

# Data Visualization Project

## Summary

This report covers a visual analysis of the baseball dataset provided by Udacity. The dataset consists of information about the name, handedness, height, weight, batting average and home runs for 1,175 baseball players.

The report explores performance of baseball players based on their batting average and home runs scored. We have used handedness, height, weight, and BMI (calculated) as parameters to explore the performance of the players.

Below are the main findings from the dataset:

- Left handed players have a higher median batting average and median home runs compared to other players
- Shorter players have a better batting average. Similarly, players with comparatively lesser weight have a better batting average
- Players with a BMI more than 25.0 have better performance on the parameter of Home Runs. Also, players with a BMI between 22.5 and 25.0 performed below average on the batting average parameter

## Design

To explore the dataset, we had to modify existing variables and calculate one new variable:

- Weight: Distribution of the weight variable revealed an inconsistent trend of high number of data entries for multiples of five. Taking this trend into account, the weight variable was recalculated with a bin size of 5
- BMI: A new variable (BMI) was calculated to check for any trends in performance of the players based on their BMI

All the analysis was visualized using bar graphs, line plots, and box plots to keep the visualizations easy to interpret

## Feedback

[Version 1](#)

[https://public.tableau.com/profile/archit6167#!/vizhome/DataVisualisationProject\\_4/Story\\_v1?publish=yes](https://public.tableau.com/profile/archit6167#!/vizhome/DataVisualisationProject_4/Story_v1?publish=yes)

[Final Version](#)

[https://public.tableau.com/profile/archit6167#!/vizhome/DataVisualisationProject\\_4/Story\\_vF?publish=yes](https://public.tableau.com/profile/archit6167#!/vizhome/DataVisualisationProject_4/Story_vF?publish=yes)

- Feedback regarding how the story seems abrupt and straight away jumps into the description of the analysis. To solve for this an additional story point was added to the beginning of the story which describes the various variables of our dataset and helps the reader get acquainted with the data
- One reader pointed out that the analysis seemed to be a basic description of various cuts on the performance variables. This led us to calculate the BMI of players and present a certain analysis that was not directly attained for the original variables

## Resources

- The online help for tableau