# Hierarchical clustering of crime data

Perform clustering for the crime data and identify the number of clusters formed and draw inferences.

Data Description:

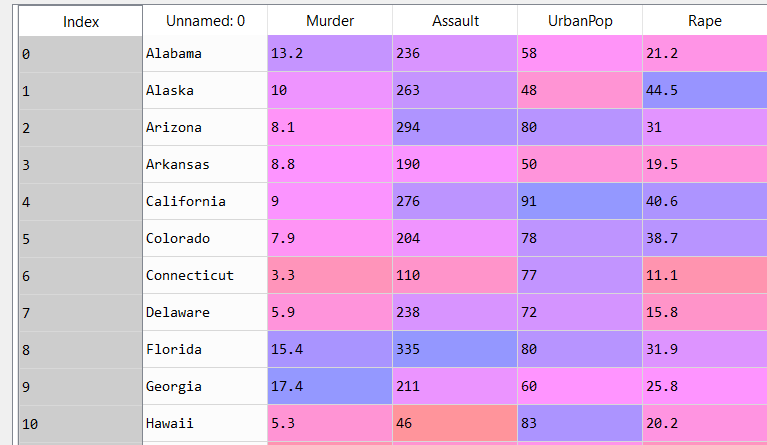
Murder -- Muder rates in different places of United States

Assualt- Assualt rate in different places of United States

UrbanPop - urban population in different places of United States

Rape - Rape rate in different places of United States

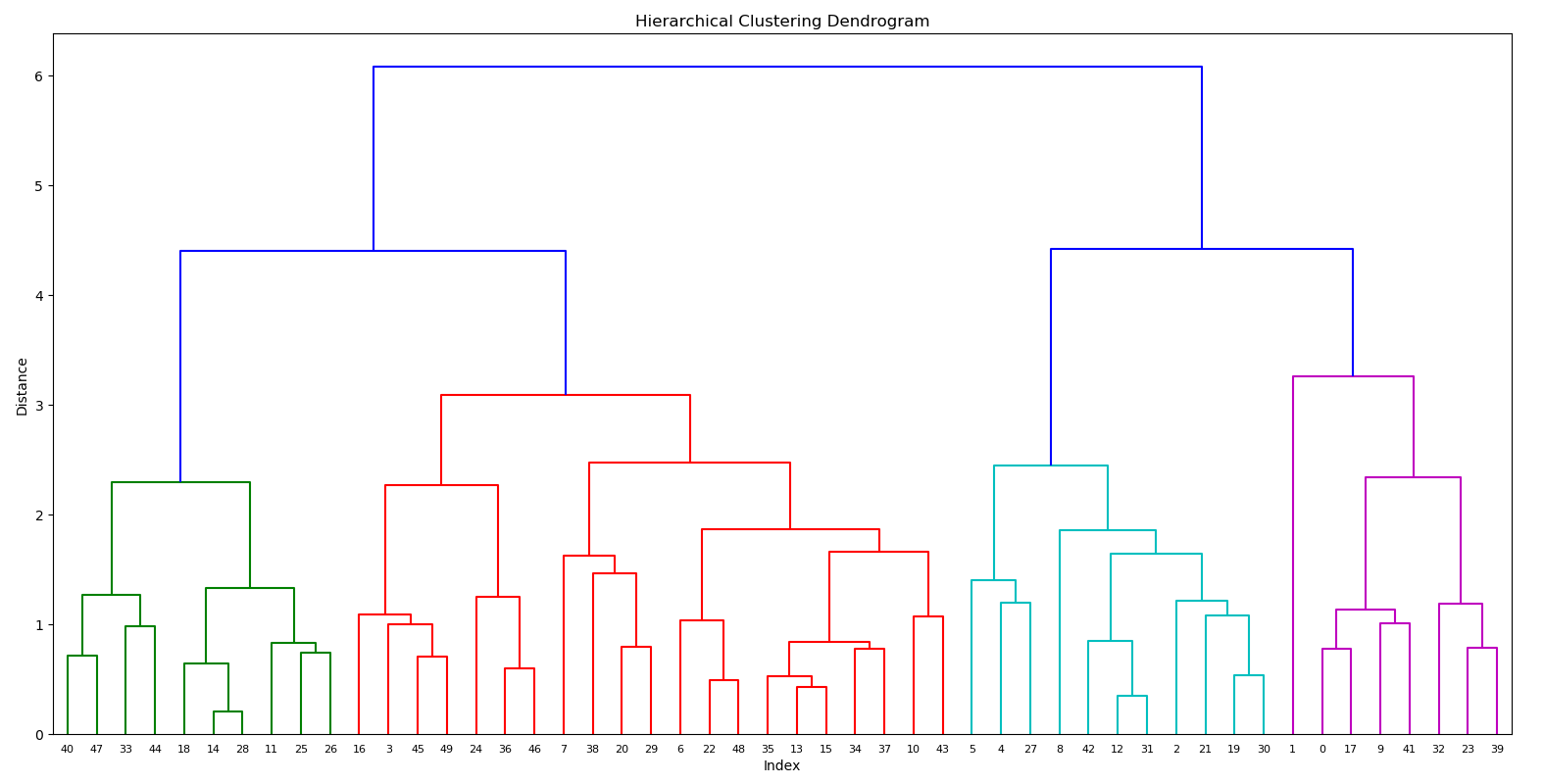
The following is the first ten data sets of crime data



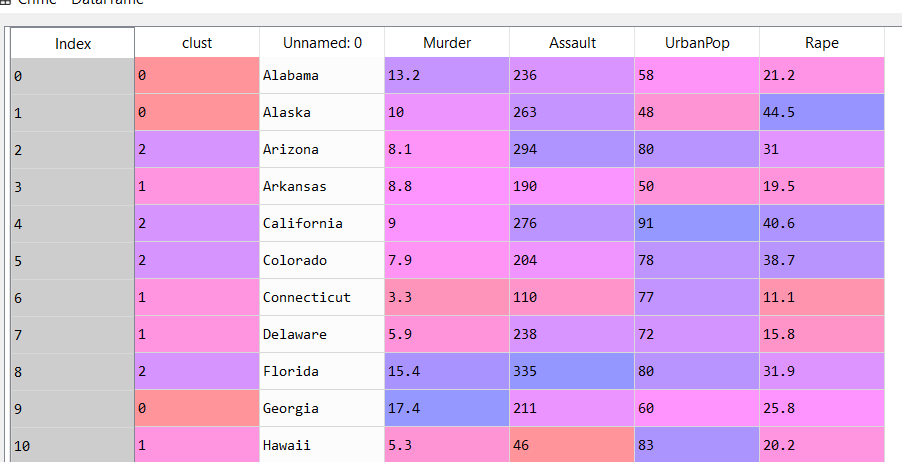
Due to drastic differences in scales we normalize the above data and obtain the following:



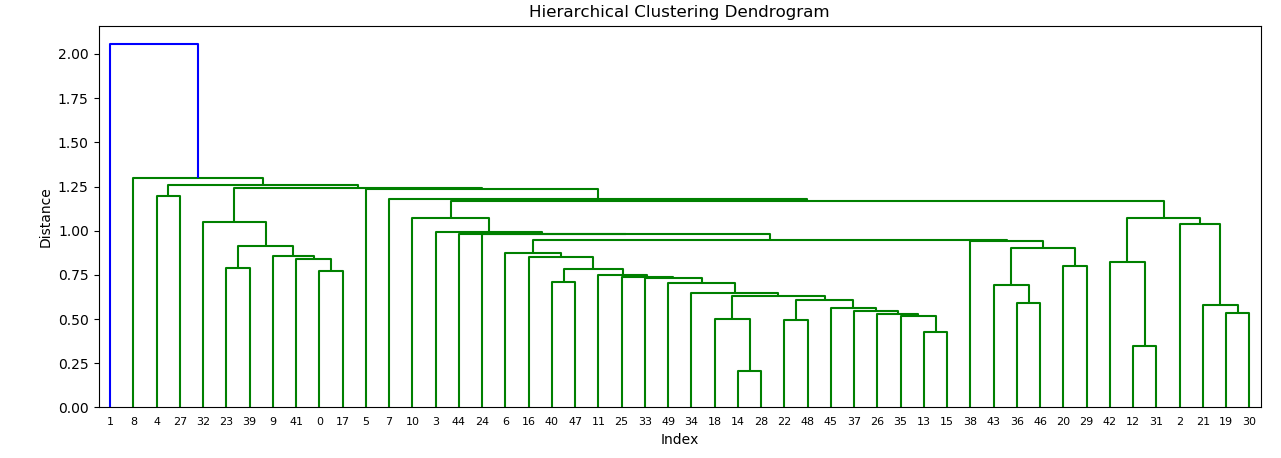
Based on Euclidian distance and complete linkage method we implement the metrics to obtain a dendrogram as follows:



Based on the above classification. We cluster the data points into four groups. The following is the classification result:

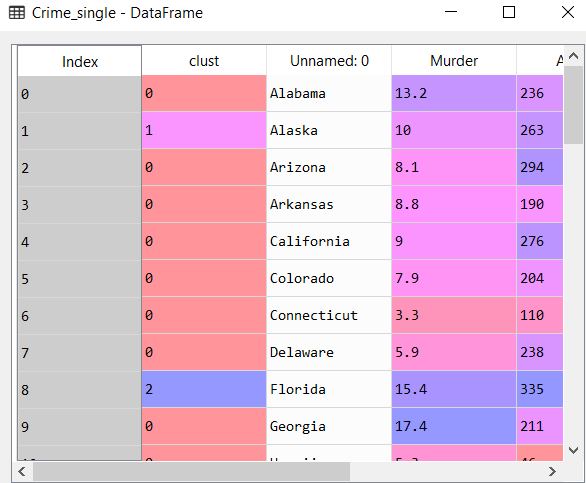


**Analysis with single Linkage and Euclidean metric**



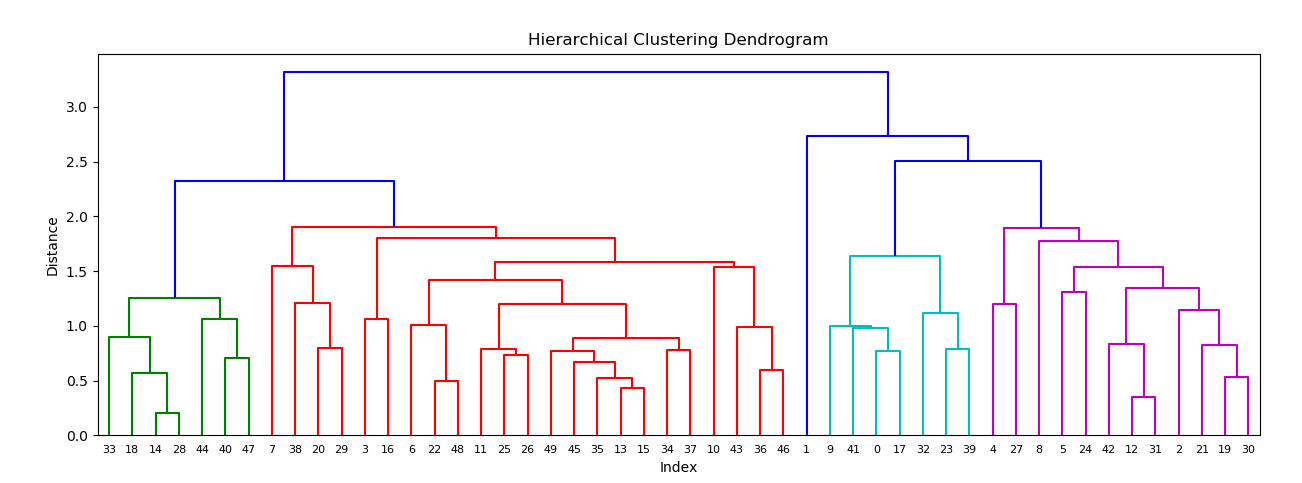
With single linkage there are predominantly two clusters from the above dendrogram.

Considering clustering with single linkage with 3 clusters, the following is the result:



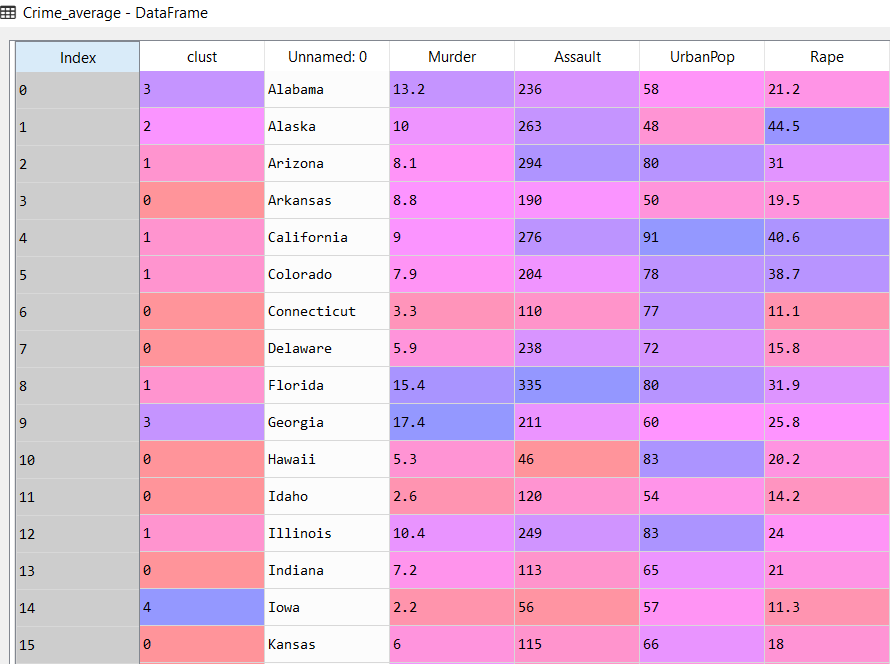
Just Alaska and Florida form individual clusters and the rest is grouped into other cluster.

**Analysis with method average linkage and Euclidean metric**

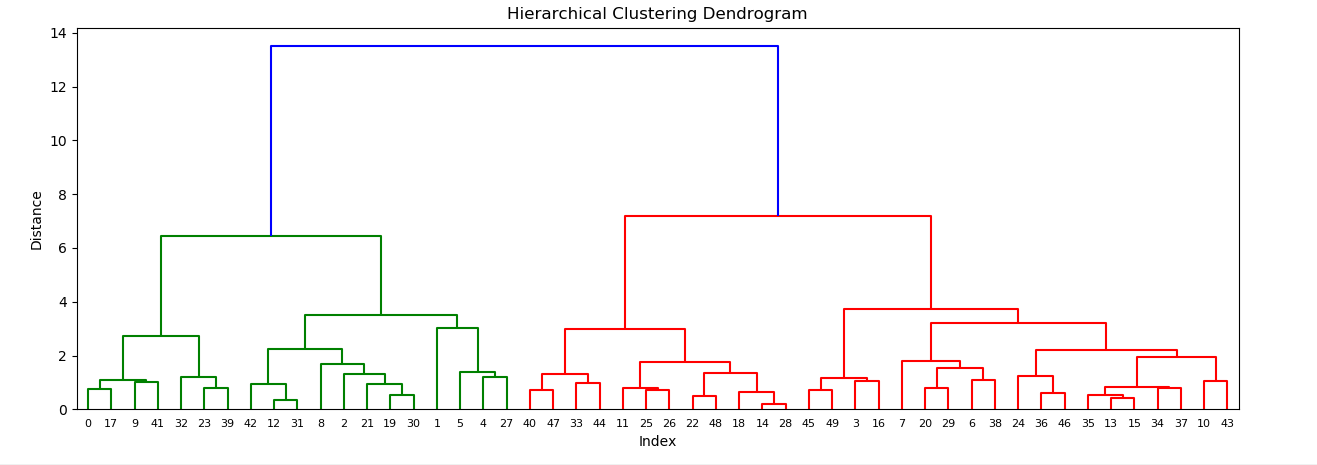


The above Dendrogram has well defined clusters of 5 indicated by blue lines.

Considering clustering with averagelinkage with 5 clusters, the following is the result:

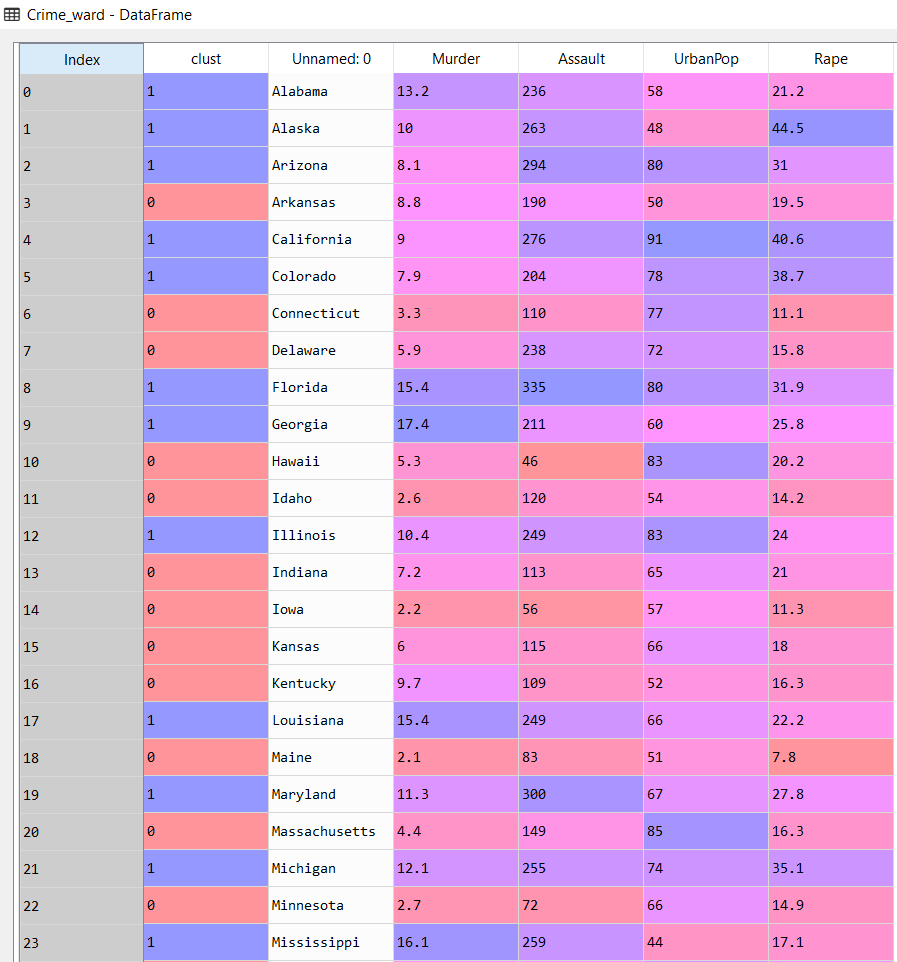


**Analysis with method ward linkage and Euclidean metric**



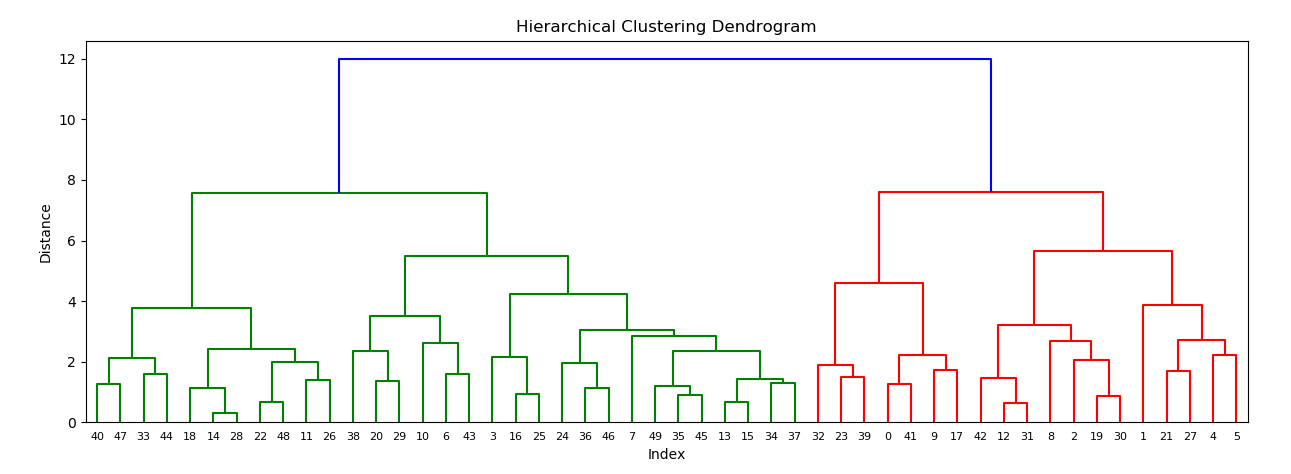
With ward linkage we have predominantly 2 linkages indicated by the blue lines.

Considering clustering with ward linkage with 2 clusters, the following is the result:



Changing the distance from Euclidean to Manhattan distance

**Analysis with method complete linkage and Manhattan metric**



The above diagram indicates that the diagram has two main clusters. As per the dendrogram we will classify two clusters.

