

Film Studio - Initial Recommendations Analysis for Microsoft

Noah-John Hizon
11/15/21

Outline

- Business Problem
- Data
- Methodology
- Results
- Conclusion

Introduction

What makes a film “successful”?

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

Business Problem



Information from online databases will be aggregated/summarized in order to find the following factors of successful movie studios:

- 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

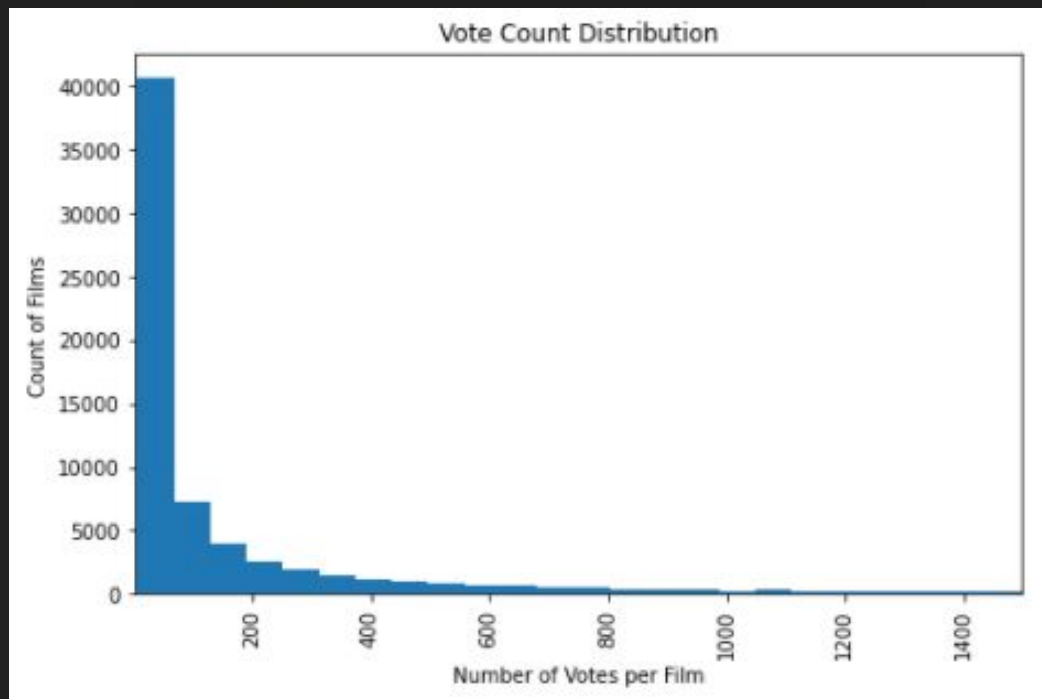
- Scale 1-10
- Represented as an aggregated average rating for each film
- ~75% of the data has less than 365 votes
- Data for films with >50 votes were kept to maximize inclusion without keeping fringe records
- 50% of the film rating data was kept

IMDb RATING

★ 6.9/10
88K

YOUR RATING

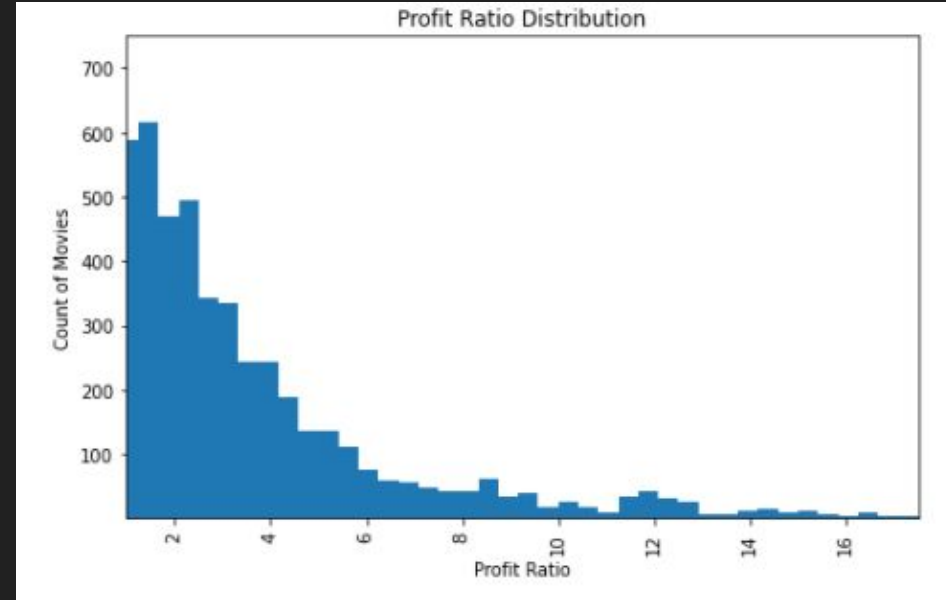
☆ Rate



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 ≥ 1 made its money back



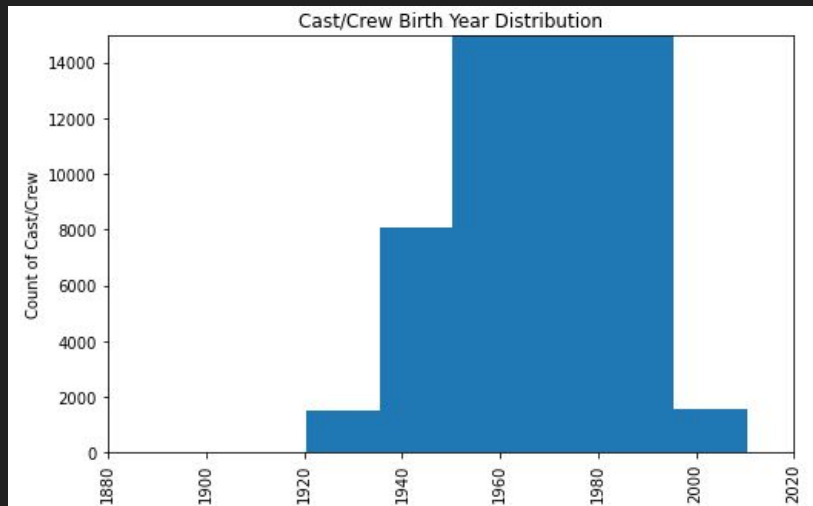
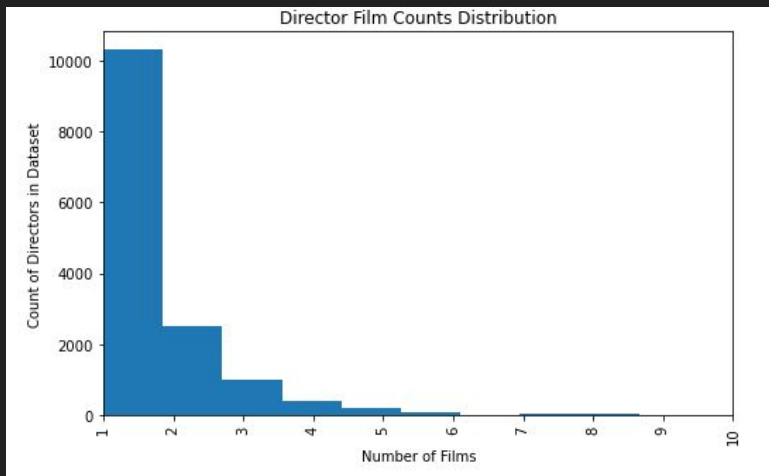
Profit Ratio = (Worldwide Gross)/(Production Budget)

Methodology - Directors

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

SORTED BY:

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

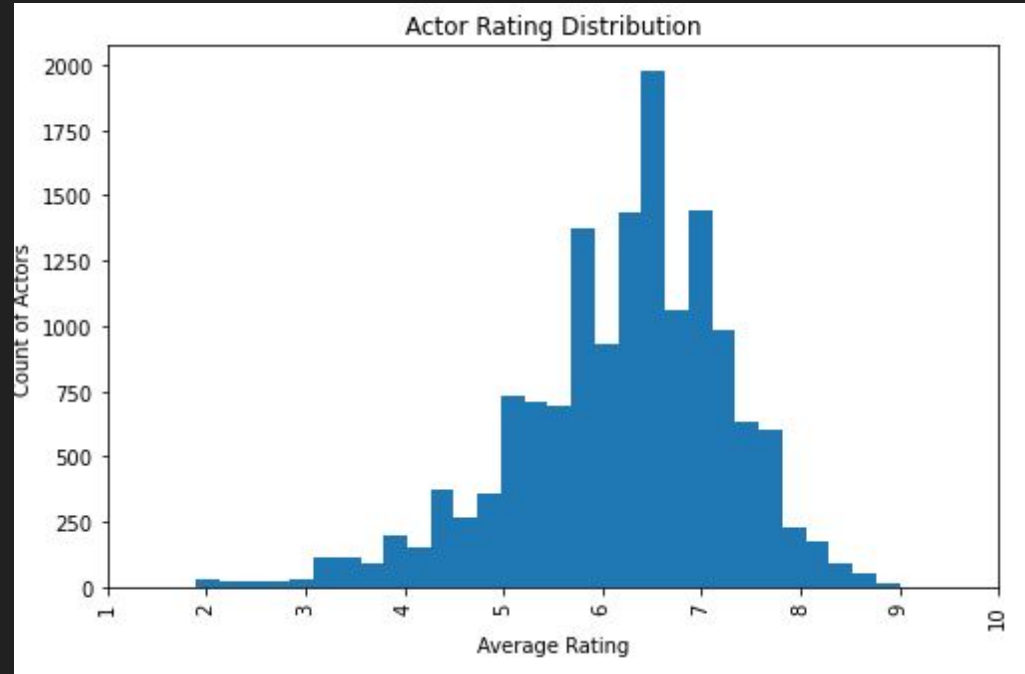


Methodology - Actors

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

SORTED BY:

- Count of films, descending
- Average production budget, descending



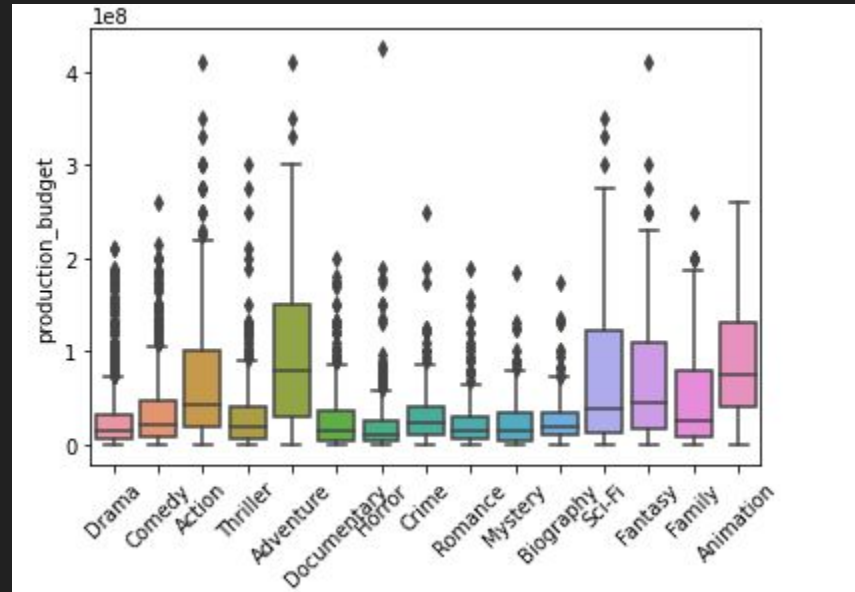
Methodology - Genres

- 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

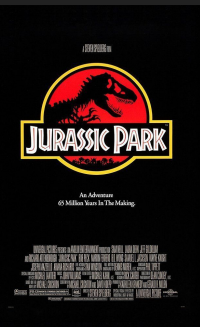
Film Budget Distribution by Genre



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	6.800377	11	1.641027e+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
Pierre Coffin	11.829330	7	7.357143e+07	3	7.133333
Peter Farrelly	3.999512	7	4.271429e+07	4	6.225000
Steven Spielberg	3.181121	7	9.214286e+07	7	7.228571
Ariel Schulman	24.501394	6	1.000000e+07	4	5.600000
Henry Joost	24.501394	6	1.000000e+07	5	5.980000
Leslie Small	8.524418	6	4.416667e+06	4	6.800000
Steven Soderbergh	6.914333	6	2.508333e+07	9	6.600000
Phil Lord	6.361456	6	5.066667e+07	3	7.333333
Christopher Miller	6.361456	6	5.066667e+07	3	7.333333
Andrew Erwin	6.243048	6	7.000000e+06	4	6.550000
Jon Erwin	6.243048	6	7.000000e+06	5	6.700000
Anthony Russo	5.196585	6	2.400000e+08	4	8.225000

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



primary_name	mean_productionbudget	count_of_films	averageIMDBrating	binnedrating
Dwayne Johnson	9.932258e+07	31	6.322581	6.5
Jennifer Lawrence	8.158214e+07	28	6.942857	7.0
Mark Wahlberg	6.985714e+07	28	6.567857	7.0
Kevin Hart	3.782143e+07	28	6.300000	6.5
Liam Neeson	6.996296e+07	27	6.359259	6.5
Channing Tatum	6.605769e+07	26	6.523077	7.0
Amy Adams	6.583333e+07	24	7.241667	7.5
Emily Blunt	6.146000e+07	24	6.658333	7.0
Steve Carell	4.704348e+07	23	6.747826	7.0
Owen Wilson	7.604348e+07	23	6.182609	6.5
Rachel McAdams	5.286364e+07	22	7.345455	7.5
Seth Rogen	3.605652e+07	22	6.609091	7.0
Bradley Cooper	5.400000e+07	21	7.176190	7.5
Ryan Reynolds	8.347560e+07	21	6.828571	7.0
Michael Fassbender	6.757143e+07	21	6.661905	7.0
Joel Edgerton	5.481905e+07	21	6.842857	7.0
Robert De Niro	3.700000e+07	21	6.514286	7.0
Johnny Depp	1.612000e+08	21	6.371429	6.5
Cate Blanchett	1.138800e+08	20	7.160000	7.5
Tom Hanks	6.650000e+07	20	7.090000	7.5

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

- Good amount of imperfect/ineligible/incomplete data amongst the selected databases (filtration)
- 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

Github: noahiz917

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