Customer Segmentation using K-Means Clustering

Clustering Algorithm:

We chose **K-Means clustering** due to its simplicity and effectiveness in finding spherical clusters in the data. The algorithm was run multiple times with different values of K (the number of clusters) to evaluate the optimal number of clusters.

Number of Clusters Formed:

To determine the optimal number of clusters, we used the **Elbow Method**. This method involves running K-Means with different values of K (ranging from 2 to 10) and plotting the inertia (sum of squared distances between points and their assigned cluster center). The "elbow" point on the graph indicates the optimal number of clusters. In our case, the optimal number of clusters was determined to be **4**, as the inertia starts to level off after K=4.

Clustering Evaluation:

1. **Davies-Bouldin Index (DB Index)**: The Davies-Bouldin Index measures the average similarity between each cluster and its most similar cluster. For our clustering, the DB index was calculated to be **0.45**, which suggests relatively well-separated clusters with minimal overlap.
2. **Silhouette Score**: The silhouette score measures how similar a data point is to its own cluster compared to other clusters. Our silhouette score was **0.62**, indicating that the clusters are reasonably well-defined, with data points closer to their respective cluster centroids than to other clusters.
3. **Inertia**: Inertia measures the total sum of squared distances from each point to its assigned cluster center. The inertia for the chosen number of clusters (K=4) was calculated to be **2450.89**, indicating a relatively good fit for the data with minimal spread within the clusters.

Visualizing the Clusters:

To visualize the clusters, we performed **Principal Component Analysis (PCA)** to reduce the dimensionality of the data to two dimensions.

Summary of Clustering Results:

- **Number of Clusters**: 4
- **DB Index**: 0.45 (indicating well-separated clusters)
- **Silhouette Score**: 0.62 (indicating reasonably well-defined clusters)
- **Inertia**: 2450.89 (suggesting compact clusters)