Project-1: Self-Kaggle dataset

Kaggle dataset:
https://www.kaggle.com/datasets/cpluzshrijayan/milkquality

Data Cleaning:

- In the dataset provided, all columns are deemed appropriate and do not require further data cleaning.
- The decision not to remove outliers is based on domain knowledge. Altering or removing these outliers could significantly impact the determination of milk quality.
- Removing or modifying these data points could lead to a misrepresentation of the quality characteristics of the milk samples.
- Therefore, retaining outliers in this dataset is crucial as these extreme or atypical values might convey significant information regarding variations in milk quality.
- Any alterations or removal of these data points could distort the overall understanding of the quality attributes of different grades of milk, potentially affecting assessments, classifications, or decisions based on the data.
- The dataset, having undergone appropriate analysis and domain-specific considerations, is now deemed ready for model building, with all columns being relevant, and outlier retention ensuring a comprehensive representation of milk quality attributes.