

march_madness_2021

Jared Klug

Introduction

In this project, we have gotten data from Kaggle, which contains extensive data of regular season games, and conferences for the NCAA Basketball seasons from 2003-2021. Using this data, I would like to predict the potential scores of the upcoming games in the 2021 March Madness Tournament

```
df = read.csv("./data/ncaam-march-mania-2021/MRegularSeasonDetailedResults.csv") %>%
  rowid_to_column("game_id") %>%
  relocate(WLoc:NumOT, .after = DayNum) %>%
  pivot_longer(
    WTeamID:LPF,
    names_to = "stat",
    values_to = "count"
  ) %>%
  mutate(outcome = case_when(
    str_detect(stat, "^W") ~ "win",
    str_detect(stat, "^L") ~ "loss"
  )) %>%
  mutate(stat = substr(stat, 2, nchar(stat))) %>%
  pivot_wider(
    names_from = stat,
    values_from = count
  ) %>%
  mutate(TeamID = as.factor(TeamID)) %>%
  unnest()

team_id = read.csv("./data/ncaam-march-mania-2021/MTeams.csv") %>%
  select(TeamID, TeamName)

df[["TeamID"]] = team_id[match(df[["TeamID"]], team_id[["TeamID"]]), "TeamName"]
```