

MLB Win Prediction

By: Mathew Katz



The Question We Seek to Answer:

Is there a way to predict how many wins each team in the MLB will have in the upcoming season? It could be very beneficial to a sports gambler but also interesting to the common fan.



Why predict Wins?

*"If you're going to play at all, you're out to win.
Baseball, board games, Jeopardy, I hate to lose."* -
Derek Jeter

WORLD CHAMPIONSHIP RINGS THROUGHOUT HISTORY





We will be using our old friend, Python.

Let's find the statistics highly correlated to wins!

SABERMETRICS, SCOUTING AND THE SCIENCE OF BASEBALL

A weekend baseball seminar featuring presentations from some of baseball's top statisticians, scouts and sports writers. All proceeds will be donated to the Jimmy Fund.

Join us May 21 and 22 at Harvard's Pfizer Lecture Hall. Register for the event and a chance to win 2 Red Sox-Cubs tickets at saberseminar.com.

wOBA Weighted on-base average Babe Ruth .500	War Wins above replacement Babe Ruth 190	Lwts Linear weights R. Henderson .2295	Ops+ Adjusted on-base + slugging Babe Ruth 206	Rs Runs Scored R. Henderson .2295	Hr Home runs Pete Rose 4256	Rbi Runs batted in Barry Bonds 762	Gp Games played Vance Anton 2297	Sb Stolen bases R. Henderson 1406	Br+ Adjusted batting runs Babe Ruth 1382	0w% Offensive wins percentage Babe Ruth .858	Sec Secondary average Barry Bonds .306	O Outs made Pete Rose 10,328	C% Contact percentage Barry Bonds .306	Uzr Ultimate zone rating E Pete Rose 5.03	+/- Plus/minus Jake Beckley .241	K Strikeouts Nolan Ryan 5714	Gs Games started Cy Young .815	Era+ Adjusted earned run average Pedro Martinez 1.816
Ops On-base + slugging Babe Ruth 1.1638	H Hits Hank Aaron 366	Tb Total bases Hank Aaron 6856	Ave Batting average Ty Cobb .366	3b Triples Sam Crawford 309	Li Leverage index Pete Rose 14.03	Ab At bats Pete Rose 15,863	Cs Caught stealing R. Henderson 335	Bxh Adjusted batting wins Babe Ruth 129.86	Xbh Extra base hits Mark Aaron 1477	Raa Runs above average Pete Rose 5.03	Gidp Grounded into double play Cal Ripken Jr. 350	Os% Outside swing percentage Pete Rose 10.328	Babip Batting average on balls in play Barry Bonds .306	A Assists R. Maravilla 8951	K/Bb Strikeout / walk Tunney Bond 5.03	Sho Shutouts Walter Johnson 110	K/9 Strikeouts / 9 innings pitched Nolan Ryan 5555	
Obp On-base percentage Tris Williams .4817	Rc Runs created Barry Bonds 2892	2b Doubles Tris Speaker 792	Wpa Win probability added Barry Bonds 127.4	Pa Plate appearances Pete Rose 15,863	Sh Sacrifice hits Eddie Collins 512	K Strikeouts R. Jackson 2597	Tob Times on-base Pete Rose 5.03	Lob Left on-base Pete Rose 5.03	Ab/Hr Walks per home run Mark McGwire 10.8	Ab/Hr Walks per home run Joe Sewell 1.6	Go/Ao/ Grounded into double play percentage Cy Young 29.565	Bf Batters faced Arthur Rhodes 215	H Hits C. Cummings 4731	H9 Hits / 9 innings pitched Cy Young 511	Bb/9 Baseball errors / 9 innings pitched C. Cummings 4731			
Sig Slugging percentage Babe Ruth .8803	Rp Runs produced Ty Cobb 4056	1b Singles Pete Rose 3215	Iso Isolated Power Barry Bonds .348	Ppa Pitches / plate appearance Eddie Murray 128	Sf Sacrifice flies Eddie Collins 512	Bb Bases on balls Barry Bonds 2558	Ibb Intentional bases on balls Barry Bonds 688	Hbp Hit by pitch H. Jennings 287	Ab/K At-bats per strikeout Willie Keeler 63.11	SwSt% Swinging-strike percentage Gus Wehring 277	Fs% First pitch strike percentage Gus Wehring 277	Bh Batters hit Tony Moline 343	Wp Wild pitches Cy Young 7356	Ip Innings pitched Cy Young 2147	Er Earned runs Cy Young 2147			

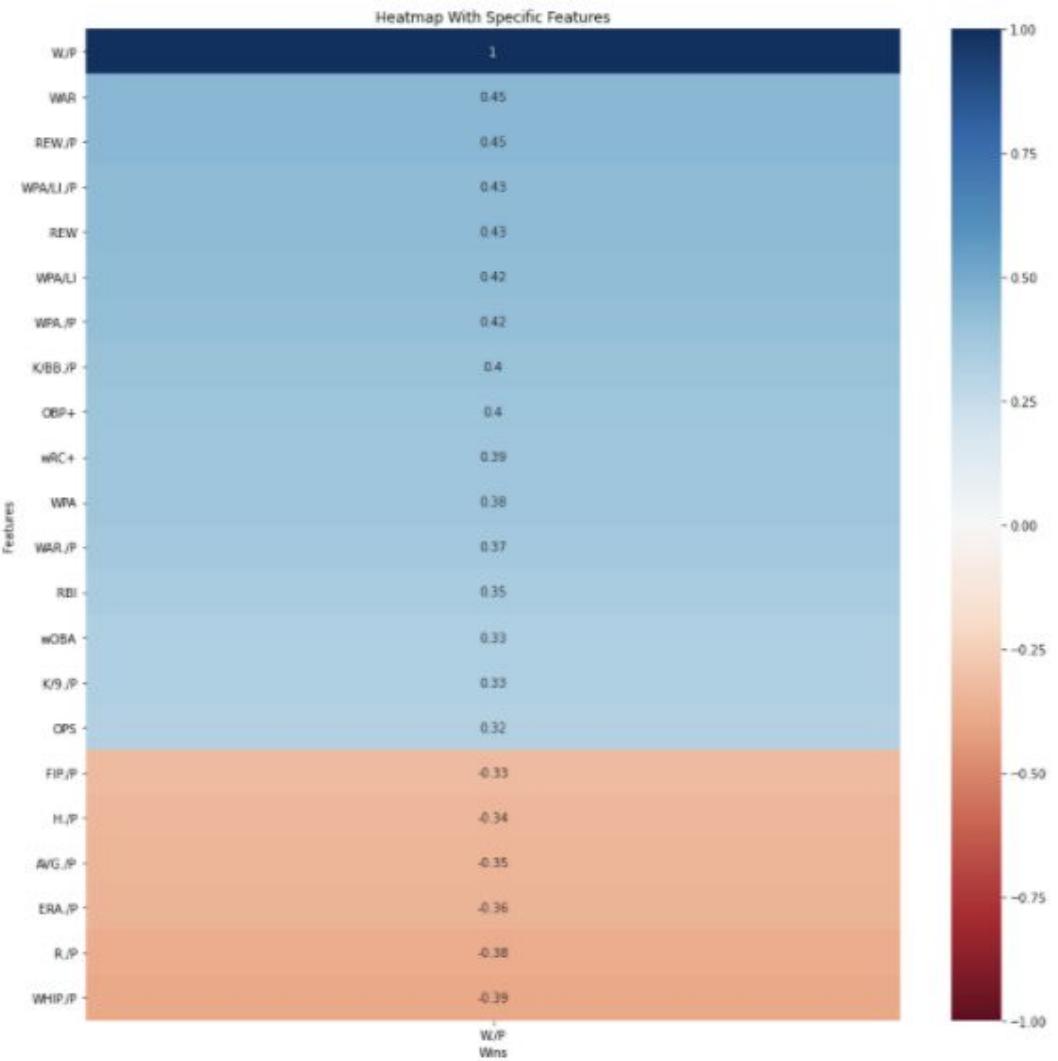
©John Kenyon 2011, john-kenyon.com

More information at saberseminar.com

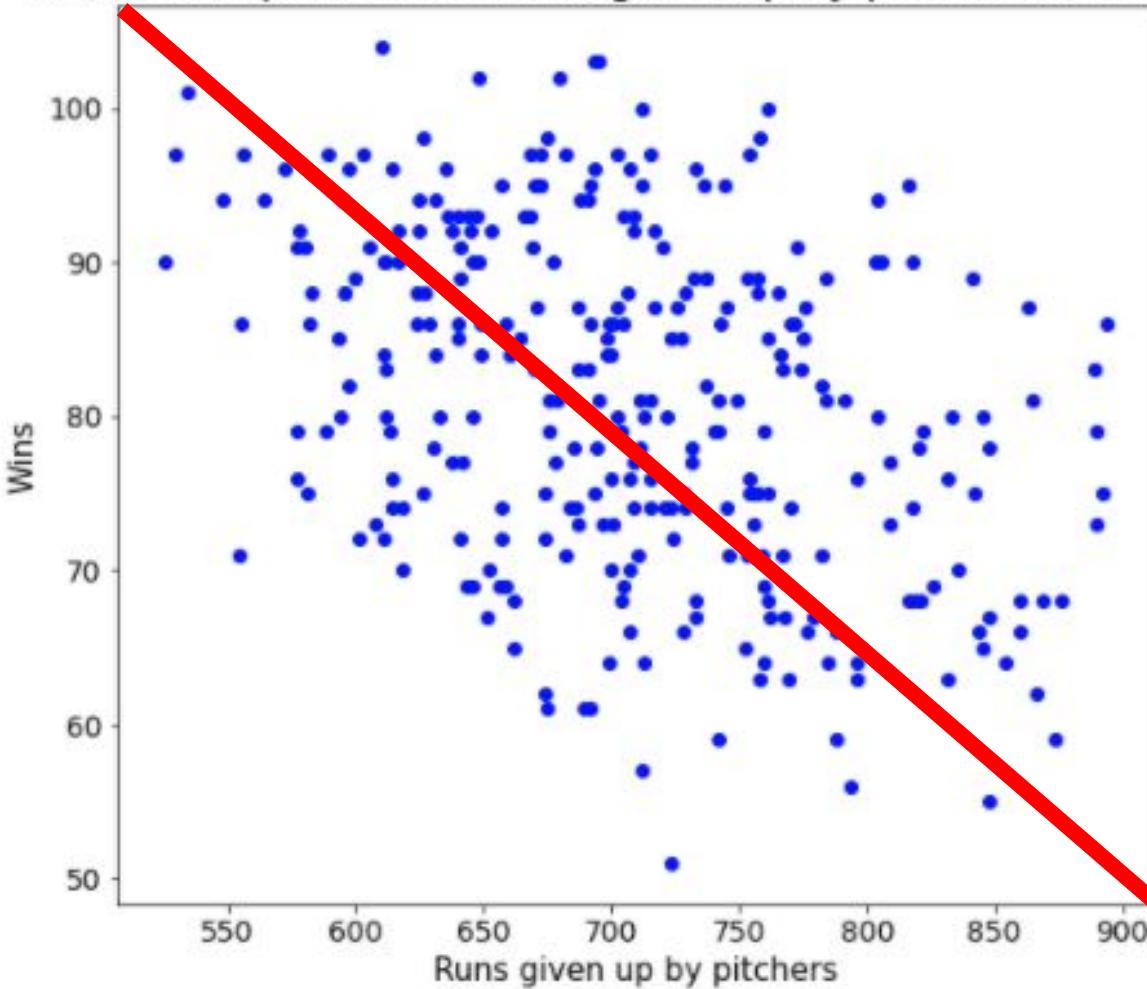


527 Statistics

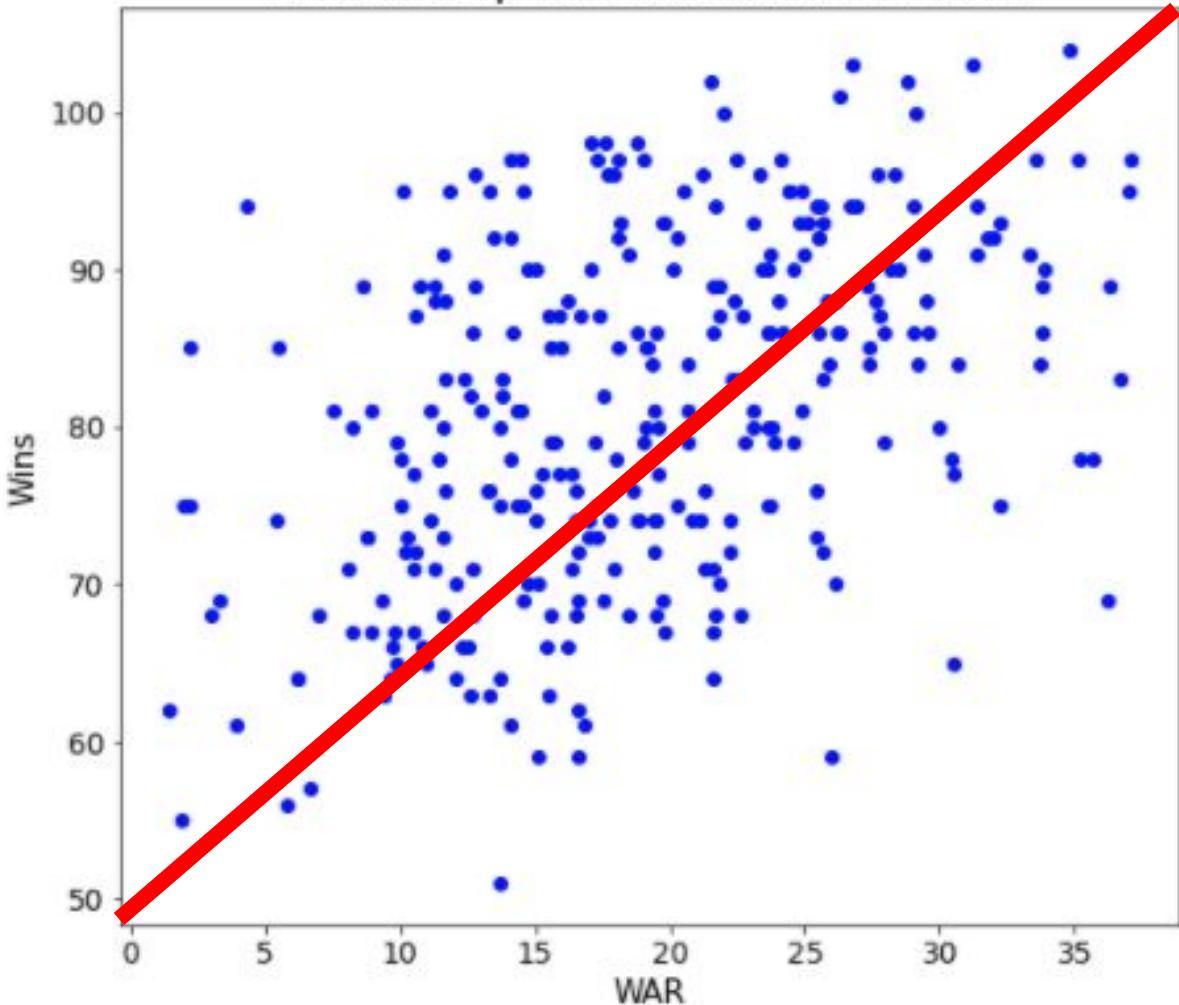
22 Statistics



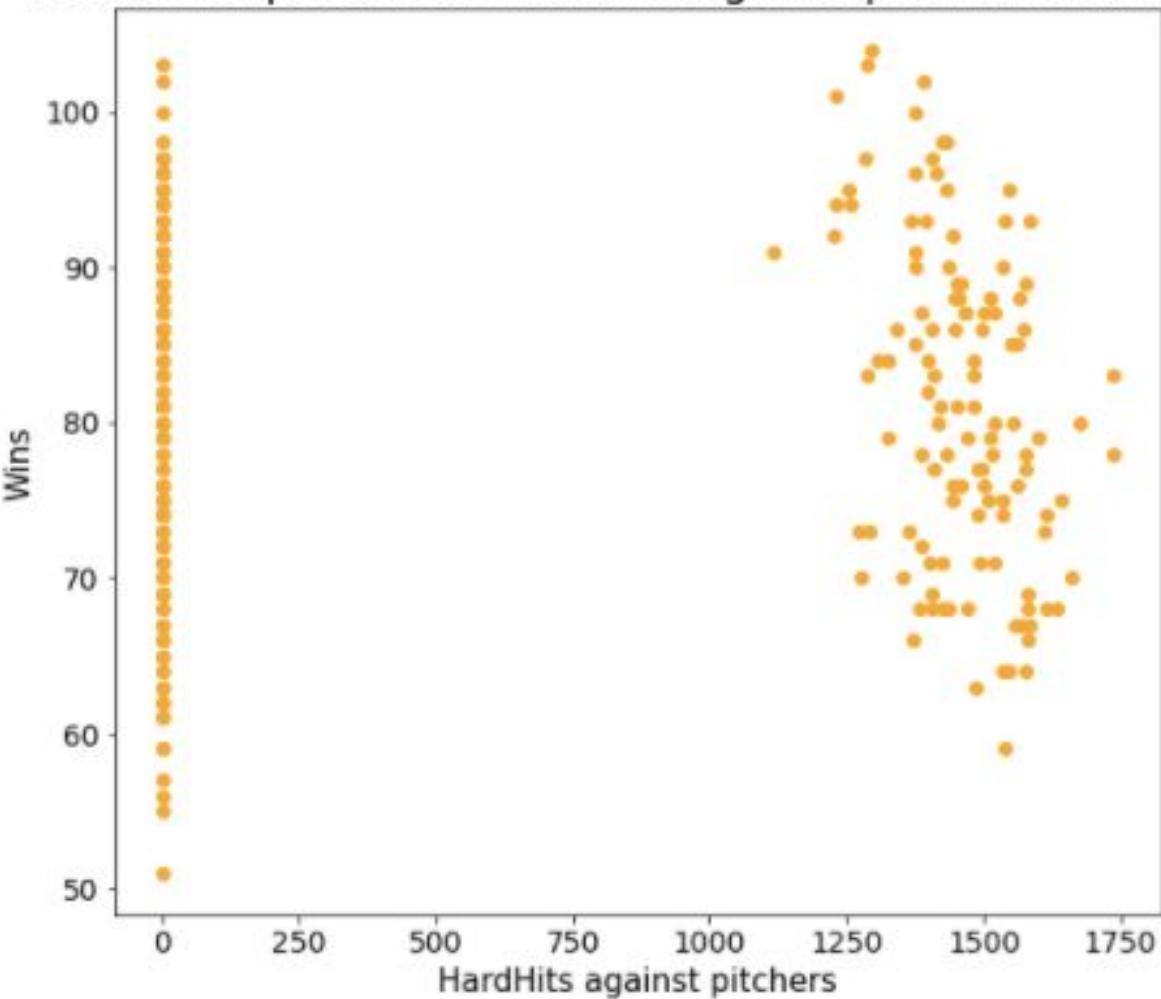
Relationship Between Runs given up by pitchers and Wins



Relationship Between WAR and Wins



Relationship Between HardHits against pitchers and Wins





Jacob deGrom

Popularity: ↓ 21%

P | B/T: L/R | 6' 4" 180LBS | Age: 32

Draft: 2010 | Rd: 9, #272, New York Mets | Stetson

2020 MLB Percentile Rankings

Exit Velocity

52

Hard Hit %

69

xWOBA

93

xERA

93

xBA

87

xSLG

80

Barrel %

19

K %

97

BB %

75

Whiff %

98

Fastball Velocity

100

Fastball Spin

90

Curve Spin

68

Poor Avg Great



'K/9', 'ERA', 'FIP', 'WAR', 'H', 'R', 'K/BB', 'AVG ', 'WHIP', 'WPA', 'REW',
'WPA/LI', 'RBI', 'wOBA', 'wRC+', 'WAR', 'OPS', 'WPA', 'REW', 'WPA/LI', 'OBP+'

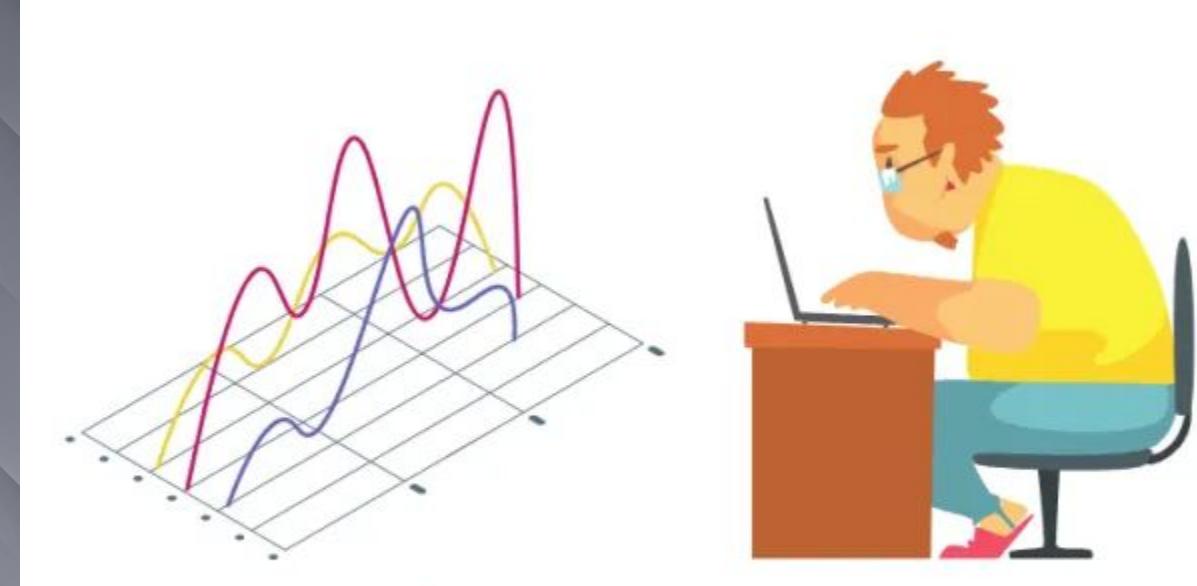


Baseline Model

Using the DummyRegressor from sklearn, it predicted the mean of the training target values. The RMSE score for the baseline model was a 13.06166647 RMSE. I knew I could beat that score.



Ridge Regression Model



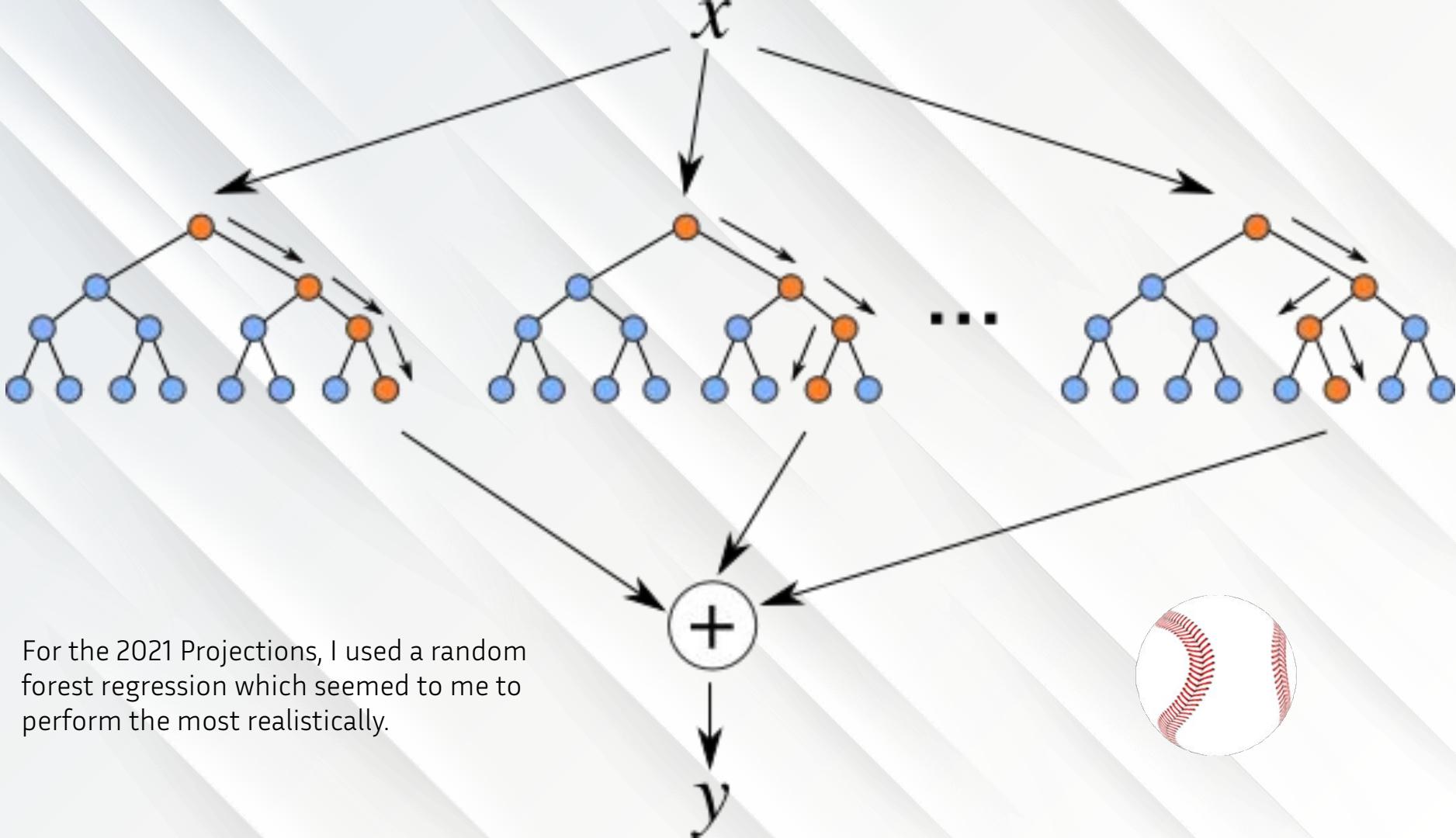
Ridge RMSE Score: 8.014876576130929

2019 Projected Wins v.s. Actual

The average difference between the Model Projections and the Actual Wins is 5.3 wins. That was the closest I could get.

Most team win projections were actually closer but due to outliers like the Marlins, Tigers, and the Twins, it became a higher average.

	W./P	Projected_Wins	Projected_-_Actual
Team			
Angels	72	77.194286	5.194286
Astros	107	101.371429	-5.628571
Athletics	97	93.988571	-3.011429
Blue Jays	67	71.240000	4.240000
Braves	97	88.182857	-8.817143
Brewers	89	88.834286	-0.165714
Cardinals	91	84.514286	-6.485714
Cubs	84	85.845714	1.845714
Diamondbacks	85	82.525714	-2.474286
Dodgers	106	101.491429	-4.508571
Giants	77	74.960000	-2.040000
Indians	93	94.057143	1.057143
Mariners	68	71.680000	3.680000
Marlins	57	70.417143	13.417143
Mets	86	83.428571	-2.571429
Nationals	93	91.417143	-1.582857
Orioles	54	63.342857	9.342857
Padres	70	71.171429	1.171429
Phillies	81	79.640000	-1.360000
Pirates	69	74.474286	5.474286
Rangers	78	74.217143	-3.782857
Rays	96	93.931429	-2.068571
Red Sox	84	92.262857	8.262857
Reds	75	74.251429	-0.748571
Rockies	71	73.548571	2.548571
Royals	59	62.782857	3.782857
Tigers	47	68.948571	21.948571
Twins	101	73.497143	-27.502857
White Sox	72	72.320000	0.320000
Yankees	103	98.897143	-4.102857



2021 MLB Projected Standings



AL East		NL East	
Team	Projected_Wins	Team	Projected_Wins
Yankees	91.5	Braves	88.5
Rays	89.0	Mets	86.5
Blue Jays	76.5	Phillies	81.0
Orioles	76.5	Nationals	80.5
Red Sox	76.5	Marlins	71.0

AL Central		NL Central	
Team	Projected_Wins	Team	Projected_Wins
Twins	91.0	Cubs	84.5
Indians	88.0	Brewers	84.5
White Sox	85.0	Reds	84.0
Royals	74.5	Cardinals	83.5
Tigers	67.5	Pirates	71.5

AL West		NL West	
Team	Projected_Wins	Team	Projected_Wins
Athletics	88.5	Dodgers	99.0
Astros	86.5	Padres	87.5
Angels	81.0	Giants	81.5
Mariners	74.5	Diamondbacks	76.0
Rangers	70.5	Rockies	69.0

Improvements





Thanks for
listening
everyone!

Are there
any
questions?