**Machine Learning**

***The Impact of Star Power and Team Quality on NBA Attendance***

As we collected data, cleaned it up, wrangled it into shape and explored it. It was then time to perform some in-depth data analysis using machine learning. The basis of my machine learning would be creating a linear regression model that will determine the specific type of variable model to use, either it be a one or a multiple variable model. To begin machine learning we split our data into two sets to get our validation results. It split into a training dataset which is 80% of our data and a testing dataset which is 20%. Then we proceeded to create the model that will find the significance levels of the different variables in our data.

The model measured the p-value, multiple R squared value to determine the significance of our independent variables on the dependent variable team rating, alongside that we also measured the root mean square error RMSE. After testing and examining the significance levels of the independent variables, we found that the model 1 consisting of just the NRTG variable showed that it presents a significant impact on the team rating with a p value of .0004916, a three star significance code equating to 0 signifying how significant it is, and R^2 value of .4033. But when creating another model called model 2 consisting of both independent variables and viewing the results it proved that the Difference variable significance code fell below three stars with a code equaling up to 1 with no stars, a p value of .002637 showing that there is little to no significance of that variable, and a R ^2 of .4278. Alongside those results, the RMSE of model 1 measured at a 2.429439, while model 2 was 2.417545.

Furthermore after examining the results we determined the success of our regression machine learning technique is decided by finding a variable that was significant enough for us to be able to use a one variable model. Based off of the R ^2 value of both independent variables NRTG and difference, we determined the best variable to build a model with would be NRTG. NRTG proved to be a better model with a R ^2 of .4278 over the difference variables R ^2 of .003664.