



Mariners 2024 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 five files – ‘train_first_half_batted_balls.csv’, ‘train_second_half_slg.csv’, ‘test_first_half_batted_balls.csv’, ‘test_submission_df.csv’, and ‘dictionary.csv’.

We have selected hitter-seasons where a batter recorded at least 100 batted balls in both the first half (March through June) and second half (July through October). For half of the hitters, we have provided you with information about each batted ball that they recorded in the first half of the season (‘train_first_half_batted_balls.csv’), and their SLG on batted balls in the second half of the season (‘train_second_half_slg.csv’). For the remaining hitters, we have given you only the information about their first-half batted balls (‘test_first_half_batted_balls.csv’).

Your objective is to predict the SLG for batted balls for the second half of the hitter-seasons listed in ‘test_submission_df.csv’ and fill out the ‘second_half_slgcon’ column in that .csv with your estimate. To assist you, we have included the ‘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 **root-mean-square error** 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.*

2. It is July 1st, 2023, and the general manager of the Seattle Mariners is interested in trading for a hitter. They can acquire either hitter xe54ef656 or hitter x084802d3. Based on the data provided, and assuming all other things being equal, write a one-page report for the general manager of the Mariners advising them on which of hitter xe54ef656 or hitter x084802d3 they should acquire.
3. 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/1j3-V48EhaGRKwS2SJSp0Yq_QRHM9nqKN?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 John Edwards at jedwards@mariners.com.