Clustering Report

Introduction :

Customer segmentation is an essential technique for understanding customer behavior and grouping them into meaningful categories based on their characteristics and transaction history. In this analysis, we used the eCommerce Transactions dataset to cluster customers. The aim is to help the business tailor strategies for different customer groups, improve targeting, and enhance customer satisfaction.

The analysis utilized customer profiles, transaction data, and clustering techniques to create distinct segments. Metrics such as the Davies-Bouldin Index were used to evaluate the clustering performance.

Methodology :

1. **Data Preparation**:
    - The datasets were merged to create customer-level features such as:
        - Total transaction value.
        - Total quantity purchased.
        - Number of transactions.
        - Most frequent region (Preferred Region).
    - The features were scaled to standardize values and encoded where necessary.
2. **Clustering Algorithm**:
    - K-Means clustering was used to segment customers.
    - The optimal number of clusters was determined using the Elbow Method, where the inertia value (within-cluster variance) was analyzed for cluster numbers ranging from 2 to 10.
3. **Evaluation**:
    - The clustering quality was assessed using the Davies-Bouldin Index, which measures the separation and compactness of clusters.

Results :

1. **Number of Clusters**:
    - Based on the Elbow Method, the optimal number of clusters was determined to be **4**. These clusters represent distinct customer segments with unique characteristics.
2. **Davies-Bouldin Index**:
    - The Davies-Bouldin Index for the clustering was calculated as **0.74** (lower is better), indicating well-separated and compact clusters.
3. **Cluster Characteristics**:
    - **Cluster 1**: Customers with high transaction volumes and total value, indicating frequent, high-value buyers. Likely to be loyal customers or businesses.

- **Cluster 2**: Customers with moderate transaction activity and balanced purchases across regions, representing regular shoppers.
- **Cluster 3**: Low transaction frequency and total value, potentially new or occasional buyers.
- **Cluster 4**: Region-specific buyers with high activity in specific product categories, indicating regional product preferences.

4. **Visualization**:
    - Clusters were visualized using PCA (Principal Component Analysis) to reduce the features to two dimensions. The scatterplot revealed clear distinctions between clusters, further supporting the quality of the segmentation.

Conclusion :

The clustering analysis successfully segmented the customers into four meaningful groups, each with unique behavioral and transactional attributes. These insights can help the business:

- Focus on high-value customers with loyalty programs (Cluster 1).
- Design campaigns for moderate and occasional buyers to increase their activity (Clusters 2 and 3).
- Tailor region-specific offers for customers with localized preferences (Cluster 4).

This segmentation provides a solid foundation for personalized marketing, improved customer retention, and strategic decision-making.