

# Bestflix and Chill



Linear Regression –  
Gross Domestic  
Revenue





BeautifulSoup



matplotlib



Data:



- Highest Grossing Movies of all time
- Sample of 1000 movies(1980-2021)
- Web-Scraped from Box Office Mojo and The Numbers

# Methodology

01

## Web Scrape

Box Office Mojo, The-Numbers



02

## Data Cleaning

Manipulating Data Set to fit the desired model



03

## Data Modeling

Linear Regression, Lasso and Ridge

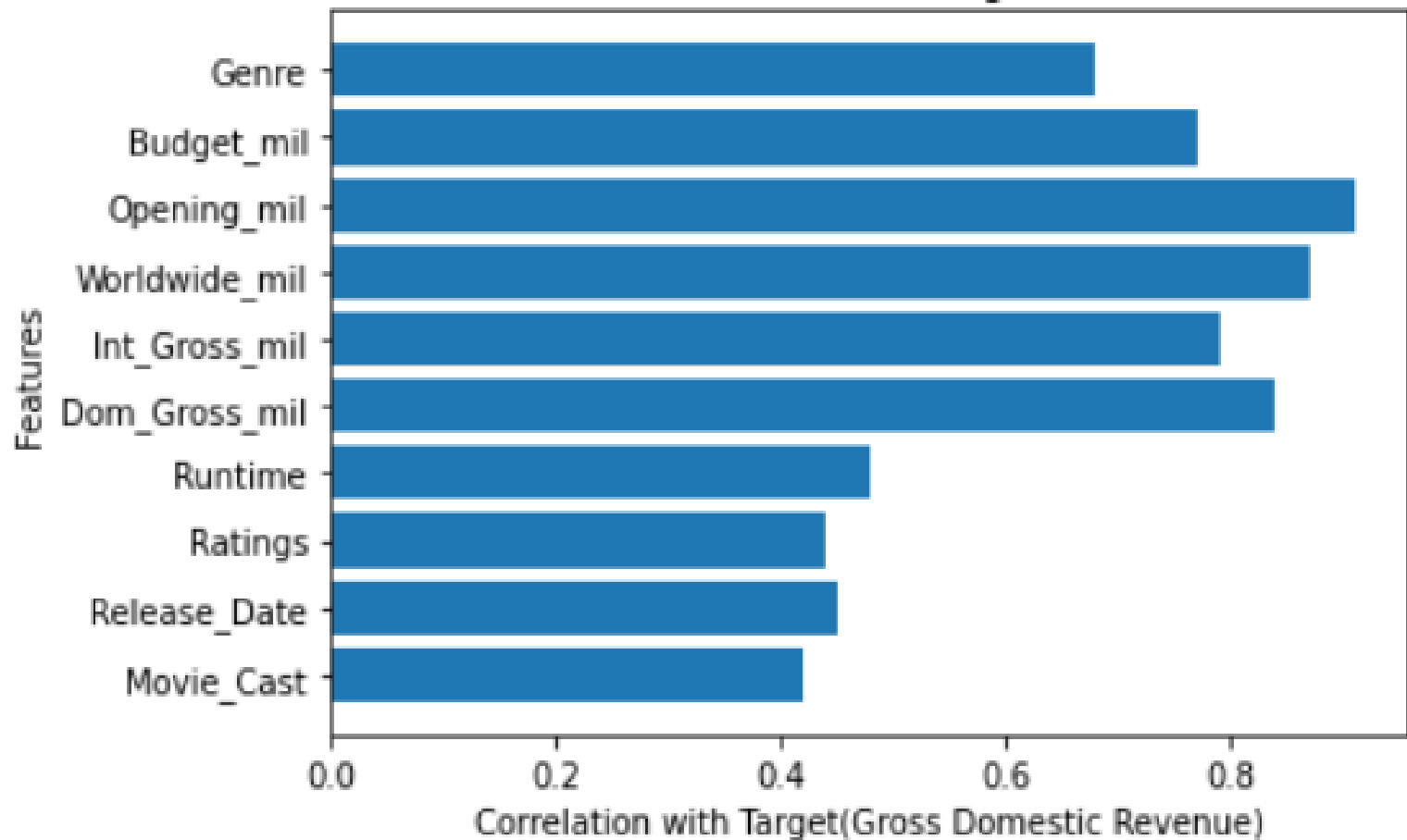
04

## Predictive Analysis

Gross Domestic Revenue (Domestic Revenue- Opening Weekend)



Features vs Target



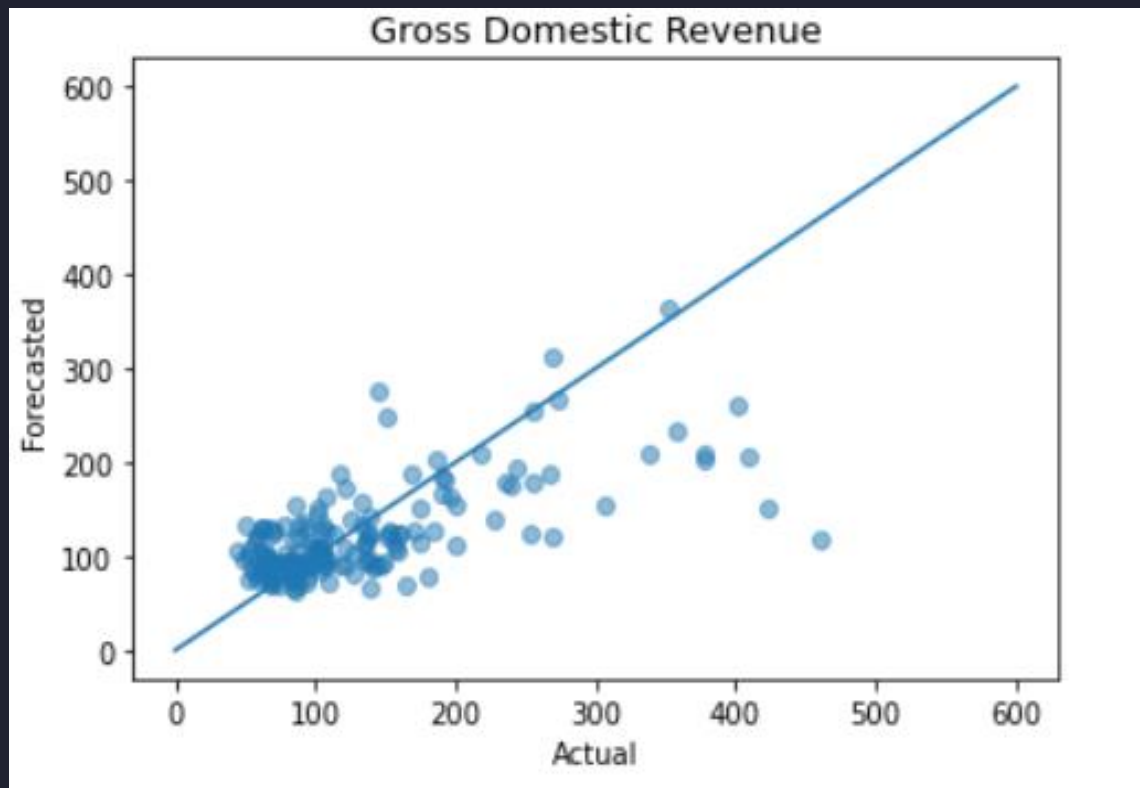
Budget  
Co-ef: -0.29

Gross  
Domestic  
Revenue

Opening  
Weekend  
Co-ef: 0.36

International  
Gross  
Co-ef: 0.27

# Final Results:



R-Squared = 0.61276  
MAE (Million)= \$ 32.13

## Findings: Model performs better on Lower Box Office Collections

Film	Actual	Predicted
Power Rangers (2017)	\$41,991,775	\$54,462,171
Alien vs. Predator (2004)	\$45,032,872	\$62,871,932
Star Wars (2015)	\$688,845,325	\$124,233,425
Avengers: Endgame (2019)	\$501,237, 993	\$69,832,517



## Future Work

- Popularity of Actors
- Marketing Initiatives(Influencers)
- Ratings



Thank You!  
Do you have any questions?  
[reynoldrdsilva@gmail.com](mailto:reynoldrdsilva@gmail.com)  
[www.reynoldddsilva.com](http://www.reynoldddsilva.com)

