Reboots & Remakes

Overview

Problem

- New IP can be risky
- Can we identify potential IP to revisit? (Either through Remakes or Reboots)
- What factors contribute the most to success? (Gross)
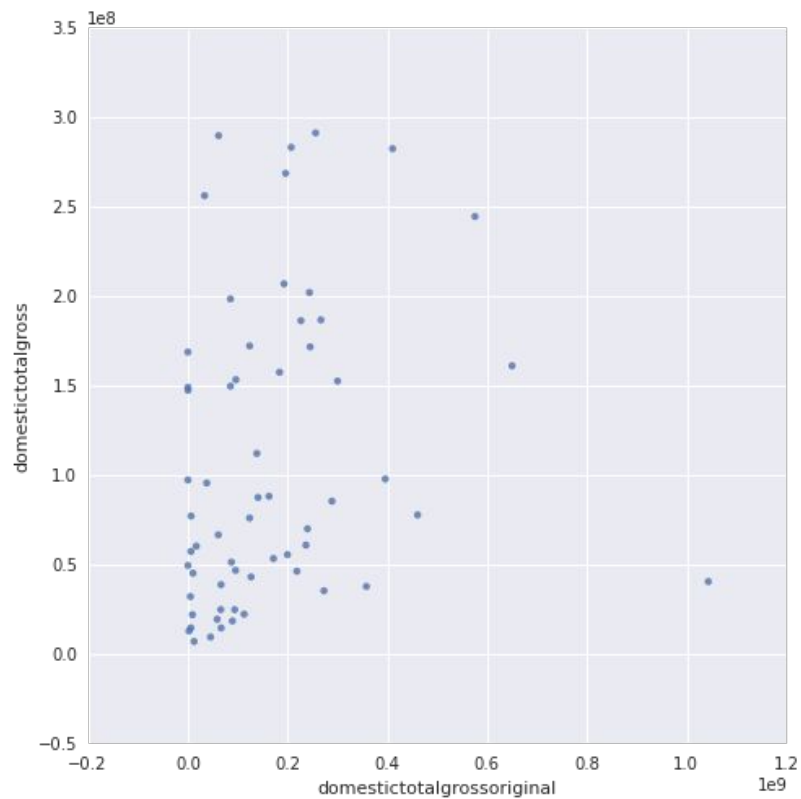
Data & Methodology

Data

- Scraped Box Office Mojo
- Collected information about movies including Gross, Budget, Release Dating, MPAA Rating, and Genre
- Ended up with 70 pairs of movies
- Lots of missing data from older movies

Methodology

- Selected features from scraped data
- Modeled Gross vs each of the features
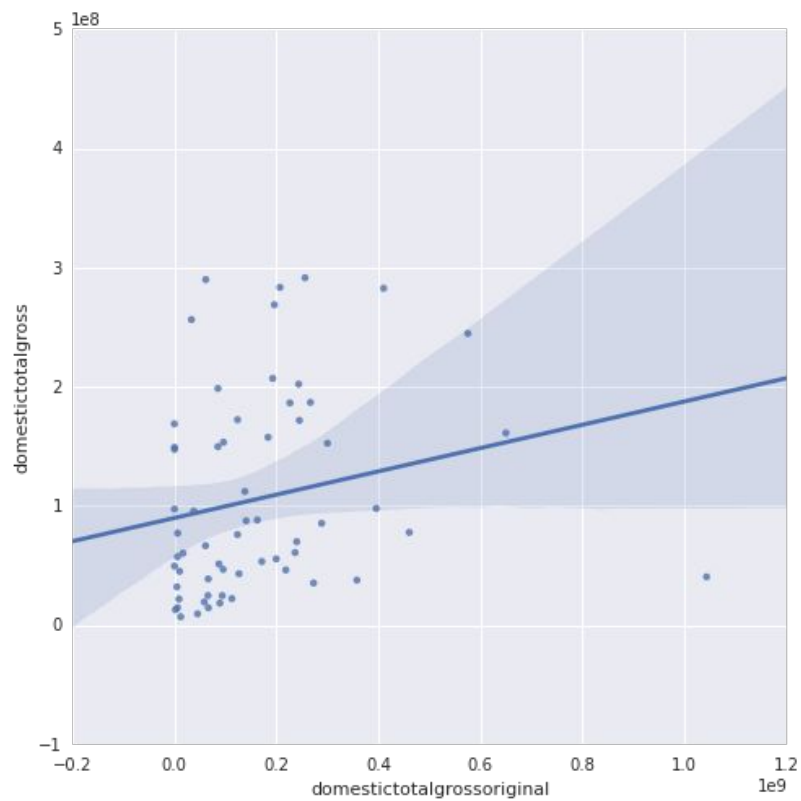- Combined Features into one model to see improvements
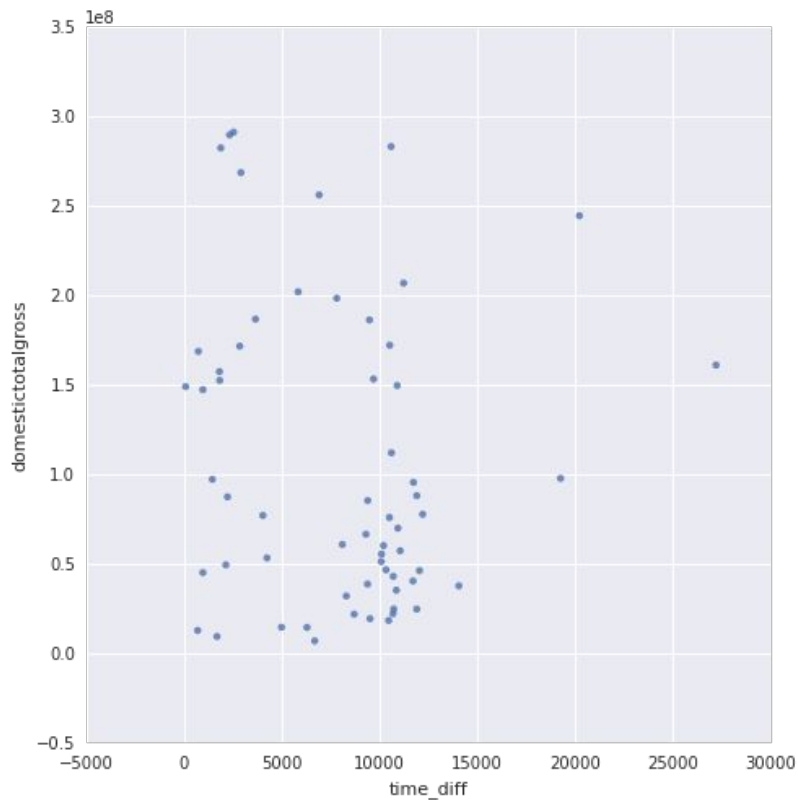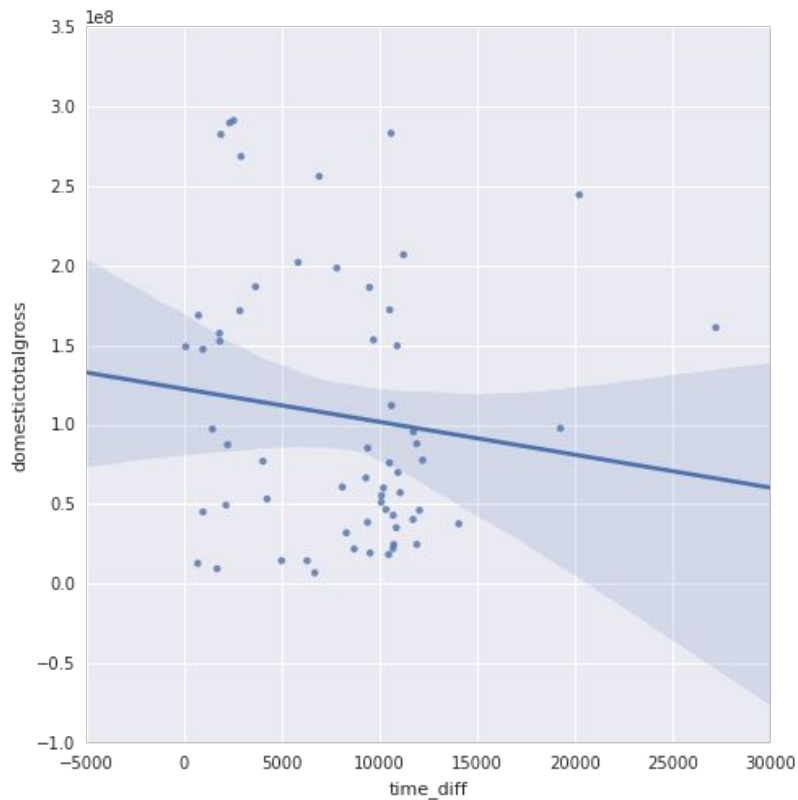
Modeling & Analysis
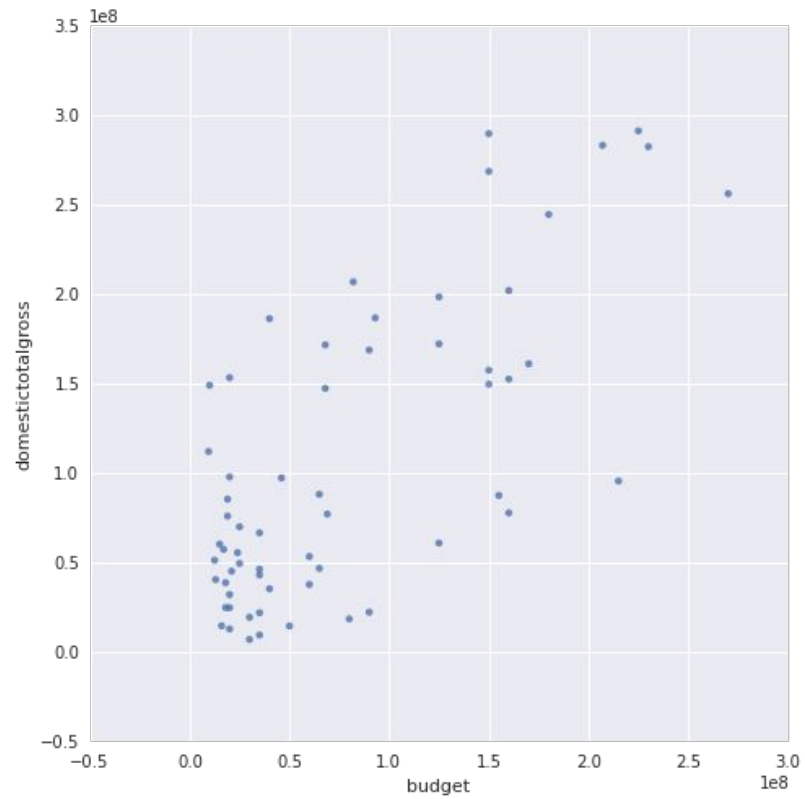
Original Gross

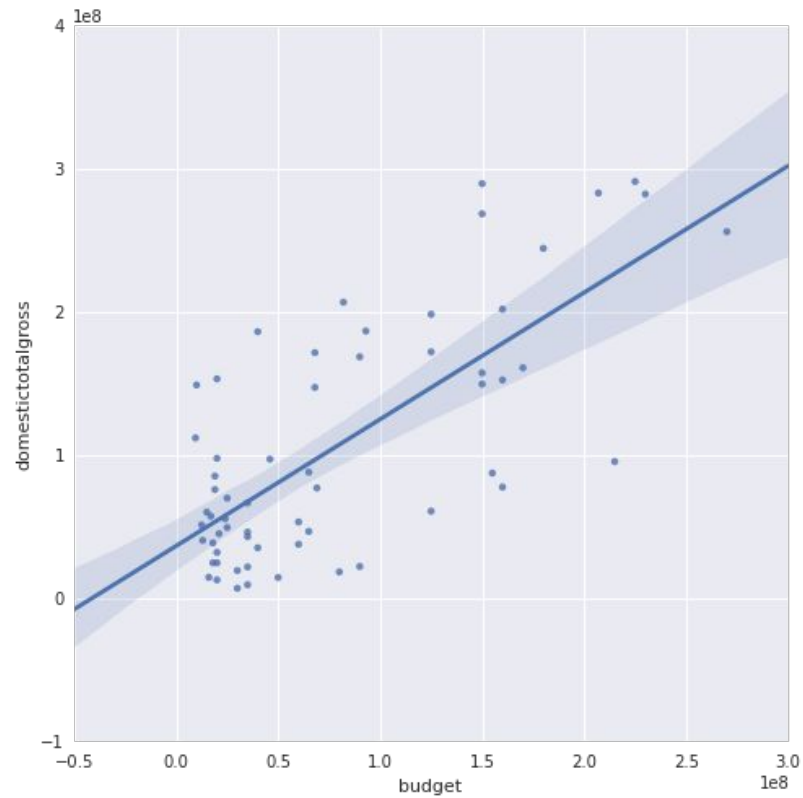# Original Gross

Time Difference Between Release Dates

Time Difference Between Release Dates

Budget

Budget

Budget vs Budget, Original DTG, and Time Diff

Features	R^2	Adjusted R^2	Mean Squared Error
Original Gross	0.046	0.030	6.55×10^{15}
Time Difference	0.017	0	6.75×10^{15}
Budget	0.528	0.520	3.24×10^{15}
Budget & Gross & Time Diff	0.552	0.528	3.07×10^{15}

Conclusions

Conclusions

- Movies are complex nonlinear systems
- Need more data!
- Budget is a somewhat reasonable indicator of DTG

Next Steps

- More Complex Models - Decision Trees/Adaboost/SVR
- Collect More Data
- Engineer better features

Questions?