

KMeans Clustering and Davies-Bouldin Index Report

Loading the Data:

- The datasets Customers.csv, Products.csv, and Transactions.csv are loaded using pandas.
 - These datasets provide essential data for customer analysis.
-

KMeans Clustering:

- The KMeans algorithm is applied to the customer data, partitioning the dataset into 5 clusters.
 - Each customer is assigned to one of these clusters based on their features.
-

Davies-Bouldin Index:

- The Davies-Bouldin Index (DBI) measures the compactness and separability of the clusters.
 - A lower DBI indicates well-separated and compact clusters, showing the quality of the clustering.
-

PCA Visualization:

- Principal Component Analysis (PCA) reduces the customer features to 2 dimensions.
 - This allows for effective visual analysis of the clusters in a scatter plot format.
-

Results:

- **Number of clusters:** 5 clusters were identified.
- **Davies-Bouldin Index:** The calculated DBI is a measure of the clustering performance.