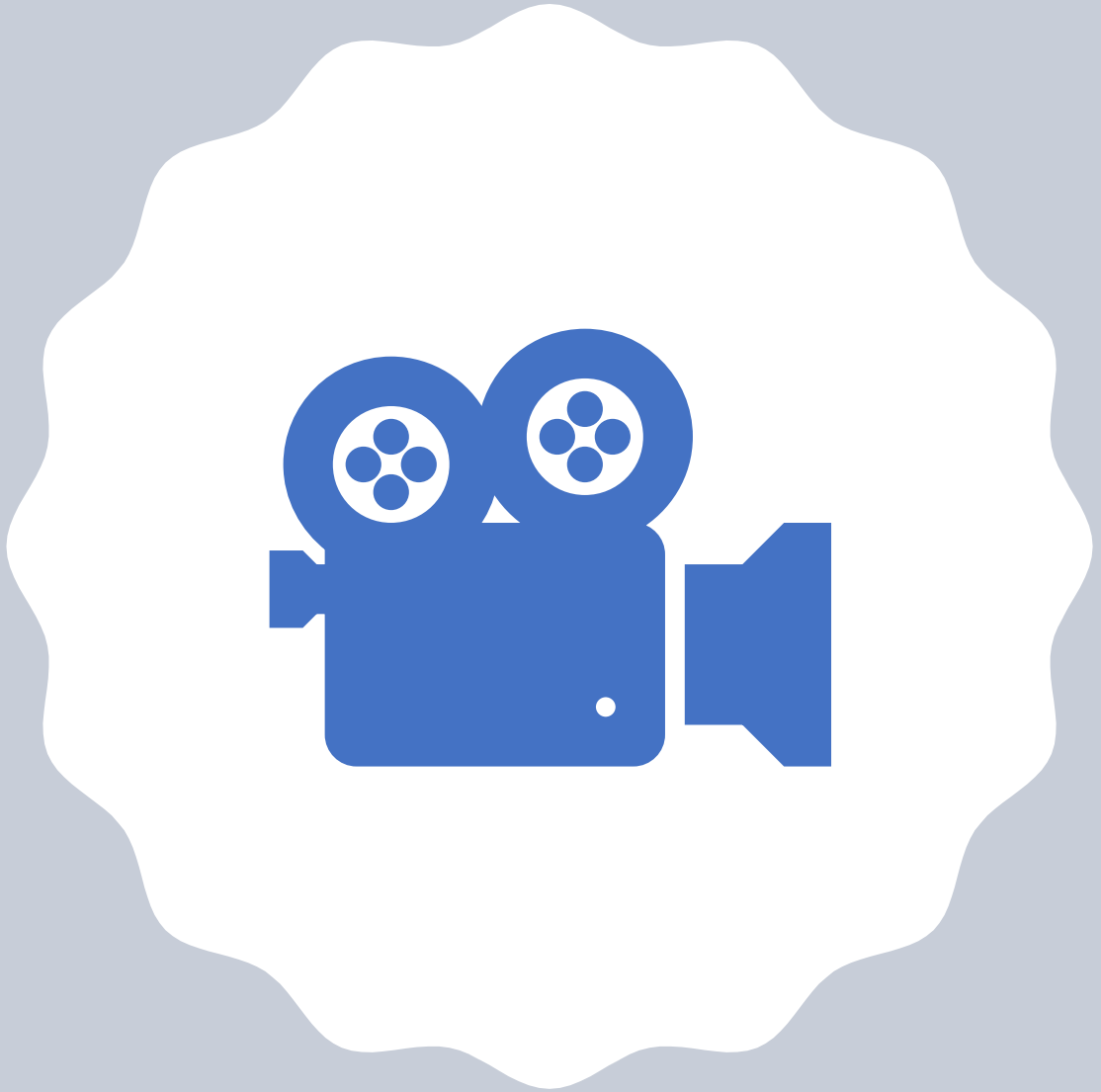
Movies Lifetime Gross Prediction



Overview

- Gross refers to all earnings of a film from all revenue sources, including box office
- The goal is to Create linear regression model that can predict the lifetime gross of a movie.

Project workflow

WEB SCRAPING & DATA GATHERING

EDA & FEATURE ENGINEERING

MODEL BUILDING & TRAINING

Data collection

Scraped box office mojo:

1. Title
2. Rank
3. Lifetime Gross
4. Overall Rank
5. Year

IMDb Dataset:

1. Title
2. Year
3. Genre
4. Duration
5. Country
6. language
7. Director
8. production company
9. Writer
10. Description
11. Actors
12. votes
13. Average vote
14. budget

EDA

1

Merge the two datasets.

2

Delete null values.

3

Delete duplicated rows and columns .

Feature engineering

1. Delete unwanted columns.
2. Convert categorical values into dummy variables
3. Spit the date into year, month and day.
4. Convert from object datatype to numeric.

The model

Features:

- Rank
- Overall Rank
- Year
- duration
- month
- day
- Average votes

Target value:

- Lifetime gross

Model Testing

R- Squared in :

- Baseline model

Training = 0.54,

Validation = 0.48

- log experiment (experiment one)

Training = 0.93,

Validation = 0.91

Model Testing

R-Squared in :

- Feature scaling (experiment two)

Training = 0.51,

Validation =0.53

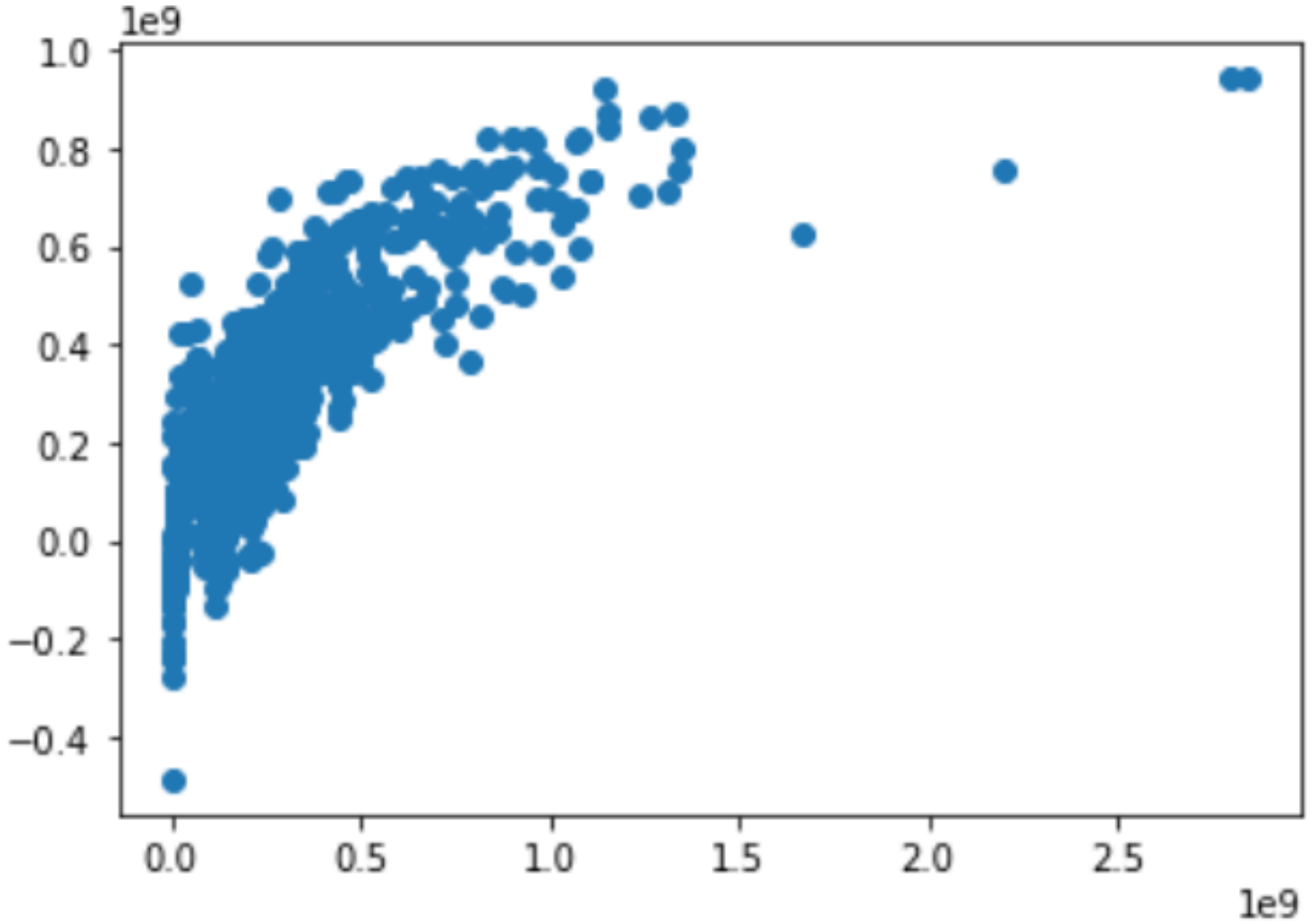
- Lasso model

Training =1,

Validation =0.99

Graph

Actual Vs Predicted



Observation:

Here the graph show the relationship between the features and our target values (Lifetime gross)

Thank You!