REPORT TITLE
Customer Segmentation Analysis

1. Executive Summary

This report presents the results of a customer segmentation analysis performed on a dataset containing customer transaction and profile information. The goal was to group customers into distinct clusters based on their purchasing behavior and demographic characteristics. The analysis used the K-Means clustering algorithm and evaluated the results using metrics such as the Davies-Bouldin Index, Silhouette Score, and Calinski-Harabasz Index. The findings provide actionable insights for targeted marketing and customer relationship management.

2. Methodology

Data Preparation:

- The dataset was cleaned and preprocessed to handle missing values and normalize the features.
- New features (Recency, Frequency, and Monetary Value) were created to better capture customer behavior.

Clustering:

- The Elbow Method was used to determine the optimal number of clusters.
- The K-Means algorithm was applied to group customers into 4 clusters.

Evaluation: Clustering quality was evaluated using the following metrics:

- Davies-Bouldin Index: 1.860
- Silhouette Score: 0.167
- Calinski-Harabasz Index: 194.44

3. Results

Optimal Number of Clusters:

The Elbow Method suggested 4 clusters as the optimal number. The graph showed a gradual decrease in inertia without a clear "elbow," indicating that the data does not have a strong natural clustering structure.

Clustering Metrics:

Davies-Bouldin Index: 1.486408692756587

Interpretation: Lower values indicate better clustering. A score of 1.860 suggests moderate clustering quality.

Silhouette Score: 0.1732435039719036

Interpretation: Values range from -1 to 1. A score of 0.167 is closer to 0, indicating weak clustering quality.

Calinski-Harabasz Index: 222.42986517780616

Interpretation: Higher values indicate better clustering. A score of 194.44 suggests moderate clustering quality.

4. Insights

Cluster Profiles:

- **Cluster 0:** Customers with moderate spending and frequency.
- **Cluster 1:** High-value, frequent customers.
- **Cluster 2:** Low-value, infrequent customers.
- **Cluster 3:** New customers with low recency and moderate spending.

Actionable Recommendations:

- Cluster 1: Offer loyalty rewards to retain these high-value customers.
- Cluster 2: Re-engage with special offers to prevent churn.
- Cluster 3: Target with personalized offers to increase engagement.
- Cluster 0: Maintain regular communication to strengthen loyalty.

5. Interpretation of Visualizations

Elbow Method Graph:

The graph shows the inertia (within-cluster sum of squares) for different numbers of clusters (2 to 10).

The absence of a clear "elbow" suggests that the data does not have a strong natural clustering structure. However, 4 clusters were chosen as a reasonable compromise between clustering quality and interpretability.

PCA Plot:

The PCA plot shows the clusters in 2D space. While some separation is visible, there is significant overlap between clusters, indicating that the clusters are not well-defined.

This suggests that the features used for clustering may not fully capture the underlying structure of the data.

t-SNE Plot:

The t-SNE plot provides a more detailed view of the clusters. It shows that the data points are spread out, with some clusters overlapping.

This further confirms that the data does not have a strong natural clustering structure and that the clusters may not be well-separated.

6. Evaluation of Metrics

- **Davies-Bouldin Index:** 1.486408692756587
- Interpretation: A lower value indicates better clustering. A score of 1.4864 suggests moderate clustering quality.
- **Silhouette Score :** 0.1732435039719036
- Interpretation: Values range from -1 to 1. A score of 0.173 is closer to 0, indicating moderate clustering quality.
- **Calinski-Harabasz Index** (222.42986517780616):
- Interpretation: Higher values indicate better clustering. A score of 222.43 suggests moderate clustering quality.

7. Conclusion

The customer segmentation analysis successfully grouped customers into 4 distinct clusters based on their purchasing behavior. While the clustering metrics indicate moderate quality, there is room for improvement. The PCA and t-SNE visualizations show that the clusters are not well-separated, suggesting that the data does not have a strong natural clustering structure. Nonetheless, the results provide valuable insights for targeted marketing strategies and customer relationship management.