Basketball Strategy and Research Project

Robert Eisner IV

Project 1: Player Evaluation Option 1

A catch and shoot hard nosed defender at the SF/PF position is scarce in the NBA. Far too often the players with above average defensive win shares, plus defensive statistics, and defensive intangibles are lacking in terms of shooting. When a player exists that meets these requirements its often the case that they are a young player, about to hit their second contract, there is an option or available team QO, or they are already compensated more than what is available under the exception. That alone removed a great majority of the potential free agents. With all of that considered, in addition to the positional and statistical requirements, the ideal player to fulfil this role is Nicolas Batum.

Of all the available hybrid forwards Batum has the third highest catch and shoot field goals made per 36 minutes. Of these opportunities roughly 90% of the shots are taken from the three. Batum also shoots 42% on C&S opportunities and 39.8% on 3P opportunities. In terms of percentage those are both within the top 5 for hybrid players I researched with more than 3 C&S opportunities a game. Batum is not a prolific rebounder, however he boasts a slightly above average defensive Win share along with an average Defensive rating of 109.8. When in comparison to NBA forwards on the whole, his defensive numbers are not anything to write home about. However when in direct comparison to other players on the free agent market, his average numbers become more above average.

As an over the cap team the current Mid-Level Non-taxpayer exception of roughly $10,000,000 with increases of roughly 5% each year. Nicolas Batum is 33 years old and currently making $3,170,029 per year with the Clippers with a player option for next season. With a player option set at this value he must have considered a couple scenarios. The first being he is older, and it is insurance if an injury were to occur this season, and in line with that, if he performed well on his contract he would be able to make more in the coming season. He is performing well relative to his contract and he will very likely turn down the player option opting to become a free agent. He has an opportunity to make roughly an additional $7 million on top of being able to compete for an NBA championship in a winning environment he has rarely experienced. Further with younger hybrid forwards entering the market in addition to a wide range of forwards with varying specialties he will likely not receive an offer greater than what the Suns are capable of offering. Looking at contracts given to aging players in recent years an offer of the full exception would not be necessary.

Project 2: Win Projection and Trade Recommendation

Part I:

I approached this project with the idea that I may need any and all data available to come up with the best information possible. I used basketball-reference to create csv files for every season from 2015 till now. These excel files had 56 columns of data pertaining to virtually all offensive statistics, defensive statistics, and an assortment of advanced statistics. With these built as individual seasonal

csv's I went through and pruned data from the data. As I brough the data over separately as pages I removed overlap between the offensive and defensive metrics prior to combining, and with the advanced statistics I removed homecourt information and margin of victory from the testing data as well. There were some other non-statistically relevant information removed as well. Once I felt confident with my data I created training data which included all the statistical information from the beginning of the 2015 season till the end of the 2021 season. I then created test data which is based upon the statistical information of the 2022 NBA season. I could have used more past seasons data to help with the analysis further.

I then decided that I wanted to find the correlation between the individual statistics and how those directly related to recording a win. My correlation matrix was able to show me the statistics that most strongly correlated to victories, and showed that statistics pertaining to efficiency (i.e. field goal percentages) were the most impactful on a victory. It also showed that poor defensive numbers correlated to recording a loss and that many of the individual statistics had varying degrees of significance when predicting whether a win has or has not occurred. I then used this correlation data to plot the most significant statistical correlations which I go into more as part of the visualization aspect. I then wanted to check varying regression tools that I have used in past projects to get a better handle on the data, I'm not sure how useful this but I found it useful to see the means and standard deviations produced by these different models.

At the end I used three models to predict wins using the 2015-2021 data as training data and the data from the 2021-2022 season for testing purposes. I first used an MPLClassifier model which ended up being too close to actual estimates, as such I believe I may have done something wrong to produce incorrect results. I also used a decision tree regression to predict wins. This resulted in skewing towards higher total wins for the best teams. And finally I used a KNN model. This produced win productions much more in the middle and I think when comparing the three different models and the differences in win total, it is useful to consider the three of them as a whole. I checked for error by using 2021 as the test data and compared victories for the 2021 season against the predicted win totals.

Part II:

I will answer this question in a few ways. If the pick from the Magic was an unprotected pick for 2022, my three predictions all produced results between 20 and 23 wins. This would result in a high lottery pick very likely in the top 5 and it would be the choice every time. However that goes in line with, if it is a lottery protected 2022 Magic first (with negligible transformation) against unprotected picks from the Raptors and the Timberwolves, then it comes down to necessity of the current asset. In all three of my predictions the Timberwolves are predicted to win as many or fewer games than the Raptors. I am also of the belief that the Timberwolves are in a more difficult conference on the whole and have worse leadership and coaching. I also have less confidence in their star players (Though Edwards is changing that) and am pessimistic towards their experience. As such I would take the Timberwolves pick over the Raptors. If the Magic pick were protected for 2023 as well I would like to take the known quantifiable asset now to assure an asset I can definitively use to make my championship quality team better right now.

Visualization:

The visualization I chose to make are pair-plots of the correlation between wins and the other statistics used to determine whether a win has occurred. It tested for this mainly because I wanted to see just what did in fact directly correlate best to a victory. So I was able to find the correlation matrix using pandas. From this correlation information I wanted to find the values that produced the highest association to winning games. As expected the most important factors/statistics all were associated with efficiencies. FG%, 3P%, 2P%, ORtg, DRtg, NRtg, TS%, and then correlating largest to a loss were DefFG%, Def3P%, Def2P%, where a higher value here brought down the opportunity to win. With all other statistical measures playing a role in the win, however there was not as high of a correlation between victories and the other statistical measures. However of note to me were that blocks had a higher correlation to victory than assists and was on par with total rebounds.