**Assignment 3**

**Name: Ashutosh Shinde**

**Roll no: 282006**

**Batch: B1**

**Problem Statement**

Visualize the data using Python by plotting the following graphs for a suitable dataset:

1. **Scatter Plot**
2. **Bar Plot**
3. **Box Plot**
4. **Pie Chart**
5. **Line Chart**

**Objective**

1. To explore and visualize data distributions.
2. To analyze relationships between variables.
3. To present insights using different types of plots.

**Resources Used**

* **Software:** Google Colab
* **Libraries:**
  + Pandas (Data Manipulation)
  + Matplotlib (Basic Visualization)
  + Seaborn (Advanced Visualization)

**Methodology**

**1. Dataset Used**

* **Customers Dataset** (Customers.csv)
  + Contains customer demographics, income, spending scores, and professions.

**2. Data Preprocessing**

* Checked for missing values (Profession had 35 missing entries).
* Filled missing values:
  + Numerical (Annual Income) → Mean imputation.
  + Categorical (Profession) → Mode imputation.

**3. Visualizations**

1. **Bar Plot:** Spending Score by Gender.
2. **Scatter Plot:** Annual Income vs. Spending Score (Gender-wise).
3. **Box Plot:** Spending Score Distribution by Profession.
4. **Pie Chart:** Gender Distribution.
5. **Line Plot:** Age vs. Spending Score (Gender-wise).

**Code Implementation & Results**

**1. Bar Plot: Spending Score by Gender**

python

Copy

plt.figure(figsize=(8, 6))

sns.barplot(x='Gender', y='Spending Score (1-100)', data=data)

plt.title('Spending Score by Gender')

plt.show()

**Insight:**

* Females have a slightly higher average spending score than males.

**2. Scatter Plot: Annual Income vs. Spending Score**

python

Copy

plt.figure(figsize=(8, 6))

sns.scatterplot(

x='Annual Income ($)',

y='Spending Score (1-100)',

data=data,

hue='Gender',

palette='Set1'

)

plt.title('Annual Income vs. Spending Score')

plt.show()

**Insight:**

* Customers with mid-range incomes (≈50k–50*k*–100k) tend to have higher spending scores.
* No clear gender-based income-spending pattern.

**3. Box Plot: Spending Score by Profession**

python

Copy

plt.figure(figsize=(10, 6))

sns.boxplot(

x='Profession',

y='Spending Score (1-100)',

data=data,

palette='Set2'

)

plt.xticks(rotation=45)

plt.title('Spending Score Distribution by Profession')

plt.show()

**Insight:**

* **Engineers** and **Healthcare** professionals show wider spending score ranges.
* **Artists** have a more concentrated spending range (50–75).

**4. Pie Chart: Gender Distribution**

python

Copy

gender\_counts = data['Gender'].value\_counts()

plt.figure(figsize=(6, 6))

gender\_counts.plot(

kind='pie',

autopct='%1.1f%%',

colors=['lightblue', 'lightcoral']

)

plt.title('Gender Distribution')

plt.show()

**Insight:**

* The dataset is **balanced** (≈50% Male, 50% Female).

**5. Line Plot: Age vs. Spending Score**

python

Copy

plt.figure(figsize=(8, 6))

sns.lineplot(

x='Age',

y='Spending Score (1-100)',

data=data,

hue='Gender',

palette='Set1'

)

plt.title('Age vs. Spending Score')

plt.show()

**Insight:**

* Younger customers (20–40) have higher spending scores.
* Spending declines with age, especially for males.

**Conclusion**

1. **Gender Analysis:** Females spend slightly more than males.
2. **Income-Spending Relationship:** Mid-income customers are the most active spenders.
3. **Profession Impact:** Engineers and artists exhibit distinct spending behaviors.
4. **Age Trend:** Younger customers drive higher spending.

These visualizations help businesses **segment customers** and tailor marketing strategies effectively.