Wil Sheffield

2 May 2024

Estimating Top Fifty NBA Player Salaries

Since its founding on June 6, 1946, the National Basketball Association has risen in prominence to be the second most popular sporting association in the United States behind the NFL, averaging 1.6 million viewers per game in 2023 (Karp). Furthermore, in 2023, the NBA garnered a whopping 4.5 billion US dollars in revenue for the 2023 season (Gough). Given the all-time high popularity of the sport and the money it generates, NBA players, especially those in the top tier of skill and viewership, are being handsomely compensated for their performance and involvement. Like any sport in the modern world, there are hundreds of statistics created and utilized for a multitude of purposes, whether it be for predicting the outcome of games, assessing the effectiveness of individual players or teams, or even for the simple enjoyment of the statistician. With that being said, can these statistics be used in a multivariable linear regression to estimate the salary of these individual players based on their in-game averages, attributes such as age and experience, and categorical variables such as position or whether or not they play for a big market team?

To begin this modeling, it was necessary to gather a sample of players that would have the most available statistics for their performance, coverage about their attributes, and significant salaries. With that being said, the most fitting (and enjoyable) group of players to assess were the most skilled and well-known players in the league. Because it is ultimately subjective who is "most skilled" or "well-known", ESPN's NBArank 2023 was used to determine which individuals qualified for the set of fifty observations, with their top fifty players for the 2023-2024 season being included on the list (ESPN NBArank Panel). By selecting a sample like this, the regression performed would be more likely to have significance in its modeling as this group of players commands significantly higher salaries compared to the rest of the player population, ultimately leading to more robust regression results.

With the dependent variable in this regression model being the annual salary of the individual players in US dollars, the most difficult part of this regression was determining what to select as independent variables. Given that this analysis was performed using Python (Pandas, NumPy, SciPy, and Statsmodels), there was no strict limit on the number of independent variables, allowing for a more flexible model-building process. To determine which variables were most relevant to an NBA player's value for a team (which is measured by their salary), it was necessary to think like a general manager of an NBA team. Conclusively, it was boiled down to the fact that an NBA player's value can be vastly measured by their ability to produce wins for their team and generate viewership and revenue for the brand of their team. After conducting thorough research about what statistics, attributes, and categorical variables most affect these measures of value, a list of twenty-five different independent variables was created. Using Python, a series of regression models were tested with different combinations of these independent variables to determine which combination produced the highest R^2 and greatest degree of statistically significant coefficients.

After this selection process was completed, the sixteen most significant independent variables were selected. Firstly, four independent variables relating to per-game statistics were selected: points per game (PPG), assists per game (APG), rebounds per game (RPG), and blocks plus steals per game (BSPG). Each of these is an extremely common metric to determine the value of a player to an individual game on average, with points per game, for example, being determined by averaging out the total points a player scored in the 2023 season by the number of games they played in the season. Secondly, three advanced statistics were selected as independent variables: true shooting percentage (TS%), player efficiency rating (PER), and win shares per 48 minutes (WS/48). True shooting percentage takes into account a player's field goal percentage, three-point field goal percentage, and free throw percentage in order to measure how effectively a player shoots the ball. Player efficiency rating is a rating system in which a player can be evaluated based on their performance per minute in a game, taking into account essentially all major individual positive and negative in-game statistics. Win shares per 48 minutes is a more recent advanced metric that estimates the number of wins a player contributes to their team per 48 minutes. Since these advanced statistics were initiated for use by the NBA in 1996, they have provided tremendous value to all facets of the game relating to numbers and predictions. Thirdly, five independent variables were used to account for various numerical attributes of an NBA player relevant to their value: age (in years), seasons played, draft position (first overall pick, etc.), All-Star game selections, and games played (out of 82 total in the 2023-2024 season). Each of these carries extreme relevance to the valuation of a player. For example, younger players have more potential for growth and are less injury-prone, but older and more experienced players may have more All-Star selections or knowledge and familiarity with the game, providing value in that manner. Finally, it was necessary to create four independent variables for three different categorical variables relating to an NBA player: position played (one dummy variable for if they are a guard and one for if they are a forward, with zeroes for both indicating that they are a center), big market (one dummy variable for if they play a team classified as a “big market”), and rookie deal (one dummy variable for if they are on a rookie deal, which is essentially a forced, smaller contract). While these variables were not initially numerical, they were essential to a proper linear regression for estimating NBA player salary as they individually play a large factor in the salary of a player (centers are typically paid less, players in big markets are typically paid more, and players on rookie deals are paid less).

After selecting the independent and dependent variables for the model, the final multivariable linear regression was performed using Python, leveraging Pandas for data handling and Statsmodels for regression analysis. The sixteen selected independent variables were used for each of the top fifty NBA players, with their 2023-2024 season salary as the dependent variable. For all independent variables besides "Rookie Deal", Basketball-Reference.com, one of the most trusted and detailed NBA player statistics and attribute databases, was referred to (Basketball-Reference.com). For the "Rookie Deal" independent variable and the 2023-2024 salary dependent variable, Spotrac's NBA salary breakdown database was used (Spotrac). After running the regression in Python on all fifty observations, using Statsmodels to compute coefficients and statistical significance, the following regression equation was achieved: Salary (USD) = -41300626.96 + 535388.5341*(PPG) + 158542.7126*(APG) + 519145.9563*(RPG) + 1784980.041*(BSPG) + 373131.5819*(TS%) - 167673.6576*(PER) - 29423770.38*(WS/48) + 1901095.983*(Age) - 1128283.826*(Season) + 462.7209591*(Draft Position) + 773311.6999*(All-Stars) - 54318.16505*(Games Played) + 3204265.194*(Guard Position Dummy) - 937602.4402*(Forward Position Dummy) - 3362414.305*(Big Market Dummy) - 20730454.34*(Rookie Deal Dummy). For the purpose of interpreting this equation, the coefficients can be explained as follows: For every All-Star selection a player has on their resume, they can expect to earn approximately $773,311 more in USD for their salary. Using Python's Statsmodels package, the model achieved an R^2 value of approximately 0.824, indicating that approximately 82.4% of the variability in salary was accounted for by the sixteen independent variables. Furthermore, the model had an F-statistic p-value of approximately

2.88265E-08, suggesting that the overall regression was statistically significant. Given the relatively high R^2 and statistical significance, it is important to dive into some of the residual analysis. When plotting the residuals for each individual variable, no pattern of residuals was observed for any of them, indicating that, in general, a linear regression model was appropriate for the data (Refer to Appendix).

Unfortunately, the individual statistical significance of each independent variable was poor. When using the conventional alpha value of 0.05, only the "Rookie Deal" independent variable came up as statistically significant, with a p-value of 1.36281E-05. Intuitively, this makes sense that the "Rookie Deal" variable would be statistically significant as a player on a rookie deal is ultimately limited to a maximum salary of around 12 million USD (regardless of talent, production, or potential), significantly less than other top fifty NBA players not on rookie deals making as much as 50 million USD annually. The "Age" and "Big Market" independent variables were the two closest to being statistically significant, with p-values of 0.072795013 and 0.109612342, respectively (Refer to Appendix). With that being said, in every case except for the "Rookie Deal" independent variable, one would have to accept the null hypothesis that the coefficient for the said independent variable is not statistically greater than zero in the regression. Although, as aforementioned, various combinations of twenty-five different independent variables were tested using Python's Statsmodels for multivariable linear regression, this was still ultimately the combination that produced the highest R^2 in the model. Using Python's Seaborn and Matplotlib for visualization, residual analysis was conducted for each independent variable. It was concluded that much of the lack of statistical significance in the other fifteen independent variables can be explained by a high degree of multicollinearity. For example, a player who averages many points per game will likely have the ball in their hands a significant amount, resulting in them generating more assists. Correspondingly, a player with a higher true shooting percentage will likely average more points per game, as they are more efficient with their shot-making ability. Additionally, a younger player will naturally tend to have fewer seasons played, and vice versa. Furthermore, a player who averages higher individual statistics or has played more seasons will have more All-Star game selections. As mentioned before, advanced statistics such as player efficiency ratings are based on some of the other independent variables, such as per-game statistics. While this is a short list of examples of collinearity, basketball is ultimately a very complex and interconnected sport in regards to individual statistics and the value they create for an individual player. With that being said, in order to obtain a set of entirely statistically significant independent variables, this regression would likely have had to be performed on an entirely different phenomenon, simply because it is nearly impossible to garner independent variables related to basketball that have no relationship or dependability on one another.

All in all, this regression model does a solid job of estimating the salary of the top fifty NBA players in the 2023-2024 season. With that being said, there are clearly limitations to it, with much of it relating to the collinearity of the independent variables. Furthermore, to better value a player's salary, a model would need to find a way to account for abstract qualities of a player such as leadership ability or marketability for a team’s brand. As mentioned before, given the complex nature of basketball and the interdependence of the majority of the variables that affect a player’s value, a different model type than a linear regression may be necessary to account for this significant presence of collinearity. At the end of the day, this was a very fun topic to focus on given my strong interest in basketball. The transition from Excel to Python allowed for a more robust and scalable analysis, but a legitimate model used to predict the salary of these players would require a way to account for the collinearity of the independent variables as well as a method of measuring abstract attributes for individual players.

Regression Appendix

Player	PPG	APG	RPG	BSPG	TS%	PER	WS/48	Age	Seasons	Draft Pos	All Stars	Games Played	Guard	Forward	Big Market	Rookie Deal	Salary
Giannis Antetokounmpo	30.4	6.5	11.5	2.3	64.9	29.9	0.246	29	10	15	8	73	0	1	0	0	$45,640,084.00
Nikola Jokic	26.4	9	12.4	2.3	65	31	0.299	28	9	41	6	79	0	0	0	0	$47,607,350.00
Joel Embiid	34.7	11	5.6	2.9	64.4	34.1	0.275	29	8	3	7	39	0	0	1	0	$47,607,350.00
Luka Doncic	33.9	9.8	9.2	1.9	61.7	28.1	0.22	24	6	3	5	70	1	0	1	0	$40,064,220.00
Stephen Curry	26.4	5.1	4.5	1.1	61.6	20.6	0.142	35	14	7	10	74	1	0	1	0	$51,915,615.00
Jayson Tatum	26.9	4.9	8.1	1.6	60.4	22.3	0.189	25	7	3	5	74	0	1	1	0	$32,600,060.00
Kevin Durant	27.1	5	6.6	2.1	62.6	21.2	0.142	35	16	2	14	75	0	1	0	0	$47,649,433.00
Shai Gilgeous-Alexander	30.1	6.2	5.5	2.9	63.6	29.3	0.275	25	6	11	2	75	1	0	0	0	$33,386,850.00
LeBron James	25.7	8.3	7.3	1.8	63	23.7	0.164	39	21	1	20	71	0	1	1	0	$47,607,350.00
Anthony Davis	24.7	3.5	12.6	3.5	62.1	25.8	0.21	30	12	1	9	76	0	0	1	0	$40,600,080.00
Devin Booker	27.1	6.9	4.5	1.3	61.1	21.9	0.181	27	9	13	4	68	1	0	0	0	$36,016,200.00
Jimmy Butler	20.8	5	5.3	1.6	62.6	22	0.214	34	13	30	6	60	0	1	0	0	$45,183,960.00
Anthony Edwards	25.9	5.1	5.4	1.8	57.5	19.7	0.13	22	4	1	2	79	1	0	0	1	$13,534,817.00
Damian Lillard	24.3	7	4.4	1.2	59	19.6	0.143	33	12	6	8	73	1	0	0	0	$45,640,084.00
Donovan Mitchell	26.6	6.1	5.1	2.3	59.5	22.1	0.159	27	7	13	5	55	1	0	0	0	$32,600,060.00
Bam Adebayo	19.3	3.9	10.4	2	57.6	19.8	0.144	26	7	14	3	71	0	0	0	0	$32,600,060.00
Jamal Murray	21.2	6.5	4.1	1.7	58.6	20.7	0.153	26	7	7	0	59	1	0	0	0	$33,833,400.00
Paul George	22.6	3.5	5.2	2	61.3	19.3	0.14	33	14	10	9	74	0	1	1	0	$45,640,084.00
Jaylen Brown	23	3.6	5.5	1.7	58	18.6	0.121	27	8	3	3	70	0	1	1	0	$31,830,357.00
Karl Anthony-Towns	21.8	3	8.3	1.4	62.5	18.8	0.148	28	9	1	4	62	0	1	0	0	$36,016,200.00
Tyrese Haliburton	20.1	10.9	3.9	1.9	60.5	23.3	0.195	23	4	12	2	69	1	0	0	1	$5,807,435.00
Domantas Sabonis	19.4	13.7	8.2	1.5	63.7	23.2	0.206	27	8	11	3	82	0	0	0	0	$28,000,000.00
De'Aaron Fox	26.6	5.6	4.6	2.4	56.7	20.1	0.117	26	7	5	1	74	1	0	0	0	$32,600,060.00
Kawhi Leonard	23.7	3.6	6.1	2.5	62.6	23.2	0.184	32	12	15	6	68	0	1	1	0	$45,640,084.00
Pascal Siakam	21.7	4.3	7.1	1.1	60.1	19.7	0.124	29	8	27	2	80	0	1	0	0	$37,893,408.00
Jrue Holiday	12.5	4.8	5.4	1.7	59.7	14.4	0.133	33	15	17	2	69	1	0	1	0	$34,954,667.00
Brandon Ingram	20.8	5.7	5.1	1.4	57.8	18.2	0.117	26	8	2	1	64	0	1	0	0	$33,833,400.00
Lauri Markkanen	23.2	2	8.2	1.4	63.1	21.5	0.163	26	7	7	1	55	0	1	1	0	$17,259,999.00
Trae Young	25.7	10.8	2.8	1.5	58.5	20.3	0.114	25	6	5	3	54	1	0	1	0	$40,064,220.00
Paolo Banchero	22.6	5.4	6.9	1.5	54.6	17.3	0.09	21	2	1	1	80	0	1	0	1	$11,608,080.00
Jaren Jackson Jr.	22.5	2.3	5.5	2.8	55.2	17.4	0.067	24	6	4	1	66	0	0	0	0	$27,102,202.00
Jalen Brunson	28.7	6.7	3.6	1.1	59.2	23.4	0.198	27	6	33	1	77	1	0	1	0	$26,346,666.00
Mikal Bridges	19.6	3.6	4.5	1.4	56	14.9	0.07	27	6	10	0	82	0	1	1	0	$21,700,000.00
Kyrie Irving	25.6	5.2	5	1.8	60.8	21.9	0.163	31	13	1	8	58	1	0	1	0	$37,037,037.00
Ja Morant	25.1	8.1	5.6	1.4	57	20.6	0.124	24	5	2	2	9	1	0	0	0	$34,005,250.00
Darius Garland	18	6.5	2.7	1.4	56	14.5	0.067	24	5	5	1	57	1	0	0	0	$34,005,250.00
Bradley Beal	18.2	5	4.4	1.5	60.7	16.3	0.107	30	12	3	3	53	1	0	0	0	$46,741,590.00
Zach Lavine	19.5	3.9	5.2	1.1	57.8	15.1	0.08	28	10	13	2	25	1	0	1	0	$40,064,220.00
DeMar Derozan	24	5.3	4.3	1.7	58.4	19.7	0.147	34	15	9	6	79	0	1	1	0	$28,600,000.00
Evan Mobley	15.7	3.2	9.4	4.6	62.6	20.1	0.174	22	3	3	0	50	0	1	0	1	$8,882,640.00
Klay Thompson	17.9	2.3	3.3	1.1	57.6	13.9	0.075	33	11	11	5	77	1	0	1	0	$43,219,440.00
Tyrese Maxey	25.9	6.2	3.7	1.5	57.3	19.8	0.147	23	4	21	1	70	1	0	1	1	$4,343,920.00
James Harden	16.6	8.5	5.1	1.9	61.2	18.6	0.163	34	15	3	10	72	1	0	1	0	$35,640,000.00
CJ McCollum	20	4.6	4.3	1.5	59.2	17.9	0.132	32	11	10	0	66	1	0	0	0	$35,802,469.00
Julius Randle	24	5	9.2	0.8	56.9	18.9	0.113	29	10	7	3	46	0	1	1	0	$25,660,800.00
Khris Middleton	15.1	5.3	4.7	1.2	59.5	17	0.109	32	12	39	3	55	0	1	0	0	$28,703,704.00
Victor Wembanyama	21.4	3.9	10.6	4.8	56.5	23.1	0.085	20	1	1	0	71	0	0	0	1	$12,160,680.00
LaMelo Ball	23.9	8	5.1	2	56.1	20.8	0.074	22	4	3	1	22	1	0	0	1	$10,900,000.00
Deandre Ayton	16.7	1.6	11.1	1.8	58.7	18.9	0.106	25	6	1	0	55	0	0	0	0	$32,459,438.00
Jarrett Allen	16.5	2.7	10.5	1.8	66.4	21.8	0.21	25	7	22	1	77	0	0	0	0	$20,000,000.00

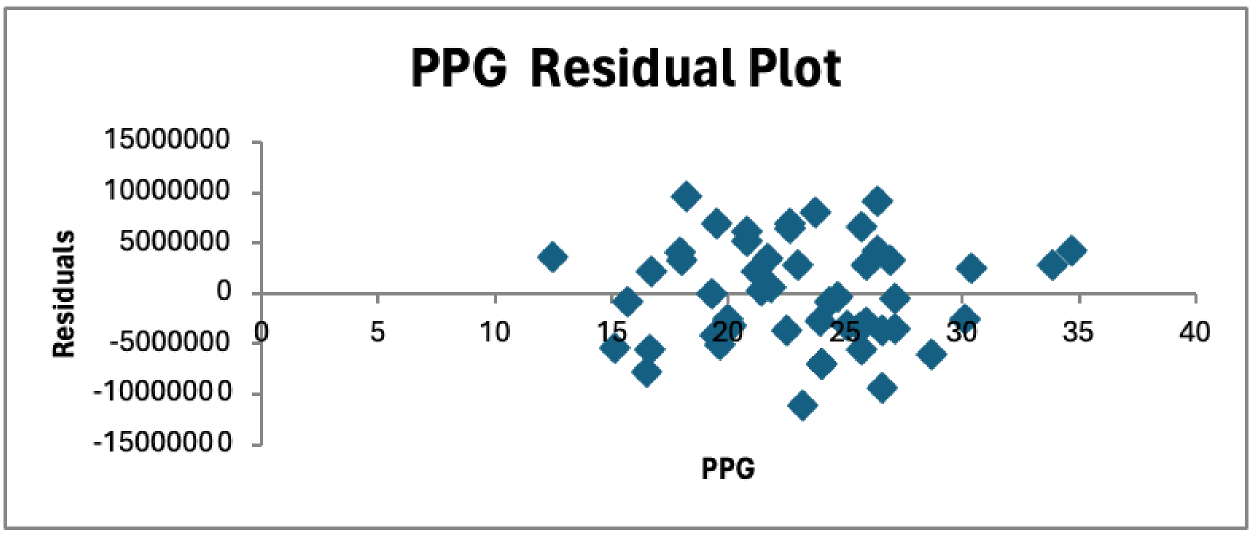
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.90783324
R Square	0.82416119
Adjusted R Sq	0.73890602
Standard Err	6200906.05
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	5.9473E+15	3.7171E+14	9.66699276	2.8826E-08
Residual	33	1.2689E+15	3.8451E+13		
Total	49	7.2162E+15			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-41300627	51632094.7	-0.7999022	0.42948825	-146346914	63745659.6	-146346914	63745659.6
PPG	535388.534	631629.267	0.84763098	0.40274959	-749670.87	1820447.94	-749670.87	1820447.94
APG	158542.713	614137.187	0.25815521	0.79789038	-1090928.8	1408014.21	-1090928.8	1408014.21
RPG	519145.956	679507.343	0.76400345	0.45029362	-863322.13	1901614.04	-863322.13	1901614.04
BSPG	1784980.04	2003852.63	0.89077411	0.37949909	-2291888.8	5861848.86	-2291888.8	5861848.86
TS%	373131.582	790528.94	0.47200243	0.64003134	-1235211.6	1981474.8	-1235211.6	1981474.8
PER	-167673.66	1224779.99	-0.136901	0.89194085	-2659507.3	2324159.96	-2659507.3	2324159.96
WS/48	-29423770	59131877.4	-0.4975957	0.62206864	-149728480	90880938.8	-149728480	90880938.8
Age	1901095.98	1025776.29	1.85332416	0.07279501	-185861.58	3988053.54	-185861.58	3988053.54
Seasons	-1128283.8	1097750.2	-1.0278147	0.3115121	-3361673.4	1105105.74	-3361673.4	1105105.74
Draft Pos	462.720959	126395.557	0.0036609	0.99710108	-256690.97	257616.414	-256690.97	257616.414
All Stars	773311.7	624847.277	1.23760114	0.22459772	-497949.64	2044573.04	-497949.64	2044573.04
Games Playe	-54318.165	69806.425	-0.7781256	0.4420386	-196340.4	87704.0745	-196340.4	87704.0745
Guard	3204265.19	4330721.65	0.73989174	0.4645972	-5606654.2	12015184.6	-5606654.2	12015184.6
Forward	-937602.44	3686063.84	-0.2543641	0.80079264	-8436955.7	6561750.82	-8436955.7	6561750.82
Big Market	-3362414.3	2044911.65	-1.6442834	0.10961234	-7522818.3	797989.734	-7522818.3	797989.734
Rookie Deal	-20730454	4062634.8	-5.1027118	1.3628E-05	-28995947	-12464962	-28995947	-12464962



Works Cited

Basketball-Reference.com. "NBA Player Directory." *Basketball-Reference.com*, Sports Reference, www.basketball-reference.com/players/. Accessed 1 May 2024.

ESPN NBArank Panel. "NBArank 2023: Player Rankings for 2023-24, from 50 to 1." *ESPN*, ESPN Enterprises, 11 Oct. 2023, www.espn.com/nba/story/_/id/38625043/nbarank-2023-player-rankings-2023-24-50-11. Accessed 1 May 2024.

Gough, Christina. "Revenue of the NBA in North America in 2023, by Type." *Statista*, 11 Apr. 2024, www.statista.com/statistics/1459408/nba-revenue-by-type/#:~:text=In%20the%202023%20season%2C%20roughly. Accessed 2 May 2024.

Karp, Austin. "NBA National Viewership up Slightly Heading into Final Months." *Sports Business Journal*, Leaders Group, 26 Feb. 2024, www.sportsbusinessjournal.com/Articles/2024/02/26/nba-viewership. Accessed 1 May 2024.

Spotrac. "NBA Contracts." *Spotrac*, USA Today Sports Media Group, www.spotrac.com/nba/contracts. Accessed 2 May 2024.