

# Understanding NBA Player Success: The Role of Physique, College, and Draft Background

Aditya Kushwaha

Kartik Gupta

Micaela DeLong

5/1/25

## Introduction

The NBA represents the best basketball players in the world and their stats, like many athletes are well documented providing an in depth description of physicality and their background. NBA team scouts and GM's are on a constant watch for different characteristics of players that could provide context to their potential in the league. We wanted to dive into these varying attributes both physical and developmental (i.e. school and background) to get a sense of what makes great performing NBA players. We extracted Data from NBA.com as well as Kaggle to create a complete dataset of 515 NBA players during the 2022-2023 season. From there we determined 2 research questions of interest:

- 1) How do height and weight determine player success in the league as measured by 'points per game' (PPG)?
- 2) Which Colleges produce the best players by role/position?

## DATA Provenance

### FAIR Principles

1. Findable
  - Metadata & Documentation: Ensure datasets are properly documented with clear descriptions ("NBA Player Statistics 2023" and "NBA Player Bios 2022-23").
  - Constant Identifiers: Use stable URLs (Kaggle dataset DOI or NBA.com's official stats page).

- Searchable Keywords: Tag datasets with relevant terms ( “NBA,” “basketball analytics,” “player performance”).

## 2. Accessible

- Open Access: The Kaggle dataset is publicly available, while NBA.com’s data may have usage restrictions (check terms of service).
- Authentication if Needed: If NBA.com requires API keys, document access procedures.

## 3. Interoperable

- Standardized Formats: Use CSV/JSON for compatibility.
- Common Variables: Merge datasets using player IDs/names for cross-analysis (e.g., linking stats to bios).
- Clear Units: Define metrics (height in cm/ft, 3P% as a decimal).

## 4. Reusable

- Licensing: Kaggle datasets often have CC licenses; NBA.com data may require attribution.
- Methodology Transparency: Document how performance metrics (“best players”) are calculated.

# CARE Principles

## 1. Collective Benefit

- Community Insights: Share findings to benefit teams, players, and fans ( how draft position correlates with success).
- Avoid Harm: Ensure analysis doesn’t reinforce biases (stereotyping players by physical attributes).

## 2. Authority to Control

- Player Consent: Acknowledge that player data is public but avoid misuse (invasive predictions about health).
- Respect NBA Policies: Follow league guidelines on data usage.

### 3. Responsibility

- Bias Mitigation: Check if conclusions favor certain demographics (does “best schools” overlook international players?).
- Transparency: Disclose limitations (stats don’t capture leadership intangibles).

### 4. Ethics

- Privacy: Anonymization isn’t needed for public stats, but avoid combining data to infer private details.
- Fair Representation: Ensure diverse player backgrounds (country, college) are analyzed equitably.

## Data Analysis

**Research Question 3: How do height and weight determine player success in the league as measured by ‘points per game’ (PPG)**

**Research Question 2: Which Colleges produce the best players by role/position**

## Conclusion

## Sources