



Weekly Assignment - I

1. Write a python program to perform the following tasks:
 - a. Load the **Iris dataset** into a Pandas DataFrame.
 - b. Calculate the mean, standard deviation, minimum, and maximum values for each quantitative attribute in the Iris dataset.
 - c. Count the frequency of each distinct value in the qualitative attribute 'class' in the Iris dataset.
 - d. Display a summary of all attributes, including quantitative and qualitative, in the Iris dataset using a table.
 - e. Compute the covariance and correlation between pairs of attributes in the Iris dataset.
 - f. Display a histogram for the 'sepal length' attribute in the Iris dataset.
 - g. Create a boxplot to show the distribution of values for each attribute in the Iris dataset.
 - h. Create scatter plots to visualize the joint distribution of each pair of attributes in the Iris dataset.
 - i. Use parallel coordinates to display all the data points simultaneously in the Iris dataset.
2. Write a program to perform the following tasks:
 - a. Load the breast cancer dataset from the UCI machine learning repository using Pandas.
 - b. Replace missing values (encoded as '?') with NaNs in the breast cancer dataset.
 - c. Count the number of missing values in each column of the breast cancer dataset?
 - d. Replace missing values in the 'Bare Nuclei' column with the median value of that column.
 - e. Discard rows with missing values from the breast cancer dataset.

3. Write a program to perform the following tasks:
- a. Load the breast cancer dataset from the UCI machine learning repository using Pandas.
 - b. Draw a boxplot to identify columns with outliers in the breast cancer dataset.
 - c. Standardize the columns of the breast cancer dataset to compute the Z-scores.
 - d. Identify and count duplicate rows in the breast cancer dataset.
 - e. Remove duplicate rows from the breast cancer dataset.
 - f. Randomly select a sample of size 3 (without replacement) from the breast cancer dataset.
 - g. Randomly select 1% of the data (without replacement) from the breast cancer dataset.
 - h. Perform sampling with replacement to create a sample whose size is equal to 1% of the entire breast cancer dataset.