A Machine Learning / Multiple Regression Model was fit using the data collected from the NBA website. This data included seasonal results for each of the 30 NBA teams over a span covering 2009-2018 seasons (starting year).

The goal of the Model was to predict the number of wins for each team. Due to a strike shortened season, the expected wins per 82 games was calculated as simply Win % multiplied by 82.

All predictor variables were based out of X out of 100 Possessions to provide a more accurate indicator of team performance --- not dependent on season length

Predictors: Field Goals and Free Throws Made per 100 Possessions. Offensive and Defensive Rebounds Per 100 Possessions. These variables were evaluated for both the team and the opposition resulting in 8 variables.

The Model:

The number of expected NBA Wins = -53.047 + 5.6866(FGM) + 2.6193(FTM) – 0.7921(OREB) + 1.0887(DREB) – 4.4802(OPP\_FGM) – 2.6903(OPP\_FTM) + .7901(OPP\_OREB) + .3555(OPP\_DREB).

This equation starts with a baseline of -53 Wins, an obviously unrealistic number! But each increase in FGM percentage adds 5.6866 expected wins. Each increase in Opposition FGM % reduces expected wins by 4.4802. As you can see the equation simply adds each variable’s contribution to the model. The higher your FGM and FTM % is compared to your opponents, the more you win! Makes sense. Interestingly, the OREB is negatively related to win. Note you can’t offensively rebound if you make the basket, so it could reflect field goal percentage.

Significance and evaluation:

Current teams’ FGM, FTM, OREB, DREB results were all significant and important contributors toward predicting total Wins.

Opponent teams’ FGM and FTM were also significantly related to total Wins.

Opponent teams’ OREB and DREB results had no relationship to total Wins, and could be excluded from a future NBA prediction model.

Prediction accuracy:

When applied to the Test data, the Training Model equation (shown above) was able to capture 78.8% of the variance in the Test season Wins. This is a very good result and shows that these variables are consistent predictors of NBA performance.