**Interpretation of Results**

**Recap of the Problem:**

In my initial Project Proposal, I went over the important question, “What causes an NBA player to be injured” and consequently “What are the factors that determine the speed at which a player can recover”. Injured players can often go through months of rehabilitation, during which they may not be able to perform simple activities like shooting or dribbling. Additionally, it can take NBA players a few weeks or a few games to get back into their form pre-injury. It can be vital to a team’s success to fast track a players rehabilitation and recovery so that they can return to the court as fast as possible. Some injuries can take longer than others to heal and a few extra weeks of rehab can make a worlds difference for the players performance, rather than rushing the recovery process and risking aggravating the previous injury. The NBA as a whole is also changing. The rise of the 3 point shot and the emphasis many NBA teams put on having as many players on the court who can shoot 3’s, may just be causing unwarranted stress on some of the games greatest athletes.

In this paper I will be going over the models and the results of the analysis I have run, up to this point. We will be going over some of the changes I have made since my preliminary analysis, starting with

**Discussion of Analyses Done:**

Nearly everything an NBA player does on the court is tracked and measured these days, down to the minute. The “holy-grail” of these statistics is the “Player Efficiency Rating” or PER, it essentially sums up all of an NBA players statistics, such as Field goals made, steals, 3 points, Free throws made, blocks, rebounds, assists, and subtracts the fouls, Field goals missed, turn overs and free throws missed in a game, and combines them to one number between 1 and 45.   
The best players of all time like Michael Jordan, Lebron James and many more have seasons with PER’s between 28.0-32.8. An NBA players PER can also be utilized to calculate how effective they were on the court pre-injury and post injury.

In the table below using the Orange Data Mining program, I created a model that would use PER, as the target variable and analyze factors that were not considered in the calculation such as minutes per game, Usage % (how often the player is involved in a play), number of games missed due to injury, and their value added off the court (how many points the team goes up or down by when the player is not on the court)Text, table

Description automatically generated

In the model I ran a K-nearest neighbors model, a decision tree, a neural network, a linear regression and an AdaBoost model, let us briefly go over the results of each of them.

**K Nearest Neighbors:**

In my K nearest Neighbors model, I have an R2 accuracy of 68.4%. this may not be the best model to build off of my PER and injury relationships, the Mean square error is also a bit high at 6.9, and it implies that this model will not be as accurate

**Decision Tree:**

My decision tree model has an R2 value of 72.2%, which implies its generally accurate, but the mean square error value is still higher than expected at 6.08. There aren’t many decisions such as if player plays with an injury that was not detected that my model has considered so there is room for improvement here

**Neural Network:**

My neural network did not return the best accuracy here, with a low 55.9% accuracy and my mean square error value is very high at 9.6, it is safe to say that my neural network model might need to be altered a bit to predict the likelihood of an NBA player playing in a game after receiving an injury. I could develop it to factor minor injuries that players can bounce back from in a relatively short lifespan (below 10 games)

**Linear Regression:**

My linear regression model returned a very nice R2 accuracy of 90%, and the mean square error is still low at 2.19. I believe my linear regression is accurate because many of the variables that I am evaluating compliment PER well, I may have an issue of overfitting that I may need to consider

**ADABOOST Model:**

The ADABoost Model returned a high accuracy of 85.2% and also has a low mean square error value at 3.22. After training and boosting this model however, I received abnormally high accuracy scores so I did not proceed with it further as I was unaware of how to tweak this particular model.

Now let us go over some of the visualizations to make better sense of the significance of an NBA player missing a game due to an injury

**Key Visualizations and their Interpretation:**

In the Key influencers graph below, we plotted what factors influence an NBA player to miss games, and what statistics go up or down, after a period of rest and rehabilitation.   
We see for instance, that their PER, goes up by an average of 4.53, this may be because they were not playing for a few games before, and when they comeback, it causes their PER to rise accordingly. Rather interesting is their Turn overs tend to go down after they return from injury, as the players may not be playing as hard to avoid re-injury, and their assists, drastically rise emphasizing the importance of team basketball. Returning players also on average play 6.6 minutes fewer per game when returning from injury, and their impact on the court is generally reduced.

Chart, scatter chart, bubble chart

Description automatically generated

In the graph below, we have a matrix chart, representing the number of games missed due to injury, along with the respective position that the player being affected, is playing. We see for instance a large number of Centers and Power Forwards (typically the tallest and biggest men on the basketball floor) suffer from ankle and Knee problems, but this problem also trickles down to the smaller more explosive athletes like the point guards and shooting guards with many of their players, also suffering from various leg based injuries ranging from ankles and knee problems, all the way to the hamstrings. We also see that NBA players miss the most number of games (over 300 each) just from Ankle and Knee problems. Chart

Description automatically generated

Is this because of the changing nature of the game? A quick look below at the list of players who have missed the most games, show that nearly all of these players play Point guard, or Shooting guard. The lone center on this list is 6’1 former point guard, Patty Mills. I mentioned earlier that a vast majority of NBA teams are emphasizing the importance of the 3 point shot and playing “Small ball” what this may have caused is an increase in the workload for explosive guards that predominantly play Point guard or shooting guard (lest the rare exceptions in forwards like Lebron and Giannis) and it causes these players to put further stress on their bodies than before.

Table

Description automatically generated

This causes an NBA player to get injured more frequently than before. We can see from the chart below that the number of Centers (a position that has historically been the most severe in terms of injury and games missed) has drastically reduced, and the highest number of injuries for players in the NBA come from the guard positions and forward positions. This does not mean that the Center position is losing its importance, contrary to popular belief, it is thriving thanks to the evolution of the game, and talented Centers embracing the 3 point shot and guard like plays, with 7 footers taking step back 3’s and driving into the paint for athletic dunks.

Chart, line chart

Description automatically generated

I think it means that with the current landscape of the NBA emphasizing the offensive side of the game more than the defensive side, and many of the leagues best talents are guards, we are bound to see more injuries for the guard players, what needs to be addressed is the dispraportionate amount of injuries affecting guards, and we may need to tweak NBA teams playbooks to utilize guards sparingly to avoid risking major injuries that can take months to recover from. We need to understand that the players are the NBA’s biggest treasure and they must be protected at all costs. Could more rest between games cause fewer injuries, absolutely but that is a difficult conversation to have with the NBA with literal billions of dollars lost when the Covid-19 pandemic hit, and many of the NBA’s biggest games had to be cancelled and moved into a bubble arena. Things are certainly getting back to normal, but I would recommend teams to play more team ball, and rely on their guards to make plays in ways that will not harm their body. The NBA has taken stricter action against players that accidentally injured another player during a basketball game, and it has caused the league and its players to respect their opponents safety much more than in the past.   
Overall I believe that the results I have gained from my preliminary analysis and post preliminary analysis have helped me get closer to my answer of what causes an NBA player to be injured, and if I tweak my data and results further, I should be able to map out exactly which specific injuries affecting the players are the NBA’s most important issues to tackle, such as Knee problems Ankle problems.