K-Means Clustering

* It uses the Logic of “Centroid Linkage”
* Here we have to initially supply the value of “K”
* Data will be provided based on your k-value here the data will be clustered into groups
* It’s is an Iterative process at Each and Every time when the data point is added or Removed from the cluster. The centroid or Center value of the cluster will be changed
* Clusters with Homogeneous (similar observations) should fall on same cluster
* Cluster with Non-Homogeneous (Dis-Similar observations) should fall on another cluster
* No Hierarchy
* No Linkages
* No Dendrograms are needed in K-Means Clustering

***How can we know exactly how many clusters are Required?***

***Here we Use a Method called Elbow method which shows how many clusters to form based on the data set that your provided***

***Note: Here we use K-Means Clustering for Large Data sets when we have more than 1000 records***

Example:

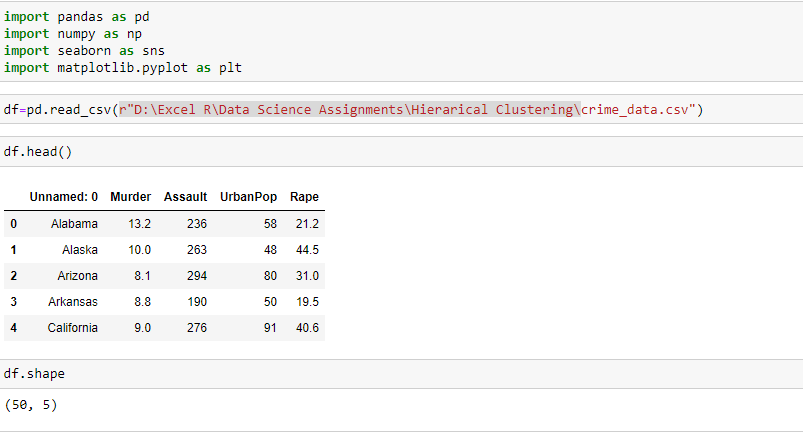
Data set is Crime Rate

* Perform Clustering(K-Means) for the crime data and identify the number of clusters formed and draw inferences.

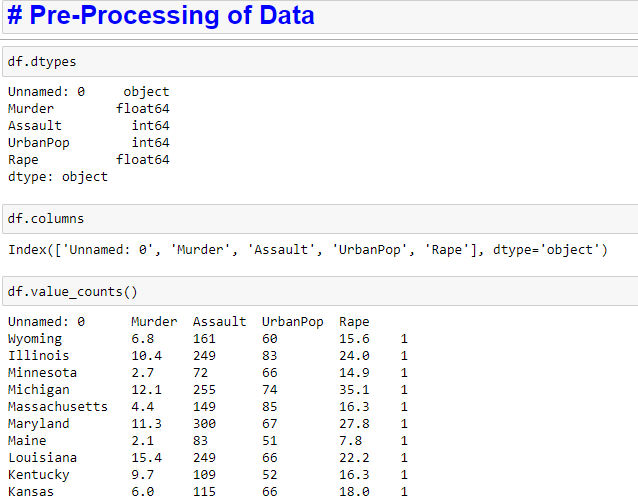
Data Description:

1. Murder -- Murder rates in different places of United States
2. Assault- Assault rate in different places of United States
3. Urban Pop - urban population in different places of United States
4. Rape - Rape rate in different places of United States

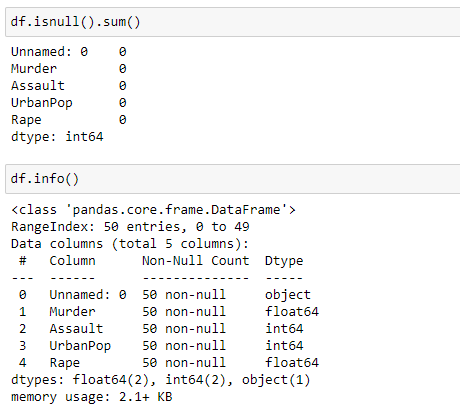
Import the data set:



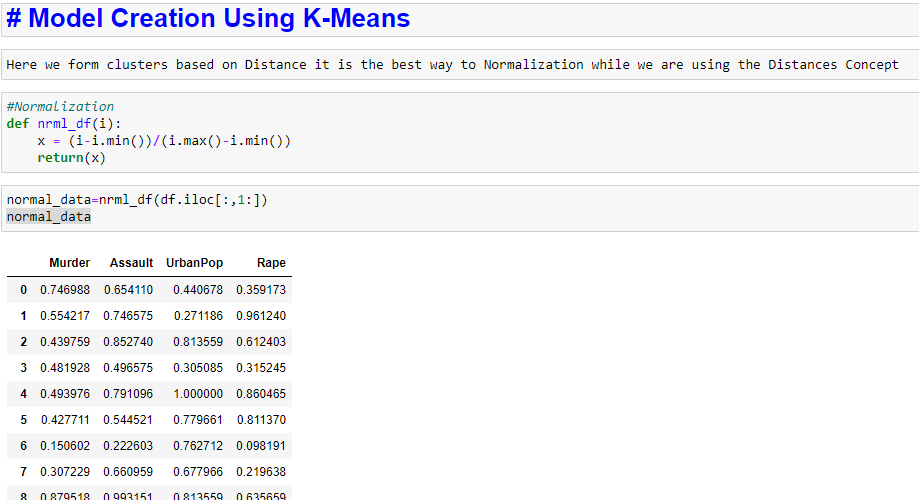
Data Pre-Processing



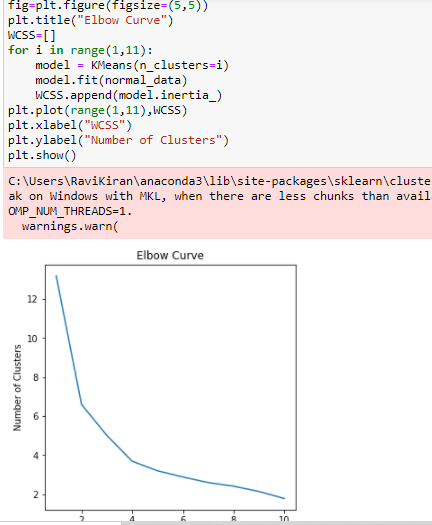
Missing Values:



Normalization of Data:



Elbow Method:



Clusters Mean:

