

**Name:** Pathan Firdos Maheraj

**Roll no:** 281073

**Batch:** A3

### **Assignment 3**

#### **Statement:**

Q. Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider a suitable dataset. Use the following types of plots:

- a) Scatter Plot
- b) Bar Plot
- c) Box Plot
- d) Pie Chart
- e) Line Chart

#### **Objective:**

1. This assignment aims to introduce data visualization techniques using Python's Matplotlib and Seaborn libraries.
2. Understand different types of plots and their use cases for analyzing data patterns and distributions.
3. Enhance our ability to interpret data insights effectively through graphical representations.

#### **Resources used:**

1. Software used: Google Colab
2. Libraries used: Matplotlib, Seaborn, Pandas

#### **Introduction to Data Visualization:**

1. Data visualization is a crucial step in data analysis as it helps to identify trends, patterns, and anomalies.
2. Various types of graphs can be used depending on the nature of data and analysis requirements.
3. Python provides several libraries for visualization, with Matplotlib and Seaborn being the most widely used.
4. These libraries allow users to create a wide range of plots, from basic to advanced, to represent data effectively.

#### **Types of Graphs Used:**

1. **Scatter Plot:** Used to visualize the relationship between two numerical variables.
  - Function: `plt.scatter(x, y)` (Matplotlib) or `sns.scatterplot(x, y, data=df)` (Seaborn)
2. **Bar Plot:** Represents categorical data using rectangular bars.
  - Function: `plt.bar(x, y)` or `sns.barplot(x, y, data=df)`
3. **Box Plot:** Displays the distribution of a dataset and highlights outliers.
  - Function: `sns.boxplot(x, y, data=df)`
4. **Pie Chart:** Used to show proportions of categorical data.
  - Function: `plt.pie(sizes, labels=labels, autopct='%1.1f%%')`
5. **Line Chart:** Represents trends over time or continuous data changes.
  - Function: `plt.plot(x, y)`

## Methodology:

### 1. Data Collection and Preparation:

- Load a suitable dataset related to assignments 1 and 2.
- Clean and preprocess the data if necessary.

### 2. Data Visualization:

- Select appropriate plots based on the dataset.
- Generate each plot using Matplotlib and Seaborn.
- Customize plots with titles, labels, and color schemes for better readability.

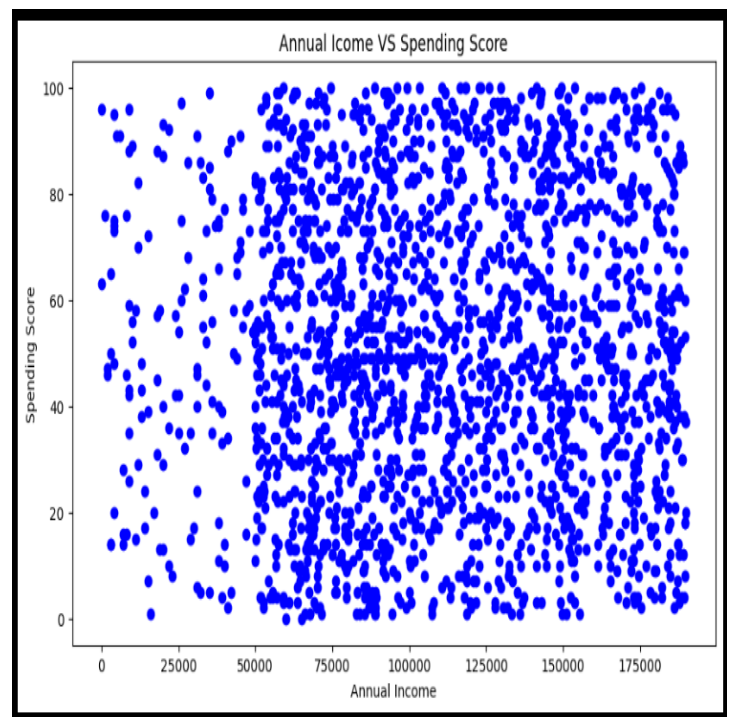
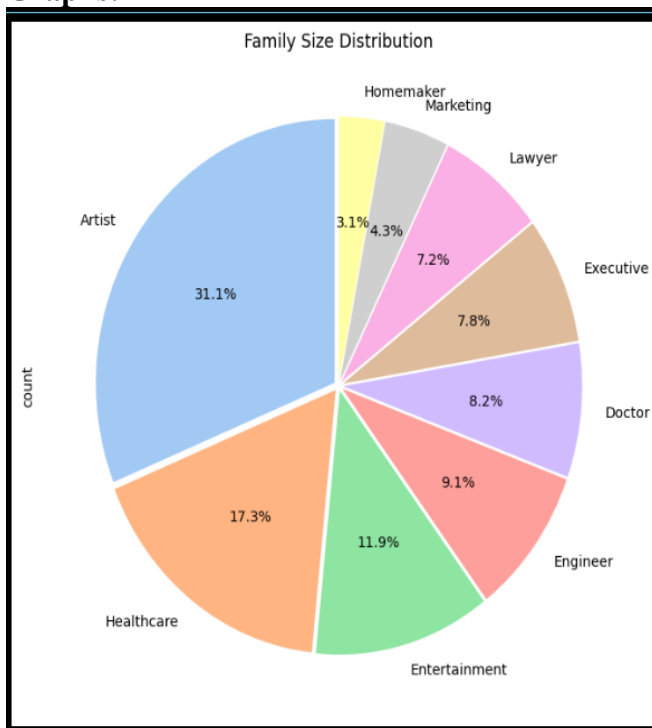
## Advantages:

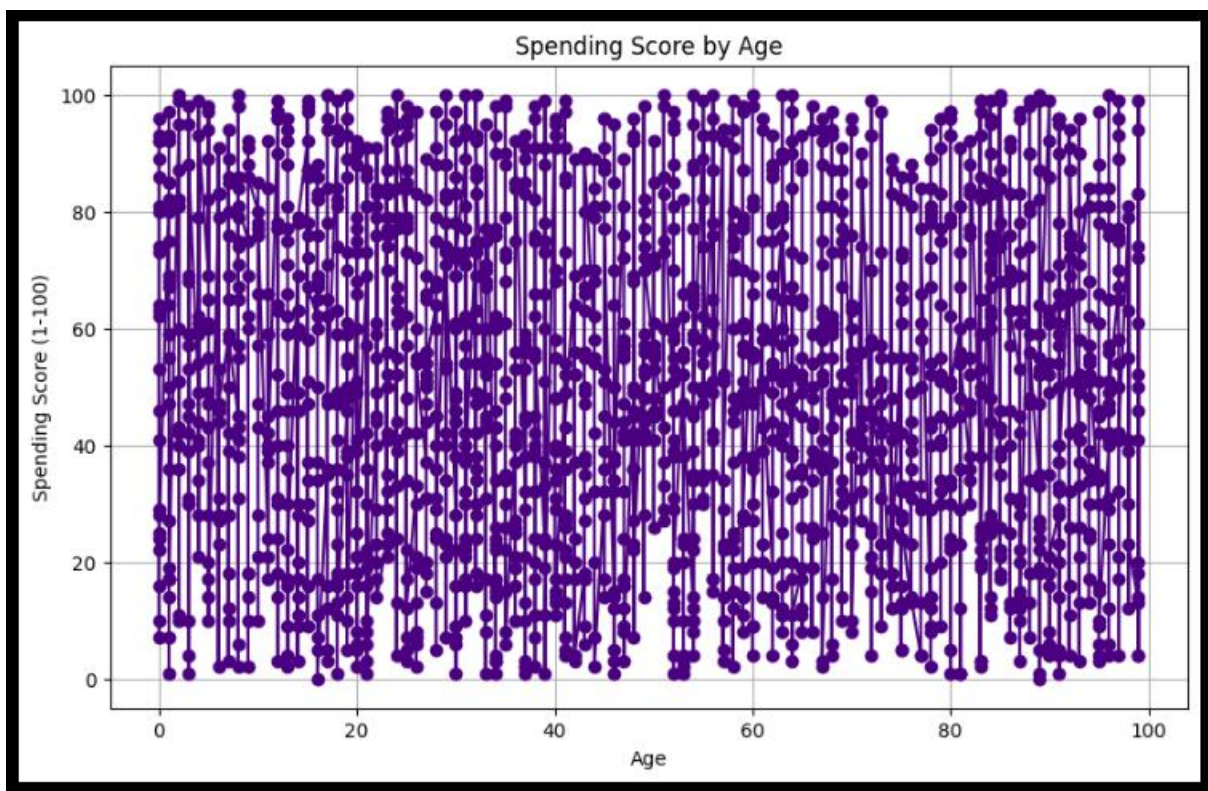
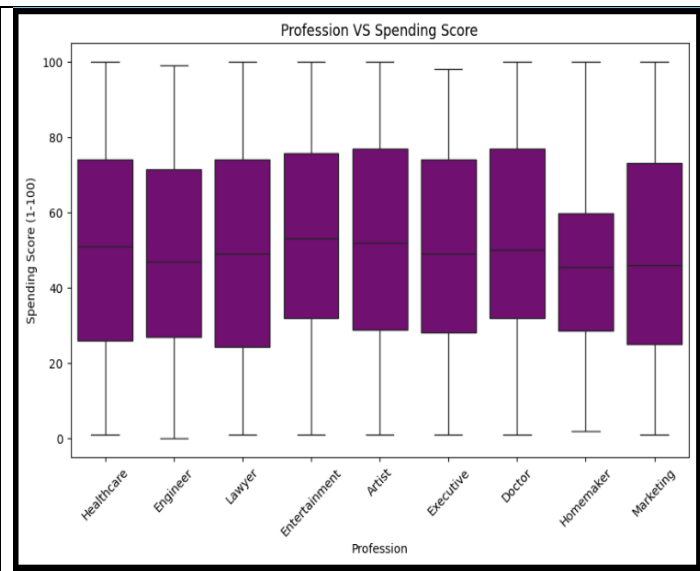
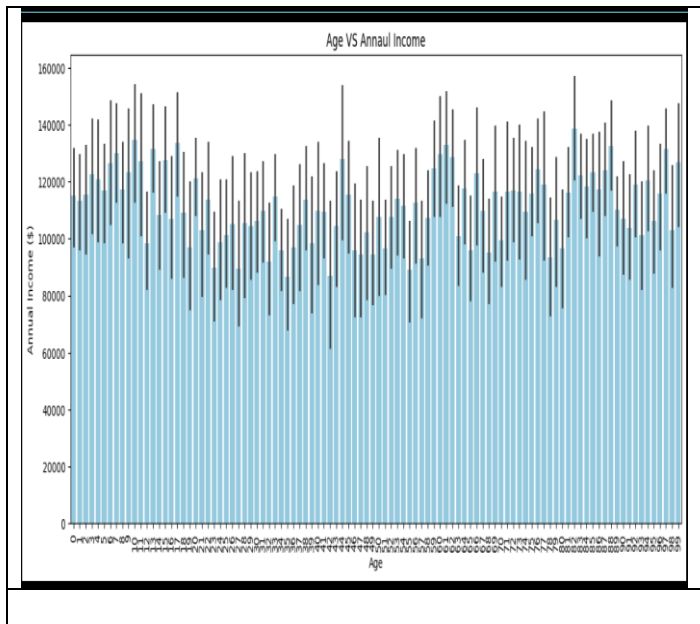
1. Enhances data interpretation by making patterns and relationships more evident.
2. Helps in identifying trends and anomalies in the dataset.
3. Provides an intuitive way to communicate data insights.

## Disadvantages:

1. Misleading graphs may lead to incorrect interpretations.
2. Large datasets may require advanced visualization techniques for better clarity.

## Graphs:





## Conclusion:

In summary, this assignment provided an overview of data visualization techniques using Python. We explored different types of graphs, including scatter plots, bar charts, box plots, pie charts, and line charts, to represent and analyze data effectively. Through practical implementation, we gained hands-on experience in selecting appropriate visualizations for different data types, enhancing our analytical skills. Mastering these visualization techniques will aid in making data-driven decisions in future projects.