Name: Rakshith B

Course: Data Science and AIML

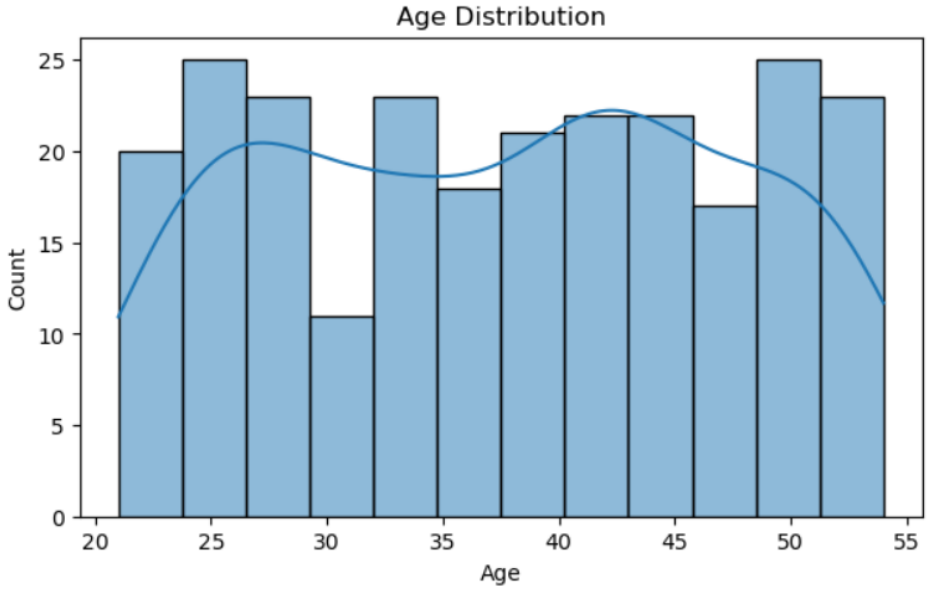
Customer Analytics EDA

Introduction:

This project focuses on **Exploratory Data Analysis (EDA)** using a Customer Analytics dataset to understand customer characteristics and behavior through data visualization and statistical analysis. The main goal of the project is to inspect the dataset, clean the data, and identify meaningful patterns and relationships between different features. Each row in the dataset represents one customer and contains information such as demographic details, income, and spending behavior. By applying EDA techniques using Python libraries like Pandas, Matplotlib, and Seaborn, the project helps in understanding data distribution, detecting trends, and generating insights that can support better business decision-making.

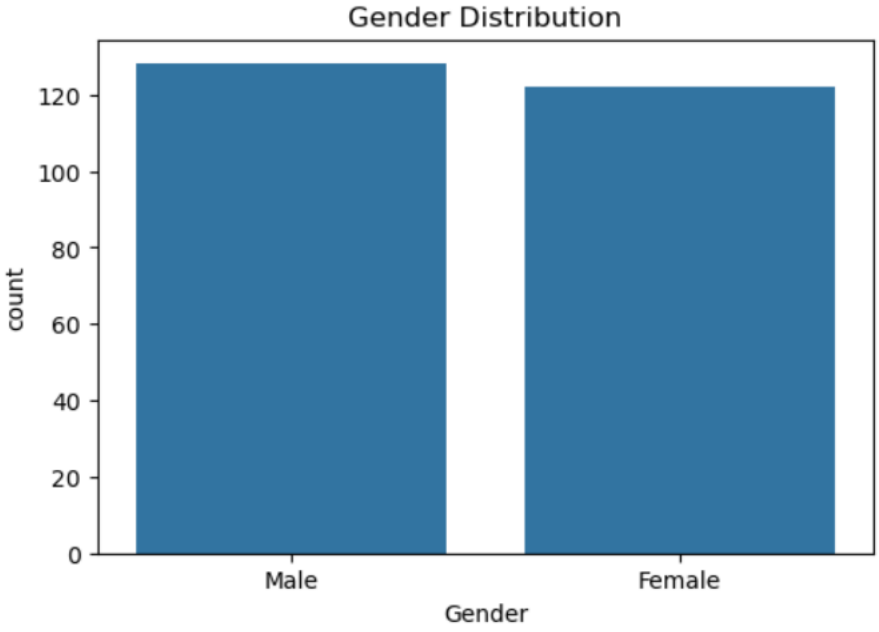
Graphs and Analysis:

1. Age Distribution (Histogram)



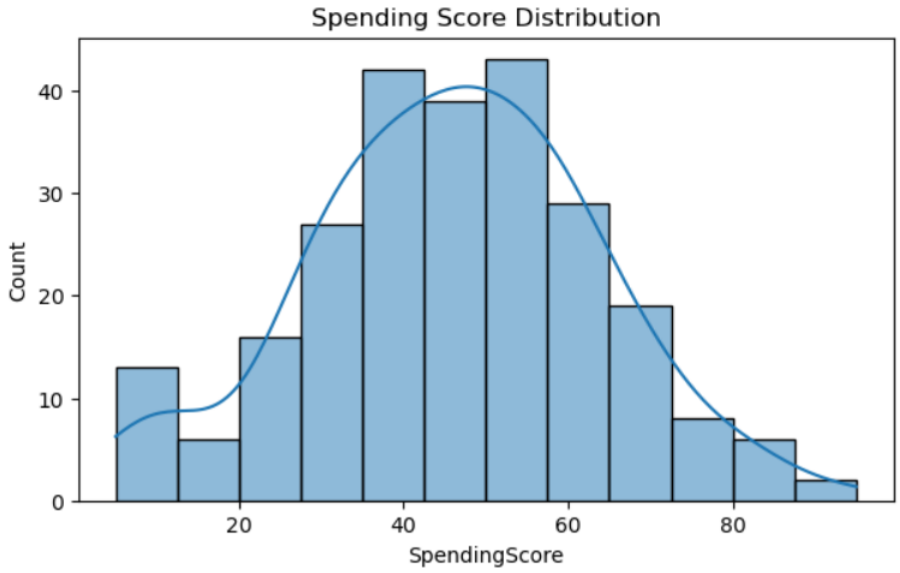
The histogram shows how customer ages are spread across different age groups. Most customers fall within the young to middle-age range, indicating that the dataset mainly represents adult customers with varied age distribution.

2. Gender Distribution (Bar Chart)



The bar chart compares the number of customers based on gender. It shows that both genders are fairly represented in the dataset, suggesting there is no major gender imbalance in the data.

3. Spending Score Distribution (Histogram)

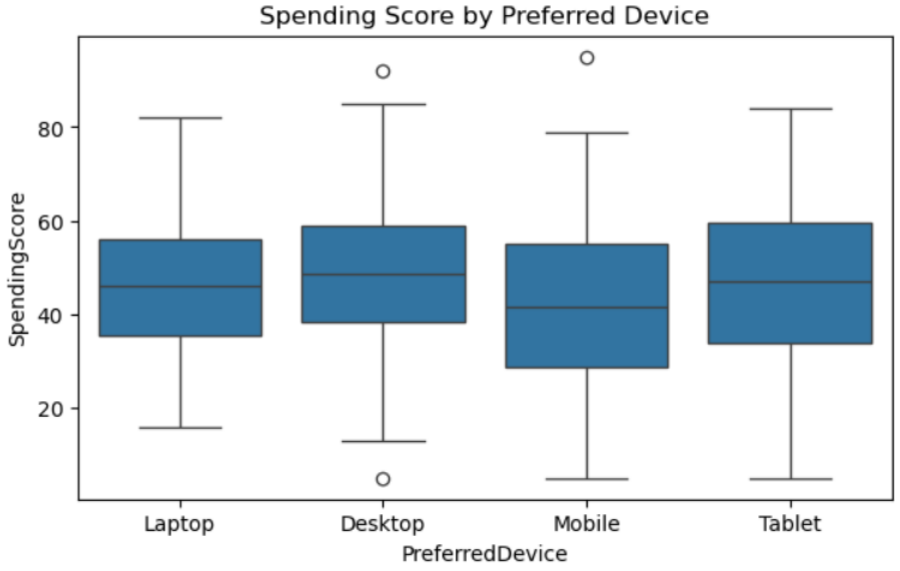


This histogram displays how spending scores are distributed among customers. The spread of values indicates that customers have different spending behaviors, ranging from low to high spenders.

4. Annual Income vs Spending Score (Scatter Plot)

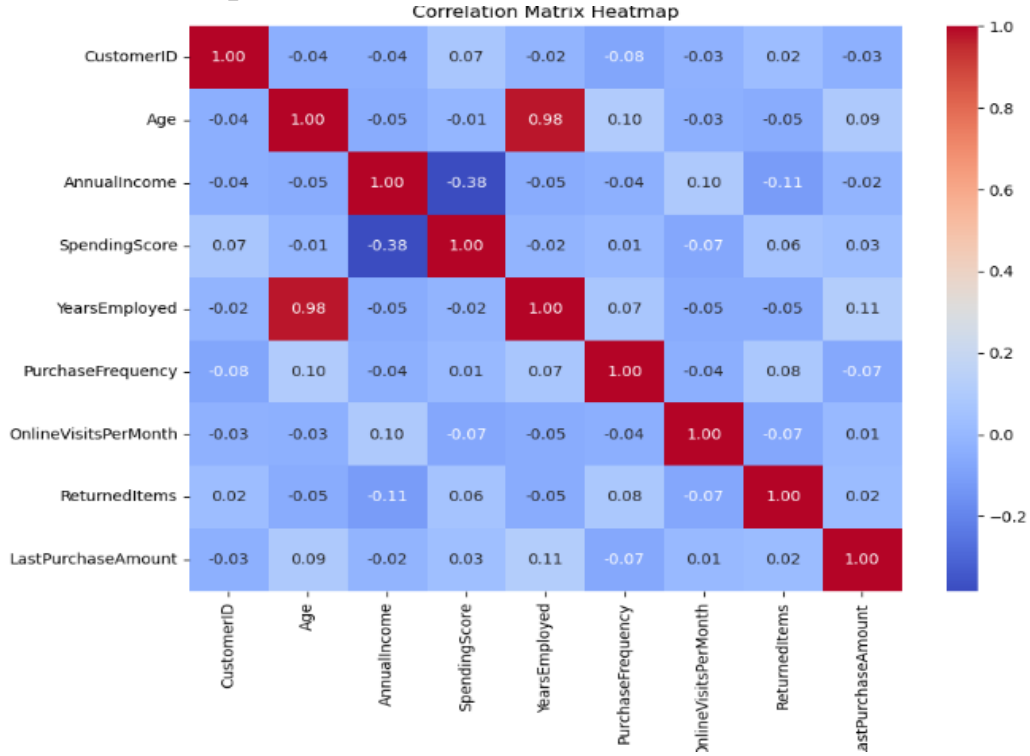
The scatter plot shows the relationship between annual income and spending score. The points are scattered without a strong pattern, suggesting that higher income does not always mean higher spending.

5. Spending Score by Preferred Device (Boxplot)



The boxplot compares spending scores across different preferred devices. It helps identify variations in customer spending behavior depending on the device they use, showing differences in median values and spread.

6. Correlation Heatmap



The correlation heatmap shows relationships between numerical variables in the dataset. Stronger colors indicate stronger relationships, helping identify which features are positively or negatively related to each other.

Code Snippet:

1. Dataset Loading

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_csv("customer_analytics.csv")

# View data
df.head()
```

2. Data Cleaning (Missing Values & Duplicates)

```
# Check missing values
df.isnull().sum()

# Fill missing values
df['Education'] = df['Education'].fillna(df['Education'].mode()[0])
df['AnnualIncome'] = df['AnnualIncome'].fillna(df['AnnualIncome'].median())

# Remove duplicate rows
df = df.drop_duplicates()
```

3. Plot Generation Code

```
# Age Distribution
sns.histplot(df['Age'], kde=True)
plt.title("Age Distribution")
plt.show()

# Gender Distribution
sns.countplot(x='Gender', data=df)
plt.title("Gender Distribution")
plt.show()

# Spending Score Distribution
sns.histplot(df['SpendingScore'], kde=True)
plt.title("Spending Score Distribution")
plt.show()

# Income vs Spending Score
sns.scatterplot(x='AnnualIncome', y='SpendingScore', data=df)
plt.title("Annual Income vs Spending Score")
plt.show()
```

```
# Spending Score by Preferred Device
sns.boxplot(x='PreferredDevice', y='SpendingScore', data=df)
plt.title("Spending Score by Preferred Device")
plt.show()

# Correlation Heatmap
corr = df.select_dtypes(include='number').corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```

Conclusion:

This Exploratory Data Analysis (EDA) project helped in understanding customer data through visualization and statistical exploration. By analyzing different graphs, we observed patterns in customer demographics, spending behavior, and relationships between variables. The analysis showed that spending habits vary across customers and are not strongly dependent only on income, indicating the importance of multiple factors in customer behavior. Overall, the project improved understanding of data cleaning, visualization, and insight generation, which are essential steps in data analysis.