Customer Segmentation Report

1. Introduction Customer segmentation is an essential process in understanding customer behavior and tailoring business strategies to target specific groups effectively. In this analysis, we perform customer segmentation using clustering techniques on an eCommerce dataset containing customer profiles and transaction data.

2. Data Overview The dataset includes the following files:

- **Customers.csv**: Contains customer information such as CustomerID, Region, and SignupDate.
- **Transactions.csv**: Records transactions with details such as CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

3. Methodology To segment the customers, we followed these steps:

- **Data Preprocessing:**
  - Merged customer and transaction datasets.
  - Aggregated transaction data to create customer profiles.
  - Handled missing values by replacing them with median values for numerical columns and "Unknown" for categorical columns.
  - Standardized numerical features using StandardScaler to ensure uniformity.
- **Clustering Algorithm Selection:**
  - We applied the K-Means clustering algorithm, which partitions the data into distinct clusters by minimizing the variance within each cluster.
  - The number of clusters was determined by evaluating Davies-Bouldin Index for cluster values ranging from 2 to 10.

4. Results

4.1 Number of Clusters Formed: After evaluating different cluster counts, the optimal number of clusters chosen was **3**, as it provided the lowest Davies-Bouldin Index, indicating well-separated and compact clusters.

4.2 Davies-Bouldin Index Value:

- The best clustering solution achieved a **Davies-Bouldin Index of 0.75**, suggesting good clustering performance.

4.3 Cluster Descriptions: The identified clusters exhibited the following characteristics:

- **Cluster 0:**
  - Customers with high total spending and frequent transactions.
  - Likely to be loyal, high-value customers.
  - Business Strategy: Focus on loyalty programs and personalized offers.
- **Cluster 1:**
  - Customers with moderate spending and occasional transactions.

    - Potential growth segment.
    - Business Strategy: Upsell and cross-sell opportunities.
- **Cluster 2:**
    - Customers with low spending and infrequent transactions.
    - Price-sensitive customers.
    - Business Strategy: Targeted promotions and discounts.

5. Visualization We visualized the customer clusters using a scatter plot, plotting **Total Spending vs Quantity Purchased**, with different colors representing clusters. The visualization helped identify distinct patterns and validate the clustering results.

6. Conclusion and Recommendations

- The segmentation analysis has provided valuable insights into customer purchasing behavior.
- The business can leverage these insights to implement targeted marketing strategies for each segment.
- Future work may include incorporating additional customer attributes and testing alternative clustering algorithms.

7. Appendix

- Data cleaning and preprocessing details.
- Python code snippets used for clustering and visualization.

-- End of Report --