


The background of the slide features a dark, moody aesthetic with two large, metallic film reels positioned in the upper half. A strip of film, showing individual frames, is draped across the scene, looping around the reels and extending towards the bottom right. The lighting is soft, highlighting the textures of the reels and the film strip.

# 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

 <h3>The Batman</h3> <p>When the Riddler, a sadistic serial killer, begins murdering key political figures in Gotham, Batman is forced to investigate the city's hidden corruption and question his family's involvement.</p>		
<div><div>Title Summary</div><div>Original Release ▾</div><div>Domestic ▾</div></div>		
<b>Grosses</b> DOMESTIC (50.2%) <b>\$302,863,177</b> INTERNATIONAL (49.8%) <b>\$300,400,000</b> WORLDWIDE <b>\$603,263,177</b>	Distributor	Warner Bros. <a href="#">See full company information</a>
	Opening	\$134,008,624 4,417 theaters
	Release Date	Mar 4, 2022
	MPAA	PG-13
	Running Time	2 hr 56 min
	Genres	Action Crime Drama Mystery
	In Release	18 days/2 weeks
	Widest Release	4,417 theaters

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

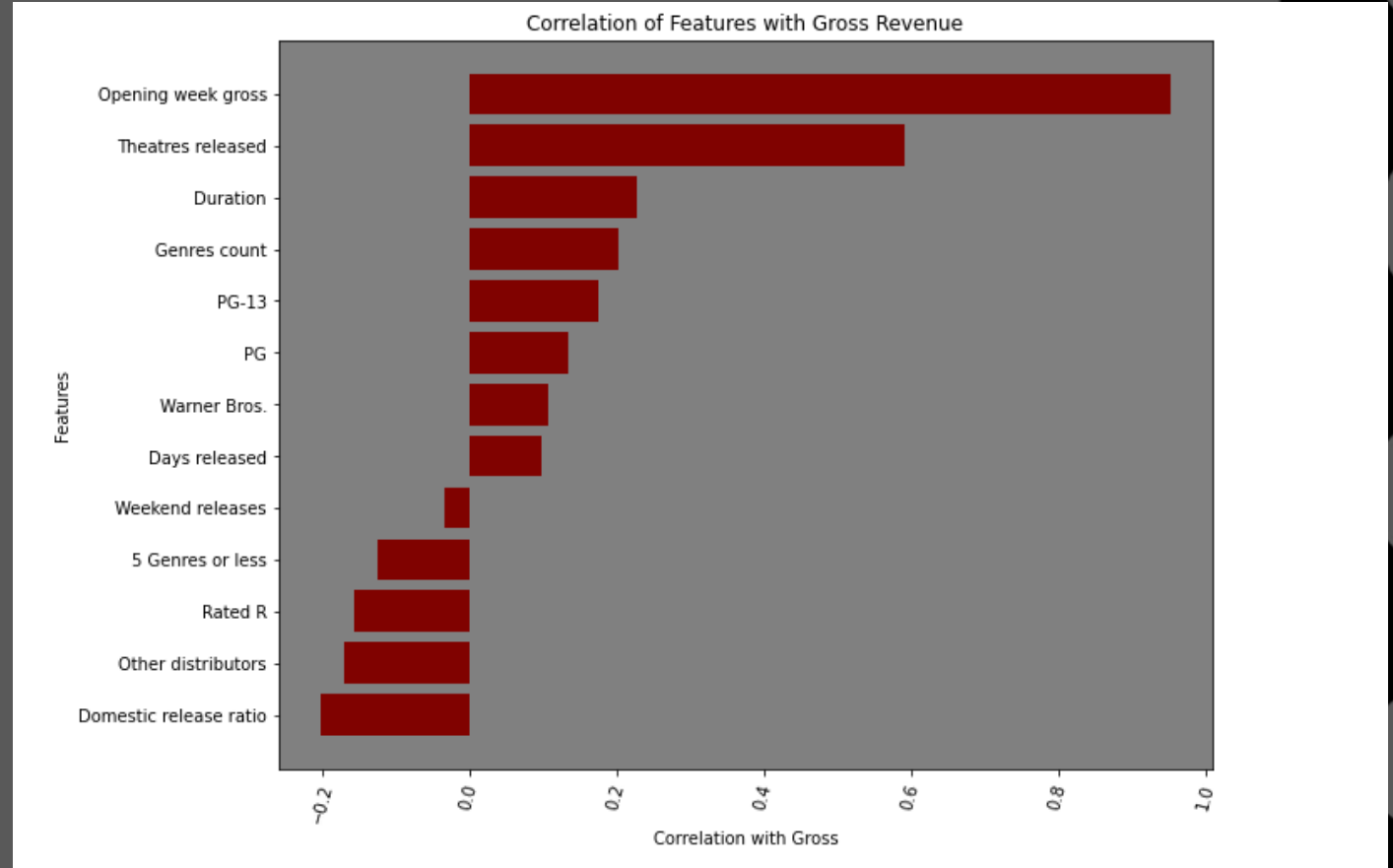
# Correlation between features and Gross. revenue

## Positively Correlated (top 3)

- Opening
- Theaters released
- Duration

## Negatively correlated (bottom 3)

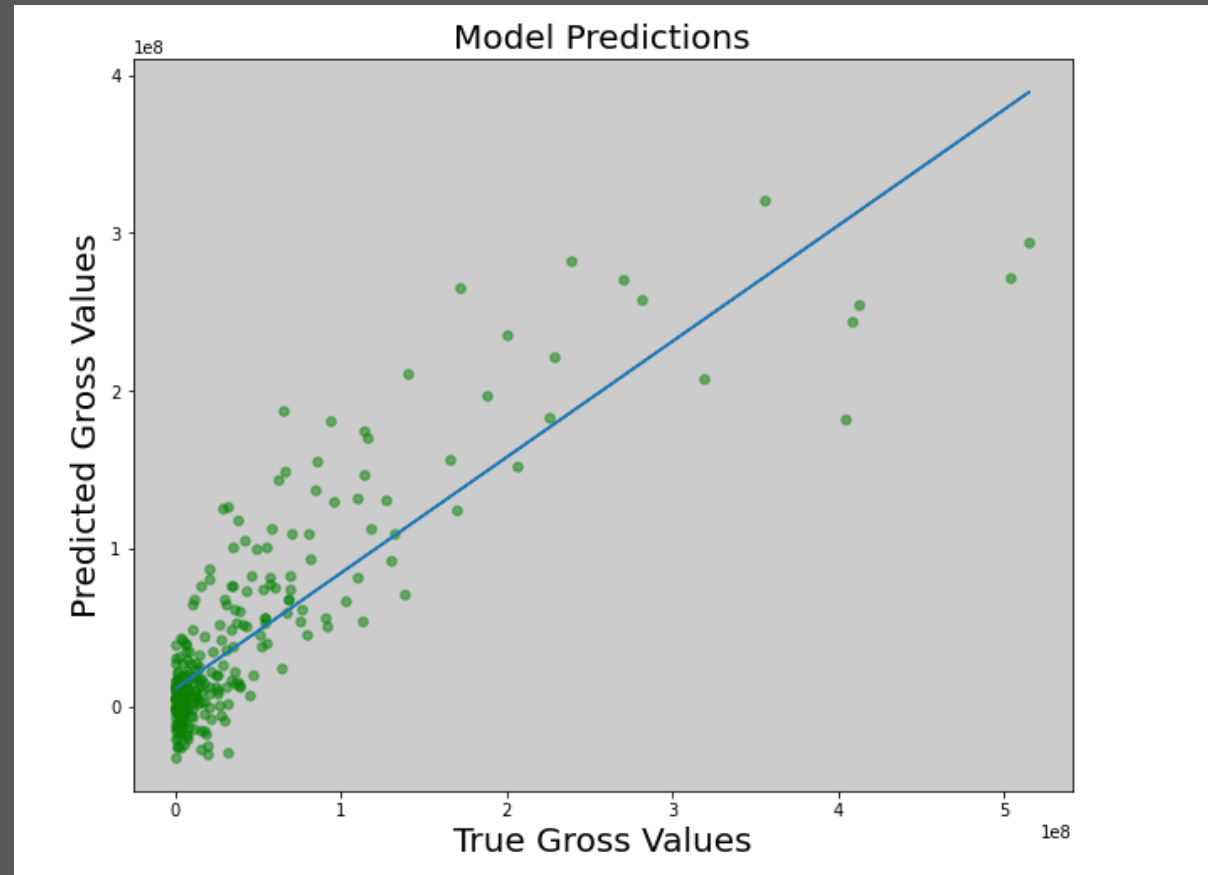
- Domestic release ratio
  - Other distributors
  - R Rated Movies
- \* Dropped Opening week gross as it is a subset of Gross revenue and that is why the high corr.



# Model Predictions

## Model Performance

- Linear Regression with Polynomial transformation.
- R Square value of 0.73
- Predictions on 'Test data'



# Model Prediction Samples

## Unhinged

Action and Thriller

- Actual Gross:  
\$20,831,465
- Predicted Gross:  
\$18,578,820
- 10% lower

## Horrible Bosses 2

Comedy and Crime

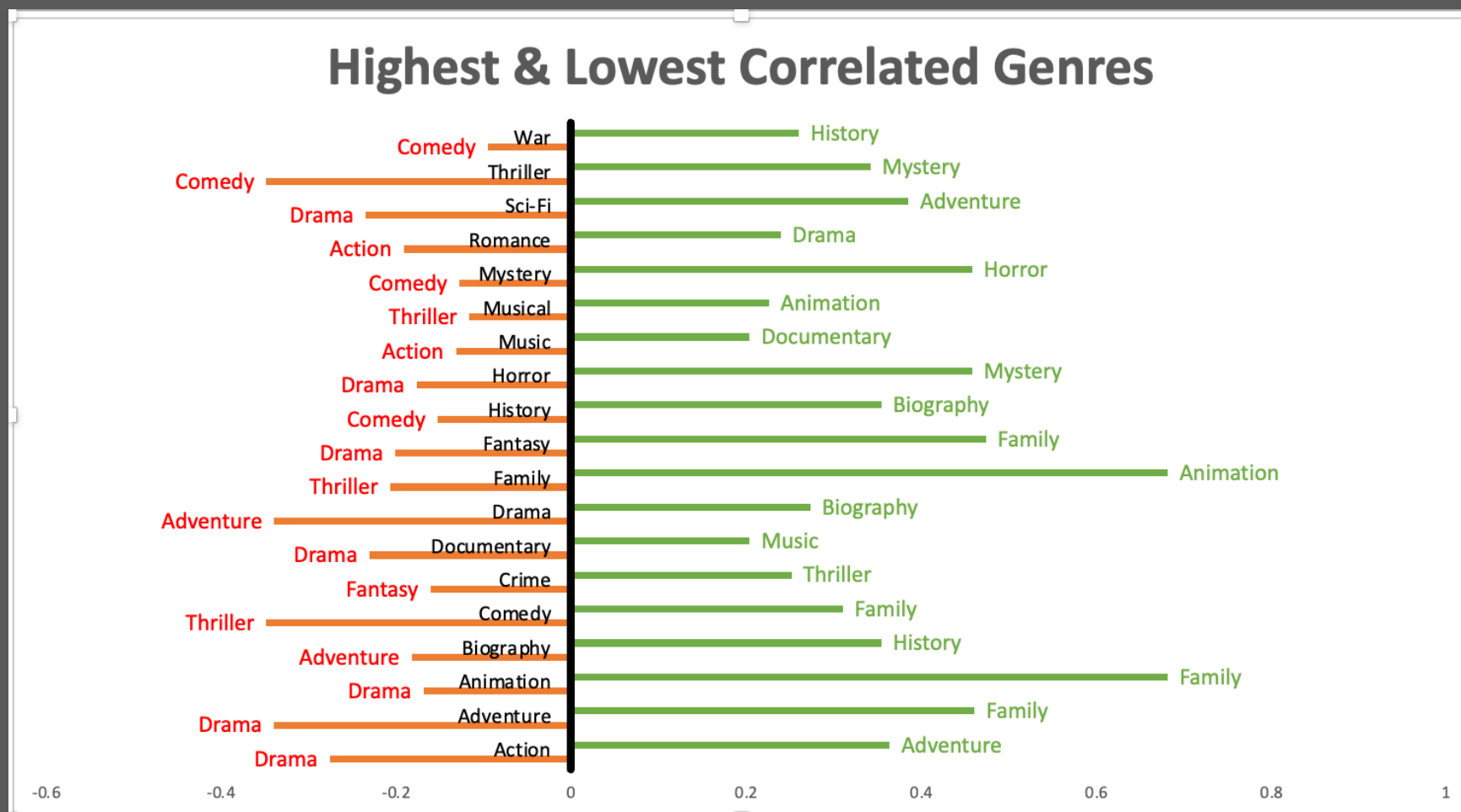
- Actual Gross:  
\$54,445,357
- Predicted Gross:  
\$56,233,540
- 3.2% higher

## Phantom Thread

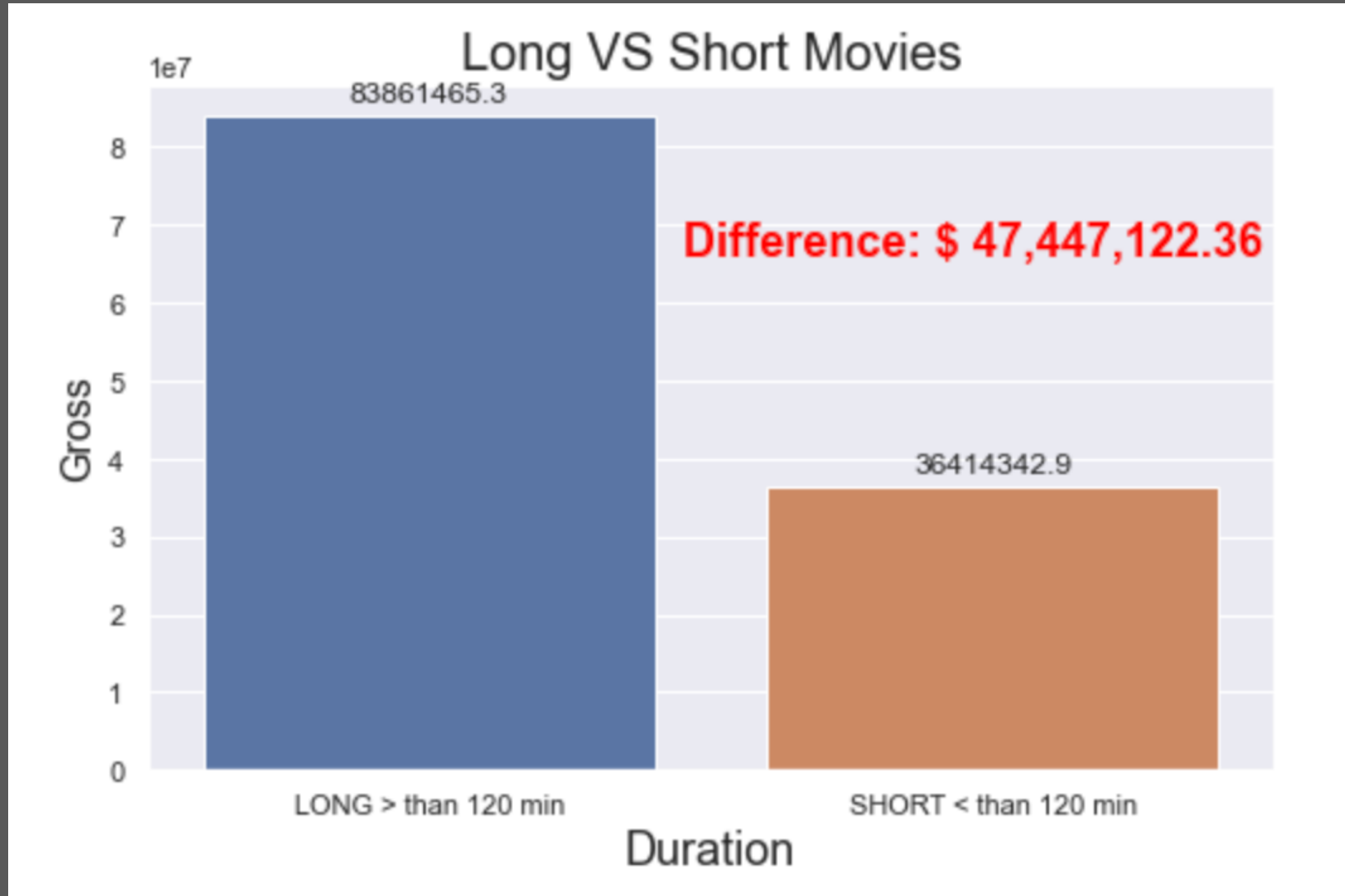
Drama and Romance

- Actual Gross:  
\$21,198,205
- Predicted Gross:  
\$19,187,860
- 9.4% lower

# Interesting Insights



# 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 it has a value of 1.
- Domestic release ratio: Proportion of releases made within the country Vs made outside the country.
- Release date: is converted to know if it was released on a weekday or weekend and assigned binary values accordingly. This feature is now called weekend release.

A photograph of a theater interior. The foreground shows rows of red upholstered seats. In the background, a stage is visible, covered by a large, vertically pleated golden curtain. A dark, semi-transparent horizontal band is overlaid across the middle of the image, containing the word "Questions" in white text.

Questions