

Capstone Project 1: Proposal

Problem Statement

The dirty secret in the movie business is that less than half of all films released will turn a profit. Everyone has their hand in the jar to take a piece of the revenue. From top ranked actors who score percentages of box office draw to distribution companies, movie houses, and promotional networks, the pot gets ever smaller before the investors can start to recoup their costs. The return on investment for a production company is barely sustainable.

One group that feeds from the movie revenue sources are the actors. Many times, perceived star power allots too much money to one individual in a film. It is a big risk to invest so heavily in the revenue generating abilities of a select few people. After all, most actors, even great ones, will put out a bomb.

Investors should wisely decide how much to invest in the leading roles of a film, while comparing this expense to how much revenue the film is expected to create. When looking at which actors could favorably tilt this balance, an actor's value should be largely based on the return on investment generated by that actor's previous films.

Can the financial success of a movie be fitted on its actors to give them an additional tangible quality? Can this value be quantified to make predictions of an actor's value? Can a production company use these predictions to make better investment decisions on who to hire for their films?

Possible Clients

The types of clients that would be interested in the results of an actor value prediction model would be anyone involved in the hiring of actors. They would include, but not be limited to casting directors, directors, and producers in the movie industry. Agents may also want to glean some insight into their competition for acting roles, as well as assess the value of their own clients.

Dataset

I will build my own dataset from API requests from the TMDb website. The people at TMDb make extracting their vast set of movie information very easy to do. They will gladly provide an API key, which can be used to collect their data. They have an extensive user guide to help navigate their website that includes the ability to test out a request to see what the response will be ahead of time. Also, they've recently lifted their request limits. So, data acquisition is worry free. Their data is extensive, as well. It includes characters, billing positions, release dates, genres, casts, crews, imdbID's, budgets, information on whether movies belong to a collection, and revenues. For actor data, they have dates of birth, imdbID's, and filmographies.

Solution

The unique idea of this project is assigning numerical quantities to actors, which can not be inferred from a quick scan of their acting credits. I will collect the return on investment figures that were seen by the investors of all the movies for each actor. These figures will be assigned to their respective actors and weighted based on the place those actors occupied in the billing order of their movies. A higher billing order position will result in a larger weight on the financial figure to be transferred to the newly created actor centric quantity. This will be the target variable of various regression models. The goal of this project is to find the best fit regression model for predicting the actor values. Comparing these values with the expected expense of hiring that particular actor will help investors in movies find their best fit actor.

Deliverable

The final product will be available through a github repository. It will contain a paper with all of the relevant analysis detailed with proper visualizations and a compelling data story, along with the companion code that produced the results.