What Makes a Movie Successful?

An analysis of factors that correlate to high rated and grossing films

In this project, I was tasked by Microsoft's budding movie studio to look at the movie industry to provide some recommendations of how to enter the movie-making scene.

Metrics of Success

What questions did I aim to answer with my analysis?

- What genre and MPAA rating combinations produce the most highly rated films?
- Which genres produce the most profitable films?
- Which director produces the highest (IMDb) rated films?
- Which director produces the highest grossing films in the box office?

2

To quantify success, I looked at two factors: IMDb rating and gross profit. My questions aimed to uncover any correlation between Genres, MPAA rating, directors and their influence on the success of a film.

Methodology

Data Collection
Data Cleaning
Modeling Results

3
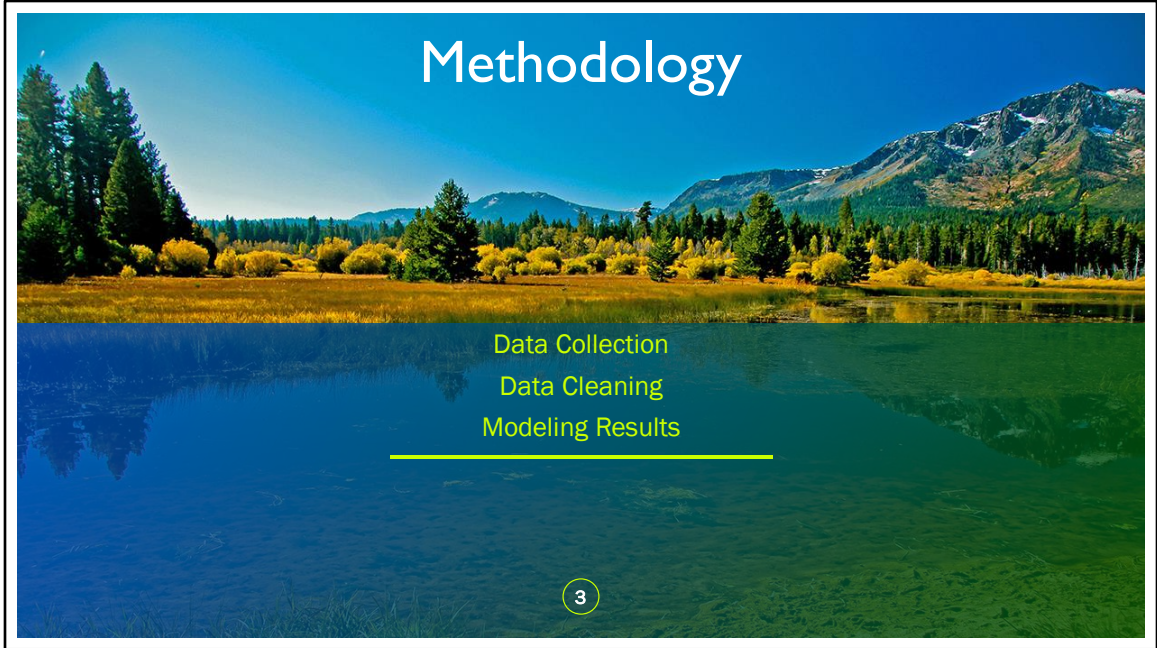
The data needed to explore these questions was gathered by a couple methods. I made use of the supplied CSVs, I webscraped IMDb's top-rated movie list, and I filled in the blanks with significant use of the OMDB API.
I then used a variety of libraries and modules to clean, organize, and model my data to come to my conclusions. Let's get into the data and what I was able to gather from it.
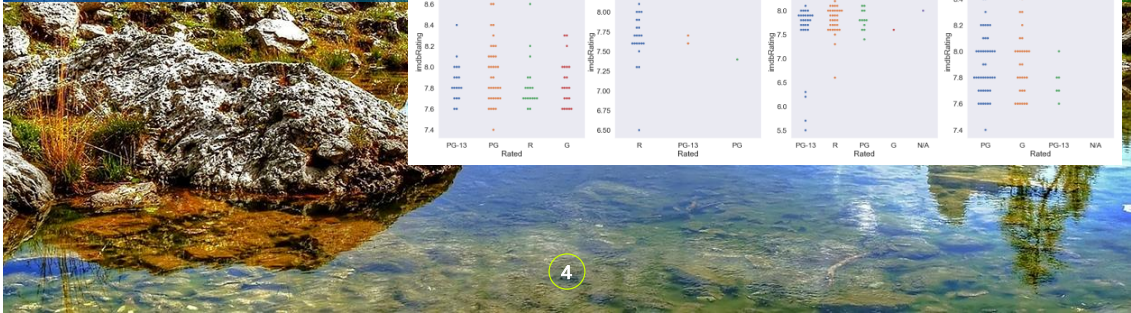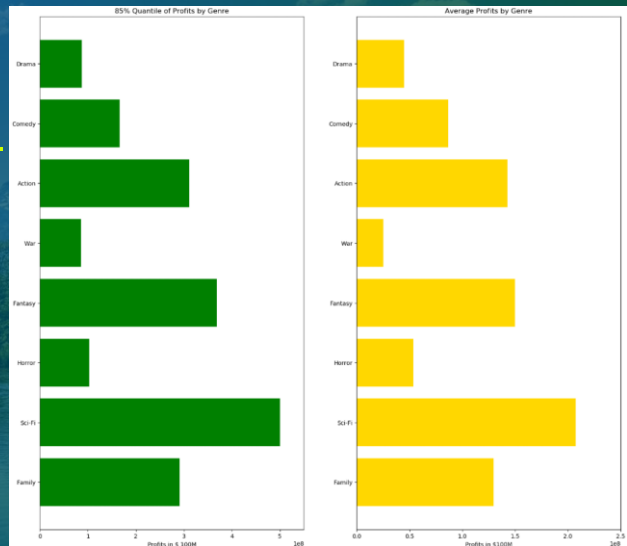
In this analysis, I looked grouped the movies together in subsets based off genre then plotted them by MPAA rating and their respective IMDb rating. As you can see, Drama, Comedy, and Action movies have the bulk of their distribution in a higher IMDb rating than other genres. There are few very low rated movies in these genres. As for rating, with the exception of Family movies, PG-13 and R rated films outperform their PG and G rated competitors.

By this analysis we can gather that it would be a sound decision to pursue either a Dramatic, Comedic, or Action-packed plot geared towards more mature viewers to see the best chances of producing a well-received film.
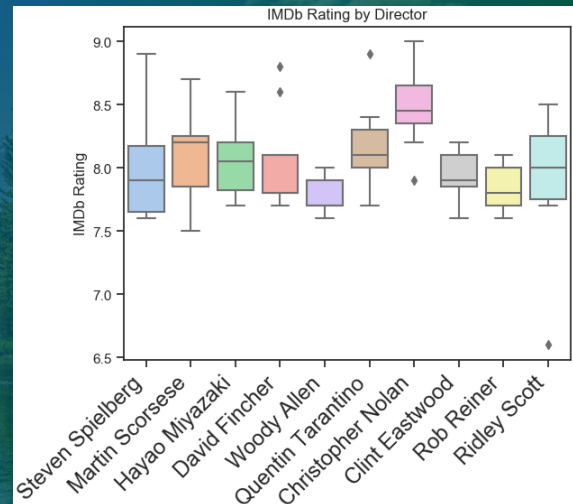
Next, I looked at how genres correlate to profit, this data is a little more unclear due to the difficulty in obtaining an extensive dataset of accurate financial data.
I've included two visualizations of the same data to broaden our vision of what's happening here. The green bars show the 85% quantile of the data, removing missing data and flops, and also removing all-time grossing blockbusters to create perhaps a more realistic outlook. The Gold bars show the mathematical means of the datasets – flops, blockbusters, and all.

It seems that Drama movies don't have a very high average profit, but this could be due to there being more drama movies than other types and therefore more flops. War movies and Horror movies seem to bring in the lowest average profits while Action, Fantasy, Sci-Fi, and Family movies bring in much higher averages.

My next look at the data was focused on figuring out which director produces the highest IMDb rated films and therefore should be pursued.
From my large set of the top 1000 movies from IMDb, I looked to see which directors were most represented on the list(and still alive and employable).

I further analyzed that shortlist directors by looking at all their movies and plotted them by their rating.

Christopher Nolan has overwhelmingly so the best-looking numbers in every metric. Martin Scorsese and Quentin Tarantino also have consistent favorable ratings but not to the degree of Christopher Nolan.

Which directors produce the highest gross box office sales?

- This slide illustrates where my data has a hole, accurate financial data is very hard to come by.
- From the data available, we can see that Christopher Nolan is consistent in creating very successful films.
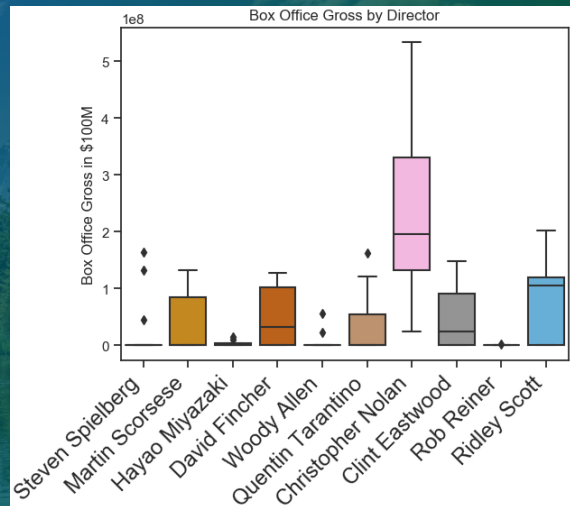
Here I continue to look at a directors influence but switch my focus to gross box office sales from IMDb rating.

This visualization illustrates one of the biggest problems I had in collecting data for this project. Accurate financial data is very hard to come by, and where it's available it's very expensive.
This is understandable as studios would likely seek to protect these exact numbers as they are pretty sensitive.

Using the data I had available to me, we can see that there are clear holes in Spielberg's, Miyazaki's, Allen's, and Reiner's data and of the remaining directors, Christopher Nolan vastly outperforms everyone.
This could be improved upon by using a more extensive financial dataset like the OPUSdata dataset.

Conclusion

My recommendations for the studio

Genres & MPAA Rating

- Drama, comedy, and action films are generally more highly praised.
- An action movie with some comedic relief and a dramatic side plot would check all the genre boxes for success.
- Aim to market films to more mature audiences, PG-13 and R rated movies vastly outperform the others.

Directors to Pursue

- If available, Christopher Nolan should be pursued for employment, he has a great track record for IMDb ratings and gross box office sales.
- Christopher Nolan also has experience in the most successful genres and ratings.
- If he's unavailable, the most suitable runner up is Quentin Tarantino.

8

In conclusion, I have the following recommendations:

Our studio should work on producing movies in the drama, comedy, or action genres with perhaps an action movie with some comedic relief and a dramatic side plot prioritized.

We should prioritize the young adult to adult market and generally pursue a more mature audience, keeping with the PG-13 to R rating.

If possible we should pursue Christopher Nolan as a director, or in the case he's not available Quentin Tarantino.

They both have respectable track records in the genres we should be working within.

Future Work

What could be done in the future to go into deeper analysis given more resources?

- Purchase a more extensive, accurate financial dataset for film budget, profit, and gross revenue. The OPUSData dataset was prohibitively expensive for the purpose of this project.
- Look into the Spotify API to see which film score producers are the most listened to and highest rated to see who should compose our soundtracks.
- Analyze which actors & actresses have the best ROI for their cost per movie to see who we should seek as our stars to maximize profit.

9

Some future work that could be useful for further analysis would be to purchase a more extensive, accurate financial dataset from OPUS or another provider.
This dataset, or others of the like were prohibitively expensive for the purpose of the project.
We could also look into using the Spotify API to see which film score producers are the most listened to and highest rated to see who we should pursue to compose our soundtracks.
And finally, we could take a deeper look into which actors and actresses have the best ROI for their cost per movie and whether or not it's better to have a star-studded cast and higher overhead,
Or some up & coming talent with less overhead but also less recognition.

THANK YOU!

Ryan Snow

Thank you! I'm happy to answer any questions that you may have about my exploration, analysis, and presentation.