



Predicting Gross Domestic Movie Sales

By Mason Ellard

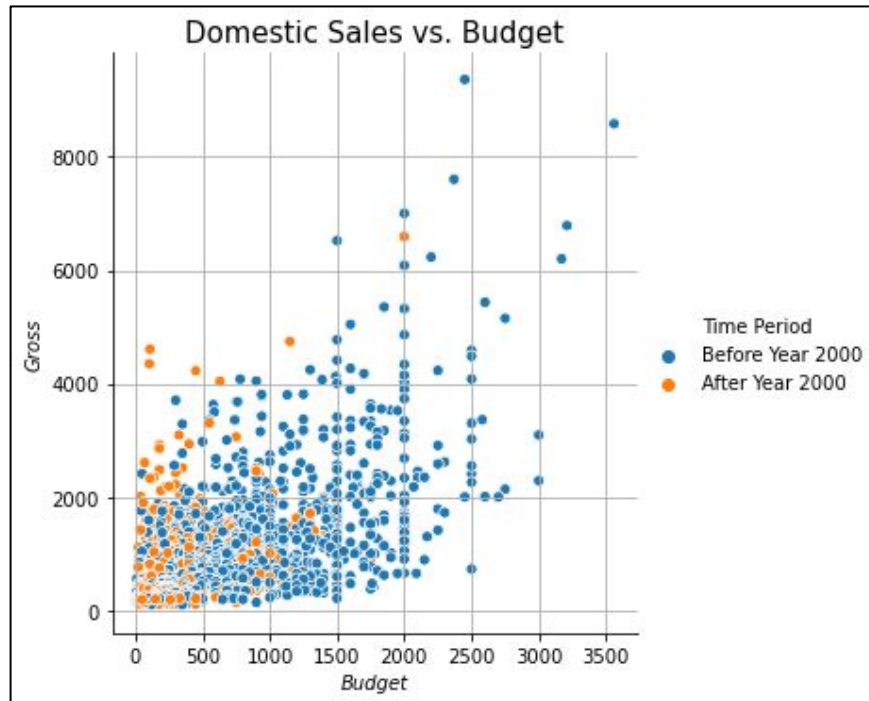
Data

- Scraped from IMDb
- 2363 Movies
- 8 Features:
 - Budget
 - Runtime (in minutes)
 - Days since release date
 - Genre
 - MPAA Rating
 - Directors
 - Writers
 - Cast



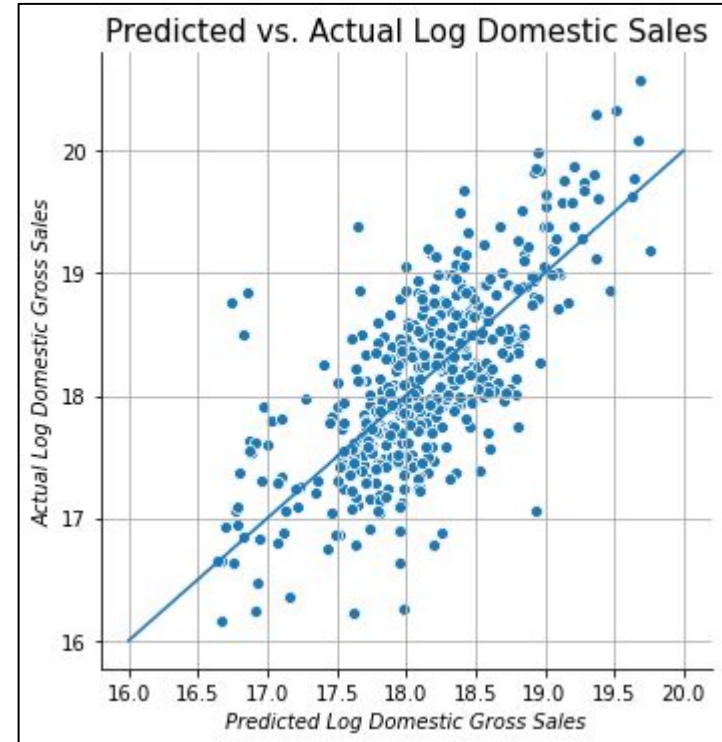
Why Do Media Companies Need Machine Learning?

- Box Office performance is difficult to predict
- The industry is getting more competitive
- Performance is getting harder to predict



Machine Learning Models Accurately Predict Gross Sales

- Much more accurate than using budget
- Executives can make more informed decisions
- Room for improvement





Model Results

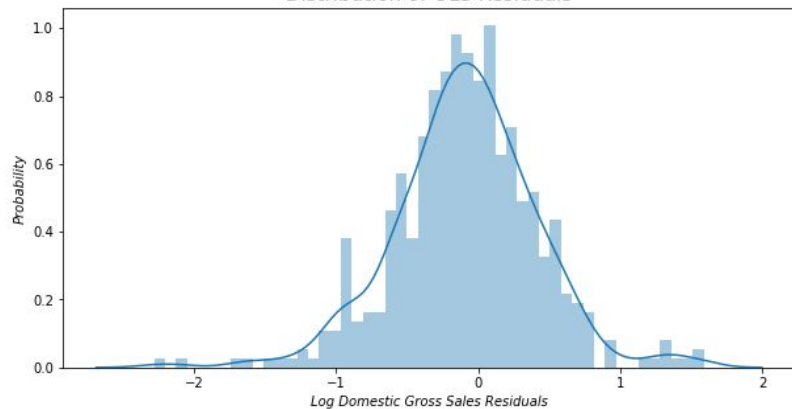
- Fit 5 different models across 4 categories
- RMSE was greater than 65M for all but one
- XGBoost vastly outperformed all other models

Model Name	RMSE	R-Squared
OLS	68.8M	.521
Lasso	65.1M	.542
Ridge	64.9M	.537
K-Nearest Neighbors	65.4M	.521
XGBoost	53.1M	.612



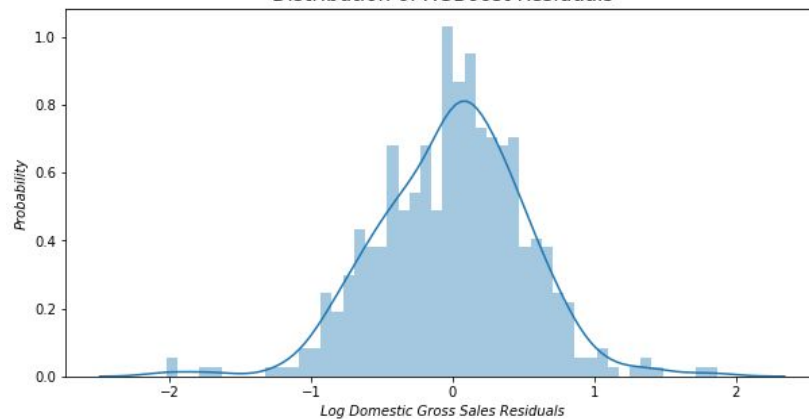
OLS

Distribution of OLS Residuals



XGBoost

Distribution of XGBoost Residuals



An Application for Quick, Easy Box Office Predictions

Domestic Gross Revenue Predictor

Enter Movie Parameters:

Budget:

0

-

+

Runtime (in minutes):

0

-

+

Number of Days After Release Date:

360

-

+

Select a Rating:

R

▼

Genre(s):

Choose an option

▼

Directors:

Choose an option

▼

Writers:

Choose an option

▼

Cast:

Choose an option

▼

Domestic Gross Revenue Prediction:

Enter Movie Parameters:

Budget:

50000000

-

+

Runtime (in minutes):

120

-

+

Number of Days After Release Date:

360

-

+

Select a Rating:

R

▼

Genre(s):

Crime X

Thriller X

Romance X

⊙ ▼

Directors:

Quentin Tarantino X

⊙ ▼

Writers:

Quentin Tarantino X

⊙ ▼

Cast:

Johnny Depp X

Channing Tatum X

⊙ ▼

Domestic Gross Revenue Prediction:

59240820.0

Questions?