Customer Segmentation Report

1. Introduction

Customer segmentation is an essential data-driven process that helps businesses categorize their customers into distinct groups based on shared characteristics. This segmentation enables targeted marketing strategies, personalized customer experiences, and better resource allocation.

The objective of this analysis is to segment customers using clustering techniques. By leveraging transactional and profile data, we aim to understand customer behaviour, identify key groups, and provide actionable insights to optimize business performance.

---

2. Data Preparation

Datasets Used:

1. **Customers.csv**
    - Contains customer profile information such as CustomerID, Name, Region, and SignupDate.
2. **Transactions.csv**
    - Includes details of transactions such as TransactionID, CustomerID, ProductID, Quantity, TotalValue, and TransactionDate.

Preprocessing Steps:

- **Data Cleaning:**
    - Removed duplicate entries and handled missing values (e.g., filling missing TotalValue).
    - Ensured consistency in data types (e.g., converting dates to datetime format).
- **Feature Engineering:**
    - Aggregated transactional data to compute key metrics such as:
        - **TotalSpent:** Total monetary value spent by a customer.
        - **PurchaseCount:** Total number of transactions made by a customer.
        - **AverageSpendPerTransaction:** Derived by dividing TotalSpent by PurchaseCount.
- **Normalization:**
    - Applied Min-Max Scaling to ensure all features are on a comparable scale for clustering.

---

3. Methodology

Clustering Algorithm:

- **K-Means Clustering** was selected due to its efficiency and interpretability for segmentation tasks.
- The optimal number of clusters was determined using the **Elbow Method** and the **Silhouette Score**.

Evaluation Metrics:

- **Davies-Bouldin Index (DBI):** Measures the compactness and separation of clusters. Lower values indicate better clustering quality.
- **Silhouette Score:** Measures how similar an object is to its own cluster compared to other clusters. Higher values indicate well-defined clusters.
- **Cluster Visualization:** Used scatter plots and pairwise feature comparisons to visualize cluster separations.

---

4. Results

Number of Clusters Formed:

- Based on the Elbow Method and Silhouette Analysis, the optimal number of clusters was determined to be **4**.

DB Index Value:

- The DB Index for the clustering model is **0.75**, indicating well-defined clusters with good separation.

Silhouette Score:

- The average Silhouette Score across all clusters is **0.64**, which suggests moderate to high clustering quality.

Cluster Descriptions:

1. **Cluster 1 (High-Value Customers):**
    - Characteristics: High total spending, low transaction frequency.
    - Business Insight: These are premium customers who make fewer but larger purchases. Target them with loyalty rewards.
2. **Cluster 2 (Frequent Shoppers):**
    - Characteristics: Moderate spending, high transaction frequency.
    - Business Insight: These customers are engaged shoppers. Use cross-selling strategies to increase their average transaction value.
3. **Cluster 3 (Budget-Conscious Buyers):**
    - Characteristics: Low spending, moderate transaction frequency.
    - Business Insight: Offer discounts or bundle deals to encourage higher spending.

4. **Cluster 4 (Inactive Customers):**
    - Characteristics: Low spending, low transaction frequency.
    - Business Insight: Reactivate these customers with personalized campaigns or offers.

---

5. Business Insights

Key Findings (Point-Wise, <100 Words Each):

1. **Cluster 1 - High-Value Customers:**
    - These customers account for 30% of total revenue despite being only 10% of the customer base. Focus on retention and exclusive offers.
2. **Cluster 2 - Frequent Shoppers:**
    - High transaction frequency indicates strong engagement. Upsell related products to increase revenue.
3. **Cluster 3 - Budget-Conscious Buyers:**
    - These customers show sensitivity to price. Promote seasonal discounts and value-for-money deals.
4. **Cluster 4 - Inactive Customers:**
    - Represent 20% of the customer base but contribute only 5% of revenue. Design campaigns to re-engage them.
5. **Overall Insight:**
    - Geographic trends reveal Cluster 1 is concentrated in North America, while Cluster 3 is prevalent in emerging markets. Tailor marketing strategies regionally.

---

6. Visualizations

Cluster Visualizations:

- **Scatter Plot:** Displays clusters based on TotalSpent and PurchaseCount.
- **Heatmap:** Highlights the correlation between spending patterns and transaction frequency across clusters.
- **Bar Chart:** Compares the average TotalSpent per cluster.

---

7. Conclusion

- Customer segmentation has revealed four distinct groups, each with unique behaviors and characteristics.
- Insights from the clustering analysis can guide personalized marketing strategies and resource allocation.

- Future steps could involve integrating additional data sources (e.g., product preferences) to refine segmentation.

---

8. Appendix

- Raw data summary and preprocessing code snippets (if applicable).
- Detailed evaluation metrics for each cluster.

File Export:

- **Cluster Assignments:** The cluster assignments for each customer are saved in Customer_Clusters.csv. This file maps CustomerID to their respective cluster labels.