

# HW2 - IMDB Analysis

2nd December 2021

Falomy Eugene (48686864), Syed Hassan Raza (48372523), Allen Mattam (48686903), Adhithya Madhavan (47705259), Srikanth Soma (48692883)

## Introduction

*Our goal is to find out what makes a movie great.*

In this report, we will be exploring the impact and the relationship between various factors to determine what makes a “good movie” given its rating score. To better understand the impact, we will consider factors such as the collection to which a movie belongs, the cast of a movie, the directors, the production house, the duration, and the popularity of the movie in question. This understanding will help us predict “good movies” and “bad movies”. We will be looking at movies from 1906 to 2014.

## Executive Summary

Using the factors mentioned above, our model suggests that having a top movie director, cast (actors), and a top production house while producing a movie, has a significant influence on the success of a movie. We also noticed the popularity of a movie can be due to the marketing efforts that the movie crew implements prior to the release of the movie. A high popularity does not necessarily translate to a high rating. We will explore the marginal effect of each of these factors below.

**Our analysis has brought to realization that:**

1. *The probability of having a good movie increases by 85% with a “good director”.*
2. *The probability of having a good movie increases by 74% with a “good cast”.*
3. *The probability of having a good movie increases by 60% with a “good production company”.*

## Understanding the Analysis

### *Movie Classification*

To determine whether a movie is good or bad, we created an IMDB formula based rating which deals with or takes into consideration certain groups that could disproportionately influence the rating. The rating is a number between 1 and 10, and we decided that **any movie with a rating greater than or equal to 6 is considered a good movie, otherwise it is classified as a bad movie.**

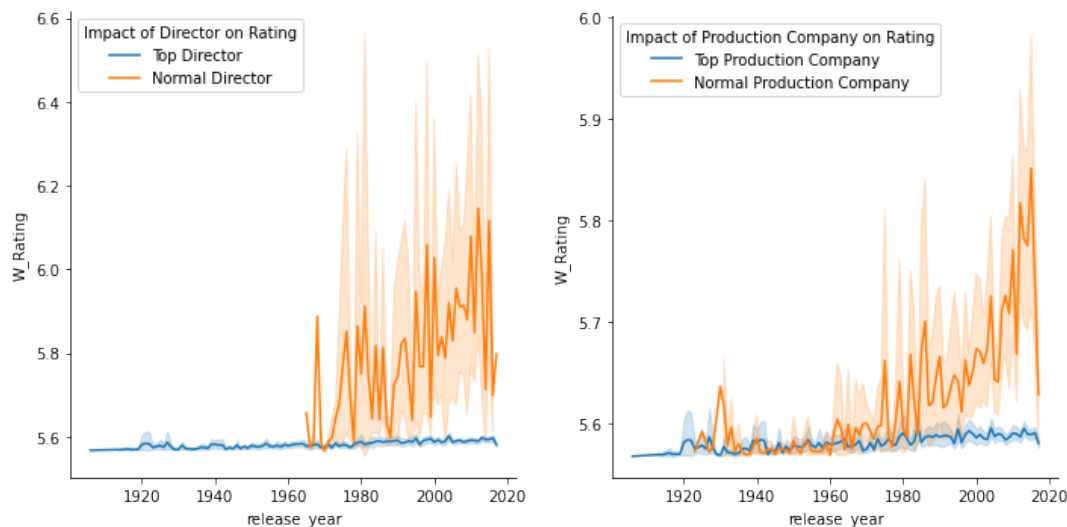
## Assumptions

We developed a number of assumptions in the hopes of capturing the causal effect of each factors. According to Aristotle, “the whole is greater than the sum of its parts”. We do realize that it’s difficult to state that a certain factor has an absolute impact on the end result without considering the impact of all the factors collectively. We also do acknowledge that certain factors not included in our analysis may have an affect on our model(endogeneity).

We assumed that the top 22 directors according to IMDB as “Good Directors” relative to other directors in the list. Also, we chose the top 70 actors according to IMDB ranking as “Good Actors”. Finally, we chose the top 16 production houses as “Good Production Company”.

## Directors/Production House and ratings

### Understanding Descriptive Relationship

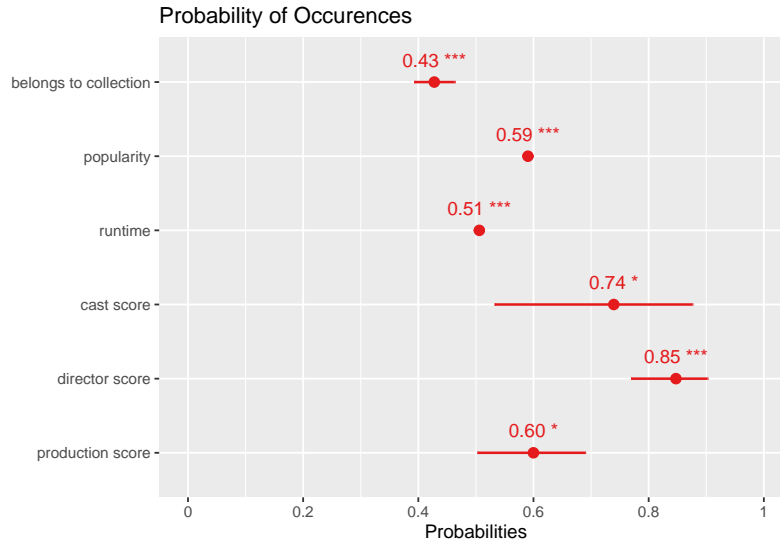


The graph on the left displays the rating of directors across the time period. The segment that we chose as “good directors” (orange line) have significantly outperformed other directors (blue line). Also, the graph on the right displays the rating of production houses across the time period. The “good production houses” (orange line) that we selected, also outperformed the other production houses (blue line). This shows a level of consistency with the assumptions that we made.

From a descriptive analysis, we segregated released movies given that movie in question is being produced by a certain type of production house and directed by a certain type of directors. We observed that the better the director and the production house, the better the rating of movies. We wish to further explore the probability of having a good movie while considering the impact of good directors and good production houses.

## Regression Analysis

While considering our model, having a top director directing the movie, and a top production house producing it, featuring star actors, its popularity, and its runtime increases the likelihood of it to turn out to be a good movie. Whereas it does not help if the movie belongs to a movie series.

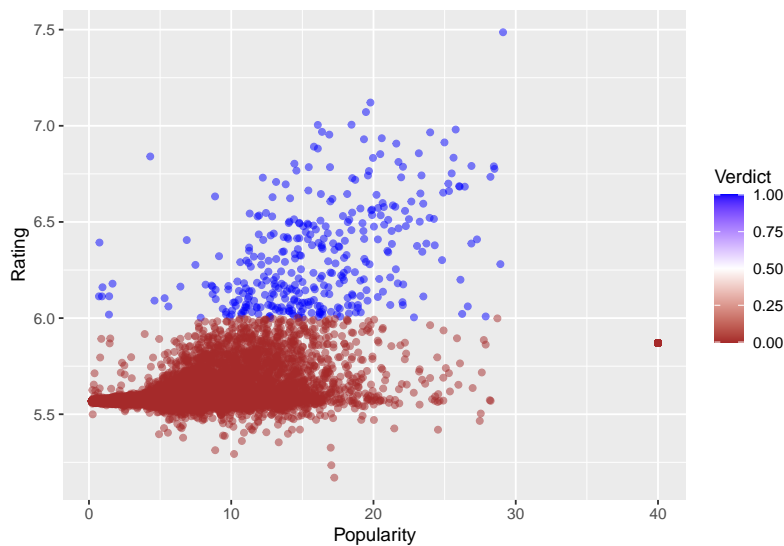


## Marginal Effect

To determine whether a movie is expected to be good movie, we considered six major factors, which will be explored below.

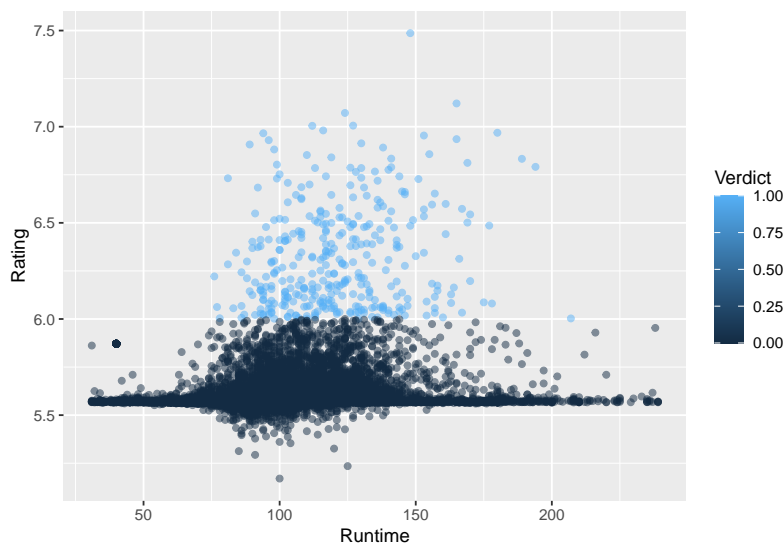
We observe that there is a negative relationship between the fact that a movie belongs to a series and the rating of the movie. Based on our model, assuming that everything remains constant, the probability that a movie is bad is 57% if it is part of a movie series. However, this goes against the common perception and there could be unaccounted factors playing into this or it is just that the count of bad rated movies belonging to series is higher than the popular ones.

Moreover, we notice that our model is highly consistent with what was anticipated earlier. The star actors in the movie, the director, and the producers of the movie have a significant influence on the success of the movie by 74% , 85%, and 60% respectively.



Another interesting observation we thought was worth mentioning; There are many movies that have very high popularity but have very low ratings and vice versa. Given that we use the calculated rating to determine

whether a movie is good or bad, we assumed that this might be due to movies that are marketed by pushing a lot of money to create a lot of hype from the start, or by having a very high rated actor in the movie, but doesn't gain a lot of viewership after the movie is actually released. However, our model suggests that the popularity of a movie means that the movie is 59% probable to be a good movie.



## Recommendations

There are certain important factors that have a huge influence on the success of a movie. Although great cast, production and direction will add value, they come at a high cost. Having a great director had a better impact than cast and production (according to our model) and could be a balanced trade off between expense and success factors. Interestingly, cost of hiring a good director is (in most cases) less than hiring a good actor. For example in Inception (top movie according to our model), the director Christopher Nolan made half (~27 million USD) compared to Leonardo Di Caprio (~50 million USD). Other good factors that contribute to the success of the movie:

- 1) Run time between 90-150 minutes
- 2) Genres: Animation, Adventure and Family Movie (for great ROI)
- 3) Popularity: IMDB calculates this variable using its proprietary hidden formula. This variable is calculated before the release of the movie and influences the marketing and movie trailer plans. So, in other words we recommend analytics based approach on marketing.

## Conclusion

Our goal was to predict whether a movie is good based on various factors. We used an IMDB ranking based formula, which takes into consideration groups that could disproportionately influence movie ratings. We observed that it is critical to have great directors to create vision, top actors and top production house for the movie's success, in that order. Popularity could be a factor we like to be cautious about as it could have correlation with other factors discussed above. However we must also keep in mind that there might be other factors that are not included in our model.