**EDA of Shooting in Basketball**

Xhoi Shyti

Bellevue University

DSC 530 Exploratory Data Analysis

The evolution of shooting in the NBA is a journey that mirrors the sport’s growth from its early days to the dynamic and high-scoring spectacle we watch on TV. From the cautious set-shots in the early days, to the breathtaking displays of long-range marksmanship, the art of shooting has undergone a remarkable transformation that reflects advancements in technique, athleticism, and strategic prowess. For better or worse (depending on who you ask), the three-pointer has revolutionized strategies, player roles, and the overall dynamics of the sport.

The purpose of this research analysis is to quantify the relationship between a team’s shooting success and their chances of winning. If we consider the target to be the probability of winning, we can then use shooting metrics around 3-pointers, 2-pointers, and free throws as predictors to evaluate this probability. We can quantify this by creating a logistic regression model. The target would be winning vs losing, and the predictors would be the various shooting metrics. Once we have created this model, we could then evaluate its accuracy by comparing its predictions against actual data. The coefficients of the predictors would emphasize which shots increase chances of winning over other kinds of shots.

The results of this study would be most useful to owners, managers, and coaching staff. If the model proves to be accurate, then they could gain insight as to what kinds of strategies to employ in-game, what drills to focus on during practice, or what kinds of players to acquire.

First interesting thing to note was that the none of the 3-point metrics where the most influential after reviewing summary. Since 3-points shots are the most rewarding, it made sense that perhaps those would have the most effect. Not the case, however, since it appears that free throw success rate has the largest coefficient. In terms of the p-value, it is not surprising that the field goal percentage scored zero (most likely near zero).

Creating a regression model with the given predictors yielded an accuracy of around 72%. This means that the model was able to predict correctly whether the team would go on to win/lose the game more often than incorrectly. What this means is that it’s a combination of 3-point, 2-point, and free throw success rate that best predicts whether the team will go on to win the game. Field goal success rate is important, but 3-point success rate is only marginally more important than 2-point success rate–i.e. it is more important to make your shots in general, wherever they are shot from.

From a coaching perspective, this might not be enough to warrant any drastic strategy changes any time soon, but it does reinforce what we already know which is that good shooting scores more points for the team and that ultimately puts them in the best possible position to win.

Although steps were taken to remove games that had missing data, no other steps were taken to identify outliers. The thought process was that since there were so many records, a few outliers would be insignificant (the final dataset contained over 87000 records).

Also, this model assumes that the observations are independent of each other, and doesn’t take into account other external factors like location (home vs away) or game start time/end time.

This was a fantastic exercise that required implementing skills at various levels in data science. Half of the work was just gathering and preparing the data. Creating the model was not the most difficult (thanks to the intuitiveness of some of the python libraries), and interpreting the results felt more natural after practicing in previous weeks. Although I thought this analysis of NBA shooting would yield more exciting results, it is usually never cut and dry, and there are always other angles to consider or better questions to ask. I enjoyed working on sports data, and because there is just so much of it, I am confident there are other patterns waiting to emerge.