Can analytics predict the tourney?

You want to win your March Madness pool.

Every March, your friend sends that email about the pool for the NCAA men's basketball tournament. Winning the pool offers not only a significant pot of money but also bragging rights for a year.

However, picking a bracket is challenging. Which 12 seed might pull off the upset over a 5 seed this year? Is there a 15 seed that will terminate a 2 seed in the round of 64? (Later, I'll discuss how neither of these matter much for winning your pool.)

Unless you cover college basketball for a living, you don't have the time to know much about all 60-some college basketball teams that make the tourney. This book will show that you need analytics and a data-driven approach to filling out your bracket.

But can analytics overcome the randomness of the tourney? They call it March Madness for a reason. You may be one of the many people who doubt the utility in using analytics to fill out a bracket. Let me tell you a story about some others.
The randomness of the tourney

In 2012, Amy Nelson of SB Nation got interested in my tourney analytics and how they could help her win her pool. She flew out to San Francisco to make a documentary on my research. We met at a sports bar there, in a city on the forefront of technology and innovation.

While I was helping Amy fill out her bracket, her crew interviewed some bar patrons. Four of these clips made the beginning of the documentary. Here’s what they said.

*There’s no way you can predict sports based on mathematics. There’s always going to be an upset. There’s always going to be a surprise. There always is every year.*

*You can’t quantify which of these guys is going to get nervous and which ones are not.*

*It’s March Madness. You can never predict what’s going to happen. That’s why it’s the madness.*

*Running numbers on March Madness takes the heart out of it.*

Yes, these quotes came from the same city in which many of Google's and Facebook's engineers live.

If you focus on a narrow range of games, they have a point. In 2014, Connecticut entered the tourney as a 7 seed and long shot to win the
tourney. The Huskies needed overtime to beat St. Joseph's in the round of 64. Then, almost out of nowhere, Connecticut won the next 5 games to win the tournament, an event that no computer or sensible human would have predicted.

In addition, the tournament always features those thrilling upsets the first weekend. In 2013, 15 seed Florida Gulf Coast looked like the Harlem Globetrotters as they won two games and earned a berth in the Sweet 16. Each year, a team seeded 10th or higher makes the second weekend of the tournament.

However, randomness looks different over a longer time horizon.

How often does the higher seeded team win?

From 2002 through 2014, the NCAA tournament has featured 844 games. In 810 of those games, the opponents had different seeds. (In recent years, the tourney has featured a First Four, a set of 4 games that feature teams with the same seed. Also, teams with the same seed could potentially meet in the Final Four.) The higher seeded team has won 71.2% of those 810 tournament games.

Let's stop to appreciate the accuracy of such a simple predictor. Each year, a selection committee, which consists of athletic directors and conference commissioners, assigns a seed to the sixty plus teams that
make the tourney. Inclusion on this committee requires no coaching experience or any background in numbers and analytics. Yet over a huge sample of games, the higher seed wins more than 2 of every 3 games.

There are always lower seeded teams that defy these odds, and we all love these upsets. In the 2014 tournament, 15 seed Mercer upset 2 seed Duke, and sports bars across the nation exploded in delight as if it started raining $100 bills. It didn't matter that most picked Duke in their pool.

Even better, Kansas entered the 2010 tourney as the clear favorite, as 41.8% of brackets on ESPN picked the Jayhawks to win it all. However, Kansas lost to Northern Iowa in the Round of 32, as Ali Farokhmanesh hit a late 3 point shot to seal the victory.

You always remember the thrilling upsets, those games in which randomness takes over and slays the big favorite. However, these games are the exception. In the big picture, the favorite usually wins, which makes it worth trying to predict the outcome of the tourney.

These numbers also show the problem with the typical bracket fill advice. Let's see if you've heard this before: The 12 seed has beaten the 5 seed in past tournaments, so look for an upset pick in these games. This advice is flawed for many reasons.

First, the 5 seed has won 66.3% of games over the 12 seed. Remember, the favorite usually wins.

Second, these early round games don't matter much in your pool. Most
pools use a scoring system that awards 1 point for a correct pick in the Round of 64, 2 points for the Round of 32, and doubling values for each subsequent round. Early round games are inconsequential compared with the championship game that's worth 32 points.

The higher seeded team has won 71.2% of 810 tournament games.

The most important choice in your bracket is the champion. Since 4 teams get a 1 seed, the seeds of teams will not help you with this decision. However, analytics can shed light on this decision through win probability calculations.

How often can analytics predict the winner of the tournament?

With the growth of sports analytics, many analytics guys calculate the win probability for each team in the tourney. You might have seen the projections of Nate Silver at FiveThirtyEight or Ken Pomeroy on his college basketball analytics site kenpom.com.

Here, we'll use my own calculations at The Power Rank since we can look at how these numbers perform over a range of years. Each year, you can
find these numbers on an interactive visual on my site, thepowerrank.com.

Calculating tourney win probabilities starts with my college basketball team rankings, which come from an algorithm I developed based on my Ph.D. research at Stanford. The method takes a statistic like margin of victory and accurately adjusts for strength of schedule. Let's look under the hood at how this works.

First, this method requires margin of victory on every college basketball game before the tourney starts. This includes over 5,000 games each season.

Then the method creates a network in which teams are nodes and games are edges that connect these nodes. This was inspired by Google's PageRank algorithm, which brought order into the complex world of web search.

From this network, the method develops a set of equations to be solved. For college basketball before the 2015 tournament, there are 351 equations for 351 unknown variables, one for each team.

The method then solves for the over 300 variables at the same time, which requires much more computer power than checking how often a 12 beats a 5 seed. The solutions to these equations give the team rankings in which the rank of each team depends on the performance of every other team in the network.
Each team gets their rank based on a rating, or an expected margin of victory against an average Division I basketball team. For any tourney game, the difference in the ratings of the two teams gives an expected margin of victory.

The expected margin of victory implies a win probability for both teams. The win probability for a team to win the tournament comes from these game-by-game numbers.

How have these tourney win probabilities performed? For the 13-year period from 2002 through 2014, the team with the highest or next highest win probability by The Power Rank has won 7 of 13 tournaments. The win probabilities for these teams ranged from 32.2% for Kansas in 2008 to 13.2% for Florida in 2007.

Team-based college basketball analytics can help you make the most important decision of your bracket: which team to pick as champion. However, these win probabilities are not enough to win your pool. You need the right strategy. Let me explain.

Contrarian strategies for your pool

Suppose you fill out your bracket according to the team with the highest win probability for each game. You will do well in your pool but probably not win.
To see why, consider the 2012 tournament. Kentucky was a huge favorite by my numbers and brackets filled out on ESPN. They seemed like the easy choice as champion.

However, many people in your pool also picked Kentucky. If they win, you and many others claim the 32 points for picking the correct champion. The winner of the pool now depends on games in earlier rounds. With so many others in contention, it’s quite likely that someone gets lucky and beats you.

Instead of picking the public favorite Kentucky, you should find a value champion that has a good chance to win but gets overlooked by others. In 2012, Ohio State was this value pick, as they had an 18.8% chance to win by my numbers. Only 4.8% of brackets on ESPN had Ohio State as champion.

If Ohio State wins, you have a great chance to win the pool. You get 32 points from picking the correct champion while few others made this same choice. The solid choices of favorites in earlier rounds give you a great chance to beat these remaining people.
This strategy didn't work in 2012. Kentucky won the tournament while Ohio State fell short against Kansas in the Final Four. However, the contrarian champion has won multiples times in recent tourneys. This book explores how you can increase your odds for winning a pool based on this contrarian strategy and gives honest odds for your chance to win.

The bracket advice in this book is much like counting cards in blackjack. In this popular casino game, the odds of winning a hand depend on which cards remain in the deck. By keeping track of dealt cards, people like Jeff Ma, an analytics guru and the inspiration behind the movie 21, made money off the casinos.

Did Jeff win every hand in blackjack? No. Did he win every time the deck gave him the best chance of winning? No way.

In the same way, you will not win your pool every year with the advice in this book. While favorites usually prevail, surprise teams do win the tourney. In 3 of the past 13 years, a team with less than 5% win probability by my numbers won the tourney.

And the NCAA tournament is particularly frustrating because it only happens once a year.

However, that doesn't mean you should pick upsets at random in your bracket. Would you hit on 18 in blackjack in hopes of getting that 3? If so, this book is not for you.

I think you're smarter than that. As random as March Madness seems,
you know data and analytics have their place in making predictions. This book shows that if you apply these contrarian strategies year after year, you'll end up way ahead of your competition.

Get the book

The remainder of *How to win your NCAA tournament pool* is available on The Power Rank. It comes with my bracket advice for the tournament every year.

To check out this offer, click [here](#).