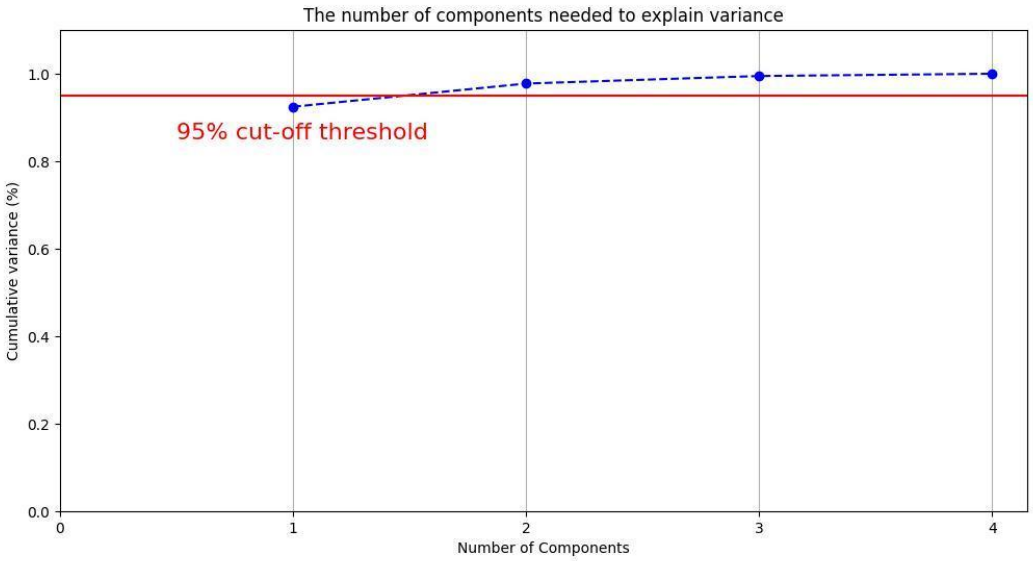
Q1

Result:

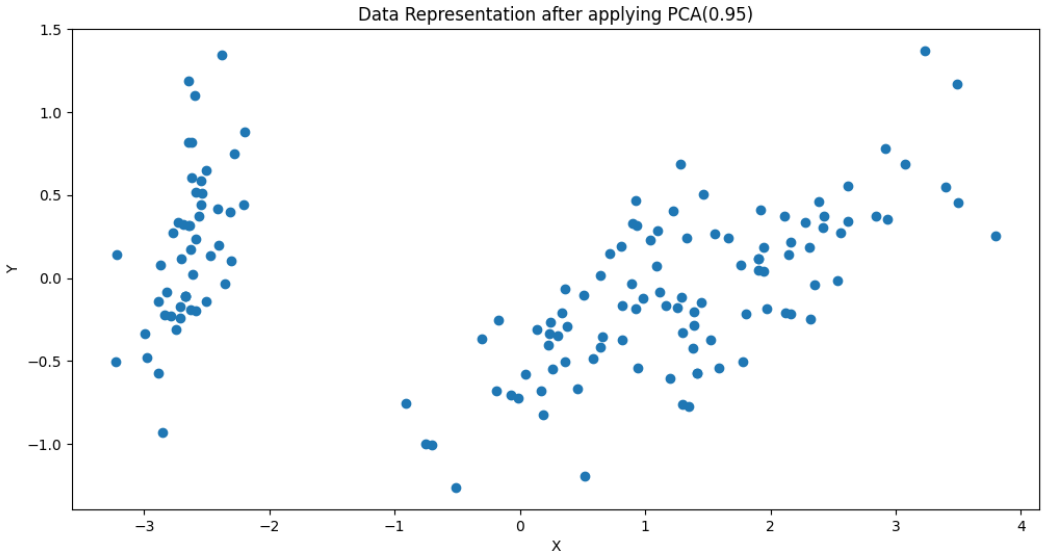
1) Cumulative Variance vs Number of component



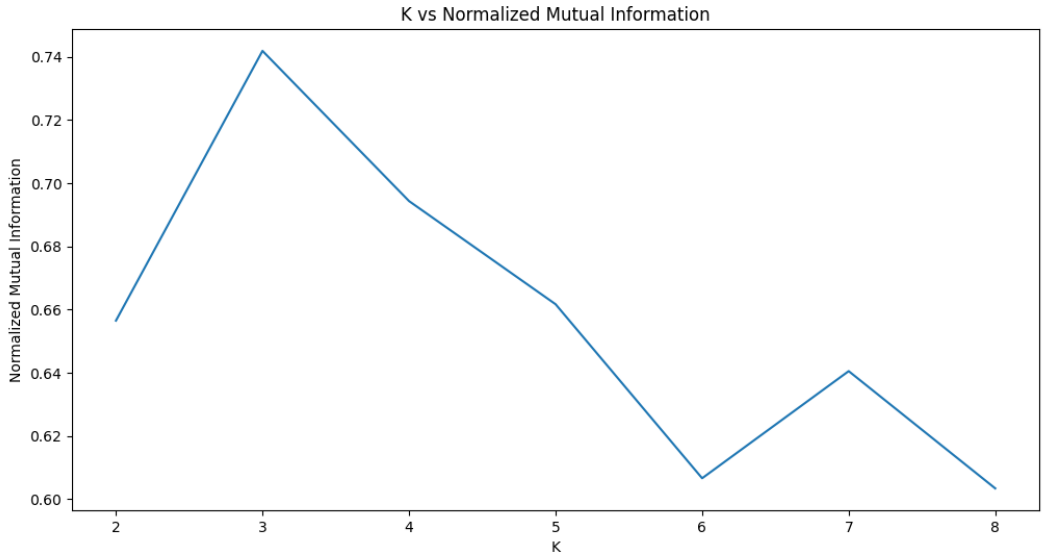
Number of components : [1 2 3 4]
Cumulative variance (%) : [0.92461621 0.97763178 0.99481691 1]

For number of component equals to 2, we get 95%+ cumulative Variance.

2) Data Representation after applying PCA with number of components equals to 2



3) K vs NMI



Random initialisation done.

Here, We get Maximum Normalized Mutual Information for **k = 3**
Values of K : [2, 3, 4, 5, 6, 7, 8]
Normalized Mutual Info : [0.6565191143081128, 0.7419116631817838, 0.7199262269988783, 0.6389005260096066, 0.615784731916942, 0.5902897118315974, 0.6287636293196424]

4) Data Representation after applying (k=3) Kmeans Clustering