Computer Science and Engineering Department, SVNIT, Surat
M.Tech.- I DS (Semester - 1)
Foundations of Data Science
(CSDS111)

Lab Assignment: 4

Data Processing:

PART A:

- Select dataset(s) without missing values in text file format and use it for the following tasks:

1. Perform encoding techniques studied in the class on the datasets for the quantitative data and observe the range of data.
2. Perform encoding techniques studied in the class on the datasets for the qualitative data and observe the size of data in Bytes.
3. Display the columns and their respective count of rows the rows and columns with missing or null values and perform the following operations:
    a. Drop rows having missing values in various formats like blank, NULL, NA, etc.
    b. Drop the column if all the values are missing
    c. Drop rows that contain less than user-given X non-missing values
    d. Replace the missing value cells with the following and display mean, median, mode, variance, and SD. Also, prepare a description of which method is the best and why:
        i. Zeros
        ii. Minimum Value
        iii. Maximum Value
        iv. The mean of the column
        v. Variance
        vi. Standard deviation
    e. Replace the missing or null value cells with the following and display mean, median, mode, variance, and SD. Also, prepare a description of which method is the best and why:
        i. With the mean of 2 backward neighbors
        ii. With the mean of 2 forward neighbors
        iii. With the mean of 2 forward and 2 backward neighbors

PART B:

1. Write a code to input the following values and validate them:
    a. Email
    b. Indian Name
    c. Mobile number with and without country code
    d. Your Admission No.
2. Write a code to take the full address with country, state, city, and pincode from the user and validate them all.
3. Write a code to convert the following fields:
    a. DD/MM/YY format to MM/DD/YY
    b. Fee amount from Rs. to $
4. Without using a built-in function to write a code to calculate the Pearson's correlation of your selected dataset.
5. Without using a built-in function to write a code to calculate the Spearman's correlation of your selected dataset.

- Data Repositories
    UCI Repository
    Kaggle Dataset
    NASDAQ Dataset
    Google Public Dataset