statcast\_stats\_2015-2019.csv

The data are a selected subset from MLB’s STATCAST data set.

<https://baseballsavant.mlb.com/leaderboard/custom?year=2018,2017,2016,2015&type=batter&filter=&sort=1&sortDir=desc&min=q&selections=xba,xslg,woba,xwoba,xobp,xiso,exit_velocity_avg,launch_angle_avg,sweet_spot_percent,barrels,barrel_batted_rate,solidcontact_percent,flareburner_percent,poorlyunder_percent,poorlytopped_percent,poorlyweak_percent,hard_hit_percent,z_swing_percent,z_swing_miss_percent,oz_swing_percent,oz_swing_miss_percent,oz_contact_percent,out_zone_swing_miss,out_zone_swing,out_zone_percent,out_zone,meatball_swing_percent,meatball_percent,pitch_count_offspeed,pitch_count_fastball,pitch_count_breaking,pitch_count,iz_contact_percent,in_zone_swing_miss,in_zone_swing,in_zone_percent,in_zone,edge_percent,edge,whiff_percent,swing_percent,pull_percent,straightaway_percent,opposite_percent,batted_ball,f_strike_percent,groundballs_percent,groundballs,flyballs_percent,flyballs,linedrives_percent,linedrives,popups_percent,popups,n_bolts,hp_to_1b,sprint_speed,&chart=false&x=xba&y=xba&r=no&chartType=beeswarm>

BaseballSavant is a site dedicated to providing player matchups, Statcast metrics, and advanced statistics in a simple and easy-to-view way. Daren Willman created the website as an employee of MLB. Instances of the data are statistics of years of play by professional baseball players in MLB, of which there are 708 in this dataset. This dataset is a subset of a larger set. The larger set is all available batting statistics for every MLB player during the timespan 2015-2019. The sample is representative of the larger set, as we set a minimum threshold for Plate Appearances (MLB qualifies players for awards at 502 Plate Appearances), so we only exclude player years that had relatively small sample sizes. The data on each instance contains Statcast statistics, normal batting statistics, name, and year. Important statistics are detailed below in the table. There is no missing information or redundancies in the dataset. The dataset relies on game files from MLB Advanced Media, L.P. There are no guarantees, but it would be highly unlikely for MLB to cease producing public game files. There are no restrictions associated with any of the external resources. The link provided by baseballsavant to external resources is <http://mlb.mlb.com/>. The data relates to people but does not provide any information about those people other than statistics about their performance during the MLB season. Some of the data was directly observable, some of the data was collected using Statcast. Statcast is a state-of-the-art tracking technology, capable of measuring previously unquantifiable aspects of the game. Set up in all 30 Major League ballparks, Statcast collects data using a series of high-resolution optical cameras along with radar equipment. The technology precisely tracks the location and movements of the ball and every player on the field, resulting in an unparalleled amount of information covering everything from the pitcher to the batter to baserunners and defensive players.

Sabr.org is a resource with many baseball research papers, some of which use statcast data.

|  |  |  |
| --- | --- | --- |
| Variable Name | Description |  |
| xba | Expected Batting Average (xBA) | xBA measures the likelihood that a batted ball will become a hit |
| xslg | Expected Slugging Percentage (xSLG) | xSLG measure the expected number of bases from a batted ball |
| woba | Weighted On-base Average (wOBA) | wOBA is a version of on-base percentage that accounts for how a player reached base -- instead of simply considering whether a player reached base |
| xwoba | Expected Weighted On-base Average (xwOBA) | xwOBA is wOBA but each way to reach base is affected by a factor considering adjusted run expectancy of a batting event in the context of the season as a whole |
| xiso | Expected Isolated Power | xISO measures a hitter’s expected extra bases per at bat |
| exit\_velocity\_avg | Average Exit Velocity (Avg EV (MPH)) | How fast, on average, in miles per hour, a ball was hit by a batter |
| launch\_angle\_avg | Average Launch Angle (°) | How high/low, on average, in degrees, a ball was hit by a batter |
| sweet\_spot\_percent | Sweet Spot % | Percentage of batted balls with a launch angle between 8 and 32 degrees |
| barrels | Barrels | A batted ball with the perfect combination of exit velocity and launch angle |
| barrel\_batted\_rate | Barrel% | Barrels per plate appearence |
| solidcontact\_percent | Solid Contact% | Percent of contact classified as solid |
| hard\_hit\_percent | Hard Hit Rate | Percent of contact with an exit velocity of 95+ MPH |
| z\_swing\_percent | Z-Swing% | Swings at pitches inside the zone / pitches inside the zone |
| z\_swing\_miss\_percent | Z-SwingMiss% | Swings and misses at pitches inside the zone / pitches inside the zone |
| oz\_swing\_percent | O-Swing% | Swings at pitches outside the zone / pitches outside the zone |
| out\_zone\_swing\_miss | O-SwingMiss% | Swings and misses at pitches outside the zone / pitches outside the zone |
| out\_zone\_swing | Out of Zone Swings | Swings at pitches outside the zone |
| out\_zone\_percent | Out Zone% | Percent of pitches outside the zone |
| out\_zone | Out of Zone | Pitches outside the zone |
| meatball\_swing\_percent | Meatball Swing % | Swings at meatballs / meatballs  Meatball is defined as a fastball over the middle of the plate |
| meatball\_percent | Meatball % | Meatballs / all pitches |
| pitch\_count\_offspeed | # Offspeed | # of offspeed pitches |
| pitch\_count\_fastball | # Fastball | # of fastballs |
| pitch\_count\_breaking | # Breaking | # of breaking balls |
| pitch\_count | # Pitches | Total # of pitches |
| iz\_contact\_percent | Z-Contact% | Number of pitches on which contact was made on pitches inside the zone / Swings on pitches inside the zone |
| in\_zone\_swing\_miss | In Zone Swing & Miss | Swings and misses at pitches inside the zone |
| in\_zone\_swing | In Zone Swings | Swings at pitches inside the zone |
| in\_zone\_percent | In Zone % | Percent of pitches inside the zone |
| in\_zone | In Zone | Number of pitches inside the zone |
| edge\_percent | Edge % | Percent of pitches on the edge of the strike zone |
| edge | Edge | Number of pitches on the edge of the strike zone |
| whiff\_percent | Whiff % | Swings and misses / pitches |
| n\_bolts | # Bolts | Number of runs where the player’s sprint speed is at least 30 ft/sec |
| hp\_to\_1b | Home to First | Time elapsed from bat on ball contact to the moment when the batter reaches first base (average on competitive runs) |
| sprint\_speed | Sprint Speed (ft/sec) | Average speed during the fastest 1 second window (ft/sec) of the top 2/3 of competitive runs |

2015\_batter\_OPS+.csv, 2016\_batter\_OPS+.csv, 2017\_batter\_OPS+.csv, 2018\_batter\_OPS+.csv, 2019\_batter\_OPS+.csv

Each row represents a player during a specific year (depending on the dataset), and the columns include information such as player name, total plate appearances that season, OPS+, and more (described in the table below). SportsReference’s Primary Aim is to answer our users' questions with the easiest-to-use, fastest, and most complete sports statistics anywhere. If they have some fun in the process, that's good too. SportsReference is a for-profit company that created this dataset. SportsReference, being a third-party company to MLB, only has stats that are visually available as well as statistics calculated using those visually available statistics. The data was not preprocessed in any way. By participating in MLB, the athletes imply consent. Raw source data can be found at these links:

<https://www.baseball-reference.com/leagues/MLB/2019-standard-batting.shtml>, <https://www.baseball-reference.com/leagues/MLB/2018-standard-batting.shtml>,

<https://www.baseball-reference.com/leagues/MLB/2017-standard-batting.shtml>,

<https://www.baseball-reference.com/leagues/MLB/2016-standard-batting.shtml>,

<https://www.baseball-reference.com/leagues/MLB/2015-standard-batting.shtml>

|  |  |  |
| --- | --- | --- |
| Variable Name | Description |  |
| Name | Name | Player full name |
| PA | # Plate Appearances | A Plate Appearance refers to a player’s turn batting |
| OBP | On-base Percentage | How frequently a batter reaches base per plate appearance, including hits, walks, and hit by pitches |
| SLG | Slugging Percentage | The total number of bases per at bat (only including hits) |
| OPS | On-base Plus Slugging | On-base Percentage plus Slugging Percentage |
| OPS+ | On-base Plus Slugging Plus | League-adjusted OPS. 100 is league average, 120 is 20% better than league average, etc. (OPS / league OPS, adjusted for park factors) x 100 |