Movieland Production Company

Analysis to predict Gross revenue of a movie

Data from www.boxofficemojo.com
Top 200 ranked domestic movies from years 2015 to 2022

Sample page

Features for Target Variable: Gross

Extracted features

- Genres
- MPAA rating
- Duration
- Number of theatres
- Number of days released
- Domestic release ratio
- Distributor
- Release date
- Opening week gross
- Movie title

Engineered features

- 5 Genres or less
- PG-13
- PG
- R
- Duration
- Number of theatres
- Number of days released
- Domestic release ratio
- Warner Bros.
- Other distributors
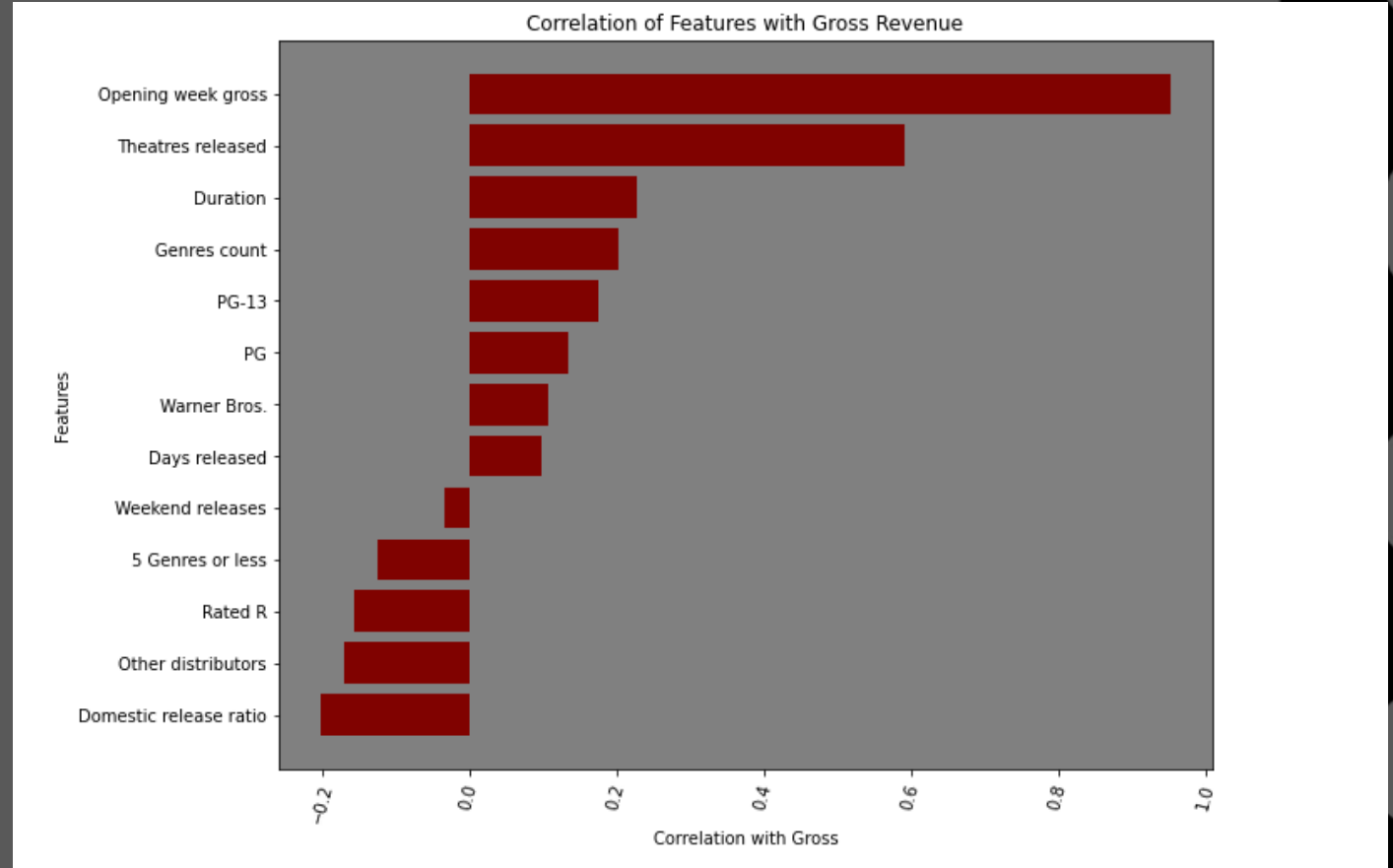- Weekend release
- Opening

Correlation between features and Gross. revenue

Positively Correlated (top 3)

- Opening
- Theaters released
- Duration
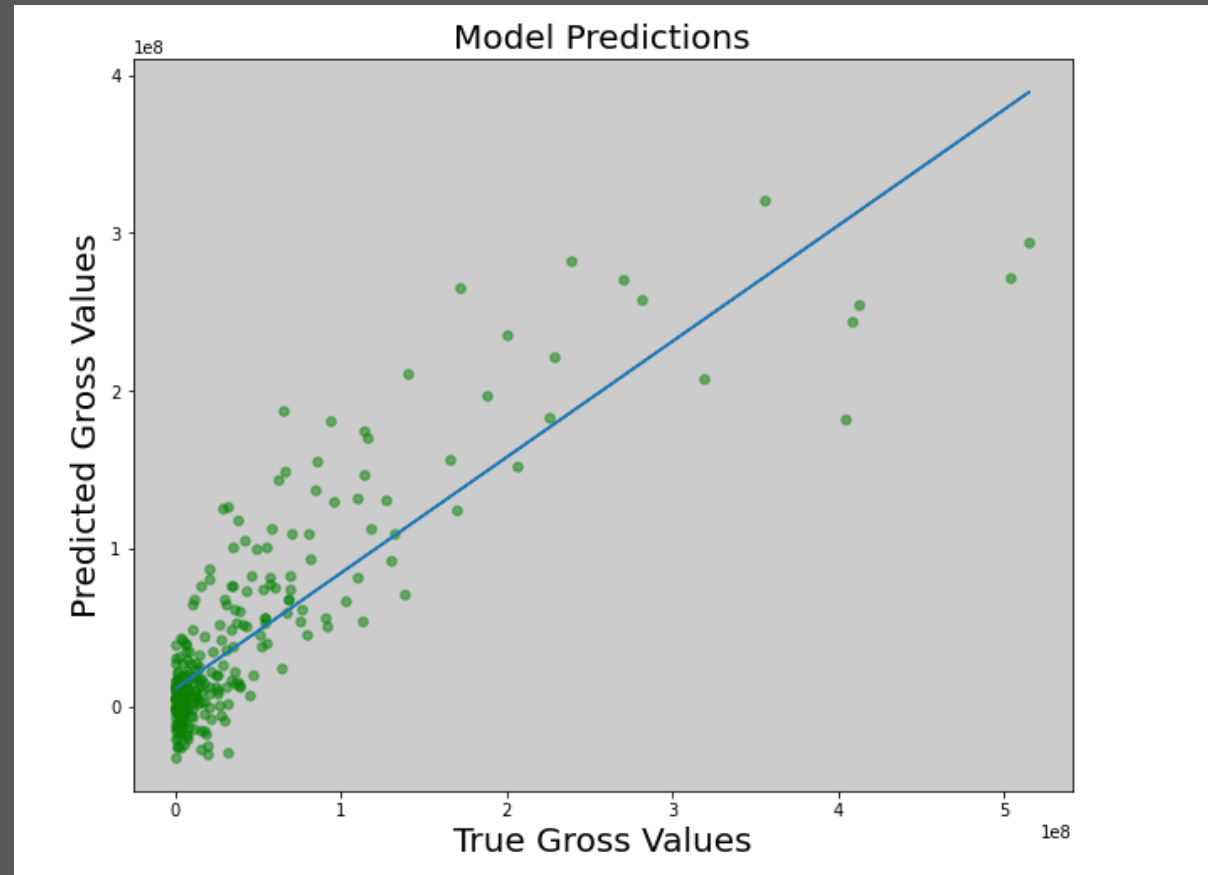
Negatively correlated (bottom 3)

- Domestic release ratio
- Other distributors
- R Rated Movies

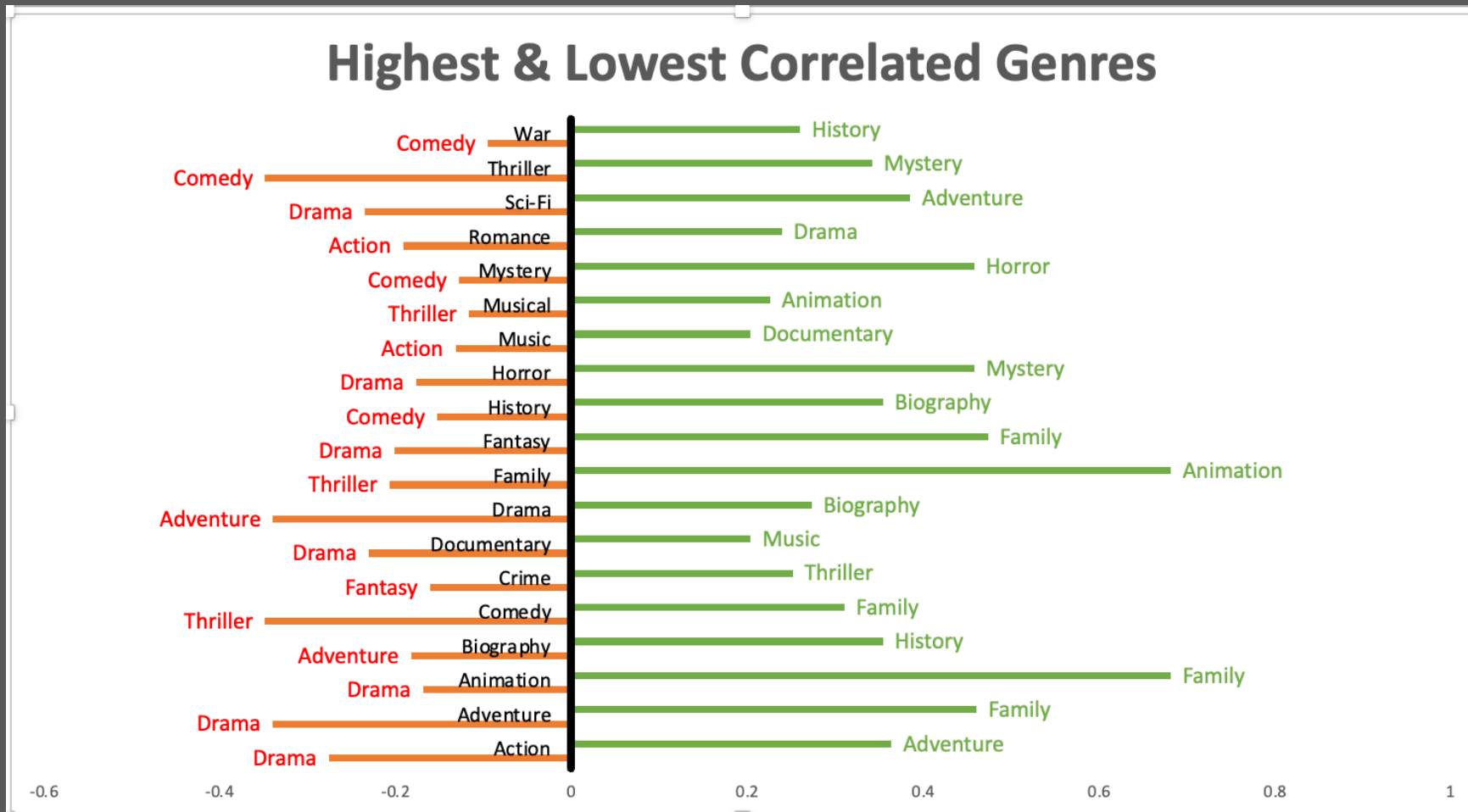- * Dropped Opening week gross as it is a subset of Gross revenue and that is why the high corr.

Correlation of Features with Gross Revenue
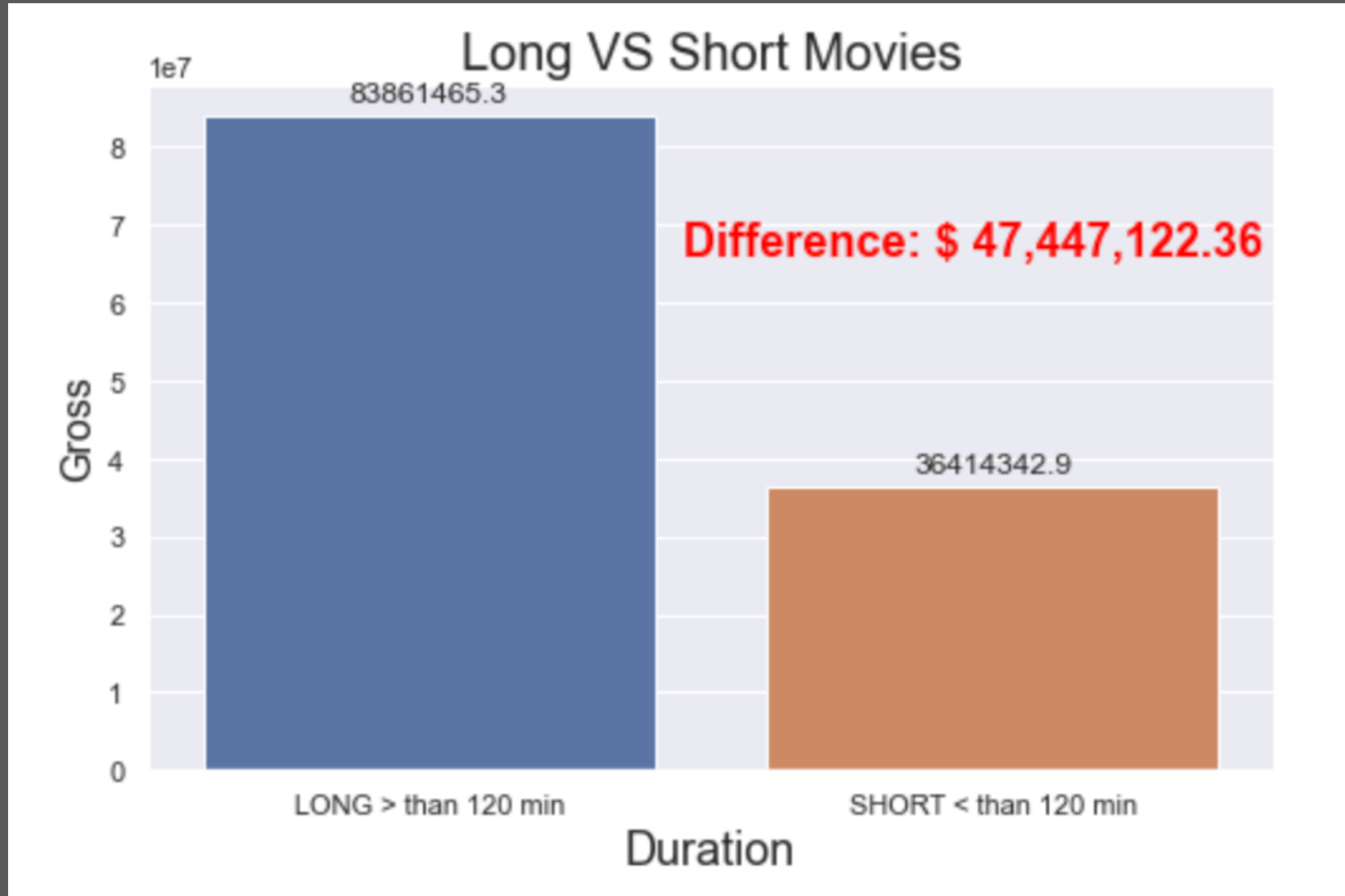
Model Predictions

Model Performance

- Linear Regression with Polynomial transformation.

- R Square value of 0.73

- Predictions on 'Test data'

Interesting Insights

Highest & Lowest Correlated Genres

Interesting Insights

Future Work

- Incorporate Budget and do Profit analysis
- Movie recommendation based on Genre and other features.
- Add more features like IMDB Rating, Directors, Lead actors and Cast.

Features Explained

- 5 Genres or less: Each movie is associated with 1 or more genre. Some have up to 9 genres. So made a column that will display binary values of 1 or 0 based on the number of genres. If 5 or less, then is has a value of 1.

- Domestic release ration: Proportion of released made with the country Vs made outside.

- Release date: is converted to know if it was release on a weekday or weekend and assigned binary values accordingly. This feature is now called weekend release.

Questions