Customer Segmentation / Clustering

- The segmentation task has been performed using K-Means clustering algorithm
- Given no of clusters to be formed, K-Means algorithm will group the nearest data points into a specified no of clusters.
- Initial points are selected randomly as per the number of clusters specified and nearest neighbours are calculated using distances. As per the formed clusters, new centroids will be updated.
- Principle Component Analysis is done to reduce the dimensionality of the data.
- The cluster's quality is being assessed by **Davies-Bouldin Index, Calinski-Harabasz Index** and **Silhouette Score.**
- The metrices for each iteration of PCA is mentioned below as desired.

No. of PCA components	Optimal K value	DB index	Silhouette Score	CH Index
1	3	0.107	0.919	6285.646
2	4	0.151	0.899	17012.527
3	**4**	**0.085**	**0.939**	**11166.134**
4	4	0.432	0.722	396.347
5	4	0.617	0.613	222.598
6	4	0.755	0.539	158.552
7	4	0.858	0.491	127.782
8	4	0.892	0.473	117.983
9	4	0.923	0.459	110.051
10	4	0.952	0.445	103.443
11	4	0.969	0.438	99.874
12	4	0.973	0.436	99.307
13	4	0.973	0.436	99.307