The NCAA Division (known informally as March Madness or the Big Dance) is a single-elimination tournament played each spring in the United States, currently featuring 68 college basketball teams from the Division I level of the National Collegiate Athletic Association (NCAA), to determine the national championship. The tournament was created in 1939 by the National Association of Basketball Coaches, and was the idea of Ohio State University coach Harold Olsen. Played mostly during March, it has become one of the most famous annual sporting events in the United States.

File descriptions

Below we describe the format and fields of the essential data files. Teams This file identifies the different college teams present in the dataset. Each team has a 4 digit id number. Seasons This file identifies the different seasons included in the historical data, along with certain season-level properties.

Season - indicates the year in which the tournament was played

Dayzero - tells you the date corresponding to daynum=0 during that season.

Region W/X/Y/Z - by convention, the four regions in the final tournament are always named W, X, Y, and Z. Whichever region's name comes first alphabetically, that region will be Region W. And whichever Region plays against Region W in the national semifinals that will be Region X. For the other two regions, whichever region's name comes first alphabetically, that region will be Region Y, and the other will be Region Z. So for this season, the W/X/Y/Z are East, Midwest, South, and West.

RegularSeasonCompactResults

This file identifies the game-by-game results for 32 seasons of historical data, from 1985 to 2015. Each year, it includes all games played from daynum 0 through 132. Each row in the file represents a single game played.

• Season - this is the year of the associated entry in seasons.csv (the year in which the final tournament occurs)

• Daynum - this integer always ranges from 0 to 132, and tells you what day the game was played on. It represents an offset from the "dayzero" date in the "seasons.csv" file.

• Wteam- this identifies the id number of the team that won the game, as listed in the teams.csv file. No matter whether the game was won by the home team or visiting team, wteam always identifies the winning team.

• Wscore - this identifies the number of points scored by the winning team.

• Lteam - this identifies the id number of the team that lost the game.

• Lscore - this identifies the number of points scored by the losing team.

• Numot- this indicates the number of overtime periods in the game, an integer 0 or higher.

• Wloc - this identifies the location of the winning team. If the winning team was the home team, this value will be "H". If the winning team was the visiting team, this value will be "A". If it was played on a neutral court, then this value will be "N".

RegularSeasonDetailedResults

This file is a more detailed set of game results, covering seasons 2003-2016. This includes team-level total statistics for each game (total field goals attempted, offensive rebounds, etc.) The column names should be self-explanatory to basketball fans (as above, "w" or "l" refers to the winning or losing team): • Wfgm - field goals made

• Wfga - field goals attempted

• Wfgm3 - three pointers made

• Wfga3 - three pointers attempted

• Wftm - free throws made

• Wfta - free throws attempted

• Wor - offensive rebounds

• Wdr defensive rebounds

• Wast - assists

• Wto - turnovers

• Wstl - steals

• Wblk - blocks

• Wpf – personal fouls

TourneyCompactResults

This file identifies the game-by-game NCAA tournament results for all seasons of historical data. The data is formatted exactly like the regular\_season\_compact\_results.csv data. TourneyDetailedResults This file contains the more detailed results for tournament games from 2003 onward. TourneySeeds

This file identifies the seeds for all teams in each NCAA tournament, for all seasons of historical data. Thus, there are between 64-68 rows for each year, depending on the bracket structure.

• Season - the year

• Seed - this is a 3/4-character identifier of the seed, where the first character is either W, X, Y, or Z (identifying the region the team was in) and the next two digits (either 01, 02, 15, or 16) tell you the seed within the region.

• Team - this identifies the id number of the team, as specified in the teams.csv file

TourneySlots

This file identifies the mechanism by which teams are paired against each other, depending upon their seeds.

• Season - the year

• Slot - this uniquely identifies one of the tournament games. For play-in games, it is a three-character string identifying the seed fulfilled by the winning team, such as W16 or Z13. For regular tournament games, it is a four-character string, where the first two characters tell you which round the game is (R1, R2, R3, R4, R5, or R6) and the second two characters tell you the expected seed of the favored team. Thus the first row is R1W1, identifying the Round 1 game played in the W bracket,

• Strongseed - this indicates the expected stronger-seeded team that plays in this game. For Round 1 games, a team seed is identified in this column (as listed in the "seed" column in the tourney\_seeds.csv file), whereas for subsequent games, a slot is identified in this column.

• Weakseed - this indicates the expected weaker-seeded team that plays in this game, assuming all favored teams have won so far. For Round 1 games, a team seed is identified in this column (as listed in the "seed" column in the tourney\_seeds.csv file), whereas for subsequent games, a slot is identified in this column.