Customer Segmentation / Clustering

1. Number of Clusters Formed:

The elbow method, applied to the inertia values across a range of cluster numbers (1 to 11), identified 8 clusters as the optimal number for this dataset. This was determined by observing a noticeable "elbow" in the plot where the rate of decrease in inertia slowed significantly. The selection of 8 clusters provides a balance between achieving granular customer segmentation and avoiding overfitting.

Each customer has been assigned to one of these clusters, allowing for in-depth segmentation based on behavior and other relevant features. These clusters will serve as the foundation for personalized marketing strategies and targeted interventions.

2. Davies-Bouldin (DB) Index:

The Davies-Bouldin Index for the clustering solution is **[DB Index Value]**. This metric evaluates clustering quality by considering both intra-cluster compactness and inter-cluster separation. A lower DB Index value indicates a better-defined clustering structure. While this value provides an initial understanding of clustering quality, further validation through additional metrics like the silhouette score is recommended for a more comprehensive evaluation.

3. Clustering Methodology:

- **Algorithm Used**: KMeans clustering was chosen for its efficiency and interpretability in handling large datasets. It partitions the data into clusters by minimizing the variance within clusters.
- **Feature Preparation**: Before clustering, features were standardized using the **StandardScaler** to ensure that all variables contributed equally to the distance calculations, as KMeans is sensitive to scale.
- **Cluster Assignment**: The dataset was processed to assign each data point (customer) to one of the 8 clusters, enabling targeted analysis of these distinct segments.

4. Additional Metrics and Observations:

1. **Inertia**: The total within-cluster sum of squares (inertia) decreased significantly with increasing clusters but plateaued after 8 clusters, confirming the choice of 8 as the optimal number.
2. **Cluster Characteristics**: Each cluster represents customers with similar purchasing behaviors, such as total spending, product preferences, or transaction frequencies. Detailed profiling will reveal specific traits of each segment.
3. **Scalability**: The chosen method scales well with larger datasets, making it suitable for potential future expansions of the dataset.

5. Practical Implications:

- **Customer Segmentation**: The formation of 8 clusters allows the business to identify distinct customer groups and tailor marketing strategies, loyalty programs, or product recommendations to each segment.
- **Resource Optimization**: Understanding the characteristics of high-value clusters enables the business to focus its resources on retaining and expanding these segments.
- **Improved Engagement**: Clusters can be used to predict customer behavior, enabling proactive engagement strategies to increase retention and lifetime value.