

Task 3: Customer Segmentation / Clustering

This involves grouping customers into distinct segments using clustering techniques based on their profile and transaction information. Here's the approach:

Steps for Task 3:

Data Preparation:

Merge Customers.csv and Transactions.csv for customer-profile + transaction-based features.

Clean and preprocess data:

Encode categorical variables (e.g., region, product category).

Normalize numerical data (e.g., transaction value, product quantity).

Feature Selection:

Profile-based features: Region, SignupDate (transformed into tenure).

Transaction-based features: Total transaction value, total quantity purchased.

Clustering:

Choose clustering algorithm (e.g., K-Means, DBSCAN).

Use 2-10 clusters and calculate metrics like the Davies-Bouldin Index (DB Index) to evaluate the clustering.

Visualization:

Use scatterplots, pair plots, and cluster heatmaps for visual representation of customer groups.

Deliverables:

Number of clusters formed.

DB Index value.

Clustering metrics and insights.