**Written Report (130 Points)**

More specifically, this written description should include:

1. (5 points) a one-paragraph abstract, giving a brief description of relevant data and any major findings.

In my shiny application, I delved into data from the 2024 March Madness tournament, examining teams’ offensive and defensive statistics alongside their win probabilities against Division 1 teams. Additionally, I analyzed another dataset spanning from 1985 to 2019, detailing tournament winners and losers for each playoff round, including score results over the years. My analysis revealed that in the 2024 March Madness tournament, teams with higher two-point percentages and lower defensive two-point percentages tended to achieve more success. I also observed that in the first round of playoffs from 1985 to 2019, most games had score differences below 20 points, with only one game exceeding a 50-point difference. The subsequent rounds generally featured score differences below 30 points, except for the Final Four, where one game in 2016 saw a 44-point difference, with Villanova defeating Oklahoma 95 to 51. Furthermore, only one winning team in these years scored over 100 points in the Final game.

1. (10 points) an introduction section, further introducing the data and giving the questions of interest you will be exploring in your project.

The data being used in my shiny application is on this past March Madness tournament, 2024 playoff, along with tournament results from each playoff round starting in the year 1985, when the league officially switched over to 64 team format, all the way to 2019, where it contains each results of every playoff round, including the winning/losing team, winner/loser score, along with a calculated difference in the two scores, to look at score variation among between the competing teams.

           For the March Madness ’24 dataset, I am interested in exploring a comparison table where you could stack multiple teams on top of each other and either compare defensive or offensive stats to see which teams are expected to put up more points or prevent points from being scored on the defensive end. More specifically, this dataset will contain the team’s rank in the tournament, their playoff seed, team name, the conference they play in, and number of games played (an indication of how far the team made it). Adjusted offensive and defensive efficiency is the number of points allowed/scored per 100 possessions against average Division One offense/defense. Barthag is a power Rating measured by a team’s chance to beat an average Division I team. EFG% and EFGD% represent the effective field goal percentage scored/allowed—offensive and defensive rebound rate, measured as a percentage. FTR and FTRD represent the free throw rate both on offense and defense and are measured by free throws attempted per field goal. There is an offensive and defensive measurement for both two- and three-point percentages, one for how many scored and the other for points allowed. Adjusted tempo is an estimate of tempo (determined by possessions per 40 minutes) a team has against a team that is playing at an average Division One tempo. Finally, wins above the bubble (WAB) mark the cut-off between making the tournament or not. The primary purpose of using this data is to explore what factors impact teams, specifically to recognize what makes a team more successful when playing in the tournament.

For the data covering tournament results from 1985 to 2019, I am interested in finding any trends in the winning/losing scores as well as differences in the scores between compete, as well finding any score records in the tournament, particularly in the fifth and sixth round of playoffs(final two rounds). This dataset will contain, the year, playoff round number, playoff seed number for both teams one and two, scores for both teams, names of both teams differentiated between one and two, then mutated the data to make a column for teams with winning and losing scores, names for each game, and lastly a column representing the difference between the score of the winning and losing team. I will be using my shiny application to produces plots of each playoff round and user can adjust the round, year, and region filters for desired outputs. This will allow the user to easy identify any games that seems to be an outlier, which would be evidence of a high-score record, if the point is found on higher y-values.

1. (110 points) a section showing relevant visualizations. Keep in mind that you do not need to write this as though it is a “step-by-step” thought-process. Instead, you should include your most interesting visualizations, along with interpretations for each. This section will likely be the longest, and, if you are answering multiple different questions of interest, you might break this section up into subsections.
   1. optionally, you can reference a Shiny app in this section, perhaps providing static visuals from your app. If you did build a Shiny app for your project, then this section is expected to be shorter than if you did not build a Shiny app. But, you should still describe major findings from the Shiny app.

Looking more in depth at the data, while observing results from the 2024 march madness, it can be concluded that the teams with higher seeds will to have higher chances of winning, however, still does not account for any potential upsets from the lower seeds. For example, in the most recent tournament, Houston was upset by North Carolina state, where Houston did technically have a higher chance of winning but by the underdogs.

Now, looking at the final four teams: UConn, Purdue, NC state, and Alabama. By offensive stats NC state had the lowest rates, and was the overall lowest seed, but still made it to the final four. Looking at UConn and Purdue the final two teams, where UConn is from the Big East conference and Purdue the big ten, both are the same playoff seed, UConn holds a higher chance of beating a D1 team, Purdue has the strong adjusted defensive efficiency, and looking at two-point shot percentage UConn is more successful, which can ultimately lead to them being able to put up more points, then finally look at wins above bubble, UConn had the better chance of making the tournament to begin with. Overall, UConn holds the stronger tournament stats, with a few areas where Purdue does excel in better but not enough to have helped them win.

Now Looking at, tournament winners and losers along with the scores from throughout the years, there is no real outstanding trend to the scores, other than when looking at only the championship rounds for 1985 to 2019, there is only one game in which the winners had scored more than 100 points in the final round, next closest score only makes it to 89 points. The winners of that highest score game was in 1990 where University of Las Vegas Nevada beat Duke, 103 pts to 73 pts.

While looking at the difference in scores between the losing and winning team found that throughout 1985 to 2019 in the first round of the playoffs there was only one game to have a score difference of over 50 points Where in 2009, Connecticut (UConn) beat Chattanooga 103 to 47, making it a 56 point difference

1. (5 points) a concluding section describing future work that you would complete if you had more time as well as any limitations to your visualizations.