Fantasy Baseball Pitcher Analysis

Nikhil Sivakumar

2024-07-25

When playing fantasy baseball, people talk about how the ranking next to a pitcher is not always accurate and how it is not a good representation for how good a pitcher is. So, I’m trying to find an ideal way to rank pitchers in fantasy baseball, using the current 2024 season. Note that this analysis is for head-to-head leagues and not rotisserie leagues.

library(tidyverse)  
library(rvest)

First, I web scrapped the rankings from CBS Sports.

Here is what the current data frame looks like:

head(df1)

## V1  
## 1 T. Skubal\n SP\n \n DET\n Tarik Skubal\n SP\n \n DET  
## 2 S. Lugo\n SP\n \n KC\n Seth Lugo\n SP\n \n KC  
## 3 C. Sale\n SP\n \n ATL\n Chris Sale\n SP\n \n ATL  
## 4 C. Burnes\n SP\n \n BAL\n Corbin Burnes\n SP\n \n BAL  
## 5 L. Gilbert\n SP\n \n SEA\n Logan Gilbert\n SP\n \n SEA  
## 6 Z. Wheeler\n SP\n \n PHI\n Zack Wheeler\n SP\n \n PHI  
## V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19  
## 1 424.0 20.2 12 3 2.35 21 21 16 — 130.0 96 11 24 154 .199 0.92 — —  
## 2 399.0 18.1 12 5 2.66 22 22 16 1 142.1 116 12 32 124 .221 1.04 — —  
## 3 386.5 20.3 13 3 2.68 19 19 11 — 117.1 83 8 25 149 .196 0.92 — —  
## 4 384.5 18.3 10 4 2.45 21 21 17 — 132.0 107 15 30 121 .217 1.04 — —  
## 5 381.0 18.1 6 6 2.72 21 21 17 — 139.0 94 15 26 132 .190 0.86 — —  
## 6 369.5 18.5 10 4 2.55 20 20 15 — 123.2 84 11 37 133 .189 0.98 — —  
## V20 V21 V22 V23 V24  
## 1 — — 1.09 10.66 1.66  
## 2 — — 1.07 7.84 2.02  
## 3 — — 1.38 11.43 1.92  
## 4 — — 1.14 8.25 2.05  
## 5 — — 0.93 8.55 1.68  
## 6 — — 0.92 9.68 2.69

There are a lot of columns that need to be renamed and variables that need to be changed from qualitative to quantitative. Also, I am going to adjust the points so that it is scored from an ESPN fantasy league

extract\_full\_name <- function(text) {  
 lines <- unlist(str\_split(text, "\n"))  
 # Find the line that contains both a first name and a last name  
 full\_name <- lines[grep(" ", lines)]  
 return(str\_trim(full\_name[5]))  
}  
df1$full\_name <- sapply(df1$V1, extract\_full\_name)  
  
yearlystats <- df1 %>%  
 mutate(`V1` = `full\_name`) %>%  
 rename("Player" = V1, "FPTS" = V2, "FPPG" = V3, "W" = V4, "L" = V5, "ERA" = V6,  
 "GP" = V7, "GS" = V8, "QS" = V9, "CG" = V10, "IP" = V11, "H" = V12,   
 "HR" = V13, "BB" = V14, "SO" = V15, "AVG" = V16, "WHIP" = V17, "SV" = V18,  
 "BS" = V19, "SHO" = V20, "HLD" = V21, "GB%" = V22, "SO9" = V23, "BB9" = V24) %>%  
 mutate("W" = as.double(W), "L" = as.double(L), "QS" = as.double(QS),   
 "SO9" = as.double(`SO9`), "BB9" = as.double(`BB9`), "GS" = as.integer(GS)) %>%  
 mutate("W" = if\_else(condition = is.na(W), true = 0, false = W)) %>%  
 mutate("L" = if\_else(condition = is.na(L), true = 0, false = L)) %>%  
 mutate("QS" = if\_else(condition = is.na(QS), true = 0, false = QS)) %>%  
 mutate("FPTS" = as.double(FPTS), "ERA" = as.double(ERA), "SO" = as.double(SO)) %>%  
 mutate("IP" = (10\*(as.double(IP)) - 7\*(round(as.double(IP))))/3) %>%  
 mutate("FPTS.ESPN" = round(FPTS - W\*7 + W\*2 + L\*5 - L\*2 - QS\*3 - (ERA \* IP / 9) + SO\*0.5)) %>%  
 mutate("FPPG.ESPN" = round(FPTS.ESPN / as.double(GP), 1)) %>%  
 mutate("KW" = round(as.double(`SO9`)/as.double(WHIP), 2)) %>%  
 select(Player, FPTS.ESPN, FPPG.ESPN, KW, matches("."), -full\_name) %>%  
 filter(GS > 7)  
head(yearlystats)

## Player FPTS.ESPN FPPG.ESPN KW FPTS FPPG W L ERA GP GS QS CG  
## 1 Tarik Skubal 368 17.5 11.59 424.0 20.2 12 3 2.35 21 21 16 —  
## 2 Seth Lugo 326 14.8 7.54 399.0 18.1 12 5 2.66 22 22 16 1  
## 3 Chris Sale 337 17.7 12.42 386.5 20.3 13 3 2.68 19 19 11 —  
## 4 Corbin Burnes 320 15.2 7.93 384.5 18.3 10 4 2.45 21 21 17 —  
## 5 Logan Gilbert 342 16.3 9.94 381.0 18.1 6 6 2.72 21 21 17 —  
## 6 Zack Wheeler 318 15.9 9.88 369.5 18.5 10 4 2.55 20 20 15 —  
## IP H HR BB SO AVG WHIP SV BS SHO HLD GB% SO9 BB9  
## 1 130.0000 96 11 24 154 .199 0.92 — — — — 1.09 10.66 1.66  
## 2 142.3333 116 12 32 124 .221 1.04 — — — — 1.07 7.84 2.02  
## 3 117.3333 83 8 25 149 .196 0.92 — — — — 1.38 11.43 1.92  
## 4 132.0000 107 15 30 121 .217 1.04 — — — — 1.14 8.25 2.05  
## 5 139.0000 94 15 26 132 .190 0.86 — — — — 0.93 8.55 1.68  
## 6 123.6667 84 11 37 133 .189 0.98 — — — — 0.92 9.68 2.69

Here, we have the adjusted points and adjusted points per game, which is supposed to score to an ESPN league settings (Only difference is that Hit by Pitches being scored as -1 was not removed). Based on the data as of 7/25/2024, the top 6 pitchers according to CBS rankings are Tarik Skubal, Seth Lugo, Logan Gilbert, Zack Wheeler, Corbin Burnes, and Aaron Nola. CBS scoring is different in many ways compared to ESPN. For example, wins and losses are more important. A pitcher can receive 7 points for a win and lose 5 for a loss, where as ESPN scores wins as +2 and losses as -2. Plus, quality starts are rewarded in CBS scoring as +3, while ESPN scoring does not reward pitchers for quality starts. Those are only a few differences between the two sites which can impact fantasy scoring a lot. This analysis chooses ESPN scoring for the analysis, but some of the ideas can also be applied for CBS scoring.

head(yearlystats %>%  
 arrange(desc(FPTS.ESPN)), 15)

## Player FPTS.ESPN FPPG.ESPN KW FPTS FPPG W L ERA GP GS QS CG  
## 1 Tarik Skubal 368 17.5 11.59 424.0 20.2 12 3 2.35 21 21 16 —  
## 2 Logan Gilbert 342 16.3 9.94 381.0 18.1 6 6 2.72 21 21 17 —  
## 3 Chris Sale 337 17.7 12.42 386.5 20.3 13 3 2.68 19 19 11 —  
## 4 Dylan Cease 332 15.1 11.78 364.0 16.5 10 8 3.50 22 22 13 1  
## 5 Seth Lugo 326 14.8 7.54 399.0 18.1 12 5 2.66 22 22 16 1  
## 6 Corbin Burnes 320 15.2 7.93 384.5 18.3 10 4 2.45 21 21 17 —  
## 7 Zack Wheeler 318 15.9 9.88 369.5 18.5 10 4 2.55 20 20 15 —  
## 8 George Kirby 303 13.8 9.07 346.5 15.8 8 7 3.03 22 22 15 —  
## 9 Garrett Crochet 297 13.5 12.47 297.0 13.5 6 8 3.23 22 22 11 —  
## 10 Tyler Glasnow 296 15.6 12.09 321.5 16.9 8 6 3.47 19 19 11 —  
## 11 Aaron Nola 291 13.9 8.19 366.0 17.4 11 4 3.44 21 21 15 1  
## 12 Cole Ragans 290 13.2 9.17 320.5 14.6 7 7 3.37 22 22 15 1  
## 13 Hunter Greene 288 13.7 9.51 313.0 14.9 7 4 2.97 21 21 10 —  
## 14 Michael King 286 13.0 8.98 316.0 14.4 9 6 3.26 22 21 10 —  
## 15 Jack Flaherty 282 15.7 11.69 303.5 16.9 7 5 2.95 18 18 11 —  
## IP H HR BB SO AVG WHIP SV BS SHO HLD GB% SO9 BB9  
## 1 130.0000 96 11 24 154 .199 0.92 — — — — 1.09 10.66 1.66  
## 2 139.0000 94 15 26 132 .190 0.86 — — — — 0.93 8.55 1.68  
## 3 117.3333 83 8 25 149 .196 0.92 — — — — 1.38 11.43 1.92  
## 4 131.0000 89 15 40 168 .191 0.98 — — 1 — 0.74 11.54 2.75  
## 5 142.3333 116 12 32 124 .221 1.04 — — — — 1.07 7.84 2.02  
## 6 132.0000 107 15 30 121 .217 1.04 — — — — 1.14 8.25 2.05  
## 7 123.6667 84 11 37 133 .189 0.98 — — — — 0.92 9.68 2.69  
## 8 130.6667 114 11 14 129 .230 0.98 — — — — 0.83 8.89 0.96  
## 9 114.3333 89 11 26 160 .207 1.01 — — — — 0.92 12.59 2.05  
## 10 114.0000 76 13 33 147 .185 0.96 — — — — 1.16 11.61 2.61  
## 11 130.6667 108 18 31 126 .222 1.06 — — 1 — 1.14 8.68 2.14  
## 12 128.3333 109 9 41 153 .226 1.17 — — — — 0.81 10.73 2.88  
## 13 124.3333 82 10 49 138 .188 1.05 — — — — 0.60 9.99 3.55  
## 14 124.3333 99 14 45 144 .214 1.16 — — — — 0.84 10.42 3.26  
## 15 106.6667 83 15 19 133 .211 0.96 — — — — 1.01 11.22 1.60

Here are the standings for ESPN standard league scoring. The two pitchers to note here is Seth Lugo and Logan Gilbert. Back in the CBS Scoring, Lugo is ahead of Gilbert in points and if you notice the column FPPG (fantasy points per game), Lugo has a higher average than Gilbert for CBS, but it’s the opposite for ESPN. This shows that there are differences in scoring between the two sites that changes the value of pitchers. For the purposes of this code, we are looking at ESPN scoring more. One thing to note is the averages vs the points itself. Some pitchers have higher scores and lower averages than others. This might be due to more starts, or an injury, or the player was called up mid-season (aka Paul Skenes). So now the question becomes what is more valuable for a pitcher, the total points or the average? I think that there should be a mix of both, but favored to the average points per start. Pitcher starts in most standard ESPN leagues are capped to a certain amount, so it is important to get the most out of each pitcher start.

Furthermore, I filtered the number of games started to 7 to remove relief pitchers (some of them “start” a match and then is followed by a bulk pitcher who plays majority of the game) and to remove players who made brief stints and do not have a big enough sample size worthy of inclusion (e.g Shane Bieber had two starts and averaged around 25 across those two starts). Here are the top pitchers based on averages:

head(yearlystats %>%  
 arrange(desc(FPPG.ESPN)), 15)

## Player FPTS.ESPN FPPG.ESPN KW FPTS FPPG W L ERA GP GS QS CG  
## 1 Paul Skenes 231 19.2 13.44 255.5 21.3 6 1 1.93 12 12 10 —  
## 2 Chris Sale 337 17.7 12.42 386.5 20.3 13 3 2.68 19 19 11 —  
## 3 Tarik Skubal 368 17.5 11.59 424.0 20.2 12 3 2.35 21 21 16 —  
## 4 Logan Gilbert 342 16.3 9.94 381.0 18.1 6 6 2.72 21 21 17 —  
## 5 Zack Wheeler 318 15.9 9.88 369.5 18.5 10 4 2.55 20 20 15 —  
## 6 Jack Flaherty 282 15.7 11.69 303.5 16.9 7 5 2.95 18 18 11 —  
## 7 Tyler Glasnow 296 15.6 12.09 321.5 16.9 8 6 3.47 19 19 11 —  
## 8 Taj Bradley 218 15.6 10.62 230.5 16.5 6 4 2.43 14 14 7 —  
## 9 Corbin Burnes 320 15.2 7.93 384.5 18.3 10 4 2.45 21 21 17 —  
## 10 Dylan Cease 332 15.1 11.78 364.0 16.5 10 8 3.50 22 22 13 1  
## 11 Seth Lugo 326 14.8 7.54 399.0 18.1 12 5 2.66 22 22 16 1  
## 12 Ronel Blanco 280 14.0 8.62 325.0 16.3 9 5 2.95 20 20 11 1  
## 13 Aaron Nola 291 13.9 8.19 366.0 17.4 11 4 3.44 21 21 15 1  
## 14 George Kirby 303 13.8 9.07 346.5 15.8 8 7 3.03 22 22 15 —  
## 15 Ranger Suarez 276 13.8 8.25 324.0 16.2 10 5 2.87 20 20 11 1  
## IP H HR BB SO AVG WHIP SV BS SHO HLD GB% SO9 BB9  
## 1 74.66667 52 8 13 97 .195 0.87 — — — — 1.38 11.69 1.57  
## 2 117.33333 83 8 25 149 .196 0.92 — — — — 1.38 11.43 1.92  
## 3 130.00000 96 11 24 154 .199 0.92 — — — — 1.09 10.66 1.66  
## 4 139.00000 94 15 26 132 .190 0.86 — — — — 0.93 8.55 1.68  
## 5 123.66667 84 11 37 133 .189 0.98 — — — — 0.92 9.68 2.69  
## 6 106.66667 83 15 19 133 .211 0.96 — — — — 1.01 11.22 1.60  
## 7 114.00000 76 13 33 147 .185 0.96 — — — — 1.16 11.61 2.61  
## 8 81.33333 55 10 27 97 .189 1.01 — — — — 0.88 10.73 2.99  
## 9 132.00000 107 15 30 121 .217 1.04 — — — — 1.14 8.25 2.05  
## 10 131.00000 89 15 40 168 .191 0.98 — — 1 — 0.74 11.54 2.75  
## 11 142.33333 116 12 32 124 .221 1.04 — — — — 1.07 7.84 2.02  
## 12 119.00000 73 18 46 114 .175 1.00 — — 1 — 0.71 8.62 3.48  
## 13 130.66667 108 18 31 126 .222 1.06 — — 1 — 1.14 8.68 2.14  
## 14 130.66667 114 11 14 129 .230 0.98 — — — — 0.83 8.89 0.96  
## 15 119.33333 100 9 27 116 .224 1.06 — — 1 — 1.52 8.75 2.04

Here, Pirates ace Paul Skenes is leading the averages with 19.2 FPPG. He got called up mid-season, so he as of now is only at 12 starts, while Tarik Skubal who is second in FPPG has 20. Another pitcher, Taj Bradley, is 10th in average despite having significantly less points because he started the season on the injured list and therefore also only has 13 games started. However, there is a certain point where less starts should not take away anything from the pitchers. Both Skenes and Bradley have double digit starts with these numbers and averages, so fantasy managers can be justified in preferring those guys over others who have more points but lower averages.

But there is more to scoring a pitcher than points and raw averages. The average pitcher start is about 5 or 6 innings and ESPN rewards 1 point for every 0.1 inning (so 3 points per 1 inning), but pitchers also get scored on strikeouts (+ 1 per strikeout) with deductions for any offensive production (-1 for walks allowed, hits allowed, hit by pitches, with -2 for every earned run conceded). I think that the best pitchers in fantasy are the ones that have the best strikeout numbers while maintaining the lowest base runners. There is a column named KW. KW is the ratio between a pitchers K/9 (Strikeout per 9 innings) and WHIP (Walk and Hits per Innings Pitched). The stat was calculated by dividing K/9 over WHIP. Here are the top pitchers for KW:

head(yearlystats %>%  
 arrange(desc(KW)), 15)

## Player FPTS.ESPN FPPG.ESPN KW FPTS FPPG W L ERA GP GS QS  
## 1 Paul Skenes 231 19.2 13.44 255.5 21.3 6 1 1.93 12 12 10  
## 2 Garrett Crochet 297 13.5 12.47 297.0 13.5 6 8 3.23 22 22 11  
## 3 Chris Sale 337 17.7 12.42 386.5 20.3 13 3 2.68 19 19 11  
## 4 Tyler Glasnow 296 15.6 12.09 321.5 16.9 8 6 3.47 19 19 11  
## 5 Dylan Cease 332 15.1 11.78 364.0 16.5 10 8 3.50 22 22 13  
## 6 Jack Flaherty 282 15.7 11.69 303.5 16.9 7 5 2.95 18 18 11  
## 7 Tarik Skubal 368 17.5 11.59 424.0 20.2 12 3 2.35 21 21 16  
## 8 Taj Bradley 218 15.6 10.62 230.5 16.5 6 4 2.43 14 14 7  
## 9 Sonny Gray 262 13.8 10.11 296.0 15.6 10 6 3.79 19 19 8  
## 10 Logan Gilbert 342 16.3 9.94 381.0 18.1 6 6 2.72 21 21 17  
## 11 Zack Wheeler 318 15.9 9.88 369.5 18.5 10 4 2.55 20 20 15  
## 12 Joe Ryan 280 13.3 9.81 308.0 14.7 6 7 3.69 21 21 12  
## 13 Luis Gil 259 12.9 9.63 296.0 14.8 10 5 3.10 20 20 9  
## 14 Yoshinobu Yamamoto 186 13.3 9.55 213.0 15.2 6 2 2.92 14 14 7  
## 15 Hunter Greene 288 13.7 9.51 313.0 14.9 7 4 2.97 21 21 10  
## CG IP H HR BB SO AVG WHIP SV BS SHO HLD GB% SO9 BB9  
## 1 — 74.66667 52 8 13 97 .195 0.87 — — — — 1.38 11.69 1.57  
## 2 — 114.33333 89 11 26 160 .207 1.01 — — — — 0.92 12.59 2.05  
## 3 — 117.33333 83 8 25 149 .196 0.92 — — — — 1.38 11.43 1.92  
## 4 — 114.00000 76 13 33 147 .185 0.96 — — — — 1.16 11.61 2.61  
## 5 1 131.00000 89 15 40 168 .191 0.98 — — 1 — 0.74 11.54 2.75  
## 6 — 106.66667 83 15 19 133 .211 0.96 — — — — 1.01 11.22 1.60  
## 7 — 130.00000 96 11 24 154 .199 0.92 — — — — 1.09 10.66 1.66  
## 8 — 81.33333 55 10 27 97 .189 1.01 — — — — 0.88 10.73 2.99  
## 9 — 111.66667 96 12 27 138 .230 1.10 — — — — 1.17 11.12 2.18  
## 10 — 139.00000 94 15 26 132 .190 0.86 — — — — 0.93 8.55 1.68  
## 11 — 123.66667 84 11 37 133 .189 0.98 — — — — 0.92 9.68 2.69  
## 12 — 126.66667 107 17 20 138 .223 1.00 — — — — 0.68 9.81 1.42  
## 13 — 107.33333 66 9 50 124 .173 1.08 — — — — 0.70 10.40 4.19  
## 14 — 74.00000 62 6 17 84 .221 1.07 — — — — 1.25 10.22 2.07  
## 15 — 124.33333 82 10 49 138 .188 1.05 — — — — 0.60 9.99 3.55

7 of the top 10 pitchers in KW are in the top 10 for FPPG, the other three pitchers (Crochet, Cease, and Gray) are in the top 15 for FPPG. While people can look at WHIP alone because that can be a indication of who is good for pitchers, KW is arguably better for fantasy purposes because strikeouts play a big role for scoring pitchers, so getting a pitcher who strikes batters out a lot, but allows some more base-runners is not the worst way to go.

Garrett Crochet is likely going to be an exception for this analysis and the rest of the season. The White Sox announced the Crochet is going to be on a pitch count as he has exceeded his innings pitched in a season and if they trade him, his suitor will also value keeping him healthy and sharp come the playoffs. His last two starts had Crochet depart after 2 innings then 4 innings which brought down his average. Looking at fantasy points per innings pitched is not a bad way to go for future seasons, but while the cap in most standard leagues is pitcher starts vs innings pitched, we have to keep in mind Crochet is almost guaranteed to not put up the same fantasy points as he did in the first half of the season.

We now should probably check outlier performances. There is an .R file Fantasy Baseball Pitching Outliers that takes the game logs of the top scoring starting pitchers and removes any outlier performances for better or worse. The data was extracted from Fox Sports and each game log does not have the decision for pitchers, so we have to temporarily remove points from wins and losses. While this does change the pitcher scores and averages, it will not drastically change the landscape of pitcher scoring and value for a pitcher should be mostly the same when comparing them to all the other pitchers.

head(yearlystats %>%  
 mutate(FPTS.ESPN = FPTS.ESPN - 2\*W + 2\*L) %>%  
 mutate(FPPG.ESPN = round(FPTS.ESPN / as.double(GP), 1)), 15)

## Player FPTS.ESPN FPPG.ESPN KW FPTS FPPG W L ERA GP GS QS CG  
## 1 Tarik Skubal 350 16.7 11.59 424.0 20.2 12 3 2.35 21 21 16 —  
## 2 Seth Lugo 312 14.2 7.54 399.0 18.1 12 5 2.66 22 22 16 1  
## 3 Chris Sale 317 16.7 12.42 386.5 20.3 13 3 2.68 19 19 11 —  
## 4 Corbin Burnes 308 14.7 7.93 384.5 18.3 10 4 2.45 21 21 17 —  
## 5 Logan Gilbert 342 16.3 9.94 381.0 18.1 6 6 2.72 21 21 17 —  
## 6 Zack Wheeler 306 15.3 9.88 369.5 18.5 10 4 2.55 20 20 15 —  
## 7 Aaron Nola 277 13.2 8.19 366.0 17.4 11 4 3.44 21 21 15 1  
## 8 Dylan Cease 328 14.9 11.78 364.0 16.5 10 8 3.50 22 22 13 1  
## 9 George Kirby 301 13.7 9.07 346.5 15.8 8 7 3.03 22 22 15 —  
## 10 Ronel Blanco 272 13.6 8.62 325.0 16.3 9 5 2.95 20 20 11 1  
## 11 Ranger Suarez 266 13.3 8.25 324.0 16.2 10 5 2.87 20 20 11 1  
## 12 Tyler Glasnow 292 15.4 12.09 321.5 16.9 8 6 3.47 19 19 11 —  
## 13 Cole Ragans 290 13.2 9.17 320.5 14.6 7 7 3.37 22 22 15 1  
## 14 Tanner Houck 272 13.0 7.29 320.0 15.2 8 7 2.79 21 21 15 1  
## 15 Bailey Ober 251 12.6 9.38 316.5 15.8 10 5 3.76 20 20 11 1  
## IP H HR BB SO AVG WHIP SV BS SHO HLD GB% SO9 BB9  
## 1 130.0000 96 11 24 154 .199 0.92 — — — — 1.09 10.66 1.66  
## 2 142.3333 116 12 32 124 .221 1.04 — — — — 1.07 7.84 2.02  
## 3 117.3333 83 8 25 149 .196 0.92 — — — — 1.38 11.43 1.92  
## 4 132.0000 107 15 30 121 .217 1.04 — — — — 1.14 8.25 2.05  
## 5 139.0000 94 15 26 132 .190 0.86 — — — — 0.93 8.55 1.68  
## 6 123.6667 84 11 37 133 .189 0.98 — — — — 0.92 9.68 2.69  
## 7 130.6667 108 18 31 126 .222 1.06 — — 1 — 1.14 8.68 2.14  
## 8 131.0000 89 15 40 168 .191 0.98 — — 1 — 0.74 11.54 2.75  
## 9 130.6667 114 11 14 129 .230 0.98 — — — — 0.83 8.89 0.96  
## 10 119.0000 73 18 46 114 .175 1.00 — — 1 — 0.71 8.62 3.48  
## 11 119.3333 100 9 27 116 .224 1.06 — — 1 — 1.52 8.75 2.04  
## 12 114.0000 76 13 33 147 .185 0.96 — — — — 1.16 11.61 2.61  
## 13 128.3333 109 9 41 153 .226 1.17 — — — — 0.81 10.73 2.88  
## 14 129.0000 110 6 33 116 .224 1.11 — — 1 — 1.62 8.09 2.30  
## 15 115.0000 89 17 27 121 .212 1.01 — — — — 0.66 9.47 2.11

Now, let’s load the csv file outlierstats.csv. To view the code for how I compiled those statistics, check there is an .R file called Fantasy Baseball Pitching Outliers in the folder. For these next statistics, we are using just the top 100 pitchers.

All statistics again were compiled as of 7/25. Let’s see who is impacted the most by outlier performances for better or for worse. These are the 15 pitchers who benefit the most from removing outlier performances.

head(outlierstats %>%  
 arrange(DiffPTS), 15)

## # A tibble: 15 × 8  
## Player FPTS.ESPN FPPG.ESPN KW FPTS.ESPN.adj FPPG.ESPN.adj KW.adj DiffPTS  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Hunter … 195 9.29 7.03 223 11.2 7.63 -28  
## 2 Lance L… 159 7.95 6.01 183 9.63 6.46 -24  
## 3 Taj Bra… 214 15.3 10.6 231 17.8 12.0 -17  
## 4 Chris S… 317 16.7 12.4 327 18.2 13.4 -10  
## 5 George … 280 13.3 9.06 285 13.6 9.15 -5  
## 6 Carlos … 202 9.62 7.77 207 9.86 7.83 -5  
## 7 Kutter … 247 12.4 8.16 251 12.0 8.13 -4  
## 8 Cristop… 209 11 5.93 212 11.2 5.97 -3  
## 9 Zack Wh… 306 15.3 9.88 307 15.4 9.98 -1  
## 10 Luis Gil 249 12.4 9.63 250 12.5 9.72 -1  
## 11 Michael… 260 12.4 8.66 261 12.4 8.73 -1  
## 12 Logan W… 229 10.4 5.8 230 10.4 5.84 -1  
## 13 Tarik S… 332 16.6 11.6 332 16.6 11.6 0  
## 14 Seth Lu… 310 14.8 7.76 310 14.8 7.76 0  
## 15 Corbin … 308 14.7 7.93 308 14.7 7.93 0

The last 3 pitchers (Skubal, Lugo, Burnes) have a DiffPTS of 0, meaning that they have not had any outlier performances for better or worse. They have been relatively consistent all season. For the other 12 pitchers, they have all have had at least one performance (likely just one) that have been so far off their season totals and averages that they have been considered outliers. There are some high profile pitchers, such as Chris Sale, Taj Bradley, and Zack Wheeler who fit have had these outlier performances. If you exclude them, their current averages (16.68, 15.29, 15.3) jump up to 18.17, 17.77, and 15.35 points per start. For Sale and Bradley, their average points per start jumps up by a point by simply taking out an outlier start. In real life and fantasy, it is harder to simply exclude these blowup starts, but in the grand scheme of an entire season full of 20+ starts, it is just one start that can be forgiven by performing at a higher level the rest of the season.

Likewise, let’s see which pitchers have had such outstanding starts, they were considered outliers.

head(outlierstats %>%  
 arrange(desc(DiffPTS)), 5)

## # A tibble: 5 × 8  
## Player FPTS.ESPN FPPG.ESPN KW FPTS.ESPN.adj FPPG.ESPN.adj KW.adj DiffPTS  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 MacKenzi… 172 8.6 7.23 167 8.35 7.18 5  
## 2 Marcus S… 176 8.8 5.17 173 8.65 5.17 3  
## 3 Joe Ryan 267 13.4 9.6 265 13.2 9.6 2  
## 4 Charlie … 185 10.3 7.21 183 10.2 7.21 2  
## 5 Tarik Sk… 332 16.6 11.6 332 16.6 11.6 0

Only four pitchers have had outlier performances that were good performances (Skubal has 0, so he’s been consistent all season). These four pitchers have had single game performances so good that they have been considered outliers, due to the fact they have been nowhere close to their numbers all season.

Let’s see the grand scheme of fantasy pitchers, how they stack up in terms of average if outlier performances were excluded.

# Sorted by stats with outliers  
head(outlierstats %>%  
 arrange(desc(FPPG.ESPN)) %>%  
 select(Player, `FPPG.ESPN`), 15)

## # A tibble: 15 × 2  
## Player FPPG.ESPN  
## <chr> <dbl>  
## 1 Paul Skenes 18.4  
## 2 Chris Sale 16.7  
## 3 Tarik Skubal 16.6  
## 4 Logan Gilbert 16.3  
## 5 Jack Flaherty 15.4  
## 6 Tyler Glasnow 15.4  
## 7 Zack Wheeler 15.3  
## 8 Taj Bradley 15.3  
## 9 Dylan Cease 14.9  
## 10 Seth Lugo 14.8  
## 11 Corbin Burnes 14.7  
## 12 Garrett Crochet 14.4  
## 13 Sonny Gray 14   
## 14 Ronel Blanco 14.0  
## 15 Cole Ragans 13.5

# Sorted by stats without outliers  
head(outlierstats %>%  
 arrange(desc(FPPG.ESPN.adj)) %>%  
 select(Player, `FPPG.ESPN.adj`), 15)

## # A tibble: 15 × 2  
## Player FPPG.ESPN.adj  
## <chr> <dbl>  
## 1 Paul Skenes 18.4  
## 2 Chris Sale 18.2  
## 3 Taj Bradley 17.8  
## 4 Tarik Skubal 16.6  
## 5 Logan Gilbert 16.3  
## 6 Jack Flaherty 15.4  
## 7 Tyler Glasnow 15.4  
## 8 Zack Wheeler 15.4  
## 9 Dylan Cease 14.9  
## 10 Seth Lugo 14.8  
## 11 Corbin Burnes 14.7  
## 12 Garrett Crochet 14.4  
## 13 Sonny Gray 14   
## 14 Ronel Blanco 14.0  
## 15 George Kirby 13.6

Again, keep in mind wins and losses have been removed temporarily for these numbers, these calculations are based solely on the pitcher skills. Taj Bradley had the jump from 8th highest average to 3rd highest average, indicating that he had a blowup start (June 1st @ Orioles, conceded 9 ER in 3.1 IP, -17 points not including the loss) that has been ruled an outlier (again, more info on the outliers calculated can be found in the Outliers .R file). None of the other pitchers were really impacted by the adjusted stats, despite guys like Wheeler and Sale also having better stats with outlier stats removed.

Note that I did create a stat called KW.adj, which recalculates the K/9 / WHIP stats. Outlier performances often blow up a pitcher’s WHIP and sometimes will stall their K/9 stats too. Here is the pitchers ranked by KW vs KW.adj:

# Sorted by KW with outliers  
head(outlierstats %>%  
 arrange(desc(KW.adj)) %>%  
 select(Player, KW), 15)

## # A tibble: 15 × 2  
## Player KW  
## <chr> <dbl>  
## 1 Paul Skenes 13.4   
## 2 Chris Sale 12.4   
## 3 Garrett Crochet 13.1   
## 4 Tyler Glasnow 12.1   
## 5 Taj Bradley 10.6   
## 6 Dylan Cease 11.8   
## 7 Jack Flaherty 11.7   
## 8 Tarik Skubal 11.6   
## 9 Sonny Gray 10.4   
## 10 Zack Wheeler 9.88  
## 11 Logan Gilbert 9.94  
## 12 Freddy Peralta 9.74  
## 13 Luis Gil 9.63  
## 14 Joe Ryan 9.6   
## 15 Yoshinobu Yamamoto 9.55

# Sorted by KW without outliers  
head(outlierstats %>%  
 arrange(desc(KW.adj)) %>%  
 select(Player, KW.adj), 15)

## # A tibble: 15 × 2  
## Player KW.adj  
## <chr> <dbl>  
## 1 Paul Skenes 13.4   
## 2 Chris Sale 13.4   
## 3 Garrett Crochet 13.1   
## 4 Tyler Glasnow 12.1   
## 5 Taj Bradley 12.0   
## 6 Dylan Cease 11.8   
## 7 Jack Flaherty 11.7   
## 8 Tarik Skubal 11.6   
## 9 Sonny Gray 10.4   
## 10 Zack Wheeler 9.98  
## 11 Logan Gilbert 9.94  
## 12 Freddy Peralta 9.74  
## 13 Luis Gil 9.72  
## 14 Joe Ryan 9.6   
## 15 Yoshinobu Yamamoto 9.55

Similar to FPPG, Sale, Bradley, and Wheeler (to a lesser degree) are the three pitchers impacted by the change, although it is not enough to change their rank. Both Sale and Bradley improved their K/9 / WHIP by over 1 with outliers removed, indicating that those two pitchers have been hurt the most by their outlier performance and that their numbers are being held back by that singular performance (For Sale, June 1st vs Athletics, 4 IP, 8 ER, -10 fantasy points).

Zack Wheeler had a few blowup starts, but his start on June 16th @ Orioles (4.1 IP, 8 ER, -10 fantasy points) seems to be the one that has been ruled far enough as the outlier. His numbers do not drastically change as much as Bradley or Sale though, which might actually highlight more of an inconsistency because Wheeler would have had to have other sub par starts to make the outlier start not that impactful on the season numbers.

So now that we have average fantasy points per game, a reasonable amount of starts for a pitcher to qualify, outlier starts removed, how should starting pitchers be ranked in fantasy baseball. With the KW factor, it is important that the top pitchers seem to have a high amount of strikeouts and a lower WHIP. I do not emphasize innings pitched as much, as the top pitchers in baseball will get around 5, 6 of 7 innings per start. The final standings could be a factor of fantasy points per game adjusted and KW adjusted which leaves with this:

head(outlierstats %>%  
 mutate("ranking\_score" = FPPG.ESPN.adj \* KW.adj) %>%   
 select(Player, ranking\_score) %>%  
 arrange(desc(ranking\_score)), 15)

## # A tibble: 15 × 2  
## Player ranking\_score  
## <chr> <dbl>  
## 1 Paul Skenes 248.  
## 2 Chris Sale 243.  
## 3 Taj Bradley 214.  
## 4 Tarik Skubal 193.  
## 5 Garrett Crochet 188.  
## 6 Tyler Glasnow 186.  
## 7 Jack Flaherty 180.  
## 8 Dylan Cease 176.  
## 9 Logan Gilbert 162.  
## 10 Zack Wheeler 153.  
## 11 Sonny Gray 146.  
## 12 Joe Ryan 127.  
## 13 Cole Ragans 125.  
## 14 George Kirby 124.  
## 15 Hunter Greene 122.

Again, keep in mind that external factors make these rankings not completely accurate projection for the rest of the season. This was made before the trade deadline, and certain pitchers (Garrett Crochet) is on a pitch count, hindering his fantasy value the rest of the season, but if we look at the season up to now, these have been the best pitchers and if one were to exclude the external factors mentioned above, these are the top 15 pitchers in fantasy baseball.

The final ranking shows Paul Skenes as the top pitcher in fantasy baseball. While is is not number 1 in points, this is because he was called up mid-season, but he has been the best pitcher in fantasy baseball since he arrived. He has the highest average points and he has the highest KW factor (he strikes out a lot of batters while not allowing many base runners), which naturally leads him to having the highest ranking score and with this analysis being named the top pitcher in fantasy baseball at this time.