

Name: Prathamesh Khoje

Roll no: 281046

Batch: A2

Assignment 3

Statement:

In this assignment, we have to visualize data using **R/Python** by plotting different types of graphs. The visualization should be based on a suitable dataset and include the following plots:

- Scatter Plot
- Bar Plot
- Box Plot
- Pie Chart
- Line Chart

Objective:

1. This assignment aims to introduce you to data visualization techniques using **Matplotlib** and **Seaborn** in Python.
2. Learn how to represent data effectively using different types of charts.
3. Understand the significance of each type of plot in data analysis and interpretation.

Resources Used:

- **Software used:** Jupyter Notebook
- **Libraries used:** Matplotlib, Seaborn, Pandas, NumPy

Introduction to Data Visualization:

Data visualization is an essential technique in data science that helps to understand the patterns, trends, and insights hidden within the data. It allows us to represent complex data in a graphical format, making it easier to interpret.

Types of Plots Used:

1. **Scatter Plot:** Used to visualize the relationship between two numerical variables.
2. **Bar Plot:** Represents categorical data with rectangular bars.
3. **Box Plot:** Displays the distribution of numerical data and identifies outliers.
4. **Pie Chart:** Shows the proportion of different categories within a dataset.
5. **Line Chart:** Used to represent trends over a period of time.

Methodology:

1. **Data Collection and Exploration:**
 - Select a suitable dataset for visualization.
 - Load the dataset into a pandas DataFrame.
2. **Preprocessing:**
 - Handle missing values if any.
 - Format the data to be suitable for visualization.
3. **Visualization:**
 - Create different plots using Matplotlib and Seaborn.
 - Customize the plots for better readability.

Program Implementation:

1. Importing Libraries:
2. Loading the Dataset:
3. Scatter Plot:
4. Bar Plot:
5. Box Plot:
6. Pie Chart:
7. Line Chart:

Advantages:

1. Data visualization helps in understanding trends and patterns.
2. Graphical representation makes complex data easier to interpret.
3. Different plots help in analyzing different aspects of the dataset.

Disadvantages:

1. Poorly designed visualizations can lead to misinterpretation of data.
2. Large datasets may require advanced visualization techniques.

Conclusion:

This assignment introduced us to data visualization techniques using Python. We explored different types of plots such as scatter plots, bar plots, box plots, pie charts, and line charts. By implementing these visualizations, we gained hands-on experience in effectively representing data for analysis and interpretation.