Problem Statement

Being such a popular sport, basketball attracts a great deal of attention from players and coaches worldwide. With so much competition in basketball within club, high school, college, and pro teams, players should always be finding ways to edge out their competitors, whether this means winning against opposing teams or beating out teammates in playing time. Because skill and athleticism play a huge role in determining the outcome of a game of basketball, basketball decision-making, often dubbed basketball IQ, can often be overlooked. A major aspect of this decision-making includes shot selection, or knowing when and how to take a shot in a game. As a result, placing a greater focus on shot selection can give coaches and players alike an advantage when facing whatever competitors they have within basketball.

Dataset

"NBA Shot Logs", retrieved from https://www.kaggle.com/dansbecker/nba-shot-logs, is a 128k by 21 dataset on Kaggle that gives information about the context of 128 thousand shots taken in the NBA. This includes the player who took the shot, the amount of dribbles before the shot, touch time before the shot, shot distance, whether it was a 2 or 3, closest defender and his distance to the shooter, and whether the shot was made or not. This dataset contains 14 numeric columns and 7 non-numeric columns, although some of the numeric columns may be better used as a categorical column. It also has the luxury of 0 missing values within the whole dataset.

To judge shot selection we would be best off by trying to predict the shot value, or how much a shot is worth in the long run. At first glance, it may seem like the column stating

whether the shot was made or not would be a useful column to predict. However, it would not completely encapsulate a shot efficiency due to the fact that some shots are worth more than others. A 3 point shot with a lower percentage could still be worth more points in the long run than a 2 point shot with a higher percentage. Therefore, we would have to look at both in conjunction to best determine a shot value, given by the column with resulting points of the shot taken.

Use Case

By inputting in the different context of a shot, such as shot distance, closest defender distance, and touch time and amount of dribbles before the shot, we can calculate a shot value. This information can be used by by coaches deciding on what specific plays to run to maximize the optimal shots being taken. A coach can record all the shots by his players in a game and see the most frequent context of a shot. A specific shot coaches can be looking for are those with a low touch time and 0 dribbles, indicating a catch and shoot, a distance of 24 feet, which is just outside the 3 point line, and a closest defender of 10 feet, indicating a fairly open shot with a high shot value. Then the coach can adjust his playbook to incorporate more plays that result in high value shots.