Davide Momi

PICK ANY PLAYER AND WRITE AN ANALYSIS OF HIS STRENGTHS, WEAKNESSES, AND WHAT YOU FEEL HIS CONTRACT VALUE & # OF YEARS SHOULD BE. ALSO IF POSSIBLE FORECAST HIS PERFORMANCE OVER THE NEXT SEVERAL YEARS

Before delving into player analysis, it's crucial to comprehend the intricate salary system of the NBA. Free agents (FAs) are players available for contracts, and the league employs mechanisms like the salary cap and luxury tax to maintain competitive balance.

The salary cap, updated yearly, restricts team spending on player salaries, ensuring financial fairness and preventing major-market teams from overpowering smaller counterparts. Teams exceeding this cap enter the luxury tax realm, which penalizes excessive spending.

While various exceptions exist, our focus centers on the salary cap and luxury tax to predict player salaries effectively

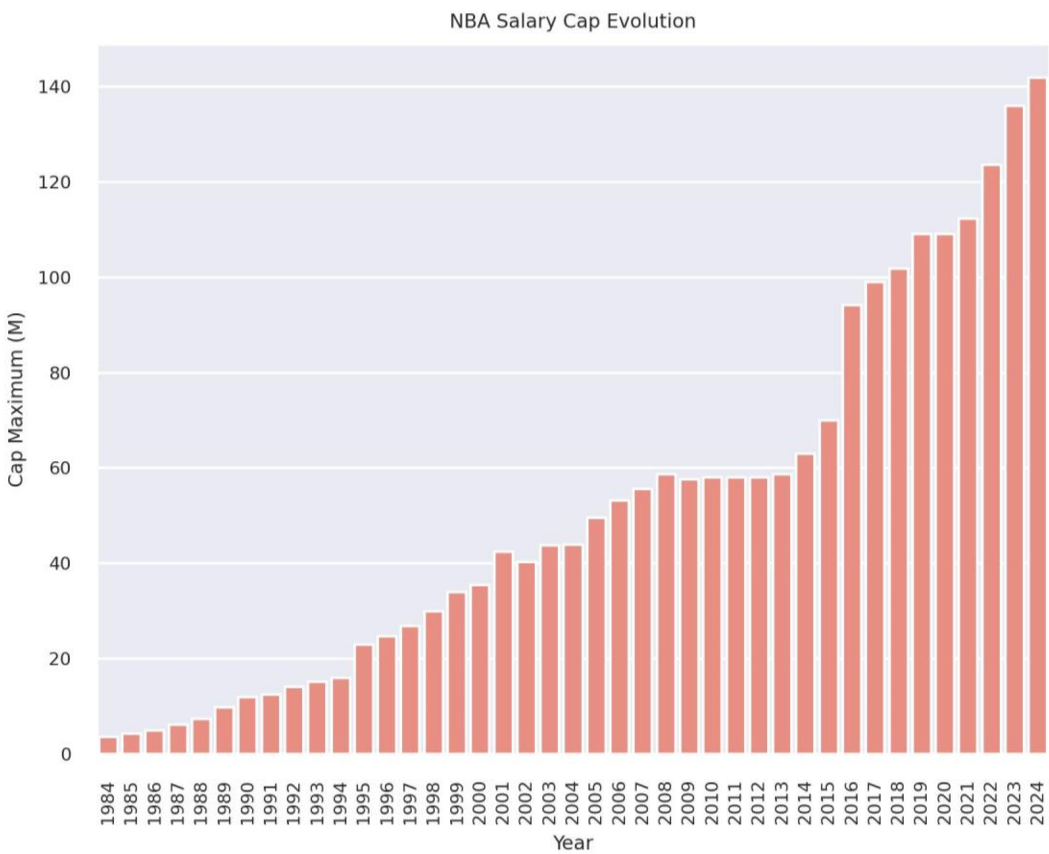

Figure 1. NBA Salary Cap Evolution from 1984 to 2023

This project utilizes individual player stats, including average and total game statistics, advanced metrics, and variables like age and position. Additionally, salary-related data from the previous season, along with the maximum cap for the current season, is used. The objective is to predict salaries based on the previous season's data. To maintain accuracy despite evolving salary caps, the analysis employs percentages of the cap as targets, ensuring applicability across seasons.

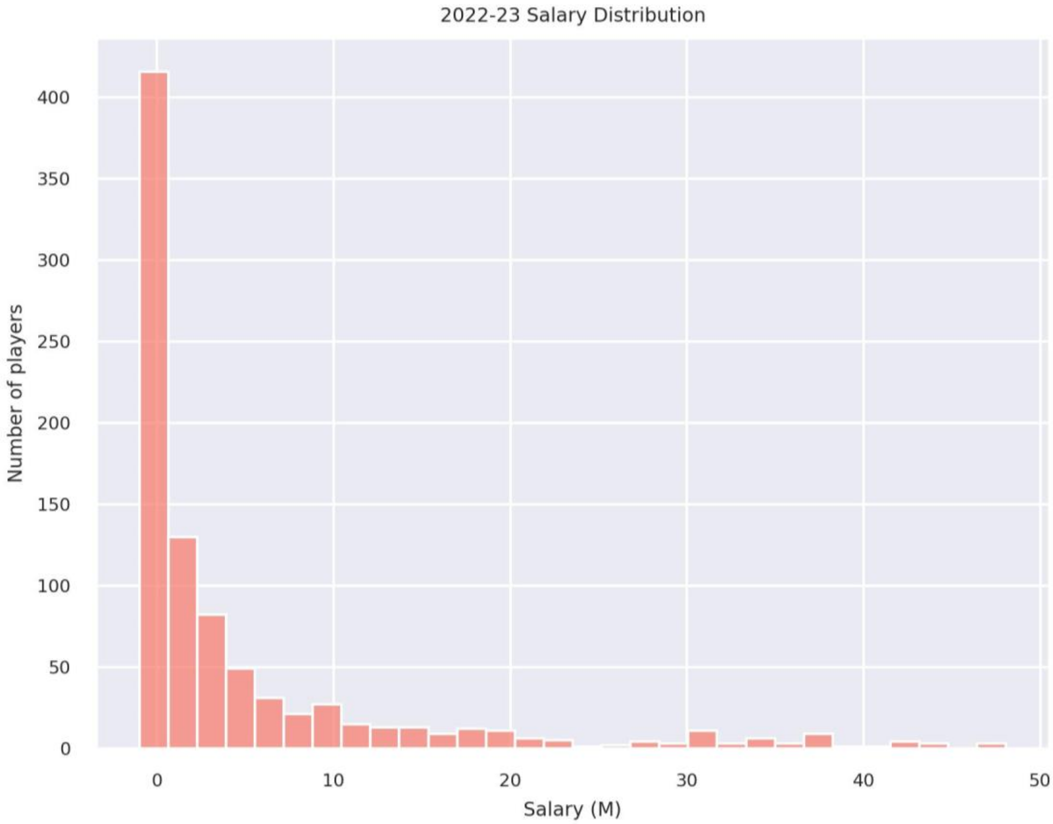

Figure 2. Salary distribution for the 2022–23 season.

To optimize model training, only free agents' data was considered. Unlike players under contract, free agents' salaries experience significant fluctuations, making them ideal for predictive modeling. Data from 2020–21 onwards was chosen, providing a substantial sample size of 426 players, including 84 free agents from 2023–24.

Several regression models were employed to predict salaries:

- **Support Vector Machines (SVM):** SVMs excel in handling complex relationships within data, making them valuable for predicting intricate salary patterns in basketball contracts.
- **Elastic Net:** Elastic Net combines the strengths of Lasso and Ridge regression, effectively handling multicollinearity and selecting essential features for accurate predictions.

- **Random Forest:** Random Forest algorithms harness the power of multiple decision trees, providing robust predictions by aggregating outputs from various trees.
- **AdaBoost:** AdaBoost focuses on weak learners, combining their outputs to form a strong model, ideal for capturing subtle salary nuances in diverse player datasets.
- **Gradient Boosting:** Gradient Boosting builds predictors sequentially, correcting errors of previous models, making it adept at capturing intricate salary structures and patterns.
- **Light Gradient Boosting Machine (LGBM):** LGBM, a gradient boosting framework, is renowned for its speed and efficiency, ensuring rapid processing of vast datasets and accurate predictions.

Each model's performance was evaluated using the Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) metrics. By employing these diverse regression models, this analysis aims to precisely predict the salaries of upcoming free agents, considering the ever-changing landscape of NBA contracts. The next steps involve delving into the results to provide comprehensive insights into player contracts, strengths, weaknesses, and future performance forecasts. Looking at the whole dataset with all seasons, the following results were obtained:

Model	RMSE	R^2
SVM	3.81	0.628
Random Forest	3.65	0.658
Elastic Net	3.82	0.627
AdaBoost	4.23	0.541
Gradient Boosting	3.54	0.679
LGBM	3.73	0.644

Table 1. RMSE and R^2 values obtained among the different models.

The models had an overall good performance, with Random Forest and Gradient Boosting obtaining the lowest RMSE and highest R^2, while AdaBoost had the worst metrics among the models used. A valuable technique for understanding the significant variables influencing the model's predictions is utilizing SHAP Values, which offer insights into how each feature impacts the model's outcomes. In a nutshell, SHAP Values provide insights into the impact of individual features on the model's predictions.

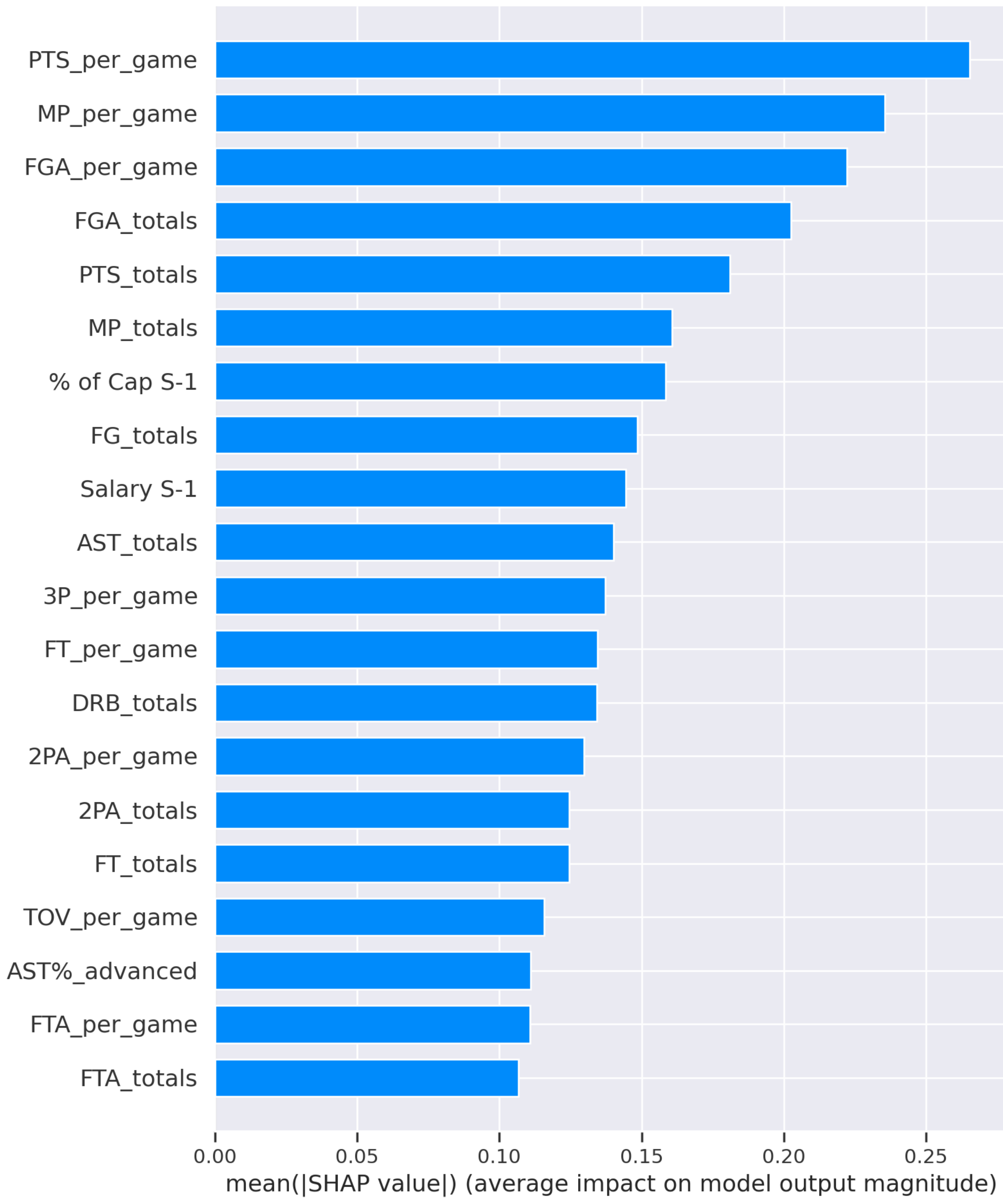

Figure 3. SHAP chart related to the Random Forest model

From the SHAP chart, several noteworthy conclusions can be drawn: Minutes per game (MP) and points per game (PTS), both total and per game, are the top three most influential features. Salary from the previous season (Salary S-1) and the percentage of the cap represented by that salary (% Cap S-1) also hold substantial impact, ranking 4th and 5th, respectively. Advanced statistics, despite their complexity, do not dominate the list of crucial features. Only two advanced metrics, Win Share (WS) and Value Over Replacement Player (VORP), make it to the list.

This outcome is surprising, because conversely to the expectation, **advanced statistics do not play a significant role in shaping SHAP's final results**. In the context of player salaries, it appears that more conventional statistics like minutes played, points scored, and games started carry considerably more weight. This revelation is unexpected because advanced statistics were specifically designed to offer a nuanced evaluation of a player's performance. **The absence of Player Efficiency Rating (PER) among the top 20 features (it ranks 43rd) is particularly noteworthy.**

This finding raises the intriguing possibility that during salary negotiations, team managers might be employing a relatively straightforward approach. They could be focusing on basic metrics, potentially overlooking the broader array of performance evaluation metrics. In essence, the issue might not be as intricate as initially thought. **Simplifying it, it appears that players who spend more time on the court and score more points tend to command higher salaries, suggesting a direct correlation between playing time, scoring, and earnings.**

An interesting result is also the prediction of the salaries divided by position.

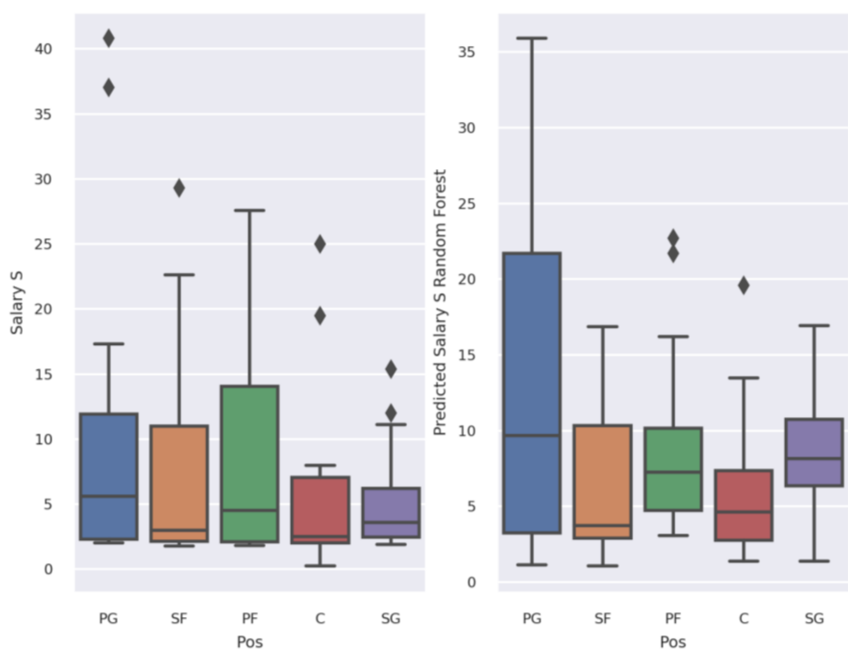

Figure 4. Predicted Salaries with the Random Forest model divided by position

Remarkably, the Point Guards seem to call higher salaries compared to the other positions.

Fred VanVleet

After providing the necessary overview as per the assignment requirements, I selected Fred VanVleet, who secured a three-year, $130 million deal with the Rockets in the recent Free Agency. According to the model's evaluations, VanVleet falls into the category of being significantly overvalued, meaning he is receiving a higher salary than what the model suggests he should be earning. It's crucial to emphasize that these assessments are exclusively derived from the model's predictions.

For the upcoming season, VanVleet is set to earn $40.01 million, which is nearly $13 million more than the model's estimated value of $28.38 million. It's noteworthy that in the NBA title-winning season of 2018/2019, VanVleet received salaries of $8.65 million and $9.34 million. During that period, his performance was stellar, leading to the new contract that raised his earnings to $21.25 million over three years.

Regrettably, his performance didn't follow a similar trajectory, remaining relatively consistent. His Player Efficiency Rating (PER) stood at 16.3 before the maximum contract, increasing marginally to 17. However, his Defensive Rating worsened, rising from 105.8 before the max contract to 114.2. It's essential to consider that the Raptors, his former team, also faced challenges and weren't as competitive compared to previous years, impacting individual player performance evaluations.