Scott Matsubara
2/22/24

Tigers Questionnaire Answers

1. This past summer I had the chance to work as a Baseball Research Analyst Fellow for the Minnesota Twins. I was closely involved with the sports science department where I led a project that leveraged data modeling to discern which hitter biomechanical features are most influenced by pitch context such as pitch location, type, velocity, and same handedness. This exploration aimed to determine whether variations in these biomechanical features were due to hitter-side adjustments or merely a response to the type of pitch encountered. For instance, an interesting finding was that a player's trail hinge angle could be predicted by the pitch type and location. Through this, we can infer that observing a greater trail hinge angle during a week that a player saw more changeups would not be a cause for concern. While I did not get a chance to interact directly with coaches and players, I gained invaluable exposure to the professional baseball environment and engaged with professionals from various departments such as Player Development, Scouting, and Core Research. I also got a chance to shadow scouts at minor league games and even be in the draft room during the MLB draft. Finally, my responsibilities also involved pitch type correction for pitcher movement plots and operating the Kinatrax motion capture system.

2. I have extensive experience with creating data visualizations, specifically using the matplotlib and seaborn packages in Python. Examples of this include simple distribution plots, box plots, and bar graphs for data exploration as well as scatter plots and partial dependency plots for model analysis. While working for the Twins, I built a Shiny app that shows the predicted value of a hitting biomechanic feature based on the pitch location and pitch type. This was displayed on a plot with a strike zone and colored based on the predicted value of that hitting feature. The app allowed the user to pick from different hitting features and different pitch types. Recently I have created many Streamlit applications in addition to the one I have submitted for this questionnaire. For my Capstone project called Hedwig.AI, I created the UI for the project that allows the user to type in an email and see how different user personas would respond to that email. I also built an application to help with my Dungeons and Dragons sessions that manages players' undead or armies by rolling attack rolls, rolling saving throws, applying buffs, and keeping track of health.

3. I have done all sorts of predictive modeling such as: using linear models to predict housing in Boston, using NLP and neural networks to predict song genres based on song lyrics, and using clustering to recommend Spotify songs to users based on their listening history. Another project I worked on with the Twins aimed to discover if any hitter

biomechanical features could predict a players exit velocity. This would be used by the player development department to help players make adjustments that could improve their performance. For this project, I used Python for my data cleaning, analysis, and modeling. I used linear models and random forests to do my analysis.. This data had over 200 variables that I had to do feature selection on. I did this by asking experts in the department about which variables could be adjusted relatively easily for players. I also chose not to use any variables that could be proxy variables for exit velocity such as stride extension. Then I filtered the dataset to only have observations that are relevant for this analysis such as 'full swings' that resulted in over 90th percentile exit velocity for an individual player. In conclusion, I used a linear regression model for my final model and was able to find very small relationships between these hitter variables and exit velocity. In the future, I would like to use bat speed as a response variable instead since it is mostly dependent on the players' biomechanics and not a variety of other factors like exit velocity.

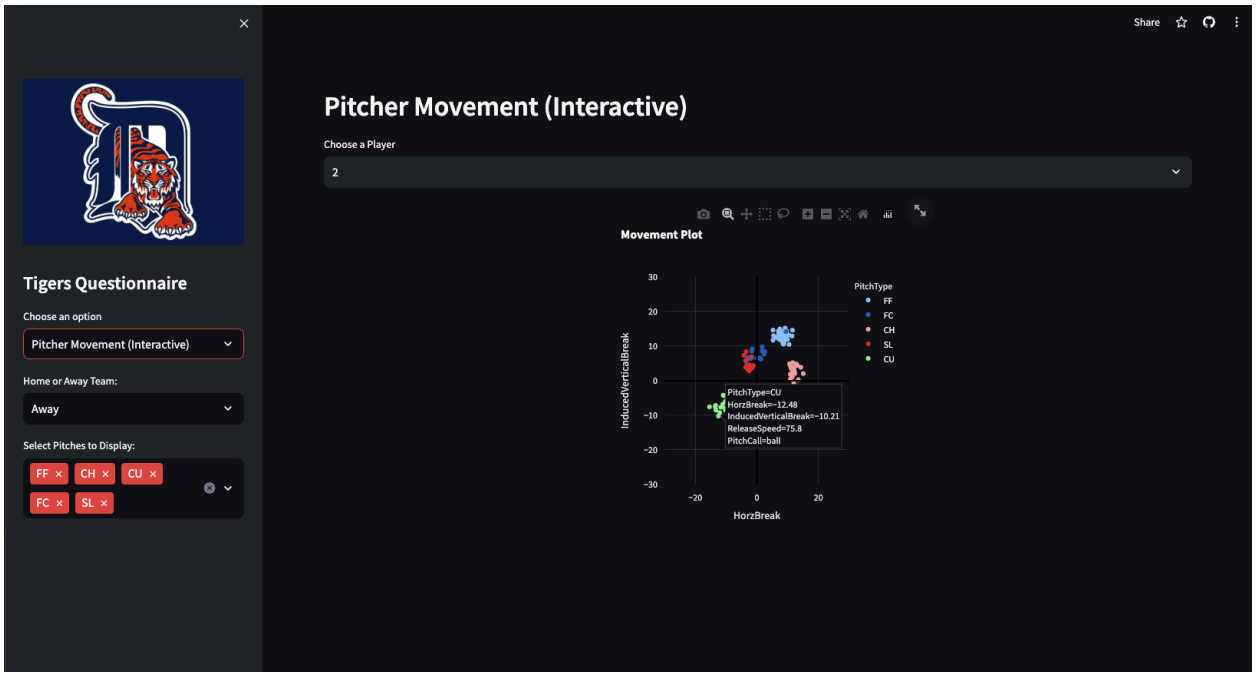
4. Now that a full season of the pitch clock has gone by, it would be great to check on how this has affected pitchers. This could be from a performance perspective and from a health/biomechanics perspective. Is less time between pitches causing a breakdown of their mechanics which could ultimately lead to injury? Additionally, which players have done worse or better under these conditions. Is there any way we can help the pitchers who did worse improve in the upcoming season? This could be done by using a simple two sample t-test to compare performance in previous seasons to performance in the 2023 season. The players with significant differences in either performance or biomechanics would be flagged to look at deeper. For these players, we would first look at if the pace between the seasons is significantly different. For example, if a player has a 5 second difference in pace between the two seasons, they would be worth looking at compared to a player with only a 1 second difference in pace. Then we could then look at appearances where they performed well to see if they did anything differently and to help them try to replicate that. We should also update any player performance models to more heavily value 2023 season data.

5. The order of preference for players would be player B, A, C. I believe that player B is the best option because he has a good balance of the hitting features. While his bat speed is less than player A, he squares up the ball more consistently and thus produces a higher exit velocity. We could help this player improve his launch angle, which should generate more home runs during the season. If we could help player A improve his squared up percentage, he would be more valuable but I believe that tweaking launch angle for player B would be easier. Vertical Attack Angle is less important for evaluation because it has more to do with pitcher matchups than overall performance. Player C is the least desirable because there is likely no improvement that can be made for this player. His

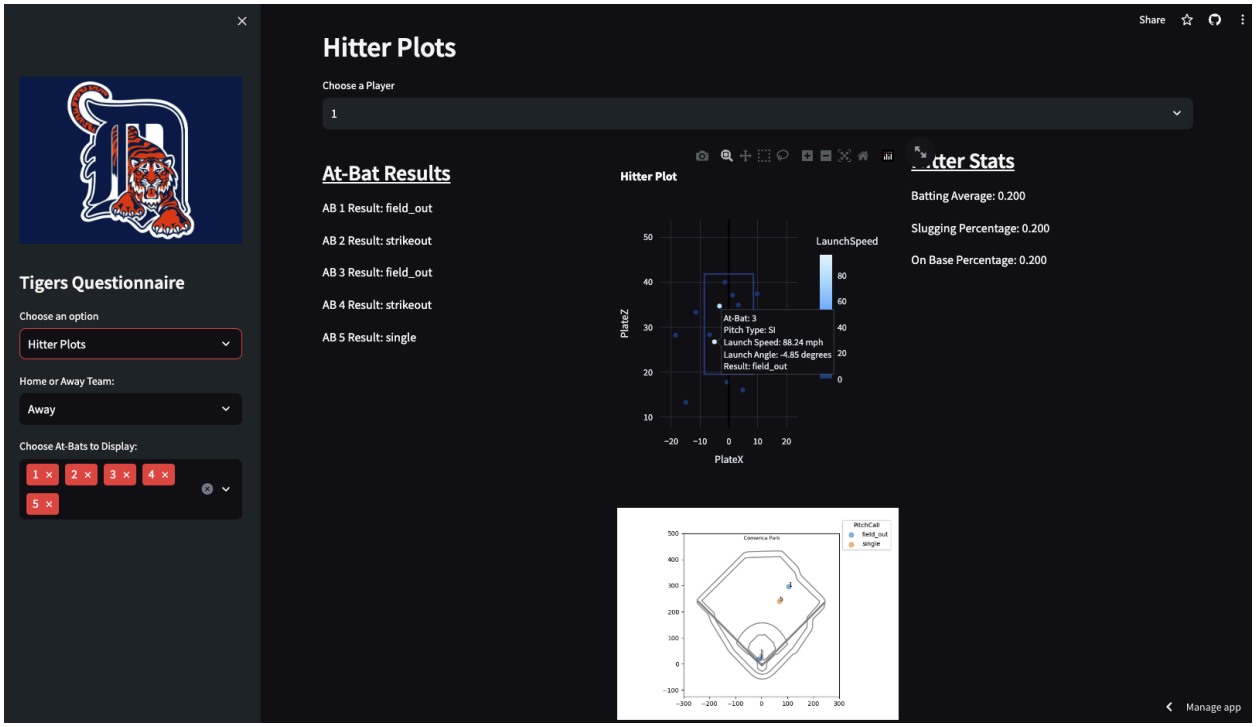
squared up percentage is higher than both players, yet given his slow bat seed his exit velocity is the same.

6. I created an app that shows visualizations of the game from a pitcher and hitter level. It can be found at the following link:
   https://baseballvisualizationsapp-5hmndxerevqedwchnxpaed.streamlit.app/
   My Github link for the code can also be found here:
   https://github.com/smatsubara15/Baseball_Visualizations_Streamlit/tree/main

The first page shows pitcher movement plots for a given game. Here you can select from any player to look at by choosing the home or away team and by choosing the player from the dropdown menu. Then you can choose which pitches to display on the plot. When you hover over a pitch in the plot, it gives you more information about the pitch such as the release speed and the result of the pitch. This can be useful for scouting against opposing teams to see how pitches move. It would also be good to see how our pitchers pitches moved in a given game.



The next page shows plots from the hitter standpoint. Again you can select any hitter front the game and you can choose which at-bat to display. The left side shows a summary of the outcome of each at-bat and the right shows their stats from the game. The first plot show the location of all pitches thrown to the hitter and the color gradient of the points is based on the exit velocity. If you hover over a pitch you can see the pitch type, launch speed, launch angle, and result of the pitch. This is useful for seeing how pitchers are pitching to hitters and what pitches hitters are swinging at. The bottom plot show the spray chart of a player during the game. Points near home plate indicate a ground ball.



Finally the last page shows all the above information but from a team standpoint. Either for the home or away team.