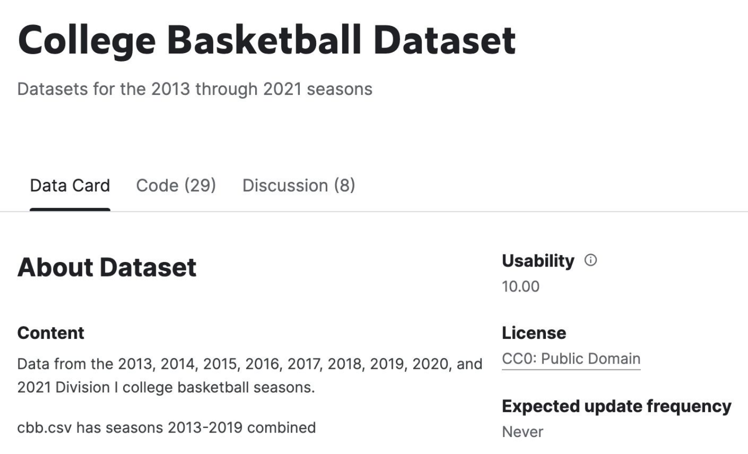
Predicting NCAA Outcomes

R-s2dio: Lukas, Kaitlyn, Jack, Angelie

DATASET

- Division I college basketball seasons from 2013-2019
  - From Kaggle & scraped from Bart Torvik
- Dataset has ***24 variables***
- The dataset includes ***2,455 observations***



College Basketball Dataset

Datasets for the 2013 through 2021 seasons

About Dataset

Content

Data from the 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, and 2021 Division I college basketball seasons.

cbb.csv has seasons 2013-2019 combined

Usability
10.00

License
CC0: Public Domain

Expected update frequency
Never

RESEARCH QUESTION

“How does regular season **adjusted offensive efficiency** and regular season **adjusted defensive efficiency** predict postseason seed?”

DEFINITIONS

adjusted offensive efficiency - points *scored* per 100 possessions against the average D-I defense

adjusted defensive efficiency - points *given* up per 100 possessions against the average D-I defense

seed - NCAA postseason ranking for teams in tournament

LITERATURE REVIEW

Overall: Some research on offensive/defensive ratings and tournament success, but effectively no research on relationship between these ratings and SEED

NCAA study, 2018 - over 9 seasons, a team's offensive rating was ~50% more important than its defensive rating in terms of NCAA tournament success

BleacherReport, 201 3 - between 2003-2013, 35/40 Final Four contestants have been in the top 25 in defensive efficiency; 33/40 have been in the top 25 in offensive efficiency

OUR HYPOTHESIS

We predicted that, in regular season, teams with higher adjusted offensive efficiency & lower adjusted defensive efficiency will be predicted to have higher seeds.

OUR METHODS

01

VISUALIZE

Created ggplot scatterplots to visualize relationships between variables

02

MODEL

Created three linear regression models to predict seeds

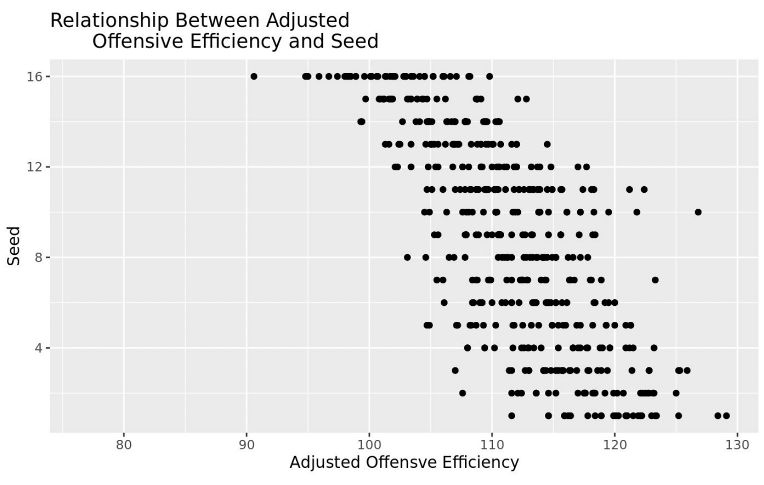
03

COMPARE

Adjusted r-squared to determine which model is best to determine correlation between variables

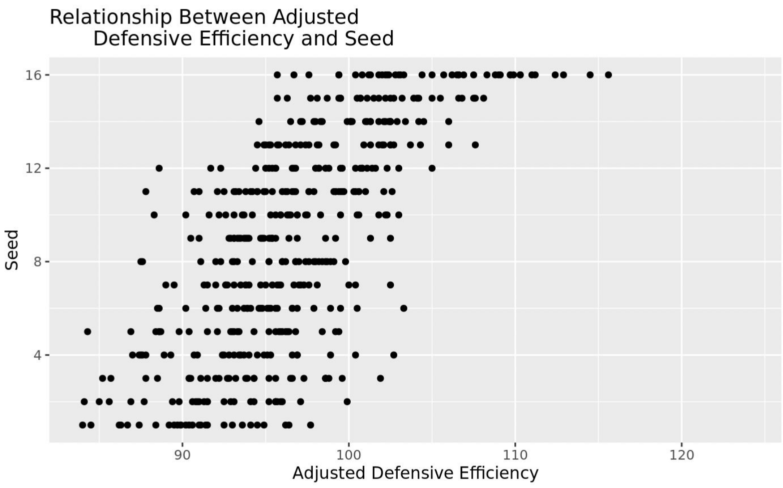
RESULTS

Model 1: ADJOE


Relationship Between Adjusted Offensive Efficiency and Seed

$$\widehat{SEED} = 69.90 - 0.55 * ADJOE$$

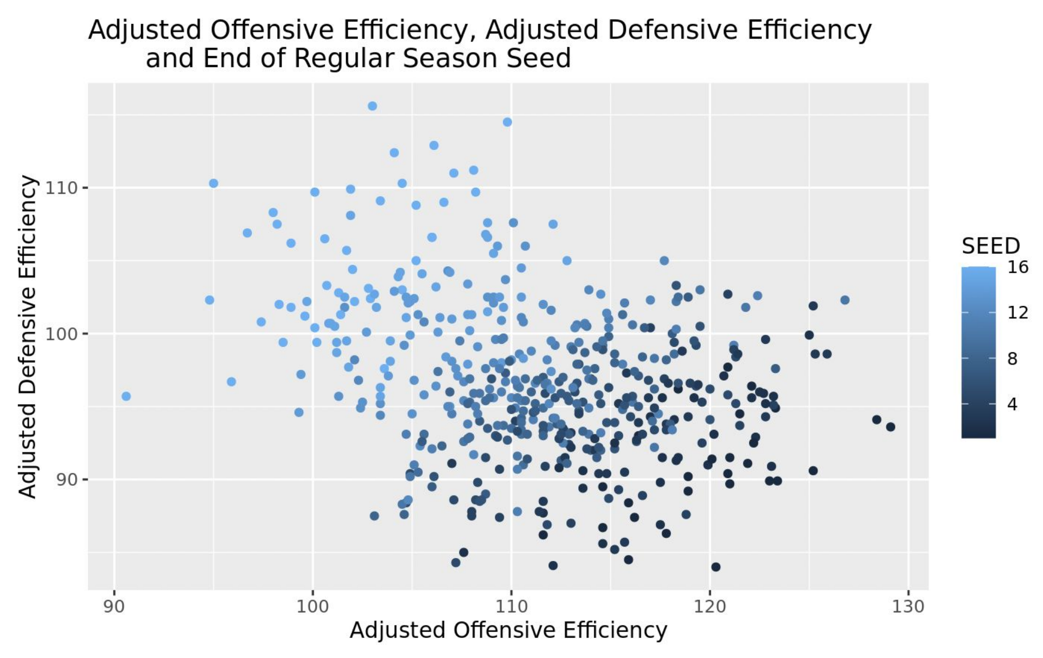
Model 2: ADJDE


Relationship Between Adjusted Defensive Efficiency and Seed

$$\widehat{SEED} = -49.52 + 0.60 * ADJDE$$

RESULTS

Model 3: ADJOE * ADJDE



Things to note

- Not meant for extrapolated data
- Graph does not coincide with the linear regression (Seed is not on the Y)

$$\widehat{SEED} = 183.59 - 1.98 * ADJOE - 1.29 * ADJDE + 0.02 * ADJOE * ADJDE$$

RESULTS

Adjusted R-Squared

Model 1: ADJOE = 0.5544491

Model 2: ADJDE = 0.4853405

Model 3: ADJOE * ADJDE = 0.8094014

AIC (Akaike Information Criterion)

Model 1: ADJOE = 2438.484

Model 2: ADJDE = 2507.12

Model 3: ADJOE * ADJDE = 2036.28

CHALLENGES/ TAKEAWAYS

- Narrowing our research question
    - Choosing variables
- Visualizations and models staying on topic with research question
- Using an Adjusted R-Squared Model
    - Justification of our models

Limitations

- Considering factors of being 1-seed
- Definition of success
- The kinds of models we could use considering variable type

CONCLUSION

- ADJOE and ADJDE correlation to seed
- Importance of AIC and Adj. R Squared
- Trying to use different models
- Combining variables

Thank you!