

Date-a Night at the Movies

A model by Karl Rudeen

GOAL

- To predict the total domestic gross for a movie once opening weekend gross is known

The Data

- 1900 movies scraped from IMDb.com, boxofficemojo.com, and the-numbers.com from 1982-2016
- Inflation data scraped from usinflationcalculator.com
- Scraped release dates, stars, directors, production budget, total gross, opening weekend data and more
- Budget and total gross are adjusted for inflation
- The plan is to use this data to predict the overall gross for a movie

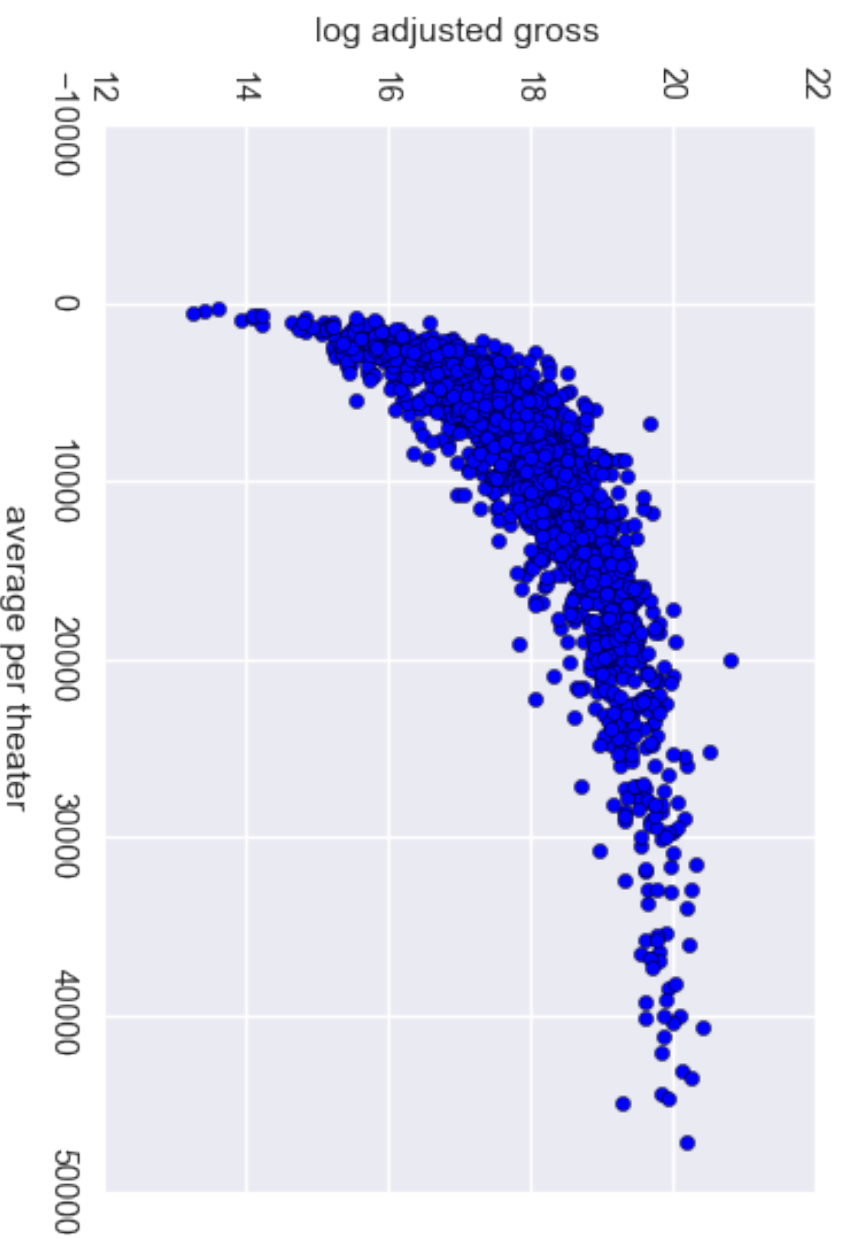
Model

- The final model incorporates metascore, average opening weekend gross per theater, and production budget
- It also incorporates director and top billed actors as dummy variables
- Uses Lasso Regression
- The R-squared value for this model is roughly .83 (on test data)

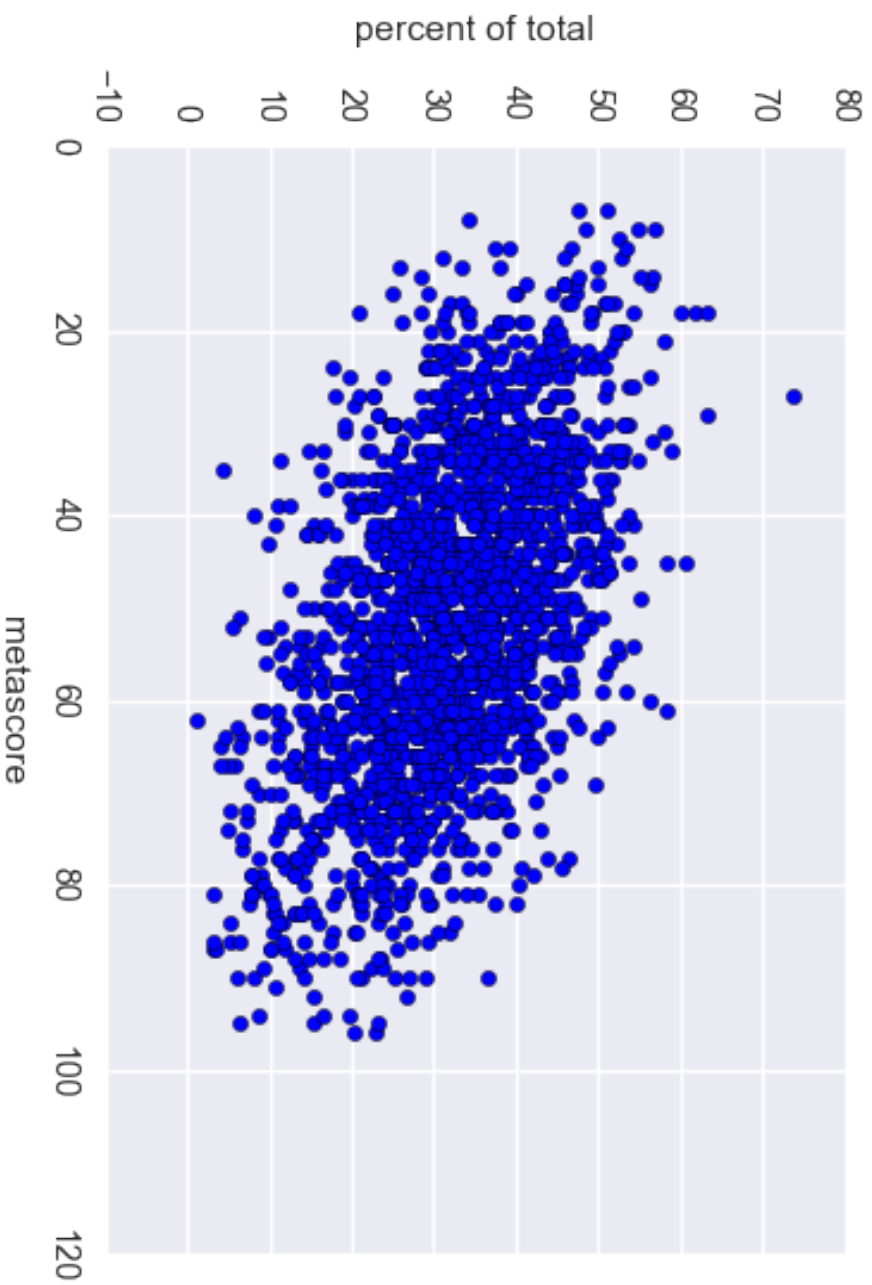
Top Five Features

- The top five features for our model are:
- Average per theater (on opening weekend)
- Production Budget
- Joel Zwick?????
- Metascore
- Bonnie Hunt?????

Average Per Theater(opening weekend) vs Gross



My Favorite Relationship



Future Improvements

- Look at descriptions/keywords
- Add categorical features for rating/genre