

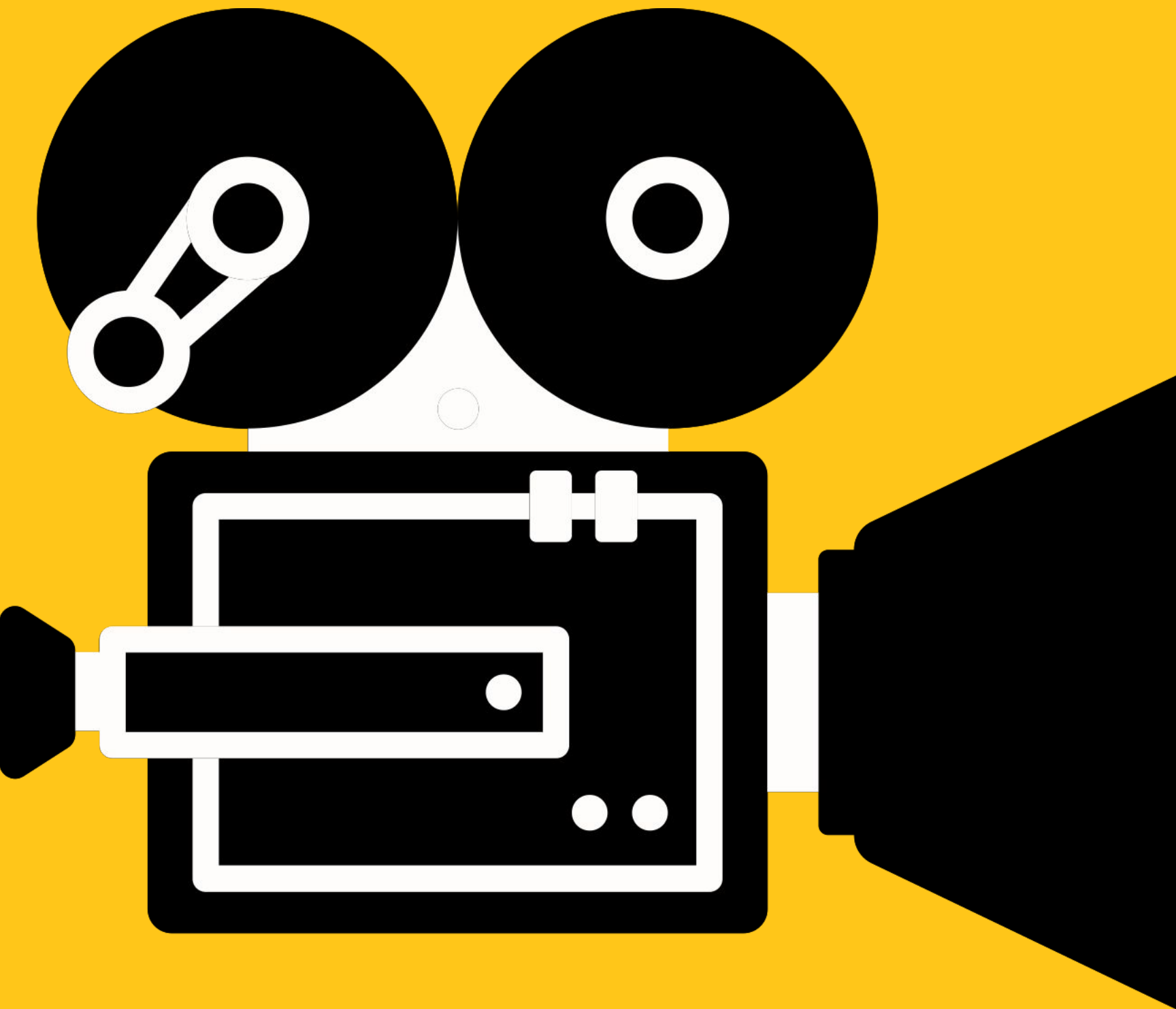


# MOVIE RATING PREDICTIONS

**Presentation by:** Rebekah Aldrich, Christopher Hornung,  
Alejandro Davila, Bryant Griessel, & Ashley Ley



# Project Flow



**1**

**Collect Data:** scrape Numbers website (Python, Pandas, BeautifulSoup, NumPy) & automated querying OMDB

---

**2**

**Process Data:** Identify potential features and target datasets  
Process data to align with approaches

---

**3**

**Machine Learning Experimentation:**  
Determine feature importance (sckit-learn & Random Forest) and determine optimal model activation

---

**4**

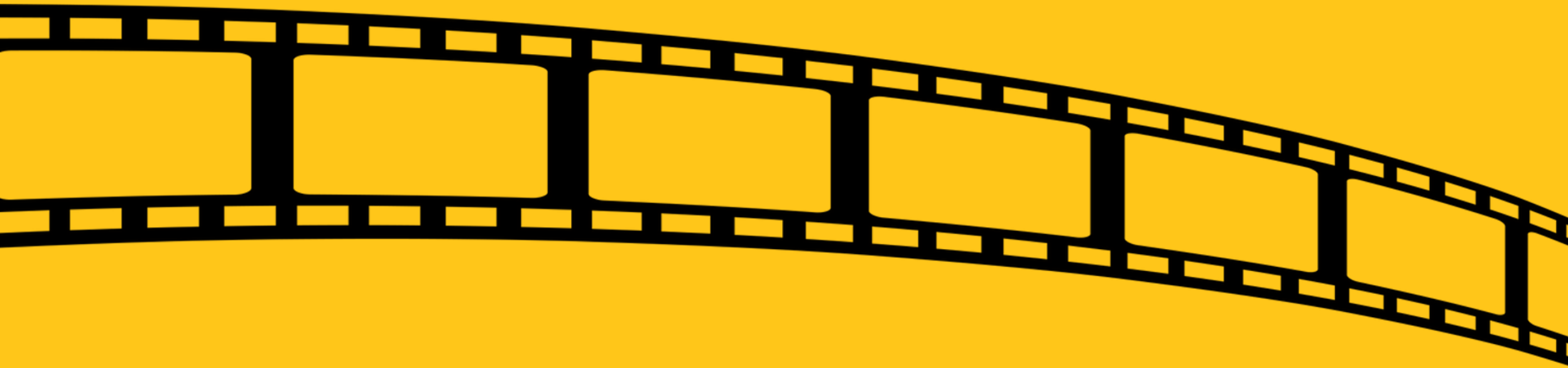
**Model Training:** Complete optimal training and document results

---

**5**

**Result Analysis:** Movie Recommendation Pages and Tableau Visualizations

---



01

# DATA COLLECTION



## Numbers Webpage - Production Budget & Worldwide Gross

Web Scraped 6,444 rows across 65 pages  
Limited overall data for model creation

---



## OMDB API - Ratings, Key Creators, Genre

API data limits & lengthy calls  
Created writer, director, actor ratings

---



## IMDB Dataset- Ratings, Key Creators, Genre

220k - 440k rows  
SQL *Significantly* Slower than Python Manipulation

---



## SciPy, Matplotlib

Preview trends, relationships to direct machine learning experiments



## Movie Budgets

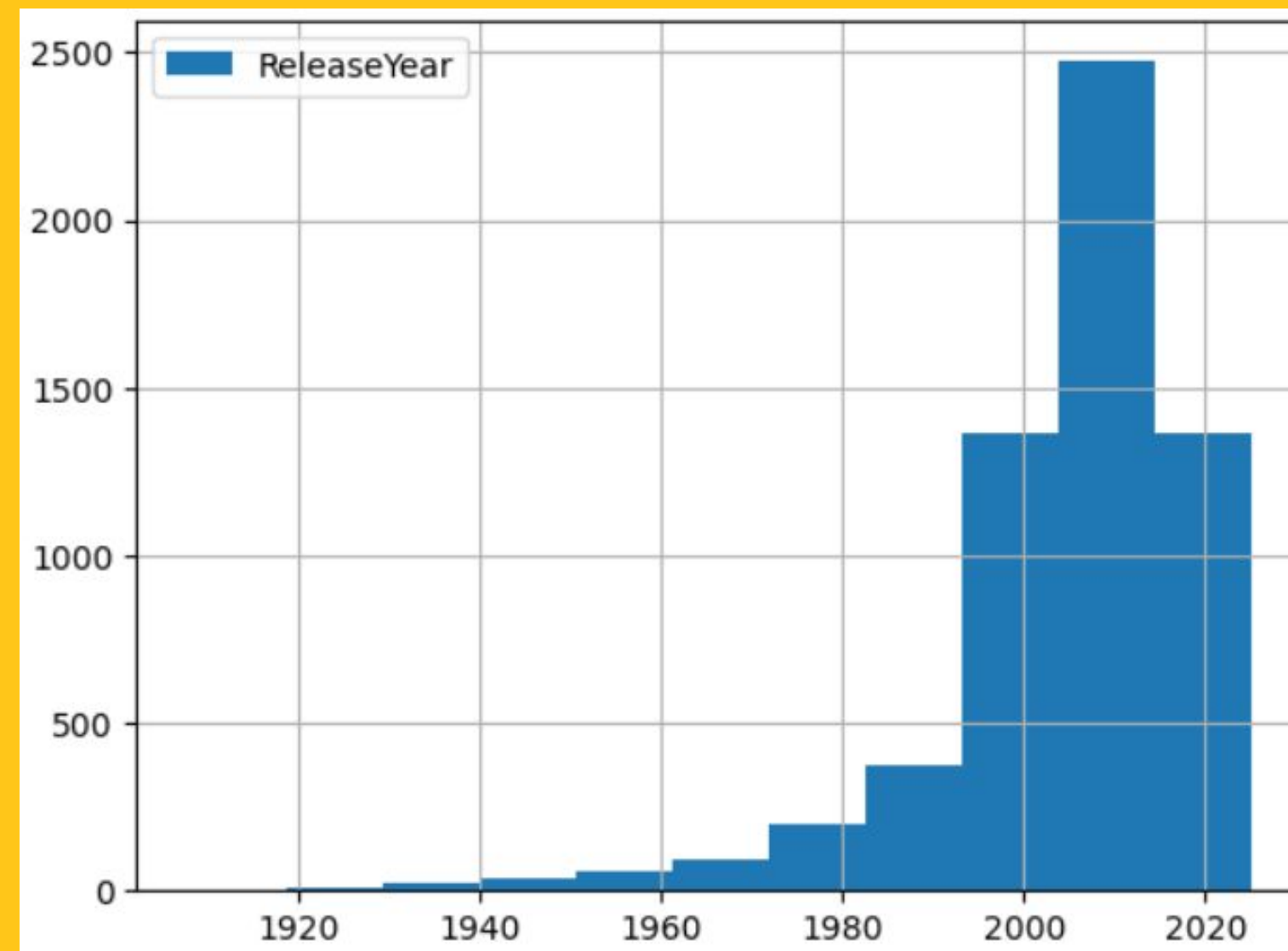
**Note:** Budget numbers for movies can be both difficult to find and unreliable. Studios and film-makers often try to keep the information secret and will use accounting tricks to inflate or reduce announced budgets.

This chart shows the budget of every film in our database, where we have it. The data we have is, to the best of our knowledge, accurate but there are gaps and disputed figures. If you have additional information or corrections, please let us know at [corrections@the-numbers.com](mailto:corrections@the-numbers.com).

Our movie profit and loss records, based on this budget information, can be found [here](#).

Release Date	Movie	Production Budget	Domestic Gross	Worldwide Gross
1 Dec 9, 2022	<a href="#">Avatar: The Way of Water</a>	\$460,000,000	\$684,075,767	\$2,319,591,720
2 Apr 23, 2019	<a href="#">Avengers: Endgame</a>	\$400,000,000	\$858,373,000	\$2,788,912,285
3 May 20, 2011	<a href="#">Pirates of the Caribbean: On Stranger Tides</a>	\$379,000,000	\$241,071,802	\$1,045,713,802
4 Apr 22, 2015	<a href="#">Avengers: Age of Ultron</a>	\$365,000,000	\$459,005,868	\$1,395,316,979
5 May 17, 2023	<a href="#">Fast X</a>	\$340,000,000	\$145,960,660	\$714,414,576
6 Dec 16, 2015	<a href="#">Star Wars Ep. VII: The Force Awakens</a>	\$306,000,000	\$936,662,225	\$2,064,615,817
7 Apr 25, 2018	<a href="#">Avengers: Infinity War</a>	\$300,000,000	\$678,815,482	\$2,048,359,754
8 May 24, 2007	<a href="#">Pirates of the Caribbean: At World's End</a>	\$300,000,000	\$309,420,425	\$960,996,492
9 Nov 13, 2017	<a href="#">Justice League</a>	\$300,000,000	\$229,024,295	\$655,945,209
10 Jun 28, 2023	<a href="#">Indiana Jones and the Dial of Destiny</a>	\$300,000,000	\$174,480,468	\$381,561,976
11 Jul 11, 2023	<a href="#">Mission: Impossible Dead Reckoning Part One</a>	\$290,000,000	\$172,135,383	\$566,325,939
12 Dec 18, 2019	<a href="#">Star Wars: The Rise of Skywalker</a>	\$275,000,000	\$515,202,542	\$1,072,767,997
13 May 23, 2018	<a href="#">Solo: A Star Wars Story</a>	\$275,000,000	\$213,767,512	\$393,151,347
14 Nov 8, 2023	<a href="#">The Marvels</a>	\$274,800,000	\$76,755,559	\$186,927,954
15 Mar 7, 2012	<a href="#">John Carter</a>	\$263,700,000	\$73,058,679	\$282,778,100
16 Mar 23, 2016	<a href="#">Batman v Superman: Dawn of Justice</a>	\$263,000,000	\$330,360,194	\$872,395,091
17 Dec 13, 2017	<a href="#">Star Wars Ep. VIII: The Last Jedi</a>	\$262,000,000	\$620,181,382	\$1,331,635,141
18 Jul 11, 2019	<a href="#">The Lion King</a>	\$260,000,000	\$543,638,043	\$1,646,106,779
19 Nov 24, 2010	<a href="#">Tangled</a>	\$260,000,000	\$200,821,936	\$582,440,151
20 May 4, 2007	<a href="#">Spider-Man 3</a>	\$258,000,000	\$336,530,303	\$894,860,230
21 Nov 9, 2022	<a href="#">Black Panther: Wakanda Forever</a>	\$250,000,000	\$453,829,060	\$853,985,546
22 Apr 22, 2016	<a href="#">Captain America: Civil War</a>	\$250,000,000	\$408,084,349	\$1,151,899,586
23 May 3, 2023	<a href="#">Guardians of the Galaxy Vol 3</a>	\$250,000,000	\$358,995,815	\$845,468,744
24 Jul 1, 2022	<a href="#">Thor: Love and Thunder</a>	\$250,000,000	\$343,256,830	\$760,928,081
25 Jul 15, 2009	<a href="#">Harry Potter and the Half-Blood Prince</a>	\$250,000,000	\$302,089,278	\$929,411,069
26 May 24, 2023	<a href="#">The Little Mermaid</a>	\$250,000,000	\$298,172,056	\$568,345,048
27 Dec 12, 2013	<a href="#">The Hobbit: The Desolation of Smaug</a>	\$250,000,000	\$258,241,522	\$959,358,436
28 Dec 10, 2014	<a href="#">The Hobbit: The Battle of the Five Armies</a>	\$250,000,000	\$255,119,788	\$940,323,039
29 Apr 7, 2017	<a href="#">The Fate of the Furious</a>	\$250,000,000	\$225,764,765	\$1,235,534,014
30 Sep 29, 2021	<a href="#">No Time to Die</a>	\$250,000,000	\$160,891,007	\$759,959,662
31 Oct 6, 2015	<a href="#">Spectre</a>	\$245,000,000	\$200,074,175	\$879,077,344
32 Dec 17, 2009	<a href="#">Avatar</a>	\$237,000,000	\$785,221,649	\$2,923,706,026
33 Jun 28, 2006	<a href="#">Superman Returns</a>	\$232,000,000	\$200,120,000	\$391,081,192
34 Jul 19, 2012	<a href="#">The Dark Knight Rises</a>	\$230,000,000	\$448,139,099	\$1,082,228,107
35 May 23, 2017	<a href="#">Pirates of the Caribbean: Dead Men Tell No Tales</a>	\$230,000,000	\$172,558,876	\$794,861,794
36 Nov 14, 2008	<a href="#">Quantum of Solace</a>	\$230,000,000	\$169,368,427	\$591,692,078
37 Apr 25, 2012	<a href="#">The Avengers</a>	\$225,000,000	\$623,357,910	\$1,515,100,211
38 Jul 6, 2006	<a href="#">Pirates of the Caribbean: Dead Man's Chest</a>	\$225,000,000	\$423,315,812	\$1,066,179,725

# WEB SCRAPING & API CALLS





# CREATOR RATING SYSTEM

Actor	Instances	Total_Production_Budget	Total_IMDB_Rating	Average_IMDB_Rating
Robert De Niro	51	\$ 2,309,200,000.00	339	6.647058824
Samuel L. Jackson	48	\$ 2,658,756,044.00	314.9	6.560416667
Nicolas Cage	44	\$ 2,314,203,077.00	269.1	6.115909091
Bruce Willis	42	\$ 2,231,243,106.00	268.8	6.4
Matt Damon	41	\$ 2,494,200,000.00	288.5	7.036585366
Mark Wahlberg	40	\$ 2,276,800,000.00	258.7	6.4675
Johnny Depp	39	\$ 3,256,200,000.00	262	6.717948718
Tom Hanks	38	\$ 2,475,800,000.00	277.6	7.305263158
Keanu Reeves	36	\$ 1,796,100,000.00	236.4	6.566666667
Brad Pitt	35	\$ 2,432,400,000.00	252.3	7.208571429
Denzel Washington	35	\$ 1,713,500,000.00	246	7.028571429
Morgan Freeman	35	\$ 1,856,800,000.00	233.9	6.682857143
Liam Neeson	35	\$ 1,639,400,000.00	226.3	6.465714286
Nicole Kidman	35	\$ 1,480,800,000.00	223.7	6.391428571
Tom Cruise	34	\$ 2,875,000,000.00	240.6	7.076470588
Julia Roberts	34	\$ 1,536,000,000.00	217.1	6.385294118
Dwayne Johnson	34	\$ 3,103,850,000.00	213.4	6.276470588
Matthew McConaughey	33	\$ 1,452,984,000.00	221.9	6.724242424
Owen Wilson	33	\$ 2,031,750,000.00	204.3	6.190909091
George Clooney	32	\$ 1,734,500,000.00	213.5	6.671875



# MOVIE CONTENT RATING SYSTEMS

```
161222
Not Rated 23575
R 15267
PG-13 5581
PG 3896
...
K-14 1
K 1
R-18+ 1
R(A) 1
M/PG 1
Name: OMDB_Rated, Length: 95, dtype: int64
```

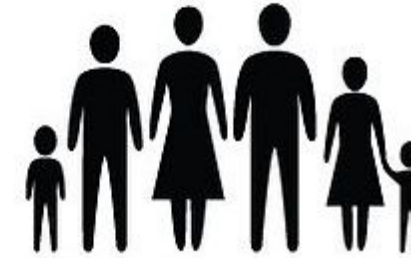
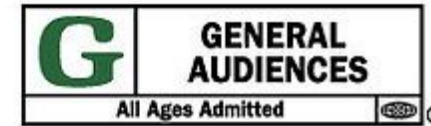
→

```
unknown 187954
adults 19128
teens 8055
kids 6796
Name: OMDB_Rated, dtype: int64
```

## THE FILM RATING SYSTEM

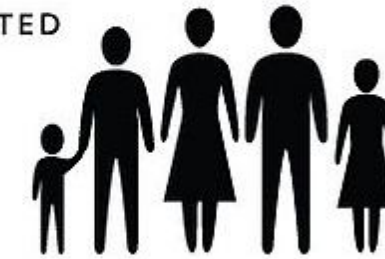
EMPOWERING FAMILIES TO MAKE INFORMED MOVIE CHOICES

### GENERAL AUDIENCES



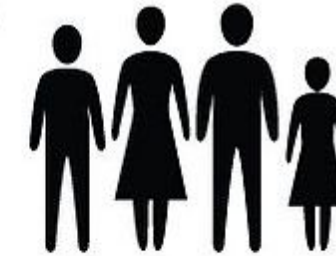
Nothing that would offend parents for viewing by children.

### PARENTAL GUIDANCE SUGGESTED



Parents urged to give "parental guidance." May contain some material parents might not like for their young children.

### PARENTS STRONGLY CAUTIONED



Parents are urged to be cautious. Some material may be inappropriate for pre-teenagers.

### RESTRICTED



Contains some adult material. Parents are urged to learn more about the film before taking their young children with them.

### NO ONE 17 AND UNDER ADMITTED



Clearly adult. Children are not admitted.

FILMRATINGS.COM

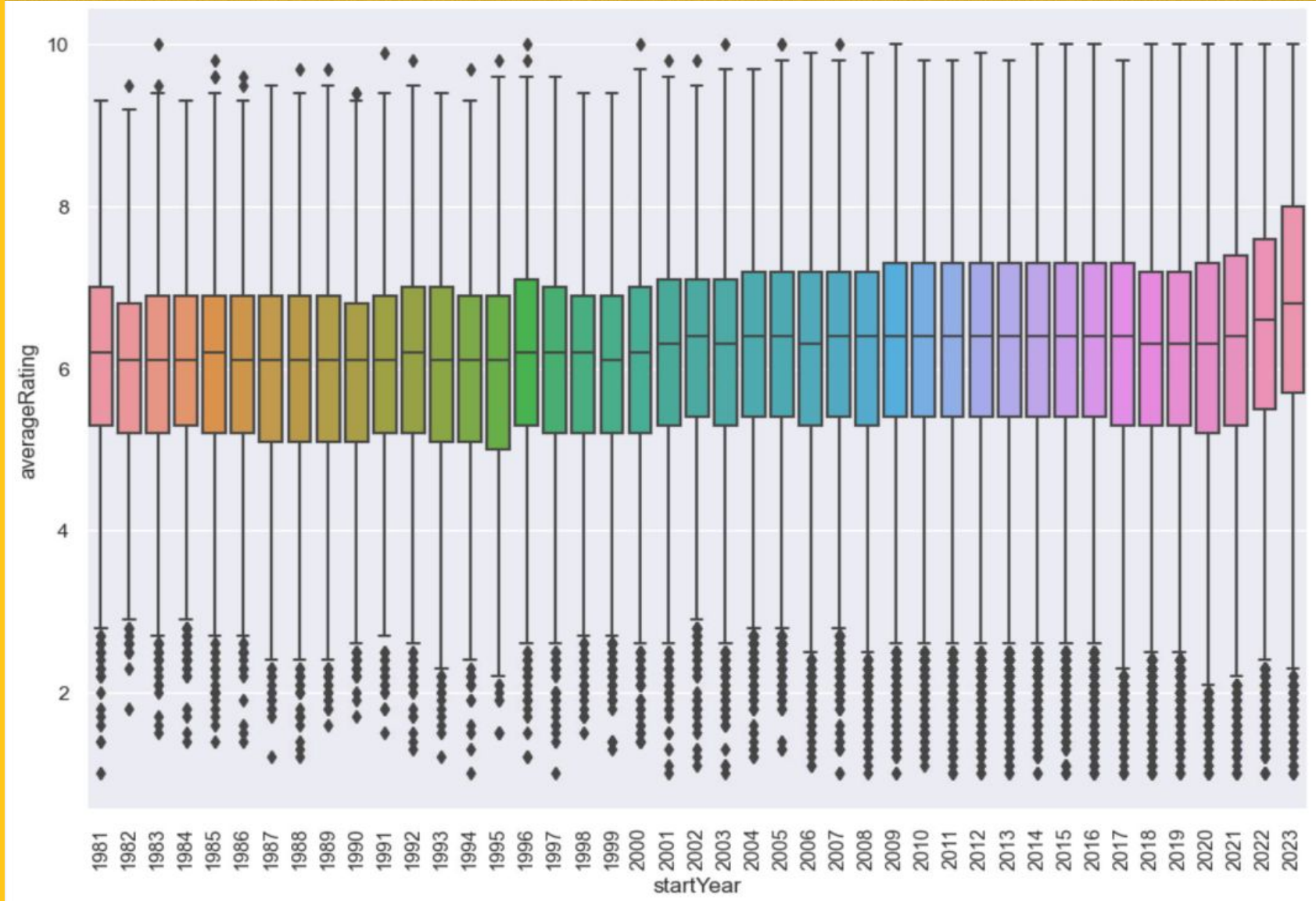




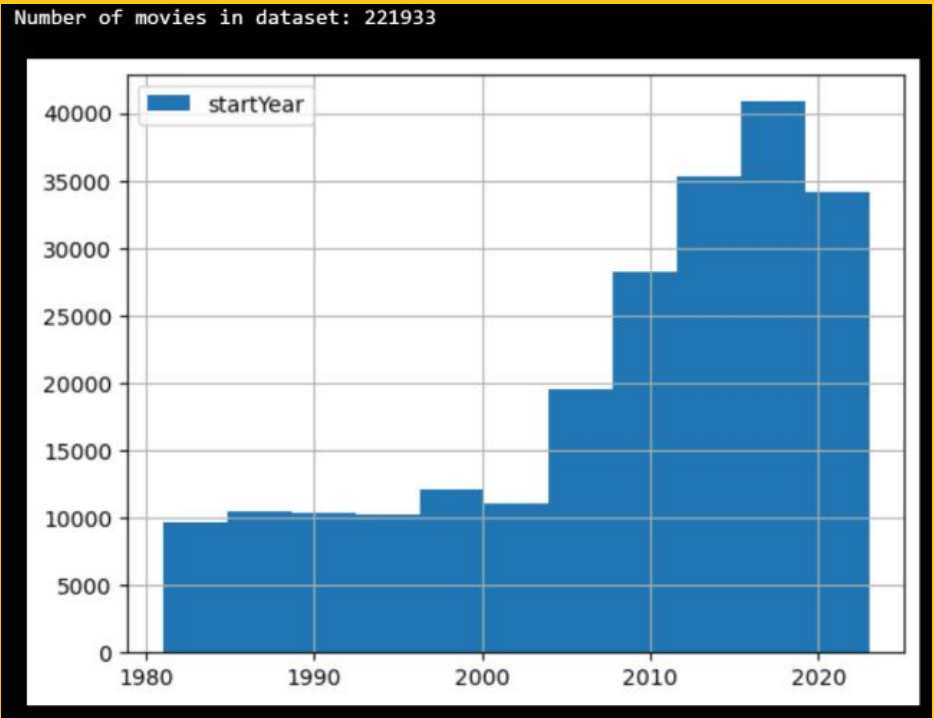


# IMDB Data & Python Cleaning

10.3 Million rows just for titles. - Reduced to 221k  
Over 4.5 GB of data available.



Name	Date modified	Type	Size
principals.tsv	11/15/2023 7:12 AM	TSV File	2,542,371 KB
titlebasics.tsv	11/14/2023 7:09 AM	TSV File	863,773 KB
namecleaned.tsv	11/15/2023 10:42 PM	TSV File	813,794 KB
name.tsv	11/14/2023 7:12 AM	TSV File	772,520 KB
crew.tsv	11/15/2023 7:08 AM	TSV File	333,125 KB
ratings.tsv	11/15/2023 7:07 AM	TSV File	23,193 KB
movies_raw_imdb.csv	11/20/2023 7:26 PM	Microsoft Excel Com...	16,335 KB
movies_formatted_imdb.csv	11/16/2023 9:35 PM	Microsoft Excel Com...	1 KB



	tconst	titleType	primaryTitle
0	tt0000001	short	Carmencita
1	tt0000002	short	Le clown et ses chiens
2	tt0000003	short	Pauvre Pierrot
3	tt0000004	short	Un bon bock
4	tt0000005	short	Blacksmith Scene
...	...	...	...
10316935	tt9916826	tvEpisode	Episode #3.10
10316936	tt9916830	tvEpisode	Episode #2.6
10316937	tt9916832	tvEpisode	Episode #3.11
10316938	tt9916834	tvEpisode	Episode #3.12
10316939	tt9916836	tvEpisode	Episode #3.14

10316940 rows × 9 columns

	tconst	averageRating	numVotes	titleType
0	tt0013274	6.8	59	movie
1	tt0015414	5.2	16	movie
2	tt0015724	6.1	27	movie
3	tt0035423	6.4	87965	movie
4	tt0036606	6.5	345	movie
...	...	...	...	...
221918	tt9914644	8.3	143	movie
221919	tt9914662	7.4	6	movie
221920	tt9914942	6.6	191	movie
221921	tt9914972	7.1	290	movie
221922	tt9915790	7.2	40	movie

221923 rows × 10 columns





# IMDB Data - SQL Cleaning

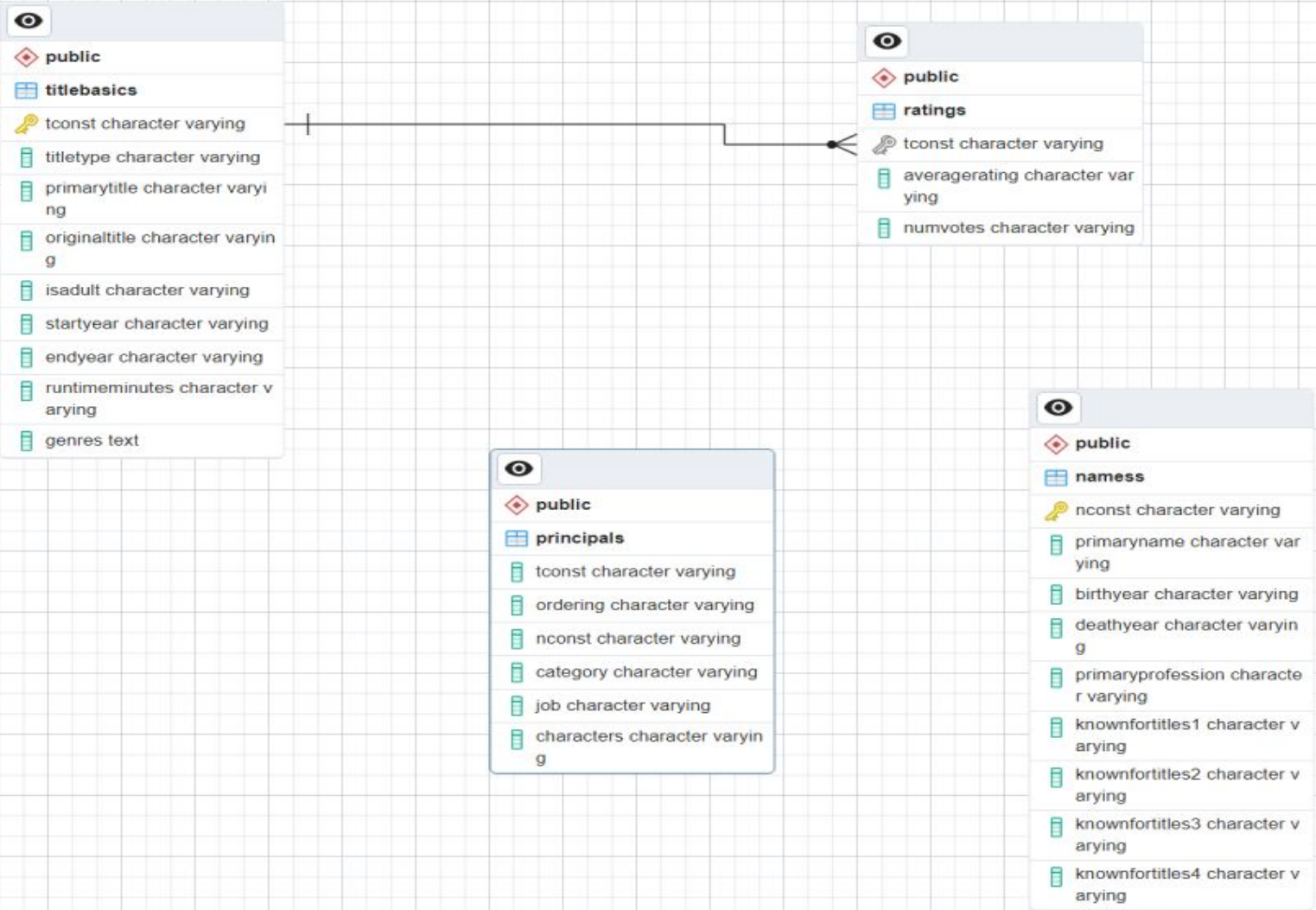
Created ERD - Reduced Original Data to 224k rows

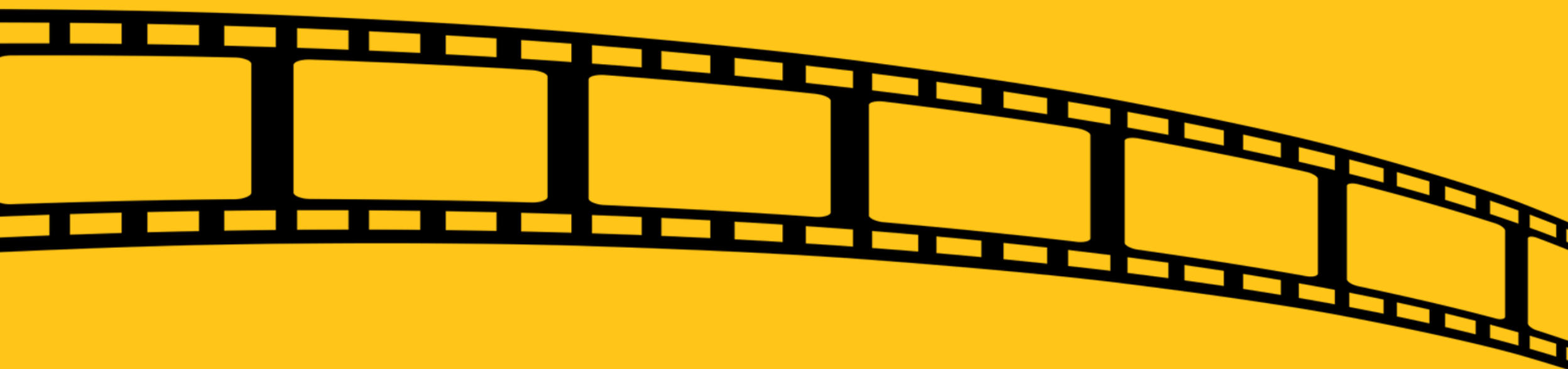
	count bigint	
1	224198	

IMDB_SQL_Ratings.csv	Microsoft Excel Com...	4,097 KB
IMDB_SQL_Titles.csv	Microsoft Excel Com...	31,709 KB

	tconst [PK] character varying	titletype character varying	primarytitle character varying	originaltitle character varying	isadult character varying	startyear character varying
1	tt10189136	movie	Paruvathin Vasalile	Paruvathin Vasalile	0	1980
2	tt10189396	movie	Kadhal Kadhal Kadhal	Kadhal Kadhal Kadhal	0	1980
3	tt10185818	movie	The Case of the Missing Consciousness	The Case of the Missing Consciousness	0	1980
4	tt10189112	movie	Uchakattam	Uchakattam	0	1980
5	tt10153656	movie	I Am Wanda	I Am Wanda	0	1980
6	tt10148002	movie	Naan Sootiya Malar	Naan Sootiya Malar	0	1980
7	tt10183584	movie	Panam Pen Paasam	Panam Pen Paasam	0	1980
8	tt10078478	movie	Allari Baava	Allari Baava	0	1980
9	tt10137318	movie	Phool Aur Pathor	Phool Aur Pathor	0	1980
10	tt10192656	movie	How to Become a Detective	How to Become a Detective	0	1980

	tconst character varying	averagerating character varying	numvotes character varying
1	tt0013274	6.8	59
2	tt0015414	5.2	16
3	tt0015724	6.1	27
4	tt0035423	6.4	87965
5	tt0036606	6.5	345
6	tt0038086	7.0	27
7	tt0038687	7.4	1980
8	tt0057461	4.5	23
9	tt0059325	6.5	257
10	tt0059900	6.4	38

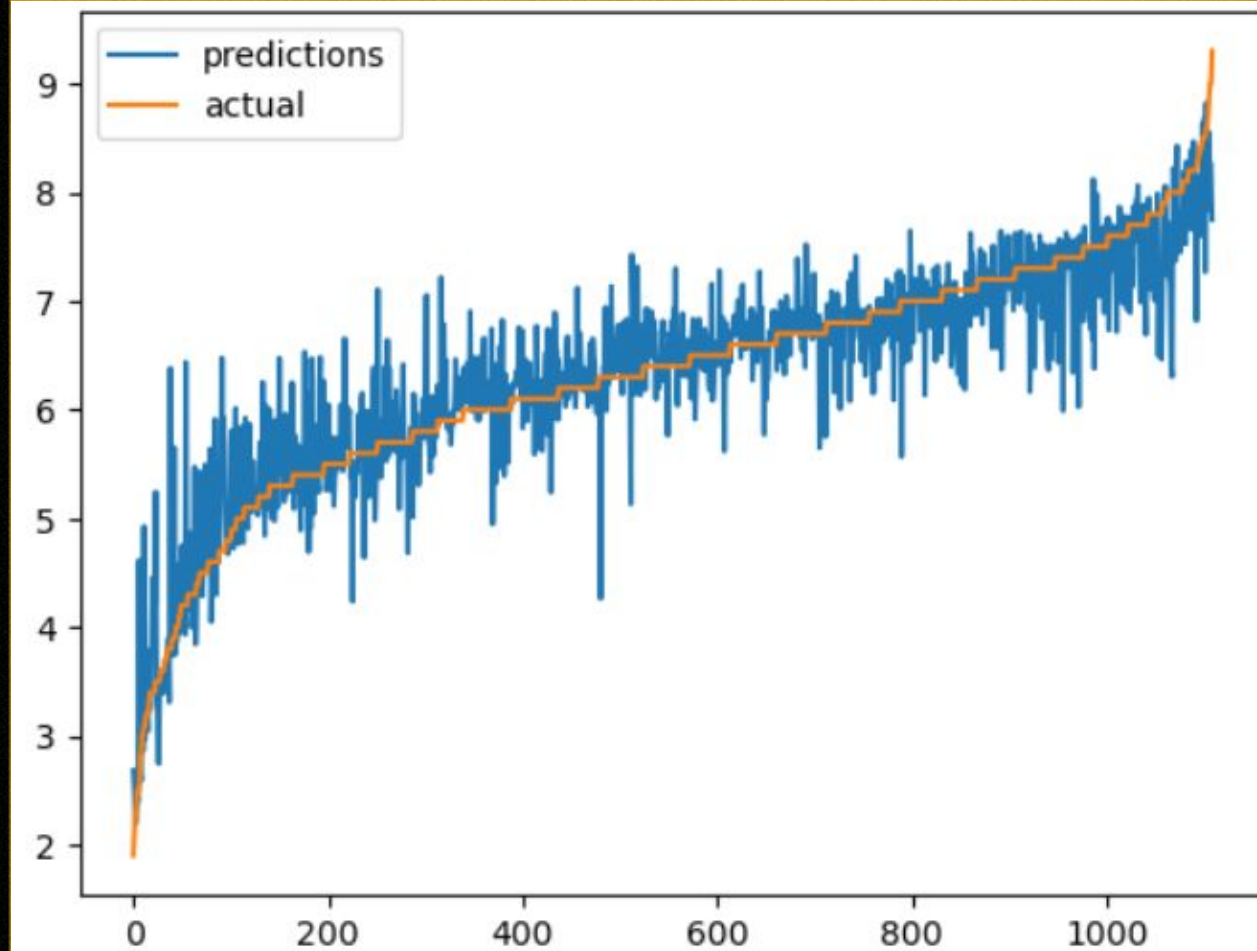




02

# MACHINE LEARNING MODELS





### Target

Movie Writer & Director predicting  
IMDB Ratings

### Actors

Average Writer Rating, Average  
Director Rating, Production Budget,  
Movie Runtime, Release Month

# MOVIE MAKERS

## PRODUCTION COMPANY

KERAS TUNER

## GENRE

relu activation,  
rmsprop optimization

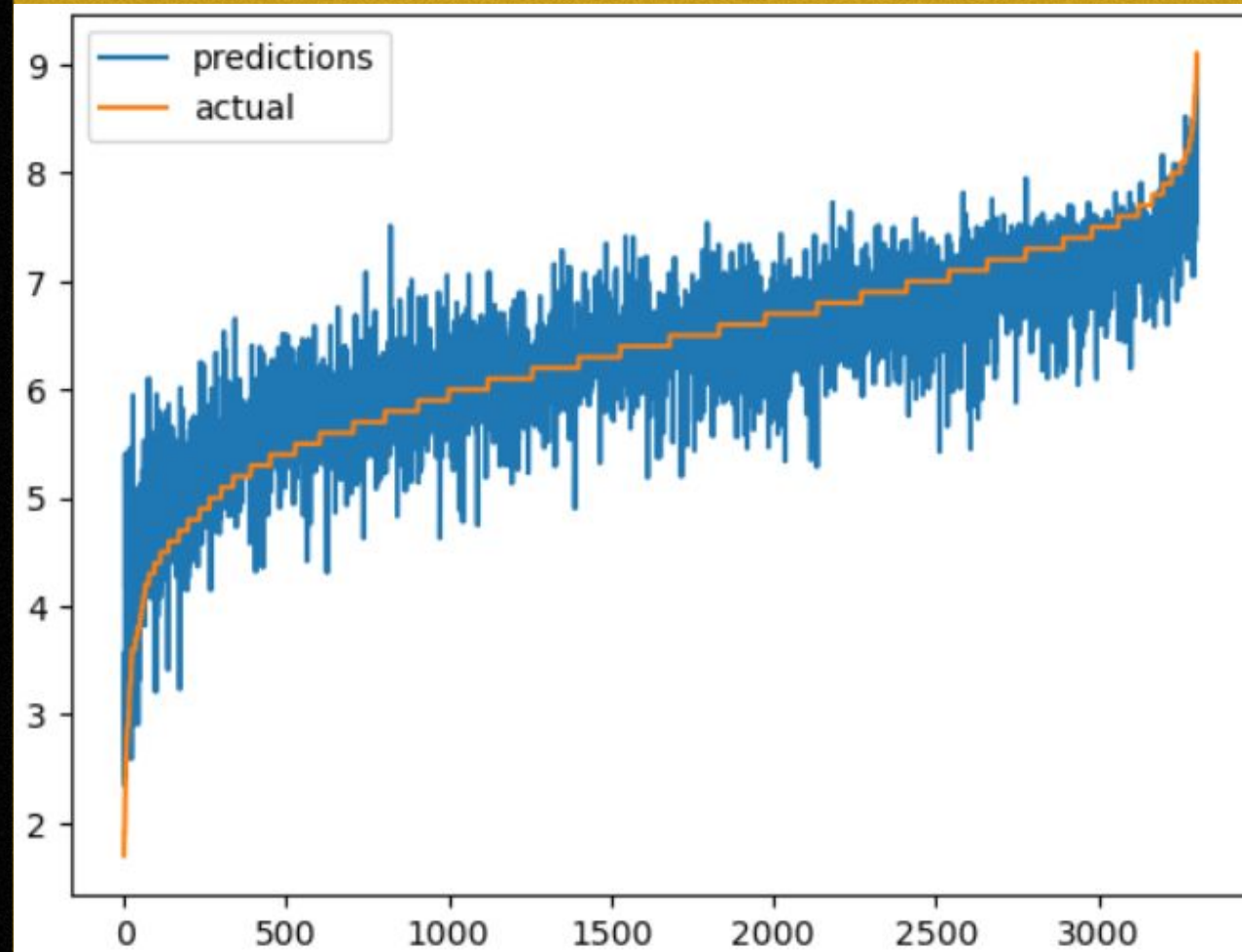
## AWARDS

Best MSE 0.19

## SYNOPSIS

- 5,542 data rows
- Other actors such as MPAA rating and typical Box Office bring-in by Writer/Director detracted from performance.
- Random Forest Productions were the top competition (MSE 0.21)





### Target

General Public predicting IMDB Ratings

### Actors

Average Writer Rating, Average Director Rating, Movie Runtime, Release Month

# MOVIE VIEWERS

## PRODUCTION COMPANY

RANDOM FOREST

## GENRE

250 estimators,  
random\_state 42

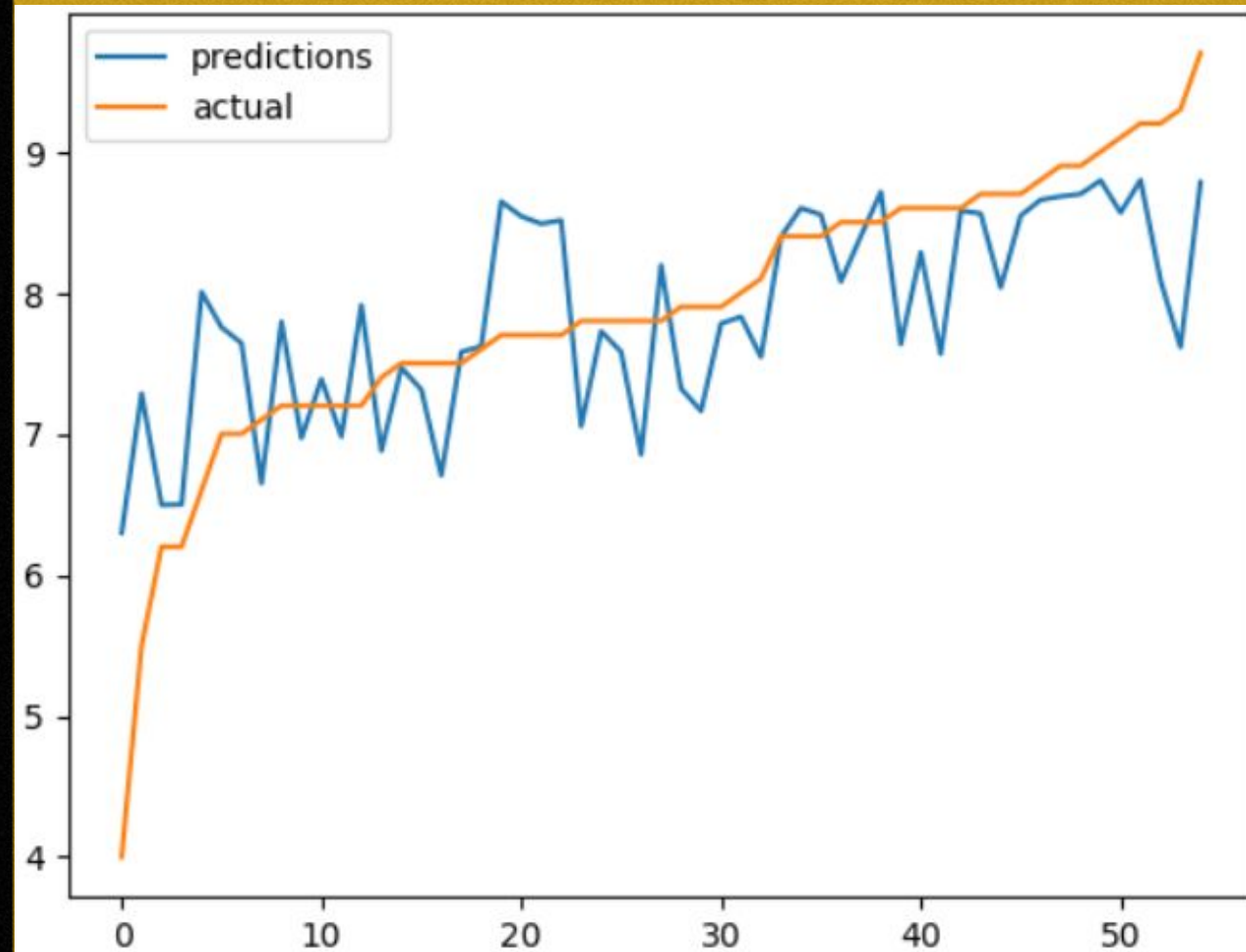
## AWARDS

Best MSE 0.21

## SYNOPSIS

- Up to 221,933 data rows
- Tried different actors and focusing on specific genres
- Keras Tuner Productions' RELU and RMSPROP subsidiaries were the top competition at MSE 0.23
- Best Random Forest != Best Keras Tuner





### Target

Single User Movie Rating 0-10

### Actors

Average IMDB Rating, Critic Rating, Movie Runtime, Production Budget, Genres

# MOVIE USER

## PRODUCTION COMPANY

RANDOM FOREST

## GENRE

75 estimators,  
random\_state 42

## AWARDS

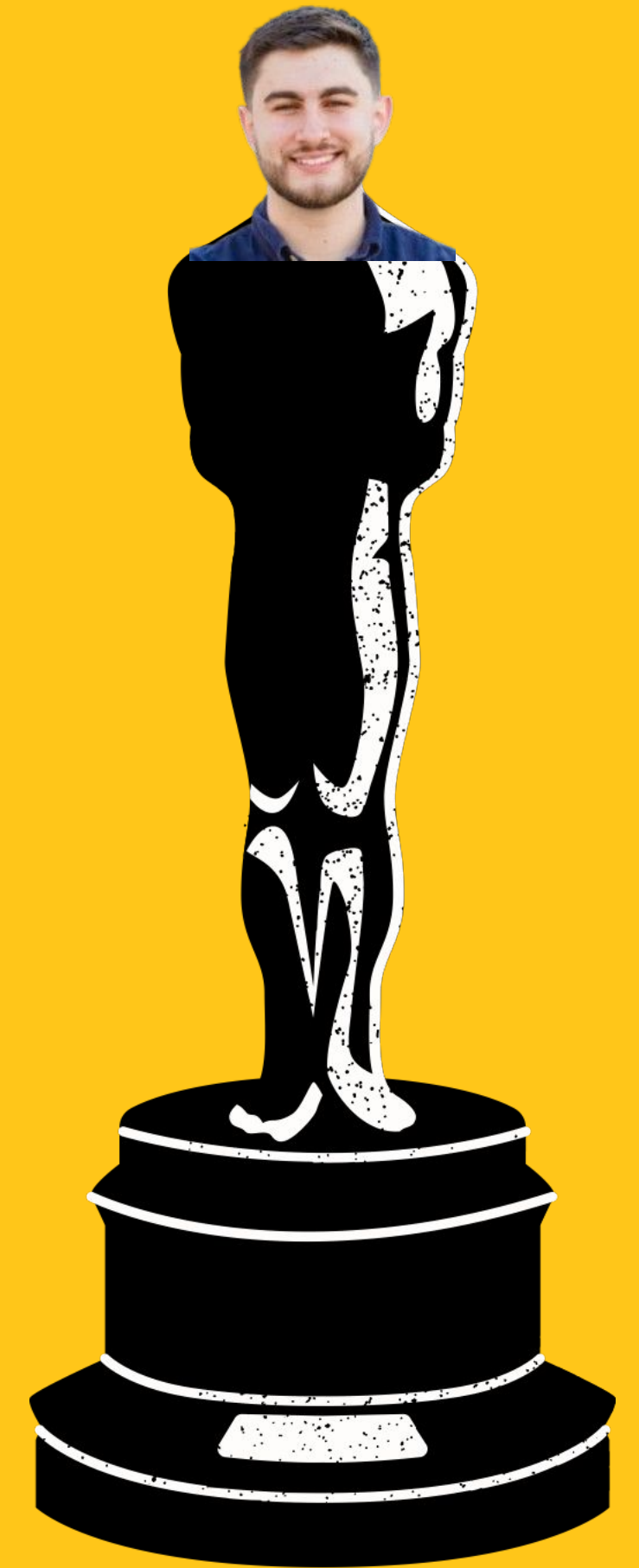
Best MSE 0.51

## SYNOPSIS

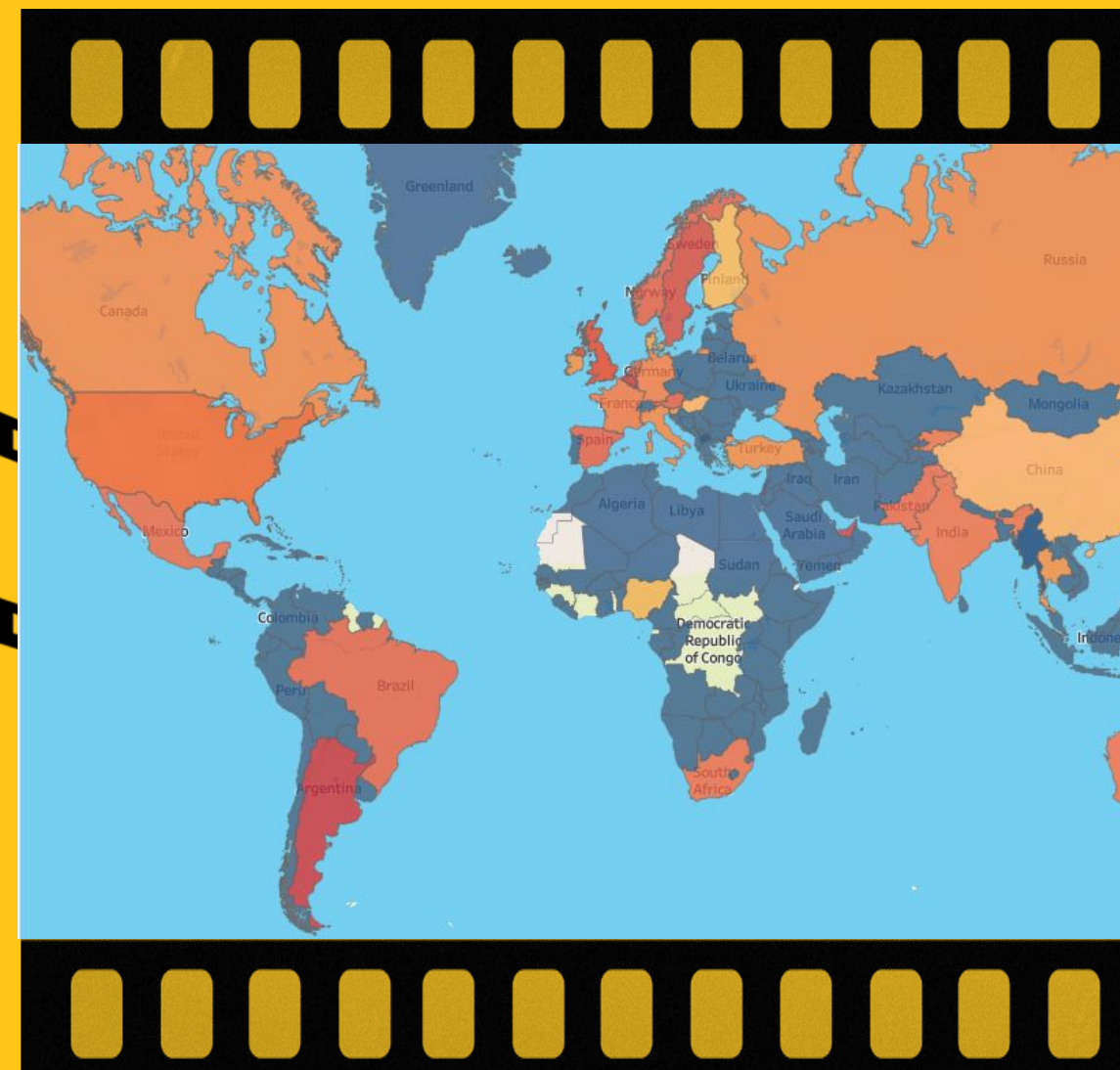
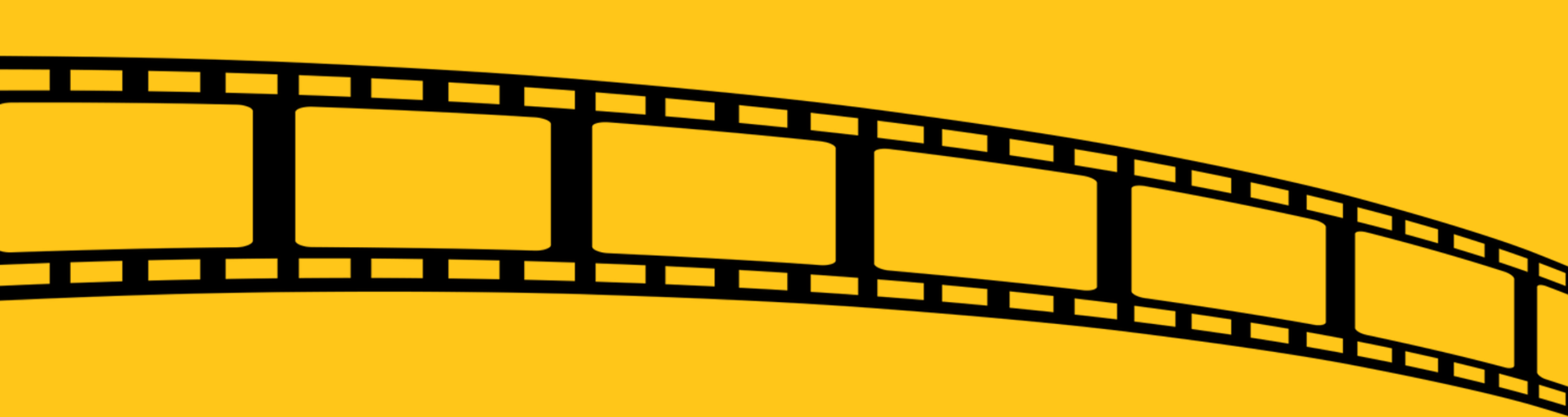
- 272 data rows
- Other actors such as movie age rating, release month, and Writer/Director/Actor ratings detracted from performance.
- Keras Tuner Productions' RELU and SGD subsidiaries were the top competition but an order of magnitude worse (MSE 48)

# Machine Learning

BEST SMALL DATA	Random Forest
BEST LARGE DATA	Keras Tuner - RELU/TANH, RMSPROP
#1 DATA IMPROVEMENT	PRODUCTION BUDGET AVAILABILITY
# 2 DATA IMPROVEMENT	PEOPLE RATINGS
# 3 DATA IMPROVEMENT	CRITIC SCORES



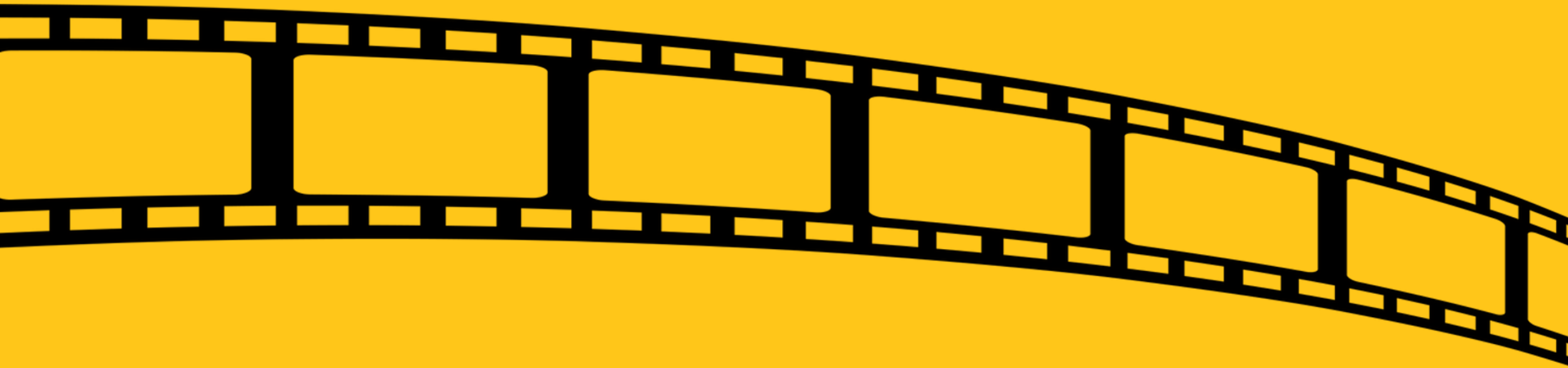




03

# **TABLEAU STORY**

# **DEMO**



04

# MOVIE USER WEBPAGES DEMO





**THANK  
YOU!**

# RESOURCE PAGE

Alejandro Davila

Ashley Yakopec-Ley

Bryant Griessel

Christopher Hornung

Rebekah Aldrich

[https://github.com/vt-bekah/Predicting\\_Movie\\_Ratings](https://github.com/vt-bekah/Predicting_Movie_Ratings)

