A Method to the Madness: Predicting the Winner of the NCAA Men's Basketball Tournament

An Honors Thesis (HONR 499)

by

Max Chenoweth

Thesis Advisor

Crystal Lorch

Ball State University
Muncie, Indiana

March 2021

Expected Date of Graduation

May 2021

Abstract

Data analysis is an emerging field that has recently been applied to the field of sports as well. From movies like *Moneyball* to the NFL Combine, everyone wants to reduce indicators of future success down to a few simple numbers. This idea is what led to my investigation of NCAA March Madness. Each year, thousands of Americans fill out brackets with their predictions of tournament results, but is there a methodical way to approach this? I analyze and interpret statistics from the years 2012 to 2019 in an attempt to create a single indicator as to who will win the tournament. In disregarding the results of preliminary rounds to solely predict the final champion, I look to combine many statistics gathered throughout the regular season into one final score for potential success.

Acknowledgements

I would like to thank Crystal Lorch for helping me throughout this entire process and for being so understanding of my schedule. I would not have been able to do this project without her and I am extremely grateful for all the time she has spent working with me.

I would like to thank my parents for instilling the values in me that have allowed me to get to this point in my education. Their constant support and investment into me and my education has truly shaped me into who I am today and I will not forget that.

Table of Contents

Process Analysis

I began this process by diving through the internet looking for any statistics I could find. I learned a lot about college basketball and what teams are consistently top performers. But I also got to learn a lot about how my problem solving process works. I began this project with a very different vision than where I ended up and I had to come up with a lot of solutions on the fly.

I came into this thinking I could research every individual team that made it into the tournament over the last ten years. I wanted to look at the roster changes, the coaching staff experience as a whole, the past tournament successes, everything I could think of. I had set the bar very high and quickly realized that was not going to be possible in one semester. I decided to cut down some variables and alter the project strictly to statistics that could be found easily and quickly. This was slightly disappointing for me as I truly believed I could find a way to predict the winner with foolproof accuracy. I thought that the combination of my mathematical background and college basketball knowledge would allow me to develop this formula, and I still believe this is true today. I knew that switching to a strictly statistics-based formula would most likely lead to being unable to predict the winner every time but I still hoped I would stumble upon a solution.

This process taught me a lot about how to look at and work with data as well. I spent a lot of time looking through spreadsheets to find ways to combine different values as efficiently as possible. This was fun because I already had the end result available to me. I knew who won the tournaments; all I needed to do was find the formulas that valued that team more highly. The fact that I was only basing my formulas on the past winners actually reduced a lot of bias that could have been involved if I was attempting to judge individual players and coaches.

I had also created a lot of formulas with very different outputs. The formulas did not produce the same outputs year to year or team to team. By this I mean that one year the average result of a formula may have been 1.5 and the next year it could have been 25. There was a lot of variance in my formulas because most of them were based on initial ratios that could vary a lot as well. I really enjoyed the process of trying to condense and simplify my results and I appreciated my statistical background in doing so.

March Madness has received its name for a reason and that is because there is a level of unpredictability to it, but at the end of the day that was the question I wanted to investigate. Was it really madness or were we just missing the few characteristics that we actually needed to look for? Ultimately, this process taught me that sometimes you cannot find the answer you are looking for in the way that you are going about it. I am a person that very much wants the simplest, driest answer possible and I hate not having the answer. This is why I enjoy mathematics: either you got the right answer or you did not. But this process resulted in me not finding the answer and I had to accept that. I spent a really large amount of time working on this and I enjoyed it but I ultimately had to understand that maybe what I am doing is not going to get the result I wanted.

Introduction

Every year, thousands of sports fans across the country spend an entire month engrossed in one tournament: NCAA March Madness. Sixty-four teams are selected to compete in the single most competitive environment in all of college basketball. Families gather together before the tournament even begins to watch Selection Sunday, anxiously waiting to hear their team's name called. The moment Selection Sunday ends is when the real fun begins though. The official bracket is announced and it is time to make predictions. The bracket challenge of March Madness takes the nation by storm within minutes and every spectator is picking their winners and losers, hoping to have the perfect bracket. Different websites offer different rewards but the draw of a big cash prize for having the perfect bracket is guaranteed across the board. From predicting winners round by round for small cash prizes to predicting the entire bracket perfectly before the tournament begins, there is money to be made here. The question is, how do you know which teams to pick? How do you know who has the best chance to win it all?

Every team that enters the tournament receives a seed of one through sixteen with one being the best. There are four regions, each composed of sixteen teams, so there are four teams with each seed. These seeds serve as the first indicator of potential success, but they can be deceiving. There is a selection committee of NCAA basketball conference commissioners and athletic directors that spend the year studying the top teams in the country. Then they sit down to create the list of every team that they believe should be in the tournament and what seed they should receive. There are two ways to get into the NCAA tournament, either through an ‘At Large Bid’ or by winning a conference tournament. Every conference has a tournament at the end of the year, just before March Madness, and the winner of each conference tournament automatically receives entry into March Madness. Because of this, some of the teams who may be one of the top 64 teams in the country may not actually get into the tournament. After the 32 spots of the conference tournament winners are filled, the rest of the spots are filled with the teams that the committee believes have proven themselves the most. Yet, somehow things are never as simple as they seem. There is something that happens in March that makes things a little more complicated. Whether it is the value of having experienced players and coaches, or the value of having just won the last five games played, or some other statistic that does not usually make its way to paper, the team that is considered to be 'better' does not always win.

This led to the investigation serving as a basis for this project. What is it that allows a team to go far in the tournament? Can some combination of the typical 'paper statistics' like rankings and record combine with untraditional statistics like experience or hot streaks to create an indicator as to who truly has the best chance to win March Madness? I wanted to research both the paper statistics and the untraditional statistics for every team involved in the last ten tournaments. This quickly presented a problem, though, in the sheer amount of research, data, and time this would require. Studying 64 teams, ten tournaments, and each team's strengths and weaknesses just was not an effective use of my limited time. I decided to look at the paper statistics for every team that made it out of the first round of the last eight tournaments. That gave me enough data to work with and create a valid attempt at finding a formula for success. I still wanted to look at the untraditional statistics, which I decided to do only for the winners. In the event that I could not find a pattern or formula within the paper statistics, I could at least create an idea of the untraditional statistics that all or most of the previous tournament winners had. I want to find the true indicators of success that seem to be consistent amongst the majority of past tournament winners.

Methods

I created an Excel spreadsheet with the data from each year in a separate sheet. For each tournament I have a list of data gathered that includes: tournament seed and finish, wins and losses before the tournament began, Quad 1 wins and losses, conference, rankings for each week of the month leading up to the tournament, KenPom rankings, BPI, strength of schedule, and strength of record. Seeding and wins and losses are pretty straightforward and are important factors in considering what the success of a tournament team may be. BPI is ESPN's power ranking of each team's strength and it is meant to be their best predictor of success in the tournament. Strength of Schedule is a ranking of how much quality competition a team has faced throughout the year, and it can be a great indicator of how a team may react to facing very talented competition late in the tournament. Strength of Record takes strength of schedule into account when looking at a team record and adjusts it accordingly to get a more accurate way to compare teams with similar records. AP rankings are the rankings released by the NCAA on a weekly basis, and KenPom rankings are the rankings released on a popular basketball analytics site that many hardcore NCAA fans follow frequently. A Quad 1 team is considered one of the top teams in the country and a team's Quad 1 record is considered to be very important when considering expected performance in the tournament. I have decided to use these factors to create a formula that will give each team a score that then indicates their chances of winning. I began to create a formula where the larger the final result, the larger the chances of winning. My goal in this was to work backwards and build the formula to result in the winner having the largest score.

First, I took the data I had and created a win:loss ratio by simply dividing wins by losses. I knew that this would be an important element to include in any calculations I would perform, but there was a problem with this. A team that played much more difficult teams and had a worse record would technically have a lower win:loss ratio than a team who only played weak teams all year.

To account for this, I then adjusted the win:loss ratio by both the Strength of Schedule (SOS) and Strength of Record (SOR) to create two new indicators.

To perform this adjustment I divided the win:loss ratio by each of these values to get an SOR adjusted win:loss ratio and an SOS adjusted win:loss ratio. This created a problem, though, in that the distribution of this adjustment is not necessarily proportional. For example, when a team is ranked first in Strength of Schedule or record, then their adjusted win:loss ratio is the same as their original win:loss ratio. However, if another team has the same win:loss ratio but is ranked just third or fourth, which is still really impressive, then their adjusted ratio would be a fourth of their original. The issue is that the sequential differences between dividing by smaller numbers can be very large and create outliers in the data. To account for this, I implemented an if statement into the adjusted win:loss formulas so that if a team was ranked first in that category, then the formula should adjust the value by dividing by two and a half instead of one. I also adjusted it so that if a team was ranked second, it should divide by two and three quarters instead of two. I played around with what numbers would effectively get rid of the outlier through guessing and checking values, and I ultimately decided that this was the closest I could get. After creating a few different fields that I felt would be good starting points for developing a formula, I began to see how I could combine them. To create what I called Formula 1, I first multiplied the SOR adjusted win:loss ratio by Quad 1 wins and dividing by Total Quad 1 games. I thought that combining these two ratios would give a better picture as to who is really doing well when it matters. The results were not fully accurate so I created Formula 2 as a continuation of this by multiplying by Total Quad 1 games and dividing by Quad 1 losses. Formula 1 effectively used the Quad 1 win percentage and Formula 2 used the win:loss ratio. At this point I realized that Strength of Record should not be the only base ranking I use, so I began to repeat every formula I had done but with Strength of Schedule as the base value instead of Strength of Record. For example, Formula 1 is based on SOR but Formula 1.2 is based on SOS. From here I had two routes of formulas going at all times that I could combine to get a more accurate picture of who was truly successful. My idea was to try to continually build on these formulas as they slowly began to account for more information.

The next two pieces of information I wanted to include were a team's conference and their average ranking in the four weeks leading up to the tournament. There are six major conferences in college basketball that regularly produce winning teams and provide consistent competition for their members. These conferences are the Big East, The Big 10, the Big 12, the Pac 12, the SEC, and the ACC. Almost every winner of the tournament in recent history has come from one of these conferences. The thought is that if every team in a conference is at least decent competition, then all of those teams will get better throughout the year simply due to playing consistently against strong competition. Whereas, a team that plays in a conference that is overall not very good will not grow as much throughout the year because they will not really be challenged. A great example of this is Gonzaga. Gonzaga plays in the WCC which is not a very competitive conference overall. Gonzaga constantly blows out most of these teams and has a really strong record, but they simply don't play very competitive teams. They are always a good team and they are frequently ranked very highly but their lack of competition throughout the year can hold them back in March. To adjust for this I created a factor that takes into account the conference and assigns a value accordingly. The value I used was one for each of the six major conferences and then one half for every other conference.

The other factor was AP ranking in the month leading up to the tournament. Since the goal of my formula is for the best team to produce the largest number, simply multiplying by their average rank for the month would actually hurt the good teams and help the bad teams. To fix this, I subtracted each week's ranking from 26, divided it by 25, and then took the average of these 5 results. The thought behind this was that a team who was ranked first would have the calculation (26-1)/25 which is just one and then everything else would be less than one as higher numbers subtracted more from 26. This allowed me to reward teams with low rankings and not those with large rankings. These two fields combined to make Formula 3 and 3.2 respectively as I took Formula 2 and 2.2 and multiplied by the Conference Adjustment and the Average Ranking Adjustment. Both Formula 3 and 3.2 represented two very different paths of information but both were important and I wanted to consolidate them back into one formula so I multiplied the two together to do this.

At this point I started to look back at the different formulas I had created and the outputs they were producing. What was surprising to me was how different the results were for different years. Some years the highest value for Formula 2 was 25 and some years it was six. The problem I was recognizing was that since so many of my formulas were just multiples of different values, those that had an extremely large value for one formula were growing much faster than those with relatively normal values. For example, if one team has a value of 25 on Formula 2 and the other has a value of six, but they each have a conference adjustment of one and average ranking adjustment of three, then the first team will have a value of 75 for Formula 3 and the second team will have a value of 18. I needed a way to counteract this. I decided to do this by normalizing the results of each formula as well as the formulas themselves. This means that for each formula I would calculate the average of all the results, the standard deviation, and then plug these values into the normal distribution function within Excel. The way this works is that it takes the value of a team, subtracts the mean, and divides by the standard deviation. Then this new result is standardized and can be compared to a premade probability table and it gives you a probability as a result. This is the probability that any random number that is produced by the original formula would be less than your specific value. For example, if I take the largest value in the set of results of Formula 2 and plug it into the normal distribution function within Excel, the result will be one. Then the rest of the results will be something below one but above zero, and the distribution of these results will resemble the distribution of the original results of Formula 2, but the maximum value will be set at one. Thus, instead of just having Formula 2 and Formula 2.2 respectively, I had Formula 2, NR2, NF2, and Formula 2.2, NR2.2, and NF2.2. NR2 represents the normalized results of Formula 2 while NF2 represents a new version of Formula 2 with the normalized results of Formula 1 as the input instead of the original results of Formula 1. I followed this path throughout the rest of the formulas as well. For each formula moving forward I had the original results, the normalized results, and then the original formula using the previous formula's normalized results as inputs.

At this point I had a fairly large number of formulas that all had different results so I decided to look through and see what formulas seemed typically to be accurate. My goal with each formula was just to have the winner produce the largest number. I did not have one formula that correctly predicted every tournament but I did have a few formulas that had the winner producing a number in the top three results. I decided to combine these formulas, and this is where I got Formula 3.4. The first two formulas that seemed to capture the winner most accurately were

Formula 1.2 and Formula 3.3, so these were the values I combined by multiplying them together. That still was only accurate for about half of the tournaments, though, so then I went back through and examined the normalized results more closely and found three more formulas that seemed especially accurate. These were NF3, NF3.2, and NF2.2. These three combined to create Formula 4, which was the closest I had come to being accurate, but there was one issue still. I had really only been using the formulas I had created and had disregarded many of the other rankings like KenPom, BPI, and seeds. I was in search of any information that could separate the winners from those currently being predicted by my formulas. So I looked through the information that I had been neglecting and found that Kenpom was actually very accurate in predicting winners. In the eight tournaments that I had data for, no winner had a KenPom ranking worse than third. So I created a formula called KP Adjusted which divided Formula 4 by their Kenpom ranking and this gave me the most accurate formula I could find. (See Appendix for formulas and 2019 Excel data and calculations.)

<u>Results</u>

My results were not far off from what I expected but they were different than what I had hoped. Ultimately, I was able to predict the correct winner with my final formula, KP Adjusted, for six of the eight tournaments I had gathered data for. Five of these tournaments were won by a 1 seed and the other was won by a 2 seed. These six tournaments had the 'correct' team win the tournament and the other two had something unexpected happen. The two years that I was not able to predict correctly were 2014 and 2015. Both of these years my calculations found that a team which was in the Final Four actually should have won the tournament. This is still close to being correct but ultimately was not perfect. In 2015, Duke won the tournament against

Wisconsin. Duke had a very impressive record and was very deserving of this win but there is a clear reason my formula was not accurate. Kentucky, who was eliminated in the Final Four, came into the tournament as a clear favorite with a perfect record. Before the tournament they were 34 and 0, and they had spent the entire month before the tournament ranked first. Their Quad 1 Win ratio was massive compared to every other team as they had not lost a game, and because of this their formula values were much larger from the beginning.

The other year that I was incorrect was 2013, and this was one of the craziest tournaments in NCAA history. The final was 7 seed UConn versus 8 seed Kentucky. No one could have predicted this final and not surprisingly, neither could I. Both teams had decent qualifications leading into the tournament but I do not see anything on paper that says that we should have seen this coming. UConn was 26 and 8 before the tournament and Kentucky was 24 and 10. They were both ranked in the low twenties or not ranked for the month leading up to the tournament and neither one had a very impressive KenPom ranking. I almost consider the tournament results to be an outlier because nothing happened the way it should have. I think it came down to individual players and statistics that cannot be measured. UConn had one of the best college guards in history in Kemba Walker, and he was incredible during the tournament, but no one could have predicted how well he or his team would perform.

Conclusions

March Madness has been around for a pretty long time but it seems like no one has really figured it out. Every year hundreds of thousands of brackets are created and massive cash prizes are offered for anyone with the perfect bracket. The odds are not exactly in your favor with 64 teams playing in a single elimination tournament. The probability of getting a perfect bracket is so small that the number is hard to even comprehend. I decided to tackle just one single game within this crazy tournament and the best I could do was 75% accuracy. From one stance, that is incredibly impressive to be able to predict the winner for six of eight tournaments, but from another perspective I could not even predict this one single game within the tournament. Now, in reality, I am predicting a lot more than one game since I am attempting to predict who will win every single game they play in the tournament, but the point still stands. From paper statistics alone, you cannot predict the winner of the NCAA tournament with 100% certainty. You can come close and you can predict quite a few tournaments correctly, but overall you cannot do it using only the paper statistics that I used. I do think it is possible to predict the winner with close to 100% accuracy but I think you would need a few things which I did not have. The first would be the statistics which cannot be found on paper. How many players does each team have with multiple years of experience? How much experience does the coach have and how much past tournament success does he have? How many star players does each team have and have those players been consistent throughout the season? There are so many statistics that are difficult to measure but have a huge impact on the success of a team.

Along with that, there are many other paper statistics which I was unable to use or just chose not to but could have improved accuracy. Things like offensive efficiency, points per game, points allowed per game, steals and turnovers, and much more. I believe that with a deep analysis of these statistics you could come much closer to correctly predicting the winner consistently. I chose not to collect these statistics due to lack of availability and difficulty to collect. I found it very difficult to find these statistics for a tournament that happened eight years ago and especially to find a way to get all these statistics in a time-efficient manner. As I only had a few months to do this project and would need to have most of the data gathered within the first month, I had to sacrifice some information and those were the statistics I chose to forego. I did this because I had hoped that the rankings and seedings would be the products of these statistics but I think there was a gap there that I was unable to close.

I would also recommend using more than just the past eight tournaments if possible. Of the eight tournaments I considered, six of them were won by 1 seeds and this makes them a little easier to predict with just paper statistics. If you just went with the rule of picking a 1 seed to win every year, you already have a 25% chance of guessing correctly. The years before 2012 had a little more variability in the seed of the winner and I would be curious to see how this formula would work for those tournaments. The last suggestion I would have is to use the data for every team in the tournament and not just those that make it out of the first round. My logic in this choice was that no team who has any real shot at winning the tournament would lose in the first round but it would make this entire process that much more accurate if you had all 64 teams' worth of data.

Just for fun, I decided to run my formulas for the data on this year's March Madness tournament. This year is incredibly different from any other year due to the limited schedules and the limited fans, so I do not know that my formula will still translate to success. However, I still wanted to give it a try and see what outcome I could predict. When running this year's numbers, my formula predicted that Michigan would win the NCAA Tournament. The only problem with this is that Michigan lost one of its best players just before March Madness began. Michigan ultimately lost in the Elite 8 in a very close game. Baylor won the tournament, and I cannot say I am surprised. The other formula I saw frequently predict the winner correctly was Formula 4, and Baylor had the highest score in that formula. The thing that hurt Baylor was their KenPom Ranking. Coming into the tournament they were ranked fourth on KenPom. This had a huge impact on their final score in my formula because whatever number they produced in Formula 4 would then be divided by four. Looking into estimating the 2021 winner actually led me to a major flaw in my formula: the historical KenPom rankings I used were reflecting the games that occurred in the tournament, not just those played before the tournament. The rest of the statistics I collected were all only based on the regular season, but KenPom did not have the option to show the rankings at a certain point in time. Thus, without knowing, I used KenPom rankings which had been adjusted to reflect tournament performance and my final formula was derived from data that could not be collected until after the tournament was finished. Now, the changes in KenPom Rankings this year from before and after the tournament are not massive, but they are enough to change the outputs of my formula. This year, Baylor came into the tournament ranked fourth on KenPom but left the tournament ranked second. When I adjusted my formula for this change in ranking, it actually predicted that Baylor would win. Granted, this may not necessarily be considered a prediction anymore, but nonetheless, when using data consistent with the past methods of collection, my formula actually was correct.

Appendix

Formulas

1. Win Ratio = $\frac{Wins}{Losses}$
2. Win Ratio Adjusted SOS = $\frac{Wins}{Losses} \div Strength\ of\ Schedule$
3. Win Ratio Adjusted SOR = $\frac{Wins}{Losses} \div Strength\ of\ Record$
4. Formula 1 = $Win\ Ratio\ Adjusted\ SOR \times \frac{Quad\ 1\ Wins}{Quad\ 1\ Total\ Games}$
5. NR1 = $\Phi\ (Formula\ 1)$ *
6. Formula 2 = $Formula\ 1 \times \frac{Quad\ 1\ Total\ Wins}{Quad\ 1\ Losses}$
7. NR2 = $\Phi\ (Formula\ 2)$ *
8. NF2 = $NR1 \times \frac{Quad\ 1\ Total\ Wins}{Quad\ 1\ Losses}$
9. Conference Adjustment =

   $if\ Conference = SEC,\ ACC,\ Big\ East,\ Big\ 10,\ Big\ 12,\ Pac\ 12 \Rightarrow 1,\ otherwise \Rightarrow .5$
10. Average Ranking Weight = $(26 \times 5 - \frac{\Sigma\ Ap\ Rankings}{25}) \div 5$
11. Normal Ave = $\Phi\ (Average\ Ranking\ Weight)$ *
12. Formula 3 = $Average\ Ranking\ Weight \times Conference\ Adjustment \times Formula\ 2$
13. NR3 = $\Phi\ (Formula\ 3)$ *
14. NF3 $= Normal\ Ave \times Conference\ Adjustment \times NF2$
15. Formula 1.2 = $Win\ Ratio\ Adjusted\ SOS \times \frac{Quad\ 1\ Wins}{Quad\ 1\ Total\ Games}$
16. NR1.2 = $\Phi\ (Formula\ 1.2)$ *
17. Formula 2.2 = $Formula\ 1.2 \times \frac{Quad\ 1\ Total\ Wins}{Quad\ 1\ Losses}$
18. NR2.2 = $\Phi\ (Formula\ 2.2)$ *

19. NF2.2 = $NR1.2 \times \frac{Quad\ 1\ Total\ Games}{Quad\ 1\ Losses}$

20. NFR2.2 = $\Phi\ (NF2.2)$ *

21. Formula 3.2 = $Average\ Ranking\ Weight \times Conference\ Adjustment \times Formula\ 2.2$

22. NR3.2 = $\Phi\ (Formula\ 3.2)$ *

23. NF3.2 = $Average\ Ranking\ Weight \times Conference\ Adjustment \times NR2.2$

24. Formula 3.3 = $Formula\ 3.2 \times Formula\ 3$

25. NR3.3 = $\Phi\ (Formula\ 3.3)$ *

26. NF3.3 = $NR3.2 \times NR3$

27. NF3.3-F = $NF3.2 \times NF3$

28. Formula 3.4 = $Formula\ 3.2 \times Formula\ 1.2$

29. NR3.4 = $\Phi\ (Formula\ 3.4)$ *

30. Formula 4 = $NF3.2 \times NF3 \times NF2.2$

31. KP Adj = $\frac{Formula\ 4}{Kenpom\ Ranking}$

***** Indicates the cumulative distribution function of a standard normal distribution.

2019 Excel Data and Calculations

Team	Finish	Seed	Wins	Losses	Win %	Win Ratio	Win Ratio Adjusted SOS	Win Ratio Adjusted	BPI	Strength of Schedule
Virginia	1	1	29	3	0.90625	9.666666667	0.878787879	3.866666667	1	11
Texas Tech	2	3	26	6	0.8125	4.333333333	0.196969697	0.866666667	6	22
Duke	Elite 8	1	29	5	0.852941176	5.8	0.966666667	2.109090909	3	6
Purdue	Elite 8	3	23	9	0.71875	2.555555556	0.851851852	0.212962963	9	3
Kentucky	Elite 8	2	27	6	0.818181818	4.5	0.25	0.642857143	7	18
Gonzaga	Elite 8	1	30	3	0.909090909	10	0.135135135	1	2	74
Michigan State	Final 4	2	28	6	0.823529412	4.666666667	1.866666667	1.555555556	4	1
Auburn	Final 4	5	26	9	0.742857143	2.888888889	0.144444444	0.262626263	12	20
UCF	Round of 32	9	23	8	0.741935484	2.875	0.040492958	0.082142857	36	71
Liberty	Round of 32	12	28	6	0.823529412	4.666666667	0.022875817	0.081871345	58	204
Maryland	Round of 32	6	22	10	0.6875	2.2	0.091666667	0.104761905	26	24
Minnesota	Round of 32	10	21	13	0.617647059	1.615384615	0.124260355	0.057692308	53	13
Iowa	Round of 32	10	22	11	0.666666667	2	0.08	0.086956522	37	25
Ohio State	Round of 32	11	19	14	0.575757576	1.357142857	0.096938776	0.039915966	42	14
Baylor	Round of 32	9	20	14	0.588235294	1.428571429	0.052910053	0.03968254	39	27
Oklahoma	Round of 32	9	19	13	0.59375	1.461538462	0.042986425	0.208791209	34	34
Kansas	Round of 32	4	25	9	0.735294118	2.777777778	1.01010101	0.213675214	18	2
Villanova	Round of 32	6	25	9	0.735294118	2.777777778	0.053418803	0.115740741	25	52
UC Irvine	Round of 32	13	30	5	0.857142857	6	0.028169014	0.127659574	85	213
Buffalo	Round of 32	6	31	3	0.911764706	10.33333333	0.103333333	0.543859649	24	100
Murray State	Round of 32	12	27	4	0.870967742	6.75	0.034974093	0.164634146	46	193
Washington	Round of 32	9	26	8	0.764705882	3.25	0.05078125	0.12037037	52	64
Florida	Round of 32	10	19	15	0.558823529	1.266666667	0.084444444	0.028787879	30	15
Wofford	Round of 32	7	29	4	0.878787879	7.25	0.07712766	0.3625	16	94
Virginia Tech	Sweet 16	4	24	8	0.75	3	0.081081081	0.1875	11	37
Florida State	Sweet 16	4	27	7	0.794117647	3.857142857	0.203007519	0.428571429	14	19
North Carolina	Sweet 16	1	27	6	0.818181818	4.5	0.9	1.125	5	5
Houston	Sweet 16	3	31	3	0.911764706	10.33333333	0.135964912	0.688888889	13	76
Michigan	Sweet 16	2	28	6	0.823529412	4.666666667	0.274509804	0.583333333	8	17
Oregon	Sweet 16	12	23	12	0.657142857	1.916666667	0.029040404	0.03422619	31	66
LSU	Sweet 16	3	26	6	0.8125	4.333333333	0.111111111	0.30952381	23	39
Tennessee	Sweet 16	2	29	5	0.852941176	5.8	0.187096774	0.966666667	10	31

Team	Strength of Record	RPI Rankings	Kenpom Rankings	Conference	Conference Tournament Finish	Q1 wins	Q1 losses	Q1 Total	AP Rankings Final
Virginia	1	1	1	ACC	Semifinals	12	3	15	2
Texas Tech	5	7	5	Big 12	Quarter Finals	8	5	13	9
Duke	2	3	4	ACC	1	11	4	15	1
Purdue	12	11	9	Big 10	Quarter Finals	7	7	14	13
Kentucky	7	4	8	SEC	Semifinals	10	5	15	7
Gonzaga	10	2	2	WCC	1	4	3	7	4
Michigan State	3	5	3	Big 10	1	13	4	17	5
Auburn	11	13	11	SEC	1	5	7	12	14
UCF	35	26	34	AAC	Quarterfinal	2	5	7	NR
Liberty	57	53	58	ASUN	1	2	0	2	NR
Maryland	21	25	24	Big 10	Second Round	6	8	14	NR
Minnesota	28	55	46	Big 10	Semifinals	5	10	15	NR
Iowa	23	40	37	Big 10	Quarter Finals	4	10	14	NR
Ohio State	34	52	44	Big 10	Quarter Finals	4	10	14	NR
Baylor	36	34	35	Big 12	Quarter Finals	4	9	13	NR
Oklahoma	7	32	32	Big 12	First round	4	10	14	NR
Kansas	13	17	17	Big 12	2	11	8	19	17
Villanova	24	28	30	Big East	1	5	6	11	23
UC Irvine	47	71	73	Big West	1	1	2	3	NR
Buffalo	19	18	22	MAC	1	2	1	3	15
Murray State	41	43	51	OVC	1	1	2	3	NR
Washington	27	44	48	Pac 12	2	2	4	6	NR
Florida	44	30	26	SEC	2	4	12	16	NR
Wofford	20	14	18	Southern	1	3	4	7	19
Virginia Tech	16	12	13	ACC	Second Round	4	8	12	16
Florida State	9	15	14	ACC	2	8	5	13	10
North Carolina	4	8	7	ACC	Semifinals	10	6	16	3
Houston	15	6	12	AAC	2	6	3	9	11
Michigan	8	10	6	Big 10	2	9	6	15	8
Oregon	56	41	28	Pac 12	1	3	5	8	NR
LSU	14	16	19	SEC	Quarter Finals	9	3	12	12
Tennessee	6	9	10	SEC	2	9	5	14	6

Team	AP Rankings 1 week til Final	AP Rankings 2 weeks til Final	AP Rankings 3 weeks til Final	AP Rankings 4 weeks til Final	Formula 1	NR1	Formula 2	NR2
Virginia	2	2	2	3	3.09333333	0.99999593	15.46666667	0.999999426
Texas Tech	7	8	11	14	0.53333333	0.61356453	1.386666667	0.515492219
Duke	5	4	3	1	1.54666667	0.97383026	5.8	0.939594686
Purdue	13	11	14	15	0.10648148	0.34195376	0.212962963	0.358149841
Kentucky	4	6	4	4	0.42857143	0.54691158	1.285714286	0.50169335
Gonzaga	1	1	1	2	0.57142857	0.63709928	1.333333333	0.508203512
Michigan State	6	9	6	10	1.18954248	0.91280417	5.055555556	0.902554311
Auburn	22	NR	NR	NR	0.10942761	0.34371887	0.187590188	0.354907579
UCF	NR	25	NR	NR	0.02346939	0.29375615	0.032857143	0.335369336
Liberty	NR	NR	NR	NR	0.08187135	0.32734701	0.16374269	0.351869769
Maryland	21	24	17	24	0.04489796	0.30589623	0.078571429	0.3410984
Minnesota	NR	NR	NR	NR	0.01923077	0.29138147	0.028846154	0.334868464
Iowa	NR	NR	22	21	0.02484472	0.29452859	0.034782609	0.335609883
Ohio State	NR	NR	NR	NR	0.01140456	0.28702056	0.015966387	0.333262093
Baylor	NR	NR	NR	NR	0.01221001	0.28746794	0.017636684	0.333470242
Oklahoma	NR	NR	NR	NR	0.05965463	0.31438335	0.083516484	0.341720361
Kansas	17	13	NR	12	0.1237067	0.35232162	0.293803419	0.368547435
Villanova	25	23	NR	17	0.05260943	0.31031877	0.096450617	0.343349173
UC Irvine	NR	NR	NR	NR	0.04255319	0.30455706	0.063829787	0.339246845
Buffalo	18	19	21	25	0.3625731	0.50410364	1.087719298	0.474639506
Murray State	NR	NR	NR	NR	0.05487805	0.31162511	0.082317073	0.341569466
Washington	NR	NR	25	NR	0.04012346	0.30317211	0.060185185	0.338789677
Florida	NR	NR	NR	NR	0.00719697	0.28468886	0.00959596	0.332468697
Wofford	20	22	24	NR	0.15535714	0.37165598	0.271875	0.365717131
Virginia Tech	16	15	20	20	0.0625	0.31603136	0.09375	0.343008839
Florida State	12	14	18	16	0.26373626	0.44006074	0.685714286	0.420197866
North Carolina	3	3	5	8	0.703125	0.71409167	1.875	0.58168443
Houston	11	12	8	9	0.45925926	0.56666219	1.377777778	0.514277728
Michigan	10	7	9	7	0.35	0.49592792	0.875	0.445706949
Oregon	NR	NR	NR	NR	0.01283482	0.28781521	0.020535714	0.333831634
LSU	9	10	13	13	0.23214286	0.41983651	0.928571429	0.452972285
Tennessee	8	5	7	5	0.62142857	0.66720734	1.74	0.56353479

Team	NF2	Confer	Average Ran	Normal Ave	Formula 3	NR3	NF3	Formula 1.2	NR1.2	Formula 2.2	NR2.2	NF2.2	NFR2.2	Formula 3.2
Virginia	4.99997964	1	4.024	0.954209467	62.23786667	1	4.771028	0.703030303	0.954327	3.515151515	0.988745	4.771636	0.999301	14.1449697
Texas Tech	1.59526777	1	2.856	0.817137628	3.96032	0.491154	1.303553	0.121212121	0.421247	0.315151515	0.408495	1.095243	0.474513	0.900072727
Duke	3.651863483	1	3.84	0.941071106	22.272	0.939257	3.436663	0.708888889	0.956121	2.658333333	0.946165	3.585454	0.983968	10.208
Purdue	0.683907513	1	2.208	0.680938718	0.470222222	0.373898	0.465699	0.425925926	0.785117	0.851851852	0.575373	1.570235	0.639509	1.880888889
Kentucky	1.640734728	1	3.496	0.90878912	4.494857143	0.509445	1.491082	0.166666667	0.479552	0.5	0.465634	1.438656	0.59503	1.748
Gonzaga	1.486564975	0.5	4.072	0.957218097	2.714666667	0.448669	0.711483	0.077220077	0.366407	0.18018018	0.367902	0.854949	0.39091	0.366846847
Michigan State	3.879417729	1	3.248	0.878428285	16.42044444	0.852364	3.40779	1.42745098	0.999973	6.066666667	0.999991	4.249885	0.996859	19.70453333
Auburn	0.589232341	1	0.64	0.280857678	0.12005772	0.362575	0.16549	0.060185185	0.345818	0.103174603	0.345356	0.592831	0.305266	0.066031746
UCF	0.411258605	0.5	0.04	0.162971406	0.000657143	0.358741	0.033512	0.011569416	0.289682	0.016197183	0.320538	0.405555	0.249732	0.000323944
Liberty	0.654694025	0.5	0.05	0.164626796	0.004093567	0.358851	0.05389	0.022875817	0.302354	0.045751634	0.328888	0.604707	0.308965	0.001143791
Maryland	0.53531841	1	0.656	0.284483033	0.051542857	0.360373	0.152289	0.039285714	0.321173	0.06875	0.335446	0.562053	0.295772	0.0451
Minnesota	0.437072209	1	0.05	0.164626796	0.001442308	0.358766	0.071954	0.041420118	0.323657	0.062130178	0.333553	0.485485	0.272764	0.003106509
Iowa	0.412340029	1	0.2	0.190756322	0.006956522	0.358943	0.078656	0.022857143	0.302332	0.032	0.324992	0.423265	0.254744	0.0064
Ohio State	0.401828783	1	0.05	0.164626796	0.000798319	0.358745	0.066152	0.027696793	0.307831	0.03877551	0.326909	0.430964	0.25694	0.001938776
Baylor	0.415231471	1	0.05	0.164626796	0.000881834	0.358748	0.068358	0.016280016	0.294931	0.023515579	0.322597	0.426011	0.255526	0.001175779
Oklahoma	0.440136687	1	0.05	0.164626796	0.004175824	0.358854	0.072458	0.012281836	0.290473	0.01719457	0.320818	0.406662	0.250044	0.000859729
Kansas	0.83676384	1	1.352	0.458891065	0.397222222	0.371528	0.383983	0.584795322	0.90403	1.388888889	0.729672	2.147072	0.807442	1.877777778
Villanova	0.568917752	1	0.352	0.219682282	0.033950617	0.359809	0.124981	0.024281274	0.303946	0.04451567	0.328537	0.557234	0.294298	0.015669516
UC Irvine	0.456835587	0.5	0.05	0.164626796	0.001595745	0.358771	0.037604	0.009389671	0.287268	0.014084507	0.319944	0.430902	0.256922	0.000352113
Buffalo	1.512310917	0.5	1.248	0.43136034	0.678736842	0.380693	0.326175	0.068888889	0.356285	0.206666667	0.375768	1.068855	0.465207	0.12896
Murray State	0.467437664	0.5	0.05	0.164626796	0.002057927	0.358786	0.038476	0.011658031	0.28978	0.017487047	0.320901	0.43467	0.258	0.000437176
Washington	0.454758169	1	0.04	0.162971406	0.002407407	0.358797	0.074113	0.016927083	0.295655	0.025390625	0.323126	0.443483	0.260531	0.001015625
Florida	0.379585142	1	0.05	0.164626796	0.000479798	0.358735	0.06249	0.021111111	0.300359	0.028148148	0.323904	0.400479	0.248305	0.001407407
Wofford	0.650397958	0.5	0.76	0.308575968	0.1033125	0.362036	0.100349	0.033054711	0.31397	0.057845745	0.33233	0.549447	0.291923	0.021981383
Virginia Tech	0.474047038	1	1.528	0.505863832	0.14325	0.363321	0.239803	0.027027027	0.307068	0.040540541	0.32741	0.460601	0.265482	0.061945946
Florida State	1.144157917	1	2.08	0.649718436	1.426285714	0.405344	0.74338	0.124927704	0.425967	0.32481203	0.411445	1.107513	0.478845	0.675609023
North Carolina	1.904244444	1	3.744	0.933128668	7.02	0.594943	1.776905	0.5625	0.891129	1.5	0.757749	2.376343	0.858085	5.616
Houston	1.699986582	0.5	2.616	0.771471377	1.802133333	0.417888	0.655745	0.090643275	0.382909	0.271929825	0.395366	1.148727	0.493411	0.355684211
Michigan	1.239819792	1	2.952	0.833686059	2.583	0.444204	1.03362	0.164705882	0.477019	0.411764706	0.438196	1.192547	0.508907	1.215529412
Oregon	0.460504344	1	0.05	0.164626796	0.001026786	0.358753	0.075811	0.010890152	0.288929	0.017424242	0.320883	0.462286	0.265972	0.000871212
LSU	1.679346056	1	2.504	0.74814281	2.325142857	0.435481	1.256391	0.083333333	0.373894	0.333333333	0.41405	1.495575	0.614459	0.834666667
Tennessee	1.868180554	1	3.288	0.883754625	5.72112	0.551267	1.651013	0.120276498	0.420061	0.336774194	0.415104	1.17617	0.503116	1.107313548

Team	NR3.2	NF3.2	Formula 3.3	NR3.3	NF3.3	NF3.3-F	Formula 3.4	NR3.4	Formula 4	KP Adj
Virginia	0.996836241	3.978708856	880.352738	0.999999707	0.996835919	18.98253098	9.944342332	0.951308	90.57773558	90.57774
Texas Tech	0.411213254	1.166663051	3.564576023	0.397747401	0.201969136	1.520807497	0.109099725	0.173223	1.665654087	0.333131
Duke	0.967989279	3.633274124	227.352576	0.860395199	0.909190255	12.48633951	7.236337778	0.925515	44.76919103	11.1923
Purdue	0.497757909	1.270423613	0.884435753	0.391563815	0.186110462	0.591635139	0.801119342	0.390798	0.929006038	0.103223
Kentucky	0.485933981	1.627856328	7.857010286	0.407703971	0.247556506	2.427267056	0.291333333	0.233031	3.492002599	0.4365
Gonzaga	0.365656434	0.749049281	0.995866907	0.391820374	0.164058776	0.532936165	0.028327942	0.133979	0.455633243	0.227817
Michigan State	0.999964148	3.24797038	323.5571949	0.951411996	0.852333938	11.06840183	28.12725542	0.999937	47.03943929	15.67981
Auburn	0.340724414	0.221027866	0.007927621	0.389547394	0.123538144	0.036577996	0.003974133	0.117829	0.021684568	0.001971
UCF	0.335365386	0.00641076	2.12877E-07	0.38952917	0.120309278	0.000214835	3.74784E-06	0.097149	8.71275E-05	2.56E-06
Liberty	0.335432048	0.008222208	4.68218E-06	0.38952918	0.120370141	0.000443096	2.61652E-05	0.101419	0.000267943	4.62E-06
Maryland	0.339013696	0.220052697	0.002324583	0.389534513	0.122171441	0.033511606	0.001771786	0.108882	0.01883531	0.000785
Minnesota	0.335591659	0.016677666	4.48054E-06	0.38952918	0.120398897	0.001200021	0.000128672	0.108616	0.000582592	1.27E-05
Iowa	0.335859557	0.064998394	4.45217E-05	0.389529272	0.120554377	0.005112544	0.000146286	0.101541	0.002163963	5.85E-05
Ohio State	0.335496694	0.016345468	1.54776E-06	0.389529173	0.120357902	0.001081282	5.36979E-05	0.103276	0.000465993	1.06E-05
Baylor	0.335434649	0.016129875	1.03684E-06	0.389529172	0.120336542	0.00110261	1.91417E-05	0.09893	0.000469724	1.34E-05
Oklahoma	0.33540895	0.016040917	3.59008E-06	0.389529178	0.120362737	0.001162297	1.0559E-05	0.097427	0.000472662	1.48E-05
Kansas	0.497481032	0.986516907	0.745895062	0.391244904	0.184827993	0.378806165	1.09811566	0.449738	0.813324221	0.047843
Villanova	0.336614002	0.115645149	0.00053199	0.389530393	0.121116603	0.014453464	0.000380476	0.102312	0.008053964	0.000268
UC Irvine	0.335367676	0.007998612	5.61882E-07	0.389529171	0.120320189	0.000300777	3.30622E-06	0.09634	0.000129606	1.78E-06
Buffalo	0.345887078	0.234479534	0.087529903	0.389730398	0.131676672	0.076481474	0.008883911	0.123234	0.081747623	0.003716
Murray State	0.335374592	0.008022514	8.99677E-07	0.389529172	0.120327639	0.000308677	5.09661E-06	0.097185	0.000134173	2.63E-06
Washington	0.335421626	0.012925042	2.44502E-06	0.389529175	0.120348272	0.000957908	1.71916E-05	0.099169	0.000424816	8.85E-06
Florida	0.335453484	0.0161952	6.75271E-07	0.389529171	0.120338976	0.001012036	2.97119E-05	0.100757	0.000405299	1.56E-05
Wofford	0.337128101	0.126285585	0.002270952	0.38953439	0.122052651	0.01267258	0.000726588	0.105848	0.006962907	0.000387
Virginia Tech	0.340390229	0.500281874	0.008873757	0.389549569	0.123671025	0.11996922	0.001674215	0.104523	0.055257988	0.004251
Florida State	0.391850375	0.855805002	0.963611497	0.391746104	0.158834139	0.636188743	0.084402284	0.166915	0.704587383	0.050328
North Carolina	0.796058778	2.837013451	39.42432	0.48235025	0.473609321	5.041103619	3.159	0.709391	11.97939079	1.711342
Houston	0.36472026	0.517138708	0.640990372	0.39100347	0.152412208	0.339111378	0.032240382	0.139655	0.389546539	0.032462
Michigan	0.438784525	1.293555757	3.139712471	0.396765402	0.194909928	1.337045718	0.200204844	0.209308	1.594489583	0.265748
Oregon	0.335409884	0.016044146	8.94548E-07	0.389529172	0.120329216	0.001216328	9.48763E-06	0.09691	0.000562291	2.01E-05
LSU	0.405546499	1.036782042	1.940719238	0.393997684	0.17660776	1.302603291	0.069555556	0.151631	1.948141186	0.102534
Tennessee	0.429285744	1.364860486	6.335073688	0.404166559	0.236651032	2.253402687	0.133183795	0.180326	2.650385163	0.265039

Documentation of Sources

2019 Pomeroy College Basketball Ratings, kenpom.com/index.php?y=2019.

"College Basketball Power Index - NCAA Tournament Predictions." *ESPN*, ESPN Internet Ventures, www.espn.com/mens-college-basketball/bpi/_/season/2019/view/overview.

Gaudios, Fred. "KenPom Rankings Explained & How to Better Evaluate Rutgers Basketball." *On the Banks*, On the Banks, 9 Dec. 2019, www.onthebanks.com/2019/12/9/21002735/kenpom-rankings-explained-how-to-better-evaluate-rutgers-basketball-big-ten-ncaa-steve-pikiell.

Matt Norlander Apr 9. "NCAA Bracket 2019: Check out How Virginia Was Able to Claim the National Championship." *CBSSports.com*, 9 Apr. 2019, www.cbssports.com/college-basketball/news/ncaa-bracket-2019-check-out-how-virginia-was-able-to-claim-the-national-championship/.

"Men's Basketball." *2014 Final AP Men's Basketball Poll - College Poll Archive - Historical College Football and Basketball Polls and Rankings*, collegepollarchive.com/mbasketball/ap/seasons.cfm?seasonid=2014#.X08Jf8hKiUk.

"NCAA Basketball Tournament Selection Process." *Wikipedia*, Wikimedia Foundation, 19 Mar. 2021, en.wikipedia.org/wiki/NCAA_basketball_tournament_selection_process#:~:text=As of the 2016–17,who receive at-large bids.

"NCAA RPI Archive // ." *NCAA RPI Archive - Home*, extra.ncaa.org/solutions/rpi/SitePages/Home.aspx.

NCAA.com. "2019 NCAA Conference Tournaments: Schedules, Brackets, Scores and Auto-Bids." *NCAA.com*, NCAA.com, 17 Mar. 2019, www.ncaa.com/news/basketball-men/2019-03-13/2019-ncaa-conference-tournaments-schedules-brackets-scores-and-auto.

NCAA.com. "DI Men's Basketball Championship History." *DI Men's Basketball Championship History | NCAA.com*, NCAA.com, 6 Apr. 2021, www.ncaa.com/history/basketball-men/d1.

"Rank - Customizable College Basketball Tempo Free Stats - T-Rank." *T*, barttorvik.com/trank.php?year=2019&sort=&lastx=0&hteam=&t2value=&conlimit=All&state=All&begin=20181101&end=20190318&top=0&quad=3&venue=All&type=All&mingames=0#.

“Seeds of NCAA Tournament Champions.” *BracketResearch.com*, 16 Mar. 2021, bracketresearch.com/the-dna-of-a-national-championship-team/seeds-of-ncaa-tournament-champions/.