STATS TO EVALUATE *THE PERFORMANCE* OF BATTERS

BY CHIEN YUAN CHANG 10-1-2021

- Statcast
  - Introduced to all 30 MLB stadiums in 2015
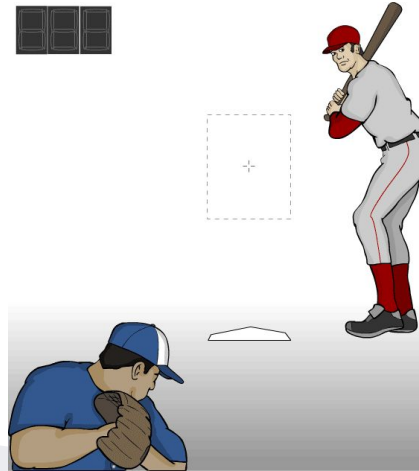
INTRODUCTION - OBJECTIVE/GOALS

- 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

- What is OPS (on-base plus slugging)?
  - OBP(on-base percentage) + SLG (slugging percentage)
- What is OPS+?
  - (OPS / league OPS, adjusted for park factors) x 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

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

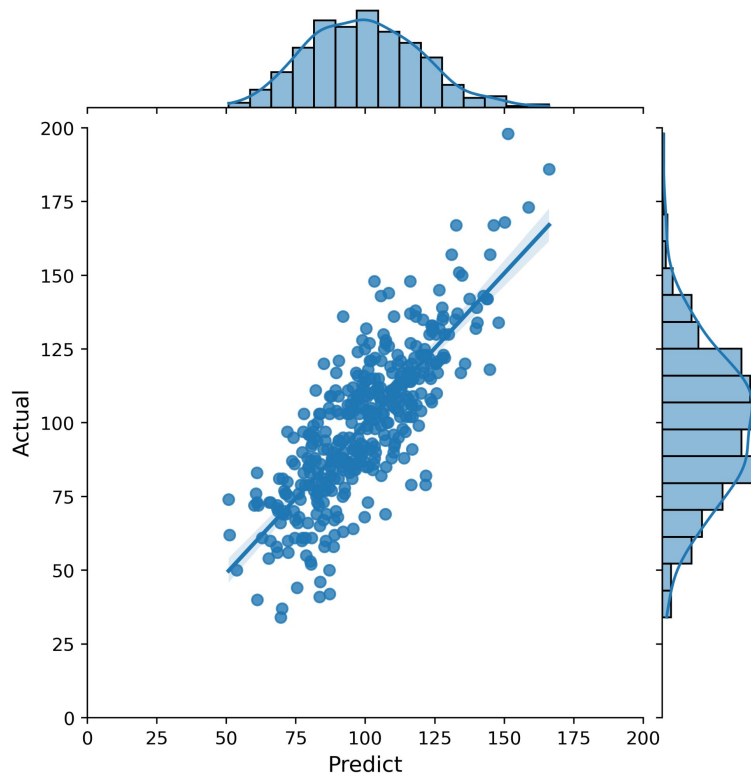Baseline Model
Adj. R^2 = 0.435

TRAIN

Training and Cross-Validation

9 FEATURES

Final Model
Adj. R^2 = 0.601

- 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% / 100	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**

Power
Higher Exit Velocity
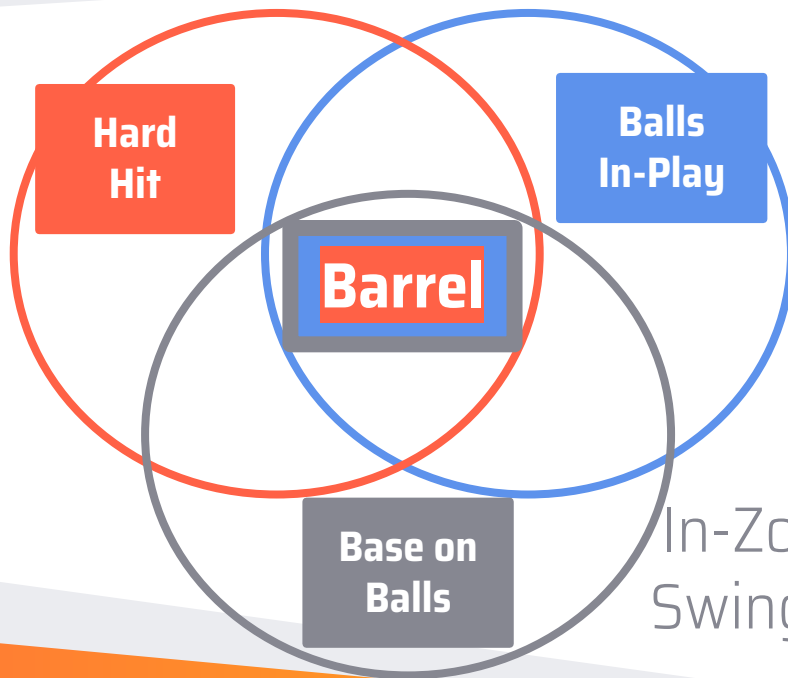
Contact
Swing and Contact
With Right Angle
(Sweet Spot)

Hard Hit

Balls In-Play

Barrel

Base on Balls

- Speed
  - Advantage
- Age
  - Small Disadvantage

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

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!

APPENDIX

RESOURCES

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

Diagnostic Plots - Test Data
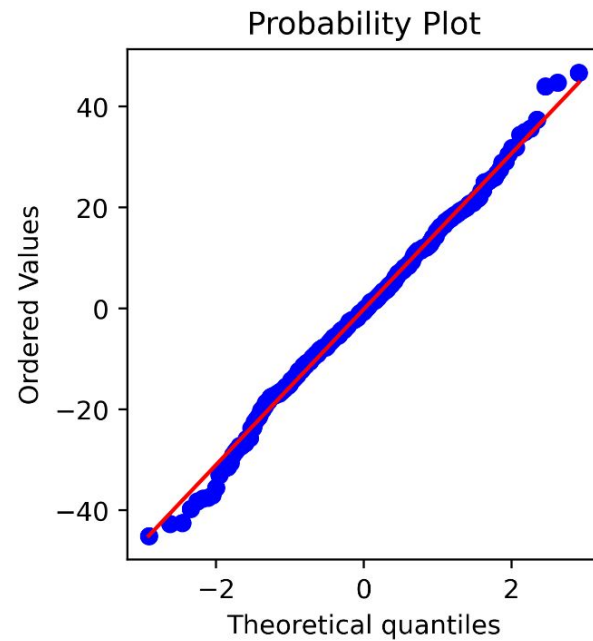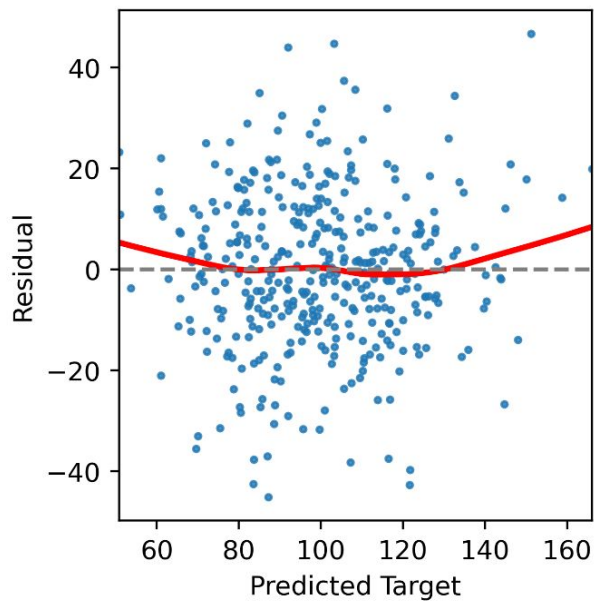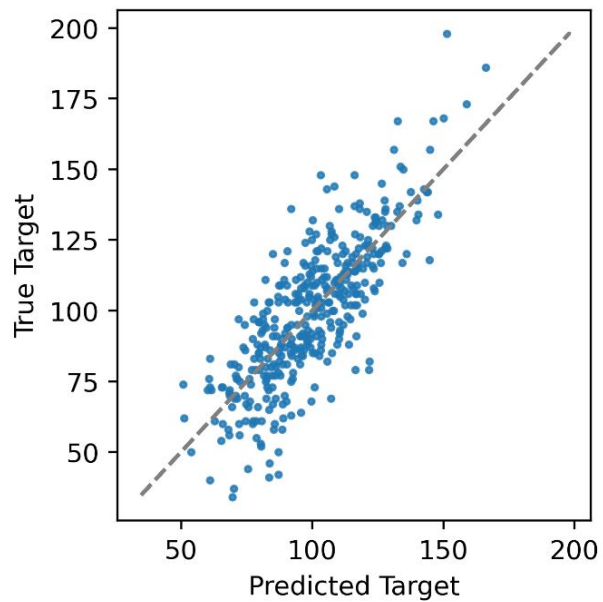
Diagnostic Plots - Train Data