

Predicting Film Success

Ryan Rappa

About me / my project

— — —

- Political scientist/economist
- Interested in film and entertainment industry
- Project: predict whether a film is profitable

Project workflow

Collect
data



Clean
data

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

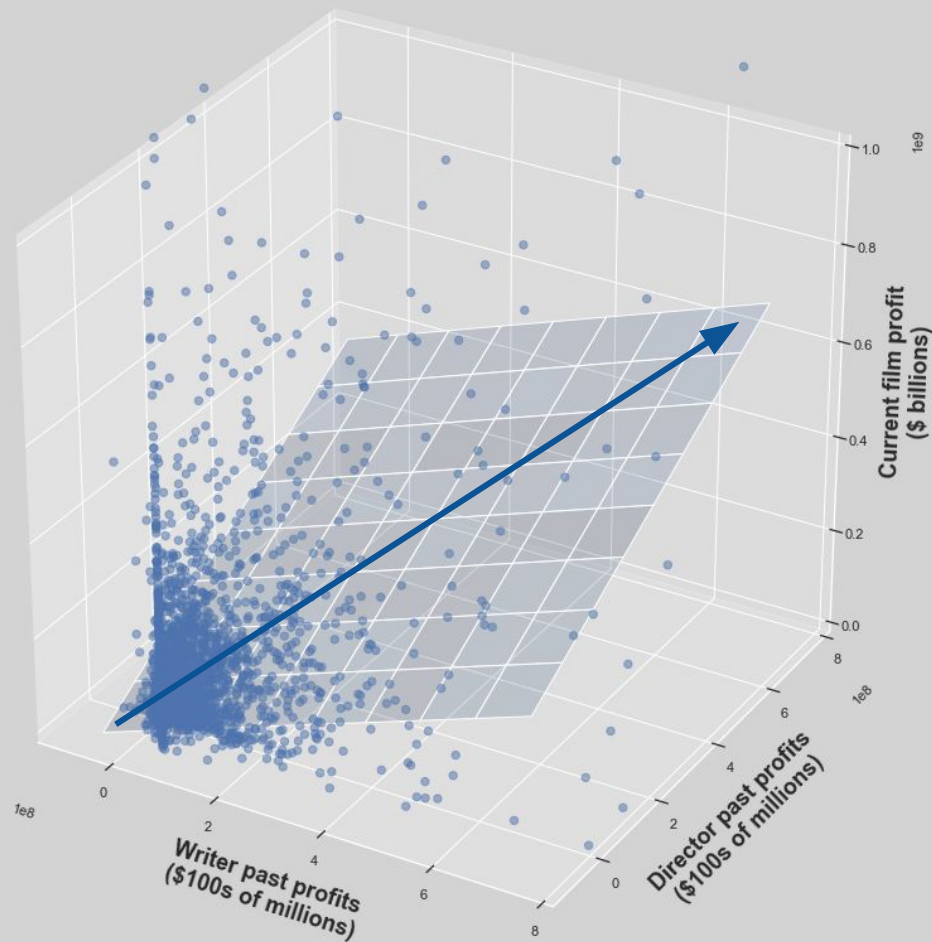
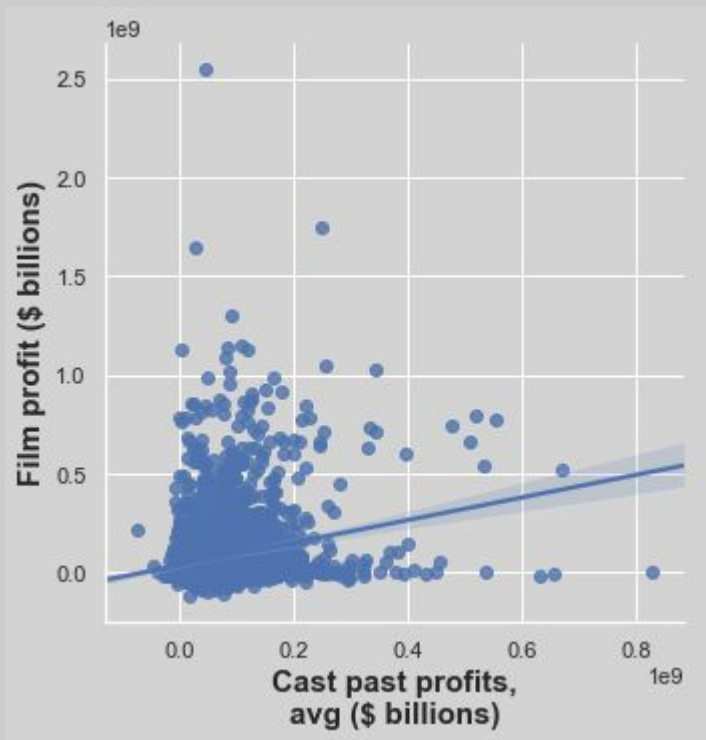
EDA

matplotlib
seaborn

Model



Visualizing the data

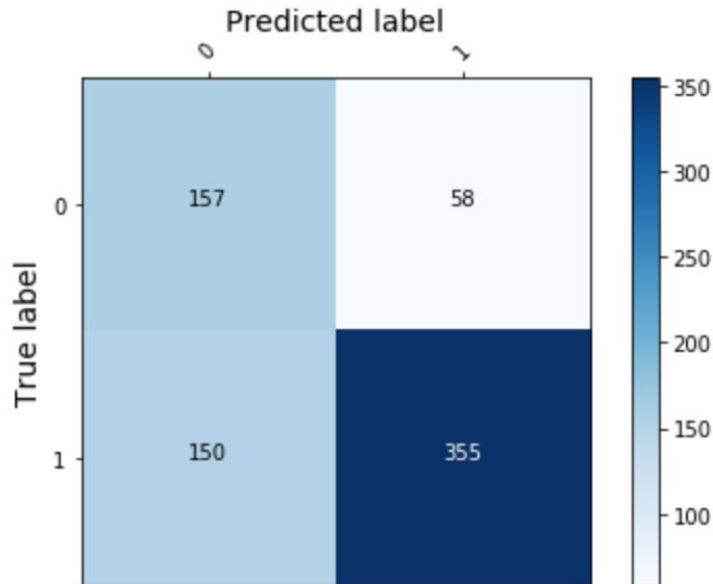


Findings

- **Major predictors of success:**
 - Number of int'l releases
 - Director past success
 - Cast past success
 - Budget and “adjusted budget”
 - Competing films: cast, director, writer, budget
 - Season (avoid Aug-Oct)
 - Runtime (aim for 90+ min)
- **Can predict with 71% accuracy and 86% precision -->**
- **Some unprofitable films are being misclassified** (top right square)

Prediction results

Gradient boost w/ 70% decision boundary:



Accuracy = 0.711
Precision = 0.860
Recall (TPR) = 0.703
Fallout (FPR) = 0.270

Next steps / variables to consider

— — —

- **Collect more data**
- Story/script
- Posters
- Trailers
- Marketing strategy
- Domestic vs. foreign box office
- Ancillary revenues
- Piracy





Contact

— — —

Ryan Rappa

ryanlrappa@gmail.com

Thank you for listening.