**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.

**S/W Packages and Libraries used:**

For the following assignment, the interpreter used was Google Collab and the Primary Library used was

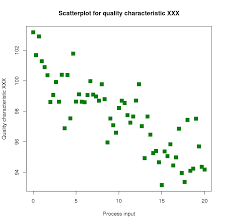
* Seaborn: Seaborn is a statistical data visualization library in Python that provides a high-level interface for creating attractive and informative visualizations. It works well with Pandas DataFrames and offers a variety of plot types and customization options.

**Theory-**

* Data Selection:
  + Choose a suitable dataset that contains numerical and categorical variables, allowing for the creation of different types of plots such as scatter plots, bar plots, box plots, and histograms.
* Data Preprocessing:
  + Load the dataset into Python using Pandas.
  + Perform any necessary data cleaning and preprocessing steps, such as handling missing values or encoding categorical variables.
* Visualization:
* Utilize the Seaborn library for creating visualizations.
* Create scatter plots to visualize the relationship between two numerical variables

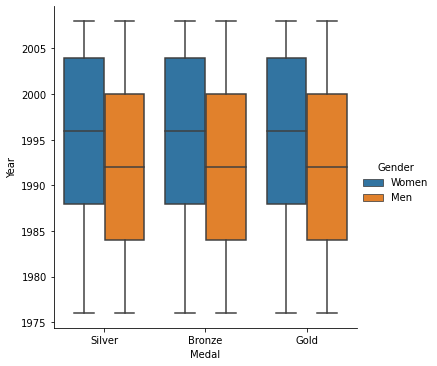
1. **Scatter Plot:**

Utilize scatter plots to visualize the relationship between two continuous variables. Each point on the plot represents an observation in the dataset.

****

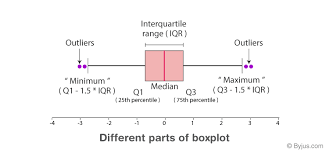
1. **Bar Plot:**

Bar plots are effective for comparing categorical data by displaying the frequency or distribution of each category using rectangular bars.



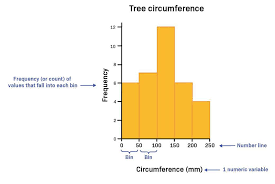
1. **Box Plot:**

Box plots provide a visual summary of the distribution of numerical data, showing the median, quartiles, and potential outliers.



1. **Histogram:**

Histograms are graphical representations of the frequency distribution of a continuous variable, displaying the data's distribution across intervals or bins.



**Advantages and Applications:**

* Data visualization helps in understanding patterns, trends, and relationships within the data, facilitating exploratory data analysis and decision-making.
* Visualizations serve as powerful tools for conveying complex information in a clear and concise manner, facilitating communication of insights to stakeholders.
* By providing visual representations of data, decision-makers can make informed decisions based on a deeper understanding of the underlying trends and patterns.
* Scatter plots are useful for identifying correlations between variables, while bar plots aid in comparing categories.
* Box plots offer insights into the spread and central tendency of data, while histograms provide an overview of data distribution.

**Limitations with Example:**

* **Overplotting in Scatter Plots**: When there are too many data points in a scatter plot, they may overlap, leading to overplotting. This can obscure patterns and make it difficult to discern relationships between variables.

**Conclusion:**

By following the outlined methodology and utilizing the Seaborn library, you can effectively visualize the selected dataset using scatter plots, bar plots, box plots, and histograms. These visualizations will help in exploring relationships, understanding distributions, and identifying patterns within the data, thereby facilitating further analysis and decision-making processes.