NBA Season Wins

Capstone: Jose Harper

Agenda

1) Business Problem and Stakeholder
2) Data Understanding
3) Model evaluation
4) Conclusion and Strategy
5) Future Analysis

Business Problem

1) Determine best metrics for winning the most games per NBA season
2) Current Research focuses on winning nightly games
3) Use mainly classification algorithms for game by game wins, loss or draw.

Stakeholder

- NBA basketball franchise Houston Rockets
- Hired to help influence draft picks and trades for upcoming season
- Current roster is above NBA salary cap but worst in wins

Rk		Salary
1	John Wall	$44,310,840
2	Eric Gordon	$18,218,818
3	Christian Wood	$13,666,667
4	Jalen Green	$8,992,200
5	D.J. Augustin	$7,000,000
6	Dennis Schröder	$5,890,000
7	David Nwaba	$4,650,000
8	Sekou Doumbouya	$3,613,680
9	Alperen Şengün	$3,214,680
10	Enes Freedom	$2,641,691
11	Usman Garuba	$2,353,320
12	Josh Christopher	$2,259,240
13	Kevin Porter Jr.	$2,130,240
14	Bruno Fernando	$1,782,621
15	Xavier Sneed	$1,563,518
16	Kenyon Martin Jr.	$1,517,981
17	Jae'Sean Tate	$1,517,981
18	Armoni Brooks	$1,489,065
19	Danuel House Jr.	$1,387,498
20	Garrison Mathews	$1,093,598

Data Understanding

1) High Multicolinearity for many features
2) 1057 rows of 35 NBA seasons
3) Data omitted : (a) Playoffs made (b) Second half of 2022 season metrics (c) Draft Picks (d) Teams average height and weight

Modeling Process

1) Data Preperation
2) Train and Tune models
3) Retrain and Validate
4) Final Model evaluations

Model Process continued:

OLS Model 1

Dep. Variable:	Rk	**R-squared (uncentered):**	0.753
Model:	OLS	**Adj. R-squared (uncentered):**	0.752
Method:	Least Squares	**F-statistic:**	1604.
Date:	Fri, 04 Mar 2022	**Prob (F-statistic):**	1.14e-320
Time:	11:19:01	**Log-Likelihood:**	-3759.5
No. Observations:	1057	**AIC:**	7523.
Df Residuals:	1055	**BIC:**	7533.

OLS Model 4

Dep. Variable:	Rk	**R-squared (uncentered):**	0.894
Model:	OLS	**Adj. R-squared (uncentered):**	0.892
Method:	Least Squares	**F-statistic:**	585.2
Date:	Fri, 04 Mar 2022	**Prob (F-statistic):**	0.00
Time:	11:20:37	**Log-Likelihood:**	-3312.1
No. Observations:	1057	**AIC:**	6654.
Df Residuals:	1042	**BIC:**	6729.
Df Model:	15		
Covariance Type:	nonrobust		

Final Results and Key features

1) Best Model: 4th OLS regression

2) Highly correlated features for winning

3) 5 features to focus on: (1) 3P%, (2) FG%, (3) STL, (4) ORB, (5) DRB

FGA	1.7302
FG%	-24.4008
3P%	-36.5593
2P%	187.7582
FT	-0.0887
FTA	0.3811
FT%	29.9185
ORB	-1.3014
DRB	-0.4625
AST	-0.1489
STL	-1.5955
BLK	0.0365
TOV	1.1003
PF	-0.0420
PTS	-1.9447

Conclusion/Recomendations

	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
Team	63	15170	2441	5427	.450	844	2454	.344	1597	2973	.537	1103	1555	.709	612	2045	2657	1456	461	284	1051	1284	6829

1) Areas that need improvement at 3P%, ORB, DFR, FT%, AST, STL, BLK, FG%
2) Since salary cap is already high trade and draft accordingly
3) Focus on trades/draft picks that have high 3P% Efficiency, athletic big men that have high 3P%, and/or Athletic guards with long wingspan

Future Analysis

1) Use website api
2) Use only 10 most recent NBA seasons for data
3) Compare salary of top teams
4) Add features such as : Age, weight, height, wingspan, experience and minutes per game played

Thanks!

Contact us:

Jose Harper
Harper.Jose@gmail.com