Task-3: Customer Segmentation / Clustering

Customer Profile and Transaction Aggregation:

- The dataset was enriched by extracting additional features like 'SignupYear' and 'SignupMonth' from the customers' signup date, helping to identify customer cohorts based on their registration timing.
- Transaction data was aggregated per customer to calculate metrics such as total spending, number of transactions, average spending, and average quantity purchased.
- The combined dataset included one-hot encoding for the 'Region' feature and filled missing transaction data with zeros, ensuring completeness for analysis.

Clustering & Feature Scaling:

- Customer segmentation was performed using K-Means clustering with 5 clusters, based on the scaled customer profile and transaction data.
- The scaled features were standardized to ensure that all variables contribute equally to the clustering process.
- Clustering evaluation metrics, such as the Davies-Bouldin index (DBI) and Silhouette score, were calculated. A lower DBI and higher Silhouette score indicate good cluster quality, providing a sound segmentation basis.

Cluster Visualizations:

- The clusters were visualized using PCA (Principal Component Analysis), reducing the data's dimensionality to 2 components. This scatter plot highlighted customer grouping based on similar spending behaviours and demographics.

- The cluster distribution showed a balanced segmentation, allowing for targeted marketing and strategy development.

Cluster Insights and Analysis:

- Summary statistics were calculated for each cluster, showcasing differences in spending behaviour, transaction frequency, and product engagement across segments.
- Visualizations like box plots and bar plots of total spending and transaction counts revealed significant differences between clusters, offering actionable insights for personalized customer strategies.