Customer Segmentation Report

Objective:
The goal of this analysis was to perform customer segmentation using clustering techniques. We aimed to identify distinct customer groups based on transaction data and profile information, which would help in formulating targeted business strategies.

Clustering Approach:
The clustering process was carried out using the KMeans algorithm, which is one of the most popular clustering techniques. To determine the optimal number of clusters, the Elbow Method was employed. The Elbow plot showed that the optimal number of clusters was 5, as the inertia (sum of squared distances between data points and their centroids) began to level off after this point.

Clustering Metrics:

- **Davies-Bouldin Index:** The Davies-Bouldin Index, which measures the average similarity ratio of each cluster with the most similar cluster, was calculated to be 1.04. A lower Davies-Bouldin Index indicates better separation between clusters. In this case, the value suggests that the clusters are relatively well separated, although there may still be some overlap.
- **Silhouette Score:** The Silhouette Score, which measures how similar a data point is to its own cluster compared to other clusters, was found to be 0.30. A score closer to 1 indicates well-defined clusters, while scores close to 0 suggest overlapping clusters. The score of 0.30 indicates moderate separation, with potential for further refinement in cluster definitions.

Cluster Analysis:
Once the clusters were formed, the following key metrics were analyzed for each cluster:

- **Monetary Value:** The total amount spent by the customers in each cluster.
- **Frequency:** The number of transactions made by the customers in each cluster.
- **Recency:** The number of days since the customer's most recent transaction.

The analysis revealed that Cluster 2 contained the highest average monetary value and frequency of transactions, indicating that these customers are high-value and frequent buyers. Conversely, Cluster 0 had the lowest average monetary value and frequency, suggesting that these customers are less engaged and may require strategies to improve their retention.

Principal Component Analysis (PCA):
To visualize the clustering results, PCA was applied to reduce the data to two dimensions. The resulting 2D scatter plot showed distinct separation between the clusters, although there was some overlap, particularly in the less-engaged customer segments. This visualization highlights the differentiation between customer behaviors, though further segmentation could be explored.

Business Implications:
This segmentation provides valuable insights for targeted marketing and customer retention strategies. For example, high-value customers in Cluster 2 could be offered exclusive promotions, while customers in Cluster 0 might benefit from re-engagement campaigns. Additionally, further analysis and fine-tuning of clusters can improve the targeting and customization of business strategies.

Conclusion:
In conclusion, the customer segmentation using KMeans clustering was successful in identifying

distinct customer groups based on their transaction and profile data. The insights derived from these clusters can guide strategic business decisions, improve customer engagement, and drive targeted marketing campaigns. The clustering process, while yielding useful results, also suggests opportunities for refinement and optimization in further iterations.