FOUNDATIONS OF DATA SCIENCE (CSD355)
LAB ASSIGNMENT - 3
DATA VISUALIZATION
(Deadline: 11-09-2025)

Download the Iris dataset from the given link - https://www.kaggle.com/uciml/iris/version/2

The Iris dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems, and can also be found on the UCI Machine Learning Repository.
It includes three iris species with 50 samples each, as well as some properties of each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.
The columns in this dataset are:
- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species

After downloading the .csv file, load its contents into a DataFrame in a Python notebook and perform preprocessing techniques as done in Lab 2, i.e., check for and handle null values if any. Now, perform the following operations:

Now, perform the following operations:

1) Print the head of the dataframe to see what the data looks like.
2) Check all the types of variables that you are dealing with and make sure they are all the appropriate types so that future mathematical operations on the data are seamless.
3) Find how many flowers of each type of species are present in the dataframe.
4) Compute basic statistics for all continuous variables in the dataframe using the describe function.
5) Find all the unique values of lengths in 'PetalWidthCm' and comment on the range of lengths in the column.
6) Copy columns 'PetalWidthCm' and 'PetalLengthCm' into another dataframe. Then, find the average 'PetalLengthCm' for each unique value of 'PetalWidthCm'.
7) Plot every variable in the dataframe against every other variable using pairplots.

8) Create a 3D scatter plot of the Iris dataset using matplotlib. Plot SepalLengthCm, SepalWidthCm, and PetalLengthCm on the three axes, and use different colors to distinguish the three Iris species. Add appropriate axis labels, a title, and a legend.

9) Plot histograms of SepalLengthCm for each species separately.

10) Find the mean, median, and mode of SepalLengthCm for the entire dataset.

11) Calculate the mean petal length (PetalLengthCm) separately for each species of Iris. Which species has the longest average petal length?