

NAME : EDE RENUKA MADHAV

REG NO: 21BCT0244

SLOT : 6:00 PM TO 8:00 PM

VIT VELLORE

Market Basket Magic: Extracting Insights for Retail Success

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

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

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
df.shape
```

```
(200, 5)
```

```
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
```

Data Preprocessing

```
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

CHECKING NULL VALUES

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

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

dtype: int64

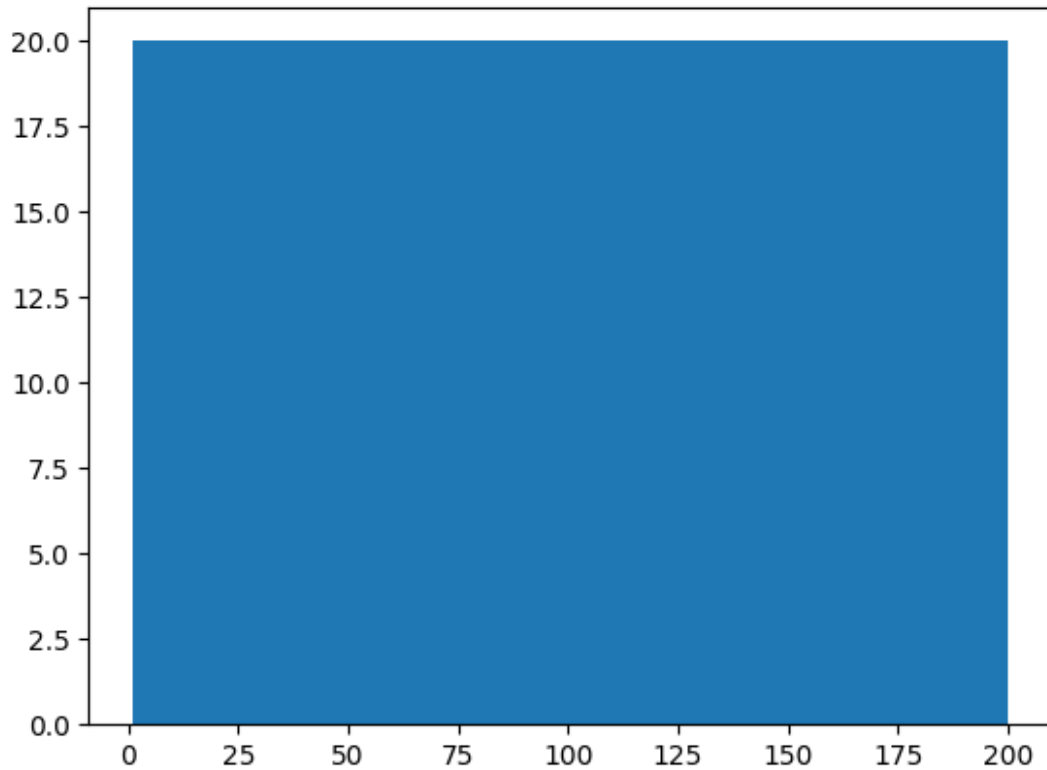
```
df['Gender'].value_counts()
```

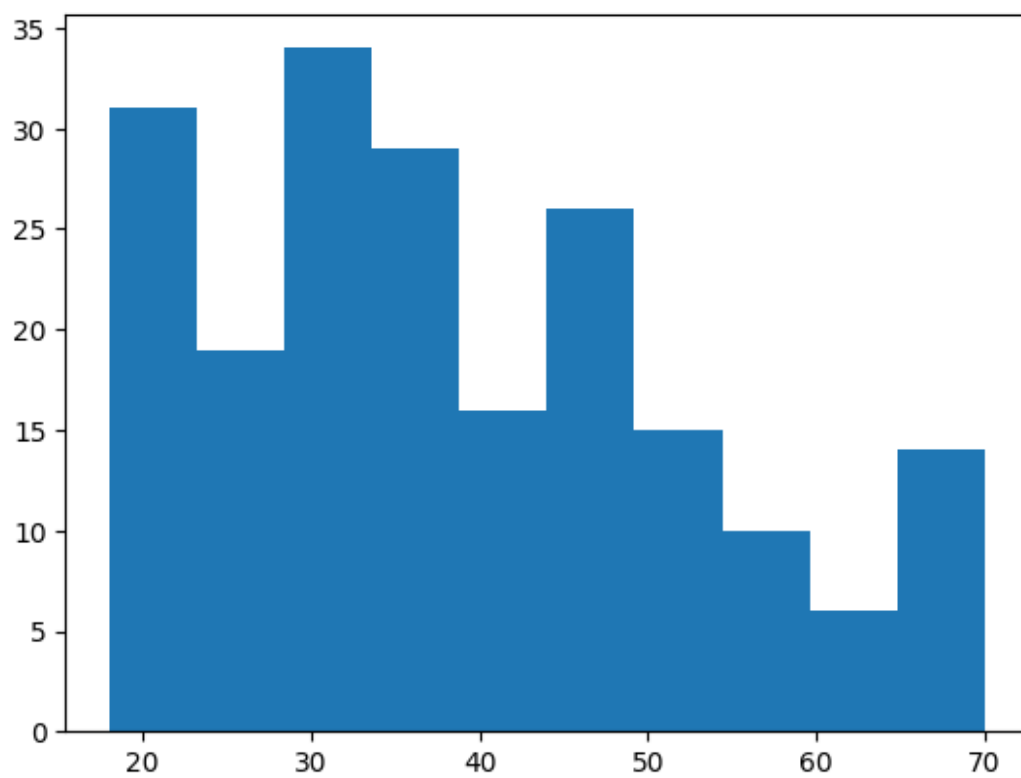
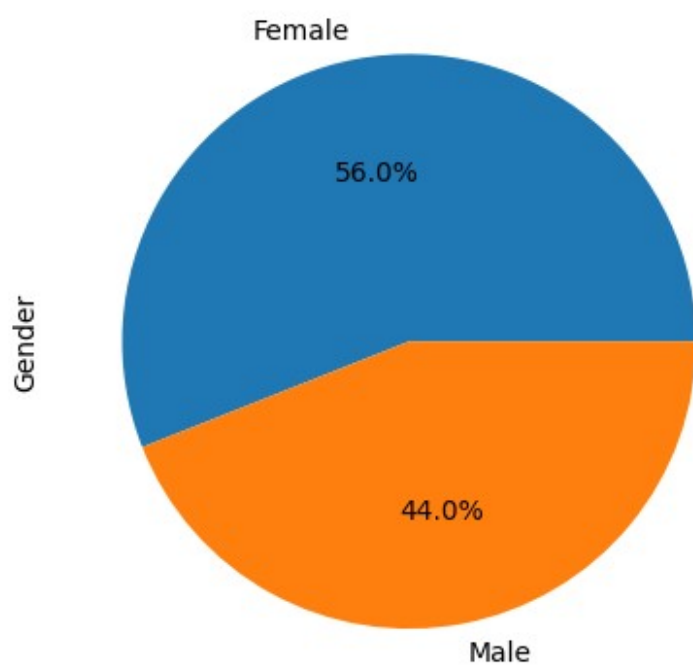
Female	112
Male	88

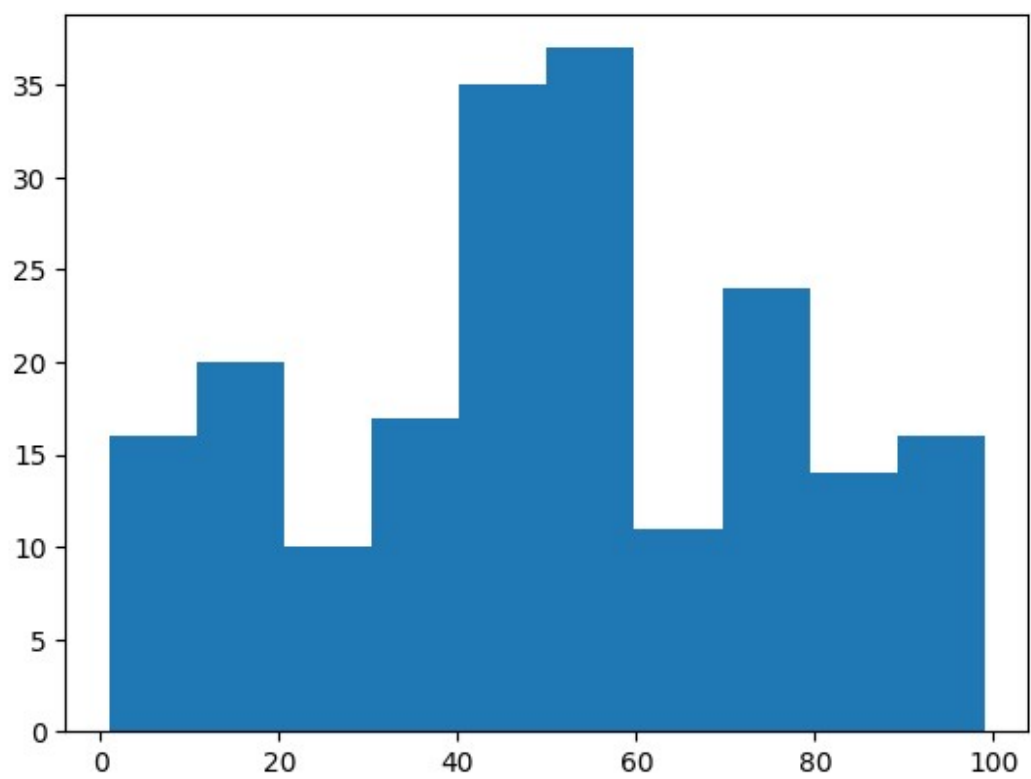
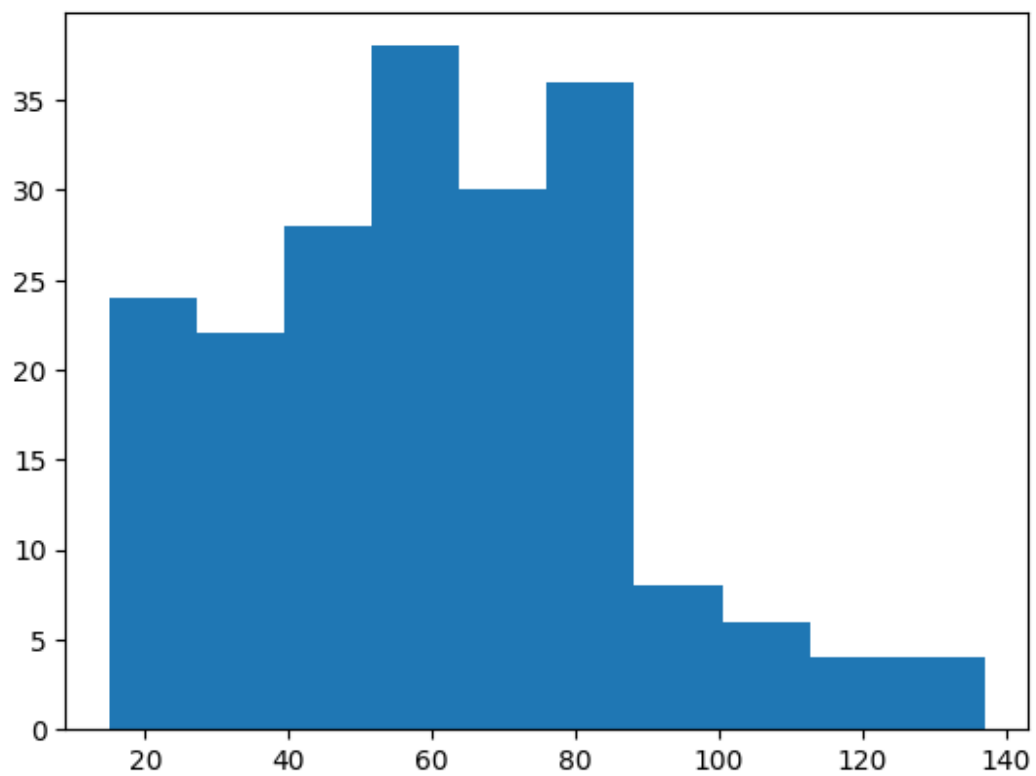
Name: Gender, dtype: int64

UNIVARIATE ANALYSIS

```
for i in df.columns:  
    if(i!='Gender'):  
        plt.hist(df[i])  
        plt.show()  
    else:  
        df[i].value_counts().plot.pie(autopct='%1.1f%%')  
        plt.show()
```

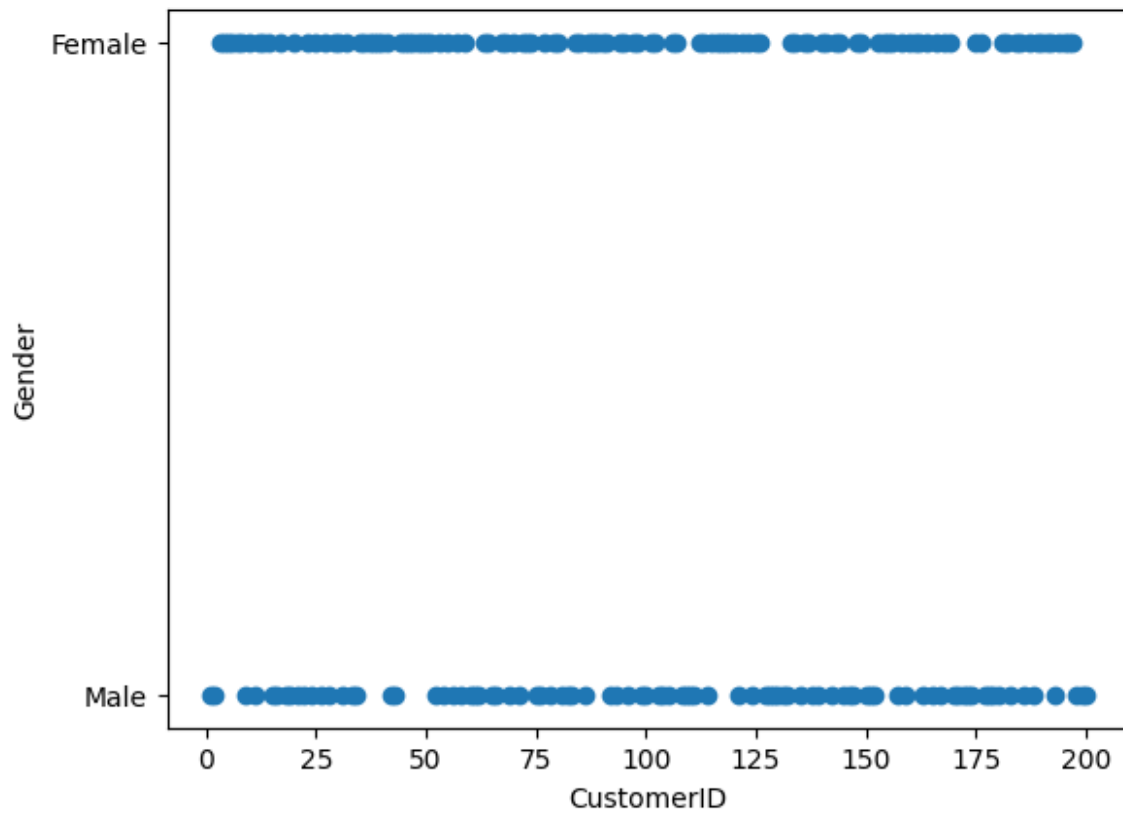


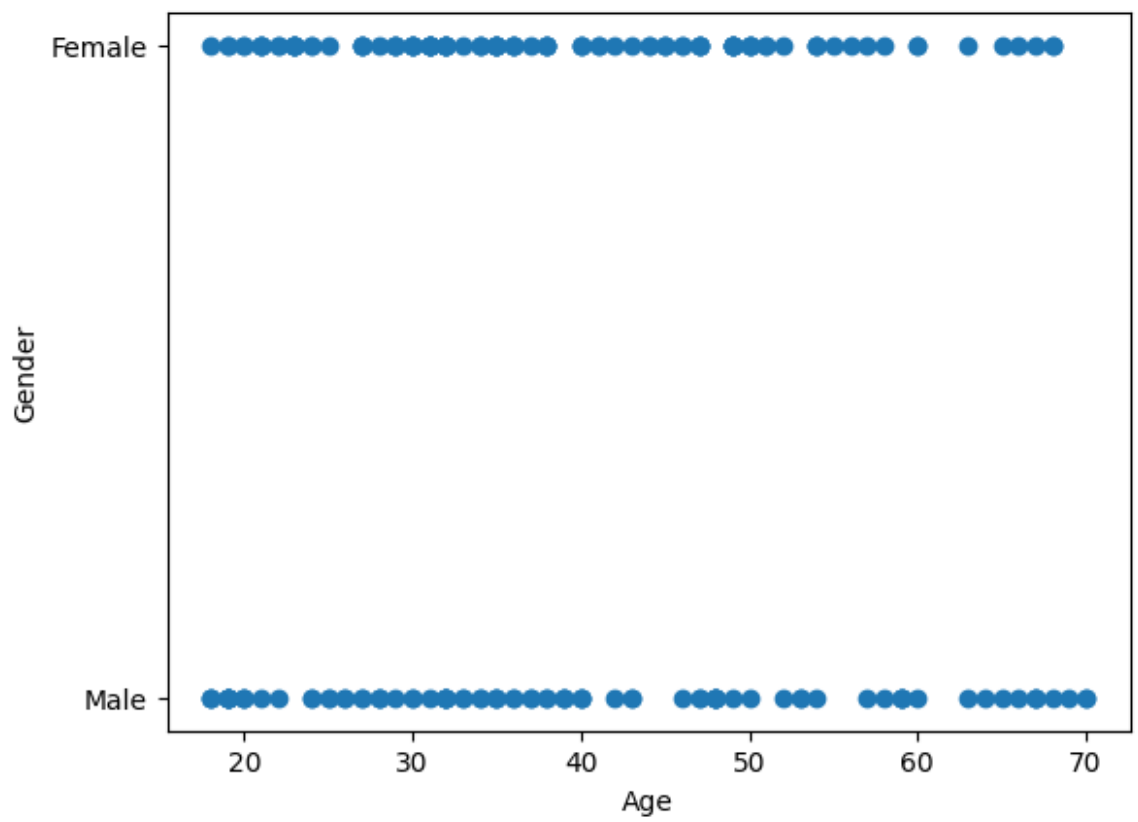
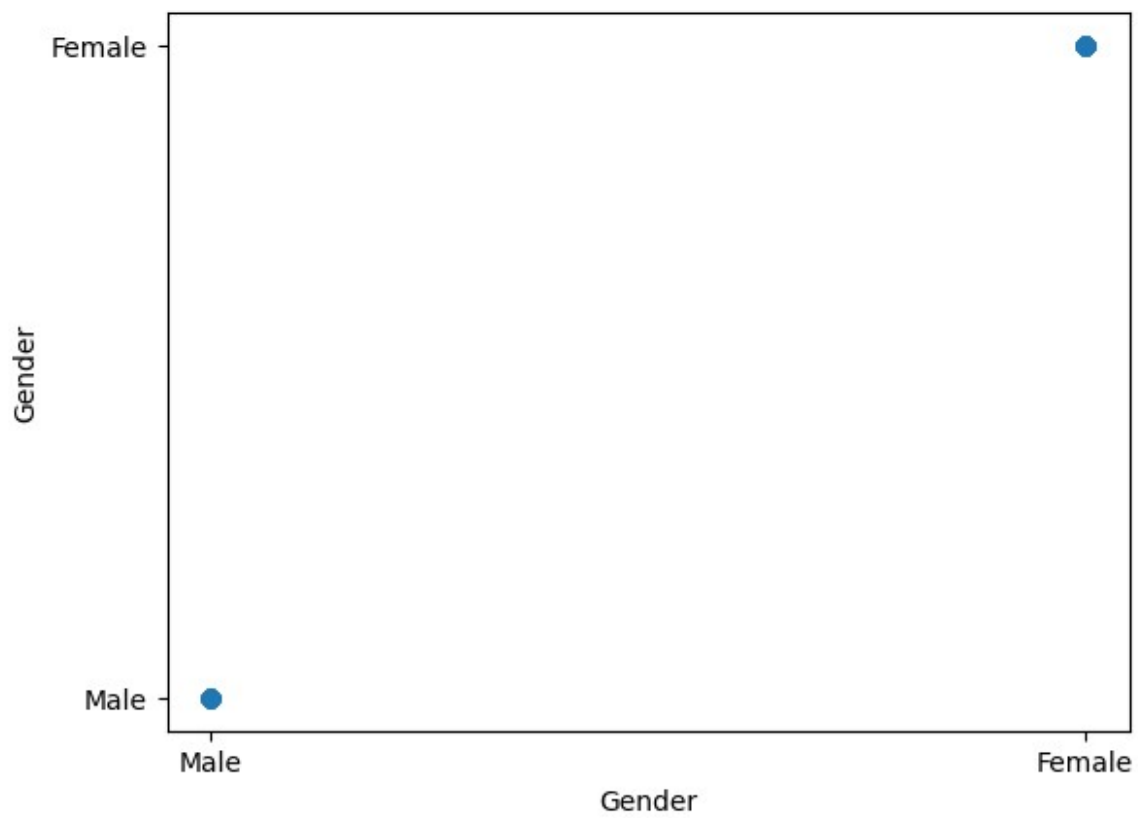


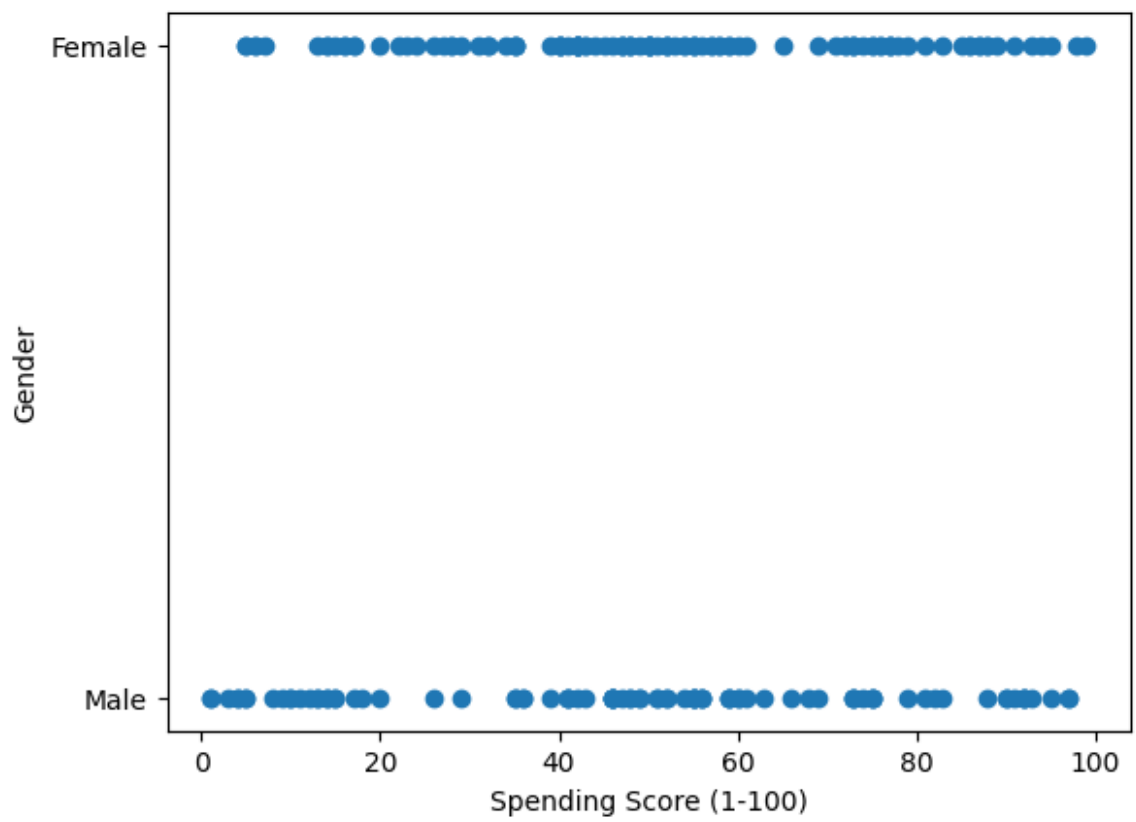
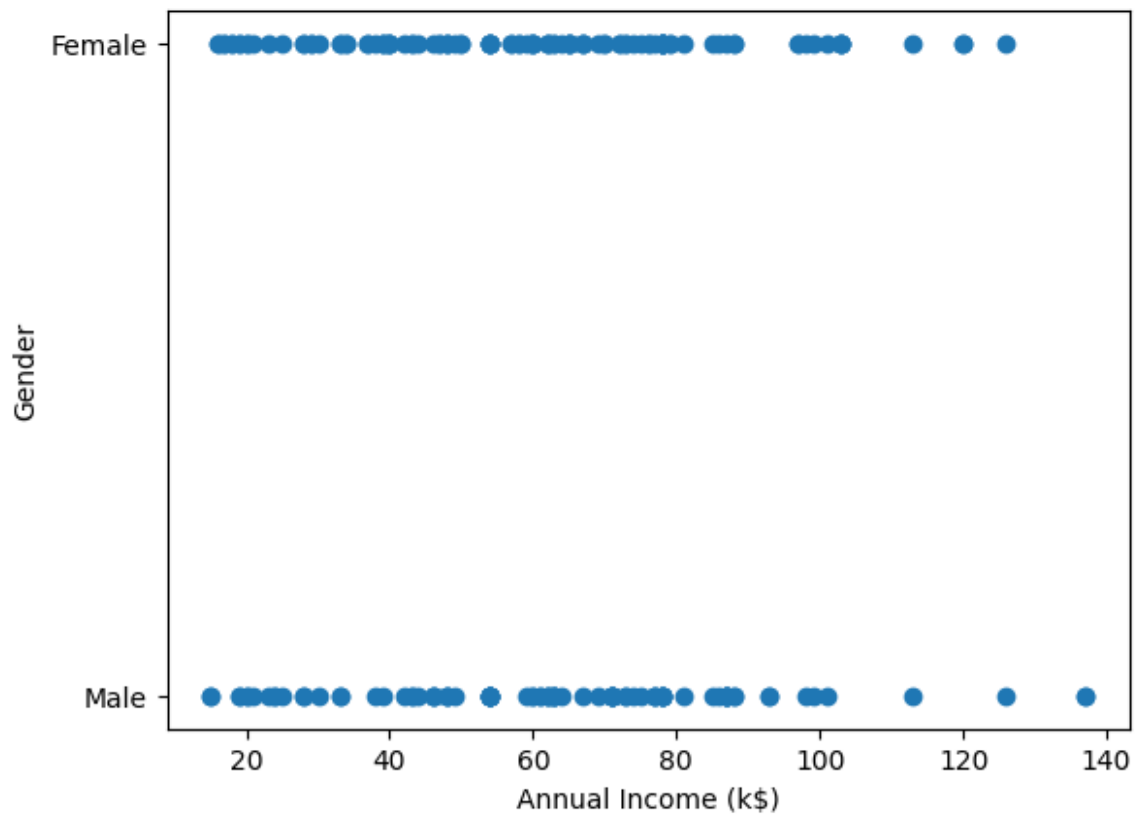


BIVARIATE ANALYSIS

```
for i in df.columns:  
    plt.scatter(df[i],df['Gender'])  
    plt.xlabel(i)  
    plt.ylabel("Gender")  
    plt.show()
```



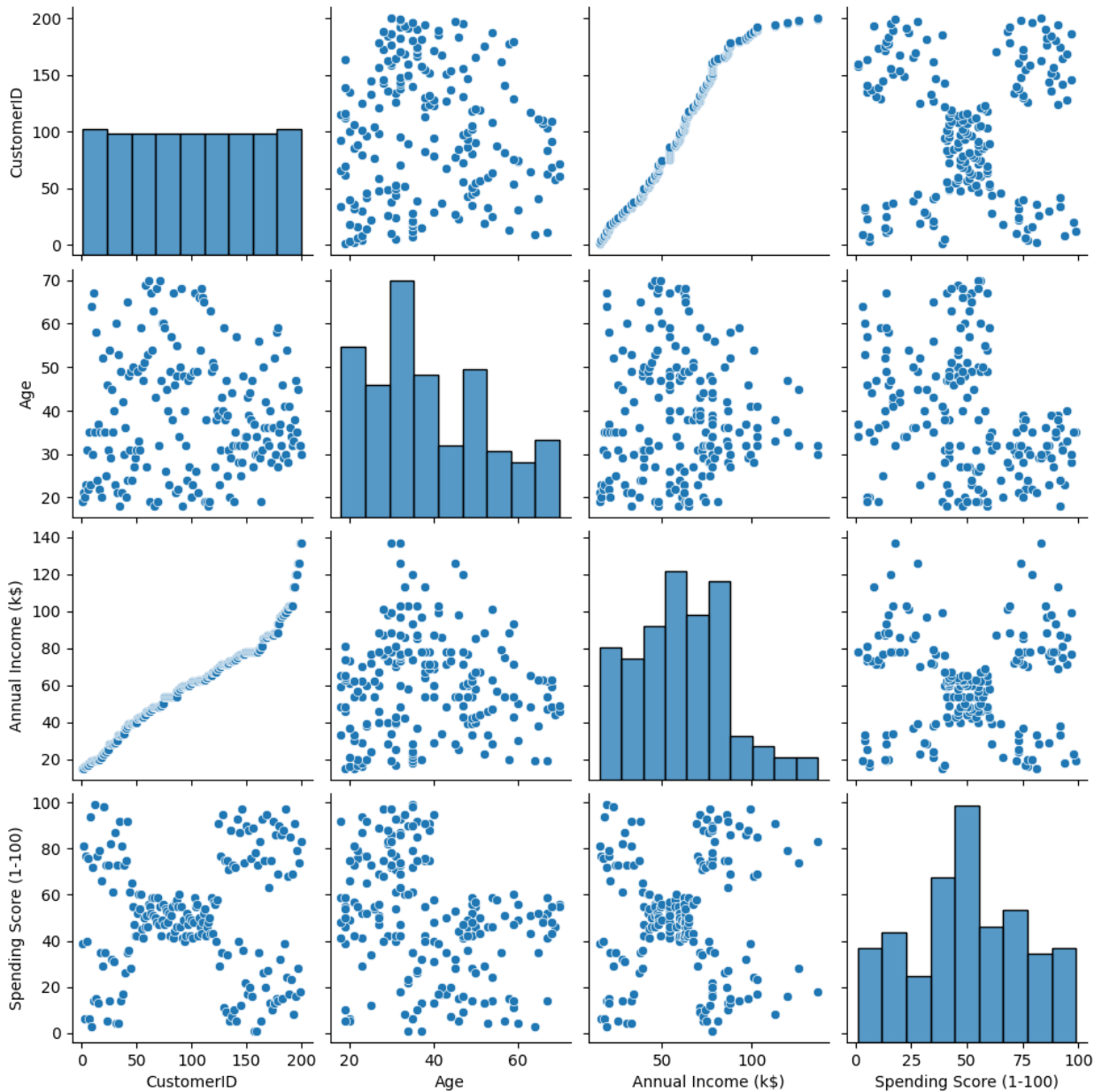




#MULTIVARIATE ANALYSIS

```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7db45e0865f0>
```



```
x = df.iloc[:,[3,4]]  
x.head()
```

	Annual Income (k\$)	Spending Score (1-100)
0	15	39
1	15	81

```
2          16          6
3          16         77
4          17         40
```

```
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-56-8df7bcac526d>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
```

```
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```



Machine Learning approach with K-Means Clustering Algorithm

```
from sklearn import cluster  
  
error=[]  
for i in range(1,11):  
    kmeans = cluster.KMeans(n_clusters=i,init = 'k-means+  
+',random_state=0)  
    kmeans.fit(x)  
    error.append(kmeans.inertia_)  
  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/  
_kmeans.py:870: FutureWarning: The default value of `n_init` will  
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly  
to suppress the warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n_init` will change from 10 to  
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the  
warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n_init` will change from 10 to  
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the  
warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n_init` will change from 10 to  
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the  
warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n_init` will change from 10 to  
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the  
warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n_init` will change from 10 to  
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the  
warning  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870  
: FutureWarning: The default value of `n init` will change from 10 to
```

```

'auto' in 1.4. Set the value of `n_init` explicitly to suppress the
warning
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870
: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the
warning
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870
: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the
warning
    warnings.warn(

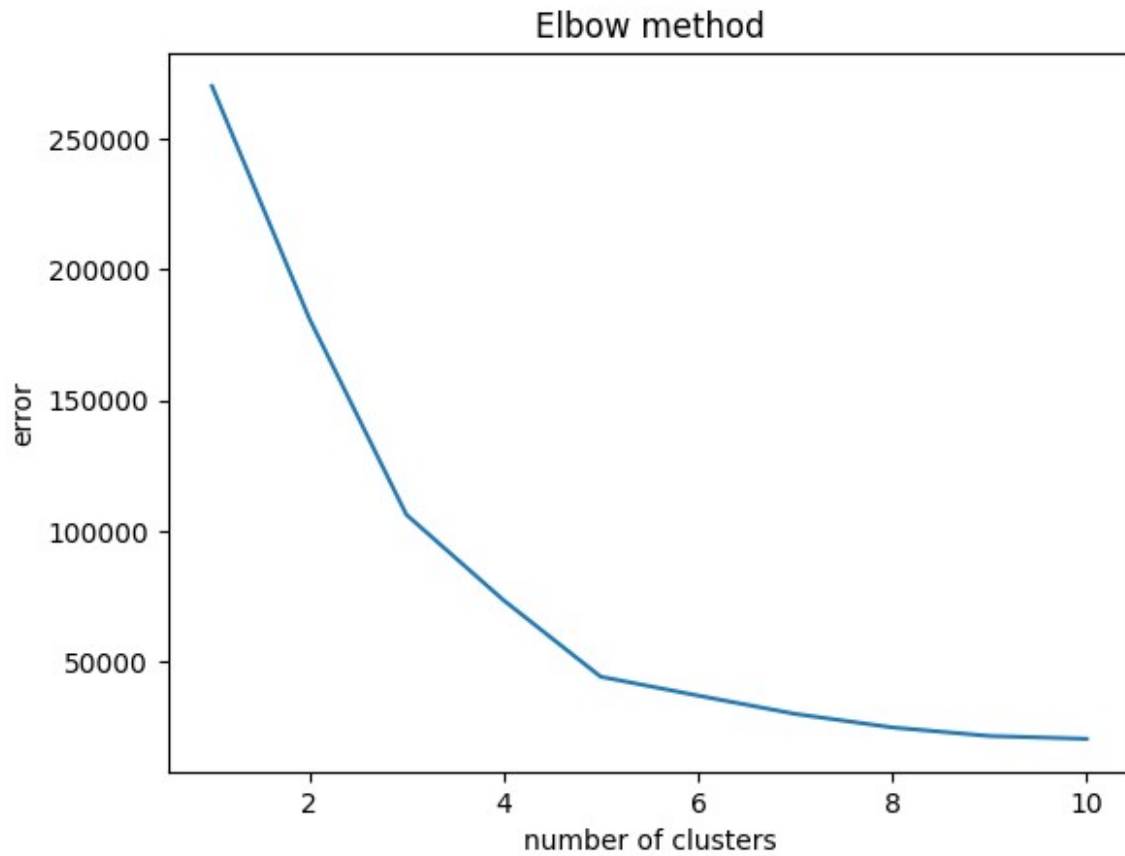
error

[269981.28,
 181363.59595959593,
 106348.37306211122,
 73679.78903948836,
 44448.4554479337,
 37265.86520484346,
 30259.65720728547,
 25095.70320999756,
 21830.041978049434,
 20736.679938924128]

import matplotlib.pyplot as plt

plt.plot(range(1,11),error)
plt.title('Elbow method')
plt.xlabel('number of clusters')
plt.ylabel('error')
plt.show()

```

[illegible]

```

1,      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 0, 2, 1, 2, 0, 2, 0,
2,      1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0,
2,      0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0,
2,      0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0,
2,      0, 2], dtype=int32)

```

TESTING WITH RANDOM VALUES

```

km_model.predict([[4.3,4.4]])

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but KMeans was
fitted with feature names
  warnings.warn(

array([4], dtype=int32)

km_model.predict([[6.1,7.2]])

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but KMeans was
fitted with feature names
  warnings.warn(

array([4], dtype=int32)

```