



# Movie budget predictions

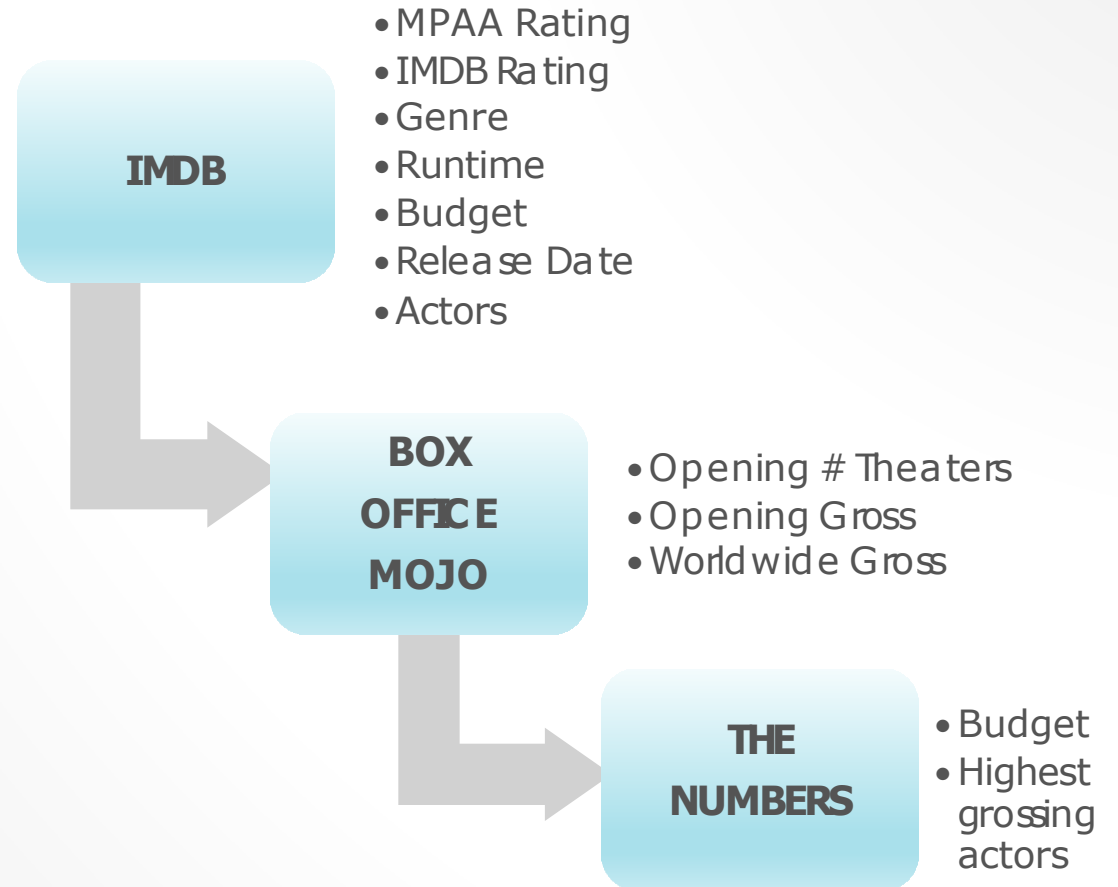
NAJLA BIN-MELHA  
MUTAB ALZUBAIR  
TAHANI ALDOSARI

# Initial hypotheses on worldwide gross

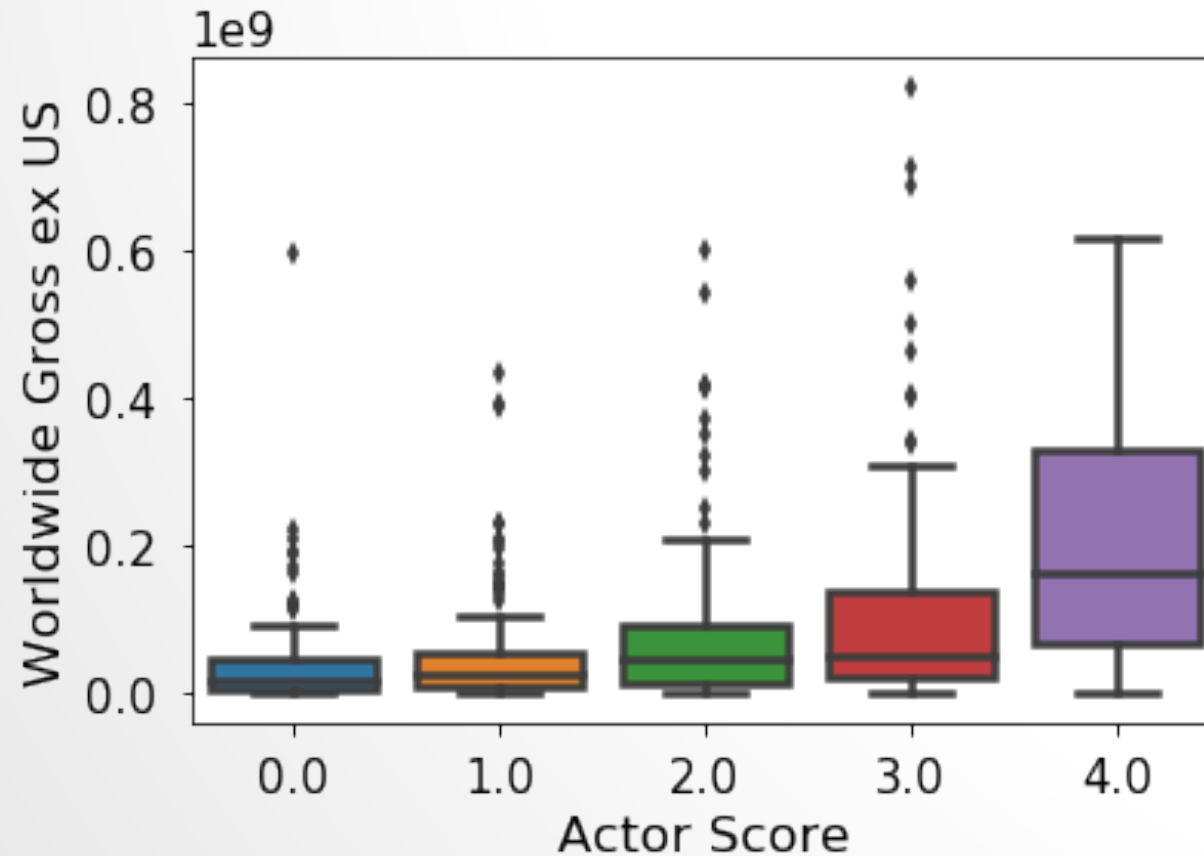
- Opening US domestic gross a strong predictor
- Starry actors play a large role
- Budget also important

# Setting up the model

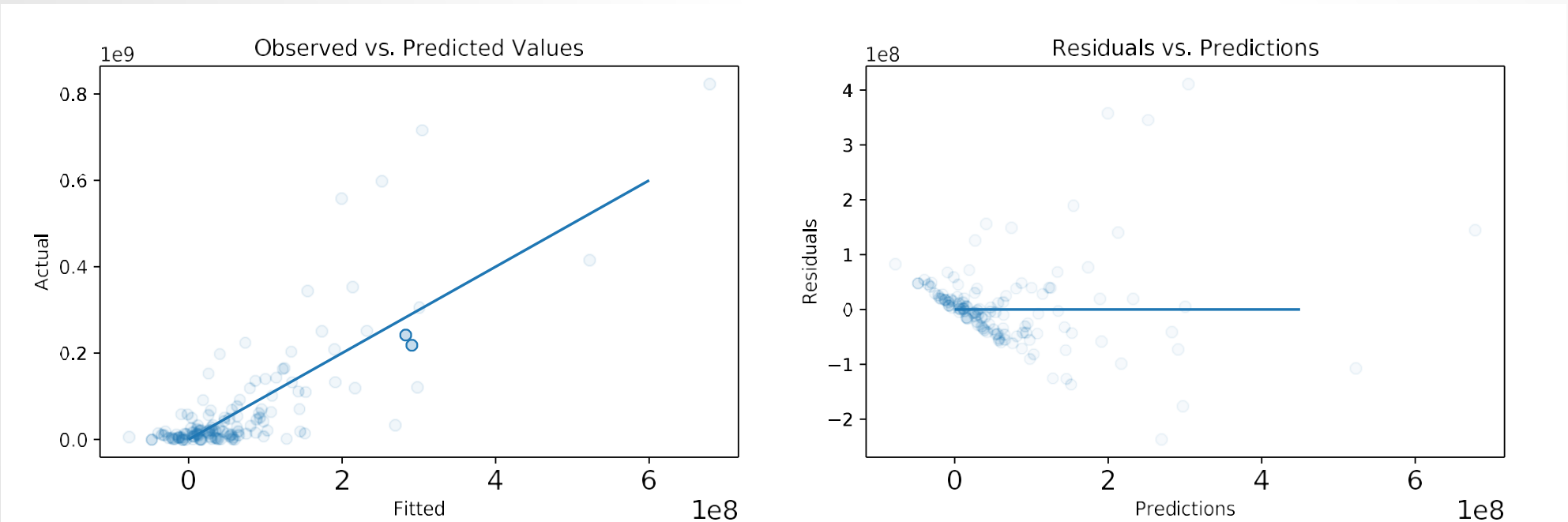
- Scraped with BeautifulSoup
- Generated an Actor Score based on highest grossing actors
- Limited data set to movies showing in 100+ theaters
- Target variable:
  - Worldwide gross ex US



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



# R<sup>2</sup> is 0.625 for linear regression model



# Ridge performs best across linear regression features

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

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

**RIDGE CV ARRAY**

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

# Conclusion

- Today's big hits earn 80% of their box office revenue internationally
- Thinking about different audiences has become important
- Next steps to improve prediction model:
  - Location of filming
  - Plot
  - Production company



THANK YOU