

# Movies & Beyond

Metis Spring 2018

Jacob Levine

# Problem

- Do movies associated with a more decorated cast/director typically perform better in terms of domestic gross total?
- How does the predictive value of 'star power' features compare to other features such as budget, rating, etc?

# Sources and Methods

- Websites Scraped:
  - BoxOfficeMojo
  - Oscars.org
- Tools used:
  - BeautifulSoup
  - Scikit-learn
  - Statsmodels



# Assumptions

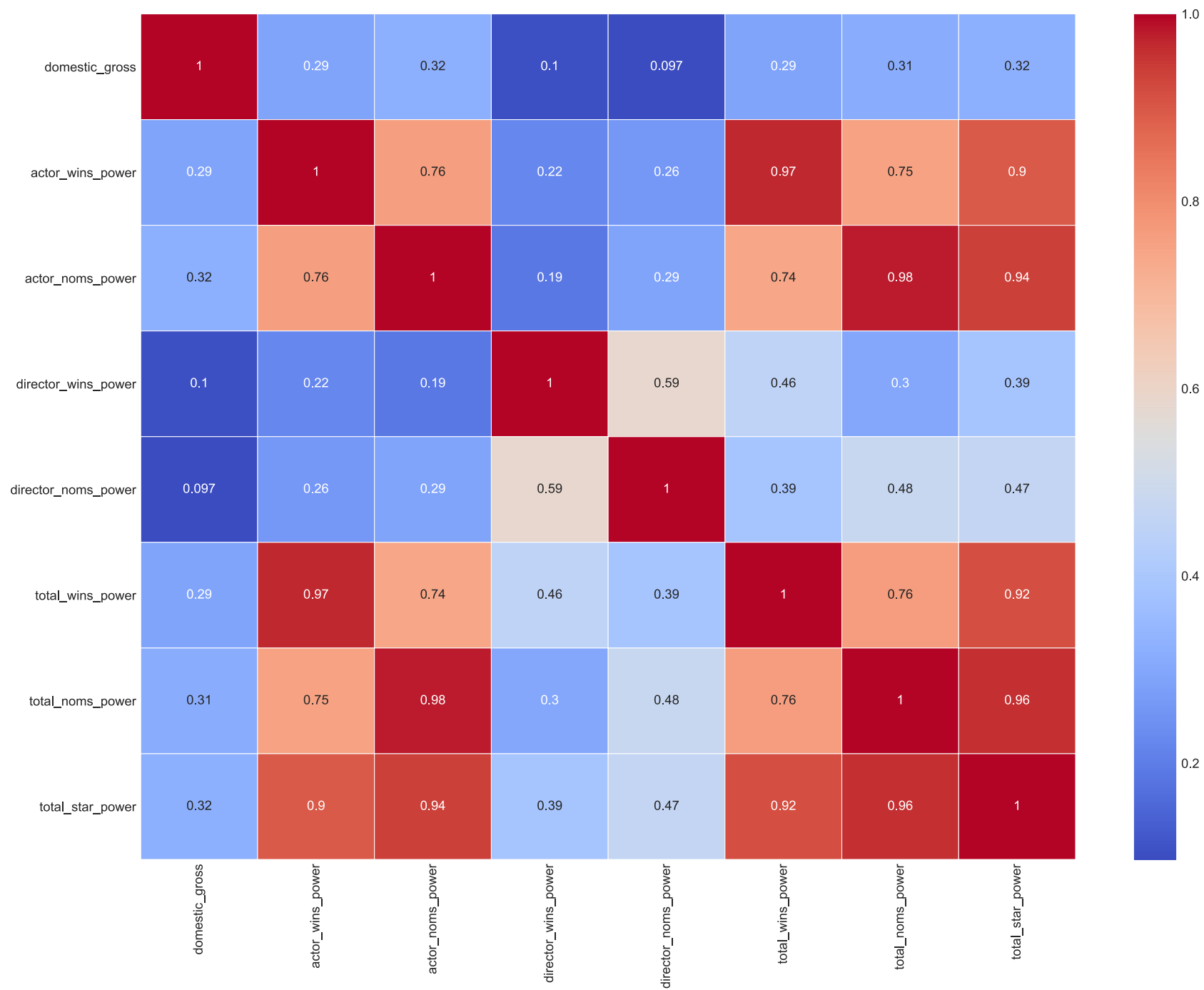
- Movies that perform better in the box office tend to feature more Oscar award winning/ Oscar nominated actors/director
- Oscar wins hold more weight than Oscar nominations
- Movies with higher domestic gross totals tend to have more information about them readily available on the internet

# Analysis – Part 1

Type	Points
Actor Win	2.5
Actor Nomination	1
Director Win	2.5
Director Nomination	1

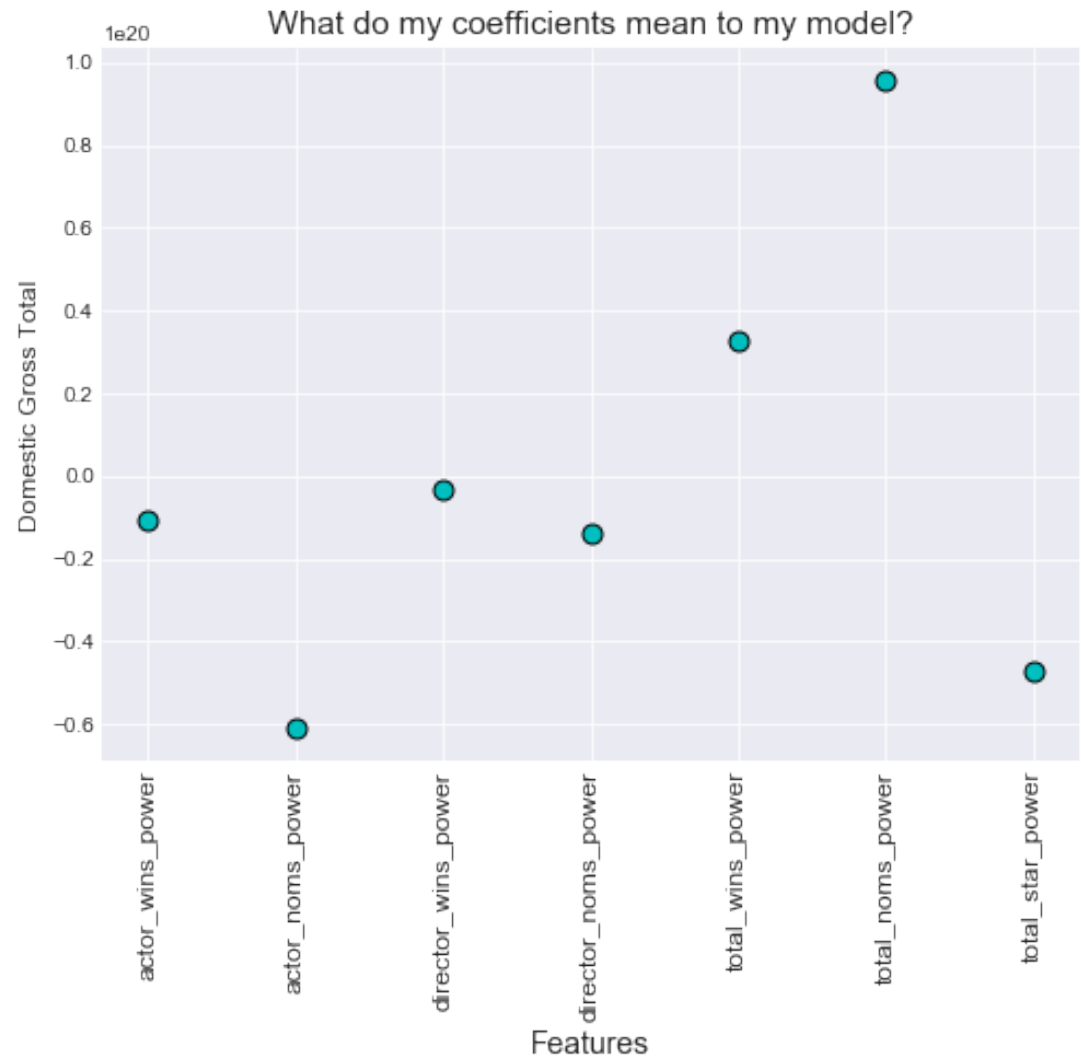
n = 6,103





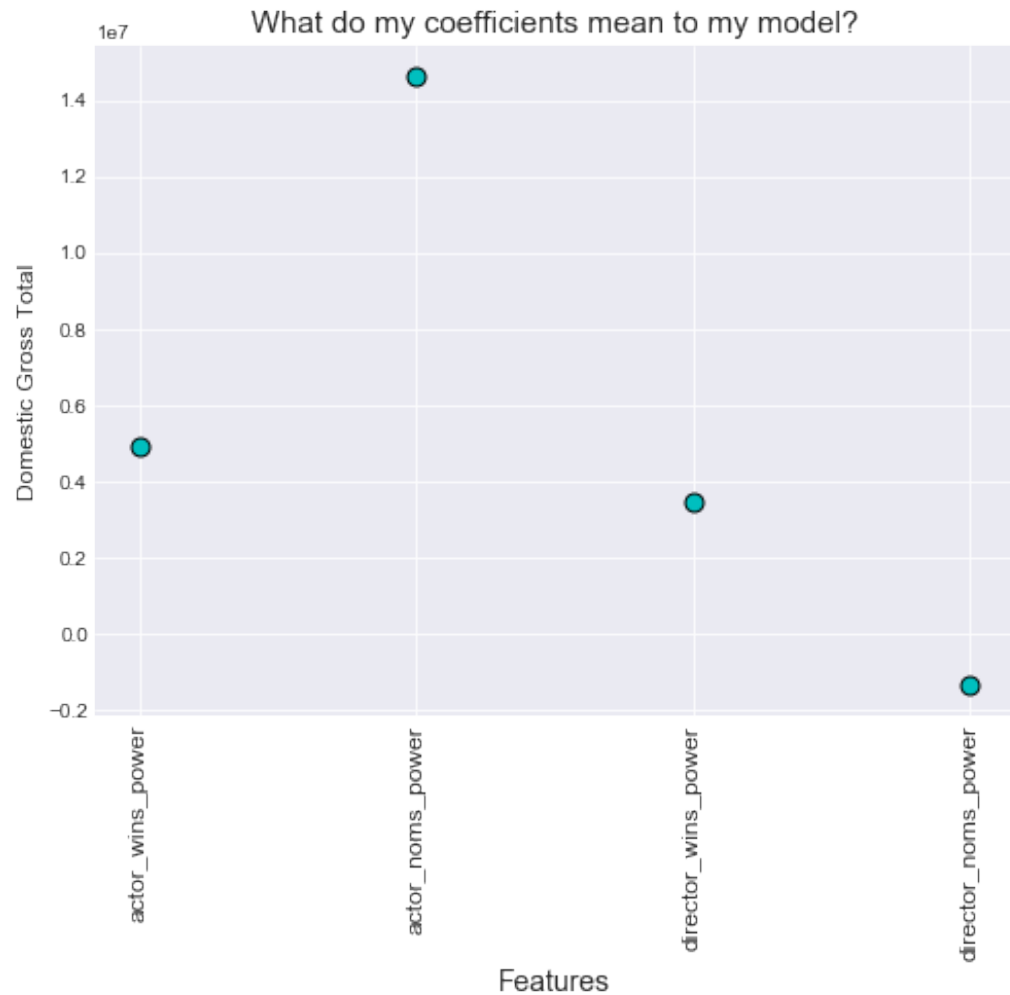
# Model Interpretation

- Test  $R^2$  – 0.127
- Total Nominations Power is the most telling feature



# Model Interpretation

- Test  $R^2$  – 0.127
- Total Nominations Power broken down
- Actor Nominations Power is up





# StatsModels

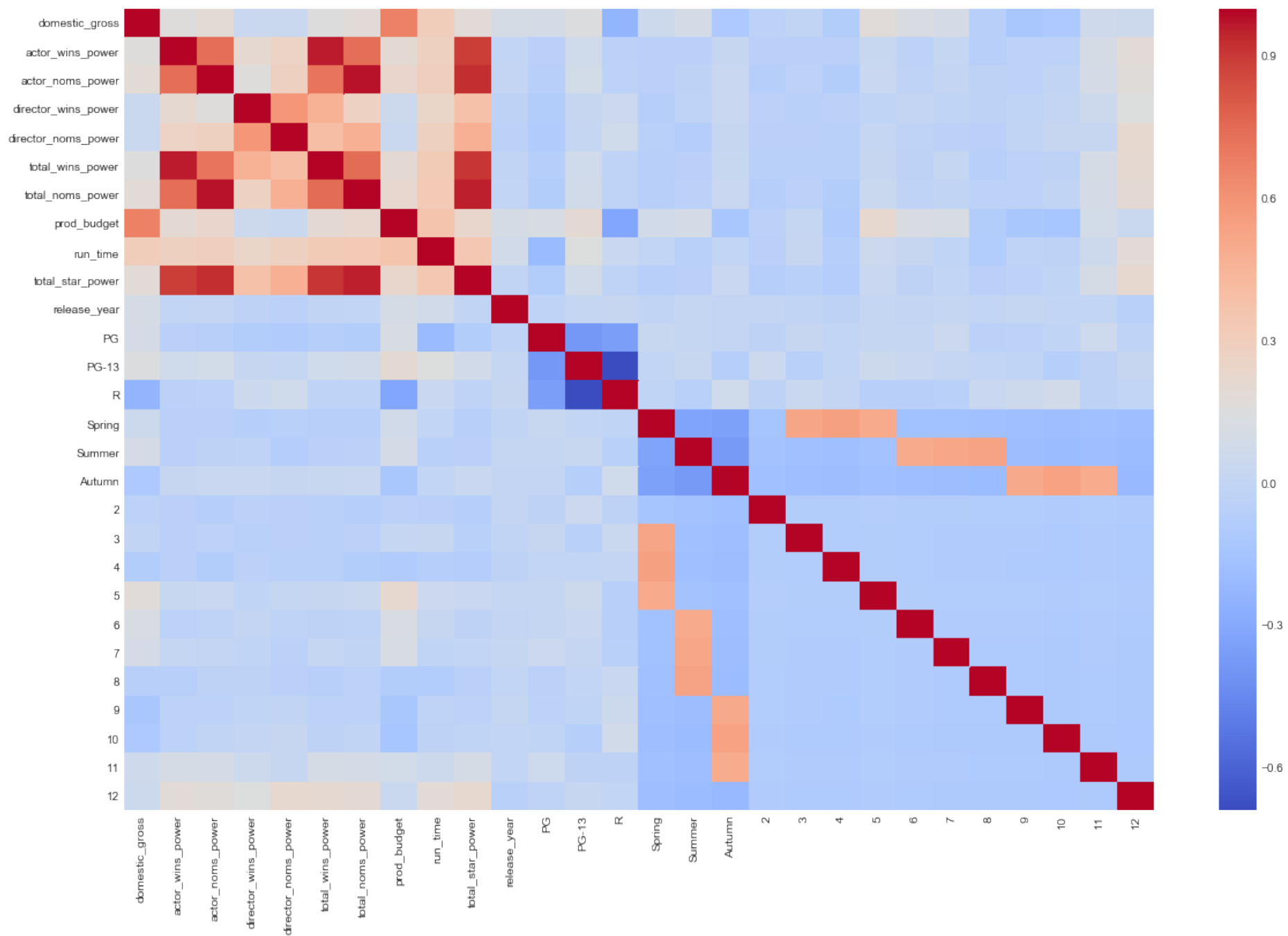
<b>Dep. Variable:</b>	y_train	<b>R-squared:</b>	0.103
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.103
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	123.1
<b>Date:</b>	Thu, 26 Apr 2018	<b>Prob (F-statistic):</b>	1.53e-99
<b>Time:</b>	17:20:51	<b>Log-Likelihood:</b>	-82298.
<b>No. Observations:</b>	4272	<b>AIC:</b>	1.646e+05
<b>Df Residuals:</b>	4267	<b>BIC:</b>	1.646e+05
<b>Df Model:</b>	4		
<b>Covariance Type:</b>	nonrobust		

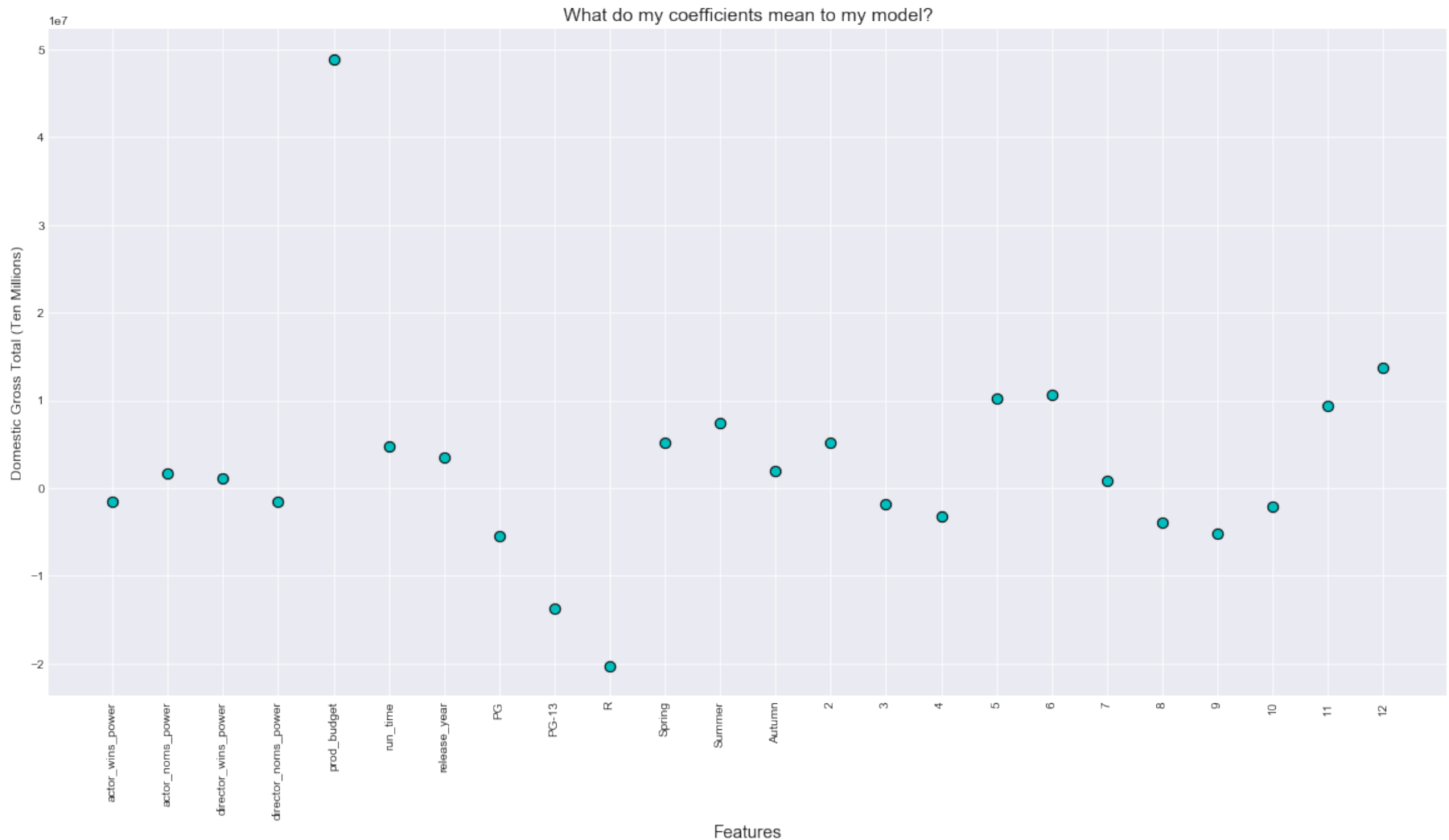
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.643e+07	8.61e+05	30.687	0.000	2.47e+07	2.81e+07
X_scaled[0]	4.898e+06	1.31e+06	3.730	0.000	2.32e+06	7.47e+06
X_scaled[1]	1.464e+07	1.31e+06	11.161	0.000	1.21e+07	1.72e+07
X_scaled[2]	3.467e+06	1.08e+06	3.209	0.001	1.35e+06	5.58e+06
X_scaled[3]	-1.365e+06	1.1e+06	-1.242	0.214	-3.52e+06	7.89e+05

<b>Omnibus:</b>	4221.834	<b>Durbin-Watson:</b>	1.959
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	288567.178
<b>Skew:</b>	4.727	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	42.138	<b>Cond. No.</b>	2.96

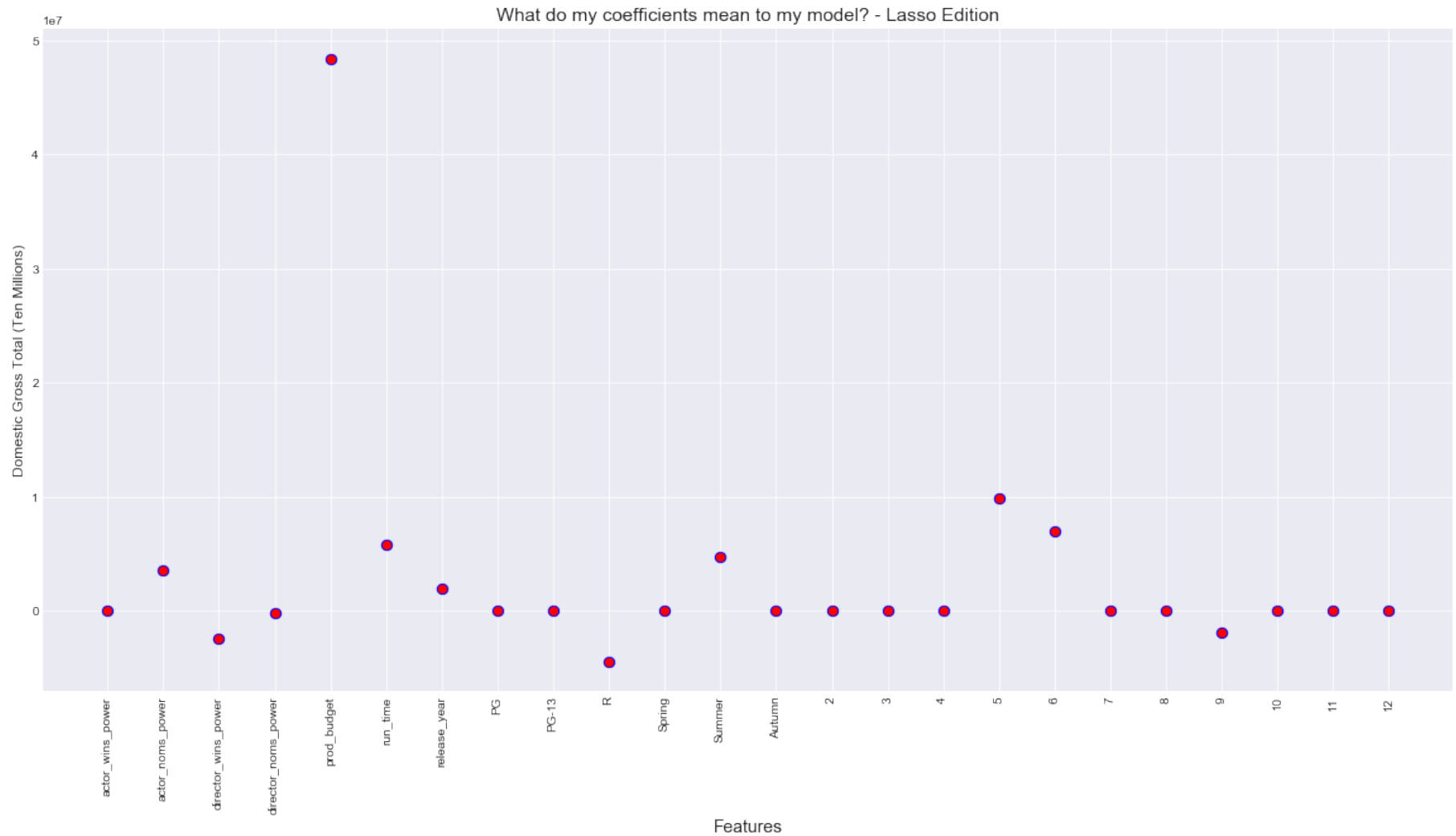
# Analysis – Part 2

- Increased number of features
- Comparison between star power features and non star power related features
- Exclusion of nulls (budget, runtime, release date, titles with foreign genre) down to 2,001 samples





Test  $R^2$  – 0.492 | Star Power Related features take backseat to Production Budget and certain seasonal features, there is still predictive power in Actor Nominations, Director Wins



Test  $R^2$  – 0.428 | Utilizing Lasso - Actor Nomination Power, runtime, certain seasonal features and of course budget hold weight.

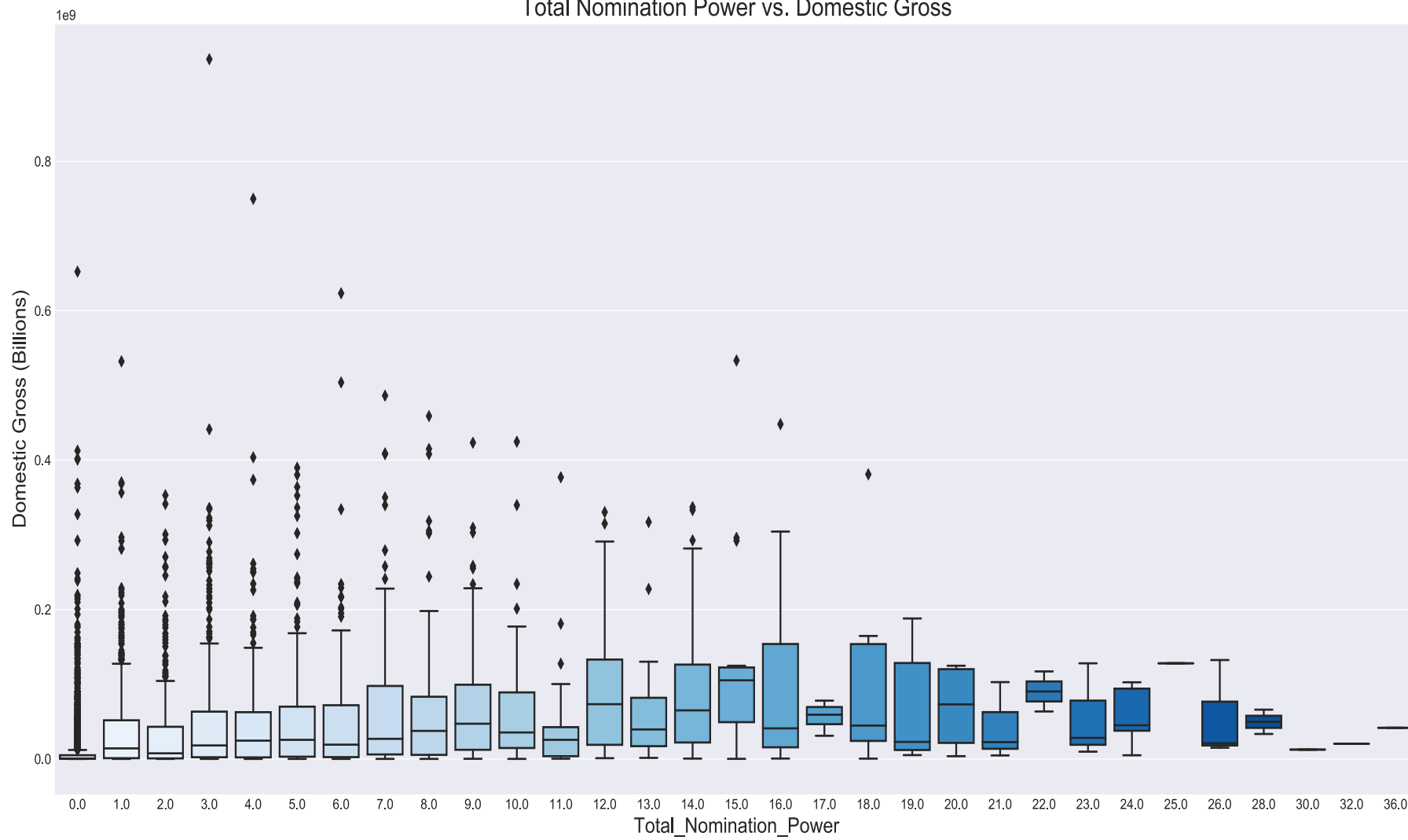
# Results

- Compared to Production Budget, star power features yield little predictive power. However, compared to other features such as runtime, year of release and season, certain star power features, particularly actor nominations power, hold their own
- Rating holds less power than I anticipated, late spring – early summer is best release period

# Results

- Movies that feature Oscar winning and/or Oscar nominated cast/directors do tend to outperform movies that do not, however, more is not always better...

Total Nomination Power vs. Domestic Gross





# Next Steps

- More Data!
- Many nulls in cast, production budget, genre, etc.
- Scrape titles from a wider range of years
- Experiment further with feature engineering and transformations

Thank You