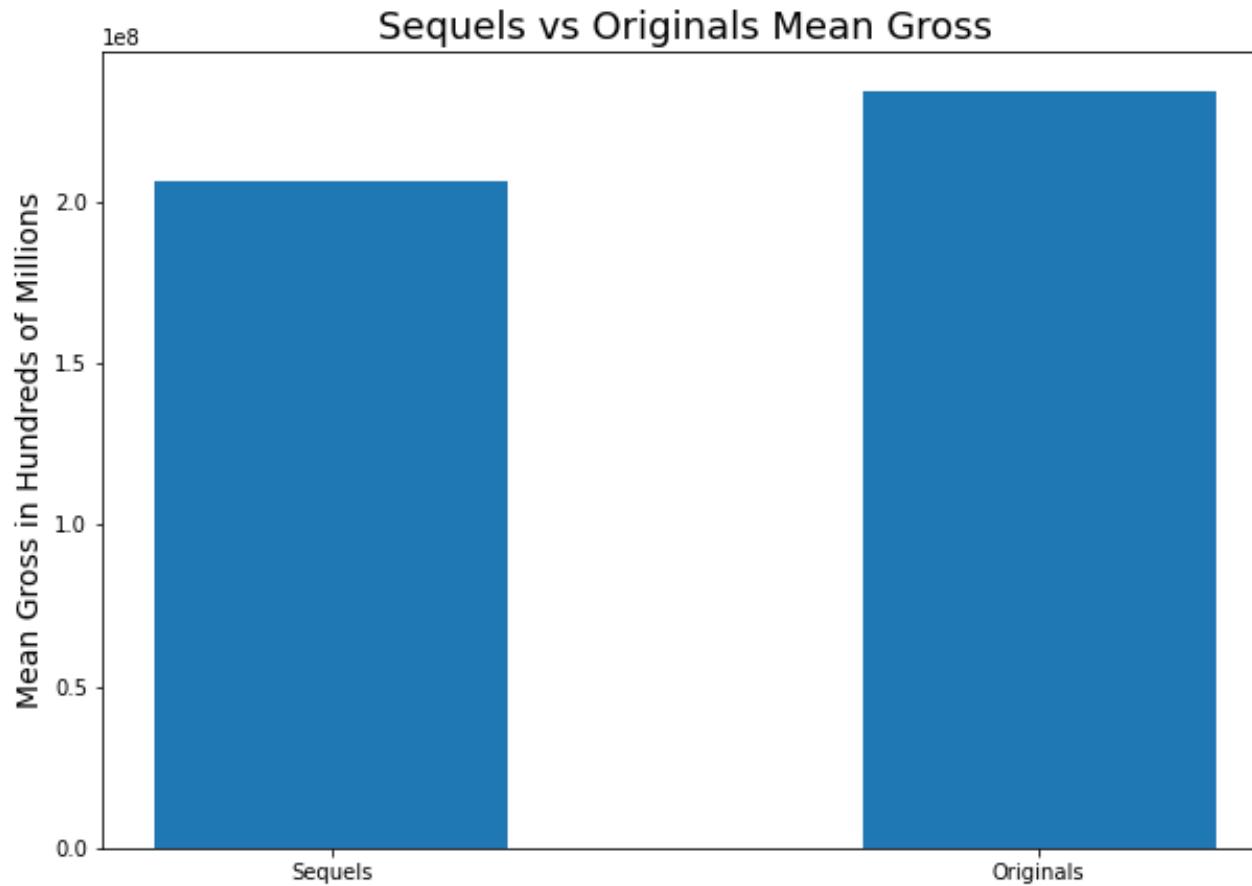


Predicting Sequel Worldwide Gross

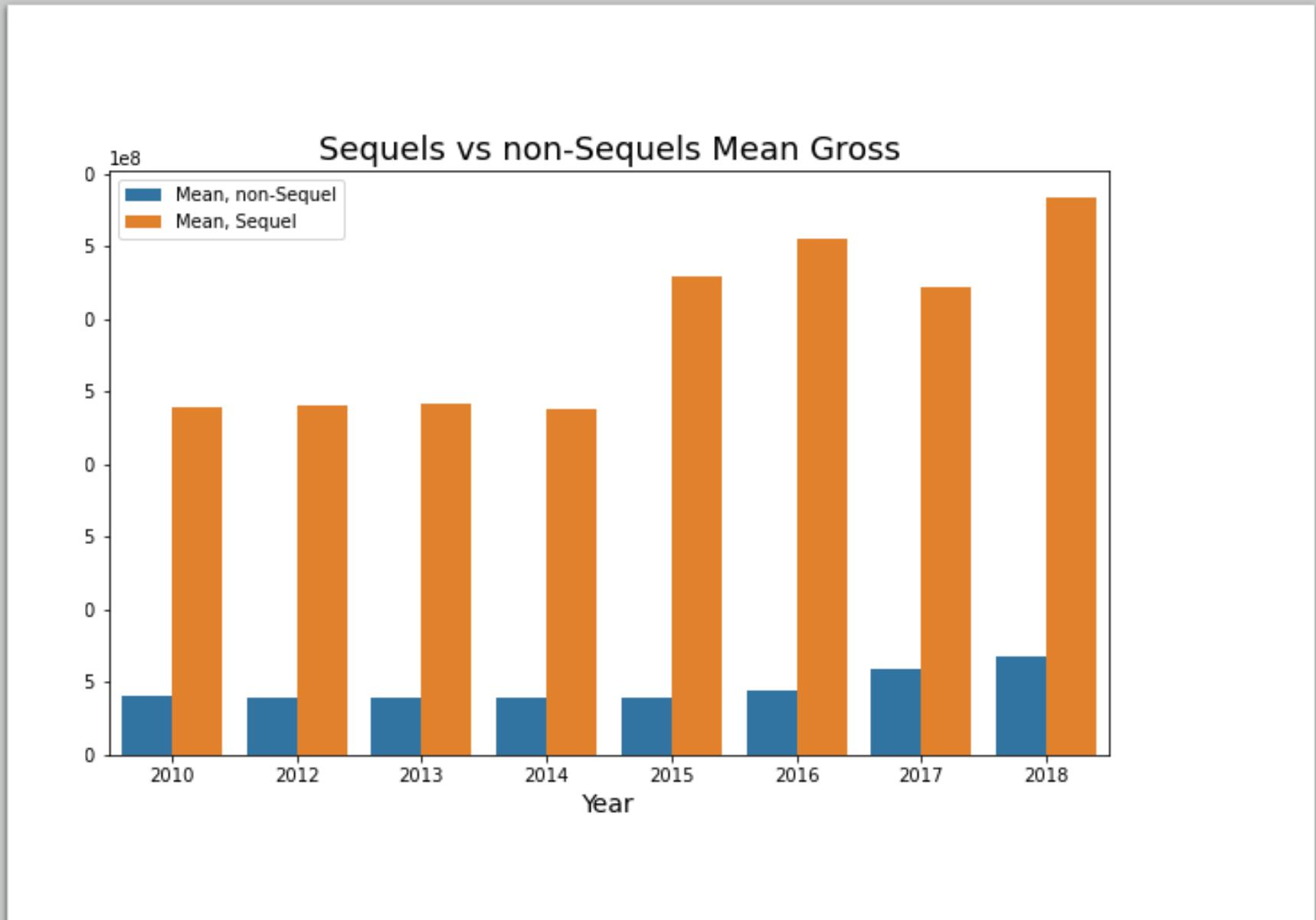
Zachary Brandt



Sequels
Perform
Worse at Box
Office Than
Their
Originals



Sequels Still
Perform
Much Better
than The
Average
Movie



Is It Worth It to Make a Sequel? Model That Predicts Worldwide Gross of a Sequel

Data scraped from IMDb:

Movies with sequels

Features:

- Box Office
 - Worldwide Gross
 - Opening Weekend Revenue
 - Budget
- Runtime
- Genre
- MPAA Rating
- IMDb Rating
- Time between release of the sequel



Target:
Worldwide Gross of Sequel

Head of Table

movie_title	worldwide_total_gross	opening_weekend_rev	budget	runtime_minutes	IMDb_rating	time_between_sequel	sequel_worldwide_total_gross
Spider-Man	825025036.0	114844116.0	139000000.0	121.0	7.3	2.160209	788976453.0
Spider-Man 2	788976453.0	88156227.0	200000000.0	127.0	7.3	2.841947	894983373.0
The Matrix	465718867.0	27788331.0	63000000.0	136.0	8.7	4.123288	741847937.0
The Matrix Reloaded	741847937.0	91774413.0	150000000.0	138.0	7.2	0.476396	427344277.0
The Lord of the Rings: The Fellowship of the Ring	887934994.0	47211490.0	93000000.0	178.0	8.8	0.996598	943282270.0

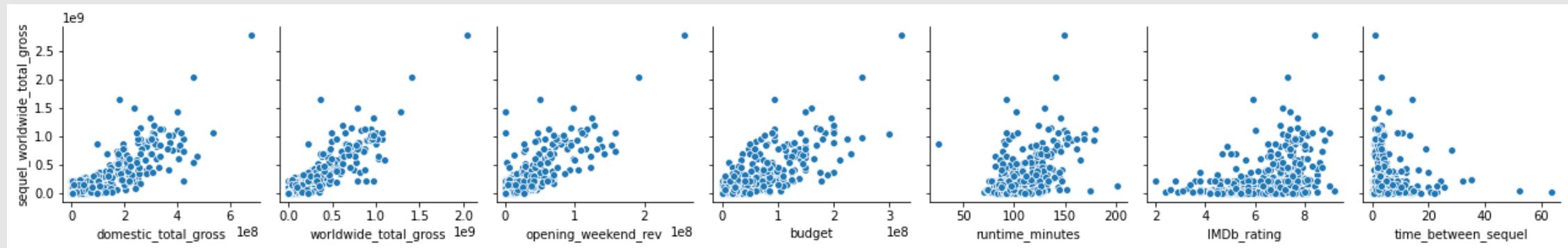
Head of Table



movie_title	worldwide_total_gross	opening_weekend_rev	budget	runtime_minutes	IMDb_rating	time_between_sequel	sequel_worldwide_total_gross
Spider-Man	825025036.0	114844116.0	139000000.0	121.0	7.3	2.160209	788976453.0
Spider-Man 2	788976453.0	88156227.0	200000000.0	127.0	7.3	2.841947	894983373.0
The Matrix	465718867.0	27788331.0	63000000.0	136.0	8.7	4.123288	741847937.0
The Matrix Reloaded	741847937.0	91774413.0	150000000.0	138.0	7.2	0.476396	427344277.0
The Lord of the Rings: The Fellowship of the Ring	887934994.0	47211490.0	93000000.0	178.0	8.8	0.996598	943282270.0

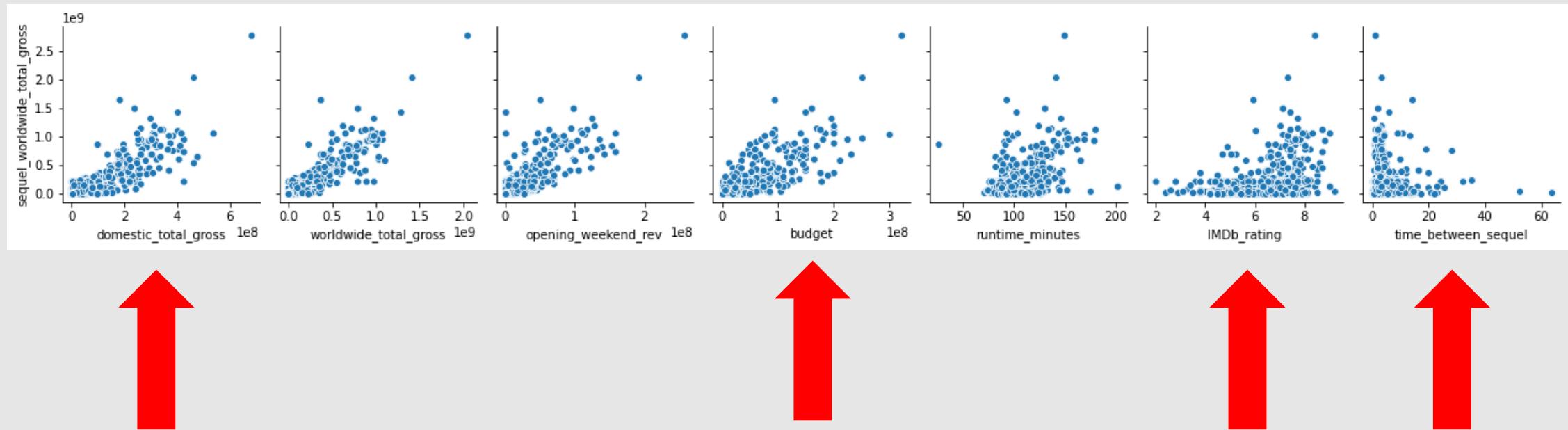
Linear Regression Model to Predict Target and Gather Insights on Relative Importance of Features

Plotted Features to Target for Feature Engineering



Linear Regression Model to Predict Target and Gather Insights on Relative Importance of Features

Plotted Features to Target for Feature Engineering



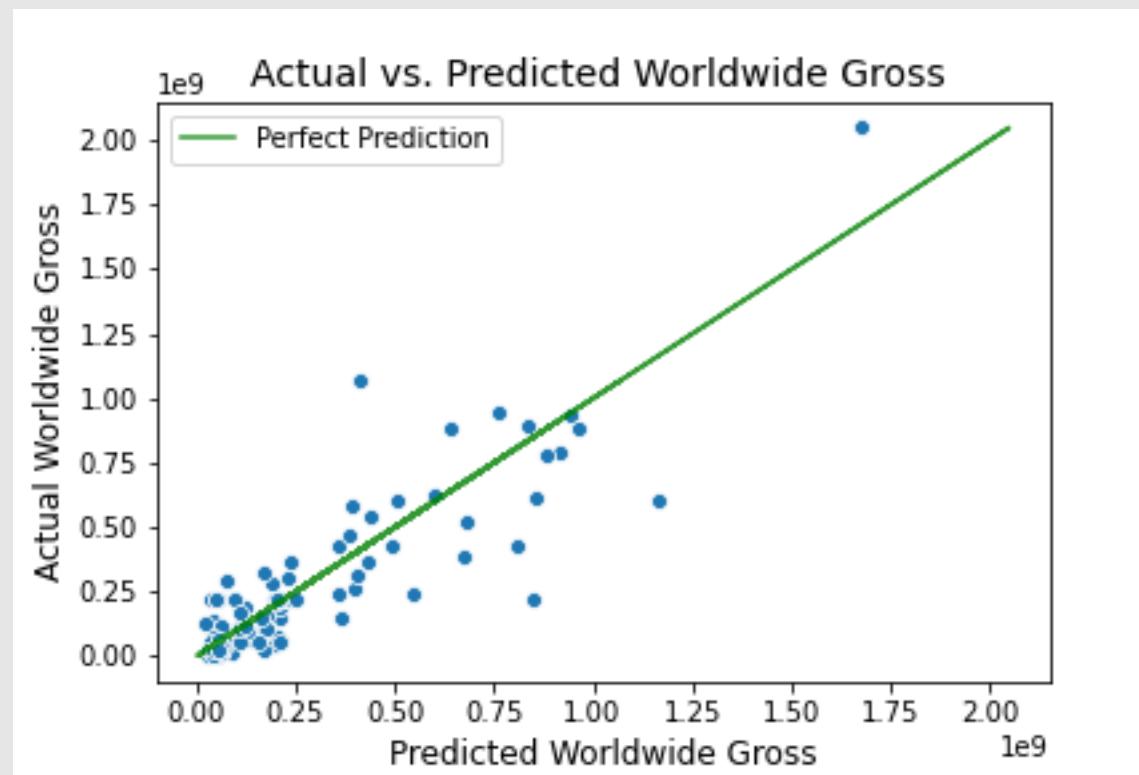
Ran 4 Regression Models

- Linear Regression
 - Train R²: 0.79
 - Test R²: 0.75
- Polynomial Regression (Interactions)
 - Train R²: 0.79
 - Test R²: -0.01
- LassoCV (with Poly)
 - Train R²: 0.81
 - Test R²: 0.73
- RidgeCV (with Poly)
 - Train R²: 0.89
 - Test R²: 0.56

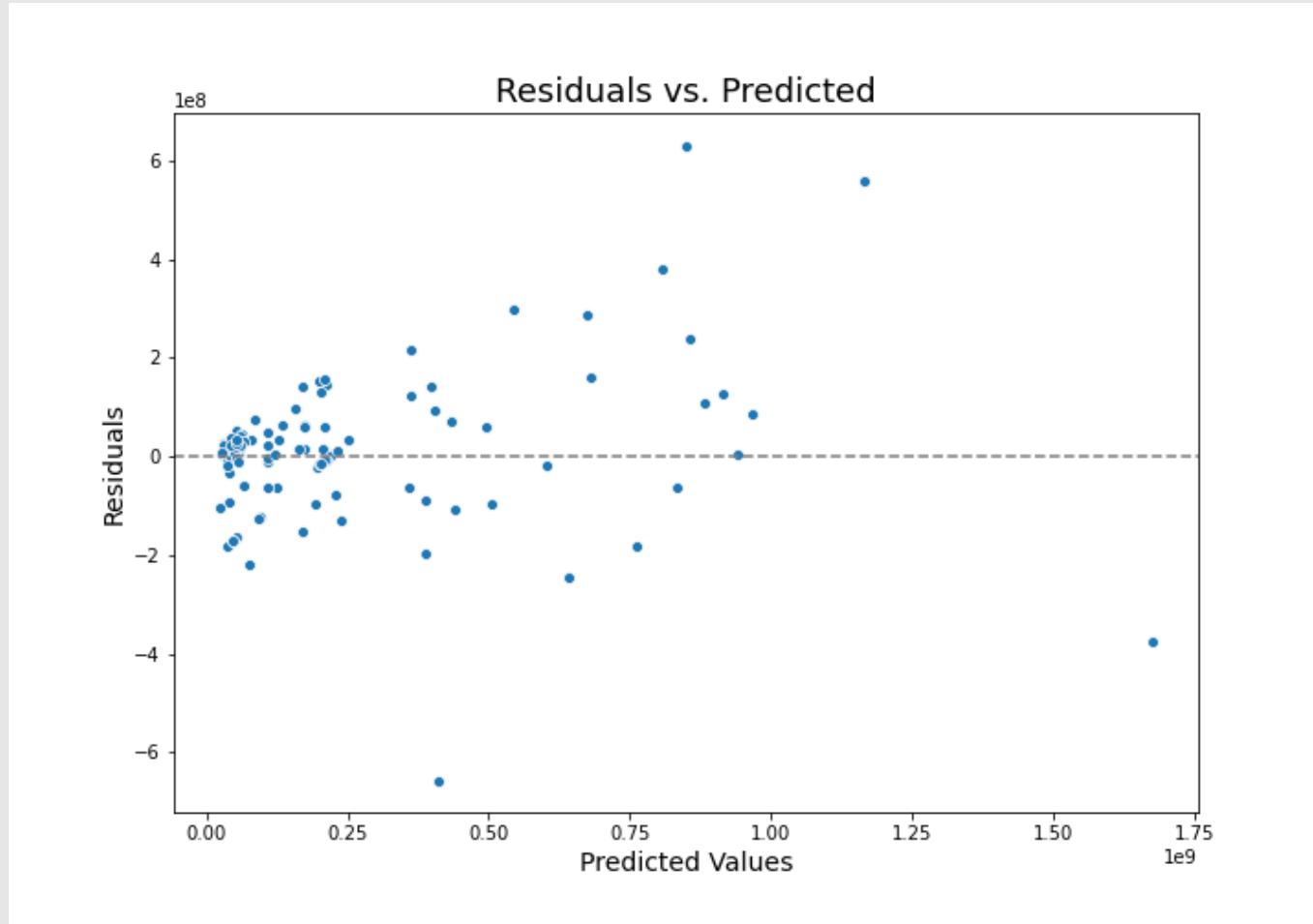
Ran 4 Regression Models

- Linear Regression
 - Train R²: 0.79
 - Test R²: 0.75
- Polynomial Regression (Interactions)
 - Train R²: 0.79
 - Test R²: -0.01
- LassoCV (with Poly)
 - Train R²: 0.81
 - Test R²: 0.73
- RidgeCV (with Poly)
 - Train R²: 0.89
 - Test R²: 0.56

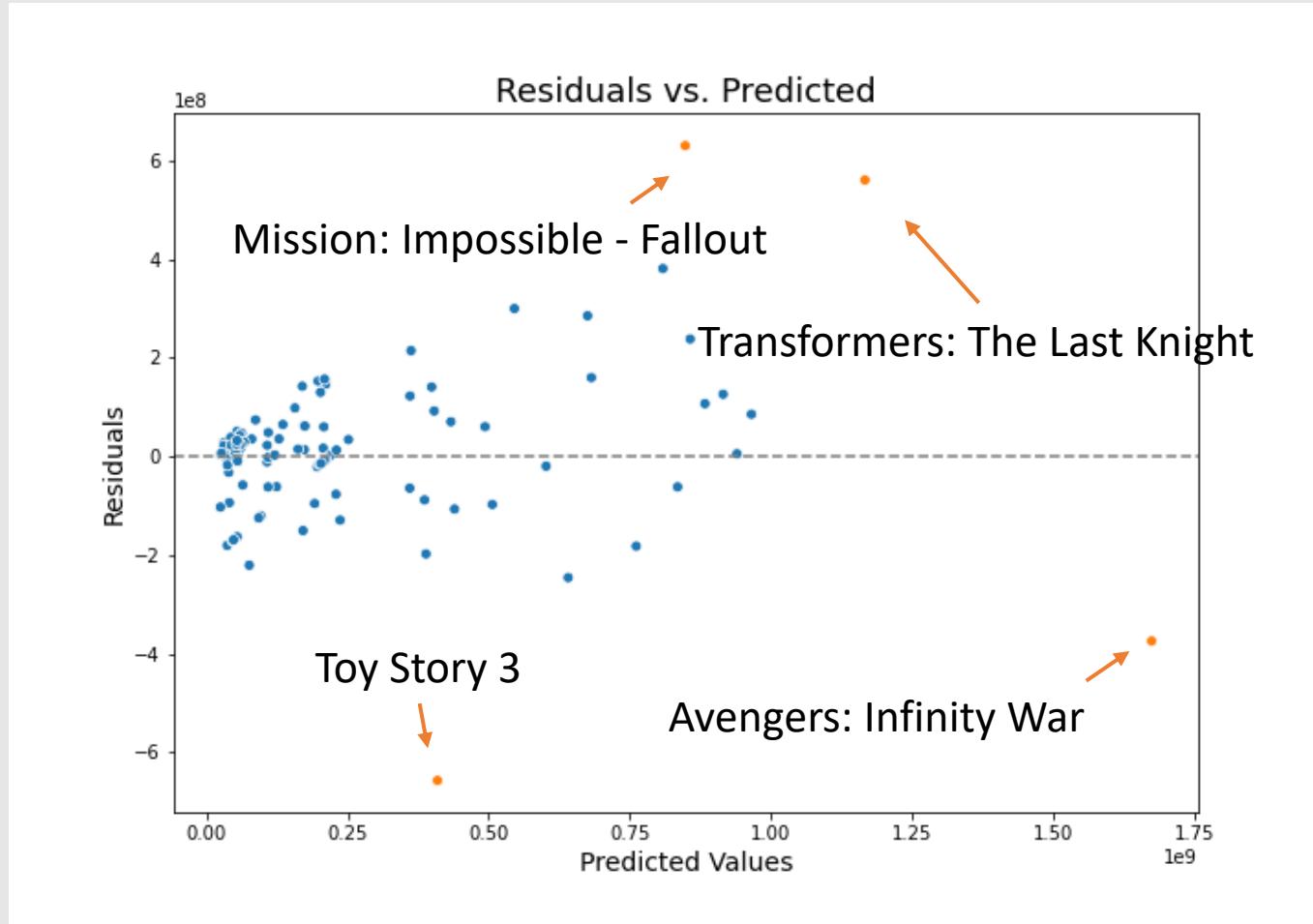
Model Predicted Better For Lower Grossing Movies



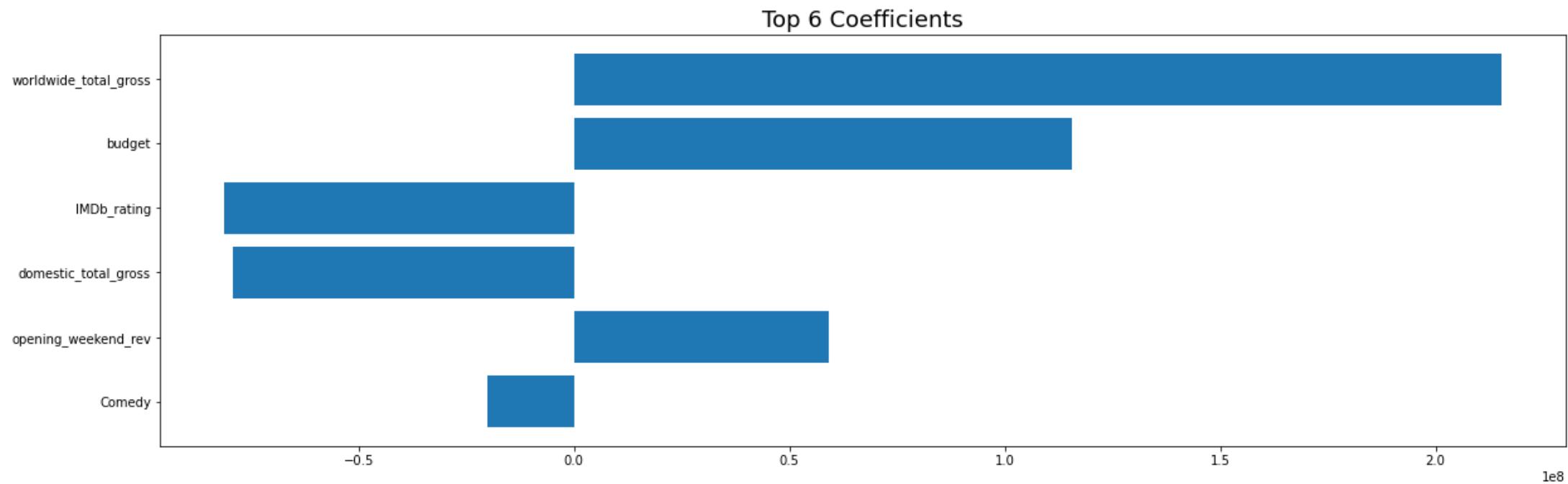
Residuals Show Model Works Best For Lower Grossing Movies

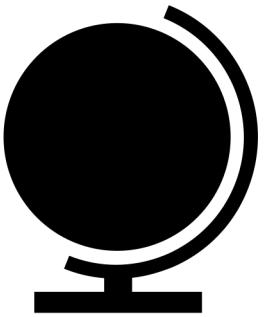
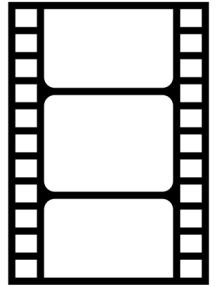


Residuals Show Model Works Best For Lower Grossing Movies



Coefficients Show that Worldwide Gross of Original is Strongest Predictor





Takeaways are two-fold

1. Can use model to predict sequel gross
 - Make call on whether it is worth it
2. Can take insights from feature coefficients
 - Worldwide gross from original
 - Genre

Questions?

Can Improve Model by Considering High Grossing Movies

- Add features:
 - Part of large series
 - Remake / Reboot

