Idelson Taveras 6/4/2023

Toronto Blue Jays Exercise

A picture containing text, screenshot, diagram, plot

Description automatically generatedA picture containing text, screenshot, diagram, map

Description automatically generated During my analysis and decision making off the Deploy.CSV there were many different decisions to be made in the analysis. First, I decided to use Python for my predictive analysis and imported different packages such as NumPy, Pandas, matplotlib, and sklearn. I used these packages for data manipulation, logistic regression analysis for predictions and for data visualizations. I then found correlations between each of the variables to determine which variables had the best relationship. Velo, spin rate, and induced vertical break had the best correlations with each other. Horizontal break didn’t have good correlations with the other three, only with Induced Vertical Break but it still wasn’t strong enough as the others. After I found correlations, I decided to plot the certain linear relationships to see how each of them worked together. The higher the velocity the better the spin rate, the better the spin rate the better the induced vertical break was, and the higher the velocity was the higher the induced vertical break was. After visualizing the data then came time to make predictions using these three variables. Missing the history of the outcomes in the deploy.csv made it hard to predict since we didn’t have previous data showing history or results of these pitches like swing or misses. I decided to generate a random binary to determine whether each pitch was put in to play or not. This was a way to generate a y variable to determine the effects of these pitches. I ran a logistical regression model to predict whether each pitch was put into play or not. I tested the model and trained it as well and came out with the outcomes.

As you can see from the plot above three of the four variables have very good relationships with each other. For the pitcher, velocity, spin rate, and induced vertical break are the ones he should be focused on for his fastball. The higher his spin rate was the more induced vertical break pitches had which makes it more difficult for hitters to get on top of fastballs and be able to hit the ball squared forward. The higher velocity with good spin rate allows fastballs to appear harder than what they are and carry better towards home plate which also makes it tough for hitters to be able to see spin and recognize pitches.

The next steps as an analyst for me would be to gather more data, this time with the same variables but add a result column of swing or miss on the fastball. This way we know exactly at what velocity, spin rate, or induced vertical break players are hitting fair or swinging and missing. For right now with the information provided I would recommend to the pitcher to work on spin rate and induced vert as these are factors that can be worked on simply by changing grips and working on spinning the ball. Gaining velocity would be tasked in the off season as needed but focusing on spin rate would be my recommendation as we gather more information on the pitchers’ fastballs and outcomes on it.