[This is a Glass Identification Data Set from UCI 1](https://archive.ics.uci.edu/ml/datasets/Glass+Identification). The overall problem of interest is to classify the glass type based on the given attributes. The data set was donated by Vina Spiehler, Ph.D., DABFT Diagnostic Products Corporationand. The glass types are defined in terms of their oxide content (weight percent in corresponding oxide).

The data set contains 10 attributes including id and target. [The target is glass type (discrete 7 values) 2](https://www.kaggle.com/datasets/uciml/glass). The attributes are:

* Id number: 1 to 214
* RI: refractive index
* Na: Sodium
* Mg: Magnesium
* Al: Aluminum
* Si: Silicon
* K: Potassium
* Ca: Calcium
* Ba: Barium
* Fe: Iron
* Target: 1 to 7

The original source papers for this data set are not available. However, the data set has been used in several studies. Vina conducted a comparison test of her rule-based system, BEAGLE, the nearest-neighbor algorithm, and discriminant analysis. The goal was to determine whether the glass was a type of "float" glass or not. The study was motivated by criminological investigation. [Previous analysis of this data has been done by Jason Brownlee1](https://machinelearningmastery.com/imbalanced-multiclass-classification-with-the-glass-identification-dataset/) where he explored how to develop and evaluate a model for the imbalanced multiclass classification. In the classification task, we hope to learn how to predict the glass type based on the given attributes. In the regression task, we hope to learn how to predict refractive index based on other attributes. It shouldn’t be neccessary to transform the data in order to accomplish those tasks.

Regarding data issues, it’s important to note that this dataset does not have any missing values.

The basic summary statistic:

HERE THE TABLE WITH MEAN,MIN,MAX,STD …

All attributes (obviously except target which is categorical and nominal) are continuous and ratio variables. The reason they are considered a ratio attributes is because they have a clear definition of zero (e.g. no Sodium present), and the difference between two values is meaningful. For example, a glass sample with 2% Sodium has twice as much Sodium as a glass sample with 1% Sodium.