Assignment 08:

Title of the Assignment:

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. Provide the codes with outputs and explain everything that you do in this step.

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:

Data Visualisation plays a very important role in Data mining. Various data scientists spent their time exploring data through visualisation. To accelerate this process we need to have a well-documentation of all the plots.
Even plenty of resources can't be transformed into valuable goods without planning and architecture

1. Seaborn Library Basics

Seaborn is a Python data visualisation library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
For the installation of Seaborn, you may run any of the following in your command line.

```
pip install seaborn
conda install seaborn
```

To import seaborn you can run the following command.

```
import seaborn as sns
```

2. Know your data

The dataset that we are going to use to draw our plots will be the Titanic dataset, which is downloaded by default with the Seaborn library. All you have to do is use the load_dataset function and pass it the name of the dataset.
The dataset contains 891 rows and 15 columns and contains information about the passengers who boarded the unfortunate Titanic ship. The original task is to predict whether or not the passenger survived depending upon different features such as their age, ticket, cabin they boarded, the class of the ticket, etc. We will use the Seaborn library to see if we can find any patterns in the data.

3. Finding patterns of data.

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

Assignment Questions:

1. List out different types of plot to find patterns of data
2. Explain when you will use distribution plots and when you will use categorical plots.
3. Write the conclusion from the following swarm plot (consider titanic dataset).
4. Which parameter is used to add another categorical variable to the violin plot,Explain with syntax and example.