

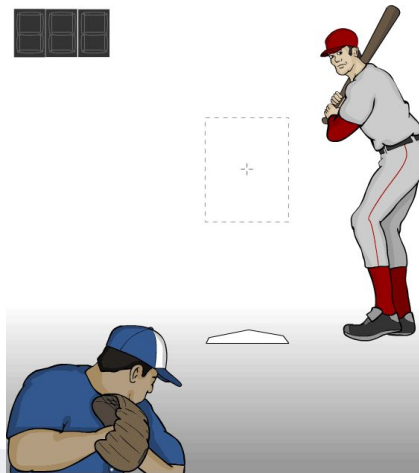
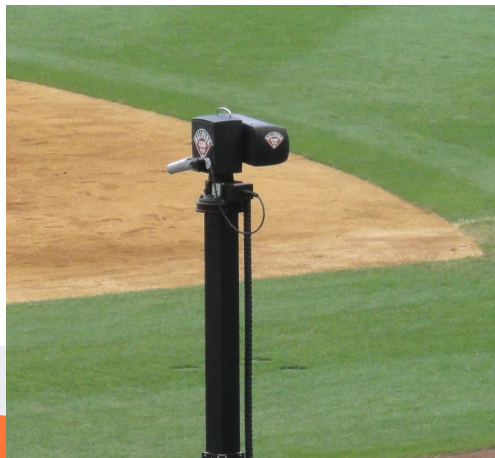


# ***STATS TO EVALUATE THE PERFORMANCE OF BATTERS***

***BY CHIEN YUAN CHANG 10-1-2021***

# INTRODUCTION - MOTIVATION

- Statcast
  - Introduced to all 30 MLB stadiums in 2015



# INTRODUCTION - OBJECTIVE/GOALS

- Objectives
  - A Linear Regression model to interpret the factors and predict the performances of the batters
- Goals
  - Baseball franchises and fantasy baseball players
    - ▶ Evaluate the batters
  - The coaches and players
    - ▶ Adjust hitting types and improve performances

# ***METHODOLOGY - OVERVIEW***

- Data
  - 2015-2019 Batters with at least 162 PA in a single season
- Tools
  - Python: BeautifulSoup, SQLAlchemy, Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Statsmodels, Scipy
  - SQLite 3

## **TARGET - ON-BASE PLUS SLUGGING PLUS (OPS+)**

- What is OPS (on-base plus slugging)?
  - OBP(on-base percentage) + SLG (slugging percentage)
- What is OPS+?
  - $(OPS / \text{league OPS, adjusted for park factors}) \times 100$
- Why OPS+?
  - Easy to understand
  - Representative of important offensive skills
  - Normalized, mean = 100



# WEB-SCRAPING FROM BASEBALL REFERENCE

1. Scraping the list of 2015-2019 players and the addresses of their player pages → 2. Scraping the pages of each player

2019 MLB Season Minor Leagues Standings Schedule Fielding Batting Pitching Managers Leaders & Awards Other Back to top

Player Standard Batting Share & Export When table is sorted by rate stats, hide non-qualifiers (min. 3.1 PA/GlgAvg) Glossary Hide Partial Rows

| Rk | Name              | Age | Tm  | Lg  | G   | PA  | AB  | R   | H   | 2B | 3B | HR | RR  | SB   | CS   | BB   | SO   | BA   | OBP  | SLG  | OPS  | OPS+ | TS  | GDP | HBP | SH | SF | IBB | Pos | Summary   |
|----|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|-----|------|------|------|------|------|------|------|------|------|-----|-----|-----|----|----|-----|-----|-----------|
| 1  | Fernando Abad*    | 33  | SEA | NL  | 18  | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 2  | Jose Abreu        | 32  | CHW | AL  | 159 | 693 | 634 | 85  | 180 | 38 | 1  | 33 | 123 | 2    | 2    | 36   | 152  | 284  | .330 | .503 | .834 | 118  | 319 | 24  | 13  | 0  | 10 | 4   |     | *3D       |
| 3  | Ronald Acuna Jr.  | 21  | ATL | NL  | 156 | 715 | 626 | 127 | 175 | 22 | 2  | 41 | 101 | 37   | 9    | 76   | 188  | 280  | .365 | .518 | .883 | 121  | 324 | 8   | 9   | 0  | 1  | 4   |     | *879/H    |
| 4  | Jason Adam        | 27  | TBR | AL  | 3   | 9   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 5  | Cristhian Adames* | 10  | 24  | 22  | 1   | 7   | 1   | 0   | 0   | 2  | 0  | 2  | 8   | .318 | .375 | .364 | .739 | 100  | 8    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 6  | Willy Adames      | 23  | TBR | AL  | 152 | 584 | 531 | 69  | 135 | 25 | 1  | 20 | 52  | 4    | 2    | 46   | 153  | 254  | .317 | .418 | .735 | 96   | 222 | 9   | 3   | 3  | 1  | 1   |     | *6/H      |
| 7  | Austin Adams      | 32  | TOT | AL  | 1   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 8  | Austin Adams      | 32  | DET | AL  | 1   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 9  | Austin Adams      | 28  | TOT | MLB | 4   | 1   | 1   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 10 | Austin Adams      | 28  | WSN | NL  | 1   | 0   | 0   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 11 | Austin Adams      | 28  | SEA | AL  | 3   | 1   | 1   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 12 | Matt Adams*       | 30  | WSN | NL  | 111 | 333 | 310 | 42  | 70  | 14 | 0  | 20 | 56  | 0    | 0    | 20   | 115  | .226 | .276 | .465 | .741 | 86   | 144 | 7   | 2   | 0  | 1  | 1   |     | 3H        |
| 13 | Jim Adcox*        | 34  | CIN | AL  | 2   | 5   | 5   | 0   | 0   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 14 | Eliane Adrianza*  | 29  | MIN | AL  | 83  | 236 | 202 | 34  | 55  | 8  | 3  | 5  | 22  | 0    | 2    | 20   | 40   | .272 | .349 | .416 | .765 | 103  | 84  | 2   | 6   | 2  | 4  | 1   |     | 563H/497I |
| 15 | Daric Adreani     | 24  | DET | NL  | 14  | 24  | 22  | 0   | 2   | 0  | 0  | 0  | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0   | 0   | 0   | 0  | 0  | 0   | 0   | 1         |
| 16 | Jesús Aguilar     | 29  | TOT | MLB | 131 | 369 | 314 | 39  | 74  | 12 | 0  | 12 | 50  | 0    | 0    | 43   | 81   | .236 | .325 | .389 | .714 | 87   | 122 | 12  | 2   | 0  | 7  | 0   |     | 3HD/5     |
| 17 | Jesús Aguilar     | 29  | TEL | NL  | 94  | 262 | 223 | 26  | 50  | 0  | 0  | 8  | 34  | 0    | 0    | 31   | 59   | .225 | .300 | .374 | .694 | 80   | 83  | 11  | 2   | 0  | 4  | 0   |     | 3H/5      |
| 18 | Jesús Aguilar     | 29  | TBR | AL  | 37  | 107 | 92  | 13  | 24  | 3  | 0  | 4  | 16  | 0    | 0    | 12   | 22   | .261 | .336 | .424 | .760 | 104  | 39  | 1   | 0   | 0  | 3  | 0   |     | 3D/H      |
| 19 | Nick Ahmed        | 29  | ARI | NL  | 158 | 625 | 556 | 79  | 141 | 33 | 6  | 19 | 82  | 8    | 2    | 52   | 113  | .254 | .316 | .437 | .753 | 93   | 243 | 15  | 4   | 1  | 12 | 2   |     | *6/H      |

Standard Batting Show Minors Games by Position Share & Export Glossary

| Year          | Age | Tm  | Lg | G    | PA   | AB   | R   | H    | 2B  | 3B | HR  | RR  | SB | CS | BB  | SO  | BA   | OBP  | SLG  | OPS  | OPS+ | TS   | GDP | HBP | SH | SF | IBB | Pos | Awards                  |
|---------------|-----|-----|----|------|------|------|-----|------|-----|----|-----|-----|----|----|-----|-----|------|------|------|------|------|------|-----|-----|----|----|-----|-----|-------------------------|
| 2014          | 27  | CHW | AL | 145  | 622  | 556  | 80  | 176  | 35  | 2  | 36  | 107 | 3  | 1  | 51  | 131 | .317 | .383 | .581 | .964 | 173  | 323  | 14  | 11  | 0  | 4  | 15  |     | *3D/H AS,MVP-5,ROY-1,SS |
| 2015          | 28  | CHW | AL | 154  | 668  | 613  | 88  | 178  | 34  | 3  | 30  | 101 | 0  | 0  | 39  | 140 | .290 | .347 | .502 | .850 | 135  | 308  | 16  | 15  | 0  | 1  | 11  |     | *3D/H MVP-21            |
| 2016          | 29  | CHW | AL | 159  | 695  | 624  | 87  | 183  | 32  | 1  | 25  | 100 | 0  | 2  | 47  | 125 | .293 | .353 | .468 | .820 | 124  | 292  | 21  | 15  | 0  | 9  | 7   |     | *3D                     |
| 2017          | 30  | CHW | AL | 156  | 675  | 621  | 95  | 189  | 43  | 6  | 33  | 102 | 3  | 0  | 35  | 119 | .304 | .354 | .552 | .906 | 141  | 343  | 21  | 15  | 0  | 4  | 6   |     | *3D/H MVP-14            |
| 2018          | 31  | CHW | AL | 128  | 553  | 499  | 68  | 132  | 36  | 1  | 22  | 78  | 2  | 0  | 37  | 109 | .265 | .325 | .473 | .798 | 117  | 236  | 14  | 11  | 0  | 6  | 7   |     | *3D/H AS,SS             |
| 2019          | 32  | CHW | AL | 159  | 693  | 634  | 85  | 180  | 38  | 1  | 33  | 123 | 2  | 2  | 36  | 152 | .284 | .330 | .503 | .834 | 118  | 319  | 24  | 13  | 0  | 10 | 4   |     | *3D AS,MVP-19           |
| 2020          | 33  | CHW | AL | 60   | 262  | 240  | 43  | 76   | 15  | 0  | 19  | 60  | 0  | 0  | 18  | 59  | .317 | .370 | .617 | .987 | 165  | 148  | 10  | 3   | 0  | 1  | 1   |     | *3D MVP-1,SS            |
| 2021          | 34  | CHW | AL | 190  | 650  | 558  | 85  | 146  | 29  | 2  | 29  | 113 | 1  | 0  | 60  | 141 | .262 | .351 | .477 | .827 | 125  | 266  | 28  | 22  | 0  | 10 | 3   |     | *3D/5                   |
| 1 Yrs         |     |     |    | 1111 | 4818 | 4345 | 611 | 1260 | 262 | 16 | 227 | 784 | 11 | 5  | 323 | 976 | .290 | .350 | .514 | .865 | 135  | 2235 | 148 | 105 | 0  | 45 | 54  |     |                         |
| 162 Same-Avg. |     |     |    | 162  | 793  | 634  | 89  | 184  | 38  | 2  | 33  | 114 | 2  | 1  | 47  | 142 | .290 | .350 | .514 | .865 | 135  | 226  | 22  | 15  | 0  | 7  | 8   |     |                         |

bold season totals indicate player led league. *italic* season totals indicate player led all major leagues.

[Player Value-Batting](#)
[WAR Explained \(v2.2\): 8+ MVP 5+ A-S, 2+ Starter, 0-2 Sub, < 0 Rep, < 0 R Share & Export](#)
[Glossary](#)

| Year | Age | Tm  | Lg | G   | PA  | Rbat | Rbaser | Rdp | Rfield | Rpos | RAA | WAR | Rep | RAR | WARwL% | cWAR | cWARwL% | Salary | Pos  | Awards                    |                                     |
|------|-----|-----|----|-----|-----|------|--------|-----|--------|------|-----|-----|-----|-----|--------|------|---------|--------|------|---------------------------|-------------------------------------|
| 2014 | 27  | CHW | AL | 145 | 622 | 53   | -2     | 0   | -10    | -9   | 32  | 3.6 | 22  | 54  | 5.8    | .523 | .521    | 6.8    | -2.0 | 64                        | \$7,000,000 *3D/H AS,MVP-5,ROY-1,SS |
| 2015 | 28  | CHW | AL | 154 | 668 | 26   | -3     | -1  | -3     | -10  | 9   | 1.2 | 23  | 3.5 | .506   | .506 | 3.7     | -1.3   | 36   | \$8,666,000 *3D/H MVP-21  |                                     |
| 2016 | 29  | CHW | AL | 159 | 695 | 22   | -3     | -3  | -1     | -9   | 6   | 0.7 | 34  | 3.1 | .504   | .504 | 3.2     | -1.0   | 31   | \$13,334,000 *3D/H MVP-14 |                                     |
| 2017 | 30  | CHW | AL | 156 | 675 | 35   | 1      | -2  | 1      | -9   | 26  | 2.6 | 24  | 49  | 4.9    | .516 | .516    | 4.8    | -0.8 | 48                        | \$10,625,000 *3D/H MVP-14           |

5. SQLite3



4. Data Cleaning by Pandas

3. CSV

# DATA FLOW

**65 COLUMNS, 1925 ROWS**

Stats of Standard Batting, Advanced Batting, and Ratio Batting  
Web-scraping from [Baseball Reference](#)

**97 COLUMNS, 1925 ROWS**

Data of Statcast and Swing and Pitches  
Download from [Baseball Savant](#)

**>125 COLUMNS, 1925 ROWS**

Data Engineering

**1540  
TRAINING DATA**

**385  
TEST DATA**

Data Splitting

**16 FEATURES**

Baseline Model  
 $\text{Adj. } R^2 = 0.435$

**TRAIN**

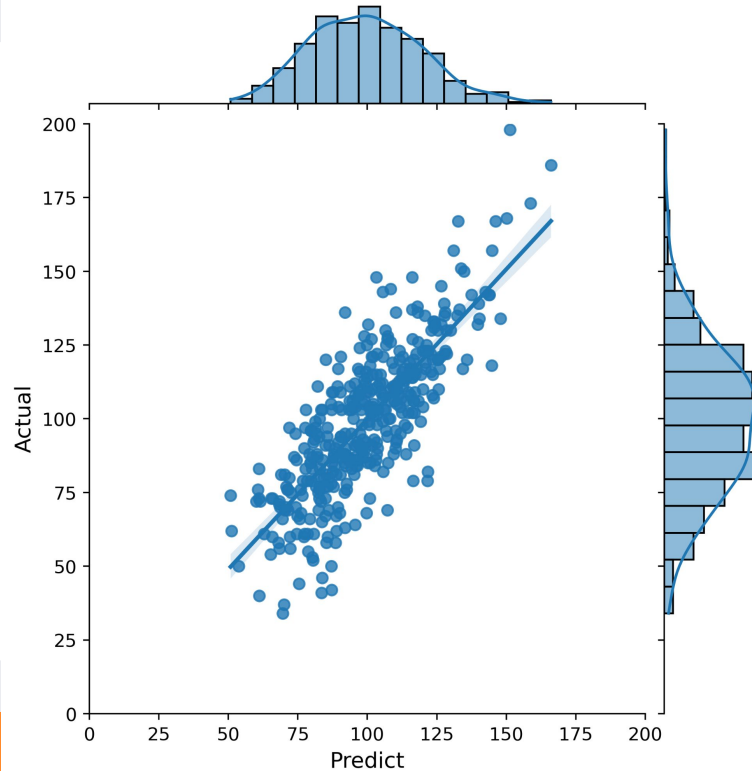
Training and Cross-Validation

**9  
FEATURES**

Final Model  
 $\text{Adj. } R^2 = 0.601$

# RESULT - OLS MODEL

- Test Data
  - Adj.  $R^2$ 
    - ▶ 0.626
  - $R^2$ 
    - ▶ 0.635
  - MAE
    - ▶ 11.963
  - RMSE
    - ▶ 15.367



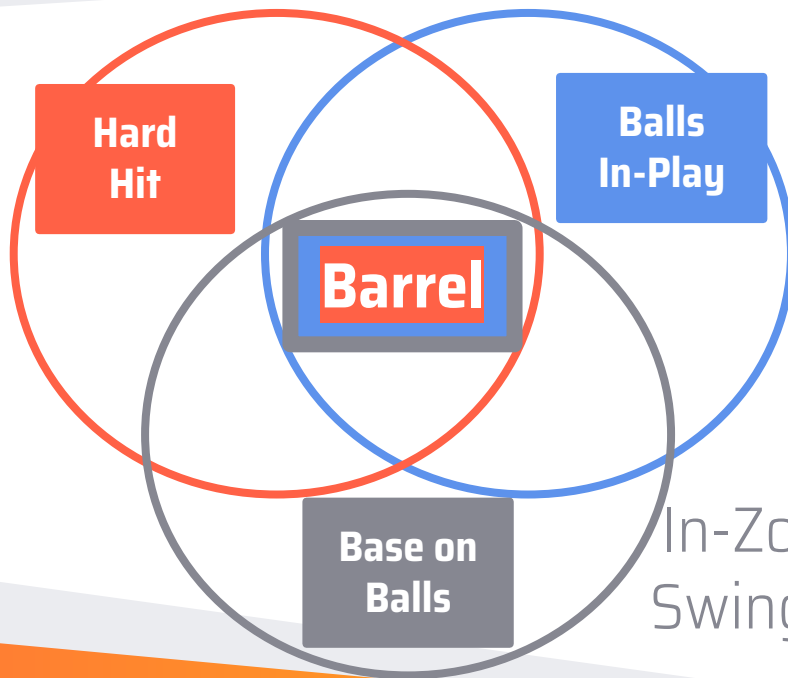


# RESULT

|   | Features                     | Coefficient | P-Value |       |       |
|---|------------------------------|-------------|---------|-------|-------|
|   | Constant                     | -41.53      | 0.001*  |       |       |
|   | Barrel Batted%               | 6.263       | <0.001* |       |       |
|   | Sprint Speed(ft/sec)         | 3.075       | <0.001* |       |       |
| → | Swing%                       | -1.568      | <0.001* |       |       |
| → | Z-Swing% * Z-Contact%        | 1.563       | <0.001* |       |       |
| → | SweetSpot% - Barrel Batted%  | 1.393       | <0.001* |       |       |
| → | Outside the Zone Swing&Miss% | -0.966      | <0.001* |       |       |
| → | Hard Hit% - Barrel Batted%   | 0.658       | <0.001* |       |       |
| → | Fly Ball%                    | -0.522      | <0.001* |       |       |
|   | Age - 17                     | -0.393      | 0.001*  |       |       |
|   | Comprehensive                | Eye         | Contact | Power | Other |

# CONCLUSIONS - STATCAST IS A GREAT TOOL

Power  
Higher Exit Velocity



Contact  
Swing and Contact  
With Right Angle  
(Sweet Spot)

Eye  
In-Zone/Out-Zone  
Swing/Don't Swing

Speed

Advantage

Age

Small Disadvantage

## ***FUTURE WORK***

- Test the model with 2020, 2021 and future data
- Dig into the swing data more
- Research the play-by-play data with the statcast
- Analyze pitcher data

# ***THANK YOU!***



This slide template is from SlidesCarnival and free to use under [Creative Commons Attribution license](#).



# ***APPENDIX***

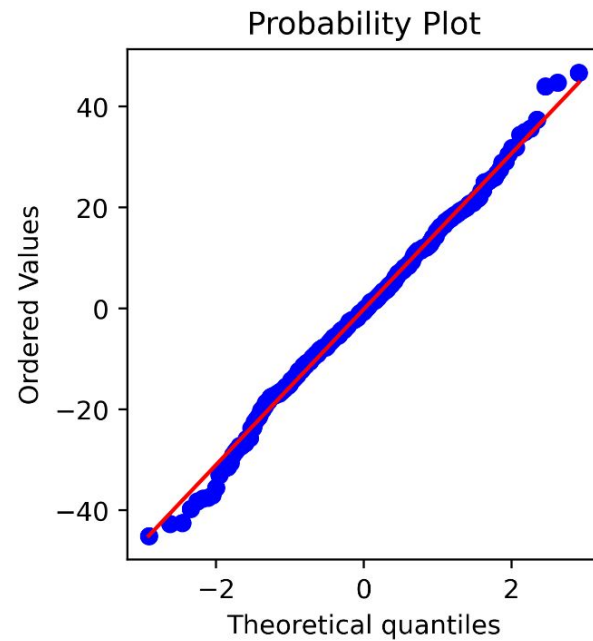
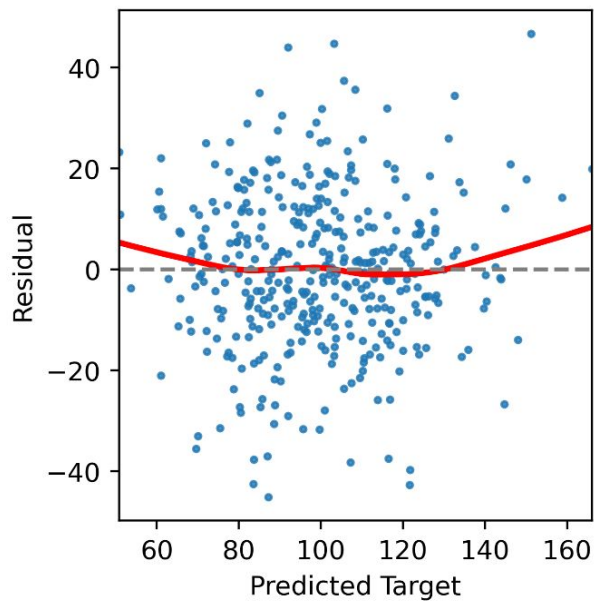
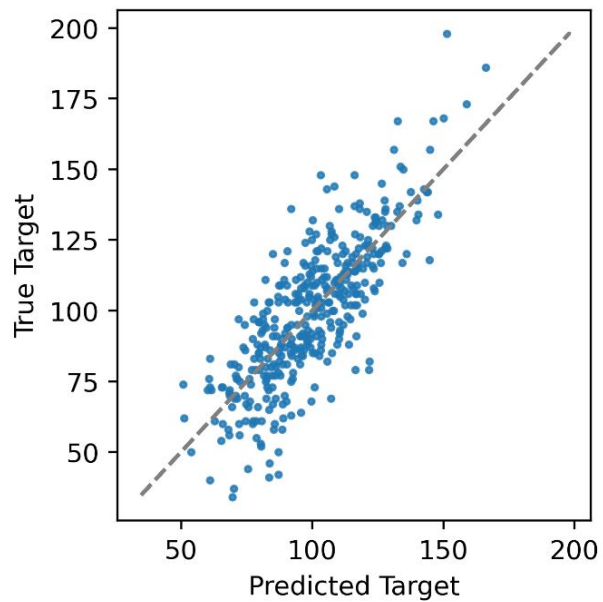


## ***RESOURCES***

- <https://www.mlb.com/>
- <https://www.baseball-reference.com/>
- <https://baseballsavant.mlb.com/>
- <https://en.wikipedia.org/>



## Diagnostic Plots - Test Data



## Diagnostic Plots - Train Data

