

TASK 3: Customer Segmentation / Clustering

DATA.

1. Customers.csv
The dataset contains customer information with name, region and signup date.
2. Products.csv
The dataset contains product information with product name, price and category.
3. Transactions.csv
The dataset contains purchase information of products and its buyer.

Inference:--

Here I have tried different algorithm like K-Means, Agglomerative Clustering and many more Clustering algorithms.

With the Number of Clusters ranging from 2-10 and have provided a vivid comparison between them

Best Value of k with respect to DB Index:

Looking at the DB Index values across algorithms, the lowest DBI is observed at 4 clusters for many clustering algorithms:

- K-Means: DB Index = 0.75 at 4 clusters
- Agglomerative: DB Index = 0.78 at 4 clusters
- GMM: DB Index = 0.77 at 4 clusters
- Birch: DB Index = 0.76 at 4 clusters
- Spectral: DB Index = 0.74 at 4 clusters

Best Value of k with respect to Silhouette Score:

The highest Silhouette Score occurs at 4 clusters for most clustering algorithms:

- K-Means: Silhouette Score = 0.72 at 4 clusters
- Agglomerative: Silhouette Score = 0.65 at 4 clusters
- GMM: Silhouette Score = 0.71 at 4 clusters
- Birch: Silhouette Score = 0.69 at 4 clusters
- Spectral: Silhouette Score = 0.72 at 4 clusters

Best Combined Value of k (Considering Both DB Index and Silhouette Score):

When considering both the lowest DBI and the highest Silhouette Score, 4 clusters seems to be the most optimal choice for the majority of clustering algorithms, as it minimizes the DB Index and maximizes the Silhouette Score.

Thus, the best value of k overall, considering both DBI and Silhouette Score together, is 4 clusters.

In conclusion:

Best k with respect to DBI: 4

Best k with respect to Silhouette Score: 4

Best k combined: 4

DB Index value: -

Number of Clusters considered	DB Index Value
2	
3	
4	
5	
6	
7	
8	
9	
10	