NBA Maximum Contract Final Report

**Problem Statement**

NBA teams are paying players more now than ever before. The best contract an NBA player can receive is a maximum contract. Players could qualify for this contract by having 7+ years of experience in the NBA. Teams usually offer the Maximum to high performing players with experience. The only ways to qualify without being in the NBA for 7+ years of experience is to become an All-NBA player, NBA MVP, or NBA defensive player of the year. The problem is that there is no way to properly quantify how one an NBA player could increase his chances of obtaining a Maximum Contract. By using data from basketballreference.com, I created a model that can predict whether an NBA team will offer a player a Maximum contract. The main person we will be focusing on is Devin Booker, who is eligible for the max contract.

**Data Wrangling**

The raw data included seasons player injured seasons. This created problems with the data due to the unique object. I found every injury season and dropped them. Every player was in the same league, so I decided to also drop the Lg column.

**Exploratory Data Analysis**

You need to understand your data to properly analyze it. The methods I use to get a feel for the data is the correlation(.corr) and heatmap function. The correlation function gives me a numerical representation of the data, while the heatmap is a more visual aid. From there you select data points that have similarities in the way they behave.

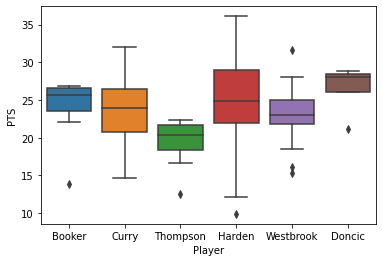
An interesting correlation was between “PTS” and “3PA”. I first used a scatterplot to see if the data had an affinity for one another. Since the scatter show linearity, I used a regplot to show the average slope of the points to see how the data clustered towards the average. From that you can find that as 3PA increases, PTS increase as well.

Chart, scatter chart

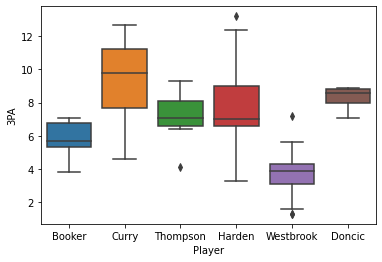
Description automatically generated

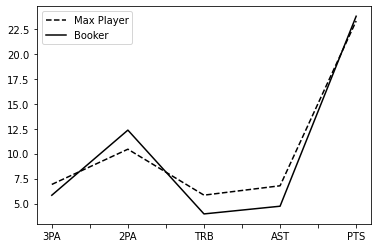
Now that we have found data points that influence each other, it is time to dive into Devin Booker’s stats.

We took Booker’s PTS and compared them with max contract players. Booker had many seasons of 22 or more PTS with many being about 26 PTS. Most max contract players have seasons clustering around the 20-25 PTS range, which bodes well for Booker. Booker has not had many seasons under his belt compared to other max players, so his future is very promising.



We then too Booker’s 3PA and compared them with max contract players. Booker had on a little below 6 3PA per game seasons, while max contract players hovered in the 6-10 3PA. The 3PA was below most max contract players and could show an area Booker could improve in. From this we can conclude that booker should be taking more 3PA to improve his PTS even more.



When you look at all the data, we can see some areas where Booker is lacking compared to other max players. His REB and AST are lacking by 1.8 and 2, respectively. Other than that, Booker stats are remarkably like other max contract players. Devin Booker should receive the max contract due to his stats being closely correlated to current max contract players. 

**Model Selection**

**Random Forest**

We tested the parameters around who was offered a maximum contract and who was not offered a maximum contract. The mindset of the NBA owners is more recall, so that is the method I focused on. The random forest classifier was the first model I built to test how well the parameters predicted a player worthy of a maximum contract. The Random Forest model had an average Recall score of 0.89. Its Precision and F1 scores were also similarly high. We may have to consider overfitting since all scores are close to or at 0.9.

**Decision Tree**

The Decision Tree model had an average Recall score of 0.69. It had a surprisingly low recall score when predicting NBA teams offering the max contract. That means the model tends to offer a maximum contract often.

**Gradient Boosting**

The Gradient Boosting model had an average Recall score of 0.84. This model had the most hyperparameter tuning. There was a sweet spot for number of estimators, random states, and max depth. The average score went from 0.66 all the way up to 0.84. After the adjustments, all scores were in the low 0.80s and performed the most consistently amongst models.

**Conclusion**

Gradient Boosting model has the most consistent results without looking overfitted like the Random Forest model. The Decision Tree model can be useful if you are trying to filter players early in their career for future max contracts. The Suns team should offer Devin Booker a maximum contract at the end of the season. The Gradient Boosting model is also the best of the three models in predicting Maximum contracts.

NBA maximum contracts first started in 1999 so case studies of players will grow. I used 11 players who played at the guard position. I would like to expand this model to include all players who were offered a maximum contract to increase the data size. Basic NBA stats were used, and more advanced stats could be added in the future. The problem with using advanced stats is that certain basic stats are favored and could run into multicollinearity issues.