



How US comedies fare worldwide

Metis Bootcamp, Fall 2018
Arianna Breslauer

Past US comedies may have travelled less well overseas



Now we have a more globalized movie marketplace



Web connected foreign consumers



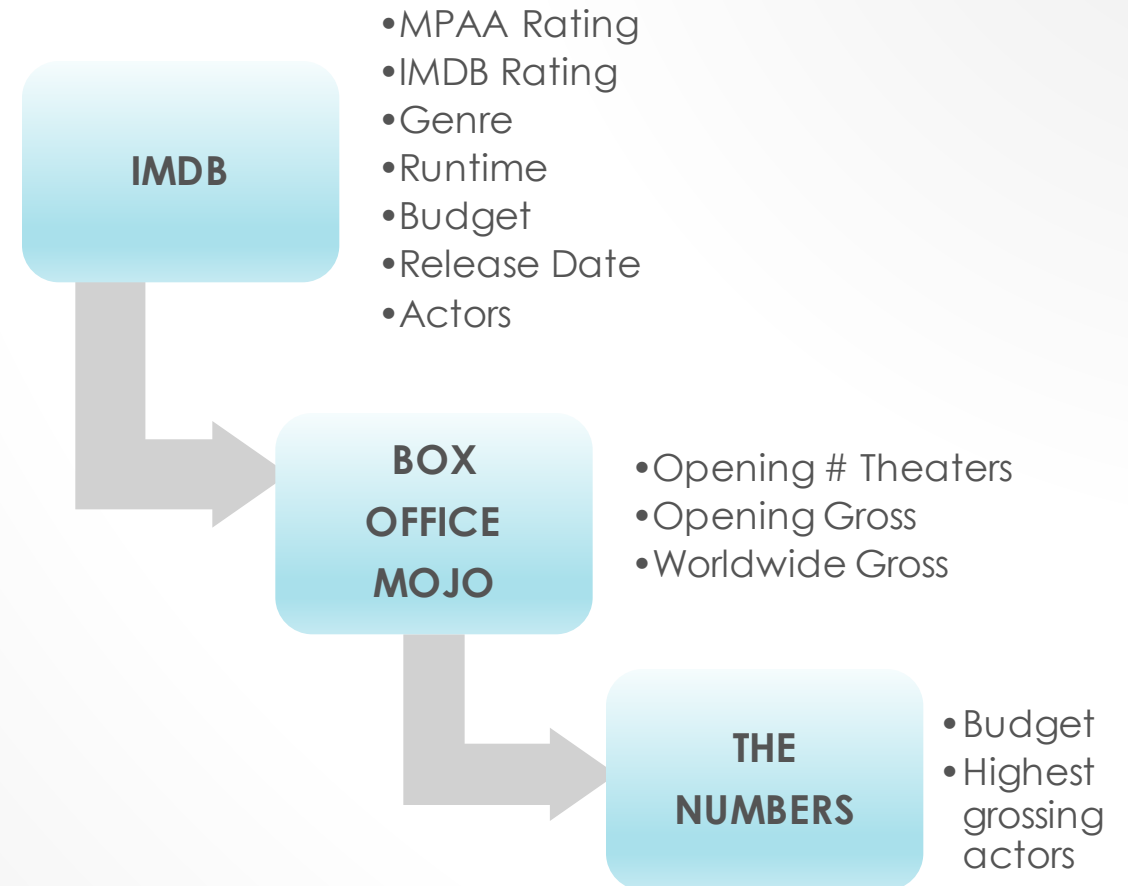
Purpose-built comedies

Initial hypotheses on worldwide gross

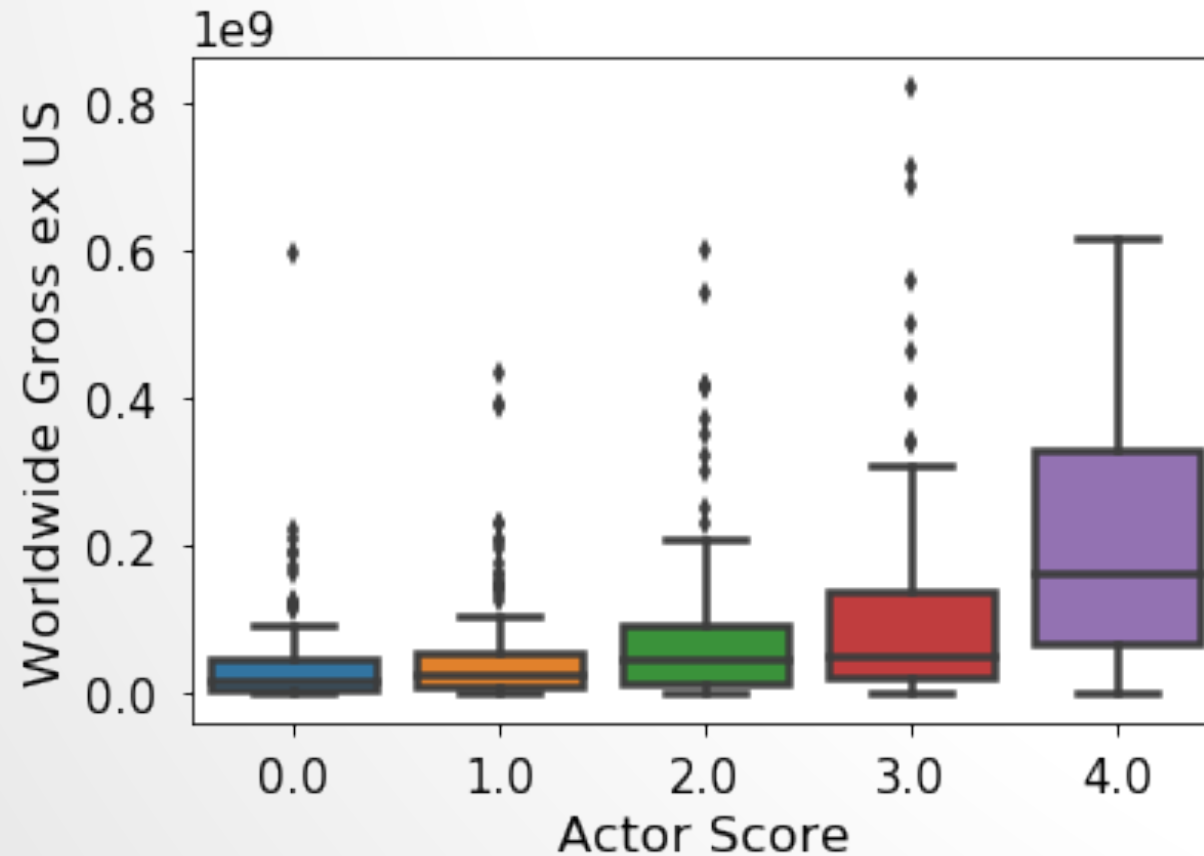
- Opening US domestic gross a strong predictor
- Starry actors play a large role
- Budget also important

Setting up the model

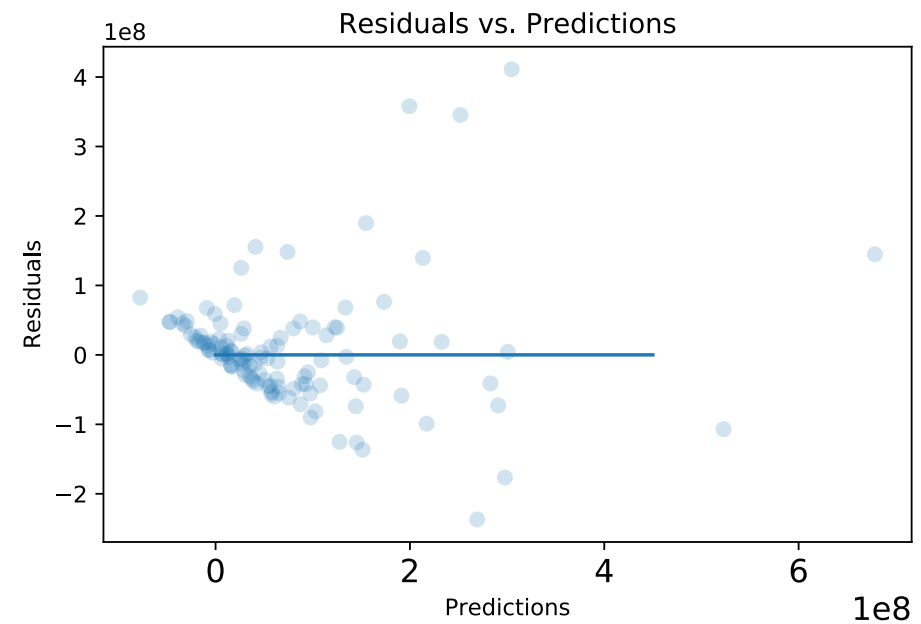
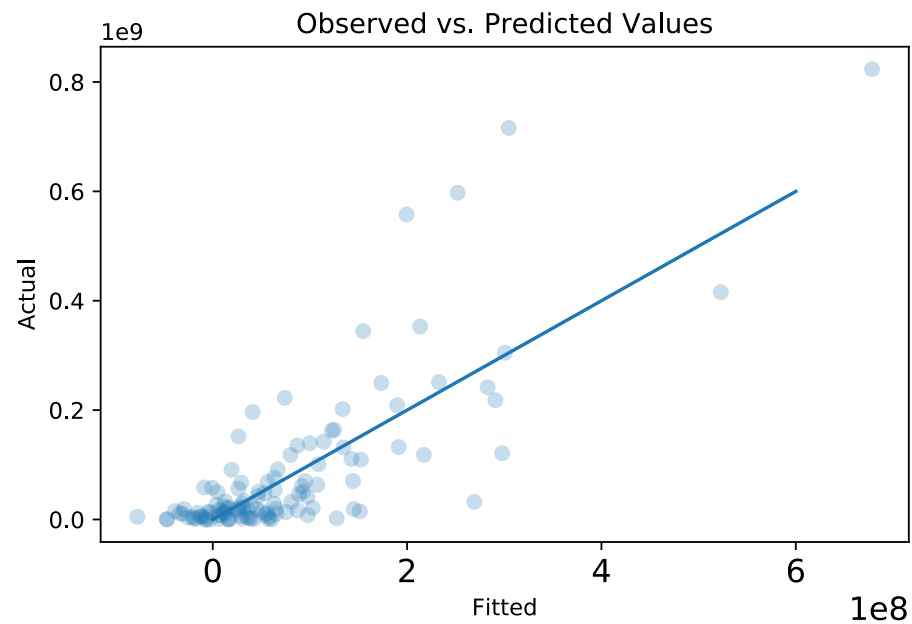
- Scraped with BeautifulSoup
- Generated an Actor Score based on highest grossing actors
- Limited data set to movies showing in 100+ theaters
- Target variable:
 - Worldwide gross ex US



Initial EDA: high actor scores appear to be correlated with worldwide gross ex US



R^2 is 0.625 for linear regression model



Ridge performs best across linear regression features

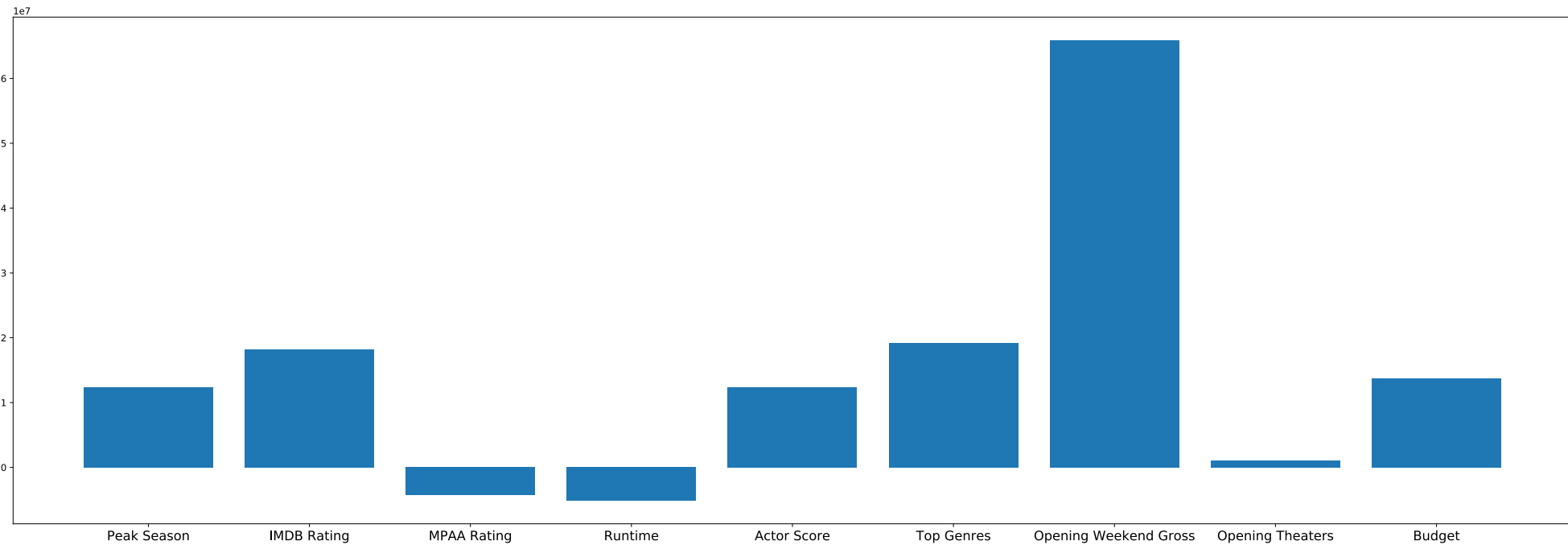
Train/Validation/Test	R ²
Linear Regression	0.62477
Ridge Regression	0.62481
Lasso Regularization	0.62472
Polynomial Features	0.44392

Cross-Validation/Test	R ²
Linear Regression	0.44575
Ridge Regression	0.44613
Lasso Regularization	0.44573
Polynomial Features	0.44575

RIDGECV ARRAY

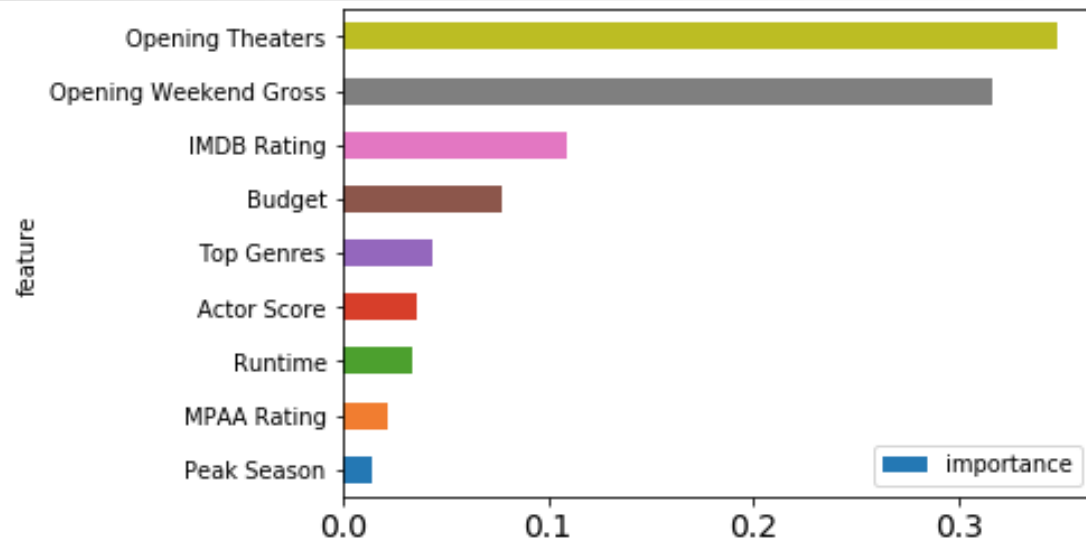
```
[0.61579173  
-0.32579784  
0.59130861  
0.69818528  
0.65115842]
```


Scaled Relative Feature Strength – Ridge; initial hypotheses seem to hold



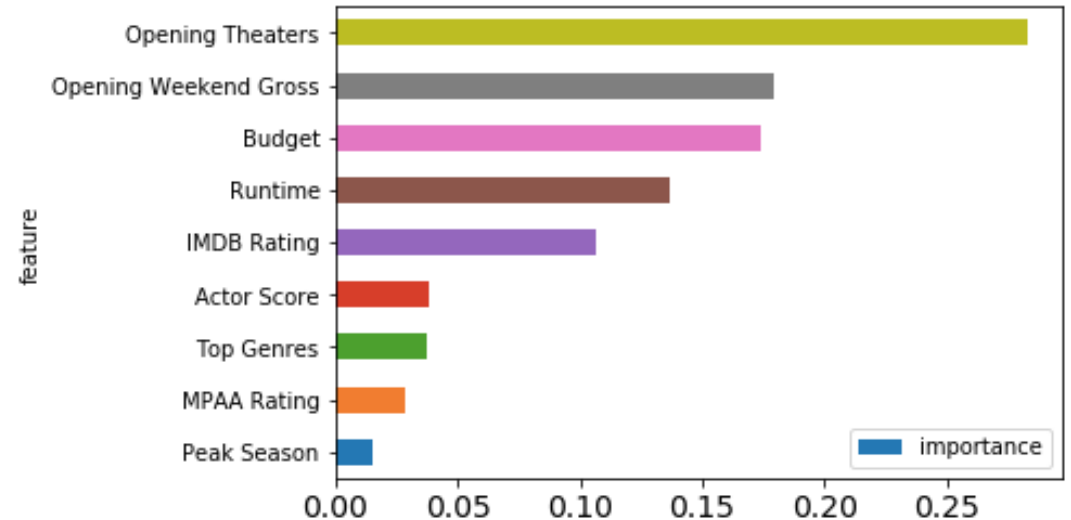
Tree based methods outperform previous models; RMSE reduced by half

RANDOMFOREST



$R^2 = 0.66044$
RMSE = 81,411,737

GRADIENTBOOSTING



$R^2 = 0.69044$
RMSE = 77,732,225

Closing remarks

- Today's big hits earn ~80% of their box office revenue internationally
- Thinking about foreign audiences has become de rigueur
- Next steps to improve prediction model:
 - Location of filming (filmed ex US?)
 - Plot (high-concept, universal appeal?)
 - Production company (large distribution network worldwide?)