

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv("Mall_Customers.csv")
df.head()
```

```
Out[1]:
```

	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

```
In [2]: df.describe()
```

```
Out[2]:
```

	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

```
In [3]: df.isnull().sum()
```

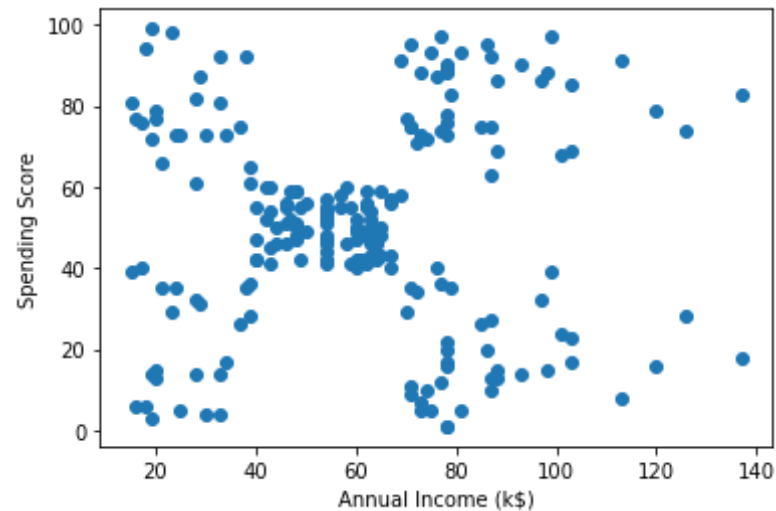
```
Out[3]: CustomerID      0
Gender      0
Age      0
```

```
Annual Income (k$)      0
Spending Score (1-100)  0
dtype: int64
```

```
In [24]: x = df.iloc[:,[3,4]].values
x[0:5]
```

```
Out[24]: array([[15, 39],
               [15, 81],
               [16,  6],
               [16, 77],
               [17, 40]], dtype=int64)
```

```
In [25]: plt.scatter(x[:,0],x[:,1])
plt.xlabel("Annual Income (k$)")
plt.ylabel("Spending Score")
plt.show()
```

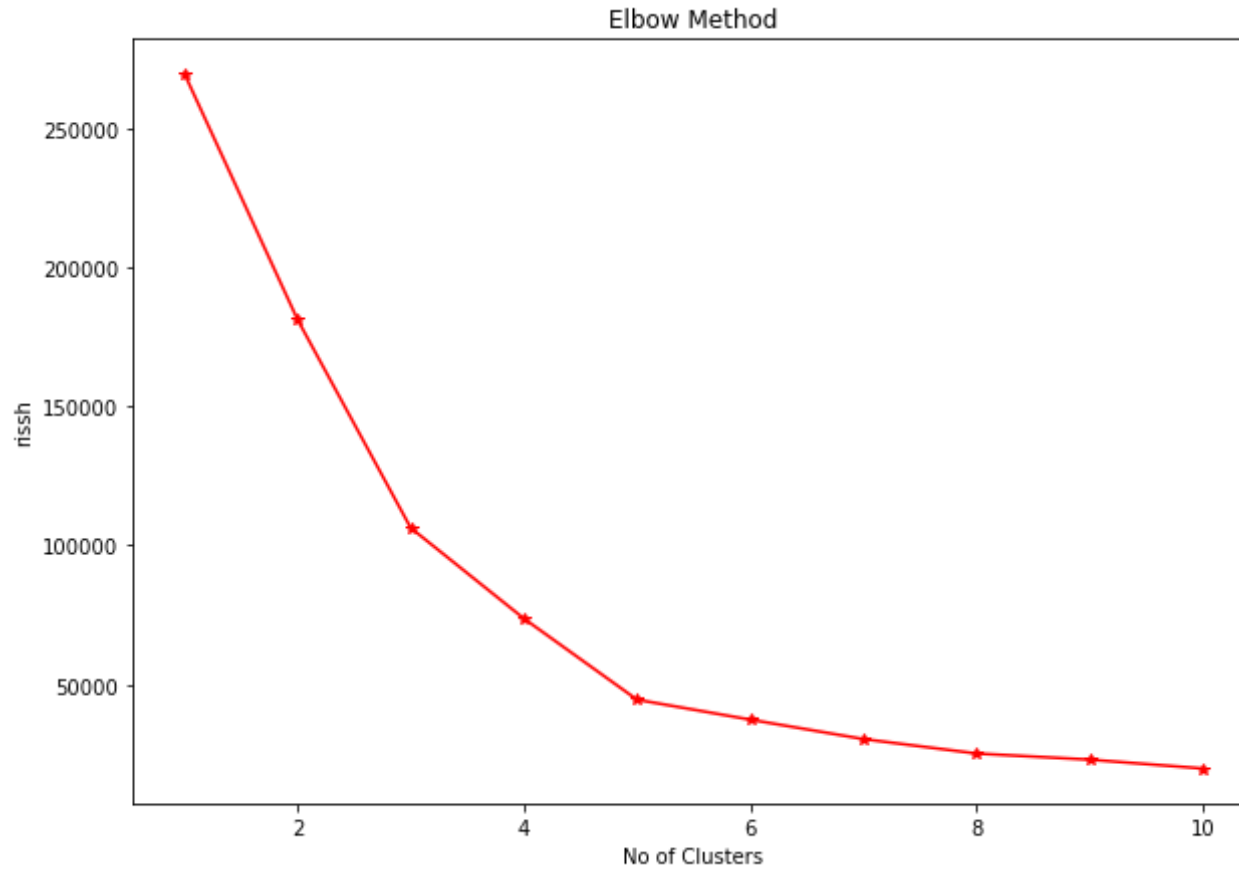


```
In [37]: from sklearn.cluster import KMeans
rissh = []
for i in range(1,11):
    kmeans = KMeans(n_clusters=i, init="k-means++")
    kmeans.fit(x)
    rissh.append(kmeans.inertia_)
```

In [38]: rissh

```
Out[38]: [269981.280000000014,  
181363.59595959607,  
106348.37306211119,  
73679.78903948837,  
44448.45544793369,  
37271.88623658948,  
30273.394312070028,  
25062.433792653766,  
22856.45429537046,  
19664.68519600554]
```

```
In [41]: plt.figure(figsize=(10,7))  
plt.plot(range(1,11),rissh,color="r",marker="*")  
plt.title("Elbow Method")  
plt.xlabel("No of Clusters")  
plt.ylabel("rissh")  
plt.show()
```



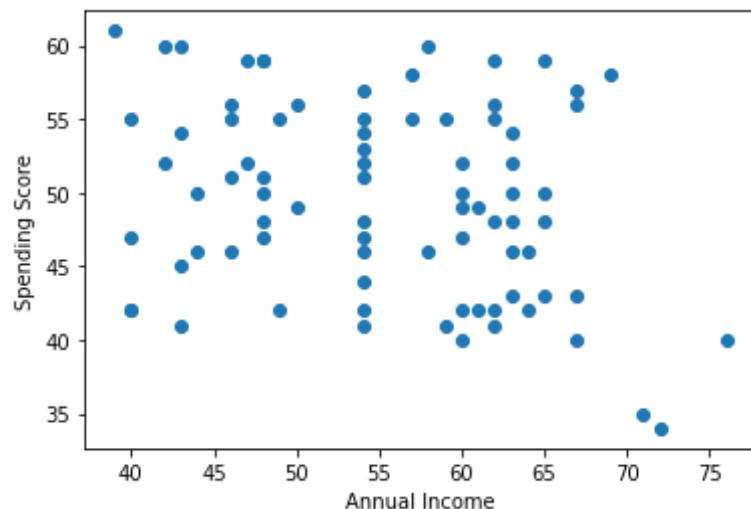
```
In [42]: k_means = KMeans(n_clusters=5,init="k-means++")
iit = k_means.fit(x)
```

```
In [43]: y_pred = iit.predict(x)
          y_pred
```

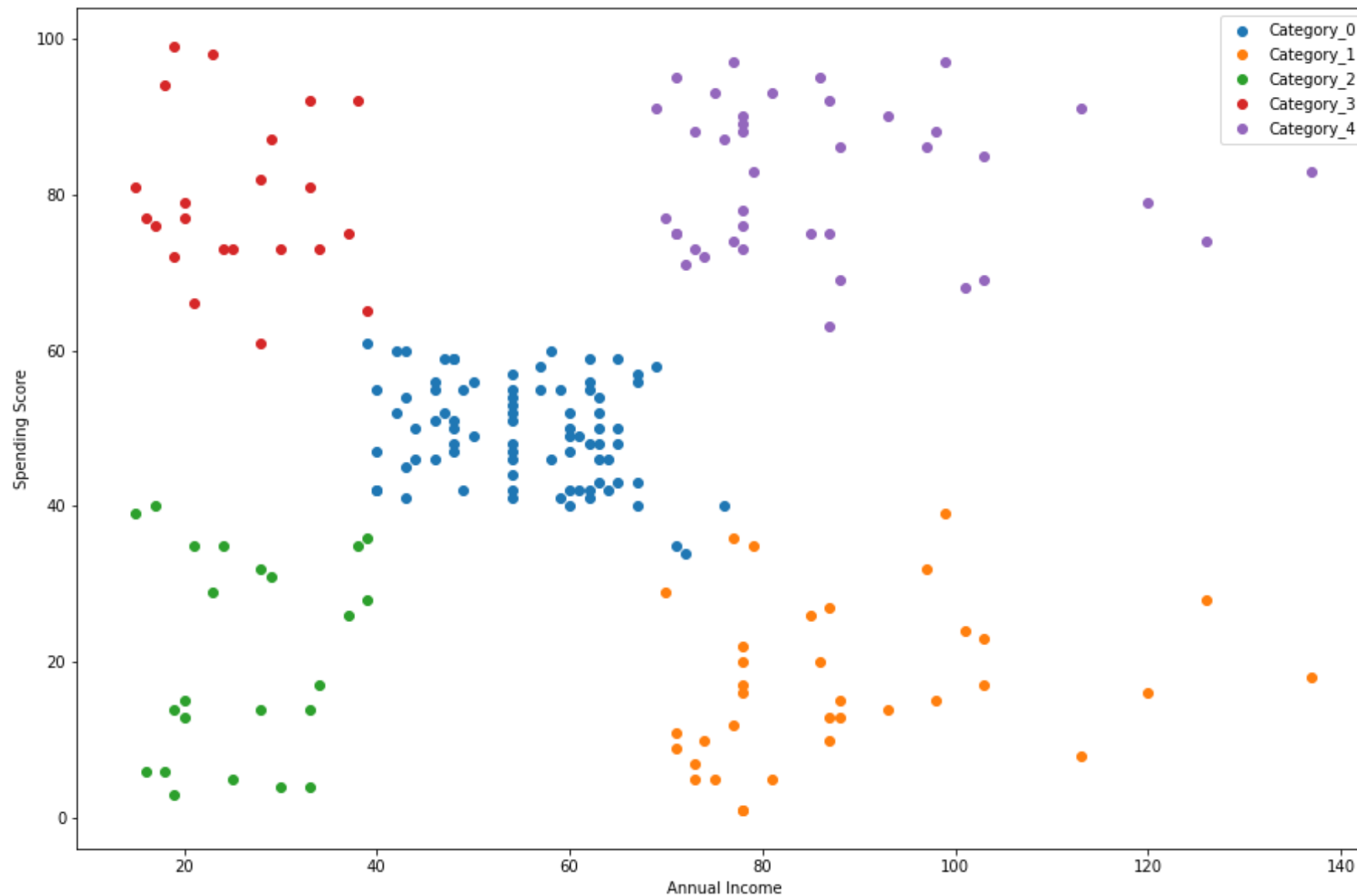
```
Out[43]: array([2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3,
                2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 0,
                2, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 1, 4, 0, 4, 1,
                4, 1, 4, 0, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4])
```

```
1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4,
1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4,
1, 4])
```

```
In [46]: plt.scatter(x[y_pred==0,0],x[y_pred==0,1])
plt.xlabel("Annual Income")
plt.ylabel("Spending Score")
plt.show()
```



```
In [50]: plt.figure(figsize=(15,10))
plt.scatter(x[y_pred==0,0],x[y_pred==0,1],label="Category_0")
plt.scatter(x[y_pred==1,0],x[y_pred==1,1],label="Category_1")
plt.scatter(x[y_pred==2,0],x[y_pred==2,1],label="Category_2")
plt.scatter(x[y_pred==3,0],x[y_pred==3,1],label="Category_3")
plt.scatter(x[y_pred==4,0],x[y_pred==4,1],label="Category_4")
plt.legend()
plt.xlabel("Annual Income")
plt.ylabel("Spending Score")
plt.show()
```



```
In [51]: df["Customer Category"] = y_pred
```

```
In [53]: df.head(10)
```

Out[53]:

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)	Customer Category
0	1	Male	19	15	39	2
1	2	Male	21	15	81	3
2	3	Female	20	16	6	2
3	4	Female	23	16	77	3
4	5	Female	31	17	40	2
5	6	Female	22	17	76	3
6	7	Female	35	18	6	2
7	8	Female	23	18	94	3
8	9	Male	64	19	3	2
9	10	Female	30	19	72	3

In [ ]: