Executive Summary

High-Level Overview

For this assignment, our group decided to focus on the domain of basketball, specifically men’s basketball in the form of the NCAA spring tournament. Because the focus of sports is, of course, to win, and because sports betting was recently made legal in multiple states in America, it only seemed fitting to analyze data related to the aforementioned topic with a focus on prediction (Licata). Our group sourced March Madness tournament data from Data World, aggregated from the years 1985-2016, including fields like “round,” “winning_seed,” and “winning_score”.

Goal of the Study

In short, the goal of the study was to use predictive analysis to identify which, if any, variables play a role in determining (read: predicting) game outcomes - the applications of which will be discussed momentarily.

Analysis Run

We converted Region and Overtime to dummy variables, removed outliers, created two new variables (Seed Margin and Margin of Victory (MoV) variable), and identified the only significant predictor of our target variable, MoV, that being Seed Margin. We ran a linear model and discovered how much MoV changed in relation to Seed Margin. Our team tried to see what
predictors had a significant impact on margin of victory. We discovered that Region was not a significant predictor. Next, we created another linear model to see if Overtime had a significant effect on the MoV. We found that Overtime is a significant predictor of MoV. Consequently, our team determined if games don’t enter overtime at all, there is a around a 6.2 MoV.

**Potential Applications**

First and foremost, the findings from our analysis confirm the intuition that better teams beat worse teams by a larger margin. Stated alternatively, the greater the disparity in seed (read: ranking) between two teams, the greater the margin of victory. Of course, many sports spectators can sympathize with the underdog - particularly those who are loyal to said teams - but our group thinks it wise to be objective when considering two teams that are set to play. While upsets (when the underdog beats the anticipated winner) certainly do occur, our analysis indicates this is usually not the case, and the more outmatched the underdog is, the more likely it is they are going to lose...by a larger amount. What this translates to in terms of practical applications seems quite clear: betting (for money or otherwise) on the higher ranked team will likely yield winning results. Furthermore, understanding that Overtime is a significant predictor of MoV provides arbitrage gambling opportunities in regards to betting on point spreads. For example, if a game is not likely to enter Overtime and the spread is +3.5 for the underdog, a recommended bet would be to place money on the underdog. This is because the predicted MoV is around 6.2 points which falls within the spread in the example above.
Bibliography