**A picture containing shape, arrow

Description automatically generatedSTANDARDIZATION & NORMALIZATION**

Instructions:

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: RAKESH SETHU N P**

**Batch Id: 11052022\_7.30PM**

**Topic: Data Pre-Processing**

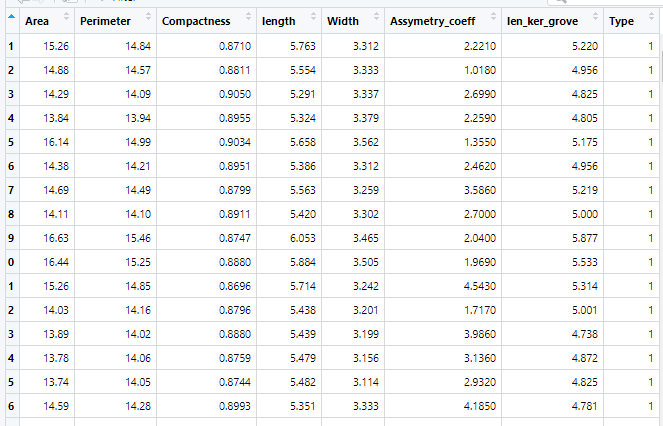
**Problem Statement:**

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

**A picture containing shape, arrow

Description automatically generated**

|  |  |  |  |
| --- | --- | --- | --- |
| Name of Feature | Description | Type | Relevance |
| Area | Area of seed | Continuous, Ratio | useful information |
| Perimeter | Perimeter of seed | Continuous, Ratio | useful information |
| Compactness | Compactness of seed | Continuous, Ratio | useful information |
| Length | Length of seed | Continuous, Ratio | useful information |
| Width | Width of seed | Continuous, Ratio | useful information |
| Assymetry\_coeff | Asymmetry coefficient of seed | Continuous, Ratio | useful information |
| len\_ker\_grove | length of kernel groove | Continuous, Ratio | useful information |
| Type | Type | Discrete, Count | useful information |

**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