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

Aditya Kushwaha   Kartik Gupta   Micaela DeLong

2025-05-01

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 different aspects of the game like 'Blocks Per Game', '3 Point Percentage', etc.

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

Before creating any visualizations, we had to clean and manipulate our data. For example, the data pulled from

Here we are using an API call to pull our first data set from NBA.com. This Data set contains vital information like Players physical attributes like Height and Weight as well as other important advanced statistics such as college, country, draft year, draft round, etc. Which can be used as an indicator to provide insights into success.

```
  PLAYER_ID   PLAYER_NAME    TEAM_ID TEAM_ABBREVIATION AGE PLAYER_HEIGHT
1   1630639   A.J. Lawson 1610612742               DAL  22           6-6
2   1631260      AJ Green 1610612749               MIL  23           6-5
3   1631100    AJ Griffin 1610612737               ATL  19           6-6
4    203932  Aaron Gordon 1610612743               DEN  27           6-8
5   1628988 Aaron Holiday 1610612737               ATL  26           6-0
  PLAYER_HEIGHT_INCHES PLAYER_WEIGHT        COLLEGE COUNTRY DRAFT_YEAR
1                   78           179 South Carolina  Canada  Undrafted
2                   77           190  Northern Iowa     USA  Undrafted
3                   78           220           Duke     USA       2022
4                   80           235        Arizona     USA       2014
5                   72           185           UCLA     USA       2018
  DRAFT_ROUND DRAFT_NUMBER GP  PTS REB AST NET_RATING OREB_PCT DREB_PCT USG_PCT
1   Undrafted    Undrafted 15   56  21   2      -20.1    0.046    0.152   0.189
2   Undrafted    Undrafted 35  154  45  22       -4.9    0.016    0.105   0.159
3           1           16 72  639 153  73        1.5    0.026     0.08   0.174
4           1            4 68 1109 446 203       12.1    0.086    0.136   0.206
5           1           23 63  247  74  89        0.9    0.028    0.059   0.129
  TS_PCT AST_PCT
1  0.589   0.032
2  0.607   0.092
3  0.577    0.07
4  0.617   0.129
5  0.528   0.135
```

Next we will import our second Data set from Kaggle which will contain important information like player statistics in different aspects of the game.

Now we will merge our data sets on key attribute 'PLAYER_NAME' to obtain our final cleaned, wrangled data set which will be used for visualization and insights

PLAYER_NAME	PLAYER_HEIGHT_INCHES	PLAYER_WEIGHT	COLLEGE	COUNTRY	DRAFT_YEAR	DRAFT_ROUND	DRAFT_NUMBER	POS	Team	GP	FG%	3P%	FT%	+/-	PPG	RPG	APG	TPG	SPG	BPG
A.J. Lawson	78	179	South Carolina	Canada	Undrafted	Undrafted	Undrafted	G	DAL	15	50.0	40.0	25.0	-46	3.7	1.4	0.1	0.2	0.1	0.0
Aaron Gordon	80	235	Arizona	USA	2014	1	4	PF	DEN	68	56.4	34.7	60.8	518	16.3	6.6	3.0	1.4	0.8	0.8
Aaron Holiday	72	185	UCLA	USA	2018	1	23	G	ATL	63	41.8	40.9	84.4	17	3.9	1.2	1.4	0.6	0.6	0.2
Aaron Nesmith	77	215	Vanderbilt	USA	2020	1	14	SF	IND	73	42.7	36.6	83.8	-211	10.1	3.8	1.3	1.0	0.8	0.5
Aaron Wiggins	77	190	Maryland	USA	2021	2	55	SG	OKC	70	51.2	39.3	83.1	9	6.8	3.0	1.1	0.8	0.6	0.2
Admiral Schofield	77	241	Tennessee	United Kingdom	2019	2	42	SF	ORL	37	45.1	32.4	91.3	-79	4.2	1.7	0.8	0.4	0.2	0.1
AJ Green	77	190	Northern Iowa	USA	Undrafted	Undrafted	Undrafted	G	MIL	35	42.4	41.9	100.0	-26	4.4	1.3	0.6	0.3	0.2	0.0
AJ Griffin	78	220	Duke	USA	2022	1	16	F	ATL	72	46.5	39.0	89.4	63	8.9	2.1	1.0	0.6	0.6	0.2
Al Horford	81	240	Florida	Dominican Republic	2007	1	3	C	BOS	63	47.6	44.6	71.4	309	9.8	6.2	3.0	0.6	0.5	1.0
Alec Burks	78	214	Colorado	USA	2011	1	12	SG	DET	51	43.6	41.4	81.4	-7	12.8	3.1	2.2	1.1	0.7	0.2

Research Question : How do height and weight determine player success in the league as measured by ‘Blocks Per Game’, (PPG), ‘3 Point Percentage’ (3P%) and ‘Rebounds Per Game’ (RPG).

Visualizations:

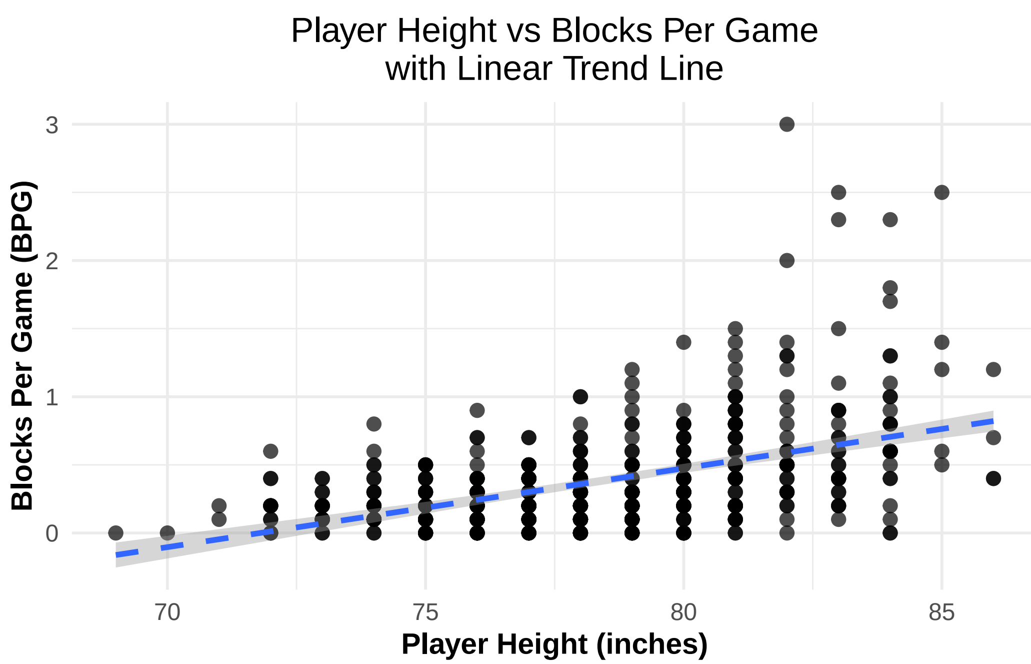

Figure 1: Player Height vs Blocks Per Game

The scatter plot depicting player height versus blocks per game (BPG) reveals a strong and intuitive relationship that underscores the physical advantages conferred by height in professional basketball. As player height increases, there is a noticeable upward trend in blocking ability, with the tallest athletes consistently outperforming their shorter counterparts in this defensive metric. Players under 75 inches (6’3”) tend to average fewer than 0.5 blocks per game, reflecting their physical limitations when contesting shots near the rim. Meanwhile, players standing 80 inches (6’8”) or taller not only appear more frequently among the higher BPG values, but many also average over 1.5 blocks per game — with some elite shot blockers exceeding 2.0. This trend is especially pronounced among centers and power forwards who often anchor their team’s interior defense. The outliers at the top right of the plot — tall players with exceptionally high block rates — represent rim protectors who serve as defensive cornerstones, altering shot trajectories and deterring drives. Overall, the visualization confirms the strong correlation between height and shot-blocking prowess, reinforcing the notion that height is one of the most critical physical attributes for effective interior defense in the NBA.

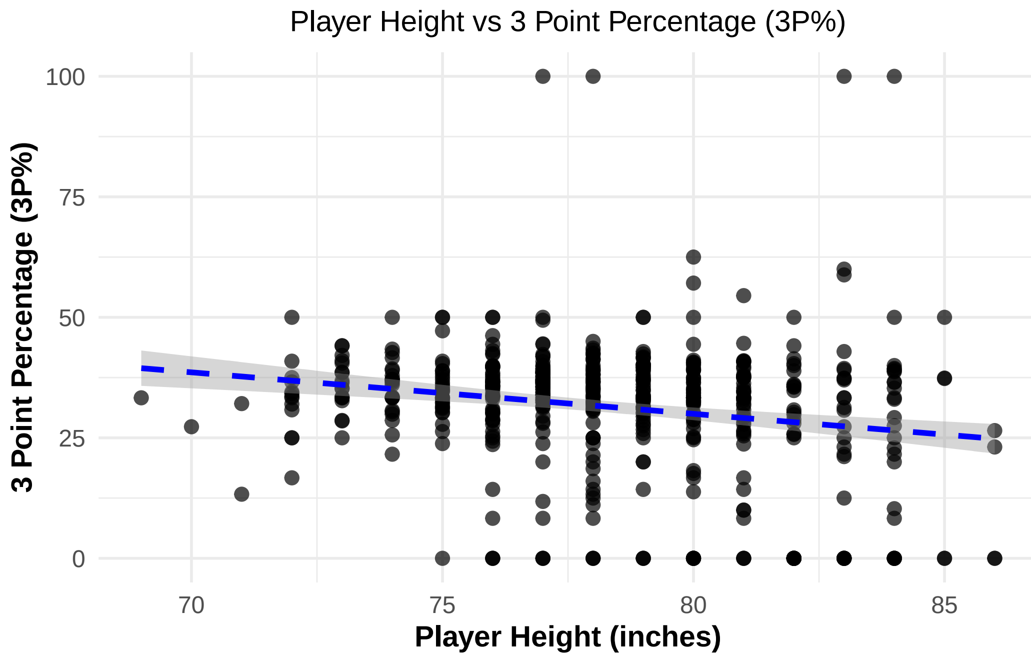

Figure 2: Player Height vs 3 Point Percentage

The scatter plot visualizing player height against 3-point shooting percentage reveals a mild negative trend, suggesting that shorter players tend to shoot more accurately from beyond the arc. Players in the 70–75 inch range show a higher concentration of 3P% values between 35% and 45%, which is considered strong in the NBA. As height increases, the spread of 3P% values becomes more erratic and includes many players with very low or even 0% from three, likely indicating that taller players such as forwards and centers are less likely to attempt or make 3-point shots. This trend aligns with the roles typically assigned in the league—shorter guards are often perimeter shooters, while taller players operate closer to the basket. Although there are some exceptions—tall players with decent shooting percentages—the data overall supports the conventional view that elite 3-point shooting is more prevalent among shorter, backcourt players.

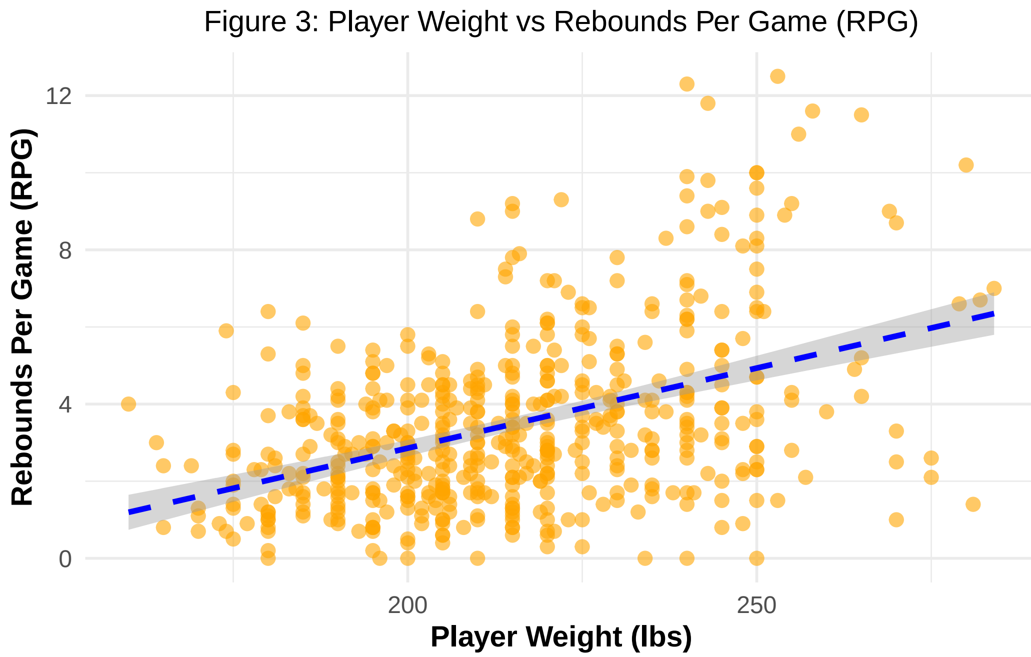

Figure 3: Player Weight vs Rebounds

This scatter plot examines the relationship between player weight and rebounds per game (RPG), revealing a clear upward trend. As weight increases, so too does a player's average number of rebounds per game. Lighter players (under 190 lbs) generally stay below 4 RPG, whereas heavier players—especially those above 240 lbs—frequently exceed 6 or more RPG, with several standout performers reaching double digits. This trend suggests that body mass and physical presence play a crucial role in rebounding success. Heavier players are often better equipped to establish position in the paint, box out opponents, and absorb contact when fighting for boards. The linear regression line reinforces this pattern, showing that weight is positively associated with rebounding output. While rebounding is also influenced by height, timing, and role, this visualization highlights how mass and strength—represented here through weight—are important physical traits for excelling in this aspect of the game.