

Name: Jiya Sharma

Roll no: 281056

PRN: 22311741

Assignment 3

Problem Statement:

Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider a suitable dataset. Use a Scatter plot, Bar plot, Box plot, Pie chart, and Line Chart.

Objective:

- To introduce and explore basic visualization techniques in Python using Seaborn and Matplotlib.
- To demonstrate how to visualize data using different plot types, including Scatter plot, Bar plot, Box plot, Pie chart, and Line chart.
- To analyze a suitable dataset using the various plot types for better insights.

Resources used:

- Software used: Visual Studio Code
- Libraries used: Pandas, Matplotlib, Seaborn

Theory:

1. Seaborn:

Seaborn is a powerful Python visualization library based on Matplotlib, designed to simplify the process of creating informative and attractive statistical graphics. It provides a high-level interface for drawing various types of plots with simple functions, while also integrating with Pandas DataFrames for seamless data handling.

I. Key features:

- Built-in themes and color palettes.
- Integration with Pandas for DataFrame compatibility.
- More attractive default plots, including categorical plots and regression plots.

2. Matplotlib:

Matplotlib is one of the most widely used Python libraries for creating static, animated, and interactive visualizations. It provides a low-level interface for creating basic plots and offers extensive customization options.

II. Key features:

- Extensive support for various plot types such as line, scatter, bar, and pie charts.
- Highly customizable, allowing fine control over figure aesthetics.
- Integration with NumPy and Pandas.

Methodology:

For this assignment, we'll visualize data using the following plot types:

- **Bar Plot:** Bar plots are used to represent categorical data with rectangular bars, where the length of each bar is proportional to the value it represents.
- **Scatter Plot:** A scatter plot is used to display the relationship between two continuous variables. Each point represents a data point on the x and y axes.
- **Box Plot:** Box plots display the distribution of a dataset, highlighting the median, quartiles, and any potential outliers. They are useful for understanding the spread of data.
- **Pie Chart:** Pie charts are used to represent the proportions of a whole. Each segment of the pie represents a category's contribution to the total.
- **Line Chart:** Line charts show trends over time or continuous data points. It is useful for visualizing time series data.

Conclusion:

Data visualization is a key aspect of data analysis and communication. Using libraries like **Seaborn** and **Matplotlib** in Python, we can easily create a variety of visualizations to make sense of complex datasets. The Scatter plot, Bar plot, Box plot, Pie chart, and Line chart each serve a specific purpose and provide unique insights into the data. Proper visualization not only makes it easier to interpret the data but also aids in decision-making and sharing results with others effectively.