Clustering Report for E-Commerce Customers

In this project, clustering was performed on the customer transaction data with the aim to identify distinct customer segments. These segments were derived based on both customer and product information, including transaction history, quantity purchased, product category, and region. The approach involved the following key steps:

Preprocessing: Data preprocessing included scaling numerical features and encoding categorical features (such as Region and Category) using methods like **StandardScaler** and **OneHotEncoder**.

Clustering Algorithm: The **KMeans** algorithm was used, with the optimal number of clusters determined via the **Elbow Method** and **Silhouette Analysis**.

Approach

Clustering allows businesses to segment customers into actionable groups, each with unique purchasing behaviors. By analyzing these segments, businesses can tailor their marketing, product offerings, and customer engagement strategies for improved sales, customer retention, and overall business growth. The approach used in this project is scalable, allowing for the segmentation of large customer bases and better targeting of marketing efforts.

Evaluation Metrics

The choice of evaluation metric depends on whether ground truth labels are available and the nature of the clusters.

Results

2]:

	n_clusters	davies_bouldin_index
0	2	0.723365
1	3	0.772685
2	4	0.865062
3	5	0.896729
4	6	0.932103
5	7	0.928462
6	8	0.921743
7	9	1.076514
8	10	1.032008

visualization of clusters: