Just to add, my goal is to be able to develop models that generalize well on unseen data (sorry if I'm stating the obvious here).

In order to do that, I'd like to have a Validation (and Test) set that is representative of the Training set. My concern is that that may not be currently be the case with the existing Validation set. So if I optimize my model on it, the model may not be able to generalize as well on future data?