# NBA Playoff Performance Modeling

Beckett Newton
NYU Data Science, Fall 2024

## Data Collection

This project uses a publicly sourced dataset from Kaggle containing NBA playoff statistics from 2000 to 2023. The dataset includes advanced metrics such as:

- Offensive Rating (ORtg)
- Defensive Rating (DRtg)
- Pace
- Free Throw Rate (FTr)
- Three-Point Attempt Rate
- True Shooting Percentage (TS%)
- Turnover Percentage (TOV%)
- Offensive and Defensive Rebound Percentage (ORB%, DRB%)

These statistics extend beyond traditional counting stats to offer deeper insight into team performance during the postseason.

## Data Preprocessing

Before analysis, the dataset was cleaned by:

- Removing columns with missing values (NAs)
- Dropping rows with duplicate headers
- Excluding the 2023 playoff data due to incomplete results
- Removing irrelevant columns such as `Rk`, `PL`, and `PW`

| | Team | Age | W | L | WL_percent | OPtg | DRtg | NRtg | Pace | FTr | ... | TS_percent | eFG_percent_offense | TOV_percent_offense | ORB_percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Portland Trail Blazers | 30.0 | 10 | 6 | 0.625 | 107.7 | 102.2 | 5.5 | 85.9 | 0.386 | ... | 0.541 | 0.481 | 12.9 | 25.9 |
| **1** | Miami Heat | 29.6 | 6 | 4 | 0.600 | 97.4 | 94.4 | 3.0 | 84.6 | 0.328 | ... | 0.493 | 0.451 | 14.1 | 25.7 |
| **2** | Indiana Pacers | 30.9 | 13 | 10 | 0.565 | 109.7 | 106.8 | 2.9 | 89.0 | 0.322 | ... | 0.545 | 0.488 | 11.4 | 25.1 |
| **3** | Los Angeles Lakers | 29.3 | 15 | 8 | 0.652 | 110.1 | 107.5 | 2.6 | 90.3 | 0.393 | ... | 0.529 | 0.498 | 11.4 | 30.6 |
| **4** | Milwaukee Bucks | 27.6 | 2 | 3 | 0.400 | 107.5 | 105.5 | 2.0 | 89.3 | 0.266 | ... | 0.520 | 0.479 | 11.3 | 28.5 |

Figure 1: Preview of the cleaned playoff dataset

## Descriptive Statistics

To better understand the dataset and establish a baseline for interpretation, we begin by examining summary statistics for key performance variables. These include Net Rating (NRtg), Offensive Rating (ORtg), True Shooting Percentage (TS%), Three-Point Attempt Rate (3PAr), Free Throw Rate (FTr), Rebounding percentages, Pace, and Turnover Percentage.

| | NRtg | OPtg | TS_percent | 3PAr | DRB_percent | FT/FGA_offense | FTr | ORB_percent | Pace | TOV_percent_offense |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 | 368.000000 |
| **mean** | -2.843478 | 105.536957 | 0.530609 | 0.269429 | 74.448913 | 0.234712 | 0.310361 | 25.110326 | 91.568478 | 12.994565 |
| **std** | 6.857285 | 6.303175 | 0.032519 | 0.087085 | 4.240167 | 0.042296 | 0.056359 | 4.382881 | 4.602024 | 1.750974 |
| **min** | -27.000000 | 87.500000 | 0.443000 | 0.087000 | 62.000000 | 0.124000 | 0.160000 | 13.500000 | 82.400000 | 8.100000 |
| **25%** | -7.200000 | 101.900000 | 0.507750 | 0.207000 | 71.675000 | 0.205500 | 0.270000 | 22.075000 | 87.900000 | 11.900000 |
| **50%** | -1.650000 | 105.500000 | 0.529000 | 0.254000 | 74.650000 | 0.231000 | 0.305500 | 24.900000 | 91.000000 | 12.850000 |
| **75%** | 2.000000 | 110.125000 | 0.551250 | 0.328500 | 77.200000 | 0.266000 | 0.345250 | 28.150000 | 94.625000 | 14.125000 |
| **max** | 13.700000 | 122.200000 | 0.616000 | 0.545000 | 85.700000 | 0.359000 | 0.469000 | 36.700000 | 104.400000 | 19.300000 |

Figure 2: Descriptive statistics of selected advanced playoff metrics (2000–2022)

## Exploratory Data Analysis (EDA)

To validate whether the data is meaningfully associated with team success, we conduct a hypothesis test on Net Rating (NRtg)—a composite metric combining offensive and defensive performance.
**Null Hypothesis:** There is no difference in Net Rating between teams with above- and below-median winning percentages.
**Alternative Hypothesis:** Teams with above-median winning percentages have significantly higher Net Ratings.

- T-statistic: 18.73
- P-value: $1.18 \times 10^{-55}$

The extremely small p-value provides strong evidence against the null hypothesis, confirming that winning teams tend to have significantly higher Net Ratings. This validates the dataset's utility for further modeling.

## Correlation Analysis

Next, we explore the relationship between each team metric and playoff winning percentage. This allows us to identify which variables may be most predictive of postseason success.

| | Variable | R_value | P_value |
|---|---|---|---|
| **2** | NRtg | 0.753711 | 1.048701e-68 |
| **0** | OPtg | 0.435343 | 1.886529e-18 |
| **6** | TS_percent | 0.416663 | 6.913391e-17 |
| **7** | eFG_percent_offense | 0.397925 | 2.058786e-15 |
| **5** | 3PAr | 0.130485 | 1.223342e-02 |
| **13** | DRB_percent | 0.112449 | 3.103254e-02 |
| **10** | FT/FGA_offense | 0.103116 | 4.808091e-02 |
| **4** | FTr | 0.095764 | 6.649966e-02 |
| **12** | TOV_percent_defense | 0.073491 | 1.594541e-01 |
| **9** | ORB_percent | 0.071297 | 1.723172e-01 |
| **3** | Pace | -0.050066 | 3.381829e-01 |
| **8** | TOV_percent_offense | -0.103298 | 4.768531e-02 |
| **14** | FT/FGA_defense | -0.180610 | 4.984555e-04 |
| **11** | eFG_percent_defense | -0.322730 | 2.291273e-10 |
| **1** | DRtg | -0.392218 | 5.551531e-15 |

Figure 3: Correlation of Team Statistics with Playoff Winning Percentage

## Data Analysis

To identify which metrics most strongly impact playoff success, we extract the correlation coefficient (R-value) for each feature relative to winning percentage.

- **Net Rating (NRtg)** shows the highest positive correlation, confirming that overall efficiency—balancing both offensive and defensive effectiveness—is the most predictive of wins.
- **Offensive Rating (OPtg)** and **True Shooting Percentage (TS_percent)** demonstrate moderate positive correlations, highlighting the importance of scoring efficiently.
- **Defensive Rating (DRtg)** and **Defensive eFG_percent** show moderate negative correlations, reinforcing the value of strong defense.
- Other variables such as **Pace**, **Offensive Rebound Percentage (ORB_percent)**, **Defensive Turnover Percentage (TOV_percent_defense)**, and **Free Throw Rate (FTr)** exhibit statistically insignificant correlations.

These results align with intuitive expectations—teams that score efficiently and limit opponent scoring are the most successful in the playoffs.

## Trends in the Data

To further contextualize these relationships, we explore how certain team metrics have evolved over time. We begin be correlating team metrics with year.

| | Variable | R_value | P_value |
|---|---|---|---|
| **5** | 3PAr | 0.774521 | 8.521001e-75 |
| **3** | Pace | 0.588665 | 1.082759e-35 |
| **11** | eFG_percent_defense | 0.566965 | 1.105338e-32 |
| **7** | eFG_percent_offense | 0.519131 | 8.795902e-27 |
| **13** | DRB_percent | 0.504551 | 3.663769e-25 |
| **6** | TS_percent | 0.452143 | 6.080698e-20 |
| **1** | DRtg | 0.397698 | 2.142469e-15 |
| **0** | OPtg | 0.368895 | 2.631193e-13 |
| **2** | NRtg | -0.019400 | 7.106869e-01 |
| **8** | TOV_percent_offense | -0.268650 | 1.671698e-07 |
| **12** | TOV_percent_defense | -0.269587 | 1.508526e-07 |
| **10** | FT/FGA_offense | -0.358256 | 1.383098e-12 |
| **14** | FT/FGA_defense | -0.394277 | 3.890071e-15 |
| **4** | FTr | -0.417421 | 5.999519e-17 |
| **9** | ORB_percent | -0.457413 | 1.989702e-20 |

Figure 4: League-Wide Trends Correlated With Year

If we look at trends in the data over time, we can see an upward we can see a strong correlation between **3-Point Attempt Rate (3PAr)** and **Pace** with year. This explains the low correlation coefficients in pace and 3PAr with wins since we can see that these features have a much higher correlation with year. As any basketball fan would agree, there were different play-styles between eras as the game evolved to focus on more efficient scoring: high quality shots close to the basket, and 3 pointers which has reduced the frequency of mid-range shots.

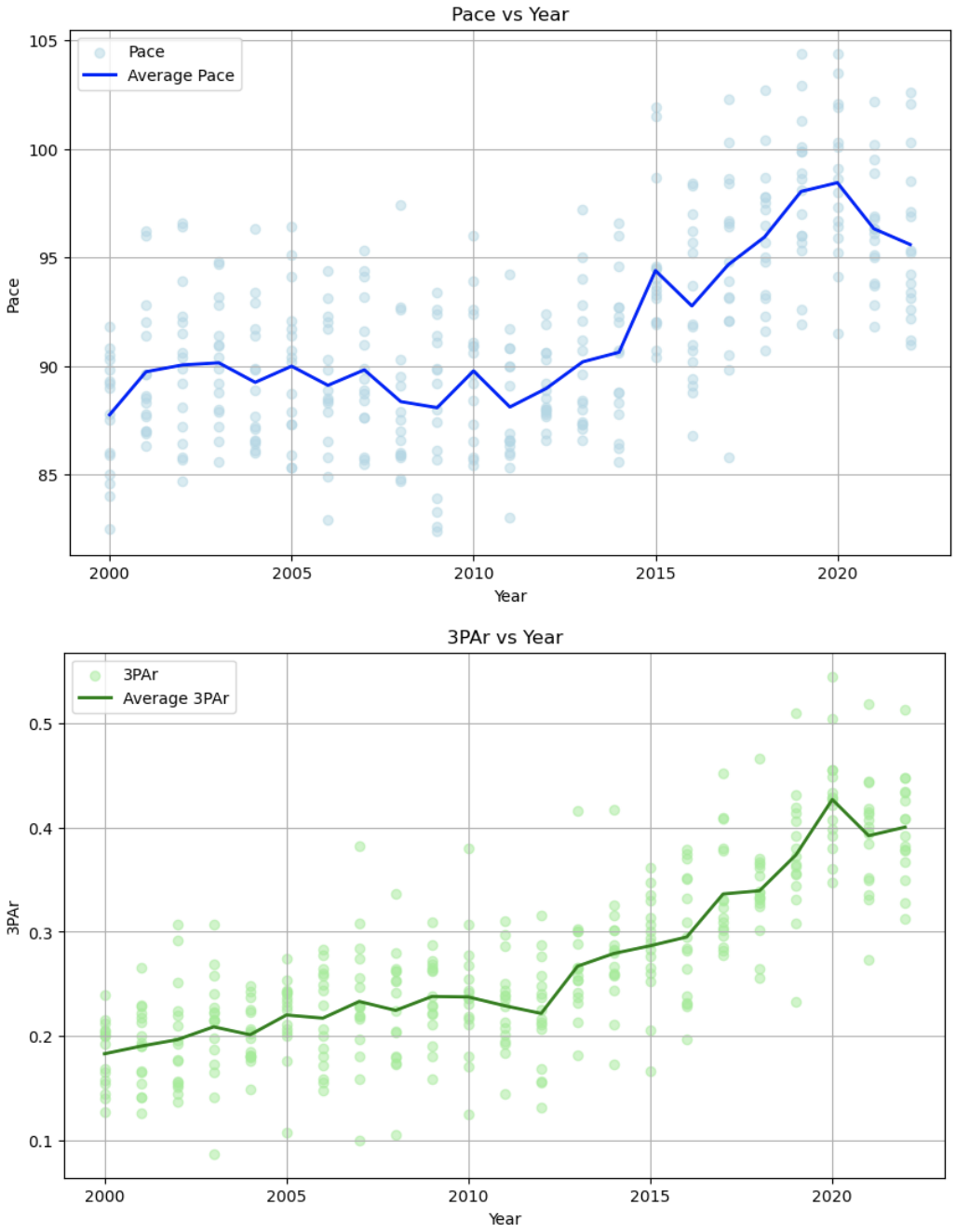


Figure 5: Trends Over Time: Pace and 3PAr (Three-Point Attempt Rate) vs. Year

- The early 2000s, often dubbed the *"dead-ball era"*, emphasized half-court offense and rugged defense, leading to slower pace and lower 3PAr.
- Post-2015, the league has increasingly favored *"pace-and-space"* offenses, resulting in rapid increases in both pace and reliance on three-point shooting.

These shifts underscore the importance of contextualizing basketball statistics across eras. While pace and 3PAr are not directly predictive of playoff success, their rise reflects a broader strategic shift in how teams approach the game.

## Feature Importance for Playoff Wins

Because features in basketball statistics are often correlated (for example, Offensive Rating and True Shooting Percentage), we built a **Ridge Regression** model to address multicollinearity. Ridge regression applies a penalty to large coefficients, stabilizing the model and providing more reliable insights into which factors are most associated with playoff wins. The ML approach also helps address collinearity concerns by shrinking correlated features like pace and 3PAr, which are strongly tied to year, reducing their undue influence in the model.

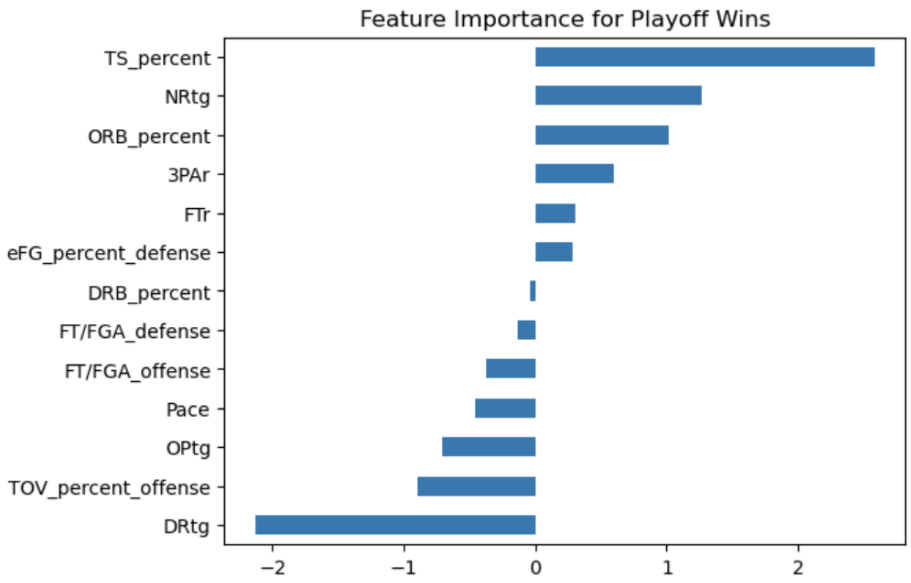


Figure 6: Feature importance for playoff wins as determined by Ridge Regression. Positive coefficients indicate a positive association with wins, while negative coefficients indicate a negative association.

**Top positive contributors:**

- TS% (True Shooting Percentage): +2.582
- NRtg (Net Rating): +1.269
- ORB% (Offensive Rebound Percentage): +1.021

**Top negative contributors:**

- DRtg (Defensive Rating): −2.128
- TOV% (Offense): −0.895
- OPtg (Opponent Offensive Rating): −0.706

**Model performance:**

- Mean Squared Error (MSE): 6.33
- $R^2$: 0.642

## Ridge Regression Analysis

We see that the model has a $R^2$ of 0.638, which means that the model explains about 64% of the variance in playoff wins, indicating a reasonably good fit.

True Shooting Percentage (1.76) has the highest positive influence. A one-standard-deviation increase in TS_percent leads to a 1.76 standard deviation increase in playoff wins, making it the most critical predictor for success. Interpretation: Teams with efficient scoring (higher TS%) tend to perform better in the playoffs, which aligns with our prior understanding that it is the most efficient teams who are most successful.

A new insight that was made using the Ridge Regression model was that a one-standard-deviation increase in Free Throw Attempts per Field Goal Attempt corresponds to a 1.16 standard deviation increase in wins. Interpretation: Drawing fouls and scoring from free throws is vital for playoff success.

Finally, a fascinating insight that the model found was that teams with slightly higher turnover percentages may still succeed. A positive coefficient in this metric (1.12) suggests that despite turnovers harming a teams offensive output, it is also an indicator of fast-paced and aggressive play styles being successful in creating scoring opportunities in the playoffs, despite the fact that they may lead to increased turnover rates. This same thinking applies to offensive rebound percentage (-1.04) where a one-standard-deviation increase in offensive rebound percentage corresponds to a 1.04 standard deviation decrease in playoff wins. Logically, more offensive rebounds corresponds to more second-chance scoring opportunities for the offense, however what the analysis tells me is that focusing too much on offensive rebounds negatively impacts winning as it likely leaves the team in a vulnerable defensive position in transition when an offensive rebound is not secured.

## Conclusion

In conclusion, I sought out to find the key metrics that are most important to success on the basketball court. I determined that the most honest data would come from the playoffs, when winning is the most vital. The playoff data that I ultimately used was sourced from publicly available data on Kaggle in the form of a CSV which is attached in the repository of this project. I then cleaned the data so that incomplete or missing data was not included in my analysis, and then ran a regression to determine the most pertinent metrics towards winning. While I was able to diagnose efficiency as a key metric towards success, I also identified trends of evolution (ie: more 3-point shooting and faster pacing) when I regressed the metrics by year. Faced with collinearity concerns, I built a ridge regression model to be able to best understand the metrics that lead to playoff success. In my analysis of the model's results, I was able to confirm my prior understanding of efficiency as a paramount component to playoff success, but I was also able to uncover a few other metrics that correlate to wins on the basketball court, leading to my final findings for optimizing wins in the playoffs:

### Prioritizing Offensive Efficiency

Prioritizing efficiency on the offensive end is necessary to be successful in the playoffs. This means taking quality, high probability shots (TS_percent aka true shooting percentage) and getting to the free throw line at a high rate (FT/FGA_offense aka free throws per shot attempt on offense).

Success in both of these metrics demonstrates offensive schemes that are both efficient and capitalize on free throw opportunities.

### The Importance of Disciplined Defense

Teams with a lower defensive rating (which is indicative of better defensive performance) are more successful in the playoffs as a one-standard-deviation decrease in defensive rating is equivalent to a 2.24 standard deviation increase in wins. This metric is based in avoiding fouls, and optimizing defensive rotations which are both grounded in playing good fundamental, disciplined defense.

### Play Style

Finally, my project has revealed that it is incredibly important for teams to avoid over-committing to offensive rebounds and instead prioritizing sound transition defense, limiting easy points for opposing teams in transition. Conversely, offenses with fast and aggressive play styles benefit favorable as is indicated by the positive correlation between offensive turnovers, a direct effect of this style of play, and wins.