

### **PRACTICAL NUMBER: 7**

**Aim:** Implement hierarchical clustering using Dendrogram

**Code:**

```
import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

import scipy.cluster.hierarchy as shc

from sklearn.cluster import AgglomerativeClustering


customer_data = pd.read_csv('shopping_data.csv')

#shape: Returns tuple of shape (Rows, columns) of dataframe/series

print("Rows,columns in data:",customer_data.shape)

#head: Displays the first five rows of the dataframe by default.

print("First five rows:",customer_data.head())

#iloc: helps us to select a specific row or column from the data set.

data = customer_data.iloc[:,3:5].values

plt.figure(figsize=(10,7))

plt.title("Customer Dendograms")

# create a dendrogram variable linkage is actually the algorithm itself of hierarchical clustering and then

# in linkage we have to specify on which data we apply and engage.

# Ward method is actually a method that tries to minimize the variance within each cluster.

# We choose Euclidean distance and ward method for our algorithm class.

dend = shc.dendrogram(shc.linkage(data,method = 'ward'))

cluster = AgglomerativeClustering(n_clusters=5, affinity='euclidean', linkage='ward')

cluster.fit_predict(data)

plt.figure(figsize=(10, 7))

plt.scatter(data[:,0], data[:,1], c=cluster.labels_,cmap='rainbow')

plt.show()
```

## shopping\_data.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	CustomerID	Genre	Age	Annual Inc	Spending Score (1-100)													
2	1	Male	19	15	39													
3	2	Male	21	15	81													
4	3	Female	20	16	6													
5	4	Female	23	16	77													
6	5	Female	31	17	40													
7	6	Female	22	17	76													
8	7	Female	35	18	6													
9	8	Female	23	18	94													
10	9	Male	64	19	3													
11	10	Female	30	19	72													
12	11	Male	67	19	14													
13	12	Female	35	19	99													
14	13	Female	58	20	15													
15	14	Female	24	20	77													
16	15	Male	37	20	13													
17	16	Male	22	20	79													
18	17	Female	35	21	35													
19	18	Male	20	21	66													
20	19	Male	52	23	29													
21	20	Female	35	23	98													
22	21	Male	35	24	35													
23	22	Male	25	24	73													
24	23	Female	46	25	5													
25	24	Male	31	25	73													
26	25	Female	54	28	14													
27	26	Male	29	28	82													

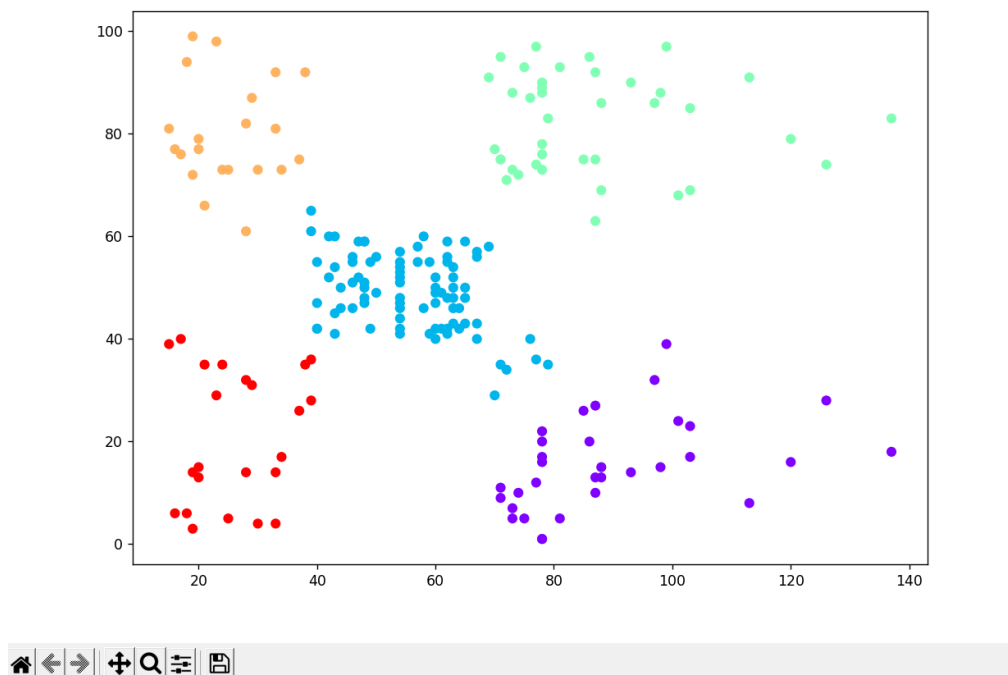
## Output:

```
Python 3.7.9 Shell*
File Edit Shell Debug Options Window Help
Python 3.7.9 (tags/v3.7.9:13c94747c7, Aug 17 2020, 18:58:18) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: D:/MSc IT/Sem3/ML Pracs/Prac7.py =====
Rows,columns in data: (200, 5)
First five rows:
CustomerID  Genre  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
```

Figure 1



Figure 2



# 1. Explain clustering techniques with hierarchical clustering