Report on Customer Segmentation Task

1. **Objective**:
   The objective of this task was to perform customer segmentation using clustering techniques, leveraging both customer profile information and transaction data. The goal was to identify distinct customer groups for targeted marketing and personalized services.
2. **Data Preparation**:
   - **Data Merging**: Customer data from the **`Customers.csv`** file was merged with transaction data from **`Transactions.csv`** to create a unified dataset containing both customer profile and transaction information.
   - **Feature Selection**: Key features like **`Age`, `Income`, `TotalSpend`,** and transaction-based attributes (e.g., frequency of transactions) were selected for clustering.
   - **Handling Non-Numeric Data**: Categorical data such as **`Gender`** and **`Region`** was encoded into numerical values using Label Encoding to make them suitable for clustering algorithms.
3. **Clustering Algorithm**:
   - **K-Means Clustering** was chosen for segmentation as it is efficient and widely used for customer segmentation tasks. The number of clusters was selected between 2 and 10 using the Elbow Method, which helps determine the optimal number of clusters by minimizing within-cluster variance.
4. **Clustering Process**:
   - The features were standardized using **StandardScaler** to ensure that all variables contribute equally to the clustering model.
   - **K-Means** was applied, and the dataset was clustered into 4 distinct groups based on customer characteristics and transaction patterns.
5. **Evaluation Metrics**:
   - **DB Index** was used to evaluate the clustering performance. A lower DB index indicates better-defined clusters, helping assess the quality of segmentation.
   - Additional metrics such as silhouette score and inertia were also calculated for validation.
6. **Visualization**:
   - **PCA** was used to reduce the dimensionality of the features to 2D, and the clusters were visualized using scatter plots for clear representation.
   - The clusters were color-coded, allowing for easy identification of the different customer segments.
7. **Conclusion**:
   The segmentation revealed meaningful patterns within customer data, allowing businesses to target each cluster with tailored strategies.