

In this task, the goal is to segment customers based on their profiles and transaction behaviors to identify distinct groups that share similar characteristics.

Data Preprocessing:

- First, we merged customer information from the Customers.csv file with their transaction history from the Transactions.csv file, using CustomerID as the common key.
- We then created useful features such as total spending, total quantity purchased, and average price per transaction to represent customer behavior.

Standardization:

- Since our features (spending, quantity, etc.) had different scales, we standardized the data. This ensures that each feature contributes equally to the clustering process.

Clustering with KMeans:

- We used the **KMeans** algorithm to group customers into clusters. To determine the optimal number of clusters, we applied the **Elbow Method**, where we observed how the total within-cluster sum of squares (inertia) changed with different values of k. The "elbow" point on the graph gave us the best number of clusters.

Cluster Evaluation:

- To evaluate the quality of the clusters, we calculated the **Davies-Bouldin Index (DB Index)**. A lower DB Index indicates that the clusters are well-separated and distinct, meaning the clustering is of high quality.

Visualization:

- We reduced the data to 2D using **PCA** (Principal Component Analysis) for easier visualization. This allowed us to plot the clusters and see how well they are separated from each other.

Results:

- **Number of Clusters:** After applying the Elbow Method, we found the optimal number of clusters.
- **DB Index:** We calculated the DB Index to confirm that the clusters were well-defined.
- **Visualization:** The 2D plot visually showed distinct customer groups.

This clustering process helps to identify different customer segments based on their behavior, which can be used for tailored marketing, personalized recommendations, or improving customer engagement strategies.