Basketball Simulation and Insights

Members

Nandini Bhattad 220693

Ankita Goyal 231080016

Gunavant Thakare 231080040

Naman Manchanda 231080061

Dr. Dootika Vats

Data Science Lab 1

Acknowledgments

I would like to express my sincere gratitude to Dr. Dootika Vats for her invaluable guidance and support throughout the Data Science Lab 1 project. Her expertise and mentorship have played a pivotal role in shaping the direction of this project.

Dr. Vats' insightful feedback, encouragement, and dedication to fostering a collaborative learning environment have been instrumental in enhancing my understanding of data science concepts and methodologies.

I am truly thankful for the opportunity to work under Dr. Vats' supervision, and I appreciate her commitment to excellence in teaching and research. This project has been a rewarding experience, largely due to her mentorship.

Table of Contents

Introduction

The Basketball Simulation and Insights project delve into the dynamic world of basketball analytics, aiming to leverage data science methodologies to gain valuable insights into player performance, team strategies, and game outcomes. Basketball, as a highly dynamic and fast-paced sport, generates a wealth of data during each match, from player statistics to in-game events, providing a rich landscape for analysis.

The primary objective of this project is to develop a comprehensive basketball simulation model that can replicate real-game scenarios, allowing us to study various factors influencing team and player performance. By harnessing the power of data science and statistical modeling, we seek to unravel patterns, trends, and key indicators that contribute to success on the basketball court.

In addition to the simulation aspect, this project aims to provide actionable insights for coaches, analysts, and enthusiasts. Through exploratory data analysis and machine learning techniques, we aspire to identify critical performance metrics, strategic plays, and potential areas for improvement. The insights derived from this project have the potential to inform coaching decisions, optimize player training regimens, and contribute to a deeper understanding of the sport.

The project will encompass a diverse range of data sources, including player statistics, game play-by-plays, and team dynamics. Our analytical approach will involve data cleaning, feature engineering, and the implementation of machine learning algorithms to model and simulate basketball scenarios.

As we embark on this basketball analytics journey, we anticipate uncovering new perspectives on the sport, fostering innovation in coaching strategies, and contributing to the growing field of sports analytics. The combination of simulation and insights aims to not only enhance our understanding of basketball dynamics but also provide practical applications for teams and enthusiasts passionate about the game.

Methodology

Data Collection

The foundation of this project involves the extraction of comprehensive player statistics from the NBA 2022-23 season. We employed web scraping techniques to gather data from the Basketball Reference website (https://www.basketball-reference.com). The website provides detailed and up-to-date information on NBA player statistics, team performance, and game results.

The primary data source for this project was the individual player statistics pages on Basketball Reference. We focused on collecting information for all 679 players who participated in the NBA during the 2022-23 season.

Web Scraping

We utilized the `rvest` package in R for web scraping, allowing us to programmatically extract player statistics. The scraping process involved navigating through player pages, retrieving summary statistics, and storing the data for further analysis.

Data

The following variables were used, Rk (Rank): The player's rank or position in the specified context, often indicating their rank among all players in a certain category or statistical measure.

Player: The name of the basketball player.

1. **Age:** The age of the player during the specified season.

2. **Tm (Team):** The abbreviation or code representing the team the player is associated with during the specified season.

3. **G (Games Played):** The number of games the player participated in during the season.

4. **GS (Games Started):** The number of games in which the player was in the starting lineup.

5. **MP (Minutes Per Game):** The average number of minutes the player spent on the court per game.

6. **FG (Field Goals Made):** The total number of successful field goals made by the player.

7. **FGA (Field Goals Attempted):** The total number of field goal attempts by the player.

8. **FG% (Field Goal Percentage):** The shooting accuracy of the player, calculated as (FG / FGA) * 100.

9. **3P (Three-Pointers Made):** The total number of successful three-point shots made by the player.

10. **3PA (Three-Pointers Attempted):** The total number of three-point shot attempts by the player.

11. **3P% (Three-Point Percentage):** The three-point shooting accuracy of the player, calculated as (3P / 3PA) * 100.

12. **2P (Two-Pointers Made):** The total number of successful two-point field goals made by the player.

13. **2PA (Two-Pointers Attempted):** The total number of two-point field goal attempts by the player.

14. **2P% (Two-Point Percentage):** The two-point shooting accuracy of the player, calculated as (2P / 2PA) * 100.

15. **eFG% (Effective Field Goal Percentage):** A modified field goal percentage that accounts for the added value of three-pointers, calculated as (FG + 0.5 * 3P) / FGA * 100.

16. **FT (Free Throws Made):** The total number of successful free throws made by the player.

17. **FTA (Free Throws Attempted):** The total number of free throw attempts by the player.

18. **FT% (Free Throw Percentage):** The free throw shooting accuracy of the player, calculated as (FT / FTA) * 100.

19. **ORB (Offensive Rebounds):** The total number of offensive rebounds grabbed by the player.

20. **DRB (Defensive Rebounds):** The total number of defensive rebounds grabbed by the player.

21. **TRB (Total Rebounds):** The total number of rebounds (both offensive and defensive) grabbed by the player.

22. **AST (Assists):** The total number of assists made by the player.

23. **STL (Steals):** The total number of steals made by the player.

24. **BLK (Blocks):** The total number of shots blocked by the player.

25. **TOV (Turnovers):** The total number of turnovers committed by the player.

26. **PF (Personal Fouls):** The total number of personal fouls committed by the player.

27. **PTS (Points):** The total number of points scored by the player.

28. **Pos (Position):** The player's position on the basketball court, such as guard (G), forward (F), or center (C).

Analysis

Player Analysis

Radial Plot
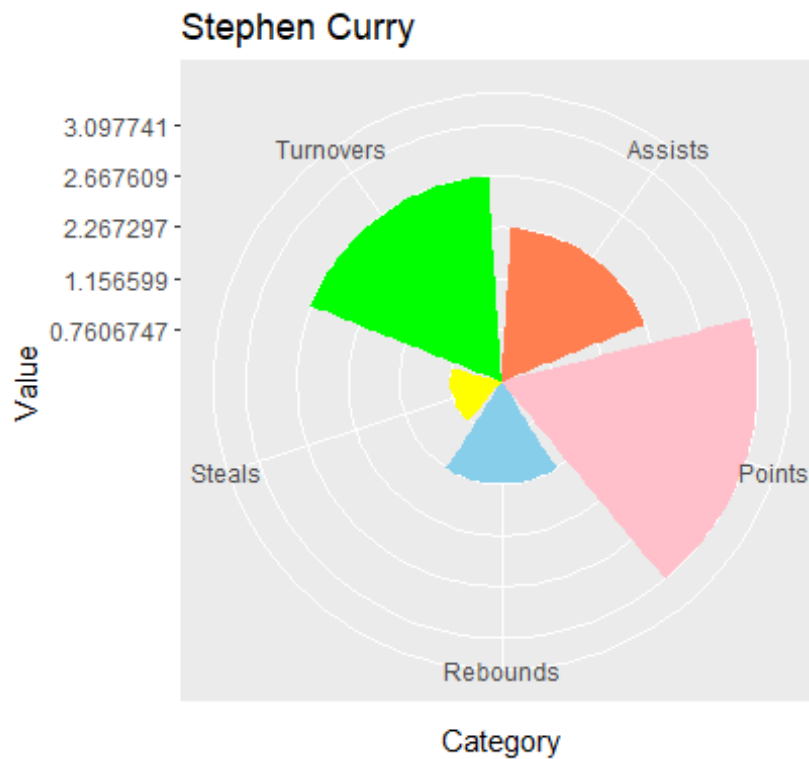
Stephen Curry

This is the Radial Plot of Stephen Curry, on the y-axis, you will find points that correspond to his normalized score in all the departments. As observed from the graph we can infer that he is above 0 in all departments, i.e. he is an above average player in every category. You can find this type of radial plot for all Players in the Shiny Web App.
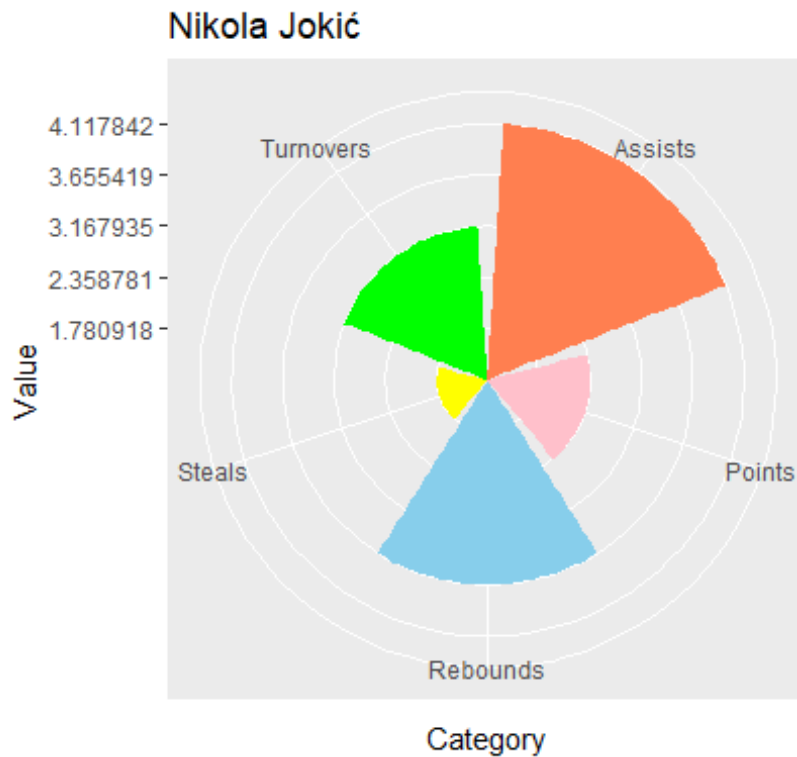
Nikola Jokić

This is the radial plot of Nikola Jokic who won the Most Valuable Player (MVP) Award in the NBA season 2022-23. As evident from the y-axis he performs above average in every category and he is at the top of the league in Assists and Rebounds and a top 20 player in Points per game.

Scatter Plot
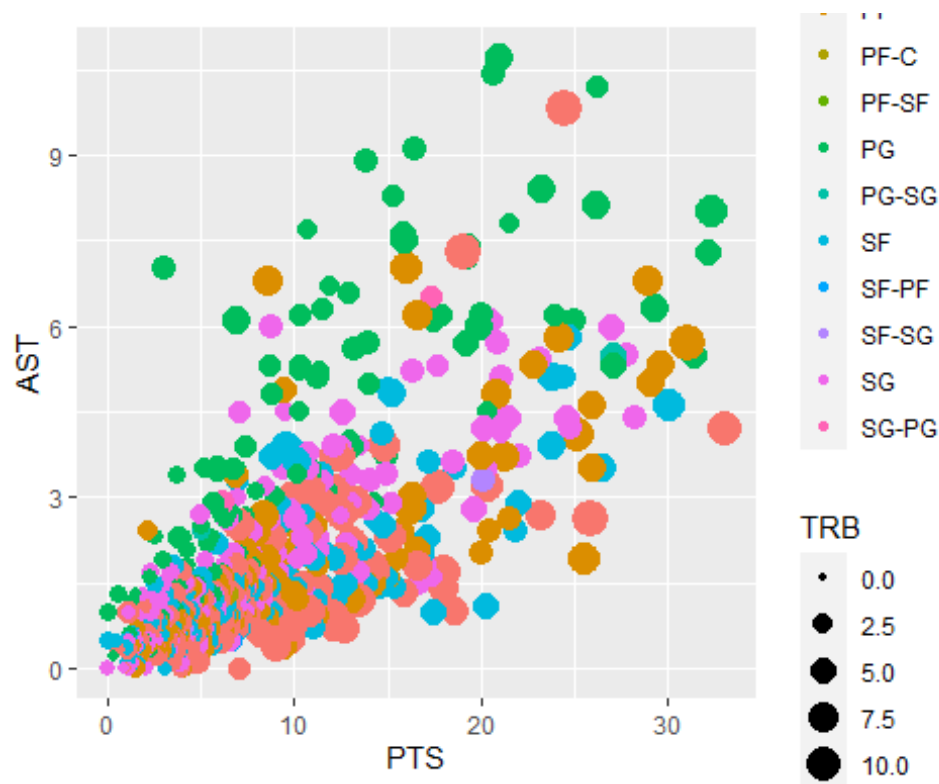
This scatter plot aims to correlate Points and Assists while grouping the data based on their Position and the number of rebounds they take,

We can observe,

- **Green:** Point guards tend to have higher points and higher assists however they lack in the department of Rebounds.

- **Pink:** Centers tend to have more rebounds and less assists , however there are some outliers like Joel Embid and Nikola Jokic who tend to perform well in every category.

- **Blue:** Shooting Guards look to score more rather than assisting other players.
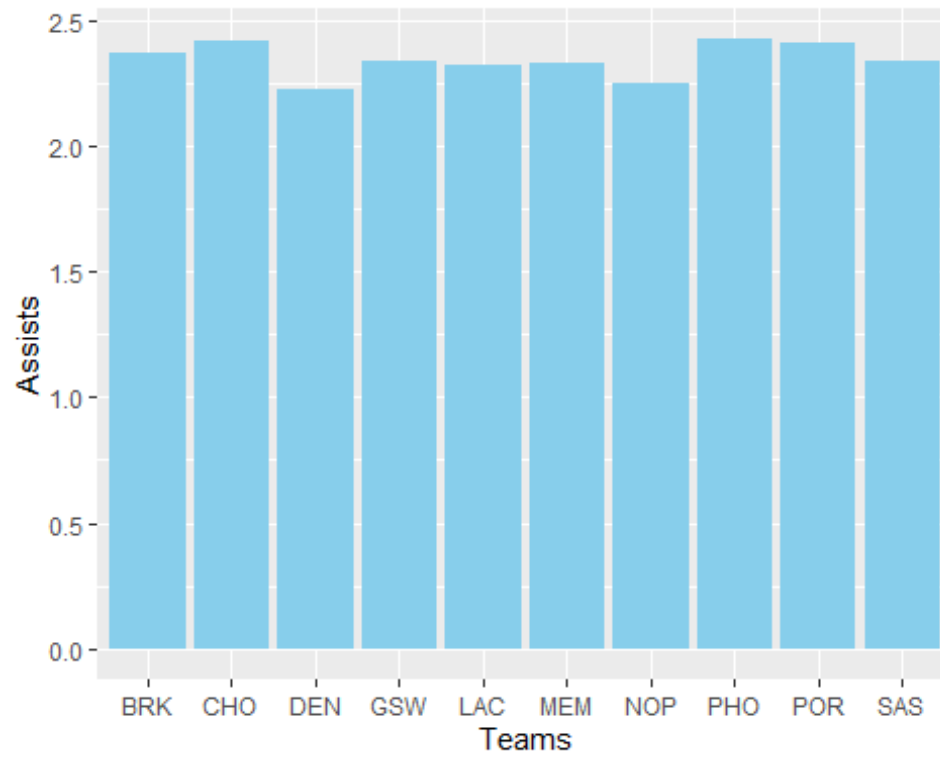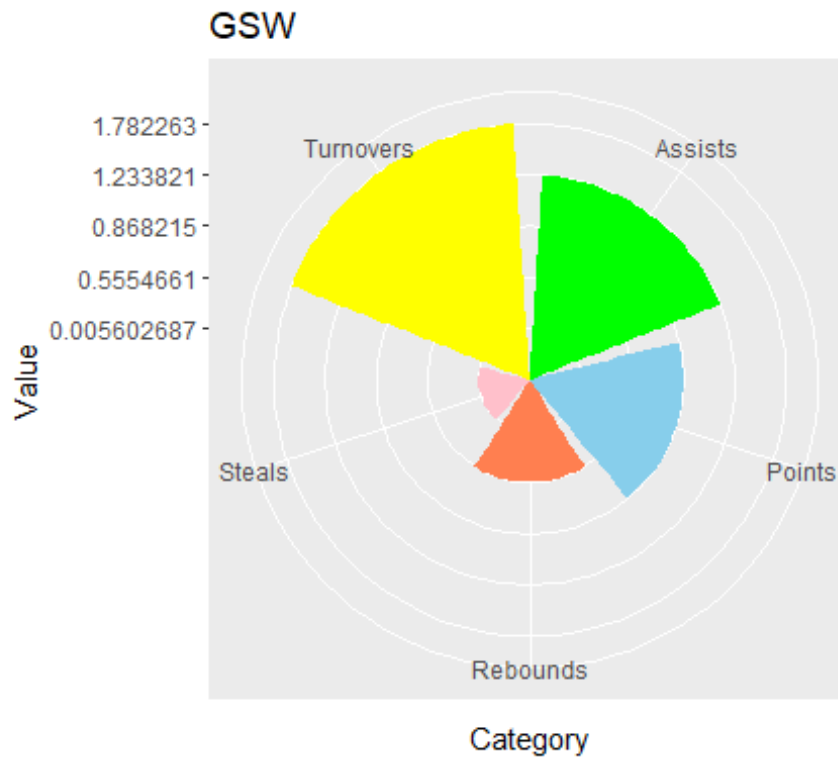
Team Analysis

Best Team in a Category
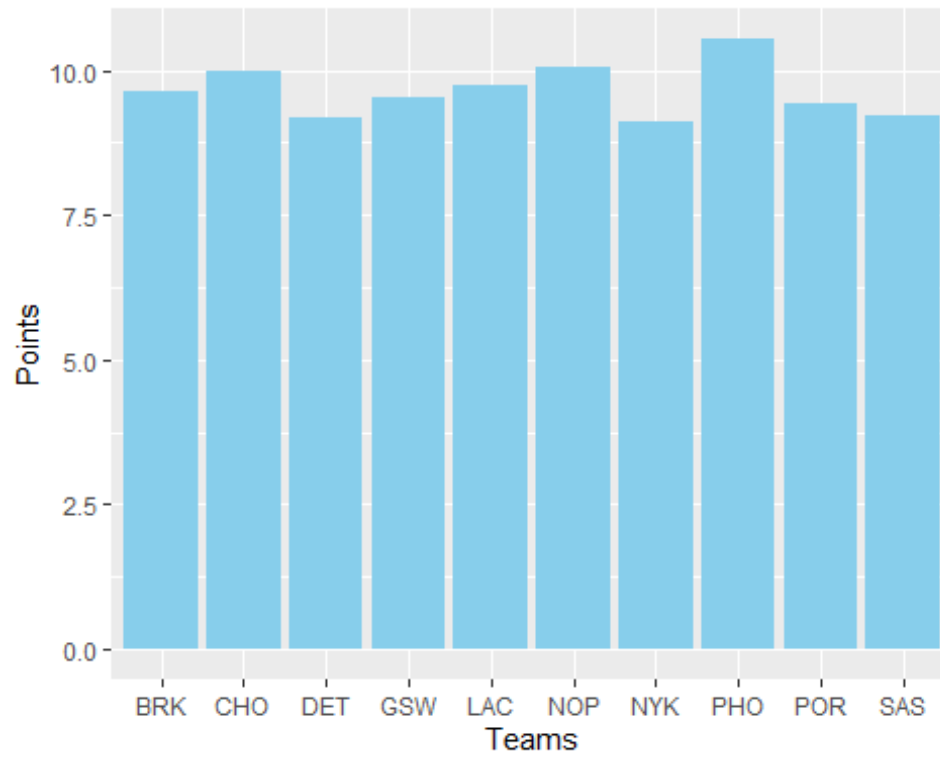
As observable from the graph , the most scoring team in the NBA season is Phoenix Suns followed up by Charlotte Hornets and Brooklyn Nets.

Here also, Phoenix Suns and Charlotte Hornets perform the best.

Radial Plot

GSW

This is the Radial Plot of Golden State Warriors , on the y-axis, you will find points that correspond to his normalized score in all the departments. As observed from the graph we can infer that this team is above 0 in all departments, i.e. GSW is an above average team in every category. You can find this type of radial plot for all teams in the Shiny Web App.

TOT
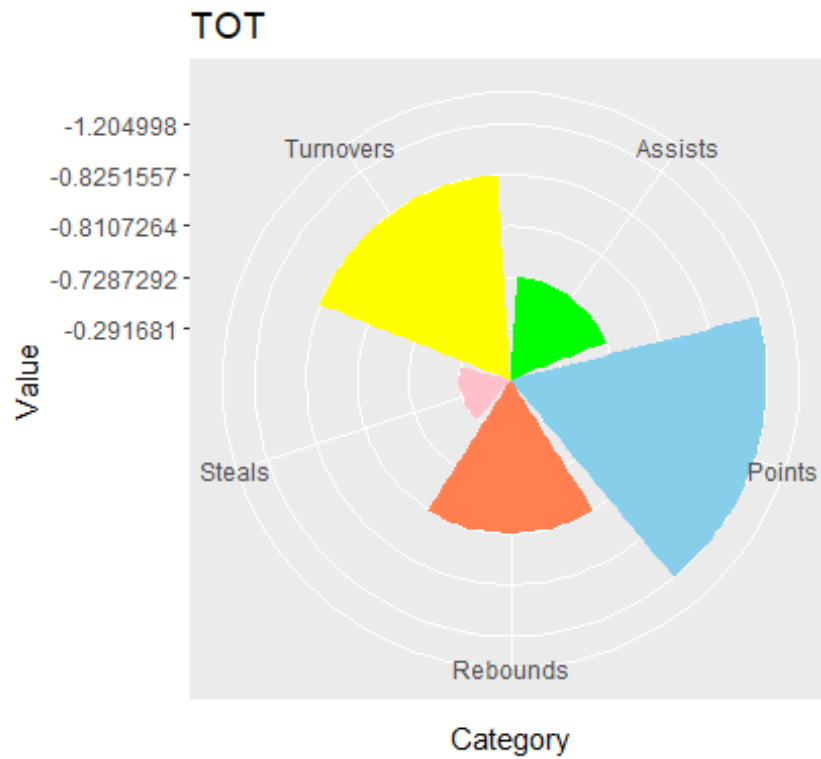
As you can see on the y-axis, Torronto Raptors perform below the league average in almost all departments, that's why they could not qualify for playoffs in the NBA season 2022-23.

Inter Team Analysis

DEN vs MIA
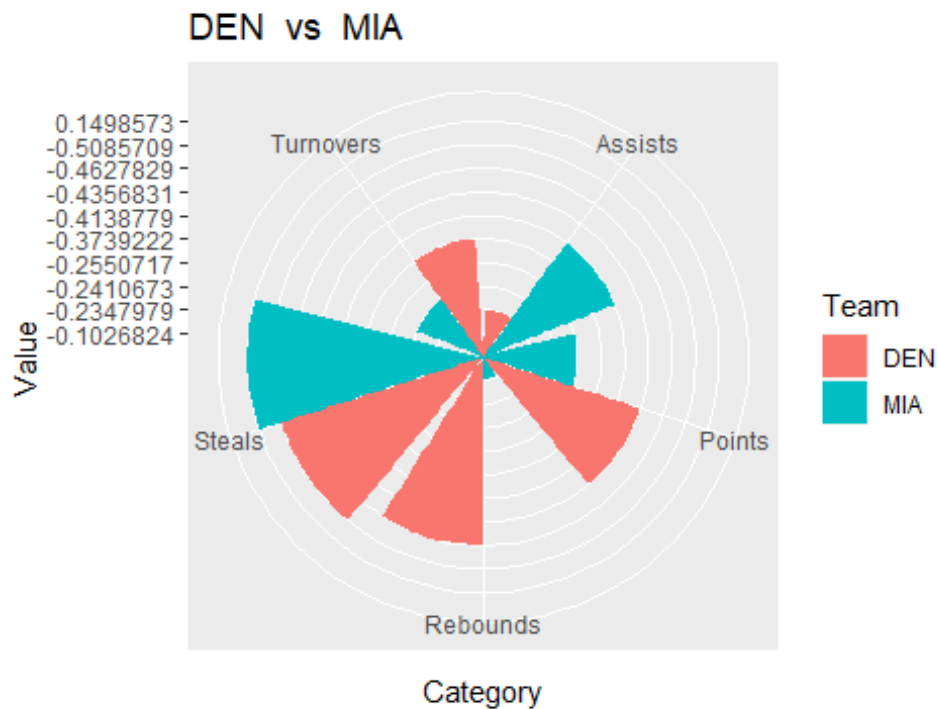
These teams were in the NBA finals last year, and as you can see Denver Nuggets outperform Miami Heat in more crucial chracterstics such as Points and Assists.

Winning Impact

We have tried to measure winning impact using the given formula

$$Winning\ Impact = \frac{Points + Assists + Rebounds + Effective\ Field\ Goal\ Percentage}{Minutes\ Played}$$

While evaluating this statistic , scaled chracterstics were used i.e. league average was subtracted and the result was divided by the league standard deviation.

The top three players as per this statistic are

1. Giannis Antetokounmpo

2. Nikola Jokić

3. Joel Embiid

And the top the players who got the most votes for MVP Award in NBA season 2022-23,

1. Joel Embiid
2. Nikola Jokić
3. Giannis Antetokounmpo

Game Stimulation

For Game Stimulation, Two teams were allowed to take 5 players each alternatively,

After that a random number was generated between -7 to 7

and this random number was added to the season averages of the first team,

This was followed up by the calculation of Game Score,

$$Game\ Score = 0.6 * Points + 0.2 * Assists + 0.2 * Rebounds + 0.1 * Steals - 0.1 * Turnovers$$

The above process was repeated to calculate the Game Score of Team B

The team whose Game Score was higher won the quater

and the team who won more quaters out of 4 won the game.

Results

- More Complete Teams tend to perform better in playoffs however some factors like Points and Assists tend to affect winning in playoffs more than other factors.

- Winning Impact can be calculated with various ways however the method highlights the value of that player to that team.

- Players who excel in there role tend to perform better however there are some outliers who perform well in all categories like Nikola Jokic and Stephen Curry.

References

1. *Basketball Reference.* (n.d.). Retrieved from https://www.basketball-reference.com

2. *NBA Official Website.* (n.d.). Retrieved from https://www.nba.com