PROJECT LUTHER:

Explaining the Mariners' 16-Year Playoff Drought

Kelly Jones | Metis | April 2018

"You've been traded for some big data,
two spreadsheets, and an algorithm."

229 rows of data

Years:
2002 to 2017

Teams:
American League (N=15)

Option 1: A Clear but Useless Model

Features: Avg runs scored, Avg runs given up

Dep. Variable:	pctwon	R-squared:	0.858
Model:	OLS	Adj. R-squared:	0.856
Method:	Least Squares	F-statistic:	473.6
Date:	Thu, 26 Apr 2018	Prob (F-statistic):	3.17e-67
Time:	19:50:07	Log-Likelihood:	348.15
No. Observations:	160	AIC:	-690.3
Df Residuals:	157	BIC:	-681.1
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.5219	0.026	19.751	0.000	0.470	0.574
R/G	0.0967	0.004	21.953	0.000	0.088	0.105
RA/G	-0.1011	0.004	-24.030	0.000	-0.109	-0.093

Omnibus:	3.314	Durbin-Watson:	2.001
Prob(Omnibus):	0.191	Jarque-Bera (JB):	3.190
Skew:	0.345	Prob(JB):	0.203
Kurtosis:	2.962	Cond. No.	80.6

Features: RBI, ERA, Saves, Fielding %

Dep. Variable:	pctwon	R-squared:	0.923
Model:	OLS	Adj. R-squared:	0.921
Method:	Least Squares	F-statistic:	462.6
Date:	Thu, 26 Apr 2018	Prob (F-statistic):	4.87e-85
Time:	19:51:31	Log-Likelihood:	393.38
No. Observations:	160	AIC:	-776.8
Df Residuals:	155	BIC:	-761.4
Df Model:	4		
Covariance Type:	nonrobust		

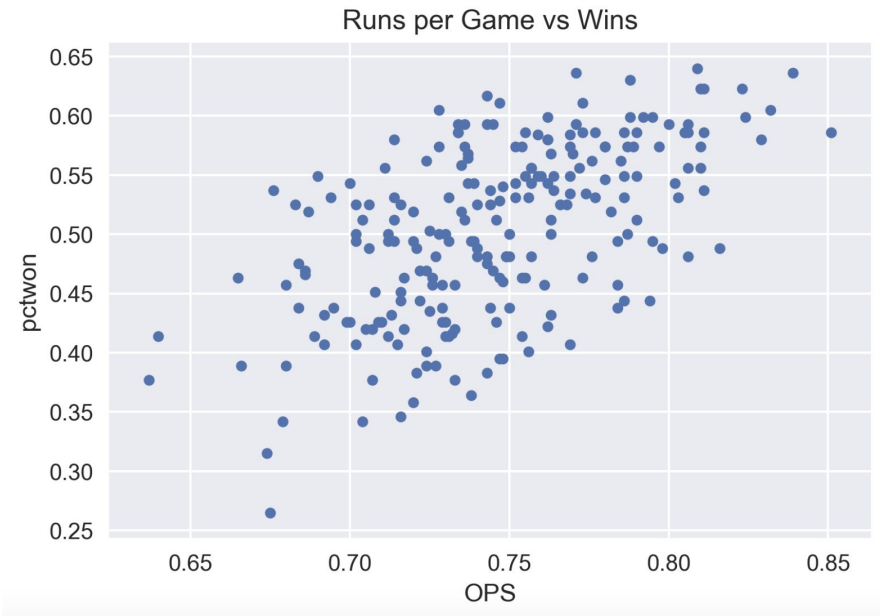
	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.2482	0.743	-3.024	0.003	-3.717	-0.780
RBI	0.0005	2.07e-05	26.347	0.000	0.001	0.001
ERA	-0.0755	0.004	-19.59	0.000	-0.083	-0.068
SV	0.0026	0.000	10.32	0.000	0.002	0.003
Fld%	2.6199	0.756	3.466	0.001	1.127	4.113

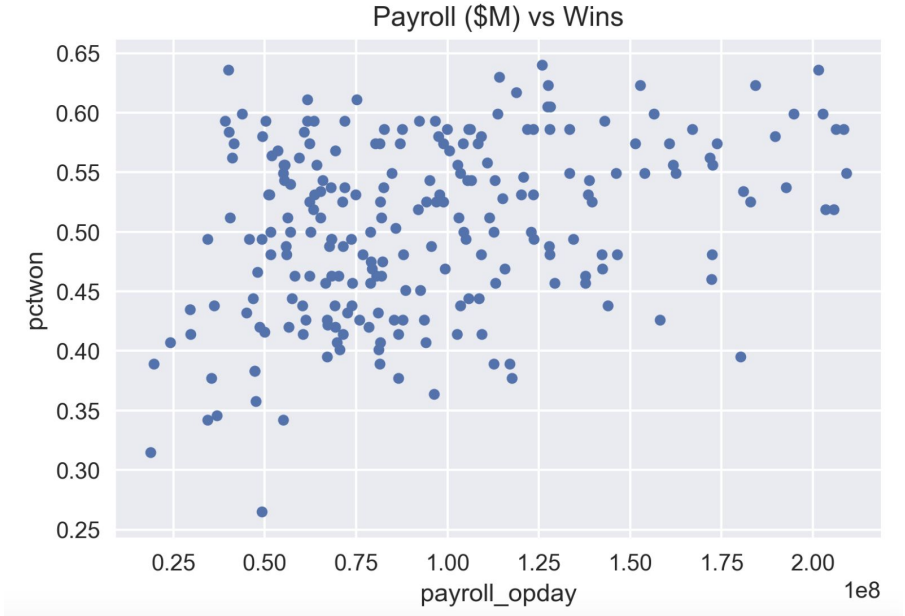
Omnibus:	5.110	Durbin-Watson:	2.143
Prob(Omnibus):	0.078	Jarque-Bera (JB):	2.800
Skew:	-0.015	Prob(JB):	0.247
Kurtosis:	2.353	Cond. No.	4.60e+05

NEWS FLASH:

To win games, just play well.

Look for Less Obvious Relationships





Option 2: A Better Model

Inputs

Pitching
Batting
Fielding
Payroll
Manager Tenure
Mariners “flag”

Features That Emerged

Small but Statistically Significant

+	–
Singles	Caught Stealing
Homeruns	Left On Base
Walks	
Hit By Pitch	
Intentional Walks	

Accuracy: 80% (+/- 8)

Improving the Model

- Tested polynomials (degrees 1, 2, 3) - did not improve model
- Regularization: Elastic Net Cross-Validation
  - Heavily weighted toward Ridge regularization
  - Many features reduced to zero
  - Mostly batting stats left - no payroll, no manager influence

So why haven't the Mariners made the playoffs in so long, then...?

Short Answer:

A lot of little things, no one big thing

No evidence of a "Mariners curse"

Future Direction

Injury Data

Who's been injured and for how long

How many $M are sitting on the bench

Top Players

How many award-winning players, and in which positions

Leadership

Changes in ownership and/or GMs

Thank you!

LINKS: Cartoon | Team Photo | Baseball Reference (Data) | Lahman Baseball Database (Data) | USA Today (Data) | Project Files (Git)