



PREDICTING MOVIE PIRACY

REGRESSION & STUFF

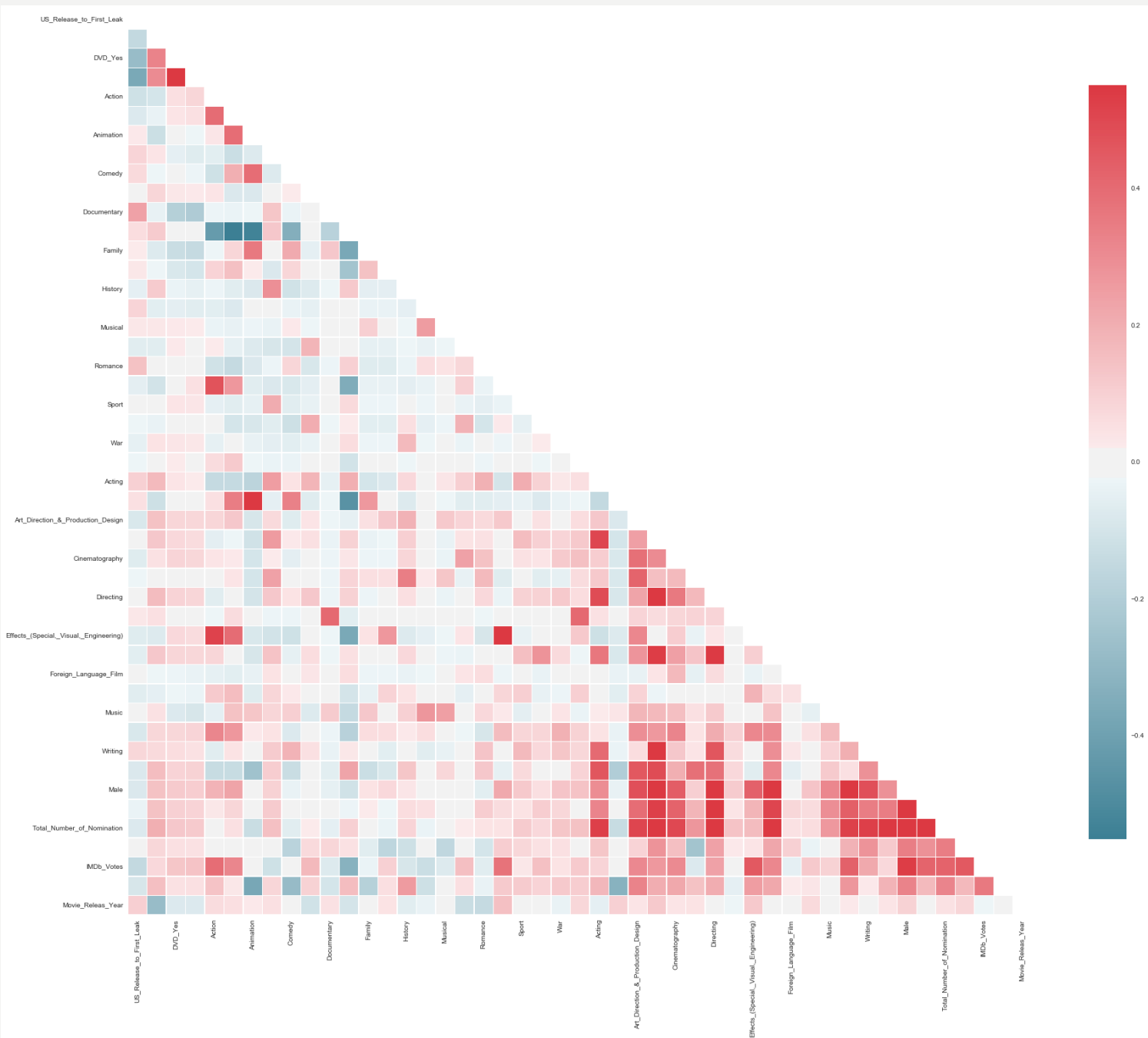
WHY IS IT IMPORTANT?

- Piracy leads to significantly lower revenues than estimated
- Make better estimations of revenue with predictions on when a leak may occur
- Become proactive about when a leak will happen and seek out preventive measures

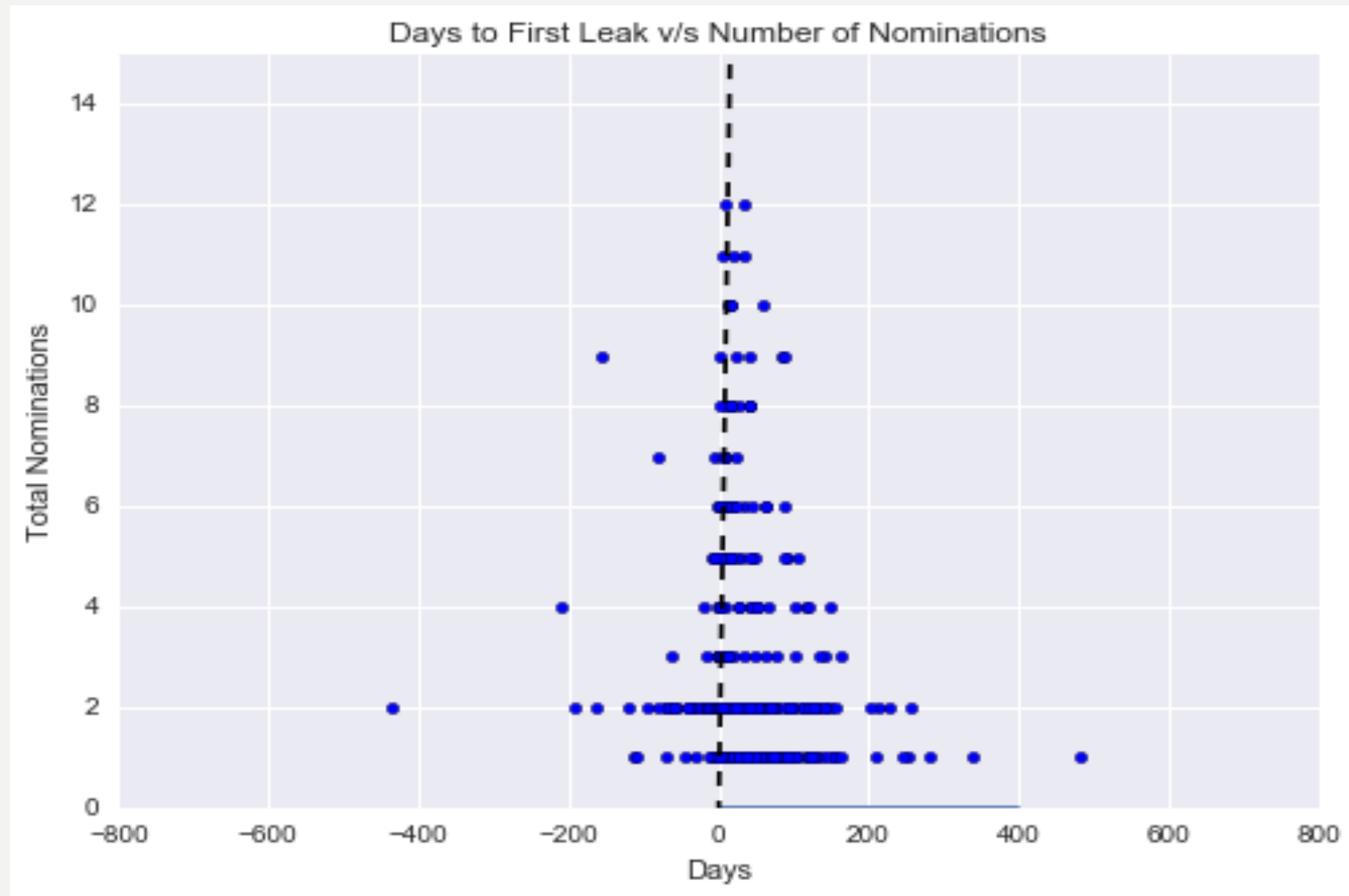
THE DATA

- Piracy Data on 650 Oscar Nominated movies from 2013-2015
- Oscar Data on Nominations, Genre, Type of Award, Sex of Nominee etc.
- Predictor Variable: [US Release to First Leak]
- 55 Variable features including Dummies

UNDERSTANDING VARIABLE CORRELATIONS



SIGNIFICANCE OF NUMBER OF NOMINATIONS



OLS REGRESSION WITH POLYNOMIALS

Adj. R square: 0.145

RMSE: 61.76

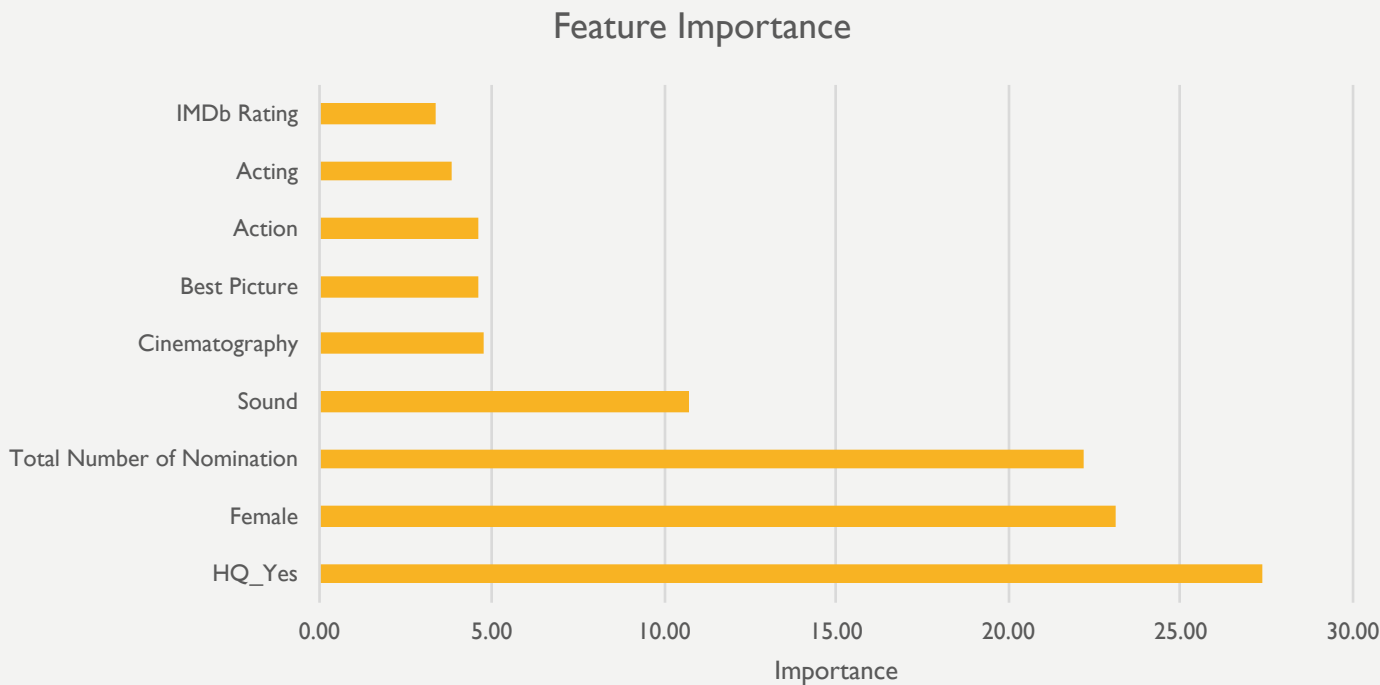
Variable	Coefficient	p-value
Intercept	25.11	0.00
DVD_Yes	36.27	0.00
HQ_Yes	-50.77	0.00
Documentary	9.90	0.00
Drama	13.11	0.00
Mystery	-5.15	0.12
Romance	6.45	0.06
SciFi	5.05	0.15
War	-5.81	0.09
Acting	11.20	0.01
Directing	6.74	0.22
Sound	12.04	0.01
Writing	27.65	0.00
I(Writing^2)	-4.78	0.03
Number of Wins	18.23	0.03
I(Number of Wins ^2)	-3.70	0.08
Total Number of Nominations	-53.79	0.00
I(Total Number of Nomination^2)	7.48	0.01
Movie Release Year	3.43	0.10

...ON TO MACHINE LEARNING

Gradient Boosting Regression offers the best scores and highest R-square Values

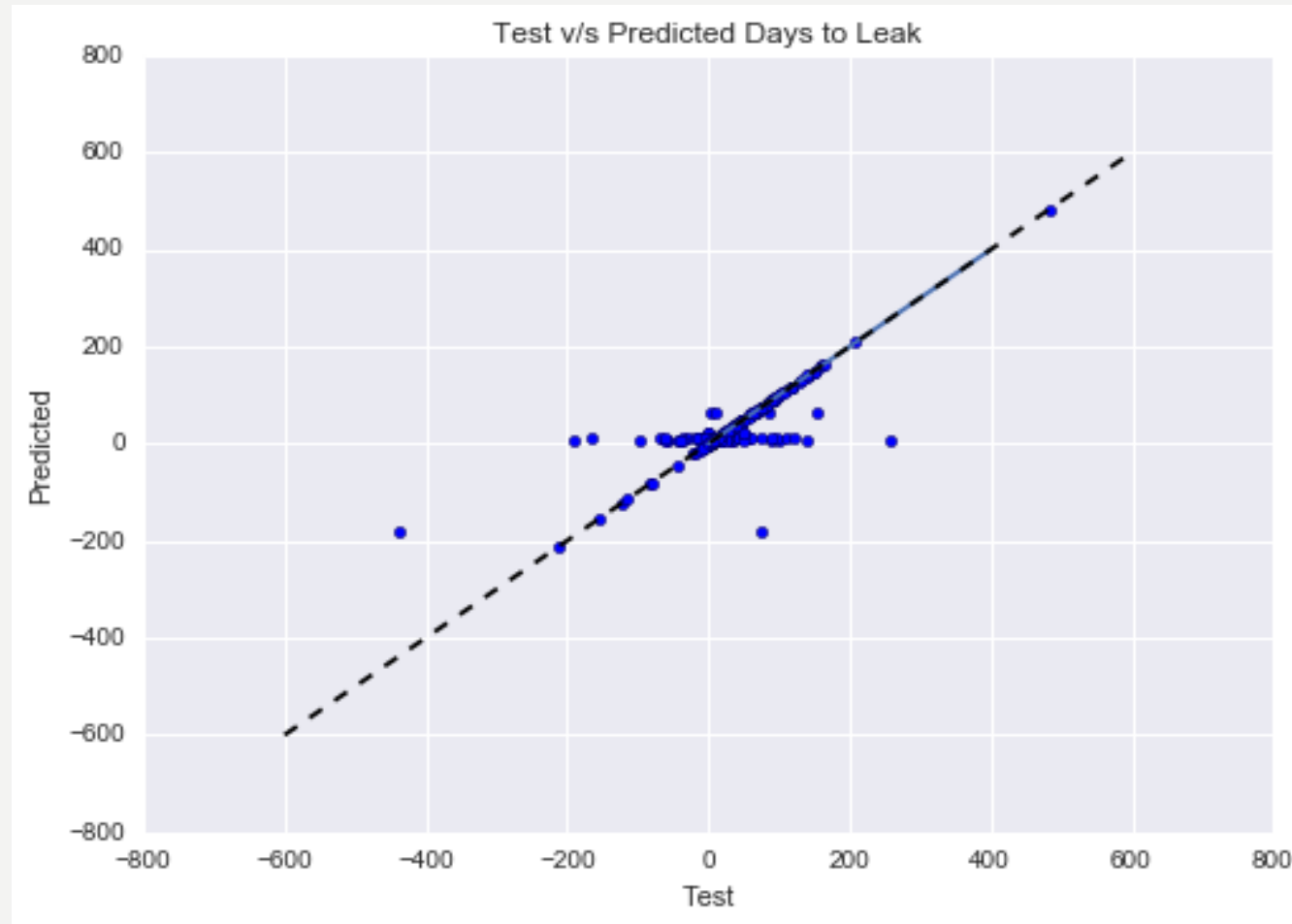
R-Square: **0.68**

RMSE: **36.6**

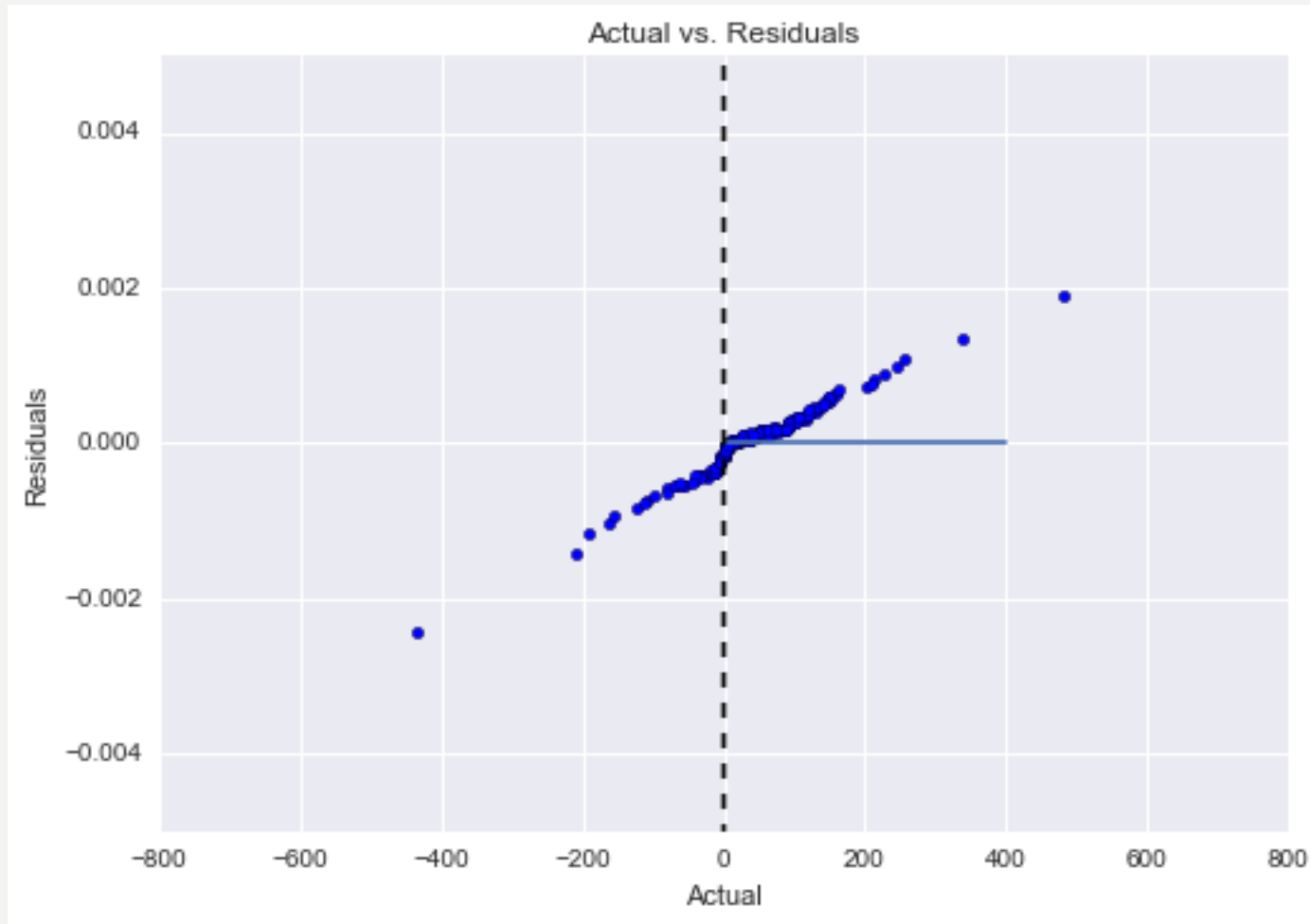


5 K-Fold Cross Validation Score: **0.94**

PREDICTED V/S ACTUAL



RESIDUALS V/S ACTUAL



IMPORTANT FEATURES

- Number of Nominations
- Number of Wins
- Type of Awards (Sound / Cinematography/ Acting)
- Sex of Nominee
- Keep an eye out for Action/War type films

FUTURE WORK

- Add variables on Budgets, Expected ROI, Seasonality
- Get information on movie piracy by hacker site (success rate)
- Add features on movie companies to predict company specific leaks
- Calculate potential Revenue lost by Leak for a given movie based on models