Analysis Topic: What factors influence the IMDB rating? Is there a good predictor?

Methodology: By calculating the correlation of rating versus various quantifiable factors available on the IMDB movie page and looking at average ratings across categorical factors, we can show if there are any strong predictors of a movie's rating.

Data set: We are looking at the Top 250 movies on IMDB. Among the data available on IMDB, the following are used for the analysis:

1) Number of votes - we want to check whether the rating is related to the number of votes for the movie, which may be an indicator of the popularity of the movie

2) Runtime (length of the movie) – is there a relationship between how long the movie runs and how well viewers rate the movie?

3) Number of awards (won) – does the number of awards influence the rating of the movie? Here we excluded nominations.

4) Gross box office (worldwide) – Gross box office is also an indicator of the popularity of the movie; does it influence the rating? To simplify the analysis, I just used movies from the last 10 years as the dollar amounts are not adjusted for inflation.

Budget was not used because of multiple missing data and amounts given in currencies other than USD, although it would be interesting to see if the spending on the film correlates with how well it is received by the audience.

5) Genre – several movies belong to multiple genres. For the average rating, a movie is included in each of the genres it belongs to, so one rating may affect the average for more than one genre.

6) Release year – the rating given for the movie could be given at different points in time hence based on this data alone I cannot assess the correlation of the rating with the 'age' of the movie from release date i.e., I could not consider it as a quantitative data but only as a qualitative data

7) Origin – similar to genre, some movies may have multiple countries of origin based on the IMDB database, hence a movie is included in each of the countries it lists.
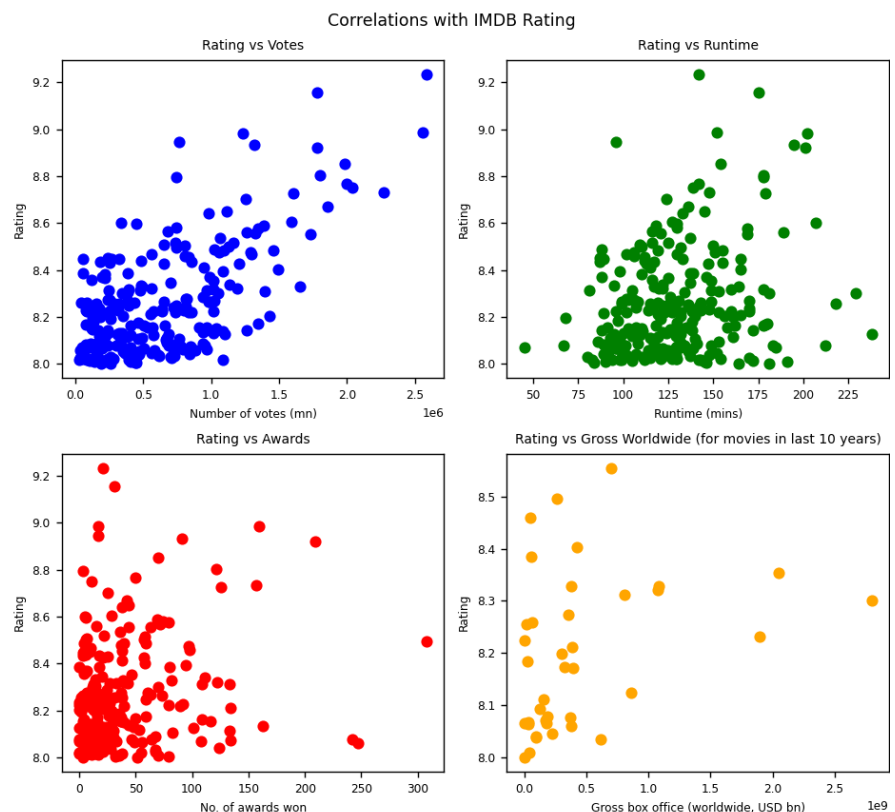
Results:

	Correlation
Rating vs number of votes	0.662262
Rating vs runtime in minutes	0.263311
Rating vs number of awards won	0.209234
Rating vs gross box office revenue (worldwide, USD)	0.334369

Data as of 21/05/2022, 23:37 (GMT+2)

Correlations with IMDB Rating

Based on correlation, there does not seem to be a significant predictor of a movie's rating. The number of votes has the highest correlation with rating, among the four factors we considered, and the scatter plots show this as well. This could mean the popularity of the movie influences the rating. However, the number of votes is not a characteristic of the movie itself that can enable one to predict how well it will be rated. The least correlated with rating is the number of awards won.

In addition, we try to look at average ratings across categorical factors of the movie: genre, release year, and origin to show whether a movie of a particular property tends to be rated higher.

The graphs show there are some genres which tend to be rated higher: Sci-Fi, Music, and then Action and Crime, although the average ratings are not significantly different.

There is no clear trend in the average ratings per year of release which could tell us generally which period produced the highest rated films. The spike in 1970's is driven by a single movie released in 1972.

The ratings per country of origin show that movies from New Zealand, on average, get the highest ratings. Upon inspection of the data, these are composed of three movies—each from the Lord of the Rings series and lists the US and New Zealand together as countries of origin. Brazil, Malta, Morocco, and Poland only consist of one movie in the Top 250 data set and share it with another country.

Conclusion:

Given the Top 250 rated IMDB movies, and data available on the IMDB database, we found no strong predictors of how high users will rate a movie. Based on correlation, rating is most related with the number of votes (correlation = 0.66), which, while not really a characteristic of the movie

Average IMDB Ratings per Genre, Year of Release, and Origin

Average IMDB rating per genre

Average IMDB rating per year of release

Average IMDB rating per country of origin

itself, can indicate how popular a movie is. Average ratings per categorical factor show that Sci-Fi, Music, Action, and Crime movies have tended to receive higher ratings, while Sport received the lowest on average. Average ratings per year of release and country of origin was not very informative, however it is important to note that the data set is only the top 250 movies.