

This project was completed as the final Phase 2 assessment in the Flatiron School's Data Science Bootcamp.

Analysis by Erin Wasserman, February 2024

Overview

This project analyzes which types of movies generate the most revenue at the box office. Descriptive analysis of movie data sets collected from Box Office Mojo, IMBD, Rotten Tomatoes, The Movie DB, and The Numbers, will highlight specific movie characteristics that increase ROI such as, genre selection, month of the film's release, and the length/runtime of the movie. This analysis can be used by the company to make decisions on these specific movie criteria and help the company's new movie studio operate with the lowest risk in their new business endeavor.

Business Problem

Analyzing several movie datasets can provide valuable insights that can lead to concrete business recommendations in various aspects of the film industry. When considering characteristics of a successful movie, the movie data was analyzed to identify movies with the greatest return on investment. Based on this analysis, which is explained below, are the following three recommendations:

Action and/or Adventure genre movies gross higher than other genres

Languages

Jupyter Notebook 100.0%

- Most successful release months are May, June, November, and December
- The most profitable movies are 85-125 minutes in length

Data Understanding

Some media databases provide public data that includes various data about historical and current films both domestically and internationally. Each database has a variety of variables some of which may match other databases. The quantity of data provided was excellent, however, the quality of these various datasets were weak.

Overview of dataset:

Independent Sources:

- Internet Movie Database (IMDb)
- The Movie Database (TMDb)
- Box Office Mojo
- The Numbers

About: Selected film industry data reported between 1915-2020. Due to the differences in the independent databases, when choosing variables several considerations were made:

- Data Relevance
- Data Quantity
- Data Quality

Decision-based Variables used:

- Profit/Profit Margin
- Average Domestic Gross
- Movie Runtime
- Movie Release Date
- · Genre of Movie

Methodology

This project uses descriptive analytics to report film financial performance across several variables using historical and current data. This analysis organizes and uses data from a variety of datasets to alert, explore, and report trends and insights into the potential gross earnings of new films based on genre, runtime, and release date.

There are two main sections:

1. Data preparation steps include:

- Import libraries
- Read and clean provided data
- Deal with missing values
- Join datasets

2. Analysis for each variable:

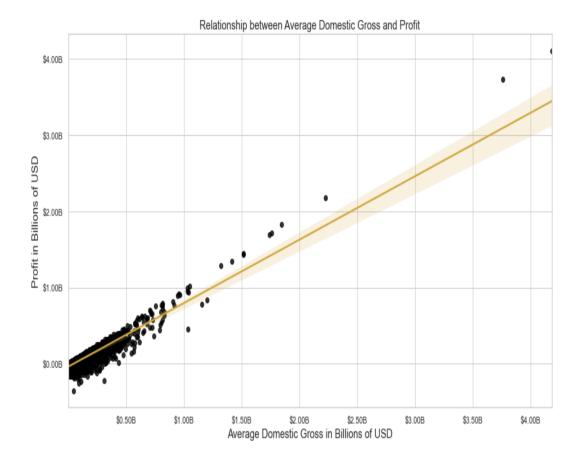
• If possible, apply feature engineering

- Visualize data
- Provide conclusions and recommendations

Key Findings

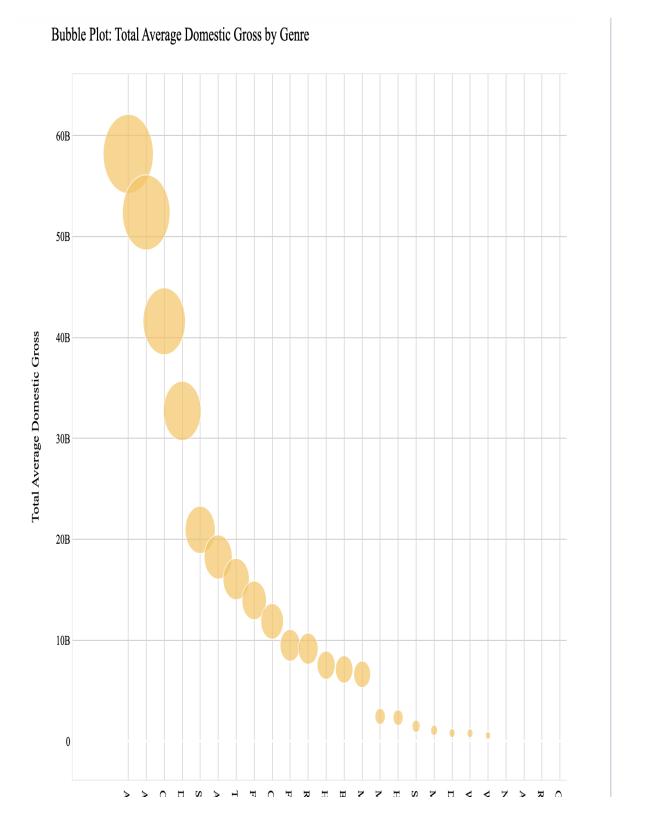
Profit and Profit Margin

- A strong, positive, linear association between profit and average domestic gross with a few outliers.
- The median profit is approximately 500 million US dollars.
- The median profit margin for a top 100 all-time grossing movie is approximately 88%.
- No correlation between budget and adjusted average domestic gross.



Genre

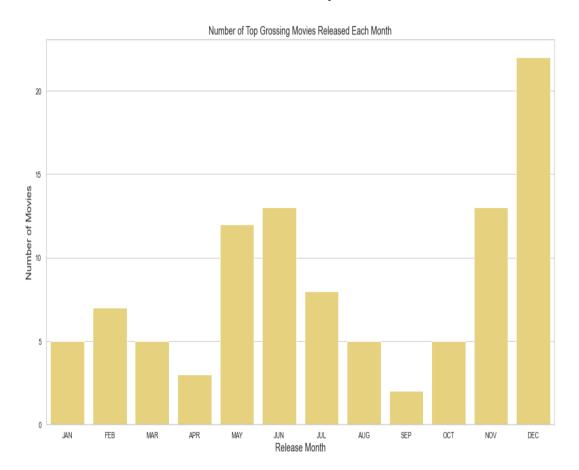
- The top three all-time grossing genres are Action, Adventure, and Comedy.
- Almost 64% of the 100 top grossing movies of all time fall into the top three, all-time grossing genres of Action, Adventure, and Comedy.
- More than 50% of all movies produced are in the genre categories of Drama, Documentary, and Comedy.





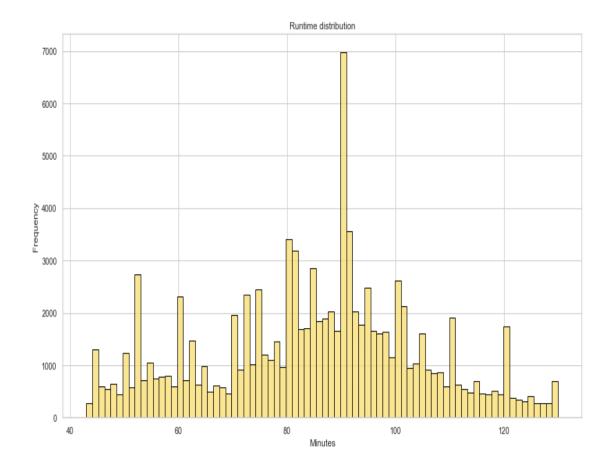
Release Date

- The most successful months are May, June, July, November, and December.
- Most movies are released on a Friday.



Runtime

- The average runtime is 87 minutes.
- Top grossing movies run between 85-125 minutes.
- Movies that run between 85-125 minutes have a larger adjusted average domestic gross.



Actionable Insights

1. Produce a movie with a runtime between 85-125 minutes.

- 2. Aim for a May/June or November/December release date.
- 3. Produce a movie with a genre of action or adventure.
- 4. Plan for a budget of \$44-\$136 million.

Next Steps

While providing valuable insights, the analysis acknowledges the need for further investigation to address nuances and ensure a comprehensive understanding of movie dynamics influencing financial success.

- 1. In-Depth Genre Analysis
- 2. Mixed-Genre Film Exploration
- 3. External Factors Impacting Release Timing
- 4. Consumer Preferences and Genre Evolution
- 5. Comprehensive ROI Forecasting Model

These next steps aim to address identified inconsistencies, enhance the robustness of the analysis, and provide more actionable insights for the film company's decision-making processes.

Author

Name: Erin Wasserman

GitHub: Cellister

Email address: cellister at gmail .com

Repository Structure

Jupyter Notebook

The <u>Jupyter Notebook</u> is the key deliverable and contains the details of my data strategy, methodology, data cleaning, visualizations, and actionable insights.

Presentation

This 5-7 minute, non-technical <u>presentation</u> was made in <u>Canva</u> and gives an impactful and brief overview of the key insights and recommendations.

Data

The data used in this analysis can be found in the 'Data' folder. Some