

## Prelim2019 RegularSeasonDetailedResults.csv

This file identifies the game-by-game results for many seasons of historical data, starting with the 2003 season (the first year the NCAA® had a 64-team tournament). For each season, the file includes all games played from daynum 0 through 132. It is important to realize that the "Regular Season" games are simply defined to be all games played on DayNum=132 or earlier (DayNum=132 is Selection Sunday, and there are always a few conference tournament finals actually played early in the day on Selection Sunday itself). Thus a game played on or before Selection Sunday will show up here whether it was a pre-season tournament, a non-conference game, a regular conference game, a conference tournament game, or whatever.

- **Season** - this is the year of the associated entry in Seasons.csv (the year in which the final tournament occurs). For example, during the 2016 season, there were regular season games played between November 2015 and March 2016, and all of those games will show up with a Season of 2016.
- **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. For example, the first game in the file was DayNum=20. Combined with the fact from the "Seasons.csv" file that day zero was 10/29/1984 that year, this means the first game was played 20 days later, or 11/18/1984. There are no teams that ever played more than one game on a given date, so you can use this fact if you need a unique key (combining Season and DayNum and WTeamID). In order to accomplish this uniqueness, we had to adjust one game's date. In March 2008, the SEC postseason tournament had to reschedule one game (Georgia-Kentucky) to a subsequent day because of a tornado, so Georgia had to actually play two games on the same day. In order to enforce this uniqueness, we moved the game date for the Georgia-Kentucky game back to its original scheduled date.
- **WTeamID** - 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, or if it was a neutral-site game, the "WTeamID" always identifies the winning team.
- **WScore** - this identifies the number of points scored by the winning team.
- **LTeamID** - this identifies the id number of the team that lost the game.
- **LScore** - this identifies the number of points scored by the losing team. Thus you can be confident that WScore will be greater than LScore for all games listed.
- **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". Sometimes it is unclear whether the site should be considered neutral, since it is near one team's home court, or even on their court during a tournament, but for this determination we have simply used the Kenneth Massey data in its current state, where the "@" sign is either listed with the winning team, the losing team, or neither team. If you would like to investigate this factor more closely, we invite you to explore Data Section 3, which

provides the city that each game was played in, irrespective of whether it was considered to be a neutral site.

- **NumOT** - this indicates the number of overtime periods in the game, an integer 0 or higher.
- **WFGM** - field goals made (by the winning team)
- **WFGA** - field goals attempted (by the winning team)
- **WFGM3** - three pointers made (by the winning team)
- **WFGA3** - three pointers attempted (by the winning team)
- **WFTM** - free throws made (by the winning team)
- **WFTA** - free throws attempted (by the winning team)
- **WOR** - offensive rebounds (pulled by the winning team)
- **WDR** - defensive rebounds (pulled by the winning team)
- **WAsst** - assists (by the winning team)
- **WTO** - turnovers committed (by the winning team)
- **WStl** - steals (accomplished by the winning team)
- **WBlk** - blocks (accomplished by the winning team)
- **WPF** - personal fouls committed (by the winning team)

(and then the same set of stats from the perspective of the losing team: **LFGM** is the number of field goals made by the losing team, and so on up to **LPF**).

## [Teams.csv](#)

This file identifies the different college teams present in the dataset. You will not see games present for all teams in all seasons, because the games listing is only for matchups where both teams are Division-I teams. There are 351 teams currently in Division-I, and an overall total of 364 teams in our team listing (each year, some teams might start being Division-I programs, and others might stop being Division-I programs). Each team has a 4 digit id number.

- **TeamID** - a 4 digit id number, from 1000-1999, uniquely identifying each NCAA® men's team. A school's TeamID does not change from one year to the next, so for instance the Duke men's TeamID is 1181 for all seasons. To avoid possible confusion between the men's data and the women's data, all of the men's team ID's range from 1000-1999, whereas all of the women's team ID's range from 3000-3999.
- **TeamName** - a compact spelling of the team's college name, 16 characters or fewer. There are no commas or double-quotes in the team names, but you will see some characters that are not letters or spaces, e.g., Texas A&M, St Mary's CA, TAM C. Christi, and Bethune-Cookman.
- **FirstD1Season** - the first season in our dataset that the school was a Division-I school. For instance, FL Gulf Coast (famously) was not a Division-I school until the 2008 season, despite their two wins just five years later in the 2013 NCAA® tourney. Of course, many schools were Division-I far earlier than 1985, but since we don't have any data included prior to 1985, all such teams are listed with a FirstD1Season of 1985.

- **LastD1Season** - the last season in our dataset that the school was a Division-I school. For any teams that are currently Division-I, they will be listed with LastD1Season=2018, and you can confirm there are 351 such teams. It has been a few years since any teams stopped being Division-I; the last was Centenary whose final Division-I year was 2011.

## [cbb.csv](#)

This file contains aggregated season statistics from the 2013-2019 Division I college basketball seasons.

- **RK** (Only in cbb20): The ranking of the team at the end of the regular season according to barttorvik
- **TEAM**: The Division I college basketball school
- **CONF**: The Athletic Conference in which the school participates in (A10 = Atlantic 10, ACC = Atlantic Coast Conference, AE = America East, Amer = American, ASun = ASUN, B10 = Big Ten, B12 = Big 12, BE = Big East, BSky = Big Sky, BStH = Big South, BW = Big West, CAA = Colonial Athletic Association, CUSA = Conference USA, Horz = Horizon League, Ivy = Ivy League, MAAC = Metro Atlantic Athletic Conference, MAC = Mid-American Conference, MEAC = Mid-Eastern Athletic Conference, MVC = Missouri Valley Conference, MWC = Mountain West, NEC = Northeast Conference, OVC = Ohio Valley Conference, P12 = Pac-12, Pat = Patriot League, SB = Sun Belt, SC = Southern Conference, SEC = South Eastern Conference, Slnd = Southland Conference, Sum = Summit League, SWAC = Southwestern Athletic Conference, WAC = Western Athletic Conference, WCC = West Coast Conference)
- **G**: Number of games played
- **W**: Number of games won
- **ADJOE**: Adjusted Offensive Efficiency (An estimate of the offensive efficiency (points scored per 100 possessions) a team would have against the average Division I defense)
- **ADJDE**: Adjusted Defensive Efficiency (An estimate of the defensive efficiency (points allowed per 100 possessions) a team would have against the average Division I offense)
- **BARTHAG**: Power Rating (Chance of beating an average Division I team)
- **EFG\_O**: Effective Field Goal Percentage Shot
- **EFG\_D**: Effective Field Goal Percentage Allowed
- **TOR**: Turnover Percentage Allowed (Turnover Rate)
- **TORD**: Turnover Percentage Committed (Steal Rate)
- **ORB**: Offensive Rebound Rate
- **DRB**: Defensive Rebound Rate Allowed
- **FTR** : Free Throw Rate (How often the given team shoots Free Throws)
- **FTRD**: Free Throw Rate Allowed
- **2P\_O**: Two-Point Shooting Percentage
- **2P\_D**: Two-Point Shooting Percentage Allowed
- **3P\_O**: Three-Point Shooting Percentage
- **3P\_D**: Three-Point Shooting Percentage Allowed

- **ADJ\_T:** Adjusted Tempo (An estimate of the tempo (possessions per 40 minutes) a team would have against the team that wants to play at an average Division I tempo)
- **WAB:** Wins Above Bubble (The bubble refers to the cut off between making the NCAA March Madness Tournament and not making it)
- **POSTSEASON:** Round where the given team was eliminated or where their season ended (R68 = First Four, R64 = Round of 64, R32 = Round of 32, S16 = Sweet Sixteen, E8 = Elite Eight, F4 = Final Four, 2ND = Runner-up, Champion = Winner of the NCAA March Madness Tournament for that given year)
- **SEED:** Seed in the NCAA March Madness Tournament
- **YEAR:** Season