MCA 5141 – Machine Learning Lab
Week - 2

EXERCISE 1: - Introduction to Pandas

Data frame creation and manipulation

1. Create a data frame with details of 10 students and columns as Roll Number, Name, Gender, Marks1, Marks2, Marks3.

2. Create a new column with total marks

3. Find the lowest marks in Marks1

4. Find the Highest marks in Marks2

5. Find the average marks in Marks3

6. Find student name with highest average

7. Find how many students failed in Marks2 (<40)

Exercise 2 – Data Preprocessing

Using the given **CEREALS** dataset, perform data preprocessing and answer the following questions.

1) Create a table with the 5-number summary of all the numeric attributes.

2) For each of the numeric attributes (proteins up to vitamins), identify and replace all missing data (indicated with -1) with the arithmetic mean of the attribute.

3) Create a table with the 5-number summary of all the numeric attributes after treating missing values. Do you think the strategy used in dealing with missing values was effective?

4) For each of the numeric attributes (proteins up to vitamins), identify and replace all noisy data with the median of attribute.

5) Create a table with the 5-number summary of all the numeric attributes after treating noisy values. Do you think the strategy used in dealing with noisy values was effective?