Thomas Preston
DSC680 Project Two Milestone 2

NBA Player Salary and Performance Analysis: 2025 Season

Introduction:

NBA players are making more money than ever, and it's natural to wonder if those big contracts match how well players actually perform on the court. For this project, I looked at data from the 2025 season to see which players are giving teams their money's worth, and which ones might not be. I used player stats, some simple machine learning, and a bunch of visuals to find out. The goal was to break down salaries, performance, and try to spot trends, like whether veterans are getting overpaid or if young stars are worth the hype.

Data and Methods:

The data included stats and salaries for NBA players from 2010 through 2025, but I kept it simple and just focused on the 2025 season. That way, everything was up-to-date and easier to work with. I picked some of the most common stats people care about: points, assists, rebounds, shooting percentages (FG%, 3P%, FT%), age, and of course, salary.
I cleaned the data to remove any players missing important info. Then I grouped players into four different categories using K-Means clustering. That's just a way to let the computer sort players into groups based on how similar their stats are. The four groups were:

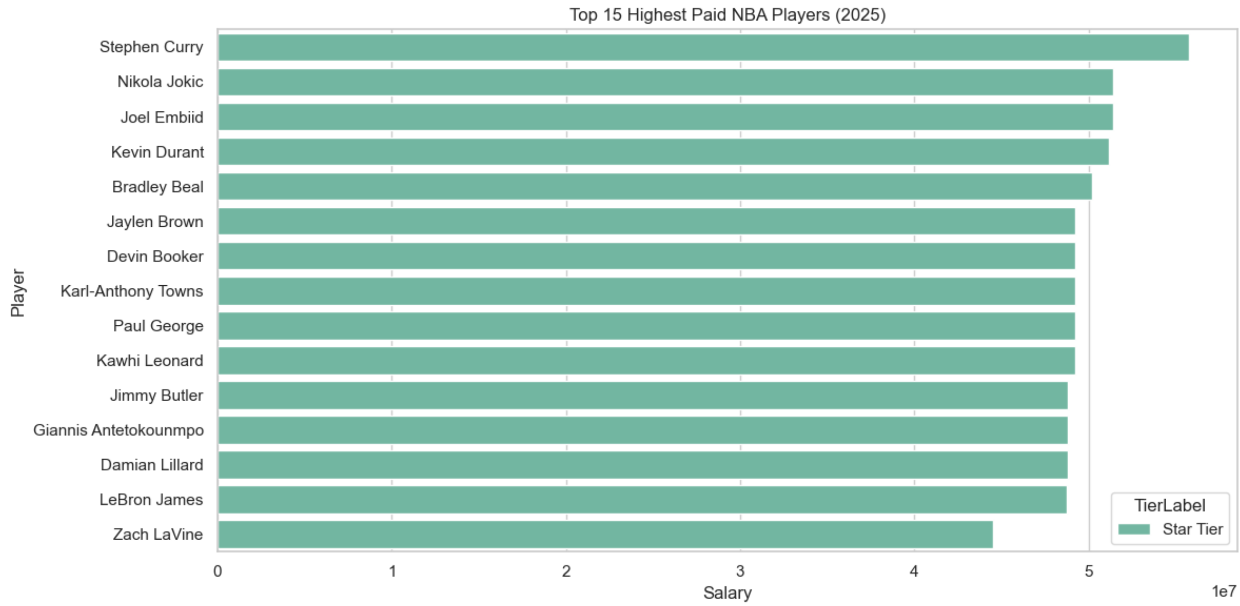
Star Tier: top performers across the board
Efficient Veterans: experienced players with strong but maybe less flashy stats
Solid Contributors: players who contribute consistently but aren't stars
Role Players: guys who fill specific needs on a team

I also tried predicting salaries using two models: Random Forest and Linear Regression. Both models used the same player stats, and I compared how well they predicted salary by checking how far off they were from actual numbers.

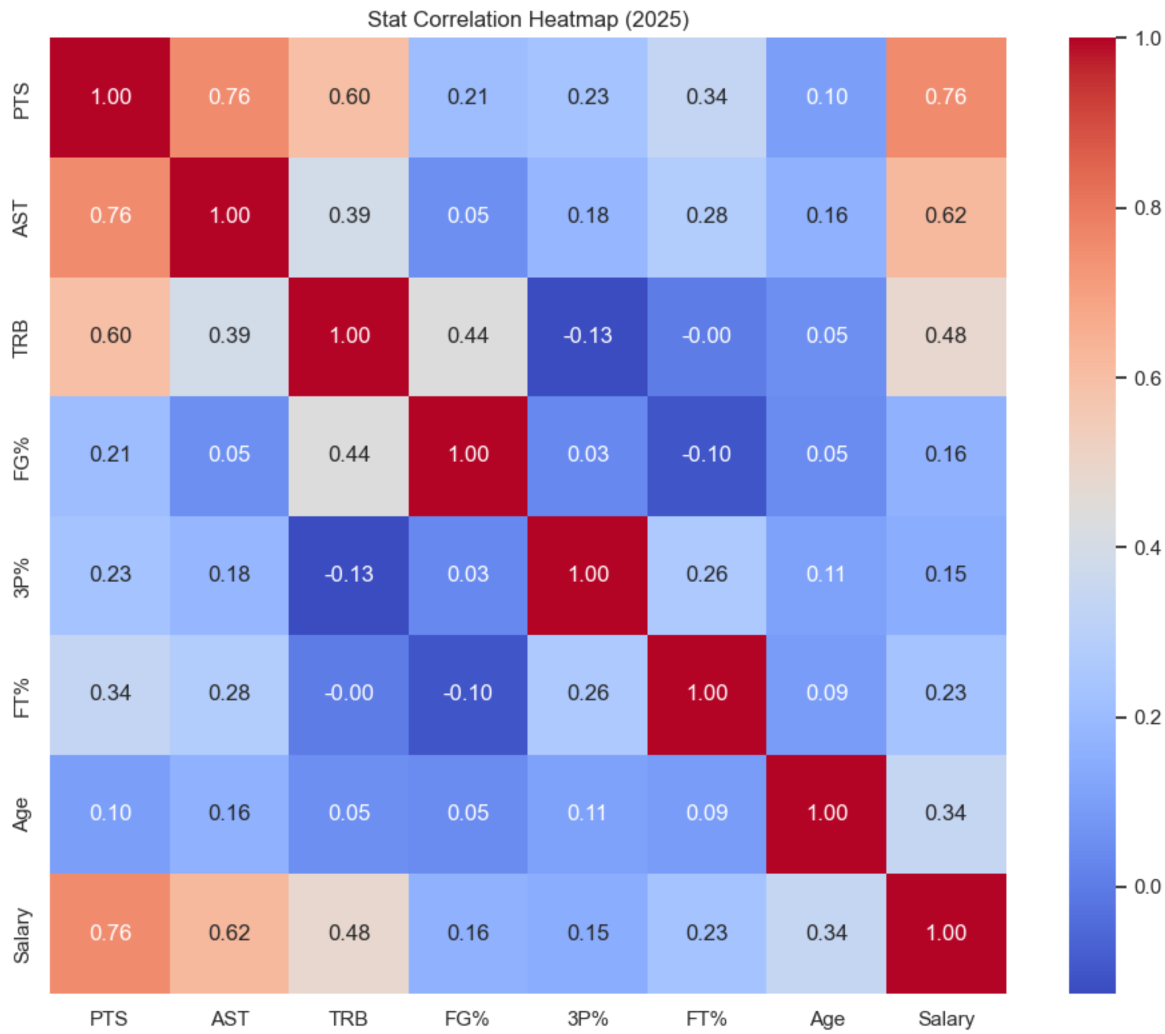
Results/Discussion:

To start, I looked at a bar chart of the top 15 highest-paid players in the league:

Thomas Preston
DSC680 Project Two Milestone 2

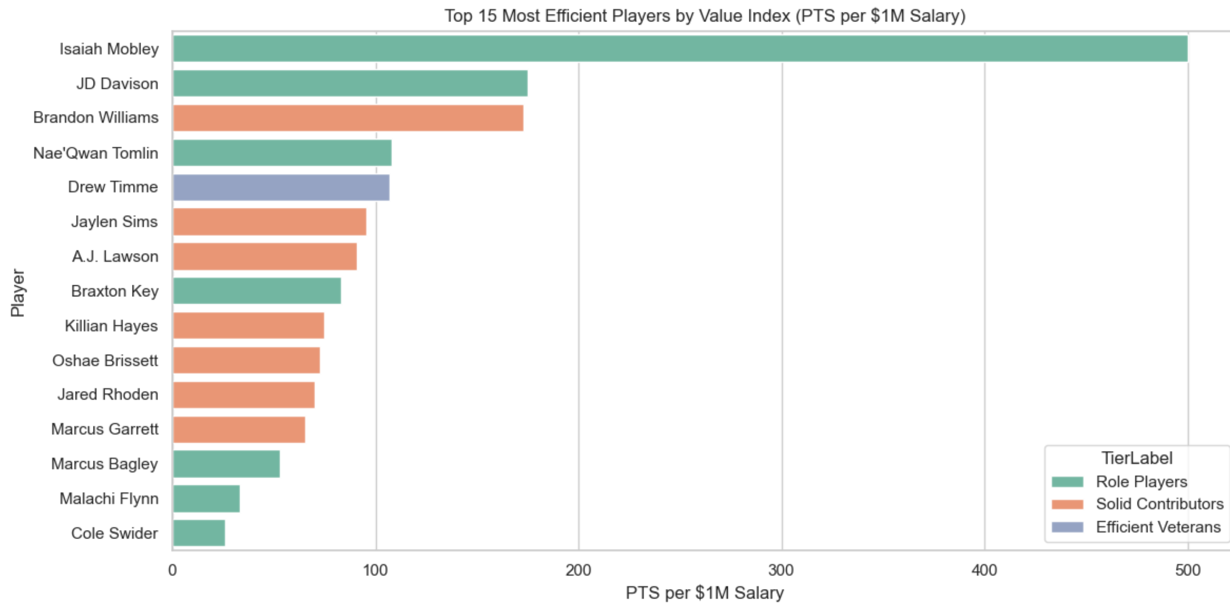


What stood out was that a few of these guys weren't even in the "Star Tier." That means some players might be getting paid based more on past success or name recognition than how well they're actually playing right now. Next, I made a heatmap to look at how different stats relate to salary:



Not surprisingly, the strongest links were between salary and stats like points, assists, and rebounds. Players who score more or help teammates score tend to get paid more. Shooting percentages like FG% and 3P% weren't as closely tied to salary. Age also didn't seem to matter much, which was interesting. Then, I came up with a simple way to measure value: Points per million dollars of salary. That gave me a "Value Index" and I used it to make another chart:

Thomas Preston
DSC680 Project Two Milestone 2



This showed the top 15 players who give teams the most scoring for their salary. While you might expect these to all be Efficient Veterans or Stars, they actually came from across several of the tiers. In fact, many were younger players on smaller contracts who were putting up big numbers. This reminded me that value isn't just about efficiency, it's also about how much teams are paying for those stats.

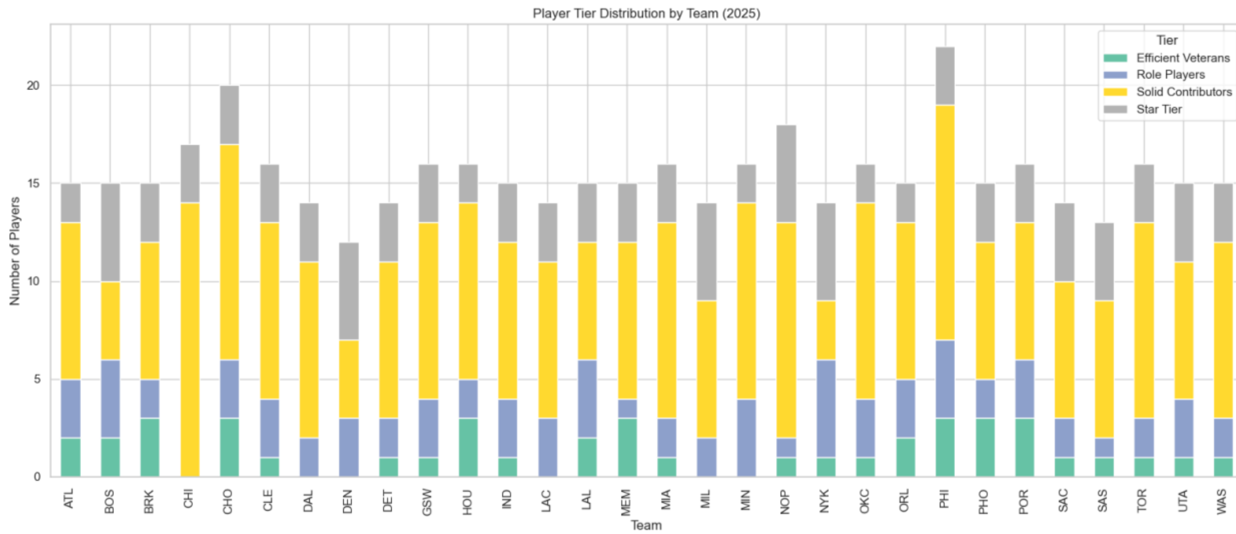
Here's how the two prediction models stacked up:

Random Forest: R^2 = 0.72, MAE ≈ \$2.1M
Linear Regression: R^2 = 0.48, MAE ≈ \$3.5M

The Random Forest model did a better job overall. It handled the ups and downs in the data better and predicted closer to the actual salaries. Linear Regression was easier to understand, but it didn't do as well at capturing the complexity of NBA contracts.

I also broke things down by team to see how different franchises are built. Some teams have lots of Role Players, while others lean heavily on a few stars. This was useful to see which teams might be getting the most value, and which ones might be top-heavy with big contracts that aren't paying off.

Thomas Preston
DSC680 Project Two Milestone 2



Conclusions:

This project showed that NBA salaries don't always line up with what players are doing on the court. Some big names are still getting paid even if their numbers have dropped, while some quiet veterans are providing solid value at a lower cost. The Random Forest model was the best at predicting salaries, and the Value Index helped highlight which players are real bargains.

If I were to keep going, I'd want to look at multiple years of data to see how performance and pay change over time. I'd also include more advanced stats like PER (Player Efficiency Rating) or Win Shares to get a deeper picture. But even just using basic stats and a few models, this project showed a lot about how NBA teams are spending their money and where they might want to think twice.

Ethical Considerations:

When looking at salaries and performance like this, it's important to remember that NBA players are people, not just numbers on a spreadsheet. A lot of what goes into a player's salary isn't just about stats, it can include leadership, locker room presence, fan appeal, or even just loyalty to a team. These are things that models can't always measure.
Also, some players might be recovering from injuries or adjusting to new roles, and those situations don't always show up in the data. So even if a model says someone is "overpaid," there could be good reasons behind it that go beyond performance stats. It's important not to let data analysis turn into unfair judgment, especially when real careers and reputations are involved.

Thomas Preston
DSC680 Project Two Milestone 2

Audience Questions:

1. How did you decide which stats to include in the model?

   I used common stats like points, assists, and rebounds because they're widely used and easy to understand.

2. Why did you choose K-Means clustering over other grouping methods?

   It's simple, fast, and good for grouping players based on numbers without needing labels.

3. Did you look at how injuries might have affected a player's stats or salary?

   Not directly, but I know injuries can impact totals. It's something I'd add in a future version.

4. How did you handle outliers like rookies on cheap contracts or supermax deals?

   I kept them in, they helped show who's really over- or under-valued.

5. Were there any surprises in the Star Tier or Efficient Veterans group?

   Yes, some stars weren't good value, and some vets were performing better than expected.

6. How accurate do you think your salary predictions would be for future seasons?

   Fairly accurate, but not perfect. Contracts and outside factors still matter a lot.

7. Did you consider team performance or wins in your analysis?

   No, I focused on individual performance only.

8. Why do you think shooting percentages had a weak correlation with salary?

   Because teams often pay more for volume scorers than for efficient ones.

9. Were there any players who ranked high in value but were almost unknown?

   Yes—some younger or lesser-known players gave great value for their salary.

Thomas Preston
DSC680 Project Two Milestone 2

10. If teams actually used your Value Index, how would it change how they build rosters?

    They might spend smarter, investing more in undervalued players and less in big names.

Kaggle. (n.d.). *NBA Player Stats and Salaries: 2010–2025* [Data set]. https://www.kaggle.com/