QUESTION BANK FOR LAB INTERNALS

1. A) What are the different functions used for data cleaning using pandas in python?
   B) Implement a program to do Data Cleaning and Data Manipulation Operations on the dataset 'datasetExample.csv'. Perform following operations:
   - Identify the datatype of each column
   - Check and remove all duplicate records
   - Check and remove all duplicate columns

2. A) How to work with missing data?
   B) Write a Python code to perform Data pre-processing task on given dataset 'pre-process_datasample.csv' (eg. Checking and Handling Missing Data, handling inappropriate data)

3. A) What is Hypothesis testing? Explain Null and Alternate Hypothesis.
   B) Implement the python code for hypothesis testing of following data:
   A sample of 25 individuals is taken, and their cholesterol levels are measured.
   Cholesterol Levels (mg/dL): 205, 198, 210, 190, 215, 205, 200, 192, 198, 205, 198, 202, 208, 200, 205, 198, 205, 210, 192, 205, 198, 205, 210, 192, 205.
   Populations Mean = 200
   Population Standard Deviation (σ): 5 mg/dL(given for this problem)

4. A) Explain Simple Linear Regression.
   B) Write a Python code to implement Simple Linear Regression on given data:

   x = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

   y = [1, 3, 2, 5, 7, 8, 8, 9, 10, 12]

5. A) What is Correlation Coefficient? Give the formula for Pearson Correlation Coefficient.
   B) Write a python program to compute correlation coefficient for given data:

   X = [15, 18, 21, 24, 27]

   Y = [25, 25, 27, 31, 32]

6. A) What are the steps to perform Hypothesis Testing?
   B) Implement the python code for hypothesis testing of following data:
   You are determining if the average age of 10 people is 30 or otherwise.