

## NBA Salaries versus Statistics

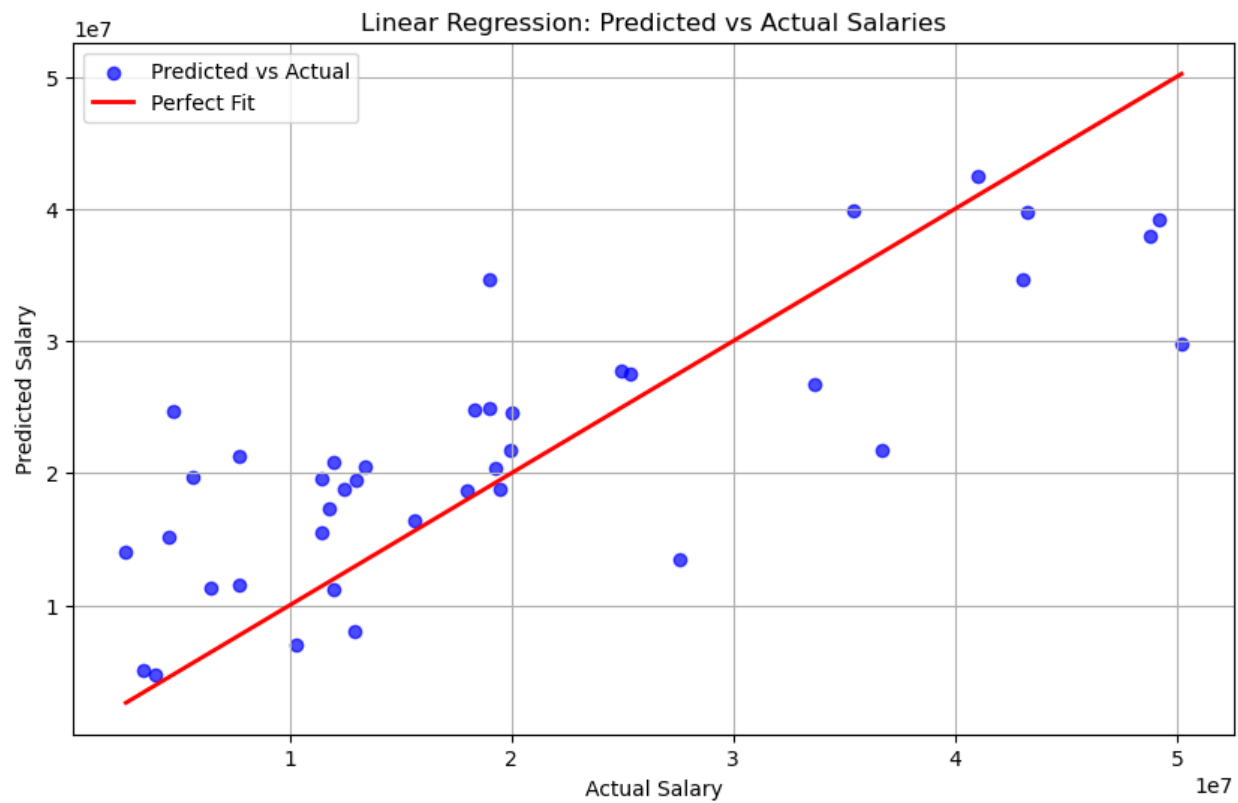
Kevin Trafford

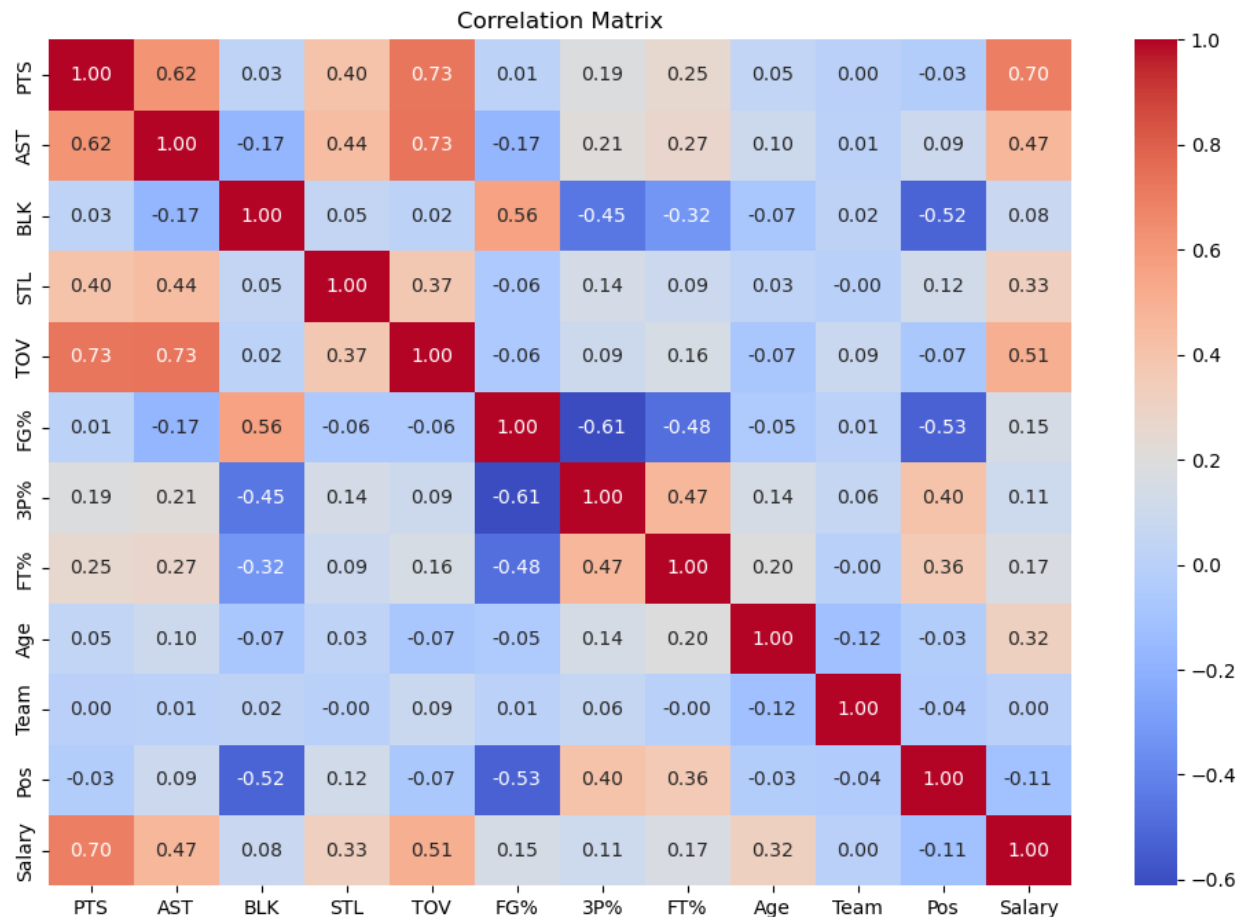
The NBA, or national basketball association, has seen player salaries soar. Players are signing deals worth hundreds of millions of dollars total, and north of 50 million dollars annually. With there being so many different types of players in the NBA, such as three-point specialists, tall post players, defensive wizards, and more, it begs the question: what exactly motivates teams to pay players these outrageous salaries? I predict that points per game is the primary indicator of a player being paid a high salary, however other statistics such as rebounds per game, assists per game, field goal percentage, and more definitely having an impact as well. By taking a sample size of the 200 highest scoring players from the previous NBA season, as well as their respective salaries, I used a wide array of their season long averages to perform a linear regression that attempts to predict their salary. As expected, points per game was the highest predictor of salary, with others playing a lesser role. Furthermore, the model's  $r^2$  value indicated that statistics are predictably a good indicator of how much a player's salary will be.

In order to begin the analysis, the first step was to create a single csv that contained a player's various statistics, as well as their salaries, as these numbers came from different sources. The statistics were from the 2023-2024 season, as at the time of this analysis the 2024-2025 season had not yet concluded. The salaries were from the current year, as the thought process was that players are paid based on their past, not current performance. The goal was to build a model that would take all these statistics and predict what a player would be making in terms of salary. I used an excel function of xLOOKUP to add the salary column to the data set containing statistics. That meant the data would need some cleaning when imported into VS code. The columns used for the xLOOKUP columns needed to be deleted along with the empty columns in

between the two that were unnamed. Furthermore, rows that were missing key statistics needed to be dropped, also some players had not attempted any three-point shots, meaning that their 3P% cells were empty, and needed to be filled with zeroes. Finally, there were two ordinal statistics (Team and Position), and they needed to be encoded using an ordinal encoder. From there, it was clear that linear regression would be the best model for this project as we would be attempting to predict values on a continuous basis. The target variable was salary, and the features were points per game, assists per game, blocks per game, steals per game, turnovers per game, three-point percentage, field goal percentage, free throw percentage, age, team, and position. The data was scaled, and the linear regression was trained, tested, and assessed based on its R-squared value. Furthermore, a correlation matrix with heatmap was used to determine which statistics were having the most impact on salary.

The results of the model are visualized with the following figures:





The  $R^2$  squared value of the linear regression is 0.605. This reveals that the model is relatively accurate, however not extremely accurate. The confusion matrix reveals that points were indeed the best predictor of salary, followed by turnovers, assists, and blocks respectively. I was relatively surprised that turnovers were such a predictor of salary, however it could be explained that players with the highest salaries play the most and hence have more opportunities for turnovers. Age was not as surprising as it makes sense due to the fact younger players earn less on their rookie contracts. As for the least predictive statistics, they were position, team blocks, and three-point percentage, respectively. Position makes sense to me as players of all positions are in high demand in the NBA, however the team was a little surprising due to the fact that some teams often opt not to have expensive players. Three-point percentage being so low was

also very surprising considering the emphasis on three-point shooting in today's league. Overall, the hypothesis of points being the most predictive was supported and the model was relatively successful in predicting salaries.

Overall, it is important to recognize the limitations of this study. 200 players is quite a large sample size, however only one season is probably not comprehensive enough of a scope. Furthermore, using the top scores as a selection process may have been a confounding factor for points being the top indicator, however using a random selection would have been far more difficult to retrieve salaries and stats, and could have ended up with players that have no stats in numerous categories. Future studies could address these problems with access to larger amounts of data. That being said, I believe this study's results are still applicable to a certain extent. It shows that the league's emphasis on scoring is very relevant to correspondingly paying players. Also, it was interesting that some lesser emphasized stats such as turnovers could have an impact on salary. Conversely, some highly regarded stats such as three-point percentage and Blocks were very interesting to not have had a massive impact on salary. I believe this study could be useful for NBA teams when attempting to determine who to give the most money to, and what kind of players to sign. In terms of successes of this project, I am happy that the model was relatively accurate and that the hypothesized points emphasis was supported. In terms of challenges, I would say that the creation of the data set as well as cleaning held the most issues with incomplete data values as well as missing ones. Overall, this study was very interesting, and I believe it lays the groundwork for future studies that teams may use to value players accurately.

## Citations

2023-24 NBA player stats: Per game. Basketball. (n.d.). [https://www.basketball-reference.com/leagues/NBA\\_2024\\_per\\_game.html](https://www.basketball-reference.com/leagues/NBA_2024_per_game.html)

HoopsHype. (n.d.). These are the salaries of all NBA players. HoopsHype.  
<https://hoopshype.com/salaries/players/>

Microsoft Corporation. (2024). *Excel (Microsoft 365 Subscription)*. [Computer Software].  
Microsoft Corporation. <https://www.microsoft.com/en-us/microsoft-365/excel#:~:text=Microsoft%20Excel%20with%20a%20Microsoft,Excel%202007%2C%20and%20Excel%202003>.