Assignment 10:

Title of the Assignment:

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the inference as:

1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a boxplot for each feature in the dataset.
4. Compare distributions and identify outliers.

Objective of the Assignment: Students should be able to understand the Seaborn library with matplotlib using Python on any open source dataset.

Prerequisite:

1. Basic of Python Programming
2. Concept of statistics such as mean, median, minimum, maximum, standard deviation etc.

Theory:

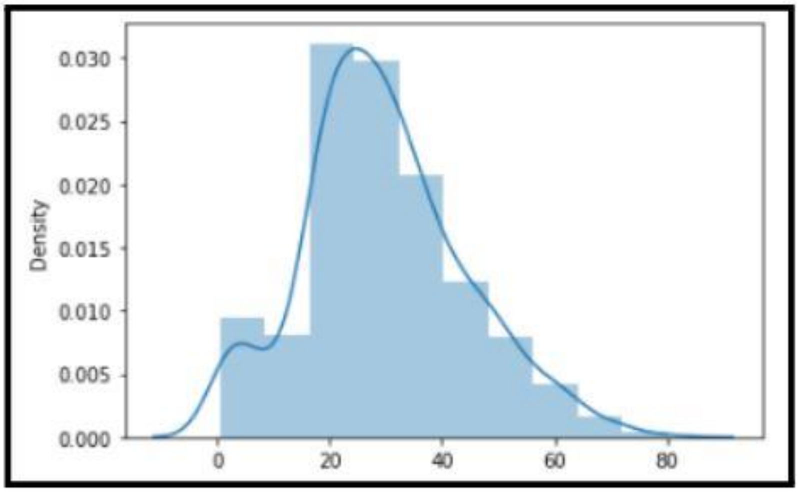
Distribution Plots:

These plots help us to visualise the distribution of data. We can use these plots to understand the mean, median, range, variance, deviation, etc of the data.

a.Distplot

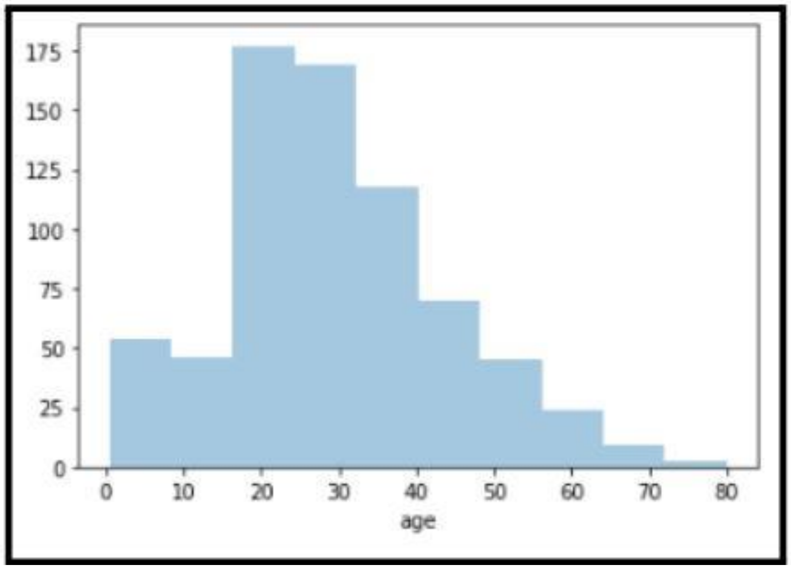
- Dist plot gives us the histogram of the selected continuous variable.
- It is an example of a univariate analysis.
- We can change the number of bins i.e. number of vertical bars in a histogram

```
import seaborn as sns
sns.distplot(x = dataset['age'], bins = 10)
```



- The line that you see represents the kernel density estimation. You can remove this line by passing False as the parameter for the kde attribute as shown below

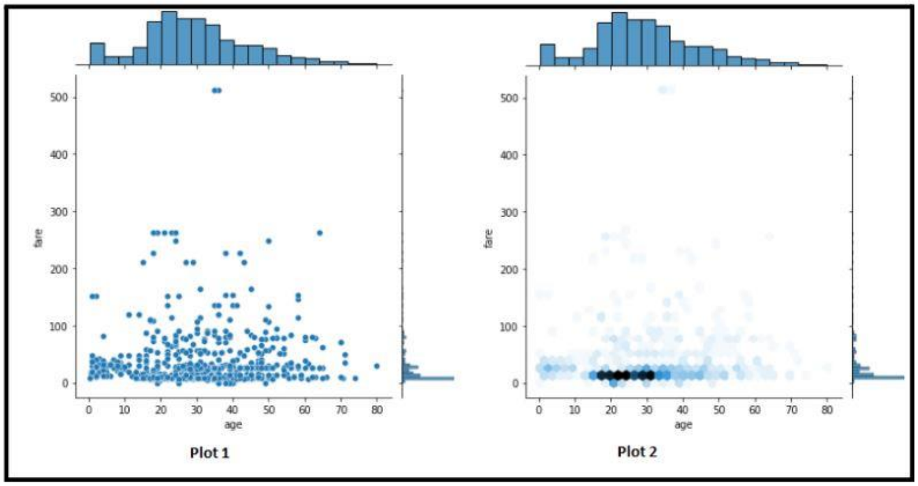
```
sns.distplot(dataset['age'], bins = 10,kde=False)import seaborn as sns
sns.distplot(x = dataset['age'], bins = 10)
```



- Here the x-axis is the age and the y-axis displays frequency. For example, for bins = 10, there are around 50 people having age 0 to 10.

b. Joint Plot

- It is the combination of the distplot of two variables.
- It is an example of bivariate analysis.
- We additionally obtain a scatter plot between the variables to reflect their linear relationship. We can customise the scatter plot into a hexagonal plot, where, the more the colour intensity, the more will be the number of observations.
        - import seaborn as sns
    - # For Plot 1
        - sns.jointplot(x = dataset['age'], y = dataset['fare'], kind ='scatter')
    - # For Plot 2
        - sns.jointplot(x = dataset['age'], y = dataset['fare'], kind = 'hex')



Plot 1

Plot 2

Assignment Questions:

1. Explain Rug plot with an example.
2. Draw inferences from the example shown in 1..