

# BestFlix and Chill



Linear Regression - Predict  
Domestic Revenues



Reynold D'Silva



Objective: To analyze factors that contribute to the box office collection of films and leverage that to predict future domestic revenues

Factors: Ratings, Budget, Genre, Runtime, Release Date, Worldwide Gross, International Gross, Domestic Gross, Opening Weekend Collection

## Tools

- Python
- Statsmodels
- Scikit-Learn
- Pandas
- NumPy
- Matplotlib
- Seaborn

# Methodology

01

Web Scrape

Box Office Mojo



02

Data Cleaning

Manipulating Data Set to  
fit the desired model



03

Predictive Analysis

Gross Domestic Revenue =  
(Domestic Gross – Opening  
Weekend)



Data Modeling

Linear Regression, Lasso  
and Ridge

04

# Linear Regression Model

$X$

International Gross  
Budget  
Opening Weekend

$y$

Gross Domestic  
Revenue

# Model R-Squared (Cross Validation)



Linear Regression

0.47821



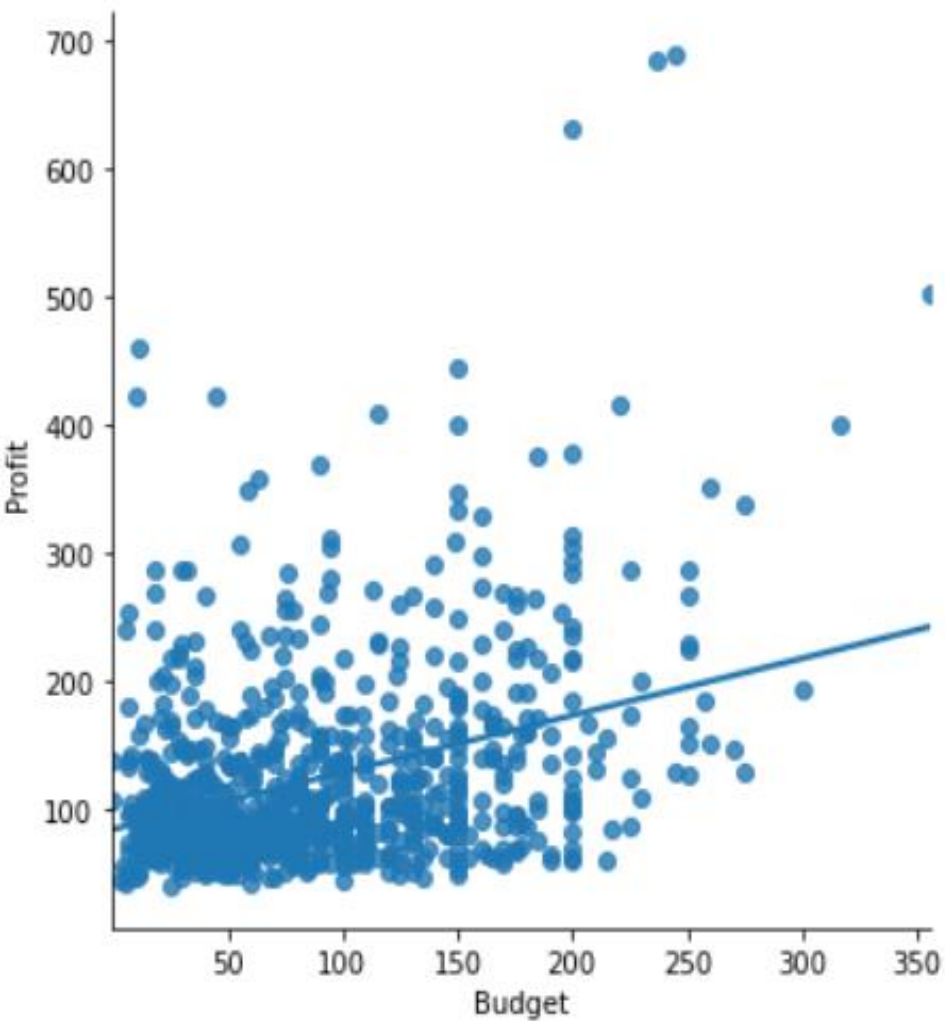
Lasso

0.47835



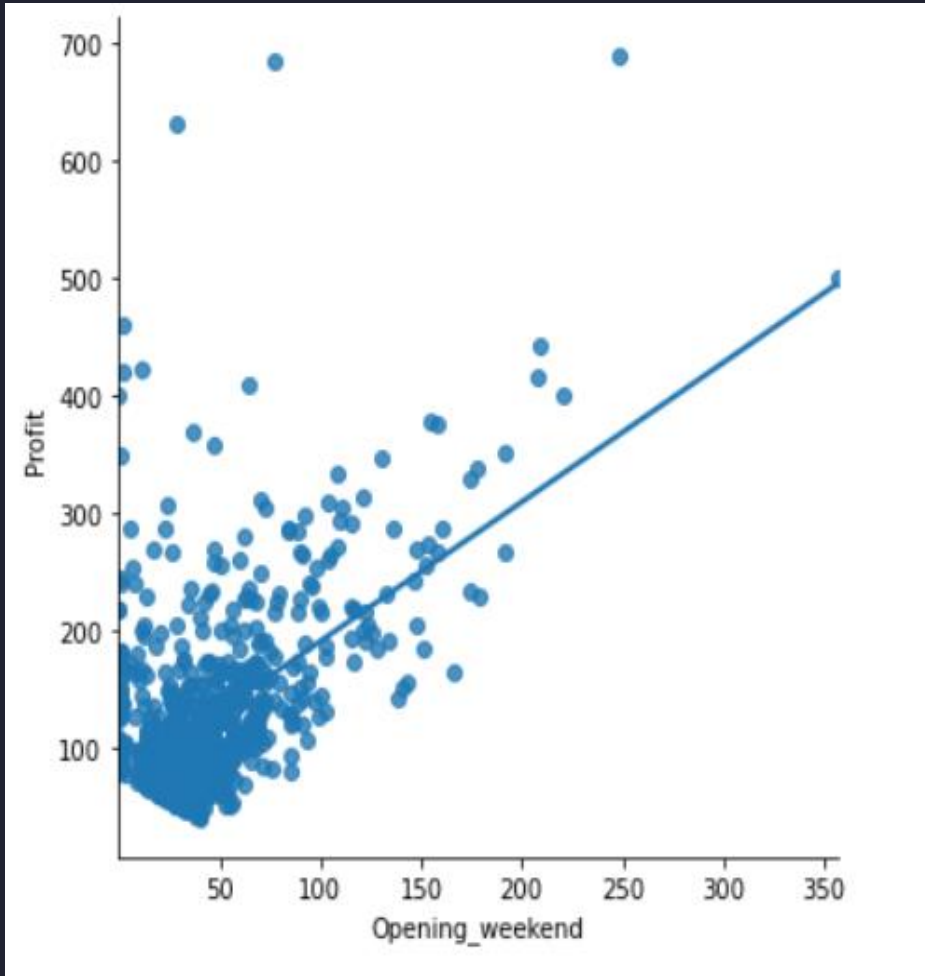
Ridge

0.47831



## Gross Domestic Revenue vs Budget

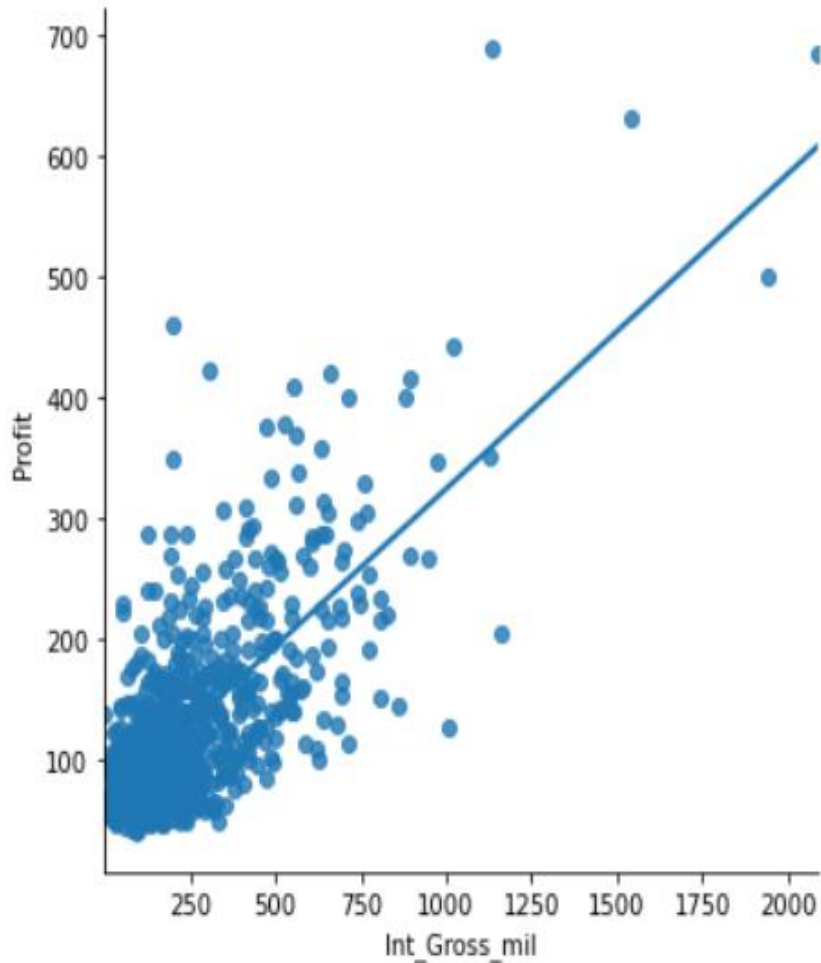
Correlation with Domestic Revenue:0.482156



## Gross Domestic Revenue vs Opening Weekend Collection

Correlation with Domestic Revenue:0.767446

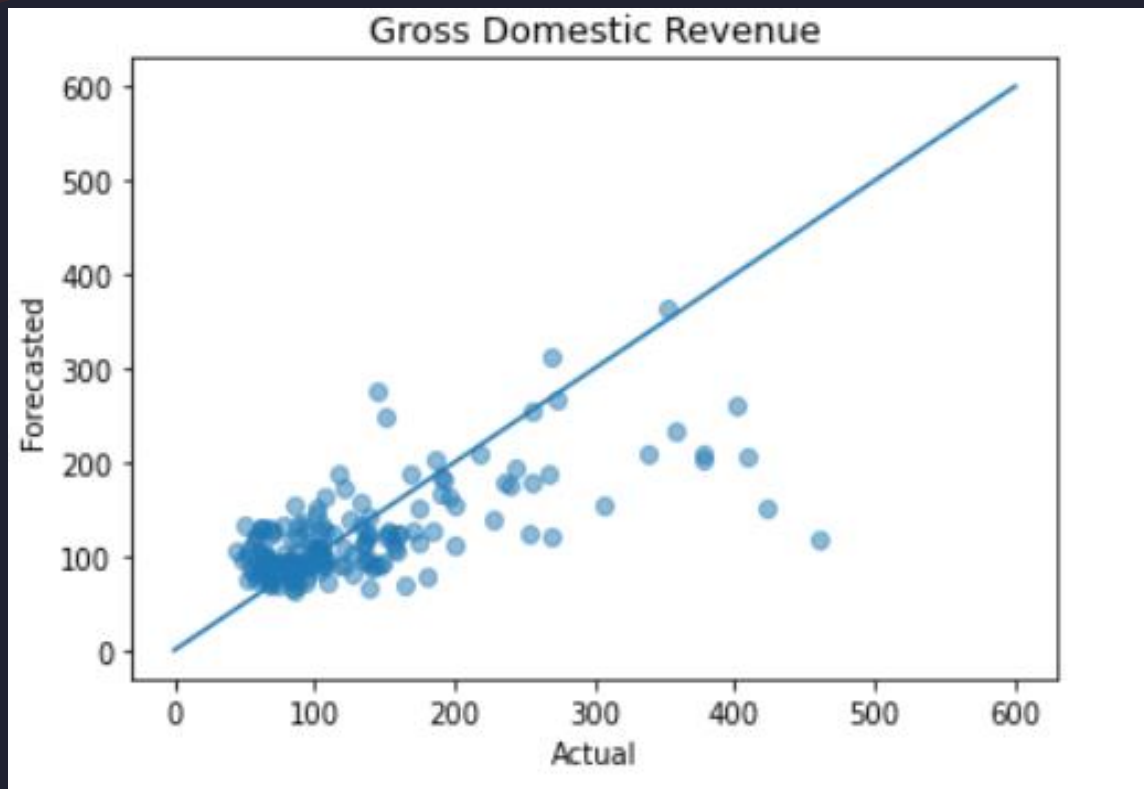




## Gross Domestic Revenue vs International Gross

Correlation with Domestic Revenue: 0.793032

## Results on Test Data: Budget, International Gross and Opening Weekend Collection



R-Squared = 0.61276

MAE = 32.1341

# Conclusion

- International Gross, Opening Weekend and Budget have a decent impact on the Domestic Revenues of movies
- Opening Weekend can give Bestflix Studios, a rough estimate of the future potential box office collection of the movie

## Future Work

- Actors
- Genre
- Marketing Initiatives(Influencers)
- Ratings



Thank You!

Do you have any questions?

[reynoldrdsilva@gmail.com](mailto:reynoldrdsilva@gmail.com)

+91-732-372-1303

[www.reynolddsilva.com](http://www.reynolddsilva.com)

