Model Summary

The only data added to the model was 2018 OBP and PA by player.

The tidymodels framework in R was used to run the models. Two models were tested, a linear regression and a regularized regression, to predict the expected MarApr OBP from the variables. Feature engineering

normalized all numeric predictors, removed highly correlated predictors and a principal component analysis was done prior to modeling.

For the regularized regression, grid search was used to choose the optimal tuning parameters. The regularized regression performed marginally better and was selected as part of the final model.

The predicted MarApr OBP was combined with 2018 stats weighted by plate appearances with 2018's PA with the weight of 2018 plate appearances reduced by 50 percent.

The rmse for the final model was .0305 and r-squared was .41.