



Creating **movielens** Data-Driven Movie Ratings

SAMMI CHANG, SERGEY DIRIN, CAESAR PHAN,
KEVIN CHENG, HUY TRUONG, DAVID PRIMROSE

The Dataset

movielens

Results

movie genre rating predictions
successful with 25% accuracy based
on rapidly designed model

The Model

```
summary(fitted.model<- clm(rating ~ userId  
  Action +  
  Adventure +  
  Animation +  
  Childrens +  
  Comedy +  
  Crime +  
  Documentary +  
  Drama +  
  Fantasy +  
  Film_Noir +  
  Horror +  
  Musical +  
  Mystery +  
  Romance +  
  Sci-Fi +  
  Thriller +  
  War +  
  western, data=training, link="logit"))
```

<https://github.com/Lxr82/movie-rating-prediction>

Survey data was collected from viewers of
9742 movies



We logged 100836 ratings and 3683 tags
across 610 users

Derived Logistic
Model Prediction

Demonstrated
comparison genre
affinity

Investigated
clustering
Basic exploratory
data analysis on
combined movie,
rating, user, genre
dataset

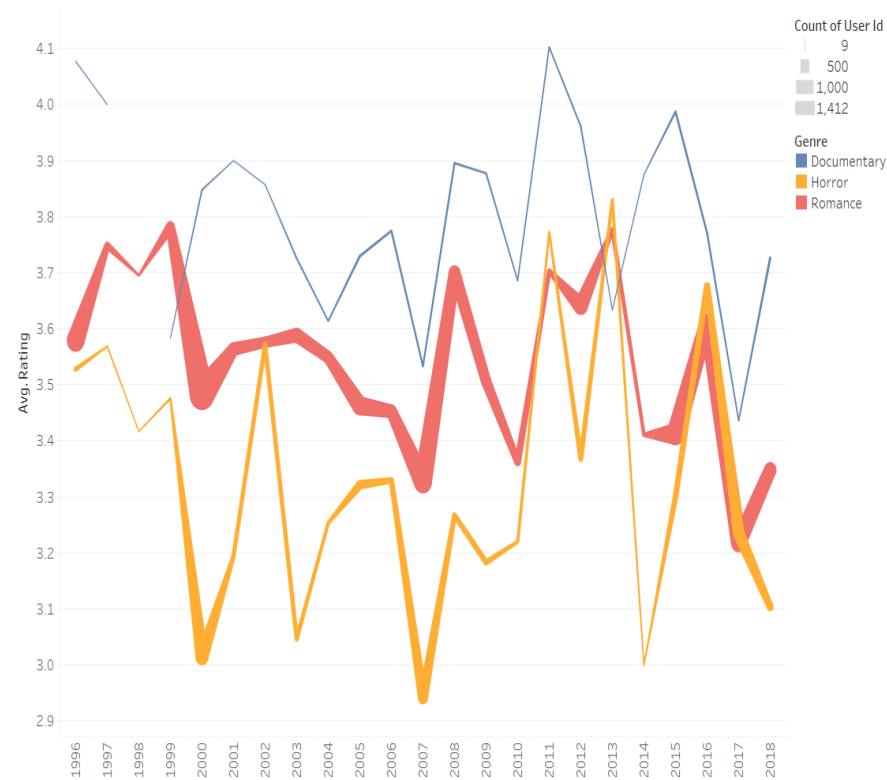
Exploratory Data Analysis

Aggregated Popularity by Genre



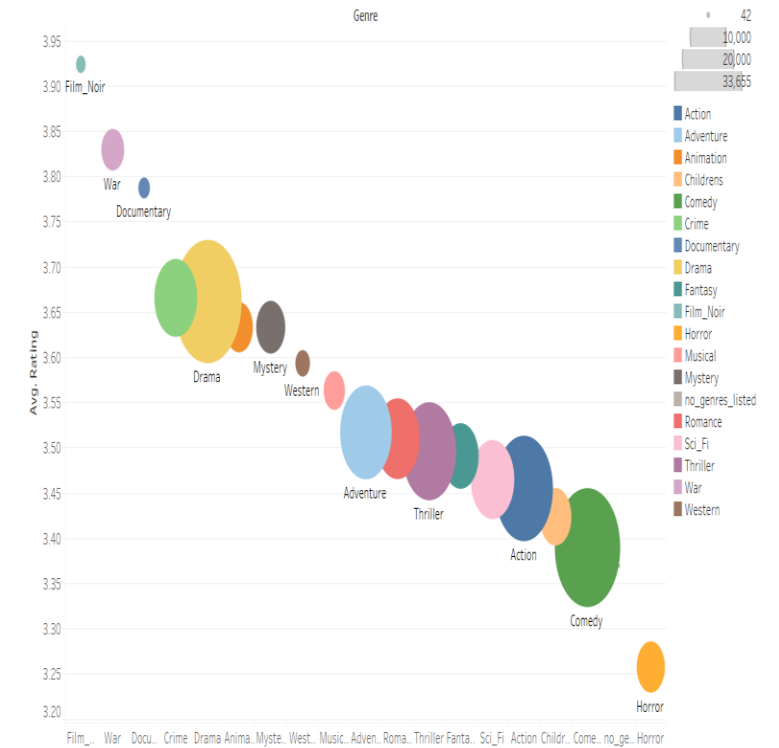
Genre. Color shows average of Rating. Size shows sum of Value. The marks are labeled by Genre.

Genre Sentiment Over Time



The trend of average of Rating for review_years Year. Color shows details about Genre. Size shows count of User Id. Details are shown for Genre. The data is filtered on Rating, which keeps all values. The view is filtered on Genre, which keeps Documentary, Horror and Romance.

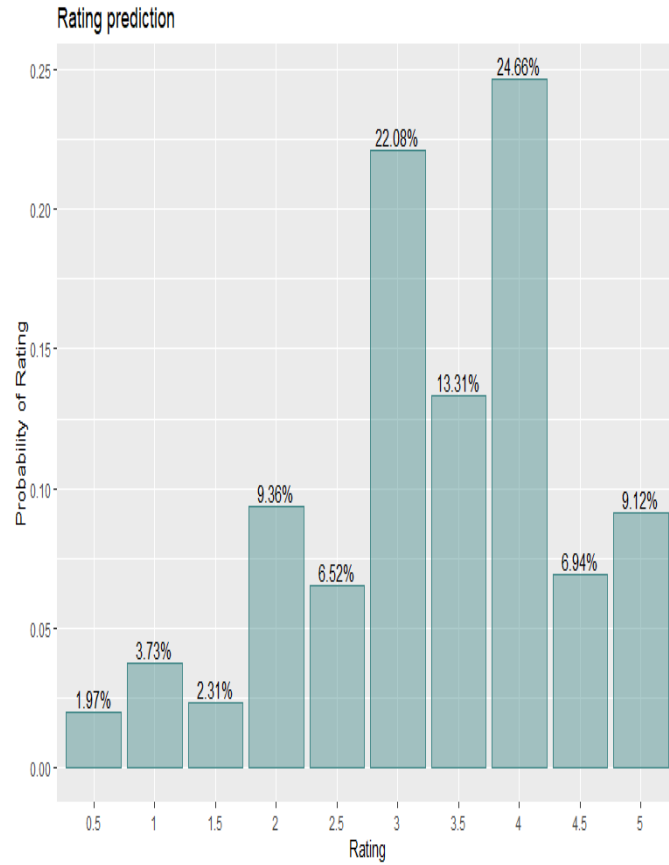
Ratings by Genre



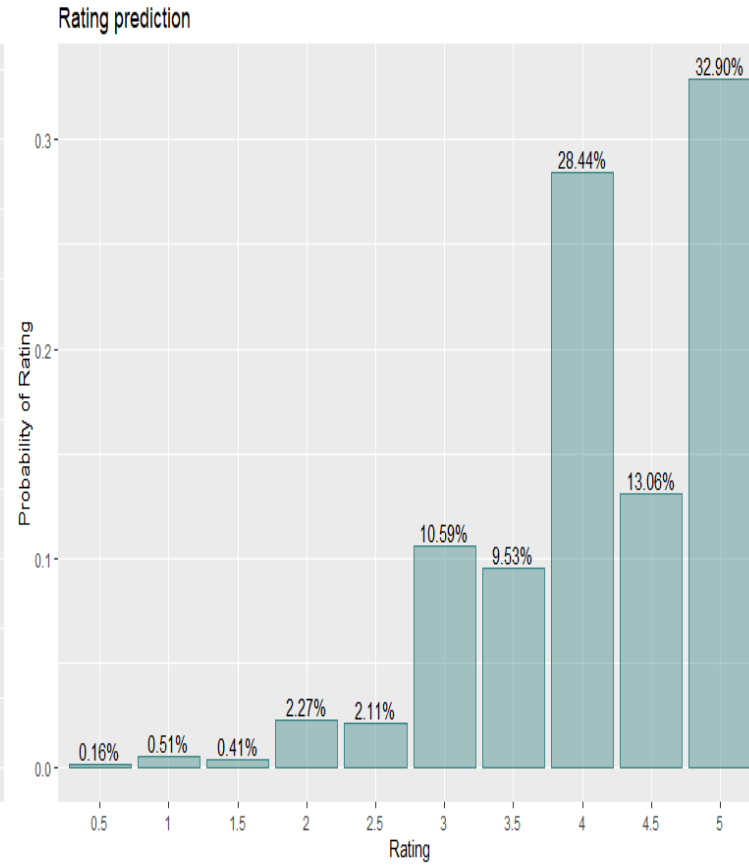
Average of Rating for each Genre. Color shows details about Genre. Size shows count of Title. The marks are labeled by Genre. The data is filtered on Rating and count of Movie Id. The Rating filter ranges from 0.5 to 5. The count of Movie Id filter ranges from 42 to 33,655. The view is filtered on Genre, which keeps 19 of 19 members.

Logistic Regression of Movie Reviews

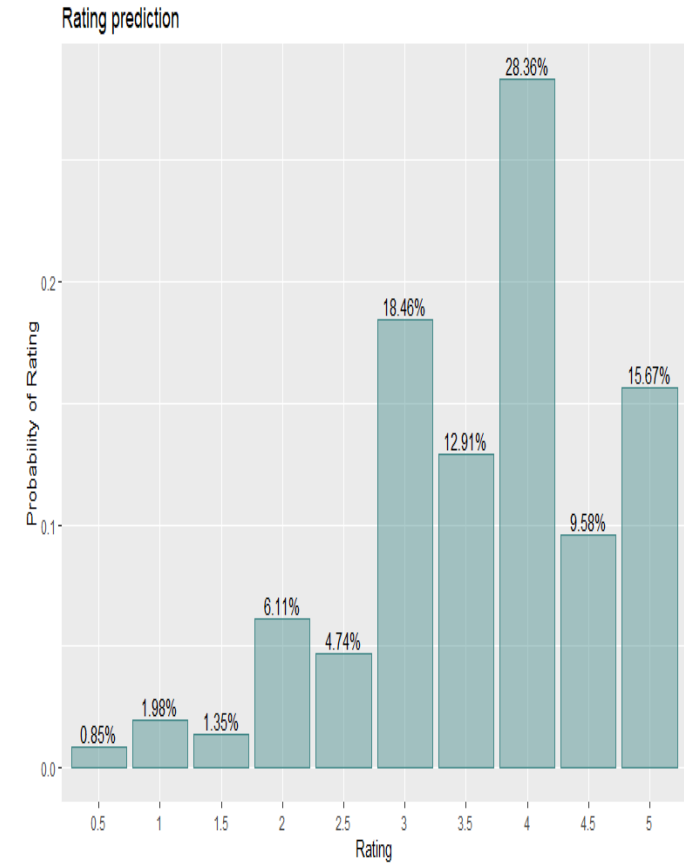
Comedy-Romance



Comedy-Crime-Noir-War

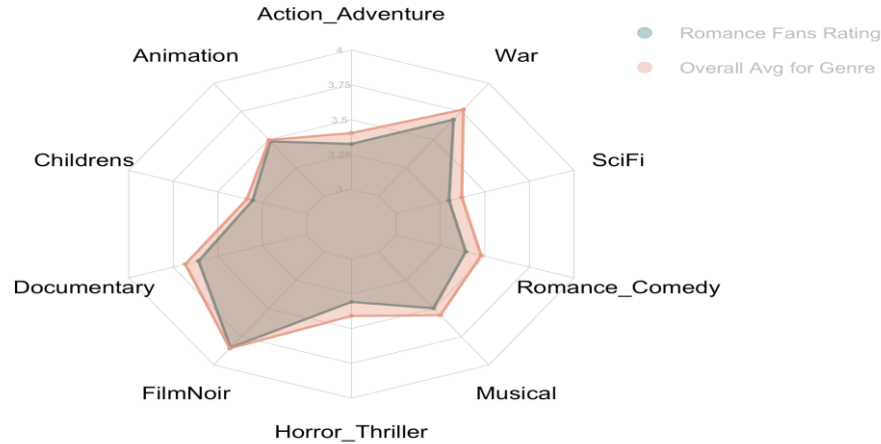


Romance-Drama



Fan Rating Tendencies across Genres

Romance Fan Tendencies



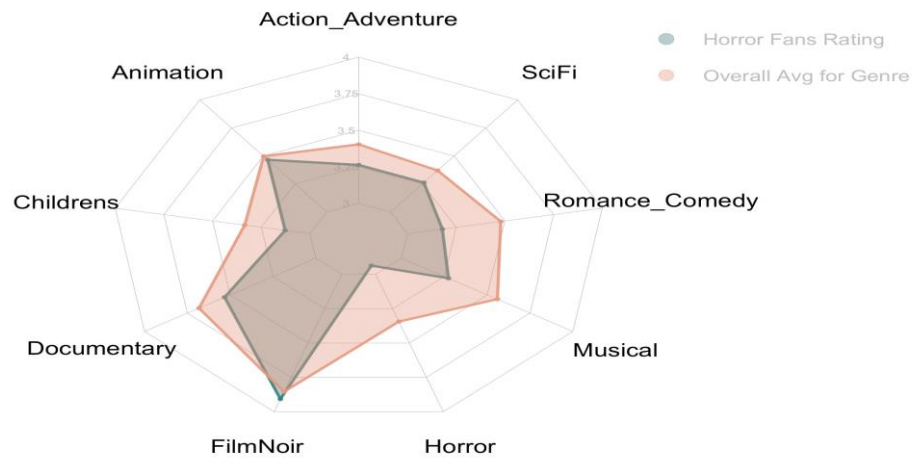
It can be said in all three cases even “expert raters” are biased in their ratings if they are a fan of a genre.

Romance Fans tend to rate higher Animation and Film Noir(classic crime movies) and rate lower Action&Horror

Horror Fans tend to like Film Noir and are very critical of Horror movies as well as Romance

Documentary Fans tend to be fair in their ratings of documentaries and are critical of Children’s and Musicals

Horror Fan Tendencies



Documentary Fan Tendencies

