# DBSCAN CLUSTERING

Till now we have learned two types of clustering techniques (K-Means, Hierarchical clustering techniques) the why DBSCAN clustering again

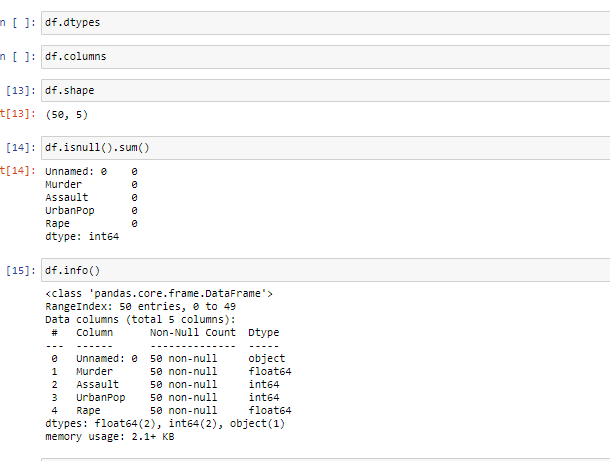
* K-Means and Hierarchical both fail in creating the arbitrary clusters of arbitrary shapes. They are not able to form clusters based on densities that’s why we need DBSCAN
* Density Based Spatial Clustering of Applications with Noise
* K-Means and Hierarchical clustering fails in detecting the Noise in the data set
* DBSCAN is used to detect the Noise and form another cluster
* ***It was proposed by Martin Ester et al. in 1996. DBSCAN is a density-based clustering algorithm that works on the assumption that clusters are dense regions in space separated by regions of lower density.***
* **The most exciting feature of DBSCAN clustering is that it is robust to outliers**.
* It doesn’t require the number-clusters to be said beforehand unlike K-Means.
* Here we have to specify the number of centroids and min points are the two parameters (epslon and min-points) required for this clustering technique.
* Epslon is the radius of the circle to be created around the datapoint to check the density
* To check the distances between the circles we use **Euclidean Distance** i.e; when we are dealing with Distance s we need to Normalize the data
* Minpoints are the minimum number of data points required in the circle for that data point is considered as a **Core Point**
* DBSCAN creates circles with the help of epslon called radius around data points and classifies them as **Core, Border, Noise** data points
* A data point is considered as a **Core data** point if the circle is around, it contains at least “min-points”
* If the number of min-points are less than min-points it is considered as a **Border Point**
* It they are no other data points around data observations within epslon or radius are treated as **Noise Points**

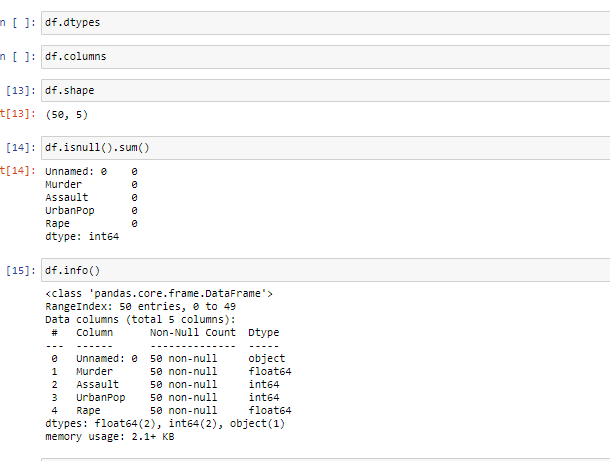
**Parameter Selection:**

* Epson: the value of Epson can be decided from the K-graph the maximum point curvature elbow in that graph tells us about the value of epslon
* Min-Points: no of features + one

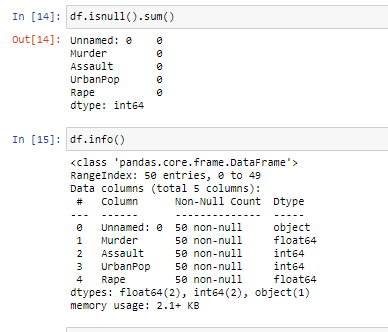
***Code: For Crime Rate***

Pre-Processing:

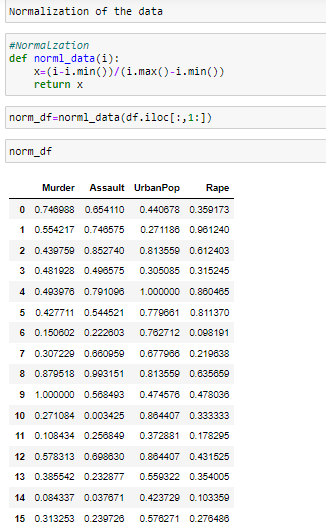




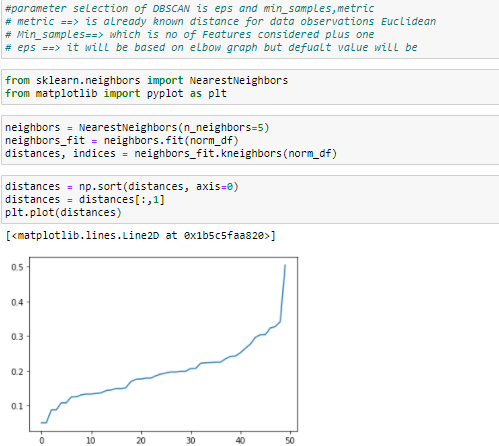
Missing values



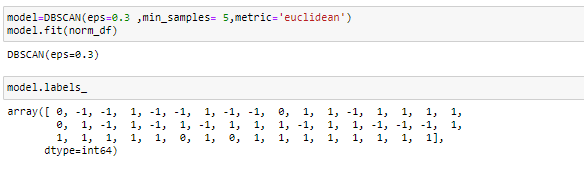
Normalization



K-Graph to check the epslon value



Model Creation



Cluster Mean

