Microsoft Film Analysis Based On ROI

Author: Andre Layton

Summary

You're looking to break into the film industry, but stuck on what sort of content to begin creating. This study analyzes movie data from IMDB and financial data from The Numbers to recommend the best types of films to create for your studio's initial content.

The main focus behind the analysis is seeing which films are best to begin creating by calculating and comparing both the median Domestic and Worldwide returns on investment (ROI), and then analyzing production budgets to estimate how much investment will be necessary.

According to the analysis:

- Adventure films are the best genre to begin creating for high gross sales and ROI.
- A minimum budget of $100 million will be necessary to begin seeing profit.
- A maximum of $170 million to invest in the initial film in order to gross $200 million domestically.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- Determine which film genre would be best to create for the new studio's initial content based on ROI data

- Estimate production costs

Data

- The dataset from The Numbers displays the budgets, domestic gross, worldwide gross, and release dates of various movies, ordered by budget in descending order.
    - The data from The Numbers will be crucial in calculating the ROI, which is equivalent to the gross income divided by the cost of the investment, or the budget in this case.

- The data located in IMDB is located in a SQLite database with various tables, one of which is "movie_basics", which contains the names, genres, and other details for each movie in the database.
    - Despite the need to calculate the ROI, the independent variable we are interested in is the movie genre and which genres yield the highest ROIs.

Methods

Action	530
Drama	456
Comedy	403
Adventure	201
Biography	154
Horror	134
Documentary	131
Crime	120
Thriller	30
Animation	19
Sci-Fi	12
Family	8
Fantasy	7
Mystery	6
Romance	5
Music	5
Musical	4
War	4
Sport	2
Western	2

In order to combine both datasets and create a single onefor analysis, I begin by preparing the movie_budgets data first. I drop any fields I do not need (i.e., 'id') before I clean up the names of the remaining features.

I also change the data type of the last three columns, based on the information given by the .info() method, as they are numeric and key to statistical analysis later in our exploratory process. In order to convert these fields to integer types, the values for these fields must be transformed to fit conversion criterion. The .head() method also reveals that there are movie titles that contain apostrophes (') that did not fully convert, and thus requires transforming.

Next step is to clean up the SQL table to fit for merging with our refined movie budgets dataset. Fortunately, this table doesn't require much work - only renaming the field names. Once that's complete, we merge the two together using the Movie Name field. After joining both tables, we begin cleaning up our final DataFrame by creating a new column representing whether the movie titles match to ensure the genres listed accurately describe the associated movie in each record.

Next, we find any missing values, and after taking a deeper look at the dataset, decide to remove these missing records as well as the records that do not have matching titles. Before dropping the records with mismatched movie titles, we create a function to change all movie names with numbers in the string to numeric form to bring a little more consistency in the movie names, and allow us to retain a few more records. We also drop all duplicates from the combined dataset to produce a final dataset for the next step in our process.

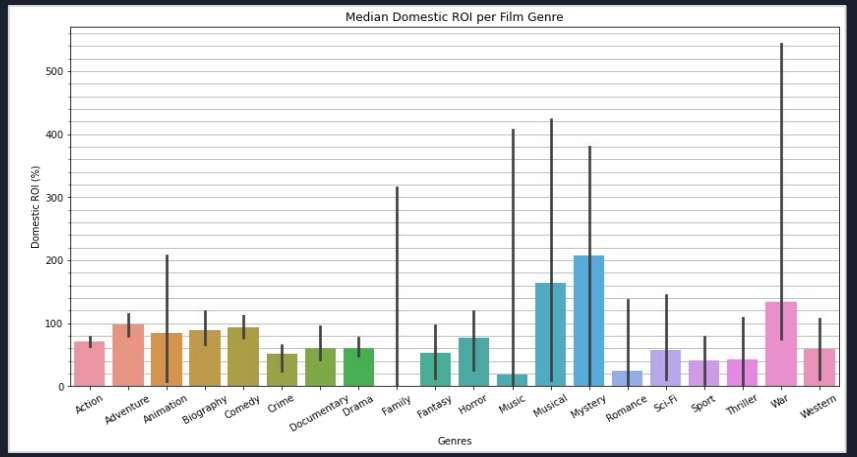
Shown to the left is a count of all genres within the final dataset. As you can see, the bottom half of genres lacked sufficient records, and as such, are susceptible to outliers present, which may skew our results. We refocus our analysis within the top 4 genres, and the next few visuals will support this refocus.

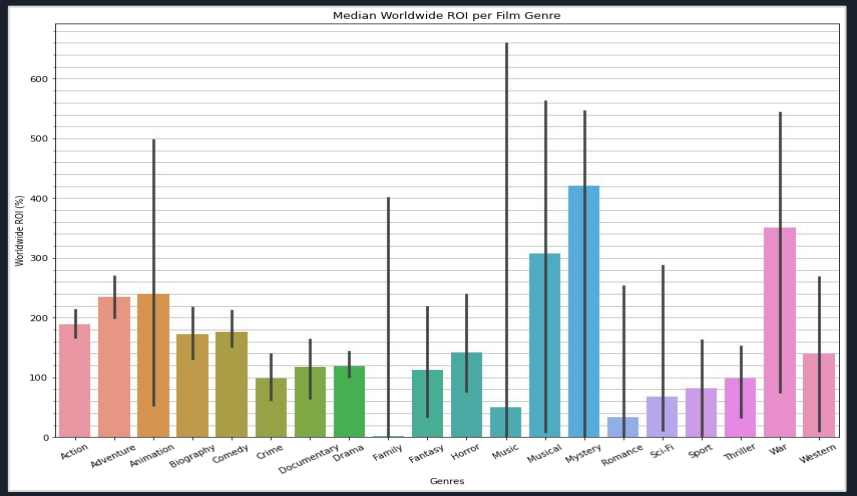
Results

Based on the statistical analysis and visual results, Adventure movies would be the best genre to begin creating for potential domestic and worldwide success, predicated on the film's return on investment (ROI).

Other genres to consider are Comedy movies, which held the second-highest domestic ROI among the top 4, and the highest average domestic ROI. Of course, these findings are formulated assuming that we only consider the genres with low margins of error (depicted by the black bars illustrated in the first few plots above).

Musicals and War movies produced high median ROIs; however, that is primarily due to the fact that there were few records for those genres, among others. As such, the results were susceptible to any outliers that may exist (the high error margin bars in the bar plots portray this effect). Therefore, we focused on the top four genres - Action, Adventure, Comedy, and Drama - that had a high number of records. There was more confidence in their results, exhibited by the low error margins.





Conclusions

Here are my recommendations based on my analysis:

1. **START WITH ADVENTURE FILMS!** Adventure films generated the highest median domestic and worldwide returns on investment, among the top four genres. We focused our analysis to those four genres due to outliers skewing the results (e.g., low budget films that found major success), and a lack of records for the other genres to confidently include them in the analysis. As such, Adventure films were the best choice to begin with, and is estimated to produce the highest median returns, as well as an average domestic ROI of approximately 130% - suggesting the film will make enough to cover production costs and even generate profit.

2. **MINIMUM BUDGET SUGGESTIONS!** According to the average budget model above, prepare to spend a minimum of $75-80 million on this film. Despite the median budget being $55 million for Adventure films, the visuals suggest the higher budget range will generate enough revenue, domestically, to break even on production costs, or more than double the investment, worldwide. However, the budget minimum I am suggesting is $100 million - where the 75th percentile of films lie, as shown in the boxplot above. This is estimated to gross well over the initial investment, and generate profit, both domestically and globally.

3. **MAXIMUM BUDGET SUGGESTIONS!** I am suggesting a maximum budget of approximately $170 million for the Adventure film. According to the models, in order to break $200 million in domestic gross, an estimated budget of about $170 million will be required. That same estimated budget will far exceed the $400 million mark, globally, and thus lead to commercial success.

The analysis above, as mentioned earlier, is formulated assuming each movie has one individual genre that defines it. Unfortunately, films can be characterized in multiple ways/genres, and the analysis would need to be revise with that factored in, in order to solve the business problem when presented with those possible datasets. By doing so, we'll be able to reach a wider audience, and further market the film for success. To further improve this project, I could also analyze ratings, see how that compares with ROI success, and observe any changes in results (or any confirmations).

Thank You!

Email: alaygt6@gmail.com
GitHub: @therookiescientist-andre
LinkedIn: linkedin.com/in/username/