Exploring Data Analytics and Interpreting Team Wins in the NBA

Context: In fall 2018, I enrolled in a Business Analytics course for my MBA at Santa Clara University. For the final project for this class, we got tasked to analyze a dataset and present any conclusions about the predictive power of various variables. My group and I decided to analyze NBA team statistics to attempt to predict wins in the Association.

Data: We obtained a dataset which covered all major NBA statistics from 2001-2011. Our data includes possibilities of using regression-based tools to analyze information from players, plays, to team based stats. We decided to look at team stats and see if certain team statistics—field goal %, offensive rebounds given up, pace, etc.—could successfully explain winning outcomes over 10 years of games.

Method: We relied on backward and stepwise linear regression models to assess our information. Linear regression is a basic tool of predictive analysis which indicates how certain predictor variables do in determining an output variable and which variables are statistically significant. We began by looking at 16 offensive and 15 defensive team statistics separately to assess relationships with our output, wins. We ran this through a Solver platform in Excel.

Results: The initial attempts allowed us to whittle down the best offensive and defensive predictors. (See Slides 4-6). While neither attempt produced statistically significant results, only an adjusted R squared of .43 and .26 respectively, this process allowed us to infer the best stat categories as a whole. After this, we handpicked the “best” categories from the 32 variables and combined them. Here, we looked at how a combination of offensive free throws made, offensive three pointer made, offensive assists, offensive points, defensive field goal attempts, defensive offensive rebounds given up, and defensive points. As a result, we returned an adjusted R squared of .92. (See Slide 7). This indicates our ability to explain 92% of the variation in NBA wins with these seven variables.

Significance: On the face, this conclusion indicates the amazing ability to explain past outcomes through data. On its face, an adjusted R squared of .92 shows a strong relationship between these seven variables and the ability to explain win variance. Further, our team immediately realized the setting of the NBA was ripe for a team to dominate the league. For example, the Golden State Warriors won three NBA titles since our dataset concluded by dominating three of the seven variables we analyzed: 3pt made, offensive assists, and offensive points for. (See Slide 12). We felt this dataset helps explain the Warriors success from a statistical perspective.

Moreover, certain organizations are now shifting to this type of model by increasing the emphasis on 3-point shooting. However, we noted this comes with a pitfall. While offensive 3 pointers made is a statistically significant category to winning games, there is no positive relationship between winning and 3 pointers attempted. In turn, we felt teams can fall to the fallacy of simply shooting 3s for the sake of it but not realizing the key is to have good shooters, not just a strategy in place. Thus, understanding the information being analyzed becomes critical for any organization.

Next Steps: When I began this in the fall, I sought to explore statistics and the predictive abilities in sports-settings—a long time passion of mine. In conducting this, I confirmed the immense power of this information which is starting to influence how NBA organizations. Going forward, I would like to fine-tune this type of modeling to more specific problems like positions and roster management with the goal of making this type of work a full-time career upon graduation.