Beyond the Hype: Predicting the 2023 National Basketball Association (NBA) Draft Class

Caleb Miller

Master's of Data Science Candidate (Drexel)

* Not affiliated with the NBA

Data Collection

- NBA Players
  - College Statistics & Player Efficiency Rating (PER)
  - Seasons included: 1982-83, 1992-93, 2002-02, 2012-13, & 2022-23
    - Represent different eras of the NBA
    - Data Source: https://www.basketball-reference.com/
- 2023 NBA Draft Class
  - College Stats collected
  - PER will be predicted by models
    - Data Source: https://basketball.realgm.com/

Data Cleaning & Preparation

- Merge NBA Player data
- Excluded:
  - Duplicate entries
  - Rows containing null values
    - i.e. Played when statistics were not tracked
- Profile URL, ‘name’, and ‘age’ columns dropped
- Outliers/unrealistic data removed (22 players)
- Cleaned dataset: 898 rows x 17 columns

NBA PER Tier Definition

Lower quartile →

Upper quartile →

Career PER	Type of Player
Less than 9	Won't stick around in the NBA
9 - 11	Fringe roster player
11 - 13	Will get a roster spot
13 - 15	Rotation player
15 - 17	Good role player
17 – 22	All-Star
22+	Superstar

← The average player from the data

Type of Player	Data Rows
Won't stick around in the NBA	85
Fringe roster player	126
Will get a roster spot	189
Rotation player	197
Good role player	131
All-Star	118
Superstar	17

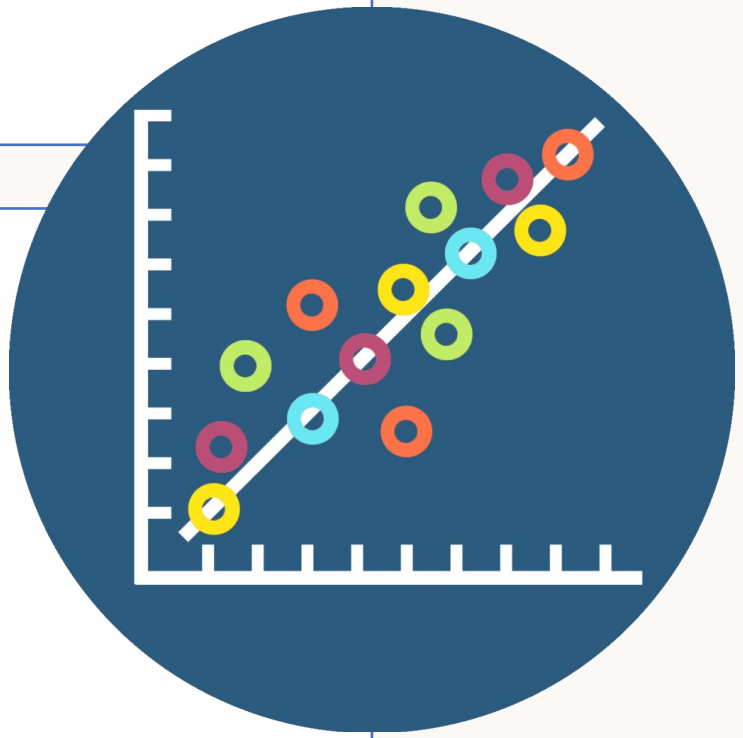
Descriptive Statistics

Normal Distribution – median PER is 13.2 and mean PER is 13.29

- 25th Percentile categorized as 'will get a roster spot'
- 50th Percentile categorized as a 'rotation player'
- 75th Percentile categorized as a 'good role player'

Correlation matrix

- Positive Correlations:
  - 1st: Field-Goal Percentage (0.4)
  - 2nd: Blocks (0.36)
  - 3rd: Rebounds (0.35)
- Negative Correlations:
  - 1st: Three-Pointers Attempted (-0.12)
  - 2nd: Three-Pointers Made (-0.1)
  - 3rd: Three-Point Percentage (-0.082) – very logical
  - 4th: Free-Throw Percentage (-0.039)

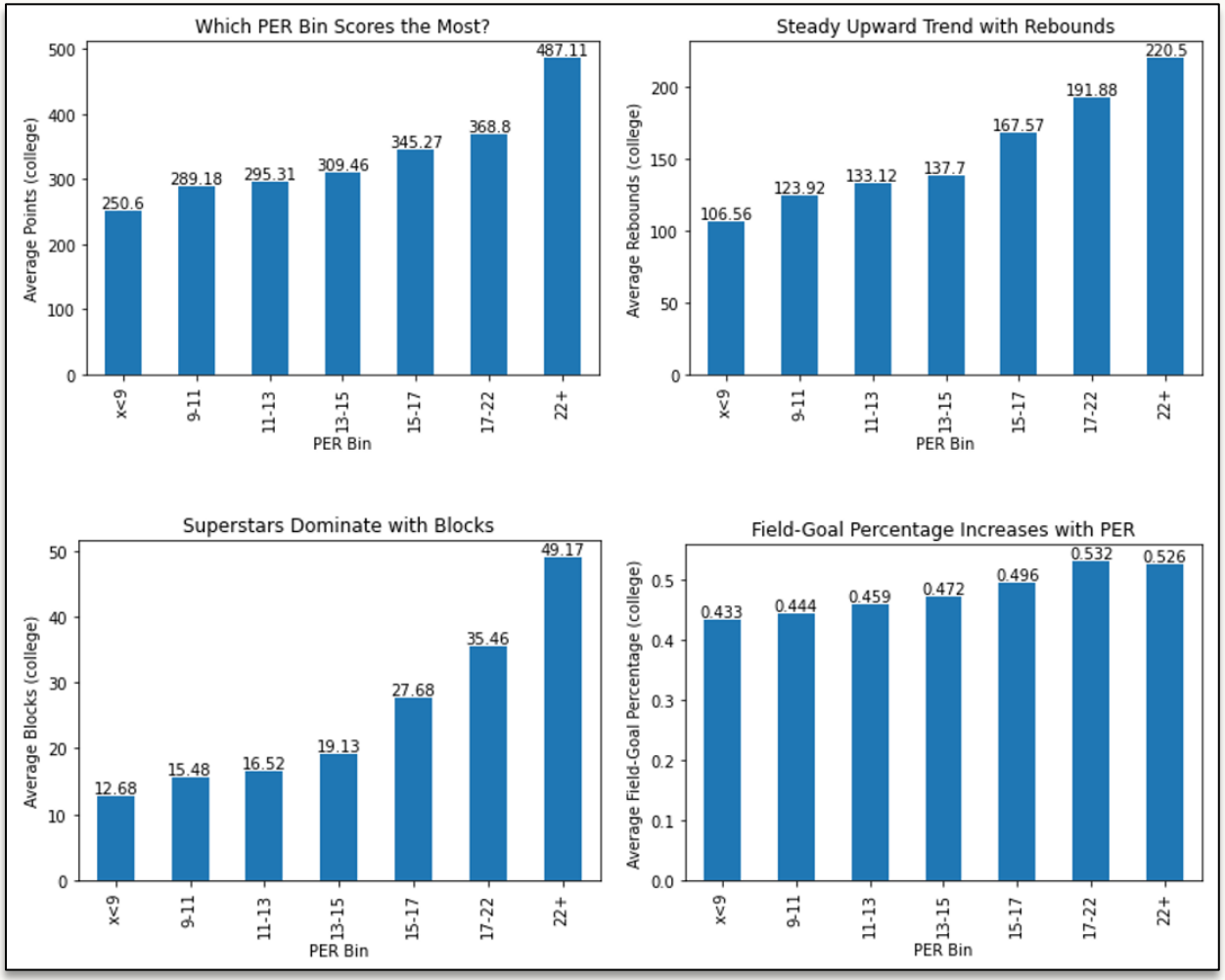


Data Analysis

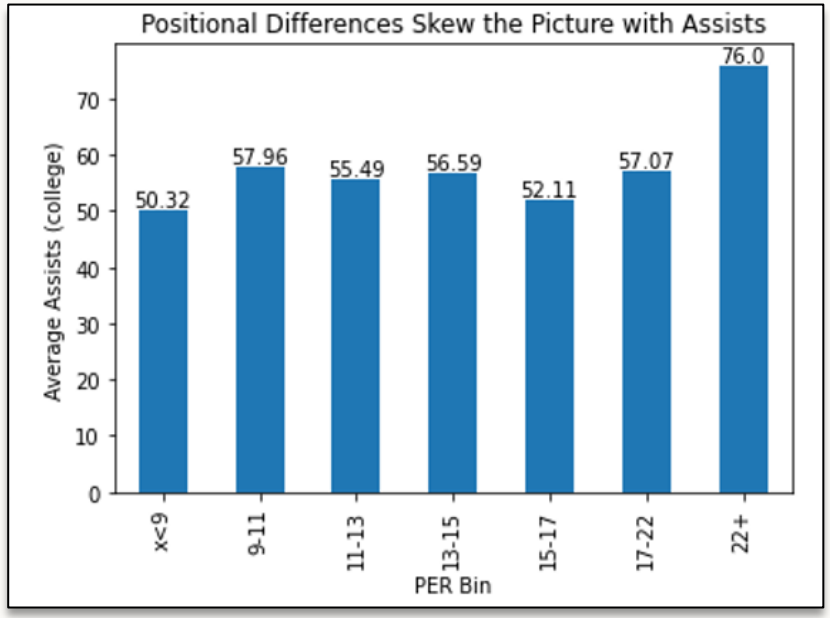
Players were assigned to bins

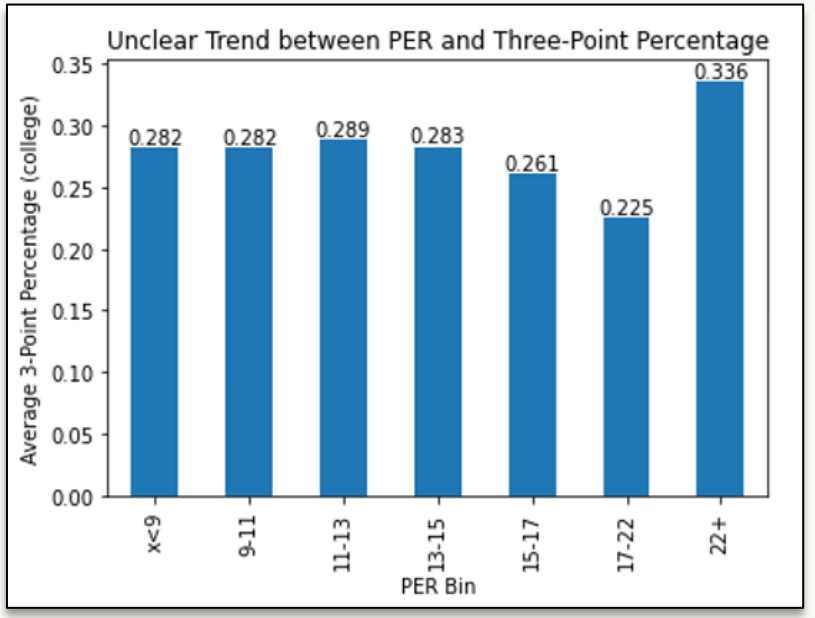
‘Superstar’ domination

Scored 29.64% more points	Grabbed 14.92% more rebounds	Collected 38.66% more blocks



Data Analysis

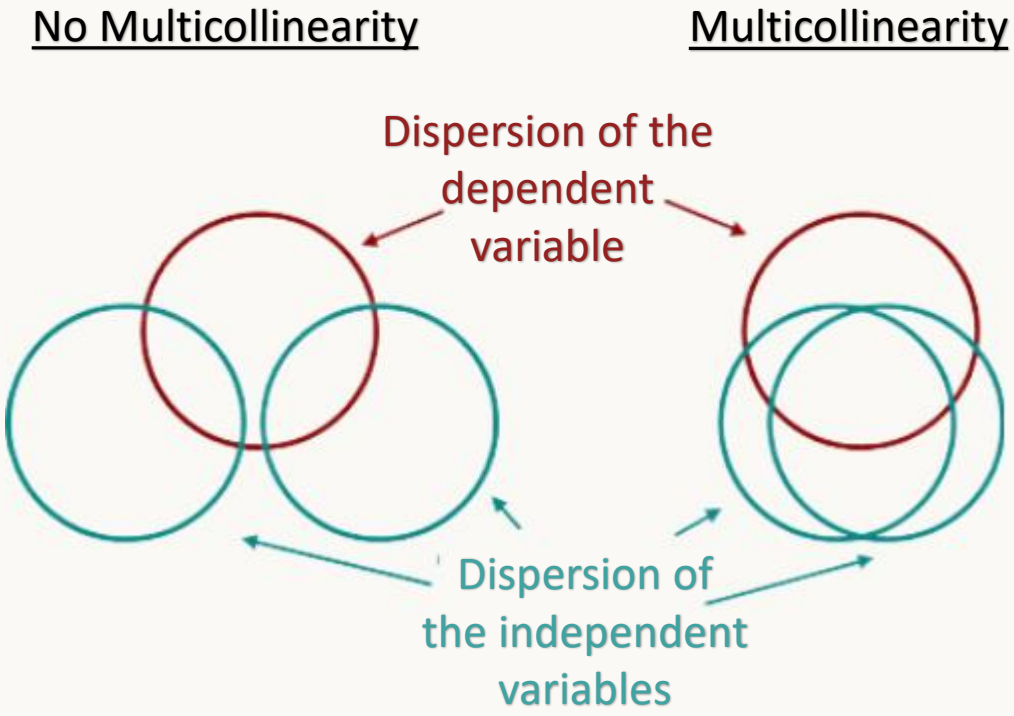




- Some relationships are less clear
  - Three-Point Percentage and PER trend negatively until the 'Superstar' bin jumps
  - Relationship between Assists and PER shifts erratically
    - Likely due to positional needs: i.e., guards account for the vast majority of assists

Data Analysis

- Need to drop variables due to multi-collinearity:
  - Minutes
  - Field-Goals Made
  - Field-Goals Attempted
  - Free-Throws Made
  - Free-Throws Attempted
  - Three-Pointers Made
  - Three-Pointers Attempted



Model Training & Performance

Data is split into two datasets:

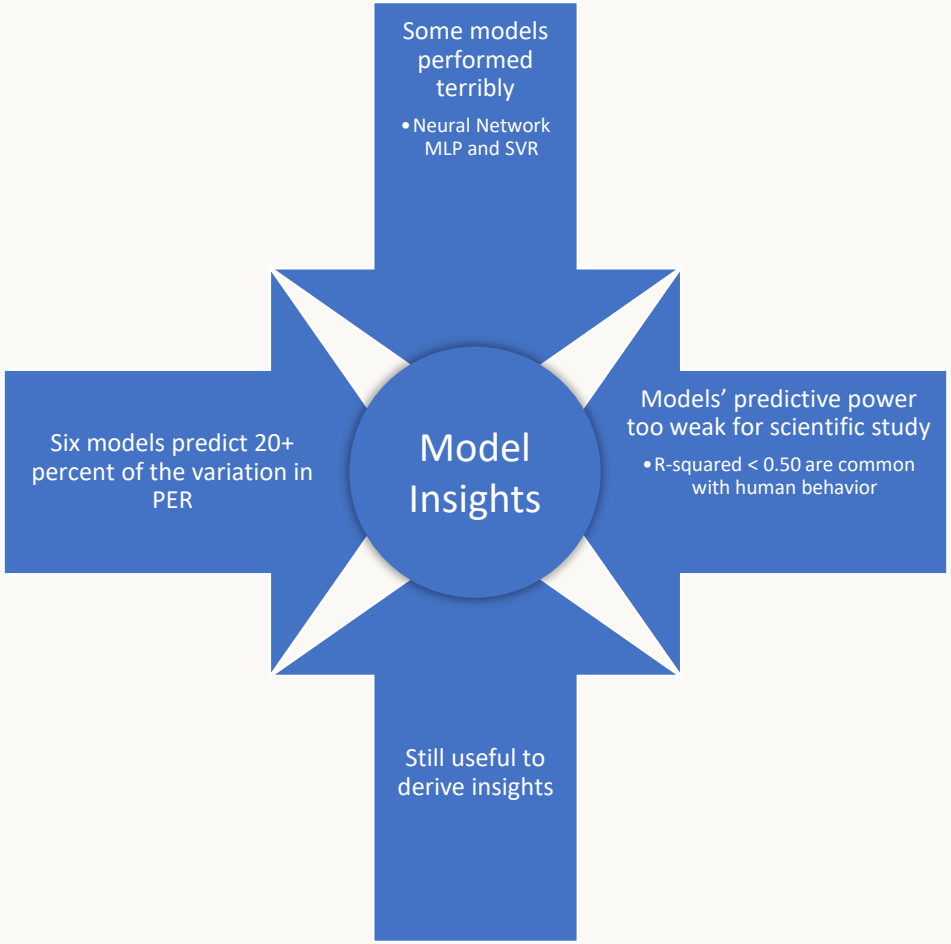
- One holds 75% of data for training models
- Other holds 25% of data for testing models' accuracy

10 models were trained using the Sklearn package within Python

Three different metrics were used to measure accuracy:

- R-Squared
- Root Mean Squared Error
- Mean Absolute Error

Model Training & Performance



Model Name	R Squared	Root Mean Squared Error	Mean Absolute Error
Quantile	0.272857	3.346237	2.582522
Lasso	0.272589	3.346854	2.583629
Linear Regression	0.262390	3.370237	2.598229
XGBoost	0.241699	3.417178	2.621969
Bayesian Ridge	0.236643	3.428552	2.643920
Random Forest	0.224650	3.455381	2.633166
KNN	0.150318	3.617222	2.713222
Decision Tree	0.116342	3.688832	2.869624
SVR	0.016237	3.892173	2.994967
Neural Network MLP	-0.035017	3.992278	2.988547

NBA Rookie Career PER Prediction: Quantile Regression

Ranking in Draft Class	Drafted	Name	Predicted Career PER
1	57	Trayce Jackson-Davis	19.39
2	1	Victor Wembanyama	18.69
3	12	Dereck Lively II	17.80
4	Undrafted	Drew Timme	17.25
5	54	Jalen Slawson	17.24
23	9	Taylor Hendricks	15.08
28	6	Anthony Black	14.57
30	8	Jarace Walker	14.39
35	7	Bilal Coulbaly	14.22
41	2	Brandon Miller	13.97
42	10	Cason Wallace	13.94
71	13	Gradey Dick	12.90
86	11	Jett Howard	12.05
90	14	Jordan Hawkins	11.97

NBA Rookie Career PER Prediction: Lasso Regression

Ranking in Draft Class	Drafted	Name	Predicted Career PER
1	57	Trayce Jackson-Davis	19.39
2	1	Victor Wembanyama	19.00
3	12	Dereck Lively II	17.56
4	54	Jalen Slawson	17.12
5	Undrafted	Drew Timme	17.09
21	9	Taylor Hendricks	15.17
27	6	Anthony Black	14.57
28	8	Jarace Walker	14.49
36	2	Brandon Miller	14.11
39	7	Bilal Coulbaly	14.09
43	10	Cason Wallace	13.90
72	13	Gradey Dick	12.91
87	11	Jett Howard	12.15
90	14	Jordan Hawkins	12.07

NBA Rookie Career PER Prediction: Linear Regression

Ranking in Draft Class	Drafted	Name	Predicted Career PER
1	57	Trayce Jackson-Davis	19.37
2	1	Victor Wembanyama	18.58
3	12	Dereck Lively II	17.89
4	Undrafted	Drew Timme	17.28
5	54	Jalen Slawson	17.28
23	9	Taylor Hendricks	15.06
28	6	Anthony Black	14.56
32	8	Jarace Walker	14.35
35	7	Bilal Coulbaly	14.25
41	10	Cason Wallace	13.95
42	2	Brandon Miller	13.92
71	13	Gradey Dick	12.90
86	11	Jett Howard	12.03
89	14	Jordan Hawkins	11.94

NBA Rookie Career PER Prediction: Standouts

Highest predicted PER < 20

- Would categorize the highest-ranked players as ‘good role players’
- Does not align with public opinion
- Could be indicative that the draft class was overhyped

Athlete’s ranking among draft class is weighted heavier than predicted PER

NBA Rookie Career PER Prediction: Standouts

- Prospect value varies between models and draft
  - The models rank some prospects much higher:

Trayce Jackson-Davis
Ranked #1 in all three models
Projected second-round pick
Drafted 57th overall

Drew Timme
Ranked #4, #4, and #5
Projected late second-round pick
Undrafted

Jalen Slawson
Ranked #4, #5, and #5
Projected second-round pick
Drafted 54th overall

  - The models rank some prospects much lower:

Brandon Miller
Ranked 36th, 41st, and 42nd
Projected top-3 pick
Drafted 2nd overall

Jordan Hawkins
Ranked 89th, 90th, and 90th
Projected first-round pick
Drafted 14th overall

Flaws with the Models

- The presence of unquantifiable statistics
  - Work ethic, overall potential, etc.
  - Common issue when predicting human behavior
- Omitted data due to null values
- No data for some athletes
  - Drafted from High School (LeBron, Kobe, etc.)
  - Drafted from G-League or overseas (Scoot Henderson, Luka Doncic)
- A better source of data would likely increase predictive power

Key Takeaways

Strongest positive correlations with PER:

1st - Field-goal percentage
2nd - Blocks
3rd - Rebounds

Strongest negative correlations with PER:

1st - Three-point percentage
2nd - Free-throw percentage

Trayce Jackson-Davis, Drew Timme, and Jalen Slawson project to be top-5 players

Victor Wembanyama projects to be #2

Brandon Miller and Jordan Hawkins don’t project to live up to the hype

Key Takeaways

Trayce Jackson-Davis might not end up being the best player from this draft.

However, he could be the steal of the draft

Drew Timme and Jalen Slawson may not become top 5 players

Yet, they could still be diamonds in the rough

Brandon Miller and Jordan Hawkins might not be busts

Although, they may not live up to the expectations of a lottery pick

Work Cited

- RealGM LLC. (n.d.). *Basketball News, rumors, scores, stats, analysis, depth charts, forums.* Basketball.RealGM.com. https://basketball.realgm.com/
- Sports Reference LLC. (n.d.). *Basketball Statistics & History.* Basketball-Reference.com. https://www.basketball-reference.com/