

I have approached this problem in two ways, using original features and reduced features using PCA

I am using same features, that I have used for Lookalike. I have explained each feature in detail in ipynb file.

Total features created were 115 in dimension

Following are the results for each method I used

1. Kmeans Without Dimentionality Reduction:

Clusters Formed – 3 , chosen using Elbow Method

DB Index Value – 2.0877

Silhouette Score – 0.1244

Calinski-Harabasz Score – 37.85

2. Kmeans With Dimentionality Reduction:

Clusters Formed – 12 , chosen using Elbow Method

DB Index Value – 1.07

Silhouette Score – 0.06

Calinski-Harabasz Score – 15.890

While experimenting, I had reached DB Index Value of 0.96 , for 13 clusters, but after plotting them I was not satisfied, so I dropped it.