Practical No. 9:

	Guru Gobind Singh Foundation's Guru Gobind Singh College of Engineering and Research Centre, Nashik	

Experiment No: 09

Title of the Assignment: Data Visualization II

1. Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not. (Column names : 'sex' and 'age')

2. Write observations on the inference from the above statistics.

Student Name:	
Class:	TE (Computer)

Div:		Batch:	

Roll No.:	
Date of Attendance (Performance):	
Date of Evaluation:	
Marks (Grade) Attainment of CO Marks out of 10	
CO Mapped	
Signature of Subject Teacher	

Group A

Assignment No: 9

Objective of the Assignment: Students should be able to perform the data Visualization operation using Python on any open source dataset

----

Prerequisite:

1. Basic of Python Programming

2. Seaborn Library, Concept of Data Visualization.

An introduction to seaborn

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures.

Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

Here's an example of what seaborn can do:

```python
# Import seaborn
import seaborn as sns

# Apply the default theme
sns.set_theme()

# Load an example dataset
tips = sns.load_dataset("tips")

# Create a visualization
sns.relplot(
    data=tips,
    x="total_bill", y="tip", col="time",
    hue="smoker", style="smoker", size="size",
)
```

A few things have happened here. Let's go through them one by one:

```
# Import seaborn
import seaborn as sns
```

Seaborn is the only library we need to import for this simple example. By convention, it is imported with the shorthand sns.

Behind the scenes, seaborn uses matplotlib to draw its plots. For interactive work, it's recommended to use a Jupyter/IPython interface in matplotlib mode, or else you'll have to call **matplotlib.pyplot.show()** when you want to see the plot.

```
# Apply the default theme
sns.set_theme()
```

This uses the matplotlib rcParam system and will affect how all matplotlib plots look, even if you don't make them with seaborn. Beyond the default theme, there are several other options, and you can independently control the style and scaling of the plot to quickly translate your work between presentation contexts (e.g., making a version of your figure that will have readable fonts when projected during a talk). If you like the matplotlib defaults or prefer a different theme, you can skip this step and still use the seaborn plotting functions.

```
# Load an example dataset
tips = sns.load_dataset("tips")
```

Most code in the docs will use the **load_dataset()** function to get quick access to an example dataset. There's nothing special about these datasets: they are just pandas dataframes, and we could have loaded them with **pandas.read_csv()** or built them by hand. Most of the examples in the documentation will specify data using pandas dataframes, but seaborn is very flexible about the data structures that it accepts.

```
# Create a visualization
sns.relplot(
    data=tips,
    x="total_bill", y="tip", col="time",
    hue="smoker", style="smoker", size="size",
)
```

Assignment Questions

1. Write down the code to use inbuilt dataset 'titanic' using seaborn library. 2. Write code to plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not.

3. Write the observations from the box plot.