**M.Tech Data Science**

**Statistical Modelling Lab – 2**

Load the dataset file named Dataset\_1.csv into a data frame and perform the following using python code:

1. Display rows **3** through **8** and all columns of the Data Frame
2. Display the total purchase amount(purch\_amt) of each customer(Each customer is represented by a unique customer ID).
3. Replace the missing values with the value 999.
4. Create a lineplot for plotting ord\_no and purch\_amt
5. Delete the last 3 rows of the data frame.

Load the dataset file named Dataset\_2.csv into a data frame and perform the following using python code:

1. Display rows 1, 3, 6 and all columns of the Data Frame
2. Replace the missing values with the mean of that column
3. Find the average height of students of each school. (School code represents the unique code for each school)
4. Perform the min-max normalization for the attributes- age,height and weight
5. Create a scatter plot for the attributes-age and weight for visualizing the relationship between them

Load the dataset file named Dataset\_3.csv into a data frame and perform the following using python code:

1. Replace the missing values with the max value of that column
2. Select the rows where the number of attempts in the examination is less than 2 and score greater than 15.
3. Find the total score of those who have qualified and those who have not qualified separately.
4. Perform the min-max normalization for the attributes- attempt and score
5. Create a scatter plot for the attributes-attempt and score for visualizing the relationship between them.