Interpreting the impact of lead actor gender and age on movie gross

Hannah Lewis

Metis, Linear Regression & Web Scraping Module August 6, 2021



Overview

Motivation

• Female protagonists—particularly, female protagonists portrayed by actors over the age of \sim 35—are significantly underrepresented in movies

Objectives

• Determine what (if any) impact these differences in actor demographics (i.e., gender and age) have on the lifetime gross of a movie

Results

• Lifetime movie gross is strongly dependent on how it performs during its opening weekend, the budget, its reception by audiences, and its genre









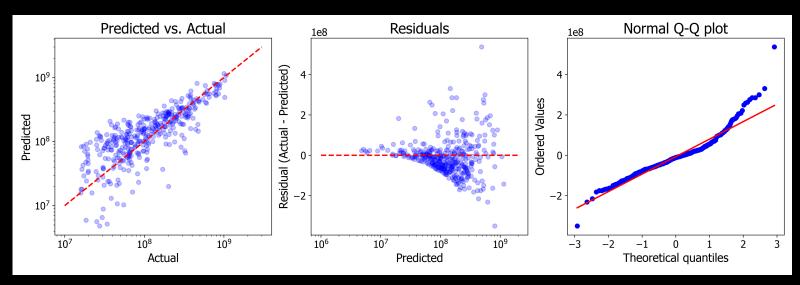
Target: worldwide lifetime gross

Features: domestic opening gross, budget, studio, MPAA rating, run time, genre, lead name, lead age, lead height, lead gender, Rotten Tomatoes

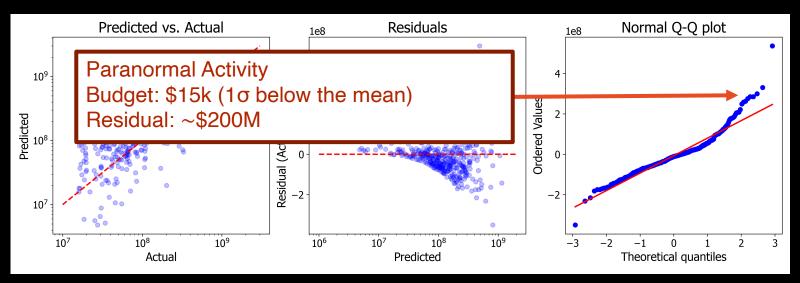
audience and Tomatometer scores, international movie?

- Baseline models trained on domestic opening gross (and its square), budget, and Rotten Tomatoes scores as features
- Select ridge regression model (validation R² ~ 0.720, MAE ~ \$68.4M)

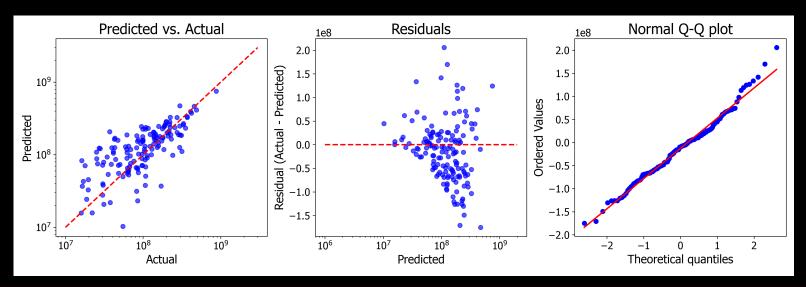
 Expanded model includes lead actor gender and age as features, as well as dummy variables for the studio that produced the movie and the genre of the movie



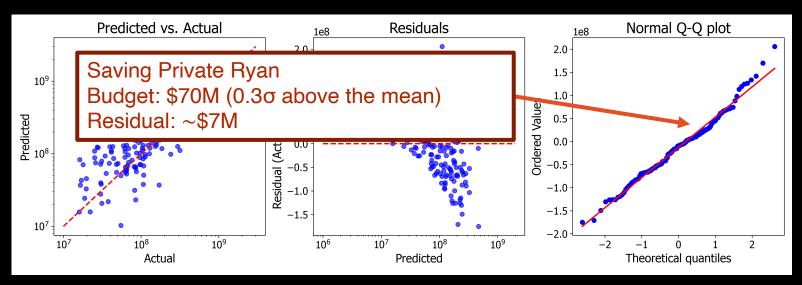
- Model does not generalize well for outliers in budget
- For any budget (\$15k up to 300M), MAE = \$63.4M



- Model does not generalize well for outliers in budget
- For any budget (\$15k up to 300M), MAE = \$63.4M



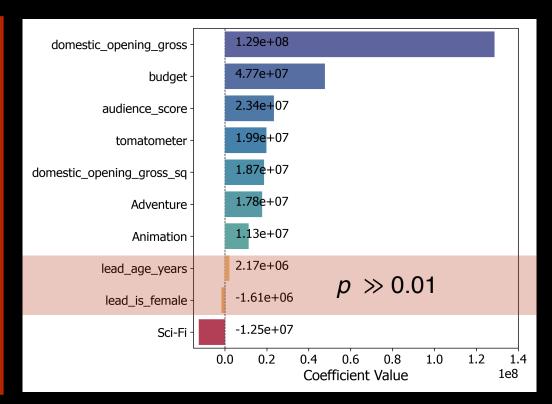
 Model works best for movies with budgets within ±0.5 standard deviations of the mean (\$30M up to \$80M), MAE = \$51.0M



 Model works best for movies with budgets within ±0.5 standard deviations of the mean (\$30M up to \$80M), MAE = \$51.0M

Results: significant features

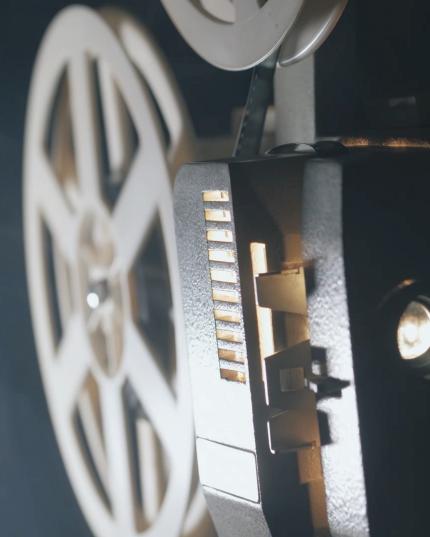
- Significant coefficients
 - *p* < 0.01
 - Reject the null hypothesis that $\beta_i = 0$
- In the final model, lead actor age and gender do not have significant coefficients



Conclusions & Recommendations

- Lead actor gender and age are not among the most important features that determine movie gross
- Studios should produce a larger fraction of movies that are led by a more gender and age-diverse range of actors



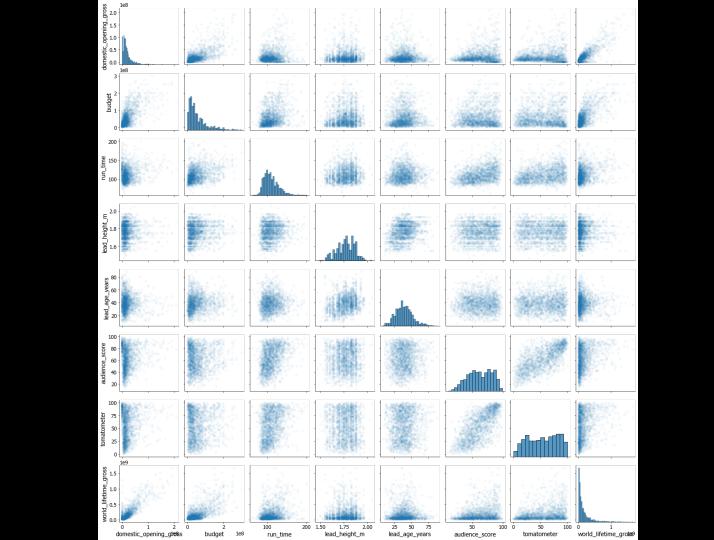


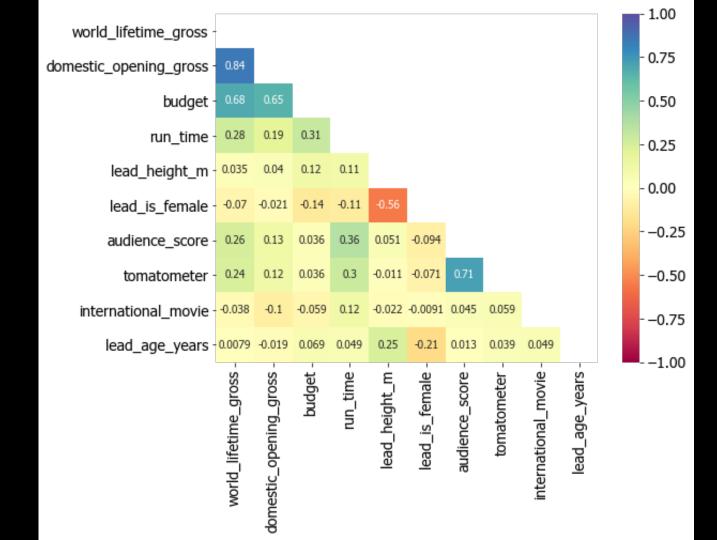
Future Work

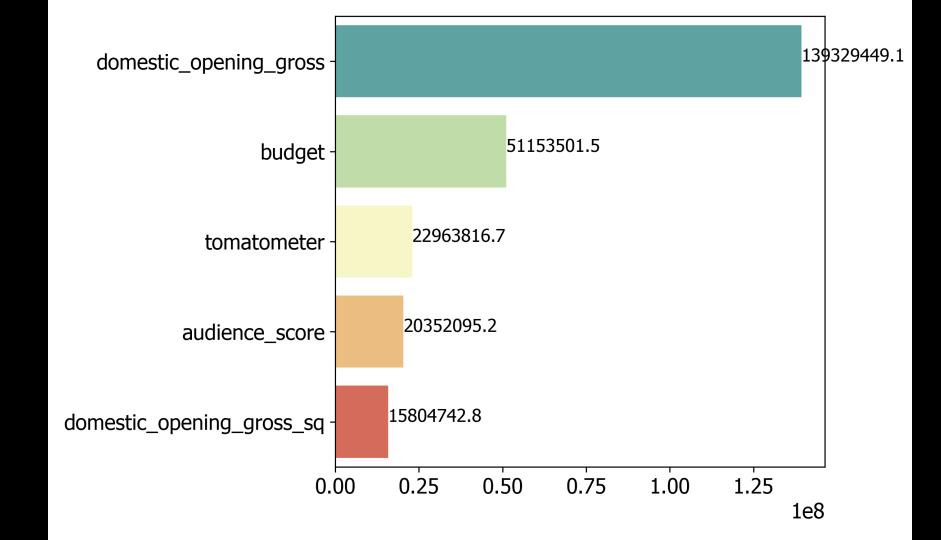
- Scrape movies with lower lifetime gross and broader range of budgets
- Subset data by genre (e.g., action vs. romance) to see in gender does become a significant feature

Questions?









domestic_opening_gross -	1.29e+08	
- pudget	4.//e+U/ 3.342.107	
andience_score	2.34e+U/	
tomatometer -	1.996+07 1.970+07	
domestic_opening_gross_sq -	1.8/e+0/ 1.86+07	
Adventure -	1.0e+0/ 1.13o+07	
Animation -	10 + 0.0 + 0	
Romance -	6.16e+U6	
Twentieth Century Fox -	6.U2e+U6	
New Line Cinema -	5.00e+06	
The Weinstein Company -	4.46e+06	
Thriller -	4.25e+06	
Walt Disney Studios Motion Pictures -	3.18e+06	
Fantasy -	2.98e+06	
Other genre -	2.75e+06	
lead_age_years -	2.1/e+06	
Universal Pictures -	2.02e+06	
Sony Pictures Entertainment (SPE) -	1.57e+06	
TriStar Pictures	6.76e+05	
Fox Searchlight Pictures -	2.14e+05	
Horror -	9.31e+04	
Other studio -	7.72e+04	
- Miramax	-1.08e+05	
STX Entertainment -	-9.55e+05	
DreamWorks Distribution -	-1.26e+06	
lead is female	-1.61e+06	
- Mystery	-2.26e+06	
Focus Features -	-2.28e+06	
Warner Bros	-2.57e+06	
Biography -	-2.72e+06	
Paramount Pictures -	-2.75e+06	
Screen Gems -	-2.81e+06	
Metro-Goldwyn-Mayer (MGM)	-2.94e+06	
Family -	-3.42e+06	
Action -	-6.70e+06	
- Crime	-6.94e+06	
Lionsdate -	-7.51e+06	
Drama -	-7.61e+06	
Comedy -	-7.82e+06	
Sci-Fi -	-1.25e+07	
0.00	0 0.25 0.50 0.75 1.00	0 1.25
	Coefficient Value	1e8