Assignment No:- 5

Problem Statement:

Visualize the data using Python by plotting the graphs for assignment no. 1 and 2. Consider a suitable data set.
a) Use Scatter plot, bar plot, Box plot and Histogram
OR
b) Perform the data visualization operations using Tableau for the given dataset.

Software Library Package:

For this task, we'll use the following Python libraries:

- `pandas` for data manipulation.
- `matplotlib` and `seaborn` for data visualization.

Theory:

i) Methodology:

- Scatter Plot: A scatter plot is used to visualize the relationship between two variables. Each point represents an observation in the dataset, with one variable on the x-axis and the other on the y-axis.

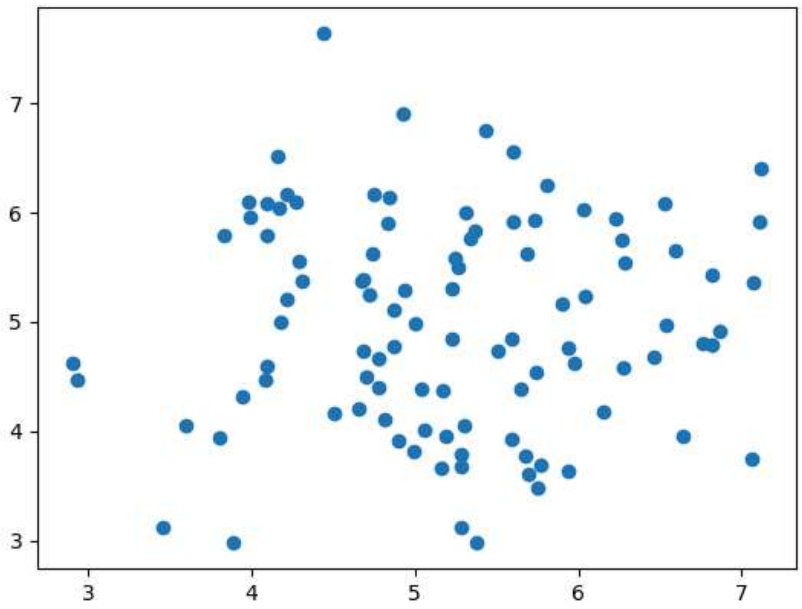

Fig 1: Scatter Plot

- Bar Plot: A bar plot is used to visualize the distribution of a categorical variable or the relationship between a categorical variable and a numerical variable. It represents the frequency or mean of each category.

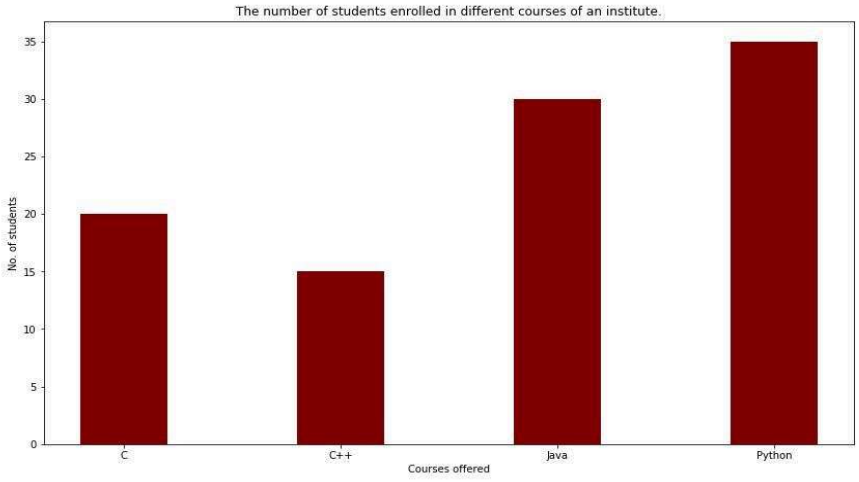

Fig 2: Bar Plot

- Box Plot: A box plot is used to visualize the distribution of a numerical variable and identify outliers. It shows the quartiles of the dataset, including the median, first quartile, third quartile, and outliers.

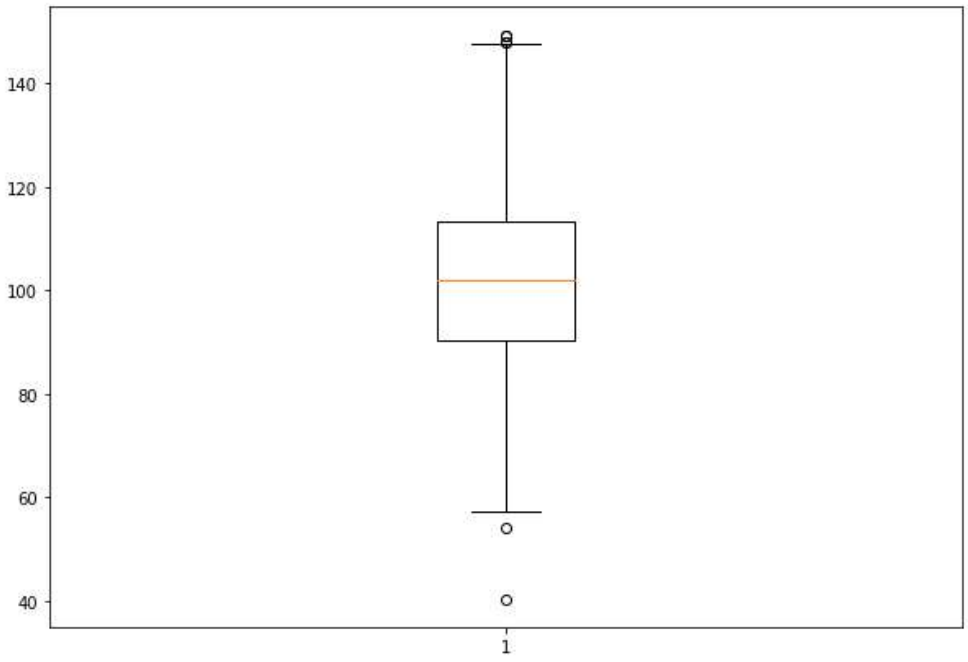

Fig 3: Box PLot

- Histogram: A histogram is used to visualize the distribution of a numerical variable. It divides the range of values into bins and shows the frequency of observations in each bin.

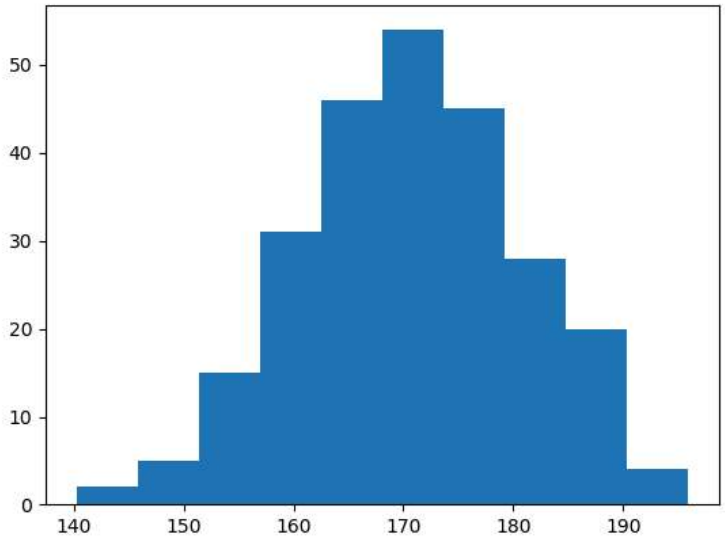

Fig 4: Histogram

ii) Advantages and Applications and Limitations/Example:

- Advantages: Python libraries such as `matplotlib` and `seaborn` offer a wide range of customizable plots for effective data visualization. Visualization enhances data exploration, analysis, and communication of insights to stakeholders.
- Applications: Data visualization is widely used in various domains such as finance, healthcare, marketing, etc., for exploratory data analysis, presentation of results, and decision-making.
- Limitations/Example: Misleading visualizations or incorrect interpretation of plots can lead to incorrect conclusions. It's important to choose appropriate visualization techniques and ensure clarity in conveying insights.

Working/Algorithm:

- Load the dataset into Python using `pandas`.
- Use `matplotlib` and `seaborn` to create scatter plots, bar plots, box plots, and histograms based on the data variables.
- Customize the plots as needed to improve readability and convey insights effectively.

Diagram:

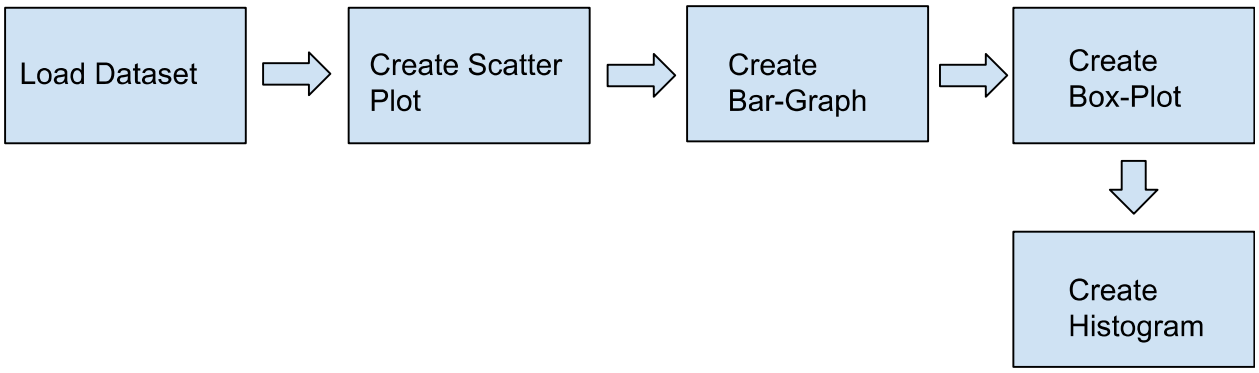

Fig 5: Workflow Diagram

Conclusion:

Data visualization using Python libraries such as `matplotlib` and `seaborn` enables effective exploration and communication of insights from the dataset. By leveraging scatter plots, bar plots, box plots, and histograms, analysts can gain deeper insights into the data distribution, relationships between variables, and identify patterns and outliers. It's important to choose appropriate visualization techniques and customize plots to effectively convey insights to stakeholders.