

March Madness

Alexander Van Roijen, Rocco Bavuso, Molly Clark

March 9, 2018

This report analyzes what plays into the determination of March Madness seeding. We are taking a look at data from multiple past seasons of play, data from past tournaments, and at winners from the past to see if there is anything specific that points to the way teams are seeded within the tournament.

Introduction

Throughout this report, we will analyze past data on March Madness teams, their seasons, and their performance in the tournament. We have gathered data from kaggle.com and broken it down to find the most useful predictors of a team winning the tournament.

Data Summary

On Kaggle's website, we found data that gave us a look at the teams who qualified for the March Madness tournament. Each team is assigned a four digit code that is consistent from season to season in each data set. This allows us to easily manipulate and analyze the data.

We begin by looking at the data that has been collected season to season and follow up by looking at the data from previous tournaments and history of success to see which factors effect a team's chances of winning.

Regular Season Data

Part of our analysis comes from regular season data of 71,241 NCAA Division 1 games from 2003 to 20016. The data available are game statistics of winning and losing teams, such as field goals made, rebounds, assists, etc. For this part of the analysis we wanted to determine particular characteristics of winning teams in the regular season visible on the stat sheet. This part was used to help determine what are the best predictors of a winning team. In order to do so, we thought the ideal comparison would be winning team statistics compared to average team statistics. We computed this difference for the

following statistics: field goal, three point, and free throw percentage, rebounds, assists, turnovers, steals, personal fouls, and blocks. The first three statistics are percentages while the rest are averages per game.

Tournament Data

With the Regular Season Data, we included data that came specifically from the March Madness Tournament. This data was pulled out of tournaments from 1985 to 2016 and includes information on both winning and losing teams throughout the tournament. The stats that are available to us through this data set include attempted and made field goals, three pointers, and free throws, as well as the seeds and slots of the teams. We analyze the differing percentages of made shots between winning and losing teams to see what may make a difference in a team's ability to win more tournament games. We took a look at past seed placement with the season and tournament data to determine potential placement in coming tournaments.

Pre-Analysis

Regular Season Data

Tournament Data

Proposed Model