Creating an Efficient 5-Man Roster Using Predictive

Analytics

Summer 2022 MSBA Capstone Presentation

By: Drew Mahoney

<u>SME:</u> Bianca Gadže (Soccer Data Analyst)

MSBA Courses That Assisted This Project

- Machine Learning for Predictive Analytics (modeling, python)
- Business Intelligence (Tableau)
- Intro to Business Analytics (How to analyze results)
- Sports Analytics

The Problem Statement

- The National Basketball Association (NBA) has developed a metric called Player Efficiency Rating (PER) that is tailored towards giving players numerical scores based on their positive/negative impacts on the game while they are playing.
- The NBA is a star's league: Meaning the players that are great are typically awarded financially, and throughout some games (getting calls in their favor, shooting more shots, etc.)

Would an NBA team that focuses mainly on recruiting players with higher Player Efficiency Ratings win more than a team that would focus on spending more money for players?
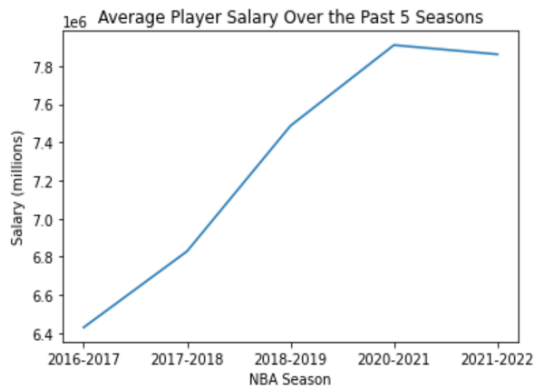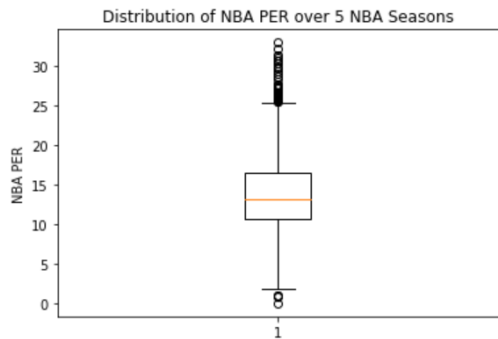
The Data

- All player statistics were pulled from basketball-reference.com
  - Collected the past 5 NBA seasons (excluding the 2019-2020 season) worth of data
    - 1,563 unique NBA players, along with 43 statistical variables
- All player and team salaries were pulled from hoopshype.com

Initial EDA

Distribution of NBA PER over 5 NBA Seasons

Average Player Salary Over the Past 5 Seasons

```
nba_data.groupby('Season')['NBA PER'].agg(np.mean)
```

```
Season
2016-2017    13.652719
2017-2018    13.647667
2018-2019    13.612737
2020-2021    13.984881
2021-2022    13.932167
Name: NBA PER, dtype: float64
```

```
nba_data.nlargest(3, 'Salary')
```

	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	...	PF	PTS	Salary	NBA PER	Season
31	Stephen Curry	PG	33	GSW	64	64	2211	535	1224	0.437	...	130	1630	45780966	21.47	2021-2022
39	James Harden	SG	32	TOT	65	65	2419	407	992	0.410	...	153	1432	44310840	20.92	2021-2022
172	Russell Westbrook	PG	33	LAL	78	78	2678	548	1233	0.444	...	235	1441	44211146	15.01	2021-2022

3 rows × 37 columns

Binned Categorical Variables

Binned Categorical Variables	Definitions
east_west_traded	1 = team in eastern conference, 2 = team in western conference, 3 = player was traded mid-season
position_bins	1 = PG, 2 = SG, 3 = SF, 4 = PF, 5 = C
age_bin	0 = (18-25), 1 = (26-32), 2 = (33+)
games_bin	0 = (10-40 GP), 1 = (41-59 GP), 2 = (60-72 GP), 3 = (73-82 GP)
Games_started_bin	0 = (0=2 GS), 1 = (3-14 GS), 2 = (15-51 GS), 3 = (52-82 GP)

Models

Ran 4 different predictive regression models to see which did the best at predicting NBA PER based on individual statistics

- Linear Regression
- Gradient Booster
- Random Forest
- Decision Tree

```
results = {'Linear Regression':[lin_regress_mse],'Gradient Booster Regressor':[gb_mse],
           'Random Forest Regresor':[rfr_mse], 'Decision Tree Regressor':[dtr_mse]}
modeling_results = pd.DataFrame(data=results,index=['MSE of Models'])
modeling_results
```

	Linear Regression	Gradient Booster Regressor	Random Forest Regressor	Decision Tree Regressor
MSE of Models	2.241644	3.305653	3.567008	7.659761

Evaluation Metric: Mean Squared Error (MSE)

Linear Regression Model

- With this model, I trained it to predict each NBA players' PER based on their basic stats each season (total points, rebounds, assists, steals, blocks, fouls, etc.), then compared the results with the actual score
- With the new scores, I grouped each player based on their position to begin forming a team from scratch

	Player	Season	NBA PER	Predicted PER	Difference
35	Giannis Antetokounmpo	2021-2022	32.12	31.350547	0.769453
119	Joel Embiid	2021-2022	31.24	30.916774	0.323226
37	Giannis Antetokounmpo	2018-2019	30.95	30.834706	0.115294
202	DeMar DeRozan	2016-2017	30.70	28.877206	1.822794
7	James Harden	2018-2019	30.62	24.681813	5.938187
...	...	...	...	...	...
1444	Mychal Mulder	2021-2022	3.72	7.947476	-4.227476
1551	Justin Robinson	2021-2022	3.64	5.181451	-1.541451
461	Gary Trent Jr.	2018-2019	3.61	5.940781	-2.330781
1291	Tim Frazier	2021-2022	3.32	5.816476	-2.496476
633	Solomon Hill	2021-2022	3.30	1.906766	1.393234

1563 rows x 5 columns

NBA PER vs Predicted PER by Primary Position
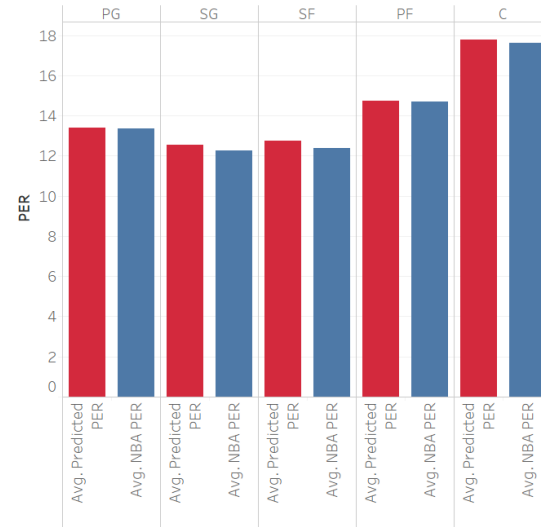
How the Teams Were Formed

- According to hoopshype.com, during the 2021-2022 NBA season, the average team payroll was ~$140 million
  - Almost $80 million of that team payroll was distributed amongst the starting 5
  - Out of all 30 NBA Teams:
    - Average starting PG salary: $18 million
    - Average starting SG salary: $14 million
    - Average starting SF salary: $14 million
    - Average starting PF salary: $16 million
    - Average starting C salary: $15 million

```
[57] pg_df = new_nba_data_2[(new_nba_data_2['Salary'] <= 19000000) & (new_nba_data_2['position_bins'] == 1) & (new_nba_data_2['Season'] == '2021-2022')]
     sg_df = new_nba_data_2[(new_nba_data_2['Salary'] <= 14000000) & (new_nba_data_2['position_bins'] == 2) & (new_nba_data_2['Season'] == '2021-2022')]
     sf_df = new_nba_data_2[(new_nba_data_2['Salary'] <= 15000000) & (new_nba_data_2['position_bins'] == 3) & (new_nba_data_2['Season'] == '2021-2022')]
     pf_df = new_nba_data_2[(new_nba_data_2['Salary'] <= 17000000) & (new_nba_data_2['position_bins'] == 4) & (new_nba_data_2['Season'] == '2021-2022')]
     c_df = new_nba_data_2[(new_nba_data_2['Salary'] <= 15000000) & (new_nba_data_2['position_bins'] == 5) & (new_nba_data_2['Season'] == '2021-2022')]
```

2021-2022 Team

```
all_possible.groupby('position_bins')[['Salary', 'NBA PER', 'Predicted PER']].agg(np.max)
```

	Salary	NBA PER	Predicted PER
position_bins			
1	18604651	25.48	26.081569
2	13347727	19.07	20.672386
3	13036364	17.62	19.581978
4	16409091	23.75	22.277792
5	14000000	23.26	25.017078

Based on NBA's PER

Player	Season	Tm	Salary	NBA PER	e
Trae Young	2021-2022	ATL	8326471	25.48	
Terry Taylor	2021-2022	IND	895125	19.07	
Desmond Bane	2021-2022	MEM	2033160	17.62	
Brandon Clarke	2021-2022	MEM	2726880	23.75	
Montrezl Harrell	2021-2022	TOT	9720900	23.26	

Based on Model's Predicted PER

Player	Season	Tm	Salary	NBA PER
Dejounte Murray	2021-2022	SAS	15428880	22.32
Tyrese Haliburton	2021-2022	TOT	4023600	18.25
Desmond Bane	2021-2022	MEM	2033160	17.62
Brandon Clarke	2021-2022	MEM	2726880	23.75
Robert Williams	2021-2022	BOS	3661976	22.10

Average PER: 21.836

Average PER: 20.808

Average NBA Team's Top 5 PER: **18.623**

Past 5 Years Team

```
all5pos = pd.concat([pgdf, sgdf, sfdf, pfdf, cdf])
all5pos.groupby('position_bins')[['Salary','NBA PER','Predicted PER']].agg(np.max)
```

	Salary	NBA PER	Predicted PER
position_bins			
1	19000000	26.59	27.262416
2	14000000	21.67	22.839702
3	15000000	26.13	28.102933
4	16409091	27.17	25.774643
5	15000000	29.70	28.331579

Based on NBA's PER

Player	Season	Tm	Salary	NBA PER
Isaiah Thomas	2016-2017	BOS	6587131	26.59
Shai Gilgeous-Alexander	2020-2021	OKC	4141320	21.67
Giannis Antetokounmpo	2016-2017	MIL	2995420	26.13
Zion Williamson	2020-2021	NOP	10245480	27.17
Boban Marjanovic	2016-2017	DET	5628000	29.70

Based on Model's Predicted PER

Player	Season	Tm	Salary	NBA PER	ea
Stephen Curry	2016-2017	GSW	12112359	24.74	
Lou Williams	2018-2019	LAC	8000000	21.30	
Giannis Antetokounmpo	2016-2017	MIL	2995420	26.13	
Zion Williamson	2020-2021	NOP	10245480	27.17	
Karl-Anthony Towns	2018-2019	MIN	7839435	26.38	

Average PER: 25.144

Average PER: 25.144

Average NBA Team's Top 5 PER: **24.123**

Well, How Did A Team Focused on Maximizing Salary Do?

2021-2022 Team

```
all_possible.groupby('position_bins')[['Salary', 'NBA PER', 'Predicted PER']].agg(np.max)
```

	Salary	NBA PER	Predicted PER
position_bins			
1	18604651	25.48	26.081569
2	13347727	19.07	20.672386
3	13036364	17.62	19.581978
4	16409091	23.75	22.277792
5	14000000	23.26	25.017078

Average NBA Team's Top 5 PER: **18.623**

Player	Season	Tm	Salary	NBA PER
Lonzo Ball	2021-2022	CHI	18604651	14.49
Luke Kennard	2021-2022	LAC	13347727	13.31
Joe Ingles	2021-2022	UTA	13036364	10.05
Aaron Gordon	2021-2022	DEN	16409091	15.38
Jonas Valanciunas	2021-2022	NOP	14000000	21.30

Average PER For A Team Trying to Maximize Salary: **14.906**

Past 5 Years Team

```
all5pos = pd.concat([pgdf, sgdf, sfdf, pfdf, cdf])
all5pos.groupby('position_bins')[['Salary','NBA PER','Predicted PER']].agg(np.max)
```

	Salary	NBA PER	Predicted PER
position_bins			
1	19000000	26.59	27.262416
2	14000000	21.67	22.839702
3	15000000	26.13	28.102933
4	16409091	27.17	25.774643
5	15000000	29.70	28.331579

Average NBA Team's Top 5 PER: **24.123**

Player	Season	Tm	Salary	NBA PER
Jonas Valanciunas	2020-2021	MEM	15000000	24.47
Eric Bledsoe	2016-2017	PHO	14000000	20.56
Taj Gibson	2018-2019	MIN	14000000	17.85
Manu Ginobili	2016-2017	SAS	14000000	13.94
Trevor Ariza	2018-2019	TOT	15000000	12.08

Average PER For A Team Trying to Maximize Salary: **17.780**

Another Step Further

- The results clearly show that PER is a compelling factor in terms of building a team, but just basic stats isn't enough
    - To predict wins? - Win Shares, Usage %, Turnover %, and other advanced metrics also impact winning and losing
- There is a clear positive correlation between Win Shares and NBA PER, and a no real correlation between Win Shares and Salary
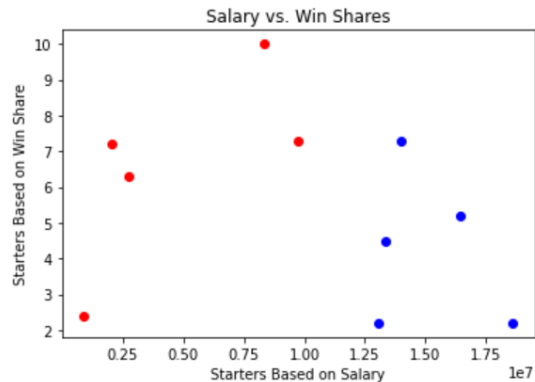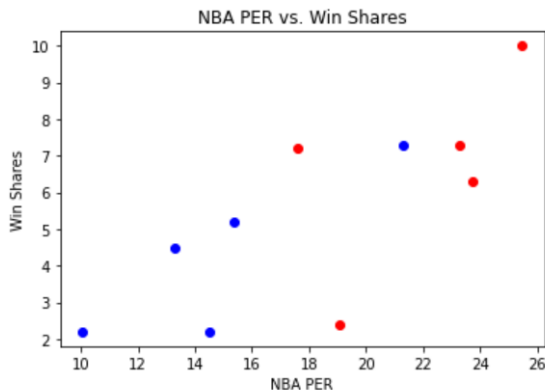
A New PER Prediction

- Using Linear Regression model once again, I was able to account for more features that would hopefully give me a more accurate prediction of PER.

Player	Salary	NBA PER	Predicted PER	New Predicted PER
Aaron Holiday	3980551	12.59	12.512259	12.207890
Armoni Brooks	1739065	8.45	8.213012	8.399359
Bones Hyland	2096880	14.44	13.661501	13.366124
Brad Wanamaker	814918	7.53	9.296516	9.441122
Brandon Goodwin	343988	11.88	10.679081	10.622992

Player	Season	Tm	Salary	NBA PER
Trae Young	2021-2022	ATL	8326471	25.48
Tyrese Haliburton	2021-2022	TOT	4023600	18.25
Desmond Bane	2021-2022	MEM	2033160	17.62
Brandon Clarke	2021-2022	MEM	2726880	23.75
Robert Williams	2021-2022	BOS	3661976	22.10

Average PER: **21.44**

Player	Season	Tm	Salary	NBA PER	ea
Stephen Curry	2016-2017	GSW	12112359	24.74	
Lou Williams	2018-2019	LAC	8000000	21.30	
Giannis Antetokounmpo	2016-2017	MIL	2995420	26.13	
Julius Randle	2018-2019	NOP	8641000	21.02	
Karl-Anthony Towns	2018-2019	MIN	7839435	26.38	

Average PER: **23.914**

NBA PER: The Most Efficient Configuration of Players
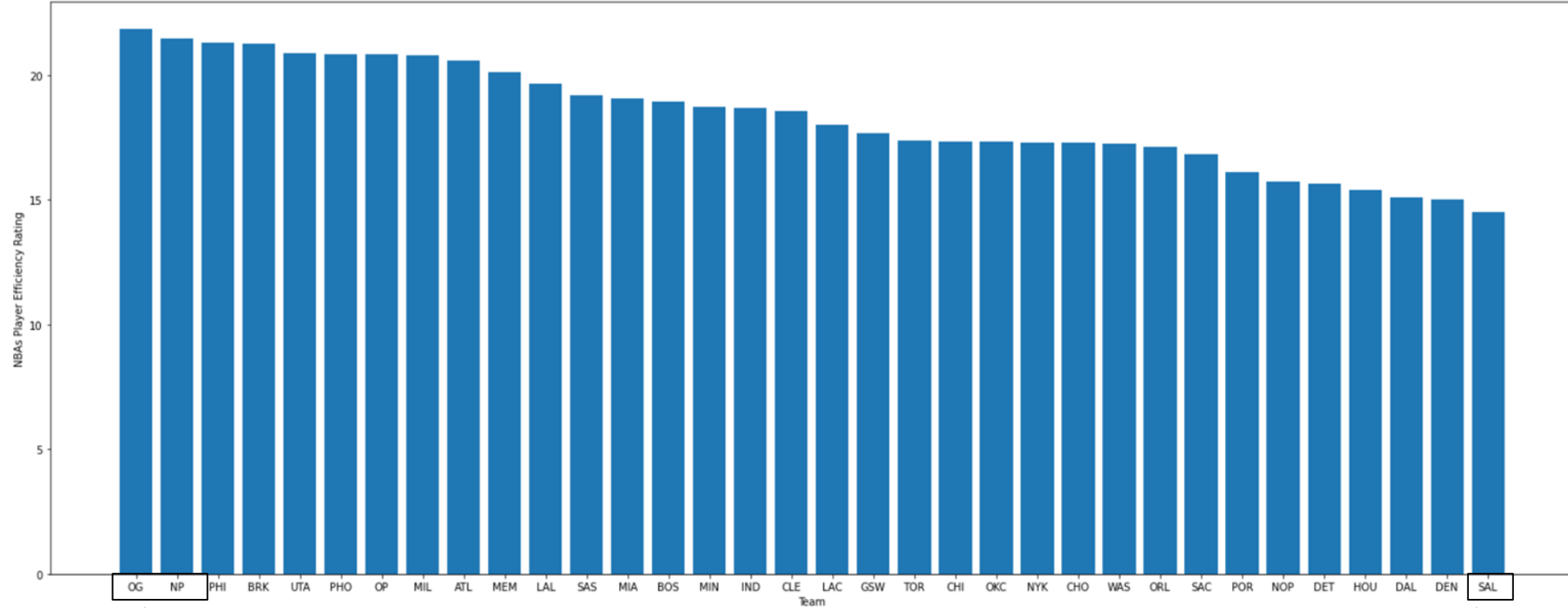
OG = Original NBA PER

NP = New NBA PER Prediction

OP = "Old" NBA PER Prediction

SAL = NBA PER based on maximizing salary

What I Learned

1. NBA PER has a direct correlation to wins (in the sense of Total Win Shares), and it is clear that it would be better for an NBA team to build based off of PER, rather than spending a lot of money to get players
2. Dealing with players that were traded

Next Steps

- Using PER to predict future salaries, so an NBA team can forecast / plan for the future