

Phase 1 Project

Shilton Soi

Summary

My project involves conducting analysis on data from IMDB and Box Office to obtain information such as revenue and ratings with respect to genre, film types and studios. This is in order to draw appropriate conclusions in order to inform the studio's decisions. For instance, I have discovered that short films bring in the most revenue, therefore one of my recommendations is that the studio begins by churning out short videos to bring revenue and to build popularity.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. Analysis needs to be done on what types of films are currently doing the best at the box office. Findings must then be translated into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Data

Datasets used:

I used three datasets;

- Title basics – Data from IMDB, contains movie titles, year, run times and genres
- Title ratings – Data from IMDB, contains ratings and number of votes of movies
- Movie gross – Data from Box office, contains movie titles, studios, revenue and year.

Data that was used from these sources include: run time, genre, average rating, domestic gross and studio

Methods

Data Preparation

- To get data from all three datasets into one data frame, I used the merge function.
- Using the drop function, I dropped unnecessary columns. After checking for missing values, I dropped them using the dropna function. Finally, I checked for duplicates using the duplicated function

Feature Engineering

- I defined two features, rating (poor to excellent) and film type (by runtime, short to extended)

Outlier Handling

Methods

Outlier Handling

- I used a boxplot and distribution plots to identify outliers, used the Z score to outline the Interquartile range, then used the loc function to drop them.

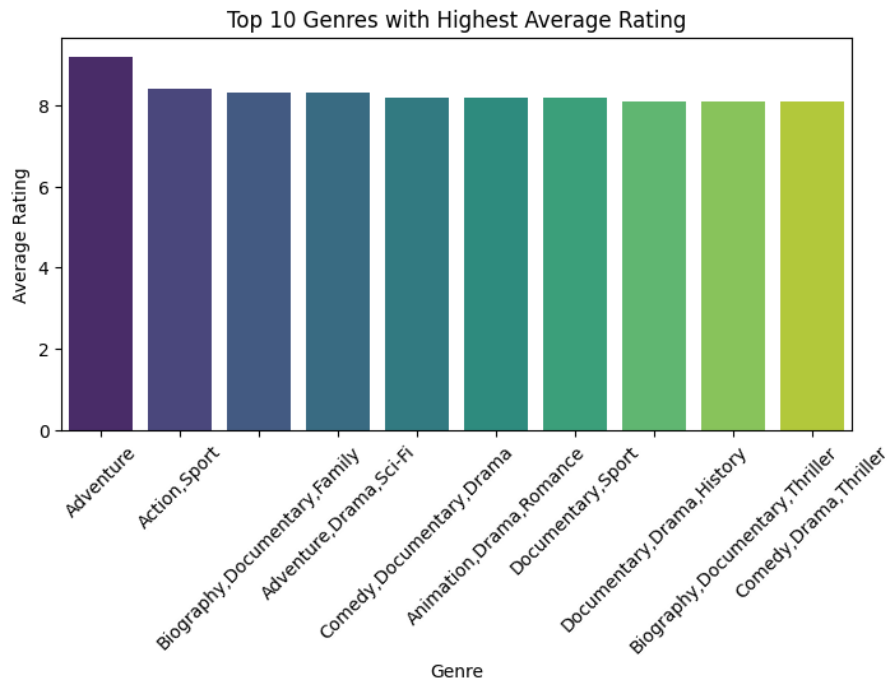
Analysis

- I grouped data genres, rating, studios and film type and sorted by average revenue and average rating

Results

Top ten genres rating

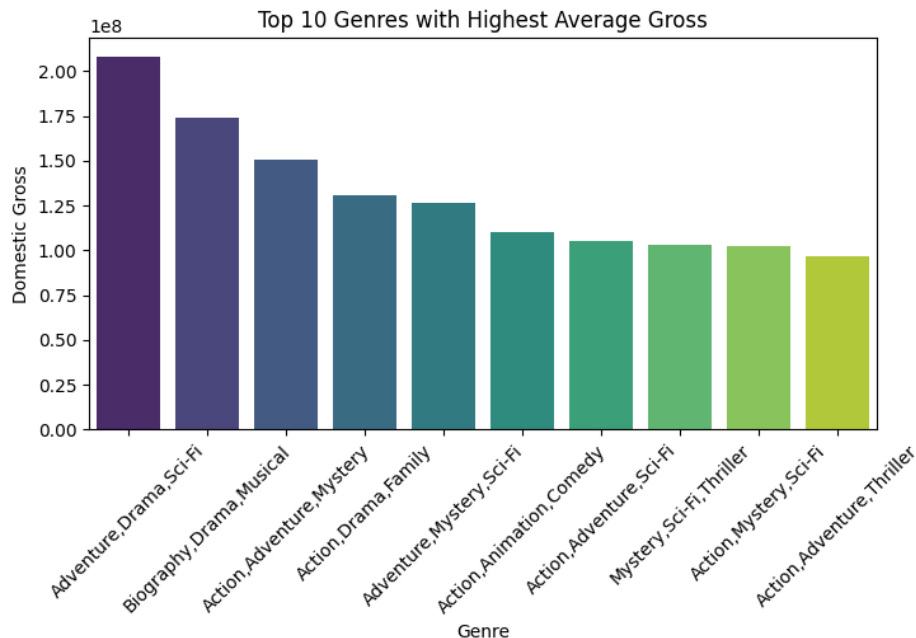
- Adventure genre had the highest average rating of 9.2
- Action, Sport genre followed at an average rating of 8.4



Results

Top ten genres revenue

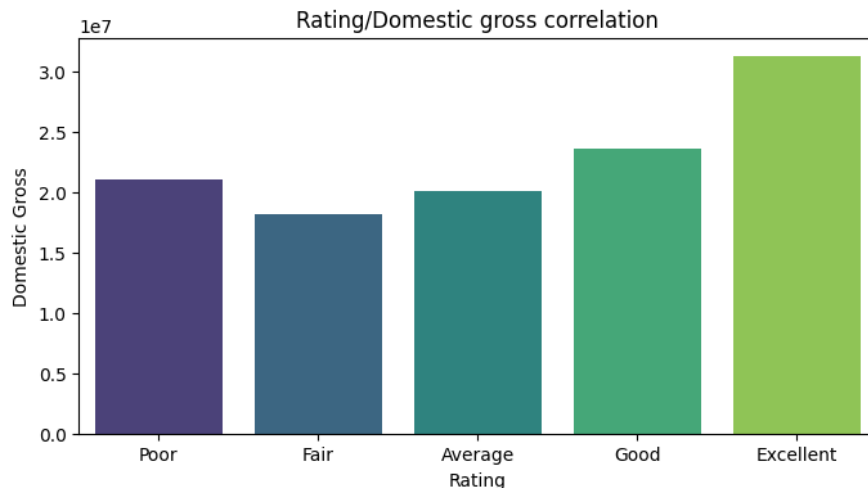
- Adventure, drama, Sci-Fi genre had the highest average gross of 208 million
- Biography, drama, musical followed at 174 million



Results

Correlation between rating and revenue

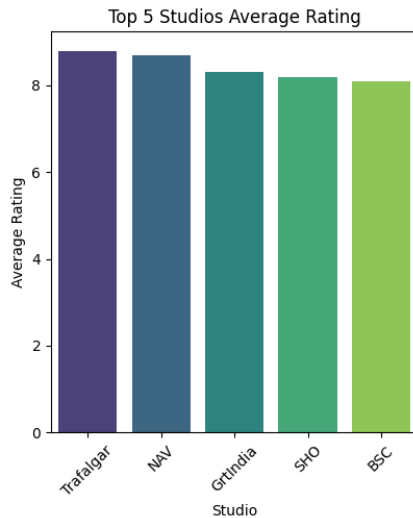
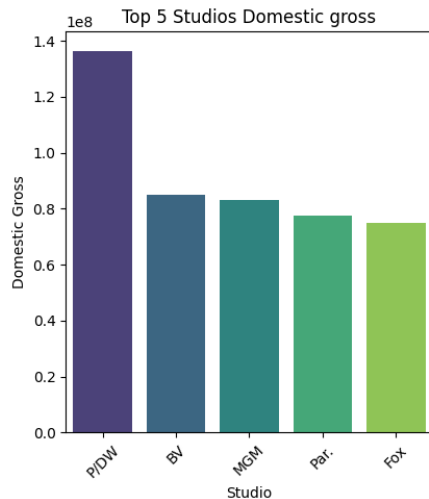
- Movies with excellent ratings had better revenue than those with lower ratings
- Movies with poor ratings had better revenue than those with fair and average ratings



Results

Top five studios revenue and rating

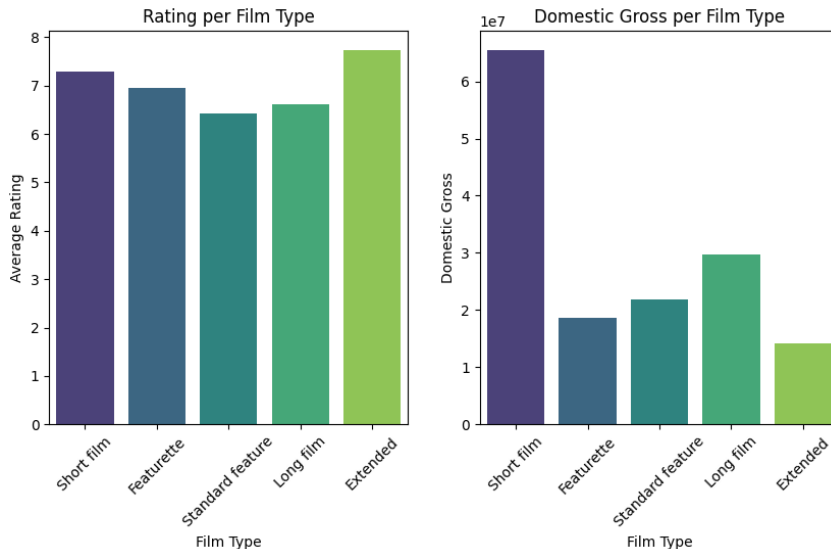
- Pixar/Disney World had the highest revenue of \$136 million
- Trafalgar had the highest average rating of 8.8



Results

Revenue and rating per film type

- Extended films had the highest rating at 7.7, closely followed by short films at 7.3.
- Short films was by far the best grossing film type at \$65 million. Long films followed at \$29 million



Conclusions

- Short films would be a good way to start releasing production from the studio. Not only are they highly rated but also, they have proven to be the most successful financially.
- I recommend creating content based on the Adventure, drama, sci-fi genre to maximize on revenue.
- I recommend bench marking with Pixar/Disney world studio as this is the highest grossing studio, beating the next best studio by over \$50 million

Conclusions

- I only used domestic gross in this analysis, therefore I cannot give accurate foreign revenue projections.
- For future improvement, I will get more data and do analysis on foreign revenue

Thank You!

Email: shiltonsoi@gmail.com

GitHub: @ShiltonTK

LinkedIn: [linkedin.com/in/ShiltonSoi/](https://www.linkedin.com/in/ShiltonSoi/)