2020 MLB Season Analysis

Which Teams Benefited From the Shortened Season?

How Covid affected the 2020 MLB Season

Normal Season	2020 Season
162 Regular Season Games	60 Regular Season Games
10 Playoff Spots	16 Playoff Spots

The Question

Had the playoff format not been expanded, which teams would have been affected the most?

The Process

Create a Linear Regression Model

Scrape data from Baseball-Reference.com

Create a model that maps a team's season statistics to their total wins

"Simulate" The Remaining Games

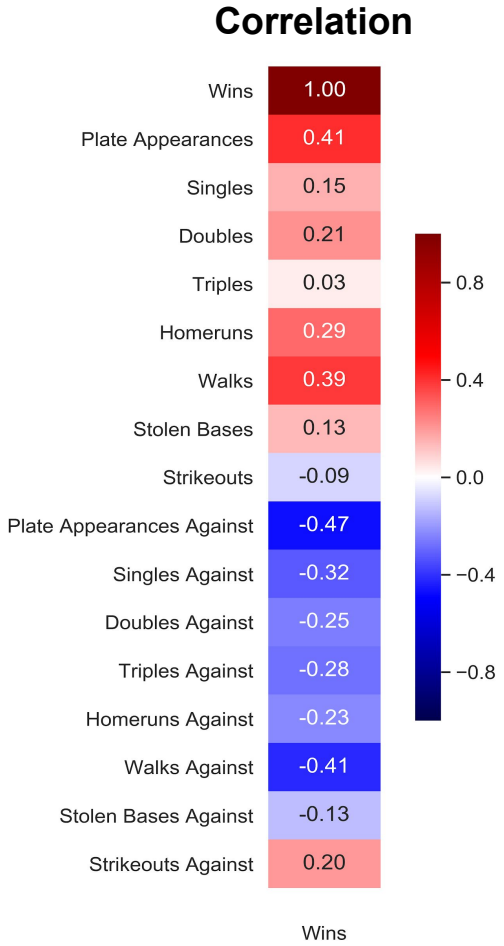
Use the model and 2020 statistics to calculate a predict 2020 win totals for each team over the remaining games

Analyze The Results

Which teams benefited most from the expanded playoff bracket

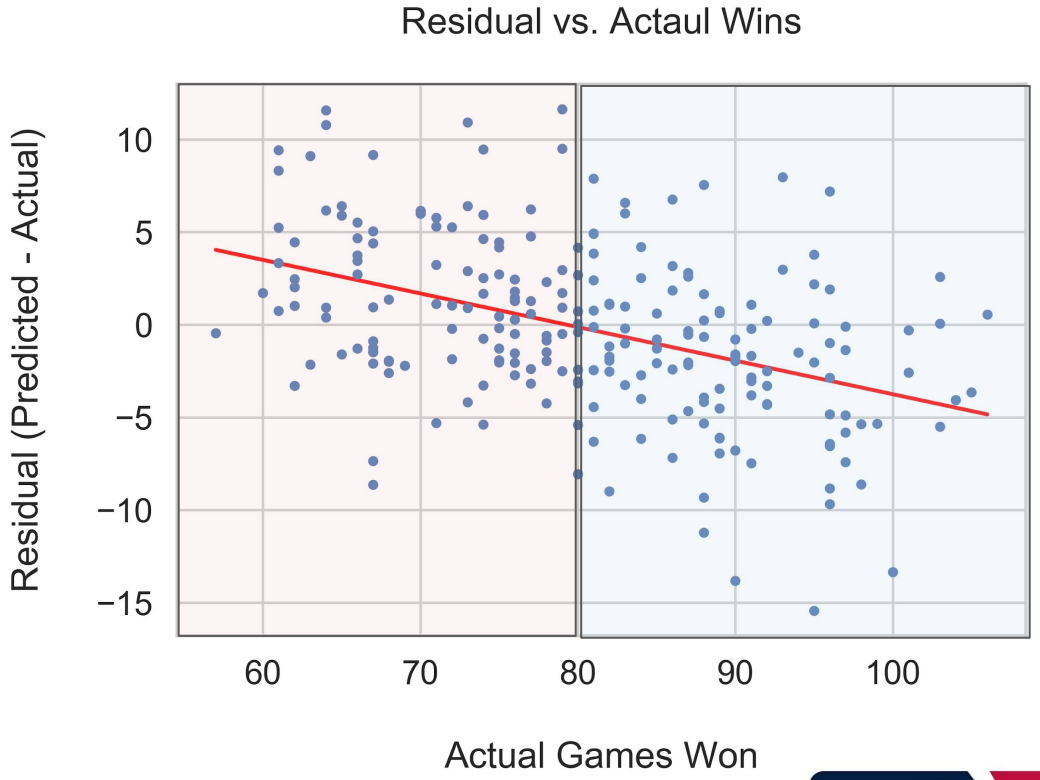
Pure Statistics

Studying the Correlation Between Different Statistics and Total Wins



Initial Model

- The Residuals are Normally Distributed
- The Residuals Have a Negative Relationship with Actual Games Won
- Model is over-predicting the worse teams and under-predicting the better teams



Engineered Features

Manipulating The Pure Stats

The Equations

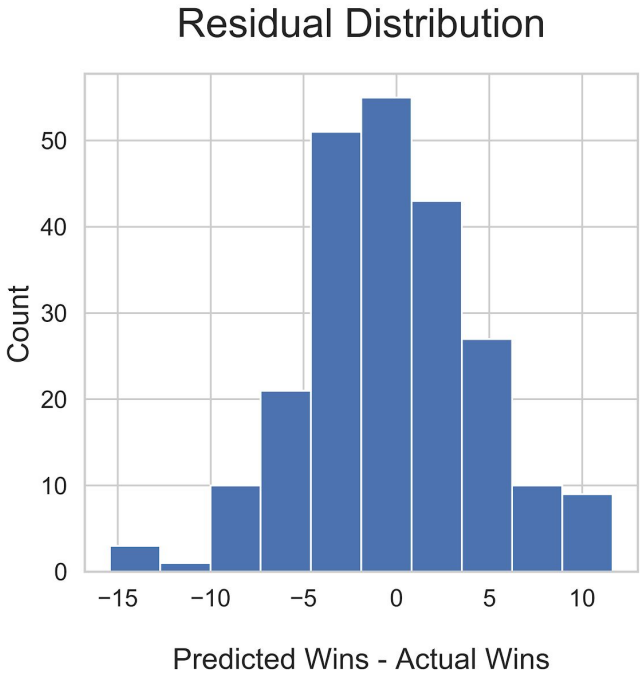
$$\text{Batting Average (BA)} = \frac{\text{(Hits)}}{\text{At-Bats}}$$

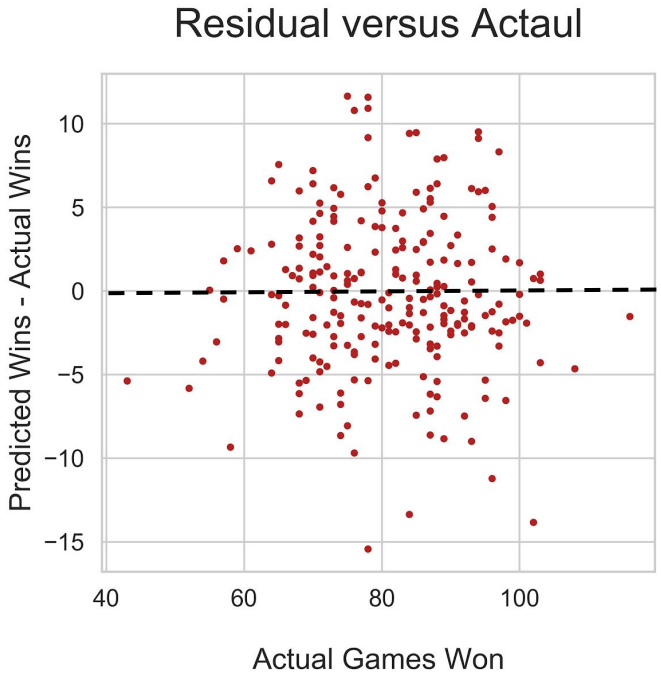
$$\text{On-Base-Percentage (OBP)} = \frac{\text{(Walks + Hits)}}{\text{Plate Appearances}}$$

$$\text{Slugging Percentage (SLG)} = \frac{\text{Singles + (2 X Doubles) + (3 X Triples) + (4 X HR)}}{\text{At-Bats}}$$

Analyzing and Evaluating the Final Model

Final Residuals



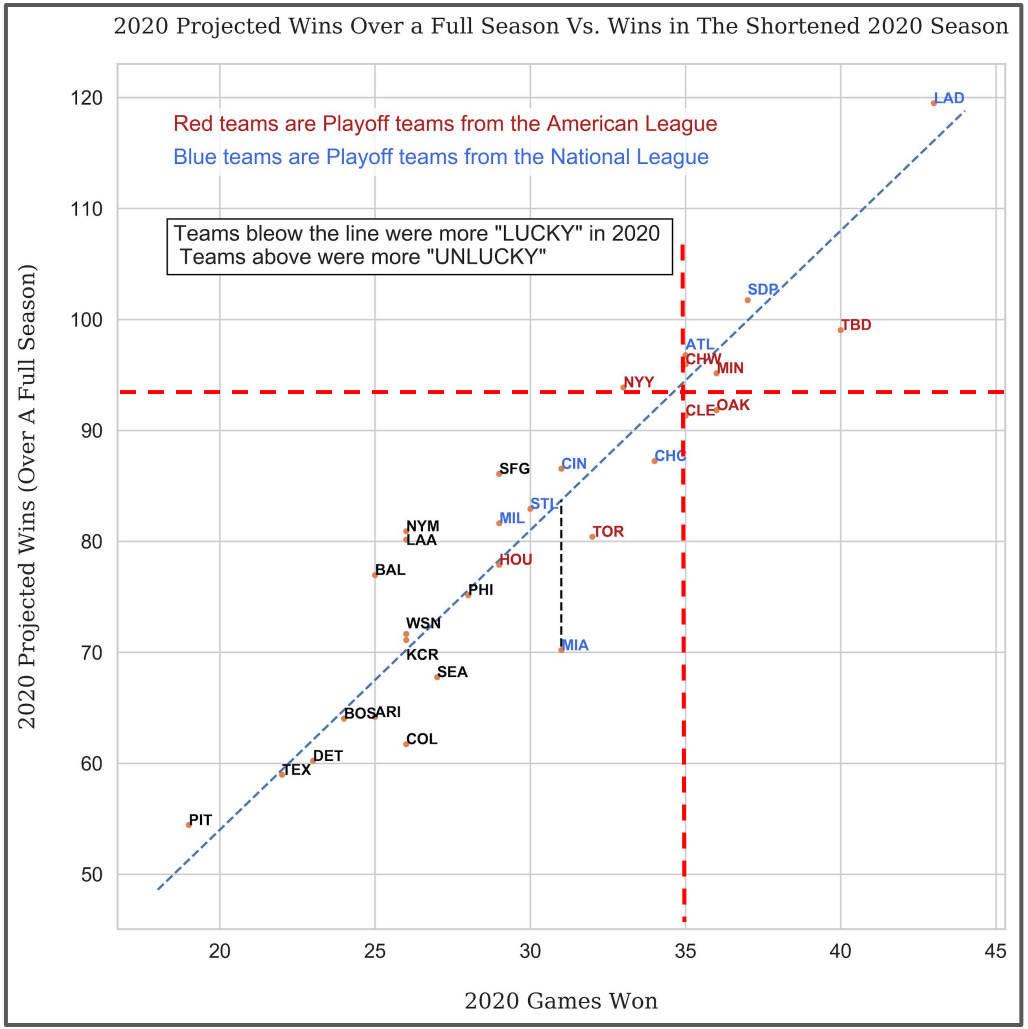


- On Average the model's prediction was within 4 wins of the target value
- 85% of the Predictions were within 6.5 wins

So What About 2020?

Key Insights

- The expanded playoffs were made for a team like the Yankees.
- The Marlins got VERY Lucky.
- If there was a “Loser” it was the Dodgers



Questions?

Scaled “Engineered” Model

- Coefficients of a scaled model tell us how much “weight” our features have
- Batting Average looks strange. Caused by Collinearity

Feature	Coefficient
Batting Average	-0.35
On-Base-Percentage	4.65
Slugging Pct.	5.08
Batting Average Against	0.66
On-Base-Percentage Against	-5.55
Slugging Pct. Against	-4.96