

CLUSTERING

Clustering performance on Dataset “Wine-Quality” using various clustering techniques on various parameters.

TABLE 1: Using K-mean

Using K-Mean Clustering																		
Parameters	No Data Preprocessing			Using Normalization			Using Transformation			Using PCA			Using T-N			Using T-N-PCA		
	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5
Silhouette	0.173	0.1623	0.1778	0.1886	0.1815	0.1819	0.3394	0.2327	0.2665	0.5312	0.4952	0.4604	0.2894	0.2366	0.2258	0.1602	0.1481	0.1441
Calinski-Harabasz	292.9428	273.1273	274.5551	375.3123	334.4627	325.6049	1710.5822	1381.9392	1252.4646	3142.1823	3169.6281	3179.3427	817.3965	700.5866	617.5399	289.6911	278.8926	253.938
Davies-Bouldin	1.846	1.7065	1.4988	1.8103	1.692	1.6334	1.0154	1.1629	1.1385	0.8503	0.7025	0.7235	1.2233	1.3931	1.3582	1.7976	1.799	1.8611

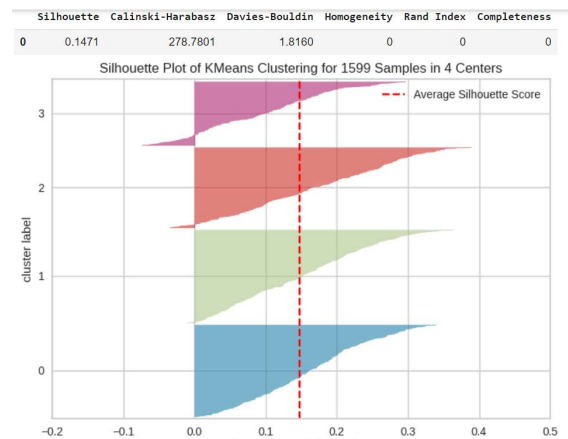
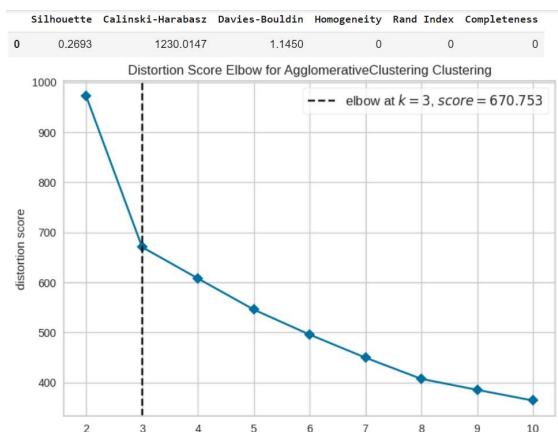


TABLE 2: USING hierarchical Clustering

Using Hierarchical Clustering																		
Parameters	No Data Preprocessing			Using Normalization			Using Transformation			Using PCA			Using T-N			Using T-N-PCA		
	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5
Silhouette	0.1247	0.1142	0.1129	0.143	0.1475	0.1296	0.3131	0.2693	0.2141	0.4589	0.3856	0.3996	0.1247	0.1142	0.1129	0.1247	0.1142	0.1129
Calinski-Harabasz	250.2404	228.2218	209.8825	228.3249	216.7685	216.5409	1596.6195	1230.0147	1072.1545	2430.6559	2332.6915	2770.6072	250.2404	228.2218	209.8824	250.2404	228.2218	209.8825
Davies-Bouldin	1.9943	2.034	2.0554	1.9956	1.6172	1.7311	1.0565	1.145	1.3181	0.7343	0.7524	0.7506	1.9943	2.034	2.0554	1.9943	2.034	2.0554

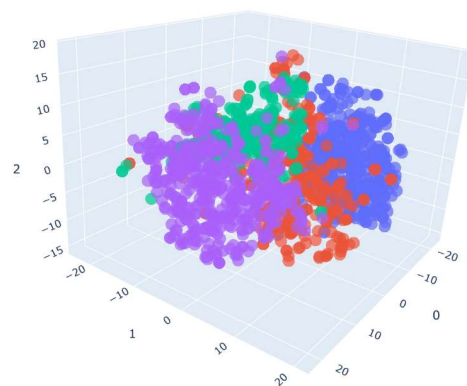
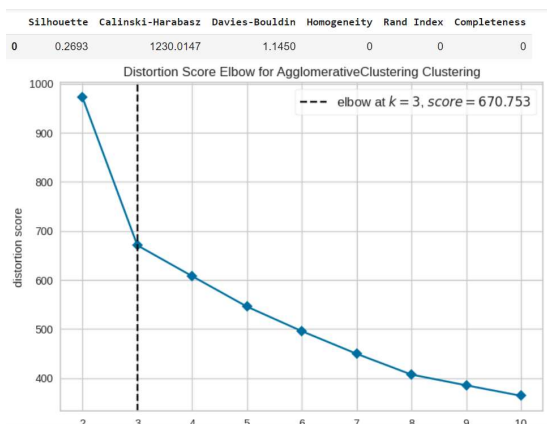
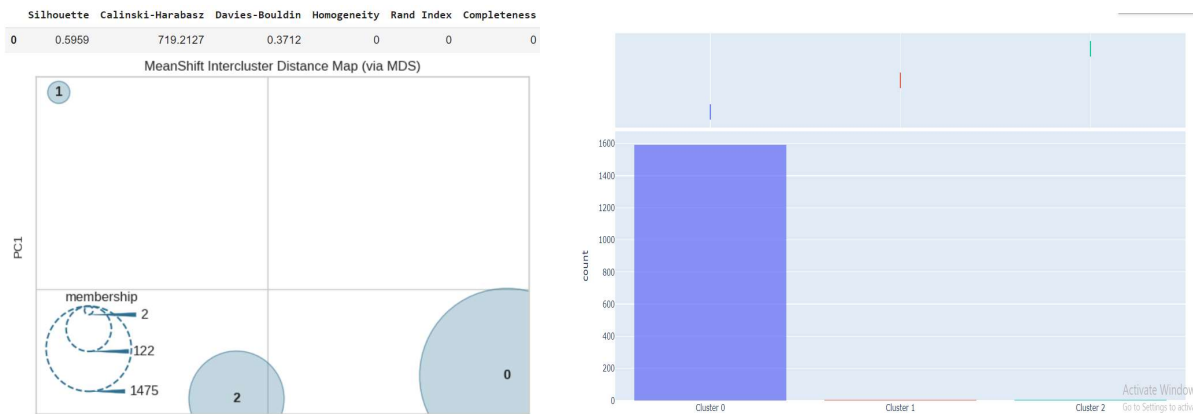


TABLE 3: Using Kmean Shift

Using K-Mean Shift Clustering																		
Parameters	No Data Preprocessing			Using Normalization			Using Transformation			Using PCA			Using T+N			Using T+N+PCA		
	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5
Silhouette	0.4035	0.4035	0.4035	0.3145	0.3145	0.3145	0	0	0	0.5959	0.5959	0.5959	0.4035	0.4035	0.4035	0.4035	0.4035	0.4035
Calinski-Harabasz	14.6834	14.6834	14.6834	35.4606	35.4606	35.4606	0	0	0	28.9787	28.9787	28.9787	14.6834	14.6834	14.6834	14.6834	14.6834	14.6834
Davies-Bouldin	0.9004	0.9004	0.9004	0.9525	0.9525	0.9525	0	0	0	0.5684	0.5684	0.5684	0.9004	0.9004	0.9004	0.9004	0.9004	0.9004



CONCLUSION

In comparing the results of K-Means, Hierarchical, and K-Means Shift clustering on the seeds dataset, several key trends emerge, particularly when considering the Silhouette score, Calinski-Harabasz index, and Davies-Bouldin score, which measure cluster cohesion, separation, and compactness.

K-Means may have performed well after data normalization, with the best performance seen when using Z-Score normalization. You likely observed optimal clustering performance when selecting **3 clusters**, corresponding to the actual number of wine classes in the dataset.

Mean Shift offers more flexibility by automatically determining the number of clusters and allowing for clusters of varying shapes. This makes it a good choice when the number of clusters is not known in advance or when you suspect that the clusters may not be spherical.