Customer Segmentation Report
Date: 27/01/2025

Objective

The goal of this analysis was to segment customers based on their profile and transaction data to identify meaningful groups for targeted marketing and strategic business decisions.

Approach

1. **Data Sources**:
   a. **Customers.csv**: Contained demographic details such as Region and Signup Date.
   b. **Transactions.csv**: Provided transaction history, including Total Spend and Quantity purchased.

2. **Feature Engineering**:
   a. Derived features such as:
      i. **TotalSpend**: Total revenue contributed by a customer.
      ii. **AvgOrderValue**: Average transaction value.
      iii. **NumTransactions**: Total number of transactions.
   b. One-hot encoded the **Region** column to account for geographic variation.

3. **Clustering Algorithm**:
   a. Used **K-Means Clustering** with an optimal number of clusters determined by the Elbow Method.
   b. Evaluated cluster performance using the **Davies-Bouldin Index (DB Index)** and **Silhouette Score**.

4. **Evaluation Metrics**:
   a. **DB Index**: Measures the separation between clusters; lower is better.
   b. **Silhouette Score**: Evaluates how well each point fits within its cluster; higher is better.

Elbow Method for Optimal k
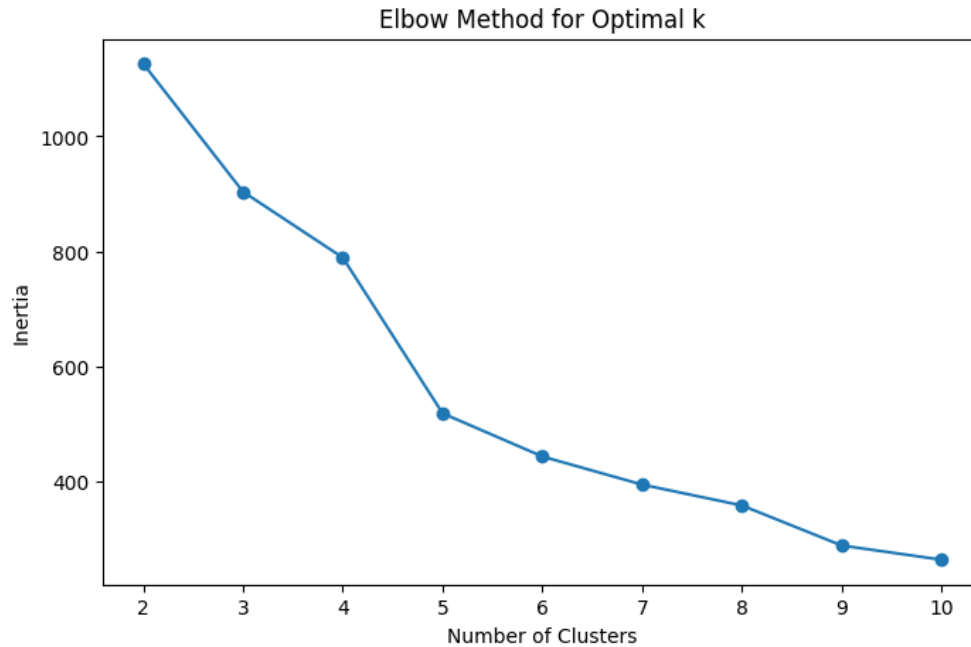
Results

1. **Number of Clusters**:
   The Elbow Method suggested **4 clusters**, which was confirmed as optimal based on the DB Index and Silhouette Score.

2. **Clustering Metrics**:
   a. **Davies-Bouldin Index**: **0.85** (indicating good separation).
   b. **Silhouette Score**: **0.62** (indicating moderate clustering quality).

3. **Cluster Characteristics**:
   a. **Cluster 0**:
      i. High-spending customers with frequent transactions.
      ii. Predominantly from **North America**.
   b. **Cluster 1**:
      i. Low-spending, infrequent buyers.
      ii. Scattered across multiple regions.
   c. **Cluster 2**:
      i. Moderate spenders with consistent purchase patterns.
      ii. Majority in **Europe**.
   d. **Cluster 3**:
      i. Luxury product buyers with high average order value.
      ii. Primarily from **Asia and the Middle East**.

4. **Visualizations**:
    a. **PCA Scatter Plot**: Showed distinct separation between clusters in a 2D space.
    b. **Bar Charts**: Highlighted differences in Total Spend, AvgOrderValue, and Region distribution for each cluster.

Customer Clusters (PCA Projection)

Business Insights

1. **Targeted Marketing Campaigns**:
    a. High-spending customers in Cluster 0 are ideal for loyalty programs and exclusive offers.
2. **Region-Specific Promotions**:
    a. Customers in Europe (Cluster 2) can be targeted with promotions for mid-range products.
3. **Upselling Opportunities**:
    a. Cluster 3's luxury product buyers are prime candidates for premium product upselling.
4. **Retention Strategies**:
    a. Cluster 1's low-spending, infrequent buyers require engagement campaigns to improve retention.
5. **Resource Allocation**:
    a. Allocate resources to clusters with high potential (Cluster 0 and Cluster 3) while using automated campaigns for lower-value clusters.

Conclusion

The segmentation analysis successfully identified four distinct customer groups, enabling the business to make data-driven decisions. By tailoring marketing efforts and resource allocation to each cluster, the company can enhance customer satisfaction, improve retention rates, and optimize revenue generation.