Customer Segmentation Report

The goal of this analysis is to perform customer segmentation using clustering techniques based on customer profile and transaction data. The insights from this segmentation can guide targeted marketing strategies and optimize customer engagement and retention.

Clustering Methodology

Preprocessing

- **Data Cleaning**: Missing values were imputed with the column mean.
- **Feature Scaling**: Normalized the data using StandardScaler to ensure uniformity across features.
- **Feature Engineering**: Merged customer profiles and transaction data into a single dataset with the following key features:
    - **Total Spend**: Sum of all transaction values per customer.
    - **Transaction Frequency**: Number of unique transactions.
    - **Average Transaction Value**: Ratio of total spend to transaction frequency.
    - **Total Quantity Purchased**: Sum of quantities in all transactions.

Clustering Approach

- **Algorithm**: K-Means clustering.
- **Cluster Evaluation**: Tested cluster sizes from **k = 2 to k = 10** using two key metrics:
    - **Davies-Bouldin Index (DB Index)**: Lower values indicate better clustering (compact and well-separated clusters).
    - **Silhouette Score**: Higher values indicate better-defined clusters.

Results

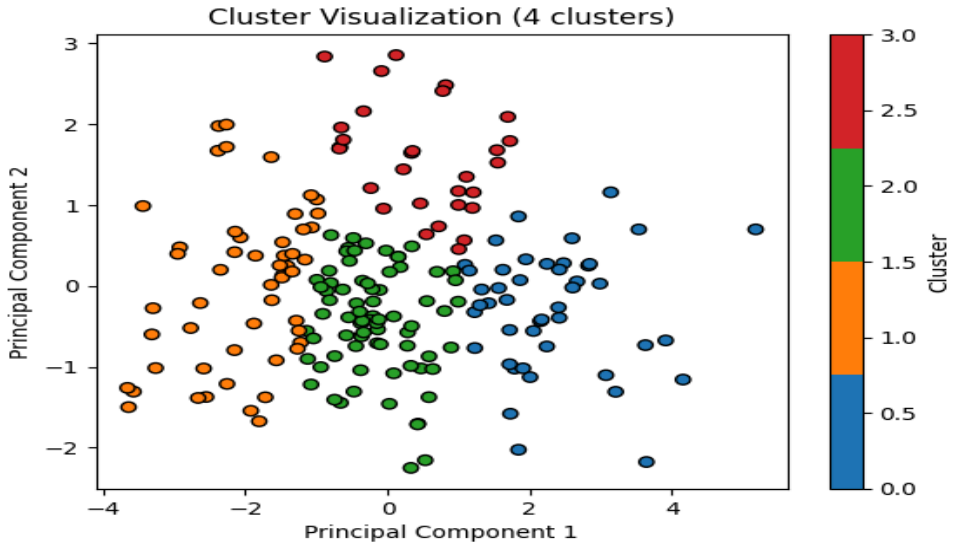
- **Optimal Number of Clusters**

Based on the evaluation metrics, **4 clusters** were selected as the optimal configuration:

- **Cluster Metrics**

Number of Clusters (k)	DB Index	Silhouette Score
2	0.942	0.395
3	1.101	0.299
4	0.930	0.346
5	0.952	0.310

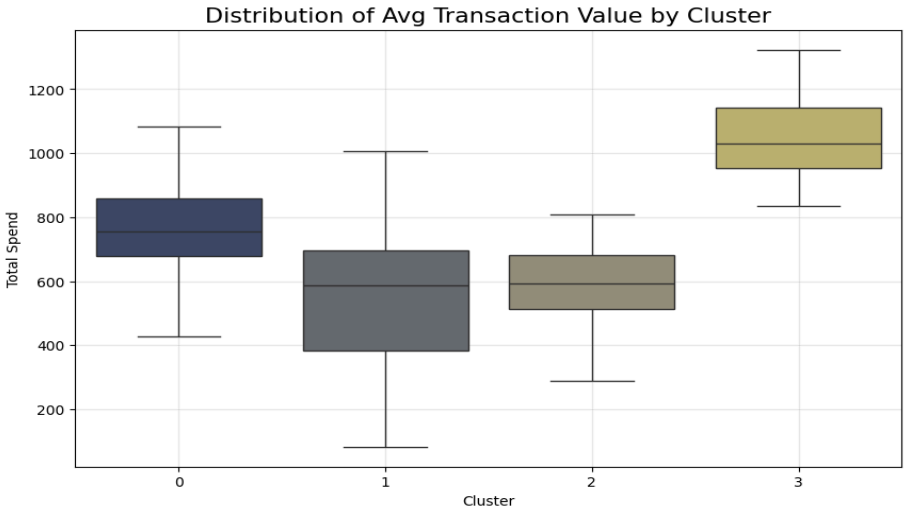
- **Visualizations**

1. **Cluster Visualization (PCA 2D Plot)**:
    - A Principal Component Analysis (PCA)-based scatter plot visualizes the separation between the 4 clusters.
    - 

    - The plot demonstrates meaningful segmentation, highlighting distinct customer groups.
2. **Distribution of Total Spend by Cluster**:

- A boxplot illustrates how the total spending varies across the clusters.

3. **Other Potential Plots**:
   - **Distribution of Average Transaction Value by Cluster**: Highlights spending behaviors within each cluster.



Business Insights

Cluster Descriptions

- **Cluster 0: High-Value Customers**
    - These customers exhibit high total spending and purchase frequency.
    - They represent the business's most valuable segment.
    - **Action**: Prioritize these customers for loyalty programs, exclusive offers, and personalized rewards.
- **Cluster 1: Moderate-Value Customers**
    - Moderate total spending with consistent purchasing behavior.
    - These customers are engaged but have potential for higher spending.
    - **Action**: Target them with promotions, discounts, and upselling opportunities.
- **Cluster 2: Low-Value Customers**
    - Low total spending and sporadic purchase behavior.
    - Likely to be disengaged or infrequent buyers.
    - **Action**: Focus on re-engagement campaigns and personalized offers to improve activity.
- **Cluster 3: Growing Customers**
    - These customers exhibit medium spending with signs of growth potential.
    - Likely to be influenced by cross-selling strategies.
    - **Action**: Use personalized marketing to nurture these customers into high-value segments.

Recommendations

- Leverage **Cluster 0** insights to create loyalty programs and deliver premium services to sustain engagement.
- For **Cluster 1**, drive increased spending through targeted promotions and upselling opportunities.
- Re-engage **Cluster 2** with retention-focused initiatives such as discounts, incentives, and tailored communication.
- Build personalized offers for **Cluster 3** to unlock growth potential and encourage them to transition into Cluster 0 or 1.

Conclusion

The customer segmentation analysis successfully identified 4 meaningful clusters, providing actionable insights for targeted marketing. By focusing on tailored strategies for each segment, the business can enhance customer satisfaction, drive revenue, and build long-term loyalty.