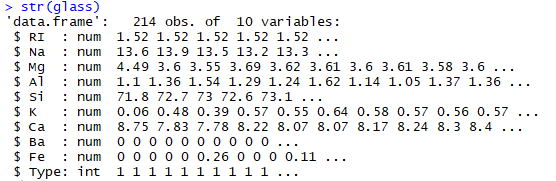
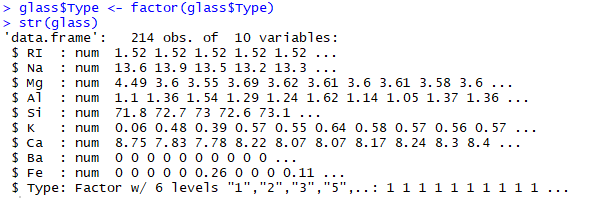
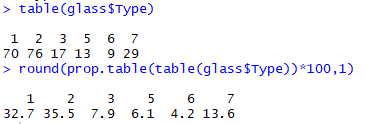
# KNN Assignment(Glass):

* Business Problem: To classify the TYPE of glass according to the given dataset
* Data Collection: glass.csv
* EDA: At First we look into the structure of the dataset, 
* We have to classify the TYPE of glass. In our dataset, “Type” is the output variable.
* “Type” is in integer format, so we have to convert into factor. We use below function and look at the structure again.



We can see that type got converted to factor.

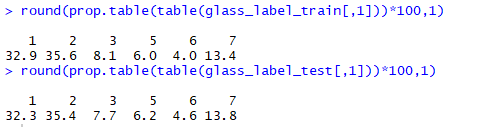
* Now we will look into number of observation belonging to different type of glass.



We can see that the dataset is imbalanced, because there is more proportion of Type 2 Glass than Type 6 Glass. So, we have to take care during splitting the dataset into Training and Test.

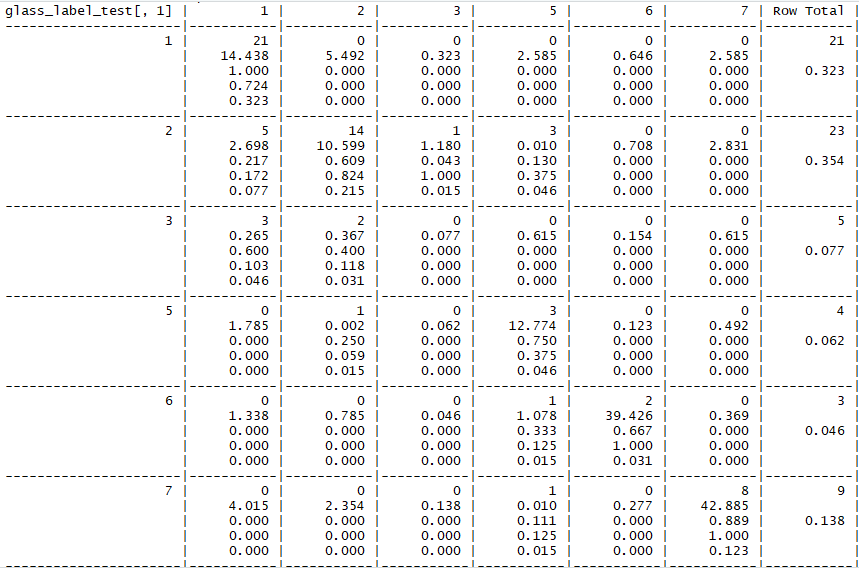
* We can also find that the dataset has scaling problem, So we normalize all the predictor variables and make a model. Next, we will standardize all the predictor variables and make another model. The model having more accuracy will be our final model.
* Next, we will split the dataset into training and test until there is approximately equal proportion of all types of glass in training and test. We will use random sampling for this.

After manually sampling the dataset we got approximately equal proportion as below.



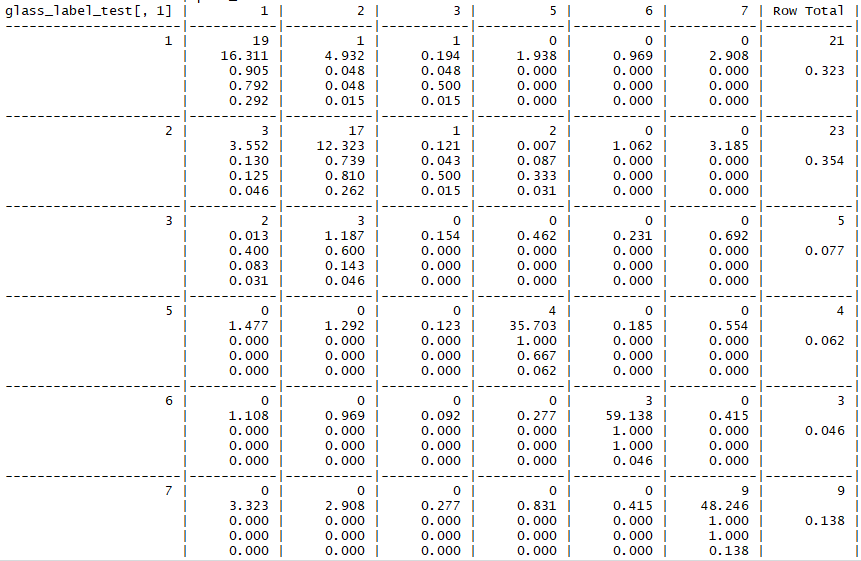
1. Data Mining:

* Next, we will Apply KNN algorithm and tweak K until we get a model with highest accuracy. Here we get highest accuracy at K = 3.



Screen Clipping

* Next, we will standardize the dataset and build a model. At K = 1, we get the highest accuracy. The output is shown below.



Screen Clipping

***So We will choose pred\_model2, with K = 1 as our final model***.