

Name: Hailey Burnette

Id: @0354283

Fill in each item on this form using a list format (instead of writing a paragraph). Each new idea should get a new bullet. Make sure to explain/give details where necessary. For example, in answering item 4, instead of just giving “bar graph”, explain with which variables, which variables on which axis, etc.

Dataset: Fish Species Sampling Data - Length and Weight

Link: <https://www.kaggle.com/datasets/taweilo/fish-species-sampling-weight-and-height-data>

1. Comment on the quality of the data.

The quality of the data was really good. There was a lot of samples and no missing values.

2. What are the primary columns of interest?

Species of sample, length of sample , weight of sample, and length:weight ratio of sample.

3. What are the filters of relevance?

No filters of relevance were needed, because all of the data was relevant to our work.

4. Give potential visualization options.

Bar graph, or histogram.

5. Discuss any limitations of the dataset.

Potential limitations include not having an exactly equal number of samples per species, which threw off the mean, median, mode and range values of the entire data set slightly.

6. Give ideas for how the dataset can be improved/expanded/deepened.

The dataset can be improved by ensuring an equal number of samples per species, and increasing the number of samples per species because the larger the dataset, the more accurate the basic operations will be compared to the actual averages of values of these animals.