

# Contributing Factors to a Movie's Overall Lifetime Domestic Gross



January 24, 2020

By: Joanna Nachman

# Motivation & End Goal

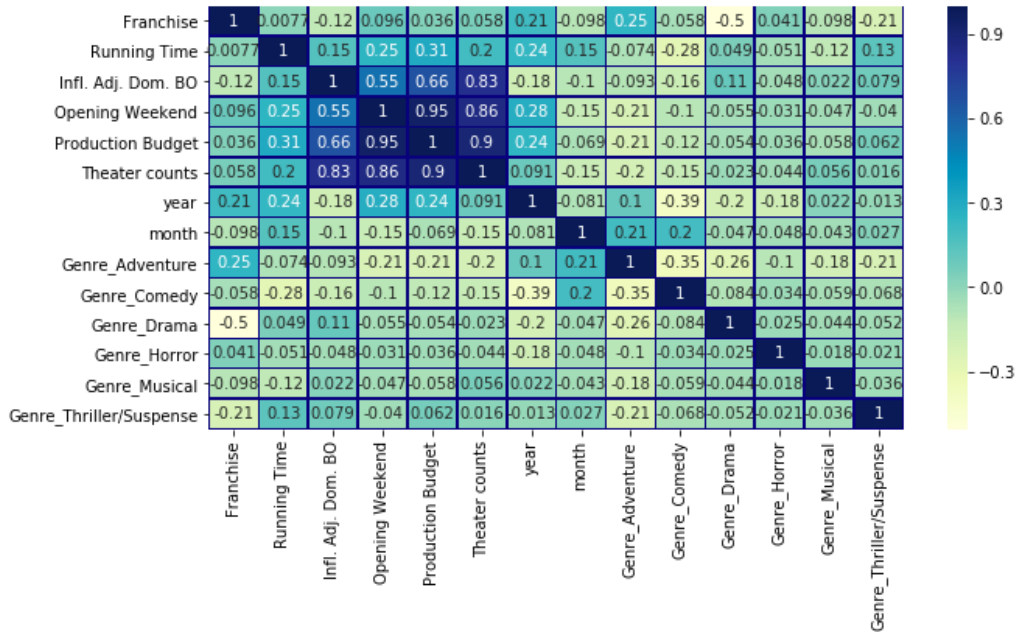
- Determine the most important factors for predicting a movie's overall lifetime gross
- Where should movie studios put their \$\$?

# Workflow

- Gather, scrape, clean movie data The Numbers
- Create baseline model, OLS, Regression
- Separate data to train, validate, and test

# Results

# Heat map



# Observations

- First iteration of LR:
  - $R^2 = 0.84$
- Simple Validation Method: Train / Validation / Test:
  - Linear Regression val  $R^2$ : 0.907
  - Ridge Regression val  $R^2$ : 0.915
  - Degree 2 polynomial regression val  $R^2$ : 1.000

# Observations

- Cross-Validation / Test:
  - Simple mean cv  $r^2$ : 0.677  $\pm$  0.000
  - Ridge mean cv  $r^2$ : 0.707  $\pm$  0.000
- Did better on the same test set:
  - Ridge Regression test  $R^2$ : 0.833



# Observations

- Feature Engineering:
  - Validation  $R^2$  score was: 0.755
  - Added interaction terms:
    - Production Budget & Franchise






# Observations

- OLS Regression Results
- Important features:
  - Theater counts\*
  - Opening weekend
  - Running Time
  - Production Budget
  - Year
  - Genres: Adventure & Drama

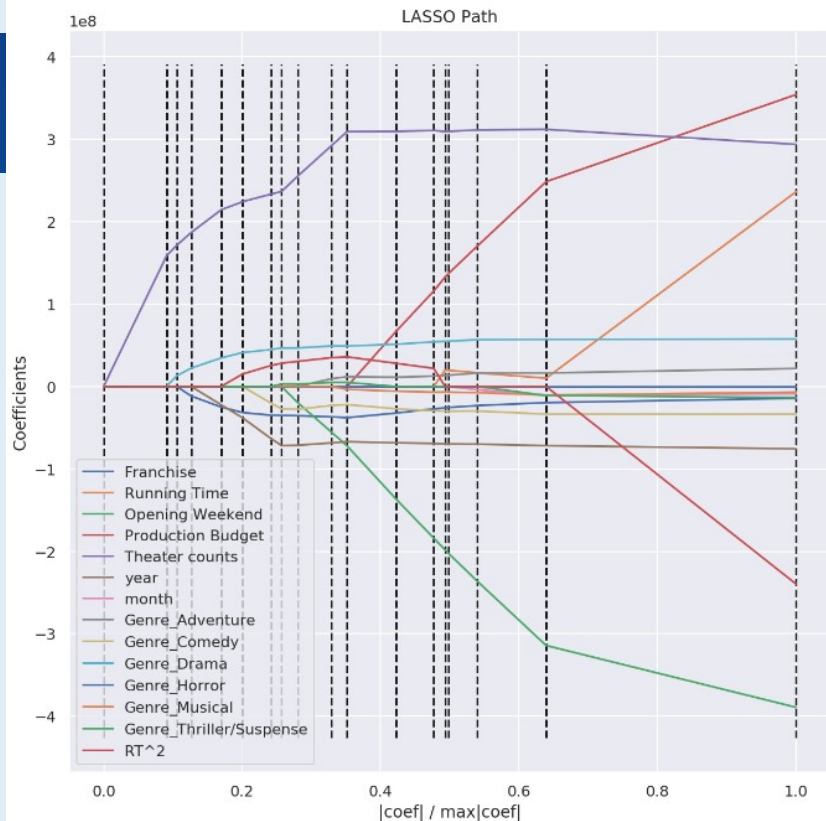
Dep. Variable:	Infl. Adj. Dom. BO	R-squared:	0.845
Model:	OLS	Adj. R-squared:	0.797
Method:	Least Squares	F-statistic:	17.54
Date:	Fri, 24 Jan 2020	Prob (F-statistic):	9.66e-14
Time:	00:45:22	Log-Likelihood:	-1206.7
No. Observations:	60	AIC:	2443.
Df Residuals:	45	BIC:	2475.
Df Model:	14		
Covariance Type:	nonrobust		



	coef	std err	t	P> t	[0.025	0.975]
const	1.222e+10	4.54e+09	2.694	0.010	3.08e+09	2.14e+10
Franchise	-2.901e+07	7.27e+07	-0.399	0.692	-1.75e+08	1.17e+08
Running Time	9.164e+06	1.05e+07	0.876	0.386	-1.19e+07	3.02e+07
Opening Weekend	-5.7654	1.585	-3.637	0.001	-8.958	-2.572
Production Budget	4.5698	1.790	2.552	0.014	0.964	8.176
Theater counts	2.344e+05	5.32e+04	4.404	0.000	1.27e+05	3.42e+05
year	-6.372e+06	2.27e+06	-2.803	0.007	-1.1e+07	-1.79e+06
month	1.796e+05	9.2e+06	0.020	0.985	-1.84e+07	1.87e+07
Genre_Adventure	4.713e+07	6.19e+07	0.761	0.451	-7.76e+07	1.72e+08
Genre_Comedy	-1.137e+08	9.99e+07	-1.138	0.261	-3.15e+08	8.75e+07
Genre_Drama	2.475e+08	1.18e+08	2.090	0.042	9e+06	4.86e+08
Genre_Horror	-4.889e-07	2.74e-07	-1.782	0.081	-1.04e-06	6.35e-08
Genre_Musical	-2.389e+07	1.1e+08	-0.216	0.830	-2.46e+08	1.99e+08
Genre_Thriller/Suspense	-4.778e+07	9.96e+07	-0.480	0.634	-2.48e+08	1.53e+08
RT^2	-3.494e+04	4.06e+04	-0.860	0.394	-1.17e+05	4.69e+04
Genre_Musical_Clone	-8754.2591	9117.389	-0.960	0.342	-2.71e+04	9609.104

# Observations

- LARS Path to Study Feature Importance:
  - Theater counts
  - Genre: Drama



# Conclusions/ Recommendations

- Theater counts
- Genres: adventure and drama
- Opening weekend
- Running Time
- Production Budget



Thank you!

# References

- The Numbers: Movie Budgets. 1997-2020 Nash Information Services, LLC. Retrieved from:
  - <https://www.the-numbers.com/movie/budgets/all>