```
In [9]: import matplotlib.pyplot as plt
In [10]: df=pd.read_csv("D:\Mall_Customers.csv")
In [11]: df
Out[11]:
           CustomerID Genre Age Annual Income (k$) Spending Score (1-100)
                  1 Male 19
                                         15
                                                         39
                                                         81
                  2 Male 21
                                         16
                                                         6
         2
                  3 Female 20
                  4 Female 23
                                                         77
                                         17
                                                         40
         4
                  5 Female 31
        195
                                        120
                                                         79
                 196 Female 35
        196
                 197 Female 45
                                        126
                                                         28
        197
                      Male 32
                                        126
                                                         74
        198
                      Male 32
                                        137
                                                         18
                                        137
                                                         83
        199
                200 Male 30
       200 rows × 5 columns
In [12]: x=df.iloc[:,3:]
In [17]: x
Out[17]:
           Annual Income (k$) Spending Score (1-100)
         0
                                      39
                      15
                                      81
         2
                                       6
                      16
                      16
                                      77
         4
                      17
                                      40
                                      79
        195
                     120
        196
                     126
                                      28
                                      74
        197
        198
                     137
                                      18
                                      83
        199
       200 rows × 2 columns
In [18]: plt.title("Unclustered Data")
       plt.xlabel("Annual Income")
       plt.ylabel("Spending Score")
       plt.scatter(x["Annual Income (k$)"],x["Spending Score (1-100)"])
Out[18]: <matplotlib.collections.PathCollection at 0x1aee8f60340>
                               Unclustered Data
         100
       Spending Score
          20
                                               100
                                                       120
                                                               140
                20
                        40
                                        80
                                  Annual Income
In [19]: from sklearn.cluster import KMeans, AgglomerativeClustering
In [20]: Km=KMeans(n_clusters=3)
In [21]: x.shape
Out[21]: (200, 2)
In [22]: Km.fit_predict(x)
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 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, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0,
              2, 0], dtype=int32)
In [23]: Km.inertia_
Out[23]: 106348.37306211118
In [26]: sse=[]
In [27]: for K in range(1,16):
           Km=KMeans(n_clusters=K)
           Km.fit_predict(x)
           sse.append(Km.inertia_)
Out[28]: [269981.28,
        183653.32894736843,
        132162.13629759202,
         73880.64496247194,
         44454.47647967974,
         38797.9027638142,
         30566.45113025186,
         25315.541822712166,
        23483.671887891876,
        21367.132190948367,
        19477.555130449615,
        17749.87056277056,
        14450.366492110039,
         13304.030806934612,
         12065.740778081567]
In [30]: plt.title("Elbow Method")
       plt.xlabel("Value of K")
       plt.ylabel("SSE")
       plt.grid()
       plt.xticks(range(1,16))
       plt.plot(range(1,16),sse,marker=".",color="red")
Out[30]: [<matplotlib.lines.Line2D at 0x1ae8d80ea10>]
                                    Elbow Method
         250000
         200000
      150000 ·
         100000
          50000
                 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
                                      Value of K
In [31]: from sklearn.metrics import silhouette_score
In [32]: silh=[]
In [33]: for K in range(2,16):
           Km=KMeans (n_clusters=K)
           labels=Km.fit_predict(x)
           score=silhouette_score(x,labels)
           silh.append(score)
In [34]: plt.title("Silhout Method")
       plt.xlabel("Value of K")
       plt.ylabel("Silhout Score")
       plt.grid()
       plt.xticks(range(2,16))
       plt.bar(range(2,16),silh,color="red")
Out[34]: <BarContainer object of 14 artists>
                                Silhout Method
         0.5
         0.4
       Silhout Score
         0.1
               2 3 4 5 6 7 8 9 10 11 12 13 14 15
                                   Value of K
In [35]: Km=KMeans(n_clusters=5, random_state=0)
In [36]: labels=Km.fit_predict(x)
In [39]: labels
3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 0,
              0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 1, 2, 1,
              0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
              2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
              2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
              2, 1], dtype=int32)
In [40]: plt.figure(figsize=(16,9))
        cent=Km.cluster_centers_
       plt.subplot(1,2,1)
       plt.title("Clustered Data")
       plt.xlabel("Annual Income")
       plt.ylabel("Spending Score")
       plt.scatter(x["Annual Income (k$)"],x["Spending Score (1-100)"],c=labels)
       plt.scatter(cent[:,0],cent[:,1],s=100,color="k")
Out[40]: <matplotlib.collections.PathCollection at 0x1ae8da64b50>
                                    Clustered Data
         100
          80
          60
       Spending Score
```

In [8]: **import** pandas **as** pd

20

```
Annual Income
In [41]: Km.inertia_
Out[41]: 44448.45544793371
In [42]: Km.labels_
3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 0,
         0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 1, 2, 1,
         0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
         2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
```

60

80

100

140

120

0

20

40

df[	labels==	4]			
•	Custome	rID Genre	Age	Annual Income (k\$) Spending S	core (1-100)
	1	2 Male	21	15	81
	3	4 Female	23	16	77
	5	6 Female	22	17	76
	7	8 Female	23	18	94
	9	10 Female	30	19	72
	11	12 Female	35	19	99
		14 Female	24	20	77
	15	16 Male	22	20	79
		18 Male	20	21	66
•	19	20 Female	35	23	98
2	21	22 Male	25	24	73
2	23	24 Male	31	25	73
2	25	26 Male	29	28	82
2	27	28 Male	35	28	61
2	29	30 Female	23	29	87
3	31	32 Female	21	30	73
3	33	34 Male	18	33	92
3	35	36 Female	21	33	81
3	37	38 Female	30	34	73
3	39	40 Female	20	37	75
		42 Male	24	38	92
4	45	46 Female	24	39	65
, ,	Km.predict	V. 106 7011)			

Out[44]: array([4], dtype=int32)