

# Context-Aware 2K-Style Ratings for NCAA Basketball

Baker Dean & Caleb Ramsey

## Introduction

### Research Question

How can we create a player rating system for men's and women's college basketball that accounts for both player performance and conference strength, allowing staff to compare athletes across divisions and evaluate transfer portal talent more effectively?

### Abstract

Player evaluation has become more complex than ever in this era of transfer portal and Name, Image, and Likeness (NIL) rules. Player movement is at an all-time high with teams often turning over more than 75% of their teams every offseason and with players able to be paid for their services coaches and general managers must be aware of the true value of a player, how much to offer him/her and how to allocate all their financial resources. With many players choosing to move between the Power 5 conferences (ACC, Big Ten, Big 12, Big East, and SEC) and the many smaller conferences, evaluating the skill and fit of a player is not as simple as simply pulling up stats and sorting by points per game. With such a wide variety of levels of talent, determining how to account for the level of competition they faced becomes an important factor alongside the performance of the player.

This project develops a video-game style player ratings system that combines box-score performance with measures of conference strength and schedule difficulty. We use existing HDI ratings (Kartes, 2024) as a baseline model for men's data, then extend, scale, and adapt that framework to the women's game. After constructing conference-level strength metrics, we apply adjustments that account for opponent difficulty and competitive environment. Our long-term goal is a transferable rating system that supports coaching staffs in scouting, recruiting, and evaluating transfer-portal talent.

## Methods

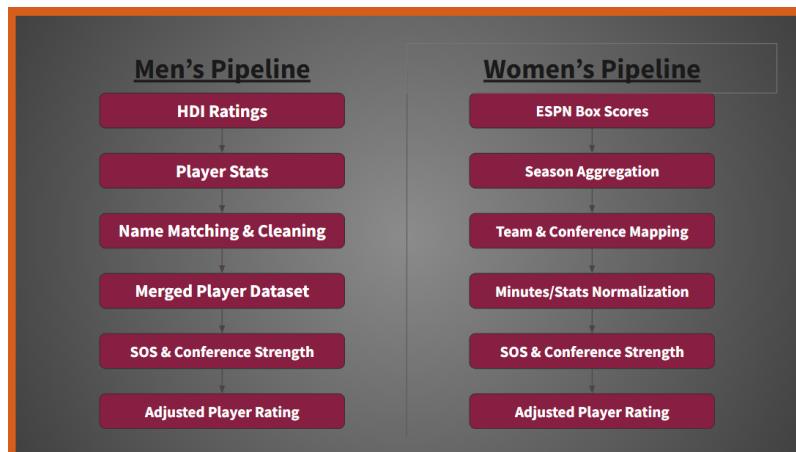


Figure 1: Pipeline diagram

Figure 1 illustrates the pipeline used in this project: data collection, cleaning, modeling, and conference adjustment.

## Data Collection

### Men's

Men's data was collected through two primary sources:

1. HDI ratings & player statistics, accessed through a Google Sheet provided by Virginia Tech Men's Basketball staff and downloaded as a CSV.
2. Barttorvik statistics (Torvik, 2024), scraped using the `cbbdata` package (Weatherman, 2024) for the corresponding seasons.

The two datasets were joined using a custom matching function designed to handle inconsistencies in naming and duplicate names across teams. Players appearing in only one dataset or with incomplete records were removed. An example of the raw dataset can be seen in Table 1 and the cleaned dataset in Table 2.

Table 1: Example of raw men's model data

Name	mpg	ppg	rpg	apg	spg	bpg	2%	3%	ft%	tov	a_to	efg	ts
A.J. Hoggard	26	9.6	2.6	4.6	1.2	0.2	39.0	27.1	73.4	2.2	2.08	39.5	44.4
A.J. Lopez	31	14.8	2.3	1.1	0.8	0.1	52.6	39.1	88.3	1.5	0.76	54.9	60.0
A.J. Staton-McCray	22	7.3	2.4	1.0	1.1	0.6	48.5	32.3	82.9	0.8	1.15	48.5	52.0
A.J. Wills	18	4.8	0.9	1.6	0.3	0.1	37.7	41.8	70.6	1.1	1.44	49.6	52.6
AC Bryant	24	14.0	4.4	1.6	1.4	0.1	50.6	40.7	67.7	1.2	1.29	54.9	57.9

Table 2: Example of cleaned men's model data

mpg	ppg	rpg	apg	spg	bpg	2%	3%	ft%	tov	a_to	efg	ts	Rtg
25	11.2	2.4	1.1	0.9	0.2	50.5	36.2	79.3	1.3	0.84	52.4	58.4	62
23	6.0	1.4	2.2	0.4	0.1	41.0	42.1	75.0	1.2	1.75	54.2	58.3	64
12	3.4	2.4	0.3	0.3	0.4	63.5	0.0	47.3	0.5	0.56	63.5	59.5	66
29	9.7	6.9	2.0	1.0	0.4	64.8	35.8	79.1	1.1	1.84	58.5	61.4	82
25	10.9	2.6	0.5	0.8	0.7	54.1	38.1	83.3	1.0	0.48	55.5	61.3	67

We use these cleaned data as the basis for the men's rating model.

### Women's

Women's data were collected from ESPN via the `wehoop` package (Gilani & Hutchinson, 2021):

- `load_wbb_player_box()` provided game-level box scores for all players in the 2025 season.
- `load_wbb_schedule` provided team IDs, opponents, and conference fields used in conference assignment.

We aggregated player box scores to season totals and standardized minutes, shooting percentages, and efficiency metrics. Teams with ambiguous conference IDs were resolved manually. We restricted the dataset to Division I teams and enforced a minimum threshold (Minutes = 200, Games Played = 10).

Table 3: Example of cleaned women's model data

player	mpg	ppg	rpg	apg	spg	bpg	efg	ts
A'riel Jackson	28	6.9	2.9	2.1	1.1	0.1	37.7	42.7
AC Markham	21	3.6	4.6	1.4	0.5	0.4	39.2	41.8
AJ Marotte	35	11.4	3.4	2.4	0.5	0.4	44.1	46.5
Aaliyah Alexander	30	10.8	3.2	1.2	0.8	0.1	50.2	53.5
Aaliyah Collins	24	8.7	1.5	2.7	2.3	0.5	55.4	57.1

These processed data serve as the basis for the women's rating model.

## Data Manipulation

### Men's Statistical Model

We trained a random forest model to predict HDI ratings using per-game and efficiency statistics, including:

- points, rebounds, assists, steals, and blocks per game (ppg, rpg, apg, spg, and bpg)
- shooting efficiency (2P%, 3P%, FT%)
- effective field goal percentage (eFG%)
- true shooting percentage (TS%)
- turnovers and assist-to-turnover ratio (A/TO)
- free-throw rate (FTR)

This model produces a statistics score that captures each player's individual production independent of conference effects. We apply this model to both the men's (Table 4) and women's (Table 5) datasets.

Table 4: Men's Data

Name	School	Rating
Ryan Kalkbrenner	Creighton	99.0
Johni Broome	Auburn	98.6
Cooper Flagg	Duke	98.2
Hunter Dickinson	Kansas	97.6
Collin Murray-Boyles	South Carolina	97.5

### Women's Role-Adjusted Model

ESPN's women's data do not include reliable positional labels, so we created three role buckets: guard, wing, center. These buckets use simple rules based on height, assist rate, rebounding profile, and shooting tendencies. The goal was not perfect positional classification, but to ensure players were evaluated relative to the expectations of their role. For each player we computed per-game and advanced metrics (TS%, eFG%, usage proxies) and converted them into percentile scores. We then combined these percentiles with role-specific weights. Finally, we integrated team strength, opponent strength-of-schedule, and conference strength to produce a composite rating scaled to 40–99.

Table 5: Women's Data

Name	School	Rating
Paige Bueckers	UConn Huskies	99.0
Ally Becki	Ball State Cardinals	98.1
Katie Dinnebier	Drake Bulldogs	97.5
Serena Sundell	Kansas State Wildcats	97.1
Sarah Strong	UConn Huskies	96.9

## Motivation for Conference Adjustment

Initial outputs revealed clear inconsistencies: players from lower-tier D1 conferences appeared near the top of the ratings despite producing those numbers against comparatively weaker opponents. This confirmed the need for a conference-based adjustment to place players on comparable footing across the country.

### Conference Strength Modeling

#### Men's

Using game results from `hoopR::load_mbb_schedule()` (Gilani, 2024), we restricted to Division I matchups and constructed a least-squares system:

$$(A^T(A) + \lambda(I))\beta = A^T(y)$$

where  $y$  is the home-away score margin and  $\lambda = 10^{-2}$  is a small penalty. The resulting coefficients were centered with mean zero and then standardized to a z-score, yielding a latent team strength measure (`team_z`) for each team.

To quantify strength-of-schedule (`sos_z`), we averaged opponent `team_z` values with weights (0.9 home, 1.1 away). We then aggregated to the conference level and standardized conference means to obtain conference strength (`conf_strength_z`) and conference schedule difficulty (`conf_sos_z`). These form a conference index that adjusts the final player ratings.

## Women's

We replicated the same least-squares approach for women's schedules using `wehoop`:

- compute `team_z` from score margins
- compute `sos_z` from weighted opponent strength
- aggregate to conferences and standardize

We merged these values back onto the women's player table and flagged Power-5 programs to incorporate conference tier effects.

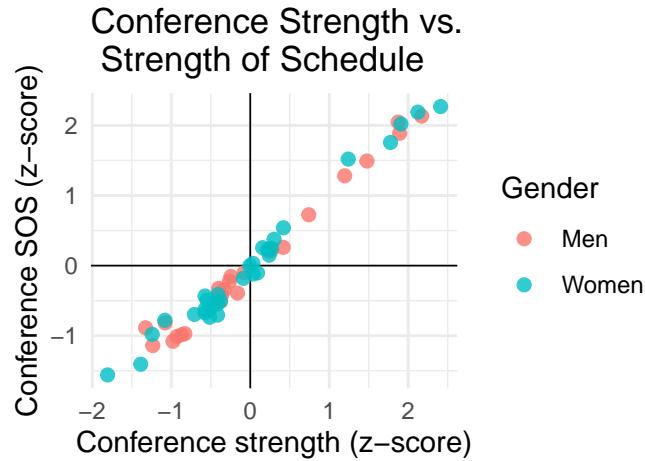


Figure 2: Conference strength across men's and women's conferences

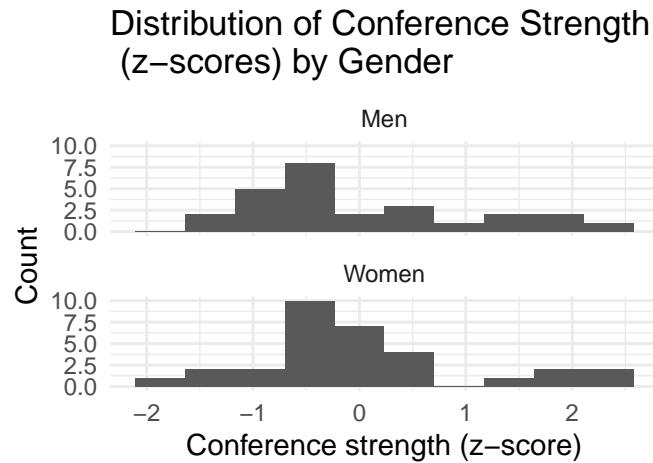


Figure 3: Conference strength vs. strength of schedule for men's and women's conferences.

As shown in Figure 2, conference strength and schedule difficulty are strongly and positively related for both men's and women's leagues. Stronger conferences cluster in the upper-right quadrant, while weaker conferences cluster in the lower-left.

Figure 3 shows that men's conference strength z-scores are more widely dispersed than women's, which are tightly clustered with only a few elite outliers. Together, Figures 2 and 3 motivate incorporating conference strength into our rating system rather than relying on raw box-score production alone.

## Analysis

After applying our conference adjustment formula to the raw stat rankings, here is an example of the final output for men's and women's players:

Table 6: Top Men's Players by Conference-Adjusted Rating

Rank	Name	School	Rating
1	Ryan Kalkbrenner	Creighton	99
2	Johni Broome	Auburn	98
3	Collin Murray-Boyles	South Carolina	97
4	Cooper Flagg	Duke	96
5	Zakai Zeigler	Tennessee	96

Table 7: Top Women's Players by Conference-Adjusted Rating

Rank	Name	School	Rating
1	Paige Bueckers	UConn Huskies	99
2	Serena Sundell	Kansas State Wildcats	98
3	Georgia Amoore	Kentucky Wildcats	98
4	Olivia Miles	Notre Dame Fighting Irish	97
5	Kymora Johnson	Virginia Cavaliers	97

These adjusted rankings seem to be much more realistic, with the top players from top conferences appearing at that, but the next step was to numerically analyze their accuracy. There are 2 ways to accomplish this: 1. Compare our men's ratings to the HDI list to ensure our results are reasonable, even though it was not our goal to duplicate them and 2. Make sure that the teams who have highly rated players were successful.

### HDI Comparison

The first way to display the relationship between the HDI rankings and ours is a scatter plot of each player's ranking in the 2 systems:

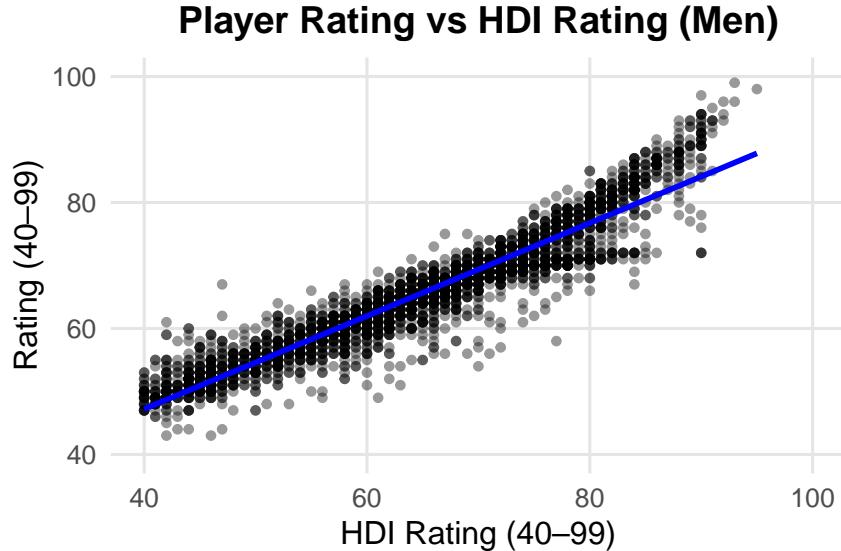


Figure 4: Player Rating vs HDI Rating (Men)

The plot shows a very encouraging trend between the two ranking as they follow a similar trendline and there are no extreme outliers. Most of the minor outliers below the trendline are midmajor players with very high stats from very bad conferences, which is reasonable given how we built the algorithm for the rankings.

To further compare our ratings to HDI, we decided it would be most fair to use the Spearman rho metric to evaluate how similar the ranks of the lists are and then also look at the top 100 players in each set and see how they compared. Here is a table outlining our results:

Table 8: ADJUSTED MEN'S RATINGS VS HDI DIAGNOSTICS

Metric	Value
Matched players	3330
Spearman rho	0.9626
Mean abs rank diff	166.9
Top-100 overlap	74 %

A Spearman rho of 0.9626 is very good for this use, that shows a very strong similarity in the ranks of the lists, suggesting reasonability, while also confirming that our list was not an exact duplication of the HDI. Having a 74% overlap in the top 100 is also within the range of similarity we were targeting, so it appears that the HDI comparison confirms that our list is very reasonable. Considering there are over 3000 players, many of whom in the lower ranks are very similar, an average difference of only 167 spots is quite good.

## Team Strength

Here are the top teams by average rotation rating, where rotation is defined as the 8 highest players on the team since those players will play the majority of the minutes:

Table 9: Top 10 Men's Teams by Average Rotation Player Rating

School	Rating
Auburn	85.5
Florida	84.4
Alabama	84.1
Kentucky	83.4
Arizona	83.1
Duke	82.9
Tennessee	82.6
Illinois	82.5
Iowa St.	82.4
Houston	82.0

Table 10: Top 10 Women's Teams by Average Rotation Player Rating

School	Rating
UConn Huskies	81.5
Baylor Bears	79.8
South Carolina Gamecocks	79.1
Georgia Tech Yellow Jackets	78.8
USC Trojans	78.1
Tennessee Lady Volunteers	77.8
Oklahoma Sooners	77.6
Notre Dame Fighting Irish	77.5
TCU Horned Frogs	77.2
Creighton Bluejays	77.1

This list of teams is very encouraging to the accuracy of our ratings because it includes both National Champions last season within the top 2 teams (Florida men and UConn women), both National Player of the Years and the runner ups (Cooper Flagg - Duke and Juju Watkins - USC), 6 of the 8 Final 4 teams, and 10 of the 16 Elite 8 teams, all very high percentages for a single-game elimination tournament with so much volatility. The men's teams included were seeded 1,1,1,1,2,2,2,3,3,4 and the women's teams were seeded 1,1,2,2,3,3,4,5,9,9.

By combining the analysis of the these methods we can conclude with confidence that our ratings are reasonable and could be applied to make decisions about the value of college basketball players.

## Limitations

### Men's-to-women's model translation

One of the problems we ran into was applying a model trained on men's data to the women's game, the original outputs were too low (top players were in the 70s) due to the different paces and styles of play. We compensated for this by scaling the ratings but the women's teams are still slightly lower in general.

### Heavy reliance on box-score stats

Both the men and women's models are constrained to box-score variables because play-by-play and advanced tracking data are unavailable for the full dataset. This limits the model's ability to capture defensive impact, screening, spacing, off-ball value, and other impacts players may have that aren't captured in the traditional box-score.

### Margin-based team strength models assume linearity

Our team strength estimates rely on least-squares systems that use scoring margins. Margin-based ratings assume linear, symmetric behavior and may overweight blowouts or undervalue teams that eke out close games, especially close wins versus good teams.

## Next Steps

This project could be further advanced in a variety of ways. The simplest future steps would be expanding to include more years worth of data in the training and analysis sections and gathering more statistics beyond the basic box score ones that are freely available. The HDI ratings were based on a wide variety of advanced metrics, so while it is encouraging that we were able to make a system that is constructed more simply there could still be value in tuning using more complex stats.

It would also be interesting to seek another women's ranking service and compare/tune based on their rankings the same way we did with HDI.

One of the big benefits to staffs is the ability to project future value for players, not only the current season so adding models to predict year-over-year change in individual rating and adjustment when transferring schools could be particularly valuable.

## References

- Weatherman A (2024). **cbbdata**: API for College Basketball Data. R package version 0.3.0.9000, <https://cbbdata.aweatherman.com/>
- Kartes, Weston. MBB Player Database 2024. HD Intelligence. <https://www.hdintelligence.com/mbb-player-database-2024>
- Kartes, Weston. MBB Player Database 2025. HD Intelligence. <https://www.hdintelligence.com/mbb-player-database-2025>
- Torvik, Bart. NCAAAM Player Stats 2024. barttorvik. <https://barttorvik.com/playerstat.php?year=2024>
- Torvik, Bart. NCAAAM Player Stats 2025. barttorvik. <https://barttorvik.com/playerstat.php?year=2025>
- Gilani & Hutchinson. **wehoop**: Women's Basketball Data for R: <https://github.com/sportsdataverse/wehoop>
- Gilani, S. **hoopr**: The SportsDataverse's R Package for Men's Basketball Data: <https://hoopR.sportsdataverse.org>