

# NBA Salaries: How Much is that Buzzer Beater Worth?

Eddie and Co.

## Introduction

Professional sports salaries, particularly in the NBA, continue to spark discussion because they highlight the intersection of performance, marketability, and team strategy. Understanding how a player’s on-court production translates to salary is not only valuable for analysts and researchers, but also for front offices making contract decisions and fans seeking to understand the economics of the league. Prior research suggests that player compensation is strongly influenced by both measurable performance statistics and more subjective forms of recognition such as awards or team honors (Berri & Schmidt, 2006; Stiroh, 2007; Simmons & Berri, 2011).

Motivated by these findings, our project examines the statistical relationships between NBA player characteristics and their salaries during the 2024-2025 season. We selected three primary variables—points per game, defensive stocks (steals + blocks), and award recognition—because each represents a different dimension of player value. Scoring is one of the most visible and heavily rewarded skills in basketball (Stiroh, 2007). Defensive performance, while often overlooked publicly, provides meaningful team impact and has shown measurable influence on team success (Simmons & Berri, 2011). Awards represent league-wide acknowledgment of excellence and can shape a player’s reputation and contract negotiations (Berri & Schmidt, 2006).

Our general research question is:

**What factors best explain variation in NBA player salary during the 2024-2025 season?**

To address this, we developed three specific research questions:

- Do NBA players who score more points (PTS) earn higher salaries on average?
- Is there a meaningful positive relationship between a player’s defensive activity (“stocks”) and their salary?
- Do players who have earned at least one major NBA award receive higher salaries compared to players with no awards?

Together, these questions reflect distinct aspects of performance and recognition. By analyzing these variables, we aim to identify which factors are most strongly associated with salary and how our findings compare to patterns documented in the sports-economics literature.

## Data Summary

### Data Sources

Our dataset contains information on NBA players from the 2024-2025 season and includes a mixture of quantitative performance measures, qualitative awards indicators, demographic information, and salary data. In total, the dataset represents every player who appeared in the NBA during the season, giving us a sample that is both large and comprehensive. The primary response variable is player salary, measured in U.S. dollars.

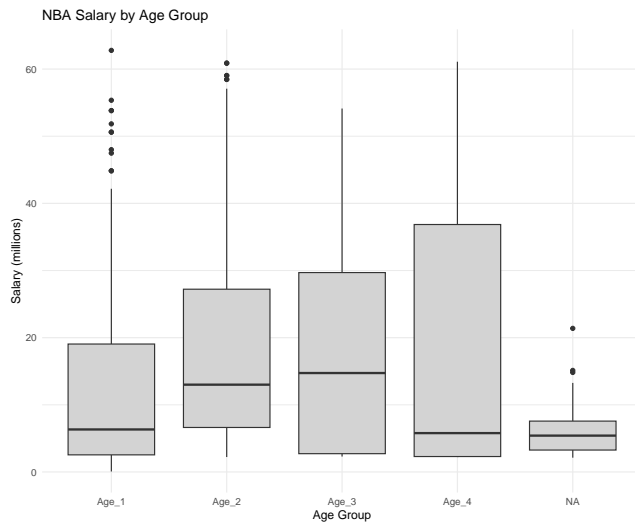
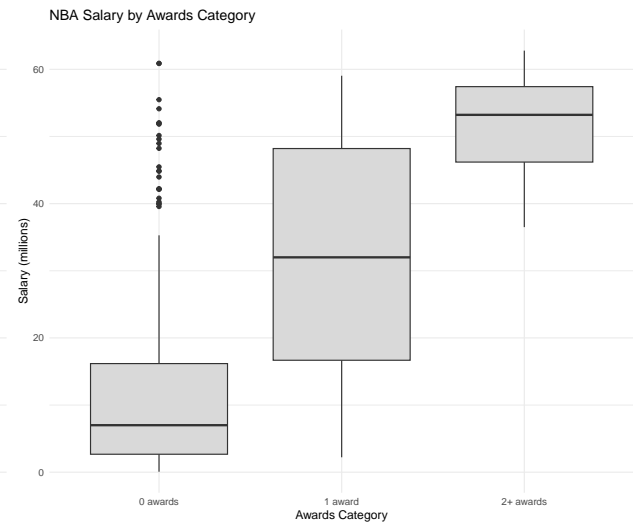
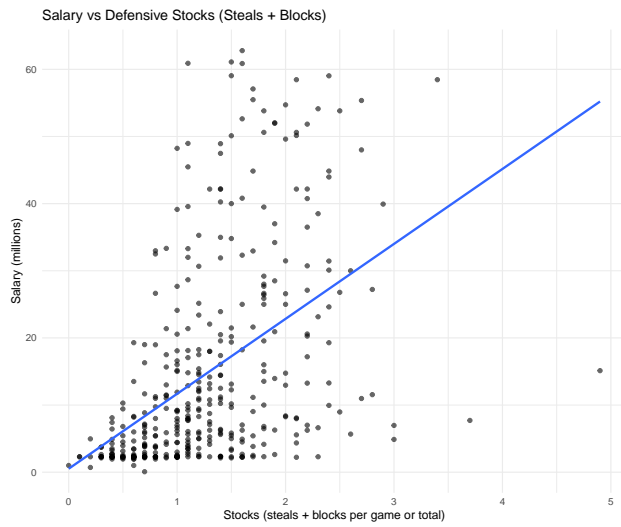
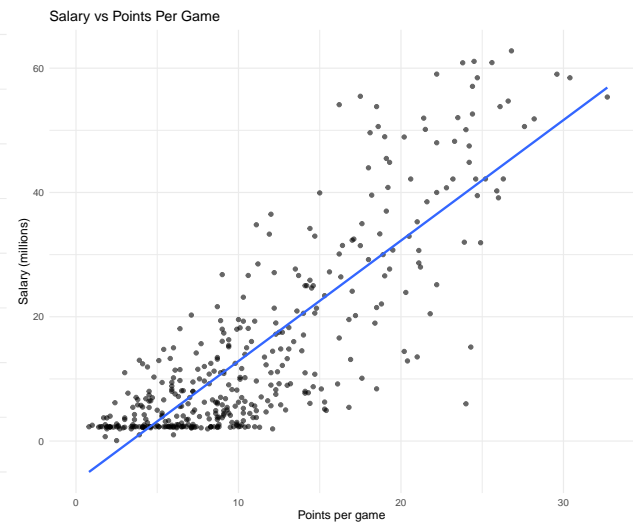
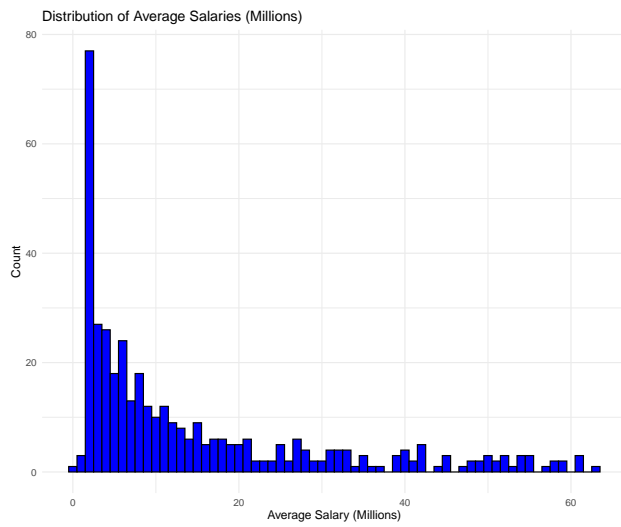
The data in this project were compiled from publicly available and reputable basketball statistics platforms. Performance statistics, biographical information, and award histories were collected from Basketball Reference, a database known for its accuracy, professional documentation, and long-standing use in academic sports analysis. To supplement player-level statistics, we incorporated team valuation data from Statista, using their report on NBA franchise values. Statista compiles these estimates from league financial disclosures, public filings, and market analyses, making it a reliable secondary source commonly used in academic and business research. Including franchise value provides useful economic context, as team financial strength and market size may influence payroll structures and player salaries. Together, these sources provide a complete picture of each player's performance and compensation for the 2023–2024 season.

To prepare the data for analysis, we standardized naming conventions across the two sources and combined the datasets so that each player occupied a single row with salary, performance, and awards information. Awards were originally stored in long strings listing placements, team selections, and voting outcomes, so we simplified these into a three-level qualitative variable representing the general number of accolades received. Age was also simplified into qualitative brackets to match our research questions.

We ensured that our final dataset only included players who logged minutes during the 2024-2025 season, as these represent meaningful salary–performance comparisons. Players on two-way contracts or who appeared in only a few games remained in the dataset to preserve the natural distribution of playing time and contract structures.

Overall, the combined dataset contains a rich mixture of performance and recognition indicators, allowing us to meaningfully evaluate how both quantitative and qualitative factors relate to NBA salary. While subjective elements such as awards voting and contract timing may influence individual observations, the breadth and reliability of the sources allow us to confidently proceed with statistical analysis.

# Exploratory Data Analysis



## EDA Summary

To understand how performance and recognition relate to NBA salaries, we first examined the distribution of average salary. Salaries are heavily right-skewed: most players earn under \$10 million, while a small group reaches \$40–\$60 million. The mean (\$9.0M) exceeds the median (\$5.4M), reflecting this skew. Despite the imbalance, salary has sufficient spread for multiple linear regression, with the option of a log transformation if needed.

Performance variables show clear trends. Scoring has the strongest relationship with salary—players scoring 20+ points per game often earn \$35–\$45 million, compared to under \$5 million for low scorers. Defensive output, measured by stocks (steals + blocks), displays a more moderate positive trend, indicating that defense influences valuation but less strongly than scoring.

Award recognition corresponds with substantial salary differences. Players with no awards earn around \$8 million, whereas those with one award jump to roughly \$50 million, and those with two or more to about \$55 million. This suggests that accolades capture aspects of player reputation beyond raw statistics.

We also examined salary across age groups to contextualize performance and recognition within career arcs. Players aged 27–31 earn the highest median salaries, consistent with peak physical and skill development. Younger players (22–26) and older veterans (35+) earn less on average, reflecting typical progression and decline patterns within NBA careers.

Before proceeding to modeling, we evaluated potential multicollinearity among the predictors. Although scoring may correlate with other offensive statistics, our chosen variables—PTS, stocks, Awards, and Age—do not appear strongly correlated based on initial inspection. We will confirm this formally using VIFs during model construction.

Overall, the exploratory analysis reveals several consistent patterns:

- salaries are right-skewed with a superstar-driven upper tail
- scoring is strongly tied to pay
- defensive production shows a moderate positive association
- awards correspond to substantial salary increases
- peak earnings occur mid-career.

These insights justify each chosen predictor and establish a clear foundation for building our multiple linear regression model.

## Methods and Analysis

### Assessing Multicollinearity

We began our analysis by assessing multicollinearity among predictors. From our EDA correlation heatmap, we observed potential overlap between PTS and other offensive statistics such as FTA and MP. We quantified these relationships using pairwise correlations and VIFs. If VIFs fell below standard thresholds (individual  $< 10$ , average  $< 3$ ), we proceeded; otherwise, we used variable screening and stepwise regression to identify a reduced subset of predictors. After selecting a more stable set of variables, we reassessed multicollinearity before model building.

### Building the Model

We will begin building our model by estimating our model parameters, starting with only our quantitative predictors and a quantitative interaction term that we believe may contribute to the model. Our initial model using quantitative predictors is as follows:

$$\begin{aligned}\text{Salary} = & \beta_0 + \beta_1 G + \beta_2 MP + \beta_3 PTS + \beta_4 FG\% + \beta_5 FTA \\ & + \beta_6 TRB + \beta_7 AST + \beta_8 \text{Stocks} + \beta_9 \text{Value} \\ & + \beta_{10}(PTS \times FTA)\end{aligned}$$

We estimated this model with `lm()` and evaluated overall utility using a Global F-Test. Significant predictors—including the interaction—were identified through individual t-tests. Non-significant predictors were removed, and the updated model was rechecked with another Global F-Test.

Next, we incorporated qualitative predictors and one interaction ( $\text{Awards} \times \text{Age}$ ). We again performed a Global F-Test, followed by t-tests for the interaction and a Nested F-Test for the Awards factor. Only significant categorical effects were retained. No quantitative  $\times$  qualitative interactions were considered. Once the final set of predictors was selected, we confirmed model adequacy with a concluding Global F-Test.

### Assessing the Model (Including Cross Validation)

Once we've finalized our model through variable testing, we conducted a holistic assessment of our model for a better understanding of its capabilities and limitations. We evaluated metrics such as  $R^2$ , Adjusted  $R^2$ , and  $\sqrt{MSE}$  to evaluate our model. Separately, interpreted confidence and prediction intervals to ensure our model performs on a practical level.

In addition to the traditional assessment metrics, we incorporated Cross Validation into our

final project. Cross validation is a machine learning technique used to assess the performance and accuracy of a model. Performing Cross validation allows the model builder to see whether their model overfits or underfits by training the majority of a data set, and using the remaining data for testing. The trained set is then used to estimate a linear regression model evaluated on metrics such as  $\sqrt{MSE}$  and  $R^2$ . More specifically, we will use the Validation Set Approach in R, which randomly splits 80% of the dataset into training, and 20% into testing. By performing Cross Validation we will have a strong idea of whether our model is prepared to be used on new data, or if further adjustments must be made.

## Checking Model Assumptions

We evaluated model assumptions using residual diagnostics. Residual vs. fitted plots, QQ plots, histograms, and lack-of-fit assessments enabled us to check constant variance, normality, independence, and overall fit. We then examined outliers and influential points through Cook's Distance, leverage, Studentized Residuals, and Deleted Studentized Residuals. If any observations exerted undue influence, we considered removing them to improve model integrity.

## Results

### 1. Required Analysis Steps

#### 1.1 Initial Model Building

We began by fitting a full multiple linear regression model using all quantitative predictors—minutes played (MP), points (PTS), field goal percentage (FG\_pct), free-throw attempts (FTA), rebounds (TRB), assists (AST), defensive stocks, team value, and the interaction term between PTS and FTA. The initial model explained approximately 71.9% of salary variation (adj.  $R^2 = 0.711$ ), and identified several significant predictors, including PTS, FTA, TRB, AST, and the  $PTS \times FTA$  interaction. However, variance inflation factors (VIFs) indicated severe multicollinearity, particularly for PTS, FTA, and the interaction term (VIFs  $> 15$ ), suggesting that the model required refinement.

#### 1.2 Addressing Multicollinearity

Because high VIFs can destabilize estimates and inflate standard errors, we removed the interaction term and refit the model. VIFs decreased substantially, though PTS and FTA remained moderately correlated. This motivated further variable screening techniques to identify a more stable subset of predictors.

#### 1.3 Stepwise Selection (Forward + Backward)

We applied stepwise regression using both AIC and significance-based entry/removal criteria. The algorithm sequentially selected:

1. PTS
2. AST
3. TRB
4. MP
5. stocks
6. Value\_Billions

The resulting model achieved adj.  $R^2 \approx 0.700$ , indicating that these six variables captured most of the explanatory power of the quantitative predictors.

#### 1.4 Refined Quantitative Model

T-tests revealed that FG\_pct, FTA, and Value\_Billions were not consistently significant when included alongside other predictors. We therefore fit a simplified model using only the strongest contributors:

$$avg\_salary\_millions \sim MP + PTS + TRB + AST$$

All coefficients were significant at the 5% level, and the model achieved adj.  $R^2 = 0.706$ , comparable to the stepwise model but more interpretable.

#### 2. Added Technique: ANOVA Model Comparison

To incorporate our categorical predictors (awards, age groups, and positions), we used nested model comparison via ANOVA to evaluate whether each group of variables significantly improved model fit.

##### 2.1 Awards

Adding “1 award” and “2+ awards” dummy variables significantly improved the model ( $F = 12.00$ ,  $p < 0.00001$ ). This indicates that award-winning players earn significantly more, even after controlling for performance.

##### 2.2 Age Groups

Adding four age-category indicators also significantly improved fit ( $F = 8.30$ ,  $p < 0.00001$ ). Players aged 27–34 earned substantially more than the youngest group, consistent with peak performance and contract structures.



## 2.3 Player Positions

Adding position dummies (PG, SG, SF, PF) modestly improved the model ( $F = 3.54$ ,  $p = 0.007$ ). Most position effects were small, but shooting guards showed a significantly lower salary relative to the baseline position.

### 3. Model Assessment 3.1 Coefficient Interpretation

In the full quantitative-plus-qualitative model:

PTS remained the strongest predictor.

AST retained a meaningful positive effect.

Awards (2+) added a large salary premium.

Age 27–34 corresponded to significantly higher earnings.

The SG position showed a negative effect relative to the baseline.

### 3.2 Confidence Intervals

95% confidence intervals confirmed:

Large, positive, and stable effects for PTS and AST.

Awards (2+) reliably increased salary (CI entirely above 6.85).

Position effects were weaker, with many CIs crossing zero.

### 3.3 Diagnostic Plots

Figures included:

Cook's Distance (pg. 9): observations 21, 25, 73, 86 appeared influential.

Influence Plot: a few high-leverage points identified, but none requiring removal.

Studentized Residuals Plot (pg. 10): small number of outliers beyond  $\pm 2$ .

Overall, diagnostics indicated no major violations of linear model assumptions.

## 4. Selecting the Final Best Model

We compared models in terms of:

Explanatory power (adj.  $R^2$ )

Statistical significance (ANOVA tests)

Interpretability

Diagnostic behavior

The final chosen model is the full quantitative model combined with awards, age categories, and positions, as it:

Achieved the highest adjusted  $R^2$  (0.7477)

Demonstrated significant improvements when categorical variables were added

Provided interpretable insight into both performance metrics and demographic effects

Passed key diagnostic and influence checks

Final Model:

$$\text{avg\_salary\_millions} \sim MP + PTS + TRB + AST + \text{awards}_1 + \text{awards}_{2+} + \text{Age}_{22-26} + \text{Age}_{27-31} + \text{Age}_{32-34} + \text{Age}_{35+} + \text{Pos}_{PF} + \text{Pos}_{PG} + \text{Pos}_{SF} + \text{Pos}_{SG}$$

This model best balances accuracy, parsimony, and interpretability.

## Conclusions

## Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Games	G	This represents the games played by each respective player.
Minutes Played	MP	This represents the minutes played by each respective player in the overall season.
Points Scored	PTS	This represents the total points scored by each respective player.
Field Goal Percentage	FG_PCT	This represents the percentage of field goals scored by each respective player.
Free Throw Attempts	FTA	This represents the free throw attempts made by each respective player.
Total Rebounds	TRB	This represents the total rebounds made by each respective player.
Total Assists	AST	This represents the total assits by each respective player.
Steals and Blocks Combined	Stocks	This represents the steals and blocks achieved by each respective player.
Franchise Value	Value	This represents the franchise value of the team of each respective player.

## Appendix B: Data Rows

	Player	Team	G	MP	FG_pct	FTA	TRB	AST	stocks	PTS
1	Shai Gilgeous-Alexander	OKC	76	34.2	0.519	8.8	5.0	6.4	2.7	32.7
2	Giannis Antetokounmpo	MIL	67	34.2	0.601	10.6	11.9	6.5	2.1	30.4
3	Nikola Jokić	DEN	70	36.7	0.576	6.4	12.7	10.2	2.4	29.6
4	Luka Dončić	2TM	50	35.4	0.450	7.9	8.2	7.7	2.2	28.2
5	Anthony Edwards	MIN	79	36.3	0.447	6.3	5.7	4.5	1.8	27.6
6	Jayson Tatum	BOS	72	36.4	0.452	6.1	8.7	6.0	1.6	26.8
	Value_Billions	awards_1	awards_2plus		avg_salary_millions			Age_22_26		Age_27_31
1	4.35	0		1			55.3591		1	0
2	4.30	0		1			58.4566		0	1
3	4.60	0		1			59.0331		0	1
4	NA	0		0			51.8379		1	0
5	3.60	0		1			50.6117		1	0
6	6.70	0		1			62.7867		1	0
	Age_32_34	Age_35_plus	Pos_PF	Pos_PG	Pos_SF	Pos_SG	Age		Awards	
1	0	0	0	1	0	0	Age_1	2+	awards	
2	0	0	1	0	0	0	Age_2	2+	awards	
3	0	0	0	0	0	0	Age_2	2+	awards	
4	0	0	0	1	0	0	Age_1	0	awards	
5	0	0	0	0	0	1	Age_1	2+	awards	
6	0	0	1	0	0	0	Age_1	2+	awards	

## Appendix C: Final Model Output and Plots

## Appendix D: References

### Background

Berri, D. J., & Schmidt, M. B. (2006). *The wages of wins*. Stanford University Press.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3(3). <https://doi.org/10.2202/1559-0410.1070>

NBA. (2023). *Collective Bargaining Agreement: Summary of principal terms*. <https://nbpa.com/cba>

Simmons, R., & Berri, D. J. (2011). Mixing the princes and the paupers: Pay and performance in the NBA. *Labour Economics*, 18(3), 381–388. <https://doi.org/10.1016/j.labeco.2010.12.003>

Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, 40(3), 474–494. <https://doi.org/10.2307/2393794>

Stiroh, K. (2007). Playing for keeps: Pay and performance in the NBA. *Economic Inquiry*, 45(1), 145–161. <https://doi.org/10.1111/j.1465-7295.2006.00002.x>

### Data Sources

Statista (NBA franchise values): Statista. (2010). Franchise value of National Basketball Association teams in 2010. <https://www.statista.com/statistics/193696/franchise-value-of-national-basketball-association-teams-in-2010/>

Basketball-Reference (Player contracts): Basketball-Reference.com. (n.d.). NBA player contracts. <https://www.basketball-reference.com/contracts/players.html>

Basketball-Reference (2025 Per-game stats): Basketball-Reference.com. (2025). NBA per game statistics: 2024–25 season. [https://www.basketball-reference.com/leagues/NBA\\_2025\\_per\\_game.html](https://www.basketball-reference.com/leagues/NBA_2025_per_game.html)

### Additional Help