Can You Stop Elly De La Cruz Jr. From Stealing Second Base?

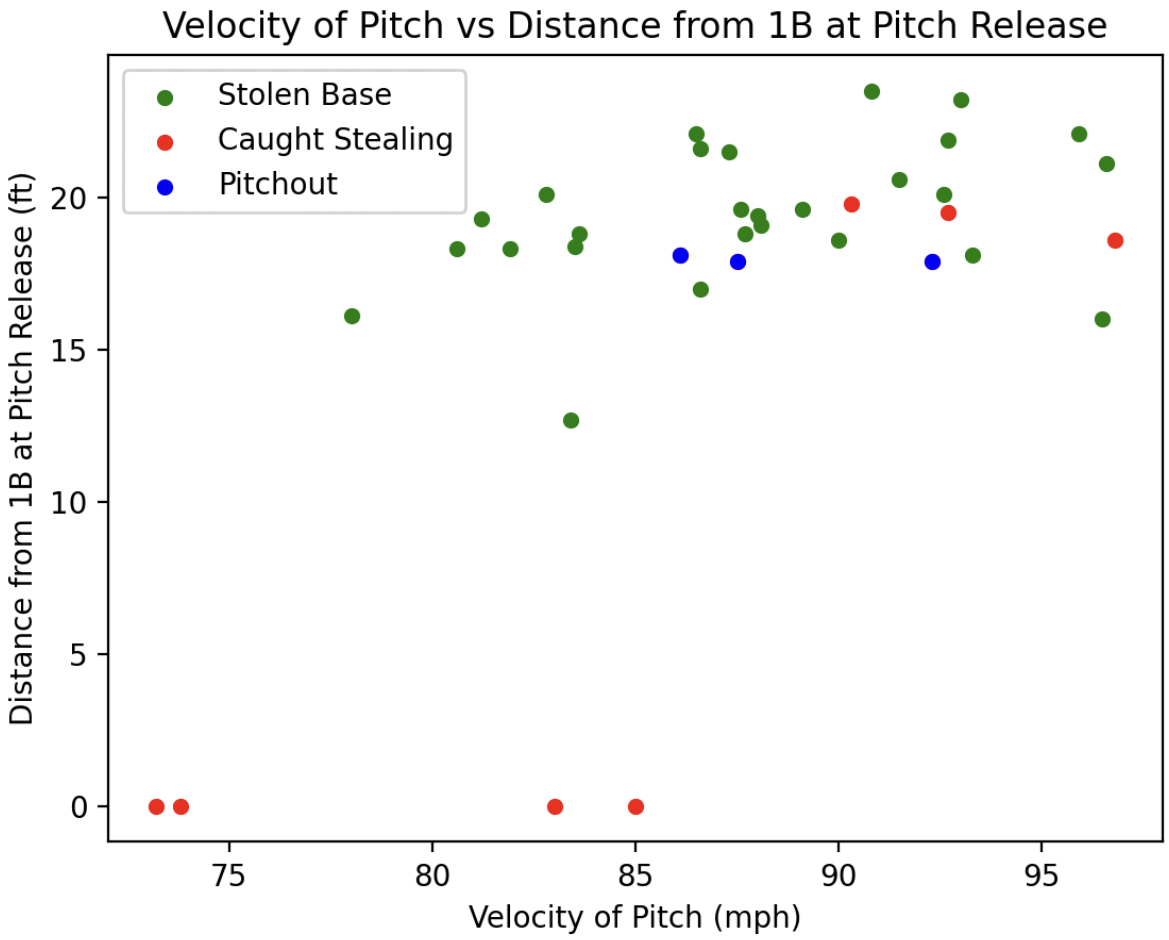
Ronan Hwang

In this project, I explored the possibility of preventing Elly De La Cruz - the stolen base leader of the 2024 MLB season and arguably one of the fastest players in the league - from stealing second base. Firstly, I sought the factors that are considered when stealing a base. The factors that I thought were important were: a pitcher's time to the plate, the time for the ball to reach home plate, the catcher's pop time, and the time for the runner to reach second base. To eliminate any bias I might have about these roles, I wanted to use a data-driven method to find factors that make Elly De La Cruz Successful in stealing second base. Then, using the obtained data, find a way to contain De La Cruz's stolen bases to second.

The first step in this process was to decide what statistics determine what factors come into a stolen base. To find my statistics, I combined Baseball Savant baserunning, catching, and pitching total season-level data. I decided that the best approach was to use his stolen base data from the 2023 and 2024 MLB seasons. One possible problem that I may come across as I continue to explore is that our sample size will not be big enough to draw a conclusion.

The key variable that I used to determine the probability of a stolen base is the runner's jump. Although some may argue that the true definition of a "jump" is how quickly a runner reacts after a pitcher's movement, I defined the "jump" as the distance separated from a pitcher's first movement to when the pitch is thrown. Getting his separation distance helped me take into account the time to plate of the pitchers and factor that into my analysis.

The first graph I created was to see if there was a relationship between the pitch velocity and how far De La Cruz gets off 1st base before the pitch and the stolen base outcome. Even if the pitch velocity and distance from first base do not relate to each other, I thought that including them would help see if those two factors affect De La Cruz's base stealing.

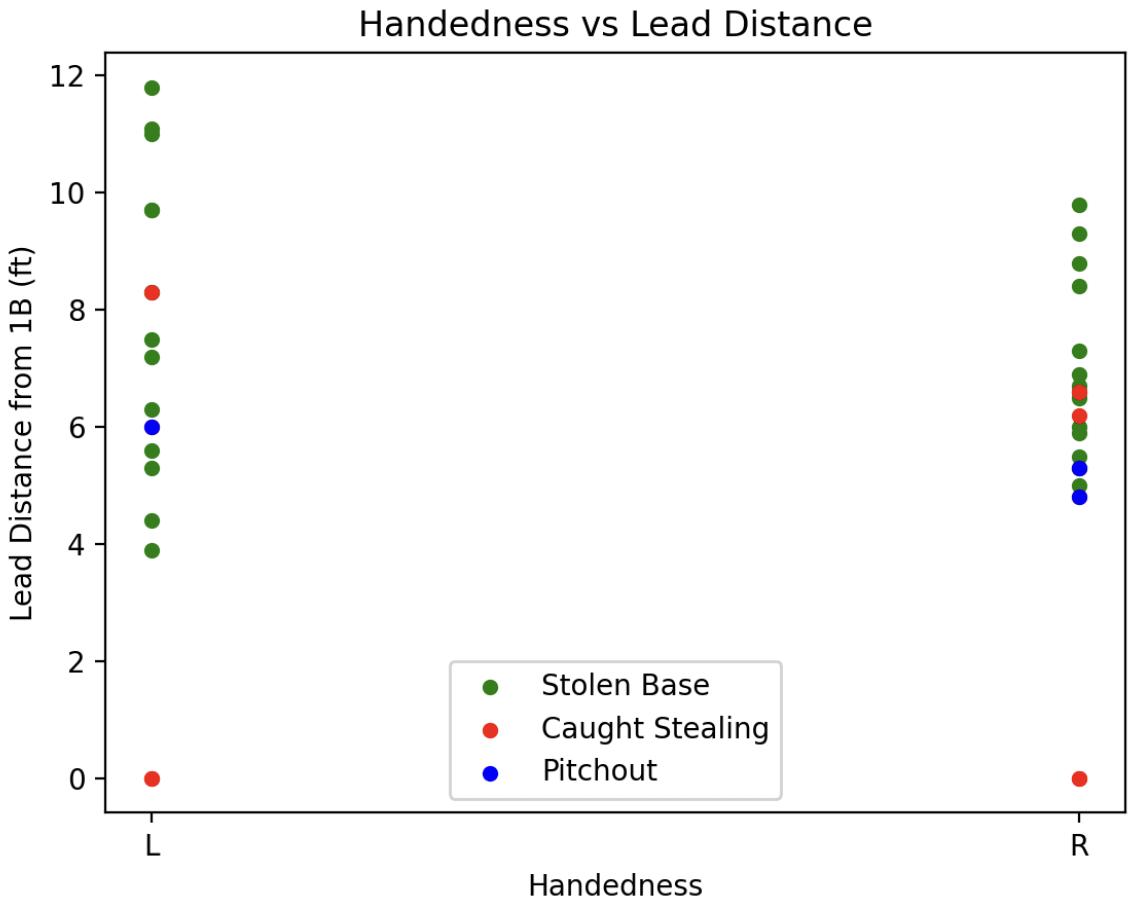


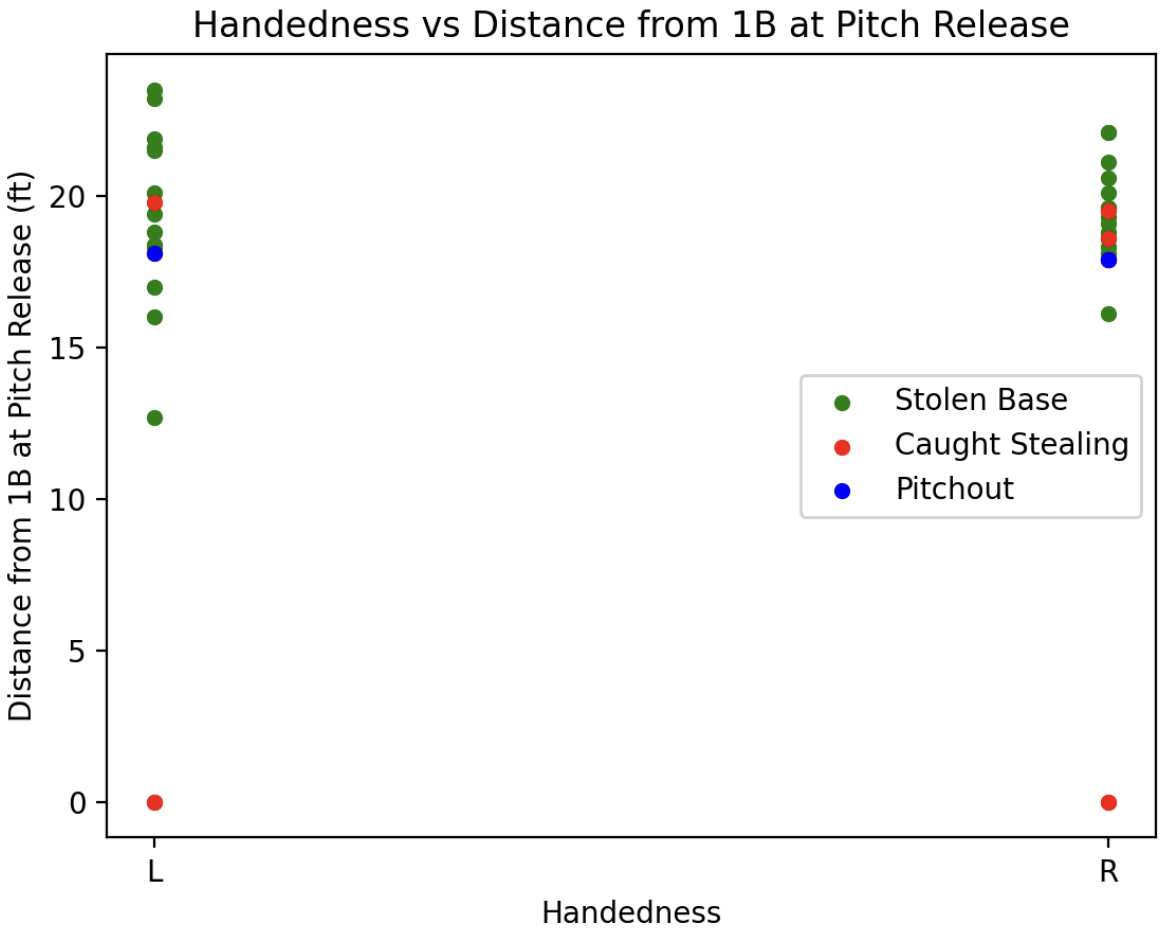
From this graph, it is visible that there is no clear relationship between the velocity of the pitch and his separation from first base with the stolen base outcome. The green dots indicate his successful stolen base outcome, the red conveys times when he was caught stealing(out), and the blue points indicate the times when there was an intentional pitch out to the plate. The blue points do not represent that he was out on a pitch out. Specifically, they show when a pitchout was attempted. This means that there was an occurrence in which they pitched out and still were

not successful in getting him out. Lastly, the data points where his distance from 1b at pitch release is 0 indicate that he was picked off by the pitcher.

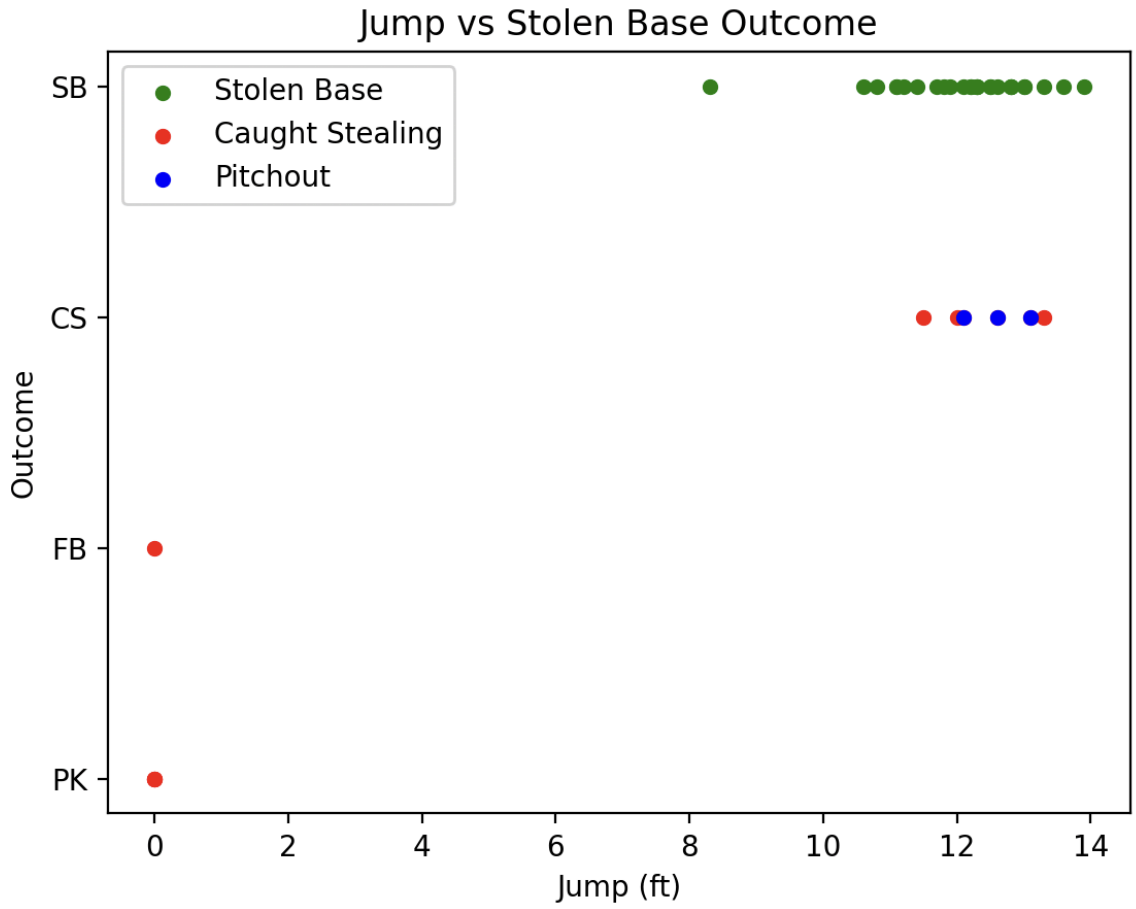
With our data, we could suggest that it is extremely likely for Elly De La Cruz to successfully steal 2nd base given that the pitch to the plate is under 85 mph.

I then tried to see if there was a relationship in the handedness of a pitcher. The graph below portrays the handedness of the pitcher and lead distance in relation to the stolen base outcome. The second graph shows the relationship between stolen base outcome, handedness, and distance from the first base at pitch release. I hypothesized that De La Cruz gets thrown out when he does not go off of the first move on a left-handed pitcher. However, I could only conclude that De La Cruz had more varied leads off of left-handed pitchers than he did with right-handed pitchers.





Lastly, I explored if the jump that Elly De La Cruz has relates to his stolen base outcome. Much like the previous three graphs, it was not possible to give a clear indication that the jump that he had was related to the outcome.



The goal for the rest of this project is to see how this data relates to the rest of the league. When the play-by-play data comes out for the 2023 season, I plan to retrieve all the stolen base data from the rest of the league from the season and match that data with Elly De La Cruz's stolen base outcomes. Using the league data and comparing it with De La Cruz, I can get a better view of whether it truly is possible to stop one of the fastest players in the league from taking an extra 90 feet. I plan on experimenting with the play-by-play data from the 2023 season and then transfer that code logic into the 2024 play-by-play data, but won't use this data to draw a conclusion. This is because I believe that the 2023 stolen data metrics will be different to the 2024 metrics(also something I can look into). Since the 2024 season is the second season since the implementation of the new pickoff/ pitch clock rules, organizations have a better idea of how

to utilize these rules to steal and prevent bases, compared to in 2023 when teams were still trying to get a feel for the new rules.