PICT, PUNE

Aarkin Karnik
31140

DSBDAL

Assignment - 9

- Title: Data Visualization II

1) Use the inbuilt dataset 'titanic' as used in 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.
2) Write Observation on the inference from the above statistics.

- Learning Objectives:

i) To be able to import an inbuilt dataset.
ii) To be able to use the seaborn library to plot various graphs.

- Learning Outcomes:

i) Student will be able to import the seaborn library.
ii) Student will be able to understand the use of various graphs & plot them.

- Hardware & software requirements:

i) Jupyter Notebook
ii) Python 3.8
iii) Ubuntu 20.4

PICT, PUNE

31140

- Theory:

Data Visualization is the graphical representation of information & data by using visual elements like charts, graphs & maps. Data Visualization tools provide an accessible way to see & understand trends, outliers & patterns in data.

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed & modeled. It must be visualized that conclusion to be made.

- Data set used: Titanic dataset

The dataset contains 891 rows & contains information about the passengers who boarded the ship. The dataset is available within the in-built dataset of seaborn library. The dataset contains features such as age, gender, class, alive, whether they survived or not.

- Seaborn library:

Seaborn is an amazing visualization library for satisfied graphics plotting in python. It provideds beautiful default styles & colour pallete to make satisfied graphs more attractive.

PICT, PUNE

811140

- Types of graphs used:

1) Bar plot:
A bar plot shows categorical data as rectangular bars with the height of bars proportional to the value they represent. It is often used to compare between values of different categories in the data.

2) Box plot:
A box plot is created to display the summary of the set of data values having properties like minimum, first quartile, median, third quartile & maximum.

3) Strip plot:
It is basically a scatter plot where the x-axis represents a categorical variable.

4) Swarm plot:
A swarm plot is a type of scatter plot that is used for representing categorical values. It is similar to the strip plot, but it avoids the overlapping of points.

5) Violin plot:
A violin plot is a hybrid of a box plot & a kernel density plot, which shows peaks in data. It is used to visualize the distribution of

numeric data. Unlike a boxplot that can only show summary statics, violin plots depict summary statics & density of each variable.

- Analysis:
From the graphs, we can conclude that in male, the age of people who survived is around 20.

The median age of male & female who survived is around 27.

The median age of people who didn't survive is around 29.

- Conclusion:
We were able to understand the use of graphs such as violin plot, swarm plot, box plot, strip plot. & We were also successfully able to plot this graphs for titanic dataset using seaborn library.