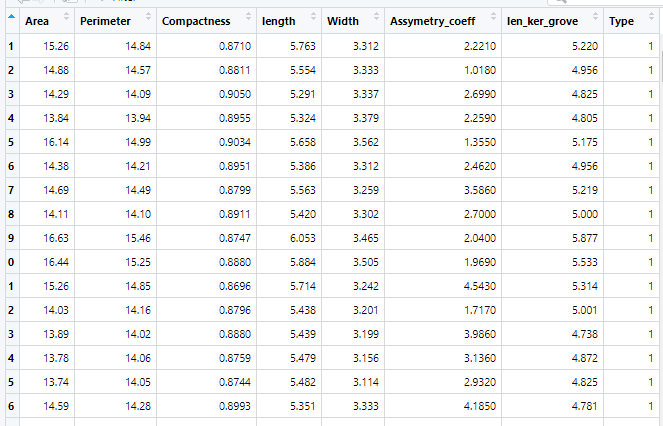
**STANDARDIZATION & NORMALIZATION**

**Problem Statement:**

Data is one of the most important assets. It is often common that data is stored in distinct systems with different formats and scales. These seemingly small differences in how the data is stored can result in misinterpretations and inconsistencies in your analytics. Inconsistency can make it impossible to deliver reliable information to management for good decision making. We have the preprocessing techniques to make the data uniform. Explore the various techniques to have reliable uniform standard data, you can go through this link:

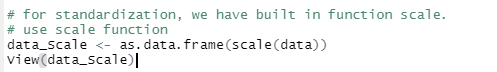
<https://360digitmg.com/mindmap-data-science>

1. Prepare the dataset by performing the preprocessing techniques, to have the standard scale to data which improves the model predictions .

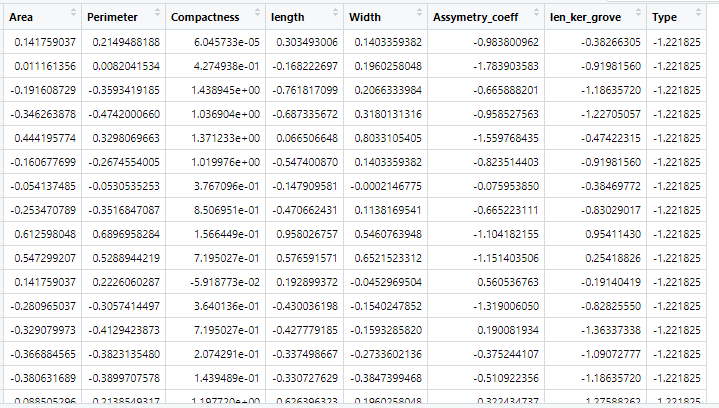


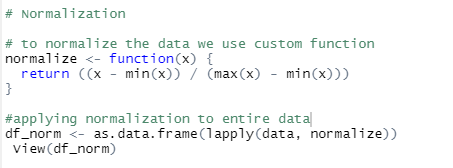
Standardization is done to remove the units of the data, so as the output remains same irrespective of the units of the data.

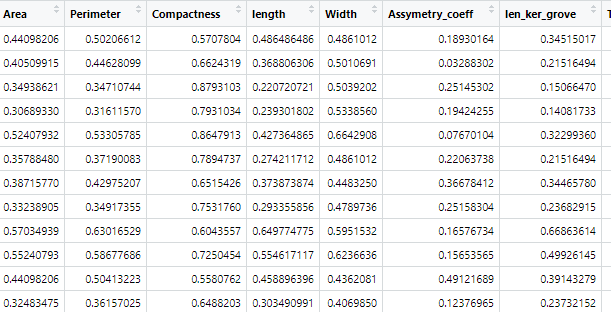
In R, we have the built in function called scale to standardize the data.



The output of the above function is as shown

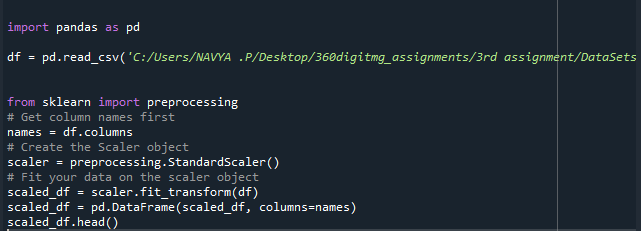




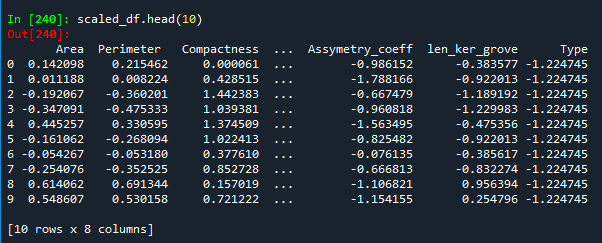


In Python, we use preprocessing function to standardize the data.

All the columns are standardized at once.



The output is as shown



**Hints:**

For each assignment, the solution should be submitted in the below format

1. Work on each feature to create a data dictionary as displayed in the image displayed below:
2. Refer to Seeds\_data.csv file
3. Research and perform all possible steps for obtaining solution
4. All the codes (executable programs) should execute without errors
5. Code modularization should be followed
6. Each line of code should have comments explaining the logic and why you are using that function