

Average Movie Ratings do not Appear to be Genre-Dependent

Diane Nicoll
DSE 200X

Dataset

For this project, I analyzed the:

IMDB Movie Dataset

Motivation

I was motivated to work with the IMDB movie dataset to solve the problem of the unknown relationship between movie genre and movie ratings. Movie ratings can have a direct impact on ticket sales and accolades, which are important for producers, directors, and the cast of movies. A better understanding of the relationship between movie genre and movie ratings would provide valuable insight for individuals in the movie profession. For example, if comedies generally tend to receive the highest ratings, a producer could use this knowledge to make sure their upcoming movie includes the comedy genre to increase his or her movies' rating, and hopefully increase box office sales and award nominations.

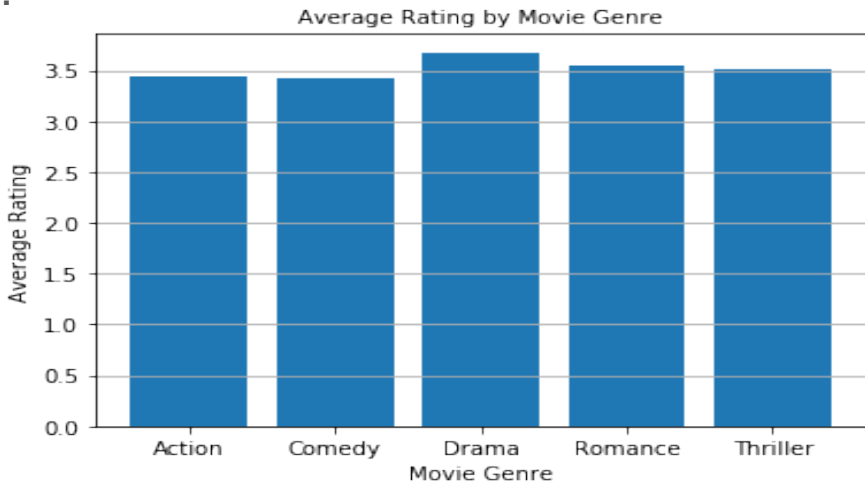
Research Question(s)

Do movies in the comedy genre receive higher ratings as compared to movies in other genres? Does this change over time?

Note: Due to the size of the dataset, only the genres with the greatest number of movies were analyzed: comedy, drama, thriller, romance, and action.

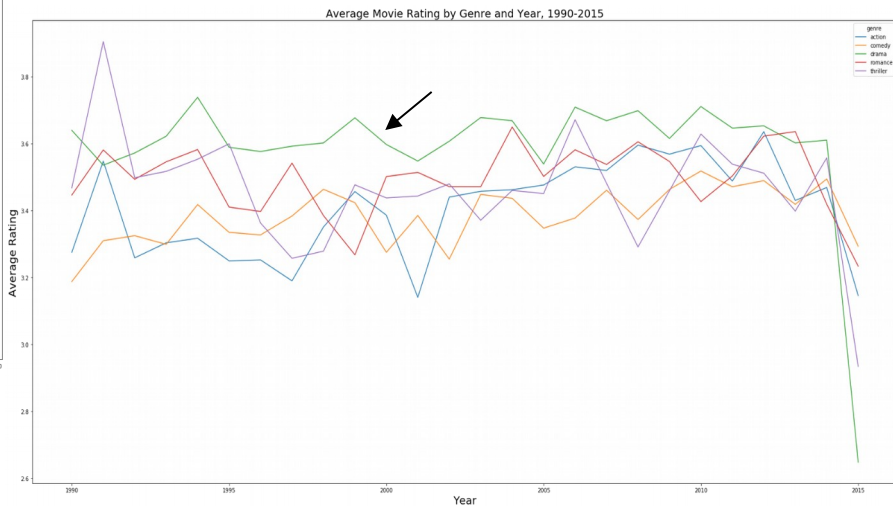
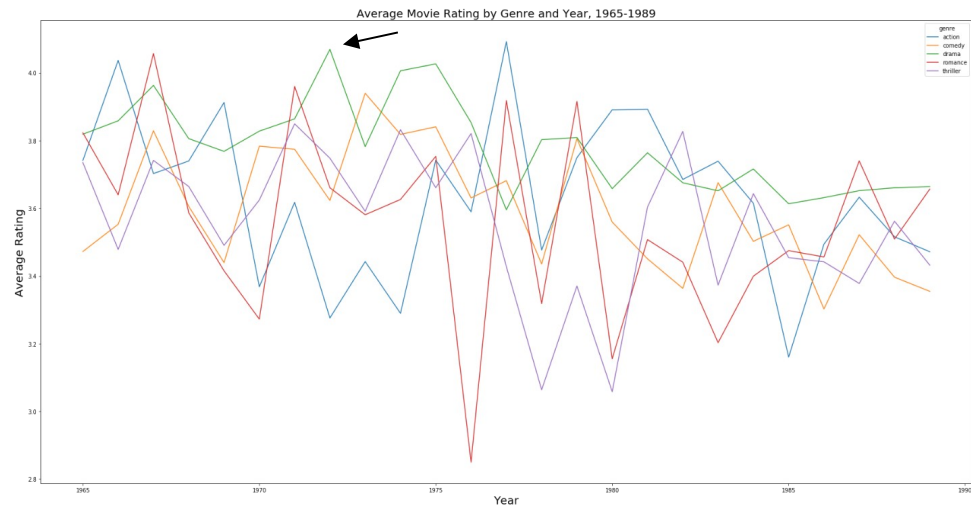
Findings

Average movie ratings do not appear to be genre-dependent. No major differences were found in average movie rating between movie genres in this dataset, when analyzed over the entire dataset (1891-2015; below) or filtered into four year-divided datasets.



Findings

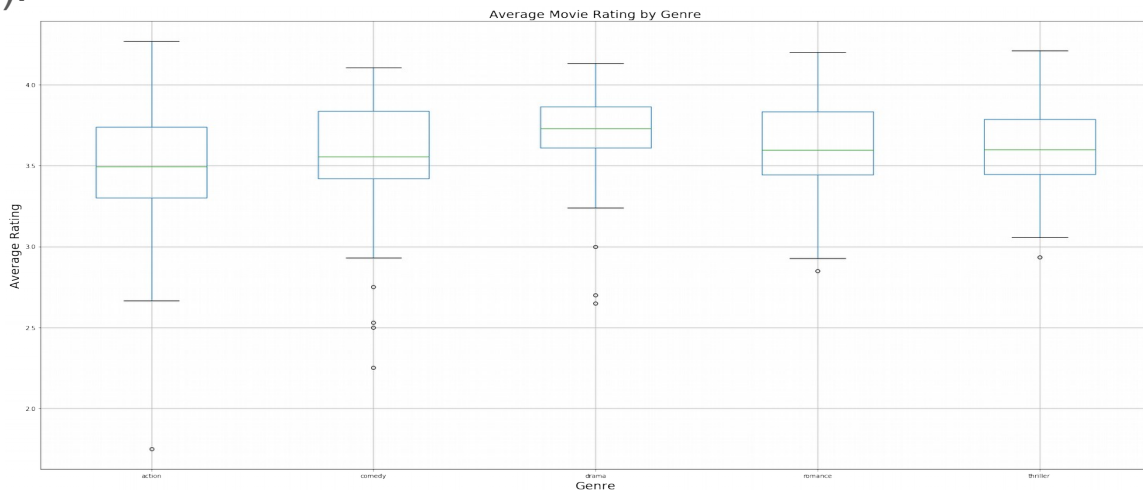
The drama genre received the highest average ratings over the entire dataset (1891-2015) as well as in two of the four year-divided datasets (note that the average ratings were minimally higher than the other genres' average ratings):



Findings

Outlier variables were present in the ratings of each of the genres analyzed.

The drama genre exhibited the least amount of variability in average ratings, while the action genre exhibited the greatest amount of variability, across the entire dataset (1891-2015):



Next Steps

The 5 genres from this project could be analyzed in smaller year-divided datasets, such as decades, in order to further assess potential relationships between movie genre and average rating.

An evaluation of each genre individually across time may lead to additional insights and understanding.

Analysis of the outliers present in this dataset could allow for a more clear understanding of the relationships between movie genre and average rating. Removal of all outliers and re-analysis of the data may lead to different findings.

Acknowledgements

I did not have anyone review or aid on this project.

References

<https://grouplens.org/datasets/movielens/>

<https://stackoverflow.com> (used to look for assistance when I received errors in code)