Analysis of College Basketball Statistics and NBA Success Prediction

John Ajit, Nathan Dorsey, Ben Kim

Abstract

This report aims to analyze college basketball statistics to predict success in the NBA, in order to determine whether a player is a success, a "bust", or somewhere in between. We developed a performance metric based on points, rebounds, assists, steals, blocks, and turnovers, a metric which we've called Player Performance Rating (PPR) and compared it with Win Share statistics for various college basketball players. We then applied the same metric to those same players, but instead evaluated their NBA performance. The findings are expected to offer insights into the predictive value of college performance for NBA success.

Introduction

The world of professional basketball is highly competitive, with college basketball serving as a key talent pool for the NBA. Understanding the correlation between a player's college performance and their subsequent NBA success is vital for scouts and teams. This study introduces a new metric based on college stats which attempts to evaluate this correlation.

Methods

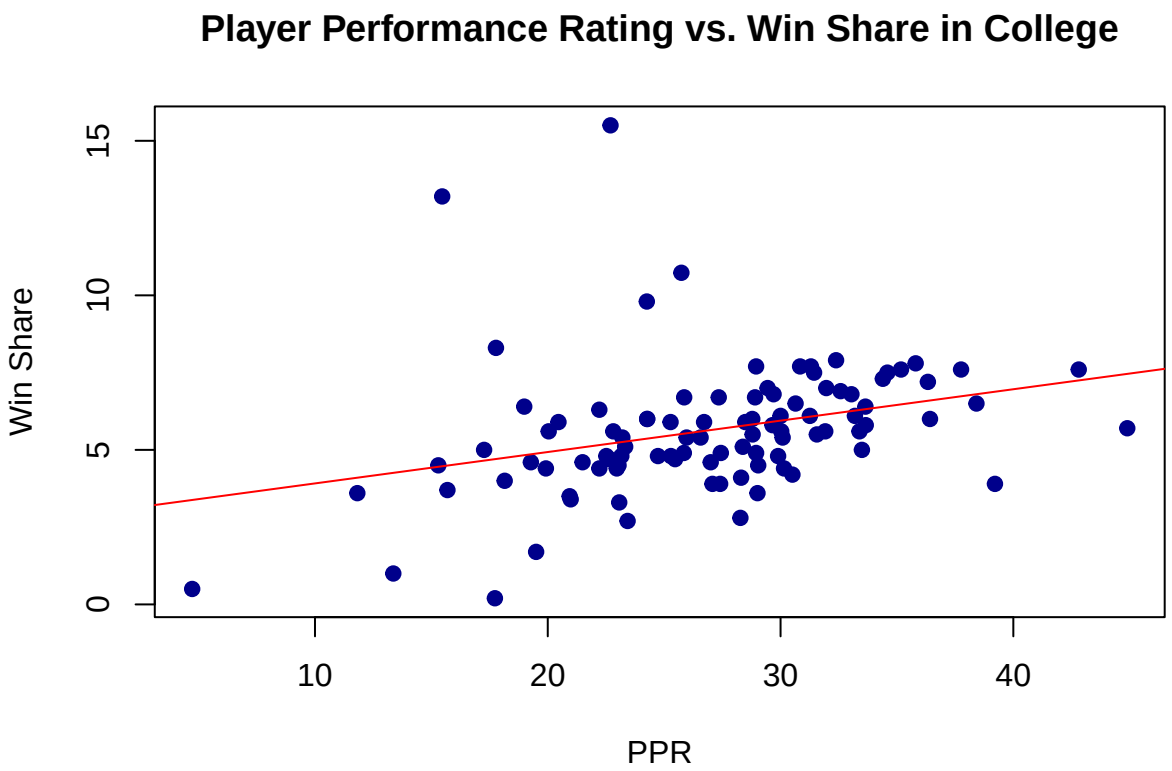
The analysis involved collecting and processing data from two primary sources: college basketball player statistics and NBA player statistics. Data was cleaned, transformed, and merged to facilitate a comparative analysis. The PPR metric was calculated for each player, and a linear regression model was used to evaluate the relationship between PPR and Win Share. We then compared a smaller sample size of a more specific group of players with a performance rating between a certain range in order to further explore the correlation.

Results

The analysis focused on comparing the Player Performance Rating (PPR) with Win Share (WS) for college basketball players and examining how these metrics correlate with their performance in the NBA.

College Basketball Players' Performance

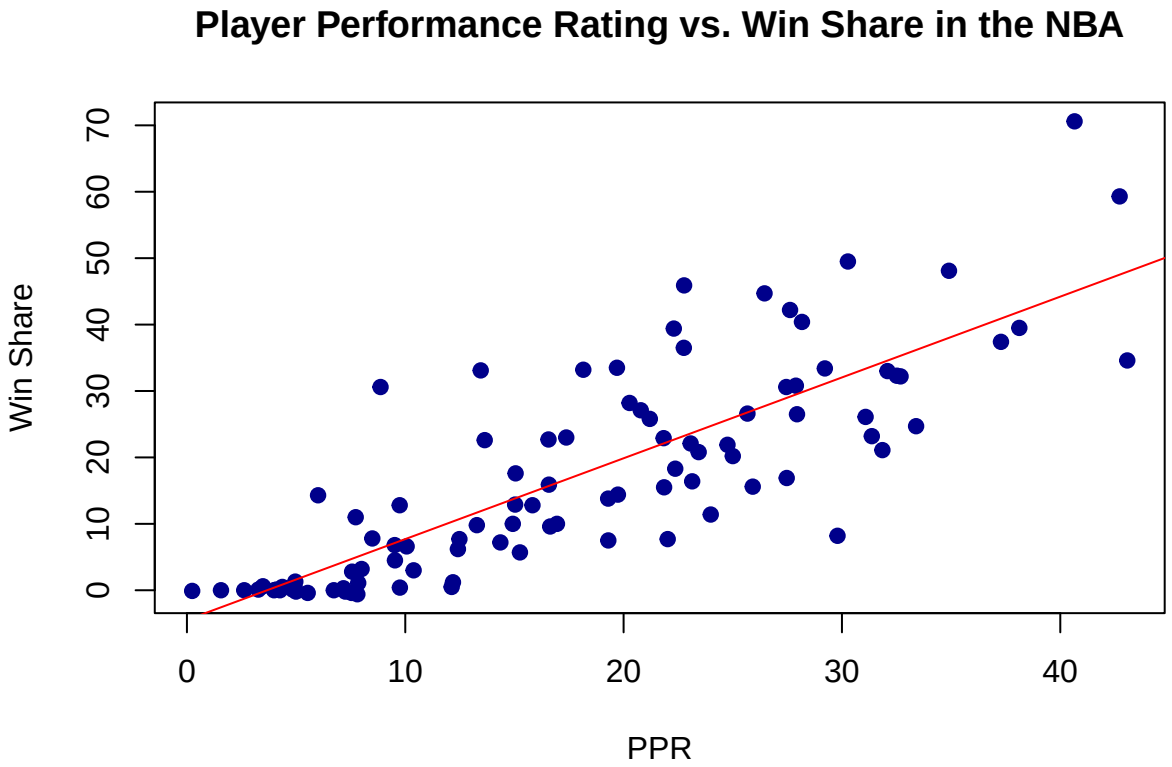
The scatter plot below shows the relationship between PER and WS for college basketball players. A linear regression line (in red) indicates the trend.



Unexpectedly, we found relatively low correlation between PPR and Win Share among college players. We believe there are a couple reasons for this: firstly, win share is a cumulative metric, meaning it is likely to increase as a player plays more games. Players simply have less games in college to allow them to raise this metric. Secondly, college basketball teams are often built around systems and can tend to involve more "role player" types, taking an emphasis off of individual player statistics as it relates to overall team performance.

NBA Players' Performance

This scatterplot below shows the same metrics plotted against each other with the same linear regression line, but for the same players' NBA stats.



The NBA, in contrast to collegiate basketball, has a significantly stronger correlation between our performance metric and Win Share, for the same reasons that it was weaker in the college data; that is, that there are more games to be played, and individual performance can have a larger impact on a team's overall performance, due to the level of skill being higher as well as the style of play being vastly different from each other at the two levels of play.

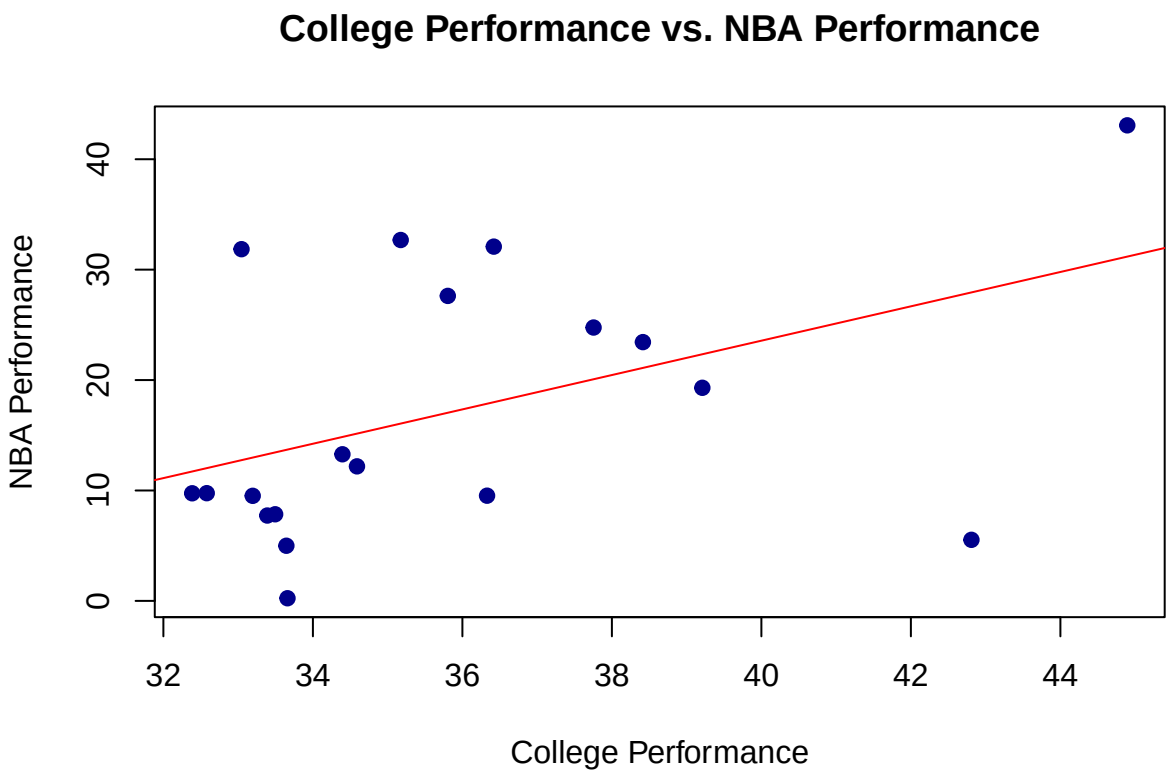
Case Study and Further Exploration

Although we found that there was little correlation between our performance metric and success at the collegiate level, we wanted to explore if we could find any determining factors which might indicate correlation.

First, let's look at an example. The top performer among both the collegiate and NBA data in terms of PPR is Trae Young. He boasts a collegiate PPR of 44.9 and an NBA PPR 43.1. His collegiate win share, however, is exceptionally low compared to his performance rating, at just 5.7, because he only spent a single year playing college basketball before being drafted to the NBA, where he has earned an overall win share of 34.6.

Similarly, many other all-stars with exceptionally high performance ratings and win shares in the NBA have low win shares in college, mostly due to their short time spent playing college basketball. Such other examples include Joel Embiid, Karl Anthony-Towns, Donovan Mitchell, and others.

Below is a graph of the performance rating of the top performers from college against their NBA performance ratings. As we can see, when we isolate many of the top performers, we can see that, although many of these players still have a relatively low collegiate win share, which is again due to the lack of time spent in college. However, we can see that, among top performers, in particular with performance ratings above around 30, there does exist a much stronger correlation.



Looking at this much smaller sample group, predicting NBA successes is somewhat consistent, as long as one considers the larger picture. However, it may be noted that there still exist some busts, players with high collegiate performance ratings and low NBA performance ratings. Take Oakland alum Kay Felder, for example. He's the dot on the far lower-right corner of the graph, performing in college with a rating of 42.8, but only managing to hold a 5.5 rating in the NBA. This, and other cases like his, are due to factors such as competition (or lack thereof) in college, physical measures like height, role/team fit, and skill set fit. This accounts for many college students to excel at the college level, but struggle at the next level. In turn, this makes predicting busts significantly more difficult, seeing as it is easier to maintain a respectable collegiate performance rating that might not transfer over to success in the NBA.

Discussion

We found that, while in some cases, college performance statistics can give good insight into how a player might perform in the NBA, it is vital when evaluating college talent pools for potential NBA prospects to take into account more variables outside of the performance statistics, such as immesurables like a player's individual situation, their competition, or other unseen factors.

However, we found that success in the NBA is pretty strongly determined by and correlated to individual performance, and can be much more reliably predicted by our performance rating, as many of the factors that would throw off our predictions of success based on college statistics aren't there.

Context is important when evaluating a player's value, and when player performance is placed within a larger context, it can help provide even more value in predicting their success.