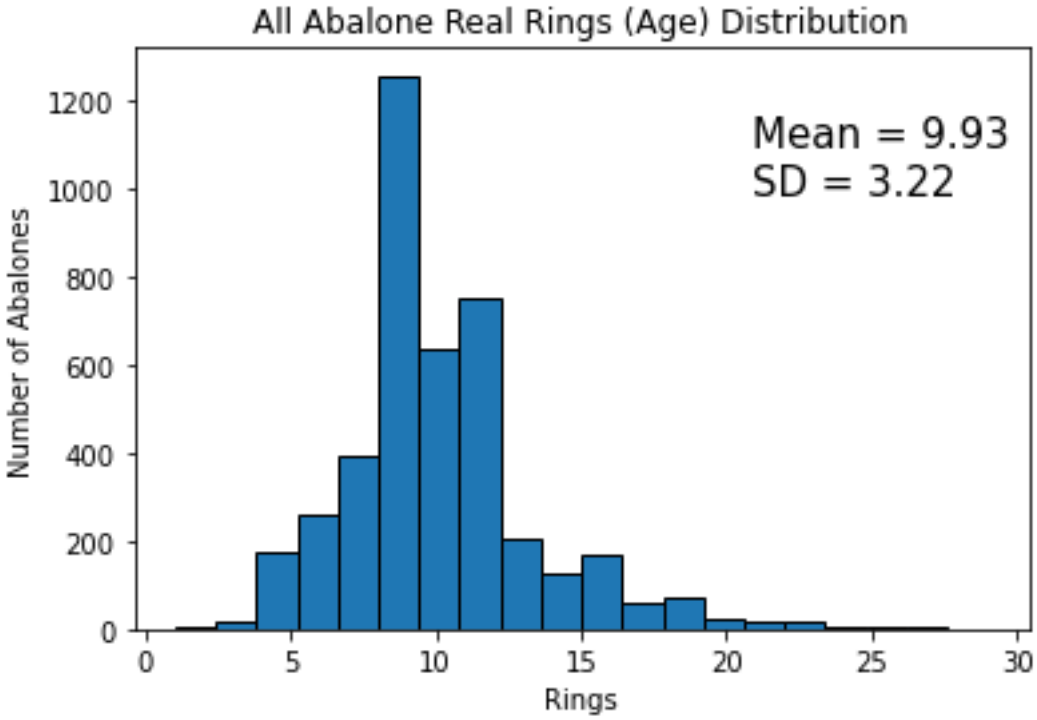
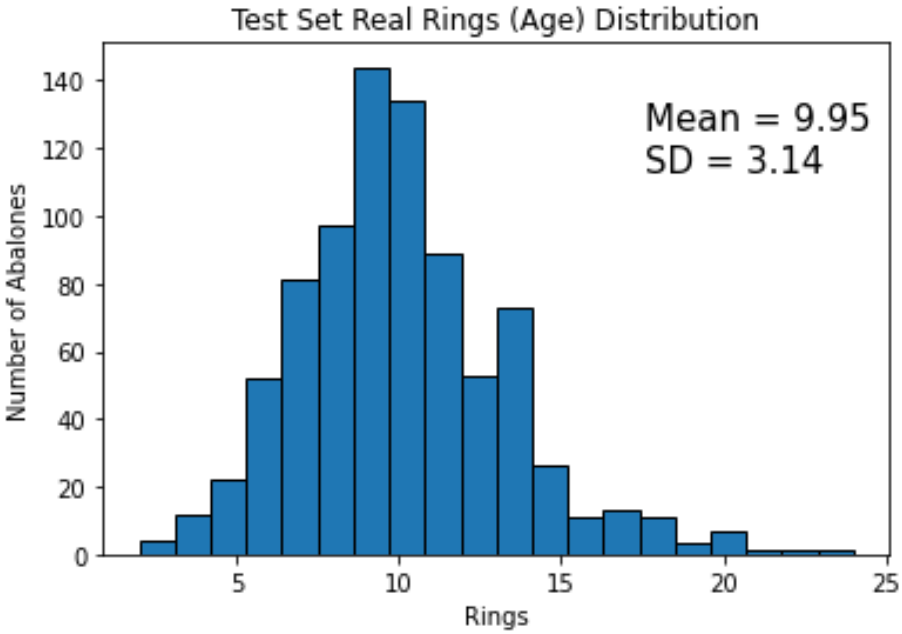
The age of an abalone can be found by cutting its shell and counting the number of rings on the shell. In the Abalone Dataset (abalone.txt), you can find the age measurements of a large number of abalones along with a lot of other physical measurements.
The physical measurements of an abalone include: "Sex", "Length", "Diameter", "Height", "Whole weight", "Shucked weight", "Viscera weight", "Shell weight", "Rings", among which “Sex” cannot be used as a features, but a label(class).
The goal of this exercise is to develop a model that can predict the age of an abalone based purely on the other physical measurements, except ‘sex’. This would allow researchers to estimate the abalone’s age without having to cut its shell and count the rings.

(First, plot the rings (ages) distribution of all abalones in the dataset and the randomly sampled 20% abalones Test Set, like: (hint: You don’t have to plot the rings (ages) distribution of the 80% Training Set.))

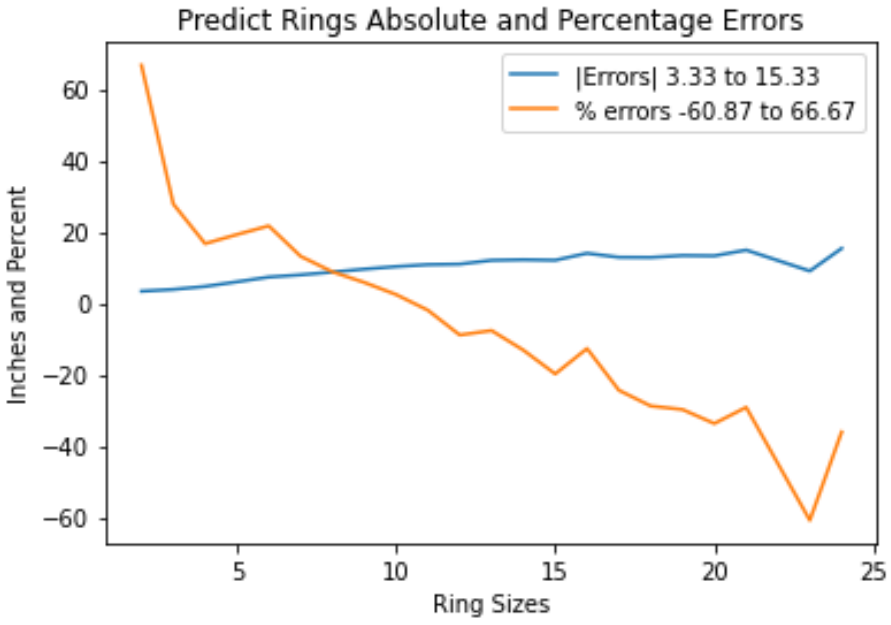




You'll be applying a **kNN** to find the closest prediction score possible. Please test various k in [3,5,7,9] to see the result. Then for each k, print the followings and plots the corresponding figures like:

Pre-Training Whole Examples Evaluation with k= 3 is: rSquare: 0.2614 Rmsd: 2.3229
Root Mean Square Deviation for k= 3 is: 2.037266464226766
After Trained Testing Using Test Set with k= 3 is: rSquare: 0.5109 Rmsd: 2.3269

(For all the samples of a ring size (age) in the Test Set, use the **kNN** to predict each of their ring size, and find the **Absolute errors of the predictions**. On the other hand, also calculate the total **prediction percentage error** to the actual ring size(age) of all the abalones of the ring size.)



(Hint: to produce the above figure, you have to count and save the number of abalones of each ring size(age) in **a list of lists** (hint: each list is for a ring size) for the Test Set)

(Finally, for each k, plot the histograms of the **Absolute Errors of the predictions** and **Percentages Errors of the predictions** of the test Set)