

## Seeds

This data set from [data.world](https://data.world) was pulled from UCI. It has the following description:

“The examined group comprised kernels belonging to three different varieties of wheat: Kama, Rosa and Canadian, 70 elements each, randomly selected for the experiment. High quality visualization of the internal kernel structure was detected using a soft X-ray technique. It is non-destructive and considerably cheaper than other more sophisticated imaging techniques like scanning microscopy or laser technology. The images were recorded on 13x18 cm X-ray KODAK plates. Studies were conducted using combine harvested wheat grain originating from experimental fields, explored at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin.”

This dataset is great for classification or cluster analysis. For clustering, delete the target column. This dataset also requires some cleaning as it contains NaN values.

It has the following attributes:

- (1) area (A)
- (2) perimeter (P)
- (3) compactness ( $C = 4\pi A/P^2$ )
- (4) length of kernel
- (5) width of kernel
- (6) asymmetry coefficient
- (7) length of kernel groove

You can use the following code to load the data into Google Colaboratory:

```
import pandas as pd
url = "https://raw.githubusercontent.com/the-codingschool/TRAIN/main/seeds/seeds.csv"
df = pd.read_csv(url,
                  sep = " ",
                  names = ["area", "perimeter", "compactness", "length", "width",
                          "asymmetry_coeff", "length_groove", "target"])
```