NCAA Basketball Analysis

Maclean Sherren & Kayla Tansiongco

Brigham Young University

December 12, 2023

Outline

Introduction

- Due to Covid, 2020 NCAA tournament did not happen.
- Question: Are there certain regular season variables that are better predictors for postseason outcomes?
- Goal: Use existing data/trends from other years and regular season data from 2020 to come up with potential methods of prediction and postseason outcomes.

Data Collection

- College Basketball Dataset folder found on Kaggle. Dataset "cbb.csv" has regular and postseason statistics from years 2013-2021. Excludes 2020 season due to lack of postseason statistics.

- Folder included datasets for each individual year, with 2020 ("cbb20.csv") having regular season stats only.

- Variables of final dataset: YEAR, TEAM, CONF, G, W, BARTHAG, POSTSEASON, SEASON_FINAL.
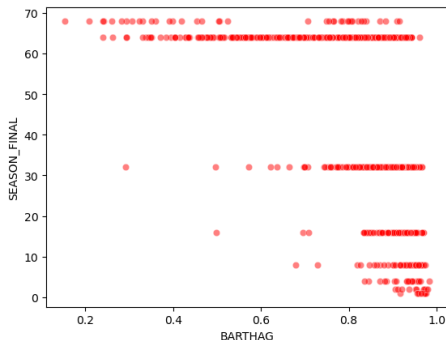
Data Collection

- Created numerically coded column, SEASON_FINAL, derived from POSTSEASON (string data type) for cbb.csv dataframe. (Ex: "Champions" = 1, "2ND" = 2, "F4" = 4)
- Initialized variables POSTSEASON and SEASON_FINAL as NaN for cbb20.csv dataframe. This is what we will fill in through our prediction methods.

Exploratory Data Analysis

- Determine what variables are most correlated with POSTSEASON outcomes.
- Used Spearman correlation since it is more appropriate for measurements taken from ordinal scales.
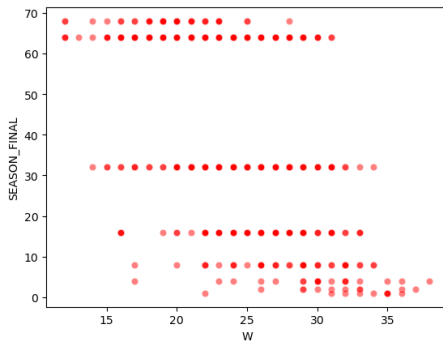
Exploratory Data Analysis

- BARTHAG vs SEASON_FINAL
- BARTHAG = Power Rating (Chance of beating an average Division 1 team)

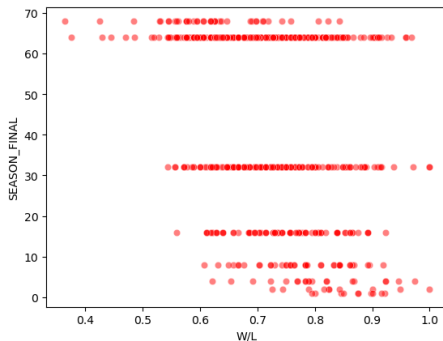Corr: -0.6461273899341722

Exploratory Data Analysis

- W vs SEASON_FINAL
- W = Number of games won

Corr: -0.5125436890870254
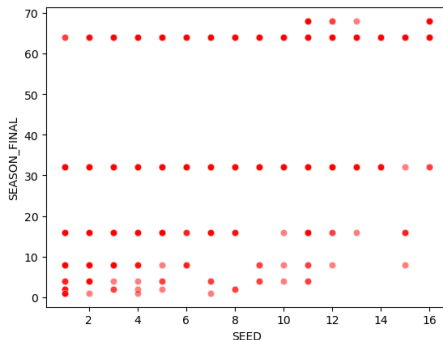
Exploratory Data Analysis

- W/L vs SEASON_FINAL
- W/L = Win-Loss Ratio

Corr: -0.3532346489556209

Exploratory Data Analysis

- SEED vs SEASON_FINAL
- SEED = Seed in the NCAA March Madness Tournament

Corr: 0.6065146756151623

Methodology

- Package contains a function that takes a season and method of prediction, then produces simulated results of an NCAA tournament.
- Compare predicted results to actual results.
- Create a prediction for 2020 NCAA postseason tournament.

Methodology

- Season: 2021
- Method: BARTHAG (Power Rating)

Predicted Final Four

- Baylor
- Michigan
- Gonzaga
- Houston

Predicted Championship

- Gonzaga vs Baylor
- Winner: Baylor

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship

- Gonzaga vs Baylor
- Winner: Baylor

Methodology

- Season: 2021
- Method: W (Wins)

Predicted Final Four

- Winthrop
- Alabama
- Gonzaga
- Loyola Chicago

Predicted Championship

- Alabama vs Gonzaga
- Winner: Alabama

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship

- Gonzaga vs Baylor
- Winner: Baylor

Methodology

- Season: 2021
- Method: SEED

Predicted Final Four

- Baylor
- Gonzaga
- Michigan
- Illinois

Predicted Championship

- Michigan vs Gonzaga
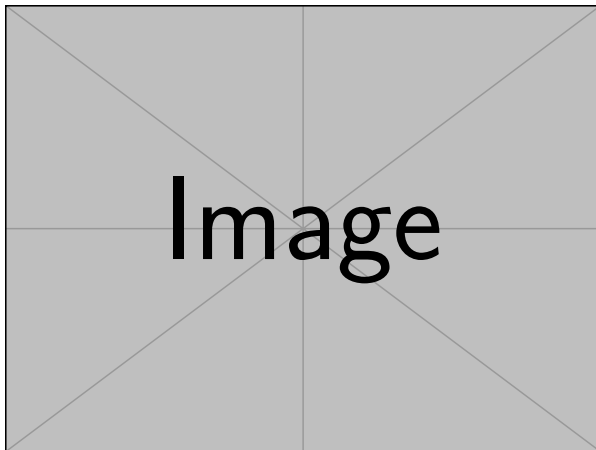- Winner: Gonzaga

Actual Final Four

- Baylor
- Gonzaga
- Houston
- UCLA

Actual Championship
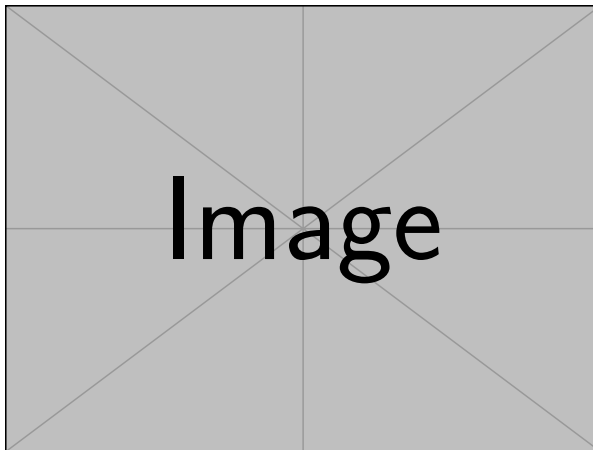
- Gonzaga vs Baylor
- Winner: Baylor

Results

- Season: 2020
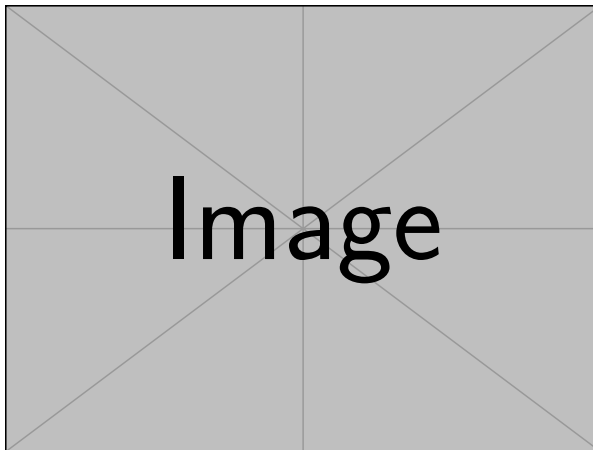- Method: BARTHAG (Power Rating)

Results

- Season: 2020
- Method: W (Wins)

Results

- Season: 2020
- Method: RK (Rank/Equivalent to SEED)

Conclusion

- Summarize the main findings and their implications.
- Reflect on the success of your project and any limitations.
- Suggest areas for future research or improvements.

Questions or comments?

Thank you! Happy Holidays!