



Movie Data Analysis



Business Case:

Microsoft is looking to enter into the movie industry, and wants to understand what the indicators of revenue and profit are in movie production.



Key Questions

Through the use of data analytics, the working group sought to answer four key questions to help Microsoft better understand profit and revenue in the movie industry:

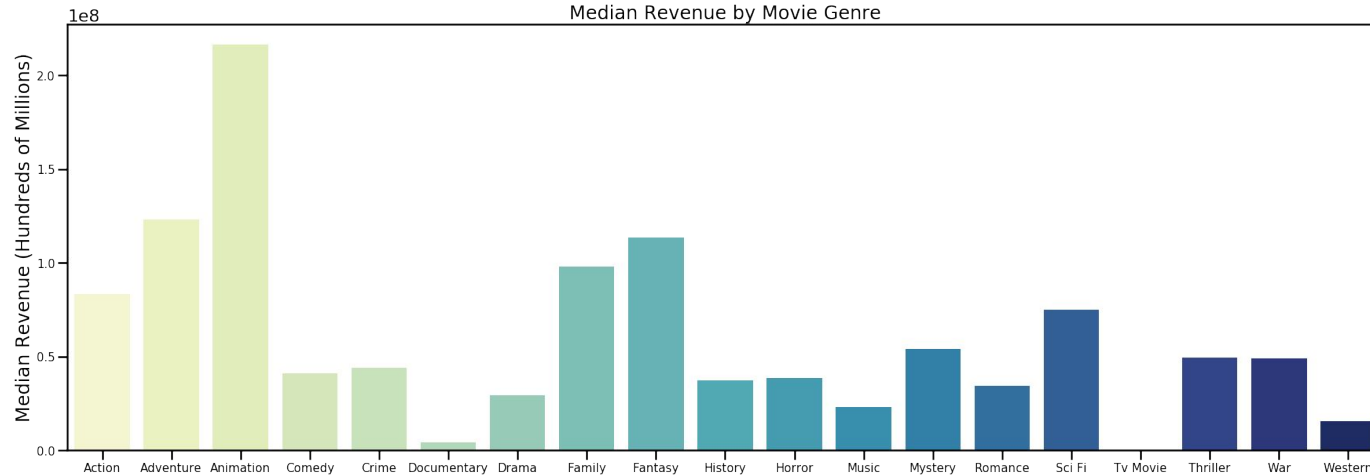
- Which genre of movies is the most profitable and which generates the highest revenue? Which genres have been the most and least profitable in the last 10 years?
- Does a higher movie budget equate to higher movie revenue, profit or ROI?
- Do higher rated movies earn more profit? More revenue? What's the relationship?
- When, throughout the year, is the most profitable time to release movies? Which month?



Methodology and Sources

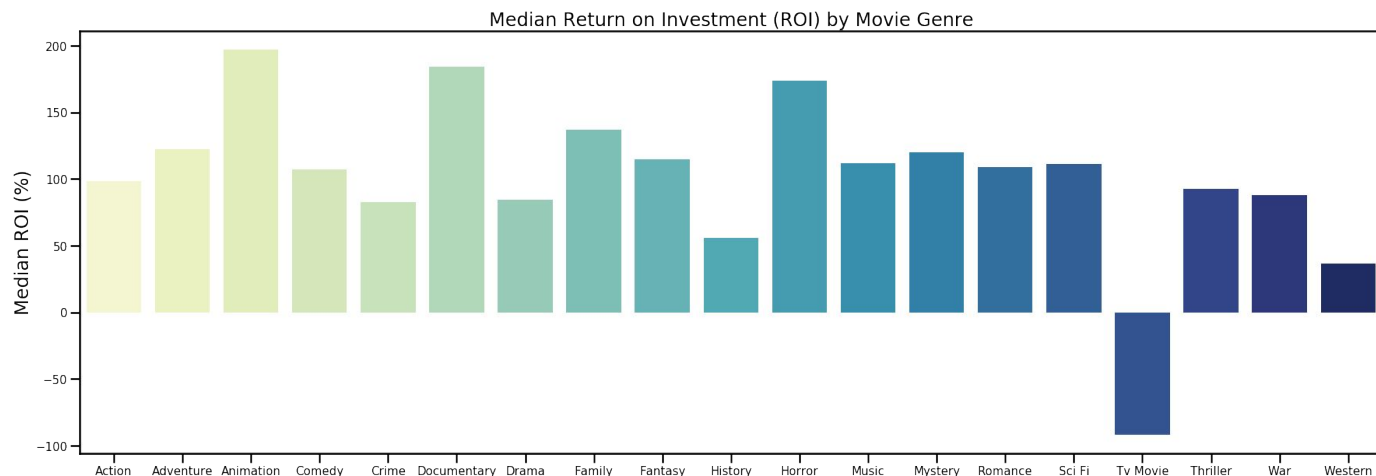
- Data was sourced from the TMDB API and a dataset provided by The Numbers
- Data was cleaned using mostly Python, the Pandas library, and the Numpy library
- Visualizations were created with Matplotlib and Seaborn
- The dataset included 4,160 movies from the years 1915 to 2019
- When applicable, the median was used as a measure of central tendency
- Profit was measured in absolute terms and using Return on Investment (ROI)

Which genre earned the highest revenue?



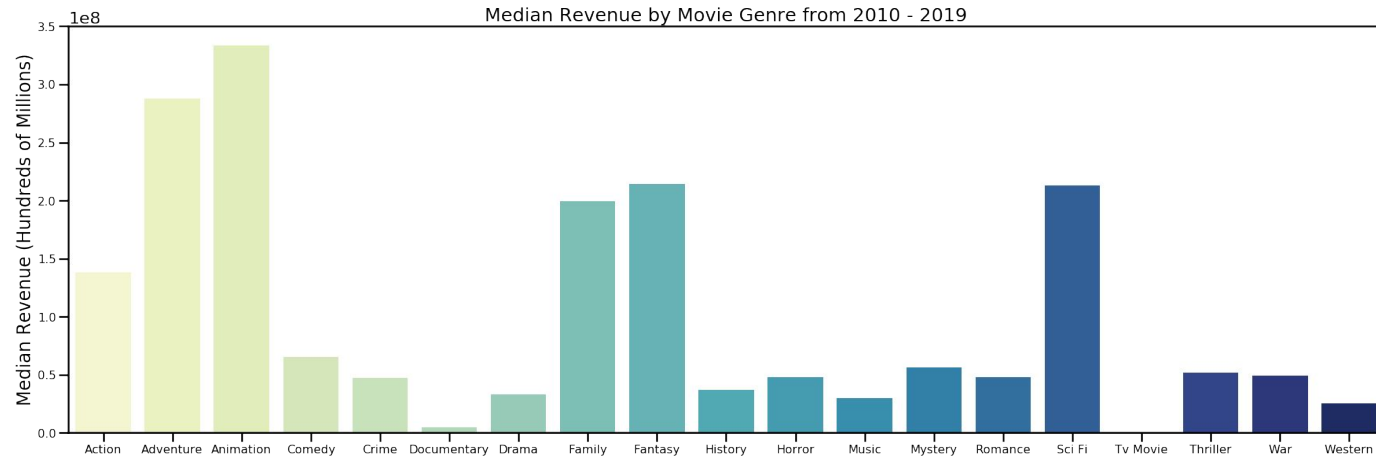
- Animation, followed by adventure and fantasy were the genres that generated the highest amount of revenue
- This could be due to animated films reaching a broader, family audience
- The genres that generated the lowest revenue were documentary, tv movies and western

Which genres are the most and least profitable in terms of ROI?



- Animation, along with generating the highest revenue, was the most profitable (median ROI of approx. 200%)
- Documentaries, while generating less revenue than other genres, were the second most profitable, potentially due to having lower production costs
- The only category to generate negative profits was TV movies

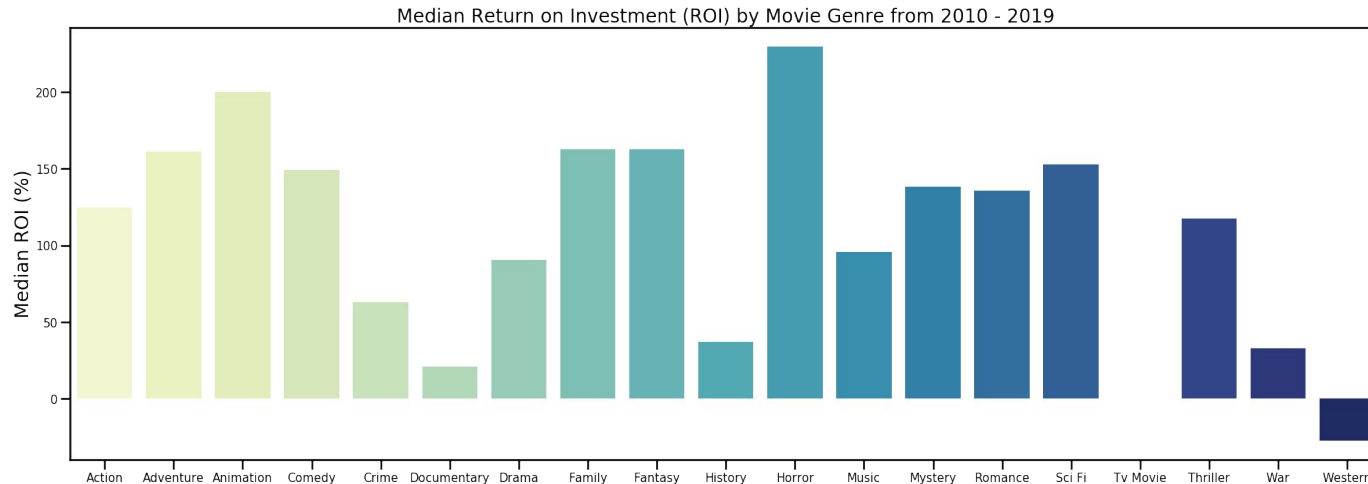
Which genres have generated the highest revenue in the last 10 years?



- Animation and adventure generated the most revenue in the last 10 years (each over 250 million)
- While documentaries have generated fairly low revenue, this may be skewed due to limited documentary movies in the dataset

Note: There was no information on TV movies from 2010 to 2019 in the dataset

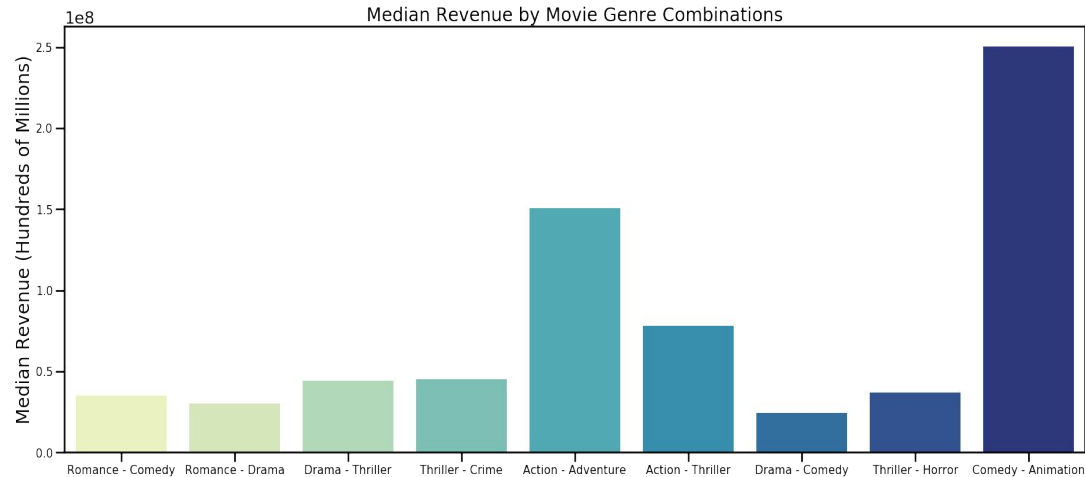
Which genres have generated the highest ROIs in the last 10 years?



- The most profitable genre in terms of ROI was horror, followed by animation and adventure
- The highest ROIs in the entire dataset were from the Blair Witch Project and Paranormal Activity, while the highest ROIs in the last 10 years were from The Gallows and The Devil Inside, all four being found footage horror movies. This could be due to the low production costs.

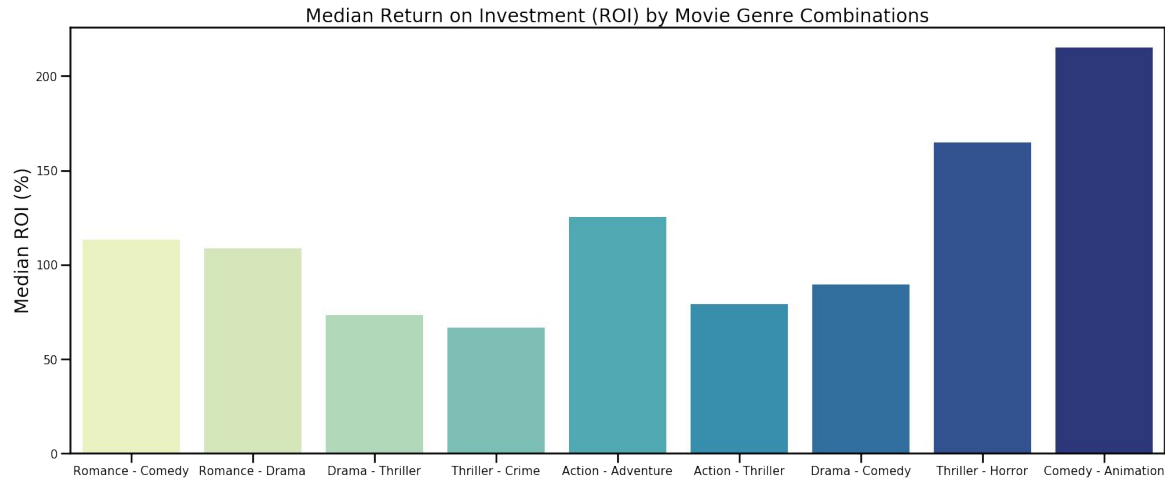
Note: There was no information on TV movies from 2010 to 2019 in the dataset

Which popular genre combinations have generated the most revenue?



- The popular genre crossover movies that earned the most revenue were comedy - animation movies, and action - adventure
- The popular genre crossovers that generated the least amount of revenue were romance - comedy and drama - comedy

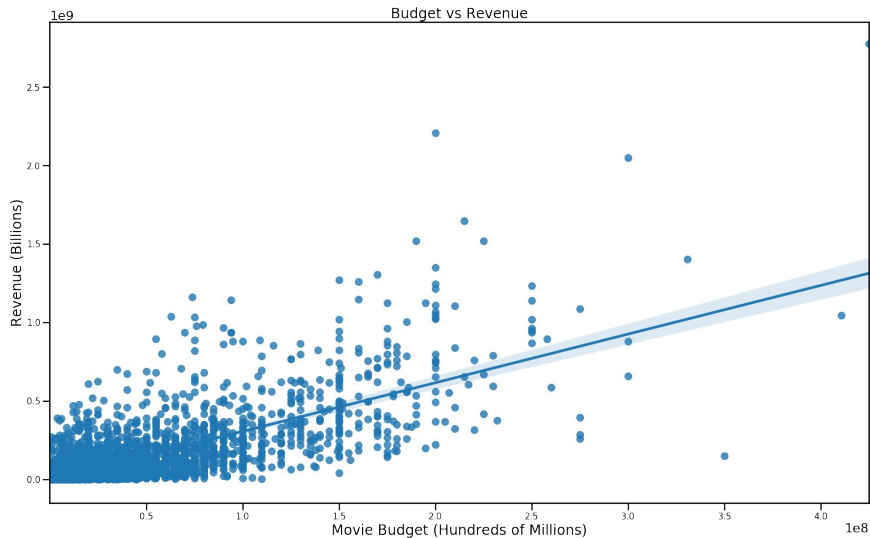
Which popular genre combinations have generated the highest ROI?



- While animation - comedy and action - adventure followed a similar trend as in revenue, the second highest ROI was generated from thriller - horror
- Romance - comedy and romance - drama categories also performed better in terms of profitability than they did in terms of revenue.

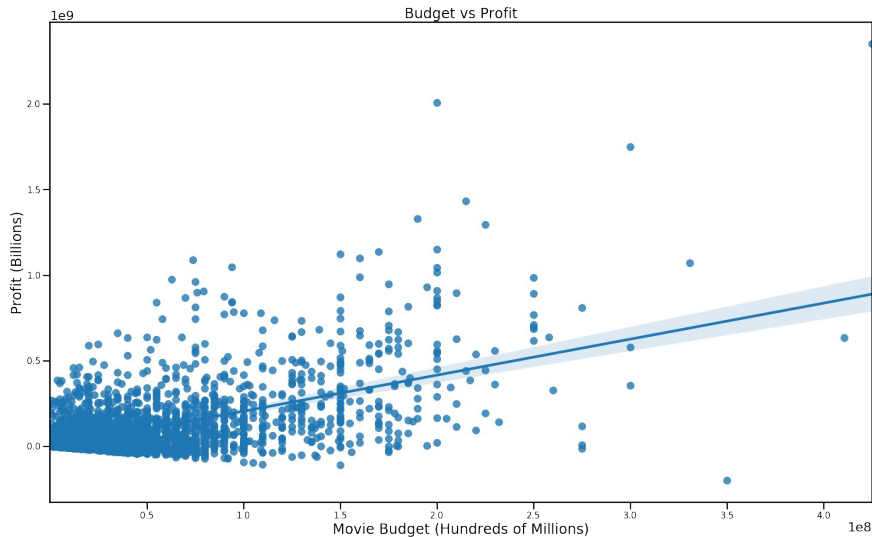
What is the relationship between a movie's budget and its revenue?

- Revenue does have a positive relationship with budget
- This is logical considering high production value movies such as Avatar and The Avengers have had hundreds of millions spent on production, but grossed billions worldwide
- This is not to say there can't be exceptions, such as in the case of early found footage horror films

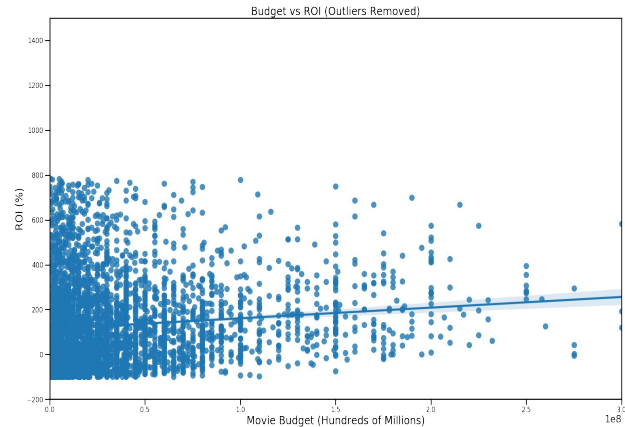
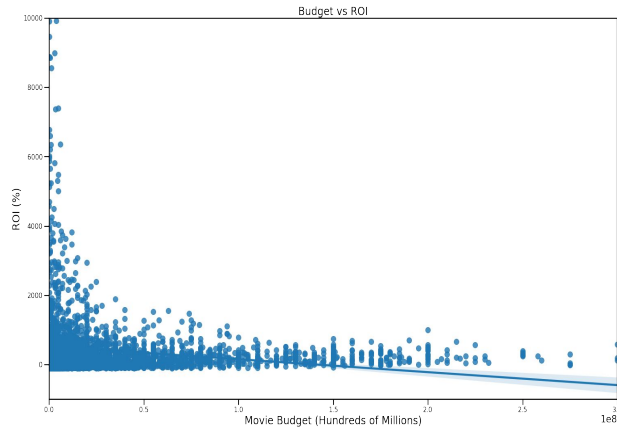


What is the relationship between a movie's budget and its profit?

- Profit in absolute terms does have a positive relationship with budget, although slightly less than revenue
- Exceptions: Some higher budgeted movies that generated negative profits include Evan Almighty, Sahara and Dark Phoenix

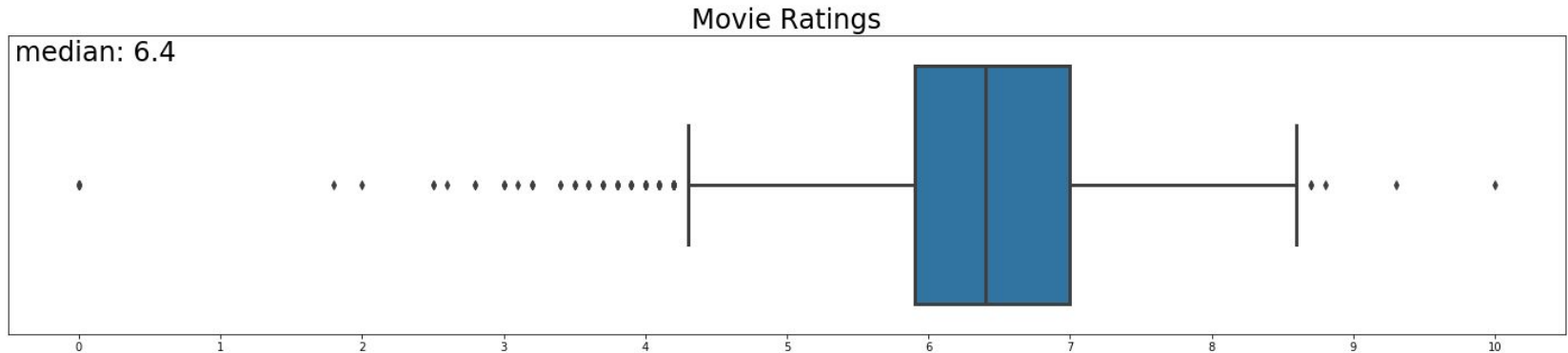


What is the relationship between a movie's budget and its ROI?



- Outlier films (such as found footage horror that have low production values but high ROI's) skew the data such that it appears that budget and ROI have no relationship
- Without such outlier films, it appears that budget and ROI have a slightly positive relationship, albeit weak

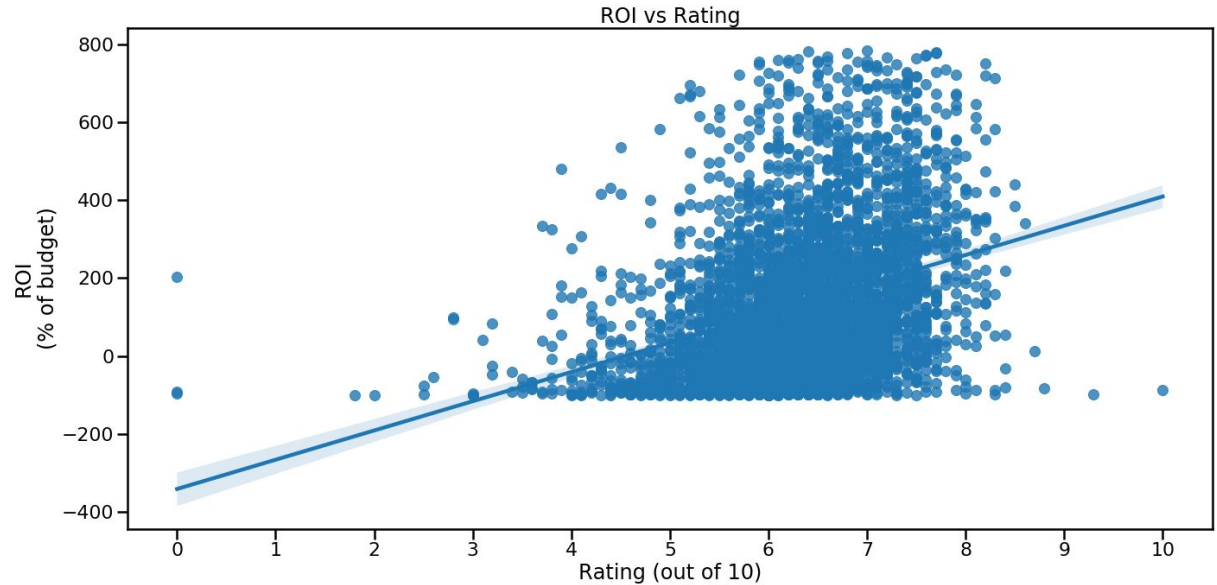
What is the relationship between a movie's rating and it's ROI?



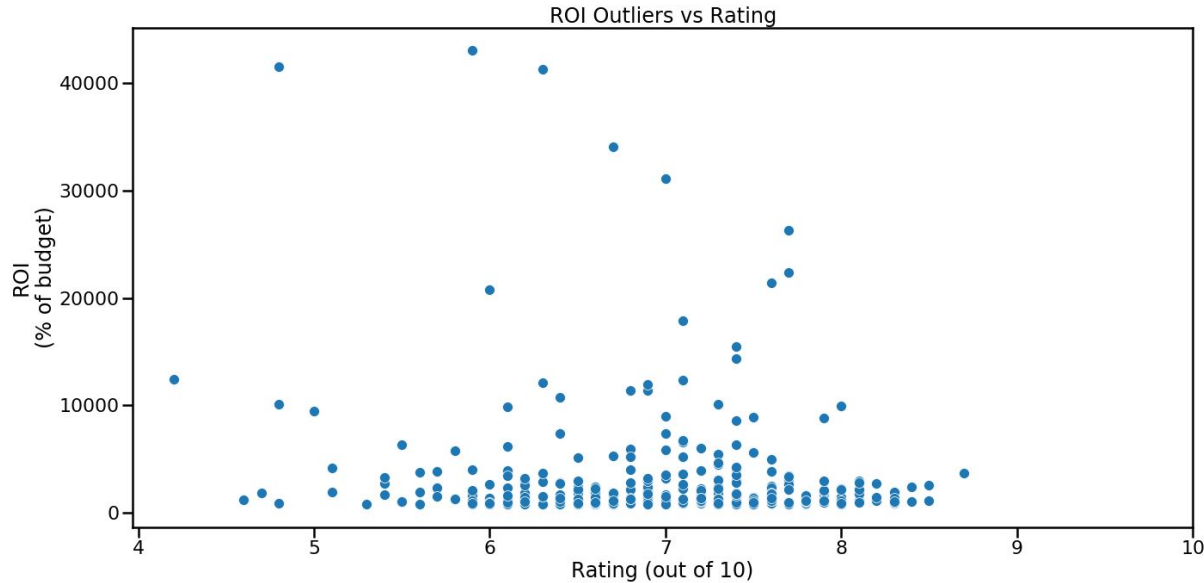
- The median rating is 6.4, while the mean is 6.36
- Not many outliers in ratings
- This is about what is expected

What is the relationship between a movie's rating and its ROI?

- ROI does have a positive correlation with rating
- The relationship is not very strong: as long as a movie is at least around average, other factors influence its ROI more
- This does not take into consideration ROI outliers

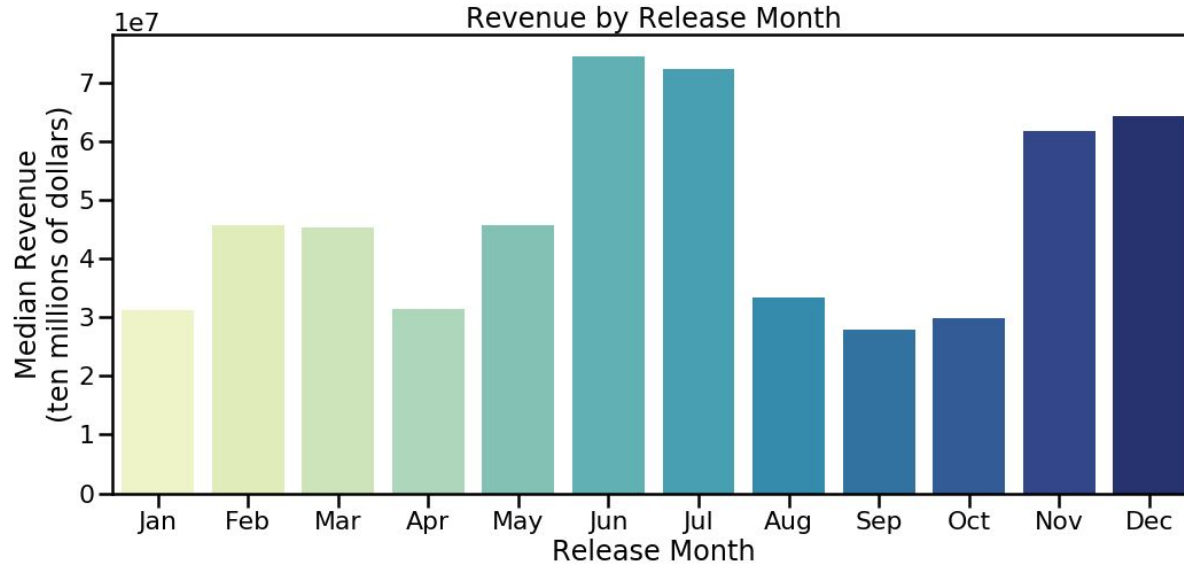


What is the relationship between a movie's rating and it's ROI?



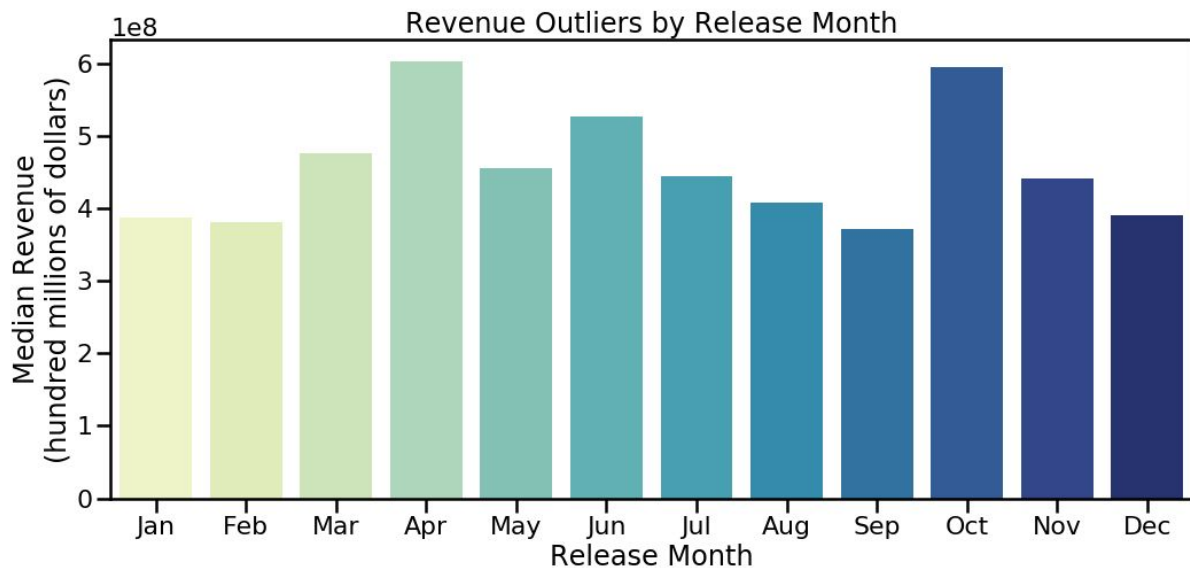
- This only looks at movies that vastly overperformed with regards to ROI
- There seems to be little to no relationship between a overperforming movie's rating and it's ROI

When in the year is it most profitable to release a movie?



- June, July, November, and December releases see much more movie revenue than other months
- These months line up with common holidays for students

When in the year is it most profitable to release a movie?



- This only looks at movies that vastly overperformed with regards to revenue
- Does not match the same pattern as most movie releases
- Breakout hit movies may be harder to predict



Key Takeaways

1. Invest in animated films, due to their tendency to have high revenues and high profits
2. A higher budget can result in higher revenues, but not necessarily higher profits
3. If the movie is at least of average quality, it can generate a high ROI
4. Release movies around school holidays during December, June and July, as movies released during these months tend to have higher revenues



Limitations

- Our data was limited to movies that were present in both the Numbers database and TMDB
- Ratings are user supplied on TMDB
- Dataset is approximately a year old, so it is missing recent data
- Certain genres had less information than others, potentially skewing that data
- There were significant outliers in certain aspects of the data, including ROI

Next Steps

- Get data from other sources, such as IMDB and OMDB
- Gather ratings from other sources, such as IMDB and RottenTomatoes
- Gather new dataset from the Numbers instead of using the supplied one
- Discount genres that have too few data entries
- Outliers were excluded when applicable using a function of the interquartile range



Thank You

