

Movie Revenue Proposal

Movie production can be extremely costly: a large investment risk with highly variable results. But what makes a blockbuster? We will produce a model for predicting world-wide movie box office revenue in advance based on a variety of movie features that can be optimized and customized pre-production in order to minimize the investment risk involved in undertaking the production of a new movie.

Production companies and studios are the most obvious stakeholders with an interest in the problem of predicting movie revenue based on the movie's objective qualities. Other individuals or interests involved in movie production would also have a stake in the problem. Our key data source comes from kaggle's collection of movie information gathered from the movie database (themoviedb.org), which also includes movie data from imdb.com. Data includes cast and crew information, plot keywords, budget, release dates, languages of production and release, production companies, countries, and more.

We are constrained in scope by the subjectivity of many movie features (what makes a movie "good"?) that cannot be included in our model. We will attempt to proxy for these by using the popularity metric, which can only be determined after release of a movie, but is the closest proxy for a subjective measure of movie quality. We are constrained within our solution space by the limits of the available information on the database: some movies, particular older movies, have less data available, and our dataset includes only 3000 movies. Additionally, the dataset is very US and anglophone-centric, which is not an issue when examining world-wide revenue but could potentially pose an issue if one were interested in examining the film industries of particular countries in depth separately.

We aim to create a model for movie revenue that highlights the modifiable features affecting it, therefore we will aim to produce a model that will predict movie revenue based primarily on factors that could be modified or taken into account before movie release to predict the return on investment expected in box office revenue, but still including a popularity metric as it is our only way to score the more subjective features of movie quality.