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 draws to distribution companies, movie houses, and promotional networks, the pot gets ever smaller. All of this happens before the movie investors can start to recoup their costs. The return on investment for a production company is barely sustainable.

One group of people that feed 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. All the while, they should be 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 revenue generated by that actor's previous films. This track record should be analyzed in the context of the specific qualities of both the actor and the actor's previous films. This will help to ensure that the actors hired for the top billing roles are the ones who are the best fits for the scope of the cinematic project.

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 to future works? 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 the 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, runtimes, proprietary review scores, review counts, genres, casts, crews, budgets, information on whether movies belong to a collection, and revenues. For actor data, they have dates of birth, locations of birth, gender, proprietary popularity scores, and filmographies.

Solution

The unique idea of this project is assigning numerical quantities to actors. These quantities can not be inferred from a quick scan of their acting credits. Most projects in this domain focus on predicting successful movies. Instead, I will aggregate the movie revenues by each actor in each film, then assign that value back to the actor. This actor centric value will be the target variable of a regression model search. The goal of this project is to find the best fit linear regression model for predicting these actor values. Weighing these values with the expense of hiring any particular actor will help movie investors 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. To accompany this paper, I will provide a slide deck and the companion code that produced the results, as well.