Assignment 3 – Agglomerative

Problem Statement:

You work in XYZ Company as a Python Developer. The company officials want you to write code for an Agglomerative Clustering Problem.

Tasks To Be Performed:

1. Load iris data from load_iris function from sklearn.datasets package
2. From the dataset extract the data property
3. Train an Agglomerative Clustering model based on the data
4. Plot dendrogram to visualize the clustering linkage

In [1]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.datasets import load_iris
from sklearn.cluster import AgglomerativeClustering
```

In [2]:
```
iris_data = load_iris()
df = pd.DataFrame(data=iris_data.data, columns=iris_data.feature_names)
df
```

Out[2]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...	...	...	...	...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

In [3]:
```
X = iris_data.data
```

In [4]:
```
clustering = AgglomerativeClustering()
clustering.fit(X)
```

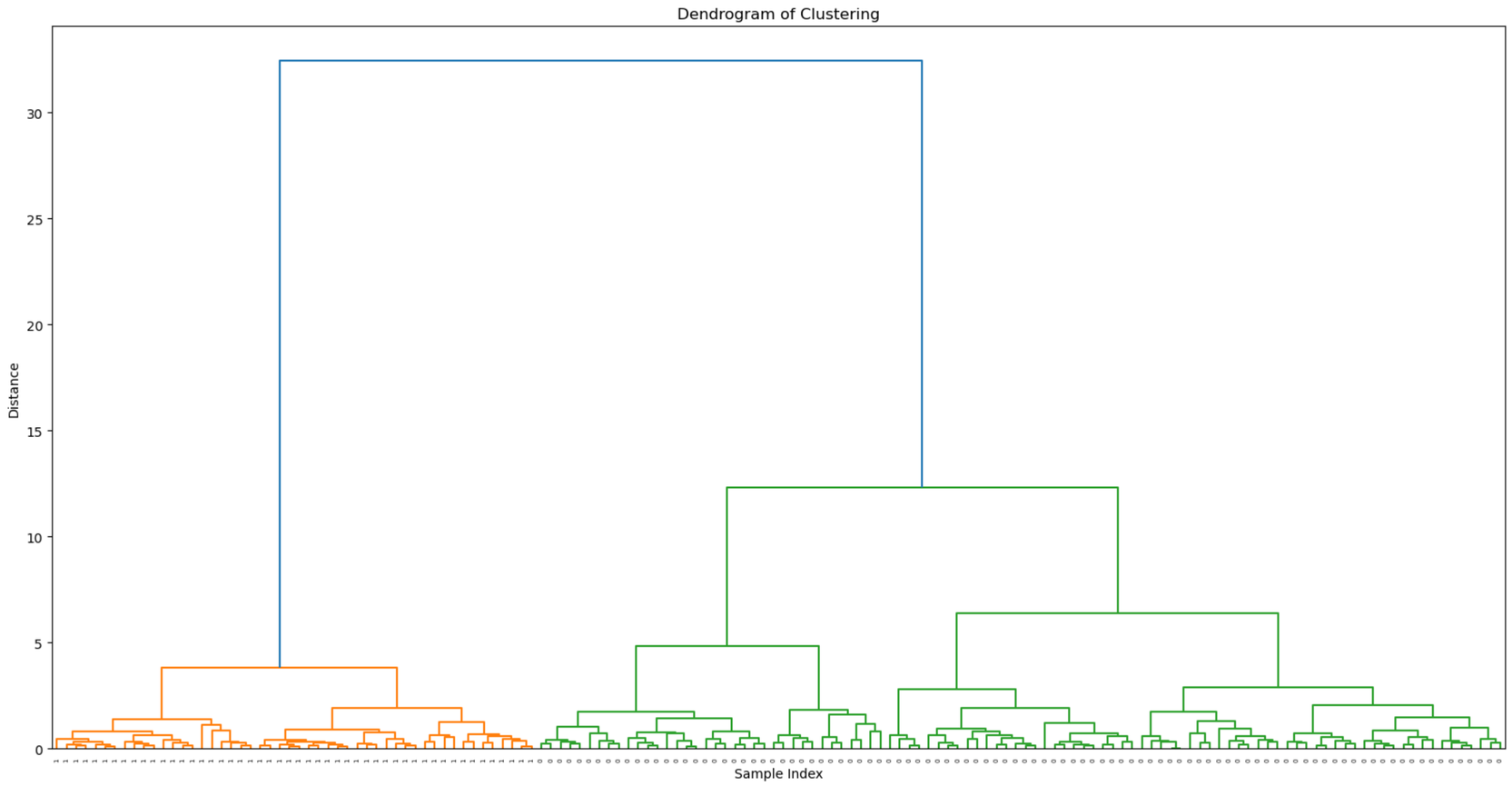
Out[4]:
```
▾ AgglomerativeClustering
AgglomerativeClustering()
```

In [5]:
```
clustering.labels_
```

Out[5]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)
```

In [6]:
```
linkage_matrix = linkage(X, method='ward')
```

In [7]:
```
plt.figure(figsize=(20, 10))
dendrogram(linkage_matrix, labels=clustering.labels_)
plt.title('Dendrogram of Clustering')
plt.xlabel('Sample Index')
plt.ylabel('Distance')
plt.show()
```


Dendrogram of Clustering
Distance
Sample Index
0
5
10
15
20
25
30

In [ ]: