

A Study of Ratings Algorithms

Kendra Ficht

La Cueva High School, Albuquerque, NM

RESEARCH QUESTION

What factors enable a movie to get high ratings on online sites such as IMDb?

MOTIVATION & INTRODUCTION

As of recent, many have started speculating as to what kinds of movies do get high ratings on online movie rating sites. These online ratings are of particular interest to people as these ensure the continued sale of DVDs of movies well beyond their release dates thereby generating revenues for the production houses. They also generate revenue for external well-known critics whose reviews are used to help people decide which movies to watch.

Earlier people would conveniently attribute high online ratings to good plot and stellar acting by the cast. However now there has been a shift in focus since over the years it has been seen that even with good acting skills, the ratings of movies can flat if the cast is not well known or the set is not well designed due to budget constraints.

In light of this observation, we set out to determine how factors extrinsic to the design of the movie influence IMDb ratings.

DATA COLLECTION METHODOLOGY

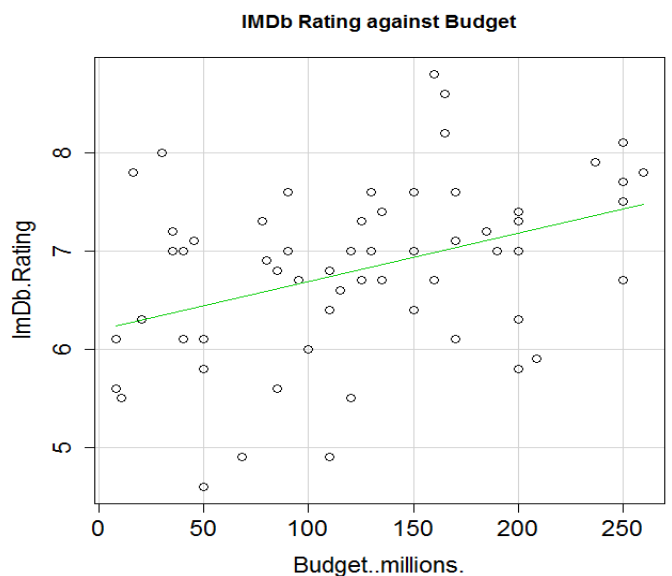
For our model, we randomly chose 60 Hollywood movies released since 2009. We restricted the time of release since factors such as budget and actor's ranking would not have been comparable otherwise. We used reliable secondary data sources to gather information on each of our variables before setting out to employ statistical techniques on them. The table below shows the variables are problem will be focusing on:

Numeric Summary of the variables

| | mean | sd | IQR | 0% | 25% | 50% | 75% | 100% |
|-----------------------|-------------|-------------|--------|----|-------|-------|--------|------|
| Actor.s.Ranking | 394.6833333 | 357.2014335 | 554.25 | 5 | 56.00 | 320.0 | 610.25 | 1137 |
| Budget..USD.Millions. | 124.4100000 | 70.4144479 | 94.50 | 8 | 75.50 | 122.5 | 170.00 | 260 |
| Countries.Released | 65.4000000 | 19.0558501 | 18.25 | 8 | 58.75 | 69.0 | 77.00 | 101 |
| Duration..mins. | 114.4666667 | 24.0729588 | 35.00 | 46 | 95.00 | 115.5 | 130.00 | 169 |
| Genre..Action. | 0.1166667 | 0.3237318 | 0.00 | 0 | 0.00 | 0.0 | 0.00 | 1 |
| Genre..Fantasy. | 0.2166667 | 0.4154502 | 0.00 | 0 | 0.00 | 0.0 | 0.00 | 1 |
| Genre..Sci.Fi. | 0.3666667 | 0.4859611 | 1.00 | 0 | 0.00 | 0.0 | 1.00 | 1 |
| Nominations | 38.7666667 | 46.4908836 | 41.50 | 1 | 9.00 | 22.5 | 50.50 | 229 |
| Wins | 15.6500000 | 24.5169087 | 18.25 | 0 | 1.00 | 7.0 | 19.25 | 143 |

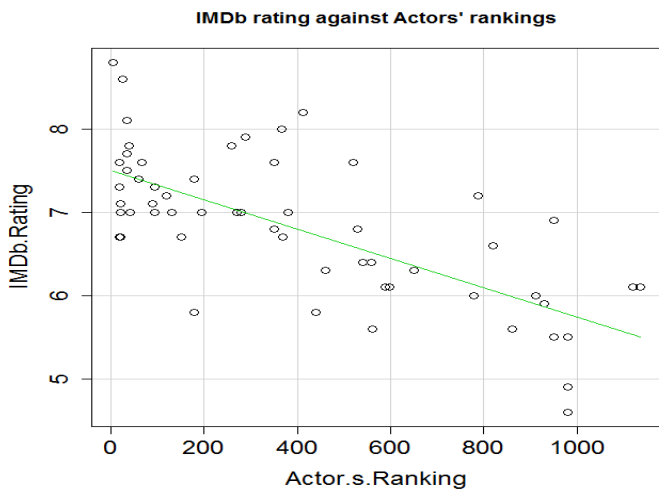
MOVIE'S BUDGET VS. IMDB RATING

the scatter plot to the left shows the relationship between a movie's IMDb rating and its budget. Data points lying below the best fit line suggests that even high budget movies are sometimes not able to garner high IMDb ratings. However overall there seems to be a positive linear relationship between the budget of a movie and its ratings. This observation was expected due to the fact that high budget means better set design, background music, 3-D effects and actors can be arranged for. These usually result in a good movie being made.

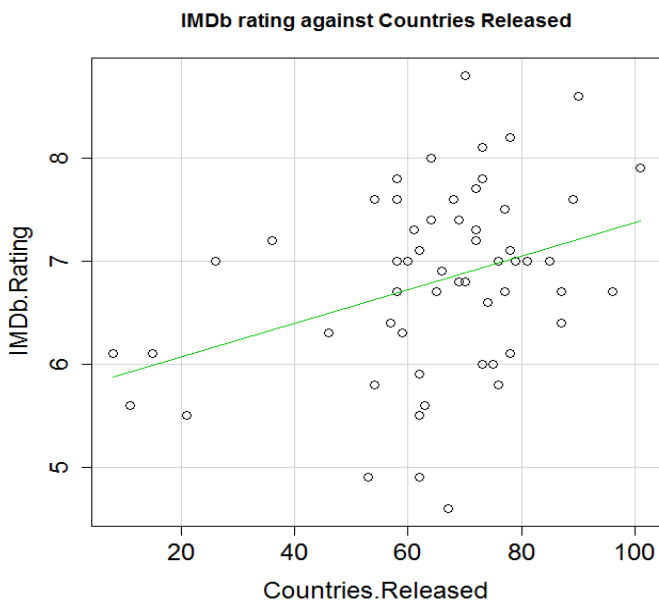


LEAD ACTORS' RANKINGS VS. IMDB RATING

The graph below shows that an inverse linear relationship between lead actors' rankings and the movie's IMDb rankings can be approximated. This conforms to a general understanding that the smaller the numeric ranking of an actor is the better does the movie in box office. This can be attributed to the fact that actors with a smaller rank number are better known to the audiences and are generally applauded for their acting skills in many award shows.



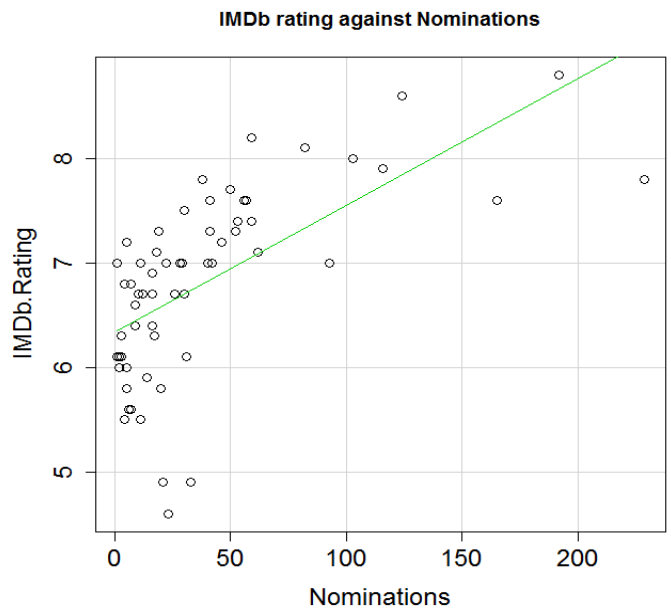
NO. OF COUNTRIES THE MOVIE IS RELEASED IN VS. IMDB RATING



As the graph above shows, most movies were released in 60 to 80 countries. Generally, the greater exposure a movie got in terms of countries it got released in, the greater were its IMDb ratings. So, a positive relationship does seem to exist between the two.

NOMINATIONS VS. IMDB RATING

The scatter plot to the right shows that most movies received 0 to 50 nominations. For movies with greater number of nominations than 50, the IMDb rating didn't fall below 7. Also, overall a positive relationship does seem to exist between the number of nominations and the movie ratings. The greater critical acclaim a movie gets the more does the audience rate it high for either depicting a topic that is close to heart or for larger than life effects.



FINAL FITTED EQUATION

Fromm Regression model given below (in exhibit), we got this final fitted equation: $\hat{Y} = 7.47 - 0.0015X_1 + 0.0015X_2 - 0.0056X_3 - 0.423X_4 - 0.401X_5 + 0.0152X_6$

Where: X_1 = Actor's Ranking

X_2 = Movie Budget in USD Millions

X_3 = Number of countries released

X_4 = Genre: Action (Yes = 1, No = 0)

X_5 = Genre: Fantasy (Yes = 1, No = 0)

X_6 = Number of award wins

INTERPRETATION

The intercept represents the static value that is to be expected even if all the X coefficients equate to 0. The coefficients of all the X variables represent the estimated change in Y to be observed due to a one unit increase or decrease in the X coefficients.

For instance if there is a one unit increase in Actor ratings, IMDb ratings would decrease by 0.0015, given that all other variables are held constant.

HYPOTHESIS TESTING

From the ANOVA Table we see that:

Residual standard error: 1.127 on 42 degrees of freedom
 Multiple R-squared: 0.6776, Adjusted R-squared: 0.3322
 F-statistic: 1.962 on 45 and 42 DF, p-value: 0.01476

INTERPRETATION OF HYPOTHESIS TESTING

Ho: $\beta_1 = 0$

Ha: $\beta_1 \neq 0$

Significance Level: 5%

$\alpha = 0.05$

P-Value: 0.01476

Rule: Reject Ho if P-Value $\leq \alpha$

Decision: Reject Ho since our p-value of 0.01476 is less than α of 0.05. Hence this variable is significant as $\beta_1 \neq 0$ at a 95% confidence interval. Since our P-value for the overall model is smaller than the significance level, we reject Ho. At least one independent variable has a significant relationship with the dependent variable.

CONCLUSION

Our model has identified the number of wins as the most important factor in the higher IMDb ratings of a movie. This suggests that viewer preferences are based on the winning popularity of the movie. Furthermore, genre also plays a significant role in decreasing the IMDb ratings, more specifically the Action and Fantasy genre. Actor's Rating is negatively related as the smallest ranking of an actor suggests more popularity and hence higher IMDb ratings.

EXHIBITS TABLE

| Variable | Variable Type | Definition and numeric code |
|---|---------------|--|
| Viewers'online rating of the movie | Response | Rating on a scale of 1-10 done on the online movie rating website named IMDb |
| Ranking of the lead actors in the movie | Explanatory | The ranking of the main actors of the movie as shown on the rankers website |
| Movie's budget | Explanatory | The budget of the movie in millions of USD |
| Genre - Fantasy | Explanatory | Whether the movie belongs to the fantasy genre Fantasy: 1 Not fantasy: 0 |

| | | |
|-------------------------------------|-------------|---|
| Genre- Sci- Fi | Explanatory | Whether the movie belongs to the Sci-Fi genre Sci-Fi: 1 Not Sci-Fi: 0 |
| Genre- Action | Explanatory | Whether the movie belongs to the action genre Action: 1 Not action: 0 |
| Countries the movie was released in | Explanatory | The number of countries the movie was released in as shown on its IMDb profile. |
| Nominations | Explanatory | The number of nominations the movie has received, as shown on its IMDb profile. |
| Wins | Explanatory | The number of awards the movie won, as shown on its IMDb profile. |
| Duration mins | Explanatory | The running time of the movie excluding intervals. |

REGRESSION MODEL

```
Residuals:
      Min       1Q   Median       3Q      Max
-1.17432 -0.26963 -0.05142  0.28307  1.14670

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.4657576  0.3389939  22.023 < 2e-16 ***
Actor.s.Ranking -0.0014624  0.0002516  -5.812 3.61e-07 ***
Budget..USD.Millions. 0.0015144  0.0012999  1.165  0.2492
Countries.Released -0.0056282  0.0047083  -1.195  0.2373
Genre..Action. -0.4232295  0.2219548  -1.907  0.0620 .
Genre..Fantasy. -0.4009941  0.1858239  -2.158  0.0355 *
Wins           0.0151676  0.0029291  5.178 3.53e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5276 on 53 degrees of freedom
Multiple R-squared:  0.686, Adjusted R-squared:  0.6504
F-statistic: 19.3 on 6 and 53 DF, p-value: 8.908e-12
```

REFERENCES

1. For actors' ratings: <<http://www.ranker.com/crowdranked-list/the-greatest-film-actors-and-actresses-of-all-time>>
2. For information about the ratings, wins and nominations of each movie: IMDB page of each individual movie <<http://www.IMDb.com/>>
3. For information about the list of movies: Excel data file. <<http://www.stephangisler.ch/docs/TheMovieExcelList.xls>>