

# How Advanced Stats Paint a Better Picture of Player Performance in the NBA

## Introduction:

The National Basketball Association or the NBA is one of the most popular leagues globally with the highest social media traction. The league has the most number of Instagram and Twitter followers and the most number of YouTube subscribers than any other sports league in the world (Tillman). The die-hard fans go to great lengths to show their support to their favourite team or player. But for the casual fan watching every moment in the long-winded 82 game season is a tough ask.

To summarize the game quickly, the box score presents the stats for the game. The box score ideally has the base level stats such as points, assists and rebounds, but this does not convey what truly takes place on the court. Sheer numbers do not showcase the impact of some players; hence taking these numbers at face value is as good as assuming the highest-scoring player in the game was the most impactful or put his team in a winning position.

Enter advanced statistics; the current age of basketball analytics has bought in new terms, stats and metrics to measure player performance during the season. Though it might have its downfalls when viewed with the entire season's sample size, it does prove to be a viable metric to be followed. Let us delve deeper into the nitty-gritty of things and find out how reliable these advanced statistics are.

Now let us have a look at our dataset. The dataset contains the per game and advanced statistics of every player in the NBA. Now let us look at the metadata of this dataset.

## NBA Per Game Glossary:

<b>Player</b>	Player Name	<b>FT</b>	Free Throws Per Game
<b>Pos</b>	Position	<b>FTA</b>	Free Throws Attempts Per Game
<b>Age</b>	Age	<b>FT%</b>	Free Throws Percentage
<b>Tm</b>	Team	<b>ORB</b>	Offensive Rebounds Per Game
<b>G</b>	Games Played	<b>DRB</b>	Defensive Rebounds Per Game
<b>GS</b>	Games Started	<b>TRB</b>	Total Rebounds Per Game
<b>MP</b>	Minutes Played	<b>AST</b>	Assists Per Game
<b>FG</b>	Field Goals Per Game	<b>STL</b>	Steals Per Game
<b>FGA</b>	Field Goal Attempts Per Game	<b>BLK</b>	Blocks Per Game
<b>FG%</b>	Field Goal Percentage	<b>TOV</b>	Turnovers Per Game
<b>3P</b>	3-Point Field Goals Per Game	<b>PF</b>	Personal Fouls Per Game
<b>3PA</b>	3-Point Field Attempted Goals Per Game	<b>PTS</b>	Points Per Game
<b>3P%</b>	3-Point Field Goal Percentage		
<b>2P</b>	2-Point Field Goals Per Game		
<b>2PA</b>	2-Point Field Attempted Goals Per Game		
<b>2P%</b>	2-Point Field Goal Percentage		
<b>eFG%</b>	Effective Field Goal Percentage; the formula is $(FG + 0.5 * 3P) / FGA$		

## NBA Advanced Stats Glossary:

<b>Player</b>	Player Name
<b>Pos</b>	Position
<b>Age</b>	Age
<b>Tm</b>	Team
<b>G</b>	Games Played
<b>MP</b>	Minutes Played
<b>PER</b>	Player Efficiency Rating (available since the 1951-52 season); PER is a rating developed by ESPN.com columnist John Hollinger. In John's words, "The PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance."
<b>TS%</b>	True Shooting Percentage; the formula is $PTS / (2 * TSA)$ . True shooting percentage is a measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws
<b>3PAr</b>	Percentage of field goal attempts from 3-Point range
<b>FTr</b>	Number of free throw attempts per FG attempt
<b>ORB%</b>	Offensive Rebound Percentage (available since the 1970-71 season in the NBA); the formula is $100 * (ORB * (Tm MP / 5)) / (MP * (Tm ORB + Opp DRB))$ . Offensive rebound percentage is an estimate of the percentage of available offensive rebounds a player grabbed while he was on the floor.
<b>DRB%</b>	Defensive Rebound Percentage (available since the 1970-71 season in the NBA); the formula is $100 * (ORB * (Tm MP / 5)) / (MP * (Tm ORB + Opp DRB))$ . Defensive rebound percentage is an estimate of the percentage of available defensive rebounds a player grabbed while he was on the floor.
<b>TRB%</b>	Total Rebound Percentage (available since the 1970-71 season in the NBA); the formula is $100 * (TRB * (Tm MP / 5)) / (MP * (Tm TRB + Opp TRB))$ . Total rebound percentage is an estimate of the percentage of available rebounds a player grabbed while he was on the floor.
<b>AST%</b>	Assist Percentage (available since the 1964-65 season in the NBA); the formula is $100 * AST / (((MP / (Tm MP / 5)) * Tm FG) - FG)$ . Assist percentage is an estimate of the percentage of teammate field goals a player assisted while he was on the floor.
<b>STL%</b>	Steal Percentage (available since the 1973-74 season in the NBA); the formula is $100 * (STL * (Tm MP / 5)) / (MP * Opp Poss)$ . Steal Percentage is an estimate of the percentage of opponent possessions that end with a steal by the player while he was on the floor.
<b>BLK%</b>	Block Percentage (available since the 1973-74 season in the NBA); the formula is $100 * (BLK * (Tm MP / 5)) / (MP * (Opp FGA - Opp 3PA))$ . Block percentage is an estimate of the percentage of opponent two-point field goal attempts blocked by the player while he was on the floor.
<b>TOV%</b>	Turnover Percentage (available since the 1977-78 season in the NBA); the formula is $100 * TOV / (FGA + 0.44 * FTA + TOV)$ . Turnover percentage is an estimate of turnovers per 100 plays.
<b>USG%</b>	Usage Percentage (available since the 1977-78 season in the NBA); the formula is $100 * ((FGA + 0.44 * FTA + TOV) * (Tm MP / 5)) / (MP * (Tm FGA + 0.44 * Tm FTA + Tm TOV))$ . Usage percentage is an estimate of the percentage of team plays used by a player while he was on the floor.
<b>WS</b>	Win Shares is a player statistic which attempts to divvy up credit for team success to the individuals on the team
<b>OWS</b>	Offensive Win Shares
<b>DWS</b>	Defensive Win Shares
<b>WS/48</b>	Win Shares per 48 minutes
<b>BPM</b>	Box Plus/Minus (available since the 1973-74 season in the NBA); a box score estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team
<b>OBPM</b>	Offensive Box Plus/Minus
<b>DBPM</b>	Defensive Box Plus/Minus
<b>VORP</b>	Value Over Replacement Player (available since the 1973-74 season in the NBA); a box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season. Multiply by 2.70 to convert to wins over replacement.

(Basketball Reference)

## Data Accumulation and Preprocessing:

After picking up the data from Basketball-Reference, it was converted to CSV, making the importing process easier. The cleaning process of the data was done on Google Sheets to make the data ready for analysis right after importing. Player name and player ID was split into different columns to enhance readability.

Player		Player	PlayerID
Precious Achiuwa\achiupr01		Precious Achiuw	achiupr01
Jaylen Adams\adamsja01		Jaylen Adams	adamsja01
Steven Adams\adamsst01		Steven Adams	adamsst01
Bam Adebayo\adebaba01		Bam Adebayo	adebaba01
LaMarcus Aldridge\aldrila01		LaMarcus Aldrid	aldrila01

The analysis is made more straightforward by initially filtering out the data by removing players who played less than 40 games. The 2020-21 season was shortened to 72 compared to the usual 82 games. Hence the 40 game threshold is feasible.

Importing libraries and the datasets, validation of the dataset:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
nba_pergame=pd.read_csv("nba_pergame_1.csv")
nba_basic=nba_pergame[nba_pergame['G']>40]
nba_basic.head()
```

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL
0	1	Precious Achiuwa	PF	21	MIA	61	4	12.1	2.0	3.7	0.544	0.0	0.0	0.000	2.0	3.7	0.546	0.544	0.9	1.8	0.509	1.2	2.2	3.4	0.5	0.3
2	3	Steven Adams	C	27	NOP	58	58	27.7	3.3	5.3	0.614	0.0	0.1	0.000	3.3	5.3	0.620	0.614	1.0	2.3	0.444	3.7	5.2	8.9	1.9	0.9
3	4	Bam Adebayo	C	23	MIA	64	64	33.5	7.1	12.5	0.570	0.0	0.1	0.250	7.1	12.4	0.573	0.571	4.4	5.5	0.799	2.2	6.7	9.0	5.4	1.2
8	7	Nickell Alexander-Walker	SG	22	NOP	46	13	21.9	4.2	10.0	0.419	1.7	4.8	0.347	2.5	5.2	0.485	0.502	1.0	1.4	0.727	0.3	2.8	3.1	2.2	1.0
9	8	Grayson Allen	SG	25	MEM	50	38	25.2	3.5	8.3	0.418	2.1	5.5	0.391	1.3	2.8	0.471	0.547	1.6	1.8	0.868	0.4	2.8	3.2	2.2	0.9

```
nba_advanced=pd.read_csv("nba_advanced_1.csv")
nba_adv=nba_advanced[nba_advanced['G']>40]
nba_adv.head()
```

	Rk	Player	Pos	Age	Tm	G	MP	PER	TS%	3PAr	FT%	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS	DWS	WS	WS/48	OBPM
0	1	Precious Achiuwa	PF	21	MIA	61	737	14.2	0.550	0.004	0.482	11.5	20.6	16.1	6.1	1.3	4.0	13.5	19.5	0.3	1.0	1.3	0.085	-3.5
2	3	Steven Adams	C	27	NOP	58	1605	15.1	0.596	0.010	0.438	14.4	20.4	17.4	9.1	1.6	2.2	17.5	11.7	2.3	1.7	4.0	0.119	-0.3
3	4	Bam Adebayo	C	23	MIA	64	2143	22.7	0.626	0.010	0.443	7.7	22.6	15.3	26.9	1.7	3.2	15.0	23.7	5.6	3.2	8.8	0.197	2.9
8	7	Nickell Alexander-Walker	SG	22	NOP	46	1007	12.5	0.522	0.478	0.144	1.4	14.1	7.8	14.7	2.2	2.1	12.4	23.2	-0.3	1.0	0.7	0.035	-1.4
9	8	Grayson Allen	SG	25	MEM	50	1259	12.8	0.586	0.662	0.220	1.6	12.0	6.7	11.5	1.7	0.6	9.6	16.8	1.5	1.2	2.7	0.101	-0.4

```
nba_adv.count()
```

```
Rk      329
Player  329
Pos      329
Age      329
Tm       329
G        329
MP       329
PER      329
TS%      329
3PAr     329
FTr      329
ORB%     329
DRB%     329
TRB%     329
AST%     329
STL%     329
BLK%     329
TOV%     329
USG%     329
OWS      329
DWS      329
WS       329
WS/48    329
OBPM     329
DBPM     329
BPM      329
VORP     329
PlayerID 329
dtype: int64
```

```
nba_adv.dropna().count()
```

```
Rk      329
Player  329
Pos      329
Age      329
Tm       329
G        329
MP       329
PER      329
TS%      329
3PAr     329
FTr      329
ORB%     329
DRB%     329
TRB%     329
AST%     329
STL%     329
BLK%     329
TOV%     329
USG%     329
OWS      329
DWS      329
WS       329
WS/48    329
OBPM     329
DBPM     329
BPM      329
VORP     329
PlayerID 329
dtype: int64
```

## Data Analysis:

In this stage, let us try to differentiate and identify which statistics give better insight into player performance and impact. To figure this out let us draw a comparison between the age-old box score metrics and the top players of the season. But before that, let us see the season's award winners, which establish a focal point for comparison.

## League Awards

Award	Winner
<a href="#">Most Valuable Player</a>	<a href="#">Nikola Jokić</a>
<a href="#">Rookie of the Year</a>	<a href="#">LaMelo Ball</a>
<a href="#">Defensive Player of the Year</a>	<a href="#">Rudy Gobert</a>
<a href="#">Most Improved Player</a>	<a href="#">Julius Randle</a>
<a href="#">Sixth Man of the Year</a>	<a href="#">Jordan Clarkson</a>

(Basketball Reference 2)

Regular Stats:

```
nba_basic[['Player','Tm','Pos','PTS']].sort_values('PTS',ascending=False).head()
```

	Player	Tm	Pos	PTS
151	Stephen Curry	GSW	PG	32.0
45	Bradley Beal	WAS	SG	31.3
396	Damian Lillard	POR	PG	28.8
189	Joel Embiid	PHI	C	28.5
17	Giannis Antetokounmpo	MIL	PF	28.1

```
nba_basic[['Player','Tm','Pos','AST']].sort_values('AST',ascending=False).head()
```

	Player	Tm	Pos	AST
675	Russell Westbrook	WAS	PG	11.7
257	James Harden	TOT	PG-SG	10.8
702	Trae Young	ATL	PG	9.4
237	Draymond Green	GSW	PF	8.9
510	Chris Paul	PHO	PG	8.9

```
nba_basic[['Player','Tm','Pos','TRB']].sort_values('TRB',ascending=False).head()
```

	Player	Tm	Pos	TRB
113	Clint Capela	ATL	C	14.3
228	Rudy Gobert	UTA	C	13.5
647	Jonas Valančiūnas	MEM	C	12.5
178	Andre Drummond	TOT	C	12.0
577	Domantas Sabonis	IND	PF	12.0

```
nbthree=nba_basic[nba_basic['3PA']>5].sort_values('3P%',ascending=False)
nbthree[['Player','Tm','Pos','3P%']].head()
```

	Player	Tm	Pos	3P%
269	Joe Harris	BRK	SF	0.475
459	Marcus Morris	LAC	PF	0.473
314	Joe Ingles	UTA	SF	0.451
527	Michael Porter Jr.	DEN	SF	0.445
535	Norman Powell	TOR	SG	0.439

The three-point attempts were filtered out to above 5 to make sure we pick high volume shooters with good percentages. Because you can attempt a low number of threes and maintain an efficient percentage but more attempts will always translate to more points and impact.

Here we can see league leaders in points, assists, rebounds and three-point percentage, but it is not always the case where the league leaders in these stats end up winning awards or contributing to a winning record for their team. The regular box score can showcase which player gets the most points or assists, but how much those individual stats translate to winning is exemplified in the advanced metrics.

#### Advanced Stats:

- **Box Plus/Minus** - Box Plus/Minus (available since the 1973-74 season in the NBA); a box score estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team (Basketball Reference). This stat shows how well the team performs(+/- points advantage) when that particular player is on the court playing for them against the other team.

```
nba_adv[['Player','Tm','Pos','BPM']].sort_values('BPM',ascending=False).head()
```

	Player	Tm	Pos	BPM
343	Nikola Jokić	DEN	C	11.7
17	Giannis Antetokounmpo	MIL	PF	8.8
151	Stephen Curry	GSW	PG	8.1
106	Jimmy Butler	MIA	SF	7.5
328	LeBron James	LAL	PG	7.5

Here we can see the league MVP (Most Valuable Player) has the highest BPM score, and this just makes sense as whenever he was on the floor, the Denver Nuggets were beating their opponents by 11.7 points and Giannis who came in second, was the eventual finals MVP. This shows how much a player matters to the team and his impact when he is on the floor for them.

- Win Shares - Win Shares is a player statistic that attempts to divvy up credit for team success to the individuals on the team.

```
nba_adv[['Player','Tm','Pos','WS']].sort_values('WS',ascending=False).head()
```

	Player	Tm	Pos	WS
343	Nikola Jokić	DEN	C	15.6
228	Rudy Gobert	UTA	C	11.3
396	Damian Lillard	POR	PG	10.4
17	Giannis Antetokounmpo	MIL	PF	10.2
106	Jimmy Butler	MIA	SF	9.3

This stat is a clear indicator of how many team wins the individual player is responsible for. If it weren't for Rudy Gobert on the Utah Jazz, they would have lost 11 more games. This stat cannot always be taken at face value, but these hold a lot more weight when taken from a bigger sample size, for example, stats accumulated over a season or more. Plus, one more thing to notice in win shares is that the leading player of this statistic was the MVP, and Rudy Gobert, who was second, was the Defensive Player of the Year (DPOY).

- PER - Player Efficiency Rating (available since the 1951-52 season); PER is a rating developed by ESPN.com columnist John Hollinger. In John's words, "The PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance." (Basketball Reference). This stat is widely accepted as the hallmark NBA player efficiency.

```
nba_adv[['Player','Tm','Pos','PER']].sort_values('PER',ascending=False).head()
```

	Player	Tm	Pos	PER
343	Nikola Jokić	DEN	C	31.3
189	Joel Embiid	PHI	C	30.3
17	Giannis Antetokounmpo	MIL	PF	29.2
688	Zion Williamson	NOP	PF	27.1
106	Jimmy Butler	MIA	SF	26.5

And it shows Nikola Jokic was the regular-season MVP, and Joel Embiid, who is second in PER, finished second in MVP voting. Taking these stats by themselves still does not paint a clear picture. Still, when multiple advanced stats are considered, it shows that it is undeniably a viable metric and backs up the eye test as well.

- VORP - Value Over Replacement Player (available since the 1973-74 season in the NBA); a box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season (Basketball Reference). If VORP is 10, it means in a world in which you share the same average teammates as everybody else; you're making a 10-point difference for your team every game. (Horning)

```
nba_adv[['Player','Tm','Pos','VORP']].sort_values('VORP',ascending=False).head()
```

	Player	Tm	Pos	VORP
343	Nikola Jokić	DEN	C	8.6
151	Stephen Curry	GSW	PG	5.5
17	Giannis Antetokounmpo	MIL	PF	5.5
171	Luka Dončić	DAL	PG	5.0
396	Damian Lillard	POR	PG	4.8

This stat clearly shows why Jokic was such a dominant force this season. The IQ of the big man and his ability to create on offence is unparalleled at his position.

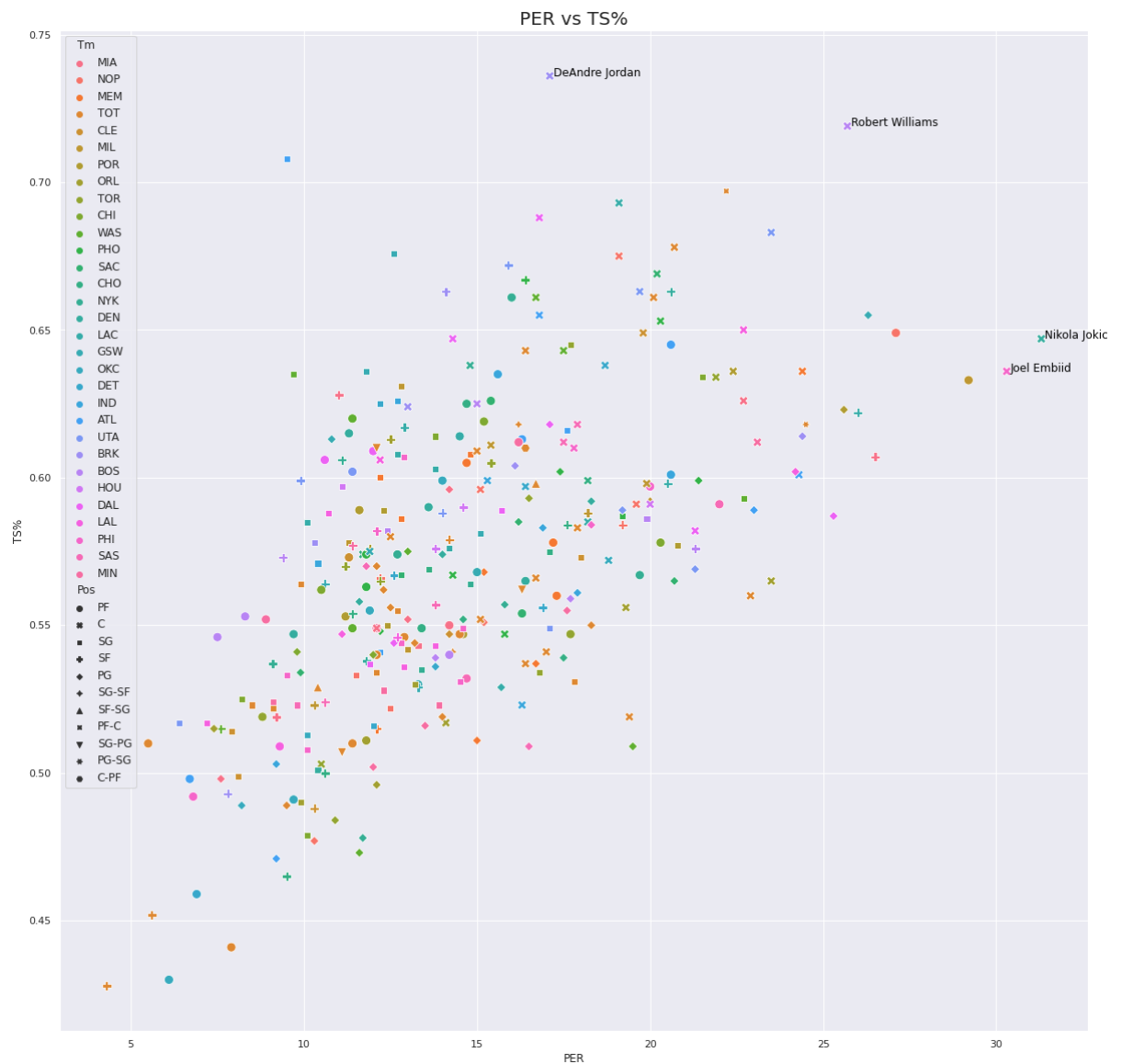


## Data Visualisation:

The tables show the comparison between the top 5 players in that particular metric but to get a broader sense of how good the league MVP and the top players are let us visualise these stats.

- PER vs TS% -

```
sns.set(rc = {'figure.figsize':(20,20)})
sns.scatterplot('PER','TS%',data=nba_adv,style='Pos',hue='Tm',legend='full',s=100)
plt.text(30.3+0.1, 0.636, "Joel Embiid", horizontalalignment='left', size='medium', color='black')
plt.text(31.3+0.1, 0.647, "Nikola Jokic", horizontalalignment='left', size='medium', color='black')
plt.text(17.1+0.1, 0.736, "DeAndre Jordan", horizontalalignment='left', size='medium', color='black')
plt.text(25.7+0.1, 0.719, "Robert Williams", horizontalalignment='left', size='medium', color='black')
plt.title("PER vs TS%", fontsize = 20)
```



TS% is True Shooting Percentage; the formula is  $PTS / (2 * TSA)$ . True shooting percentage is a measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws (Basketball Reference).

This graph showcases which players have been the most efficient while also being impactful. But as we can see from the outliers this gives a lot of credit to players who are efficient but a player can be efficient and not be very impactful for the team.

Centres tend to be very efficient as they take high percentage shots close to the basket.

Hence we can see DeAndre Jordan and Robert Williams as outliers. But the MVP Jokić has a high PER while also maintaining a good TS%. This observation is key.

## ● WS vs BPM -

```
sns.set(rc = {'figure.figsize':(20,20)})
sns.scatterplot('WS', 'BPM', data=nba_adv, style='Pos', hue='Tm', legend='full', s=100)
plt.text(15.6+0.1, 11.7, "Nikola Jokic", horizontalalignment='left', size='medium', color='black')
plt.text(10.2+0.1, 8.8, "Giannis Antetokounmpo", horizontalalignment='left', size='medium', color='black')
plt.title("WS vs BPM", fontsize = 20)
```



No graph shows how impactful Nikola Jokic was to the Denver Nuggets than this one. We can see clearly how much of an outlier Jokic is and how much of a boost he gives to his team when he is on the floor. This solidifies the MVP's run as one of the most impressive feat by a centre in the NBA.

**Conclusion:**

Advanced metrics and statistics in the NBA were created to better showcase a player's value and derive how influential the player is to the team's success. Though not well received by the masses these stats provide the data analysts who work for individual teams and the league itself a better understanding of a player's playstyle and prowess.

- Nikhil Thomas A  
2139446

## Works Cited

- Alpaydin, Umut. "Nba 2020-2021 Season Player Stats." *Kaggle*, 2021,  
<https://www.kaggle.com/umutalpaydn/nba-20202021-season-player-stats>.
- Basketball Reference. "Glossary." *Basketball Reference*,  
<https://www.basketball-reference.com/about/glossary.html>.
- Basketball Reference 2. "2020-21 NBA Season Summary." *Basketball Reference*,  
[https://www.basketball-reference.com/leagues/NBA\\_2021.html#all\\_all\\_awards](https://www.basketball-reference.com/leagues/NBA_2021.html#all_all_awards).
- Horning, Clay. "VORP may settle the debate between Michael, LeBron." *The Daily Item*,  
2020,  
[https://www.dailyitem.com/sports/vorp-may-settle-debate-between-michael-lebron/article\\_bad6b808-a422-5c3f-90d4-935adfc26ae5.html](https://www.dailyitem.com/sports/vorp-may-settle-debate-between-michael-lebron/article_bad6b808-a422-5c3f-90d4-935adfc26ae5.html).
- NBA Stuffer. "Per-Minute Ratings." *NBA Stuffer*,  
<https://www.nbastuffer.com/analytics101/per-minute-ratings/>.
- Tillman, Johnathan. "TV Ratings Are Down, but the NBA's Popularity Isn't SHARE." *Boardroom*, 09 07 2021, <https://boardroom.tv/nba-popularity-tv-ratings/>.
- Dataset from: <https://www.basketball-reference.com/>