MiLB Pitcher Evaluations Documentation

Reasoning

The purpose of the minor league (MiLB) Pitcher Evaluations report is to compare current prospects to their most similar current big-league pitchers at the same point in their career. This can give an idea of what one might expect from a pitcher at the MLB level, as well as what those similar pitchers may have changed to improve their success. It has four components: 1) A table with qualitative rankings for a handful of the focus pitcher's KPIs at their current level 2) A table with the most similar big league pitchers to the focus pitcher, based on a variety of factors (explained further below) 3) A pie chart with a role breakdown for those most similar pitchers at the big league level and 4) A percent likelihood of making MLB based on the most similar pitchers.

Determining Similar Pitchers

To find the most similar pitchers, a variety of qualitative and performance characteristics were found for every pitcher by season from A-MLB dating back to 2018 (2020 not included). These characteristics were:

- Competition level
- Throwing hand
- Age
- Binary variables indicating whether each pitcher threw the following pitch types
    - Fastball
    - Sinker
    - Slider
    - Curveball
    - Changeup
    - Splitter
    - Cutter
    - Knuckleball
    - Screwball
- A 7-level factor for each of the following characteristics (to classify continuous variables into 7 groups)
    - Release height
    - Release side
    - Extension
    - Games started %
- A 7-level factor for each of the following KPIs (levels being a ranking system from 1-7: Awful, Poor, Below Average, Average, Above Average, Great, Elite)
    - Chase %
    - K %

- BB %
- First Pitch Strike %
- GB %
- xwOBA
- FB Stuff
- BB Stuff
- OS Stuff
- FB LVA
- BB LVA
- OS LVA

For pitch types to qualify in a pitcher's repertoire, he must have thrown the pitch at least 5% of the time at the focus level and year. For a pitcher to qualify at a certain level and year, he must have thrown at least 400 pitches at that level and year. Pitcher extension is only based on pitches tagged as fastballs. For a pitcher to be considered as an "MLB pitcher" he must have both made it to the big leagues and faced at least 100 batters at the big-league level.

To begin evaluating the most similar pitchers, a focus pitcher is chosen, along with competition level and year. To easily explain, let's say for example I want to look at Winston Santos at AA in 2024. This will automatically filter the data to only include right-handed pitchers at AA between 20-22 years old from 2018-2019 and 2021-2023 (the most recent year before the focus year). The similarity score is then found using a combination of Manhattan distances and Euclidean distances.

First, Euclidean distances are found between the focus pitcher's "release point" (release height/side) and all other pitchers in the filtered data. This distance is added to the table and the separate release height/side metrics are dropped. Then, Manhattan distances are found across all characteristics between the focus pitcher and each pitcher in the filtered data. The smaller this distance, the more "similar" a pitcher is to the focus pitcher. These "similarity scores" are rescaled on a 0-100 scale where 100 is perfectly similar to the focus pitcher.

MiLB Pitcher Evaluations Documentation

Report Components

For this section, we'll continue to use Winston Santos' 2024 season at AA as our focus.

1) *KPI Rankings –*

Winston Santos's 2024 Performance Rankings

TEX / RHP / AA / 21 Y.O. / FB, SL, CH / 5.5 Ht / -2 Sd / 6.5 Ext

KPI	Rank	Pitch Grades	Rank
Chase %	Above Average	FB Stuff	Great
K %	Great	BB Stuff	Elite
BB %	Above Average	OS Stuff	Above Average
FPS %	Average	FB LVA	Average
GB %	Poor	BB LVA	Average
xwOBA	Below Average	OS LVA	Average

The first component of the report is the KPI rankings table. These rankings are determined by converting the raw values to z-scores and grouping into categories based on the scores. This is done at each level and year. This gives a more intuitive idea of how well the pitcher performed in each KPI at that point in their career.

2) *Closest MLB Pitchers –*

Closest MLB Pitchers

min. 100 BF at MLB

	At Same Point In Career												MLB Career Statistics									
Pitcher	Year	Level	Age	Throws	Rel Height	Rel Side	Extension	Repertoire	Similar KPIs	Better KPIs	Worse KPIs	Similarity %	G	GS	IP	FIP	K%	BB%	GB%	WAR	Pitches Added	Pitches Dropped
Connor Seabold	2018	AA	22	R	5.6	-3.1	6	FB, SL, CH	Chase, FPS, GB, OS STUFF, LVA (FB,OS)	xwOBA, BB LVA	K, BB, STUFF (FB,BB)	90	33	19	108.2	5.88	16.6	7.3	33.3	0		
Justin Dunn	2018	AA	22	R	5.8	-2.1	6.4	FB, SL, CU, CH	Chase, FPS, LVA (FB,BB,OS)	GB, xwOBA	K, BB, STUFF (FB,BB,OS)	89	32	32	133.2	6.25	19.3	14.7	33	-0.6		CU, CH
Jared Jones	2023	AA	21	R	5.6	-2.2	6.5	FB, SL, CH	BB, STUFF (FB,OS), LVA (FB,OS)	FPS, GB, xwOBA, BB LVA	Chase, K, BB STUFF	89	22	22	121.2	4.05	26.2	7.7	39	1.6	CU	
Jack Leiter	2023	AA	22	R	5.6	-1.2	6.8	FB, SL, CU	K, FPS, GB, xwOBA, LVA (FB,OS)		Chase, BB, STUFF (FB,BB,OS), BB LVA	88	9	6	35.2	5.59	17.9	9.8	38.8	-0.3	CH	
Jonathan Hernandez	2019	AA	22	R	5.8	-1.6	6.5	FB, SL, CH	FPS, xwOBA, LVA (FB,BB,OS)	GB	Chase, K, BB, STUFF (FB,BB,OS)	87	127	3	153	4.32	22.3	12	52.3	0.1	SI	FB
Nate Pearson	2019	AA	22	R	6	-1.9	6.7	FB, SL, CU, CH	FPS, STUFF (FB,OS), BB LVA	GB, xwOBA	Chase, K, BB, BB STUFF, LVA (FB,OS)	87	112	6	142	4.83	24.7	10.3	35.6	0		CH
Cody Bolton	2019	AA	20	R	5.3	-3.1	6.6	FB, SI, SL, CH	GB, xwOBA, FB STUFF, LVA (BB,OS)		Chase, K, BB, FPS, STUFF (BB,OS), FB LVA	87	33	0	40	5.05	20.5	12.6	40.5	-0.3	CT	FB
Spencer Strider	2021	AA	22	R	5.9	-1.8	6.3	FB, SL	Chase, K, GB, LVA (FB,BB)	xwOBA, STUFF (FB,OS)	BB, FPS, BB STUFF, OS LVA	86	67	54	329.2	2.54	36.9	8.1	37.1	11.5	CU	
Grayson Rodriguez	2021	AA	21	R	5.9	-2.3	6.4	FB, SL, CU, CH	Chase, BB, FB STUFF, LVA (FB,BB,OS)	K, FPS, GB, xwOBA, OS STUFF	BB STUFF	86	43	43	238.2	3.76	25.7	7.8	43.9	4.3		
Ty Madden	2022	AA	22	R	5.9	-1.7	6	FB, SL, CH	K, BB, OS STUFF, LVA (FB,BB,OS)	Chase, FPS, GB, xwOBA	STUFF (FB,BB)	86	6	1	23	3.92	16.8	7.9	44	0.2	SP, CT	CH
TOTALS	-	-	-	-	-	-	-	-	-	-	-	-	648	324	2068.1	4.27	24.7	9.9	43.5	-	-	-

1–10 of 13 rows Previous 1 2 Next

MiLB Pitcher Evaluations Documentation

The key aspect of the report is the closest MLB pitchers table. This is where the similarity scores come into play – as they can be seen highlighted on a green to red color scale in the middle of the table. The table has two spanners: “At Same Point in Career” and “MLB Career Statistics”.

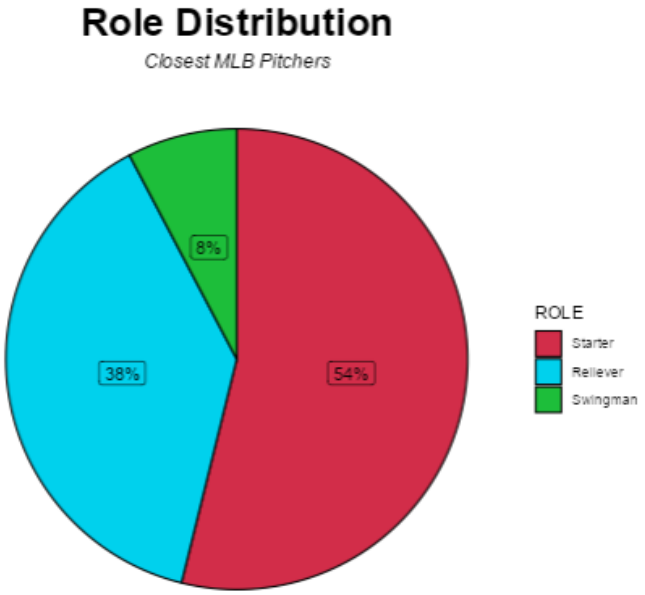
The first section contains characteristics of the most similar pitchers at the same point in their career, including release and extension, repertoire, and whether their KPIs were similar (same factor level), better (better factor level), or worse (worse factor level). This is good to get an idea of what the most successful pitchers on the list do well (and vice versa) and helps the focus pitcher set goals for himself.

The second section is the career MLB statistics of the most similar pitchers, which can be used in conjunction with the first section to determine if there is a pattern in certain characteristics that lead to success at the big-league level. To further support that claim is the pitches added/dropped column, indicating if the similar pitchers have changed their repertoire since being at the same level as the focus pitcher – and whether this has helped or hurt them.

There is a totals row at the bottom containing grand summary statistics across all pitchers included in the table.

You may notice the table has multiple pages with only 10 of the 13 pitchers being displayed. Some pitchers have several similar contemporaries while others are quite unique in their characteristics. This can lead the similarity scores the fluctuate from pitcher to pitcher. To maintain brevity and simplicity, a threshold has been set where if the 10th most similar pitcher has a similarity score less than 85, only the 10 closest players are shown. If the 10th most similar pitcher has a similarity score GREATER than 85, the table will continue displaying similar pitchers until either the score of a pitcher dips below 85 or a maximum of 20 pitchers are displayed.

3) *Role Breakdown –*



MiLB Pitcher Evaluations Documentation

The role distribution pie chart breaks down the pitchers in the closest MLB pitchers table by their primary role at the MLB level. The roles are starter (GS >= 60%), swingman (30% <= GS < 60%), and reliever (GS < 30%). This helps give an impression of what kind of role to expect the focus pitcher to fit into once at the big-league level.

4) *Likelihood of Making MLB –*

Likelihood of Making MLB %

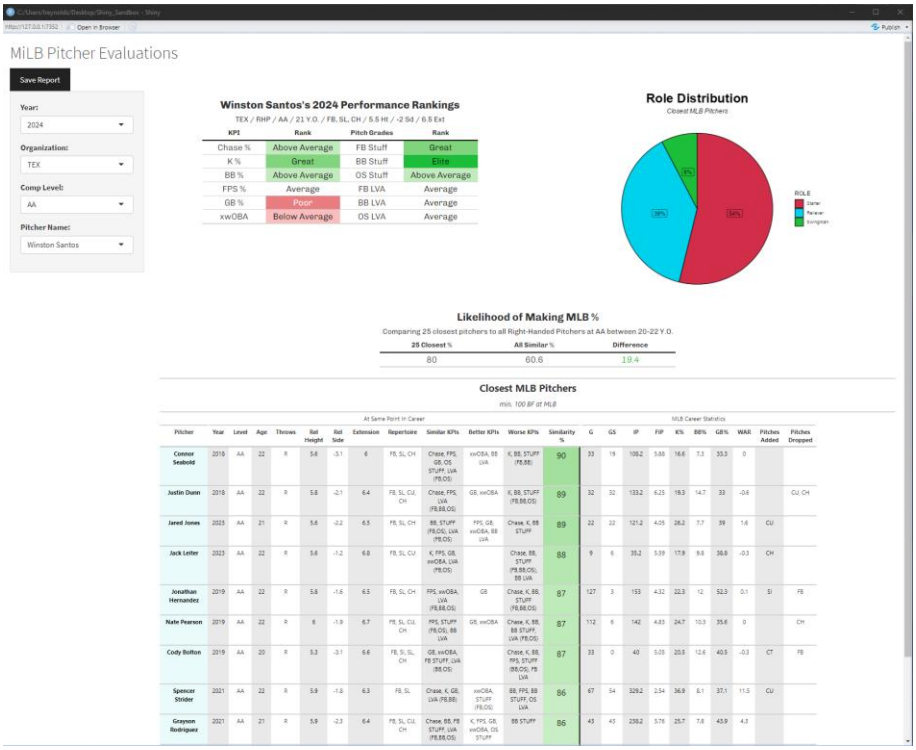
Comparing 25 closest pitchers to all Right-Handed Pitchers at AA between 20-22 Y.O.

25 Closest %	All Similar %	Difference
80	60.6	19.4

The final component of the report is the likelihood of making MLB percentage. It compares the percentage of the 25 most similar pitchers that made it to the big leagues to the percentage of all pitchers in the filtered data that made it to the big leagues. The difference in the two is also shown. In theory, the more positive the difference is, the more confident we are in a pitcher with a profile like that of the focus pitcher making it to MLB.

Shiny App

The report is being implemented into the Shiny server in Lonestar for users to have the ability to generate a report on any pitcher they want. The user will have the option to select year, organization, competition level and player, along with a button they can click to save the report as an image. Below is an example of what the interface looks like.



MiLB Pitcher Evaluations Documentation

Message to Code Manager

The files for the Shiny App are located in Databricks in `shiny-server/pd`. The file to update the pitcher characteristics table can be found in `databricks_analytics/Lance/Pitcher_Evaluations`. You will need to set up a job to run daily in the morning to maintain the most up-to-date information on each pitcher.