Customer Segmentation Report

Objective: The goal is to segment customers based on their transaction data and profile, using clustering techniques.

1. Data Preprocessing

The customer and transaction data were merged using `CustomerID`. Missing values were handled by filling them with zeros. Key features for clustering were created:

- **Total Spend**: Sum of total value spent by each customer.
- **Quantity**: Sum of total quantity bought by each customer.

2. Clustering Methodology

KMeans was used for clustering. The optimal number of clusters was determined through the **Elbow Method**, which suggested **3 clusters**. The clustering quality was evaluated using the **Davies-Bouldin Index**, which yielded a value of **0.7087**, indicating good separation between clusters.

3. Cluster Characteristics

- **Cluster 0**: High spenders, large quantity purchases (Avg Total Spend: 5876.53, Quantity: 21.04).
- **Cluster 1**: Low spenders, fewer purchases (Avg Total Spend: 1510.97, Quantity: 6.35).
- **Cluster 2**: Moderate spenders, medium quantity purchases (Avg Total Spend: 3459.86, Quantity: 12.50).

4. PCA Visualization

PCA was applied to reduce data dimensions and visualize clusters. A scatter plot showed clear separation of the clusters.

5. Conclusion

Three distinct customer groups were identified based on their spending behavior. This segmentation can guide targeted marketing strategies and improve customer engagement.