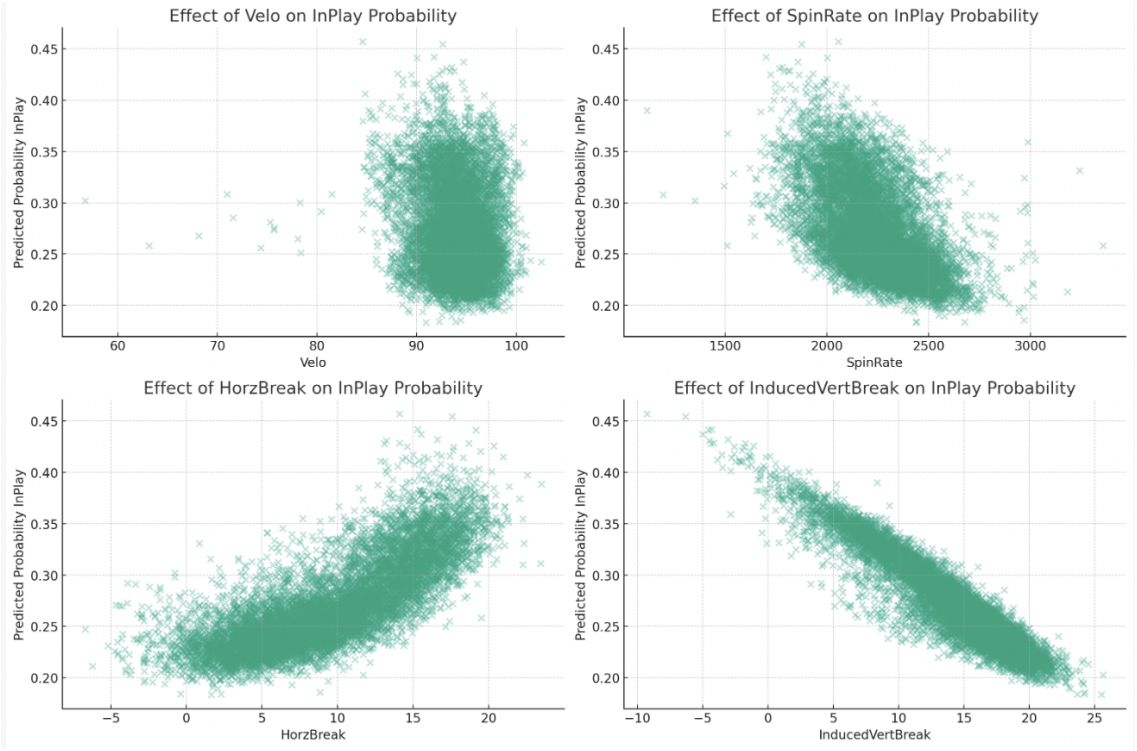
2. In this analysis, I started by loading and cleaning the training dataset. I removed rows with missing values. I then trained a logistic regression model using the four pitch-related features to predict the binary outcome of whether a ball would be put in play. After training the model, I used it to predict the probabilities of each pitch being put in play in a separate "deploy" dataset, which was also cleaned to remove rows with missing values. The predictions were then saved to a CSV file and further visualized through scatter plots to interpret the impact of each feature on the outcome. I chose a logistic regression model mainly because the binary nature of hitting a ball into play aligns well with logistic regression's suitability for such scenarios. Other benefits are that it offers clear insights and has a good computational efficiency.

3. Increasing the horizontal break of your pitch is likely to make it easier for the batter to put the ball in play. On the other hand, increasing the pitch velocity, spin rate, or induced vertical break will make it more challenging for the batter.



4. If I had another week to work on this question, I would work on refining the logistic regression model by incorporating more advanced techniques like feature engineering and hyperparameter tuning. Additionally, I would consider evaluating other machine learning models, such as Random Forest or Gradient Boosting, to compare their performance and potentially improve prediction accuracy. As for the analysis of the results, I would conduct a deeper dive into the variable importance to identify which factors are most influential in affecting a batter's ability to put the ball in play. This could involve advanced statistical tests or machine learning interpretability tools. Also, I would prepare a comprehensive report or dashboard to present these findings in an easily digestible format.