Name:-Adrija Sengupta 3c41

Clustering-Assignment

Dataset used - Seeds

KMeans

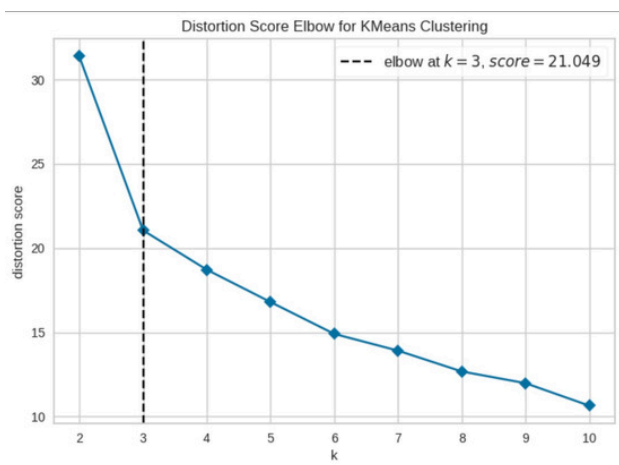| Parameters						Using KMeans Clustering											
	No Data Processing			Using Normalization			Using Transform			Using PCA			Using T+N			T+N+PCA		
	C=3	C=4	C=5	C=3	C=4	C=5	C=3	C=4	C=5	C=3	C=4	C=5	C=3	C=4	C=5	C=3	C=4	C=5
Silhouettee	0.4540	0.3962	0.4021	0.3991	0.3326	0.2779	0.5198	0.5343	0.502	0.4572	0.4042	4021	0.3960	0.3173	0.2780	0.3960	0.3280	0.2572
Calinski-Harabasz	324.495	245.880	237.81	213.5749	168.7474	137.363	488.6652	653.3190	606.1324	324.37	275.4213	237.8020	196.5351	152.9984	127.1048	196.5351	179.2579	58.8120
Davies-Bouldin	0.789	1.0350	0.8625	0.9647	1.3136	1.4062	0.5032	0.5414	0.5659	0.7981	0.8602	0.8625	0.9848	1.2858	1.2895	0.9483	1.1101	1.2872

For No Data Processing
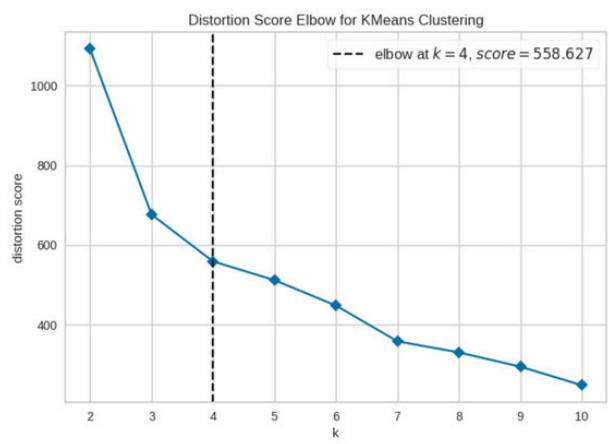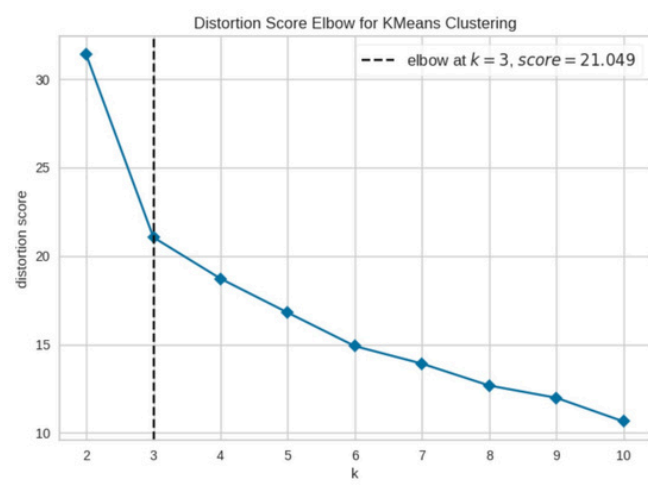

## For Using Normalisation
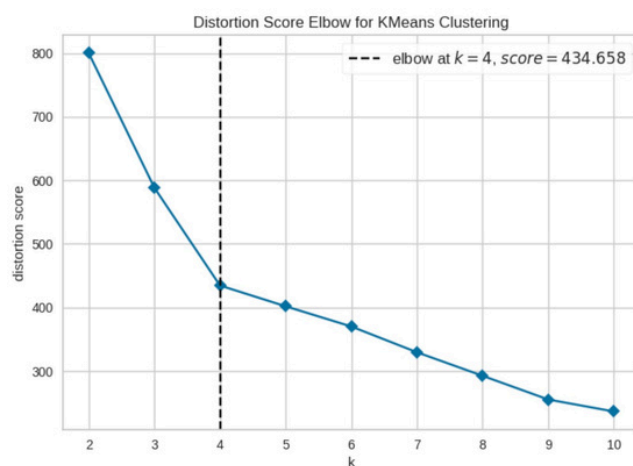
For Using T+N

For Using T+N+PCA

Hierarchical Clustering

| Parameters							Using Hierichal Clustering												
	No Data Processing			Using Normalization			Using Transform			Using PCA			Using T+N			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
Silhouettee	0.5064	0.5169	0.4994	0.3737	0.3755	0.2961	0.5064	0.5129	0.4994	0.4066	0.3914	0.3734	0.3822	0.3870	0.3590	0.3822	0.3870	0.3590
Calinski-Harabasz	438.9554	608.5274	630.6559	189.7498	154.0536	136.1257	438.9552	608.5274	630.6559	253.3361	256.1087	231.3250	188.0839	149.23292	125.584	188.0304	149.232	125.5824
Davies-Bouldin	0.5312	0.5506	0.5332	1.0376	0.9478	1.0402	0.5312	0.5506	0.5332	0.8383	0.9231	0.9811	1.036	1.3004	1.0750	1.0136	1.3004	1.0750

For No Data Processing

For Using Normalisation
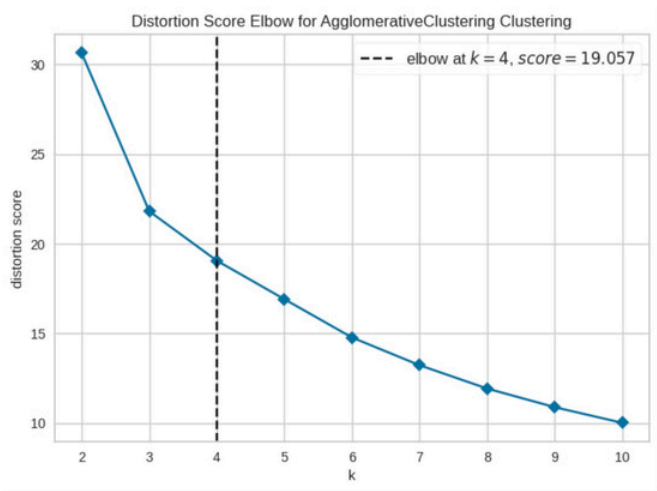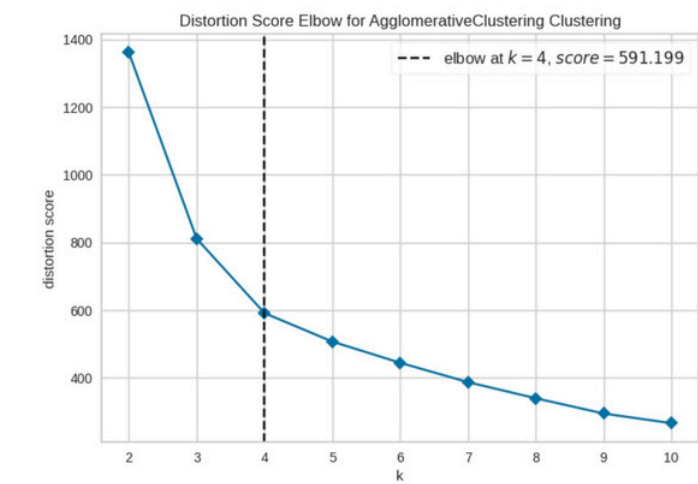
For Using T+N
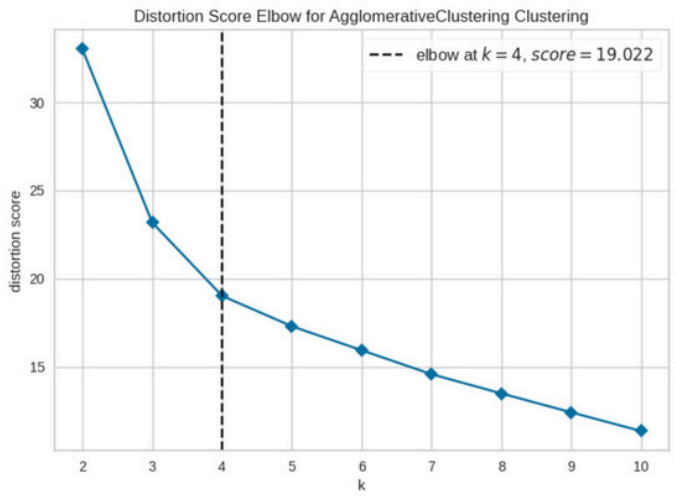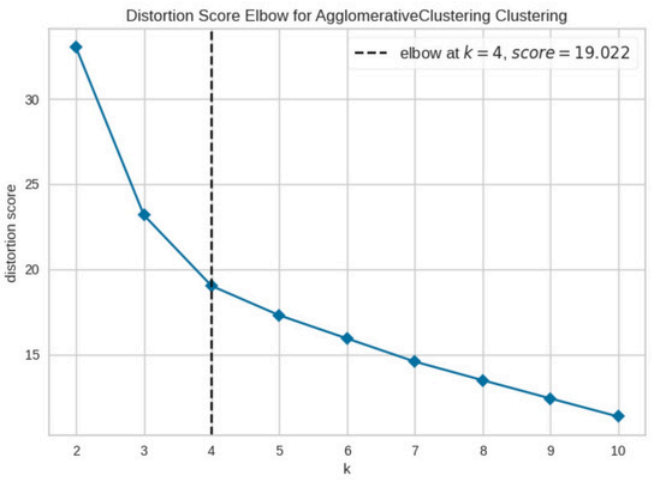
For Using T+N+PCA

KMeans Shi< Clustering

Parameters	No Data Processing			Using Normalization			Using Transform			Using PCA			Using T+N			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
Silhouettee	0.4401	0.4401	0.4401	0.4505	0.4505	0.4505	0	0	0	0.4724	0.4724	0.4724	0.3997	0.3997	0.3997	0.4115	0.4115	0.4115
Calinski-Harabasz	93.4796	93.4796	93.4796	74.0335	74.0335	74.0335	0	0	0	93.4795	93.4795	93.4795	55.7264	55.7264	55.7264	62.2176	62.2176	62.2176
Davies-Bouldin	0.4805	0.4805	0.4805	0.4979	0.4979	0.4979	0	0	0	0.4805	0.4805	0.4805	0.5739	0.5739	0.5739	0.5382	0.5382	0.5382

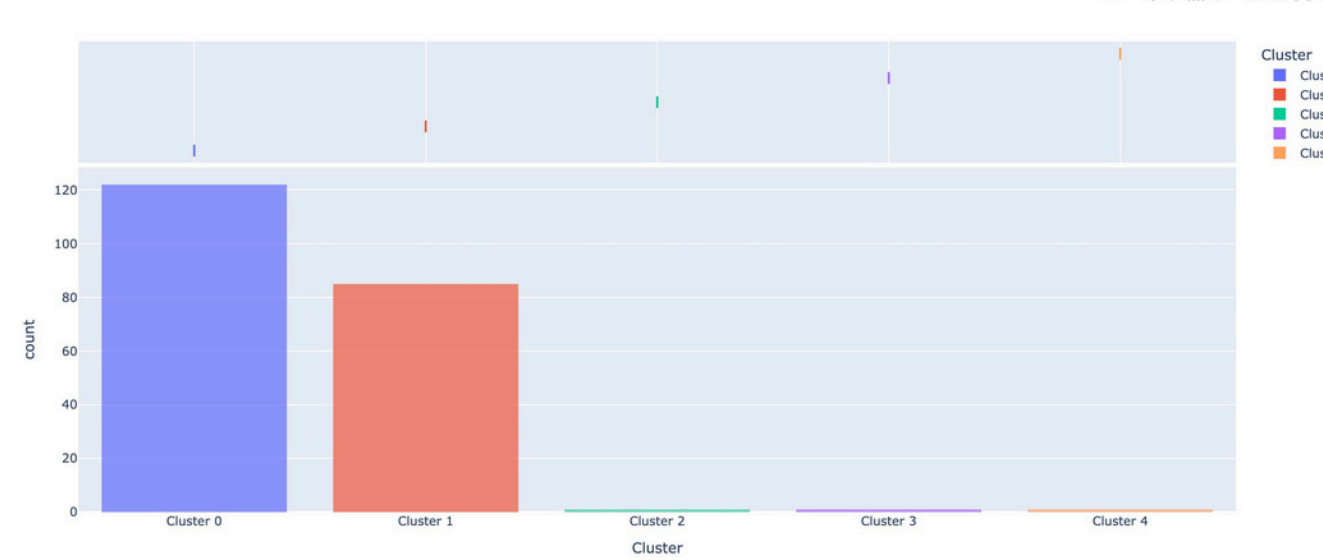(Table heading above columns: Using KMean Shift Clustering)
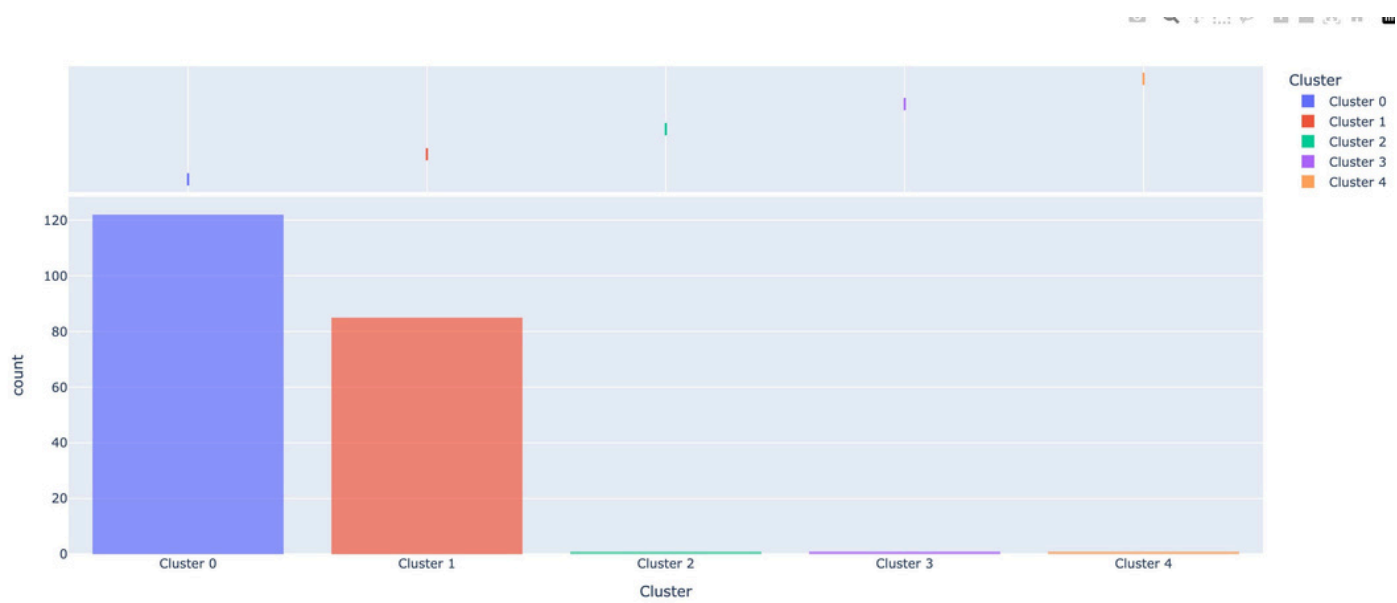
For No Data Processing

For Using Normalisation
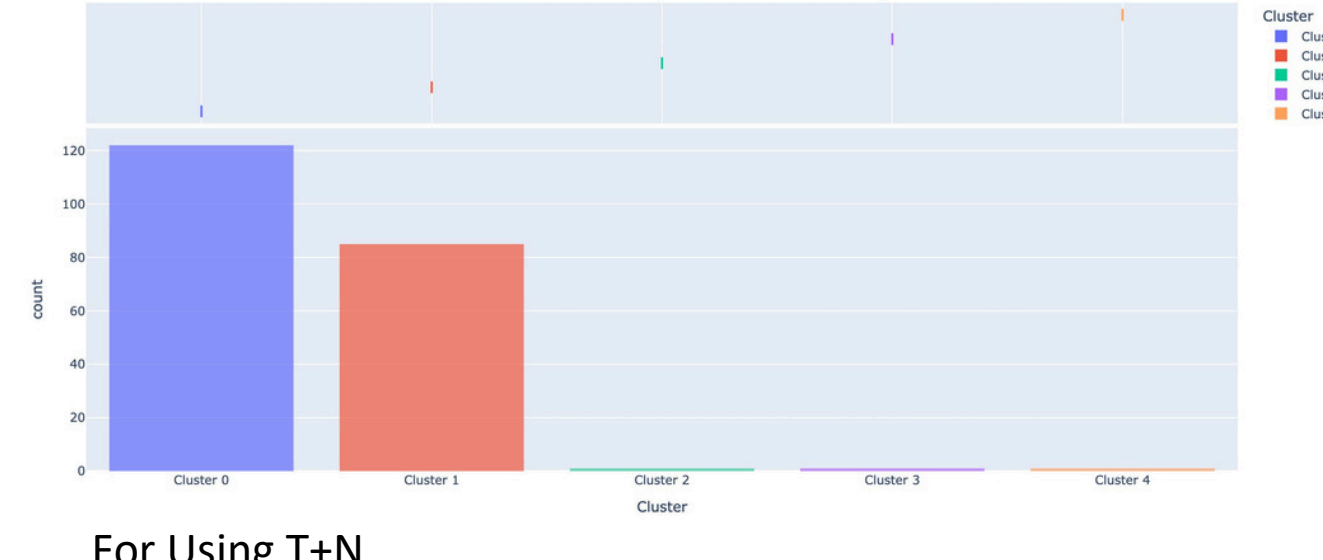
For Using T+N

For Using T+N+PCA

Conclusion:-

K–Means provided a good balance of efficiency and performance with preprocessing, while Hierarchical clustering needed normalization for better results. K–Means Shift offered flexibility with raw data but at computational cost. Overall, the choice depends on the specific dataset and computational resources. K–Means with preprocessing appeared to be the most consistent in terms of performance.