Python Proficiency Exercise

Last Updated January 17, 2020

The data set for this exercise is found in the file `Baseball Data.` It shows the outcome of the 2016 Fantasy Baseball League. Your tasks are as follows:

1. Import the database as a Pandas dataframe.
    a. Which player had the highest batting average? The lowest? Make sure to print out your results.
    b. Which player had the highest # of home-runs per game?
    c. Do players with higher batting averages tend to score more home runs per game? Create a scatter plot and determine if a relationship exists (make sure to include labels!).
2. Create a numpy array of Games Played.
    a. What is the mean number of games played? The median?
    b. Plot a histogram of this data, choosing an appropriate bin size, and observe the distribution.
3. Isolate players who play on 1st Base, and those who play outfield.
    a. To which position does the highest home-run scoring player belong to?
    b. Compare the means and medians of batting averages. Can you conclude that one group hits more successfully than the other?
4. Submit your work as a Jupyter notebook (.ipynb file)