Assignment No 8:

Title: **Data Visualization I**

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.
2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

Theory: -

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions

The advantages and benefits of good data visualization:

Data visualization is another form of visual art that grabs our interest and keeps our eyes on the message. When we see a chart, we quickly see trends and outliers. If we can see something, we internalize it quickly.

Patterns of data can be find out with the help of different types of plots

Types of plots are:

A. Distribution Plots

a. Dist-Plot

b. Joint Plot

d. Rug Plot

B. Categorical Plots

a. Bar Plot

b. Count Plot

c. Box Plot

d. Violin Plot

C. Advanced Plots

a. Strip Plot

b. Swarm Plot

D. Matrix Plots

a. Heat Map

b. Cluster Map

A. Distribution Plots:

These plots help us to visualize 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

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.

c. The Rug Plot

The rugplot() is used to draw small bars along the x-axis for each point in the dataset. To plot a rug plot, you need to pass the name of the column. Let's plot a rug plot for fare.

Algorithm:

Step 1: Download the data set of Titanic (https://www.kaggle.com/uciml/iris)

Step 2: Importing Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
dataset = sns.load_dataset('titanic')
dataset.head()
```

Step 3: Draw distributional plot

```
sns.distplot(dataset['fare'])
```

Step 4: Removal of Kernal Density line

```
sns.distplot(dataset['fare'], kde=False)
```

Step 5: Draw histogram

```
sns.histplot(data=dataset, x="fare",binwidth=30)
```

Step 6: The Joint Plot

```
sns.jointplot(x='age', y='fare', data=dataset)
```

Step 7: The Pair Plot

```
sns.pairplot(dataset)
```

Question: Draw Histogram, Joint Plot.

Conclusion: Implemented successfully Simple Data visualization techniques using Python on Titanic dataset.