Semester	T.E. Semester V – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	Lab 312 A

Student Name	Deep Salunkhe	
Roll Number	21102A00014	
Grade and Subject Teacher's Signature		

Experiment Number	06	
Experiment Title	Implementation of Hierarchical Clustering	
Resources / Apparatus Required	Hardware: Computer system	Software: Python
Description	**Dendrograms in Hierarchical Clustering:** Dendrograms are a visual representation commonly used in hierarchical clustering analysis. Hierarchical clustering is a method for grouping similar data points together in a hierarchical manner. Dendrograms provide a way to visualize the hierarchical structure of these groupings. **Components of a Dendrogram:** ● **Leaves**: The leaves of a dendrogram represent individual data points or observations. Each data point is depicted as a leaf node. ● **Nodes:** The points at which branches merge in a dendrogram are called nodes. Nodes represent the merging of clusters or groups of data points. ● **Branches:** Branches in a dendrogram connect nodes and represent the merging of clusters or the distance between clusters. **Construction of a Dendrogram:** Dendrograms are constructed based on the linkage method chosen for hierarchical clustering. The most common linkage methods include:	

- **Single Linkage:** This method merges clusters based on the minimum pairwise distance between their members. It often leads to long, "stringy" clusters.

- **Complete Linkage:** This method merges clusters based on the maximum pairwise distance between their members. It tends to create compact, spherical clusters.

- **Average Linkage:** This method merges clusters based on the average pairwise distance between their members.

- **Ward's Linkage:** This method minimizes the increase in the sum of squared distances when merging clusters. It tends to create balanced and compact clusters.

Interpreting a Dendrogram:

Interpreting a dendrogram involves understanding how clusters are formed and the distances between clusters. Some key points to consider when interpreting a dendrogram include:

- **Height of Nodes:** The height at which nodes merge in a dendrogram represents the distance or dissimilarity between the merged clusters. Longer branches indicate greater dissimilarity.

- **Cutting the Dendrogram:** To obtain a specific number of clusters, you can cut the dendrogram at a certain height. The height at which you cut the dendrogram depends on your desired number of clusters.

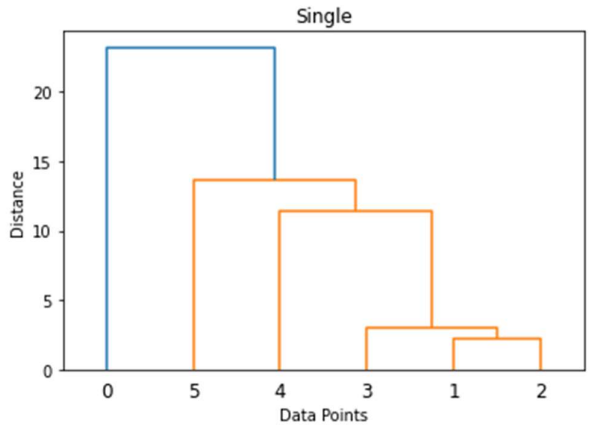
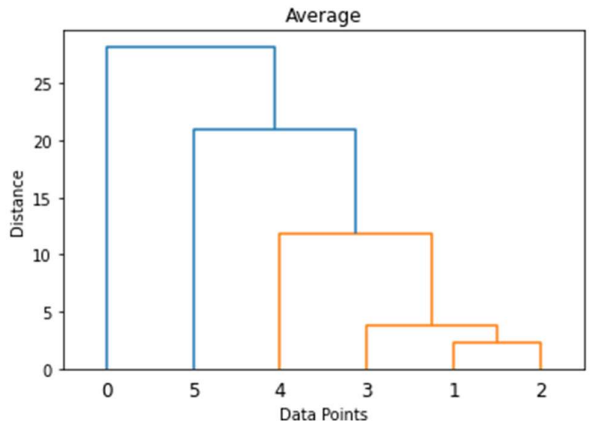
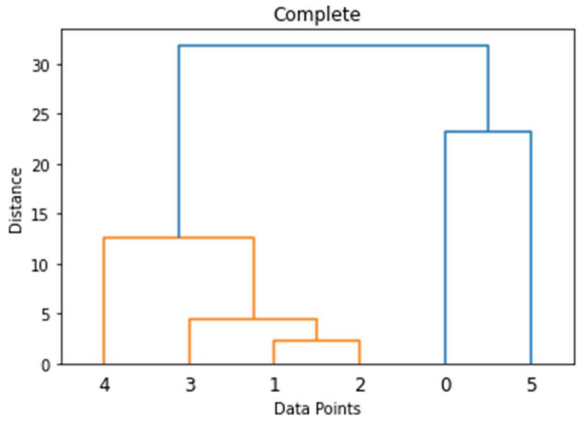
Use Cases:

- Dendrograms are useful in various fields, including biology (for gene expression analysis), social sciences (for clustering similar individuals or groups), and data analysis (for segmenting data into meaningful groups). They provide a visual representation of hierarchical relationships in the data.

Program

```
import pandas as pd
import matplotlib.pyplot as plt
from scipy.cluster import hierarchy

# Load the data from the file
data = pd.read_csv('data.csv')

# Compute the linkage matrix
linkage_matrix = hierarchy.linkage(data, method='single')


# Create a dendrogram
dendrogram = hierarchy.dendrogram(linkage_matrix)

# Show the dendrogram
plt.title('Single')
plt.xlabel('Data Points')
plt.ylabel('Distance')
plt.show()
```

```
linkage_matrix = hierarchy.linkage(data, method='average')


# Create a dendrogram
dendrogram = hierarchy.dendrogram(linkage_matrix)

# Show the dendrogram
plt.title('Average')
plt.xlabel('Data Points')
plt.ylabel('Distance')
plt.show()


linkage_matrix = hierarchy.linkage(data, method='complete')

# Create a dendrogram
dendrogram = hierarchy.dendrogram(linkage_matrix)

# Show the dendrogram
plt.title('Complete')
plt.xlabel('Data Points')
plt.ylabel('Distance')
plt.show()
```

Output	  
Conclusion:	In summary, dendrograms are valuable tools for understanding hierarchical clustering results. They help researchers and analysts visualize how data points or observations are grouped into clusters at different levels of similarity or dissimilarity. Proper interpretation of dendrograms can assist in making informed decisions about the number and composition of clusters in hierarchical clustering analysis.