Customer Segmentation Clustering Report

Objective

The primary objective of this analysis was to segment customers based on various features related to their activity and profile, including region, signup age, total spending, transaction history, and recency. By employing KMeans clustering, we aimed to group customers into clusters with similar behaviors, facilitating tailored marketing strategies, product recommendations, and retention efforts.

Clustering Methodology

1. Preprocessing

- **Categorical Encoding**: The customer region was numerically encoded to ensure compatibility with the clustering algorithm.
- **Date Conversion**: The signup date was transformed into **SignupAge**, representing the number of days since signup, indicating customer activity duration.
- **Transaction Summary**: Transaction data was aggregated at the customer level, calculating:
    - **Total Spent**: The total amount spent by a customer.
    - **Total Transactions**: The total number of transactions associated with each customer.
    - **Recency**: The number of days since the customer's last transaction, indicating recent activity.

2. Standardization

All features were standardized using **StandardScaler** to ensure each feature had a mean of 0 and a standard deviation of 1. This step was crucial to prevent any single feature from disproportionately influencing the clustering results due to differing scales.

3. Optimal Cluster Selection

- The **Davies-Bouldin Index (DB Index)** and **Silhouette Score** were utilized to evaluate clustering quality for different values of k (number of clusters).
- After assessing clustering performance for k ranging from 2 to 10, **k=4** was selected as the optimal number of clusters based on the lowest DB Index and the highest Silhouette Score.

Clustering Results

1. Number of Clusters Formed

The clustering algorithm successfully formed **4 clusters**.

2. DB Index Value

The **Davies-Bouldin (DB) Index** for **k = 4** is **1.35299345470687**. A lower DB index value indicates that the clusters are well-separated and compact. The value suggests moderate separation between clusters in this case.

3. Cluster Summary

The following table summarizes the cluster centroids based on key metrics:

Cluster	Avg Total Spent	Avg Total Transactions	Avg Recency	Avg Signup Age
0	8.087	8.087	80.087	615.83
1	2.050	2.050	264.800	578.15
2	4.563	4.563	93.859	509.17
3	4.242	4.242	82.274	564.95

4. Additional Clustering Metrics

- **Compactness and Separation**: Based on the DB index, the clusters exhibit moderate compactness and separation, indicating some overlap but not excessive, which is typical for clustering with this dataset.
- **Cluster Characteristics**:
    - **Cluster 0**: Highest average total spent and transactions, moderate recency, and the longest signup age.
    - **Cluster 1**: Low values for spending, transactions, and recency, with a slightly lower signup age, indicating potential churn.
    - **Cluster 2**: Moderate values across all metrics, with a relatively high recency, suggesting recent engagement.
    - **Cluster 3**: Similar to Cluster 2, with slightly higher spending than Cluster 1 and lower recency.

Insights and Recommendations

Customer Engagement Strategies

- **Cluster 0 (Moderate Activity, Long Tenure)**: Focus on retention programs or loyalty schemes to maintain engagement.
- **Cluster 1 (Low Activity, Long Dormancy)**: Target with re-engagement campaigns, such as special promotions or personalized offers, to rekindle interest.
- **Clusters 2 and 3 (Moderate Activity, Recent Engagement)**: Implement strategies like cross-selling or up-selling to enhance customer lifetime value.

Future Improvements

- Consider refining the current clustering by exploring advanced techniques (e.g., hierarchical clustering) or incorporating additional features like customer preferences.

- The DB Index and Silhouette Score suggest that further refinement could help achieve better separation between clusters, reducing overlap.

Conclusion

By targeting the specific needs of each cluster, the organization can improve the overall customer experience and maximize retention and engagement. The insights derived from this clustering analysis provide a foundation for developing tailored marketing strategies and enhancing customer relationships.