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-01**

PART: A) ASSIGNMENT NO: 01

Title: Data Wrangling-I

Perform the following operations using Python on any open source dataset (e.g., data.csv)

Import all the required Python Libraries.

1. Locate open source data from the web (e.g. https://www.kaggle.com).
2. Provide a clear description of the data and its source (i.e., URL of the web site).
3. Load the Dataset into the pandas data frame.
4. Data Preprocessing: check for missing values in the data using pandas insult(), describe() function to get some initial statistics. Provide variable descriptions. Types of variable etc. Check the dimensions of the data frame.
5. Data Formatting and Data Normalization: Summarize the types of variables by checking the data types (i.e., character, numeric, integer, factor, and logical) of the variables in the data set. If variables are not in the correct data type, apply proper type conversions.
6. Turn categorical variables into quantitative variables in Python.

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 data preprocessing in details
2. Explain Data Frame with Suitable example.
3. What is the limitation of the label encoding method?
4. What is the need of data normalization?
5. What are the different Techniques for Handling the Missing Data?