# MODULE 10

# **Package**

**Package Used in R**

mlogit - To run the Multinomial Regression

nnet - To run the test perdition on the train data

**Package Used in Python**

Pandas - Data Manipulation

Seaborn - Data Visualization

Matplotlib.pyplot - Data Visualization

Sklearn - Used to import other Sklearn features

Train\_test\_split - Split the data for training and test

Logistic Regression - To run the regression model

Accuracy\_score - Calculate the accuracy of the model

**Loading the Data set**

**Plot**

Box plot showed no outliers in the dataset

No Scaling issues found in scores variable

Majority of the students has scored average marks

Only few has scored below 35 in the scores

No has scored above 80 marks except in Read scores

**EDA**

1. No Nan Values found in the dataset
2. No Outliers found in the dataset
3. Typecasting ses Variables to numeric
4. Typecasting schtyp , honors Variables to 0 and 1
5. Removing unwanted Variables like X, id, female.
6. Rearranging the Prog Variables as the 1st Variables in the dataframe
7. The data set is the imbalanced dataset.
8. Academic has double the entry compared to General and Vocation

**Data Partitioning**

Dataset is split into training and test data with random sampling method.

Data is split into 80:20 Ratio.

**MODELING**

mdata1 <- multinom(prog ~ ses +schtyp +read+ write+ math +science+ honors,data=mdata)

#AIC 352.44

mdata2 <- multinom(prog ~ schtyp +read+ write+ math +science+ honors, data=mdata)

#AIC 355.89

mdata3 <- multinom(prog ~ ses +schtyp + math +science+ honors, data=mdata)

#AIC: 354.74

mdata4 <- multinom(prog ~ write+ math +science, data=mdata)

#AIC: 358.44

mdata5 <- multinom(prog ~ schtyp + math+ science , data=mdata)

#AIC: 357.64

mdata6 <- multinom(prog ~ read+ write+ math+ science, data=mdata)

#AIC: 356.82

Model 1 has better AIC value compared to other models.

**Accuracy**

Predicting the selected model with the test data and accuracy is 0.50

Now Predicting the model with train data and accuracy is 0.64

Dataset is imbalanced dataset. Academic has double the entry compared to General and Vocation. In the Prediction the Academic has been predicated more than General and Vocation.

