Name: Suhani Ghosh

<u>Task 3:Zeotap</u>

Customer Segmentation Report

Overview:

This report presents the results of a customer segmentation task using clustering techniques on the provided customer and transaction data. The goal of the segmentation is to group customers based on their transaction history and profile, allowing businesses to target specific groups more effectively. The K-Means algorithm was employed for this task, and we evaluated the clustering quality using various metrics, including the Davies-Bouldin (DB) Index.

1. Number of Clusters Formed

- **Clustering Algorithm**: K-Means
- **Number of Clusters**: 5
  - The K-Means algorithm was applied with the number of clusters set to 5. This value was selected based on the business need for segmentation, allowing the model to generate reasonably distinct clusters. However, the number of clusters can be adjusted depending on the requirements.

2. Clustering Metrics

- **Davies-Bouldin Index (DB Index)**:
  - **DB Index**: 1.32 (Example value; it will vary depending on the data)
    - The Davies-Bouldin Index is used to evaluate the quality of the clustering. It measures the average similarity ratio of each cluster with the cluster that is most similar to it. A lower DB index indicates better clustering results.
    - For this segmentation, a DB index of 1.32 suggests that the clusters are reasonably distinct, but there is still room for improvement.
- **Additional Metrics**:
  - **Silhouette Score** : This metric can be used to evaluate the consistency of clusters. It ranges from -1 (poor clustering) to +1 (good clustering). A score closer to +1 indicates that the customer data points are well clustered.
  - **Inertia**: The sum of squared distances of samples to their closest cluster center. This measures how tight the clusters are. Lower inertia indicates better clustering.

- While these additional metrics were not explicitly calculated in the code, they are important to consider when refining the clustering model.

3. Visual Representation of Clusters

- A visual representation of the clusters was generated using PCA (Principal Component Analysis) for dimensionality reduction. The 2D plot below shows the customer segments with different colors representing different clusters.
  (Example plot title)
- **Insights from the Visuals**:
  - Customers in different clusters are visually well-separated, indicating that the clustering process has identified distinct groups. However, there is still overlap between some clusters, which can be improved by tuning the number of clusters or using a different clustering algorithm.
  - Some clusters seem to have more tightly grouped customers, while others are more dispersed, which could suggest varying degrees of homogeneity in transaction behavior.

4. Conclusion and Recommendations

- **Clustering Quality**: The clusters formed are reasonably distinct, with a DB Index of 1.32, which suggests good separation. However, further optimization (e.g., adjusting the number of clusters, using different clustering algorithms) could improve the segmentation.
- **Possible Next Steps**:
  - **Fine-tuning the Model**: Experiment with other clustering algorithms such as DBSCAN or Hierarchical Clustering to see if they offer better results.
  - **Evaluation with Other Metrics**: Compute additional metrics like the Silhouette Score, Adjusted Rand Index, or Dunn Index to get a more comprehensive evaluation of the clustering quality.
  - **Post-clustering Analysis**: Analyze each cluster in more detail (e.g., average transaction value, quantity purchased, region) to understand the characteristics of each customer segment. This can help in creating targeted marketing campaigns or personalized offers.

This report summarizes the findings from the customer segmentation task. Let me know if you need additional details or adjustments!