

Movie Classification & Analytics

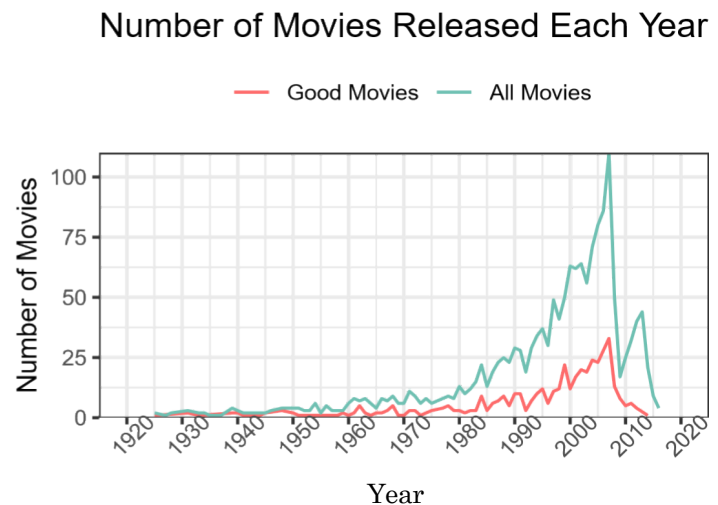
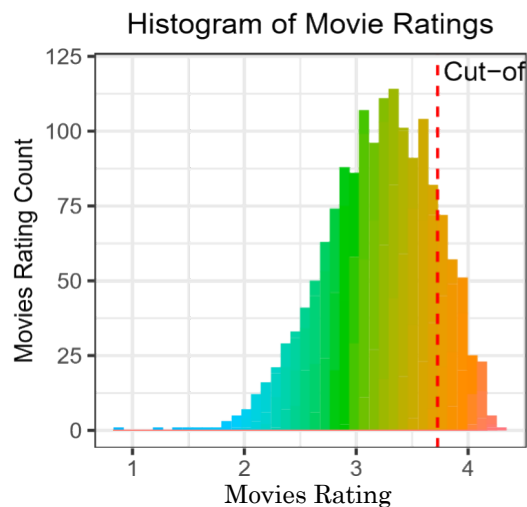
Authors: Yin, Karolina, Emma, Rachana, Nupur, Sarah

2023-02-18

Raw Data Summary and Cleaning

The Full MovieLens Dataset has 26 million ratings and 750,000 tag applications from 270,000 users on all the 45,000 movies. Before training predictive models, we cleaned the raw data by dropping irrelevant columns or columns with over 50% missing values, eliminating visually extreme values and created new metrics/features (weighted rating, ROI, profitability, etc.)

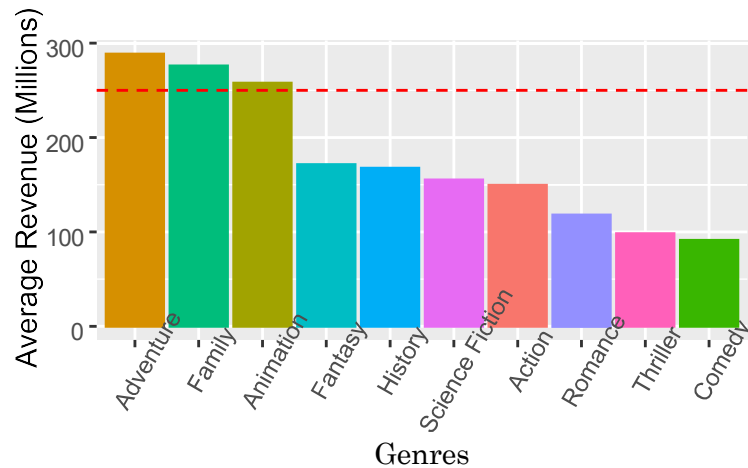
Data Analytics



Understanding Movies Rating and History

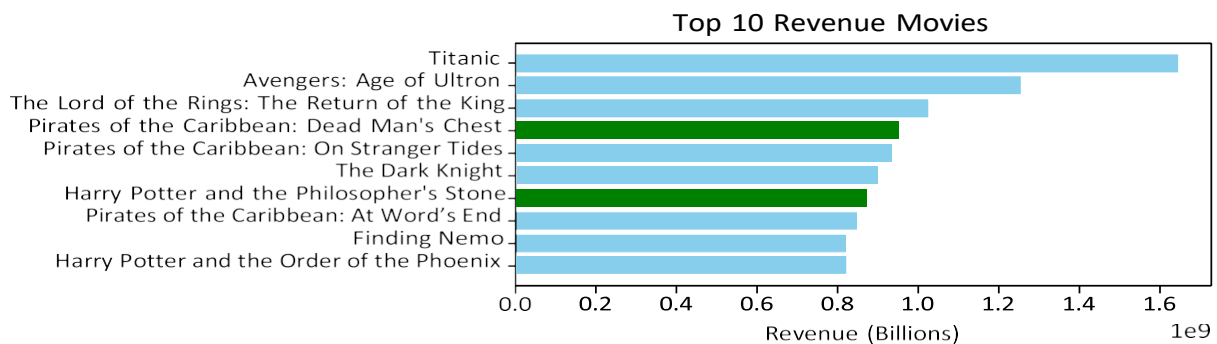
Based on IMDB's weighted rating (wr) formula, we created a new fairer feature "weighted_rating" which takes number and average of voter ratings into consideration during calculation. Having observed below histogram, it is safe to argue that vast majority of movies rating are above 3.0. Therefore, in order to set apart "good" movies, we need to establish a criteria determining ultimate result, which is "if weighted_rating > cut-off 3.54, and vice versa.) Specifically, the cut-off is derived from 85th quantile of weighted rating, meaning if a movie's weighted rating > 3.54, then it is better than 85% all movies, thus rendering itself as "good movies". To identify factors that may influence a movie's rating, we considered various explanatory variables, such as the cast, budget, production company, release date, number of competing movies released at the same time, and movie length (runtime)

Top 10 Genres by Average Revenue



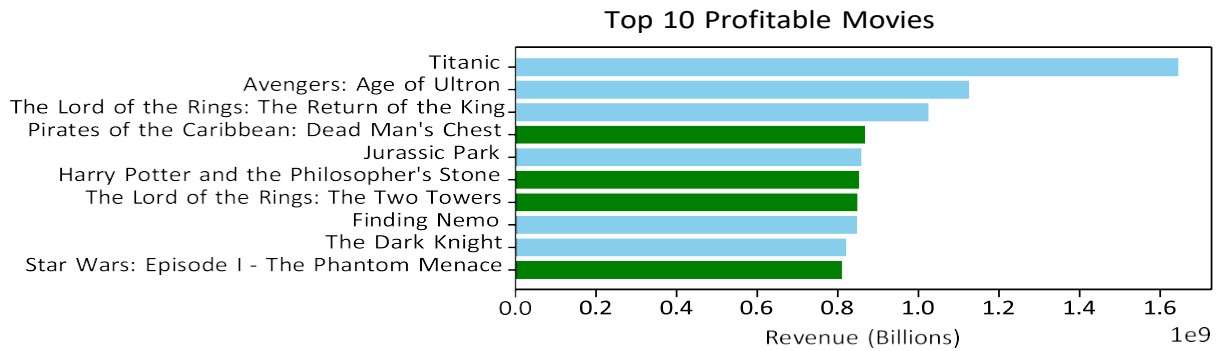
Findings on revenue vs genre:

Adventure, Family, Animation genres generally generate the most amount of revenue. It is interesting to notice that Action, Adventure, and Drama genres also score top 10% in terms of the length of the movie, which can result in higher popularity due to length preferences in comparison with other genres. This probably resulted in increase in revenue.



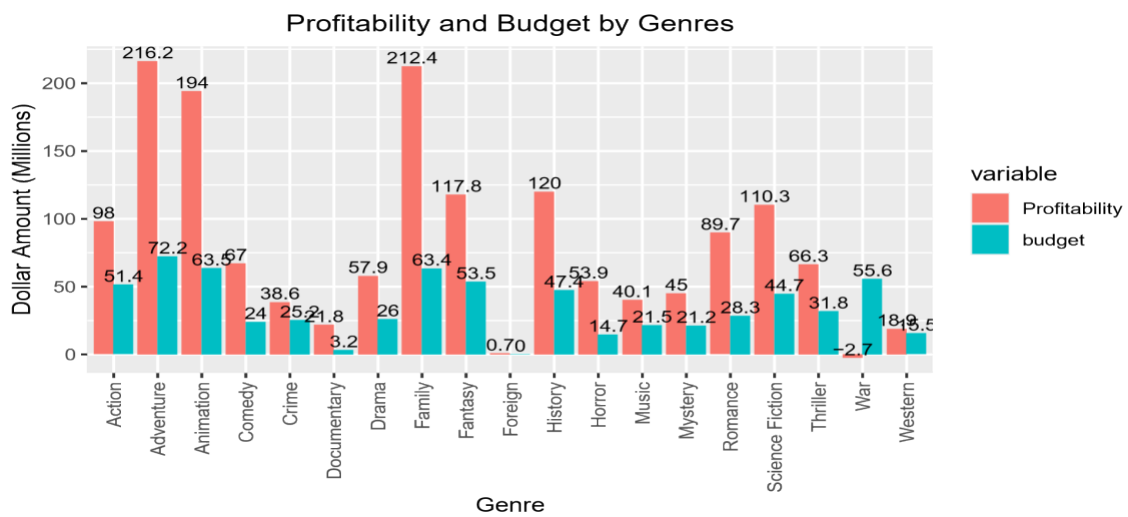
Findings on revenue vs movies:

For Top Revenue Movies it turned out that Titanic and Avengers scored highest for the revenue but 2 movie Pirates of the Caribbean Dead's Man's Chest, Harry Potter and the Philosopher's Stone had highest ratings above 3.54 considering they are a good movie. This can be attributed to the cause of different audience demographics with varying preferences but interestingly Pirates of the Caribbean might have highest rating due to its genre adventure and its length which can be related to the insight in 2.2.1



Findings on profitability vs movies:

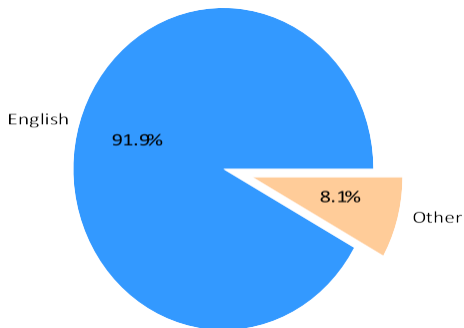
Top Profitable Movies that present four good movies with rating > 3.54 are marked in green. These four movies fall under the genre of Adventure. Interestingly, Adventure genre again appears to be on top in terms of Profitability supporting insights 2.2.1 and 2.3.1.



Findings on Profitability/budget vs genres:

Adventure, Family, and Animation genres have the highest profitability. Interestingly, Adventure genre again shows that not just in tops in revenue but by spending small budget- profit and revenue can be easily maximized in this genre. This again supports all the above insights.

Movies Count by English Language vs Other



Movies Count by Other Languages

