



# FILM EVALUATION VIA IMDB WEBSCRAPING

Matheus Sartortt & Ricardo Veloso

# Today's Discussion

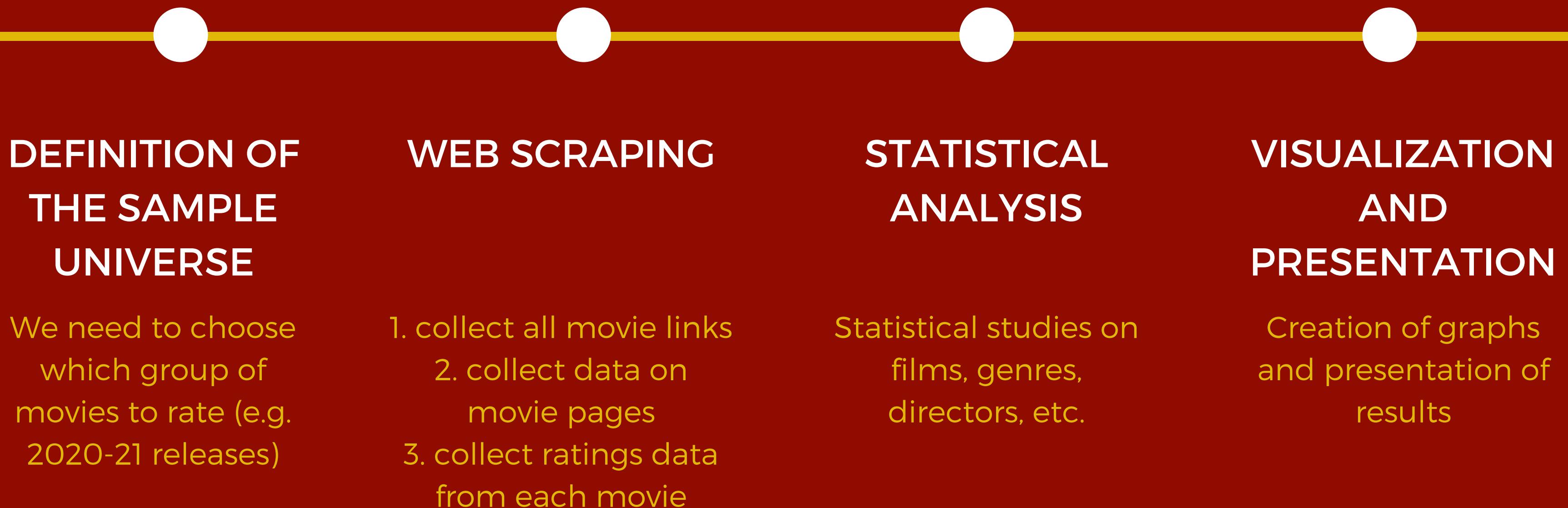
## Topics to Cover

- Project timeline
- Definition of movie universe/sampling
- Techniques used
- Results obtained
- Project difficulties
- Next steps

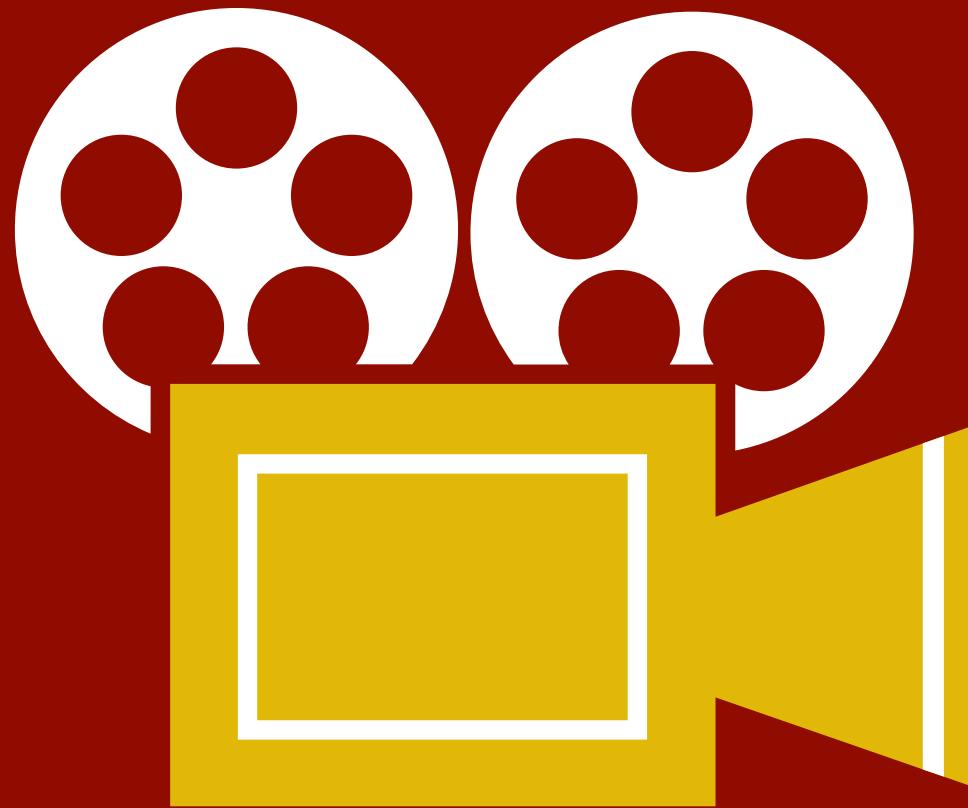


# Project Timeline

## How our activity flow works

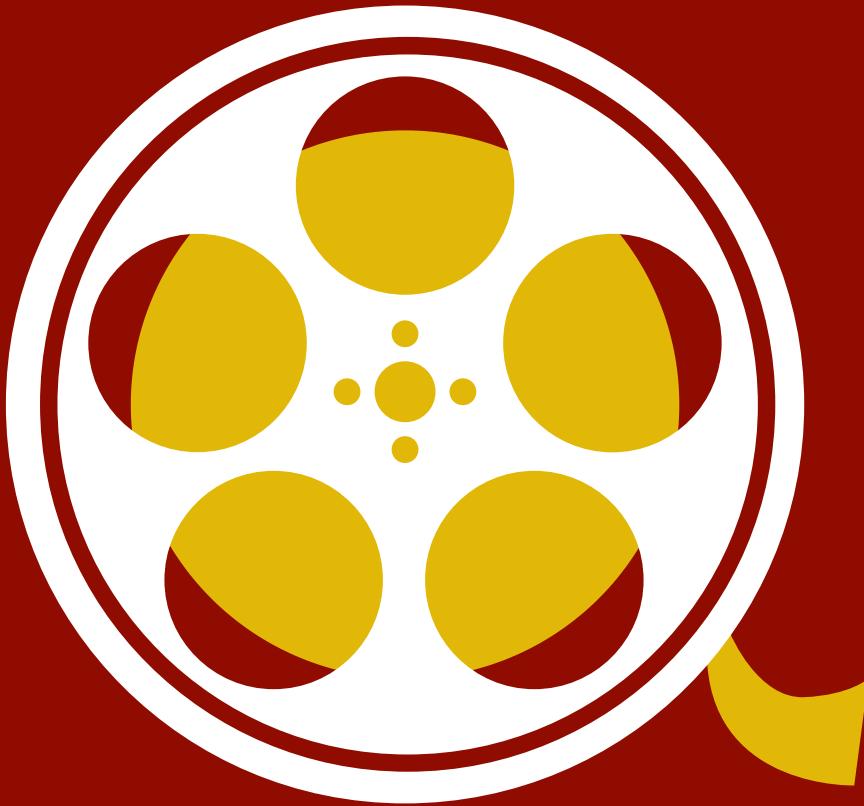


# Sample Definition



## KAGGLE

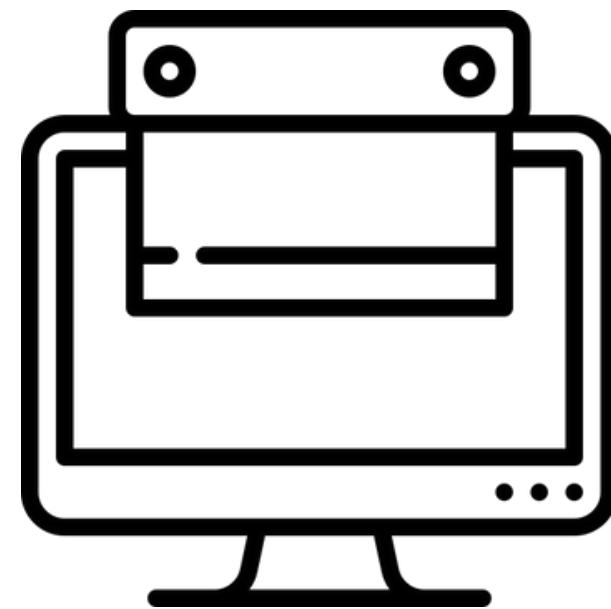
Through Kaggle we found a random listing of 5,000 movies.



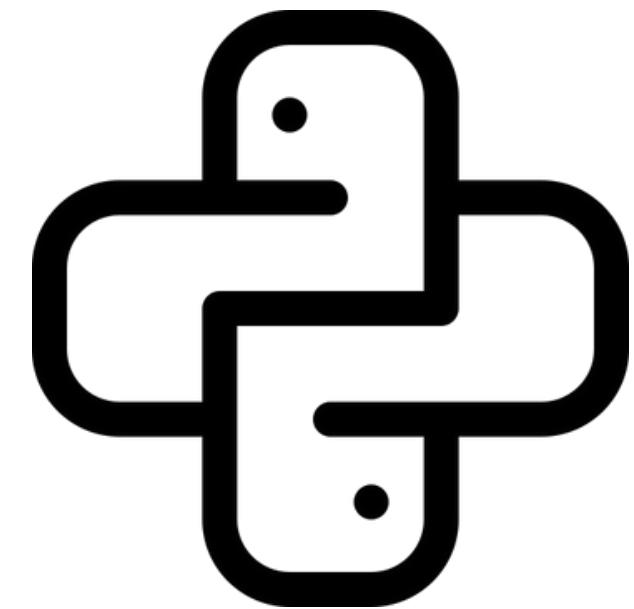
## IMDB

Através do IMDB é possível filtrar a busca de filmes por datas, gêneros, entre outros critérios.

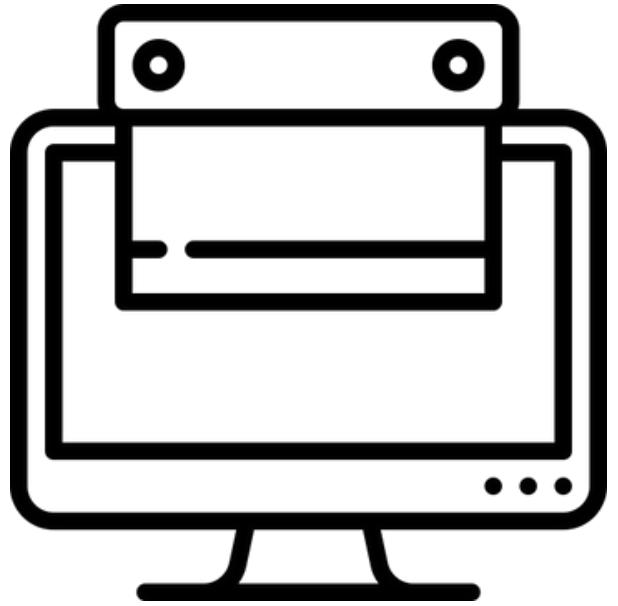
# Techniques used



**Web Scraping**



**Statistical  
analysis**



# Web Scraping

IMDb ☰ Menu All Search IMDb IMDbPro + Watchlist Matheus

## Eternals

2021 · PG-13 · 2h 36m

IMDb RATING 6.5/10 YOUR RATING POPULARITY Rate 1 ▲ 43

Cast & crew · User reviews · Trivia · IMDbPro All topics Share



28 VIDEOS

99+ PHOTOS

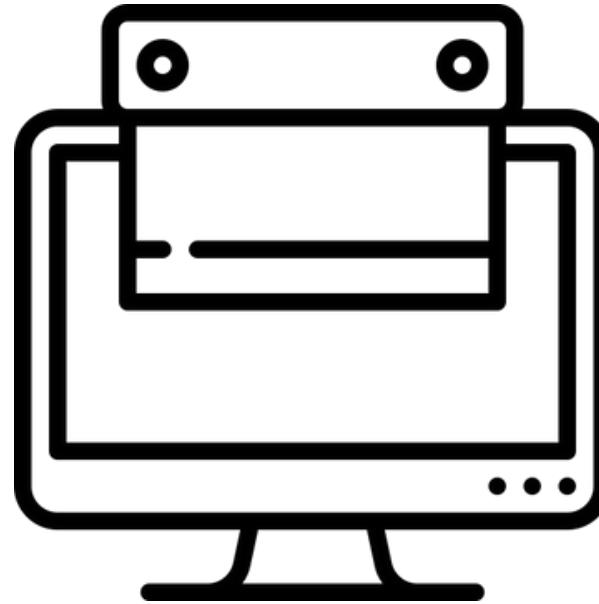
Play trailer 0:31

Action Adventure Fantasy

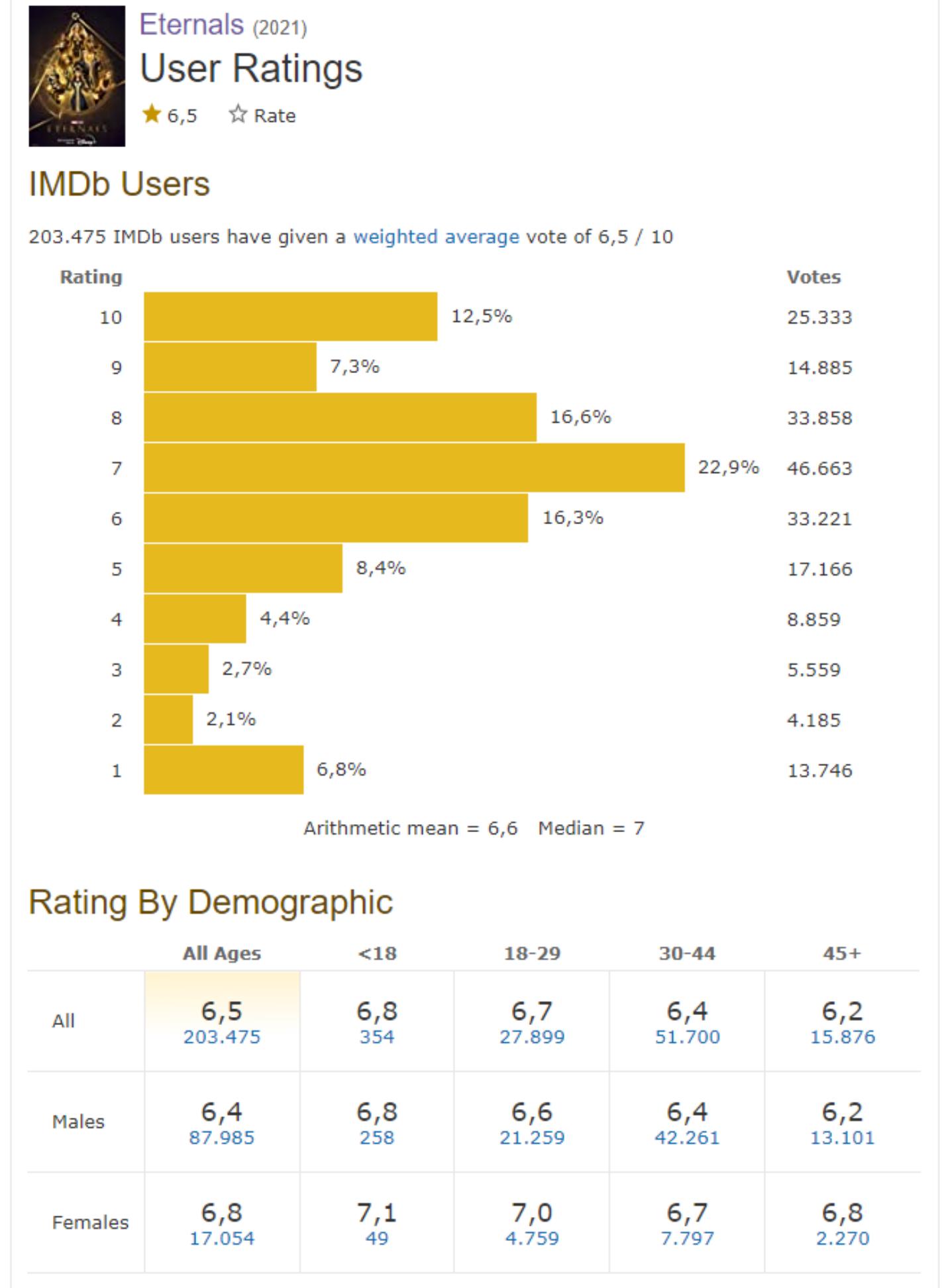
The saga of the Eternals, a race of immortal beings who lived on Earth and shaped its history and civilizations.

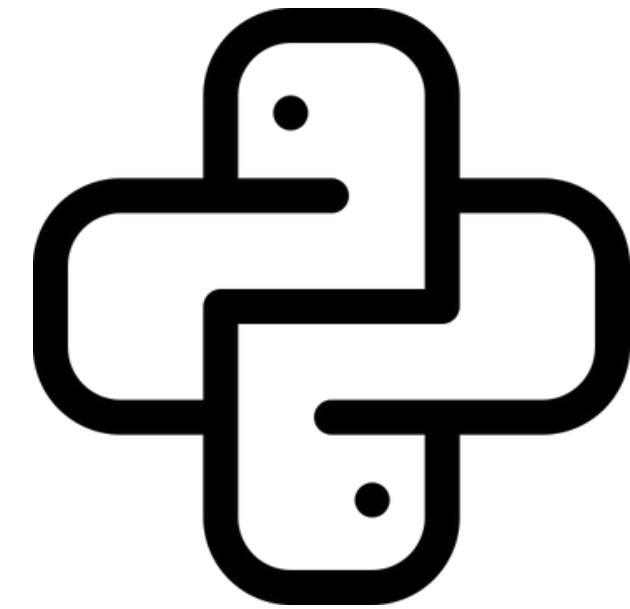
Director Chloé Zhao

2.8K User reviews 306 Critic reviews 52 Metascore



# Web Scraping

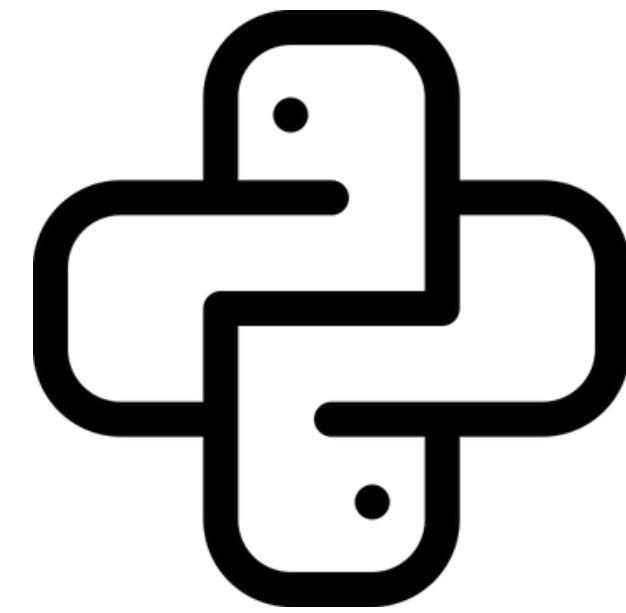




# Statistical analysis

	imdbID	movie_name	year	imdb_rating	rating_link	user_votes	director	oscars	awards_won	awards_nominated	genres	country	languages	boxoffice_budget	gross_worldwide
0	tt0499549	Avatar (2009)	2009	7.8	<a href="http://www.imdb.com/title/tt0499549/ratings?ref_=tt_urll">http://www.imdb.com/title/tt0499549/ratings?ref_=tt_urll</a>	1.2M	James Cameron	3	89	131	[Action, Adventure, Fantasy]	[United States]	[English, Spanish]	237000000	2847246203
1	tt0449088	Piratas das Caraíbas - Nos Confins do Mundo (2...	2007	7.1	<a href="http://www.imdb.com/title/tt0449088/ratings?ref_=tt_urll">http://www.imdb.com/title/tt0449088/ratings?ref_=tt_urll</a>	624K	Gore Verbinski	2	23	51	[Action, Adventure, Fantasy]	[United States]	[English]	300000000	960996492
2	tt2379713	007 Spectre (2015)	2015	6.8	<a href="http://www.imdb.com/title/tt2379713/ratings?ref_=tt_urll">http://www.imdb.com/title/tt2379713/ratings?ref_=tt_urll</a>	425K	Sam Mendes	1	8	37	[Action, Adventure, Thriller]	[United Kingdom, United States]	[English, Spanish, Italian, German, French]	245000000	880681519
3	tt1345836	O Cavaleiro das Trevas Renasce (2012)	2012	8.4	<a href="http://www.imdb.com/title/tt1345836/ratings?ref_=tt_urll">http://www.imdb.com/title/tt1345836/ratings?ref_=tt_urll</a>	1.6M	Christopher Nolan	0	39	103	[Action, Crime, Drama]	[United Kingdom, United States]	[English, Arabic]	250000000	1061142612
4	tt5289954	"Disneyember" Star Wars: Episode VII - The Fo...	2015	6.8	<a href="http://www.imdb.com/title/tt5289954/ratings?ref_=tt_urll">http://www.imdb.com/title/tt5289954/ratings?ref_=tt_urll</a>	31	Doug Walker	0	0	0	[Documentary, Short]				
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5037	tt3000844	"Signed, Sealed, Delivered" Second Chances (TV...	2013	7.7	<a href="http://www.imdb.com/title/tt3000844/ratings?ref_=tt_urll">http://www.imdb.com/title/tt3000844/ratings?ref_=tt_urll</a>	1.5K	Scott Smith	0	0	2	[Comedy, Drama, Mystery]	[Canada]	[English]		
5038	tt2071645	Os Seguidores (TV Series 2013–2015)	2013	7.4	<a href="http://www.imdb.com/title/tt2071645/ratings?ref_=tt_urll">http://www.imdb.com/title/tt2071645/ratings?ref_=tt_urll</a>	86K	Kevin Williamson	0	3	14	[Crime, Drama, Horror]	[United States]	[English]		
5039	tt2107644	A Plague So Pleasant (2013)	2013	5.5	<a href="http://www.imdb.com/title/tt2107644/ratings?ref_=tt_urll">http://www.imdb.com/title/tt2107644/ratings?ref_=tt_urll</a>	84	Benjamin Robards	0	0	0	[Drama, Horror, Thriller]	[United States]	[English]	1400	
5040	tt2070597	Destino: Shanghai (2012)	2012	6.2	<a href="http://www.imdb.com/title/tt2070597/ratings?ref_=tt_urll">http://www.imdb.com/title/tt2070597/ratings?ref_=tt_urll</a>	1.7K	Daniel Hsia	0	7	1	[Comedy, Drama, Romance]	[United States, China]	[English, Mandarin, Shanghainese]		10400
5041	tt0378407	O Meu Encontro com Drew (2004)	2004	6.4	<a href="http://www.imdb.com/title/tt0378407/ratings?ref_=tt_urll">http://www.imdb.com/title/tt0378407/ratings?ref_=tt_urll</a>	4.6K	Jon Gunn	0	11	0	[Documentary]	[United States]	[English]	1100	262770

4918 rows × 15 columns



# Statistical analysis

	Movie Id	Gender	All Ages	<18	18-29	30-44	45+
0	tt0499549	All	7.8	7.6	7.7	7.8	7.9
1	tt0499549	Males	7.8	7.5	7.7	7.8	7.9
2	tt0499549	Females	7.7	7.6	7.6	7.7	7.9
3	tt0449088	All	7.1	7.7	7.4	7.0	6.7
4	tt0449088	Males	7.0	7.7	7.3	7.0	6.6
...	...	...	...	...	...	...	...
15115	tt2070597	Males	6.2	-	6.5	6.2	6.2
15116	tt2070597	Females	6.3	-	6.3	6.3	6.5
15117	tt0378407	All	6.4	-	6.2	6.5	6.4
15118	tt0378407	Males	6.5	-	6.2	6.5	6.4
15119	tt0378407	Females	6.3	-	6.4	6.3	6.3

15120 rows × 7 columns

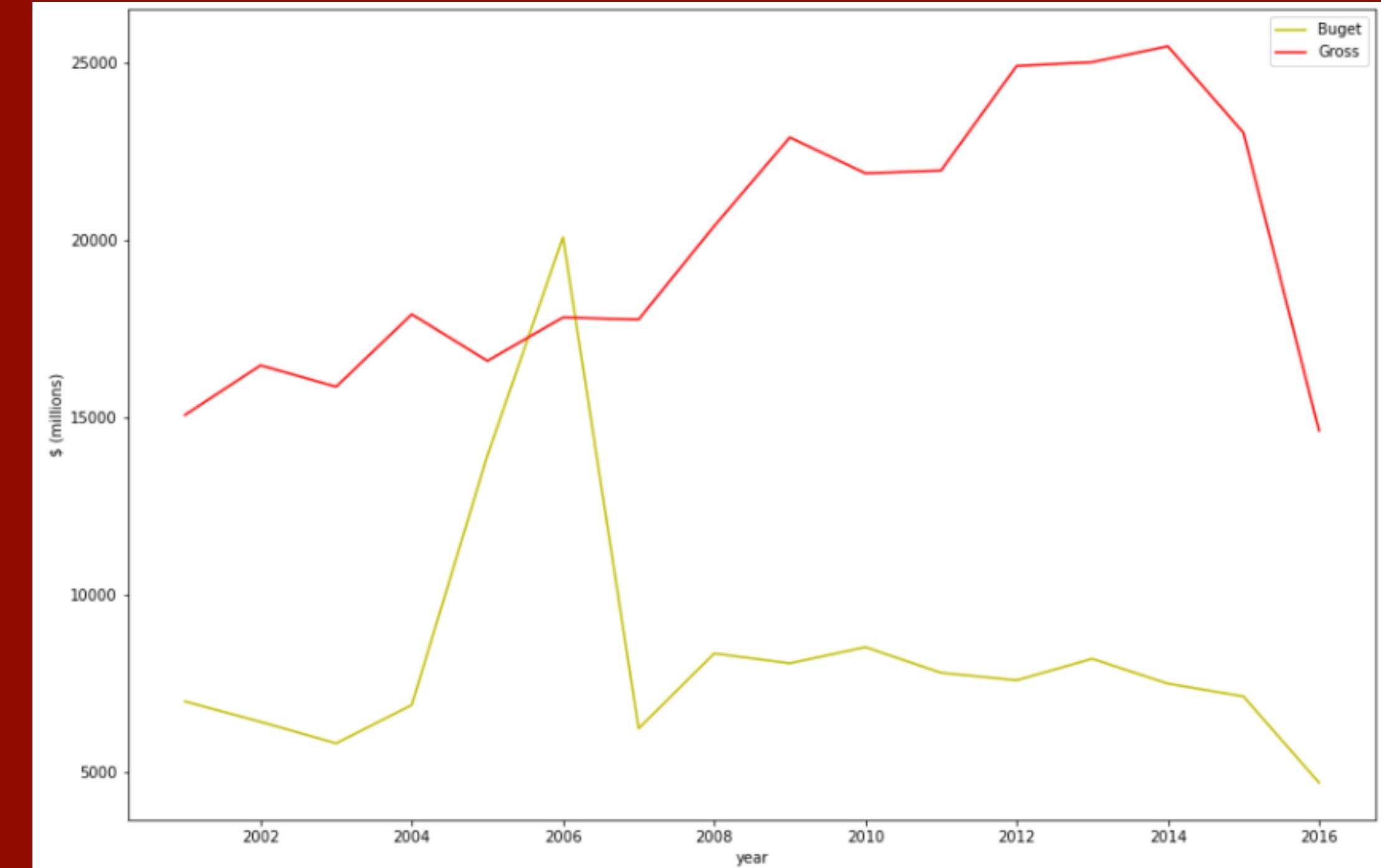


# A deep dive in movies

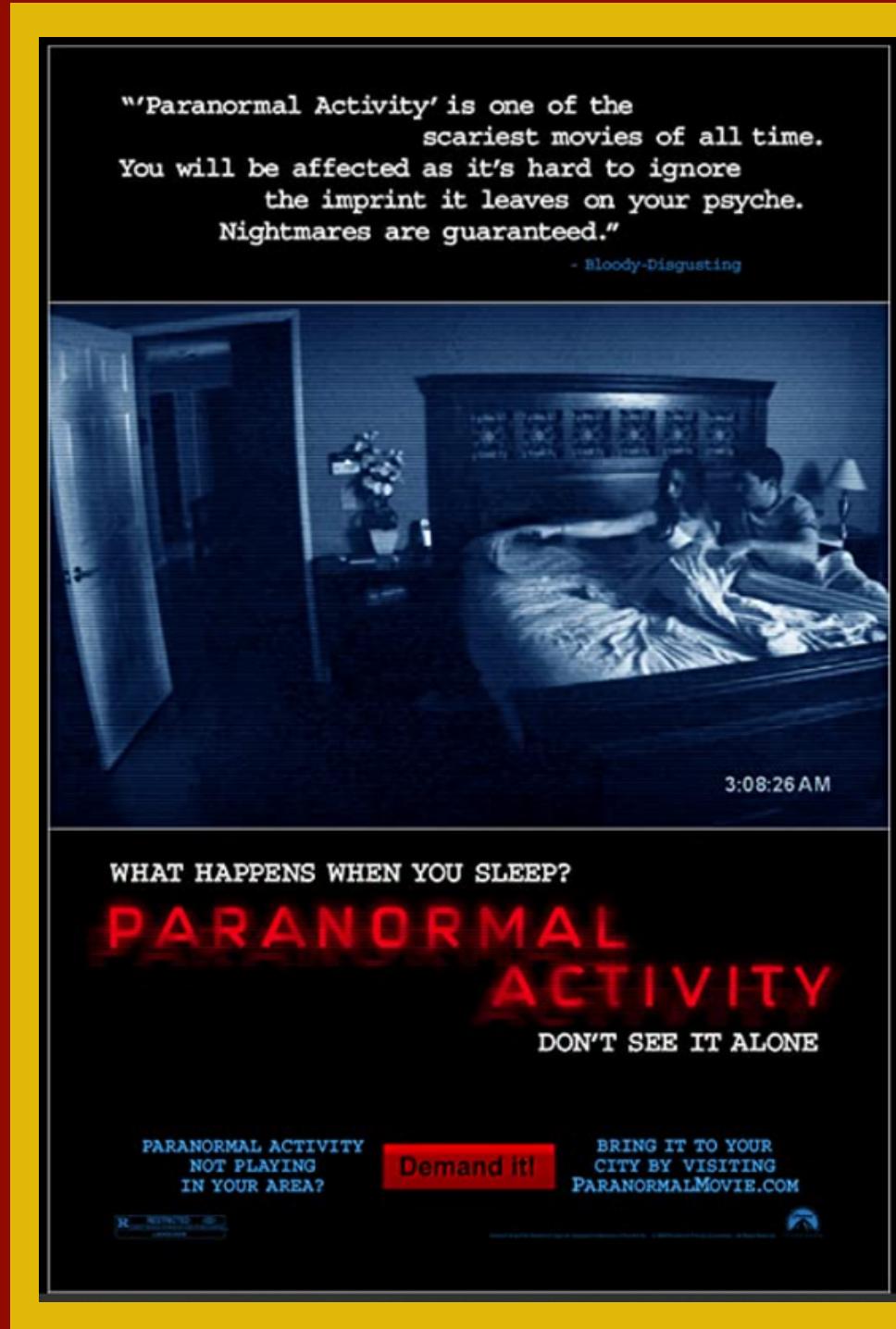
OUR RESULTS

# Annual Survey

**Total budget  
and gross  
worldwide in  
millions of  
dollars**



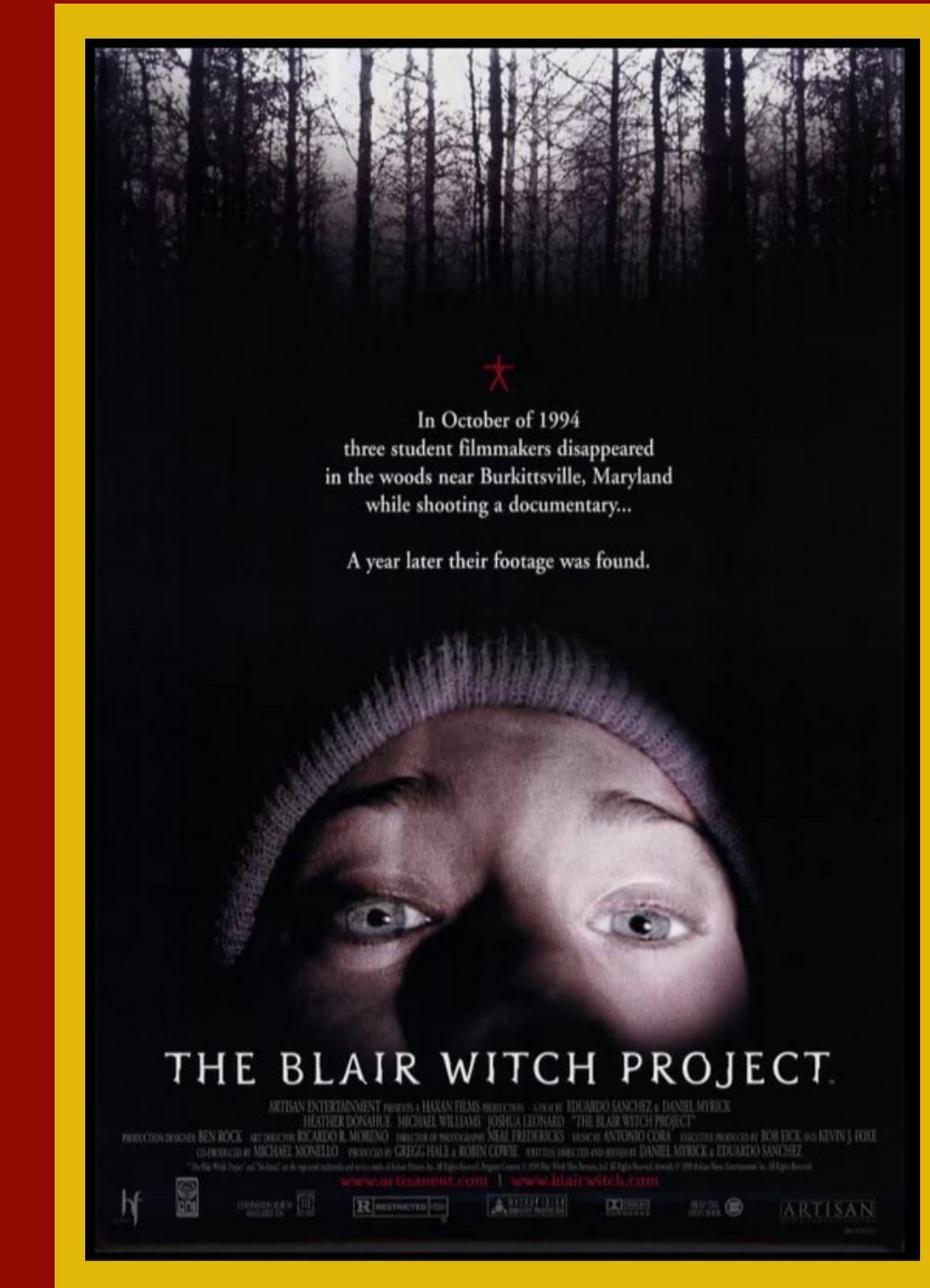
# TOP 2 BEST MOVIES BY ROI



## Box office

Budget  
\$15,000 (estimated)  
Gross US & Canada  
\$107,918,810  
Opening weekend US & Canada  
\$77,873 · Sep 27, 2009  
Gross worldwide  
\$193,355,800

**ROI of  
12889.39%**



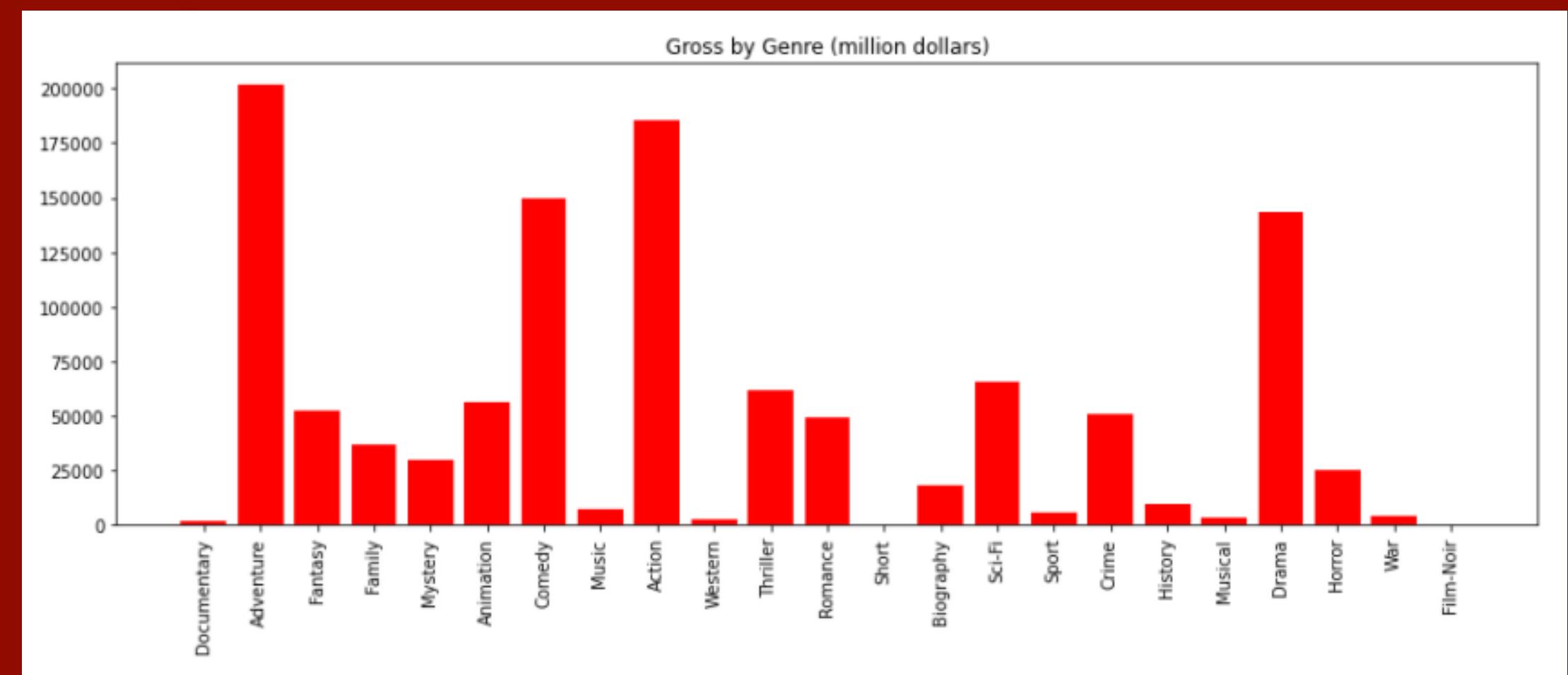
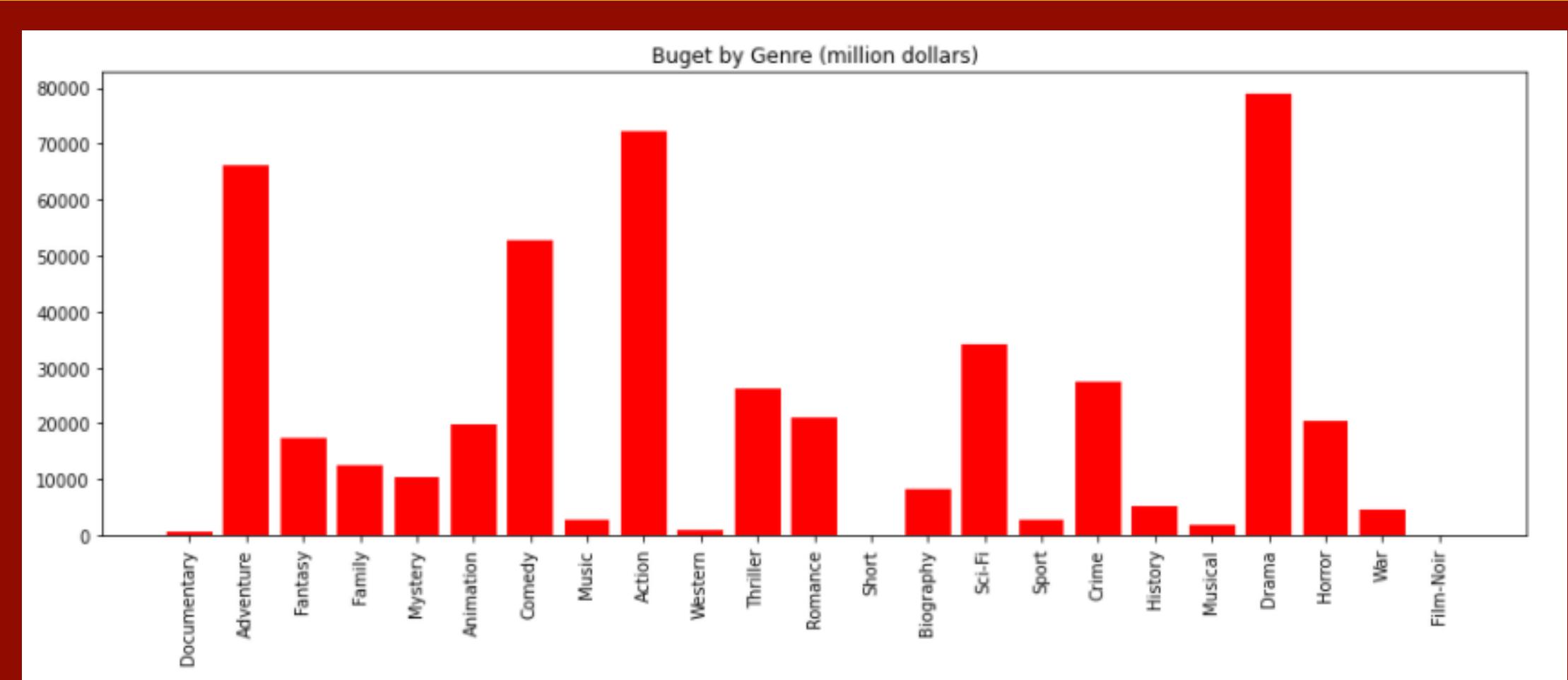
**ROI of  
4142.98%**

## Box office

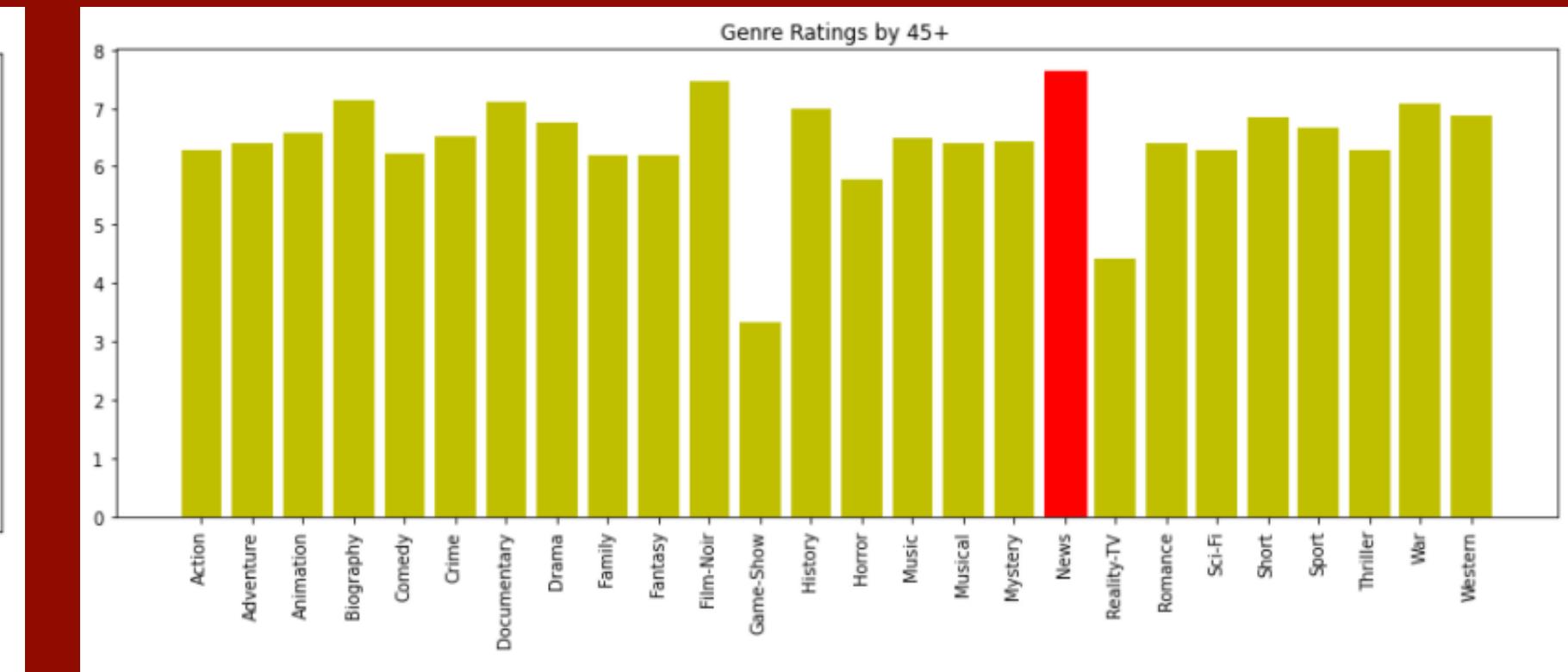
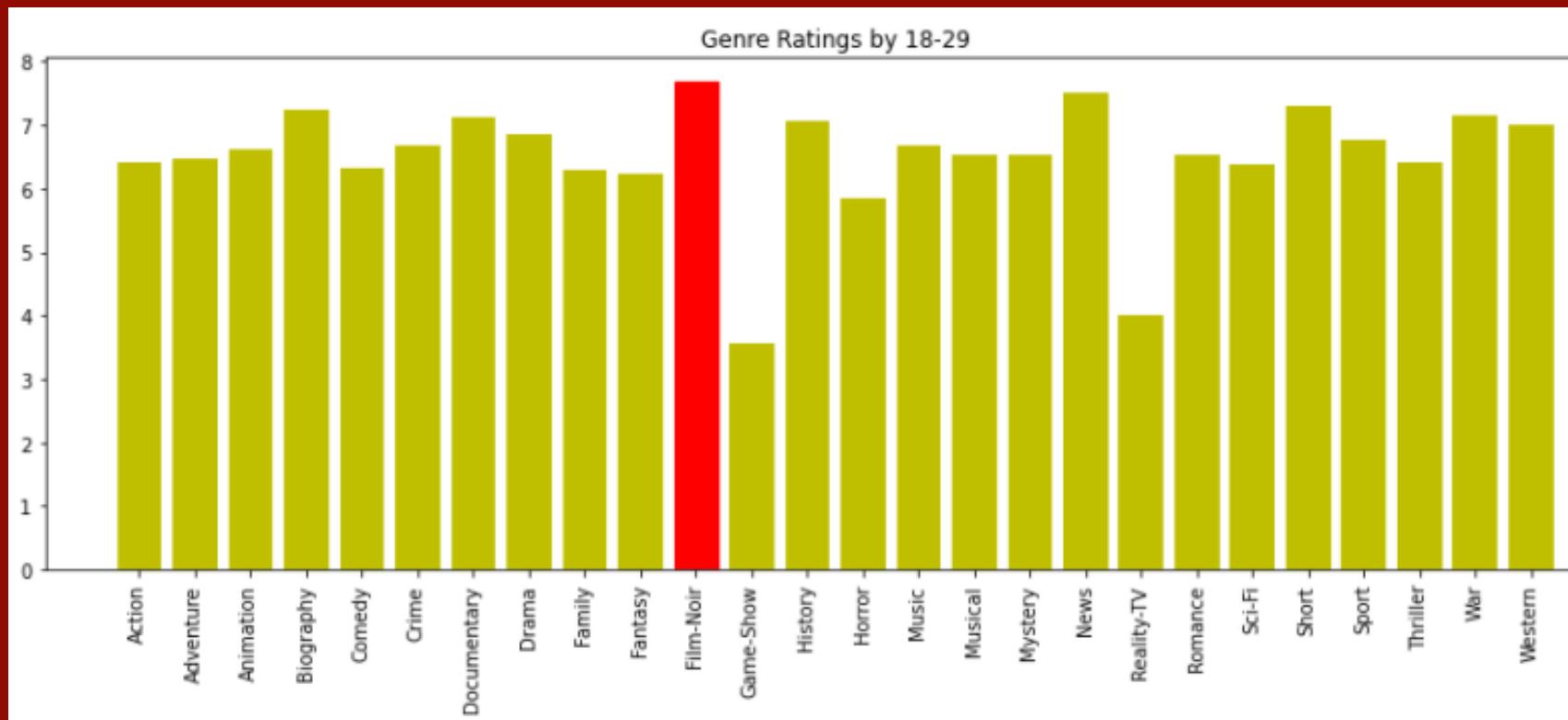
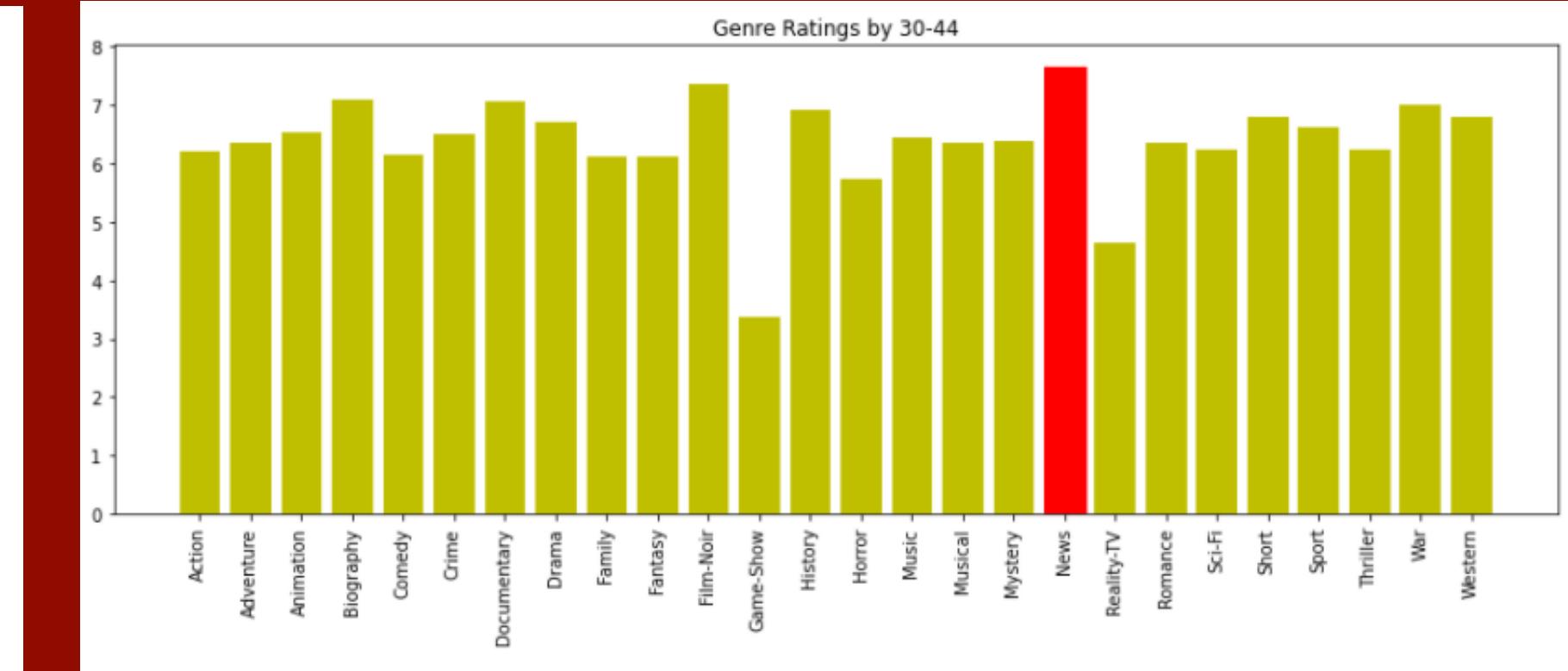
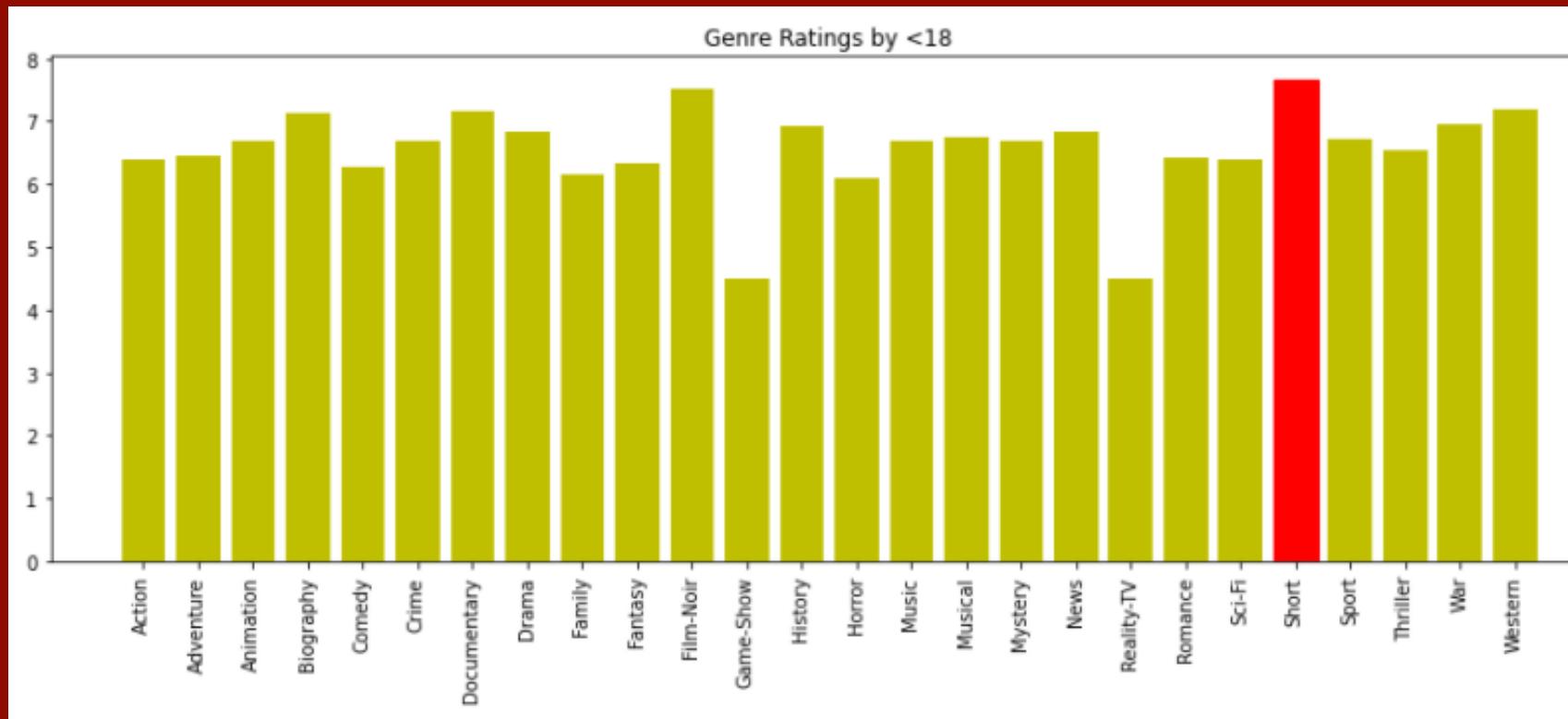
Budget  
\$60,000 (estimated)  
Gross US & Canada  
\$140,539,099  
Opening weekend US & Canada  
\$1,512,054 · Jul 18, 1999  
Gross worldwide  
\$248,639,099

# Genre Survey

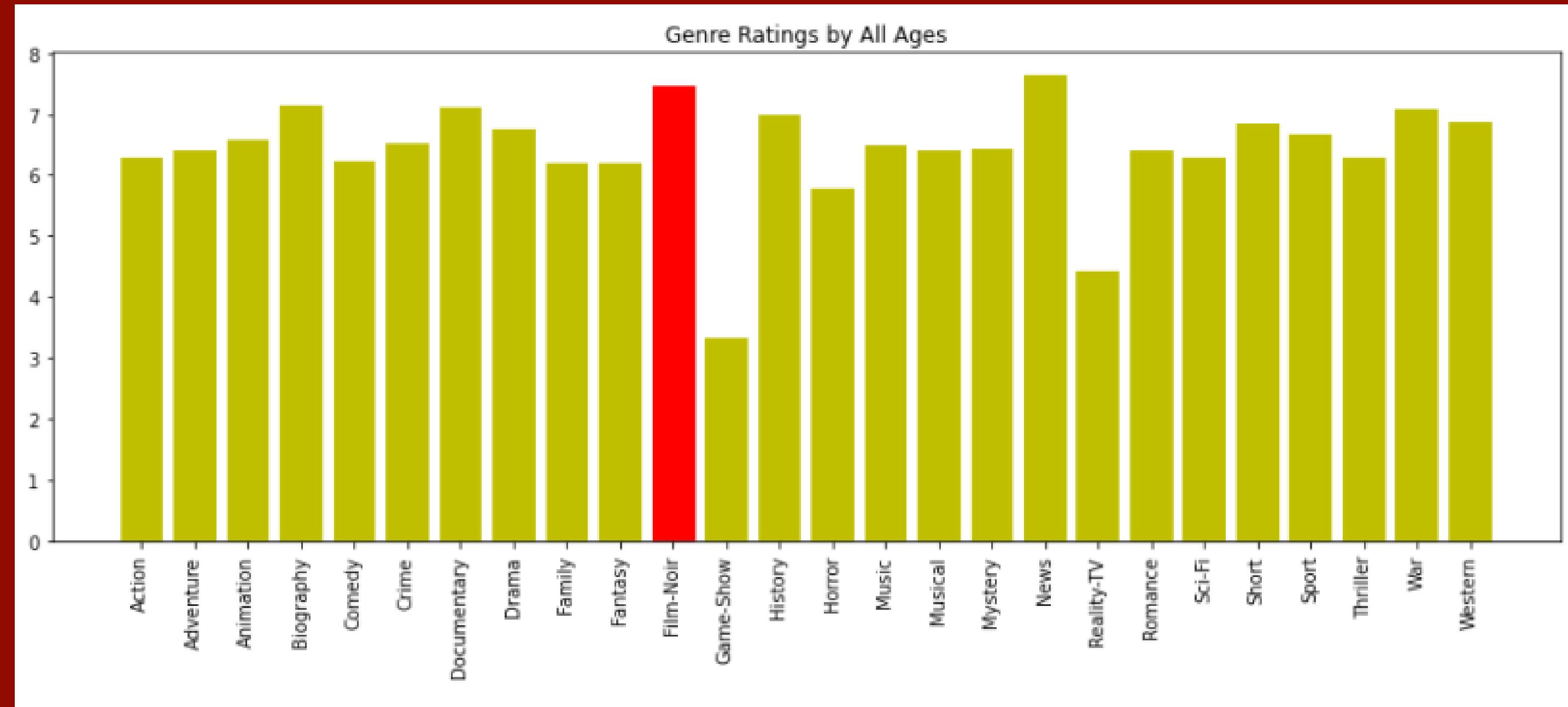
## Total budget and gross worldwide in millions of dollars



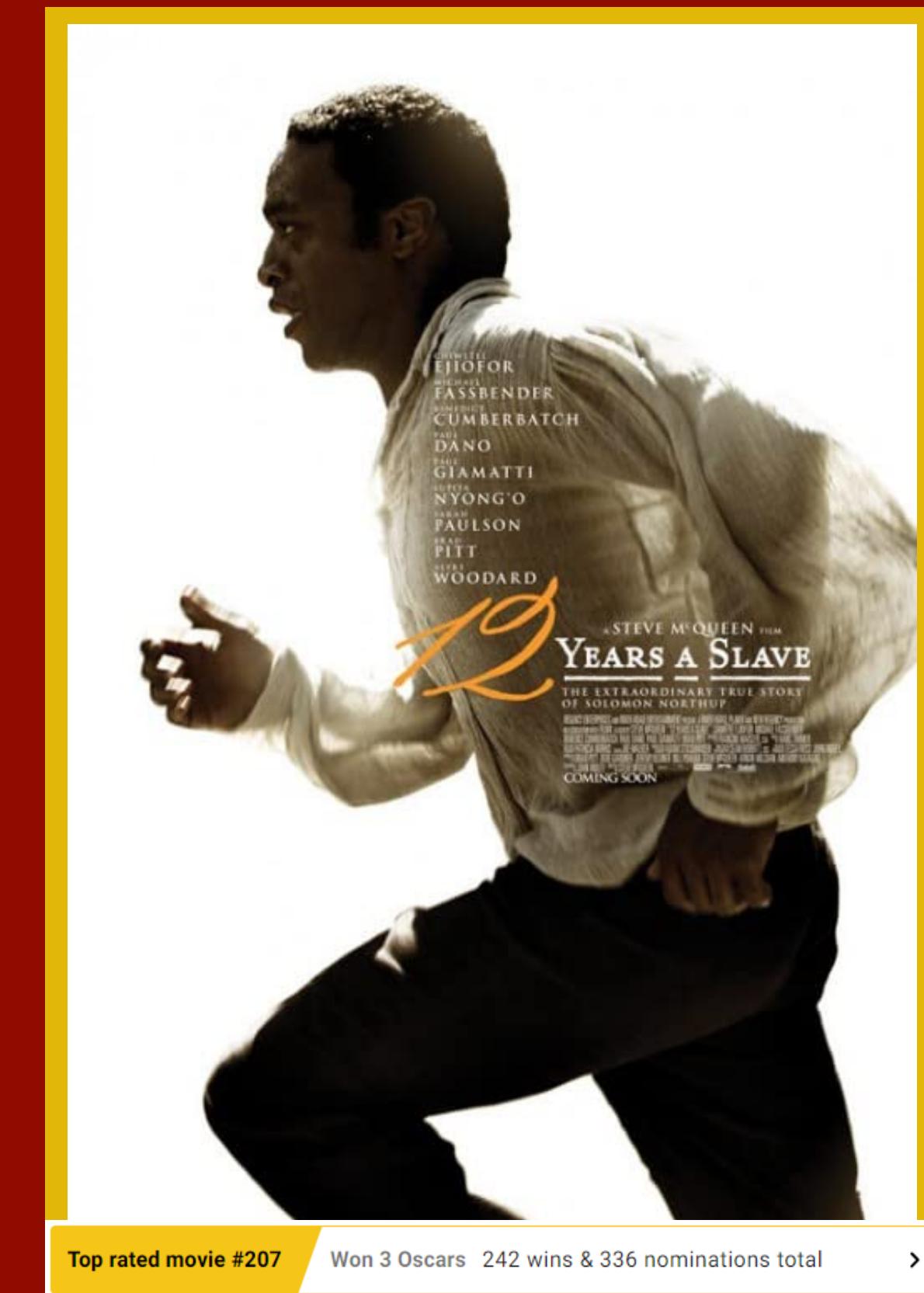
# Genre Ratings by Age Group



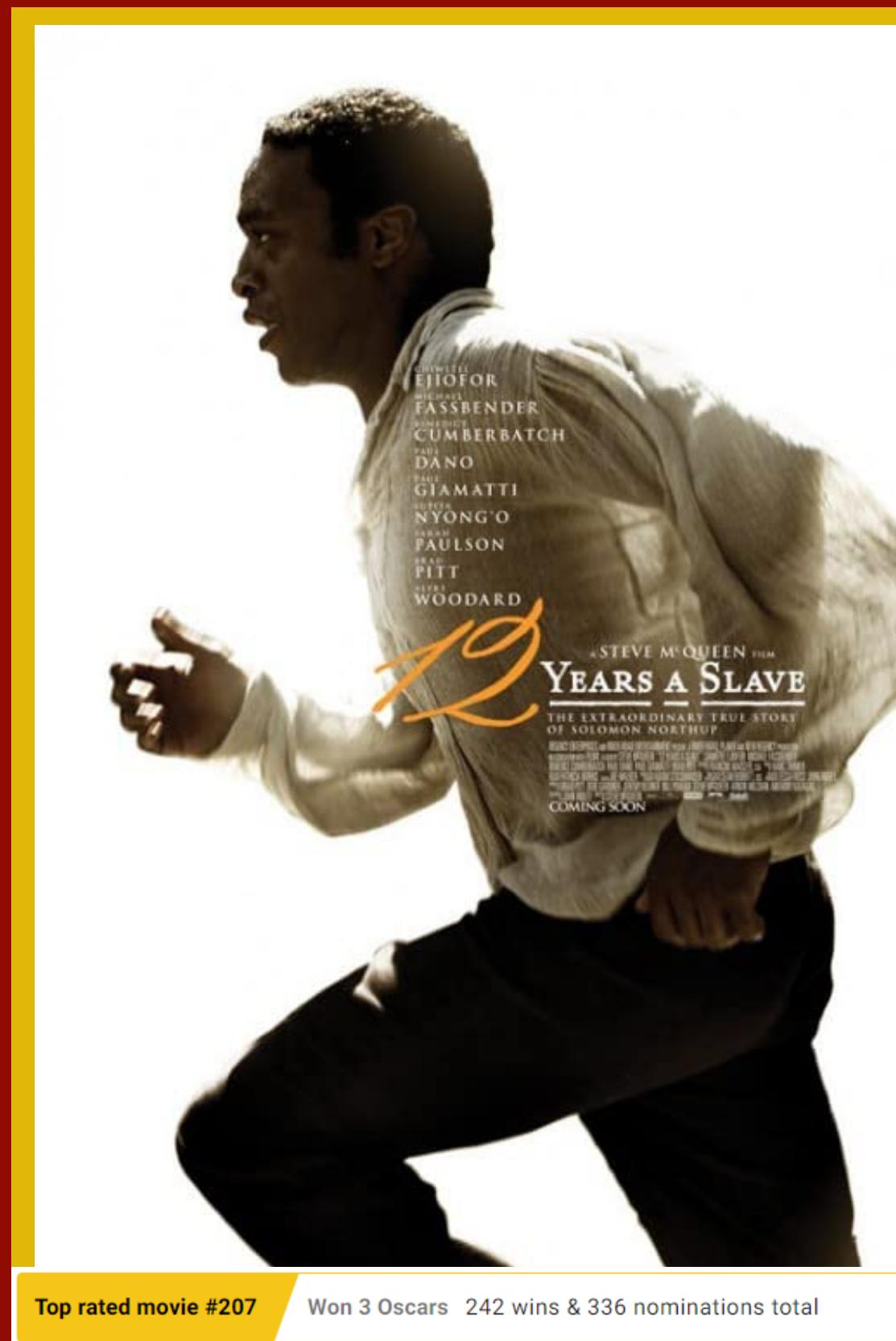
# Genre Ratings by Age Group



# TOP 2 FILMS WITH MOST AWARDS WON



# TOP 2 FILMS WITH MOST NOMINATIONS



Top rated movie #207

Won 3 Oscars 242 wins & 336 nominations total



Won 4 Oscars 193 wins & 291 nominations total

# **Project difficulties**

- FIND AN APPROPRIATELY SIZED MOVIE LISTING
- WEBSCRAPING PROCESSING DELAY
- CONSTANT SERVER CRASHES
- COVID-19

## Next steps

- OPTIMIZE WEB SCRAPING PROCESSING TIME
- IMPROVE DATA VISUALIZATION (TABLEAU,  
OTHER PYTHON PACKAGES)
- DIVERSIFY AND EXPLORE OTHER STATISTICS  
AND ANALYTICS

“  
If you can film an idea in your mind,  
follow that film idea shot for shot, scene  
for scene, that idea is worth making.

CRAIG MAPP

# Any questions?

**MATHEUS SARTORT**

[linkedin.com/in/msartortt/](https://linkedin.com/in/msartortt/)

**RICARDO VELOSO**

[linkedin.com/in/ricardo-veloso-10a544167/](https://linkedin.com/in/ricardo-veloso-10a544167/)

