Assignment No. 8

Name - Kuldharan Sumit Dattatraya
Class - TE
Div - 4
Subject - DSBDAL

Problem statement -

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows & 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 -

1) Introduction to data visualization.

→ Data visualization has two terms, data & visualization. Data means information & visualization means pictorial or graphical visualization. So the data visualization term is defined as the pictorial representation of some information so that the user can analyze data quickly.

Data finding with visualization-based method provides the ability to merge different data from different source to make various customized analytical views.

2) Types of Data Visualization.

→
1. Table
2. Histogram
3. Scatter Plot
4. Various charts
5. Timeline
6. Various diagrams

① Table -

Data is represented in collection of rows & columns.

② Histogram -

A huge data in histogram is represented in vertical bars.

③ Scatter plot -

Scatter plots are also known as X-Y plots. It is used to represent correlation & distribution analysis.

④ Various chart -

We can visualize data in various charts such as Line chart, Bar chart, Pie chart, Area chart, Flow chart, Bubble chart.

⑤ Timeline -

A Timeline is a pictorial representation of a series of events in chronological sequence along with drawing straight line.

⑥ Various diagrams -

We can represent data in various diagrams such as Venndiagram, Data flow diagram, Entity relationship diagram.

3) Explain histogram with example.

→ A histogram is basically used to represent data provided in a form of some groups. It is accurate method for the graphical representation of numerical data distribution. It is a type of bar plot where X-axis represents the bin ranges while Y-axis gives information about frequency.

Example -

```
from matplotlib import pyplot as plt
import numpy as np
a= np.array([22,87,5,43, 56, 73,55, 54,11, 20,51,5,79,
             31 ,27])
fig,ax = plt.subplots(figsize=(10,7))
ax.hist(a, bins = [0,25, 50,75, 100])
plt.show()
```

4) Explain hist() with different parameters used.

→ Following parameters accepted by matplotlib.pyplot.hist() function-

1) x - array or sequence of array
2) bins - optional parameter contains integer or sequence or strings
3) density - optional parameter contains boolean value.
4) range - optional parameter represents upper & lower range of bins.
5) histtype - optional parameter used to create type of histogram [bar, barstacked, step, stepfilled], default is "bar".
6) align - optional parameter controls the plotting of histogram [left, right, mid]
7) weights - optional parameter contains array of weights having same dimensions as X.
8) bottom - location of the baseline of each bin.
9) rwidth - optional parameter which is relative width of the bars with respect to bin width.
10) color - optional parameter used to set color or sequence of color specs.
11) label - optional parameter string or sequence of string to match with multiple datasets.
12) log - optional parameter used to set histogram axis on log scale.

conclusion-

① used seaborn library to find patterns in the titanic dataset.
② plotted histogram for pair column.