Jack Harenchar

Professor Martin Barron

ITAO-40520

25 February 2025

*Batting Across the Bay: Analyzing the Rays’ Offensive Evolution at Steinbrenner Field*

1. **Introduction**

In the early hours of October 10th, as Hurricane Milton made its way across the state of Florida, the storm’s fierce wind and piercing rain tore to shreds the fiberglass dome of Tropicana Field, home of MLB’s Tampa Bay Rays. Tropicana Field, built in 1990, was due for replacement by 2028, but the damage rendered it unusable for the upcoming 2025 season. As the city of St. Petersburg, Pinellas County, and the Rays work to replace the fiberglass roof, the team will play their 2025 home games at George M. Steinbrenner Field. The stadium currently plays host to the New York Yankees for Spring Training and the Tampa Tarpons, the Single-A affiliate of the Yankees, during the minor league season. Besides being across Tampa Bay from the Rays current home, the stadium’s field is also larger and most notably, outside.

Considering the Rays shift from the domed, climate-controlled Tropicana Field to the outdoor Steinbrenner field, exposed to the heat and unpredictable storms of south Florida, the problem I am considering is how will this shift impact the Rays’ performance at home. Of course, in a smaller, open-air stadium, the team may feel less of a home field advantage, but more importantly, will this transition help or harm the team’s ability to score runs? I think it will be fascinating to see how a different sized stadium impacts the team as well as how playing so many more games under the hot Florida sun impacts the team daily and over the course of the season.

The inputs that I will use for this project are pitch by pitch data that I scraped from Statcast for all home games for the Rays and the Tarpons. I filtered that data down further to include only when the home team, Rays and Tarpons, were batting in their respective parks. For most of my models, I also focused on the summer months of June through August since I think these warm weather months may have the greatest impact on offensive performance when transitioning from playing baseball inside to outside.

Using the Tarpons home ball in play data to train a model and the Rays home ball in play data to test it, my intended output for this project was to see how predicted hit results would change if the Rays balls in play were under the same conditions as the Tarpons. I took the Rays’ offensive performance from 2024 and tried to predict how it would change if they were playing outside at Steinbrenner Field and their batted balls were affected by the elements that they will come to face this upcoming season. My goal once I receive this output data would be to offer insights as to what aspects of their offensive game the Rays should focus on to help counterbalance their transition from playing inside to out.

1. **Related Work**

Few teams have ever shifted from an indoor stadium to an outdoor stadium from one season to the next. Even fewer, if any, have made that transition in such an intense climate as south Florida has during the summer. While I’m sure this is an analysis both the Tampa Bay Rays and Major League Baseball conducted at a deeper level than the scope of this project, few people in the past have had to look at what would happen to a team if they had to shift the playing environment of their 81 home games from one season to the next. Just recently though, ESPN published a piece intended to help fantasy baseball players that discusses their projections for the Rays this season at Steinbrenner Field. It discussed how several factors impact the flight of a baseball including outfield fences, the batter’s eye, foul territory, temperature, humidity, wind, and altitude. It ultimately concluded that over the course of the season both home runs and batting average would be expected to increase for Rays left and right-handed batters at Steinbrenner field. Unfortunately for pitchers, their ERA would also be expected to increase playing in this warmer outdoor climate.[[1]](#footnote-1)

1. **Methods**

To predict the hit results of the 2024 Rays were they to play under the conditions of the 2024 Tampa Tarpons at Steinbrenner Field, I used an XGBoost model. I chose to use an XGBoost model because I wanted a classification model that would be able to iterate through different trials, focusing on the errors, in order to output the best predictions under my given conditions. I ran this model 3 times on various variables and training sets to get a good idea of how the conditions at Steinbrenner Field may affect the results of Rays balls in play. As mentioned previously, the data I used to train and test these models was pitch by pitch data scraped from statcast using the statcast\_search function for Major League data and statsapi\_gameinfo for Minor League data. I then filtered both data frames down to Rays and Tarpons at home as well as bottom of the inning to get home team batting. I then looked only at balls in play because I was concerned with predicting results for balls batted into play.

When running my XGBoost models, I kept my model settings constant for num\_class, nrounds, verbose, print\_every\_n, objective, and eval\_metric across each model. I set my num\_class, or the number of distinct classes for classification to the length of my factorized model response list. I set my number of rounds to 100 and set verbose to 1 to print out every fit. My print\_every\_n was set to 20 to print the 20th iteration. For my objective, I chose to use multi:softmax because I am doing a multiclass classification with factorized classes. I used merror as my eval\_metric to get the multiclass classification error rate which the model would seek to minimize in each iteration.

For the first iteration of my XGBoost model, I trained the model using Tarpons home balls in play from June through August and I used the columns of x and y hit location as well as trajectory. Trajectory included the values of fly ball, ground ball, line drive, and pop up, so I dummy coded these values for use in my model. For my test data to make my prediction on, I used balls in play for the Rays at home from June through August, and likewise used x and y hit location as well as trajectory columns.

For the second iteration of my model, I trained the model using the same Tarpons summer balls in play data, but this time added the columns of hit distance, launch angle, and exit velocity to my three previous training variables. Before adding these three columns, I replaced their NAs with the column means. I then trained my model using the aforementioned parameters and then tested and predicted my data on the Rays home summer balls in play using the same 6 variables as I did for my model training.

For the third iteration of my model, I made some significant changes including encoding additional columns into my data and expanding my model training set. Because the team is making a transition from playing in a climate-controlled stadium to playing outside, I wanted to try to include temperature and wind variables to see if they have any effect on the hit results. I created an average wind and an average temperature column for the Rays ball in play data and made the temperature a constant 72 degrees Fahrenheit and the wind a constant 0 miles per hour. For the Tarpons data, I also made an average temperature and average wind column which I encoded with the monthly average temperature and the monthly average wind speed for each month of games. To train my model, I also expanded my data from the Tarpons home balls in play for June through August to the entire season so I could try to maximize seeing how the temperature and wind variables affect hit results. I once again tested and predicted my model on Rays home summer balls in play using the same column variables that I trained my model on.

1. **Discussion**

I found it fascinating that as I trained my model on more variables and larger datasets, the predicted change in Rays hit outcomes varied less and less from their offensive performance at Tropicana field in 2024. Ultimately in each of my models, the number of predicted outs was greater than actual outs, but that seemed to be due to singles becoming outs rather than doubles, home runs, or triples becoming outs. These extra base hits saw an increase, potentially due to the effects of the new stadium dimensions and playing in the open Florida air. One of my biggest takeaways from running these models is how much randomness is likely to play a part in hit results due to defensive player positioning or the random conditions at the time of any given batted ball.

**Figure 1 Figure 2**

A number and text on a white background

AI-generated content may be incorrect. A graph of a bar graph

AI-generated content may be incorrect.

My first model saw the most significant change between the acutal Rays hit results and the new predicted outcomes under Steinbrenner Field conditions. As can be seen in Figure 1, outs increased by 33 overall, but extra base hits saw an increase as several outs became doubles, triples, and homeruns while some singles and doubles became doubles and home runs, respectively. Most increases we see are single base increases with singles and doubles becoming doubles and triples, but some doubles did become home runs. In Figure 2, we can see how the x and y location of the hits seemed to play the biggest role in how they were ultimately classified.

**Figure 3 Figure 4**

A number and numbers on a white background

AI-generated content may be incorrect. A number and numbers on a white background

AI-generated content may be incorrect.

Models 2 and 3, whose outputs can be seen in Figures 3 and 4, respectively, varied noticeably from the first iteration of my model, but less so among themselves. In Figure 3, you can see how when running my model the second time after adding hit distance, exit velo, and launch angle variables, far fewer singles became outs as many more singles stayed as singles. Fewer home runs also become doubles, and no home runs were lost to outs. Only 15 actual hits in 2024 were predicted to become outs in model 2 versus 33 the first time I ran my model. In Figure 4, one can see how even fewer hit results are predicted to change when model 3 is ran after adding average temperature, average wind, and expanding my training set. Fewer balls in play are predicted to be reclassified as doubles, home runs, or singles and far more hit results are predicted the same as what actual happened to the Rays balls in play at Tropicana field in 2024.

Ultimately, when looking at where the changes from actual to predicted hit results occurred, as well as the variable importance in my models, distance and launch angle seem to be the two factors that players can control the most that the Rays should focus on. By working with their players on strength training in the offseason and spring training, the Rays can help their players to hit the ball harder and farther to adjust to farther fences. Additionally, the team can focus on finding the ideal launch angle where they keep the ball off the ground, preventing singles from becoming outs, and hitting line drives that take advantage of an outwardly wind to travel further.

1. **Conclusion**

Recognizing that a lot of factors may be out of the control of the team and vary pitch by pitch, I recommend that the Tampa Bay Rays spend their spring training working with batters on strength training and lowering launch angles to hit the ball further and on a line. This should help the team adjust for further fences, and by focusing on line drives, the team can work to limit singles predicted to potentially be outs instead, and getting the ball airborne to maximize extra base hits. So much of hit results is determined by where the ball ends up and the position of the defense when any given pitch is thrown, so besides hitting it out of the park, batters can do little to try to influence whether their batted ball is able to drop for a hit. By focusing on the factors such as the distance the ball travels and the angle it is hit at, the batters can do their best to try to influence the factors shown to have the biggest impact on predicted outcome. It is also important to note that throughout this project, the training data has been that of minor league batters. The increased offensive power and better defensive positioning and IQ of major league players would certainly play a role in predicting the difference in offensive output from indoors to outdoors but is something I am unable to account for in this project.

If given more time and resources, I would love to try to figure out how these hit result predictions would change the expected runs of a team and dive deeper into the specific effects of temperature, wind, and humidity on the ball from inside to outside.

**BIBLIOGRAPHY**

Zola, Todd. “Fantasy Baseball Impact of New Homes for Rays, Athletics.” *ESPN*, ESPN Internet Ventures, 18 Feb. 2025, www.espn.com/fantasy/baseball/insider/story/\_/id/43812146/fantasy-baseball-new-2025-home-stadiums-athletics-rays-stats-projections-stats.

1. Todd Zola, “Fantasy Baseball Impact of New Homes for Rays, Athletics,” ESPN, 18 Feb. 2025. [↑](#footnote-ref-1)