**Week 1 Report**

By: Jai Arora, Robert Banas, Bailey Harrison, and Simran Mallik

Data Cleaning

The data analyzed was collected from *FanGraphs*. There were a total of four datasets that were cleaned and analyzed (MiLB 2012 Batters, MiLB 2012 Pitchers, MiLB 2013 Batters, MiLB 2013 Pitchers) in R, JMP, and Excel. The metrics analyzed for the 2012 and 2013 batters included the On-base Plus Slugging (OPS), Weighted Runs Created + (wRC+), and Home Run to Fly Ball Rate (HR/FB) metrics. The metrics analyzed for the 2012 and 2013 pitchers included the Home Run Per Nine Innings (HR/9), Expected Fielding Independent Pitching (xFIP), and Walks and Hits Per Inning Pitched (WHIP) metrics.

Both the Advanced and Batted datasets were needed for the batters since each had the necessary batter metric(s), whereas only the Advanced dataset was used for the pitchers. *Fangraphs* included an automatic filter that only displayed the statistics for players who were ‘qualified’ based off of the minimum Plate Appearance and Innings Pitched (PA/IP). The dataset was filtered for players who were in the Upper Level (AAA/AA) league since players from this league are more likely to make it to the MLB. Additionally, the dataset was filtered for players who were between the ages of 18-25. Players who were 16-17 years old would be more expensive to maintain and require more equipment since they would remain in the MiLB for a longer period of time than older players. The younger player prospects also tend to be international players, which means that the expense is even larger. Players aged 25 in the MiLB were no longer deemed as a viable prospect and were unlikely to reach Major League Baseball (MLB) player status, and therefore were not included in our cleaned dataset.  Once all the filtering had been completed, each dataset consisted of a random sample of 75 players.

Using Microsoft Excel, each players’ score for each metric was multiplied by a weight that our team designated for each specific metric. For each batter, the scores for the OPS and wRC+ metrics were each multiplied by .40, and the score for the HR/FB metric was multiplied by .20. This method created a weighted OPS, wRC+, and HR/FB score for each player.

For each pitcher, the scores for the xFIP and WHIP metrics were multiplied by 0.40, and the score for the HR/9 metric was multiplied by .20. This method created a weighted xFIP, WHIP, and HR/9 score for each player. These weighted scores for each player were summed to get a final score. Then, the sort and filter function on Excel was applied to ensure that the list of players were being updated per individual metric. The ranking of the scores depended upon what type of metrics were chosen for batters and pitchers. The higher the score for each metric for the batters, the better their rank would be. The lower the score for each metric for the pitchers, the better their rank would be. This resulted in the batters being ranked from highest score to lowest score, while the pitchers were ranked from lowest score to highest score. The players towards the top of the batter and pitcher list were more likely to move to the MLB.

Variables Selected

**Hitters:** As a team, we thought that On-base Plus Slugging (OPS), Weighted Runs Created + (wRC+), and Home Run to Fly Ball Rate (HR/FB) were the best metrics used to evaluate a hitters potential to make it to the MLB. We wanted to use OPS because it combines a hitters on-base percentage and slugging percentage. Meaning how well a hitter can reach base, with how well he can hit for average and for power. Next, we used wRC+ as our second metric to evaluate hitters. wRC+ takes the statistic Runs Created and adjusts that number to account for important external factors, like ballpark or era. It is useful because wRC+ quantifies run creation and normalizes it. Then, our last metric we used to evaluate hitters was HR/FB. Home Run to Fly Ball rate is the ratio of home runs a player hits out of their total number of fly balls. A hitters HR/FB ratio can be useful in providing context about how sustainable their power is. In our formula we weighted OPS and wRC+ more than HR/FB because we felt like those metrics were more important. To be a good baseball player you do not need to hit a lot of homeruns. For example, you can be a high batting average hitter with speed and have a low HR/FB and still be very good.

**Pitchers:** As a team we thought that Expected Fielding Independent Pitching (xFIP), Walks and Hits Per Inning Pitched (WHIP), and Home Run Per Nine Innings (HR/9) were the best metrics to evaluate a pitchers potential to make it to the MLB. We used xFIP because it is a better version of ERA and FIP. It shows whether the pitcher performs well at limiting runs scored, independent of his team’s poor or positive fielding. We chose xFIP over regular FIP as it, “replaces a pitcher’s home run total with an estimate of how many home runs they should have allowed given the number of fly balls they surrendered while assuming a league average home run to fly ball percentage.” Next, we used WHIP as our second metric. WHIP shows how well a pitcher has kept runners off the basepaths. The formula is the sum of a pitcher's walks and hits, divided by his total innings pitched. Then, our last metric we used was HR/9. Home Run Per Nine Innings represents the average number of home runs allowed by a pitcher on a nine-inning scale. HR/9 is a statistic that is in the control of the pitcher, because defensive positioning plays no factor. In our formula we weighted xFIP and WHIP more than HR/9 because we felt like those metrics were more important. Certain pitchers can have success with a high HR/9 rate, as long as they manage to limit their baserunners otherwise, leading to fewer multi-run home runs. However, some pitchers have mastered the art of allowing fly balls but not many home runs.

Final Formulas

The higher the score, the better for hitters.

The lower the score, the better for pitchers.

After plugging in the data to our formula you can see from our rankings below that our formula does a good job identifying future MLB players and MLB All-Star players. At the top of the lists, you can see notable names like Jedd Gyorko, Dan Straily, Zach Wheeler, Chris Archer, Collin McHugh, Joc Pederson, Mark Canha, Kyle Hendricks, and Taijuan Walker to just name a few.

2012 Hitters Rankings

Rank   Score    Name

1 62.240 Hunter Morris

2 59.879 Marc Krauss

3 58.693 Tim Fedroff

4 58.547 Ben Guez

5 58.054 Thomas Neal

6 57.702 Ryan Wheeler

7 55.954 Jedd Gyorko

8 54.460 Corban Joseph

9 53.832 Eric Campbell

10 52.998 Chun-Hsiu Chen

11 52.873 Starling Marte

12 51.209 Austin Wates

13 50.510 Kole Calhoun

14 49.724 Bryce Brentz

15 49.536 Robbie Grossman

16 49.487 John Whittleman

17 49.454 Matt Clark

18 48.899 Kyle Jensen

19 47.458 Mike McDade

20 46.927 Josh Bell

21 46.899 Cole Figueroa

22 46.787 Juan Carlo Perez

23 46.619 Addison Maruszak

24 46.614 L.J. Hoes

25 46.217 Omar Luna

26 45.579 Ezequiel Carrera

27 45.428 Andy Wilkins

28 44.892 Kolten Wong

29 44.514 Nolan Arenado

30 44.512 Matt Long

31 44.419 Lars Anderson

32 44.095 Luis Jimenez

33 43.207 Brett Jackson

34 42.892 Anthony Gose

35 42.462 Rafael Ynoa

36 42.289 Ryan Goins

37 42.086 Luis Nunez

38 41.953 Ryan LaMarre

39 41.658 Juan Diaz

40 40.846 Josh Prince

41 40.816 Cesar Hernandez

42 40.770 Conor Gillaspie

43 40.626 Joey Terdoslavich

44 40.388 Engel Beltre

45 39.891 Ramon Cabrera

46 39.873 Scooter Gennett

47 39.833 Chase d'Arnaud

48 39.598 Brandon Laird

49 39.351 Brad Glenn

50 39.106 Ben Paulsen

51 39.069 Jonathan Schoop

52 39.028 Hak-Ju Lee

53 38.252 Tyler Saladino

54 37.277 Cedric Hunter

55 36.695 Jefry Marte

56 36.413 Tony Plagman

57 35.823 Evan Bigley

58 35.783 Ryan Strausborger

59 35.628 Caleb Gindl

60 35.226 Josh Phegley

61 34.374 Sebastian Valle

62 33.438 Zack Cox

63 32.837 James Beresford

64 32.791 Che-Hsuan Lin

65 32.741 Rey Navarro

66 32.581 Dusty Coleman

67 32.456 Phil Gosselin

68 31.733 Kyle Hudson

69 31.135 Dave Sappelt

70 31.118 David Nick

71 29.836 Jeudy Valdez

72 29.670 Francisco Peguero

73 29.080 Keenan Wiley

74 28.434 Darwin Perez

75 25.167 Bryan Anderson

2012 Pitchers Rankings

Rank    Score   Name

1 1.679 Chris Heston

2 1.737 Dan Straily

3 1.812 Matt Magill

4 1.857 Adam Wilk

5 1.874 Andrew Werner

6 1.892 Phil Irwin

7 1.939 Zack Wheeler

8 1.953 Chris Archer

9 1.958 Collin McHugh

10 1.981 Brandon Maurer

11 2.010 Justin Grimm

12 2.020 Jeff Locke

13 2.025 Trevor Bauer

14 2.041 Vidal Nuno

15 2.059 Hiram Burgos

16 2.069 Tyler Skaggs

17 2.076 Allen Webster

18 2.092 Nick Struck

19 2.093 Edwar Cabrera

20 2.104 Mike Kickham

21 2.104 Tyler Cloyd

22 2.110 Paolo Espino

23 2.126 Ethan Martin

24 2.142 Daniel Corcino

25 2.167 TJ House

26 2.177 Giovanni Soto

27 2.177 Charlie Leesman

28 2.205 Jake Pettit

29 2.206 Tanner Roark

30 2.210 Barret Loux

31 2.218 Gonzalez Germen

32 2.230 Stolmy Pimentel

33 2.232 Taijuan Walker

34 2.236 Justin Wilson

35 2.241 Danny Hultzen

36 2.253 Tyler Lyons

37 2.256 David Hale

38 2.257 Jarred Cosart

39 2.260 Brett Marshall

40 2.282 Chris Schwinden

41 2.282 Jeurys Familia

42 2.287 Kyle Lobstein

43 2.298 Casey Crosby

44 2.334 Keyvius Sampson

45 2.337 Brooks Raley

46 2.340 Jose Cisnero

47 2.351 A.J. Schugel

48 2.373 D.J. Mitchell

49 2.397 Brett Oberholtzer

50 2.408 Darin Gorski

51 2.418 Jake Brigham

52 2.424 Chad Bell

53 2.437 Trevor May

54 2.438 Scott Gorgen

55 2.445 Ariel Pena

56 2.477 Jake Odorizzi

57 2.486 Wily Peralta

58 2.491 Johnny Hellweg

59 2.499 Omar Poveda

60 2.505 John Gast

61 2.513 Aaron Miller

62 2.513 Rob Scahill

63 2.514 Kelvin De La Cruz

64 2.530 Brad Hand

65 2.584 Juan Carlos Sulbaran

66 2.618 Shelby Miller

67 2.624 Julio Rodriguez

68 2.639 Paul Demny

69 2.655 Charles Brewer

70 2.656 Andy Oliver

71 2.767 Brad Peacock

72 2.787 Neil Ramirez

73 2.830 Forrest Snow

74 2.838 Ryan Verdugo

75 3.008 Dellin Betances

2013 Hitters Rankings

|  |  |  |
| --- | --- | --- |
| Rank | Score | Name |
| 1 | 62.357 | Joc Pederson |
| 2 | 56.850 | Mark Canha |
| 3 | 55.248 | Domingo Santana |
| 4 | 53.580 | Cody Asche |
| 5 | 53.191 | Arismendy Alcantara |
| 6 | 52.779 | Brett Nicholas |
| 7 | 52.071 | Wilmer Flores |
| 8 | 50.872 | Christian Villanueva |
| 9 | 50.858 | Alex Dickerson |
| 10 | 50.783 | Carlos Moncrief |
| 11 | 50.713 | Rubi Silva |
| 12 | 48.809 | Jaff Decker |
| 13 | 48.762 | Jarrett Parker |
| 14 | 47.586 | Yamaico Navarro |
| 15 | 46.752 | Kyle Roller |
| 16 | 46.324 | Brian Goodwin |
| 17 | 46.106 | Ryan Schimpf |
| 18 | 45.715 | Chris McGuiness |
| 19 | 45.400 | Eury Perez |
| 20 | 45.048 | Matt Szczur |
| 21 | 44.769 | Dee Strange-Gordon |
| 22 | 44.334 | Tyler Collins |
| 23 | 43.674 | Greg Garcia |
| 24 | 43.433 | Rafael Ynoa |
| 25 | 43.125 | C.J. Cron |
| 26 | 43.091 | Dan Brewer |
| 27 | 42.249 | Rey Navarro |
| 28 | 41.832 | Slade Heathcott |
| 29 | 41.830 | Ramon Flores |
| 30 | 41.374 | Mel Rojas Jr. |
| 31 | 41.162 | Jonathan Meyer |
| 32 | 40.970 | Neftali Soto |
| 33 | 40.924 | Mike Walker |
| 34 | 40.818 | Shane Peterson |
| 35 | 40.614 | Ender Inciarte |
| 36 | 40.514 | Justin Bloxom |
| 37 | 40.489 | Brandon Laird |
| 38 | 39.971 | D'Arby Myers |
| 39 | 39.661 | Anthony Hewitt |
| 40 | 39.395 | Rico Noel |
| 41 | 39.230 | Chih-Hsien Chiang |
| 42 | 38.870 | Tyler Holt |
| 43 | 38.726 | Edward Salcedo |
| 44 | 38.038 | Engel Beltre |
| 45 | 37.817 | Todd Cunningham |
| 46 | 36.902 | Carlos Triunfel |
| 47 | 36.788 | Todd Glaesmann |
| 48 | 36.768 | Jermaine Curtis |
| 49 | 36.706 | Nick Shaw |
| 50 | 36.364 | Christian Colon |
| 51 | 36.250 | Logan Watkins |
| 52 | 36.214 | Ryan Goins |
| 53 | 35.928 | Yangervis Solarte |
| 54 | 35.499 | Juan Diaz |
| 55 | 35.422 | Jose  Ramirez |
| 56 | 35.056 | Jayson Langfels |
| 57 | 34.308 | Zach Collier |
| 58 | 34.100 | Anthony Gose |
| 59 | 34.006 | Shannon Wilkerson |
| 60 | 33.688 | Tim Wheeler |
| 61 | 33.196 | Billy Hamilton |
| 62 | 32.839 | Ronny Rodriguez |
| 63 | 32.560 | Travis Witherspoon |
| 64 | 32.458 | Robby Price |
| 65 | 32.277 | Miguel Rojas |
| 66 | 31.172 | Nick Ahmed |
| 67 | 31.011 | Leandro Castro |
| 68 | 30.893 | Ryan Strausborger |
| 69 | 28.875 | Gary Brown |
| 70 | 26.728 | Destin Hood |
| 71 | 26.079 | Kaleb Cowart |
| 72 | 25.920 | Yolmer Sanchez |
| 73 | 25.874 | Jio Mier |
| 74 | 25.308 | Leon Landry |
| 75 | 16.721 | Hector Gomez |

2013 Pitchers Rankings

|  |  |  |
| --- | --- | --- |
| Rank | Score | Name |
| 1 | 1.715 | Kyle Hendricks |
| 2 | 1.725 | Erik Johnson |
| 3 | 1.741 | Zach Lee |
| 4 | 1.811 | Jose Alvarez |
| 5 | 1.832 | Gus Schlosser |
| 6 | 1.906 | Aaron Northcraft |
| 7 | 1.910 | Rafael Montero |
| 8 | 1.922 | Taijuan Walker |
| 9 | 1.926 | Jake Buchanan |
| 10 | 1.936 | Matt Andriese |
| 11 | 1.943 | Andrew Chafin |
| 12 | 1.949 | Brian Flynn |
| 13 | 1.949 | Nate Karns |
| 14 | 1.966 | Jake Odorizzi |
| 15 | 2.001 | Eric Jokisch |
| 16 | 2.006 | Kyle Lobstein |
| 17 | 2.038 | Cody Martin |
| 18 | 2.055 | Yordano Ventura |
| 19 | 2.056 | Tim Crabbe |
| 20 | 2.058 | David Martinez |
| 21 | 2.074 | Josh A. Smith |
| 22 | 2.088 | Donn Roach |
| 23 | 2.098 | Spencer Arroyo |
| 24 | 2.100 | Matt Packer |
| 25 | 2.118 | Jameson Taillon |
| 26 | 2.136 | TJ House |
| 27 | 2.150 | Omar Poveda |
| 28 | 2.154 | Caleb Clay |
| 29 | 2.167 | Carlos Pimentel |
| 30 | 2.184 | Chad Rogers |
| 31 | 2.189 | Sean Murphy |
| 32 | 2.204 | Lay Batista |
| 33 | 2.220 | Mike Bolsinger |
| 34 | 2.233 | Jesse Biddle |
| 35 | 2.252 | David Bromberg |
| 36 | 2.257 | Jake Thompson |
| 37 | 2.259 | Pat Dean |
| 38 | 2.262 | Enny Romero |
| 39 | 2.274 | Merrill Kelly |
| 40 | 2.276 | Keith Couch |
| 41 | 2.282 | Deck McGuire |
| 42 | 2.293 | Keyvius Sampson |
| 443 | 2.294 | Seth Rosin |
| 44 | 2.296 | Andy Moye |
| 45 | 2.306 | Zach Neal |
| 46 | 2.311 | Marcus Walden |
| 47 | 2.312 | Murphy Smith |
| 48 | 2.322 | James Paxton |
| 49 | 2.322 | Will Roberts |
| 50 | 2.341 | Ariel Pena |
| 51 | 2.351 | Matt Crouse |
| 52 | 2.381 | Victor Mateo |
| 53 | 2.381 | Danny Rosenbaum |
| 54 | 2.386 | Boone Whiting |
| 55 | 2.390 | Jason Adam |
| 56 | 2.396 | Nik Turley |
| 57 | 2.401 | Brett Brach |
| 58 | 2.407 | Casey Sadler |
| 59 | 2.411 | Josh Stinson |
| 60 | 2.429 | Devin Jones |
| 61 | 2.434 | David Buchanan |
| 62 | 2.441 | Charles Brewer |
| 63 | 2.443 | Seth Blair |
| 64 | 2.449 | Caleb Cotham |
| 65 | 2.464 | Sugar Ray Marimon |
| 66 | 2.465 | Stephen McCray |
| 67 | 2.510 | Richard Castillo |
| 68 | 2.514 | Brett Marshall |
| 69 | 2.515 | Ethan Martin |
| 70 | 2.524 | Chris Hernandez |
| 71 | 2.536 | Nick McCully |
| 72 | 2.587 | Zach Nuding |
| 73 | 2.802 | Chris Dwyer |
| 74 | 2.847 | Ross Seaton |
| 75 | 2.934 | Daniel Corcino |
|  |  |  |