+02
25%	-1.000000e+02
50%	-5.265678e+01
75%	1.578497e+02
max	1.239638e+09

dtype: float64

budget(예산)이 0인 행이 존재하므로
이를 제거하고 해결

3) ROI 분석 - revenue(수익)과 budget(예산)을 활용



1. Film in Suda
2. Lions Gate
3. United Artists

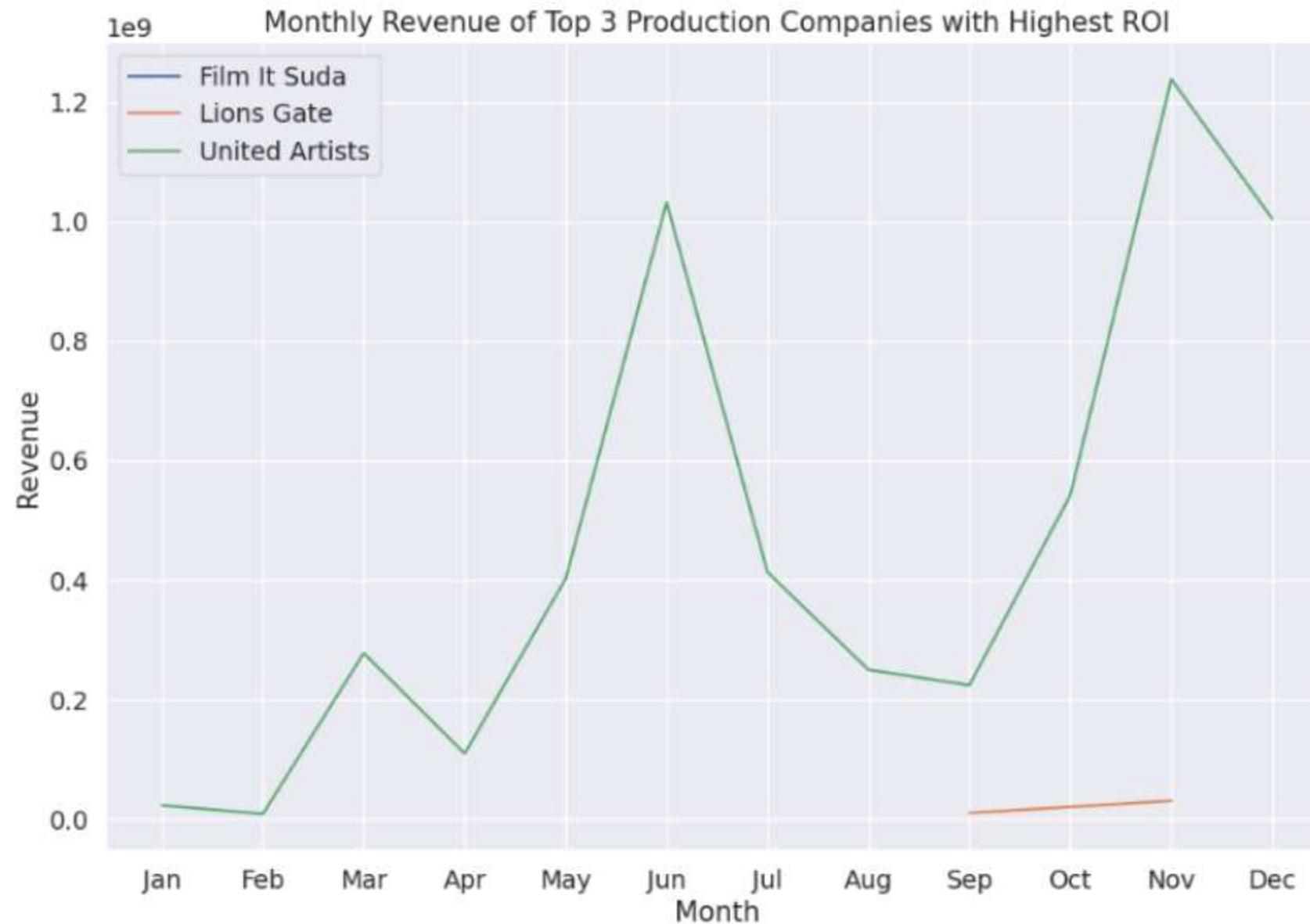
위 3개의 제작사가 ROI지수가 높은 것으로 확인
-> 3개의 제작사의 특성 확인

한계점

예산(budget)이 공개되지 않은 제작사의 정보는 제외하고 확인했기 때문에 위 3개의 제작사가 ROI지수의 TOP3라고 얘기하기는 어렵다.

04. 데이터 분석

3) ROI 분석 - 3개의 제작사 월별 ROI 분석



United Artists 제외한 나머지 두 개의 제작사는
그래프에 표현이 잘 되지 않음



```
1 print(f'Film It Suda Count : {film_it_suda_data.shape}')
2 print(f'Lions Gate Count : {lions_gate_data.shape}')
3 print(f'United Artists Count : {united_artists_data.shape}')
```

Film It Suda Count : (3, 24)
Lions Gate Count : (98, 24)
United Artists Count : (309, 24)



3개의 제작사의 실제 데이터 수를 확인해보니
Film It Suda : 3개
Lions Gate Count : 98개
United Artists Count : 309개
있기 때문에 그래프 자체로 3가지 전부 확인 불가능

3) ROI 분석 - United Artists 월별 ROI 분석



United Artists만 확인해본다면



6월, 11월이 가장 ROI(투자 대비 수익률)이 높은 것을
확인할 수 있다.



이는 흥행요인 분석에서 확인한 5,6월 / 11,12월에
가장 많은 수익을 올린 분석과 일치한다.

05. 분석 결과 & 결론

분석 결과

- ★ 흥행 요인 분석 - 온 가족이 볼 수 있는 영화가 대체적으로 성공 확률이 높아보임
- ★ 여러 인기 영화들이 개봉하는 4,5,6월에 영화를 개봉하여 대중들에게 인지도를 쌓는 방법
- ★ 1,2,8,9,10월과 같이 상대적으로 인기 영화들이 덜 개봉하는 월을 선택하여 경쟁 상대를 줄이는 방법 두 가지를 고려해 볼 수 있다.
- ★ ROI지수가 높은 영화 제작사(United Artists)를 참고한다면 6월,11월에 개봉하는 영화가 수익에 유리할 수 있다.



결론

- 창현 엔터테인먼트는
1. 온 가족이 볼 수 있는 영화를 만든다.(Adventure, Fantasy, Family 등의 Genres)
 2. 영화 주제는 Love, Life, Family 와 같은 키워드의 영화면 유리할 수 있다.
 3. 종합해 봤을 때 6월, 11월에 개봉하는 것이 인지도 측면과 수익적인 측면에서 유리할 수 있다.

아쉬운점

가장 큰 문제 : 벼락치기.

아쉬운점

- 상관관계를 확인하지 않았다.
- ROI 분석에서 전처리를 조금 더 잘 했더라면 유의미한 결과를 도출 할 수 있었을것 같다.
- movie_dataset 자체에 있는 다른 csv파일을 활용하지 못했다. (rating.csv, credits.csv 등)

지금까지 저의
발표를 들어주셔서
감사합니다.

