MES Wadia College of Engineering Pune-01

Department of Computer Engineering

Name of Student:	**Class:**
Semester/Year:	**Roll No:**
Date of Performance:	**Date of Submission:**
Examined By:	**Experiment No: Part A-02**

PART: A) ASSIGNMENT NO: 02

Title: Data Wrangling-II

Create an "Academic performance" dataset of students and perform the following operations using Python.

1. Scan all variables for missing values and inconsistencies. If there are missing values and/or inconsistencies, use any of the suitable techniques to deal with them.
2. Scan all numeric variables for outliers. If there are outliers, use any of the suitable techniques to deal with them.
3. Apply data transformations on at least one of the variables. The purpose of this transformation should be one of the following reasons: to change the scale for better understanding of the variable, to convert a non-linear relation into a linear one, or to decrease the skewness and convert the distribution into a normal distribution.

Reason and document your approach properly.

OBJECTIVES:

- Students should be able to perform the data wrangling operation using Python on any open source dataset

PREREQUISITE:

- Basic of Python Programming
- Concept of Data Preprocessing, Data Formatting, Data Normalization and Data Cleaning.

APPRATUS:

- Programming Language: Python.
- Dataset: Kaggle Dataset (e.g. https://www.kaggle.com).

CONCLUSION:

QUESTIONS:

1. Explain the methods to detect the outlier.
2. Explain data transformation methods.
3. Write the algorithm to display the statistics of Null values present in the dataset.
4. Write an algorithm to replace the outlier value with the mean of the variable.