Movies Go Abroad

Objective

How well can we predict international box office based off of movie pre-release features and domestic box office information?

What features have a stronger correlation to box office success?

Data Collected

IMDb®

Box Office Mojo by IMDbPro

THE NUMBERS
Where Data and the Movie Business Meet

- 1,750 movies
- US wide released from 2000-2020
- Top grossing

Model Selection and Workflow

Cross Validation Tests	R^2
Linear Regression	0.535

Poly LASSO	0.626

Runtime

Budget

Release Month

Metacritic Score

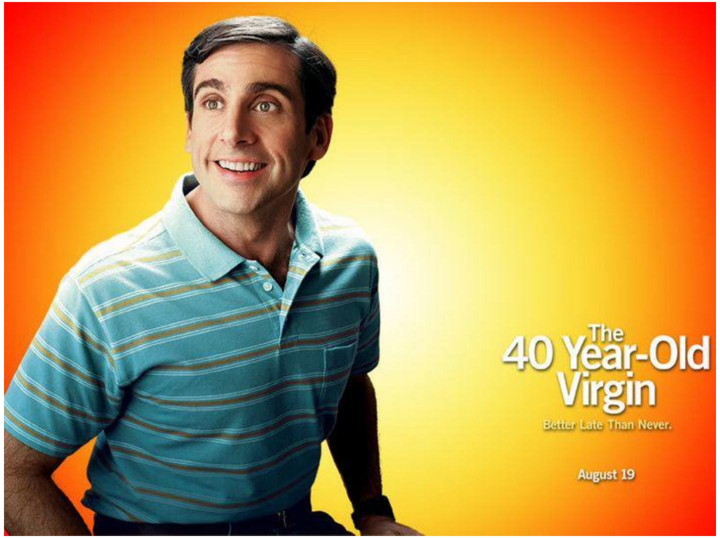
“Adventure” Genre

Director

Dropping Outliers

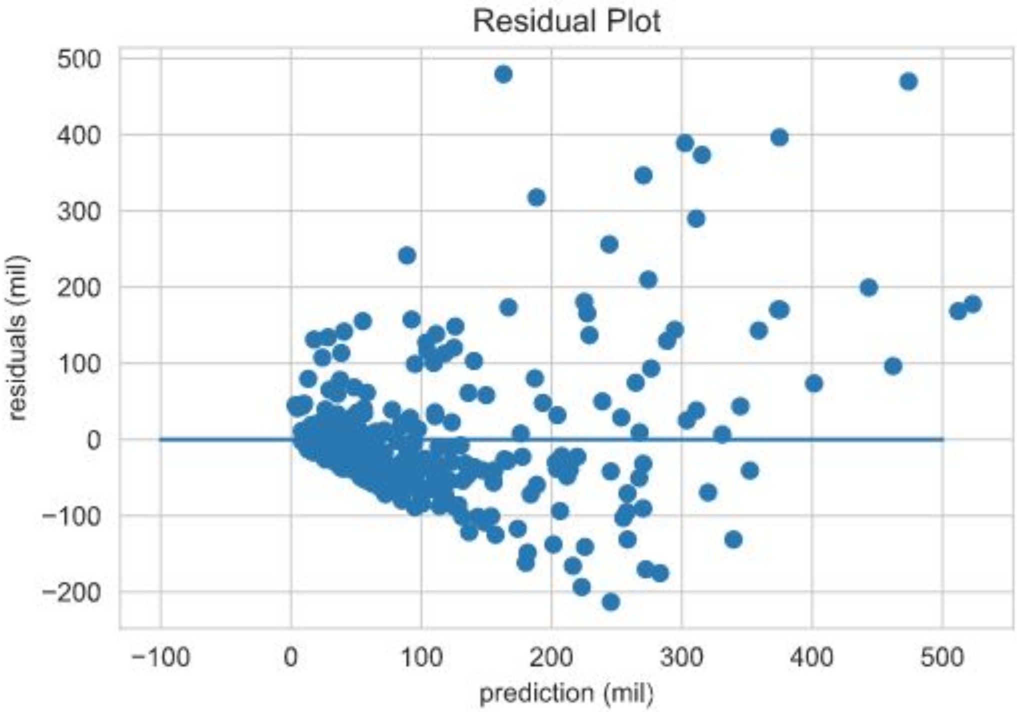
Movie Predictions

Film Title	Actual (mil)	Predicted (mil)	Error (mil)
The 40 Year Old Virgin	$67.93	$66.16	$1.77
Rat Race	$28.88	$25.70	$3.18
Transformers	$691.28	$302.40	$388.88
Avengers	$943.80	$474.03	$469.78



$R^2 = 0.626$
RMSE: 95.75

Final Model Residual Plot

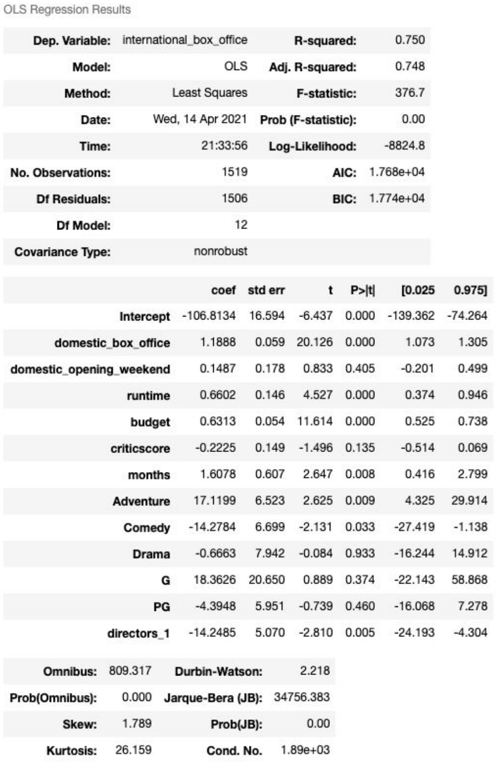


R² = 0.626
RMSE: 95.75

Future Work

- Features:
  - Franchise
  - Popular int’l actors or directors
  - Popular int’l genres
- Log Regression
  - Too skewed?

Appendix 1

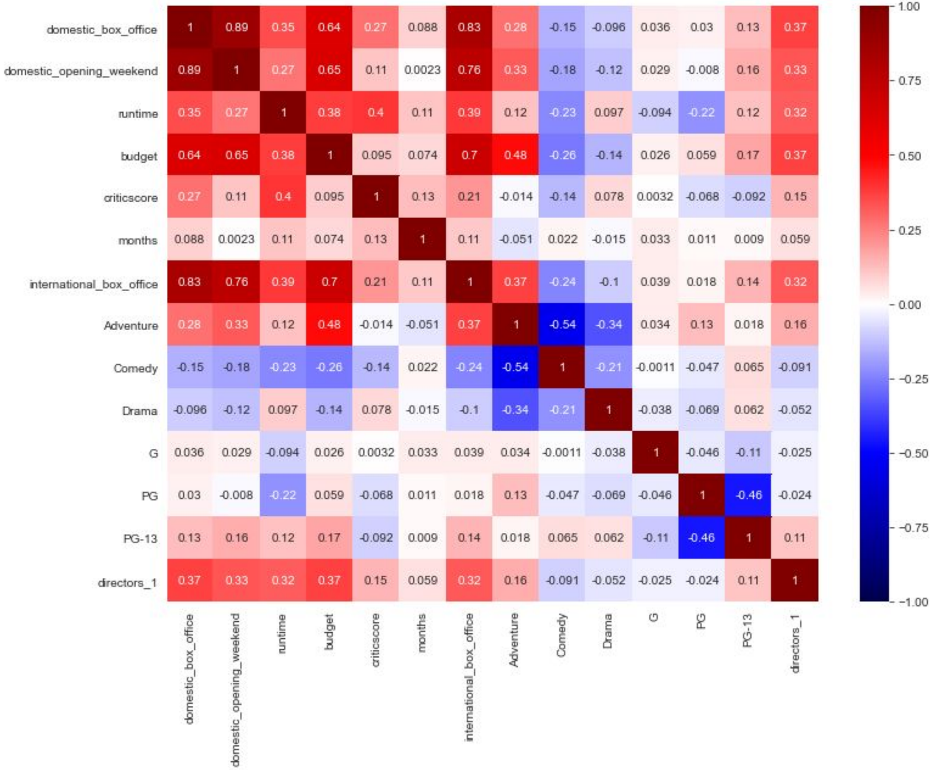


OLS Regression Results

Dep. Variable:	international_box_office	**R-squared:**	0.750
Model:	OLS	**Adj. R-squared:**	0.748
Method:	Least Squares	**F-statistic:**	376.7
Date:	Wed, 14 Apr 2021	**Prob (F-statistic):**	0.00
Time:	21:33:56	**Log-Likelihood:**	-8824.8
No. Observations:	1519	**AIC:**	1.768e+04
Df Residuals:	1506	**BIC:**	1.774e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-106.8134	16.594	-6.437	0.000	-139.362	-74.264
domestic_box_office	1.1888	0.059	20.126	0.000	1.073	1.305
domestic_opening_weekend	0.1487	0.178	0.833	0.405	-0.201	0.499
runtime	0.6602	0.146	4.527	0.000	0.374	0.946
budget	0.6313	0.054	11.614	0.000	0.525	0.738
criticscore	-0.2225	0.149	-1.496	0.135	-0.514	0.069
months	1.6078	0.607	2.647	0.008	0.416	2.799
Adventure	17.1199	6.523	2.625	0.009	4.325	29.914
Comedy	-14.2784	6.699	-2.131	0.033	-27.419	-1.138
Drama	-0.6663	7.942	-0.084	0.933	-16.244	14.912
G	18.3626	20.650	0.889	0.374	-22.143	58.868
PG	-4.3948	5.951	-0.739	0.460	-16.068	7.278
directors_1	-14.2485	5.070	-2.810	0.005	-24.193	-4.304

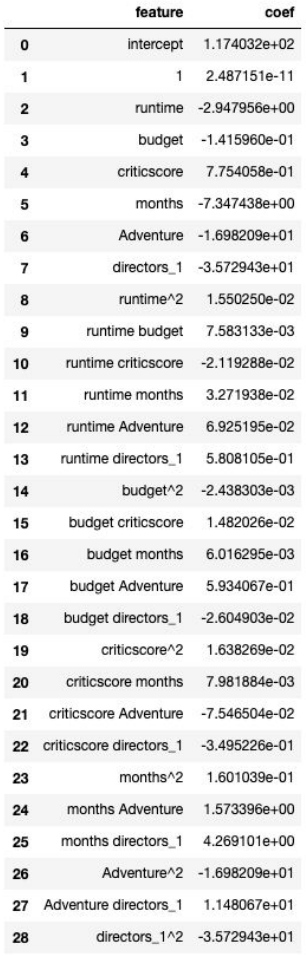
Omnibus:	809.317	**Durbin-Watson:**	2.218
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	34756.383
Skew:	1.789	**Prob(JB):**	0.00
Kurtosis:	26.159	**Cond. No.**	1.89e+03



	domestic_box_office	domestic_opening_weekend	runtime	budget	criticscore	months	international_box_office	Adventure	Comedy	Drama	G	PG	PG-13	directors_1
domestic_box_office	1	0.89	0.35	0.64	0.27	0.088	0.83	0.28	-0.15	-0.096	0.036	0.03	0.13	0.37
domestic_opening_weekend	0.89	1	0.27	0.65	0.11	0.0023	0.76	0.33	-0.18	-0.12	0.029	-0.008	0.16	0.33
runtime	0.35	0.27	1	0.38	0.4	0.11	0.39	0.12	-0.23	0.097	-0.094	-0.22	0.12	0.32
budget	0.64	0.65	0.38	1	0.095	0.074	0.7	0.48	-0.26	-0.14	0.026	0.059	0.17	0.37
criticscore	0.27	0.11	0.4	0.095	1	0.13	0.21	-0.014	-0.14	0.078	0.0032	-0.068	-0.092	0.15
months	0.088	0.0023	0.11	0.074	0.13	1	0.11	-0.051	0.022	-0.015	0.033	0.011	0.009	0.059
international_box_office	0.83	0.76	0.39	0.7	0.21	0.11	1	0.37	-0.24	-0.1	0.039	0.018	0.14	0.32
Adventure	0.28	0.33	0.12	0.48	-0.014	-0.051	0.37	1	-0.54	-0.34	0.034	0.13	0.018	0.16
Comedy	-0.15	-0.18	-0.23	-0.26	-0.14	0.022	-0.24	-0.54	1	-0.21	-0.0011	-0.047	0.065	-0.091
Drama	-0.096	-0.12	0.097	-0.14	0.078	-0.015	-0.1	-0.34	-0.21	1	-0.038	-0.069	0.062	-0.052
G	0.036	0.029	-0.094	0.026	0.0032	0.033	0.039	0.034	-0.0011	-0.038	1	-0.046	-0.11	-0.025
PG	0.03	-0.008	-0.22	0.059	-0.068	0.011	0.018	0.13	-0.047	-0.069	-0.046	1	-0.46	-0.024
PG-13	0.13	0.16	0.12	0.17	-0.092	0.009	0.14	0.018	0.065	0.062	-0.11	-0.46	1	0.11
directors_1	0.37	0.33	0.32	0.37	0.15	0.059	0.32	0.16	-0.091	-0.052	-0.025	-0.024	0.11	1

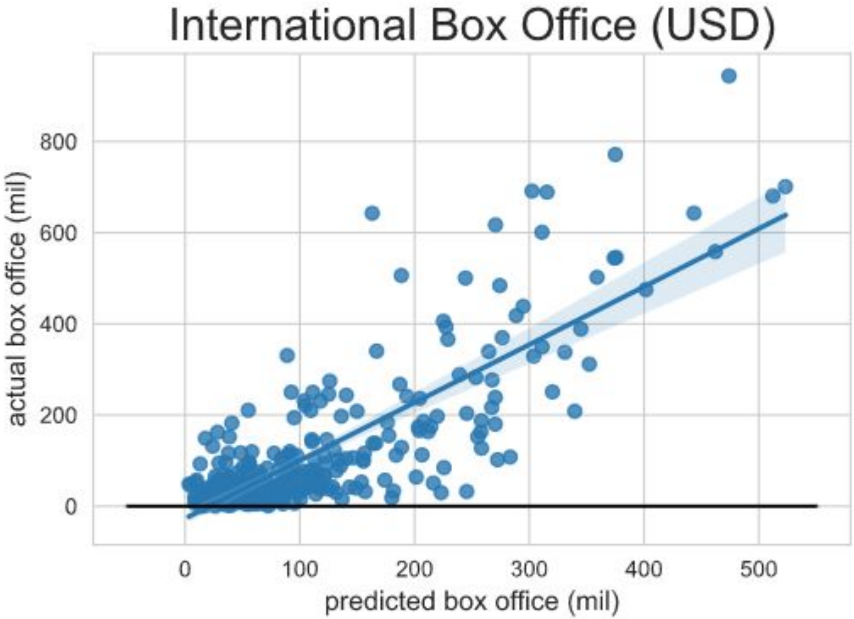
Appendix 2

final model regression formula



	feature	coef
0	intercept	1.174032e+02
1	1	2.487151e-11
2	runtime	-2.947956e+00
3	budget	-1.415960e-01
4	criticscore	7.754058e-01
5	months	-7.347438e+00
6	Adventure	-1.698209e+01
7	directors_1	-3.572943e+01
8	runtime^2	1.550250e-02
9	runtime budget	7.583133e-03
10	runtime criticscore	-2.119288e-02
11	runtime months	3.271938e-02
12	runtime Adventure	6.925195e-02
13	runtime directors_1	5.808105e-01
14	budget^2	-2.438303e-03
15	budget criticscore	1.482026e-02
16	budget months	6.016295e-03
17	budget Adventure	5.934067e-01
18	budget directors_1	-2.604903e-02
19	criticscore^2	1.638269e-02
20	criticscore months	7.981884e-03
21	criticscore Adventure	-7.546504e-02
22	criticscore directors_1	-3.495226e-01
23	months^2	1.601039e-01
24	months Adventure	1.573396e+00
25	months directors_1	4.269101e+00
26	Adventure^2	-1.698209e+01
27	Adventure directors_1	1.148067e+01
28	directors_1^2	-3.572943e+01

Appendix 3



final model regplot