Name: Brandon Wisniewski

School: NC State University

I sourced my data from the HoopR package in R. I used the espn\_mbb\_team\_stats() function to output season data from each team from the last 4 regular seasons. I then used the espn\_mbb\_scoreboard() function to obtain the results from each game in each of the last 3 tournaments. I then split up the Results datasets to separate Home (higher seeded) and Away (lower seeded) teams. The purpose of splitting them apart was to create a new dataset to calculate differences in regular season stats between matchups, and later using a model to predict the outcome based on these differences. My model contained 170 total games from the 2021-23 tournaments (I had to remove some observations due to missing data and outliers).

I employed a Logistic Regression Model using the Following Predictors:

* Rebounds per game
* Assists per game
* Scoring Efficiency
* Field Goal Percentage
* Free Throws
* Free Throw Percentage
* 3 pt Percentage
* Assist/Turnover Ratio
* Steal/Foul Ratio
* Pace (Estimated Possessions per game)
* Points per Possession
* Assists\*Field Goal Percentage
* Scoring Efficiency\*Pace
* Free Throws\*Free Throw Percentage
* 3 pt Percentage\*3 pt Attempts

Once the model from 2021-23 was fitted, I applied the coefficients to the same differences in stats in all possible matchups in this year’s tournament.

Note: McNeese St. and St. Peter’s had incorrect data in the following categories: Scoring Efficiency, Field Goal Percentage, Free Throw Percentage, Free Throws, 3 pt Percentage, Pace. I manually edited those numbers using figures from CBB reference, but the Model was still not able to accommodate those teams, assigning Win Probability Values of 0 or NA to each of their matches. For St. Peter’s this probably will not matter, but since McNeese has more potential to make a run, this shortcoming was unfortunate.

Happy March!