16/9/25

Ex : 7

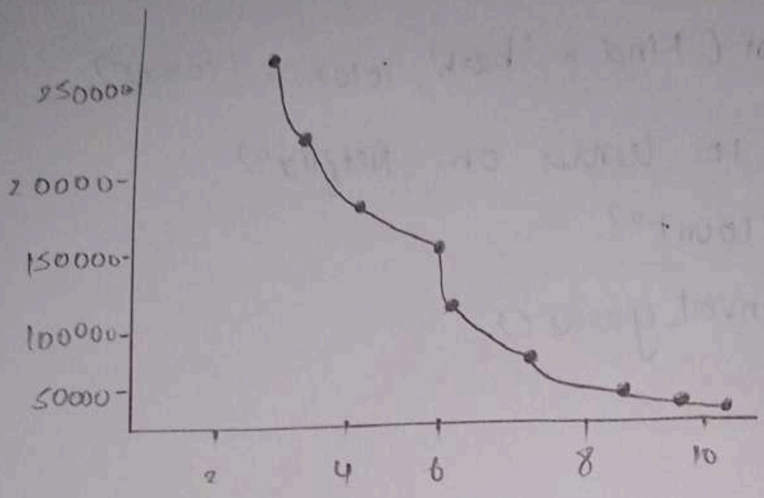
clustering

AIM :

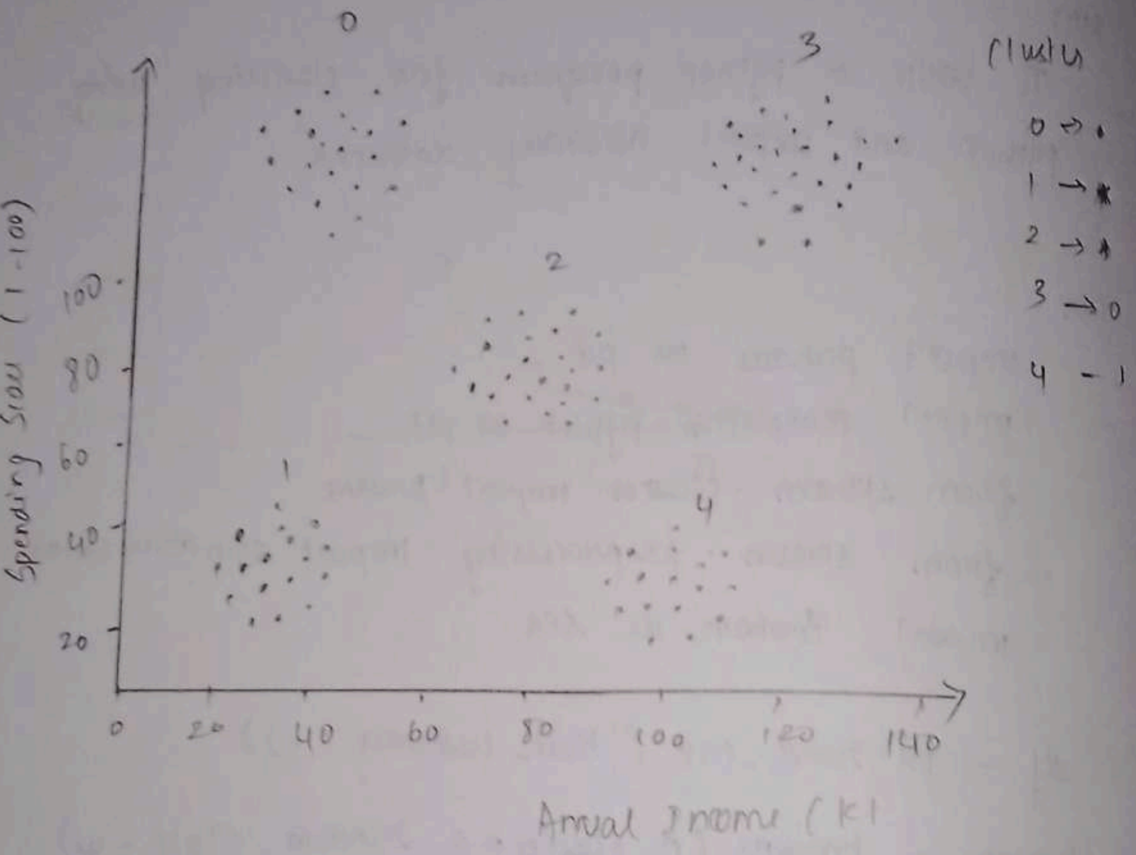
To write a python program for clustering using python and import necessary dataset.

CODE :

```
import pandas as pd
import matplotlib . pyplot as plt
from sklearn . cluster import kmeans
from sklearn . preprocessing import standardscalar
import seaborn as sns .

df = pd . read_csv ('Mall_customer . csv)
kmeans = kmeans (n_clusters = 5, random_state = 42)
df ['cluster'] = kmeans . fit_predicted (df ['AnnualIncome
                                              (K$)
                              spending . score)

disortions = [ ]
for i in range (1,11):
    km = kmeans (n_cluster = i)
    km . fit (df ['annual income (K$) ' , '
                       spending score (1-100)']])

    disortion . append (km . inertia)
  plt . plot (range (1,11) , disortion, marker = 'o')
    plt . title ('Elbow method')
    plt . xlabel (inertia)
    plt . ylabel (No . of clusters)
```


250000
20000
150000
100000
50000
2
4
6
8
10

```
from sklearn . metrices import silhoutte . score
frome sklearn . cluster import spectral clustering
import matplotlib . pyplot as plt
import numpy as np
import pandas as pd

# Load data
   wine = load_wine()
     x = pd. dataframe ( wine . data, column = wine.
                                        feature_names)
     x_scaled = standardscale() . fit_transform(x)

 # Generate base clustering
     base_clustering = []
   for k in [2,4,5]:
        km = kmeans (n_cluster = x . random_stab=42)
        base_clustering . append (km. fit_predicted
                                          (x_scaled))

  # Apply ensemble
           ensemble_labels = cspa_ensemble (base_clusterin

  # Evaluate
        print (" silhoutte score : " silhoutte_score())

 # plot clusters
        plt . figure ( fig_size = (10,6))
       plt . scatter ( x_pca [:,0] , x_pca [:,1],
                   c= ensemble_labels,
             cmap= 'viridis', s=50, edge color
```


0
3
Cluster
0 → •
1 → *
2 → ▲
3 → o
4 - 1
2
1
4
Spending Score (1-100)
100
80
60
40
20
0
20
40
60
80
100
120
140
Annual Income (K)

```
plt.title ("CSPA ensembling cluster on wine Dataset")
plt. xlabel ("PCA component 1")
plt. ylabel ("PCA component 2")
plt. colorbar (label = 'cluster label')
plt. grid (True)
plt. show()
```

Result:

Therefore the required program for clustering has been executed successfully.