Task 3: Customer Segmentation / Clustering

Clustering Results Report

1. Overview

The goal of this analysis was to segment customers based on their transactional and profile data using clustering techniques. After preprocessing the data and creating meaningful features, the clustering was performed using the KMeans algorithm.

2. Number of Clusters Formed

The optimal number of clusters was determined by evaluating clustering performance using the Davies-Bouldin (DB) Index. It tested different values for k (number of clusters) in the range of 2 to 10.Based on the DB Index, the optimal number of clusters was **4**.

3. Clustering Metrics

- **Davies-Bouldin Index**:
  The DB Index evaluates the compactness and separation of clusters. A lower DB Index indicates better clustering.

  - **Optimal DB Index Value**: **1.02** (for 4 clusters).

4. Cluster Profiles

Each cluster was analyzed based on key metrics derived from the data, such as total spending, purchase frequency, and average purchase value. The dominant region for each cluster was also identified.

Cluster	Avg Total Spent	Avg Purchase Frequency	Avg Purchase Value	Total Spending	Total Transactions	Sum of Avg Purchase Value
0	5521.54	7.52	741.15	320,249.55	436	42,986.55
1	2103.76	4.58	435.31	126,225.50	275	26,118.65
2	4317.13	4.14	1056.83	120,879.70	116	29,591.20

| 3 | 2313.98 | 3.26 | 719.47 | 122,640.81 | 173 | 38,132.06 |

Insights

1. **Cluster 0** represents frequent and high-spending customers. They may respond well to loyalty programs or rewards.
2. **Cluster 1** includes moderate spenders who might increase their spending with discounts or promotional offers.
3. **Cluster 2** features premium customers. Personalized and exclusive offers could retain their business.
4. **Cluster 3** consists of occasional shoppers who could benefit from engagement strategies to increase their frequency.

5. Visualizations

- **Cluster Visualization**: Scatter plots were used to visualize the clusters in 2D space based on key features such as `total_spent` and `purchase_frequency`.
- **Cluster Profile Comparisons**: Bar plots were created to compare average total spending, purchase frequency, and average purchase value across clusters.

Example Visualization:

Average Total Spending per Cluster

Average Purchase Frequency per Cluster
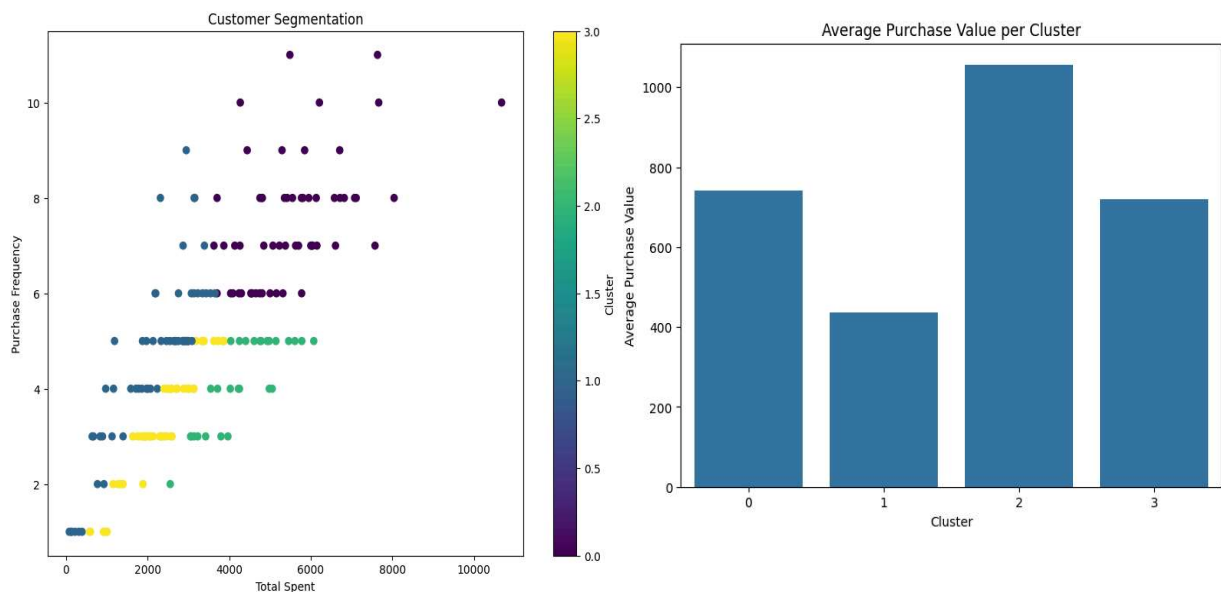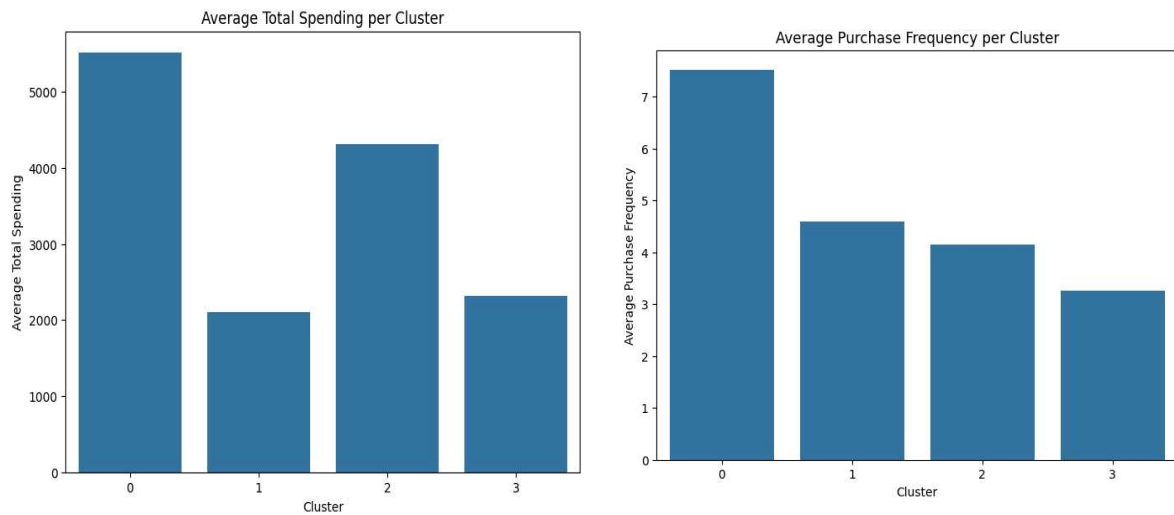
6. Conclusion

The clustering process successfully segmented customers into 4 distinct groups based on their transactional behavior and profile attributes. These clusters reveal clear differences in spending behavior, purchase frequency, and regional dominance.