To Do:

* Data preprocessing:
  + ~~Season averages:~~
    - ~~Offensive efficiency~~
    - ~~Defensive efficiency~~
    - ~~Effective technical shooting percentage~~
  + Last 14 day averages:
    - ~~Win ratio~~
    - ~~Scoring gaps~~
    - ~~Offensive efficiency~~
    - ~~Defensive efficiency~~
    - ~~Effective technical shooting percentage~~
  + Hotness metric:
    - Hotness should start at 0, and decrease after round 2
    - Maximum when seed = 16, round = 2
      * 16 seed is hottest right after they beat a 1 seed
    - Local max when seed > 8, round >= 3
      * Teams that are low seeds (bottom half, >8) that make it to at least the 3rd round increase their hotness in each round
    - Minimum when seed = 1, round = 1
  + Target data:
    - ~~Probability of 0 for the losing team, 1 for the winning team → predict probabilities with regression~~
    - ~~Predict score differential → positive = team A wins, negative = team B wins~~
    - Classify each winning team as 1, losing team as 0 → predict classes with classification, extract probabilities
* Model training:
  + ~~Regression:~~
    - ~~Linear regression~~
    - ~~Elastic-net~~ 
      * ~~Good for multiple correlated features~~
    - ~~KNN regression~~
    - ~~Random forest regression~~
    - ~~AdaBoost regression~~
      * ~~Good for weak learners~~
    - ~~DNN~~
  + Classification:
    - Logistic regression
      * Testing accuracy: 0.7036
      * Testing Brier score: 0.4353
    - KNN classifier
      * Testing accuracy: 0.6716
      * Testing Brier score: 0.4294
    - Random forest classifier
      * Testing accuracy: 0.7049
      * Testing Brier score: 0.3535
    - Gradient tree boosting
    - DNN
* Model prediction:
  + Predict on all possible matchups in all possible rounds
    - 64 teams, 6 rounds
    - (64 choose 2) \* 6 = 4032 possible matchups for the tournament
  + Function where I give it two team ID’s and a round → outputs probabilities that each team wins (or just which team wins)

1-16 → 0

2-15 → 3

3-14 → 2

4-13 → 1

5-12 → 1

6-11 → 2

7-10 → 3

8-9 → 0