**ASSIGNMENT: - 05**

**Problem Statement: -**

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, Library and Package:**

1. Software:

* Python
* Jupyter Notebook (for interactive visualization within Jupyter)

1. Libraries:

* pandas: For data manipulation and handling
* matplotlib: For creating visualizations such as scatter plot, bar plot, box plot, and histogram
* seaborn (optional): For enhanced statistical visualizations

1. Packages:

* NumPy: For numerical computations and array operations

**Theory:**

Methodology:

1. Scatter Plot: A scatter plot is used to visualize the relationship between two variables. Each point on the plot represents a data point with values for both variables.
2. Bar Plot: A bar plot is used to compare categorical data by showing the value of each category as bars of different heights.
3. Box Plot: A box plot (box-and-whisker plot) is used to display the distribution of a dataset and identify outliers, quartiles, and median.
4. Histogram: A histogram is used to represent the frequency distribution of continuous data by dividing it into bins and showing the number of data points in each bin.

Advantages:

1. Scatter Plot: Shows patterns and relationships between variables.
2. Bar Plot: Effective for comparing categories and displaying trends.
3. Box Plot: Provides insights into the spread and central tendency of data.
4. Histogram: Visualizes data distribution and helps identify data skewness.

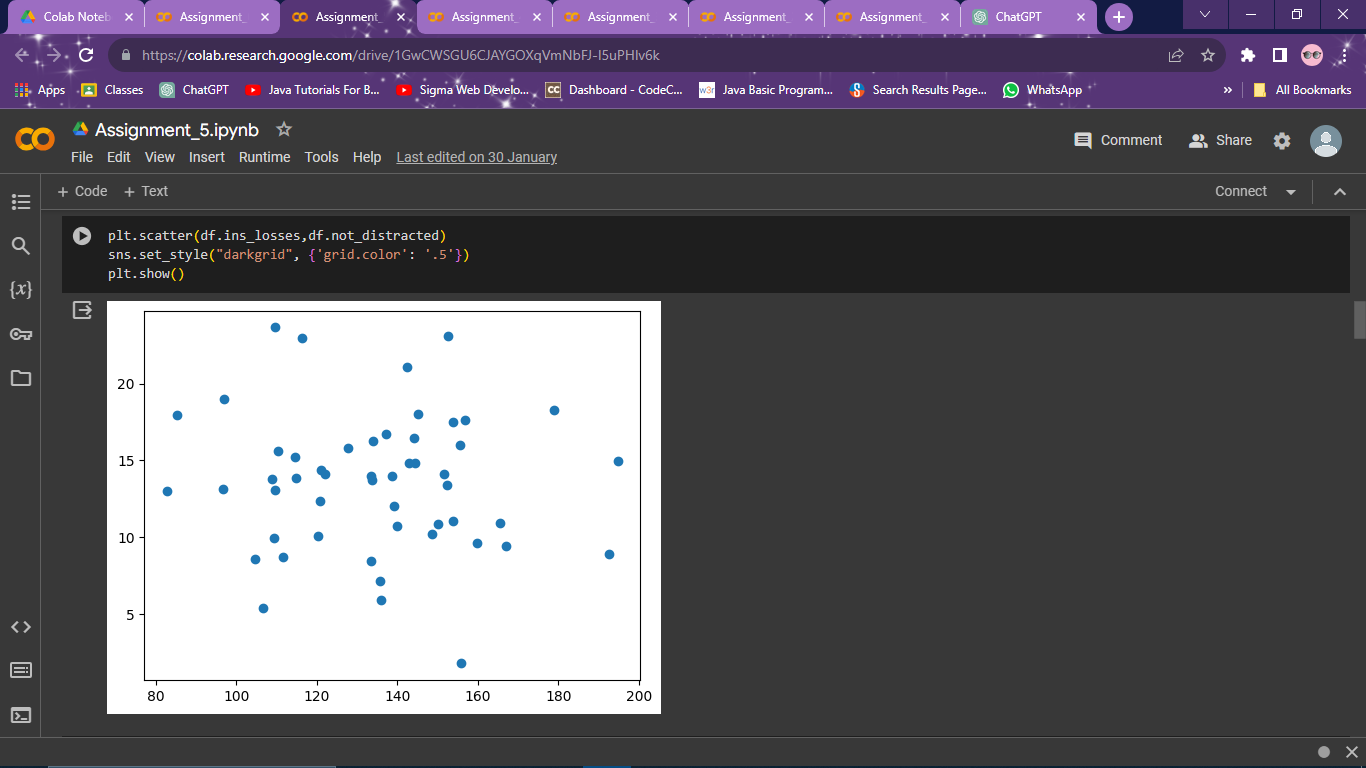
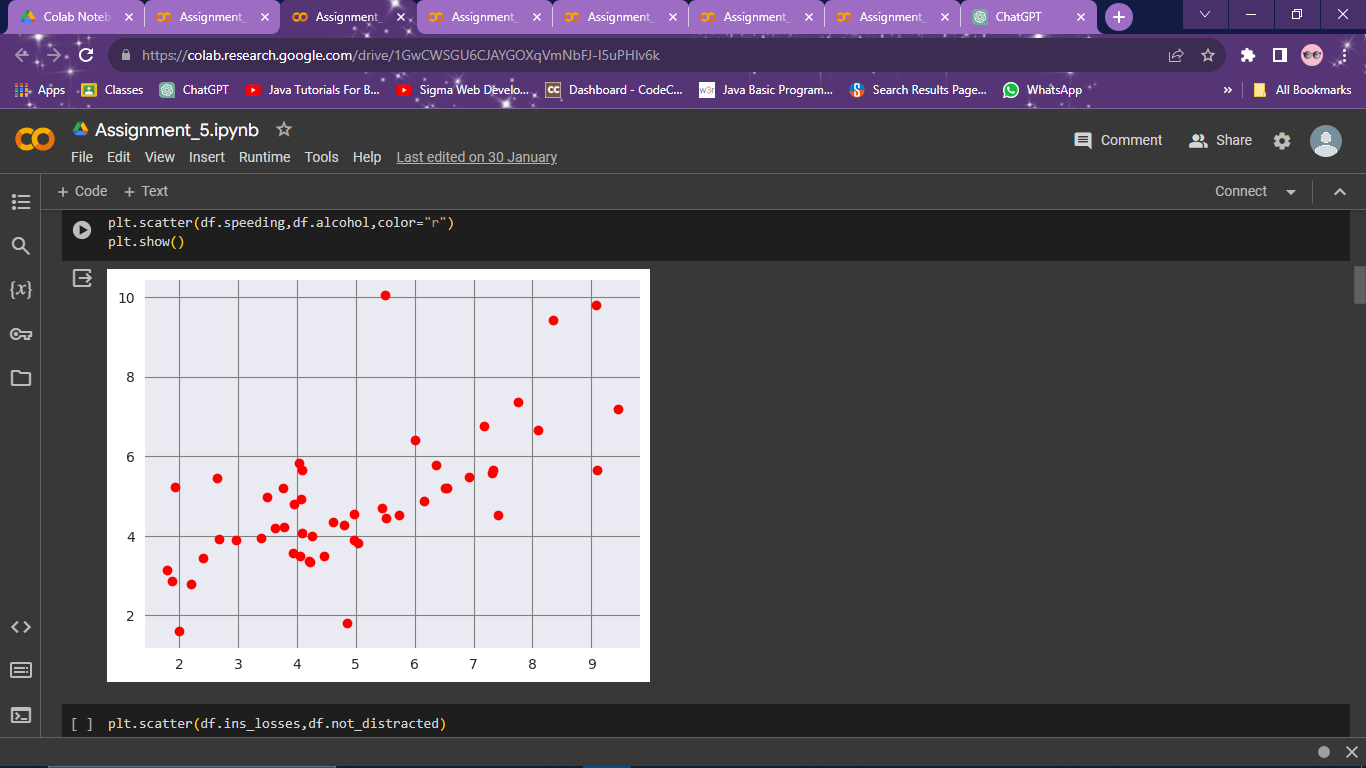
Applications:

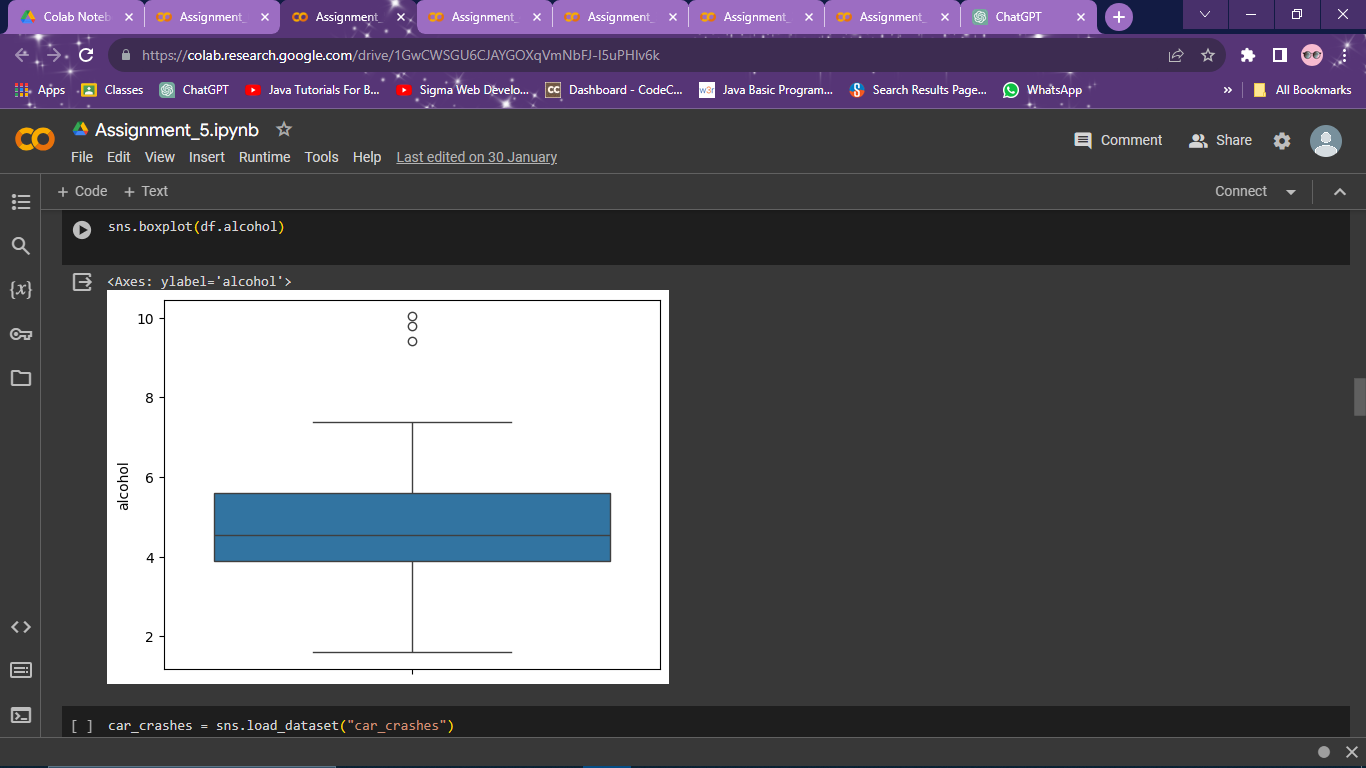
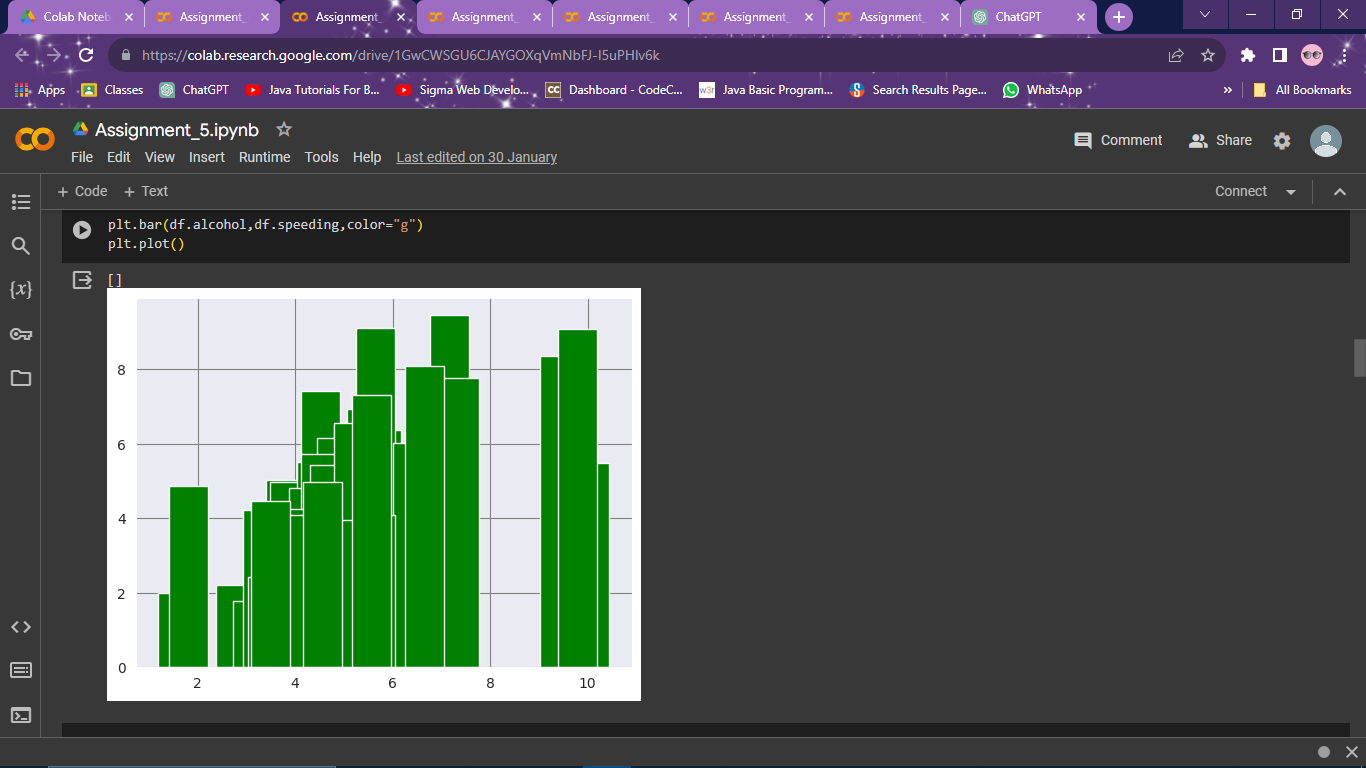
1. Scatter Plot: Used in regression analysis, correlation studies, and identifying clusters or patterns in data.
2. Bar Plot: Used in market research, sales analysis, and comparing performance across categories.
3. Box Plot: Used in statistical analysis, outlier detection, and comparing distributions across groups.
4. Histogram: Used in data preprocessing, understanding data distribution, and identifying data anomalies.

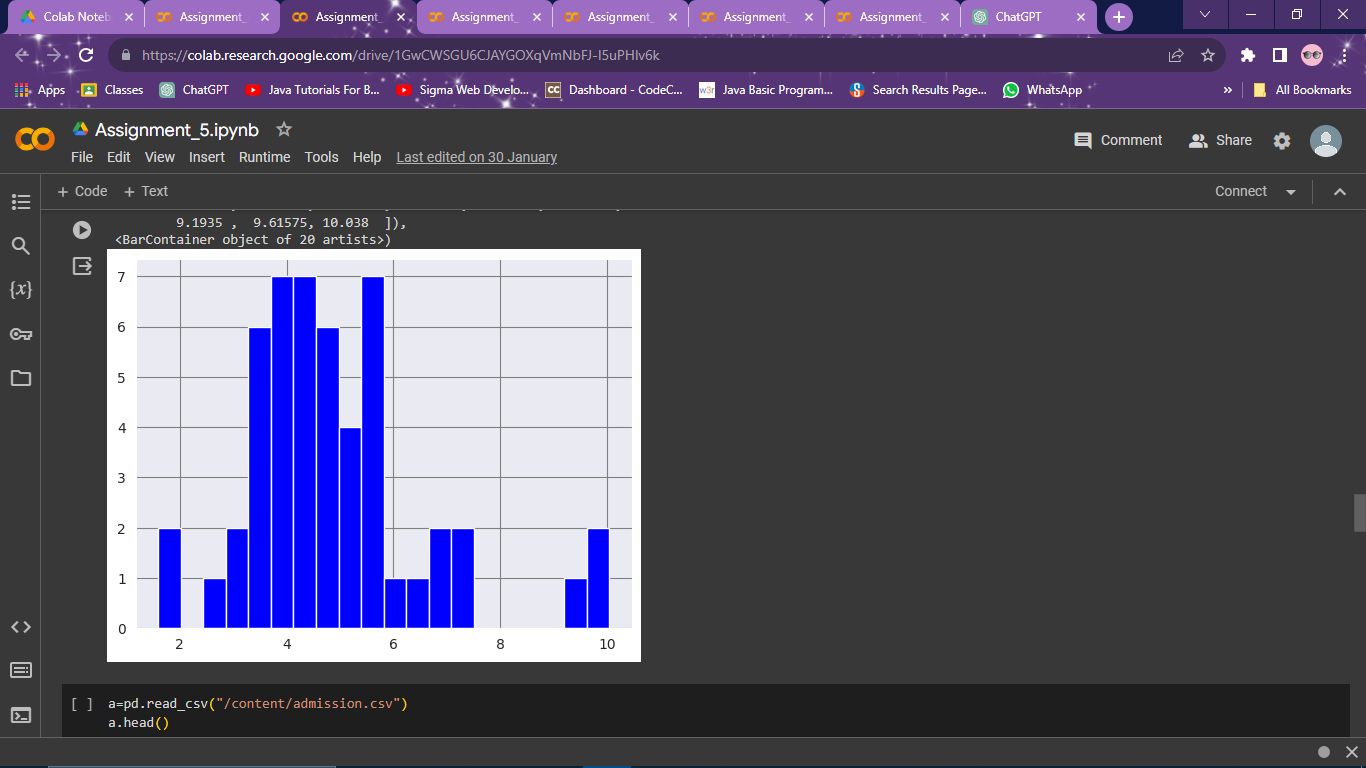
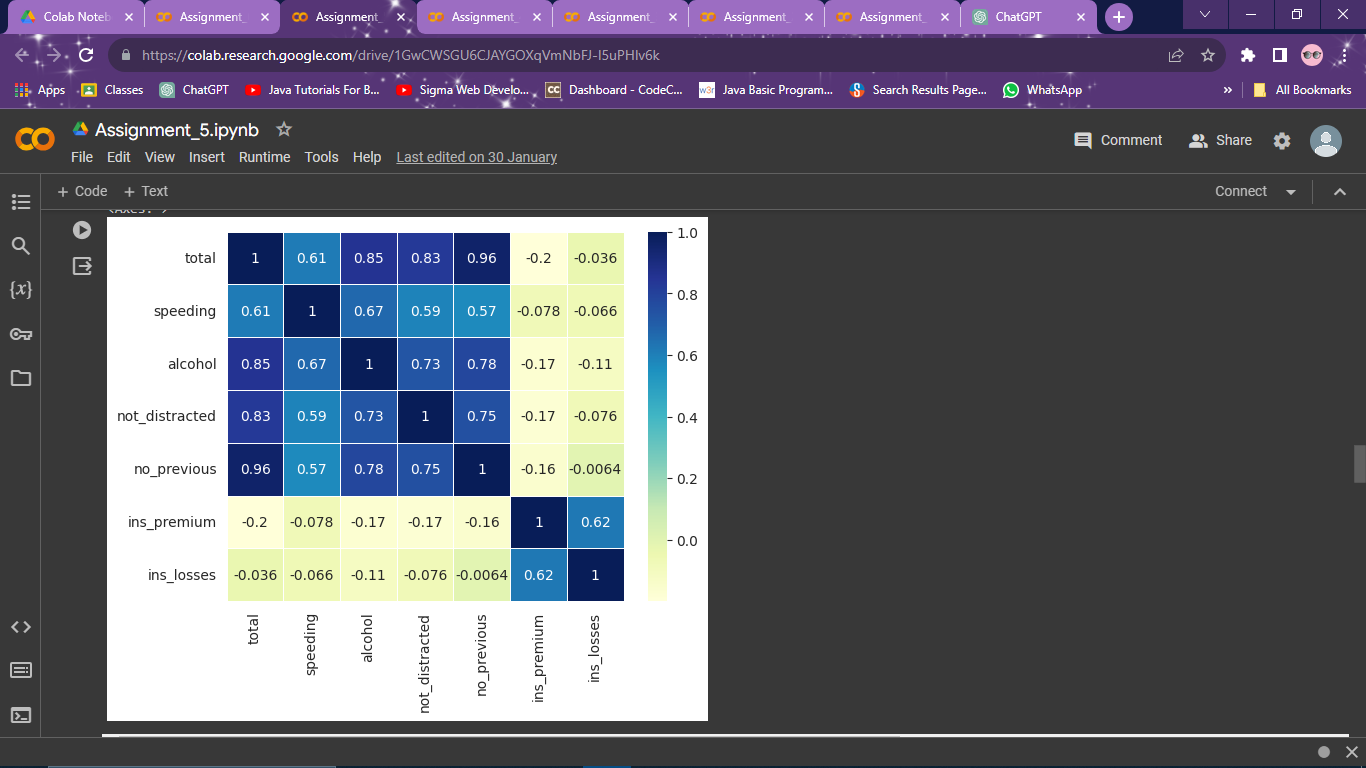
Limitations:

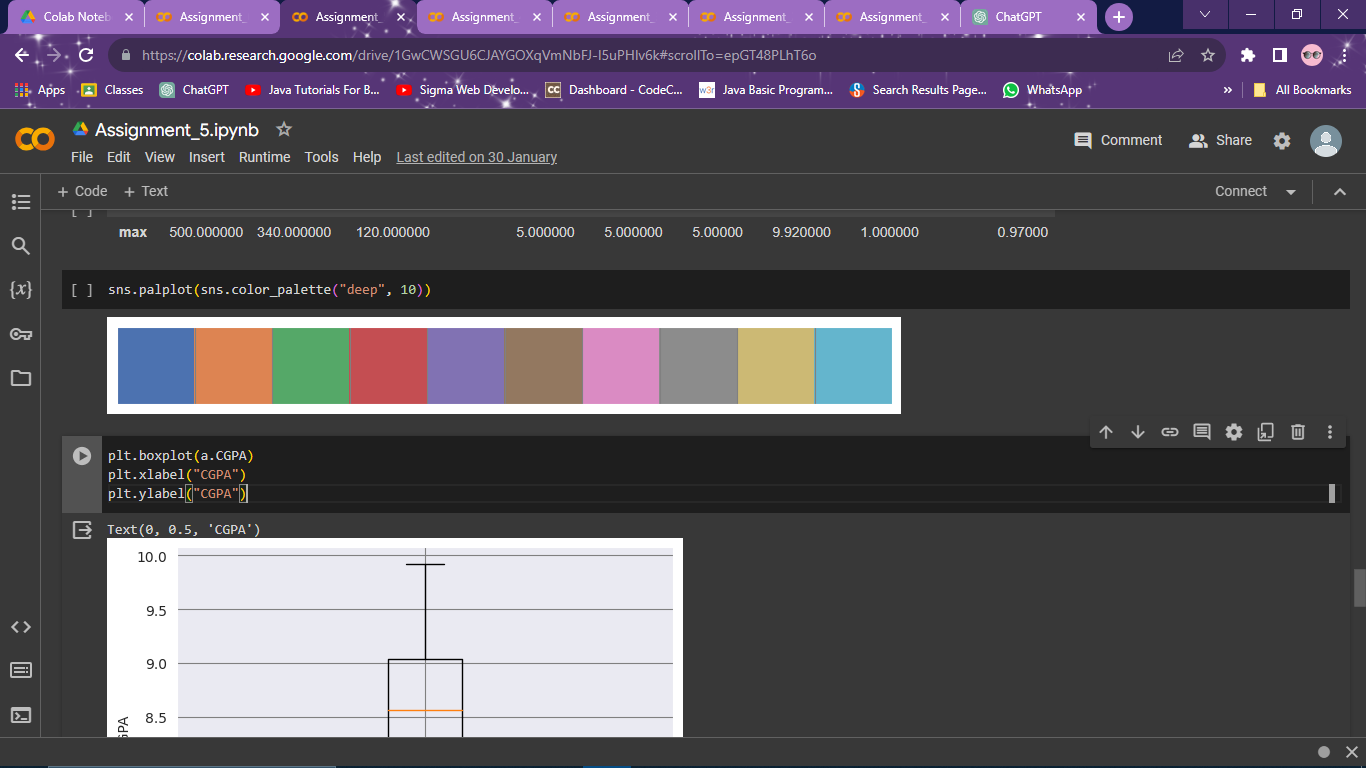
1. Scatter Plot: Limited in displaying multiple variables simultaneously.
2. Bar Plot: Not suitable for continuous data visualization.
3. Box Plot: May not show detailed data distribution compared to other plots.
4. Histogram: Sensitive to bin size, and interpretation can vary based on binning method.

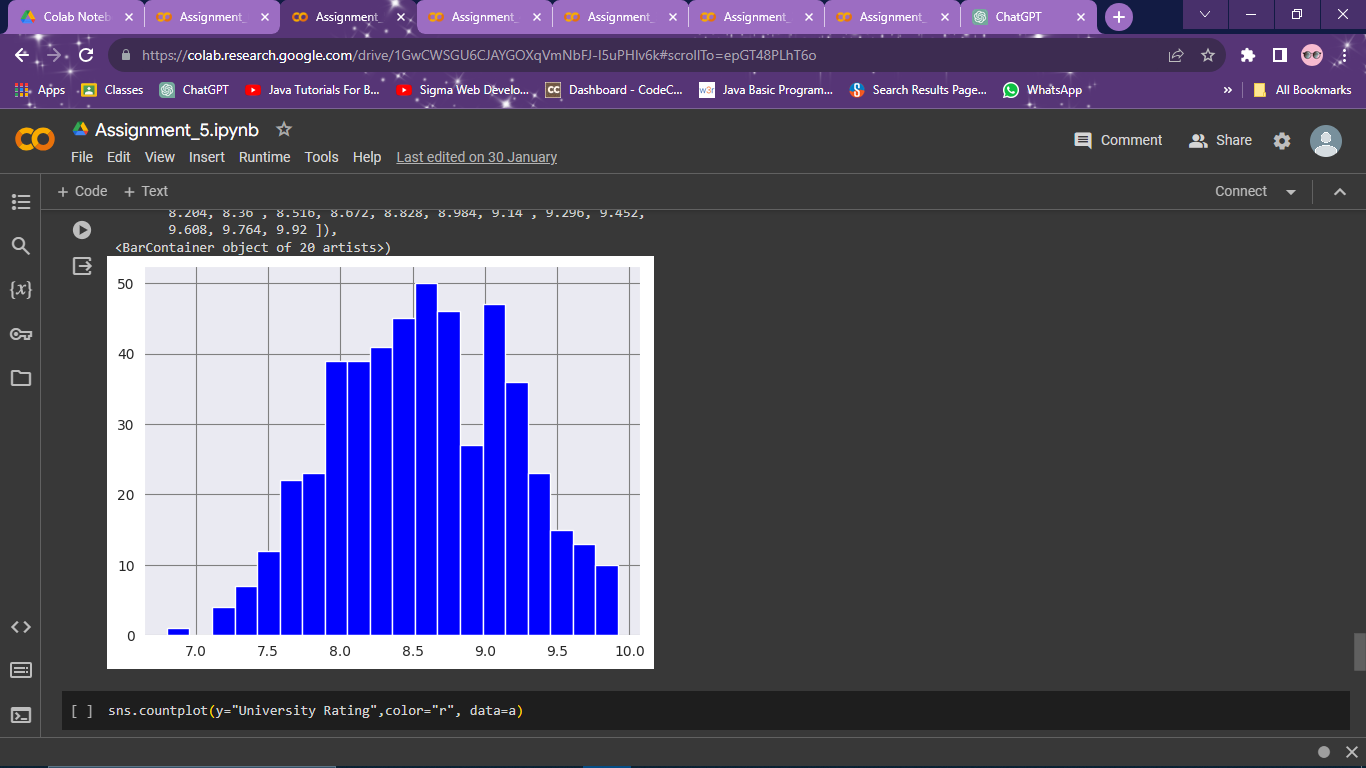
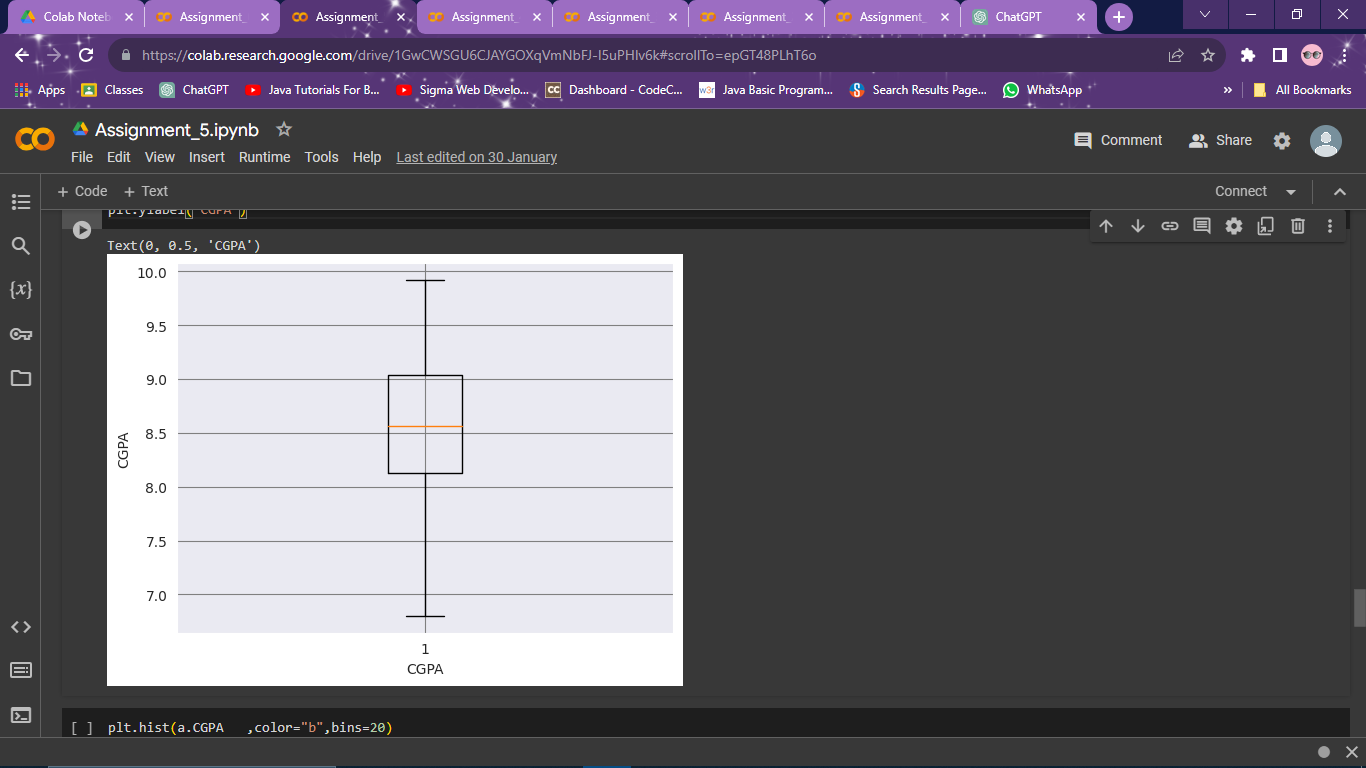
**OUTPUT :**

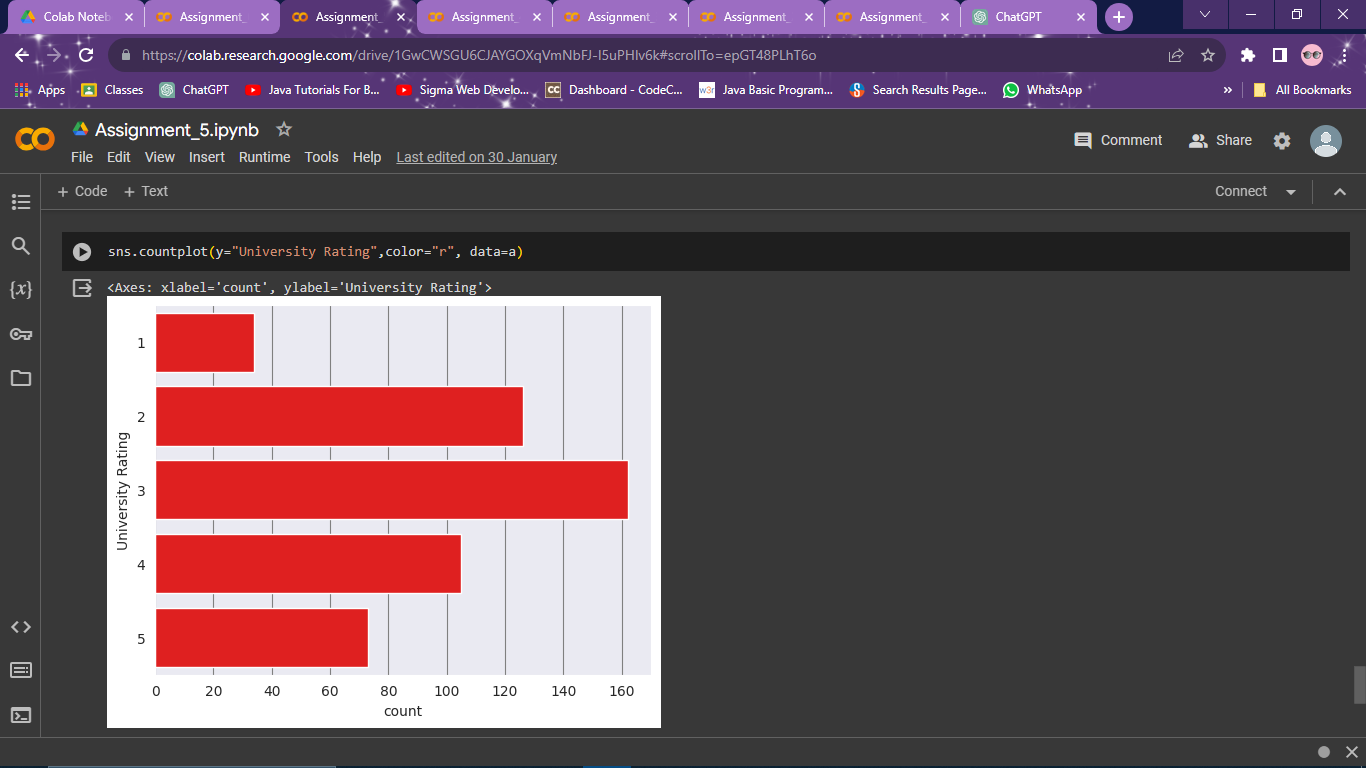
****

****

****

****

****

****

**Conclusion:**

The visualizations reveal key insights:

1. Scatter Plot: Shows relationships between variables; no clear pattern observed in the sample data.
2. Bar Plot: Compares categorical data; category C has the highest value in the sample data.
3. Box Plot: Illustrates data distribution and outliers; category E has a higher median value compared to others.
4. Histogram: Depicts the frequency distribution of numerical data; values are evenly distributed across bins in the sample data.

Overall, these visualizations provide a comprehensive understanding of the dataset's characteristics, including relationships, distributions, and outliers, aiding in data analysis and decision-making.