The Effect of Famous Actors on Summer Blockbuster Grosses
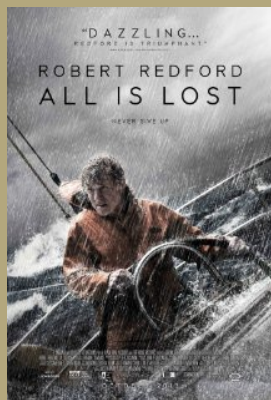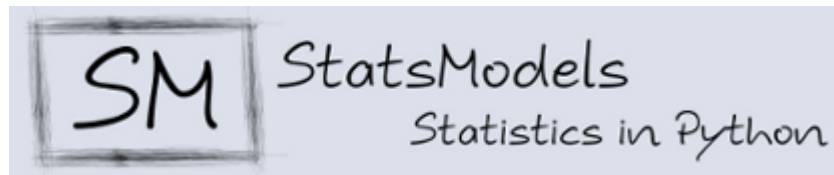
John Gilling

Original Question:
Could your summer blockbuster benefit from having one or more "superstars" in it?
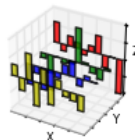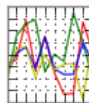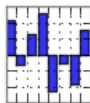

Subsequent Questions:
- Should you choose stars who have been in more films, or stars who gross more money in their films on average?
- Is there a tipping point?
- When do your returns start to diminish?

Tools

python

Requests

scikits learn
machine learning in Python

SM StatsModels
Statistics in Python

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

matplotlib

Pydata/patsy

NumPy

Beautiful Soup

Assumptions

How do we define our terms?

- A "famous actor" is someone who has been in at least 10 films and is in the top 100 for average gross over these films (data from BoxOfficeMojo).

- A "summer blockbuster" is a movie that came out between the first weekend in May and Labor Day weekend, and was one of the top 50 grossing movies that year (years 2011-2015).

Process

1. Scrape BoxOfficeMojo for blockbusters and famous actors.

2. Scrape IMDB for lists of actors in blockbusters.

3. Scrape deeper on IMDB for historical grosses (pre-2011) for each actor from the blockbusters.

Building a model

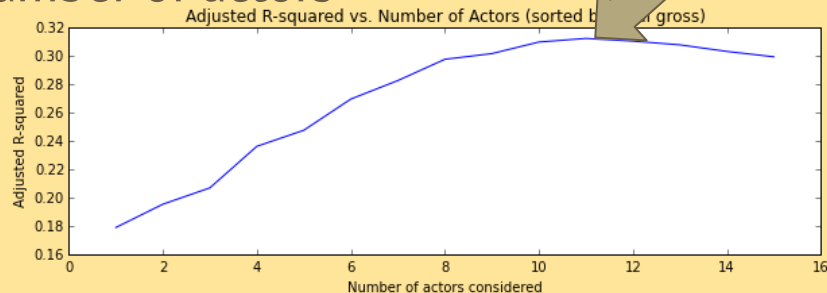First, we want to figure out how many actors from each blockbuster to consider.
Should we take the top N actors in terms of …
- Total number of movies?
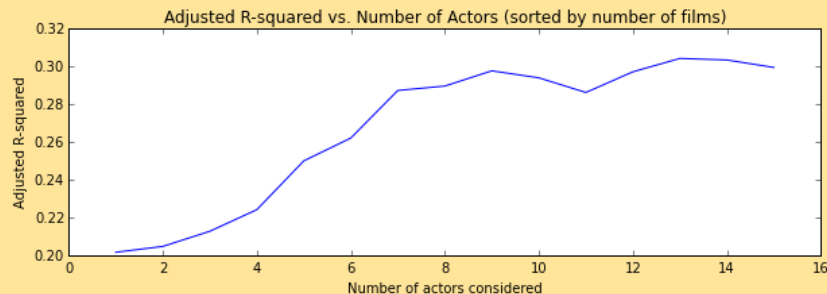- Total historical gross?
- Average gross per movie?

I built a number of different models based on each of these assumptions, varying the number of actors considered in each model:
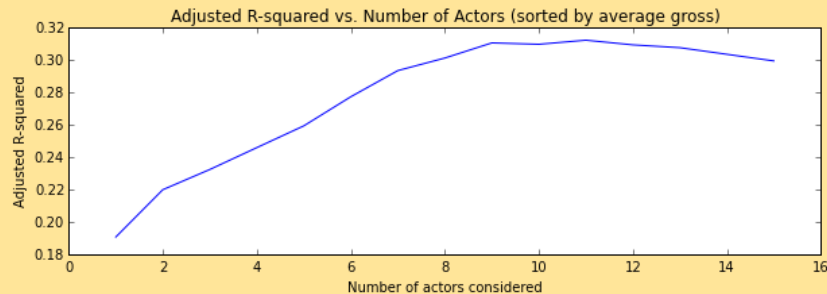
TOTAL GROSS

TOTAL # OF FILMS

AVERAGE GROSS

Adjusted R-squared vs. Number of Actors (sorted by total gross)

Adjusted R-squared vs. Number of Actors (sorted by number of films)

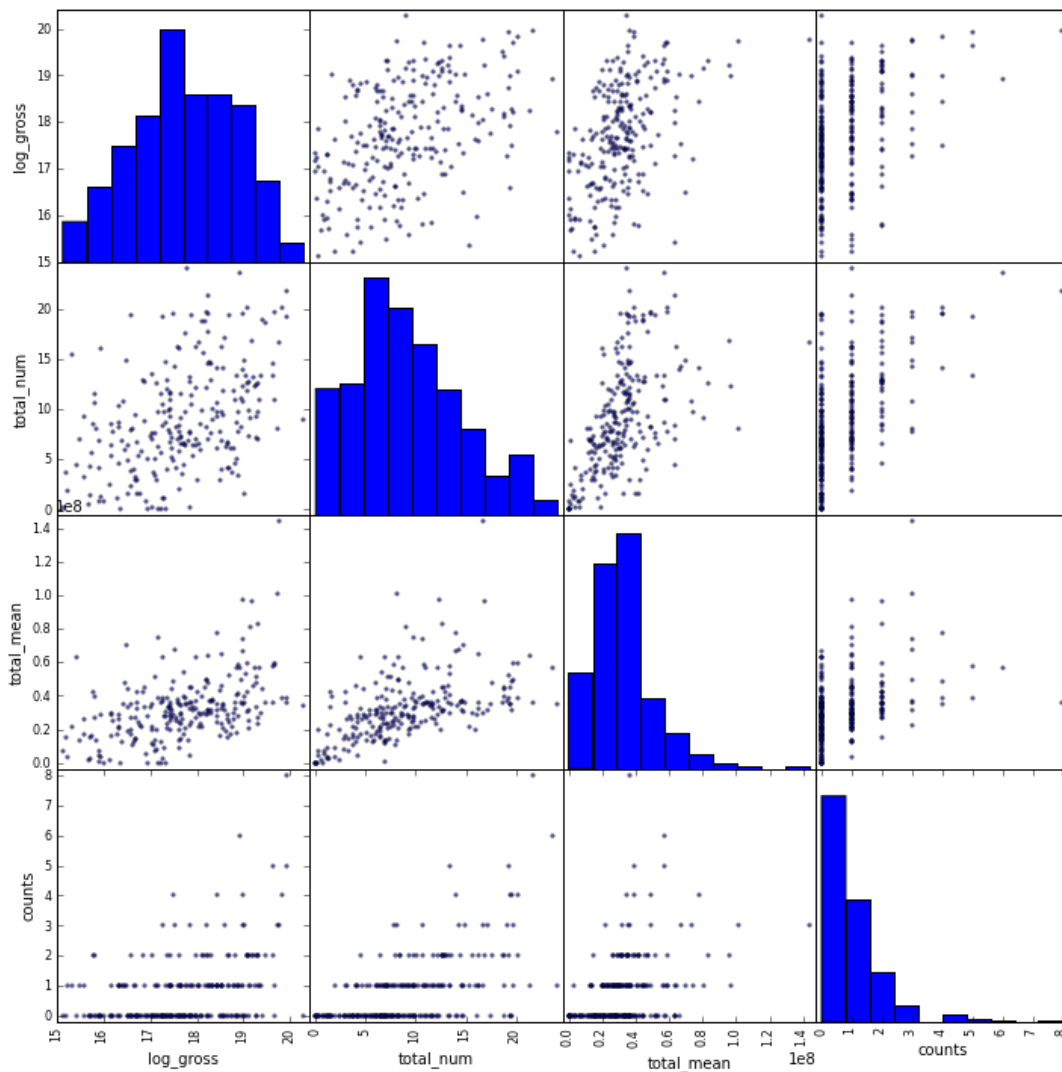Adjusted R-squared vs. Number of Actors (sorted by average gross)

Final Model
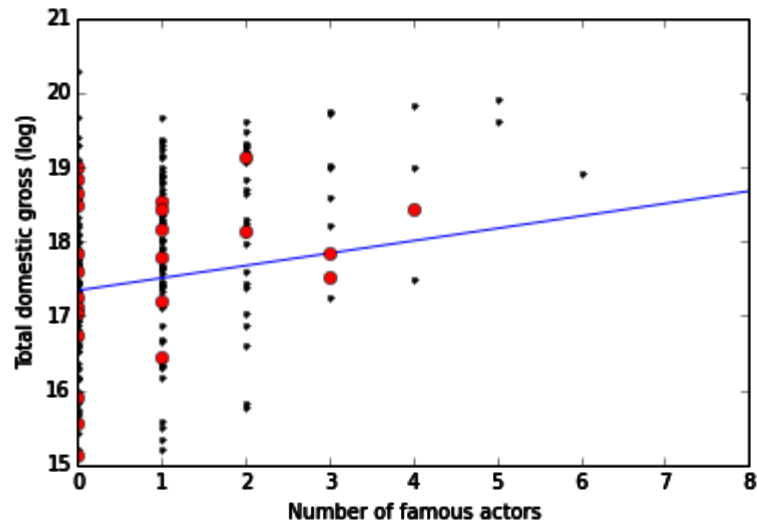
The four histograms show distributions of:

★ Blockbuster Gross

★ Actor Experience
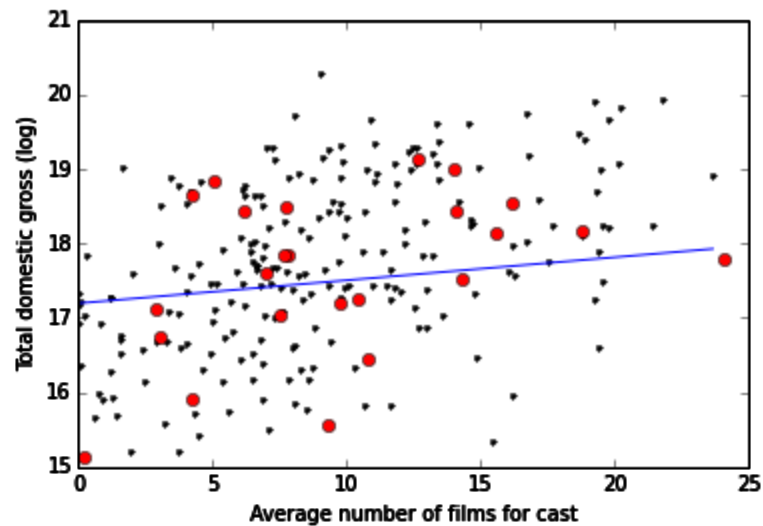
★ Actor Success

★ Number of "famous actors"

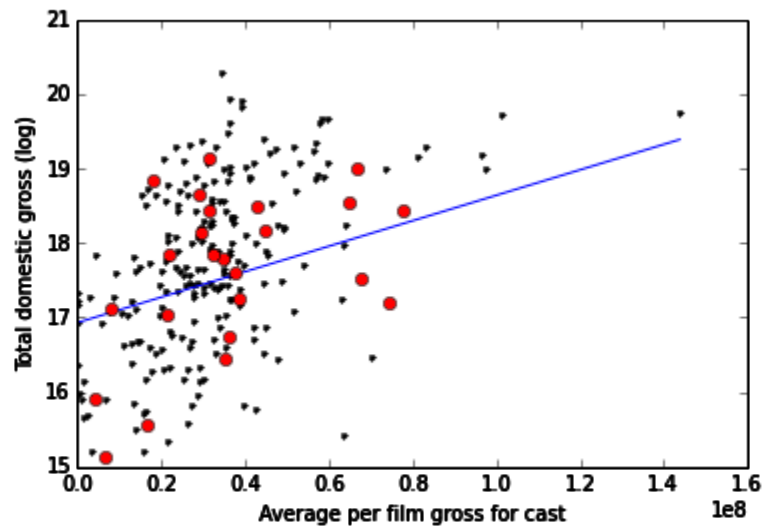Let's zoom in on the top row of relationships, i.e. how our features do at predicting total blockbuster gross.

Number of famous actors

Experience of actors

Success of actors

Conclusions

- If you want to *guarantee* a high grossing summer blockbuster, you should include at least 3 famous actors from my nifty list. Results may vary (and in pretty similar ways) when you only include 0-2 of these actors.
- The total number of movies your cast has been in is not a great indicator of the success for your movie (but the final model is more accurate with a more experienced cast!)
- Averaging the top 11 actors may have its own drawbacks, as viewers may only care about the first three or four billed actors …

Next Steps

- ❏ Take a look at other features (e.g. directors, amount spent on special effects, genre, etc.)
- ❏ Expand the range of years and/or shrink the number of movies considered each year.
- ❏ Take ROI/cost of actors into consideration.

THANKS Y'ALL!

Who's got questions?