Sinhgad Institutes | Date

Name: Pallavi Vijay Patil
Roll No.: COTB028
Subject: DSBDA

Assignment No. 8.

- **Aim:** Data Visualization I:
  i) Use inbuilt dataset 'titanic'. The dataset contains 891 rows and contains info about passengers who boarded unfortunate Titanic ship. Use seaborn library to see if we can find any patterns in data.
  ii) Write a code to check how the price of ticket by each passenger is distributed by plotting histogram.

- **Introduction:**
  Seaborn is extremely useful library for data visualization in Python. The seaborn library is built on top of Matplotlib and offers many advanced data visualization capabilities.
  Though, the seaborn library can be used to draw variety of charts such as matrix plots, grid plots, regression plots, etc.

- **Distributional plots:**
  Distributional plots, as the name suggests are types of plots that show the statistical distribution of data. In this section we will see some of most commonly used distribution plots seaborn.

- The Dist Plot:
The distplot() shows the histogram distribution of data for a single column. The column name passed as parameter to displot() function.

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

- The Joint Plot:
The jointplot() is used to display the mutual distribution of each column. You need to pass three parameters to jointplot(). The first parameter is the column name for which you want to display the distribution of data on x-axis. The second parameter is the column name for which you want to display distribution of data on y-axis.

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

- The Rug Plot:
The rugplot() is used to draw small bars along x-axis for each point in the dataset to plot a rug plot, we need to pass the name of column.

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

- Categorical Plots:
As name suggests are normally used to

plot categorical data. The categorical plots plot the values in the categorical column / against another categorical column / numeric column.

- The Bar Plot:

The barplot() is used to display the mean value for each value in a categorical column, against a numeric column. The first parameter is the categorical column, second parameter is numeric column, while third parameter is dataset.

```
sns.barplot(x='gender', y='age',
            data=dataset)
```

- The Count Plot():

The countplot() is similar to barplot(), however it display the count of categories in specific column. For instance, if we want to count the no. of males and women passenger we can do so using count as follows:

```
sns.countplot(x='gender', y=data=
              dataset)
```

- The Box Plot():

The boxplot() is used to display distribution of categorical data in the form of quartiles.

The center of box shows the median value. The value from lower whisker to bottom of box shows first quartiles. From bottom of box to middle box lies second quartile.

```
sns.boxplot(x='gender', y='age', data=dataset)
```

- The Violin plot:
The violinplot() is similar to boxplot(), however the violin plot allows us to display all components that actually correspond to data point. It is used to plot violin plot. Like the box plot, the

```
sns.violinplot(x='gender', y='age', data=dataset)
```

- The Strip Plot:
The strip plot draws a scatter plot where one of the variables is categorical. We have 2 numeric variables.
The stripplot() fun' is used to plot violin plot.

```
sns.stripplot(x='sex', y='age', data=dataset)
```

- Combining Swarm and Violin Plots:
Swarm plots are not recommended if you have huge dataset since they have to

each data point. If you really like swarm plot, a better way is to combine two plots.

```
sns.violinplot(x='sex', y='gender', data=
        dataset)
sns.swarmplot(x='sex', y='age', data=
    dataset, color='black')
```

- Conclusion:

Seaborn is an advanced data visualization library built on top of Matplotlib library. In this practical, we looked at how we can draw distributional and categorical plots using seaborn library.