

Predicting MLB Salaries



Jason Salazer-Adams



Can rookie stats predict a player's salary in 6 years?

Contract negotiations

Retain talent

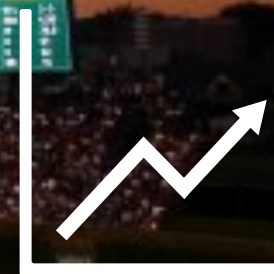


Rookie stats 1985 - 2017



Filtered to only batters and
played for at least 6 seasons

Built linear models to predict
salary in the 6th year

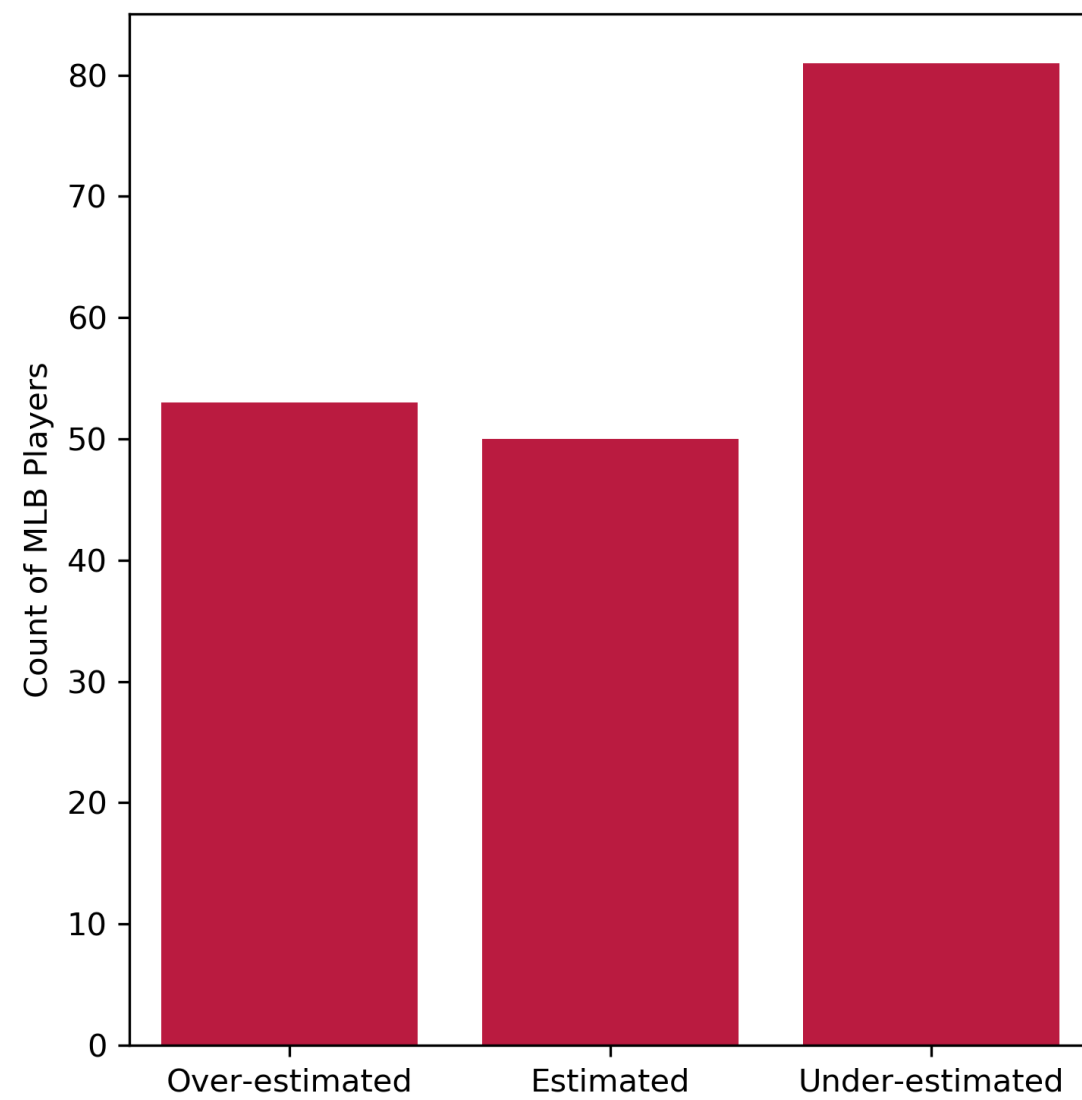




Los Angeles Times

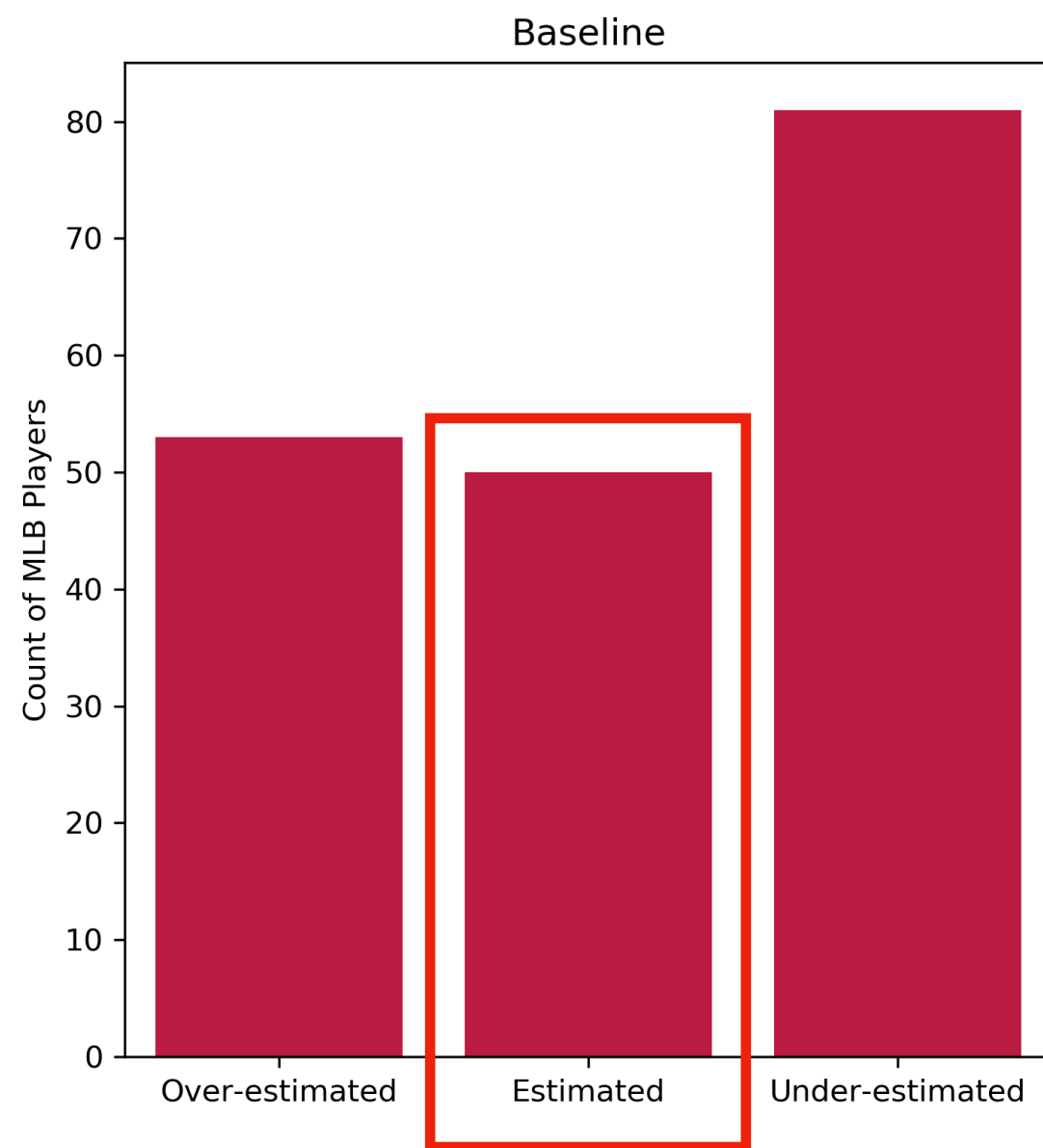
6th-Year Salary Prediction Accuracy

Baseline



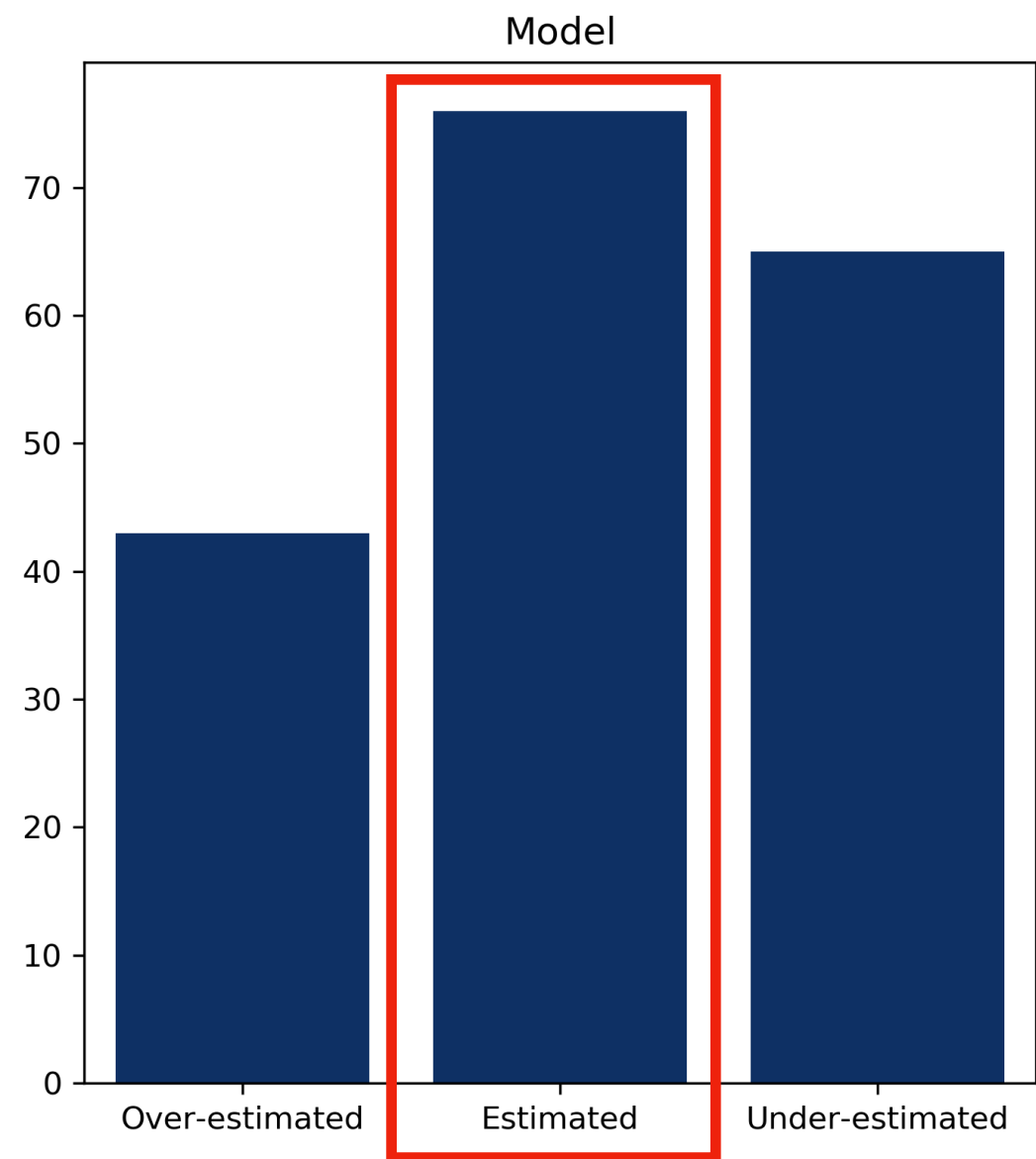
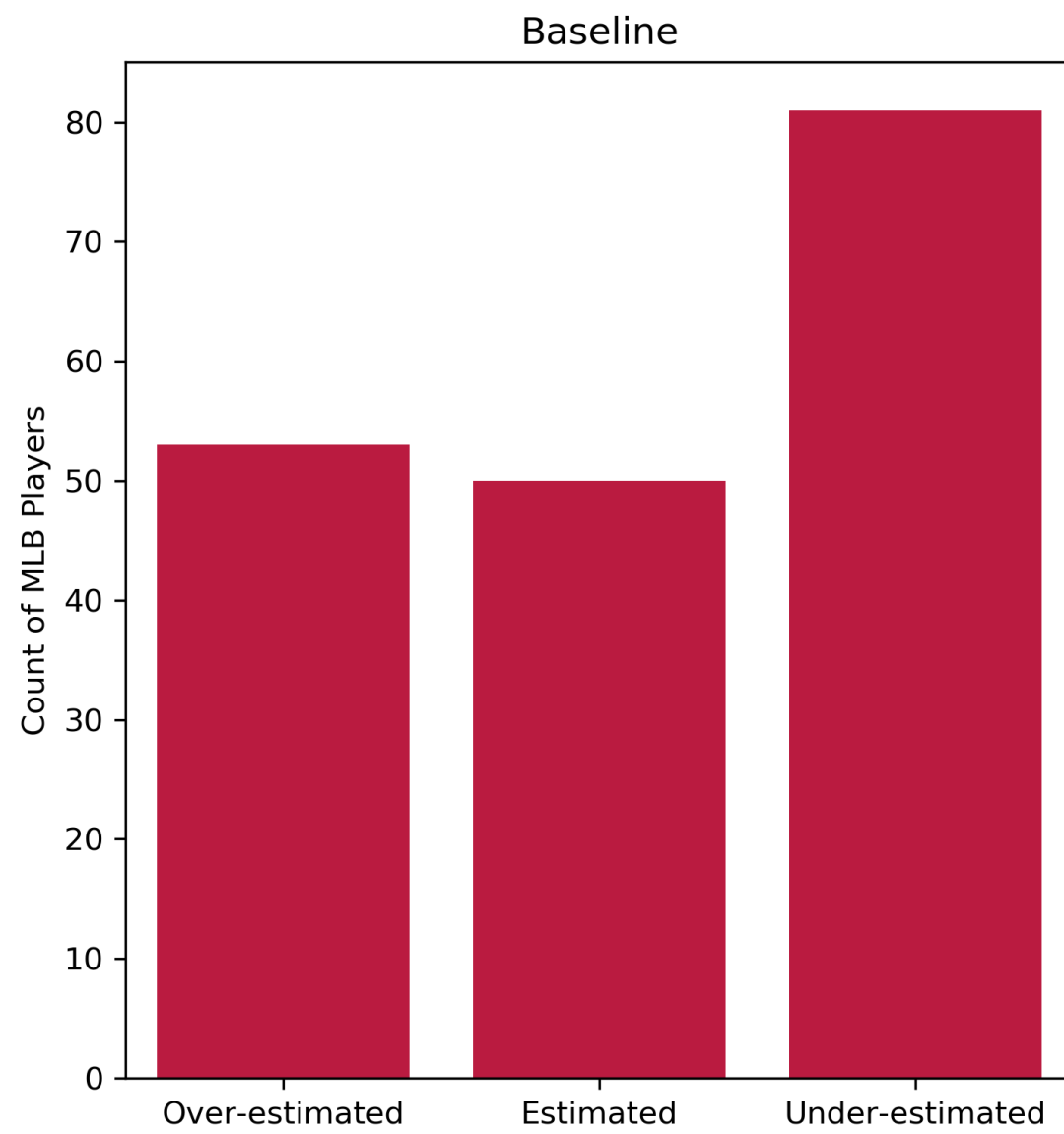


6th-Year Salary Prediction Accuracy





6th-Year Salary Prediction Accuracy



Future Work



Consider saber metric data

Build a model by position

Consider first 2 season in MLB

An aerial photograph of Safeco Field, a baseball stadium. The stadium is filled with green seats, and the field is visible in the center. The roof is a complex, metallic structure. Various advertisements are visible on the outfield walls, including "GoMariners", "ROOT SPORTS", and "Holland America Line". The text "Thank You" is overlaid in the center of the field.

Thank You

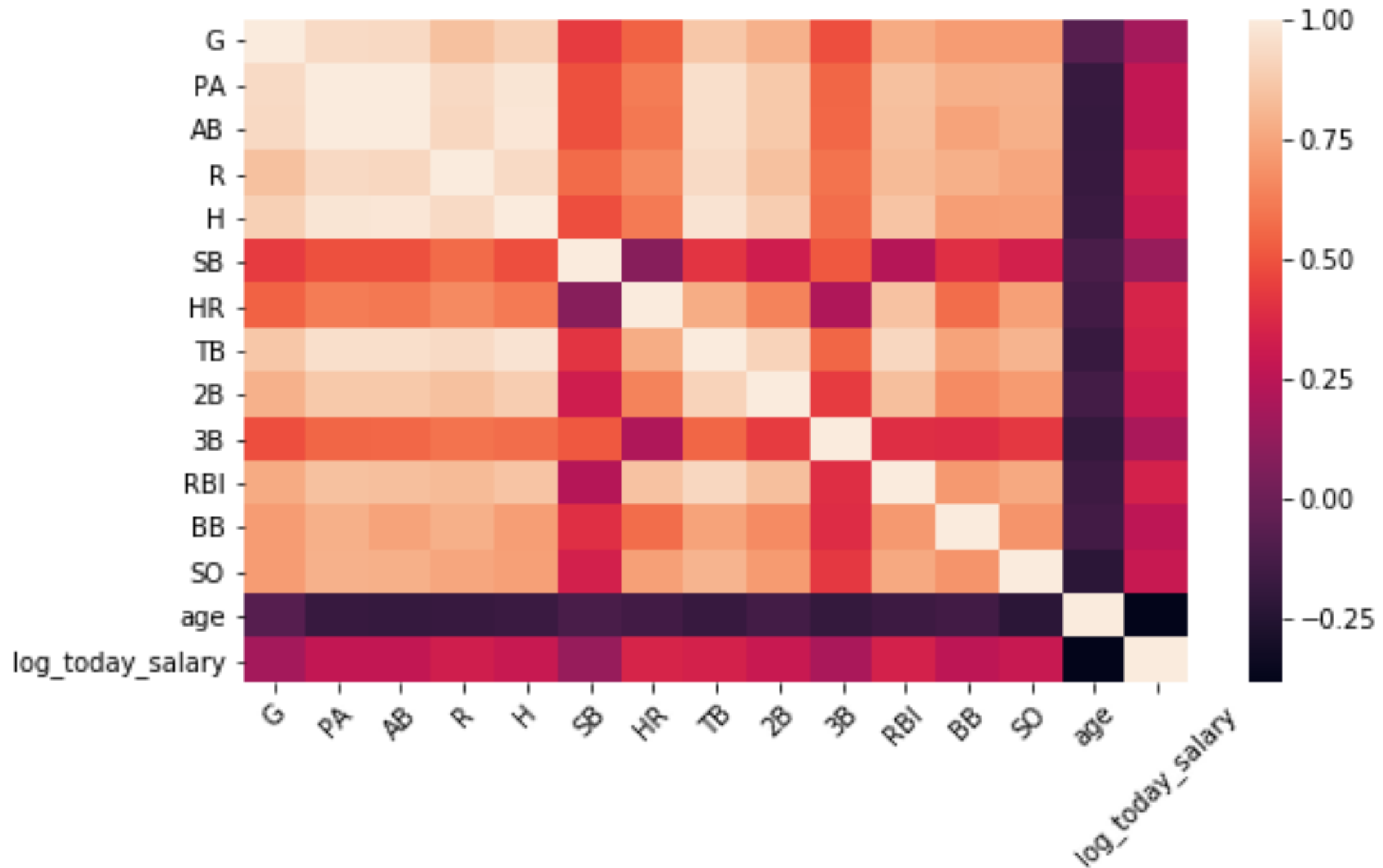
SAFECO FIELD

Appendix

Train Data Statistics

	G	PA	AB	R	H	SB	HR	TB	2B	3B	RBI	BB	SO	age
count	735	735	735	735	735	735	735	735	735	735	735	735	735	735
mean	87	308	277	36	73	6	7	112	14	2	32	24	52	24
std	36	160	144	23	42	9	6	66	9	2	21	16	31	2
min	3	9	8	0	0	0	0	0	0	0	0	0	1	19
25%	57	179	161.5	19.5	41	1	2	61	7	0	16.5	12	29	23
50%	85	275	244	31	64	3	5	95	12	2	27	20	46	24
75%	115	429.5	384	51	101	8	9	154.5	19	3	44	32	70	25
max	162	734	684	122	209	110	37	365	47	12	130	91	185	32

Correlation Matrix

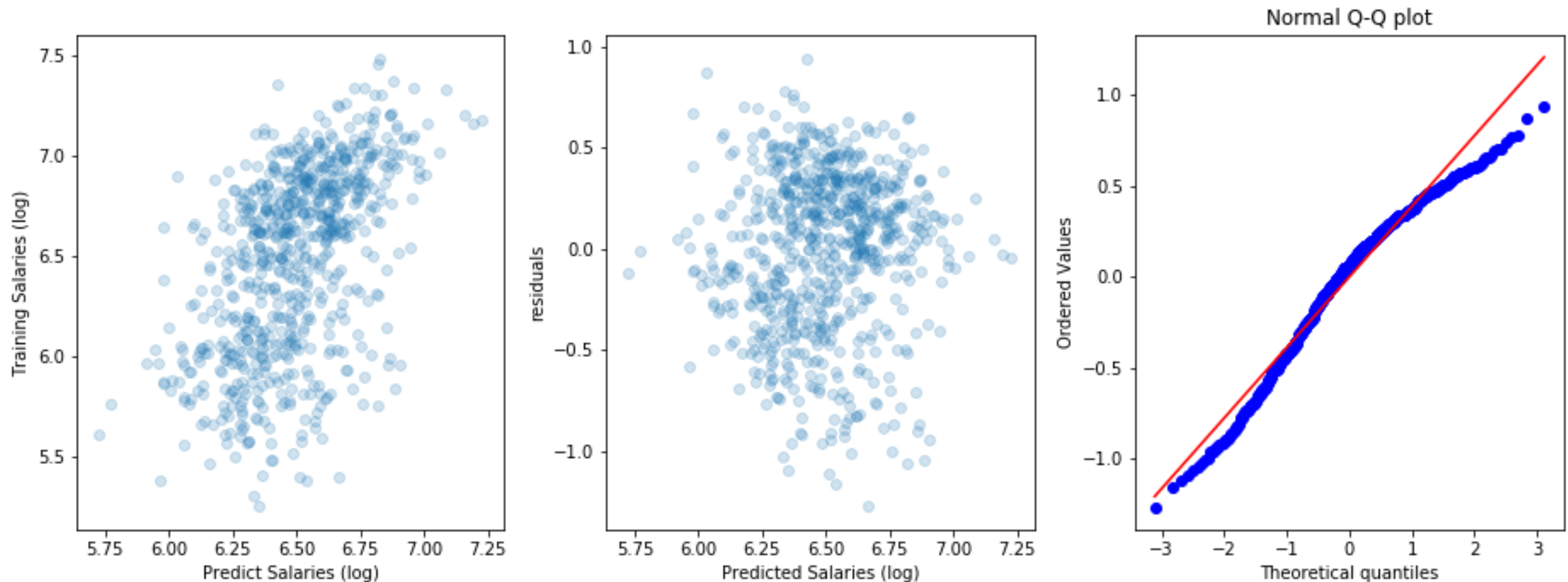


Model Selection

- Utilized 10-fold cross validation comparing standard OLS with Lasso regularization.
- Features were standardized and polynomial features were added to find the best general fit.

Run	Model	MSE
0	lasso poly 3 cv - 0.01	0.161218
1	lasso poly 2 cv - 0.01	0.161348
2	lasso poly 4 cv - 0.01	0.161483
3	lasso poly 5 cv - 0.01	0.162328
4	lasso cv - 0.01	0.162541
5	linear model - cv10	0.163087

Best Model Goodness of Fit



Best Model Coefficients

	Coef
age	-0.174874
G^2	-0.034479
G age	-0.028781
R age^2	0.032096
SB^2 HR	0.009415
SB 3B BB	0.01288
SB BB^2	0.007684
HR age^2	0.098267
TB age^2	0.024179
2B age^2	0.026535
3B BB age	0.005021
RBI age^2	0.01861
SO^3	-0.013046