Adams_FinalProject

Grace Adams

2024-04-16

A handful of fyi stuff:

- cbb.csv has seasons 2013-2019 and seasons 2021-2023 combined
- The 2020 season's data set is kept separate from the other seasons, because there was no postseason due to the Coronavirus.
- cbb24 contains data as of 3/18/2024. This dataset will be updated and added to cbb.csv at the conclusion of the Tournament.

Data Dictionary:

- 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: Offensive 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

Loading packages

#install.packages("tidyverse")

library(tidyverse)

Reading in the data

Set the directory where your .csv files are located

folder_path <- "data_raw/cbb"

List all .csv files in the folder

file_list <- list.files(folder_path, pattern = "cbb[0-9]{2}\.csv")

Loop through each file and read it into the environment

I don't think this was taught in class, need to find a way to do this with the things taught in class??

for (file in file_list) { # Extract the two-digit number from the file name file_number <- gsub("cbb|\.csv", " ", file)

Read the .csv file into a data frame with a name based on the file number assign(paste0("cbb", file_number), read.csv(file.path(folder_path, file))) }

Data Exploration

summary(cbb)

What kind of stuff do I have here?

- 4 character variables (TEAM, CONF, POSTSEASON, SEED)
- 20 numeric variables (G, W, ADJOE, ADJDE, BARTHAG, EFG_O, EFG_D, TOR, TORD, ORB, DRB, FTR, FTRD, X2P_O, X2P_D, X3P_O, X3P_D, ADJ_T, WAB, YEAR)
  - Very happy that there does not appear to be any NAs in the numeric stuff or in the character stuff