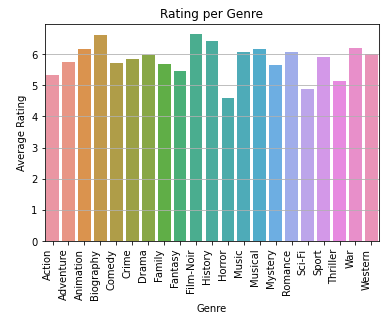
Movie Analytics

Daniel Chang, Luis Jovel

How do genre, director, or budget affect the average rating of movies? Studios could use the results to prioritize movies production. The public could also use it to gain insights into a movie.

We obtained IMDb’s data from Kaggle. This dataset was one of the few datasets that included budgets. We filtered the data to movies released in the USA, and ignored movies that didn’t have a country specified.

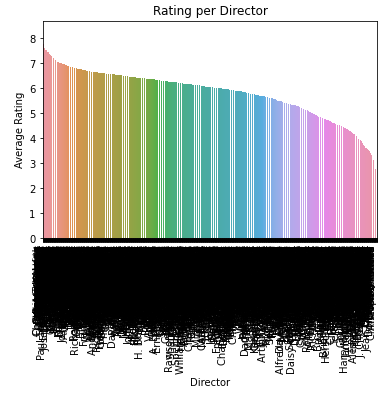
We first took a look at the ratings per genre. After we averaged the movie ratings into groups of genres, we found several genres with only one movie. After removing the genres with only one movie, we plotted the following bar plot.



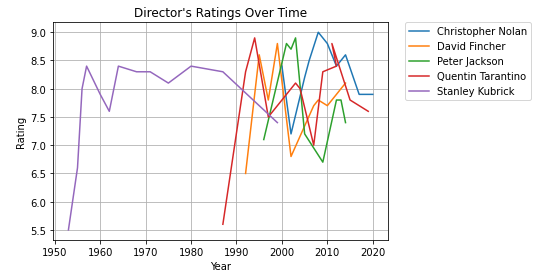
From the above plot, we can quickly see that most of the genres are fairly even, with exception to Horrors doing noticeably worse on average.

We could delve more deeply into how genre can affect ratings, but there are more extensive studies into how genre affect ratings and gross revenue, but didn’t include how directors or budget can affect the ratings. The other studies focused on business aspects such as gross revenue.  **<COME BACK TO THIS>**

Next up we looked at directors. After taking a quick look at all the directors we found a few outlier directors that had almost perfect ratings, but only had one or two movies. We filtered out directors that had less than 5 movies to their name and sorted them in descending order. Taking a look at the data frame, we narrowed the directors to a measly 1516 directors. Trying to plot all of them resulted into a pretty good-looking mess.

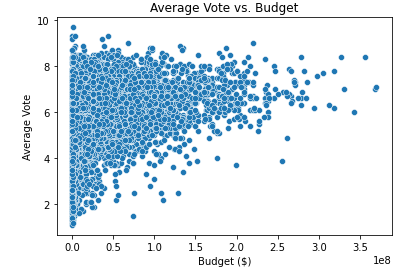


From the above mess, we took a look at the top 5 directors, and plotted their average movie ratings over their career.

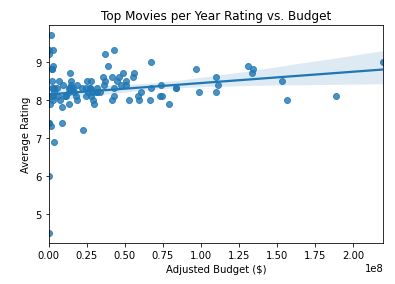


Christopher Nolan is the top-rated director according to our IMDb dataset, but had a shorter career compared to the rest of the top 5. The director’s ratings appear that their first couple of movies aren’t that great, but then they have fairly consistent results.

To see if there was a correlation between budget and movie ratings. We found a lot of movies were missing budget data so we removed them since we weren’t going to be able to find the budget. We then had to adjust for inflation using the movie release year. We used the library cpi to adjust the budget for inflation using Consumer Pricing Index data. However, the CPI data only consisted of data from 1913 to 2019, so we had to drop movies released prior to 1913, and assumed 2019 and 2020 budget to be close enough. Plotting the all the movies vs their budget resulted in the below graph.



Quickly looking at the plot above, we see several movies being made for somewhat unreasonably small budgets. Like a 90-minute film with a budget of $3. We see that there isn’t a general correlation between the budget and the average rating. We thought to repeat the plot using the best movie per year from 1913 to 2020.



The regression line looks fairly flat, so we quickly computed the correlation between budget and average rating for a correlation of 0.283.

In summary, it appears the only really consistent way to get a highly-rated movie on IMDb is to have a top-rated director for your movie, and avoid Horror genre.