# M.Sc. IT Semester - 1

**PRACTICAL ASSIGNMENT**

Subject: Fundamental of Data Science

# 

Download given CSV of Water quality dataset and perform following operations.

Url: 10.25.50.252/water\_potability.csv

* 1. Import csv file and perform exploratory data analysis by identifying size of the datasets, column information and null values.
  2. Identify column wise missing values in the dataset and replace them with mean values of the respective column.
  3. Perform co-relation analysis for the given columns and generate co-relation matrix for the same.
  4. Implement Naive Bayes Classifier to predict Water Potability.

# 

Download given CSV of DMart Products and perform following operations on the dataset.

Url: 10.25.50.252/DMart.csv

* 1. Import csv file and perform exploratory data analysis by identifying size of the datasets, column information and missing values.
  2. Identify Top 10 Most Product Categories having highest products

and plot graph for the same.

* 1. Perform following operations. (Any 2)

1. Fill the missing value of Brand with value ‘Others’.
2. Find out 10 brands which are having highest products.
3. Identify and Remove columns which are not useful for data analysis.
   1. Implement Naive Bayes classifier to predict product category. Perform lable encoding if required.