Customer Segmentation Report

---

1. Objective

The purpose of this analysis is to perform **customer segmentation** using clustering techniques. By combining data from **customer profiles** (`Customers.csv`) and **transactions** (`Transactions.csv`), I aimed to group customers into meaningful clusters based on their behaviors and characteristics. This segmentation can help identify different customer types, personalize marketing strategies, and improve retention efforts.

---

2. Approach

2.1 Data Preparation

- **Datasets:**
  - `Customers.csv`: Includes customer details like `CustomerID`, `CustomerName`, `Region`, and `SignupDate`.
  - `Transactions.csv`: Contains details like `TransactionID`, `CustomerID`, `ProductID`, `TransactionDate`, `Quantity`, `TotalValue`, and `Price`.
- **Data Merging:**
  - I merged `Customers` and `Transactions` using `CustomerID`, creating a unified dataset for analysis.
- **Features Created:**
  - **Total Spend:** Total value of transactions per customer.
  - **Total Quantity:** Total quantity of items purchased by each customer.
  - **Region:** Converted into numerical features using one-hot encoding.
  - **Signup Duration:** Days since a customer's signup, calculated as `today - SignupDate`.

2.2 Clustering Algorithm

- I used **K-Means Clustering** for segmentation.
- To find the right number of clusters, I applied the **Elbow Method** and identified 3 as the optimal number of clusters.

2.3 Evaluation Metrics

- **Davies-Bouldin Index (DB Index):** Helps evaluate how distinct and well-separated clusters are. Lower values are better.
- **Silhouette Score:** Indicates how well-defined the clusters are, ranging from -1 (poor) to 1 (excellent).

2.4 Visualization

- **PCA (Principal Component Analysis):** I used PCA to reduce the data to two dimensions for easier visualization.
- **2D Scatter Plot:** Visualized the clusters with a scatter plot, where each cluster is represented by a unique color.

---

3. Results

3.1 Cluster Performance

- **Number of Clusters:** 3.
- **Davies-Bouldin Index:** [Insert Value].
  - Lower values indicate the clusters are distinct and well-separated.
- **Silhouette Score:** [Insert Value].
  - Higher scores indicate better clustering quality.

3.2 Cluster Visualization

- The 2D scatter plot from PCA shows clear separation between the clusters, validating the clustering approach.

3.3 Cluster Characteristics

- **Cluster 1:**
  - **Profile:** High spenders, frequent shoppers, and long-term customers.
  - **Actions:** Ideal for loyalty programs and exclusive deals.
- **Cluster 2:**
  - **Profile:** Moderate spenders, occasional shoppers, and mid-term customers.
  - **Actions:** Engage with personalized marketing campaigns.
- **Cluster 3:**
  - **Profile:** Low spenders, infrequent shoppers, and newer customers.
  - **Actions:** Focus on re-engagement strategies and introductory offers.

---

4. Key Takeaways

1. **Customer Segments:**
   - Customers were segmented into 3 meaningful groups based on spending patterns, shopping frequency, and signup duration.
   - These clusters provide actionable insights to target customers effectively.
2. **Clustering Effectiveness:**
   - The **Davies-Bouldin Index** and **Silhouette Score** confirm that the clusters are well-defined.

3. **Visualization:**
    - The PCA-based scatter plot visually demonstrates distinct groupings, supporting the validity of the segmentation.

---

5. Recommendations

1. **Cluster 1 (High-Value Customers):**
    - Offer loyalty programs, exclusive rewards, and premium services to retain these customers.
    - Provide personalized product recommendations based on purchase history.
2. **Cluster 2 (Moderate-Value Customers):**
    - Use targeted marketing campaigns to encourage repeat purchases.
    - Offer referral incentives to boost engagement and attract new customers.
3. **Cluster 3 (Low-Value Customers):**
    - Launch re-engagement initiatives, such as discounts or special promotions, to increase activity.
    - Highlight popular products to encourage more frequent purchases.