NCAA BASKETBALL TEAMS' OFFENSIVE AND DEFENSIVE EFFICIENCY

Panda Ballers: Audrey Sompie, Anirudh Dhawan, Sanjana Holla, Naomi Webster, Dane Nighswander, Ira Joshi

PROBLEM

- Understanding whether defensive or offensive capabilities matter more for winning games

DATA COLLECTION

- The dataset was scraped from barttorvik.com, covering NCAA Division I men's basketball teams from 2013-2023.

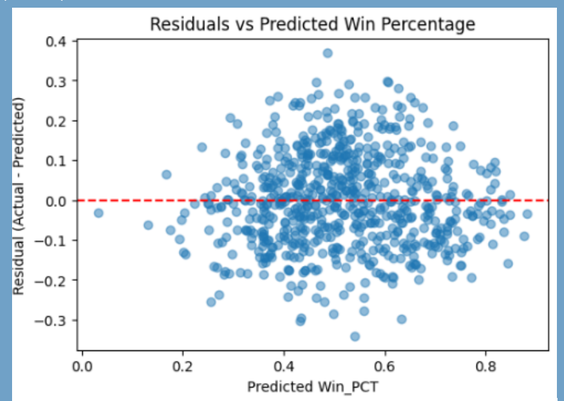
DATA PREPARATION AND EXPLORATION

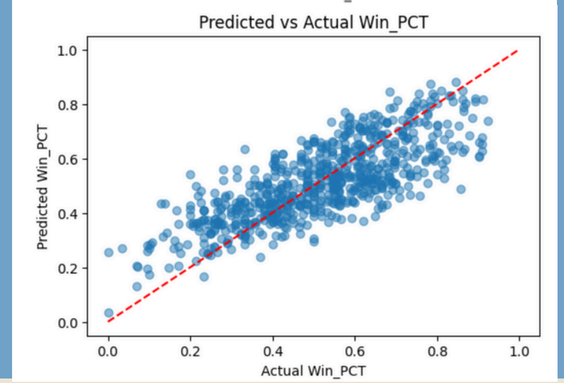
- Rows with missing values in critical variables (W, ADJOE, ADJDE, ADJ_T, ORB) were removed
- POSTSEASON, SEED, and other contextual variables were removed
- ADJOE => Higher = Better
- ADJDE => Lower = Better
- Engineered new columns for win percentage and team styles (Win_PCT, High_Defense, High_Offense, Style)
- EDA indicated fairly linear relationships between the predictors and win percentage

TEAM	ADJOE	ADJDE	ADJ_T	High_Offense	High_Defense	Style	Win_PCT
North Carolina	123.3	94.9	71.7	True	True	Both_high	0.825000
Wisconsin	129.1	93.6	59.3	True	True	Both_high	0.900000
Michigan	114.4	90.4	65.9	True	True	Both_high	0.825000
Texas Tech	115.2	85.2	67.5	True	True	Both_high	0.815789
Gonzaga	117.8	86.3	71.5	True	True	Both_high	0.948718
...	...	...	...	...	...	...	...
Toledo	119.9	109.6	69.5	True	False	Offense_only	0.794118
Liberty	111.4	97.3	64.4	True	True	Both_high	0.818182
Utah Valley	107.1	94.6	69.8	True	True	Both_high	0.823529
UAB	112.4	97.0	70.7	True	True	Both_high	0.763158
North Texas	110.0	93.8	58.7	True	True	Both_high	0.861111

MODEL BUILDING

- **Trained linear regression** model to **predict** team **win percentage**
- **MinMax scaling** applied to features
- ADJOE, ADJDE, ADJ_T => Win_PCT
- Variables capture distinct components of team performance: offense, defense, and play style.
- Ran **Permutation Hypothesis Test**
- **H0** (**null**): no difference in Win_PCT
- **H1** (**alt**): difference in Win_PCT





MODEL EVALUATIONS

- R^2 = 0.60, RMSE = 1.4%
- ADJOE = +0.64, ADJDE = -0.45, ADJ_T = +0.085
- Coefficients for ADJOE and ADJDE showed that offense and defense have strong linear relationships with winning.
- ADJ_T not as significant as ADJOE & ADJDE
- Model neither underfits nor overfits and no class imbalance
- p-value = 0.0005 = 0.05%
- Offense wins more than defense by 2.6%