Name: Sejal Sampat Godbole

Roll No: 281030

Batch: A2

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.

Objectives:

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

Resources Used:

1. **Software used:** Visual Studio Code
2. **Libraries used:** Pandas, Matplotlib, Seaborn

Theory:

- **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.

  Key features:

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

- **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.

  Key features:

  1. Extensive support for various plot types such as line, scatter, bar, and pie charts.
  2. Highly customizable, allowing fine control over figure aesthetics.

3. Integration with NumPy and Pandas.

Methodology:

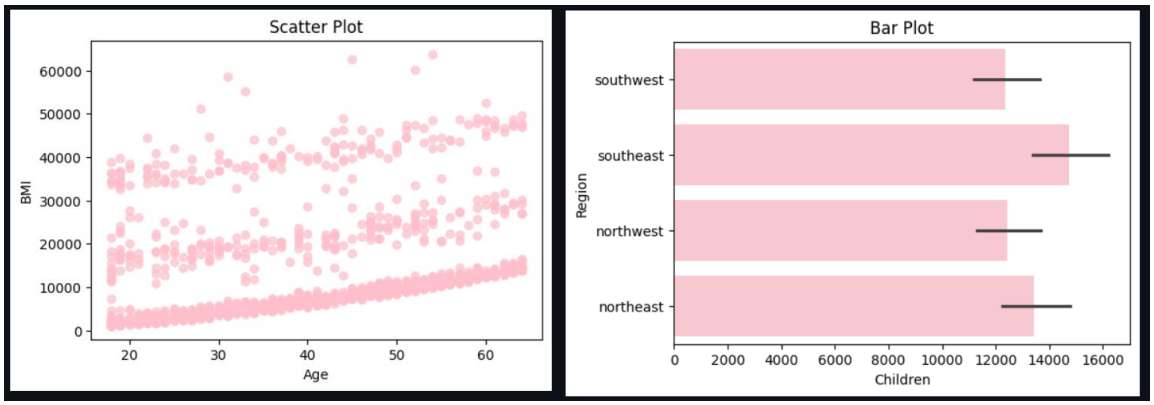
For this assignment, we'll visualize data using the following plot types:

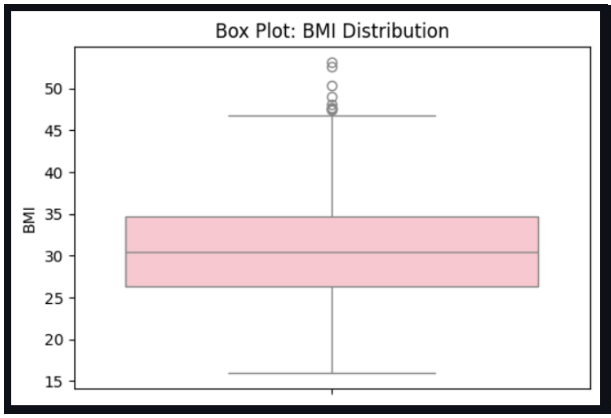
1. **Bar Plot:** Bar plots are used to represent categorical data with rectangular bars, where the length of each bar is proportional to the value it represents.
2. **Scatter Plot:** A scatter plot is used to display the relationship between two continuous variables. Each point represents a data point on the x and y axes.
3. **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.
4. **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.
5. **Line Chart:** Line charts show trends over time or continuous data points. It is useful for visualizing time series data.

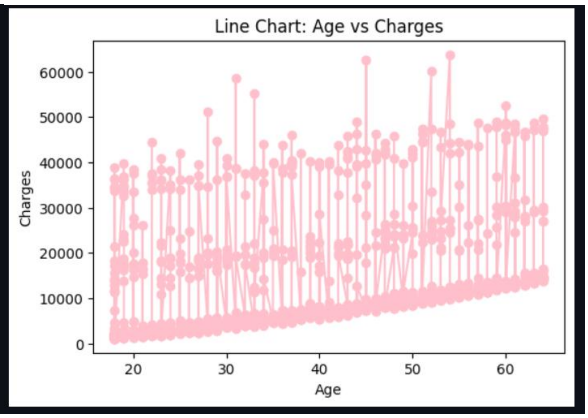
Conclusions:

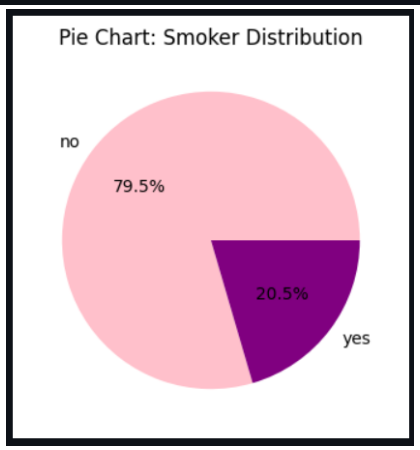
Data visualization is a key aspect of data analysis and communication. Using libraries like **Seaborn** and **Matplotlib** in Python, we can easily create a variety of visualizations to make sense of complex datasets. The Scatter plot, Bar plot, Box plot, Pie chart, and Line chart each serve a specific 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.

Results:




Box Plot: BMI Distribution
BMI
50
45
40
35
30
25
20
15


Line Chart: Age vs Charges
Charges
60000
50000
40000
30000
20000
10000
0
20
30
40
50
60
Age


Pie Chart: Smoker Distribution
no
79.5%
20.5%
yes