In this project I look to see how accurately one can model the IMDb ratings of any movie based on its basic properties. This problem showcases multiple simpler questions that I wanted to ask, such as how does the region (country where movie was made/released) affect ratings? is there a particular region more inclined to certain types of movies? What kind of effects do R-ratings on movies have? Is there a preferred length for a movie? Do certain genres do better than others?

To answer such questions, I looked at 4 IMDb datasets

1. The basics dataset: Contains the basic information about each movie including language, type, attributes of the movie
2. The titles dataset: Contains extra details including title, release year, runtime, genre, adult-rating
3. The ratings dataset: Contains the average rating and number of votes for each movie
4. The crew dataset: Contains the directors and writes of the movies.

To focus on a particular question, I would be looking at which movies are the best to make, from a ratings perspective. In doing so, I would explore what kind of movie I should be making, where I should release it and which crew members are likely to give me the best results.

Disclaimer: As someone who loves the worldbuilding and storytelling aspects of movies, I understand that there is much more to the art than what this analysis can explain. Some of my favorite movies are never given much acclaim and some are downright rejected by many of these rating communities. It is important to understand that movies are subjective, and many are much more appreciated later on by varied audiences. There is an X-factor in movies that is lost when they are made purely for ratings and profits, and thus this analysis is not meant to discourage or encourage any particular type of movies, but rather see the trends and predict what movie will get a fictional studio the best rating.