

# Film Studio - Initial Recommendations Analysis for Microsoft

Noah-John Hizon  
11/15/21

# Introduction

What makes a film “successful”?

- Financial Standpoint
  - Budget vs Profit
  - Marketing, Streaming, Box Office
- Audience Satisfaction
  - Reviews
  - Genre-Defining

# Outline

- Business Problem
- Data
- Methodology
- Results
- Conclusion

# Business Problem



Information from online databases will be aggregated/summarized in order to find the following:

- Optimal directors to work on films
- Optimal actors/actresses to cast in films
- Optimal genres to pursue, to maximize profitability

## Data - Names/Identifiers

- Primary Name (corresponds to 'nconst' in dataset)
- Primary Title (corresponds to 'tconst' in dataset)
- Category
  - Actor, director, writer
- Genre of film
  - Multiple, in some cases
- Release Date
  - Analysis limited to 1980

# Data - User Ratings

## SOURCE: IMDB

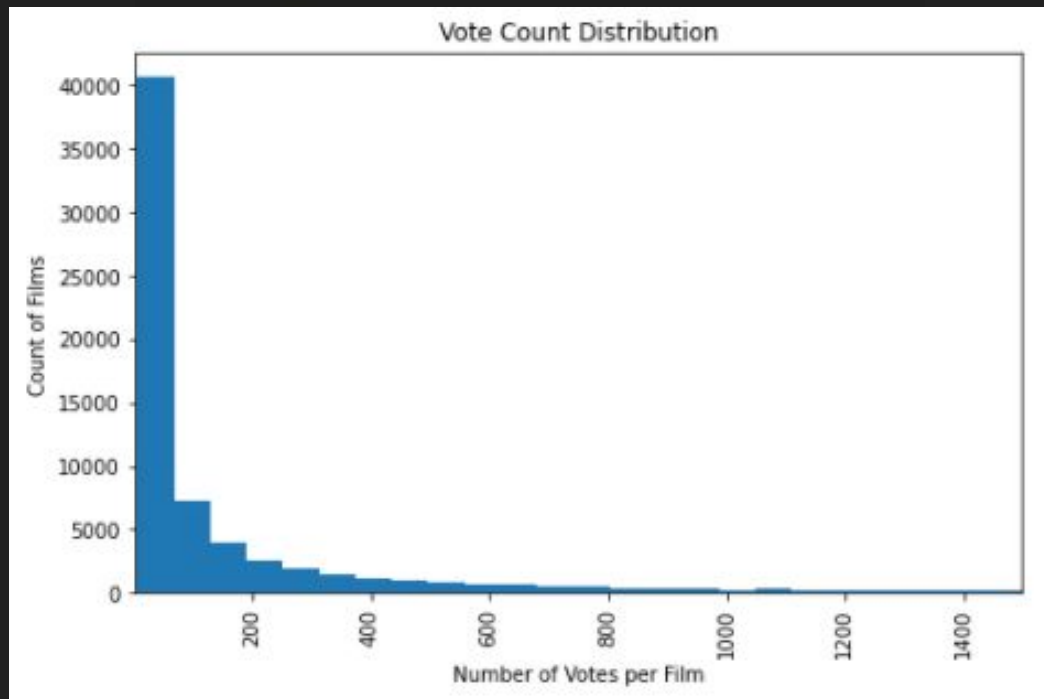
- Scale 1-10
- Represented as an aggregated average rating for each film
- ~25% of the data has less than 365 votes
- Data for films with >200 votes were kept to maximize inclusion without keeping fringe records

IMDb RATING

★ **6.9/10**  
88K

YOUR RATING

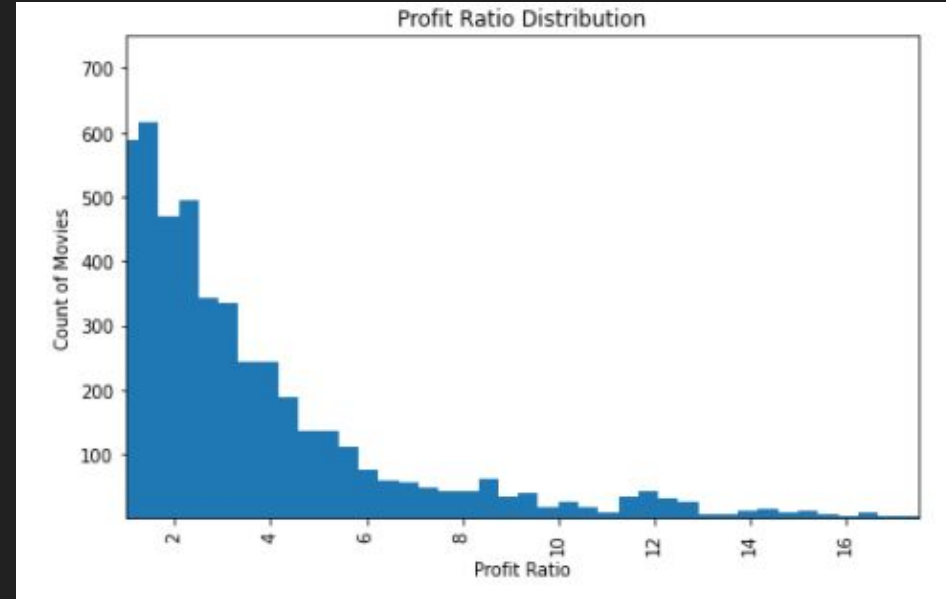
☆ **Rate**



# Data - Financial Data

## SOURCE: The Numbers

- **Production Budget**
  - The amount (in \$) that was allocated to the film's creation
- **Worldwide Gross**
  - The total amount (in \$) that was made by the film across the globe, domestic country included
- **Profit Ratio**
  - Used to directly identify success of a movie based on both of its financial metrics
  - Any movie w/ratio  $\geq 1$  made its money back



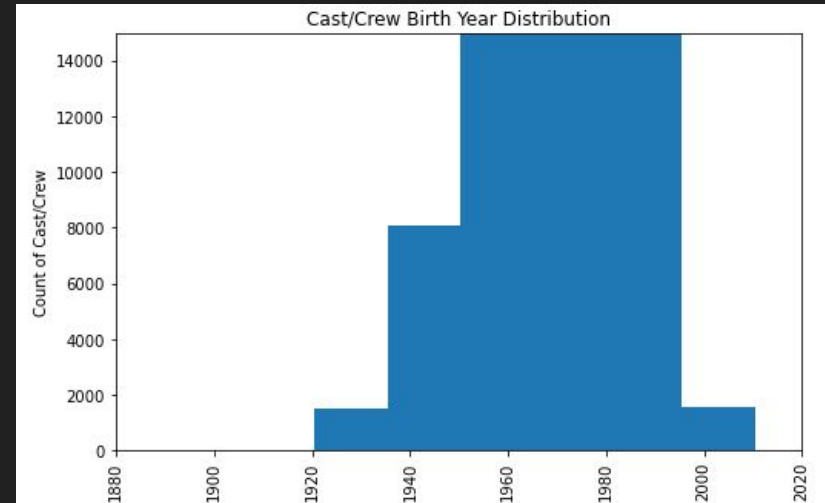
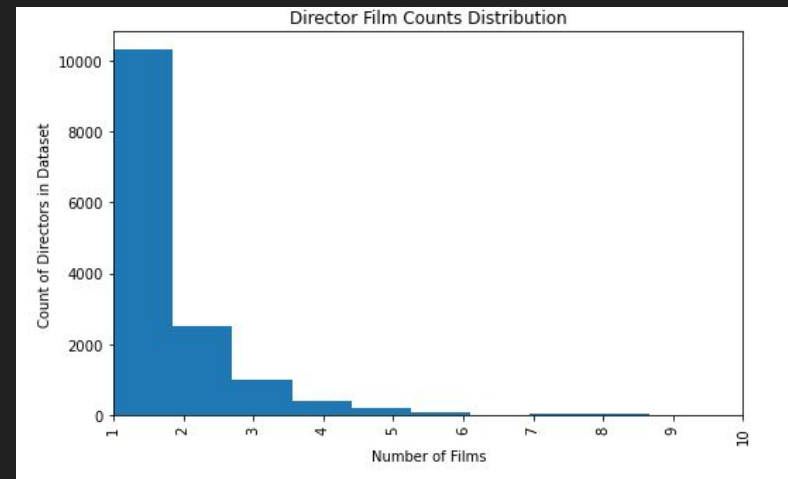
Profit Ratio = (Worldwide Gross)/(Production Budget)

# Methodology - Directors

- Ratings, binned
- Total films with available ratings  $\geq 3$
- Average profit ratio
- Average budget
- Birth year  $> 1936$

## SORTED BY:

- Count of films, descending
- Average profit ratio, descending



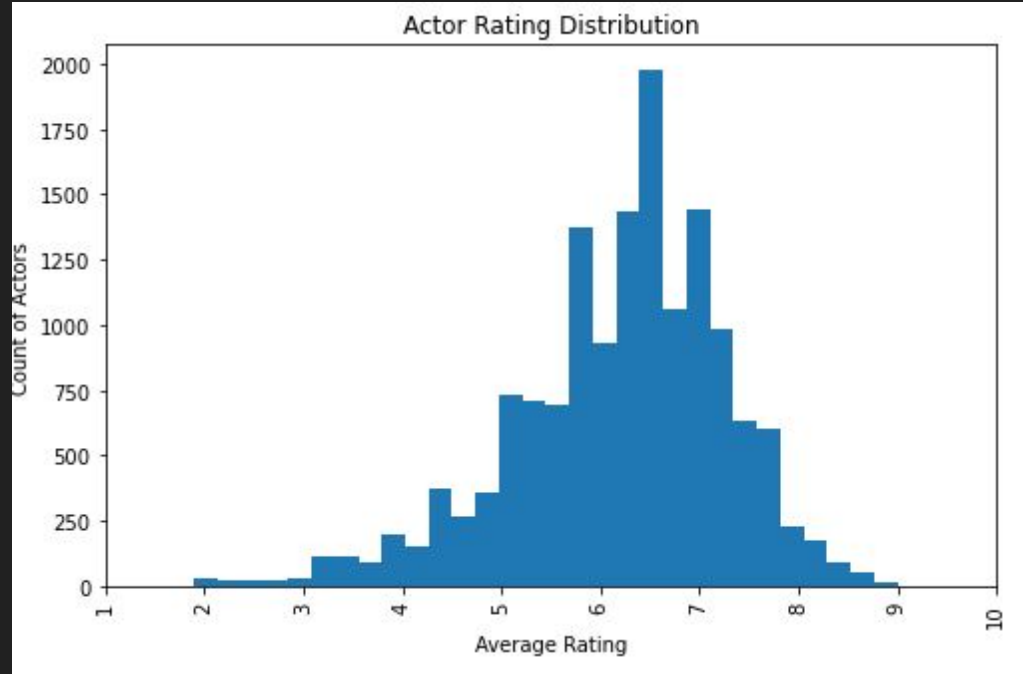


# Methodology - Actors

- Ratings, binned
- Total films with available ratings  $\geq 3$
- Average profit ratio
- Average budget
- Birth year  $> 1936$

SORTED BY:

- Count of films, descending
- Production budget, descending



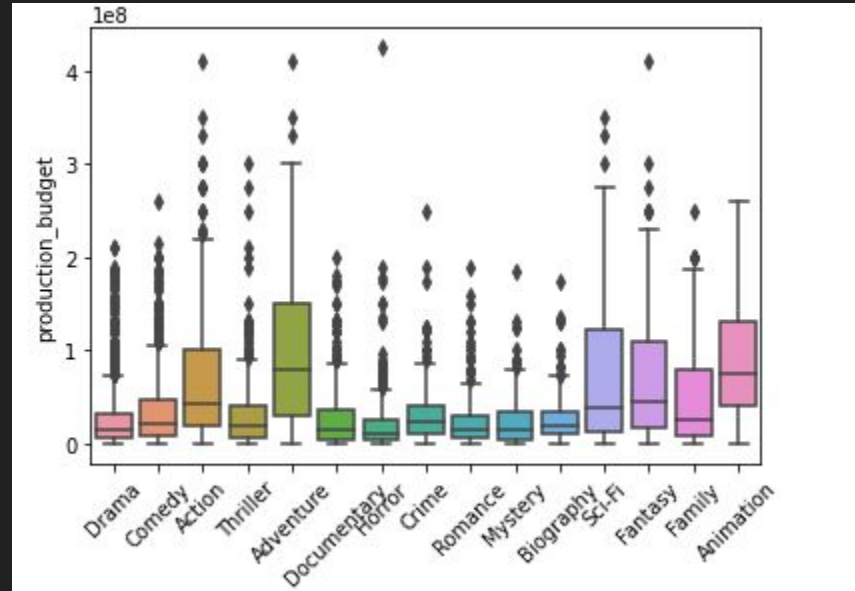
# Methodology - Genres

- Each film/genre combination stored as a unique value
  - 1 movie may be classified by 3 genres (e.g. Horror, Adventure, and Drama)
- Genres with total film count > 100

SORTED BY:

- Average profit ratio
- Average budget

**Film Budget Distribution by Genre**



# Results - Optimal Directors

Top 15 listed by ranked order

- Red selections are primarily profit oriented
- Green selections, while limited, are chosen for user rating

primary_name	average_profitratio	count_of_films	mean_productionbudget	count_filmsratings	average_imdbrating
Tim Story	6.235346	12	1.754274e+07	7	6.371429
Chris Renaud	8.487083	8	7.500000e+07	5	6.920000
Don Hall	3.531010	8	1.237500e+08	3	7.533333
Seth Gordon	1.788564	8	1.806250e+07	4	6.150000
Pierre Coffin	11.829330	7	7.357143e+07	3	7.133333
Steven Soderbergh	5.958500	7	2.152857e+07	9	6.600000
David Gordon Green	4.759385	7	1.971429e+07	8	6.225000
Peter Farrelly	3.999512	7	4.271429e+07	4	6.225000
Steven Spielberg	3.181121	7	9.214286e+07	7	7.228571
Ridley Scott	2.699194	7	1.148571e+08	7	6.585714
Ariel Schulman	24.501394	6	1.000000e+07	4	5.600000
Henry Joost	24.501394	6	1.000000e+07	4	5.600000
Leslie Small	8.524418	6	4.416667e+06	3	6.700000
Christopher Miller	6.361456	6	5.066667e+07	3	7.333333
Phil Lord	6.361456	6	5.066667e+07	3	7.333333

# Results - Optimal Actors

Top 20 listed by ranked order

- Red selections are primarily profit oriented, actors that can operate on high-budget movies
- Green selections, are chosen for user rating

	mean_productionbudget	count_of_films	averageIMDBrating	binnedrating
primary_name				
Mark Wahlberg	6.282857e+07	33	6.527273	7.0
Dwayne Johnson	9.335455e+07	33	6.333333	6.5
Robert De Niro	3.224848e+07	33	6.330303	6.5
Joel Edgerton	4.978065e+07	31	6.783871	7.0
Steve Carell	4.106452e+07	31	6.716129	7.0
Jennifer Lawrence	9.947667e+07	30	6.880000	7.0
Matthew McConaughey	4.363333e+07	30	6.856667	7.0
Liam Neeson	6.476333e+07	30	6.423333	6.5
Channing Tatum	5.826333e+07	30	6.466667	6.5
Michael Fassbender	7.872414e+07	29	6.444828	6.5
Amy Adams	6.250000e+07	28	7.128571	7.5
Emily Blunt	6.396571e+07	28	6.607143	7.0
Owen Wilson	6.714286e+07	28	6.171429	6.5
Kevin Hart	3.613333e+07	28	6.300000	6.5
Nicole Kidman	2.106667e+07	28	6.342857	6.5
Nicolas Cage	4.476429e+07	28	5.460714	5.5
Matt Damon	7.112593e+07	27	6.696296	7.0
John Goodman	3.631481e+07	27	6.877778	7.0
Anna Kendrick	3.039037e+07	27	6.518519	7.0
Johnny Depp	1.420444e+08	27	6.148148	6.5

# Results - Optimal Genres

Top genres listed by ranked order

- Red selections are chosen for average profit ratio
- Green selections are chosen for potential audience reception

genresplit	mean_profitratio	binnedratio	mean_productionbudget	count_of_films	meanIMDBrating
Horror	7.745018	8.0	2.753423e+07	396	5.407885
Mystery	7.442447	7.5	2.731016e+07	234	6.050000
Thriller	5.455271	5.5	3.654652e+07	562	5.886480
Family	5.101057	5.5	5.273897e+07	180	6.195370
Documentary	4.992545	5.0	3.129773e+07	428	6.975510
Animation	4.943798	5.0	7.429697e+07	138	6.533613
Romance	4.095652	5.0	2.776343e+07	353	6.292164
Fantasy	3.949041	4.0	6.676907e+07	181	6.050318
Biography	3.722467	4.0	2.907384e+07	233	6.958434
Drama	3.569462	4.0	3.320581e+07	1690	6.463312
Comedy	3.451865	4.0	4.403527e+07	817	6.163651
Sci-Fi	3.264354	4.0	6.890724e+07	217	6.281461
Adventure	3.236764	4.0	8.339076e+07	466	6.395923
Action	2.971202	4.0	6.922477e+07	656	6.155292
Crime	2.298395	4.0	3.070973e+07	379	6.247604

# Conclusion

Microsoft should pursue directors, actors, and genres from a variety of budget ranges using the corresponding lists as guidelines

## LIMITATIONS

- User rating is inherently subjective/imperfect metric
- Wide variety of review aggregation metrics can be found across the internet, impossible to get a true holistic, unbiased view.

# Conclusion

## NEXT STEPS

- Finding each director's/actor's best performing genre
- Aggregate review data from other sources should also be considered
- Other roles such as editors/writers should be identified as well

# Thank You!

Email: [noahiz917@gmail.com](mailto:noahiz917@gmail.com)

Github: noahiz917

LinkedIn: [linkedin.com/in/noah-john-hizon/](https://www.linkedin.com/in/noah-john-hizon/)