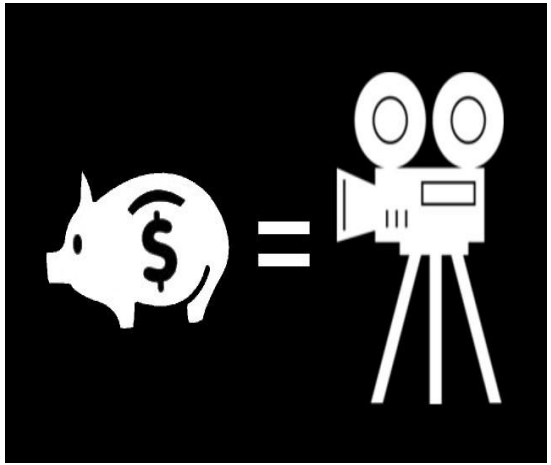
Domestic Revenue Prediction Model for Movies

Pramila Chaudhary

BACKSTORY



- Prediction model for an Investment Firm

- Interested in Knowing the Domestic Revenue of movies before release

Data Scraped

From : https://www.boxofficemojo.com

Target: Total Domestic Revenue

For : Movies (2011 - 2020)

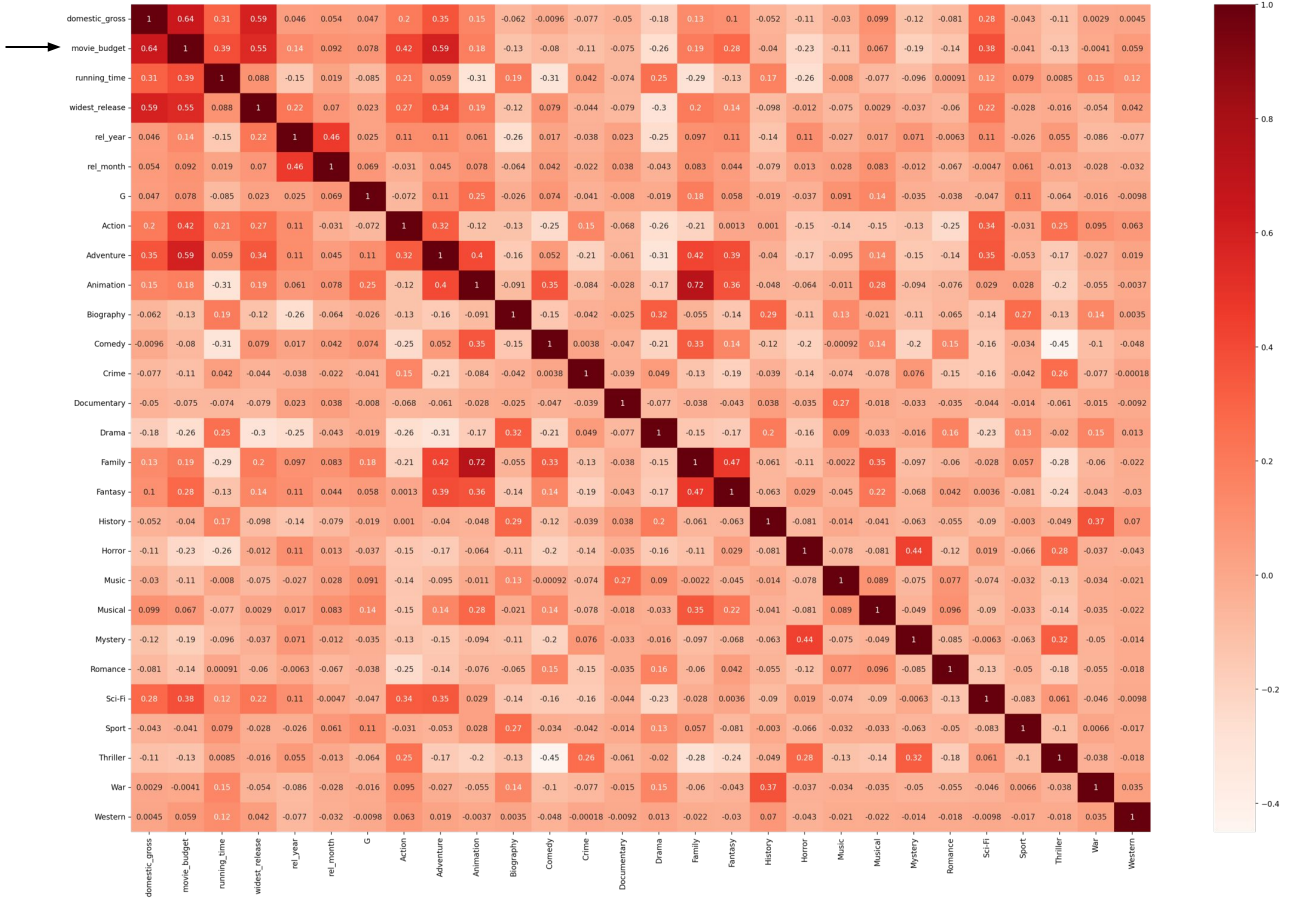
Scraped Data points: 2000

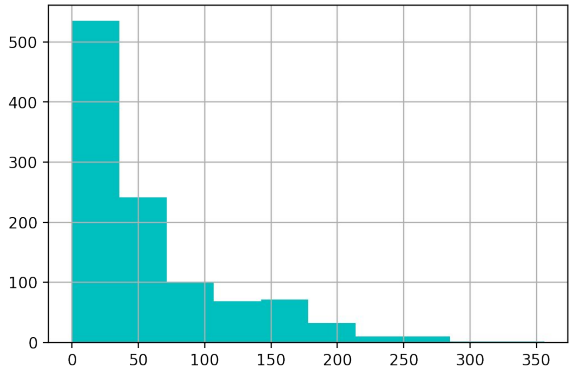
Features Scraped : 12

Process:

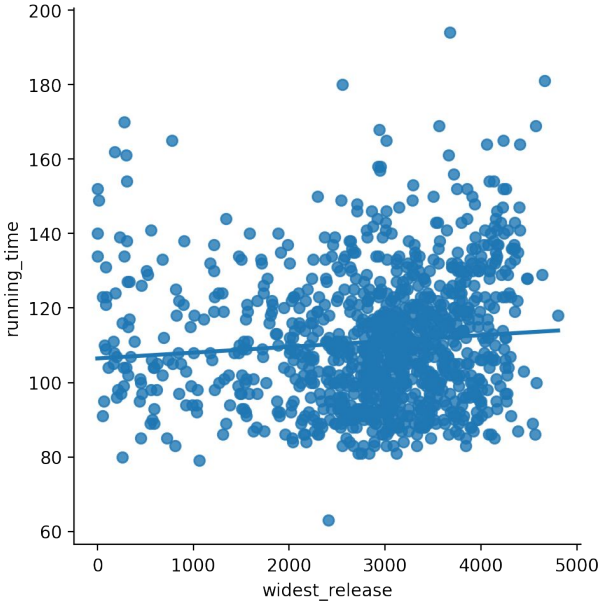
1. Web Scrape
2. EDA to select features
3. Regression Analysis
4. Prediction

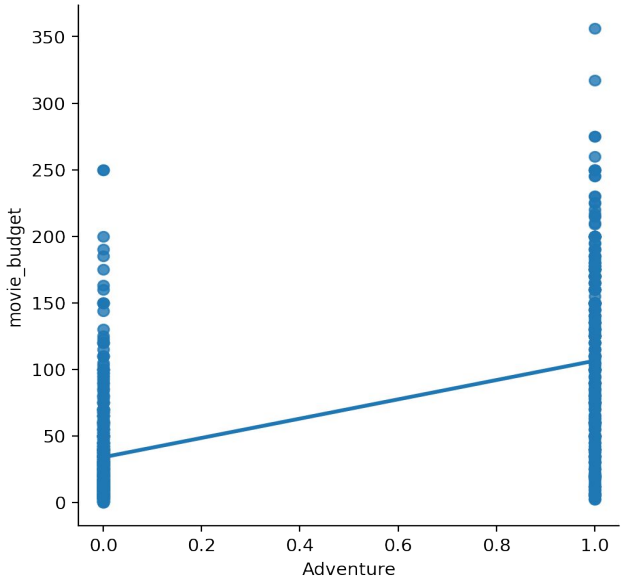
Movie budget



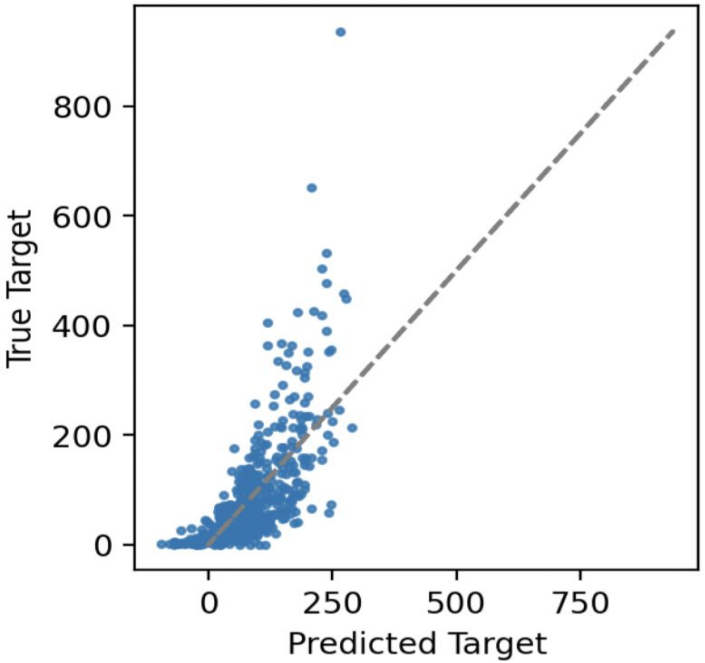


Feature to Feature relationship

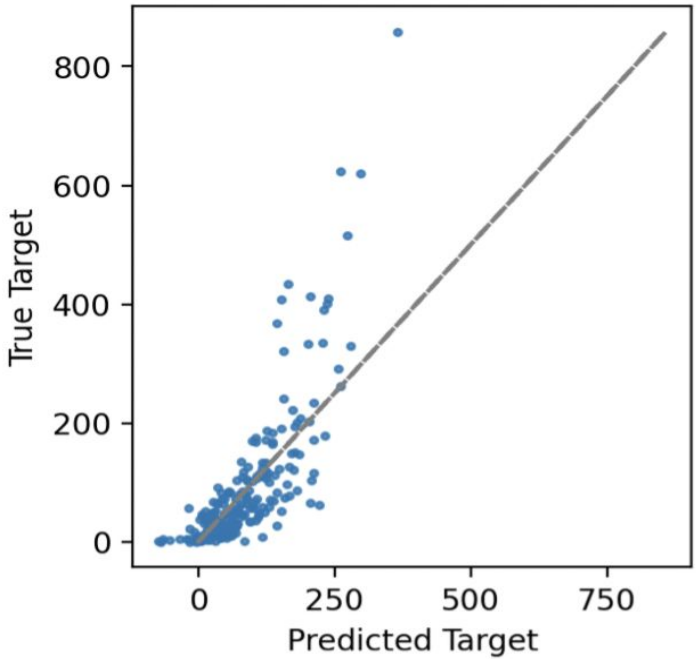




Linear Regression model on train/ validation data

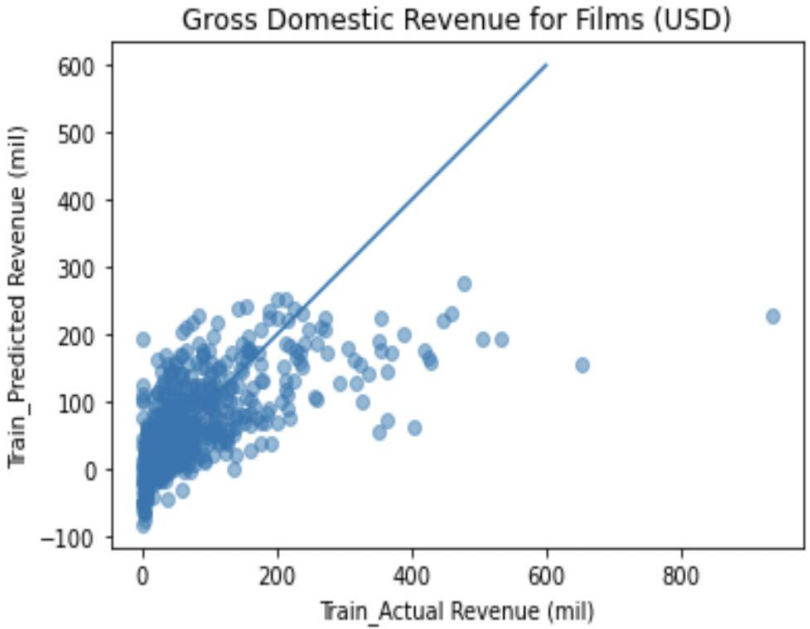


R2 for train = 0.49
MSE = 66.87

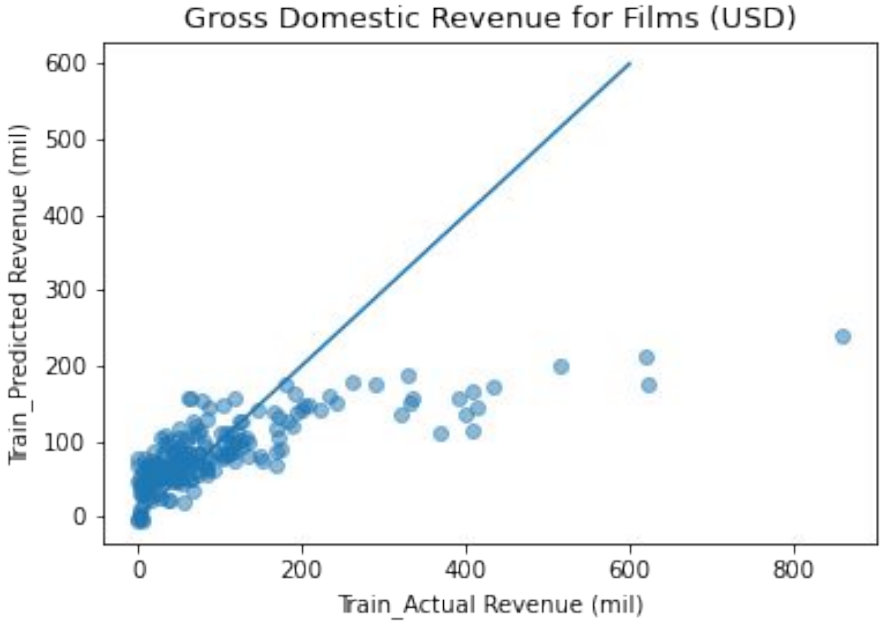


R2 for val = 0.56
MSE = 77

Lasso model on train /validation data set

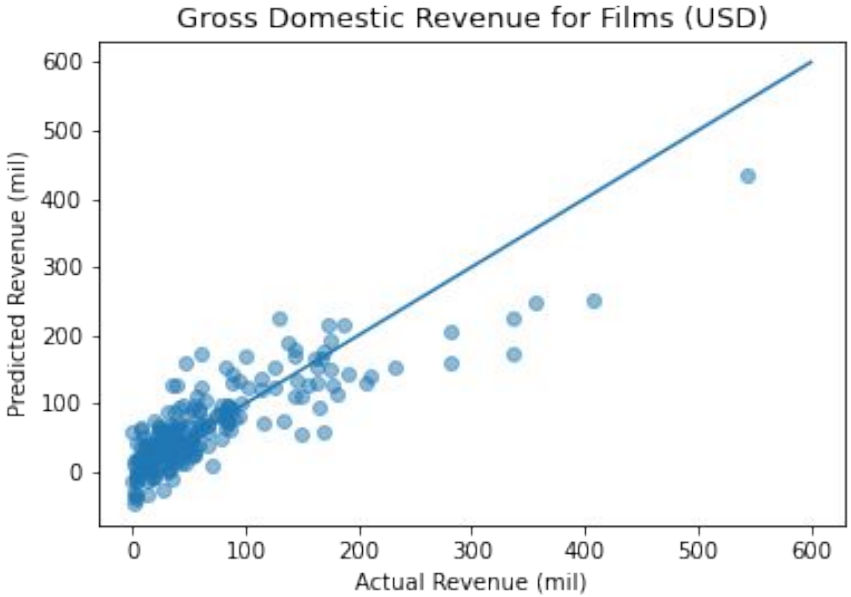


R2 for train = 66
MSE = 74.38



R2 for train = 77
MSE = 91

Final prediction on test data set using Linear Regression model



R2	71
MSE	40

Conclusions

- Prediction with linear regression showing better result than lasso model
- R2 = 71
- Mean Squared Error (MSE) = 40

- Future Direction:
    - Focus on feature selection to avoid overfitting
    - minimize error
    - Including more feature