

In [53]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.show()
```

In [50]:

```
pwd
```

Out[50]:

```
'C:\\Users\\RUSHIKESH SUNIL'
```

In [51]:

```
df = pd.read_csv("C:/Users/RUSHIKESH SUNIL/Documents/Internship/customers.csv")
df.head()
```

Out[51]:

	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 [31]:

```
df.shape
```

Out[31]:

```
(200, 5)
```

In [32]:

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

Out[32]:

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

In [37]:

```
X = df.iloc[:, [3,4]].values
X
```

Out[37]:

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
       [ 20,  15],
       [ 20,  77],
       [ 20,  13],
       [ 20,  79],
       [ 21,  35],
       [ 21,  66],
       [ 23,  29],
       [ 23,  98],
       [ 24,  35],
       [ 24,  73],
       [ 25,   5],
       [ 25,  73],
       [ 28,  14],
       [ 28,  82],
       [ 28,  32],
       [ 28,  61],
       [ 29,  31],
       [ 29,  87],
       [ 30,  41],
```

```
[ 30, 73],  
[ 33,  4],  
[ 33, 92],  
[ 33, 14],  
[ 33, 81],  
[ 34, 17],  
[ 34, 73],  
[ 37, 26],  
[ 37, 75],  
[ 38, 35],  
[ 38, 92],  
[ 39, 36],  
[ 39, 61],  
[ 39, 28],  
[ 39, 65],  
[ 40, 55],  
[ 40, 47],  
[ 40, 42],  
[ 40, 42],  
[ 42, 52],  
[ 42, 60],  
[ 43, 54],  
[ 43, 60],  
[ 43, 45],  
[ 43, 41],  
[ 44, 50],  
[ 44, 46],  
[ 46, 51],  
[ 46, 46],  
[ 46, 56],  
[ 46, 55],  
[ 47, 52],  
[ 47, 59],  
[ 48, 51],  
[ 48, 59],  
[ 48, 50],  
[ 48, 48],  
[ 48, 59],  
[ 48, 47],  
[ 49, 55],  
[ 49, 42],  
[ 50, 49],  
[ 50, 56],  
[ 54, 47],  
[ 54, 54],  
[ 54, 53],  
[ 54, 48],  
[ 54, 52],  
[ 54, 42],  
[ 54, 51],  
[ 54, 55],  
[ 54, 41],  
[ 54, 44],  
[ 54, 57],  
[ 54, 46],  
[ 57, 58],  
[ 57, 55],  
[ 58, 60],  
[ 58, 46],  
[ 59, 55],  
[ 59, 41],  
[ 60, 49],  
[ 60, 40],  
[ 60, 42],  
[ 60, 52],  
[ 60, 47],  
[ 60, 50],  
[ 61, 42],  
[ 61, 49],  
[ 62, 41],  
[ 62, 48],  
[ 62, 59],  
[ 62, 55],  
[ 62, 56],  
[ 62, 42],  
[ 63, 50],  
[ 63, 46].
```

[63, 43],
[63, 48],
[63, 52],
[63, 54],
[64, 42],
[64, 46],
[65, 48],
[65, 50],
[65, 43],
[65, 59],
[67, 43],
[67, 57],
[67, 56],
[67, 40],
[69, 58],
[69, 91],
[70, 29],
[70, 77],
[71, 35],
[71, 95],
[71, 11],
[71, 75],
[71, 9],
[71, 75],
[72, 34],
[72, 71],
[73, 5],
[73, 88],
[73, 7],
[73, 73],
[74, 10],
[74, 72],
[75, 5],
[75, 93],
[76, 40],
[76, 87],
[77, 12],
[77, 97],
[77, 36],
[77, 74],
[78, 22],
[78, 90],
[78, 17],
[78, 88],
[78, 20],
[78, 76],
[78, 16],
[78, 89],
[78, 1],
[78, 78],
[78, 1],
[78, 73],
[79, 35],
[79, 83],
[81, 5],
[81, 93],
[85, 26],
[85, 75],
[86, 20],
[86, 95],
[87, 27],
[87, 63],
[87, 13],
[87, 75],
[87, 10],
[87, 92],
[88, 13],
[88, 86],
[88, 15],
[88, 69],
[93, 14],
[93, 90],
[97, 32],
[97, 86],
[98, 15],
[98, 88],
[99, 39].

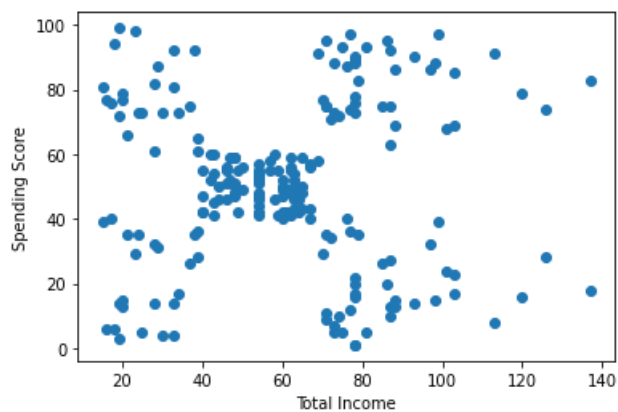
```
[ 99, 97],
[101, 24],
[101, 68],
[103, 17],
[103, 85],
[103, 23],
[103, 69],
[113, 8],
[113, 91],
[120, 16],
[120, 79],
[126, 28],
[126, 74],
[137, 18],
[137, 83]], dtype=int64)
```

```
plt.scatter(X[... ,0],X[... ,1])
plt.xlabel('Total Income')
plt.ylabel('Spending Score')
plt.show
```

In [40]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

Out[40]:



In [43]:

```
from sklearn.cluster import KMeans
wcss = []
for i in range (1,11):
    kmeans = KMeans(n_clusters=i,init='k-means++',random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

In [44]:

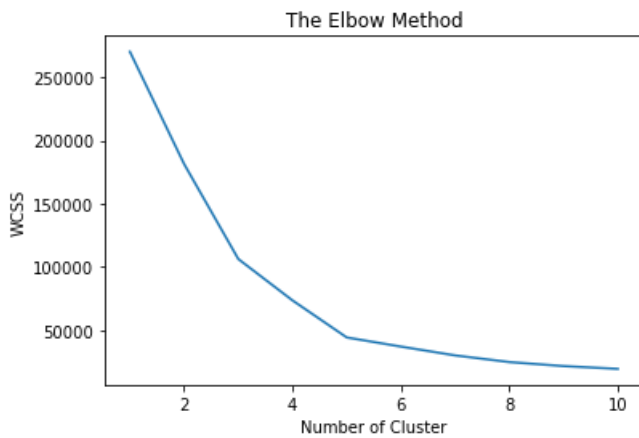
```
wcss
```

Out[44]:

```
[269981.280000000014,
181363.59595959607,
106348.37306211119,
73679.78903948837,
44448.45544793369,
37239.83554245604,
30273.394312070028,
25018.576334776328,
21850.16528258562,
19664.68519600554]
```

In [56]:

```
plt.plot(range(1,11),wcss)
plt.title("The Elbow Method")
plt.xlabel("Number of Cluster")
plt.ylabel("WCSS")
plt.show()
```



In [58]:

```
KMeans(n_clusters=5,init='k-means++',random_state=0)
Y_kmeans = kmeans.fit_predict(X)
Y_kmeans
```

Out[58]:

```
array([[7, 0, 4, 0, 7, 0, 4, 0, 4, 0, 4, 0, 4, 0, 7, 0, 7, 0, 7, 0,
        4, 0, 4, 0, 7, 0, 7, 0, 4, 0, 4, 0, 4, 0, 4, 0, 7, 0, 7, 0, 7, 8,
        7, 8, 8, 8, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8,
        8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 1, 8, 8, 1, 1, 8, 8, 8, 8,
        8, 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, 1, 6, 1, 2, 5, 6, 5, 6,
        1, 6, 5, 2, 5, 6, 5, 6, 5, 2, 1, 2, 5, 2, 1, 6, 5, 2, 5, 2, 5, 6,
        5, 2, 5, 6, 5, 6, 1, 2, 5, 2, 5, 6, 5, 2, 5, 6, 5, 6, 5, 2, 5, 2,
        5, 6, 5, 2, 3, 9, 3, 9, 3, 9, 3, 9, 3, 9, 3, 9, 3, 9, 3, 9, 3, 9,
        3, 9])
```

In [62]:

```
X[Y_kmeans==0,0]
```

Out[62]:

```
array([15, 16, 17, 18, 19, 19, 20, 20, 21, 23, 24, 25, 28, 28, 29, 30, 33,
        33, 34, 37, 38], dtype=int64)
```

In [63]:

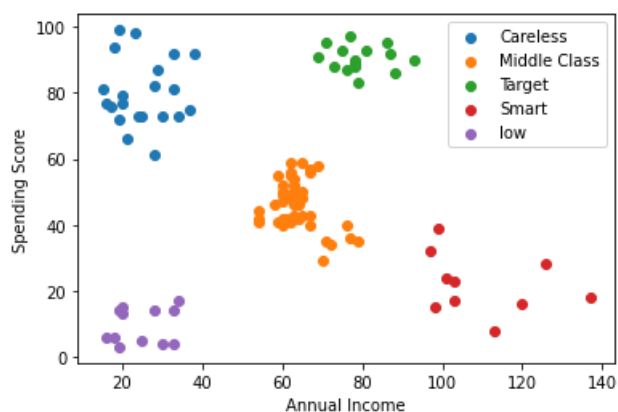
```
X[Y_kmeans==0,1]
```

Out[63]:

```
array([81, 77, 76, 94, 72, 99, 77, 79, 66, 98, 73, 73, 82, 61, 87, 73, 92,
        81, 73, 75, 92], dtype=int64)
```

In [79]:

```
plt.scatter(X[Y_kmeans==0,0],X[Y_kmeans==0,1],label = 'Careless')
plt.scatter(X[Y_kmeans==1,0],X[Y_kmeans==1,1],label = 'Middle Class')
plt.scatter(X[Y_kmeans==2,0],X[Y_kmeans==2,1],label = 'Target')
plt.scatter(X[Y_kmeans==3,0],X[Y_kmeans==3,1],label = 'Smart')
plt.scatter(X[Y_kmeans==4,0],X[Y_kmeans==4,1],label = 'low ')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.legend()
plt.show()
```



In [85]:

```
df['Target'] = Y_kmeans
```

In [83]:

```
df
```

Out[83]:

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)	Target
0	1	Male	19	15	39	7
1	2	Male	21	15	81	0
2	3	Female	20	16	6	4
3	4	Female	23	16	77	0
4	5	Female	31	17	40	7
...
195	196	Female	35	120	79	9
196	197	Female	45	126	28	3
197	198	Male	32	126	74	9
198	199	Male	32	137	18	3
199	200	Male	30	137	83	9

200 rows × 6 columns

In []:

In []: