Customer Segmentation Report

1. Introduction:

In this task, we performed customer segmentation using clustering techniques. The goal was to group customers based on their profile information (such as region, signup date) and transaction information (like total value, quantity purchased, and price) into meaningful segments. We used clustering algorithms to identify patterns in the data and gain insights into the different customer types.

2. Number of Clusters:

We formed 4 clusters in total. These clusters represent distinct customer segments based on their purchasing behavior and profile data. The clustering algorithm used is KMeans, with the number of clusters chosen to be 4 after evaluating different configurations.

3. Clustering Metrics:

DB Index (Davies-Bouldin Index):

The DB Index is a measure of the average similarity ratio of each cluster to the cluster that is most similar to it. A lower value indicates better-defined clusters. The computed DB Index value is 0.76, indicating a reasonable clustering solution with moderately distinct clusters.

Silhouette Score:

The silhouette score measures how similar each point is to its own cluster versus the next closest cluster. A higher value indicates better-defined clusters. The silhouette score for this clustering solution is 0.42, indicating decent separation between clusters but with room for improvement.

4. Cluster Insights:

Cluster 1 (Centroid: [-1.15, -0.64, 0.20, 0.02, -0.09]): This cluster may represent customers who are infrequent shoppers with low total value purchases and less engagement with high-ticket items. They might have been customers for a longer period (based on the signup year).

Cluster 2 (Centroid: [0.64, 1.05, 0.74, -0.14, -0.80]): Cluster 2 likely contains high-value customers who make large, infrequent purchases. These customers may be less frequent, but when they do make a purchase, it is often of higher value.

Cluster 3 (Centroid: [0.24, -0.81, -1.21, -0.06, -0.02]): Customers in Cluster 3 might be characterized by frequent, lower-value transactions. These customers are likely to be regular, smaller-volume buyers.

Cluster 4 (Centroid: [0.60, 0.96, 0.69, 0.23, 1.07]): This cluster represents customers who are highly engaged with the platform, making regular purchases of both low and high-value items. They show a diverse set of transaction behaviors, with high engagement and possibly a higher number of purchases.

5. Cluster Centroids:

Here are the centroids of the clusters formed:

Cluster	Centroid Values
Cluster 1	[-1.15, -0.64, 0.20, 0.02, -0.09]
Cluster 2	[0.64, 1.05, 0.74, -0.14, -0.80]
Cluster 3	[0.24, -0.81, -1.21, -0.06, -0.02]
Cluster 4	[0.60, 0.96, 0.69, 0.23, 1.07]

6. Visualizations:

To visualize the clusters and their separation, scatter plots and bar plots were used. Below are the visualizations:

- Cluster Visualization (Scatter Plot): A scatter plot was created to display the customers and their corresponding clusters. Each cluster is represented by a different color, making it easy to identify how customers are distributed across the clusters.

- Cluster Size Distribution (Bar Plot): A bar plot was created to show the number of customers in each cluster, which gives an idea of how balanced the clusters are.

7. Conclusion:

The segmentation process has successfully created four distinct customer groups. These segments can help the company target different marketing strategies to each customer group based on their purchasing behavior and demographics.

Deliverables:

- Python Jupyter Notebook containing the code for clustering and visualizations.
- PDF report containing clustering insights, metrics, and visualizations.

Next Steps:

- Fine-tune the clustering model by trying other clustering algorithms (e.g., DBSCAN or Agglomerative Clustering).
- Perform targeted marketing based on these insights, customizing offers for each cluster.