****

**Mariners 2025 Analytics Intern Problem Set**

Below is the set of problems we would like you to complete. You will be evaluated not just on your predictions, but also on your ability to communicate clearly and the quality of your process.

1. In the Google Drive folder linked below you will find three files: `data-train.csv`, `data-test.csv`, and `data-dictionary.csv’.

We have provided the 2023 season of contacted balls for two Minor League levels, along with play metadata and Trackman hit tracking information, pre-filtered to non-infield plays. This **includes** foul balls and home runs not necessarily hit into playable territory. The `train.csv` file includes two additional columns: on balls caught for outs, `is\_airout` will be 1 and `first\_fielder` will give the player id responsible for the out. On balls not caught for outs, `is\_airout` will be 0 and `first\_fielder` will be null.

Your objective is to predict the air out probability for batted balls in the `data-test.csv` file and fill out the `p\_airout` column in that .csv with your estimate. To assist you, we have included the ‘data-dictionary.csv’ file to explain what each column in the attached datasets represents.

You may use whatever method or language you like, but **you must submit the code you used to generate and evaluate your predictions**. You will be evaluated on both the **log loss** score of your predictions and on **your process in generating the predictions.** Please also include a brief explanation of your process for an R&D audience and what steps you would take to improve on this model if you had more resources.

1. It is October 1st, 2023, and with the conclusion of the Minor League season the Director of Player Development of the Seattle Mariners is interested in 15411’s outfield defense. Based on the data provided, and assuming all other things being equal, write a one-page report for a coaching audience breaking down this player’s defensive performance and abilities.
2. In approximately 300 words, what is a recent mistake the Mariners organization has made, and why do you consider it a mistake?

Data for problems 1 and 2 can be found here: https://drive.google.com/drive/folders/132Svb8NETigiTtjHiMWh\_V0ffRKeX60-?usp=sharing

Due to the large size of the train and test datasets, consider either zipping the file before emailing or storing it in a public cloud storage site like Google Drive or Dropbox.

If you have any questions, or need clarification of any kind, please message Forrest Diamond at [fdiamond@mariners.com](mailto:fdiamond@mariners.com).