AN ANALYSIS OF THEATRICAL MOVIE RELEASES:

A SUMMARY OF PREDICTORS IN THE PRODUCTION OF MOVIES WITH HIGH RETURN ON INVESTMENT

05/03/21

Matt LeGro

Presentation Roadmap

- Outline the business objective
- Describe the data source and preparation methods
- Summarize results and discuss interpretations
- Provide actionable insights

Business Objective

- Global Box Office Revenue hit a new record in 2019 at $42.5 billion and has been growing every year
- Netflix is reported to invest $15 billion on new content; Prime Video and Apple TV+ at $6 billion; Hulu, Disney+ and HBO Max are all coming in around $3 billion
- Enable Microsoft to join the media industry as a production company and grow shareholder value
- Determine features of successful theatrical movie releases to guide Microsoft in producing films that will generate profit

Microsoft

Data Preparation

- The data used in the analysis comes from The Numbers (TN), The Movie Database (TMDb), and the Internet Movie Database (IMDb)
- Being interested in profitable movies, as many records as possible containing budget and revenue were collected, but to determine features of successful films, we need other attributes as well
- Our initial TN dataset has both financial fields, but lacks other categorical information
- Our TMDb and IMDb datasets have plenty of descriptive information, but not financial items

Data Preparation

- To improve the TMDb dataset, it was found that information about individual movies could be easily retrieved using a unique movie ID via their web API. Thus, for each movie in the initial TMDb set, additional information including budget, revenue, and importantly, the IDMb unique ID was retrieved
- To improve the TN dataset, the unique IMDb was assigned from the IMDb tables based on the title and release year of movies in the TN dataset. Once an IMDb ID was assigned, it could easily be related to the other tables in which IMDb is an identifier
- Finally, the expanded TN and TMDB datasets were combined and checked for duplicate records for a final collection of movies with the required budgetary information and maximal categorical attributes

Data Preparation

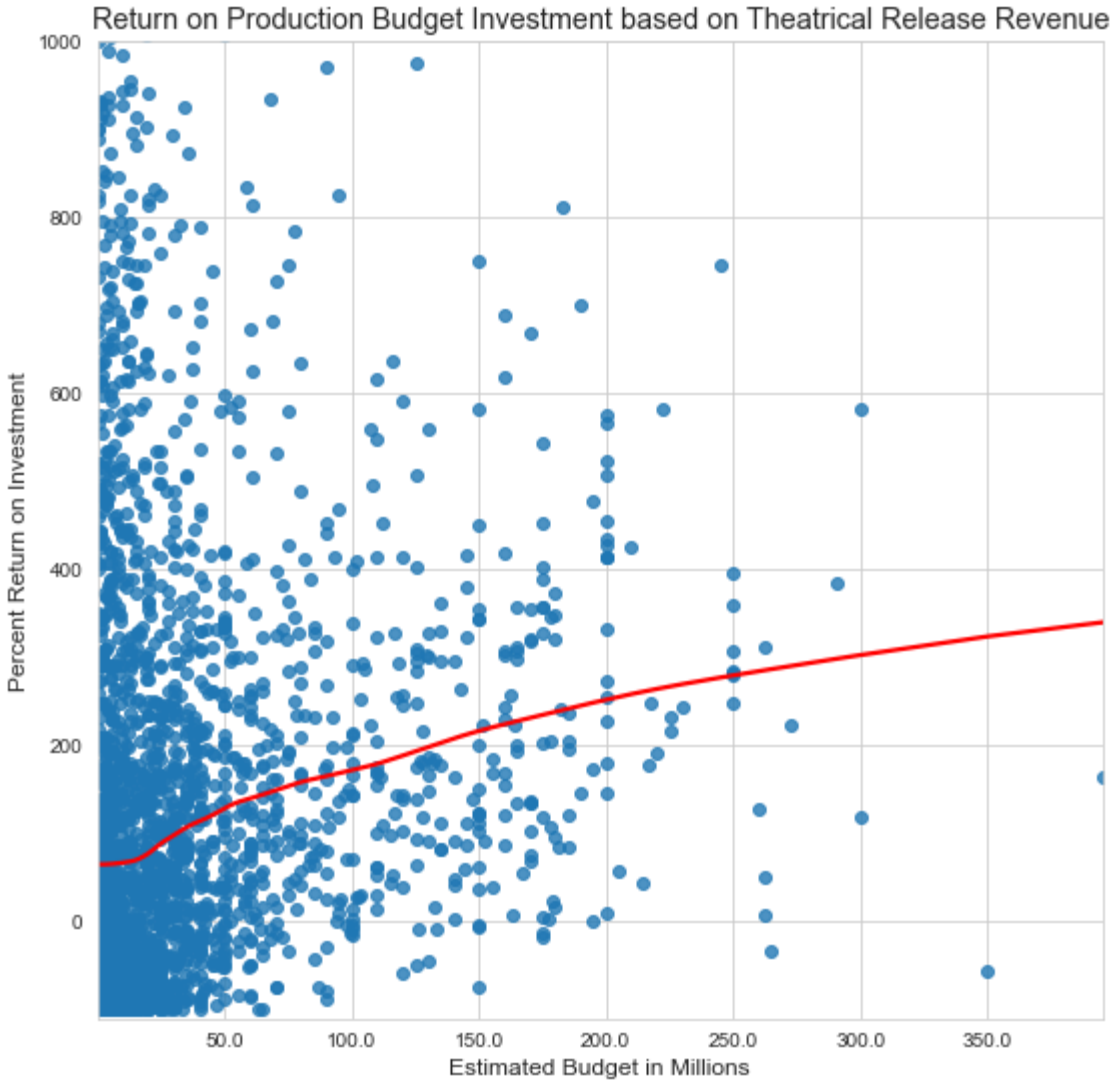
tconst	Title	Genres	runtime_ minutes	vote_ average	vote_count	budget_est	worldwide _gross	release_ week_day	release_ day	release_ month	release_ year
tt0892769	How to Train Your Dragon	['Fantasy', 'Adventure', 'Animation', 'Family']	98	8.095082	618909	165000000.0	494874875.5	4	26	3	2010
tt1228705	Iron Man 2	['Adventure', 'Action', 'Science Fiction']	124	6.996308	670058	185000000.0	622544860.0	4	7	5	2010

- Note: For duplicate rows when merging TN and TMDb, if records existed in both, all numeric columns were averaged and categorical columns were kept from the TMDb dataset
- In analysis, additional aggregate categories such as Return on Investment were included (ROI), plus binning categories were defined, such as ranges for budget, and ROI percentage

Questions for Review

- How much should be spent on production?
- What type of movie is most often high in return on production budget?
- Once the movie is produced, when should it be released?

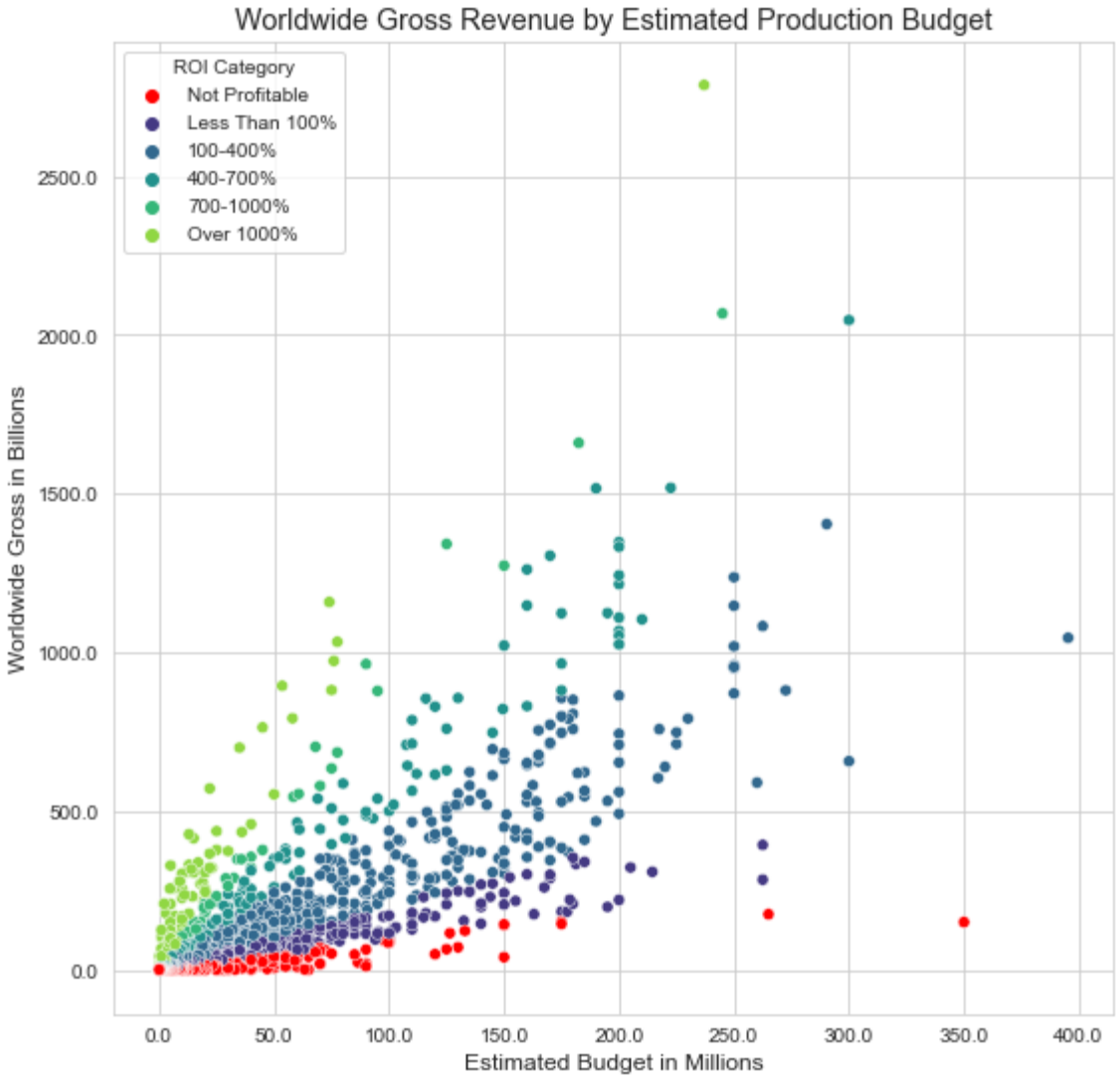
What should be spent on production?

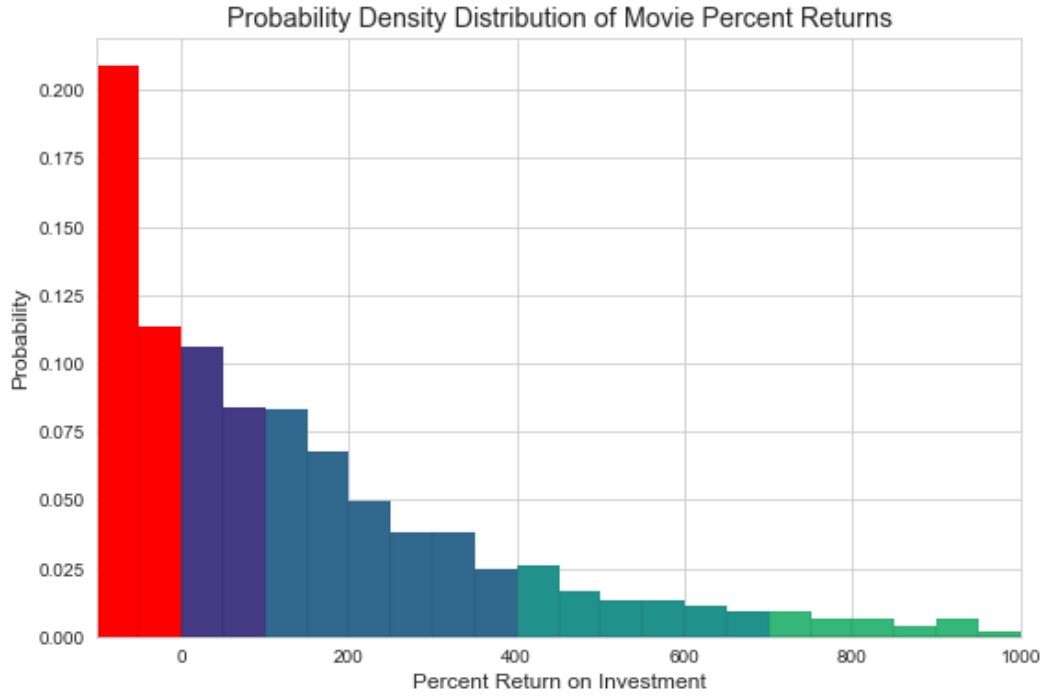


- Red line is a locally weighted linear regression (LOWESS)
- Except for budgets less than ~20mil, return on investment trends upward with increased budget
- Ultra low budget region has very wide spread
- What subset of movies should we look at?

Microsoft

What should be spent on production?

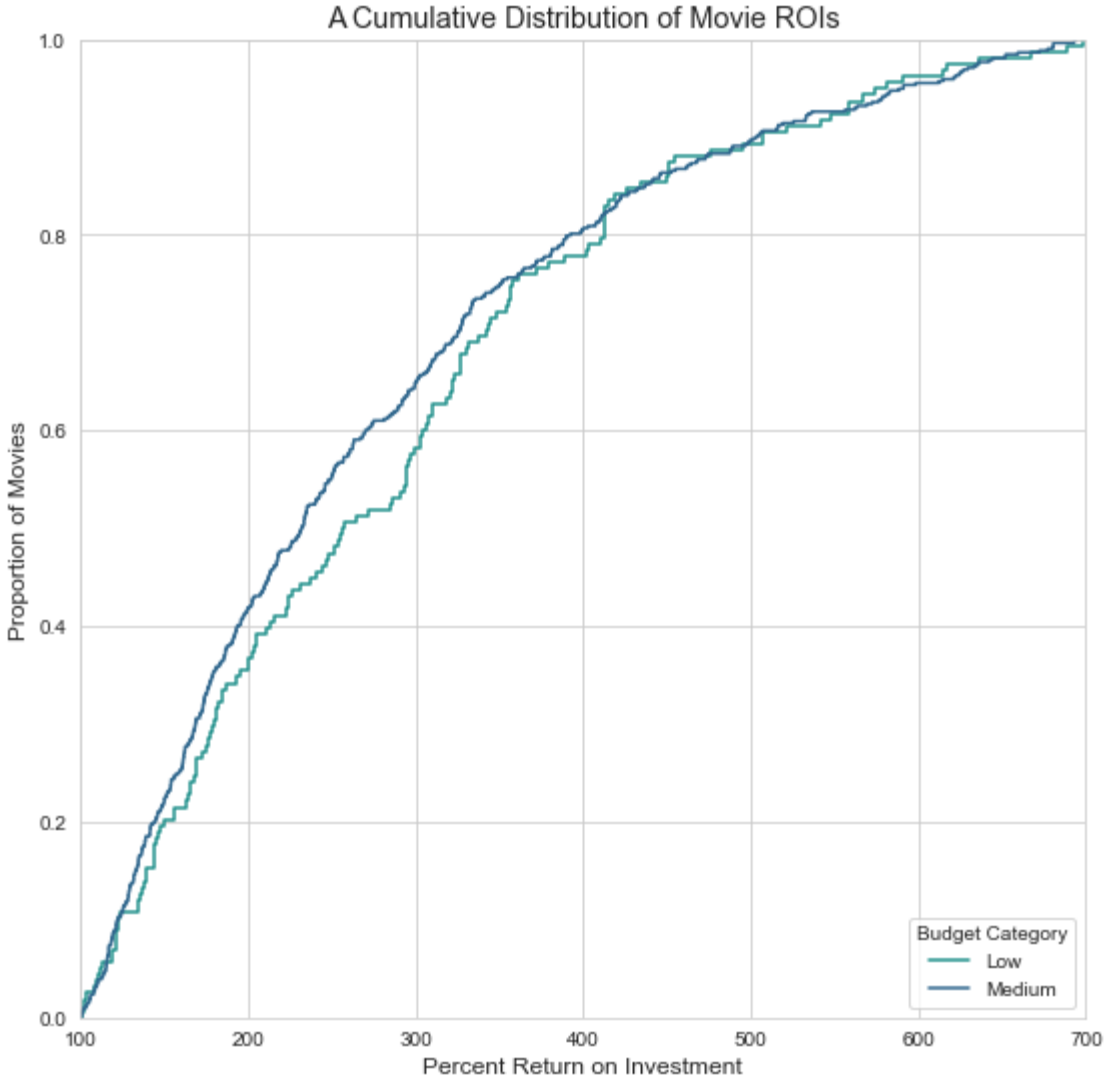


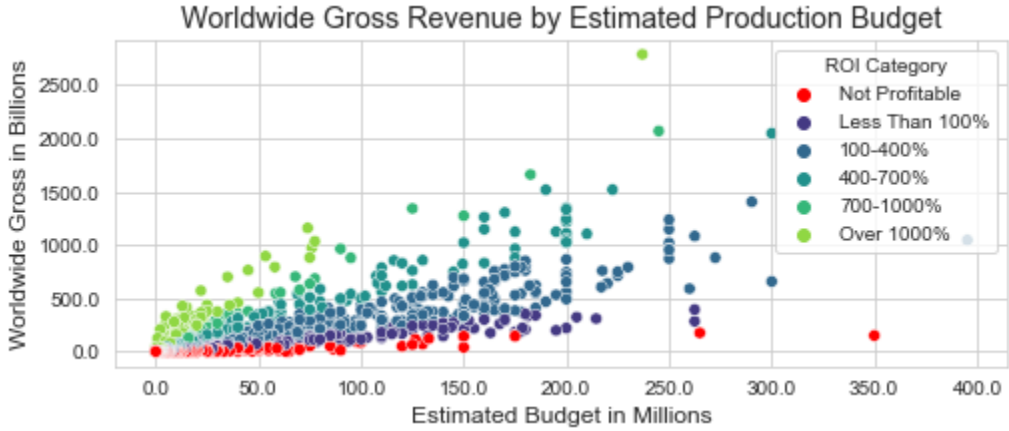


- Note: x-axis of distribution plot has been limited to a Percent ROI of 1000
- Movies with exceptional ROI tend to have minimal budget; compare Gross Revenue of higher budget movie with lower ROI
- Choose to look at movies from the middle of the distribution for sample size in finding patterns

Microsoft

What should be spent on production?

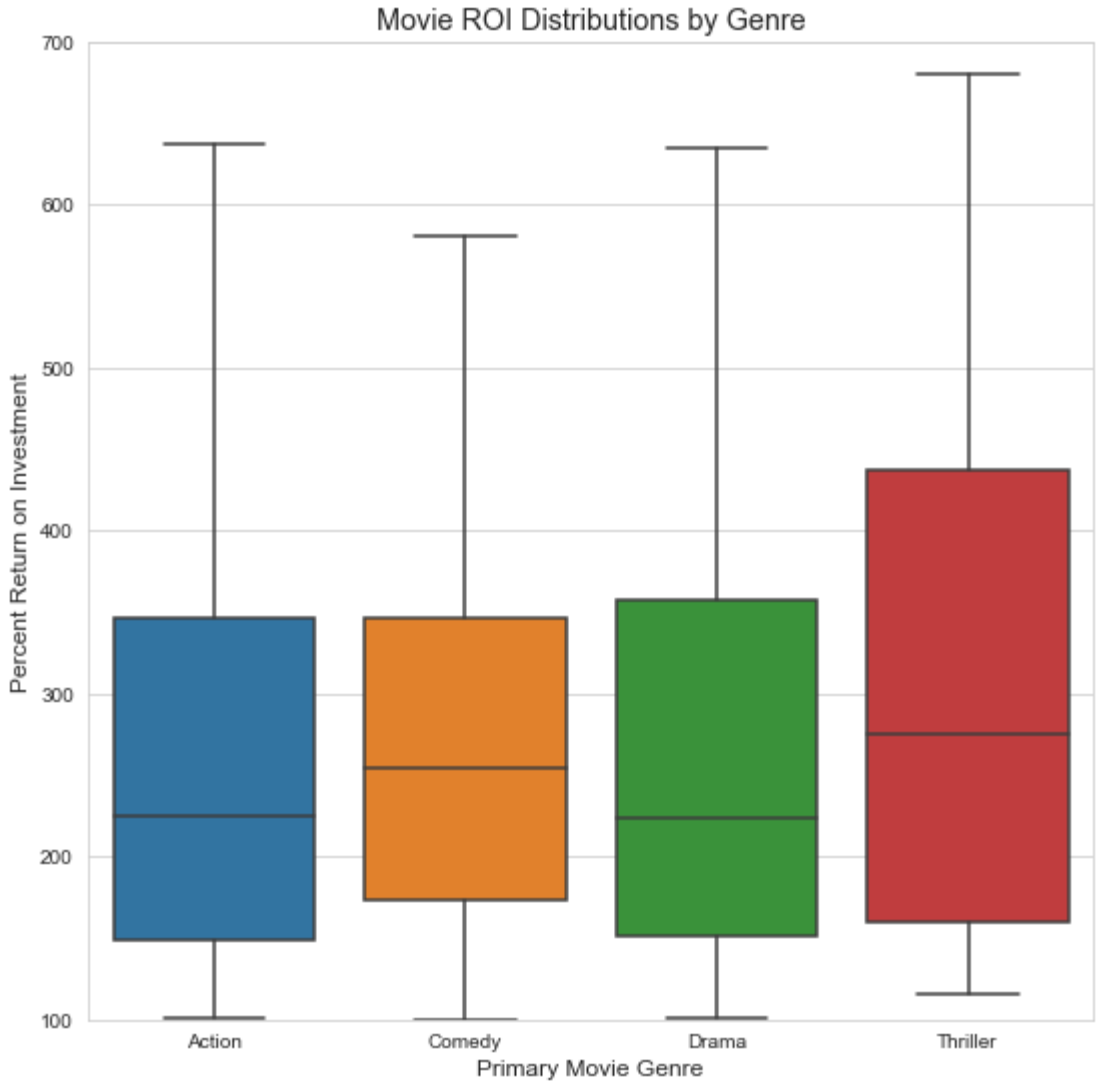




- As in previous plots, can divide data into budget categories like for ROI category
- Define: High > 200,000,000; Medium >= 100,000,000; Low >= 10,000,000; Ultra Low
- Movies in the High and Ultra Low categories (in addition to outer ROI categories) have been thrown out from here on
- No real observed difference in ROI distribution between the two middle budget categories

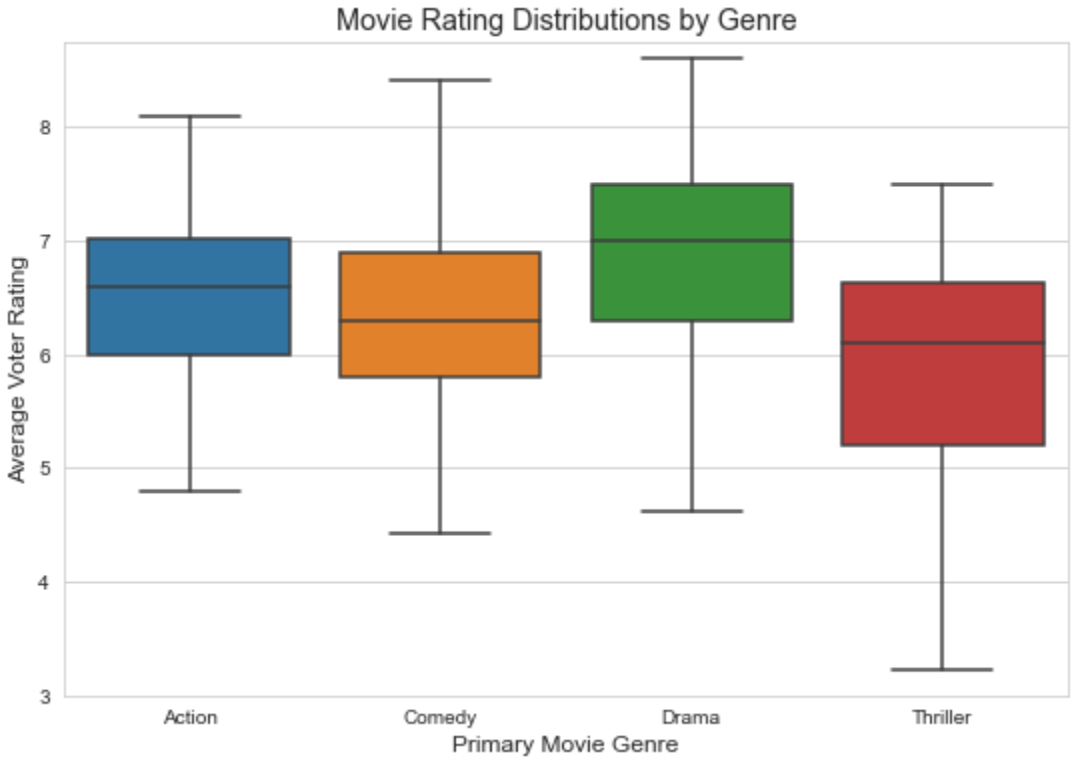
Microsoft

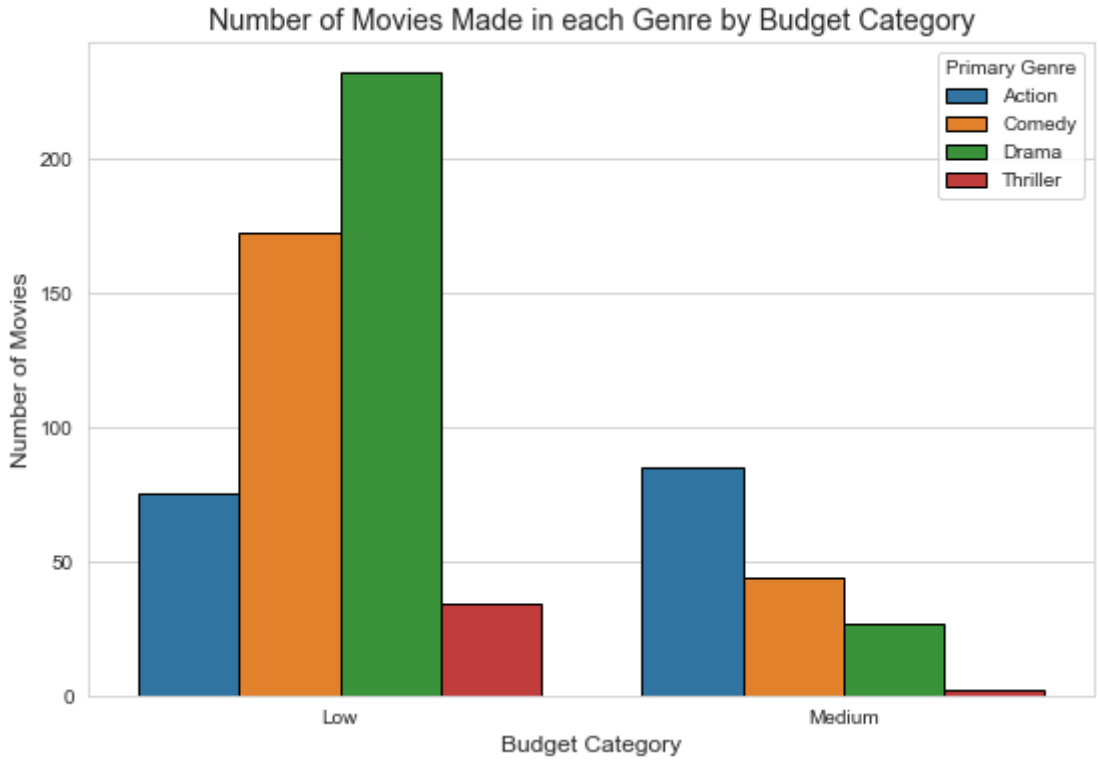
What types of movies produce high returns?



- There are many genres in the movie collection, and some movies have more than one genre attributed
- To be able to more easily look at movies by genre, a new field is defined for each movie called the primary genre
- First a count of all the unique genres in the collection was made. Then, each movie was assigned to one of the four that were already the most popular, if that wasn't already their only attributed genre
- Notable / common examples include Adventure -> Action and Horror -> Thriller

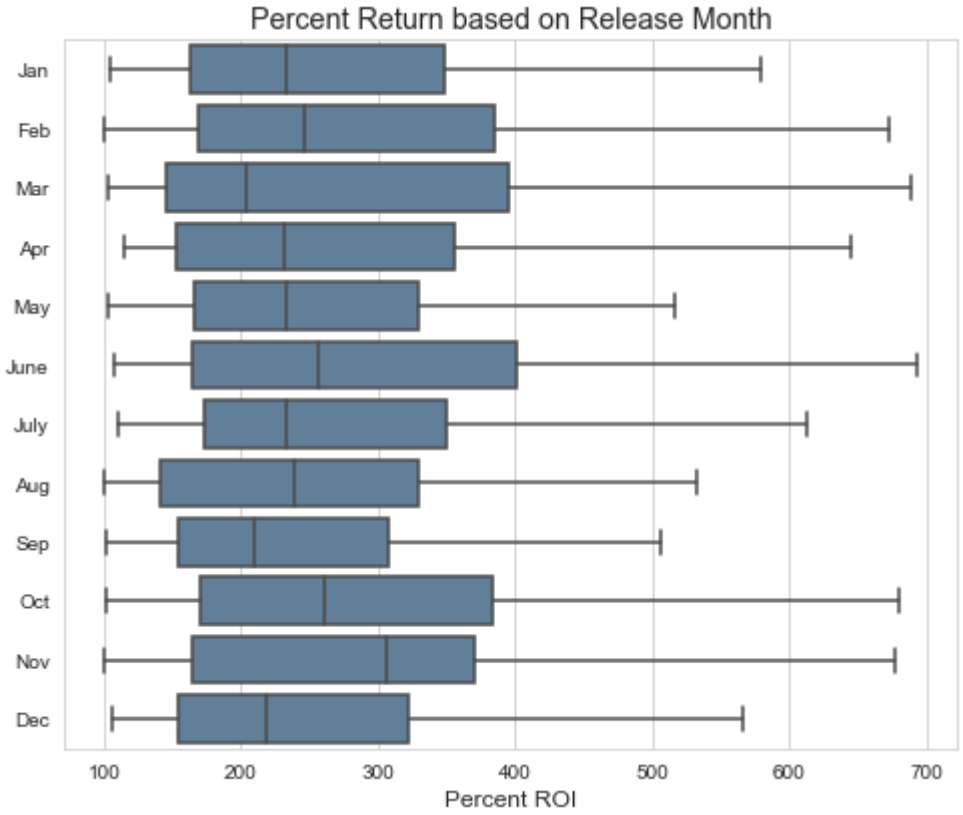
What types of movies produce high returns?

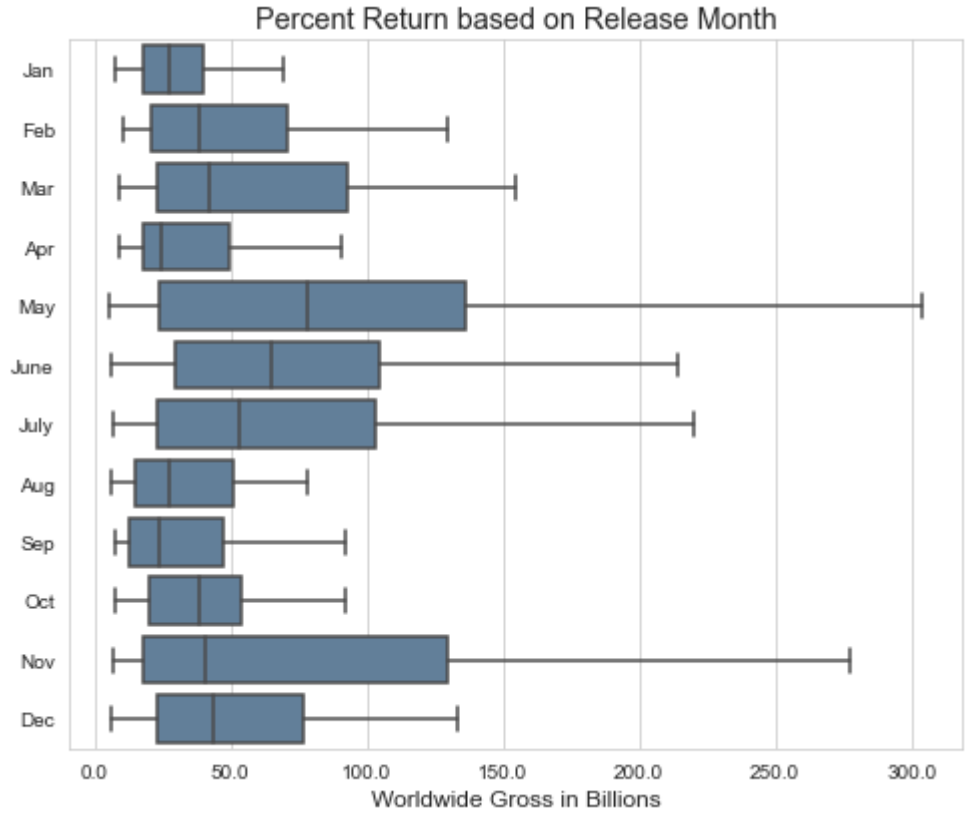




- While we saw on the last slide thriller wasn't edging out the other categories by much, but we can see here thrillers seem to be produced less and may do well on release independent of viewer reception
- This could indicate an easy niche to break into, getting consistent returns without having to worry about production quality, meaning less can be spent on big name cast and crew members

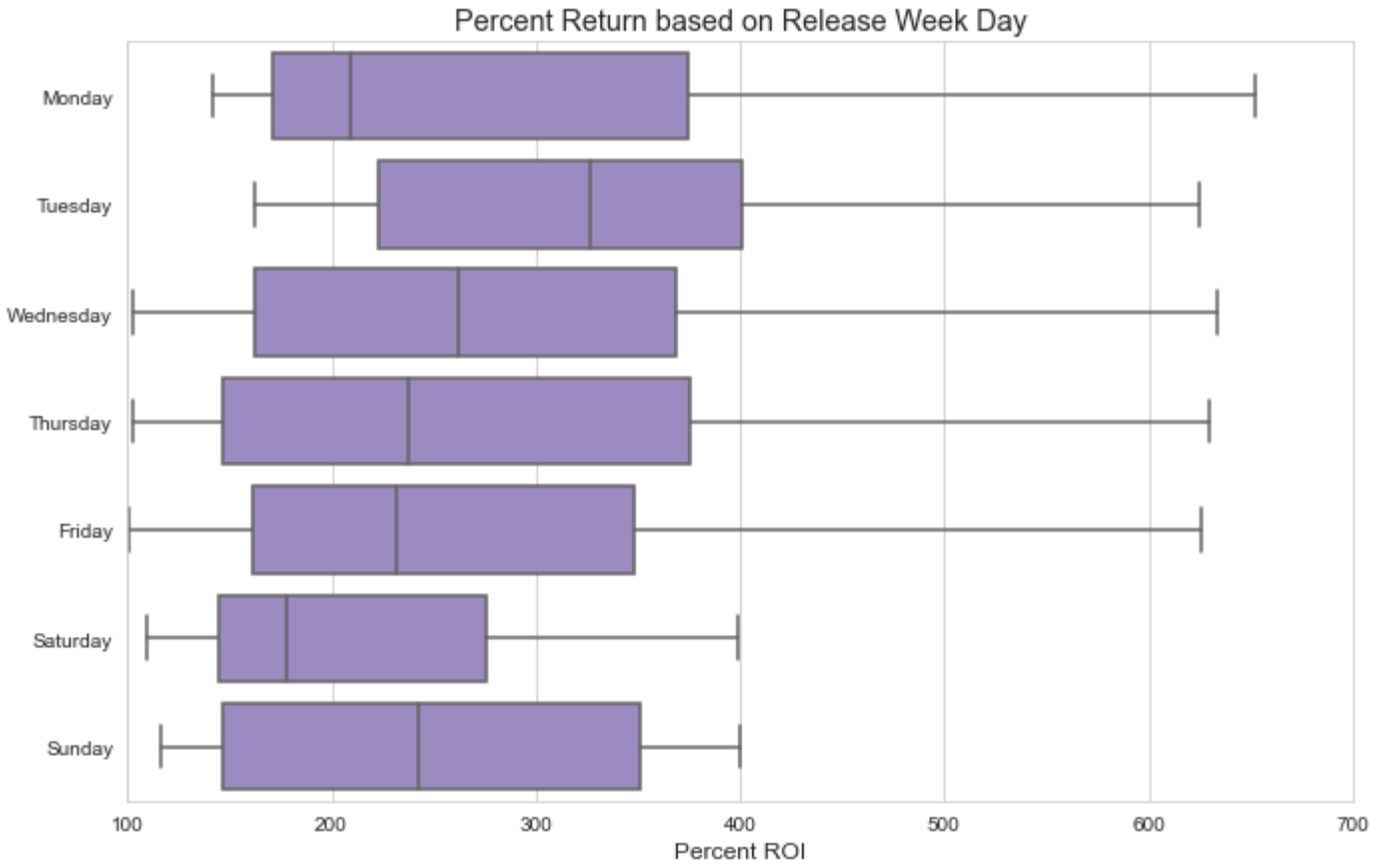
When should the movie be released?





- Hard to determine release month based on ROI distribution, though when placed side by side with Worldwide Gross, we can see that to give the movie a best chance at solid ROI while increasing the ceiling for gross revenue, the movie should be released either going into summer or in November

When should the movie be released?



- Tuesday is the strongest performing day of the week in terms of percent returns

Summary

- In general, spending more on production will increase the return on the original investment, but based on our dataset, it is recommended to spend between $10 million and $100 million
- Of the different movie genres, Thriller movies are least often produced but tend to yield the highest ROI. Additionally, ROI may be less affected by production quality, decreasing production budget
- To maximize availability on theatrical release, release movies on a Tuesday evening in one of the early summer months

Future Avenues of Exploration

- These datasets and analysis only investigates returns based on reported worldwide box office gross numbers and estimated production budgets. The theatrical release of a blockbuster film does not account for a majority of movie-related income; there is home entertainment sales (DVDs), Video on Demand and Streaming Services, Television airtime, and merchandising. On the other hand, costs related to these are also not explored; manufacturing DVDs and merchandise, deals with streaming service providers and television providers, etc. Similarly, ROI's presented here have the potential to be hugely overinflated because it is unclear exactly what is included in the production budget estimate. It may be that marketing and other costs related solely to the theatrical release have not been accounted for.
- We primarily looked at a subset of our movie collection where we required movies to make a profit (have ROI > 0). In the future, we can double check that the described features hold true only for profitable movies, increasing our confidence
- We have access to data connecting movie cast and crew to IMDb ID's which we have for every movie in our collection dataset. While for Thrillers it may not be as important, it may be that there are a subset of the production crew that is a better predictor of high return movies
- As it relates to the first concern, the pandemic, and the rise of streaming services.. Is theatrical release worth it as a new production company with plenty of capital?

Thank You!

- Questions?

- Github Repo: https://github.com/mattlegro/dsc-phase-1-project
- Email: mlegro93@gmail.com
- LinkedIn: https://www.linkedin.com/in/matthew-legro-8b60211b6/

Microsoft