Name: **Rohit Singla**   Roll No: **102203804**   Class: **3CO18**

Clustering Performance on Datasets *"seeds"* using various clustering techniques on various parameters

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	0.454	0.382	0.3713	0.3867	0.3974	0.3437	0.5234	0.5343	0.5067	0.4572	0.403	0.415	0.3916	0.3581	0.277	0.3958	0.333	0.2935
Calinski - Harabasz	324.4954	273.3039	256.7087	155.3634	164.6173	142.5587	492.833	653.319	605.707	324.3747	275.5164	235.9164	195.7777	150.3618	127.304	196.4406	144.991	124.6828
Davies - Bouldin	0.7987	0.947	0.9549	1.0247	0.8704	1.0435	0.5664	0.5415	0.5637	0.7981	0.8609	0.788	0.9936	1.2999	1.5029	0.9925	1.4963	1.4113

No Data Pre-processing
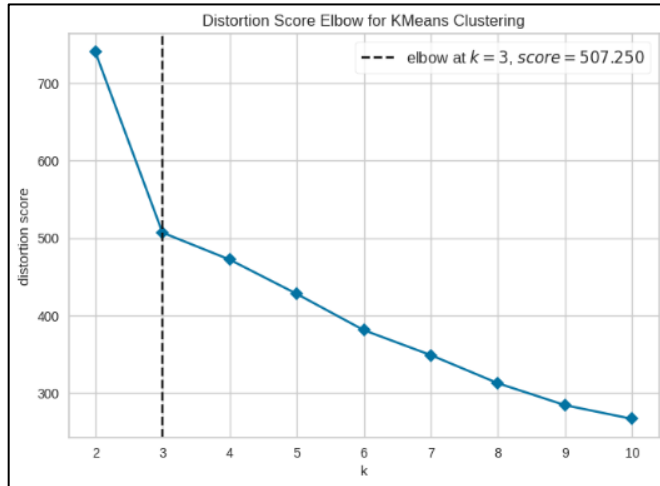
Preprocessed Data (T+N+PCA)
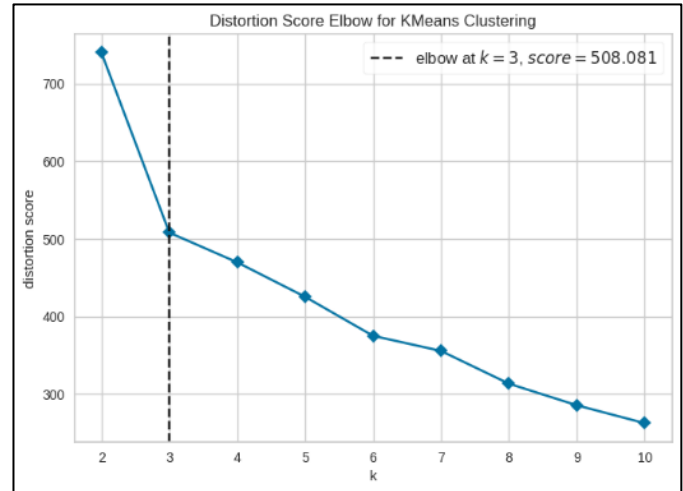
Using Hierarchical Clustering																	
Parameters	No Data Preprocessing			Using Normalization			Using Transformation			Using PCA			Using T+N			Using T+N+PCA		
	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5
Silhouette	0.4066	0.3914	0.3734	0.3794	0.3624	0.3006	0.5064	0.5129	0.4994	0.4066	0.3914	0.3734	0.3822	188.0838	1.0136	0.3822	188.0838	1.0136
Calinski - Harabasz	253.3361	256.1087	231.325	144.008	149.012	132.3875	438.9553	608.5274	630.6559	253.3361	256.1087	231.325	0.387	149.2392	1.3004	0.387	149.2392	1.3004
Davies - Bouldin	0.8383	0.9231	0.9811	1.121	0.8866	1.0281	0.5312	0.5506	0.5332	0.8383	0.9231	0.9811	0.359	125.5824	1.075	0.359	125.5824	1.075

No Data Pre-processing
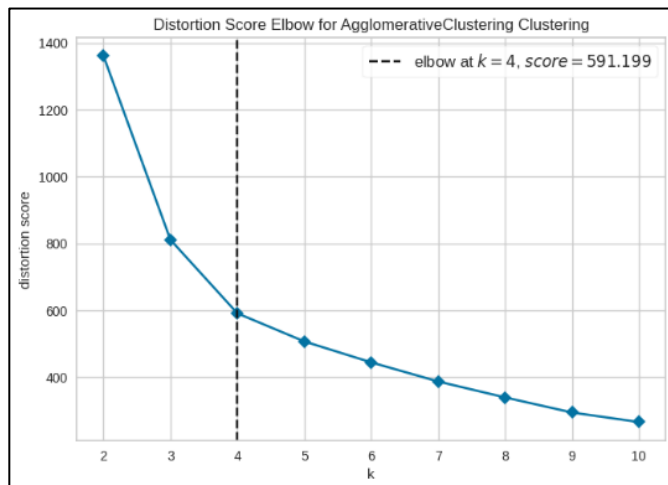
Preprocessed Data (T+N+PCA)
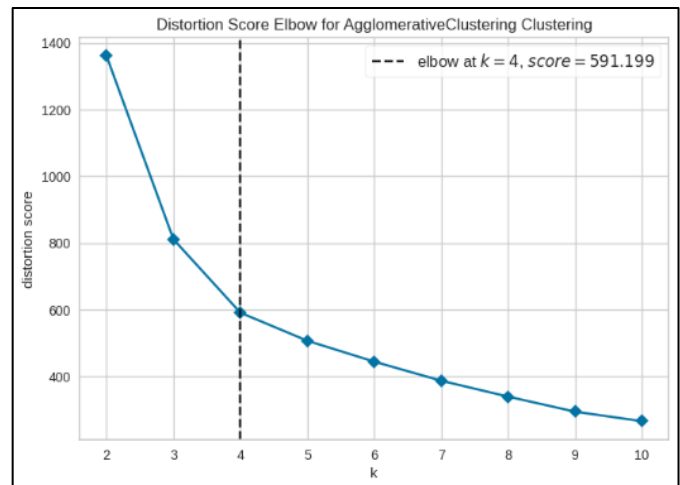
Parameters	Using K-Mean Shift Clustering																	
	No Data Preprocessing			Using Normalization			Using Transformation			Using PCA			Using T+N			Using T+N+PCA		
	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5	c=3	c=4	c=5
Silhouette	0.4401	0.4401	0.4401	0.6812	0.6812	0.6812	0	0	0	0.4724	0.4724	0.4724	0.3997	0.3997	0.3997	0.4115	0.4115	0.4115
Calinski - Harabasz	74.5656	74.5656	74.5656	18.8635	18.8635	18.8635	0	0	0	93.4795	93.4795	93.4795	55.7264	55.7264	55.7264	62.2176	62.2176	62.2176
Davies - Bouldin	0.4598	0.4598	0.4598	0.2186	0.2186	0.2186	0	0	0	0.4805	0.4805	0.4805	0.5739	0.5739	0.5739	0.5382	0.5382	0.5382

No Data Pre-processing
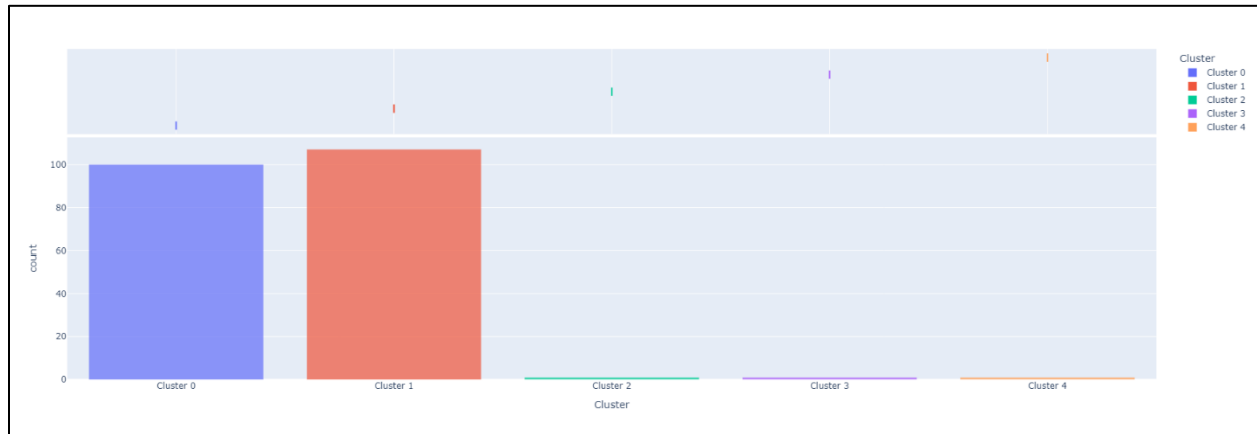
Preprocessed Data (T+N+PCA)
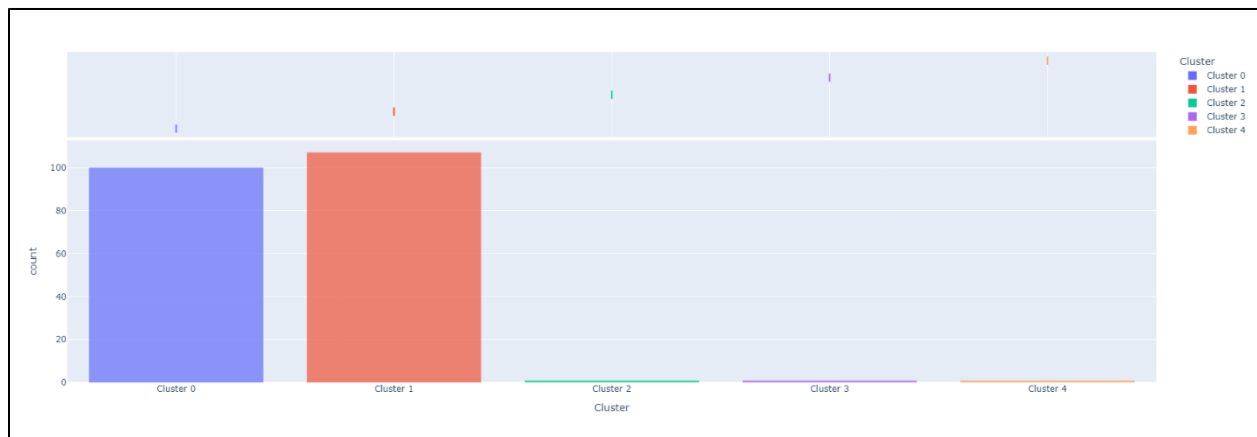
Conclussion:-

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 clustering emerges as the most robust method, particularly when data transformation is applied, achieving a good balance of cohesion and separation. Hierarchical clustering can be competitive, especially when transformations are applied, but its performance is slightly weaker overall. K-Means Shift clustering, while showing excellent cohesion with normalization, struggles with cluster separation, making it less ideal for this dataset