Name: Siddhant Thakur

Detroit Tigers Baseball Analytics Questionnaire

Return upload your responses to box using this link. If you have any issues or questions, please reach out to analyticsstaffing@tigers.com. Please adhere to the word limits and include your full name in all filenames.

1. **Do you have any experience working with a sports team before? If so, please elaborate on your experiences, especially any interactions with coaches or players. [250 words]**
    - While I have not worked with a sports team in an official capacity, I did work with the Texas A&M Football team on a semester-long project where we developed a computer vision model and an analytical platform to assess player routes. This technology was designed to assist coaches and players in refining their in-game strategies and practice routines. My primary interactions were with the Director of Sports Science & Analytics at Texas A&M Football. We met biweekly to discuss the technology stack for the platform and to delve into the nuances of the coach's playbooks and player routes. These interactions were highly productive, given the Director's expertise in both the technical and domain aspects of the project. The project's culmination presented a unique challenge. Explaining the capabilities and limitations of the computer vision model to the coaches and players, who lacked a technical background, was not straightforward. However, the analytical platform proved to be a bridge in this communication gap. Its intuitive and user-friendly design made it an effective tool for conveying the insights derived from the vision model to the coaches and players. This experience not only honed my technical skills but also enhanced my ability to translate complex technical concepts into understandable terms for a non-technical audience.

2. **Please describe your experience with creating data visualizations and developing Shiny apps. Include any examples if you have them (they do not need to be baseball related). [250 words]**
    - During my semester abroad at University of Wisconsin-Madison, I developed a Shiny web app to provide a visualization aspect to our project which modeled aphid diversity using crop and weather data for the Midwestern states. Since the statistical model was a little complex for everyone to understand, I used Shiny to plot crop distribution as well as the predicted aphid diversity across these states. Apart from developing webapps in R, I have also worked on web apps developed using Flask and Dash in Python. From what I have experienced, Shiny is actually more intuitive and easier to work with than Flask, where you can set up the UI and Server with the same code.
    - Aphids - https://github.com/sidthakur08/Aphids

3. **Please give a brief overview of your experience with predictive modeling as well as a description of an analytical project that you have completed. Include the purpose of the project, the methods or models chosen, any additional methods or models that were tested, and what tools or programming languages you used to complete the project. This project does not need to relate to baseball. [250 words]**

- My experience with predictive modeling and analytical projects is exemplified in a project I completed titled "Analysis of Dew Point Affected Pitches." The purpose of this project was to estimate whether a given pitch was affected by Dew point or not, without having the dew point or temperature data.
- In the data exploration phase, I analyzed 9,889 pitches thrown by 37 pitchers, focusing on variables like horizontal break, release side, and horizontal approach angle, among others. I used Python for data preprocessing, including handling NaN values and outliers, and for correlation checks among variables.
- The methodology was centered around a mixed-effect model, incorporating fixed effects like release characteristics, approach angle, and spin rate, and random effects such as pitchers and their pitch types. The model hypothesized that if we can account for all the effects on induced vertical break and horizontal break using the fixed and random effects, any deviation from prediction should indicate an effect of any external factor not accounted for, one of which could be dew point. The residuals from these models were then used to infer the likelihood of external factors, such as dew point, affecting the pitches.
- For implementation, I developed a Python class with methods for scaling data, building models, and computing residuals. These residuals were assumed to be coming from a normal distribution, which then gave us the probability that the pitch was affected by dew point, by effectively asking the question “How unlikely it is to see this value of residual in the model”. The model was applied separately for induced vertical and horizontal break, and the average probabilities from both models were used for final assessment.
- To validate the model, I checked the normality of residuals through graphical methods, confirming the assumption of normal distribution in the mixed-effect model. This project was a comprehensive exercise in applying predictive modeling techniques to a unique and challenging problem in sports analytics.

4. **What is a project that you believe would add substantial value to a baseball team? Please describe the project and provide an overview of how you would complete it. [250 words]**
    - A project that would significantly benefit a baseball team is leveraging biomechanical data for pitching analysis. The project can be divided into four key sub-projects:
        - Fastball Velocity Analysis: This involves constructing models to understand how pitchers generate their fastball velocity. We'll identify factors leading to velocity loss and suggest improvement strategies, crucial for developing young pitchers. Collaboration with pitching coaches will be vital to discern variables that are modifiable versus those that are inherent.
        - Comprehensive Velocity and Movement Modeling: Extending the analysis to all pitch types, this sub-project will predict velocity and movement, estimating performance with currently unused pitches. This could reveal undervalued pitchers who might benefit from adjusting their pitch selection.
        - Fatigue Metrics Development: We plan to develop a fatigue metric correlated with biomechanical data, identifying contributing factors. This will include analyzing individual pitch types to find specific movements or pitches that lead to increased fatigue.
        - Injury Prevention Model: The final aspect focuses on developing a model to understand and prevent injuries, particularly in the elbow

and shoulder. By pinpointing key risk factors, we can strategize to mitigate injury risks.

- The execution of this project will involve close collaboration with coaches, integrating their insights with our data-driven models. We'll employ advanced statistical and machine learning techniques to analyze the biomechanical data, ensuring our models are both accurate and actionable.

5. For Question 5, please refer to the table below. [250 words]
    - The following pitches all come from the same pitcher. Rank them in regards to quality ("Stuff") from 1 to 3 (1 = Best, 3 = Worst) and explain your reasoning. Assume a consistent release point across all three.
        - To rank the pitches based on "Stuff" using the metrics provided in the table and referencing the principles from Driveline Baseball's article on quantifying pitches with pitch models, we'll consider factors such as velocity, spin rate, spin direction, spin efficiency, and movement.
            - Changeup (CH): This pitch ranks the highest in quality. The velocity of 84.5 mph provides a sufficient difference from the pitcher's fastball velocity. It also has the highest spin efficiency at 86% and substantial horizontal (13.3 inches) and vertical movement (9.1 inches), which should make it difficult for hitters to track and contact effectively.
            - Fastball (FB): The fastball comes in second. It has the highest velocity (92.7 mph) though could use an increase in velocity by roughly 2-3 mph, though it does have a good spin rate (2145 rpm), and an 80% spin efficiency. The pitch has strong vertical movement (13.4 inches), which is a key indicator of a quality fastball. Its horizontal movement (8.5 inches) also suggests it could have a tailing action to one side, potentially increasing its effectiveness.
            - Slider (SL): The slider ranks third primarily due to its low spin efficiency (38%). While it has the highest spin rate (2675 rpm), this does not translate effectively into movement, with negative horizontal movement (-8.5 inches) and minimal vertical movement (-1.6 inches), indicating that the pitch may not have the desired break, making it easier for hitters to recognize and hit. Also, the velocity drop compared to the fastball is 11 mph, which is too high, and hence the slider needs to be thrown harder.
    - What adjustments would you recommend, whether to individual pitches or the arsenal as a whole? What existing pitch would you adjust?
        - Improving the Slider: Work should be done to decrease the spin efficiency of the slider, so that it acts like a gyro-slider. This might involve changing the grip or the pitcher's wrist action, making the pitch more deceptive. Another option could be converting this slider into a sweeper and gain more horizontal movement.
        - Overall Arsenal: The arsenal could benefit from adding a pitch with a different speed or movement profile, like a curveball with a high vertical drop or a cutter with late horizontal movement. This would give hitters a different look and could complement the existing pitches.
        - Fastball Tuning: Small adjustments to increase the spin rate or tweak the spin axis might help to make the fastball more dominant,

especially if it can be thrown with a higher perceived velocity or with more 'rise' to challenge hitters at the top of the strike zone.
- These recommendations aim to maximize the pitcher's effectiveness by enhancing the deception and movement of their pitches, making them harder to hit and potentially improving their strikeout rates.

Pitch Type	Velocity	Spin Rate	Observed Spin Dir.	Spin Eff.	Horizontal Movement	Vertical Movement
FB	92.7	2145	1:25 (228°)	80%	8.5	13.4
SL	81.4	2675	8:47 (84°)	38%	-8.5	-1.6
CH	84.5	1760	1:43 (230°)	86%	13.3	9.1

6. For Question 6, please refer to the table below. [150 words]
    - Players A, B, and C are available to acquire (for this exercise assume positions are inconsequential, they are all the same handedness, that they are the same age and of similar cost). Please rank them from the player you are most interested in, to least interested in. Explain your reasoning.
        - Considering the batting metrics provided for Players A, B, and C, we can evaluate their performance based on Exit Velocity (EV), Launch Angle (LA), Swing%, Z-Contact% (contact rate on pitches within the strike zone), O-Swing% (swing rate on pitches outside the strike zone), and O-Contact% (contact rate on pitches outside the strike zone).
            - Player A: This player is most desirable due to a high contact rate both inside (95%) and outside (90%) the strike zone, indicating good plate discipline and contact skills. Although the EV is slightly lower, the combination of high contact rates with decent swing decisions (40% Swing%, 30% O-Swing%) suggests a consistent and disciplined hitter.
            - Player B: The second preference goes to Player B, who has a slightly higher EV and LA, suggesting potential for more power. However, the lower Z-Contact% (85%) and O-Contact% (60%) compared to Player A might lead to more strikeouts or less consistent contact.
            - Player C: Despite having the highest EV, which suggests power potential, Player C ranks third due to a lower contact rate both inside (75%) and outside (55%) the strike zone, coupled with the highest swing rate (50%) and the highest O-Swing% (33%). This profile indicates a more aggressive approach that could result in more strikeouts and potentially less consistent on-base performance.
        - In summary, Player A's high contact rates and disciplined approach make them the most attractive, Player B offers a balance between power and contact, and Player C, while potentially the most powerful, carries a higher risk due to lower contact rates and less disciplined swing decisions.

	EV	LA	Swing%	Z-Contact %	O-Swing%	O-Contact %
Player A	88.5	12.5	40%	95%	30%	90%
Player B	89.0	13.0	45%	85%	25%	60%
Player C	93.0	12.0	50%	75%	33%	55%

7. Attached are pitch-level data from two games (AnalyticsQuestionnairePitchData.csv). Please produce self-explanatory graphs and tables to summarize the players' performance in those two games. Summarize your findings with bullet points. Include all code/workbooks.

```
In [165]: data.head(8)
```

Out[165]:

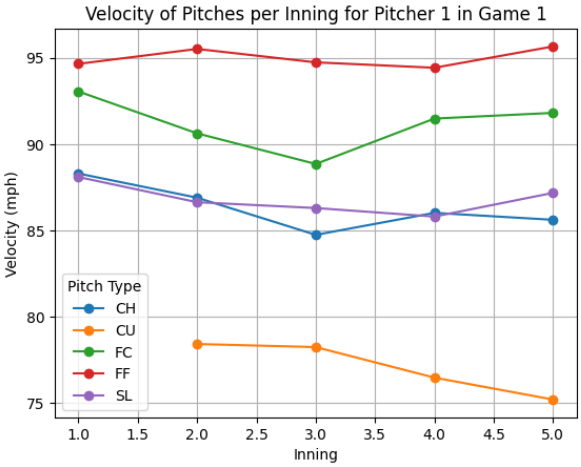
	PitchId	GamePk	PitcherHand	PitchCall	PitchType	BatterId	PitcherId	BatterSide	Inning	IsTop	Balls	Strikes	Outs	PostBalls
0	231	1	L	called_strike	CH	19	11	R	1	1	0	0	0	0
1	384	1	L	ball	CU	19	11	R	1	1	0	1	0	1
2	616	1	L	field_out	SL	19	11	R	1	1	1	1	0	1
3	503	1	L	ball	SI	24	11	L	1	1	0	0	1	1
4	322	1	L	swinging_strike	CU	24	11	L	1	1	1	0	1	1
5	502	1	L	single	SI	24	11	L	0	1	1	1	1	1
6	606	1	L	ball	CH	21	11	R	1	1	0	0	1	1
7	579	1	L	called_strike	CH	21	11	R	1	1	1	0	1	1

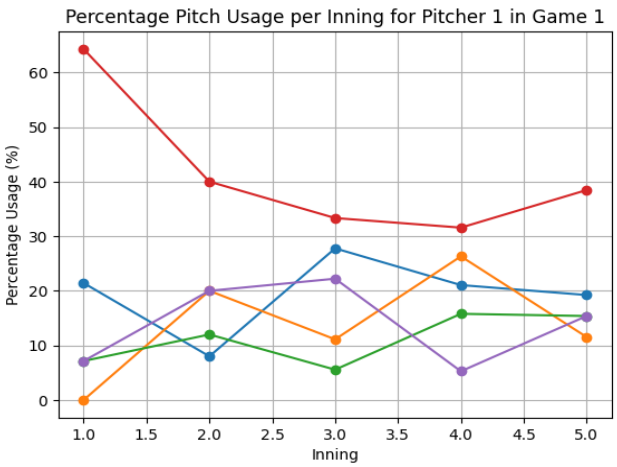
Description of the Data:

The given data is from 2 games, with 17 different pitchers throwing 620 pitches. For the pitch level we have release data, movement data, spin data and the trajectory of the pitch. To produce the graphs and game report, I am focusing on pitcher id 1 who pitched 198 pitches over 2 games, and I would be looking at each of those two games for that pitcher. Similar game reports can be drawn for the rest of pitchers, which due to time crunch I have not shown here in the document, but the self explanatory graphs can be seen in the jupyter notebook/code attached.

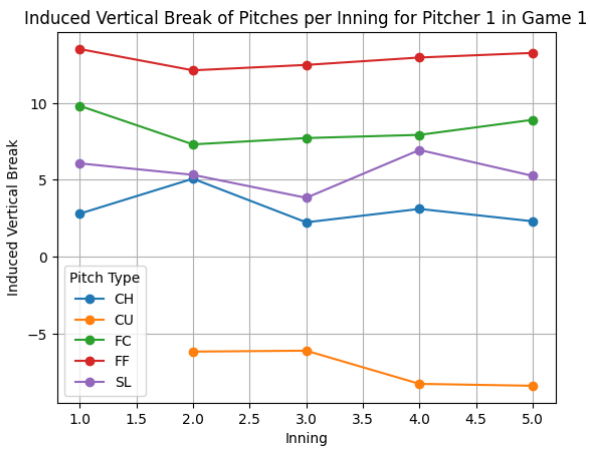
Pitcher ID 1 in Game 1:

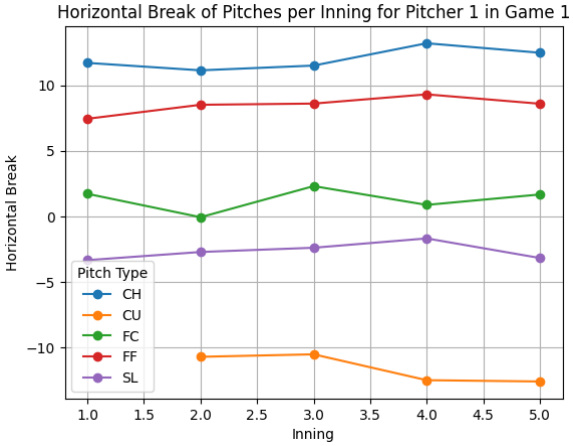
Here are the velocities of different pitch types, for the pitcher ID 1 in game 1 as the inning progressed.



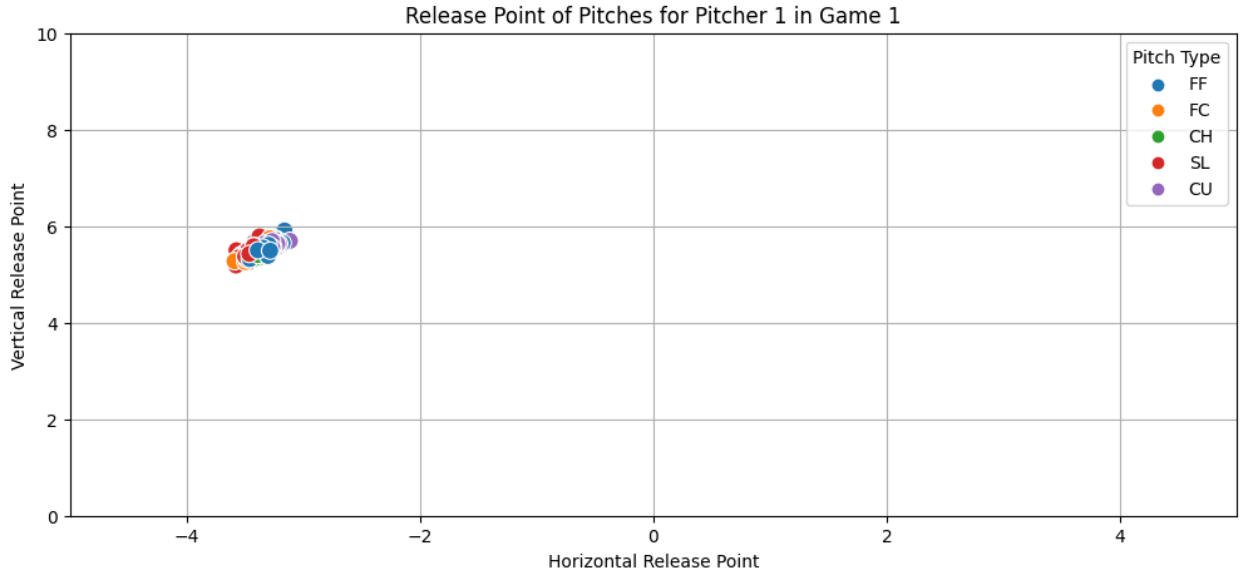


We see that this pitcher's fastball velocity remains consistent in this outing, while there is a dip in his Changeup and Cutter velocity. Though after the 3rd inning, we see a little bit of jump in the velocity. We also see that he didn't use a curveball in the first inning but the usage increased as the game went on, and the velocity greatly decreased by the 5th inning. We also see a decrease in Slider velocity as the game progresses.

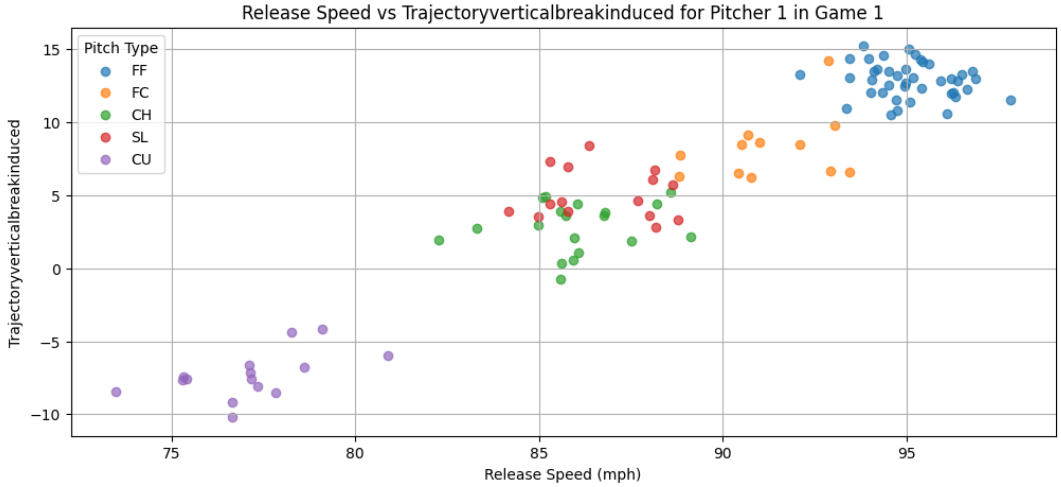


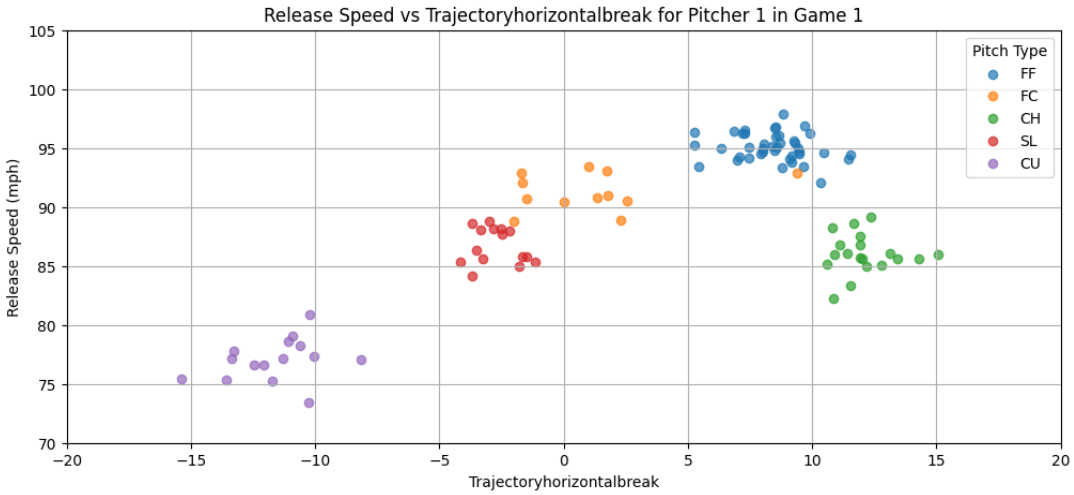


From the above graphs, we see that his Induced Vertical Break and Horizontal Break for all of the pitches were relatively consistent, though we did see a slight increase in horizontal break for both his changeup and fastball.

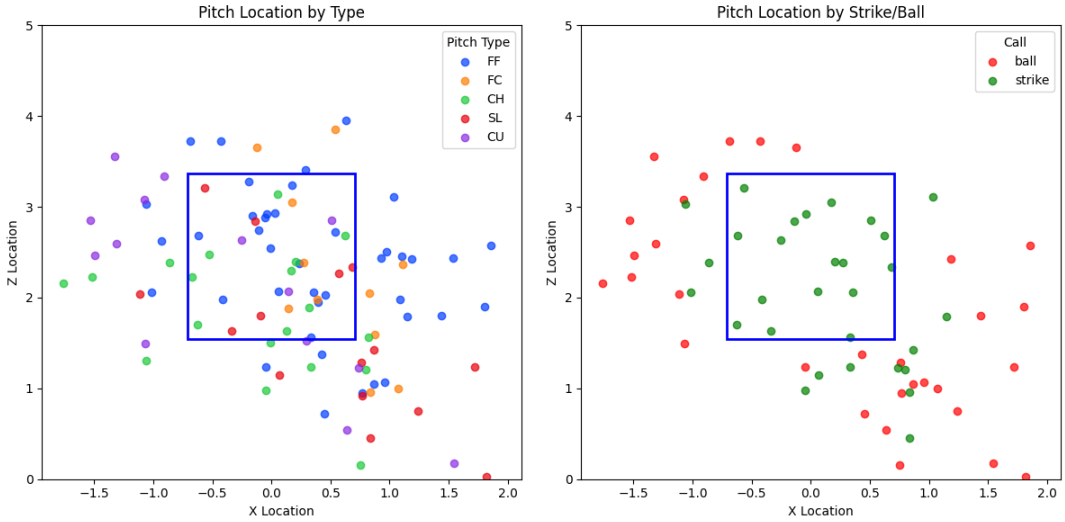


The release points for this pitcher for all his pitch types looks tightly clustered, so we then look at the movement of his pitches with respect to the velocity.

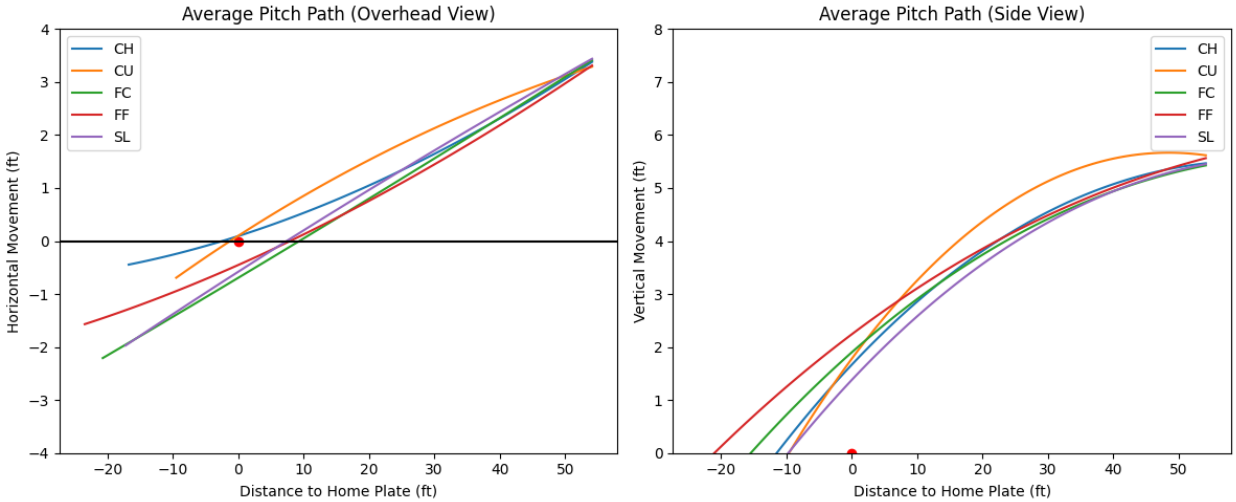




We see that his curveball has a very high vertical and horizontal break, making it a very good curveball. Though his slider has little to no movement in both vertical and horizontal directions. His fastball, though, has sufficiently high speed and vertical break. Looking at his changeup, we see that it is roughly 8mph less than his fastball with a lot of horizontal break. Looking at this outing, it seems he has a good fastball, changeup and curveball combo, but not a very good slider and cutter.



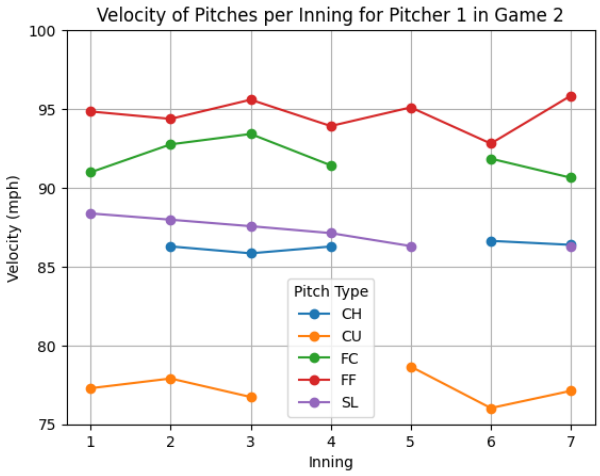
Looking at his strike zone and pitch types, it seems like he misses a lot of fastballs to the right and away of the strike zone (when looking from the pitcher's perspective). This could be due to the high horizontal break on his fastball discovered from the previous graph. His changeup, however, stays low and results in a strike outside the zone.

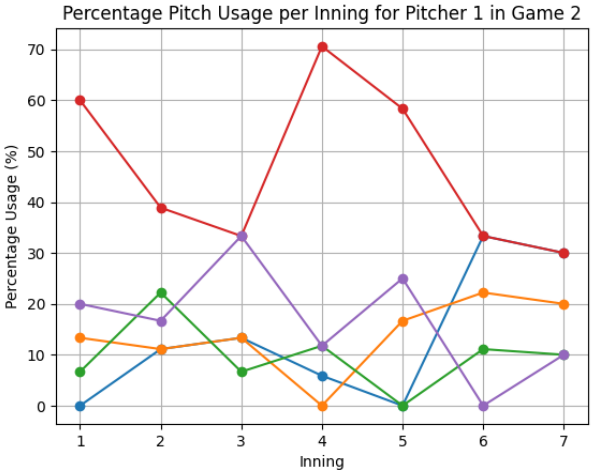


Finally, we look at his pitch trajectories arranged by his pitch types, and we see that his curveball has a beautiful dip, but has a different pitch path compared to the rest of his pitch types. Hence, we see that the other pitch types based on the pitch path break late in the delivery and closer to the home plate.

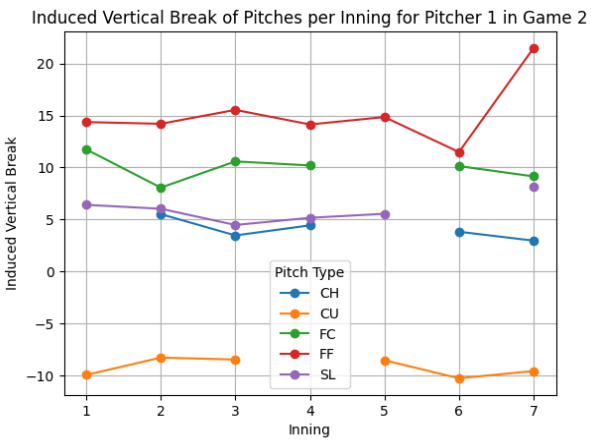
Pitcher ID 1 in Game 2:

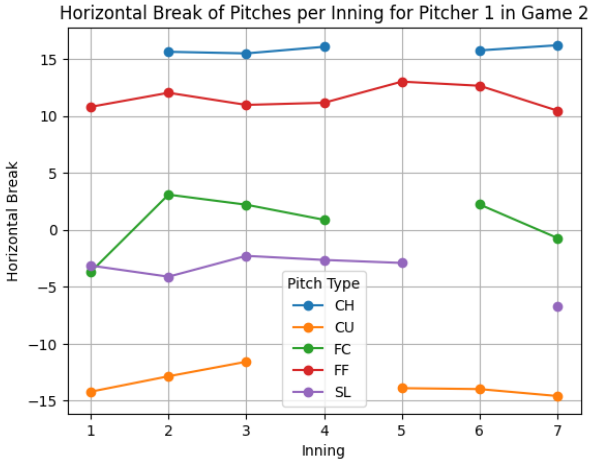
Now we look at the same pitcher ID 1, in a different game (game id 2) to see if this outing might be a little different.



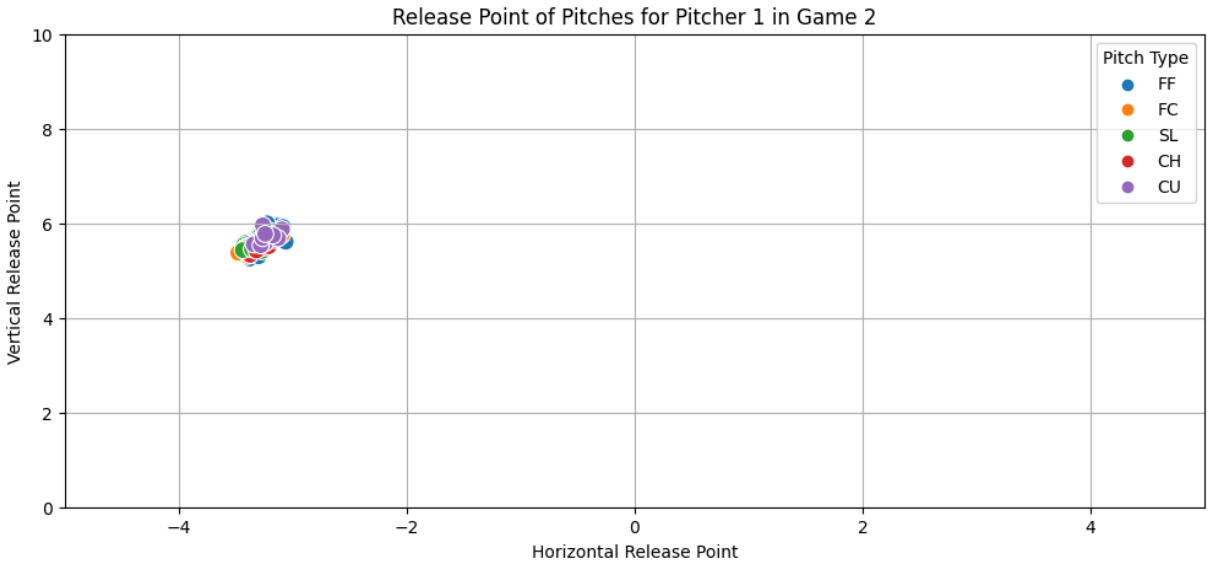


In the second game, we see more variance in fastball velocity compared to game 1 and a dip in his velocity in the sixth inning. Also, his slider velocity seems to be decreasing as the innings progressed, similar to game 1. His pitch usage is also different compared to game 1, where in this game he had a peak usage of fastballs in inning 4 and slowly diminishing by the sixth inning.

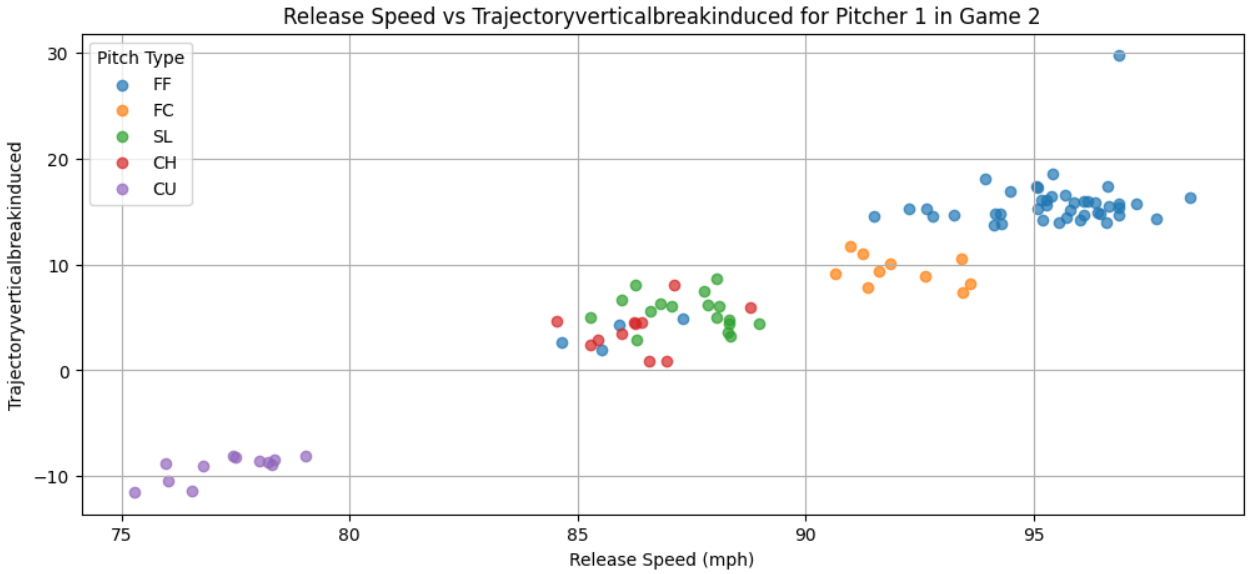


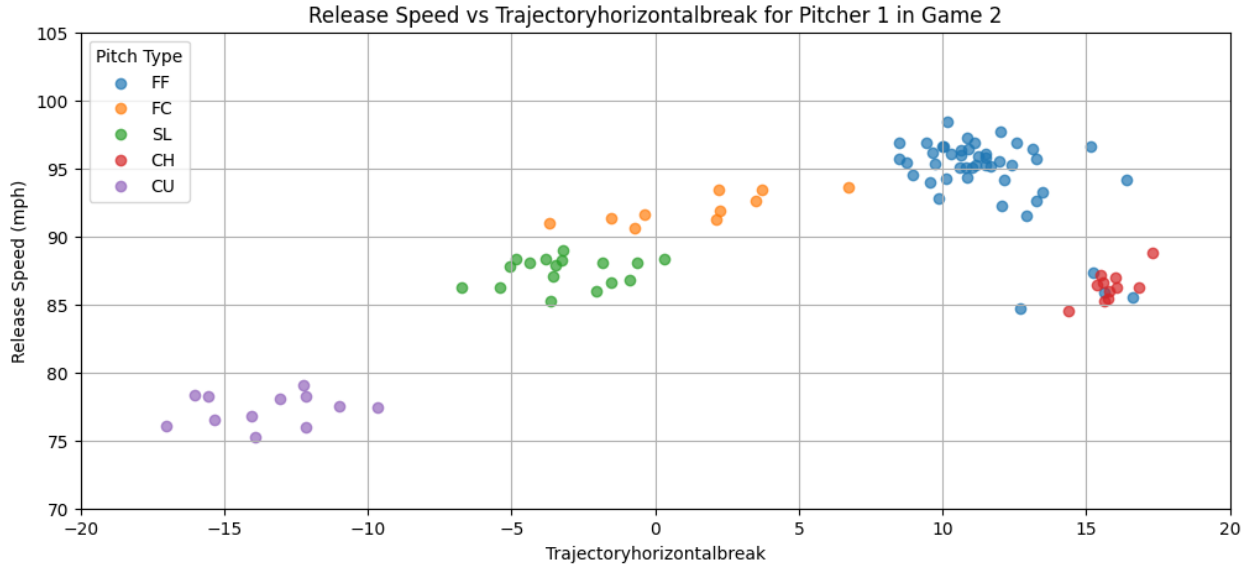


Through the inning both IVB and HB remain consistent, but we do see a slightly higher horizontal break for his fastballs compared to game 1 (He seems to be hovering around 12 inches of horizontal break in game 2, compared to around 7 inches in game 1)

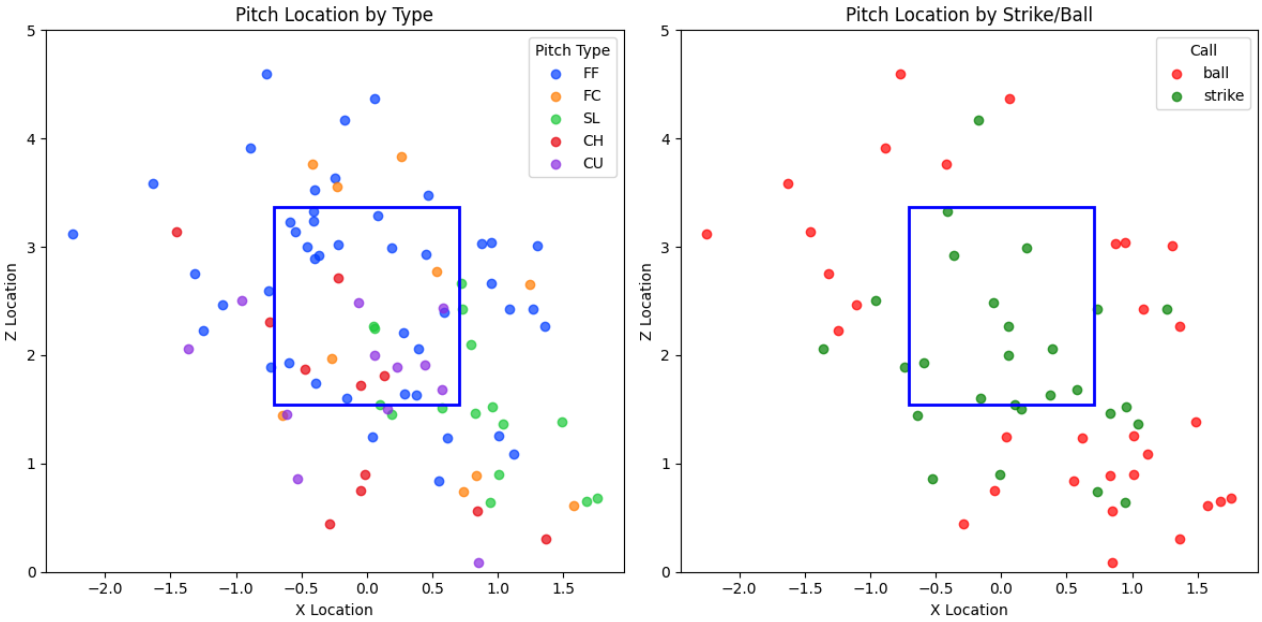


We see that even in this outing he has a consistent release point that has not changed since game 1.

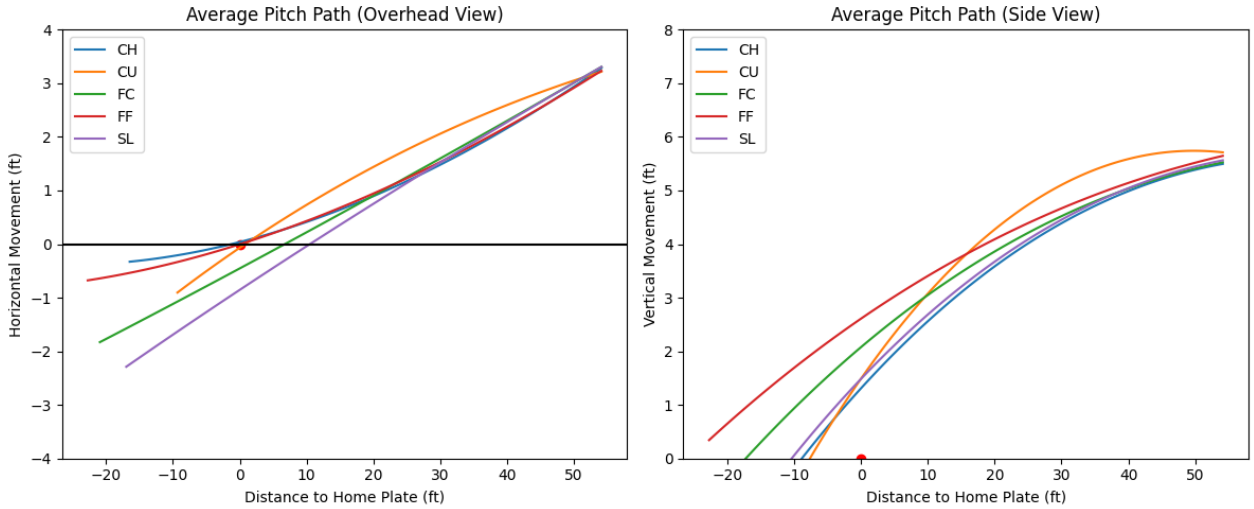




One big notable difference in this outing is that his fastball remains at the same speed, but achieves a slightly higher vertical break and horizontal break compared to his outing in game 1. The changeup also has a slightly higher horizontal break in game 2. Rest of the pitches remain consistent throughout the two games, with his sliders not having a lot of movement both vertically and horizontally.



In this outing, we see that his Fastball is comparatively more center focused when compared to game 1.



Another notable observation is that in this outing, it seems that his fastball and changeup follow a similar path when we look at it from overhead view. This has the potential to confuse the batter, as the plane might look similar, but the changeup's depth might get them to swing and miss.

For other pitchers, we can see a similar game report in the jupyter notebook attached with the submission.