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

```
df=pd.read_csv("/content/Mall_Customers.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Gender                  200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
df.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

```
df.shape
```

```
(200, 5)
```

```
df.isnull().sum()
```

	0
CustomerID	0
Gender	0
Age	0
Annual Income (k$)	0
Spending Score (1-100)	0

dtype: int64

```
df.dtypes
```

	0
CustomerID	int64
Gender	object
Age	int64
Annual Income (k$)	int64
Spending Score (1-100)	int64

dtype: object

```
df.duplicated().sum()
```

```
np.int64(0)
```

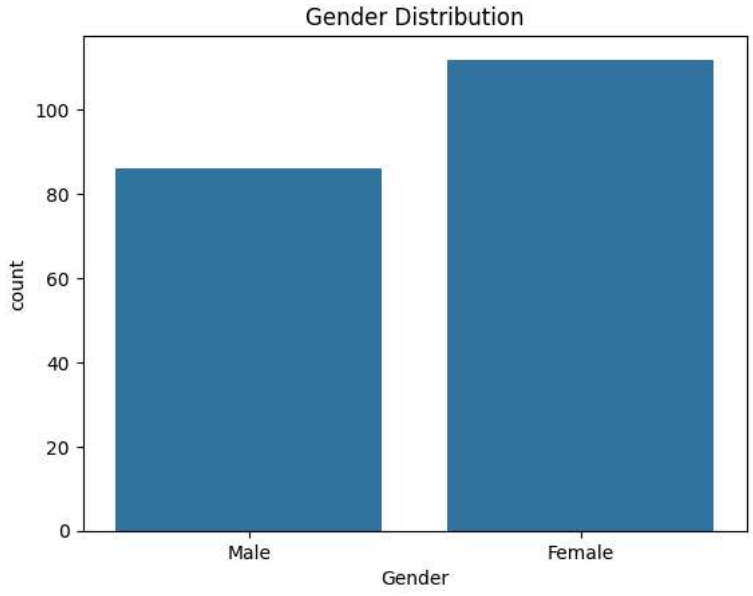
```
Q1 = df["Annual Income (k$)"].quantile(0.25)
Q3 = df["Annual Income (k$)"].quantile(0.75)
IQR = Q3 - Q1

lower = Q1 - 1.5 * IQR
upper = Q3 + 1.5 * IQR

df = df[(df["Annual Income (k$)"] >= lower) &
        (df["Annual Income (k$)"] <= upper)]

df.shape
```

```
(198, 5)
```

VISUALIZATION

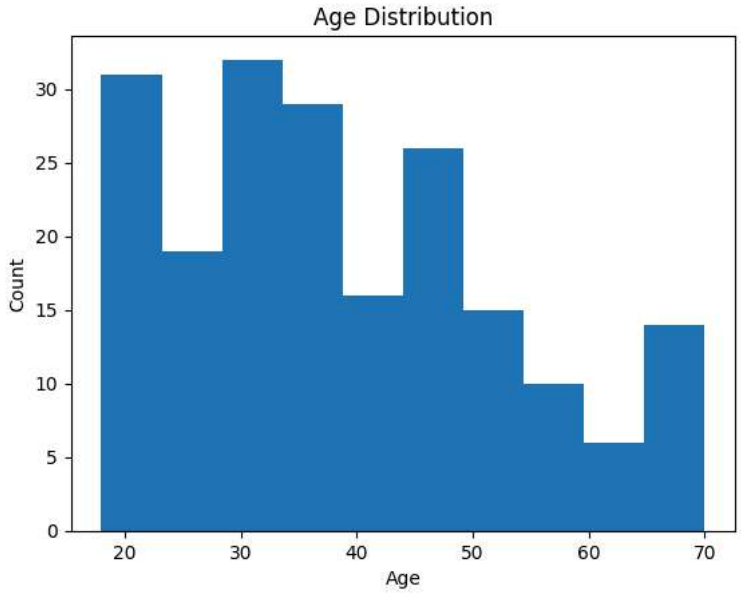
1.Gender Distribution

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



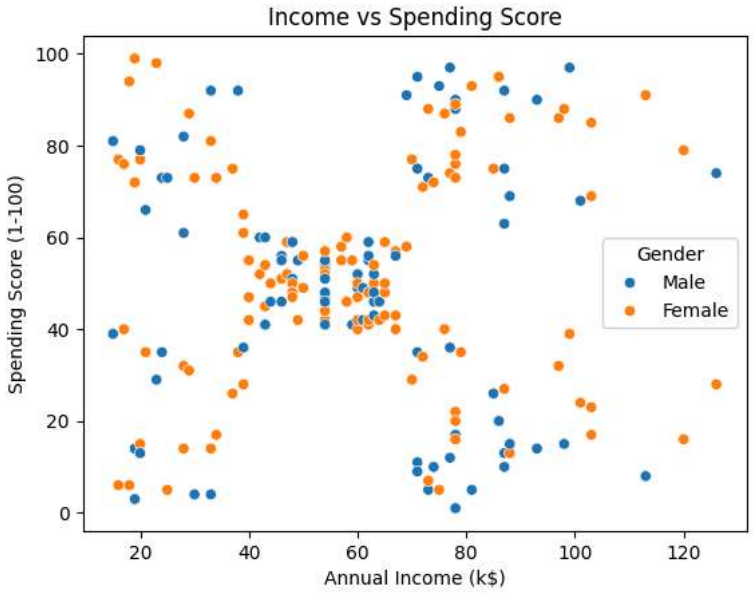
2. Age Distribution

```
plt.figure()
plt.hist(df["Age"], bins=10)
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```



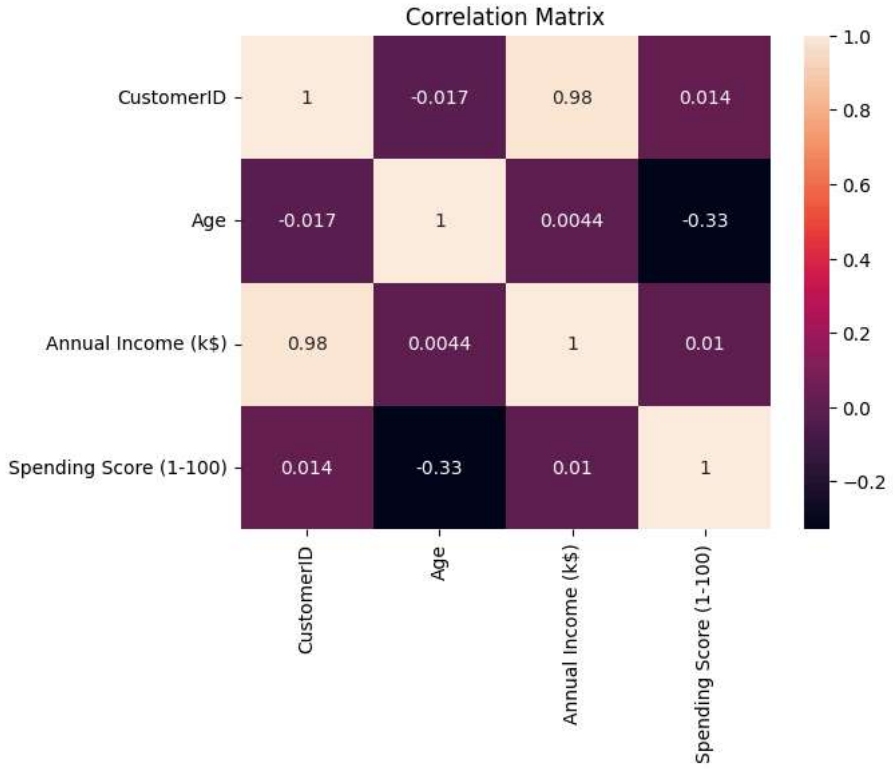
3. Income vs Spending Score

```
plt.figure()
sns.scatterplot(x="Annual Income (k$)",
                y="Spending Score (1-100)",
                hue="Gender",
                data=df)
plt.title("Income vs Spending Score")
plt.show()
```



Correlation Heatmap

```
plt.figure()
sns.heatmap(df.corr(numeric_only=True), annot=True)
plt.title("Correlation Matrix")
plt.show()
```



OBSERVATIONS

Most customers are in the age group of 20–40.

Spending score varies across income levels.

Both male and female customers contribute significantly.

Some customers with high income have low spending score and vice versa.