```
In [14]: import pandas as pd
                  import numpy as np
                  import seaborn as sns
                  import matplotlib.pyplot as plt
                  import warnings
                  warnings.filterwarnings("ignore")
In [15]: df = pd.read_csv("Kmeans data.csv")
In [16]: df.head()
                       CustomerID Genre Age Annual Income (k$) Spending Score (1-100)
Out[16]:
                                              Male
                                                                                                                          39
                                                        19
                                                                                       15
                                                                                       15
                                                                                                                          81
                                      2 Male
                                                       21
                                      3 Female
                                                                                       16
                                                                                       16
                                                                                                                          77
                                                        23
                                      4 Female
                                                                                       17
                                     5 Female 31
                                                                                                                           40
In [17]: df.shape
                  (200, 5)
Out[17]:
In [18]: df.columns
Out[18]: Index(['CustomerID', 'Genre', 'Age', 'Annual Income (k$)',
                                'Spending Score (1-100)'],
                             dtype='object')
In [19]: df1 = df.drop(columns=['CustomerID', 'Genre', 'Age'])
In [20]: df1.head()
                       Annual Income (k$) Spending Score (1-100)
Out[20]:
                                                                                 39
                                              15
                                              15
                                                                                 81
                  2
                                              16
                                                                                   6
                                              16
                                                                                  77
                                              17
                                                                                 40
In [21]: sns.scatterplot(data = df1, x='Annual Income (k$)',y='Spending Score (1-100)')
                  <AxesSubplot:xlabel='Annual Income (k$)', ylabel='Spending Score (1-100)'>
Out[21]:
                      100
                  Spending Score (1-100)
                        60
                        40 -
                        20 -
                                  20
                                                                      80
                                                                                             120
                                                                                                        140
                                                                                100
                                                         Annual Income (k$)
In [22]: from sklearn.cluster import KMeans
 In [23]: wcss = []
                  for i in range(1,10):
                          kmeans = KMeans(i)
                          kmeans.fit(df1)
                          wcss.append(kmeans.inertia_)
                  number_clusters = range(1,10)
                  plt.plot(number_clusters,wcss)
                  plt.title('The Elbow title')
                  plt.xlabel('Number of clusters')
                  plt.ylabel('WCSS')
Out[23]: Text(0, 0.5, 'WCSS')
                                                               The Elbow title
                       250000
                       200000
                   S 150000
                       100000
                        50000
                                                              Number of clusters
In [28]: # selecting no. of clusters to be 3
                   Kmeans = KMeans(5)
                  Kmeans.fit(df1)
                  KMeans(n_clusters=5)
In [29]: Kmeans.labels_
Out[29]: array([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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 1, 4, 1, 4,
                               4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 2,
                               2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 0, 3, 2, 3, 0, 3, 0, 3,
                               2, 3, 0, 3, 0, 3, 0, 3, 0, 3, 2, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3,
                               0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3,
                               0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3, 0, 3,
                               0, 3])
```