

# **Training Day-13 Report:**

## **Data Preprocessing in Python:-**

In order to derive knowledge and insights from data, the area of data science integrates statistical analysis, machine learning, and computer programming. It entails gathering, purifying, and converting unstructured data into a form that can be analysed and visualised. Data scientists process and analyse data using a number of methods and tools, such as statistical models, machine learning algorithms, and data visualisation software. Data science seeks to uncover patterns in data that can help with decision-making, process improvement, and the creation of new opportunities. Business, engineering, and the social sciences are all included in this interdisciplinary field.

## **Data Preprocessing:-**

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

## **Steps in Data Preprocessing:-**

Step 1: Import the necessary libraries.

Step 2: Load the dataset

Step 3: Statistical Analysis

Step 4: Check the outliers:

Step 5: Correlation

Step 6: Separate independent features and Target Variables

Step 7: Normalization or Standardization