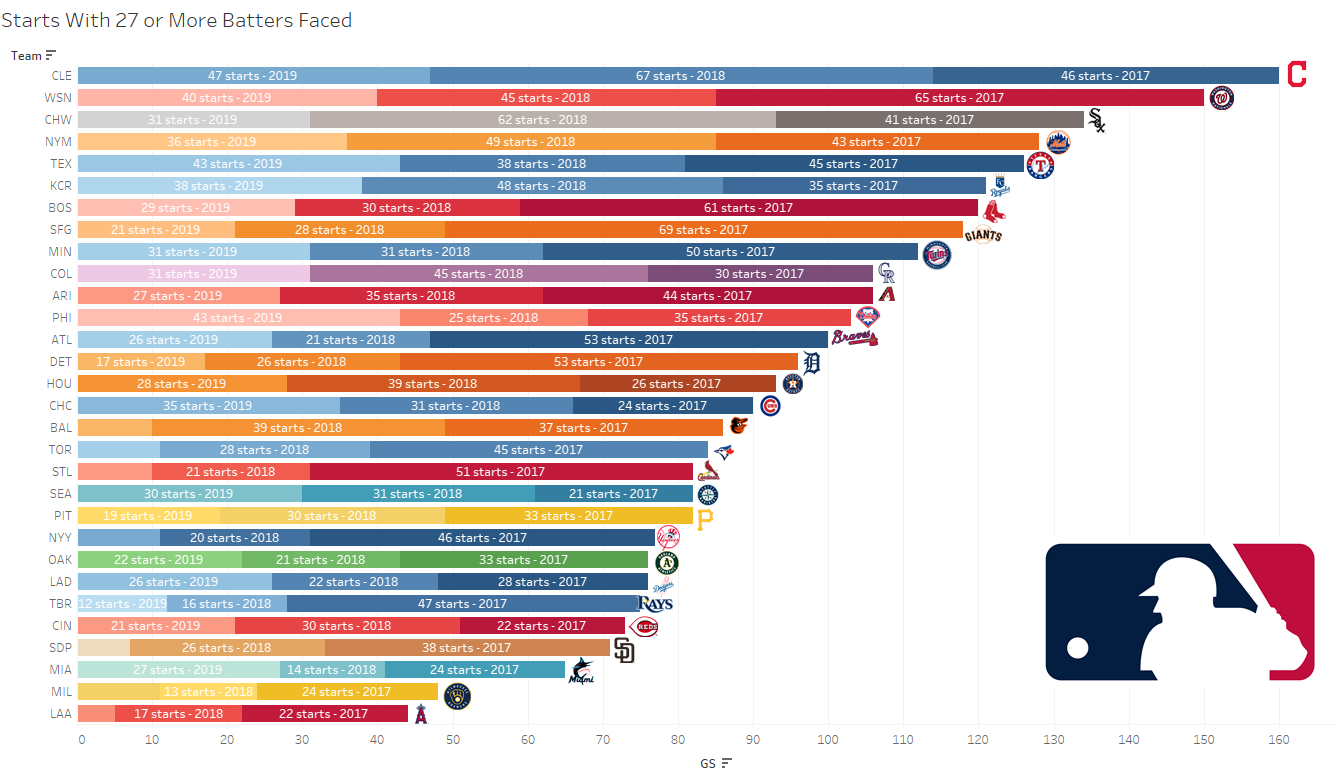
What makes a Good Start a Great Start?

**Zach Greer**

The last few years have seen teams revolutionize the usage of both starters and relievers to mitigate the penalty that starters often experience the third time that a batter has seen them. Teams are pulling starters faster than ever before, often to their benefit. However, this decreases the likelihood of watching a starter work deep into the game when he just has "it". Complete games (and no-hitters/perfect games) might be a thing of the past if trends continue.

|  |  |
| --- | --- |
| Year | Starts with >=27 Batters Faced |
| 2017 | 1191 |
| 2018 | 948 |
| 2019 | 745 |

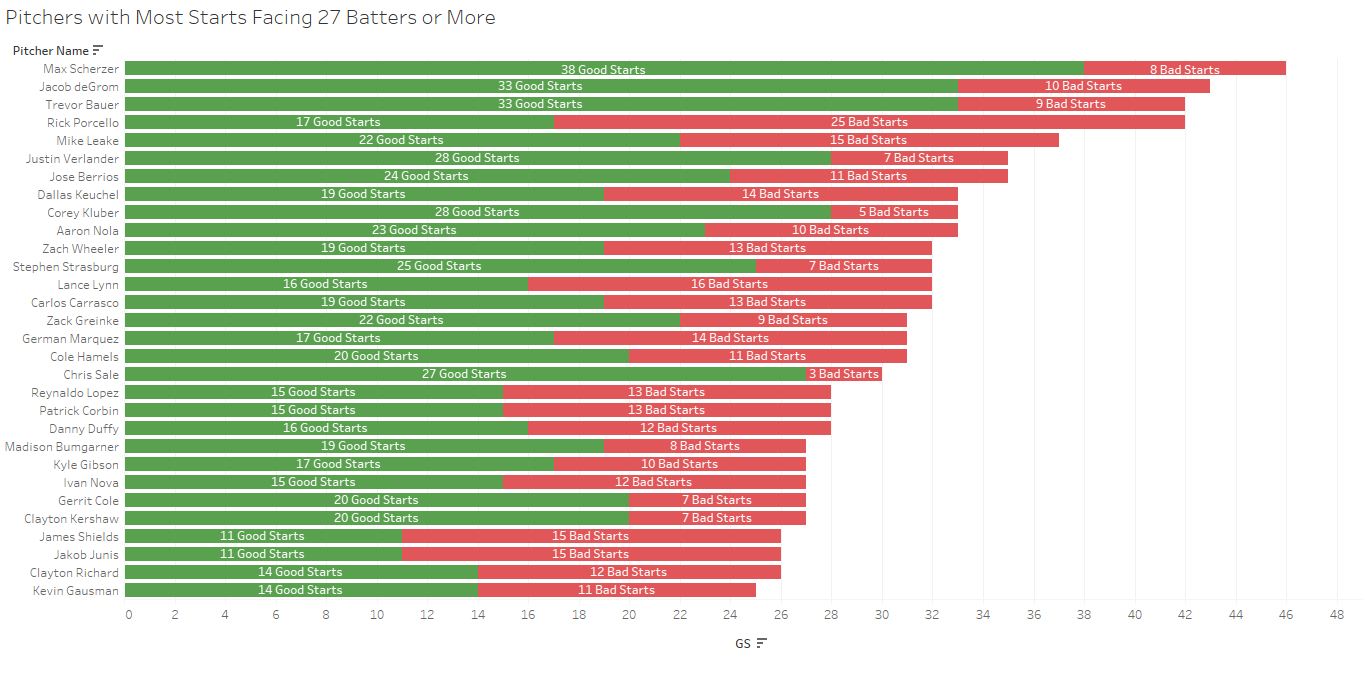
After grouping by team, patterns started to emerge. Teams such as Cleveland and Washington routinely let their starters work deep into the game. Other teams such as Boston and San Francisco have seen a shift in philosophy with new administration.



Some of these starts are the best of the best, but the range of outcomes varies drastically. Pitchers realistically could be facing their 27th batter anywhere from the fourth inning up until the end of the game. To make better sense of the data, one group of good starts and another group of bad starts were formed through a process called k-means clustering.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Start Type | Games | Win % | Innings | Pitches per AB | Hits | HR | K | BB | BABIP | ERA | FIP | xFIP | wOBA | Whiffs per 100 Swings |
| Good Start | 1694 | 75% | 7.34 | 3.58 | 5.21 | 0.5 | 6.73 | 1.45 | 0.242 | 1.90 | 2.95 | 3.68 | 0.243 | 23.44 |
| Bad Start | 1190 | 31% | 5.95 | 3.54 | 7.79 | 1.09 | 4.83 | 2.35 | 0.348 | 6.02 | 5.31 | 4.99 | 0.391 | 19.71 |

The big three of Max Scherzer, Jacob deGrom, and Trevor Bauer at the top are expected, given their team’s willingness to let them work deeper into games. After those three comes Rick Porcello, rarely grouped with his predecessors when it comes to performance. Porcello is being asked to turn through the lineup multiple times, regardless of outcome. On the contrary, another pitcher for the Red Sox from 2017-2019 with a similar number of total starts gave his manger no reason to have a quick hook. Chris Sale had the highest percentage of good starts with 90% while Porcello had the lowest at 40%.

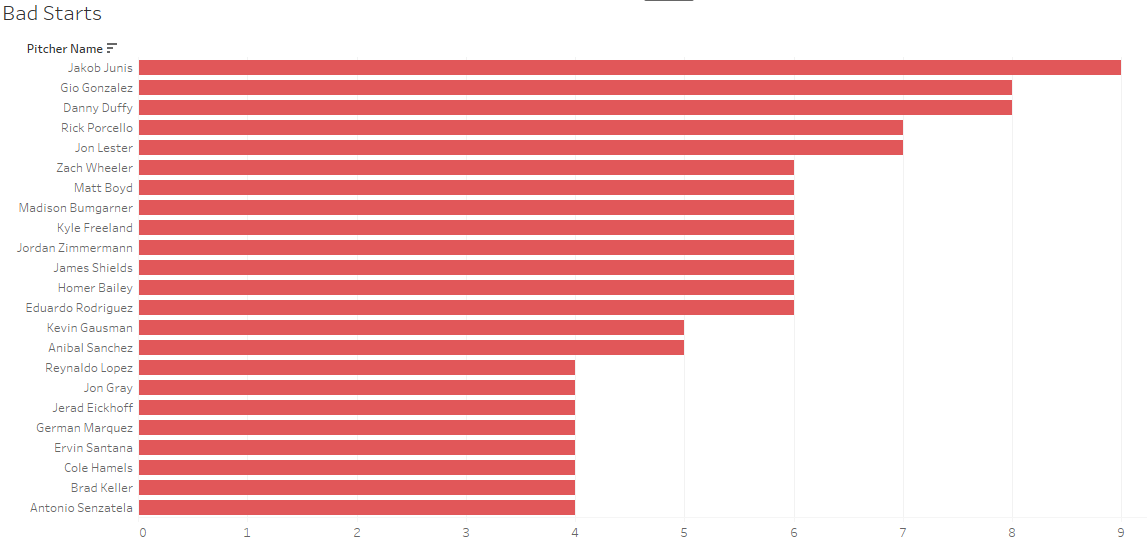


Now that good starts are defined, the composition of those starts can be determined. The two most relevant pitches were identified for each pitcher. The average velocity, x-movement, y-movement, and spinrate of these pitches were then calculated. The percentage difference gap between the first time through the order and the pitcher’s overall metrics was calculated for each game. The same was calculated for the difference between the second time through the order and the first time through the order. Other variables like the First Pitch Strike % and the usage rate for primary pitches were also included.

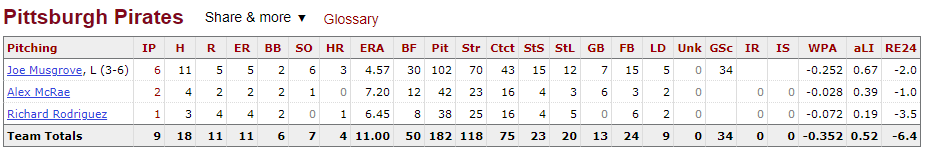
After the first iteration of clustering with pitch metrics included, one cluster emerged with 97% of the group being classified as a “bad start”. These starts were pulled and separated from the sample.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Start Type | Games | Win % | Innings | Pitches per AB | Hits | HR | K | BB | BABIP | ERA | FIP | xFIP | wOBA | Whiffs per 100 Swings |
| Bad Start | 313 | 22% | 5.70 | 3.58 | 8.58 | 1.37 | 5.11 | 2.56 | 0.386 | 7.24 | 6.03 | 5.14 | 0.431 | 21.24 |

These are the starts where the starter often has to shelter the bullpen from extended use. While these starters have been somewhat unlucky with a BABIP of .386 and an xFIP nearly a run/9 lower than FIP and two runs/9 lower than ERA, it’s not great to have HR/9 have a value higher than the K/BB ratio.

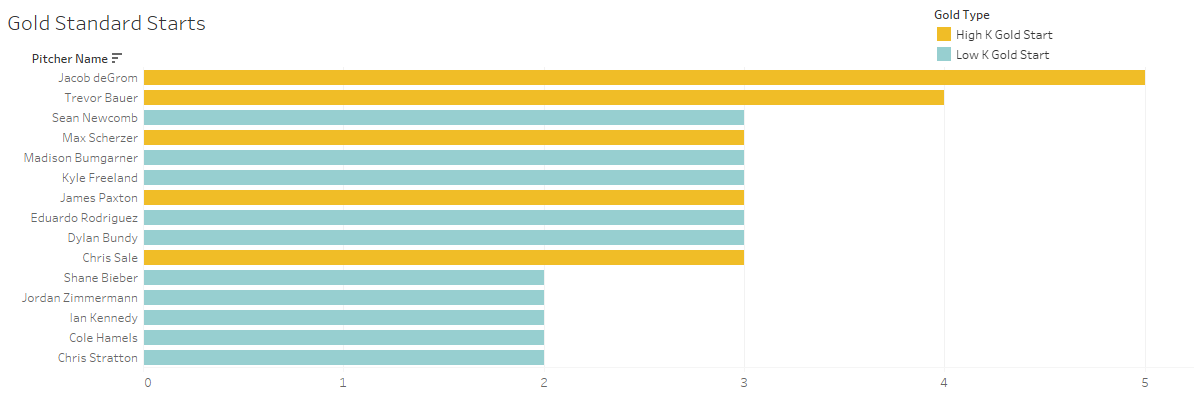


Take this start by Joe Musgrove on May 30, 2019. The Brewers were seeing the ball well, with every starter but one getting a hit off of Musgrove; with Mike Moustakas connecting for 2 HRs. The Pirates bullpen had to cover 16 innings over the previous days after playing a double header on the 27th. They needed someone to work deeper into the game, and Musgrove redeemed himself at the end by retiring 5 of his final 6 opponents, with 3 coming via swinging strikeouts.



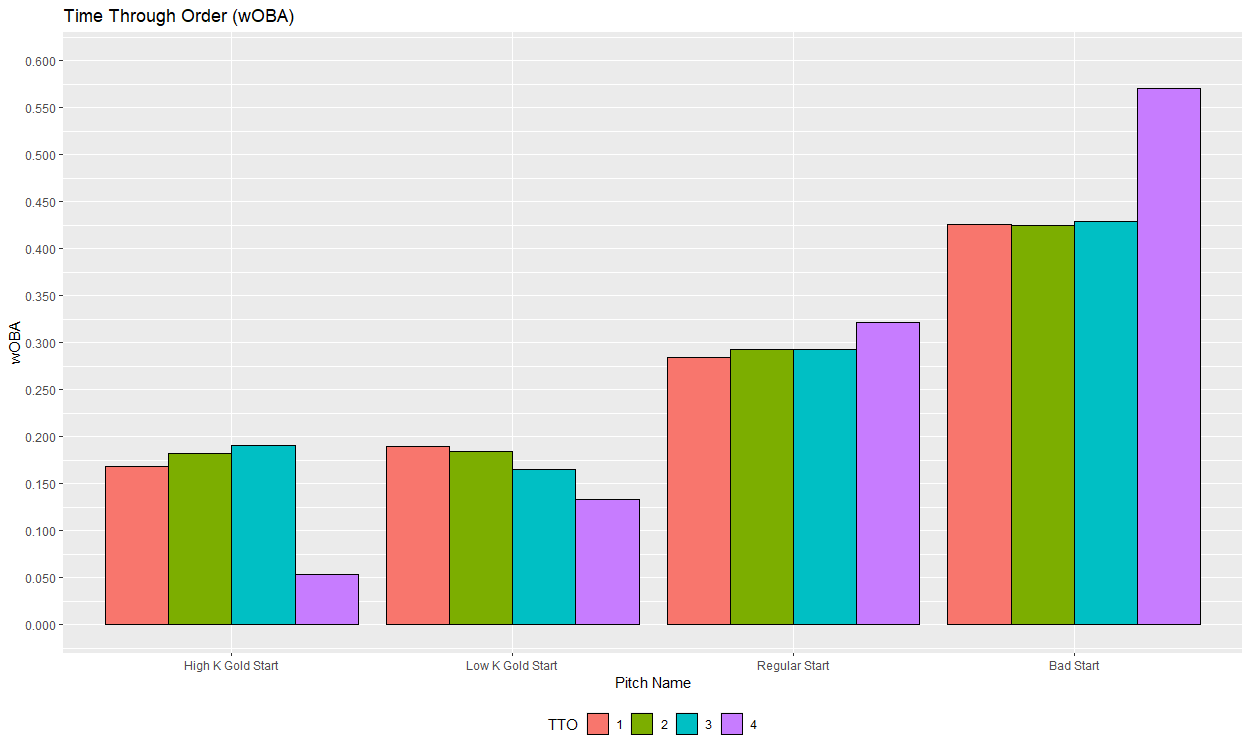
After more iterations of clustering, two groups of exceptional starts began to form.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Start Type | Games | Win % | Innings | Pitches per AB | Hits | HR | K | BB | BABIP | ERA | FIP | xFIP | wOBA | Whiffs per 100 Swings |
| High K Gold Standard | 28 | 92% | 7.98 | 3.69 | 3.64 | 0.25 | 9.86 | 1.39 | 0.201 | 0.81 | 1.80 | 2.92 | 0.173 | 31.15 |
| Low K Gold Standard | 56 | 85% | 7.64 | 3.67 | 3.61 | 0.23 | 6.21 | 1.54 | 0.170 | 1.10 | 2.65 | 4.00 | 0.180 | 22.51 |



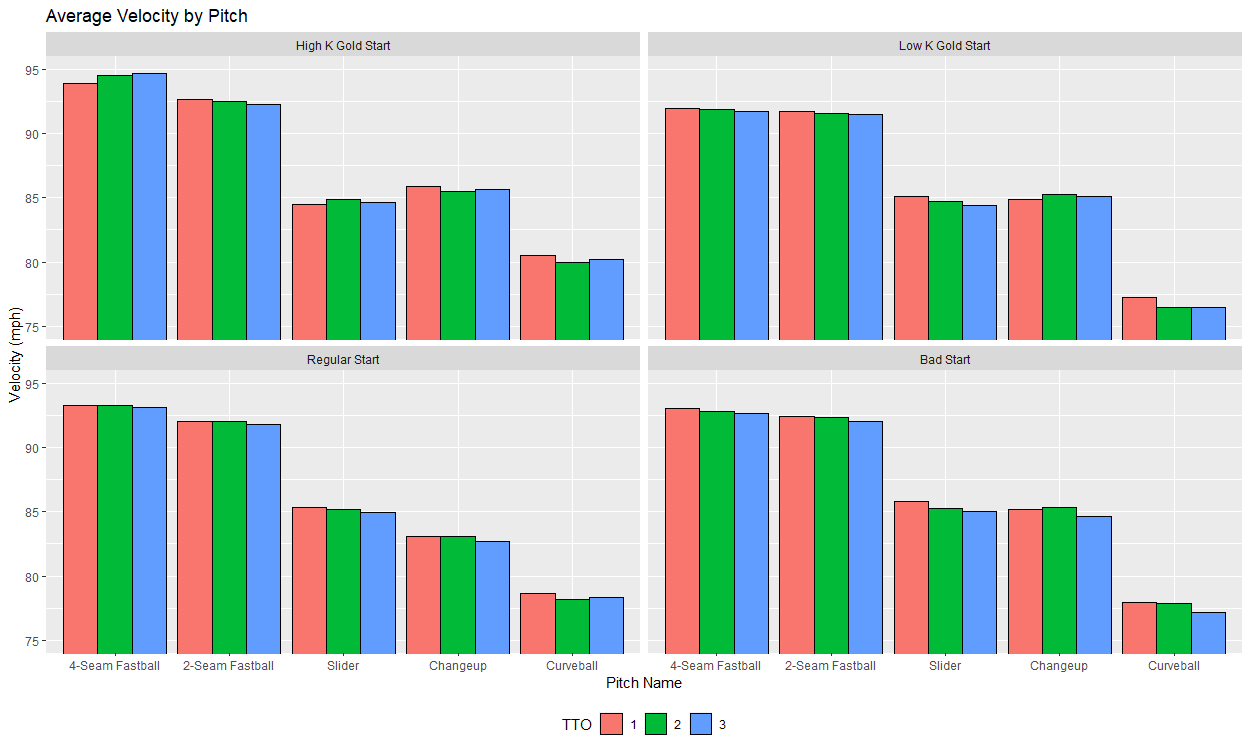
Looking at starters with more than 1 Gold Standard Start further reinforce the divide between the two groups. The High K group includes fireballers like deGrom, Bauer, and Scherzer while the Low K group includes crafty pitchers like Madison Bumgarner and Cole Hamels. More surprising names include Dylan Bundy, Jordan Zimmermann and Ian Kennedy (pre-bullpen conversion) as they performed at a league average level (if that) over the 2017-2019 sample.

All but the Low K group saw gradual increases in wOBA as the game progressed. The Low K group got progressively better deeper into the game. The sample size for the fourth TTO is limited, but the noticeable drop for High K Gold Standard starts could suggest that the starter shifted into another gear to finish the game out.

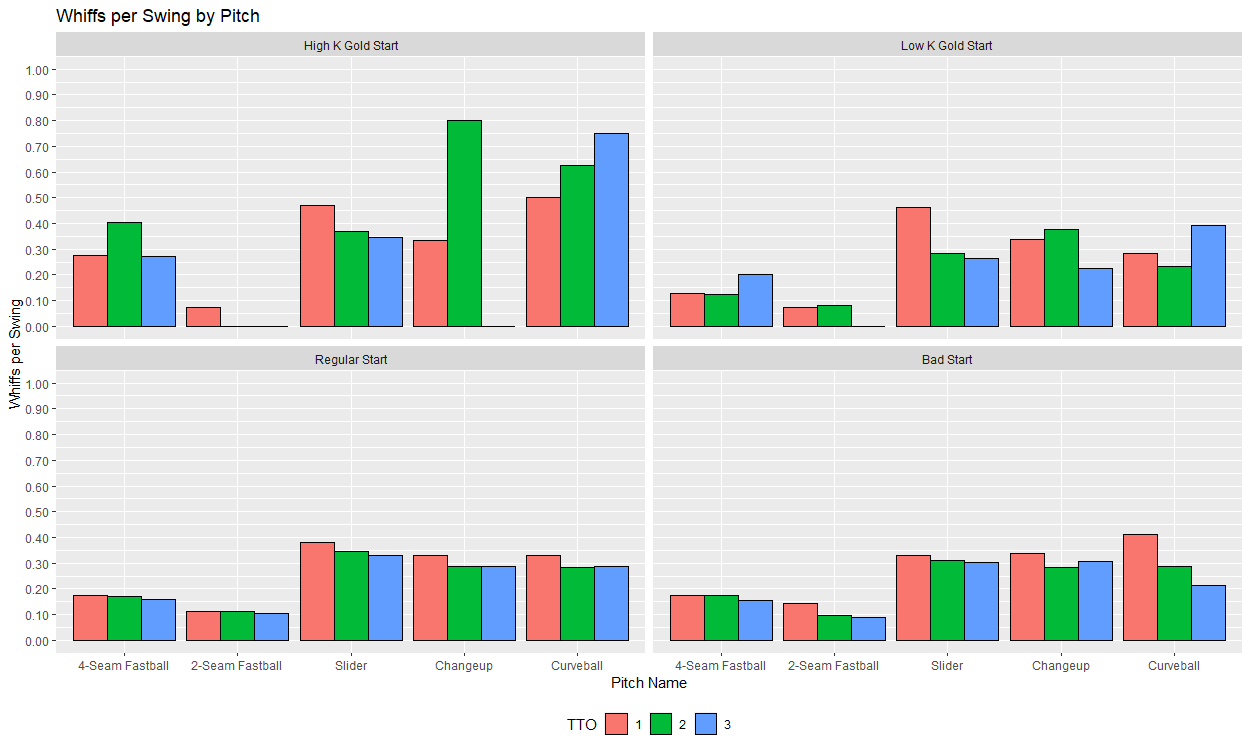


Using this information, is it possible to distinguish whether a pitcher has “it” in their first two trips through the order? If so, does that trend continue through the 27th batter?

In the High K Gold Standard starts, there was a small bump in velocity the second time through the order for 4-Seam Fastballs and Sliders. High K Gold Standard 4-Seam Fastballs had another bump for the third time through the order. The only other pitches that had a bump the second time through the order was the Changeup for both Low K Gold Standard Starts and Bad Starts.



The High K Gold Standard Starts again stand out for the increases as the game progresses. 4-Seam Fastballs, Changeups, and Curveballs all had significant increases the second time through the order. The only other pitch to have a greater Whiffs per Swing value the second time through the lineup is the Changeup for Low K Gold Standard starts.

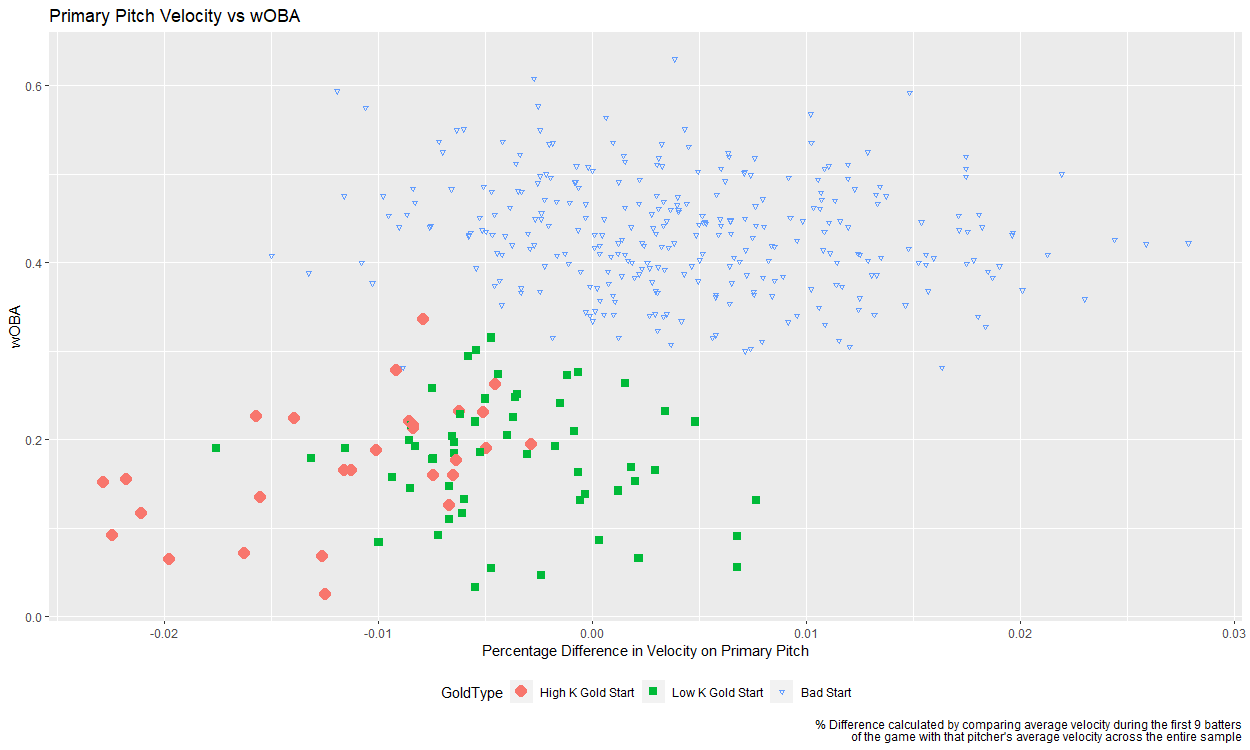


High K Gold Standard starts are beginning to separate themselves from the pack. Those pitchers are seeing an increase in velocity and whiffs per swing as the game goes on. Using this information, are there any indicators as the game progresses that separate these starts further? By looking at the correlations (relationship strength) of the metrics used to cluster these starts, a story begins to take shape.

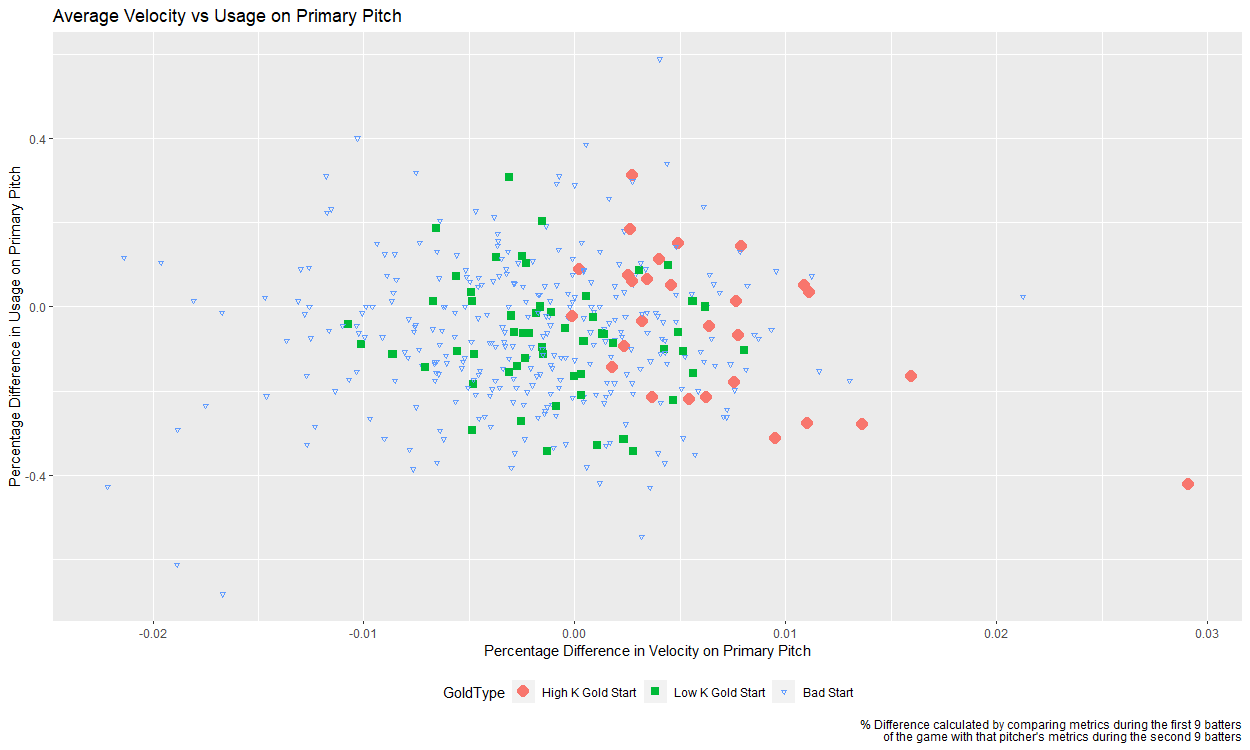
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable | Independent Variable | High K Gold Starts | Low K Gold Starts | Bad Starts |
| wOBA | Average Velocity (First TTO) | 0.5153 | -0.1387 | -0.1547 |
| Primary Pitch Usage (Second TTO) | Average Velocity (Second TTO) | -0.5804 | -0.1437 | 0.0394 |
| Whiffs per Swing (Second TTO) | Average X Movement (First TTO) | 0.5123 | 0.0086 | -0.0291 |

All variables used other than wOBA are the delta comparing to a previous baseline. Variables from the first time through the order were compared with the sample average. Variables from the second time through the order are then compared to the first time through the order.

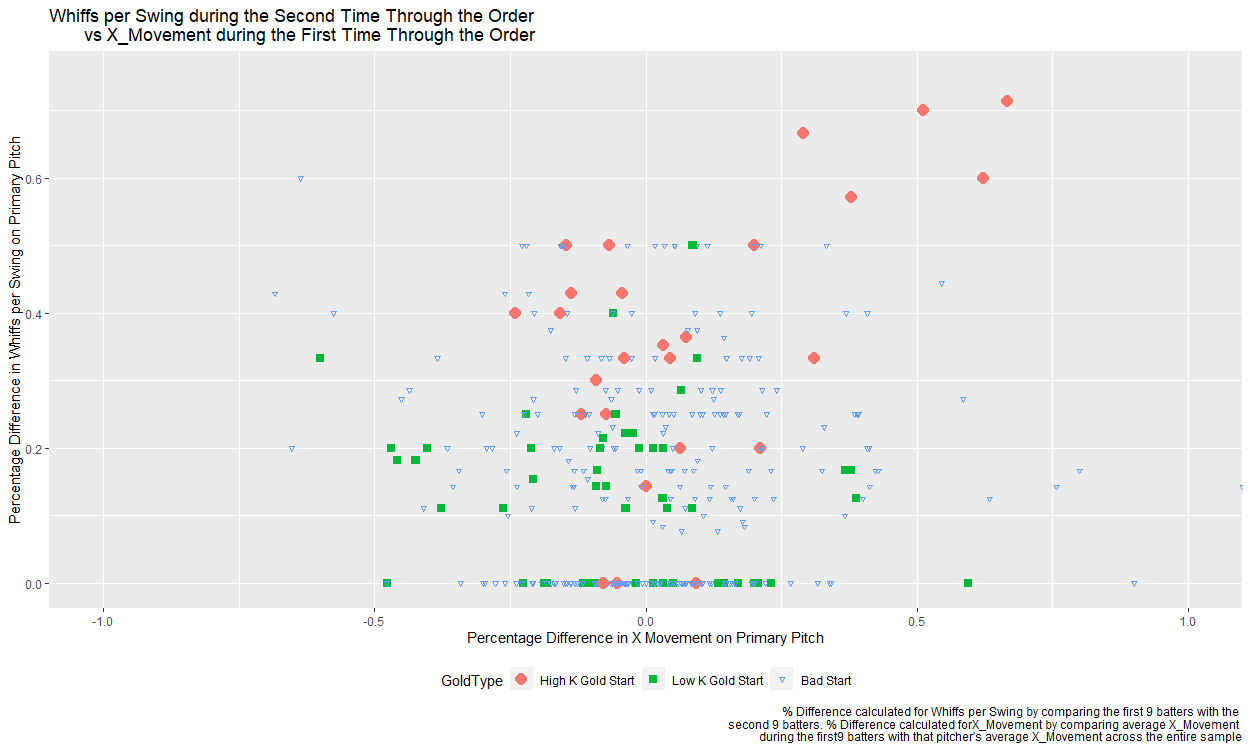
The graph below plots the relationship between a pitcher’s average velocity on their primary pitch compared to the wOBA allowed in that game. The simple meaning of the graph is that in the High Strikeout Gold Standard starts, pitchers are throwing slightly below their normal velocity in the first 9 batters, resulting in a decrease in wOBA. This is supported by graphs above, indicating that the starter is getting stronger as the game progresses.



Again, velocity is a significant variable. This time, the graph is showing that the larger the increase in velocity from the second 9 batters compared to the first 9 batters, the less the starter will use that pitch. Most pitchers are fastball first, meaning that these starts have the starter throwing more offspeed pitches the second time they see a batter. Is there a reason why and what effect does that have on performance?



We might have our answer. This graph is showing that if a starter is getting more break on their primary pitch over the first 9 batters, they will see a subsequent increase in whiffs per swing during the second 9 batters. Primary pitches (mainly fastballs) having more break than normal could make it more difficult for the hitter to recognize the pitch. This increased confusion results in more whiffs per swing as the game progresses.



At surface level, a High Strikeout Gold Standard start involves pitchers working deep into the game. They start the game with more break on their primary pitch than normal, combined with an average velocity slightly less than average. As that velocity trends upward over the course of the start, the usage decreases. The change in both gameplan and pitch metrics results in batters having more trouble as the game progresses.

In conclusion, High Strikeout Gold Standard starts can be identified as when a starter has “it”. That “it” factor shows that these starters improve the deeper into the game they work. They have a feel for their primary pitch from the beginning of the game, resulting in more confusion for the batters. Future work building off of this study could involve building predictive models to help decide whether to leave a starter in after that 18th batter. Linear trends began to emerge from variables that describe a pitcher’s performance in game. Incorporating other variables that deal with the strength of opponent could further refine decision making. Deeper analysis into “bad starts” could also prove beneficial as similar trends may begin to emerge.