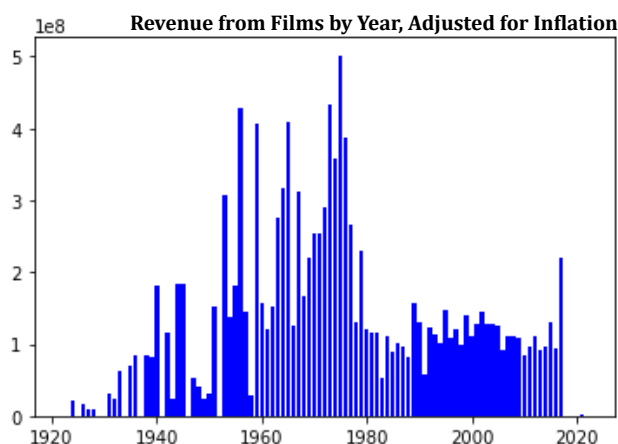


Translating Box Office Outcomes

Erdős Project Group *Basswood*, Fall 2022 Cohort
Sridhar Venkatesh, Kaitlin Engelbrecht McClamrock, Reebhu Bhattacharyya

Would you be interested in knowing which films are likely to be box office hits? Movies generate billions (recently nearly \$42 billion in 2018) but in a changing ecosystem, predicting hits is even more important to investors and production companies than ever before. Using machine learning techniques, we work to predict overall worldwide box office revenues for past films in order to develop a model that could help stakeholders determine which films are likely successes.



Making use of data on thousands of films via The Movie Database, we develop models predicting box office revenue with variables such as the actors involved, genres of the films, budget, release dates, production companies, and more. We use box office revenue as the outcome of interest as this is considered the prototypical reference for film success and find that revenue is highly dependent on films' budgets. In extreme outliers we see films with lower (reported) budgets perform well at the box office, but most important to box office success seems to be investment in the film's production via budget. Prediction of revenue is done using polynomial regression where our estimates can be sensitive to multiple outliers (which are present in our data). Here we demonstrate via heatmap which predictors are most associated with a film's box office success.

