Springboard Capstone 1

Goal: Explore NBA advanced metrics from the perspective of an NBA General Manager, by building models that explain the connections between metrics such as team total and wins over the course a season.
Determine if raising the team total in some of these advanced statistics is correlated to increased wins over the course of the season. If possible also explore the relationship of salary to these advanced metrics. Essentially the goal is to answer the question “If I am a an NBA general manager how am I most likely (based on advanced metrics) to improve my win total in the coming season.”

Data: All data are available via basketballreference.com (hereafter referred to as BBR). BBR provides a wealth of NBA data. While advanced metrics could be generated from the base statistics provided, BBR also provides many advanced NBA metrics so this step should, most likely be unnecessary. Some of the better known advanced metrics that should be worth exploring will include Box Plus MInus (BPM), VORP (Value OVER Replacement Player), Wins Over Replacement Player (WORP), and Player Efficiency Rating (PER).

Methodology: The statistics will be looked at in a multi-dimensional space. Team metrics provide the rows of data and each of our statistical variables will provide the columns. It is my initial belief that because many of these advanced metrics incorporate the same base statistics, “double counting” (which might translate to collinearity) may actually lead to a less useful final product. Regression models will be applied to these statistics to see how they affect total team wins on a year to year basis. The coefficients should provide useful data as to the relationship between each statistic.
It may also be useful to look at teams in groups based on win totals for previous years for the analysis. This might provide insight into specific examples of where these metrics might or might not be useful as a solution to the business problem of increasing win totals. For example we might find that adding VORP helps teams in the bottom two-thirds of the league, but might have a little or no effect at the top.

Potential Pitfalls: There are some obvious potential problems that may need workarounds. I propose that we start to look at the data and then address the best way to adjust for these issues. Some of the problems may include player injuries and players playing more or less minutes per game in a new team. Also the addition of rookies is going to affect team wins but can not be accounted for by a previous season statistic (college statistics do not translate to pro

statistics in a useful way). It is also possible that single season advanced metrics may be a less useful indicator than a multi-year average. The downside of a multi-year average is that we will have more data gaps to fill. For instance if we use a 3-year average, any player that has been in the NBA for less than 3 years (or injured for one of those years) will not have proper data for evaluation. Age may also have an effect on future statistics. It is possible that it might also be worthwhile to adjust predictive advanced metrics along a curve fitted for player age.

The Case of Rookies: As a potential solution to the problem of rookies who may dramatically affect team wins but can not be accounted for by previous year metrics, we could potentially use one of our advanced statistics “Wins Above Replacement Player” (WORP) to adjust team wins. The WORP statistic essentially says that a player accounts for a certain amount of wins more (or less) than an average player at that position would have. So for instance if a team increases its win total by 3 wins year-on-year, and the WORP of their rookies is +3, their win total would effectively be unchanged.

Deliverables: As required, I will submit all the Jupyter notebooks I will develop, a final report, and a presentation slide deck.