Exploring Trends of Movie Audience Ratings Using Recorded Timestamps

Aliasghar Mehdizadeh Dastjerdi

Dataset:

- IMDB Movie Dataset was used for data analysis.
- This dataset describes 5-star rating and free-text tagging activity from MovieLens.
- It contains 25000095 ratings and 1093360 tag applications across 62423 movies.
- These data were created by 162541 users between January 09, 1995 and November 21, 2019.
- For data analysis, the two files of “movies.csv” and “ratings.csv” were only used.

Motivation

Audience reviews and ratings can play a significant role in consumer decision-making processes and accordingly, influence key movie performance measures such as movie revenues. Previous studies showed that people's preferences and rating behavior depend on wide ranging factors such as audience characteristics (e.g., age), movie attributes (e.g. genre), cast, director etc. However, there is a lack of knowledge investigating how audience movie ratings change over time. Therefore, this study aims to explore temporal trends in movie rating behavior through data analysis and visualization. More specifically, this study will firstly investigate non-random patterns such as upward and downward trends and then provide insights regarding the interpretation and comparison. This is valuable since it will pave the way for predict consumer behavior and its better explanation.

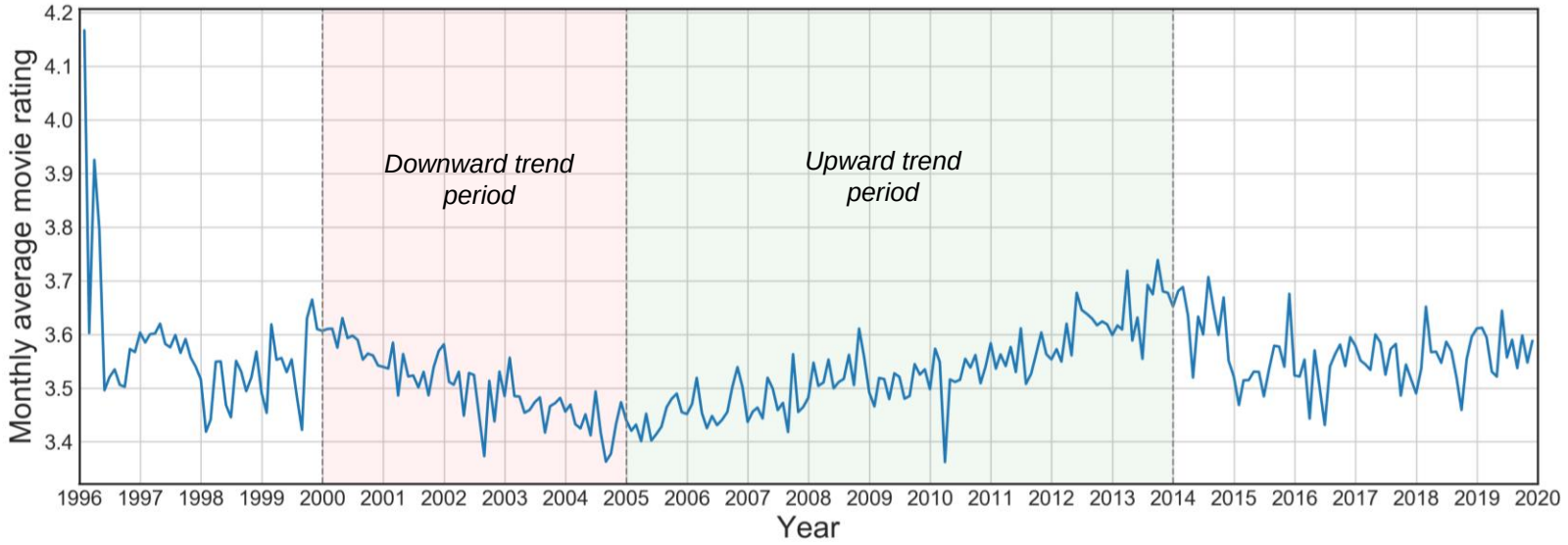
Research Question(s)

Can any non-random pattern be recognized in recorded movie ratings over time? I hypothesized that there are some long period of times showing upward/downward trend. Therefore, the focus was on identifying such trends and finding insights into why they occurred using only IMDB dataset.

In this regard, I extracted the date of recorded ratings using timestamps. Monthly average movie ratings were calculated (from 1996 till 2019) and plotted against date. To explain observed trends, I investigated the difference between distributions of movie year of release as well as favorite movie genre over two periods.

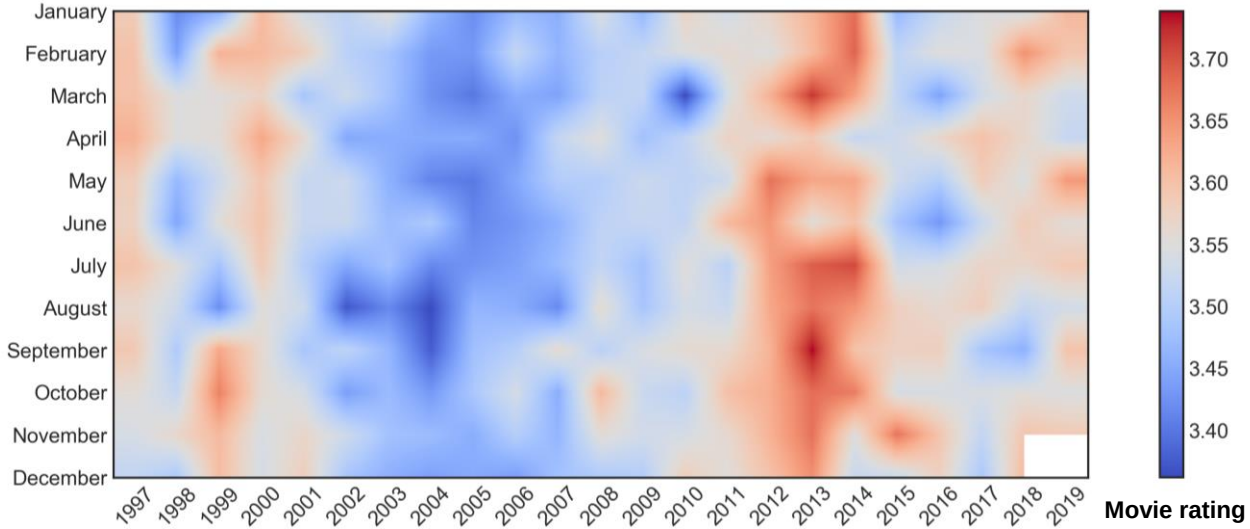
Findings:

- The line plot shows how audience movie ratings change over time.
- Two trends are identified and presented below.
- Although they do not consistently increase or decrease, they are considered as non-random patterns (Brownlee, 2017).
- This plot addresses the research question by showing both increasing and decreasing trends.



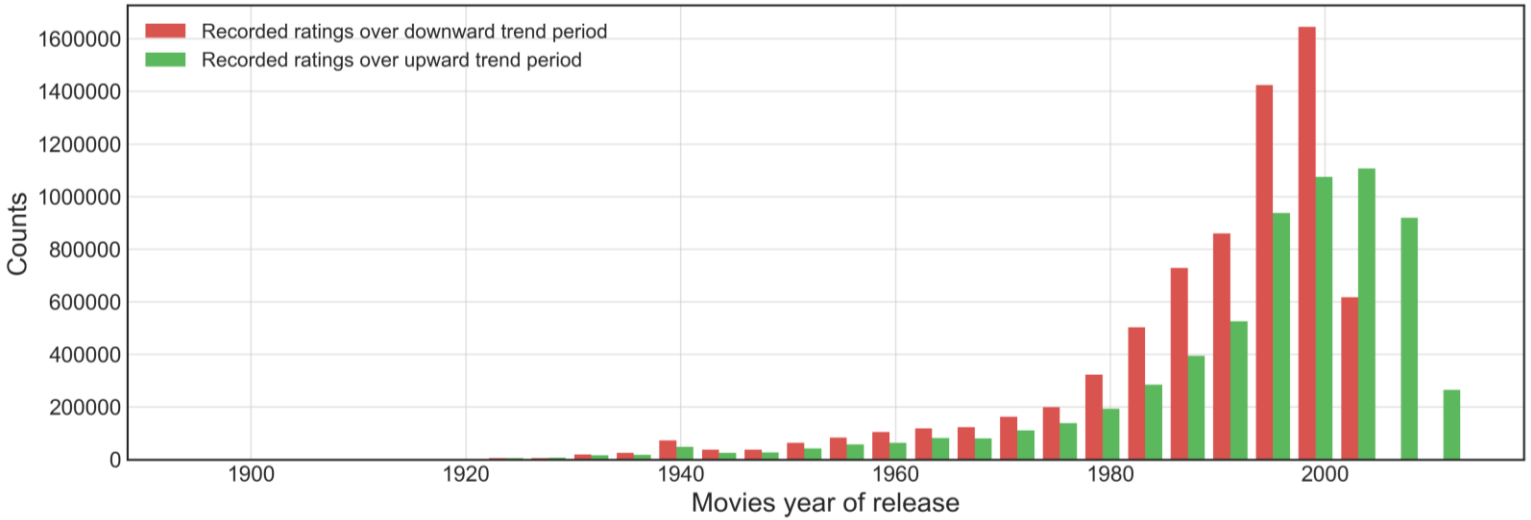
Findings:

- This heat map are created by arranging the observations into a matrix of year-columns and month-rows, with average movie ratings in the cell for each month (Note: white patches are missing data)
- The heat map does not show a clear monthly pattern. However, its yearly pattern agrees with the last plot i.e. higher ratings over upward trend period (more reddish) and lower ratings over downward trend period (more blueish)



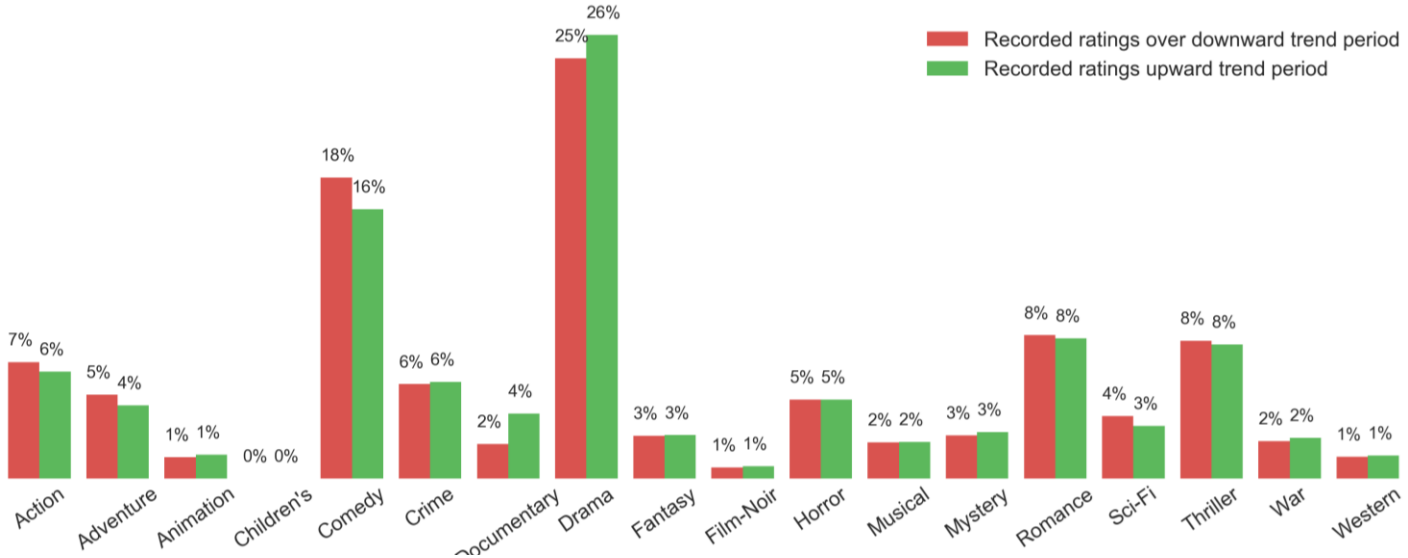
Findings:

- The bar plot shows how movies with different release year are distributed over both upward and downward trend periods
- The proportions (counts) are different in two periods in particular movies came out after 2000.
- Increasing/decreasing trend in movies ratings could be explained by this quality; however, it needs to be further explored using other datasets.



Findings:

- The plot shows audience preferred movie genres in two periods, where the movies rated below 4 were excluded for analysis and visualization.
- There is no significant difference in any category in both time periods. The shares are mostly comparable, thus increasing/decreasing trend in movies ratings can not be related to changes in favorite movie genres.



Acknowledgements

I would like to acknowledge that no one gave me any feedback to improve this assignment. However, the main idea (i.e. time series approach) was inspired by Brownlee, J. (2017).

References

Brownlee, J. (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.

The following websites were also used to determine syntax and debug my code.

- pandas.pydata.org
- doc.python.org
- stackoverflow.com