Movies Rating Prediction Analysis

Zimu Su/ Metis Business project

Nonnegligible impact of movie user score

- woven into the ticket-buying behavior.
- affect box-office performance.
- influence the decision to make a movie.
- important in recommendation system

....

Overview of Data-driven Rating model

Movie Production Features

Genres, Release date, Title, Budget, Cast, Director...

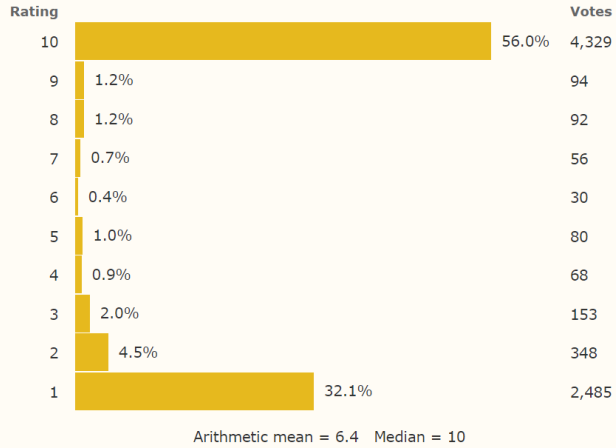
predict

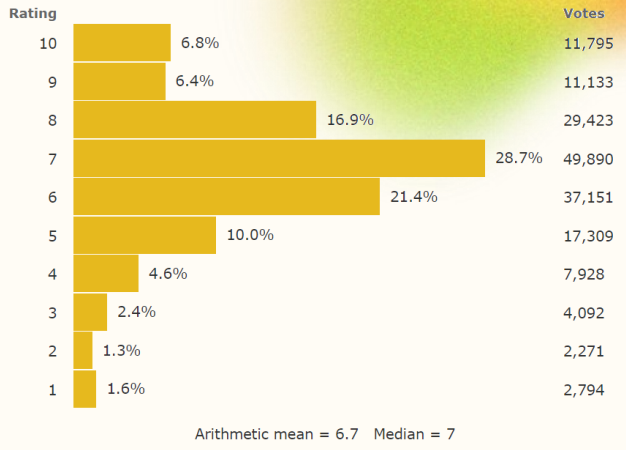
Movie score

Expectation of popularity, box office performance ...

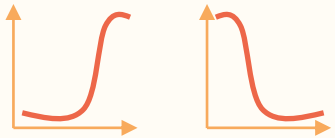
- **Underlying Problem:** Score metrics based on single measure (like average) removes distribution information.

- Different rating distributions can have the same average score.



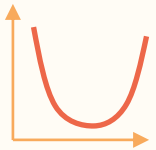


- Distribution-based score analysis provide elaborate information about audience's feedback.



Concentrated good/bad

Evenly good/bad rating



Polarized

Solution path

Identify rating distribution

Identify relationship between movies features and distribution pattern.

Use probability distribution function to characterize the shape.

Regression analysis between features and distribution shape.

>40,000 movies info collected from Grouplens website

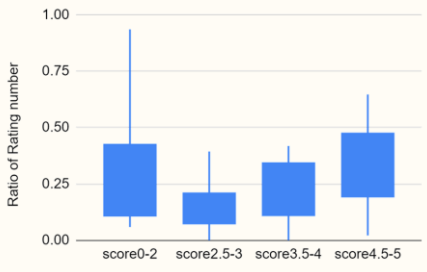
Filtered with total ratings > 40

~14,000movies left

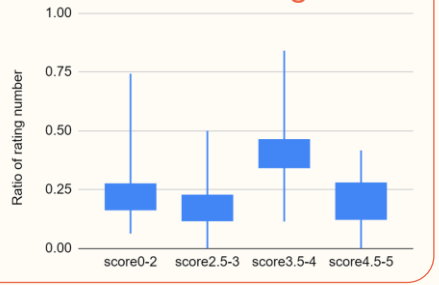
Distribution analysis

Controversial

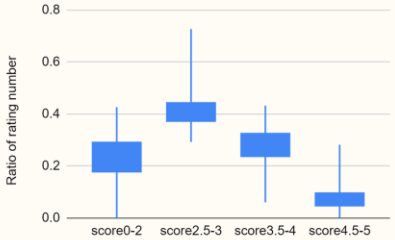
Polarized



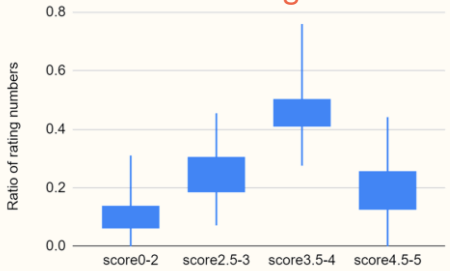
Wave rating

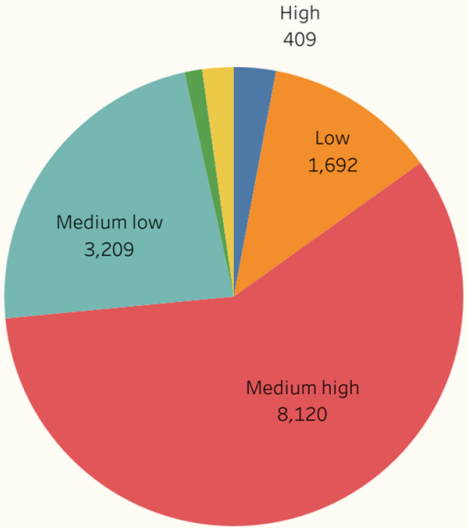


Medium low

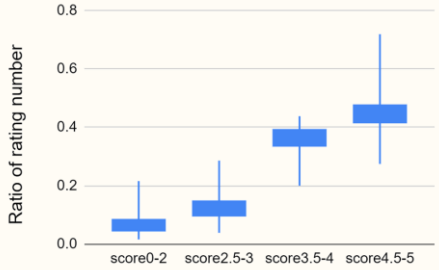


Medium high

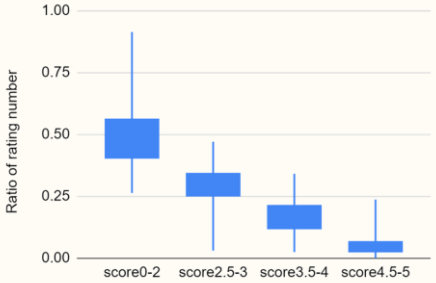




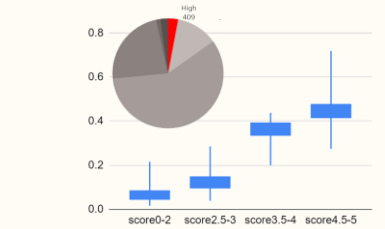
High

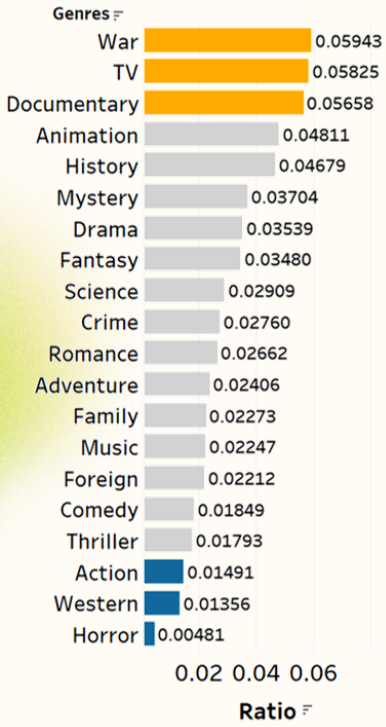


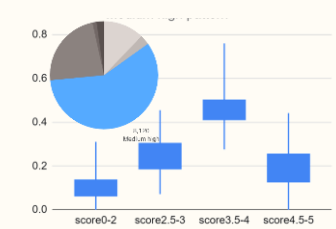
Low

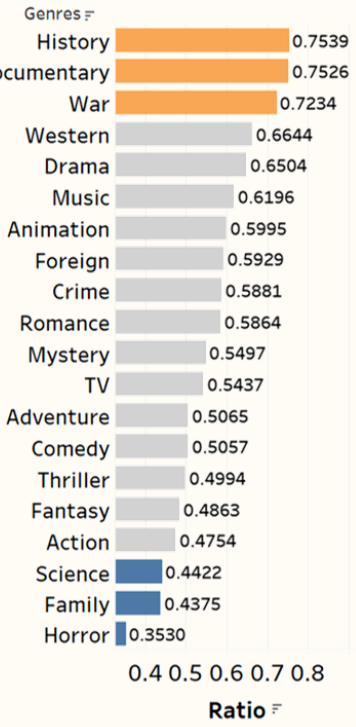


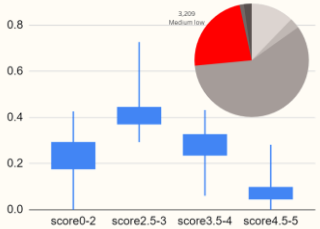
Genre analysis

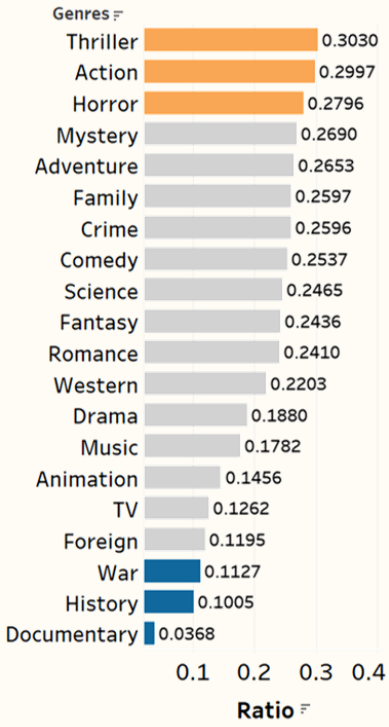


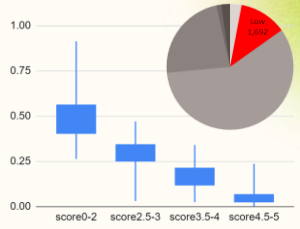


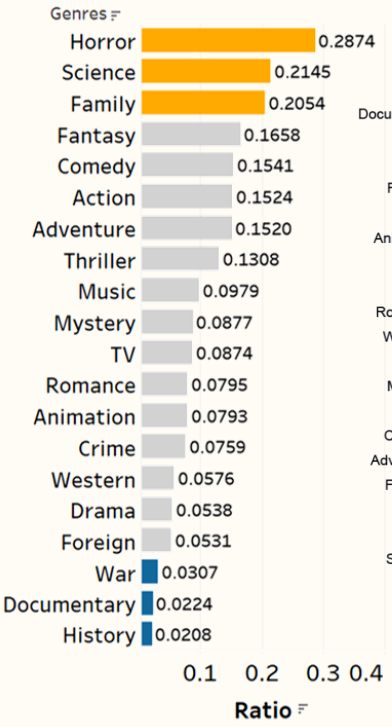


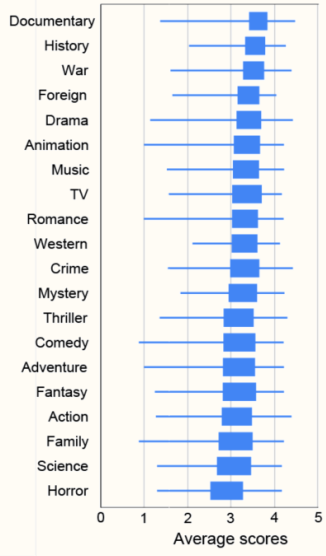






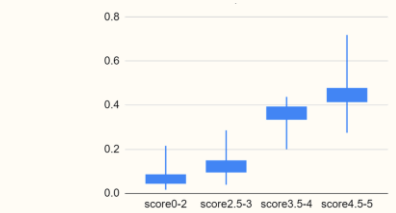




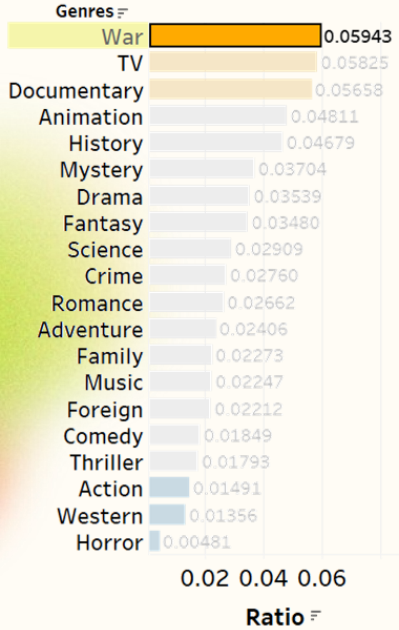


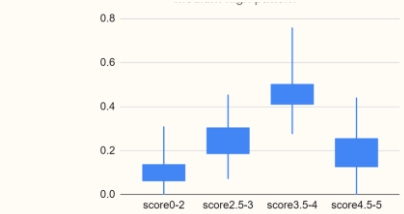
Consistent good rating

Other examples: Documentary, animation, history

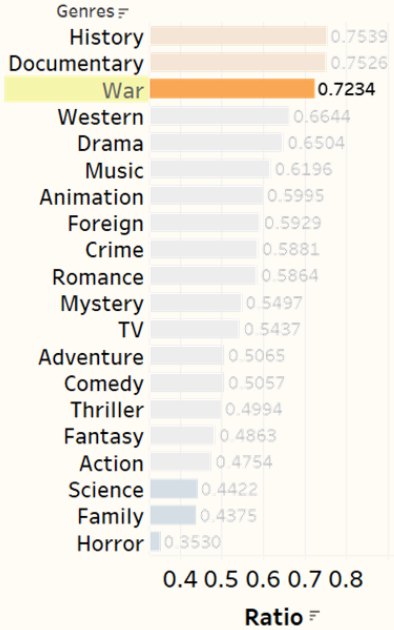


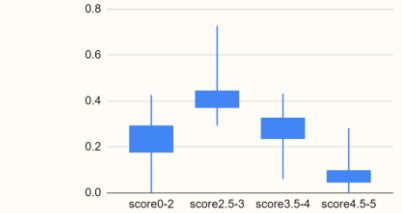
High rating



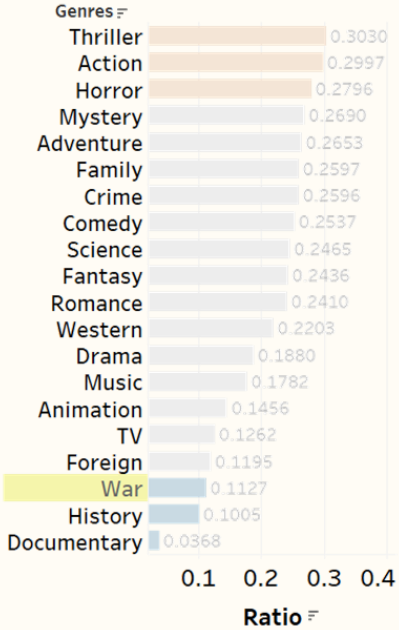


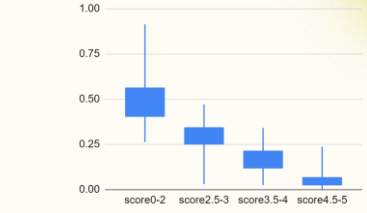
Medium high rating



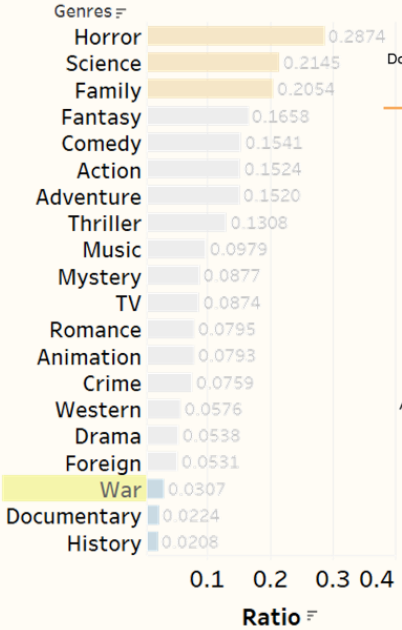


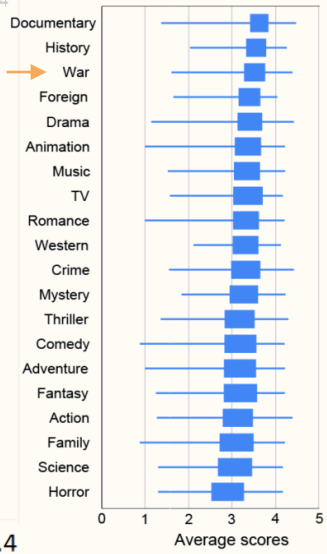
Medium low rating





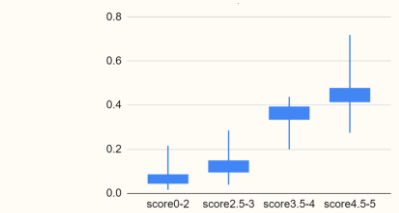
Low rating

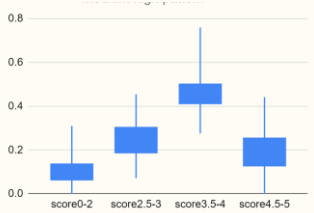


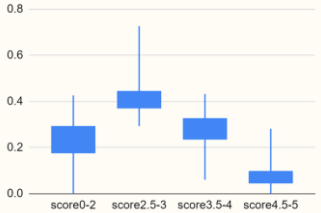


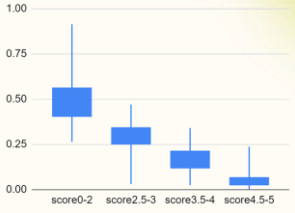
Consistent bad rating

Other examples: Horror, Thrill

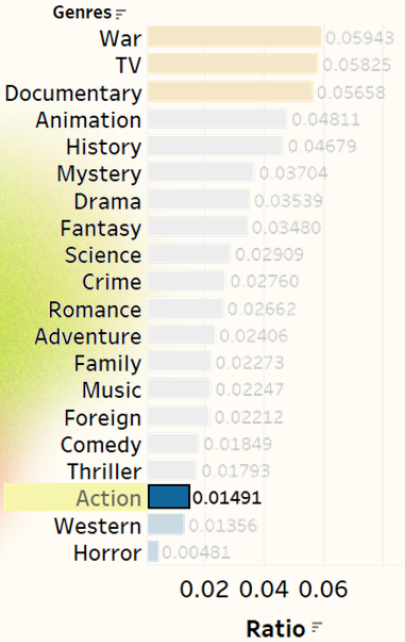




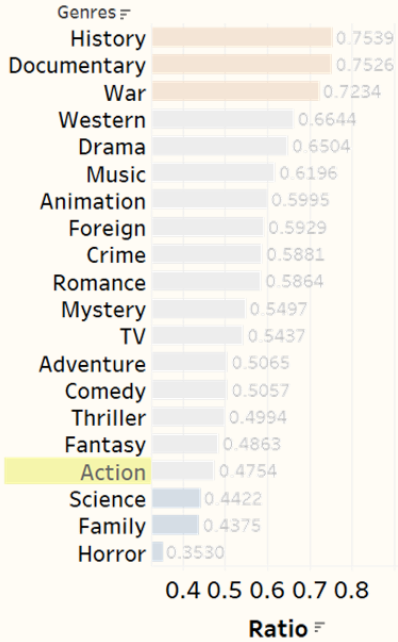




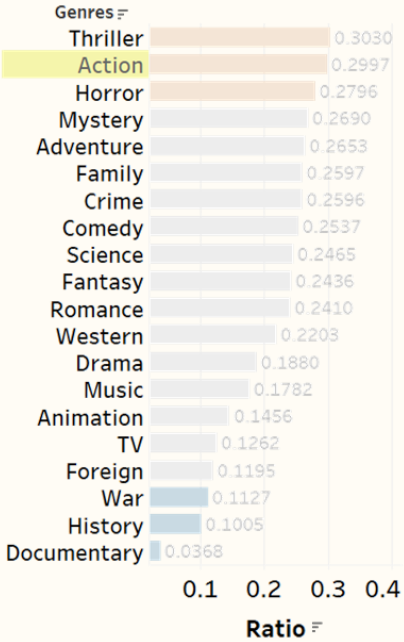
High rating



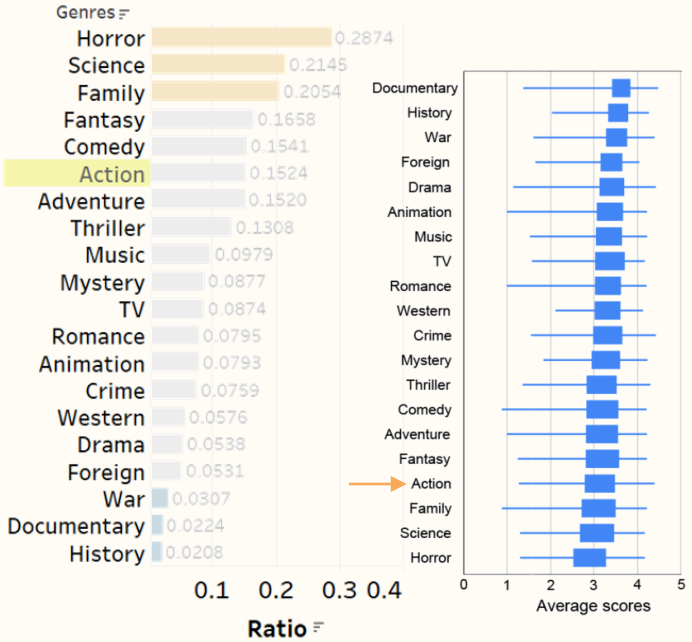
Medium high rating



Medium low rating

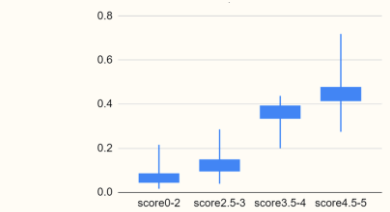


Low rating

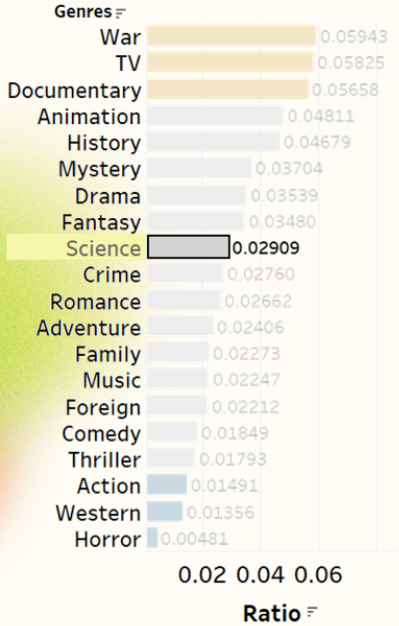


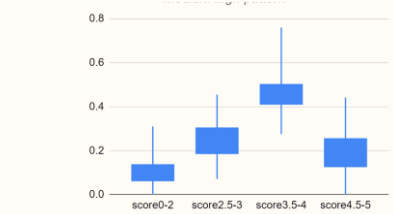
Some are excellent but most of them are bad

Other examples: Fantasy, Mystery, Adventure

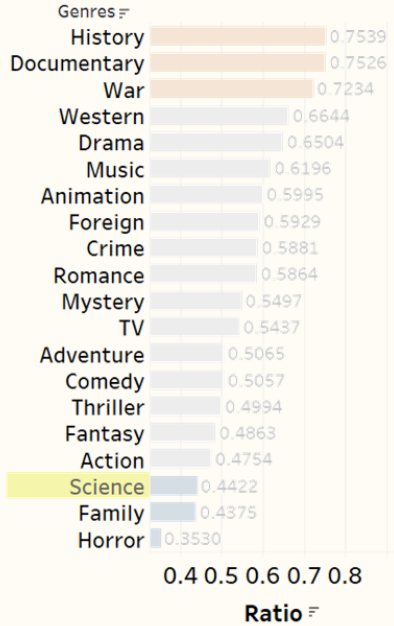


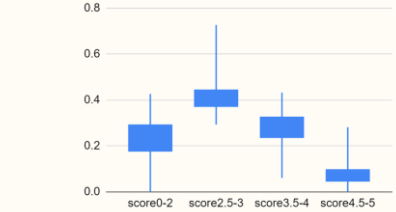
High rating



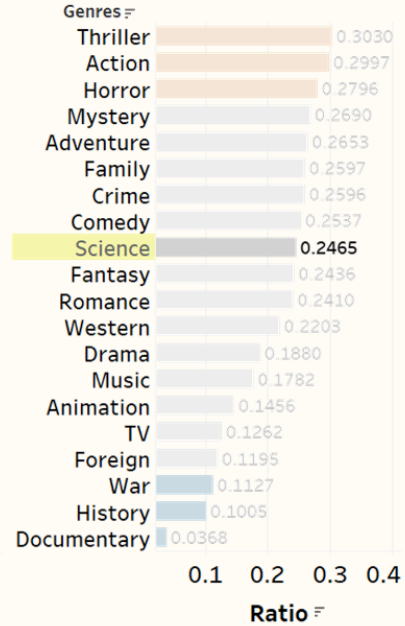


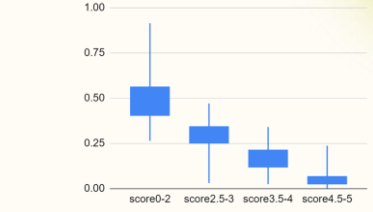
Medium high rating



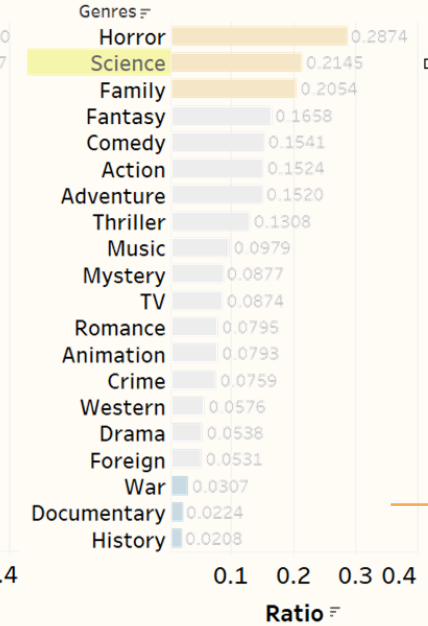


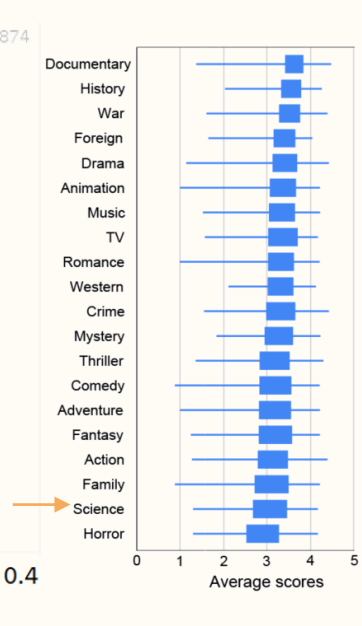
Medium low rating



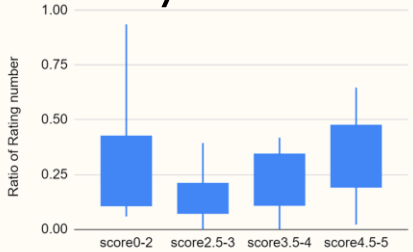


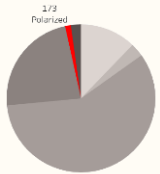
Low rating

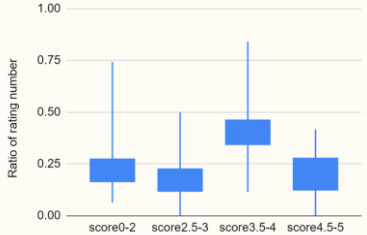


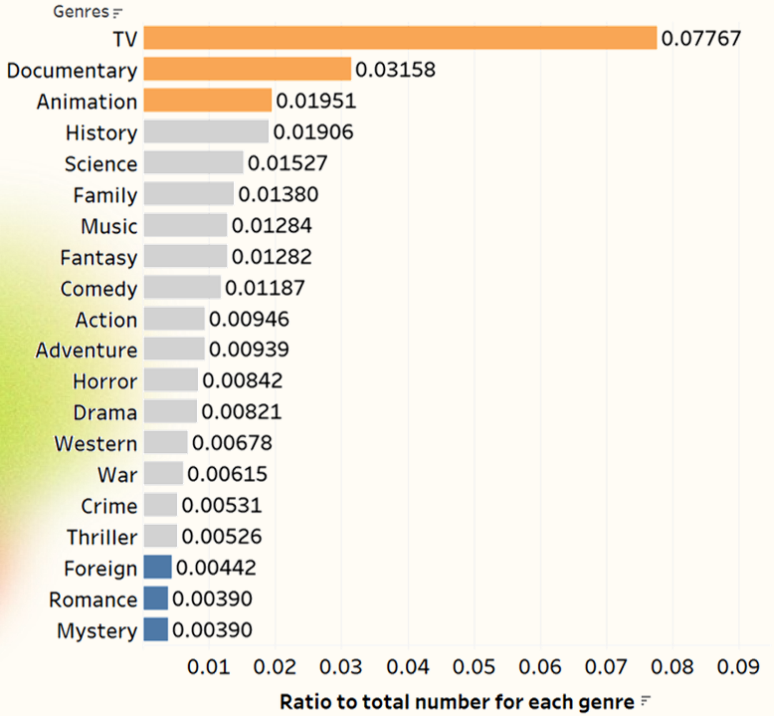


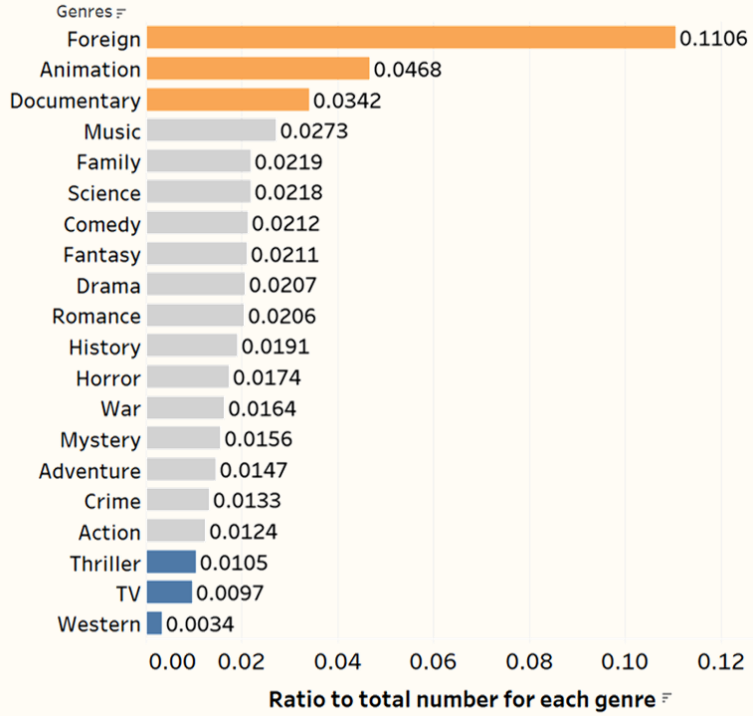
Genre analysis of controversial movies

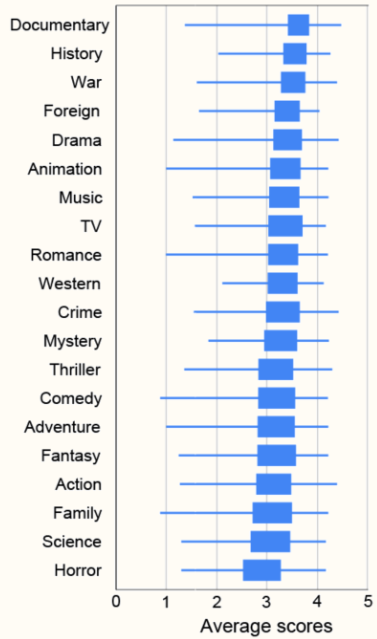












Next step:

- Incorporate more features such as title, budget, cast, director, etc.
- Parameterize the distribution shape.
- Build regression model between features and distribution parameters.

Thank you