INTRODUCTION

Overview of the Task : This report presents the results of the customer segmentation analysis performed using clustering techniques. The objective was to group customers based on their transactional behavior and profile information to uncover insights that could drive business strategies.

Data Used : The analysis leverages customer profile data (from Customers.csv) and transaction data (from Transactions.csv), which were merged to create aggregated customer transaction data.

CLUSTERING ALGORITHMS

	KMeans Clustering	Hierarchical Clustering	Gaussian Mixture Model
No. of Clusters	4	4	2
Davies-Bouldin Index (DB Index)	0.935	1.031	1.227
Silhouette Score	0.328	0.296	0.275

EVALUATION OF CLUSTERING QUALITY

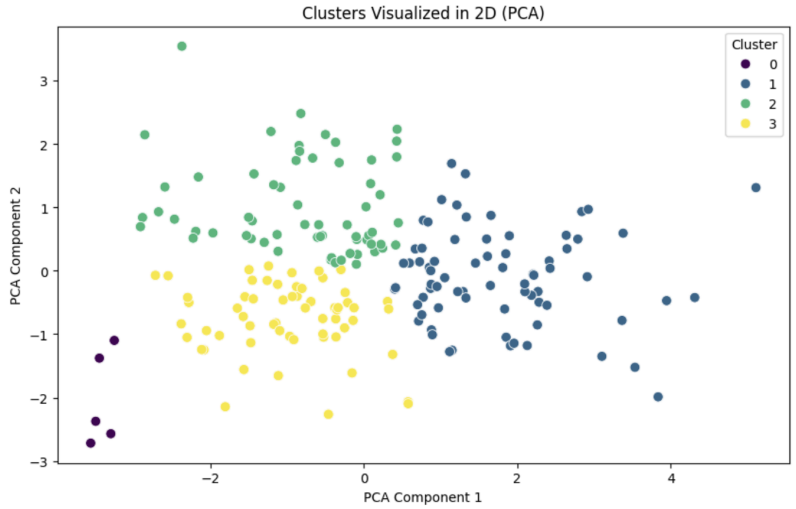
Davies-Bouldin Index (DB Index) : A **lower DB Index** indicates better clustering. ***KMeans*** performs the best with the lowest DB Index of 0.935, suggesting that the clusters formed are **relatively compact and well-separated**.

Silhouette Score : A **higher silhouette score** indicates better-defined clusters. ***KMeans*** also has the highest silhouette score of 0.328, but all models show moderate clustering quality, implying that the **data may contain overlapping clusters**(real-world data).

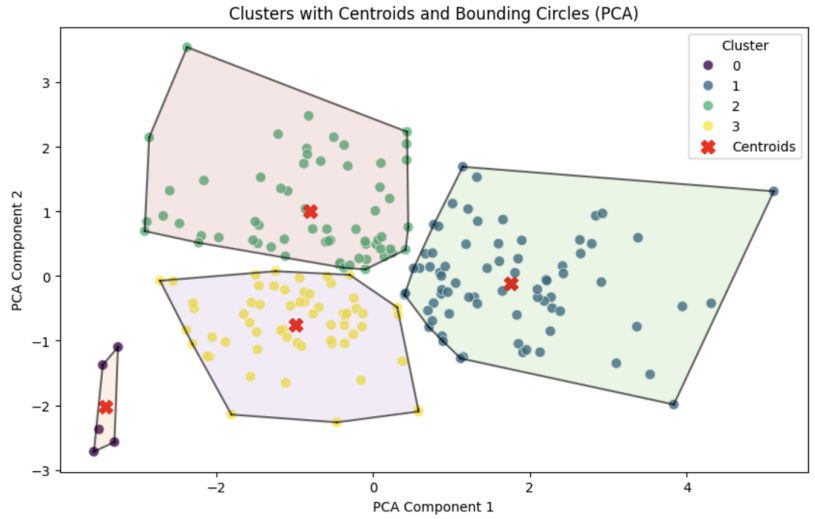
Therefore,
KMeans perform the best on the dataset with k=4

VISUALIZATION

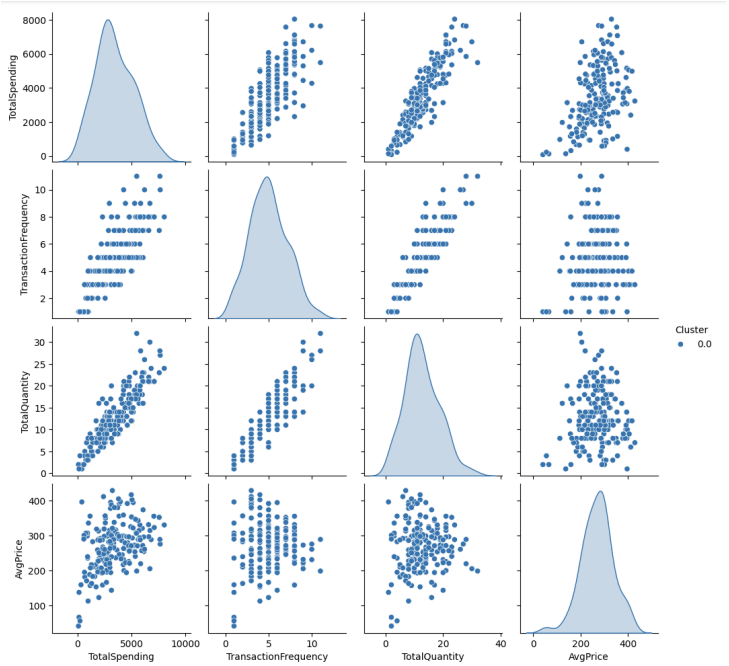
PCA-based 2D Visualization of KMeans Clusters : A scatter plot showing the 2D projection of the customer data (after applying PCA for dimensionality reduction) with the clusters visualized by different colors.



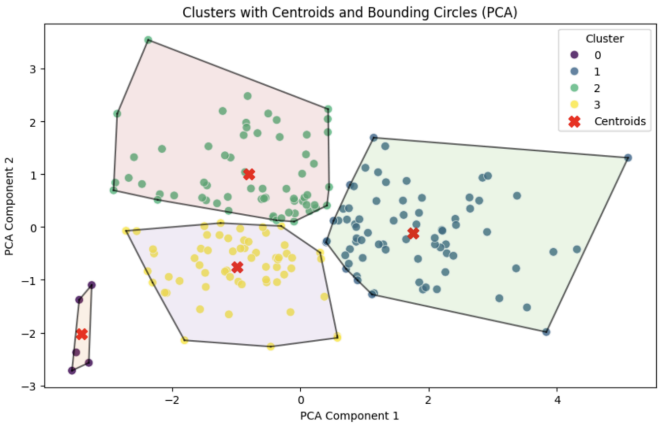
Clusters with Centroids and Bounding Circles : A plot showing the clusters with the centroids marked in red. Bounding circles represent the convex hulls that encapsulate each cluster, highlighting their shape and size.



Pairplot of Clusters : A pairwise scatterplot matrix of the features used for clustering (e.g., total spending, transaction frequency, total quantity, average price). Each plot in the matrix shows how the clusters are separated based on the features.



BUSINESS & ACTIONABLE INSIGHTS



- **Cluster 0**: Represent high-spending, frequent buyers.
- **Cluster 1**: Represent low-spending, occasional buyers.
- **Cluster 2**: Represent customers who buy in high quantities but at lower prices.
- **Cluster 3**: Represent customers with average spending but purchasing more expensive products.

High-value customers (e.g., Cluster 0) could be targeted with loyalty programs or premium offerings.

Occasional or low-value customers (e.g., Cluster 1) might benefit from promotional offers or discounts to increase transaction frequency.

High-quantity customers (e.g., Cluster 2) could be incentivized with bulk purchase discounts or bundle deals.