# Stat184_FP_Report

Rymer Popovich, Ben Tartaglia

## NBA Dataset Report

The NBA season is a thrilling showcase of basketball excellence, where players compete for recognition through their individual and team performances. From March 1st, 2024, onward, numerous athletes have demonstrated remarkable skills and impact on the court. This report evaluates players' performances and predicts the recipients of three prestigious awards: Most Valuable Player (MVP), Six Man of the Year, and Defensive Player of the Year. Through statistical exploration, visualization, and regression analysis, we provide insights backed by a handful of evidence to demonstrate standout performers across these categories.

### Research Questions

This analysis focuses on answering three key questions. First, who is the Most Valuable Player, considering the highest overall performance in points, assists, and rebounds? Second, which first-year player stands out as the Six Man of the Year based on significant contributions to their team's success? Finally, who excels as the Defensive Player of the Year by demonstrating superior performance in defensive metrics such as blocks, steals, and defensive win shares?

### Data Provenance

The dataset used in this report comprises NBA player statistics for games played after March 1, 2024. It was sourced from Kaggle (multiple NBA datasets were available, we decided to pick the most broad dataset in terms of stat-types, from March 1st onward) and includes attributes such as points per game (pts), assists (ast), rebounds (trb), blocks (blk), steals (stl), and defensive win shares (dws). Each row represents a player's game-by-game averages, providing a powerful foundation for analyzing individual performances. The data was collected with the purpose of evaluating key performance metrics to highlight exceptional players.

### FAIR Principles

This dataset follows the FAIR principles, making it accessible and easy to use for analysis. First, the data is findable because it's well-organized and labeled, so it's easy to locate what we need. It's also accessible since it's stored in a common format like CSV, which most tools can read without any special software. The dataset is interoperable because it uses a standard structure that works with R and other programming languages, allowing seamless analysis. Finally, the data is reusable because it's clearly documented with explanations of each column, making it simple for anyone to understand and use for future projects.

### Methodology

The analysis began with cleaning the data to remove missing values and calculating additional metrics essential for award predictions. Specifically, Total Contributions was derived as the sum of points, assists, and rebounds, while Defensive Impact was calculated as the sum of blocks and steals. Visualizations and statistical summaries were then used to identify trends and rank players for each award category. Finally, regression analysis was employed to understand the key factors influencing player contributions.

### Results

The Most Valuable Player (MVP) award prioritizes overall performance in scoring, playmaking, and rebounding. Total Contributions (pts + ast + trb) was the primary metric used to evaluate candidates. Visualization of the top 10 MVP contenders highlights Joel Embiid as the leading candidate, showcasing their balanced contributions across all key areas. Their performance demonstrates a notable ability to influence the game offensively and defensively.

For Six Man of the Year, analysis focused on non-starting players and their contributions to team success. A filtered dataset revealed that Malik Monk leads among non-starters with exceptional scoring and playmaking metrics. Their consistency and ability to adapt quickly to the professional level distinguish them from their peers.

The Defensive Player of the Year award evaluates players excelling in defensive metrics. By analyzing Defensive Impact (blocks and steals) and Defensive Win Shares, Victor Wembanyama emerged as the standout candidate. They demonstrated an unparalleled ability to disrupt opponents' plays and contribute to their team's defensive success. A scatterplot of the top 10 defensive players further illustrates their dominance in key defensive metrics.

A regression analysis was conducted to identify significant predictors of Total Contributions. Points per game (pts) emerged as the most impactful variable, followed by Player Efficiency Rating (per). These results confirm that scoring ability and overall efficiency are critical determinants of a player's contribution to their team's success.

## Conclusion

This analysis provides data-driven predictions for the 2024 NBA season awards. Joel Embiid is projected as the MVP due to their well-rounded performance in scoring, assisting, and rebounding. Six Man of the Year is likely to go to Malik Monk, who has shown remarkable productivity in their first season. Lastly, Defensive Player of the Year is predicted to be awarded to Victor Wembanyama for their superior defensive metrics and impact. While the findings are definitely strong, the analysis is limited by the exclusion of qualitative factors such as leadership and game-winning moments, which are often considered in award decisions yet very difficult to track in the form of a player stat - they would be considered extra information for this dataset, and are relatively difficult to record.

## References


- Dataset: https://www.kaggle.com/datasets/bryanchungweather/nba-player-stats-dataset-for-the-2023-2024
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2023). Introduction to Econometrics with R. Springer.