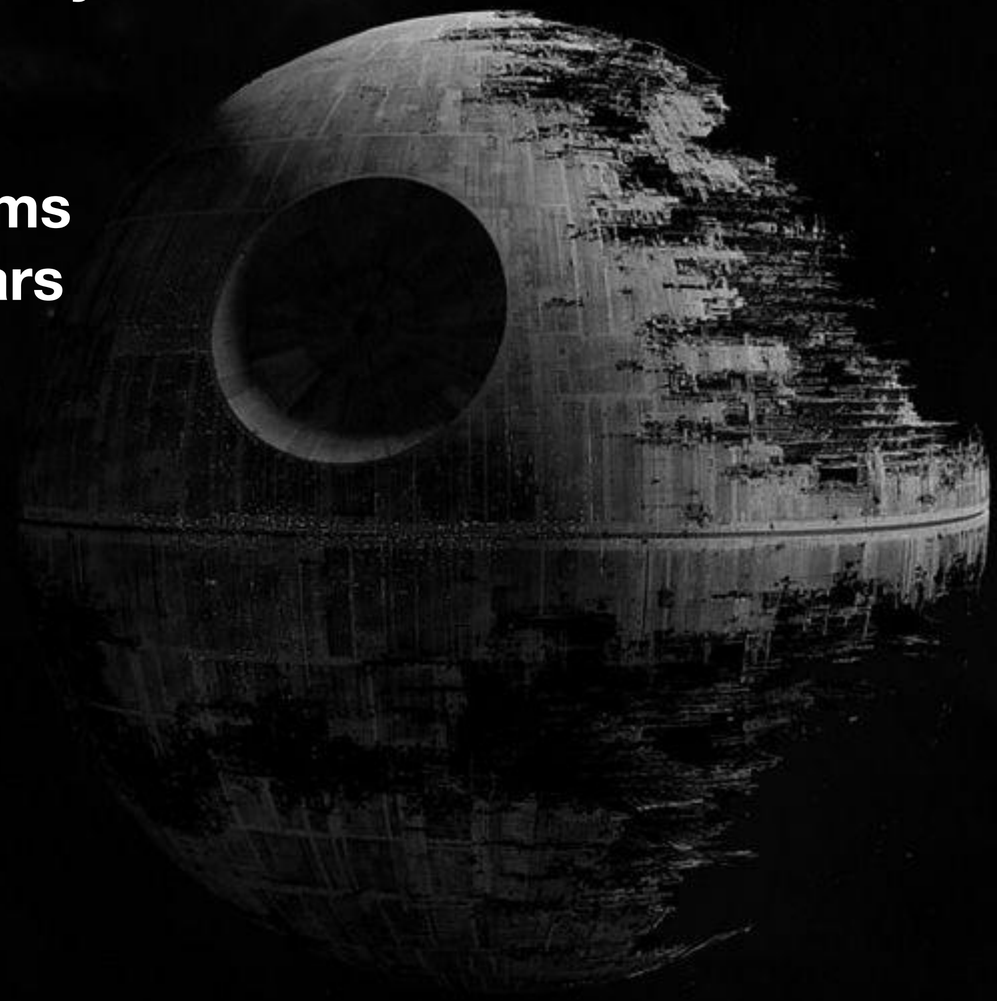
Modeling profits for the most successful films that never made it to #1 on the Box Office

Metis Project 2

By Nathan Huvelle

Virtually every weekend the Box Office chart is dominated
By yet another Super Hero sequel pumped out by one of the
Few powerhouse franchises in the country.

10 out of the top 20 highest lifetime grossing films
By domestic gross revenue are either a Star Wars
Or Avengers film.




PERFECTLY UNBALANCED
AS ALL THINGS SHOULD BE
imgflip.com

Introduction

Goal: Model and Interpret various features and their Effects on movie profit at the box office

Data was scraped from boxofficemojo.com
Using BeautifulSoup and analyzed in Pandas.
Data about the top 1000 movies that never Made it to #1 on the Box Office chart was Used as the dataset.

The main modeling technique used was multiple Linear Regression using an assortment of quantitative And categorical features. The target was profit.

Results

Dep. Variable:	profit	**R-squared:**	0.758
Model:	OLS	**Adj. R-squared:**	0.756
Method:	Least Squares	**F-statistic:**	369.8
Date:	Fri, 09 Jul 2021	**Prob (F-statistic):**	5.33e-179
Time:	00:27:24	**Log-Likelihood:**	-10914.
No. Observations:	595	**AIC:**	2.184e+04
Df Residuals:	589	**BIC:**	2.187e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.301e+09	3.44e+08	6.685	0.000	1.62e+09	2.98e+09
Year	-1.174e+06	1.74e+05	-6.745	0.000	-1.52e+06	-8.32e+05
widest	1.454e+04	2857.514	5.088	0.000	8926.915	2.02e+04
opening	2.6426	0.123	21.471	0.000	2.401	2.884
budget	-1.1086	0.029	-38.137	0.000	-1.166	-1.052
runtime_minutes	3.738e+05	5.93e+04	6.302	0.000	2.57e+05	4.9e+05

Omnibus:	267.583	**Durbin-Watson:**	1.126
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1785.516

Features analyzed: Year of release, month of year, budget, movie length, genre, distributor, rating, widest release, opening weekend gross.

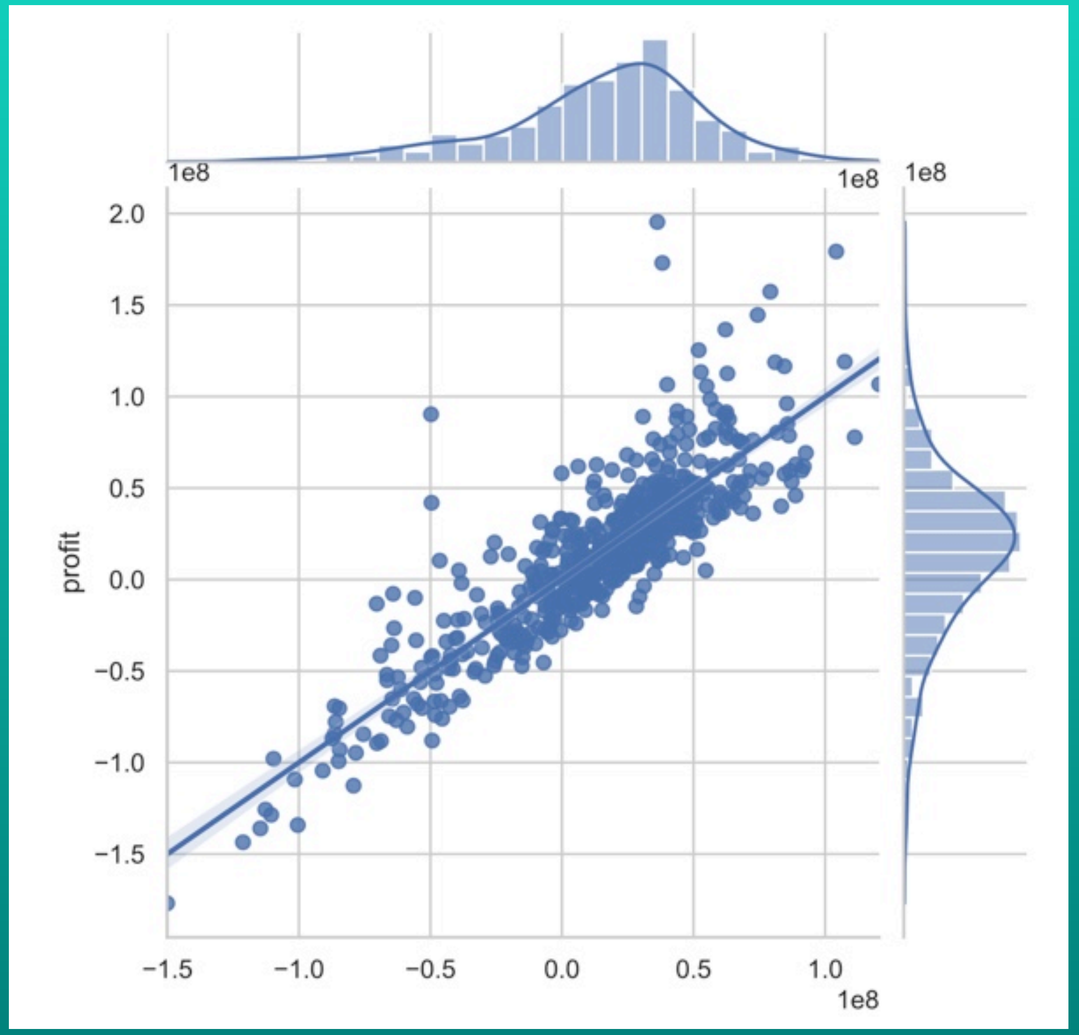
Most influential features: Budget (by far), Year, Month, Rating.

Best months: December, June.
Worst months: January, October.

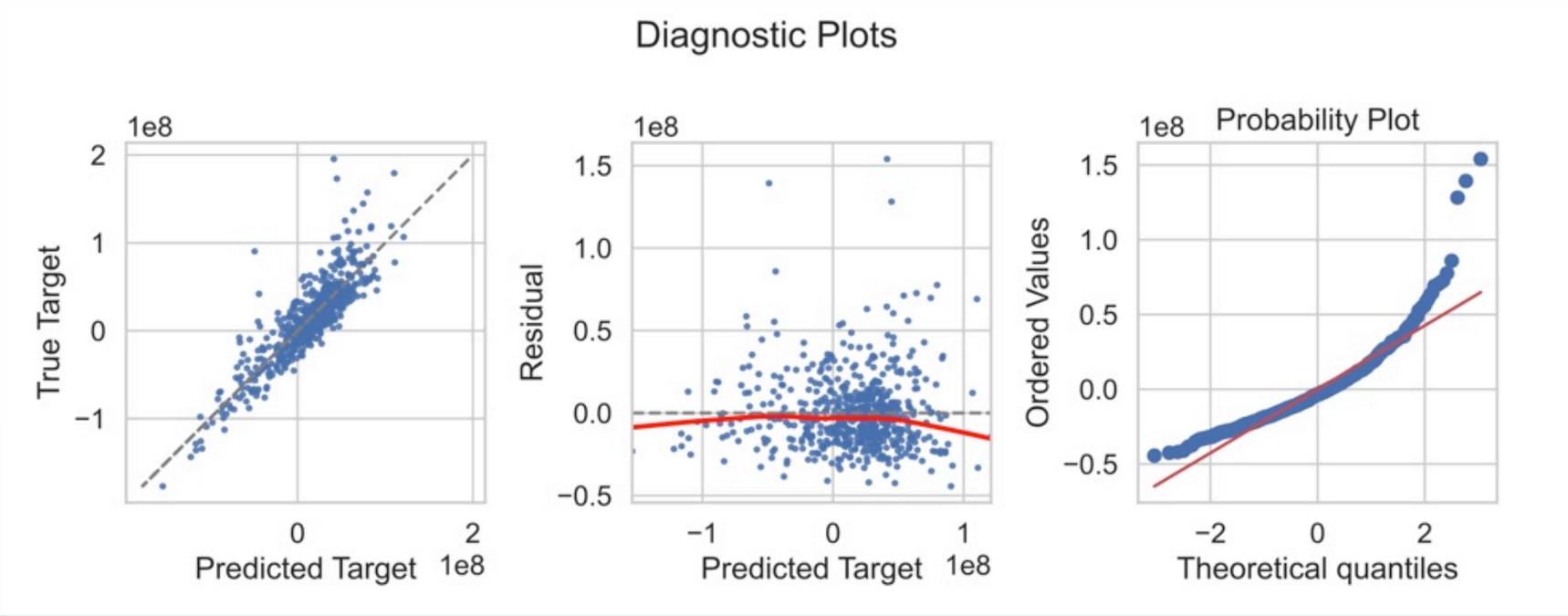
Best Rating: G

Least influential features: Distributor, Runtime, widest release

Highest R^2 obtained: 0.83 using Linear Regression and a simple Train, validate, test approach.
Technically got a 0.845 using Ridge Regression,
But not confident in that one.



Analyzing Residuals



Conclusions

Remember the goal was to interpret features that led to highest profits
Among movies that don’t top the Box office charts. And to provide
Actionable advice for the dreamers out there who have an awesome
Movie script and want to compete with the likes of Marvel/DCU et al.

Step 1: Go back to 1990

Step 2: Release a G-rated family film, in December, on a low budget.

Step 3: ??????????

Step 4: Profit!

Future Steps