ASSIGNMENT NO. 5

Title: Visualize the data using R/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.

S/W Packages and H/W apparatus used: Linux OS: Ubuntu/Windows ,

Jupyter notebook.

Theory/Methodology:

Data visualization is a critical aspect of data analysis, allowing us to explore relationships, patterns, and distributions within the data. In this practical, we'll utilize Python libraries like Matplotlib and Seaborn to create visualizations such as scatter plots, bar plots, box plots, and histograms.

Scatter Plot:

Purpose: A scatter plot displays the relationship between two continuous variables. Each point on the plot represents a single observation.

Usage: Scatter plots are useful for visualizing patterns, trends, and correlations between variables. They help identify relationships such as linear, non-linear, positive, negative, or no correlation.

Insights: Scatter plots can reveal clusters of data points, outliers, and the overall distribution of the data.
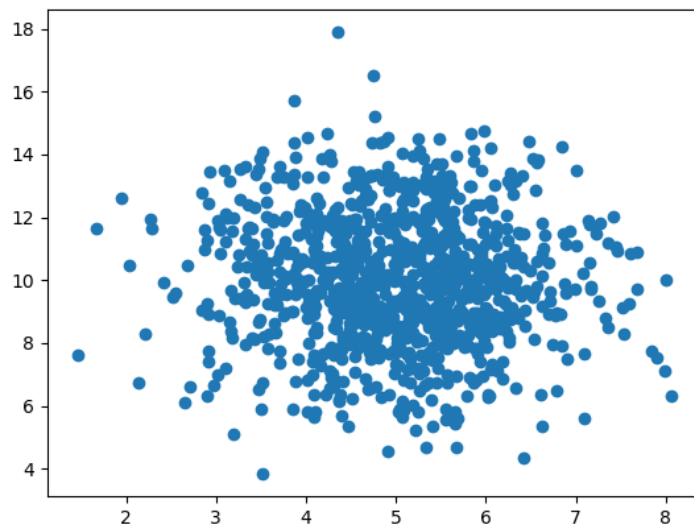
Bar Plot:

Purpose: A bar plot displays the distribution of categorical variables or the relationship between a categorical variable and a continuous variable.

Usage: Bar plots are commonly used to compare the frequency or proportion of different categories. They can also show the mean or median value of a continuous variable across different categories.

Insights: Bar plots make it easy to compare values across categories, identify dominant categories, and visualize trends or patterns in categorical data.
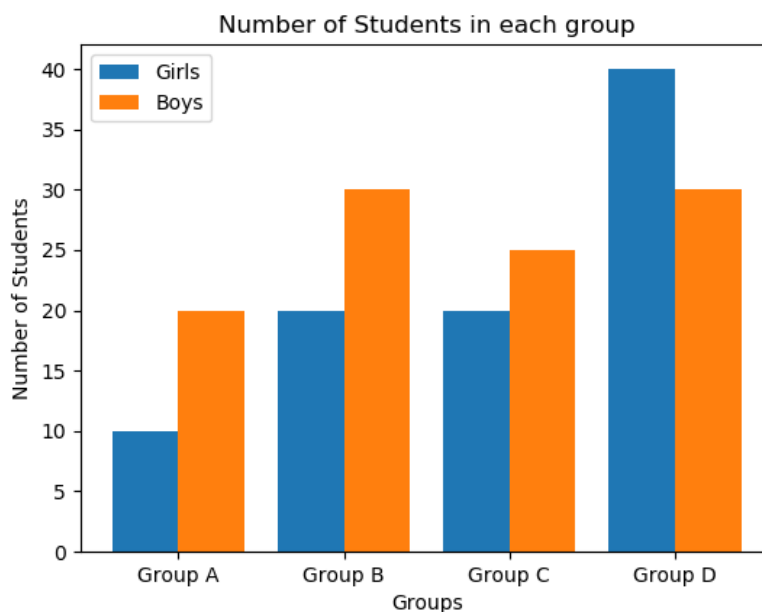

Box Plot (Box-and-Whisker Plot):

Purpose: A box plot displays the distribution of a continuous variable and highlights important summary statistics such as the median, quartiles, and outliers.

Usage: Box plots are useful for visualizing the spread and variability of data, detecting outliers, and comparing distributions between different groups or categories.

Insights: Box plots provide a visual summary of the central tendency, spread, and skewness of the data distribution. They help identify variability, symmetry, and the presence of extreme values.
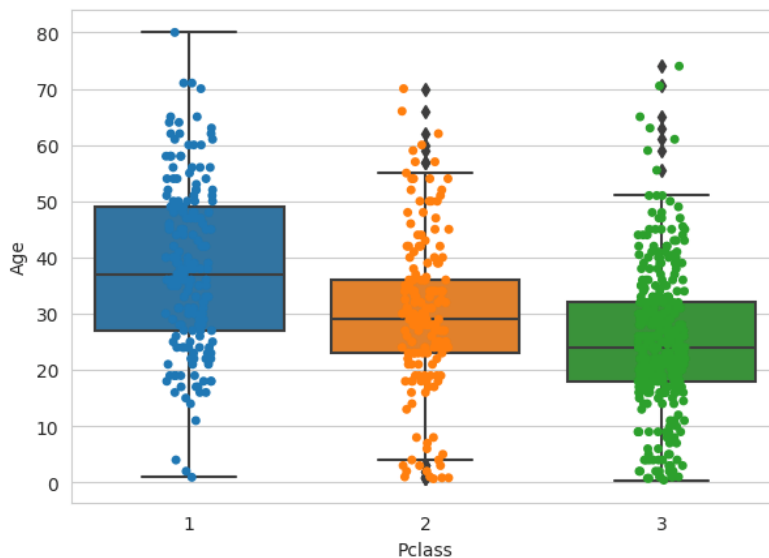

Histogram:

Purpose: A histogram displays the distribution of a single continuous variable by dividing the data into bins and counting the frequency of observations in each bin.

Usage: Histograms are used to visualize the shape, central tendency, and spread of a data distribution. They help identify patterns such as normality, skewness, multimodality, or outliers.

Insights: Histograms provide insights into the frequency and density of data values across the range of the variable. They are particularly useful for exploring the overall distribution and detecting deviations from expected patterns.
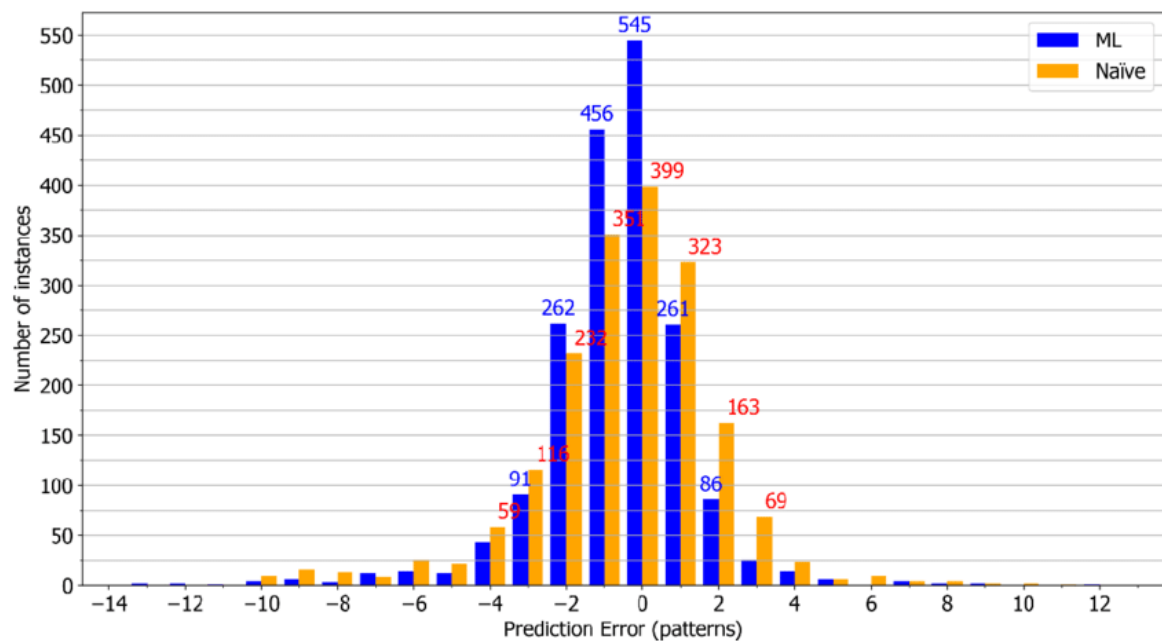
Advantages:

Feature Analysis:

Visualizations help in understanding the relationships between input features and the target variable. Scatter plots can reveal linear or nonlinear relationships, while box plots can highlight differences in distributions between classes or categories.

Model Evaluation:

Visualizations aid in evaluating model performance. For example, histograms of model predictions can reveal if the model is biased towards certain predictions or if it is well-calibrated.

Model Diagnostics:

Box plots and scatter plots can help diagnose issues with models, such as heteroscedasticity (unequal variance) or outliers in residuals.

Working/Algorithm::

1. Scatter Plot: Display the relationship between two numerical variables using points on a 2D plane.

2. Bar Plot: Represent categorical data with rectangular bars, where the length of each bar corresponds to the value of a variable.

3. Box Plot: Visualize the distribution of a numerical variable through quartiles, median, and outliers.

4. Histogram: Display the distribution of a numerical variable by dividing the data into intervals (bins) and counting the number of observations in each bin.

Conclusion:

By utilizing Python libraries like Matplotlib and Seaborn, we can create informative visualizations to explore and communicate insights from the data effectively. Each type of plot offers unique advantages in representing different aspects of the data, allowing for a comprehensive analysis. However, it's essential to choose the appropriate visualization techniques based on the nature of the data and the insights we aim to convey.