

# **Application: WHSDSC25-2728026835**

Rebecca Tees - rtees@mcdonogh.org

2025 Wharton High School Data Science Competition: Basketball Tournament Predictions

## **Summary**

**ID:** WHSDSC25-2728026835

**Status:** Your team has been eliminated. Thank you for participating!

**Last submitted:** Jan 31 2025 08:35 AM (EST)

## **Phase 1: Women's College Basketball Playoff Game Predictions**

Completed - Mar 3 2025

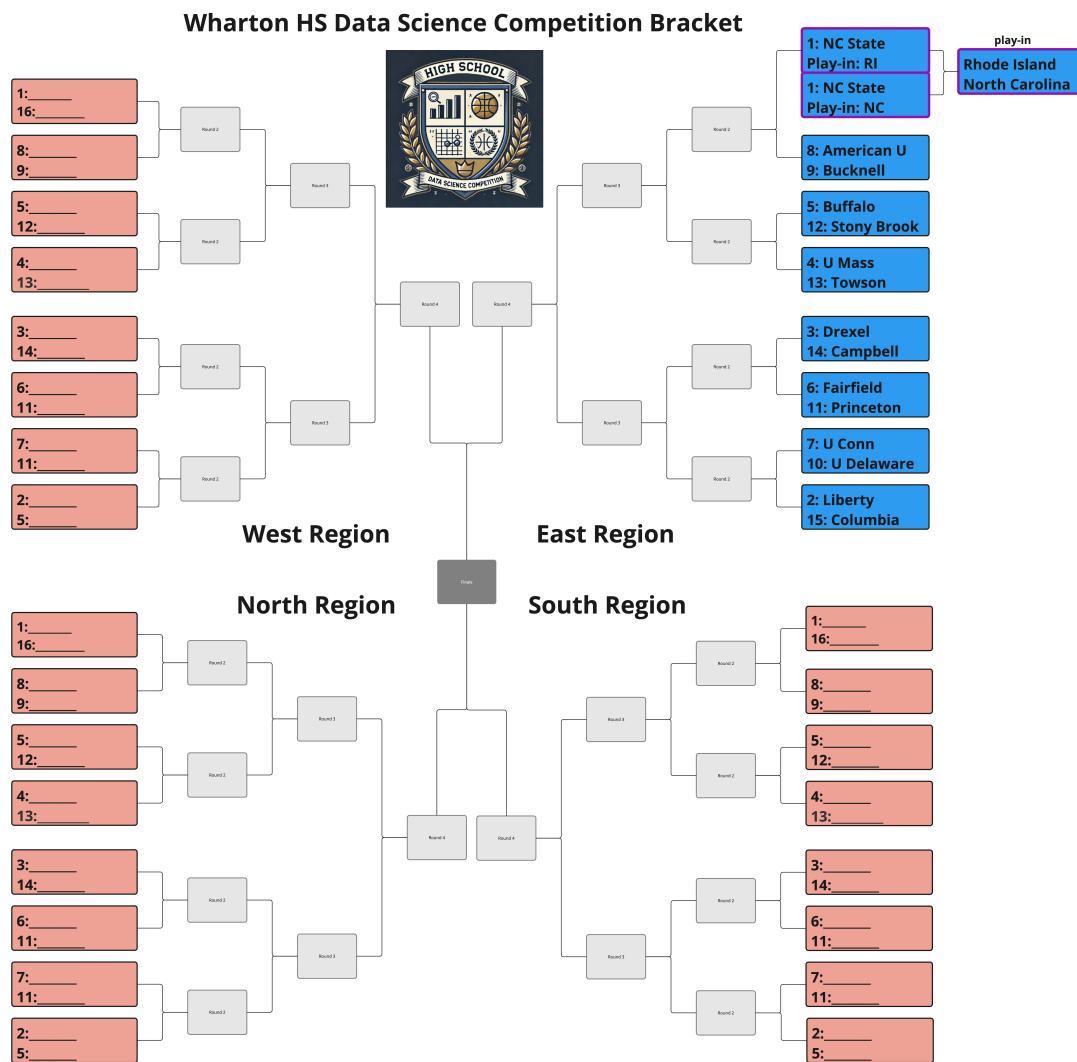
## **Phase 1: Basketball Tournament Game Predictions**

Phase 1: Basketball Tournament Game Predictions

# Phase 1: Basketball Tournament Game Predictions

You're the analytics staff for a basketball team, the group behind the numbers that drive performance. It's the offseason, and your task is clear: analyze the previous season to uncover trends, rethink strategies, and predict future success. You and your team have been hard at work analyzing and using the data to come up with the best answers. Now it's time to share your predictions!

The bracket below summarizes the Main Competition tasks.



**1a:** Your analytics staff's first challenge is to rank the top 16 teams in three of four regions—West, North, and South—based on their regular season performance.

**1b:** You'll predict the winning probabilities for 10 first-round games in the East Region, including a play-in game.

**1c:** Tell your story and summarize your methodology. Craft a concise but comprehensive explanation of your approach.

You and your team have been hard at work analyzing and using the data to come up with the best answers. Now it's time to share your predictions!

## Phase 1a: Ranking the Teams

Your analytics staff's first challenge is to rank the top 16 teams in three of four regions—West, North, and South—based on their regular season performance. But this isn't just about win-loss records. Like the Coaches Polls, your rankings should reflect the overall strength and quality of the teams.

To help you get started, each team has already been assigned to one of four regions (East, West, North, South).

Note: this setup differs from the actual NCAA tournament to make your work easier and more streamlined.

Now, it's time to make your picks—who are the top contenders?

For each of the three regions (West, North, and South), you will first select the top 16 teams in any order. Then, you will rank the 16 selected teams based on your predictions.

## West Tournament Rankings

### Rank Your Top Teams From WEST

Based on your analyses, select the teams you predict will rank in the top 16 of the WEST Region. [Rank the 16 teams, with 1st place being the “best” team.](#) The team ranking should reflect the overall quality of the team. Be sure to double-check that the 16 teams are the ones you wish to rank and that there are no duplicate teams! Teams must be listed in their correct ranking order.

---

Select the team you predict will rank #1 (first place) in the WEST Region.

Stanford Cardinal

**Select the team you predict will rank #2 (second place) in the WEST Region.**

Baylor Bears

---

**Select the team you predict will rank #3 (third place) in the WEST Region.**

Gonzaga Bulldogs

---

**Select the team you predict will rank #4 (fourth place) in the WEST Region.**

Texas Longhorns

---

**Select the team you predict will rank #5 (fifth place) in the WEST Region.**

BYU Cougars

---

**Select the team you predict will rank #6 (sixth place) in the WEST Region.**

Colorado Buffaloes

---

**Select the team you predict will rank #7 (seventh place) in the WEST Region.**

Nebraska Cornhuskers

---

**Select the team you predict will rank #8 (eighth place) in the WEST Region.**

South Dakota State Jackrabbits

---

**Select the team you predict will rank #9 (ninth place) in the WEST Region.**

Oklahoma Sooners

---

**Select the team you predict will rank #10 (tenth place) in the WEST Region.**

UT Arlington Mavericks

---

**Select the team you predict will rank #11 (eleventh place) in the WEST Region.**

Utah Utes

---

**Select the team you predict will rank #12 (twelfth place) in the WEST Region.**

Washington State Cougars

---

**Select the team you predict will rank #13 (thirteenth place) in the WEST Region.**

Creighton Bluejays

---

Select the team you predict will rank #14 (fourteenth place) in the WEST Region.

UNLV Lady Rebels

Select the team you predict will rank #15 (fifteenth place) in the WEST Region.

San Francisco Dons

Select the team you predict will rank #16 (sixteenth place) in the WEST Region.

California Baptist Lancers

## North Tournament Rankings

Select Your Top Teams From the NORTH

Based on your analyses, select the teams you predict will rank in the top 16 of the NORTH Region. Rank the 16 teams, with 1st place being the “best” team. The team ranking should reflect the overall quality of the team. Be sure to double-check that the 16 teams are the ones you wish to rank and that there are no duplicate teams! Teams must be listed in their correct ranking order.

Select the team you predict will rank #1 (first place) in the NORTH Region.

South Carolina Gamecocks

**Select the team you predict will rank #2 (second place) in the NORTH Region.**

Florida Gulf Coast Eagles

---

**Select the team you predict will rank #3 (third place) in the NORTH Region.**

UCF Knights

---

**Select the team you predict will rank #4 (fourth place) in the NORTH Region.**

LSU Tigers

---

**Select the team you predict will rank #5 (fifth place) in the NORTH Region.**

Belmont Bruins

---

**Select the team you predict will rank #6 (sixth place) in the NORTH Region.**

South Florida Bulls

---

**Select the team you predict will rank #7 (seventh place) in the NORTH Region.**

Ole Miss Rebels

---

**Select the team you predict will rank #8 (eighth place) in the NORTH Region.**

Stephen F Austin Ladyjacks

---

**Select the team you predict will rank #9 (ninth place) in the NORTH Region.**

Georgia Lady Bulldogs

---

**Select the team you predict will rank #10 (tenth place) in the NORTH Region.**

Middle Tennessee Blue Raiders

---

**Select the team you predict will rank #11 (eleventh place) in the NORTH Region.**

Tennessee Lady Volunteers

---

**Select the team you predict will rank #12 (twelfth place) in the NORTH Region.**

Florida State Seminoles

---

**Select the team you predict will rank #13 (thirteenth place) in the NORTH Region.**

Florida Gators

---

Select the team you predict will rank #14 (fourteenth place) in the NORTH Region.

Stetson Hatters

Select the team you predict will rank #15 (fifteenth place) in the NORTH Region.

Jackson State Lady Tigers

Select the team you predict will rank #16 (sixteenth place) in the NORTH Region.

Troy Trojans

## South Tournament Rankings

Select Your Top Teams From the SOUTH

Based on your analyses, select the teams you predict will rank in the top 16 of the SOUTH Region. Rank the 16 teams, with 1st place being the “best” team. The team ranking should reflect the overall quality of the team. Be sure to double-check that the 16 teams are the ones you wish to rank and that there are no duplicate teams! Teams must be listed in their correct ranking order.

Select the team you predict will rank #1 (first place) in the SOUTH Region.

Iowa Hawkeyes

**Select the team you predict will rank #2 (second place) in the SOUTH Region.**

Louisville Cardinals

---

**Select the team you predict will rank #3 (third place) in the SOUTH Region.**

Iowa State Cyclones

---

**Select the team you predict will rank #4 (fourth place) in the SOUTH Region.**

Kentucky Wildcats

---

**Select the team you predict will rank #5 (fifth place) in the SOUTH Region.**

IU Indianapolis Jaguars

---

**Select the team you predict will rank #6 (sixth place) in the SOUTH Region.**

Indiana Hoosiers

---

**Select the team you predict will rank #7 (seventh place) in the SOUTH Region.**

Michigan Wolverines

---

**Select the team you predict will rank #8 (eighth place) in the SOUTH Region.**

Dayton Flyers

---

**Select the team you predict will rank #9 (ninth place) in the SOUTH Region.**

Ohio State Buckeyes

---

**Select the team you predict will rank #10 (tenth place) in the SOUTH Region.**

Virginia Tech Hokies

---

**Select the team you predict will rank #11 (eleventh place) in the SOUTH Region.**

Notre Dame Fighting Irish

---

**Select the team you predict will rank #12 (twelfth place) in the SOUTH Region.**

Murray State Racers

---

**Select the team you predict will rank #13 (thirteenth place) in the SOUTH Region.**

Toledo Rockets

---

Select the team you predict will rank #14 (fourteenth place) in the SOUTH Region.

Missouri State Lady Bears

Select the team you predict will rank #15 (fifteenth place) in the SOUTH Region.

Cleveland State Vikings

Select the team you predict will rank #16 (sixteenth place) in the SOUTH Region.

Northern Iowa Panthers

## Phase 1b: Predicting Winning Probabilities

Your team is busy and working overtime; it's getting intense! You'll predict the winning probabilities for 10 first-round games in the East Region, including a play-in game. This is your chance to flex your data-driven strategy skills!

Here's the breakdown:

- **One play-in game:** Two teams face off for a chance to enter the tournament.
- **Two play-in possibilities:** Predict how either winner of the play-in game would perform against the #1 seed in the next round.
- **First-round matchups:** Seven games featuring top-ranked teams in the East Region.

But don't be fooled by the seeding. The highest-seeded team isn't always the strongest! Your predictions should focus on the stats and strategies that truly matter, not just the seed number.

**When you submit your predictions, your submission for each game should be a numeric probability between 0 and 1. Think of it as assigning the higher-seed team a “chance to win” based on your analysis.**

## Play-In Games

For each game, estimate and input the probability that the team listed first wins. Each of your answers should be a numeric value between 0 and 1. For example, if you think a team has a 67% chance of winning, write .67 as your answer.

---

### Game 1: Rhode Island Rams - North Carolina Tar Heels

Estimate the probability that the **Rhode Island Rams** wins:

0.16

---

### Game 2: North Carolina State Wolfpack - Rhode Island Rams

Estimate the probability that the **North Carolina State Wolfpack** wins:

0.99

---

### Game 3: North Carolina State Wolfpack - North Carolina Tar Heels

Estimate the probability that the **North Carolina State Wolfpack** wins:

0.51

## East Region Tournament Matchups

For each game, estimate and input the probability that the team listed first wins. Each of your answers should be a numeric value between 0 and 1. For example, if you think a team has a 67% chance of winning, write .67 as your answer.

---

### **Game 1: Liberty Flames - Bucknell Bison**

Estimate the probability that the **Liberty Flames** wins:

0.69

---

### **Game 2: Drexel Dragons - Delaware Blue Hens**

Estimate the probability that the **Drexel Dragons** win:

0.85

---

### **Game 3: Massachusetts Minute Women - Princeton Tigers**

Estimate the probability that the **Massachusetts Minute Women** wins:

0.83

---

### **Game 4: Buffalo Bills - Stony Brook Seawolves**

Estimate the probability that the **Buffalo Bills** wins:

0.81

### **Game 5: Fairfield Stags - Towson Tigers**

Estimate the probability that the **Fairfield Stags** wins:

0.28

---

### **Game 6: UConn Huskies - Campbell Fighting Camels**

Estimate the probability that the **UConn Huskies** wins:

0.93

---

### **Game 7: American University Eagles - Columbia Lions**

Estimate the probability that the **American University Eagles** wins:

0.56

## **Phase 1c: Tell Your Story - Summarize Your Methodology**

Approach this section as if your team is presenting its findings to the head coach and front office executives, breaking down how your insights can shape strategy and drive success.

Your task is to **craft a concise but comprehensive explanation of your approach**—how you analyzed the raw data, identified key drivers of performance, and predicted game outcomes to inform critical decisions.

Ensure your explanation is detailed enough that your methodology could be replicated by another analytics team, showcasing your work's rigor and transparency.

---

## Process

How did you clean or transform the raw data before analysis? Please describe it in ~50 words.

First, we averaged out each statistic for each team by sorting, then for each game, we computed the past statistics for each team and deleted current statistics to prevent leakage. Then, we merged the rows from the same game into 1 row and calculated the difference between team A's stats and team B's stats.

Did you create any additional variables? Please describe in ~25 words.

We created variables such as:

Field-Goal%

Assist to turnover ratio

Team contribution

PIE (Player Impact Estimate)

Average Turnover%

Average Effective Field-Goal%

Win%

Average PPG

## Phase 1c: continued

Approach this section as if your team is presenting its findings to the head coach and front office executives, breaking down how your insights can shape strategy and drive success.

Your task is to **craft a concise but comprehensive explanation of your approach**—how you analyzed the raw data, identified key drivers of performance, and predicted game outcomes to inform critical decisions.

Ensure your explanation is detailed enough that your methodology could be replicated by another analytics team, showcasing your work's rigor and transparency.

# Tools & Techniques

## Responses Selected:

AI (Code interpreter, ChatGPT, Copilot, etc.)

Google Sheets

Python

Other, please specify...: Geeks for Geeks, Stack Overflow

How did you use the selected software? For what purpose? Please describe in ~50 words.

We used ChatGPT to clarify complex concepts in ML. We leveraged ChatGPT to troubleshoot errors in our model such as dataframe merging issues etc. We used google colab and libraries such as pandas, sklearn, numpy, and seaborn since they are powerful tools that helped us visualise and manipulate data, make statistical-connections, and leverage powerful algorithms.

What statistical methods did you employ? Please describe in ~100 words.

For ranking, we used the elo rating system that is used widely in [Chess.com](#). We used averaging and simple math for feature engineering, but also used intricate formulas used in the NBA's advanced-statistics. Next, we used Min-Max scaling for normalization so larger values won't impact training performance. In addition, we split the data into train, validation, and test in order to evaluate different algorithms accurately without potential overfitting. Lastly, we used a sequential neural network with 2 hidden layers with 16 and 8 neurons that is equipped with L2 weight decay and LeakyRelu to prevent overfitting. By carefully looking at a loss graph we fine tuned the parameters.

## Phase 1c: continued

Approach this section as if your team is presenting its findings to the head coach and front office executives, breaking down how your insights can shape strategy and drive success.

Your task is to **craft a concise but comprehensive explanation of your approach**—how you analyzed the raw data, identified key drivers of performance, and predicted game outcomes to inform critical decisions.

Ensure your explanation is detailed enough that your methodology could be replicated by another analytics team, showcasing your work's rigor and transparency.

## Predictions

How did you create team rankings? Please describe in ~50 words.

We used an elo ranking system. We start with default parameters and we calculate probability a team with elo\_A would beat a team with elo\_B. If a team wins, they gain certain amount of elo based on different factors such as strength of schedule and difference between the teams ratings.

How did you determine game-winning probabilities? Please describe in ~50 words.

We fitted our pre-processed data to a neural network and calculated our accuracy and loss, then we looked at a loss graph. We fine tuned our parameters such as layers, neurons, learning rate, and regularization based on if the model was overfitting or underfitting. We did this until both training and validation converged similarly.

### What statistical methods did you use to produce your predictions?

(check all that apply)

#### Responses Selected:

Averaging

Grouping

Logistic Regression

Other, please specify...: Neural Network

## Phase 1c: continued

Approach this section as if your team is presenting its findings to the head coach and front office executives, breaking down how your insights can shape strategy and drive success.

Your task is to **craft a concise but comprehensive explanation of your approach**—how you analyzed the raw data, identified key drivers of performance, and predicted game outcomes to inform critical decisions.

Ensure your explanation is detailed enough that your methodology could be replicated by another analytics team, showcasing your work's rigor and transparency.

## Your Insights

How did you assess your model performance? Please describe in ~50 words.

We assessed our model's performance by analyzing a training and validation graph with X axis being number of epochs and y axis being loss. Throughout the process we identified when the validation began to rise, then tweaked a parameter and tested again until loss was low while training and validation loss stayed near each other.

Did you use generative AI tools (ChatGPT, etc.)? If so, explain how in ~25 words.

We used ChatGPT to understand complex techniques/algorithms such as regularization, fine-tuning, and assessing model performance, but also with debugging and troubleshooting.

Did you use any additional outside data sources? If so, list them and explain why they were relevant in 50 words or less.

No

**Did you use any of the educational materials provided for this competition?**

No

## Final Step - Verify Your Team Information

Before submitting your answers for Phase 1, please reconfirm your team member details. We understand that teams may have changed since the initial registration period, so this is your opportunity to ensure that the correct names are listed.

These will be the official team details used for the remainder of the competition and for generating certificates of participation—so please double-check that everything is entered correctly!

# High School

Please enter your high school details.

High School Name	McDonogh
High School Country	United States
High School Zip Code	21117

## Team Composition

Please select the number of students on your team.

5

## Student Team Leader

**Please be sure to add the new team leader as a collaborator on your application, and remove the old one!**

Please use their school email address, if they have one. This should be the same email address you use to invite this student as a collaborator.

First Name	Huy
Last Name	Le
Email Address	<a href="mailto:hgle@mcdonogh.org">hgle@mcdonogh.org</a>
Graduation Year	2027

## Student Team Member #2

Please use their school email address, if they have one.

First Name	Chidi
Last Name	Onwuanibe
Email Address	<a href="mailto:cgonwuanibe@mcdonogh.org">cgonwuanibe@mcdonogh.org</a>
Graduation Year	2027

## Student Team Member #3

Please use their school email address, if they have one.

First Name	Arvin
Last Name	Bozorgi
Email Address	<a href="mailto:abozorgi@mcdonogh.org">abozorgi@mcdonogh.org</a>
Graduation Year	2027

## Student Team Member #4

Please use their school email address, if they have one.

First Name	Jonathan
Last Name	Afolabi
Email Address	<a href="mailto:ojafolabi@mcdonogh.org">ojafolabi@mcdonogh.org</a>
Graduation Year	2027

## Student Team Member #5

Please use their school email address, if they have one.

First Name	Logan
Last Name	Walsh
Email Address	<a href="mailto:lpwalsh@mcdonogh.org">lpwalsh@mcdonogh.org</a>
Graduation Year	2027

## Phase 1: Submit Your Answers

**After submitting this form, you will have officially completed Phase 1 of the Wharton High School Data Science Competition!** Please ensure all your answers are entered correctly, as you will not be able to re-submit once your responses are submitted.

Teams advancing to the semi-finals will be announced at **9:00 AM EST on Monday, March 10th**.

---

Please check the box to confirm the submitted information is accurate and that the team members listed are the students who have completed the analysis and work for the Wharton High School Data Science Competition.

### Responses Selected:

As the team leader, I confirm