Customer Segmentation Using Clustering Techniques

Introduction:

Customer segmentation is a critical process in understanding different customer groups based on their behavior and transactions. By clustering customers based on their profile information and transaction data, we can identify distinct segments, which can help in personalized marketing, sales strategies, and improving customer satisfaction. In this task, we applied a clustering technique to segment customers into distinct groups.

Clustering Process:

For this analysis, we used the K-means clustering algorithm, which is a popular unsupervised learning technique for partitioning data into clusters based on similarity. We chose K-means due to its simplicity and efficiency. The features selected for clustering were Total Value and Quantity of transactions from the customer's transaction history.

- **Number of Clusters**: We experimented with several numbers of clusters ranging from 2 to 10. After evaluation, we selected **4 clusters** as the optimal choice.
- **Preprocessing**: The data was standardized using StandardScaler to ensure each feature had equal weight in the clustering process.

Evaluation Metrics:

The clustering performance was evaluated using the Davies-Bouldin (DB) Index, a metric that quantifies the compactness and separation of the clusters. The lower the DB index, the better the clustering solution.

DB Index for 4 clusters: **0.835**. This indicates that the clusters are well-separated with minimal overlap.

Clustering Results: After applying K-means clustering, we obtained the following customer segments:

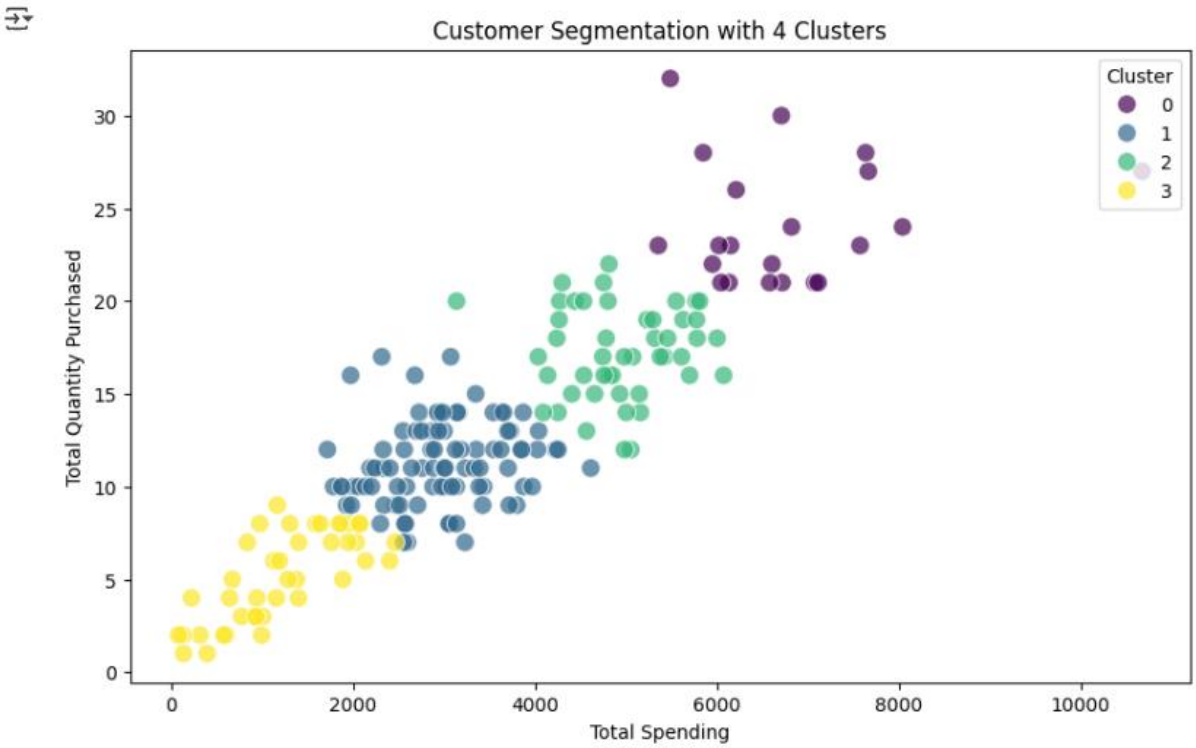
- **Cluster 1**: Customers with high transaction value but moderate quantity of transactions, likely representing premium customers.
- **Cluster 2**: Customers with moderate transaction value and high transaction frequency, possibly indicating budget-conscious, regular shoppers.
- **Cluster 3**: Customers with low transaction values and low quantity, possibly sporadic or low-value customers.
- **Cluster 4**: Customers with high quantity but lower total transaction value, potentially indicating frequent buyers with smaller transactions.

Visualizations:

The clustering results were visualized using a **scatter plot**, where customers are plotted based on their **Total Value** and **Quantity**. Each cluster is color-coded for easy identification, showing clear separation between customer groups.

Visualizations:

The clustering results were visualized using a **scatter plot**, where customers are plotted based on their **Total Value** and **Quantity**. Each cluster is color-coded for easy identification, showing clear separation between customer groups.



Conclusion:

The customer segmentation results provide valuable insights into different customer behaviors. The distinct customer segments allow businesses to target marketing and sales efforts more effectively. For example:

- **Cluster 1** could be targeted with premium offers and loyalty programs.
- **Cluster 2** may benefit from promotions encouraging more frequent purchases.
- **Cluster 3** may require re-engagement strategies to increase transaction value.

This segmentation helps businesses understand customer profiles better, which can lead to increased customer satisfaction and targeted marketing efforts.

Business Insights Derived from Exploratory Data Analysis (EDA)

1. **Top Customers by Transaction Value**

The top 5 customers account for a large portion of the company's revenue. Targeting these high-value customers with personalized offers or loyalty programs could improve retention and increase sales. Understanding their common traits (e.g., location, preferences) enables tailored marketing strategies.

2. Product Category Popularity

Certain product categories exhibit higher total sales. By focusing on these high-performing categories, the company can optimize inventory and marketing efforts. Promotions targeting popular products could drive increased sales, and bundling strategies can further enhance cross-selling opportunities.

3. Regional Sales Distribution

Sales show regional disparities, with some areas contributing significantly more to total revenue. This suggests the need for targeted marketing in underperforming regions. Additionally, localized product offerings and strategic partnerships can help boost sales where necessary.

4. Transaction Frequency vs. Total Spend

Customers who make frequent purchases tend to spend more overall. This insight suggests that incentivizing customers to increase transaction

frequency (e.g., through loyalty programs) could lead to higher lifetime value per customer and stronger revenue growth.

5. Impact of Quantity Purchased on Sales

Customers who purchase in larger quantities often have higher transaction values. Introducing bulk purchase discounts or offers could help increase average transaction value and drive sales growth, particularly for high-demand products or categories.

Conclusion

These insights highlight key areas for business growth, such as focusing on high-value customers, optimizing product offerings, and targeting specific regions for better performance. Utilizing this data effectively can guide marketing strategies, improve customer engagement, and ultimately drive increased revenue.

CustomerID	Recommendations
C0001	C0085: 1.00, C0042: 1.00, C0089: 1.00
C0002	C0157: 1.00, C0166: 1.00, C0029: 1.00
C0003	C0111: 0.99, C0160: 0.99, C0147: 0.99
C0004	C0162: 1.00, C0165: 1.00, C0090: 1.00
C0005	C0080: 1.00, C0167: 1.00, C0177: 1.00
C0006	C0079: 1.00, C0117: 1.00, C0196: 0.99
C0007	C0146: 1.00, C0125: 1.00, C0061: 1.00
C0008	C0109: 1.00, C0136: 1.00, C0124: 1.00
C0009	C0015: 1.00, C0131: 1.00, C0193: 1.00
C0010	C0176: 1.00, C0027: 0.99, C0030: 0.98
C0011	C0139: 1.00, C0100: 1.00, C0023: 1.00
C0012	C0101: 1.00, C0156: 1.00, C0093: 1.00
C0013	C0067: 1.00, C0138: 1.00, C0141: 1.00
C0014	C0192: 1.00, C0159: 1.00, C0186: 1.00
C0015	C0009: 1.00, C0131: 1.00, C0120: 1.00
C0016	C0154: 1.00, C0040: 1.00, C0026: 0.99
C0017	C0075: 1.00, C0179: 1.00, C0039: 1.00
C0018	C0045: 1.00, C0041: 1.00, C0068: 1.00
C0019	C0142: 1.00, C0174: 1.00, C0055: 1.00
C0020	C0058: 1.00, C0193: 1.00, C0198: 1.00