

An aerial photograph of a baseball field. The green grass of the outfield is visible, along with the brown dirt of the infield. A pitcher in a white uniform is on the mound, and a batter in a blue uniform is at home plate. The text "Predicting length of MLB careers based on minor league stats" is overlaid in orange on the left side of the image.

Predicting length of MLB careers based on minor league stats

Samir Thanedar

Why is this important?

General Managers need to make many decisions:

- Who are my best players now?
- What will my team look like in 5 years?
- What are my team's biggest weaknesses?
- Who should I trade from my team? & Who should I trade for?
- Who should I not trade?
 - Babe Ruth, Jeff Bagwell & Miguel Cabrera were all traded before the age of 25 and became great.

Methodology/Workflow

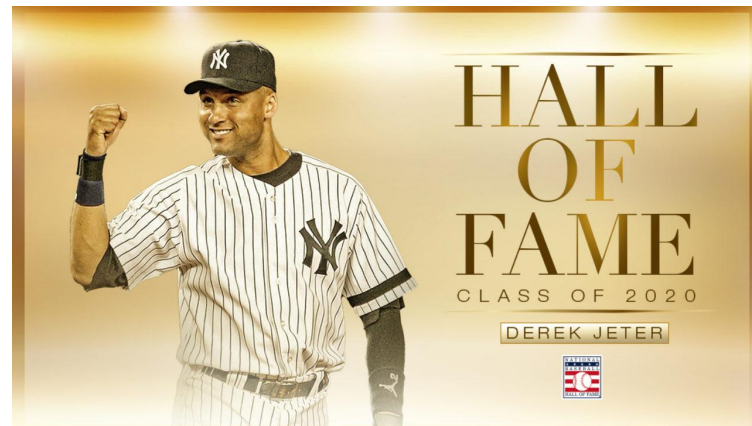
- Data was scraped from baseball-reference.com
- Data includes 1151 batters (no pitchers!)
 - Data includes all players who were playing some sort of professional baseball in 2000.
- Conclusions derived from 3 Linear regression Models: Lasso, Ridge, OLS

Target variable: MLB At-bats Per Season

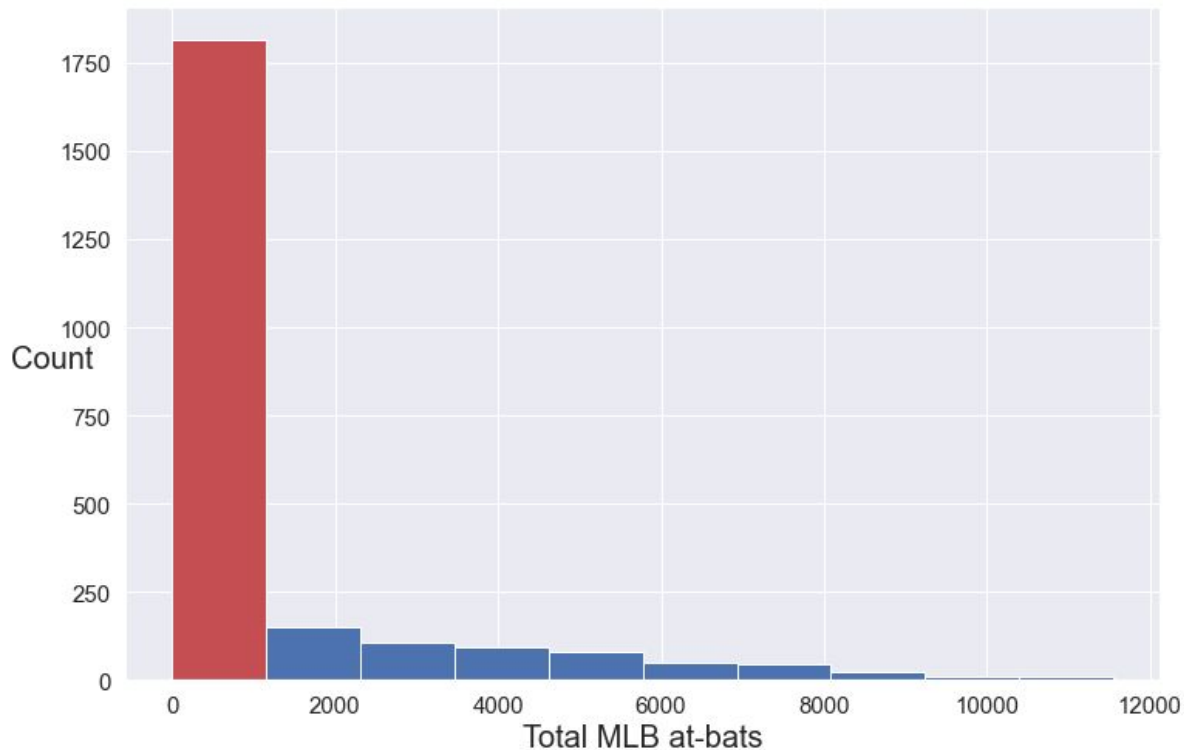
- How good a player is not easily encapsulated in a single statistic
- GMs want players that will not only be good but be consistently good for a long time.
- At-bats is a good way to measure all-around good players.



?



Initial Roadbump: most players don't make the MLB



Findings and Conclusions

Final R squared: 0.40

Three statistics had the highest correlation with MLB ABs:

- 1. Strikeouts per AB**
- 2. OPS**
- 3. Runs per AB**

Feature	Impact Rank
Debut Age	1
Strikeouts per AB	2
OPS	3
Runs per AB	4
School	5
Draft	6
BMI	7
Position = Catcher	8
Position = Shortstop	9
Bats = Both	10
Position = First Baseman	11
Position = Second Baseman	12
Bats = Left	13
Position = Outfielder	14
Bats = Right	15
Position = Third Baseman	16

Findings and Conclusions Part 2

It was just as interesting to find out that the following had low impact:

1. Round player was drafted
2. Player position
3. Left handed or right handed

Feature	Impact Rank
Debut Age	1
Strikeouts per AB	2
OPS	3
Runs per AB	4
School	5
Draft	6
BMI	7
Position = Catcher	8
Position = Shortstop	9
Bats = Both	10
Position = First Baseman	11
Position = Second Baseman	12
Bats = Left	13
Position = Outfielder	14
Bats = Right	15
Position = Third Baseman	16

Future Work

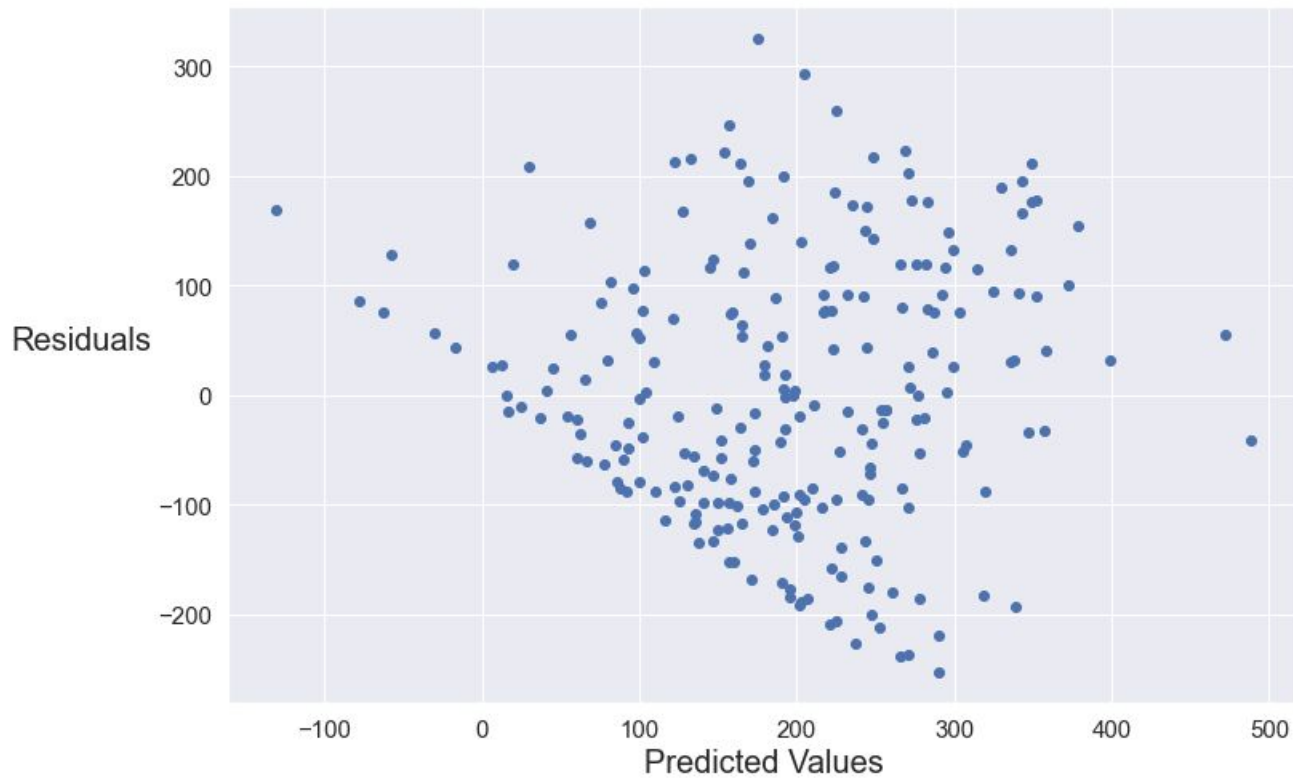
1. Make a classification model for whether or not player will make it to pros
2. Add fielding stats.
3. Apply this pitchers as well

An aerial photograph of a baseball field during a game. The pitcher's mound is in the upper right, with a pitcher in a white uniform with red accents in mid-throw. The batter's box is in the lower right, with a batter in a blue uniform swinging a bat. Behind the batter are the catcher and umpire. The infield is reddish-brown dirt, and the outfield is green grass. White lines mark the bases and pitcher's mound. The word "Questions?" is superimposed in the center of the field.

Questions?

Appendix

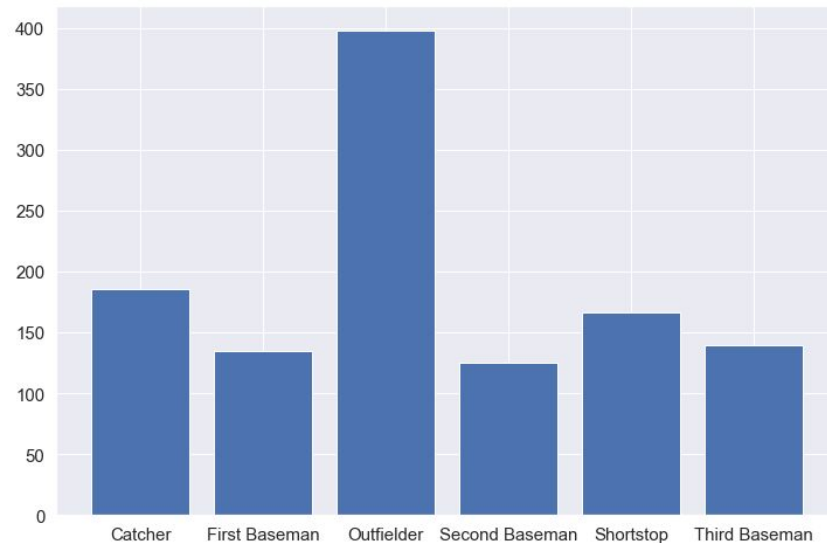
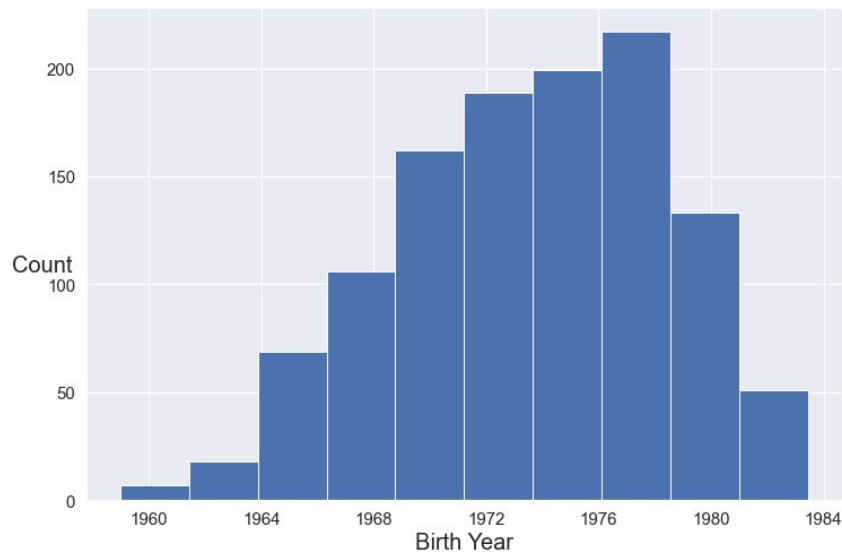
Residual Plot



Summary of Regression Results

Data with Absolute Value Coefficients									
Feature	Lasso	Ridge	OLS	Lasso Rank	Ridge Rank	OLS Rank	Average Rank	Average Coefficient	Rank of Coeff
Strikeouts per AB	39.33956444	39.3433013	749.7253891	2	2	2	2	276.136085	1
Runs per AB	17.32502936	17.32597847	764.6245963	4	4	1	3	266.4252014	2
OPS	25.95188962	25.95565996	334.0712593	3	3	6	4	128.659603	3
Bats = Both	3.596741472	2.953023042	373.929191	9	10	3	7.333333333	126.8263185	4
Bats = Left	0.484165822	1.315538935	370.1571529	15	13	4	10.66666667	123.9856192	5
Bats = Right	0	0.909951191	363.3325669	16	15	5	12	121.4141727	6
Position = Shortstop	3.229632735	3.185978124	197.4597006	10	9	7	8.666666667	67.95843715	7
Position = Second Baseman	1.806553401	1.76862869	187.6827176	12	12	8	10.66666667	63.75263323	8
Position = First Baseman	2.506231266	2.467183741	186.2457017	11	11	9	10.33333333	63.73970556	9
Position = Third Baseman	0.6844002821	0.7357271305	183.8767113	14	16	10	13.33333333	61.7656129	10
Position = Outfielder	0.8527063189	0.9215998561	182.4739026	13	14	11	12.66666667	61.41606959	11
Position = Catcher	4.80070234	4.855989016	169.680177	8	8	12	9.333333333	59.77895613	12
Debut Age	71.48177651	71.48616173	30.26774724	1	1	14	5.333333333	57.74522849	13
School	16.77717591	16.78983972	36.00470411	5	5	13	7.666666667	23.19057325	14
BMI	10.23882379	10.24652659	5.322494446	7	7	15	9.666666667	8.602614939	15
Draft	11.1728033	11.18358342	0.8849306858	6	6	16	9.333333333	7.747105802	16

Understanding Our Sample



Actual vs. Predicted Values

