Assignment 4 – Dendrogram

Problem Statement:

You work in XYZ Company as a Python developer. The company officials want you to write code for an Agglomerative Clustering Problem. Data: [[5,3], [10,15], [15,12], [24,10], [30,30], [85,70], [71,80], [60,78], [70,55], [80,91],]

Tasks To Be Performed:

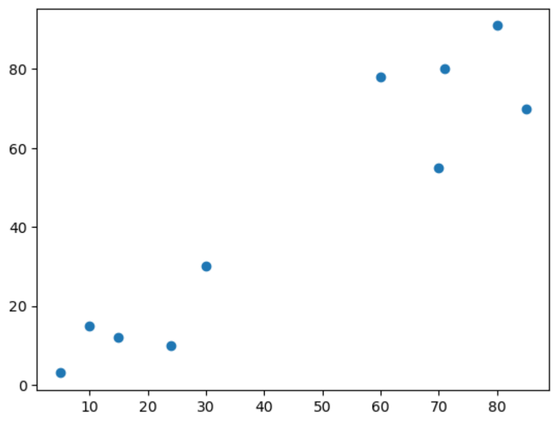
1. Using the np.array function create an np array from the data given above
2. Generate a scatter plot for the data
3. Plot dendrogram to visualize the clustering linkage

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from scipy.cluster.hierarchy import dendrogram, linkage
```

In [2]:
```
arr = np.array([[5,3], [10,15], [15,12], [24,10], [30,30], [85,70], [71,80], [60,78], [70,55], [80,91],])
arr
```

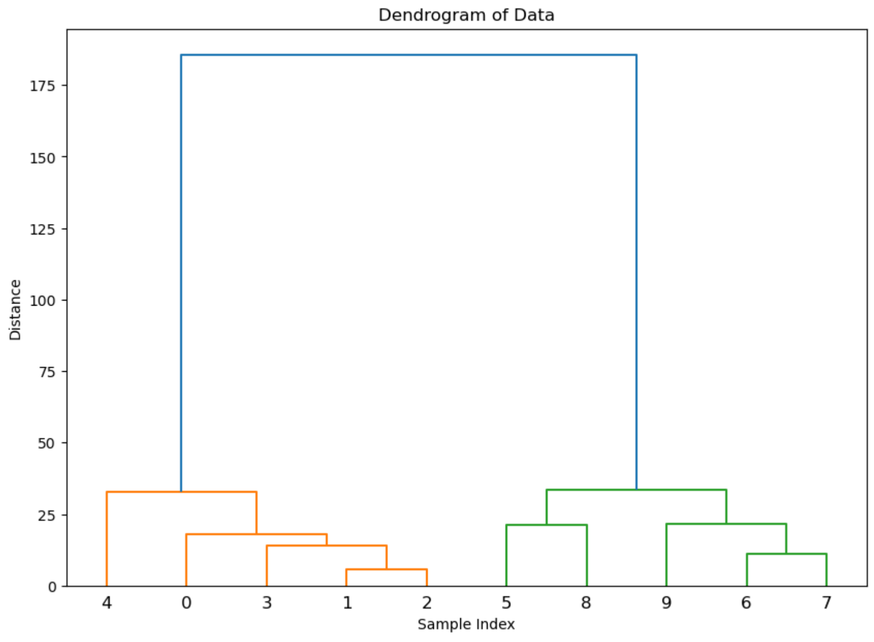
Out[2]:
```
array([[ 5,  3],
       [10, 15],
       [15, 12],
       [24, 10],
       [30, 30],
       [85, 70],
       [71, 80],
       [60, 78],
       [70, 55],
       [80, 91]])
```

In [3]:
```
plt.scatter(arr[:,0],arr[:,1])
plt.show()
```



In [4]:
```
linkage_matrix = linkage(arr, method='ward')
```

In [5]:
```
plt.figure(figsize=(10, 7))
dendrogram(linkage_matrix)
plt.title('Dendrogram of Data')
plt.xlabel('Sample Index')
plt.ylabel('Distance')
plt.show()
```


Dendrogram of Data
Distance
Sample Index
0
25
50
75
100
125
150
175
4
0
3
1
2
5
8
9
6
7

In [ ]: