Assignment I.a: Dateset Familiarization Using Python

q1. Python program to load the iris data from a given CSV file into a data frame (data_iris)

- print the shape of the data (number of samples, number of columns)
- type of the data
- first five rows with labels.
- find the keys
- find number of features and feature names
- description of the Iris data.
- number of categories/class labels in the data
- get the missing values and nan values (with feature and class label indexing).
- view basic statistical details like percentile, mean, std etc. of iris data

Assignment I.a: Dateset Familiarization Using Python

q1. continues...

- get number of observations of each species (setosa, versicolor, virginica) from iris data and create sub data frames for each species (data_iris_setosa, data_iris_versicolor, data_iris_virginica)
- drop Id column from data_iris Dataframe and create a new data frame with this modified part
- create two date frames data_iris_feature (only with feature columns) and data_iris_species (only with species column).
- create a Bar plot and a Pie plot to get the frequency of the three species of the Iris data
- create a graph to find the relationship between the sepal length and sepal width of different species (scatter plot , give different color labels for different species)

Assignment I.a: Dateset Familiarization Using Python

q1. continues...

- create a graph to see how the feature (SepalLength, SepalWidth, PetalLength, PetalWidth) are distributed for each species. Comment on it.
- create a pairplot of the iris data set and check which flower species seems to be the most separable. Comment on it.
- find the correlation between variables of iris data (between features and between feature and species). Also create a heatmap using Seaborn to present their relations. Comment on it.