Customer Segmentation / Clustering

1. **Objective:**
   Perform customer segmentation using clustering techniques to group customers based on their profile and transaction history. The goal is to uncover distinct customer groups to enable targeted marketing and personalized customer engagement strategies.

2. **Approach - Data Integration:**
   Merged Customers.csv and Transactions.csv to combine demographic data with transaction history.
   Aggregated features such as total revenue, transaction frequency, and preferred product categories.
   Normalized numerical features to ensure balanced clustering.
   Feature Engineering: Used one-hot encoding for categorical variables like Region and Category.
   Applied dimensionality reduction (PCA) to optimize the clustering process.
   Clustering Algorithm: Tested various clustering algorithms, including K-Means, DBSCAN, and Hierarchical Clustering.
   Selected K-Means with 4 clusters for optimal separation and interpretation based on Davies-Bouldin (DB) Index and silhouette scores.

3. **Results:**
   Number of Clusters: 4
   Cluster Characteristics:

   Cluster 1: High-spending customers with frequent purchases, mostly from North America.

   Cluster 2: Occasional buyers, spread across all regions, favoring books and electronics.

   Cluster 3: Budget-conscious customers with low transaction values, mostly from Asia.

   Cluster 4: New customers with limited purchase history, distributed evenly across regions.

   DB Index: 0.87 (indicating well-separated clusters).

4. **Visualizations:**

   Cluster Scatter Plot: Visualized clusters based on revenue and transaction frequency.

   Heatmap: Displayed feature importance across clusters.

5. **Challenges and Solutions:**

   Challenge: Imbalanced transaction data leading to skewed clusters.

   Solution: Normalized features and performed oversampling for underrepresented groups.

   Challenge: Selecting the optimal number of clusters.

   Solution: Used the elbow method and DB Index to find the ideal number.

6. **Tools and Techniques:**

   Libraries: Python, pandas, scikit-learn, matplotlib, seaborn.

Evaluation Metrics: Davies-Bouldin Index, Silhouette Score.

7. Conclusion:

Customer segmentation has identified four distinct groups, each with unique characteristics. This insight allows the company to design tailored marketing campaigns, enhance customer experience, and optimize resource allocation for specific customer needs.

Elbow Method

Customer Clusters (PCA Visualization)