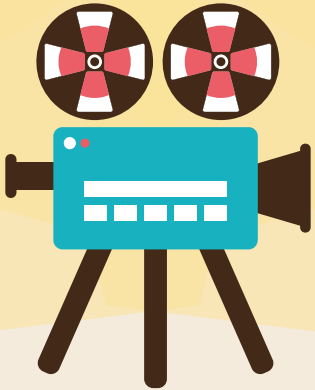
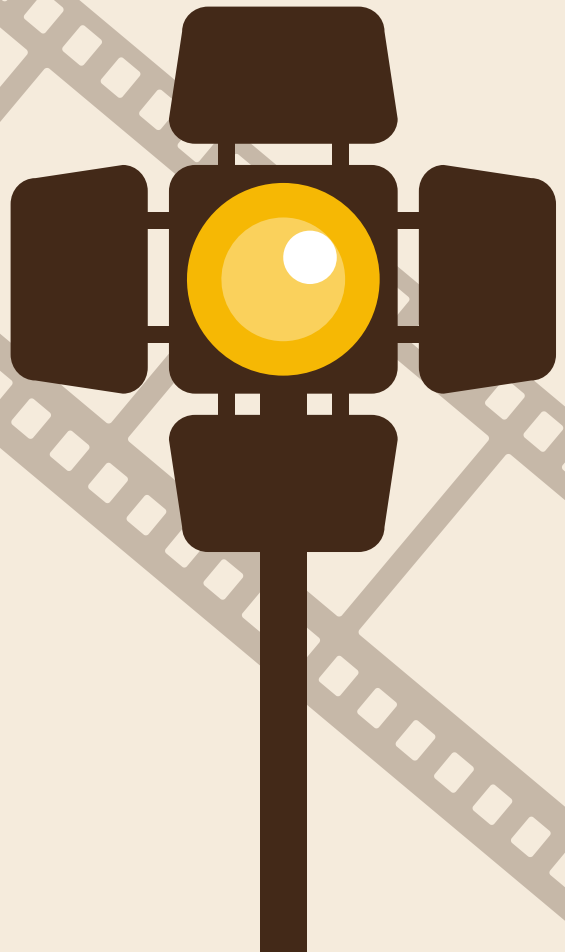
Predicting An Actor's Career Number of Film Credits



Motivation

- Advise prospective and newly working film actors on which projects to pick to maximize their odds of future success in the industry

Methodology



Gathering Data

Step 1

- Web scraped list of actors from IMDB (1,500 names)
- All from US made films released in early 2000

Gathering Data

Step 1

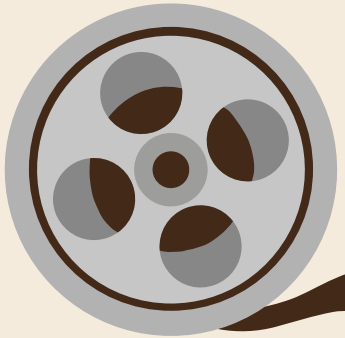
- Web scraped list of actors from IMDB (1,500 names)
- All from US made films released in early 2000

Step 2

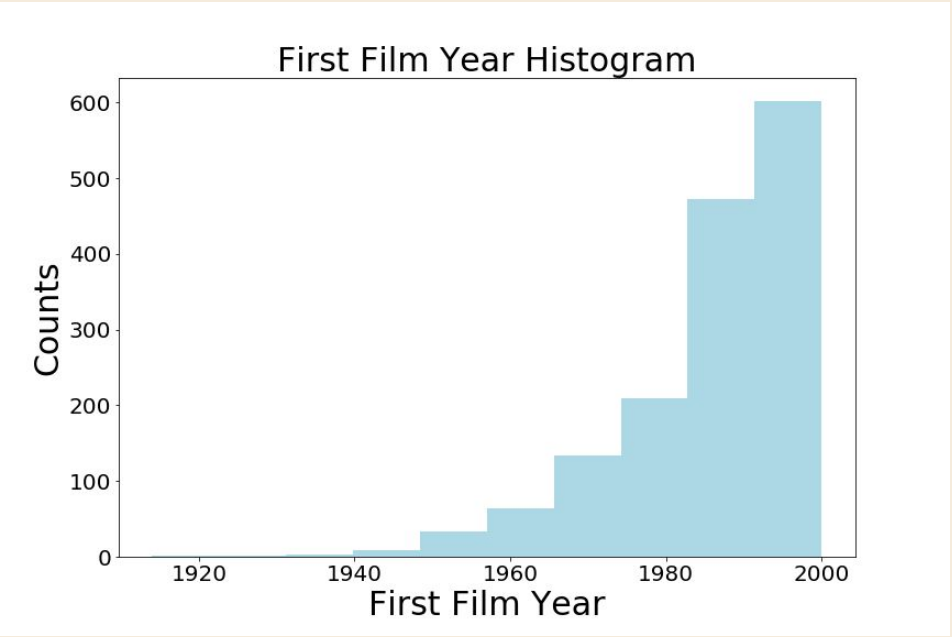
- Collected data from the first three films of each actor from IMDB Pro
- **Features**:
  - Year of first film release
  - Average rating of films
  - Genres of the films

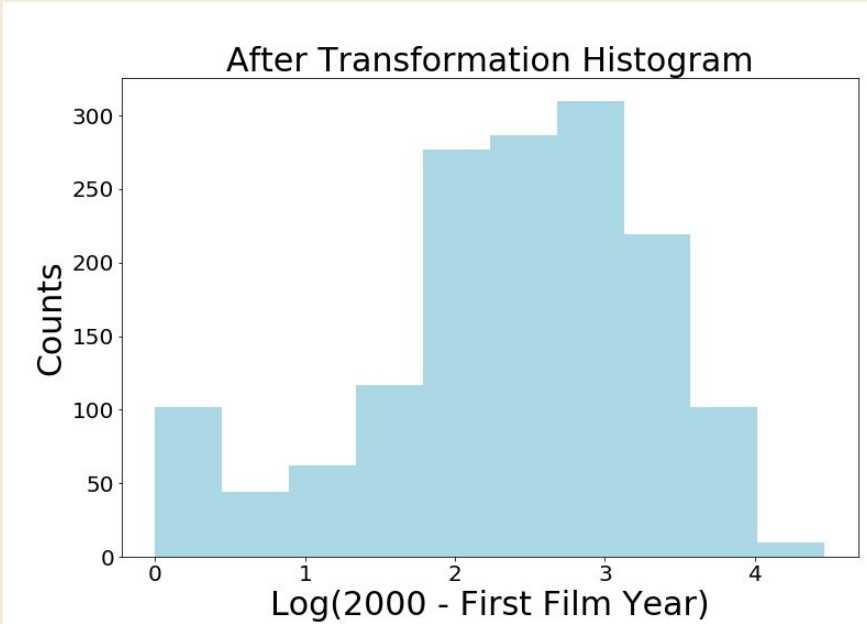
Modeling the Data

- Removed outliers from dataset with credits over 300
- Added as a feature the logarithm of 2000 minus the First Film Year (referred to as Log-Diff) to try normalize it's distribution and capture non-linearity



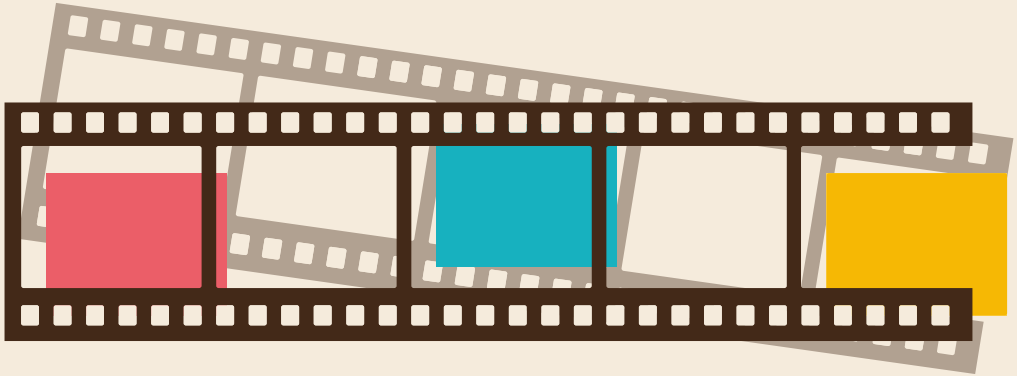
Modeling the Data





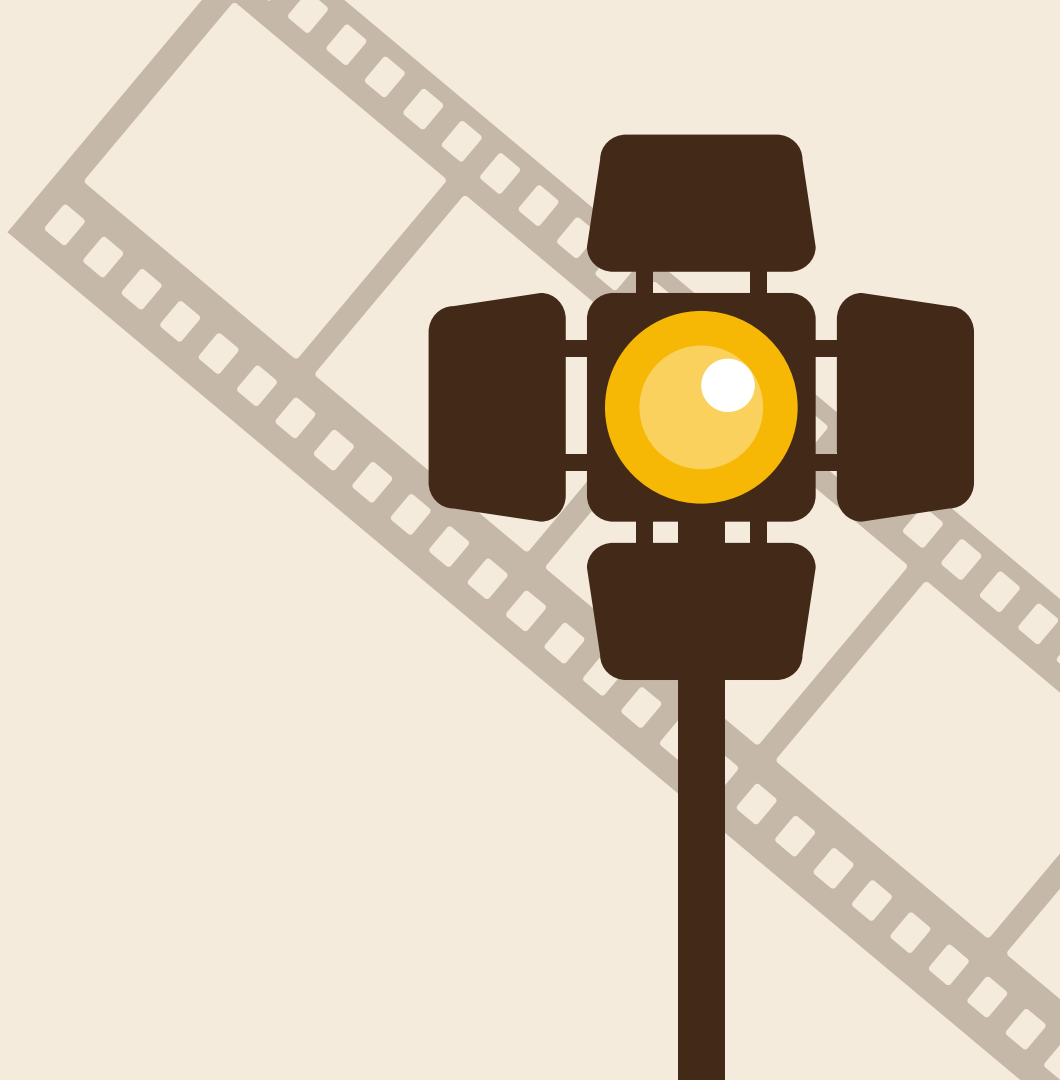
Modeling the Data

- Lasso Regression to see which features were assigned coefficients of zero
- Trained new model on only non-zero features using Lasso with cross validation (5 folds) and optimized alpha

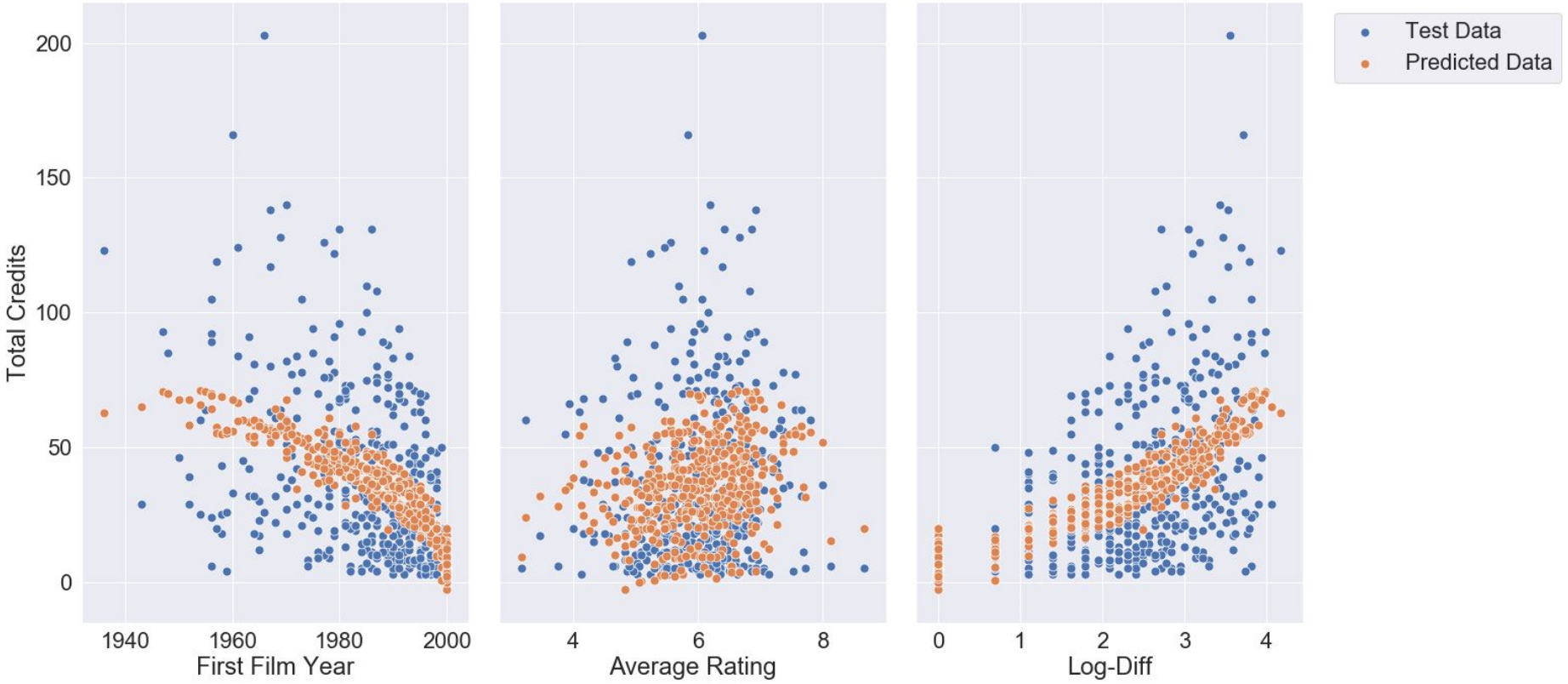


Kept Features	Coefficients
First Film Year	-4.47
Average Rating	0.90
Log-Diff	9.48
Comedy	-0.80
Fantasy	-0.97
Thriller	-2.16
Musical	-0.52
Documentary	-0.86
Animation	1.10

Results



RESULTS


Test Data
Predicted Data
Total Credits
200
150
100
50
0
1940
1960
1980
2000
First Film Year
4
6
8
Average Rating
0
1
2
3
4
Log-Diff

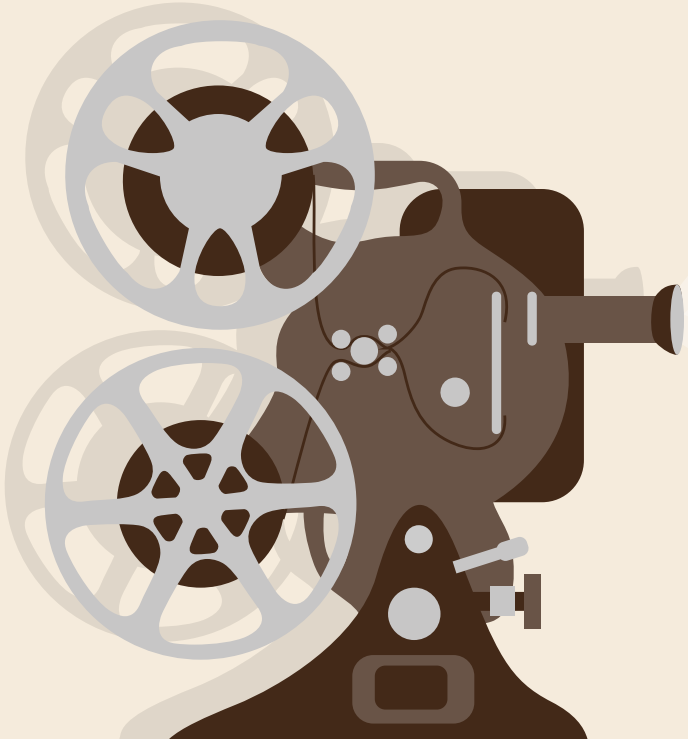
RESULTS

- Model ultimately did not have enough features to make accurate predictions
- Too much variance in the data unaccounted for
- Getting offers = Getting future offers

Metrics	Training Set	Test Set
R^2	0.25	0.24
MAE	20.9	19.8

Future Work

- Potentially Valuable Features:
- Study found 73% of working actors from middle/upper class
- Most actors reach fame by 29
- Race and gender correlation to total credits



Thank you

Questions?

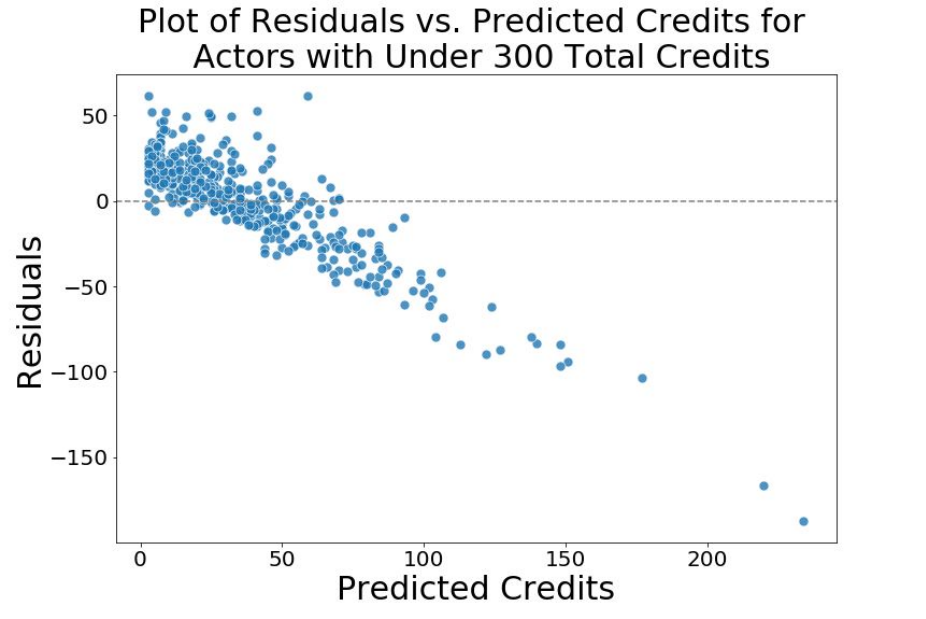
CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

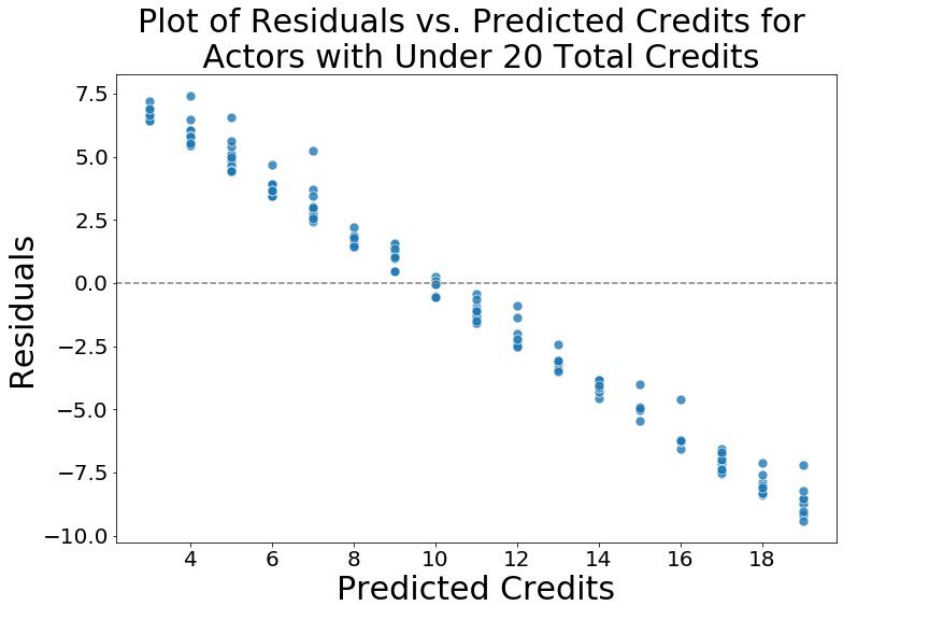

DIRECTOR
PRODUCCION
ESCENA
TOMA
PLANO
CAMARA
INT
SONIDO
DIA
NOCHE
EXT
FECHA

Appendix

Plots of the Model's Residuals vs. Predicted Credits

- These plots show the negativity sloped nature of the residuals of my model for the entire dataset (left) and for only actors with 20 or less credits (right)





Heat Map of Kept Features