Fill In The Gap Productions

An Investigation into Maximizing Domestic Gross

Trevor Bedrin
9/19/2014

The Question

Is it possible to determine the best possible projects for the client to pursue based on very limited data?

- Budget
- Release Month
- Runtime
- Genre
- MPAA Rating

The Premise

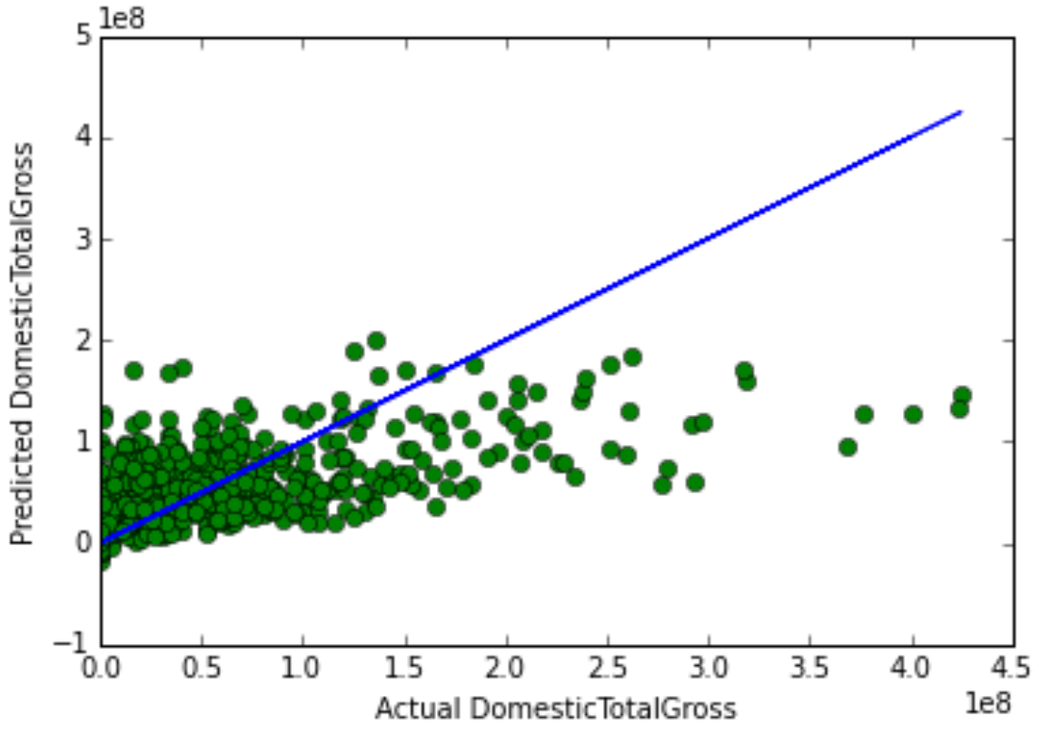
- We started with the assumption that there is an underlying relationship between our selected features and the actual Domestic Gross that can be well represented by a linear regression model.

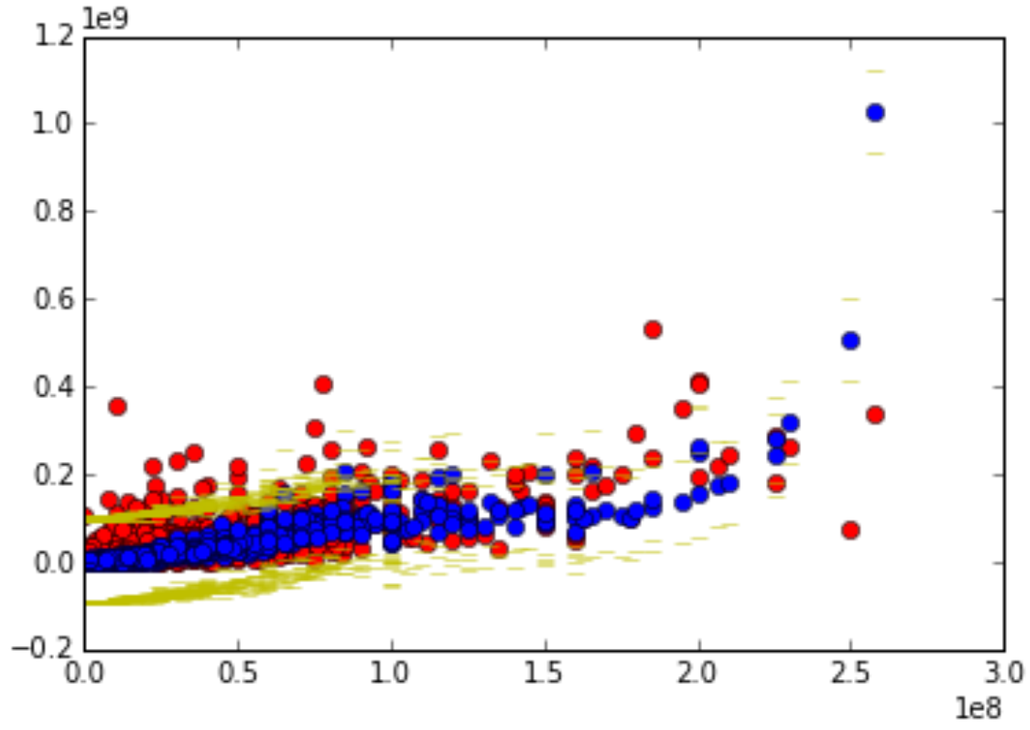
- We ended up with a model that used a combination of Runtime, Release Month, Genre, and Rating as well as interactions between Genre and Month and Genre and Rating.

- Where is Budget? It turns out that removing Budget from the model allows it to predict values that more closely matched the actual Domestic Gross of the test set.

Performance

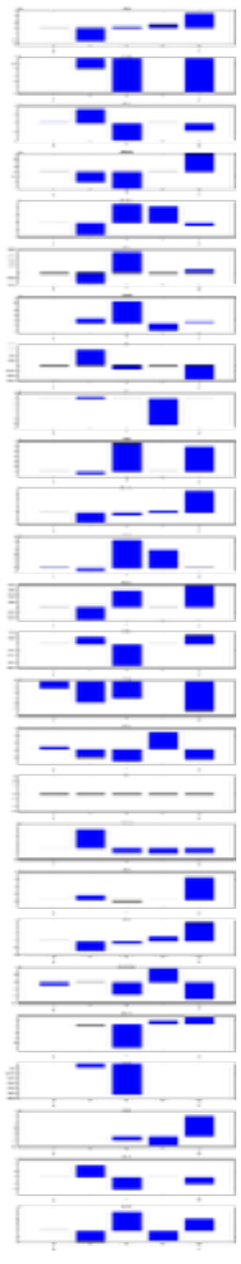
The model, while unable to exactly predict the Domestic Gross based on the provided parameters, is able to predict, with 95% confidence, the range that the Domestic Gross will fall in. It turns out that this interval is the predicted value +/- $52,138,469.57.



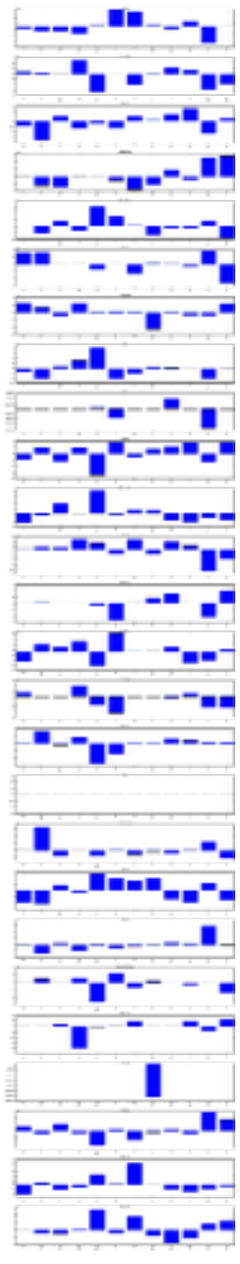


What else do we know?

Rating



Month



- Because the predicted value of the model is the Domestic Gross, we can use the model to get a sense of how many dollars each parameter actually contributes to the total.
- The charts on the right, which are way too small to actually make out in this presentation, plot the value each genre/rating (left chart) or genre/month (right chart)
- How is this useful information? Using these values we can determine the best release month and rating to release a movie of a specific genre.

Target These Projects

Maximizing Domestic Gross by Genre

	Month	Month Value	Rating	Rating Value			Month	Month Value	Rating	Rating Value
Sci-Fi	Jun	$39,624,597.10	PG-13	$12,541,979.82		Romantic	Jun	$19,247,041.44	PG-13	$2,969,235.01
Unknown	Apr	$13,307,792.61	R	-$3,616,569.18		Foreign	Apr	$10,348,345.61	G	$0.00
Crime	Oct	$13,080,512.51	R	$8,549,565.80		Drama	Feb	$16,662,982.72	G	$12,446,593.88
Romance	Dec	$26,428,308.47	PG-13	$19,769,412.09		IMAX	Mar	$0.00	N/A	$0.00
Animation	May	$63,630,290.58	PG	$27,005,070.83		Historical	Feb	$63,623,563.91	R	$31,300,013.65
Music	Nov	$8,961,344.50	PG	$3,668,162.29		Action	May	$28,361,487.40	PG-13	$16,292,634.18
Epic	Sep	$4,125,056.38	R	$783,152.99		Sports	Nov	$30,525,101.31	PG-13	$13,004,315.02
War	May	$22,462,374.78	R	$6,297,152.75		Documentary	Jun	$6,156,618.39	G	$9,582,533.15
Comedy	Jan	$6,248,980.46	PG	$10,298,011.82		Musical	Dec	$26,095,855.03	PG-13	$16,011,548.27
Horror	Jun	$9,415,106.50	PG	$10,997,339.13		Concert	Mar	$0.00	R	-$410,263.38
Adventure	May	$57,393,169.91	PG-13	$37,521,581.42		Family	Nov	$29,640,746.36	PG-13	$20,903,531.05
Thriller	Apr	$12,921,880.26	PG	$14,655,093.92		Period	Jul	$25,371,064.40	R	$10,012,757.50
Western	Dec	$10,356,453.17	PG-13	$4,966,078.55		Fantasy	May	$53,427,158.55	PG	$32,098,201.78

Moving Forward

With more time to pursue this project it would be possible to pull in additional data points from sources other than Box Office Mojo.

Directors, Producers and Featured Actors are all features that could have a real impact on the predictive ability of the model. The number of other, similar projects slated for release during the same month would also have an affect.

This analysis also only looked at Domestic Gross. Incorporating foreign gross into the model would allow it to much better predict what projects offer the best chance of being profitable.