

DSC 423 - Winter 2021

Sports/Video Game Group: El Dorado

Ken Thomas
Chunnan Liu

Wei Tong Su (James Su)
Devin Carroll

Zhong Xie

Context: In the NBA, the statistics of each game are a large reflection of individual player performance. Of all the statistics in basketball, none is more prestigious than points. Few things can validate a player more than scoring many points because it is the end objective of any offensive possession. Elite players are often defined and measured by the number of points they score. Therefore, we used points as a response variable to indicate the strengths and weaknesses of players.

The specific problem: Despite being a team sport, basketball is a sport built around individual player performance. Each player has a role to play - whether defensive, offensive, or both - and the statistics surrounding these roles have in many ways become as valuable as the players themselves. There is a lot of money on the line for all stakeholders in the NBA and front offices have tough decisions to make that can result in millions of dollars in gains or losses. How do you organize a roster in such a way that player's strengths and weaknesses complement one another? Who do you trade for whom? How do you utilize your draft picks? How much money is player x worth vs. player y? These are all questions driven by player statistics. This project specifically examines the ways in which player statistics affect points scored but throughout our analysis we will examine many other interesting findings.

The pitch: For this project, we will investigate the correlation associated with offensive and defensive stats contributing to total points in a season through regression analysis. We want to develop a model that can easily predict a player's total performance based on their offensive and defensive contributions.

The data: The NBA is 74 years old, so it has amassed a huge collection of statistics surrounding player performance. However, the game at the beginning was different from now. For starters, the three pointer was first introduced in the late 70's, early 80's. In the 2002-2003 season, the NBA allowed teams to play zone defense, significantly altering how the game is played. As a result of these dramatic shifts in how the game is played, we decided to narrow our dataset to include games played after the 2001-2002 season. We are also currently considering restricting the data to include only those who meet a certain playing time threshold. We also removed statistics accrued during playoff season because they do not apply to everyone, every season. We used data from the website Kaggle that was scraped from the website RealGM using Python, Pandas packages. The second dataset we used "NBA Players stats since 1950", was also obtained from Kaggle. Each of these data scraping efforts was collected into separate files that we merged to create our complete dataset. Our current dataset contains 9171 observations and 59 features.

Goals It is our intention to divide our regression analysis up by each player's position, creating a separate model for each position group. Using the combination of these two datasets split in this way, we hope to answer these questions: What factors most influence a player's scoring output for a season? How do those factors differ between positions? By answering these questions, we hope to develop a model that is a

simple tool for NBA front offices and coaches to use as a preliminary scouting and modeling tool. Part of a coach's job is to direct a player's training effort to improve the overall quality of their performance. By identifying which factors most heavily impact a player's scoring, it is our goal to provide a straightforward list of areas that players could focus on to improve their scoring output for a season. It is also our goal that these models provide a way for Coaches or General Managers to simulate various playing time and play calling distributions to find the most productive distribution for their team. The model could also be used by General Managers in the NBA to identify players who are significantly under or over performing according to the model. Once identified, they could make decisions about whether to acquire or trade away those players based on other models capturing change in player performance or their intuition on whether that trend will continue. Similarly, the model could provide a guideline for identifying which player attributes to target based on their predictive power in scoring output.

Ken - The position of point guard primarily resides at the perimeter of play. It is through the point guard that most offense is initiated so having a clear vision of the court is essential. The point guard will generally have more control time over the ball than any other player on the court, so ball handling skills are essential. As well, one of the main goals of the point guard is to get the ball into the right players hands, at the right moment, so assists become an important statistic for point guards. Because ball handling and assists are often prioritized over scoring, point guards tend to have a smaller, but athletic stature compared to other players on the team. However, in modern play many point guards have adjusted their play to become more of a shooting threat. Ken will focus his analysis on the role of the point guard and specifically those attributes that highlight a point guards' effectiveness on the court. Special attention will be paid to input variables that naturally result from a player who handles the ball more and resides at the perimeter of play; for example, assists, turnovers and three-point attempts.

Chunnan - The shooting guard is a fixed position in the basketball lineup. Generally, players in this position have extremely high requirements for body shape, physical fitness, and psychological quality, and have advantages in speed and explosive power. They are often held by the players with the strongest scoring ability in the team, and their style of play is very comprehensive. A shooting guard often has two things to do. The first is to have a good gap to shoot outside, so his outside shooting and stability must be good. The second is to find a gap in a small gap to shoot outside, so he must shoot quickly. Michael Jordan is considered the best shooting guard in NBA history because he clarified the function of this position. Other notable shooting guards include Kobe Bryant, Dwyane Wade, Jerry West, Tracy McGrady, Vince Carter, Allen Iverson, and so on. Chunnan wants to find out clearly what factors can define a good shooting guard. Is shooting rate a criterion for judging a shooting guard? Do shooting guards who are good at three-pointers make more points than those who are good at two-pointers? Chunnan will attempt to answer those questions by developing a multiple regression model using a subset of shooting guard data.

Devin - As NBA rules have changed throughout the years, speed and creating space on the exterior have replaced strength and skill under the basket as the primary avenue for scoring. The small forward could be considered the median position, making it one of the most versatile positions on the court. That versatility has helped Small Forward to grow into one of the most important positions on the court. Many of the all-time great players have spent time at the Small Forward position: LeBron James, Scottie Pippen,

Kobe Bryant, Michael Jordan, Kevin Durant, to name a few. By focusing a model on this position, Devin Carroll will explore what factors contribute to make a good “swingman” or “wing” player. Is it important to assist other players opening driving and shoot lanes to make scoring easier? Or is it more important for a Small Forward to grab defensive rebounds, pushing the ball down court quickly for easy transition points? How important is it for Small Forwards to draw fouls and get to the free throw line? Maybe there is a specific ratio of assists to field goal attempts that is key in elevating point totals. Devin will attempt to answer these questions by developing a multiple regression model for point totals on the subset of Small Forwards and comparing results with the models for other position groups.

Zhong - Power forward is traditionally categorized as an offensive position that focuses on scoring under the basket or position themselves within the two-point zone. However, as the game progressively strategizes beyond the three-point line, we are experiencing a trend of power forwards comfortably shooting three pointers and driving toward the paint for an easy layup. In the modern era, players like Anthony Davis, Kevin Durant, Giannis Antetokounmpo, and Kristaps Porzingis were labeled as borderline “unguardable” for their ability to shoot three when switching off from a screen or drive to the paint when their opponent is providing help defense. Even before we have Dirk Nowitzki, the pioneer of the strategy, we later saw other stars in that era adapt the playstyle; players like Tim Duncan, Pau Gasol, and LaMarcus Aldridge to name a few. By focusing on modeling the power forward position, Zhong Xie will explore what factors contribute to make a good power forward. Does having high three-point attempts/percentages contribute to an increase in points per season? Was defense being neglected as the power forward position moved further to the three-point line? How their stats differ from traditional power forward? These are questions that Zhong will attempt to answer by developing a multiple regression model using a subset of power forward data.

James - When watching the NBA, we find that top centers are quite different than before. The top centers today need to be able to shoot the three pointer. If the center only has skills in paint, they would be eliminated, such as Hasheem Thabeet. He is 7 feet 3 inches tall and was drafted second in the first round by the Memphis Grizzlies in 2009. However, he was eliminated from the NBA and played basketball in a lot of countries. If he had played in earlier times, he would still be in the NBA and maybe dominate the paint. In the past, they used a tactic called “Hack-a-Shaq” which was created to restrict the former top center Shaquille O’Neal’s performance. He was not good at shooting, and during those times, the Center did not need to have great shooting skills. However, these days we will not see those kinds of tactics employed, because they do not play only in the paint and top centers need to have good shooting skills outside the paint, such as Nikola Jokic. He plays very differently from historically famous centers. He has good shooting skills and good vision to pass and create assists. He is also particularly good at grabbing rebounds. If the other team wants to foul him, they pay the price. Wei Tong will work to find out the impact of the center being playing more outside the paint. He will also look to examine the relationship between assist and points, since top centers like Nikola Jokic are good at both.

Conclusion. Scoring is the prime consideration when evaluating NBA players. This study will gauge a number of predictors of scoring totals for a player’s season for each primary position in basketball: point guard, shooting guard, small forward, power forward, and center. We will assess which factors and metrics

are the best predictors of scoring in the NBA for each position to gain insights as to what makes a player good at each position. We will also attempt to put numbers to the differences between the positions by comparing models. The models will be trained using a dataset that was created by combining two datasets created by others and found on Kaggle.com. The models will provide a preliminary framework for assessing player performance and predicting future performance. The model has potential to integrate into more complex models for simulating a team's performance over a season and guide coaching decisions on play time and play call distributions.