



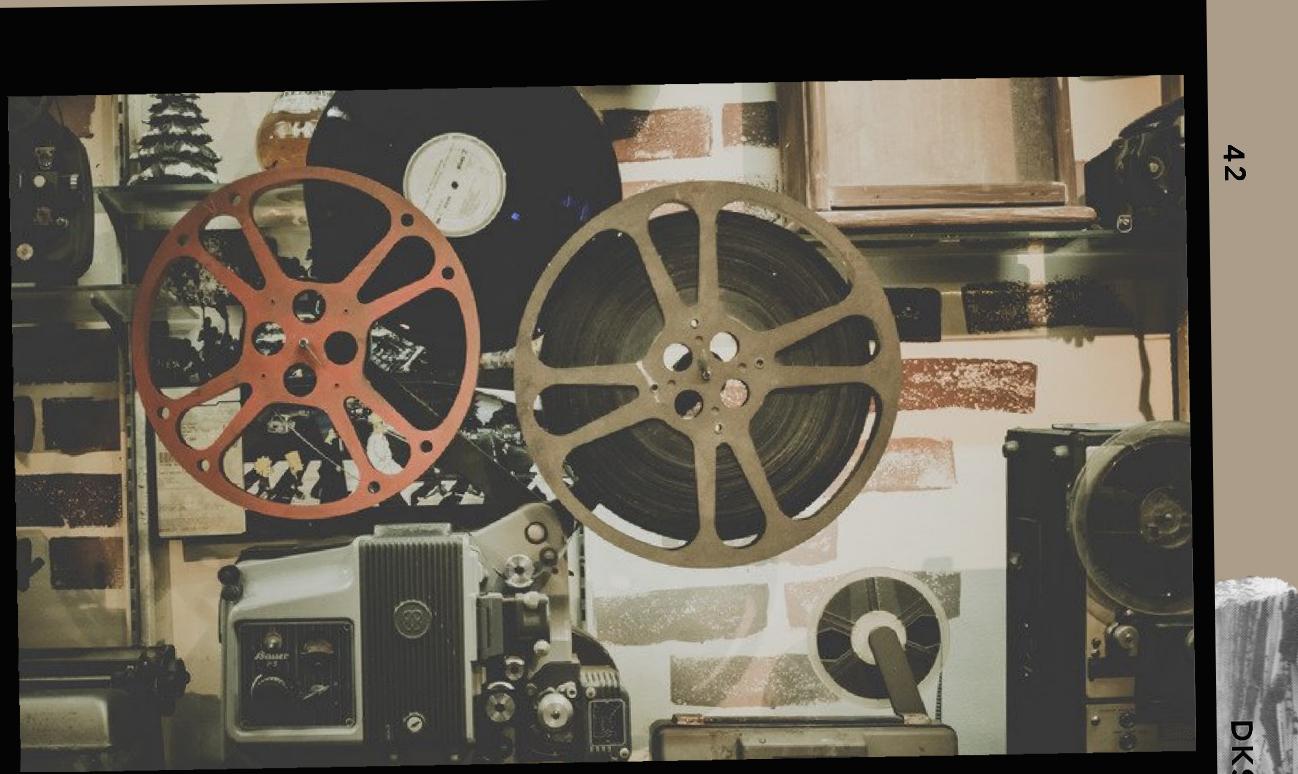
# Linear Regression With Movies Based On Novels.

# Outline:

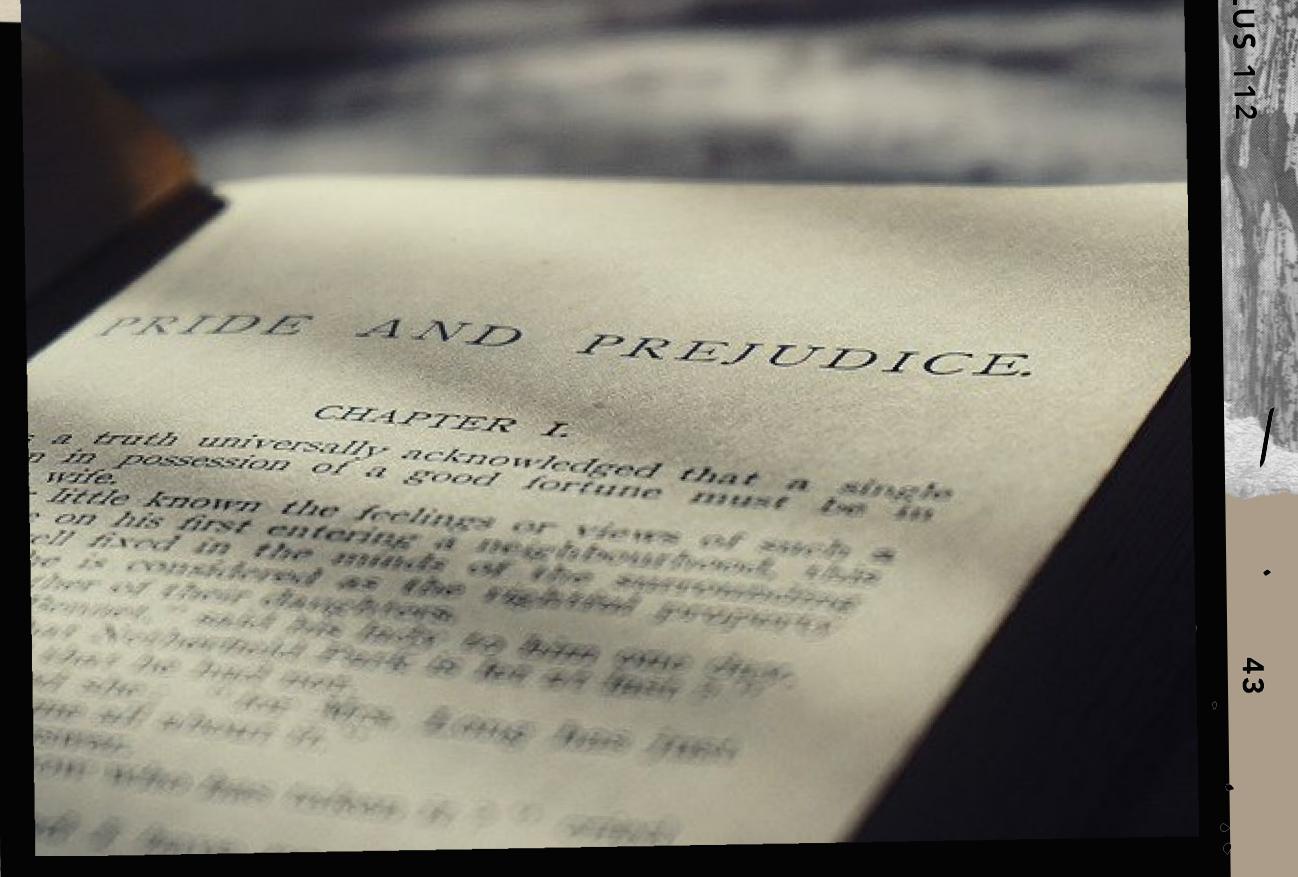
- Problem Statement and Motivation.
- Data Collection.
- Data Cleaning.
- Methodology.
- Result.
- Future Steps.

# Motivation

Create a Linear regression model that can predict World Wild Gross of movies based on novels by determining the features most influential to their success.



42



DKS PLUS 112

43

# Data Collection:

We scraping three web site to collect our data by merge the flowing :

## 1- IMDB website:

[Https://www.imdb.com/search/keyword/?keywords=based-on\\_novel](https://www.imdb.com/search/keyword/?keywords=based-on_novel)

Title, Rating, Runtime, IMDB scores, Genre.

Total of 3450 dataset.

## 2-BoxOfficeMojo website.

<https://www.boxofficemojo.com>

The theaters, Total gross.

Total of 999 dataset.

## 3- The Numbers website.

<https://www.the-numbers.com>

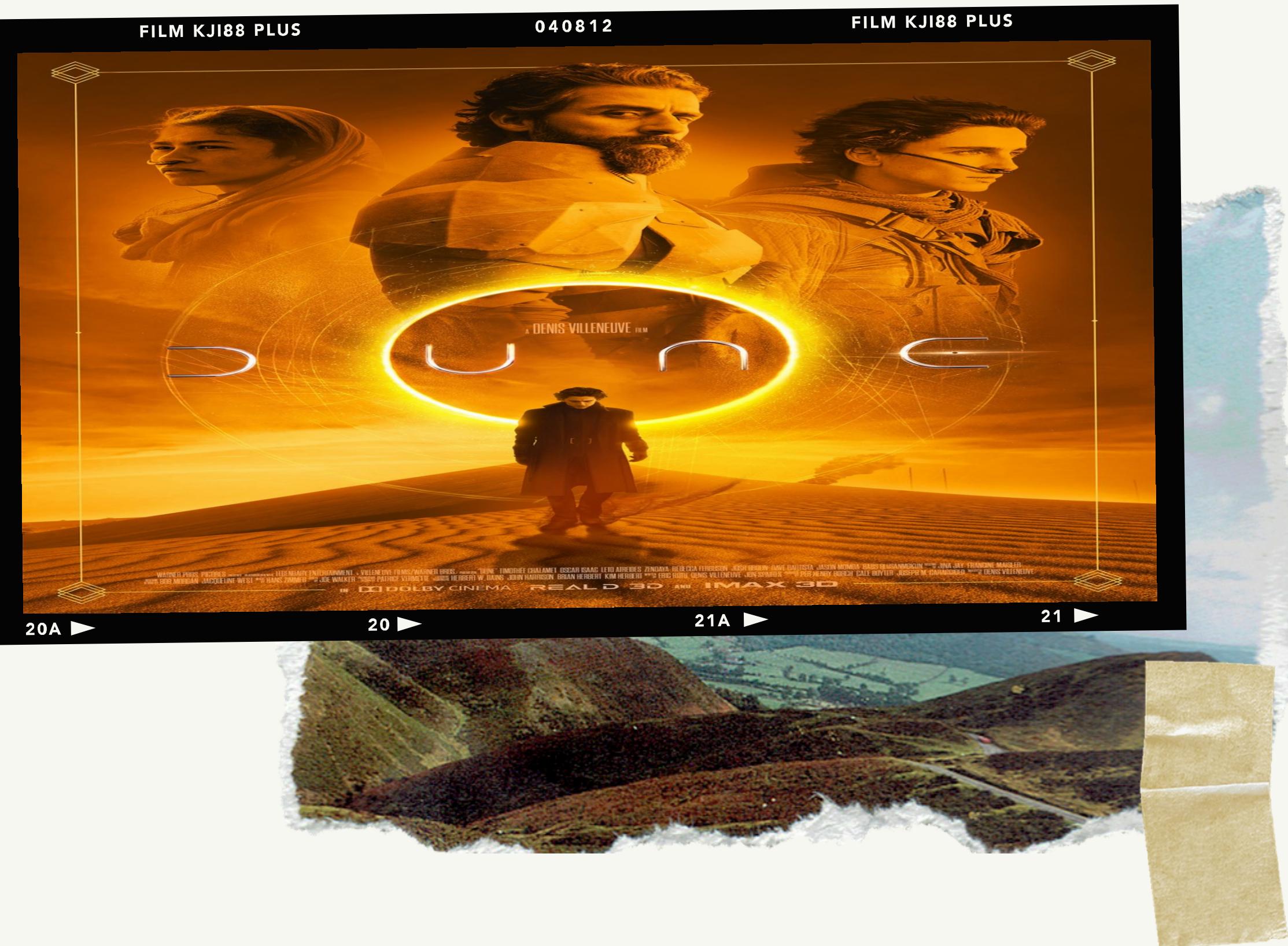
Title, Budget, Worldwide gross.

Total of 6100 dataset.

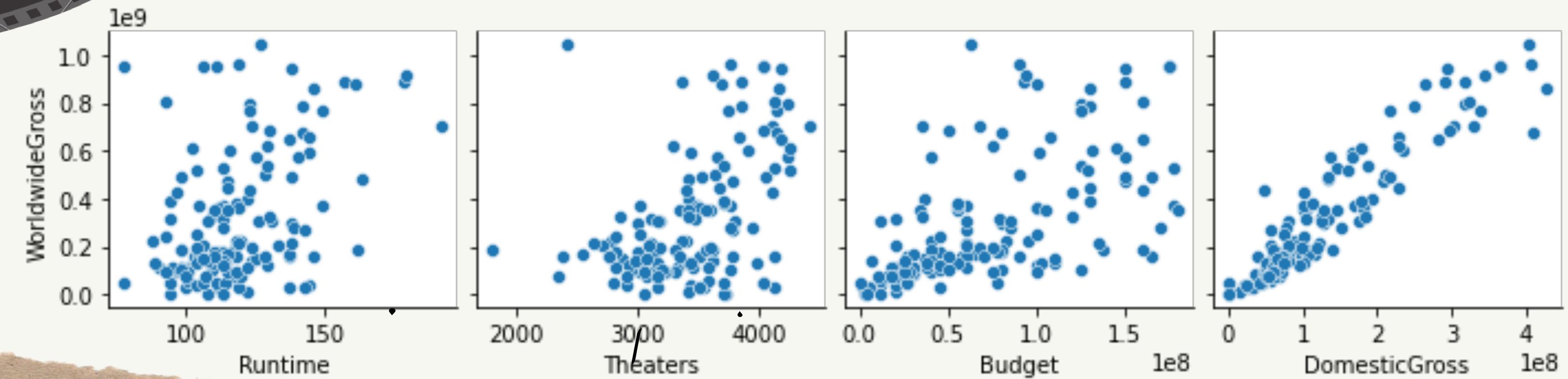


# Data clean

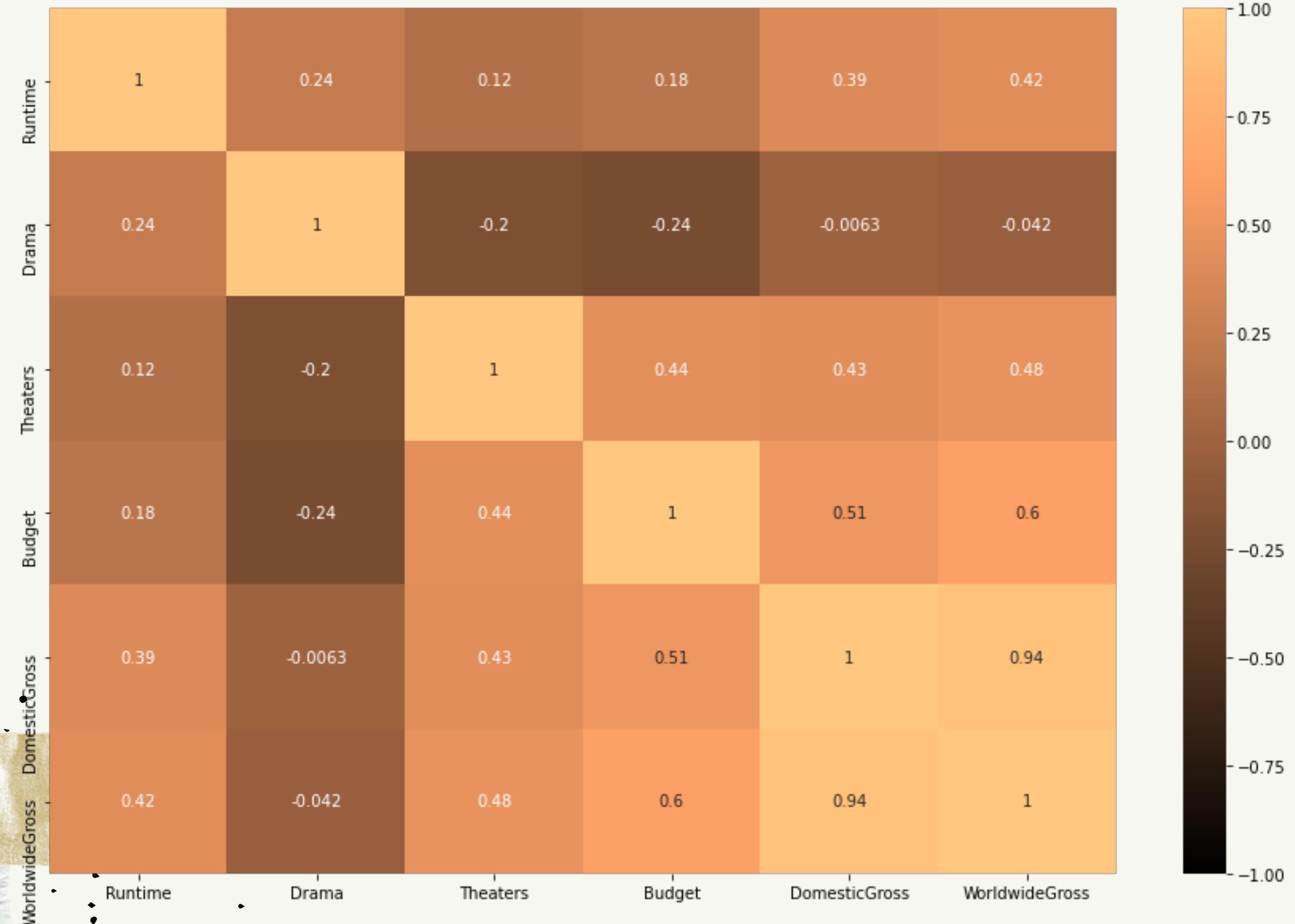
1. Handling miss value
2. Remove duplicated value and nan value
3. Convert formatting for runtime ,date resident features



## Features pair plot



### Correlation Heatmap

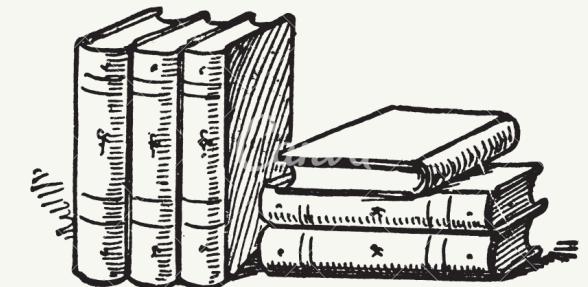




# Methodology

- split data train ,test, validation
- Liner regression
- LASSO Regularization
- Feature engineering

Create new variable "Primetime", indicating if a movie was released during a typically successful time of year (summer or holiday season).



# Result

X= Rating, Runtime, IMDB\_score, Budget, Theaters, Primetime.

Y= Worldwide Gross.

the metrics	Score	R <sup>2</sup>
Linear Regression	0.54	0.31
Lasso	0.54	0.31

the movie was unclear. I had signs of the island being brought back to life anyway, I had

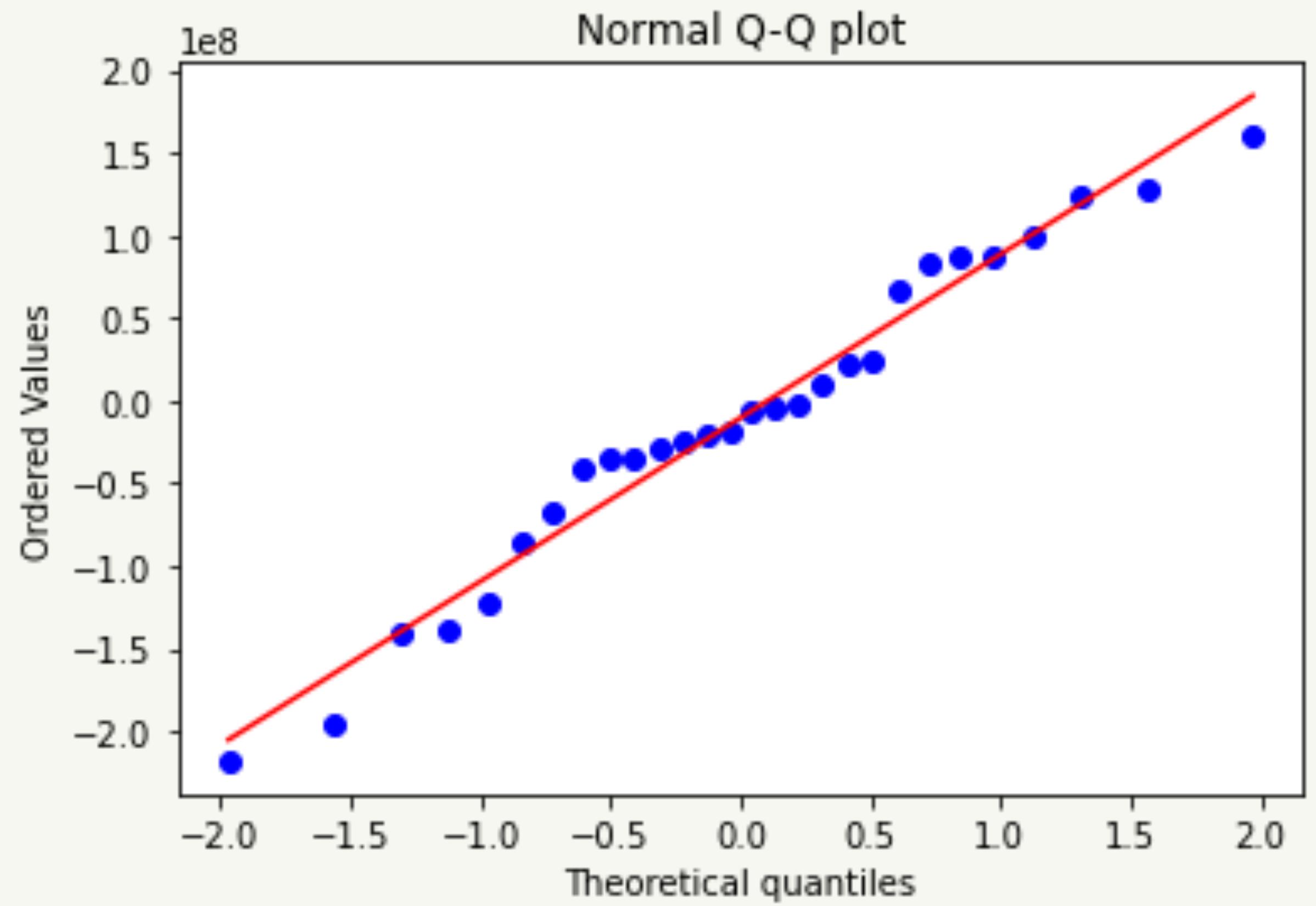
# Result

X= Budget, Primetime, Theaters, Domestic Gross.

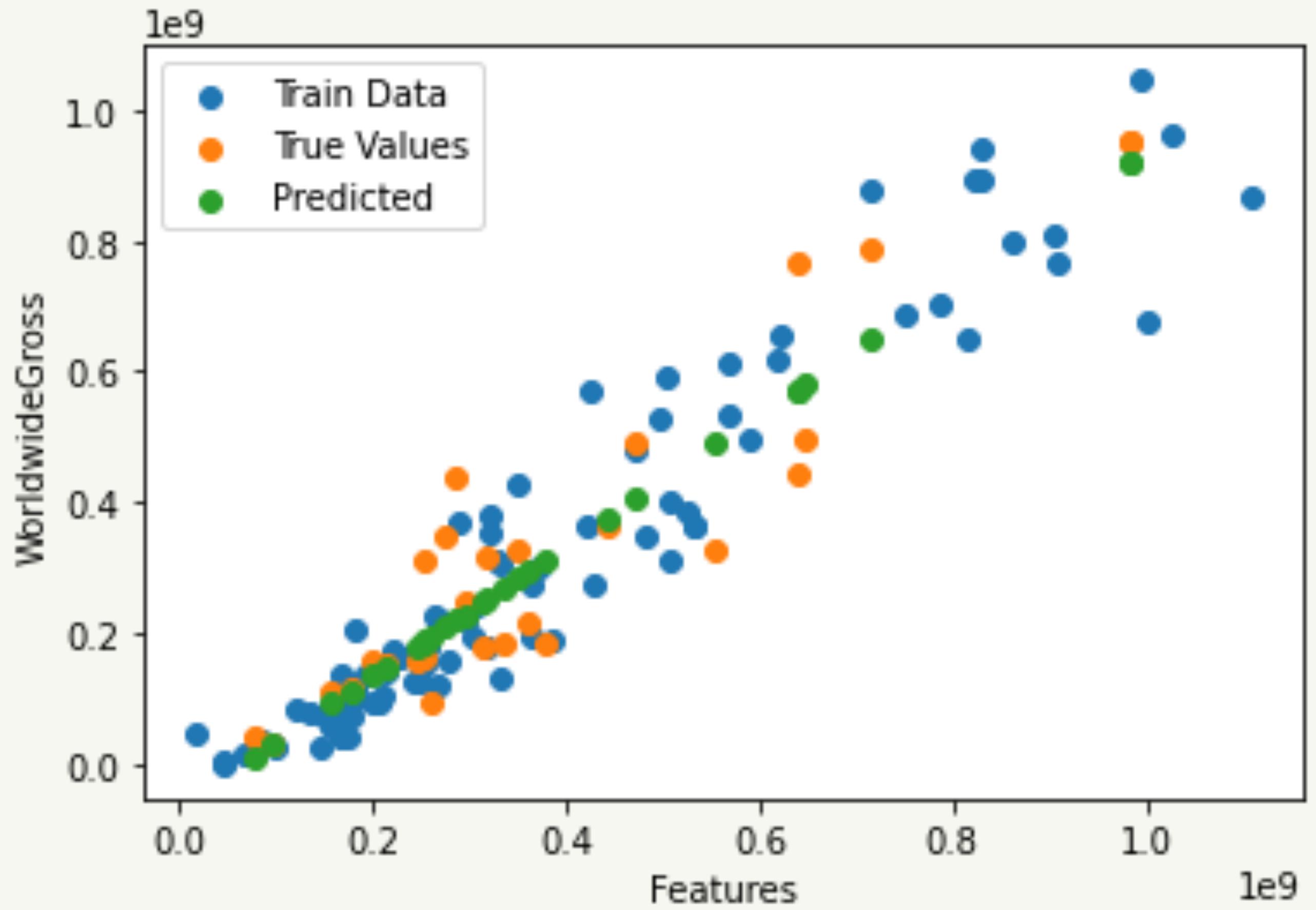
Y= Worldwide Gross.

the metrics	Score	R <sup>2</sup>
Linear Regression	0.92	0.85
Lasso	0.92	0.85

It was unclear if there were signs of the island being brought back to life anyway, I had

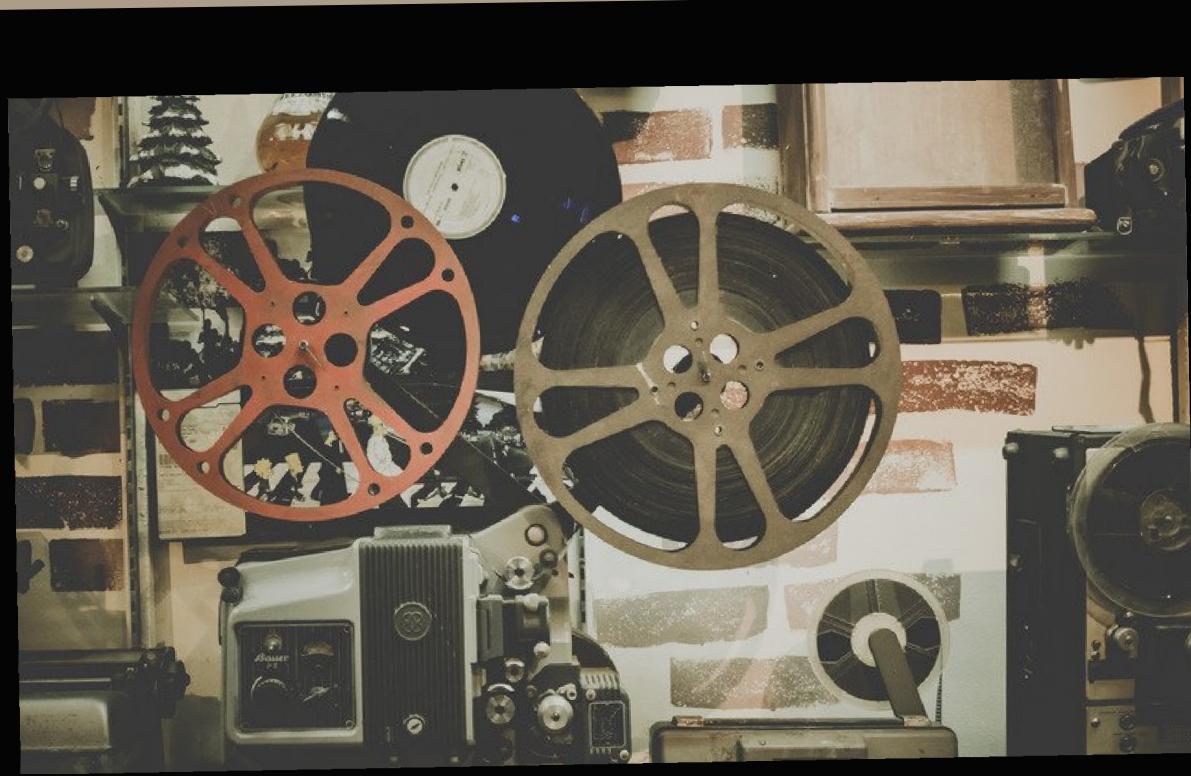


... was unclear. I had signs of the island being brought back to life anyway, I had



... was unclear. I had signs of the island being brought back to life anyway, I had

unclear. I never  
to life anyway, I had

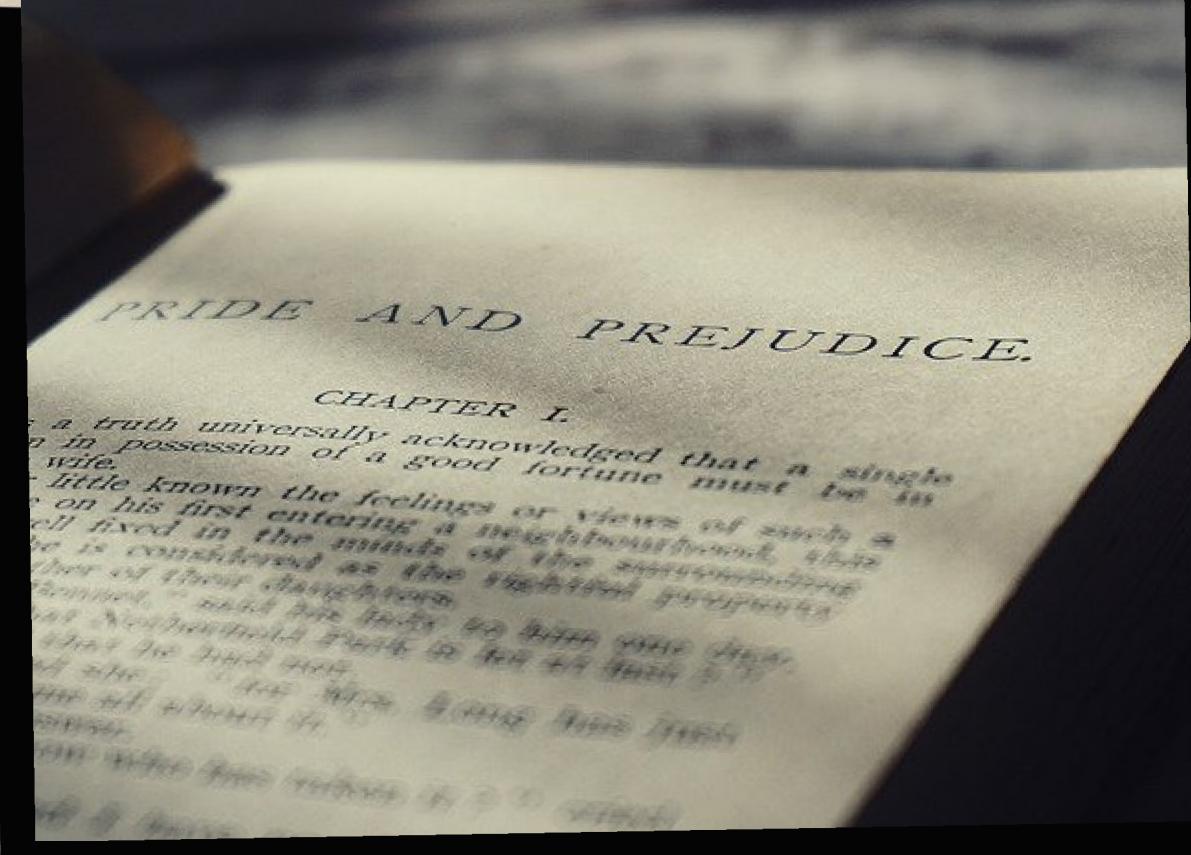


42

## Future steps

We will improve our prediction model by add more new features

Define different movie's category



DKS PLUS 112

43

1

# THANK YOU

by:

Rana Alqahtani  
Ohoud Albabtain  
Khulud Amer

