Predicting Success in the NBA

January 27, 2017

Malu Dee

What to predict?

Wins / losses by season

How?

Starting five
2006 - 2016
All teams

Source?

basketball-reference.com

Feature Selection

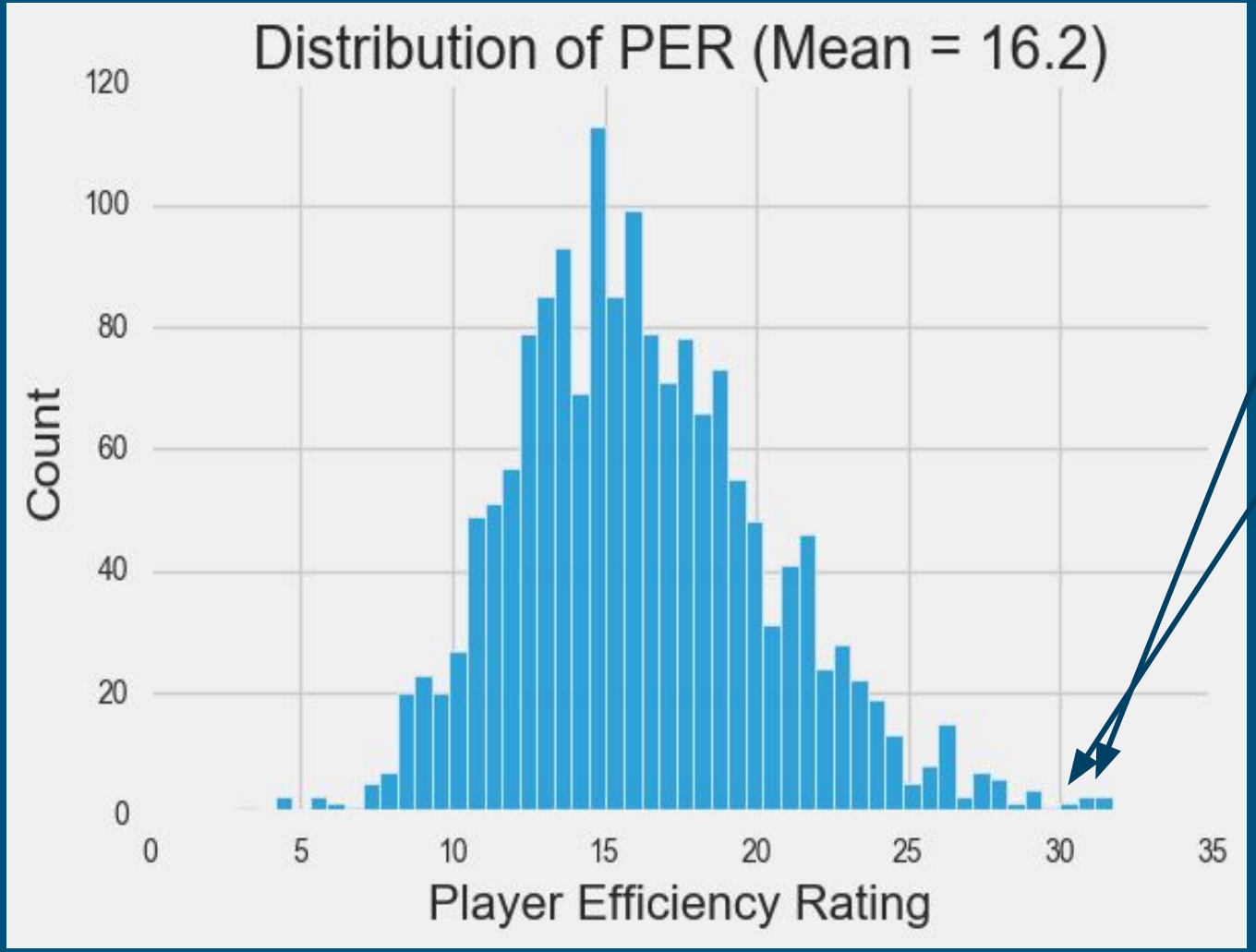
General questions -

Better to have strong offensive or defensive stats?

Better to have a balanced starting five, or a very strong player?

Averaged all stats by starting five (per team)

- Field goals
- PER (player efficiency rating)
- Age
- Salary
- Field goal success rate
- Field goal attempts
- Consistency of starting five
- Opponent field goal success
- Three pointer success rate
- Pace
- Steals
- Weak link on starting five
- Strong link on starting five



2006 - 2016

Steph Curry, 2016

LeBron James, 2016

Is it better to have a star player in the starting five?

+ 63%

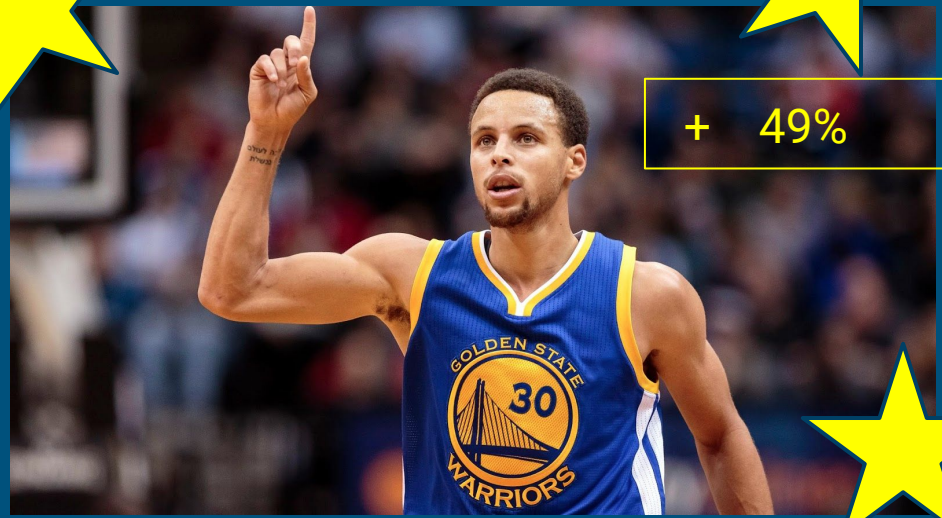
Salary: $22.9 Million
PER: 30
Points/game: 25

LeBron James

Salary: $11.3 Million
PER: 31.5
Points/game: 30

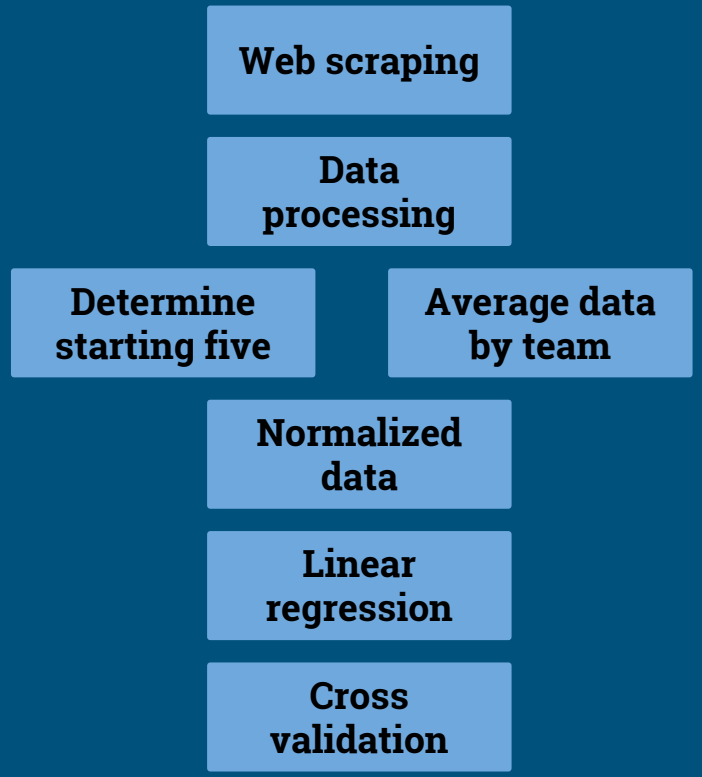
Stephen Curry

+ 49%

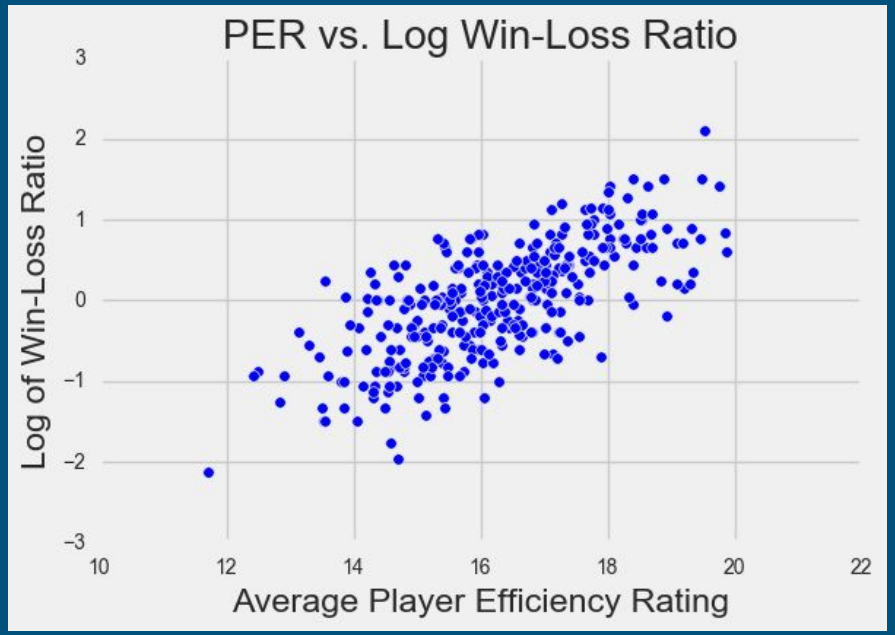


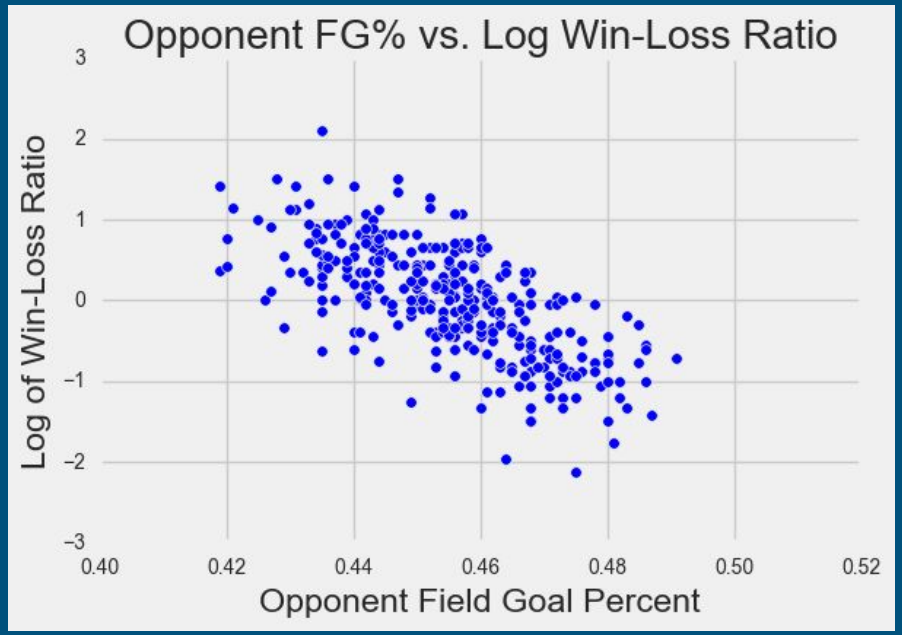
Model # 1: Exploration

- Started with many features
- Used data from the entire season
- Explored which features best predicted win/loss ratio for the end of the season



PER and Opponent Field Goal % vs. W/L Ratio





Takeaways from model # 1 (interpretation):

PER and opponent attack stats were most predictive.

Player age, the consistency of the starting five, and having a strong/weak link also had an impact.

Would these factors predict win/loss ratio if we only took data from the first games of the season?

OLS Regression Results

Dep. Variable:	log_avg_wl	**R-squared:**	0.785
Model:	OLS	**Adj. R-squared:**	0.780
Method:	Least Squares	**F-statistic:**	156.3
Date:	Thu, 26 Jan 2017	**Prob (F-statistic):**	2.12e-69
Time:	22:54:43	**Log-Likelihood:**	261.22
No. Observations:	220	**AIC:**	-510.4
Df Residuals:	214	**BIC:**	-490.1
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
opp_fg_pct	-0.3183	0.029	-11.031	0.000	-0.375 -0.261
avg_per	0.3505	0.032	11.058	0.000	0.288 0.413
intercept	0.3552	0.035	10.267	0.000	0.287 0.423
avg_pct_started	0.1985	0.032	6.291	0.000	0.136 0.261
avg_age	0.1387	0.029	4.739	0.000	0.081 0.196
weak_link	-0.1084	0.024	-4.531	0.000	-0.156 -0.061

Model # 2: Early Season Games

OLS Regression Results

Dep. Variable:	log_avg_wl	**R-squared:**	0.385
Model:	OLS	**Adj. R-squared:**	0.372
Method:	Least Squares	**F-statistic:**	31.24
Date:	Thu, 26 Jan 2017	**Prob (F-statistic):**	3.28e-20
Time:	19:08:36	**Log-Likelihood:**	136.73
No. Observations:	205	**AIC:**	-263.5
Df Residuals:	200	**BIC:**	-246.8
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
avg_per	0.4628	0.060	7.677	0.000	0.344 0.582
avg_age	0.1987	0.056	3.577	0.000	0.089 0.308
avg_opp_fg_pct	-0.1623	0.053	-3.073	0.002	-0.266 -0.058
intercept	0.1835	0.053	3.476	0.001	0.079 0.288
strong_link	0.1434	0.061	2.348	0.020	0.023 0.264

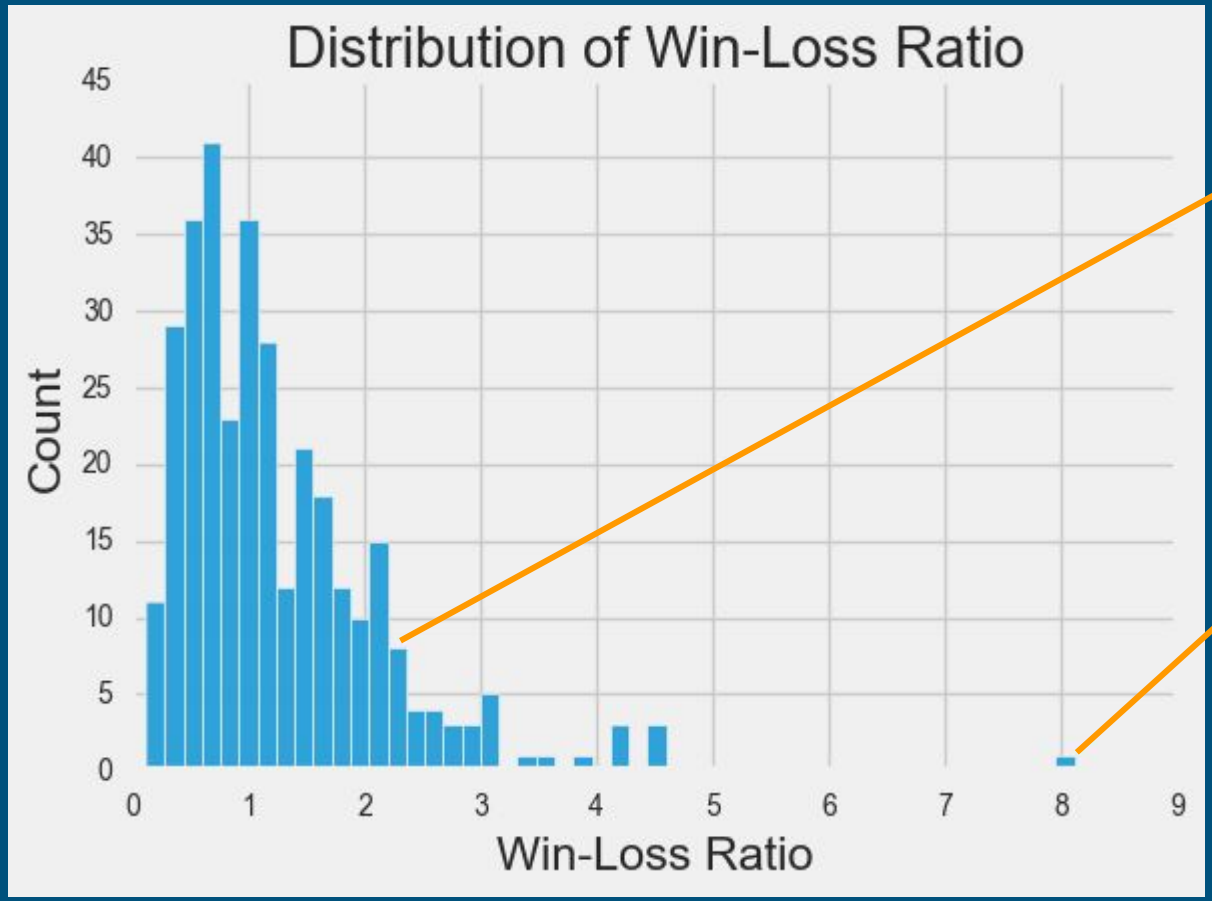
Takeaways:

Harder to predict season-end outcomes based on just the first few games.

PER, player age, and opponent field goal success rate impacted win/loss ratio.

Having a strong player in the starting five also had an impact.

2006 - 2016



Cleveland Cavaliers, 2016
2.28
57 wins - 25 losses

Golden State Warriors, 2016
8.11
73 wins - 9 losses

LeBron's Cavs stun Warriors in thriller to clinch title

NBA Finals: Cleveland Cavaliers stun Golden State Warriors 93-89 in Game Seven to take first title

	team	season	predicted_wl	avg_wl	wl_diff	avg_per	avg_age	opp_fg_pct	avg_salary	weak_link	strong_link
0	Cleveland_Cavaliers	2016	2.06	2.28	0.22	18.68	28.00	0.45	13,433,400.00	-0.48	0.63
1	Golden_State_Warriors	2016	3.19	8.11	4.93	19.52	26.20	0.43	11,408,836.80	-0.48	0.49

Sports

Cavaliers end o
sports heartbre
champio

PRO BASKETBALL | CAVALIERS 93, WARRIORS 89 | CLEVELAND WINS SERIES, 4-3

The Cavaliers' brilliant title turnaround shattered the NBA landscape

Steph Curry is fallible, the Warriors can fail and Kevin Love can defend. The NBA finals proved much of what we thought about the league is wrong

LeBron,

June 19, 2016 10:5

Most predictive features:

- PER (offensive)
- Opponent field goals (defensive)

Also:

- Age
- Having a star player
- Having a consistent starting five

Conclusions

Next steps:

& Questions?

- More years
- Data from 1st month, 2nd month, 3rd, and so on
- Exploring even more features:
  - VORP?
  - More defensive stats
  - Data from other websites
- Injuries
- Transfers
- Non-starting five