Michael Pallante
Task 2: Option B
Initial_Findings_Report

After performing an Exploratory Data Analysis on Pitchers 1, 2, and 3, we are able to reveal some initial findings that provide us with insight on the research questions of this study:

- Why did Pitcher 1 struggle after the all-star break?
- Why did Pitcher 2 improve after the all-star break?
- Why did Pitcher 3 improve after the all-star break?

While I offered analysis for each variable within each pitcher notebook, below, I will summarize what I believe to be the key initial findings for each individual pitcher, and then offer conclusions and recommendations based on that.

Pitcher 1 Analysis

- Pitcher 1 had his highest success rate (called strike percentage) with his Four-Seam Fastball/Slider/Curveball combination before the all-star break. His decrease in percent change in usage of his Slider and even his Four-Seam Fastball are not recommended. He should place more emphasis on throwing these pitches as opposed to his Changeup and Cutter, which have both been rather ineffective pitches for him the entire season, according to his low success rate with both pitches. He was wise in increasing in Curveball usage though, as this was his most effective pitch before the all-star break.
- However, it is important to note that even his most effective pitches dropped in effectiveness after the all-star break, which suggests that there is more to his story than the above. There were no physical attributes of his pitching that were of major concern, although we did see some fairly significant fluctuation with his horizontal pitch movement (pfx_x). His Curveball and Slider movement actually seemed to improve, while his other pitches did not.
- There is not a drastic difference in terms of which catcher is catching for Pitcher 1. He has a slightly higher called strike percentage with Catcher 6, than with Catcher 1. There is some pretty significant fluctuation with his called strike percentage due to umps though. He earned over 30+% better called strike percentage with some umps than he did with others, perhaps suggesting he could have some bad luck there.

Pitcher 2 Analysis

- Pitcher 2 had his highest success rate with his Four-Seam Fastball and breaking ball combination of his Slider and Curveball before the all-star break. He increased his usage of his Four-Seam Fastball and Slider, which is justified by the fact that his success rate with both after the all-star break went up, especially with the Four-Seam Fastball where he saw a near 6% increase in called strike percentage. Pitcher 2 also saw an increase in his Curveball success rate, despite using less of it after the all-star break. Pitcher 2 also upped his usage of his Changeup after the all-star break, which is not recommended given the drop in his success rate in the pitch after the all-star break.
- One item that stood out among the physical attributes of Pitcher 2's pitching was the medium increase in his horizontal release for all of his pitches, which suggests that perhaps he is threw from a better arm slot after the all-star break.
- Another item of note was that Pitcher 2 had a large increase in his Changeup spin rate, which is interesting considering he did not throw it very successfully after the all-star break. Perhaps this suggests that we could maximize this pitch more by ensuring he uses this pitch in the proper situations, and also in the proper locations in the strike zone.

Pitcher 3 Analysis

Michael Pallante
Task 2: Option B
Initial_Findings_Report

-

Certainly, the overall analysis of these three pitchers can drastically improve with the addition of more data sources, specifically result data so we can analyze what baseball game events happened after each pitch. With data on hits, strikeouts, walks, and just game or play-by-play data in general, we are able to offer a much bigger picture on evidence that would support why each pitcher either struggled or improved over the course of the season. One statistic that immediately comes to my mind is the use of FIP to analyze a pitcher at a deeper level than just his ERA, which sometimes does not tell the full story. If afforded with the luxury of home runs allowed, walks, and strikeouts in the dataset, we can calculate FIP and see just how well the pitcher is truly pitching, due to FIP's independence of balls in play and fielding errors. For example, Pitcher 1's 4.80 ERA after the all-star break could be elevated due to factors out of his control and if his FIP was lower (let's say his FIP is 3.35), we would know that he might have some bad luck and that if he continues pitching at that level, we can feel confident that eventually his ERA will come down and meet closer to his FIP.

From a modeling perspective, the addition of more result-type of variables will also give us more conclusive variables to build a model on. Right now, given the current dataset, we have a hard time selecting which of our variables is of most importance in predicting pitcher success. I'd likely select to build a model based on predicting called strikes, as this could suggest that with a rise in called strikes, perhaps the pitcher can get more outs.

Also, a key note that I mentioned in my data cleaning notebook is that due to the fact that results-type of data is unavailable to us, situational variables are of less use to us than they could be, including ball-strike count and game state (or the variables I created to show what I'd use in a model including out count, baserunner count, what bases are occupied). We'd be highly interested in using those variables in our model even if we simply just knew strikeout and walk data, where we can offer pitchers suggestions on specific ways they should pitch in situational baseball and what pitch types and pitch locations they should lean on in these events. We are unable to do these things with the data in its current state, but we have the potential to do so much more if we were able to obtain results-type data.