**M.Tech Data Science**

**Statistical Modelling Lab – 1**

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

1. Delete the 'SNo' column from the DataFrame.
2. Fill the missing values in the Maxpulse column with the value 120 and the missing values in Calories column with its mean.
3. Add the following row as a last row in the existing DataFrame

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 30 | '2020/12/11' | 103 | 145 | 319.3 |

1. Sort the data frame by 'Maxpulse' in descending order
2. Create a boxplot for plotting Date vs Pulse. The labels of x-axis and y-axis need to be displayed in the graph.

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

1. Delete the ' AdjClose ' column from the DataFrame
2. Fill the missing values in the High column with the **mean** and the missing values in Low column with its min value
3. Create a histogram plot of opening, closing, high, low stock prices
4. Select the rows where the Open is between 1206 and 1221(inclusive)
5. Perform z-score normalisation for all attributes other than Date

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

1. Replace the missing values of total\_bill column with its median and replace the missing values of day with mode.
2. Display the maximum and minimum tip given by the non-smokers.
3. Change the **day** of the second row to **Sat**.
4. Create a scatterplot for plotting total\_bill vs tip.
5. Delete the column size from the dataframe.