

Hello!

Below is the set of problems we would like you to complete. As stated in an earlier email, you will be evaluated not just on your accuracy or correctness, but also on your ability to communicate clearly and potential for growth.

The data can be found here:

<http://drive.google.com/drive/folders/1U4qdlxuBplBEtzE6vicG6K0zTP8IIRsO>

If you have any questions or need clarification don't hesitate to ask!

#### Coding Challenge

- 1) In the google drive folder you will find two files, train.csv and test.csv. We have given you the specific pitch results in the training dataset ('pitch\_call'), but they are not available in the testing dataset. Your objective is to fill in the 'is\_strike' column in the test dataset as accurately as possible, 1 for strike or 0 for ball. Please note that the same mix of 'pitch\_call' may appear in the test dataset! You may use whatever method or language you like, but you must submit the code you used to generate and evaluate your 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) The front office has brought in catcher f06c9fdf this offseason. Using the data provided, please prepare a report **for a coaching staff audience** to brief them on the player's current strengths and weaknesses as a pitch receiver. The report (explanations and any visuals) should be less than two pages long. Please also submit the code used to generate the brief, though your process here is less important than the way you communicate your finding in the report.

#### Written Response

- 1) What is a recent (within the past one to two years) mistake the Mariners organization has made, and why do you consider it a mistake?
- 2) Propose a rule change MLB could make that would make baseball more fun.

Because of the large size of the test dataset, I would recommend either zipping the file before emailing or storing it in a public google drive or dropbox or github or similar cloud storage site. You may also choose to keep only the most important columns (for our evaluation purposes we need pitch\_id and is\_strike).