Predicting Reviews by Roger Ebert

Project Luther
Zach Heick

Problem and Model

- Can we analyze and predict how Roger Ebert would review movies today?
- This can be done with Linear Regression!
- My model:
  - Mean Squared Error (Squared) 0.71
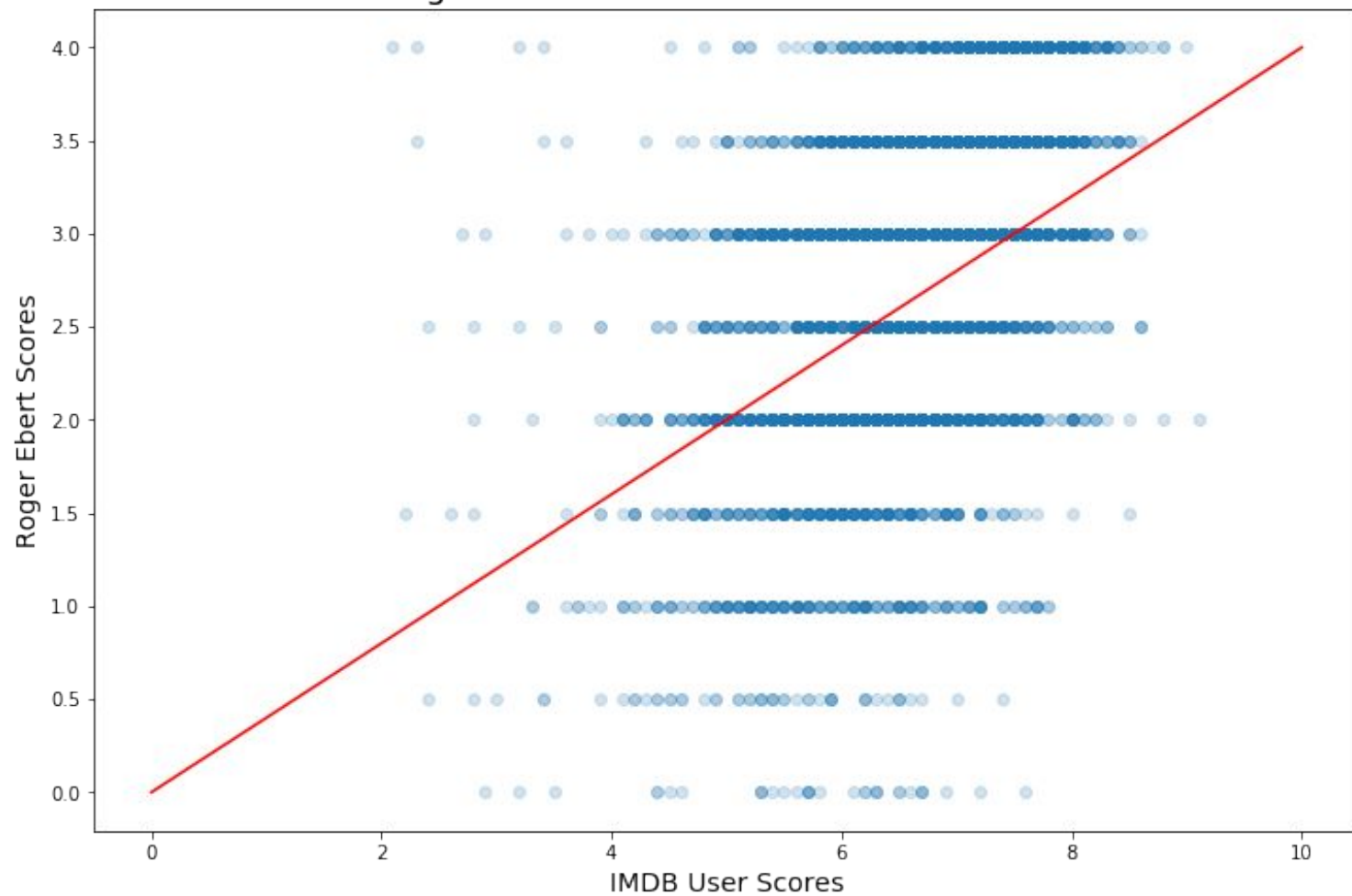 - Adjusted R Squared 0.939
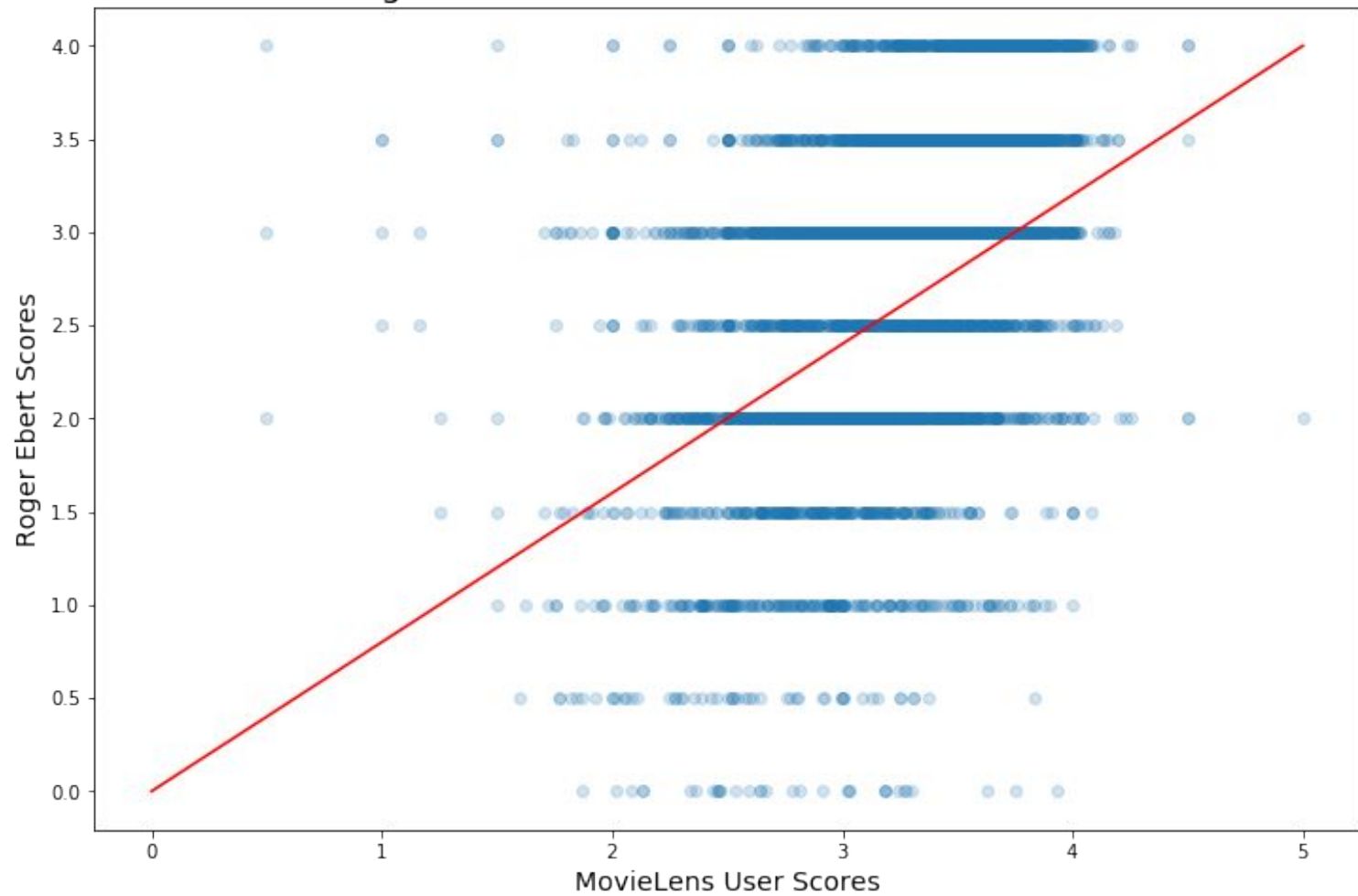 - 19 Features

The Data

- Scraped from Roger Ebert's website and IMDb.com using Python's BeautifulSoup and Selenium
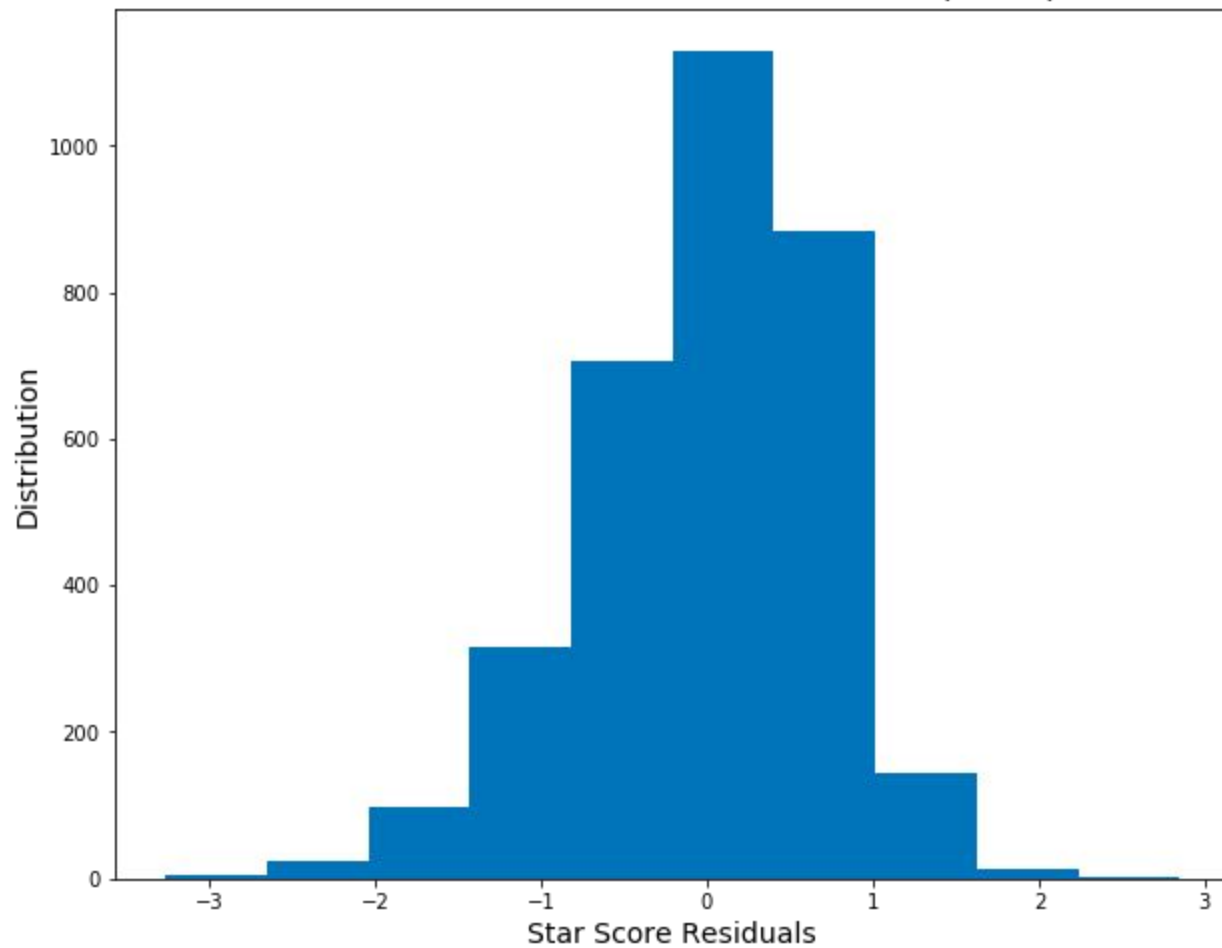- MovieLens data downloaded as a flat file
- 4749 total movie reviews

Roger Ebert Scores vs IMDB User Scores

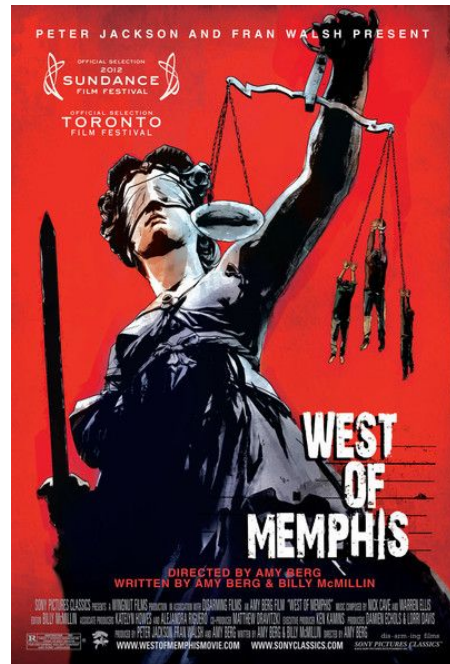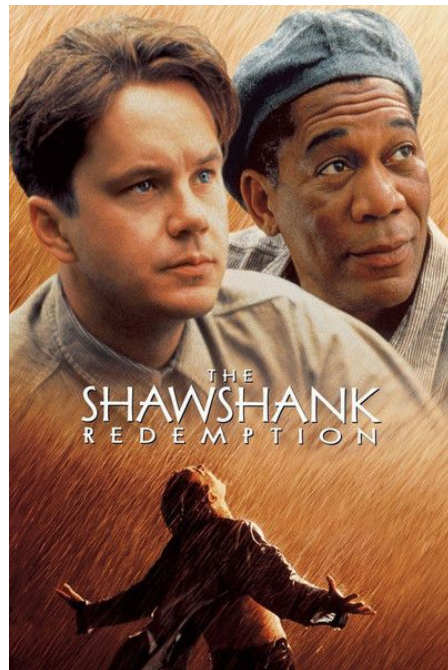Roger Ebert Scores vs MovieLens User Scores

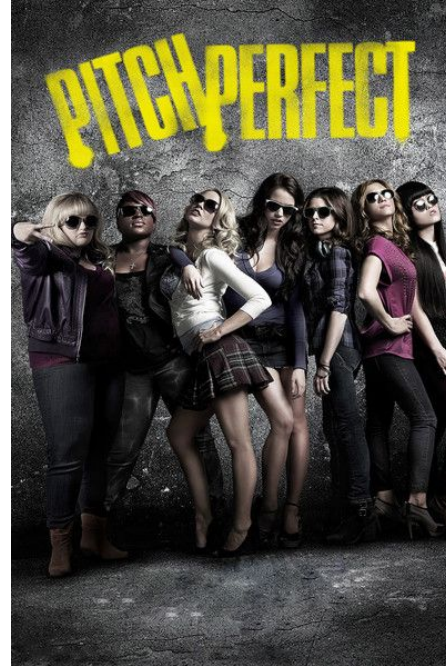Distribution of Star Score Residuals (Train)

Ebert's Likes

- Dramas
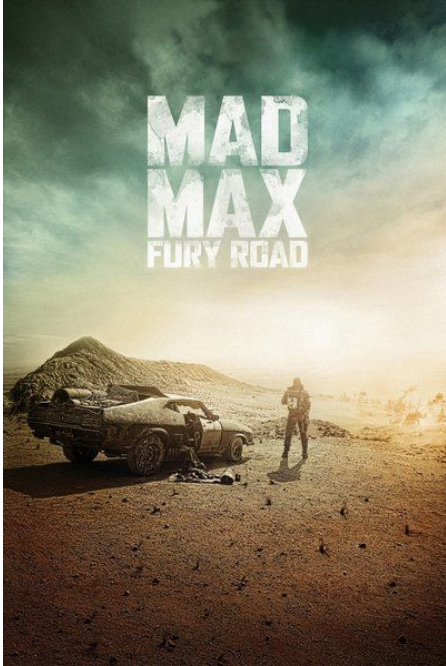
- Documentaries

- Foreign Films

Ebert's Dislikes

- Comedies

- PG-13 and PG

- Horror Films

Predicted Ratings

- Dunkirk (2017) : 3.61 / 4.00

- Emoji Movie (2017) : 1.35 / 4.00

- La La Land (2016) : 3.14 / 4.00

- Mad Max Fury Road (2015) : 3.81 / 4.00

Thank you!