Predicting the Revenue of Hollywood Movies

Purpose: The purpose of my model is to analyze movie data and to check if there is a correlation between certain factors in a movie and its success rate (in terms of revenue). My hypothetical client here is Bestflix Studios, and they are looking for quantifiable variables, that’ll increase the probability of success (in terms of revenue) in the American Movies market.

Data: I’ll be web scrapping data using Beautiful Soup and Selenium from Box Office Mojo and IMDB. As mentioned by Chris in class, I’ll base this analysis on dataset of 1000 movies, with an aim to understand the effects of actors, movie location and content on the gross revenues collected by domestic movies. I will use a Linear Regression model for this project, to correlate multiple variables to the revenue collection of movies on my dataset.

Tools:

I’ll use libraries like Pandas and NumPy for Data Manipulation. I’ll use Matplotlib and Tableau to Visualize my findings. Statsmodels and Scikit Learn for testing my Linear Regression Model/s.