Un-Supervised Learning:

* From Now we are dealing with Un-Supervised Machine Learning here we are not aware of Dependent Variable or Target Variable or “Y” then we can say it as Un-Supervised Machine Learning Models
* **Clustering**: Here clustering means Grouping of Similar Items. Or forming groups based on their Categories
* Eg: Blood group’s means 🡺 A+, B+, AB+, AB- grouping of similar blood packets
* Cluster Analysis: Data Segmentation is on Exploratory Method for identify the Homogeneous groups of Clustering
* (i.e); Similar Records belong to Same set of Clusters and Non similar Records doesn’t belong to Same set of Clusters
* Now we deal with ***Hierarchal Clustering or Agglomerative Clustering***
* Here we use a concept of***Dendrogram***
* ***Dendrogram*** is a Tree like Structure here we are finding the Similarity Between two Data Points
* Here Dendrogram uses Bottom-up-Approach

Here How can we say that data points between Similarity and Dis Similarity

* The Two Closest Records or Observations must be Merged with one cluster
* It is an Iterative Process by adding one to another we form clusters or a group of clusters
* Here we say we use a technique to measure closest records called ***Euclidian’s Distance***
* ***Euclidian’s Distance 🡺 it is responsible for finding the distance between two data points***
* With the help of the Euclidian’s distance, we can the data points are similar and Dis-Similar to each other.

Forming Clusters:

* Here the clusters will be formed based on the Dendrogram
* In clusters we need to cut the dendrogram based on the threshold value
* Based on the threshold value we form clusters

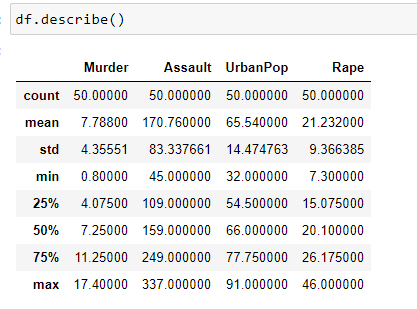
*Here we have a data set of Crime Data:*

Description of the data is

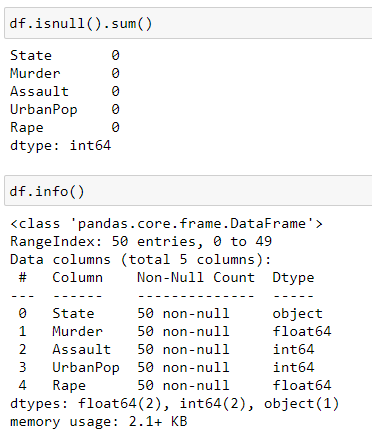
* *Data Description*:
* Murder -- Murder rates in different places of United States
* Assault- Assault rate in different places of United States
* Urban Pop - urban population in different places of United States
* Rape - Rape rate in different places of United States



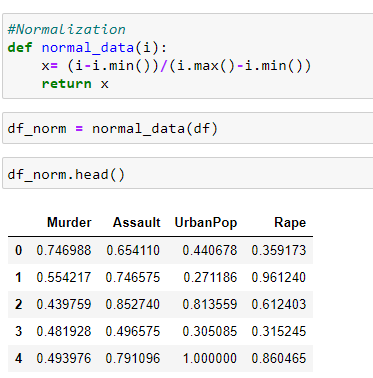
Data Describe ():description of the data



Missing Values of data:



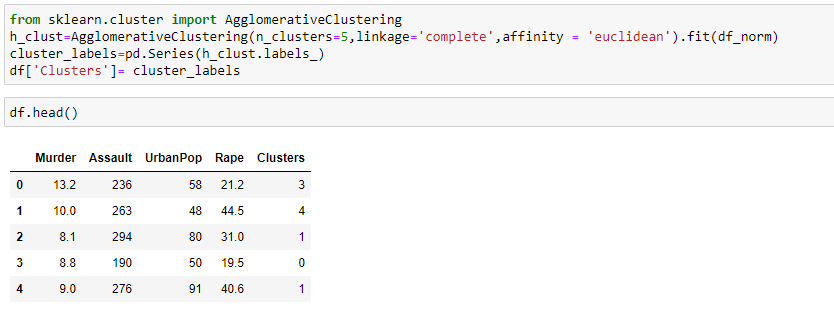
Normalization of Data:



After the process of normalization of the data Normalization converts the data into 0’s and 1’s

Dendrogram for complete linkage

Cluster Creations Agglomerative Clustering

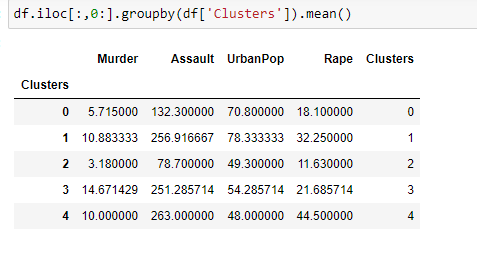


* Above we created a data frame and combined with original data set with Clusters the last column

Which contains 5 clusters 0, 1, 2, 3, 4.

* And we used Linkage=” Complete”, affinity=Euclidean which is distance metrics

Let’s see the data have been fall into what number of clusters



The above I cut dendrogram near to 0.5 which passes through 5 vertical lines which means clusters= 5.