**Ella Heath**

**Project Proposal**

Analyze how team seeds in March Madness affect their tournament performance and overall tournament upsets.

**Introduction**

March Madness is a yearly 3-week long single elimination tournament in NCAA Division 1 Men’s and Women’s basketball. The tournament consists of 68 teams that compete in 7 rounds to become national champions. The March Madness tournament has taken place every year (except the 2019-2020 season due to Covid-19) since 1939. There are two ways teams can participate in the March Madness tournament. First, if a team wins their division postseason conference tournament they receive an automatic bid into the tournament; there are 31 Division 1 conference, so there are 31 automatic bids. The remaining and second way a team can play in the tournament if by receiving an at-large bid through an invitation; a selection committee on “Selection Sunday” decides 37 teams who were not automatic qualifiers based on stats and rankings to decide who will receive an invitation. Once the 68 teams are selected, four of these teams are eliminated in the opening round (Play-In or First Four) of the tournament which leaves the tournament with 64 teams. These teams are given a seed in one of four regions; each region has a 1-16 seed decided by the Selection Committee based on regular season and conference tournament performance (Wilco, 2025).

The tournament then starts with the top seeded team (No. 1) playing the lowest seeded team (No.16). This continues with the next highest team playing the next lowest team, leaving team No.8 playing team No.9. Single elimination leaves 6 rounds (Round of 64, Round of 32, Sweet 16, Elite Eight, Final Four, National Champtionship) for a team to be considered the national champion. March Madness if a highly anticipated sports event where millions of people create brackets in hopes of creating the perfect bracket (predicting the winner of each of the 68 games correctly). The chances of predicting a perfect bracket if you have some prior knowledge of basketball is about 120.2 billion. However, a perfect bracket has never been made, the longest perfect bracket was in 2019 when an individual correctly predicted 49 games until the 50th game after an upset (No.3 Purdue beats No.2 Tennessee). An upset in March Madness happens when a lower seeded team beats a higher seeded team (Benzie, 2025). Because of this, March Madness fans have become obsessed with trying to create a perfect bracket and correctly predicting upsets based on teams seedings going into the tournament (Wilco, 2025).

**Data**

The main data source I plan to use is a compilation of March Madness games spanning from 1985-2024 games. This data source keeps track of the year the game was played, the round the game was played, and information about both the winning and losing teams. With this data set I can explore what teams have won the most tournament games, what seed has won the most tournament games, and vice versa with losing tournament games. I will be able to compare these results with the other data sources I plan to use.

<https://github.com/shoenot/march-madness-games-csv/blob/main/csv/combined.csv>

This data source has information about games results for only seeds and is not team specific. I can scrape this site to pull the table to see the best performance each seed has as well as their win-loss percentage. I will be able to visualize this data and show what seeds win the most and those who lose the most.

<https://www.ncaa.com/news/basketball-men/article/2025-02-05/records-every-seed-march-madness-1985-2024>

This data source will be my main form of web scraping. This source tracks every upset that has happened in the March Madness tournament from 1985-2024. I can scrape the team names for all upsets as well as their seeds. I can also scrape things like how the upset happened or what it meant for the rest of the tournament that year.

<https://www.ncaa.com/live-updates/basketball-men/d1/mens-march-madness-upset-tracker>

**Data Dictionary**

*Data Dictionary for March Madness Games, 1985-2024*

|  |  |  |
| --- | --- | --- |
| **Field** | **Type** | **Description** |
| year | Numeric | The year the game was played |
| round\_of | Numeric | Round the game was played |
| winning\_team\_name | Text | Name of the team that won the game |
| winning\_team\_seed | Numeric | Seed of the team that won the game |
| winning\_team\_score | Numeric | Score of the team that won the game |
| losing\_team\_name | Text | Name of the team that lost the game |
| losing\_team\_seed | Numeric | Seed of team that lost the game |
| losing\_team\_score | Numeric | Score of the team that lost the game |

*Data Dictionary for Records of every seed in March Madness, 1985-2024*

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| Seed | Text | Seeds (1-16) |
| Overall Record | Text or Numeric | Record of the seed |
| Best Finish | Text | Best finish of the seed |
| Seed VS. Seed | Numeric | Seed vs. Seed options |
| W-L | Text or Numeric | The seed vs. seed options win-loss record |
| PCT. | Numeric | Seed vs. Seed win-loss percentage |

**Proposed Analysis**

1. Is there a higher seed that is more likely to lose as a result of an upset/ is there a lower seed that’s more likely to win as a result of an upset?
2. What teams have been seeded No.1 the most, how many times have they won the tournament?
3. What seeds are most likely to make to the most anticipated round of the tournament, the Final Four?

# Combine feature names with coefficients

equation\_parts = [

f"{coef:.2f} \* {feature}" for coef, feature in zip(coefficients, X.columns)

]

# Build full equation string

equation = f"Predicted Score = {intercept:.2f} + " + " + ".join(equation\_parts)

print(equation)

**Works Cited**

“The Longest an NCAA Bracket Has Ever Stayed Perfect.” *NCAA.Com*, NCAA.com, 24 Mar. 2025, www.ncaa.com/news/basketball-men/bracketiq/2025-03-23/longest-ncaa-bracket-has-ever-stayed-perfect.

“Records for Every Seed in March Madness from 1985 to 2024.” *NCAA.Com*, NCAA.com, 5 Feb. 2025, www.ncaa.com/news/basketball-men/article/2025-02-05/records-every-seed-march-madness-1985-2024.

Shoenot. “March-Madness-Games-CSV/CSV/Combined.Csv at Main · Shoenot/March-Madness-Games-CSV.” *GitHub*, github.com/shoenot/march-madness-games-csv/blob/main/csv/combined.csv. Accessed 26 Mar. 2025.

“We’re Tracking Every Upset in the NCAA Men’s Tournament.” *NCAA.Com*, NCAA.com, 20 Mar. 2024, www.ncaa.com/live-updates/basketball-men/d1/mens-march-madness-upset-tracker.

“What Is March Madness: The NCAA Tournament Explained.” *NCAA.Com*, NCAA.com, 22 Jan. 2025, www.ncaa.com/news/basketball-men/bracketiq/2025-01-22/what-march-madness-ncaa-tournament-explained.