Quest for Perfection: A Statistical Analysis on the NCAA Tournament in College Basketball

Jarrett Karnibad

MSSM 620

Mount St. Mary's University

May 6, 2024

Abstract:

The NCAA Tournament, also known as March Madness, is the pinnacle of collegiate sports in the United States. Each year, millions and millions of people fill out brackets to try to accurately predict the winners of the 63 college basketball games that crown the champion. These brackets are entered into pools against family and friends, coworkers around the office, and occasionally, strangers with lots of money on the line. There is one thing that is the dream of every college basketball fan that has remained untouched: the perfect bracket. The odds of completing a perfect bracket are estimated to be over 1 in 9.2 quintillion, and thus no has ever filled one out, and it is likely that no one will, at least in the next fifty years. However, through the use of statistical analysis, this study monitors and analyzes the success of four self-created, and heavily tested and improved models for 2022-2023 and 2023-2024 NCAA college basketball season. The four models are XPoint, AFT, ZSCORE, and COMBO. This study will additionally compare the four self-created models to the models of research teams with similar goals and purposes.

Note: All statistics used in this study come from Sports-reference.com, Warrennolan.com, and teamrankings.com. This research would not be possible without them, and their work is very much appreciated.

Disclaimer: All bets placed in this study are imaginary and are used simply to help provide another layer of analysis to the models. No actual bets were placed during the process of this study.

1. Introduction:

The NCAA Basketball Tournament transcends the boundaries and reach of a typical collegiate, or even professional, sporting event. The tournament is watched by millions of people each year and is known for its unpredictability and amazing storylines. However, the thing that separates the NCAA Tournament from all other American sporting events, is the bracket. Each year millions upon millions of people fill out brackets trying to predict the winners of all 63 tournament games, and eventually crown the National Champion. This ritual of filling out brackets occurs in a five-day period from Selection Sunday, the Sunday before the third Thursday of March and the day that the teams are selected into the tournament and the official bracket is created, and noon of the third Thursday of March, when the tournament games begin. In those, roughly, 88 hours, many will fill out a bracket or two as a part of a family or office pool, but a small few, will spend those 88 hours compiling data and statistics in hopes doing what has never been done before: creating a perfect bracket. While the odds of accomplishing a perfect bracket are infinitesimal, around 1 in 9.2 quintillion, it is still sought after by both the casual fan, the die-hard basketball fan, and the statistical analysts (Wilco, 2024). The reason that the odds of filling out a perfect bracket are 1 in 9.2 quintillion is because every single matchup has to be correct, meaning the odds are 2 raised to the 63^{rd}. Imagine there was only a four-team bracket, as is the case in college football. There are only three games, the two semi-finals and the championship. The way to calculate the number of possible unique brackets is found by raising the number of possible outcomes per game, in this case it is two: Team A or Team B, to the number of total games in the bracket, in this case three. Therefore, the odds of creating a perfect bracket in a tournament with four teams is 1 in 8 or 12.5%. Now some argue that the odds are not actually that high because with statistical information and historical trends, some of the 63 games are easier to predict than others. However, this can be misleading, because even the 1-seeds can lose in the first round. It has happened twice in the history of the tournament: UMBC knocking off 1-seed Virginia in 2016, and Fairleigh Dickinson defeating 1-seed Purdue in 2023. These two upsets prove that even with all of the data and statistical analysis available, crazy things happen in March.

This study analyzes the 2023 and 2024 NCAA Tournaments through the creation and application of four self-created models, as well as three comparison models from known statisticians and researchers. The self-created models use a plethora of data, coming from multiple sports statistical databases such as sports-reference.com, kenpom.com, and warrennolan.com. The models all try to analyze the game through different methods, often providing different predictions on which team will win. The four models are XPoint, AFT, ZSCORE, and COMBO. The names of the comparison models are 4Facts PS, 4Facts LR, and KENPOM.

2. Related Literature:

The quest to analyze basketball games, NCAA tournament games in particular, has been undertaken by many, both individuals such as myself, and full research teams such as Magel and Unruh, and Lopez and Matthews. These are just two studies performed that use statistical analysis, free of analyst bias, to try to predict the winners of college basketball games. Magel and Unruh's work is going to the basis for two of the comparison models used in this study to measure the success of my personal models. Rhonda Magel and Samuel Unruh's paper "Determining factors influencing the outcome of college basketball games" was written in 2013 and tries to pinpoint the statistics responsible for determining the winner of a given game. The study analyzed twelve different independent variables. Their research dictated that there were four independent variables that most greatly affected. From this point, the team created two models. The first model was the points spread model. This model used the least squares regression model to find the coefficients for each of the four variables, ranked in importance according to their coefficients: turnovers, defensive rebounds, assists, and free throw attempts. All four of these variables are the difference between the two teams. This means that if Team A, or as they call them, the Team of Interest has 15 turnovers, and Team B, or the Opposing team, has 10 turnovers, then the number that goes into the model is 5. The equation for determining the points spread for a given game is:

$$\text{Points Spread Model: } Y=0.062(FTA)+1.49(DR)+0.587(A)-1.601(TO)$$

The coefficients of the equation correlate to how many points the team of interest either gains or loses per change in corresponding variable. For example, a change by +1 in the difference of defensive rebounds correlates to an increase of 1.49 points for the team of interest. If the equation provides a positive number, then the team of interest is predicted to win the number of points that the equation provides. This is why it is called the points spread model. It is providing a point spread for a game between any two teams. This particular model, over 100 randomly selected regular season games from the 2011-2012 regular season, correctly predicted the winner in 64 of the 100 games.

The second model from Magel and Unruh's study is the logistic regression model. This model uses the same four statistics as the point spread model. However, this model gives the percentage chance a team wins based upon the variables. The equation is:

$$\text{Logistic Regression Model: } \pi(FTA,DR,A,TO) = \exp(0.123(FTA)+0.488(DR)+0.363(A)-0.474(TO))/D^*$$

$$D^* = 1 + \exp(0.123(FTA)+0.488(DR)+0.363(A)-0.474(TO))$$

The given percentage is the probability that the team of interest wins a given game. This model correctly predicted the winner of 68 of 100 randomly selected games from the 2011-2012 regular season. The two models both correctly predicted games at over 60%. This is very decent results, and both models will be used as comparisons in this study (Magel & Unruh, 2013).

One thing to note about the work of Magel and Unruh is that much of their work was analyzing games using the results from that game. This is how they were able to select their variables: the four statistics most likely to determine the winner of a game. However, it is not as impressive to correctly pick the winner of the game based on information from that game. That is like predicting which student performed better on a test while looking at their multiple-choice answers in that same test. However, when Magel and Unruh tried to predict the winner of a game using data from previous games, they were not nearly as successful. This is because predicting games before they start is much more difficult than "predicting" the winner of a game using the final box score. Finally, Magel and Unruh used only the last four games as their statistical base to predict future games, but in this study, 4Facts PS and 4Facts LR will use each team's full-year statistics in order to fit a team's ability and performance more accurately.

The goal of Lopez and Matthews' study is very similar to this one, theirs titled: "Building an NCAA Men's Basketball Predictive Model and Quantifying Its Success." Their actual research is fairly similar to the research accomplished in this study in that it discusses the impact of possessions on predicting games, and they also use Ken Pomeroy's website as a source for some of their data. Pomeroy is a huge inspiration for this project, and his specific predictive method (KENPOM) is used as a comparison for the self-created models of this project. In conjunction with using Ken Pomeroy's possession-based data, Lopez and Matthews interestingly decide to use the points spread of a game as given by sportsbook. Their argument is that the points spread is a fairly strong prediction of a final score on its own, and thus should have some predictive success in determining the winner of future games. Their points spread model was built using logistic regression involving the points spreads available for the first-round games. The points spread for the second-round and beyond are unavailable until the first-round games are over, and thus Lopez and Matthews used a linear regression model built from the regular season results of that season in order to predict the points spreads beyond the first round. The efficiency model built from Pomeroy's data was created through logistic regression of 5 variables: Home and Away Offensive Efficiency, Home and Away Defensive Efficiency, and a Neutral Location Indicator. After analyzing the data and simulating it 10,000 to ensure a large enough sample size, the authors concluded that it is very difficult to find a predictive model that will outperform the points spread. This will be something to research further in the future (Lopez & Matthews, 2015).

The final study in this section comes from Paul Fearnhead and Benjamin M. Taylor from the Department of Mathematics and Statistics at Lancaster University in the United Kingdom. This is just another example of how popular March Madness is, that even statistics departments in Europe follow and analyze it. Fearnhead and Taylor are concerned with the current, as of 2010, system used to analyze strength of schedule in selecting teams for invites to the NCAA Tournament. Strength of Schedule has, and will continue to be, a contentious topic in the world of college sports. Because there are a limited number of automatic bids to the national championships, the rest of the invites go to “at-large bids.” These bids can go to any team. What makes these bids so difficult to award to teams in the difference in levels of competition across the country. The Southeastern Conference, or SEC, is a lot stronger than the Northeastern Conference, or NEC. We know this to be the case because the SEC constantly recruits higher ranked recruits, and consistently outperforms the NEC in out-of-conference play, and in national tournaments. So, how can a team in the SEC be compared to a team in the NEC? The answer is through strength of schedule. The current formula in place for determining strength of schedule is called the NET rankings. The NET ranking is a complex formula built from a multitude of data, but the formula that was in place at the time of Fearnhead and Taylor's study was the Ratings Percentage Index, or RPI. This system is actually now used as a part of one of the models, and it will be discussed at length later. Essentially, it combines a team's win record, its opponents' record, and the record of their opponents' opponents. While it is a good statistic to use for the purposes of models like mine, it actually struggles to compare two teams against each other, as Fearnhead and Taylor point out. They wrote, "The first problem is that the ranking of teams does not allow a simple and direct comparison of different team's records. For example, which record was better in the 2008/09 season: Wisconsin's win-loss record of 19-13 with the 42nd hardest schedule, or St. Mary's one of 24-6 with the 118th hardest schedule? Answering this question is not possible without a way of quantifying, in terms of wins/losses, how much harder the first schedule is than the second" (Fearnhead & Taylor, 2010). They bring up an excellent point, when allocating the at-large bids, it often comes down a handful of teams being compared to each other on, what is known as, "the bubble," which represents the border from being in the tournament to being left on the outside. Additionally, the teams on the bubble, all often come from different conferences with perceived differences in ability and talent.

Essentially, the work of Fearnhead and Taylor can be explained by predicting the number of wins on one team's schedule and comparing it to the team's actual number of wins. They use the formula of actual wins minus predicted wins to argue which teams should have deserved a bid. This is an early concept of "strength of record," a concept that has grown in popularity recently for its use in college football. They argue that teams that outperform their predicted win probability are more deserving of a spot in the tournament. While their research is sound, and the data is reasonable, they are discounting the scores of games in their calculation. Some may argue that this is the correct way to go about this issue, while others will say that it matters if a team gets blown out. Their research applies to which teams actually make the NCAA Tournament, and thus it does not have many applications once the tournament draw is selected. However, their strength of schedule calculation requires them to calculate win probabilities for every game that the teams play in order to calculate the predicted number of wins for the season. This could be useful moving forward, as it is similar to Magel and Unruh's study of calculation win percentage based on the four factors using a linear regression model.

3. History:

This study is a continuation of a lifelong pursuit of perfection in the March Madness tournament. Since the 8th grade, I have been creating formulas and models and power rankings to try and achieve the impossible: a perfect bracket. While the odds have always been against those who strive for perfection, I have always persevered and continued the journey. My first real attempt for perfection came as a sophomore in high school. Before this year, I had been doing the calculations by hand with a pencil and a calculator. This was the first time I had used Google Sheets to aid the process and make things significantly easier. I had created a power ranking that I thought would be highly successful. Michigan State was the highest team by far according to my formula. They were ranked significantly higher than even the second-place team, and thus I felt confident filling out the bracket. Michigan State was a 2-seed that year and became just the 8th 2-seed to ever lose a first-round game to a 15-seed. My heartbreak was unimaginable, and yet, I felt more desire than ever to get back, do more research, and hopefully, eventually come close to perfection. Just two years later, XPoint was created. The first model of its kind, rather than producing a power ranking, meaning the team with the higher number is better, and thus will win a hypothetical game, XPoint is a formula that examines the exact matchup between two teams. Additionally, it is the first model to give a predicted final score of any given matchup. XPoint is still being used 5 years later, though it has evolved, and it will be examined in this project. Just one year after the Michigan State fiasco, another power ranking formula was created, called AFT, which stands for Assists, Frees (throws), and Three's. This formula showed a lot of success in the 2019 tournament, and thus was brought back for both the 2023 and 2024 tournaments. Leading into the 2020 NCAA Tournament, confidence was high due to the creation of a new type of formula using the sum of z-scores of five different statistics in order to create a new power ranking. However, due to the Coronavirus pandemic and lockdown and the creation of additional models, this one was never used until 2023 and 2024. In this study, it will be known as ZSCORES. In 2022, I completed a project similar to this one discussing the successes, or lack thereof, of two models: XPoint and one called KTheory. Both models were matchup-based, meaning the score predictions change based upon opponent. While neither saw great success, a lot was learned from analyzing the two models, and XPoint was improved, while KTheory went into the vault for now. While I am not the first, nor will I be the last, to attempt a perfect bracket using only statistics, this journey has been incredibly fun, and I plan to continue it for many years, whether I eventually succeed or not.

4. The Models:

This section is going to explain both the background and details of each of the models used in this study, as well as the inspiration for each.

a) XPoint and SOS:

This model dates back to 2019 and was my first self-created matchup model. It is designed to give the expected final score between any two college basketball teams if they were to play. The model is based on predicting how many shot attempts a team will take, and what percent of those shots they will make. In order to predict how many shot attempts Team A will take in a given game, the average must be taken from Team A's shot attempts and the amount of shot attempts that Team B allows per game. This must be done for free throws, 2-pointers, and 3-pointers. This will provide the shot attempts for Team A in a hypothetical game against Team B. Then, the provided shot attempts are multiplied by the shooting percentages of Team A in order to predict how many of those shot attempts Team A will make. This provides three numbers, the amount of made shots for free throws, 2-pointers, and 3-pointers. Finally, these numbers are multiplied by how much they are worth in basketball: free throws are multiplied by one, 2-pointers by two, and 3-pointers by three. This results in a predicted score for Team A against Team B. The same process is followed to obtain Team B's predicted score against Team A.

The nice thing about a model such as this, is that Team A's predicted score will be different against Team B than it will be against Team C. This is the basic skeleton of XPoint and has been unchanged in the last four years. Changes have been implemented, however, on how to involve the Strength of Schedule (SOS) metric.

The SOS metric is also self-designed and revolves around the Quadrant system used to create a team's NET rank, which is used to help determine which teams are selected for the NCAA tournament. The quadrant system is shown in Figure 1 below. In this figure, it shows how any

Quadrant 1 (Q1)	**Quadrant 2 (Q2)**	**Quadrant 3 (Q3)**	**Quadrant 4 (Q4)**
Home (1-30)	Home (31-75)	Home (76-160)	Home (161-358)
Neutral (1-50)	Neutral (51-100)	Neutral (101-200)	Neutral (201-358)
Away (1-75)	Away (76-135)	Away (136-240)	Away (241-358)

Figure 1: The Quadrant system, used in the NET Rankings, and my personal Strength of Schedule model.

game falls into one of the four quadrants. The rankings shown in the figure are not the AP rankings, but rather the RPI rankings. RPI stands for ratings percentage index, and it can be found "by adding a team's winning percentage by its opponents' winning percentage and its opponents' opponents' win percentage. The actual formula is: (0.25*WP)+(0.5*OWP)+(0.25*OOWP)" (Karnibad, 2023). The self-designed SOS metric used in XPoint uses a descending series of multipliers for the number of games each team plays in each quadrant. Teams with a higher SOS play a stronger schedule, and thus their statistics are "stronger." One major issue when using regular season statistics to determine tournament success is that each team does not play on an equal playing field. A school like University of Kentucky which plays in traditionally one of the strongest conferences, the SEC, will play a tougher schedule than a school like Gonzaga University, which plays a very weak conference, the West Coast Conference. This allows the possibility of "inflated" statistics for schools like Gonzaga because they are playing against poor competition. The goal was to even the playing field. To accomplish this, a major change was made to XPoint for the 2023 tournament. Instead of applying the SOS multiplier after the final scores have already been compiled, the SOS metric was implemented to every single individual statistic. Up until then, the final scores of every game were multiplied by the SOS multiplier. The theory was that this would get rid of the inflated stats and level the playing field. The SOS descending multiplier scale used to produce the final SOS metric was adjusted to have a smaller range that mostly revolved around the number 1. This was so that adjustments made would not have as large of an impact on the predicted total score of the game. Finally, instead of multiplying the SOS metric by each individual statistic, it was used as an exponent. The reason for this decision to use the SOS metric as an exponent instead of as a multiplier was that it would decrease the amount of change to each statistic.

b) AFT:

The second model used in this study is AFT, which again stands for Assists, Free's, and Three's. These are the three main statistics used in this power-ranking model. AFT is unlike XPoint, KENPOM, and 4Facts because it provides a power ranking. To determine the winner of a hypothetical game, the team with the higher number is the winner, however a team's number does not change based on who they are playing. The first step of AFT is calculated by adding together a team's number of assists, free throws made, and three pointers made, and then dividing by the number of rebounds they allow. Then the team's defensive efficiency, divided by 100, is subtracted from the first total. The defensive efficiency is taken from Ken Pomeroy's website, and it represents the amount of points a team gives up per 100 possessions. The penultimate step in calculating the AFT power ranking is multiplying the second total by each team's Strength of Schedule number from Jeff Sagarin, a writer and sports statistician from USA Today. Unfortunately, after the 2023 NCAA Tournament, Sagarin retired and stopped updating his website with all of his statistics and research. Therefore, 2024 AFT had to use a different strength of schedule metric to replace Sagarin's. Additionally, 2023 AFT was reanalyzed using the new strength of schedule metric and using the pre-tournament statistics rather than including the tournament games. The new SOS metric comes from the ratings percentage index. In the ratings percentage index, 75% of the formula comes from the win-loss records of a team's opponents (66% of the 75%), and the win-loss record of their opponents' opponents (33% of the 75%). This number, multiplied by 100 for scaling purposes, what is multiplied by the second total calculated earlier. The last step is adding "seed points." Four points are awarded to all of the one-seeds, three points to two-seeds, two points to three-seeds, and finally one point to four-seeds.

c) ZSCORE:

ZSCORE, like AFT, is a power ranking model that uses the statistical measurement: z-score, which allows statisticians to compare different statistics on a like scale. ZSCORE uses 5 different statistics, hand-picked to try to find the most balanced teams, and finds the sum of the five z-scores for each team. A z-score is found by subtracting the mean of the population, in this case it is the 68 tournament teams, from the observed statistic, and then dividing that by the standard deviation of the population, again it is the 68 tournament teams. The five statistics are steals, average scoring margin, total rebounds, free throw percentage, and Ken Pomeroy's strength of schedule metric. Steals is used in order to gauge how aggressive a defense is. Typically, defenses that are aggressive and steal the ball perform better in March Madness. Every champion in the last 21 years has been in the top 25 in Ken Pomeroy's adjusted defensive efficiency metric. Average scoring margin measures how a team performs on average, and it is found by subtracting points allowed per game from points scored per game. This helps show which teams do not get blown out in losses. Teams with a high average scoring margin are typically among the best teams in the country. Total rebounds are fairly straightforward. There was an interview after a tournament game in 2016 in which 12-seeded Yale upset 5-seeded Baylor, where a reporter asked Baylor forward, and now NBA player Taurean Prince, how Yale could outrebound Baylor. The reporter was asking how a small team like Yale could outrebound a strong Big-12 team like Baylor. Prince responded by explaining how rebounds work. He said, "You go up and grab the ball of the rim when it comes off, and then you grab it with two hands, and you come down with it, and that's considered a rebound. So, they got more of those than we did" (Prince, 2016). While Prince's answer was not what the reporter was looking for, the main point remains, how does a team like Yale, outrebound Baylor and upset them in the first round of the tournament? One of the most common denominators in tournament upsets is lower-seeded teams winning the rebound battle. That is why it is important to include it in this model. Free throw percentage is another fairly straightforward addition. Any basketball coach will tell you how important it is to make free throws. Teams that make higher percentages of free throws tend to succeed in the tournament because they can close out games when the losing team fouls to stop the clock. The last statistic is Ken Pomeroy's strength of schedule margin. Pomeroy uses an un-traditional strength of schedule metric. His metric is found by finding the margin between the average of all of one team's opponents adjusted offensive and defensive efficiency. For example, Team A will play 5 games against Teams B, C, D, E, and F. They each have their own offensive and defensive efficiencies based on the games that they play. To find Team A's strength of schedule, take the average of the defensive efficiency of Teams B, C, D, E, and F, and subtract it from the average of the offensive efficiency of Teams B, C, D, E, and F. This gives the adjusted efficiency margin of Team A's opponents, and thus their strength of schedule.

d) COMBO and Upset Indicators.

This serves as a first for my models. Instead of COMBO being its own model, it acts as an aggregate between the previous three models: XPoint, AFT, and ZSCORE. Those three models all measure very different things, and thus favor different teams. XPoint is the matchup model of the group, analyzing how a particular matchup could play out in a game. AFT and ZSCORE are both power rankings but use entirely different statistics to produce their final findings. In fact, there is not one statistical overlap between the two, and thus they should provide different insights into the relative strength of teams. The way COMBO works is very simple: whichever team is predicted in at least two out of three models is chosen to win the game. There is however, one override situation, and it comes from a self-created criteria called Upset Indicators.

This acts as a way to identify which teams are prone to being upset in the first two rounds of the tournament. It serves no purpose beyond that. In the four years of its existence, it has identified 24 possible highly seeded teams to lose in the first two rounds of the tournament. Of those 24 teams, 19 were correctly identified as losing on opening weekend. This is a 79% hit probability which is incredibly impressive. Obviously, there were highly seeded teams that were upset that Upset Indicators did not predict, however this model has proven successful in its four years. Upset Indicators revolves around Ken Pomeroy's adjusted efficiencies and tempo, as well as a team's KenPom rank compared to their actual tournament seed. Upset Indicators is looking for "red flags" in a team. Everyone is familiar with the term, "red flag," but if not, it is something that usually signals warning or danger. In this situation, if a team has two or more red flags, then they are predicted to lose in the first two rounds. There are four categories in which a team can have a red flag. The first two categories are offensive and defensive efficiency. If a team has an offensive or defensive efficiency that is outside of the top 68 teams in the country, then they are given red flags for each that is outside (68 is used because that is how many teams make the tournament). If Team A has rank 80 offensive efficiency and rank 15 defensive efficiency, then they are given 1 red flag, in this case for the offensive efficiency. The third category is tempo, which is the number of possessions per 40 minutes. If a team has lower tempo rank than 170, then that is a red flag. While teams that play slow are not always susceptible to being upset, if they do trail early in a game, it is more difficult for them to come back. The final category is team's seed is lower (higher in numerical value, lower in seeding) than its expected seed according to Ken Pomeroy's adjusted efficiency margin rankings, then that is a red flag. If a team has more than two red flags, then they are predicted to lose within the first two rounds of the tournament.

Going back to COMBO, if the aggregate predicts Team A (4-seed) to beat Team B (13-seed) in the first round, but Upset Indicators shows two red flags for Team A, then the initial result is overridden, and Team B is predicted as the winner. For teams seeded 5 or 6, two or more red flags means a predicted first round loss. For teams seeded 1-4, three red flags mean a predicted first round loss, while two red flags mean a predicted second round loss. While this does not impact predictions when looking at games individually, it makes a big difference when filling out the bracket before the tournament begins.

e) KENPOM

As one might have noticed throughout this study, Ken Pomeroy was and continues to be an inspiration for many of these models. It seemed only right to research how Pomeroy predicts games, and test how it performs in relation to my models. After digging through the FAQ section and other various pages on the kenpom.com website, his method for predicting games using his statistics was found. As is the case for much of his research, his predicting model revolves around offensive and defensive efficiency, and tempo. First, in any given game, there is an offense going against a defense. While it may not be as obvious as in football or baseball, but offense and defense in basketball can be measured by points scored and points allowed. Pomeroy adjusts these numbers to convey how each team actually performed in every game. In order to predict a final score, a game efficiency must first be calculated. This translates to points per possession in this particular game. The difference between Team A's offensive efficiency and the national average, plus Team B's defensive efficiency and the national average, is added to the national average for offensive efficiency, and then divided by 100. This gives points per possession for this game. Pomeroy explains, “For example, let’s say that team A has an offensive efficiency (OE) of 120 and team B has a defensive efficiency (DE) of 120. Keeping our round number principle, I’ll use a national average for OE of 100. For a game between A and B, A’s offensive efficiency is expected to be 140. This is arrived at because both teams deviate from the norm by +20. So the sum of the deviations is 40, and that gets added to the nationwide average of 100” (Pomeroy, 2004). For the purposes of predicting the score in a given game, the offensive efficiency of 140 becomes 1.4 points per possession. This is the game efficiency for Team A and the first part of predicting a final score. The second part involves predicting the tempo of a given game, meaning how many possessions will each team have. The pace is determined using a similar method as the efficiency. The difference between Team A’s possessions and the national average, plus the difference between Team B’s possessions and the national average, are then added to the national average to find the game pace for this particular matchup. When the game efficiencies for each team are multiplied by the game pace, a final score is given. Points per possession multiplied by possessions produces points. This is a matchup model, and while it is not self-created, it is fascinating, and I will be experimenting with ways I can personalize in in the future.

I did, however, do extensive regular season testing on this model to analyze its validity. 100 games were the sample size of the test games that took place during the regular season, in the month of February. All but three of the games selected to be tested were within 10-point spreads, in order to try to predict the toughest games that could go either way. When strictly predicting the winner of the game, the model garnered a win percentage of 74%, which is incredibly impressive. When predicting games against the spread, the model was accurate at 55%. This is just enough to earn a profit. The general rule is that gamblers must be above 54% in order to turn a profit. This is because the sportsbooks take a percentage of all winnings. Finally, when analyzing games against the game total, predicting games to go over or under, the model was exactly at 50% wins and 50% losses. This is slightly below profitable, and quite surprising, considering how this model is designed, and its purpose. If this model continues to be this successful in the NCAA Tournament, then it will be tough for my models to beat.

f) 4Facts

Like the model KENPOM, this is not self-created, but rather the work of Rhonda Magel and Samuel Unruh from the statistics department at North Dakota State University. These models were discussed at length in the **Related Literature** section. They will serve as comparison models for the self-created models of this study.

5. Methodology:

This study has a three-pronged approach to how success is viewed for each of the models. The first stage will be filling out a bracket. Each model will predict the winners of the basketball games before the start of the tournament, and the results cannot be changed after the tournament starts. Success will be determined using the ESPN Tournament Challenge points structure, and bracket percentiles. A correct prediction in the first round is worth 10 points. A correct prediction in the second round is worth 20 points. A correct prediction in the Sweet 16 is worth 40 points. A correct prediction in the Elite 8 is worth 80 points. A correct prediction in the Final Four is worth 160 points, and finally a correct prediction in the National Championship is worth 320 points. There is a maximum of 320 points that can be won in each round of the tournament. The maximum number of points that can be earned is 1920. This would signify a perfect bracket. The bracket percentiles are how each bracket performs against all other brackets filled out on ESPN

Tournament Challenge. The higher the percentile, the better the bracket performed against the public.

The second stage will go on as the tournament progresses. I will run every game through the formulas and models to predict a winner. Therefore, even if I incorrectly predict the winner of a game in the bracket, I will still use my formulas to predict the game in the next round. For example: if Texas plays Oregon State first round, and one of my models predicts Oregon State will win, and they lose, I will still predict the winner of Texas and their next opponent. This is important because I will have a guaranteed sample size of 63 games. The bracket stage will show worse results because as the tournament goes on, it is harder to accurately predict the winning teams from the beginning.

The third and final stage of the project is analyzing the games against the odds set for sports gambling. This will include both “against the spread” and the “game total” odds. All odds were taken from ESPN Bet. The “against the spread” selections for each model are determined by the following process. The matchup models (XPoint, KENPOM, 4Facts PS) give a predicted final score or points spread. That spread is then compared to the actual points spread given by the sportsbook. There are three scenarios: A) the predicted spread is larger than the sportsbook spread, B) the predicted spread is less than the sportsbook spread, C) the underdog, according to the sportsbook, is predicted to win. In Scenario A, the favorite is predicted to cover the spread, this means that they will win by more points than what the sportsbooks predict. In Scenarios B and C, the underdog is predicted to cover the spread, meaning either the underdog will win the game, or they will lose by less points than the sportsbooks predict. The “game total” selections are chosen by whether or not the predicted final score from the models is over or under the sportsbooks’ line. All money used in this study is imaginary and serves the purpose of showing whether or not these models could be used as a part of a profitable system in sports gambling. This will show if the models have any uses outside of head-to-head predictions. This stage will not be used for every model, because they do not predict a final score or a point spread. Instead, these models produce power rankings, and the team with the higher number is determined as the winner. Stage 3 will only apply to XPoint, KENPOM, and 4Facts. Through these three stages, each model can be thoroughly and accurately analyzed for its success.

6. Results of Comparison Models:

This section is going to show the results of the comparison models for the 2024 NCAA Tournament so that they can be used as comparisons for the models of this study.

	KENPOM		2024
Stage 1:	Bracket Points		1230
Stage 1:	Bracket Percentile		93.80%
Stage 2:	Total Win/Loss		45-18
Stage 3:	ATS Win/Loss		30-33
Stage 3:	Game Total Win/Loss		26-37

Figure 3: KENPOM 2024 Results

	4Facts PS Model		2024
Stage 1:	Bracket Points		420
Stage 1:	Bracket Percentile		17.9%
Stage 2:	Total Win/Loss		35-28
Stage 3:	ATS Win/Loss		26-37

Figure 2: 4Facts PS 2024 Results

	4Facts LR Model		2024
Stage 1:	Bracket Points		700
Stage 1:	Bracket Percentile		53.5%
Stage 2:	Total Win/Loss		35-28

Figure 4: 4Facts LR 2024 Results

The results from the three comparison models can be seen on Figures 2, 3, and 4. The most successful, by far, of all of the comparison models was KENPOM, correctly predicting the winner in 45 of the possible 63 games in the tournament. It was also very successful in Stage 1 with 1,230 points in ESPN's Tournament Challenge. The next best of the three comparison models was the 4Facts Logistic Regression (LR) model. While using the same four statistics as the 4Facts Points Spread (PS) model, it was far more successful in Stage 1, outperforming it by 280 points. Interestingly, the two models predicted the same number of games correctly: 35. This was the lowest of any of the models analyzed for either 2023 or 2024.

7. Results and Analysis:

The following section is grouped by model. Each model will be analyzed for its results in 2023 and in 2024, and the comparison between the two years, before moving to the next model.

a) XPoint

Being the oldest of any of my models, there were high hopes for XPoint going into this study. It had been upgraded from the 2022 NCAA Tournament, and it performed well in regular season testing. XPoint continued that success by scoring very well in both years of research. Solely analyzing Stage 1 results, it was the third best of the four models, and yet it still performed very well: outperforming 75% of all brackets in both years, and over 90% of all brackets in 2023. This was a very encouraging sign. An interesting thing to note is that in 2023, the Final Four consisted

	XPoint	2023	2024
Stage 1:	Bracket Points	570	1050
Stage 1:	Bracket Percentile	90.3%	78.9%
Stage 2:	Total Win/Loss	43-20	42-21
Stage 3:	ATS Win/Loss	28-35	27-33
Stage 3:	Game Total Win/Loss	34-29	23-40

Figure 5: Two year results for XPoint

of four teams that very few people correctly predicted: Connecticut (4-seed), San Diego State (5-seed), University of Miami (5-seed), and Florida Atlantic (9-seed). Because ESPN Tournament Challenge's point system exponentially rewards wins as they get deeper into the tournament, the 2023 bracket points and percentiles are going to look way off from the 2024 results where there were two 1-seeds that made the Final Four and National Championship. In 2023, XPoint had only 570 points, yet it fell in the 91st percentile, when in 2024 it had 1,050 points, and yet it only fell in the 79th percentile. This trend is going to continue when analyzing the rest of the models as well. The percentiles for 2024 correspond to much higher bracket points due to more people correctly predicting the results in 2024.

Moving on to Stage 2, analyzing the models' predictions in each of the 63 games regardless of bracket predictions, XPoint performed the second-best of my models, correctly predicting the winner in 43 and 42 games in 2023 and 2024 respectively. These are encouraging results, as both years, XPoint was able to correctly predict at least two thirds of all of the tournament games correctly. Additionally, this is encouraging due to the differences in how the tournaments progressed. In 2023, zero 1-seeds made the Elite 8, while in 2024, two 1-seeds made the Elite 8, Final 4, and even the National Championship game. While all success is encouraging, it was expected that a matchup model, such as XPoint, would perform better than a power-ranking model. This is due to the fact that basketball, and sports in general are two-sided. Teams cannot play against no-one. The defense has an influence and an impact on how the offense performs. Therefore, some teams are going to matchup better against others. An example of this can be tempo. Some teams, such as Alabama and Kentucky, play at a very high pace, and thus they have more possessions per game, while other teams, such as Saint Mary's and Virginia, play at a very slow pace, limiting possessions per game. A team that typically plays slow, would probably prefer to play another team that plays slow, rather than a team that is going to play really fast. Conversely, a team that plays very fast, is not going to want to play against a team that plays slow, because it will prevent them from playing their preferred play style.

Finally, Stage 3 analyzes how XPoint performed in terms of sports gambling. The first aspect of Stage 3 is XPoint's results "against the spread." Described earlier in the **Methodology** section, the selections made for XPoint in this stage are comparisons to the lines set by sportsbooks. The success of XPoint in Stages 1 and 2 was not replicated in Stage 3. XPoint was consistent however, correctly predicting 28 and 27 games in the past two years respectively. While XPoint's results against the spread were less than ideal, it was still able to outperform 4Facts PS model, which only correctly predicted the winner of the game against the spread 26 out 63 times in 2024. It did however not perform as well as KENPOM against the spread which correctly predicted 30 games correctly in 2024. One of the reasons that XPoint struggled so much against the spread is that it still, after years of tinkering the formula, undervalues a team's strength of schedule, or more accurately in this case, strength of statistic. For example, a team that makes 35% of their three pointers against teams from the BIG 10, traditionally one of the strongest conferences in the country, is probably a better 3-point shooting team than a team that makes 35% of their three-pointers against teams from the NEC, traditionally one of the weakest conferences in the country. More on this in a later section, but moving forward, there needs to be a stronger adjustment for strength of statistic. In the second half of Stage 3, analyzing XPoint's success on the "game total," the results were mixed. The two-year record was not good: 57-69. However, in 2023, the model correctly predicted 34 games correctly, and it was in 2024, that the model performed poorly, only correctly predicting 23 games. When looking at this as a whole, it is slightly confusing how the model could perform so well one year, and so poorly the next.

When looking at the data even closer, the picture becomes even more confusing. In 2023, the model predicted 54 out of 63 games to go “Under” the game total. That alone, is surprising, but when compared to 2024, it becomes even more of a surprise. In 2024, the model predicted only 18 out of 63 games to go “Under.” Normally, there would never be this large of a disparity in statistical bias from year to year. After seeing the results and predictions from 2023, most people would have expected XPoint to predict a majority of the games to go “under” the game total, but the opposite was the case. The model was not changed from 2023 to 2024, which makes this discrepancy even more of a mystery. When compared against its main competitor: KENPOM, it was less successful in correctly predicting the game total to go over or under, XPoint’s 23 correct predictions versus KENPOM’s 26.

Some final noteworthy points about XPoint over the course of these past two years include the mystery of McNeese State, an underdog bias against the spread, and a bias towards strong offenses. Starting with McNeese St, this team had almost no impact on the 2024 NCAA Tournament, and yet it will be a focus heading into further research. The reason this team will be analyzed so closely is that XPoint predicted that McNeese State would make the championship game before bowing out to the champion: University of Connecticut, which was correctly predicted as champion. McNeese State was a 12-seed that lost to Gonzaga by 21 points in the first round of the tournament. After initially filling out the bracket, I thought for sure there was a mistake in their data, yet when I looked back and checked it, each individual statistic was accurate, and no mistake had been made in any of the formulas or functions along the way. The next step was looking at the statistics themselves, what was it that made McNeese State so special that they were going to make the championship game as a 12-seed. After combing through the data, there was no obvious answer, as to why XPoint favored them so much. They were a top 10 team in 3-point shooting percentage, but they were bottom 25 in the country in 3-point shots allowed per game. However, when looking at the bracket as a whole, it was clear that XPoint overvalues the 3-point shooting percentage. There were eight teams that made the 2024 NCAA Tournament that were also in the top 25 in the country in 3-point shooting percentage. Of those eight teams, seven were predicted to make the Sweet 16, and the only one that was not predicted to make the Sweet 16, Dayton, was predicted to lose to another top eight team in 3-point shooting percentage, Arizona. This is a trend that is going to have to carefully analyzed moving forward, and it is possible that the importance of 3-point shooting percentage may have to be lowered in the future.

Similar to the bias of 3-point shooting percentage, there seems to be bias towards teams with strong offenses. Solely looking at offensive efficiency, five of the top eight teams were predicted to make the Elite 8 in XPoint in 2024. If a simple statistic such as points per game is used instead of offensive efficiency, all 8 teams in the Elite 8 of XPoint's bracket were in the top 50 in the country in 2024. Looking at 2023, the same trend rings true. All 8 teams in XPoint's Elite 8 were in the top 50 in the country in points per game. This is very fascinating, as this is a new connection that had gone unnoticed in years past. Going forward, the importance of defense may have to be increased.

Finally, there is certainly an underdog bias when applying XPoint to make predictions against the spread. As previously stated, XPoint was not created with the intention of having success in terms of sports gambling, but there is no reason that it cannot be applied to it. However, as shown above, there was very little success in the past two years when applied to sports gambling, particularly against the spread. One of the reasons for this is that strength of statistic is very difficult to measure. As described when discussing the Stage 3 results, the regular season competition matters when looking at raw, unadjusted statistics. This is one reason that Ken Pomeroy has had such success: his efficiency numbers are adjusted for the opponent. Moving forward, strength of statistic is going to be the main focus for improving XPoint.

Overall, XPoint had very promising results in two of the three stages. XPoint was not created with the intention of having success in sports gambling, but rather it is an attempt to accurately predict how different play styles can create different results based on matchups. KENPOM is another statistician's take on this same principle, and its results will be discussed later in this section. The next two models are both power ranking models, and thus Stage 3 does not apply.

b) AFT

As explained in the **Models** section, AFT is short for Assists, Free throws, and Three-pointers, the three main components of the model. Additionally, AFT is a power ranking model, where the best teams typically have a score of above 40, and the worst teams have a score close to 0. This includes the adjustment required on the 2023 data due to the lack of the Sagarin strength of schedule metric in 2024. The highest score recorded in the two years was 2024 Purdue. Interestingly, Purdue was the highest ranked team in both years of data. This is probably due to the fact that Purdue had one of, if not the best, player in college basketball, Zach Edey. With his towering size, seven-foot-four-inches, he was able to limit opposing teams' rebounds, which is quite advantageous for this model. As opposing team rebounds decreases, a teams' AFT score increases. Furthermore, Purdue played the 2nd hardest schedule in 2024 which leads to a higher AFT score.

The average score is 2023 was 19.75, and the average score in 2024 was 21.23. The main reason the average was higher in 2024 was because Purdue had a ridiculous score of 51.96. No team was within ten points of this score in either year of research. The standard deviation was 8.06 and 9.13 in the two years respectively.

In Stage 1, AFT had two successful years in brackets, earning 810 points in 2023, and 980 points in 2024. Unfortunately, due to an error in the input of the bracket onto ESPN's Tournament Challenge, there is no percentile data for 2023. The percentile for 2024 was 74.6%. The error was noticed when looking back at the 2023 submission into ESPN's Tournament Challenge for AFT and realizing games were entered incorrectly. The correct results were inputted post-factum, and 810 points is the accurate score for the correct AFT bracket that should have been submitted. In the interest of full transparency, the incorrect submission for AFT was created using Upset Indicators to help predict the bracket. The bracket should have been submitted purely using data and predictions from the AFT model. The incorrect bracket submission scored 740 points and was in the 97th percentile of all brackets submitted on ESPN. Therefore, the correct bracket submission for AFT would have been in the at least the top 3% of all brackets submitted, seeing as its score is 70 points higher than the incorrect bracket submission.

The success of AFT in 2023 is very surprising considering that the predicted champion, Purdue, lost first round, becoming only the second 1-seed ever to lose in the first round of the NCAA Tournament. As previously noted, the tournament was much more unpredictable in 2023, meaning losing its predicted champion in the first round had less of an impact than in other years. Compared to XPoint, AFT had a 100-point advantage when combining the two years,

	AFT	2023	2024
Stage 1:	Bracket Points	810	980
Stage 1:	Bracket Percentile	---	74.6%
Stage 2:	Total Win/Loss	44-19	39-24

Figure 6: Two year results for AFT

1790 vs 1620, and AFT ranked better compared to the public over the two-year stretch.

Stage 2 was a mixed bag of results. 2023 AFT tied for the most wins of the models that year with 44, however 2024 AFT had the fewest wins of my models with only 39 wins. It is interesting that this model could perform well the first year, and not in the second year. What this data shows is that either AFT performs better when there is more unpredictability in the tournament, or the statistics used in AFT were more important in determining the winner of games in 2023 than in 2024. Trends can change from year to year, and that seems to be the case in this situation.

Overall, AFT can be seen as a success. While its champion was eliminated in the first round in 2023, the model performed exceptionally well in spite of that. Truthfully, that loss by the 1-seed Purdue against the 16-seed Fairleigh Dickinson can be seen as an anomaly, due to the fact that this is only the second time in NCAA Tournament history a 1-seed has lost in the first round of the tournament.

c) ZSCORE

The following model is another power-ranking model that was created with hopes of finding a well-rounded team that can win the championship. ZSCORE uses the sum of 5 z-scores in five different statistics, and the higher the score the better the team is perceived to be. The range of the scores in the two years is -6.72 (2023 Texas Southern) to 5.35 (2024 Arizona). The average score for both years is zero, as it should be by the definition of z-scores. As shown in Figure 7, the data for the two years is very similar, and almost identical. This is not necessarily surprising considering z-scores by definition are trying to compare statistics on the same scale.

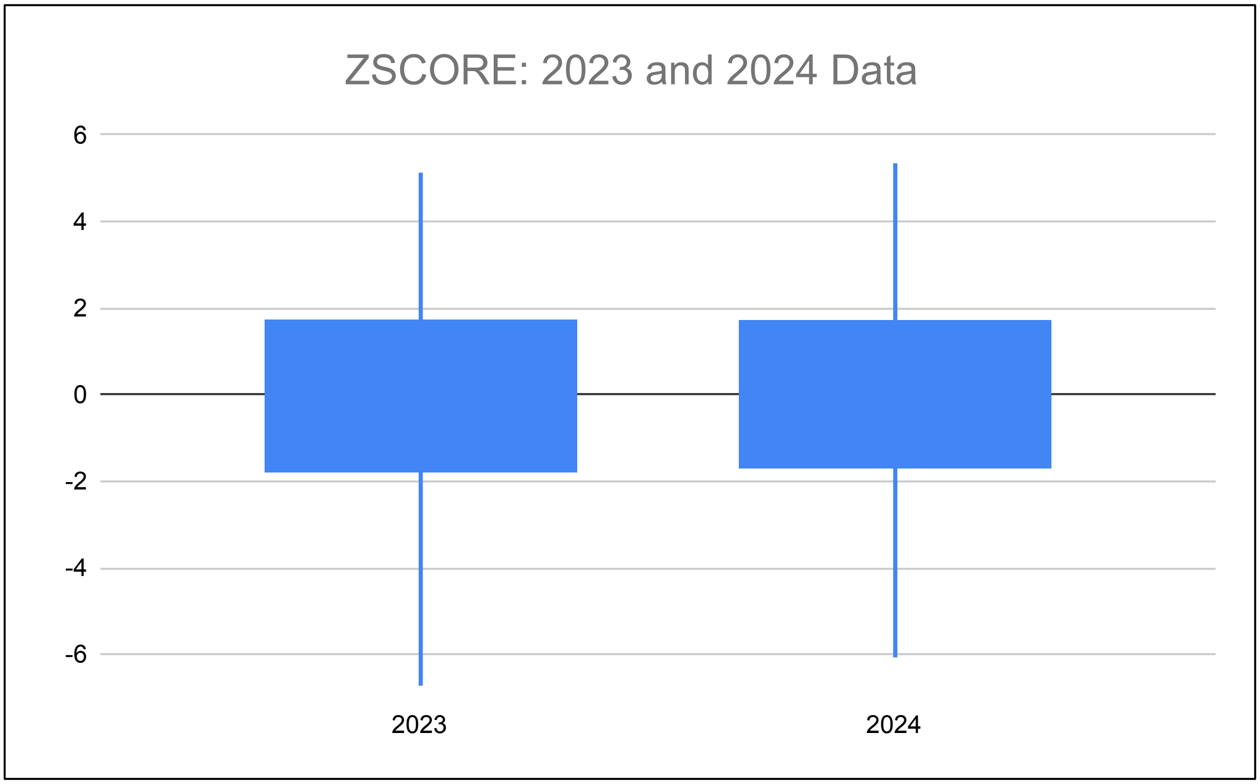

Figure 7: Statistical Distribution for ZSCORE by year

In Stage 1, the results were very different. ZSCORE had the biggest disparity between the two years of data. In 2023, it earned 630 points, while in 2024, it only earned 520 points. This was the only model to score less points in 2024 than 2023. The 2023 bracket fell in the 94th percentile, while the 2024 submission fell all the way down in the 31st percentile. This was a shocking revelation, and it was certainly unexpected. Over the two years, ZSCORE scored the fewest points of the models only scoring 1150 points. What is even more surprising is how it performed in Stage 2, because despite the poor results in 2024, it still managed to correctly predict 41 out of 63 games correctly in both years. While still the lowest two-year total of any of the models, it is above 65% which is still promising results. And, in spite of the poor bracket results in 2024, it managed to predict more games correctly than AFT.

The stark shift is success from 2023 to 2024 was very surprising, and it is hard to pinpoint the exact cause of this decline. The main theory is that simply the statistics used in 2023 were better indicators of success than they were in 2024. It is likely that the success in 2023 was coincidental, and the statistics used in this model, or the model itself, have no real predictive ability. Having only one out of the four models be classified as a bust is still a very good thing overall. However, ZSCORE cannot be scrapped completely because it is one third of the most successful model in this study: COMBO.

d) COMBO

As previously discussed, COMBO is the first aggregate model that I have created, and it saw immediate success. The main difference of COMBO is that it is not actually a single model, but the combination of the three previous models: XPoint, AFT, and ZSCORE. The theory behind this was that each team is going to have strengths in different models due to the different statistics used. Furthermore, teams are going to be ranked differently in AFT and ZSCORE, the two power-rankings models. Finally, XPoint provides the essential tiebreaker, looking at the actual matchup between the two teams, and seeing how two different play styles will clash. Whichever team is predicted to win in at least two of the three models is selected to win the game. When using COMBO to create a bracket, it was combined with Upset Indicators, a criteria-based model to identify teams that are likely to lose in the first two round of the NCAA Tournament. Once again, Upset Indicators overrules any predictions made by COMBO for the purpose of filling out a bracket, thus the Stage 1 results. However, when analyzing the wins and losses of COMBO, in this study: Stage 2, Upset Indicators was not factored in. Stage 2 was analyzed like this in order to purely look at the results of COMBO without any outside influences.

In Stage 1, COMBO was far and away the best of the models in both years of research. With a combined score of 2550 points over the two years, it scored over 800 points higher than the second-best model. Not only was it the only model to score over 1,000 points in both years of the ESPN Tournament Challenge, but it was the only model to correctly pick the national champion in both years. To add to its success, COMBO scored in the top 2% of all ESPN brackets in 2024, while in 2023, it finished in the top 12,000, which is within the top 0.1%.

	COMBO	2023	2024
Stage 1:	Bracket Points	1160	1390
Stage 1:	Bracket Percentile	99.9%	98.4%
Stage 2:	Total Win/Loss	44-19	43-20

Figure 8: Two year results for COMBO

The Stage 2 results were just as encouraging, bringing in the most correct picks of any of the models over the two years: 87 wins and 39 losses. In 2023 it tied for the most wins of any of my models with AFT at 44 wins. In 2024, it outperformed all of my other models, but had two less wins than KENPOM, 43 versus 45 wins respectively. Ken Pomeroy created an excellent model for predicting the winner of games. However, very rarely does his model predict upsets, in fact the KENPOM bracket, while successful: 1,230 points, it rarely predicts upsets to occur. I consider this to be a strength of XPoint, that it does not always select the perceived better team to win every game. And while that does produce a bit of a wild prediction every now and then: see McNeese State in 2024, I believe that XPoint is better for COMBO than KENPOM would be. In fact, had the COMBO bracket been submitted using KENPOM instead of XPoint as the matchup model in 2024, it would have scored only 1,350 points, which is forty points lower than the original. This was a very encouraging result to see, as it was affirmation that XPoint does have merit as a model, and it is a vital part of COMBO.

The overall results over the two years for each of the models were very promising. COMBO was far and away the best model, which was not entirely surprising due to the fact it is comprised of all three of the other models from this study. The second-best model was split between XPoint and AFT, the former had a better two-year record and correctly predicted the champion in one of the years, while the latter had a better combined score in the two-years of brackets. ZSCORE was the least successful of the models of this study, while still having decent results. All four models outperformed the 4Facts models from Magel and Unruh, while only COMBO was able to outperform KENPOM in the bracket. KENPOM outperformed all models in predicting the games in Stage 2. Pomeroy's model for predicting games, while very accurate, does not predict many upsets, and thus would most likely not perform very well in a tournament with lots of upsets and unpredictability. Just as they have before, each of the models will be looked at during the offseason, and there will be improvements and adjustments in order to try to make them more accurate.

8) Improvements for the Future

These two years of research has led to many conclusions, as well as shed light on many things that can be improved. This section is dedicated to improvements that will be made before the next NCAA Tournament, and also first opinions on possible new models going forward.

The first improvement moving forward takes place in XPoint, and it involves altering the estimation of shots per game. In the current version of XPoint, a team's shots per game against a given opponent are estimated by averaging the shots per game a team attempts and the shots per game the opponent allows. However, after extensive research, and analysis for how Ken Pomeroy estimates a game's pace, a change will be made. Similar to how Pomeroy estimates the possessions in a given game, XPoint will now use the same principle. The average shots per game of the 68 tournament teams will be taken, in each of the three categories: free throws, two-pointers, three-pointers, as well as the average shots allowed per game for the 68 tournament teams in the same three categories. When calculating free throws per game for Team A against Team B, the difference of Team A's free throw attempts and the average free throws attempted per game is taken. Secondly, the difference of Team B's free throw attempts allowed, and the average free throw attempts allowed is taken. These two are added together and added to the average free throw attempts per game. This will give the average free throw attempts per game for Team A. The same process is completed for both teams, for two-pointers and three-pointers as well. This method uses the same methodology that Pomeroy uses, although it focuses on the individual statistic in the matchup rather than one statistic that will be the same for both teams: possessions per game. This improvement should create more accurate shot attempt predictions based on matchups.

The second improvement also is for XPoint, and this one should hopefully address how poorly it performed on predicting game totals. As previously discussed, XPoint was not accurate in correctly predicting the games to over or under in 2024, with a win percentage of only 36.5%. However, in analyzing the final scores of the tournament games, and comparing with tournament teams' points per game leading up to the tournament, there was a clear discrepancy. Using the last four tournaments, teams have scored 6.74% less per game in the tournament than in the regular season. The average points per game of tournament teams prior to the NCAA Tournament from 2021 to 2024 was 75.08 points. The average points per game in the tournament over the same span, was only 70.02 points per game. Additionally, when the data is analyzed yearly, as opposed to the average of the four years, the same trend exists.

	Points Per Game			
Year	**tourn**	**reg sea**	**delta**	**%change**
2024	72.54	76.92	-4.39	-5.70%
2023	69.52	74.98	-5.46	-7.28%
2022	68.87	73.99	-5.13	-6.93%
2021	69.14	74.41	-5.27	-7.08%
Average	70.02	75.08	-5.06	-6.74%

Figure 9: Points per Game in Tournament vs Regular Season

As shown in Figure 9, each year has had at least a five and a half percent decrease in scoring from the regular season to the tournament. The average decrease in points over the course of the past four years is 5.06 points per game. This equates to 6.74%. This trend can be attributed to a number of factors: the pressure of playing in a NCAA Tournament game, playing in a new stadium away from their home court, teams playing with more effort defensively due to the stakes of the game, and others. Whatever the case may be, the trend exists, and thus adjustments need to be made moving forward. In the 2025 NCAA Tournament, each team will have an adjustment of a 6.74% decrease to their predicted final score in XPoint. The hope is that this can make XPoint more accurate when predicting the final score of games, and thus perform better on game totals.

The next improvement for future research is a new model. This model comes after watching some of the upsets take place in the 2024 Tournament. There is a saying in sports that goes: “Live by the sword, die by the sword.” This applies to when a team relies on one thing so much, that when they are having an off night in that area, they are in deep trouble and have trouble winning. In the case of basketball, that most often applies to shooting, in particular, 3-point shooting. Teams like BYU, Arizona, and Kentucky all relied heavily on the three-point shot. BYU in particular was second in the country in percent of points from three-pointers. 40% of their points per game came from threes. Furthermore, when they shot below 30% from the three-point line, they lost every game: 0 wins and 7 losses. When BYU shot 33.3% or less, making one out of every three attempts, they had 3 wins and 9 losses. In their first round NCAA Tournament game against Duquesne, they shot 33.3% from three, and was upset by 11-seed Duquesne. BYU is the perfect example of dying by the sword in the NCAA Tournament. The new model will be called SWORD, and while the exact details are not finalized, it will involve three-point shooting percentages, the volatility of a team's shooting percentages, and other factors to be decided. The same type of model can be constructed for any statistic in theory. The early thought process involves using the standard deviation of that team's game-by-game statistics. This can be applied to any number of statistics, although the current belief is that an off-night in three-point shooting has the biggest impact on determining the outcome of the game. Other statistics that this type of model might be applied to is rebounding, turnovers, and free throws per game. Turnovers in particular is something that a lot of teams struggle with when it comes tournament time. Teams that are prone to turnovers can find themselves in trouble when the pressure is on. While most tournament teams have relatively low turnover per possession rates, these numbers can be exacerbated when playing aggressive defenses at an away from home location with the season on the line. An example of this is the first-round matchup between the 5-seed Saint Mary's Gaels and the 12-seed Grand Canyon Antelopes. While the Gaels were 54th in the country at turnovers per game with 10.3, they were all the way down at 99th nationally in turnovers per possession at 15.6%. This shows that they are susceptible to turnovers because they are also the 5th slowest team in the country. To the untrained eye, someone may say that Saint Mary's only has 10 turnovers a game, and thus they are not in danger of losing games due to turnovers. However, when comparing it to their turnovers per possession, and factoring in their slow pace of play, one can see that this is not the case. In their first-round matchup against Grand Canyon, a team that is ranked 43^{rd} in turnovers forced per possession, the Gaels had 13 turnovers, which led to 17 points for the Antelopes, an efficiency of 1.31 points per turnover forced which is an incredible rate. Four of Saint Mary's seven losses going into the NCAA Tournament occurred in which they record above their season average in turnovers.

Another possible new model moving forward has to do with matchups and categories. This is a model that I have attempted in other sports, with moderate success. It involves putting teams into specific categories based upon their play style and abilities. This model was originally built for the English Premier League. Teams were put into categories that had predetermined advantages built in against other categories. The same will be true for basketball. For example, a strong defensive team that rebounds well, is a bad matchup for a team that typically plays a smaller lineup and relies on their shooting. While the categories have not been determined yet, the early choices for categories would be "Sharp Shooters," which will be for teams that rely heavily on the three-point shot. Another category will be "Bruisers," which will encompass teams that rebound well, but maybe tend to foul a lot. "Roadrunners" can apply to teams that play really fast and have lots of possessions per game. The idea is that certain categories of teams will matchup better against other categories, and thus provide an advantage. There will certainly be more categories as basketball teams are complex and there are a lot of different identities that teams can take on.

9) Other Applications

This section is dedicated to other applications for the research and findings from this study. While this study has been dedicated to predicting the winners of games in the NCAA Tournament, there can be other applications for this research. As discussed throughout the study so far, there are obviously sports gambling applications for this research when picking games against the spread or on the game total. Furthermore, with power ranking models such as AFT or ZSCORE, they can give some predictions or selections for "future" bets. These are bets that one can place for the future, and in most cases, they apply to picking the winner of the tournament or league before the championship game. In the case of the NCAA Tournament, there are also Final Four future bets, and others. COMBO predicted the correct winner of the NCAA Tournament in each of the two years of its existence. This shows that there is some value in using these models for future bets. While the Stage 3 results throughout this study were not exactly the most positive,

Outside of the sports gambling sector, there are definitely applications in coaching. While some takeaways from this research are obvious, such as the better statistics a team has, the better they will perform in models, others are more nuanced. One takeaway from XPoint, is that limiting three-point attempts can prove successful. In today's game, this can prove especially difficult. However, if a team has a low number of three-point attempts allowed per game, they are taking away the valuable scoring opportunities from their opponent.

An interesting concept is points per attempt. There are obviously three ways to score points in a basketball game: free throws, two-pointers, and three-pointers. But to a coach, it is important to know which is the most efficient. The way to find this is by calculating the points per attempt in each of the three categories. In the 2024 NCAA Basketball season, the average points per attempt were 0.718 for a single free throw attempt, 1.018 for two-pointers, and 1.016 for three-pointers. This is fascinating because some might assume that three-pointers would be the most valuable, but actually the most efficient use of a possession is two attempt a two-point basket. Albeit the difference between the two-pointer and the three-pointer is very small, only two thousandths of a point. Therefore, pure analytics would say that teams should attempt two-pointers at a higher rate than three-pointers. However, this is the NCAA average, teams like Northwestern, Howard, and Virginia, this is definitely the wrong strategy. These three teams all made the NCAA Tournament while scoring more than one point per attempt on three-pointers and scoring less than one point per two-point attempt. For these three teams, it would be better to attempt more three-point shots than two-point shots. Using Northwestern in particular, they averaged 36.6 two-point shot attempts per game and 21 three-point shot attempts per game. Their points per attempt respectively were 0.985 and 1.169. They average 0.184 more points per attempt on three-pointers. Thus, the coach of Northwestern should tell his team to attempt more three-pointers per game. On the other hand, BYU is an interesting case. As said earlier when discussing the future model: SWORD, over 40% of their points per game comes from three-pointers. However, they are 0.112 points more efficient on two-point attempts than three-point attempts. BYU has the 21st biggest difference between two-pointers and three-pointers in the country. Using this knowledge, along with their propensity to lose games when shooting poorly from the three-point line, BYU's coaches should preach more two-point attempts per game, and less three-point attempts.

Using statistics and analytics in this way can greatly enhance a team's chances to win games and increase their efficiency. The models in this study probably serve more as indicators of what the team is good at, instead of how they can improve. This is because the models were designed with the intention of accurately predicting future games, rather than how to become a better team.

10) Conclusion

Basketball is an amazing sport because there are so many things that go into determining the winner of a game. And when the season is on the line, and the athletes on these teams are competing for what could be the last time in their college careers, truly anything can happen. That is what makes the NCAA Tournament so special. As shown by the research from this study, it is incredibly difficult to only use statistics to create a model that accurately predicts college basketball games. At the end of the day, the players on the court that actually determine the outcome of the game are human. And humans are complex and creative, and they do not always live up or play down to expectations. Sports are so special and important because of their unpredictability. They would be a lot more boring, and less people would watch them if the outcome was pre-determined, and the higher ranked team won every game. That unpredictability is what separates sports from other forms of entertainment. It is also what makes predicting the winners of games so difficult. There are so many factors that cannot be measured by statistics that can influence the outcome of NCAA basketball games. A few examples are the players' class schedules and workload, the travel distance to the arena for the tournament games, and finally the mental state of the players themselves. These factors can make predicting games very difficult. With that being said, each of the models contributed something, whether it was something that did not work, or it was an amazing success. COMBO was incredibly successful, and it will continue to be used moving forward. The three models that makeup COMBO each had their own merits. XPoint, after years of adjustments and perceived improvements, had two successful years of results, albeit not necessarily in the world of sports gambling. It faces another adjustment this offseason that should hopefully remedy its woes in that realm. AFT was another successful model, performing strongly in each of the two stages of its analysis. After its strength of schedule improvement due to the lack of data moving forward, it should remain fairly the same moving forward, as it has proved an important aspect of COMBO. Finally, although ZSCORE did not perform well on its own, it was valuable as a part of COMBO. Thus, ZSCORE will continue to be used, most likely solely for its contributions to COMBO. I am looking forward to the improvements for each of the current models, as well as the models that are to come next year and in the future. The quest for the perfect bracket remains arduous, and while the destination of 63 wins in a row seems impossible to reach, the journey has, and will continue to be, a lot of fun.

References:

Fearnhead, P., & Taylor, B. M. (2010). Calculating strength of schedule, and choosing teams for March Madness. *The American Statistician*, *64*(2), 108-115.

Karnibad, J (2023). Is there a Method to March Madness? *Mount St. Mary's SPARQ Festival.* Click here to download

Lopez, M. & Matthews, G. (2015). Building an NCAA men's basketball predictive model and quantifying its success. *Journal of Quantitative Analysis in Sports*, *11*(1), 5-12. https://doi.org/10.1515/jqas-2014-0058

Magel, R., & Unruh, S. (2013). *Determining factors influencing the outcome of college basketball games*. SCIRP. https://www.scirp.org/journal/PaperInformation?PaperID=35927

Pomeroy, K. (2004, December 27). *National efficiency: The kenpom.com blog*. KenPom.com. https://kenpom.com/blog/national-efficiency/

Wilco, D. (2024, March 5). *The absurd odds of a perfect NCAA bracket*. NCAA.com. https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-16/perfect-ncaa-bracket-absurd-odds-march-madness-dream